Metadata Extraction from Tables and Charts in Scientific Publications: A Systematic Literature Review

Journal Title
XX(X):1–36

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Erick Cedeño[1], Daniel Garijo[1] and Oscar Corcho[1]

Abstract

Extracting metadata from tables and charts in scientific publications is essential for enabling structured knowledge representation, automated retrieval of experimental results, and large-scale evidence synthesis. Such metadata includes descriptive information that goes beyond structural parsing and captures the roles and relationships of elements such as headers, units, variables, legends, and axis labels. Although numerous methods have been proposed for table extraction and chart understanding, prior work remains fragmented, meaning that most approaches are developed and evaluated independently for a single modality and therefore lack mechanisms for connecting information across tables and charts. This specific approach to each modality has led to heterogeneous processes and inconsistent assessment practices, limiting comparability across studies. In this systematic literature review, we analyze 68 peer-reviewed studies using a unified evaluation framework specifically designed to examine metadata extraction capabilities in both tables and charts. Guided by explicit research questions, we compare these systems in terms of their task definitions, model architectures, metadata outputs, and reporting practices. Rather than reproducing implementations, our analysis evaluates the extent to which each method supports metadata identification, variable interpretation, and multimodal alignment. The findings highlight unresolved challenges in linking related information across modalities (e.g., associating table headers with chart axes), interpreting variables beyond their superficial textual labels, and establishing standardized benchmarks that measure correctness at the metadata level.

Keywords

Metadata Extraction, Document Retrieval, LLM, Document Understanding, Benchmark, Dataset, Multimodal.

Introduction

As the volume of scientific publications continues to grow at an unprecedented rate, the ability to automatically extract meaningful information from such documents has become essential for information retrieval, natural language processing (NLP), machine learning, and digital library systems Bornmann et al. (2021); Dagdelen et al. (2024); Jonnalagadda et al. (2015). Scientific articles frequently include rich visual and tabular content: tables and charts that encode key elements of experimental design, variable definitions, measurement units, comparative analyses, and numerical results. Unlike narrative text, which can be handled using established NLP pipelines, tables and charts encode information through structured layouts and visual encodings that require dedicated processing methods Desai et al. (2021); Zheng et al. (2021); Ma et al. (2021); Ramos et al. (2020).

Metadata extraction refers to the process of identifying, classifying, and organizing the components that define the structure and meaning of tables and charts (e.g., headers, units, axes, legends, labels). In scientific publications, this process spans two primary modalities. For tables, it involves tasks such as table detection, structure recognition (rows, columns, multi-span headers), cell classification, and content labeling Zheng et al. (2021). For charts, it includes detecting visual marks (bars, lines, points), mapping them to axis values, and reconstructing the underlying data represented in the visual encoding Shivasankaran et al. (2023); Ma et al. (2021). These tasks support downstream applications such as question answering, summarization, information retrieval, reproducible science, and multimodal document understanding Masry et al. (2022); Kantharaj et al. (2022).

Despite sustained progress in both areas, current approaches remain largely disconnected from one another. That is, methods designed for table extraction rarely interact with systems for chart understanding, even though tables and charts frequently describe overlapping aspects of the same dataset in scientific documents. For example, a table may list specific numerical values while a chart summarizes their distribution or trends. Without mechanisms to align and jointly interpret these modalities, systems fail to construct coherent representations of the underlying experimental information. This limitation has direct implications for knowledge graph construction from scientific publications, where capturing complete and consistent factual knowledge requires integrating information expressed across multiple modalities Kabongo et al. (2024). When table and chart derived metadata are processed in isolation, knowledge graphs risk containing fragmented, duplicated, or partially

[01]Ontology Engineering Group, Universidad Politécnica de Madrid, Madrid, Spain
0
Corresponding author:
Erick Cedeño, erick.cedeno@upm.es

inconsistent representations of the same experimental entities, measurements, or relationships. Furthermore, although benchmarks exist for table tasks (e.g., PubTables-1M Smock et al. (2022), TabLeX Desai et al. (2021)) and chart tasks (e.g., ChartInfo Ma et al. (2021), UnchartIt Ramos et al. (2020)), there is still no standardized evaluation protocol or shared dataset that assesses both modalities together under a consistent methodology. This lack of multimodal benchmarks further limits the development and evaluation of extraction pipelines capable of producing unified representations of scientific knowledge suitable for knowledge graph construction.

Existing surveys have examined isolated components of this problem, often focusing exclusively on table structure recognition Shah and Patel (2004) or chart image analysis Huang et al. (2025). However, these reviews generally do not address the complete metadata extraction workflow: element detection, structure recognition, content labeling, and data reconstruction, nor do they consider how information extracted from tables and charts can reinforce each other within the same publication. As a result, the multimodal dimension of scientific metadata extraction remains insufficiently explored.

This work addresses these limitations by conducting a systematic literature review of metadata extraction methods for tables and charts in scientific publications. The primary scope of the review is document-level extraction from tables and charts embedded in scientific publications, including PDF tables, chart images, and visually structured artifacts. Semantic Web approaches for semantic table interpretation, ontology alignment, and knowledge graph population are considered related downstream methods when they assume that the tabular or graphical content has already been extracted into a structured representation. This distinction allows us to focus the primary corpus on extraction and reconstruction systems, while still analyzing how their outputs can support semantic integration.

Our work offers a comprehensive and comparative perspective that evaluates existing approaches to metadata and content extraction from tables and charts considering not only task performance in areas such as table detection, structure recognition, and chart data reconstruction, but also their ability to support cross-modal reasoning and integrated document understanding.

To support this contribution, we analyzed 68 peer-reviewed publications by adapting the research question approach proposed by Kitchenham et al. (2009) into an evaluation framework for this review. This framework enables a comparison of current systems, accounting for the type and depth of extracted information, the underlying model architectures (e.g., CNNs, Transformers, and hybrid designs), the availability of datasets and open-source tools, and the extent to which each system facilitates multimodal document analysis.

The key contributions of this systematic literature review are summarized as follows:

- We present a comprehensive systematic literature review that jointly analyzes table and chart modalities, bridging two research areas that have traditionally evolved independently within the document analysis community.
- We introduce a structured evaluation framework guided by explicit research questions and grounded in established SLR methodology, enabling a reproducible and multidimensional comparison of existing systems.
- We provide a detailed and annotated review of 68 peer-reviewed publications, identifying methodological trends, recurring limitations, and understudied aspects such as variable interpretation, multimodal linking, and alignment between table and chart representations.
- Based on this comparative analysis, we outline future research directions toward the development of explainable, benchmark-driven, and interoperable frameworks for multimodal data extraction in the scientific literature.

The remainder of the paper is structured as follows. The Background section introduces the foundational concepts and definitions related to metadata extraction from tables and charts. The Methodology section presents the systematic literature review methodology, including the research questions, paper selection strategy, and the design of the evaluation framework. The Results section provides an aggregated comparison of the 68 reviewed systems across multiple analytical dimensions. The RQ-Based Literature Review and Analysis section offers a detailed analysis organized according to the research questions. The Discussion section discusses the broader implications of the findings and highlights open challenges for future research. Finally, the Conclusion section summarizes the main insights of the review and outlines promising directions for advancing multimodal metadata extraction.

Background

Understanding metadata extraction from tables and charts requires clarifying the types of information these visual and tabular elements encode and how this information is represented in scientific publications. This section provides context on metadata for scientific publications, outlines the structural and visual characteristics of tables and charts, and identifies the main technical challenges faced by automated metadata extraction systems, such as handling irregular layouts, interpreting visual encodings, and establishing meaningful relationships between textual and graphical elements.

Metadata in Scientific Publications

In this review, we define *metadata* as structured descriptive information that identifies, contextualizes, and relates the components of tables and charts in scientific publications Zheng et al. (2021); Masry et al. (2022). Unlike bibliographic metadata, such as title, authors, venue, or keywords, the focus of this study is on structural, semantic, and contextual metadata embedded in visual and tabular artifacts. This includes elements such as table headers, row and column labels, cell roles, units, chart axes, legends, tick labels, plotted variables, visual marks, and the relationships that connect these elements to numerical values.

This definition distinguishes metadata from raw extracted content. For example, a numerical value extracted from a table cell or a bar height reconstructed from a chart is considered data, whereas the header, unit, variable name, axis label, or legend entry that explains what the value represents is considered metadata. When systems explicitly associate such descriptors with values, they support metadata-level interpretation rather than only structural extraction.

Accurate extraction of such metadata is essential for enabling structured representations of scientific evidence. For example, correctly interpreting a table requires not only detecting its grid structure but also identifying how header cells define the meaning of corresponding data entries. Similarly, in a chart, the position or height of a bar can only be understood when mapped to its respective axis value and legend category. Such structural relationships often encode key information that is not explicitly described in the accompanying text.

A practical illustration appears frequently in scientific publications. For example, UnchartIt Ramos et al. (2020) show cases in which a bar chart visualizes the same numerical results previously reported in a table within the same article. In these settings, the chart encodes aggregated values through bar heights, while the corresponding table lists the exact numeric entries. Correctly aligning these two representations requires identifying shared metadata elements such as variable names, units, or axis labels and determining how the chart's visual marks correspond to specific rows or columns in the table.

A similar situation is illustrated in the study of Ma et al. (2021), where several scientific publications contain a table reporting detailed measurements while a related chart summarizes the same variables visually using bars or lines. In these examples, understanding the connection between modalities depends on recognizing consistent metadata across representations, demonstrating why structural and contextual metadata are essential for multimodal interpretation in scientific publications.

Table and Chart Structures

Tables and charts constitute two structurally distinct yet complementary modalities within scientific publications. Both play a central role in communicating quantitative information but differ in how such information is organized and visually encoded.

Tables follow a grid-based layout composed of rows and columns, often including multi-span headers, merged cells, nested structures, and hierarchical relationships. Structural recognition typically involves detecting table boundaries, segmenting cells, and assigning logical roles (e.g., header, stub, and data regions). Because table formats and layouts vary considerably across publishers and domains, building generalizable extraction models remains a significant challenge Desai et al. (2021); Shah and Patel (2004).

Charts, in contrast, refer to data-driven visualizations such as bar, line, scatter, or pie plots that encode structured or numerical information. Other visual elements such as conceptual diagrams, flowcharts, or illustrations fall outside the scope of this study, as they do not convey measurable or structured data. Understanding charts involves identifying primitive visual components (bars, lines, points) and mapping them to their corresponding variables, axis values, and legend entries Ma et al. (2021). Unlike tables, which present values explicitly, charts express information implicitly through position, length, color, or area, requiring higher-level reasoning to reconstruct the underlying data.

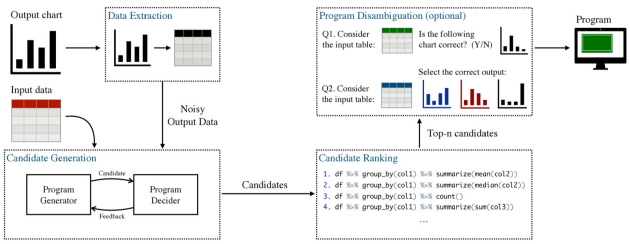

Figure 1. Architecture of UnchartIt Ramos et al. (2020), illustrating the alignment between tabular input data and chart representations. The pipeline combines visual extraction, candidate generation, and program ranking to reconstruct structured data from charts based on shared metadata such as variables, aggregation operations, and axis roles.

A practical illustration of this appears in several papers within our reviewed corpus, for example, ChartSense Jung et al. (2017) examines scientific publications in which charts visually summarize numerical results that are implicitly represented as structured data. In these cases, the chart encodes quantities through visual attributes such as bar height, line position, or point coordinates, while the underlying values can be reconstructed as tables to support data extraction and analysis. Similarly, LineEX Shivasankaran et al. (2023) analyzes scientific publications where line charts plot continuous measurements that correspond directly to tabular records reported in adjacent sections. In both situations, correctly interpreting the chart requires identifying shared elements such as variables, axis labels, or units and determining how visual marks correspond to specific numerical entries. These examples show why structural and contextual metadata are essential for connecting tabular and chart representations in scientific publications.

Figure 1 illustrates a concrete example of this multimodal relationship, adapted from UnchartIt Ramos et al. (2020). The figure shows how a chart and its corresponding tabular data are jointly processed to reconstruct structured representations of the underlying numerical information. In this architecture, the extraction of visual elements from the chart is explicitly linked to the input table through shared metadata, such as variable identifiers, column roles, and aggregation functions. Candidate programs are generated and ranked based on how well they reconcile visual marks with tabular values, highlighting the central role of structural and contextual metadata in aligning table and chart representations within the same scientific publication. In this context, column roles denote the functional interpretation of table columns (e.g., identifiers, categorical attributes, or numerical measurements), while axis roles describe the function of chart axes in encoding variables and values (e.g., independent or dependent variables).

Methodology

Our methodology adopts the main stages proposed by Kitchenham et al. (2009) for conducting systematic literature reviews, ensuring transparency, reproducibility, and consistency throughout the process of data collection, analysis, and reporting. We detail below how we have followed each of the methodology steps.

Research Questions

We formulated a set of research questions (RQs) designed to capture the breadth of existing methods, tasks, models, and integration strategies across both modalities. Each question is accompanied by an explanation of its purpose within the review to ensure clarity and methodological transparency.

- **RQ1a:** What are the current approaches, tools, and models for data extraction from tables and charts in scientific publications?. Purpose: To provide an overview of the functional capabilities addressed in the literature (e.g., detection, structure recognition, component extraction), revealing which parts of the pipeline have received the most or least attention.
- **RQ1b:** What models and architectures (e.g., CNNs, Transformers, hybrid systems) are used for tabular data extraction?. Purpose: To analyze methodological trends and identify how architectural choices influence performance, generalization, and reproducibility in table understanding.
- **RQ1c:** How do existing systems transform tables from unstructured document formats (e.g., PDFs) into machine-readable representations (such as CSV or JSON, i.e., structured formats that can be directly processed, queried, and reused by computational systems)?. Purpose: To examine end-to-end pipelines for layout detection and structure reconstruction, and to assess how reliably these workflows support downstream data analysis.
- **RQ1d:** What tools and models are used to extract data from charts?. Purpose: To identify the techniques used for chart component detection, text association, and data reconstruction, highlighting differences across chart types (e.g., bar, line, scatter).
- **RQ1e:** How do current methods reconstruct the underlying data represented in charts into structured tabular formats?. Purpose: To understand chart-to-table conversion processes and their relevance for applications such as scientific comparison, meta-analysis, and machine-readable data extraction.
- **RQ1f:** How do existing methods identify and interpret variables within tables and charts?. Purpose: To assess how systems detect column or row headers that define measurement dimensions, or map graphical elements such as bars or points to their corresponding axis values and legend categories.
- **RQ2a:** How do current methods identify and align related information across tables and charts within the same scientific publication?. Purpose: To examine alignment mechanisms (e.g., text matching, embedding similarity, vision–language models) and to assess their robustness in scientific publications.

These research questions are designed to uncover both the technical diversity and the methodological gaps that characterize current systems for table and chart understanding. Addressing them is crucial for three reasons. First, they provide an overview of the evolution of extraction methods, from rule-based and heuristic pipelines to modern deep learning architectures, and how these trends have shaped the reproducibility and scalability of current solutions. In this context, rule-based approaches refer to systems that rely on explicitly defined rules or patterns. For example, locating tables based on border lines, whitespace distribution, or predefined header positions. Heuristic methods use simplified, task-specific decision strategies that approximate reasoning without formal learning (e.g., assigning a detected label to the nearest data region). Such approaches dominated early work in table and chart recognition before the emergence of large annotated datasets and machine learning models Embley et al. (2006); Huang and Tan (2007).

Second, these research questions make explicit how progress in this area depends on dataset availability, evaluation practices, and the integration of multimodal information.

Third, answering these questions enables a clearer identification of missing capabilities, such as multimodal alignment or standardized benchmarking, thereby guiding future research toward more comprehensive, explainable, and better integrated frameworks for metadata extraction from tables and charts.

Review Design

This section describes the methodological design of the systematic review, including the search strategy, inclusion and exclusion criteria, and the selection procedure.

Source Selection and Search Strategy. We queried multiple scholarly databases and academic search engines covering the fields of computer vision, natural language processing, and document analysis. Our goal was to identify publications related to table extraction, chart understanding, and multimodal document processing. The search strategy relied on sources that provide stable metadata, support keyword-based retrieval, and index peer–reviewed literature.

Primary scholarly databases. These platforms were queried directly because they index peer–reviewed journals and conferences and support structured search:

- **Web of Science**[1]: Multidisciplinary citation database indexing peer–reviewed journals and conferences, including research in computer vision, machine learning, document analysis, and related scientific fields.
- **Scopus**[2]: Large multidisciplinary database providing citation data and access to peer–reviewed publications across science, engineering, and the humanities.
- **IEEE Xplore**[3]: Digital library providing access to journals, conference proceedings, and standards,

[1] https://www.webofscience.com
[2] https://www.scopus.com
[3] https://ieeexplore.ieee.org

including research on document analysis, figure extraction, computer vision, and related engineering domains.
- **ACM Digital Library**[4]: Digital library covering peer–reviewed publications in computer science, including work on multimodal learning, document parsing, and information extraction.
- **SpringerLink**[5]: Publisher platform providing access to peer–reviewed journals, books, and conference proceedings across a wide range of scientific disciplines.

To broaden coverage, we also used *Publish or Perish* (PoP)[6], which issues queries to several academic search engines and metadata aggregators. PoP was used to run the same keyword queries across:

- **Google Scholar**: Used to identify additional publications not indexed in curated databases.
- **Semantic Scholar**: Queried for citation-enhanced keyword retrieval.
- **Lens.org**: Used to obtain additional scholarly metadata via PoP.
- **Crossref**: Queried for DOI-based metadata resolution.
- **Scopus**: PoP's Scopus connector was used to replicate and validate results obtained from direct Scopus queries whenever institutional access was available.

Search Queries and Keywords. The search strategy was defined using a fixed set of keyword combinations representing core concepts related to table and chart data extraction from scientific documents.

For bibliographic databases providing advanced query interfaces (e.g., Web of Science, Scopus, IEEE Xplore), these keyword combinations were expressed using explicit logical operators.

For academic search engines accessed via Publish or Perish (e.g., Google Scholar, Semantic Scholar, Lens.org), the same keyword combinations were submitted as quoted keyword phrases and simplified logical expressions, in accordance with the capabilities of each interface. In these cases, logical connectors (e.g., `AND`, `OR`) guided relevance-based retrieval rather than acting as strict filtering constraints.

The final set of keyword combinations used in this study included:

- `"table extraction" AND "chart extraction"`
- `"chart understanding" AND "table understanding"`
- `"figure extraction" OR "table extraction"`
- `"multimodal" AND ("scientific publications" OR "scientific articles") AND ("tables" OR "charts")`
- `"chart data reconstruction" AND "scientific publications"`

Each keyword combination was executed independently across all platforms, with minor syntactic adaptations when required (e.g., quotation handling or operator support). When supported by the platform, results were restricted to peer–reviewed venues. The search covered all publications available up to April 1, 2025.

Inclusion and Exclusion Criteria. To ensure the relevance, quality, and methodological consistency of the selected studies, we defined a set of inclusion (IC) and exclusion (EC) criteria. These criteria were applied across the title, abstract, and full-text screening phases.

Inclusion Criteria (IC). We included studies that met all of the following conditions:

- **IC1:** Present a method, model, or pipeline for extracting structured information from tables, charts, or both.
- **IC2:** Provide a concrete technical contribution relevant to data extraction (e.g., an algorithm, model architecture, system pipeline, or benchmark dataset) that can be analyzed within our study.
- **IC3:** Include sufficient methodological detail and evaluation results to allow meaningful comparison with other approaches.
- **IC4:** Be peer–reviewed journal articles or conference papers.
- **IC5:** Be published or publicly available up to April 1, 2025.

Exclusion Criteria (EC). We excluded studies that met any of the following conditions:

- **EC1:** Focused solely on document layout detection or segmentation without addressing data extraction or interpretation.
- **EC2:** Address only textual metadata (e.g., bibliographic or citation parsing) without involving visual components.
- **EC3:** Addressed non-quantitative or illustrative figures (e.g., conceptual diagrams, flowcharts, or schematic illustrations).
- **EC4:** Works not published in peer-reviewed venues, including theses, abstracts, workshop summaries without full papers, or unpublished preprints without an accepted peer reviewed version.

