

АНАЛИЗ ОСОБЕННОСТЕЙ ФУНКЦИОНАЛЬНОГО СОСТОЯНИЯ ОРГАНИЗМА СПОРТСМЕНОВ И НЕ СПОРТСМЕНОВ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

Р.М. Хуснутдинов^{1,2}, А.М. Ситдинов², М.И. Галяутдинов², Л.М. Ахметшина^{3,4}

¹Федеральное государственное бюджетное образовательное учреждение высшего образования «Казанский государственный энергетический университет», г. Казань, Россия

²Федеральное государственное бюджетное образовательное учреждение высшего образования «Поволжский государственный университет физической культуры, спорта и туризма», г. Казань, Россия

³Государственное автономное учреждение здравоохранения «Межрегиональный клинико-диагностический центр», г. Казань, Россия

⁴Государственное автономное учреждение здравоохранения «Зеленодольская центральная районная больница», г. Зеленодольск, Россия

Аннотация. Целью работы является сравнительный анализ функционального состояния спортсменов и не спортсменов с помощью методов машинного обучения. В качестве входных факторов использовались экспериментальные данные по физиологическим параметрам и характеристикам испытуемых, полученные в научно-исследовательском центре ФГБОУ ВО «Поволжский государственный университет физической культуры, спорта и туризма». Отбор значимых факторов осуществлялся с помощью метода главных компонент. Детальный анализ экспериментальных данных был выполнен с помощью моделей искусственной нейронной сети и случайного леса, для которых были найдены оптимальные конфигурации и значения гиперпараметров. Установлено, что оба метода позволяют с высокой степенью вероятности (свыше 90%) идентифицировать спортсменов и не спортсменов по их физиологическим параметрам и характеристикам.

Ключевые слова: метод главных компонент, искусственная нейронная сеть, метод случайного леса, функциональное состояние.

ANALYSIS OF THE FUNCTIONAL STATE OF ATHLETES AND NON-ATHLETES BY MACHINE LEARNING METHODS

R.M. Khusnutdinov^{1,2}, A.M. Sitdikov², M.I. Galyautdinov², L.M. Akhmetshina^{3,4}

¹Kazan State Power Engineering University, Kazan, Russia

²Volga Region State University of Physical Culture, Sports and Tourism, Kazan, Russia

³Interregional Clinical Diagnostic Center, Kazan, Russia

⁴Zelenodolsk Central District Hospital, Zelenodolsk, Russia

Abstract. The aim of the study – a comparative analysis of the functional state of athletes and non-athletes using machine learning methods. The input factors were experimental data on the physiological parameters and characteristics of the subjects obtained at the research center of the Volga Region State University of Physical Culture, Sport and Tourism. The selection of significant factors was carried out using the principal component analysis method. A detailed analysis of the experimental data was performed using artificial neural network and random forest models, for which optimal configurations and hyperparameter values were found. It was found that both methods allow identifying athletes and non-athletes by their physiological parameters and characteristics with a high degree of probability (over 90%).

Keywords: principal component analysis, artificial neural network, random forest method, functional state.

Введение. В современном мире методы интеллектуального анализа данных находят широкое применение в различных областях человеческой деятельности. Так, методы машинного обучения позволяют обрабатывать и анализировать огромные объемы информации, выявлять закономерности и строить прогнозы, порой невозможные для человека. Кроме того, искусственно-интеллектуальные

системы обладают такими важными свойствами как самообучение и принятие автономных решений на основе данных без какого-либо явного сопровождения со стороны человека и программного задания. Применение искусственного интеллекта в описании, предсказании, классификации и интерпретации сложных наборов данных является широко используемой на сегодняшний день

технологией. В частности, в медицине эти методы используются в диагностике и анализе, а также для принятия эффективных решений [1]. В сфере спорта эта технология меняет аналитику, предлагая более глубокое и точное понимание каждого аспекта спортивных результатов [2]. Машинное обучение позволяет проводить комплексный анализ, выходящий за рамки возможностей традиционных методов.

