APPENDIX A

Table 1. Summary of reviewed papers on ML-based adaptive and personalized systems for health and wellness.

Ref.	Domain	Target Audience	Modality	Development Process	Model & Performance	Adaptation Approach
(Zhou et al., 2020b)	Medical diagnosis	General audience	Single Modal	Bag-of-words using TF-IDF; Deep learning for extracting sentence-level features; TruncatedSVD for feature selection; Multiclass Classification; 174760 patients, 164692 questions, 1300 kinds of drugs	CRNN (accuracy = 88.63%)	Using the model, the recommender system guides a patient to the appropriate healthcare department according to his/her question or suggests the names of the drugs according to the patient's symptoms.
(Auffenberg et al., 2019)	Disease management (Prostate cancer)	Male Adults	Single Modal	Feature selection using permutation method; Multiclass Classification; Derivation cohort (n=5016), Validation cohort (n=2527); Train/Test split	RF (AUC = 0.81)	System determines the probability of a patient receiving a given primary treatment (e.g., radical prostatectomy, radiation therapy, etc.), and then offers personalized treatment decisions based on similar patients from the clinical registry.
(W. Yang et al., 2020)	Mental health (Anxiety)	College Students	Single Modal	Limit feature set (dimensionality reduction); Vocabulary size; Classification (Character Sequence); There are 10213 training and 2553 testing instances in the Dailydialog dataset; 87146 training and 21788 testing instances in the RedditCausal dataset, and 164 training and 116 testing instances in the Depression dataset	Hybrid-VHRED (BLEU=51.12%)	*Mental Mentor* chatbot leverages the model to generate tailored responses including guided exercises and study tips based on utterances and emotions of users (students).
(Martín et al., 2011)	Assistive healthcare (Human fall detection)	Older Adults	Single Modal	Classification	DT (MQE=0.16)	Using the model, multiagent system detects the occurrence of a fall and automatically places an emergency call to the closest medical center based on user location/profile. The system also sends SMS to the user's contact person.
(Abd et al., 2017)	Disease management (Sickle cell disease (SCD))	SCD Patients	Single Modal	Binary Classification	LogitBoost (accuracy = 99.5984%)	System classifies patients into those with sickle cell trait and those without using the model. For patients with sickle cell trait, the system further determines if the situation is critical or not; if critical, the patient

						receives personalized recommendations and treatment directly, otherwise the system contacts a physician for indirect support.
(J. Yang et al., 2020)	Disease management (Heart disease)	Patients with Heart Disease	Single Modal	Regression; Train/Test split (75/25)	LSTM (validation loss = 0.0054)	*HEARTlistener* system employs the model to predict patient's health condition based on breath rate and heartbeat. System automatically triggers first-aid strategy (e.g., contacting an emergency doctor online, requesting an ambulance) if heart attack is imminent.
(Forman, Goldstein, Crochiere, et al., 2019)	Disease management (Obesity)	Overweight/obese Adults	Single Modal	Classification	Ensemble model of four algorithms (sensitivity = 69.2%, specificity = 83.8%)	*OnTrack* app leverages the model to predict the risk of dietary lapse and then deliver tailored interventions when risk is elevated.
(Cheerla & Gevaert, 2017)	Medical diagnosis (Cancer)	General audience	Single Modal	Data imputation, thresholding, Normalization; correlation-based feature selection algorithm, recursive feature elimination algorithm; SMOTE for oversampling; Multiclass Classification (21 cancer types); 5229 samples from 21 different tumor and normal tissues; 10-fold cross-validation (repeated 10 times)	SVM for diagnosis (accuracy = 97.2%); SVM for prognosis (accuracy = 85%)	System classifies different cancer types the diagnosis model, and then with the help of a prognosis model recommends three personalized treatment regimens to each patient.
(Akbulut et al., 2018)	Medical diagnosis (Fetal health)	Pregnant Women	Single Modal	Tree-based feature selection algorithm; Binary Classification (healthy and anomaly); 10-fold cross-validation; Train/Test split (80/20); Hyperparameter tuning	DF (accuracy = 89.5%)	System leverage model to predict fetal anomaly status, and then recommend suitable physical activities (and a schedule) to perform during pregnancy depending on the predicted status.
(Gorbonos et al., 2018)	Dietary management (Recipe recommendation)	General audience	Single Modal	Regression	ANN (accuracy = 62.63%); NMF (result not reported)	The *NutRec* system uses the models to detect interactions between ingredients and their proportions within recipes with the aim of finding recipes that best fit a set of ingredients and follow healthy eating guidelines for the purpose of offering appropriate recommendations.
(Rachakonda et al., 2020)	Dietary management (Food monitoring)	General audience	Single Modal	Image segmentation, Feature selection using Principal Component Analysis (PCA) technique; Multiclass Classification (130	MobileNet (accuracy = 98%)	*iLog* system first uses the model to detect, classify, and quantify food objects on the user's plate and then use this quantification to determine user's eating behaviour (normal-eating and stress-eating). The

