LEARNING @ SCALE 2022
CORNELL TECH, NYC

Meta Transfer Learning for Early Success Prediction in MOOCs

VINITRA SWAMY, MIRKO MARRAS, TANJA KÄSER

EPFL

Vinitra Swamy

University of Cagliari

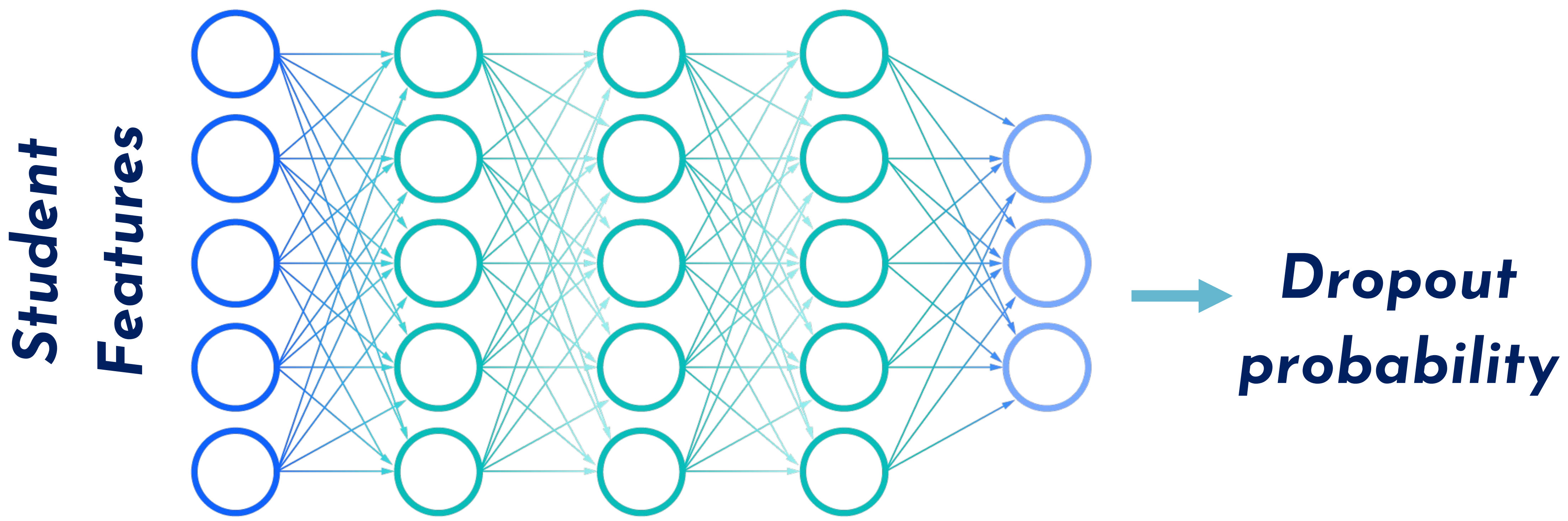
Mirko Marras

EPFL

Tanja Käser

Deep Learning has been increasingly researched in digital learning environments

(MOOCs) Dropout Prediction



(LMS) Autograding, Plagiarism detection

(OELEs) Student Knowledge Tracing

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

Real-world challenges

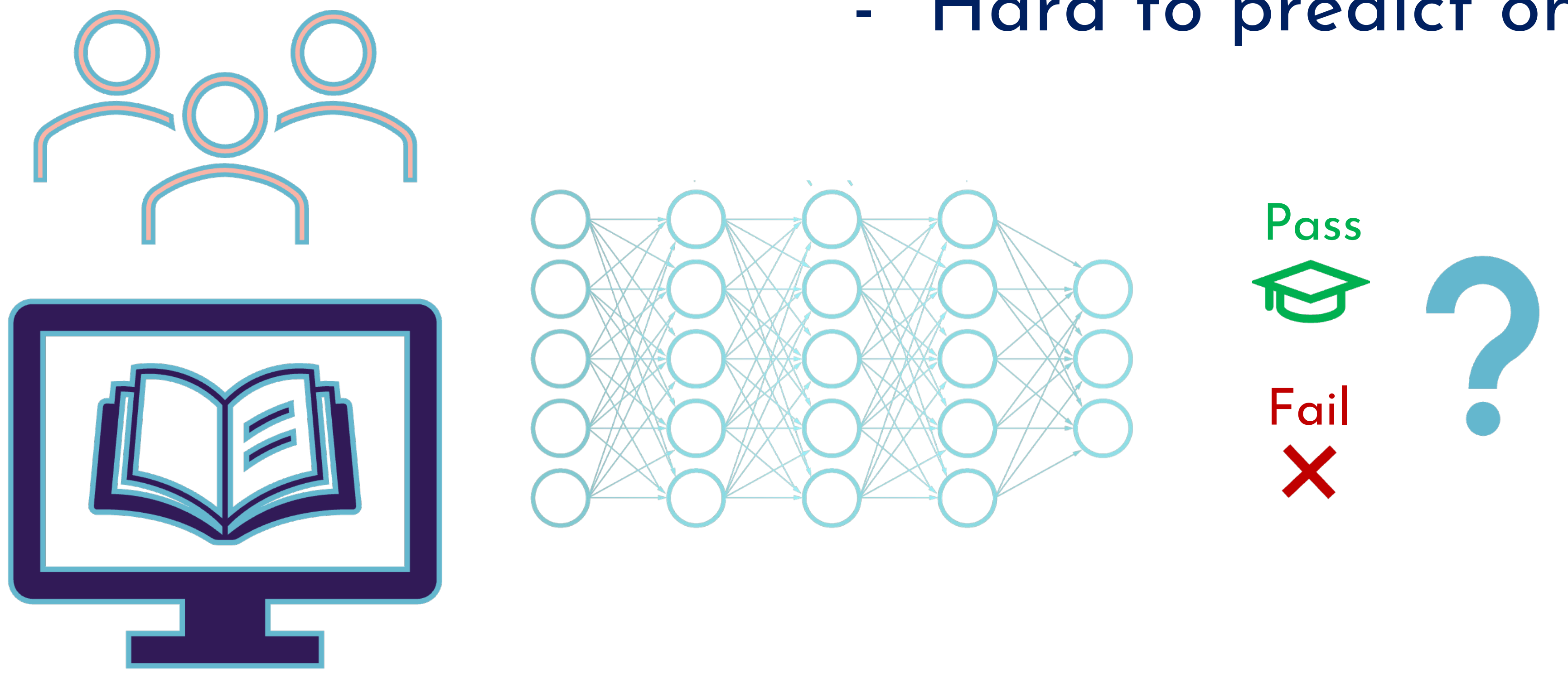
DEEP LEARNING IN EDUCATION

Ongoing or first-time courses

- Hard to make prediction for first students
- Zero-shot learning solutions are not common

Small dataset sizes

- Classrooms of 20, 30, 50, 100 students
- Hard to predict on without overfitting



In many settings, it is impossible to train an ideal model from scratch for early student success prediction.

