

# APPLYING MACHINE LEARNING FOR PREDICTION OF CARDIOVASCULAR DISEASES ON SMALL DATA SETS

E. A. Kaledina, O. E. Kaledin, T. I. Kulyagina\*

Ogarev Mordovia State University,  
430005, Saransk, Russia

\*IE Wilhelm,  
430005, Saransk, Russia

---

DOI: 10.24412/2073-0667-2022-1-66-76

As a result of increasing computing power and generating large amounts of data, artificial intelligence algorithms are currently being actively used to perform a wide range of medical tasks. One of the most important areas in which the use of artificial intelligence can be useful is the diagnosis of various diseases and the prediction of their possible outcomes.

Cardiovascular diseases are one of the main factors of mortality and disability in most countries of the world, including the Russian Federation. The most important risk factor for two major cardiovascular diseases (myocardial infarction and cerebral stroke) is arterial hypertension. Therefore, the main task of primary prevention of cardiovascular diseases (CVD) is the timely detection of a high risk of fatal CVD in patients with diagnosed uncomplicated arterial hypertension. The use of machine learning algorithms can solve this problem and significantly improve the accuracy of predicting cardiovascular diseases and their complications. Machine learning methods are the main tool of artificial intelligence, the use of which allows you to automate the processing and analysis of big data, identify hidden or non-obvious patterns on this basis, and extract new knowledge.

This article describes the process of using machine learning algorithms to predict the risk of developing adverse cardiovascular events in patients with diagnosed arterial hypertension in the next 12, 24 and 36 months. The analysis included 16 predictors, which are a combination of both standard indicators of the risk of cardiovascular diseases (age, male sex, smoking, elevated cholesterol, impaired uric acid metabolism), and some specific indicators. A distinctive feature of this task is the use of local data collected in a separate region of the Russian Federation as a training data set. This feature can improve the adaptability of the predictive model to possible local features of the development of cardiovascular diseases, however, it also has a significant drawback — a small amount of training data, which contributes to model retraining and, as a result, a decrease in its ability to generalize.

The target feature in the study is a binary predictive vector of major adverse cardiovascular events at three reference time points. Due to the fact that censoring, as well as some of the considered cardiovascular diseases, can occur simultaneously or be repeated throughout all or part of the observation period, the study is formally presented as a solution to the multilabel classification problem. The paper indicates the stages of forming a data set and explores predictive machine learning algorithms on small sets to create a model for calculating the risks of cardiovascular diseases. The advantages and disadvantages of individual ensemble methods of machine learning machine learning methods (binary relevance, multioutput classifier, label powerset, MLkNN, classifier chain) for the development of predictive algorithms in the conditions of the problem are shown.

From the results of the study, we can say that the machine learning algorithms — multioutput classifier and labelpowerset on a small dataset showed the best result among all the analyzed methods

for assessing the development of cardiovascular diseases. This fact makes it relevant to study the application of this method on samples of large volumes, with the inclusion of a larger set of risk factors.

**Key words:** machine learning algorithms, data analysis, cardiovascular disease prediction.

## References

1. Balanova YU.A., Shal'nova S. A., Imayeva A. E., Kapustina A. V., Muromtseva G. A., Yevstifeyeva S. Ye., Tarasov V. I., Red'ko A. N., Viktorova I. A., Prishchepa N. N., Yakushin S. S., Boytsov S. A., Drapkina O. M. Rasprostranennost' arterial'noy gipertonii, okhvat lecheniyem i yego effektivnost' v Rossiyskoy // Ratsional'naya Farmakoterapiya v Kardiologii. 2019. N 15 4. S. 450–466.
2. Simerzin V. V., Gagloyeva I. V., Gar'kina S. V. Sovremennaya kontsepsiya profilaktiki serdechno-sosudistykh zabolеваний // Vestnik SamGU. Yestestvennayauchnaya seriya. 2007. N 9/1 59. S. 296–306.
3. Conroy R. M., Pyorala K., Fitzgerald A. P. et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project // European Heart Journal. 2003. N 24. P. 987–1003.
4. Gusev A. V., Gavrilov D. V., Korsakov I. N., Serova L. M., Novitskiy R. E., Kuznetsova T. YU. Perspektivy ispol'zovaniya metodov mashinnogo obucheniya dlya predskazaniya serdechno-sosudistykh zabolеваний // Iskusstvennyy intellekt v zdorovookhranenii. 2019. N 3. S. 41–47.
5. Poplin R., Varadarajan A. V., Blumer K., et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning // Nat Biomed Eng. 2018. V. 2. N 3. R. 158–164.
6. Quesada J. A., Lopez-Pineda A., Gil-Guillen V. F., et al. Machine learning to predict cardiovascular risk // The international journal of clinical practice. 2019. V. 73. N 10. e13389.
7. Zack C. J., Senecal C., Kinar Y. et al. Leveraging Machine Learning Techniques to Forecast Patient Prognosis After Percutaneous Coronary Intervention // JACC Cardiovasc Interv. 2019. V. 12. N 14. R. 1304–1311.
8. Tai F., Lin H.-T. Multilabel Classification with Principal Label Space Transformation. // Neural Comput., 24–9, 2012. P. 2508–2542.
9. Gibaja E., Ventura S. A Tutorial on Multilabel Learning // ACM Computing Surveys. 2015. Vol. 47, N. 3. P. 1–38.
10. Tsoumakas G., Katakis I., Vlahavas I. Data Mining and Knowledge Discovery Handbook. Springer, 2010. R. 667–685.
11. Zhang M.-L., Zhou Z.-H. A Review on MultiLabel Learning Algorithms // IEEE Transactions on Knowledge and Data Engineering. 2014. Vol. 26, Iss. 8. P. 1819–1837.



# ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРЕДСКАЗАНИЯ СЕРДЕЧНО-СОСУДИСТЫХ ЗАБОЛЕВАНИЙ НА МАЛЫХ НАБОРАХ ДАННЫХ

Е. А. Каледина, О. Е. Каледин, Т. И. Кулягина\*

ФГБОУ ВО Мордовский государственный университет им. Н. П. Огарева,  
430005, Саранск, Россия

\*ИП Вильгельм,  
430005, Саранск, Россия

УДК 004.89

DOI: 10.24412/2073-0667-2022-1-66-76

Благодаря увеличению вычислительной мощности и генерации больших объемов данных алгоритмы искусственного интеллекта в настоящее время активно используются для выполнения широкого круга медицинских задач. Одной из важнейших областей, в которых может быть полезно применение искусственного интеллекта, в частности, методов машинного обучения, — это диагностика заболеваний и прогнозирование их возможных исходов. В настоящей статье описывается процесс использования алгоритмов машинного обучения для предсказания риска развития неблагоприятных сердечно-сосудистых событий у пациентов с диагностированной артериальной гипертензией в ближайшие 12, 24 и 36 месяцев. Отличительной чертой данной задачи является использование в качестве обучающего дата-сета локальных данных, собранных в отдельном регионе РФ. Данная особенность может улучшить адаптированность предсказательной модели к возможным локальным особенностям развития сердечно-сосудистых заболеваний, однако и имеет существенный недостаток — малое количество обучающих данных. В работе рассмотрена одна из возможных формальных математических постановок рассматриваемой задачи, указаны этапы формирования дата-сета и исследованы предсказательные алгоритмы машинного обучения на малых наборах для создания модели расчета рисков сердечно-сосудистых заболеваний. Представленные результаты позволяют говорить о возможности использования моделей машинного обучения для повышения точности прогнозирования сердечно-сосудистых рисков.

**Ключевые слова:** алгоритмы машинного обучения, анализ данных, предсказание сердечно-сосудистых заболеваний.

**Введение.** Артериальная гипертензия (АГ) — наиболее распространенное хроническое заболевание в России и важнейший фактор риска основных сердечно-сосудистых заболеваний — инфаркта миокарда и мозгового инсульта. Основной задачей первичной профилактики сердечно-сосудистых заболеваний (ССЗ) является своевременное выявление у пациентов с неосложненной АГ высокого риска развития смертельных ССЗ [1].

Научной основой профилактики ССЗ является концепция факторов риска, которые были выявлены в эпидемиологических исследованиях [2]. В настоящий момент для оценки риска смертельных ССЗ в России и странах Европы применяется шкала SCORE [3], учитывающая такие факторы риска, как возраст, мужской пол, курение, повышенный уровень холестерина, нарушение обмена мочевой кислоты, и др. Однако, данная шкала имеет

слишком общий характер и не подходит для оценки риска развития острого коронарного синдрома, инсульта, ишемической болезни сердца для пациентов с диагностированной АГ. Низкая точность предсказания сердечно-сосудистых событий имеет ряд причин [4]: 1) оценка суммарного риска не адаптирована к региональным особенностям. Структура смертности от ССЗ в РФ варьируется в зависимости от региона и места проживания; 2) в шкалах рассматривается небольшое число основных признаков, влияющих на развитие ССЗ и зачастую не учтены существенные для клинические состояния. Данное обстоятельство не позволяет определить фактический риск наступления неблагоприятных сердечно-сосудистых событий; 3) данные, которые были использованы для составления шкал, были получены 30–50 лет назад и не соответствуют современным реалиям.

