

A literature review on disease detection with automated machine learning

Ersin Enes Eryılmaz and Erdal Kılıç

Department of Computer Engineering, Ondokuz Mayıs University, Samsun, Turkey

ABSTRACT

This study reviews disease detection with Automated Machine Learning (AutoML), aiming to identify gaps and evaluate AutoML's impact in this field. In this study, seven review articles published in Q1- or Q2-quartile journals between 2020 and 2025 were analyzed. The reviews were assessed using ten academic criteria, covering AutoML performance, data strategies, feature techniques, noise reduction, model selection, training/testing methods, and frameworks for disease detection. Additionally, test reliability, patient selection, reference standards, and application processes were evaluated with the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool. A literature review was conducted using 11 different databases; however, due to limited functionality in four of them, the research primarily relied on seven digital databases, which initially yielded 552 studies. The study selection and screening processes were performed in accordance with the Preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines. Next, 40 studies published outside the 2020–2025 period were removed, followed by the exclusion of 117 studies that were not journal articles. An additional 145 studies were eliminated because they were reviews, books, conference proceedings, posters, editorial notes, *etc.*, and seven studies were excluded as they did not pertain to human diseases. After these elimination processes, 243 articles remained for full-text review. Out of these, 214 articles were read in full and assessed for relevance, leading to 29 articles deemed suitable for inclusion in this review on disease detection using AutoML. After removing five duplicate articles, a final total of 24 studies were included in the review. The research questions of the study include questions such as which disease detection models AutoML methods are preferred more, the input features and data sets used, the effects of feature extraction and selection methods on model performance, how often noise reduction methods are used in disease data, and what the AutoML model evaluation metrics are. The results show that AutoML methods are effectively used on disease detection and that different AutoML techniques, data sets, and model selection processes make significant contributions to success. This review provides an important resource for making AutoML applications for disease detection more efficient and for eliminating the deficiencies in the literature.

Submitted 5 May 2025
Accepted 14 August 2025
Published 22 September 2025

Corresponding author
Ersin Enes Eryılmaz,
enes.eryilmaz@bil.omu.edu.tr

Academic editor
Shibiao Wan

Additional Information and
Declarations can be found on
page 26

DOI 10.7717/peerj-cs.3193

© Copyright
2025 Eryılmaz and Kılıç

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Artificial Intelligence, Data Mining and Machine Learning, Data Science, Text Mining, Neural Networks

Keywords Machine learning, AutoML, Automated machine learning, Deep learning, Disease detection, Feature engineering, Feature selection, Artificial intelligence, Disease diagnosis, Disease prediction

INTRODUCTION

Automated machine learning (AutoML) simplifies machine learning processes, enabling a wider audience to benefit from this technology. By automating complex steps such as data preprocessing, model selection, and hyperparameter tuning, it helps even non-experts develop effective models. This allows data scientists to spend their time on more strategic and creative tasks. Additionally, AutoML tools increase the consistency of results using existing best practices ([Salehin et al., 2024](#)). The primary objective of AutoML is to enhance the accessibility and efficiency of machine learning workflows. It enables data scientists to speed up and optimize the model development process by including steps such as automatic data preprocessing, model or algorithm selection, and hyperparameter optimization. At the same time, by automating these processes, it becomes easier for non-machine learning experts to benefit from machine learning. This democratization allows AutoML to accelerate innovations and technological developments in different sectors, while also supporting the implementation of machine learning projects with fewer resources.

AutoML also minimizes the need for manual tuning in machine learning through automatic optimization methods that try the best parameter combinations to improve model performance, helping to make machine learning applications more reliable, accurate, and sustainable.

Several widely utilized platforms in this domain include Google Cloud AutoML, H2O AutoML Framework (H2O), artificial intelligence (AI), and Tree-Based Pipeline Optimization Tool (TPOT), which are among the prominent tools employed for implementing AutoML techniques in disease detection and related biomedical applications. These tools provide fast and effective solutions by allowing users to simply upload their data and start processing. The development of AutoML increases the accessibility of machine learning, making it available in more industries ([Romero et al., 2022](#)). Abbreviations and expansions are shown in [Table S1](#) and are presented in [Supplemental Files](#).

AutoML enables the automation of machine learning workflows, including data preprocessing, feature engineering, model selection, and hyperparameter tuning. This reduces the need for manual intervention and allows both experts and non-experts to develop high-performance models efficiently. Widely used AutoML platforms such as H2O, TPOT, and Google Cloud AutoML provide scalable and accessible solutions across various domains, particularly in healthcare. The main objective of this review is to evaluate the effects of AutoML methods on disease detection and to reveal the gaps existing in the literature in this field.

AUTOML

AutoML distinguishes itself from conventional machine learning approaches by automating key stages of the pipeline, including data preprocessing, feature engineering, model selection, and hyperparameter optimization ([Bergstra et al., 2015](#)). AutoML uses methods such as neural architecture search, transfer learning, and reinforcement learning to identify and use the most suitable model, reducing the need for manual trial and error.

While in traditional ML processes, these stages are performed manually by experts, AutoML streamlines these steps, making model development faster and more efficient. Additionally, AutoML platforms often include automated evaluation and optimization mechanisms, ensuring that models are fine-tuned for optimal performance with minimal human intervention.

AutoML is more accessible and can be used by the masses, including non-experts and domain specialists who may not have extensive machine learning expertise, while traditional methods require deep theoretical and practical knowledge. However, while AutoML provides less human intervention, scalability, and faster deployment, traditional ML offers more control, flexibility, and customization for domain-specific requirements. In highly specialized applications where expert-driven feature engineering and fine-tuning are necessary, traditional ML may still be preferred. Nonetheless, the rapid advancements in AutoML continue to bridge the gap between automation and expert-driven customization, making it an increasingly powerful tool in machine learning workflows.

In recent years, due to the increasing data volume and demand for machine learning solutions, automatic data processing has gained great importance and studies on these have begun to appear in the literature. The studies primarily addressed individual solutions such as data preprocessing (cleaning, missing data completion, labeling, categorical coding, etc.), data augmentation, and feature engineering (feature extraction, generation, and selection). Additionally, the study explored integrated methodologies that consolidate all processing stages into a cohesive, end-to-end deep learning framework. It also offered an in-depth analysis of the fundamental features of general AutoML frameworks tailored for large-scale data applications and critically assessed potential advancements that could improve the efficiency of automated data analysis (Mumuni & Mumuni, 2024).

AutoML systems automate the following steps:

Data ingestion and preprocessing: Filling in missing values, data normalization, coding categorical variables.

Feature engineering: Deriving new features and selecting the most meaningful ones.

Model selection: Determining the most appropriate algorithms (e.g., tree-based methods, deep learning models).

Hyperparameter optimization and fine tuning: Finding the best combination of parameters that will improve model performance.

Evaluation and validation: Evaluating the performance of the model and making necessary improvements.

The main use of AutoML is to assist the developer by automating model selection and hyperparameter optimization. Non-experts can use AutoML with minimal or no code. Some AutoML systems include meta learning and transfer learning approaches. AutoML makes it effortless to build machine learning pipeline processes. An AutoML framework pipeline are shown in Fig. 1.

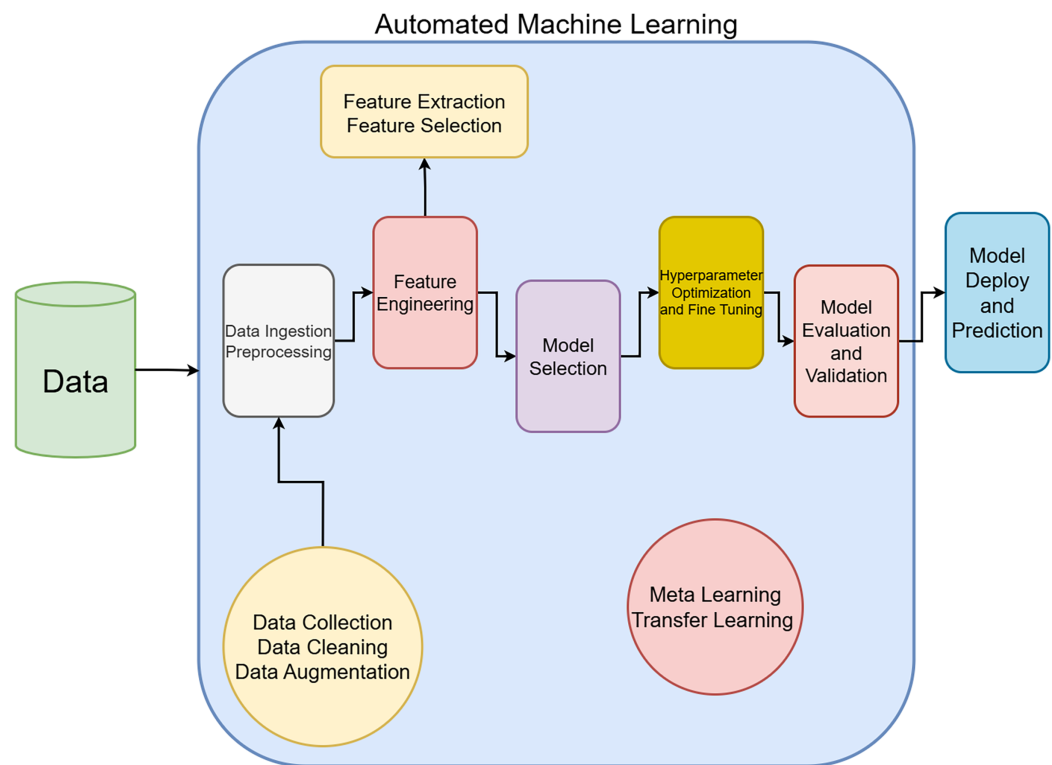


Figure 1 AutoML pipeline.

Full-size DOI: [10.7717/peerj-cs.3193/fig-1](https://doi.org/10.7717/peerj-cs.3193/fig-1)

The main use of AutoML is to assist the developer by automating model selection and hyperparameter optimization. Non-experts can use AutoML with minimal or no line of code. AutoML makes it effortless to create machine learning pipeline processes. In an AutoML process, incoming data is first processed, then feature engineering is performed to extract features. Model selection and HPO are performed to reach the result. AutoML tries to guarantee high accuracy, efficiency, scalability, and flexibility while performing these processes.

In data preprocessing steps, data collection, data cleaning and data augmentation (Mumuni & Mumuni, 2022) techniques are used. The basic feature engineering steps in the AutoML workflow are; selection and preprocessing of features, feature transformation and encoding, feature generation, feature selection, automatic optimization of features, and alignment of features with the model (Hutter, Kotthoff & Vanschoren, 2019). Some automatic feature engineering methods include AutoFeat, Deep & Cross Network Version 2 (DCN-V2), Open Feature Engineering (OpenFE), Subsampled Approximate Feature Engineering (SAFE), Automated Feature Crossing (Autocross), Feature Engineering by Constrained Optimization Heuristic (FETCH), and Feature Construction Tree (FCTree) (Mumuni & Mumuni, 2024). By automating feature engineering with minimal human intervention, AutoML delivers a faster and more efficient model development process.