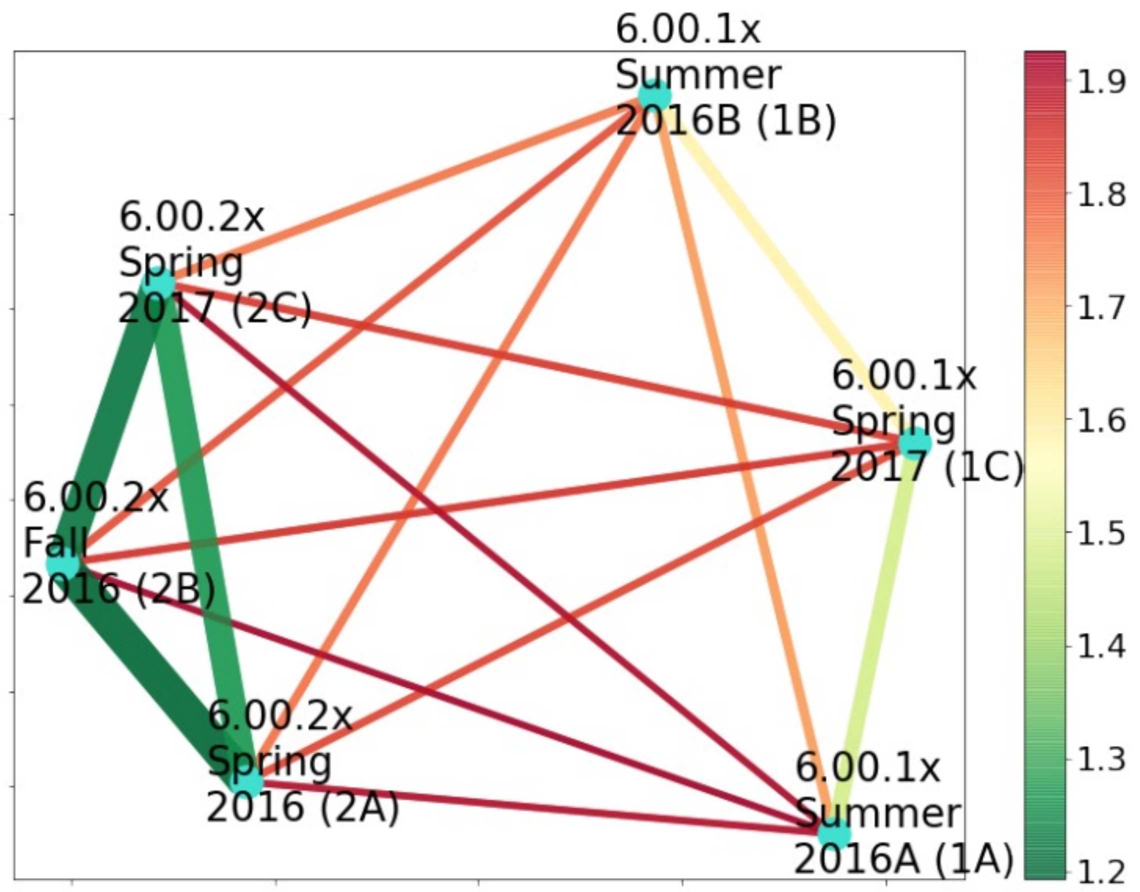
Previous Work

MOTIVATION

Train on past iterations[1,2]

- Deep Knowledge tracing
- Dropout Prediction
- Performance Prediction

Transfer across few courses[3]



[1] Wang, Lisa, et al. "Deep knowledge tracing on programming exercises." *Proceedings of the fourth (2017) ACM conference on learning@ scale*. 2017.
[2] David John Lemay and Tenzin Doleck. Grade prediction of weekly assignments in MOOCs: mining video-viewing behavior. Education and Information Technologies, 25(2):1333–1342, 2020.
[3] Ding, Mucong, et al. "Transfer learning using representation learning in massive open online courses." *Proceedings of the 9th international conference on learning analytics & knowledge*. 2019.

Objectives

MOTIVATION

The objective of this paper is therefore
to use **meta transfer learning** to develop **early success prediction**
models that can be **applied across diverse domains**

Large Model Diverse Dataset

Dataset: 26 MOOCs, 145,000 enrollments, millions of interactions

Meta Learning + Transfer Learning = Meta Transfer Learning

Research Questions

MOTIVATION

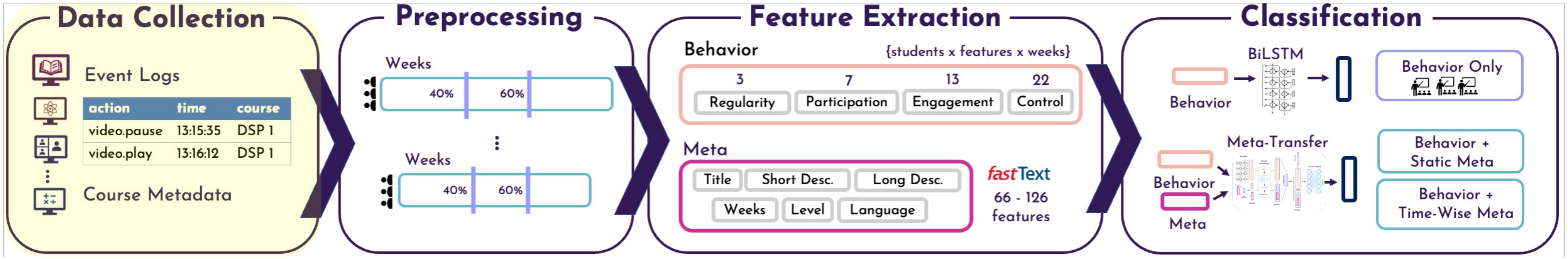
1) Can student behavior transfer across iterations of the same course and across different courses?

2) Is a meta learning model trained on a combination of behavior and course metadata information more transferable?

3) Can fine-tuning a combined model on past iterations of an unseen course lead to better transferable models?

Pipeline

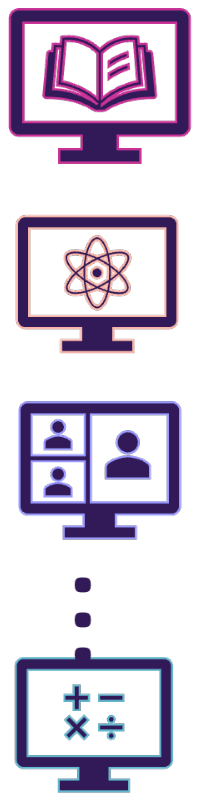
METHODOLOGY



Pipeline

METHODOLOGY

Data Collection



Event Logs

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

Course Metadata

Easy-to-Predict: Filter out easy-to-predict failing students, as there is no need for a complex model if a LogReg is sufficient!

Data Collection

METHODOLOGY

Digital Signal Processing

Villes Africaines

Structures

Functional Programming

. . .

