APPENDIX

This supplementary file includes all tables, Tables S1-S9. You can also see explanations of some tables after the table titles.

Table S1 Abbreviations and expansions

Abbreviation	Expansion
ADNI	Alzheimer's Disease Neuroimaging Initiative
AE	Autoencoders
AEH	Atypical Endometrial Hyperplasia
AutoML	Automated Machine Learning
AWS	Amazon Web Services
BA	Balanced Accuracy
BM	Bookmaker Informedness
BPH	Benign Prostatic Hyperplasia
BRFSS	Behavioral Risk Factor Surveillance System
CHD	Cleveland Heart Disease Dataset
CKD	Chronic Kidney Disease
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Networks
CT	Computed Tomography
DRD	Diabetes Readmission Dataset
DRG	Diagnosis-Related Group
DSD	Diabetes Surveillance Dataset
EC	Endometrial Cancer
EDA	Exploratory Data Analysis
FN	False Negative
FP	False Positive
FSL	Few-Shot Learning
GA	Genetic Algorithm
HCV	Hepatitis C Virus
HHO	Harris Hawk Optimization
HPO	Hyperparameter Optimization
IBD	Inflammatory Bowel Disease
ICD-10	International Classification of Diseases, 10th Revision
LIME	Local Interpretable Model-Agnostic Explanations
LUTS	Lower Urinary Tract Symptoms
mAML	Microbiome-based Automated Machine Learning
MALDI-MS	Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry
MCC	Matthews Correlation Coefficient
MK	Markedness
ML	Machine Learning
MLP	Multi-Layer Perceptron
MOGAHHO	Multi-Objective Genetic Algorithm and Harris Hawk Optimization
MRI	Magnetic Resonance Imaging
MS	Mass Spectrometry
NACC	National Alzheimer's Coordinating Center

NAEH	Non-Atypical Endometrial Hyperplasia
NUTS	Nomenclature of Territorial Units for Statistics
OAI	Osteoarthritis Initiative
OWL	Web Ontology Language
PCA	Principal Component Analysis
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RA	Rheumatoid Arthritis
ROI	Region of Interest
SHAP	SHapley Additive exPlanations
SVD	Singular Vector Decomposition
TN	True Negative
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
TP	True Positive
TPOT	Tree-based Pipeline Optimization Tool
TSE	Turbo Spin Echo
TVU	Transvaginal Ultrasound
T2D	Type 2 Diabetes
UCI	University of California, Irvine (ML Repository)
XAI	eXplainable Artificial Intelligence

Table S2 Popular AutoML platforms and tools

AutoML platform	Access to data	Supported data types
Amazon SageMaker Autopilot	Registered, chargeable	Tabular, Time series, Image, Text
Apple Create ML	Free	Image, Text, Tabular
ATM	Free	Tabular, Image, Text
AutoGluon (Amazon)	Registered, chargeable	Image, Text, Tabular data
Auto-Keras	Free	Text, Image, Tabular
AutoPrognosis	Free	Tabular
Auto-Sklearn	Free	Tabular
Auto-WEKA	Free	Tabular
BigML OptiML	Free	Tabular, Image, Text, Time series
Darwin	Registered, chargeable	Tabular
DataBricks	Registered, chargeable	Tabular, Text, Time series
DataRobot	Registered, chargeable	Time series, Image, Text
FEDOT	Free	Tabular, Time series, Image, Text
FLAML (Microsoft)	Registered, chargeable	Tabular, Text, Time series
Google Cloud AutoML	Registered, chargeable	Tabular, Image, Text and video data
Google Vertex AI	Registered, chargeable	Tabular, Image, Text and video data
H2O AutoML	Registered, chargeable	Tabular, Text
H2O Driverless AI	Registered, chargeable	Time series, Tabular, Image, Text
JADBIO	Registered, chargeable	Tabular
IBM Watson AutoAI	Registered, chargeable	Tabular, Text, Time series
Ludwig	Free	Tabular, Text, Time series
MLBox	Free	Tabular
MS Azure AutoML	Registered, chargeable	Time series, Image, Text

