

Научная секция "Искусственный интеллект и машинное обучение"

# Assessment of the reliability of spectral interpretation in hyperspectral 3D-CNN models under dimensionality reduction

Авторы:  
Смирнов И. К.  
Васильев Е. П.

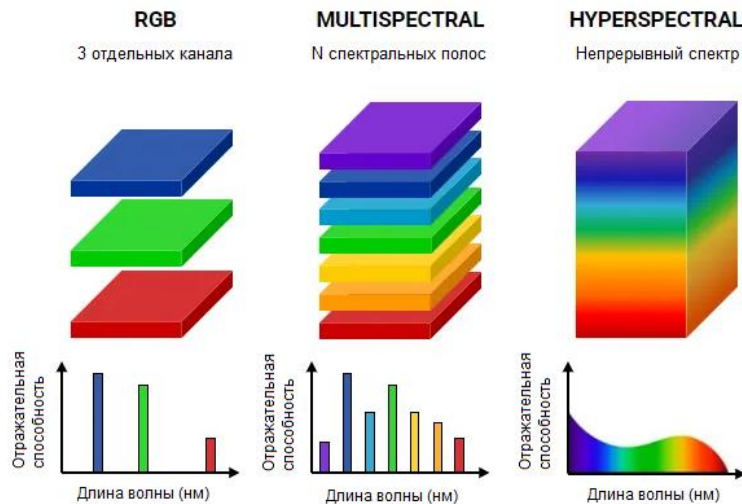
Москва, 2026 г.

# Гиперспектральные данные: преимущество и вычислительный барьер

В отличие от RGB, ГС камера сканирует сотни узких спектральных каналов, фиксируя скрытые физико-химические свойства объектов.

**Вычислительный барьер** - трехмерная структура куба данных ( $H \times W \times C$ ) и высокая межканальная корреляция создают колоссальную нагрузку на VRAM/RAM.

Для работы с ГСИ необходимы эффективные параллельные алгоритмы сжатия и оптимизированные нейросетевые архитектуры.



# Постановка задачи и вклад работы

**Цель исследования:** разработка, оптимизация и верификация легкой 3D-CNN архитектуры для классификации ГСИ с механизмом интерпретации предсказаний.

**Научно-практический вклад работы:**

- Сравнительное исследование линейного PCA и Kernel PCA в условиях аппаратных ограничений по памяти.
- Создание 3D-сверточной модели, адаптированной под параллельные GPU-вычисления
- Сравнение результатов с теоретическими границами Байесовского классификатора.
- Обоснование применения векторов Шепли в исходном спектральном пространстве.

# Сравнение датасетов

| Название           | Структура              | Диапазон волн (нм) | Экземпляры                   | Специфика                            | Калибровка сигнала                                |
|--------------------|------------------------|--------------------|------------------------------|--------------------------------------|---------------------------------------------------|
| WeedCube           | 3 культуры + 4 сорняка | 400-1000           | 160 калиброванных гиперкубов | <b>Контролируемые условия съёмки</b> | <b>Полная с помощью белого и черного эталонов</b> |
| Agricultural plant | 16 классов растений    | 420-979            | 385 гиперкубов из 59 снимков | Дисбаланс классов                    | Частичная (белый эталон)                          |
| Weed Species ID    | 3 вида сорняков        | 400-1000           | 5 сцен за 6 дней             | Только ранние стадии роста           | Отсутствует                                       |
| HLJ Dataset        | С/х культуры           | 350-1050           | Масштабные снимки (2 сцены)  | Низкая детализация растений          | Атмосферная коррекция                             |

# Обоснование выбора исследуемых архитектур

| Модель / Подход                         | Требуемый тип входа                               | Теоретическое преимущество                                                                      | Ограничение                                                                                      |
|-----------------------------------------|---------------------------------------------------|-------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|
| 1D-CNN                                  | Спектральный вектор (1 пиксель)                   | Минимальные вычислительные затраты, быстрая скорость обучения.                                  | Полностью игнорирует пространственную форму листьев и текстуру.                                  |
| Гибридная 2D-3D CNN                     | Пространственный патч ( $19 \times 19 \times C$ ) | Раздельное извлечение признаков.                                                                | Жесткое разделение этапов может терять сложные совместные пространственно-спектральные паттерны. |
| PyFormer (Transformer)                  | Пространственный патч ( $7 \times 7 \times C$ )   | Способность моделировать глобальные зависимости и перекрытия листьев через Self-Attention.      | Экстремальная вычислительная сложность на эпоху при высокой размерности данных.                  |
| <b>3D-CNN</b><br>(Разработанная модель) | Пространственный патч ( $19 \times 19 \times C$ ) | Совместное непрерывное извлечение пространственно-спектральных признаков; высокая стабильность. | Повышенная вычислительная сложность по сравнению с гибридными моделями.                          |

# Предобработка данных

## Для одномерной модели (1D-CNN [1]):

- Извлечение изолированных спектральных сигнатур.
- Применение SNV фильтра для устранения мультипликативных помех.
- Фильтрация Савицкого-Голея для подавления высокочастотного шума.

## Для пространственно-спектральных моделей (2D–3D-CNN [2], PyFormer [3]):

- Нарезка гиперкубов на пространственные патчи ( $19 \times 19 \times C$  или  $7 \times 7 \times C$ ).
- Поканальная стандартизация - приведение каждого спектрального канала к  $\mu = 0, \sigma = 1$ .