Одним из способов решения проблемы недостаточной точности результатов расчета сердечно-сосудистого риска может являться использование алгоритмов машинного обучения (МО). Подобные алгоритмы повышают эффективность прогнозирования рисков за счет использования объемных хранилищ данных при независимой идентификации дополнительных факторов риска и сложных взаимодействий между ними [5–7]. Кроме того, использование моделей МО, обученных на данных отдельных регионов, может получить более индивидуализированную, чем при использовании шкалы SCORE, оценку риска развития смертельных ССЗ, а следовательно, лучшую адаптацию управления рисками к отдельным пациентам.

**1. Подготовка данных.** Целью исследования является решение задачи предсказания риска развития неблагоприятных сердечно-сосудистых событий у пациентов с диагностированной артериальной гипертензией в ближайшие 36 месяцев с использованием алгоритмов машинного обучения на основе локальных данных Республики Мордовия. В качестве неблагоприятных сердечно-сосудистых событий представлены острый коронарный синдром, инсульт, ишемическая болезнь сердца, смерть. Контрольные даты прогнозирования рисков наступления ССЗ — 12, 24 и 36 месяцев с постановки диагноза артериальная гипертензия.

Для достижения поставленной цели были решены следующие задачи:

1. Формирование дата-сета для исследования, представленного данными по пациентам с необходимыми и достаточными признаками для расчетов риска ССЗ.

2. Проведение анализа имеющихся данных, формального описания и математической постановки задачи.

3. Исследование предсказательных алгоритмов машинного обучения и их готовых реализаций, создание модели расчета рисков ССЗ для полученного дата-сета и оценка ее эффективности.

Данные были предоставлены сотрудниками медицинского института НИ МГУ им. Н. П. Огарева (г. Саранск) и представляют собой сводную таблицу о наблюдении пациентов с АГ. Базовая выборка пациентов имеет средний возраст  $57.8 \pm 10.4$  года, с 34 до 84 лет. Исходная дата наблюдения была установлена 1 января 2017 года, что позволило всем пациентам в когорте находиться под наблюдением в течение 3 лет, дата конца периода наблюдения была определена как 1 января 2020. Наблюдения представляют собой совокупность как стандартных показателей риска ССЗ (возраст, курение и т. д.), так и специфических показателей таких как интерлейкин 6, неоптерин, омега-3 и др. Полный список наблюдаемых признаков представлен в таблице 1.

При формировании дата-сета отсутствующие значения были заменены на среднее по признаку. Так, например, отсутствующие значения длительности АГ были заменены на

Таблица 1  
Наблюдаемые признаки

Признак	Ед. изм.	Сред. знач.	СКО	Min	Max
Возраст пациента	лет	57.78	10.37	34	84
Пол пациента		1 — муж., 0 — жен.			
Курение		1 — да, 0 — нет			
Длительность АГ	годы	7.53	3.943	1	20
Систолическое АД	мм.рт.ст.	141.85	110.56	1	170.52
Диастолическое АД	мм.рт.ст.	82.98	13.25	58.49	113.35
Холестерин общий	ммоль/л.	5.54	0.7	3.7	6.9
Интерлейкин 6	пг/мл	14.19	1.48	10.52	17.93
Фактор некроза опухолей	пг/мл.	11.47	1.9	7.15	14.97
Неоптерин	нмоль/л.	6.0	1.79	2.09	9.84
Альфа-линоленовая кислота	нмоль/л.	110.08	20.0	78.67	140.01
Арахидоновая кислота	нмоль/л.	1243.57	121.8	1054.6	1433.34
Концентрация малонового диальдегида	нмоль/г.	3.24	0.77	1.51	5.46
Активность супероксиддисмутазы	усл. ед.	3.31	0.86	1.58	6.08
Индекс массы миокарда левого желудочка	г/м <sup>2</sup>	112.02	11.06	87.02	146.72
Характер ночного снижения АД		0 — нормальное, 1 — недостаточное, 2 — ночная гипертония			
Результаты контрольного наблюдения на сроках 12, 24, 36 месяцев		0 — цензурирован; 1 — острый коронарный синдром; 2 — инсульт; 3 — ишемическая болезнь сердца; 4 — смерть			

среднее значение в возрастной категории 50–60 лет. Категориальные переменные (пол, курение, результаты контрольного наблюдения) были переведены в бинарные. Далее, проведен корреляционный анализ рассматриваемых показателей, результат которого представлен на рис. 1.

