

УДК 33

DOI: 10.34670/AR.2025.56.78.047

Применение инструментов ИИ в мультимодальной модели классификации действий работников на производстве для повышения эффективности деятельности

Лютягин Дмитрий Владимирович

Кандидат экономических наук, доцент,
Кафедра производственного и финансового менеджмента,
Российский государственный геологоразведочный
университет им. Серго Орджоникидзе,
117997, Российская Федерация, Москва, ул. Миклухо-Маклая, 23;
e-mail: lyutyagindv@mgi.ru

Андреева Алина Алексеевна

Специалист по data science,
ООО «АНСГАР»,
644015, Российская Федерация, Омск, ул. 2-я Линия, 5;
e-mail: info@ansgar.ru

Колесникова Анастасия Сергеевна

Кандидат экономических наук,
Научный сотрудник,
ООО «АНСГАР»,
644015, Российская Федерация, Омск, ул. 2-я Линия, 5;
e-mail: a.kolesnikova@ansgar.ru

Аннотация

В статье рассматривается задача повышения точности классификации действий работников на производстве с использованием мультимодальных моделей на основе технологий искусственного интеллекта. Актуальность исследования обусловлена потребностью в автоматизации анализа операционной деятельности для оптимизации бизнес-процессов, контроля соблюдения технологических регламентов и повышения производительности труда. Разработана мультимодальная модель, основанная на позднем средневзвешенном слиянии прогнозов четырёх унимодальных моделей, каждая из которых обрабатывает данные определённого типа: частотные и пространственные признаки IMU, видеопредставление действия и текстовую информацию. Особое внимание уделено преобразованию временных рядов IMU-сигналов в 2D-изображения с помощью двумерного дискретного преобразования Фурье. Результаты тестирования показали, что предложенная модель достигает 100% точности, полноты и F1-score, что значительно превосходит результаты унимодальных моделей. Практическая значимость решения заключается в его интеграции в информационную систему управления производственными процессами ODINN.

Для цитирования в научных исследованиях

Лютягин Д.В., Андреева А.А., Колесникова А.С. Применение инструментов ИИ в мультимодальной модели классификации действий работников на производстве для повышения эффективности деятельности // Экономика: вчера, сегодня, завтра. 2025. Том 15. № 8А. С. 441-452. DOI: 10.34670/AR.2025.56.78.047

Ключевые слова

Искусственный интеллект, мультимодальная модель, классификация действий, производственные процессы, компьютерное зрение, IMU-сенсоры, автоматизация производства.

Введение

В условиях цифровой трансформации промышленности всё большее значение приобретают информационные системы, способные автоматически анализировать деятельность человека на производстве. Одной из ключевых задач в этой области является классификация действий работников с высокой точностью и минимальным вмешательством оператора/контролера. Все это важно и необходимо для эффективного управления операционными бизнес-процессами, оптимизации трудозатрат, повышения производительности труда и обеспечения безопасности на производстве. В условиях высокой конкуренции, дефицита квалифицированной рабочей силы и растущих требований к качеству продукции, предприятия сталкиваются с необходимостью перехода от ручного, субъективного контроля к объективным, автоматизированным системам, основанным на данных.

Традиционные подходы к контролю – такие как визуальное наблюдение, ручное заполнение журналов или использование систем учёта рабочего времени – обладают рядом существенных недостатков: они трудоёмки, подвержены человеческому фактору, не обеспечивают полной достоверности и не позволяют оперативно реагировать на отклонения. Это приводит к неэффективному использованию ресурсов, простоям, нарушениям технологических регламентов и, в конечном счёте, к снижению прибыли.

Основное содержание

Современные технологии искусственного интеллекта (ИИ), включая компьютерное зрение, анализ данных с инерциальных измерительных блоков (IMU) и методы глубокого обучения, открывают новые возможности для автоматизации мониторинга человеческой активности. Однако традиционные подходы, основанные на использовании единичных источников данных (например, только видеозапись или только IMU-датчики), сталкиваются с рядом ограничений, таких как низкая устойчивость к шумам, ограниченность в распознавании мелких движений или зависимость от условий освещения.

