

«Murojaatlar» portali jismoniy va yuridik shaxslardan tashkilotga kelayotgan murojaatlar bilan ishlashda xodimlarning ishini avtomatlashtirishga, shuningdek murojaat bo'yicha qabul qilingan qarorlar, rezolyutsiyalar ustidan nazoratni ta'minlashga imkon beradi.

Adabiyotlar ro'yxati

1.O'zbekiston Respublikasining "Elektron hukumat to'g'risida" 2015-yil 9-dekabrda O'RQ-395-son Qonuni.

2.O'zbekiston Respublikasining "Jismoniy va yuridik shaxslarning murojaatlari to'g'risida" 2017-yil 11-sentabrda O'RQ-445-son Qonuni.

3.O'zbekiston Respublikasi Prezidentining «Jismoniy va yuridik shaxslarning murojaatlari bilan ishlash tizimini tubdan takomillashtirishga doir chora-tadbirlar to'g'risida» 2016-yil 28-dekabrda PF-4904-son Farmoni.

4.O'zbekiston Respublikasi Oliy Majlisi Senatining "Aholi murojaatlari bilan ishlash bo'yicha parlament nazoratini kuchaytirish to'g'risida" 2021-yil 31-iyulda SQ-359-IV-son qarori.

5.O'zbekiston Respublikasi Oliy majlisi qonunchilik palatasi kengashining "O'zbekiston Respublikasi Oliy Majlisi Qonunchilik palatasida jismoniy va yuridik shaxslarning murojaatlari bilan ishlash tartibi to'g'risidagi nizom tasdiqlash haqida" 2021-yil 11-oktabrda 1386-IV-son qarori.

6.O'zbekiston Respublikasi Vazirlar Mahkamasining "Fuqarolarning o'zini o'zi boshqarish organlarida jismoniy va yuridik shaxslarning murojaatlari bilan ishlash tartibi to'g'risida namunaviy nizomni tasdiqlash haqida" 2018-yil 5-yanvarda 7-son qarori.

7.O'zbekiston Respublikasi Vazirlar Mahkamasining "Davlat organlarida, davlat muassasalarida va davlat ishtirokidagi tashkilotlarda jismoniy va yuridik shaxslarning murojaatlari bilan ishlash tartibi to'g'risidagi namunaviy nizomni tasdiqlash haqida" 2018-yil 7-mayda 341-son qarori.

8.O'zbekiston Respublikasi Vazirlar Mahkamasining 2016-yil 3-iyunda 188-son qarori bilan tasdiqlangan Internet tarmog'ida O'zbekiston Respublikasining Hukumat portali to'g'risida nizom talablari hisobga olingan holda davlat organlarida, davlat muassasalarida va davlat ishtirokidagi tashkilotlarda jismoniy va yuridik shaxslarning murojaatlari bilan ishlash tartibi to'g'risidagi namunaviy Nizomi.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧАХ ОНКОПРОФИЛАКТИКИ И ДИАГНОСТИКИ

Хидирова Чарос Муродиллоевна

Доцент Ташкентского информационного технологического университета

Абдуллахужаев Саидисломбек Нематиллаевич

Магистр Ташкентского информационного технологического университета

saidislomabdullakhodjaev@gmail.com

Аннотация: В статье представлен сравнительный анализ алгоритмов машинного обучения, применяемых для ранней диагностики онкологических заболеваний. В работе рассмотрены алгоритмы Random Forest, XGBoost, AdaBoost и другие, протестированные на различных клинических задачах, таких как рак шейки матки, лёгких и кожи. Особое внимание уделено ансамблевым методам, которые продемонстрировали наивысшую точность, особенно алгоритму Random Forest (случайный лес). Подчёркивается

универсальность изучаемых методов, их адаптивность к разнородным медицинским данным и потенциал для разработки интеллектуальных систем поддержки принятия клинических решений.

Ключевые слова: машинное обучение, диагностика, прогнозирование риска рака, Random Forest, ансамблевые методы, клинические задачи, ранняя диагностика, онкологические заболевания, рак шейки матки.

