Learning Predictive Checklists from Continuous Medical Data

Yukti Makhija[1], Edward De Brouwer[2], Rahul G. Krishnan[3]

1. Indian Institute of Technology, Delhi 2. ESAT-STADIUS, KU Leuven, Belgium 3. University of Toronto and Vector Institute

Motivation

- **Predictive checklists** are widely used in the medical domain to assist complex decision-making and patient triaging.
- They are **discrete linear classifiers** which are **highly interpretable** and promote reliability. A patient is classified as positive if **M out of N** rules are satisfied.
- Currently, they are **designed manually** after the collection and assessment of medical evidence.
- With the widespread employment of EMRs, this process becomes **ineffective** and **tedious**.

Sepsis if 4+ items are checked

- ☐ Bilirubin_direct_sd > 0.346
- ☐ FiO2_sd > 0.0296
- ☐ FiO2_mean > -0.0178
- ☐ EtCO2_sd > 0.533
- ☐ FiO2_last > -0.017
- ☐ Alkalinephos_sd > 0.447
- ☐ pH_sd > 0.421
- ☐ TroponinI_sd > 1.082
- ☐ Bilirubin_direct_last > -0.486

Existing Approaches

- Recent work by **Zhang et al. (2021)**, formulated an Integer Linear Program for learning checklists from **categorical data**.
- Clinical data such as images or time series is not categorical by nature and limits the applicability of this approach.

Our Method

- We propose a novel approach which relaxes the previous boolean assumption and **learns checklists from continuous-values medical data**.
- Considering a dataset with **n** patients, **(X_i,y_i)** for **$i \in [n]$**, we let **X_i** be the continuous medical variables and **$y_i \in \{0, 1\}$** the label. The checklist prediction for a patient writes:

$$\boxed{\hat{y}_i = (w^T C(X_i) \geq M)}$$

Here, **w** are the learnable binary weights and **$C(X_i)$** are the binary concepts derived from **X_i** by learning thresholds **t_j**

- **Mixed Integer Program**

$$\min_{w,z,M,t} \; l^+ + \lambda l^- + \epsilon_N N + \epsilon_M M$$

s.t.

$$A_j C_{i,j} > X_{ij} - t_j \qquad X_{i,j} > t_j$$

$$A_j C_{i,j} < t_j - X_{i,j} \qquad X_{i,j} \leq t_j$$

$$B_i z_i \geq M - w^T C_i \qquad i \in I^+$$

$$B_i z_i \geq w^T C_i - M + 1 \qquad i \in I^-$$

$$l^+ = \sum_{i \in I^+} z_i \qquad l^- = \sum_{i \in I^-} z_i$$

$$N = \sum_{j=1}^{d} w_j \qquad w_j \in \{0, 1\}$$

$$C_i \in \{0, 1\}^d \qquad i \in [n]$$

$$M \leq N$$

Results

- Comparison with **baselines on the PhysioNet 2019 Early Sepsis Prediction**

Model	Accuracy	Precision	Recall	Specificity	N	M
Dummy Classifier	62.77	0	0	-	-	
MLP (non-interpretable)	**64.96 ± 2.59**	0.57 ± 0.05	0.48 ± 0.07	0.76 ± 0.06	-	-
Logistic Regression	62.56 ± 1.65	**0.62 ± 0.05**	0.14 ± 0.04	0.94 ± 0.03	-	-
Unit Weighting	58.28 ± 3.58	0.52 ± 0.09	0.44 ± 0.3	0.69 ± 0.25	9.6 ± 0.8	3.2 ± 1.16
SETS Checklist	56.48 ± 7.88	0.52 ± 0.11	**0.66 ± 0.30**	0.49 ± 0.32	10 ± 0	6 ± 0.63
ILP mean thresholds	62.99 ± 0.82	0.54 ± 0.087	0.12 ± 0.09	0.93 ± 0.32	4.4 ± 1.01	2.8 ± 0.75
MIP (*ours*)	63.69 ± 2.44	0.56 ± 0.05	**0.40 ± 0.08**	0.79 ± 0.06	8 ± 1.09	3.6 ± 0.8