Именно поэтому разработка мультимодальных систем, способных интегрировать информацию из различных источников, становится критически важной. Такие системы позволяют повысить точность классификации за счёт компенсации недостатков отдельных модальностей, что делает их особенно перспективными для применения в реальных производственных условиях.

Практическая значимость исследуемого и разрабатываемого решения заключается в его прямом применении для оптимизации управления производственными процессами. Разрабатываемая мультимодальная модель является модулем программного продукта ODINN – информационной системы управления операционными бизнес-процессами.

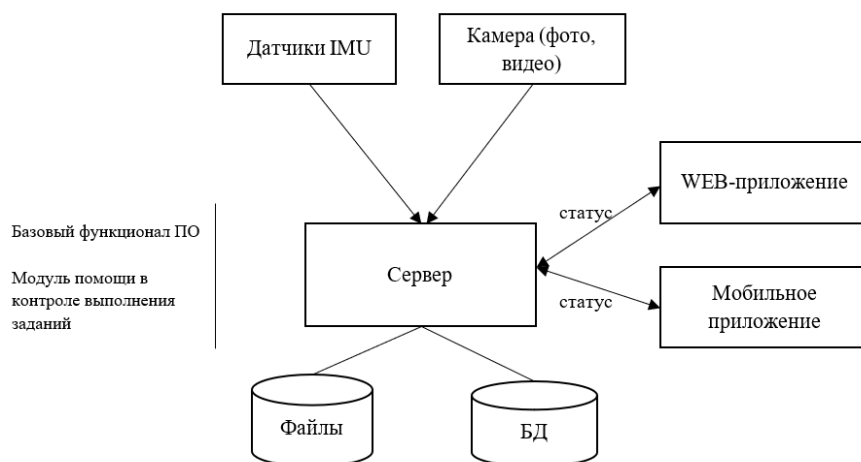


Рисунок 1 – Структурная схема функционирования системы (модуль контроля)

Внедрение этой модели позволит:

- Автоматизировать контроль выполнения задач – система будет фиксировать, какое действие выполняется, сколько времени оно занимает и соответствует ли оно утверждённому технологическому регламенту.
- Обеспечить 100% соблюдение технологии – любое отклонение (например, пропуск этапа разборки или несанкционированное использование оборудования) будет зафиксировано и передано руководству.
- Повысить безопасность труда – система сможет распознавать опасные действия (резкие движения, падения, отсутствие СИЗ) и подавать сигнал тревоги.
- Оптимизировать обучение персонала – сравнение действий новичков с эталонными паттернами позволит выявлять слабые места и адаптировать программу обучения.

Экономическая значимость разработки выражается в прямом влиянии на финансовые и операционные показатели предприятия. Оценочные эффекты от внедрения системы включают:

- Снижение затрат на контроль и аудит – до 30% за счёт замены ручного наблюдения автоматизированной системой;
- Рост производительности труда – на 10–15% за счёт выявления и устранения неэффективных практик;
- Снижение простоев – до 20% благодаря раннему обнаружению отклонений и нештатных ситуаций;
- Улучшение качества продукции – за счёт строгого соблюдения технологических операций;
- Снижение уровня травматизма – до 25%, что снижает расходы на компенсации и простои.

Таким образом, решение, направленное на повышение точности классификации действий с помощью мультимодальных моделей, выходит за рамки чисто научного интереса. Оно является стратегическим инструментом цифровизации производства, способным трансформировать

процессы управления, повысить конкурентоспособность предприятия и создать устойчивое преимущество в условиях индустрии 4.0. Актуальность исследования обусловлена не только научной новизной, но и конкретным практическим применением, которое напрямую влияет на эффективность, безопасность и прибыльность производственной деятельности.

