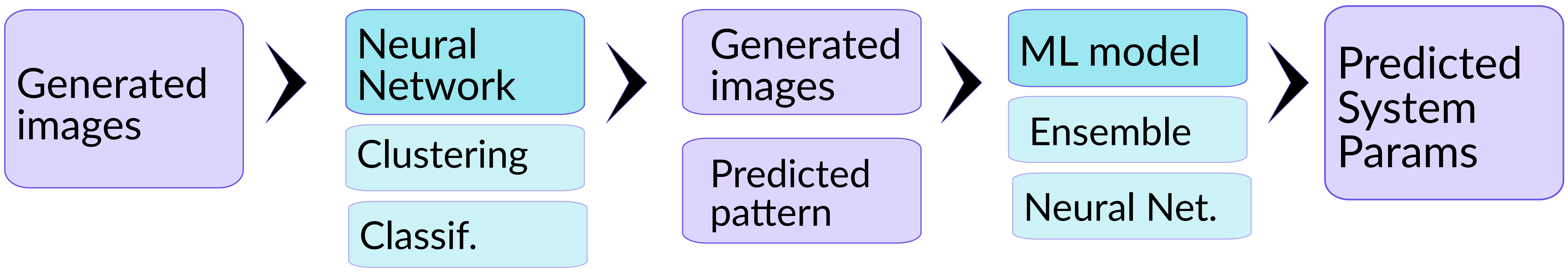
# Optimizing spatiotemporal models

Ilya Ivanov, Veronika Novgorodskaya, Vera Terenteva, Laura Avinyó

In development biology we retrieve lots of image data. We create lots o models as well, but we still do not have an standarized way to relate and the image data and the models to find out what are the specific values of the model to generate a specific set of data

In this work, we considered the reaction-diffusion model, which describes the formation of patterns, and the morphology of limbs in one of its versions

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




## Neural Networks approach

We used **neural networks** to cluster the images, to classify the images into patterns and to predict parameters using regression.

In order to reduce the dimension of our input data we used an **autoencoder neural networks**. This type of neural network consisting of two elements an encoder and a decoder. The encoder transforms the input data in a lower dimension vector, and the decoder transforms it back to the same dimension as the input data. The model aims to minimize the difference between the input and output data. Therefore the encoder reduces the dimension with minimal loss of information.

Using a **convolutional neural network** we have perform a classification task. *Fig 2.*

Next, we used convolutional neural networks (1D and 2D) to predict the value of parameters from the source images or encoded by the autoencoder, respectively. *Fig 3.*

Moreover we used the predicted class as an additional feature to the regressor.

## Ensemble approach

Another way to make predictions is using and **Ensemble of models**. This is a method, that in a specific way combines predictions of several models to achieve the best results. In our work we used bagging methods finding a **weighted mean** of each models' predictions.

We have splitted data into **bags** (randomly or by a specific value of a given column). To achieve the best accuracy we also added some weights that define the importance of each model's output — so the result of the best model is prioritized.
модели имеет приоритет.
A particular issue found is that some patterns are formed faster than the others and so that the last time points lack information for some patterns but not for others. Based on that, we have tried different tactics of splitting the data and changing the weights: (1) depending on exact **time intervals** where the images has being stored and (2) the **amount of chemical** as a proxy of pattern development. The last approach have shown the best result.

## Results

1. We can train machine learning models to predict parameters for doctors to use.
2. Some models behave better than others for this specific data set.
3. The use of ensemble using different smart ways to split the data is promising but still requires work.
4. The methods need to be more robust to be able to be used for other models, but we successfully have finished the proof of concept of the pipeline.