Целью работы является сравнительный анализ функционального состояния спортсменов и не спортсменов с помощью методов машинного обучения.

Методы и организация исследования. С целью выявления особенностей функционального состояния различных групп испытуемых (спортсменов и не спортсменов) на основе экспериментальных данных использовались два метода машинного обучения с учителем: искусственная нейронная сеть и случайный лес. Полносвязная многослойная нейронная сеть состояла из четырех слоев (входного, двух скрытых и выходного). Входной слой содержал 11 нейронов, где на входе принимались параметры испытуемых, характеризующие функциональное состояние и один шумовой параметр. Первый и второй скрытые слои состояли из 32 и 64 нейронов, соответственно. Методом перебора были определены гиперпараметры, обеспечивающие модели наибольшую точность прогнозирования и оптимальные значения весов. Использован стохастический градиентный спуск [3] с разбиением на минибатчи по 10 обучающих выборок и коэффициентом скорости обучения $\gamma=0,001$. Элементарные преобразователи данной сети, включающие входной и скрытые слои, имели нелинейную функцию активации ReLU (Rectified Linear Unit), в то время как сигнал выходного слоя преобразовывался с помощью функции Softmax. Для оценки функции потерь на каждой итерации рассчитывалась среднеквадратическая ошибка между точными и предсказанными значениями. Набор данных был разделен на обучающую и тестовую выборку в соотношении 50% к 50%. При оценке эффективности модели на тестовой выборке использовался критерий качества – перекрестная энтропия.

В качестве второго метода машинного обучения использовался метод случайного

леса (random forest), предложенный Лео Брейманом и Аделем Катлером [4]. Алгоритм применяется для задач классификации, регрессии и кластеризации. Основная идея метода заключается в использовании большого ансамбля решающих деревьев, который сочетает в себе метод бэггинга Бреймана и метод случайных подпространств. В настоящей работе нами было выполнено обучение нескольких моделей случайного леса с различным количеством деревьев. Было установлено, что для стабильной работы алгоритма достаточно $N=10$ деревьев.

Экспериментальные данные представляли собой датасет, состоящий из физиологических параметров и характеристик испытуемых обоих полов в возрасте от 10 до 27 лет, которые включали: возраст, максимальное потребление кислорода (МПК), физическую активность (час/нед.), вес тела (кг), массу жировой ткани (кг), мышечную массу (кг), массу костной ткани (кг), протеин, индекс массы тела (ИМТ), массовую долю воды (кг); параметры, определяющие основной обмен; один шумовой параметр, генерируемый датчиком случайных чисел. Последний параметр использовался с целью проверки стабильности предсказательной способности применяемых в работе методов машинного обучения. В исследованиях участвовали две группы испытуемых: экспериментальная (126 человек, или 82,89% от общего числа испытуемых) и контрольная (26 человек, 17,11%). В экспериментальную группу входили спортсмены, профессионально занимающиеся определенными видами спорта (легкая атлетика, художественная гимнастика, бадминтон, большой теннис, футбол, волейбол, борьба, восточные единоборства, академическая гребля, гребля на байдарках, плавание и др.), а также имеющие спортивные разряды и звания. В контрольную группу входили молодые, здоровые люди, занимающиеся умеренной физической активностью от 2 до 10 часов в неделю. Все исследования проводились на базе научно-исследовательского центра Поволжского государственного университета физической культуры, спорта и туризма [5].

Результаты исследования и их обсуждение. В качестве начального этапа анализа экспериментальных данных мы рассчитали корреляционную матрицу Пирсона для рассматриваемых признаков с целью

определения корреляций между различными параметрами и характеристиками испытуемых (рис. 1). Как видно из рисунка, ряд ингерентных физиологических параметров имеют сильную корреляцию между собой, такие как, например, физическая активность и доля мышечной ткани, массовая доля воды и масса костной ткани и другие. Для выявления значимых факторов мы выполнили

сокращение размерности данных с помощью метода главных компонент (рис. 2). Так как все входные параметры характеризуются различными величинами и диапазонами значений, то в качестве предварительной обработки данных была выполнена процедура калибровки: таким образом, чтобы значения входных параметров варьировались в диапазоне [0;1].

