EPFL

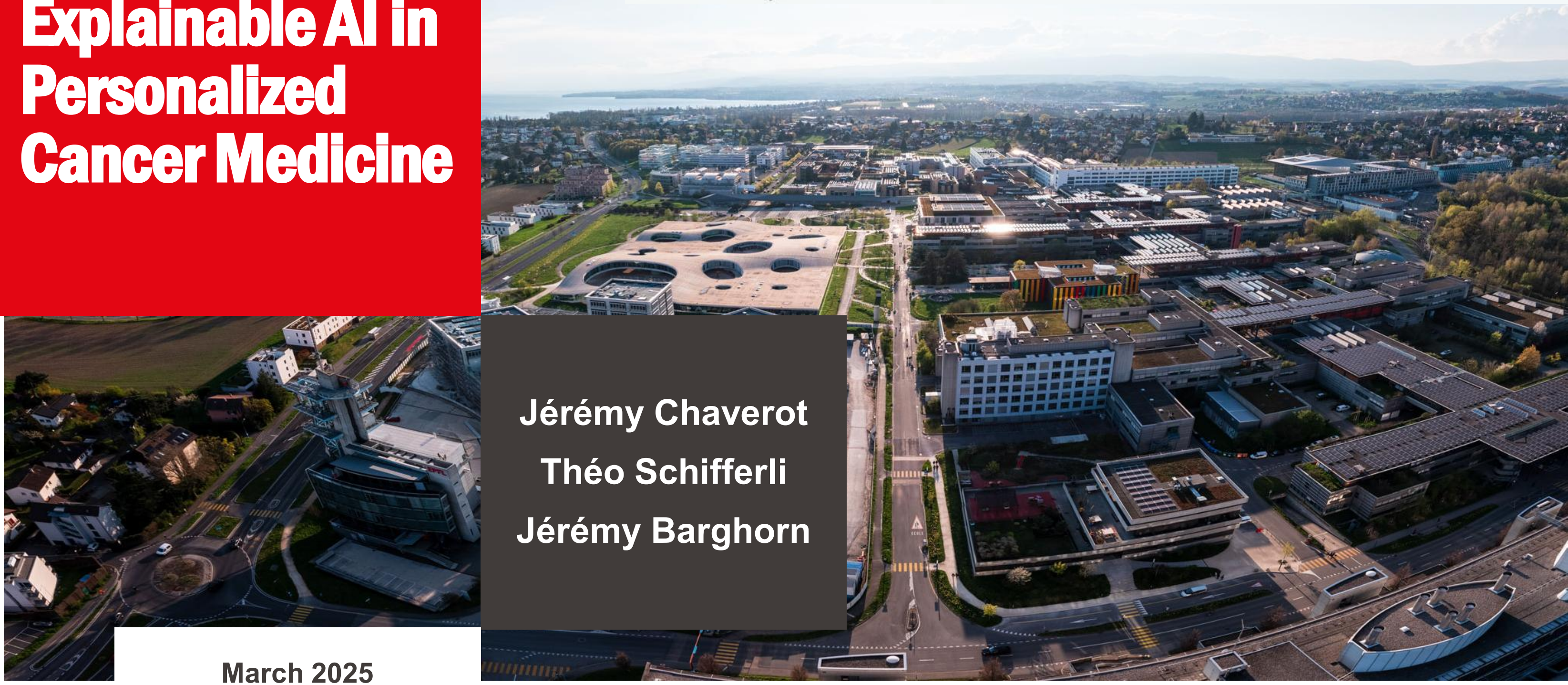
zhaw

2025 IEEE Symposium Series on Computational Intelligence

Explainable AI in Personalized Cancer Medicine

Jérémy Chaverot
Théo Schifferli
Jérémy Barghorn

March 2025



École polytechnique fédérale de Lausanne

About Us

Jérémy Barghorn

Théo Schifferli

Jérémy Chaverot

- Master's students in Data Science
- EPFL, Switzerland

- CS-433 Machine Learning course project by Prof. Martin Jaggi
- Supervised by Prof. Jasmina Bogojeska (ZHAW)

Jérémy Barghorn

O U T L I N E

Dataset curation and models

Results

Explainability

Challenges and future work

Jérémy Barghorn

Data Processing : Possibilities and Limitations

Cancer genomic data for Erlotinib

19 920 features

742 samples — Train data

304 samples — Test data

Specificity

- ⚠️ Specific to Erlotinib
- ⚠️ AAC as a label
- ⚠️ Gene alignment

💡 **Data augmentation**

CCLE dataset [1]

1460 features

1061 samples — Train data

304 samples — Test data

Limitations

- ✔ Gene expression level
- ❌ RNA sequence
- ❌ Phenotype data

- Single cell foundation models
- GenePT
- Graph models with prior knowledge

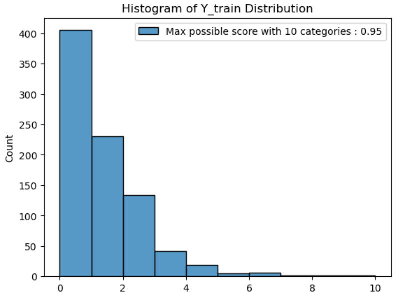
[2], [3], [4],[5], [6]

