Title

Use of machine learning to predict medication adherence in individuals at risk for atherosclerotic cardiovascular disease

Authors

Mirzadeh, Seyed Iman
Arefeen, Asiful
Ardo, Jessica
et al.

Publication Date

2022-12-01

DOI

10.1016/j.smhl.2022.100328

Peer reviewed

HHS Public Access

Author manuscript

Smart Health (Amst). Author manuscript; available in PMC 2023 May 09.

Published in final edited form as:

Smart Health (Amst). 2022 December ; 26: . doi:10.1016/j.smhl.2022.100328.

Use of machine learning to predict medication adherence in individuals at risk for atherosclerotic cardiovascular disease

Seyed Iman Mirzadeh[a], **Asiful Arefeen**[b,*], **Jessica Ardo**[c], **Ramin Fallahzadeh**[d], **Bryan Minor**[a], **Jung-Ah Lee**[c], **Janett A. Hildebrand**[e], **Diane Cook**[a], **Hassan Ghasemzadeh**[b], **Lorraine S. Evangelista**[f]

[a]School of Electrical Engineering & Computer Science, Washington State University, Pullman, WA, 99163, USA

[b]College of Health Solutions, Arizona State University, Phoenix, AZ, 85004, USA

[c]Sue & Bill Gross School of Nursing University of California Irvine, Irvine, CA, 92697, USA

[d]Department of Biomedical Data Sciences, Stanford University, Stanford, CA, 94305, USA

[e]Department of Nursing at the School of Social Work, University of Southern California, Los Angeles, CA, 90089, USA

[f]School of Nursing, University of Nevada, Las Vegas, NV, 89154, USA

Abstract

Background: Medication nonadherence is a critical problem with severe implications in individuals at risk for atherosclerotic cardiovascular disease. Many studies have attempted to predict medication adherence in this population, but few, if any, have been effective in prediction, sug-gesting that essential risk factors remain unidentified.

Objective: This study's objective was to (1) establish an accurate prediction model of medication adherence in individuals at risk for atherosclerotic cardiovascular disease and (2) identify significant contributing factors to the predictive accuracy of medication adherence. In particular, we aimed to use only the baseline questionnaire data to assess medication adherence prediction feasibility.

Methods: A sample of 40 individuals at risk for atherosclerotic cardiovascular disease was recruited for an eight-week feasibility study. After collecting baseline data, we recorded data from

*Corresponding author: aarefeen@asu.edu (A. Arefeen).

Credit author statement

Seyed Iman Mirzadeh Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing **Asiful Arefeen** Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing **Jessica Ardo** Supervision, Writing – review & editing **Ramin Fallahzadeh** Supervision, Writing – review & editing **Bryan Minor** Supervision, Writing – review & editing Supervision, Writing – review & editing **Janett A. Hildebrand** Supervision, Writing – review & editing **Diane Cook** Supervision, Writing – review & editing **Hassan Ghasemzadeh** Supervision, Writing – review & editing **Lorraine S. Evangelista** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

a pillbox that sent events to a cloud-based server. Health measures and medication use events were analyzed using machine learning algorithms to identify variables that best predict medication adherence.

Results: Our adherence prediction model, based on only the ten most relevant variables, achieved an average error rate of 12.9%. Medication adherence was closely correlated with being encouraged to play an active role in their treatment, having confidence about what to do in an emergency, knowledge about their medications, and having a special person in their life.

Conclusions: Our results showed the significance of clinical and psychosocial factors for predicting medication adherence in people at risk for atherosclerotic cardiovascular diseases. Clini-cians and researchers can use these factors to stratify individuals to make evidence-based decisions to reduce the risks.

Keywords

Medication adherence machine learning cardiovascular diseases cloud computing Surveys and questionnaires

1. Introduction

Medication adherence, termed by Food and Drug Administration (FDA) as the extent to which patients take medication as prescribed by their doctors (FDA (2019)), is a significant factor in the efficacy of medical therapies and the secondary prevention of atherosclerotic cardiovascular disease. Despite the significance, up to 50% of patients fail to take their medications as directed, which has an immediate impact on clinical conditions. (Huber et al. (2019); von Wyl et al. (2020); Lauffenburger et al. (2018)). In the US, the cost of prescription non-adherence ranges from $100 billion to $290 billion (El-Saifi et al. (2018); Conn and Ruppar (2017); Capoccia et al. (2016)). For each person with atherosclerotic cardiovascular disease, the adjusted yearly cost of nonadherence medication ranges from $6943 to $16,124 (Cutler et al. (2018)). Other studies have shown that raising medication adherence can save medical expenses while also enhancing clinical results in the fight against atherosclerotic cardiovascular disease (Mathews et al. (2018); Arnold et al. (2020); Du et al. (2017); Bansilal et al. (2016)).

The importance of medication adherence has led to research to identify modifiable predictors of medication adherence in individuals at risk for atherosclerotic cardiovascular disease. The goal of these studies was to develop therapies that take clinical and psychological factors into account in order to lessen or eliminate the negative effects of pharmaceutical nonadherence (Kumamaru et al. (2018); Hope et al. (2019); Omezzine et al. (2019); Khayyat et al. (2018)). Studies currently available on predicting medication adherence have a few drawbacks, though. First, most existing studies focus on predicting medication adherence as a quantitative variable (i.e., whether or not adherence is above a certain threshold) and approaching adherence estimation as a classification issue. We assert that this issue may be extended to a regression problem where consistency is a constant rather than a single variable. Medication adherence would be treated as a continuous variable that can be modeled in a more informative manner in this situation. Second,

while certain factors during the study would influence medication adherence and potentially predict improved adherence, we believe it is more important to determine medication adherence for clinical and research purposes to scrutinize those at higher risk for medication nonadherence need to be closely monitored.

Since medication adherence is important in lowering the risk of atherosclerotic cardiovascular disease, a greater understanding of the factors associated with medication adherence in individuals deemed high-risk for this condition is required to develop effective strategies to improve medication adherence (primary prevention). Many who have had cardiovascular disease should also be scrutinized for medication adherence behaviors (secondary prevention). This study aims to overcome the above limitations and predict medication adherence with high accuracy, utilizing standard variables that can be collected quickly. This study is aimed to-

1. Establish an accurate prediction model of medication adherence in individuals at risk for atherosclerotic cardiovascular disease and
2. Identify variables showing the highest levels of contribution to the prediction accuracy of medication adherence.