				classes); 1000 images; Train/Test split (80/20)		system provides personalized suggestions on when, what, and how much to eat if stress-eating (i.e., uncontrollable cravings) is confirmed.
(Spanakis et al., 2017)	Dietary management (Eating behaviour assessment)	Adults	Single Modal	Classification + hierarchical agglomerative clustering; 47 participants	DT (increase in healthy eating from 8 users in first four weeks of intervention to 20 in the last 2 weeks)	*ThinkSlim* app utilized model to extract significant rules that indicate what combinations of states are predictive of unhealthy or healthy eating. Based on these rules, the application warns individual users prior to a probable unhealthy eating event.
(Xu et al., 2019)	Dietary management (Diet assessment)	Children	Single Modal	Image segmentation, Derived attributes; Regression	LR (Precision > 80%)	System employs model to estimate food weight and nutrient load, and then provides personalized dietary recommendations to the user based on the estimations.
(P. Chiang & Dey, 2019)	Health monitoring (Blood pressure)	Adults	Multimodal	Removal of missing values, Data imputation; RF was used for feature selection after ranking the features; Regression; 5-fold cross-validation; Train/Test split (80/20)	RFFS (MAE = 5.18 and 4.30 for Systolic BP and Diastolic BP respectively)	Framework uses model to predict systolic blood pressure (BP) and diastolic BP of individual users. Based on this prediction, user receives personalized health behaviour recommendations such as increasing exercise or going to bed earlier.
(W. Gu, 2017b)	Health monitoring (Blood glucose)	Diabetic patients	Multimodal	Multiclass Classification; Train/Test (80/20)	Multi-RNN (accuracy = 82.14%)	*BGMonitor* system utilizes model to detect abnormal blood glucose (BG) events, and then reminds the user to double-check their BG using a clinical continuous glucose monitoring (CGM) device or a finger pricking method.
(Zeevi et al., 2015)	Health monitoring (Postprandial glycemic response)	Adults	Multimodal	Regression; leave-one-out cross validation; Training set, Validation set	GB ($r \leq 0.80$)	Using the mode, the framework predicts individual's postprandial glycemic response (PPGR) to real-life meals, and then offers personalized dietary suggestions based on the PPGR.
(Nosakhare & Picard, 2020)	Health monitoring	College Students	Multimodal	Binning, bag-of-words + one-hot, filtering, Normalization; Regression; 5-fold cross validation	sLDA (**Stressed-Calm**: F1-score = 72%; **Sad-Happy**: F1-score = 68%; **Sick-Healthy**: F1-score = 68%)	Recommender system utilized the model to predict three categories of students' well-being – stress, mood, and physical health. Based on the predicted state for each category, students receive tailored behavioural suggestions that are both achievable and that might improve a future outcome, such as changing the bedtime, length of sleep, and social interactions planned for next day.
(Scherzer et al., 2020)	Substance use (Opoid-use	Patients with OUDs	Single Modal	Derived attributes; Binary and Multiclass classification; Train/Test split	Model is based on natural language processing (NLP) technique (F1-	The *Marigold* app uses model to predict whether a given peer chat or message from user needs moderator intervention (in

	disorder (OUD))				score ≥ 85% for binary classification task, and F1-score ≈ 70% for multiclass)	cases of self-harm, harm to others, or risk of relapse) or not, and also predicts the severity of the message. The app flags contents that are suicidal or homicidal which, in turn, are automatically sent to a clinician for review, while other message types are sent to a moderator. The system also indicates if the user needs intensive care resources.
(Bae et al., 2018)	Substance use (Alcohol-use disorder)	Young Adults	Multimodal	Sliding window; Correlation and Information Gain feature selection techniques; SMOTE for oversampling; Multiclass classification; Cross-Validation; Train/Test/Validation split - 60/20/20	RF (AUC = 96%)	System utilized model to classify time periods as *non-drinking*, *low-risk*, and *high-risk drinking* with the aim of triggering just-in-time behavioural interventions once a high-drinking is detected (such as delivering supportive messaging or contacting supportive friends or family members).
(T. Chen et al., 2018)	Smoking cessation	Adults	Multimodal	Dimensionality reduction; Multiclass Classification	LSTM (F1-score = 86.3%)	System uses model to detect smoking motion (out of six possible motions) in real-time. If smoking motion is detected, the system sends an alert message that includes a quit plan to the user and subscribers (such as doctors or family members).
(Priyadarshi & Saha, 2020b)	Remedy finding (homoeopathy)	General audience	Single Modal	Spelling correction, Bag-of-words+TF-IDF, Feature selection using information gain, mutual information, and chi-square test; Binary and Multiclass Classification; 10-fold cross-validation	SVM 1 (accuracy = 99.04%); SVM 2 (accuracy = 93.42%)	System utilizes models to analyze a patient's question and determine whether the question is seeking remedy and part of five shortlisted diseases. If confirmed, the system forms relevant queries to search the web for homeopathy remedy (medicines) related to the disease and then recommends top-ranked medicine name to the user.
(Abdo et al., 2020)	Patient location privacy	General audience	Multimodal	Multiclass Classification; 9 cases for up to 100 users; 4500 samples for training, 500 samples for testing; 10-fold cross-validation	DT (accuracy = 95.4%)	The framework leverages the model to classify users' health state into *urgency*, *illness*, or *healthy*. In case of urgency, user's location is sent accurately to the nearest medical centre for immediate dispatch of an ambulance. If predicted state is *healthy*, the user's location will not be shared. However, if state is *illness*, user's location is perturbed and obfuscated based on the privacy level preferred by the user/patient.
(Edara et al., 2020)	Health apps recommendation	General audience	Single Modal	Preprocessing using NLTK; Multiclass Classification	DNN-LSTM (F1-score = 97.9%)	The recommender system leverages the hybrid model to recommend mobile health (mHealth)