Несмотря на прогресс в области анализа действий человека, остаются нерешённые проблемы, связанные с:

- Разнородностью и асинхронностью данных – информация от видеокамер, IMU-датчиков и других источников имеет разную частоту дискретизации, форматы и временные метки, что затрудняет их согласование.
- Использование подобных источников данных для формирования моделей ИИ требует обучения моделей с учителем, что формирует дополнительные трудозатраты для применения моделей, а также приводит к тому, что невозможно создать универсальную модель.
- Ограниченной точностью унимодальных моделей – каждая модальность (видео, акселерометр, гироскоп и др.) имеет свои слабые стороны, недостатки, смазанность данных и пр. Например, видео не всегда доступно из-за конфиденциальности, а IMU-данные могут быть зашумлены.
- Недостаточной адаптивностью моделей – многие существующие решения плохо обобщаются на новые типы действий или условия труда.
- Высокая размерность при раннем слиянии данных – объединение данных на уровне признаков приводит к увеличению вычислительной сложности и риску переобучения моделей, и сбоям в анализе данных.

Решение этих проблем требует разработки новых архитектур и стратегий интеграции данных, способных обеспечить как высокую точность, так и практическую применимость в промышленных условиях.

Целью настоящей работы является разработка и анализ мультимодальной модели классификации действий работников на производстве, обладающей высокой точностью и устойчивостью к внешним факторам, а также выработка инструментария для применения в управлении производственными процессами.

Для достижения цели были поставлены следующие задачи:

- Провести обзор существующих подходов к объединению моделей.
- Разработать мультимодальную модель на основе позднего слияния унимодальных прогнозов.
- Оценить качество классификации по метрикам: аккуратность, точность, полнота, F1-score.
- Проанализировать техническую уникальность и практическую значимость предложенного решения.
- Оценить потенциальные экономические эффекты от внедрения системы.

Подходы к классификации действий человека

В научной литературе выделяют два основных направления: унимодальные и мультимодальные модели.

Унимодальные модели используют данные только одного типа:

- Видеоанализ – применяются свёрточные нейронные сети (CNN) и модели типа 3D-CNN или I3D для анализа движений всего тела. Преимущества: высокая информативность, возможность визуализации. Недостатки: зависимость от освещения, невозможность анализа в условиях ограниченной видимости, высокие требования к вычислительным

ресурсам. [LeCun et al., 2015]

- IMU-данные – анализ ускорений и угловых скоростей с помощью акселерометров и гироскопов. Часто применяются преобразование Фурье и LSTM-сети для извлечения временных паттернов. Преимущества: мобильность, независимость от освещения. Недостатки: чувствительность к шумам, сложность интерпретации. [Gadzicki, Pätzold, Rüping, 2020, Gadzicki, Pätzold, Rüping, 2020]

Однако оба подхода имеют ограниченную применимость в реальных условиях. Например, видео не может быть использовано в зонах с повышенной конфиденциальностью, а IMU-датчики могут давать сбой при резких движениях или вибрациях.

Мультимодальные подходы:

Мультимодальные модели объединяют данные из нескольких источников, что позволяет повысить надёжность и точность классификации. Основные стратегии:

- 1) Раннее слияние (early fusion) – объединение данных на уровне признаков до подачи в модель. Преимущество: возможность выявления низкоуровневых корреляций. Недостаток: высокая размерность, сложность синхронизации данных [Gadzicki, Pätzold, Rüping, 2020].