Scope of Semantic Web and Ontology-based approaches. Because this review focuses on metadata and data extraction from tables and charts in scientific publications, Semantic Web and ontology-based approaches were considered for inclusion in the primary corpus only when they addressed the extraction or reconstruction of information directly from document-level artifacts, such as PDF tables, chart images, scanned pages, or visually structured scientific documents. Approaches that assume already structured inputs, such as CSV files, HTML tables, relational tables, or web tables, and focus primarily on semantic table interpretation, entity linking, column type annotation, or knowledge graph population Liu et al. (2023a); Nguyen et al. (2024); Limaye et al. (2010);

[4] https://dl.acm.org
[5] https://link.springer.com
[6] https://harzing.com/resources/publish-or-perish

Venetis et al. (2011); Ritze et al. (2016) were treated as related downstream work rather than as primary extraction systems. This distinction explains why some Semantic Web approaches are discussed in the review as contextual and integration-oriented contributions, but are not counted among the 68 primary extraction systems.

Selection Procedure. The paper selection process was conducted in four sequential phases. To ensure methodological transparency and reproducibility, we explicitly distinguish between (i) searches performed directly in curated databases (Web of Science, Scopus, IEEE Xplore, ACM DL, SpringerLink) and (ii) supplementary searches executed through Publish or Perish (PoP), which acts as an aggregator and provides unified access to sources such as Google Scholar, Crossref, Semantic Scholar, and Scopus.

1. **Initial retrieval:** Search queries were executed first in curated publisher databases (Web of Science, Scopus, IEEE Xplore, ACM Digital Library, SpringerLink), using their respective advanced search interfaces. To increase coverage, we additionally ran keyword-based searches through Publish or Perish (PoP), which retrieved records from Google Scholar, Crossref, Semantic Scholar, and Scopus. Across all sources, the combined retrieval yielded 8,398 records. After deduplication, 720 duplicate records were removed, leaving 7,678 unique records for screening.
2. **Title review:** Titles were reviewed for relevance to table and chart extraction in scientific publications. This stage excluded 7,113 records that were clearly outside the scope of the review, leaving 565 records for abstract-based eligibility assessment.
3. **Abstract review:** Abstracts were examined to ensure alignment with the focus of the review on structured data extraction from tables and charts in scientific publications. This stage served as the main eligibility filtering step and excluded 497 records according to the exclusion criteria defined above, further narrowing the pool of relevant articles to 68.
4. **Full-text evaluation and final inclusion:** The remaining 68 works were reviewed in full text to confirm their eligibility and to extract the detailed information required by the evaluation framework. No additional papers were excluded at this stage. The final corpus therefore consists of 68 peer-reviewed publications, which form the basis of the comparative evaluation presented in Results.

Figure 2 summarizes the identification, deduplication, title screening, abstract screening, full-text evaluation, and inclusion stages of the review using a PRISMA-style flow diagram. Table 1 reports the number of records excluded at each stage and clarifies the role of full-text evaluation as a final confirmation and annotation step.

Each selection phase was documented explicitly to ensure transparency and reproducibility. All screening decisions (title, abstract, and full-text) were recorded in a shared annotation sheet, and borderline cases were revisited to maintain consistency in the application of inclusion and exclusion criteria. A reproducible record of the retrieved records, filtering decisions, annotation files, and supplementary materials is provided through the GitHub repository[7], which is also archived in Zenodo Cedeño and Garijo (2025).

Reviewer involvement and annotation validation. The screening, selection, and annotation process followed an expert-reviewed consensus procedure. The initial screening and annotation were carried out by one reviewer, who applied the inclusion and exclusion criteria and annotated the selected studies according to the dimensions of the evaluation framework, including modality coverage, task coverage, model family, dataset and benchmark use, evaluation metrics, artifact availability, and level of extracted information. The resulting annotations were then reviewed by a second expert reviewer. This secondary review focused on verifying the consistency of the screening decisions, the application of the inclusion and exclusion criteria, and the assignment of annotation categories. Borderline cases and potentially ambiguous annotations were discussed in joint meetings. These cases mainly concerned modality coverage, multimodal scope, model family assignment, task coverage, and the level of information extracted. Disagreements or uncertainties were resolved through consensus, using the definitions of the evaluation framework and the inclusion and exclusion criteria as reference. The final annotation table available in the supplementary Zenodo repository Cedeño and Garijo (2025) corresponds to the expert-reviewed and reconciled version used for the quantitative analysis reported in this paper.

Evaluation Framework Design. To enable a systematic and comparable analysis of the selected methods, we defined an evaluation framework that organizes the information extracted from each paper into a consistent set of analytical characteristics. The framework supports comparison across technical scope, modelling choices, dataset usage, evaluation practices, and extracted information levels. These characteristics are also used to derive a structured categorization of the reviewed methods based on the patterns observed across the selected studies. This categorization supports comparison beyond task coverage by organizing the systems according to their methodological family, output representation, and degree of semantic or multimodal integration. Each publication was analyzed according to the following characteristics, which are aligned with the research questions defined in Research Questions:

- **Modality Coverage**: Identifies whether the approach targets tables, charts, or both. This directly contributes to **RQ1a** (existing approaches and tools) and **RQ2a** (integration across modalities).
- **Task Coverage**: Specifies which extraction tasks are addressed by the system, such as table detection, table structure recognition, chart component detection, axis and legend extraction, data reconstruction, or table–chart linking. This characteristic is used to analyze how different parts of the extraction pipeline are covered across studies and directly supports **RQ1a–RQ1e**.

[7] https://github.com/erick4556/metadata-extraction-slr

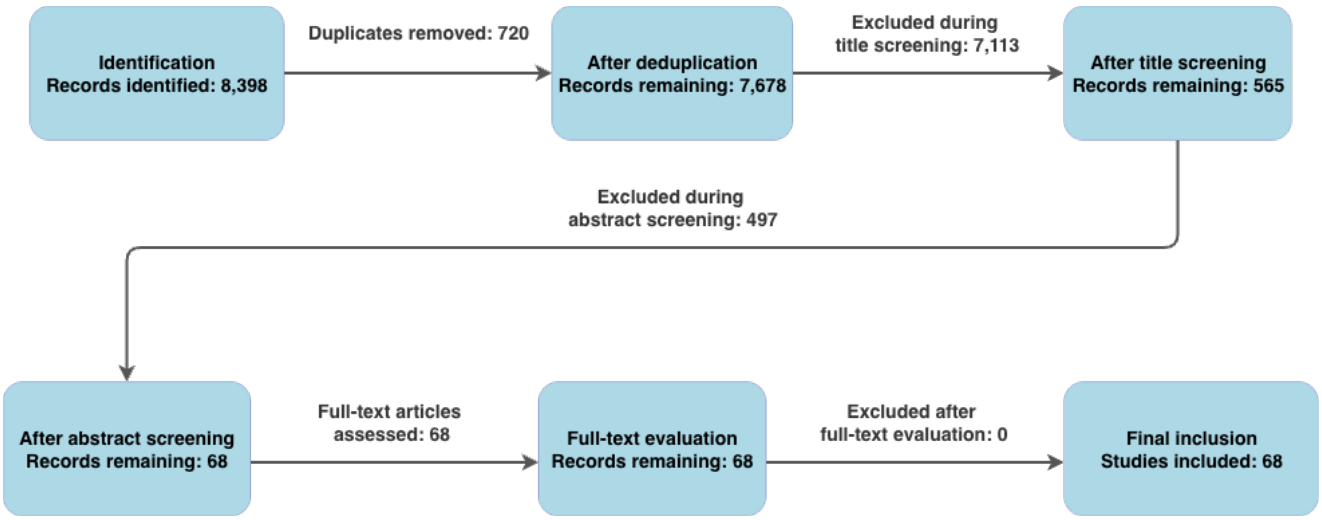

Figure 2. PRISMA-style flow diagram summarizing the identification, deduplication, screening, eligibility assessment, full-text evaluation, and final inclusion stages of the systematic literature review.

Table 1. Records excluded at each stage of the selection process

Selection stage	Basis for exclusion or confirmation	Records excluded
Deduplication	Duplicate records identified across databases and academic search engines.	720
Title screening	Records clearly outside the scope of table or chart data extraction in scientific publications.	7,113
Abstract screening	Records excluded according to the exclusion criteria, including works focused only on general document layout, bibliographic metadata, non-quantitative figures, non-peer-reviewed records, or insufficient methodological and evaluation detail. This stage served as the main eligibility filtering step.	497
Full-text evaluation and final inclusion	The 68 papers retained after abstract-based eligibility assessment were reviewed in full text to confirm eligibility and support detailed annotation according to the evaluation framework. No additional papers were excluded at this stage.	0

Note: The complete selection process is illustrated in Figure 2. Full-text evaluation was used as a final confirmation and detailed annotation stage. The main eligibility filtering was completed during abstract review, which explains why no additional papers were excluded during full-text evaluation.

- **Model Family**: Describes the computational paradigm used by each system, including rule-based or heuristic methods Embley et al. (2006), CNN-based models Paliwal et al. (2019); He et al. (2016), graph-based approaches Gemelli et al. (2022), Transformer-based architectures Vaswani et al. (2017); Smock et al. (2022), multimodal large language models (LLMs) and vision–language models (VLMs) Han et al. (2025); Zhang et al. (2024); Masry et al. (2023), and hybrid pipelines Luo et al. (2021); Zhang et al. (2021). This dimension responds to **RQ1b** and **RQ1d** by enabling comparison across methodological families rather than only across extraction tasks.
- **Dataset and Benchmark Use**: Specifies whether public or proprietary datasets are used and the presence of benchmark standardization. This dimension is linked to **RQ1b** and **RQ1c**, providing insight into the data and evaluation resources supporting each method.
- **Evaluation Metrics and Reported Performance**: Records the evaluation metrics explicitly reported by each study and the level at which performance is assessed. Based on the annotation in the evaluation framework, this includes metrics such as precision, recall, F_1-score, accuracy, intersection-over-union (IoU), numeric error or deviation, and task-specific measures (e.g., cell-level accuracy, axis parsing accuracy, or value reconstruction error). This characteristic also captures whether results are reported at the component level (e.g., table detection, header identification, chart element detection), at the structure level (e.g., row/column reconstruction, chart-to-table conversion), or at the end-to-end system level. This dimension primarily supports **RQ1a**, **RQ1b**, **RQ1c**, **RQ1d**, **RQ1e**, and **RQ1f** by enabling a systematic comparison of how different tasks, models, and pipelines are evaluated. In addition, it indirectly informs **RQ2a** by highlighting the limited and heterogeneous evaluation practices reported for multimodal alignment between tables and charts.
- **Availability of Artifacts**: Indicates whether code, datasets, trained models, or demos are publicly available. This characteristic complements **RQ1a** by assessing practical reproducibility and transparency.

- **Level of Information Extracted**: Describes the level at which information is extracted, distinguishing between: (i) layout-level extraction (e.g., detection of rows, columns, cells, or chart components), (ii) label-level extraction (e.g., headers, axis titles, legends, units), and (iii) value-level association (e.g., explicit linking of labels or variables to numerical values). This characteristic supports **RQ1f** and **RQ2a** by clarifying whether systems move beyond structural detection towards interpretable data representation.

The annotation process of each paper was carried out manually across all dimensions using the information reported in the publication, supplementary materials, and, where available, associated open-source repositories or demo pages. When a criterion was not explicitly reported, it was marked as "Not Specified." This manual curation ensured consistency and traceability across the 68 papers included in the SLR. To support transparency and reproducibility of the review process, the complete annotation table, screening decisions, scripts, data processing steps, and auxiliary materials are provided through the GitHub repository[8], which is also archived in Zenodo Cedeño and Garijo (2025). The repository includes the full spreadsheet used for analysis, documenting all annotated characteristics and screening decisions for each paper, as well as preprocessing routines, aggregation scripts, and instructions to reproduce the quantitative analysis reported in this study.

Results

This section presents the aggregate comparison derived from the methodology used, following the results of our evaluation framework for the 68 annotated papers.

Modality, Dataset, and Content Coverage Summary

In this review, the notion of data extraction covers multiple tasks that vary in scope and across the literature. These tasks include: (i) detection of tables or charts within documents; (ii) reconstruction of their internal structure (e.g., rows, columns, cells, axes, and visual marks), (iii) identification of descriptive elements such as headers, stubs (row or column labels that identify the entities or categories represented in a table), axis labels, legend entries, and measurement units, (iv) reconstruction of the underlying numerical data into structured representations and (v) alignment or linking of related information across tables and charts. Not all reviewed systems address the same subset of these tasks, and many focus on only a limited portion of the extraction pipeline.

Tables and charts remain largely independent research targets, with limited integration across modalities. To characterize these tendencies, Table 2 and Table 3 summarize modality coverage, dataset and tool availability, and the level at which extraction is performed.

Table 2 presents the distribution of the reviewed systems by modality across the final corpus of 68 studies. Most approaches focus on a single modality, with 47.1% dedicated exclusively to charts and 41.2% addressing tables only. Only a smaller subset of works (11.8%) explicitly handles both modalities, confirming that joint modeling of tabular and graphical information remains comparatively less common.

Table 2. Distribution of reviewed systems by modality

Modality	Count	Percentage (%)
Chart only	32	47.1
Table only	28	41.2
Both	8	11.8

Table 3. Availability of public datasets and open-source tools among reviewed systems

Availability aspect	Yes (n)	No (n)
Public dataset available	57	11
Open-source tool released	65	3

Table 3 reports on the availability of datasets and software tools across the reviewed systems. Dataset availability indicates whether a study uses a publicly accessible dataset, introduces a new dataset and releases it publicly, or provides a public link to the dataset it relies on, regardless of whether it is used for training, evaluation, or both. Because many papers either reuse the same dataset for training and evaluation or do not explicitly separate these stages in their experimental setup, Table 3 aggregates dataset availability into a single indicator.

Analysis of the underlying dataset annotations reveals a strong concentration around a limited set of widely reused benchmarks. In particular, chart-focused datasets such as PlotQA Methani et al. (2020), and ChartQA Masry et al. (2022) are common across a substantial fraction of chart extraction studies, often in combination with one another. Similarly, table-oriented benchmarks including PubTables-1M Smock et al. (2022), PubLayNet Zhong et al. (2019), TableBank Li et al. (2020), and ICDAR 2013 Göbel et al. (2013) are repeatedly employed across table extraction and table structure recognition systems. Several works rely on these shared datasets either directly or alongside synthetic or task-specific extensions, indicating a high degree of benchmark reuse within both modalities.

The Appendix provides a breakdown by article, including the names and URLs of the datasets and tool repositories. In particular, the Summary of Comparison Matrix of Reviewed Systems subsection documents the comparison matrix, while the Datasets and Tools and Systems subsections provide the curated dataset and tool lists. Overall, 57 studies make use of or release public datasets, while 65 provide open-source software, indicating that code availability remains more common than dataset publication. In contrast, a non-negligible subset of works relies on private or unavailable datasets or does not release any implementation, which continues to limit experimental reproducibility and cross-study comparability.

Table 4 summarizes the capabilities of the reviewed systems with respect to content recognition within individual modalities and alignment across tables and charts. Most approaches (66 out of 68) include mechanisms to identify

[8] https://github.com/erick4556/metadata-extraction-slr

and recognize descriptive content elements, such as table headers, row or column labels, chart axes, tick labels, or legend entries. This indicates that content recognition at the level of variables and labels beyond pure layout or region detection is widely supported across current systems. The remaining two systems focus primarily on extracting structured regions or visual elements without explicitly identifying or interpreting variables, and therefore provide more limited content-level analysis while still satisfying the inclusion criterion IC1.

In contrast, only 18 systems explicitly include mechanisms to connect or align information between tables and charts within the same document, while 50 systems do not address this capability. In this SLR, table–chart linking refers to document-level alignment strategies that associate related variables, labels, or numerical values across modalities. Typical examples include matching table headers with chart axis labels, linking chart series or bars to corresponding table rows or columns, or identifying that a table and a chart report results derived from the same experiment or dataset described in the paper. These links are established using information available within the document itself, such as textual labels, captions, figure references, or spatial and structural cues, where structural cues are layout-based signals (e.g., alignment, grouping, or ordering) that indicate relationships between visual elements in tables and charts.

Representative examples include UnchartIt Ramos et al. (2020), which links bar chart elements to corresponding rows in tables reporting the same experimental measurements, and TTC-QuAli Dong et al. (2024b), which explicitly models quantitative correspondences between text, tables, and charts to support cross-element consistency checking. These systems illustrate that multimodal alignment is typically achieved through shared labels, numerical consistency, caption analysis, or document-level context, rather than through external resources.

Overall, recognition of individual chart content components such as axes, legends, and graphical marks is now widely supported across the reviewed studies. As summarized in Table 13, 35 studies explicitly address chart component detection, and 14 studies report OCR-based extraction of textual elements such as axis labels and legends. In contrast, more advanced capabilities related to the reconstruction and interpretation of chart data remain less prevalent, with only 15 studies reporting end-to-end chart understanding and a substantially smaller subset of systems supporting higher-level reasoning beyond component-level extraction.

It is important to note that modality coverage and multimodality linking capture different aspects of system capabilities. The "Both" category in Table 2 refers to systems that explicitly process tables and charts as primary inputs. In contrast, the "Table–Chart Linking" capability reported in Table 4 reflects whether a system establishes relationships between tabular and graphical information within the same document, regardless of whether both modalities are fully processed as standalone inputs. As a result, some systems that operate primarily on a single modality (e.g., charts) may still align their outputs with information contained in tables, leading to a higher count for table–chart linking than for systems categorized as covering both modalities. The main extraction and interpretation tasks identified for each modality are examined in more detail in RQ-Based Literature Review and Analysis, where the reviewed approaches are analyzed according to the research questions.

Architectural Paradigms of Reviewed Systems

Table 5 classifies the analyzed systems according to their underlying architectural paradigm. This categorization distinguishes between convolutional neural network (CNN)-based systems, graph neural network (GNN) approaches, non-LLM transformer architectures, multimodal large language and vision–language models (LLMs/VLMs), unspecified deep learning methods, and rule-based or heuristic systems. In the resulting categorization, VLM-based systems are explicitly represented within the multimodal LLM/VLM family because these approaches jointly process visual and textual inputs and are commonly applied to chart question answering Masry et al. (2022); Han et al. (2025), chart-to-table conversion and visual data extraction Han et al. (2025); Zhang et al. (2024), and multimodal scientific document understanding Li et al. (2024); Khalighinejad et al. (2025).

A subset of the reviewed systems adopts multimodal large language or vision–language models, representing 19.1% of the final corpus. These approaches integrate visual and textual inputs within a unified multimodal framework, enabling joint reasoning over charts, tables, and surrounding document context. In this review, VLM-based systems are considered part of this family when they jointly encode visual content and textual information for tasks such as chart question answering, chart-to-table translation, visual data extraction, or multimodal scientific document understanding. Representative examples include ChartQA Masry et al. (2022), ChartLlama Han et al. (2025), TinyChart Zhang et al. (2024), and TTC-QuAli Dong et al. (2024b).

Transformer-based models that do not fall under the category of large language models account for 8.8% of the reviewed systems. These approaches employ the Transformer encoder–decoder mechanism Vaswani et al. (2017) for specialized visual, spatial, or structural analysis tasks, rather than for natural language understanding or generation. Representative examples include DETR architectures used for table or chart detection and region localization Carion et al. (2020), Table Transformer (TATR) models for table structure recognition and document layout analysis Zheng et al. (2021), and transformer-based parsers designed for cell segmentation or structural decomposition of tables and charts Yang et al. (2022). In contrast to multimodal LLMs, these models are tightly coupled to specific extraction tasks defined in this work, such as table detection, table structure recognition, chart element detection, and spatial segmentation (RQ1a–RQ1e). They operate primarily on visual features and positional encodings to reason about object boundaries, grid structure, and geometric relationships, without performing instruction following, open-ended text generation, or multimodal dialogue. As a result, transformer-based models are typically integrated into task-specific pipelines focused on object-level extraction and spatial reasoning, often serving as core components within larger table or chart extraction systems.

CNN-based models remain widely used (in 25.0% of the studies), particularly in early-stage pipelines for chart element recognition and table grid extraction. These

Table 4. Content-level and multimodality capabilities of analyzed systems

Capability type	Description and representative systems	Count
Variable and label identification	Identifies and interprets variables or descriptive elements such as table headers, row/column labels, chart axes, or legends (e.g., PubTables-1M Smock et al. (2022), ChartOCR Luo et al. (2021), TableLab Wang et al. (2021b))	66/68
Table–chart linking	Aligns or associates related variables or values between tables and charts within the same document (e.g., UNCHARTIT Ramos et al. (2020), TTC-QuAli Dong et al. (2024b))	18/68

Table 5. Distribution of architectural paradigms across reviewed systems

Architecture type	Number of systems	Percentage (%)
CNN-based	17	25.0
Multimodal LLM / VLM-based	13	19.1
Pure unspecified / Not reported	16	23.5
Graph neural network (GNN)	7	10.3
Transformer-based (non-LLM)	6	8.8
Unspecified deep learning	5	7.4
Rule-based / Heuristic	4	5.9
Total	68	100.0

approaches typically rely on convolutional backbones such as ResNet He et al. (2016) or EfficientNet Tan and Le (2019) and operate on pixel-level feature maps to capture local visual patterns and spatial cues relevant for detection and structural segmentation tasks.