						apps with positive reviews to individuals based on their health condition.
(Tuti et al., 2020b)	Clinician knowledge representation (emergency neonatal care)	Healthcare Providers	Single Modal	Sliding window; Classification; 697 participants; Train/Test split - 80/20; 50% of randomly selected samples from the training dataset was used as the validation dataset	LSTM (accuracy ≤ 88.32%)	To provide personalized training on emergency neonatal care, the system employs model to predict learners' future performance and forgetting curves based on sequence embeddings of learning task attempts from specific healthcare providers. Based on this prediction, the system provides timely interventions that support self-regulated personalized learning at scale.
(Stark & Samarah, 2019)	Dental and oral health	Adults	Multimodal	Feature selection using correlation analysis, ANOVA, Principal Component Analysis (PCA); Multiclass Classification; 2000 samples; Train/Test split - 70/30; Hyperparameter tuning (e.g., number of trees)	RF (accuracy = 99.71% and 99.63% for **left-handed** and **right-handed** users respectively)	Smart toothbrush leverages model to detect the tooth and surface brushed, and then provides comprehensive and real-time feedback (via an app) to individual users, informing them about their brushing behaviours including suggestions to improve tooth brushing. Users are also reminded to floss and clean their tongue as part of good oral hygiene.
(Sowah et al., 2020)	Disease management (Diabetes)	Type-2 diabetic patients	Single Modal	Filtering (removing ambiguous and duplicate samples); Multiclass Classification (25 classes); 300 images; Train/Validation/Test split (80/10/10)	CNN (based on Inception network) – accuracy = 95%; KNN (result not reported)	System applies CNN model to recognize food and KNN to suggest the right meals for breakfast, lunch, and dinner based on a patient's profile.
(Sarwar & Javed, 2019)	Assistive healthcare (Human activity recognition)	Adults	Multimodal	Sliding windows, trimming; Multiclass Classification; Running samples (n=7000) + Walking samples (n=11700) + Sleeping samples (n=10000); cross-validation	NB (accuracy = 90%)	System utilizes model to recognize the current activity a user is performing (such as sleeping, walking, etc.), and then suggests an optimal care plan for health improvements based on the activity.
(Dijkhuis et al., 2018)	Physical activity	Workers/ Employees	Single Modal	Transformation (reformatting), Removal of missing values, Derived attributes, Normalization; Binary Classification; Train/Test split (70/30); 5-fold cross-validation; grid search	RF (accuracy = 93%)	System utilized the model to estimate the probability of users meeting their individual daily physical activity goals on an hourly basis. Based on the estimation, individual users receive personalized feedback throughout the day to help them to achieve their goal.
(Kadri et al., 2020)	Physical activity	Adults	Single Modal	Multiclass Classification	DT (F1-score = 62%); BiLSTM (F1-score = 59%)	Model predicts physical activity behaviour of users (jogging, sitting, standing, upstairs, downstairs, and walking). Based on the predicted activities, the recommender system informs individual users

						about their health behaviour including calorie suggestions.
(Z. Li et al., 2019b)	Physical activity	General audience	Multimodal	Noise removal, data imputation, thresholding; Multiclass Classification and Binary Classification; 505 fitness tracking device users, out of which 203 were females	RF (AUC ≥ 0.8)	System leverages model to generate an hour-by-hour activity plan (steps goal) based on the user's probability of adhering to the plan. If probability is less than a threshold, the system suggests an alternative plan; otherwise, the plan is changed and a new activity plan is recommended.
(Stamate et al., 2017)	Disease management (Parkinson's Disease (PD))	PD patients	Single Modal	Normalization; Binary Classification; 256 samples and 512 samples; 10 iterations of leave-one-out method (random weight initialization applied at the start of each iteration); Early Stopping	MLP (accuracy = 76.9%); DT (result not reported)	Using MLP model, system is able to detect non-adherence movement protocol and then offer personalized support in form of audio, video and textual media guide to boost adherence. The system also provides personalized quick tests using the DT model.
(Ghandeharioun et al., 2019)	Mental health	General audience	Multimodal	Sliding windows, transformation, normalization; Binary Classification (Negative/Positive, Low/High); Regression; Train/Validation/Test split; 10-fold cross-validation	RF (accuracy = 82.4%); AdaBoost (accuracy = 67%)	Personal assistant/agent predicts user's emotional state (valence and arousal) using the models and, based on the predicted state, suggests micro-interventions that fall into one of the following psychotherapy categories: positive psychology, cognitive behavioral, meta-cognitive, or somatic interventions.
(Suh et al., 2012)	Physical activity	General audience	Multimodal	Classification	DT (accuracy = 88.71%)	System uses the model to generate a personalized exercise schedule based on individuals' exercise sessions, and then recommends conditions allowing the user to maximize the amount of exercise within a given heart rate range without fatigue.
(Alharthi et al., 2019)	Mental health (Stress)	Adults	Multimodal	Conversion (discrete to continuous), Sliding windows; Multiclass Classification	NB (accuracy = 78.3%)	Based on users' stress state (relaxed, normal, or stressed) predicted by the model, the system automatically triggers relief interventions (such as relaxation techniques) if the predicted state is *stressed*.
(Paredes et al., 2014)	Mental health (Stress)	Adults	Multimodal	Regression	RF (percentage of interventions adopted by the users overall is 97%)	Recommender system leverages the model predicts the expected stress reduction of each intervention for an individual at a given context. Based on these estimates, the system selects the best interventions that promote self-awareness of stress, lower depression-related symptoms and knowledge of stress coping strategies.