2. Methods

2.1. Recruitment and procedures

This study has been reviewed and approved by the appropriate Institutional Review Board (IRB). Participants were recruited from a single outpatient tertiary care facility in Southern California through primary care provider referrals. The eligibility of the participants was verified before obtaining their informed consent. Inclusion criteria included:

1. ≥ 21 years of age;
2. at risk for atherosclerotic cardiovascular disease (e.g., overweight, obesity, hypertension, diabetes, metabolic syndrome, or hyperlipidemia);
3. taking drugs for their condition;
4. Living independently;
5. Willing to carry the study phone with them at all times.

Individuals who had extreme comorbidities, cognitive impairment, serious uncorrected vision or hearing loss, or were unable to read and write English were not eligible for the study. Considering all the inclusion and exclusion criteria, we recruited 40 participants who completed baseline and eight-week follow-ups. Participants' demographic information is summarized in Table 1. In addition, participants were provided with adequate training on using the pillbox, Android smartphone device, and answering the activity prompts. The research team met privately with each participant to obtain informed consent, address any research-related queries, and collect baseline measures. The patient's clinical status (blood pressure, heart rate, height, weight, and waist circumference) was measured and reported after consenting to participate. Finally, participants completed a survey consisting of questionnaires discussed in detail in the following paragraphs.

2.2. Data collection

All participants were asked to complete a survey packet composed of several questionnaires. The survey provided sociodemo-graphic information, including age, sex, gender, race/ethnicity, the number of medications/pills taken every day, and various clinical and psychosocial parameters. The questionnaires included in the survey included:

Charlson Comorbidity Index (CCI)—Charlson Comorbidity Index (CCI), which includes 19 disorders weighted according to their 1-year mortality association (Buck et al. (2013)). The stated goal of the instrument was to control for sicker individuals in longitudinal clinical trials. The CCI had the advantage of simplicity and ease of use over previous methods of comorbidity measurement. The index score is constructed by summing the weights assigned to each disease based on the magnitude of the relative risk of mortality associated with each disease. Scores can range from 0 to 34. Illness severity can be divided into: “not ill,” “mildly ill,” “moderately ill,” and “severely ill.” Previous validation of the CCI with correlation coefficients of >0.40 showed a “good” test-retest reliability and “moderate to good” interrater reliability (Charlson et al. (1987)). In more recent studies, the CCI was an appropriate prognostic indicator for in-hospital and one-year outcomes in patients with acute coronary syndromes (Radovanovic et al. (2013)).

The Hospital Anxiety and Depression Scale (HADS)—The Hospital Anxiety and Depression Scale (HADS) is a 14-item scale that uses seven days as the reference period. Seven items relate to anxiety (HADS-A), and seven relate to depression (HADS-D). This outcome measure was developed explicitly to avoid reliance on elements of these conditions that are often typical somatic symptoms of illness (e.g., fatigue, insomnia) (Herrero et al. (2003)). The HADS is a well-known self-report questionnaire for mental distress widely used in chronic illness and other physical health settings (e. g., cardiology, brain injury, general medicine). It was originally designed to screen patients’ emotional discomfort in non-psychiatric settings by identifying the two most common distress components: anxiety and depression (Löwe et al. (2004); Terluin et al. (2009)). According to a recent analysis of cancer patients, the optimal cut-off for the HADS-A was greater than nine units, and the optimal cut-off for the HADS-D was greater than seven units. The area under the Receiver Operating Characteristic for HADS-A was 0.90, while for HADS-D, it was 0.84 (Annunziata et al. (2019)). A separate analysis of a larger general primary health care population found that the area under the curve values for both measures was well above 0.70 (Terluin et al. (2009)).

The Edmonton Symptom Assessment Scale (ESAS)—The Edmonton Symptom Assessment Scale (ESAS) evaluates nine symptoms (e.g., pain, fatigue, drowsiness, lack of appetite, shortness of breath, nausea, anxiety, depression, and wellbeing) using a scale from 0 to 10, with 10 being the worst score; the scores are further classified as mild (0–3), moderate (4–6), and severe (7–10) (Bruera et al. (1991)). The ESAS has been psychometrically validated and translated into over 20 languages since its inception in 1991 (Hui and Bruera (2017)). It is now widely used for acute symptom screening and long-term patient monitoring in patients seen by palliative care, oncology, nephrology, and other

disciplines in inpatient and outpatient environments. In addition, it has provided insight into patients' future symptom trajectories (Milton et al. (2020)).

Short Form 12, Version 2 (SF-12v2)—Short Form 12, Version 2 (SF-12v2), was used to measure participants' quality of life (Sansom et al. (2020)). The SF-12v2 taps eight health concepts: physical functioning, bodily pain, role limitations due to physical health problems, role limitations due to personal or emotional problems, emotional well-being, social functioning, energy/fatigue, and general health perceptions. It also contains a single object that reflects the perceived change in well-being. The score of each subscale was calculated using standard scoring algorithms from Ware et al. (Ware et al. (1996)). The SF-12 summary scores, which included both the physical and mental portion summaries, ranged from 0 to 100, with a higher score suggesting better self-reported health. The Physical Component Summary ranged from 0.43 to 0.93 (median = 0.67) in 14 validity tests involving physical criteria, while the Mental Component Summary ranged from 0.60 to 107 (median = 0.97) in 6 tests involving mental criteria (Ware et al. (1996)). The SF-12v2 has been translated into various languages, with a reliability coefficient that has consistently been greater than 0.70, indicating high reliability (Sansom et al. (2020); Shou et al. (2015)).

Patient Activation Measure (PAM)—Patient Activation Measure (PAM) is a 13-item tool to evaluate patients' ability to take self-rated preventive measures, manage symptoms, find and use appropriate medical care, and collaborate with care providers. Higher scores show higher activation and correlate with better self-management for chronic diseases (Evangelista et al. (2015)). The PAM can be scored using a Rasch score table that converts curvilinear summated raw scores to linear interval scores. The measure has good psychometric properties indicating that it can be used at the individual patient level to tailor intervention and assess changes. The Rasch person reliability for the preliminary 21-item measure was between 0.85 (real) and 0.87 (model). Cronbach's alpha was .87 (Hibbard et al. (2004)).

Multidimensional Scale of Perceived Social Support Scale (MSPSSS)—Multidimensional Scale of Perceived Social Support Scale (MSPSSS) is a 12-item self-reporting measure of how you see your social aid system, including social support sources for an individual (i.e., family, friends, and others) (Liu et al. (2016)). The sum of PSSS is assessed using a seven-point Likert scale, with responses ranging from strongly disagree (=1) to agree (=7) strongly. The total/-cumulative score ranges from 12 to 84. Since no item response theory calibration was applied to the instrument, the scores are interpreted: the higher the score, the greater available social support (Dambi et al. (2018)). The original PSSS produced a three-factor structure, high internal consistency (0.88), stability (0.85), and moderate construct validity, as SS scores were negatively linked to anxiety (r = 0.18; p 0.01) and depression scores (r = 24; p 0.01) (Dahlem et al. (1991)).