The final step in AutoML is model selection and hyperparameter optimization. Today, there are various AutoML approaches developed for different purposes. Table S2 provides

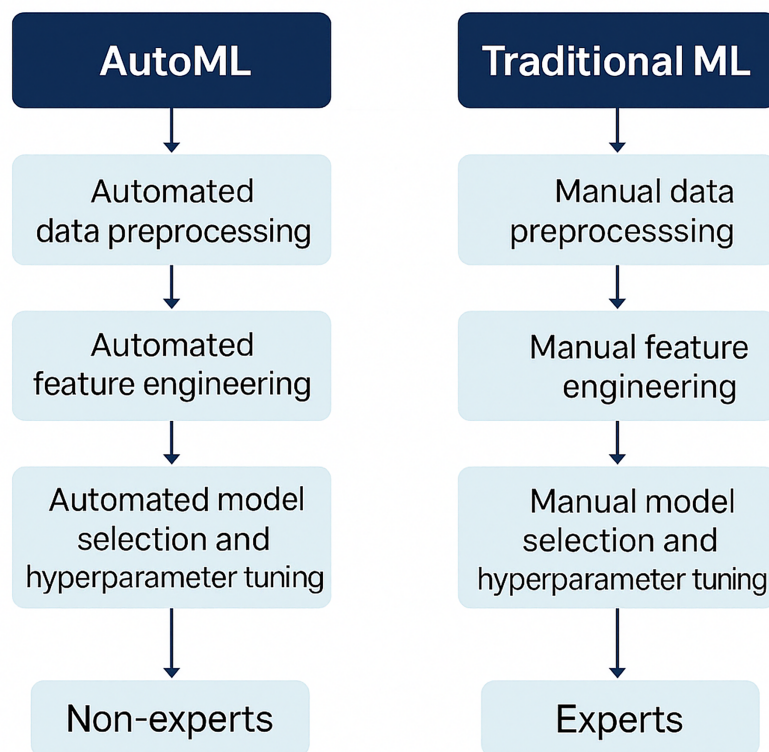


Figure 2 The differences between AutoML and traditional ML.

Full-size DOI: [10.7717/peerj-cs.3193/fig-2](https://doi.org/10.7717/peerj-cs.3193/fig-2)

the basic features of these AutoML platforms, tools and are presented in [Supplemental Files](#) ([Mustafa & Rahimi Azghadi, 2021](#); [Mumuni & Mumuni, 2024](#)).

This study presents a literature review on the application of AutoML in disease detection. To this end, a comprehensive search was performed across eleven distinct digital databases utilizing predefined search strategies to ensure an extensive retrieval of pertinent studies. The initial pool of studies was refined through the application of clearly defined inclusion and exclusion criteria, with emphasis placed on publication timeframe, thematic relevance, methodological soundness, and the integration of AutoML techniques. In the final phase, the selected studies were subjected to an in-depth analysis, focusing on elements such as the datasets employed, model efficacy, evaluation methodologies, and the specific AutoML algorithms implemented. The outcomes of this review offer valuable insights into prevailing trends, existing challenges, and prospective avenues for future research within the domain of AutoML-based disease detection. The differences between AutoML and traditional ML are shown in [Fig. 2](#).

MOTIVATION AND CONTRIBUTIONS

Based on the literature review conducted in this study, it was observed that a comprehensive assessment encompassing all relevant criteria for evaluating model performance—specifically in the context of data acquisition and analysis, feature

extraction and selection, noise reduction, model selection, architectural design, and model training—remains limited within the current literature on AutoML for disease detection. To bridge this research gap, a focused investigation was carried out on review articles addressing AutoML applications. In this regard, seven review articles published in journals in the Q1 or Q2 quarter between 2020 and 2025 were meticulously examined.

These reviews were evaluated against ten predefined criteria, the outcomes of which are presented in [Table S3](#) and are presented in [Supplemental Files](#). The selection process was restricted to English-language review articles published between 2020 and 2025, specifically targeting the application of AutoML in the detection of human diseases. During the screening phase, only articles categorized as reviews within academic databases and available as open-access, full-text documents were considered. The search queries utilized to retrieve the relevant literature are detailed in [Table S4](#) and are presented in [Supplemental Files](#).

- Q1: How was the literature search process carried out across relevant databases for the studies included in the review? Were any statistical details reported regarding the selection process? In other words, was the review conducted in a systematic manner?
- Q2: Have the studies reviewed been individually summarized, and have the predictive models proposed in these studies been articulated in general terms, along with a discussion of their contributions to the existing body of literature?
- Q3: Has there been an extensive investigation on the input features utilized in predictive models?
- Q4: Has information been provided regarding the feature selection or feature extraction techniques employed for prediction models?
- Q5: Has there been a significant emphasis on suggested methodologies for mitigating or minimizing noise in disease-related data?
- Q6: Is there a comprehensive analysis of the various training and testing methodologies employed in predictive modeling?
- Q7: Has any information been provided regarding the specific AutoML techniques that the proposed prediction models in the studies emphasize?
- Q8: Are the performance evaluation metrics for the proposed models clearly defined?
- Q9: Has there been a detailed AutoML research on general human diseases?
- Q10: Have searches been conducted in many well-known digital databases?

[Preity & Shahnawazuddin \(2024\)](#) covers automatic AI-based methods for eye vessel segmentation and disease detection. Diseases diagnosed by retinal vessel analysis are examined by comparing deep learning with traditional image processing models. The role of models such as CNN, U-Net and GAN in the detection of eye diseases is discussed. The study reveals current challenges and future research directions in the context of datasets, evaluation metrics and clinical applications.

In [Yuan et al. \(2024\)](#), 118 articles were analyzed and AutoML's data preparation, feature engineering and model development processes were examined, and its advantages over

classical ML were demonstrated with case studies. Interpretation methods were summarized under the titles of feature interaction, data dimensionality reduction, internal interpretable models and rule extraction, and the use of AutoML in image, text, table, signal, genome and multimodality data was discussed. While interpretable AutoML increases the trust in ML in healthcare, it is recommended that issues such as data preparation and integration of basic models be further investigated in the future.

[Mishra, Pandey & Malhotra \(2024\)](#) examines the role of deep learning in combating neglected vector-borne diseases. Deep learning techniques have surpassed traditional methods in disease transmission risk prediction, vector detection, parasite classification, and treatment optimization. Convolutional neural networks and AutoML algorithms facilitate early detection and disease surveillance, especially in resource-limited regions. Interdisciplinary integration and smartphone-based applications have the potential to improve global health outcomes.

The advantages of codeless deep learning (CFDL) over bespoke DL were evaluated in five ophthalmological tasks, but it was determined that most of such discussions were one-dimensional and that there were wide applicability gaps. It was noted that high-quality assessment of the applicability of CFDL over bespoke DL required a context-specific, weighted assessment of clinician intent, patient acceptance, and cost-effectiveness. It was concluded that CFDL and bespoke DL are unique in their own right and cannot be substituted for each other ([Wong et al., 2024](#)).

This review examines the key challenges and best practices in ML approaches in microbiome research ([Papoutsoglou et al., 2023](#)). The high dimensionality, heterogeneity, and noisy nature of microbiome data complicate the accuracy and generalizability of ML models. The study highlights issues such as data preprocessing, feature selection, model evaluation, and XAI. It notes that tools such as JADBIO can be used to analyze microbiome data. It also recommends best practices such as using open data, standardized methodologies, and reproducibility.

In their article, [Mustafa & Rahimi Azghadi \(2021\)](#) examine the role of AutoML in the healthcare sector, especially how it is used in the analysis of clinical notes. It emphasizes the benefits it provides such as interpreting clinical texts with natural language processing (NLP) techniques, accelerating the diagnostic process and reducing the workload of healthcare professionals. Challenges such as data complexity, explainability and reliability are also addressed. Future research areas and opportunities for AutoML to become more effective in the healthcare field are discussed.

A total of 82 studies conducted in accordance with the PROSPERO protocol and covering studies in Cochrane, Embase, MEDLINE, and Scopus databases up to July 11, 2022 evaluated 26 different AutoML models. Brain and lung diseases were the most frequently investigated domains, yet the performance of AutoML systems exhibited considerable variability (AUCROC: 35.0–100.0%; F1-score: 16.0–99.0%; AUPRC: 51.0–100.0%). In the majority of trials, AutoML systems achieved optimal performance with an AUCROC of 75.6%, an F1-score of 42.3%, and an AUPRC of 83.3%. Among the tools evaluated, AutoPrognosis (for structured data) and Amazon Rekognition (for unstructured data) delivered the best performance outcomes. However, the overall

reporting quality was suboptimal, with a median DECIDE-AI score of 14 out of 27, underscoring the need for improved standards in clinical applications ([Thirunavukarasu et al., 2023](#)).

As a result of the reviews, most of the seven review studies summarized the studies one by one, and the prediction models proposed in these studies were outlined. The evaluation metrics employed to assess model performance were outlined, and the contributions of the studies to the existing body of literature were examined. It was observed that the focus was on model performance evaluation, and different training and testing models were evaluated. Although the majority of the review studies were systematic, few specifically addressed many of the most prominent digital databases, and many failed to emphasize methods for noise removal and reduction. Although most of the studies detected human diseases with AutoML, some studies detected animal and plant diseases with AutoML. Some studies were on automated disease detection instead of disease detection with AutoML.

The audience it is intended for

Healthcare professionals such as doctors, radiologists, and hospital administrators; academic and clinical researchers such as geneticists and bioinformatics researchers; startups and medical device manufacturers developing health technologies; businesses in the health sector such as insurance companies; and patients who want to monitor their individual health and healthy living enthusiasts. Furthermore, graduate students and postdoctoral researchers engaged in research related to disease detection and diagnosis modeling may find value in the thorough review and analysis provided in this work. In conclusion, this study is also aimed at an academic audience seeking to enhance their comprehension and application of AutoML methodologies in the field of disease detection.

Organization

The remainder of this study is organized as follows: the ‘Methodology of Survey Research’ section offers a detailed account of the procedures used to identify and select the article incorporated into this survey. The ‘Findings and Results’ section begins with a critical synthesis of the reviewed literature, followed by an extensive evaluation aligned with the predefined research questions. Finally, the ‘Conclusions and Recommendations’ section offers a synthesis of the studies reviewed and presents recommendations for future research directions.

METHODOLOGY OF SURVEY RESEARCH

This literature begins with the identification of research questions and advances by concentrating on the review objectives. Subsequently, database searches are conducted to select studies that can provide relevant answers to these questions. These searches are executed using query phrases formulated based on predefined inclusion and exclusion criteria. In the final phase, the process of reviewing the selected studies and excluding those that do not meet the criteria is undertaken. This stage involves assessing the studies’ alignment with the established criteria and their relevance to the review’s objectives. With

this approach, a comprehensive analysis of the literature is presented and a solid foundation is established for future research studies.