(Wahle et al., 2016)	Mental health (Depression)	Adults	Multimodal	Sliding window; Binary classification (depression vs non-depression); leave-one-out cross validation;	SVM (accuracy = 59.4%); RF (accuracy = 61.5%)	*MOSS* app offers just-in-time cognitive behaviour therapy tailored to individual patients depending on their depression level, as predicted by the two ML models.
(Pelle et al., 2019)	Disease management (Osteoarthritis (OA))	Patients with knee and/or hip OA	Single Modal	Reinforcement Learning; 25 participants (for training); 161 participants for testing	MAB (sensitivity = 75%, specificity = 89%)	Using the model, *dr. Bart* app dynamically learns and suggests top 5 goals (covering exercise, sleep, and nutrition) tailored to individual users based on their personal and contextual data.
(Yates & Islam, 2019)	Mental health (Depression)	Adolescents and Adults	Multimodal	Sliding window; Regression; Behavioural data: 1440 data points per 24-hour period for each user	DT (accuracy = 96.6%)	*Mindful* app applies the model to predict patients' depression level, and then provides a proactive warning about change in mental health state followed by suggestions on lifestyle changes to improve well-being.
(Forman, Goldstein, Zhang, et al., 2019)	Disease management (Obesity)	Overweight/obese Adults	Single Modal	Data balancing, Data Imputation; Multiclass Classification; 44 participants provided data; leave-one-fold-out cross-validation	Ensemble model of four algorithms (specificity = 83.8%)	Using the model, system determines the risk of dietary lapse and then offer personalized interventions if risk is high.
(Sookrah et al., 2019)	Disease management (Hypertension)	Hypertensive patients	Single Modal	Binary Classification; 10 hypertensive patients (above the age of 40)	MLP (accuracy = 99%)	Recommender system applies the model to classify food-related parameters into *positive* or *negative*. Foods classified as *positive* are then used to create or recommend a personalized meal plan to hypertensive individuals.
(Barata et al., 2016)	Disease management (Asthma)	Asthmatic patients	Single Modal	Binary Classification (cough vs non-cough)	SVM (accuracy = 83.3%)	*MobileCoach* app uses the model to detect coughs in real-time and then provides personalized intervention based on number of coughs detected.
(Kariyawasam et al., 2019)	Disease management (Dyslexia, Dysgraphia, Dyscalculia)	Children	Single Modal	Feature extraction was done using Mel-Frequency Cepstral Coefficient (MFCC), Binarization, segmentation; Binary Classification	**Dyslexia**: CNN (accuracy = 65%) and KNN (result not reported); **Dysgraphia**: CNN (accuracy = 85%) and RF (accuracy = 90%) and SVM (result not reported); **Dyscalculia**: SVM (accuracy = 90%)	Pubudu system leverages various models to assess children for dyslexia, dysgraphia, dyscalculia and automatically provide tailored interventions to support the child's learning process (reading, writing, and mathematical skills) depending on the disease.
(Chin et al., 2020)	Assistive healthcare (Human fall detection)	General audience	Single Modal	Binary Classification	SlowFast Net (test loss = 0.477)	Robot equipped with an iPhone leverages model running on the edge (phone) and on a 2D camera installed in the room to track, detect, and assess a fall. Once a fall is detected and confirmed, a caregiver is notified for quick rescue.
(Rabbi et al., 2015)	Disease managem	Adults	Single Modal	Reinforcement Learning	MAB (cohen-d ≤ 0.84)	System dynamically learns and influences user's physical activity

	ent (Obesity)					and dietary behaviours by suggesting actions that maximize the chances of achieving calorie loss goals.
(Burns et al., 2011)	Mental health (Depression)	Adults	Multimodal	Classification and Regression; 10-fold cross-validation; Train/Test split (90/10)	DT (accuracy ranges from 60% to 91%)	*Mobilyze* app predicts the contextual states of depressive disorder patients using the model, and then offer personalized interventions (e.g., lessons, tools) and graphical feedback on individual patients' states. Personalized coaching-based therapeutic support are also provided.
(Koren et al., 2019)	Medical diagnosis	Adults	Single Modal	Embeddings (Word2Vec), Bag-of-words+TF-IDF; Binary Classification (Multilabel); 1085 patients in MHS, and 130 patients in Integrity Health Services	Ensemble of Bayes Net, XGBoost, LogisReg, and ANN (accuracy = 83.9%)	Diagnostic system first determines the cohort of similar people with a similar set of symptoms using the model. Next, the patient is shown the cohort's path to treatment, including the various conditions with which they were diagnosed, along with the cohort's full course of action including the types of medical professionals seen, tests ordered, medications prescribed, and expected recovery time.
(Khumrin et al., 2018)	Medical diagnosis	Medical students	Single Modal	Multiclass Classification; 20 clinical scenarios	NB (accuracy = 60%)	Using the model, *DrKnow* system adapts feedback to individual students as they work through a clinical scenario (or virtual cases) to help them review and reflect on their diagnostic rationale and inform next steps.
(M. Chen et al., 2018)	Disease management (Diabetes)	General audience	Multimodal	Filtering (removing irrelevant and duplicate samples), Data imputation, Conversion, Normalization; Deep learning for feature extraction; Binary Classification; 469 diabetes patients and 9081 normal patients	Ensemble of DT, SVM, and ANN (accuracy = 94%)	Model predicts if a patient has a higher risk of diabetes or not. If diabetes risk is confirmed, he/she receives tailored suggestions for meals, as well as personalized reminders on medications or insulin injections.
(Lopez-Guede et al., 2015)	Health monitoring (Human activity recognition)	Older Adults	Multimodal	Sliding window; Classification and Clustering; Cross-validation;	DT (accuracy = 81.8%); LOF (result not reported)	*Lynx* system leverages models to detect deviations from users' normal daily tasks (wake-up times, sleep habits, diary strolls, etc.) and health situations. For every unusual situation detected, the system sends notifies the family members, care center, or medical agents.
(Kesavan & Arumugam, 2020b)	Medical diagnosis	Adults	Multimodal	Deep learning for feature extraction (temporal and spatial features); Classification; Levy Flight-based	ADCNN (F1-score = 95%)	System leverages the model to classify patients' health condition into *normal*, *sensitive*, or *critical* severity level, and then the notifies the

