



Review

Machine Learning Tools and Platforms in Clinical Trial Outputs to Support Evidence-Based Health Informatics: A Rapid Review of the Literature

Stella C. Christopoulou

Department of Business Administration and Organizations, University of Peloponnese, 24100 Kalamata, Greece; stellachristop@gmail.com

Abstract: Background: The application of machine learning (ML) tools (MLTs) to support clinical trials outputs in evidence-based health informatics can be an effective, useful, feasible, and acceptable way to advance medical research and provide precision medicine. Methods: In this study, the author used the rapid review approach and snowballing methods. The review was conducted in the following databases: PubMed, Scopus, COCHRANE LIBRARY, clinicaltrials.gov, Semantic Scholar, and the first six pages of Google Scholar from the 10 July–15 August 2022 period. Results: Here, 49 articles met the required criteria and were included in this review. Accordingly, 32 MLTs and platforms were identified in this study that applied the automatic extraction of knowledge from clinical trial outputs. Specifically, the initial use of automated tools resulted in modest to satisfactory time savings compared with the manual management. In addition, the evaluation of performance, functionality, usability, user interface, and system requirements also yielded positive results. Moreover, the evaluation of some tools in terms of acceptance, feasibility, precision, accuracy, efficiency, efficacy, and reliability was also positive. Conclusions: In summary, design based on the application of clinical trial results in ML is a promising approach to apply more reliable solutions. Future studies are needed to propose common standards for the assessment of MLTs and to clinically validate the performance in specific healthcare and technical domains.

Keywords: machine learning; evidence-based health informatics; clinical trials; RCT; automated screening; artificial intelligence; automation; machine learning tools



Citation: Christopoulou, S.C. Machine Learning Tools and Platforms in Clinical Trial Outputs to Support Evidence-Based Health Informatics: A Rapid Review of the Literature. *Biomedinformatics* **2022**, *2*, 511–527. <https://doi.org/10.3390/biomedinformatics2030032>

Academic Editor: José Manuel Ferreira Machado

Received: 15 August 2022

Accepted: 10 September 2022

Published: 14 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Evidence-based health informatics (EBHI) can be defined as the conscious, explicit, and judicious use of current best evidence to support a health care decision that employs information technologies (ITs) [1]. Towards this direction, clinical trials are considered as a well-established experimental clinical tool suitable not only to evaluate the effectiveness of interventions, but also to support the conduct of an adequately designed systematic review [2]. Furthermore, meta-analysis is a systematic approach for understanding a phenomenon by analyzing the results of many previously published clinical trials [3]. Meta-analysis applied to clinical trials is a central method for quality evidence generation. In particular, meta-analysis is gaining speedy momentum in the growing world of quantitative information [4]. Thus, both EBHI and clinical trials are currently at the forefront of supporting clinicians in clinical decision making.

Precedence Research announced that the global clinical trials market size was valued at USD 51.05 billion in 2021, and is forecast to hit USD 84.43 billion by 2030 with a registered compound annual growth rate (CAGR) of 5.7% during the forecast period of 2022 to 2030 (<https://www.precedenceresearch.com/clinical-trials-market>, accessed on 10 August 2022).

At the same time, the increasing volume of patient admissions due to the increase in various chronic diseases and the rapidly increasing aging population worldwide is fueling the growth of artificial intelligence (AI) in the healthcare market.

Related research by Precedence announced that the size of the global AI in the healthcare market was estimated at USD 11.06 billion in 2021, and is expected to exceed approximately USD 187.95 billion by 2030, growing at a CAGR of 37% during the forecast period of 2022 to 2030 (<https://www.precedenceresearch.com/artificial-intelligence-in-healthcare-market>, accessed on 10 August 2022). The clinical trials segment generated revenue of over 24.2% in 2021 and dominated the global healthcare AI market.

However, even if we consider the field of research, we find that during this time period, according to the search engine semantic scholar, defining the search key “health”, 6,060,000 articles are circulating on the internet, of which 749,050 are reviews and 5290 are clinical studies (Figure 1).

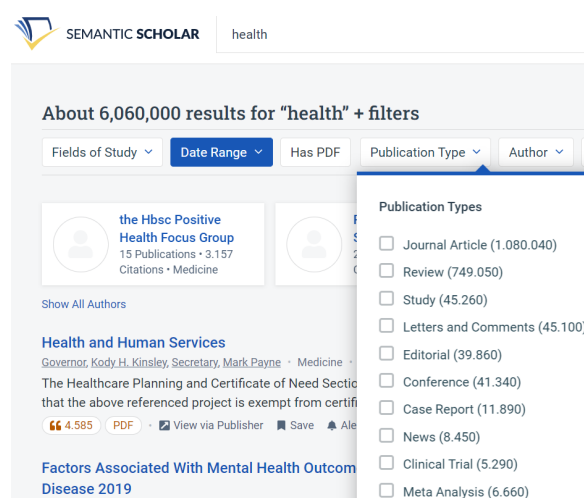


Figure 1. Search with the search key “health” in the semantic scholar search engine (source: <https://www.semanticscholar.org/>, accessed on 10 August 2022).

The information available on the internet is growing dramatically every year. For example, searching for “clinical trials” with the semantic scholar search engine found 32,099 related articles in 2001, 82,250 related articles in 2011 (an increase of 256%), and 139,008 related articles in 2021 (an increase of 169%) (Figure 2).

However, because of the large and complex collection of datasets derived from clinical trials, it often becomes impossible to fully exploit and apply them in health and they are difficult to process with traditional data processing applications [5]. Furthermore, because this amount of information is growing rapidly, the ability to apply machine learning tools (MLTs) to automate knowledge extraction is now more critical than ever.

Currently, the application of digital technology in clinical trials is studied, proposed, promoted, and implemented in some studies [6,7].

Clinical trials are a fundamental tool used to evaluate the effectiveness and safety of new drugs and medical devices and other health system interventions. The traditional clinical trial system acts as a reliable tool for the development and implementation of new drugs, devices, and interventions in the health system. However, digital tools can be used to analyze and optimize clinical trials, and finally, in the future, it will be possible to support them with digital tools completely by implementing virtual tests and experiments using even virtual human models [8].



Figure 2. Search for “clinical trials” with the semantic scholar search engine (source: <https://www.semanticscholar.org/>, accessed on 10 August 2022).

At the present stage, however, the management of clinical trials results in drawing conclusions, and selecting the appropriate treatment with the application of AI and machine learning (ML) is an extremely topical and critical issue. In this way, it will be possible to make the most of clinical studies, thus achieving a more rational and more economical application in daily clinical practice compared with classical methods.

ML was defined by Arthur Samuel, a ML pioneer, as “a field of study that gives computers the ability to learn without being explicitly programmed” [9]. A broader domain of ML is AI. AI refers, in general, to the simulation of human intelligence in machines that are programmed to think in the same way as humans and mimic their actions.

Although there is a skepticism regarding the practical application and interpretation of results from ML-based approaches in healthcare settings, the inclusion of these approaches is growing at a rapid pace [10].

More in detail, recent developments in AI and ML technology have brought on substantial strides in issues such as the prediction and detection of health emergencies, the treatment of diseases and immune response problems [10], the diagnosis of diseases, living assistance, biomedical information processing, biomedical research [11], automated treatment, disease recommendation, automated robotic surgery, and drug discovery and development [12].

