AAAI 2023
WASHINGTON DC

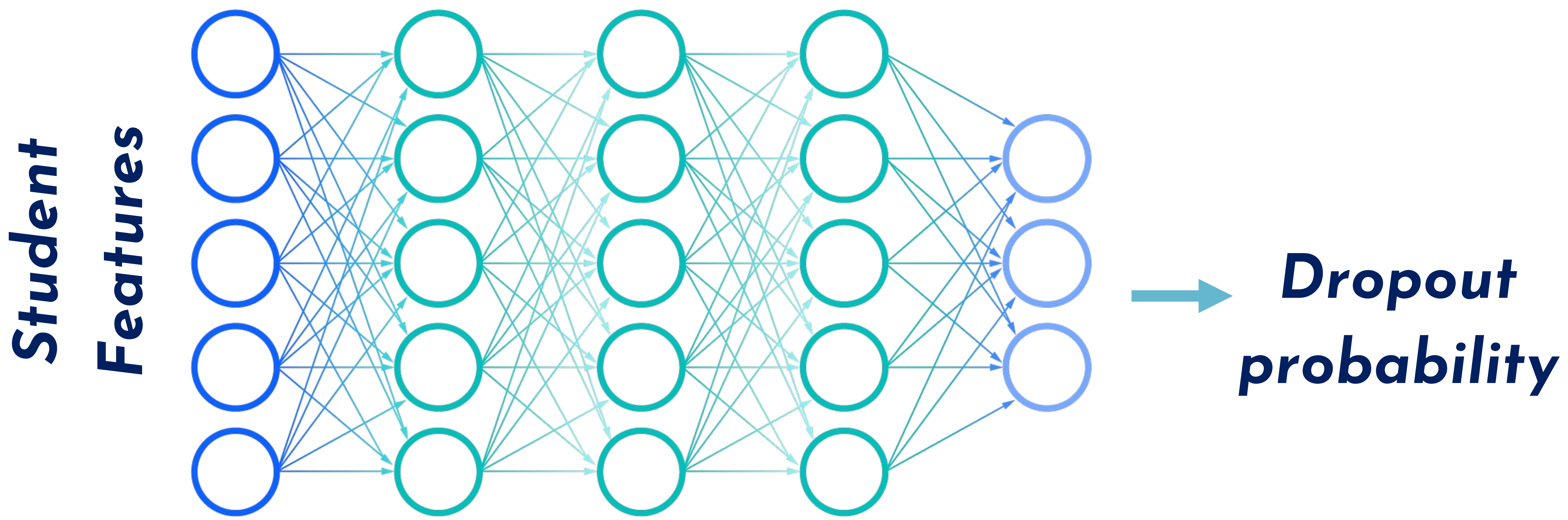
RIPPLE: Concept-Based Interpretation for Raw Time Series Models in Education

EPFL

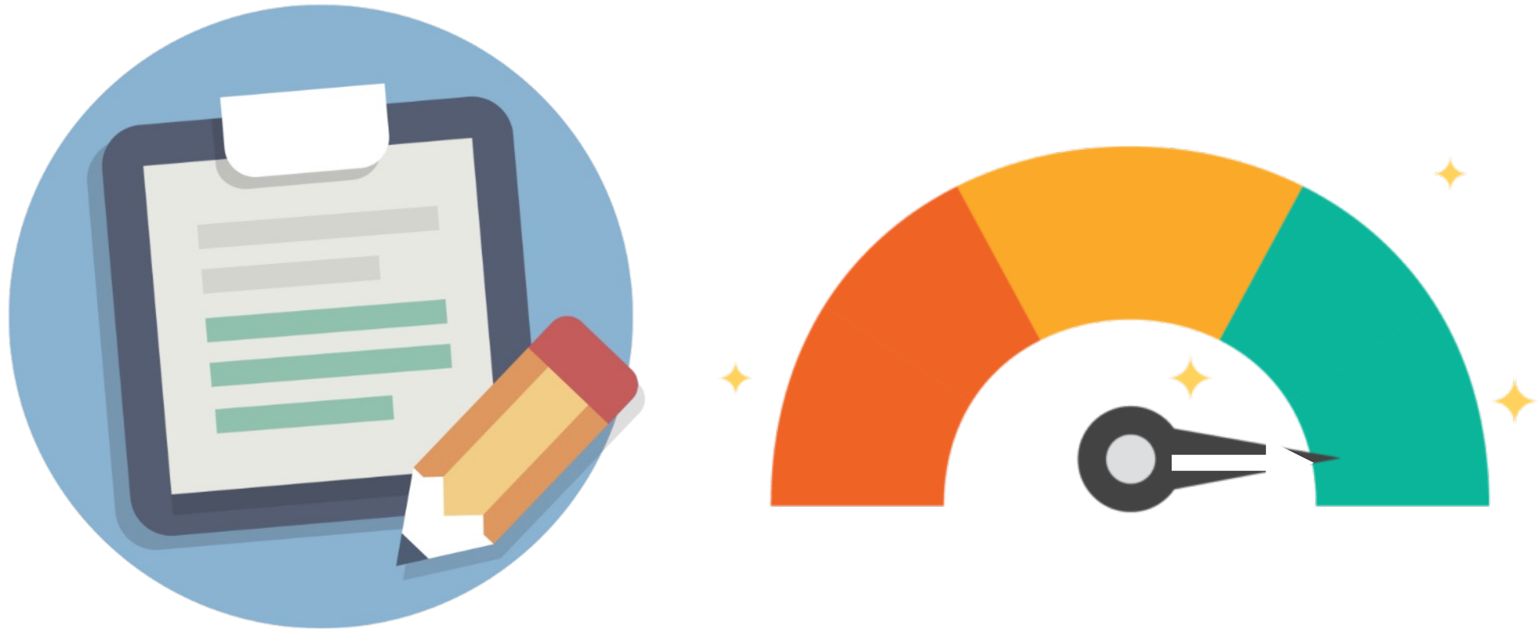
MOHAMAD ASADI, VINITRA SWAMY, JIBRIL FREJ, JULIEN VIGNOUD, MIRKO MARRAS, TANJA KÄSER

