# Modellalapú gépi tanulás a jelfeldolgozásban

Bognár Gergő

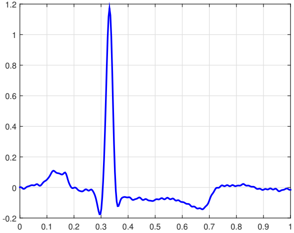
ELTE IK Numerikus Analízis Tanszék

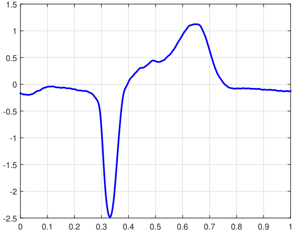
Analízis és Alkalmazásai Workshop, 2024

# Outline

# Outline

# Motivations: healthcare



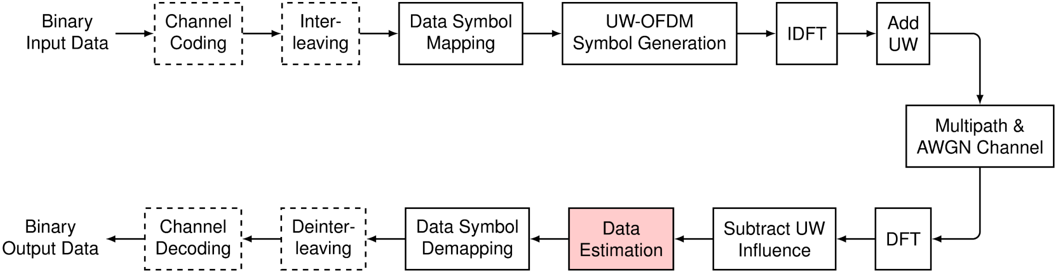

## Tasks

- Biomedical signal processing via modeling and machine learning
- *ECG heartbeat classification for arrhythmia detection*

## Expectations

- Accuracy, efficiency, explainability

# Motivations: telecommunication




## Tasks

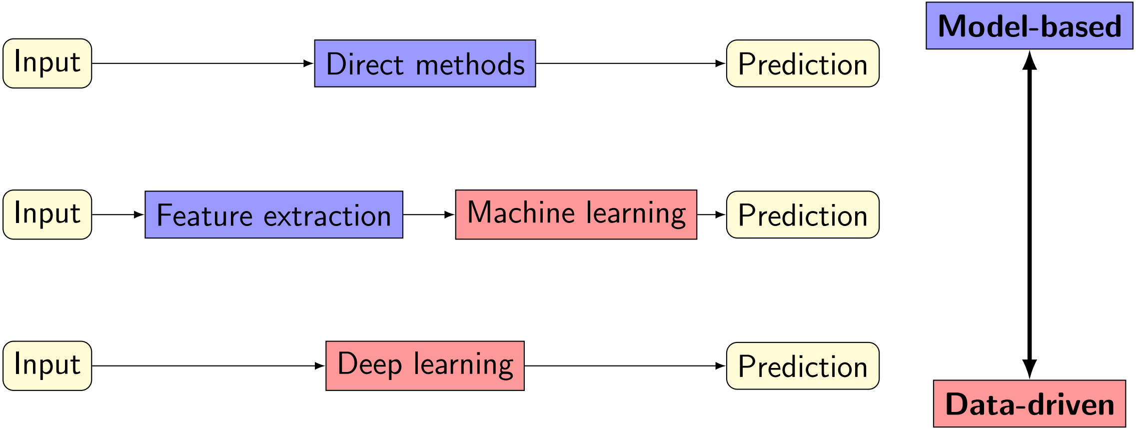
- Physical layer transmission in wireless communication
- *Data estimation in UW-OFDM systems*

