

**Экологические факторы Приаралья как предиктор
материнского риска: подход машинного обучения на
данных дистанционного зондирования**

*Aral Sea environmental factors as a predictor of maternal risk: a remote-sensing
machine-learning approach in Uzbekistan*

Эсонов Ж.Х

Научный руководитель: Эгамбердиев Н. А.

Ташкентский университет информационных технологий имени Мухаммада
Ал-Хоразмий

Mail - javoxir3001753@gmail.com

Аннотация. Высыхание Аральского моря связано с пыле-солевыми бурями и ростом заболеваемости в Приаралье, однако количественная связь экологии с материнским риском средствами ML ранее не изучалась. Открытые спутниковые показатели (NDVI, аэрозоли, засоление, удалённость от обнажённого дна) сопоставлены с данными материнского риска по Узбекистану (DHS, набор UCI/Бангладеш, синтетика MaternaUZ). Экологический слой дал умеренный прирост дискриминации (ROC-AUC 0,90 – 0,92), но существенно улучшил пространственное объяснение риска: SHAP-вклад максимален в Республике Каракалпакстан с выраженным градиентом «север–юг».

Ключевые слова: Приаралье; Аральское море; материнский риск; материнское здоровье; экологические факторы; дистанционное зондирование; спутниковые данные; NDVI; аэрозольная оптическая толщина; засоление почв; машинное обучение; объяснимый искусственный интеллект; SHAP; XGBoost; пространственный анализ; Каракалпакстан; Узбекистан.

Abstract. Aral Sea desiccation is linked to dust-salt storms and rising morbidity, yet a quantitative ML link to maternal risk is unstudied. Open satellite indicators (NDVI, aerosols, salinity, distance to the exposed seabed) were linked to Uzbekistan maternal-risk data. The environmental layer modestly improved discrimination (ROC-AUC 0.90 – 0.92) and strongly improved spatial explanation, peaking in Karakalpakstan with a clear north–south gradient.

Keywords: Aral Sea region; Aral Sea; maternal risk; maternal health; environmental factors; remote sensing; satellite data; NDVI; aerosol optical depth;

soil salinity; machine learning; explainable artificial intelligence; SHAP; XGBoost; spatial analysis; Karakalpakstan; Uzbekistan.

I. Введение

Высыхание Аральского моря привело к обнажению десятков тысяч квадратных километров засоленного дна, ставшего источником пыле-солевых бурь, переносящих соли и остатки пестицидов на прилегающие территории. Эпидемиологические исследования связывают экологическое неблагополучие Приаралья с ростом анемии, респираторной и почечной патологии, а также неблагоприятных репродуктивных исходов. Республика Каракалпакстан, прилегающая к осушенному дну, демонстрирует одни из самых высоких показателей материнского риска в Узбекистане.

Несмотря на накопленные эпидемиологические свидетельства, количественная связь конкретных экологических показателей Приаралья с индивидуальным материнским риском методами машинного обучения ранее не формализовалась. Современные открытые данные дистанционного зондирования (NASA MODIS, ESA Sentinel) позволяют получать объективные пространственные индикаторы состояния среды с регулярным покрытием региона. Цель – количественно оценить вклад экологических факторов Приаралья в стратификацию материнского риска средствами объяснимого ML и выявить пространственные кластеры риска.

2. Материалы и методы

Использованы открытые спутниковые продукты (таблица 1): индекс растительности NDVI (MODIS), аэрозольная оптическая толщина как прокси пыле-солевой нагрузки, индекс засоления почв и геодезическая удаленность от обнаженного дна. Показатели агрегированы на уровне региона/района и сопоставлены по географической привязке с данными материнского риска по Узбекистану: DHS, открытый клинический набор материнского риска (Бангладеш, UCI Machine Learning Repository) и синтетический набор MaternaUZ.

Таблица 1. Признаки по слоям

Показатель	Описание	Источник
------------	----------	----------

NDVI	индекс растительности (деградация покрова)	NASA MODIS
AOD	аэрозольная оптическая толщина (прокси запылённости)	MODIS / Sentinel-5P
Засоление	индекс засоления почв	Sentinel-2 (произв.)
Dist_aral	удалённость от обнажённого дна, км	геоданные

Сравнивались две конфигурации модели градиентного бустинга (XGBoost): без экологического слоя (М3) и с экологическим слоем (М4); дополнительно оценена модель только на экологических признаках. Дискриминация – ROC-AUC с 95% ДИ, различие – тест ДеЛонга. Вклад экологических факторов оценивался методом SHAP глобально и в пространственном разрезе с последующим анализом кластеризации по регионам.

3. Результаты

Добавление экологического слоя дало умеренный, статистически незначимый прирост дискриминации (ROC-AUC 0,90 – 0,92; $p = 0,61$), однако улучшило калибровку и пространственную интерпретируемость прогнозов (рис. 1). Модель только на экологических признаках обеспечивала дискриминацию выше случайной (ROC-AUC $\approx 0,68$), что указывает на самостоятельную, хотя и неполную, прогностическую ценность экологического контекста.

