

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Аржаник Александра Алексеевна

Выпускная квалификационная работа

**Сравнительный анализ прогностических моделей для
поддержки принятия врачебных решений**

Уровень образования:

Направление *01.04.02 «Прикладная математика и информатика»*

Основная образовательная программа: *ВМ.5691 «Прикладная математика и информатика в задачах медицинской диагностики»*

Научный руководитель:

старший преподаватель кафедры ТСУЭФА,

к.ф.-м.н.,

Гончарова Анастасия Борисовна

Рецензент:

профессор кафедры медицинской

и биологической кибернетики СибГМУ,

д.м.н., доцент

Карась Сергей Иосифович

Санкт-Петербург

2021 год

Оглавление

Введение	3
Постановка задачи	5
Существующие методы решения задачи	6
1. Исходные данные	8
2. Подготовка данных	9
2.1. Выявление факторов, значимых для прогнозирования	9
2.1.1. Количественные переменные	10
2.1.2. Качественные переменные	11
2.2. Разбиение базы на обучающую и тестовую выборки	12
2.3. Преобразование количественных данных в бинарные	14
2.4. Подсчет частот	16
3. Прогностические модели	18
3.1. Модели, построенные по всем данным	18
3.1.1. Байесовский классификатор	18
3.1.2. Логистическая регрессия	21
3.2. Модели, построенные по параметрам CURB-65	24
3.2.1. Шкала CURB-65 и ее модификация	24
3.2.2. Байесовский классификатор	28
3.2.3. Логистическая регрессия	29
Выводы	32
Заключение	33
Список использованных источников	35

Введение

Для корректной постановки диагноза врачу необходимо обработать большой объем информации о пациенте, включающий в себя данные о жалобах и анализах, сопутствующих заболеваниях, результатах медицинских исследований и другие. К тому же медицина стремительно развивается, появляются новые методики исследований, растет объем знаний, но время, отведенное на прием одного пациента, остается неизменным. Обработка такого количества информации в ограниченное время – это сложная задача даже для опытных врачей, не говоря о молодых специалистах. В связи с этим растет число врачебных ошибок, которые приводят к различным негативным явлениям – от снижения эффективности работы систем здравоохранения до роста смертности [1]. Помочь предотвращать такие ошибки может система поддержки принятия решений.

В общем виде под термином «система поддержки принятия решений» понимается компьютерная система, которая путем сбора и анализа информации может влиять на процессы принятия решений в различных областях человеческой деятельности, в том числе – в медицине. В здравоохранении такие системы называются уточненным термином «системы поддержки принятия врачебных решений» (СППВР) [2].

Основываясь на результатах, полученных с помощью СППВР, врачами могут приниматься решения о госпитализации или помещении пациента в отделение интенсивной терапии, также с помощью них можно оценивать риски развития осложнений.

Некоторые компании занимаются разработкой таких систем. Например, компания «К-скай» на основе анализа данных о пациенте строит прогнозы развития некоторых заболеваний [3]. Аналогичный продукт реализован в компании «Соцмедика». В проекте «Медицинские скрининг системы» реализована СППВР в радиологии – с помощью нее можно детектировать малозаметные признаки различных патологий.

Также в 2019-м году был издан указ президента Российской Федерации «О развитии искусственного интеллекта» с целью обеспечения его ускоренного развития. Согласно указу, должны быть созданы условия для эффективного взаимодействия государства, организаций и граждан в сфере развития искусственного интеллекта (ИИ), что может позволить российским технологиям ИИ занять значительную долю мирового рынка [4].

Таким образом, можно отметить, что процесс реализации систем поддержки принятия решений в медицине не стоит на месте, с каждым годом появляются новые данные, на основе которых можно разрабатывать новые продукты или дополнять уже имеющиеся.

В основе каждой СППВР лежит некоторая прогностическая модель. Прогностические модели играют важную роль в медицине. С помощью них можно прогнозировать тяжесть заболевания, осложнения, исход и некоторые другие параметры.

Подходов к построению прогностических моделей существует огромное множество, у всех есть свои преимущества и недостатки и пределы применимости. Также в медицинских исследованиях важны не только оценки прогноза, но и параметры, значимые для прогнозирования, так как они показывают, на что стоит обращать особенное внимание.

Цель данной работы – провести сравнительный анализ прогностических моделей для поддержки принятия врачебных решений.

Постановка задачи

Для всех моделей постановку задачи можно сформулировать следующим образом: имеется ряд параметров $\{X_i\}$ ($i = \overline{1, n}$), в них могут входить:

- симптомы;
- анамнез заболевания;
- сопутствующие заболевания;
- другие данные о пациенте.

Также имеется ряд диагнозов $\{D_j\}$ ($j = \overline{1, m}$). Под диагнозами могут подразумеваться непосредственно диагнозы, исходы заболевания, стратегии лечения.

Необходимо на основе данных параметров поставить диагноз, то есть сопоставить набору параметров X один или несколько наиболее вероятных диагнозов D_j .

Существующие методы решения задачи

Анализ публикаций в научной литературе по теме СППВР показал, что разработки и исследования в этой области ведутся во всем мире и в различных направлениях не менее 30 лет. Причем в последние два года в России наблюдается существенный рост в динамике публикаций по этой теме [1].

Данную задачу можно решать с помощью методов машинного обучения.

Один из методов машинного обучения, часто используемых в медицине – *логистическая регрессия*. Данный метод используется для прогнозирования вероятности принадлежности объекта к классу.

С помощью множественной логистической регрессии в [5] (2011) выделены факторы риска предиабета, в [6] (2010) были выделены факторы риска внебольничной пневмонии. В работе [7] модель для принятия хирургических решений строится на основе бинарной логистической регрессии и реализуется с помощью нейронных сетей. Авторы решали задачу с небольшим количеством входных и выходных данных (3 варианта исхода). Процент верной классификации для всех моделей превышает 90%.

В [8] (2013) приведен сравнительный анализ методов машинного обучения для обнаружения пневмонии у детей (метод k-средних показал лучшую точность, наивный байесовский – самую низкую).

В [9] (2015) разные методы были применены к данным по болезням легких у детей (наибольшая чувствительность у случайного леса, метода опорных векторов и k-средних, наименьшая – у Байеса). В [10] (2013) сравниваются нейронная сеть, дерево решений и метод опорных векторов на данных по диабету (лучший результат показала нейронная сеть).

В статье [11] рассматривается реализация системы через дерево принятия решений для пациентов с артериальной гипертонией. Процент правильной постановки диагноза около 77%, однако, выборка пациентов, для которых определялся диагноз, довольно мала (всего 18 человек).

Аналогичный подход к реализации СППР в области дерматовенерологии предлагается в статье [12].

Также предпринимаются попытки построения СППР на основе *нечетких множеств*, например, в работах [13] и [14]. Такие модели не всегда удобно использовать на практике, но подход довольно хорошо подходит для медицинских данных, так как имеется возможность учитывать важность симптомов и вероятности их появления при определенных болезнях.

В работе [15] рассматривается реализация системы поддержки принятия решений через *запросы в базу знаний* для больных с онкопатологиями. Преимуществом такого подхода является постоянное пополнение базы, но среди недостатков можно выделить определение «*существенных и несущественных*» параметров, так как в реальности несущественный симптом может оказаться существенным для постановки диагноза. Также при таком подходе довольно трудно работать с качественными признаками, оценивать их важность, а также учитывать все возможные синонимы, если симптом задается текстовыми переменными. Результаты тестирования данного подхода не приводятся.

В работах [16], [17] реализация системы сводится к задаче *принятия решения в условиях риска*. Такой подход легко реализуется на небольшом количестве входных и выходных данных, однако сложен, если данных много. В статьях не приводятся результаты тестирования данного подхода.