Inheritent Characteristics

- ⚠️ AAC a continuous metric
- ⚠️ Skewed data
- ⚠️ High dimensional

💡 **Target categorization**

💡 **Dimensionality reduction**



- PCA at the rescue ?
- Chi-square or ANOVA F-test

Models

Base dataset

Augmented dataset

Categorized targets and reduced dataset

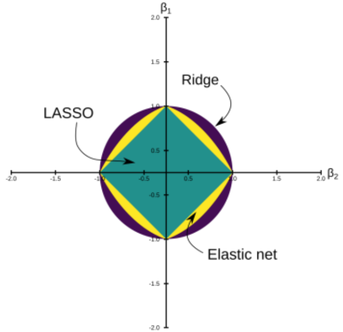
Models

Categorized targets and reduced dataset

Linear models

- Linear regression
- Ridge / Lasso regression
- Elastic Net

Baselines



Models

Categorized targets and reduced dataset

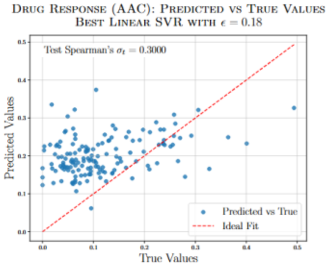
Linear models

Support Vector Regressors (SVR) models

- Linear SVR
- Sigmoïd SVR
- Polynomial / RBF SVR

Kernel methods for regression tasks

Tuned with grid search and cross validation



Jérémy Barghorn

Models

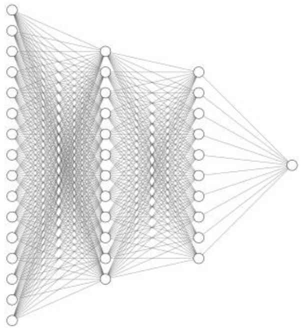
Categorized targets and reduced dataset

Linear models

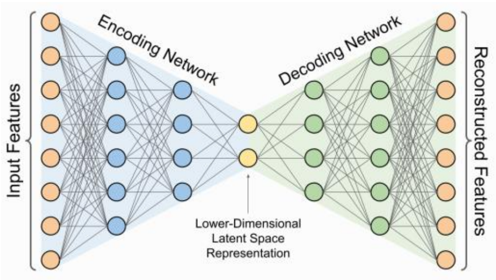
Support Vector Regressors (SVR) models

Neural Networks

- **Neural networks**
- Encoder Decoder models
- WPFS



- Stronger models that need more data
- How many layers for our **small** data?
- Data augmentation with noisy samples



Models

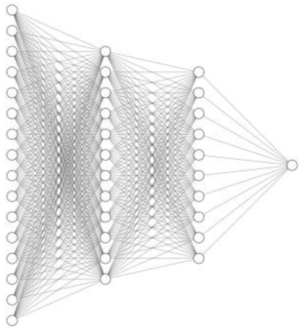
Categorized targets and reduced dataset

Linear models

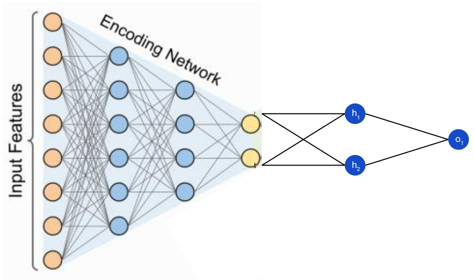
Support Vector Regressors (SVR) models

Neural Networks

- Neural networks
- **Encoder Decoder models**
- WPFS



- Stronger models that need more data
- How many layers for our **small** data?
- Data augmentation with noisy samples