Так, зависимым признаком можно считать показатель альфа-линоленовой кислоты (омега-3, ALA), который коррелирует с показателями диастолического артериального давления (DAD), неоптерина, арахидоновой кислоты (AA), индекса массы миокарда левого желудочка (IMMLG). Данный признак был удален из рассмотрения.

Следующий шаг в создании data-сета — нормализация данных. В машинном обучении нормализацией называют метод предобработки числовых признаков в обучающих наборах данных с целью приведения их к некоторой общей шкале без потери информации о различии диапазонов. Необходимость нормализации вызвана тем, что разные признаки обучающего набора данных представлены в разных масштабах и изменяются в разных диапазонах. В этом случае возникает нарушение баланса между влиянием входных переменных, представленных в разных масштабах, на выходную переменную. Т.е. это влияние обусловлено не реальной зависимостью, а изменением масштаба. В результате обучающая модель может выявить некорректные зависимости. В работе использована следующая формула нормализации:

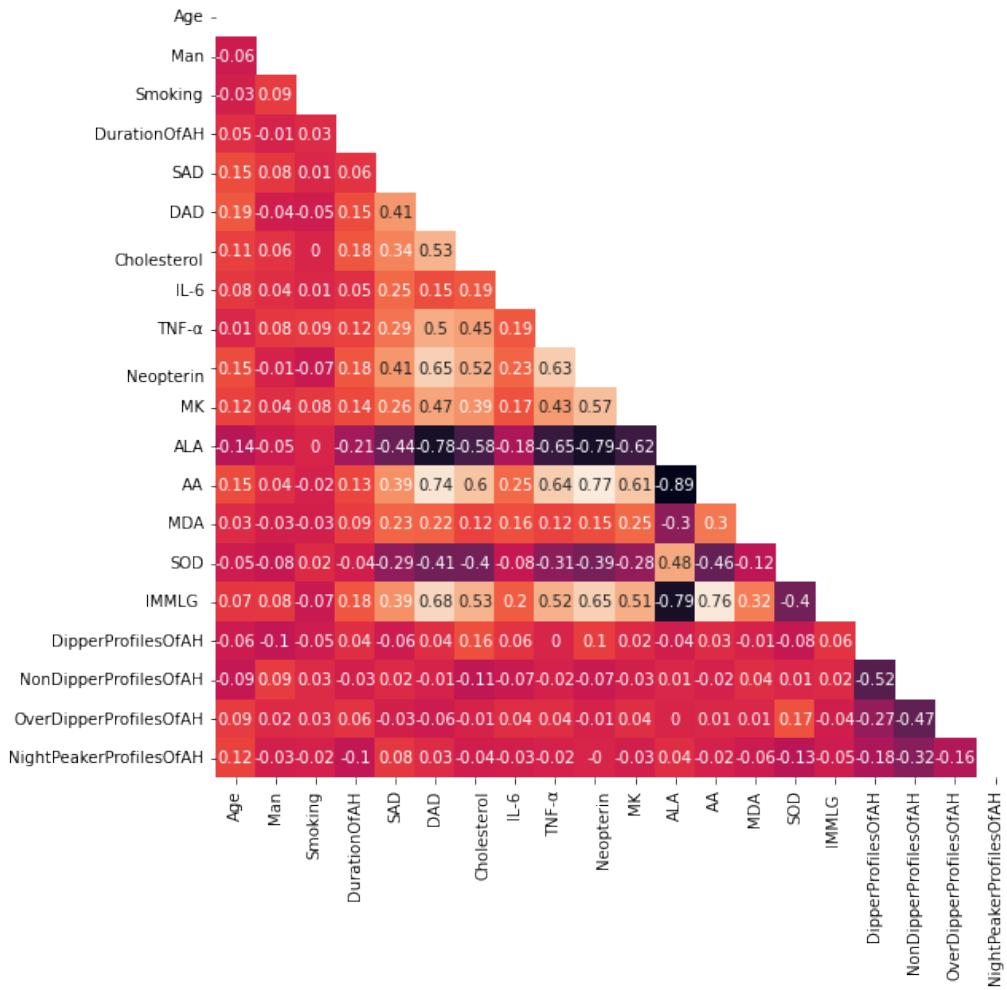


Рис. 1. Корреляционный анализ признаков

$$x_{norm} = \frac{(x - x_{min})}{(x_{max} - x_{min})},$$

где  $x_{min}$  и  $x_{max}$  — минимальное и максимальное значения признака соответственно. Данное масштабирование обеспечивает устойчивость к небольшим стандартным отклонениям функций и сохраняет нулевые записи в разреженных данных.

