

Научная статья

ТЕХНОЛОГИИ МАШИННОГО ОБУЧЕНИЯ ПРИ ОПРЕДЕЛЕНИИ ФАКТОРОВ, СВЯЗАННЫХ С ДОЛЕЙ ЖИРА В ОРГАНИЗМЕ ВЬЕТНАМСКИХ ПОДРОСТКОВ

Нгуен Тхи Тхонг Хан¹, Ле Тхи Туйет¹, Нгуен Тхи Трунг Тху¹, До Тхи Нху Транг¹,
Дуонг Тхи Анх Дао¹, Ле Тхи Тхуй Дунг², Данг Ксуан Тхо³

¹Национальный университет образования Ханоя, Вьетнам, г. Ханой, ул. Ксуан Туй, 136

²Главная больница Бинх Донг, Вьетнам, г. Бинх Донг, Хьеп Тханг, Пхам Нгок Тхак, 5

³Академия политики и развития, Вьетнам, г. Ханой, район Хуай Дук

С применением технологий машинного обучения для прогнозного анализа изучены факторы, влияющие на процент жира в организме вьетнамских подростков в возрасте от 11 до 15 лет.

В исследовании приняли участие 1208 подростков (598 мальчиков и 610 девочек), обучающихся в девяти средних общеобразовательных школах столицы Вьетнама. Изучение состава тела осуществлялось при помощи прибора HBF 375 (Отгюп) методом биоимпедансного анализа. Опросник, примененный в исследовании, был первоначально утвержден Национальным институтом питания. Он включал в себя вопросы, связанные с пищевыми привычками, частотой приема пищи, уровнем физической активности, склонностью к сидячему образу жизни и знаниями о принципах правильного питания. В исследовании применялись технологии машинного обучения с использованием алгоритма дерева решений для определения основных детерминант, имеющих наиболее значимую корреляционную связь с долей жира в организме.

Авторам исследования удалось установить шесть основных групп предикторов, связанных с долей жира в организме подростков с помощью модели дерева решений со значениями среднеквадратической ошибки и средней абсолютной ошибки, равными 4,8 и 3,8 соответственно. Среди данных предикторов факторами, оказывающими наибольшее влияние на долю жира в организме подростков, являются частота употребления фруктов, привычки, связанные с перекусом, способ передвижения по дороге в школу и обратно и экранное время (компьютер и / или смартфон). Комбинация данных факторов во взаимодействии с полом и стадией полового созревания могут определять долю жира в организме вьетнамских подростков.

Данное исследование проливает определенный свет на комплексные и разнообразные факторы, влияющие на долю жира в организме вьетнамских подростков. Результаты подтверждают необходимость формирования приверженности здоровым пищевым привычкам и физическим упражнениям среди подростков, а также дают понимание важности данных вопросов для родителей и руководителей школ для разработки более эффективных стратегий деятельности образовательных организаций.

Ключевые слова: машинное обучение, доля жира в организме, предсказуемость, факторы влияния, пищевые привычки, физическая активность, вьетнамские подростки, дерево решений.

© Нгуен Тхи Тхонг Хан, Ле Тхи Туйет, Нгуен Тхи Трунг Тху, До Тхи Нху Транг, Дуонг Тхи Анх Дао, Ле Тхи Тхуй Дунг, Данг Ксуан Тхо, 2024

Нгуен Тхи Тхонг Хан – кандидат наук, преподаватель кафедры биологии (e-mail: hanhnhth@hnue.edu.vn; тел.: +84 (902) 80-11-83; ORCID: <https://orcid.org/0000-0002-1392-8081>).

Ле Тхи Туйет – кандидат наук, преподаватель физиологии человека и животных, кафедра биологии (e-mail: tuyetlt@hnue.edu.vn; тел.: +84 (968) 79-55-55; ORCID: <https://orcid.org/0000-0002-3308-5886>).

Нгуен Тхи Трунг Тху – кандидат наук, преподаватель кафедры биологии (e-mail: trungthu@hnue.edu.vn; тел.: +84 (983) 42-09-85; ORCID: <https://orcid.org/0000-0002-9800-2287>).

До Тхи Нху Транг – магистр, помощник преподавателя кафедры биологии (e-mail: trangdtn@hnue.edu.vn; тел.: +84 (986) 70-38-10; ORCID: <https://orcid.org/0009-0009-5461-4554>).

Дуонг Тхи Анх Дао – кандидат наук, доцент, преподаватель кафедры биологии (e-mail: daodangduc@gmail.com; тел.: +84 (886) 81-67-66; ORCID: <https://orcid.org/0009-0003-4407-8942>).

Ле Тхи Тхуй Дунг – кандидат наук, доктор медицины, педиатрическое отделение (e-mail: letono2002@gmail.com; тел.: +84 (987) 00-89-14; ORCID: <https://orcid.org/0000-0002-8855-7801>).

Данг Ксуан Тхо – кандидат наук, главный преподаватель факультета цифровой экономики (тел.: +84 (912) 62-93-83; ORCID: <https://orcid.org/0000-0002-7654-5942>).

Доля жира в организме является общепризнанным точным и эффективным способом измерения статуса ожирения. Предыдущими исследованиями доказано, что данный показатель может также способствовать пониманию рисков сердечно-сосудистых заболеваний и других метаболических нарушений, поскольку адипокины, выделяемые жировой тканью, могут оказать влияние на многие метаболические функции, включая распределение жира, аппетит, расход энергии, чувствительность к инсулину, выделение инсулина, глюкозный и липидный обмен [1]. Взаимосвязь между долей жира в организме и факторами риска ожирения является довольно противоречивой при рассмотрении популяций, различных по этническому признаку. Помимо этого, в литературе имеется довольно ограниченная информация об эффективности данных методов оценки ожирения в странах Азии. В азиатских популяциях в особенности существуют отличия в относительных вкладах мускульной массы, скелетной массы и объема жидкости в общий вес тела по сравнению с европейцами. Данные отличия обусловлены влиянием культурных подгрупп, социоэкономических условий, а также инфраструктурных характеристик [2]. В то время как доля жира в организме играет важнейшую роль в оценке распространенности избыточного веса и ожирения, в литературе отсутствуют исчерпывающие данные относительно распределения жира в подростковых популяциях, особенно в азиатских регионах.

На глобальном уровне урбанизация значительно ускоряет процесс изменения пищевых привычек и физической активности, и Вьетнам не является исключением [3]. Пищевые привычки и уровни физической активности считаются ключевыми факторами изменения доли жира в организме [4]. Результаты исследования, выполненного в Тегеране, показали, что всесторонние изменения образа жизни в сторону его оздоровления привели к снижению доли жира в организме подростков с ожирением на 1,81 % спустя 12 недель эксперимента [5]. В исследовании с участием 764 итальянских школьников в возрасте 10–18 лет было показано, что подростки со здоровыми пищевыми привычками и регулярной физической активностью имели более низкую долю жира в организме, чем их сверстники с менее здоровыми пищевыми привычками и нерегулярной физической активностью [6]. Далее в исследовании с участием 70 школьников в возрасте 14–15 лет также была обнаружена значимая взаимосвязь между физической активностью и долей жира в организме [7]. Анализ факторов риска, связанных с пищевыми привычками и физической активностью, является основой для выработки рекомендаций по своевременным мерам, направленным на сдерживание роста доли жира в организме в раннем возрасте, а также на предотвращение избыточного веса и ожирения в школьном возрасте [8–10]. Машинное обучение за-

