Peer Review File

Article information: https://dx.doi.org/10.21037/jmai-21-26

Reviewer A

The manuscript "Application of machine learning in predicting medication adherence of patients with cardiovascular diseases: a systematic review of the literature" has the aim to systematically review the literature from 2000 to August 2021 for studies that use machine learning for predicting medication adherence in patients with cardiovascular disease (those with coronary syndromes and heart failure) and in patients with cardiovascular risk factors (hypertension and hypercholesterolemia). Of 32 studies reviewed and 12 studies selected for critical appraisal, there were 2 studies on heart failure, 3 on coronary syndromes, and 7 studies that evaluated medication adherence on risk factors hypertension and hypercholesterolemia. The authors looked for accuracies of the models built with machine learning and also for comparisons with models from more conventional methods (mostly logistic regression analysis). One side of the conclusion is that studies using machine learning for predictive models of medication adherence may have the advantage that a larger number of predictive variables may be used than with conventional prediction models. On the other hand, machine learning models for medication adherence do not seem to improve in accuracy compared to conventional models and lack external validation.

The manuscript is written at a time when machine learning still has to prove its value for clinical use, and therefore the review is interesting.

Main comment:

Comment 1.

My main comment is that the discussion is not a comprehensive evaluation of findings, but a repeated results analysis. These results and critical appraisal should already be in the tables, some of them to be highlighted in the text of results.

There is an obvious difficulty at the start of your discussion to decide what the further aim of the review should be, whether to distinguish clinical conditions or discuss how many variables may be present in a method, or whether to discuss the comparison with more conventional statistical methods. It would be helpful if the authors would decide on these items and their order of discussion and still try to comprehensively discuss the findings without repeating the results section too much.

My suggestion (and only to provide some help) would be to start discussion with the main conclusion: medication adherence does not seem to be better predicted by ML techniques than by conventional statistical techniques in those studies that predefined the number of variables that could be tested for each model. Where ML methods seemed to be better, variables are added such as previous refill data (at least, that is what you suggest, but make it clear that this explained a number of inconsistent comparisons and name the studies as refs). Then discuss the various interpretations/ definitions of medication adherence and the way it may influence the performance of the models, are there examples you found (studies ref 32, and are medication adherence

and medication persistence something else: time period between initiation and last dose, as defined in reference 35, so a term used for several other reasons of non-adherence) and do we agree on the best definition for medication adherence (reference 16, 17 and 18 in your reference 35, but also a reference on diabetes medications * see bottom of this commentary). Then discuss the advantage of ML techniques for the possibility to include more variables (your reference 35 has a nice discussion to compare with). Was this the case in the studies, that more variables were included in the ML methods? Was there any difference in the kind of variables to be used in ML techniques compared to conventional statistics? Then it remains possible to discuss the clinical conditions for whom the models are made, is this an important distinction only from the perspective of outcome and association with medication adherence? Is seems that patients using more medications are better in medication adherence, possibly by some learning effect. And you could have added studies on diabetes medications (for example * at the bottom of this commentary), which is actually a minor limitation of your study, since you may need more studies that apply ML methods to have some insight in larger data sets to see how many variables would be useful.

Reply 1.

We appreciate the comment and made changes accordingly.

The first paragraph of the discussion was replaced with two paragraphs evaluating the studies based on their ML model comparison with conventional methods and the other paragraphs focused on the effect of method used for medication adherence on the accuracy of the ML models.

Regarding your suggestion for adding studies on diabetes, we decided not to change our inclusion criteria due to the large number of studies on diabetes.

We also checked the suggested study but in addition to the different disease state, it did not meet other inclusion criteria.

Changes in the text:

The following part was removed and replaced with two other paragraphs on pages 12-14 to cover the reviewer suggestions.

