

Table S1: Characteristics of the included studies

Author	Study Focus	Data Source	Population	Predictors	Machine Learning Algorithm(s)	Evaluation Metrics	Model Validation Techniques	Explainability	Summary of Findings
Yadav et al. [32] India 2022	General Health Risk	Secondary data: The dataset was obtained from the UCI machine learning repository	Not specified	Age, sex, resting blood pressure, serum cholesterol levels, fasting blood sugar, resting electrocardiogram (ECG) results, maximum heart rate, exercise-induced angina (chest pain or discomfort), and specific ECG parameters	Decision Tree	Accuracy	Not specified	No	93.7% accuracy was reported for health risk prediction based on the introduced ensemble learning-based approach. The implemented models were shown to efficiently predicts health risks using labelled training data and converges in 30 iterations. To enhance the model, the study is considering future research exploring clustering methods and dimensionality reduction techniques like PCA
Singh et al. [33]	General Health Risk	Primary data: The data was collected	Asians located in North Indian. The	16 attributes or the indicators of	SVM	Accuracy and AUC	The internal validation procedures	No	The study highlights that feature

India 2019		using Omron body composition machine	dataset contains 500 instances	the health risk were collected, including birth age, height, weight, gender, body fat, blood pressure and fasting blood sugar			involved balancing the dataset with equal proportions of feature rows from different classes (Fitness Level 1 and Fitness Level 2) to prevent bias.		selection and dataset quality are crucial for effective machine learning classification and prediction. It finds that the Medium Radial SVM kernel performed the best and emphasizes that SVMs handle noisy and overlapping data more effectively than regression methods. The research concludes that Graph theory and social network analysis are effective for disease risk prediction, providing a viable alternative for utilizing administrative data in healthcare.
Khan et al. [53] Australia 2019	Type 2 Diabetes	Secondary data: Obtained from the by private healthcare funds based in Australia.	Hospital admission information for about 749,000 deidentified patients. There are two cohorts: the cohort of patients diagnosed with T2D and the cohort who are non- chronic—that	Demographic, behavioural (age, sex, behaviour) and graph/health trajectory based (graph node match, graph pattern match and network cluster match)	Logistic Regression	Accuracy	10-Fold Cross Validation	No	

Ashwath et al. [34] India 2024	Cardiovascular Health Risk	Secondary data: Three datasets: Framingham Dataset, Cleveland Heart Study Dataset (KEEL) and Svetlano Ulianova Heart Study Dataset	is, do not have T2D Not specified. Only the samples size for Svetlano Ulianova Heart Study dataset was reported as 70,000	The attributes collected include age, height, weight, gender, blood pressure, heart rate, and blood sugar.	Ensemble of Decision trees, random forests, logistic regression, support vector machines, and neural networks.	Accuracy	K-Fold Cross-validation	No	With the implemented model achieving an accuracy of up to 99.0% on different datasets, the study advocates for integrating advanced machine learning in health risk assessments, with future efforts aimed at improving model accuracy and user interface design.
Chen et al. [39] China 2021	General Health Risk through Posture Recognition	Not Specified	Not Specified	Not Specified	Decision trees	Accuracy	Not specified	No	A dynamic health risk assessment model developed using sports gesture recognition and daily habit monitoring demonstrated effective performance in evaluating physical health risks.

Li et al. [40] China 2023	Asthma	Primary data: Collected from asthma patients in the Affiliated Shuguang Hospital of Shanghai Traditional Chinese Medicine University.	The dataset includes 152 samples with the majority aged 50-69 years, and a gender distribution of 40% males and 60% females.	Twenty-four indicators, including complete blood count differentials and red blood cell indices, were collected.	New model, SVM, support vector regression, Logistic regression, Random Forest and Lasso	Accuracy and AUC	2-Fold cross-validation	Somewhat: Using a correlation heatmap to understand feature relationships.	Based on the evaluation of three different feature subsets, it was revealed that the proposed model achieved an accuracy of up to 78.18%, with performance across all subsets ranging from 72.92% to 78.18%. This performance result from the integration of an affinity graph which enhanced the model's ability to capture intricate interactions within the data and improve overall prediction performance
Thakkar et al. [35] India 2024	Maternal Health Risk	Secondary data: The maternal health risk factors dataset was obtained from	Majority of the samples in the datasets represents pregnant women who pose minimal	The features collected include age, blood pressure, blood glucose level (BS),	Ensemble of Random Forest, XG Boost, Support Vector Classifier	Accuracy	Not specified	No	The random forest algorithm with hyperparameter tuning was more effective than other