2.3. Setting up intervention-prompts & reminders

The smart medication bottle from Pillsy Inc. (Pillsy) was used to acquire the log of medication events (i.e., open/close event). The Pillsy mobile app communicates with the device to obtain medication events and transmits such data to the back-end server. This

information allows researchers to know when the bottle was opened and medicine was taken. The pillbox must be positioned in close proximity to the user's smartphone to guarantee data transmission via Bluetooth without any interruption. On the other hand, the smartphone records the entries on a server, and researchers can access the app's data. We developed a mobile app that communicated with the server through the Pillsy API to obtain medication events. While smart pill boxes are highly reliable (Boyd (2019)), no information is available regarding their accuracy, and their usage does not require FDA approval.

A study team member taught participants to use health devices (Android smartphone, Bluetooth pill bottle), react to activity prompts, charge the phone every night, and position the phone within three feet of the pill bottle for optimal data synchronization and transmission. The research team organized the patient's medication warning in the pharmacy tracking program when the person typically took their medication. We also quantify medication time surrounding the smart pill bottle in the app. The researchers sent each participant an activity prompt as a check and analyzed their ability to respond. We advised participants to respond every day to as many prompts as necessary and stop responding or using the phone while driving or operating heavy machinery. We note, however, that the activity data collected in this paper were not used for analysis and served other purposes, such as robust activity surveillance in uncontrolled environments.

For the first two weeks of the study, participants responded to the activity prompts sent to their mobile devices via the activity learning app. Prompts were sent every 2 h from 8:00 a.m. to 8:00 p.m. Participants received static or fixed-time medication reminders at weeks three and four and dynamic (active-aware) reminders during weeks five to eight. Upon completing the study (eight weeks of follow-up), participants met with clinicians to assess their health status, complete the questionnaire packet, and return health devices.

2.4. Data analyses

2.4.1. Machine learning approach—We treated the medication adherence problem as a regression task rather than a binary classification task. The medication usage log, acquired using the pillbox and mobile application, is the basis of the target domain data for this project. We employed a standard supervised regression framework to build a prediction system. This approach can be modeled as a supervised learning problem, where we have access to both input (i.e., questionnaire) and output/target (i.e., medication usage log). Since each participant's adherence is a continuous variable (between 0 and 100), we modeled our problem as a regression problem. We made sure to have various machine learning algorithms for regression, including more powerful decision-tree-based algorithms and baseline algorithms such as linear regression (Chen et al. (2019)). We agree that medication adherence can be defined in many ways. For example, adherence can be presented as a continuous variable that measures the amount of delay in taking medication. Nonetheless, in this study, we were interested in overall medication adherence as defined by the percentage of the days during which the participants took their medication at any time during the day.

First, we entered data from the completed questionnaires. However, roughly 7% of participant responses to baseline questionnaires were not always accurate. Therefore we had to pre-process the raw data. The pre-processing step included filling all the inaccurate and

missing data fields with the mean of similar values or with zeros, depending on the function type (Jakobsen et al. (2017)). Then, we chose a subset of questions in the feature extraction phase and removed the others. For instance, when studying the effectiveness of a particular questionnaire, we only selected the questions in that specific questionnaire during the feature extraction phase. The processed data was then inputted into the adherence prediction model.

Second, we calculated medication adherence using the medication usage log. Adherence was defined as a value between 0 and 100, representing the percentage of the study days a participant consumed the prescribed medication in time. Finally, we used several regression models and compared their regression errors using the K-fold cross-validation technique (Poldrack et al. (2019)). Fig. 1 shows an illustration of our machine learning system. The code for analysis was written in Python programming language, and we used the open-source Python libraries for data analysis. Additionally, we used Pandas and Numpy (Harris et al. (2020)) library for data loading and preprocessing and Scikit-Learn (Abraham et al. (2014)) library to implement regression algorithms.

2.4.2. Regression models—We used several known machine learning algorithms for the regression problem and the linear regression model for our baseline predictor. The Decision Tree Regression model uses CART (Classification and Regression Trees) (Reynolds et al. (2019)), which constructs binary trees using the features and threshold that yield the most important information gained at each node. We used mean absolute error as our metric for assessing the quality of splits. The mean absolute error minimizes the L1 loss using the median of each terminal node (Poldrack et al. (2019)). The Support Vector Regression model (Huang et al. (2018)), an extended version of the Support Vector Classifier, is a machine learning regression algorithm to maximize the model's generalization error. The Adaboost Regression (Patterson et al. (2019)) and Gradient Tree Boosting (Zhang et al. (2019)) models are meta-estimators that utilize ensemble learning to enhance a learning algorithm's performance. We began by fitting the base regressor to the original dataset and then attaching additional regressor instances to the same dataset while changing the weights according to the prediction error. This learning technique led the predictor to focus more on difficult data points. We used the decision tree regressor as the base regressor for both Adaboost and Gradient Tree Boosting algorithms (Patterson et al. (2019); Zhang et al. (2019)).

We have performed 4-fold cross-validation, which partitioned 25% of the participants as our validation set. Such splitting rules out the risk of bias to a larger extent, and all samples can equally contribute to both training and testing. Hyperparameters like learning rate and max depth were selected from a bag of options, and the set that produced the best result was finally chosen.

2.4.3. Selecting the most informative subset of questions—In this step, we expanded our research to answer the question, "What are the best N questions (among all queries) that are most useful to the prediction model?" Initially, this issue appeared to be an extension of the previous step as we can pick a subset of questions used in each questionnaire. However, the number of possible subsets of all objects was very large to

compute. Although if we have $\binom{Q}{N}$ questions and want to find N questions, we can use to pick these N questions in various ways, each combination requires us to train regression models separately, which is an impractical approach. This process is known as a feature selection challenge in the machine learning literature, and we address it here with two standard feature selection algorithms (Haq et al. (2020)).

First, we performed the univariate linear regression tests and selected the top N features with the highest F-values (Radovic et al. (2016)). Second, we used a more advanced technique, known as Recursive Feature Elimination (RFE) to select features by recursively evaluating smaller and smaller feature sets. This technique starts with an initial regression model fed with all data points to instigate n-fold cross-validation and uses regression-dependent feature importance to remove the least important feature in each selection iteration. This iterative process continues until the desired number of features have been achieved.

3. Results

3.1. Study sample

Forty individuals (16 male, 24 female) met the eligibility criteria and participated in the study between October 2018 and April 2019. The participants' average age was 52.2 years (SD, 14.8). Half of the participants were Asians, 30% were Caucasians, 10% were Hispanic, and 10% were African Americans. Sixty-seven percent of the participants were university graduates (Table 1). The average number of medications per day was 2.8 pills (SD 0.5).