COMPARATIVE ANALYSIS OF MACHINE LEARNING METHODS IN ONCOPREVENTION AND DIAGNOSIS

Annotation: The paper presents a comparative analysis of machine learning algorithms applied for the early diagnosis of oncological diseases. The paper examines algorithms such as Random Forest, XGBoost, AdaBoost, and others, tested on various clinical tasks including cervical, lung, and skin cancer. Special attention is given to ensemble methods, which demonstrated the highest accuracy, particularly the Random Forest algorithm. The study emphasizes the versatility of the methods, their adaptability to heterogeneous medical data, and their potential for developing intelligent clinical decision support systems.

Keywords: machine learning, diagnosis, cancer risk prediction, Random Forest, ensemble methods, clinical tasks, early diagnosis, oncological diseases, cervical cancer.

ONKOLOGIK KASALLIKLARNI OLDINI OLISH VA DIAGNOSTIKA VAZIFALARIDA MASHINA O'QITISH USULLARINING QIYOSIY TAHLILI

Annotatsiya: Maqolada onkologik kasalliklarning erta tashxisida qo'llaniladigan mashina o'qitish algoritmlari solishtirma tahlili taqdim etilgan. Maqolada Random Forest, XGBoost, AdaBoost va boshqa algoritmlar turli klinik vazifalarda, masalan, bachadon bo'yi saratoni, o'pka saratoni va teri saratoni kabi kasalliklarda sinovdan o'tkazilgan. Ansambl usullari, ayniqsa Random Forest (tasodifiy o'rmon) algoritmi, eng yuqori aniqlik ko'rsatkichlarini namoyon etgani sababli, alohida e'tibor markazida bo'lib, ushbu usullarning universalligi va turli xil tibbiy ma'lumotlarga moslashuvchanligi aniqlangan. Shuningdek, maqolada intellektual tizimlar yordamida shifokor qarorlarini qo'llab-quvvatlash imkoniyatlari yoritilgan.

Kalit so'zlar: mashina o'qitish, tashxislash, saraton xavfini bashorat qilish, tasodifiy o'rmon, ansambl usullari, klinik vazifalar, erta tashxislash, onkologik kasalliklar, bachadon bo'yni saratoni.

Развитие цифровых технологий и рост объёмов клинических данных открывают широкие возможности для применения методов машинного обучения в медицинской практике. Особенно перспективным считается использование алгоритмов машинного обучения для ранней диагностики и прогноза онкологических заболеваний, где точность и своевременность выявления играют решающую роль.

Сегодня активно применяются как классические алгоритмы (Logistic Regression, Decision Tree, SVM), так и современные ансамблевые и нейросетевые подходы (Random Forest, XGBoost, CatBoost, Deep Learning). Однако эффективность каждого из них зависит от типа данных, цели задачи, структуры признаков и доступного объёма обучающей выборки.

Целью данной статьи является сравнительный анализ алгоритмов машинного обучения, применяемых в задачах онкопрофилактики и диагностики, с оценкой их точности, универсальности и практической применимости. Анализ основан на реальных

клинических сценариях, включая рак шейки матки, рак лёгких, меланому, рак желудка и другие формы злокачественных новообразований.

Научная новизна работы заключается в систематизации и сопоставлении результатов более 40 моделей, протестированных в различных исследованиях. Это позволило выявить наиболее надёжные и точные алгоритмы, среди которых алгоритмы Random Forest и ансамблевые методы стабильно демонстрируют высокую производительность и адаптивность к различным условиям.

Johanna backman и другие [1] демонстрируют эффективность применения алгоритмов машинного обучения для раннего выявления рака шейки матки на основе электронных медицинских записей из Швеции. Исследование показало, что наиболее информативными признаками стали клинические сущности, извлечённые из текстовых заметок врачей с помощью lstm и per. Среди протестированных моделей наилучшие результаты достиг random forest, показав auc до 0.97 за день до диагноза. Также он продемонстрировал высокую точность и специфичность, что делает его перспективным для использования в клинической практике.

Иван Петров и другие [2] описывают создание информационной системы для прогноза и профилактики заболеваний с использованием нейросетевой модели Multilayer Perceptron, обученной на массиве из 600 000 медицинских записей. Модель достигла точности 99,17%, показав высокую применимость в клинической практике. Для выявления скрытых закономерностей применялся метод кластеризации K-means, эффективно сегментируя заболевания по полу и симптомам. Несмотря на тестирование альтернативных алгоритмов, таких как Decision Trees и SVM, они уступили по точности нейросетевому подходу. Разработанный интерфейс системы позволяет получать персонализированные прогнозы на основе введённых данных.

