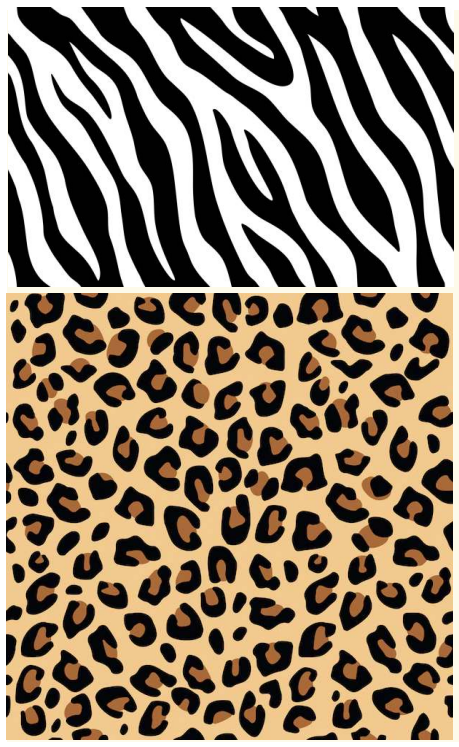


Optimizing spatiotemporal models

Ilya Ivanov, Veronika Novgorodskaya, Vera Terenteva, Laura Avinyó



В биологии развития мы получаем множество изображений в качестве данных. Мы также создаем множество моделей, но у нас все еще нет стандартизированного способа связать данные изображения и модели, чтобы выяснить, каковы конкретные значения модели для создания определенного набора данных.

В этой работе мы рассмотрели реакционно-диффузионную модель, которая описывает формирование паттернов и в одной из своих вариаций морфологию конечностей.

$$\begin{cases} \frac{\partial u}{\partial t} = D_u \nabla^2 u - uv^2 + F(1 - u) \\ \frac{\partial v}{\partial t} = D_v \nabla^2 v + uv^2 - (F + k)v \end{cases}$$



Нейронные сети

Мы использовали нейронные сети для кластеризации изображений, классификации изображений по паттернам и прогнозирования параметров с помощью регрессии.

Чтобы уменьшить размерность наших входных данных, мы использовали **автоэнкодер**. Это тип нейронной сети, состоящий из двух элементов — кодировщика и декодерова. Первый преобразует входные данные в вектор меньшего размера, а второй преобразует их обратно в ту же размерность, которую имеют входные данные. Цель модели — свести к минимуму разницу между входными и выходными данными. Таким образом, кодировщик уменьшает размер с минимальной потерей информации.

Используя **свёрточную нейронную сеть**, мы выполнили задачу классификации. Рис. 2.

Затем мы использовали свёрточные нейронные сети (1D и 2D) для прогнозирования значений параметров из исходных изображений или закодированных автоэнкодером соответственно. Рис. 3.

Более того мы использовали предсказанный класс в качестве дополнительных данных для регрессора.

Ансамблевый метод

Другой способ делать прогнозы - это использовать **ансамбль моделей**. Это метод, который определённым образом объединяет прогнозы нескольких моделей для достижения наилучших результатов. В нашей работе мы использовали методы **бэггинга**, находя **взвешенное** среднее значение прогнозов каждой модели.

Мы разделили данные на пакеты (случайным образом или по определённому значению данного столбца). Для достижения наилучшей точности мы также добавили некоторые веса, которые определяют важность выходных данных каждой модели, поэтому результат наилучшей модели имеет приоритет.

Отдельная обнаруженная проблема заключается в том, что некоторые паттерны формируются быстрее, чем другие, и поэтому в последние моменты времени некоторые паттерны дают меньше информации. Исходя из этого, мы испробовали различные подходы к разбиению данных и изменению весов: (1) рассмотрение **точных временных интервалов**, и (2) **количество вещества** как показатель развития паттерна. Последний подход показал наилучший результат.

