

Оценка качества бинарного классификатора в научных исследованиях

© А. А. Ковалев¹, Б. К. Кузнецов¹, А. А. Ядченко¹, В. А. Игнатенко²

¹УО «Гомельский государственный медицинский университет», г. Гомель, Республика Беларусь
²ГП «Институт биохимии биологически активных соединений НАН Беларуси», г. Гродно, Республика Беларусь

РЕЗЮМЕ

В статье рассмотрены основные принципы оценки качества бинарного классификатора посредством возможностей ROC-анализа. Описан алгоритм ROC-анализа и интерпретация полученных результатов. Разобрана методика расчета основных параметров ROC-кривой, а также указания для оценки их статистической значимости.

Ключевые слова: бинарный классификатор, диагностическая модель, ROC-анализ, ROC-кривая, площадь под ROC-кривой, статистическая значимость площади под ROC-кривой, оценка эффективности диагностического теста.

Вклад авторов: Ковалев А.А., Кузнецов Б.К., Ядченко А.А., Игнатенко В.А.: концепция и дизайн статьи, сбор теоретического материала, оформление графического материала, редактирование, обзор публикаций по теме статьи, проверка критически важного содержания.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Источники финансирования: исследование проведено без спонсорской поддержки.

ДЛЯ ЦИТИРОВАНИЯ:

Ковалев АА, Кузнецов БК, Ядченко АА, Игнатенко ВА. Оценка качества бинарного классификатора в научных исследованиях. *Проблемы Здоровья и Экологии*. 2020;4:105–113.

Assessment of the quality of a binary classifier in research

© Aleksey A. Kovalev¹, Boris K. Kuznetsov¹, Aleksey A. Yadchenko¹,
Valeriy A. Ignatenko²

¹Gomel State Medical University, Gomel, Republic of Belarus

²Institute of Biochemistry of Biologically Active Compounds of the National Academy of Sciences of Belarus, Grodno, Republic of Belarus

ABSTRACT

The article deals with the basic principles of the assessment of the quality of a binary classifier using the potential of the ROC analysis and describes a ROC analysis algorithm and the interpretation of the obtained results. The work discusses the methodology for the calculation of the main parameters of the ROC curve, as well as the instructions for the assessment of their statistical significance.

Key words: binary classifier, diagnostic model, ROC analysis, ROC curve, area under the ROC curve, statistical significance of the area under the ROC curve, assessment of the effectiveness of a diagnostic test.

Author contributions: Kovalev A.A., Kuznetsov B.K., Yadchenko A.A., Ignatenko V.A.: concept and design of the article, collection of theoretical material, design of graphic material, editing, review of publications on the topic of the article, critically important content.

Conflict of interest: authors declare no conflict of interest.

Funding: study conducted without sponsorship.

FOR CITATION:

Kovalev AA, Kuznetsov BK, Yadchenko AA, Ignatenko VA. Assessment of the quality of a binary classifier in research. *Health and ecology Issues*. 2020;4:105–113. (in Russ.)

Введение

Бинарным классификатором принято называть такое решающее правило, кото-

рое позволяет отнести наблюдение или группу наблюдений к одному из двух заведомо известных или экспериментально определенных классов с некоторой точно-

стью. В зависимости от типа классификатора возможно получение как вероятностной количественной оценки принадлежности к классу (примером может служить логистическая регрессия), так и оценки в абсолютной шкале, согласно используемым размерностям. Преимущество классификации посредством количественной оценки выражается в возможности варьирования порога отсечения, что позволяет выбрать наиболее оптимальное решение поставленной задачи. Ярким примером применения бинарных классификаторов являются медицинские диагностические тесты и диагностические модели (на основе ряда тестов). Целью диагностического теста является выявление наличия или отсутствия заболевания в зависимости от результатов анализов, необходимых для постановки диагноза. Таким образом, решаются следующие важные для клинициста задачи: сведение результатов анализов, зачастую сложно поддающихся визуальной интерпретации, к простому и понятному разделению на два класса (например, болен/здоров), а также выявление коррелирующих с диагнозом факторов и придание им весовых оценок. Данный подход позволяет определить вероятностную оценку наличия или отсутствия заболевания, что дает возможность уменьшить степень неопределенности в принятии решений [1].

На практике зачастую предпочтительнее применение более простых, пусть и менее точных тестов в сравнении с имеющимся «золотым стандартом». Это может быть вызвано сложными процедурами, сопряженными с применением последнего, его дороговизной или возможными неже-

лательными последствиями для пациента. Подобный подход ведет к необходимости количественной оценки схожести альтернативного теста с эталонным. А в случае разработки новых методов диагностики (например, в случае отсутствия эталонного метода) требуется некий объективный показатель их качества.