# Условия эксперимента и инфраструктура

**НПС-среда:** Google Colab, GPU NVIDIA Tesla T4 (16 GB VRAM).

**Программный стек:** Python, Keras / TensorFlow (с cuDNN ускорением), Scikit-learn.

**Параметры:** Оптимизатор Adam:  $lr = 10^{-4} - 10^{-5}$  (подбор под архитектуру), функция потерь Sparse Categorical Cross-Entropy

**Скрининг:** 10% репрезентативной стратифицированной выборки для быстрой оценки ресурсов.

# Архитектуры базовых моделей

1D-CNN полностью игнорирует пространственную геометрию и форму объектов, обучаясь только на векторе отражения.

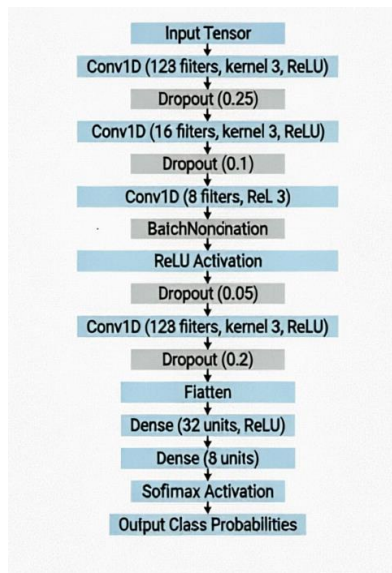


Рис.1. Архитектура 1D-CNN

PyFormer оценивает глобальных взаимосвязей между каналами через механизм внимания. Обладает большой вычислительной сложностью.

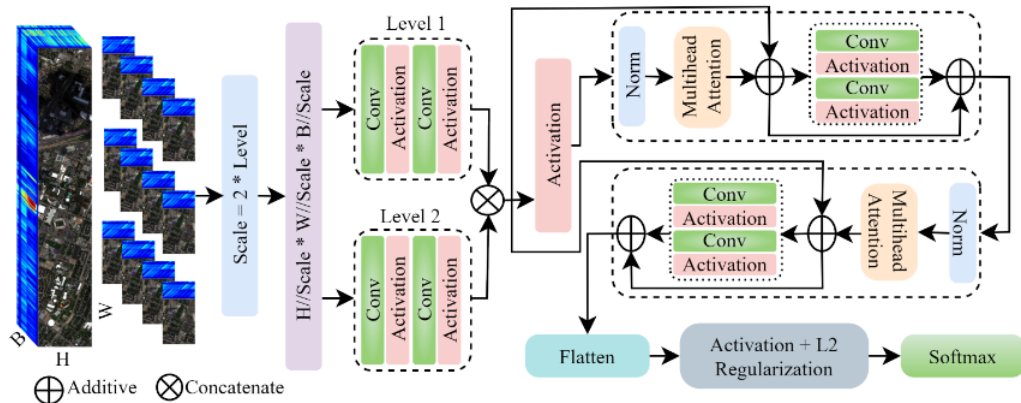


Рис.2. Архитектура PyFormer

# Архитектуры базовых моделей

2D-3D-CNN одновременно анализирует текстуры листа и формы спектральной кривой. Имеет наилучший баланс точности и скорости

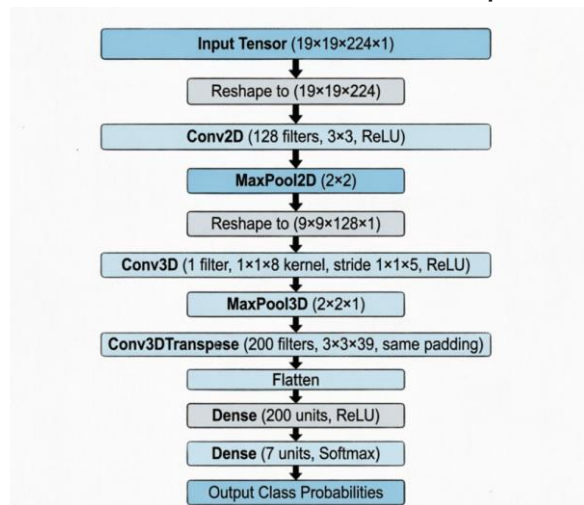
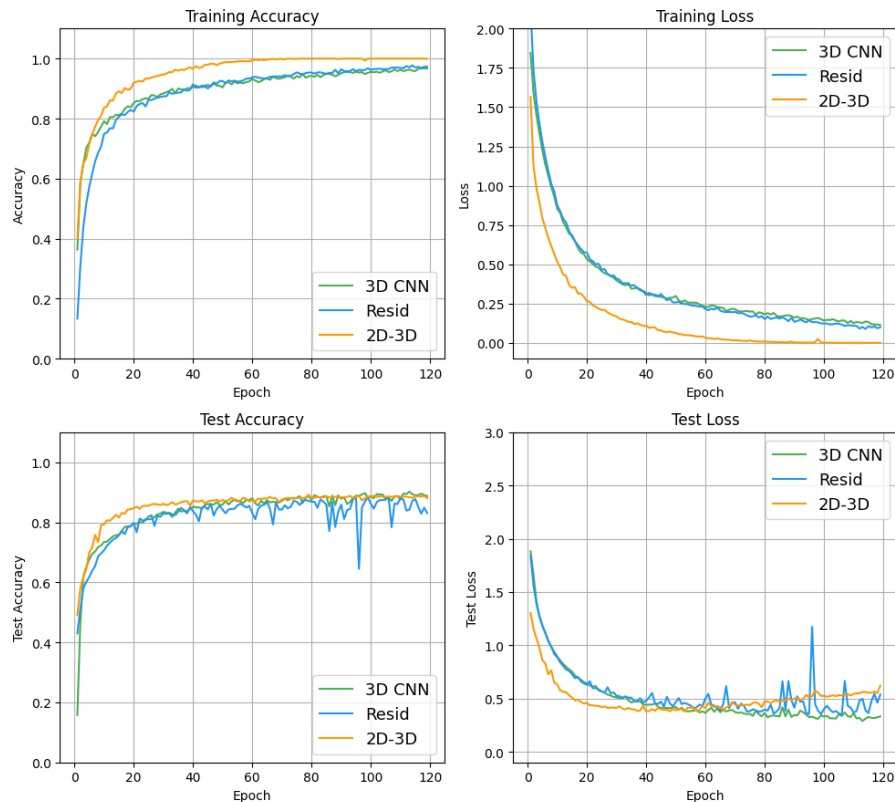