The PRISMA protocol ([Moher et al., 2009](#)) was used as a basis when conducting a literature review on disease detection with AutoML. For this purpose, our research question was determined first. Then, the literature search strategy was determined with the databases to be used, search keywords, publication range, language restrictions and inclusion/exclusion criteria. Inclusion criteria (studies that detect diseases with AutoML, academic articles excluding review articles, articles containing clinical data) and exclusion criteria (studies that use only traditional ML, articles that do not provide sufficient performance metrics) were determined for literature selection. In accordance with the PRISMA flow chart; the articles found were listed, duplicates were eliminated, inappropriate ones were removed by title-abstract review, and the final included articles were determined by full-text analysis. For relevant studies; general information, disease type, AutoML tool used, comparative methods and performance metrics were summarized. Risk assessment (bias analysis); Bias (bias) assessment can be performed with tools such as the Cochrane Risk of Bias Tool or QUADAS-2. The QUADAS-2 method was used in this literature review ([Whiting et al., 2011](#)). Graphs and tables were created comparing the performance of AutoML methods. Our literature article was written in accordance with the PRISMA 2020 Reporting Guidelines ([Page et al., 2021](#)). The PRISMA checklist is provided in the [Supplemental Files](#).

QUADAS-2 implementation steps

1. Define study selection criteria

- Determine the characteristics of the studies you will include in your article. (Criteria are determined by query expressions.)
- Must be related to diagnostic tests and include the use of AutoML. (Studies that did not use AutoML were eliminated.)

2. Examine the four main areas of QUADAS-2

Patient selection: Has the appropriate patient population been selected for the study? (Only human diseases were searched; other living beings were not included in the study.)

- Index test: Was the test (*e.g.*, AutoML model) used appropriately in the diagnosis process?
- Reference standard: Was the diagnostic test compared to a reliable gold standard?
- Flow and timing: Were all patients subjected to the same diagnostic tests and reference standard?

3. Evaluate risk of bias and applicability

- Determine whether there is low, high or uncertain risk for each area. (Shown with table, graphic legends and summaries of studies.)
- Show this assessment with a table or graph.

4. Report results

- QUADAS-2 assessment findings are presented with tables and graphs.
- Explain how bias risk was managed: The most cited studies in the last 5 years that detected human diseases with AutoML were selected.

The strategy of survey research

This review aims to accomplish the stated objective by addressing the following research questions:

- RQ1: Which AutoML techniques are most commonly utilized in disease detection models?
- RQ2: What are the primary input features and datasets emphasized by the proposed predictive models?
- RQ3: To what extent are feature extraction, feature selection, and feature engineering techniques employed in disease prediction models, and what impact do these methods have on model performance?
- RQ4: Which techniques are predominantly recommended for denoising disease-related data, and how frequently are they preferred in the existing studies?
- RQ5: What is the most commonly employed approach for training-testing data splitting in AutoML-based prediction models?
- RQ6: What performance evaluation metrics are typically used to assess the effectiveness of prediction models?
- RQ7: What hyperparameter optimization and model evaluation methods are used?

After identifying the research questions, comprehensive literature searches were performed across eleven distinct digital repositories: IEEE Xplore, Scopus, PubMed, Web of Science, ACM Digital Library, Wiley Online Library, ScienceDirect—Elsevier, Google Scholar, SpringerLink, Hindawi and Taylor & Francis. The search space was broadened by employing the “AND” operator for various keywords and the “OR” operator for synonyms of the keywords. The query statements applied in this process, along with details about the sources where they were searched, are outlined in [Table S4](#) and are presented in [Supplemental Files](#). It was determined that an effective filtering could not be done in Google Scholar, SpringerLink, and Taylor & Francis. Because the queries in this database are very general and do not allow advanced search phrases and query statements. Hindawi journals have joined Wiley’s open access journal portfolio. Therefore, Google Scholar, SpringerLink, Hindawi and Taylor & Francis searches were ignored. Searches were conducted across seven digital repositories based on the data in [Table S4](#) presented in the [Supplemental Files](#).

As a result, the studies screened in seven different data from January 2020 to March 2025 are filtered by elimination criteria, inclusion and exclusion criteria, summarized in [Table S5](#), presented in the [Supplemental Files](#). As a result of these screens, 552 studies were found with queries created in seven different digital databases, 40 studies that were not

published between 2020–25 were eliminated, 117 studies that were not published in the journal were eliminated later, 145 studies were eliminated because they were reviews, books, conference proceedings, posters, editorial notes, *etc.*, and seven studies were eliminated because they were not on human diseases. At the end of all these elimination processes, 243 articles were included in the full-text review.

A total of 243 articles were selected, including six conference articles in IEEE Xplore, 18 articles in Scopus, seven articles in Pubmed, 23 articles in WoS, four articles in ACM, 60 articles in Wiley, and 125 articles in ScienceDirect-Elsevier digital libraries. The query expressions specified in [Table S4](#) were used to select the articles. Then, the title, keyword, abstract and article content will be read and finally it will be decided how many articles will be included in this survey content.

Article selection and exclusion

As a result of the queries, inclusion and exclusion criteria were defined to determine whether the 243 studies obtained were suitable for evaluation within the scope of the review. A total of 29 studies from seven different databases were deemed worthy of review. It was determined that some of the studies were included in more than one database. Therefore, duplicates were removed to ensure that each study was listed only once. A total of 5 duplicate studies were removed. In the final stage, a total of 24 studies were subjected to review in our article and abstract, full-text and title readings were performed, and the scanned and selected articles are shown in [Fig. 3](#) according to the digital databases. Additionally, the criteria for screening and selection were established as follows to determine whether the identified studies would be considered for evaluation within the scope of this review:

• Inclusion criteria

- Research articles published between 2020 and 2025
- Articles published in Q1 (journals ranked in the top 25% based on impact factor) or Q2 (journals with an impact factor between the 25th and 50th percentiles)
- Studies focusing on disease detection, diagnosis, and healthcare with AutoML
- Research articles published in the English language

• Exclusion criteria

- Studies that do not concentrate on disease detection and diagnosis using AutoML.
- Studies such as review articles, conference proceedings, books or chapters, poster presentations
- Studies not focusing on human diseases
- Duplicate articles

The PRISMA flow diagram depicts the process of study selection, beginning with 552 identified records, narrowing to 243 screened articles, 29 eligible studies after applying inclusion criteria, and finally including 24 studies in the review. A total of 214 articles were

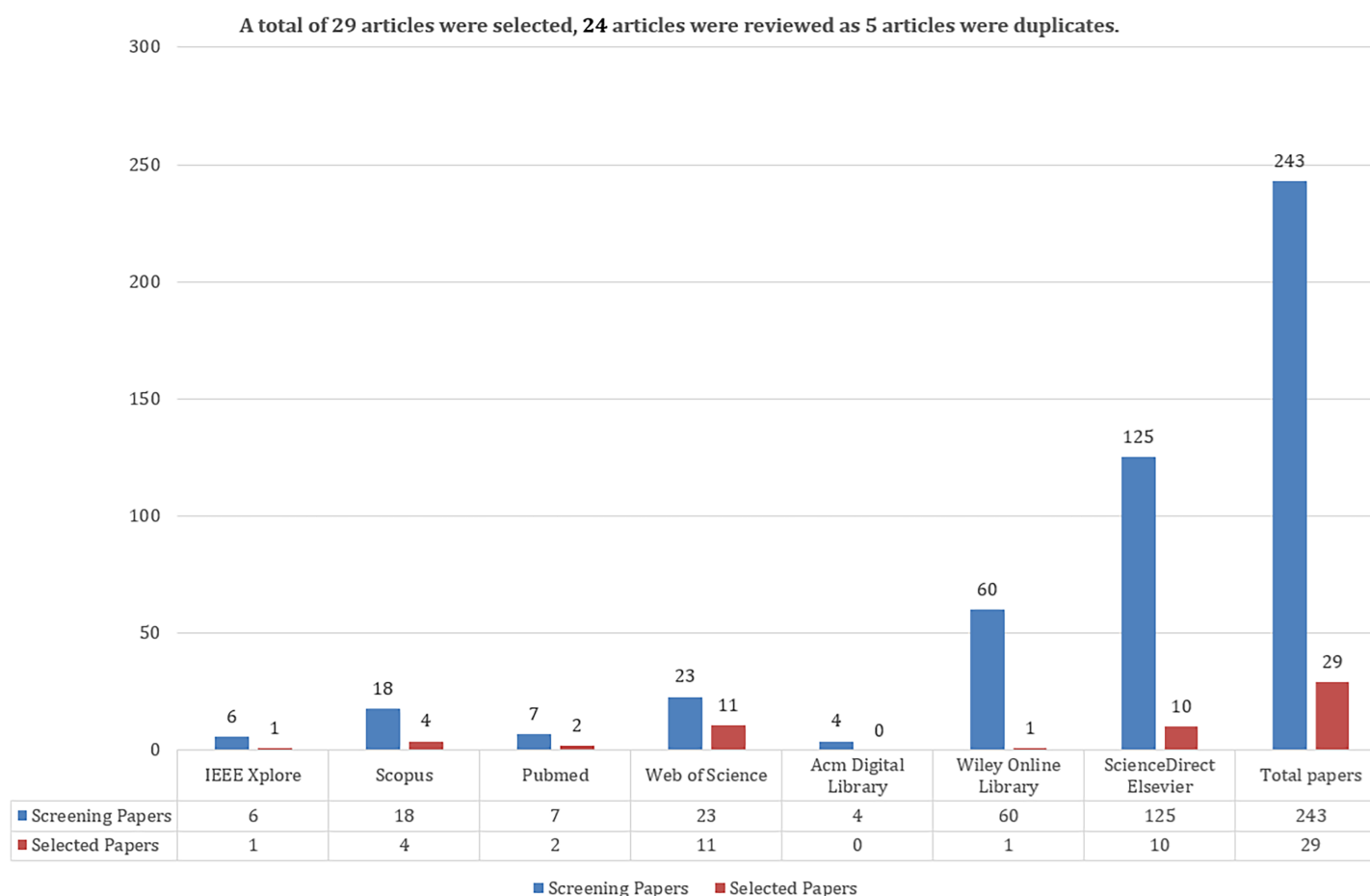


Figure 3 Number of studies selected and included in the databases as a result of query expressions. Following the initial search, inclusion and exclusion criteria were applied to assess the suitability of the 243 retrieved studies for this review. After screening, 29 studies from seven different databases met the criteria for further evaluation. Since some studies appeared in multiple databases, five duplicates were removed to avoid redundancy. Ultimately, 24 unique studies were included in the final review. These studies were evaluated based on their titles, abstracts, and full texts. [Full-size DOI: 10.7717/peerj-cs.3193/fig-3](https://doi.org/10.7717/peerj-cs.3193/fig-3)

read in full and assessed for relevance. A total of 29 articles were considered appropriate for inclusion in this review on disease detection utilizing AutoML. These selected articles are distributed across digital databases as follows: IEEE Xplore: 1, Scopus: 4, Web of Science: 11, PubMed: 2, ACM: 0, Wiley: 1, ScienceDirect (Elsevier): 10. [Figure 4](#) shows the number of articles according to digital databases. The PRISMA flow diagram of our article is shown in [Fig. 5](#).

FINDINGS AND RESULTS

In this section, a detailed summary of each of the 24 reviewed articles is provided, with an in-depth explanation of the proposed prediction models. This enables researchers to gain a thorough understanding of each study, facilitate comparisons, and pinpoint existing gaps in the literature. Then, the studies are analyzed separately within the framework of each research question. Detailed articles and dataset information are included in [Table S6](#) in the [Supplemental Files](#).