Одним из способов оценки бинарного классификатора является ROC-анализ. Метод достаточно популярный и часто применяется в исследовательской практике. Также ведется работа в направлении усовершенствования методов классического ROC-анализа, что позволяет преодолеть некоторые его ограничения [2]. В данной статье ставится цель изложить методику проведения ROC-анализа, некоторые его ограничения, а также указать на возможные ошибки интерпретации результатов.

Постановка задачи

Оценка классификатора осуществляется проверкой его способности классифицировать участников контрольной выборки (группы), результаты классификации которых заведомо известны (например, определены с помощью эталонного теста). Возьмем за положительный исход наличие определяемого моделью (тестом) заболевания, отрицательный исход – его отсутствие.

В ходе проверки алгоритм для каждого наблюдения дает свою количественную оценку классификации (например, вероятность). Эта оценка сопоставляется с меткой класса в контрольной группе. Результат работы классификатора можно представить в виде таблицы.

Таблица 1 — Фрагмент таблицы результатов тестирования модели

№	Оценка алгоритма (вероятность)	Интерпретация результата (порог 0,5)	Истинное значение класса в контрольной группе
1	0,75	Болен	Болен
2	0,41	Здоров	Болен
3	0,45	Здоров	Здоров
4	0,83	Болен	Болен
5	0,91	Болен	Болен
6	0,82	Болен	Болен
7	0,33	Здоров	Здоров
8	0,25	Здоров	Здоров
9	0,61	Болен	Здоров

При достаточно большой выборке результаты работы модели можно изобразить в виде распределения значений оценок алгоритма в соответствующих классах обучающей выборки. В этом случае возможны

варианты практически полного, частичного или отсутствия пересечения множеств (рисунок 1 — слева).

Если распределения не пересекаются (рисунок 1а), то это говорит о том, что зна-

чения оценок четко разделяются на две группы, что позволяет безошибочно классифицировать наблюдения (идеальный классификатор). При полном или практически полном пересечении множеств (рисунок 1г) модель не разделяет пациентов и оказывается бесполезной при любом пороге отсечения, что соответствует «случайному гаданию», как если бы врач ставил диагноз, подбрасывая монету [1]. Однако чаще встречаются ситуации, когда некоторые значения оценок, выдаваемые алгоритмом для наблюдений одного класса, оказываются в области значений оценок другого класса, в итоге алгоритм совершает ошибочную классификацию (наблюдения 2 и 9 в таблице 1). Распределения значений оценок алгоритма будут частично пересекаться, «накладываться» друг на друга (области FN или FP на рисунке 1б, в). То есть какие-то значения оценок модели (непересекающиеся области на рисунке 1б, в) позволяют осуществлять правильную классификацию, но в области пересекающихся значений классификация может быть ошибочной. Варианты частичного пересечения множеств соответствуют большинству реальных диагностических тестов. При этом, если мы будем варьировать величину порога отсечения (в таблице 1 он принят рав-

ным 0,5), классификатор будет определять одни и те же наблюдения, находящиеся в зоне пересечения множеств, либо к одной, либо ко второй группе. Например, изменив порог в таблице 1 с 0,5 на 0,4, получим смену интерпретации результатов для наблюдений 2 и 3 на противоположные значения (диагноз «болен» вместо «здоров»). Таким образом, варьируя порог, возможно осуществление более точной классификации. Но подобная процедура может привести к увеличению ошибок алгоритма: возрастание количества больных в группе здоровых (ложноотрицательные результаты FN) и наоборот (ложноположительные результаты FP). Например, можно установить такое значение порога, при котором классификатор будет правильно классифицировать все наблюдения с меткой «положительный диагноз» (сделать порог очень низким). В этом случае количество ложноотрицательных результатов будет равняться нулю и алгоритм не пропустит больного, однако количество ложноположительных ошибок возрастет: в группу больных попадет большое количество здоровых (заштрихованная область FP на рисунке 2).

Довольно часто результаты работы модели представляют в виде таблицы 2×2 (таблица 2).