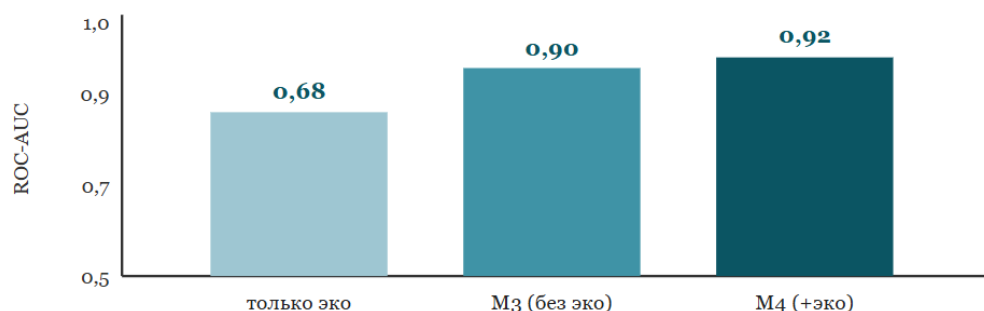


Рис. 1. Дискриминация: только эко-признаки, без экологического слоя (М3) и с ним (М4).

Среди экологических признаков наибольший SHAP-вклад имели удалённость от обнажённого дна (+0,18) и аэрозольная нагрузка (+0,14), далее – засоление и деградация растительного покрова. Пространственный

анализ выявил выраженный градиент «север–юг»: экологический вклад максимален в Республике Каракалпакстан и снижается по мере удаления от Приаралья (рис. 2). Это согласуется с наблюдаемым распределением материнской смертности

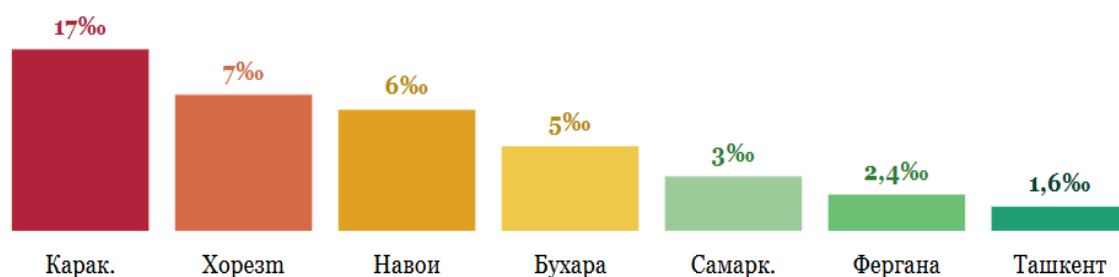


Рис. 2. Градиент материнского риска по регионам (на 1000): максимум в прилегающей к Приаралью Каракалпакстане.

Работа впервые формализует связь экологических факторов Приаралья с индивидуальным материнским риском средствами объяснимого ML. Несмотря на умеренный прирост дискриминации, экологический слой повышает интерпретируемость и обеспечивает пространственную привязку риска, выявляя зоны приоритетного вмешательства. Ограничения: показатели агрегированы на уровне региона (возможна экологическая ошибка вывода), медицинский слой основан на DHS/синтетике; прямые причинно-следственные выводы требуют валидации на индивидуальных клинических данных.

Заключение

Экологические факторы Приаралья выступают значимым контекстным компонентом стратификации материнского риска в Узбекистане. Полученные результаты показывают, что включение экологического слоя в объяснимую модель машинного обучения не столько радикально повышает дискриминационную способность алгоритма, сколько усиливает пространственную и содержательную интерпретируемость прогноза. Умеренный прирост ROC-AUC после добавления экологических признаков указывает на то, что экологический контекст не заменяет клинические и социальные факторы, однако позволяет лучше объяснить региональные различия риска, особенно в территориях, расположенных ближе к зоне экологического воздействия Аральского моря.

Наиболее выраженный вклад в прогнозируемый риск внесли удалённость от обнажённого дна Аральского моря, аэрозольная нагрузка, засоление почв и деградация растительного покрова. Пространственный анализ выявил закономерный градиент «север–юг»: максимальный экологический вклад наблюдается в Республике Каракалпакстан и постепенно снижается по мере удаления от Приаралья. Это подтверждает, что спутниковые показатели могут быть полезны не только для экологического мониторинга, но и для построения объяснимых моделей общественного здравоохранения.

Практическая значимость предложенного подхода заключается в возможности более точного выделения территорий, где беременные женщины могут нуждаться в усиленном дородовом наблюдении, профилактике анемии, контроле артериального давления и улучшенной маршрутизации в учреждения более высокого уровня. При этом результаты не следует трактовать как доказательство прямой причинно-следственной связи между экологическими факторами и материнскими исходами, поскольку использованные показатели агрегированы на региональном уровне, а медицинский слой основан на DHS-совместимых данных и синтетическом наборе MaternaUZ.

Дальнейшие исследования должны быть направлены на проверку модели на индивидуальных клинических данных, включая данные DMED, с учётом точной географической привязки пациенток и временной динамики экологических воздействий. При успешной внешней валидации предложенный подход может стать основой для создания экологически чувствительной системы поддержки принятия решений в сфере охраны материнства, ориентированной на раннее выявление зон повышенной уязвимости и адресное распределение профилактических ресурсов в регионах Узбекистана.

Литература

- [1] O'Hara SL, Wiggs GFS, Mamedov B, et al. Exposure to airborne dust contaminated with pesticide in the Aral Sea region. *Lancet*. 2000;355(9204):627–628.
- [2] Micklin P. The Aral Sea disaster. *Annu Rev Earth Planet Sci*. 2007;35:47–72.
- [3] Crighton EJ, et al. Impacts of an environmental disaster on health in Karakalpakstan. *Soc Sci Med*. 2003;56(3):551–567.

- [4] Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. In: NeurIPS. 2017;30.
- [5] Didan K. MODIS Vegetation Index Products (NDVI). NASA EOSDIS LP DAAC; 2015.
- [6] Ahmed M, et al. Maternal Health Risk Data Set (Bangladesh). UCI Machine Learning Repository; 2020.