**2. Обучение моделей.** Главной особенностью дата-сета является его малый размер вследствие локального сбора данных по региону. Данное обстоятельство может существенно снизить способность модели к обобщению. Проблемы с малыми наборами данных связаны с высокой дисперсией и включают в себя переобучение, когда модель не только запоминает тренировочные данные, но подгоняется под тестовый набор. Кроме того, выбросы, т. е. объекты, сильно отличающиеся по некоторым признакам от признаков основной массы, становятся намного опаснее, так как сильнее влияют на итоговую модель.

Одним из подходов к тому, чтобы сделать пространство ввода более гладким и легким для изучения, является добавление шума во вводные данные. В рассматриваемой задаче точность решения существенно (около 10 % точности) повысил такой способ аугментации данных как добавление гауссовского шума с математическим ожиданием  $\mu=0$  и средне-

квадратичным отклонением  $\sigma=0,05$  к числовым признакам. Отметим, что к бинарным признакам запутывание не применялось.

Целевым признаком в исследовании является бинарный вектор предсказания основных неблагоприятных сердечно-сосудистых событий в трех контрольных временных точках. Наличие последовательных моментов времени не позволяет интерпретировать проблему как задачу многоклассовой классификации. При этом, контрольных промежутков всего три, поэтому рассматривать задачу как прогнозирование временного ряда также нецелесообразно. В связи с тем, что цензурение, а также некоторые рассматриваемые ССЗ могут происходить одновременно или повторяться на протяжении всего или части периода наблюдения, было решено представить исследование как решение задачи классификации с пересекающимися классами (мультиметочная классификация).

В данных задачах  $X = \mathbb{R}^D$  — множество объектов, где  $D$  — число признаков (атрибутов) каждого объекта,  $Y = \{0, 1\}^L$  — множество ответов. При этом каждому объекту  $x \in X$  соответствует вектор  $y \in Y$ , показывающий, к каким классам объект относится. Обучающая выборка представляет собой совокупность пар  $D = \{(x^{(i)}, y^{(i)})\}_{i=1}^N$ , где  $x^{(i)} = [x_1, x_2, \dots, x_D]$  представляет собой один экземпляр данных,  $y^{(i)} = [y_1, y_2, \dots, y_L]$  — вектор меток, причем

$$y_j = \begin{cases} 1, & \text{если данный экземпляр относится к классу } j, \\ 0, & \text{в противном случае.} \end{cases}$$

В рассматриваемой задаче число классов  $L = 15$ , количество атрибутов объектов  $D = 20$ , размер обучающего множества данных после аугментации  $N = 900$ .

Большинство алгоритмов машинного обучения созданы для задач бинарной классификации, поэтому для классификации с пересекающимися классами используются ансамблевые подходы. Данные алгоритмы сводят исследуемую задачу к последовательности бинарных классификаторов, учитывающих возможную корреляцию между классами [8–9]. Кратко опишем используемые в работе методы.

Независимая классификация (binary relevance) предполагает, что все классы независимы, и определяет принадлежность объекта к каждому отдельным классификатором. Исходная задача преобразована в 15 отдельных задач бинарной классификации (по одной для каждой метки), где в качестве классификатора использован метод опорных векторов. К недостаткам данного алгоритма можно отнести то, что он не учитывает возможные зависимости между метками, а также низкую точность предсказаний при дисбалансе классов.

Классификация с несколькими выходами (MultiOutput classifier) — обобщение задачи мультиметочной классификации, когда объект классифицируется по нескольким критериям (целям). Алгоритм состоит в подгонке одного классификатора для каждой цели, что позволяет использовать несколько классификаций целевых переменных. Это простая стратегия расширения классификаторов, которые изначально не поддерживают многоцелевую классификацию. В качестве используемого классификатора в работе использовался алгоритм случайный лес.

LabelPowerset преобразует задачу мультиметочной классификации в задачу классификации с несколькими классами, состоящими из всех уникальных комбинаций меток, обнаруженных в обучающих данных. Для моделирования корреляции классов решается единственная мультиклассовая задача с  $2^L$  возможными значениями класса. Используемый в работе классификатор — многоклассовый метод опорных векторов. К сложностям

данного алгоритма относятся множество полученных классов, при этом на каждую уникальную комбинацию меток приходится немного обучающих примеров, что повышает риск переобучения модели.