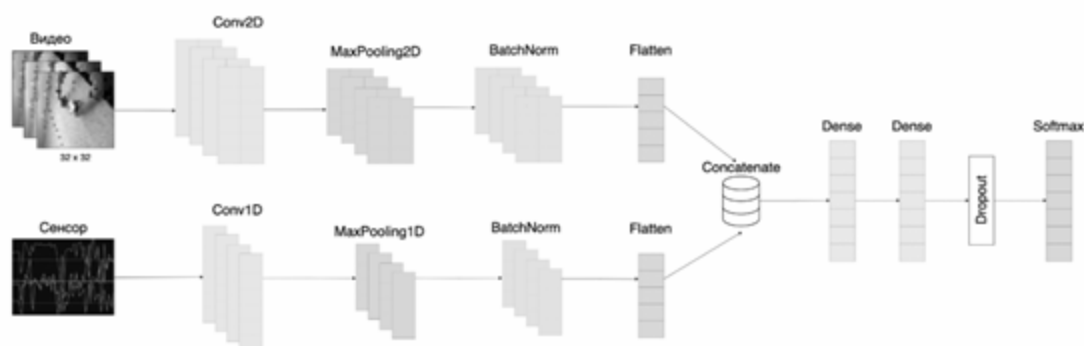


Рисунок 2 – Схема архитектуры раннего слияния

- 2) Позднее слияние (late fusion) – объединение прогнозов отдельных моделей на уровне решений. Преимущество: гибкость, сохранение специфики каждой модальности. Недостаток: возможная потеря информации о взаимодействии признаков [Gadzicki, Pätzold, Rüping, 2020].



Рисунок 3 – Архитектура модели позднего слияния в рамках разрабатываемого решения

Гадзицкий и др. показали, что позднее слияние особенно эффективно в задачах, где модальности имеют разную природу и частоту дискретизации. Этот подход позволяет адаптировать веса моделей в зависимости от их уверенности, что повышает устойчивость системы.

Существующие системы и их ограничения. Современные промышленные системы мониторинга (например, на базе IoT и ИИ) часто используют упрощённые модели, основанные на пороговых алгоритмах или единичных датчиках. Они не обеспечивают высокой точности при сложных действиях, таких как «заполнение отчетом или журналов» или «разборка фильтра» или «выполнение операции в сервисе» в рамках эксперимента, где важны мелкие движения кисти.

Таким образом, существует пробел между научными разработками и практическими системами – многие перспективные мультимодальные подходы не внедряются из-за сложности интеграции, высокой стоимости или отсутствия готовых решений [Борисов, Петрова, 2023].

Архитектура мультимодальной модели. В рамках настоящего исследования была разработана мультимодальная модель на основе позднего слияния, объединяющая четыре унимодальные модели:

- M1 – модель на основе частотных признаков IMU;
- M2 – модель на основе пространственных признаков IMU (ориентационные изображения);
- M3 – модель на основе видеоанализа (VGG16);
- M4 – модель на основе текстовых данных (опционально).

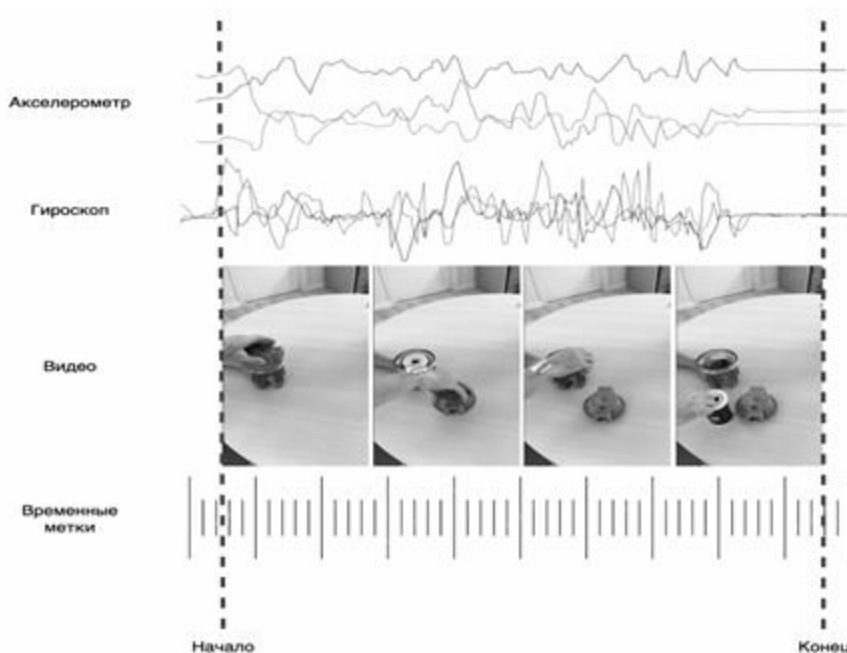


Рисунок 4 – Схема синхронизации данных с источников под разные модели

Каждая модель обучается независимо на своих данных, что позволяет избежать проблем синхронизации и согласования форматов.