Finally, a limited number of systems (5.9%) rely primarily on rule-based or heuristic approaches. These methods employ deterministic logic and predefined rules to support tasks such as layout refinement or rule-driven table parsing. In contrast, a substantial portion of the reviewed works (23.5%) do not clearly specify their underlying architectural paradigm. This category includes systems that either omit architectural details entirely or describe their approach at a high level without sufficient technical information to determine whether the method is based on deep learning, traditional machine learning, or non-learning-based techniques. The relatively large size of this group highlights persistent gaps in methodological reporting, which limits reproducibility and makes systematic comparison more difficult.

To complement the quantitative distribution of model architectures reported in Table 5, Table 6 summarizes the categories identified through the application of the evaluation framework to the reviewed methods. These categories are organized around five analytical dimensions: input modality, extraction level, model family, output representation, and integration capability. The resulting categorization supports comparison beyond task-level descriptions by clarifying what each system processes, how it models the extraction problem, what type of output it produces, and whether it supports semantic or multimodal integration.

Comparison of Standard and Proprietary Benchmark Adoption Among Reviewed Systems

Evaluation benchmark adoption varies considerably across the reviewed systems. As shown in Table 7, 32 systems evaluate their methods using benchmarks that are reused across multiple studies. To provide concrete evidence of which benchmarks are most commonly adopted, Table 8 details the most frequently used benchmarks in the reviewed corpus, together with the number of works that use them, the extraction or understanding tasks they target, and a brief description of their scope. ChartQA Masry et al. (2022), PlotQA Methani et al. (2020), and DVQA Kafle et al. (2018) emerge as the most recurrent chart-oriented benchmarks, primarily supporting tasks such as chart question answering, visual and numerical reasoning, and data digitization from plots. In contrast, PubTabNet Zhong et al. (2019), TableBank Li et al. (2020), and the ICDAR 2013 competition dataset Göbel et al. (2013) are the most commonly reused benchmarks for table-related tasks, with a focus on table detection and table structure recognition. The repeated use of these benchmarks enables partial comparability across independent works, particularly for mid-level extraction and understanding tasks such as chart QA, numerical reasoning, table detection, and structure reconstruction.

Table 7 also shows that 15 systems (22.1%) rely on proprietary or custom-built datasets. Importantly, the use of proprietary data does not necessarily imply that experiments are not reproducible, as several studies release their datasets or make them available upon request. However, because such datasets are typically adopted by a single study or by a very limited number of works, they offer limited support for systematic cross-paper comparison and cumulative evaluation.

The remaining systems either employ public datasets that are not reused by other studies in the corpus or do not clearly specify the benchmark used for evaluation. Rather than indicating a lack of methodological rigor, this observation reflects the diversity of task formulations, document types, and evaluation objectives currently present in multimodal table and chart extraction research. However, this diversity complicates direct comparison across systems when benchmark reuse, annotation conventions, or evaluation protocols differ.

Overall, the combined evidence from Table 7 and Table 8 indicates that, while a small set of benchmarks is repeatedly adopted across the literature, evaluation practices remain distributed across multiple datasets rather than converging on a single reference benchmark. Even the most frequently used benchmark (ChartQA Masry et al. (2022)) appears in only a limited fraction of the reviewed systems, reflecting the diversity of evaluation setups across modalities and tasks.

RQ-Based Literature Review and Analysis

This section provides a detailed synthesis of existing work organized according to the research questions defined in

Table 6. Main dimensions and corresponding categories across the evaluation framework

Dimension	Categories identified in the assessed works	Representative systems
Input modality	Table only; chart only; tables and charts	PubTables-1M; ChartOCR; TTC-QuAli
Extraction level	Detection; structure reconstruction; label or variable identification; value association; table–chart alignment	GTE; LineEX; UnchartIt; MATVIX
Model family	Rule-based or heuristic; CNN-based; GNN-based; Transformer-based (non-LLM); multimodal LLM/VLM-based; hybrid	TableNet; TATR; ChartLlama; TinyChart
Output representation	Bounding boxes; reconstructed tables; CSV or JSON; textual answers; semantically enriched structured representations	ChartReader; QURMA; TTC-QuAli
Integration capability	Component-level extraction; intra-modal association; cross-modal linking; semantic integration	ChartOCR; PubTables-1M; UnchartIt; MMSci; MATVIX; QURMA

Table 7. Comparison of standard and proprietary benchmark adoption among reviewed systems

Benchmark category	Yes (n)	Yes (%)	No (n)	No (%)
Uses standard benchmarks	32	47.1	36	52.9
Uses proprietary benchmarks	15	22.1	53	77.9

Table 8. Most frequently used benchmarks for table and chart extraction and understanding tasks

Benchmark	Modality	Tasks	No. of works	Brief description
ChartQA Masry et al. (2022)	Charts	Chart QA, reasoning	11	Benchmark for chart understanding and numerical reasoning using chart images paired with natural language questions.
PlotQA Methani et al. (2020)	Charts	Data digitization, QA	8	Dataset focused on numerical value extraction and quantitative reasoning over plots.
DVQA Kafle et al. (2018)	Charts	Visual QA	4	Visual question answering benchmark over bar charts with compositional and multi-step questions.
PubTabNet Zhong et al. (2019)	Tables	Structure recognition	4	Benchmark for table structure recognition with HTML-based ground truth annotations.
TableBank Li et al. (2020)	Tables	Detection, structure recognition	3	Large-scale dataset of labeled tables from Word and LaTeX documents for table detection and structure recognition.
ICDAR 2013 Göbel et al. (2013)	Tables	Detection, recognition	3	Competition dataset for table detection and recognition in document images.

the Methodology and the analysis builds upon the aggregated summaries reported in Results.

RQ1a: What are the current approaches, tools, and models for data extraction from tables and charts in scientific publications?

As shown in Table 2, most systems in the corpus focus on a single modality, either charts (47.1%) or tables (41.2%), while only a smaller subset (11.8%) explicitly supports both within the same document. The descriptions recorded in the study for RQ1a confirm that these works focus on concrete tasks for extracting information from tables or charts in scientific publications, and that they introduce a variety of approaches (e.g., deep learning models, hybrid methods, rule-based components) and tools (interactive systems, pipelines, and libraries) that operate mainly within a single modality.

In the case of tables, the results show that many works propose complete pipelines for table detection, table structure recognition, and conversion of PDF tables into structured tables. Representative examples include PubTables-1M Smock et al. (2022), which provides a large-scale dataset and metrics for table extraction from unstructured documents, and TabLeX Desai et al. (2021), which focuses on table detection in scientific documents. Other works in the study address end-to-end extraction from PDFs, combining deep learning and symbolic analysis for digital table extraction Zhang et al. (2021), tools such as TableLab Wang et al. (2021b), and page layout analysis

to refine table extraction Mikhailov and Shigarov (2021). Several entries describe methods based on graph neural networks or clustering over lines and text for table extraction in PDF documents Gemelli et al. (2022); Boutalbi et al. (2023), as well as pipelines oriented to knowledge base population Nugumanova et al. (2022) or to specific domains such as financial and invoice tables Peng and Li (2025), which we cite here as representative examples of application-driven extraction settings. Our evaluation framework also includes toolkits and utilities such as PyTabby Mikhailov et al. (2021), and pdf2gtfs Yildiz et al. (2005), that provide practical support for table extraction in applied environments, such as scientific PDFs and documents from specific application domains. Surveys and overviews on table processing and table recognition Embley et al. (2006); Shah and Patel (2004); Shigarov (2022) complement these contributions by organizing common task formulations and methodological categories reported in the evaluation framework, such as table detection, structure recognition, and end-to-end extraction. In addition, recent evaluations of PDF extraction tools and parsing methods Meuschke et al. (2023) assess the performance of document-to-table extraction pipelines across multiple document types and application domains. Overall, the entries marked as "Table Only" in Table 2, or otherwise associated with table-related tasks under RQ1a, show a strong focus on detecting tables within document layouts, recovering their internal structure by identifying rows, columns, and cells, and producing structured tabular outputs that can be directly reused in downstream applications.

Chart extraction is represented in our analysis by works that focus on extracting structured information from chart images and visualizations, predominantly in scientific and technical documents. The systems categorized as "Chart Only" in Table 2 address a range of recurring chart-related tasks, including the detection of chart components, extraction of axes and tick values, recognition of legends and textual labels using OCR, and the digitization of visual marks such as lines, bars, or points. To clarify the scope of these tasks, Table 9 summarizes the main chart extraction tasks identified in the reviewed literature, together with brief descriptions and representative systems. Early contributions and survey on chart analysis and understanding Huang and Tan (2007); Farahani et al. (2023); Huang et al. (2025) provide background on these tasks, while subsequent systems address them using a variety of approaches, including interactive tools for chart data extraction Jung et al. (2017); Zhao et al. (2022); Chai et al. (2021), learning-based methods for structuring chart content Al-Zaidy and Giles (2017), and automatic generation of textual or numeric representations from charts Dai et al. (2018); Mishra et al. (2022b); Kantharaj et al. (2022). Additional studies focus on specific chart families, such as line charts Sohn et al. (2021); Shivasankaran et al. (2023); Lal et al. (2023), bar charts Daggubati et al. (2022); Zhou et al. (2021), scatter plots Dadhich et al. (2021), or charts with circular structures Daggubati and Sreevalsan-Nair (2022). Benchmarks such as PlotQA Methani et al. (2020), ChartQA Masry et al. (2022), and ChartInfo Ma et al. (2021) are repeatedly used to support tasks related to chart question answering, data extraction, and the evaluation of chart understanding models.

Inspection of the results also shows that only a small subset of works explicitly address both tables and charts or position themselves as multimodal. These include benchmarks and systems such as TTC-QuAli Dong et al. (2024b), MATVIX Khalighinejad et al. (2025), MMSci Li et al. (2024), and MaCBench Alampara et al. (2024). These works indicate a growing interest in multimodal document understanding, although most of them still focus on coarse document-level or task-level integration rather than explicit variable–value alignment across tables and charts. Table 10 summarizes the benchmarks identified for RQ1a by modality and application domain.

In addition to extraction pipelines focused on tables and charts as standalone artifacts, a smaller but relevant subset of works explicitly connects data extraction tasks with ontology-based representations and knowledge graph (KG) construction. These approaches typically build upon the outputs of table and chart extraction systems to populate structured, semantically enriched representations that support integration, comparison, and reuse across documents. For instance, pipelines oriented toward knowledge base population, such as QURMA Nugumanova et al. (2022), explicitly target the transformation of extracted tables into structured entities and relations suitable for downstream semantic integration. Similarly, systems and frameworks discussed in the semantic web literature emphasize semantic table interpretation, variable disambiguation, and entity linking as essential post-extraction steps for KG construction Liu et al. (2023a); Nguyen et al. (2024); Limaye et al. (2010).

Large-scale initiatives such as the Open Research Knowledge Graph (ORKG) Kabongo et al. (2024) rely on structured representations of scholarly contributions and variables, but typically assume that table and figure extraction has already been performed by upstream tools rather than relying directly on specific chart or table benchmarks such as PubTables-1M Smock et al. (2022) or ChartQA Masry et al. (2022). In this sense, most benchmarks reviewed under RQ1a primarily evaluate low- and mid-level extraction capabilities (e.g., detection, structure recognition, and data reconstruction), while ontology-driven integration and knowledge graph population are addressed at a later stage in the processing pipeline.

Overall, the information collected for RQ1a shows that current research on structured data extraction from scientific publications covers a broad range of tasks, approaches, and tools. For tables, most works focus on detecting table regions, recovering internal structure, and converting content into structured formats using deep learning and layout analysis, while comparatively few address semantic reconciliation tasks such as aligning headers, units, and contextual descriptions with extracted values. For charts, the emphasis lies on component detection, OCR-based label extraction, digitization of visual marks, and data reconstruction, supported by dedicated benchmarks and interactive or semi-automatic systems. Only a limited subset of studies explicitly integrates tables and charts through multimodal benchmarks or models, as summarized in Table 2. Even in these cases, multimodal approaches primarily address numerical or visual alignment rather

Table 9. Common Chart Extraction Tasks Identified

Task	Brief description	Representative systems
Chart component detection	Identification of chart regions and visual elements such as axes, bars, lines, points, and legends.	ChartOCR; ChartDetective; Yan et al. Yan et al. (2023)
Axis and scale extraction	Extraction of axis lines, tick marks, and numeric scales to support value reconstruction.	ChartQA; ChartInfo; Sohn et al. Sohn et al. (2021)
Text and label recognition	Recognition of textual elements including titles, axis labels, and legends using OCR and layout cues.	ChartOCR; ChartDecoder Dai et al. (2018)
Visual mark digitization	Conversion of graphical elements (e.g., bars, lines, scatter points) into numerical data.	PlotQA; LineEX Shivasankaran et al. (2023); ReverseBar Zhou et al. (2021)
Chart question answering	Answering natural language questions over charts by combining visual and textual cues.	ChartQA; PlotQA
Chart-to-text or table generation	Generation of textual descriptions or structured tables from chart images.	Chart2Text Kantharaj et al. (2022); ChartVI Mishra et al. (2022b)

Table 10. Summary of benchmarks by modality and covered domains

Modality	Number of benchmarks	Main domains
Tables	17	Information retrieval; PDF accessibility; general-purpose table understanding.
Charts	24	Scientific analysis (e.g., materials science); vision–language models; chart understanding and data extraction; model evaluation and instruction tuning.
Multimodal (tables + charts)	6	AI evaluation; scientific domains (chemistry, laboratory knowledge, materials science); performance assessment; information retrieval and quantitative reasoning.

than explicit semantic representation. Ontology-driven integration and knowledge graph population are therefore typically treated as downstream processes that rely on the outputs of table and chart extraction systems, as exemplified by large-scale initiatives such as the Open Research Knowledge Graph. Based on the categories identified through the evaluation framework, the main divide in the literature is not only between table-oriented and chart-oriented systems but also between systems focused on structural extraction and systems that attempt semantic or multimodal interpretation. Most table-oriented methods focus on detection, structure reconstruction, and machine-readable table generation, whereas most chart-oriented methods emphasize component detection, OCR-based label extraction, and value digitization. Systems that explicitly support cross-modal linking or semantically enriched outputs for downstream knowledge graph construction remain comparatively rare. This comparison shows that current progress is focused on the component and structure levels, while higher-level integration capabilities are still less mature.

RQ1b: What models and architectures are used for tabular data extraction?

As shown in Table 2, 36 studies address table-related extraction tasks, either as table-only systems (n=28) or as systems covering both tables and charts (n=8). These entries provide detailed descriptions of the models and architectures proposed for detecting tables, identifying rows and columns, recovering cell boundaries, converting tables from PDF or image formats into structured outputs, and, in some cases, supporting higher-level reasoning over tabular or multimodal document content. The information recorded in the study indicates that most table extraction systems rely on deep learning models that operate on visual layout cues, OCR results, or both. Many approaches combine document image processing with region detection, while others incorporate sequence models, graph-based models, transformer-based components, or multimodal LLM/VLM-based systems designed to handle complex layouts and visual–textual reasoning. Table 11 provides a structured summary of the main table-related extraction tasks and the representative model families commonly used to address them across the reviewed studies.

Representative architectures corresponding to the task and model-family categories summarized in Table 11 span a range of modeling strategies aligned with different stages of the table extraction pipeline. Convolutional neural network models are commonly used for table detection and region segmentation, as exemplified by TableNet Paliwal et al. (2019). For table structure recognition, several approaches rely on top-down and bottom-up cues to infer rows and columns, including models such as TabStructNet Raja et al. (2020), as well as transformer-based systems like TableFormer Yang et al. (2022), which employ splitting-based strategies to recover table structure. Other works adopt hybrid pipelines that integrate visual analysis with rule-based or OCR-driven processing, particularly for cell boundary recovery. Representative examples include

Table 11. Table-related extraction tasks and representative model families

Task	Model family	Model architectures / processing strategy	Representative systems
Table detection	CNN-based and region-based models	Convolutional detectors and region-based models operating on document layout images to identify table regions.	TableNet Paliwal et al. (2019)
Table structure recognition	Transformer-based models	Transformer-based and splitting-based architectures for recovering rows, columns, and table structure.	TableFormer Yang et al. (2022), TabStructNet Raja et al. (2020)
Cell boundary recovery	Hybrid visual–OCR and rule-assisted models	Pipelines combining visual layout analysis, OCR outputs, and rule-assisted processing to recover cell boundaries and table grids.	GTE Zheng et al. (2021), Integrated PDF pipelines Zhang et al. (2021); Mikhailov and Shigarov (2021)
Implicit table structure recovery	Graph-based and clustering-based models	Graph or clustering methods that leverage spatial alignment, text proximity, and layout relationships when explicit ruling lines are absent.	GNN-based table extraction Gemelli et al. (2022), clustering-based methods Boutalbi et al. (2023)
Table layout reconstruction	Clustering and heuristic models	Clustering approaches combining line detection, text grouping, and spatial heuristics to organize the global table layout.	Clustering-based methods Boutalbi et al. (2023)
Pipeline-based table extraction	Integrated extraction pipelines	End-to-end or modular pipelines combining table detection, structure recovery, and export into structured formats.	QURMA Nugumanova et al. (2022), PyTabby Mikhailov et al. (2021)
Higher-level table and multimodal document understanding	Multimodal LLM/VLM-based models	Vision–language and multimodal LLM-based systems that jointly process visual and textual inputs to support table reasoning, chart-to-table translation, visual data extraction, or multimodal document understanding. These models are less commonly used for low-level table structure extraction, but are increasingly relevant for semantic interpretation and document-level reasoning.	TTC-QuAli Dong et al. (2024b), MATVIX Khaligh-inejad et al. (2025), MMSci Li et al. (2024)

GTE Zheng et al. (2021) and pipelines specialized for digital PDF extraction Zhang et al. (2021); Mikhailov and Shigarov (2021). Implicit table structure recovery in the absence of explicit ruling lines is addressed by graph-based or clustering-based models that leverage spatial alignment and text proximity to infer relationships between table components, as illustrated by GNN-based approaches Gemelli et al. (2022) and clustering-based methods Boutalbi et al. (2023). Related clustering strategies are also employed for table layout reconstruction, where line detection, text grouping, and spatial heuristics are combined to organize the global table structure. The study additionally includes systems that provide complete, pipeline-based table extraction solutions, such as QURMA Nugumanova et al. (2022) and PyTabby Mikhailov et al. (2021), which integrate detection, structure recovery, and export functionalities to support the processing of real-world documents.

The revised model-family organization also makes explicit the role of multimodal LLM/VLM-based systems. These approaches are not primarily designed for low-level table structure extraction tasks such as cell boundary recovery or row and column segmentation. Instead, they jointly process visual and textual inputs to support higher-level reasoning over tables, charts, and surrounding document context. In this review, such systems are therefore represented as a separate model family associated with higher-level table and multimodal document understanding. Representative examples include TTC-QuAli Dong et al. (2024b), MATVIX Khaligh-inejad et al. (2025), and MMSci Li et al. (2024), which address multimodal reasoning, quantitative alignment, or scientific document understanding across textual, tabular, and graphical elements.

A review of the modeling strategies across these entries shows two dominant tendencies. First, systems built specifically for document layout tasks generally rely on visual detectors, region segmentation networks, or document-structure transformers trained on table-rich datasets such as PubTables-1M Smock et al. (2022), TableBank Li et al. (2020), TabLeX Desai et al. (2021), and related benchmarks. Second, some works adapt architectures originally developed for general object detection, such as convolutional detectors or transformer encoders, into table detection or structure recognition tasks. This pattern is visible in methods that apply edge prediction Liu et al. (2023b), span detection, or column and row grouping based on visual similarity. Several studies also highlight

that performance is influenced by OCR quality and layout irregularities, especially in documents containing merged cells, missing borders, or multi-line headers.

The analysis of the reviewed corpus also indicates that the level of architectural detail varies considerably across publications. Some entries provide precise descriptions of model components and training procedures, while others lack detailed specifications, making replication more difficult. A number of studies evaluate their models using proprietary datasets or domain-specific collections, and several entries note that differences in annotation conventions across datasets cause inconsistencies when comparing results or reusing pretrained models.