3.2. Predicting adherence using all questionnaires

We used the first model to predict medication adherence using all baseline questionnaire information. We used the Mean Absolute Error (MAE) (Turnbull et al. (2019)), a standard regression comparison measure for our models. The recorded values are the MAE's mean and standard deviations over all cross-validation folds (Table 2). The other regression methods (support vector regression and gradient tree regression) outperform linear regression, with the support vector regression model outperforming the others with an error rate of 16.6%.

3.3. Predicting adherence using selected questionnaires

One drawback of using all questionnaires was that the number of features used for regression was higher than the number of data points (i.e., number of participants). As a result, the least square calculations will have a low bias if the number of data points is higher than the number of functions. Conversely, when the number of data points is much lower than the number of characteristics, there may be a lot of variability in the least-squares, resulting in over-fitting and, therefore, weak predictions (Chen et al. (2019)). We used filtering methods to define a subset of features for use in the training and validation of machine learning algorithms (Haq et al. (2020)).

The aim was to see how good each questionnaire was at predicting medication adherence. This method will potentially inform future studies in which only a small number of

instruments are administered, and the system's feature extraction module is modified so that only the questions from each questionnaire are used. We then used the regression models described in the previous section on this dataset, and the best performance model is shown in Table 3. Of all other indicators, the CCI was the most informative.

We tracked back the leading factors behind the selected set of questionnaires. Our findings of the most accurate predictors of medication adherence and nonadherence were intuitive. Medication adherence was closely related to-

1. participants' ability to play an active role in their treatment (i.e., empowerment, engagement),
2. confidence about actions to take in case of an emergency;
3. knowledge about their medications,
4. Having a specific individual as a source of social support.

3.4. Most informative subset of questions

As mentioned earlier, we applied univariate linear regression tests to identify top contributing features with maximum F-values. Table 4 shows that Support Vector Regression, out of all the aforementioned models, produced the best result on the selected top 10 features having highest F-values with an average MAE of 14.17.

Additionally, we set the number of desired features to 10 and used all the regression models described above for RFE. As shown in Table 4, Linear Regression model brought the best performance out of these 10 top contributing features and recorded an average MAE of 12.88. Table 5 lists the most informative features and their relative rankings as well.

4. Discussion

Medication adherence is a complex phenomenon with many causes and associations (El-Saifi et al. (2018); Conn and Ruppar (2017); Capoccia et al. (2016); Patel et al. (2019); Juste et al. (2018); Mondesir et al. (2018)). To allow successful approaches for improving medication adherence, a detailed understanding of facilitators and barriers to medication adherence among people at risk of atherosclerotic cardiovascular disease is necessary. However, most medication adherence trials in this population have had limitations that restricted their usefulness. The lack of multivariate research approaches to evaluate medication adherence is one drawback. We addressed this issue with three machine learning approaches to identify medication adherence predictors in this at-risk population.

As we listed, empowerment, engagement, confidence during emergencies, medication knowledge and social support have dominant impact in medication adherence. In the subsequent sections, we will go through each of these predictors in greater depth and provide healthcare recommen-dations.

Empowerment, which occurs when individuals assume responsibility for their well-being, is vital to shared decision-making. They can then learn to solve their problems with the help of their healthcare provider (Probst et al. (2018)). Empowerment starts with the

provider recognizing that individuals are essentially in charge of their treatment and strives to improve their capacity to think objectively and make independent, informed health decisions (Flynn et al. (2012)). Encouraging individual engagement in healthcare outside of the clinic environment will help consumers appreciate the dynamic interplay of lifestyle decisions, medications, and disease (Kambhampati et al. (2016)). Shared decision making has several benefits and includes the ability to integrate facts and patients' expectations into a consultation; enhancing patient awareness, risk perception accuracy, improving patient-provider communication; and eliminating decisional conflict, feeling uninformed, and improper use of tests and treatments (Hoffmann et al. (2014)). Healthcare providers may be guided through the procedure using a variety of methods. The PARTNER structure proposed by Vale et al. (2003) (Vale et al. (2003)) summarizes the variables essential for successful chronic disease management collaboration (i.e., hypertension). These care management roles include:

1. communicating with patients, staff, and the community;
2. coordinating follow-up care;
3. Overcoming adherence barriers;
4. Monitoring treatment response and progress and 5. informing and involving patients in self-management.

Targeting and enhancing these functions could be necessary to achieve higher drug adherence and improved clinical outcomes.

Our findings suggest that confidence and knowledge are strong predictors of medication adherence, which is not surprising. Similarly, in a study of older adults with chronic disease, self-efficacy – described as the process by which individuals understand their role, are given the knowledge and skills to perform a task by their healthcare provider, and engage in their treatment – was among the top predictors of medication adherence (Patel et al. (2019); Juste et al. (2018); Mondesir et al. (2018); López-Campos et al. (2019)). Strategies that provide clear instructions, guidance, and answers to an individual's most pressing questions will increase their confidence while increasing their willingness to follow the treatment plan. Likewise, recognizing and adhering to medication regimens has long been recognized as requiring knowledge of one's medica-tions (Hoffmann et al. (2014)). However, healthcare providers need to recognize that individuals have varying expectations about the amount of information received and information delivery. Consequently, satisfaction with medication details is a key criterion for assessing an intervention's quality (Hedegaard et al. (2015)). As a result, the field of adherence research has shifted toward new methods that include individualized rather than structured adherence interventions and team-based treatment (Nieuwlaat et al. (2014); Kini and Ho (2018)). Counseling focused on motivational interviewing is one method with increasing evidence for improving medication adherence (Palacio et al. (2016)).

One of this study's key contributions was to explore the impact of social support or having a particular person in one's life on medication adherence. Instinctively, having a particular person in one's life was among the top ten predictors of medication adherence in our

model. Social support comes from siblings, partners, and confidantes who affect health-related decision-making; patients who receive their support will experience psychological relaxation and cope better with health issues. A study in patients with heart failure showed that medication adherence and social support independently and in combination predicted cardiac event-free survival (Wu et al. (2013)). In addition, several studies in patients with chronic conditions confirmed that having multiple friends and relatives was associated with better medication adherence and recovery than those who were less socially integrated (López-Campos et al. (2019); Saffari et al. (2019); Sousa et al. (2019); Ruppar et al. (2015)). Furthermore, social support is often related to subjective well-being; if individuals perceive and receive social support from others (e.g., they feel worried for, welcomed, valued, cared for by other people), then their subjective well-being rapidly increases; they will therefore take a positive attitude toward their chronic diseases (Osborn and Egede (2012)). Interventions that combine systemic and functional elements with psycho-social support can help support medication adherence. However, the mechanism of this relationship warrants further investigation, particularly in our target population of individuals at risk for atherosclerotic cardiovascular disease.