Мультиметочный подход ленивого обучения (MLkNN) относится к группе так называемых адаптированных алгоритмов. Данный метод использует метод  $k$  ближайших соседей для поиска ближайших примеров к тестируемому классу и использует байесовский вывод для выбора назначенных меток.

Цепной классификатор (chain classifier) строит цепочку двоичных классификаторов  $c_0, c_1, \dots, c_L$ , где классификатор  $c_i$  использует предсказания предыдущего классификатора  $c_j$ , где  $j < i$  в качестве дополнительного атрибута признаков. В качестве бинарного классификатора также использовался метод опорных векторов. Для тестового объекта  $\tilde{x}$  цепной классификатор рассчитывает значения  $\tilde{y}_1 = c_1(\tilde{x}), \tilde{y}_2 = c_2(\tilde{x}, \tilde{y}_1), \dots, \tilde{y}_L = c_L(\tilde{x}, \tilde{y}_1, \dots, \tilde{y}_{(L-1)})$  и в качестве результата возвращает вектор  $\tilde{y} = [\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_L]$ .

**3. Результаты.** Для оценки результатов обучения были использованы следующие метрики [10–11]:

- Точный коэффициент соответствия (accuracy score) — самый строгий показатель, показывающий процент образцов, для которых все их ярлыки классифицированы правильно. Рассчитывается по формуле

$$\text{accuracy score}(a, X) = \frac{1}{l} \sum_{i=1}^l \frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|},$$

где  $Y_i$  — множество классов, которым объект  $x_i \in X$  принадлежит на самом деле, а через  $Z_i$  — множество классов, к которым объект был отнесен алгоритмом  $a(x)$ . Недостаток этой метрики состоит в том, что она не учитывает частичные совпадения классов.

- Хэммингово расстояние (hamming loss) — доля классов, факт принадлежности которых угадан неверно. Данную метрику необходимо минимизировать и ее расчет происходит по формуле

$$\text{hamming loss}(a, X) = \frac{1}{l} \sum_{i=1}^l \frac{|Y_i \setminus Z_i| + |Z_i \setminus Y_i|}{K},$$

- F-мера — гармоническое среднее точности и полноты

$$F = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}},$$

достигающее оптимального значения при 1 и худшего значения при 0. При микросреднении (Fbeta score micro) характеристики точности и полноты усредняются по всем классам, а затем вычисляется итоговая метрика. При макро-усреднении (Fbeta score macro) сначала вычисляется итоговая метрика точности и полноты для каждого класса, а затем результаты усредняются по всем классам.

Результаты обучения рассмотренных моделей представлены в таблице 2.

Сравнительная оценка метрик моделей показала, что наиболее близкие к оптимальным значениям метрик качества обучения имеет ансамблевый метод классификации с несколькими выходами (MultiOutput classifier), где в качестве бинарного классификатора используется алгоритм случайный лес. Близкое к нему значение точного коэффициента соответствия имеет алгоритм LabelPowerset. Точность предсказания каждого класса для моделей-„победителей“ представлена на рис. 2.

Таблица 2

Показатели метрик качества классификации с пересекающимися классами

Модель	Accuracy score	Hamming loss	Fbeta score macro	Fbeta score micro
Независимая классификация	0.637	0.0487	0.3577	0.8913
Классификация с несколькими выходами	0.7188	0.0335	0.5536	0.9284
LabelPowerset	0.6875	0.0446	0.5339	0.8958
Мультиметочный подход ленивого обучения	0.5750	0.0603	0.2376	0.8646
Цепной классификатор	0.6063	0.0616	0.3149	0.8562