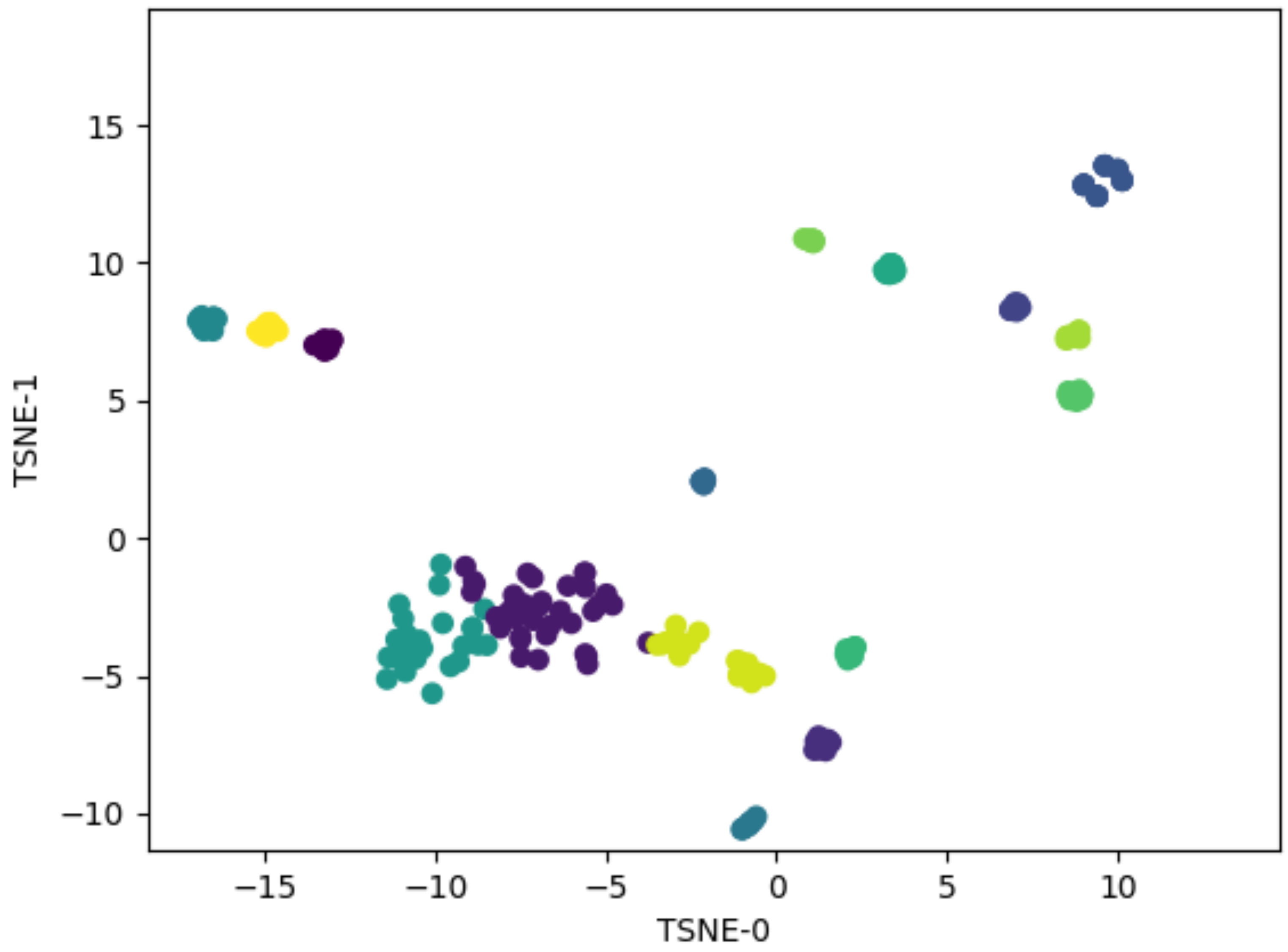



*Fig. 1:* Clustering based on encoded diffusion images

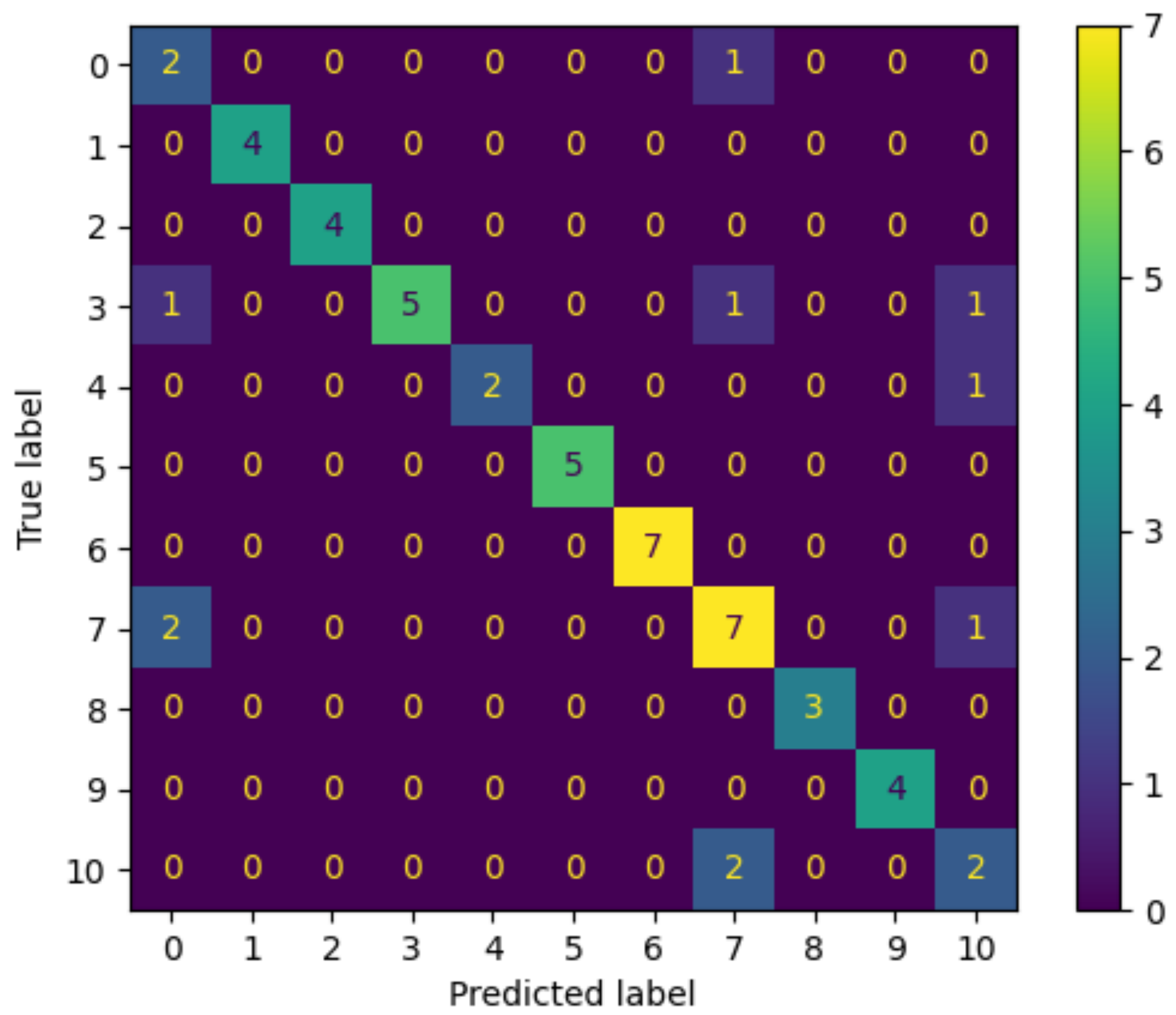



*Fig. 2:* Confusion matrix for the pattern classification task (using convolutional neural network)

