

ВЫЯВЛЕНИЕ МЕТАСТАЗОВ РАКА ШЕЙКИ МАТКИ С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ

Т. Н. Сушинская¹, А. И. Кузнецов², Е. В. Щепкина³, Н. И. Стуклов⁴, С. В. Епифанова⁵

¹ МНИОИ им. П. А. Герцена — филиала ФГБУ «НМИЦ радиологии» Минздрава России

² Московский Авиационный Институт (МАИ)

³ Институт информационных технологий (факультет), РТУ-МИРЭА, Россия, Москва

⁴ Медицинский институт РУДН

⁵ ФГБУ «ЦКБ с поликлиникой УДП РФ»

Цель исследования: Разработать метод диагностики метастатического поражения регионарных лимфатических узлов у больных раком шейки матки (РШМ) на основе определенных показателей крови методами машинного обучения. Машинное обучение — это эффективный способ классификации данных, особенно в области медицины, где оно широко используется в диагностике.

Материал и методы: 495 пациенток МНИОИ им. П. А. Герцена, филиала ФГБУ НМИЦ радиологии Минздрава России в возрасте 18–82 лет (42 лет), с верифицированным диагнозом РШМ были разделены на 2 группы: с метастазами в регионарных лимфатических узлах и без метастазов. Исследование полученных у каждой больной РШМ 21 показателей крови определило, что наиболее адекватно пациентки были визуально разделены по следующим семи показателям: СОЭ, эритроциты, гемоглобин, фибриноген, Д-димер, агрегация тромбоцитов с АДФ и РФМК.

Для проведения машинного обучения были выбраны следующие алгоритмы: Дерево решений (DT); Случайный лес (RF); К ближайшим соседям (KNN); методы опорных векторов (SVM); Adaboost и XGBoost.

Результаты. Точность построенных моделей на обучающей и тестовой выборках (70 % и 30 %, соответственно) показала, что наиболее точными являются модели Adaboost (100 %, 98 %) и XGBoost (100 %, 98 %) (Рис 2). Оценка моделей по ROC-AUC показала, что XGBoost (100 %, 99,8 %) незначительно лучше, чем Adaboost (100 %, 98,8 %). Однако, R2 показала, что доля объясняющей дисперсии лучше у XGBoost (100 %, 80,5), чем у Adaboost (84,9 %, 66,1 %).

Выводы. Экспериментальные результаты показывают, что «Метод опорных векторов» (SVM) и XGBoost обеспечивают высочайшую точность (98 %) при максимальной объясняющей доле дисперсии (82,9 % и 80,5 %, соответственно).

«Метод опорных векторов» (SVM) является наиболее точным (98 %) с максимальной объясняющей способностью (82,9 %) для диагностики наличия или отсутствия метастазов РШМ в регионарных лимфатических узлах на основе показателей крови.

Ключевые слова: рак шейки матки, РШМ, метастазы, машинное обучение, алгоритмы, модель прогнозирования, диагностика, Accuracy, ROC-AUC, R².

DETECTION OF CERVICAL CANCER METASTASIS USING MACHINE LEARNING ALGORITHMS

T. N. Sushinskaya¹, A. I. Kuznetsov², E. V. Shchepkina³, N. I. Stuklov⁴, S. V. Epifanova⁵

¹ P. A. Gertsen Moscow Scientific Research Oncologic Institute — a Branch of Federal State Budgetary Institution "National Medical Research Center of Radiology" of the Ministry of Healthcare of the Russian Federation

² Moscow Aviation Institute (MAI)

³ Institute of Information Technologies (Faculty), Russian Technological University — Moscow Institute of Radio Engineering, Electronics and Automation, Russia, Moscow

⁴ Medical Institute of Peoples' Friendship University of Russia

⁵ Federal State Budgetary Institution "Central Clinical Hospital with Outpatient Clinic of the Administrative Directorate of the President of the Russian Federation"

Objective of the study is to develop a method of diagnosis of regional lymph node metastasis in patients with cervical cancer (CC) based on certain hematological parameters using machine learning algorithms. Machine learning algorithm is an effective technique of data classification, particularly in medical field where it is extensively used for diagnostic purposes.

Materials and methods. 495 patients of P. A. Gertsen Moscow Scientific Research Oncologic Institute — a Branch of Federal State Budgetary Institution “National Medical Research Center of Radiology” of the Ministry of Healthcare of the Russian Federation aged from 18 to 82 years (42 years) with verified diagnosis of cervical cancer were divided into 2 groups according to the presence or absence of regional lymph node metastasis: with metastases and without metastases. An examination of 21 parameters of blood samples collected from each patient with cervical cancer determined that the patients were most appropriately visually separable based on 7 components: ESR (erythrocyte sedimentation rate), erythrocytes, hemoglobin, fibrinogen, D-dimer, platelet aggregation with ADP (adenosine diphosphate) and SFMC (soluble fibrin monomer complex).

The following algorithms were chosen to conduct machine learning: Decision Tree (DT), Random Forest (RF), K-nearest Neighbors (KNN), Support Vector Machine (SVM), Adaboost and XGBoost.

Results. Accuracy of the models constructed on training and testing samples (70 % and 30 % respectively) highlighted that Adaboost (100 %, 98,8 %) and XGBoost (100 %, 99,8 %) models are the most accurate (Fig.2). The assessment of the models based on ROC-AUC showed that XGBoost (100 %, 99,8 %) is slightly better than Adaboost (100 %, 98,8 %). However, R^2 revealed that XGBoost (100 %, 80,5 %) provides more accurate proportion of explained variance than Adaboost (84,9 %, 66,1 %).

Conclusion. Experimental findings indicate that “Support Vector Machine” technique (SVM) ensures the highest grade of accuracy (98 %) with maximum proportion of explained variance (82,9 % and 80,5 %, respectively).

“Support Vector Machine” is the most precise technique (98 %) with maximum explanatory power (82,9 %) for the diagnosis of the presence of cervical cancer metastases in regional lymph nodes based on hematological parameters.

Keywords: cervical cancer, CC, metastases, machine learning, algorithms, predictive modeling, diagnosis, accuracy, ROC-AUC, R^2 .

Вступление

По данным Всемирной организации здравоохранения (ВОЗ), число заболевших РШМ в мире за последние 20 лет выросло более чем на 110 000 и к 2020 году составило 604 127 человек [1, 2]. За аналогичный период в России число заболевших РШМ увеличилось более чем на 3 000 человек и, по последним статистическим данным, составило 17 503 человека в 2019 году, из которых на I стадии РШМ впервые выявлен у 37,2 %, на II стадии — у 28,9 %, на III стадии — у 22,8 %. В 9,8 % случаев установлена IV стадия заболевания [4].

Максимальный прирост заболеваемости РШМ в России за последние 20 лет у женщин репродуктивного возраста. В возрастной группе 20–24 года прирост составил 127 %, 25–29 лет — 182 %, 30–34 года — 35 %, 35–39 лет — 103 %.

Основными причинами смертности от РШМ является позднее выявление заболевания, связанное с несвоевременным обращением больных. В 2019 году от РШМ умерло 6 389 больных, что составило 36,5 % от числа впервые выявленных случаев [3].