Overall, the information collected for RQ1b shows that tabular data extraction relies on a diverse set of model families, including convolutional networks, graph-based models, Transformer-based architectures, multimodal LLM/VLM systems, and hybrid OCR–vision pipelines. The categorization derived from the evaluation framework helps clarify how these families contribute to different stages and output levels of the extraction process. CNN-based systems are mainly used for visual detection and segmentation tasks, graph-based approaches model spatial relationships among document elements, Transformer-based architectures support more flexible structural reasoning over layouts, and hybrid OCR–vision pipelines combine textual and visual cues to recover table structure. These systems have achieved reliable performance on core tasks such as table region detection and the recovery of rows, columns, and cell boundaries. However, the comparison also shows that most table extraction systems remain concentrated at the structural or component level. While CNN-, GNN-, Transformer-based, and hybrid systems often produce outputs such as detected regions, reconstructed rows and columns, or cell-level structures, they rarely generate explicit semantic mappings between headers, variables, units, and values. Recent multimodal LLM/VLM systems improve joint visual–textual reasoning and can support question answering or higher-level interpretation, but they still rarely produce verifiable mappings between variables, values, units, and semantic representations. As a result, while accurate recovery of table structure is relatively well established, semantic richness, consistency in evaluation protocols, dataset variability, and completeness in architectural reporting remain important challenges for future progress.

RQ1c: How do existing systems transform tables from unstructured document formats into machine-readable representations?

The analysis of the reviewed corpus indicates that the transformation of tables from unstructured document sources, particularly PDF files, into machine-readable representations is a central focus of the works analyzed. Many entries describe pipelines that begin with page-level detection of table regions and subsequently recover internal structure through row and column identification, cell segmentation, or boundary refinement. Representative systems include convolutional and hybrid approaches such as TabStructNet Raja et al. (2020), TableFormer Yang et al. (2022), and graph- and clustering-based methods for table organization recovery Gemelli et al. (2022); Boutalbi et al. (2023), as well as integrated PDF-oriented pipelines Zhang et al. (2021); Mikhailov and Shigarov (2021). These approaches rely heavily on visual cues to infer table layout, including explicit ruling lines, consistent spacing between cells, alignment of textual elements into rows and columns, and rectangular grouping of content blocks. Such visual indicators are especially prominent in works that operate directly on document images or scanned PDFs, where explicit structural markup is unavailable, and are used to approximate table boundaries and internal divisions prior to generating structured, machine-readable outputs.

Studies addressing table conversion from scientific publications reveal a consistent processing pattern. Most approaches rely on an initial parsing stage to reconstruct text positions, bounding boxes, and layout information from PDF pages, which serves as the foundation for subsequent table reconstruction.

The counts reported in Table 12 are computed over the subset of studies that explicitly address table conversion or table extraction from PDF documents (n=17). This subset is derived from the broader group of table-related studies reported in Table 2, which includes table-only studies and studies covering both tables and charts (n=36). The denominator is therefore narrower than the general modality distribution because not every table-related paper addresses PDF-to-table conversion or machine-readable table reconstruction. A single study may contribute to multiple categories because modern table extraction systems often combine several processing stages.

As summarized in Table 12, visual layout analysis emerges as the most prevalent approach, appearing in 12 studies. These approaches exploit ruling lines, spacing regularities, visual context, bounding boxes, and alignment patterns to infer rows, columns, table regions, or cell boundaries. PDF text and layout parsing and rule-based or heuristic processing are each reported in 9 studies. PDF parsing is commonly used to recover text coordinates, bounding boxes, text layers, font attributes, or reading order from document pages, while rule-based or heuristic processing supports the handling of merged cells, spanning headers, irregular layouts, or table boundaries. Integrated PDF table extraction pipelines appear in 8 studies and typically integrate parsing, detection, structure recovery, and structured export within a single workflow. Page layout refinement is reported in 6 studies, usually as a post-processing step to improve bounding boxes, resolve overlaps, or correct row and column boundaries. Clustering-based structure recovery appears in 1 study, where spatial proximity and alignment between lines and text are used to reconstruct table organization. Despite these advances, several studies highlight persistent challenges related to merged cells, irregular borders, multi-line content, and tables embedded within complex page layouts, which often limit the robustness of visually driven conversion methods when clear ruling lines or consistent alignment cues are absent.

Only a subset of the systems reviewed for RQ1c explicitly document the generation of fully structured outputs such as JSON, XML, or CSV. In addition, a substantial body of work has investigated the transformation of tabular data

Table 12. Techniques used for table conversion from PDF documents

Technique	Number of studies	Description
Visual layout analysis	12	Use of ruling lines, spacing regularities, visual context, bounding boxes, and alignment patterns to infer rows, columns, table regions, or cell boundaries.
Clustering-based structure recovery	1	Grouping of text blocks, line segments, or bounding boxes based on spatial proximity and alignment to reconstruct table organization.
PDF text and layout parsing	9	Extraction of text coordinates, bounding boxes, text layers, font attributes, or reading order from PDF pages as a preprocessing step for table reconstruction.
Rule-based and heuristic processing	9	Application of manually defined rules, symbolic processing, heuristics, or document-structure assumptions to handle merged cells, spanning headers, irregular layouts, or table boundaries.
Integrated PDF table extraction pipelines	8	Integrated systems combining parsing, detection, structure recovery, and export into structured formats.
Page layout refinement	6	Post-processing techniques that refine detected table regions by adjusting bounding boxes, resolving overlaps, correcting misalignments, or improving row and column boundaries.

Note: Counts are computed over the subset of studies that explicitly address PDF-based table extraction or reconstruction into machine-readable representations (n=17). This subset is narrower than the table-related group reported in Table 2 (n=36), because not all table-related studies address PDF-to-table conversion or machine-readable table reconstruction. A single study may contribute to multiple techniques because table extraction systems often combine several processing stages.

into RDF representations to support semantic integration and knowledge graph construction Fiorelli and Stellato (2021). However, as highlighted in this survey, most approaches that lift tables to RDF assume that the tabular data is already available in structured or semi-structured form (e.g., HTML tables, CSV files, or web tables), and do not operate as end-to-end pipelines starting directly from unstructured document formats such as PDFs. Several tools and extraction pipelines in the reviewed corpus indicate that their extracted tables can be directly reused by downstream applications, including systems such as QURMA Nugumanova et al. (2022), as well as domain-focused extraction pipelines for scientific or financial documents Circi et al. (2024); Peng and Li (2025); Balsiger et al. (2024). In many other cases, while the extracted table content is implicitly reusable, the target structured format and the intended integration pathway (e.g., RDF generation, ontology alignment, or knowledge graph population) are not described in detail.

In parallel, a substantial body of work from the Semantic Web and data management communities has investigated the interpretation, annotation, and reuse of tabular data, particularly in the context of large collections of web tables. Prior studies have explored the identification of entities, column types, and relationships in tables Limaye et al. (2010); Venetis et al. (2011), the profiling of tabular data for integration and reuse across domains Ritze et al. (2016), and the construction of large-scale table corpora to support downstream analysis Lehmberg et al. (2016). More recent work has also investigated structured representations of tables for retrieval and downstream reasoning tasks Wang et al. (2021a). However, these lines of work are less frequently adopted or explicitly discussed in recent table extraction pipelines reviewed under RQ1c, which tend to focus primarily on layout reconstruction rather than on the end-to-end specification of machine-readable structured outputs.

This distinction helps delimit the scope of the review. Approaches centered on semantic table interpretation, entity linking, column type annotation, semantic annotation, or knowledge graph population are discussed as related work because they generally operate on tabular data that has already been extracted or structured Liu et al. (2023a); Nguyen et al. (2024); Limaye et al. (2010); Venetis et al. (2011); Ritze et al. (2016). Their contribution is therefore most relevant to the subsequent representation, alignment, and reuse of extracted information, including variables, units, entities, and relations. Such studies are included among the 68 reviewed papers only when they also address the extraction or reconstruction of information directly from unstructured scientific document artifacts.

Although dataset and benchmark availability is summarized in Results, the reviewed RQ1c systems reveal that reproducibility also depends on the documentation of PDF parsing configurations, preprocessing steps, and conversion procedures. Several studies provide limited detail on these implementation choices, which makes it difficult to compare table conversion pipelines even when public datasets or open-source tools are available.

Overall, the information collected for RQ1c shows that most systems demonstrate strong capabilities in detecting tables and recovering their visual layout, but fewer works fully address the challenges of converting these structures into complete machine-readable representations. The reviewed entries consistently highlight the dependency on document parsing, the sensitivity to layout irregularities, and the variability of dataset conventions. As a result, the conversion of unstructured table formats into structured representations remains an area where progress is steady, but where implementation consistency and robustness across

diverse document layouts in scientific publications continue to pose significant challenges.

RQ1d: What tools and models are used to extract data from charts?

The results include a broad range of studies dedicated to chart data extraction, covering tasks such as identifying chart components, recovering textual labels, and reconstructing values encoded visually.

The counts reported in Table 13 are computed over the subset of chart-related studies (n=40). This subset is derived from the modality distribution reported in Table 2, and includes chart-only studies (n=32) and studies covering both tables and charts (n=8). This denominator is used because the table summarizes extraction tasks that are specific to chart processing, rather than tasks observed across the full corpus of 68 studies. A single study may contribute to multiple extraction tasks because modern chart understanding systems often combine several processing stages within the same pipeline.

As summarized in Table 13, most approaches focus on chart component detection, data point digitization, and chart-type specific extraction, while fewer systems explicitly address end-to-end chart understanding. Text and label extraction through OCR appears as a supporting component in a smaller subset of studies. The works in this group typically describe pipelines that combine optical character recognition (OCR) with detection of visual elements such as bars, lines, points, axes, legends, and tick labels. Representative systems documented in the corpus include ChartOCR Luo et al. (2021), LineEX Shivasankaran et al. (2023), Reverse-Engineering Bar Charts Zhou et al. (2021), ScatterPlotAnalyzer Dadhich et al. (2021), and tensor-field methods for bar and scatter charts Sreevalsan-Nair et al. (2021). A detailed list of systems and their associated evaluation metrics is provided in Appendix A (Table 16).

Figure 3 provides a high-level overview of the main tool categories and modeling paradigms used for chart data extraction, synthesizing the processing pipelines discussed throughout this research question. Arrows indicate the conceptual flow of information from the input chart image toward structured chart data, rather than a strict or mandatory execution order. The colored blocks represent distinct methodological families commonly employed in chart extraction systems. OCR-based tools focus on extracting textual elements such as axis labels and legends, visual detectors identify graphical primitives including bars, lines, and points, and rule-based or heuristic models align visual elements with numeric values through axis calibration and coordinate mapping. More recent multimodal vision–language models jointly process visual representations of charts and associated textual information to support integrated reasoning over graphical elements, labels, and underlying data Masry et al. (2022); Cheng et al. (2023); Han et al. (2025); Zhang et al. (2024). Colors are used solely to visually distinguish these methodological categories and do not imply exclusivity, as in practice most systems combine several of these components within the same pipeline to produce structured chart data suitable for downstream analysis.

Several entries in the study introduce interactive tools designed to support chart extraction through visual highlighting and user guidance. Representative examples include UNCHARTIT Ramos et al. (2020), ChartSense Jung et al. (2017), CrowdChart Chai et al. (2021), and ChartDetective Masson et al. (2023). These systems typically assist in the identification and extraction of chart variables, legends, axis labels, and data series by combining automated detection with user interaction, rather than relying on fully automated pipelines. Other systems specialize in particular chart types, such as line charts Sohn et al. (2021); Shivasankaran et al. (2023), bar charts Daggubati et al. (2022); Zhou et al. (2021), or circular and radial chart structures Daggubati and Sreevalsan-Nair (2022). These focused approaches explicitly extract chart components such as variables, legends, axes, and data points for the targeted chart families, and tend to achieve strong accuracy within those categories. However, according to the results, they often report reduced performance when applied to unfamiliar chart designs.

As summarized in Results, chart extraction studies frequently reuse benchmarks such as PlotQA Methani et al. (2020), ChartQA Masry et al. (2022), and ChartInfo Ma et al. (2021). In RQ1d, these benchmarks are relevant mainly because they shape the types of chart extraction tasks being evaluated, particularly component detection, OCR-based text association, data digitization, and chart question answering. However, their coverage remains concentrated around a limited set of chart types, which partly explains why several systems report reduced robustness when applied to unfamiliar or irregular visual designs.

Recent works highlight a growing adoption of multimodal and vision–language models for chart analysis. Systems such as ChartReader Cheng et al. (2023), ChartLlama Han et al. (2025), TinyChart Zhang et al. (2024), ChartAssistant Meng et al. (2024), ChartMoE Xu et al. (2025), and ChartCitor Goswami et al. (2025) apply transformer-based encoders and multimodal training strategies to jointly process chart images and associated textual content. Despite architectural differences, these models share several common preprocessing steps. Most pipelines include chart image normalization and resizing, detection or cropping of the plotting area, OCR-based extraction of axis labels and legends, and alignment of recognized text with visual regions. Common OCR configurations typically include standard text recognizers applied to high-resolution graphic regions, often with heuristic filtering to remove background text or elements that are not part of the graphic. In addition, several models perform axis normalization, which refers to mapping visual positions of ticks, bars, or points into a unified numerical coordinate system based on detected axes, scales, and units, enabling consistent value reconstruction across charts with different resolutions or aspect ratios. While some approaches incorporate chart-to-table pretraining or instruction tuning to improve reasoning and text association, others focus on tightly coupling OCR outputs with visual embeddings. Although these shared preprocessing choices enable effective multimodal learning, the reviewed studies differ substantially in how these steps are implemented and reported, particularly with respect to OCR settings, axis scaling assumptions, and normalization

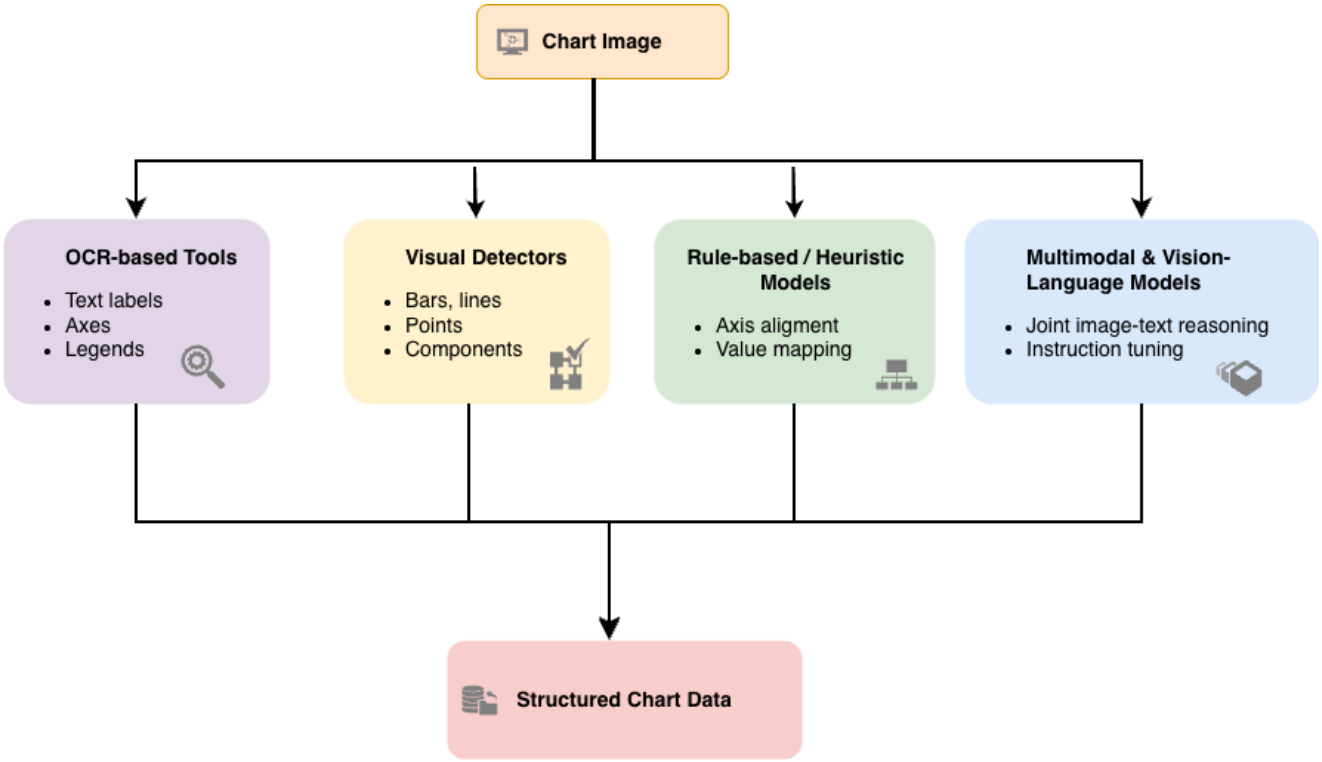

Figure 3. Overview of the main tools and modeling paradigms used for chart data extraction. Arrows indicate dataflow.

Table 13. Chart data extraction tasks and representative systems or approaches

Extraction task	Number of studies	Representative systems / approaches
Chart component detection	35	ChartOCR Luo et al. (2021); Reverse-engineering bar charts Zhou et al. (2021); ScatterplotAnalyzer Dadhich et al. (2021)
Data point digitization	34	LineEX Shivasankaran et al. (2023); ScatterplotAnalyzer Dadhich et al. (2021); tensor-field-based extraction methods Sreevalsan-Nair et al. (2021)
Chart-type specific extraction	33	LineEX Shivasankaran et al. (2023) (line charts); bar chart reconstruction approaches Zhou et al. (2021)
End-to-end chart understanding	15	Integrated chart understanding pipelines (e.g., Masry et al. (2022); Cheng et al. (2023))
Text and label extraction (OCR)	14	ChartOCR Luo et al. (2021); LineEX Shivasankaran et al. (2023); OCR-based chart extraction pipelines (e.g., Masry et al. (2022))

Note: Counts are computed over the subset of chart-related studies (n=40), derived from the chart-only studies (n=32) and studies covering both tables and charts (n=8) reported in Table 2. A single study may contribute to multiple extraction tasks because modern chart understanding systems often combine several processing stages within the same pipeline.

heuristics. This variability limits reproducibility and complicates direct comparison across multimodal chart extraction systems. Across the reviewed systems, many approaches still rely on manually defined rules to convert detected visual elements into numerical values. In this context, handcrafted rules refer to explicitly defined decision rules that map visual measurements (e.g., pixel distances, bounding-box positions, or relative alignment) to numeric values without being learned end-to-end from data. Typical examples include fixed thresholds for matching bars or points to axis ticks, normalization of pixel coordinates based on detected axis scales, and rule-based association of labels with nearby visual marks. Such heuristics are explicitly reported in works such as the reverse-engineering of bar charts by Zhou et al. Zhou et al. (2021) and the tensor-based extraction approach in ScatterPlotAnalyzer Dadhich et al. (2021), where geometric thresholds such as distance tolerances, slope constraints, or spatial proximity rules play a central role in reconstructing chart values. Similar rule-based alignment strategies are described across multiple chart extraction systems summarized in Table 13, particularly within tasks such as data point digitization and chart-type specific extraction, indicating that this design choice is widespread rather than isolated. While these strategies are effective for conventional chart layouts with regular axes and clear visual separation, many studies report reduced robustness when applied to charts with irregular scaling, stylistic variation, overlapping text, or multiple axes. This limitation is reflected in the performance differences observed between systems primarily evaluated on bar and line charts and those applied to more complex chart types such as scatterplots or circular visualizations.

Overall, the evidence collected for RQ1d shows that chart extraction systems rely on a recurring set of components, including OCR modules for textual labels, visual detectors for chart elements, heuristic alignment procedures for value reconstruction, and, more recently, multimodal and vision–language models for joint visual–textual processing. As summarized in Table 13, which shows a high prevalence of chart component detection, data point digitization, and chart-type–specific extraction tasks, and supported by the benchmark usage patterns in Table 8, most approaches focus on a limited set of chart types, primarily bar and line charts, evaluated on a small number of widely used benchmarks. While recent multimodal models improve integration of visual and textual cues, the reviewed systems remain sensitive to layout variation and chart design diversity, indicating that robust generalization across chart types is still an open challenge. From the perspective of knowledge representation and knowledge graph construction, none of the chart extraction systems reviewed under RQ1d explicitly target the transformation of extracted chart information into semantic representations such as RDF or ontology-aligned graph structures. Existing approaches focus primarily on reconstructing numerical values and tabular representations (e.g., CSV or JSON) suitable for downstream analysis, but stop short of semantic modeling or formal integration into knowledge graphs. While ontologies and semantic models for tabular data and scientific knowledge exist in the broader semantic web literature Fiorelli and Stellato (2021), their adoption in chart extraction pipelines remains largely unexplored. This reveals a clear gap between chart understanding research and ontology-driven knowledge integration, and highlights an open opportunity for future work to bridge chart data extraction with semantic representation and knowledge graph population.

