

УДК 004.93'1

Булгаков Денис Геннадьевич

студент 4 курса бакалавриата, Национальный исследовательский ядерный университет «МИФИ», Россия, г. Москва

Заика Михаил Алексеевич

студент 4 курса бакалавриата, Национальный исследовательский ядерный университет «МИФИ», Россия, г. Москва

РАЗРАБОТКА АЛГОРИТМОВ И ПРОГРАММНОЙ СИСТЕМЫ ДЛЯ ПОВЫШЕНИЯ ПРОДУКТИВНОСТИ КРУПНОГО РОГАТОГО СКОТА

Аннотация. В статье рассматривается разработанная интеллектуальная программная система для повышения продуктивности крупного рогатого скота, объединяющая анализ продуктивных, сенсорных и визуальных данных. Для прогнозирования удоя и классификации вкуса молока были применены градиентные бустинги LightGBM и CatBoost с интерпретацией результатов через SHAP. Для раннего выявления заболеваний по сенсорным данным реализованы модели контролируемого обучения и LSTM-автоэнкодер для обнаружения аномалий. Для анализа видеопоследовательностей разработан конвейер компьютерного зрения на основе YOLO, HRNet, CNN и LSTM с механизмом внимания. Все модули объединены в единую масштабируемую архитектуру, обеспечивающую гибкость и поддержку принятия решений в животноводстве.

Ключевые слова: крупный рогатый скот, продуктивность, заболевания, машинное обучение, LightGBM, CatBoost, LSTM.

Annotation. The article presents a developed intelligent software system for increasing cattle productivity, integrating the analysis of productive, sensory, and visual data. For milk yield prediction and milk taste classification, gradient boosting methods LightGBM and CatBoost were applied, with results interpreted using SHAP. For early disease detection based on sensory data, supervised learning models and an LSTM autoencoder for anomaly detection were implemented. For video sequence analysis, a computer vision pipeline based on

YOLO, HRNet, CNN, and LSTM with an attention mechanism was developed. All modules are combined into a unified scalable architecture, providing flexibility and decision support in livestock farming.

Keywords: cattle, productivity, diseases, machine learning, LightGBM, CatBoost, LSTM.

Введение

Повышение продуктивности крупного рогатого скота (КРС) требует комплексного подхода, объединяющего контроль продуктивных показателей (удой, качество молока) и раннее выявление заболеваний, которые приводят к снижению продуктивности. Традиционные методы мониторинга основаны на редких лабораторных анализах и визуальных осмотрах, что делает их реактивными и трудозатратными [4, 7]. Современные фермы генерируют большие объёмы данных (показатели удоев, активности, температуры, а также видеопотоки), которые часто не используются в полной мере. Существующие коммерческие системы (Afimilk, Lely, DeLaval) привязаны к дорогостоящему оборудованию и имеют ограниченную гибкость, что затрудняет расширение их функционала [5]. Таким образом, разработка интеллектуальной системы, обеспечивающей автоматизированный анализ продуктивных, сенсорных и визуальных данных для комплексного мониторинга состояния животных является актуальной и перспективной задачей современного животноводства.

Целью работы является разработка методов и программной системы для повышения продуктивности КРС за счёт автоматизированного анализа трёх типов данных: продуктивных (удой, состав молока, рацион), сенсорных физиолого–поведенческих (температура, активность, время лежания) и визуальных (изображения, видеопоследовательности, поза животных).

Прогнозирование продуктивности

Для прогнозирования среднегодового удоя решалась задача регрессии. В качестве исходных данных использовались биохимические показатели молока (содержание жира, белка, соматических клеток), параметры рациона (энергетические кормовые единицы, содержание сырого протеина), а также порода, возраст животного и тип пастбища.

Для решения задачи регрессии (определение среднегодового удоя) был выбран алгоритм LightGBM (Light Gradient Boosting Machine). Данный выбор обусловлен высокой производительностью на табличных данных, устойчивостью к пропущенным значениям, а также встроенным механизмом регуляризации, снижающим риск переобучения. Гиперпараметры модели оптимизировались с помощью поиска по сетке с кросс-валидацией.

Для определения вкуса молока решалась задача бинарной классификации с категориями «вкусно» и «не вкусно». В качестве модели использован CatBoost, который оптимизирован для работы с категориальными признаками и не требует их предварительного кодирования. Модель обучалась с использованием бинарной кросс-энтропии в качестве функции потерь.