ключается в применении искусственного интеллекта, который помогает системам автоматически понимать данные с учетом обучения без необходимости дополнительного специального программирования. По сравнению с классической статистикой, новизна данного метода заключается в умении машины эффективно делать выводы и обучаться на выборках данных с использованием подходящих алгоритмов и компьютерных способов обработки больших массивов данных. Таким образом, машинное обучение считается передовой дисциплиной в современной статистике и, в более широком смысле, в сфере науки о данных. Поэтому с развитием больших данных необходимым является применение алгоритмов машинного обучения для извлечения информации с целью получения интеллектуальных данных. Исследование с участием 7162 человек, проведенное в Китае с использованием 11 алгоритмов машинного обучения, показало, что применение алгоритма произвольной пространственной классификации помогает достигать большой точности результатов и области под кривой. В исследовании с применением оценки ИМТ было показано наличие сильной взаимосвязи между продолжительностью физической активности с высокой интенсивностью в неделю и продолжительностью физической активности с умеренной интенсивностью в неделю и ожирением [11]. В исследовании авторов В. Oladapo et al. (2020) изучалась применимость машинного обучения для более эффективного прогнозирования массы тела при применении определенной диеты [12]. В исследовании с участием 40 032 британцев было обнаружено, что модели машинного обучения, основанные на двухмерной проекции, позволяют получать близкие к совершенству оценки объема жировой ткани [13].

Однако в настоящее время в литературе можно найти очень ограниченное количество работ, в которых описывается применение машинного обучения для анализа взаимосвязи между пищевыми привычками, физической активностью и долей жира в организме.

Цель исследования – определение факторов риска и факторов защиты, связанных с долей жира в организме, и их предсказуемости в репрезентативной выборке, состоящей из школьников в возрасте от 11 до 15 лет, с применением технологий машинного обучения. Результаты данного исследования могут послужить основой для выработки рекомендаций по суточному потреблению пищи, а также разработки стратегий и корректировки профилактических программ с целью контроля возрастающей распространенности ожирения среди данной группы населения.

Материалы и методы. В исследовании приняли участие 1208 подростков (598 мальчиков и 610 девочек), обучающихся в девяти средних общеобразовательных школах столицы Вьетнама (выбраны из 583 школ Ханоя методом случайной выборки). Группы исследования учащихся в каждой школе сформиро-

рованы при помощи программного обеспечения Epi info 6. В исследование включались здоровые школьники при условии отсутствия диагностированных хронических заболеваний, которые не принимали никакие лекарства, способные повлиять на состав тела (например, бета-блокаторы или диуретики).

Протокол исследования был одобрен локальным комитетом по биомедицинской этике кафедры физиологии человека и животных факультета биологии Национального университета образования Ханоя. Процедуры сбора и хранения данных соответствовали этическим требованиям и полностью обеспечивали конфиденциальность участников. Исследователи дали подробные объяснения директорам школ, учителям, участникам и их родителям относительно потенциальных преимуществ и рисков, связанных с данным исследованием. Родители или законные представители подростков предоставили письменное добровольное информированное согласие на их участие в исследовании. В течение всего исследования у подростков было право отказаться от ответа на любой вопрос или прервать свое участие в любой момент.

Измеряемые антропометрические параметры включали в себя вес (Weight), рост (Height), окружность талии (WC) и окружность бедер (HC); измерения выполнялись в соответствии со стандартными методиками, разработанными Национальным институтом питания. Во время измерения участники были одеты в легкую одежду, предварительно сняв обувь и аксессуары для волос (при наличии). Рост, окружность талии и бедер измерялись с точностью до 1 мм, а вес тела измерялся с точностью до 100 г с помощью стандартных медицинских весов. Окружность талии измерялась на уровне посередине между подвздошным гребнем и нижним ребром, в окружность бедер – в самом широком месте ягодиц с помощью неэластичной мерной ленты. Каждое измерение выполнялось дважды для каждого ребенка, для последующего анализа применялось среднее значение. Индекс массы тела (ИМТ – BMI) рассчитывался по стандартной формуле делением веса в килограммах на квадрат роста в метрах. Пищевой статус школьников оценивался с помощью критериев, установленных Специальной рабочей группой по проблемам ожирения (International Obesity Task Force (IOTF)).

Биоэлектрический импедансный анализ (БИА): состав тела участников данного исследования был определен с использованием устройства HBF 375 (Omron). Исследовалась доля жировой массы (BFP). Перед проведением данных измерений участники воздерживались от интенсивной физической активности по меньшей мере в течение 12 ч, от приема воды или напитков в течение как минимум 3 ч, а также помочились и совершили акт дефекации в течение последних 30 мин до измерения. Участницы женского пола исключались из измерений во время менструации. Все измерения в рамках БИА выполнялись утром.

Опросник, примененный в исследовании был предварительно утвержден Национальным институтом питания. Он включал в себя вопросы на тему пищевых привычек, частоты основных приемов пищи, физической активности, склонности к сидячему образу жизни и знаний о принципах здорового питания. Общая информация включала в себя возраст и пол участников, а также место проживания. Характеристики пищевых привычек включали в себя 18 пунктов: наличие завтрака; скорость потребления пищи; количество приемов пищи в день; привычка перекусывать; тип пищевых продуктов, используемых для перекуса; обычное время дня для перекуса; вкусовые предпочтения (овощи / фрукты / жирная пища / сладкая пища / газированные напитки / фаст-фуд / субпродукты); частота употребления овощей / фруктов / жирной пищи / сладкой пищи / газированных напитков / фаст-фуда / субпродуктов. Физическая активность (ФА) измерялась по 27 параметрам: предпочитаемые виды ФА, способ передвижения по дороге в школу и обратно, частота пеших или велосипедных прогулок, участие в игровых видах спорта, частота занятий физкультурой, время, проведенное в малоподвижности, продолжительность сна, частота интенсивных, умеренных или легких физических нагрузок. Знания о принципах здорового питания оценивались по 9 параметрам, а именно наличие представления о питании как необходимости потреблять полезные для здоровья продукты, определении ожирения, адекватных методах контроля веса.

Алгоритмы машинного обучения. Алгоритм дерева решений был использован для прогнозирования целевых значений в задачах регрессии или целевых классов в задачах классификации. Основное значение придавалось решению задач регрессии, где терминальные узлы дерева решений, также известные как краевые узлы, могут накапливать непрерывные значения, как правило, в виде действительных чисел.

Процесс конструирования дерева регрессии подразумевает итерационное разделение набора данных на все более мелкие подмножества с одновременным прогрессивным расширением структуры дерева решений. Окончательным результатом является полностью выросшее дерево, объединяющее узлы решений и краевые узлы. Узлы решений (например, «пол») разделяются в точке бифуркации на две ветви или более (например, «мужской» или «женский»), каждая из которых представляет возможные значения атрибута. С другой стороны, краевые узлы (например, «доля жира в организме») означают конечное решение, относящееся к числовому целевому значению. Самый верхний узел решения на дереве, который соответствует наиболее информативному предиктору, считается корневым узлом. В течение всего экспериментального исследования использовались обширные возможности библиотеки Scikit-learn, набора инструментов для машинного обучения на основе языка Python [14].