Главными задачами клинициста-онколога являются установление точного диагноза, определение прогноза заболевания и планиро-

вание максимально эффективной стратегии лечения, для чего требуется объективная оценка анатомической распространенности опухолевого процесса. И эти задачи необходимо решить на догоспитальном этапе, до начала специального лечения.

К сожалению, несмотря на использование многочисленных методов лабораторного и инструментального обследований, точность диагностики РШМ с метастазами в регионарных лимфатических узлах на догоспитальном этапе далека от идеальной.

В доступной литературе мы обнаружили ряд исследований, подтверждающих данное утверждение. По опубликованным данным разных авторов, ошибки диагностики метастатического поражения лимфатических узлов у больных РШМ встречаются в 7,9 % — 32 % случаев [4, 5, 6].

Методы машинного обучения и интеллектуального анализа данных — это простые и эффективные способы понимания и прогнозирования. В то же время онколог осматривает пациента и анализирует представленные данные обследования самостоятельно, а затем принимает решение совместно с другими экспертами (консилиум онкологов). Этот процесс требует времени, а результаты зависят от знаний и опыта врачей.

Поэтому исследовательское сообщество предложило автоматизированную систему под названием САД (Компьютерная диагностика) для лучшей классификации наличия метастазов, получения точных результатов и более быстрого внедрения, без необходимости в других узких специалистах. Алгоритмы машинного обучения (MLS) являются одним из вариантов человеческого видения и опыта для помощи врачу в принятии окончательных решений с высокой точностью.

В этом проекте мы использовали данные МНИОИ им. П. А. Герцена, филиала ФГБУ НМИЦ радиологии Минздрава России, о состоянии лимфатической системы у больных РШМ по результатам гистологического исследования после проведения хирургического стадирования. Эти данные применили в алгоритмах машинного обучения для предсказания наличия или отсутствия метастазов. Затем, используя комплексный подход, подходящий для демонстрации и прямой интерпретации их результатов, сравнили алгоритмы.

Обзор литературы

Автоматическая диагностика рака любой локализации является важным перспективным направлением в современном мире. Многочисленные методы современной диагностики злокачественных опухолей не всегда дают удовлетворяющий врачей полноценный результат. Врачи умеют достаточно точно определять размер первичной опухоли и стадировать процесс на основе этих размеров. Однако, когда дело касается метастазов, наличие которых у пациента может существенно влиять на выбор метода специального лечения, часто встречаются ошибки в их диагностике. Ошибки диагностики наличия регионарных метастазов ограничивают возможности органосохраняющего лечения, которое может стать как недостаточным, так и избыточным. Таким образом, наличие/отсутствие регионарных метастазов является «краеугольным камнем» в процессе диагностики. К сожалению, диагностика регионарных метастазов у больных РШМ в настоящее время не достаточно достоверна. Так, по данным разных авторов, в 7,9%–32% случаев [4, 5, 6] на-

личие регионарных метастазов не выявляется при клиническом обследовании, а значит, клиническое стадирование оказывается неверным, и, следовательно, неверным (недостаточным или избыточным) может оказаться и выбранный метод лечения. Поэтому определение метастазов РШМ является проблемой, которая требует комплексного и, по возможности, автоматизированного, решения.

Поиск решения этой проблемы за последнее десятилетие привел к появлению компьютеризированных диагностических инструментов, призванных помочь врачу разобраться в сложных переплетениях данных [7].

Так, на основе данных, полученных в медицинской клинике при университете штата Висконсин (WBCD), опубликовано много исследований по медицинской диагностике рака молочной железы (РМЖ) с использованием различных алгоритмов искусственного интеллекта [8–17], в том числе «Дерево решений» (DT), «Случайный лес» (RF), «k-ближайших соседей» (KNN), «Метод опорных векторов» (SVM), «Adaboost», «XGBoost», нейросети и др. Точность результатов разных алгоритмов обучения, по приведенным данным, составляет от 92,2% до 99,98%. Например, в одном из этих исследований авторы сравнивали метод опорных векторов (SVM) и нейросети с использованием 4 разных наборов данных о РМЖ, раке печени, раке простаты и раке яичников. Исследователи продемонстрировали, что метод опорных векторов (SVM) дал лучшие результаты точности, чем нейросети, по РМЖ (99,51% против 98,54%) и раку печени (63,11% против 57,28%). Нейросети лучше сработали при обнаружении рака простаты (82,47% против 78,35%) и рака яичников (78,95% против 64,47%) [8].

Авторы всех приведенных исследований утверждают, что используемые алгоритмы можно считать подходящими для прогнозирования РМЖ [8–17].

Обзор доступной литературы показал, что способы выявления метастазов при РШМ методами машинного обучения изучены слабо. Так, в одном крупном исследовании авторы изучили 858 больных РШМ по 33 признакам

и результатам биопсии (на наборе данных UCI по РШМ). Классификация была выполнена с использованием следующих методов: «Много-слойный перцептрон», «Наивный Байес» и «к-ближайших соседей» (KNN). Наилучшие результаты показал метод KNN [18].

В другом исследовании авторами использовались следующие алгоритмы МО: «Логистическая регрессия», «Случайный лес» (RF) и «Дерево решений» (DT) на наборе данных UCI по РШМ. Наибольшую точность показала модель, комбинирующая все 3 метода. В результате была достигнута точность 96,73 % [19].

Еще в одном исследовании использовались три типа алгоритмов МО, чтобы классифицировать набор данных UCI (University of California, Irvine, 1987) РШМ: «XGBoost», «Случайный лес» (RF) и «Метод опорных векторов» (SVM). Было обнаружено, что «Случайный лес» (RF) наиболее точно дифференцирует злокачественные и доброкачественные процессы [20].

В российской литературе описывается несколько методов, позволяющих автоматизировать процесс диагностики РШМ [21–24]. Однако, лабораторные параметры, используемые в расчетных формулах, не относятся к рутинным, или требуют специального оборудования, следовательно, не могут быть исследованы в любой лаборатории, а сами способы не позво-

ляют уточнить наличие регионарных метастазов РШМ. Следует также заметить, что ни один из указанных выше методов прогнозирования стадии РШМ не использует при построении модели алгоритма машинного обучения.

Ранее нами предложен способ диагностики метастазов при РШМ на основе алгоритма машинного обучения «Многофакторная Логистическая регрессия» [25]. В данной работе мы решили расширить спектр применения алгоритмов машинного обучения.

Методология исследования

Описание набора данных

Мы получили данные результатов гистологического исследования у больных РШМ после проведения хирургического стадирования в МНИОИ им. П. А. Герцена, филиале ФГБУ НМИЦ радиологии Минздрава России. Для анализа были использованы данные 495 пациентов в возрасте 18–82 лет (42 [34; 51] лет). Каждая из них имела параметр наличия/отсутствия метастазов, и мы строили модель, которая будет автоматически для каждого конкретного случая прогнозировать этот параметр по 7 показателям клинического анализа крови и коагулограммы. Эти показатели периферической крови были выбраны из 21 исходных, как наиболее статистически значимо изменяющиеся в зависимости от наличия/отсутствия регионарных метастазов [6]. Те показатели крови, значения которых статистически значимо не менялись, средний объем эритроцита (норма 80–100), среднее содержание гемоглобина (норма 26,0–34,0), тромбоциты (норма 180–350), АЧТВ (норма 26,00–37,00), протромбиновое время (норма 10,1–13,7), тромбиновое время (норма 14,6–22,0), плазминоген (75,0–140,0 %), активность антитромбина III (%), в построении модели прогнозирования не участвовали.