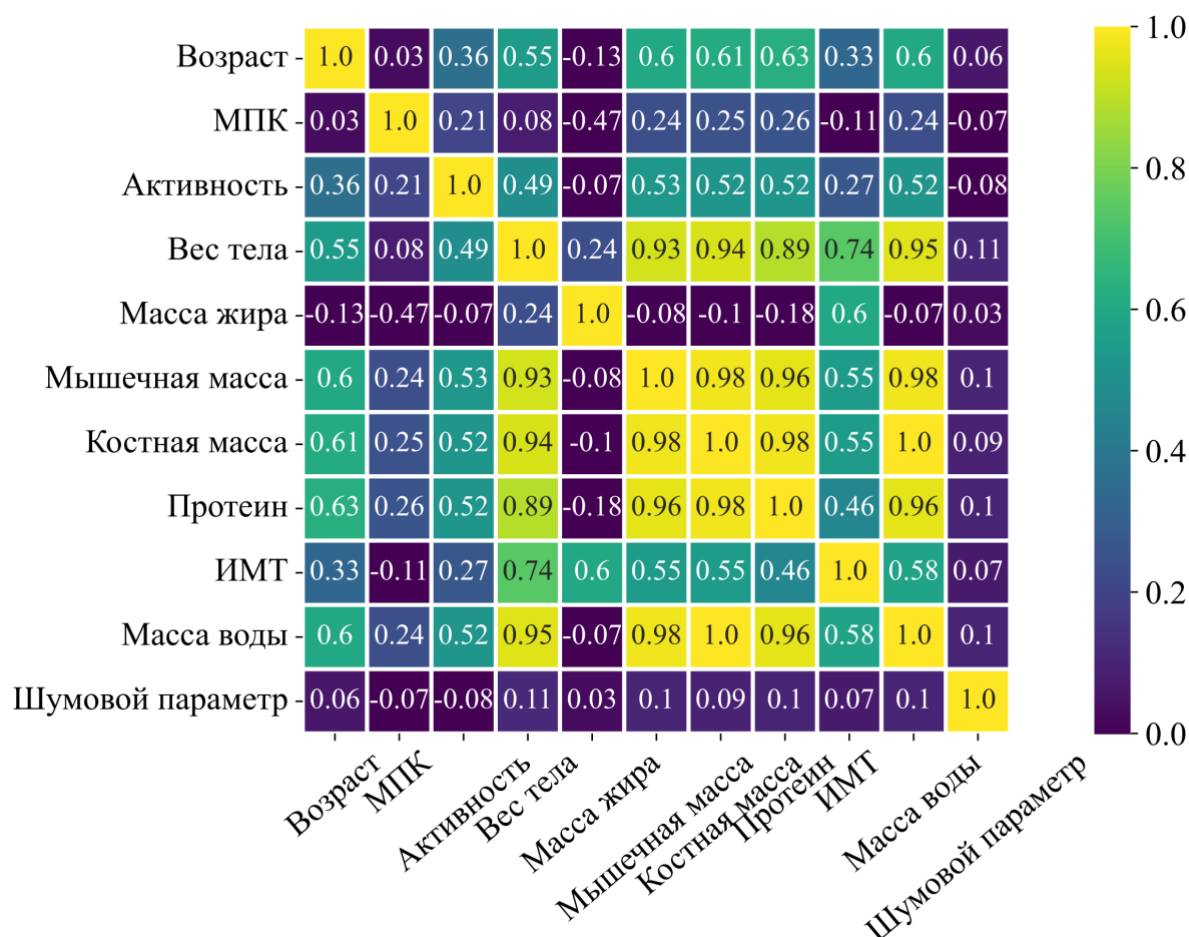


Рис. 1. Матрица корреляций Пирсона для рассматриваемых признаков

Из значения кумулятивного процента объясненной дисперсии (см. левую колонку рис. 2) видно, что 4 первые главные компоненты объясняют 94,4% дисперсии (при рекомендуемых 60% дисперсии). Первая компонента объясняет 60,8% от дисперсии, соответствующей 11 переменным; вторая – 22,9%; третья – 6,0%; четвертая – 4,66%. В правой колонке рисунка 2 представлено распределение экспериментальных данных для спортсменов и не спортсменов в плоскости первых двух компонент. Как видно из рисунка, на данной диаграмме выделяются два кластера: кластер с компактным распределением точек