NNI (Microsoft)	Free	Tabular, Image
PyCaret	Free	Tabular, Text, Time series
TPOT	Free	Tabular, Text, Image

Table S3 Evaluation of review

The review investigates literature on disease prediction with AutoML by addressing ten key aspects: (Q1) whether a structured search and selection process was applied across databases; (Q2) if individual studies and their predictive models were summarized and contextualized; (Q3) whether input features in models were extensively analyzed; (Q4) if feature selection or extraction methods were reported; (Q5) whether techniques to reduce noise in disease data were emphasized; (Q6) if training and testing methodologies were thoroughly examined; (Q7) whether specific AutoML techniques were identified; (Q8) if performance metrics were clearly defined; (Q9) whether there was a broad focus on AutoML research for general human diseases; and (Q10) whether searches covered multiple major digital databases.

Name of review	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Automated computationally intelligent methods for ocular vessel segmentation and disease detection: A review (Preity et al. 2024)	✗	✓	✓	✓	✓	✓	✗	✓	✗	✗
Automated machine learning with interpretation: A systematic review of methodologies and applications in healthcare (Yuan et al. 2024)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗
Deep learning in neglected vector-borne diseases: a systematic Review (Mishra et al. 2024)	✓	✓	✗	✗	✗	✓	✗	✓	✗	✗
Comparing code-free and bespoke deep learning approaches in ophthalmology (Wong et al. 2024)	✓	✓	✓	✗	✗	✓	✓	✓	✗	✗
Machine learning approaches in microbiome research: challenges and best practices (Papoutsoglou et al. 2023)	✗	✓	✓	✓	✗	✓	✗	✓	✓	✗
Clinical performance of automated machine learning: A systematic review (Thirunavukarasu et al. 2023)	✓	✗	✗	✓	✗	✓	✓	✓	✓	✗
Automated machine learning for healthcare and clinical notes analysis (Mustafa & Rahimi Azghadi 2021)	✗	✓	✗	✓	✗	✓	✓	✓	✓	✗
A literature review on disease detection with automated machine learning (Our Papers)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table S4 Query statements

The query statements used in this process, along with the corresponding source details. It was observed that effective filltering could not be achieved in Google Scholar, SpringerLink, and Taylor & Francis due to the general nature of their search functionalities, which do not support advanced search phrases or complex query statements. Additionally, since Hindawi journals have been integrated into Wiley's open access journal portfolio, searches conducted through Google Scholar, SpringerLink, Hindawi, and Taylor & Francis were excluded from the analysis. Consequently, as shown in Table S4, the final searches were carried out using seven selected digital repositories.

Databases	Query statements
IEEE Xplore	("All Metadata":"disease detection" OR "All Metadata":"disease diagnosis" OR "All Metadata":"disease prediction" OR "All Metadata":"human healthcare") AND ("All Metadata":"automated machine learning" OR "All Metadata":"AutoML")
Scopus	TITLE-ABS-KEY ("disease detection" OR "disease diagnosis" OR "human healthcare") AND ("automl" OR "automated machine learning")
Web of Science	ALL= (("disease detection" OR "disease diagnosis" OR "disease prediction" OR "human healthcare") AND ("AutoML" OR "automated machine learning"))
PubMed	(("disease detection" OR "disease diagnosis" OR "human healthcare") AND ("AutoML" OR "automated machine learning"))
Acm Digital Library	[[All: "disease detection"] OR [All: "disease diagnosis"] OR [All: "disease prediction"] OR [All: "human healthcare"]] AND [[All: "automated machine learning"] OR [All: "automl"]]
Wiley Online Library	""disease detection" OR "disease diagnosis" OR "disease prediction" OR "human healthcare"" anywhere and ""automated machine learning" OR "automl"" anywhere
ScienceDirect Elsevier	("disease detection" OR "disease diagnosis" OR "disease prediction" OR "human healthcare") AND ("AutoML" OR "automated machine learning")

Table S5 Digital databases filtering results

As a result of the screening process conducted across seven different digital databases from January 2020 to March 2025, studies were filtered based on predefined inclusion and exclusion criteria. Initially, 552 studies were identified. Of these, 40 studies published outside the 2020-2025 time frame were excluded. Subsequently, 117 studies not published in peer reviewed journals were removed. An additional 145 studies were excluded for being reviews, books, conference proceedings, posters, editorial notes, or similar publication types. Finally, 7 studies were excluded for not focusing on human diseases. After applying all elimination criteria, 243 studies were retained for full-text review.