				Grey Wolf Optimization (LFGWO) technique for optimizing model weights		doctor or practitioner immediately if condition is not normal.
(Aujla et al., 2019)	Medical diagnosis	General audience	Multimodal	Feature selection using Singular Value Decomposition (SVD); Multiclass Classification; 100 patients, 20 doctors	DT (RMSE = 0.424, MAPE = 4.64%), CNN (accuracy of recommendations ranges from 70% to 100%)	Recommender system leverages the DT model to classify individuals' health data into one of *k* diseases, and then a ranked list of doctors that can provide the right treatments for the disease is recommended using CNN. Each patient can choose a nearby or remote doctor based on his/her ranking.
(El Barachi et al., 2019)	Disease management (Epilepsy)	Epileptic patients	Single Modal	Transformation/resampling, thresholding, segmentation, Normalization; Binary Classification (Seizure/Non-seizure); Cross-validation; Train/Test split (70/30)	SVM (accuracy = 86%)	*EpiSense* app applies the model to detect epileptic seizures in real-time, and then alerts the patient's caretakers of the occurrence of a seizure for quick intervention.
(Hermens et al., 2014)	Mental health (Stress)	General audience	Multimodal	Regression	MR (r^2=0.481 for one of the subjects)	Coaching system adapts stress detection model to individual users to achieve the most reliable estimation of stress.
(Delmastro et al., 2020)	Mental health (Stress)	Older Adults	Multimodal	Sliding window, MATLAB and Ledalab for feature extraction; Correlation-based feature selection (CFS), Information gain ratio-based Feature Selection using Principal Component Analysis (PCA); SMOTE for oversampling; Binary Classification; 10 repetitions of the 10-fold cross validation scheme; grid search	RF (accuracy = 85.3%); AdaBoost (accuracy = 85.5%)	App detects the stress level of older adults during a cognitive training session using both models, and then creates a personalized training activity that helps to reduce the stress level (if above expected limit) while improving cognitive ability.
(Neloy et al., 2019)	Medical diagnosis	Patients in Intensive Care Unit (ICU)	Single Modal	Transformation, Normalization, Feature selection via correlation analysis; Multiclass Classification; 5000 sample data; 6-fold cross-validation; Train/Test/Validation - 60/20/20; Grid search	Bagging SVM achieved the highest accuracy of 92% compared to other ensemble models	CPMS system utilizes model to predict the health condition of individual patients. If the condition is worse, the system sends tailored SMS to the appropriate health professionals for immediate action.
(Alfian et al., 2018)	Disease management (Diabetes)	Pima Indian Women	Multimodal	Derived attributes; Classification and Regression; 768 patients out of which 500 tested negative (normal) while 268 of them	MLP (accuracy = 77%); LSTM ($r \leq 0.999$)	*uHealthFit* app leverages the models to predict the presence or absence of diabetes in individual users and their blood glucose level. Based on the prediction, the app

				tested positive; 10-fold cross-validation;		presents personalized suggestions covering healthy diet, weight loss, and physical activity.
(Khalaf et al., 2016)	Disease management (Sickle Cell Disease (SCD))	SCD Patients	Single Modal	Regression	MLP (MAPE = 0.1345)	System utilized the model to predict the correct amount of Hydroxycarbamide drug/liquid with the aim of providing the right therapeutic support that fit the patient.
(Javed et al., 2021)	Assistive healthcare (Cognitive impairment)	Cognitive Impaired People	Multimodal	Normalization; Multiclass Classification; 5-fold, 10-fold, and leave-one-fold-out cross-validation	HT (F1-score = 90.2%); LogisReg (F1-score = 90.2%)	PP-SPA framework employs models to detect the current activity or task a user is performing and then provides a personalized real-time support (such as prompts showing functional aid) to help the user to complete the task
(Kajiwara & Kimura, 2019)	Assistive healthcare (Visual impairment)	Visually impaired People	Multimodal	Extract image features using OpenPose tool; Classification; 10-fold cross-validation	RF (accuracy = 92%)	*Follow Me!* app leverages model to provide smart navigation support by detecting people walking in the same direction as the visually impaired user, oncoming pedestrians, steps, and other “unknown” objects, and then recommend a safe route to the user.
(Mrozek et al., 2020)	Assistive healthcare (Human fall detection)	Older Adults	Multimodal	Sliding windows; Binary Classification (Fall vs Non-fall); 5-Fold Cross-validation; Hyperparameter tuning (maximum number of leaves per tree, minimum number of samples per leaf node, learning rate, number of trees)	BDT (*edge-based detection*: accuracy = 99.2%; *cloud-based detection*: accuracy = 99.8%)	*Whoops* app utilized model deployed on user’s smartphone (edge) and cloud to detect falls. If a fall is detected and the user confirms it, the system places an emergency call to the caregiver immediately.
(Asthana et al., 2017)	Medical diagnosis	Adults	Multimodal	Bag-of-words+TF-IDF; Multiclass Classification; 135,000 data points, 50 target classes (disease risks)	DT (RMSE = 0.1066)	Framework predicts a user’s at-risk health conditions, and then suggests appropriate wearable devices and measurements that can help the user to evaluate health risks and monitor them.
(Popp et al., 2019)	Disease management (Diabetes)	English- or Hebrew-speaking audiences	Multimodal	Regression	GB (r = 0.6267)	Using the model, personalized PPGR (postprandial glycemic response) is calculated per patient for every meal and snack based on their nutrient composition. Individual patients are then advised to make different choices or food substitutions if their meal is neither excellent nor good using meal ratings obtained from calorie-adjusted quintile cutoffs of the PPGR.
(da Silva, Souza, et al., 2019)	Disease management	Older Adults	Multimodal	Transformation, duplicate filters; Binary	DT (accuracy = 95.1%)	System leverages model to predict adherence or non-adherence to