Качество моделей оценивалось с помощью пятикратной кросс-валидации со стратификацией по целевой переменной для классификационной задачи. Для регрессионной модели достигнуты следующие значения метрик: $RMSE=(218,8\pm37,5)$ кг, $MAE=(180,5\pm28,7)$ кг, $R^2 = 0,805\pm0,067$. Полученное значение коэффициента детерминации свидетельствует о том, что модель объясняет около 80% дисперсии целевой переменной. Для модели классификации получены: $Accuracy=0,714\pm0,033$, $Precision=0,706\pm0,035$, $Recall=0,707\pm0,038$, $F1-score=0,706\pm0,036$. Сбалансированные значения полноты и точности указывают на отсутствие существенного дисбаланса в распознавании положительного и отрицательного классов.

Для интерпретации результатов и выявления наиболее значимых признаков использовался SHAP-анализ (SHapley Additive exPlanations). Наибольший вклад в прогноз удоя вносят сахарно-протеиновое отношение и процент жира

в молоке (Рис. 1); для классификации вкуса — сахарно-протеиновое отношение и тип пастбища (Рис. 2).

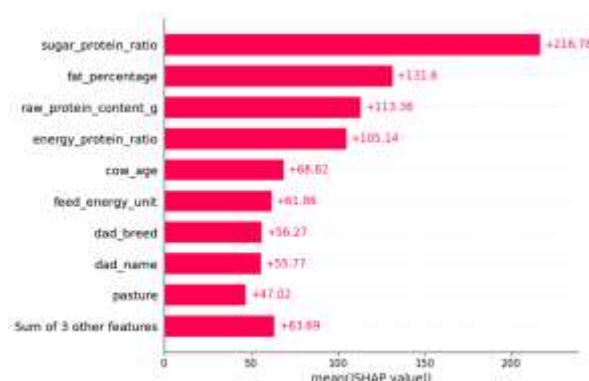


Рисунок 1 – Bar-график важности признаков для модели регрессии

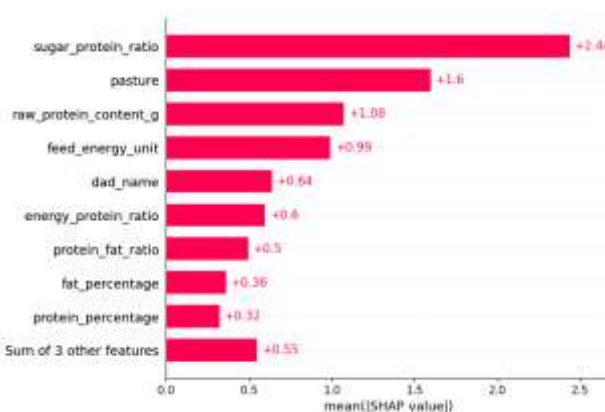


Рисунок 2 – Bar-график важности признаков для модели классификации

Прогнозирование заболеваний по сенсорным данным

Для решения задачи раннего выявления заболеваний крупного рогатого скота использовались данные мониторинга, включающие физиологические, поведенческие и продуктивные параметры: температуру тела, двигательную активность, время лежания, суточный удой, а также климатические параметры (температура и влажность воздуха в помещении). Все показатели регистрировались с частотой от 1 секунды (активность) до 1 суток (удой) и были синхронизированы по временной шкале.

Рассматривались следующие заболевания и патологические состояния, наиболее распространённые в молочном животноводстве и имеющие выраженные физиолого-поведенческие проявления: мастит (воспаление

молочной железы) [2, 12, 13], кетоз (метаболическое нарушение в ранний период лактации) [9], ламинит (воспаление копыт, приводящее к хромоте) [8, 11], тепловой стресс (нарушение терморегуляции при повышенной температуре и влажности) [1, 6, 10], респираторные заболевания (пневмония и воспалительные процессы дыхательных путей), а также метрит (послеродовое воспаление матки). Каждое из перечисленных заболеваний сопровождается характерными изменениями наблюдаемых параметров: повышением температуры, снижением активности, увеличением времени лежания, падением удоев или изменением пищевого поведения.