## Expectations

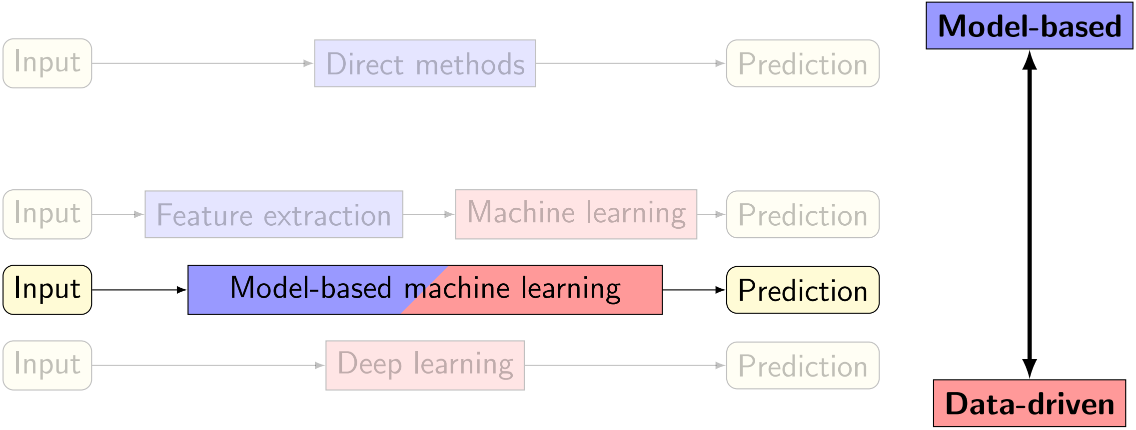
- Accuracy, efficiency, theoretically optimal solution

# Outline

# Prediction techniques



# Prediction techniques



# Model-based machine learning

## Advantages

- *Bridge between model-based direct methods and machine learning*
- Domain knowledge incorporation
- Model-based representation learning
- Compact, low-dimensional, optimized representation
- Interpretable parameters, explainable representation

## Challenges

- *Why?* – modeling vs. learning
- *What?* – model selection, parametrization, mathematical description
- *How?* – specialized architecture development

# Outline

# Least squares problems

## Least squares data estimation

- General modeling problem:

$$x \approx \hat{x} = \hat{x}(\theta), \qquad \|x - \hat{x}\|_2^2 \to \min_{\theta}$$

$\theta$: linear or nonlinear system parameters

## Gradient-based optimization

- Gradient descent iteration:

$$\theta^{(k+1)} := \theta^{(k)} - \delta \cdot \nabla_{\theta} \|x - \hat{x}\|_2^2$$

- Projected gradient descent:

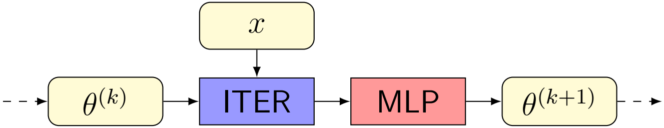
$$\theta^{(k+1)} := \Pi \left(\theta^{(k)} - \delta \cdot \nabla_{\theta} \|x - \hat{x}\|_2^2\right)$$

# Deep unfolding

## Concept

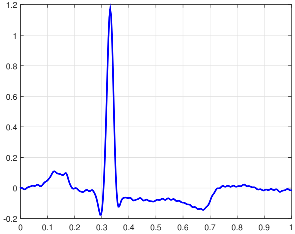
- Projected gradient descent:
$$\theta^{(k+1)} := \Pi\left(\theta^{(k)} - \delta \cdot \nabla_\theta \|x - \hat{x}\|_2^2\right)$$
- Unfolding iterations to NN layers
$$\theta^{(k+1)} := \mathrm{MLP}\left(\theta^{(k)} - \delta \cdot \nabla_\theta \|x - \hat{x}\|_2^2\right)$$
- Representation learning, combination with dense layers

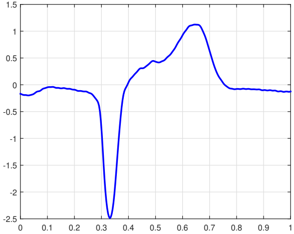
## Deep unfolding layer structure



# Outline

# Application: Healthcare




## Task

- ECG heartbeat classification on MIT-BIH Arrhythmia Database (PhysioNet)
- 5 AAMI classes, inter-patient paradigm (DS1 and DS2)[1]