Рис.2. Архитектура 2D-3D-CNN

| Модель    | Ассурасу (тест) | Относительная стоимость по времени | Количество эпох обучения |
|-----------|-----------------|------------------------------------|--------------------------|
| 1D-CNN    | 86%             | 1.0×<br>(baseline)                 | 500                      |
| 2D-3D-CNN | 92%             | 1.8×                               | 150                      |
| PyFormer  | 90%             | 2.3×                               | 50                       |

Таблица 1. Результаты скрининга на 10% выборки

# Графики сравнения качества доработанных моделей.



2D-3D-модель склонна к переобучению, а 3D-модель демонстрирует более устойчивое поведение.

Модель с остаточными связями показывает метрики, сопоставимые с другими вариантами, однако характеризуется выраженными флуктуациями на тесте

3D-CNN обеспечила оптимальный баланс: гладкую сходимость без осцилляций и максимальную обобщающую способность на тестовой выборке

# Результаты анализа первого запуска 3D CNN

Большинство ошибок модель допускает на границах изображений и в области фона.

Решение – использовать перекрывающиеся патчи

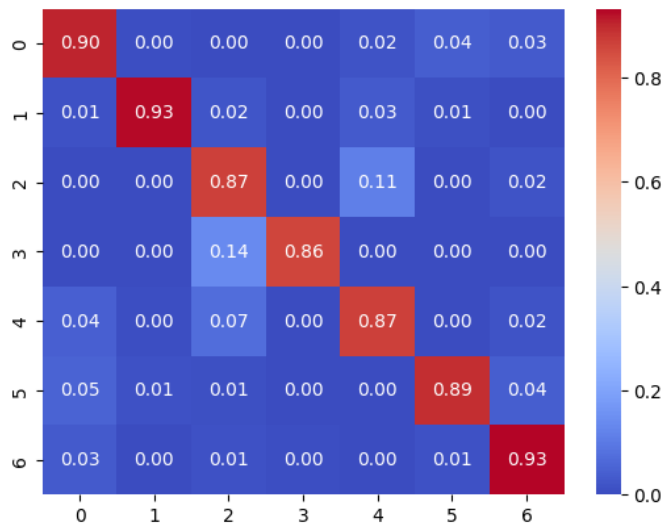


Рис. 5. Матрица ошибок первой версии 3D-CNN

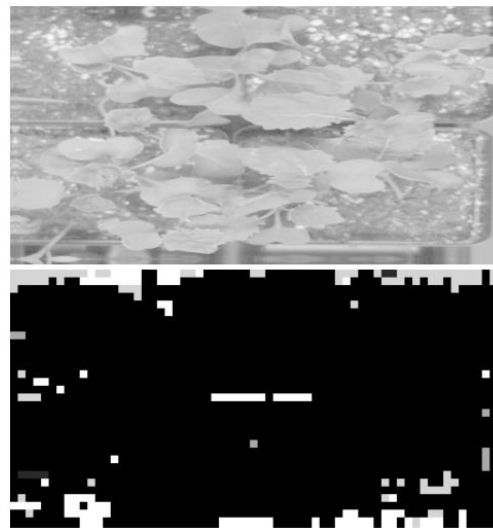


Рис. 6. Примеры неверно классифицированных патчей

# Сжатие данных

Из-за неоднородности дисперсии необходимо привести данные к нулевому математическому ожиданию и единичной дисперсии

Критерий качества:

$$\varepsilon = \frac{||X - \tilde{X}||_F}{||X||_F}$$

$X$  – исходные данные  
 $\tilde{X}$  – восстановленные

## Результаты оптимизации:

- Число главных компонент:  $k = 70$  (из 224 исходных).
- Доля объясненной дисперсии: **99.98%**.
- Относительная ошибка реконструкции:  $\varepsilon \approx 6.44 * 10^{-5}$

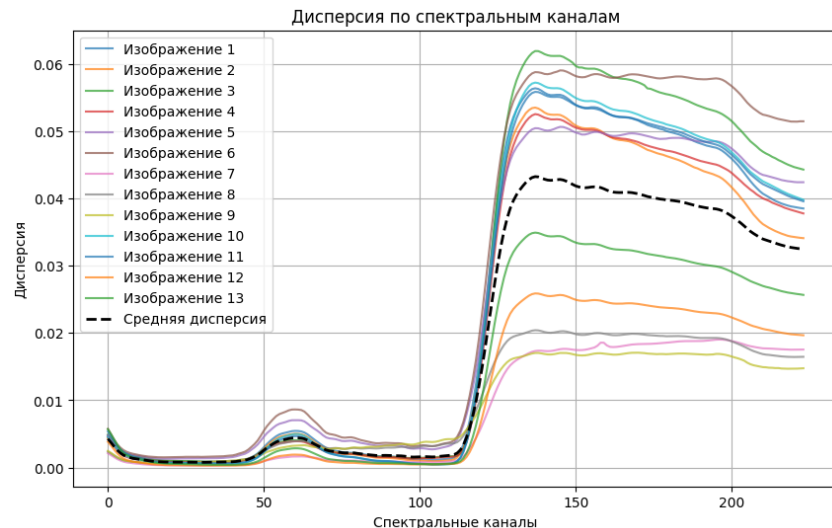


Рис.4. Зависимость дисперсии от номера канала

## Экспериментальный анализ нелинейного снижения размерности (Kernel PCA)

Прямое вычисление ядерной матрицы  $K \in R^{N \times N}$  для датасета WeedCube невозможно из-за ограничений RAM. Вычислительная сложность классического Kernel PCA –  $O(N^3)$  по операциям и  $O(N^2)$  по памяти.