Venture

Geomatique

--- only 1 iteration ---

Languages: English / French

Student Level: Prop / BSc / MSc

 # Students: 95 - 19k

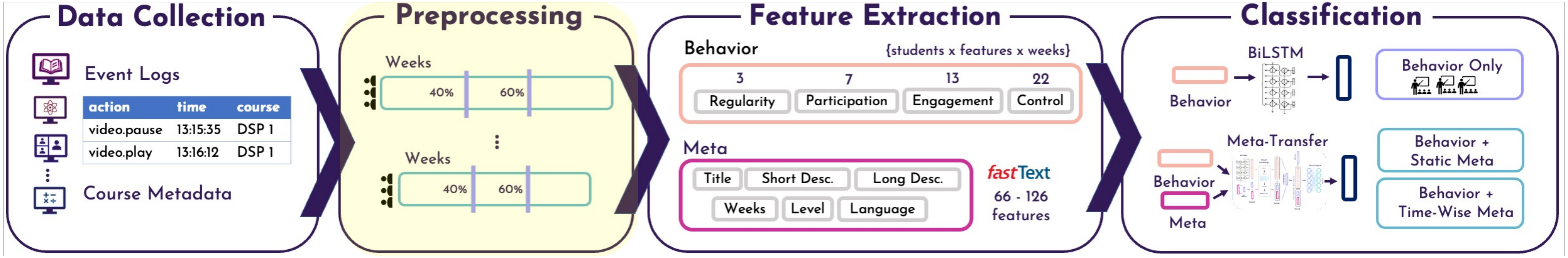
 # Weeks: 5 - 15

Pass Ratio: 1% - 65%

 # Quizzes: 4 - 38

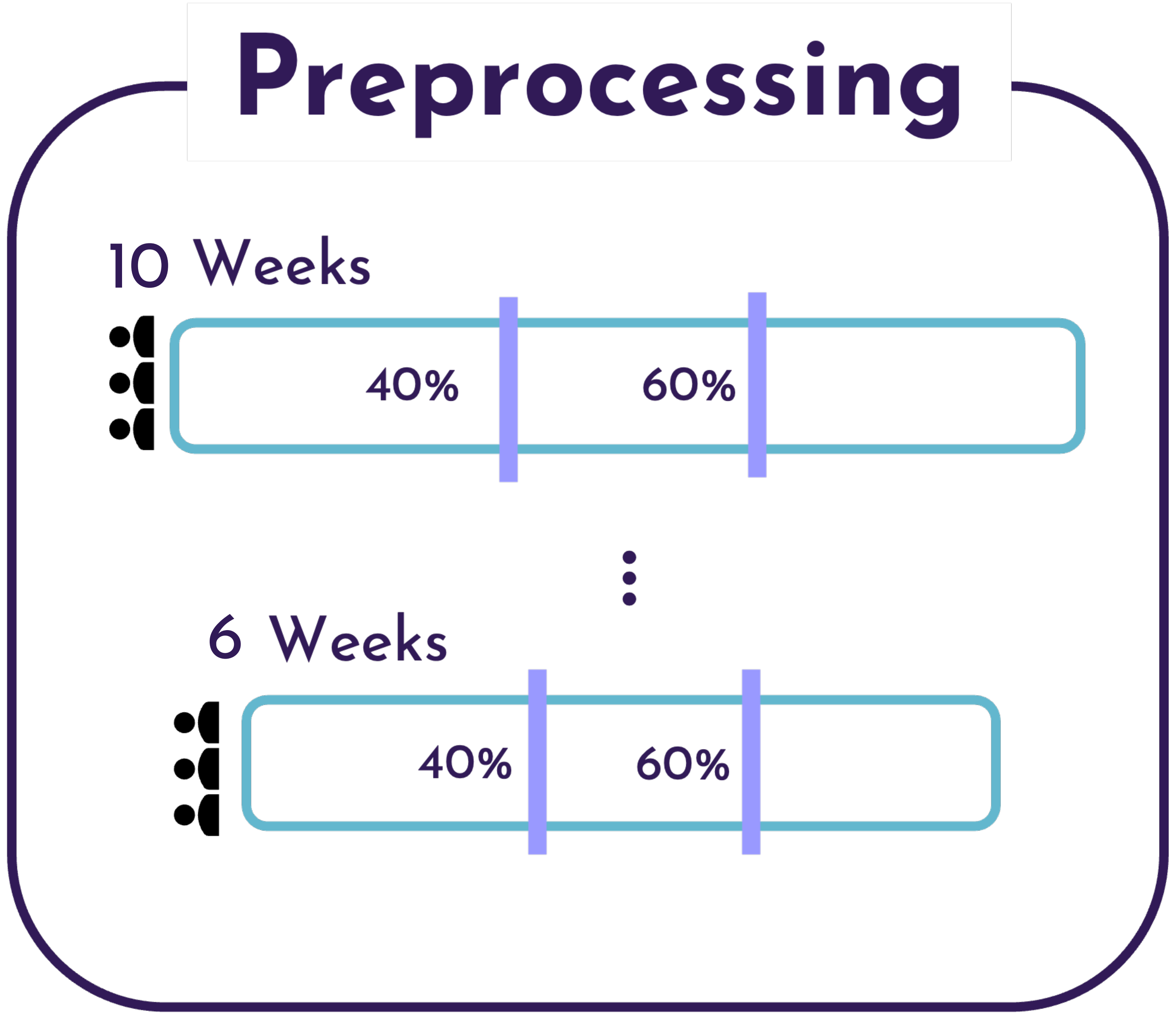
Pipeline

METHODOLOGY



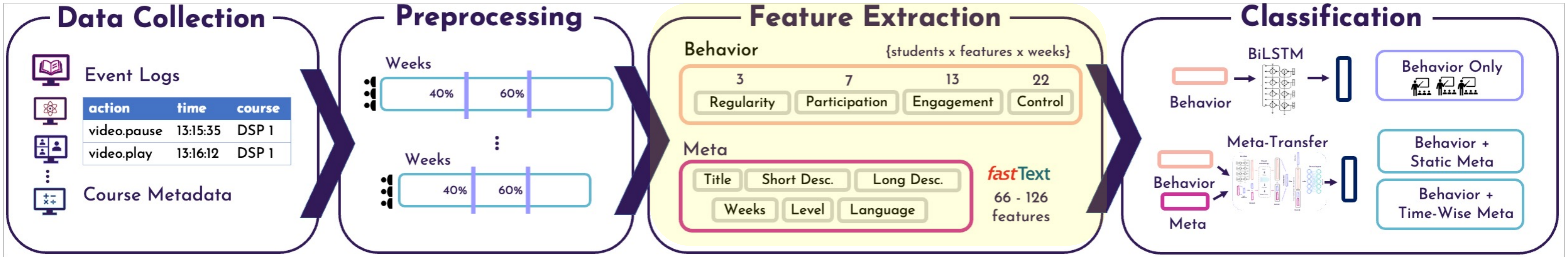
Pipeline

METHODOLOGY



Pipeline

METHODOLOGY



Pipeline

METHODOLOGY

Feature Extraction

Behavior {students x features x weeks}

Regularity	Participation	Engagement	Control
3	7	13	22

Meta

Title, Short Desc., Long Desc.