		the UCI ML public source.	health risks. Out of 1,014 rows and 7 columns in the dataset, 40% of pregnant women were classified as low risk, 33.1% as medium risk, and 26.8% as high risk.	heart rate, and risk level.	(SVC), Decision Tree, and Logistic Regression.				models in predicting maternal health risks.
Shifa et al. [48] Bangladesh 2024	Maternal Health Risk	Primary data: Clinical data was collected from Kurigram General Hospital, Bangladesh,	Although demographic information was not specified, 1102 distinct samples were collected	11 important features were collected. It includes age, systolic and diastolic blood pressure measures, blood sugar levels, body temperature, body mass index (BMI), previous health issues.	XGBoost, Decision Tree, Catboost, Random Forest, Gaussian Processes, SVM with Kernel, Naïve Bayes, Perception, and Ridge	Accuracy	Holdout Validation	Somewhat: Using a correlation heatmap to understand feature relationships.	Machine Learning Effectiveness: It was found that machine learning models, particularly XGBoost and Random Forest, significantly enhance maternal health risk prediction accuracy. Interpretability Challenges: While the models achieved high accuracy, issues with interpretability and model explainability

									remain, highlighting the need for improvements to support clinical decision-making. The models using demographic, lifestyle, family history, and personal health device data, with or without laboratory data, demonstrated similar predictive performance. Feature importance analysis revealed that modifiable lifestyle factors are crucial for developing effective health guidelines, supporting the practicality of using non-laboratory data for health
Park et al. [43] South Korea 2024	General Health Risk	Primary data: Collected from the National Sample Cohort of the Korean National Health Insurance Service (NHIS).	425,148 NHIS members who underwent medical check-ups between 2009 and 2012.	Four categories of predictor variables were collected: demographic (e.g., age), lifestyle (e.g., smoking and alcohol usage), family history (e.g., history of hypertension), personal health (e.g., blood pressure measurements), and laboratory data (e.g., total cholesterol).	XGBoost and logistic regression	Area under the curve (AUC), accuracy, and F1-score.	10-fold cross-validation	Yes, using feature importance algorithm	

Shomorony et al. [49] USA 2020	General Health Risk	Primary data: Collected from 1,253 self-assessed healthy individuals in a clinical research facility, utilizing tools and techniques such as whole genome sequencing (WGS) and microbiome sequencing.	Data were collected from 1253 individuals initially and then 1083 individuals from the longitudinal cohort were used to test and validate the results.	1385 data features were collected from diverse modalities, including metabolome, microbiome, genetics, and advanced imaging,	Unsupervised learning method (Gaussian Processes)	p value	External Validation	Somewhat: Using a correlation heatmap to understand feature relationships.	risk assessments. The research identified novel biomarkers, including specific metabolites for diabetes, gut microbiome health, and lean mass. It also found early signatures for hypertension and poor metabolic health, and new associations between a uremic toxin and microbiome genera.
Sahoo et al. [50] USA 2024	Mental Health Risk	Primary data: collected via interviews and survey	Participants included 120 Sri Lankan Tamil parent-adolescent dyads, with adolescents aged 13-18, selected through stratified purposive sampling from	Mental and cognitive status was assessed using the MINI Mental Health exam with basic psychology questions	Glmboost algorithm, which is a variant of linear regression	RMSE	10-Fold Cross-validation	No	The proposed Glmboost model effectively captured non-linear relationships and identified key predictors of declining mental well-being, such as family