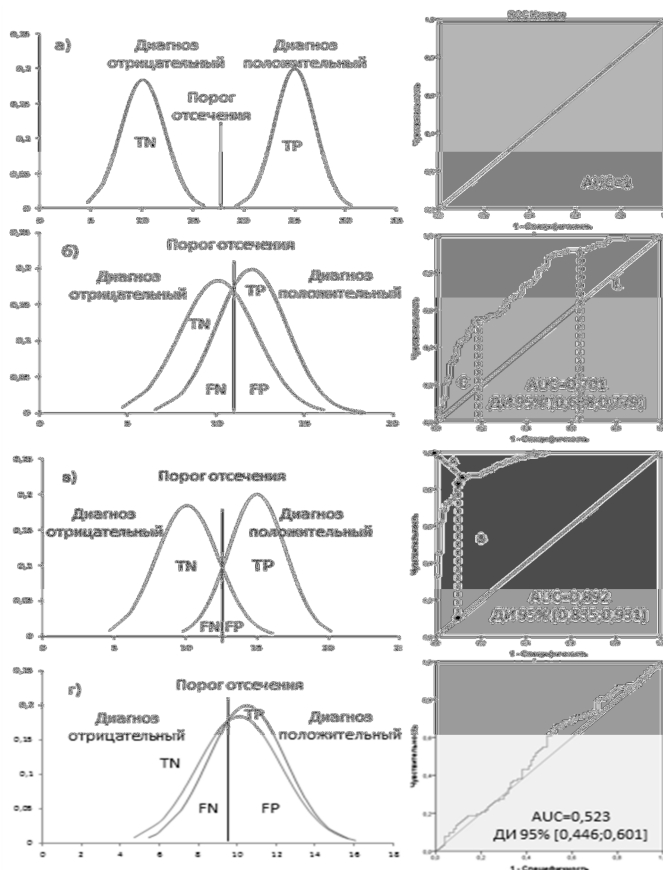


Рисунок 1 — Распределение результатов диагностического теста. Соответствующие ROC-кривые

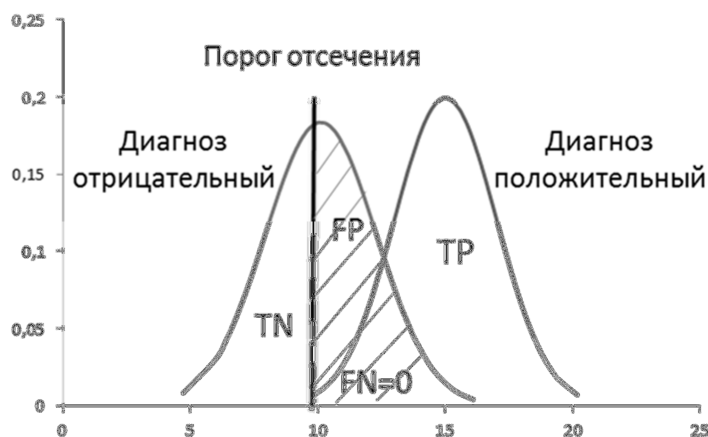


Рисунок 2 — Распределение результатов диагностического теста с изменением порога отсечения в сторону уменьшения количества ложноотрицательных результатов

Таблица 2 — Общий вид сводной таблицы результатов тестирования

Модель (предсказанные моделью результаты)	Фактические (наблюдаемые) результаты (результаты эксперимента)	
	положительно	отрицательно
Положительно	TP	FP
Отрицательно	FN	TN

Примечания: **TP** (True Positives, истинно положительные случаи) — верно классифицированные положительные примеры, то есть определение моделью наличия заболевания, когда оно на самом деле есть; **TN** (True Negatives, истинно отрицательные случаи) — верно классифицированные отрицательные примеры, алгоритм определил отсутствие заболевания в случае его истинного отсутствия; **FN** (False Negatives, ложно отрицательные примеры) — положительные примеры, классифицированные как отрицательные (ошибка I рода), то есть пациент болен, но классифицирован как здоровый; **FP** (False Positives, ложно положительные случаи) — отрицательные примеры, классифицированные как положительные (ошибка II рода), или ложное обнаружение, то есть при отсутствии заболевания ошибочно выносится решение о его наличии.

Таблица 2 по смыслу соответствует графикам на рисунке 1, но является более наглядной, представляя результаты тестирования модели в числовой форме. Очевидно, что при изменении порога отсечения будут изменяться и данные в ячейках таблицы.

Несмотря на наглядность таблицы сопряженности, она не дает полной оценки работы алгоритма, сложна в применении при сравнении двух моделей, а также не позволяет выбрать оптимальный порог отсечения. Для решения подобных задач предпочтительно воспользоваться методами ROC-анализа.

Алгоритм построения ROC-кривой

Ядром ROC-анализа является построение ROC-кривой (или кривой ошибок). Она характеризует зависимость количества (доли) верно классифицированных положительных результатов (эту величину часто называют чувствительностью Se) от количества (доли) неверно классифицированных отрицательных результатов (количества ложноположительных результатов, что часто записывается, как $1-Sp$, где Sp — специфичность теста, то есть доля верно

классифицированных тестом отрицательных случаев) при варьировании порога отсечения (cut-off).