Пациентки были разделены на две группы по критерию наличия или отсутствия метастазов в лимфатических узлах таза.

Значения объектов представлены в числовом формате. «Цель» означает пациентку, у которой метастазы «есть» (= 1) или «нет» (= 0).

Таблица 1

«Цель»

Тип пациента	Цель
Есть метастазы	1
Нет метастазов	0

Таблица 2

Параметры крови

Код	Название параметра и референсные значения параметра
BF	СОЭ (норма до 30 мм/час)
BH	Эритроциты (норма 4,00–5,50)
BI	Гемоглобин (норма 120–165 г/л)
BS	Фибриноген (норма 2,00–4,00 г/л)
BT	Д-димер ФЭЕ (норма 0–0,55 ФЭЕ мг/л)
BU	Агрегация тромб с АДФ (пределы 80–120 %)
BY	РФМК (норма 0–4,0)

Визуализация данных

Данные визуализированы по двум категориям: 1) есть метастазы 0) нет метастазов.

Визуализация данных показывает, что полученные результаты довольно хорошо подразделяются по выбранным параметрам (показателям крови).

Если изображать параметры крови в одномерном пространстве, то они существенно не

различимы. Для большей наглядности мы сделали двухмерное изображение базы данных параметров крови. На графике изображены точки, каждая из которых обозначает одного пациента (в нашем случае, больную РШМ). График показывает пациентов в семимерном пространстве, позволяя видеть каждую точку (пациента) по отдельности, по 2 показателя крови на каждом графике. Оранжевые точки — больные

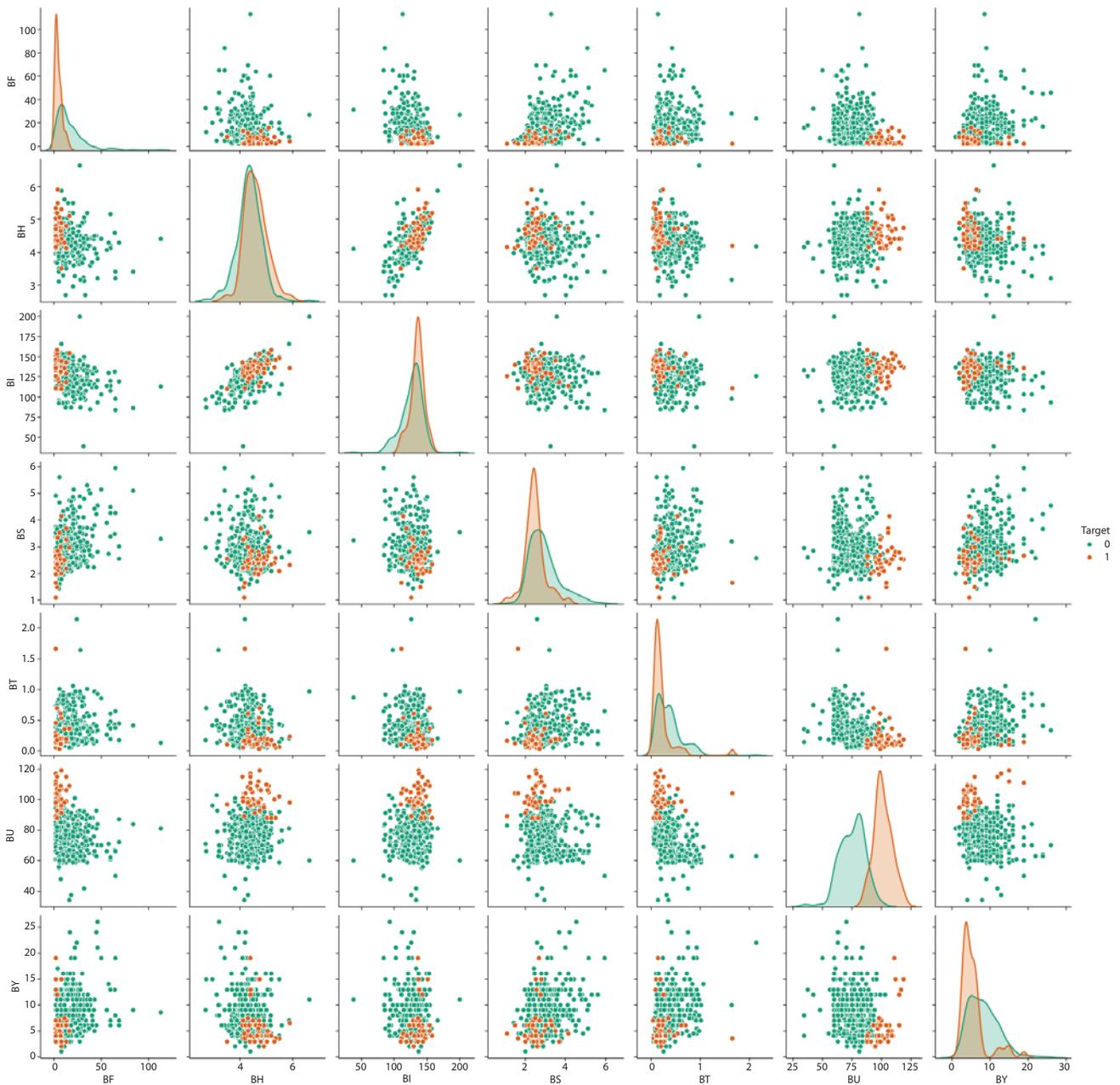


Рис. 1. Парный График Объектов [BF — CO₂, BH — эритроциты, BI — гемоглобин, BS — фибриноген, BT — Д-димер, BU — агрегация тромбоцитов с АДФ, BY — РФМК]. Зеленые объекты — без метастазов, оранжевые — с метастазами

РШМ с гистологически подтвержденными метастазами в регионарных лимфатических узлах таза, зеленые — больные без метастазов.

На большинстве двумерных графиков оранжевые точки (пациенты с метастазами) образуют своеобразные «облака», что заметно отделяет их от «облаков» зеленых точек (пациенты без метастазов). Визуальный диссонанс точек (пациентов с метастазами и без) говорит о том, что и математически точки могут быть разделены с помощью методов машинного обучения.

Результаты построения моделей для прогнозирования

Мы использовали Jupyter Notebook (Anaconda), Python 3.8 в качестве платформы для кодирования и построения моделей для прогнозирования.

Выбор модели — самый важный шаг в машинном обучении. Алгоритмы машинного обучения делятся на следующие: «обучение с учителем» и «обучение без учителя». Для проекта нам потребовались алгоритмы «обучение с учителем», наиболее часто использующиеся при построении моделей прогнозирования в медицине: алгоритмы машинного «обучения с учителем» — методы классификации, такие как «Дерево решений» (DT), «Случайный лес» (RF), «k-ближайших соседей» (KNN), «Метод опорных векторов» (SVM), «Adaboost» и «XGBoost» (табл. 3).

Набор данных содержал объекты, которые довольно сильно различались по величинам и единицам измерения. Поэтому мы привели все характеристики к одному и тому же уровню величин; применив стандартное масштабирование с использованием библиотеки `v-SKLearn` (StandardScaler).