Результаты

1. Мы можем обучать модели машинного обучения для прогнозирования параметров, и это может использоваться врачами.
2. Некоторые модели ведут себя лучше, чем другие, для данного конкретного набора данных.
3. Использование ансамбля различными способами разделения данных является многообещающим подходом, но всё ещё требует работы.
4. Методы должны быть более надёжными, чтобы их можно было использовать для других более сложных моделей, но они работают и идейно верны.

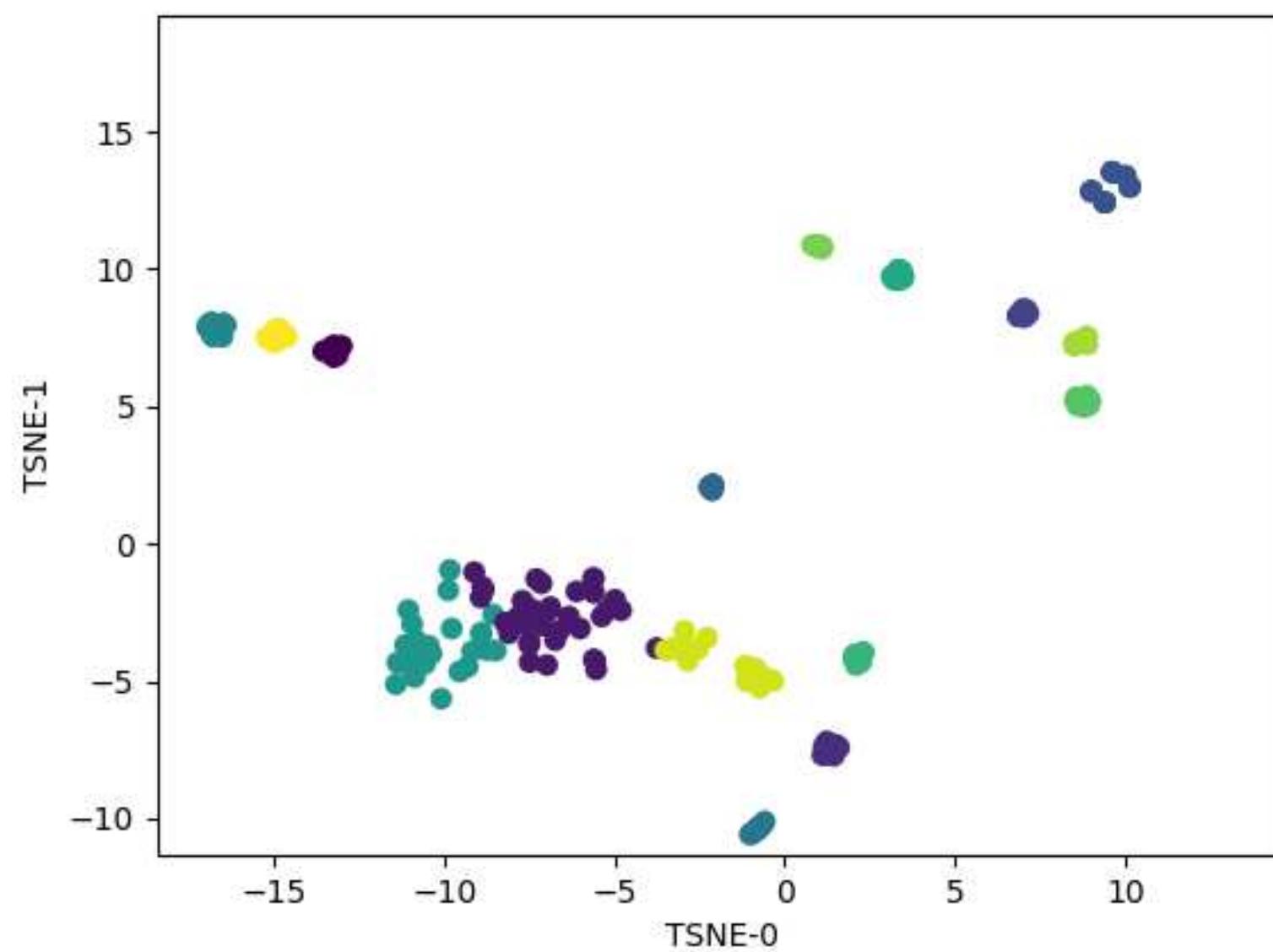


Рис. 1: Кластеризация, основанная на кодировании исходных изображений

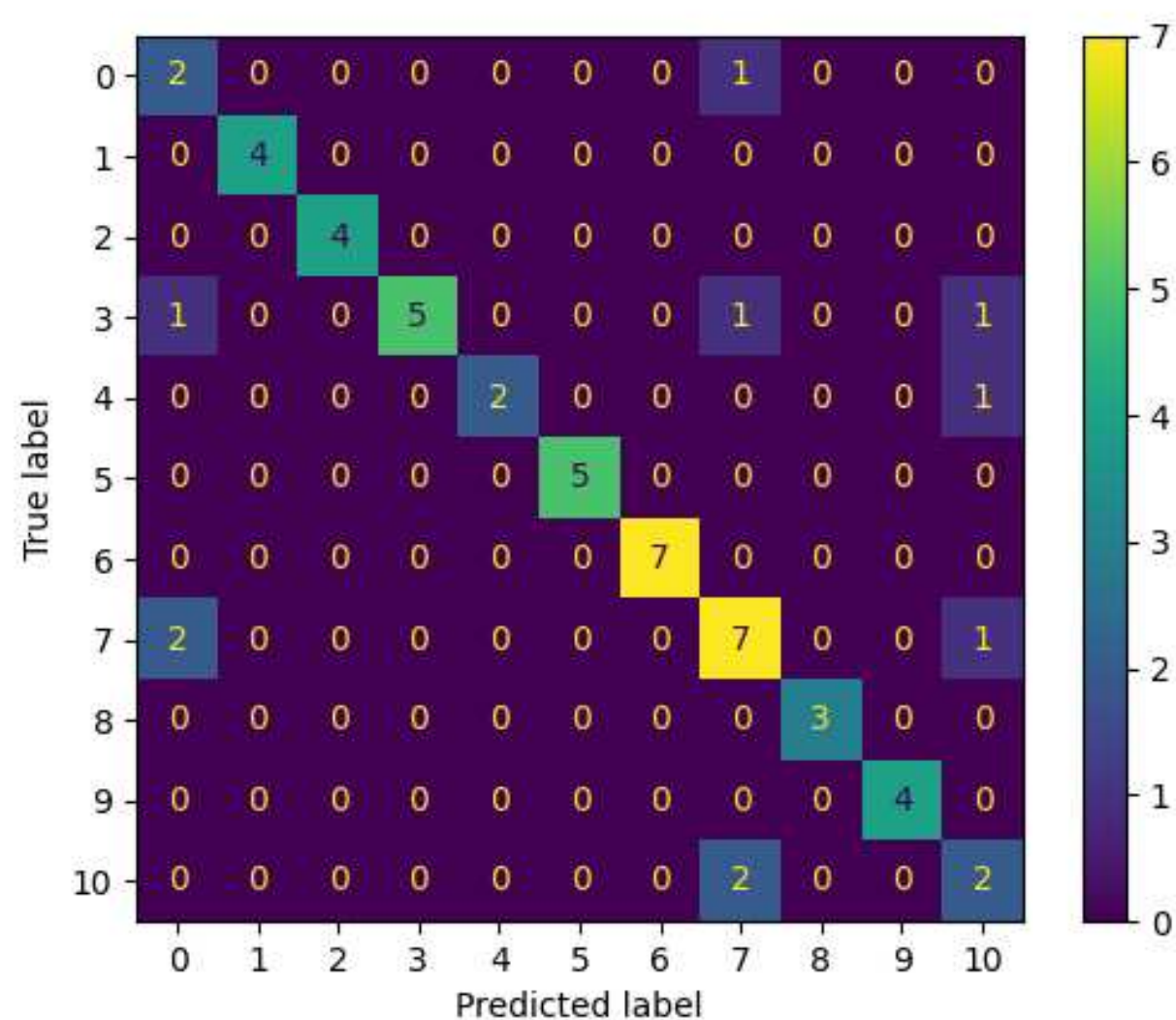


Рис. 2: Матрица оценки для задачи классификации с применением свёрточной нейронной сети