Для построения кривой необходимо объединить обе группы обучающей выборки (больных (n) и здоровых (m)) в общий массив и упорядочить его по убыванию значений результатов оценки алгоритма. Начало кривой лежит в точке $(0; 0)$. Каждое последующее значение ответа алгоритма из упорядоченного ряда принимается в качестве порога отсечения (или задается фиксированный шаг, например $0,01$). В зависимости от текущего значения порога пациент определяется в одну из двух групп: ниже порогового значения — к группе здоровых; включительно и выше порога — к группе больных. Для каждого значения выбранного порога отсечения подсчитывается доля истинно положительных (Se) и ложноположительных ($1-Sp$) результатов теста по отношению к заведомо определенным классам. Полученные результаты наносятся на график (рисунок 3 и ROC-кривые на рисунке 1), график заканчивается в точке $(1; 1)$ и дополняется прямой $y = x$.

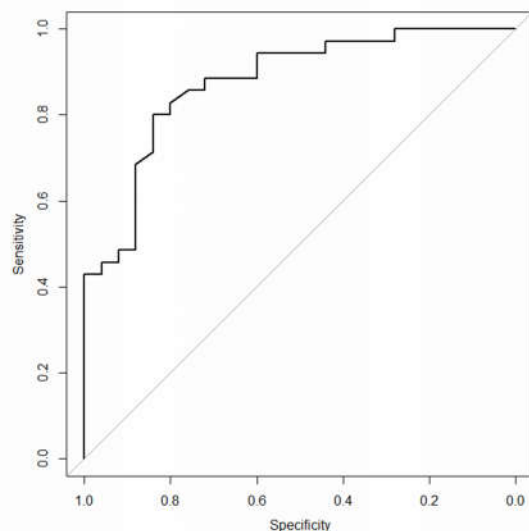


Рисунок 3 — ROC-кривая в среде R

ROC-кривая как для выборок, так и для распределений всегда имеет ступенчатый вид, однако в случае с большими выборками и при очень маленьком значении шага порога отсечения она может визуально сглаживаться.

В случае отсутствия порога отсечения (модель дает бинарный ответ) кривая строится по трем точкам: (0; 0), (Se; 1-Sp), (1; 1) (рисунок 4).

Интерпретация ROC-кривой

ROC-кривая дает наглядное представление качества работы алгоритма. При идеальном классификаторе, когда алгоритм не совершает ошибок, кривая проходит через верхний левый угол (точка (0;1)) (рисунок 1а). Если же кривая приближается к диагональной линии (или лежит на ней), это свидетельствует о низком качестве модели или о бесполезном классификаторе, в это случае модель случайным образом дискриминирует пациентов по группам (рисунок 1г). Наиболее часто кривая лежит между этими экстремальными точками ROC-пространства (рисунок 1б, в).

Модели, располагающиеся вблизи левого нижнего угла, дают малый процент ложноположительных результатов — «консервативные» (область С на рисунке 1б), они имеют низкое значение специфичности. Такие модели определяют доподлинно больных, но могут классифицировать больного как здорового. Тесты в районе верхнего правого угла — «либеральные» [2] (область L на рисунке 1б), их применение снижает риск пропуска больного, при этом определяются доподлинно здоровые. Если кривая проходит ниже диагональной прямой, то модель чаще относит больных к группе здоровых и наоборот, что считается худшим результатом, нежели чем «случайное угадывание». Но такой тест можно адаптировать путем замены стратегии принятия решений на противоположную [2, 3].

Площадь под ROC-кривой (AUC)

Объективной оценкой качества классификатора принято считать площадь под ROC-кривой (AUC). Ее можно рассчитать по формуле трапеций [3]:

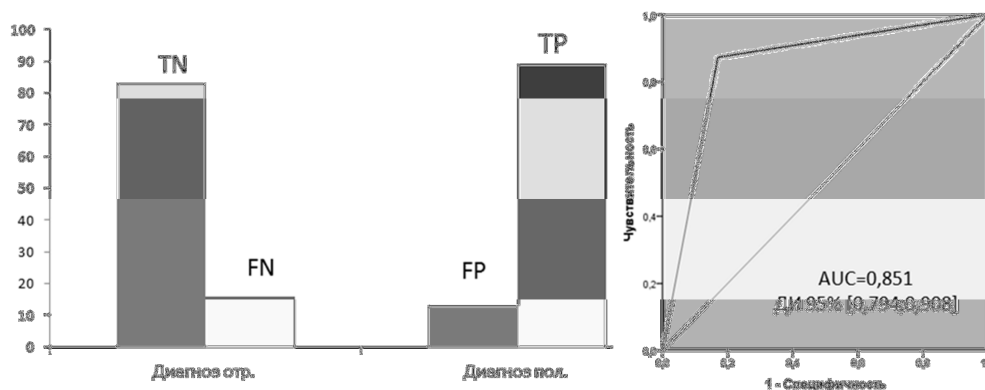


Рисунок 4 — ROC-кривая в случае бинарного ответа алгоритма

$$AUS = \frac{1}{2} \times \sum_{j=1}^{n+m-1} (Se_i + Se_{i+1}) \times (Sp_i - Sp_{i-1}) \tag{1}$$

AUC может принимать значения в интервале от 0,5 до 1. При значении AUC, близком или равном 0,5, говорят об отсутствии дискриминирующей способности у теста, при значении, равном 1, тест считается иде-

альным. Следовательно, чем ближе величина AUC к 1, тем выше прогностическая способность модели. Довольно часто приводится экспертная шкала оценок AUC (таблица 3), по которой можно судить о качестве модели [4].