Interventions to enhance medication adherence have been studied extensively for decades; however, even complex interventions have only a modest effect (Nieuwlaat et al. (2014)). One possible reason is that nonadherence is multifactorial, making it impossible to implement a completely successful intervention (Andrews et al. (2017)). A better understanding of the multi-factorial predictors, including modifiable factors of medication adherence, may provide healthcare providers with better leverage to enhance medication adherence. We identified modifiable factors such as uncon-trolled diabetes, the prevalence of persistent discomfort and elevated pain levels, dyspnea, and multiple comorbidities associated with nonadherence to medication. Such observations are not extraordinary or exceptional. Over the past two decades, access to medicines, polypharmacy, multiple comorbidities, and complex treatment regimens have been well-documented and informative predictors of medication nonadherence across the lifespan (Conn and Ruppar (2017); Ruppar et al. (2015); Dibao-Dina et al. (2018)). Additional modifiable predictors of medication adherence that we did not observe in our research included depression, fatigue, health literacy, and impaired patient-provider communication (Hope et al. (2019); Viswanathan et al. (2012)). This result is probably related to the study participants' higher than usual educational preparedness. Nonetheless, our findings confirm what is already well-documented; medication adherence is a complex phenomenon that depends on an interaction of sociodemo-graphic, medical, and psychological factors (El-Saifi et al. (2018); Conn and Ruppar (2017); Capoccia et al. (2016); Hope et al. (2019); Juste et al. (2018); Mondesir et al. (2018); López-Campos et al. (2019); Ruppar et al. (2015); Osborn and Egede (2012); Dibao-Dina et al. (2018); Viswanathan et al. (2012); Moon et al. (2017)).

On a different note, our approach prioritizes overall adherence. We could have studied momentary variables like stress, anxiety or workload etc., with higher priority. We note that objective assessment of such contextual factors is non-trivial and requires significant infrastructure de-velopment, user study design, data collection, and algorithm design. Such an effort was not the focus of our work in this project. However, the self-reported measures (as explained before) may represent various contextual factors that impact a person's daily

life routine and may impact med-ication adherence. For example, the Hospital Anxiety and Depression Scale (HADS) somehow stems from anxiety and depression. The Edmonton Symptom Assessment Scale (ESAS) is influenced by pain, fatigue, drowsiness, and Short Form 12, Version 2 (SF-12v2) is correlated to emotional health and social life. All these modalities are connected to the momentary traits of a person. Nevertheless, some other momentary attributes (e.g., workload, solicitude) which might aid the tendency to forget to take the medication in time have been completely overlooked here, which can be counted as a drawback of this study. A broader understanding of these interactions will facilitate the development of tailored interventions to improve medication adherence, quality of care, and performance for these at-risk individuals.

To our understanding, our research is the first to examine predictors of medication adherence and nonadherence using machine-learning techniques in people at risk for atherosclerotic cardiovascular disease. However, we acknowledge several limitations to our research. First, this study's main limitation is its small sample size, making it very difficult for any endpoints to reach statistical significance. Likewise, half of our small sample were Asians. A larger sample size and heterogeneous group may be appropriate for examining medication adherence predictors and generalizable findings across large populations. Furthermore, the cross-sectional design does not allow the causal relationship between variables to be identified. The third limitation was that we did not measure medication adherence directly. Further research would be conducted to examine the use of multiple adherence assessment methods, such as the evaluation by the health care provider and the Morisky Medication Adherence Scale (Zhang et al. (2016)), which could be compared and aggregated to obtain a single adherence estimate. Finally, our research primarily illustrates the potential utilization of the three machine learning techniques. Further studies should be carried out where the differential strength of our models is contrasted with that of the widely employed logistic regression, the Bayesian network, and the neural network models (Kangi and Bahrampour (2018)).

5. Conclusion

Nonadherence to medication is an important issue with severe consequences for persons at risk of chronic atherosclerotic disease. Linear regression, support vector regression, and Adaboost regression algorithms proved to be more effective than the other methods we tested for predicting medication adherence in individuals at risk for atherosclerotic cardiac disease when many input variables are relative to the number of available observations. Our results have demonstrated clinical and psychosocial factors that can reduce medication nonadherence in this population. These factors can be used by professional clinicians and researchers in stratifying people to reduce the risk of evidence-based decision-making. Predictive models can also be a theoretically important method of treating drug adherence trials in more extensive clinical studies to deal with reduced success/completion rates. Implementing and managing machine learning methods in complex diseases can help researchers improve compliance rates in similar types of studies.

Acknowledgments

This work was supported in part by the National Institute of Health under grant 1R21NR015410-01. However, any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations.

Data availability

The authors do not have permission to share data.

References

Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, Gramfort A, Thirion B, & Varoquaux G (2014). Machine learning for neuroimaging with scikit-learn. Frontiers in Neuroinformatics, 8.

Andrews AM, Russell CL, & lin Cheng A (2017). Medication adherence interventions for older adults with heart failure: A systematic review. Journal of Gerontological Nursing, 43(10), 37–45. [PubMed: 28556872]

Annunziata MA, Muzzatti B, Bidoli E, Flaiban C, Bomben F, Piccinin M, Gipponi K, Mariutti G, Busato S, & Mella S (2019). Hospital anxiety and depression scale (hads) accuracy in cancer patients. Supportive Care in Cancer, 28, 3921–3926. [PubMed: 31858249]

Arnold S, de Lemos JD, Liu Y, Mues K, Bhatt DL, Cannon C, & Kosiborod M (2020). Adherence to guideline medication recommendations to prevent atherosclerotic cardiovascular disease progression among adults with prior myocardial infarction. JAMA Network Open, 3(4), Article e203032.

Bansilal S, Castellano J, Garrido E, Wei HG, Freeman A, Spettell C, García-Alonso F, Lizano I, Arnold RJG, Rajda J, Steinberg G, & Fuster V (2016). Assessing the impact of medication adherence on long-term cardiovascular outcomes. Journal of the American College of Cardiology, 68(8), 789–801. [PubMed: 27539170]

Boyd C (2019). Smart pill bottle which reminds patients to take tablets 'is 100% reliable. URL: https://www.dailymail.co.uk/health/article-7754815/Smart-pill-bottle-reminds-patients-tablets-100-reliable.html.

Bruera E, Kuehn N, Miller MJ, Selmser P, & Macmillan K (1991). The edmonton symptom assessment system (esas): A simple method for the assessment of palliative care patients. Journal of Palliative Care, 7, 6–9. [PubMed: 1714502]

Buck H, Akbar JA, Zhang S, & Bettger J (2013). Measuring comorbidity in cardiovascular research: A systematic review. Nursing Research and Practice, 2013.