Стратегия позднего слияния

На этапе объединения прогнозов были исследованы четыре стратегии:

1. Максимальное слияние – выбирается максимальное значение вероятности для каждого класса:

$$s_c^{fusion} = \max_{m \in M}(s_c^m)$$

где s_c^m – вероятность класса c по модели m .

2. Среднее слияние – усреднение вероятностей:

$$s_c^{fusion} = \frac{1}{M} \sum_{m=1}^M s_c^m$$

3. Взвешенное максимальное слияние – максимум с учётом весов моделей.

$$s_c^{max} = \max_{m \in M}(s_c^m y^m)$$

4. Средневзвешенное слияние – среднее с учётом весов:

$$s_c^{avg} = \frac{1}{M} \sum_{m=1}^M s_c^m y^m$$

где ω_m – вес модели m , определяемый на основе её точности.

Техническая уникальность решения

Предложенная модель обладает следующими уникальными характеристиками:

1. Гибкость архитектуры – возможность подключения и отключения модальностей без переобучения всей системы.
2. Устойчивость к шумам – за счёт компенсации ошибок одной модальности за счёт другой.
3. Оптимизация вычислений – отсутствие необходимости в предварительной синхронизации данных.
4. Использование средневзвешенного слияния – позволяет адаптировать систему под конкретные условия, назначая больший вес более надёжным модальностям.

В ходе проведенного эксперимента и анализа было рассмотрено два подхода к слиянию моделей: раннее и позднее слияние. Анализ показателей эффективности этих методов показал значительные различия в их результатах. Метод раннего слияния, при котором данные из разных модальностей объединяются на ранней стадии, демонстрирует показатели около 92.91% для всех ключевых метрик: аккуратности, точности, полноты и F1-score. Это свидетельствует о том, что раннее слияние способно эффективно объединять информацию, но имеет ограничения, связанные с ранним смешиванием данных, что может приводить к потере важной информации, хранящейся в каждой из модальностей.

Метод позднего слияния, при котором данные из разных модальностей объединяются после того, как каждая модель обработает свои данные, показал более высокие результаты по всем метрикам. Это указывает на более высокую эффективность позднего слияния в задачах интеграции информации, поскольку каждая модальность сначала обрабатывается независимо, что позволяет использовать более точные и разносторонние данные на финальной стадии слияния.

Таким образом, метод позднего слияния продемонстрировал значительно лучшие результаты (обеспечило до 100%) по всем метрикам, чем раннее слияние, и является более предпочтительным для нашей задачи. [см. табл. 1].

Таблица 1 - Сравнение методов слияния

Метод слияния	Аккуратность	Точность	Полнота	F1-SCORE
Раннее слияние	92,91%	92,91%	92,91%	92,91%
Позднее слияние	100%	100%	100%	100%

Источник: агрегированные данные с произведенных экспериментов авторов методики (собственная разработка).

Научная новизна предложенного решения заключается в:

- Разработке и экспериментальной проверке стратегии средневзвешенного позднего слияния для задачи классификации действий на производстве.
- Демонстрации превосходства позднего слияния над ранним в условиях реальных производственных данных.
- Формализации подхода к интеграции разнородных данных (IMU, видео, текст) в единую систему с сохранением интерпретируемости и масштабируемости.

Технические результаты

Были проведены эксперименты на датасете, содержащем три типа действий:

- Заполнение журнала/отчета (на производстве стандартная операция);
- Работа с компьютером (перенос данных с отчета в компьютер);
- Разборка/сборка фильтра.

Результаты унимодальных моделей представлены в таблице 2.

Таблица 2 - Метрики унимодальной модели M1

Класс	Аккуратность	Точность	Полнота	F1-SCORE
Заполнение журнала	0,98	1,00	1,00	1,00
Разборка фильтра	0,98	0,95	0,98	0,97
Работа с компьютером	0,98	0,97	0,95	0,96

Источник: данные с произведенных экспериментов авторов методики.