At the same time, ML and AI have been developing rapidly in recent years in terms of software algorithms, hardware implementation, and applications in a huge number of areas [11].

However, the authors of [13] found no unified information extraction framework tailored to the systematic review process, and published reports focused on a limited (1–7) number of data elements. Biomedical natural language processing techniques have not been fully utilized to fully or even partially automate the data extraction step of systematic reviews.

Nevertheless, it is estimated that natural language processing (NLP) will emerge as the most effective tool for generating structured information from unstructured data, which is commonly found in clinical trial texts. In the research article [14], the bibliometric analysis of the annual publication trend showed that there has been a dramatic increase in research interests in NLP-enhanced clinical trial research.

Moving in this direction, the author of this article deals with the application of ML through appropriate tools to extract the results from the application of clinical trials so

that they can be properly applied in daily clinical practice [15]. Thus, initially, the author searched for relative work and analytically described it below in Section 3.1.

More analytically, the author in this article performed a rapid review exploring MLTs and approaches in the field of clinical trials to support EBHL.

The main research question was as follows:

- RQ1. What MLTs and platforms are reported in the literature to derive results through clinical trial implementations?

The secondary research questions were as follows:

- RQ2. What are the main categories of these MLTs?
- RQ3. What are the results, benefits, and experience gained from their implementation and what are the inherent difficulties in implementing them and the main observations for future work and challenges to be overcome?

The rest of this study is organized as: Section 2 discusses a group of related articles. Section 3 presents the Materials and Methods of this study. Section 4 summarizes the results. Section 5 discusses the key issues arising from this study. Section 6 concludes the study and presents future directions.

2. Related Work

There are some notable studies in the field, but only a limited number deal with the subject thoroughly. Automation has been proposed or used to expedite most steps of the systematic review process of clinical studies, including searching, screening, and data extraction.

Marshall and Wallace [16] provided an overview of the current machine learning methods that have been proposed to expedite evidence synthesis. They also offer guidance on which of these are ready for use, their strengths and weaknesses, and how a systematic review team might go about using them in practice.

In addition, Tsafnat et al. [17] detailed a survey designed to support or automate individual tasks in the systematic review, and in particular systematic reviews of randomized controlled clinical trials, which revealed the trends that see the convergence of several parallel research projects. This survey described each of the systematic review tasks in detail. Each task was described along with the potential benefits of its automation. In addition, the technology systems (up to 2014) that automate or support the tasks are listed in detail.

Many significant studies refer to algorithms [14,18,19] and strategies to automate data and knowledge extraction from reviews [20]. Finally, many studies focus on the evaluation of ML methods through a specific tool [21–23].

3. Materials and Methods

3.1. Study Design

In this study design, the author used the rapid review approach [24]. A rapid review can be defined as a form of knowledge synthesis that is produced within a short timeframe using limited resources by streamlining or omitting a number of methods for producing evidence [24].

Moreover, the forward and backwards snowball method was used [25]. It has been proposed that in systematic reviews of complex or heterogeneous evidence in the field of health services research, “snowball” methods of forward (citation) and backwards (reference) searching are especially powerful. This method allows researchers using the references and citations of an article to find specific literature on a topic quickly and relatively easily. Experimentations with this methodology yielded positive results and have also been presented in [26].

Finally, the *SF/HIT* model was used as a template to define specific keywords in order to identify the impacts and outcomes resulting from the use of digital tools in the healthcare domain [27].

3.2. Search Strategy and Eligibility Criteria

The review was conducted using the following databases: PubMed, Scopus, COCHRANE LIBRARY, clinicaltrials.gov, Semantic Scholar and the first six pages of Google Scholar for the 10 July–15 August 2022 period.

Reviews that observed the main objective of describing the MLTs that may extract information and knowledge from the data of clinical trials and the assessments of them were included. It was decided not to restrict the search field in order to collect as much information as possible. Restrictions were related to the language (only English articles were included). Snowballing was undertaken, starting from the included citations and from the references of each article.

3.3. Data Screening

The reference manager Qiqqa version v.76s and Excel were used to export and manage the results.

A two-stage review process (Figure 3) was performed by the author, (a) initially excluding assignments based on the titles and their abstracts, and (b) then the remaining assignments were reviewed based on reading the full text of the article.

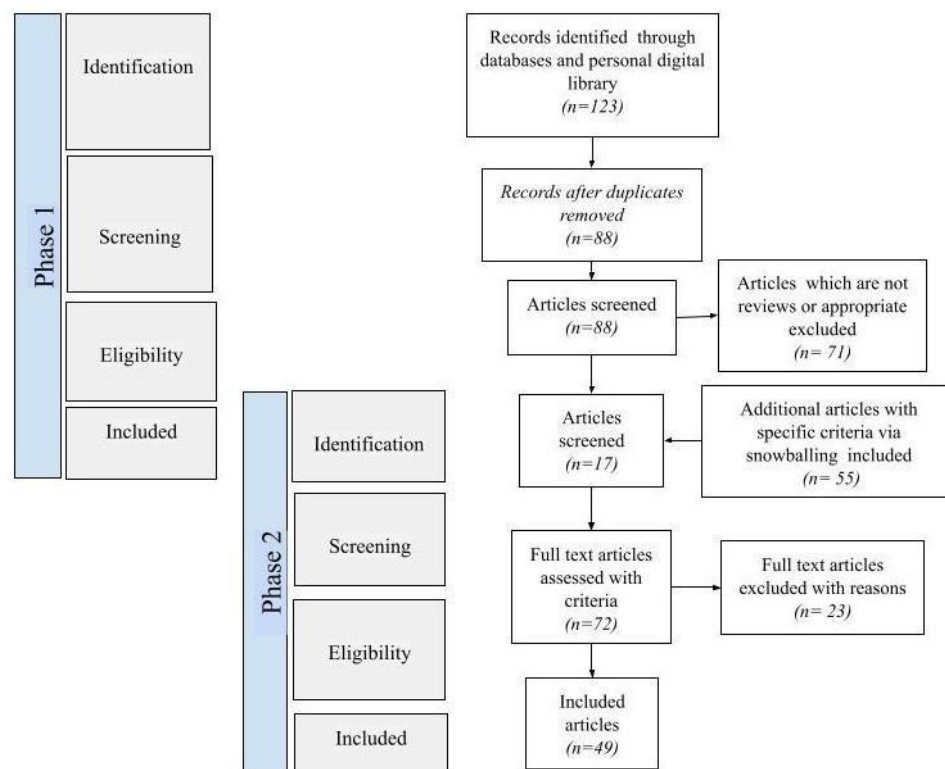


Figure 3. Flow diagram of the literature search.

Specifically, the first stage included two phases.

In the first phase, the articles were extracted based on the following acceptance characteristics: ML methods OR ML approaches OR machine learning tool OR machine learning systems OR ML techniques AND ((RCT OR clinical trial) AND review).

In this phase, only reviews of clinical studies/trials were selected, as the aim was to search for tools in which ML was applied and compared in trials.

Thus, the collected articles were studied based on their titles and abstracts, and the selected ones were included in the second phase.