Для многоклассовой классификации заболеваний использовались два алгоритма: многослойный перцептрон (MLP) и случайный лес (Random Forest). MLP представляет собой полносвязную нейронную сеть с одним скрытым слоем из 64 нейронов и функцией активации ReLU; выходной слой использует функцию softmax для получения вероятностей принадлежности к каждому из классов (норма + шесть заболеваний). Случайный лес состоял из 100 деревьев решений, использовал критерий Джини для оценки качества разбиения и поддерживал взвешивание классов для учёта дисбаланса выборки. Обе модели обучались на признаках, агрегированных к суточным временным интервалам, с разделением данных на обучающую (80%) и тестовую (20%) выборки со стратификацией по классам.

Для выявления аномалий в случаях, когда размеченные данные отсутствовали или были крайне разрежены, применялся автоэнкодер на основе рекуррентной нейронной сети LSTM (LSTM-AE). Модель обучалась на последовательностях признаков, соответствующих здоровым животным, и училась восстанавливать входные данные после сжатия через скрытое пространство. Ошибка реконструкции (среднеквадратичное отклонение между входным и восстановленным вектором) служила аномалийным баллом: превышение порога, определённого по 95-му перцентилю распределения ошибок на обучающей выборке, интерпретировалось как аномальное состояние, потенциально связанное с развитием заболевания. Такой подход

позволил обнаруживать отклонения без необходимости разметки, что особенно ценно для редких или слабо выраженных патологий.

Качество моделей оценивалось с помощью пятикратной кросс-валидации. Для модели классификации на примере мастита (как наиболее распространённого заболевания) были получены следующие значения метрик: Accuracy=0,79±0,06, Precision=0,91±0,03, Recall=0,76±0,04, F1-score=0,77±0,08. Высокая точность (Precision=0,91) свидетельствует о низкой доле ложных срабатываний – система редко ошибочно сигнализирует о мастите при его отсутствии. Более низкая полнота (Recall=0,76) указывает на то, что около четверти случаев заболевания могут быть пропущены, что является допустимым компромиссом для системы раннего оповещения, где важнее не пропустить тяжёлый случай, но при этом не перегружать фермера ложными срабатываниями системы предупреждений. Для аномальной модели на основе LSTM–AE значение Precision@10 (доля истинных аномалий среди первых 10 объектов с наибольшим аномальным баллом) составило 0,88, а частота ложных срабатываний - 0,004, что подтверждает способность автоэнкодера выделять нетипичные состояния без предварительного обучения на размеченных примерах.

Полученные результаты демонстрируют, что комбинация методов контролируемого обучения (MLP, Random Forest) для классификации распространённых заболеваний и аномального подхода (LSTM–AE) для обнаружения нетипичных состояний обеспечивает эффективное решение задачи раннего выявления патологий крупного рогатого скота на основе сенсорных данных.

Прогнозирование заболеваний по визуальным данным.

Для анализа видеопоследовательностей разработан конвейер компьютерного зрения, включающий несколько этапов (Рис. 3). На первом этапе выполняется детекция животных с использованием архитектуры YOLO, позволяющей выделять ограничивающие прямоугольники для каждой коровы

на кадре. На втором этапе производится оценка позы с помощью HRNet – сети, определяющей координаты ключевых точек тела (голова, спина, конечности). Это позволяет перейти от анализа сырых изображений к структурированному представлению вида $s(t) = \{(x_1, y_1), (x_2, y_2) \dots, (x_m, y_m)\}$, где m – число ключевых точек. Такое представление абстрагируется от фона и условий съёмки. На третьем этапе из полученных областей интереса с помощью свёрточной нейронной сети (CNN) извлекаются пространственные признаки. Далее формируются временные последовательности признаков фиксированной длины, и для моделирования временной динамики используется рекуррентная нейронная сеть LSTM с механизмом внимания. Механизм внимания позволяет модели фокусироваться на наиболее информативных временных интервалах. Финальная классификация заболеваний (мастит, ламинит, тепловой стресс, общее угнетённое состояние) выполняется многослойным перцептроном с сигмоидной функцией активации для каждого класса независимо (многометочная классификация).

Поведенческие состояния распознаются с Accuracy=0,91. Для выявления аномального поведения получено AUROC=0,68, Precision@10=0,81. Итоговые метрики классификации заболеваний на основе визуальных данных: Accuracy=0,85±0,04, Precision=0,85±0,06, Recall=0,72±0,06, F1–score=0,74±0,03.