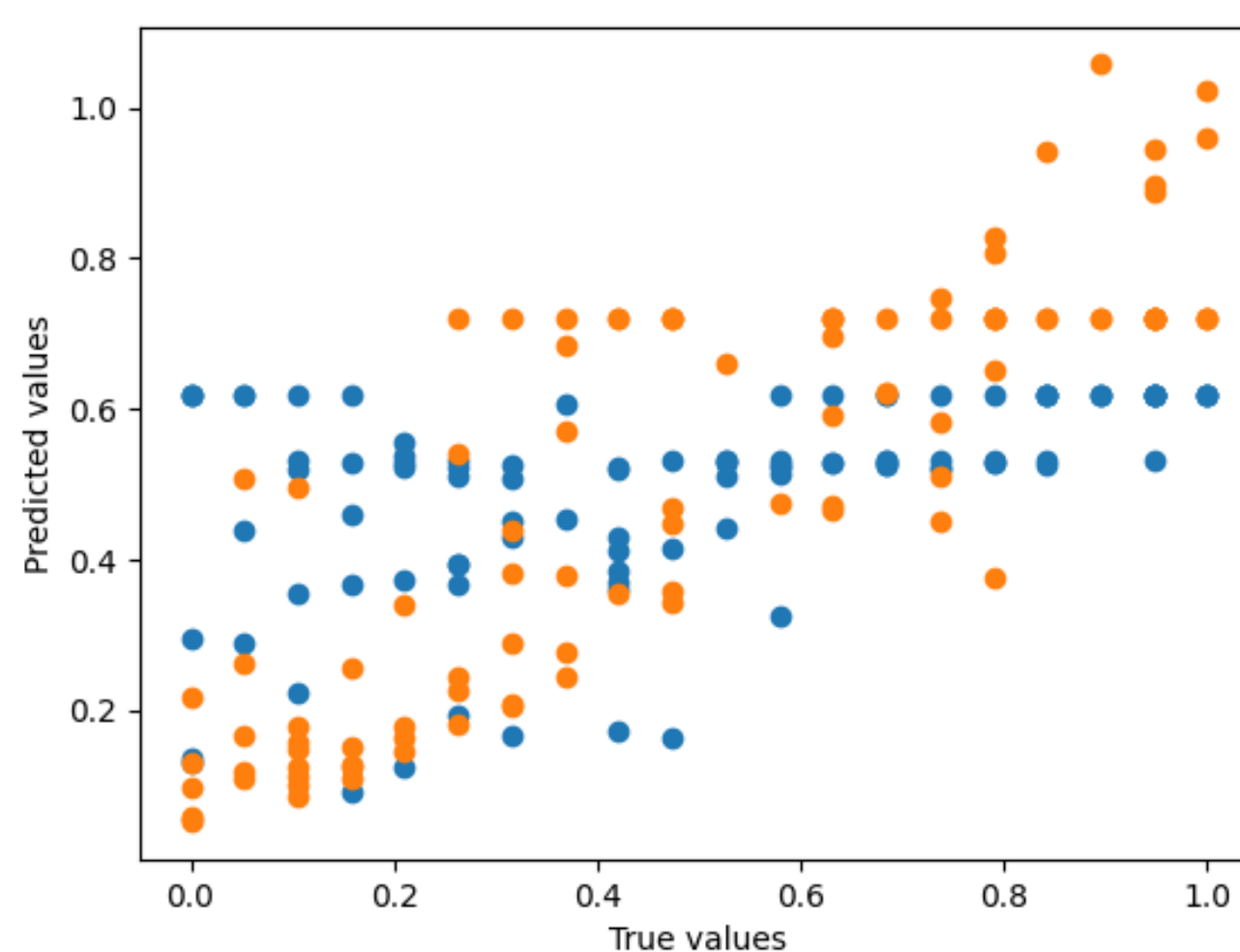


Fig. 3: Предсказанные значения feed rate и kill rate (регрессия с использованием 1D свёрточной нейронной сети)