	Searching	Excluding					Screening
Digital Library	**Identification**	**2020-2025 or not?**	**Journal or not?**	**Article or not?**	**Language = English**	**Humans or not?**	**Papers**
IEEE Xplore	6	0	6	No searching filter	No searching filter	No searching filter	**6***
Scopus	37	1	17	1	0	No searching filter	**18**
Pubmed	18	4	No searching filter	Lots of searching filter	0	7	**7**
Web of Science (WoS)	32	1	No searching filter	8	0	No searching filter	**23**
ACM Digital Library	105	6	90	5	No searching filter	No searching filter	**4**
Wiley Online Library	90	14	16	No searching filter	No searching filter	No searching filter	**60**
ScienceDirect-Elsevier	264	14	No searching filter	125	No searching filter	No searching filter	**125**
*IEEE Conferences included		Number of articles that underwent full text, keyword and title review					**243**

Table S6 Papers and Dataset

A comprehensive summary of all 24 reviewed papers is presented, including detailed explanations of the proposed prediction models. This thorough overview allows researchers to understand each study in depth, draw comparisons across different approaches, and identify gaps in the existing literature. Following the summaries, the studies are individually analyzed in relation to each research question. Detailed information about the papers and datasets used is provided in Table S6.

#	Papers	Dataset
1	mAML: an automated machine learning pipeline with a microbiome repository for human disease classification (Yang & Zou 2020)	GMrepo Microbiome Learning repository
2	Setting up an Easy-to-Use Machine Learning Pipeline for Medical Decision Support: A Case Study for COVID-19 Diagnosis Based on Deep	Collected from COVID-19-related publications on medRxiv and bioRxiv

	Learning with CT Scans (Sakagianni et al. 2020)	
3	Time Series Analysis and Forecasting with Automated Machine Learning on a National ICD-10 Database (Olsavszky et al. 2020)	National ICD-10 Database
4	Heart disease prediction using hyper parameter optimization (HPO) tuning (Valarmathi & Sheela 2021)	CHD and Z-Alizadeh Sani dataset
5	Benchmarking AutoML frameworks for disease prediction using medical claims (A. Romero et al. 2022)	lung cancer, prostate cancer, RA, T2D, IBD, and CKD Medicare or Commercial plans
6	Automated Machine Learning for Prediction of Type 2 Diabetes and Its Major Complications: A Comparative Study (Mallikarachchi et al. 2023)	2015 US BRFSS dataset
7	CloudAISim: A toolkit for modelling and simulation of modern applications in AI-driven cloud computing environments (Bhowmik et al. 2023)	UCI Breast Cancer Wisconsin (Diagnostic), UCI Heart Disease Cleveland dataset, National Institute of Diabetes and Digestive and Kidney Diseases, and “Covid-19” (Chowdhury et al. 2022) dataset
8	Evaluating the performance of automated machine learning (AutoML) tools for heart disease diagnosis and prediction (Paladino et al. 2023)	Cleveland Heart Disease and UCI Hungarian Heart disease datasets
9	Clinical Hematochemical Parameters in Differential Diagnosis between Pediatric SARS-CoV-2 and Influenza Virus Infection: An Automated Machine Learning Approach (Dobrijević et al. 2023)	A total of 268 pediatric patients (133 diagnosed with SARS-CoV-2 and 135 with Influenza) were included in the study conducted by the Institute for Child and Youth Health Care of Vojvodina
10	Explainable coronary artery disease prediction model based on AutoGluon from AutoML framework (Wang et al. 2024a)	five distinct data sets: Cleveland (303 observations), Hungary (294 observations), Switzerland (123 observations), VA Long Beach (200 observations), and Statlog (270 observations)
11	Enhanced abnormal data detection hybrid strategy based on heuristic and stochastic approaches for efficient patients rehabilitation (Khan et al. 2024)	A clinical dataset containing exercise data (including Borg RPE and TUG indicators) from real patient clinical data is acquired from JNU, Republic of Korea
12	A User-friendly Approach for the Diagnosis of Diabetic Retinopathy Using ChatGPT and Automated Machine Learning (Mohammadi & Nguyen 2024)	Methods to Evaluate Segmentation and Indexing Techniques in the field of Retinal Ophthalmology (Messidor-2), an open-source dataset
13	Machine learning model matters its accuracy: a comparative study of ensemble learning and AutoML using heart disease prediction (Rimal et al. 2024)	UCI Heart Disease Dataset (303 samples, 14 features)
14	AutoML-Driven Insights into Patient Outcomes and Emergency Care During Romania's First Wave of COVID-19 (Simon et al. 2024)	DRG database of COVID-19 patients in Romania (825,698 cases of disease)