Deep Learning has been increasingly researched in digital learning environments

(MOOCs) Dropout Prediction

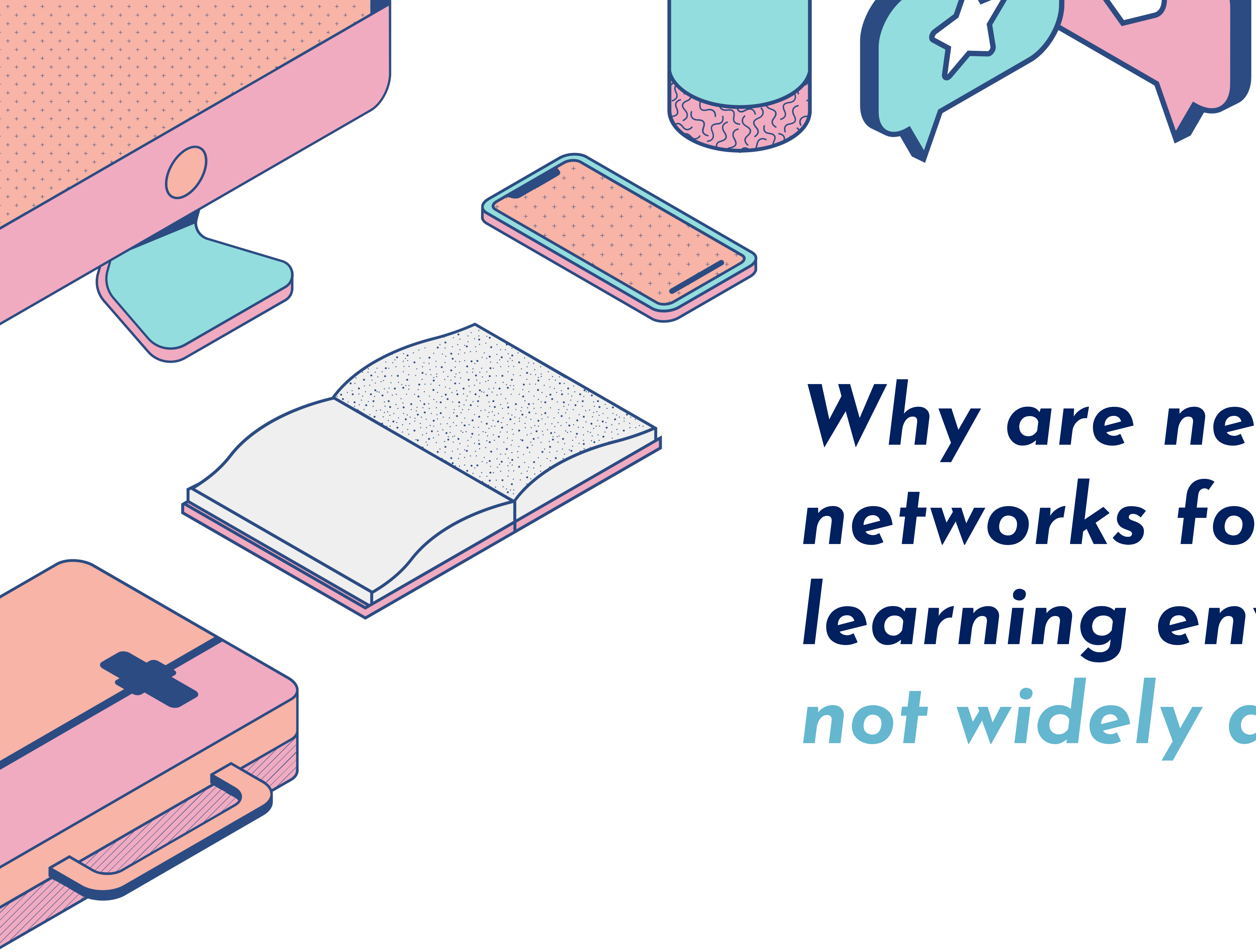


(LMS) Autograding, Plagiarism detection



(OELEs) Student Knowledge Tracing

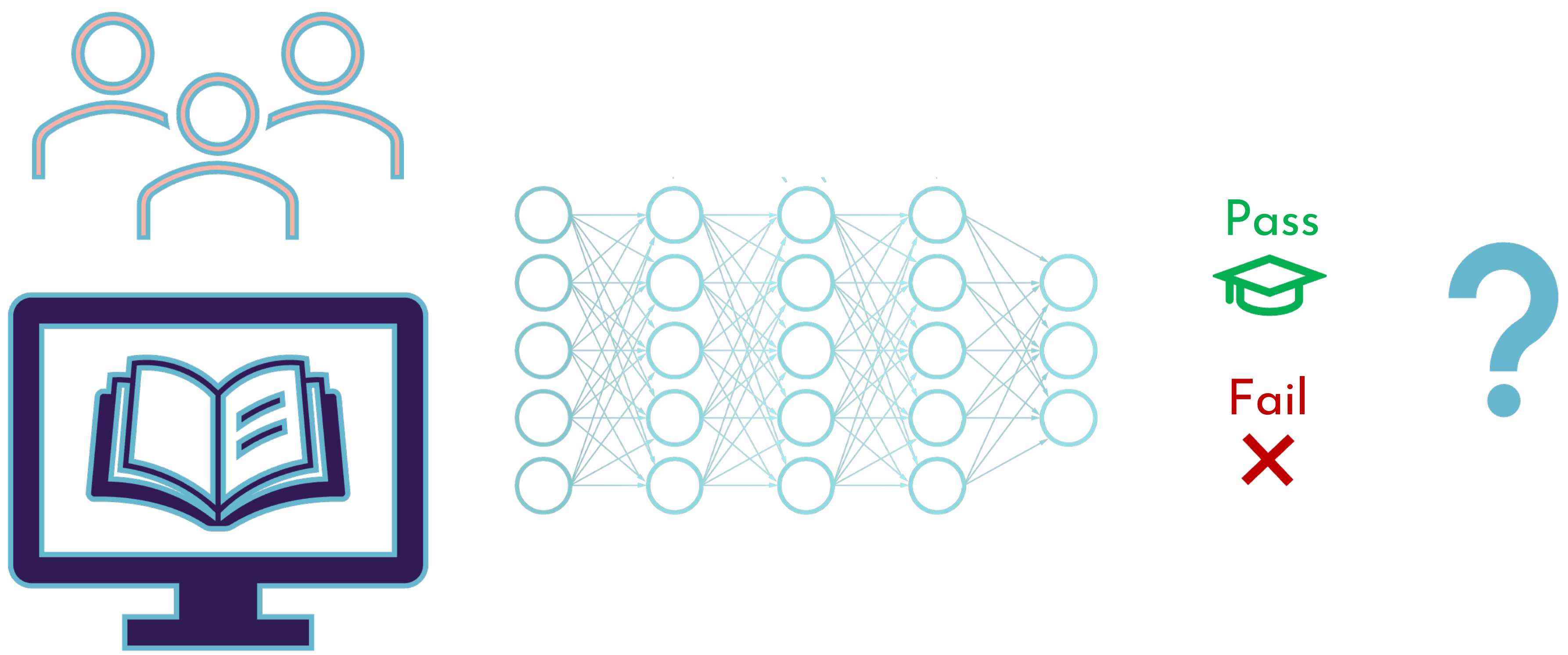
Imran et al. ICCAI 2019; Xing and Du, Journal of Edu Computing Research 2019; Piech et al. NeurIPS 2015.



Why are neural networks for digital learning environments not widely adopted?

The human-centric cost of neural networks

DEEP LEARNING IN EDUCATION



Problem: extracting hand-crafted features is hard

Identifying "why" is important for effective, personalized interventions

Problem: deep learning trades transparency for accuracy

Objectives

MOTIVATION

The objective of this work is therefore
to develop models for early success prediction that

(1) beat SoTA baselines with raw time series clickstreams
(2) provide interpretable insights for personalized intervention

Dataset: 23 MOOCs, >100k enrollments, millions of interactions
6 learner-centric behavioral dimensions
(Regularity, Effort, Consistency, Proactivity, Control, Assessment)

