

---

This is the **submitted version** of the preprint:

Estrella Arraez, Antonio [et al.]. «Machine learning for the analysis of healthy lifestyle data : a scoping review and guidelines (Preprint)». *JMIR Human Factors*, 2025, p. 1-65

---

This version is available at <https://ddd.uab.cat/record/323957>

under the terms of the  license.

# Machine learning for the analysis of healthy lifestyle data: a scoping review and guidelines

Tony Estrella<sup>1,2,\*</sup>, Lluís Capdevila<sup>1,2</sup>, Carla Alfonso<sup>1,2</sup>, Josep-Maria Losilla<sup>2,3</sup>

<sup>1</sup> Department of Basic Psychology, Universitat Autònoma de Barcelona (UAB), Bellaterra, Barcelona, Spain.

<sup>2</sup> Sport Research Institute, Universitat Autònoma de Barcelona (UAB), Bellaterra, Barcelona, Spain.

<sup>3</sup> Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, Bellaterra, Spain

\*Corresponding author: Tony Estrella, Department of Basic Psychology, Universitat Autònoma de Barcelona (UAB), Bellaterra, Barcelona, Spain, [antonio.estrella@uab.cat](mailto:antonio.estrella@uab.cat)

## Abstract

**Background:** Advances in data science and technology have transformed lifestyle studies by enabling the integration of multimodal information and generation of large volumes of data. Despite the growing interest in machine learning (ML) in health behaviour research, significant methodological gaps remain.

**Objectives:** The study aims to systematically review the applications of supervised ML algorithms in analyzing healthy lifestyle (HL) data, with a specific focus on the methodological approach employed. The specific objectives are to explore the types and sources of data used in health outcomes, examine the ML processes employed, including explainability artificial intelligence (XAI) methods, and review the software tools utilized. Additionally, this review aims to provide practical guidelines to enhance the quality and transparency of future ML research in health.

**Methods:** Following the PRISMA-ScR recommendations, the search was conducted across PubMed, PsychINFO, and Web of Science, resulting in 48 studies that meet the inclusion criteria.

**Results:** Most studies (37, 77%), integrated multidomain data from physical activity, diet, sleep, and stress. Data sources were split between self-acquired (25, 52.08%) and health repositories (23, 47.92%). Single items measurements were common, particularly for physical activity, diet and sleep. Despite a multimodel approach in 28 studies, random forest was the most frequently used algorithm. Only 10 studies (20.83%) employed XAI methods, with 9 using SHapley Additive exPlanation (SHAP) values and 1 using Local Interpretable Model-agnostic Explanations (LIME). R was the most widely used software, with variations in the libraries employed.

**Conclusion:** This review highlights methodological gaps in the application of supervised ML to HL data. The ML workflow should span from data acquisition to explainability, with iterative steps to improve the process. Multidomain approaches in data acquisition enhance understanding of health issues related to lifestyle but are constrained by low data representativeness due to methodological limitations in acquisition. While random forest was prevalent, a multimodel approach is recommended for comprehensive comparison. Lifestyle components consistently ranked among the top features in studies that incorporated XAI. Integrating XAI methods into the ML pipeline can support personalized interventions, provided the data is accurately collected. The R metapackage *tidymodels* facilitates process evaluation through unified syntax, improving replicability. Methodological and reporting guidelines are provided to enhance transparency and replicability in multidisciplinary ML research.

**Keywords:** machine learning; artificial intelligence; healthy lifestyle; physical activity; diet; sleep; stress; review; data analysis; XAI

## Introduction

There is a growing interest in understanding the effects of synergistic relationships among lifestyles behaviors and their effect on health outcomes [1,2]. Traditionally healthy lifestyles (HL) research has primarily focused on physical activity and diet. However, recent studies increasingly include sleep and stress management as critical components of lifestyle [3,4]. For instance, stress has been shown to negatively influence physical activity, sleep, and dietary habits [5], which in turn have an overall impact on health and well-being. Thereby adopting a HL has emerged as a cost-effective strategy to mitigate the incidence of noncommunicable diseases like cardiovascular and metabolic diseases [6].

Technological advances, including wearable devices and lifelogging processes, have significantly enhanced the capability to collect multimodal, high-frequency, and ecological lifestyle data [7,8]. This wealth of data provides valuable contextual information and insights for researchers and users [9]. However, the vast amount and complexity of behavioural and physiological data expose significant analytical challenges. Traditional statistical models often struggle with the high-dimensionality, heterogeneity, and non-linearity typical of lifestyle studies. Recent progress in computational power, and artificial intelligence (AI), particularly machine learning (ML), has contributed to address these limitations [10].

ML models are capable of analyzing complex data types and generate insights and knowledge to improve decision-making [11,12]. Furthermore, ML algorithms can flexibly handle non-linear relationships among features and outcomes. While the boundary between classical statistics and ML is not clear, ML algorithms are recognized for their flexible data-driven approach, avoiding the imposition of a predetermined relational structure between variables [13–15]. Additionally, prioritizing algorithms that maximize the generalizability to new data, often referred as scalability in the big data context, is crucial to face new health challenges [16,17]. These characteristics make ML analysis a suitable methodology for predictive modeling and feature extraction in health-related lifestyle research.

ML models are broadly classified into supervised learning (SL) and unsupervised learning (UL). In SL the model is trained with labelled data, where each observation has an associated response measurement, to predict known outcomes such as disease risk or behavioral adherence [17]. The goal of SL is to fit a model that will be able to predict the response when applied to new data. When the response value is continuous, this is known as a “regression problem”; when the response is categorical it is known as a “classification problem”. In contrast, in UL models, the goal is to discover patterns rather than predict outcomes since there is not an associated response to the input, and the model seeks relationships and similarities between observations. In the health domain, where diagnosis and detection are key focuses, SL and particularly classification tasks, are more prevalent due to their ability to evaluate these predictions [18,19]. Clinical applications of SL include triage systems, prognosis prediction and disease classification using rapid testing [20]. Consequently, SL methods are standard in epidemiology to enhance clinical decisions based on input-output relationships [21]. Since prediction and explainability are central concerns in health research, this scoping review focuses specifically on SL methods.

Despite the growing attention to ML in health behaviour research, there remain significant methodological gaps. Prior reviews have focused primarily on outcome effectiveness or AI chatbots interventions, often providing limited detail about the ML process involved [22]. A recent scoping review on ML methods used in health promotion and behavioral change found that the main interventions studied are those related to physical activity, while other crucial aspects of a HL were overlooked, revealing an imbalance in the current literature [23]. Similarly, Lai et al. [24] reviewed the applications of large language models in exercise recommendations and physical activity, highlighting methodological limitations associated with these AI models. In sum, these studies underscore the need for a more comprehensive review study to include a

holistic concept of HL. Furthermore, methodological details such as data preprocessing, model evaluation, and interpretability are often underreported, hindering transparency, reproducibility, and interdisciplinary collaboration.

To address the lack of interpretability, the eXplainable Artificial Intelligence (XAI) approach has emerged focusing on understanding AI algorithms making them more transparent. XAI aims to provide human-understandable explanations for the decisions made by machine learning models [25]. In HL research, XAI can be directed to identify the set of behaviours that significantly influence on health, thereby enhancing transparency and interpretability. It is important to distinguish between interpretability and explainability in the AI context. While interpretability refers to understanding the influence of each feature in the original model, explainability involves deriving actionable human insights from the model's prediction [26]. Therefore, reporting the method of explainability used in the ML projects is crucial for interpreting how lifestyle factors interact with health outcomes.

Therefore, the present study aims to systematically review the applications of supervised ML algorithms in analyzing healthy lifestyle data, with a specific focus on the methodological aspects employed in these studies, rather than their results. The specific objectives are to explore: a) the specific lifestyle data used in health outcome studies; b) the sources and types of data subjected to analysis; c) the characteristics of the ML models, including XAI methods and interpretability; and d) programs and libraries used for ML implementation. Additionally, based on the findings of this scoping review, we aim to provide practical guidelines to enhance the quality and transparency of future ML research in lifestyle science. A scoping review is the type of systematized review (i.e., systematic, transparent and replicable) most appropriate for addressing these objectives [27].

## Method

To maximize the reporting quality of the scoping review, we followed the “PRISMA Extension for Scoping Reviews” (PRISMA-ScR) recommendations [28]. The protocol for this scoping review was registered at the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) (doi: 10.37766/inplasy2023.3.0065). All data generated in this review is provided in the supplementary file and accessible in the institutional repository [29].

### Search strategy

In this scoping review, we searched for primary studies in the three principal health databases: PubMed by National Centre by Biotechnology Information (NCBI), PsychINFO by ProQuest and Web of Science by Clarivate. The search strategy followed the Peer Review of Electronic Search Strategies (PRESS) [30] and PRISMA for Searching (PRISMA-S) guidelines [31], and it consisted of two groups of search terms referring to: a) healthy lifestyle, and b) machine learning. We also added a third group of terms preceded by the boolean operator “NOT” to improve the specificity of the search strategy.

The search strategy was adapted to the specific syntax of each database (S1 Table 1). The research was conducted in April 2024, with language restriction (English and Spanish), but no restriction on publication years.

### Study selection

Studies were included in the review if the following criteria were met:

- Employed supervised machine learning models for analyzing lifestyle data.
- Analyzed lifestyle behaviors as either inputs or outputs of the ML models.
- Used data from real individuals (not simulations).
- Published in English or Spanish.

Studies were excluded based on the following criteria:

- Focused on UL without connection to SL modeling.
- Focused on mathematical formulation or guidelines for implementing machine learning models in health.
- Used simulated data or aimed to develop a chat bot or app based on machine learning.
- Primarily addressed substance abuse, such as alcohol intake or smoking cessation.
- Focused exclusively on classical statistical regression algorithms, such as lineal or logistic regression, which were not considered ML on their own in this review.

Justification of exclusions: UL algorithms were excluded because there is not an associated response to inputs, thereby lacking performance evaluation. Classical statistical regression algorithms, such as lineal or logistic regression, are not considered ML on their own in this review. Consequently, studies focusing exclusively on this type of statistical algorithms were excluded. However, we acknowledged that the use of a model ensemble approach allows for the inclusion of these statistical algorithms to assess the performance of different algorithms during the evaluation step. Studies on substance use disorders were excluded as they involve distinct behavioral and neurobiological mechanisms that differ substantially from the domains of physical activity, diet, sleep, and stress, which are the core components of healthy lifestyle behaviors as defined in this review.

Two reviewers (TE, CA) independently screened in the first phase title and abstract, and then in the next phase the full text. Discrepancies were resolved by consensus, with the participation of a third reviewer (JML, LC) when necessary. Agreement between reviewers during the selection process was analyzed by calculating Cohen's Kappa.

#### Data management

Mendeley was used as reference manager software, results of the search strategy were entered, and duplicates were merged or removed. An ad hoc checklist was used to extract the information from the papers included. The checklist was divided into 5 sections:

- General information: authors, title, year, and country of affiliation.
- Methodological data: type of study, aim, year of data collection, form of data acquisition, sample, and countries represented in the data.
- Study variables: health issues, lifestyle features, and model's input/output variables.
- Software: Statistical programming language, libraries, and packages.
- Model aspects: Type of problem, stages of ML analysis, ML methods, model evaluation, evaluation metrics, and interpretability.

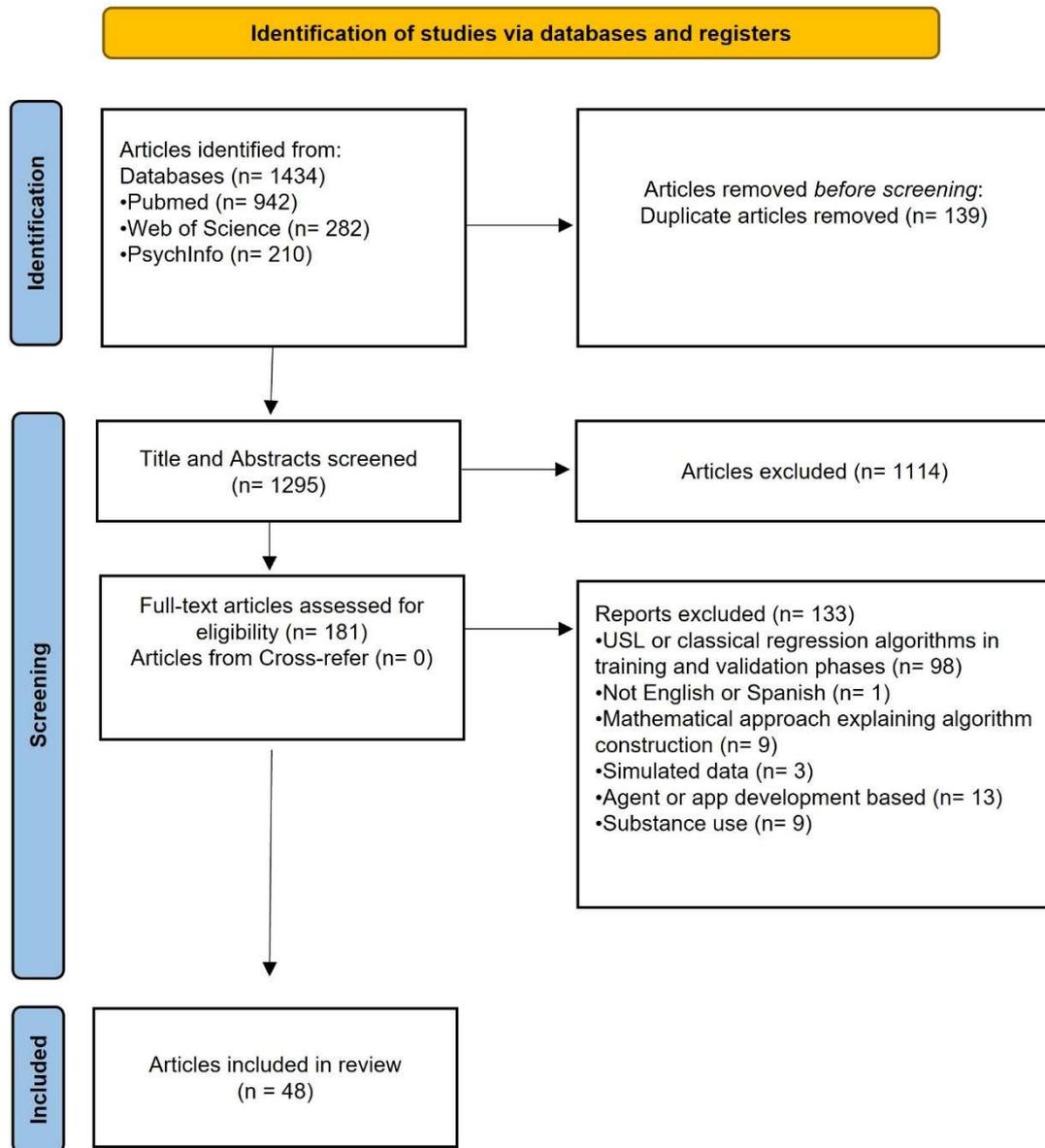
#### Strategy of data synthesis

The review was presented as a narrative synthesis, and the information was summarized in tables and figures. The information extracted from the studies were divided into three blocks: type of data, ML process, and software. For extracting data information, we focused on the components of lifestyles, health outcomes, sources of data, acquisition methods, and data typology. Regarding the ML process, we focused on the whole process, consisting of preprocessing, modeling, validation, evaluation, and interpretability. Finally, we focused on the software used in each study.