Jérémy Barghorn

Models

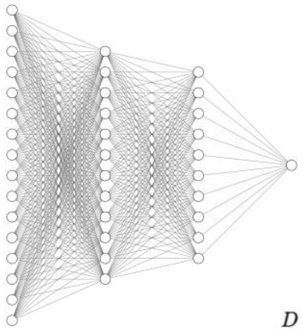
Categorized targets and reduced dataset

Linear models

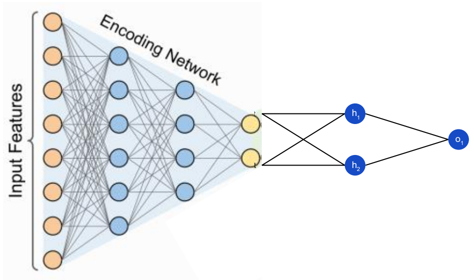
Support Vector Regressors (SVR) models

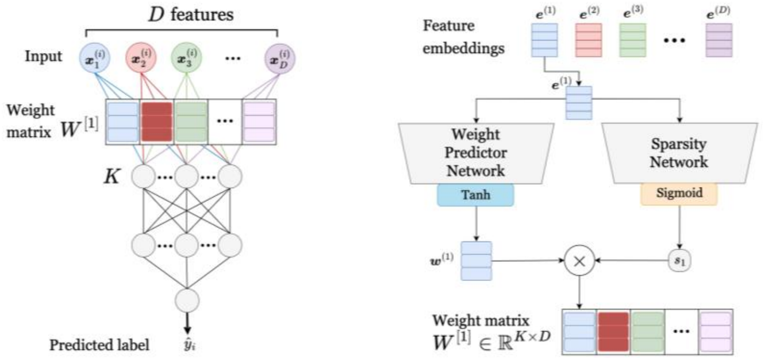
Neural Networks

- Neural networks
- Encoder Decoder models
- **WPFS**

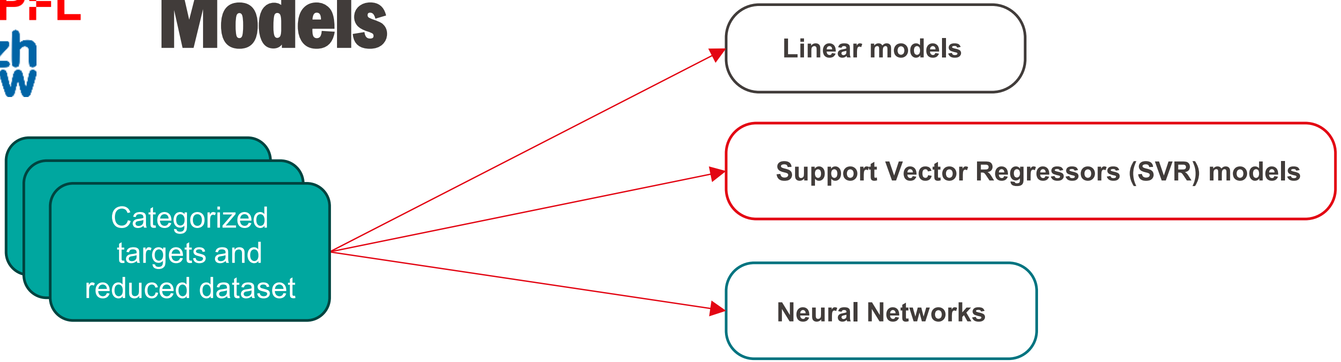


- Stronger models that need more data
- How many layers for our **small** data?
- Data augmentation with noisy samples





Models



Let’s see how the models performed !

OUTLINE

Dataset curation and models

Results

Explainability

Challenges and future work

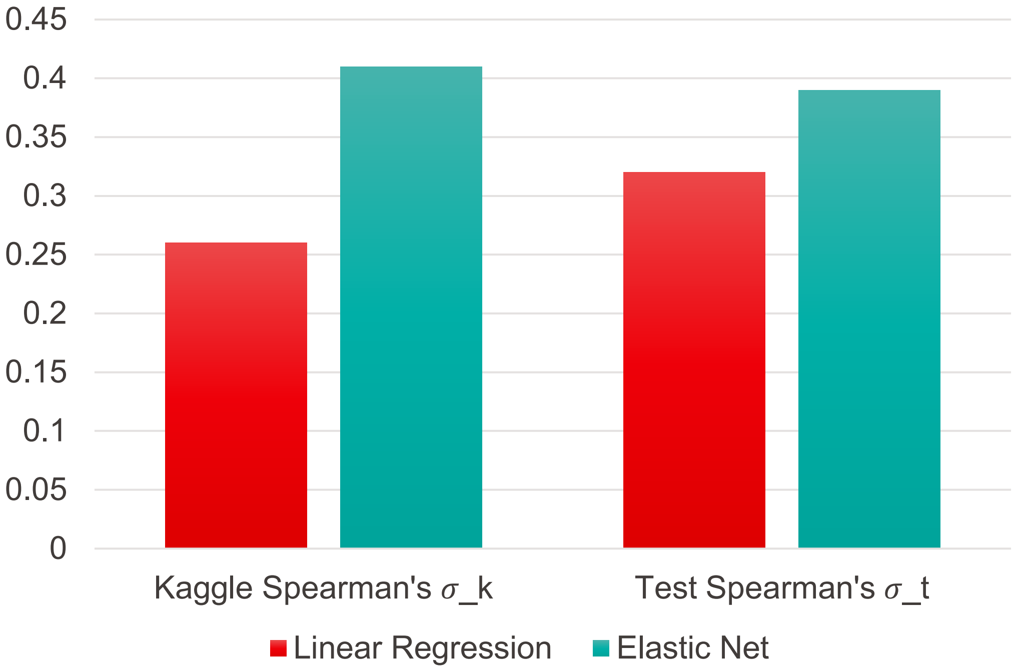
Results Table

Model	Kaggle Spearman's σ_k	Test Spearman's σ_t
Linear Regression	0.26	0.32
Elastic Net	0.41	0.39
Neural Network	0.39	**0.61**
Neural Network (w/ classes)	0.34	0.58
Encoder Decoder	0.16	0.21
WPFS	0.35	0.32
Linear SVR, $\epsilon = 0.18$	**0.58**	0.30
Linear SVR*, $\epsilon = 0.18$	-0.20	0.19
Sigmoid SVR, $\epsilon = 0.23$	0.56	0.30
Sigmoid SVR*, $\epsilon = 0.23$	0.53	0.36
Linear SVR, $\epsilon = 0.28$, $k = 500$	0.49	0.42

Models marked with * are trained on the augmented dataset.

TABLE I: RESULTING PERFORMANCE IN THE KAGGLE COMPETITION AND LOCALLY FOR THE BEST MODEL IN EACH CATEGORY.

Results: Linear Models



- Regularization effective for noisy data.
- Baseline with simple architecture.