Таким образом, после анализа публикаций можно сказать, что на разных данных одни и те же методы могут давать разные результаты – невозможно однозначно выбрать метод, который будет лучшим для всех задач, и в каждом исследовании имеет смысл использовать несколько подходов для выбора оптимального.

1. Исходные данные

Модели будут разрабатываться на основе базы данных по всем пациентам, госпитализированным с внебольничной пневмонией во все стационары Томской области в 2017 году.

Внебольничные пневмонии являются одной из основных причин смерти во всем мире. Инфекции нижних дыхательных путей занимают четвертое место по смертности в мире (данные Всемирной организации здравоохранения 2016 г. [18]). В этой связи актуальным является проведение исследования поиска факторов летальности и осложнений внебольничной пневмонии.

Исходная база содержит данные о 239 различных параметрах 1412 пациентов.

В качестве прогнозируемой переменной будет использоваться летальный исход.

Таким образом, необходимо на основе информации о пациенте при поступлении спрогнозировать исход заболевания.

2. Подготовка данных

Пневмония – стремительно развивающееся заболевание, которое может за короткий срок привести к состоянию, угрожающему жизни. Поэтому при поступлении пациента в стационар, необходимо правильно оценить его состояние и определить необходимость госпитализации. Следовательно, прогнозировать исход необходимо на основе данных пациента, полученных в момент его поступления в стационар.

В базе содержится информация о лечении и об анализах, проведенных уже после госпитализации, в прогностической модели эти данные использоваться не будут.

Подготовка данных включает в себя следующие этапы:

1. выявление факторов, значимых для прогнозирования;
2. разбиение базы на обучающую и тестовую выборки для построения и проверки моделей соответственно;
3. перевод количественных данных в бинарные для моделей, в которых возможно использование только бинарных данных, а также расчет частот встречаемости.

2.1. Выявление факторов, значимых для прогнозирования

Исходная база содержит данные о большом количестве параметров, измеренных у 1412 пациентов. Очевидно, что не все параметры стоит учитывать при прогнозировании исхода. Необходимо снизить размерность данных, то есть выделить те данные, которые влияют на летальность.

Пациенты были разбиты на две группы: умершие и выжившие. Сравнения проводились по каждому параметру.

Для определения различий между группами использовались:

1. критерий Манна-Уитни для количественных переменных;
2. критерий Пирсона для качественных переменных.

2.1.1. Количественные переменные

Так как исследуются две независимые выборки (две группы пациентов) и в группах нет нормального распределения, для исследования различий между группами использовался критерий Манна-Уитни. Данный критерий используется для оценки различий между двумя независимыми выборками по уровню количественного признака [19].

Расчеты проведены в пакете Statistica 10. В таблице 1 представлены значения статистики и p -уровня для каждого параметра.

Таблица 1. Результаты расчета критерия Манна-Уитни

	U	p -уровень
Возраст	23786,0	<0,0001
Индекс массы тела	23306,5	0,0129
Систолическое давление	18605,5	<0,0001
Диастолическое давление	18000,0	<0,0001
Частота сердечных сокращений	25305,5	<0,0001
Частота дыхания	15115,0	<0,0001
Сатурация	9144,5	<0,0001
Температура	14319,5	<0,0001
Гемоглобин	28806,0	0,0144
Эритроциты	26421,5	0,0006
Тромбоциты	21534,0	0,0001
Сегментоядерные	10976,0	<0,0001
Лейкоциты	28531,5	0,0164
Палочкоядерные + юные	10787,0	0,0002
Гематокрит	15988,5	0,0043
Билирубин	13536,0	0,0023
Общий белок	8371,5	<0,0001
Глюкоза	28230,0	0,0269
Аланинаминотрансфераза (АЛТ)	22312,5	<0,0001
Аспаратаминотрансфераза (АСТ)	23301,5	0,0002
Мочевина	9096,0	<0,0001
Креатинин	17289,5	<0,0001
Na+	9150,5	0,8454
K+	9808,0	0,9846
Альбумин	877,0	0,0006

Для тех параметров, у которых значения p -уровня меньше 0,05, различия являются статистически значимыми, следовательно, их имеет смысл использовать для прогнозирования.

2.1.2. Качественные переменные

Так как в базе содержатся данные о значительном количестве пациентов, то есть объем выборки большой, а также – выборки независимые (две группы пациентов), для исследования различий между группами по качественным переменным использовался критерий Пирсона.

Критерий χ^2 Пирсона – это непараметрический метод, который позволяет оценить значимость различий между фактическим количеством качественных характеристик выборки, попадающих в каждую категорию, и теоретическим количеством, которое можно ожидать в изучаемых группах при справедливости нулевой гипотезы (различий нет). Метод позволяет оценить статистическую значимость различий двух относительных показателей [20].

В таблице 2 приведены результаты расчета критерия χ^2 в пакете Statistica 10 для тех показателей, для которых значение p -уровня меньше 0,05, то есть различия являются статистически значимыми. Данные показатели имеет смысл учитывать при прогнозировании.

Таблица 2. Результаты расчета критерия χ^2

	χ^2	p-уровень
Кашель	82,1271	<0,0001
Лихорадка	47,7773	<0,0001
Боль в грудной клетке	19,0682	<0,0001
Пневмония в данном году впервые	4,3859	0,0362
Острое начало	18,3417	<0,0001
Переохлаждение	5,2995	0,0213
Ишемическая болезнь сердца	53,9358	<0,0001
Сахарный диабет	8,7788	0,0031
Хроническая сердечная недостаточность	56,8728	<0,0001
Заболевания почек	19,4182	<0,0001
Хроническая болезнь почек (стадии 3-5)	16,0488	<0,0001
Острые нарушения мозгового кровообращения в анамнезе	33,6630	<0,0001
Алкоголизм	22,3613	<0,0001
Неврологические нарушения	86,6407	<0,0001
Курение в анамнезе	8,0549	0,0178
Нарушение сознания в связи с пневмонией	340,3197	<0,0001

2.2. Разбиение базы на обучающую и тестовую выборки

База была разбита на обучающую и тестовую выборки в соотношении 2:1. Разбиение производилось следующим образом:

1. Была сгенерирована выборка из равномерного распределения объема, соответствующего количеству пациентов (1412 значений от 0 до 1).
2. Каждому значению из выборки был сопоставлен пациент.
3. Все пациенты, которым было присвоено число из промежутка $[0; 1/3]$, были определены в тестовую группу, остальные пациенты – в обучающую.

При разбиении должно выполняться условие – между группами не должно быть статистически значимых различий. Для проверки наличия различий были рассчитаны критерии Манна-Уитни (для количественных переменных) и χ^2 (для качественных) для параметров, определенных как значимые для прогнозирования.

Результаты расчетов представлены в таблицах 3 и 4.