Для аппроксимации ядра применяется теорема Бохнера для перехода к случайным тригонометрическим признакам (Random Fourier Features – RFF\*)

$$k(x, y) \approx z(x)^T z(y), \quad z(x) = \frac{1}{\sqrt{D}} (\cos(w_1 x), \sin(w_1 x), \dots, \cos(w_D x), \sin(w_D x))^T$$

где частоты  $w_k$  сэмплируются из преобразования Фурье ядерной функции. Это позволило свести задачу к линейному PCA в пространстве RFF. Аппроксимация RFF снимает ограничение по RAM, но не дает прироста точности

**Результаты экспериментов:** прирост целевых метрик классификации отсутствует. Аппроксимация нелинейных эффектов реализуется последующими слоями глубокой нейросети. Дополнительное нелинейное преобразование на этапе предобработки избыточно.

## Итоговая архитектура 3D CNN и метрики классификации

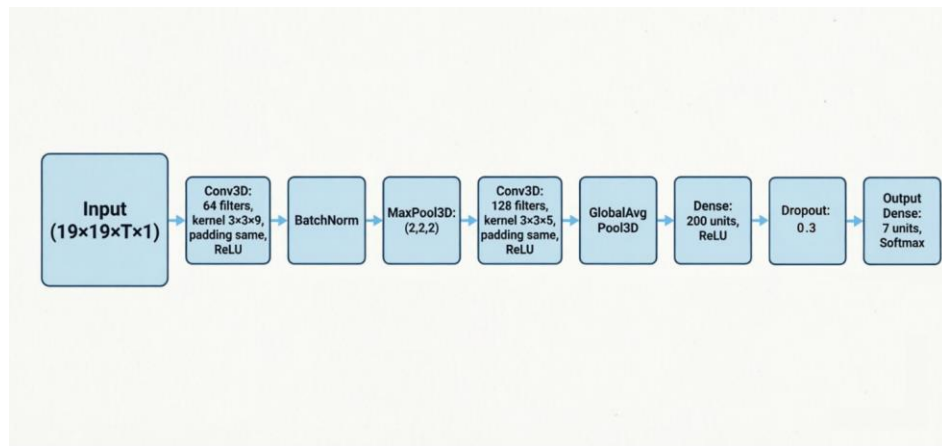


Рис.7. Итоговая архитектура 3D CNN

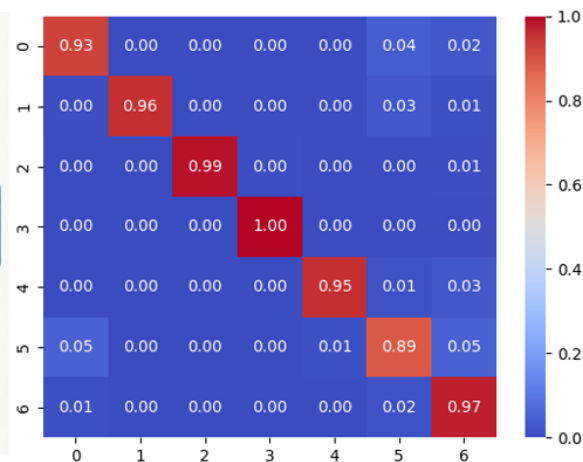


Рис. 8. Матрица ошибок 3D-CNN

|         | Precision | Recall | F1-score | Accuracy |
|---------|-----------|--------|----------|----------|
| Average | 0.95      | 0.96   | 0.95     | 0.94     |