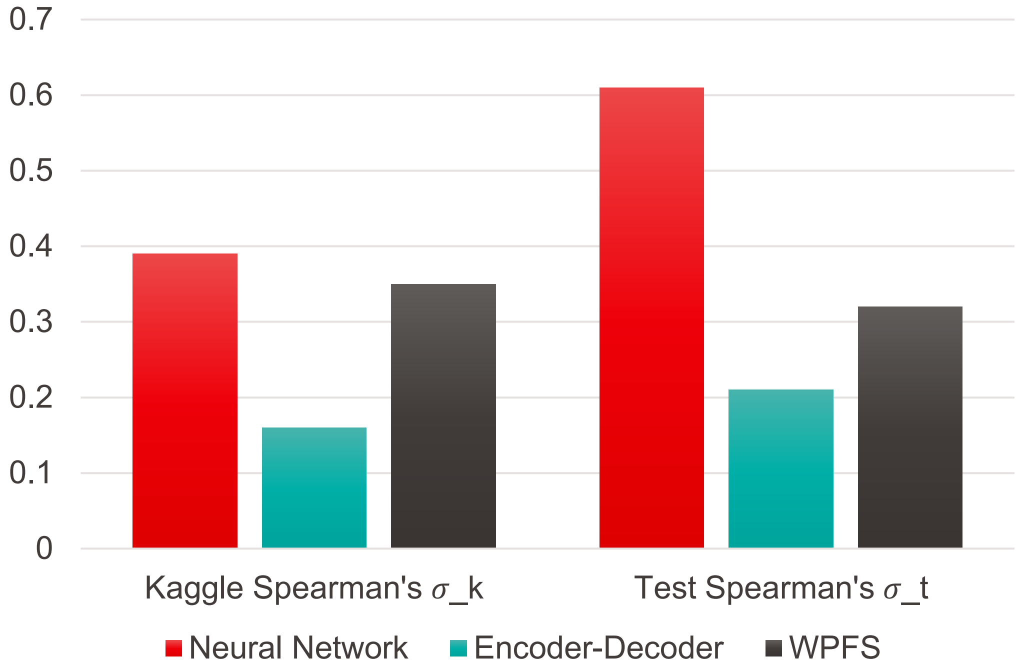
Model	Kaggle Spearman's σ_k	Test Spearman's σ_t
Linear Regression	0.26	0.32
Elastic Net	0.41	0.39
Neural Network	0.39	**0.61**
Neural Network (w/ classes)	0.34	0.58
Encoder Decoder	0.16	0.21
WPFS	0.35	0.32
Linear SVR, $\epsilon = 0.18$	**0.58**	0.30
Linear SVR*, $\epsilon = 0.18$	-0.20	0.19
Sigmoid SVR, $\epsilon = 0.23$	0.56	0.30
Sigmoid SVR*, $\epsilon = 0.23$	0.53	0.36
Linear SVR, $\epsilon = 0.28$, $k = 500$	0.49	0.42

Models marked with * are trained on the augmented dataset.

TABLE I: RESULTING PERFORMANCE IN THE KAGGLE COMPETITION AND LOCALLY FOR THE BEST MODEL IN EACH CATEGORY.

Jérémy Chaverot

Results: Neural Networks



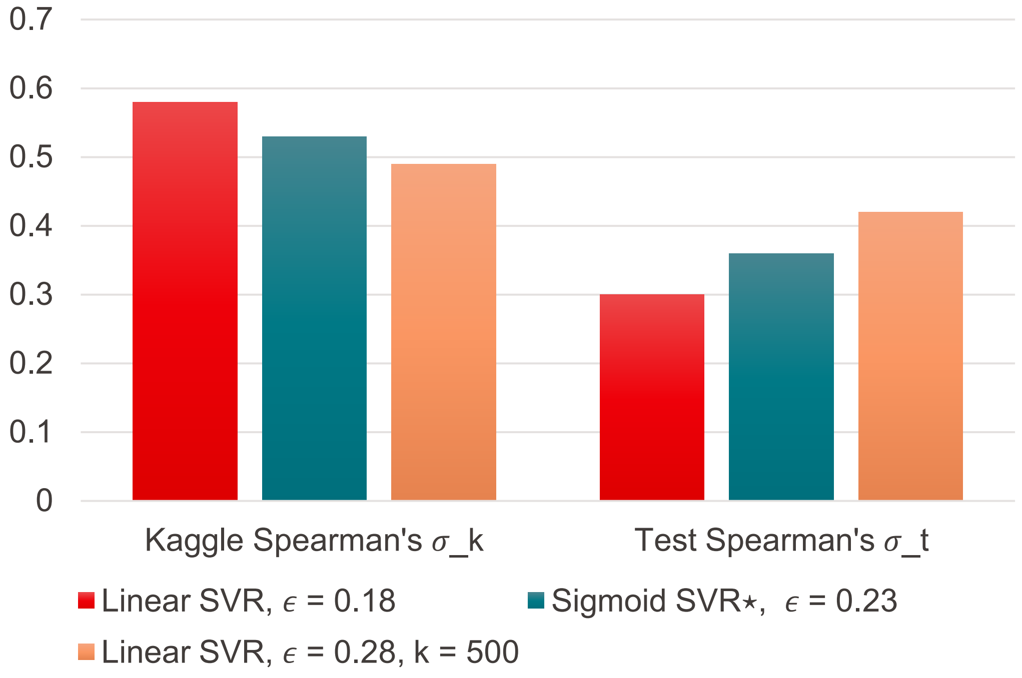
- Fast overfitting tendency.
- Deep models struggle with HDFS data.

Model	Kaggle Spearman's σ_k	Test Spearman's σ_t
Linear Regression	0.26	0.32
Elastic Net	0.41	0.39
Neural Network	0.39	**0.61**
Neural Network (w/ classes)	0.34	0.58
Encoder Decoder	0.16	0.21
WPFS	0.35	0.32
Linear SVR, $\epsilon = 0.18$	**0.58**	0.30
Linear SVR*, $\epsilon = 0.18$	-0.20	0.19
Sigmoid SVR, $\epsilon = 0.23$	0.56	0.30
Sigmoid SVR*, $\epsilon = 0.23$	0.53	0.36
Linear SVR, $\epsilon = 0.28$, $k = 500$	0.49	0.42

Models marked with * are trained on the augmented dataset.

