

ПРОГНОЗИРОВАНИЕ ИНТЕНСИВНОСТИ ПОРАЖЕНИЯ ЗУБОВ КАРИЕСОМ ВОСПИТАННИКОВ ОБЩЕОБРАЗОВАТЕЛЬНЫХ ОРГАНИЗАЦИЙ МИНОБОРОНЫ РОССИИ С ПОМОЩЬЮ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

И.К. Солдатов^{1*}, Г.А. Гребнев¹, А.Н. Солдатов², А.С. Багненко¹

¹ ФГБВОУ ВО «Военно-медицинская академия им. С.М. Кирова» Минобороны России, Санкт-Петербург

² ФГАО УВО «Национальный исследовательский университет ИТМО», Санкт-Петербург

PREDICTION OF THE INTENSITY OF DENTAL CARIES AMONG STUDENTS OF EDUCATIONAL ESTABLISHMENTS OF THE MINISTRY OF DEFENSE OF THE RUSSIAN FEDERATION BY ALGORITHMS OF THE MACHINE LEARNING

I.K. Soldatov^{1*}, G.A. Grebnev¹, A.N. Soldatov², A.S. Bagnenko¹

¹ Military Medical Academy named after S.M. Kirov, Saint Petersburg, Russia

² ITMO University, Saint Petersburg, Russia

*E-mail: ivan-soldatov@mail.ru

Аннотация

Прогнозирование стоматологической заболеваемости является неотъемлемой частью и актуальной задачей при планировании лечебно-профилактической деятельности в рамках медицинского обеспечения воспитанников на протяжении всего периода обучения в образовательной организации. Разработка и внедрение прогностических моделей на основе алгоритмов машинного обучения обеспечивает более детальный персонализированный подход, снижает риск развития основных стоматологических заболеваний и повышает качество оказания медицинской помощи. **Цель исследования** – разработка модели прогнозирования интенсивности поражения зубов кариесом воспитанников общеобразовательных организаций Министерства обороны Российской Федерации (далее – Минобороны России) при помощи алгоритмов машинного обучения. **Материалы и методы.** В результате диспансеризации получены данные о состоянии общесоматического и стоматологического здоровья 3701 воспитанника в возрасте 10–18 лет, а также другие медицинские данные за предыдущие периоды обучения с помощью выкопировки из личных медицинских книжек и официальных форм медицинской отчетности (форма 13К/МЕД). Проанализированы результаты анкетирования 2242 кадетов об уровне знаний и навыках в области профилактической стоматологии. Все вышеуказанные сведения обобщены в созданные базы данных. Рассчитаны группы здоровья воспитанников, уровень гигиены полости рта, а также распространенность и интенсивность основных стоматологических заболеваний и других факторов, влияющих на достоверность прогноза. Прогнозирование индексов пораженности кариесом зубов для смешанного и постоянного прикусов проводили при помощи алгоритмов машинного обучения CatBoost, Decision Tree и Random Forest с определением эффективности, надежности и достоверности прогноза. **Результаты.** Выявлен высокий уровень распространенности (72.9%) и интенсивности КПУ + кп = 3.96 ± 0.06 («К» – постоянный кариозный зуб; «П» – постоянный пломбированный зуб; «У» – постоянный удаленный зуб; «к» – молочный кариозный зуб; «п» – молочный пломбированный зуб) в популяции воспитанников. При сравнении алгоритмов машинного обучения модель CatBoost показала наилучший прогноз (MSE = 1.53, R² = 0.82). Отмечены наиболее значительные признаки, вносящие наивысший вклад в прогностическую модель, – индекс КПУ за предыдущий год (66.3%), число интактных (пораженных) зубов (5.9%), потребность в неотложной медицинской помощи (5.1%).

Ключевые слова: заболеваемость, воспитанник, прогнозирование, кариес, индекс КПУ, машинное обучение.

Abstract

Forecasting dental morbidity is an integral part and urgent task when planning treatment and preventive activities as a part of medical care of students in educational organizations throughout the entire period of their studies. Development and implementation of predictive models based on machine learning algorithms provides a more detailed, personalized approach, helps to reduce the risk of major dental diseases and to improve the quality of medical care. **Purpose.** To develop a model for predicting the intensity of dental caries spread in students of general educational organizations of the Ministry of Defense of the Russian Federation using machine learning algorithms. **Materials and methods.** Due to the regular follow-up of 3 701 students, aged 10–18, the following information was obtained: their general somatic and dental health, as well as other medical information for the previous stages of their education taken from copies of their personal medical records and official medical reporting forms (form 13K/MED). 2 242 cadets were invited to take part in the survey to evaluate their knowledge and skills in preventive dentistry. All of the obtained information was summarized in the created databases. Groups of cadets' health, levels of oral hygiene, prevalence and intensity of major dental diseases and other factors influencing the prognosis reliability were calculated. Prediction of dental caries indices for mixed and permanent dentition was made using machine learning algorithms CatBoost, Decision Tree and Random Forest which could calculate the effectiveness, reliability and validity of the prognostic model. **Results.** A high level of prevalence (72.9%) and intensity of КПУ + кп = 3.96 ± 0.06 (where "K" is a permanent carious tooth; "P" is a permanent filled tooth; "Y" is a permanent extracted tooth; "k" is a milk carious tooth; "n" is a filled milk tooth) was revealed in students' population. When comparing machine learning algorithms, the CatBoost model showed the best

prediction ability (MSE = 1.53, R² = 0.82). The most significant signs that make the highest contribution to the prognostic model are the following: КПУ index for the previous year (66.3%); number of intact (affected) teeth (5.9%), need for emergency medical care (5.1%).

Keywords: morbidity, cadet, forecasting, tooth decay, DMFT index, machine learning.

Ссылка для цитирования: Солдатов И.К., Гребнев Г.А., Солдатов А.Н., Багненко А.С. Прогнозирование интенсивности поражения зубов кариесом воспитанников общеобразовательных организаций Минобороны России с помощью алгоритмов машинного обучения. Кремлевская медицина. Клинический вестник. 2024; 2: 58–63.

Введение

Здоровье детей и подростков является приоритетной задачей любого государства. Сохранение и укрепление стоматологического здоровья является важной составной частью жизни человека в целом [1].

Кариес зубов является вторым по распространенности заболеванием во всем мире и самым распространенным среди стоматологических заболеваний, достигая 100%, и не имеет тенденций к снижению [2]. Проводимые лечебно-профилактические мероприятия и достижения современной науки в последнее десятилетие не влияют на снижение стоматологической заболеваемости [3].

Индексная оценка кариозного процесса при обследовании детей и подростков позволяет более качественно оценивать распространенность и интенсивность основных стоматологических заболеваний, а также проводить мониторинг, определять нуждаемость в оказании медицинской помощи и анализировать ее эффективность [4].

В субъектах Российской Федерации под эгидой Всемирной организации здравоохранения (ВОЗ) трижды проводились эпидемиологические стоматологические обследования. Завершенные в 1999, 2009 и 2019 гг. обследования легли в основу ряда научных работ и подтверждают влияние различных факторов на развитие стоматологических заболеваний. Таким образом, разработка и внедрение профилактических мероприятий по сохранению и укреплению здоровья становятся особенно актуальными [5].

Прогнозирование основных стоматологических заболеваний во многом зависит от различных климатогеографических, социальных факторов, а также общесоматических и генетических предрасположенностей. В клинической практике уже применяются модели прогнозирования: Cariogram, CAMBRA, PreViser, NUS-CRA, CAT и CRAF, в которых анализируются от девяти до 14 факторов. Однако доказательство повышения эффективности прогнозирования, профилактики и лечения кариеса зубов пока недостаточно [6].

Воспитанники общеобразовательных организаций Минобороны России являются обособленным ведомственным контингентом, отобранном из различных регионов, культур и социальных слоев; на них распространяются не только нормативные правовые документы Минздрава России, но и ведомственные документы Минобороны России.

Эпидемиологическое стоматологическое обследование воспитанников общеобразовательных организаций Минобороны России в соответствии с критериями и рекомендациями ВОЗ в рамках научно-исследовательской работы № VMA/01.01/04/1819/0040 шифр «Кадет» позволило получить данные о состоянии здоровья полости рта у воспитанников 10–18 лет на всем протяжении их обучения в училищах [7].

Цель исследования – разработка модели прогнозирования интенсивности поражения зубов кариесом воспитанников общеобразовательных организаций Минобороны России при помощи алгоритмов машинного обучения.

Материалы и методы

В ходе проведения ежегодной диспансеризации получены данные о состоянии общесоматического здоровья и зубочелюстной системы 3701 воспитанника (9550 осмотров-наблюдений) с первого по седьмой курс в возрасте 10–18 лет из девяти общеобразовательных организаций Минобороны России за 2013/14–2019/20 учебные годы.

Группу здоровья воспитанников определяли в соответствии с нормативными правовыми документами Минобороны России: первая – здоровые воспитанники с нормальным развитием и нормальным уровнем основных функций; вторая – имеющие функциональные или морфологические отклонения после острых заболеваний, часто болеющие воспитанники; третья – больные хроническими заболеваниями в состоянии компенсации с сохранением функциональных возможностей организма. Группы состояния здоровья воспитанников представлены в относительных величинах (%).

Осмотр полости рта проводили в условиях стоматологического кабинета медицинского пункта с использованием стандартного набора инструментов (зеркало и зонд), результаты оценки стоматологического статуса воспитанников регистрировали в модифицированной авторской карте, рекомендованной ВОЗ. Данные стоматологических осмотров за предыдущие годы внесены с помощью выкопировки из медицинской документации (медицинская книжка воспитанника – форма 2). В последующем проведена оцифровка регистрационных карт и внесение данных в созданный единый реестр стоматологического статуса воспитанника [7].

В дальнейшем производили расчет распространенности и интенсивности основных стоматологических заболеваний с проведением индексной оценки пораженности кариесом зубов (КПУ, КПУ + кп), а также отдельных элементов индекса («К» – постоянный кариозный зуб; «П» – постоянный пломбированный зуб; «У» – постоянный удаленный зуб; «к» – молочный кариозный зуб; «п» – молочный пломбированный зуб). Распространенность кариеса рассчитывали в процентах как отношение числа обследуемых воспитанников, имеющих признаки кариозного процесса, к общему числу обследуемых кадетов. Интенсивность кариеса определяли как сумму клинических признаков кариозного процесса у каждого воспитанника как индивидуально, так и в возрастных группах, а также среднее по образовательной организации в целом.

Также определяли интенсивность зубного налета на зубах у воспитанников при помощи упрощенного индекса гигиены полости рта (ИГР-У). Изучали состояние шести ключевых зубов (международное обозначение обследуемых зубов: вестибулярная поверхность – 1.1, 3.1; щечная поверхность – 1.6, 2.6; язычная поверхность – 3.6, 4.6). Индекс оценивали в баллах от 0 до 3.

Выявляли особенности распространенности и интенсивности кариеса в различных возрастных группах воспитанников с разным уровнем гигиены полости рта.

После агрегации данных из единого реестра стоматологического статуса проведен перенос информации в базу данных и устранены опечатки, методом «средних» заполнены недостающие значения.

Результаты исследования обрабатывали с помощью пакета стандартных компьютерных программ для статистического анализа Excel Statistica for Windows 6.0. Определяли показатели вариационного анализа: среднее арифметическое (M), среднее квадратичное отклонение (Sx), ошибку среднего арифметического (m). Достоверность различий исследуемых выборочных данных определяли при помощи критерия Стьюдента (t). Различия считались значимыми при $p < 0.05$. Анализ прогнозирования проводили при помощи алгоритма, написанного на языке Python.

На основе сформированного датасета производили расчет спрогнозированного индекса КПУ при помощи различных алгоритмов машинного обучения с определением эффективности, надежности и достоверности прогноза.

Для прогнозирования интенсивности поражения кариесом рассмотрены два подхода машинного обучения: классификация и регрессия. В задаче классификации предсказывали факт увеличения индекса КПУ при помощи алгоритмов CatBoost Classifier, Decision Tree Classifier, Random Forest Classifier и KNeighbors Classifier. Данный вариант предсказания актуален только для пациента, так как отсутствуют данные количественного прироста для планирования лечебно-профилактической помощи, важные для врача и характеризующие степень активности кариеса.

В задаче регрессии предсказывали значение индекса КПУ на предстоящий год при помощи алгоритмов CatBoost, Random Forest и Gradient Boosting Decision Trees (GBDT).

CatBoost – алгоритм градиентного бустинга, который может использоваться для задач классификации, регрессии и ранжирования. Одним из главных преимуществ данного алгоритма является его способность обрабатывать категориальные переменные без необходимости вручную кодировать их. В основу заложен метод Ordered Boosting для обработки категориальных переменных, что позволяет автоматически обрабатывать данные [8].

Random Forest – это алгоритм машинного обучения, используемый для решения задач классификации и регрессии. Он основывается на концепции «ансамбля обучения», то есть объединяет несколько простых моделей для повышения точности прогнозов. Основным преимуществом Random Forest является то, что он позволяет интерпретировать результаты прогнозирования, а также обеспечивает высокую эффективность этого процесса [9].

GBDT – еще один популярный алгоритм машинного обучения. Он формирует ансамбль деревьев решений, используя градиентный бустинг. Алгоритм предназначен для решения задач как классификации, так и регрессии, и его эффективность подтверждена результатами множества исследований. GBDT является одним из наиболее популярных алгоритмов машинного обучения благодаря способности работать с большим объемом данных различных типов. К преимуществам данного алгоритма можно отнести особенность использования градиентного бустинга для улучшения производительности модели и создания ансамбля деревьев решений, где каждое последующее дерево решений исправляет ошибки предыдущего дерева, что, в свою очередь, позволяет GBDT создавать более точные модели, чем простые деревья решений [10].

Для оценки качества регрессионных моделей использовали следующие метрики:

- Mean Squared Error (MSE) – это среднее значение квадрата разности между прогнозируемыми и фактическими значениями. Чем меньше значение MSE, тем лучше модель:

$$MSE = \frac{1}{n} \sum_{i=1}^n (a(x_i) - y_i)^2;$$

- Mean Absolute Error (MAE) – это среднее значение абсолютной разности между прогнозируемыми и фактическими значениями. MAE также используется для измерения точности модели, и чем меньше значение MAE, тем лучше модель:

$$MAE = \frac{1}{n} \sum_{i=1}^n |a(x_i) - y_i|;$$

- R-квадрат (R^2) – это коэффициент детерминации, который показывает, насколько хорошо модель соответствует данным. R^2 может принимать значения от 0 до 1, где 1 означает, что модель идеально соответствует данным:

$$R^2 = 1 - \frac{\sum_{i=1}^n (a(x_i) - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}.$$

Результаты и обсуждение

В результате проведенного эпидемиологического стоматологического обследования воспитанников общеобразовательных организаций Минобороны России распространенность кариеса зубов в популяции среди воспитанников составила 72.9%. В возрастных когортах с 10 до 18 лет имеется закономерный рост с 52.9 до 88.8% пораженности кариесом, что не противоречит общероссийским данным.

Интенсивность кариеса среди воспитанников – индекс КПУ + кп = 3.96 ± 0.06 . В возрастных когортах с 10 до 18 лет также имеется закономерный рост индекса КПУ с 1.91 до 5.41 соответственно. Полученные данные согласуются с общероссийскими показателями, не превышая их, но выше среднего по критериям ВОЗ.

Проведенный анализ интенсивности кариеса у воспитанников с разными зубочелюстными аномалиями показал, что у лиц со скученностью зубов наблюдался самый высокий уровень интенсивности кариеса – 5.31 ± 0.36 , и наоборот: у лиц с наличием тремы (диагностемы) КПУ составил 2.62 ± 0.73 ; при мезиальном и перекрестном типах прикуса КПУ – 4.24 ± 0.55 и 4.64 ± 0.41 соответственно, а самым низким КПУ оказался при глубоком типе прикуса – 3.43 ± 0.18 единицы.

Проанализировано значение индекса при проведении и отсутствии профилактических мероприятий до поступления в училище. Так, у лиц, которым не была проведена в школе профилактика стоматологических заболеваний до поступления в училище, средний индекс КПУ за популяцию составил 4.01 ± 0.07 , а у лиц, кому профилактика была проведена, – 3.32 ± 0.07 . Так, у лиц, подвергающихся регулярным профосмотрам в школе с целью проведения своевременных лечебно-профилактических мероприятий, индекс КПУ составил 3.42 ± 0.09 , а у лиц, которым преподавали уроки гигиены полости рта, – 2.08 ± 0.19 .

Проведено сравнение индекса с различным состоянием уровня гигиены полости рта. Среди детей, у которых гигиенический индекс (ГИ) = 0, пораженность составила 3.77 ± 0.06 , ГИ = 1 – 3.46 ± 0.10 , ГИ = 2 – 3.77 ± 0.26 , и при ГИ = 3 индекс КПУ был наивысшим – 5.49 ± 0.48 единицы.

Также проведено сравнение индекса и в группах с разным состоянием физического здоровья воспитанников. У лиц первой группы (56.6%) воспитанников индекс КПУ составил -3.06 ± 0.07 , второй (39.2%) – 3.49 ± 0.08 и третьей группы (4.2%) – 3.93 ± 0.23 . Таким образом, можно предположить, что общесоматическая патология также негативно сказывается на интенсивности кариозного процесса у воспитанников.

В дальнейшем для построения и обучения предсказательных моделей провели закладку вышеуказанных данных и признаков, влияющих на прогноз. Так, в объединенную группу признаков вошли 16 признаков из карт эпидемиологического стоматологического обследования воспитанников, один признак из анкет социологического обследования и ноль признаков из данных учетно-отчетной документации формы 13К/МЕД.

С целью оценки качества моделей и выбора наилучших значений гиперпараметров набор данных разделен на две группы: обучающий (80% выборки данных) и тестовый (20% выборки данных). Далее обучающий набор был разделен на две подгруппы: обучающий (80% обучающего набора) и проверочный (20% обучающего набора).

В настройке гиперпараметров модели машинного обучения использовали алгоритм подбора гиперпараметров по сетке (grid search) [11]. Данный алгоритм позволил перебрать все возможные комбинации гиперпараметров из заданного диапазона значений и выбрать те, которые дают наилучший результат (табл. 1).

После отбора признаков и обучения моделей на выбранных алгоритмах машинного обучения оценки качества MSE, MAE, R^2 имеют решающее значение для выбора подходящего алгоритма для итоговой модели прогнозирования. Лучший результат показала модель на основе CatBoost, значение MSE составило 1.52 с коэффициентом детерминации – 0.82 (табл. 2).

Графическое отображение зависимости фактического индекса КПУ от предсказанного различными моделями представлено на рис. 1.

Таблица 1

Результаты значений подобранных гиперпараметров

Модель	Параметр	Значение
CatBoost	iterations	358
	l2_leaf_reg	3
	learning_rate	0.06
	depth	6
Random Forest	max_depth	10
	min_samples_split	10
	n_estimators	150
GBDT	n_estimators	50
	max_depth	10
	min_samples_leaf	4
	learning_rate	0.05
	subsample	0.5
	max_features	auto

Таблица 2

Результаты значений метрик выбранных моделей

Модель	MSE	MAE	R^2
CatBoost	1.52	0.86	0.82
GBDT	1.56	0.88	0.81
Random Forest	1.63	0.90	0.80

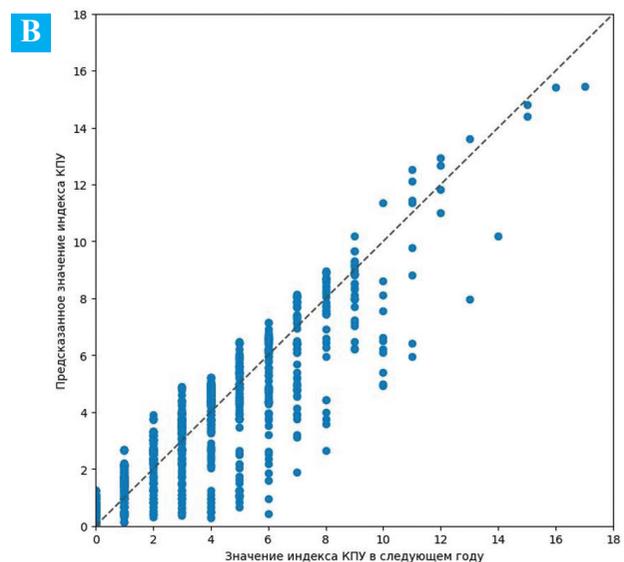
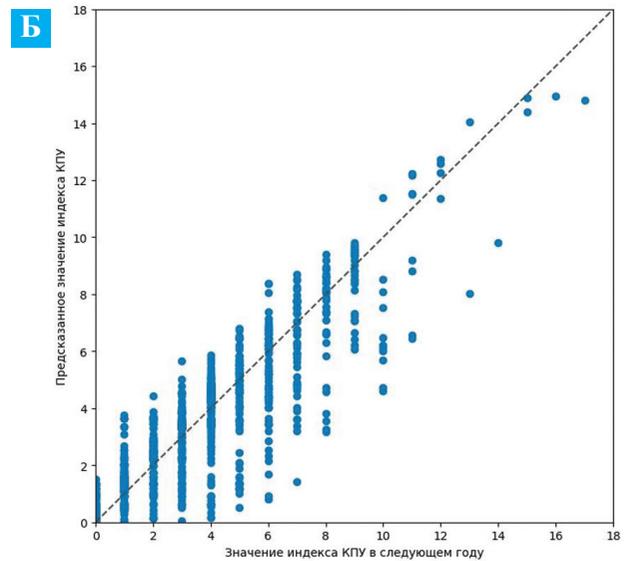
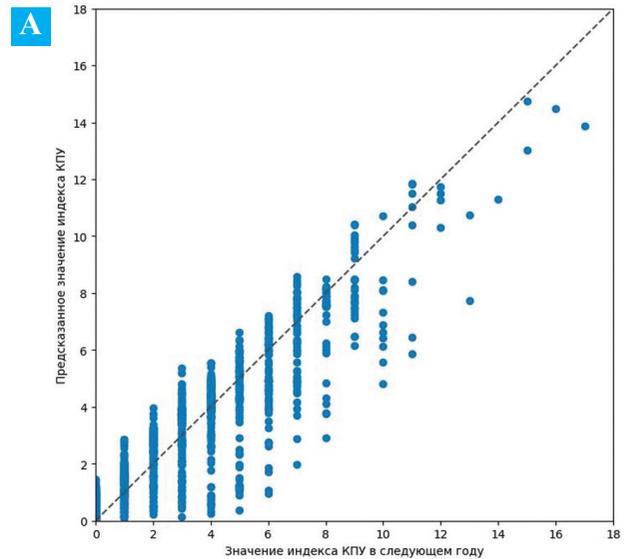


Рис. 1. Зависимость индекса КПУ на предстоящий год с предсказанными значениями КПУ моделями алгоритмов машинного обучения: А – CatBoost; Б – Decision Trees; В – Random Forest

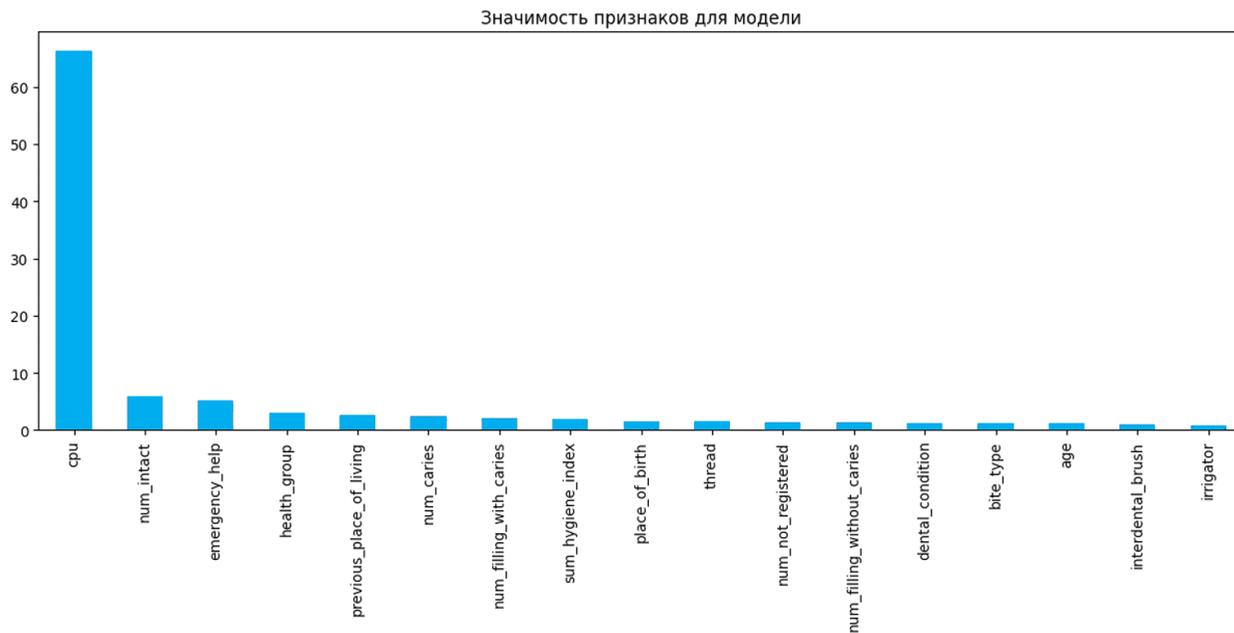


Рис. 2. Распределение признаков по значимости вклада в предсказательной модели, %

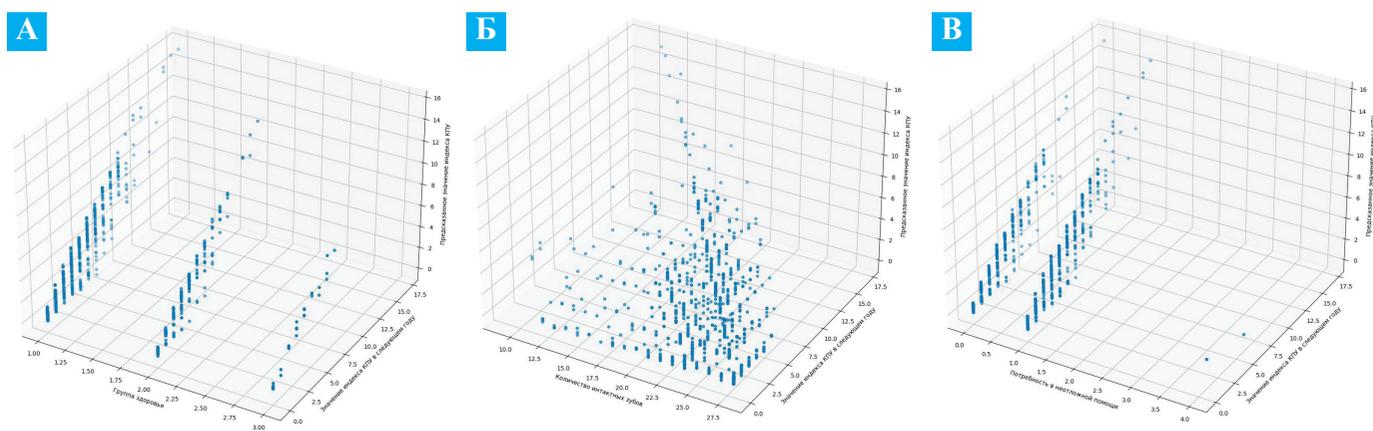


Рис. 3. Зависимость предсказанного и фактического индексов КПУ от основных доминирующих признаков: А – группы здоровья; Б – число здоровых зубов; В – медицинская помощь

В последующем для проведения расчетов прогнозирования использовали модель CatBoost как самую достоверную. При обучении модели выявили, что наиболее значимыми признаками для прогнозирования индекса КПУ были индекс КПУ за предыдущий год – 66.3%; число интактных зубов – 5.9%, потребность в неотложной медицинской помощи – 5.1%, группа состояния соматического здоровья – 3.0%; предыдущее место жительства до поступления в училище – 2.6%. Данные признаки вносят самый существенный вклад в предсказательную модель. Значимость всех признаков для модели представлена на рис. 2.

В дальнейшем провели сравнение предсказанного и реального индексов КПУ. Зависимость предсказаний представлена на рис. 3.

Таким образом, после обучения модели и интерпретации результатов выявлено, что наиболее значимые признаки, такие как число здоровых и кариозных зубов,

а также индекс КПУ за предыдущий год вносят существенный вклад в предсказательную модель. Напротив, признак места жительства воспитанников не вносит существенного вклада в решение прогностической задачи данной модели.

Заключение

Результаты эпидемиологического стоматологического обследования показали высокий уровень распространенности и интенсивности кариеса у воспитанников всех возрастов. Во всей популяции воспитанников 10–18 лет ($n = 3701$) распространенность кариеса составила 72.9% при интенсивности КПУ + кп = 3.96 ± 0.06 .

Индекс КПУ достоверно отличался у лиц, имеющих зубочелюстные аномалии, у лиц с различным уровнем гигиены полости рта, состоянием соматического здоровья, а также проводимыми лечебно-профилактическими мероприятиями.

Разработанная прогностическая модель на основе алгоритмов машинного обучения CatBoost имела самый высокий уровень точности предсказания ($R^2 = 0.82$). После обучения модели выявлено, что наиболее значимыми признаками, влияющими на уровень предсказания, являлись возраст воспитанников, число здоровых зубов и нуждаемость в неотложной медицинской помощи.

Данная модель реализуема в практической медицине и предназначена для врачей-стоматологов в целях принятия врачебных решений в рамках диспансерного наблюдения воспитанников и проведения лечебно-профилактических мероприятий.

Литература

1. Кузьмина Э.М. Роль деятельности ВОЗ в оценке здоровья полости рта населения на основе мониторинга стоматологической заболеваемости // Dental Forum. – 2015. – № 1 – С. 2–4. [Kuzmina E.M. The role of the WHO in assessing the oral health of the population based on monitoring dental morbidity // Dental Forum. – 2015. – No 1. – P. 2–4. In Russian].
2. Кузьмина Э.М. и др. Тенденции распространенности и интенсивности кариеса зубов среди населения России за 20-летний период // Dental Forum. – 2020. – Т. 78. – № 3. – С. 2–8. [Kuzmina E.M. et al. Trends in the prevalence and intensity of dental caries among the population of Russia over a 20-year period // Dental Forum. – 2020. – V. 78. – No 3. – P. 2–8. In Russian].
3. Волошина И.М. и др. Кариес зубов высокой степени риска и комплаентность пациента // Эндодонтия Today. – 2020. – Т. 18. – № 2. – С. 41–44. [Voloshina I.M. et al. High-risk dental caries and patient compliance // Endodontics Today. – 2020. – V. 18. – No 2. – P. 41–44. In Russian]. DOI: 10.36377/1683-2981-2020-18-2-41-44.
4. Авраамова О.Г. и др. Оценка эффективности лечебно-профилактических мероприятий у детей 10–12 лет с разной степенью интенсивности кариеса и уровнем резистентности эмали // Стоматология. – 2023. – Т. 102. – № 1. – С. 82–85. [Avraamova O.G. et al. Evaluation of the effectiveness of therapeutic and preventive measures in children aged 10–12 years with different degrees of caries intensity and enamel resistance // Stomatologia. – 2023. – V. 102. – No 1. – P. 82–85. In Russian]. DOI: 10.17116/stomat202310201182.
5. Кузьмина Э.М. и др. Стоматологическая заболеваемость населения России. – Москва: Практическая медицина. – 2019. – С. 303. [Kuzmina E.M. et al. Dental morbidity in the population of Russia. – Moscow: Practical medicine. – 2019. – P. 303. In Russian].
6. Модринская Ю.В. Модели оценки риска кариеса зубов: Стоматология вчера, сегодня, завтра: сборник научных трудов юбилейной научно-практической конференции с международным участием, посвященной 60-летию стоматологического факультета, Минск, 02–03 апреля 2020 г. / Под общей редакцией Т.Н. Тереховой. – Минск: Белорусский государственный медицинский университет, 2020. – С. 340–343. [Modrinskaya Y.V. Models for assessing the risk of dental caries: Dentistry Yesterday, Today, Tomorrow: collection of scientific papers of the jubilee scientific and practical conference with international participation dedicated to the 60th anniversary of the dental faculty, Minsk, April 02–03, 2020 / Edited by T.N. Terekhova. – Minsk: Belarusian State Medical University, 2020. – P. 340–343. In Russian].
7. Солдатов А.Н. и др. База данных состояния стоматологического здоровья воспитанников общеобразовательных организаций Минобороны России. Свидетельство о государственной регистрации базы данных № 2023622392 РФ. № 2023622089: заявл. 03.07.2023; опублик. 13.07.2023. [Soldatov A.N. et al. Database of dental health status of students in general education institutions of the Ministry of Defense of Russia. Certificate on Official Registration of the Database RU 2023622022, published 20/06/2023. In Russian].
8. Prokhorenkova L. et al. CatBoost: unbiased boosting with categorical features // arXiv preprint arXiv:1706.09516v5. – 2017. DOI: 10.48550/arXiv.1706.09516.
9. Anghel A. et al. Benchmarking and optimization of gradient boosting decision tree algorithms // arXiv preprint arXiv:1809.04559. – 2018. DOI: 10.48550/arXiv.1809.04559.
10. Khoshgoftaar T.M. et al. An Empirical study of learning from imbalanced data using random forest: in proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007). – 2007. – V. 2. – P. 310–317. DOI: 10.1109/ICTAI.2007.46.
11. Liashchynskiy P. et al. Grid search, random search, genetic algorithm: a big comparison for NAS // arXiv preprint arXiv:1912.06059. – 2019. DOI: 10.48550/arXiv.1912.06059.