Таблица 3. Результаты расчета критерия Манна-Уитни для обучающей и тестовой выборок

	U	p-value
Возраст	212074	0,1977
Индекс массы тела	192253	0,4814
Систолическое давление	217494	0,9856
Диастолическое давление	214143	0,6063
Частота сердечных сокращений	204494	0,2036
Частота дыхания	194843	0,6264
Сатурация	140588	0,4621
Температура	132843	0,0990
Гемоглобин	204276	0,1917
Эритроциты	202823	0,2202
Тромбоциты	187122	0,9816
Сегментоядерные	112758	0,8598
Лейкоциты	208296	0,5680
Палочкоядерные + юные	98172,5	0,5077
Гематокрит	111238	0,6754
Билирубин	76734,5	0,9370
Общий белок	81535,5	0,3315
Аланинаминотрансфераза (АЛТ)	189432	0,8680
Аспартатаминотрансфераза (АСТ)	188064	0,7049
Мочевина	176938	0,3353
Креатинин	194143	0,9324
Альбумин	14321,5	0,9850

Таблица 4. Результаты расчета критерия χ^2 для обучающей и тестовой выборок

	χ^2	p-уровень
Кашель	3,7414	0,0531
Лихорадка	0,0148	0,9032
Боль в грудной клетке	0,1773	0,6737
Пневмония в данном году впервые	0,6573	0,4175
Острое начало	0,0004	0,9838
Переохлаждение	0,0000	0,9944
Ишемическая болезнь сердца	0,8983	0,3432
Сахарный диабет	2,8780	0,0898
Хроническая сердечная недостаточность	0,8653	0,3523
Заболевания почек	0,0058	0,9394
Хроническая болезнь почек (стадии 3-5)	0,0221	0,8817
Острые нарушения мозгового кровообращения в анамнезе	0,5456	0,4601
Алкоголизм	0,1408	0,7074
Неврологические нарушения	0,9723	0,3241
Курение в анамнезе	0,2554	0,8801
Нарушение сознания в связи с пневмонией	0,0051	0,9432

По полученным результатам можно отметить, что ни по какому из параметров значимых различий между тестовой и обучающей выборками не наблюдается. Следовательно, данное разбиение приемлемо. В обучающей выборке 942 пациента, в тестовой – 470.

2.3. Преобразование количественных данных в бинарные

В ряде случаев для использования СППР необходимы только бинарные данные. Также их проще хранить, передавать, они лучше воспринимаются человеком.

Существуют различные способы переводы количественных данных в бинарные. Например, можно использовать референсные значения – это те значения, в пределах которых результаты исследований считаются нормой для здорового человека. Эти значения получены путем обследования большого количества здоровых людей с наблюдениями за нормальными для них показателями результатов.

Для множества показателей не существует единственного диапазона нормальных результатов, который можно было бы применить к каждому человеку, так как на эти показатели могут влиять разные факторы, такие как сопутствующие заболевания, место проживания человека и многие другие [21]. Также недостатком данного подхода является то, что не для любого параметра существует понятие «нормы», например, для возраста такого понятия нет.

Другой способ перевода – это использование оценочных шкал. Для пациентов с внебольничной пневмонией существуют оценочные шкалы, которые позволяют разделять пациентов на разные группы риска летальности. Разные шкалы имеют свои преимущества и недостатки, они были разработаны в разных социально-экономических и географических условиях, поэтому ни одна из них не является общепринятой в мире [22].

Референсные значения рассчитываются для здоровых людей и не всегда применимы к больным, кроме того они не учитывают специфику

заболевания. Оценочные шкалы, напротив, учитывают специфику, однако в большинстве своем они разработаны зарубежными исследователями, также шкалы не учитывают региональные и социально-экономические особенности, поэтому не универсальны. В связи с этим в данной работе для перевода количественных данных в бинарные используется ROC-анализ – такой подход не обладает перечисленными выше недостатками.

Характеристические кривые, или ROC-кривые (Receiver Operating Characteristic curve) отражают взаимную зависимость ложноположительных и истинно положительных результатов [23]. ROC-анализом называется процесс построения данных кривых [24]. С помощью ROC-анализа можно выбрать порог для разделения классов – это значение называется *точкой разделения*.

Характеристическая кривая строится по мерам чувствительности и специфичности. Под **чувствительностью** (Se – sensitivity) бинарной модели понимается доля истинно-положительных классификаций в общем числе положительных наблюдений. Это есть доля правильно классифицированных положительных наблюдений. Под **специфичностью** (Sp – specify) модели понимается доля истинно-отрицательных классификаций в общем числе отрицательных наблюдений. По вертикальной оси графика ROC-кривой представляется чувствительность, а по горизонтальной – величина, равная $1 - Sp$.

В качестве точки разделения выбирается такая точка на ROC-кривой, для которой отличие ординат ROC-кривой и диагонали максимально, то есть [24]:

$$J = \max(Se + Sp - 1).$$

Число J называется индексом Йодена (J – Youden index).

В качестве прогнозируемой переменной использовался летальный исход. Для обучающей выборки были рассчитаны точки разделения по значимым показателям (табл. 5).

Таблица 5. Точки разделения для количественных показателей

	Точка разделения
Возраст, лет	>72
Индекс массы тела	≤24,8
Систолическое давление, мм рт. ст.	≤105
Диастолическое давление, мм рт. ст.	≤65
Частота сердечных сокращений, в минуту	>97
Частота дыхания, в минуту	>22
Сатурация, %	≤92
Гемоглобин, г/л	≤122
Эритроциты, ·10¹²	≤4,06
Тромбоциты, ·10⁹	≤191
Сегментоядерные, %	>71
Лейкоциты, ·10¹²	>13,7
Палочкоядерные + юные, %	>6
Гематокрит, %	≤36,8
Билирубин, мкмоль/л	>22,5
Общий белок, г/л	≤63
АЛТ, ед./л	>33
АСТ, ед./л	>34
Мочевина, ммоль/л	>9,5
Креатинин, мкмоль/л	>114
Альбумин, г/л	≤35

В соответствии с точками разделения данные были переведены в бинарный вид.

2.4. Подсчет частот

Для бинарных данных были рассчитаны частоты встречаемости в соответствии с исходными данными, для количественных – в соответствии с данными, переведенными в бинарный вид с использованием ROC-анализа.

Результаты представлены в таблице 6.

Таблица 6. Частоты встречаемости

Количественные параметры, переведенные в бинарные			Качественные параметры		
	Умершие	Выжившие		Умершие	Выжившие
Возраст (>72)	57,47%	27,05%	Кашель	50,57%	87,70%
Индекс массы тела ($\leq 24,8$)	68,57%	51,05%	Сахарный диабет	22,99%	11,83%
Систолическое давление (≤ 105)	55,17%	8,83%	Боль в грудной клетке	5,75%	27,05%
Диастолическое давление (≤ 65)	52,87%	11,31%	Ишемическая болезнь сердца	74,71%	34,54%
Частота сердечных сокращений (>97)	56,32%	27,69%	Пневмония в данном году впервые	5,75%	2,11%
Частота дыхания (>22)	51,85%	8,55%	Переохлаждение	4,60%	13,11%
Сатурация (≤ 92)	68,00%	16,06%	Острое начало	67,82%	43,79%
Гемоглобин (≤ 122)	52,44%	36,79%	Лихорадка	45,98%	79,16%
Эритроциты ($\leq 4,06 \cdot 10^{12}$)	54,88%	32,46%	Хроническая сердечная недостаточность	58,62%	21,78%
Тромбоциты ($\leq 191 \cdot 10^9$)	50,67%	27,58%	Заболевания почек	22,99%	8,32%
Сегментоядерные (>71)	74,58%	40,83%	Хроническая болезнь почек (стадии 3-5)	29,89%	13,70%
Лейкоциты ($> 13,7 \cdot 10^{12}$)	41,98%	23,45%	Острые нарушения мозгового кровообращения в анамнезе	22,99%	5,85%
Палочкоядерные + юные (>6)	43,40%	14,09%	Неврологические нарушения	28,74%	3,75%
Гематокрит ($\leq 36,8$)	61,19%	40,59%	Алкоголизм	19,54%	5,85%
Билирубин ($> 22,5$)	47,89%	22,40%	Курение в анамнезе	9,30%	22,37%
Общий белок (≤ 63)	64,71%	20,50%	Нарушение сознания в связи с пневмонией	52,87%	1,76%
АЛТ (>33)	63,75%	37,47%			
АСТ (>34)	70,00%	41,94%			
Мочевина (>9,5)	72,50%	15,73%			
Креатинин (>114)	66,25%	21,90%			
Альбумин (≤ 35)	75,00%	30,22%			

3. Прогностические модели

В работе рассматриваются следующие подходы к построению моделей:

1. Наивный байесовский классификатор.
2. Логистическая регрессия.
3. Шкала для оценки прогноза тяжести пневмонии (на примере CURB-65 и ее модификации).