RIPPLE

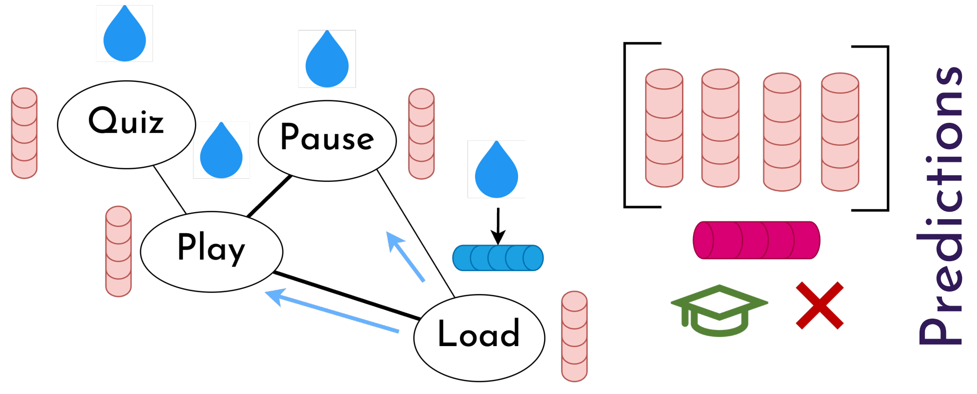
METHODOLOGY

Data Collection

Raw Time Series

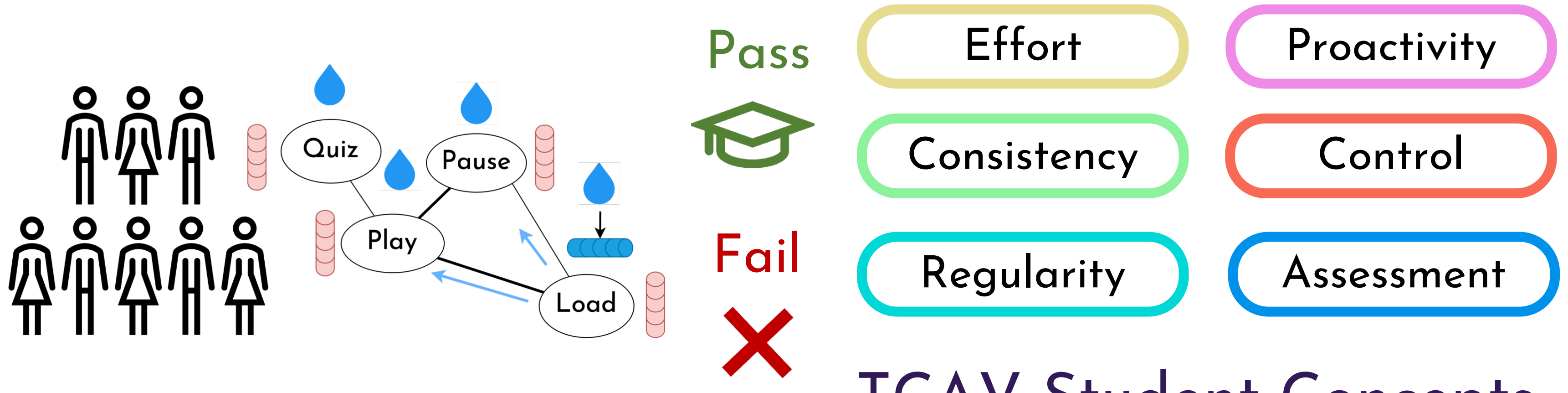
action	time	course
video.pause	13:15:35	DSP 1
video.play	13:16:12	DSP 1

Time-Series Classification



RAINDROP

Concept Interpretation



TCAV Student Concepts

Data Collection

METHODOLOGY

Digital Signal Processing

Villes Africaines

Structures

Household Water Treatment

. . .

Venture

Geomatique

Languages: English / French

Student Level: Prop / BSc / MSc

 # Students: 452 - 19k

 # Weeks: 5 - 11

Pass Ratio: 3% - 82%

 # Quizzes: 3 - 38

Raw Time Series

action	time	course
video.pause	13:15:35	DSP 1
video.play	13:16:12	DSP 1

RIPPLE

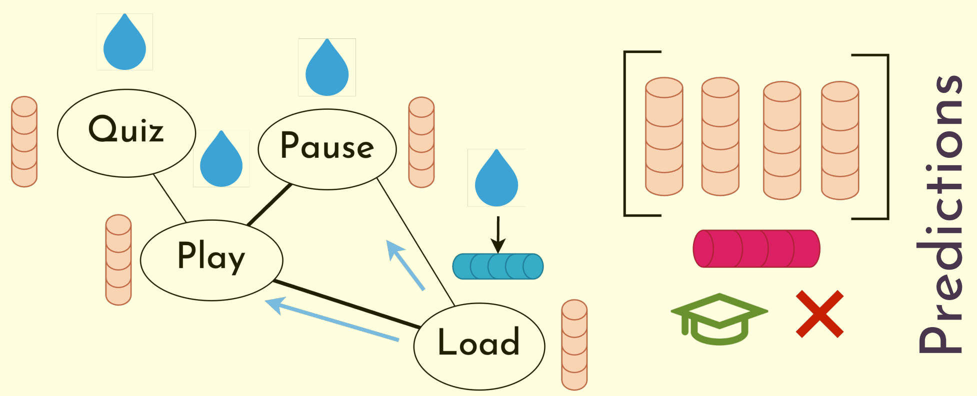
METHODOLOGY

Data Collection

Raw Time Series

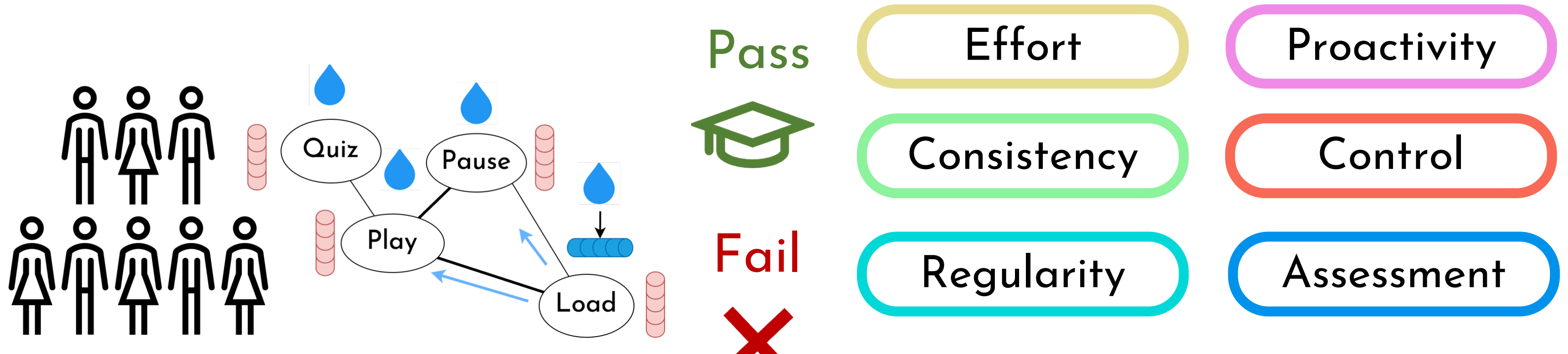
action	time	course
video.pause	13:15:35	DSP 1
video.play	13:16:12	DSP 1