Weeks, Level, Language

fastText

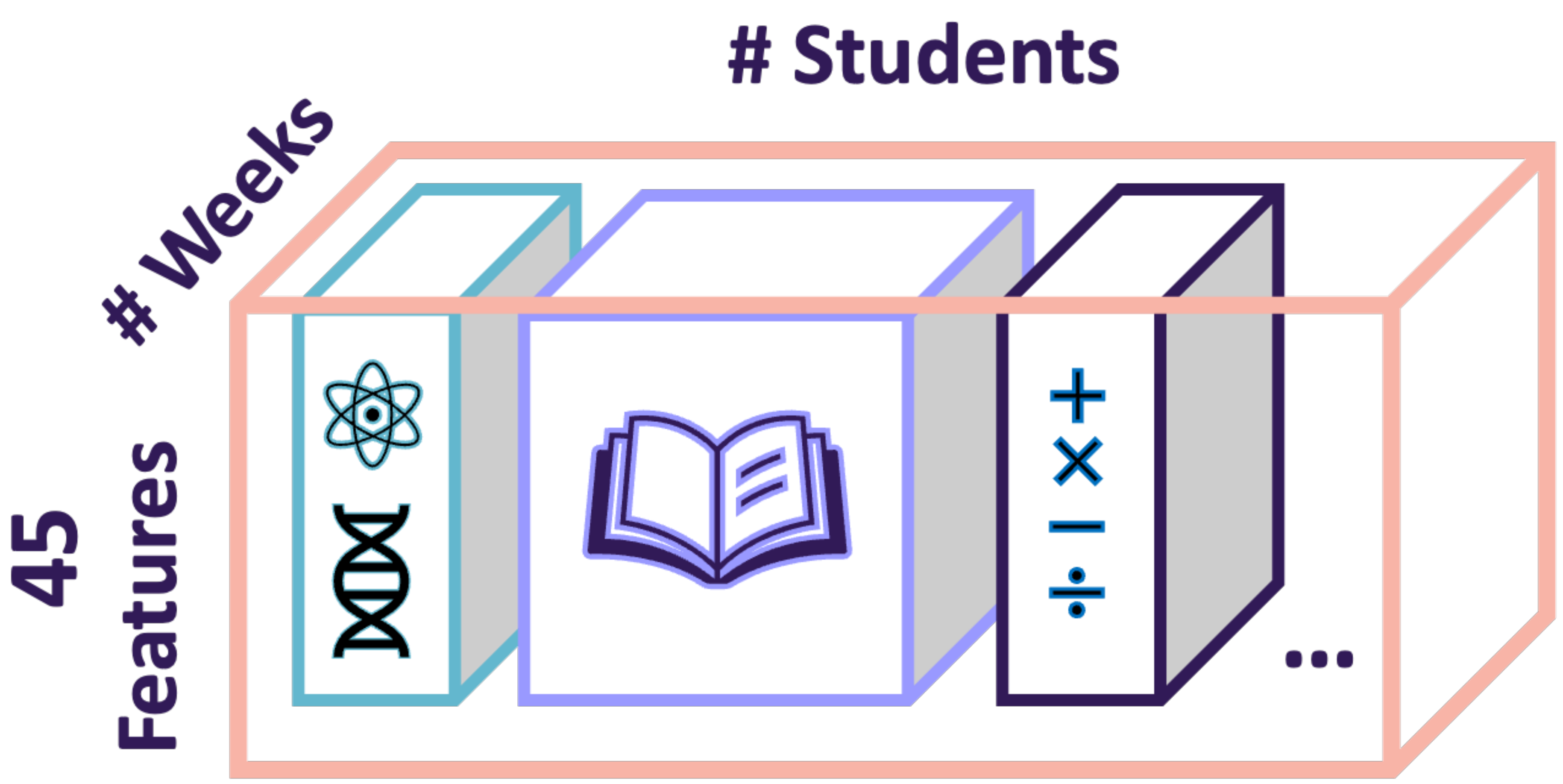
66 - 126 features

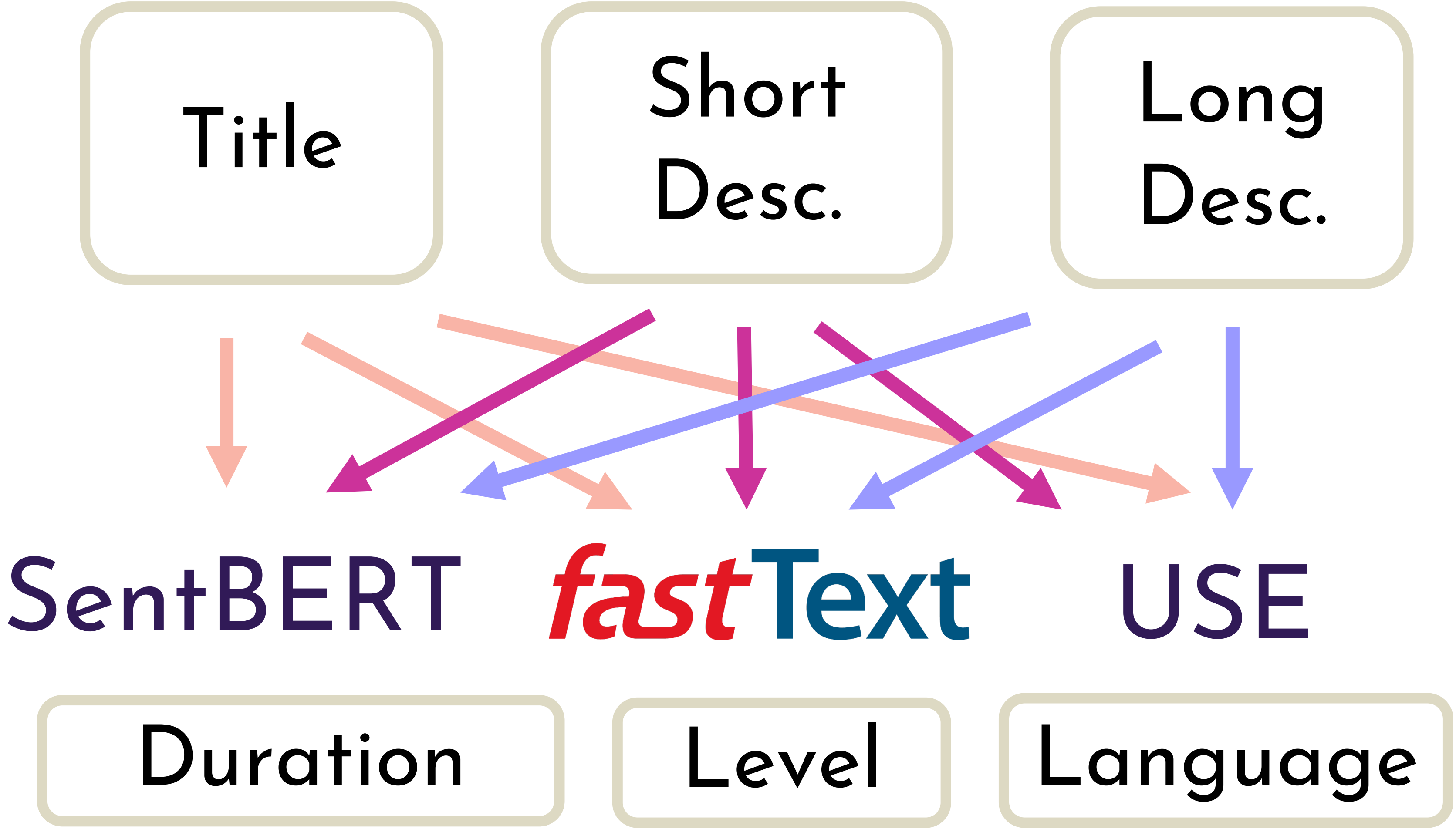
All features are derived from previous work.

(Boroujeni et al., Marras et al., Chen Cui, Lalle Conati)

Feature Extraction

METHODOLOGY





Behavior Features

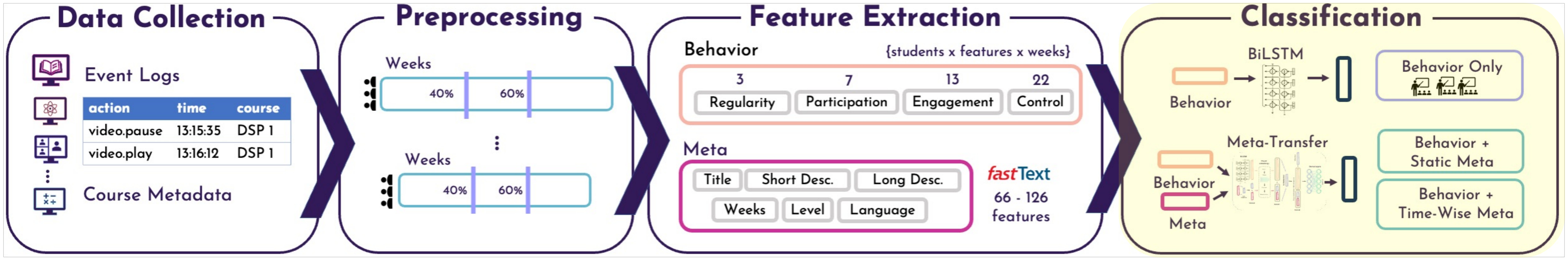
Features derived from student clickstream.

Meta Features

Features derived metadata about the course.