James DeJohn и другие [3] анализируют потенциал применения методов машинного обучения в сочетании с ультразвуковыми исследованиями для диагностики онкологии головы и шеи. Обзор охватил 34 публикации, где основное внимание уделялось задачам диагностики, реже – оценке лечения и прогнозу токсичности. Наиболее эффективными оказались модели на базе SVM, достигая точности до 99% AUC. Комбинирование клинических и радиомических данных позволило превзойти существующие рекомендации на 10–15%. Однако большинство работ носили ретроспективный характер, что указывает на потребность в дальнейших проспективных исследованиях.

Maria J. Fernandez и другие [4] исследуют применение различных алгоритмов машинного обучения для прогноза рака шейки матки на данных из открытого набора UCI. В выборке использовались 32 признака, включая демографические и клинические факторы риска. Среди шести протестированных моделей наивысшую точность (100%) показали Random Forest, Decision Tree и Adaptive Boosting. SVM также продемонстрировал высокие результаты, в то время как K-Nearest Neighbors оказался наименее надёжным из-за низкой полноты. Авторы подчёркивают эффективность ансамблевых методов для ранней диагностики заболевания.

Xiaoqing Li и другие [5] рассматривают построение моделей для прогнозирования риска рака лёгких в китайской популяции с учётом как эпидемиологических факторов, так и генетических маркёров. В исследовании сравнивались алгоритмы XGBoost и Logistic Regression. Хотя Logistic Regression показал наилучший результат (AUC = 83.3%) при прогнозе плоскоклеточного рака, в остальных сценариях более стабильной и точной оказалась модель на XGBoost. Особенно хорошо она справлялась с учётом сложных

взаимодействий между признаками, что делает её предпочтительной для универсальных моделей раннего выявления.

V. Saravanan и другие [6] предлагают гибридный подход для диагностики меланомы кожи, основанный на сочетании методов машинного и глубинного обучения. Исследование использует данные дерматоскопии и объединяет предсказания моделей KNN, SVM, CNN и Naïve Bayes с помощью ансамблевого метода Majority Voting. Последний показал наивысшую точность – 99.23%. Среди отдельных моделей лучшими оказались Naïve Bayes и SVM. Благодаря высокой эффективности, подход может быть применён в системах поддержки принятия решений в дерматологической практике.

Yue Qiu и другие [7] рассматривают применение алгоритмов машинного обучения для прогнозирования онкологических рисков у пациентов с сахарным диабетом 2 типа на основе лабораторных показателей. Из 1381 признака было отобрано 107 ключевых, включая опухолевые маркёры и биохимические параметры. Среди 16 протестированных моделей наилучшие результаты продемонстрировал CatBoost, показав AUC = 85.2% и высокие значения по всем метрикам. Это подтверждает его применимость для раннего выявления злокачественных процессов у диабетиков.

Jiancheng Wu и другие [8] исследуют применение методов машинного обучения для прогноза выживаемости пациентов с раком желудка после операции. Модель обучалась на базе SEER и прошла валидацию на китайской выборке. Среди протестированных подходов Deep Learning показал наилучшие результаты по AUC на всех этапах прогноза (1, 3 и 5 лет), превзойдя как традиционные модели, так и систему TNM. Кроме того, DL-модель обеспечила лучшую калибровку и клиническую полезность, подтверждая свой потенциал в персонализированной онкопрактике.

Rahul Sharma и другие [9] рассматривают применение машинного обучения для ранней диагностики сердечно-сосудистых заболеваний с акцентом на важность настройки гиперпараметров. Используя два датасета и шесть алгоритмов, авторы показывают, что оптимизация моделей через GridSearchCV значительно повышает точность предсказаний. Особенно высокие результаты были достигнуты с ансамблевыми методами – Random Forest и XGBoost, которые при увеличении объёма данных демонстрировали до 99% точности. Это подчёркивает важность объёма выборки и грамотного тюнинга параметров в клинических задачах.