Time-Series Classification



RAINDROP

Concept Interpretation



TCAV Student Concepts

Time Series Classification

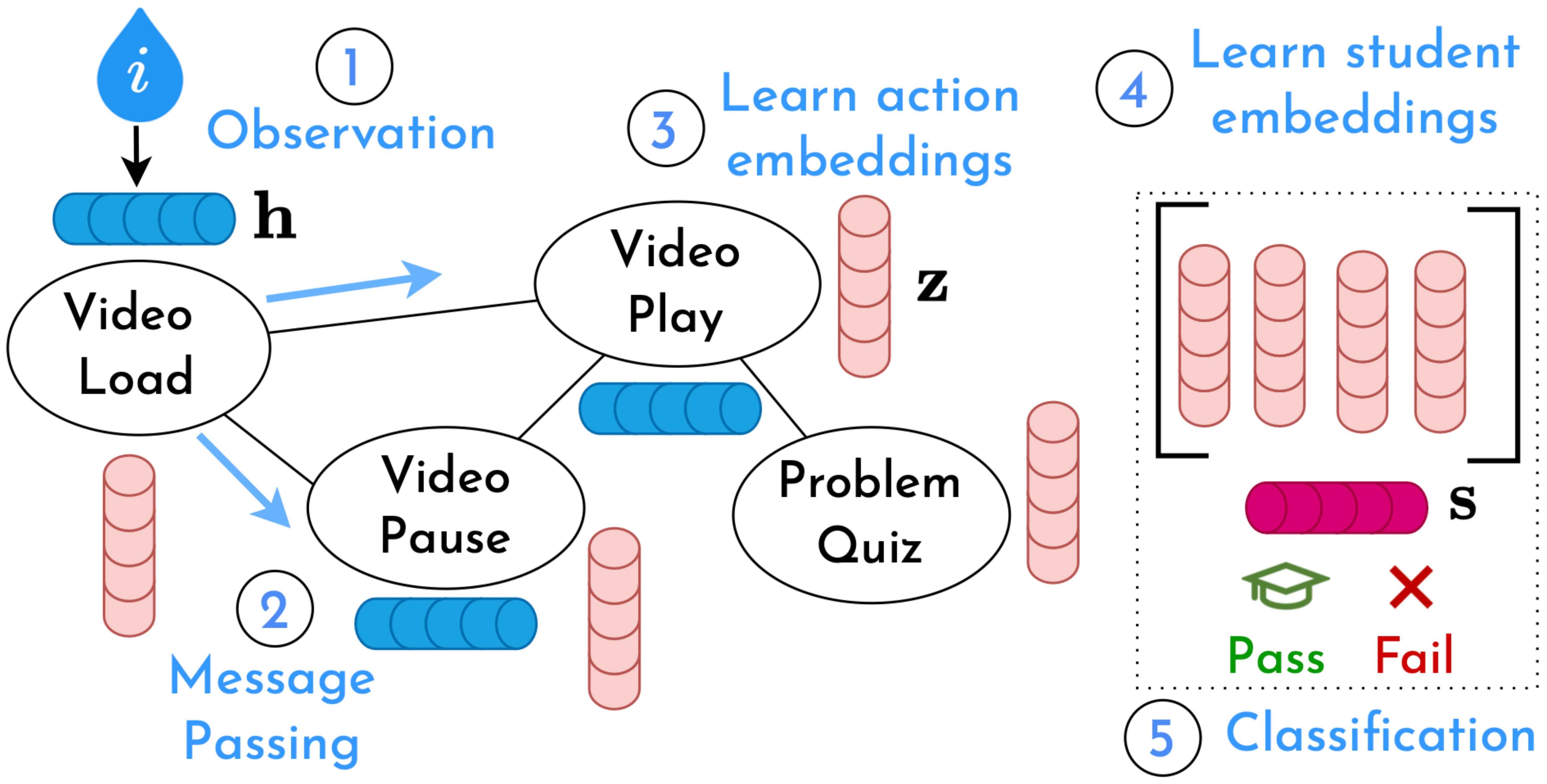
METHODOLOGY

task:
early prediction of student success (pass/fail)

challenge:
raw time series

RAINDROP

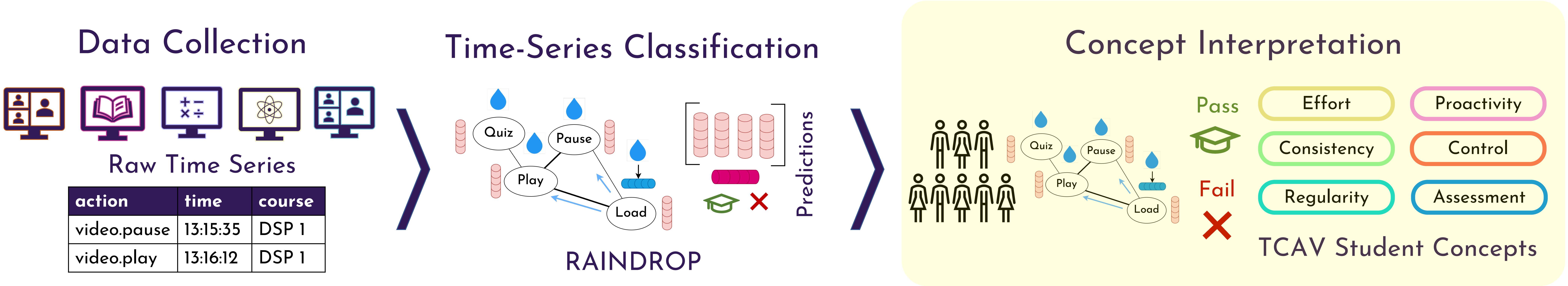
Graph NN with message passing



Zhang et al. ICLR 2021

RIPPLE

METHODOLOGY



action	time	course
video.pause	13:15:35	DSP 1
video.play	13:16:12	DSP 1

RIPPLE: Raindrop InterPretability PipeLine for Education

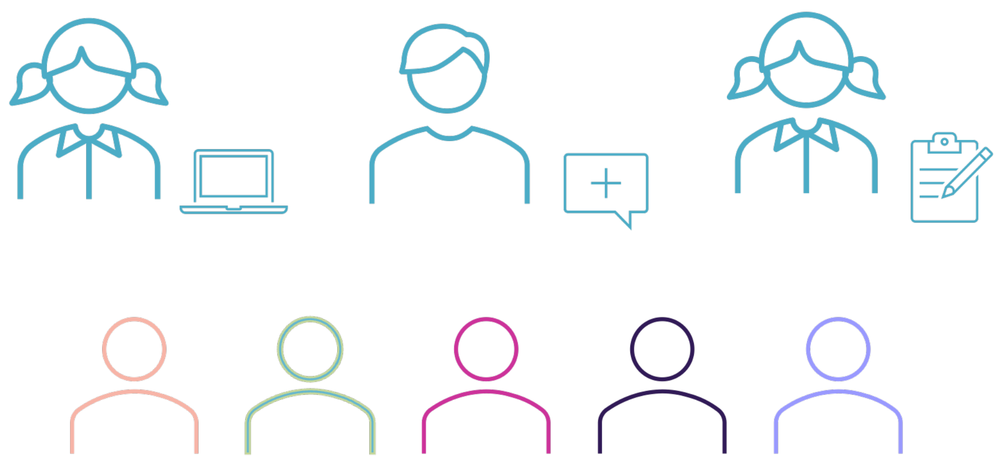
TCAV

METHODOLOGY

Can we obtain interpretability on raw multivariate time series?

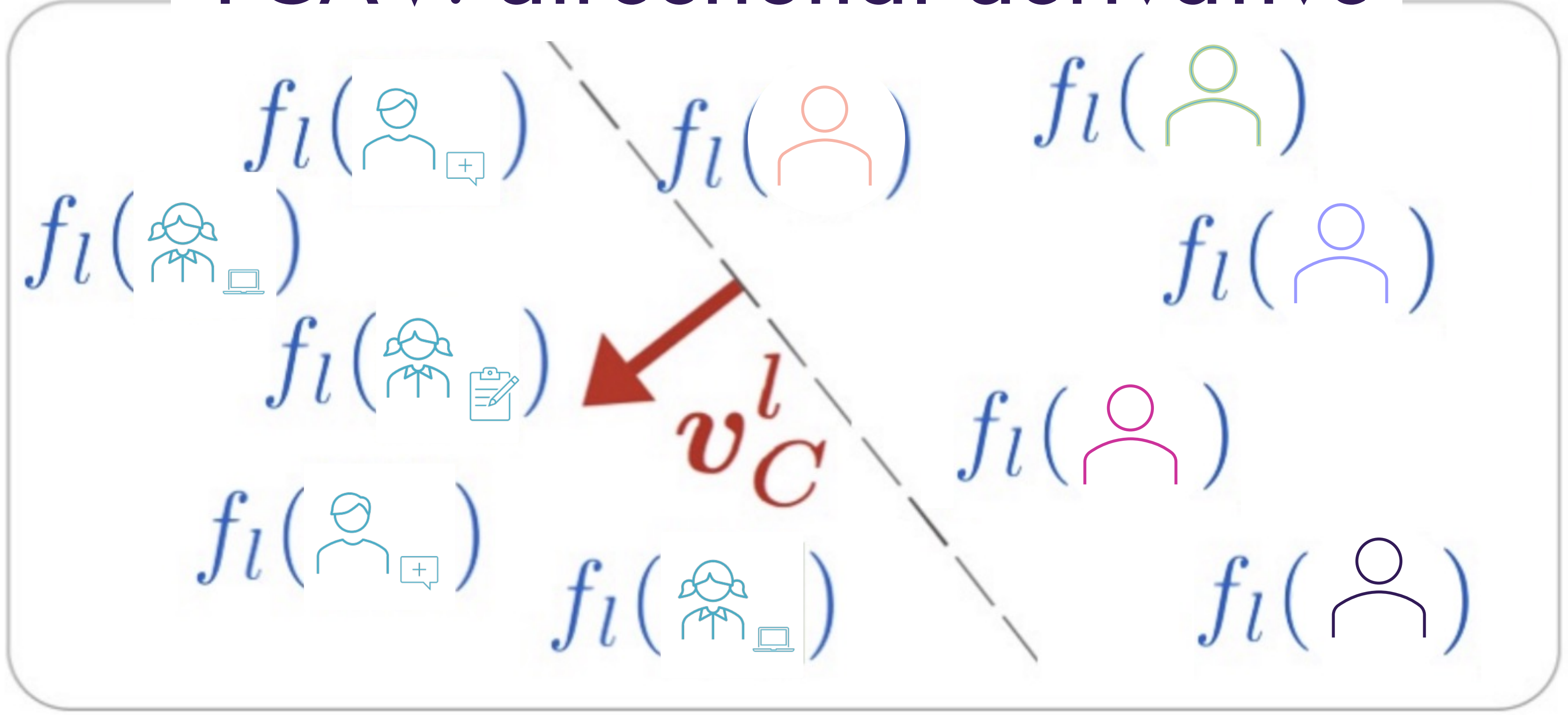
Concept Activation Vectors

concept: high effort



random students

TCAV: directional derivative



Advantages: global and local scale, user-specified example-based concepts, directly using model activations (accuracy)