[1] P. de Chazal, M. O'Dwyer, R. B. Reilly: Automatic classification of heartbeats using ECG morphology and heartbeat interval features, IEEE Trans Biomed Eng, 2004

# Variable projections[2] (VP, VarPro)

## Separable non-linear least squares

- Parametric function system: $\Phi_k(\theta) \in \mathbb{R}^m$, $\theta$: non-linear system parameters
- Non-linear modeling problem:

$$x \approx \hat{x} = \sum_{k=1}^{n} c_k \Phi_k(\theta) = \Phi(\theta)c, \qquad r(c,\theta) := \|x - \Phi(\theta)c\|_2^2 \to \min_{c,\theta}$$

- VP functional, Hilbert space approximation:

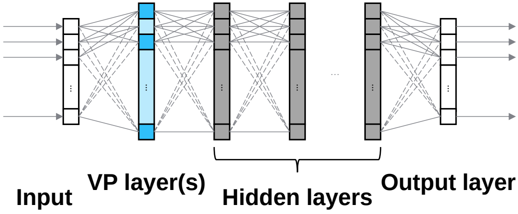
$$r_2(\theta) := \|x - \Phi(\theta)\Phi^+(\theta)x\|_2^2 \to \min_{\theta}, \qquad c = \Phi^+(\theta)x$$

$\Phi^+(\theta)$: Moore–Penrose pseudoinverse of matrix $\Phi(\theta)$

- Gradient-based optimization possible (gradient descent, Gauss–Newton, Levenberg–Marquardt, . . . )

[2] G. H. Golub, V. Pereyra: The Differentiation of Pseudo-Inverses and Nonlinear Least Squares Problems Whose Variables Separate, SIAM Journal on Numerical Analysis, 1973

# VPNet: Variable Projection Networks[3]




- Model-based neural network with VP representation learning
- VP layers: VP projections for feature learning:

$$x \mapsto f^{(\mathrm{vp})}(x) = \Phi^{+}(\theta)x = c \qquad \text{(classification), or}$$

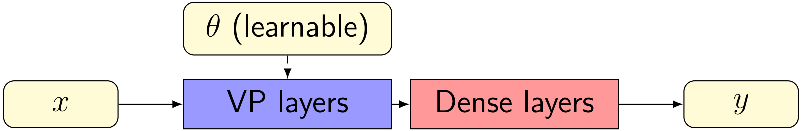
$$x \mapsto f^{(\mathrm{vp})}(x) = \Phi(\theta)\Phi^{+}(\theta)x = \hat{x} \qquad \text{(regression)}$$

- Different variants: autoencoder, spiking NN, SVM, ...

[3] P. Kovács, G. Bognár, C. Huber, M. Huemer: VPNet: Variable Projection Networks, International Journal of Neural Systems, 2022

# Deep unfolding VP

## Original VPNet

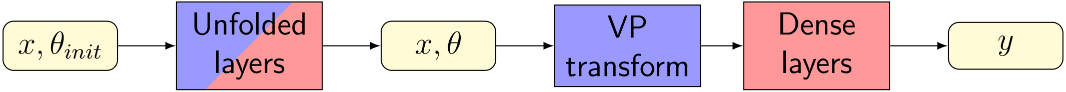



## Deep unfolding variable projection network

- Motivation: expand VPNet to learn to learn (sic!) system parameters $\theta$
- Unfolding the VP gradient iteration:

$$\theta^{(k+1)} := \text{MLP}\left(\theta^{(k)} + 2\delta\left(x - \Phi(\theta)\Phi^{+}(\theta)x\right)^{T}\mathbf{D}\Phi(\theta)\Phi^{+}(\theta)x\right)$$