During the second phase, the MLTs were identified and selected. Next, relevant studies describing these tools in detail were searched using the Snowball citations method.

Finally, these studies were screened based on the title and abstract, and the appropriate ones were included in the pool of selected articles.

One researcher reviewed the articles.

3.4. Data Extraction and Analyses

The following data were extracted from the included studies: authors, type of article, and summary of the article.

The type of article was one of the following:

- Review (selected from the first phase);
- Tools assessment (selected from the first either second phase);
- Automated tool (article selected from the first phase);
- Book either book chapter (selected from the first or second phase).

The heterogeneity and the difficulty in finding analytical and similar descriptive data of the tools made it difficult to carry out a rigorous and standardized analytical record.

Specifically, the results of this review were classified into 12 tasks (categories) in accordance with their type of use and are presented in the results section. This classification relied heavily on the classification of tasks developed in the study by Tsafnat et al. [17].

Specifically, these are the following:

- Design systematic search
- Run systematic search
- Deduplicate
- Obtain full texts
- Snowballing
- Screen abstracts
- Data extraction and text mining tool
- Automated bias assessments
- Automated meta-analysis
- Summarize/synthesis of data (analysis)
- Write up
- Data miner/analysis of data for general purpose.

4. Results

Finally, 49 articles met the criteria and were included in this review; 17 of them were identified and gathered in the first phase of the study (Table 1) and 32 tools were identified and gathered in the second phase of the study (described in Section 4.2).

Table 1. Review articles relative to MLTs for extracting clinical trial outputs.

Author(s)	Tools
(J. Clark et al., 2021) [28]	Polyglot Search, Translator, Deduplicator, SRA-Helper, and SARA
(Clark et al., 2020) [29]	Word Frequency Analyzer, The Search Refiner, Polyglot Search Translator, De-duplicator, SRA Helper, RobotSearch, Endnote, SARA, RobotReviewer, SRA—RevMan Replicant
(Marshall & Wallace, 2019) [16]	RobotSearch, Cochrane, Register of Studies, RCT tagger, Thalia, Abstrackr, EPPI reviewer, RobotAnalyst, SWIFT-Review, Colandr, Rayyan, ExaCT, RobotReviewer, NEMine, Yeast MetaboliNER, AnatomyTagger
(Khalil et al., 2022) [30]	LitSuggest, Rayyan, Abstrackr, BIBOT, R software, RobotAnalyst, DistillerSR, ExaCT and NetMetaXL

Table 1. *Cont.*

Author(s)	Tools
(Erickson et al., 2017) [31]	Caffe, Deeplearning4j, Tensorflow, Theano, Keras, MXNet, Lasagne, Cognitive Network Toolkit (CNTK), DIGITS, Torch, PyTorch, Pylearn2, Chainer, Nolearn, Sklearn-theano and scikit-learn to work with the Theano library, Paddle, H2O
(Pynam et al., 2018) [5]	RapidMiner, Weka, R Tool, KNIME and Orange
(Cleo et al., 2019) [32]	Covidence, SRA-Helper for EndNote, Rayyan and RobotAnalyst
(Wang et al., n.d.) [18]	The authors selected nine mainstream ML algorithms and implemented them in the response-adaptive randomization (RAR) design to predict treatment response.
(Tsafnat et al., 2014) [17]	Quick Clinical, Sherlock, Metta, ParsCit, Abstrackr, ExaCT, WebPlotDigitizer, Meta-analyst, RevMan-HAL, PRISMA Flow Diagram Generator
(Shravan, 2017) [33]	Weka, Rapid Miner, Orange, Knime, DataMelt, Apache Mahout, ELKI, MOA, KEEL, Rattle Mining tasks: Pre-processing, Clustering, Classification, Outlier analysis, Regression, Summarisation Techniques: pattern recognition, statistics, ML, etc.
(Ratra & Gulia, 2020) [34]	WEKA and Orange
(Altalhi et al., 2017) [35]	ADaM, ADAMS, AlphaMiner, CMSR, D.ESOM DataMelt, ELKI, GDataMine, KEEL, KNIME, MiningMart, ML-Flex, Orange RapidMiner, Rattle, SPMF, Tanagra, V.Wabbit, WEKA
(Dwivedi et al., 2016) [36]	WEKA and Salford System
(Naik & Samant, 2016) [37]	RapidMiner, Weka, R Tool:, KNIME and Orange
(Zippel & Bohnet-Joschko, 2021) [38]	RobotSearch, Cochrane Register of Studies, RCT tagger, Thalia, Abstrackr, EPPI reviewer, RobotAnalyst, SWIFT-Review, Colandr, Rayyan, ExaCT, RobotReviewer, NEMine.Yeast MetaboliNER, AnatomyTagger
Systematic Review Toolbox (Marshall and Sutton 2016) [39]	Many tools are presented on web (http://systematicreviewtools.com/about.php , accessed on 5 August 2022)
(Felizardo and Carver 2020) [20]	An overview of strategies researchers have developed to automate the Systematic Literature Review (SLR) process. We used a systematic search methodology to survey the literature about the strategies used to automate the SLR process in SE

These articles describe in detail the MLTs and platforms applied to the automatic extraction of clinical trial data and outputs.

4.1. Review Articles on MLTs for Extracting Clinical Trial Results

Systematic reviews, the cornerstone of evidence-based medicine, are not produced quickly enough to support clinical practice. Production costs, availability of the required expertise, and timeliness are often cited as major factors for this delay. The following reviews and surveys (Table 1) were designed to support or automate individual tasks of reviews, and systematic reviews of randomized controlled clinical trials, and to reveal trends, applied algorithms, and tools while highlighting the convergence of many parallel research projects [17].

4.2. Articles Relative to MLTs for Extracting Clinical Trial Outputs

This section lists state-of-the-art tools that automate tasks that support knowledge extraction from clinical trial outputs.

Analytically, during the second phase of this research, 32 MLTs were identified and were selected.

The most important of these MLTs with a brief description of them are listed below. Many of the tools present characteristics that place them in more than one category. Thus, if deemed necessary, they are recorded in all categories. Otherwise, this is simply stated in their description.

These tools are classified in accordance with 12 functional characteristics (tasks) (e.g., design systematic search, run systematic search, deduplicate, obtain full texts, etc.) (Figure 4).

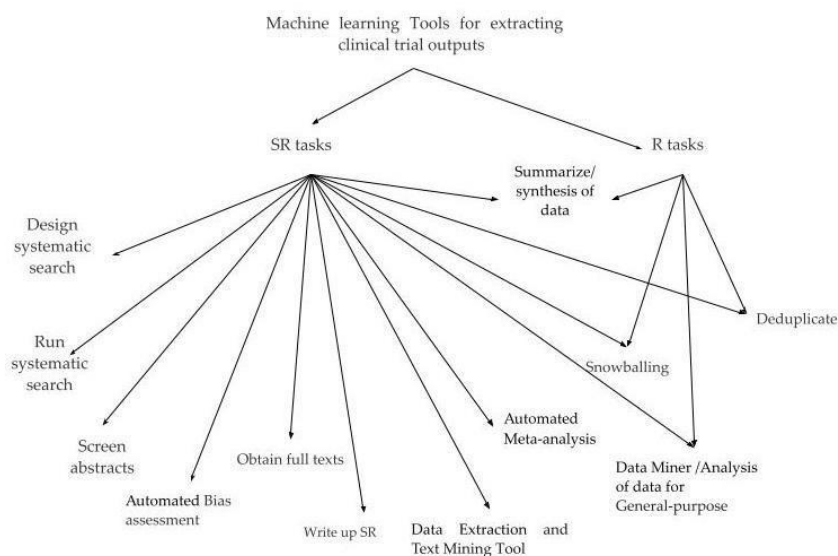


Figure 4. Classification of the ML tools in accordance with their functional characteristics.