Capoccia K, Odegard PS, & Letassy N (2016). Medication adherence with diabetes medication: A systematic review of the literature. The Diabetes Educator, 42, 34–71. 10.1177/0145721715619038. URL. doi:10.1177/0145721715619038. [PubMed: 26637240]

Charlson M, Pompei P, Ales K, & Mackenzie C (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. Journal of Chronic Diseases, 40(5), 373–383. [PubMed: 3558716]

Chen J, de Hoogh K, Gulliver JS, Hoffmann B, Hertel O, Ketzel M, Bauwelinck M, van Donkelaar A, Hvidtfeldt UA, Katsouyanni K, Janssen NAH, Martin RV, Samoli E, Schwartz PE, Stafoggia M, Bellander T, Strak M, Wolf K, Vienneau D, … Hoek G (2019). A comparison of linear regression, regularization, and machine learning algorithms to develop europe-wide spatial models of fine particles and nitrogen dioxide. Environment International, 130, Article 104934.

Conn V, & Ruppar T (2017). Medication adherence outcomes of 771 intervention trials: Systematic review and meta-analysis. Preventive Medicine, 99, 269–276. [PubMed: 28315760]

Cutler R, Fernandez-Llimos F, Frommer M, Benrimoj CSI, & Garcia-Cardenas V (2018). Economic impact of medication non-adherence by disease groups: A systematic review. BMJ Open, 8.

Dahlem NW, Zimet GD, & Walker RR (1991). The multidimensional scale of perceived social support: A confirmation study. Journal of clinical psychology, 47 (6), 756–761. [PubMed: 1757578]

Dambi JM, Corten L, Chiwaridzo M, Jack HE, Mlambo T, & Jelsma J (2018). A systematic review of the psychometric properties of the cross-cultural translations and adaptations of the multidimensional perceived social support scale (mspss). Health and Quality of Life Outcomes, 16.

Dibao-Dina C, Angoulvant D, Lebeau JP, Peurois J, Hirtsi KAE, & Lehr-Drylewicz AM (2018). Patients' adherence to optimal therapeutic, lifestyle and risk factors recommendations after myocardial infarction: Six years follow-up in primary care. PLoS One, 13.

Du L, Cheng Z, Zhang Y, Li Y, & Mei D (2017). The impact of medication adherence on clinical outcomes of coronary artery disease: A meta-analysis. European Journal of Preventive Cardiology, 24, 962–970. [PubMed: 28436725]

El-Saifi N, Moyle W, Jones C, & Tuffaha H (2018). Medication adherence in older patients with dementia: A systematic literature review. Journal of Pharmacy Practice, 31, 322–334. [PubMed: 28539102]

Evangelista LS, Lee J-A, Moore AA, Motie M, Ghasemzadeh H, Sarrafzadeh M, & Mangione CM (2015). Examining the effects of remote monitoring systems on activation, self-care, and quality of life in older patients with chronic heart failure. Journal of Cardiovascular Nursing, 30, 51–57. [PubMed: 24365871]

Flynn D, Knoedler MA, Hess EP, Murad MH, Erwin PJ, Montori VM, & Thomson RG (2012). Engaging patients in health care decisions in the emergency department through shared decision-making: A systematic review. Academic Emergency Medicine : Official Journal of the Society for Academic Emergency Medicine, 19 8, 959–967. [PubMed: 22853804]

Haq AU, Li J, Khan J, Memon MH, Nazir S, Ahmad S, Khan GA, & Ali A (2020). Intelligent machine learning approach for effective recognition of diabetes in e-healthcare using clinical data. Sensors, 20. Basel, Switzerland.

Harris CR, Millman KJ, van der Walt S, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del R'io JF, Wiebe M, Peterson P, … Oliphant TE (2020). Array programming with numpy. Nature, 585, 357–362. [PubMed: 32939066]

Hedegaard U, Kjeldsen LJ, Pottegård A, Henriksen JE, Lambrechtsen J, Hangaard J, & Hallas J (2015). Improving medication adherence in patients with hypertension: A randomized trial. The American Journal of Medicine, 128 12, 1351–1361. [PubMed: 26302142]

Herrero MJ, Blanch J, Peri J, de Pablo J, Pintor L, & Bulbena A (2003). A validation study of the hospital anxiety and depression scale (hads) in a Spanish population. General Hospital Psychiatry, 25(4), 277–283. [PubMed: 12850660]

Hibbard JH, Stockard J, Mahoney ER, & Tusler M (2004). Development of the patient activation measure (pam): Conceptualizing and measuring activation in patients and consumers. Health Services Research, 39(4 Pt 1), 1005–1026. [PubMed: 15230939]

Hoffmann TC, Ĺegaŕe F, Simmons MB, McNamara KP, McCaffery KJ, Trevena LJ, Hudson B, Glasziou PP, & Mar CD (2014). Shared decision making: What do clinicians need to know and why should they bother? Medical Journal of Australia, 201.

Hope H, Binkley GM, Fenton S, Kitas G, Verstappen S, & Symmons D (2019). Systematic review of the predictors of statin adherence for the primary prevention of cardiovascular disease. PLoS One, 14.

Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, & Xu W (2018). Applications of support vector machine (svm) learning in cancer genomics. Cancer genomics & proteomics, 15 1, 41–51. [PubMed: 29275361]

Huber C, Meyer M, Steffel J, Blozik E, Reich O, & Rosemann T (2019). Post-myocardial infarction (mi) care: Medication adherence for secondary prevention after mi in a large real-world population. Clinical Therapeutics, 41(1), 107–117. [PubMed: 30591287]

Hui D, & Bruera E (2017). The edmonton symptom assessment system 25 years later: Past, present, and future developments. Journal of Pain and Symptom Management, 53(3), 630–643. [PubMed: 28042071]

Jakobsen JC, Gluud C, Wetterslev J, & Winkel P (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. BMC Medical Research Methodology, 17.