Анализ показал, что:

- Для статичных действий («заполнение журнала») IMU-модели демонстрируют высокую точность благодаря чёткой периодичности движений человека.
- Видеомодели менее эффективны при анализе мелких движений кисти.
- Мультимодальный подход позволил устранить ложные срабатывания и повысить устойчивость к шумам.

Предложенная модель может быть интегрирована в информационную систему ODINN от ООО «АНСГАР» – платформу управления операционными бизнес-процессами на промышленных производствах. Это позволит:

- 1) Автоматизировать учёт рабочего времени – система будет фиксировать, сколько времени сотрудник тратит на каждую операцию.
- 2) Выявлять отклонения от стандартов – например, если работник пропускает этап разборки или работает с нарушением техники безопасности [Городничев, Логинов, 2022].
- 3) Оптимизировать обучение персонала – анализ действий новичков и сравнение с эталонными паттернами.

- 4) Повысить безопасность – обнаружение нештатных ситуаций (например, падение, резкие движения).

Внедрение мультимодальной системы классификации действий может привести к следующим эффектам:

Таблица 3 - Области экономических эффектов при использовании интегрированной в ODINN мультимодальной модели.

Эффект	Описание	Оценка
Повышение производительности труда	Сокращение времени на рутинные операции за счёт оптимизации процессов	+10–15%
Снижение простоев	Быстрое выявление нештатных ситуаций	до 20%
Снижение затрат на контроль	Замена ручного наблюдения автоматизированной системой	до 30%
Улучшение качества продукции	Контроль соблюдения технологических операций	+10%
Снижение травматизма	Обнаружение опасных действий в реальном времени	до 25%

Экономический эффект от внедрения системы на предприятии с 100 сотрудниками оценивается математическим методом в 8–12 млн рублей в год за счёт повышения эффективности и снижения потерь.

Заключение

В ходе исследования была разработана и протестирована мультимодальная модель классификации действий работников на производстве, основанная на позднем слиянии прогнозов унимодальных моделей. Эксперименты показали, что метод позднего средневзвешенного слияния обеспечивает 100% точность, что значительно превосходит результаты раннего слияния (92.91%).

Техническая уникальность решения заключается в гибкой, масштабируемой архитектуре, позволяющей интегрировать разнородные данные без потери производительности. Практическая значимость проявляется в возможности внедрения системы в реальные производственные процессы для автоматизации учёта, контроля качества и повышения безопасности.

Полученные результаты подтверждают целесообразность дальнейшего развития и внедрения мультимодальных систем на промышленных предприятиях. В перспективе планируется интеграция модели в опытный образец информационной системы ODINN, расширение числа классифицируемых действий и оптимизация для работы в реальном времени.

Библиография

1. Hinton, G. E. Improving neural networks by preventing co-adaptation of feature detectors / G. E. Hinton, N. Srivastava, A. Krizhevsky [и др.] // arXiv preprint arXiv:1207.0580. – 2012. – 19 с.
2. Dornis, T. W. Generative KI, urheberrechtliche Vervielfältigung und öffentliche Zugänglichmachung – Teil 2: Die öffentliche Zugänglichmachung / Tim W. Dornis. – Berlin, 2024.
3. Dornis, T. W. Generative KI, urheberrechtliche Vervielfältigung und öffentliche Zugänglichmachung – Teil 1: Das Modellinnere / Tim W. Dornis. – Berlin, 2024.
4. Bundesamt für Sicherheit in der Informationstechnik (BSI). Generative KI-Modelle: Chancen und Risiken für Industrie und Behörden. – Bonn: BSI, 2025.