	(Hypertension)			Classification (adherence vs non-adherence); 10-fold cross-validation		prescribed medications. In case of non-adherence, the system triggers tailored messages alerting individual patients, doctors, caregivers, and relatives.
(Sansrimahachai, 2020b)	Physical activity	Older Adults	Multimodal	Derived attributes, Normalization; Regression; Train/Validation;	ANN (MAE = 16.78, MSE = 15.3)	*WalkPal* system utilizes model to predict weekly walking exercise minutes, and then generates a challenging but realistic personalized walking plan for each user based on the predicted minutes.
(Afreen et al., 2019)	Disease management (Diabetes)	Type-2 diabetic patients	Single Modal	Normalization; Regression; Trained on a dataset of 8069 patients and validated on 2018 patient data; Train/Test split (80/20); Random selection on various hyperparameters (random grid search)	XGBoost (r^2 = 0.9799); MLP (r^2 = 0.9522)	Recommender system leverages the models to predict total calories consumed and % of carbohydrate, protein, and fat in a diet, and then generates personalize diet chart for the diabetic patient.
(Pathinarupothi et al., 2018)	Medical diagnosis	General audience	Multimodal	Sliding (time) windows; Binary Classification; Group of 35 patients (called group H) who had AHE during their stay in ICU, and another 35 patients (called group G) who did not have AHE during their ICU stay; 5-fold cross-validation (70/30 split)	SVM (F1-score > 88%) for detecting Acute Hypotensive Episodes (AHE)	RASPRO framework provides personalized patient monitoring, precision diagnostics, and preventive criticality alerts (3Ps). A prototype system uses SVM model to detect Acute Hypotensive Episodes (AHE) and criticality, and then trigger alerts to physician for interventional attention if condition is critical.
(Prabhu et al., 2018b)	Disease management (Cardiovascular Disease)	Patients who suffered a cardiac event and are in Phase III of the recovery process	Single Modal	Sliding window; Binary Classification (LME exercise vs Random movement); Each exercise is performed by 6 participants; cross-validation (80/20); Regularization	SVM (accuracy ≥ 98%)	*MedFit* system leverages the model to detect up to 14 local muscular endurance (LME) exercises completed by the patient and repetitions. The system then offers personalized video exercise classes and feedback to each patient.
(Van Woensel et al., 2020)	Assistive healthcare (Human activity recognition)	Adults	Single Modal	Sliding time window; Classification; 2012 samples for training; 841 samples for testing; 10-fold cross-validation (repeated 10 times)	DT (accuracy = 84%); hierarchical clustering (accuracy = 90%)	*iLocate* framework uses models to identify individuals' discrete location (room-level estimation) and region-level estimation, as well as the semantics of identified locations (e.g., waiting room). An eHealth system leverages *iLocate* to deliver context-sensitive care activities (e.g., pathfinding aids and health warnings) based on the patient's inferred current location.
(Hassan et al., 2019b)	Assistive healthcare (Human	Older Adults	Single Modal	Sliding window; Multiclass Classification (Fall, Standing, Lying),	CNN-LSTM (F1-score = 97%)	Using model which runs on user's smartphone (edge), the system detects human falls in real-time.

	fall detection)			Binary Classification (Fall, Non-fall); 7671 samples for Experiment 1, 4000 samples for Experiment 2		Once a fall is detected, the system triggers an indoor sound alert to family members through a wireless access point at home and an outdoor SMS alert is sent to a hospital or caregiver through a mobile network base station.
(Stamate et al., 2018)	Disease management (Parkinson's disease (PD))	Patients with PD	Single Modal	Feature extraction using Deep Learning, Spike 2 tool for biosignal analysis; Classification; 10 cycles of leave-one-(session)-out cross-validation; early stopping	RCNN (accuracy = 78%); DT (result not reported)	cloudUPDRS system utilizes the RCNN model to detect failures to follow the movement protocol, and then provides tailored support (such as audio, video and textual media guide) to help patients adhere to prescribed movements and to reduce test duration (via personalized quick tests created using DT model).
(Kang, 2021)	Health monitoring	Military Personnel	Single modal	Samples 17007 (Heart rate dataset); Regression; Train/Test split (70/30)	Levenberg-Marquardt algorithm (accuracy = 95.6%)	A human performance management system was developed to monitor the fitness status of warfighters in real-time using physiological data and ML. The system offers personalized fitness training based on individuals' predicted health index and fitness level.
(Luštrek et al., 2021)	Disease management (Congestive heart failure)	Patients with Congestive heart failure	Multimodal	Preprocessing using MATLAB and R: Speech normalization, noise removal (low-pass filtering) from accelerometer reading, band-pass filtering (remove gravitational component), data segmentation (window size of 2); Classification; Regression; Train/Test split (70/30); Leave-one-subject-out cross validation	RF for BP estimation (MAE = 9.0 for SBP, MAE = 7.0 for DBP); SVM for psychological profile detection (accuracy=88.6%)	HeartMan mobile application monitors patients with congestive heart failure (CHF) by utilizing ML models for continuous blood pressure estimation, physical activity monitoring, and psychological profile recognition. Based on model prediction/estimation, the application offers personalized exercise programs (such as endurance and resistance exercises), nutrition advice, CBT for anxiety and depression management, and mindfulness-related contents.
(Ganju et al., 2021)	Health awareness	Children	Single modal	Exclusion of certain features with little or no predictability of the target and by identifying highly correlated features; Classification; Regression	RF (accuracy = 93%; RMSE: 25.09, $r^2 = 0.91$)	An RF model to predict user churn and user lifetimes (i.e., number of days a user will stay on the *Saathealth* application before uninstalling). Model output was used to incentivize users with optimized and more personalized/targeted offers and omni-channel nudges, as well as augmented in-app experiences.
(Y. Zhang et al., 2022)	Disease surveillance	General audience	Single modal	Convert all text to lower case; Remove punctuations and	Linear Regression ($r^2 > 0.90$; fine-tuned	A web-based COVID-19 surveillance system utilizes fine-tuned BERT model for classifying