при отрицательных значениях первой главной компоненты – характеризует данные контрольной группы, в то время как кластер с рассеянным распределением точек в области положительных значений первой главной компоненты характеризует данные экспериментальной группы.

На основе выполненной предобработки данных была построена модель искусственной нейронной сети. На рисунке 3 представлены функция потерь (а) и точность модели (б) в зависимости от номера итерации. Как видно из рисунка, точность модели на тренировочной выборке составила более 93%.

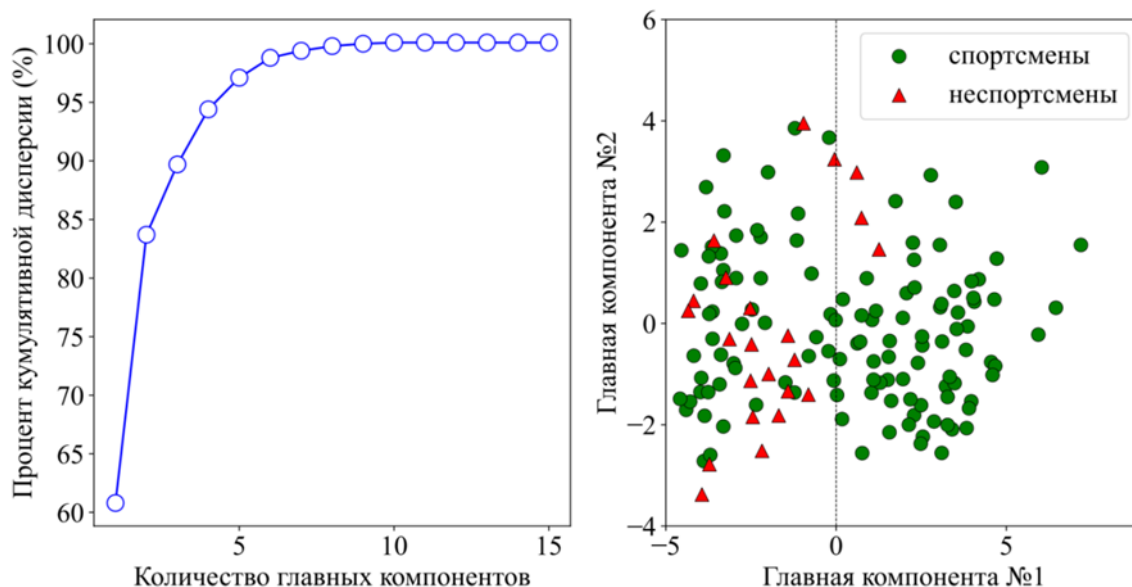


Рис. 2. Метод главных компонент: процент кумулятивной дисперсии в зависимости от количества главных компонент (левая колонка); распределение данных по экспериментальной (кружки) и контрольной (треугольники) группам в плоскости первых двух компонент (правая колонка)