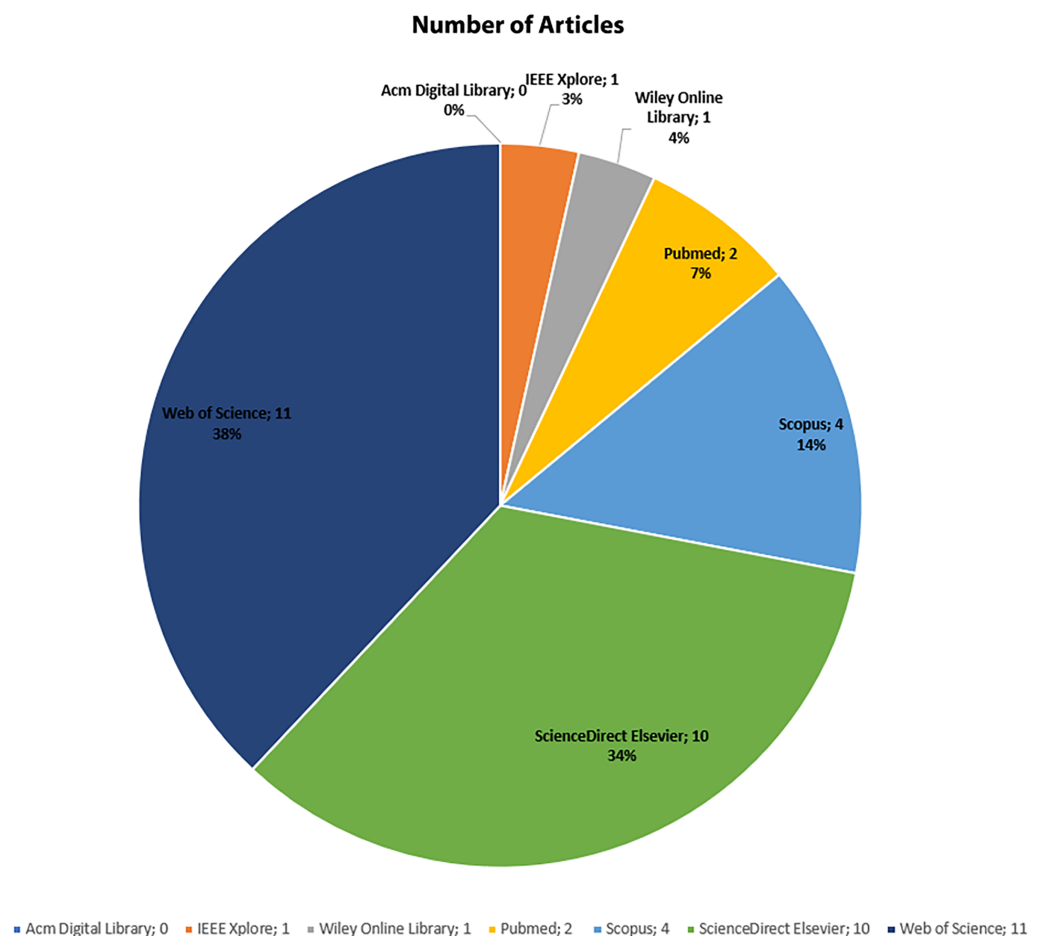


Figure 4 Number of articles according to digital databases.

Full-size DOI: [10.7717/peerj-cs.3193/fig-4](https://doi.org/10.7717/peerj-cs.3193/fig-4)

The annual distribution of publications illustrates a growing research interest in applying AutoML techniques to disease detection. Beginning with a limited number of articles in 2020, there is a marked increase in publication frequency, particularly in 2024 and 2025. This upward trend suggests a maturation of the field, likely influenced by both advancements in AutoML frameworks and increased demand for scalable AI-driven healthcare solutions in the post-pandemic era. The number of articles by year is shown in Fig. 6.

A categorization of studies by disease type indicates a predominant focus on cardiovascular diseases, diabetes, and COVID-19. These conditions are among the most widespread and clinically significant, which may explain their prioritization in AutoML applications. Nonetheless, the inclusion of other areas such as oncology, neurology, and gynecology reflects the adaptability of AutoML methods across diverse clinical domains. The current distribution also reveals opportunities for further research in underrepresented disease categories. Figure 7 illustrates the distribution of studies across various disease categories.

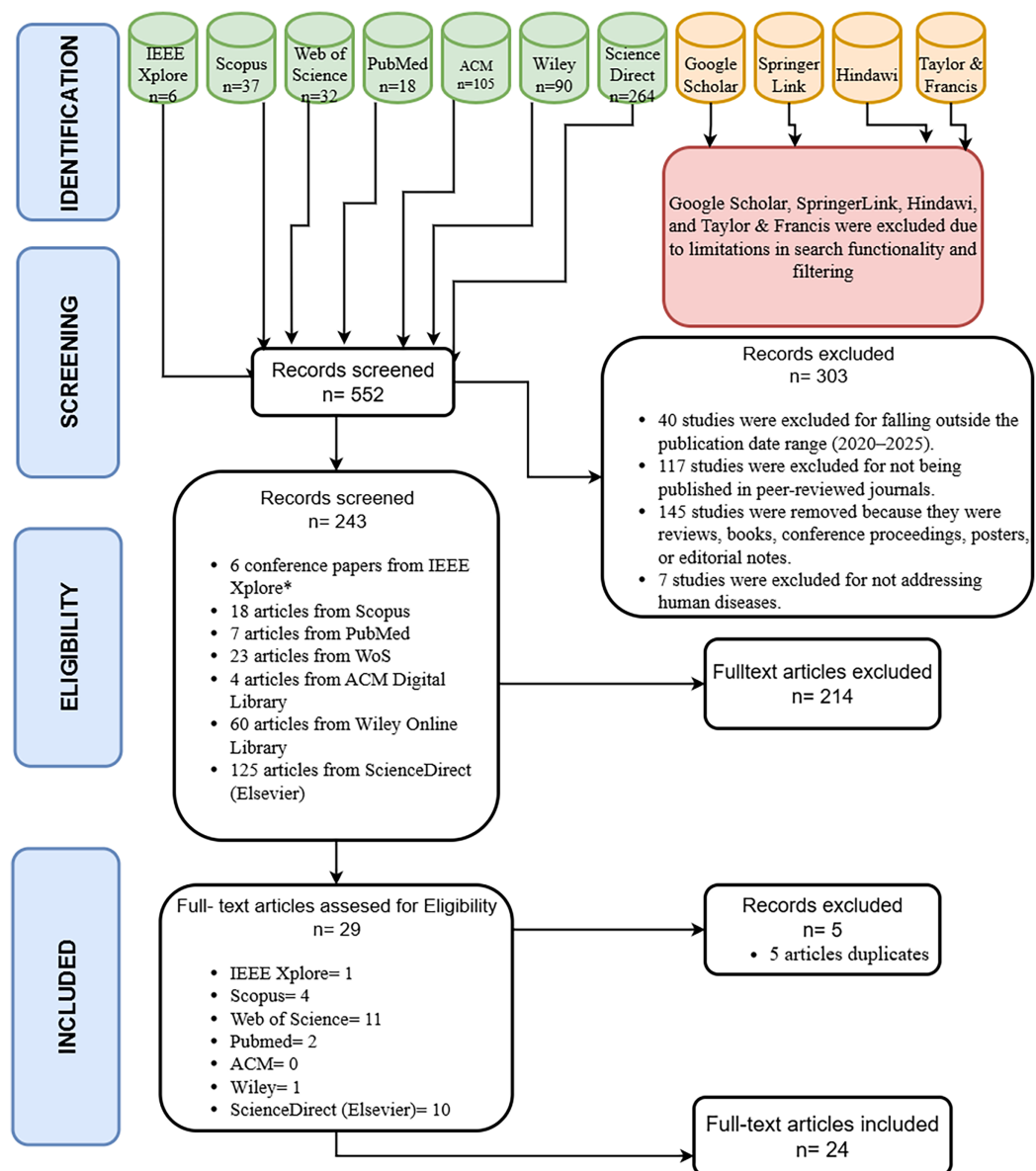


Figure 5 PRISMA flow chart of our article.

Full-size DOI: 10.7717/peerj-cs.3193/fig-5

Analysis of data modalities reveals that tabular datasets are most commonly utilized, likely due to their structured format and prevalence in clinical records and surveys. Image-based data also constitutes a significant portion, particularly in studies involving radiological and ophthalmological diagnostics. Although less frequently used, time-series, genomic, and multimodal datasets are beginning to emerge, signaling an expansion in the types of data being leveraged. This trend underscores the necessity for AutoML tools capable of accommodating a wide variety of biomedical data formats. The distribution according to data types is shown in [Fig. 8](#).

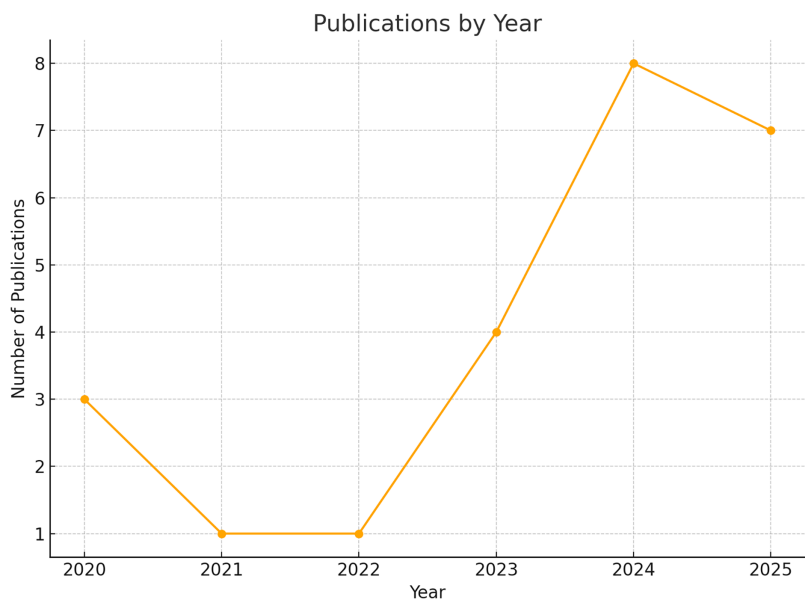


Figure 6 Publications by year. Presents the number of articles published each year.
 [Full-size](#) DOI: 10.7717/peerj-cs.3193/fig-6

