EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Faculty of Science

IKP STUTTGART

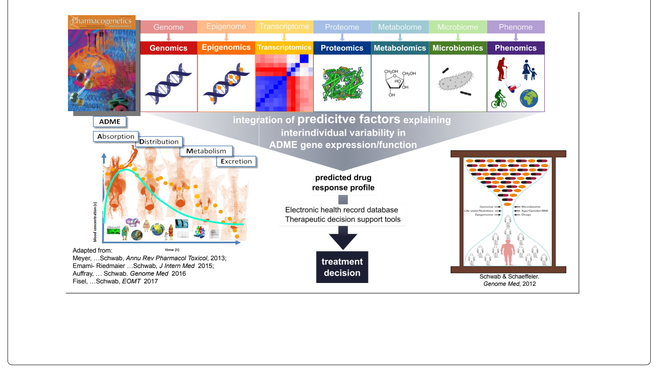
Jonas C. Ditz, Matthias Schwab, Nico Pfeifer

Extending Deep Kernel Approaches

Improving Prediction and Understanding in Precision Medicine

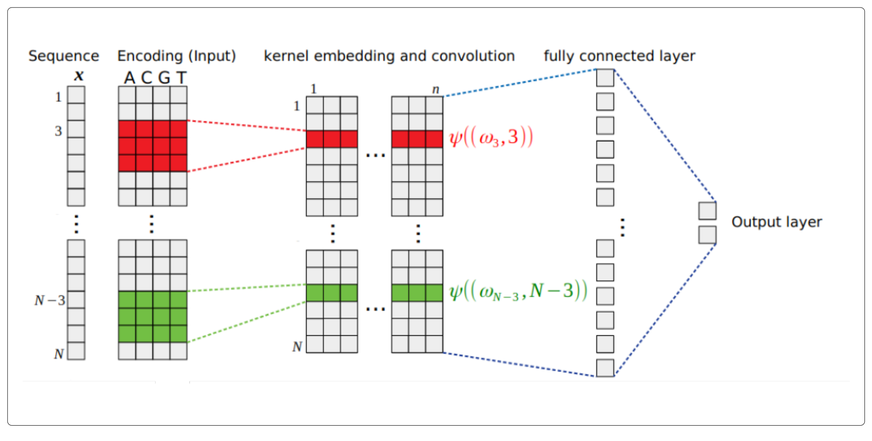
1 Background

Precision medicine is understood as a medical approach in which patients are stratified based on their disease subtype, risk, prognosis, or treatment response using specialized diagonstic tests. For example, genetic variations in ADME genes (up to 300) affect the drug response of patients and mutations of protein sequences determine the resistance of HI viruses against antiretroviral drugs. Interpretable machine learning can improve precision medicine by improving predictions and strengthening understanding.



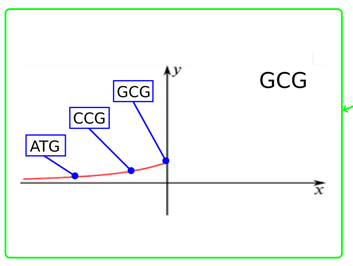
2 Methods

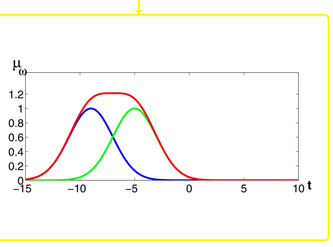
We developed Convolutional Motif Kernel Networks (CMKNs), an architecture that utilizes the expressiveness of our position-aware motif kernel together with the Nyström method to create models that provide interpretable end-to-end learning on biological sequences. Our method can be robustly trained on small datasets (< 1000 samples), but also scales to large datasets (> 100,000 samples). Additionally, CMKNs provide global and local interpretations without the need for *post-hoc* models.



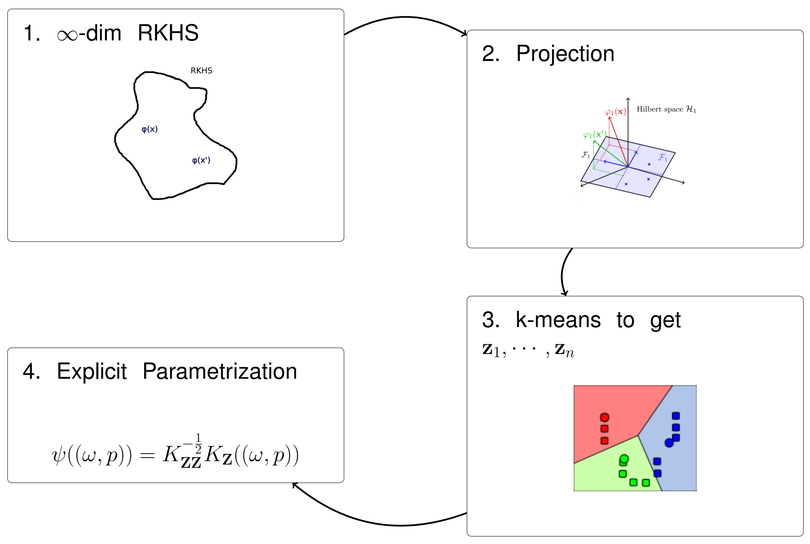
1. New Kernel Function

$$K((\omega,p),(\omega',q)) = \exp\left(\alpha\left(\omega^T\omega' - k\right) + \frac{\beta}{2\sigma^2}\left(\mathbf{p}_{\mathbf{x}}^T\mathbf{q}_{\mathbf{x}'} - 1\right)\right)$$