4.2.1. Design Systematic Search (Includes Two Tools)

Accelerates the design of a search by counting the number of times a word or phrase appears in a selected group of articles either by checking the recall and precision for each term in the search string and then displaying it visually.

Some of these tools are described below:

- SRA—Word Frequency Analyzer [28], (<http://sr-accelerator.com/#/help/wordfreq>, accessed on 10 August 2022)

Accelerates the design of a search by counting the number of times a word or phrase appears in a selected group of articles. Words that appear frequently should be used in the systematic search.

- The Search Refiner [28]

Accelerates designing a search by checking the recall (number of relevant studies found) and precision (number of irrelevant studies found) for each term in the search string and then displays it visually. Used to quickly determine which terms should be removed from the search string.

4.2.2. Run Systematic Search (Includes Two Tools)

Allows for the search of specific concepts.

Two of these tools are described below:

- Polyglot Search Translator (<http://sr-accelerator.com/#/polyglot>, accessed on 10 August 2022), [28,40]

Accelerates running a search by converting a PubMed or Ovid Medline search to the correct syntax to be run in other databases.

- Thalia (<http://nactem-copious.man.ac.uk/Thalia/>, accessed on 10 August 2022), [16]
Allows for the search of PubMed for concepts (i.e., chemicals, diseases, drugs, genes, metabolites, proteins, species, and anatomical entities).

4.2.3. Deduplicate (Includes One Tool)

Automates most of the deduplication process.

One relative tool is described below:

- De-duplicator (<http://sr-accelerator.com/#/help/dedupe>, accessed on 8 August 2022)
Automates most of the deduplication process by identifying and removing the same study from a group of uploaded records. It is designed to be cautious so some duplicates will remain, which will require removal manually.

4.2.4. Obtain Full Texts (Includes Three Tools)

Screen abstracts and obtain full texts.

Some of these tools are described below:

- SRA Helper (<http://sr-accelerator.com/#/sra-helper>, accessed on 8 August 2022)
Accelerates screening and obtaining full texts by assigning groups to be performed with a hotkey. Hotkeys are also assigned to search a list of prespecified locations to attempt to find the full text of articles.
- SARA (<http://sr-accelerator.com/>, accessed on 8 August 2022)
Automates requesting full-text articles to the library by requesting all of the needed full texts with a single request, for which normally these requests need to be processed and sent one at a time (available within SRA).

- ASH [41]

The ASH tool allows users to download the full text of articles and perform a full-text search. The tool provides a meta-search interface that allows users to obtain much higher search completeness, unifies the search process across all digital libraries, and can overcome the limitations of individual search engines.

4.2.5. Snowballing (Includes One Tool)

These MLTs apply the method for automatic citation snowballing.

One of these tools is described below:

- ParsCit [42]

The proposed tool for automatic citation snowballing is accurate and is capable of obtaining the full texts or abstracts for a substantial proportion of the scholarly citations in review articles.

4.2.6. Screen Abstracts (Includes Six Tools)

Screening abstracts automatically sort a search retrieval by relevance.

Some of these tools are described below:

- RobotSearch (<https://robotsearch.vortext.systems/>, accessed on 8 August 2022), [9]
It is a front-end for a ML model that identifies reports of randomized controlled trials. Moreover, automate citation screening by identifying the studies that are obviously not randomized controlled trials (RCTs) from a group of search results. Removes them, leaving a pool of potential RCTs to be screened.
- Abstrackr (<http://abstrackr.cebm.brown.edu>, accessed on 8 August 2022), [16,43]

The authors in [43] described the ongoing development of an end-to-end interactive ML system. More specifically, they developed abstrackr, an online tool for the task of citation screening for systematic reviews. This tool provides an interface to our ML methods. The main aim of this work is to provide a case study for deploying cutting-edge ML methods that will actually be used by experts in a clinical research setting.

- EPPI reviewer (<https://eppi.ioe.ac.uk/cms/er4>, accessed on 8 August 2022), [16,44]

EPPI reviewer is an application plus a web-based software program for managing and analyzing data in literature reviews. It was developed for all types of systematic reviews (meta-analysis, framework synthesis, thematic synthesis, etc.), but also has features that would be useful in any literature review.

- SWIFT-Review (<https://www.sciome.com/swift-review/>, accessed on 8 August 2022), [16]

SWIFT-Review (Sciome Workbench for Interactive computer-Facilitated Text-mining) provides several features that can be used to search, categorize, and prioritize large (or small) bodies of literature in an interactive manner. Moreover, it utilizes statistical text mining and ML methods that allow users to uncover over-represented topics within the literature corpus and to rank order documents for manual screening.

- Colandr (<https://www.colandrapp.com>, accessed on 8 August 2022), [16]

Colandr is a web-based, open access platform for conducting evidence reviews. Colandr can be used by collaborative teams and provides an organizational structure to manage information throughout the entire evidence review process. Among others, it provides collaborative team working, citation upload in common bibliographic formats (e.g., BibTex and RIS), de-duplication of citations, citation screening using the title and abstract powered by ML, data extraction from full texts powered by natural language processing, and the export of screening decisions and extracted data in comma-separated value format.

- Rayyan (<https://rayyan.qcri.org>, accessed on 8 August 2022), [16,45]

Rayyan is a free web and mobile app that helps expedite the initial screening of abstracts and titles using a process of semi-automation while incorporating a high level of usability.

4.2.7. Data Extraction and Text Mining Tool (Includes Six Tools)

These systems automatically extract data elements (e.g., sample sizes, descriptions of PICO elements).

Some of these tools are described below:

- ExaCT [16,21,22], (<http://exactdemo.iit.nrc.ca>, accessed on 8 August 2022)

ExaCT is a prototype ML and text mining tool that helps to automatically extract study characteristics from the full-texts of RCTs. It also aims to help efficiency compared with manual data extraction.

- RobotAnalyst [46], (<http://www.nactem.ac.uk/robotanalyst/>, accessed on 8 August 2022), [16]

RobotAnalyst is a web-based software system that combines text-mining and ML algorithms for organizing references by their content and actively prioritizing them based on a relevancy classification model that is trained and updated throughout the process.

RobotAnalyst and SWIFT-Review also allow for topic modeling, where abstracts related to similar topics are automatically grouped, allowing the user to explore the search retrieval.

- Dextr [47]

Dextr provides a similar performance to manual extraction in terms of recall and precision and greatly reduces data extraction time. Unlike other tools, Dextr provides the ability to extract complex concepts (e.g., multiple experiments with various exposures and doses within a single study), properly connect the extracted elements within a

study, and effectively limit the work required by researchers to generate machine-readable, annotated exports.