Juste AM, Miguel AG, Plou BP, Rubio FG, Pascual-Salcedo MMA, Menditto E, & Torres AP (2018). Adherence to treatment of hypertension, hypercholesterolaemia and diabetes in an elderly population of a Spanish cohort. Medicina Clínica, 153(1), 1–5. [PubMed: 30503066]

Kambhampati S, Ashvetiya T, Stone NJ, Blumenthal RS, & Martin SS (2016). Shared decision-making and patient empowerment in preventive cardiology. Current Cardiology Reports, 18, 1–7. [PubMed: 26694723]

Kangi AK, & Bahrampour A (2018). Predicting the survival of gastric cancer patients using artificial and bayesian neural networks. Asian Pacific Journal of Cancer Prevention : Asian Pacific Journal of Cancer Prevention, 19, 487–490. [PubMed: 29480983]

Khayyat S, Mohamed MMA, Alhazmi RSH, Korani M, Allugmani EB, Saleh S, Mansouri D, Lamfon QA, Beshiri OM, & Hadi MA (2018). Association between medication adherence and quality of life of patients with diabetes and hypertension attending primary care clinics: A cross-sectional survey. Quality of Life Research, 28, 1053–1061. [PubMed: 30470970]

Kini V, & Ho PM (2018). Interventions to improve medication adherence: A review. JAMA, 320, 2461–2473. [PubMed: 30561486]

Kumamaru H, Lee MP, Choudhry N, Dong Y-H, Krumme A, Khan NF, Brill G, Kohsaka S, Miyata H, Schneeweiss S, & Gagne J (2018). Using previous medication adherence to predict future adherence. Journal of managed care & specialty pharmacy, 24(11), 1146–1155. [PubMed: 30362915]

Lauffenburger J, Franklin J, Krumme A, Shrank W, Matlin O, Spettell C, Brill G, & Choudhry N (2018). Predicting adherence to chronic disease medications in patients with long-term initial medication fills using indicators of clinical events and health behaviors. Journal of managed care & specialty pharmacy, 24(5), 469–477. [PubMed: 29694288]

Liu L, Gou Z, & Zuo J (2016). Social support mediates loneliness and depression in elderly people. Journal of Health Psychology, 21, 750–758. [PubMed: 24925547]

López-Campos JL, Gallego EQ, & Hernández LC (2019). Status of and strategies for improving adherence to copd treatment. International Journal of Chronic Obstructive Pulmonary Disease, 14, 1503–1515. [PubMed: 31371936]

Löwe B, Spitzer R, Gräfe K, Kroenke K, Quenter A, Zipfel S, Buchholz C, Witte S, & Herzog W (2004). Comparative validity of three screening questionnaires for dsm-iv depressive disorders and physicians' diagnoses. Journal of Affective Disorders, 78(2), 131–140. [PubMed: 14706723]

Mathews R, Wang WT, Kaltenbach L, Thomas L, Shah R, Ali MJ, Peterson E, & Wang TY (2018). Hospital variation in adherence rates to secondary prevention medications and the implications on quality. Circulation, 137, 2128–2138. [PubMed: 29386204]

Milton L, Behroozian T, Coburn NG, Trudeau ME, Razvi Y, McKenzie E, Karam I, Lam H, & Chow E (2020). Prediction of breast cancer–related outcomes with the edmonton symptom assessment scale: A literature review. Supportive Care in Cancer, 29, 595–603. [PubMed: 32918128]

Mondesir FL, Carson AP, Durant RW, Lewis MW, Safford MM, & Levitan EB (2018). Association of functional and structural social support with medication adherence among individuals treated for coronary heart disease risk factors: Findings from the reasons for geographic and racial differences in stroke (regards) study. PLoS One, 13.

Moon SJ, Lee W-Y, Hwang JS, Hong YP, & Morisky DE (2017). Accuracy of a screening tool for medication adherence: A systematic review and meta-analysis of the morisky medication adherence scale-8. PLoS One, 12.

Nieuwlaat R, Wilczynski NL, Navarro T, Hobson N, Jeffery RA, Keepanasseril A, Agoritsas T, Mistry N, Iorio A, Jack SM, Sivaramalingam B, Iserman EC, Mustafa RA, Jedraszewski D, Cotoi C, & Haynes RB (2014). Interventions for enhancing medication adherence. Cochrane Database of Systematic Reviews, 11. CD000011.

Omezzine RG, Akkara A, Koubaa AA, Sriha AB, Rdissi A, & Amamou K (2019). Predictors of poor adherence to hypertension treatment. La Tunisie medicale, 97 4, 564–571. [PubMed: 31729707]

Osborn CY, & Egede LE (2012). The relationship between depressive symptoms and medication nonadherence in type 2 diabetes: The role of social support. General Hospital Psychiatry, 34(3), 249–253. [PubMed: 22401705]

Palacio AM, Garay D, Langer B, Taylor J, Wood BA, & Tamariz L (2016). Motivational interviewing improves medication adherence: A systematic review and meta-analysis. Journal of General Internal Medicine, 31, 929–940. [PubMed: 27160414]

Patel S, Abreu M, Tumyan A, Adams-Huet B, Li X, & Lingvay I (2019). Effect of medication adherence on clinical outcomes in type 2 diabetes: Analysis of the simple study. BMJ Open Diabetes Research & Care, 7.

Patterson BW, Engstrom C, Sah VK, Smith MA, Mendonça EA, Pulia MS, Repplinger MD, Hamedani AG, Page D, & Shah MN (2019). Training and interpreting machine learning algorithms to evaluate fall risk after emergency department visits. Medical Care, 57, 560–566. [PubMed: 31157707]

Pillsy ( ). Connected medication platform, with smart pill bottle, app and secure cloud-based platform. URL: https://www.pillsy.com/.

Poldrack RA, Huckins G, & Varoquaux G (2019). Establishment of best practices for evidence for prediction: A review. JAMA Psychiatry.

Probst MA, Noseworthy PA, Brito JP, & Hess EP (2018). Shared decision-making as the future of emergency cardiology. Canadian Journal of Cardiology, 34 (2), 117–124. [PubMed: 29289400]

Radovanovic D, Seifert B, Urban P, Eberli FR, Rickli H, Bertel O, Puhan MA, & Erne P (2013). Validity of charlson comorbidity index in patients hospitalised with acute coronary syndrome. insights from the nationwide amis plus registry 2002–2012. Heart, 100, 288–294. [PubMed: 24186563]

Radovic MD, Ghalwash MF, Filipovic ND, & Obradovic Z (2016). Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. BMC Bioinformatics, 18.

Reynolds EL, Callaghan BC, & Banerjee M (2019). Svm–cart for disease classification. Journal of Applied Statistics, 46, 2987–3007. [PubMed: 33012942]

Ruppar TM, Delgado JM, & Temple J (2015). Medication adherence interventions for heart failure patients: A meta-analysis. European Journal of Cardiovascular Nursing, 14, 395–404. [PubMed: 25669661]

Saffari M, Lin C-Y, Chen HF, & Pakpour AH (2019). The role of religious coping and social support on medication adherence and quality of life among the elderly with type 2 diabetes. Quality of Life Research, 1–11.

Sansom GT, Kirsch KR, & Horney JA (2020). Using the 12-item short form health survey (sf-12) to assess self rated health of an engaged population impacted by hurricane harvey, houston, tx. BMC Public Health, 20.