RQ1e: How do current methods reconstruct the underlying data represented in charts into structured tabular formats?

Our results show that only a limited subset of works explicitly address the reconstruction of charts into structured tabular representations, as reflected by the relatively low prevalence of end-to-end chart understanding tasks in Table 13, despite the widespread adoption of data point digitization approaches. Systems such as UNCHARTIT Ramos et al. (2020), LineEX Shivasankaran et al. (2023), and approaches documented in ChartInfo Ma et al. (2021) and PlotQA Methani et al. (2020) aim to recover the numerical values encoded in chart images and convert them into machine-readable outputs. These systems typically isolate graphical primitives such as bars, points, or lines, extract textual labels through OCR, and align both components with parsed axis scales to generate representations such as CSV or JSON. The resulting tables aim to approximate the underlying data series originally used to produce the chart, enabling downstream analytical processing and reuse Kantharaj et al. (2022). Figure 4 illustrates the main stages involved in reconstructing structured tabular data from chart images, as identified in RQ1e. The pipeline highlights the sequential dependency between visual component detection and textual extraction, followed by axis alignment and value mapping, which are necessary to translate visual positions into numeric values. These steps enable data point digitization and, ultimately, the reconstruction of machine-readable tabular representations such as CSV or JSON, which may later serve as input for downstream semantic representations.

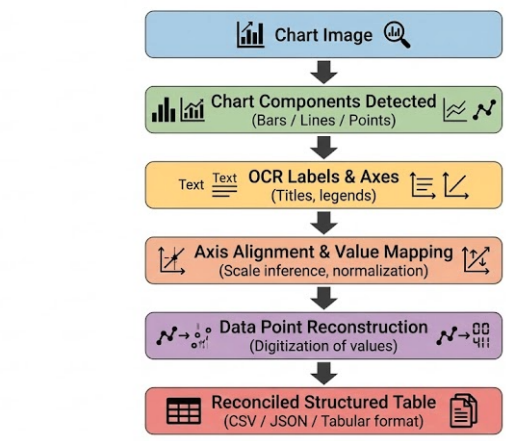

Figure 4. Conceptual pipeline for reconstructing structured tabular data from chart images, synthesized from common processing stages reported in the reviewed studies.

Across the reviewed systems, reconstruction methods tend to specialize in a narrow set of chart types, particularly bar and line charts Ramos et al. (2020); Shivasankaran et al. (2023). In this context, reconstruction refers to the combination of data point digitization and the reconciliation of visual marks with textual labels and axis scales, tasks that, while widely addressed at the level of individual data point extraction, are less frequently extended to full end-to-end reconstruction, as reflected in Table 13.

Bar and line charts are particularly well suited for reconstruction because their visual elements show simple, well-defined spatial relationships with the underlying numerical values. Specifically, values can be inferred by mapping the geometric position of bars, line segments, or points to axis coordinates using linear or piecewise-linear transformations derived from the detected axes and tick marks. In addition, these chart types typically follow standardized layouts, with uniformly spaced categories or continuous axes, which simplifies the alignment between visual elements and their corresponding tabular representations. By contrast, scatterplots, stacked charts, circular charts, and mixed chart compositions receive substantially less attention, as indicated by the lower prevalence of end-to-end reconstruction-oriented systems evaluated on benchmarks such as PlotQA Methani et al. (2020) and ChartQA Masry et al. (2022). These chart types introduce challenges such as overlapping elements, variable marker density, non-uniform spacing, and non-linear or implicit scaling, which complicate both the mapping between visual coordinates and reconstructed table values and the reconciliation of visual elements with textual descriptors. Consequently, reconstruction performance varies considerably across chart categories, and reported accuracy often depends on clean layouts and explicitly defined axis structures.

The results also document substantial variability in the representation formats produced by chart reconstruction systems. Some approaches output simple listings of data pairs or value sequences Ramos et al. (2020); Methani et al. (2020), while others attempt to preserve richer structural information, such as multi-series organization, grouped bar relationships, or explicit associations between visual marks and axis semantics Shivasankaran et al. (2023); Nugumanova et al. (2022). However, these output formats are rarely standardized or formally specified, which complicates their integration with downstream tabular extraction pipelines, analytical workflows, or semantic data integration processes.

This lack of standardization also represents a limitation for knowledge graph generation, as reconstructed chart data are typically not exposed using machine-interpretable schemas or semantic representations. Although standards for tabular data publication and interchange exist, such as the W3C CSV on the Web recommendation, which defines metadata models to describe the structure, semantics, and provenance of CSV data, these standards are not explicitly adopted by the chart reconstruction systems reviewed in this study. As a result, the transformation of reconstructed chart data into interoperable representations suitable for semantic integration or knowledge graph population remains largely unaddressed in current end-to-end chart extraction pipelines.

Overall, the evidence collected for RQ1e shows that chart-to-table reconstruction remains a technically demanding but underrepresented task within chart analysis. Existing systems demonstrate that numerical recovery from charts is feasible, yet their performance remains limited by narrow chart-type coverage, sensitivity to visual irregularities, and inconsistencies in evaluation procedures and output formats. While promising progress has been made in controlled settings, the field has not yet established unified methodologies or robust benchmarks capable of supporting comprehensive, cross-domain reconstruction of chart-derived datasets. Recent advances in large language models and multimodal vision–language models suggest a promising direction for addressing these limitations, as such models are increasingly capable of jointly reasoning over visual content, textual context, and implicit structural patterns. However, within the reviewed corpus, the application of LLM-based approaches to full chart-to-table reconstruction remains largely unexplored, indicating a clear opportunity for future work at the intersection of chart understanding and structured data reconstruction.

RQ1f: How do existing methods identify and interpret variables within tables and charts?

Our results show that variable and label identification, understood as the detection and localization of visible textual elements, is broadly supported across the surveyed systems. As summarized in Table 4, the majority of reviewed approaches include components for identifying table headers, row or column labels, axis titles, legend entries, and tick labels. This capability is well established in both table- and chart-oriented systems, supported by mature OCR and layout analysis techniques. For example, ACciRo Daggubati and Sreevalsan-Nair (2022) reports a success rate of 96% (F1-score > 0.8) for variable and axis identification on circular charts using the Chart-AC and Chart-FC benchmarks Daggubati and Sreevalsan-Nair (2022), while ChartOCR Luo et al. (2021) achieves a similarity score of 0.962 on datasets such as ExcelChart400K Luo et al. (2021), FQA Methani et al. (2020), and WebData Luo et al. (2021) for label and text extraction. Similarly, interactive systems such as UNCHARTIT Ramos et al. (2020) report bar chart extraction accuracies above 90% when evaluated using WebPlotDigitizer benchmarks Ramos et al. (2020), indicating reliable identification of chart variables and associated labels. However, despite the widespread ability to locate textual elements, only a limited subset of systems explicitly attempt to interpret variables in a way that connects labels to the underlying data values they describe. As reflected in Table 4, fewer than one-third of the reviewed works support table–chart linking or variables and values association. Systems that go beyond detection to perform such interpretation include UNCHARTIT Ramos et al. (2020), TTC-QuAli Dong et al. (2024b), MATVIX Khalighinejad et al. (2025), and MMSci Li et al. (2024), which explicitly align visual elements, textual descriptors, and extracted numeric values. This distinction highlights that, while variable identification is largely resolved at the detection level, variable interpretation and reconciliation with data values remain comparatively underexplored.

Most systems address variable identification as a standalone step, detecting header regions or axis labels without modeling how these labels correspond to individual data values. In the table domain, works such as PubTables-1M Smock et al. (2022) and related table extraction systems accurately identify header rows and labeled regions but often do not establish explicit mappings between header text and the individual cell values beneath them, particularly in tables with multi-level headers or merged regions. Similarly, chart extraction methods referenced in the study, including PlotQA Methani et al. (2020) and ChartQA Masry et al. (2022), detect axes, legends, and text labels but face challenges when linking these labels to the correct visual elements in the chart, especially when multiple variables share similar colors, line styles, or marker shapes.

Some systems provide partial solutions by incorporating axis parsing, legend interpretation, or structured header analysis. For example, PubTables-1M Smock et al. (2022) includes annotations that distinguish between table regions, and ChartQA Masry et al. (2022) uses OCR and alignment strategies to associate textual elements with numerical values. However, these processes remain bounded to the specific chart types or table structures represented in the datasets on which the systems are trained. Other works in the study, such as ChartInfo Ma et al. (2021) and LineEX Shivasankaran et al. (2023), incorporate procedures to align detected labels with graphical marks, yet such alignment depends heavily on visual regularity and is sensitive to layout variation.

The difficulty of connecting variables to their corresponding values emerges recurrently across the systems compared in the reviewed corpus, rather than from explicit error analyses reported by a single study. Several common factors can be observed when contrasting system designs and reported limitations. First, variation in table formatting or

chart design introduces ambiguity, as variables may appear in multiple header regions, multi-level column headings, or legend entries, while their associated values may be distributed across different visual blocks or axes Smock et al. (2022); Masry et al. (2022). Second, textual and visual cues are frequently spatially separated, requiring alignment mechanisms that go beyond simple geometric proximity. This challenge is explicitly discussed in chart extraction systems such as UNCHARTIT Ramos et al. (2020), ChartOCR Luo et al. (2021), and ScatterPlotAnalyzer Dadhich et al. (2021), where mismatches between OCR outputs and detected visual elements are reported as a common source of error. Third, inconsistencies in units, abbreviations, and scientific notation further complicate variables and values association, particularly in scientific documents. Such issues are noted in systems targeting scientific charts and tables, including ChartInfo Ma et al. (2021), PlotQA Methani et al. (2020), and TTC-QuAli Dong et al. (2024b), where correct interpretation requires normalizing units or resolving abbreviated variable names across modalities.

Another limitation is that current datasets and benchmarks rarely explicitly annotate correspondences between variables and values. While resources such as PubTables-1M Smock et al. (2022) and ChartQA Masry et al. (2022) provide detailed structural or textual labels, they do not consistently define how each variable relates to its corresponding numerical cells or plotted points. As a result, most systems report accuracy metrics for isolated tasks, such as header identification or axis detection, rather than evaluating the correctness of complete variables and values interpretations.

At the same time, it is important to acknowledge that the problem of variable reconciliation and interoperability across tables has been extensively studied in the Semantic Web and data integration communities. Prior work has explored mapping table variables to shared conceptual identifiers, ontologies, or knowledge bases in order to determine whether tables from different sources describe the same underlying concepts and can be meaningfully compared or integrated Limaye et al. (2010); Venetis et al. (2011); Ritze et al. (2016); Wang et al. (2021a). More recent surveys further systematize these efforts by categorizing semantic table interpretation tasks and methods aimed at knowledge graph construction and cross-table integration Liu et al. (2023a). In this context, initiatives such as I-ADOPT Magagna et al. (2022) propose a community-driven framework for harmonizing the naming and conceptualization of observable properties across scientific domains. The goal is to reduce fragmentation caused by proliferating, poorly aligned vocabularies for measured, simulated, counted, or observed quantities, thereby improving interoperability and reuse of tabular data. However, these approaches typically assume that variables and values have already been extracted into structured representations and are therefore rarely integrated into the end-to-end table and chart extraction pipelines reviewed in this study, which primarily focus on visual detection, layout reconstruction, and low-level data extraction.

A remaining challenge for the Semantic Web community is to connect the structured representation of tables and charts with the broader context of the scientific article. Existing approaches support the semantic representation and integration of tabular content and scientific results in knowledge graphs Liu et al. (2023a); Kabongo et al. (2024). However, the reviewed literature provides limited support for explicitly linking extracted variables, values, and units to the sections, methodological descriptions, experimental settings, captions, and claims that provide their meaning and provenance. Addressing these cross-section and cross-modal links would enable knowledge graphs to preserve not only the extracted results, but also the scientific context in which those results were produced, reported, and interpreted.

Overall, the evidence collected for RQ1f indicates that variable detection is supported by the vast majority of reviewed systems, with 66 out of 68 works including mechanisms for identifying variable names through table headers, axis titles, legend entries, or tick labels (Table 4). In contrast, explicit variable interpretation remains limited. While most approaches can reliably localize variable names, they rarely establish complete mappings that connect these identifiers to their corresponding values in tables or charts. This does not imply that variable reconciliation is entirely absent. A small subset of systems, including UNCHARTIT Ramos et al. (2020), TTC-QuAli Dong et al. (2024b), MATVIX Khalighinejad et al. (2025), and MMSci Li et al. (2024), explicitly address variables and values association or multimodal alignment. The limited adoption of these mechanisms constrains higher-level analysis, including cross-table comparison, trend extraction, and integration across visual and textual data sources.

RQ2a: How do current methods identify and align related information across tables and charts within the same scientific publication?

The analysis of the annotated corpus indicates that explicit alignment between tables and charts remains relatively uncommon. As summarized in Table 4, only 18 of the 68 reviewed systems provide mechanisms that may support associations between these two modalities.

Table–chart alignment is relevant because scientific publications frequently use tables and charts to present complementary views of the same experimental results, measurements, or variables, even when the relationship between them is not explicitly stated. Identifying these correspondences can support variable reconciliation, consistency checking, and evidence integration across representations.

Among the reviewed approaches, some vision–language systems jointly encode textual and visual information. Systems such as ChartReader Cheng et al. (2023) and ChartLlama Han et al. (2025) illustrate how axis labels, legends, captions, and visual marks can be processed within a shared representation. Although these capabilities are relevant to table–chart alignment, the reviewed systems generally do not establish explicit correspondences between individual table values and specific chart elements. Their associations are more commonly inferred from captions, surrounding text, document layout, or semantic similarity.

Other approaches rely on heuristic or rule-based mechanisms. For example, chart extraction systems may associate axis labels and legend entries obtained through OCR with variables or textual mentions using string matching, spatial

proximity, or lexical similarity Masry et al. (2022); Shivasankaran et al. (2023). These strategies are effective when labels and graphical elements follow regular layouts, but they become less reliable when labels are abbreviated, OCR outputs are incomplete, several data series have similar visual attributes, or charts contain complex axes and legends.

In document-level pipelines, alignment is often an auxiliary capability rather than a dedicated modeling objective. Text embeddings and joint vision–language representations can associate figure captions, table titles, and variable mentions located in different parts of a document Cheng et al. (2023); Sun et al. (2025). However, these associations usually operate at the document, region, or caption level and do not define verifiable links between individual table cells and chart marks. Interactive and semi-automated systems such as ChartSense Jung et al. (2017), CrowdChart Chai et al. (2021), and ChartDetective Masson et al. (2023) similarly depend on user guidance or visual cues instead of fully automated table–chart correspondence models.

Agentic and question-answering architectures are also relevant because they demonstrate document-level reasoning over text, tables, and figures. For instance, DocAgent Sun et al. (2025) combines information from multiple document elements to answer questions about scientific publications. Nevertheless, these systems typically rely on implicit relationships inferred from text, captions, or layout rather than producing explicit and inspectable mappings between table entries and chart elements. Within the reviewed corpus, we did not identify agentic LLM systems specifically designed to establish and evaluate verifiable table–chart correspondences.

The difficulty of this task arises from the different representational structures of tables and charts. Tables organize values through rows, columns, headers, and cells, whereas charts encode information through position, shape, color, scale, and other visual properties. Alignment therefore requires connecting textual variables and units to graphical marks and subsequently linking those marks to the corresponding numerical values. Differences in layout, notation, label placement, scale, and visual design introduce additional ambiguity.

Some recent multimodal resources and systems, including MMSci Li et al. (2024), MATVIX Khalighinejad et al. (2025), and TTC-QuAli Dong et al. (2024b), provide joint representations of text, tables, charts, or visually rich document elements. However, their alignment capabilities generally remain at the document, quantity, region, or caption level rather than establishing complete mappings between specific table values and chart marks.

Progress is also constrained by the limited availability of benchmarks and evaluation metrics designed specifically for table–chart alignment. Existing resources commonly evaluate tables and charts independently, using metrics for OCR accuracy, component detection, structure recognition, value reconstruction, or chart question answering Luo et al. (2021); Smock et al. (2022); Methani et al. (2020). Such metrics do not directly assess whether a table variable, value, and unit have been correctly associated with their corresponding chart series, axis, legend entry, or visual mark.

Based on the integration capability considered in the evaluation framework, the approaches observed in the reviewed corpus can be analyzed through four table–chart alignment strategies. Lexical alignment associates captions, headers, axis labels, legends, and textual mentions through string matching or embedding similarity. Structural alignment connects table rows, columns, or cells with chart marks using spatial organization, ordering, scales, and layout relationships. Numerical consistency checking compares values extracted from tables with values reconstructed from charts. Finally, semantic alignment represents variables, values, units, provenance, and their relationships in structured forms suitable for downstream reuse or knowledge graph construction.

Most reviewed systems remain at the lexical or coarse document-level alignment stage. Structural alignment is generally restricted to specific chart types and regular layouts, while numerical consistency checking depends on accurate value reconstruction and compatible units and scales. Semantic alignment remains rarely implemented in end-to-end extraction systems.

Table 14 summarizes these strategies according to the evidence they use and their main limitations. It further develops the integration capability summarized in Table 6 by showing how relationships between tables and charts are established, or remain unresolved, in the reviewed systems.

Discussion

Before discussing the broader implications of the review, Table 15 provides a synthesized answer to each research question and links it to the supporting evidence reported in the Results and RQ-based analysis sections. This summary makes explicit how the findings address each RQ and serves as a bridge between the detailed analysis in RQ-Based Literature Review and Analysis and the thematic discussion that follows.

Building on the synthesized answers in Table 15, the following discussion focuses on the cross-cutting challenges that emerge across the research questions. A central challenge when extracting knowledge from scientific publications derives from the structural heterogeneity of scientific tables and charts. Tables and charts differ fundamentally in the way they encode quantitative information Smock et al. (2022); Masry et al. (2022). Tables express values explicitly through textual and numeric cells, whereas charts rely on visual encodings such as lines, points, and bars to convey trends or distributions Ma et al. (2021). Most existing systems are designed to process only one of these representations, resulting in separate methodological branches for table extraction and chart extraction. This separation has produced specialized progress within each modality but limited transfer of techniques, shared components, or representational frameworks across them Zheng et al. (2021).

A key finding emerging from the synthesis of RQ1 is that many approaches emphasize visual and structural processing stages such as grid segmentation, axis identification, and bounding-box extraction while giving comparatively less attention to the interpretation of the variables that these structures express Yang et al. (2022). While such structural recognition is necessary, it is not sufficient

Table 14. Comparison of table–chart alignment strategies

Alignment strategy	Description	Typical evidence used	Main limitation
Lexical alignment	Associates textual descriptors across tables and charts, including headers, captions, axis labels, legends, and variable mentions.	String overlap, OCR text, captions, titles, surrounding text, and embedding similarity.	Sensitive to abbreviations, synonyms, OCR errors, and inconsistent terminology.
Structural alignment	Associates table rows, columns, or cells with chart marks such as bars, points, or lines using layout and visual structure.	Spatial proximity, ordering, axis scales, visual marks, and row or column structure.	Usually restricted to specific chart types and regular layouts.
Numerical consistency checking	Compares values extracted from tables with values reconstructed from charts to identify agreement or inconsistency.	Cell values, digitized chart values, units, scales, and quantitative correspondences.	Requires accurate value reconstruction and compatible units and scales.
Semantic alignment	Represents correspondences between variables, values, units, and provenance in a structured form for downstream reuse.	Variable–value mappings, units, entity links, semantic identifiers, and provenance.	Rarely implemented and evaluated in end-to-end table and chart extraction systems.