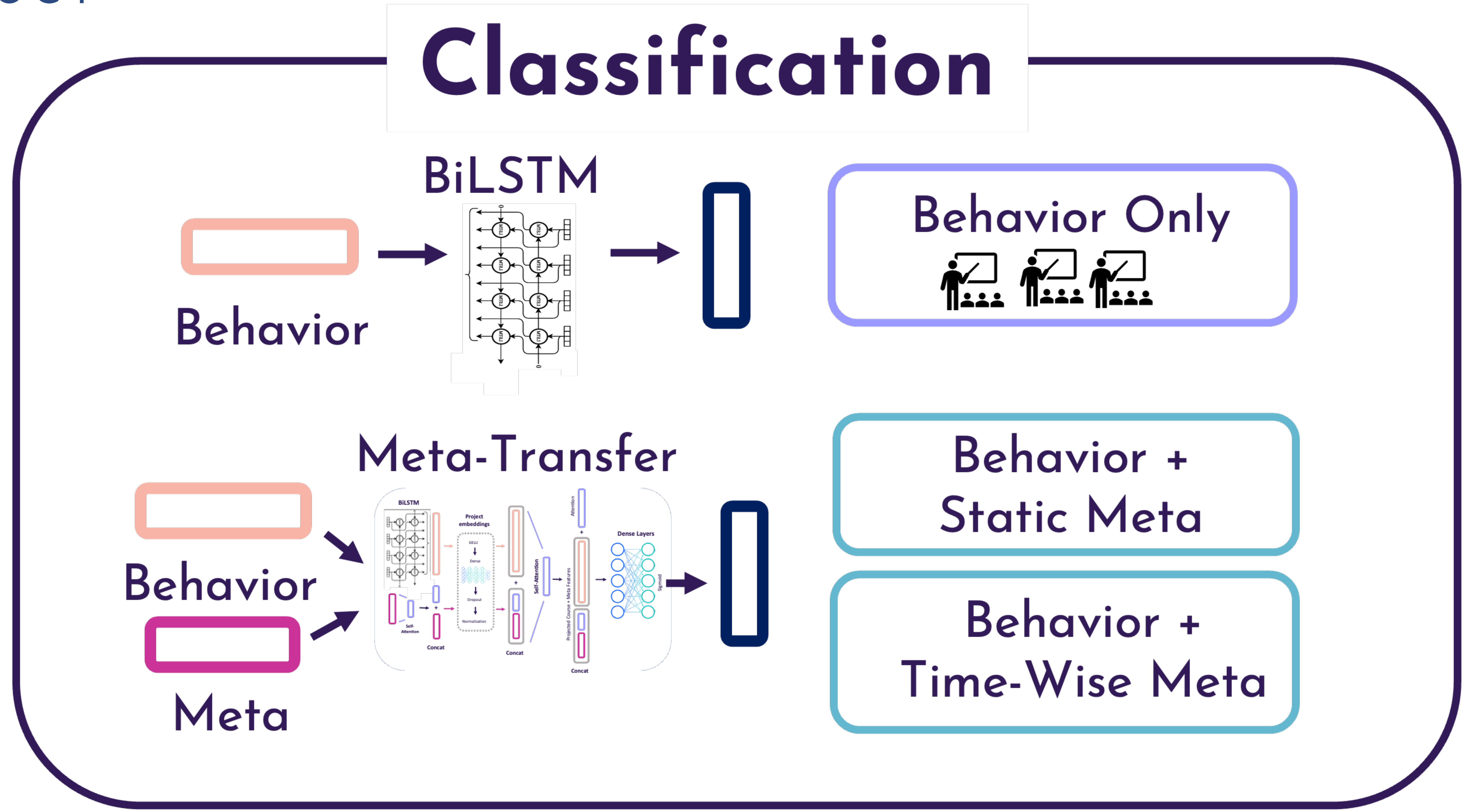
Pipeline

METHODOLOGY



Pipeline

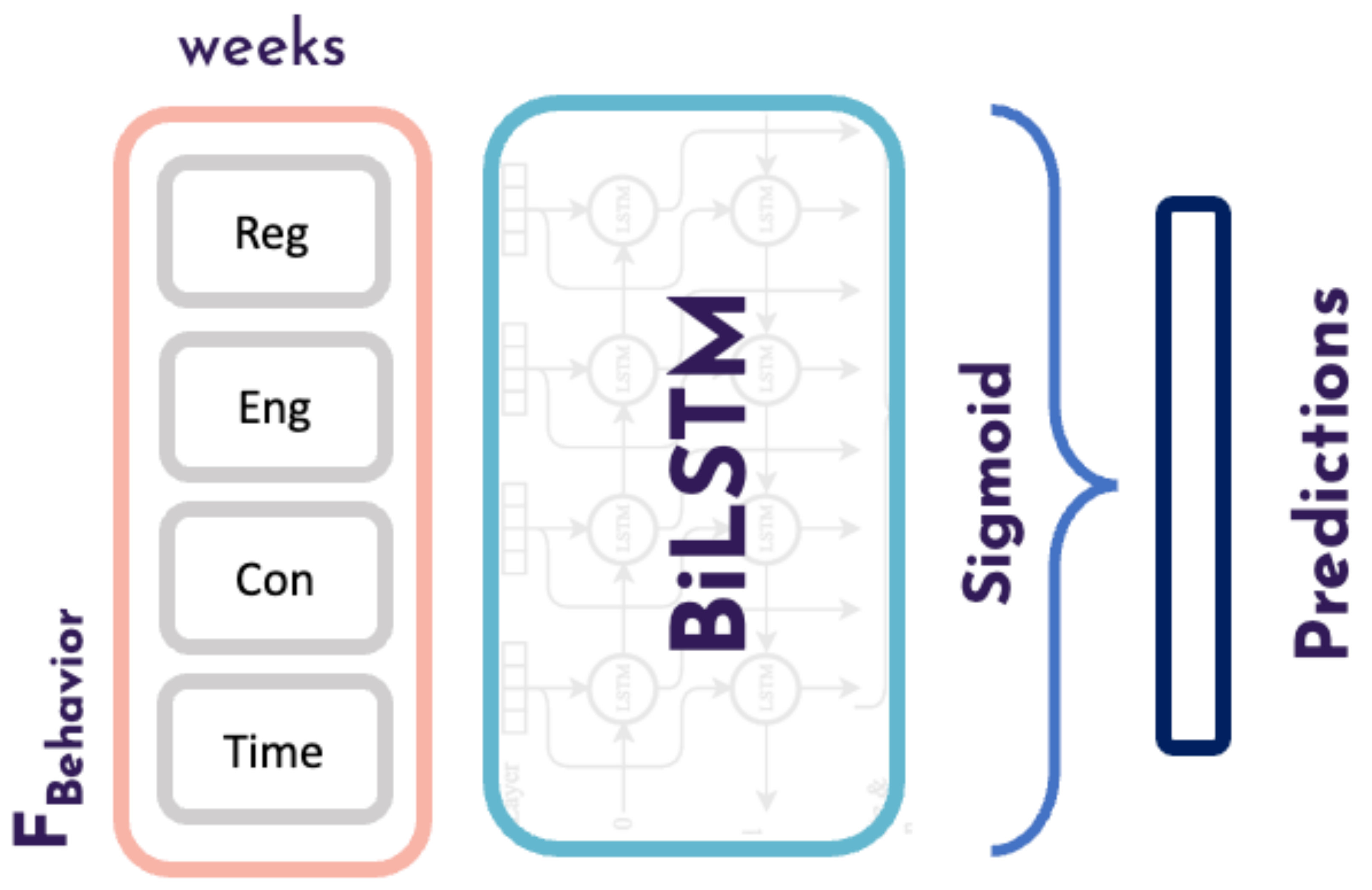
METHODOLOGY



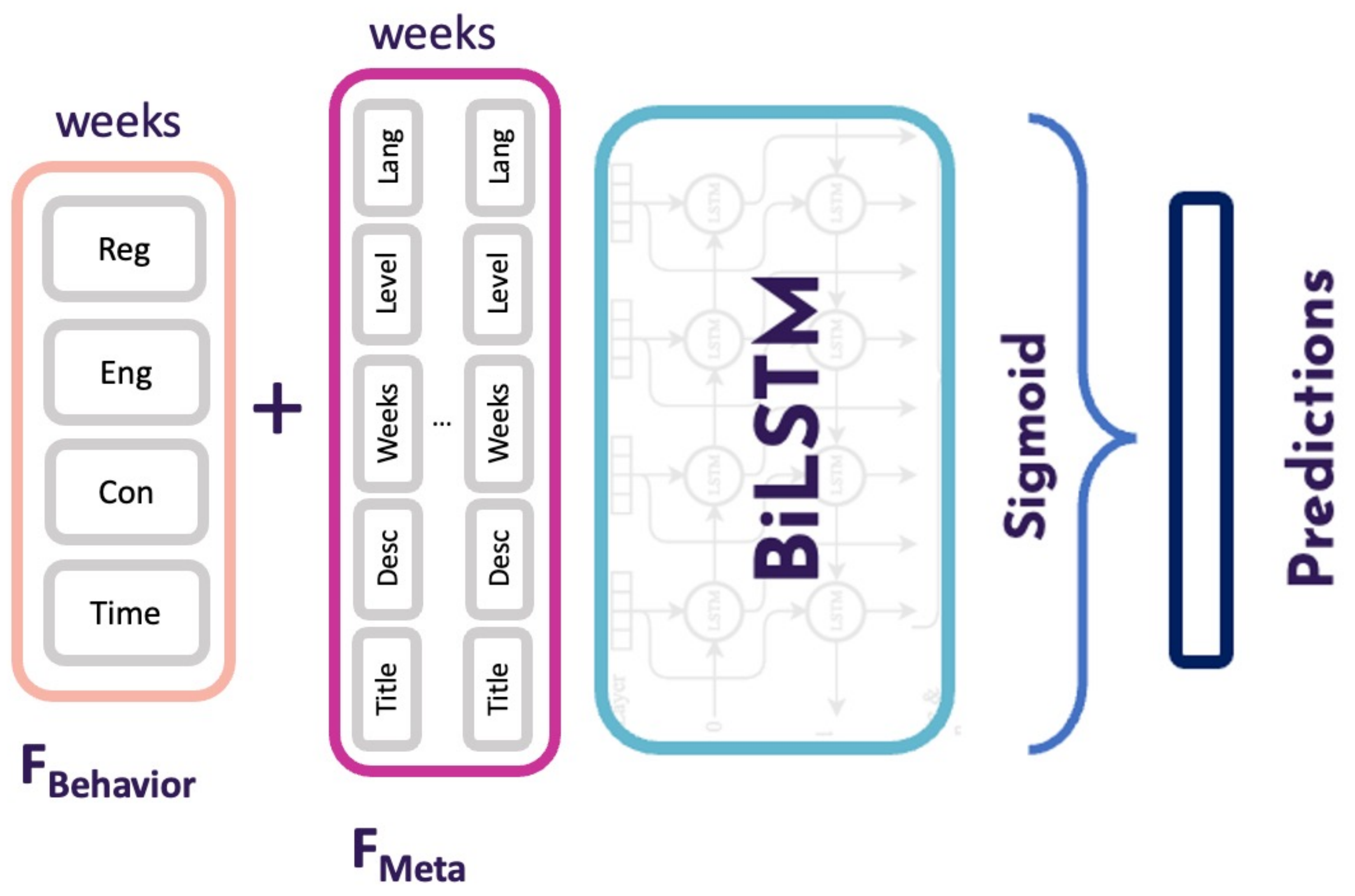
Evaluation metric: Balanced Accuracy (BAC)