Общие характеристики участников исследования

Параметр	Мальчики, N = 598	Девочки, N = 610	Значение P
Возраст, лет ^b	13,0 (12,1–13,8)	12,9 (11,9–13,7)	0,406
Вес, кг ^b	44,8 (37,3–52,2)	43,6 (37,8–49,0)	0,03
Рост, см ^a	154,9 ± 10,3	153,1 ± 6,6	< 0,0001
ИМТ, кг/м ² ^b	18,5 (16,6–20,8)	18,4 (16,5–20,2)	0,097
Окружность талии, см ^b	67,0 (62,5–73,1)	65,3 (62,0–69,6)	< 0,0001
Соотношение талия – бедра ^b	0,81 (0,78–0,86)	0,78 (0,75–0,81)	< 0,0001
Пищевой статус			
– ожирение, %	4,5	0,8	0,001
– избыточный вес, %	13,5	11,4	
– нормальный вес, %	67,6	73,9	
– недостаток веса, %	14,4	13,9	
ДЖ, % ^b	18,0 (13,3–22,6)	21,1 (18,4–23,5)	< 0,0001
Доля подкожного жира, % ^a	12,5 ± 4,3	17,9 ± 4,0	< 0,0001
Доля мышечной массы, % ^b	36,3 (33,5–38,7)	29,5 (28,1–31,0)	< 0,0001

Примечание: ИМТ – индекс массы тела; ДЖ – доля жира; ^a – данные представлены как среднее ± SD; ^b – данные представлены в виде медианы (интерквартильный размах). Значения P получены с помощью t-критерия Стьюдента или критерия Манна – Уитни или критерия хи-квадрат. Значимые различия между случаями и контролями выделены жирным.

Оценка модели. При оценке регрессионной модели мы применили два необходимых показателя: среднеквадратическая ошибка (RMSE) и средняя абсолютная ошибка (MAE) [15]. Они служат показателями точности модели и величины отклонения от действительных значений. Технически среднеквадратическая ошибка RMSE является квадратным корнем среднего значения квадратичных ошибок, в то время как средняя абсолютная ошибка MAE является средним значением абсолютных ошибок. В данном контексте ошибка представляет собой расхождение между прогнозными величинами (оцененными с помощью регрессионной модели) и истинными значениями переменной. В данном конкретном случае были применены следующие формулы:

$$RMSE = \sqrt{\frac{\sum (y_i - y_p)^2}{n}}, \quad (1)$$

$$MAE = \frac{\sum |y_i - y_p|}{n}, \quad (2)$$

где y_i – действительное значение; y_p – прогнозное значение, n – количество наблюдений.

Результаты и их обсуждение. В табл. 1 представлены характеристики участников исследования в разрезе пола. Средний возраст мальчиков и девочек был сходным – 13,0 и 12,9 г. соответственно. Однако мальчики имели значимо больший вес, рост, окружность талии, соотношение «талия – бедра» и долю мышечной массы, по сравнению с девочками ($P < 0,001$). И наоборот, доля жира и подкожного жира были ниже у мальчиков, чем у девочек. У мальчиков средняя доля жира составляла 18,0 %, а средняя доля подкожного жира – 12,5 %, в то время как у девочек значения данных показателей достигали 21,1 и 17,9 % соответственно. Также была отмечена разница

в пищевом статусе по оценке согласно критериям Международной рабочей группы по проблемам ожирения IOTF между подростками мужского и женского пола ($P = 0,001$), а именно доля подростков с избыточным весом и ожирением была значительно выше среди мальчиков, чем среди девочек (18 против 12,2 %).

При изучении взаимосвязей между долей жира в организме и некоторыми антропометрическими переменными у вьетнамских подростков было выявлено несколько значимых связей. Сильная положительная корреляция была обнаружена между долей жира в организме и индексом массы тела (ИМТ) ($r = 0,51$), окружностью талии ($r = 0,42$) и окружностью бедер ($r = 0,31$). И наоборот, отрицательная корреляция наблюдалась между долей жира в организме и ростом ($r = -0,30$) (рис. 1).

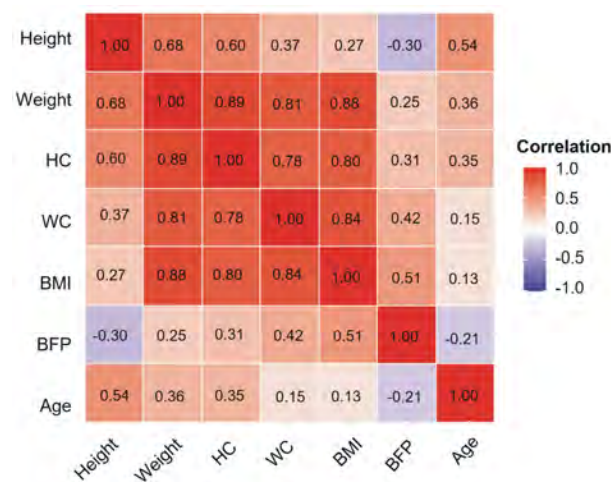


Рис. 1. Взаимосвязь между долей жира в организме и некоторыми антропометрическими переменными

Т а б л и ц а 2

Оценка модели дерева решений, N = 1208

Категория	Обучение (80 %: 966)	Проверка (20 %: 242)
RMSE	4,20	4,80
MAE	3,29	3,80

Примечание: RMSE – среднеквадратическая ошибка; MAE – средняя абсолютная ошибка.

Разработка прогнозной модели для подростков 11–15 лет. Анализ с использованием дерева решений. Согласно результатам эксперимента, оцененным с помощью модели дерева решений и приведенным в табл. 2, для обучающего набора данных получены значения RMSE и MAE, равные 4,20 и 3,29, а для проверочного набора данных – 4,80 и 3,80 соответственно. Различия в значениях среднеквадратической и средней абсолютной ошибки, полученных для двух наборов данных, являются относительно небольшими. Следовательно, прогнозная модель, разработанная в данном исследовании, может считаться высокостабильной и эффективной.

Дерево решений позволило спрогнозировать значение доли жира в организме подростков на основании взаимодействия с внешними факторами, такими как пол, стадия полового созревания и факторы образа жизни. В модель дерева решений входят семь терминальных узлов. В особенности четыре фактора образа жизни среди всех терминальных узлов вносят весомый вклад в повышение доли жира в организме. В разрезе пола эти четыре узла выглядят следующим образом: девочки, не склонные к перекусам (прогнозируемая доля жира в организме

составляет 22,85); девочки, не склонные к перекусам и проводящие более 2 ч в день у экрана (21,815); мальчики, которые употребляют фрукты менее трех раз в неделю (20,749); мальчики, которые употребляют фрукты менее трех раз в неделю и добираются до школы, используя транспорт, а не пешком или на велосипеде (26,641) (рис. 2).

Оценка модели. Далее для более исчерпывающей оценки эффективности метода дерева решений мы сравнили его результаты с таковыми, полученными при помощи линейной регрессии как альтернативного метода.

Результаты, полученные при тестовой выборке и приведенные на рис. 1, также подтверждают, что метод дерева решений позволяет добиться относительно низких значений RMSE и MAE – 4,80 и 3,80 соответственно. И наоборот, применение метода линейной регрессии связано с более высокими значениями: 5,12 для RMSE и 4,20 для MAE (табл. 3). Следовательно, можно сделать вывод, что метод дерева решений позволяет добиться более достоверных результатов по сравнению с методом линейной регрессии.

В данном исследовании была применена методология машинного обучения для прогнозирования доли жира в организме на основании пищевых привычек и физической активности подростков в возрасте 11–15 лет. Результаты исследования указывают, что в данной возрастной группе несколько факторов образа жизни оказывают наиболее значительное влияние на долю жира в организме. Эти факторы включают в себя потребление фруктов, привычки, связанные с перекусом, способ добираться до школы и продолжительность использования компьютера или смартфона в течение дня.