Kim et al. ICML 2018

Concepts

METHODOLOGY

Dimensions	Measures	Patterns
Effort	Total time online Total video clicks	Higher intensity Lower intensity
Consistency	Mean session duration Relative time online Relative video clicks	Uniform First half Second half
Regularity	Periodicity of week day Periodicity of week hour Periodicity of day hour	Higher peaks Lower peaks
Proactivity	Content anticipation Delay in lecture view	Anticipated Delayed
Control	Fract. time spent (video) Pause action frequency Average change rate	Higher intensity Lower intensity
Assessment	Competency strength Student shape	Higher intensity Lower intensity

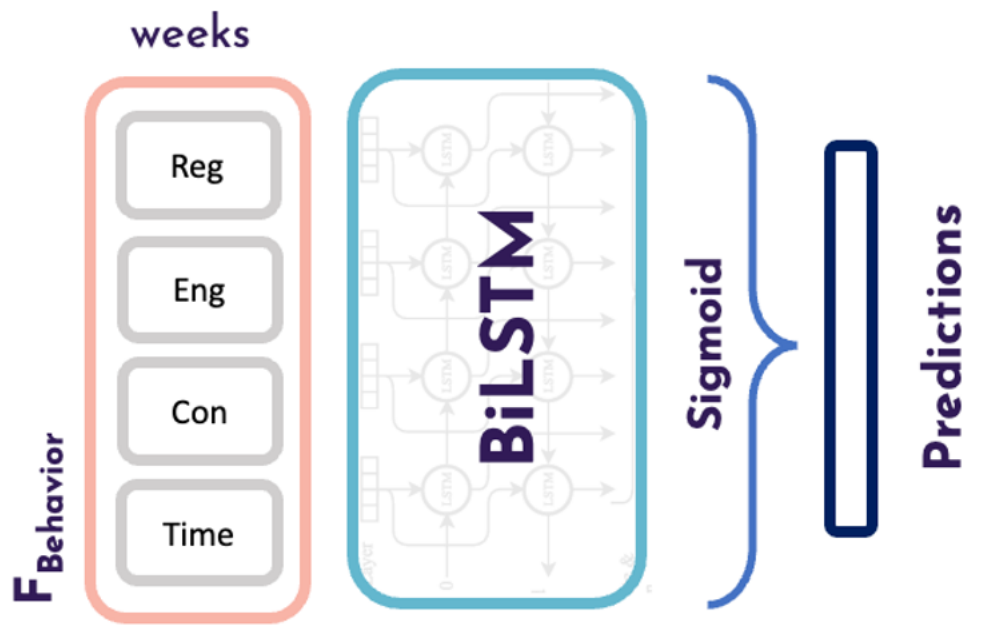
Table 1: Learning dimensions from Mejia-Domenzain et al. (2022) used as concepts for interpretability in our study.

RQ1: Performance

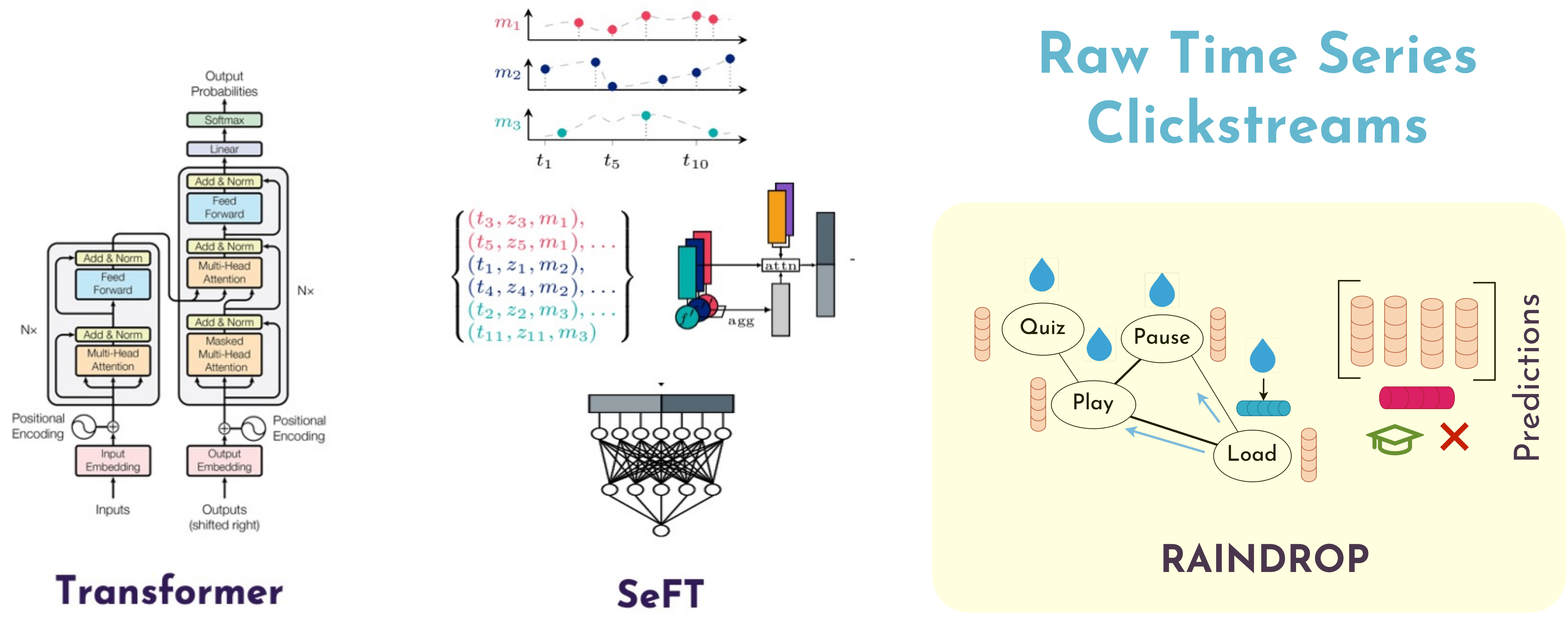
RESULTS

Can we use **raw time series as input** and achieve comparable performance to hand-crafted features?