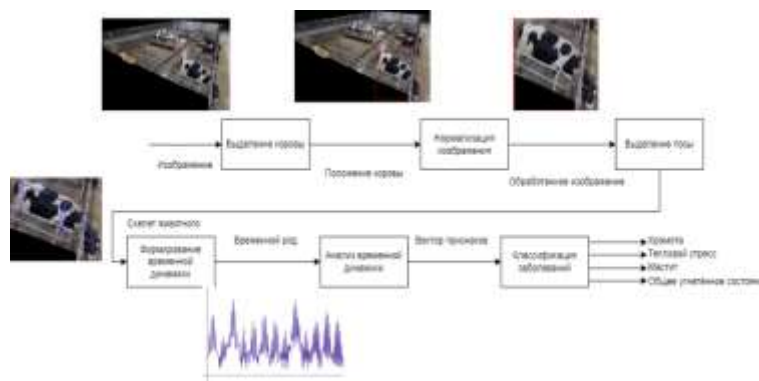


Рисунок 3 – Общая схема работы классификатора заболеваний на основе последовательности изображений

Архитектура системы.

Для объединения разработанных модулей спроектирована модульная архитектура (Рис. 4), включающая следующие подсистемы: загрузки и объединения данных (поддержка CSV, изображений, видео), предварительной обработки (очистка, нормализация, синхронизация временных рядов), извлечения признаков (CNN для изображений, агрегация сенсорных данных), генерации производных признаков (отклонения от индивидуальной нормы, индексы активности), анализа временной динамики (LSTM), классификации/регрессии (LightGBM, CatBoost, Random Forest, MLP), интерпретации (SHAP) и визуализации результатов. Архитектура обеспечивает гибкость (возможность замены отдельных моделей), масштабируемость (добавление новых источников данных и заболеваний) и интерпретируемость результатов за счёт SHAP-анализа.

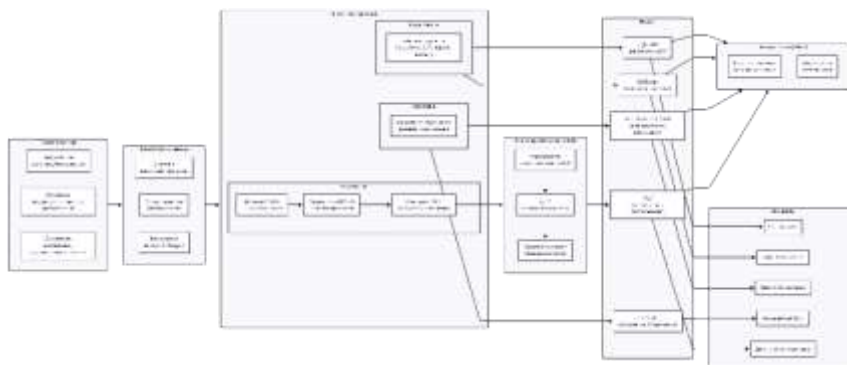


Рисунок 4 – Архитектура системы

Выводы

В работе разработаны методы и программная система для повышения продуктивности крупного рогатого скота на основе анализа продуктивных, сенсорных и визуальных данных. Для прогнозирования удоя и вкуса молока применены LightGBM и CatBoost ($R^2=0,805$, $Accuracy=0,714$). Для выявления заболеваний по сенсорным данным использованы MLP, Random Forest и LSTM-автоэнкодер ($F1-score=0,77$ для мастита). Для анализа видеопоследовательностей разработан конвейер из YOLO, HRNet, CNN, LSTM с механизмом внимания, достигнута $Accuracy=0,85$. Все модули объединены в единую масштабируемую архитектуру с SHAP-

интерпретацией. Система может быть использована для ранней диагностики и поддержки принятия решений в животноводстве. Направления дальнейшего развития включают интеграцию с IoT, расширение набора заболеваний и разработку веб–интерфейса.

