

**АВТОМАТИЧЕСКАЯ ОЦЕНКА КЛИНИЧЕСКОЙ
КОМПЕТЕНТНОСТИ СТУДЕНТОВ - МЕДИКОВ ПО ДАННЫМ
ТЕЛЕМЕТРИИ ВИРТУАЛЬНОЙ РЕАЛЬНОСТИ**

**AUTOMATIC ASSESSMENT OF MEDICAL STUDENTS' CLINICAL
COMPETENCE BASED ON VIRTUAL REALITY TELEMETRY DATA**

Talipova Ozoda Khabirovna

Tashkent University of Information Technologies Senior Lecturer,
Department of "Multimedia Technologies"

Nazirov Abror Shodmonqulovich

Tashkent University of Information Technologies Senior Lecturer,
Department of "Multimedia Technologies"

Juraev Azizbek Isroil ugli

Tashkent University of Information Technologies Assistant,
Department of "Multimedia Technologies"

Fozilov Samandar Dilshod ugli

student of the "Central Asian University"

Талипова Озода Хабировна

Ташкентский Университет Информационных Технологий
Старший преподаватель кафедры «Мультимедийные технологии»

Назиров Аброр Шодмонкулович

Ташкентский Университет Информационных Технологий
Старший преподаватель кафедры «Мультимедийные технологии»

Жураев Азизбек Исроил угли

Ташкентский Университет Информационных Технологий
Ассистент кафедры «Мультимедийные технологии»

Фозилов Самандар Дилшод угли

студент «Среднеазиатского университета»

ABSTRACT

Conventional assessment of clinical skills such as OSCE and Mini-CEX is resource-intensive and inevitably subjective, whereas virtual-reality environments record every learner action as high-frequency objective telemetry, and this study proposes and validates a method for the automatic assessment of medical students' clinical competency from behavioural VR data. A four-stage pipeline acquires a 60 Hz telemetry stream (hand position, gaze direction, reaction times), extracts four feature groups - procedural accuracy, communication, movement economy, and safety - combines them through a weighted sum, and estimates the entrustment level by logistic regression; the method was tested on 90 sessions, each independently rated by two OSCE examiners on a 1-5 scale. Agreement between

the automatic score and averaged expert ratings was high (Pearson $r = 0.86$; intraclass correlation ICC = 0.84; mean absolute error 0.34 points), and inter-rater agreement between the two experts (**Cohen's $\kappa = 0.79$**) was comparable to the system - expert agreement, indicating that automated scoring performs within the range of normal inter-rater variability. These results show that behavioural telemetry carries sufficient information for objective, reproducible competency assessment at a level acceptable for formative evaluation, reducing routine grading load and providing immediate feedback while remaining advisory in high-stakes decisions where expert judgement should prevail.

Keywords: *automatic assessment, telemetry, virtual reality, clinical competency, feature extraction, logistic regression, OSCE.*

АННОТАЦИЯ

Традиционная оценка клинических навыков, такая как OSCE и Mini-CEX, является ресурсоемкой и неизбежно субъективной, в то время как среды виртуальной реальности фиксируют каждое действие обучающегося как высокочастотную объективную телеметрию, и данное исследование предлагает и проверяет метод автоматической оценки клинической компетентности студентов-медиков на основе поведенческих данных VR. Четырехступенчатый трубопровод получает телеметрический поток частотой 60 Гц (положение рук, направление взгляда, время реакции), извлекает четыре группы признаков - точность процедур, коммуникацию, экономию движения и безопасность - объединяет их через взвешенную сумму и оценивает уровень доверия с помощью логистической регрессии; метод был протестирован на 90 сессиях, каждая из которых была независимо оценена двумя экзаменаторами ОБСЕ по шкале 1-5. Согласованность между автоматическим баллом и усредненными экспертными рейтингами была высокой (Pearson $r = 0,86$; внутриклассовая корреляция ICC = 0,84; средняя абсолютная ошибка 0,34 балла), а межрейтерское согласие между двумя экспертами (Cohen's $\kappa = 0,79$) было сопоставимо с системой - экспертным соглашением, что свидетельствует о том, что автоматизированное

оценивание выполняется в пределах нормальной межрейтерной вариабельности. Эти результаты показывают, что поведенческая телеметрия содержит достаточную информацию для объективной, воспроизводимой оценки компетентности на уровне, приемлемом для формирующей оценки, снижая нагрузку на рутинное оценивание и обеспечивая немедленную обратную связь, сохраняя при этом консультативную значимость при принятии важных решений, где преобладает экспертное суждение.