- RobotReviewer (<https://robotreviewer.vortext.systems>, accessed on 8 August 2022), [16]

RobotReviewer is an open-source ML system that supports semi-automated bias assessments. It accelerates assessing risk of bias on four of the seven risk of bias domains by highlighting the supporting phrases in the PDF of the original paper. A check of the assessments is recommended, although the process is drastically speeded up.

- NaCTeM [16], (<http://www.nactem.ac.uk/software.php>, accessed on 8 August 2022)

NaCTeM is a text mining tool for automatically extracting concepts relating to genes and proteins (NEMine).

- Trialstreamer [48]

Trialstreamer continuously monitors PubMed and the World Health Organization International Clinical Trials Registry Platform and looks for RCTs using a validated classifier. It combines ML and rule-based methods to extract information from the RCT abstracts.

4.2.8. Automated Bias Assessments (Includes One Tool)

These tools support automatic assessment of the biases in the reports of RCTs.

The systems are recommended for semi-automatic use (i.e., with human reviewer checking and correcting the ML suggestions).

- RobotReviewer [49], (<https://robotreviewer.vortext.systems>, accessed on 8 August 2022), [16]

RobotReviewer also supports automated bias assessment processes. As previously discussed, this is an open-source ML system that also supports semi-automates bias assessments.

4.2.9. Automated Meta-Analysis (Includes Three Tools)

Meta-analysis is a systematic approach for understanding a phenomenon by analyzing the results of many previously published experimental studies. Unfortunately, meta-analysis involves great human effort, rendering a process that is extremely inefficient and vulnerable to human bias. To overcome these issues, researchers are working toward automating meta-analysis [3].

Some of these ML automated tools are described below:

- SAMA (Ajji et al., 2022) [50]

This tool provides semi-automated meta-analysis (SAMA).

- MetaCyto (<http://bioconductor.org/packages/release/bioc/html/MetaCyto.html>, accessed on 8 August 2022), [51]

The authors Hu et al. [51] developed the MetaCyto Tool for automated meta-analysis of both flow and mass cytometry (CyTOF) data.

- PythonMeta [4]

PythonMeta package performs the meta-analysis on an open-access dataset from Cochrane.

4.2.10. Summarize/Synthesis of Data (Analysis) (Includes One Tool)

Although software tools have long existed to support the data synthesis component of reviews (especially to perform meta-analyses), methods for automating them are beyond the capabilities of the available ML and NLP tools [16]. However, research in these areas continues apace. Thus, a related tool, recently developed, is described below:

- Visae [52]

Visae is an app developed in R that uses correspondence analysis to help summarize data on adverse events from clinical trials. It is built on the underlying approach of applying

stacked correspondence analysis and contribution biplots to help explore differences in adverse events among interventions within clinical trials

4.2.11. Write Up (Includes Two Tools)

These tools help with the auto-generation of the abstract, results, and discussion sections of a review.

Some of these tools are described below:

- Endnote (<https://endnote.com/>, accessed on 8 August 2022)

Endnote screen abstracts, obtains full texts, and writes up SR. Accelerates multiple tasks and it assists with reference management. Useful for storing search results, finding full texts, sorting into groups during screening, and to insert references into the manuscript.

- RevManHAL [53]

RevManHAL is an add-on program, which helps auto-generate the abstract, results, and discussion sections of RevMan-generated reviews in multiple languages.

4.2.12. Data Miner/Analysis of Data for General-Purpose (Includes Five Tools)

These are toolkits that support ML and data mining processes.

Some of these tools are described below:

- RapidMiner [5,37]

RapidMiner supports predictive analysis with its user-friendly, rich library of data science and ML algorithms through its all-in-one programming environments such as RapidMiner Studio. Besides the standard data mining features such as data cleansing, filtering, clustering, etc., the software also features built-in templates, repeatable work flows, a professional visualization environment, and seamless integration with languages.

- WEKA [5,37,54–57]

WEKA is a widely used toolkit for ML and data mining that was originally developed. It contains a large collection of state-of-the-art ML and data mining algorithms written in Java. WEKA contains tools for regression, classification, clustering, association rules, visualization, and data pre-processing.

- KNIME [5,58]

KNIME is an open source data analysis platform. It allows the user to create workflows for processing and analyzing almost any kind of data. Written in Java and built upon Eclipse, its access is through a GUI that provides options to create the data flow and conduct data pre-processing, collection, analysis, modeling, and reporting.

- COKE [23,24]

The COKE (COVID-19 Knowledge Extraction framework for next generation discovery science) project involves the use of machine reading and deep learning to design and implement a semi-automated system that supports and enhances the SLR and guideline drafting processes. Specifically, the authors propose a framework for aiding in the literature selection and navigation process that employs natural language processing and clustering techniques for selecting and organizing the literature for human consultation, according to PICO (Population/Problem, Intervention, Comparison, and Outcome) elements.

- KEEL (<http://keel.es/>, accessed on 8 August 2022), [59–63]

KEEL (Knowledge Extraction for Evolutionary Learning) is a Java-based open source tool. It is powered by a well-organized GUI that lets you manage (import, export, edit, and visualize) data with different file formats, and to experiment with the data (through its data pre-processing, statistical libraries, and some standard data mining and evolutionary learning algorithms).

Summarizing the results obtained from this study, it is worth mentioning that advances in technology have revolutionized the healthcare sector. ML has helped create tools and methods for the effective management of data in healthcare [64].

Data mining, also known as knowledge discovery from databases, is a process of mining and analyzing enormous amounts of data and extracting information from it [33].

The growing interest in the extraction of useful knowledge from data with the aim of being beneficial for the data owner has given rise to multiple data mining tools [35].

More specifically, this review produced the following results:

- Using MLTs to assist with data extraction resulted in performance gains compared with using manual extraction.
- At the same time, the use of MLTs has enough flexibility and can speed up and further improve the results of meta-analyses.
- In summary, there are a number of data mining tools available in the digital world that can help researchers with the evaluation of the clinical trials outputs [34]. Evaluations from applying ML to datasets and clinical studies show that this approach could yield promising results.

Evaluations of these tools were found in a number of articles identified by this study.

Specifically, the initial use of automated tools resulted in modest [21] to satisfactory time [29] savings compared with manual management.

In addition, the evaluation of the performance, functionality, usability, user interface, and system requirements also yielded positive results [35].

Moreover, the evaluation of some tools in terms of acceptance [50], feasibility [50], precision [49], accuracy [42], efficiency [21], efficacy [65], and reliability [21] was also positive.

5. Discussion

The whole idea of developing ML is associated with achieving faster, more efficient, and more reliable results in the health sector. ML mainly comprises algorithms that, when put together, have the power to diagnose, display results, and feed data into databases faster than the traditional method of entering the data manually. Nowadays, as more clinically relevant datasets are available electronically, researchers have applied ML techniques to a wide range of clinical tasks [64].

As reported in the literature, many benefits arise from MLTs in the field of extracting clinical trial results.

More specifically, ML has been described as “the key technology” for the development of precision medicine [4]. ML uses computer algorithms to build predictive models based on complex patterns in data. ML can integrate the large amounts of data required to “learn” the complex patterns required for accurate medical predictions. ML has excelled in automated meta-analysis, extraction of data from clinical trials and text mining, semi-automates bias assessments, and in specific medical domains.