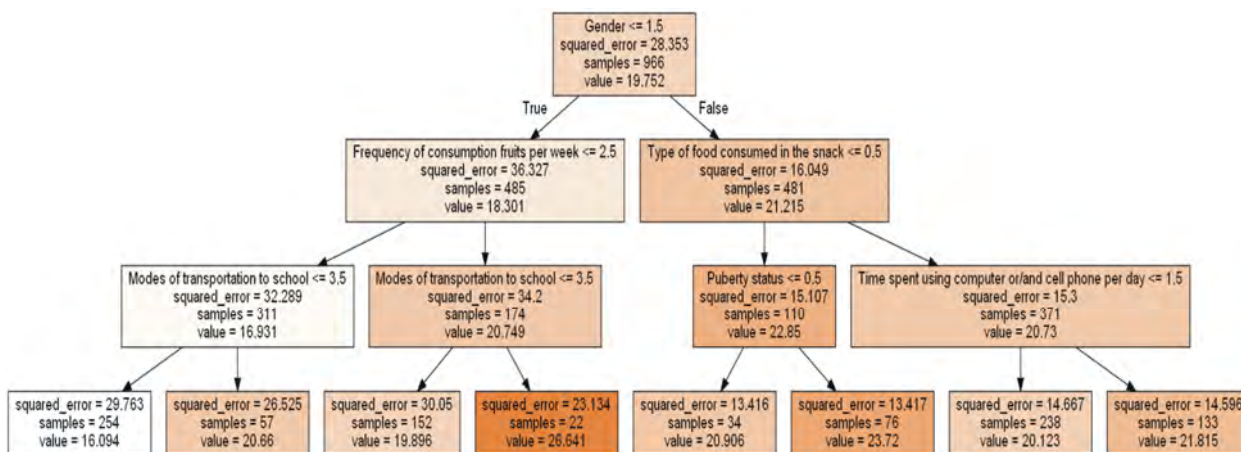


Рис. 2. Модель дерева решений помогает идентифицировать группы риска с высоким ИМТ среди подростков в возрасте 11–15 лет: Value (значение) – прогнозируемая доля жира; Gender (пол): 1 – мужской; 2 – женский; Frequency of consumption fruits per week (частота потребления фруктов в неделю): 1 – более 5 раз/неделю; 2 – 3–5 раз/неделю; 3 – менее 3 раз/неделю; Types of food consumed in the snack (тип пищевых продуктов для перекуса): 0 – ничего; 1 – молоко и молочные продукты; 2 – высокоуглеводные продукты; 3 – фастфуд; Modes of transportation to school (способ передвижения по дороге в школу и обратно): 1 – пешком; 2 – велосипед; 3 – автобус; 4 – мотоцикл; 5 – машина; Puberty status (Стадия полового созревания): 0 – 1, 2, 3 по шкале Таннера; 1 – 4, 5 по шкале Таннера; Time spent using computer or/and cell phone per day (Время, проведенное за компьютером или смартфоном в день): 1 – менее 2 ч/день; 2 – 2 ч/день и более

Таблица 3

Показатели результативности алгоритмов машинного обучения

Метод	RMSE	MAE
Линейная регрессия	5,12	4,20
Дерево решений	4,80	3,80

Примечание: RMSE – среднеквадратическая ошибка; MAE – средняя абсолютная ошибка.

В контексте данного исследования для измерения доли жира в организме применялся биоимпедансный анализ (БИА). Данный метод широко используется в эпидемиологических и клинических исследованиях для оценки состава тела [16]. БИА отличается высокой безопасностью, особенно в исследованиях подростков, а также стабильностью, что и обуславливает его популярность. Несмотря на неотъемлемый предел погрешности, присущий методу, во многих исследованиях указывается, что этот предел остается на сравнительно минимальном уровне [17, 18]. В исследовании с участием 200 здоровых волонтеров обнаружено, что коэффициент корреляции между БИА (с применением HBF 359) и измерением доли жира при помощи двухэнергетической рентгеновской абсорбциометрии составлял 0,89 ($P < 0,001$) [19].

Результаты нашего исследования показали, что метод дерева решений позволяет добиться более достоверных результатов по сравнению с линейной регрессией. Дерево решений представляет собой один из наиболее распространенных алгоритмов машинного обучения, который дает возможность эффективно решать как задачи регрессии, так и задачи классификации. Метод дерева решений представляет собой действенный инструмент прогнозирования избыточного веса и ожирения на основании анализа набора разнообразных факторов, включая образ жизни, пищевые привычки, генетику и т.д. [20, 21]. Применение широкого спектра алгоритмов машинного обучения описывается в многочисленных исследованиях, где они позволили добиться значительной точности в прогнозировании ожирения [22]. Исследования, в которых для оценки доли жира применяется машинное обучение, предлагают перспективный способ понимания сути данной проблемы, а также обладают потенциалом повышения точности и эффективности прогнозирования содержания жира в организме. С другой стороны, применение деревьев решений может прогнозировать долю жира в организме на основании анализа совокупности взаимодействующих многочисленных факторов риска или профилактики вместо оценки влияния каждого фактора по отдельности. Точная оценка доли жира в организме с использованием машинного обучения имеет огромное значение для охраны здоровья населения [22]. Однако прогнозирование доли жира в организме подростков в зависимости от влияния внешнесредовых факторов остается малоизученным.

В нашем исследовании самым точным предиктором оказались потребление фруктов, а также наличие и состав перекусов. В многочисленных исследованиях изучается взаимосвязь между рационом питания и ИМТ, ожирением и избыточным весом у подростков в разных странах. Однако результаты данных исследований не были последовательно соотнесены друг с другом. В предыдущем исследовании, проведенном в 34 странах, было отмечено отсутствие значимой взаимосвязи между избыточным весом и потреблением фруктов или овощей [23]. Далее в проспективном когортном исследовании, проведенном в США с участием 15 900 детей, выявлено, что потребление фруктов и фруктового сока не может считаться предиктором изменений в ИМТ [24]. Согласно данным, полученным в Международном исследовании астмы и аллергии у детей, в котором участвовали 201 871 подросток, у испытуемых, которые употребляли фрукты, овощи, бобовые и орехи три раза в неделю или более, определялось меньшее значение ИМТ, по сравнению с теми, кто употреблял данные виды пищевых продуктов редко или время от времени [25].

Результаты нашего исследования также показали, что девочки-подростки, которые не перекусывают в течение дня, имеют более высокие значения доли жира в организме, чем те, у которых перекус является частью образа жизни. Результаты многочисленных исследований указывают, что перекусы могут привести к повышению суточного потребления калорий вместо его снижения. Тем не менее перекус является неотъемлемым компонентом общей стратегии по снижению веса. Оптимальными вариантами перекуса в рамках такой стратегии являются продукты питания, богатые сложными углеводами, белками и волокном, поскольку они продляют ощущение сытости [26]. Согласно анализу данных китайского исследования здоровья и питания, собранных в 2006, 2009 и 2011 гг., наибольшее сокращение в z -параметре ИМТ (-2,1) было выявлено в верхнем терциле выборки детей с избыточным весом в возрасте 7–13 лет по такому параметру, как перекус ($P < 0,05$) [27].

В данном исследовании было обнаружено, что время, проведенное за компьютером или со смартфоном, являлось значимым предиктором доли жира в организме подростков. В нескольких исследованиях продемонстрировано, что непропорциональное использование электронных устройств, в особенности среди подростков, оказывает значительное влияние на физическое, физиологическое и социальное благополучие данной когорты населения [28]. В оценочном исследовании, проведенном на популяционной выборке близнецов из Финляндии ($n = 4098$) была обнаружена взаимосвязь между продолжительным временем, проведенным за домашним компьютером, и повышенным риском избыточного веса. Более того, наблюдался положительный линейный тренд между использо-

ванием смартфона и ИМТ, с бета-коэффициентом, равным 0,18 (95 % ДИ: 0,06–0,30) [29]. Результаты изучения 44 исследований при помощи мета-анализа двойного класса также показали, что вероятность ожирения или избыточного веса была в 1,27 раза выше для подростков с самым длинным периодом экранного времени ($P < 0,001$) [30].

За последние несколько десятилетий во Вьетнаме изменения способов передвижения по дороге в школу и обратно привели к большему распространению элементов малоподвижного образа жизни. В существующей литературе до сих пор нет единого мнения по поводу возможной взаимосвязи между способом передвижения по дороге в школу и обратно и составом тела. Метаанализ 13 исследований был сконцентрирован исключительно на взаимосвязи между активным способом передвижения по дороге в школу и обратно и индексом массы тела детей и подростков. Среди 13 исследований, вошедших в окончательную выборку, в трех была выявлена определенная взаимосвязь. В трех других была обнаружена частичная взаимосвязь, ограниченная определенными подгруппами или обусловленная социальными / географическими факторами. В семи исследованиях значимых корреляций выявлено не было [31].

Сильные стороны и ограничения. Некоторые факторы, связанные с долей жира в организме подростков, как было установлено в предыдущих исследованиях, например, уровень физической активности, не были включены в дерево решений. К результатам данного исследования стоит относиться с осторожностью, поскольку отсутствие данных переменных в модели совершенно не означает, что они не связаны с долей жира в организме. При интерпретации результатов данного исследования необходимо учитывать несколько ограничений. Во-первых, возможность экстраполяции результатов на другую этническую группу может оказаться проблематичной, поскольку база данных, использованная в исследовании, является характерной для подростков, живущих в Ханое. Во-вторых, пусть даже машинное обучение может обнаруживать взаимосвязи, оно может не установить причинную обусловленность. Тем не менее у данного исследования есть несколько сильных сторон. Во-первых, в нем предпринята попытка раскрыть сложные взаимосвязи между переменными-предикторами с помощью анализа, основанного на использовании дерева решений. Во-вторых, применение технологий машинного обучения дает возможность углубленного исследования факторов, связанных с долей жира в организме, с учетом многих нюансов и на основании реальных данных. Потенциально это может помочь

обнаружить нелинейные взаимосвязи, которые могут оказаться недоступными для обнаружения методами конвенциональной статистики.

Практическая значимость. Результаты данного исследования имеют большую практическую значимость. Прежде всего, они подчеркивают важность формирования здоровых пищевых привычек и стимулирования физической активности у подростков. Кроме того, родители и сотрудники школ могут использовать результаты данного исследования в программах воспитания. Для поддержания доли жира в организме детей на здоровом уровне рекомендуется потребление полноценных порций фруктов не реже трех раз в неделю, а также введение в рацион детей ежедневных перекусов полезными для здоровья продуктами в дополнение к трем основным приемам пищи. Помимо этого, детей следует поощрять добираться до школы активным способом и сокращать экранное время.

Выводы. В заключение хотелось бы отметить, что в данном исследовании были успешно идентифицированы шесть групп предикторов, связанных с долей жира в организме подростков, с использованием модели дерева решений. Важно и то, что подобная методика позволяет учитывать многочисленные переменные одновременно, что повышает прогностическую точность модели. Среди выделенных предикторов наиболее значимыми факторами, связанными с долей жира в организме, следует считать частоту потребления фруктов, привычки, связанные с перекусом, способ передвижения по дороге в школу и обратно и экранное время. Данное исследование проливает свет на сложные и разнообразные факторы, которые оказывают влияние на долю жира в организме вьетнамских подростков. Комбинация этих факторов, а также их взаимодействие с полом и стадией полового созревания могут обуславливать долю жира в организме вьетнамских подростков. Однако необходимо признать тот факт, что между данными предикторами могут существовать пока еще не выявленные взаимосвязи. Поэтому будущие исследования должны быть направлены на более глубокое изучение данных взаимосвязей, а также лежащих в их основе сложностей, существующих в этой многоаспектной области исследования.

Благодарность. Авторы выражают благодарность доцентам Буи Тхи Нхунг и Тран Куан Бхин, а также коллегам из Национального института питания и Национального университета образования Ханоя за их всеобъемлющую помощь и поддержку.

Финансирование. Исследование не имело спонсорской поддержки.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Список литературы

1. Performance of body mass index and percentage of body fat in predicting cardiometabolic risk factors in Thai adults / S. Vanavanan, P. Srisawasdi, M. Rochanawutanon, N. Kumproa, K. Kruthkul, M.H. Kroll // *Diabetes Metab. Syndr. Obes.* – 2018. – Vol. 11. – P. 241–253. DOI: 10.2147/DMSO.S167294
2. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies // *Lancet.* – 2004. – Vol. 363, № 9403. – P. 157–163. DOI: 10.1016/S0140-6736(03)15268-3
3. Mendez M.A., Popkin B.M. Globalization, urbanization and nutritional change in the developing world // *Journal of Agricultural and Development Economics.* – 2004. – Vol. 1, № 2. – P. 220–241.
4. Associations between diet, physical activity and body fat distribution: a cross sectional study in an Indian population / L. Bowen, A.E. Taylor, R. Sullivan, S. Ebrahim, S. Kinra, K.V.R. Krishna, B. Kulkarni, Y. Ben-Shlomo [et al.] // *BMC Public Health.* – 2015. – Vol. 15. – P. 281. DOI: 10.1186/s12889-015-1550-7
5. Indicator for success of obesity reduction programs in adolescents: body composition or body mass index? Evaluating a school-based health promotion project after 12 weeks of intervention / N. Kalantari, N.K. Mohammadi, S. Rafieifar, H. Eini-Zinab, A. Aminifard, H. Malmir, N. Ashoori, S. Abdi [et al.] // *Int. J. Prev. Med.* – 2017. – Vol. 8, № 73. – P. 128–132. DOI: 10.4103/ijpvm.IJPVM_306_16
6. BFP and its correlation with dietary pattern, physical activity and life-style factors in school going children of Mumbai, India / J. Madan, N. Gosavi, P. Vora, P. Kalra // *Obes. Metab. Res.* – 2014. – Vol. 1, № 1. – P. 14–19.
7. Dewi R.C., Rimawati N., Purbodjati P. Body mass index, physical activity, and physical fitness of adolescence // *J. Public Health Res.* – 2021. – Vol. 10, № 2. – P. 2230. DOI: 10.4081/jphr.2021.2230
8. Steele R.G., Van Allen J. The treatment of pediatric obesity: Bringing contexts and systems into focus // *Children's Health Care.* – 2011. – Vol. 40, № 3. – P. 171–178. DOI: 10.1080/02739615.2011.590384
9. Briggs D. The roles of managers in addressing Sustainable Development Goals and addressing the burden of chronic disease // *Asia Pacific Journal of Health Management.* – 2018. – Vol. 13, № 2. – P. 1–3. DOI: 10.24083/apjhm.v13i2.17
10. Alfaleh G., Huffman F.G., Li T., Vaccaro J.A. Child Obesity Prevention Intervention in Kuwaiti Summer Camps Targeting Health Behaviors in Nutrition, Physical Activity, and Screen Time // *Journal of Health Science and Medical Research.* – 2021. – Vol. 39, № 2. – P. 85–99. DOI: 10.31584/jhsmr.2020765
11. Does physical activity predict obesity – A machine learning and statistical method-based analysis / X. Cheng, S.-Y. Lin, J. Liu, S. Liu, J. Zhang, P. Nie, B.F. Fuemmel, Y. Wang, H. Xue // *Int. J. Environ. Res. Public Health.* – 2021. – Vol. 18, № 8. – P. 3966. DOI: 10.3390/ijerph18083966
12. A machine learning approach to short-term body weight prediction in a dietary intervention program / B. Oladapo, H. Tawfik, A. Palczewska, A. Gorbenko, A. Arne, J.A. Martinez, J.-M. Oppert, T.I.A. Sorensen // *Computational Science – ICCS 2020: 20th International Conference.* – The Netherlands, Amsterdam, June 3–5, 2020. – P. 441–455. DOI: 10.1007/978-3-030-50423-6_33
13. Association of machine learning-derived measures of body fat distribution with cardiometabolic diseases in > 40,000 individuals / S. Agrawal, M.D. Klarqvist, N. Diamant, T.L. Stanley, P.T. Ellinor, N.N. Mehta, A. Philippakis, K. Ng [et al.] // *MedRxiv.* – 2021. DOI: 10.1101/2021.05.07.21256854
14. Scikit-learn: Machine learning in Python / F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Muller [et al.] // *Journal of Machine Learning Research.* – 2011. – Vol. 12. – P. 2825–2830.
15. Comparative study of regressor and classifier with decision tree using modern tools / J.S. Kushwah, A. Kumar, S. Patel, R. Soni, A. Gawande, S. Gupta // *Materials Today Proceedings.* – 2022. – Vol. 56, № 6. – P. 3571–3576. DOI: 10.1016/j.matpr.2021.11.635
16. Assessment of body composition in health and disease using bioelectrical impedance analysis (BIA) and dual energy X-ray absorptiometry (DXA): a critical overview / M. Marra, R. Sammarco, A. De Lorenzo, F. Iellamo, M. Siervo, A. Pietrobelli, L.M. Donini, L. Santarpia [et al.] // *Contrast Media Mol. Imaging.* – 2019. – Vol. 2019. – P. 3548284. DOI: 10.1155/2019/3548284
17. Performance of bioelectrical impedance analysis compared to dual X-ray absorptiometry (DXA) in Veterans with COPD / P.N. Cruz Rivera, R.L. Goldstein, M. Polak, A.A. Lazzari, M.L. Moy, E.S. Wan // *Sci. Rep.* – 2022. – Vol. 12, № 1. – P. 1946–1953. DOI: 10.1038/s41598-022-05887-4
18. Comparison of body composition assessment by DXA and BIA according to the body mass index: A retrospective study on 3655 measures / N. Achamrah, G. Colange, J. Delay, A. Rimbart, V. Folope, A. Petit, S. Grigioni, P. Déchelotte, M. Coëffier // *PLoS One.* – 2018. – Vol. 13, № 7. – P. e0200465. DOI: 10.1371/journal.pone.0200465
19. Comparison of two bioelectrical impedance analysis devices with dual energy X-ray absorptiometry and magnetic resonance imaging in the estimation of body composition / J.-G. Wang, Y. Zhang, H.-E. Chen, Y. Li, X.-G. Cheng, L. Xu, Z. Guo, X.-S. Zhao [et al.] // *J. Strength Cond. Res.* – 2013. – Vol. 27, № 1. – P. 236–243. DOI: 10.1519/JSC.0b013e31824f2040
20. Risk factors for obesity among children aged 24 to 80 months in Korea: A decision tree analysis / I. Lee, K.-S. Bang, H. Moon, J. Kim // *J. Pediatr. Nurs.* – 2019. – Vol. 46. – P. e15–e23. DOI: 10.1016/j.pedn.2019.02.004
21. A systematic literature review on obesity: Understanding the causes & consequences of obesity and reviewing various machine learning approaches used to predict obesity / M. Safaei, E.A. Sundararajan, M. Driss, W. Boulila, A. Shapi'i // *Comput. Biol. Med.* – 2021. – Vol. 136. – P. 104754–104780. DOI: 10.1016/j.combiomed.2021.104754
22. A review of machine learning in obesity / K. DeGregory, P. Kuiper, T. DeSilvio, J. Pleuss, R. Miller, J. Roginski, C.B. Fisher, D. Harness [et al.] // *Obes. Rev.* – 2018. – Vol. 19, № 5. – P. 668–685. DOI: 10.1111/obr.12667
23. Comparison of overweight and obesity prevalence in school-aged youth from 34 countries and their relationships with physical activity and dietary patterns / I. Janssen, P.T. Katzmarzyk, W.F. Boyce, C. Vereecken, C. Mulvihill, C. Roberts, C. Currie, W. Pickett [et al.] // *Obes. Rev.* – 2005. – Vol. 6, № 2. – P. 123–132. DOI: 10.1111/j.1467-789X.2005.00176.x
24. Association between fruit and vegetable intake and change in body mass index among a large sample of children and adolescents in the United States / A.E. Field, M.W. Gillman, B. Rosner, H.R. Rockett, G.A. Colditz // *Int. J. Obes. Relat. Disord.* – 2003. – Vol. 27, № 7. – P. 821–826. DOI: 10.1038/sj.ijo.0802297
25. Association between frequency of consumption of fruit, vegetables, nuts and pulses and BMI: analyses of the International Study of Asthma and Allergies in Childhood (ISAAC) / C.R. Wall, A.W. Stewart, R.J. Hancox, R. Murphy,

I. Braithwaite, R. Beasley, E.A. Mitchell, ISAAC Phase Three Study Group // *Nutrients*. – 2018. – Vol. 10, № 3. – P. 316–326. DOI: 10.3390/nu10030316

26. Snacking behaviors, diet quality, and body mass index in a community sample of working adults / T.L. Barnes, S.A. French, L.J. Harnack, N.R. Mitchell, J. Wolfson // *J. Acad. Nutr. Diet.* – 2015. – Vol. 115, № 7. – P. 1117–1123. DOI: 10.1016/j.jand.2015.01.009

27. Taillie L.S., Wang D., Popkin B.M. Snacking is longitudinally associated with declines in body mass index z scores for overweight children, but increases for underweight children // *J. Nutr.* – 2016. – Vol. 146, № 6. – P. 1268–1275. DOI: 10.3945/jn.115.226803

28. Kwong C.K.Y., Fong B.Y. Promotion of appropriate use of electronic devices among Hong Kong adolescents // *Asia Pacific Journal of Health Management*. – 2019. – Vol. 14, № 1. – P. 36–41. DOI: 10.24083/apjhm.v14i1.199

29. Are computer and cell phone use associated with body mass index and overweight? A population study among twin adolescents / H.-R. Lajunen, A. Keski-Rahkonen, L. Pulkkinen, R.J. Rose, A. Rissanen, J. Kaprio // *BMC Public Health*. – 2007. – Vol. 7. – P. 24. DOI: 10.1186/1471-2458-7-24

30. Screen time increases overweight and obesity risk among adolescents: a systematic review and dose-response meta-analysis / P. Haghjoo, G. Siri, E. Soleimani, M.A. Farhangi, S. Alesaeidi // *BMC Prim. Care*. – 2022. – Vol. 23, № 1. – P. 161. DOI: 10.1186/s12875-022-01761-4

31. Masoumi H.E. Active Transport to School and Children's Body Weight: A Systematic Review // *TeMA – Journal of Land Use, Mobility and Environment*. – 2017. – Vol. 10, № 10. – P. 95–110. DOI: 10.6092/1970-9870/4088

Технологии машинного обучения при определении факторов, связанных с долей жира в организме вьетнамских подростков / Нгуен Тхи Тхонг Хан, Ле Тхи Туйет, Нгуен Тхи Трунг Тху, До Тхи Нху Транг, Дуонг Тхи Анх Дао, Ле Тхи Тхуй Дунг, Данг Ксуан Тхо // Анализ риска здоровью. – 2024. – № 1. – С. 158–168. DOI: 10.21668/health.risk/2024.1.16

UDC 612.3-004.42

DOI: 10.21668/health.risk/2024.1.16.eng



Research article

IDENTIFYING THE FACTORS RELATED TO BODY FAT PERCENTAGE AMONG VIETNAMESE ADOLESCENTS USING MACHINE LEARNING TECHNIQUES

Nguyen Thi Hong Hanh¹, Le Thi Tuyet¹, Nguyen Thi Trung Thu¹, Do Thi Nhu Trang¹, Duong Thi Anh Dao¹, Le Thi Thuy Dung², Dang Xuan Tho³

¹Hanoi National University of Education, 136 Xuan Thuy St., Hanoi, Vietnam

²Binh Duong General Hospital, 5 Pham Ngoc Thach, Hiep Thanh, Binh Duong, Vietnam

³Academy of Policy and Development, Hoai Duc district, Hanoi, Vietnam

The aim of this study was to investigate the factors influencing Body Fat Percentage (BFP) among Vietnamese adolescents aged 11 to 15 employing machine learning techniques for predictive analysis.

© Nguyen Thi Hong Hanh, Le Thi Tuyet, Nguyen Thi Trung Thu, Do Thi Nhu Trang, Duong Thi Anh Dao, Le Thi Thuy Dung, Dang Xuan Tho, 2024

Nguyen Thi Hong Hanh – PhD., Lecturer at the Department of Biology (e-mail: hanhnhth@hnue.edu.vn; tel.: +84 (902) 80-11-83; ORCID: <https://orcid.org/0000-0002-1392-8081>).

Le Thi Tuyet – PhD., Lecturer for Human and Animal Physiology at the Department of Biology (e-mail: tuyetlt@hnue.edu.vn; tel.: +84 (968) 79-55-55; ORCID: <https://orcid.org/0000-0002-3308-5886>).

Nguyen Thi Trung Thu – PhD., Lecturer at the Department of Biology (e-mail: trungthu@hnue.edu.vn; tel.: +84 (983) 42-09-85; ORCID: <https://orcid.org/0000-0002-9800-2287>).

Do Thi Nhu Trang – M.S., Assistant Lecturer at the Department of Biology (e-mail: trangdtn@hnue.edu.vn; tel.: +84 (986) 70-38-10; ORCID: <https://orcid.org/0009-0009-5461-4554>).

Duong Thi Anh Dao – PhD., Associate Professor, Lecturer at the Department of Biology (e-mail: daodang-duc@gmail.com; tel.: +84 (886) 81-67-66; ORCID: <https://orcid.org/0009-0003-4407-8942>).

Le Thi Thuy Dung – PhD., M.D. at the Department of Pediatrics (e-mail: letono2002@gmail.com; tel.: +84 (987) 00-89-14; ORCID: <https://orcid.org/0000-0002-8855-7801>).

Dang Xuan Tho – PhD., Main Lecturer at the Faculty of Digital Economics (tel.: +84 (912) 62-93-83; ORCID: <https://orcid.org/0000-0002-7654-5942>).

A total of 1,208 adolescents, comprising 598 boys and 610 girls, drawn from nine junior high schools in Vietnam's capital, were enrolled in the study. Body composition measurements were conducted using the HBF 375 (Omron) device by Bioelectrical Impedance Analysis method. The study questionnaire, initially validated by The National Institute of Nutrition, encompassed inquiries related to dietary behaviors, meal frequencies, physical activities, sedentary habits, and nutritional knowledge. A machine learning methodology employing a decision tree algorithm was employed to discern the primary determinants most significantly correlated with BFP.

This study successfully identified six distinct predictor groups associated with BFP among adolescents, leveraging the decision tree model, with Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) values of 4.80 and 3.80, respectively. Among these predictors, frequency of fruit consumption, snacking habits, mode of transportation to school, and screen time (computer and/or cell phone usage) emerged as the most influential factors linked to BFP in adolescents. The combination of these factors and interactions with gender and pubertal status can BFP in Vietnamese adolescents.

This research sheds light on the complex and diverse factors impacting BFP in Vietnamese adolescents. This study's results underscore the practical importance of promoting healthy eating and exercise habits among adolescents, offering valuable insights for parents and schools to enhance their childcare strategies.

Keywords: machine learning, body fat percentage, predictability, influencing factors, eating habits, physical activity, Vietnamese adolescents, the decision tree.

References

1. Vanavanan S., Srisawasdi P., Rochanawutanon M., Kumproa N., Kruthkul K., Kroll M.H. Performance of body mass index and percentage of body fat in predicting cardiometabolic risk factors in Thai adults. *Diabetes Metab. Syndr. Obes.*, 2018, vol. 11, pp. 241–253. DOI: 10.2147/DMSO.S167294
2. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet*, 2004, vol. 363, no. 9403, pp. 157–163. DOI: 10.1016/S0140-6736(03)15268-3
3. Mendez M.A., Popkin B.M. Globalization, urbanization and nutritional change in the developing world. *Journal of Agricultural and Development Economics*, 2004, vol. 1, no. 2, pp. 220–241.
4. Bowen L., Taylor A.E., Sullivan R., Ebrahim S., Kinra S., Krishna K.V.R., Kulkarni B., Ben-Shlomo Y. [et al.]. Associations between diet, physical activity and body fat distribution: a cross sectional study in an Indian population. *BMC Public Health*, 2015, vol. 15, pp. 281. DOI: 10.1186/s12889-015-1550-7
5. Kalantari N., Mohammadi N.K., Rafieifar S., Eini-Zinab H., Aminifard A., Malmir H., Ashoori N., Abdi S. [et al.]. Indicator for success of obesity reduction programs in adolescents: body composition or body mass index? Evaluating a school-based health promotion project after 12 weeks of intervention. *Int. J. Prev. Med.*, 2017, vol. 8, no. 73, pp. 128–132. DOI: 10.4103/ijpvm.IJPVM_306_16
6. Madan J., Gosavi N., Vora P., Kalra P. BFP and its correlation with dietary pattern, physical activity and life-style factors in school going children of Mumbai, India. *Obes. Metab. Res.*, 2014, vol. 1, no. 1, pp. 14–19.
7. Dewi R.C., Rimawati N., Purbodjati P. Body mass index, physical activity, and physical fitness of adolescence. *J. Public Health Res.*, 2021, vol. 10, no. 2, pp. 2230. DOI: 10.4081/jphr.2021.2230
8. Steele R.G., Van Allen J. The treatment of pediatric obesity: Bringing contexts and systems into focus. *Children's Health Care*, 2011, vol. 40, no. 3, pp. 171–178. DOI: 10.1080/02739615.2011.590384
9. Briggs D. The roles of managers in addressing Sustainable Development Goals and addressing the burden of chronic disease. *Asia Pacific Journal of Health Management*, 2018, vol. 13, no. 2, pp. 1–3. DOI: 10.24083/apjhm.v13i2.17
10. Alfaleh G., Huffman F.G., Li T., Vaccaro J.A. Child Obesity Prevention Intervention in Kuwaiti Summer Camps Targeting Health Behaviors in Nutrition, Physical Activity, and Screen Time. *Journal of Health Science and Medical Research*, 2021, vol. 39, no. 2, pp. 85–99. DOI: 10.31584/jhsmr.2020765
11. Cheng X., Lin S.-Y., Liu J., Liu S., Zhang J., Nie P., Fuemmeler B.F., Wang Y., Xue H. Does physical activity predict obesity – A machine learning and statistical method-based analysis. *Int. J. Environ. Res. Public Health*, 2021, vol. 18, no. 8, pp. 3966. DOI: 10.3390/ijerph18083966
12. Oladapo B., Tawfik H., Palczewska A., Gorbenko A., Arne A., Martinez J.A., Oppert J.-M., Sørensen T.I.A. A machine learning approach to short-term body weight prediction in a dietary intervention program. *Computational Science – ICCS 2020: 20th International Conference*. Amsterdam, The Netherlands, June 3–5, 2020, pp. 441–455. DOI: 10.1007/978-3-030-50423-6_33
13. Agrawal S., Klarqvist M.D., Diamant N., Stanley T.L., Ellinor P.T., Mehta N.N., Philippakis A., Ng K. [et al.]. Association of machine learning-derived measures of body fat distribution with cardiometabolic diseases in > 40,000 individuals. *MedRxiv*, 2021. DOI: 10.1101/2021.05.07.21256854
14. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Muller A. [et al.]. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 2011, vol. 12, pp. 2825–2830.
15. Kushwah J.S., Kumar A., Patel S., Soni R., Gawande A., Gupta S. Comparative study of regressor and classifier with decision tree using modern tools. *Materials Today Proceedings*, 2022 vol. 56, no. 6, pp. 3571–3576. DOI: 10.1016/j.matpr.2021.11.635
16. Marra M., Sammarco R., De Lorenzo A., Iellamo F., Siervo M., Pietrobelli A., Donini L.M., Santarpia L. [et al.]. Assessment of body composition in health and disease using bioelectrical impedance analysis (BIA) and dual energy X-ray absorptiometry (DXA): a critical overview. *Contrast Media Mol. Imaging*, 2019, vol. 2019, pp. 3548284. DOI: 10.1155/2019/3548284
17. Cruz Rivera P.N., Goldstein R.L., Polak M., Lazzari A.A., Moy M.L., Wan E.S. Performance of bioelectrical impedance analysis compared to dual X-ray absorptiometry (DXA) in Veterans with COPD. *Sci. Rep.*, 2022, vol. 12, no. 1, pp. 1946–1953. DOI: 10.1038/s41598-022-05887-4
18. Achamrah N., Colange G., Delay J., Rimbart A., Folope V., Petit A., Grigioni S., Déchelotte P., Coëffier M. Comparison of body composition assessment by DXA and BIA according to the body mass index: A retrospective study on 3655 measures. *PLoS One*, 2018, vol. 13, no. 7, pp. e0200465. DOI: 10.1371/journal.pone.0200465

19. Wang J.-G., Zhang Y., Chen H.-E., Li Y., Cheng X.-G., Xu L., Guo Z., Zhao X.-S. [et al.]. Comparison of two bio-electrical impedance analysis devices with dual energy X-ray absorptiometry and magnetic resonance imaging in the estimation of body composition. *J. Strength Cond. Res.*, 2013, vol. 27, no. 1, pp. 236–243. DOI: 10.1519/JSC.0b013e31824f2040
20. Lee I., Bang K.-S., Moon H., Kim J. Risk factors for obesity among children aged 24 to 80 months in Korea: A decision tree analysis. *J. Pediatr. Nurs.*, 2019, vol. 46, pp. e15–e23. DOI: 10.1016/j.pedn.2019.02.004
21. Safaei M., Sundararajan E.A., Driss M., Boulila W., Shapi'i A. A systematic literature review on obesity: Understanding the causes & consequences of obesity and reviewing various machine learning approaches used to predict obesity. *Comput. Biol. Med.*, 2021, vol. 136, pp. 104754–104780. DOI: 10.1016/j.combiomed.2021.104754
22. DeGregory K., Kuiper P., DeSilvio T., Pleuss J., Miller R., Roginski J., Fisher C.B., Harness D. [et al.]. A review of machine learning in obesity. *Obes. Rev.*, 2018, vol. 19, no. 5, pp. 668–685. DOI: 10.1111/obr.12667
23. Janssen I., Katzmarzyk P.T., Boyce W.F., Vereecken C., Mulvihill C., Roberts C., Currie C., Pickett W. [et al.]. Comparison of overweight and obesity prevalence in school-aged youth from 34 countries and their relationships with physical activity and dietary patterns. *Obes. Rev.*, 2005, vol. 6, no. 2, pp. 123–132. DOI: 10.1111/j.1467-789X.2005.00176.x
24. Field A.E., Gillman M.W., Rosner B., Rockett H.R., Colditz G.A. Association between fruit and vegetable intake and change in body mass index among a large sample of children and adolescents in the United States. *Int. J. Obes. Relat. Disord.*, 2003, vol. 27, no. 7, pp. 821–826. DOI: 10.1038/sj.ijo.0802297
25. Wall C.R., Stewart A.W., Hancox R.J., Murphy R., Braithwaite I., Beasley R., Mitchell E.A., ISAAC Phase Three Study Group. Association between frequency of consumption of fruit, vegetables, nuts and pulses and BMI: analyses of the International Study of Asthma and Allergies in Childhood (ISAAC). *Nutrients*, 2018, vol. 10, no. 3, pp. 316–326. DOI: 10.3390/nu10030316
26. Barnes T.L., French S.A., Harnack L.J., Mitchell N.R., Wolfson J. Snacking behaviors, diet quality, and body mass index in a community sample of working adults. *J. Acad. Nutr. Diet.*, 2015, vol. 115, no. 7, pp. 1117–1123. DOI: 10.1016/j.jand.2015.01.009
27. Taillie L.S., Wang D., Popkin B.M. Snacking is longitudinally associated with declines in body mass index z scores for overweight children, but increases for underweight children. *J. Nutr.*, 2016, vol. 146, no. 6, pp. 1268–1275. DOI: 10.3945/jn.115.226803
28. Kwong C.K.Y., Fong B.Y. Promotion of appropriate use of electronic devices among Hong Kong adolescents. *Asia Pacific Journal of Health Management*, 2019, vol. 14, no. 1, pp. 36–41. DOI: 10.24083/apjhm.v14i1.199
29. Lajunen H.-R., Keski-Rahkonen A., Pulkkinen L., Rose R.J., Rissanen A., Kaprio J. Are computer and cell phone use associated with body mass index and overweight? A population study among twin adolescents. *BMC Public Health*, 2007, vol. 7, pp. 24. DOI: 10.1186/1471-2458-7-24
30. Haghjoo P., Siri G., Soleimani E., Farhangi M.A., Alesaeidi S. Screen time increases overweight and obesity risk among adolescents: a systematic review and dose-response meta-analysis. *BMC Prim. Care*, 2022, vol. 23, no. 1, pp. 161. DOI: 10.1186/s12875-022-01761-4
31. Masoumi H.E. Active Transport to School and Children's Body Weight: A Systematic Review. *TeMA – Journal of Land Use, Mobility and Environment*, 2017, vol. 10, no. 10, pp. 95–110. DOI: 10.6092/1970-9870/4088

Nguyen Thi Hong Hanh, Le Thi Tuyet, Nguyen Thi Trung Thu, Do Thi Nhu Trang, Duong Thi Anh Dao, Le Thi Thuy Dung, Dang Xuan Tho. Identifying the factors related to body fat percentage among Vietnamese adolescents using machine learning techniques. *Health Risk Analysis*, 2024, no. 1, pp. 158–168. DOI: 10.21668/health.risk/2024.1.16.eng

Получена: 25.01.2024

Одобрена: 29.02.2024

Принята к публикации: 20.03.2024