	(COVID-19)			stop words; Apply lemmatization to convert all words to their basic form; Apply tokenization; Classification; Regression; Train/Test split (70/30)	BERT: F1-score = 98%;)	tweets into COVID-19-related or otherwise. Classified tweets are then analyzed to forecast epidemic using LR model. The system provides early earning messages based on the forecast.
(Zingg et al., 2021)	Mental health (Depression)	Patients with PPD (Peripartum Depression)	Single modal	Remove unwanted characters, white spaces; Make all text lowercase; Stopwords removal, stemming; Randomly selected 850 user posts for manual labeling; Classification; Train/Test split (67/33)	RF (average F1-score=77.2%)	Digilego adaptive framework applied ML to analyze social media posts to inform digital features of a mobile intervention (MomMind) that promotes Peripartum Depression (PPD) prevention and self-management.
(Hu et al., 2021)	Medication management	General audience	Single modal	Data augmentation of tongue images using an image data generator; Tongue positioning and cutting of non-tongue regions; Transfer learning; Normalization of the names of herbs; Classification	CNN+MLP+AUX_LDA (precision similarity = 46.77%, recall similarity = 37.17%)	Prescription system used ML models to predict and recommend appropriate herbs for individuals based on their tongue image.
(Barbosa et al., 2021)	Medical diagnosis (COVID-19)	General audience	Single modal	Handling missing data (missing data were filled in with the average value of each of the exams); Transforming categories into numerical classes; Classification	RF for diagnosis (accuracy = 92.891%, sensitivity = 93.6%, precision = 92.3%, specificity = 92.1%, AUC=98.4%); RF for hospitalization prediction (accuracy>99%)	Heg.IA system detects the presence or absence of COVID-19 virus from clinical data (e.g., blood tests) using RF model. The model is also used to predict/recommend the best type of hospitalization for the patient: regular ward, semi-ICU or ICU.
(Mishra et al., 2021)	Physical activity	General audience	Single modal	Undersampling of training data (adaptive model) using the Instance Hardening Threshold (IHT) method; 10-minute window to define the receptivity metrics; Classification; Train/Test split (80/20); 5-fold Leave-one-group-out cross-validation	SVM (F1-score=36%); Logistic Regression (F1-score=38%)	Ally+ chat-based mobile application employed two ML models (SVM as static model and LR as adaptive model) trained using contextual information about a person (such as time, battery state and level, phone state, and activity) to predict when a person is more receptive and then triggered interventions at that moment.
(Lee et al., 2021)	Medical specialty recommendation	General audience	Single modal	Duplicate and missing data were eliminated; Ambiguous sentences removal; Removal of sentences with ≤2 words or those not related to medical consultations; Stopword removal; Deep learning for feature extraction;	LSTM (precision=80.5%, recall=68.6%, F1-score=73.9%, AUC=96.5%)	A web-based AI chatbot was developed to help patients understand which medical specialty is appropriate for the treatment of their current symptoms and then make an appointment with the appropriate specialist.

				Classification; 10-fold cross-validation		
(Mahyari & Pirolli, 2021)	Mental health (Stress)	General audience	Single modal	Data vectorization using Word2Vec with Skipgram; Association analysis; Classification; Leave-one-subject-out cross validation	First LSTM (accuracy=80%); Second LSTM (result not reported)	Adaptive recommender system aimed to reduce perceived stress by suggesting exercises that have a high probability of being pursued and achieved for each individual on each day using two interconnected RNNs.
(Coutinho et al., 2021)	Mental health (Depression)	General audience	Single modal	Data standardization; Applied the Boruta feature selection algorithm on each fold to optimise the input feature set; 1802 songs belonging to 14 musical styles and Arousal and Valence annotations for each track; Regression; nested cross-validation procedure	SVR with Laplacian Kernel (RMSE=0.047); RF (RMSE = 0.110)	POLYHYMNIA Mood web app empowers people to use music effectively for coping with depression in everyday life. Based the estimated emotion (valence and arousal) using ML, a personalized playlist comprising 14 tracks (about 45 minutes of music) was automatically generated to elevate individual users' mood and reduce depressive symptoms.
(S. Wang et al., 2021)	Physical activity	General audience	Single modal	Pre-learned a generalized initial delivery strategy in the Reinforcement Learning model from a large-scale empirical running data	Markov decision process (MDP)-based Reinforcement Learning model (83.3% of reminders sent at adaptive moments were able to elicit user reaction within 50 min; 66.7% of physical activities in the intervention week were performed within 5 hours of the delivery of a reminder)	Reinforcement learning agent (named PAUL) adaptively selects the optimal strategy for delivering physical activity reminders with respect to the momentary context of this user (time and calendar).
(Arpaia et al., 2022)	Disease management (Hyperthension)	Hypertensive patients	Multimodal	Classification; Train/Test split (80/20)	LSTM Autoencoder (F1-score=96%, accuracy=93%, AUC=81%)	An adaptive telemonitoring system was developed to provide an alert when patients' vitals exceed certain thresholds using ML.
(Elvitigala et al., 2021)	Mental health (Stress)	General audience	Single modal	Time window of 10 seconds; Classification; Leave-one-user-out cross validation	Linear-discriminant analysis (F1-score = 84.11%; accuracy = 84.32%)	*StressShoe* app estimates stress level by sensing behavioural changes based on sitting posture using sensor. The app delivers just-in-time interventions if the stress level estimated by a linear-discriminant analysis model exceeds a user-defined threshold.
(Jamil et al., 2021)	Physical activity	General audience	Single modal	Remove duplication, blank space, change the text to upper/lower case, and spell-check; Normalization; Fill the missing values	SVM (accuracy = 92.1%, F1-score = 87.2%)	An adaptive blockchain-based fitness app was developed to collect fitness data via IoT devices and then apply an SVM model in its inference engine to recommend personalized

				using the probabilistic model that is based on maximum likelihood; Classification; Train/Test split (80/20); 10-fold cross-validation		diet and workout plan to individuals.
(P.-H. Chiang et al., 2021)	Health monitoring	General audience	Multimodal	Feature extraction using Autoregressive Integrated Moving Average (ARIMA); Feature selection based on a pre-trained RF model and Shapley value; Regression	RF with Shapley Value: *SBP estimation* (MAE=5.34, RMSE=8.24, MAPE=4.19%, r^2=0.51); *DBP estimation* (MAE=3.80, RMSE=6.05, MAPE=4.83%, R^2=0.52)	BP prediction and recommendation system uses ML model to predict a user's current BP level using his/her historical BP readings as well as activity, sleep and heart rate data. The system also identifies the most important lifestyle features/factors that impact the individual's BP trend, and then recommend the next top feature to the user.
(Rajasekaran & Kousalya, 2022)	Assistive healthcare	Older adults	Single modal	Classification	DC-DBN (TPR=95.66%, FPR=0.04%)	Virtual nursing system used Deep Continuous Deep Belief Network (DC-DBN) with Restricted Boltzmann Machines (RBM) to process video streams that track elderly patients' well-being. An alarm is transmitted to the patient's relatives by the system if the patient's motions alter abruptly (such as during a fall).

APPENDIX B

Table 2. Public datasets utilized in the reviewed literature for training ML models.