***Ключевые слова:** автоматическая оценка, телеметрия, виртуальная реальность, клиническая компетентность, извлечение признаков, логистическая регрессия, OSCE*

ВВЕДЕНИЕ

Традиционная оценка клинических навыков (OSCE, Mini-CEX) требует значительных ресурсов экспертов и неизбежно содержит элемент субъективности [1,2]. Среда виртуальной реальности открывает принципиально новую возможность: каждое действие обучающегося фиксируется в виде объективной телеметрии высокой частоты, которую можно превратить в количественные показатели компетентности. Идея оценивать мастерство по кинематике движений уже подтвердила состоятельность в хирургическом обучении, где параметры траектории инструментов надёжно различают уровни экспертизы [3,4]. Настоящая работа переносит этот подход в более широкий контекст клинической подготовки и описывает метод автоматической оценки компетентности по VR-телеметрии, а также проверяет его согласованность с суждением экспертов.

МЕТОД ОЦЕНКИ

Конвейер оценки состоит из четырёх ступеней (рис. 1): сбор сырой телеметрии, извлечение признаков, нормировка и агрегация, вычисление интегрального балла. На вход поступает поток частотой 60 Гц, содержащий координаты рук, направление взгляда и временные метки действий.

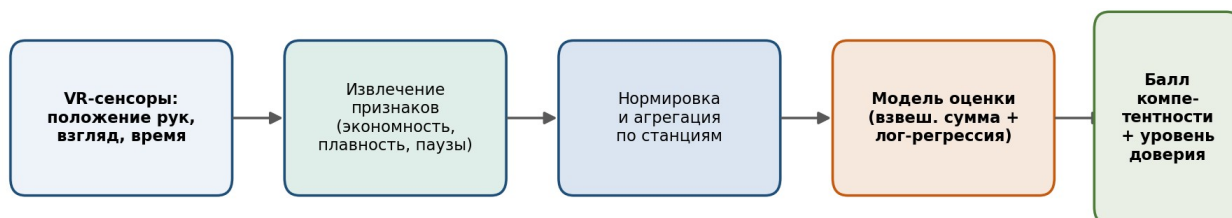


Рис. 1. Блок-схема конвейера автоматической оценки компетентности

Из телеметрии вычисляются четыре группы признаков: процедурная точность A_{proc} , качество коммуникации A_{comm} , экономность (эффективность) действий A_{effic} и безопасность A_{safe} . Интегральный балл формируется как взвешенная сумма:

$$S = 100(w_1A_{proc} + w_2A_{comm} + w_3A_{effic} + w_4A_{safe}) \quad (1)$$

Показатель экономности отражает, насколько траектория действий студента близка к оптимальной; он убывает экспоненциально с ростом избыточной длины пути:

$$A_{effic} = \exp\left(-\lambda \frac{L_{path} - L_{opt}}{L_{opt}}\right) \quad (2)$$

где L_{path} - фактическая длина траектории (или число шагов), L_{opt} - эталонная длина, λ - коэффициент чувствительности. Наконец, вероятность того, что студенту можно доверить самостоятельное выполнение задачи (уровень доверия ЕРА), оценивается логистической регрессией по агрегированным признакам:

$$P(entrust) = \frac{1}{1 + e^{-(\beta_0 + \sum_k \beta_k A_k)}} \quad (3)$$

Веса w_k в (1) и коэффициенты β_k в (3) откалиброваны по обучающей выборке сессий с известными экспертными оценками. Перечень признаков и их источники приведены в табл. 1.

Таблица 1. Признаки, извлекаемые из VR-телеметрии

Группа признаков	Показатель	Источник в телеметрии
Процедурная точность	доля верных шагов, порядок действий	последовательность событий
Коммуникация	своевременность реплик, зрительный контакт	взгляд, аудио-метки
Экономность	длина траектории рук, число лишних действий	координаты рук (60 Гц)
Безопасность	число опасных/пропущенных действий	маркеры сценария
Темп	время до ключевых решений	временные метки

РЕЗУЛЬТАТЫ

Метод апробирован на 90 сессиях, каждая из которых параллельно оценивалась двумя независимыми экспертами по шкале OSCE (1–5). Согласованность автоматической оценки с усреднённой оценкой экспертов оказалась высокой: коэффициент корреляции Пирсона $r = 0,86$, внутриклассовый коэффициент корреляции ICC = 0,84 (рис. 2). Согласие двух экспертов между собой (κ Коэна = 0,79) было сопоставимо с согласием «система - эксперт», что говорит о том, что автоматическая оценка не уступает межэкспертной вариабельности.