Table 15. Summary of answers to each research question

RQ	Synthesized answer	Supporting evidence	Main remaining gap
RQ1a	Current systems cover a broad range of table and chart extraction tasks, including detection, structure recognition, component extraction, OCR-based label recognition, data reconstruction, and chart question answering. However, most systems remain modality-specific.	Table 2; Table 9; Table 10.	Limited joint processing of tables and charts.
RQ1b	Table extraction relies on CNN-based, Transformer-based, GNN-based, clustering-based, hybrid, and multimodal LLM/VLM-based systems, each supporting different stages of the extraction pipeline.	Table 5; Table 11; Table 6.	Incomplete architectural reporting and limited semantic output.
RQ1c	PDF-to-table conversion is mainly based on visual layout analysis, PDF text and layout parsing, rule-based processing, end-to-end extraction pipelines, page layout refinement, and, less frequently, clustering-based structure recovery.	Table 12.	Robust conversion of irregular tables into complete machine-readable representations.
RQ1d	Chart extraction systems commonly combine OCR, visual component detection, data point digitization, chart-type specific processing, heuristic alignment, and increasingly multimodal vision–language models.	Figure 3; Table 13; Table 8.	Generalization across chart types and irregular visual designs.
RQ1e	Chart-to-table reconstruction is feasible for common chart types, especially bar and line charts, but remains underrepresented as a full end-to-end task and lacks standardized output formats.	Figure 4; Table 13.	Standardized reconstruction outputs and evaluation of reconstructed values.
RQ1f	Most systems identify visible variables and labels, such as table headers, row or column labels, axis titles, and legends. However, fewer systems establish explicit mappings between variables and their corresponding values.	Table 4.	Semantic interpretation of variables, units, and value associations.
RQ2a	Explicit table–chart alignment remains uncommon. Existing systems mostly rely on lexical matching, captions, document context, heuristic associations, or coarse multimodal representations rather than verifiable variable-level mappings.	Table 4; Table 14.	Verifiable cross-modal alignment at variable and value level.

to determine what the extracted values represent, which units they use, or how they relate to the experimental context described in the publication. As a result, systems that rely primarily on geometric heuristics, OCR-based extraction, or layout templates often degrade when faced with complex or irregular scientific tables and charts,

particularly those embedded in multi-column PDF layouts or documents lacking consistent structural cues Paliwal et al. (2019). More importantly, this imbalance between geometric precision and contextual interpretation has direct implications for knowledge integration. Without explicit identification of variables, units, and semantic relationships, the extracted outputs cannot be reliably mapped to formal representations such as ontologies or knowledge graphs. As a consequence, although many systems succeed at producing visually accurate reconstructions of tables or charts, their outputs remain difficult to integrate into KG-based infrastructures for scientific data integration, comparison, and reuse. Addressing this gap requires moving beyond purely structural extraction toward representations that explicitly capture variable semantics and relationships, which is a prerequisite for automated KG generation from scientific publications.

Moreover, most pipelines process textual and visual features separately, preventing models from learning how these features relate to each other. Without mechanisms that explicitly connect visual cues (e.g., bar height or point location) to textual descriptors (e.g., axis labels or headers), models cannot establish explicit and verifiable multimodal associations. This limitation is particularly problematic in scientific publications, where tables and charts frequently present complementary information about the same variables or experimental settings, rather than identical datasets. For example, a table may report precise numerical measurements, while a chart visualizes trends, distributions, or comparisons derived from those values. Relating these complementary representations is essential to support variable reconciliation, consistency checking, and integrated analysis across modalities. While large language models operating over text can sometimes infer such relationships implicitly, the reviewed systems rarely provide explicit, structured representations that enable consistent and verifiable reasoning across visual and textual formats.

Reproducibility further complicates the landscape. As both RQ1 and RQ2 indicate, many studies omit architectural details or rely on proprietary datasets, making it difficult to replicate experiments or compare systems. Even when resources are released, inconsistencies in annotation schemes, preprocessing steps, and evaluation scripts hinder comparability. The absence of standardized protocols slows cumulative progress and restricts the ability to benchmark multimodal systems reliably Masry et al. (2022).

Interpretability represents an additional underdeveloped area, giving rise to two recurring challenges:

- **Lack of transparent reasoning.** Most systems function as black boxes: they produce structured outputs but do not explain how variables, labels, and numeric values are connected during the extraction process.
- **Limited traceability of extracted values.** Since extracted results are often used for downstream scientific synthesis or meta-analysis, the absence of explicit alignment evidence makes it difficult to verify whether an extracted value faithfully corresponds to a specific table cell or chart element.

Overcoming these challenges requires shifting from isolated, task-specific models toward unified frameworks capable of reasoning jointly over visual and textual cues. Based on the findings from RQ1 and RQ2, several strategies may guide this transition:

Unified multimodal architectures. End-to-end architectures that jointly encode visual and textual inputs represent a promising direction. Transformer-based models that process image regions, OCR text, and layout cues within a shared representation allow visual elements (e.g., table cells or chart marks) and textual descriptors (e.g., headers or axis labels) to be modeled together, improving robustness to layout variability. Recent advances in multimodal and vision–language large language models (LLMs), such as ChartQA Masry et al. (2022), ChartLlama Han et al. (2025), and TinyChart Zhang et al. (2024), further extend this paradigm by enabling higher-level reasoning over tables and charts through joint visual–textual representations. While LLM-based systems improve flexibility, summarization, and question answering, they typically operate over pre-extracted visual and textual inputs and rely on implicit associations learned from captions, surrounding text, or instruction tuning, rather than enforcing explicit alignments between variables, values, and their visual encodings across tables and charts. As a result, LLMs currently complement but do not replace structured multimodal encodings that explicitly model relationships between visual elements, textual variables, and numeric values. Agentic frameworks based on LLMs constitute a promising extension of this line of work. By combining multimodal perception modules with iterative reasoning, tool invocation, and verification steps, agentic systems may potentially support explicit table–chart alignment, hypothesis testing, and consistency checking across representations.

Multimodal training data and benchmarks. Progress in multimodal data extraction is strongly influenced by the availability of datasets that jointly represent tables, charts, and surrounding text. While high-quality benchmarks exist for individual subtasks, such as table structure recognition with PubTabNet Zhong et al. (2019) or chart question answering with ChartQA Masry et al. (2022) and PlotQA Methani et al. (2020), these resources typically evaluate each modality in isolation. The main gap is therefore not the absence of benchmarks in general, but the lack of datasets that explicitly link tables and charts as complementary representations of the same variables, values, units, or experimental conditions within a document. This limits the ability of models to learn and evaluate cross-modal correspondences, even when they are pretrained on visually rich scientific documents or synthetic chart data.

Evaluation and reproducibility. Current evaluation practices remain focused on component-level metrics such as detection accuracy, OCR performance, structure recognition scores, or chart question answering accuracy. These metrics are useful for assessing individual extraction tasks, but they do not measure whether extracted variables, values, and relationships are consistently aligned across tables and charts. At the same time, reproducibility depends on clear reporting of preprocessing steps, dataset splits, model configurations, OCR settings, parsing parameters, and scoring procedures. Therefore, future evaluations should

combine stronger reporting standards with metrics that assess multimodal alignment, data fidelity, and cross-representation consistency.

Actionable benchmark and metric requirements. The findings of this review suggest that future benchmarks should evaluate capabilities that are not captured by current table-only or chart-only datasets. A new benchmark for multimodal metadata extraction should include scientific documents containing paired tables and charts that describe related variables, values, units, or experimental conditions. Each instance should provide annotations at multiple levels: (i) document-level links between related tables, charts, captions, and surrounding text; (ii) component-level annotations for table cells, headers, chart axes, legends, visual marks, and tick labels; (iii) variable–value mappings that specify which header, axis, legend, or unit gives meaning to each numerical value; and (iv) provenance annotations indicating the exact table cell, chart mark, or document region from which each extracted value was obtained. Such a benchmark would make it possible to evaluate not only whether a table or chart component was detected, but also whether the extracted data are semantically grounded and correctly aligned across modalities.

In addition to new datasets, future evaluation protocols should include metrics that capture correctness beyond component detection or OCR accuracy. Examples include *variable–value mapping accuracy*, which measures whether each extracted numerical value is linked to the correct variable, header, axis, or legend; *unit normalization accuracy*, which evaluates whether units are correctly detected, normalized, and associated with values; *cross-modal consistency*, which measures whether values reconstructed from charts agree with corresponding values reported in tables within an acceptable numerical tolerance; *provenance accuracy*, which assesses whether each extracted value can be traced back to the correct table cell, chart mark, or document region; and *semantic alignment accuracy*, which evaluates whether extracted variables, units, and values are mapped to appropriate ontology terms or knowledge graph entities. These metrics would support evaluation at the metadata and semantic levels, rather than only at the level of bounding boxes, text recognition, or reconstructed CSV/JSON outputs. In particular, semantic alignment accuracy should assess whether extracted variables, values, units, and contextual information can be reliably connected to semantic identifiers and represented using existing Semantic Web approaches for semantic table interpretation, RDF generation, ontology alignment, and knowledge graph construction Fiorelli and Stellato (2021); Liu et al. (2023a); Nguyen et al. (2024); Limaye et al. (2010); Venetis et al. (2011).

Knowledge Integration and Linking.

The analysis across RQ1 and RQ2 points toward the need for integrated document understanding systems that reason jointly over tables, charts, and surrounding text. Rather than detecting structures in isolation, future systems should recover the meaning of the data itself, including variables, measurement units, quantitative relationships, etc., and link them in structured representations. This would enable models to infer, for example, that a table reporting "Accuracy across models" and a corresponding chart visualizing the same models refer to the same underlying experiment.

Related efforts in the scientific knowledge management community already move in this direction. For example, the Open Research Knowledge Graph (ORKG) and recent extensions such as ORKG-Leaderboards aim to represent and compare empirical research results in a machine-actionable form Kabongo et al. (2024). ORKG-Leaderboards focus on the automated extraction of leaderboard-style results, defined as task–dataset–metric tuples, from LaTeX and PDF papers, achieving high extraction accuracy in this constrained setting. The extracted results are then integrated into the ORKG to support comparison and aggregation across publications. However, this workflow targets a specific and well-structured class of results and relies on predefined schemas and task formulations. More broadly, the integration of arbitrary tables, charts, and heterogeneous experimental results into knowledge graphs remains largely dependent on manual curation or semi-automated processes.

Bridging this gap requires tighter integration between multimodal extraction pipelines and structured knowledge representations. Automated alignment of variables (e.g., with structured representations such as I-ADOPT Magagna et al. (2022)), values, and experimental contexts across tables and charts would reduce manual effort and enable scalable population of research knowledge bases.

From a Semantic Web perspective, this challenge also concerns the semantic grounding of extracted table and chart content. Existing work on semantic table interpretation, entity linking, type annotation, and semantic annotation demonstrates how already structured tabular data can be enriched and integrated into knowledge graphs Liu et al. (2023a); Nguyen et al. (2024); Limaye et al. (2010); Venetis et al. (2011); Ritze et al. (2016). Initiatives such as ORKG-Leaderboards further show how extracted scientific results can be organized as machine-actionable knowledge for comparison across publications Kabongo et al. (2024). However, the reviewed extraction systems usually produce CSV, JSON, bounding boxes, reconstructed tables, or textual answers without consistently linking these outputs to semantic identifiers, provenance, experimental settings, or the sections and claims that provide their scientific meaning. Addressing this gap requires tighter integration between document-level extraction, semantic annotation, and knowledge graph construction.

Overall, the future of multimodal data extraction lies in moving beyond geometry-driven perception toward reasoning-based understanding. End-to-end multimodal architectures, combined with robust alignment mechanisms and appropriate evaluation criteria, will become essential for enabling systems that can support scientific comparison, verification, and large-scale evidence synthesis.

Conclusion

This study presents a comprehensive and comparative review of information extraction methods that address both tabular and chart information in scientific publications. Guided by explicit research questions and a structured evaluation framework, we systematically analyze 68 peer-reviewed studies to identify methodological trends, research challenges, and emerging directions in variable identification, multimodal alignment, and system interoperability.

Our findings show that while structural detection and layout analysis are well developed, higher-level data interpretation and integration across modalities are still limited. Differences in dataset construction, evaluation protocols, and resource accessibility continue to affect reproducibility and hinder comparison between studies.

Future research should prioritize the development of multimodal architectures that jointly process visual, textual, and structural information within unified learning frameworks. Such systems should move beyond isolated detection tasks toward integrated reasoning over tables, charts, and accompanying text, enabling consistent interpretation of variables, values, and relationships across representations. Equally important is the design of explainable and flexible extraction systems capable of adapting to the diversity of scientific publications while tracking the provenance of the obtained results for transparency. More specifically, benchmarks should move beyond isolated table or chart detection and include annotations for variable–value mappings, units, provenance, and cross-modal correspondences between tables and charts. Likewise, evaluation metrics should assess whether extracted values are not only visually detected or numerically reconstructed, but also correctly grounded in their metadata and reusable in semantic or knowledge graph-oriented representations.

Our study identifies the main challenges that currently limit multimodal document analysis and outlines concrete directions for future research. In particular, we highlight the need for improved integration across tables and charts, more robust alignment of variables and values, and stronger evaluation practices. Recent efforts point toward promising directions, including multimodal scientific benchmarks such as MMSci Li et al. (2024), which jointly model text, tables, and figures; domain-specific multimodal extraction systems such as MATVIX Khalighinejad et al. (2025), designed for materials science documents; task-oriented frameworks such as TTC-QuAli Dong et al. (2024b), which focus on quantitative alignment across textual, tabular, and graphical elements; and knowledge-oriented infrastructures such as the Open Research Knowledge Graph (ORKG) Kabongo et al. (2024), which aim to represent and integrate structured scientific knowledge derived from heterogeneous sources.

Acknowledgments

The work was supported by the Predoctoral Grant PIPF-2024/COM-35132 of the Consejería de Educación, Ciencia y Universidades de la Comunidad de Madrid, Spain.

References

Al-Zaidy R and Giles CL (2017) A machine learning approach for semantic structuring of scientific charts in scholarly documents. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31. pp. 4644–4649. DOI:10.1609/aaai.v31i2.19088. URL https://doi.org/10.1609/aaai.v31i2.19088.

Alampara N, Mandal I, Khetarpal P, Grover HS, Schilling-Wilhelmi M, Krishnan NMA and Jablonka KM (2024) Macbench: A multimodal chemistry and materials science benchmark. In: *AI for Accelerated Materials Design, NeurIPS 2024 Workshop*. URL https://openreview.net/forum?id=Q2PNocDcp6.

Arshad A, Moetesum M, Ul Hasan A and Shafait F (2024) Enhancing multimodal information extraction from visually rich documents with 2d positional embeddings. In: *Proceedings of the International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. pp. 561–568. DOI:10.1109/DICTA63115.2024.00087. URL https://doi.org/10.1109/DICTA63115.2024.00087.

Balsiger D, Dimmler HR, Egger-Horstmann S and Hanne T (2024) Assessing large language models used for extracting table information from annual financial reports. *Computers* 13: 257. DOI:10.3390/computers13100257. URL https://doi.org/10.3390/computers13100257.

Berenguer A, Mazón JN and Tomás D (2024) Word embeddings for retrieving tabular data from research publications. *Machine Learning* 113(4): 2227–2248. DOI:10.1007/s10994-023-06472-0. URL https://doi.org/10.1007/s10994-023-06472-0.

Bornmann L, Haunschild R and Mutz R (2021) Growth rates of modern science: A latent piecewise growth curve approach to model publication numbers from established and new literature databases. *Humanities and Social Sciences Communications* 8(1): 224. DOI:10.1057/s41599-021-00903-w. URL https://doi.org/10.1057/s41599-021-00903-w.

Boutalbi K, Sylejmani V, Dardouillet P, Le Van O, Salamatian K, Verjus H, Loukil F and Telisson D (2023) A clustering approach combining lines and text detection for table extraction. In: *Document Analysis and Recognition – ICDAR 2023 Workshops, Proceedings, Part II*. pp. 139–145. DOI:10.1007/978-3-031-41501-2_10. URL https://doi.org/10.1007/978-3-031-41501-2_10.

Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko S (2020) End-to-end object detection with transformers. In: *Computer Vision – ECCV 2020*, Lecture Notes in Computer Science. Springer, pp. 213–229. DOI:10.1007/978-3-030-58452-8_13. URL https://doi.org/10.1007/978-3-030-58452-8_13.

Cedeño E and Garijo D (2025) Metadata extraction – slr. DOI:10.5281/zenodo.18072492. URL https://doi.org/10.5281/zenodo.18072492. Archived GitHub repository and supplementary materials.

Chai C, Li G, Fan J and Luo Y (2021) Crowdchart: Crowdsourced data extraction from visualization charts. *IEEE Transactions on Knowledge and Data Engineering* 33(11): 3537–3549. DOI:10.1109/TKDE.2020.2972543. URL https://doi.org/10.1109/TKDE.2020.2972543.

Chen C, Lee B, Wang Y, Chang Y and Liu Z (2024) Mystique: Deconstructing svg charts for layout reuse. *IEEE Transactions on Visualization and Computer Graphics* 30(1): 447–457. DOI:10.1109/TVCG.2023.3327354. URL https://doi.org/10.1109/TVCG.2023.3327354.

Chen W, Wang H, Chen J, Zhang Y, Wang H, Li S, Zhou X and Wang WY (2020) Tabfact: A large-scale dataset for table-based fact verification. In: *Proceedings of the 8th International Conference on Learning Representations (ICLR 2020)*. Addis Ababa, Ethiopia: OpenReview. URL https://openreview.net/forum?id=rkeJRhNYDH.

Cheng ZQ, Dai Q and Hauptmann AG (2023) Chartreader: A unified framework for chart derendering and comprehension

without heuristic rules. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. pp. 22145–22156. DOI:10.1109/ICCV51070.2023.02029. URL https://doi.org/10.1109/ICCV51070.2023.02029.

Circi D, Khalighinejad G, Chen A, Dhingra B and Brinson L (2024) Extracting materials science data from scientific tables. In: *ACL 2024 Workshop on Language + Molecules*. URL https://openreview.net/forum?id=KzVJFYLCtt.

Dadhich K, Daggubati SC and Sreevalsan-Nair J (2021) Scatterplotanalyzer: Digitizing images of charts using tensor-based computational model. In: *Computational Science – ICCS 2021*. pp. 70–83. DOI:10.1007/978-3-030-77977-1_6. URL https://doi.org/10.1007/978-3-030-77977-1_6.

Dagdelen J, Dunn A, Lee S, Walker N, Rosen AS, Ceder G, Persson KA and Jain A (2024) Structured information extraction from scientific text with large language models. *Nature Communications* 15(1): 1418. DOI:10.1038/s41467-024-45563-x. URL https://doi.org/10.1038/s41467-024-45563-x.

Daggubati SC and Sreevalsan-Nair J (2022) Acciro: A system for analyzing and digitizing images of charts with circular objects. *Lecture Notes in Computer Science* 13352: 605–612. DOI:10.1007/978-3-031-08757-8_50. URL https://doi.org/10.1007/978-3-031-08757-8_50.

Daggubati SC, Sreevalsan-Nair J and Dadhich K (2022) Barchartanalyzer: Data extraction and summarization of bar charts from images. *SN Computer Science* 3(6). DOI:10.1007/s42979-022-01380-x. URL https://doi.org/10.1007/s42979-022-01380-x.

Dai W, Wang M, Niu Z and Zhang J (2018) Chart decoder: Generating textual and numeric information from chart images automatically. *Journal of Visual Languages & Computing* 48: 101–109. DOI:10.1016/j.jvlc.2018.08.005. URL https://doi.org/10.1016/j.jvlc.2018.08.005.

Del Bimbo D, Gemelli A and Marinai S (2022) Data augmentation on graphs for table type classification. In: *Structural, Syntactic, and Statistical Pattern Recognition (S+SSPR 2022)*. Montreal, QC, Canada: Springer, pp. 242–252. DOI:10.1007/978-3-031-23028-8_25. URL https://doi.org/10.1007/978-3-031-23028-8_25.

Desai H, Kayal P and Singh M (2021) Tablex: A benchmark dataset for structure and content information extraction from scientific tables. In: *Document Analysis and Recognition – ICDAR 2021*, *Lecture Notes in Computer Science*, volume 12810. Springer, pp. 554–569. DOI:10.1007/978-3-030-86331-9_36. URL https://doi.org/10.1007/978-3-030-86331-9_36.

Dhote A, Javed M and Doermann DS (2023) A survey and approach to chart classification. In: *Document Analysis and Recognition – ICDAR 2023 Workshops*. San José, CA, USA: Springer, pp. 67–82. DOI:10.1007/978-3-031-41498-5_5. URL https://doi.org/10.1007/978-3-031-41498-5_5.

Dong H, Hu M, Xu Q, Wang H and Hu Y (2024a) Opente: Open-structure table extraction from text. In: *ICASSP 2024 – IEEE International Conference on Acoustics, Speech and Signal Processing*. pp. 10306–10310. DOI:10.1109/ICASSP48485.2024.10448427. URL https://doi.org/10.1109/ICASSP48485.2024.10448427.

Dong H, Wang H, Zhou A and Hu Y (2024b) Ttc-quali: A text-table-chart dataset for multimodal quantity alignment. In: *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*. pp. 181–189. DOI:10.1145/3616855.3635777. URL https://doi.org/10.1145/3616855.3635777.

Embley DW, Hurst M, Lopresti D and Nagy G (2006) Table-processing paradigms: A research survey. *International Journal of Document Analysis and Recognition* 8(2): 66–86. DOI:10.1007/s10032-006-0017-x. URL https://doi.org/10.1007/s10032-006-0017-x.