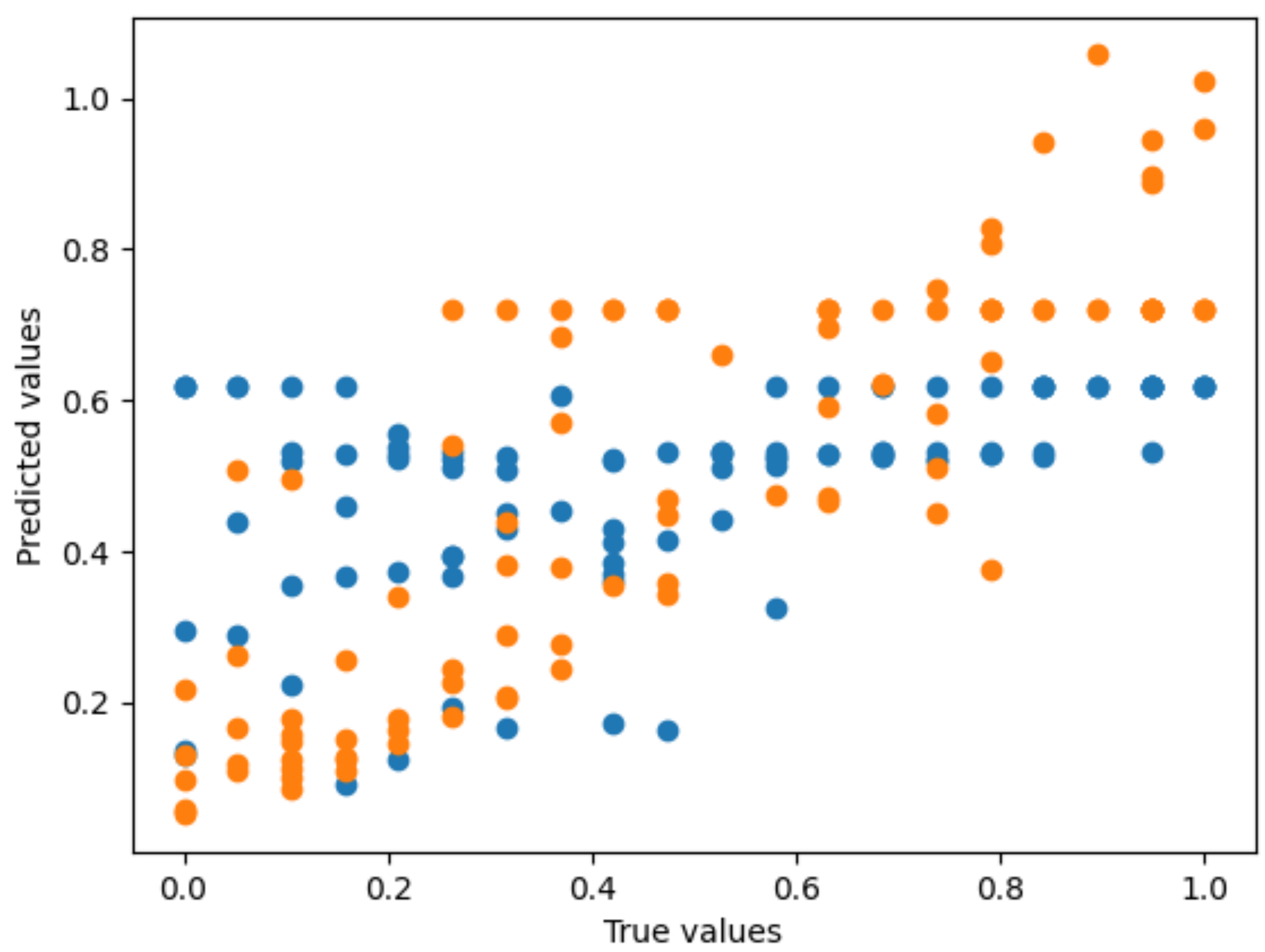



*Fig. 3:* Predicted values of feed rate and kill rate (encoded data regression using1D convolutional NN)