Все выбранные алгоритмы использовались в построении модели для прогнозирования результата и расчета точности прогноза (рис. 2).

Как видно из графика на рисунке 2, стандартное масштабирование данных не приводит к явному улучшению показателей точности, которая либо сохраняется, либо падает. Поэтому был сделан вывод, что от системы стандартного масштабирования данных в данном случае можно отказаться.

Матрица ошибок и точность модели

Матрица ошибок использовалась для оценки производительности модели классификации. Она сравнивает фактические целевые значения с прогнозируемыми, полученными с помощью модели машинного обучения, что показывает, каким образом работает прогнозистическая модель.

График на рисунке 3 демонстрирует, что лучше всего построенные модели функционируют при построении их алгоритмами «Метод опорных векторов», Adaboost и XGBoost. Точность на обучающей выборке у них падает только на 2 %: со 100 % до 98 %.

График на рисунке 4 демонстрирует, что по критерию точности ROC-AUC лучше всего модели работают при построении алгоритмами «Случайный лес» (RF), «Метод опорных векторов» (SVL) и XGBoost. Точность на обучающей выборке у них падает только на 0,1–0,2 %.

График на рисунке 5 демонстрирует, что по доле объясняющей дисперсии лучше всего работает модель, построенная с помощью алгоритма «Метод опорных векторов» (RF). Следующей наиболее пригодной для клинической практики следует считать модель, постро-

Таблица 3

Сравнение между методами

Методы	Точность без стандартной шкалы	Точность со стандартной шкалой
Дерево решений (DT)	95,3 %	95,3 %
Случайный лес (RF)	96,6 %	96,0 %
Метод К ближайших соседей (KNN)	98,7 %	96,6 %
Метод опорных векторов (SVM)	98,0 %	98,0 %
Adaboost	98,0 %	98,0 %
XGboost	98,0 %	98,0 %

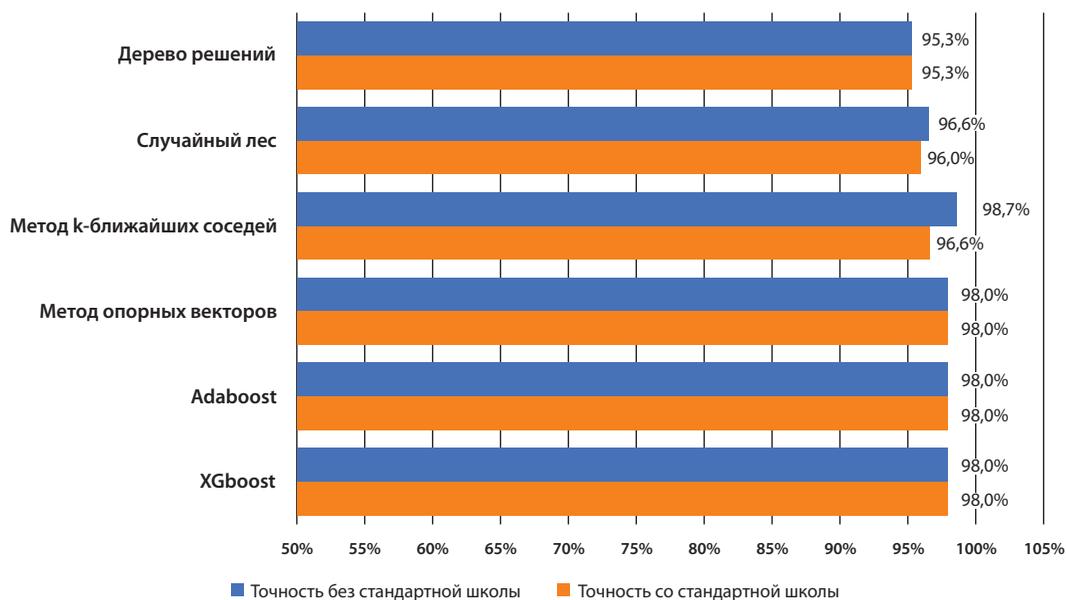


Рис. 2. Сравнение точности на тестовых данных

енную алгоритмом XGBoost. Максимальная объясняющая доля дисперсии на тестовой выборке в этом случае составляет 82,9 % и 80,5 %, соответственно.

Предлагаемая архитектура системы

Из представленной диаграммы следует, что мы сначала загрузили набор данных о наличии/отсутствии метастазов из представленных историй болезни. Затем была проведена предварительная обработка этих данных и применена

модель машинного обучения для прогнозирования метастазирования РШМ (рис. 6).

Применение построенной модели прогнозирования (SVM) на клинических примерах

Пример № 1.

Пациентка А., 31 год, обратилась в поликлинику МНИОИ им. П. А. Герцена, филиала ФГБУ НМИЦ радиологии Минздрава России,