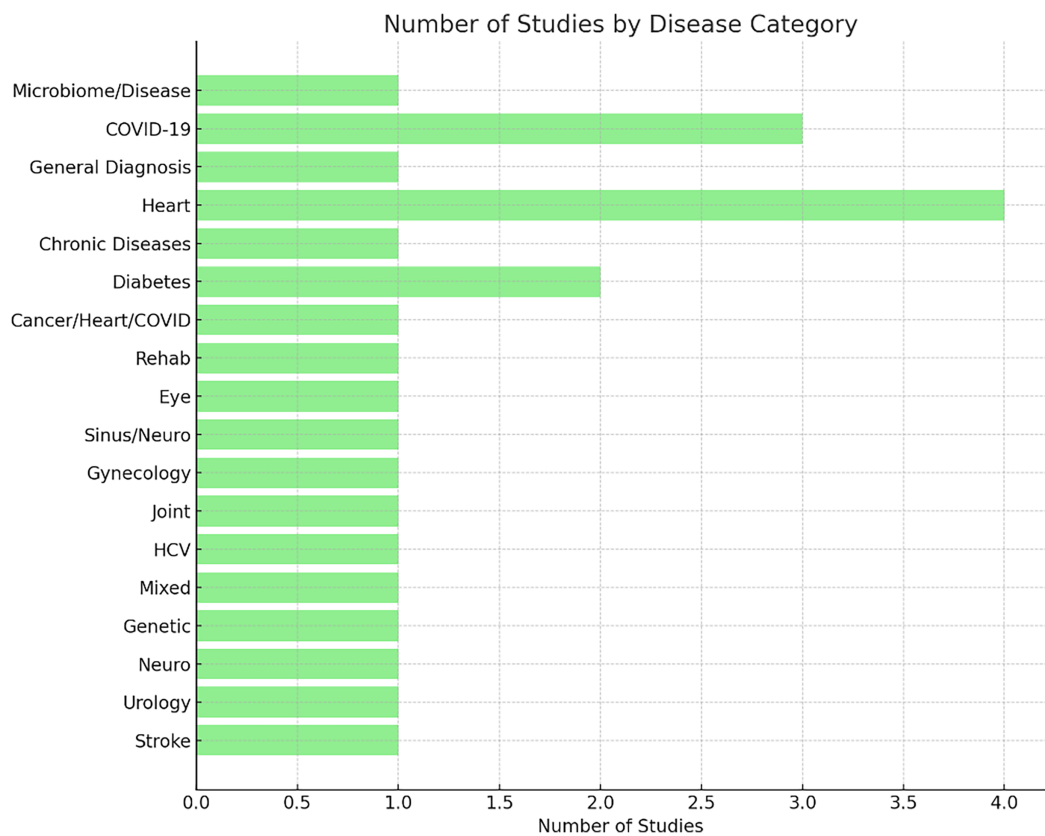


Figure 7 Research studies by disease classification.
 [Full-size](#) DOI: 10.7717/peerj-cs.3193/fig-7

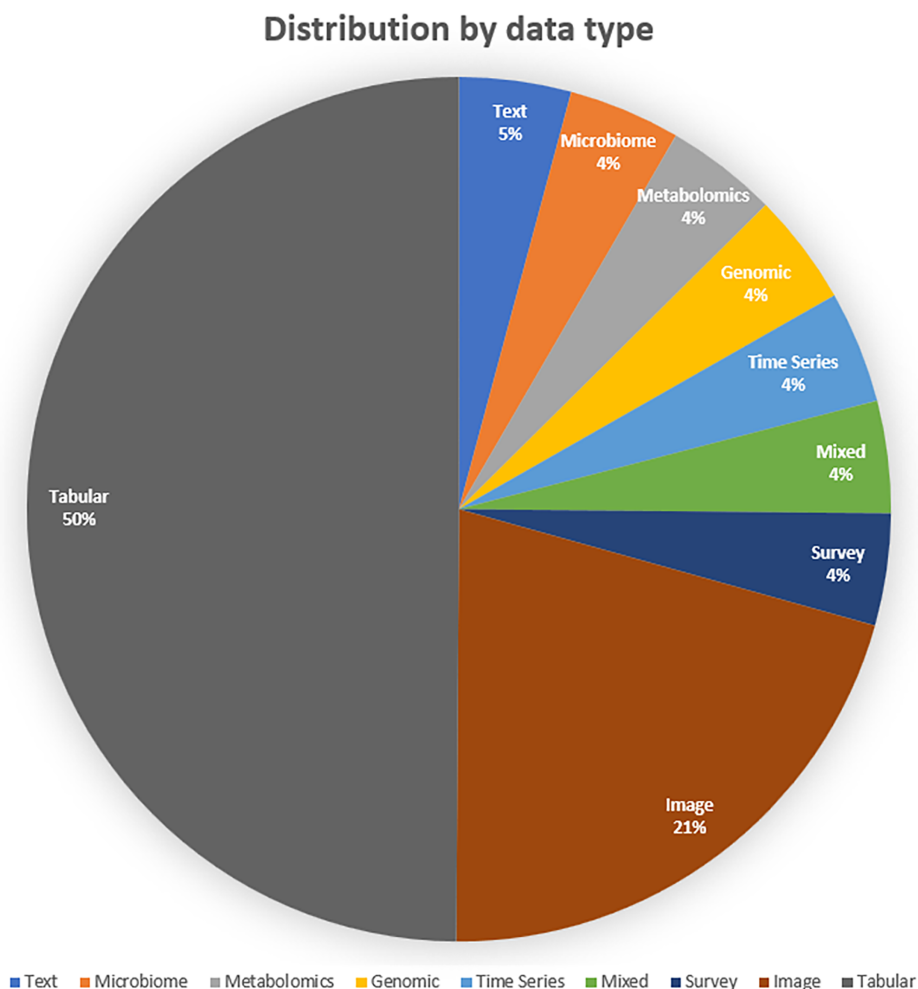


Figure 8 Distribution by data type. Presents the distribution based on data types.

Full-size DOI: [10.7717/peerj-cs.3193/fig-8](https://doi.org/10.7717/peerj-cs.3193/fig-8)

Summaries of articles

Model-Agnostic Meta-Learning (mAML) is an AutoML framework designed for the classification of human diseases utilizing microbiome data. It generates reproducible, automated, efficiently optimized, and interpretable models tailored to personalized microbiome-based classification problems. Implemented as a web-based platform, mAML features a scalable architecture. The system exhibits robust performance across 13 benchmark datasets, encompassing both binary and multi-class classification tasks. Furthermore, to support the application of mAML and enhance the microbiome learning landscape, the GMrepo ML repository has been developed. This resource comprises 120 microbiome-based classification tasks associated with 85 distinct human disease phenotypes, built upon a comprehensive dataset of 12,429 metagenomic and 38,643 amplicon samples (Yang & Zou, 2020).

COVID-19 remains a persistent global health challenge, associated with considerable morbidity and mortality. The disease often exhibits distinctive patterns on chest computed tomography (CT) scans, which can facilitate its early identification. Prompt and precise diagnosis is critical for the effective clinical management of affected individuals. This study aimed to assess the diagnostic capabilities of an AutoML algorithm for detecting COVID-19-related pneumonia based on chest CT imaging. The diagnostic performance was assessed using the area under the receiver operating characteristic curve (AUC), sensitivity, and positive predictive value. The method achieved an average precision of 0.932 (Sakagianni et al., 2020).

By investigating the use of ML in time series analysis, the study introduces the automatic time series (AutoTS) machine learning method. Time series models were created based on the ICD-10 dataset using Romanian hospitalization data between 2008 and 2018. Highly accurate predictions were generated for the ten most fatal diseases, and projections for the years 2019–2020 were developed at the NUTS 2 regional level. AutoTS automatically tries different models to select the best forecast model and accelerates time series analysis (Olsavszky et al., 2020).

Coronary artery disease prediction is one of the most complex and critical tasks encountered in the healthcare field. In this study, a predictive model is developed for the detection of heart disease. The performance of RF and XGBoost classifiers is enhanced through the application of three distinct hyperparameter optimization (HPO) techniques: grid search, random search, and genetic programming, implemented via the TPOT classifier. Model performances are evaluated using the CHD and Z-Alizadeh Sani dataset. The RF model optimized with TPOT provided the highest success by achieving 97.52% accuracy in the CHD dataset. In addition, the random forest model optimized with random search detected Least Absolute Deviation (LAD), Linear Cross Feature Model (LCX), and Randomized Clustering Algorithm (RCA) vessel stenoses in the Z-Alizadeh Sani dataset with 80.2%, 73.6% and 76.9% accuracy, respectively. The obtained results show that the proposed models provide higher accuracy compared to the studies in the existing literature. These findings reveal HPO methods play an important role in improving the performance of heart disease prediction systems (Valarmathi & Sheela, 2021).

This research evaluates the performance of prominent AutoML frameworks—Google AutoML, H2O.ai AutoML, Auto-Sklearn, and TPOT—for disease prediction tasks using medical claims data. In the analysis conducted using large-scale insurance claims data, techniques were applied to address missing data filling, categorical variable transformation, and imbalanced data problems. In the study where the models were evaluated with metrics such as accuracy, F1-score, sensitivity, and precision, it was determined that TPOT provided the highest accuracy on some data sets with genetic algorithms, Auto-Sklearn was successful in model selection and hyperparameter optimization, H2O.ai AutoML offered better scalability with large data sets, and Google AutoML Tables produced robust predictions with minimal user input. The results show that AutoML frameworks save time compared to manual model development and are an

effective tool in health data analytics; however, it is emphasized that the most appropriate framework should be selected according to different usage scenarios ([Romero et al., 2022](#)).

In this study ([Mallikarachchi et al., 2023](#)), AutoML and traditional machine learning approaches are compared for the prediction of T2D, CKD, and Ischemic Heart Disease (IHD). The results show that AutoML frameworks (TPOT, Auto-Sklearn, H2O) provide superior performance compared to traditional models. Auto-Sklearn provides the best accuracy (0.868) for T2D, while TPOT achieves high accuracies for CKD (0.99646) and IHD (0.7456). AutoML simplifies model development by requiring less manual work and expertise.

CloudAISim is a toolkit that aims to develop effective and explainable machine learning techniques in the healthcare field. It provides a prototype web application that provides data visualization and explainability by identifying accurate models for chronic and infectious diseases. It increases real-time efficiency with cloud computing support. Its architectural components include data entry, EDA, feature engineering, model building, hyperparameter tuning, and prediction explanation with LIME. 96–98% accuracy rate was achieved on breast cancer, heart disease, diabetes, and COVID-19 datasets ([Chowdhury et al., 2022](#)) using Auto-Keras. An interactive web application was developed with Streamlit to provide a use that does not require technical knowledge ([Bhowmik et al., 2023](#)).

In this study ([Paladino et al., 2023](#)), three distinct AutoML frameworks—AutoKeras, AutoGluon, and PyCaret—were evaluated using heart disease data across three datasets: the Cleveland dataset, the Hungarian dataset, and a combined dataset integrating both. In addition, 10 traditional ML models built with sklearn were evaluated as a reference. While the accuracy rate of these models remained between 55–60%, AutoML tools provided higher accuracies. The most successful tool was AutoGluon (78–86% accuracy), PyCaret’s success varied depending on the dataset (65–83%), and AutoKeras gave the most unstable results (54–83%). These findings show that AutoML is more effective than traditional methods in heart disease prediction.