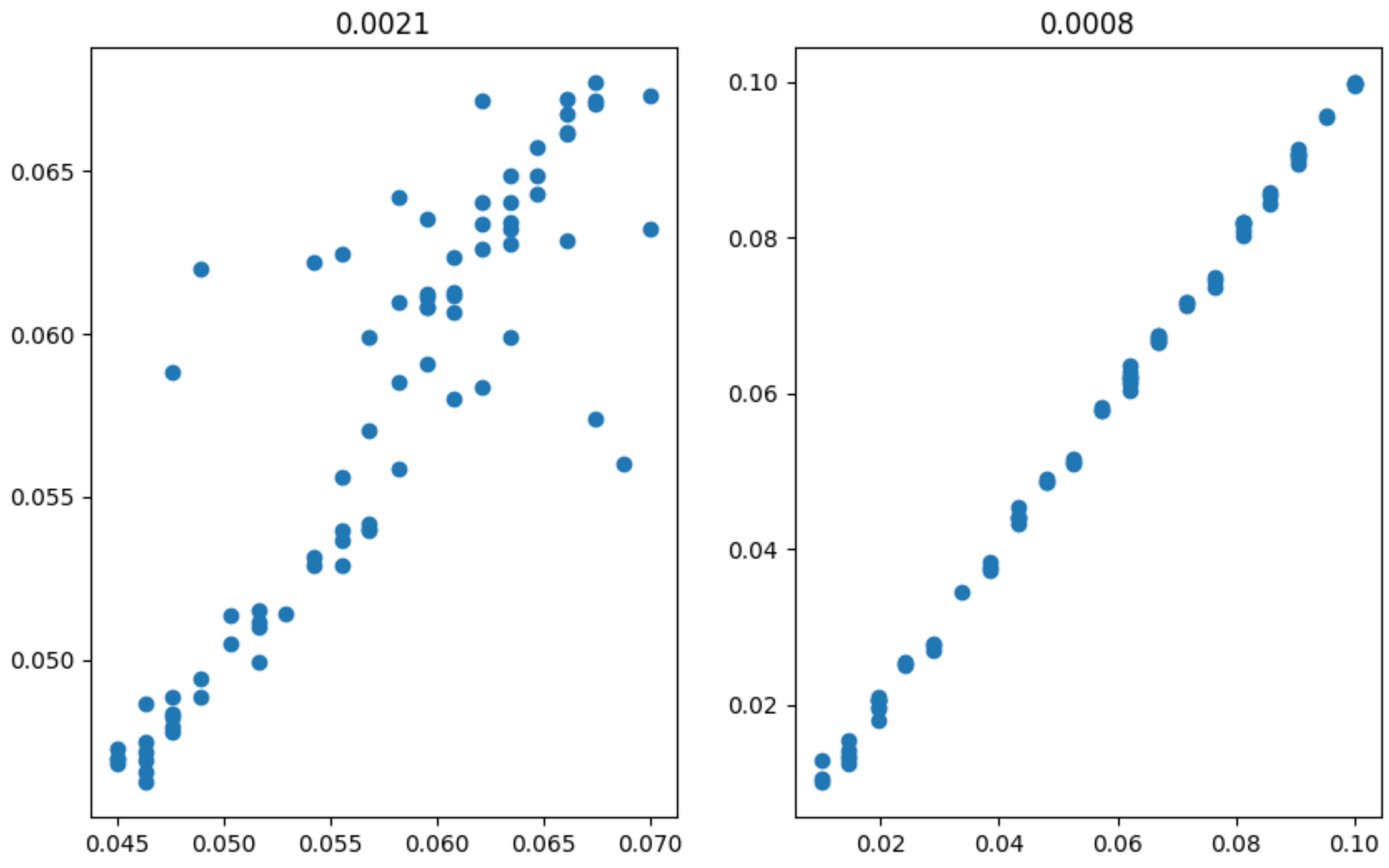



*Fig. 4:* 3D Random forest regression