3.1. Модели, построенные по всем данным

Модели разрабатывались на двух наборах входных данных:

1. Все значимые параметры, отобранные в пункте 5.1.
2. Параметры, входящие в шкалу CURB-65.

3.1.1. Байесовский классификатор

В основе байесовского классификатора лежит формула Байеса:

$$P(D_j | S) = \frac{P(D_j) \cdot P(S | D_j)}{\sum_{j=1}^d P(D_j) \cdot P(S | D_j)}, j = \overline{1, d},$$

где

$P(D_j | S)$ – вероятность заболевания при данном симптомокомплексе;

$P(D_j)$ – априорная вероятность наличия заболевания D_j ;

$P(S | D_j)$ – вероятность наблюдения симптомокомплекса S для заболевания D_j ;

d – количество возможных заболеваний/исходов [25].

Под симптомом подразумевается любая информация о пациенте, в том числе сопутствующие заболевания, результаты лабораторных анализов и другие показатели.

Цель классификации состоит в том, чтобы понять к какому классу принадлежит объект. Байесовский классификатор использует оценку апостериорного максимума для определения наиболее вероятного класса. То есть необходимо рассчитать вероятность для всех классов и выбрать тот класс, который обладает максимальной вероятностью [26].

В данной задаче имеются два класса – выжившие и умершие пациенты. Необходимо на основе имеющихся данных рассчитать вероятность, что пациент попадет в тот или иной класс.

Для реализации подхода использовались найденные ранее частоты встречаемости симптомов. Расчеты были проведены в пакете MATLAB R2020a. Результаты расчетов для тестовой выборки приведены в таблице 7.

Таблица 7. Результаты применения байесовского классификатора

		Результат классификации	
		Умершие	Выжившие
Летальность	Умершие	33	7
	Выжившие	44	386

Таким образом, чувствительность классификатора составила 0,825, а специфичность – 0,898.

ROC-кривая классификатора представлена на рисунке 1.

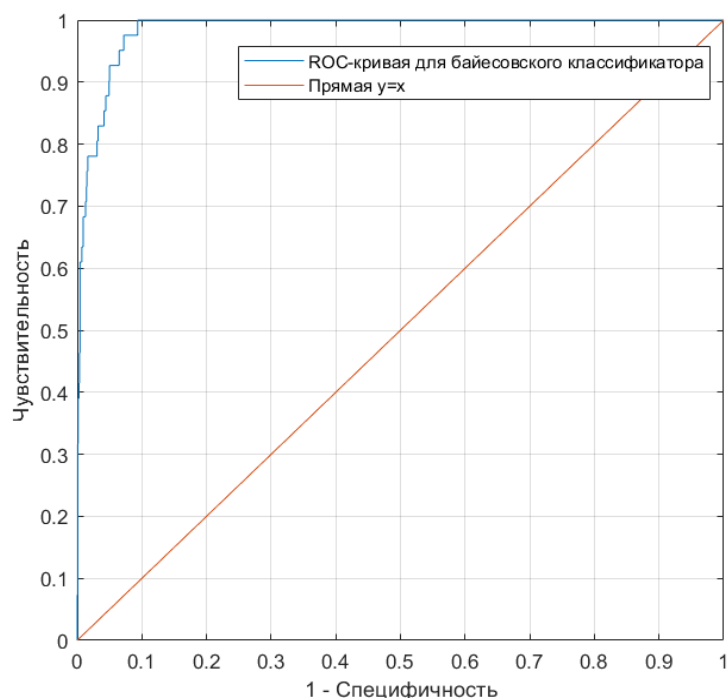


Рис. 1. ROC-кривая для байесовского классификатора в сравнении с прямой $y=x$

Одним из показателей качества классификации является площадь под ROC-кривой (area under curve – AUC). Площадь под кривой изменяется в пределах от 0 до 1. Значение 1 соответствует наилучшему классификатору, 0,5 – случайному, 0 – наихудшему. Чем больше показатель AUC, тем лучшей прогностической силой обладает модель.

AUC байесовского классификатора составила 0,952 (доверительный интервал 0,928–0,969) – это говорит о высоком качестве классификации.

Также были рассчитаны прогностические ценности положительного и отрицательного результатов (ПЦПР и ПЦОР), они составили 0,429 и 0,982 соответственно. Эти результаты можно интерпретировать следующим образом: ПЦПР показывает вероятность летального исхода среди тех, кто имеет положительный результат теста; ПЦОР показывает вероятность отсутствия летального исхода среди тех, кто имеет отрицательный результат теста [27]. Высокий ПЦОР говорит о том, что количество летальных исходов в базе пациентов не высоко.

3.1.2. Логистическая регрессия

Логистическая регрессионная модель предназначена для решения задач предсказания значения непрерывной зависимой переменной, принимающей значения на интервале от 0 до 1.

Используется следующее уравнение регрессии:

$$y = \frac{1}{1 + \exp\left(-\left(b_0 + \sum_{i=1}^n b_i x_i\right)\right)},$$

где

y – вероятность, что произойдет интересующее событие или же что случай будет отнесен к конкретному классу,

x_i – значения независимых переменных,

b_i – коэффициенты, расчет которых является задачей регрессии, $i = \overline{1, n}$.

Предсказанные значение y всегда будут лежать в диапазоне от 0 до 1. Если полученное значение больше 0,5, то случай относится к одному классу, если меньше – к другому [28].

В логистическую модель были включены не все значимые показатели. Критерии отбора параметров, включаемых в модель:

1. Среди параметров с высоким попарным коэффициентом корреляции выбирался только один, чтобы избежать мультиколлинеарности.

2. Некоторые параметры были измерены у небольшого числа пациентов. Включение таких параметров в модель сокращает выборку, так как эти пациенты при обучении модели не будут учитываться.

В таблице 8 представлены значения коэффициентов для каждого параметра, а также их уровни значимости, рассчитанные для обучающей выборки.

Таблица 8. Коэффициенты уравнения регрессии и уровни значимости

	Коэффициент	p-уровень
Острое начало	-1,5463	0,0116
Заболевания почек	-1,5870	0,0301
Алкоголизм	-2,3839	0,0007
Неврологические нарушения	-1,8721	0,0234
Диастолическое давление	0,0439	0,0889
Частота сердечных сокращений	0,0014	0,9397
Частота дыхания	-0,1488	0,0337
Сатурация	0,1225	0,0643
Нарушение сознания	-2,8278	<0,0001
Эритроциты	0,3568	0,2741
Палочкоядерные нейтрофилы	-0,0367	0,3464
Мочевина	-0,1879	0,0003

На рисунке 2 приведена линия регрессии.