The aim of this article is to discover data mining tools used in EBHI and to provide the research community with an extensive study based on a wide set of features that any tool should satisfy. In this paper, the author addresses the interest of data mining and describes the most popular mining tools used in EBHI, and especially to extract clinical trial results.

Although there is no tool that can automate the entire knowledge extraction process, the author identified a broad evidence base of publications describing the overview of (semi)automated data-extraction literature in order to extract the results from the clinical trials.

However, the lack of publicly available gold-standard data for evaluation, and the lack of application thereof, makes it difficult to draw conclusions about which is the best-performing system for each data extraction target [66].

This review aims to present the appropriate MLTs that will allow for faster and more reliable extraction of information and knowledge from clinical trials and related to the prognosis, diagnosis, treatment, and drug use, as related studies are limited.

There are a limited number of relevant studies. However, these either refer to algorithms and techniques or study the performance of an MLT with applications in clinical trials. The reviews that focus on the subject of data extraction from clinical trial data either present a small sample of MLTs or deal with a specialized task.

Thus, the contribution of this study is the renewal of existing knowledge by presenting a large number of older and more modern tools for extracting information and knowledge from the outputs of clinical studies. MLTs of more general use are also presented, i.e., tools that are not limited to the management of RCTs, but that can be used in them as well.

Nevertheless, this review aims first to explore the options available for automating information and knowledge extraction in this domain. A more detailed and in-depth review will follow in the future.

In addition, the present study has some methodological limitations. Initially, the author had some difficulty in identifying suitable articles. This limitation was partially addressed through the use of snowballing methods. Secondly, the author included articles written only in English.

In addition, it was not possible to present a consolidated list with a common rating. This happened because each author adopted different evaluation criteria for the tools they presented.

6. Conclusions and Future Directions

Evidence-based knowledge synthesis in medicine, i.e., clinical trials, is rapidly becoming unfeasible due to the extremely rapid increase in evidence production. At the same time, limited resources (in cost, human resources, time, and money) can be better used with computational assistance and automation to significantly improve the process of extracting knowledge from clinical trials. In addition, advances in the automation of systematic reviews of clinical trials will provide clinicians with more evidence-based answers and thus enable them to provide higher quality information [17].

Sequentially, ML is the fastest growing field in computer science and in accordance with Health Informatics, one of the biggest challenges has become providing improvements in medical diagnoses, disease analysis, and drug development in the future [67].

In summary, design based on the application of clinical trial outputs in ML is a promising approach to implement more effective solutions.

However, more studies are needed in the future for clinical and technical validation of the performance of ML tools in the health sector. Among other things, future research should focus on studying the assessment characteristics in order to propose common measurement standards and assessment mechanisms for these MLTs.

It is also important to conduct a systematic review analytically and precisely evaluate, and to apply strict evaluation criteria to the MLTs. In this way, it becomes possible to choose the right MLT for each case.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data from this research are not available elsewhere. Please contact the author for more information, if required.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

AI	Artificial intelligence
ML	Machine learning
MLT	Machine learning tool
SE	Software engineering
SLR	Systematic literature review
SR	Systematic review
R	Review
RCT	Randomized controlled trial

References

1. Ammenwerth, E.; de Keizer, N. A viewpoint on evidence-based health informatics, based on a pilot survey on evaluation studies in health care informatics. *J. Am. Med. Inform. Assoc.* **2007**, *14*, 368–371. [[CrossRef](#)] [[PubMed](#)]
2. Sargeant, J.M.; Kelton, D.F.; O'Connor, A.M. Study Designs and Systematic Reviews of Interventions: Building Evidence Across Study Designs. *Zoonoses Public Health* **2014**, *61*, 10–17. [[CrossRef](#)] [[PubMed](#)]
3. Cheng, L.; Katz-Rogozhnikov, D.A.; Varshney, K.R.; Baldini, I. Automated meta-analysis: A causal learning perspective. *arXiv* **2021**. [[CrossRef](#)]
4. Masoumi, S.; Shahraz, S. Meta-analysis using Python: A hands-on tutorial. *BMC Med. Res. Methodol.* **2022**, *22*, 193. [[CrossRef](#)]
5. Pynam, V.; Spanadna, R.R.; Srikanth, K. An Extensive Study of Data Analysis Tools (Rapid Miner, Weka, R Tool, Knime, Orange). *Int. J. Comput. Sci. Eng.* **2018**, *5*, 4–11. [[CrossRef](#)]
6. Steinhubl, S.R.; Wolff-Hughes, D.L.; Nilsen, W.; Iturriaga, E.; Califf, R.M. Digital clinical trials: Creating a vision for the future. *npj Digit. Med.* **2019**, *2*, 126. [[CrossRef](#)]
7. Rosa, C.; Marsch, L.A.; Winstanley, E.L.; Brunner, M.; Campbell, A.N.C. Using digital technologies in clinical trials: Current and future applications. *Contemp. Clin. Trials* **2021**, *100*, 106219. [[CrossRef](#)]
8. Inan, O.T.; Tenaerts, P.; Prindiville, S.A.; Reynolds, H.; Dizon, D.S.; Cooper-Arnold, K.; Turakhia, M.; Pletcher, M.J.; Preston, K.L.; Krumholz, H.M.; et al. Digitizing clinical trials. *Npj Digit. Med.* **2020**, *3*, 107. [[CrossRef](#)]
9. McClendon, L.; Meghanathan, N. Using Machine Learning Algorithms to Analyze Crime Data. *Mach. Learn. Appl. Int. J.* **2015**, *2*, 2101. [[CrossRef](#)]
10. Habehh, H.; Gohel, S. Machine Learning in Healthcare. *Curr. Genom.* **2021**, *22*, 291–300. [[CrossRef](#)]
11. Rong, G.; Mendez, A.; Assi, E.B.; Zhao, B.; Sawan, M. Artificial Intelligence in Healthcare: Review and Prediction Case Studies. *Engineering* **2020**, *6*, 291–301. [[CrossRef](#)]
12. Kumar, Y.; Mahajan, M. 5. Recent advancement of machine learning and deep learning in the field of healthcare system. In *Computational Intelligence for Machine Learning and Healthcare Informatics*; Walter de Gruyter GmbH & Co KG: Berlin, Germany, 2020; pp. 77–98. [[CrossRef](#)]
13. Jonnalagadda, S.R.; Goyal, P.; Huffman, M.D. Automating data extraction in systematic reviews: A systematic review. *Syst. Rev.* **2015**, *4*, 78. [[CrossRef](#)]
14. Chen, X.; Xie, H.; Cheng, G.; Poon, L.K.M.; Leng, M.; Wang, F.L. Trends and Features of the Applications of Natural Language Processing Techniques for Clinical Trials Text Analysis. *Appl. Sci.* **2020**, *10*, 2157. [[CrossRef](#)]
15. Harrer, S.; Shah, P.; Antony, B.; Hu, J. Artificial Intelligence for Clinical Trial Design. *Trends Pharmacol. Sci.* **2019**, *40*, 577–591. [[CrossRef](#)]
16. Marshall, I.J.; Wallace, B.C. Toward systematic review automation: A practical guide to using machine learning tools in research synthesis. *Syst. Rev.* **2019**, *8*, 163. [[CrossRef](#)]
17. Tsafnat, G.; Glasziou, P.; Choong, M.K.; Dunn, A.; Galgani, F.; Coiera, E. Systematic review automation technologies. *Syst. Rev.* **2014**, *3*, 74. [[CrossRef](#)]
18. Wang, Y.; Carter, B.Z.; Li, Z.; Huang, X. Application of machine learning methods in clinical trials for precision medicine. *JAMIA Open* **2022**, *5*, ooab107. [[CrossRef](#)]
19. Izonin, I.; Tkachenko, R.; Kryvinska, N.; Tkachenko, P. Multiple Linear Regression Based on Coefficients Identification Using Non-iterative SGTm Neural-like Structure. In *Advances in Computational Intelligence*; Springer: Cham, Switzerland, 2019; pp. 467–479.
20. Felizardo, K.R.; Carver, J.C. Automating Systematic Literature Review. In *Contemporary Empirical Methods in Software Engineering*; Springer: Cham, Switzerland, 2020; pp. 327–355. [[CrossRef](#)]
21. Gates, A.; Gates, M.; Sim, S.; Elliott, S.A.; Pillay, J.; Hartling, L. Creating efficiencies in the extraction of data from randomized trials: A prospective evaluation of a machine learning and text mining tool. *BMC Med. Res. Methodol.* **2021**, *21*, 169. [[CrossRef](#)]
22. Kiritchenko, S.; De Bruijn, B.; Carini, S.; Martin, J.; Sim, I. ExaCT: Automatic extraction of clinical trial characteristics from journal publications. *BMC Med. Inform. Decis. Mak.* **2010**, *10*, 56. [[CrossRef](#)]
23. Golinelli, D.; Nuzzolese, A.G.; Sanmarchi, F.; Bulla, L.; Mongiovì, M.; Gangemi, A.; Rucci, P. Semi-Automatic Systematic Literature Reviews and Information Extraction of COVID-19 Scientific Evidence: Description and Preliminary Results of the COKE Project. *Information* **2022**, *13*, 117. [[CrossRef](#)]