#### Results

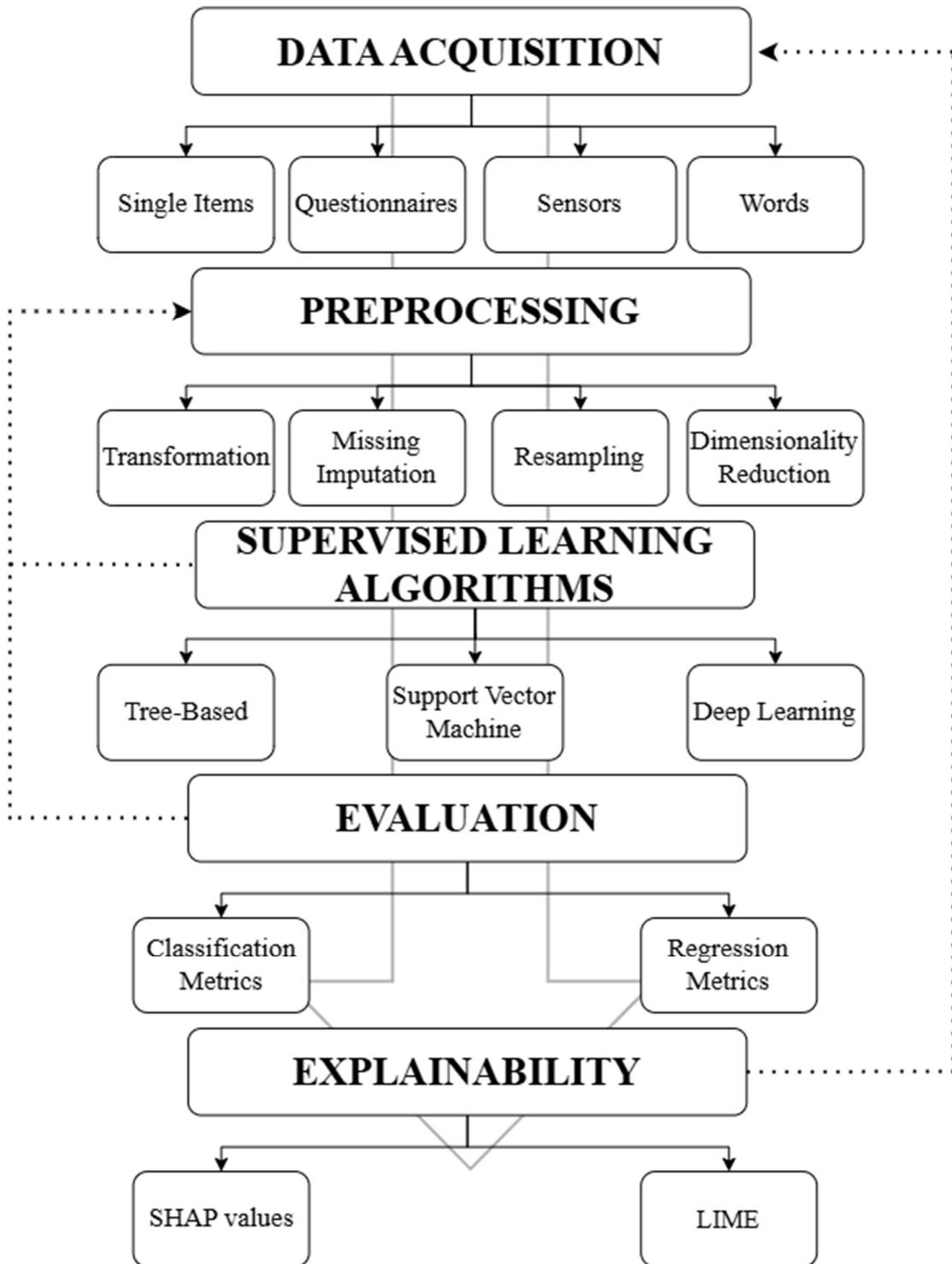
A total of 1434 articles were retrieved from the databases, and 48 studies met the eligibility criteria and were included in this scoping review (see Figure 1). There was very good agreement between reviewers during the selection process: 96% (Kappa=0.84; CI 95%: 0.61,

1.0) in the screening by title and abstract, and 94% (Kappa=0.88; CI 95%: 0.71, 1.0) in the full-text screening.



[Insert **Figure 1**. PRISMA flow diagram of the scientific literature search and selection.]

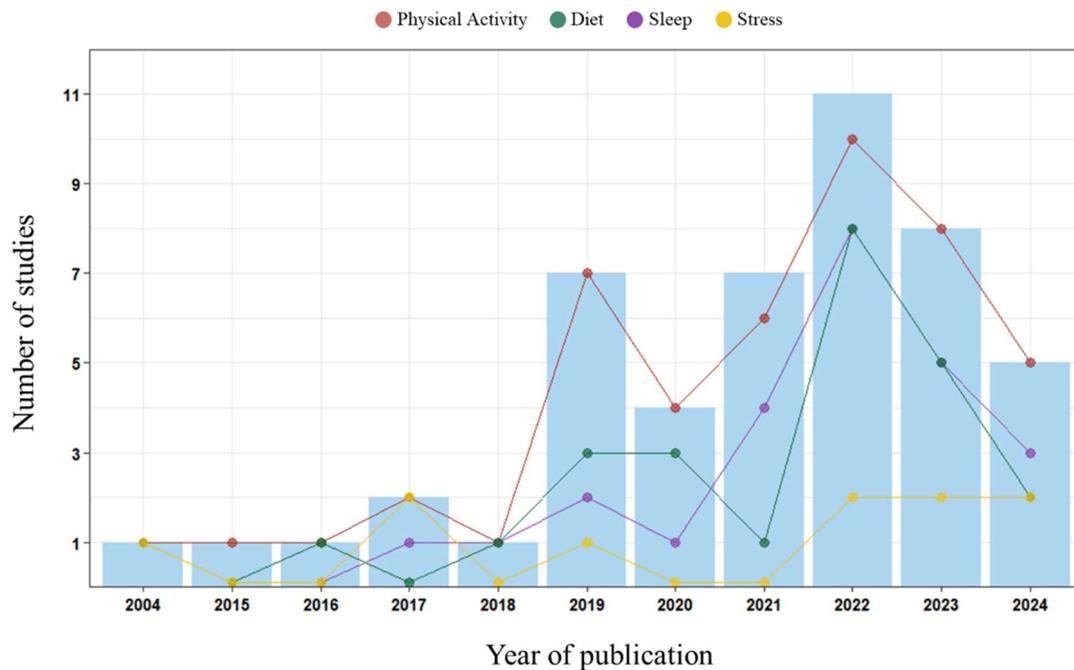
From this point forward, the results section is structured following the ML workflow depicted in Figure 2. Figure 2 illustrates the five key steps in the ML pipeline. The process begins with data acquisition, followed by preprocessing to prepare the data. Then, SL algorithms are applied and evaluated to determine their effectiveness. Finally, explainability techniques are used to understand the models. The dashed lines indicate that modeling, evaluation and explainability can improve earlier stages making the process iterative. Each stage of the process corresponds to a subsection. Finally, we examined the software employed throughout the entire process in the included studies.



[Insert **Figure 2**. Overview of the machine learning workflow, spanning from data acquisition to explainability. The dashed lines represent iterative feedback loops within the process.]

Data acquisition: collection modes, data structures, lifestyle variables and health outcomes

The 48 articles included in this review were published between 2004 and 2024, 42 of them (87.5 %) from 2019, Figure 3 shows the annual productivity distributed by lifestyle components. The studies were carried out in several geographical regions, representing 4 continents (S1 Table 2). The average sample across the included studies was 2,9524.13 participants, with the smallest study having 8 participants and the largest having 470,778 participants.



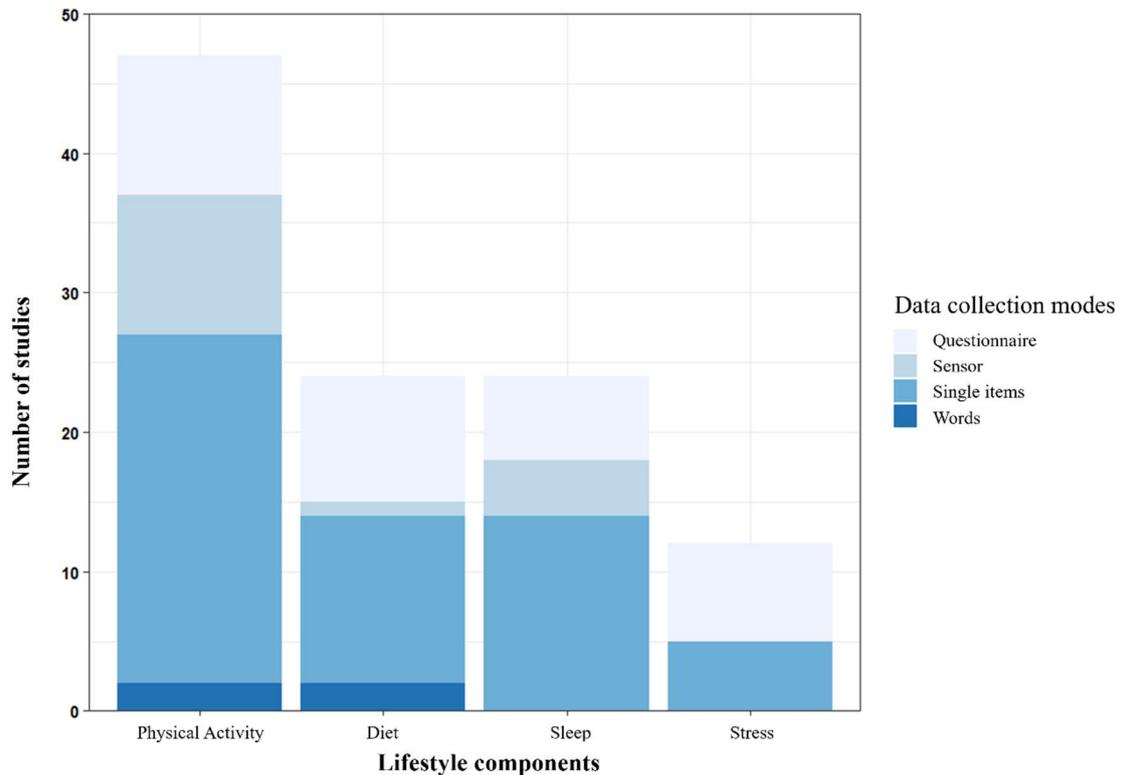
[Insert **Figure 3**. Bar graph showing the productivity over publication year, with each line and dot representing a lifestyle component.]

Four domains of lifestyle were identified in the review: physical activity, diet, sleep, and stress. Most studies, 37 (77%), integrated multi-domain data from more than one of the mentioned lifestyles (see Table 1). The most studied component was physical activity, appearing in 47 out of the 48 included studies. Diet and sleep each appeared in 24 (50%) studies, while stress appeared in 12 (25%) (see Figure 3). Only 11 studies (23%) focused exclusively on one domain (10 on physical activity [32–41], and 1 on sleep [42]).

Regarding data sources, 25 (52.08%) articles relied on self-acquired data, while 23 (47.92%) utilized either private or public health datasets, such as UK Biobank [43]. Among the studies employing self-acquired data, the mean sample size was 1,591.32. Five of them focused exclusively on females [38,39,44–46] and 17 reported a female distribution ranging from 41% to 83%. Studies utilizing health repositories exhibited a greater mean sample size of 61,265.95. Fourteen studies reported a female distribution ranging from 50% to 70%, while 9 did not report the distribution.

Figure 4 summarizes the different data collection modes used for each of the four lifestyle domains identified (S1 Table 3 contains a detailed description of each measure). Single items were utilized for measurement in half of the studies that collected data on physical activity and

diet, and over half of those focused on sleep. Despite the heterogeneity of these items, distinct categories emerged within each component.



[Insert **Figure 4.** Staked bar chard summarizing the data acquisition methodology for each lifestyle component.]

For physical activity, these categories included engagement in physical activities [39,44,47–50], intensity [8,34,41,43,51,52], frequency [35,36,53–65], and environmental factors [66,67]. Within the domain of diet, categories encompassed the frequency of consuming [44,50,65], types of products [55,56,59,62], environmental factors [66], and consumption habits [45,47,51,64]. Regarding sleep studies the categories were, sleep duration [43,47,50,51,57,58,60,62,63,65,66], perceived sleep quality [48], and sleep-related problems [52,59]. Finally, the assessment of stress typically focused on the stress level [8,43,62,64,68].