Shou J, Ren L, Wang H, Yan F, Cao X, Wang H, Wang Z, Zhu S, & Liu Y (2015). Reliability and validity of 12-item short-form health survey (sf-12) for the health status of Chinese community elderly population in xujiahui district of shanghai. Aging Clinical and Experimental Research, 28, 339–346. [PubMed: 26142623]

Sousa H, Ribeiro Ó, Paúl C, Costa E, Miranda V, Ribeiro F, & Figueiredo D (2019). Social support and treatment adherence in patients with end-stage renal disease: A systematic review. Seminars in Dialysis, 32, 562–574. [PubMed: 31309612]

Terluin B, Brouwers EPM, van Marwijk H, Verhaak PF, & van der Horst HE (2009). Detecting depressive and anxiety disorders in distressed patients in primary care; comparative diagnostic accuracy of the four-dimensional symptom questionnaire (4dsq) and the hospital anxiety and depression scale (hads). BMC Family Practice, 10, 58–58. [PubMed: 19698153]

Turnbull AMJ, Crawford GJ, & Barrett GD (2019). Methods for intraocular lens power calculation in cataract surgery after radial keratotomy. Ophthalmology.

Vale MJ, Jelinek MV, Best JD, Dart AM, Grigg LE, Hare DL, Ho BP, Newman RW, & McNeil JJ (2003). Coaching patients on achieving cardiovascular health (coach): A multicenter randomized trial in patients with coronary heart disease. JAMA Internal Medicine, 163, 2775–2783.

Viswanathan M, Golin CE, Jones CD, Ashok M, Blalock SJ, Wines RCM, Coker-Schwimmer EJL, Rosen DL, Sista P, & Lohr KN (2012). Interventions to improve adherence to self-administered

medications for chronic diseases in the United States. Annals of Internal Medicine, 157, 785–795. [PubMed: 22964778]

Ware J, Kosinski M, & Keller SD (1996). A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. Medical care, 34 3, 220–233. [PubMed: 8628042]

Wu J-R, Frazier SK, Rayens MK, Lennie TA, Chung ML, & Moser D (2013). Medication adherence, social support, and event-free survival in patients with heart failure. Health Psychology : Official Journal of the Division of Health Psychology, American Psychological Association, 32(6), 637–646. [PubMed: 22746258]

von Wyl V, Ulytė A, Wei W, Radovanovic D, Grübner O, Brüngger B, Bähler C, Blozik E, Dressel H, & Schwenkglenks M (2020). Going beyond the mean: Economic benefits of myocardial infarction secondary prevention. BMC Health Services Research, 20.

Zhang X, Yuan Z, Ji J, Li H, & Xue F (2016). Network or regression-based methods for disease discrimination: A comparison study. BMC Medical Research Methodology, 16.

Zhang Z, Zhao Y, Canes A, Steinberg D, & Lyashevska O (2019). Predictive analytics with gradient boosting in clinical medicine. Annals of Translational Medicine, 7(7), 152. [PubMed: 31157273]

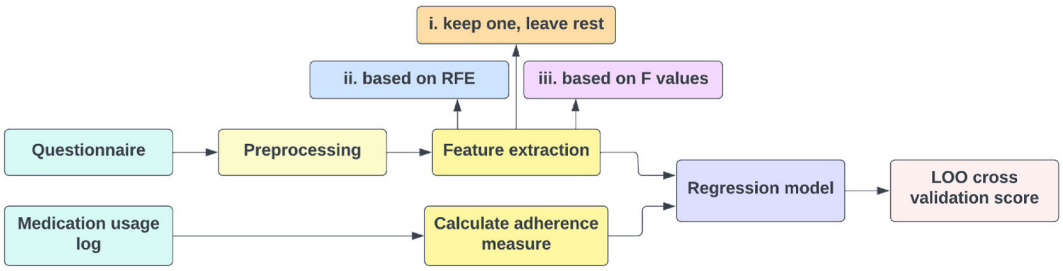

Fig. 1.

Machine learning pipeline for adherence prediction. Based on the requirements, we applied three types of feature extraction: i. *keep one, leave rest* was employed to study the individual effect of each questionnaire, ii. *RFE* and *F values* based rankings were done to identify the relative contribution of the features among all.

Table 1

Participant demographics.

	Mean ± SD	
Age	52.2 ± 14.8	
	Frequency	**Percentage**
Gender		
Male	16	40.0
Female	24	60.0
Ethnicity		
White or Caucasian	12	30.0
Asian	20	50.0
Hispanic/Latino	4	10.0
Mixed	4	10.0
Marital Status		
Single	9	22.5
Married	24	60.0
Divorced/Separated	4	10.0
Widowed	3	7.5

Table 2

Comparison of different regression models.

Method	Mean Absolute Error (Mean ± SD)
Support Vector Regression	16.61 ± 3.89
Random Forest Regression	18.911 ± 4.49
Gradient Tree Boosting Regres-sion	19.42 ± 4.65
AdaBoost Regression	20.35 ± 4.06
Decision Tree Regression	21.98 ± 4.01
Linear Regression	36.63 ± 7.62

Table 3

Comparison of selected questionnaires.

Questionnaire	Mean Absolute Error Mean ± SD.	Best Performing Model
CCI	14.7 ± 2.97	AdaBoost Regression
PAI	15.16 ± 3.51	Gradient Boosting Regression
Demo	16.47 ± 3.93	Support Vector Regression
PSS	16.64 ± 5.1	Support Vector Regression
QOL	16.69 ± 3.93	Support Vector Regression
HADS	16.77 ± 3.75	Support Vector Regression
ESAS	16.81 ± 3.72	Support Vector Regression

Table 4

Comparison of different feature selection methods.

Feature Selection Algorithm	Min MAE	Best Performing Model
Recursive Feature Elimination	12.88 ± 4.58	Linear Regression
Top 10 F-Value Scores	14.17 ± 5.4	Support Vector Regression

Table 5

Most informative features.

Rank	Feature	Description
1 – 3	Confide PAI2 PAI3	Demographic form question: Do you have someone to confide in?
		Taking an active role in my health care is an essential factor in determining my health and ability to function I am confident that I can take actions that will help prevent or minimize some symptoms or problem associated with my health
4 – 5	CCI11	DCCI form question: I have/had diabetes
	PAI4	I know what each of my prescribed medications does
6–10	CCI total ESAS_pain	The overall number of comorbidities ESAS form: the degree of feeling pain ESAS form: the degree of feeling dyspnea
	ESAS_dyspnea PSS10	PSS Form: There is a special person in my life who care about my feelings
	CCI3	CCI form: I have/had Unresolved pain