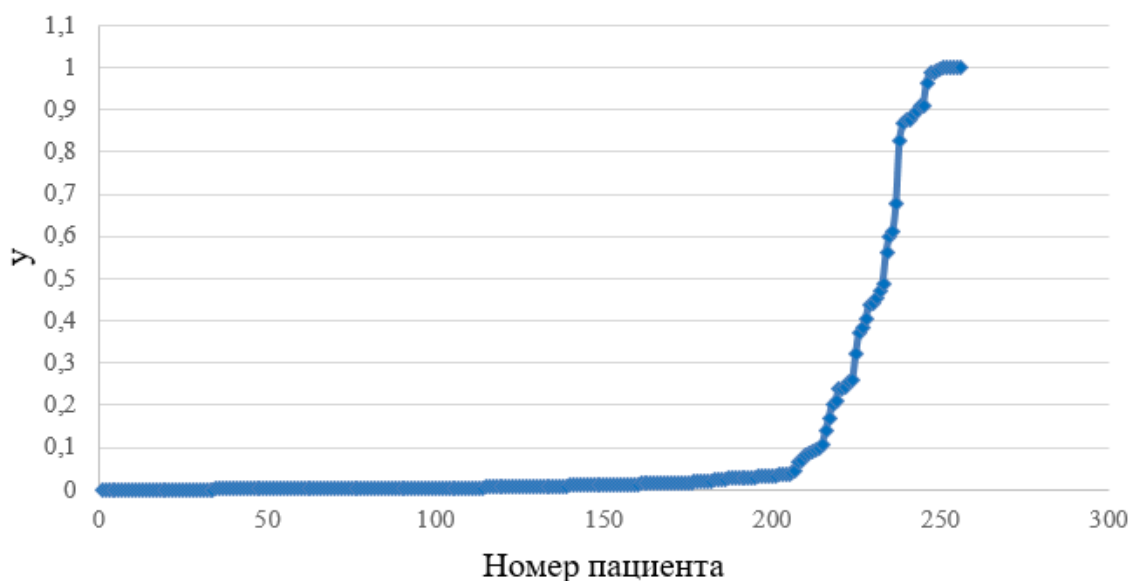


Рис. 2. Линия логистической регрессии

В классической логистической регрессии принято использовать в качестве точки разделения значение 0,5. В таблице 9 представлены результаты применения логистической регрессии с данным пороговым значением.

Таблица 9. Результаты применения логистической регрессии с пороговым значением 0,5

		Результат классификации	
		Умершие	Выжившие
Летальность	Умершие	16	6
	Выжившие	7	227

С точкой разделения 0,5 чувствительность составляет 0,727, специфичность – 0,970.

В данном случае использование такого порога приводит к очень высокой специфичности и низкой чувствительности. Также по рисунку 2 можно отметить, что значение 0,5 не будет давать адекватных результатов. Это может быть связано с тем, что в исходной базе соотношение умерших и выживших не 1:1, а 1:9. Чтобы улучшить классификационные качества модели, пороговое значение было вычислено с помощью ROC-анализа на обучающей выборке, оно составило 0,06.

Результаты применения логистической регрессии на тестовой выборке приведены в таблице 10. Учитывались только те пациенты, у которых имелась информация по всем параметрам, входящим в модель (256 человек).

Таблица 10. Результаты применения логистической регрессии

		Результат классификации	
		Умершие	Выжившие
Летальность	Умершие	20	2
	Выжившие	29	205

Чувствительность составила 0,909, специфичность – 0,876.

ROC-кривая для логистической регрессии представлена на рисунке 3.

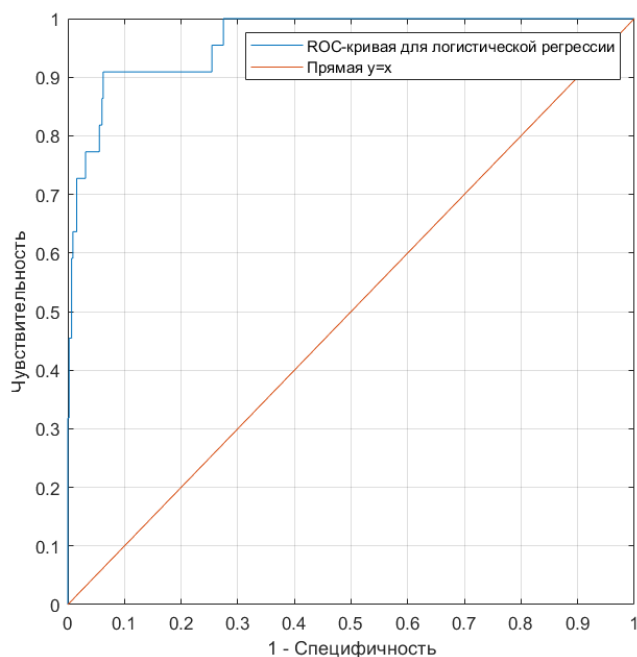


Рис. 3. ROC-кривая для логистической регрессии, построенной по всем параметрам, в сравнении с прямой $y=x$

Площадь под ROC-кривой классификатора равна 0,930. ПЦПР и ПЦОР составили 0,408 и 0,990 соответственно.

3.2. Модели, построенные по параметрам CURB-65

При поступлении пациента в больницу, необходимо как можно быстрее принять решение о госпитализации. Однако не все параметры могут быть измерены достаточно быстро – некоторые требуют времени для проведения лабораторных анализов, а некоторые биохимические показатели могут быть корректно измерены только натошак. В связи с этим было принято решение обучить модели на ограниченном наборе данных – на параметрах, входящих в шкалу CURB-65.

3.2.1. Шкала CURB-65 и ее модификация

Оценку прогноза на основании клинических рекомендаций Минздрава Российской Федерации по внебольничной пневмонии по всем пациентам рекомендуется делать по шкале CURB-65 [29].

Шкала CURB-65 была разработана Lim W.S. et al. [30]. Данная шкала включает в себя показатели, которые могут быть быстро измерены у пациента при поступлении в стационар.

Имеются исследования, подтверждающие эффективность шкалы CURB-65. В исследовании Falcone M. et al. [31] эффективность шкалы показана на 223 пациентах, в исследовании Aronsky D. и Dean N. [32] – на 883 пациентах.

Однако некоторые исследования показывают, что CURB-65 необходимо использовать с осторожностью, так как имеются пациенты, классифицированные как легкие, однако впоследствии умершие [33]. Например, в исследовании Pg A. et al. [34] в группе пациентов, классифицированных как легкие, смертность составила 0,6%.

Шкала CURB-65 включает анализ 5 признаков:

- 1) нарушение сознания, обусловленное пневмонией;
- 2) повышение уровня азота мочевины > 7 ммоль/л;
- 3) частота дыхания ≥ 30 /мин;
- 4) снижение систолического артериального давления < 90 мм рт.ст. или диастолического ≤ 60 мм рт.ст.;
- 5) возраст больного ≥ 65 лет.

Наличие каждого признака оценивается в 1 балл, общая сумма может варьировать от 0 до 5 баллов, риск летального исхода возрастает по мере увеличения суммы баллов.

Для каждого пациента был рассчитан балл по шкале CURB-65 (табл. 11).

Таблица 11. Статистический анализ выборки по шкале CURB-65

Балл	Умершие	Выжившие
0	0 (0,00%)	195 (100,00%)
1	5 (3,62%)	133 (96,38%)
2	12 (12,50%)	84 (87,50%)
3	14 (46,67%)	16 (53,33%)
4	7 (77,78%)	2 (22,22%)
5	2 (100,00%)	0 (0,00%)

Также полученные с помощью ROC-анализа по всем пациентам значения были использованы для модификации шкалы.

Для модифицированной шкалы признаки следующие:

- 1) нарушение сознания, обусловленное пневмонией;
- 2) повышение уровня азота мочевины $> 9,5$ ммоль/л;
- 3) частота дыхания > 22 /мин;
- 4) снижение систолического артериального давления ≤ 105 мм рт.ст.

или диастолического ≤ 65 мм рт.ст.;

- 5) возраст больного > 72 лет.

Для всех пациентов были посчитаны суммы баллов по модифицированной шкале. Результаты представлены в таблице 12.