Questionnaires were used to collect physical activity data in 10 studies. The standardized questionnaires were the *Global Physical Activity Questionnaire* (GPAQ) [65,69], the lifetime total physical activity questionnaire (exercise and sport sub-scale) [70], the *Indian Migration Study Physical Activity Questionnaire* (IMS-PAQ) [71], physical fitness test [33], the *Short Questionnaire to Assess Health-enhancing Physical Activity* (SQUASH) [53,54], the physical activity scale from the active living index [67], the pregnancy physical activity questionnaire [46], and the physical activity scale for the elderly [72]. Nine studies as for diet, the *Food Frequency Questionnaire* (FFQ) [43,53,54,61,67,69,71], the *PrimeScreen* questionnaire [46], and a non-standardized questionnaire consisting of items from different questionnaires [70]. The standardized questionnaires used to measure sleep were the *Pittsburg Sleep Quality Index* (PSQI) [8,46,64,69,72], the *Munich Chronotype* questionnaire and *Sleep Disturbance Scale for Children* [49], and the *Epworth Sleepiness Scale* [72]. To measure stress the specific subscale from the *Depression Anxiety Stress* (DASS) [70], the *INTERHEART STRESS* questionnaire [45], the *Psychosocial Well-being Index-Short Form* [69], the *Perceived Stress Scale* (PSS) [46,49,67],

and the *Profile Of Mood States* (POMS) [60] were used. Regarding data collection through sensors, most studies used wearables, one study used the smartphone to obtain points of interest for physical activity and diet [73], and a sleep study used polysomnography [42]. Finally, two studies used words related to physical activity and diet, one from Google Trends [74], and the other from Twitter [75].

Concerning the inputs modelled, 39 (80.56%) studies utilized multimodal data. The input modalities were lifestyle (43; 89.6%), sociodemographic (32; 66.67%), clinical (20; 41.67%), anthropometric (11; 22.91%), psychological (10; 19.44%), physical (3; 6.25%), environmental (9; 22.92%), physiological (3; 6.25%), and behavioral (2; 4.17%). The model outcomes included lifestyle domains in 12 studies (25%; 5 physical activity [32,36–38,68], 4 sleep [8,40,64,76], 2 diet [47,50], and 1 stress [46]) and other health outcomes in 36 (75%; with mental health, cancer, cardiovascular diseases, and diabetes being the most frequent categories) (see Table 1).

Cross-sectional data were acquired in 23 (48%) studies [34–36,43,44,46,47,50,53,54,57–59,63,65,71], longitudinal data in 14 (29%) [8,33,42,45,49,52,60,62,72], time series data in 7 (15%) [37,38,40,55,56,76,77], longitudinal combined with time series data in 1 [68], textual data in 2 [74,75], and cross-sectional combined with geographical data in 1 [78].

**Table 1.** Summary of the ML workflow from data acquisition to explainability in the included studies.

Study	Physical Activity	Diet	Sleep	Stress	Health outcome/s	Preprocess	ML algorithm	Model Evaluation	Explainability
[65]	Questionnaire (Standardized) and Single items (Frequency)	Single items (Frequency)	Single items (Sleep hours)		Mental health	Missing imputation, Resampling, Dimensionality reduction	<sup>a</sup> RF, <sup>b</sup> ANN, <sup>c</sup> NB, <sup>d</sup> KNN	Train and test	
[44]	Single items (Engagement)	Single items (Frequency)			Cancer	Transformation, Missing imputation, Dimensionality reduction	<sup>e</sup> DT, <sup>f</sup> MLPNN, <sup>g</sup> RBFNN, <sup>h</sup> FNN, <sup>i</sup> PNN, KNN	10-fold cross-validation	
[70]	Questionnaire (Standardized) and Single items (Frequency)	Questionnaire (non-Standardized)		Questionnaire (Standardized)	Alzheimer	Transformation, Missing imputation, Dimensionality reduction	RF, <sup>j</sup> SVM	nested cross-validation	
[66]	Single items (Environment)	Single items (Environment)	Single items (Sleep hours)		Obesity	Missing imputation, Dimensionality reduction	RF, DT	2-fold cross-validation	Local interpretable model-agnostic explanations (LIME)

[45]	Sensor (Wearable)	Single items (Habits)		Questionnaire (Standardized)	Cardiovascular disease	Transformation, Missing imputation, Dimensionality reduction	RF, DT, KNN, NB	Leave-one-out cross-validation (LOOCV)	
[71]	Questionnaire (Standardized)	Questionnaire (Standardized)			Diabetes	Resampling, Dimensionality reduction	RF	Train and test	
[47]	Single items (Engagement)	Single items (Habits)	Single items (Sleep hours)		Lifestyle (Diet)	Transformation, Dimensionality reduction	DT	no reported	
[32]	Sensor (Wearable)				Lifestyle (Physical activity)	Transformation	RF	5-fold cross-validation	
[33]	Questionnaire (Standardized)				Successful aging	Transformation, Dimensionality reduction	RF, <sup>k</sup> GBM, ANN	10-fold cross-validation	
[68]	Sensor (Wearable)			Single items (Stress level)	Lifestyle (Physical activity)	Dimensionality reduction	RF, DT	no reported	
[77]	Sensor (Wearable)		Sensor (wearable)		Blood Pressure	Transformation, Missing imputation, Dimensionality reduction	RF, GBM, MLPNN, <sup>l</sup> LSTM-RNN, SVM	5-fold cross-validation. Online Weighted-Resampling	

[53]	Questionnaire (Standardized) and Single items (Frequency)	Questionnaire (Standardized)			Cancer	Transformation, Missing imputation, Resampling, Dimensionality reduction	RF, SVM	Train and test. Training set 10-fold cross-validation	
[54]	Questionnaire (Standardized) and Single items (Frequency)	Questionnaire (Standardized)			Cancer	Transformation, Resampling, Dimensionality reduction	RF, GBM, SVM	5-fold cross-validation	
[39]	Single items (Engagement)				Cancer	Missing imputation, Dimensionality reduction	RF, DT, XGBoost, ANN	10-fold cross-validation	
[55]	Single items (Frequency)	Single items (Type of products)			Diabetes	Transformation, Missing imputation	LSTM-RNN, ANN, KNN	Train, validation, and test	
[56]	Single items (Frequency)	Single items (Type of products)			Cardiometabolic disease		RF	Leave-one-out cross-validation (LOOCV)	SHAP values
[58]	Single items (Frequency)		Single items (Sleep hours)		Cardiovascular disease	Missing imputation, Dimensionality reduction	RF, XGBoost, BART	5-fold cross-validation	

[57]	Single items (Frequency)		Single items (Sleep hours)		Cardiovascular disease	Missing imputation, Dimensionality reduction	RF, BART	no reported	
[48]	Single items (Engagement)		Single items (Sleep quality)		Cognitive impairment	Missing imputation, Resampling	RF, BoostDT, XGBoost, LSTM-RNN	Train, validation, and test	SHAP values
[69]	Questionnaire (Standardized)	Questionnaire (Standardized)	Questionnaire (Standardized) and Single items (Sleep hours)	Questionnaire (Standardized)	Quality of life	Transformation, Resampling	RF, DT, XGBoost, SVM, NB, KNN	6-fold cross-validation	SHAP values
[79]	Sensor (Wearable)		Sensor (Wearable)		Alzheimer	Dimensionality reduction	SVM	5-fold cross-validation	
[49]	Single items (Engagement)		Questionnaire (Standardized) and Single items (Sleep hours)	Questionnaire (Standardized)	Mental health	Transformation, Missing imputation, Dimensionality reduction	XGBoost	nested cross-validation	SHAP values
[8]	Single items (Intensity)		Questionnaire (Standardized) and Single items (Sleep quality)	Single items (Stress level)	Lifestyle (Sleep)	Transformation, Missing imputation, Dimensionality reduction	RF, DT	Train and test	

[34]	Single items (Intensity)				Osteoarthritis	Transformation, Missing imputation, Resampling, Dimensionality reduction	°FFNN	Train and test	
[43]	Single items (Intensity)	Questionnaire (Standardized)	Single items (Sleep hours)	Single items (Stress level)	Chronic kidney disease	Missing imputation, Dimensionality reduction	GBM	Train and test. Test set 10-fold cross-validation	
[59]	Single items (Frequency)	Single items (Type of products)	Single items (Sleep problems)		Mental health	Missing imputation	XGBoost	no reported	SHAP values
[40]	Sensor (Wearable)				Lifestyle (Sleep)	Transformation	MLPNN	Train, validation and test	
[67]	Questionnaire (Standardized) and Single items (Environment)	Questionnaire (Standardized) and Single items (Environment)		Questionnaire (Standardized)	Cardiovascular disease	Missing imputation	RF, ANN	10-fold cross-validation	SHAP values
[50]	Single items (Engagement)	Single items (Frequency)	Single items (Sleep hours)		Lifestyle (Diet)	Dimensionality reduction	FFNN	Train and test	

[60]	Single items (Frequency)		Single items (Sleep hours)	Questionnaire (Standardized) and Sensor (Wearable)	Fatigue	Transformation, Missing imputation, Dimensionality reduction	RF	10-fold-cross validation	
[46]	Questionnaire (Standardized)	Questionnaire (Standardized)	Questionnaire (Standardized)	Questionnaire (Standardized)	Lifestyle (Stress)	Transformation, Missing imputation, Resampling, Dimensionality reduction	SVM	Train and test	
[74]	Words (Google Trends)	Words (Google Trends)			Obesity	Dimensionality reduction	RF, GBM, SVM	out-of-sample	
[61]	Single items (Frequency)	Questionnaire (Standardized)			Visceral fat	Transformation, Missing imputation, Dimensionality reduction	RF, XGBoost, ANN	Train, validation, and test	SHAP values
[62]	Single items (Frequency)	Single items (Type of products)	Single items (Sleep hours)	Single items (Stress level)	Diabetes	Missing imputation, Resampling	MLPNN	Train and test	
[41]	Single items (Intensity)				Adverse health event	Dimensionality reduction	XGBoost	Train and test	

[52]	Single items (Intensity)		Single items (Sleep problems)		Mortality	Transformation, Missing imputation, Dimensionality reduction	RF	Train and test	
[51]	Single items (Intensity)	Single items (Habits)	Single items (Sleep hours)		Mental health	Missing imputation, Dimensionality reduction	RF, DT, XGBoost, SVM, ANN, KNN	10-fold cross-validation	
[36]	Single items (Frequency)				Diabetes	Missing imputation, Resampling	RF, GBM, AdaBoost	10-fold cross-validation	
[35]	Single items (Frequency)				Lifestyle (Physical activity)	Missing imputation, Resampling	RF, GBM, AdaBoost	10-fold cross-validation	
[76]	Sensor (Wearable)		Sensor (wearable)		Lifestyle (Sleep)	Missing imputation	MLPNN, PCNN, SETRNN, LSTM-RNN	Train, validation and test	
[37]	Sensor (Wearable)				Lifestyle (Physical activity)	Transformation	RF, DT, ANN, SVM	Leave-one-out cross-validation (LOOCV)	
[75]	Words (Twitter)	Words (Twitter)			Inflammatory bowel disease		RF, GBM, AdaBoost, SVM	Train and test. Training set 10-fold cross-validation	

[72]	Questionnaire (Standardized)		Questionnaire (Standardized) and Single items (Sleep hours)		Mortality		RF	no reported	
[42]			Single items (Sleep hours) and Sensor (Polysomnography)		Mortality	Missing imputation, Dimensionality reduction	RF	External dataset	
[63]	Single items (Frequency)		Single items (Sleep hours)		Mental health	Dimensionality reduction	RF	Train and test. Test set 10-fold cross-validation	SHAP values
[64]	Single items (Frequency)	Single items (Habits)	Questionnaire (Standardized)	Single items (Stress level)	Lifestyle (Sleep)	Resampling	RF, DT, XGBoost, SVM, ANN, KNN	External dataset	SHAP values
[38]	Sensor (Wearable)				Lifestyle (Physical activity)	Transformation	SVM	out-of-sample	
[73]	Sensor (Phone)	Sensor (Phone)			Obesity	Missing imputation, Dimensionality reduction	RF, ANN	10-fold cross-validation	

<sup>a</sup>Random Forest (RF)

<sup>b</sup>Artificial Neural Network (ANN)

<sup>c</sup>Naive Bayes (NB)

---

<sup>d</sup>K-Nearest Neighbors (KNN)

<sup>e</sup>Decision Tree (DT)

<sup>f</sup>Multilayer Perceptron Neural Networks (MLPNN)

<sup>g</sup>Radial Basis Function Neural Network (RBFNN)

<sup>h</sup>Fuzzy Neural Networks (FNN)

<sup>i</sup>Probabilistic Neural Networks (PNN)

<sup>j</sup>Support Vector Machine (SVM)

<sup>k</sup>Gradient Boosting Machine (GBM)

<sup>l</sup>Long Short-term Memory (LSTM-RNN),

<sup>m</sup>Extreme Gradient Boosting (XGBoost)

<sup>n</sup>Bayesian Additive Regression Trees (BART)

<sup>o</sup>Feed-Forward Neural Network (FFNN)

<sup>p</sup>Convolutional Neural Network (CNN)

<sup>q</sup>Simple Elman-type Recurrent Neural Network (SETRNN)

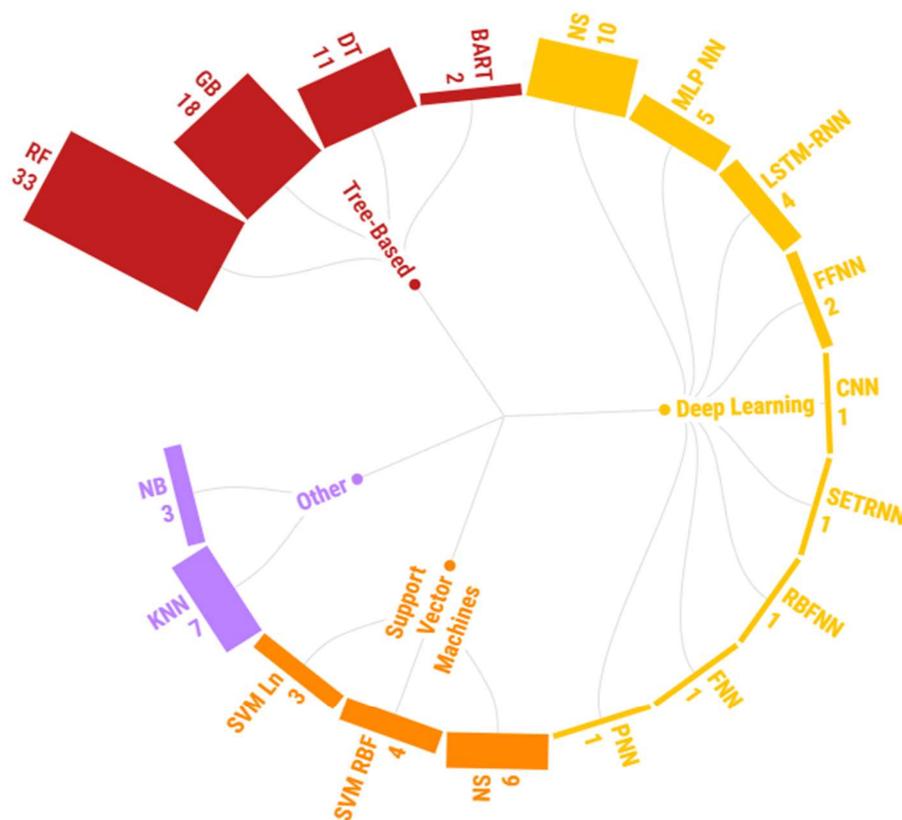
<sup>r</sup>Adaptive Boosting (AdaBoost)

## Families and characteristics of Machine Learning models

In this section, the general components of ML models are described, beginning with problem formulation and algorithm families. Subsequently, preprocessing and evaluation components are mentioned. While these stages are common in the ML workflow, their application is tailored to specific cases; therefore, detailed exploration is beyond the scope of this review.