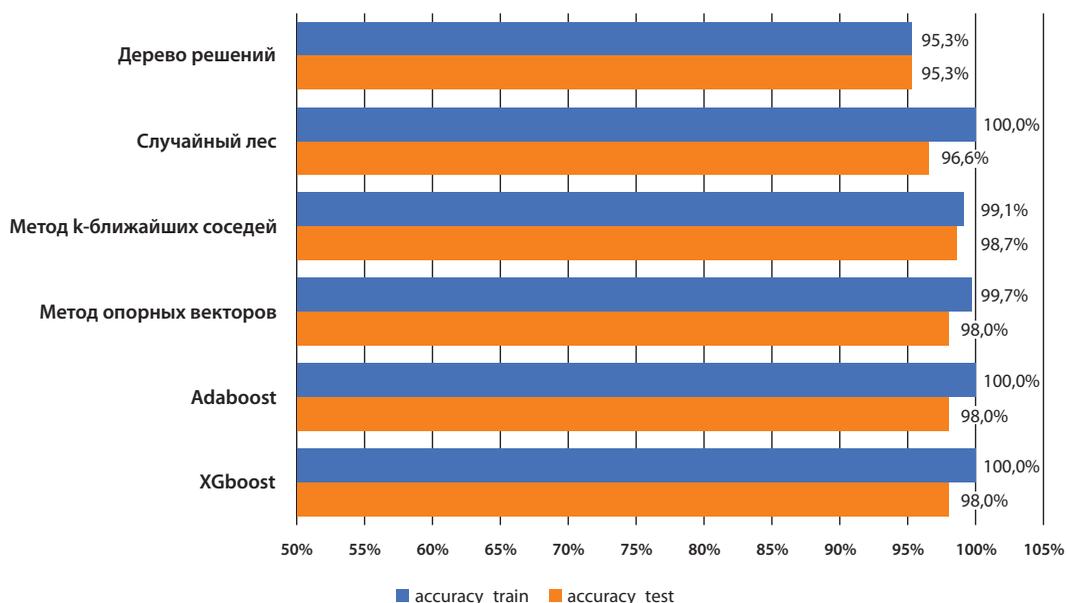


Рис. 3. Сравнение точности (Accuracy) на обучающей и тестовой выборках

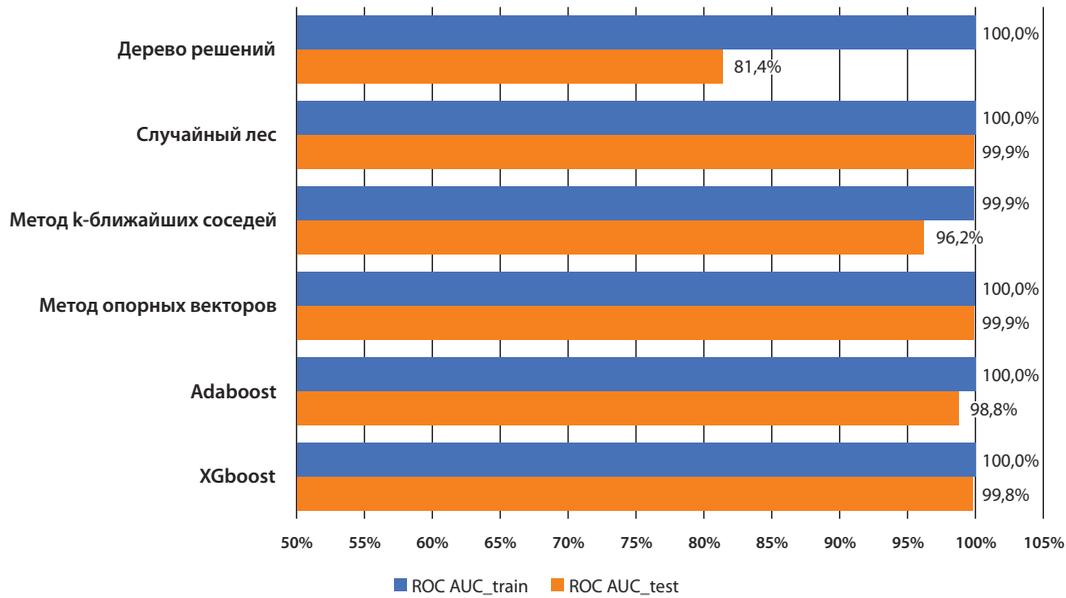


Рис. 4. Сравнение точности (ROC-AUC) на обучающей и тестовых выборках

с диагнозом «Дисплазия шейки матки 3 степени» (CIN III, HSIL) цитологически. При осмотре заподозрен РШМ.

Влагалищный статус при первичном осмотре. Наружные половые органы сформированы правильно. Наружное отверстие уретры без патологии. Слизистая влагалища не изменена. Шейка матки с выраженным сосудистым рисунком (предположительно, инфильтративные опухолевые изменения), общим размером около 3 см, контактно кровоточит. Тело матки

обычной формы и размеров, плотное, подвижное, безболезненное. Придатки справа не определяются, область их безболезненна. Придатки слева не определяются, область их безболезненна. Сводья справа без особенностей, слева в своде незначительная тяжесть. Макроскопически перехода опухоли на влагалище не определяется. Маточно-прямокишечное углубление свободно. Per gestum: слизистая прямой кишки на высоте пальца гладкая, подвижная.

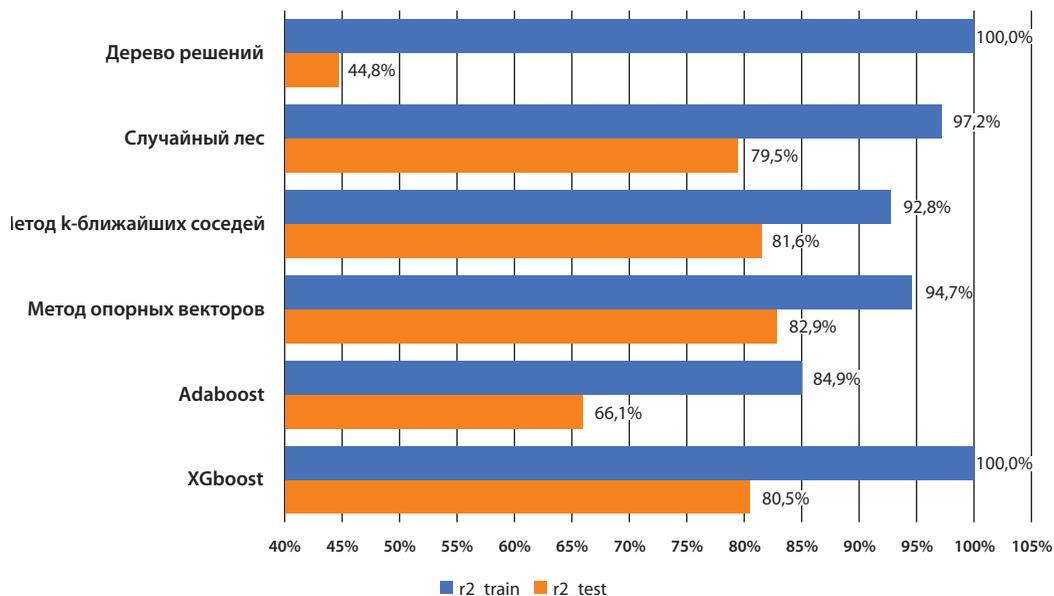


Рис. 5. Сравнение точности (R^2) на обучающей и тестовой выборках

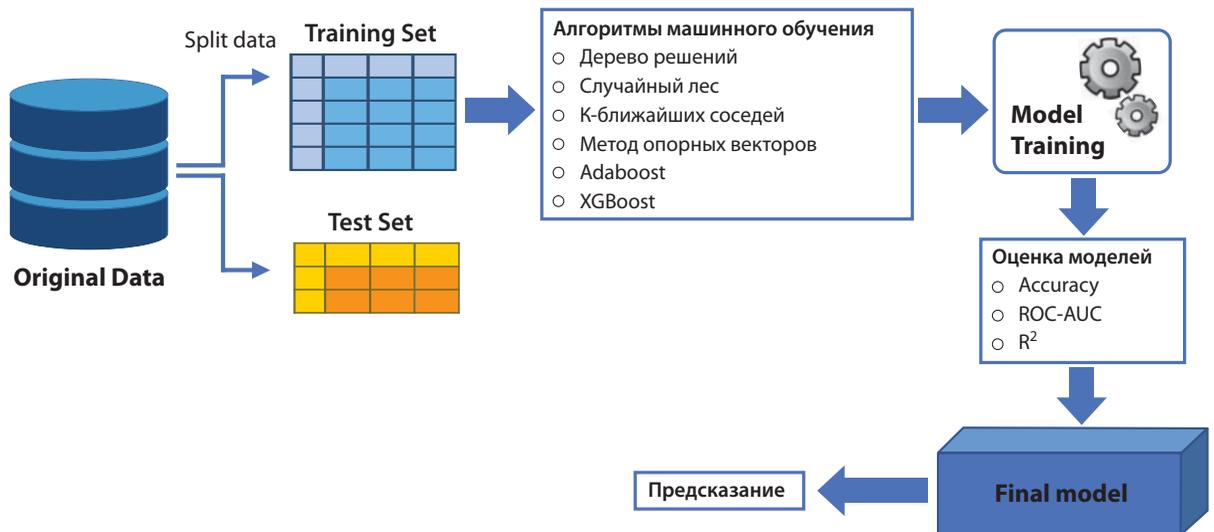


Рис. 6. Архитектура системы

Повторное цитологическое исследование.
 Заключение: цитологическая картина соответствует плоскоклеточному раку с наклонностью к орогованию.