- Exact gradient (and gradient of gradient) computation for numerical stability



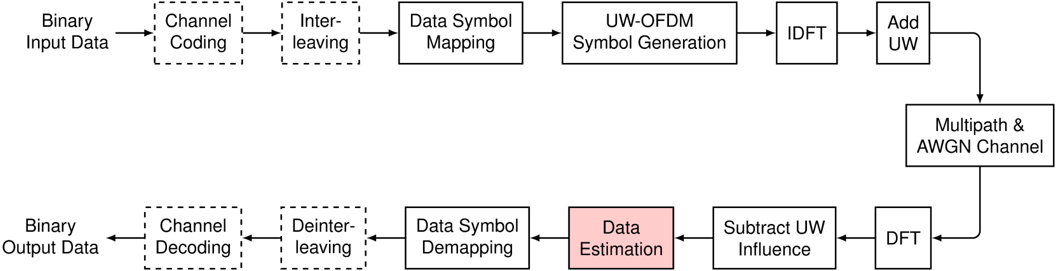
# ECG classification results

| Method | Description | | Accuracy |
|---|---|---|---|
| de Chazal et al. | Waveform + RR | LD | 86.1% |
| Llamado et al. | Waveform + wavelet + RR | LD | 93% |
| Ye et al. | Wavelet + ICA + RR | SVM | 86% |
| Dózsa et al. | Hermite VP (LC + NLP + PRD) + RR | SVM | 93.6% |
| Bognár et al. | Rational VP (LC + NLP) + RR | SVM | 94.5% |
| | Hermite VPNet | | 91.9% |
| | Hermite VPNet + RR | | 93.2% |
| **Bognár et al.**[4] | Hermite VP Unfold | | 93.5% |
| | Hermite VP Unfold + RR | | **94.7%** |

[4] G. Bognár, P. Kovács: ECG Classification with Deep Unfolding Variable Projection Network, Computing in Cardiology Conference, 2024

# Outline

# Application: Telecommunication

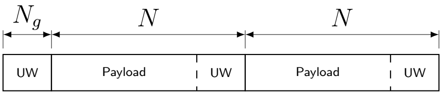



## Task

- Data estimation in UW-OFDM systems
- Multipath, additive white Gaussian noise channel (IEEE 802.11a)

# Unique Word OFDM[5]

**Transmit sequence in time domain:**




**System model:**

$$\mathbf{y} = \underbrace{\tilde{\mathbf{H}}\mathbf{G}}_{\mathbf{H}}\mathbf{d} + \mathbf{w}$$

- $\mathbf{y} \in \mathbb{C}^N$: received vector
- $\mathbf{d} \in \mathbb{S}^{N_d} \subset \mathbb{C}^{N_d}$: data symbol vector ($\mathbb{S}$: modulation alphabet)
- $\tilde{\mathbf{H}} \in \mathbb{C}^{N \times N}$: channel frequency response matrix (IEEE 802.11a)
- $\mathbf{G} \in \mathbb{C}^{N \times N_d}$: UW-OFDM generator matrix
- $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, N\sigma_n^2\mathbf{I})$

[5] M. Huemer, C. Hofbauer, and J. B. Huber: The Potential of Unique Words in OFDM, 15th International OFDM Workshop, 2010

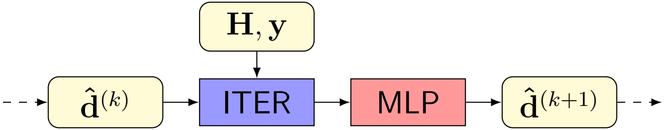
# Deep unfolding data estimation

## Deep unfolding data estimation

- Motivation: data-driven asymptotically optimal estimator
- Estimation: $\mathbf{y} = \mathbf{H}\mathbf{d} + \mathbf{w}, \quad (\mathbf{H}, \mathbf{y}) \longrightarrow \hat{\mathbf{d}}$
- Unfolding the gradient iteration:

$$\hat{\mathbf{d}}^{(k+1)} := \mathsf{MLP}\left(\hat{\mathbf{d}}^{(k)} + \delta_k \left(\mathbf{H}^T\mathbf{y} - \mathbf{H}^T\mathbf{H}\hat{\mathbf{d}}^{(k)}\right)\right)$$

## DetNet layer structure[6] (simplified)




[6] N. Samuel, T. Diskin, A. Wiesel: Learning to Detect, IEEE Trans. Sign. Proc., 2019