This study explores the potential of hematochemical parameters for distinguishing between SARS-CoV-2 and influenza virus infections through the application of AutoML techniques. The dataset consists of 268 pediatric patients, including 133 diagnosed with SARS-CoV-2 and 135 with influenza. Ten distinct hematochemical features were employed to construct various machine learning models. Traditional algorithms, such as logistic regression, neural networks, k-nearest neighbors, support vector machines, and random forests, yielded classification accuracies ranging from 53.8% to 60.7%. In contrast, the AutoML approach significantly outperformed these methods, attaining an accuracy of 98.4%. These findings underscore the potential of AutoML as a powerful tool for the rapid and precise identification of SARS-CoV-2 and influenza infections in pediatric populations ([Dobrijević et al., 2023](#)).

It aims to develop an explainable and high-performance machine learning model using AutoGluon, an AutoML framework, for the prediction of coronary artery disease (CAD). Five different open data sets were combined, missing and outlier data were cleaned, and the modeling process with meaningful medical features was started. AutoGluon’s 4-fold

bagging (four bag-fold) and single-level stacking (one stack-level) structure was used in the modeling phase, thus combining the outputs of multiple basic models (random forest, Gradient Boosting Machine (GBM), artificial neural network (ANN), k-nearest neighbors (KNN), *etc.*) to form a strong ensemble model. HPO was performed automatically by AutoGluon, and the model achieved 91.67% accuracy and 0.9562 AUC score. In addition, the decision processes of the model were made explainable using SHapley Additive exPlanations (SHAP) analysis, thus presenting a transparent prediction system that increases reliability in clinical applications ([Wang et al., 2024a](#)).

An advanced hybrid strategy combining heuristic and stochastic methods is developed for the detection of abnormalities in clinical data encountered in patient rehabilitation processes. In the proposed approach, patients' routine exercise data were clustered using the optimal k-means clustering method and then abnormal data points were determined using an interquartile range (IQR) based stochastic method. In this way, reliable and effective data was provided to medical experts and the processing of high-dimensional and inconsistent clinical data was facilitated. In addition, an optimal regression model was developed using the AutoML paradigm and the effectiveness of the proposed strategy was evaluated with statistical error measures. Experimental results show that the model successfully predicts Borg relative prediction error (RPE) and Timed Up and Go (TUG) health indicators with 98.55% and 98.50% R² scores, respectively. These findings reveal that the developed hybrid strategy makes a significant contribution to abnormal data detection and reliable data analysis in patient rehabilitation ([Khan et al., 2024](#)).

This study presents an innovative approach based on ChatGPT and AutoML to facilitate diabetic retinopathy (DR) diagnosis. Using AutoML techniques in the analysis of retinal images, 92–95% accuracy, 88–93% sensitivity and 90–94% specificity rates were achieved. Thanks to ChatGPT integration, model outputs are interpreted with natural language processing and presented as user-friendly reports, facilitating the clinical use of the system. The developed method provides early diagnosis, especially in regions with limited medical resources, and is a more accessible and effective alternative compared to existing screening methods ([Mohammadi & Nguyen, 2024](#)).

This study compares the performance of Ensemble Learning and AutoML methods for heart disease prediction. A total of 18 different models (eight Ensemble, 10 AutoML) were tested using the heart disease dataset (303 samples, 14 features). Among the ensemble models, SVM and logistic regression provided 80% accuracy, while among the AutoML models, GLM showed the best performance with 88% accuracy. In the study, a deep learning based ANN model also achieved 89.6% accuracy ([Rimal et al., 2024](#)).

This research presents an AutoML-based analysis leveraging patient data collected during the initial phase of the COVID-19 pandemic in Romania, from January to September 2020. The dataset, derived from the DRG system, encompasses 825,698 COVID-19 cases and includes detailed sociodemographic characteristics, medical histories, hospitalization records, and geographic information. Data preprocessing was performed using the Polars library, where feature engineering included renaming variables, applying one-hot encoding, and introducing new Boolean indicators relevant to COVID-19. The AutoML models developed for this period achieved an F1-score of 96.44%

for predicting discharge outcomes and 75.45% for mortality prediction. Additionally, an accuracy of 98.84% was recorded in classifying cases as “urgent or acute.” The analysis of feature importance indicated that older age, specific hospitals, and oncology departments exhibited a weaker association with patient recovery. Conversely, higher mortality rates were correlated with abnormal laboratory results and pre-existing cardiovascular conditions. Additionally, patients admitted without referrals, as well as those from central and capital regions of Romania, demonstrated a higher likelihood of being classified as acute cases ([Simon et al., 2024](#)).

A total of 1,376 coronal T2-weighted MRI head images from the OASIS-3 dataset were labeled by expert radiologists as showing sinonasal disease (777 images) or not (599 images). The dataset was split into training, validation, and testing sets (80/10/10). A single-label image classification model was trained using Google Cloud Vertex AI with 10-fold cross-validation. The best model achieved 91.3% sensitivity, 92.8% specificity, and 92% accuracy, with 63 true positives, 64 true negatives, five false positives, and six false negatives. Average training time was 158.5 min. The final model is available upon request ([Cheong et al., 2024](#)).

This study presents a ML-based modeling process to predict cardiovascular health risks in patients with diabetes. Data cleaning, feature encoding, and scaling were performed on DSD and DRD datasets using the PyCaret platform. The LightGBM model showed high performance with AUC 0.8302 and precision 0.7320. The most influential features were GenHlth, hypertension, BMI, and age. LightGBM is strong in overall accuracy, but XGBoost outperforms it in F1 and Kappa metrics. This indicates that XGBoost offers a better balance between recall and precision is more reliable against random agreements. The model predicted “more than 30-day readmission” 84% correctly, but tends to underestimate short-term admissions. This indicates that a careful balance should be struck between false positives and negatives in clinical applications ([Jose et al., 2024](#)).

The few-shot learning (FSL) method is employed for the detection of Auto-Encoding Hyperparameters (AEH), Neural Architecture and Evolutionary Hyperparameter Optimization (NAEH), and evolutionary computation (EC) using a limited number of total variation uncertainty (TVU) images. TVU images from pathologically confirmed NAEH, AEH, and EC patients were split into support (SS) and query (QS) sets. Eigenvectors of size 1×64 were extracted using a dual pre-trained ResNet50 V2 model. Subsequently, Euclidean distances between each TVU image in the QS and the nine images in the SS were calculated, and diagnosis was performed using the KNN algorithm. The results indicate that the overall accuracy and macro precision of the proposed FSL model are 0.878 and 0.882, respectively. Furthermore, the model achieved the highest performance in EC recognition, with precision (0.964), recall (0.900), and F1-score (0.931). Additionally, various models, including the H2O AutoML (ensemble), the traditional ResNet50 V2 model, the FSL model combining the dual pre-trained ResNet50 V2 eigenvector extractor, and the junior and senior sonographer models, were employed in the study ([Wang et al., 2024b](#)).

The AutoML Models (Autoprognosis V.2.0) developed to estimate the rapid progress of the knee osteoarthritis (OA) were examined. Models with all features provide the highest

accuracy, while only simpler models trained only with clinical data have performed strongly with AUC-PRC 0.727 in multi-class estimates and AUC-PRC 0.764 in binary forecasts. Multi-class models have succeeded especially in early stage OA patients, while binary models have given more reliable results in individuals under 60 years of age. Patient-reported symptoms and MRI data are identified as the most influential attributes; web-based tools have also been developed with personalized predictions ([Castagno et al. 2025](#)).

Hepatitis C virus (HCV) is a major infection that causes chronic liver diseases worldwide. Early diagnosis and effective management are critical to prevent complications. In this study, class imbalances were corrected and additional features were added to the dataset obtained from the UCI ML Repository for the prediction of HCV. Then, modeling was performed using seven different AutoML tools and high accuracy rates between 99.29% and 100% were achieved. The obtained results show that AutoML-based models are effective tools that can assist physicians in the diagnosis of HCV ([Değer & Can, 2025](#)).

The study proposes a hybrid multi-objective feature selection system called MOGAHHO, combining genetic algorithm (GA) and Harris hawk optimization (HHO), to improve disease risk analysis and prediction. Using TPOT AutoML, it selects the most accurate ML model (e.g., logistic regression (LR), decision tree (DT), multilayer perceptron (MLP), KNN, and random forest (RF)) across ten datasets. MOGAHHO optimizes feature selection by balancing GA's global search with HHO's exploration-exploitation phases, aiming to maximize classification accuracy while minimizing feature count. Selected features are ranked using TOPSIS, highlighting key disease markers. The system outperforms methods like PCA, SVD, and autoencoders in classification accuracy and feature reduction across multiple medical datasets ([Kuanr & Mohapatra, 2025](#)).

TF-IDF-Kmer and DNA composition components were used to extract effective features from DNA sequences and these data were combined with physicochemical information. Feature selection, model determination and HPO processes were automated via the AutoGluon platform. The developed model was tested on three different datasets and achieved 97.14%, 79.71% and 98.73% accuracy rates, respectively, outperforming the existing leading models by 2%, 2.56% and 4%. In addition, the decision processes of the model were analyzed using interpretability tools such as SHAP and the explainability of the predictions was increased. Finally, a prediction application was developed to provide easy access to researchers ([Ye et al., 2025](#)).

A comprehensive machine learning and deep learning pipeline for brain age estimation is presented. T1-weighted MRI images were processed with FastSurfer, converted into a form suitable for deep learning, and phenotypes were extracted for machine learning models. The models developed using the UK Biobank, ADNI, and NACC datasets were evaluated both for age estimation in healthy individuals and as biomarkers for neurodegenerative diseases. With the designed statistical framework, the age estimation performance of the models, their consistency across different data sources and demographic groups, and their ability to distinguish neurodegenerative diseases were analyzed. The results show that especially the penalized linear models developed with

Zhang's methodology perform with high accuracy (<1 year MAE) and generalizability; the AUROC value is up to 0.90 in distinguishing diseases such as dementia (Capó *et al.*, 2025).

In order to support early and accurate diagnosis of BPH, a nanoparticle-assisted MALDI-MS based metabolomics platform was used to analyze metabolic profiles obtained from urine and serum samples. In the two-stage analysis process, healthy individuals were first separated from those with LUTS by UMP, and then BPH cases in the LUTS group were identified with high accuracy ($AUC = 0.830$) by serum metabolic patterns (SMP). The study also identified eight potential metabolic biomarkers that were distributed independently of age groups, emphasizing the clinical value of this approach for early diagnosis and personalized treatment of BPH. In addition, an experimental cohort dataset was specifically created within the scope of this study, including samples from BPH, LUTS and healthy individuals (Xu *et al.*, 2025).

A semantic ontology (StrokeOnto) based on TPOT AutoML and OWL is used to predict stroke deterioration and improve recommendations. Secure and autonomous data control is provided by adhering to the principle of digital sovereignty in patient data and compliance with local data privacy laws is supported. In addition, the classifications are explained with LIME to determine the importance of attributes. The proposed model aims to provide tailored interventions according to individual patient profiles. The model was validated using a publicly available stroke dataset, which was also employed in the development of the associated ontology. Among the supervised machine learning models tested in TPOT, the pipeline combining a decision tree classifier with a variance threshold achieved the highest performance, demonstrating an accuracy of 95.2%, surpassing other models (Chatterjee, 2025).

Summaries of the studies according to the research questions are shown in detail in Table S7 in the Supplemental Files. The table explains the preprocessing steps, denoising and cleaning, feature extraction and feature extraction, prediction models, performance metrics, and hpo methods.