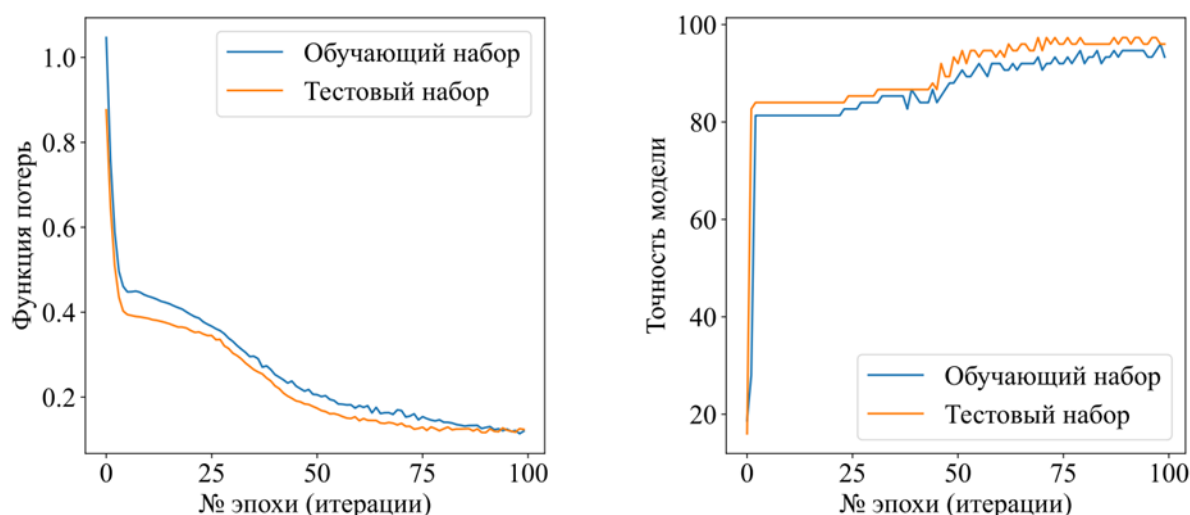


Рис. 3. Функция потерь (левая колонка) и точность модели (правая колонка) в зависимости от номера итерации

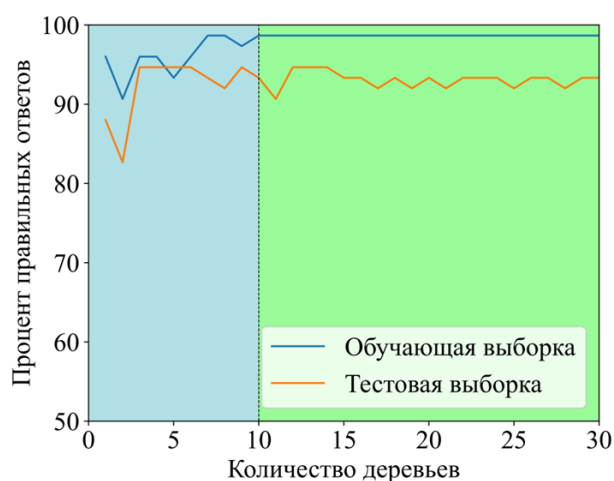


Рис. 4. Процент правильных ответов в зависимости от количества деревьев

Следующим этапом нашего исследования явилось применение модели случайного леса (ансамбль деревьев решений, используемый для задач классификации и регрессии) к анализу полученных экспериментальных данных. Мы выполнили обучение нескольких моделей случайного леса с различным количеством деревьев (рис. 4). Как видно из рисунка, на обучающей выборке процент правильных ответов выходит на плато при значениях $N=10$ деревьев, что указывает на стабильную работу алгоритма

при минимальном количестве деревьев $N=10$. На рисунке 5 представлено сравнение предсказательной способности двух методов машинного обучения: представленной модели искусственной нейронной сети и модели случайного леса с количеством деревьев $N=10$. Как видно из рисунка, оба метода приводят к одинаковым результатам, а именно позволяют с высокой вероятности (свыше 90%) идентифицировать спортсменов и не спортсменов по их физиологическим параметрам и характеристикам.