# NN-based data estimation for UW-OFDM[7]

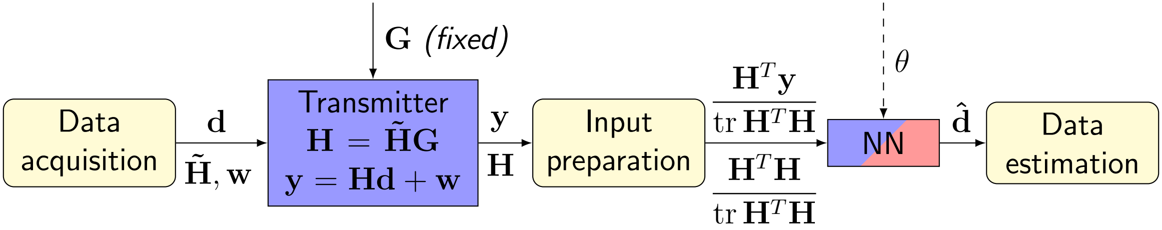



Figure: Simulation framework for NN-based data estimation

[7] S. Baumgartner, G. Bognár, O. Lang, and M. Huemer: Neural Network Approaches for Data Estimation in Unique Word OFDM Systems, IEEE Trans. Vehicular Technology, 2024

# NN-optimal UW-OFDM[8]

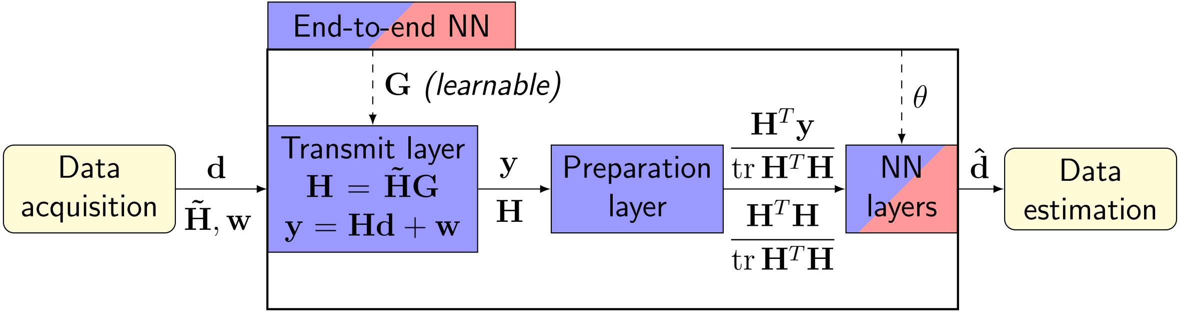
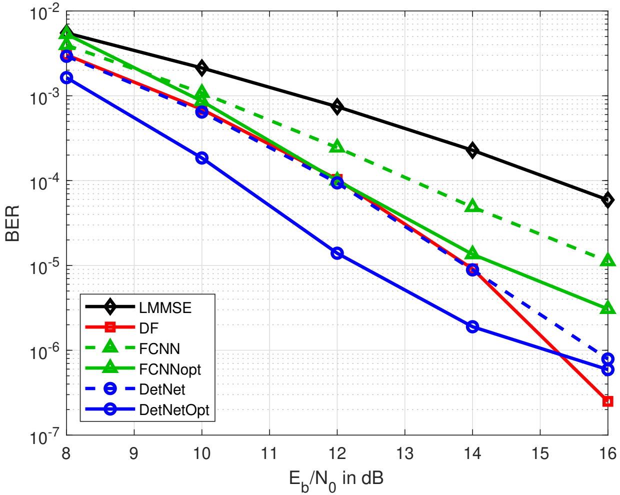



Figure: NN-optimal end-to-end framework for data estimation

[8] G. Bognár, S. Baumgartner, O. Lang, and M. Huemer: Neural Network Optimal UW-OFDM, Asilomar Conference, 2021

# NN-optimal data estimation results



# Outline

# Summary

- Model-based NN architectures based on deep unfolding
- Close to optimal data estimation
- Compact, low-dimensional representation learning
- Explainable representation, interpretable parameters

**Thank you for your attention!**