~~In this review, we identified 12 studies that utilized ML methods to investigate MA among patients with CVD (CAD and HF) and their risk factors HTN and HCL. The earliest study was published in 2010 and used a Support Vector Machine (SVM) model to evaluate MA among HF patients (31). Later in 2016, there was another study on HF using ML (30). Five studies evaluated MA among HTN patients (37–41). Three studies evaluated MA among patients recently hospitalized due to ACS or MI (34–36). Two studies evaluated MA among patients using statins for HCL treatment (32,33). One study evaluated MA in a sample of elderly suffering from various chronic conditions including HTN (37).~~

ML models predicting MA compared to conventional methods

Based on our review, it seems that ML methods do not have better predictive values

compared to conventional statistical techniques. In studies that used similar predictive variables, there was no noticeable difference between the accuracy of the models (32,35,36). In one of the studies that the accuracy of the ML model was significantly higher than the conventional methods (0.97 vs 0.71), there was an internal validation issue (37). In Lee et al. study, they tested the ML model with the same dataset previously used for model training and it resulted in a biased high accuracy. In two other studies that the ML model had higher accuracies, several variables related to patient's past medical history were added to ML models that were not included in the LR models (34,36). At first glance it seems that using ML models provided the opportunity of adding several predictive factors to the model. However, evaluation of previous literature indicates that utilization of patients' MA history as a predictor of MA had previously been used in LR models. One of the studies in this review that used logistic regression with and without MA history also reported that inclusion of MA history will enhance the predictive accuracy of the model even without utilization of ML methods (32).

ML models predicting MA based on MA measurement tools

Several methods with different levels of accuracy were used for MA measurment in studies. Five studies used PDC which is among the most common measured parameters in pharmacy claim databases for MA calculation (32,33,35,39,40). Among these five studies, only two of them compared their ML model accuracy with the conventional methods and did not report any significant differences (32,35). One study used medication refill gap of >120 days as a definition for non-adherence (36) and reported higher accuracies in their ML models. First, studies have shown that refill gap as a MA measurement tool is less strongly correlated with patient outcomes in comparison with PDC (45). On the other hand, a more careful comparison of the variables included in their models indicated that several variables based on patients' MA history were added to the ML model but not the standard LR model (36). Four studies used self-reported methods (31,34,37,38) that generally tend to overestimate adherence due to recall or reporting bias (44). Among these four studies, only one of them compared their ML model accuracy to a LR model (37). While they reported a dramatically higher accuracy for their ML model in comparison to their LR model, their finding was affected by an internal validation issue. Lee et al. tested the accuracy of their ML model using the same dataset previously used for model training (37). Two studies used healthcare provider's opinion and report to estimate MA among patients, but none of them compared their model accuracy to other statistical methods (30,41).

Comment 2.

Introduction: page 4 lines 10 – 21, page 5 lines 1-17, are redundant, explaining the prevalence, mortality, and costs of the clinical conditions. The manuscript is about medication adherence and models, so what does the literature about medication adherence in these conditions reveal: this starts at page 5 line 17.

Reply 2.

Thank you for the great comment. We removed the information about the disease mortality and cost. Instead, we added some information about the association between non-adherence and the outcomes of each disease state on pages 4 and 5.

Changes in the text:

"American Heart Associate reported that CAD leading to acute coronary syndrome (ACS), was responsible for 365,914 deaths in the US in 2017 (4). A metanalysis done in 2017 indicated that good adherence to evidence-based medications for CAD, was associated with a significantly lower probability of all-cause mortality (risk ratio 0.56; 95% CI: 0.45-0.69), mortality due to cardiovascular disease (risk ratio 0.66; 95% CI: 0.51-0.87), and cardiovascular hospitalization (risk ratio 0.61; 95% CI: 0.45-0.82) (5).
HF is a growing health concern affecting more than 20 million people globally (6). The results of a study on 219 HF patients indicated that HF patients with poor medication adherence had higher risk of reporting dyspnea or ankle swelling (p=.05) (7). Using a cox model, We et al. reported a worse cardiac event-free survival associated with poor medication adherence (p=.006) (7).
HTN is a major risk factor for CVD (8). About half of adults in the US (108 million) have HTN or are using medication for HTN while only one out of four HTN patients have their blood pressure in the acceptable range (9). A recent metanalysis (54,349 patients) assessed the association of antihypertensive medication adherence to the recurrence of CVD events (10). They found a 9% reduction in the risk of CVD events for each 20% increase in adherence to antihypertensive medications (10).
HCL is a very common condition in the US. According to the CDC, 73.5 million (31.7%) of US adults have LDL on higher-than-normal range (11). A metanalysis that included 710,504 patients assessed the incidence of stroke among patients using statins (12). Xu T et al. did a dose-response analysis that indicated an 8% reduction in the total risk of stroke for each 20% improvement in statin adherence (12)."

Comment 3.

Please define medication adherence in line 19 page 5.

Reply 3.