Таблица 12. Статистический анализ выборки по модифицированной шкале CURB-65

Балл	Умершие	Выжившие
0	0 (0,00%)	247 (100,00%)
1	5 (3,87%)	124 (96,13%)
2	12 (21,05%)	45 (78,95%)
3	10 (47,62%)	11 (52,38%)
4	7 (77,78%)	2 (22,22%)
5	6 (85,71%)	1 (14,29%)

Затем был проделан ROC-анализ для обеих шкал, прогнозируемая переменная – летальный исход. ROC-кривые для обеих шкал представлены на рисунке 4.

Площадь под кривой для модифицированной шкалы (AUC=0,928) больше, чем для исходной (AUC=0,893). Это говорит о том, что, используя новые пороговые значения, можно лучше классифицировать пациентов.

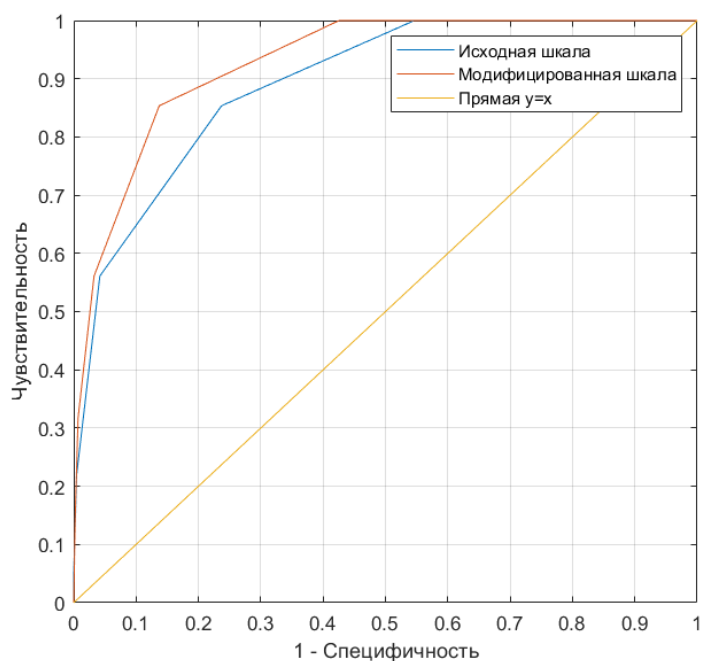


Рис. 4. ROC-кривые для исходной и модифицированной шкал CURB-65

В таблицах 13 и 14 представлены результаты применения исходной и модифицированной шкал на тестовой выборке.

Таблица 13. Результаты применения шкалы CURB-65

		Результат классификации	
		Умершие	Выжившие
Летальность	Умершие	35	5
	Выжившие	102	328

Таблица 14. Результаты применения модифицированной шкалы CURB-65

		Результат классификации	
		Умершие	Выжившие
Летальность	Умершие	35	5
	Выжившие	59	371

Для исходной шкалы чувствительность составила 0,875, специфичность – 0,763. Для модифицированной чувствительность не изменилась, однако специфичность выше – 0,863.

В данной задаче целесообразно максимизировать чувствительность – то есть правильно определять пациентов с высоким риском летального исхода. Если сравнивать по чувствительности, то данные модели не отличаются. Модифицированная шкала лучше классифицирует пациентов, но это обусловлено повышением специфичности.

3.2.2. Байесовский классификатор

Для реализации подхода использовались все параметры, входящие в шкалу CURB-65, а именно:

- систолическое и диастолическое давление;
- мочевины;
- частота сердечных сокращений;
- потеря сознания;
- возраст.

Результаты расчетов для тестовой выборки приведены в таблице 15.

Таблица 15. Результаты применения байесовского классификатора

		Результат классификации	
		Умершие	Выжившие
Летальность	Умершие	35	5
	Выжившие	47	383

ROC-кривая классификатора представлена на рисунке 5.

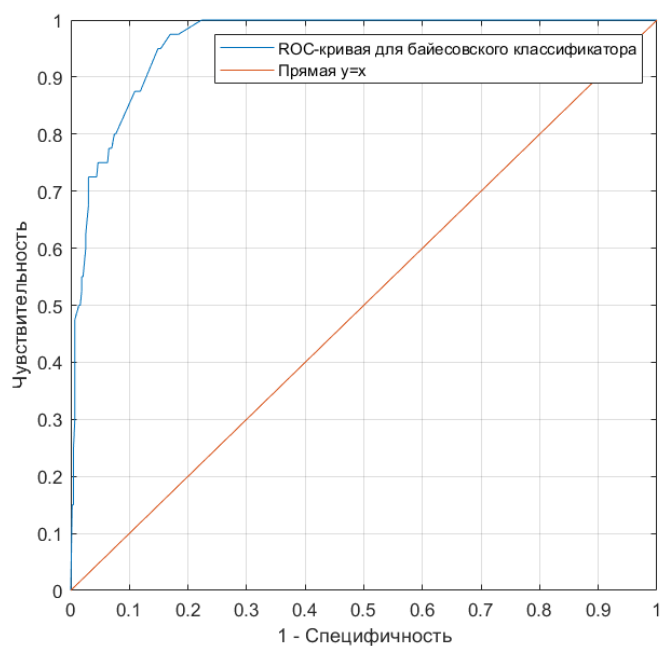


Рис. 5. ROC-кривая для байесовского классификатора, построенного по параметрам CURB-65, в сравнении с прямой $y=x$

Таким образом, чувствительность классификатора составила 0,875, а специфичность – 0,891. AUC классификатора 0,961 – это говорит о высоком качестве классификации.

ПЦПР и ПЦОР составили 0,427 и 0,987 соответственно.

3.2.3. Логистическая регрессия

В логистическую модель были включены показатели, входящие в шкалу CURB-65. В таблице 16 представлены значения коэффициентов для каждого параметра, а также их уровни значимости. Все коэффициенты значимы, следовательно, будут включены в модель.

Таблица 16. Коэффициенты уравнения регрессии и уровни значимости

Параметр	Коэффициент	р-уровень
Возраст	0,0219	0,0184
Систолическое давление	-0,0397	<0,0001
Частота дыхания	0,1717	0,0001
Мочевина	0,1358	<0,0001
Нарушение сознания	2,7235	<0,0001

На рисунке 6 приведена линия регрессии.

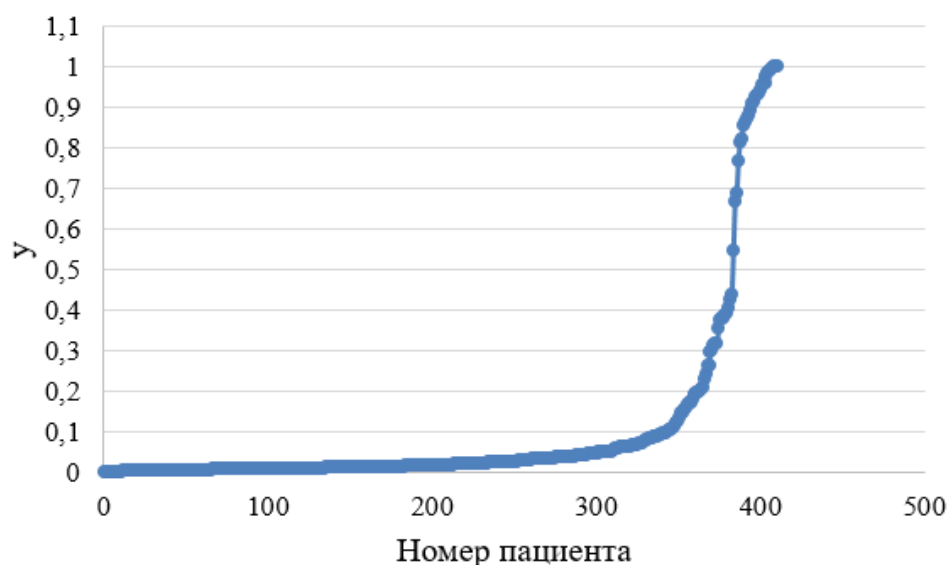


Рис. 6. Линия логистической регрессии

Пороговое значение было вычислено с помощью ROC-анализа на обучающей выборке, оно составило 0,04.