24. Khangura, S.; Konnyu, K.; Cushman, R.; Grimshaw, J.; Moher, D. Evidence summaries: The evolution of a rapid review approach. *Syst. Rev.* **2012**, *1*, 10. [\[CrossRef\]](#)
25. Greenhalgh, T.; Peacock, R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources. *BMJ* **2005**, *331*, 1064–1065. [\[CrossRef\]](#)
26. Manktelow, M.; Iftikhar, A.; Bucholc, M.; McCann, M.; O’Kane, M. Clinical and operational insights from data-driven care pathway mapping: A systematic review. *BMC Med. Inform. Decis. Mak.* **2022**, *22*, 43. [\[CrossRef\]](#)
27. Christopoulou, S.C.; Kotsilieris, T.; Anagnostopoulos, I. Assessment of Health Information Technology Interventions in Evidence-Based Medicine: A Systematic Review by Adopting a Methodological Evaluation Framework. *Healthcare* **2018**, *6*, 109. [\[CrossRef\]](#)
28. Clark, J.; McFarlane, C.; Cleo, G.; Ramos, C.I.; Marshall, S. The Impact of Systematic Review Automation Tools on Methodological Quality and Time Taken to Complete Systematic Review Tasks: Case Study. *JMIR Med. Educ.* **2021**, *7*, e24418. [\[CrossRef\]](#)
29. Clark, J.; Glasziou, P.; del Mar, C.; Bannach-Brown, A.; Stehlik, P.; Scott, A.M. A full systematic review was completed in 2 weeks using automation tools: A case study. *J. Clin. Epidemiol.* **2020**, *121*, 81–90. [\[CrossRef\]](#)
30. Khalil, H.; Ameen, D.; Zarnegar, A. Tools to support the automation of systematic reviews: A scoping review. *J. Clin. Epidemiol.* **2022**, *144*, 22–42. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Erickson, B.J.; Korfiatis, P.; Akkus, Z.; Kline, T.; Philbrick, K. Toolkits and Libraries for Deep Learning. *J. Digit. Imaging* **2017**, *30*, 400–405. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Cleo, G.; Scott, A.M.; Islam, F.; Julien, B.; Beller, E. Usability and acceptability of four systematic review automation software packages: A mixed method design. *Syst. Rev.* **2019**, *8*, 145. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Shravan, I.V. Top 10 Open Source Data Mining Tools. Open Source For You, CreateSpace Independent Publishing Platform, Delhi NCR, India. 2017. Available online: <https://www.opensourceforu.com/2017/03/top-10-open-source-data-mining-tools/> (accessed on 1 September 2022).
34. Ratra, R.; Gulia, P. Experimental Evaluation of Open Source Data Mining Tools (WEKA and Orange). *Int. J. Eng. Trends Technol.* **2020**, *68*, 30–35. [\[CrossRef\]](#)
35. Altalhi, A.H.; Luna, J.M.; Vallejo, M.A.; Ventura, S. Evaluation and comparison of open source software suites for data mining and knowledge discovery. *WIREs Data Min. Knowl. Discov.* **2017**, *7*, e1204. [\[CrossRef\]](#)
36. Dwivedi, S.; Kasliwal, P.; Soni, S. Comprehensive study of data analytics tools (RapidMiner, Weka, R tool, Knime). In Proceedings of the 2016 Symposium on Colossal Data Analysis and Networking (CDAN), Indore, India, 18–19 March 2016. [\[CrossRef\]](#)
37. Naik, A.; Samant, L. Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime. *Procedia Comput. Sci.* **2016**, *85*, 662–668. [\[CrossRef\]](#)
38. Zippel, C.; Bohnet-Joschko, S. Rise of Clinical Studies in the Field of Machine Learning: A Review of Data Registered in ClinicalTrials.gov. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5072. [\[CrossRef\]](#)
39. Marshall, C.; Sutton, A. Systematic Review Toolbox. *Value Health* **2016**, *19*, A398. [\[CrossRef\]](#)
40. Clark, J.M.; Sanders, S.; Carter, M.; Honeyman, D.; Cleo, G.; Auld, Y.; Booth, D.; Condrón, P.; Dalais, C.; Bateup, S.; et al. Improving the translation of search strategies using the Polyglot Search Translator: A randomized controlled trial. *J. Med. Libr. Assoc.* **2020**, *108*, 195–207. [\[CrossRef\]](#)
41. Sośnicki, M.; Madeyski, L. ASH: A New Tool for Automated and Full-Text Search in Systematic Literature Reviews. In *Computational Science—ICCS 2021*; Springer: Cham, Switzerland, 2021; pp. 362–369. [\[CrossRef\]](#)
42. Choong, M.K.; Galgani, F.; Dunn, A.G.; Tsafnat, G. Automatic evidence retrieval for systematic reviews. *J. Med. Internet Res.* **2014**, *16*, e223. [\[CrossRef\]](#)
43. Wallace, B.C.; Small, K.; Brodley, C.E.; Lau, J.; Trikalinos, T.A. Deploying an interactive machine learning system in an evidence-based practice center. In Proceedings of the 2nd ACM SIGHIT symposium on International health informatics—IHI ’12, Miami, FL, USA, 28–30 January 2012. [\[CrossRef\]](#)
44. Shemilt, I.; Khan, N.; Park, S.; Thomas, J. Use of cost-effectiveness analysis to compare the efficiency of study identification methods in systematic reviews. *Syst. Rev.* **2016**, *5*, 140. [\[CrossRef\]](#)
45. Ouzzani, M.; Hammady, H.; Fedorowicz, Z.; Elmagarmid, A. Rayyan—a web and mobile app for systematic reviews. *Syst. Rev.* **2016**, *5*, 210. [\[CrossRef\]](#)
46. Przybyła, P.; Brockmeier, A.J.; Kontonatsios, G.; Le Pogam, M.-A.; McNaught, J.; von Elm, E.; Nolan, K.; Ananiadou, S. Prioritising references for systematic reviews with RobotAnalyst: A user study. *Res Synth. Methods* **2018**, *9*, 470–488. [\[CrossRef\]](#)
47. Walker, V.R.; Schmitt, C.P.; Wolfe, M.S.; Nowak, A.J.; Kulesza, K.; Williams, A.R.; Shin, R.; Cohen, J.; Burch, D.; Stout, M.D.; et al. Evaluation of a semi-automated data extraction tool for public health literature-based reviews: Dextr. *Environ. Int.* **2022**, *159*, 107025. [\[CrossRef\]](#)
48. Marshall, I.J.; Nye, B.; Kuiper, J.; Noel-Storr, A.; Marshall, R.; Maclean, R.; Soboczenski, F.; Nenikova, A.; Thomas, J.; Wallace, B.C. Trialstreamer: A living, automatically updated database of clinical trial reports. *J. Am. Med. Inform. Assoc.* **2020**, *27*, 1903–1912. [\[CrossRef\]](#)
49. Soboczenski, F.; Trikalinos, T.A.; Kuiper, J.; Bias, R.G.; Wallace, B.C.; Marshall, I.J. Machine learning to help researchers evaluate biases in clinical trials: A prospective, randomized user study. *BMC Med. Inform. Decis. Mak.* **2019**, *19*, 96. [\[CrossRef\]](#)
50. Ajiji, P.; Cottin, J.; Picot, C.; Uzunali, A.; Ripoche, E.; Cucherat, M.; Maison, P. Feasibility study and evaluation of expert opinion on the semi-automated meta-analysis and the conventional meta-analysis. *Eur. J. Clin. Pharmacol.* **2022**, *78*, 1177–1184. [\[CrossRef\]](#)