Таблица 1. Результаты сравнительного анализа методов машинного обучения в задачах профилактики и диагностики онкологических заболеваний

Алгоритмы	Область применения	Точность (AUC-ROC) %
<i>Adaptive Boosting</i>	<i>Классификация рака шейки матки</i>	<i>100</i>
CatBoost	Прогнозирование риска рака	85
Convolutional Neural Network	Классификация меланомы	97.88
	Диагностика рака груди, легких, кожи, мозга, COVID-19	98
Deep Learning	Прогноз послеоперационной выживаемости рака (SEER)	82
<i>Decision Tree</i>	<i>Классификация рака шейки матки</i>	<i>100</i>
	Прогнозирование сердечных заболеваний (Cleveland, расширенный набор)	97

Алгоритмы	Область применения	Точность (AUC-ROC) %
	Классификация заболеваний по системам	50
Extreme Gradient Boosting	Прогнозирование сердечных заболеваний (Cleveland, расширенный набор)	99
Gradient Boosting	Прогнозирование риска рака	82
K-Nearest Neighbors	Прогноз рецидивов опухоли	81
	Прогнозирование сердечных заболеваний (Cleveland, расширенный набор)	87
	Классификация заболеваний по системам	50
	Классификация рака шейки матки	95
	Классификация меланомы	96.11
LightGBM	Прогнозирование риска рака	83
Logistic Regression	Классификация рака шейки матки	98
	Общая модель риска рака лёгких (с учетом SNP и эпидемиологии)	74.2
	Прогноз плоскоклеточного рака лёгкого (SCC)	83.3
	Прогнозирование сердечных заболеваний (Cleveland, расширенный набор)	87
Majority Voting (ансамбль всех моделей)	Итоговая модель (объединение всех моделей)	99.23
Multi-Task Logistic Regression	Прогнозирование послеоперационной выживаемости при раке	78
Multivariate Logistic Regression	Предоперационный диагноз рака / комбинированные модели	93
Naive Bayes	Предсказание болезни Альцгеймера (AD) и легких когнитивных нарушений (MCI)	84
	Классификация рака шейки матки	95
	Прогноз рецидивов опухоли	83
Random Forest	Прогнозирование риска рака	83
	Классификация риска рака шейки матки	97
	Прогнозирование сердечных заболеваний (Cleveland, расширенный набор)	98
	Классификация рака шейки матки	100
Random Survival Forest	Прогнозирование послеоперационной выживаемости при раке	76
Support Vector Machine	Диагностика злокачественности узлов, лимфоузлов рака	99
	Классификация рака шейки матки	99
	Прогнозирование сердечных заболеваний (Cleveland, расширенный набор)	90
	Классификация меланомы	97.93
	Классификация заболеваний по системам	50

Алгоритмы	Область применения	Точность (AUC-ROC) %
	Классификация с высокими размерностями признаков	80
XGBoost	Общая модель риска рака лёгких (с учетом SNP и эпидемиологии)	75.9
	Прогноз риска рака у мужчин	79.1
	Прогнозирование риска рака	83
	Прогноз риска рака у курильщиков	78.5
	Прогноз аденокарциномы лёгкого (ADC)	69.9

Результаты исследования показывают, что алгоритмы AdaBoost (адаптивный бустинг), Decision Tree (дерево решений) и Random Forest (случайный лес) – достигли максимальной точности (100% AUC-ROC) в задаче классификации рака шейки матки. Это говорит об их исключительной эффективности и потенциале для дальнейшего применения в клинической практике при ранней диагностике заболевания.

Алгоритм Random Forest продемонстрировал высокие показатели точности в ряде задач: от 83% при прогнозе риска рака до 97–100% при классификации рака шейки матки. Это подтверждает универсальность модели при работе с разнородными медицинскими данными. Метод отличается стабильной производительностью и может быть применён в клинической практике для поддержки диагностики и прогнозирования онкологических заболеваний. Благодаря адаптивности к различным признаковым структурам, он подходит для интеграции в медицинские информационные системы.

Проведённый сравнительный анализ показал, что методы машинного обучения обладают высоким потенциалом в задачах раннего выявления и прогноза онкологических заболеваний. Среди исследованных моделей наивысшую точность и стабильность продемонстрировали ансамблевые алгоритмы, такие как Random Forest, AdaBoost, а также Design Tree. Особенно эффективными они оказались в задачах классификации рака шейки матки, где точность достигала 100%.