Depending on the purpose of the ML data analysis, papers were grouped as classification or regression when the objective was prediction, and as feature selection when the goal was explanation [17]. Most papers, 32 (66.67%), were focused on classification, 7 (14.58%) on regression, 2 (4.17%) on both classification and regression, and 7 (14.58%) on feature selection.

Five families of algorithms emerged from the studies: tree-based, deep learning, support vector machines, k-nearest neighbors (KNN), and naïve Bayes. The approach adopted across 28 studies was multimodal, with 20 of them incorporating diverse algorithmic families. In contrast, out of these 28, 8 performed distinct algorithms within the same family. Figure 5 provides a comprehensive taxonomy of the specific algorithms implemented. The following subsections detail the application of specific algorithms within the three major families (tree-based, deep learning and support vector machines) in relation to the type of data used.



[Insert **Figure 5.** Machine learning families and algorithms taxonomy.] Random Forest (RF), Gradient Boosting (GB), Decision Tree (DT), Bayesian Additive Regression Trees (BART),

Multilayer Perceptron Neural Networks (MLP NN), Long Short-term Memory (LSTM-RNN), Feed-Forward Neural Network (FFNN), Convolutional Neural Network (CNN), Simple Elman-type Recurrent Neural Network (SETRNN), Radial Basis Function Neural Network (RBFNN), Fuzzy Neural Networks (FNN), Probabilistic Neural Networks (PNN), Support Vector Machine Radial Basis Function (SVM RBF), Support Vector Machine kernel Linear (SVM Ln), K-Nearest Neighbors (KNN), Naive Bayes (NB), Not Specified (NS).

#### *Tree-based algorithms*

Tree-based algorithms were applied in 39 studies (81.25%), covering all data types. Random forest (RF) was used in 33 (68.75%) out of 39 studies. Specifically, RF was implemented in 16 cross-sectional studies [32,35,36,39,51,53,54,57,58,63–66,69–71], 10 longitudinal studies [8,33,42,45,48,52,60,61,67,72], 3 time series studies [37,56,77], 2 textual [74,75], 1 study with both cross-sectional and geographical data [73], and 1 study with both longitudinal and time series data [68].

Different versions of the gradient boosting (GB) algorithm were performed in 18 studies, including GB machines (GBM), extreme GB (XGBoost), adaptive boosting (AdaBoost), and light GB machine (LGBM) [33,35,36,39,41,43,48,49,51,54,58,59,61,64,69,74,75,77]. Finally, decision-tree algorithms (DT) were implemented in 11 studies [8,37,39,44,45,47,51,64,66,68,69], and Bayesian additive regression trees (BART) in 2 studies [57,58] (Table 1; Figure 4).

#### *Deep Learning algorithms*

Neural Networks (NN) are considered the cornerstone of the deep learning algorithms. Various NN architectures were applied in 18 (37.5%) of the reviewed studies. Multilayer perceptron (MLP NN) was used in cross-sectional [44], longitudinal [62], and time series data [40,76,77]. Long short-term memory (LSTM-RNN) [48,55,76,77] was applied to longitudinal [48] and time series data [55,76,77]. Feed-forward (FFNN) was used for cross-sectional data [34,50]. Convolutional (CNN) and simple Elman-type (SETRNN) were applied in a time series study [76]. Radial basis function (RBFNN), fuzzy (FNN), and probabilistic (PNN) were used in a cross-sectional study [44]. In contrast, 10 studies did not specify the ANN architecture employed [33,37,39,51,55,61,64,65,67,73] (Table 1; Figure 4).

#### *Support Vector Machine algorithms*

Support Vector Machine (SVM) algorithms were utilized in 13 (27.08%) studies, applied across various data types. SVM was implemented in 7 cross-sectional studies, 1 longitudinal study, 3 time series studies, and 2 studies that analyzed textual data. Configurations included radial basis function as the kernel type (SVM RBF) [46,51,77,79] or linear kernel (SVM Ln) [53,70,75]. Six studies did not report the type of kernel used [37,38,54,64,69,74] (Table 1; Figure 4).

#### *Preprocessing*

The preprocessing phase was divided into variable transformation, missing imputation, resampling, and dimensionality reduction. At least one of these preprocessing phases was reported by 43 (89.58%) studies (see Table 1 and S1 Table 4 for more detail).

There were 17 (35.41%) studies that reported normalization or other arithmetic or statistical transformations of variables before the modeling phase [32,33,37,40,44,45,47,49,52–55,60,61,69,70,77]. Six (12.5%) recodified categorical variables into quantitative variables (3 used one-hot encoding [8,49,77] and 1 used principal component analysis with quantile transformer scaler [34]).

Missing imputation was reported in 29 (60.42%) studies. Twelve papers simply removed cases with missing data [8,34–36,46,49,57–59,70,73,76], while others applied cutoff percentages for missing values (e.g., 10% [66], 30% [53] or greater than 50% [44,65]), and one removed

observations with missing values in the output [77]. Techniques included single imputation (mean, median or mode) [44,45,51,61,67], multiple imputation by chained equations (MICE) [43,48,53,65], K Nearest Neighbor (KNN) [77], regression-based algorithm [43,60], random forest multiple imputation [42,52], imputation comparing to the peers in a similar health profile group [62], imputation with training data [79], and replacement of missing value with the last available data [55].

Resampling techniques were reported in 12 (25%) articles. Ten balanced datasets using methods like minority class as a reference, undersampling the majority class [34,46] or the Synthetic Minority Oversampling Technique (SMOTE) [35,36,48,64,65,69,71], and one stabilized variations in underrepresented outcome class using bootstrap [62]. Regarding cancer studies where cases are fewer than controls, two different strategies were applied to the same dataset: sample-size equalization by randomly grouping cancer-free participants based on the number of cancer survivors [54], while another study matched cases and controls by sex, age, and education level, then selected a random sample resulting in 50% cases and 50% controls [53].

Dimensionality reduction was employed in 30 studies (62.5%) through three approaches. The first involved assessing the relationship between features and outcomes by removing redundant information [42,46,49,51,61,63,66,70,73,74]. Other methods included factor analysis [8], and principal components analysis [34,45,52]. The second approach optimized models for lower prediction error [44,50,79]. The third method involved automatic selection of predictors during training [33,41,43,47,53,54,57,58,60,65,68,71,77]. For more information see S1 Table 4.

### Evaluation

The model evaluation procedure was explicitly reported in 43 (89.58%) reviewed studies, with various strategies for cross-validation being the most applied. These included train-test split (also known as “hold-out”) [8,34,41,46,50,52,62,65,71], k-fold cross-validation [32,33,35,36,39,44,51,54,58,60,66,67,69,73,79], nested cross-validation [49,70], leave-one-out cross-validation (for small datasets with  $n < 150$ ) [37,45,56], and dividing the dataset into train, validation, and test for time series data [40,48,55,61,76]. Four studies combined initial train-test split with subsequent k-fold cross-validation on training data [43,53,63,75] (Table 4). Two studies used external datasets for model performance assessment [42,64]. For time series data, 2 studies divided the dataset based on the time of acquisition, keeping an out-of-sample dataset for model evaluation [38,74], while one study [77] utilized 5-fold cross-validation in an offline setting followed by online weighted-resampling methodology to address drift.

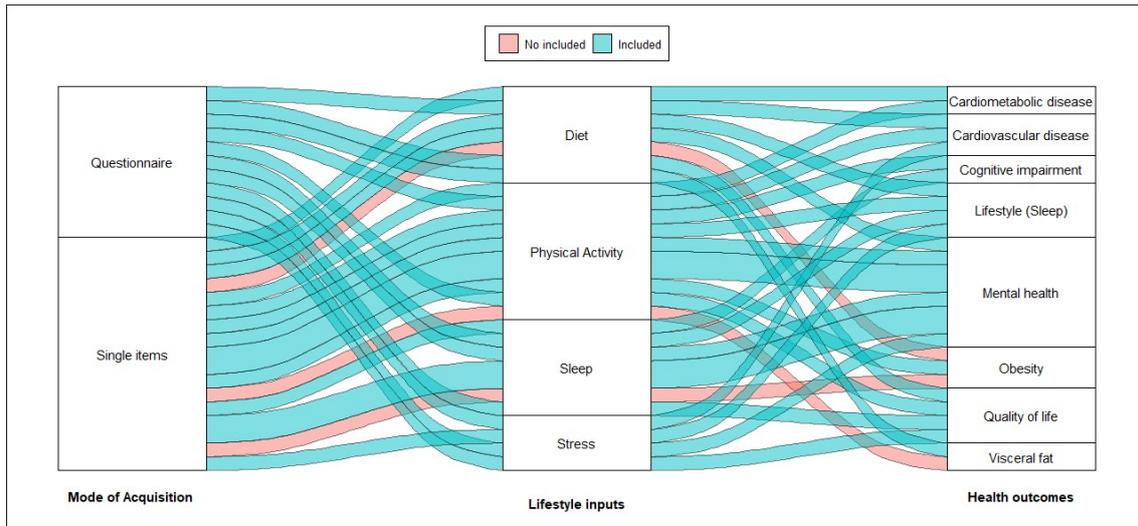
Of the 9 studies addressing regression problem, the evaluation metrics were as follows: mean absolute error (MAE) [36,55,66,77], mean squared error (MSE) [36,55], root mean square error (RMSE) [36,37,58,73,74,77], mean absolute percentage error (MAPE) [77], and coefficient of determination ( $R^2$ ) [36,60,73,74].

For classification problem, the evaluation metrics included specificity, sensitivity (recall) [33–36,39,40,44,48,50,56,62–65,69–71,79], precision and recall [32–35,40,46,51,59,64,75,76], confusion matrix [34,40,59], error rate as the proportion of misclassified observations (1-accuracy) [40,44,65,68], Cohen’s Kappa [44,65], F1-score [33,35,41,45,51,59,64,69,75,76,79], and model training time [44]. The most frequently used metrics were accuracy [8,33–37,39–41,44–46,48,50,51,59,61–65,67,69,70,76], and area under the receiver operating characteristic curve (AUC-ROC) [33–35,38–41,44,45,48,51,53,54,56,61,63–65,69–71,75,76,79] (S1 Table 5).

### Explainability of the results of Machine Learning models

To enhance interpretability, only 10 (20.83%) of the studies implemented specific XAI methods to elucidate the contribution of each predictor and the direction of the relationship with the

outcome (see Table 1). Among these, 9 studies utilized SHapley Additive exPlanation (SHAP) values, while 1 employed Local Interpretable Model-agnostic Explanations (LIME). In all 10 studies, lifestyle variables consistently ranked among the top contributors to the models (see Table 1; Figure 6). Figure 6 shows an alluvial plot that visualizes the application of XAI algorithms to model inputs. The plot is segmented into three boxes: the left box represents the mode of data acquisition, the middle box indicates the lifestyle variables included as model inputs, and the right box depicts the corresponding health outcomes. Arrows or flow lines between these boxes indicate the result of the interpretation algorithm, with red showing instances where lifestyle inputs were not considered significant, and in blue indicating the cases where these inputs were among the top-ranked contributors.



[Insert **Figure 6. Alluvial plot depicting the application of eXplainable Artificial Intelligence (XAI) algorithms to the model inputs (n = 10).**]

In a study of cardiometabolic disease [56], the ML solution was first interpreted for individual participants to provide specific behavioral feedback, and then for all participants to reveal the ranking of features for succeeding in behavioral changes. In both models' interpretations, physical activity and diet variables were among the top contributors. In three mental health studies [49,59,63], SHAP values were used to rank the contribution of each feature. In all three studies, lifestyle variables were among the top contributors. In [59], physical activity and fruit and vegetable consumption; in [49], sleep variables were the top predictors of stress in young adolescents during COVID-19; and in [63], sleep duration was identified as important for predicting depression. Sleep quality and exercise were in the top 20 features for predicting mild cognitive impairment in older adults [48]. For predicting quality of life [69], stress, sleep quality and physical activity emerged as the features with the most predictive power, with the eating index appeared in the top 15 variables. In a study on sleep disturbance [64], the SHAP values ranked stress score, vegetable consumption, and sedentary time among the top 5 variables. In [67], "favorable" food stores and global stress were identified as the top variables for predicting incident of cardiovascular disease, with the availability of outdoor activities ranking in the top 5. In a longitudinal study investigating the association of diet with long-term reduction in waist circumference, SHAP values highlighted the importance of high-quality components in reducing visceral fat [61]. This study also measured exercise with a single item of frequency, which was not included in the top ranked. Finally, physical inactivity emerged as the most important feature in explaining county-level obesity using LIME [66]. In this study, the food environment and insufficient sleep, both measured as single items, were not included as top predictors of obesity prevalence.