TABLE I: RESULTING PERFORMANCE IN THE KAGGLE COMPETITION AND LOCALLY FOR THE BEST MODEL IN EACH CATEGORY.

Results: Support Vector Regressors



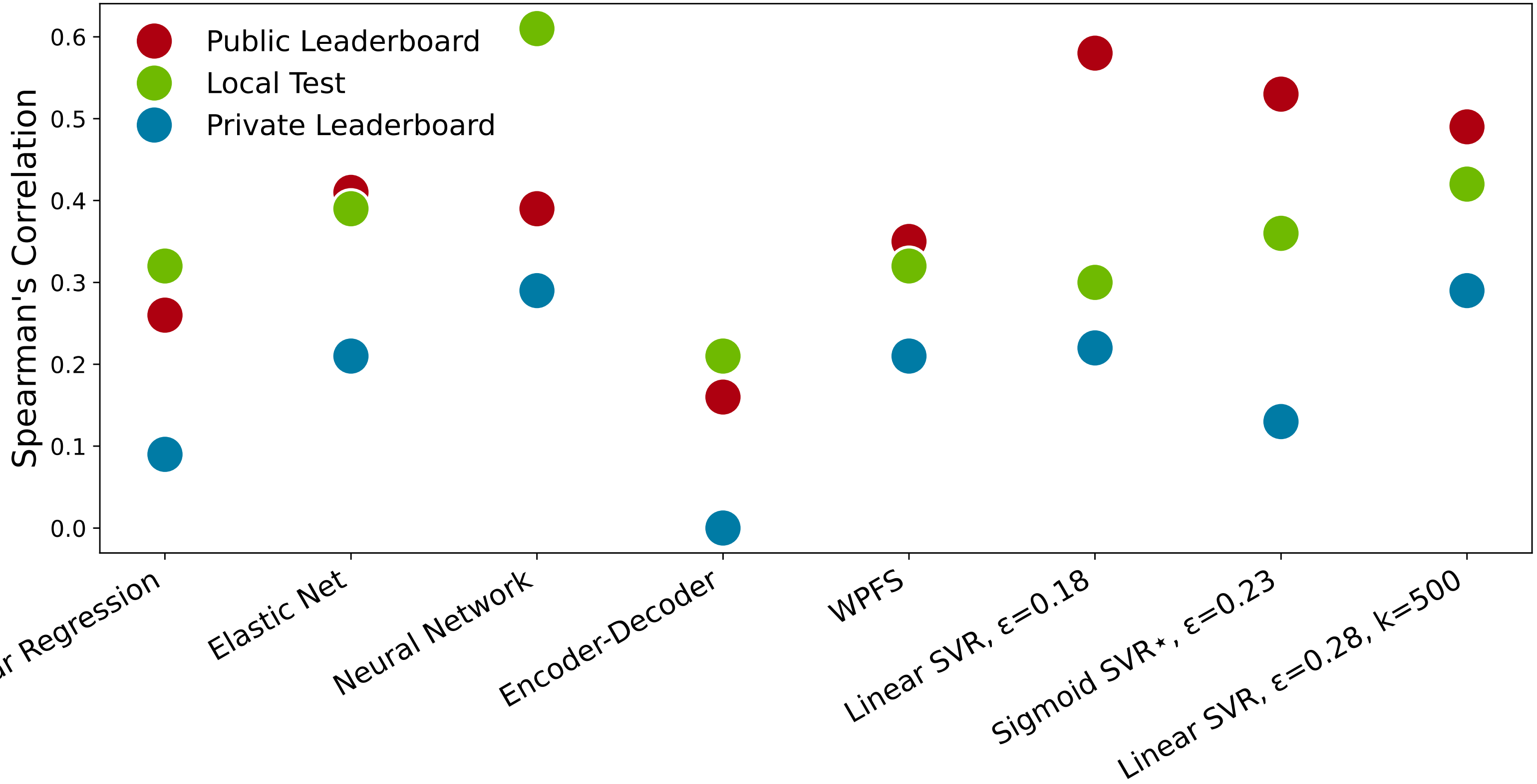
- Kernel-based methods are well suited for complex small-sized dataset.
- Data augmentation showed limited success.

Model	**Kaggle Spearman's σ_k**	**Test Spearman's σ_t**
Linear Regression	0.26	0.32
Elastic Net	0.41	0.39
Neural Network	0.39	**0.61**
Neural Network (w/ classes)	0.34	0.58
Encoder Decoder	0.16	0.21
WPFS	0.35	0.32
Linear SVR, $\epsilon = 0.18$	**0.58**	0.30
Linear SVR*, $\epsilon = 0.18$	-0.20	0.19
Sigmoid SVR, $\epsilon = 0.23$	0.56	0.30
Sigmoid SVR*, $\epsilon = 0.23$	0.53	0.36
Linear SVR, $\epsilon = 0.28$, $k = 500$	0.49	0.42