Classification

METHODOLOGY



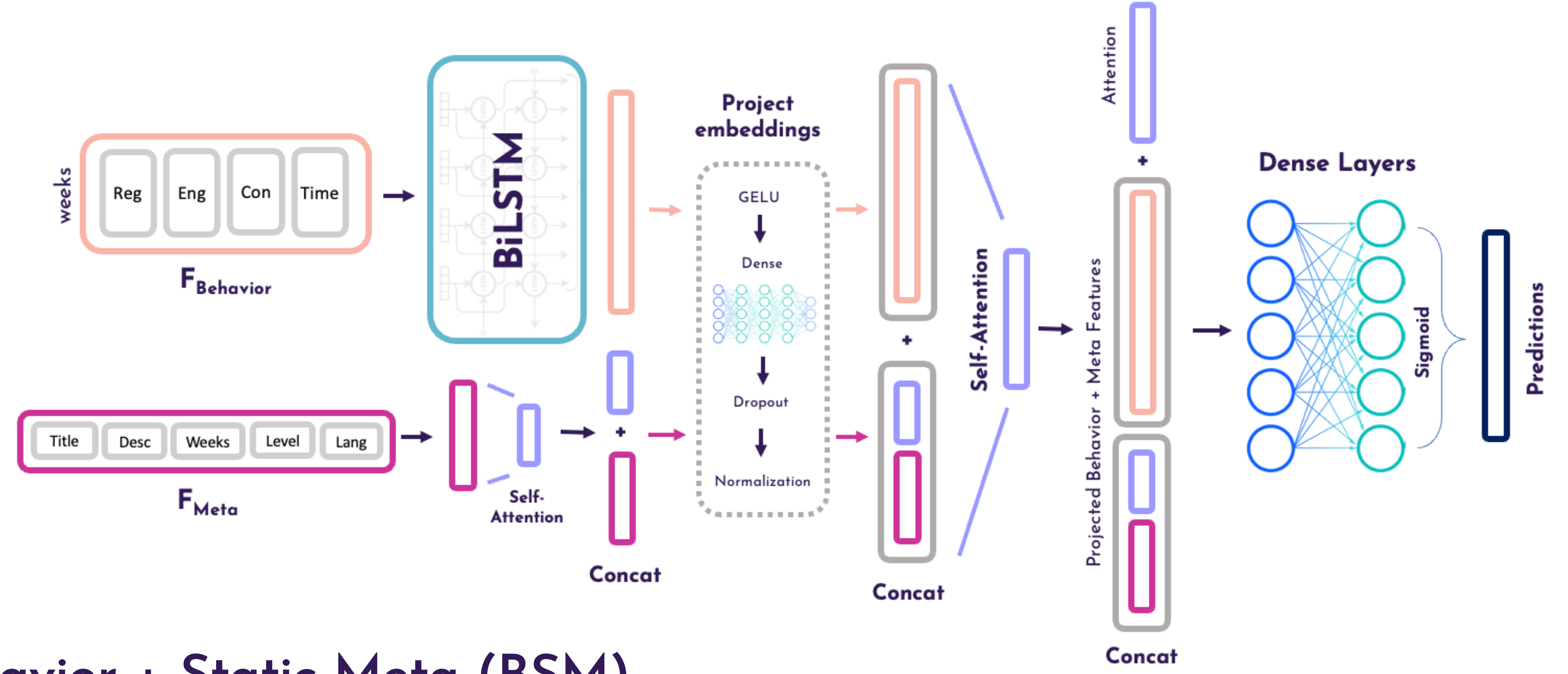
Behavior Only (BO)



Behavior + Time-wise Meta (BTM)

Classification

METHODOLOGY



Behavior + Static Meta (BSM)

RQ1: Baselines

RESULTS

Can **student behavior transfer** across iterations of the same course and across different courses?

Train on:

1) Same course
2) Different course
3) Previous iterations
4) 20 courses

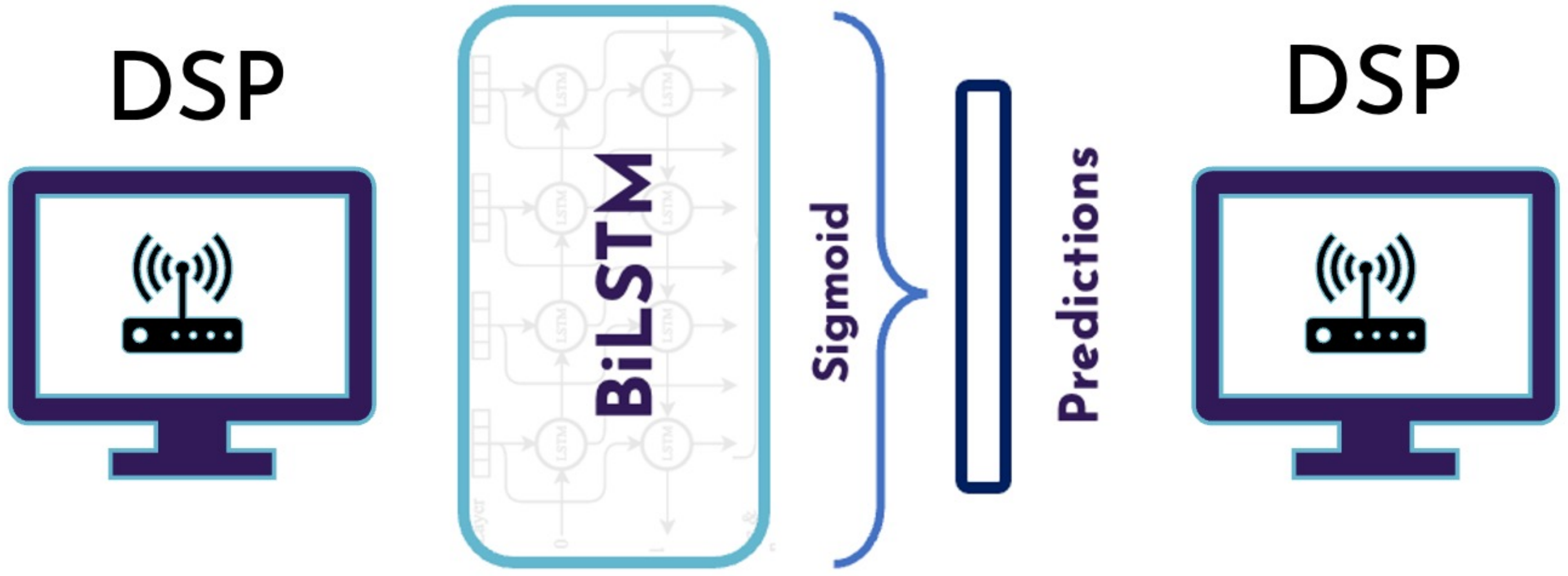
Transfer on 6 courses

RQ1: Baselines

RESULTS

Can **student behavior transfer** across iterations of the same course and across different courses?

	60% Early Prediction Level
	1-1 Same
DSP	92.7
Villes Africaines	82.9
Structures	55.2
ProgFun	50.8
Ventures	54.9
Geomatique	79.5