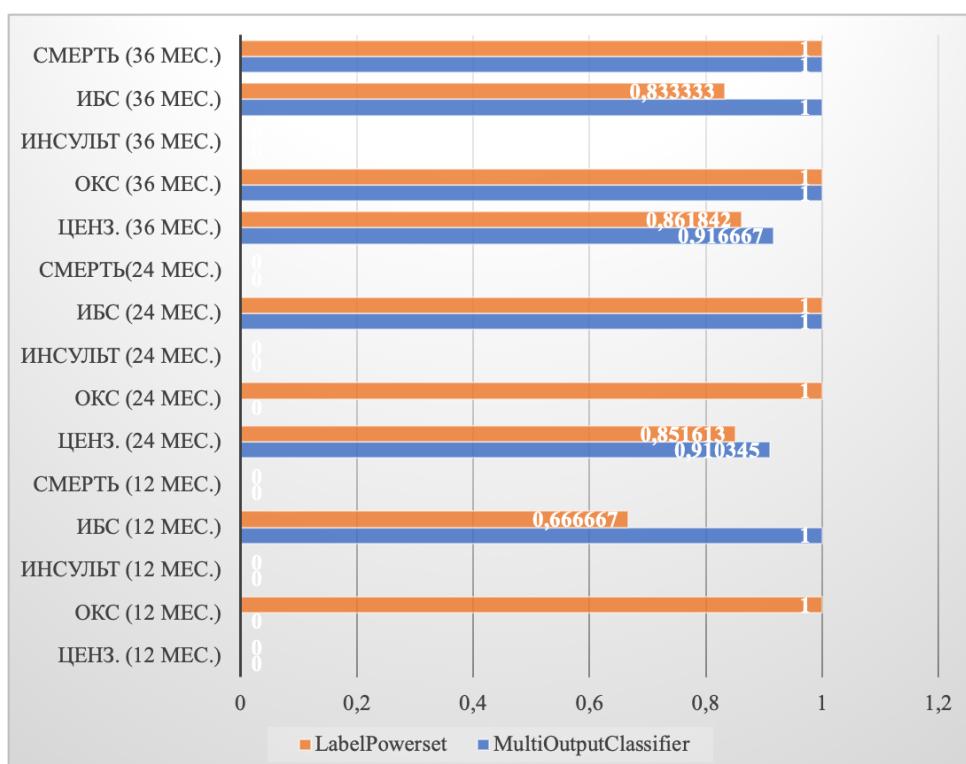


Рис. 2. Точность предсказания каждого класса для MultiOutput classifier и LabelPowerset

Итак, алгоритмы машинного обучения оказались наиболее эффективными и точными способами расчета риска на малом наборе региональных данных для таких ССЗ как острый коронарный синдром и ишемическая болезнь сердца 12, 24 и 36 месяцев. Также с максимальной точностью алгоритм прогнозирует цензурированные наблюдения. Данные результаты позволяют говорить о возможности использования моделей машинного обучения для повышения точности прогнозирования указанных заболеваний. Таким образом, подходы машинного обучения открывают перспективу достижения более индивидуализированной оценки риска ССЗ и лучшей адаптации управления рисками к отдельным пациентам. Практическое внедрение полученных моделей в клиническую практику может улучшить качество первичной профилактики, что поможет предупредить или

отсрочить развитие ССЗ. В работе с давно наблюдающимися пациентами полученные модели также могут служить эффективным инструментом контроля динамики таких модифицируемых факторов как избыточная масса тела, сахарный диабет, недостаточная физическая активность и др.

## Список литературы

1. Баланова Ю. А., Шальнова С. А., Имаева А. Э., Капустина А. В., Муромцева Г. А., Евстифеева С. Е., Тарасов В. И., Редько А. Н., Викторова И. А., Прищепа Н. Н., Якушин С. С., Бойцов С. А., Драпкина О. М. Распространенность артериальной гипертонии, охват лечением и его эффективность в Российской // Рациональная Фармакотерапия в Кардиологии. 2019. № 15 4. С. 450–466.
2. Симерзин В. В., Гаглоева И. В., Гарькина С. В. Современная концепция профилактики сердечно-сосудистых заболеваний // Вестник СамГУ. Естественнонаучная серия. 2007. № 9/1 59. С. 296–306.
3. Conroy R. M., Pyorala K., Fitzgerald A. P. et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project // European Heart Journal. 2003. N 24. P. 987–1003.
4. Гусев А. В., Гаврилов Д. В., Корсаков И. Н., Серова Л. М., Новицкий Р. Э., Кузнецова Т. Ю. Перспективы использования методов машинного обучения для предсказания сердечно-сосудистых заболеваний // Искусственный интеллект в здравоохранении. 2019. № 3. С. 41–47.
5. Poplin R., Varadarajan A. V., Blumer K., et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning // Nat Biomed Eng. 2018. V. 2. N 3. P. 158–164.
6. Quesada J. A., Lopez-Pineda A., Gil-Guillen V. F., et al. Machine learning to predict cardiovascular risk // The international journal of clinical practice. 2019. V. 73. N 10. e13389.
7. Zack C. J., Senecal C., Kinar Y. et al. Leveraging Machine Learning Techniques to Forecast Patient Prognosis After Percutaneous Coronary Intervention // JACC Cardiovasc Interv. 2019 V. 12. N 14. P. 1304–1311.
8. Tai F., Lin H.-T. Multilabel Classification with Principal Label Space Transformation. // Neural Comput., 24–9, 2012. P. 2508–2542.
9. Gibaja E., Ventura S. A Tutorial on Multilabel Learning // ACM Computing Surveys. 2015. Vol. 47, N. 3. P. 1–38.
10. Tsoumakas G., Katakis I., Vlahavas I. Data Mining and Knowledge Discovery Handbook. Springer, 2010. P. 667–685.
11. Zhang M.-L., Zhou Z.-H. A Review on MultiLabel Learning Algorithms // IEEE Transactions on Knowledge and Data Engineering. 2014. Vol. 26, Iss. 8. P. 1819–1837.