Результаты применения логистической регрессии на тестовой выборке приведены в таблице 17. Учитывались только те пациенты, у которых имелась информация по всем параметрам, входящим в модель (411 человек).

Таблица 17. Результаты применения логистической регрессии

		Результат классификации	
		Умершие	Выжившие
Летальность	Умершие	37	0
	Выжившие	97	277

Чувствительность составила 1,0, специфичность – 0,741.

ROC-кривая для логистической регрессии представлена на рисунке 7.

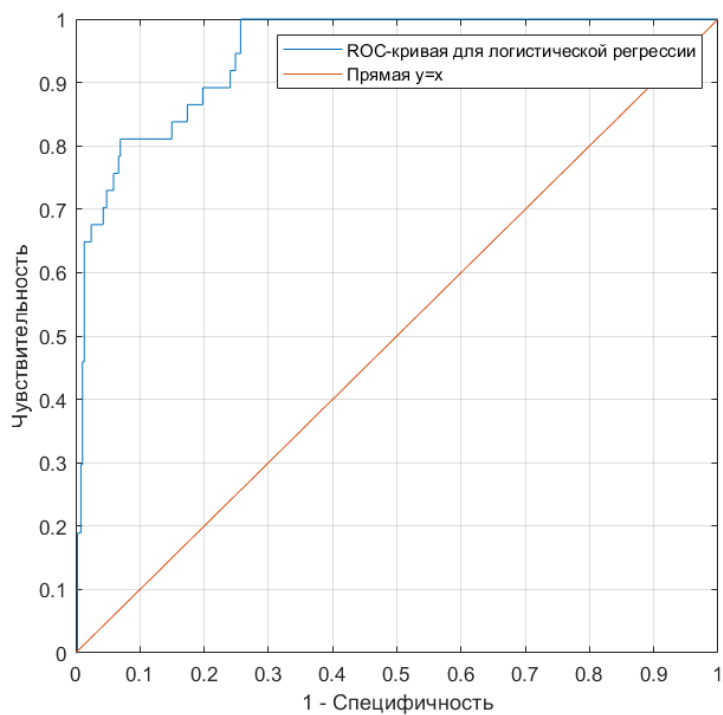


Рис. 7. ROC-кривая для логистической регрессии в сравнении с прямой $y=x$

AUC классификатора составила 0,945.

ПЦПР и ПЦОР составили 0,276 и 1,0 соответственно.

Выводы

В ходе работы были реализованы 4 модели: байесовский классификатор, логистическая регрессия, шкала CURB-65 и ее модификация.

Модели разрабатывались на двух наборах входных данных: все значимые параметры, отобранные в пункте 2.1; параметры, входящие в шкалу CURB-65.

Все характеристики моделей сведены в таблицы 18 и 19.

Таблица 18. Характеристики моделей, построенных на всех значимых параметрах

	Чувствительность	Специфичность	AUC
Байесовский классификатор	0,825	0,898	0,952
Логистическая регрессия	0,909	0,876	0,930

Таблица 19. Характеристики моделей, построенных на параметрах, входящих в шкалу CURB-65

	Чувствительность	Специфичность	AUC
CURB-65	0,875	0,763	0,868
CURB-65 – модификация	0,875	0,863	0,906
Байесовский классификатор	0,875	0,891	0,961
Логистическая регрессия	1,000	0,741	0,945

Если сравнивать модели по набору данных в целом, то лучший результат показывают модели, построенные на основе параметров, входящих в шкалу CURB-65.

Если оценивать качество классификатора по площади под ROC-кривой, то лучше всего классифицирует больных байесовский подход. Но, так как в данной задаче целью является максимизация чувствительности, лучшие результаты показывает логистическая регрессия с измененным пороговым значением, построенная на параметрах, входящих в шкалу CURB-65.

Заключение

Системы поддержки принятия решений играют важную роль в различных областях человеческой деятельности, особенно в медицине. Опираясь на результаты, полученные с помощью таких систем, можно выбирать подходы к дальнейшему лечению пациентов, ставить предварительные диагнозы, прогнозировать исходы болезней.

Существует множество подходов к построению систем, в разных задачах одни и те же подходы могут давать разные результаты. Также каждый подход имеет свои преимущества и недостатки.

В данной работе было проведено сравнение прогностических моделей на базе данных по всем пациентам, госпитализированным с внебольничной пневмонией во все стационары Томской области в 2017 году. Прогнозируемой переменной являлся летальный исход.

На первом этапе были отобраны параметры, значимые для прогнозирования исхода. Все пациенты были разбиты на две группы (умершие и выжившие), были выделены параметры, по которым наблюдались статистически значимые различия. Для реализации этапа использовался программный пакет STATISTICA 10.

Следующим этапом для сравнения моделей база была разбита на тестовую и обучающую выборки. Все модели применялись к тестовой выборке.

Так как для некоторых моделей требуются бинарные данные, все количественные параметры были переведены в бинарный вид с использованием точек разделения, полученных с помощью ROC-анализа. ROC-анализ проводился в пакетах MedCalc и MATLAB R2020a.

Затем были реализованы несколько моделей на разных наборах данных: байесовский классификатор, логистическая регрессия, шкала CURB-65 и ее модификация. Модели были реализованы в пакетах STATISTICA и MATLAB. Лучший результат показала логистическая регрессия с измененным пороговым значением, построенная на параметрах, входящих в

шкалу CURB-65. С помощью нее удалось добиться чувствительности 1,0 – то есть правильно классифицировать всех умерших пациентов.

Результаты работы были представлены на следующих конференциях:

- LI Международная научная конференция аспирантов и студентов «Процессы управления и устойчивость» Control Processes and Stability (CPS'20) (статья «Сравнение способов преобразования количественных данных в бинарные при предсказании рисков осложнений внебольничной пневмонии»), апрель 2020 г.

- IV Международная конференция "Устойчивость и процессы управления", посвященная 90-летию со дня рождения чл.-корр. РАН В.И. Зубова: памяти профессора В.И. Зубова (доклад «Detection of the Community-Acquired Pneumonia Factors Leading to Death»), октябрь 2020 г.

- LII Международная научная конференция аспирантов и студентов «Процессы управления и устойчивость» Control Processes and Stability (CPS'21) (доклад «Сравнительный анализ прогностических моделей для поддержки принятия врачебных решений»), апрель 2021 г.

В дальнейшем можно рассмотреть другие подходы, например, нейронные сети или системы, построенные на нечеткой логике.