15	Enhancing paranasal sinus disease detection with AutoML: efficient AI development and evaluation via magnetic resonance imaging (Cheong et al. 2024)	1313 unique non-TSE T2w MRI head sessions collected by Washington University in St Louis (WUSTL) Knight Alzheimer's Disease Research Center (ADRC) from the OASIS-3 repository
16	Cardiovascular health management in diabetic patients with machine-learning-driven predictions and interventions (Jose et al. 2024)	DSD, which is a clean and organized version of the 2015 BRFSS dataset and DRD, a UCI dataset encompassing a decade (1999–2008) of clinical care data from 130 hospitals across the United States.
17	Few-shot learning to identify atypical endometrial hyperplasia and endometrial cancer based on transvaginal ultrasonic images (Wang et al. 2024b)	Dataset 1 TVU images representing NAEH, AEH, and EC, with 100 samples per class, acquired from the First Affiliated Hospital of Soochow University. Dataset 2 consisted of TVU images of NAEH, AEH, and EC, each with 33 instances per category, obtained from the Suzhou Hospital of Traditional Chinese Medicine. Dataset 3 included TVU images of normal uterine conditions and uterine myoma, with 300 images per class, sourced from the First Affiliated Hospital of Soochow University.
18	Predicting rapid progression in knee osteoarthritis: a novel and interpretable automated machine learning approach, with specific focus on young patients and early disease (Castagno et al. 2025)	Osteoarthritis Initiative (OAI)
19	A predictive study on HCV using automated machine learning models (Değer & Can 2025)	HCV dataset from UCI Machine Learning Repository
20	A recommender system with multi-objective hybrid Harris Hawk optimization for feature selection and disease diagnosis (Kuanr & Mohapatra 2025)	Diabetic, ILPD, Dermatology, Cervical Cancer, Risk Classification, Breast Cancer W, Cardiotocography, Hepatitis, Heart, Lymphography, COVID-19
21	The Prediction of Recombination Hotspot Based on Automated Machine Learning (Ye et al. 2025)	The S1 dataset comprises 490 recombination hotspots and 591 recombination coldspots. The S2 dataset consists of 3,480 sequence samples representing recombination hotspots and 3,471 sequence samples corresponding to recombination coldspots. The S3 dataset, generated by Liu in 2021, closely resembles the S2 dataset and also originates from budding yeast.
22	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 (Capó et al. 2025)	ADNI, and NACC from UK Biobank
23	Decoding Benign Prostatic Hyperplasia: Insights from Multi-Fluid Metabolomic Analysis (Xu et al. 2025)	An experimental cohort dataset specifically created for the study, including samples from BPH, LUTS and healthy individuals.
24	Sovereignty in Automated Stroke Prediction and Recommendation System with Explanations and Semantic Reasoning (Chatterjee 2025)	Public stroke dataset: Sample Count 5,110 records (4,861: No stroke [Class 0], 249: Stroke [Class 1])

Table S7 Summaries of studies by research questions

Table S7 provides detailed summaries of the studies based on the research questions, covering preprocessing steps, denoising and cleaning, feature extraction, prediction models, performance metrics, and hyperparameter optimization (HPO) methods.