Thank you for the great comment. We added the definition that was provided by WHO on page 5.

Changes in the text:

"Medical conditions explained above are modifiable with medications. However, maximum benefit is not possible unless patients have enough adherence to medications. According to the World Health Organization, medication adherence is "the extent to which a person's behavior – taking medication, following a diet, and/or executing lifestyle changes, corresponds with agreed recommendations from a health care provider" (13).

Comment 4.

What does it mean that a 'condition is under control'? (line 22-23 page 5) What percentage of the 80% that is not under control is explained by medication nonadherence?

Reply 4.

Thank you for the thoughtful comment. We agree with the reviewer about this sentence not being clear and removed it.

Changes in the text:

~~They found that only 20% of CVD patients have the condition under control (3).~~

Comment 5.

What are most important factors determining medication nonadherence (lines 1-3 page 6, from reference 16). You repeat mentioning these factors In lines 9-13 page 6 (from references 19-21, 22,23), and therefore these sentences can be merged? Since these factors may have some value, are they also to be compared in your present review: were all factors deemed important also present in the studies of your manuscript, and then the question may arise whether the number of important factors present in the studies of your manuscript determined the accuracies more than the methods of the prediction models?

Reply 5.

We agree with the reviewer and removed some repetitive parts. The repetitive parts were merged. Regarding the adherence predictive factors, we did not notice any association between the accuracy of the models and the number of predictive factors, or the machine learning method used. However, we noticed and explained that inclusion of patients' adherence history resulted in an increase in the accuracy of models. You may see the change on page 5.

Changes in the text:

"Instead, it seems that the role of factors not related to medications are more prominent (14)."

Comment 6.

The critical appraisal items are mentioned in "Model validation and risk of bias" page 10 lines 5-8. In your PRISMA 2020 Checklist you refer 'The Risk of bias of studies'(item 18) to page 12 (11-21) and page 13 (1-4). On page 13 the 'Risk of bias' as part of results is presented, but this result is about lack of validation problems. Risk of bias would mean: selection bias (Medicaid data, age selection, selection of patients previously treated with same medication), reporting bias, attrition bias- were all patients still present for analysis. You do not have to present all, but in any case, define what you mean by bias. As a part of the critical appraisal, I would suggest to mention the number of variables per study that are already known to predict medication adherence.

Reply 6.

We thank the reviewer for the comment. To address the suggestions, we modified the related paragraph. Because the study was focused more on the modeling and the predictive accuracy of each model, we were more focused on the variables included in each model and the validation methods. We did add in the limitations that sampling bias was not comprehensively assessed. Reader can go back to each study and evaluate other biases. You may see the added parts on page 9 line 12-14, and page 12 lines 5-12.

Changes in the text:

~~To assess the risk of bias in each study, sample size, validation techniques, alternative statistical methods used in each study, and adherence measurement tools in each study were evaluated.~~

To assess the validation in each study, validation techniques and alternative statistical methods used in each study were evaluated. To estimate the risk of bias in each study, we evaluated sample size, inclusion criteria, length of follow up, MA measurement methods, and number and types of variables used in each study.

~~None of the studies were externally validated and only five studies attempted more than one method, specifically LR, as a comparison for the ML model (32,34–37). Among these five studies, one of them found a 0.97 accuracy in the ML model which was 0.26 higher than the accuracy of their LR model (0.71) (37). The huge difference between the performance of the two models was most likely attributable to the utilization of the same data for training and testing the model. Two other studies that performed simple cross-validation and used separate data for training and testing the model, observed about 0.09 higher accuracy in ML models compared to LR model (34,36). Careful evaluation of these two studies revealed that several variables used in ML were missing from the LR model. In one of these studies by Hu et al., several variables regarding patients' previous MA were included in ML model, while no data regarding previous MA was included in the LR model (36). In Franklin et al. study using exact similar variables, they found no significant improvement after using boosting techniques to improve the model performance (32). Lastly, in Zullig et al. study with an incredibly large sample size (n=11,969) and using identical variables, they were not able to identify any significant superiority for their ML model compared to their LR model (35).~~

Risk of bias

The main source of bias in the studies was the MA measurement methods. Several methods with different levels of accuracy were used for MA calculation in studies. Five studies used PDC which is among the most common measured parameters in pharmacy claim databases for MA calculation (32,33,35,39,40). Four studies used self-reported surveys (31,34,37,38) that generally tend to overestimate adherence due to recall or reporting bias (44). One study used healthcare provider's opinion (30), and one study

used healthcare staff report to estimate the MA (41). Lastly, one study used medication refill gap of >120 days as a definition for non-adherence (36).