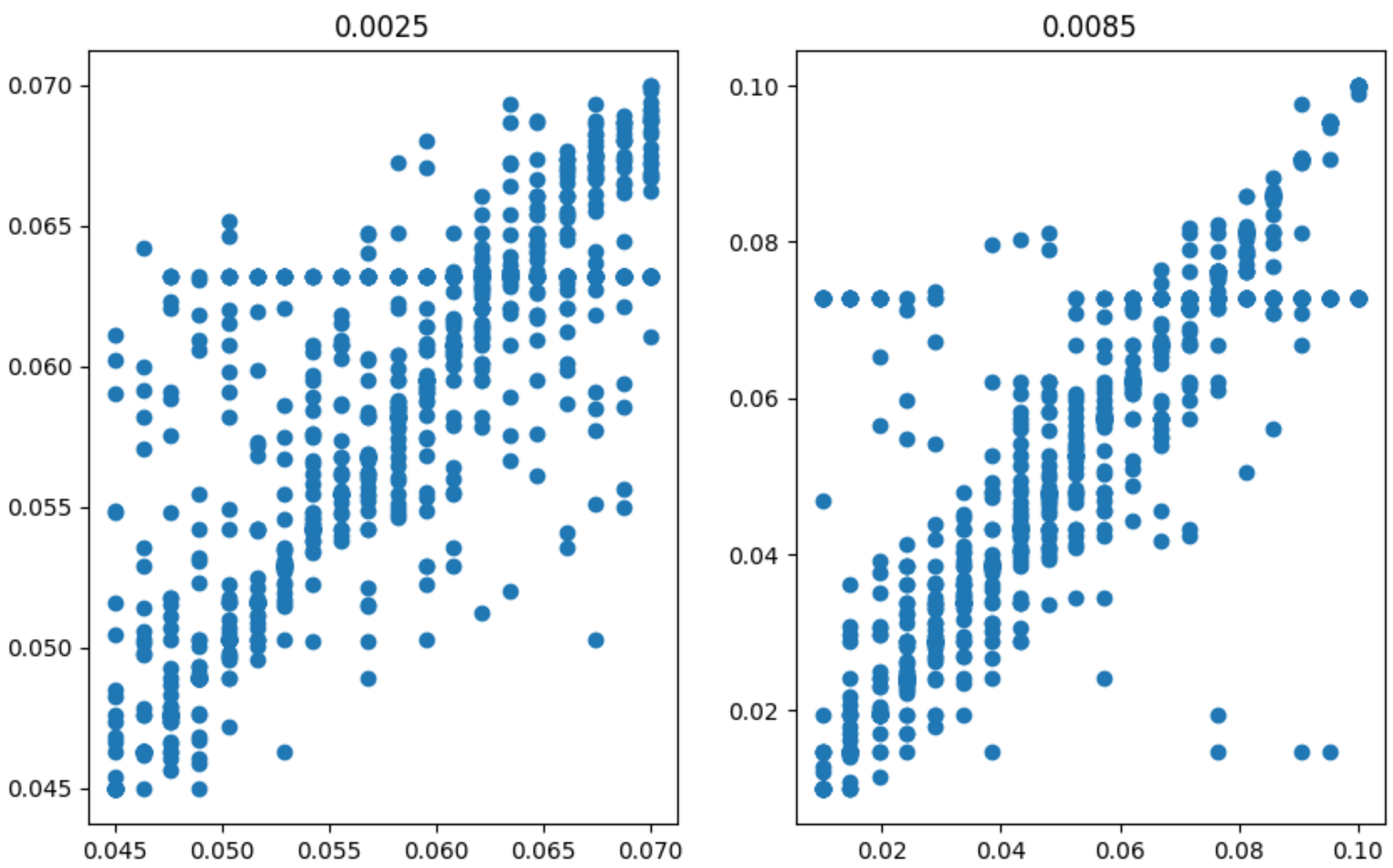



*Fig. 5:* The best result for 2D data
(Expert ensemble by amount of A)