Studies	Feature selection	Feature extraction	Remove noise or Denoising	Prediction Model	Performance metrics	Hyperparameter optimization
(Yang & Zou 2020)	distal_DBA, HFE, mRMR, UnivariateFS	ADASYN, SMOTE, RandomOverSampler	Normalizer, Scaler, LogIp, Transformer	mAML: tree based classfier, LinearSVC, GaussianNB, LogisticRegression, SGD, KNN	Accuracy, F1 score, precision, recall, roc-auc, log loss	Parallel grid search
(Sakagianni et al. 2020)	Google AutoML Cloud Vision	Google AutoML Cloud Vision	Google AutoML Cloud Vision	Google AutoML Cloud Vision	Recall, precision, roc-auc, F1score	Google AutoML Cloud Vision (name is Vertex AI now)
(Olsavszky et al. 2020)	Parallel Heuristic Search Process of AutoTS, NUTS 2 Hospital Regions	Derivation Window, Time Series Function, Automatic Feature Derivation	Stationarity, moving averages,	AutoTS	Gamma Deviance,RMSE, R-Squared, MAE, MAPE	-
(Valarmathi & Sheela 2021)	Sequential Forward Selection with ten-fold cross validation	with AutoML	with AutoML	Random Forest model optimized with TPOT	F1-Score, Precision, Sensitivity, Specificity, ROC-AUC	Grid Search, Random Search and Genetic Programming (TPOT Classifier)
(A. Romero et al. 2022)	with AutoML	with AutoML	with AutoML	Auto-Sklearn, TPOT, H2O.ai AutoML ve Google AutoML	Accuracy, F1 score, sensitivity, and precision	with AutoML
(Mallikarachchi et al. 2023)	ANOVA, chi-squared test, and covariance analyses, Recursive Feature Elimination (RFE)	with AutoML	with AutoML	Traditional ML model, TPOT, H2O (XGBoost)	Accuracy, F1 score, precision, recall,	Grid search cross-validation, Gradient Boosting, Random Forest, and XGBoost.
(Bhowmik et al. 2023)	Featuretools	Featuretools	Exploratory data analysis (EDA) with Pandas Profiling	CloudAISim with Auto-Keras using Google Cloud Platform	Accuracy, F1 score, precision, recall,	AutoKeras hyperparameters optimization
(Paladino et al. 2023)	With AutoML	With AutoML	removed to ensure data integrity	AutoKeras v1.1.0, PyCaret v3.0, and AutoGluon v0.7.0	Accuracy, F1 score,	Random Forest, AdaBoost, Gradient Boosting, XGBoost with Grid Search

(Dobrijević et al. 2023)	With AutoML	With AutoML	With AutoML	SVM, RF, LR, kNN, ANN wit Weka and AutoML	Accuracy, Sensitivity Specificity	With AutoML
(Wang et al. 2024a).	With AutoGluon and explainability is ensured with SHAP	With AutoGluon	With AutoGluon	AutoGluon, Amazon Web Services (AWS) with "4 bag-fold"to reduce overfitting and "1 stack-level1 for meta-learner	Accuracy, F1 score, precision, recall, AUC	Automatically implemented within the AutoGluon framework
(Khan et al. 2024)	With AutoML	With AutoML	k-means clustering + IQR (interquartile range) based stochastic filtering.	Heuristic (k-means) + Stochastic (IQR) analysis + AutoML regression process	R2 Score, MAE, MSE, RMSE and MAPE	With AutoML
(Mohammadi & Nguyen 2024)	R and Python scripts with ChatGPT	R and Python scripts with ChatGPT	CLAHE technique by Fiji software	ChatGPT and Vertex AI is an AutoML	Accuracy, Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value, F1 Score	Vertex AI
(Rimal et al. 2024)	Separation of Dependent and Independent Variables	Normalization, Fit and Transform (StandardScaler)	mean value imputation and one-hot encoding	SVM, LR, ANN, KNN, RF, GaussianNB, DT, Perceptron, and H2O AutoML	Accuracy, Precision, Recall, F1-Score, AUC-ROC, LogLoss, Gini, MSE, RMSE	H2O package
(Simon et al. 2024)	Analysis of important variables	New boolean properties added	Removed unnecessary columns, translated from Romanian to English	Light Gradient Boosted and Generalized Linear Model with AutoML	F1 Score, Accuracy, AUC, LogLoss, RMSE, FVE Binomial, Max MCC	Early Stopping, ElasticNet, XGBoost,
(Cheong et al. 2024)	Selected specific imaging modalities and slices with LabelBox	A single meaningful 2D slice was extracted from the 3D MRI; semantic labels were also extracted by experts (Yes/No).	Visually or radiologically inadequate images were omitted	Google Cloud AutoML Vertex AI	Sensitivity, specificity, accuracy	Google Cloud AutoML Vertex AI