2. Computational Feasibility

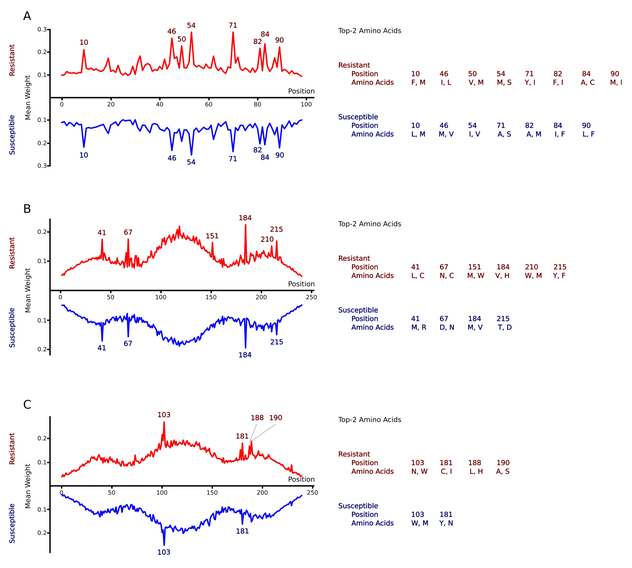


3 Results

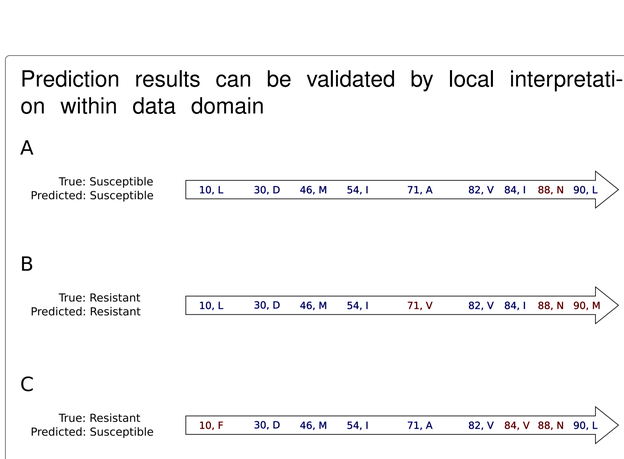
Similar or better performance compared to previous methods on HIV data

Drug	Type	Model	Accuracy	F1 Score	auROC	MCC
PI		SVM_{poly}	0.90 ± 0.04	0.83 ± 0.09	0.95 ± 0.03	0.75 ± 0.10
		SVM_{oligo}	**0.92** ± **0.03**	0.86 ± 0.09	**0.97** ± **0.03**	**0.81** ± **0.09**
		RF	**0.92** ± **0.04**	0.85 ± 0.13	**0.97** ± **0.03**	0.79 ± 0.13
		CNN	0.91 ± 0.3	0.84 ± 0.11	0.94 ± 0.05	0.77 ± 0.11
		CKN_{seq}	0.84 ± 0.05	0.72 ± 0.12	0.88 ± 0.05	0.60 ± 0.11
		CMKN	**0.92** ± **0.03**	**0.87** ± **0.09**	0.96 ± 0.03	**0.81** ± **0.10**
NRTI		SVM_{poly}	0.86 ± 0.06	0.82 ± 0.09	0.90 ± 0.05	0.70 ± 0.12
		SVM_{oligo}	0.88 ± 0.05	0.85 ± 0.09	**0.94** ± **0.03**	0.75 ± 0.10
		RF	0.88 ± 0.06	0.84 ± 0.12	**0.94** ± **0.04**	0.74 ± 0.15
		CNN	0.88 ± 0.05	0.85 ± 0.09	0.93 ± 0.04	0.74 ± 0.12
		CKN_{seq}	0.79 ± 0.06	0.73 ± 0.12	0.85 ± 0.05	0.54 ± 0.13
		CMKN	**0.89** ± **0.05**	**0.86** ± **0.09**	0.93 ± 0.05	**0.76** ± **0.11**
NNRTI		SVM_{poly}	0.82 ± 0.06	0.76 ± 0.11	0.84 ± 0.06	0.63 ± 0.14
		SVM_{oligo}	0.89 ± 0.05	0.86 ± 0.11	0.94 ± 0.05	0.79 ± 0.12
		RF	0.88 ± 0.05	0.85 ± 0.09	0.93 ± 0.07	0.75 ± 0.12
		CNN	0.89 ± 0.04	0.86 ± 0.08	0.94 ± 0.06	0.78 ± 0.10
		CKN_{seq}	0.73 ± 0.06	0.63 ± 0.16	0.78 ± 0.08	0.42 ± 0.15
		CMKN	**0.91** ± **0.03**	**0.89** ± **0.06**	**0.95** ± **0.05**	**0.81** ± **0.08**

Biologically meaningful global interpretation



Prediction results can be validated by local interpretation within data domain



4 Further Work

- Utilize CMKN for different tasks and patient cohorts (e.g. tuberculosis, ADME phenotypes; TCGA, CYPTAM)
- Extend kernel networks to omics data
- Interpretable kernel methods for health care

University of Tübingen · Faculty of Science
Methods in Medical Informatics
Sand 14 · 72076 Tübingen · Germany
Telefon +49 7071 29-70482 · pfeiferlab.org
Dr. Margarete Fischer Bosch-Institute of Clinical Pharmacology
Bosch Health Campus · Stuttgart · Germany

References:

- Ditz, Jonas C., Bernhard Reuter, and Nico Pfeifer. "Convolutional Motif Kernel Networks." arXiv preprint arXiv:2111.02272 (2021).
- Mairal, Julien. "End-to-end kernel learning with supervised convolutional kernel networks." Advances in neural information processing systems 29 (2016).
- Fröhlich, Holger, et al. "From hype to reality: data science enabling personalized medicine."BMC medicine 16.1 (2018): 1-15.