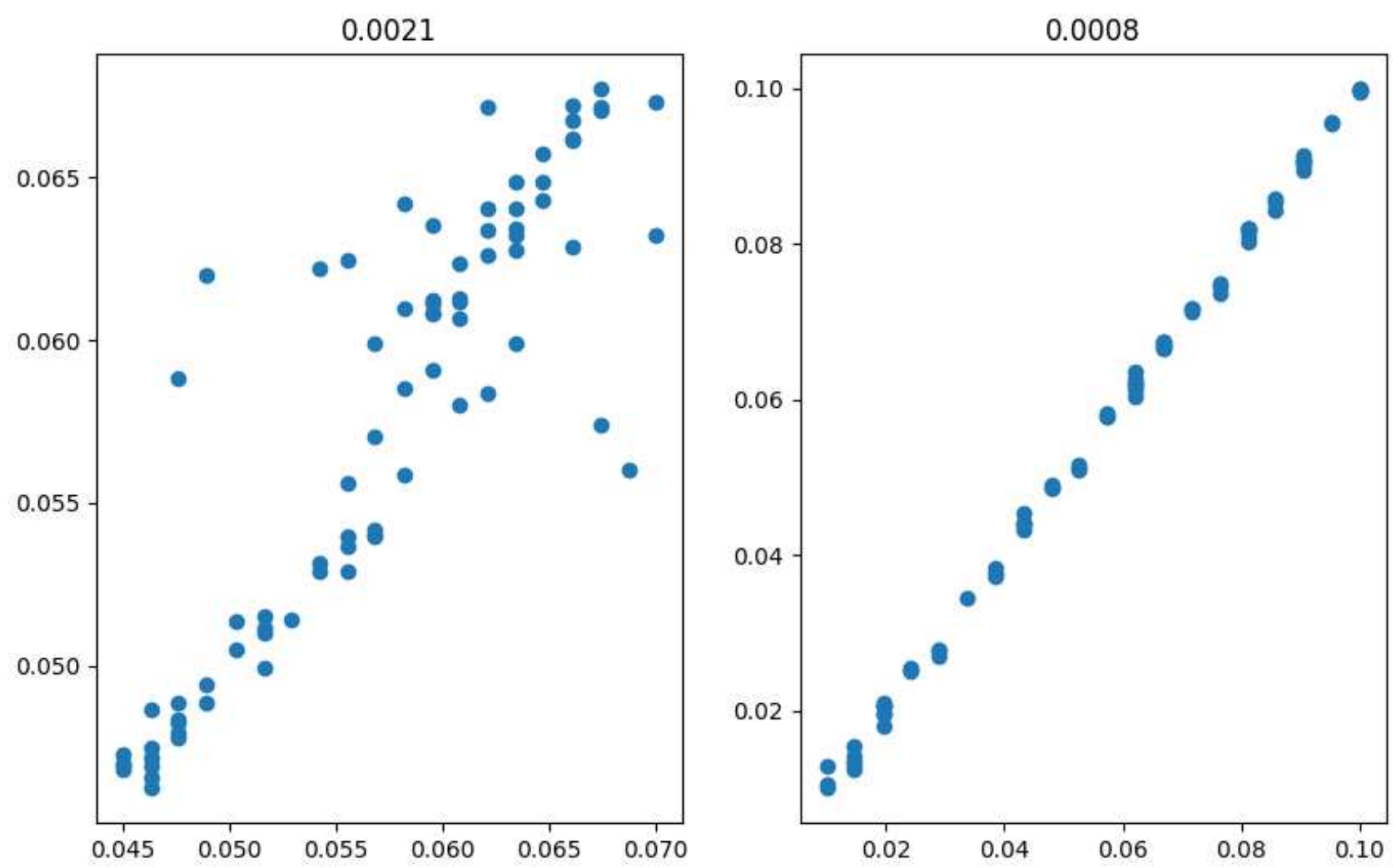


Рис. 4: 3D Random forest регрессия

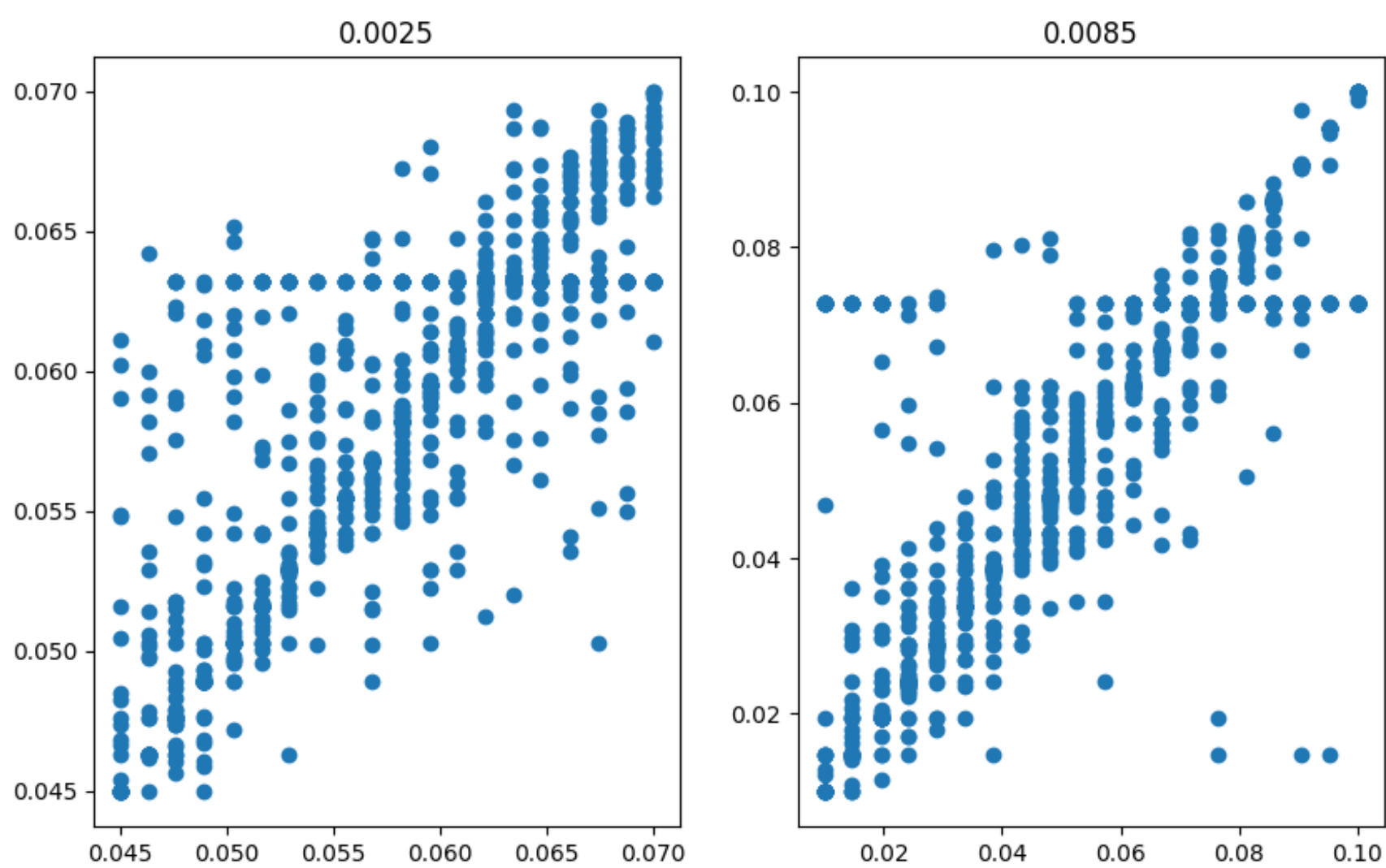


Рис. 5: Лучший результат на 2D данных (ансамбль с разделением по количеству A)