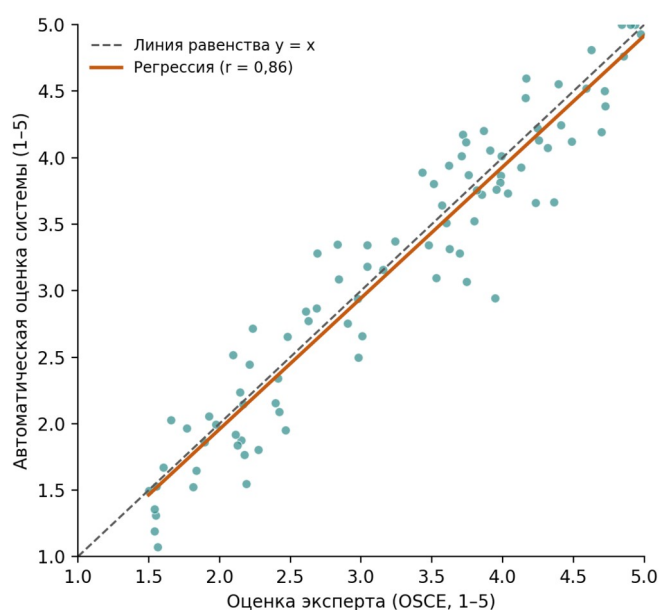


Рис. 2. Согласованность автоматической оценки системы и оценки экспертов (90 сессий)

Показатели согласованности вычислялись по классическим формулам - каппа Коэна для категориальных оценок и внутриклассовый коэффициент для непрерывных:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (4)$$

$$ICC = \frac{\sigma_{subj}^2}{\sigma_{subj}^2 + \sigma_{err}^2} \quad (5)$$

Таблица 2. Метрики согласованности автоматической и экспертной оценки

Показатель согласованности	Значение	Интерпретация
Корреляция Пирсона (r)	0,86	сильная связь
ICC (система - эксперт)	0,84	высокое согласие
κ Коэна (эксперт - эксперт)	0,79	существенное согласие
Средняя абс. ошибка (MAE)	0,34 балла	в пределах шкалы OSCE

ОБСУЖДЕНИЕ И ЗАКЛЮЧЕНИЕ

Полученные результаты показывают, что поведенческая телеметрия VR-среды несёт достаточную информацию для объективной оценки клинической компетентности: согласованность с экспертами (ICC = 0,84) соответствует уровню, приемлемому для формирующего оценивания [4]. Практическая ценность метода в том, что он снимает с преподавателя рутинную нагрузку и обеспечивает мгновенную, воспроизводимую обратную связь. Вместе с тем следует признать ограничения: выборка охватывала одну клиническую тему, а «эталонная» траектория L_opt задавалась экспертно и может нуждаться в уточнении. Кроме того, автоматическая оценка не должна полностью заменять экспертное суждение в решениях с высокими ставками — она призвана его дополнять. Дальнейшая работа связана с интеграцией метода в целостную систему и её внедрением в учебный процесс, чему посвящена заключительная статья цикла.

СПИСОК ЛИТЕРАТУРЫ

1. Norcini J.J., Blank L.L., Duffy F.D., Fortna G.S. The mini-CEX: a method for assessing clinical skills // *Annals of Internal Medicine*. – 2003. – Vol. 138, No. 6. – P. 476–481.
2. Epstein R.M. Assessment in medical education // *New England Journal of Medicine*. – 2007. – Vol. 356, No. 4. – P. 387–396.
3. Winkler-Schwartz A., Bissonnette V., Mirchi N. et al. Artificial intelligence in medical education: best practices using machine learning to assess surgical expertise // *Journal of Surgical Education*. – 2019. – Vol. 76, No. 6. – P. 1681–1690.
4. J. Azizbek, K. Uchqun, D. Nurjabova and T. Ozoda, “Deep Genomic Feature Extraction Using Multimodal Neural Networks and Alzheimer’s Disease Susceptibility Prediction Role of Blood Vessels in Alzheimer’s Disease,” 2026 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 2026, pp. 1-7, doi: 10.1109/ESCI68015.2026.11493337.
5. Koo T.K., Li M.Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research // *Journal of Chiropractic Medicine*. – 2016. – Vol. 15, No. 2. – P. 155–163.
6. Cohen J. A coefficient of agreement for nominal scales // *Educational and Psychological Measurement*. – 1960. – Vol. 20, No. 1. – P. 37–46.
7. Cook D.A., Hatala R., Brydges R. et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis // *JAMA*. – 2011. – Vol. 306, No. 9. – P. 978–988.
8. Топол Э. Искусственный интеллект в медицине / Пер. с англ. – М.: Альпина Паблишер, 2022. – 512 с.