## Сопоставление результатов с границами ошибки Байеса

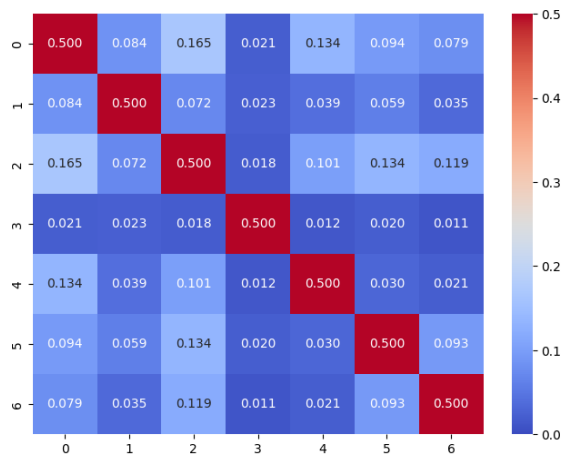


Рис. 9. Попарные нижние оценки ошибки Байеса

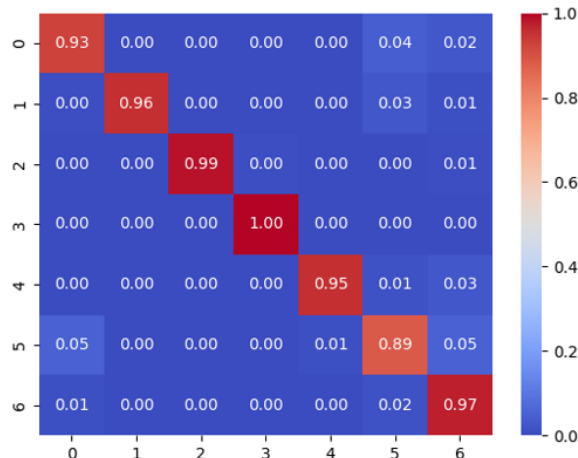


Рис. 10. Матрица ошибок 3D-CNN

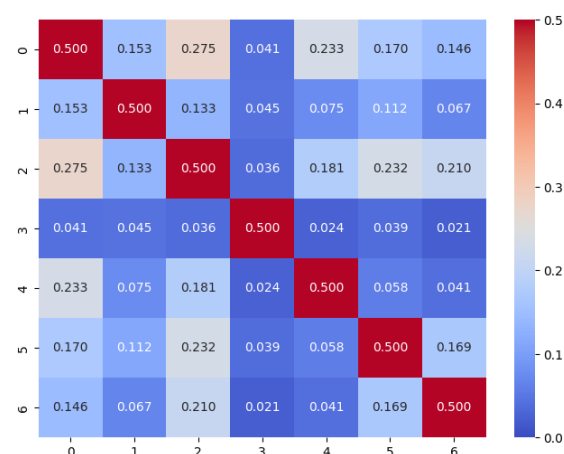


Рис. 11. Попарные верхние оценки ошибки Байеса

Границы ошибки Байеса рассчитывались на основе статистики Фридмана-Рафского [4]. Распределение ошибок разработанной 3D-CNN модели локализовано внутри теоретических границ Байесовского классификатора, что подтверждает оптимальность архитектуры.

## Метод SHAP [5,6]

Для аппроксимации значений SHAP и объяснения предсказания  $f(x)$  строится модель  $g(z')$ :

$$g(z') = \phi_0 + \sum_{i=1}^M \phi_i z'_i$$

$$\gamma(z') = \frac{M-1}{C_M^{|z'|} |z'| (M - |z'|)}$$

$z' \in \{0,1\}^M$  - бинарный вектор коалиции признаков  
 $h(z'): \{0,1\}^M \rightarrow \mathbb{R}^{19 \times 19 \times M}$ ,  $M \in \{70, 224\}$  - отображение коалиции в исходное пространство  
 $\gamma(z')$  - весовой коэффициент регрессии

Задача регрессии значений SHAP принимает вид:

$$\min_{\phi} \left( \sum_{z' \in Z} \gamma(z') \left( f(h(z')) - g(z') \right)^2 \right)$$

*Реализация:* адаптирован алгоритм решения регрессии из библиотеки shap для случая гиперспектральных данных с использованием аксиомы эффективности значений SHAP

# Подходы к вычислению SHAP

Задача регрессии SHAP может быть поставлена:

- В исходном пространстве объектов  $x$
- В пространстве PCA-проекций  $y = xW^T$ , где  $W$  – ортогональная матрица главных компонент PCA

**Стратегии замещения признаков:**

**Маргинальное:** Предполагает независимость каналов. Замещение средними значениями из фоновой выборки.

**Условное:** Учитывает корреляции каналов через распределение  $P(x_{\text{пропущенные}} | x_{\text{присутствующие}})$ . Вычислительно крайне сложно.

# Направления эксперимента

**Путь А:** Расчет маргинального SHAP непосредственно в пространстве исходных признаков  $x$ .

**Путь Б:** Расчет маргинального SHAP на независимых компонентах  $y$  с последующей обратной проекцией. Для корректности этого способа должны быть поставлены и проверены 2 гипотезы

**Гипотеза 1:** Признаки исходных гиперспектральных сигнатур  $x$  удовлетворяют многомерному нормальному распределению.

**Гипотеза 2:** Полученные SHAP-значения из ортогонального PCA-пространства  $y$  можно корректно спроецировать обратно в пространство исходных длин волн  $x$

**Анализ Гипотезы 1** проведен с помощью многомерного статистического теста Хенце-Зирклера [7]

**Особенность сырых данных:** Критерий возвращает значение  $p\text{-value} = 0.5$ . Данный результат является следствием плохой обусловленности матрицы ковариации. Поэтому выполнено прореживание признаков: взят каждый пятый спектральный канал. Повторное тестирование показало, что гипотеза о многомерной нормальности отвергается  $p\text{-value} \approx 0$

## Формулирование Гипотезы 2

Используем эвристику, основанную на том, что матрица проекции  $W$  из пространства  $X$  в пространство  $Y$  состоит из ортонормированных векторов. Тогда при проекции  $\phi^x = \phi^{*y} \tilde{W}$ ,  $\tilde{W} = (w_{ij}^2)$  будет следовать равенство:

$$\sum_{i=1}^{224} \phi_i^x = \sum_{j=1}^{70} \phi_j^{*y}$$

$\phi^{*y}$  - решение задачи для объектов  $y$ .  
 $\phi^x$  – решение задачи для объектов  $x$ ,  
аппроксимируемое с помощью  $\phi^{*y}$

# Решение задачи в пространстве $Y$

Моделирование сценария с обнулением признаков после 51-го канала привело к падению точности классификации 3D-CNN до 26%. Следовательно, обратное проецирование, основанное на гипотезе 2 некорректно. Данное искажение вызвано спецификой матрицы  $W$ , чьи наибольшие абсолютные значения сконцентрированы в первых столбцах