## Software to implement Machine Learning models

Of the included articles in the review, 20 (41.67%) used R software [80] for data analysis [33,37–39,41,42,52–54,57–59,65,66,68,70–72,74,76] (Table 3). The R packages used were *data.table* for data manipulation, *tidyverse* [81] as a general package for data science, *MICE* (Multivariate Imputation by Chained Equations) [82] for missing data imputation, *missMDA* for performing multiple imputation with principal component analysis, *FactoMineR* for exploratory data analysis and principal component analysis, *Boruta* [83] for feature selection through a wrapper algorithm, *caret* (Classification And REgression Training) [84] for creating models, *randomForest* for random forest analysis, *randomForestSRC* for random forest for survival, regression, and classification analysis, *rpart* for recursive partitioning and regression trees, *xgboost* for extreme gradient boosting, *bartMachine* for bayesian additive regression trees, *kernlab* and *e1071* for support vector machines, *survival* for survival analysis, *lime* for the local interpretable model-agnostic explanations, and finally, *SuperLearner* [85] to choose the optimal learner for a given prediction problem with a k-fold cross-validation algorithm.

In contrast, 15 (31.25%) studies programmed the models in Python [34,44,46,49,51,56,60,61,64,67,69,73,75,77,79], using the following libraries: *Scikit-learn*, used in all studies for predictive data analysis; *pandas*, for manipulating tabular data; *Numpy*, for mathematical functions; *Keras* and *TensorFlow*, for implementing deep learning; *lightgbm*, for performing light gradient boosting machine; *SHAP*, to explain machine learning solutions; *Crème*, for online machine learning; *bayesian-optimization*, as a global optimization package to find the maximum value of an unknown function in as few iterations as possible; *imbalanced-learn*, to combine either under-sample or over-sample methods; and *TextBlob*, *emoji*, *nlk*, and *profanity*, for processing and analyzing textual data.

Finally, two studies used both Python and R [43,48], and 5 studies used other software programs such as SPSS [47], MATLAB [40,50] and KNIME [35,36]. Six articles did not report the software used [8,32,45,55,62,63].

**Table 2.** Software used in the reviewed studies to perform ML algorithms. ( $n = 31$ ).

Software used	Number of studies	Study references
R	$n = 20$	[33,37–39,41,42,52–54,57–59,65,66,68,70–72,74,76]
Python	$n = 15$	[34,44,46,49,51,56,60,61,64,67,69,73,75,77,79]
R + Python	$n = 2$	[43,48]
SPSS	$n = 1$	[47]
MATLAB	$n = 2$	[40,50]
KNIME	$n = 2$	[35,36]

## Discussion and conclusion

This scoping review of 48 studies provides the current state of the application of supervised ML algorithms to the analysis of lifestyle data, which has been increasing since 2019, indicating a noteworthy area of study. The diversity in the sample origin, along with the accessibility to new AI tools and new methods for monitoring health outcomes, such as wearables, denotes global attention to lifestyle. This section addresses the methodological shortcomings found in the reviewed studies.

### About data acquisition

In relation to lifestyle data, we found that the general approach of studies is multidomain, integrating more than just a single component. This approach enhances and facilitates the understanding of health problems related to four components of people's lifestyles: physical activity, diet, sleep, and stress. The distribution of lifestyle found in this review was similar to that found in a scoping meta-review [86], although we observe that the sleep component has been gaining prominence in recent years, being similar to the distribution of the diet component. This imbalance limits the capacity of the studies to model and understand completely the interaction between the four components of HL and their combined impact on health.

Concerning the data acquisition process, over half of the studies acquired their own data. This acquisition process implies control over variables and reduces the time required for cleaning [87]. Interestingly, both self-acquired datasets and those sources from private or public health repositories demonstrated gender parity in the analyzed datasets. However, we detected a major limitation in this part of the process regarding the methodology of data acquisition. In most studies, data is collected through single items, such as regular physical activity (response "yes" or "no") [44] or usual time of waking up and going to bed [47], resulting in low representativeness of the construct being measured. The result of this acquisition method is a great heterogeneity in measures that hinders their generalizability. Therefore, the quality of data must be one of the challenges to be addressed, and specifically the consistency in measures [88,89].

Nevertheless, the current accessibility and precision of health sensors such as wearables [90] and the Internet of Things (IoT) [91] may contribute in transferability and actionability to the population [92]. The growth in technology allows the integration of different forms of data as well as more objective measure of lifestyle, importantly reducing the impact of retrospective bias by tracking real-time data in an ecological situation [93]. Therefore, merging questionnaire and sensor data may be the key to identifying relationships between lifestyle measurements and personalizing interventions or changes in specific behaviors. This integration would include physiological, psychological and behavioral factors, which are the most common analysis in the machine learning community to extract clinical insights [94].

### About characteristics of Machine Learning models

Regarding the analysis of ML, two different approaches emerged in the reviewed studies: one focused on prediction through classification and regression problems, and the other focused on explanation through feature selection. The first one is already an acknowledged approach, while the second one typically constitutes an important component of the ML process, specifically during the preprocessing stage. However, feature selection studies do not employ model evaluation metrics, which can limit their statistical validity and generalizability of results. Remarkably, the family of ML algorithms most closely related to feature selection is tree-based because it provides indexes of the importance of each variable. Although most papers in this scoping review combined different families of algorithms and compared their results, the most common model family was tree-based which was applied for each data structure identified. Specifically, random forest is the most used algorithm, which may be due to its robustness to handle missing values, the consideration of complex interaction in the data [95], and less sensitivity to variables scales [96]. In this review, DL algorithms were underutilized across the reviewed studies. This result may reflect a gap in expertise or access to computational resources among lifestyle researchers, potentially limiting the application of more complex models. With ongoing advances in computational power and algorithmic efficiency, it is expected that use of DL algorithms will become more widespread in the near future [97].

Regarding the preprocessing stage, most research papers detailed some phases of the process, but there is no consensus on the description of this stage of ML. Variable transformation is a crucial step for certain algorithms, particularly for SVM and specific architectures of DL that exhibit sensitivity to the variable raw form. In this review, 8 out of 13 studies that performed SVM algorithm and 8 out of 18 studies focused on DL reported variable transformation. Additionally, it is worth noting that these algorithms cannot handle missing values, requiring imputation before the modeling phase. Among the SVM studies, 5 out of 13 reported techniques for missing imputation, and 13 out of 18 DL studies explicitly addressed this. In contrast, tree-based algorithms are less sensible to variables scales and missing values, yet incorporating these feature engineering steps could enhance model performance [96]. It is noteworthy that the preprocessing of datasets, specifically how missing values are addressed, has been identified as a potential concern for transparency. This procedural aspect could introduce sampling biases, thereby influencing the generalizability and comprehension of the dataset context [98].

Resampling techniques, aiming to balance the dataset, are commonly implemented in classification problems. SMOTE has been the most widely used technique in this review, especially because it achieves better results than a simple under-sampling of the majority class. In the health domain, imbalanced datasets are common, SMOTE oversample the minority class with synthetic examples and randomly under sample the majority class to balance the dataset [99].

Finally, dimensionality reduction enables capturing the most relevant information for the outcome while eliminating noise and redundant information. In this review, dimensionality reduction was the most frequent preprocessing step, appearing in 30 studies. Notably, not only SVM and DL models could benefit from removing irrelevant predictors, but also in Tree-Based algorithms, dimensionality reduction minimize model complexity, resource consumption and data acquisition costs [96].

The division of the original dataset is an essential step for assessing the performance of the ML solution. In this review, five studies did not report how they split their data to assess the model, which denotes a lack of generalization of their results. Train and test division, also known as “hold out”, is a method with considerable variability due to the use of a unique random data distribution. Therefore, other methods might be more suitable. For example, leave-one-out cross-validation, which trains the model in  $n-1$  observations and makes predictions in the remaining one. Although effective for small datasets, it is computational intensive with large datasets [100]. K-fold cross-validation involves randomly division the original dataset into  $k$  groups. K-fold cross-validation not only offers computational advantage over leave-one-out, but also gives more accurate estimations due to the bias-variance trade-off [100]. In time series data, only one paper [77] introduced a different form of data split, considering the dependencies of the entire series. Ideally, this type of data should be treated with a method called rolling forecast origin resampling, which estimates the model with historical data and evaluate it with the most recent [101]. In other words, the training set should ideally comprise by observations that occurred prior to those in the test set, which we did not find in this review.

Regarding the evaluation of ML models, the choice of metrics depends on the nature of the problem, whether regression or classification. Specific evaluation metrics tailored to each problem are crucial for correct evaluation, aligning with the priorities and needs of each field. For instance, in medical studies where the cost of treatment in terms of health is high, it becomes crucial to identify true patients over false positives. In contrast, if the treatment has minimal side effects and has demonstrated benefits, sensitivity might not be as important as specificity. It is worth highlight that in field of data science, precision and recall are more

commonly used, whereas in medical fields, specificity and sensitivity are more prevalent [16]. These differences may cause misunderstandings between the two domains.

#### About the XAI of the machine learning results

When it comes to interpreting models, only 10 studies performed an interpretation algorithm (9 SHAP values and 1 LIME) as a specific step in the ML process. In this review SHAP values and LIME were the only XAI methods applied to lifestyle data, these two methods were the most common in a recent systematic review on XAI methods [102]. In our review only 10 studies used XAI methods representing a relatively small number of studies, considering that interpretability is crucial for gaining insights and taking actions based on predictions. Although tree-based algorithms, especially decision tree, are known for facilitating interpretation, SHAP values can be applied to any type of model [103]. The adoption of XAI in lifestyle studies remains low. One possible explanation for this is that explainability algorithms are often not integrated into standard ML pipeline, thereby increasing the technical complexity of the workflow. However, some efforts are being made by R and Python developers to incorporate XAI algorithms into pipelines using libraries such as H2O [104].

In this review, lifestyle components (physical activity, diet, sleep and stress) were among the top-ranked features in all the studies that interpreted their results. These findings are consistent with previous research emphasizing the integration of diverse lifestyle components [105,106]. However, the level of interpretability achieved also depends on the quality of data used in the models, which in some studies did not meet expected standard. For instance, as illustrated in Figure 6, the diet component, expectedly linked to obesity, was not consistently among the top-ranked features [66]. In this county-level study, a food environment index was measured as single items, potentially inadequately representing the diet component of lifestyle.

Therefore, integrating interpretability into the ML process could enable tailored interventions based on model results, provided that measures are collected accurately. This approach can also facilitate the translation of results to stakeholders and health systems, thereby enhancing decision-making.

#### About the software for implementing machine learning models

The competition between Python and R for ML software dominance in data science is currently intense. Both Python and R, are freely distributed, object-oriented software with large and active communities. Python, as a programming language, offers specific implementations through libraries tailored for statistical analysis, including, ML and DL. In contrast, R is a statistical software that integrates fundamental statistics into its base functionalities. While Python requires libraries for each stage of analysis, its well-established libraries streamline the process. On the other hand, R faces challenges due to its heterogeneous libraries, which hinder replicability and expertise gained to varying syntax across packages. To address this concern, the meta-package *tidymodels* resolves these issues by integrating all necessary packages for each ML step, employing a unified syntax. Additionally, *tidymodels* integrate user-friendly interfaces and promote good methodological practice, thereby preventing user errors [107].

#### Methodological and reporting guidelines

In the basis of the review's results and to enhance transparency and replicability in multidisciplinary sciences [108], we provide comprehensive methodological and reporting guidelines for ML projects. Although various studies have proposed guidelines and checklists [109,110], the rapid expansion of ML algorithms in health domains necessitates iterative evaluation to incorporate new steps into the ML research workflow. The guideline is based on the five stages of the ML workflow, as depicted in Figure 2, with added software tools.

##### a) *Data acquisition:*

The integration of multidomain data enhances the comprehension of real-world problems. Using appropriate methods to collect data ensures representativeness. We recommend the use of standardized questionnaires and validated sensors. Regarding health repositories, we recommend providing information about data characteristics such as gender distribution, sample size, and variable description [111].

b) *Preprocessing:*

Reporting the preprocessing methods used in the data analysis is particularly crucial for ensuring replicability. While preprocessing contributes to improving data quality, different preprocessing methods can lead to different results. We propose the following recommendations for each preprocessing step, although not all steps need to be performed in every ML project.

- *Transformation:* Categorical data should be encoded using methods such as one-hot encoding and dummy variables. Continuous data should be transformed using normalization or rescaling features with different units to ensure algorithm performance, particularly for those sensitive to raw form of variables [96].
- *Missing imputation:* Some algorithms cannot handle missing data and require imputation prior to modelling. Depending on the number of observations and the data distribution, imputation with mean, median or mode is typical. For time series data, imputation with the last or next observation is preferred, though rolling statistics imputation or interpolation may offer better solutions [112].
- *Resampling:* Imbalanced datasets can bias models, resulting in poor performance on underrepresented classes [113]. While the synthetic minority oversampling technique (SMOTE) is an effective technique for handling imbalanced datasets, it is not without its limitations. However, when the class imbalance ratio is extremely high SMOTE can potentially bias the model performance by overfitting the minority class. This issue is particularly pronounced in datasets containing noise, as synthetic observations may replicate these artefacts. To mitigate this challenges, recent studies have proposed tree algorithms which have shown effectiveness in handling class imbalance [114]. Furthermore, due to their robustness the use of tree-based algorithms is increasingly recommended when working with class imbalance [115].
- *Dimensionality reduction:* Removing noise from the dataset and retaining features directly related to the outcome can enhance both data acquisition and modelling efficiency. Removing correlated features is particularly beneficial [96].

c) *Modeling:*

Our guidelines focused on SL algorithms for classification and regression problems. The choice of algorithms depends on the measurement of the outcome. We recommend using multiple algorithms to compare results and select the best fit given that there is significant variability across problems [116]. Additionally, comparing different families of algorithms is also advisable, as some improvements exist within the same family. When it comes to replicability and transparency, reporting algorithm hyperparameters is crucial in ML problems, as different configurations can yield varying results [117].

d) *Evaluation:*

Avoiding overfitting requires dividing the original dataset appropriately. This division depends on data typology; we identified three different data structures sensible to this stage based on whether time is an implicit factor in the data acquisition process.