**Каледин Олег Евгеньевич** — канд. физ.-мат. наук, доцент кафедры прикладной математики, дифференциальных уравнений и теоретической механики Мордовского государственного университета им. Н. П. Огарева.

В 2006 году окончил МГУ им. Н. П. Огарева, в 2011 году защитил диссертацию на соискание степени кандидата физико-математических наук по специальности 05.13.18 „Математическое моделирование, численные методы и комплек-

сы программ“. Имеет 23 научные публикации и 4 свидетельства о государственной регистрации программ на ЭВМ. Область научных интересов: разработка программного обеспечения, математическое моделирование.

E-mail: [kaledinoe@gmail.com](mailto:kaledinoe@gmail.com). Почтовый адрес: 431751, Республика Мордовия, с. Большие Березники, ул. Мостовая, д. 16. Контактный телефон: 89276406007

**Kaledin Oleg Evgenievich** — graduated from Ogarev Mordovia State University of Faculty of Mathematics and Information Technology in 2006. Ph. D. in Physical and Mathematical

Sciences in 2011. Associate Professor of the Department of Applied Mathematics, Differential Equations and Theoretical Mechanics, Ogarev Mordovia State University of Faculty of Mathematics and Information Technology.

Research interests: software development, mathematical modeling.

E-mail: [kaledinoe@gmail.com](mailto:kaledinoe@gmail.com). Address: 431751, Republic of Mordovia, Big Berezniki v., Mostovaya st., 16. Phone: 89276406007



**Каледина Елена Александровна** — канд. физ.-мат. наук, доцент кафедры прикладной математики, дифференциальных уравнений и теоретической механики Мордовского государственного университета им. Н. П. Огарева. E-mail: [elena.lizina@gmail.com](mailto:elena.lizina@gmail.com).

В 2009 году окончила МГУ им. Н. П. Огарева, в 2015 году защитила диссертацию на соискание степени кандидата физико-математических наук по специальности 05.13.18 „Математическое моделирование, численные методы и комплексы программ“. Имеет 25 научных публикаций. Область научных интересов: математическая теория устойчивости, теория управления, машинное обучение.

Почтовый адрес: 430009, Республика Мордовия, г. Саранск, ул. Девятаева, д. 13, кв. 71.

Контактный телефон: 89876993780

**Kaledina Elena Alexandrovna** — graduated from Ogarev Mordovia State University of Faculty of Mathematics and Information Technology in 2009. Ph.D. in Physical and Mathematical Sciences in 2015.

Associate Professor of the Department of Applied Mathematics, Differential Equations and Theoretical Mechanics, Ogarev Mordovia State University of Faculty of Mathematics and Information Technology.

Research interests: mathematical theory of stability, control theory, machine learning.

E-mail: [elena.lizina@gmail.com](mailto:elena.lizina@gmail.com). Address: 430009, Republic of Mordovia, Saransk, st. Devyataeva, 13, apt. 71. Phone: 89876993780



**Кулягина Таисия Ивановна** — инженер-программист ИП Вильгельм. E-mail: [t.kulyagina@headsandhands.ru](mailto:t.kulyagina@headsandhands.ru).

В 2021 году окончила магистратуру Мордовского государственного университета им. Н. П. Огарева по направлению 01.04.02 „Прикладная математика и информатика“. Область научных интересов: разработка программного обеспечения.

Почтовый адрес: 4300016, Республика Мордовия, г. Саранск, ул. Богдана Хмельницкого, д. 61а, кв. 7. Контактный телефон: 89022308016

**Kulyagina Taisiya Ivanovna** — software engineer, IE Wilhelm.

Graduated from the magistracy in the direction 01.04.02 „Applied Mathematics and Informatics“ of the Ogarev Mordovia State University of Faculty of Mathematics and Information Technology in 2021.

E-mail: [t.kulyagina@headsandhands.ru](mailto:t.kulyagina@headsandhands.ru) Address: 4300016, Republic of Mordovia, Saransk, st. Bohdan Khmelnitsky, d. 61a, apt. 7. Phone: 89022308016

Дата поступления — 01.02.2022