При выборе алгоритма машинного обучения для задач онкопрофилактики следует учитывать специфику данных, целевую метрику и возможность интерпретации модели. Алгоритмы Random Forest и AdaBoost можно рекомендовать как универсальные и надёжные инструменты для внедрения в клинические информационные системы.

Список литературы

1. Backman J., Eriksson R., Franzén O., et al. Using machine learning for predicting cervical cancer from Swedish electronic health records by mining hierarchical representations // *PLOS ONE*. – 2020. – Vol. 15, No. 8. – DOI: [10.1371/journal.pone.0237911](https://doi.org/10.1371/journal.pone.0237911)
2. Петров И.П., Сидоров А.А. Информационная система решения задач в области прогнозирования и профилактики болезней // *Известия ТулГУ. Технические науки*. 2023. Вып. 12. -с. 552-557.
3. DeJohn J., Ahmad H., Romitti M., et al. Application of machine learning methods to improve the performance of ultrasound in head and neck oncology: A literature review // *Cancers*. – 2022. – Vol. 14, No. 3. – P. 665. – DOI: [10.3390/cancers14030665](https://doi.org/10.3390/cancers14030665)
4. Fernandez M.J., Liu X., Rahman M. A model for predicting cervical cancer using machine learning algorithms // *Sensors*. – 2022. – Vol. 22, No. 11. – P. 4132. – DOI: [10.3390/s22114132](https://doi.org/10.3390/s22114132)

5.Li X., Fang Y., Liu Y., et al. Prediction of lung cancer risk in Chinese population with genetic-environment factor using extreme gradient boosting // *Cancer Medicine*. – 2022. – Vol. 11, No. 10. – P. 2045–2055. – DOI: [10.1002/cam4.4800](https://doi.org/10.1002/cam4.4800)

6.Saravanan V., Kumaresan S., Ramesh T. Prediction and classification of skin melanoma cancer using active hybrid machine learning technique // *Journal of Physics: Conference Series*. – 2023. – Vol. 2580. – P. 012039. – DOI: [10.1088/1742-6596/2580/1/012039](https://doi.org/10.1088/1742-6596/2580/1/012039)

7.Qiu Y., Zhang L., Chen X., et al. A machine learning prediction model for cancer risk in patients with type 2 diabetes based on clinical tests // *Technology and Health Care*. – 2024. – Vol. 32, No. 2. – DOI: [10.3233/THC-230385](https://doi.org/10.3233/THC-230385)

8.Wu J., Chen Y., Wang D., et al. Development and validation of a deep learning model for predicting postoperative survival of patients with gastric cancer // *BMC Public Health*. – 2024. – Vol. 24. – Article No. 18221. – DOI: [10.1186/s12889-024-18221-6](https://doi.org/10.1186/s12889-024-18221-6)

9.Sharma R., Goyal N., Chauhan S. Influence of optimal hyperparameters on the performance of machine learning algorithms for predicting heart disease // *Processes*. – 2023. – Vol. 11, No. 3. – Article 734. – DOI: [10.3390/pr11030734](https://doi.org/10.3390/pr11030734)

SEMI-SUPERVISED LEARNING OF LOGISTIC REGRESSION IN CLASSIFICATION PROBLEM

Doshanova Malika Yuldashovna

Associate Professor of the Department of SOIT, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, yulduzxon_85@mail.ru

Otaxanova Baxrihon Ibragimovna

Associate Professor of the Department of SOIT, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi

Aliyev Raxim Raxmanbergen ugli

Student, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi

Abstract: This article discusses a method of semi-supervised learning of logistic regression, for which a special quality functional is introduced for an unlabeled collection, which allows to significantly increase the completeness of the classification.

Keywords: logistic regression, machine learning, semi-supervised learning, linear classifiers, text classification.

ПОЛУ-КОНТРОЛИРУЕМОЕ ОБУЧЕНИЕ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ В ЗАДАЧЕ КЛАССИФИКАЦИИ

Аннотация: В данной статье рассматривается метод полуконтролируемого обучения логистической регрессии, для которого вводится специальный функционал качества для немаркированной коллекции, позволяющий существенно повысить полноту классификации.

Ключевые слова: логистическая регрессия, машинное обучение, полуконтролируемое обучение, линейные классификаторы, классификация текстов.