(Jose et al. 2024)	Education with less but more meaningful data	Label encoding, One-hot encoding and Feature scaling	Some features have been excluded due to their complexity or the absence of data.	PyCaret library	F1-Score, Accuracy, Precision, Recall, AUC-ROC, Kappa, MCC	Fine-tuning the max_depth hyperparameter to avoid overfitting
(Wang et al. 2024b)	Grey level co-occurrence matrix, Gabor filter, Gauss Markov Random Field, Tamura features, grey histogram.	ResNet50 V2 and Xception 1*64 eigenvector	Masked as the ROI with Matlab	H2O AutoML, ResNet-50, Few-shot learning (FSL), sonographer model	F1-score Accuracy, recall, precision	Automatic hyperparameter tuning and K-fold cross-validation
(Castagno et al. 2025)	7 feature selection algorithms with AutoPrognosis V.2.0	7 feature scaling algorithms AutoPrognosis V.2.0	7 feature selection algorithms with AutoPrognosis V.2.0	23 classification algorithms with AutoPrognosis V.2.0	AUC-PRC, AUC-ROC, F1 score, Precision, Recall	100 iterations of Bayesian optimization, KernelSHAP
(Değer & Can 2025)	Automatic feature selection with AutoML	Average Derivative (AD) and Polynomial Features (PF)	Random Over-Sampling (ROS)	PyCaret, H2OAutoML, TPOT, AutoGluon, FlaML, Mljar, PyTorch	Accuracy, precision, recall, F1-score,	hyperparameters automatically tuned with AutoML
(Kuanr & Mohapatra 2025)	MOGA, MOHHO, MOGAHHO	PCA, SVD, AE	Top-3 features, best classifier, for the reduced dataset,	Tree-based Pipeline Optimization Tool (TPOT)	BA, MCC, BM, MK, Precision@Top-N	TOPSIS, HHO, GA
(Ye et al. 2025)	Automatic (built-in methods in AutoML)	TF-IDF-Khmer, DNA composition components	Intra-model importance assessments (SHAP)	AutoGluon (ensemble-based models)	Sensitivity, Specificity F1, MCC, Acc, auPRC, and auROC	AutoML (automatic optimization of AutoGluon)
(Capó et al. 2025)	No statistical or algorithmic method for feature selection is mentioned.	All MRI volumes harmonized, Volumetric brain region features extracted with FastSurfer	FastSurfer compatible format and normalization, unsigned 8-bit UCHAR images, (256^3) and $1mm^3$ isotropic voxel used	Ridge, Elastic Nets, LASSO and OLS, TPOT, FLAML and XGBoost/LGBM	MAE, R^2, MSE, Mean PAD per bin, corr, AUROC	LASSO and Ridge, TPOT and FLAML, and Hyperopt (Bergstra et al. 2015)
(Xu et al. 2025)	PCA, t-SNE, UMAP, ElasticNet, Lasso and Ridge	Feature Detection with LDI-MS, Intensity and Range Filtering	Polynomial fitting, wavelet denoising, mean loading, ComBat, batch normalization, Gaussian filter.	ElasticNet, Lasso and Ridge regression, and PyCaret AutoML platform	AUC, Accuracy, F1 Score, Kappa Score, MSE, Precision, Recall, Sensitivity, Specificity	Random search, grid search, and more advanced methods like Bayesian optimization