Dataset	Description	References	Download Location
SNAPSHOT dataset	Multimodal dataset that contains information about students' sleep, social networks, affect, performance, stress, and health. The dataset currently includes approximately 145,000 hours of data from 201 participants at one university.	(Nosakhare & Picard, 2020)	Available on Request
Dailydialog	Human annotated dataset consisting of dialogs that resemble day-to-day conversations across diverse topics.	(W. Yang et al., 2020)	http://yanran.li/dailydialog
RedditCasual	Contains narrative sentences scraped from Reddit	(W. Yang et al., 2020)	Available on Request
Depression dataset	Consists of posts with responses extracted from the Reddit's depression subgroup	(W. Yang et al., 2020)	https://github.com/Inusette/Identifying-depression
USDA database	US Department of Agriculture (USDA) National Nutrient Database containing nutritional values for recipes	(Gorbonos et al., 2018; Sookrah et al., 2019)	http://www.ars.usda.gov/ba/bhnrc/ndl
Allrecipes	Online repository containing diverse foods/recipes and their corresponding nutritional values	(Gorbonos et al., 2018)	https://www.allrecipes.com
Yummly	Online repository containing diverse foods/recipes and their corresponding nutritional values	(Gorbonos et al., 2018)	https://www.yummly.com

PIMA Indians Diabetes Database	Contains diabetes-related data of 768 patients such as the number of pregnancies, 2 h glucose tolerance, diastolic blood pressure, triceps skinfold, 2 h serum insulin, body mass index, diabetes pedigree function, and age.	(Alfian et al., 2018)	Available on Request
Blood Glucose (BG) dataset	Contains data of 70 diabetic patients for different time intervals: breakfast, lunch, dinner, and bedtime	(Alfian et al., 2018)	https://archive.ics.uci.edu/ml/datasets/diabetes
Blood Glucose (BG) dataset	Contains the BG levels of a Type-1 diabetic patient recorded every 5 min using a continuous glucose monitoring device	(Alfian et al., 2018)	https://choens.github.io/blood-sugars
Demographics+ Diseases dataset	Dataset of diseases that led to deaths grouped by age and gender	(Asthana et al., 2017)	https://www.worldlifeexpectancy.com/usa-cause-of-death-by-age-and-gender
IoT database	Captures wearables, IoT solutions, as well as healthcare applications and services	(Asthana et al., 2017)	Available on Request
Pediatric EEG Dataset	Contains Electroencephalogram (EEG) recordings from pediatric subjects with intractable seizures	(El Barachi et al., 2019)	Available on Request
UR Fall Detection dataset	Comprises image sequences with falls performed by 5 people from standing and sitting positions	(Chin et al., 2020)	http://fenix.univ.rzeszow.pl/mkepski/ds/uf.html
Multiple Cameras Fall dataset	Captures video sequences of falls and normal activities viewed from various cameras and performed by one subject	(Chin et al., 2020)	Available on Request
UCF101 action dataset	Comprises 101 human actions with over 13,000 clips and 27 hours of video sequences	(Chin et al., 2020)	https://www.crcv.ucf.edu/data/UCF101.php
MobiAct	Contains sensor data collected via smartphone gyroscopes, accelerometers, and orientation sensors of 67 subjects. The dataset captures four types of falls, 11 types of daily living activities, and a lying activity after a fall recorded.	(Hassan et al., 2019b)	https://bmi.hmu.gr/the-mobifall-and-mobiact-datasets-2/
Cancer Genome Atlas' (TCGA) database	Contains miRNA samples for 34 different tumor types	(Cheerla & Gevaert, 2017)	https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga
Gene Expression Omnibus (GEO) database	Made up of several miRNA datasets	(Cheerla & Gevaert, 2017)	http://www.ncbi.nlm.nih.gov/geo
MIMIC II database	Contains various body sensor values including electrocardiogram (ECG), ABP (Arterial Blood Pressure), Heart Rate (HR), etc. collected from more than 20,000 ICU patients	(Pathinarupothi et al., 2018)	https://archive.physionet.org/mimic2/
Wireless Sensor Data Mining (WISDM) dataset	Comprises 1,098,207 accelerometer sensor readings obtained from the smartphones of 36 subjects	(Kadri et al., 2020)	Available on Request
MNIST handwritten digits dataset	Contains images of handwritten digits	(Kariyawasam et al., 2019)	http://yann.lecun.com/exdb/mnist/
SisFall dataset	Contains accelerometer and gyroscope sensor readings collected from 38 people performing daily activities (walking, jogging, sitting, entering a car, and jumping) and simulating falls	(Mrozek et al., 2020)	http://sistemic.udea.edu.co/en/investigacion/proyectos/english-falls/
MyFitnessPal database	A large food database consisting over 300 million items	(Rabbi et al., 2015; Sowah et al., 2020)	http://www.myfitnesspal.com
Food-a-Pedia	Database consisting large variety of foods and their corresponding nutritional values	(Rachakonda et al., 2020)	https://catalog.data.gov/dataset/food-a-pedia
Weigh-Less book	Online repository showing food items and meal plans for a healthy weight while focusing on Mauritian cuisines	(Sookrah et al., 2019)	https://weighlessmauritius.com/recipes
ImageNet	An image database organized according to the WordNet hierarchy (nouns only) in which each node of the hierarchy is depicted by hundreds and thousands of images.	(Hu et al., 2021)	https://image-net.org/
COVID-19 clinical dataset	A public database with information on hospitalized patients at Hospital Israelita Albert Einstein in Sao Paulo, Brazil. The database consists of data from 5644 patients with symptoms similar to COVID-19.	(Barbosa et al., 2021)	https://www.kaggle.com/datasets/einsteindata4u/covid19
DEAM dataset	MediaEvalDatabase for Emotional Analysis of Music (DEAM) includes 1802 songs	(Coutinho et al., 2021)	https://cvml.unige.ch/databases/DEAM/

	belonging to 14 musical styles and Arousal and Valence annotations for each track.		