			180 refugee camp households.						functioning and support, which were overlooked by traditional linear regression, demonstrating the model's superiority in complex analyses.
Diamantoulaki et al. [56] Greece 2022	General Health Risk	Secondary data: Collected from three public repository including Kaggle	Not Specified	Several risk factors were extracted for different health conditions. For example, age, gender, resting blood pressure, and cholesterol were collected to assess heart failure, while blood sugar levels were gathered to evaluate diabetes	Deep neural network	Binary accuracy, precision, recall, AUC, specificity, and F1-score	Not specified	No	The implemented federated learning (FL) model achieved performance comparable to existing centralized learning model, suggesting that investigating future health risk assessments using FL is crucial.
Rashid et al. [44] South Korea 2022	General Health Risk	Secondary data: Collected from public repository	Not specified	Not explicitly specified	SVM, random forest, k nearest neighbour, deep neural network, neural network, naive bayes, and	Accuracy	10-Fold Cross Validation	No	The study revealed that particle swarm optimization effectively improves prediction accuracy and efficiency for

					logistic regression				diagnosing multiple chronic diseases.
Parvataneni et al. [51] USA 2023	Fetal Health Risk	Secondary data: Collected from the University of California and the University of Porto, Irvine	Records of 2126 pregnant women	2126 cardiocotograph recordings	Catboost, Gradient Boosting, LightGBM, XGBoost and Random Forest	Precision, Accuracy, Recall, Precision, F1	Holdout and Cross-validation	No	The study found that the Catboost model achieved the highest accuracy of 95.3%, while the Random Forest Classifier in Python had a higher accuracy (94%) compared to the IBM Watson-trained model (93.3%)
Chung et al. [45] South Korea 2018	General Health Risk	Primary data: Collected from the Korean National Health and Nutrition Examination Survey	Not Specified	Pulse rate, height, weight, ECG, body temperature, body max index, eyesight, hearing, blood pressure, blood sugar	Deep neural network	RMSE	Not specified	No	The health risk alert service showed strong potential for chronic disease management, with expected improvements in reliability as data expands. Future efforts is recommended to enhance data integration

Pawar et al. [36] India 2022	Maternal Health Assessment	Secondary data: Collected from the UCI machine learning repository.	Not Specified	Age, Systolic Blood Pressure, Diastolic BP, Blood Sugar, Body Temperature, Heart Rate	Decision table, Naive Bayes Multilayer Perceptron, Decision tree, Random Forest, Bagging, Logistic Model Tree, Reduced Error Pruning Tree	Accuracy and AUC	K-fold Cross Validation for k= 0-10	No	and algorithm efficiency. The implemented models shows relative robustness and reliability to support healthcare services.
Hossain et al. [47] Saudi Arabia 2020	Diabetes	Primary data: Collected from the patients using direct questionnaire from Sylhet Diabetes Hospital of Sylhet, Bangladesh	505 instances of early-stage diabetes risk prediction	A number of 15 attributes including age, polyuria, polydipsia, gender, sudden weight loss, visual blurring, and obesity were collected	Naive bayes, logistic regression, decision tree and random forest	Accuracy, precision, recall and F- measure	Not specified	No	The experiments show that ignoring a sign or symptom (considered an epidemiologic al factor) can significantly impact classifier performance. The implemented MLP model outperformed other regression models in predicting pulmonary aerosol deposition, with higher deposition efficiency (DE) in the
Islam et al. [54] Australia 2022	Human Respiratory Tract	Secondary data: computational fluid dynamics (CFD) data via simulation	Not specified	Not specified	Random forest, k- nearest neighbour, multi-layer perceptron and Gaussian process regression	RMSE, Mean Percentage Error and Mean Absolute Percentage error	Not specified	No	

Bayati et al. [52] USA 2018	General Health Risk	Secondary data: Taken from a publicly available website operated by Healthone Labs	Not specified	Not Specified	New model	AUC	5-Fold Cross validation	No	right lung. It also identified the right lower lobe as having the highest DE, providing insights for enhancing targeted drug delivery to the lower airways. The multitask learning model accurately predicted the onset of nine diseases at a low cost, outperforming the single task learning model in accuracy, cost, and complexity. However, comparisons with common disease scoring methods were limited due to unavailable biomarkers in the dataset. The findings suggest that personalized treatment for PAC can be
Kurysheva et al. [57] Russia 2023	Glaucomatous optic neuropathy	Primary data	76 Caucasian patients aged 41 to 80 years, examined from January	gender, Intraocular pressure, axial length, and	New model	Accuracy	Procrustes cross-validation	No	