Hand-crafted Features



Raw Time Series Clickstreams



Evaluation metric: Balanced Accuracy (BAC)

RQ1: Performance

RESULTS

	Early 40%							Early 60%						
	Raindrop	SeFT		TF		BiLSTM		Raindrop	SeFT		TF		BiLSTM	
	BAC	BAC	R	BAC	R	BAC	R	BAC	BAC	R	BAC	R	BAC	R
CPP*	**0.57**	0.46	2/2	0.54	2/2	0.56	2/2	**0.55**	0.53	1/2	0.52	2/2	**0.55**	2/2
DSP*	**0.81**	0.72	5/5	0.59	5/5	0.80	4/5	**0.91**	0.82	5/5	0.62	5/5	**0.91**	4/5
ProgFun*	**0.76**	0.63	2/2	0.53	2/2	0.63	2/2	**0.75**	0.69	2/2	0.56	2/2	0.67	2/2
AnNum	**0.66**	0.51	3/3	0.51	3/3	0.62	3/3	0.55	0.57	3/3	0.51	3/3	**0.69**	1/3
Geomatique*	0.50	0.45	1/1	**0.56**	0/1	0.47	1/1	**0.77**	0.55	1/1	0.45	1/1	0.76	1/1
HWTS	0.61	0.55	2/2	0.55	1/2	**0.71**	1/2	0.62	0.62	1/2	0.56	2/2	**0.73**	0/2
Micro	0.74	0.70	2/4	0.58	4/4	**0.81**	1/4	**0.78**	0.76	2/4	0.63	2/4	**0.78**	2/4
Ventures*	**0.77**	0.64	1/1	0.64	1/1	0.50	1/1	**0.88**	0.73	1/1	0.56	1/1	0.60	1/1
VA*	**0.88**	0.75	3/3	0.63	3/3	0.80	3/3	**0.90**	0.72	3/3	0.68	3/3	0.83	3/3

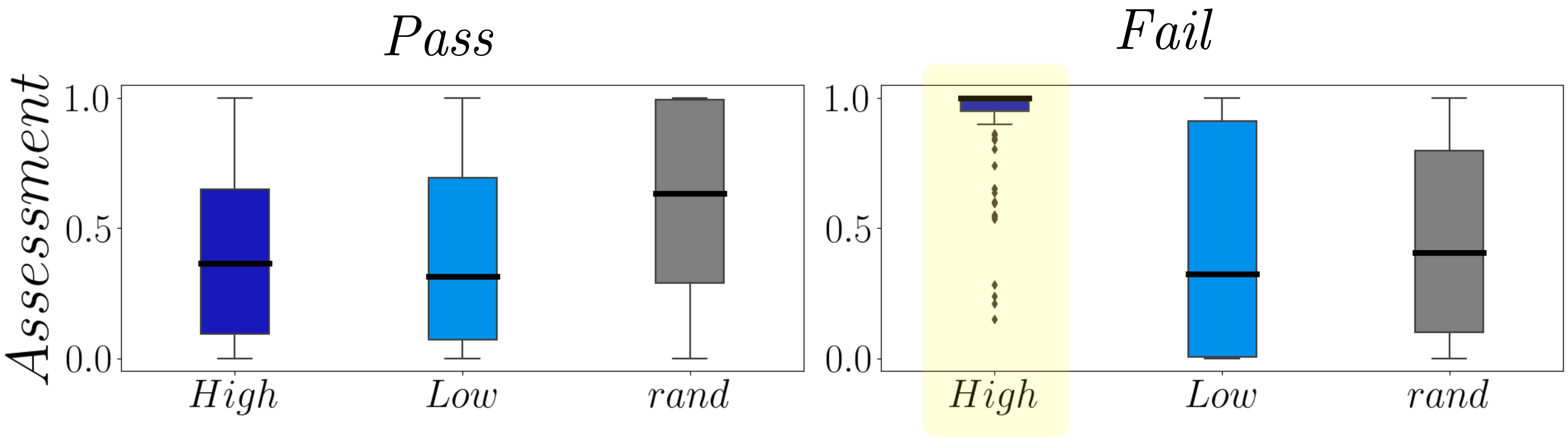
The best model for each course type and early prediction level is marked in **bold**. Course types where `Raindrop` had comparable or better performance to BiLSTM on both early prediction levels are marked in (*).

RAINDROP ≥ hand-crafted features: 18 out of 23 courses
RAINDROP ≥ SoTA time-series (SeFT, TF): 21 out of 23 courses

RQ2: Interpretation

RESULTS

TCAV plots: **random** vector **vs. learner-centric cluster** vectors



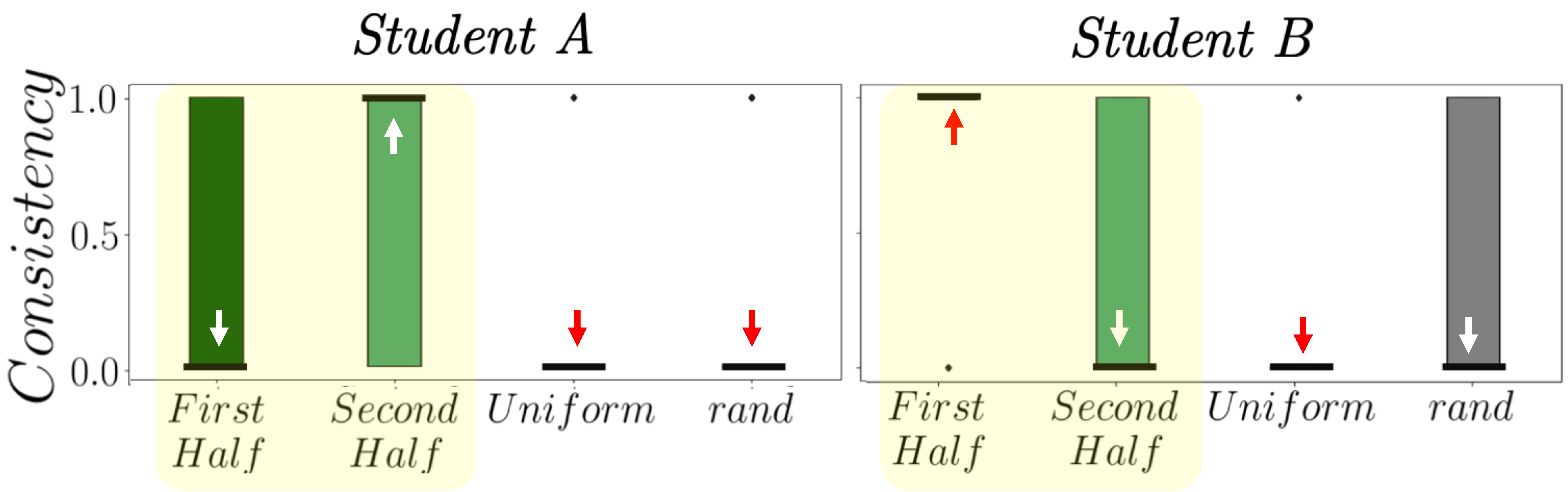
High assesment is important for predicting failing students