- Cross-sectional: Data is collected at a single point in time, with no temporal dependencies. Leave-one-out cross-validation is effective for small datasets ( $n < 150$ ) typically seen in life science, otherwise the computational complexity increases [118]. K-fold cross-validation is recommended for large datasets to balance the bias-variance trade-off [100]. Cross-validation with bootstraps resampling can also be used to evaluate the performance of the models and estimating confidence intervals of performance metrics [119].
- Longitudinal: In longitudinal studies researchers collect repeated measures, potentially with dependencies between observations and high correlation that can bias the model [107]. Data should be divided by grouping individual participant's information. The methods used are similar to those for cross-sectional data but with consideration of partitioning.
- Time series: Time series is a sequence of data points in chronological order. Rolling forecast origin resampling is suitable for this data [101]. The training set should include observations occurring before those in the test set.
- Evaluation metrics: Choose appropriate metrics based on ML problem (regression or classification) [120] and the study field [16]. Additionally, it is recommended to compute evaluation metrics repeatedly across the cross-validation samples and use non-parametric tests, such as Wilcoxon signed-rank test and Friedman's test, to evaluate the performance of each ML model [120].

#### e) *Explainability*

Reporting the method of explainability used in the ML projects is essential. It is important to distinguish between interpretability and explainability. While explainability refers to understanding the effect of each feature on the original model, interpretability involves deriving actionable insights from the model's prediction. While tree-based methods include importance metric, they do not indicate the direction of relationships with the outcome. Incorporating explainable resources like SHAP values [121] or LIME [122] enhances the interpretability of the results, providing both actionability and transparency, and transforming black-boxes into glass boxes models. The H2O package in R offers XAI algorithms that are compatible with the framework tidymodels, enabling a unified workflow for modelling and explainability. For a more detailed taxonomy of XAI packages in R, we refer the reader to [123].

#### f) *Software*

R is the most widely used software in our review. However, the variety of libraries can complicate the process. To address this concern, the metapackage *tidymodels* unifies syntax, enhancing replicability [107].

### Limitations

This review has identified several limitations within the studies that need addressing. Firstly, ensuring data quality remains a significant challenge that must be addressed by carefully selecting data acquisition methods to build reliable and robust models. Secondly, the evaluation process is crucial for preventing overfitting, and using hold-out cross-validation can lead to high variance partitioning. Therefore, it is recommended to implement k-fold cross-validation at various stages, such as during validation; for time series data, rolling forecast origin resampling is recommended [101].

However, this work is not without limitations. The term "stress" was not included in the search strategy syntax since it represents a specific field of study. Stress can both influence other lifestyle components and be influenced by them. However, the results underscore the

importance of including stress as a lifestyle component alongside physical activity, diet, and sleep. Secondly, the exclusion of UL algorithms limits the scope of ML algorithms applied to lifestyle data. Consequently, this scoping review covers one part of the studies employing ML algorithm. Future research is needed to explore the current application of UL in the context of lifestyle data.

In conclusion, our study provides a comprehensive analysis of lifestyle using ML models and serves as a guideline for future research. While the relationship between lifestyle and health is well-established, ongoing efforts are needed to refine how we measure lifestyle to create robust models. It is essential to focus not only on model performance but also on data representativeness, which is closely related to the granularity established during data collection. Although random forest algorithms are prominent in lifestyle data analysis, it is recommended to compare its performance with other algorithms within and across families. Future research should also incorporate SHAP values to enhance interpretability within the ML workflow. Additionally, the *tidymodels* metapackage (R software) with H2O for XAI can assist researchers in evaluating process quality with unified syntax, thereby contributing to the replicability.

### Acknowledgments

This work was supported from Grants PID2019-107473RB-C21 and Grant PID2022-141403NB-I00 funded by MCIN/AEI/10.13039/501100011033/FEDER, UE, and 2021SGR-00806, funded by the Catalanian Government.

### Author contribution

TE: Conceptualization, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Visualization. LC: Conceptualization, Writing - Review & Editing, Supervision, Project administration, Funding acquisition. CA: Validation, Writing - Review & Editing, Supervision. JML: Conceptualization, Methodology, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

### Conflicts of Interests

The authors declare no competing interests.

### Data availability

The data generated in this study are included in the manuscript and S1. Furthermore replication data were published in CORA Repositori de Dades de Recerca [29] (<https://dataverse.csuc.cat/dataset.xhtml?persistentId=doi:10.34810/data1088>).

### Abbreviations

RF: Random Forest

ANN: Artificial Neural Network

NB: Naive Bayes

KNN: K-Nearest Neighbors

DT: Decision Tree

MLPNN: Multilayer Perceptron Neural Networks

RBFNN: Radial Basis Function Neural Network

FNN: Fuzzy Neural Networks

PNN: Probabilistic Neural Networks

SVM: Support Vector Machine  
GBM: Gradient Boosting Machine  
LSTM-RNN: Long Short-term Memory  
XGBoost: Extreme Gradient Boosting  
BART: Bayesian Additive Regression Trees  
FFNN: Feed-Forward Neural Network  
CNN: Convolutional Neural Network  
SETRNN: Simple Elman-type Recurrent Neural Network  
AdaBoost: Adaptive Boosting

## References

1. Furihata R, Konno C, Suzuki M, Takahashi S, Kaneita Y, Ohida T, Uchiyama M. Unhealthy lifestyle factors and depressive symptoms: A Japanese general adult population survey. *Journal of Affective Disorders* 2018 Jul;234:156–161. doi: 10.1016/j.jad.2018.02.093
2. Nudelman G, Shiloh S. Connectionism and Behavioral Clusters: Differential Patterns in Predicting Expectations to Engage in Health Behaviors. *Annals of Behavioral Medicine* 2018 Sep 13;52(10):890–901. doi: 10.1093/abm/kax063
3. Nudelman G, Shiloh S. Mapping health behaviors: Constructing and validating a common-sense taxonomy of health behaviors. *Social Science & Medicine* 2015 Dec;146:1–10. doi: 10.1016/j.socscimed.2015.10.004
4. Kaminsky LA, German C, Imboden M, Ozemek C, Peterman JE, Brubaker PH. The importance of healthy lifestyle behaviors in the prevention of cardiovascular disease. *Progress in Cardiovascular Diseases* 2022 Jan;70:8–15. doi: 10.1016/j.pcad.2021.12.001
5. Braun C, Foreyt JP, Johnston CA. Stress: A Core Lifestyle Issue. *American Journal of Lifestyle Medicine* 2016 Jul;10(4):235–238. doi: 10.1177/1559827616642400
6. Cerf ME. Healthy lifestyles and noncommunicable diseases: Nutrition, the life-course, and health promotion. *Lifestyle Medicine* 2021 Apr;2(2):e31. doi: 10.1002/lim2.31
7. Gurrin C, Smeaton AF, Doherty AR. LifeLogging: Personal Big Data. *FNT in Information Retrieval* 2014;8(1):1–125. doi: 10.1561/15000000033
8. Lim J, Jeong CY, Lim JM, Chung S, Kim G, Noh KJ, Jeong H. Assessing Sleep Quality Using Mobile EMAs: Opportunities, Practical Consideration, and Challenges. *IEEE Access* 2022;10:2063–2076. doi: 10.1109/ACCESS.2021.3140074
9. Kline A, Wang H, Li Y, Dennis S, Hutch M, Xu Z, Wang F, Cheng F, Luo Y. Multimodal machine learning in precision health: A scoping review. *npj Digit Med* 2022 Nov 7;5(1):171. doi: 10.1038/s41746-022-00712-8
10. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. *JAMA* 2018 Apr 3;319(13):1317. doi: 10.1001/jama.2017.18391

11. Maleki F, Ovens K, Najafian K, Forghani B, Reinhold C, Forghani R. Overview of Machine Learning Part 1. *Neuroimaging Clinics of North America* 2020 Nov;30(4):e17–e32. doi: 10.1016/j.nic.2020.08.007
12. Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P. The role of artificial intelligence in healthcare: a structured literature review. *BMC Med Inform Decis Mak* 2021 Dec;21(1):125. doi: 10.1186/s12911-021-01488-9
13. Breiman L. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statist Sci* 2001 Aug 1;16(3). doi: 10.1214/ss/1009213726
14. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to Machine Learning, Neural Networks, and Deep Learning. *Transl Vis Sci Technol* 2020 Feb 27;9(2):14. PMID:32704420
15. Efron B. Prediction, Estimation, and Attribution. *Int Statistical Rev* 2020 Dec;88(S1). doi: 10.1111/insr.12409
16. Bruce PC, Bruce A, Gedeck P. *Estadística práctica para ciencia de datos con R y Python: más de 50 conceptos esenciales*. 2ª ed. Barcelona: Marcombo; 2022. ISBN:978-84-267-3443-3
17. James G, Witten D, Hastie T, Tibshirani R. *Statistical Learning. An Introduction to Statistical Learning* New York, NY: Springer US; 2021. p. 15–57. doi: 10.1007/978-1-0716-1418-1\_2 ISBN:978-1-07-161417-4
18. Saqib K, Khan AF, Butt ZA. Machine Learning Methods for Predicting Postpartum Depression: Scoping Review. *JMIR Ment Health* 2021 Nov 24;8(11):e29838. doi: 10.2196/29838
19. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. *Psychol Med* 2019 Jul;49(09):1426–1448. doi: 10.1017/S0033291719000151
20. Ono S, Goto T. Introduction to supervised machine learning in clinical epidemiology. *ACE* 2022;4(3):63–71. doi: 10.37737/ace.22009
21. Bi Q, Goodman KE, Kaminsky J, Lessler J. What is Machine Learning? A Primer for the Epidemiologist. *American Journal of Epidemiology* 2019 Oct 21;kwz189. doi: 10.1093/aje/kwz189
22. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial Intelligence–Based Chatbots for Promoting Health Behavioral Changes: Systematic Review. *J Med Internet Res* 2023 Feb 24;25:e40789. doi: 10.2196/40789
23. Goh YS, Ow Yong JQY, Chee BQH, Kuek JHL, Ho CSH. Machine Learning in Health Promotion and Behavioral Change: Scoping Review. *J Med Internet Res* 2022 Jun 2;24(6):e35831. doi: 10.2196/35831
24. Lai X, Chen J, Lai Y, Huang S, Cai Y, Sun Z, Wang X, Pan K, Gao Q, Huang C. Using Large Language Models to Enhance Exercise Recommendations and Physical Activity in Clinical and Healthy Populations: Scoping Review. *JMIR Med Inform* 2025 May 27;13:e59309–e59309. doi: 10.2196/59309

25. Mersha M, Lam K, Wood J, AlShami A, Kalita J. Explainable Artificial Intelligence: A Survey of Needs, Techniques, Applications, and Future Direction. arXiv; 2024; doi: 10.48550/ARXIV.2409.00265
26. Ali S, Abuhmed T, El-Sappagh S, Muhammad K, Alonso-Moral JM, Confalonieri R, Guidotti R, Del Ser J, Díaz-Rodríguez N, Herrera F. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion* 2023 Nov;99:101805. doi: 10.1016/j.inffus.2023.101805
27. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol* 2018 Dec;18(1):143. doi: 10.1186/s12874-018-0611-x
28. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Moher D, Peters MDJ, Horsley T, Weeks L, Hempel S, Akl EA, Chang C, McGowan J, Stewart L, Hartling L, Aldcroft A, Wilson MG, Garritty C, Lewin S, Godfrey CM, Macdonald MT, Langlois EV, Soares-Weiser K, Moriarty J, Clifford T, Tunçalp Ö, Straus SE. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med* 2018 Oct 2;169(7):467–473. doi: 10.7326/M18-0850
29. Estrella T, Capdevila L, Alfonso C, Losilla J-M. Replication data for Machine Learning for the analysis of healthy lifestyle data: a scoping review. [object Object]; 2024. doi: 10.34810/DATA1088
30. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *Journal of Clinical Epidemiology* 2016 Jul;75:40–46. doi: 10.1016/j.jclinepi.2016.01.021
31. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, Koffel JB, PRISMA-S Group, Blunt H, Brigham T, Chang S, Clark J, Conway A, Couban R, de Kock S, Farrah K, Fehrman P, Foster M, Fowler SA, Glanville J, Harris E, Hoffecker L, Isojarvi J, Kaunelis D, Ket H, Levay P, Lyon J, McGowan J, Murad MH, Nicholson J, Pannabecker V, Paynter R, Pinotti R, Ross-White A, Sampson M, Shields T, Stevens A, Sutton A, Weinfurter E, Wright K, Young S. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Syst Rev* 2021 Jan 26;10(1):39. doi: 10.1186/s13643-020-01542-z
32. Butkevičiūtė E, Bikulčienė L, Žvironienė A, Butkeviciute E, Bikulciene L, Zvioniene A, Butkevičiūtė E, Bikulčienė L, Žvironienė A, Butkeviciute E, Bikulciene L, Zvioniene A. Physiological State Evaluation in Working Environment Using Expert System and Random Forest Machine Learning Algorithm. *HEALTHCARE* 2023 Jan;11(2). PMID:36673588
33. Cai T, Long J, Kuang J, You F, Zou T, Wu L. Applying machine learning methods to develop a successful aging maintenance prediction model based on physical fitness tests. *Geriatrics & Gerontology International* 2020 Jun;20(6):637–642. PMID:32358851
34. Lim J, Kim J, Cheon S. A Deep Neural Network-Based Method for Early Detection of Osteoarthritis Using Statistical Data. *International journal of environmental research and public health* 2019 Apr;16(7). PMID:30974803