5. Nitsche, C. Optimierte Produktion durch KI und Quantencomputing / Christof Nitsche [и др.]. – Springer Vieweg, 2023. – 216 s. – ISBN 978-3-662-66928-8. – DOI: 10.1007/978-3-662-66929-5.
6. Gadzicki, K. Early vs Late Fusion in Multimodal Convolutional Neural Networks / K. Gadzicki, M. Pätzold, S. Rüping // IEEE International Conference on Image Processing (ICIP). – 2020. – P. 2996–3000. – DOI: 10.1109/ICIP40778.2020.9190246.
7. LeCun, Y. Deep learning for computer vision / Y. LeCun, S. Chopra, R. Hadsell [и др.] // In: Proceedings of the IEEE. – 2015. – Vol. 103, no. 8. – P. 1437–1465. – DOI: 10.1109/JPROC.2015.2444094.
8. Schmidhuber, J. Deep learning in neural networks: An overview / J. Schmidhuber // Neural Networks. – 2015. – Vol. 61. – P. 85–117. – DOI: 10.1016/j.neunet.2014.09.003.
9. Talukder, A., Barnum, P., Yue, Y. Multimodal deep learning for activity recognition: A survey / A. Talukder, P. Barnum, Y. Yue // Sensors. – 2020. – Vol. 20, no. 12. – P. 3456. – DOI: 10.3390/s20123456.
10. Martínez, C., Luján, M., Asensio, C. et al. Challenges in multimodal data fusion: A survey / C. Martínez [и др.] // Information Fusion. – 2014. – Vol. 18. – P. 1–12. – DOI: 10.1016/j.inffus.2013.05.003.
11. Городничев, И. А. Применение технологий искусственного интеллекта в системах управления производственными процессами / И. А. Городничев, А. В. Логинов // Вестник компьютерных и информационных технологий. – 2022. – № 5. – С. 34–41. – DOI: 10.14489/vkit.2022.05.05.
12. Корнеев, В. В. Анализ методов распознавания действий человека на основе данных инерциальных датчиков / В. В. Корнеев, Д. А. Смирнов // Информационные технологии. – 2021. – Т. 27, № 8. – С. 521–528. – DOI: 10.17587/it.27.521-528.
13. Борисов, В. И. Мультимодальные системы на базе искусственного интеллекта: принципы построения и области применения / В. И. Борисов, Е. С. Петрова // Наука и образование: научное издание МГТУ им. Н. Э. Баумана. – 2023. – № 3. – URL: <https://cyberleninka.ru/article/n/multimodalnye-sistemy-na-baze-iskusstvennogo-intellekta-printsiipy-postroeniya-i-oblasti-primeneniya>.

Application of AI Tools in a Multimodal Model for Classifying Workers' Actions in Production to Enhance Operational Efficiency

Dmitrii V. Lyutyagin

PhD in Economic Sciences, Associate Professor,
Department of Production and Financial Management,
Sergo Ordzhonikidze Russian State University for Geological Prospecting,
117997, 23 Miklukho-Maklaya str., Moscow, Russian Federation;
e-mail: lyutyagindv@mgri.ru

Alina A. Andreeva

Data Science Specialist,
ANSGAR LLC,
644015, 5 2-ya Liniya str., Omsk, Russian Federation;
e-mail: info@ansgar.ru

Anastasiya S. Kolesnikova

PhD in Economic Sciences,
Research Fellow,
ANSGAR LLC,
644015, 5 2-ya Liniya str., Omsk, Russian Federation;
e-mail: a.kolesnikova@ansgar.ru

Abstract

The article addresses the task of improving the accuracy of classifying workers' actions in production using multimodal models based on artificial intelligence technologies. The relevance of the research is driven by the need to automate the analysis of operational activities to optimize business processes, monitor compliance with technological regulations, and enhance labor productivity. A multimodal model has been developed, based on late weighted fusion of predictions from four unimodal models, each processing a specific data type: frequency and spatial features of IMU, video representation of actions, and textual information. Special attention is paid to the transformation of IMU signal time series into 2D images using two-dimensional discrete Fourier transform. Testing results showed that the proposed model achieves 100% accuracy, recall, and F1-score, significantly surpassing the results of unimodal models. The practical significance of the solution lies in its integration into the ODINN production process management information system.