Table S7 presents a comprehensive comparison of recent studies that integrate AutoML systems into machine learning (ML) pipelines, with specific attention to various preprocessing strategies, modeling techniques, performance metrics, and optimization procedures.

Inferences

1. Feature engineering techniques

Feature selection and extraction methods varied widely among the studies. Classical statistical approaches (e.g., ANOVA, chi-squared, PCA, and recursive feature elimination) were common in traditional ML workflows (Kuanr & Mohapatra, 2025; Mallikarachchi *et al.*, 2023), while more advanced AutoML systems employed automated or embedded techniques, including built-in feature importance assessments (Ye *et al.*, 2025) and SHAP explainability tools (Castagno *et al.*, 2025; Wang *et al.*, 2024a). A notable trend was the use of domain-specific methods, such as TF-IDF-Khmer in genomic data (Ye *et al.*, 2025) or grey-level texture features in imaging studies (Wang *et al.*, 2024b).

AUTOML MODELS WERE USED IN ARTICLES

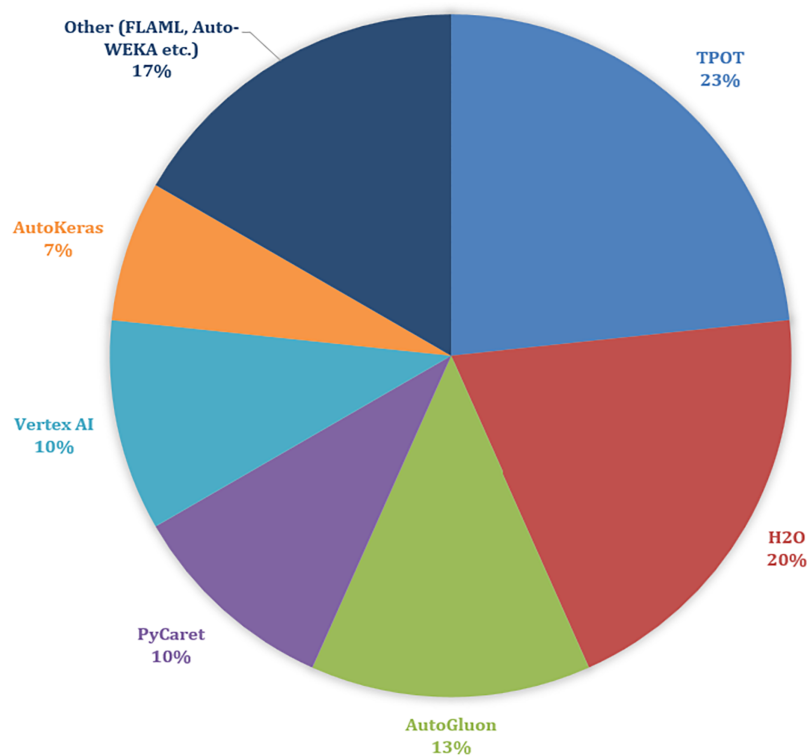


Figure 9 Number of times AutoML methods were used in articles.

Full-size DOI: 10.7717/peerj-cs.3193/fig-9

2. Noise removal and data cleaning

Noise removal or denoising, while essential, was not consistently reported. Only a few studies such as (Mohammadi & Nguyen, 2024) and (Xu et al., 2025) explicitly mentioned denoising techniques, including CLAHE and wavelet-based methods. Data normalization, transformation, and imputation were more commonly addressed (Rimal et al., 2024; Simon et al., 2024) reflecting standard preprocessing in automated pipelines.

3. Prediction models and autoML tools

The diversity of models employed spans traditional ML classifiers (e.g., SVM, RF, logistic regression), ensemble methods (e.g., XGBoost, Gradient Boosting), and AutoML frameworks like TPOT, H2O, AutoGluon, and PyCaret. AutoML platforms not only automated model selection but also facilitated meta-learning strategies (Wang et al., 2024a), with varying degrees of human intervention.

4. Evaluation metrics

A wide spectrum of performance metrics was employed, indicating the varied nature of tasks (classification, regression, etc.). Accuracy and F1-score were the most consistently reported metrics across studies. For classification tasks, recall, precision, specificity, and

sensitivity were frequently used, while regression-focused studies emphasized root mean square error (RMSE), mean absolute error (MAE), Coefficient of Determination (R^2), mean absolute percentage error (MAPE) (Capó et al., 2025; Khan et al., 2024). The use of area-under-curve metrics (AUC-ROC, AUC-PRC) was common, especially in medical and high-stakes prediction domains.

5. Hyperparameter optimization

Random search and grid search were still employed, particularly when paired with TPOT or other genetic programming frameworks (Chatterjee, 2025; Valarmathi & Sheela, 2021). However, Bayesian optimization (Castagno et al., 2025), ElasticNet regularization, and AutoML-integrated optimization were increasingly reported. The trend indicates a shift toward hands-off, scalable optimization embedded directly within AutoML systems.

Based on the analysis of article mentions, TPOT appears as the most frequently cited AutoML platform, with seven mentions, indicating its popularity and widespread use in the literature. It is closely followed by H2O, mentioned six times, which also suggests strong recognition within the research community. AutoGluon, PyCaret, and Vertex AI have moderate visibility, with four and three mentions, respectively, indicating growing interest. AutoKeras is mentioned less frequently (two times), while other tools such as FLAML, Auto-WEKA, and similar platforms collectively account for five mentions. This distribution highlights the diversity of AutoML tools being explored and suggests that while a few platforms dominate in popularity, there is ongoing interest in a variety of alternative solutions. The number of times AutoML methods and models are used in articles is shown in Fig. 9.

Table S8 in the Supplemental Files shows dataset/data type, disease/task, author (year), ML/AutoML method, best accuracy/performance.

The highest accuracies reported are generally above 90%, with some datasets reaching as high as 99–100% accuracy, such as hepatitis C and chronic kidney disease. AutoML frameworks like TPOT, AutoGluon, Auto-Sklearn, H2O, and AutoKeras have consistently demonstrated superior performance compared to traditional methods. The most frequently studied diseases include heart disease, COVID-19, diabetes, kidney disease, and various types of cancer. Additionally, several studies have enhanced accuracy through specialized hybrid approaches and hyperparameter optimization techniques, such as TPOT combined with hyperparameter optimization and genetic algorithms integrated with Harris hawks optimization (GA + HHO).

CONCLUSIONS AND RECOMMENDATIONS

This binary table provides a structured overview of 24 selected studies focusing on disease detection using AutoML methods. Each row represents a study, while each column corresponds to one of seven key research questions (RQ1–RQ7), covering topics such as AutoML usage, input features, feature engineering, denoising, model validation, performance metrics, and hyperparameter tuning.

A value of 1 indicates that the study addresses the corresponding research aspect, while 0 indicates that it does not. The table clearly shows that most studies incorporate AutoML

methods and performance evaluation metrics (RQ1 & RQ6). However, input feature selection, feature engineering, and data denoising (RQ2–RQ4) are less consistently applied across the literature.

This analysis helps identify current research trends and highlights areas that are underexplored, suggesting opportunities for future work to improve robustness and generalizability in disease prediction models. The binary contributions of the selected studies according to the research questions are shown in [Table S9](#) presented in the [Supplemental Files](#).

This review highlights several important trends in the application of AutoML for disease detection. First, there is a clear increase in the number of relevant publications over time, with a significant concentration of studies appearing in 2024 and 2025. This rise reflects both technological advancements in AutoML frameworks and a growing demand for scalable diagnostic solutions, particularly in the wake of the COVID-19 pandemic. Second, cardiovascular diseases, diabetes, and COVID-19 emerged as the most frequently studied conditions. These diseases are highly prevalent and often require timely diagnosis, making them prime candidates for machine learning applications. Nevertheless, the presence of research in less-explored areas such as neurology, ophthalmology, and gynecology demonstrate the flexibility of AutoML methods and reveals opportunities for further exploration in these domains. Lastly, the review shows that tabular data is the most commonly used data type, likely due to its widespread availability in structured clinical records. Image data also features prominently, especially in studies involving radiological imaging. Although less common, the use of genomic, time-series, and multimodal datasets is gradually increasing, pointing to a growing need for AutoML tools capable of handling complex and diverse biomedical data types. Strategies such as data subset selection, model compression, and low-precision computing make the use of deep learning sustainable, especially in healthcare institutions with limited resources ([Yuan, 2025](#)). Domain-specific pretraining provides more effective results compared to foundation models, especially in cold-start scenarios ([Yuan et al., 2025](#)). This suggests that integrated strategies with lighter models such as AutoML may be preferred, especially in the field of medical imaging. As a result of the survey;

- General evaluation: The survey has shown how effective AutoML techniques are in predicting various diseases in the field of health. Most of the models used for different diseases have reached high accuracy rates, especially the predictions made on diseases such as type 2 diabetes, heart diseases, COVID-19 are quite successful.
- AutoML and traditional methods comparison: It has been determined that AutoML methods generally provide faster and more accurate results compared to traditional machine learning methods. Especially when the data sets are large and complex, the automation provided by AutoML provides a great advantage.
- Clinical applications and data integrity: AutoML's ability to overcome difficulties such as missing data, class imbalances, and transformation of categorical data is also important when working on health data. This is a critical element for the processing of clinical data and the reliability of the results.

The recommendations of the survey can be listed as follows.

- Expansion of application areas: AutoML's application areas in the field of health can be further expanded. Especially when working with rare diseases or small data sets, the advantages of AutoML may be more pronounced. In this context, it is recommended to develop AutoML-based models for more diseases.
- Developing user-friendly tools: In order to increase the use of AutoML-based tools, user-friendly interfaces should be developed where healthcare professionals can get accurate results with less intervention in model development processes. This will save time and resources, especially in clinical settings.
- Explainability and transparency: Explainability of models is important for clinical applications. In order to provide transparent and reliable results, methods such as LIME and SHAP should be used to increase model explainability.
- Data security and ethics: Data security and patient privacy are of great importance in the application of AutoML systems in the healthcare field. Therefore, the integration of secure data management systems that comply with local data privacy laws is important.
- Future research: New algorithms and optimization techniques should be investigated to make AutoML methods more efficient. In addition, more validation studies can be conducted in clinical use to increase the accuracy of such systems in a wider patient group.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Ersin Enes Eryılmaz conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Erdal Kılıç analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

Raw data was not generated in this literature review.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.3193#supplemental-information>.