Farahani AM, Adibi P, Ehsani MS, Hutter HP and Darvishy A (2023) Automatic chart understanding: A review. *IEEE Access* 11: 76202–76221. DOI:10.1109/ACCESS.2023.3298050. URL https://doi.org/10.1109/ACCESS.2023.3298050.

Fauceglia N, Canim M, Gliozzo A, Liang J, Wang N, Burdick D, Mihindukulasooriya N, Castelli V, Feigenblat G, Konopnicki D, Katsis Y, Florian R, Li Y, Roukos S and Sil A (2021) Kaapa: Knowledge aware answers from pdf analysis. *Proceedings of the AAAI Conference on Artificial Intelligence* 35: 16029–16031. DOI:10.1609/aaai.v35i18.18002. URL https://doi.org/10.1609/aaai.v35i18.18002.

Fiorelli M and Stellato A (2021) Lifting tabular data to rdf: A survey. In: *Knowledge Graphs and Semantic Web*. Cham: Springer. ISBN 978-3-030-71902-9, pp. 85–96. DOI:10.1007/978-3-030-71903-6_9. URL https://doi.org/10.1007/978-3-030-71903-6_9.

Gemelli A, Vivoli E and Marinai S (2022) Graph neural networks and representation embedding for table extraction in pdf documents. In: *Proceedings of the 26th International Conference on Pattern Recognition (ICPR)*. pp. 1719–1726. DOI:10.1109/ICPR56361.2022.9956590. URL https://doi.org/10.1109/ICPR56361.2022.9956590.

Göbel M, Hassan T, Oro E and Orsi G (2013) Icdar 2013 table competition. In: *Proceedings of the 2013 12th International Conference on Document Analysis and Recognition*. USA: IEEE Computer Society, pp. 1449–1453. DOI:10.1109/ICDAR.2013.292. URL https://doi.org/10.1109/ICDAR.2013.292.

Goswami K, Mathur P, Rossi R and Dernoncourt F (2025) Chartcitor: Answer citations for chartqa via multi-agent llm retrieval. In: *Companion Proceedings of the ACM Web Conference 2025*. pp. 1668–1671. DOI:10.1145/3701716.3716886. URL https://doi.org/10.1145/3701716.3716886.

Gu C, Tian Y, Hou J and Li XY (2024) Ai-blueprint: A real-time system for automated identification and analysis of electrical blueprints. In: *Proceedings of the IEEE International Conference on Parallel and Distributed Systems (ICPADS)*. pp. 415–422. DOI:10.1109/ICPADS63350.2024.00061. URL https://doi.org/10.1109/ICPADS63350.2024.00061.

Han Y, Zhang C, Chen X, Yin F, Yang X, Wang Z, Yu G, Fu B and Zhang H (2025) A multimodal llm for chart understanding and generation. In: *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*. pp. 1–8. DOI:10.1109/IJCNN64981.2025.11227777. URL https://doi.org/10.1109/IJCNN64981.2025.11227777.

He K, Zhang X, Ren S and Sun J (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 770–778. DOI:10.1109/CVPR.2016.90. URL https:

//doi.org/10.1109/CVPR.2016.90.

Hogan A, Blomqvist E, Cochez M, D'Amato C, De Melo G, Gutierrez C, Kirrane S, Labra Gayo JE, Navigli R, Neumaier S, Ngonga Ngomo AC, Polleres A, Rashid SM, Rula A, Schmelzeisen L, Sequeda J, Staab S and Zimmermann A (2021) Knowledge graphs. *ACM Computing Surveys* 54(4). DOI:10.1145/3447772. URL https://doi.org/10.1145/3447772.

Horadi KV, Pujari J and Bhat NP (2022) Deep learning based model for table detection content and layout analysis in compressed document images: A comprehensive approach. *Journal of Theoretical and Applied Information Technology* 100(14): 5212–5222. DOI:10.5281/zenodo.6983145. URL https://doi.org/10.5281/zenodo.6983145.

Hu L, Wang D, Pan Y, Yu J, Shao Y, Feng C and Nie L (2024) Novachart: A large-scale dataset towards chart understanding and generation of multimodal large language models. In: *Proceedings of the 32nd ACM International Conference on Multimedia*. pp. 3917–3925. DOI:10.1145/3664647.3680790. URL https://doi.org/10.1145/3664647.3680790.

Huang KH, Chan HP, Fung M, Qiu H, Zhou M, Joty S, Chang SF and Ji H (2025) From pixels to insights: A survey on automatic chart understanding in the era of large foundation models. *IEEE Transactions on Knowledge and Data Engineering* 37(5): 2550–2568. DOI:10.1109/TKDE.2024.3513320. URL https://doi.org/10.1109/TKDE.2024.3513320.

Huang W and Tan CL (2007) Extraction of vectorized graphical information from scientific chart images. In: *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*. DOI:10.1109/ICDAR.2007.4378764. URL https://doi.org/10.1109/ICDAR.2007.4378764.

Iida H, Thai D, Manjunatha V and Iyyer M (2021) Tabbie: Pretrained representations of tabular data. In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. pp. 3446–3456. DOI:10.18653/v1/2021.naacl-main.270. URL https://doi.org/10.18653/v1/2021.naacl-main.270.

Iqbal MZ, Garg N and Ahmed SB (2025) Table extraction with table data using vgg-19 deep learning model. *Sensors* 25(1). DOI:10.3390/s25010203. URL https://doi.org/10.3390/s25010203.

Jain A, Paliwal S, Sharma M and Vig L (2021) Tsr-dsaw: Table structure recognition via deep spatial association of words. In: *Proceedings of the European Symposium on Artificial Neural Networks (ESANN)*. pp. 257–262. DOI:10.14428/esann/2021.ES2021-109. URL https://doi.org/10.14428/esann/2021.ES2021-109.

Jonnalagadda SR, Goyal P and Huffman MD (2015) Automating data extraction in systematic reviews: A systematic review. *Systematic Reviews* 4(1): 78. DOI:10.1186/s13643-015-0066-7. URL https://doi.org/10.1186/s13643-015-0066-7.

Jung D, Kim W, Song H, Hwang Ji, Lee B, Kim B and Seo J (2017) Chartsense: Interactive data extraction from chart images. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. DOI:10.1145/3025453.3025957. URL https://doi.org/10.1145/3025453.3025957.

Kabongo S, D'Souza J and Auer S (2024) Orkg-leaderboards: A systematic workflow for mining leaderboards as a knowledge graph. *International Journal on Digital Libraries* 25(1): 41–54. DOI:10.1007/s00799-023-00366-1. URL https://doi.org/10.1007/s00799-023-00366-1.

Kafle K, Price B, Cohen S and Kanan C (2018) Dvqa: Understanding data visualizations via question answering. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 5648–5656. DOI:10.1109/CVPR.2018.00592. URL https://doi.org/10.1109/CVPR.2018.00592.

Kantharaj S, Leong RT, Lin X, Masry A, Thakkar M, Hoque E and Joty S (2022) Chart-to-text: A large-scale benchmark for chart summarization. In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. pp. 4005–4023. DOI:10.18653/v1/2022.acl-long.277. URL https://doi.org/10.18653/v1/2022.acl-long.277.

Kasem MS, Abdallah A, Berendeyev A, Elkady E, Mahmoud M, Abdalla M, Hamada M, Vascon S, Nurseitov D and Taj-Eddin I (2024) Deep learning for table detection and structure recognition: A survey. *ACM Computing Surveys* 56(12). DOI:10.1145/3657281. URL https://doi.org/10.1145/3657281.

Khalighinejad G, Scott S, Liu O, Anderson KL, Stureborg R, Tyagi A and Dhingra B (2025) Matvix: Multimodal information extraction from visually rich articles. In: *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies*. pp. 3636–3655. DOI:10.18653/v1/2025.naacl-long.185. URL https://doi.org/10.18653/v1/2025.naacl-long.185.

Kitchenham BA, Brereton OP, Budgen D, Turner M, Bailey J and Linkman S (2009) Systematic literature reviews in software engineering: A systematic literature review. *Information and Software Technology* 51(1): 7–15. DOI:10.1016/j.infsof.2008.09.009. URL https://doi.org/10.1016/j.infsof.2008.09.009.

Kumar A and Ganu T (2023) Chartparser: Automatic chart parsing for print-impaired. In: *CEUR Workshop Proceedings*, volume 3656. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85189152061&partnerID=40&md5=84065eabaaabc25f6ca4c283ec18aa79. Conference paper.

Kwon H, An J, Lee D and Shin WY (2022) Data: Domain adaptation-aided deep table detection using visual–lexical representations. *Knowledge-Based Systems* 258: 109946. DOI:10.1016/j.knosys.2022.109946. URL https://doi.org/10.1016/j.knosys.2022.109946.

Lal J, Mitkari A, Bhosale M and Doermann D (2023) Lineformer: Line chart data extraction using instance segmentation. In: *Document Analysis and Recognition – ICDAR 2023*. pp. 387–400. DOI:10.1007/978-3-031-41734-4_24. URL https://doi.org/10.1007/978-3-031-41734-4_24.

Lehmberg O, Ritze D, Meusel R and Bizer C (2016) A large public corpus of web tables containing time and context metadata. In: *Proceedings of the 25th International Conference Companion on World Wide Web*, WWW '16 Companion. Republic and Canton of Geneva, CHE: International World Wide Web

Conferences Steering Committee. ISBN 9781450341448, p. 75–76. DOI:10.1145/2872518.2889386. URL https://doi.org/10.1145/2872518.2889386.

Li M, Cui L, Huang S, Wei F, Zhou M and Li Z (2020) Tablebank: Table benchmark for image-based table detection and recognition. In: *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, pp. 1918–1925. URL https://aclanthology.org/2020.lrec-1.236/.

Li Z, Yang X, Choi K, Zhu W, Hsieh R, Kim H, Lim JH, Ji S, Lee B, Yan X, Petzold LR, Wilson SD, Lim W and Wang WY (2024) Mmsci: A multimodal multi-discipline dataset for phd-level scientific comprehension. In: *AI for Accelerated Materials Design – Vienna 2024*. URL https://openreview.net/forum?id=gZJTkPXvkP.

Limaye G, Sarawagi S and Chakrabarti S (2010) Annotating and searching web tables using entities, types and relationships. *Proceedings of the VLDB Endowment* 3(1–2): 1338–1347. DOI:10.14778/1920841.1921005. URL https://doi.org/10.14778/1920841.1921005.

Liu J, Chabot Y, Troncy R, Huynh VP, Labbé T and Monnin P (2023a) From tabular data to knowledge graphs: A survey of semantic table interpretation tasks and methods. *Journal of Web Semantics* 76: 100761. DOI:10.1016/j.websem.2022.100761. URL https://www.sciencedirect.com/science/article/pii/S1570826822000452.

Liu Y, Zhang G and Shen T (2023b) A novel approach with fusion edge prediction strategy for table structure recognition. In: *Proceedings of the 11th International Conference on Information Systems and Computing Technology (ISCTech)*. pp. 46–50. DOI:10.1109/ISCTech60480.2023.00016. URL https://doi.org/10.1109/ISCTech60480.2023.00016.

Long R, Wang W, Xue N, Gao F, Yang Z, Wang Y and Xia GS (2021) Parsing table structures in the wild. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. pp. 924–932. DOI:10.1109/ICCV48922.2021.00098. URL https://doi.org/10.1109/ICCV48922.2021.00098.

Luo J, Li Z, Wang J and Lin CY (2021) Chartocr: Data extraction from chart images via a deep hybrid framework. In: *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*. pp. 1916–1924. DOI:10.1109/WACV48630.2021.00196. URL https://doi.org/10.1109/WACV48630.2021.00196.

Ma W, Zhang H, Yan S, Yao G, Huang Y, Li H, Wu Y and Jin L (2021) Towards an efficient framework for data extraction from chart images. In: *Document Analysis and Recognition – ICDAR 2021*, Lecture Notes in Computer Science. Springer, pp. 583–597. DOI:10.1007/978-3-030-86549-8_37. URL https://doi.org/10.1007/978-3-030-86549-8_37.

Magagna B, Moncoiffé G, Devaraju A, Stoica M, Schindler S, Pamment A and Group RIAW (2022) Interoperable descriptions of observable property terminologies (i-adopt): Working group outputs and recommendations. DOI:10.15497/RDA00071. URL https://doi.org/10.15497/RDA00071. Research Data Alliance (RDA).

Masry A, Kavehzadeh P, Do XL, Hoque E and Joty S (2023) Unichart: A universal vision-language pretrained model for chart comprehension and reasoning. In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Singapore: Association for Computational Linguistics, pp. 14662–14684. DOI:10.18653/v1/2023.emnlp-main.906. URL https://aclanthology.org/2023.emnlp-main.906/.

Masry A, Long DX, Tan JQ, Joty S and Hoque E (2022) Chartqa: A benchmark for question answering about charts with visual and logical reasoning. In: *Findings of the Association for Computational Linguistics: ACL 2022*. DOI:10.18653/v1/2022.findings-acl.177. URL https://doi.org/10.18653/v1/2022.findings-acl.177.

Masson D, Malacria S, Vogel D, Lank E and Casiez G (2023) Chartdetective: Easy and accurate interactive data extraction from complex vector charts. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. DOI:10.1145/3544548.3581113. URL https://doi.org/10.1145/3544548.3581113.

Meng F, Shao W, Lu Q, Gao P, Zhang K, Qiao Y and Luo P (2024) Chartassistant: A universal chart multimodal language model via chart-to-table pre-training and multitask instruction tuning. In: *Findings of the Association for Computational Linguistics: ACL 2024*. pp. 7775–7803. DOI:10.18653/v1/2024.findings-acl.463. URL https://doi.org/10.18653/v1/2024.findings-acl.463.

Methani N, Ganguly P, Khapra MM and Kumar P (2020) Plotqa: Reasoning over scientific plots. In: *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*. pp. 1516–1525. DOI:10.1109/WACV45572.2020.9093523. URL https://doi.org/10.1109/WACV45572.2020.9093523.

Meuschke N, Jagdale A, Spinde T, Mitrović J and Gipp B (2023) A benchmark of pdf information extraction tools using a multi-task and multi-domain evaluation framework for academic documents. In: *Information for a Better World: Normality, Virtuality, Physicality, Inclusivity: 18th International Conference, IConference 2023, Proceedings, Part II*. pp. 383–405. DOI:10.1007/978-3-031-28032-0_31. URL https://doi.org/10.1007/978-3-031-28032-0_31.

Mikhailov A and Shigarov A (2021) Page layout analysis for refining table extraction from pdf documents. In: *2021 Ivannikov Ispras Open Conference (ISPRAS)*. pp. 114–119. DOI:10.1109/ISPRAS53967.2021.00021. URL https://doi.org/10.1109/ISPRAS53967.2021.00021.

Mikhailov AA, Shigarov A and Kozlov IS (2021) Pytabby: A docreader's module for extracting text and tables from pdf with a text layer. In: *CEUR Workshop Proceedings*, volume 2984. pp. 120–126. URL https://ceur-ws.org/Vol-2984/.

Mishra P, Kumar S and Chaube MK (2022a) Evaginating scientific charts: Recovering direct and derived information encodings from chart images. *Journal of Visualization* 25(2): 343–359. DOI:10.1007/s12650-021-00800-z. URL https://doi.org/10.1007/s12650-021-00800-z.

Mishra P, Kumar S, Chaube MK and Shrawankar U (2022b) Chartvi: Charts summarizer for visually impaired. *Journal of Computer Languages* 69: 101107. DOI:10.1016/j.cola.2022.101107. URL https://doi.org/10.1016/j.cola.2022.101107.

Mustafa O, Ali MK, Moetesum M and Siddiqi I (2023) Charteye: A deep learning framework for chart information extraction. In:

2023 International Conference on Digital Image Computing: Techniques and Applications (DICTA). pp. 554–561. DOI: 10.1109/DICTA60407.2023.00082. URL https://doi.org/10.1109/DICTA60407.2023.00082.

Méndez GG, Nacenta MA and Vandenheste S (2016) ivolver: Interactive visual language for visualization extraction and reconstruction. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. pp. 4073–4085. DOI: 10.1145/2858036.2858435. URL https://doi.org/10.1145/2858036.2858435.

Nguyen P, Kertkeidkachorn N, Ichise R and Takeda H (2024) Mtab4d: Semantic annotation of tabular data with dbpedia. *Semantic Web* 15(6): 2613–2637. DOI:10.3233/SW-223098. URL https://journals.sagepub.com/doi/abs/10.3233/SW-223098.

Noy N, Gao Y, Jain A, Narayanan A, Patterson A and Taylor J (2019) Industry-scale knowledge graphs: Lessons and challenges. *Communications of the ACM* 62(8): 36–43. DOI: 10.1145/3331166. URL https://doi.org/10.1145/3331166.

Nugumanova AB, Apayev KS, Baiburin YM, Mansurova M and Ospan AG (2022) Qurma: A table extraction pipeline for knowledge base population. *Journal of Mathematics, Mechanics and Computer Science* 114(2): 92–103. DOI: 10.26577/JMMCS.2022.v114.i2.08. URL https://doi.org/10.26577/JMMCS.2022.v114.i2.08.

Paliwal S, Patwardhan M and Vig L (2023) Generalization of fine granular extractions from charts. In: *Document Analysis and Recognition – ICDAR 2023, Proceedings, Part II*. pp. 94–110. DOI:10.1007/978-3-031-41679-8_6. URL https://doi.org/10.1007/978-3-031-41679-8_6.

Paliwal SS, D V, Rahul R, Sharma M and Vig L (2019) Tablenet: Deep learning model for end-to-end table detection and tabular data extraction from scanned document images. In: *2019 International Conference on Document Analysis and Recognition (ICDAR)*. pp. 128–133. DOI:10.1109/ICDAR.2019.00029. URL https://doi.org/10.1109/ICDAR.2019.00029.

Pandey M, Arora M, Arora S, Goyal C, Gera VK and Yadav H (2023) Ai-based integrated approach for the development of intelligent document management system (idms). *Procedia Computer Science* 230: 725–736. DOI:10.1016/j.procs.2023.12.127. URL https://doi.org/10.1016/j.procs.2023.12.127.

Peng W and Li X (2025) An ante hoc enhancement method for image-based complex financial table extraction. *Applied Sciences* 15: 370. DOI:10.3390/app15010370. URL https://doi.org/10.3390/app15010370.

Qiao L, Li Z, Cheng Z, Zhang P, Pu S, Niu Y, Ren W, Tan W and Wu F (2021) Lgpma: Complicated table structure recognition with local and global pyramid mask alignment. In: *Document Analysis and Recognition – ICDAR 2021*. Springer, pp. 99–114. DOI:10.1007/978-3-030-86549-8_7. URL https://doi.org/10.1007/978-3-030-86549-8_7.

Raja S, Mondal A and Jawahar CV (2020) Table structure recognition using top-down and bottom-up cues. In: *Computer Vision – ECCV 2020*, Lecture Notes in Computer Science. Springer, pp. 70–86. DOI:10.1007/978-3-030-58604-1_5. URL https://doi.org/10.1007/978-3-030-58604-1_5.

Ramos D, Pereira J, Lynce I, Manquinho V and Guimarães F (2020) Unchartit: An interactive framework for programmatic chart understanding. In: *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (UIST)*. pp. 1350–1362. DOI:10.1145/3324884.3416613. URL https://doi.org/10.1145/3324884.3416613.

Reddy R, Ramesh R, Deshpande A and Khapra MM (2019) Figurenet: A deep learning model for question-answering on scientific plots. In: *2019 International Joint Conference on Neural Networks (IJCNN)*. pp. 1–8. DOI:10.1109/IJCNN.2019.8851830. URL https://doi.org/10.1109/IJCNN.2019.8851830.

Ritze D, Lehmberg O, Oulabi Y and Bizer C (2016) Profiling the potential of web tables for augmenting cross-domain knowledge bases. In: *Proceedings of the 25th International Conference on World Wide Web*, WWW '16. International World Wide Web Conferences Steering Committee, pp. 251–261. DOI:10.1145/2872427.2883017. URL https://doi.org/10.1145/2872427.2883017.

Saout T, Lardeux F and Saubion F (2023) A two-stage approach for tables extraction in invoices. In: *Proceedings of the IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*. pp. 10–15. DOI:10.1109/ICTAI59109.2023.00010. URL https://doi.org/10.1109/ICTAI59109.2023.00010.

Shah K and Patel D (2004) A survey on table recognition. *Document Analysis and Recognition* 7(1): 1–16. DOI:10.1007/s10032-004-0120-9. URL https://doi.org/10.1007/s10032-004-0120-9.

Shigarov A (2022) Table understanding: Problem overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 13. DOI:10.1002/widm.1482. URL https://doi.org/10.1002/widm.1482.