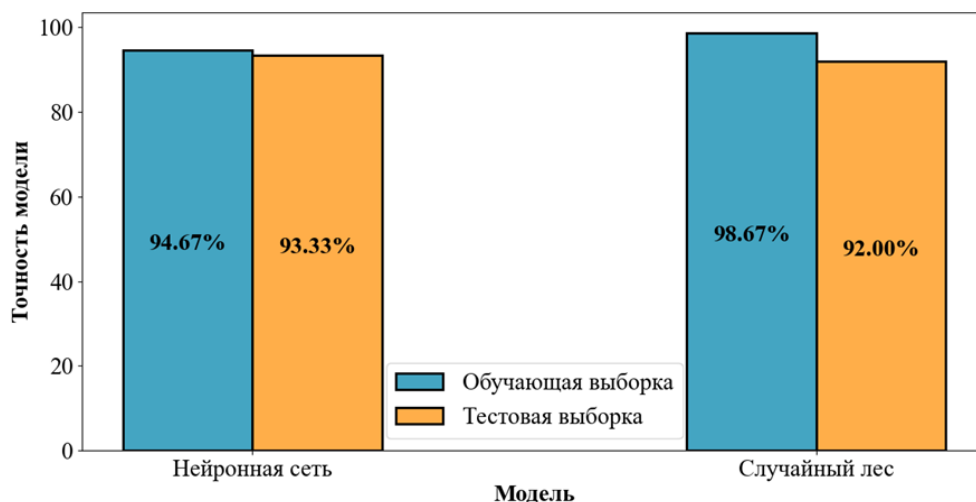


Рис. 5. Сравнение предсказательной способности двух методов машинного обучения: модели искусственной нейронной сети и модели случайного леса

Заключение. В настоящей работе выполнено систематическое исследование функционального состояния спортсменов и не спортсменов с помощью двух методов машинного обучения: искусственной нейронной сети и модели случайного леса. В качестве входных факторов использовались экспериментальные данные по физиологическим параметрам и характеристикам испытуемых. Отбор значимых факторов осуществлялся с помощью метода главных компонент. Установлено, что четыре первые главные компоненты объясняют 94,4% дисперсии (при рекомендуемых 60% дисперсии). Для моделей искусственной нейронной сети и случайного

леса были найдены оптимальные конфигурации и значения гиперпараметров. Установлено, что оба метода позволяют с высокой вероятностью (свыше 90%) идентифицировать спортсменов и не спортсменов по их физиологическим параметрам и характеристикам.

Таким образом, в настоящей работе были разработаны модели нейронной сети и случайного леса, позволяющие по набору физиологических параметров определять функциональное состояние испытуемых и идентифицировать их как спортсменов и не спортсменов — молодых, здоровых людей, занимающихся умеренной физической активностью.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Conflict of interest. The authors declare no conflict of interest.

СПИСОК ЛИТЕРАТУРЫ

1. Гифт, Н. Прагматичный ИИ. Машинное обучение и облачные технологии / Н. Гифт; пер. с англ. И. Пальти. — СПб: Питер, 2019. — 300 с.

2. Васильев, А.Н. Нейросетевое моделирование. Принципы. Алгоритмы. Приложения / А.Н. Васильев, Д.А. Тархов. — СПб: Изд-во Политехн. Ун-та, 2009. — 527 с.

3. Хайкин, С. Нейронные сети: полный курс. 2-е изд. / С. Хайкин. – М.: «Вильямс», 2006. – 1104 с.
4. Breiman, L. Random Forests / L. Breiman // *Machine Learning*. – 2001. – № 45. – Р. 5–32.
5. Сравнительный анализ функционального состояния организма спортсменов и не спортсменов по данным антропометрии, морфологии и МПК / Р.М. Хуснутдинов, А.М. Ситдинов, И.Р. Фаткуллов [и др.] // *Ученые записки университета имени П.Ф. Лесгафта*. – 2023. – № 12 (226). – С. 208–213.

REFERENCES

1. Gift N. Pragmatic AI. Machine learning and cloud technologies: scientific. ed. trans. from English by I. Palti. St. Petersburg: Piter,

2019, 300 p. (in Russ.)