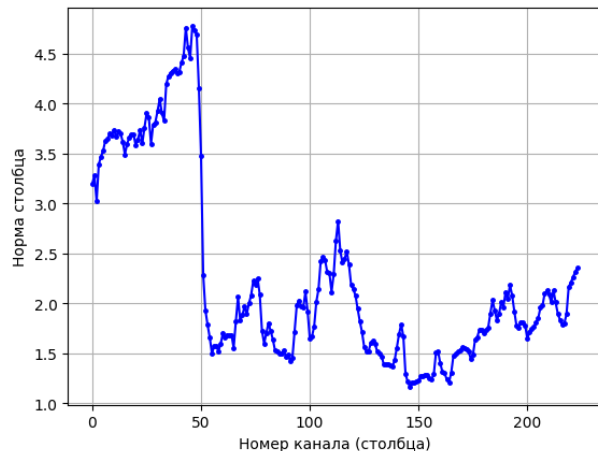


Рис.14. Сумма абсолютных значений векторов матрицы  $W$

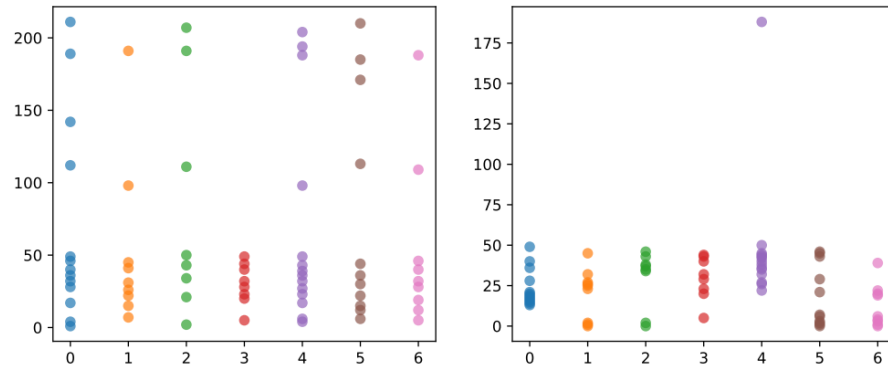


Рис.15. Распределение значимых спектральных пиков в пространстве  $Y$ . Левый график демонстрирует распределение  $N$  пиков для сырых векторов SHAP; правый график - распределение пиков после процедуры сглаживания

# Решение задачи в пространстве $X$

К сырым векторам SHAP применено скользящее сглаживание, а локальные экстремумы (пики) определяются с помощью адаптивного порога:  $N = 0.1 \cdot \max(\phi_{\text{сглаж}})$

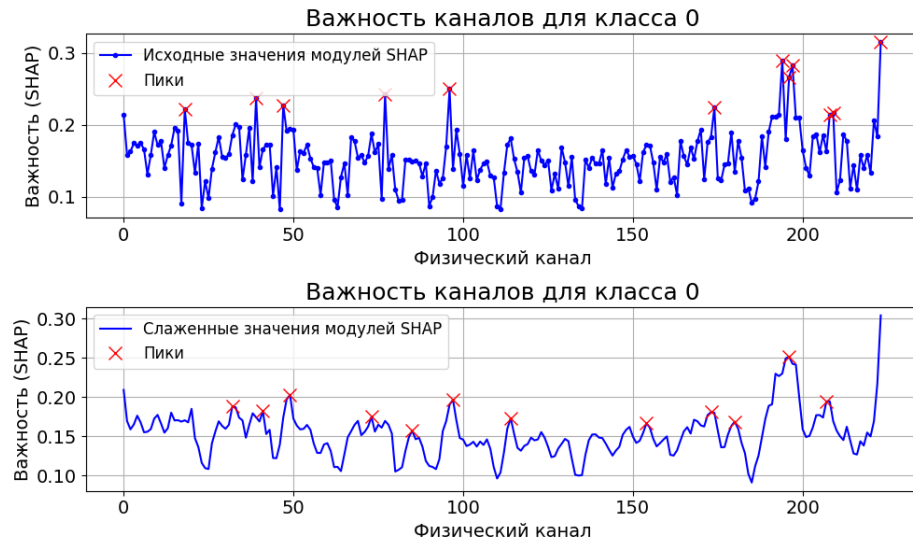


Рис.12. Пример SHAP-вектора для одного изображения с выделенными пиками

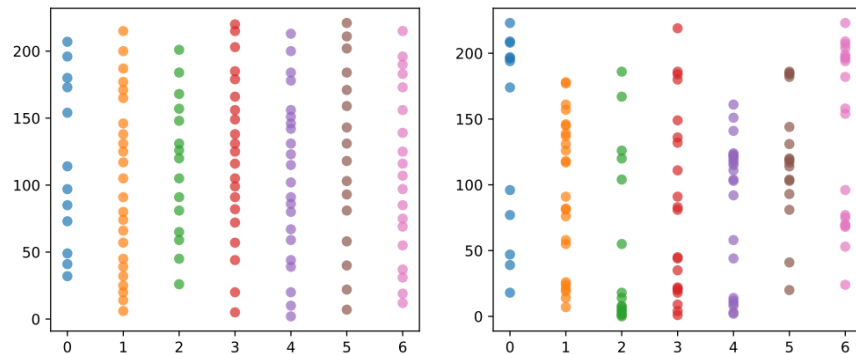


Рис.13. Распределение значимых спектральных пиков в пространстве  $X$ . Левый график демонстрирует распределение  $N$  пиков для сырых векторов SHAP; правый график - распределение пиков после процедуры сглаживания

# Выводы

Эффективность 3D-CNN архитектуры:

- Разработанная 3D-CNN достигла **Accuracy = 94%** и **F1-Score = 95%**, находясь в теоретической оценке ошибки Байеса.
- Применение перекрывающихся патчей убрало краевые ошибки фоновых объектов.