Minor comments:

Comment 7.

Page 7, lines 8-9: this sentence is repeated in lines 12-13, therefore maybe skip lines 8-9

Reply 7.

Thank you for catching the mistake. We removed lines 8 and 9.

Changes in the text:

~~We present the following article in accordance with the PRISMA reporting checklist.~~

Comment 8.

Page 9, lines 4-7: this sentence "Alternative search terms…. Appendix 1" should be on page 8 after line 15.

Reply 8.

Thank you for the comment. The following part was removed and placed on the suggested line. You may see the change on page 8 lines 4 and 5.

Changes in the text:

"Alternative search terms used for ML were obtained from previous literature (29). A list of all search terms used for ML and MA is provided in Appendix 1."

Comment 9.

Page 20, Limitations: this study has limitations, to underline the present discussion.

Reply 9.

Thank you for your comment. We changed the sentence and added to the list of limitations. You may see the added limitations on page 20 lines 11-14.

Changes in the text:

~~Several limitations should be considered regarding studies using ML to predict MA including small sample sizes in many of the studies, unreliable validation techniques used for internal validation, unreliable methods used for MA measurement, and above all, lack of external validation for all studies.~~

This study has limitations that needs to be addressed. First, this review was only focused on publications in English. Second, due to the small number of studies with several differences in variables and techniques in each health state, a limited comparison between different ML models in each health state was carried out. This review was based on the accuracy of the ML models used in each study. Hence, the evaluation of

studies was limited to sample size of the studies, validation techniques used for internal validation, reliability of methods used for MA measurement, and above all, presence of external validation.

Comment 10.
Page 21, line 6:
"We found limited studies": what it means: a limited number of studies, or studies with limitations.

Reply 10.
Thank you for the comment. We meant a limited number of studies and made changes accordingly.

Changes in the text:
"We found a limited number of studies that developed predictive models using ML to predict medication adherence among patients with cardiovascular diseases and its risk factors namely hypertension and hypercholesterolemia."

Comment 11.
Table 1, abbreviations legend: EBCD should be EBCM

Reply 11.
Thank you for catching the mistake. We made changes accordingly.

Changes in the text:
"EBCD" was changed to "EBCM"

Comment 12.
Table2: the ML method on top is not what is finally in the columns, which also contain the compared statistical methods. Maybe adapt the top row term?

Reply 12.
We agree with the reviewer. The column covers the information about the modeling method used and in some cases the information about standard regression is also included in this column. Hence, we decided to change the column name to *Modeling method details.*

Changes in the text:
"ML method" was changed to "Modeling method details"

* Lo-Ciganic WH, Donohue JM, Thorpe JM, Perera S, Thorpe CT, Marcum ZA, Gellad WF. Using machine learning to examine medication adherence thresholds and risk of hospitalization. Med Care. 2015 Aug;53(8):720-8. doi: 10.1097/MLR.0000000000000394. PMID: 26147866; PMCID: PMC4503478.

Reviewer B

This manuscript is a targeted review of English literature undertaken in PubMed and Google Scholar from January 1, 2000, to August 9, 2021. All studies used machine learning to predict medication adherence among patients with chronic cardiovascular diseases. The topic is interesting and may contribute to the body of knowledge. The paper is well written and will certainly attract the attention of many other researchers.
I would like to know why the authors did not search for references in IEEE, ACM, or MDPI bases. These bases are not directly linked to medicine, but they have many good papers about machine learning medical applications. As an example, the article https://doi.org/10.3390/s19204539 has a long analysis of related works covering medication adherence.
Despite that, in my opinion, the manuscript is good. Congratulations to the authors.

Reply:

We do appreciate the comments. We used Pubmed because Pubmed is the most commonly used database in medical field. We also added Google Scholar to provide a better coverage on studies in the field of computer science. The databases that are mentioned by the reviewer are not known among researchers in medical field and are mostly used in computer science.

Regarding the paper that was suggested by the reviewer, we checked it and found that the study does not meet the inclusion criteria of our review. The paper “Commercial Devices-Based System Designed to Improve the Treatment Adherence of Hypertensive Patients” is about a device-based system to enhance medication adherence.