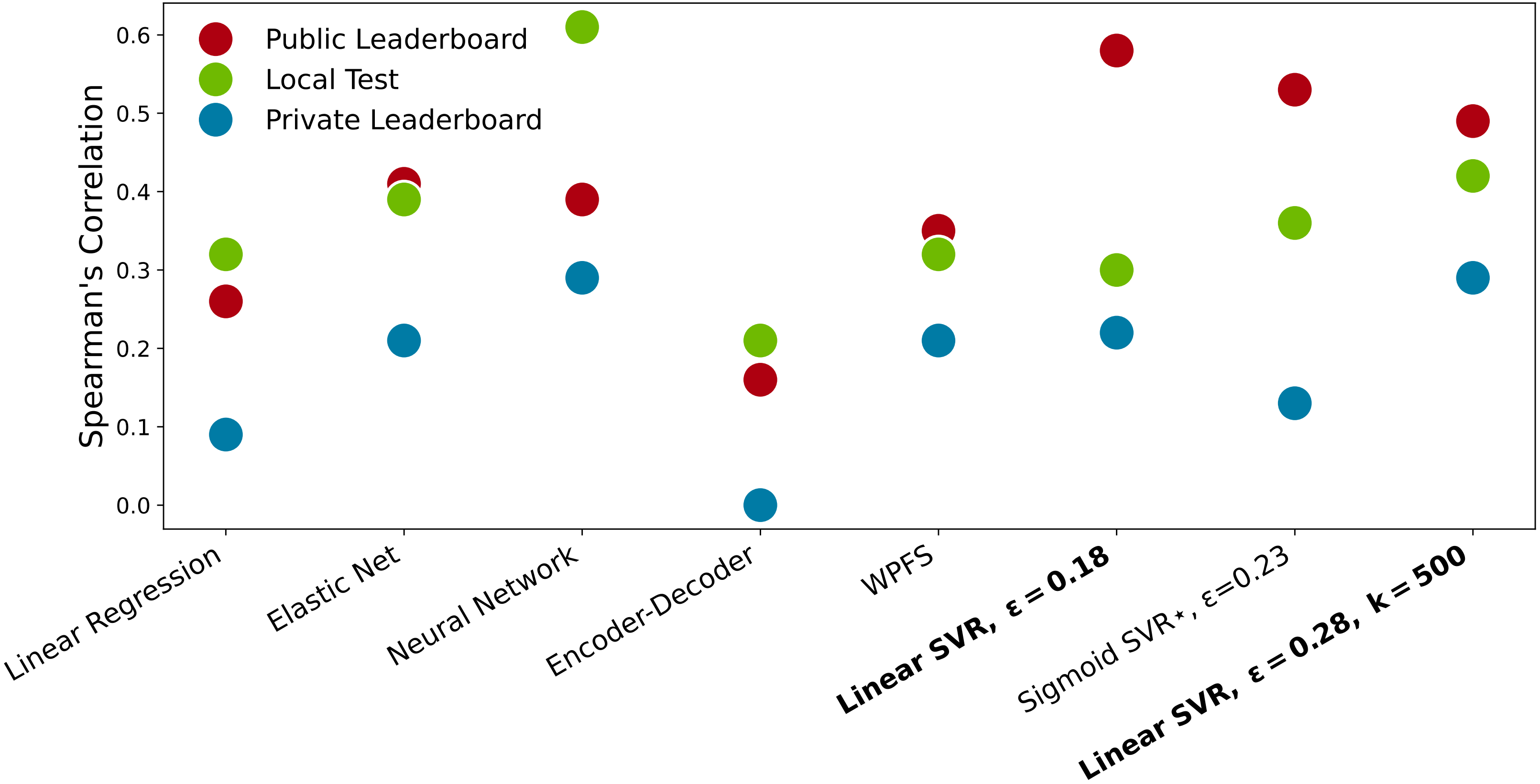
Models marked with * are trained on the augmented dataset.

TABLE I: RESULTING PERFORMANCE IN THE KAGGLE COMPETITION AND LOCALLY FOR THE BEST MODEL IN EACH CATEGORY.