35. Recenti M, Ricciardi C, Edmunds KJ, Gislason MK, Sigurdsson S, Carraro U, Gargiulo P. Healthy Aging Within an Image: Using Muscle Radiodensitometry and Lifestyle Factors to Predict Diabetes and Hypertension. *IEEE journal of biomedical and health informatics* 2021 Jun;25(6):2103–2112. PMID:33306475
36. Recenti M, Ricciardi C, Edmunds K, Jacob D, Gambacorta M, Gargiulo P. Testing soft tissue radiodensity parameters interplay with age and self-reported physical activity. *European journal of translational myology* 2021 Jul;31(3). PMID:34251162
37. Staudenmayer J, He S, Hickey A, Sasaki J, Freedson P. Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements. *Journal of applied physiology (Bethesda, Md : 1985)* 2015 Aug;119(4):396–403. PMID:26112238
38. Zhou M, Fukuoka Y, Goldberg K, Vittinghoff E, Aswani A. Applying machine learning to predict future adherence to physical activity programs. *BMC medical informatics and decision making* 2019 Aug;19(1):169. PMID:31438926
39. Dianati-Nasab M, Salimifard K, Mohammadi R, Saadatmand S, Fararouei M, Hosseini KS, Jiavid-Sharifi B, Chausalet T, Dehdar S. Machine learning algorithms to uncover risk factors of breast cancer: insights from a large case-control study. *FRONTIERS IN ONCOLOGY* 2024;13:1276232. PMID:38425674
40. Matta SC, Sankari Z, Rihana S. Heart rate variability analysis using neural network models for automatic detection of lifestyle activities. *BIOMEDICAL SIGNAL PROCESSING AND CONTROL* 2018;42:145–157. doi: 10.1016/j.bspc.2018.01.016
41. Park H, Jung SY, Han MK, Jang Y, Moon YR, Kim T, Soo-Yong S, Hwang H. Lowering Barriers to Health Risk Assessments in Promoting Personalized Health Management. *Journal of Personalized Medicine* KakaoHealthCare Corp., Seongnam-si 13529, Gyeonggi-do, Republic of Korea; hpark.park@kakahealthcare.com (H.P.); saylor.moon@kakahealthcare.com (Y.R.M.); ray0601@snu.ac.kr (T.K.); sean.shin@kakahealthcare.com (S.-Y.S.); drhwang.spike@kakahealthcare.com: MDPI AG PP - Basel; 2024;14(3):316. doi: <https://doi.org/10.3390/jpm14030316>
42. Wallace ML, Coleman TS, Mentch LK, Buysse DJ, Graves JL, Hagen EW, Hall MH, Stone KL, Redline S, Peppard PE. Physiological sleep measures predict time to 15-year mortality in community adults: Application of a novel machine learning framework. *Journal of sleep research* 2021 Dec;30(6):e13386. PMID:33991144
43. Luo W, Gong L, Chen X, Gao R, Peng B, Wang Y, Luo T, Yang Y, Kang B, Peng C, Ma L, Mei M, Liu Z, Li Q, Yang S, Wang Z, Hu J. Lifestyle and chronic kidney disease: A machine learning modeling study. *Frontiers in nutrition* 2022;9:918576. PMID:35938107
44. Afrash MR, Bayani A, Shanbehzadeh M, Bahadori M, Kazemi-Arpanahi H. Developing the breast cancer risk prediction system using hybrid machine learning algorithms. *Journal of education and health promotion* 2022;11:272. PMID:36325225
45. Alshurafa N, Sideris C, Pourhomayoun M, Kalantarian H, Sarrafzadeh M, Eastwood J-A. Remote Health Monitoring Outcome Success Prediction Using Baseline and First Month Intervention Data. *IEEE Journal of Biomedical and Health Informatics* 2017 Mar;21(2):507–514. PMID:26780823

46. Nichols ES, Pathak HS, Bgeginski R, Mottola MF, Giroux I, Van Lieshout RJ, Mohsenzadeh Y, Duerden EG. Machine learning-based predictive modeling of resilience to stressors in pregnant women during COVID-19: A prospective cohort study. *PLoS one* 2022;17(8):e0272862. PMID:35951588
47. Bôto JM, Marreiros A, Diogo P, Pinto E, Mateus MP, Boto JM, Marreiros A, Diogo P, Pinto E, Mateus MP, Bôto JM, Marreiros A, Diogo P, Pinto E, Mateus MP. Health behaviours as predictors of the Mediterranean diet adherence: a decision tree approach. *PUBLIC HEALTH NUTRITION* 2022 Aug;25(7):1864–1876. PMID:34369348
48. Huang YC, Huang ZS, Yang QR, Jin HJ, Xu TK, Fu YT, Zhu Y, Zhang XY, Chen C. Predicting mild cognitive impairment among Chinese older adults: a longitudinal study based on long short-term memory networks and machine learning. *FRONTIERS IN AGING NEUROSCIENCE* 2023;15. doi: 10.3389/fnagi.2023.1283243
49. Kiss O, Alzueta E, Yuksel D, Pohl KM, de Zambotti M, Müller-Oehring EM, Prouty D, Durley I, Pelham WE 3rd, McCabe CJ, Gonzalez MR, Brown SA, Wade NE, Marshall AT, Sowell ER, Breslin FJ, Lisdahl KM, Dick AS, Sheth CS, McCandliss BD, Guillaume M, Van Rinsveld AM, Dowling GJ, Tapert SF, Baker FC. The Pandemic's Toll on Young Adolescents: Prevention and Intervention Targets to Preserve Their Mental Health. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine* 2022 Mar;70(3):387–395. PMID:35090817
50. Mousavi H, Karandish M, Jamshidnezhad A, Hadianfard AM. Determining the effective factors in predicting diet adherence using an intelligent model. *Scientific reports* 2022 Jul;12(1):12340. PMID:35853992
51. Qasrawi R, Polo SV, Abu Khader R, Abu Al-Halawa D, Hallaq S, Abu Halaweh N, Abdeen Z. Machine learning techniques for identifying mental health risk factor associated with schoolchildren cognitive ability living in politically violent environments. *FRONTIERS IN PSYCHIATRY* 2023;14. doi: 10.3389/fpsyt.2023.1071622
52. Puterman E, Weiss J, Hives BA, Gemmill A, Karasek D, Mendes WB, Rehkopf DH. Predicting mortality from 57 economic, behavioral, social, and psychological factors. *Proceedings of the National Academy of Sciences of the United States of America* 2020 Jul;117(28):16273–16282. PMID:32571904
53. Cortés-Ibañez FO, Nagaraj SB, Cornelissen L, Navis GJ, van der Vegt B, Sidorenkov G, de Bock GH, Cohort P, Sidorenkov G, Bock GHD. Prediction of Incident Cancers in the Lifelines Population-Based Cohort. *Cancers* 2021 Apr;13(9):2133. doi: 10.3390/cancers13092133
54. Cortés-Ibañez FO, Nagaraj SB, Cornelissen L, Sidorenkov G, de Bock GH, Bock GHD. A Classification Approach for Cancer Survivors from Those Cancer-Free, Based on Health Behaviors: Analysis of the Lifelines Cohort. *CANCERS* 2021;13(10). doi: 10.3390/cancers13102335
55. Faruqui SHA, Du Y, Meka R, Alaeddini A, Li C, Shirinkam S, Wang J. Development of a Deep Learning Model for Dynamic Forecasting of Blood Glucose Level for Type 2 Diabetes Mellitus: Secondary Analysis of a Randomized Controlled Trial. *JMIR mHealth and uHealth* 2019 Nov;7(11):e14452. PMID:31682586

56. Guthrie NL, Carpenter J, Edwards KL, Appelbaum KJ, Dey S, Eisenberg DM, Katz DL, Berman MA. Emergence of digital biomarkers to predict and modify treatment efficacy: machine learning study. *BMJ open* 2019 Jul;9(7):e030710. PMID:31337662
57. Hu L, Liu B, Li Y. Ranking sociodemographic, health behavior, prevention, and environmental factors in predicting neighborhood cardiovascular health: A Bayesian machine learning approach. *Preventive medicine* 2020 Dec;141:106240. PMID:32860821
58. Hu L, Liu B, Ji J, Li Y. Tree-Based Machine Learning to Identify and Understand Major Determinants for Stroke at the Neighborhood Level. *Journal of the American Heart Association* 2020 Nov;9(22):e016745. PMID:33140687
59. Majcherek D, Kowalski AM, Lewandowska MS. Lifestyle, Demographic and Socio-Economic Determinants of Mental Health Disorders of Employees in the European Countries. *International Journal of Environmental Research and Public Health* 2022 Sep;19(19):11913. PMID:36231214
60. Mun E-Y, Geng F. Predicting post-experiment fatigue among healthy young adults: Random forest regression analysis. *Psychological test and assessment modeling* 2019 Nov;61(4):471–493. PMID:32038903
61. Park S. Association of a High Healthy Eating Index Diet with Long-Term Visceral Fat Loss in a Large Longitudinal Study. *Nutrients* 2024 Feb;16(4). PMID:38398858
62. Park J, Edington DW. Application of a prediction model for identification of individuals at diabetic risk. *Methods of information in medicine* 2004;43(3):273–281. PMID:15227557
63. Xin Y, Ren X. Predicting depression among rural and urban disabled elderly in China using a random forest classifier. *BMC psychiatry* 2022 Feb;22(1):118. PMID:35168579
64. Zhang L, Zhao S, Yang Z, Zheng H, Lei M. An Artificial Intelligence Platform to Stratify the Risk of Experiencing Sleep Disturbance in University Students After Analyzing Psychological Health, Lifestyle, and Sports: A Multicenter Externally Validated Study. *Psychology research and behavior management* 2024;17:1057–1071. PMID:38505352
65. Abdul Rahman H, Kwicklis M, Ottom M, Amornsriwatanakul A, H. Abdul-Mumin K, Rosenberg M, Dinov ID. Machine Learning-Based Prediction of Mental Well-Being Using Health Behavior Data from University Students. *Bioengineering* 2023 May 10;10(5):575. doi: 10.3390/bioengineering10050575
66. Allen B. An interpretable machine learning model of cross-sectional US county-level obesity prevalence using explainable artificial intelligence. *PLOS ONE* 2023;18(10). doi: 10.1371/journal.pone.0292341
67. Morris MC, Moradi H, Aslani M, Sims M, Schlundt D, Kouros CD, Goodin B, Lim C, Kinney K. Predicting incident cardiovascular disease among African-American adults: A deep learning approach to evaluate social determinants of health in the Jackson heart study. *PLOS ONE* 2023;18(11). doi: 10.1371/journal.pone.0294050
68. Cheung YK, Hsueh P-YSYS, Qian M, Yoon S, Meli L, Diaz KM, Schwartz JE, Kronish IM, Davidson KW. Are Nomothetic or Ideographic Approaches Superior in Predicting Daily

Exercise Behaviors? Analyzing N-of-1 mHealth Data. *METHODS OF INFORMATION IN MEDICINE* 2017 Feb;56(06):452–460. PMID:29582914

69. Kim J, Jeong K, Lee S, Baek Y. Machine-learning model predicting quality of life using multifaceted lifestyles in middle-aged South Korean adults: a cross-sectional study. *BMC public health* 2024 Jan;24(1):159. PMID:38212741
70. Ai M, Morris TP, Zhang J, de la Colina AN, Tremblay-Mercier J, Villeneuve S, Whitfield-Gabrieli S, Kramer AF, Geddes MR, Aisen P, Anthal E, Appleby M, Bellec P, Benbouhoud F, Bohbot V, Brandt J, Breitner JCS, Brunelle C, Chakravarty M, Cheewakriengkrai L, Collins L, Couture D, Craft S, Dadar M, Daoust L-A, Das S, Dauar-Tedeschi M, Dea D, Desrochers N, Dubuc S, Duclair G, Dufour M, Eisenberg M, El-Khoury R, Etienne P, Evans A, Faubert A-M, Ferdinand F, Fonov V, Fontaine D, Francoeur R, Frenette J, Gagné G, Gauthier S, Gervais V, Giles R, Gonneaud J, Gordon R, Greco C, Hoge R, Hudon L, Ituria-Medina Y, Kat J, Kazazian C, Kligman S, Kostopoulos P, Labonté A, Lafaille-Magnan M-E, Lee T, Leoutsakos J-M, Leppert I, Madjar C, Mahar L, Maltais J-R, Mathieu A, Mathotaarachchi S, Mayrand G, McSweeney M, Meyer P-F, Michaud D, Miron J, Morris JC, Multhaupt G, Münter L-M, Nair V, Near J, Newbold-Fox H, Nilsson N, Pagé V, Pascoal TA, Petkova M, Picard C, Binette AP, Pogossova G, Poirier J, Rajah N, Remz J, Rioux P, Rosa-Neto P, Sager MA, Saint-Fort EF, Savard M, Soucy J-P, Sperling RA, Spreng N, St-Onge F, Tardif C, Thérroux L, Thomas RG, Toussaint P-J, Tuwaig M, Vachon-Preseau E, Vallée I, Venugopalan V, Wan K, Wang S. Resting-state MRI functional connectivity as a neural correlate of multidomain lifestyle adherence in older adults at risk for Alzheimer’s disease. *Scientific Reports* (Nature Publisher Group) Northeastern University, Department of Psychology, Boston, USA (GRID:grid.261112.7) (ISNI:0000 0001 2173 3359) ; Northeastern University, Department of Physical Therapy, Movement and Rehabilitation Sciences, Boston, USA (GRID:grid.261112.7) (ISNI:0000 000: Nature Publishing Group PP - London; 2023;13(1):7487. doi: <https://doi.org/10.1038/s41598-023-32714-1>
71. Birk N, Matsuzaki M, Fung TT, Li Y, Batis C, Stampfer MJ, Deitchler M, Willett WC, Fawzi WW, Bromage S, Kinra S, Bhupathiraju SN, Lake E. Exploration of Machine Learning and Statistical Techniques in Development of a Low-Cost Screening Method Featuring the Global Diet Quality Score for Detecting Prediabetes in Rural India. *The Journal of Nutrition* 2021 Oct;151(12 Suppl 2):110S-118S. PMID:34689190
72. Wallace ML, Buysse DJ, Redline S, Stone KL, Ensrud K, Leng Y, Ancoli-Israel S, Hall MH. Multidimensional Sleep and Mortality in Older Adults: A Machine-Learning Comparison With Other Risk Factors. *The journals of gerontology Series A, Biological sciences and medical sciences* 2019 Nov;74(12):1903–1909. PMID:30778527
73. Zhou RZ, Hu Y, Tirabassi JN, Ma Y, Xu Z. Deriving neighborhood-level diet and physical activity measurements from anonymized mobile phone location data for enhancing obesity estimation. *International journal of health geographics* 2022 Dec;21(1):22. PMID:36585658
74. Oladeji O, Zhang C, Moradi T, Tarapore D, Stokes AC, Marivate V, Sengeh MD, Nsoesie EO. Monitoring Information-Seeking Patterns and Obesity Prevalence in Africa With Internet Search Data: Observational Study. *JMIR public health and surveillance* 2021 Apr;7(4):e24348. PMID:33913815