Данные инструментального обследования

УЗИ малого таза и брюшной полости.
 Заключение: эхо-картина полипов эндометрия. Зона неоднородности в шейке матки неясного генеза. При исследовании парааортальной области, по ходу общих подвздошных сосудов измененные лимфоузлы не определяются. Необходимо дообследование малого таза другими методами /МРТ/.

Магнитно-резонансная томография органов малого таза с контрастированием.
 Заключение: МР — картина опухоли шейки матки с распространением на верхнюю треть влагалища, признаками выхода в параметрий, внутритазовые лимфоузлы не изменены, эндометриоидная киста правого яичника.

Гистологическое заключение: (после проведенной диагностической РХ биопсии шейки матки, выскабливания цервикального канала): фрагменты ткани шейки матки с ростом инвазивной опухоли, представленной солидными структурами из атипичных эпителиальных клеток с плоскоклеточной дифференцировкой, выраженным ядерным полиморфизмом, наличием фигур митозов, без кератинизации, с выраженной лимфоцитарно-плазмоцитарной

инфильтрацией окружающей стромы. Глубина инвазии опухоли в пределах исследованного материала не менее 2,5 мм. Среди слизи и форменных элементов крови в соскобе обнаружены фрагменты вышеописанной опухоли.

По результатам обследования установлен диагноз: рак шейки матки IB1 стадии (T1b1N0M0).

В таблице 4 представлены показатели клинического анализа крови и коагулограммы пациентки А. при первичном обращении.

По этим показателям был произведен расчет «Наличия метастазов» (N) по построенной модели прогнозирования XGBoost.

Модель предсказала значение = 1, следовательно, у данной пациентки имеются метастазы в лимфатических узлах таза.

После расчета уточнен клинический диагноз: рак шейки матки IIC1 стадии (T1b1N1M0).

Предсказанный диагноз был подтвержден результатами морфологического исследования после стадирующей лапароскопической лимфодиссекции (планирование органосохраняющего лечения).

Результаты гистологического исследования послеоперационного материала. Клетчатка с лимфатическими узлами справа — фрагменты двадцати семи лимфоузлов, из них в одном — метастаз плоскоклеточного рака. Клетчатка с лимфатическими узлами слева — девятнадцать фрагментов лимфоузлов, в пяти

Показатели клинического анализа крови и коагулограммы пациентки А.

Показатель	Сокращение	Результат	Референсный интервал	Ед. измерения
D-димер ФЭЕ		0,31	0,00–0,55	ФЭЕ мг/л
Эритроциты	RBC	3,81	3,90–4,70	
Фибриноген по Клауссу		2,87	2,00–4,00	г/л
Агрегация тромбоцитов с АДФ		53	69–88	%
СОЭ по Вестергрену	СОЭ	13	2–15	мм/ч
Гемоглобин	HGB	110,8	120,0–140,0	г/л
РФМК		6	0–4	

из них метастазы плоскоклеточного рака со слабо выраженным ороговением.

Заключение. В одном лимфоузле клетчатки справа из двадцати семи и пяти лимфоузлах клетчатки слева из девятнадцати метастазы плоскоклеточного рака со слабо выраженным ороговением.

Таким образом, наличие метастазов в лимфатических узлах малого таза у данной пациентки было предсказано точно. Иными методами метастазы в лимфатических узлах малого таза обнаружены не были.

Пример № 2.

Пациентка К., 30 лет, направлена в поликлинику МНИОИ им. П. А. Герцена, филиал ФГБУ НМИЦ радиологии Минздрава России, с предположительным диагнозом РШМ (цитологически).

Влагалищный статус при первичном обращении. Наружные половые органы сформированы правильно. Наружное отверстие уретры без патологии. Слизистая влагалища не изменена. Шейка матки: видимая часть экзоцервикса с переходом на заднюю и правую боковую стенки — неравномерно окрашенная очаговая гиперемия с неровным наружным контуром, переходящая в цервикальный канал (предположительно, инфильтративная опухоль). Общий размер поражения более 4 см². Тело матки обычной формы и размеров, плотное, подвижное, безболезненное. Придатки справа не определяются, область их безболезненна. Придатки слева не определяются, область их безболезненна. Своды свободны. Инфильтраты в пара-

метриях не определяются. Дугласов карман свободен.

Per rectum: слизистая прямой кишки на высоте пальца гладкая, подвижная.

Данные инструментального обследования: протокол УЗИ брюшной полости, забрюшинного пространства, малого таза (ТВУЗИ). Заключение: эхографическая картина опухоли шейки матки без признаков поражения лимфатических узлов.

Магнитно-резонансная томография органов малого таза с контрастным усилением. Заключение: МР картина объемного образования шейки матки без убедительных признаков распространения на параметральную клетчатку, без поражения регионарных лимфатических узлов таза.

В таблице 5 представлены показатели клинического анализа крови и коагулограммы пациентки К. при первичном обращении.

По этим показателям был произведен расчет «Наличия метастазов» (N) по построенной модели прогнозирования XGBoost.

Модель предсказала значение = 0, следовательно, у данной пациентки регионарных метастазов нет.

После расчета предположительный клинический диагноз рак шейки матки IV стадии (T1b2N0M0).

Диагноз подтвержден результатами гистологического исследования после проведенного органосохраняющего лечения (влагалищной трахелэктомии с лапароскопической подвздошно-обтураторной лимфаденэктомией).

Показатели клинического анализа крови и коагулограммы пациентки К.

Показатель	Сокращение	Результат	Референсный интервал	Ед. измерения
D-димер ФЭЕ		0,25	0,00–0,55	ФЭЕ мг/л
Эритроциты	RBC	4,42	3,90–4,70	
Фибриноген по Клауссу		1,79	2,00–4,00	г/л
Агрегация тромбоцитов с АДФ		100	69–88	%
СОЭ по Вестергрену	СОЭ	6	2–15	мм/ч
Гемоглобин	HGB	131	120,0–140,0	г/л
РФМК		7,5	0–4	

Гистологическое заключение: 1. Клетчатка справа в лимфоузлах без метастазов карциномы; 2. Клетчатка слева — в лимфоузлах без метастазов карциномы; 3. Узел имеет строение простой лейомиомы; 4. Шейка матки: Край резекции — интактный.

В толще стенки шейки выявлен рост инвазивной плоскоклеточной ороговевающей карциномы, G1, с фокусами ороговения, с распространением по ходу цервикального канала вверх на 1,9 см, и инвазией в толщу стенки на 1,3 см более 1/2 толщины шейки. Опухолевые клетки умеренно-полиморфные, с умеренной митотической активностью. Выявлены единичные фокусы сосудистой инвазии. Обнаружены более плотноклеточные зоны без ороговения. В пласте плоского эпителия вблизи опухоли признаки *in situ* и HSIL.

5 ц/к эндо- и экзоцервикс. Кусочки плоского эпителия без подлежащей ткани, без изменений. Частицы цервикальной слизи. Мелкие фрагменты цервикальных желез.

