Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View -- Checklist

Adapted from: Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, Shilton A, Yearwood J, Dimitrova N, Ho TB, Venkatesh S. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. Journal of medical Internet research. 2016 Dec 16;18(12):e5870. doi: 10.2196/jmir.5870

For each row, indicate the page number where the topic is addressed, or enter "NA" if not applicable.

Item Number	Section	Topic	Checklist Item	Page
1	Title	Nature of study	Identify the report as introducing a predictive model	Title page
2	Abstract	Structured summary	Include abstract formatted correctly for *Stroke*, identifying data sources, performance metrics of the predictive models (including point estimates and confidence intervals), and conclusion including the practical value of the models	1-2
3	Introduction	Rationale	Identify the clinical goal, including how the prediction problem may benefit the clinical goal	Page 3, line 78-90
4		Rationale	Review the current practice and prediction accuracy of any existing models	Page 4, line 91-99
5		Objectives	State the nature of study being predictive modeling, defining the target of prediction	Page 4, line100-102
6	Methods	Data availability	Report the conditions under which other investigators can access the models and data, following AHA TOP guidelines at https://www.ahajournals.org/submission-requirements	Page 9, line 202-205
7		Describe the setting	Identify the clinical setting for the target predictive model, including facility type, size, volume, and time period.	Page 4, line 106-120
8			Determine that the study is retrospective or prospective.	Page 4, line 108-109; Page 5, line 114-116
9		Define the prediction problem	Determine the form of the prediction model: (1) classification if the target variable is categorical, (2) regression if the target variable is continuous, (3) survival prediction if the target variable is the time to an event.	Page 6, line 138-151
10			Define the success criteria for prediction (e.g., based on metrics in internal validation or external validation) in the context of the clinical problem	Page 7, line 173-180
11		Prepare data for model building	Identify relevant data sources for derivation and validation and quote the ethics approval number for data access.	Page 6, line 152-172
12			State the inclusion and exclusion criteria for data.	Page 4, line 111-112; page

				5, 117-119.
13			Describe the time span of data and the sample size.	Page 4, line 110; page 9, 208-214
14			Define the observational units on which the response variable and predictor variables are defined.	Page 7, line 173-180
15			Define the predictor variables. Caution is needed to prevent information leakage from the response variable to predictor variables	n/a
16			Describe the data preprocessing, including data cleaning and transformation. State any criteria used for outlier removal.	Page 6, line 152-172
17			State how missing values were handled and the amount of missing data.	n/a
18			Describe the basic statistics of the datasets, including derivation and validation, particularly of the response variable. These include the ratio of positive to negative classes for a classification problem and the distribution of the response variable for regression problem.	Page 7, line 173-180
19			Define the model validation strategies. Internal validation is the minimum requirement; external validation should also be performed whenever possible.	Page 7, line 173-180
20			Define the validation metrics. For regression problems, use the normalized root-mean-square error. For classification problems, the metrics should include sensitivity, specificity, positive predictive value, negative predictive value, area under the ROC curve, and calibration plot	Page 7, line 173-180
21		Build the predictive model	Report the number of independent variables, the number of positive examples, and the number of negative examples.	Page 10, line 226-238
22			Determine a set of candidate modeling techniques (e.g., logistic regression, random forest, or deep learning). If only one type of model was used, justify the decision for using that model.	Page 6, line 152-172
23			Define the performance metrics to select the best model	Page 7, line 167-172
24			Specify the model selection strategy. Common methods include K-fold validation or bootstrap to estimate the lost function on a grid of candidate parameter values. For K-fold validation, proper stratification by the response variable is needed.	n/a
25			For model selection, include discussion on (1) balance between model accuracy and model simplicity or interpretability, and (2) the familiarity with the modeling	n/a

			techniques of the end user.	
26	Results	Report the final model and performance	Report the predictive performance of the final model in terms of the validation metrics specified in the methods and identify which dataset were used to generate the data.	Page 10, line 220-225; page 10, line 240- 226
27			If possible, report the parameter estimates in the model and their confidence intervals. When the direct calculation of confidence intervals is not possible, report nonparametric estimates from bootstrap samples.	Page 10, line 220-225; page 10, line 240- 226
28			If possible, report what variables were shown to be predictive of the response variable. State which subpopulation has the best prediction and which subpopulation is most difficult to predict.	n/a
29	Discussion	Clinical implications	Report the clinical implications derived from the obtained predictive performance. For example, report the dollar amount that could be saved with better prediction. How many patients could benefit from a care model leveraging the model prediction? And to what extent?	Page 12, line 282-294
30		Limitations	Report limitations including assumptions regarding data formats, pitfalls in interpretation, potential for biases, unexpected signs of coefficients, and generalizability.	Page 15, line 352-375
31		Comparisons	Compare performance to other models in the literature. Base comparisons on confidence intervals, whenever possible.	Page 15, line 335-345