Оптимизация размерности данных:

- Линейный PCA сократил спектральную размерность с 224 до 70 каналов при сохранении 99.98% объясненной дисперсии.
- Kernel PCA (RFF) показал ограничение по памяти  $O(N^2)$  и избыточность нелинейного сжатия для глубоких 3D-сетей.

Обоснование интерпретируемости:

- Обратное проецирования весов SHAP из PCA-пространства некорректно - экспериментальный тест обнуления каналов показал падение точности до 26%.
- Единственным математически и физически корректным методом признан прямой расчет маргинального SHAP в исходном пространстве длин волн ( $X$ ).

# Литература

1. Cordeiro, J., Raimundo, A., Postolache, O., Sebastião, P.: Neural Architecture Search for 1D CNNs—Different Approaches Tests and Measurements.
2. Yu, C., Han, R., Song, M., Liu, C., Chang, C.: A Simplified 2D-3D CNN Architecture for Hyperspectral Image Classification Based on Spatial-Spectral Fusion.
3. M. Ahmad, M. Hassaan F. Butt, M. Mazzara, S. Distifano Pyramid Hierarchical Transformer for Hyperspectral Image Classification / Muhammad Ahmad, Muhammad Hassaan Farooq Butt, Manuel Mazzara, Salvatore Distifano
4. Visar Berisha, Alan Wisler, Alfred O. Hero, and Andreas Spanias Empirically Estimable Classification Bounds Based on a New Divergence Measure
5. Molnar, C.: Interpretable Machine Learning. A Guide for Making Black Box Models Explainable.
6. Lundberg, S.M., Lee, S.I. A unified approach to interpreting model predictions // Advances in Neural Information Processing Systems
7. N. Henze & B. Zirkler (1990) A class of invariant consistent tests for multivariate normality, Communications in Statistics - Theory and Methods

**Спасибо за внимание!**

**Формула SNV:**

$$SNV(x_i) = \frac{x_i - \mu}{\sigma}$$

где:

$x_i$  - значение сигнатуры в точке  $i$ ,

$\mu$  – среднее значение одной спектральной сигнатуры по всем длинам волн,

$\sigma$  – стандартное отклонение одной спектральной сигнатуры по всем длинам волн.

Принцип работы SG-фильтра: сглаживание спектральной упорядоченной последовательности выполняется во временном/спектральном окне фиксированной ширины  $2m + 1$  путём локальной полиномиальной аппроксимации методом наименьших квадратов :

$$sg(x_k) = \sum_{i=-m}^m a_i x_{i+k}$$

$x_k$  – исходное значение отражения в канале  $k$

$sg(x_i)$  - сглаженное значение

$a_i$  - нормированные весовые коэффициенты, аналитически вычисляемые для полинома заданной степени  $d$

## Оценка ошибки Байеса

Для оценки ошибки применяется метод, основанный на статистике Фридмана-Рафского, описанный в [4]. Суть метода в построении минимального остового дерева на данных и подсчете количества ребер, соединяющих объекты разных классов -  $C(X_0, X_1)$ .

Это количество ребер есть статистика приближающая, введенную в [4] дивергенцию  $\tilde{D}_p(f_0, f_1)$ , позволяющую оценить ошибку как:

$$\frac{1}{2} - \frac{1}{2} \sqrt{\tilde{D}_p(f_0, f_1)} \leq \varepsilon^{bayes} \leq \frac{1}{2} + \frac{1}{2} \tilde{D}_p(f_0, f_1)$$

$$1 - \frac{C(X_0, X_1)}{N_0 + N_1} \xrightarrow{\text{п.н.}} \tilde{D}_p(f_0, f_1), \text{ при } N_0 \rightarrow \infty, N_1 \rightarrow \infty, \quad \frac{N_0}{N_0 + N_1} \rightarrow p, \frac{N_1}{N_0 + N_1} \rightarrow q$$

$X_k$  – выборка объектов класса  $k$

$N_k$  – количество элементов в выборке

$p, q$  – априорные вероятности классов 0,1

$f_k(x) = p(x|k)$  – условная плотность распределения класса  $k$

$k = 0; 1$  – классы

# Тест Хенце-Зирклера

Тест формулируется для проверки сложной нулевой гипотезы: распределение случайного вектора  $X$  является невырожденным нормальным распределением.

Предположения:

- Исходные данные  $X_1, \dots, X_n$  должны быть выборкой независимых и одинаково распределенных случайных векторов в  $\mathbb{R}^d$ .
- Истинное распределение не должно быть сконцентрировано в какой-либо гиперплоскости размерности  $d - 1$ .