Таблица 3 — Экспертная шкала оценок AUC

Интервал AUC	Качество модели
0,9–1,0	Отличное
0,8–0,9	Очень хорошее
0,7–0,8	Хорошее
0,6–0,7	Среднее
0,5–0,6	Неудовлетворительное

В клинической диагностике, и в особенности в случаях применения ROC-анализа для оценки тяжести состояния или прогнозирования летального исхода, значение AUC должно принимать значения не ниже 0,9, а тесты с AUC от 0,8 до 0,9 принимаются во внимание в качестве допол-

нительной информации. Тесты же с AUC менее 0,8 не используются для прогнозирования летального исхода [5].

Для вычисления стандартной ошибки AUC и ее двухстороннего доверительного интервала используются следующие формулы [3]:

$$SE(AUS) = \sqrt{\frac{AUS(1-AUS) + (n-1)(Q_1 - AUS^2) + (m-1)(Q_2 - AUS^2)}{n \times m}}, \tag{2}$$

где Q_1 и Q_2 — $AUS/(2 - AUS)$ и $2AUS_2/(1 + AUS)$ соответственно.

$$DI_{auc} = (AUS - \Psi((1 + \beta)/2) \times SE(AUS); AUS + \Psi((1 + \beta)/2) \times SE(AUS)), \tag{3}$$

где $\Psi(.)$ — обратная функция стандартного нормального распределения; β — доверительный уровень в долях (0,99; 0,95 и т.д.).

Помимо доверительного интервала для AUC часто строится доверительный интервал для самой кривой (рисунок 5).

Показатель AUC является критерием качества модели (теста) достаточно условно. Более подходящим будет его применение при срав-

нении ROC-кривых для разных моделей, так как визуальное их сравнение достаточно проблематично. Для сравнения площадей под ROC-кривыми, а также для оценки значимости AUC в сравнении с AUC для «бесполезного» классификатора используется z-статистика [3]:

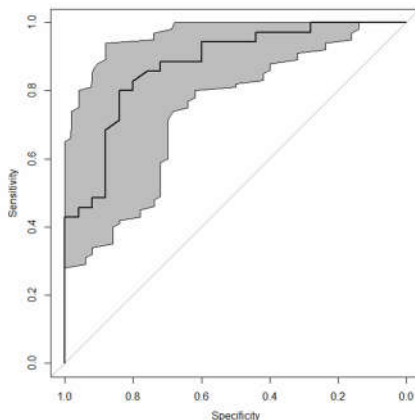


Рисунок 5 — ROC-кривая и ее 95 % доверительный интервал

$$Z = \frac{|AUC_1 - AUC_2|}{\sqrt{SE(AUC_1)^2 + SE(AUC_2)^2}}, \quad (3)$$

где AUC_1 и AUC_2 — площади под сравниваемыми кривыми (для «бесполезного» классификатора $AUC = 0,5$).

Довольно часто ROC-кривые разных классификаторов могут пересекаться, при этом иметь одинаковое значение AUC. Сравнивая две модели посредством площади под ROC-кривыми, необходимо иметь в виду случаи, когда в определенной (возможно, значимой для практического применения) области (или диапазоне значений) одна кривая

может быть эффективнее другой при равных значениях AUC [6]. В этом случае наилучшим способом оценки будет парциальный индекс площади (Partial Area Index), вычисляющийся под определенным участком для каждой кривой (рисунок 6), в зависимости от выбранной допустимого количества ложноположительных результатов [5, 7]. Например, на рисунке 6 в выбранном диапазоне значений 1-Sp более предпочтительной является первая модель (Partial AUC = 0,221).