(Chatterjee 2025)	SelectKBest, LIME	Variance Threshold	Missing data analysis and filling, outlier control.	TPOT (AutoML) VarianceThreshold (0.1) + DecisionTreeClassifier	Accuracy, Precision, Recall, F1-Score, MCC	Genetic algorithm-based TPOT has automated HPO

Table S8 Dataset, disease, AutoML-ML method and performance

The table summarizes key information from studies reviewed in the paper. "Dataset / Data Type" refers to the datasets and the type of data used in the study. "Disease / Task" indicates the specific disease targeted and the machine learning task performed, such as classification or prediction. "Author (Year)" shows the name of the study's author(s) and its publication year. "ML/AutoML Method" lists the machine learning or automated machine learning methods applied in the study. "Best Accuracy / Performance" presents the highest accuracy or performance metric reported, reflecting the effectiveness of the applied methods.

Dataset / Data Type	**Disease / Task**	**Author (Year)**	**ML/AutoML Method**	**Best Accuracy / Performance**
13 Benchmark Microbiome Datasets	Multi-class disease classification	(Yang & Zou 2020)	mAML	High performance (exact value not given)
Chest CT Scans	COVID-19 pneumonia detection	(Sakagianni et al. 2020)	AutoML (unspecified)	Average precision 0.932
Romanian Hospitalization ICD-10 Time Series	Forecasting top 10 fatal diseases	(Olsavszky et al. 2020)	AutoTS	High accuracy (not specified)
Cleveland Heart Disease (CHD) Dataset	Coronary heart disease prediction	(Valarmathi & Sheela 2021)	RF + TPOT (Hyperparameter Optimization)	97.52% accuracy
Z-Alizadeh Sani Dataset	Coronary artery stenosis prediction	(Valarmathi & Sheela 2021)	RF + Random Search	80.2% (LAD), 73.6% (LCX), 76.9% (RCA)
Large-scale Insurance Claims Data	General disease prediction	(A. Romero et al. 2022)	TPOT, Auto-Sklearn, H2O.ai, Google AutoML	TPOT highest accuracy (not specified)
T2D, CKD, IHD Datasets	Type 2 Diabetes, Chronic Kidney Disease, Ischemic Heart Disease	(Mallikarachchi et al. 2023)	Auto-Sklearn, TPOT, H2O	T2D Auto-Sklearn: 86.8%, CKD TPOT:

				99.65%, IHD TPOT: 74.56%
Breast Cancer, Heart Disease, Diabetes, COVID-19	Multi-disease classification	(Bhowmik et al. 2023)	Auto-Keras	96% - 98% accuracy range
Cleveland, Hungarian, Combined (Heart)	Heart Disease	(Paladino et al. 2023)	AutoGluon (78–86%), PyCaret (65–83%), AutoKeras (54–83%), Traditional ML (55–60%)	78–86% (AutoGluon)
Hematochemical Parameters (Pediatric Patients)	SARS-CoV-2 vs Influenza infection	(Dobrijević et al. 2023)	AutoML (unspecified)	98.4% accuracy
Combined Heart Disease Datasets	Coronary artery disease prediction	(Wang et al. 2024a)	AutoGluon (ensemble + HPO)	91.67% accuracy, AUC 0.9562
Patient Rehabilitation Data	Health indicators prediction (Borg RPE, TUG)	(Khan et al. 2024)	AutoML + Hybrid (k-means + regression)	R^2: 98.55% and 98.50%
Diabetic Retinopathy	Image-based diagnosis	(Mohammadi & Nguyen 2024)	AutoML + ChatGPT	92%–95% accuracy
Heart Disease Dataset (303 samples)	Heart disease prediction	(Rimal et al. 2024)	AutoML (GLM best), ANN deep learning	88% (GLM), 89.6% (ANN)
Romania COVID-19 Dataset	COVID-19 outcome prediction	(Simon et al. 2024)	AutoML (unspecified)	F1=96.44%, Accuracy=98.84%
Sinonasal Disease MRI Images	Sinonasal disease detection	(Cheong et al. 2024)	Vertex AI (Google Cloud)	92% accuracy
Cardiovascular Risk in Diabetic Patients	Cardiovascular risk prediction	(Jose et al. 2024)	PyCaret + LightGBM, XGBoost	AUC 0.83 (LightGBM), balanced performance with XGBoost
AEH, NAEH, EC TVU Images	Endometrial disease diagnosis	(Wang et al. 2024b)	H2O AutoML, FSL, ResNet50 V2 + KNN	87.8% accuracy
Knee Osteoarthritis (OA)	OA progression prediction	(Castagno et al. 2025)	AutoML (unspecified)	AUC-PRC 0.727 (multi-class), 0.764 (binary)
Hepatitis C Virus (UCI ML Dataset)	HCV diagnosis	(Değer & Can 2025)	7 different AutoML tools	99.29% - 100% accuracy range