Список использованных источников

1. Гусев А. В., Зарубина Т. В. Поддержка принятия врачебных решений в медицинских информационных системах медицинской организации // Врач и информационные технологии. 2017. – № 2. – С. 60–72.
2. Greenes R.A. Clinical decision support: the road ahead / R.A. Greenes. – Boston: Elsevier Academic Press, 2007. – 581 p.
3. Система поддержки принятия врачебных решений Webiomed.DHRA [Электронный ресурс]. – Режим доступа: <https://webiomed.ai/products/webiomed-dhra/> (дата обращения: 20.03.20).
4. Указ Президента Российской Федерации от 10.10.2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации» [Электронный ресурс]. – Режим доступа: <http://www.kremlin.ru/acts/bank/44731> (дата обращения: 20.03.20).
5. R. M. Anjana. Prevalence of diabetes and prediabetes (impaired fasting glucose and/or impaired glucose tolerance) in urban and rural India: Phase I results of the Indian Council of Medical Research–INdia DIABetes (ICMR–INDIAB) study. *Diabetologia* (2011) 54:3022–3027 DOI: 10.1007/s00125-011-2291-5.
6. Gau JT, Acharya U, Khan S, Heh V, Mody L, Kao TC. Pharmacotherapy and the risk for community-acquired pneumonia. *BMC Geriatr*. 2010 Jul 6;10:45. DOI: 10.1186/1471-2318-10-45.
7. Дарвин В.В., Егоров А.А., Микшина В.С., Суровов А.А. Интеллектуальная информационная система поддержки принятия решений хирурга по выбору способа завершения операции // Современные проблемы науки и образования. – 2011. – № 5.
8. Sousa Rafael, Marques Oge, Soares Fabrizzio, Sene Iwens, Oliveira, Leandro, Spoto Edmundo. (2013). Comparative Performance Analysis of Machine Learning Classifiers in Detection of Childhood Pneumonia Using Chest Radiographs. *Procedia Computer Science*. 18. 2579-2582. DOI: 10.1016/j.procs.2013.05.444.

9. Rother A-K, Schwerk N, Brinkmann F, Klawonn F, Lechner W, Grigull L (2015) Diagnostic Support for Selected Paediatric Pulmonary Diseases Using Answer-Pattern Recognition in Questionnaires Based on Combined Data Mining Applications—A Monocentric Observational Pilot Study. PLoS ONE 10(8): DOI: 10.1371/journal.pone.0135180.

10. M.R.K. Mookiah, U. Rajendra Acharya, Roshan Joy Martis, Chua Kuang Chua, C.M. Lim, E.Y.K. Ng, Augustinus Laude, Evolutionary algorithm based classifier parameter tuning for automatic diabetic retinopathy grading: A hybrid feature extraction approach, Knowledge-Based Systems, Volume 39, 2013, Pages 9-22, ISSN 0950-7051, DOI: 10.1016/j.knosys.2012.09.008.

11. Ance Kancere, Ieva Markovica, Zigurds Markovics. Computer system for diagnosis in medicine domain //Technologies of Computer Control. 2013. p. 44-72.

12. Разводин Р.А. Интеллектуальная система поддержки принятия врачебных решений в дерматовенерологии // Проблемы медицинской микологии. 2014. Т. 16. № 3. С. 59–65.

13. Коробова Л.А., Гладких Т.В. Разработка модели принятия решения для постановки диагноза заболеваний на основе нечеткой логики // Вестник ВГУИТ. 2018. Т. 80. № 4. С. 80–89. DOI: 10.20914/2310-1202-2018-4-80-89.

14. Локтюхин В.Н., Черепнин А.А. Поддержка принятия решений на основе нейро-нечеткой технологии при диагностике патологий желудочно-кишечного тракта // Биотехносфера. №2. С-Пб.: Политехника. 2009. С. 20-23.

15. Гуляев Ю.В., Никитов Д.С., Пономарев Н.В. Исследование возможности создания и применения системы поддержки принятия решений при постановке диагноза в области онкопатологии // Радиотехника и электроника. 2006. – №11. – С. 1387-1397.

16. Егоров А. А. Модель принятия решения хирурга / А. А. Егоров, В. С. Микшина // Вестн. новых мед. Технологий, 2011. – Т. 7, № 4. – С. 178 – 181.

17. Мокина Е.Е., Марухина О.В., Шагарова М.Д., Дубинина И.А. Использование методов Data mining при принятии медицинских диагностических решений // Fundamental research. 2016. – № 5. – С. 269-274.

18. 10 ведущих причин смерти в мире [Электронный ресурс]. – Режим доступа: <https://www.who.int/ru/news-room/fact-sheets/detail/the-top-10-causes-of-death> (дата обращения: 10.12.2019).

19. Сидоренко Е.В. Методы математической обработки в психологии / Е.В. Сидоренко. – СПб.: Речь, 2003. – 350 с.

20. Критерий Хи-квадрат Пирсона [Электронный ресурс]. – Режим доступа: <https://medstatistic.ru/methods/methods4.html> (дата обращения: 06.12.2020).

21. Анализы: референсные значения = норма? [Электронный ресурс]. – Режим доступа: <http://www.laboratorii.ru/reference-range/> (дата обращения: 20.01.20).

22. Фесенко О.В., Синопальников А.И. Тяжелая внебольничная пневмония и шкалы оценки прогноза. Практическая пульмонология. 2014; 2: 20-26 / Fesenko O., Sinopalnikov A. Severe community-acquired pneumonia and prognostic scores. Practical pulmonology. 2014, 2: 20-26.

23. Королюк И.П.. Медицинская информатика: Учебник / – 2 изд., перераб. и доп. - Самара : ООО «Офорт» : ГБОУ ВПО «СамГМУ».2012.- 244 с. 2012.

24. Анализ с помощью характеристической кривой [Электронный ресурс]. – Режим доступа: <https://wiki.loginom.ru/articles/roc-analysis.html> (дата обращения: 05.01.2020).

25. Бейли Н. Т. Дж. Математика в биологии и медицине: Пер. с англ. – М.: Мир, 1970. – 326 с.

26. Наивный байесовский классификатор [Электронный ресурс]. – Режим доступа: <http://bazhenov.me/blog/2012/06/11/naive-bayes.html> (дата обращения: 06.12.2020).

27. Красько О. Статистический анализ данных в медицинских исследованиях: учеб. метод. пособие. Минск 2014. – 127 с.

28. Логистическая регрессия [Электронный ресурс]. – Режим доступа: <http://statistica.ru/theory/logisticheskaya-regressiya/> (дата обращения: 10.01.2021).

29. Министерство здравоохранения Российской Федерации Клинические рекомендации. Внебольничная пневмония (МКБ 10: J13-J18) 2018 г. [Электронный ресурс] <http://spulmo.ru> (дата обращения: 05.01.2020).

30. Lim W.S., van der Eerden M.M., Laing R. *et al.* Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax*. 2003; 58(5): 377-82. DOI: 10.1136/thorax.58.5.377.

31. Falcone M, Corrao S, Venditti M, Serra P, Licata G. Performance of PSI, CURB-65, and SCAP scores in predicting the outcome of patients with community-acquired and healthcare-associated pneumonia [published correction appears in *Intern Emerg Med*. 2013 Sep;8(6):549]. *Internal and Emergency Medicine*. 2011;6(5):431-436. DOI: 10.1007/s11739-011-0521-y.

32. Aronsky, Dominik & Dean, Nathan. (2004). Admission Decision For Pneumonia: A Validation Study of the CURB-65 Criteria. *CHEST Journal*. DOI: 10.1378/chest.126.4_MeetingAbstracts.738S-a.

33. Ananda-Rajah MR, Charles PG, Melvani S, Burrell LL, Johnson PD, Grayson ML. Comparing the pneumonia severity index with CURB-65 in patients admitted with community acquired pneumonia. *Scandinavian Journal of Infectious Diseases*. 2008; 40(4):293-300. DOI: 10.1080/00365540701663381.

34. Ilg A., Moskowitz A., Konanki V., *et al.* Performance of the CURB-65 Score in Predicting Critical Care Interventions in Patients Admitted With Community-Acquired Pneumonia. *Ann Emerg Med*. 2019; 74(1): 60-68. DOI: 10.1016/j.annemergmed.2018.06.017.