75. Stemmer M, Parmet Y, Ravid G. Identifying Patients With Inflammatory Bowel Disease on Twitter and Learning From Their Personal Experience: Retrospective Cohort Study. *Journal of medical Internet research* 2022 Aug;24(8):e29186. PMID:35917151
76. Sathyanarayana A, Joty S, Fernandez-Luque L, Ofli F, Srivastava J, Elmagarmid A, Arora T, Taheri S. Sleep Quality Prediction From Wearable Data Using Deep Learning. *JMIR Mhealth Uhealth* 2016 Nov 4;4(4):e125. doi: 10.2196/mhealth.6562
77. Chiang P-H, Dey S. Offline and Online Learning Techniques for Personalized Blood Pressure Prediction and Health Behavior Recommendations. *IEEE Access* 2019;7:130854–130864. doi: 10.1109/ACCESS.2019.2939218
78. Zhong J, Liu W, Niu B, Lin X, Deng Y. Role of Built Environments on Physical Activity and Health Promotion: A Review and Policy Insights. *Front Public Health* 2022 Jul 12;10:950348. doi: 10.3389/fpubh.2022.950348
79. Kimura N, Aota T, Aso Y, Yabuuchi K, Sasaki K, Masuda T, Eguchi A, Maeda Y, Aoshima K, Matsubara E. Predicting positron emission tomography brain amyloid positivity using interpretable machine learning models with wearable sensor data and lifestyle factors. *ALZHEIMERS RESEARCH & THERAPY* 2023;15(1). doi: 10.1186/s13195-023-01363-x
80. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2022. Available from: <https://www.R-project.org/>
81. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. Welcome to the Tidyverse. *JOSS* 2019 Nov 21;4(43):1686. doi: 10.21105/joss.01686
82. van Buuren S, Groothuis-Oudshoorn K. **mice** : Multivariate Imputation by Chained Equations in R. *J Stat Soft* 2011;45(3). doi: 10.18637/jss.v045.i03
83. Kurska MB, Rudnicki WR. Feature Selection with the **Boruta** Package. *J Stat Soft* 2010;36(11). doi: 10.18637/jss.v036.i11
84. Kuhn M. Building Predictive Models in R Using the **caret** Package. *J Stat Soft* 2008;28(5). doi: 10.18637/jss.v028.i05
85. Van Der Laan MJ, Polley EC, Hubbard AE. Super Learner. *Statistical Applications in Genetics and Molecular Biology* 2007 Jan 16;6(1). doi: 10.2202/1544-6115.1309
86. Castro R, Ribeiro-Alves M, Oliveira C, Romero CP, Perazzo H, Simjanoski M, Kapczynski F, Balanzá-Martínez V, De Boni RB. What Are We Measuring When We Evaluate Digital Interventions for Improving Lifestyle? A Scoping Meta-Review. *Front Public Health* 2022 Jan 3;9:735624. doi: 10.3389/fpubh.2021.735624
87. Pasquetto IV, Borgman CL, Wofford MF. Uses and Reuses of Scientific Data: The Data Creators' Advantage. *Harvard Data Science Review* 2019 Nov 15;1(2). doi: 10.1162/99608f92.fc14bf2d

88. Aldoseri A, Al-Khalifa KN, Hamouda AM. Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges. *Applied Sciences* 2023 Jun 13;13(12):7082. doi: 10.3390/app13127082
89. Kim K-H, Oh SW, Ko SJ, Lee KH, Choi W, Choi IY. Healthcare data quality assessment for improving the quality of the Korea Biobank Network. Meri A, editor. *PLoS ONE* 2023 Nov 20;18(11):e0294554. doi: 10.1371/journal.pone.0294554
90. Iqbal SMA, Mahgoub I, Du E, Leavitt MA, Asghar W. Advances in healthcare wearable devices. *npj Flex Electron* 2021 Apr 12;5(1):9. doi: 10.1038/s41528-021-00107-x
91. Mohanta B, Das P, Patnaik S. Healthcare 5.0: A Paradigm Shift in Digital Healthcare System Using Artificial Intelligence, IOT and 5G Communication. 2019 International Conference on Applied Machine Learning (ICAML) Bhubaneswar, India: IEEE; 2019. p. 191–196. doi: 10.1109/ICAML48257.2019.00044
92. Kamel Boulos MN, Koh K. Smart city lifestyle sensing, big data, geo-analytics and intelligence for smarter public health decision-making in overweight, obesity and type 2 diabetes prevention: the research we should be doing. *Int J Health Geogr* 2021 Dec;20(1):12, s12942-021-00266-0. doi: 10.1186/s12942-021-00266-0
93. Perez-Pozuelo I, Zhai B, Palotti J, Mall R, Aupetit M, Garcia-Gomez JM, Taheri S, Guan Y, Fernandez-Luque L. The future of sleep health: a data-driven revolution in sleep science and medicine. *npj Digit Med* 2020 Mar 23;3(1):42. doi: 10.1038/s41746-020-0244-4
94. Reinertsen E, Clifford GD. A review of physiological and behavioral monitoring with digital sensors for neuropsychiatric illnesses. *Physiol Meas* 2018 May 15;39(5):05TR01. doi: 10.1088/1361-6579/aabf64
95. Breiman L. Random Forests. *Machine Learning* 2001;45(1):5–32. doi: 10.1023/A:1010933404324
96. Kuhn M, Johnson K. Feature engineering and selection: a practical approach for predictive models. Boca Raton London New York: CRC Press, Taylor & Francis Group; 2020. ISBN:978-1-138-07922-9
97. Thompson NC, Ge S, Manso GF. The Importance of (Exponentially More) Computing Power. *arXiv*; 2022. doi: 10.48550/ARXIV.2206.14007
98. Arora A, Alderman JE, Palmer J, Ganapathi S, Laws E, McCradden MD, Oakden-Rayner L, Pfohl SR, Ghassemi M, McKay F, Treanor D, Rostamzadeh N, Mateen B, Gath J, Adebajo AO, Kuku S, Matin R, Heller K, Sapey E, Sebire NJ, Cole-Lewis H, Calvert M, Denniston A, Liu X. The value of standards for health datasets in artificial intelligence-based applications. *Nat Med* 2023 Nov;29(11):2929–2938. doi: 10.1038/s41591-023-02608-w
99. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. *jair* 2002 Jun 1;16:321–357. doi: 10.1613/jair.953
100. James G, Witten D, Hastie T, Tibshirani R. Resampling Methods. An Introduction to Statistical Learning New York, NY: Springer US; 2021. p. 197–223. doi: 10.1007/978-1-0716-1418-1\_5 ISBN:978-1-07-161417-4

101. Hyndman R, Athanasopoulos G. *Forecasting: Principles and Practice*. 2nd ed. OTexts; 2018. Available from: <https://otexts.org/fpp2/>
102. Saarela M, Podgorelec V. Recent Applications of Explainable AI (XAI): A Systematic Literature Review. *Applied Sciences* 2024 Oct 2;14(19):8884. doi: 10.3390/app14198884
103. Rodríguez-Pérez R, Bajorath J. Interpretation of Compound Activity Predictions from Complex Machine Learning Models Using Local Approximations and Shapley Values. *J Med Chem* 2020 Aug 27;63(16):8761–8777. doi: 10.1021/acs.jmedchem.9b01101
104. H2O.ai. H2O: Scalable machine learning platform. 2024. Available from: <https://docs.h2o.ai/h2o/latest-stable/h2o-docs/index.html>
105. Hosker DK, Elkins RM, Potter MP. Promoting Mental Health and Wellness in Youth Through Physical Activity, Nutrition, and Sleep. *Child and Adolescent Psychiatric Clinics of North America* 2019 Apr;28(2):171–193. doi: 10.1016/j.chc.2018.11.010
106. Kris-Etherton PM, Sapp PA, Riley TM, Davis KM, Hart T, Lawler O. The Dynamic Interplay of Healthy Lifestyle Behaviors for Cardiovascular Health. *Curr Atheroscler Rep* 2022 Dec;24(12):969–980. doi: 10.1007/s11883-022-01068-w
107. Kuhn M, Silge J. *Tidy Modeling with R: a framework for modeling in the Tidyverse*. First edition. Beijing Boston Farnham Sebastopol Tokyo: O'Reilly; 2022. ISBN:978-1-4920-9648-1
108. Han S, Olonisakin TF, Pribis JP, Zupetic J, Yoon JH, Holleran KM, Jeong K, Shaikh N, Rubio DM, Lee JS. A checklist is associated with increased quality of reporting preclinical biomedical research: A systematic review. Boltze J, editor. *PLoS ONE* 2017 Sep 13;12(9):e0183591. doi: 10.1371/journal.pone.0183591
109. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, Shilton A, Yearwood J, Dimitrova N, Ho TB, Venkatesh S, Berk M. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. *J Med Internet Res* 2016 Dec 16;18(12):e323. doi: 10.2196/jmir.5870
110. Al-Zaiti SS, Alghwiri AA, Hu X, Clermont G, Peace A, Macfarlane P, Bond R. A clinician's guide to understanding and critically appraising machine learning studies: a checklist for Ruling Out Bias Using Standard Tools in Machine Learning (ROBUST-ML). *European Heart Journal - Digital Health* 2022 Jul 6;3(2):125–140. doi: 10.1093/ehjdh/ztac016
111. Stevens GA, Alkema L, Black RE, Boerma JT, Collins GS, Ezzati M, Grove JT, Hogan DR, Hogan MC, Horton R, Lawn JE, Marušić A, Mathers CD, Murray CJL, Rudan I, Salomon JA, Simpson PJ, Vos T, Welch V. Guidelines for Accurate and Transparent Health Estimates Reporting: the GATHER statement. *The Lancet* 2016 Dec;388(10062):e19–e23. doi: 10.1016/S0140-6736(16)30388-9
112. Moritz S, Sardá A, Bartz-Beielstein T, Zaefferer M, Stork J. Comparison of different Methods for Univariate Time Series Imputation in R. *arXiv*; 2015. doi: 10.48550/ARXIV.1510.03924
113. Mohsen F, Al-Absi HRH, Yousri NA, El Hajj N, Shah Z. A scoping review of artificial intelligence-based methods for diabetes risk prediction. *npj Digit Med* 2023 Oct 25;6(1):197. doi: 10.1038/s41746-023-00933-5

114. Safi SK, Gul S. An Enhanced Tree Ensemble for Classification in the Presence of Extreme Class Imbalance. *Mathematics* 2024 Oct 16;12(20):3243. doi: 10.3390/math12203243
115. Velarde G, Weichert M, Deshmunkh A, Deshmane S, Sudhir A, Sharma K, Joshi V. Tree boosting methods for balanced and imbalanced classification and their robustness over time in risk assessment. *Intelligent Systems with Applications* 2024 Jun;22:200354. doi: 10.1016/j.iswa.2024.200354
116. Caruana R, Niculescu-Mizil A. An empirical comparison of supervised learning algorithms. *Proceedings of the 23rd international conference on Machine learning - ICML '06 Pittsburgh, Pennsylvania: ACM Press; 2006.* p. 161–168. doi: 10.1145/1143844.1143865
117. Arnold C, Biedebach L, Küpfer A, Neunhoeffler M. The role of hyperparameters in machine learning models and how to tune them. *PSRM* 2024 Feb 5;1–8. doi: 10.1017/psrm.2023.61
118. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. *arXiv; 2009*; doi: 10.48550/ARXIV.0907.4728
119. Tsamardinos I, Greasidou E, Borboudakis G. Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation. *Mach Learn* 2018 Dec;107(12):1895–1922. doi: 10.1007/s10994-018-5714-4
120. Rainio O, Teuvo J, Klén R. Evaluation metrics and statistical tests for machine learning. *Sci Rep* 2024 Mar 13;14(1):6086. doi: 10.1038/s41598-024-56706-x
121. Lundberg S, Lee S-I. A Unified Approach to Interpreting Model Predictions. *arXiv; 2017*. Available from: <http://arxiv.org/abs/1705.07874> [accessed Apr 22, 2024]
122. Ribeiro MT, Singh S, Guestrin C. Model-Agnostic Interpretability of Machine Learning. *arXiv; 2016*. doi: 10.48550/ARXIV.1606.05386
123. Maksymiuk S, Gosiewska A, Biecek P. Landscape of R packages for eXplainable Artificial Intelligence. *arXiv; 2020*. doi: 10.48550/ARXIV.2009.13248

## Supplementary Files

**S1 Table 1. Search strategy.**

**S1 Table 2. Continents and Countries.**

**S1 Table 3. Data acquisition methodology for each lifestyle component and health outcomes across the included studies.**

**S1 Table 4. Preprocessing.**

**S1 Table 5. Evaluation metrics**

**S2 PRISMA 2020 checklist**