2. Vasil'ev A.N., Tarkhov D.A. Neural network modeling. Principles. Algorithms. Applications. St. Petersburg: Publishing house Polytechnic. Univ., 2009, 527 p. (in Russ.)
3. Khaikin S. Neural networks: a complete course. 2nd ed. Moscow: Williams, 2006, 1104 p. (in Russ.)
4. Breiman, L. Random Forests. *Machine Learning*, 2001, vol. 45, pp. 5–32.
5. Khusnutdinova R.M., Sitdikov A.M., Fatkulov I.R., Galyautdinov M.I., Akhmetshina L.M. Comparative analysis of the functional state of the body of athletes and non-athletes according to anthropometry, morphology and BMD. *Scientific Notes of the P.F. Lesgaft University*, 2023, no. 12, pp. 208–213. (in Russ.)

СВЕДЕНИЯ ОБ АВТОРАХ:

Рамиль Миннегазович Хуснутдинов – кандидат физико-математических наук, доцент кафедры информационных технологий и интеллектуальных систем, ФГБОУ ВО «КГЭУ»; доцент кафедры информационных систем и фиджитал спорта, ФГБОУ ВО «ПГУФКСИТ», Казань, e-mail: khrm@mail.ru

Алмаз Муллаянович Ситдинов – старший преподаватель кафедры информационных систем и фиджитал спорта, ФГБОУ ВО «ПГУФКСИТ», Казань, e-mail: cam77@mail.ru

Марат Ильдарханович Галаяутдинов – кандидат физико-математических наук, доцент кафедры информационных систем и фиджитал спорта, ФГБОУ ВО «ПГУФКСИТ», Казань, e-mail: gmarat_68@mail.ru

Лилия Мустакимовна Ахметшина – врач-кардиолог ГАУЗ «МКДЦ», Казань; врач-терапевт, ГАУЗ «ЗЦРБ», Зеленодольск, e-mail: alm717@mail.ru

INFORMATION ABOUT THE AUTHORS:

Ramil M. Khusnutdinoff – Candidate of Physical and Mathematical Sciences, Associate Professor of the Department of Information Technologies and Intelligent Systems, Kazan State Power Engineering University; Associate Professor of the Department of Information Systems and Phygital Sports, Volga Region State University of Physical Culture, Sports and Tourism, Kazan, e-mail: khrm@mail.ru

Almaz M. Sitdikov – Senior Lecturer of the Department of Information Systems and Phygital Sports, Volga Region State University of Physical Culture, Sports and Tourism, Kazan, e-mail: cam77@mail.ru

Marat I. Galyautdinov – Candidate of Physical and Mathematical Sciences, Associate Professor of the Department of Information Systems and Phygital Sports, Volga Region State University of Physical Culture, Sports and Tourism, Russia, Kazan, e-mail: gmarat_68@mail.ru

Lilia M. Akhmetshina – Cardiologist, Interregional Clinical Diagnostic Center, Kazan; General Practitioner, Zelenodolsk Central District Hospital, Zelenodolsk, e-mail: alm717@mail.ru

Для цитирования: Анализ особенностей функционального состояния организма спортсменов и не спортсменов методами машинного обучения / Р.М. Хуснутдинов, А.М. Ситдинов, М.И. Галаяутдинов, Л.М. Ахметшина // *Российский журнал спортивной науки: медицина, физиология, тренировка*. – 2025. – Т. 4. – № 1(13). DOI: 10.24412/2782-6570-2025_04_01_2

For citation: Khusnutdinoff R.M., Sitdikov A.M., Galyautdinov M.I., Akhmetshina L.M. Analysis of the functional state of athletes and non-athletes by machine learning methods. *Russian Journal of Sports Science: Medicine, Physiology, Training*, 2025, vol. 4, no. 1(13). DOI: 10.24412/2782-6570-2025_04_01_2