Список литературы

1. Аннамария Б. Тепловой стресс. Контроль состояния дойных коров // Эффективное животноводство. – 2021. – № 3 (169).
2. Войтенко Л. Г. и др. Мастит. Диагностика. Методы лечения // Ветеринарная патология. – 2013. – № 4. – С. 9–13.
3. Воронов Д. В., Бобёр Ю. Н. Внутренние болезни у высокопродуктивных коров на молочно-товарных комплексах. – Гродно : ГГАУ, 2018. – С. 12–23.
4. Загидуллин Л. Р., Хисамов Р. Р., Шайдуллин Р. Р. Цифровизация молочного скотоводства на примере системы роботизированного доения // Техника и технологии в животноводстве. – 2021. – № 4(44). – С. 17–22.
5. Иванов Ю. Г., Лапкин А. Г. Сравнительная оценка энерго, трудо- и эксплуатационных затрат при переводе коров с доения в молокопровод на робот «Lely Astronaut» // Техника и технологии в животноводстве. – 2013. – № 3(11). – С. 188–190.
6. Кузьминова Е. В., Семененко М. П., Абрамов А. А. и др. Проблема теплового стресса в молочном животноводстве // Ветеринария Кубани. – 2020. – № 3. – С. 10–11.
7. Овсянкина Н. М., Прозоров А. А. Использование АСУ в производстве молока // Молочнохозяйственный вестник. – 2011. – № 1. – С. 81–86.
8. Симонов Ю. И., Симонова Л. Н. Ламинит коров и упитанность // Актуальные проблемы инновационного развития животноводства. – Брянск : Брянский государственный аграрный университет, 2019. – С. 156–160.

9. ICAR Guidelines for the Monitoring and Management of Ketosis in Dairy Cattle. – Rome : International Committee for Animal Recording, 2019.
10. ICAR Guidelines on Heat Stress Indicators in Dairy Cattle. – Rome : International Committee for Animal Recording, 2017. – 76 p.
11. ICAR Guidelines on Lameness Monitoring in Dairy Cattle. – Rome : International Committee for Animal Recording, 2020. – 88 p.
12. ICAR Guidelines on Somatic Cell Counts in Milk. – Rome : International Committee for Animal Recording, 2018. – 85 p.
13. Supper D. The AfiMilk system in modern farming // Сборник статей III Всероссийской (Национальной) научно-практической конференции. – Саратов, 2022. – С. 93–96.

References

1. Annamaria B. Heat stress. Monitoring the condition of dairy cows // Effective Animal Husbandry. – 2021. – No. 3 (169).
2. Voitenko L. G. et al. Mastitis. Diagnosis. Treatment methods // Veterinary Pathology. – 2013. – No. 4. – P. 9–13.
3. Voronov D. V., Bobyor Yu. N. Internal diseases in high-yielding cows at dairy commercial complexes. – Grodno : GGAU, 2018. – P. 12–23.
4. Zagidullin L. R., Khisamov R. R., Shaydullin R. R. Digitalization of dairy cattle farming on the example of a robotic milking system // Machinery and Technologies in Animal Husbandry. – 2021. – No. 4(44). – P. 17–22.
5. Ivanov Yu. G., Lapkin A. G. Comparative assessment of energy, labor, and operating costs when transitioning cows from pipeline milking to the Lely Astronaut robot // Machinery and Technologies in Animal Husbandry. – 2013. – No. 3(11). – P. 188–190.
6. Kuzminova E. V., Semenenko M. P., Abramov A. A. et al. The problem of heat stress in dairy farming // Veterinary Medicine of Kuban. – 2020. – No. 3. – P. 10–11.

7. Ovsyankina N. M., Prozorov A. A. Use of automated control systems in milk production // Dairy Farming Bulletin. – 2011. – No. 1. – P. 81–86.
8. Simonov Yu. I., Simonova L. N. Cow lameness and body condition // Current Problems of Innovative Development of Animal Husbandry. – Bryansk : Bryansk State Agrarian University, 2019. – P. 156–160.
9. ICAR Guidelines for the Monitoring and Management of Ketosis in Dairy Cattle. – Rome : International Committee for Animal Recording, 2019.
10. ICAR Guidelines on Heat Stress Indicators in Dairy Cattle. – Rome : International Committee for Animal Recording, 2017. – 76 p.
11. ICAR Guidelines on Lameness Monitoring in Dairy Cattle. – Rome : International Committee for Animal Recording, 2020. – 88 p.
12. ICAR Guidelines on Somatic Cell Counts in Milk. – Rome : International Committee for Animal Recording, 2018. – 85 p.
13. Supper D. The AfiMilk system in modern farming // Proceedings of the III All-Russian (National) Scientific and Practical Conference. – Saratov, 2022. – P. 93–96.