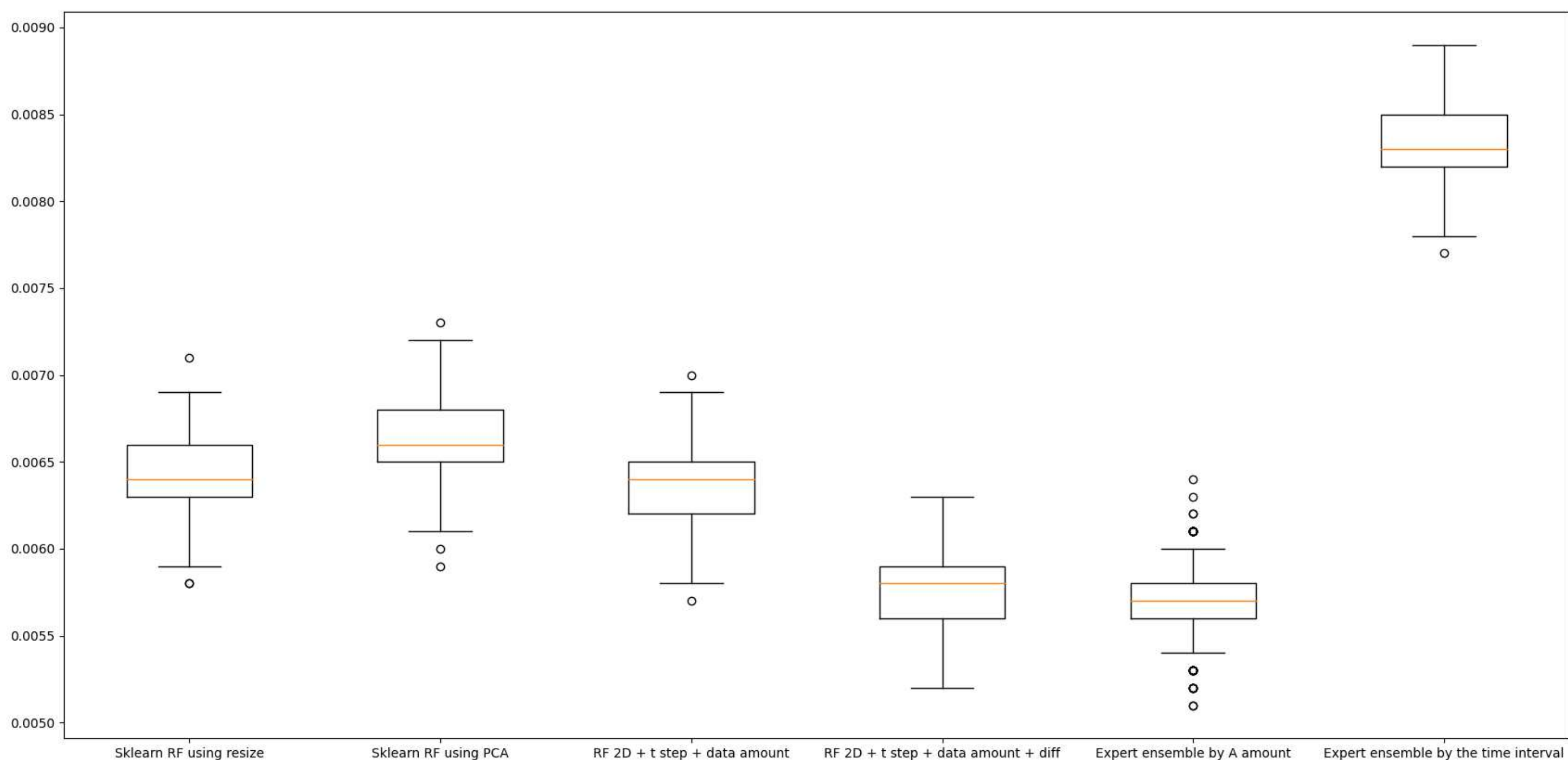


Рис. 6: Сравнение подходов к 2D данным