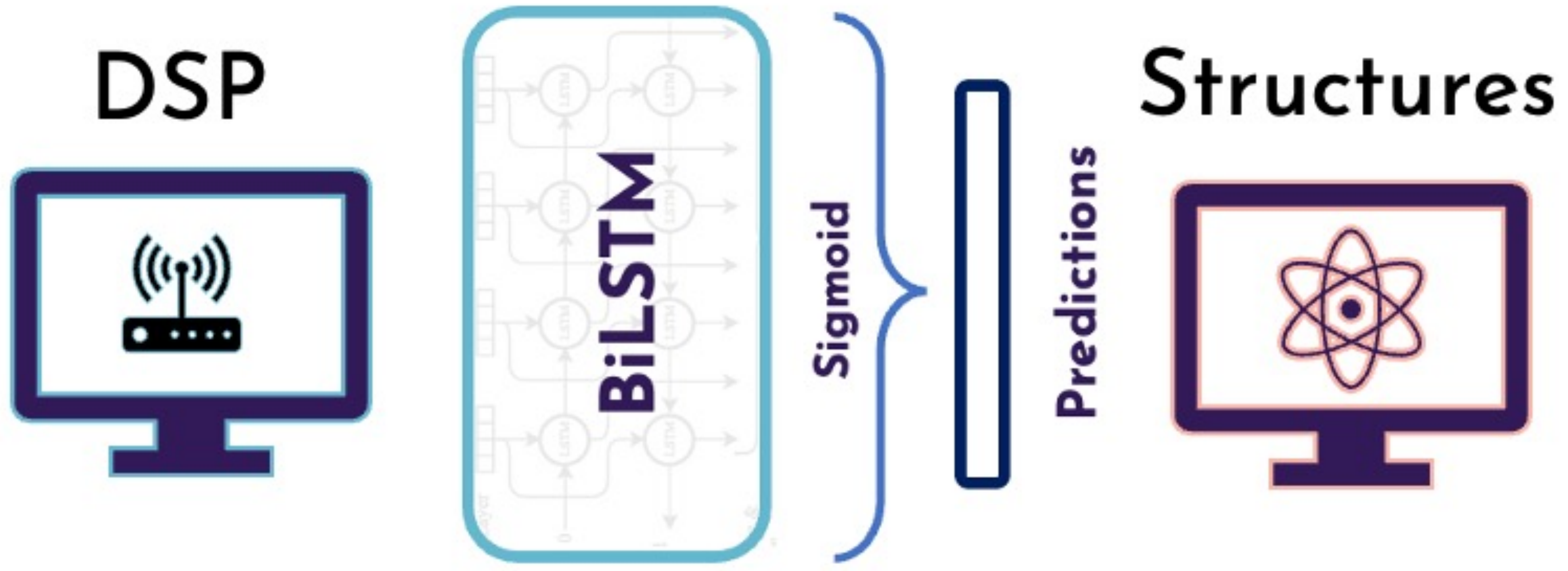
1-1 Same: predict on same course as training

RQ1: Baselines

RESULTS

Can student behavior transfer across iterations of the same course and across different courses?

	60% Early Prediction Level 1-1 Diff[2]
DSP	65.3
Villes Africaines	67.0
Structures	51.3
ProgFun	51.0
Ventures	60.2
Geomatique	57.6



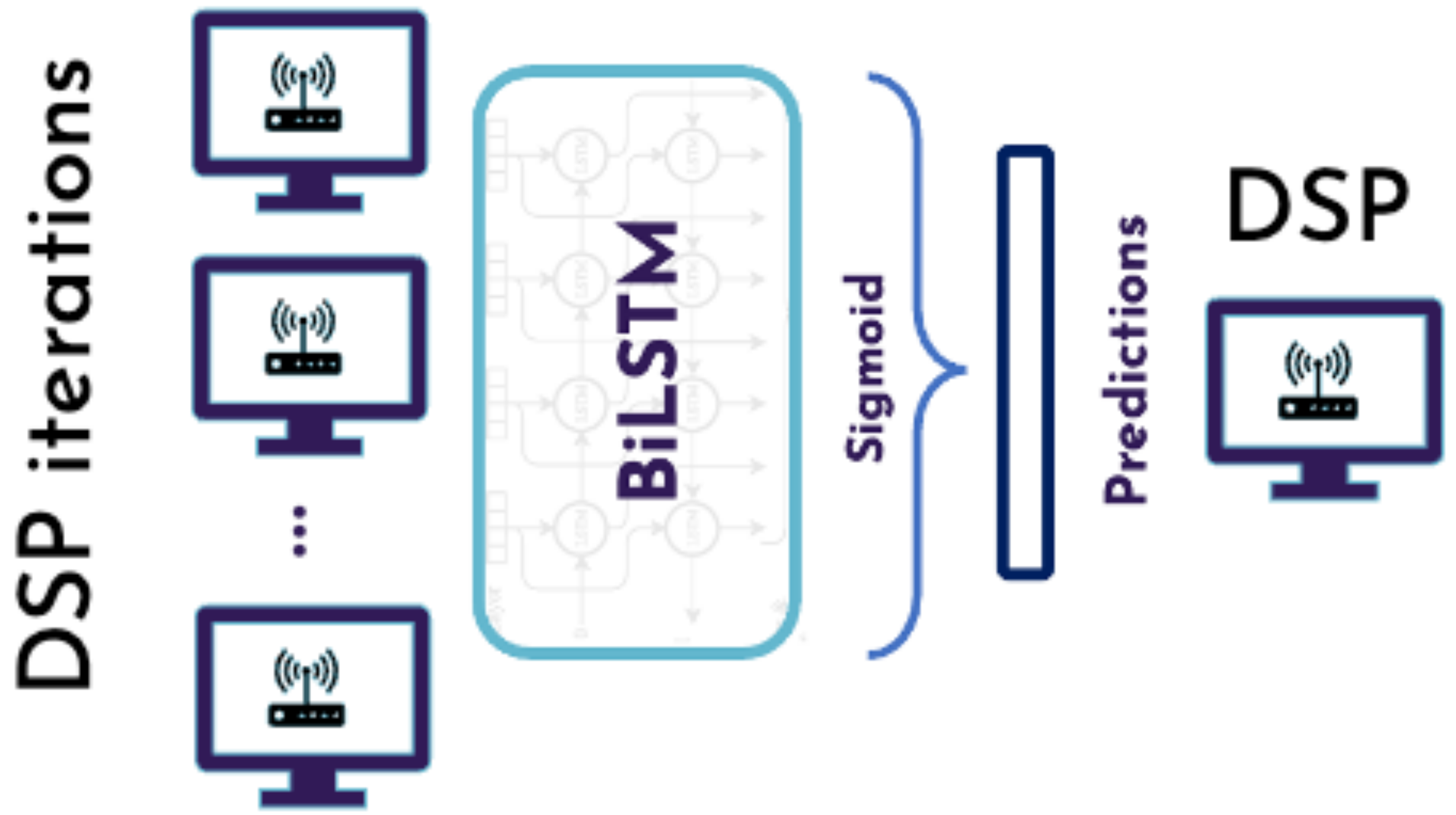
1-1 Diff: predict on different course from training

RQ1: Baselines

RESULTS

Can student behavior transfer across iterations of the same course and across different courses?

	60% Early Prediction Level N-1 Same
DSP	**91.8**
Villes Africaines	80.7
Structures	50.4
ProgFun	**62.3**
Ventures	-
Geomatique	-