			2019 to December 2021	anterior chamber depth					effectively guided by predictive models comparing IOP reduction from LE and LPI. A shorter model using only four parameters (gender, IOP, AL, and ACD) performs well, simplifying clinical application.
Salian et al. [37] India 2023	General Health Risk	Secondary data: Collected from three public repository including UCI	Not specified	Not specified	K-Nearest Neighbour, Logistic Regression, Stochastic Gradient Boosting, Voting Ensemble and Decision Tree	Accuracy, Recall, Precision and F1-score.	Not specified	No	The proposed models achieved high prediction accuracy for heart attack (93.44% with Decision Tree), diabetes (98.55% with Decision Tree), and obesity (96.6% with Stochastic Gradient Boosting) using various datasets.
Ge et al. [41] China 2020	General Health Risk	Secondary data: Collected from public repository	19733 patients with an average age of 67 years. The proportion of	Not specified	New model	Hamming loss, ranking loss, average precision, one	Not specified	No	The multi- label neural network algorithm significantly

		such as MIMIC-II published by Beth Israel Deaconess Medical Centre	men and women was 56% and 44%			error and coverage			outperformed traditional multi-label learning methods in identifying complications across 10 chronic diseases. The model also demonstrated competitive running time efficiency due to the effectiveness of the multi-label loss function.
Anbu et al. [38] India 2017	Women Health Risk	Secondary data: Collected from the Melinda Gates Foundation	9000 women aged between 15 to 30 from 9 underdeveloped regions were surveyed when they visited the clinics.	50 attributes were collected. 2 were identifiers, 3 were target attributes, and the remainder included lifestyle factors (e.g., use of mobile, fridge, computer) and sexual health factors (e.g., condom use)	Random forest, neural network, XGBoost, Conditional Inference Tree	Accuracy	K Fold-Cross validation	Yes, using feature importance algorithm	The study demonstrated the effectiveness of machine learning approach in classifying women into health risk segments and subgroups, enabling the creation of customized awareness programs that save time and money. Future work could extend this

Lin et al. [42] China 2023	Obesity	Primary data: Collected from Chinese subjects from Ningde City in Fujian province, southeast China,	10,905 Chinese subjects from Ningde City in Fujian province, southeast China,	Features collected include blood pressure, blood samples and body mass index	SVM, random forest, decision tree, K-nearest neighbour, Catboost, Gradient boosting machine, Gaussian and neural network and Logistic regression	AUC	5-Fold cross-validation	Yes, using the SHAP algorithm	classification to include all age groups and both genders It was highlighted that the Catboost machine learning model, combined with SHAP visualization, effectively identifies and quantifies risk factors for overweight, significantly outperforming traditional methods. The research demonstrated that AI detection technology for predicting hypertension risk shows high consistency with biochemical test data, indicating its feasibility and potential for effective risk
Yang et al. [46] Taiwan 2023	Hypertension	Primary data: Collected through a study or experiment involving direct recruitment of participants	12 white-collar participants (6 males and 6 females), aged 20 and above, were selected, regardless of hypertension history	Hypertension Vascular, Dysfunction Arterial, Stiffness Index, Vascular Stiffness Index, Visceral Fat, Body Fat	New model	Accuracy	Not specified	No	

Liu et al. [55] Australia 2023	General Health Risk	Secondary data: MIMIC- III and eICU are publicly available databases of electronic health records collected from intensive care units	21,139 and 38,056 patient journeys were extracted from MIMIC-III and eICU, respectively	Features extracted include Diastolic Blood Pressure, Glucose, Heart Rate, Oxygen Saturation, and Temperature	Deep neural network	AUC and Area Under the Precision- Recall Curve	Holdout validation (70:15:15 ratio)	Yes, using feature importance algorithm	warning and preventive health management. The introduced Stacked Attention- based Network design facilitated enhanced interpretability , offering clear insights into the model's decision- making process, as evidenced by various analytical experiments and case studies.
--------------------------------------	------------------------	---	---	--	------------------------	---	--	--	---