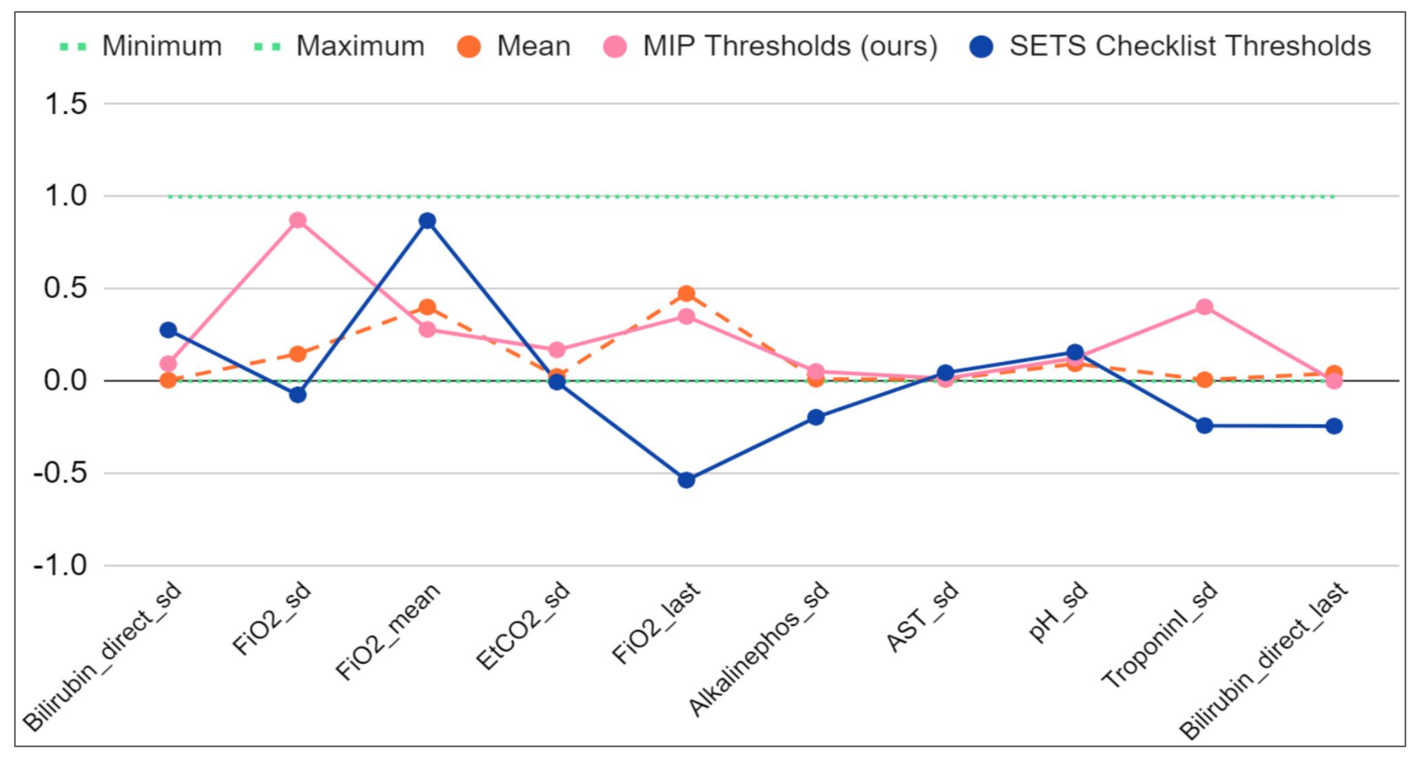
A **significant improvement in recall** was observed when the thresholds are learnt, as opposed to the original ILP, where mean binarization is applied.

- Comparison with **logistic regression**

	P@R=0.403	R@P=0.563
Logistic Regression	0.545 ± 0.052	**0.468 ± 0.23**
MIP (*ours*)	**0.563 ± 0.05**	0.403 ± 0.08

Checklists have a **lower capacity than logistic regression** due to their binary weights but are **more practical and interpretable**.

- Investigating the **learnt rules**



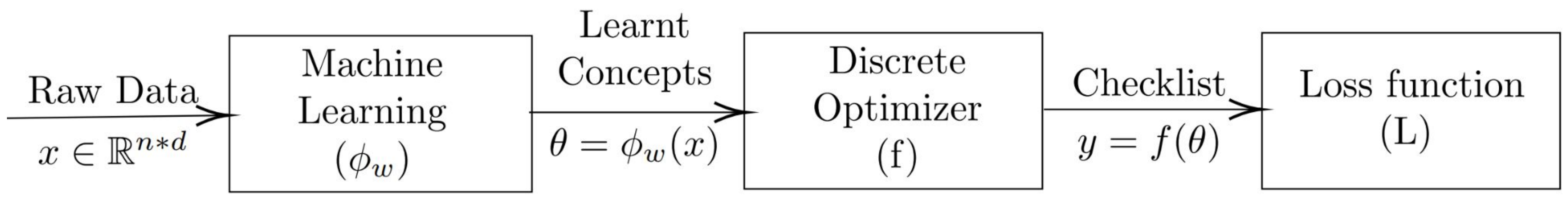
- **Balancing sensitivity and specificity**

By tuning λ, we can modify the cost function and attain the clinically-appropriate operating point.

Cost Function	Model	Accuracy	Precision	Recall	Specificity	N	M
Minimize 01 ($\lambda = 1$)	ILP mean thresholds	61.678	0.4727	0.2549	0.8314	4	2
	MIP (ours)	67.518	0.57143	0.5098	0.7733	9	4
Minimize FNR ($\lambda = 1/\lvert I^+\rvert$)	ILP mean thresholds	49.635	0.4189	0.9117	0.25	7	1
	MIP (ours)	38.321	0.3764	1	0.0174	7	2
Minimize FPR ($\lambda = \lvert I^-\rvert$)	ILP mean thresholds	63.139	1	0.0098	**1**	3	3
	MIP (ours)	62.773	0.5	0.0098	0.9942	8	3

Conclusion

- Our approach still lacks the ability to process more **complex data modalities**, such as histopathological images, surgical videos, and multimodal datasets.



- This can be achieved by building end-to-end hybrid architectures:
  - **ML layers** (for learning concepts from raw data)
  - **Combinatorial Optimization Layer** (for learning checklists)
- Recent papers have defined differentiation techniques for discrete optimizers that facilitate gradient approximation for backpropagation.

References

H. Zhang, Q. Morris, B. Ustun, and M. Ghassemi. Learning optimal predictive checklists. Advances in Neural Information Processing Systems, 34:1215–1229, 2021.