RQ2: Interpretation

RESULTS

Comparing students with differing profiles

For student A, *second half consistency* is important



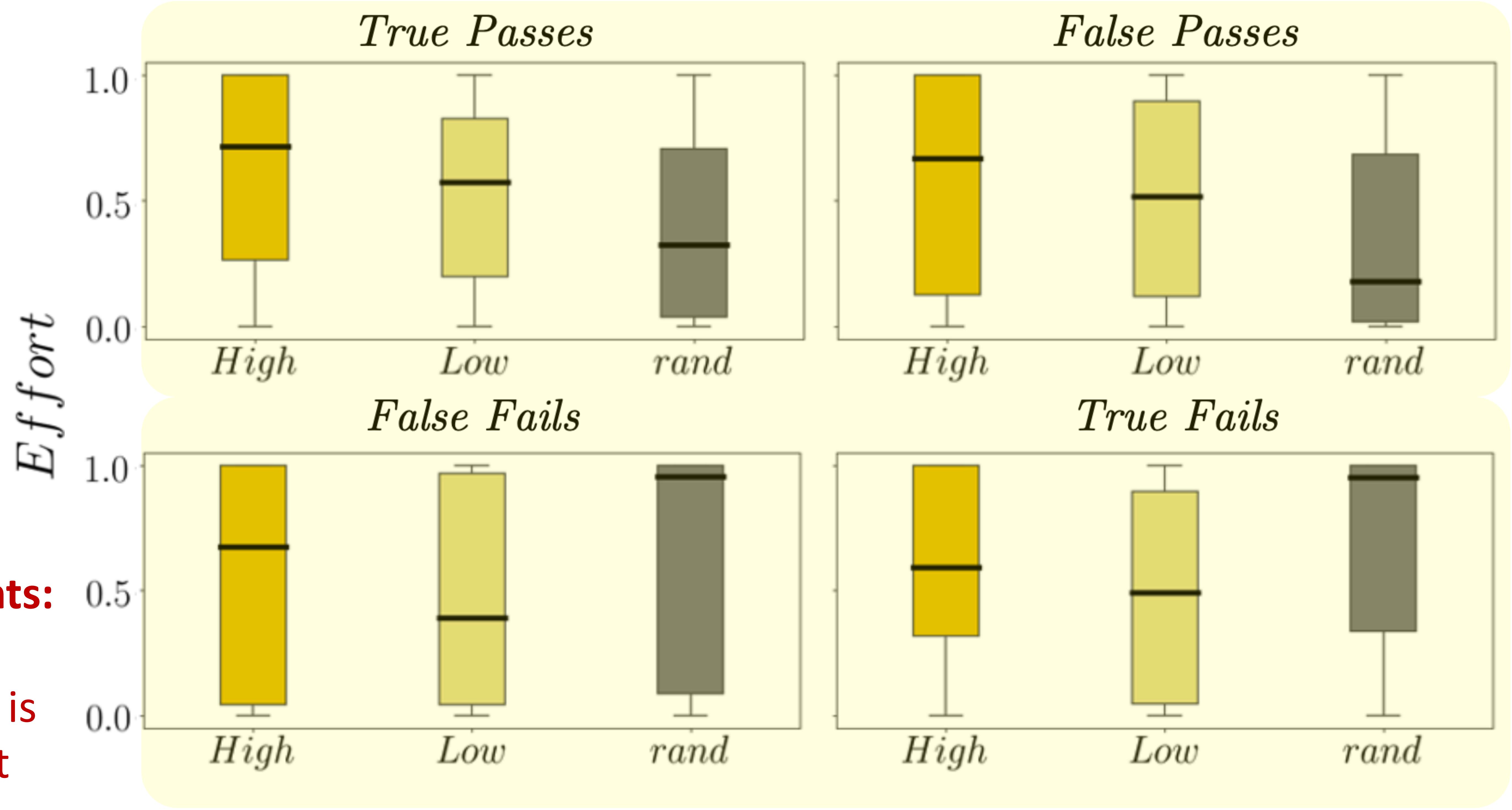
rand is low, so *consistency* is important

For student B, *first half consistency* is important

RQ2: Interpretation

RESULTS

Confusion Matrix Analysis: when does the model make mistakes?

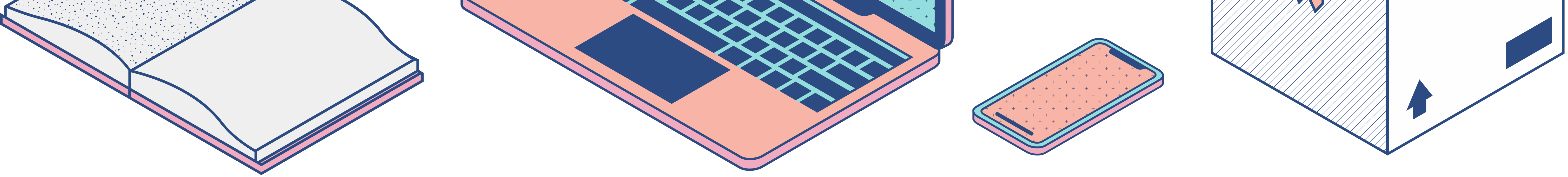


Passing students: model uses *high effort* as a proxy for *pass,* but sometimes gets it wrong

Failing students: high *rand* means *effort* is insignificant

next steps?

- automated concepts (d-tcav)
- generalization



Main Takeaway

RIPPLE: CONCEPT-BASED INTERPRETATION FOR RAW TIME SERIES MODELS IN EDUCATION

epfl-ml4ed/ripple

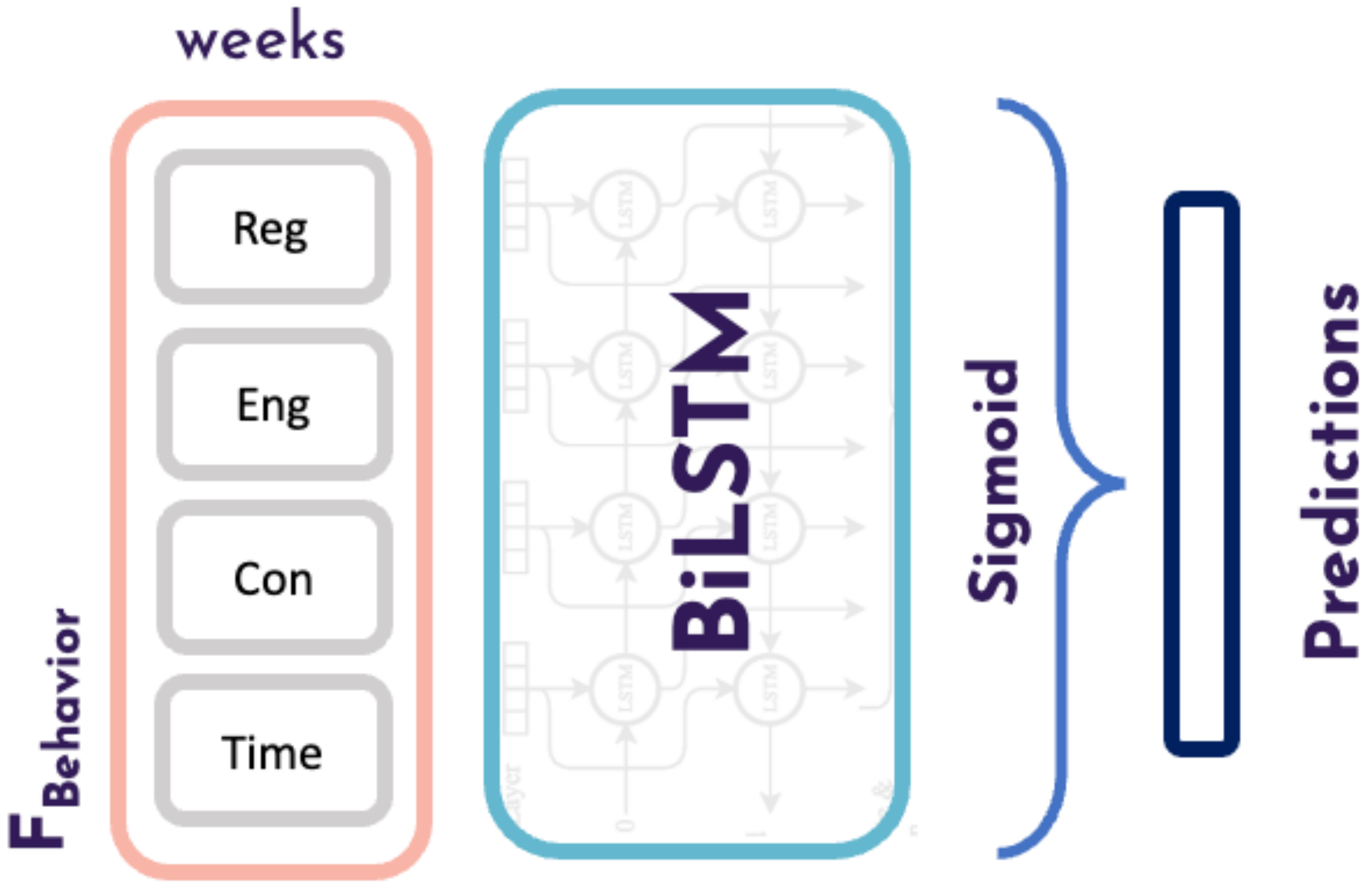
Transparency does not have to come at the cost of **accuracy** or **ease-of-use**

Vinitra Swamy
vinitra.swamy@epfl.ch

Thank you!