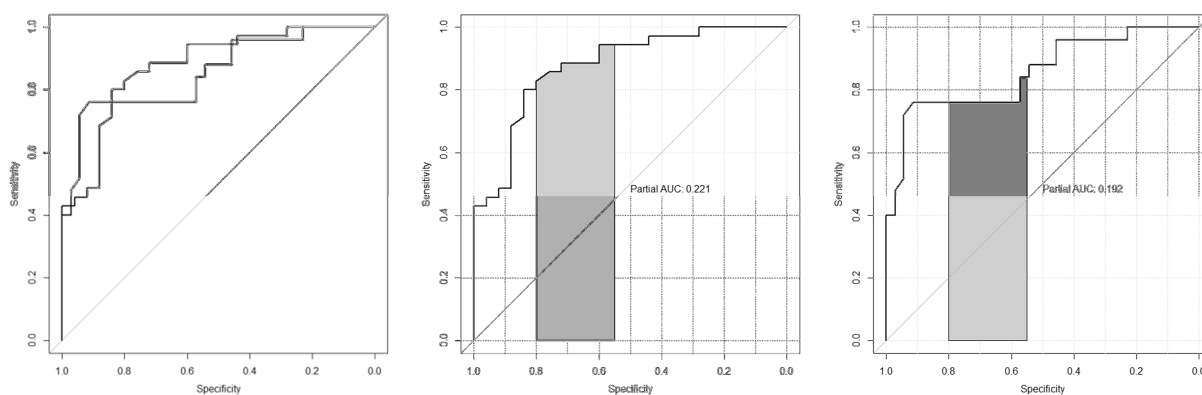


Рисунок 6 — Пересекающиеся ROC-кривые с выделенными областями и рассчитанными парциальными индексами площади этих областей (0,221 и 0,192 соответственно)

При оценке AUC важным моментом является размер выборки. Применяя ROC-анализ для малых выборок, важно помнить, что в этом случае неизменно увеличивается стандартная ошибка AUC (SE) и, соответственно, ширина доверительного интервала как самой кривой, так и площади под ней. Тем самым доверительный интервал кривой и AUC может оказаться достаточно большим, как следствие, возрастает величина среднеквадратичной ошибки модели. А это означает, что модельная оценка состояния новых наблюдений может значительно отличаться от их истинного состояния. Алгоритм, неплохо справлявшийся на примере небольшой тестовой выборки, возможно, будет делать серьезные ошибки при работе с новыми наблюдениями. Применение такой модели на практике нежелательно. Таким образом, малые выборки приводят к большим неточностям в оценке параметров ROC-кривой [8]. На рисунке 7 изображена зависимость экспериментальных значений $AUC \pm SE$ для разного количества выборок. Как

видно, для выборок общим количеством наблюдений в обоих классах менее 100 стандартная ошибка AUC (и соответственно, доверительный интервал) принимает большие значения. Следовательно, для объективной оценки AUC и возможности применения этого параметра в процессе анализа работы алгоритма предпочтительны достаточно большие выборки.

Выбор порога отсечения

Порог отсечения (cut-off) — некоторое выбранное значение в ряду всех возможных результатов отклика модели, которое является границей разделения двух классов. При этом ставится цель получения наименьшего общего количества ошибок классификации или сведения к минимуму ошибок при классификации к той или иной группе (недопущение пропуска больных или наоборот). Предположим, алгоритм выдает вероятностную оценку наличия заболевания в интервале от 0 до 1 (0–100 %), чем эта величина ближе к 1, тем вероятнее наличие заболевания у конкретного пациента. Задача заключается в том,

чтобы выбрать такое оптимальное значение этой вероятности, при котором алгоритм не будет классифицировать больных пациентов как здоровых и наоборот. При идеальном классификаторе эта задача достаточно тривиальна, но, как было показано выше, большинство моделей имеют некоторый процент ошибок, и задача сводится к их минимизации в зависимости от желаемых значений чувствительности и

специфичности. Это может быть требование к минимальной (или какой-то иной определенной в зависимости от поставленных задач) величине чувствительности (специфичности). Или выбирается баланс между этими величинами ($Se \approx Sp$): минимальное расстояние от ROC-кривой до точки (0;1), отрезок А в ROC-пространстве на рисунке 1в, которое рассчитывается для каждой точки разделения по формуле 4 [9]:

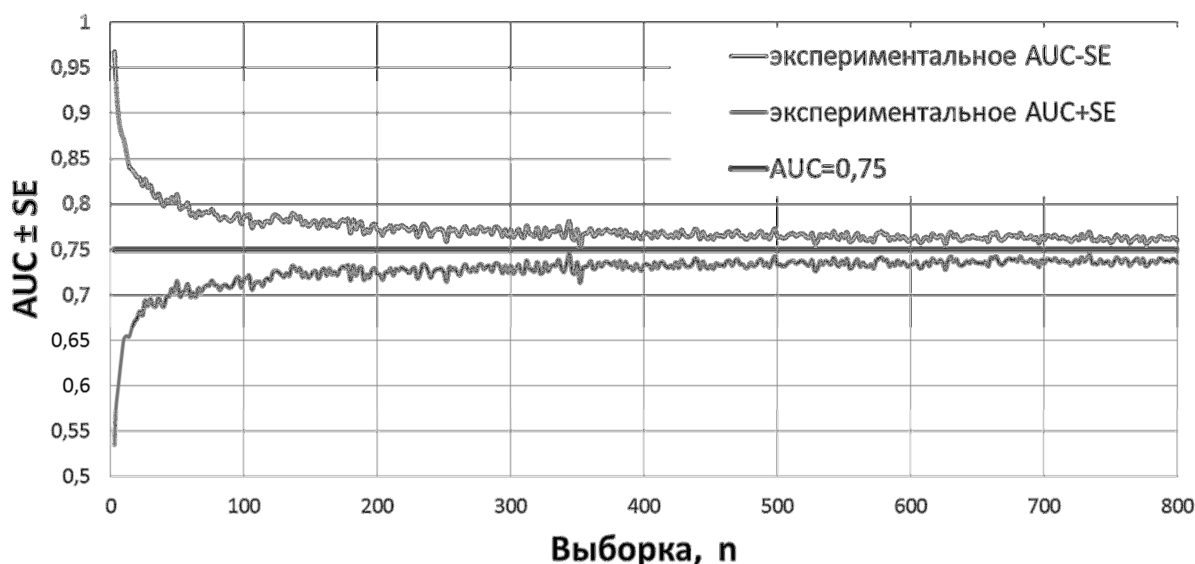


Рисунок 7 — График зависимости экспериментальных значений AUC ± SE для разного количества выборок

$$\sqrt{(1 - Se)^2 + (1 + Sp)^2} \quad (4)$$

Также это может быть точка, где доля истинно положительных результатов (Se) принимает наибольшее значение при наименьшем количестве ложноположительных результатов (1-Sp): максимальное значение суммы чувствительности и специфичности. В этом случае принято использовать индекс Йодена (Youden's index), характеризующий максимальное отличие ординат ROC-кривой и прямой $y = x$ [2], отрезок В в ROC-пространстве на рисунке 1в.

В иных случаях предпочтительнее выбрать порог таким образом, чтобы получить более «либеральный» классификатор, не позволяющий пропустить больного, но при этом включающий в группу носителей признака большой процент наблюдений с его отсутствием. Тем самым увеличивается шанс не пропустить опасную, но излечимую болезнь [1]. В случае же если ложноположительный результат теста ведет к необходимости последующих сложных, опасных или финансово затратных для пациента процедур, то предпочтительнее выбрать тест с вы-

соким значением специфичности, который точнее определит доподлинно больных.

Заключение

В заключение следует отметить, что программная реализация модуля ROC-анализа имеется во многих специализированных пакетах, таких как «Statistica» (модуль нейронные сети), SPSS, MedCalc, Attestat и других. Также в сети Интернет имеются ресурсы для обработки данных посредством ROC-анализа, например, по адресу: <http://www.rad.jhmi.edu/jeng/javarad/roc/JROCFITi.html>. Подробно изучив методику проведения анализа, не составляет труда осуществить его посредством применения возможностей электронных таблиц MS Excel. Однако для сравнения кривых, определения AUC, статистической значимости AUC и других важных для анализа параметров имеет смысл воспользоваться возможностями среды для обработки данных R, например, применив пакеты pROC или ROCR.

ЛИТЕРАТУРА

1. Флетчер Р, Флетчер С, Вагнер Э. Клиническая эпидемиология. Основы доказательной медицины. Москва, РФ: Медиа Сфера; 1988. 352 с.
2. Файнзил'берг ЛС, Жук ТН. Гарантированная оценка эффективности диагностических тестов на основе усиленного ROC-анализа. *Управляющие Системы и Машины*. 2009;5:3-13.
3. Гайдьшев ИП. Оценка качества бинарных классификаторов. *Вестник Омского Университета*. 2016; 1(79): 14-17. <https://cyberleninka.ru/article/n/otsenka-kachestva-binarnyh-klassifikatorov/viewer> (in Russ.)
4. Литвин АА, Калинин АЛ, Тризна НМ. Использование данных доказательной медицины в клинической практике. *Проблемы Здоровья и Экологии*. 2008 Vol 18(4):12-19.
5. Пацай ДИ, Блахов НЮ. ROC-анализ оценки тяжести состояния и качества прогноза у больных острым панкреатитом. *Военная Медицина*. 2011;42(3):60-66.
6. David J. Hand. Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine Learning*. 2009;77:103-23. doi: 10.1007/s10994-009-5119-5
7. Seong Ho Park, Jin Mo Goo, Chan-Hee J. Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists *Korean Journal of Radiology*. 2004; 5(1):11-18. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2698108/>
8. Hanczar B, Hua J, Chao Sima, Weinstein J, Bittner M, Edward R. Dougherty Small-sample precision of ROC-related estimates. *Bioinformatics*. 2010;26(6):822-30. <https://doi.org/10.1093/bioinformatics/btq037>
9. Королюк ИП. ROC-анализ (операционные характеристики наблюдателя): базовые принципы и применение в лучевой диагностике. *Медицинская Визуализация*. 2013;6:113-23.