Shivasankaran VP, Hassan MY and Singh M (2023) Lineex: Data extraction from scientific line charts. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. pp. 6202–6210. DOI:10.1109/WACV56688.2023.00615. URL https://doi.org/10.1109/WACV56688.2023.00615.

Smock B, Pesala R and Abraham R (2022) Pubtables-1m: Towards comprehensive table extraction from unstructured documents. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 4634–4642. DOI: 10.1109/CVPR52688.2022.00459. URL https://doi.org/10.1109/CVPR52688.2022.00459.

Sohn C, Choi H, Kim K, Park J and Noh J (2021) Line chart understanding with convolutional neural network. *Electronics* 10: 749. DOI:10.3390/electronics10060749. URL https://doi.org/10.3390/electronics10060749.

Sreevalsan-Nair J, Dadhich K and Daggubati SC (2021) Tensor fields for data extraction from chart images: Bar charts and scatter plots. In: *Topological Methods in Data Analysis and Visualization VI*. pp. 219–241. DOI:10.1007/978-3-030-83500-2_12. URL https://doi.org/10.1007/978-3-030-83500-2_12.

Sun L, He L, Jia S, He Y and You C (2025) Docagent: An agentic framework for multi-modal long-context document understanding. In: *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*. Suzhou,

China: Association for Computational Linguistics. ISBN 979-8-89176-332-6, pp. 17712–17727. DOI:10.18653/v1/2025.emnlp-main.893. URL https://aclanthology.org/2025.emnlp-main.893.

Sviatov K, Yarushkina N and Sukhov S (2021) Data extraction of charts with hybrid deep learning model. In: *Computational Science and Its Applications – ICCSA 2021*. Cagliari, Italy: Springer, pp. 382–393. DOI:10.1007/978-3-030-87013-3_29. URL https://doi.org/10.1007/978-3-030-87013-3_29.

Tan M and Le Q (2019) Efficientnet: Rethinking model scaling for convolutional neural networks. In: *Proceedings of the 36th International Conference on Machine Learning (ICML)*, *Proceedings of Machine Learning Research*, volume 97. PMLR, pp. 6105–6114. URL https://proceedings.mlr.press/v97/tan19a.html.

Thome B, Hertweck F, Jonas L, Yasar S and Conrad S (2024) Automated extraction of icon-based tables. In: *INFORMATIK 2024*. Bonn: Gesellschaft für Informatik, pp. 2003–2005. DOI:10.18420/inf2024_173. URL https://doi.org/10.18420/inf2024_173.

Tsutsui S and Crandall DJ (2017) A data driven approach for compound figure separation using convolutional neural networks. In: *Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. pp. 533–540. DOI:10.1109/ICDAR.2017.93. URL https://doi.org/10.1109/ICDAR.2017.93.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł and Polosukhin I (2017) Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NeurIPS '17. Red Hook, NY, USA: Curran Associates Inc., pp. 6000–6010. URL https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Venetis P, Halevy A, Madhavan J, Paşca M, Shen W, Wu F, Miao G and Wu C (2011) Recovering semantics of tables on the web. *Proceedings of the VLDB Endowment* 4(9): 528–538. DOI:10.14778/2002938.2002939. URL https://doi.org/10.14778/2002938.2002939.

Wang F, Sun K, Chen M, Pujara J and Szekely P (2021a) Retrieving complex tables with multi-granular graph representation learning. In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '21. Association for Computing Machinery, pp. 1472–1482. DOI:10.1145/3404835.3462909. URL https://doi.org/10.1145/3404835.3462909.

Wang NXR, Burdick D and Li Y (2021b) Tablelab: An interactive table extraction system with adaptive deep learning. In: *Companion Proceedings of the 26th International Conference on Intelligent User Interfaces*. pp. 87–89. DOI:10.1145/3397482.3450718. URL https://doi.org/10.1145/3397482.3450718.

Widodo S and Krisnadhi A (2024) Enhancing table-to-text generation with numerical reasoning using graph2seq models. *International Journal of Innovation in Enterprise System* 8: 11–21. DOI:10.25124/ijies.v8i02.236. URL https://doi.org/10.25124/ijies.v8i02.236.

Wu S, Xie C, Huang Y, Tang G, Liao Q, Wang J, Chen B, Li H, Chang X, Li H, Ding K, Huang Y and Jin L (2021) Improving machine understanding of human intent in charts. In: *Document Analysis and Recognition – ICDAR 2021*. pp. 676–691. DOI:10.1007/978-3-030-86334-0_44. URL https://doi.org/10.1007/978-3-030-86334-0_44.

Xu Z, Qu B, Qi Y, Du S, Xu C, Yuan C and Guo J (2025) Chartmoe: Mixture of diversely aligned expert connector for chart understanding. In: *International Conference on Learning Representations (ICLR)*. pp. 63452–63474. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-105010235588&partnerID=40&md5=7f18c86e9c755b0c82db25f987c60bb2.

Yan P, Ahmed S and Doermann D (2023) Context-aware chart element detection. In: *Document Analysis and Recognition – ICDAR 2023*. pp. 218–233. DOI:10.1007/978-3-031-41676-7_13. URL https://doi.org/10.1007/978-3-031-41676-7_13.

Yang J, Gupta A, Upadhyay S, He L, Goel R and Paul S (2022) Tableformer: Robust transformer modeling for table-text encoding. In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. pp. 528–537. DOI:10.18653/v1/2022.acl-long.40. URL https://doi.org/10.18653/v1/2022.acl-long.40.

Yildiz B, Kaiser K and Miksch S (2005) pdf2table: A method to extract table information from pdf files. In: *Proceedings of the 2nd Indian International Conference on Artificial Intelligence (IICAI 2005)*. Pune, India: IICAI. ISBN 0-9727412-1-6, pp. 1773–1785. URL https://api.semanticscholar.org/CorpusID:260837170.

Zhang L, Hu A, Xu H, Yan M, Xu Y, Jin Q, Zhang J and Huang F (2024) Tinychart: Efficient chart understanding with program-of-thoughts learning and visual token merging. In: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. pp. 1882–1898. DOI:10.18653/v1/2024.emnlp-main.112. URL https://doi.org/10.18653/v1/2024.emnlp-main.112.

Zhang M, Perelman D, Le V and Gulwani S (2021) An integrated approach of deep learning and symbolic analysis for digital pdf table extraction. In: *2020 25th International Conference on Pattern Recognition (ICPR)*. pp. 4062–4069. DOI:10.1109/ICPR48806.2021.9413069. URL https://doi.org/10.1109/ICPR48806.2021.9413069.

Zhao J, Fan M and Feng M (2022) Chartseer: Interactive steering exploratory visual analysis with machine intelligence. *IEEE Transactions on Visualization and Computer Graphics* 28(3): 1500–1513. DOI:10.1109/TVCG.2020.3018724. URL https://doi.org/10.1109/TVCG.2020.3018724.

Zheng X, Burdick D, Popa L, Zhong X and Wang NXR (2021) Global table extractor (gte): A framework for joint table identification and cell structure recognition using visual context. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. pp. 697–706. DOI:10.1109/WACV48630.2021.00074. URL https://doi.org/10.1109/WACV48630.2021.00074.

Zhong X, Tang J and Jimeno Yepes A (2019) Publaynet: Largest dataset ever for document layout analysis. In: *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*. pp. 1015–1022. DOI:10.1109/ICDAR.2019.00166. URL https://doi.org/10.1109/ICDAR.2019.00166.

Zhou F, Zhao Y, Chen W, Tan Y, Xu Y, Chen Y, Liu C and Zhao Y (2021) Reverse-engineering bar charts using neural networks. *Journal of Visualization* 24(2): 419–435. DOI: 10.1007/s12650-020-00702-6. URL https://doi.org/10.1007/s12650-020-00702-6.

Zhou M, Fung Y, Chen L, Thomas C, Ji H and Chang SF (2023) Enhanced chart understanding via visual language pre-training on plot table pairs. In: *Findings of the Association for Computational Linguistics: ACL 2023*. pp. 1314–1326. DOI: 10.18653/v1/2023.findings-acl.85. URL https://doi.org/10.18653/v1/2023.findings-acl.85.

Zhu H, Yin J, Chu C, Zhu M, Wei Y, Pan J, Han D, Tan X and Chen W (2024) A visual analysis approach for data transformation via domain knowledge and intelligent models. *Multimedia Systems* 30(3): 1–22. DOI:10.1007/s00530-024-01331-x. URL https://doi.org/10.1007/s00530-024-01331-x.

Appendix

This appendix compiles the extended empirical resources supporting the analyses presented in Results and RQ-Based Literature Review and Analysis. It includes the summarized comparison matrix of all reviewed systems, an overview of application domains with representative tools, and comprehensive lists of publicly available datasets and tool repositories referenced in the studied works. Together, these materials provide a transparent foundation for replication and future research on table and chart data extraction.

Summary of Comparison Matrix of Reviewed Systems

Table 16 presents a summarized comparison matrix of representative systems included in the review. It reports each system's modality, main extraction tasks, and associated evaluation metrics, providing a compact overview of the methodological and evaluation diversity observed across the reviewed corpus.

The complete version of the matrix, which includes all annotated attributes, is provided as an open supplementary resource to ensure transparency and reproducibility. It can be found at the following link:

https://doi.org/10.5281/zenodo.18072492

Table 16. Representative systems derived from the comparison framework with associated evaluation metrics

System	Year	Modality	Tasks	Evaluation metrics
PubTables-1M Smock et al. (2022)	2021	Table	Table detection (TD), table structure recognition (TSR)	AP, AP50, AP75, AR, AccCont, GriTS_Top, GriTS_Cont, GriTS_Loc, GriTS_Adj
TableLab Wang et al. (2021b)	2021	Table	Table structure recognition (TSR)	Structure recognition accuracy
GTE Zheng et al. (2021)	2021	Table	Table detection, cell structure recognition, structure recognition	Character-level recall, precision, F1-measure, IoU, TEDS
TabLeX Desai et al. (2021)	2021	Table	Table structure recognition (TSR)	TEDS
BarChartAnalyzer Daggubati et al. (2022)	2022	Chart	Bar chart data extraction	Precision, recall, F1, nMAE, MAPE
ChartOCR Luo et al. (2021)	2021	Chart	Chart-to-data extraction and text recognition (bar, pie, line)	Precision, recall, F1, chart similarity score, mean error
UnchartIt Ramos et al. (2020)	2020	Chart	Bar chart data extraction and program recovery (chart-to-table conversion)	Accuracy, MAE, RMSE, success rate
LineEX Shivasankaran et al. (2023)	2023	Chart	Line chart data extraction and axis mapping (keypoint detection and line segmentation)	IoU, precision, recall, F1
ChartLlama Han et al. (2025)	2023	Chart	Chart data generation, chart figure generation, instruction data generation, chart-to-chart, text-to-chart, chart editing	Accuracy, precision, F1, GPT score
TinyChart Zhang et al. (2024)	2024	Both	ChartQA, chart-to-text, chart-to-table, OpenCQA, ChartX-Cognition	Accuracy, BLEU-4, RMS_F1, augmented accuracy (Aug), human accuracy (Hum), average (Avg)
ChartAssistant Meng et al. (2024)	2024	Chart	Chart-to-table translation, chart numerical QA, chart referring QA, chart open-ended QA, chart summarization	Relaxed correctness, BLEU, RMS_F1
ChartEye Mustafa et al. (2023)	2024	Chart	Chart information extraction	Accuracy, IoU, precision, recall
TTC-QuAli Dong et al. (2024b)	2024	Both	Unified text–table–chart quantity alignment	Accuracy
MATVIX Khalighinejad et al. (2025)	2024	Chart	Multimodal information extraction (polymer nanocomposite – PNC, polymer biodegradation – PBD)	Precision, recall, F1, headers, curves, CAS
AI-Blueprint Gu et al. (2024)	2024	Table	Circuit diagram classification, component detection and classification	Accuracy, mAP50
ChartReader Cheng et al. (2023)	2023	Chart	Chart-to-table extraction, ChartQA, chart-to-text	Accuracy, BLEU4
Mystique Chen et al. (2024)	2024	Chart	Axis and legend detection, GREC-based chart deconstruction, chart reproduction (user study)	Detection accuracy, deconstruction accuracy, qualitative user reproduction success
ChartMoE Xu et al. (2025)	2025	Chart	ChartQA, ChartBench, ChartFC, ChartCheck	Relaxed accuracy@0.05, accuracy

Application Domains and Representative Tools

Table 17 summarizes the main application domains represented among the 68 reviewed systems, highlighting a selection of representative tools and models within each area. Rather than an exhaustive classification, this summary provides an illustrative overview of how table and chart data extraction methods have been adapted to diverse contexts from scholarly publishing and document analysis to specialized fields such as finance, engineering, and materials science. Each domain groups systems that share similar objectives or operate on comparable data sources, reflecting the range of real world problems addressed by current research efforts.

The observed domain distribution highlights the interdisciplinary scope of metadata extraction research. Consistent with prior surveys on document understanding and visual information extraction Desai et al. (2021); Shah and Patel (2004), most systems are developed for scholarly and general document analysis, while a smaller subset targets specialized domains such as finance, materials science, and engineering. Recent multimodal approaches, particularly those based on vision–language architectures such as TinyChart Zhang et al. (2024), suggest a gradual trend toward domain-adaptive and cross-context models capable of generalizing across multiple application areas.

Datasets

Table and chart extraction research builds upon a variety of public datasets spanning scientific, financial, educational, and multimodal domains. The following list compiles all accessible datasets mentioned across the reviewed works and supplementary resources, excluding proprietary or unavailable ones. These resources are commonly used for table structure recognition, chart understanding, and visual question answering tasks.

- https://huggingface.co/datasets/bsmock/pubtables-1m
- https://paperswithcode.com/dataset/tablex
- https://chartinfo.github.io/toolsanddata.html
- https://huggingface.co/datasets/liminghao1630/TableBank
- https://huggingface.co/datasets/bsmock/ICDAR-2013.c
- https://github.com/webdataset/webdataset
- https://github.com/soap117/DeepRule
- https://github.com/adobe-research/CHART-Synthetic/releases
- https://pmc.ncbi.nlm.nih.gov/tools/openftlist/
- https://github.com/ibm-aur-nlp/PubLayNet
- https://github.com/ibm-aur-nlp/PubTabNet
- https://github.com/cy-sohn/LCUdataset_generator
- https://github.com/Maluuba/FigureQA
- https://paperswithcode.com/dataset/vega-lite-chart-collection
- https://automeris.io/
- https://github.com/doc-analysis/TableBank
- https://github.com/Academic-Hammer/SciTSR
- https://huggingface.co/datasets/bsmock/FinTabNet.c
- https://github.com/Shiva-sankaran/LineEX
- https://github.com/doc-analysis/DocBank
- https://huggingface.co/datasets/HuggingFaceM4/ChartQA
- https://github.com/NiteshMethani/PlotQA
- https://github.com/vis-nlp/Chart-to-text
- https://github.com/vis-nlp/UniChart
- https://github.com/kushalkafle/DVQA_dataset
- https://github.com/allenai/figureseer
- https://github.com/tingyaohsu/SciCap
- https://github.com/anita76/LineCapDataset
- https://www.gapminder.org/data/
- https://github.com/vis-nlp/ChartQA
- https://github.com/vis-nlp/OpenCQA
- https://github.com/xdwang0726/MeSHup
- https://github.com/DS4SD/DocLayNet
- https://github.com/ghazalkhalighinejad/matvix
- https://github.com/Leezekun/MMSci
- https://github.com/lamalab-org/macbench
- https://github.com/Elucidator-V/NovaChart
- https://github.com/OpenGVLab/ChartAst
- https://github.com/OSU-slatelab/MapQA
- https://github.com/Future-House/paper-qa
- https://huggingface.co/datasets/SincereX/ChartBench
- https://huggingface.co/datasets/MMInstruction/ArxivQA
- https://github.com/google-research-datasets/ToTTo
- https://github.com/microsoft/HiTab
- https://huggingface.co/datasets/ethanbradley/synfintabs
- https://github.com/RamiKrispin/covid19Italy
- https://huggingface.co/datasets/anhdang000/ChartQA-PoTQuery
- https://www.pewresearch.org/tools-and-datasets/

Table 17. Key application domains and representative tools for table and chart data extraction

Application domain	Representative tools / systems
Scientific and academic publishing	PubTables-1M Smock et al. (2022); TableLab Wang et al. (2021b); ChartQA Masry et al. (2022)
Financial and business reporting	UnchartIt Ramos et al. (2020); ChartReader Cheng et al. (2023)
Engineering and circuit design	AI-Blueprint Gu et al. (2024); ChartAssistant Meng et al. (2024); MATVIX Khalighinejad et al. (2025)
Materials and chemical sciences	MATVIX Khalighinejad et al. (2025); TTC-QuAli Dong et al. (2024b)
Education and infographics	TinyChart Zhang et al. (2024); ChartReader Cheng et al. (2023); ChartMoE Xu et al. (2025)
Biology and life sciences	PubTables-1M Smock et al. (2022); TableLab Wang et al. (2021b)
General document analysis	ChartOCR Luo et al. (2021); TableBank Li et al. (2020); TableFormer Yang et al. (2022); GTE Zheng et al. (2021)

- https://www.kaggle.com/datasets/ritvik1909/marmot-dataset
- https://materialsmine.org/
- https://allenai.org/data/cord-19
- https://www.kaggle.com/datasets/xhlulu/covidqa
- https://github.com/matsa-ai/demo
- https://github.com/karmaresearch/tab2know
- https://chartinfo.github.io/index_2022.html
- https://tc11.cvc.uab.es/datasets/ICDAR-CHART-2019-PMC_1
- https://vis.stanford.edu/papers/revision
- https://github.com/vega/vega-datasets
- https://cvit.iiit.ac.in/usodi/Docfig.php
- https://github.com/zmykevin/ChartOCR
- https://guillaumejaume.github.io/FUNSD/
- https://huggingface.co/datasets/aharley/rvl_cdip
- https://huggingface.co/datasets/chiragtubakad/chart-to-table
- https://github.com/RhoInc/Webcharts
- https://ourworldindata.org/
- https://www.oecd.org/en/data/datasets.html
- https://huggingface.co/datasets/kasnerz/numericnlg
- https://github.com/TabbyXL/TabbyXL

Tools and Systems

The tools and models listed below represent the most frequently used resources in multimodal table and chart understanding. They include open-source libraries for OCR, layout parsing, document segmentation, visual-language modeling, and multimodal reasoning. All duplicates and inoperative links have been removed.

- https://automeris.io/
- https://github.com/tesseract-ocr/tesseract
- https://github.com/facebookresearch/detr
- https://github.com/microsoft/table-transformer
- https://github.com/facebookresearch/ResNeXt
- https://github.com/zmykevin/ChartOCR
- https://github.com/rbgirshick/py-faster-rcnn
- https://research.ibm.com/publications/tablelab-an-interactive-table-extraction-sys
- https://github.com/DevashishPrasad/CascadeTabNet
- https://github.com/xuewenyuan/TGRNet
- https://github.com/facebookresearch/detectron2
- https://developer.adobe.com/document-services/docs/overview/pdf-extract-api/
- https://tika.apache.org/
- https://camelot-py.readthedocs.io/en/master/
- https://github.com/kermitt2/grobid
- https://pymupdf.readthedocs.io/en/latest/
- https://github.com/allenai/science-parse
- https://tabula.technology/
- https://huggingface.co/docs/transformers/en/model_doc/pix2struct
- https://github.com/tingxueronghua/ChartLlama-code
- https://github.com/haotian-liu/LLaVA

- https://github.com/ejlee95/Graph-based-TSR
- https://github.com/sachinraja13/TabStructNet
- https://github.com/Irene323/GFTE
- https://damienmasson.com/ChartDetective/home.html
- https://github.com/UKPLab/sentence-transformers
- https://github.com/google-research/bert
- https://github.com/facebookresearch/fairseq
- https://github.com/allenai/scibert
- https://github.com/facebookresearch/fastText
- https://github.com/allenai/specter
- https://github.com/PaddlePaddle/PaddleOCR
- https://github.com/OpenGVLab/ChartAst
- https://github.com/vis-nlp/ChartInstruct
- https://github.com/QwenLM/Qwen-VL
- https://github.com/google-research/pix2struct
- https://github.com/clovaai/donut
- https://github.com/microsoft/unilm/tree/master/beit
- https://github.com/matterport/Mask_RCNN
- https://github.com/NiteshMethani/PlotQA
- https://github.com/Cvrane/ChartReader
- https://github.com/Shiva-sankaran/LineEX
- https://github.com/TheJaeLal/LineFormer
- https://github.com/ultralytics/yolov5
- https://opencv.org/
- https://pandas.pydata.org/