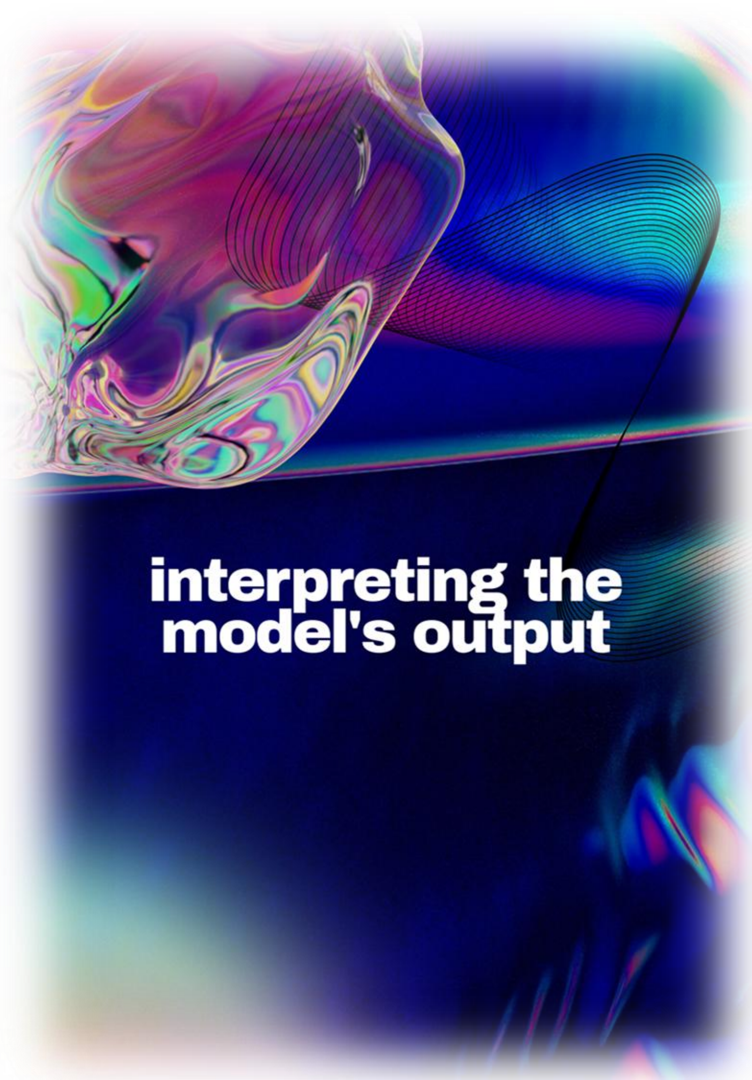
Results: Summary



Results: Best Models



Théo Schifferli



OUTLINE

Dataset curation and models

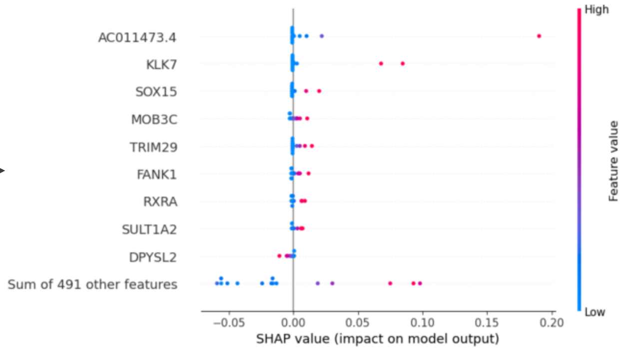
Results

Explainability

Challenges and future work

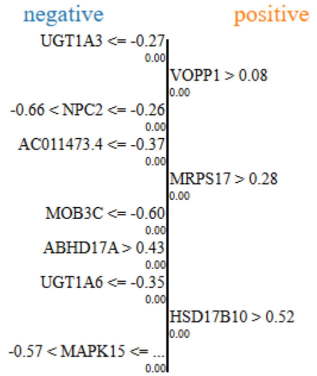
Two Approaches

Model Output → SHAP →



Model Output → LIME →

Local Focus Limitation

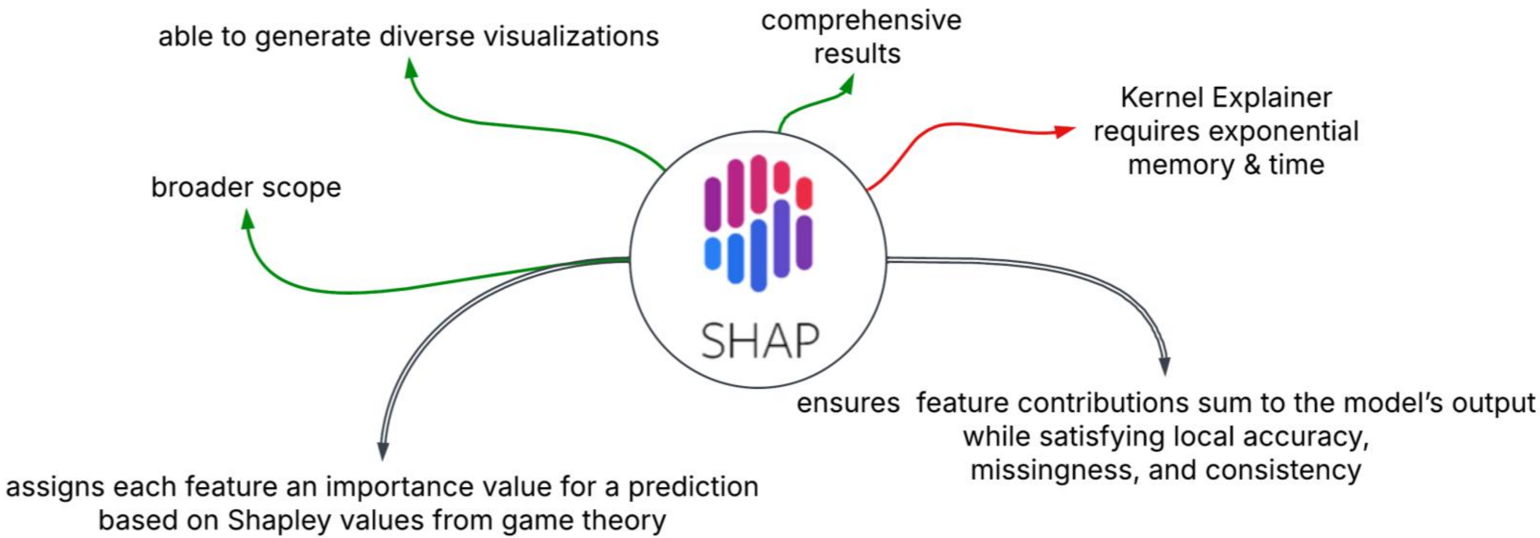


Feature	Value
UGT1A3	-0.27
VOPP1	0.73
NPC2	-0.48
AC011473.4	-0.37
MRPS17	0.82
MOB3C	-0.70
ABHD17A	1.19
UGT1A6	-0.35
HSD17B10	2.22
MAPK15	-0.56

More info available at : https://shap.readthedocs.io/en/latest/index.html
https://github.com/marcotcr/lime/tree/master