51. Hu, Z.; Jujjavarapu, C.; Hughey, J.J.; Andorf, S.; Lee, H.-C.; Gherardini, P.F.; Spitzer, M.H.; Thomas, C.G.; Campbell, J.; Dunn, P.; et al. MetaCyto: A Tool for Automated Meta-analysis of Mass and Flow Cytometry Data. *Cell Rep.* **2018**, *24*, 1377–1388. [CrossRef]
52. Diniz, M.A.; Gresham, G.; Kim, S.; Luu, M.; Henry, N.L.; Tighiouart, M.; Yothers, G.; Ganz, P.A.; Rogatko, A. Visualizing adverse events in clinical trials using correspondence analysis with R-package visae. *BMC Med. Res. Methodol.* **2021**, *21*, 244. [CrossRef]
53. Torres, M.T.; Adams, C.E. RevManHAL: Towards automatic text generation in systematic reviews. *Syst. Rev.* **2017**, *6*, 27. [CrossRef]
54. Frank, E.; Hall, M.; Trigg, L.; Holmes, G.; Witten, I.H. Data mining in bioinformatics using Weka. *Bioinformatics* **2004**, *20*, 2479–2481. [CrossRef]
55. Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. *Data Mining: Practical Machine Learning Tools and Techniques*; Morgan Kaufmann: Burlington, MA, USA, 2016.
56. Witten, I.H.; Frank, E.; Trigg, L.; Hall, M.; Holmes, G.; Cunningham, S.J. *Weka: Practical Machine Learning Tools and Techniques with Java Implementations*; University of Waikato, Department of Computer Science: Hamilton, New Zealand, 1999.
57. Weka 3—Data Mining with Open Source Machine Learning Software in Java. Available online: www.cs.waikato.ac.nz/ml/weka/ (accessed on 14 August 2022).
58. Meinel, T.; Jagla, B.; Berthold, M.R. Integrated data analysis with KNIME. In *Open Source Software in Life Science Research*; Elsevier: Amsterdam, The Netherlands, 2012; pp. 151–171. [CrossRef]
59. Alcalá-Fdez, J.; Sánchez, L.; García, S.; Del Jesus, M.J.; Ventura, S.; Garrell, J.M.; Otero, J.; Romero, C.; Bacardit, J.; Rivas, V.M.; et al. KEEL: A software tool to assess evolutionary algorithms for data mining problems. *Soft Comput.* **2009**, *13*, 307–318. [CrossRef]
60. Fernández, A.; Luengo, J.; Derrac, J.; Alcalá-Fdez, J.; Herrera, F. Implementation and Integration of Algorithms into the KEEL Data-Mining Software Tool. In *Intelligent Data Engineering and Automated Learning—IDEAL 2009*; Springer Science & Business Media: Berlin, Germany, 2009; pp. 562–569. [CrossRef]
61. Alcalá-Fdez, J.; García, S.; Berlanga, F.J.; Fernández, A.; Sánchez, L.; del Jesus, M.; Herrera, F. KEEL: A data mining software tool integrating genetic fuzzy systems. In *Proceedings of the 2008 3rd International Workshop on Genetic and Evolving Systems*, Witten-Bommerholz, Germany, 4–7 March 2008. [CrossRef]
62. Triguero, I.; González, S.; Moyano, J.M.; García, S.; Alcalá-Fdez, J.; Luengo, J.; Fernández, A.; del Jesús, M.J.; Sánchez, L.; Herrera, F. KEEL 3.0: An Open Source Software for Multi-Stage Analysis in Data Mining. *Int. J. Comput. Intell. Syst.* **2017**, *10*, 1238–1249. [CrossRef]
63. Available online: http://www.keel.es/software/KEEL_template.zip (accessed on 10 August 2022).
64. Rashid, S.; Kathuria, N. Machine Learning in Clinical Trials. In *Big Data and Artificial Intelligence for Healthcare Applications*; CRC Press: Boca Raton, FL, USA, 2021; pp. 69–82. [CrossRef]
65. Margulis, E.; Dagan-Wiener, A.; Ives, R.S.; Jaffari, S.; Siems, K.; Niv, M.Y. Intense bitterness of molecules: Machine learning for expediting drug discovery. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 568–576. [CrossRef] [PubMed]
66. Schmidt, L.; Olorisade, B.K.; McGuinness, L.A.; Thomas, J.; Higgins, J.P.T. Data extraction methods for systematic review (semi)automation: A living systematic review. *F1000Research* **2021**, *10*, 401. [CrossRef] [PubMed]
67. Holzinger, A. *Machine Learning for Health Informatics: State-of-the-Art and Future Challenges*; Springer: Cham, Switzerland, 2016.