

## Medical ML Model Template:

**Why was this ML model made?** Why was it necessary to create this model? What does it provide that we do not already have?

**What techniques were used to create this model?** Is this model made with Support Vector Machines, Logistic Regression, a Decision Tree, Random Forest, Gradient Boosting, or something else? What is this specific method being used? How does this model address fairness, explainability, and bias?

**How many people were in the study?** How big was this dataset? How does it compare to datasets of other models? Is this dataset still growing? Is the training and testing split 80-20 or a different metric?

**What were the results of this model?** Report using the standard methods described below. Focus on the confusion matrix, recall, precision, and accuracy to create a standardized explanation.

**How does this model compare to other nonalgorithmic assessments and existing algorithms?** Does this model demonstrate improvement compared to humans and other models?

**Is this a viable method?** Is this model a replacement for humans or a supplement? How would this be used as a supplement when with human intervention? Should it be reported with and without human intervention?

**Standardizing Terminology - for experts as well as novices.** These all must be reported at the top of the paper to give a clear understanding of the results of this model.

True Positive (TP): This is an outcome where the model correctly predicts the positive class.

Reported True Positives:

False Positive (FP): Also known as a Type I error, this is an outcome where the model incorrectly predicts the positive class.

Reported False Positives:

True Negative (TN): This is an outcome where the model correctly predicts the negative class.

Reported True Negatives:

False Negative (FN): Also known as a Type II error, this is an outcome where the model incorrectly predicts the negative class.

Reported False Negatives:

Confusion Matrix: This is used to organize and display the TP, FP, TN, and FT. This is a 2x2 grid with TP on the top left, FP on the top right, FN on the bottom right, and TF on the bottom left.

TP:	FP:
FN:	TN:

Accuracy: This is the proportion of true results (both true positives and true negatives) among the total number of cases examined. It's calculated as:

Accuracy=(True Positives + True Negatives) / Total Predictions

Reported Accuracy:

Precision: Also known as positive predictive value, this metric is the ratio of true positives to the sum of true and false positives. It's a measure of a classifier's exactness. A low precision indicates a high number of false positives.

Precision = True Positives / (False Positives + True Positives)

Reported Precision:

Recall: Also known as sensitivity, this metric is the ratio of true positives to the sum of true positives and false negatives. It's a measure of a classifier's completeness. A low recall indicates a high number of false negatives.

Recall = True Positives / (False Negatives + True Positives)

Reported Recall: