



УДК 797.1:004.032.26

Программный комплекс прогноза результатов соревнований пловцов на основе языковых моделей нейросетей глубокого обучения

А. А. Болотов, В. Г. Барчуков, А. С. Галузин, Е. Н. Жирнов

Федеральное государственное бюджетное учреждение «Государственный научный центр Российской Федерации – Федеральный медицинский биофизический центр имени А.И. Бурназяна» Федерального медико-биологического агентства, Москва, Россия

Аннотация

Актуальность Для спортсмена наряду с эффективной тренировочной подготовкой актуальным при выступлении на соревнованиях являются оценка его функционального состояния и готовность показать лучший результат. При этом оценка состояния спортсмена должна быть для него быстрой, малозатратной и без отвлечения от соревновательного процесса.

Цель. Разработка инновационного программного комплекса, использующего методы глубокого обучения и обработку естественного языка NLP (Natural Language Processing), для прогнозирования спортивных результатов пловцов. Комплекс предназначен для поддержки эффективных решений по отбору сильнейших спортсменов для участия в командных и индивидуальных заплывах на основе анализа их функциональных состояний посредством автоматизированного осознанного диалога и учета ключевых гематологических показателей.

Методы. Для обучения системы были использованы реальные данные о спортсменах сборной команды РФ по плаванию: всего было проанализировано 100 случаев по восьми ключевым гематологическим параметрам, влияющим на спортивные показатели. Для создания комплекса использовались нейросети глубокого машинного обучения на основе библиотеки с открытым исходным кодом Keras и языка программирования Python.

Результаты. Разработанный комплекс включает три основных модуля: модуль «Обучение» – создание языковой модели на основе библиотеки Keras и Python; модуль «Чат+Прогноз» – поддержка осмысленного диалога с пользователем (тренером/врачом/спортсменом) относительно текущего состояния здоровья и уровня готовности; модуль «Тест-Прогноз»: массовая проверка эффективности разработанной модели на конкретных примерах. Точность модели достигла 90% во время обучения, основанного на анализе истории обучения модели. Реальное тестирование подтвердило эффективность подхода, показав совпадение прогноза с результатами в 87% случаев.

Заключение. Разработанный комплекс обеспечивает быструю оценку состояния спортсмена, минимизируя затраты и позволяя эффективно выбирать участников для различных соревнований. Возможности для повышения точности сохраняются путем расширения набора данных. Для получения различных данных по предметной области используется чат-бот с запросами на естественном языке.

Ключевые слова: программный комплекс, искусственная нейросеть, машинное обучение, гематологические показатели, пловцы, соревнования, чат-бот, Keras, языковая QA-модель, классификация

Software package for predicting swimming competition results based on language models of deep learning neural networks

A. A. Bolotov, V. G. Barchukov, A. S. Galuzin, E. N. Zhirnov

State Research Center – Burnasyan Federal Medical Biophysical Center of Federal Medical Biological Agency, Moscow, Russia

Abstract

Relevance. For an athlete, it is not only important to have effective training but also to assess their functional state and readiness to perform at their best during competitions. The assessment of the athlete’s condition should be quick, low-cost, and not distract from the competitive process.

Objective. The development of an innovative software package that utilizes deep learning techniques and natural language processing (NLP) to accurately forecast the athletic performance of swimmers. This package aims to support effective decision-making in selecting the strongest athletes for team and individual swimming events, based on an analysis of their physiological states through automated dialogue and consideration of key blood parameters.

Methods. We used real data on athletes from the Russian national swimming team to train our system. We analyzed a total of 100 cases for eight key hematological parameters that affect athletic performance. To create the system, we used deep machine learning neural networks based on the Keras open-source library and the Python programming language.

Results. The developed system consists of three main components: “Training”: This component creates a language model using the Keras and Python libraries; “Chat+Forecast”: This component supports a meaningful dialogue with users (trainers, doctors, athletes) about their current health status and readiness level; “Test–Forecast”: This component tests the effectiveness of the model on specific examples and provides mass testing. During training, the accuracy of the model reached 90% based on the model’s training history analysis. Real-world testing confirmed the approach’s effectiveness, with a forecast matching results in 87% of cases.

Conclusions. The developed system provides a quick assessment of an athlete’s condition, minimizing costs and allowing for effective selection of participants for different competitions. The accuracy of the system can be improved by expanding the dataset. A chatbot that uses natural language queries can be used to gather various data on the topic area.

Keywords: software package, artificial neural network, machine learning, hematological parameters, swimmers, competitions, chatbot, Keras, language QA-model, classification

Введение

Для спортсмена наряду с эффективной тренировочной подготовкой актуальной при выступлении на соревнованиях является оценка его функционального состояния и готовность показать лучший результат. При этом оценка состояния спортсмена должна быть для него быстрой, малозатратной и без отвлечения от соревновательного процесса. Нами было предложено в соревновательный период использовать для этого гематологические показатели крови спортсмена, забираемой из пальца накануне соревнований. Гематологические показатели позволяют одновременно оценить состояние кислородотранспортной (гемоглобин, эритроциты, гематокрит и производные) и иммунной системы (лейкоциты и их подгруппы). Кислородотранспортная система связана с энергообеспечением, что важно в ходе прохождения дистанции с ускорением на финише, а также для восстановления при повторных стартах в течение соревнований. Достаточное количество энергии также обеспечивает эффективную работу иммунной системы, которая, как известно, страдает на пике тренировочного процесса, что проявляется увеличением простудных заболеваний, приводящих к срыву стартов по причине болезни [1, 2, 3, 4, 5, 6, 7].

Обработка естественного языка – это область искусственного интеллекта (ИИ), которая обеспечивает связь между людьми и машинами, позволяя компьютерам понимать и генерировать человеческий язык. В современном мире одним из самых актуальных направлений в области обработки естественного языка является создание систем «вопрос-ответ» (QA-системы) [8, 9, 10, 11, 12, 13, 14, 15]. QA (Question Answering) системы основаны на искусственном интеллекте и предназначены для автоматического понимания вопросов пользователей и предоставления релевантных и точных ответов на них. Датасет «вопрос-ответ» (QA-модели) для ИИ обычно состоит из пар: вопросов-ответов, соответствующих друг другу. Основой их работы являются языковые модели, которые позволяют анализировать и понимать естественный язык. Основное применение таких систем: вопросно-ответные чат-боты и виртуальные ассистенты, автоматизация поддержки клиентов, техническая помощь. Современные QA-системы используют комплекс технологий глубокого обучения и обработки естественного языка, позволяющий обеспечить высокое качество взаимодействия с пользователями и предоставление точной и актуальной информации. Анализ таких систем показал, что они также успешно решают задачи классификации, когда по набору признаков в запросе, представленных в виде слов, определяется тот или иной класс. Такой же механизм лежит в основе работы QA-систем. Это обстоятельство привело авторов к идее использования подходов QA для прогноза успешности выступления спортсменов на основе гематологических показателей. Предложенный подход позволяет в одной модели обеспечить пользователей разнообразной информацией о системе, а также решать задачи классификации (прогноза) в рамках интеллектуального чат-бота. Для этого представляется целесообразным использовать нейросети глубокого машинного обучения на основе библиотеки с открытым исходным кодом Keras [8, 9, 10, 11, 12, 13, 14, 15]. Возможности этой библиотеки позволяют быстро и с высокой точностью (эффективно) строить вопросно-ответные системы (типа чат-ботов) [13, 14, 15]. Библиотека Keras позволяет с помощью встроенного программного интерфейса пользователя легко и просто сформировать и настроить структуру нейросети, определить ее свойства, обеспечить эффективное (быстрое и управляемое) машинное обучение на основе широкого набора гиперпараметров и таким образом сформировать собственную модель системы, позволяющую вести диалог на естественном языке [14]. Keras предлагает согласованный и простой API, минимизирует количество действий пользователя, необходимых для решения задач множественной

классификации. Возможность легко добавлять новые объекты, модули и функции через датасет делает Keras отличным средством для проведения исследований.

Целью данной работы являлась разработка инновационного программного комплекса, использующего методы глубокого обучения и обработку естественного языка (NLP), для прогнозирования спортивных результатов пловцов в составе чат-бота с возможностью быстрого получения дополнительной информации по предметной области на основе запросов пользователей на естественном языке. Комплекс предназначен для поддержки эффективных решений по отбору сильнейших спортсменов для участия в командных и индивидуальных заплывах на основе анализа их функционального состояния посредством автоматизированного осознанного диалога и учета ключевых гематологических показателей. Решение реализовано с применением открытого инструментария библиотеки Keras на языке программирования Python, обеспечивая перспективность дальнейшей оптимизации и масштабируемость модели.

Методы

Для оценки функционального состояния спортсменов с использованием автоматизированного биохимического анализатора [16, 17] были собраны данные 18 гематологических показателей пловцов в период проведения соревнований с оценкой успешности их выступления. Оценка успешности выступления в соревновании проводилась путем сравнения наилучших результатов спортсмена за предыдущий сезон с лучшими результатами текущего соревнования по баллам калькулятора FINA (ФИНА), дающего возможность единой оценки результатов в различных дисциплинах плавания [18]. Если результат соревнования был лучше результатов предыдущего сезона, то результат считался успешным, иначе – неуспешным. Кровь у спортсменов забиралась с утра до завтрака перед соревнованиями.

Фактическая база данных состояла из 100 примеров выступлений российских пловцов-мужчин членов сборной команды РФ в возрасте от 18 до 27 лет, включающих их гематологические показатели и итоговые результаты состязаний. Полностью детализированная статистика представлена в источнике [19]. Среди этих случаев выделены 65 неудачных выступлений и 35 успешных. Для повышения качества прогноза успешности выступлений спортсменов была проведена аналитика, направленная на выявление факторов наибольшего значения среди гематологических показателей. Методы вероятностного анализа позволили выделить восемь наиболее значимых критериев из исходных восемнадцати. Это сделано с целью максимального снижения информационной избыточности и фокусировки на факторах, оказывающих существенное влияние на спортивный успех.

Оценка важности показателей проводилась на основе методики, изложенной в [19], по двум видам значимости:

- интегральная значимость: выделение главных признаков, непосредственно влияющих на результативность прогноза;
- градационная значимость: определение степени влияния конкретного значения каждого показателя на общий прогноз успеха, а также для объяснения результата прогноза.

Применение такой методологии существенно повысило информативность данных и снизило количество используемых характеристик до оптимального минимума, обеспечивая надежность прогноза. В итоге для оценки функционального состояния спортсменов с целью прогноза успешности их выступлений использовался набор фактических данных по восьми наиболее значимым гематологическим показателям пловцов с оценкой по двум классам успешности/неуспешности выступления спортсменов, полученных в период проведения

соревнований. В него вошло 100 случаев выступления пловцов мужчин, членов сборной команды РФ, при этом гематологические показатели представлены в дискретной форме: признак-градация, согласно методике, приведенной в [19, 20]. Набор таких признаков с их градационной информативностью представлен в [таблице 1](#).

Таблица 1: Наиболее значимые признаки для системы прогноза успешности выступления

№	Название признака	Балл	Интервал значений	Информативность
1	Лейкоциты, $\times 10^9/\text{л}$	1	< 3,587	0,01
		2	3,587-6,708	1,3
		3	> 6,708	0,01
2	Гемоглобин, г/л	1	< 142,6	0,69
		2	142,6-148,0	1,34
		3	> 148,0	2,2
3	Средний объем эритроцита, фл	1	< 75,133	0,61
		2	75,133-83,633	2,14
		3	> 83,633	0,74
4	Среднее содержание гемоглобина в эритроците, пг	1	< 25,533	2,85
		2	25,533-28,893	1,62
		3	> 28,893	0,31
5	Средняя концентрация гемоглобина в эритроцитной массе, г/л	1	< 292	0,3
		2	292-344	1,36
		3	> 344	0,2
6	Распределение тромбоцитов по их объему, %	1	< 12,467	0,37
		2	12,467-13,4	1,74
		3	> 13,4	0,36
7	Лимфоциты, %	1	< 26,88	0,01
		2	26,88-36,987	2,11
		3	> 36,987	0,69
8	Гранулоциты, %	1	< 53,74	0,81
		2	53,74-58,9	2,86
		3	> 58,9	0,95

Для каждой градации признаков в 4-й колонке [таблицы 1](#) представлена информативность градации признака: она показывает, во сколько раз увеличивается или уменьшается вероятность успешного выступления спортсмена. Эта информация важна для системы, т.к. она может использоваться для объяснения прогноза, поиска и анализа прецедентов, позволяет искать границы между классами решений и др. Поэтому информативность признаков была занесена в датасет QA-модели системы для обеспечения информационной поддержки пользователей с использованием запросов на естественном языке.

Для построения QA-модели использовался последовательный (Sequential) тип нейросетевой модели в библиотеке Keras, предназначенной для решения широкого спектра задач, включая классификацию, регрессию и другие виды предсказательных моделей. Последовательная модель представляет собой линейную последовательность слоев, соединенных последовательно один за другим. Каждый слой получает на вход выход предыдущего слоя и передает свой результат следующему слою. Такая структура идеальна для простых и классических архитектур нейронных сетей, таких как многослойные перцептроны (MLP), сверточные сети (CNN) и рекуррентные сети (RNN).

Преимущества последовательной модели:

- простота и удобство использования, такую архитектуру легко проектировать и понимать благодаря линейному порядку слоев;
- высокая производительность, т.к. библиотека оптимизирует работу последовательных моделей, что делает их обучение быстрым и эффективным;
- легкость настройки гиперпараметров, поскольку простая архитектура облегчает эксперименты с количеством слоев, их настройкой, размером фильтров и т.п.

Результаты

Для создания системы прогноза разработаны три программы: «Обучение», «Чат+Прогноз» и программа «Тест–Прогноз». Для обучения системы необходимо было разработать нейросетевую структуру модели с возможно меньшим количеством слоев, позволяющую обеспечить высокую точность ответов, минимальный объем модели и быстроту ее ответа без потери точности ответа по прогнозу и в режиме «вопрос-ответ». Для определения рациональной структуры нейросети нами были разработаны, настроены и практически апробированы 3 структуры: с 4, 6 и 8 слоями. 4-х слойная нейросеть содержала 1 входной слой, один скрытый, один дроп слой и один выходной. Однако такая модель оказалась недостаточно точной. 8-слойная модель показала хорошие результаты, но значительно дольше обучалась, чем 4-х слойная. Достаточно хорошие результаты по обучению и точности показала 6-слойная модель, состоящая из входного, двух скрытых, двух дроп-слоев и одного выходного слоя. Как показано на [рис. 1](#), данная модель была настроена с помощью гиперпараметров и использовалась для создания языковой QA-модели системы.

Каждый нейрон сети получает один или несколько входных сигналов от нейронов предыдущего слоя. Первый скрытый слой получает данные из входного потока. Начиная со второго скрытого слоя, на вход каждому нейрону подается взвешенная сумма нейронов из предыдущего слоя. На втором (скрытом) слое происходит активация входных данных за счет специальной нелинейной функции. С помощью функции активации становится доступной реализация обратного распространения ошибки, которая помогает скорректировать веса нейронов. Методом обратного распространения ошибки происходит обучение модели, при котором оптимизируется критерий потерь, в качестве которого используется кросс-энтропия [11, 12, 13].

Для входных слоев была выбрана простая, но мощная функция активации «ReLU», широко используемая в нейронных сетях. Если входное значение X положительное, то функция возвращает это значение, а если X отрицательное, то функция возвращает 0:

$$f(x) = \max(0, x). \quad (1)$$

Дроп-слои позволяют осуществлять фильтрацию сигналов, задавая порог значений, меньше которого сигнал обнуляется. Это дает возможность ускорить обучение системы и убрать нежелательные колебания потерь и точности при обучении.

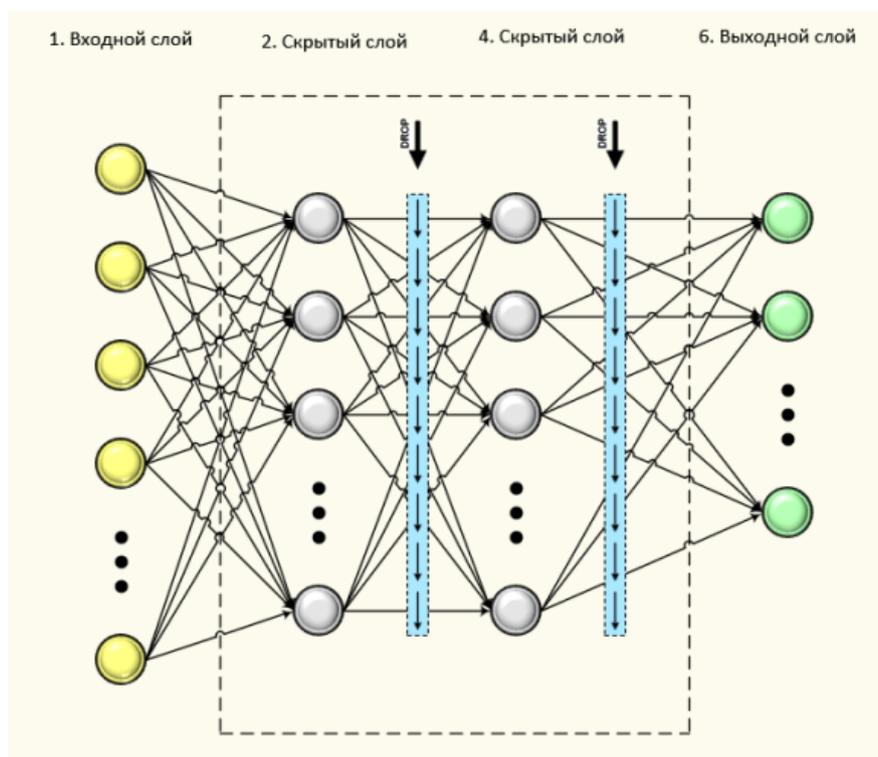


Рис. 1: Структура нейросети для чат-бота и прогноза успешности выступления пловцов
 Fig. 1: Neural network structure for a chatbot and predicting the success of swimmers' performance

Для выходного слоя выбран метод активации «Softmax», поскольку данная функция активации широко используется в нейронных сетях, особенно в контексте задач классификации. Эта функция активации принимает в себя K вещественных чисел, которые вернула ей нейронная сеть, после чего возвращает K чисел в интервале от 0 до 1 и дающих в сумме единицу. Таким образом, каждое значение, получаемое функцией активации Softmax, интерпретируется как вероятность принадлежности к соответствующему классу:

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}, \quad (2)$$

где z_i – выходной сигнал i -го нейрона (до применения функции активации), e – основание натурального логарифма, K – общее количество классов (нейронов в последнем слое).

Для машинного обучения был создан датасет формата «json» в виде двух частей:

1) содержит набор фактических данных гематологических показателей с оценкой по двум классам успешности и неуспешности выступления спортсменов;

2) информационно-справочная часть, касающаяся описания работы системы, аспектов ее функционирования, информации о том как работать в системе, представление о ее основных функциях, а также текстовая информация по используемым признакам, содержащая описание каждого признака, его влияние на прогноз успешности признаков (информативность) и др. в виде набора вопросов и ответов системы для обеспечения диалога с пользователем на естественном языке.

Вид фрагмента такого датасета представлен в [таблице 2](#).

Таблица 2: Вид датасета в формате «json» для прогноза и информационно-справочной части системы (ИСЧ)

Пример датасета для прогноза	Пример датасета ИСЧ для диалога
<pre>{ "tag": "1", "name_tag": ["3"], "patterns": ["1:3 2:3 3:3 4:3 5:3 6:1 7:3 8:1", "Лейкоциты:3 Гемоглобин:3 СреднОбъемЭритроцита:3 СреднСодержГмбВЭритроците:3 СреднКонцентрГмбВЭритроцитнойМассе:3 РаспредТромбоцитовПоИхОбъему:1 ЛимфоцитыПроц:3 ГранулоцитыПроц:1"], "responses": ["Класс 1 Прогноз НЕУСПЕШНЫЙ"] }</pre>	<pre>{ "tag": "s.1", "name_tag": ["Справка"], "patterns": ["Справка по системе", "Справка", "Информация по системе", "Что такое СПУ СТАРТ", "СПУ СТАРТ", "Справка по СПУ СТАРТ"], "responses": ["Я, система 'СПУ-СТАРТ', могу помочь Вам по вопросам оценки прогноза успешности выступления спортсменов на основе гематологических признаков, а также по другим вопросам обеспечения этого процесса.\nПримеры запросов:\n1. Лейкоциты\n2. Гемоглобин\n3. Значимость признаков"] }</pre>

При этом "tag" определяет класс системы, "name_tag" для прогноза содержит номер строки из фактических данных обучения по гематологии, для информационно-справочной части – название конкретного вида информации.

Раздел "patterns" содержит вопросы для нашей QA-модели. В нем гематологические данные из обучающей выборки определяют конкретный вопрос в виде образца исходных данных с соответствующим классом прогноза. В информационно-справочной части датасета может быть использована любая текстовая информация, связанная с данной предметной областью системы в виде набора вопросов к ней и ответов системы.

Раздел "responses" обеспечивает тот или иной ответ модели. Для прогнозной части в этом разделе представляется класс конкретного случая из обучающей выборки. Для информационно-справочной части – ответ на все вопросы из раздела «patterns».

По всем наборам вопросов, представленных в разделе «patterns», формируются поисковые вектор-запросы и словарь модели, которые используются для машинного обучения. Список классов создается из раздела «tag». Для автоматизированного создания прогнозной части датасета использовались макросы Excel таблицы исходных данных для 8 признаков с установленным классом. Затем с помощью текстового редактора к этим данным добавлялись вопросы-ответы информационно-справочной части системы. Сформированный датасет подается на вход системы обучения, в результате чего создается файловый состав языковой QA-модели Kegas, который содержит 4 файла следующего вида:

- model1.h5 – обученная языковая QA-модель;
- classes.pkl – классы (группы ответов) модели (в бинарном виде);
- words.pkl – словарь модели (в бинарном виде);

- dataset.json – датасет модели.

Результаты обучения системы представлены на рис. 2 – графиках с оценкой функции потерь и точности. Функция потерь (Loss) минимизируется в процессе обучения. Для этого была выбрана категориальная кросс-энтропия. Она вычисляет потерю кросс-энтропии между истинными классами и предсказанными классами, имеет вероятностную интерпретацию, позволяет модели выводить вероятности каждого класса. В качестве точности выбрана метрика "accuracy". Она рассчитывает, как часто прогнозы совпадают с обучающими данными по классам (меткам), заданных полем "tag" в датасет. Для построения графиков рис. 2 использовалась так называемая «история обучения» библиотеки Keras, в которой сохраняются массивы данных каждой эпохи обучения. При этом можно автоматически разделить данные датасета на обучающую и тестируемую части. Тестируемая часть не используется в процессе обучения и служит для получения оценки точности, весьма близкой к реальной оценке точности работы системы. Таким образом, уже на этапе обучения системы можно судить о качестве ее работы, а также экспериментально осуществлять ее настройку, подбирая гиперпараметры нейросети.

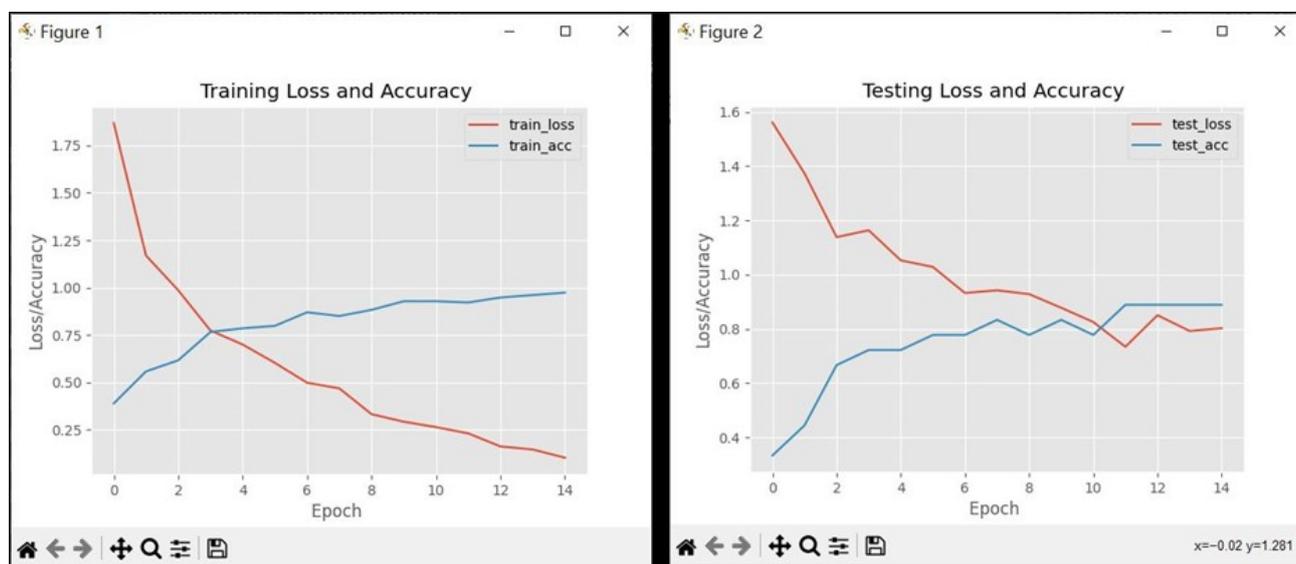


Рис. 2: Графики оценки функции потерь и точности при обучении системы
 Fig. 2: Graphs for estimating the loss function and accuracy during system training

В программе «Чат-Прогноз», названной «СПУ Старт», для получения любой информации используется поле ввода запроса системы. Для прогноза в поле ввода запроса вводится информация в виде строки, содержащей набор слов специального вида – в форме номера признака и его градации, например, 1:2 2:3 3:1 и т.д., либо в виде текстового описания с градацией, например, лейкоциты:2, гемоглобин:3 и т.д. Такой подход позволяет нейросети с высокой точностью разделять ответы на прогнозную и информационную части, т.к. при этом на выходе нейросети активируются только те выходы, которые отвечают за прогноз. Дополнительно предусмотрен графический интерфейс для быстрого и удобного ввода признаков для прогноза. Программа «СПУ Старт» использует QA языковую модель как для прогноза, так и для общения с пользователем на естественном языке по всей информации, имеющейся в модели, обеспечивая при этом достаточно высокую общую точность ответов.

Следует отметить несколько особенностей предложенного подхода.

1. Можно обучить модель только для прогноза успешности, создав датасет исклю-

чительно из фактических данных по успешности/неуспешности. При этом программное обеспечение не требует каких-либо изменений.

2. Модель не зависит от числа классов. Можно было бы с таким же успехом использовать 3 класса для прогноза, например, 1 – неуспешно, 2 – неопределенный результат, 3 – успешный результат. Для этого в фактических данных должно быть 3 класса примеров. Программное обеспечение в этом случае также не требует изменений, но фактических данных должно быть значительно больше, чем в случае 2-х классов.

3. Подход позволил создать несколько QA-моделей: малую, среднюю и большую. Малая модель содержит небольшое число слоев (до 4) и обеспечивает только прогноз. Средняя модель соответствует рассмотренному в статье материалу. Большая модель содержит больше слоев, чем средняя и обеспечивает дополнительную информационную поддержку в ответе системы – появляется возможность давать ответ на запрос не только в тексте, но и в виде изображений, аудио, видео, текстовых документов большого объема и др.

4. Среди небольших недостатков для прогнозной части системы можно отметить необходимость задания строго определенного запроса в виде набора гематологических показателей в формате признак:градация. Далее рассмотрим несколько примеров работы «СПУ Старт».

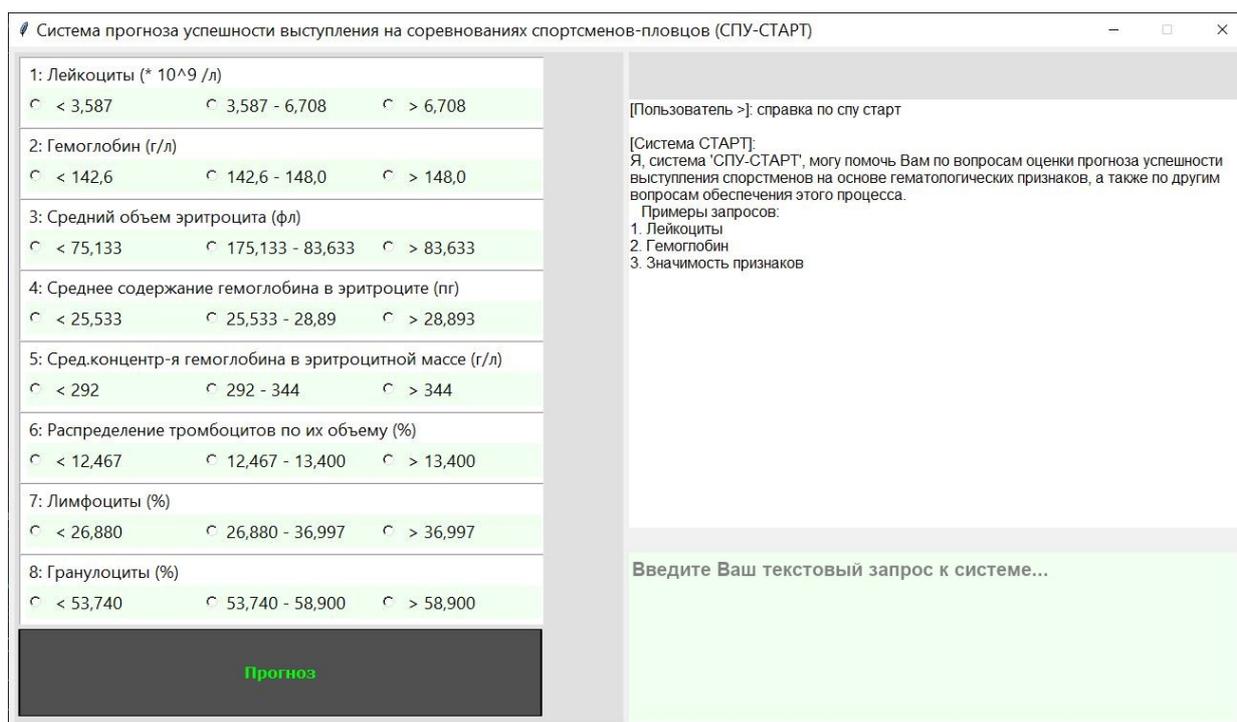


Рис. 3: Пример реализации справки по запросу пользователя

Fig. 3: An example of providing assistance at the user's request

На рис. 3 представлен пример реализации ответа системы по запросу пользователя, представленного в виде набора слов «справка по СПУ Старт», а на рис. 4 – пример прогноза успешности по графическому интерфейсу и запроса на получение информации по значимости признаков.

Система прогноза успешности выступления на соревнованиях спортсменов-пловцов (СПУ-СТАРТ)

1: Лейкоциты (* 10 ⁹ /л)	< 3,587	3,587 - 6,708	> 6,708
2: Гемоглобин (г/л)	< 142,6	142,6 - 148,0	> 148,0
3: Средний объем эритроцита (фл)	< 75,133	175,133 - 83,633	> 83,633
4: Среднее содержание гемоглобина в эритроците (пг)	< 25,533	25,533 - 28,89	> 28,893
5: Сред.концентр-я гемоглобина в эритроцитной массе (г/л)	< 292	292 - 344	> 344
6: Распределение тромбоцитов по их объему (%)	< 12,467	12,467 - 13,400	> 13,400
7: Лимфоциты (%)	< 26,880	26,880 - 36,997	> 36,997
8: Гранулоциты (%)	< 53,740	53,740 - 58,900	> 58,900

Прогноз

[Система СТАРТ]:
Класс 2 Прогноз УСПЕШНЫЙ

[Пользователь >]: 1:2 2:2 3:3 4:2 5:2 6:3 7:1 8:2

[Система СТАРТ]:
Класс 2 Прогноз УСПЕШНЫЙ

[Пользователь >]: Значимость признаков

[Система СТАРТ]:
Оценка влияния гематологических показателей на определение функционального состояния (в целях прогноза успешности выступлений) спортсменов на соревнованиях основана на базе вероятностных подходов. Рассматриваются два вида оценок значимости показателей: интегральная оценка и градационная оценка влияния значений каждого признака на результат прогноза. Интегральная оценка значимости предназначена для отбора признаков, наиболее сильно влияющих на результат прогноза, а градационная оценка позволяет определить влияние значений каждого показателя на прогноз успешности и служит для объяснения результата прогноза. Градации гематологического показателя, у которых оценка его значимости больше единицы увеличивают вероятность успешного прогноза и наоборот - градации показателя, у которых оценка значимости меньше единицы, уменьшают вероятность успешного прогноза.

Введите Ваш текстовый запрос к системе...

Рис. 4: Пример прогноза успешности по графическому интерфейсу и запроса на получение информации по значимости признаков

Fig. 4: An example of predicting success using a graphical interface and requesting information on the significance of features

Система прогноза успешности выступления на соревнованиях спортсменов-пловцов (СПУ-СТАРТ)

1: Лейкоциты (* 10 ⁹ /л)	< 3,587	3,587 - 6,708	> 6,708
2: Гемоглобин (г/л)	< 142,6	142,6 - 148,0	> 148,0
3: Средний объем эритроцита (фл)	< 75,133	175,133 - 83,633	> 83,633
4: Среднее содержание гемоглобина в эритроците (пг)	< 25,533	25,533 - 28,89	> 28,893
5: Сред.концентр-я гемоглобина в эритроцитной массе (г/л)	< 292	292 - 344	> 344
6: Распределение тромбоцитов по их объему (%)	< 12,467	12,467 - 13,400	> 13,400
7: Лимфоциты (%)	< 26,880	26,880 - 36,997	> 36,997
8: Гранулоциты (%)	< 53,740	53,740 - 58,900	> 58,900

Прогноз

Лейкоциты:> 6,708->значимость=0,01
Показатель значимости градации признака более 1 увеличивает вероятность успешного прогноза во столько раз

[Пользователь >]: 1:1 2:2 3:3 4:3 5:1 6:3 7:2 8:3

[Система СТАРТ]:
Класс 1 Прогноз НЕУСПЕШНЫЙ

[Пользователь >]: признаки

[Система СТАРТ]:
Признаки в системе СПУ-СТАРТ используются в формате признак:градация, например Лейкоциты:3 Гемоглобин:3 СреднОбъемЭритроцита:3 СреднСодержГмбВЭритроците:3 СреднКонцентрГмбВЭритроцитнойМассе:3 РаспредТромбоцитовПоИхОбъему:1 ЛимфоцитыПроц:3 ГранулоцитыПроц:1

[Пользователь >]: Лейкоциты:3 Гемоглобин:3 СреднОбъемЭритроцита:3 СреднСодержГмбВЭритроците:3 СреднКонцентрГмбВЭритроцитнойМассе:3 РаспредТромбоцитовПоИхОбъему:1 ЛимфоцитыПроц:3 ГранулоцитыПроц:1

[Система СТАРТ]:
Класс 1 Прогноз НЕУСПЕШНЫЙ

Введите Ваш текстовый запрос к системе...

Рис. 5: Пример использования запроса по признакам системы и прогноза на основе текстового представления признаков

Fig. 5: An example of using a query based on system features and a forecast based on a text representation of features

Признаки (верхняя левая часть экрана) подсвечиваются определенным цветом. Зеленый цвет соответствует успешному прогнозу. Для реализации этой функциональности используется информация о градационной значимости признаков.

Пример использования запроса по признакам системы и прогноза на основе текстового представления признаков представлен на [рис. 5](#).

Для оценки точности подсистемы «Прогноз» была дополнительно разработана программа «Тест-Прогноз», которая позволяет оценить точность классификации (прогноза) на массиве тестовых данных, записанных в Excel файле при использовании обученной QA-модели. Тестовые данные представлены набором из 8 признаков в цифровом формате «признак:градация» и номером класса, предложенного экспертом. При работе этой программы по набору признаков вычислялся прогноз (класс) системы, который сравнивался с классом, заданным экспертом. Проверка на реальных 25 тестовых случаях прогноза успешности выступления спортсменов – членов сборной команды РФ по плаванию показала следующие результаты, представленные в [таблице 3](#).

Таблица 3: Оценка точности работы программы «Тест-прогноз» с матрицей путаницы

Наименование показателя	1 класс-неуспешный	2 класс-успешный	Всего
Исходные тестовые данные	18	7	25
Количество верных прогнозов	16	6	22
Количество ошибок прогноза	2	1	3
Процент ошибок прогноза	11,1	14,3	12,7

Таким образом, достаточно высокое соответствие прогноза и исходной оценки успешности в 87% случаев (при ошибках 1-го рода в 11,1% и 14,3% ошибках 2-го рода) позволяет сделать вывод о работоспособности предложенного подхода. При этом следует заметить, что успешность выступления спортсмена зависит не только от его физического состояния, но и от психологического, от его настроенности на победу, поэтому для более точного прогноза целесообразно еще иметь и такую подсистему прогноза.

Заключение

Разработанный программный комплекс, основанный на технологиях искусственного интеллекта и глубоком машинном обучении, обеспечивает достаточно хорошие перспективы в плане практического применения в спортивной медицине и подготовке профессиональных спортсменов. Система успешно решает задачу оперативного прогнозирования успешности выступлений пловцов на соревнованиях, используя данные гематологического профиля. Благодаря встроенным механизмам взаимодействия на естественном языке и дополнительным сервисам системы повышается уровень доверия пользователей к результатам прогноза.

Тестирование программы подтверждает ее надежность и способность выдавать точные прогнозы даже на ограниченном объеме данных. Полученные результаты свидетельствуют о достаточно высокой эффективности метода и открывают широкие горизонты для дальнейших исследований и его совершенствования на основе дополнительных фактических данных и применения в других спортивных дисциплинах.

В целом, данная работа представляет собой важный вклад в развитие технологий спортивной аналитики и предлагает эффективный инструмент для принятия обоснован-

ных решений при формировании состава сборных команд и планирования тренировок спортсменов высокого класса.

Список литературы

1. *Нехвядович А.И., Будко А.Н.* Оценка эффективности тренировочного процесса спортсменов на основе вариабельности показателей крови: Практическое пособие. Минск: БГУФК, 2019. 40 с.
2. *Трищенкова С.Н., Екимовских А.В., Егоров Г.Е.* Интегральные гематологические показатели у спортсменов [Электронный ресурс] // Спортивная медицина. URL: <https://www.sportmedicine.ru/medforsport-2011-papers/trischenkova.php> (Дата обращения: 27.03.2025)
3. *Суздальницкий Р.С., Левандо В.А.* Иммунологические аспекты спортивной деятельности человека // Теория и практика физической культуры. 1998. № 10. С. 43-46.
4. *Даутова А.З., Яньшиева Г.Г., Якубов Р.Ю., Назаренко А.С., Зверев А.А.* Взаимосвязь гематологических и биохимических параметров крови у спортсменов разных возрастных групп // Наука и спорт. 2022. № 3 (10). С. 14-21. DOI: <https://doi.org/10.36028/2308-8826-2022-10-3-14-21>
5. *Бойков В.Л.* Комплексная характеристика автономной регуляции сердечного ритма, гематологического и биохимического профиля у пловцов высокой квалификации: автореф. дис. ... канд. биол. наук: 03.03.01 Физиология. Ярославль, 2021. 21 с.
6. *Рязанцев А.И., Сафонова С.Н., Голомедов М.Р.* Гематологические маркеры состояния функциональных систем высококвалифицированных пловцов в условиях среднегорной подготовки // Физическая культура. Спорт. Здоровье. 2024. № 1. С. 82-89. URL: <https://repo.nspu.ru/handle/nspu/5103>
7. *Anshik.* AI for Healthcare with Keras and TensorFlow 2.0. Berkeley, CA: Apress, 2021. 381 p. DOI: <https://doi.org/10.1007/978-1-4842-7086-8>
8. *Элбон К.* Машинное обучение с использованием Python. Сборник рецептов. СПб.: БХВ-Петербург, 2019. 384 с.
9. *Machine Learning with Python Tutorial.* [Электронный ресурс] URL: <https://www.geeksforgeeks.org/machine-learning-with-python/> (Дата обращения: 15.03.2025)
10. *Python.* [Электронный ресурс]. URL: <https://www.python.org/> (Дата обращения: 15.03.2025)
11. *Best Python libraries for Machine Learning.* [Электронный ресурс] URL: <https://www.geeksforgeeks.org/best-python-libraries-for-machine-learning/> (Дата обращения: 15.03.2025)
12. *TensorFlow Core.* [Электронный ресурс]. URL: <https://www.tensorflow.org/guide/keras?hl=ru> (Дата обращения: 15.03.2025)
13. *Шолле Ф.* Глубокое обучение на Python. 2-е междунар. изд. СПб.: Питер, 2022. 576 с.
14. *Keras. Deep Learning for humans.* [Электронный ресурс]. URL: <https://keras.io/> (Дата обращения: 27.03.2025)
15. *Keras: библиотека глубокого обучения на Python.* [Электронный ресурс]. URL: <https://ru-keras.com/home/> (Дата обращения: 27.03.2025)
16. *Руководство по обслуживанию. Принципы работы. Описание электронного оборудования Chemwell-T.* [Электронный ресурс]. URL: <https://www.manualslib.com/manual/1863455/Awareness-Technology-Chemwell-T-4600-Series.html?page=38#manual> (Дата обращения: 15.03.2025)
17. *Компания Td-lab, дилер продукции Awareness Technology в РФ.* [Электронный ресурс]. URL: <https://td-lab.ru/product/avtomaticheskij-bioximicheskij-analizator-chemwell-t> (Дата обращения: 15.03.2025)

18. *Таблица очков World Aquatics (FINA)*. [Электронный ресурс]. URL: <https://russwimming.ru/fina-points> (Дата обращения: 27.02.2025)
19. Болотов А.А., Барчуков В.Г., Галузин А.С., Тен А.М., Онопченко О.В. Использование гематологических показателей для прогноза результатов соревнований спортсменов-пловцов на основе вероятностных оценок // Клинический вестник ФМБЦ им. А. И. Бурназяна. 2023. № 4. С. 10-18. DOI: <https://doi.org/10.33266/2782-6430-2023-4-10-18> EDN: <https://www.elibrary.ru/uafhzv>
20. Болотов А.А., Барчуков В.Г., Галузин А.С., Тен А.М., Онопченко О.В. Прогноз результатов соревнований спортсменов-пловцов по гематологическим показателям // Сборник тезисов VII Всероссийской с международным участием научно-практической конференции «День спортивной информатики». 2024. С. 31-34. DOI: <https://doi.org/10.62105/2949-6349-2024-1-S1-31-34> EDN: <https://www.elibrary.ru/ktlcm>

References

1. Nekhyadovich A.I., Budko A.N. Evaluation of the effectiveness of athletes' training process based on blood parameter variability: Practical guide. Minsk: BSUPC, 2019, 40 p. (in Russ.)
2. Trishchenkova S.N., Ekimovskikh A.V., Egorov G.E. Integral hematological parameters in athletes. *Sports Medicine*. [Electronic resource]. URL: <https://www.sportmedicine.ru/medforsport-2011-papers/trishchenkova.php> (in Russ.)
3. *Suzdalnitsky R.S., Levando V.A.* Immunological aspects of human sports activity. Theory and Practice of Physical Culture, 1998, (10), pp. 43-46. (in Russ.)
4. Dautova A.Z., Yanysheva G.G., Yakubov R.Yu., Nazarenko A.S., Zverev A.A. Relationship between hematological and biochemical blood parameters in athletes of different age groups. *Science and Sport*. 2022, 3 (10), pp. 14-21. DOI: <https://doi.org/10.36028/2308-8826-2022-10-3-14-21> (in Russ.)
5. Boikov V.L. Comprehensive characteristics of autonomous heart rate regulation, hematological and biochemical profile in highly qualified swimmers: Abstract of dissertation ... Candidate of Biological Sciences: 03.03.01 Physiology. Yaroslavl, 2021. 21 p. (in Russ.)
6. Ryazantsev A.I., Safonova S.N., Golomedov M.R. Hematological markers of the functional systems state in highly qualified swimmers under mid-mountain training conditions. *Physical Culture. Sport. Health*. 2024, (1), pp. 82-89. URL: <https://repo.nspu.ru/handle/nspu/5103> (in Russ.)
7. Anshik. *AI for Healthcare with Keras and TensorFlow 2.0*. Berkeley, CA: Apress, 2021. 381 p. DOI: <https://doi.org/10.1007/978-1-4842-7086-8>
8. Elbon K. *Machine Learning with Python Cookbook*: Transl. from English. St. Petersburg: BHV-Petersburg, 2019. 384 p. (in Russ.)
9. *Machine Learning with Python Tutorial*. [Electronic resource]. URL: <https://www.geeksforgeeks.org/machine-learning-with-python/> (Accessed: 15.03.2025)
10. *Python*. [Electronic resource]. URL: <https://www.python.org/> (Accessed: 15.03.2025)
11. *Best Python libraries for Machine Learning*. [Electronic resource]. URL: <https://www.geeksforgeeks.org/best-python-libraries-for-machine-learning/> (Accessed: 15.03.2025)
12. *TensorFlow Core*. [Electronic resource]. URL: <https://www.tensorflow.org/guide/keras?hl=ru> (Accessed: 15.03.2025)

13. Chollet F. *Deep Learning with Python*. 2nd international ed. St. Petersburg: Piter, 2022. 576 p. (in Russ.)
14. Keras. *Deep Learning for humans*. [Electronic resource]. URL: <https://keras.io/> (Accessed: 27.03.2025)
15. Keras: Python deep learning library. [Electronic resource]. URL: <https://ru-keras.com/home/> (Accessed: 27.03.2025)
16. *Maintenance Manual. Operating Principles. Description of Chemwell-T Electronic Equipment*. [Electronic resource]. URL: <https://www.manualslib.com/manual/1863455/Awareness-Technology-Chemwell-T-4600-Series.html?page=38#manual> (Accessed: 15.03.2025)
17. *Td-lab Company, dealer of Awareness Technology products in Russia*. [Electronic resource]. URL: <https://td-lab.ru/product/avtomaticheskij-bioximicheskij-analizator-chemwell-t> (in Russ.) (Accessed: 15.03.2025)
18. *World Aquatics (FINA) Points Table*. [Electronic resource]. URL: <https://russwimming.ru/fina-points> (Accessed: 27.02.2025)
19. Bolotov A.A., Barchukov V.G., Galuzin A.S., Ten A.M., Onopchenko O.V. Using hematological parameters to predict competition results of swimmer athletes based on probabilistic estimates. *Clinical Bulletin of Burnazyan Federal Medical Biophysical Center*, 2023, (4), pp. 10-18. (in Russ.) DOI: <https://doi.org/10.33266/2782-6430-2023-4-10-18> EDN: <https://www.elibrary.ru/uafhzv>
20. Bolotov A.A., Barchukov V.G., Galuzin A.S., Ten A.M., Onopchenko O.V. Prediction of competition results of swimmer athletes by hematological parameters. *Collection of abstracts of the VII All-Russian scientific-practical conference with international participation "Sports Informatics Day"*, 2024, pp. 31-34. (in Russ.) DOI: <https://doi.org/10.62105/2949-6349-2024-1-S1-31-34> EDN: <https://www.elibrary.ru/ktlcm>

Сведения об авторах

Болотов Александр Александрович – кандидат технических наук, ФГБУ ГНЦ ФМБЦ им. А.И. Бурназяна ФМБА России, Москва, Россия.

Е-mail: abolotov@bk.ru

Барчуков Валерий Гаврилович – доктор медицинских наук; ФГБУ ГНЦ ФМБЦ им. А.И. Бурназяна ФМБА России, Москва, Россия.

Е-mail: barchval@yandex.ru

Галузин Александр Сергеевич – ФГБУ ГНЦ ФМБЦ им. А.И. Бурназяна ФМБА России, Москва, Россия.

Е-mail: alexserg_n@mail.ru

Жирнов Евгений Николаевич – кандидат медицинских наук, ФГБУ ГНЦ ФМБЦ им. А.И. Бурназяна ФМБА России, Москва, Россия.

Е-mail: eugzhirnov@mail.ru

Для цитирования:

Болотов А.А., Барчуков В.Г., Галузин А.С., Жирнов Е.Н. Программный комплекс прогноза результатов соревнований пловцов на основе языковых моделей нейросетей глубокого обучения // Российский журнал информационных технологий в спорте. – 2025. – Т. 2, № 1. – С. 22–37. DOI: <https://doi.org/10.62105/2949-6349-2025-2-1-22-37> EDN: UXEFIO

For citation:

Bolotov A.A., Barchukov V.G., Galuzin A.S., Zhirnov E.N. Software package for predicting swimming competition results based on language models of deep learning neural networks. *Russian Journal of Information Technology in Sports*, 2025, 2 (1), pp. 22–37. (in Russ.). DOI: <https://doi.org/10.62105/2949-6349-2025-2-1-22-37> EDN: UXEFIO

Статья поступила в редакцию: 15.04.2025

Статья принята в печать: 15.06.2025

Статья опубликована: 30.06.2025