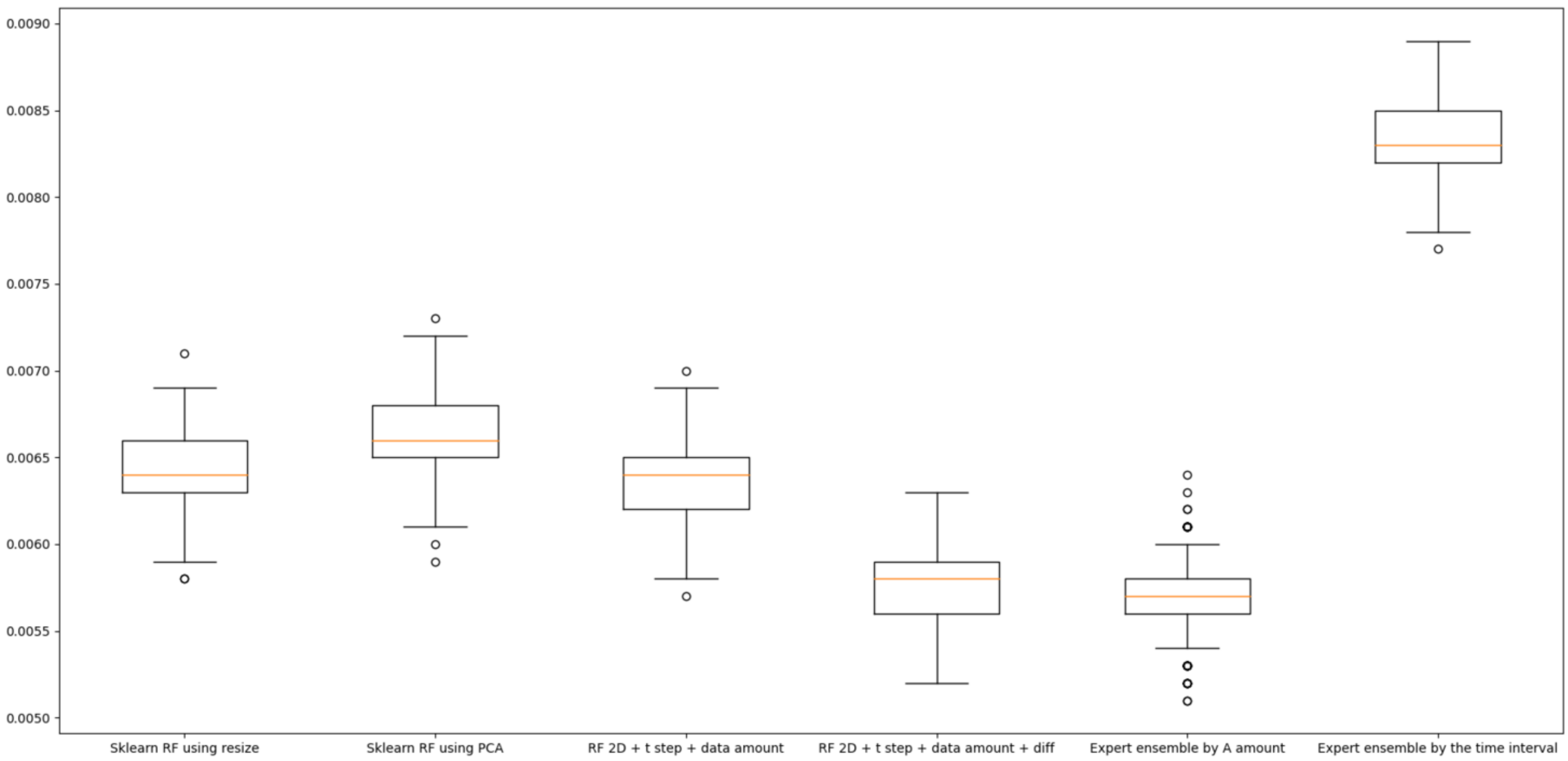



*Fig. 6:* Comparison of 2D data approaches