REFERENCES

- Bergstra J, Komer B, Eliasmith C, Yamins D, Cox DD. 2015.** Hyperopt: a python library for model selection and hyperparameter optimization. *Computational Science & Discovery* 8:014008 DOI 10.1088/1749-4699/8/1/014008.
- Bhowmik A, Sannigrahi M, Chowdhury D, Dey A, Gill SS. 2023.** CloudAISim: a toolkit for modelling and simulation of modern applications in AI-driven cloud computing environments. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations* 3(4):100150 DOI 10.1016/j.tbench.2024.100150.
- Capó M, Vitali S, Athanasiou G, Cusimano N, García D, Cruickshank G, Patel B, AsDN I. 2025.** UK biobank MRI data can power the development of generalizable brain clocks: a study of standard ML/DL methodologies and performance analysis on external databases. *NeuroImage* 308(3):121064 DOI 10.1016/j.neuroimage.2025.121064.
- Castagno S, Birch M, van der Schaar M, McCaskie A. 2025.** Predicting rapid progression in knee osteoarthritis: a novel and interpretable automated machine learning approach, with specific focus on young patients and early disease. *Annals of the Rheumatic Diseases* 84(1):124–135 DOI 10.1136/ard-2024-225872.
- Chatterjee A. 2025.** Sovereignty in automated stroke prediction and recommendation system with explanations and semantic reasoning. *Procedia Computer Science* 254(5):201–210 DOI 10.1016/j.procs.2025.02.079.
- Cheong RCT, Jawad S, Adams A, Campion T, Lim ZH, Papachristou N, Unadkat S, Randhawa P, Joseph J, Andrews P. 2024.** Enhancing paranasal sinus disease detection with AutoML: efficient AI development and evaluation via magnetic resonance imaging. *European Archives of Oto-Rhino-Laryngology* 281(4):2153–2158 DOI 10.1007/s00405-023-08424-9.
- Chowdhury D, Poddar S, Banarjee S, Pal R, Gani A, Ellis C, Arya RC, Gill SS, Uhlig S. 2022.** CovidXAI: explainable AI assisted web application for COVID-19 vaccine prioritization. *Internet Technology Letters* 5:e381 DOI 10.1002/itl2.381.
- Değer SU, Can H. 2025.** A predictive study on HCV using automated machine learning models. *Computers in Biology and Medicine* 188(2):109897 DOI 10.1016/j.combiomed.2025.109897.
- Dobrijević D, Antić J, Rakić G, Katanić J, Andrijević L, Pastor K. 2023.** Clinical hematochemical parameters in differential diagnosis between pediatric SARS-CoV-2 and influenza virus infection: an automated machine learning approach. *Children* 10(5):761 DOI 10.3390/children10050761.
- Hutter F, Kotthoff L, Vanschoren J. 2019.** *Automated machine learning: methods, systems, challenges*. Berlin, Germany: Springer Nature.
- Jose R, Syed F, Thomas A, Toma M. 2024.** Cardiovascular health management in diabetic patients with machine-learning-driven predictions and interventions. *Applied Sciences* 14(5):2132 DOI 10.3390/app14052132.
- Khan MA, Iqbal N, Jamil H, Qayyum F, Jang J-H, Khan S, Kim J-C, Kim D-H. 2024.** Enhanced abnormal data detection hybrid strategy based on heuristic and stochastic approaches for efficient patients rehabilitation. *Future Generation Computer Systems* 154(4):101–122 DOI 10.1016/j.future.2023.11.036.
- Kuanr M, Mohapatra P. 2025.** A recommender system with multi-objective hybrid Harris Hawk optimization for feature selection and disease diagnosis. *Healthcare Analytics* 7(6):100384 DOI 10.1016/j.health.2025.100384.
- Mallikarachchi D, Rathnayake D, Abegunawardana D, Van-Hoff S, Kasthurirathna D, Gamage A. 2023.** Automated machine learning for prediction of type 2 diabetes and its major

- complications: a comparative study. In: *2023 5th International Conference on Advancements in Computing (ICAC)*. Piscataway: IEEE, 466–471.
- Mishra A, Pandey A, Malhotra R. 2024.** Deep learning in neglected vector-borne diseases: a systematic review. *International Journal of System Assurance Engineering and Management* **16**(5):1–13 DOI [10.1007/s13198-024-02380-1](https://doi.org/10.1007/s13198-024-02380-1).
- Mohammadi SS, Nguyen QD. 2024.** A user-friendly approach for the diagnosis of diabetic retinopathy using ChatGPT and automated machine learning. *Ophthalmology Science* **4**(4):100495 DOI [10.1016/j.xops.2024.100495](https://doi.org/10.1016/j.xops.2024.100495).
- Moher D, Liberati A, Tetzlaff J, Altman DG, the PRISMA Group*t. 2009.** Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine* **151**(4):264–269 DOI [10.7326/0003-4819-151-4-200908180-00135](https://doi.org/10.7326/0003-4819-151-4-200908180-00135).
- Mumuni A, Mumuni F. 2022.** Data augmentation: a comprehensive survey of modern approaches. *Array* **16**(6):100258 DOI [10.1016/j.array.2022.100258](https://doi.org/10.1016/j.array.2022.100258).
- Mumuni A, Mumuni F. 2024.** Automated data processing and feature engineering for deep learning and big data applications: a survey. *Journal of Information and Intelligence* **3**(2):113–153 DOI [10.1016/j.jiixd.2024.01.002](https://doi.org/10.1016/j.jiixd.2024.01.002).
- Mustafa A, Rahimi Azghadi M. 2021.** Automated machine learning for healthcare and clinical notes analysis. *Computers* **10**(2):24 DOI [10.3390/computers10020024](https://doi.org/10.3390/computers10020024).
- Olsavszky V, Dosius M, Vladescu C, Benecke J. 2020.** Time series analysis and forecasting with automated machine learning on a national ICD-10 database. *International Journal of Environmental Research and Public Health* **17**(14):4979 DOI [10.3390/ijerph17144979](https://doi.org/10.3390/ijerph17144979).
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE. 2021.** The PRISMA, 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* **372**:n71 DOI [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71).
- Paladino LM, Hughes A, Perera A, Topsakal O, Akinci TC. 2023.** Evaluating the performance of automated machine learning (AutoML) tools for heart disease diagnosis and prediction. *AI* **4**(4):1036–1058 DOI [10.3390/ai4040053](https://doi.org/10.3390/ai4040053).
- Papoutsoglou G, Tarazona S, Lopes MB, Klammersteiner T, Ibrahimi E, Eckenberger J, Novielli P, Tonda A, Simeon A, Shigdel R. 2023.** Machine learning approaches in microbiome research: challenges and best practices. *Frontiers in Microbiology* **14**:1261889 DOI [10.3389/fmicb.2023.1261889](https://doi.org/10.3389/fmicb.2023.1261889).
- Preity BAK, Shahnawazuddin S. 2024.** Automated computationally intelligent methods for ocular vessel segmentation and disease detection: a review. *Archives of Computational Methods in Engineering* **31**(2):701–724 DOI [10.1007/s11831-023-09998-7](https://doi.org/10.1007/s11831-023-09998-7).
- Rimal Y, Paudel S, Sharma N, Alsadoon A. 2024.** Machine learning model matters its accuracy: a comparative study of ensemble learning and automl using heart disease prediction. *Multimedia Tools and Applications* **83**(12):35025–35042 DOI [10.1007/s11042-023-16380-z](https://doi.org/10.1007/s11042-023-16380-z).
- Romero ARA, Deypalan YMN, Mehrotra S, Jungao JT, Sheils NE, Manduchi E, Moore JH. 2022.** Benchmarking AutoML frameworks for disease prediction using medical claims. *BioData Mining* **15**:15 DOI [10.1186/s13040-022-00300-2](https://doi.org/10.1186/s13040-022-00300-2).
- Sakagianni A, Feretzakis G, Kalles D, Koufopoulou C, Kaldis V. 2020.** Setting up an easy-to-use machine learning pipeline for medical decision support: a case study for COVID-19 diagnosis based on deep learning with CT scans. In: *The Importance of Health Informatics in Public Health during a Pandemic*. Amsterdam: IOS Press, 13–16.
- Salehin I, Islam MS, Saha P, Noman S, Tuni A, Hasan MM, Baten MA. 2024.** AutoML: a systematic review on automated machine learning with neural architecture search. *Journal of Information and Intelligence* **2**(1):52–81 DOI [10.1016/j.jiixd.2023.10.002](https://doi.org/10.1016/j.jiixd.2023.10.002).

- Simon SC, Bibi I, Schaffert D, Benecke J, Martin N, Leipe J, Vladescu C, Olsavszky V. 2024. AutoML-driven insights into patient outcomes and emergency care during Romania's first wave of COVID-19. *Bioengineering* 11(12):1272 DOI 10.3390/bioengineering11121272.
- Thirunavukarasu AJ, Elangovan K, Gutierrez L, Hassan R, Li Y, Tan TF, Cheng H, Teo ZL, Lim G, Ting DSW. 2023. Clinical performance of automated machine learning: a systematic review. *medRxiv* 2(5):266 DOI 10.1101/2023.10.26.23297599.
- Valarmathi R, Sheela T. 2021. Heart disease prediction using hyper parameter optimization (HPO) tuning. *Biomedical Signal Processing and Control* 70(11):103033 DOI 10.1016/j.bspc.2021.103033.
- Wang M, Liu W, Gu X, Cui F, Ding J, Zhu Y, Bian J, Chen Y, Zhou J. 2024b. Few-shot learning to identify atypical endometrial hyperplasia and endometrial cancer based on transvaginal ultrasonic images. *Heliyon* 10(16):e36426 DOI 10.1016/j.heliyon.2024.e36426.
- Wang J, Xue Q, Zhang CW, Wong KKL, Liu Z. 2024a. Explainable coronary artery disease prediction model based on AutoGluon from AutoML framework. *Frontiers in Cardiovascular Medicine* 11:1360548 DOI 10.3389/fcvm.2024.1360548.
- Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MMG, Sterne JAC, Bossuyt PMM, the QUADAS-2 Group*Q. 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 155(8):529–536 DOI 10.7326/0003-4819-155-8-201110180-00009.
- Wong CYT, O'Byrne C, Taribagil P, Liu T, Antaki F, Keane PA. 2024. Comparing code-free and bespoke deep learning approaches in ophthalmology. *Graefes Archive for Clinical and Experimental Ophthalmology* 262(9):2785–2798 DOI 10.1007/s00417-024-06432-x.
- Xu X, Tan H, Zhang W, Liu W, Chen Y, Zhang J, Gu M, Yang Y, Chen Q, Wang Y. 2025. Decoding benign prostatic hyperplasia: insights from multi-fluid metabolomic analysis. *Small Methods* 9:2401906 DOI 10.1002/smt.202401906.
- Yang F, Zou Q. 2020. mAML: an automated machine learning pipeline with a microbiome repository for human disease classification. *Database* 18:baaa050 DOI 10.1101/2020.02.11.943316.
- Ye D-X, Yu J-W, Li R, Hao Y-D, Wang T-Y, Yang H, Ding H. 2025. The prediction of recombination hotspot based on automated machine learning. *Journal of Molecular Biology* 437(6):168653 DOI 10.1016/j.jmb.2024.168653.
- Yuan H. 2025. Overcoming computational resource limitations in deep learning for healthcare: strategies targeting data, model, and computing. *Medicine Advances* 3(1):42–45.
- Yuan H, Yu K, Xie F, Liu M, Sun S. 2024. Automated machine learning with interpretation: a systematic review of methodologies and applications in healthcare. *Medicine Advances* 2(3):205–237 DOI 10.1002/med4.75.
- Yuan H, Zhu M, Yang R, Liu H, Li I, Hong C. 2025. Rethinking domain-specific pretraining by supervised or self-supervised learning for chest radiograph classification: a comparative study against ImageNet counterparts in cold-start active learning. *Health Care Science* 4(2):110–143.