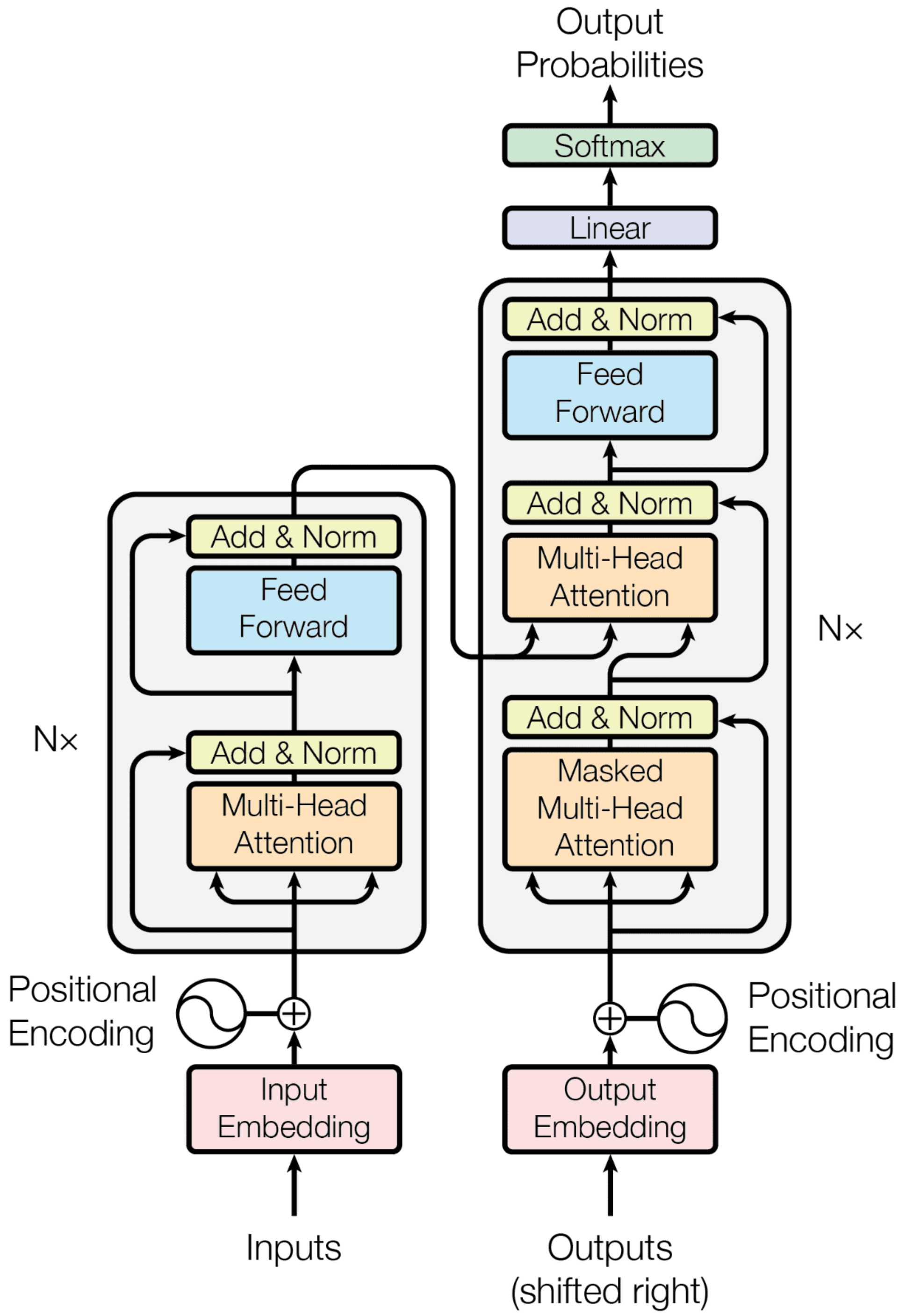
Baselines

METHODOLOGY

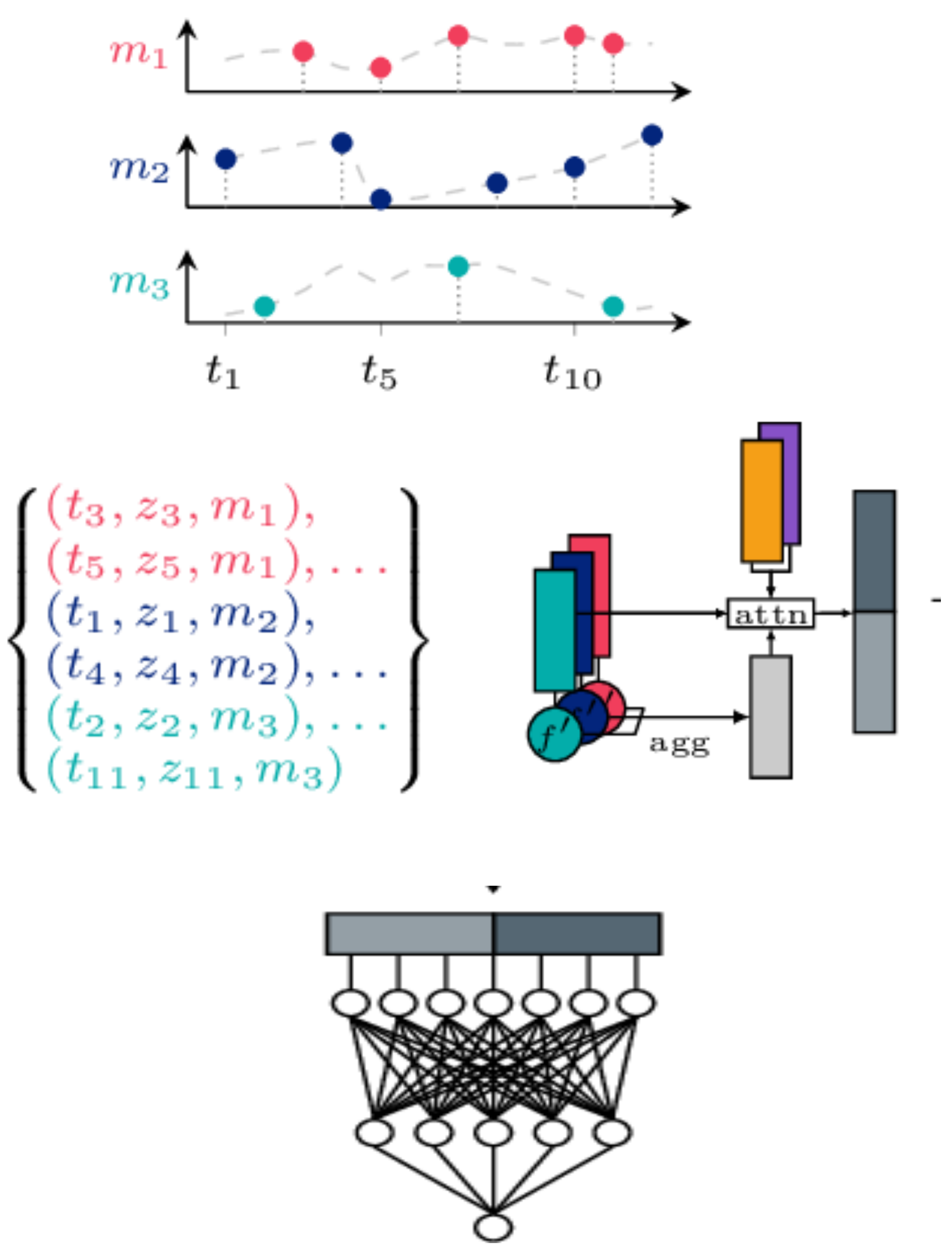


Hand-crafted features
(Swamy et al.)

All features are derived from previous work.



Transformer



SeFT
(Set Functions for Time Series)

Swamy et al. EDM / L@S 2022, Vaswani et al. NeurIPS 2017, Horn et al. ICLR 2020