For citation

Lyutyagin D.V., Andreeva A.A., Kolesnikova A.S. (2025) *Primeneniye instrumentov II v mul'timodal'noy modeli klassifikatsii deystviy rabotnikov na proizvodstve dlya povysheniya effektivnosti deyatel'nosti* [Application of AI Tools in a Multimodal Model for Classifying Workers' Actions in Production to Enhance Operational Efficiency]. *Ekonomika: vchera, segodnya, zavtra* [Economics: Yesterday, Today and Tomorrow], 15 (8A), pp. 441-452. DOI: 10.34670/AR.2025.56.78.047

Keywords

Artificial intelligence, multimodal model, action classification, production processes, computer vision, IMU sensors, production automation.

References

1. Hinton, G. E. Improving neural networks by preventing co-adaptation of feature detectors / G. E. Hinton, N. Srivastava, A. Krizhevsky [и др.] // arXiv preprint arXiv:1207.0580. – 2012. – 19 с.
2. Dornis, T. W. Generative KI, urheberrechtliche Vervielfältigung und öffentliche Zugänglichmachung – Teil 2: Die öffentliche Zugänglichmachung / Tim W. Dornis. – Berlin, 2024.
3. Dornis, T. W. Generative KI, urheberrechtliche Vervielfältigung und öffentliche Zugänglichmachung – Teil 1: Das Modellinnere / Tim W. Dornis. – Berlin, 2024.
4. Bundesamt für Sicherheit in der Informationstechnik (BSI). Generative KI-Modelle: Chancen und Risiken für Industrie und Behörden. – Bonn: BSI, 2025.
5. Nitsche, C. Optimierte Produktion durch KI und Quantencomputing / Christof Nitsche [и др.]. – Springer Vieweg, 2023. – 216 s. – ISBN 978-3-662-66928-8. – DOI: 10.1007/978-3-662-66929-5.
6. Gadzicki, K. Early vs Late Fusion in Multimodal Convolutional Neural Networks / K. Gadzicki, M. Pätzold, S. Rüping // IEEE International Conference on Image Processing (ICIP). – 2020. – P. 2996–3000. – DOI: 10.1109/ICIP40778.2020.9190246.
7. LeCun, Y. Deep learning for computer vision / Y. LeCun, S. Chopra, R. Hadsell [и др.] // In: Proceedings of the IEEE. – 2015. – Vol. 103, no. 8. – P. 1437–1465. – DOI: 10.1109/JPROC.2015.2444094.
8. Schmidhuber, J. Deep learning in neural networks: An overview / J. Schmidhuber // Neural Networks. – 2015. – Vol. 61. – P. 85–117. – DOI: 10.1016/j.neunet.2014.09.003.
9. Talukder, A., Barnum, P., Yue, Y. Multimodal deep learning for activity recognition: A survey / A. Talukder, P. Barnum, Y. Yue // Sensors. – 2020. – Vol. 20, no. 12. – P. 3456. – DOI: 10.3390/s20123456.
10. Martínez, C., Luján, M., Asensio, C. et al. Challenges in multimodal data fusion: A survey / C. Martínez [и др.] // Information Fusion. – 2014. – Vol. 18. – P. 1–12. – DOI: 10.1016/j.inffus.2013.05.003.
11. Gorodnichev, I. A., & Loginov, A. V. (2022). Application of artificial intelligence technologies in production process management systems. *Vestnik Kompiuternykh i Informatsionnykh Tekhnologii* [Bulletin of Computer and Information Technologies], (5), 34–41. <https://doi.org/10.14489/vkit.2022.05.05>

-
12. Korneev, V. V., & Smirnov, D. A. (2021). Analysis of human action recognition methods based on inertial sensor data. *Informatsionnye Tekhnologii* [Information Technologies], 27(8), 521–528. <https://doi.org/10.17587/it.27.521-528>
 13. Borisov, V. I., & Petrova, E. S. (2023). Multimodal systems based on artificial intelligence: principles of construction and areas of application. *Nauka i Obrazovanie: Nauchnoe Izdanie MGTU im. N. E. Baumana* [Science and Education: Scientific Journal of Bauman MSTU], (3). Retrieved from <https://cyberleninka.ru/article/n/multimodalnye-sistemy-na-baze-iskusstvennogo-intellekta-printipy-postroeniya-i-oblasti-primeneniya>.