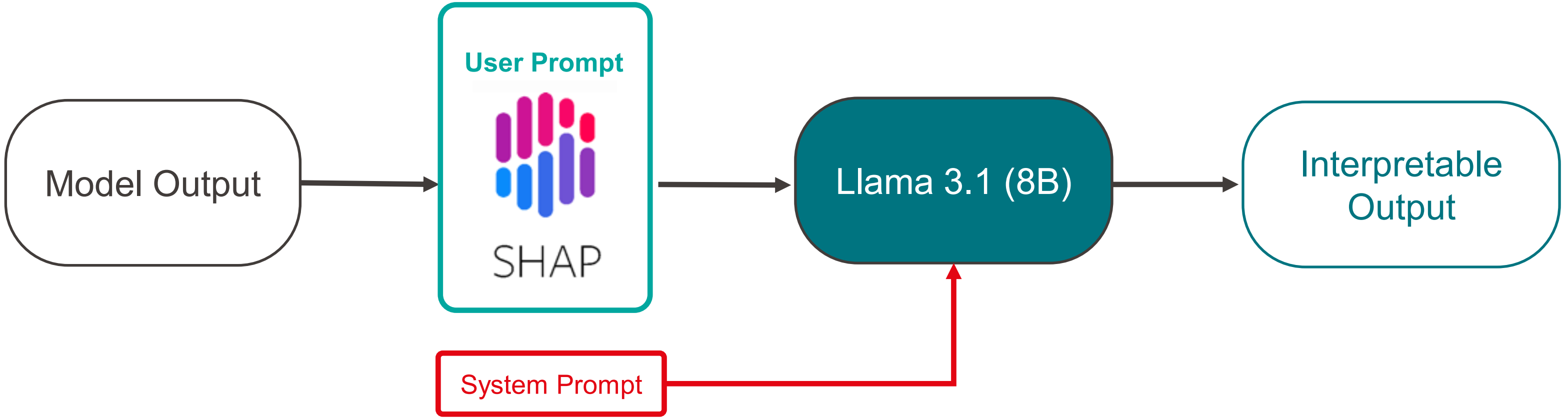
Théo Schifferli

Why SHAP?

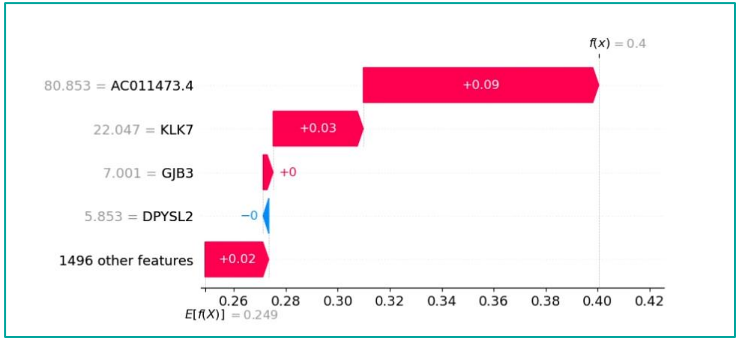

able to generate diverse visualizations
comprehensive results
Kernel Explainer requires exponential memory & time
broader scope
SHAP
ensures feature contributions sum to the model's output while satisfying local accuracy, missingness, and consistency
assigns each feature an importance value for a prediction based on Shapley values from game theory

Final Automated Pipeline and Example

Théo Schifferli



Concrete Example : TS28



Convert To Text

f(x) = Model Output : AAC

System Prompt:

- **Behave** like a medical expert
- Use **complex** and **specialized** vocabulary
- **Formal** and **precise** answers required

User Prompt :

- **Task :** explain significance of top features

Feature	SHAP Value	Feature Value
AC011473.4	0.0906273	80.8526
KLK7	0.0345304	22.0465
GJB3	0.0036523	7.00119
DPYSL2	-0.00210705	5.8528

Prediction : {AAC}
- Output **formatting**
- **Example** of output to follow (CoT manner)

Output Generation :

- **AC011473.4 (SHAP: 0.0906273):** Strong positive contribution, suggesting its high expression (80.8526) plays a key role in drug sensitivity, potentially indicating a favorable response to erlotinib.
- **KLK7 (SHAP: 0.0345304):** Moderate positive effect, with elevated expression (22.0465) potentially influencing tumor response, suggesting a possible synergistic interaction with erlotinib. […]
- **DPYSL2 (SHAP: -0.00215705):** Slight negative influence, suggesting reduced expression (5.8528) marginally lowers the AAC, indicating a potential antagonist effect on erlotinib response. […]

Jérémy Chaverot

OUTLINE

Dataset curation and models

Results

Explainability

Challenges and future work

Jérémy Chaverot

Challenges Faced

Challenges Faced

- **Severe Class Imbalance:** Most AAC values < 0.2, none above 0.7, limiting generalization.
- **High-Dimensional, Small Dataset:** With 19,920 features but only 743 samples, the model architecture choice and feature selection were critical.
- **Limited Generalization:** Models struggled to generalize, with larger networks overfitting and even the best SVR model performing poorly on the private leaderboard.
- **Ineffective Data Augmentation:** External datasets and added Gaussian noise had small impact.
- **Explainability:** SHAP Kernel Explainer required excessive memory & time (even on 384 GB RAM).

Future Work

- **Expert validation:** Work with clinicians and biologists to demonstrate the trustworthyness of the automatic tools (SHAP/LLaMA).
- **Curate more data** to improve model performances.
- **LLM Fine-Tuning & Alignment:** Improve SHAP result interpretation by fine-tuning on medical data (SFT) for greater domain knowledge, while applying preference alignment techniques (RLHF, DPO) to ensure results are presented in a user-preferred format.

Thank you

Questions

École
polytechnique
fédérale
de Lausanne