Заключение C53.1. Инвазивная плоскоклеточная ороговевающая карцинома шейки матки, G1, с распространением по ходу цервикального канала вверх на 1,9 см и инвазией в толщу стенки на 1,3 см (более 1/2 толщины шейки). Простая лейомиома. В представленных срезах лимфатических узлов таза метастазов рака нет, гиперплазия лимфоидной ткани, гистиоцитоз синусов.

Таким образом, отсутствие метастазов в лимфатических узлах малого таза у данной

пациентки было предсказано по формуле точно, что подтверждено гистологическим исследованием послеоперационного материала.

Заключение и перспективные направления исследования

В статье нами были рассмотрены различные алгоритмы машинного обучения для выявления метастазов РШМ. Целью нашего исследования был анализ клинических данных больных РШМ, обследованных и получивших лечение в МНИОИ им. П. А. Герцена, филиале ФГБУ НМИЦ радиологии Минздрава России, путем их визуализации и оценки прогнозов моделей машинного обучения. С помощью этой исследовательской работы мы смогли сравнить работу классификаторов «Дерево решений» (DT), «Случайный лес» (RF), «k-ближайших соседей» (KNN), Метод опорных векторов (SVM), Adaboost, и XGBoost. Мы пришли к выводу, что Метод опорных векторов (SVM) и XGBoost являются наиболее точными алгоритмами для получения наилучшего точного результата для выявления метастазов РШМ с точностью 98 %. В будущем мы бы хотели добавить набор данных для валидации построенных моделей и проверить их эффективность и масштабируемость. Чувствительность и специфичность в данном исследовании были заменены критерием ROC-AUC, который является комбинацией чувствительности и специфичности.

ЛИТЕРАТУРА

1. Torre L. A., Bray F., Siegel R. L., Ferlay J., Lortet-Tieulent J., Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015 Mar;65(2):87–108. DOI: 10.3322/caac.21262. Epub 2015 Feb 4. PMID: 25651787, <https://gco.iarc.fr/today/data/factsheets/cancers/23-Cervix-uteri-fact-sheet.pdf>
2. Каприн А. Д., Старинский В. В., Шахзадова А. О. Злокачественные новообразования в России в 2019 году (заболеваемость и смертность). М.: МНИОИ им. П. А. Герцена — филиал ФГБУ «НМИЦ радиологии» Минздрава России, 2020. 251 с.
3. Каприн А. Д., Старинский В. В., Шахзадова А. О. Состояние онкологической помощи населению в России в 2019 году. М.: МНИОИ им. П. А. Герцена — филиал ФГБУ «НМИЦ радиологии» Минздрава России, 2020. 238 с.
4. Dostalek L., Slama J., Fisherova D. et al. Impact of sentinel lymph node frozen section evaluation to avoid combined treatment in early-stage cervical cancer. *Int J Gynecol Cancer.* 2020; 30(6):744–748. DOI: 10.1136/ijgc-2019-001113
5. Tsunoda A. T., Marnitz S., Soares Nunes J. et al. Incidence of histologically proven pelvic and para-aortic lymph node metastases and rate of upstaging in patients with locally advanced cervical cancer: results of a prospective randomized trial oncology. 2017;92(4):213–220. DOI: 10.1159/000453666
6. Сушинская Т. В., Епифанова С. В., Щепкина Е. В., Кузнецов А. И., Стуков Н. И. Ошибки диагностики предраковых заболеваний и рака шейки матки // Онкогинекология. 2020. № 1.
7. Pena-Reyes C. A., Sipper M. A fuzzy-genetic approach to breast cancer diagnosis, *Artificial Intelligence in Medicine*, 131–155, 1999.
8. Albrecht A. A., Lappas G., Vinterbo S. A., Wong C. K., Ohno-Machado L. Two applications of the LSA machine, *ICONIP* 2002.
9. Pena-Reyes C. A., Sipper M. Fuzzy CoCo: A cooperative coevolutionary approach to fuzzy modeling, *IEEE Transactions on Fuzzy Systems*, 9(5):727–737, 2001.
10. Setiono R. Generating concise and accurate classification rules for breast cancer diagnosis, *Artificial Intelligence in Medicine*, 18(3): 205–217, 2000.
11. Tüba Kiyandı Tülay Yildirim (2004), Breast cancer diagnosis using statistical neural networks, *Journal of electrical & electronics engineering*, vol.4, pp.1149–1153.
12. Sivapriya J., Aravind Kumar V., Siddarth Sai S., Sriram S. Breast Cancer Prediction using Machine Learning, ISSN: 2277–3878, Volume-8 Issue-4, November 2019.
13. Manav Mangukiya, Anuj Vaghani, Meet Savani. Breast cancer detection using machine learning algorithms, *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, ISSN: 2321–9653; Volume 10 Issue II, February 2022, DOI: 10.22214/ijraset.2022.40204
14. Hafizah S., Ahmad S., Sallehuddin R., and Azizah N. Cancer Detection Using Artificial Neural Network and Support Vector Machine: A Comparative Study. *J. Teknol*, vol. 65, pp. 73–81, 2013.
15. Divyavani M., Dr. Kalpana G. An Analysis on SVM & ANN using Breast Cancer Dataset, *Aegaeum Journal*, SSN NO: 0776–3808, January 2021
16. Keles, M. Kaya. Breast Cancer Prediction and Detection Using Data Mining Classification Algorithms: A Comparative Study. *Tehnicky Vjesnik — Technical Gazette*, vol. 26, no. 1, 2019, p. 149.
17. Gc S., Kasaudhan R., Heo T. K., and Choi H. D. Variability Measurement for Breast Cancer Classification Mammographic adaptive and convergent systems (RACS), Prague, Czech Republic, 2015, pp. 177–182.
18. Muhammed Fahri Unlersen, Kadir Sabanci, Muciz Özcan, Determining Cervical Cancer Possibility by Using Machine Learning Methods, ISSN: 2454–5031, *International Journal of Latest Research in Engineering and Technology (IJLRET)*, Volume 03 — Issue 12. December 2017. PP. 65–71.
19. Riham Alsmariy, Graham Healy, Hoda Abdelhafez. Predicting Cervical Cancer using Machine Learning Methods, *International Journal of Advanced Computer Science and Applications*, Vol. 11, No. 7, 2020.
20. Deng X., Luo Y., and Wang C. Analysis of Risk Factors for Cervical Cancer Based on Machine Learning Methods. *Proceedings of 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems, CCIS 2018*. pp. 631–635, 2019.
21. Патент № 2728675 С1 Российская Федерация, МПК G01N 33/50. Способ диагностики рака шейки матки: № 2019133539: заявл. 21.10.2019: опубл. 31.07.2020 / Е. В. Каюкова, П. П. Терешков; заявитель Федеральное государственное бюджетное образовательное учреждение высшего образования Читинская государственная медицинская академия Министерства здравоохранения Российской Федерации.

22. Патент № 2645111 С2 Российская Федерация, МПК G01N 33/48. Способ стадирования рака шейки матки: № 2016126853: заявл. 04.07.2016: опубл. 15.02.2018 / Т. П. Генинг, Д. Р. Долгова, Т. В. Абакумова [и др.]; заявитель федеральное государственное бюджетное образовательное учреждение высшего образования «Ульяновский государственный университет».