Multiple Disease Datasets (10 datasets)	General disease prediction	(Kuanr & Mohapatra 2025)	TPOT + GA + HHO (MOGAHHO)	Highest accuracy (not specified)
DNA Sequences	Bioinformatics classification	(Ye et al. 2025)	AutoGluon	97.14%, 79.71%, 98.73% accuracy rates
Brain Age Estimation (UK Biobank, ADNI, NACC)	Neurodegenerative disease biomarkers	(Capó et al. 2025)	Penalized linear models	AUROC ~0.90
Benign Prostatic Hyperplasia (BPH)	BPH diagnosis	(Xu et al. 2025)	MALDI-MS + Metabolomics + AutoML	AUC = 0.830
Stroke	Stroke deterioration prediction	(Chatterjee 2025)	TPOT + Decision Tree + Variance Threshold	95.2% accuracy

Table S9 Binary contributions of selected studies according to research questions

Study	**RQ1 AutoML Methods**	**RQ2 Input Features**	**RQ3 Feature Engineering**	**RQ4 Denoising Methods**	**RQ5 Train-Test Split**	**RQ6 Performance Metrics**	**RQ7 Hyperparameter Tuning**
(Yang & Zou 2020)	1	1	1	1	1	1	1
(Sakagianni et al. 2020)	1	0	0	0	0	1	0
(Olsavszky et al. 2020)	1	1	1	1	1	1	0
(Valarmathi & Sheela 2021)	1	0	0	0	1	1	1
(A. Romero et al. 2022)	1	0	0	0	1	1	0
(Mallikarachchi et al. 2023)	1	0	0	0	1	1	1
(Bhowmik et al. 2023)	1	1	1	1	1	1	1
(Paladino et al. 2023)	1	0	0	1	1	1	1
(Dobrijević et al. 2023)	1	0	0	0	1	1	0
(Wang et al. 2024a).	1	0	0	0	1	1	1
(Khan et al. 2024)	1	1	1	1	1	1	1
(Mohammadi & Nguyen 2024)	1	1	1	1	1	1	1

(Rimal et al. 2024)	1	1	1	1	1	1	1
(Simon et al. 2024)	1	1	1	1	1	1	1
(Cheong et al. 2024)	1	1	1	1	1	1	1
(Jose et al. 2024)	1	1	1	1	1	1	1
(Wang et al. 2024b)	1	1	1	1	1	1	1
(Castagno et al. 2025)	1	1	1	1	1	1	1
(Değer & Can 2025)	1	1	1	1	1	1	1
(Kuanr & Mohapatra 2025)	1	1	1	1	1	1	1
(Ye et al. 2025)	1	1	1	1	1	1	1
(Capó et al. 2025)	1	1	1	1	1	1	1
(Xu et al. 2025)	1	1	1	1	1	1	1
(Chatterjee 2025)	1	1	1	1	1	1	1