REFERENCES

1. Fletcher R, Fletcher S, Vagner E. *Klinicheskaya epidemiologiya. Osnovy dokazatel'noj mediciny*. Moskva, RF: Media Sfera; 1988. 352 p. (in Russ.)
2. Fajnzil'berg LS, Zhuk TN. Garantirovannaya ocenka effektivnosti diagnosticheskikh testov na osnove usilennogo ROC-analiza. *Upravlyayushchie Sistemy i Mashiny*. 2009;5:3-13. (in Russ.)
3. Gajdyshev IP. Ocenka kachestva binarnyh klassifikatorov. *Vestnik Omskogo Universiteta*. 2016; 1(79): 14-17. <https://cyberleninka.ru/article/n/otsenka-kachestva-binarnyh-klassifikatorov/viewer> (in Russ.)
4. Litvin AA, Kalinin AL, Trizna NM. Ispol'zovanie dannyh dokazatel'noj mediciny v klinicheskoy praktike. *Problemy Zdorov'ya i Ekologii*. 2008 Vol 18(4):12-19. (in Russ.)
5. Pacaj DI, Blahov NYU. ROC-analiz ocenki tyazhesti sostoyaniya i kachestva prognoza u bol'nyh ostrym pankreatitom. *Voennaya Medicina*. 2011;42(3):60-66. (in Russ.)
6. David J. Hand. Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine Learning*. 2009;77:103-23. doi: 10.1007/s10994-009-5119-5
7. Seong Ho Park, Jin Mo Goo, Chan-Hee J. Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists *Korean Journal of Radiology*. 2004;5(1):11-18. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2698108/>
8. Hanczar B, Hua J, Chao Sima, Weinstein J, Bittner M, Edward R. Dougherty Small-sample precision of ROC-related estimates. *Bioinformatics*. 2010; 26(6):822-30. <https://doi.org/10.1093/bioinformatics/btq037>
9. Korolyuk IP. ROC-analiz (operacionnye harakteristiki nablyudatelya): bazovye principy i primeneniye v luchevoj diagnostike. *Medicinskaya Vizualizaciya*. 2013;6:113-23. (in Russ.)

Поступила 27.05.2020

Received 27.05.2020

Принята в печать 24.09.2020

Accepted 24.09.2020

Сведения об авторе:

Ковалев Алексей Алексеевич — заведующий сектором ЭТОРОИМ УО «Гомельский государственный медицинский университет»; e-mail: etorom@gsmu.by; <https://orcid.org/0000-0001-9148-487X>

Кузнецов Борис Киприянович — к.б.н., доцент, зав. кафедрой медицинской и биологической физики УО «Гомельский государственный медицинский университет»; <https://orcid.org/0000-0002-3121-5595>

Ядченко Алексей Александрович — доктор физико-математических наук, профессор кафедры медицинской и биологической физики УО «Гомельский государственный медицинский университет»; <https://orcid.org/0000-0002-3366-1601>

Игнатенко Валерий Андреевич — к.б.н., заведующий отраслевой лабораторией по мониторингу пищевого (микронутриентного) статуса населения и разработки технологий его коррекции с использованием функциональных продуктов и биологически активных добавок ГП «Институт биохимии биологически активных соединений НАН Беларуси»; <https://orcid.org/0000-0001-5370-865X>

Автор, ответственный за переписку:

Ковалев Алексей Алексеевич — e-mail: etorom@gsmu.by

Information about authors:

Aleksey A. Kovalev — Head of the Sector of Operational Maintenance and Repair of Equipment of the EI «Gomel State Medical University»; e-mail: etorom@gsmu.by; <https://orcid.org/0000-0001-9148-487X>

Boris K. Kuznetsov — Candidate of Biological Science, Associate Professor, Head of the Department of Medical and Biological Physics of the EI «Gomel State Medical University»; <https://orcid.org/0000-0002-3121-5595>

Aleksey A. Yadchenko — Doctor of Physics and Mathematics, Professor at the Department of Medical and Biological Physics of the EI «Gomel State Medical University»; <https://orcid.org/0000-0002-3366-1601>

Valeriy A. Ignatenko — Candidate of Biological Science, Head of the Branch Laboratory for Monitoring the Nutritional (Micronutrient) Status of the Population and Development of Technologies for Its Correction Using Functional Products and Biologically Active Supplements, SE «Institute of Biochemistry of Biologically Active Compounds of the National Academy of Sciences of Belarus»; <https://orcid.org/0000-0001-5370-865X>

Corresponding author:

Aleksey A. Kovalev — e-mail: etorom@gsmu.by