23. Патент № 2343485 С1 Российская Федерация, МПК G01N 33/53. Способ диагностики рака шейки матки: № 2007123039/15: заявл. 14.06.2007: опубл. 10.01.2009 / А. А. Булгаков, О. М. Родионова, В. И. Апанасевич [и др.]; заявитель Государственное образовательное учреждение высшего профессионального образования «Владивостокский государственный медицинский университет Федерального агентства по здравоохранению и социальному развитию» (ГОУ ВПО ВГМУ Росздрава), Тихоокеанский институт биоорганической химии Дальневосточного отделения Российской академии наук (ТИБОХ ДВО РАН), Институт биологии моря Дальневосточного отделения Российской академии наук (ИБМ ДВО РАН).

24. Патент № 2582979 С1 Российская Федерация, МПК G01N 33/48. Способ уточнения стадии распространенного рака шейки матки: № 2015109694/15: заявл. 19.03.2015: опубл. 27.04.2016 / Т. П. Генинг, Т. В. Абакумова, Д. Р. Долгова [и др.]; заявитель федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Ульяновский государственный университет».

25. Патент № 2750769 С2 Российская Федерация, МПК G01N 33/49. Способ диагностики метастазов в лимфатические узлы малого таза у больных резектабельным раком шейки матки: № 2020142386: заявл. 22.12.2020: опубл. 02.07.2021 / А. Д. Каприн, Н. И. Стуков, Т. В. Сушинская [и др.]; заявитель Федеральное государственное бюджетное учреждение «Национальный медицинский исследовательский центр радиологии» Министерства здравоохранения Российской Федерации.

АВТОРЫ

Сушинская Татьяна Валентиновна, кандидат медицинских наук, старший научный сотрудник отдела опухолей репродуктивных и мочевыводящих органов МНИОИ им. П. А. Герцена, филиала ФГБУ «НМИЦ радиологии» Минздрава России, 125284, Россия, Москва, 2-й Боткинский проезд, д. 3, ORCID iD 0000-0001-8812-9105.

Sushinskaya Tatyana V., C.Sc. (Med.), senior researcher, P. Hertsen Moscow Oncology Research Institute — Branch of the National Medical Research Radiological Centre, 3, 2nd Botkinskiy lane, Moscow, 125284, Russian Federation; ORCID iD 0000-0001-8812-9105.

Кузнецов Антон Игоревич, студент факультета «Системы управления, информатика и электроэнергетика» МАИ, 125993, Россия, Москва, Волоколамское шоссе, д. 4; ORCID iD 0000-0003-2182-5792.

Kuznetsov Anton I., student of the Faculty «Control Systems, Informatics, & Electricity», Moscow Aviation Institute (National Research University), 4, Volokolamskoe road, Moscow, 125993, Russian Federation; ORCID iD 0000-0003-2182-5792.

Щепкина Елена Викторовна, кандидат соц. наук, доцент кафедры № 256 — медицинских информационных систем, Институт информационных технологий (факультет), РТУ-МИРЭА. Россия, Москва, проспект Вернадского, 78, стр. 4. Директор по статистике, ООО «Техдепартамент», 141407, Россия, г. Химки, ул. Германа Титова, д. 2, ORCID iD 0000-0002-2079-1482.

Schepkina Elena V., C. Sc. (Soc.), docent, Department of Medical Information Systems, Russian Technological University — MIREA. 78, Vernadsky Avenue., Moscow, 119454, Russian Federation. Director for Statistics, LLC «Techdepartament»; 2, German Titov str., Khimki, 141407, Russian Federation; ORCID iD 0000-0002-2079-1482.

Стуков Николай Игоревич, доктор медицинских наук, профессор кафедры госпитальной терапии с курсами эндокринологии, гематологии и клинической лабораторной диагностики Медицинского института РУДН, 117198, Москва, ул. Миклухо-Маклая, д. 8; руководитель курса гематологии, главный научный сотрудник отделения высокодозной химиотерапии с блоком трансплантации костного мозга МНИОИ им. П. А. Герцена — филиала ФГБУ «НМИЦ радиологии» Минздрава России, 125284, Россия, Москва, 2-й Боткинский проезд, д. 3, ORCID iD 0000-0002-4546-1578.

Stuklov Nikolay I., Dr. Sc. (Med.), professor of the Department of Hospital Therapy with the Courses of Endocrinology, Hematology, & Clinical Laboratory Diagnostics, Medical Institute of the Russian University of Peoples' Friendship; 8, Miklukho-Maklaya str., Moscow, 117198, Russian Federation; Head of the Course of Hematology; principal researcher, P. Hertsen Moscow Oncology Research Institute — Branch of the National Medical Research Radiological Centre; 3, 2nd Botkinskiy pass., Moscow, 125284, Russian Federation; ORCID iD 0000-0002-4546-1578.

Епифанова Светлана Викторовна, кандидат медицинских наук, врач-рентгенолог отделения рентгеновской диагностики и томографии ФГБУ «ЦКБ с поликлиникой УДП РФ», 121359, Россия, Москва, ул. Маршала Тимошенко, д. 15, ORCID iD 0000-0002-7591-5120.

Epifanova Svetlana V., C. Sc. (Med.), radiologist of the Department of X-ray Diagnostics & Tomography, Central Clinical Hospital and Polyclinic of the Department for Presidential Affairs, 15, Marshal Timoshenko str., Moscow, 121359, Russian Federation; ORCID iD 0000-0002-7591-5120.



**НМИЦ
ОНКОЛОГИИ**
им. Н. Н. Блохина

Федеральное государственное бюджетное учреждение
«Национальный медицинский исследовательский
центр онкологии им. Н. Н. Блохина»
Министерства здравоохранения РФ

5–7 СЕНТЯБРЯ СОСТОИТСЯ ФОРУМ «ИННОВАЦИОННАЯ ОНКОЛОГИЯ»

Коллеги, приглашаем вас принять участие в крупнейшем научно-практическом мероприятии для специалистов в сфере медицины и здравоохранения — Международном форуме «Инновационная онкология».

Форум пройдет 5-7 сентября 2022 года в НМИЦ онкологии им. Н.Н. Блохина при поддержке и непосредственном участии профессиональных сообществ онкологов, а также экспертов международного класса, зарубежных коллег.

Впервые мероприятие такого масштаба пройдет в live-формате — работа крупнейшего в России и мире онкологического центра будет синхронизирована с программой форума, чтобы каждый его участник получил доступ к современным технологиям оказания онкологической помощи на практике, осуществляемой ведущими экспертами профессиональных сообществ, специалистами мирового уровня.

Участие в Международном форуме «Инновационная онкология» станет серьезным вкладом в профессиональный рост российских специалистов, будет способствовать развитию междисциплинарных, межцентровых, межрегиональных и международных связей в онкологии.

Ознакомиться с программой форума «Инновационная онкология»-2022 и зарегистрироваться для участия можно по ссылке: <https://inno-onco.ru/>

Лучшие эксперты обсудят методы профилактики рака, достижения хирургии, новейшие методы лекарственного лечения, лучевой терапии и диагностики, открытия фундаментальной онкологии.

Приглашаем вас на форум «Инновационная онкология» в НМИЦ онкологии им. Н.Н. Блохина!

Координатор проекта:

Фролова Оксана

E-mail: af@blokhinfond.ru

Сайт форума: <https://inno-onco.ru/>