Статистика:  $T_{n,\beta} = n(4I\{S_n \text{ is singular}\} + D_{n,\beta}I\{S_n \text{ is nonsingular}\})$ ,  $I$  - индикаторная функция.

$$D_{n,\beta} = \int_{\mathbb{R}^d} \left| \psi_n(t) - \exp\left(-\frac{1}{2} \|t\|^2\right) \right|^2 \phi_\beta(t) dt \quad \phi_\beta(t) = (2\pi\beta^2)^{-\frac{d}{2}} \exp\left(-\frac{1}{2\beta^2} \|t\|^2\right)$$

$$\psi_n(t) = \frac{1}{n} \sum_{j=1}^n \exp(itY_j) \quad Y_j = S_n^{-\frac{1}{2}}(X_j - \bar{X}) \quad S_n = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^T (X_j - \bar{X})$$

Если распределение не нормальное, то  $T_{n,\beta} \xrightarrow[n \rightarrow \infty]{\text{п.н.}} \infty$ , а если нормальное то  $T_{n,\beta} \xrightarrow[n \rightarrow \infty]{d} T_\beta = \sum_{k=1}^{\infty} \delta_k(\beta) \chi_{1,k}^2$ ,  $\delta_k(\beta)$  - собственные числа введенного в [7] оператора

## Формализация гипотезы о сохранении локальной важности признаков при переходе между пространствами $X$ и $Y$ через веса матрицы $W$

Составим вектор  $\hat{z}$  - спроецированную маску:  $\hat{z}_k = \sum_{i=1}^{224} z_i w_{ki}^2$  для объекта в пространстве  $Y$ . Тогда вторая гипотеза примет вид:  $g_{\phi^x}(z) = g_{\phi^y}(\hat{z})$ .

Откуда:

$$\begin{aligned} \sum_{i=1}^{224} \phi_i^x z_i &= \sum_{j=1}^{70} \phi_j^{*y} \hat{z}_j \Rightarrow \sum_{i=1}^{224} \phi_i^x z_i = \sum_{j=1}^{70} \phi_j^{*y} \sum_{i=1}^{224} z_i w_{ji}^2 \\ \sum_{i=1}^{224} \phi_i^x z_i &= \sum_{i=1}^{224} \left( \sum_{j=1}^{70} \phi_j^{*y} w_{ji}^2 \right) z_i \quad \forall z \in \{0,1\}^{224} \Rightarrow \phi_i^x = \sum_{j=1}^{70} \phi_j^{*y} w_{ji}^2 \quad i = \overline{1,224} \\ \sum_{i=1}^{224} \phi_i^x &= \sum_{j=1}^{70} \phi_j^{*y} \sum_{i=1}^{224} w_{ji}^2 \Rightarrow \sum_{i=1}^{224} \phi_i^x = \sum_{j=1}^{70} \phi_j^{*y} \end{aligned}$$

# Снижение размерности задачи регрессии SHAP для соблюдения аксиомы эффективности

Аксиома эффективности:  $f(x) - E[f(X)] = \sum_{i=1}^M \phi_i$

$$\phi_M = f(x) - E[f(X)] - \sum_{i=1}^{M-1} \phi_i$$

$$g(z') = E[f(X)] + \sum_{i=1}^{M-1} \phi_i z'_i + \left( f(x) - E[f(X)] - \sum_{i=1}^{M-1} \phi_i \right) z'_M$$

$$\min \left( \sum_{z' \in Z} \gamma(z') \left( f(h(z')) - E[f(X)] - (f(x) - E[f(X)])z'_M - \sum_{i=1}^{M-1} \phi_i (z'_i - z'_M) \right)^2 \right)$$

Тогда  $(f(h(z')) - E[f(X)] - (f(x) - E[f(X)])z'_M)$  - новая целевая переменная, а  $(z'_i - z'_M)$  - новые признаки.

## Доказательство независимости признаков после PCA

Пусть  $x_1, \dots, x_n \in \mathbb{R}^B$  получены из нормального  $B$  мерного распределения, тогда

$$f(x_i) = \text{const} * \exp\left(-\frac{1}{2}(x_i - 0)C^{-1}(x_i - 0)^T\right) = \text{const} * \exp\left(-\frac{n}{2}x_i(X^T X)^{-1}x_i^T\right)$$

$$f(x_i) = \text{const} * \exp\left(-\frac{n}{2}x_i(V\Sigma^T\Sigma V^T)^{-1}x_i^T\right) = \text{const} * \exp\left(-\frac{n}{2}y_i\Sigma^{-1}\Sigma^{-T}y_i^T\right)$$

$$f(x_i) = \text{const} * \exp\left(-\frac{n}{2}\left(\frac{1}{\sigma_1^2}y_{1i}^2 + \frac{1}{\sigma_2^2}y_{2i}^2 + \dots + \frac{1}{\sigma_B^2}y_{Bi}^2\right)\right) = g(z_1) \dots g(z_B)$$

$f(x)$  – плотность нормального многомерного распределения,  $C$  – матрица ковариации признаков.

$g(z)$  – одномерное нормальное распределение

$$x_i = y_i V^T,$$

$$z_k = \frac{1}{\sigma_k^2} y_{ki}^2.$$

## Выбор количества главных компонент

Доля объясненной дисперсии для компоненты  $i$  по формуле

$$\frac{\lambda_i}{\sum_j \lambda_j}$$

где  $\lambda$  – собственные числа матрицы ковариации  $C = \frac{1}{n} X^T X$ , и поскольку  $X^T X = V \Sigma^T \Sigma V^T$ , то

$\lambda_i = \frac{\sigma_i^2}{n}$ , а значит

$$\frac{\lambda_i}{\sum_j \lambda_j} = \frac{\sigma_i^2}{\sum_j \sigma_i^2}$$

По теореме Экарта-Янга\* наилучшим приближением по норме Фробениуса будет матрица  $\tilde{X}$ , получаемая из сингулярного разложения матрицы  $X$  по формуле:  $\tilde{X} = U \tilde{\Sigma} V^T$ ,  $\tilde{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_k, 0, \dots, 0)$ . Откуда получаем ошибку реконструкции данных:

$$\varepsilon = \frac{||X - \tilde{X}||_F}{||X||_F}$$

\* Eckart C., Young G. The approximation of one matrix by another of lower rank [Текст] / Eckart C., Young G. // Psychometrika. — 1936. — № 3. — С. 211–218.