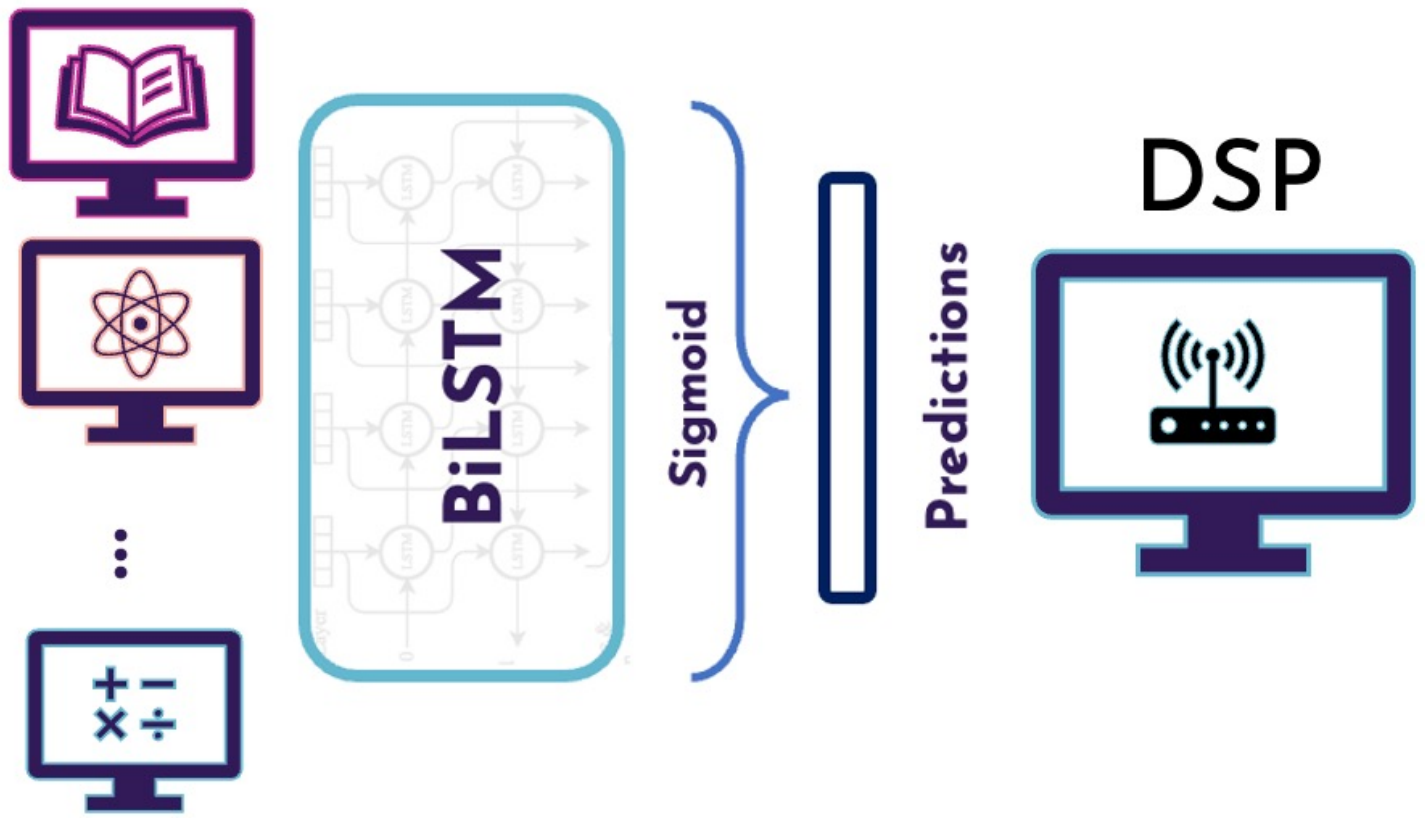
N-1 Same: train on previous iterations

RQ1: Baselines

RESULTS

Can student behavior transfer across iterations of the same course and across different courses?

	60% Early Prediction Level N-1 Diff
DSP	87.8
Villes Africaines	**82.7**
Structures	**54.4**
ProgFun	62.0
Ventures	**71.8**
Geomatique	**65.5**



N-1 Diff: train on 20 courses, predict on 6

RQ1: Baselines

RESULTS

Can student behavior transfer across iterations of the same course and across different courses?

	60% Early Prediction Level			
	1-1 Same[1]	1-1 Diff[2]	N-1 Same	N-1 Diff
DSP	92.7	65.3	$\boxed{\textbf{91.8}}$	87.8
Villes Africaines	82.9	67.0	80.7	$\boxed{\textbf{82.7}}$
Structures	55.2	51.3	50.4	$\boxed{\textbf{54.4}}$
ProgFun	50.8	51.0	$\boxed{\textbf{62.3}}$	62.0
Ventures	54.9	60.2	-	$\boxed{\textbf{71.8}}$
Geomatique	79.5	57.6	-	$\boxed{\textbf{65.5}}$

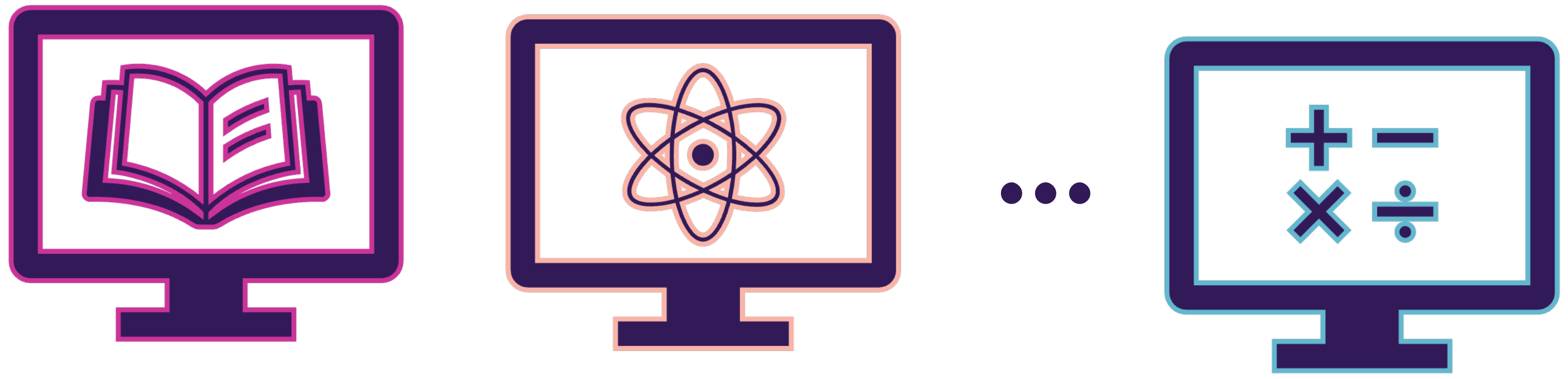
Yes! Courses with previous iterations, high # of students, and high passing rate benefit from models trained on previous iterations (N-1 Same).

Otherwise, training on different courses (N-1 Diff) is better.

RQ2: Meta Models

RESULTS

Is a meta learning model trained on a **combination of behavior and course metadata** information more transferable?



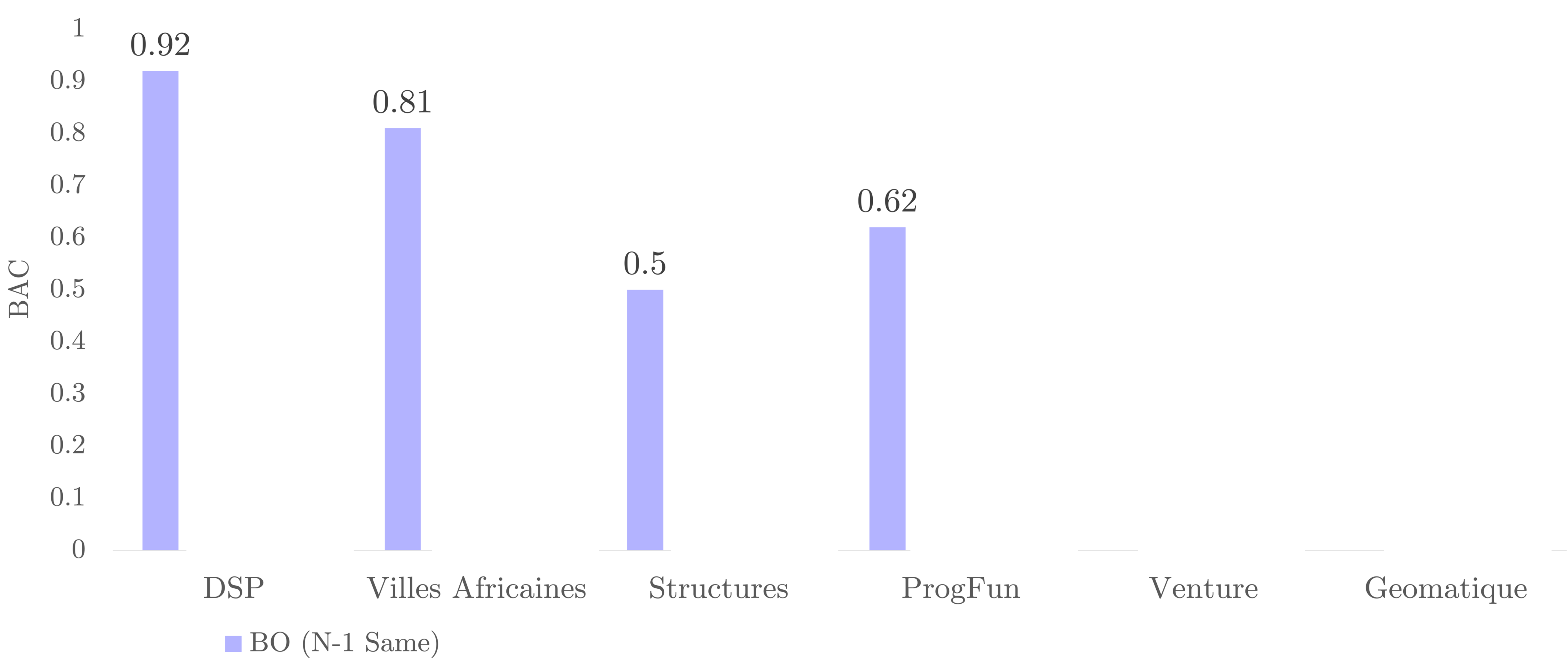
Train on 20 courses
(behavior + meta)

Transfer on 6 courses

RQ2: Meta Models

RESULTS

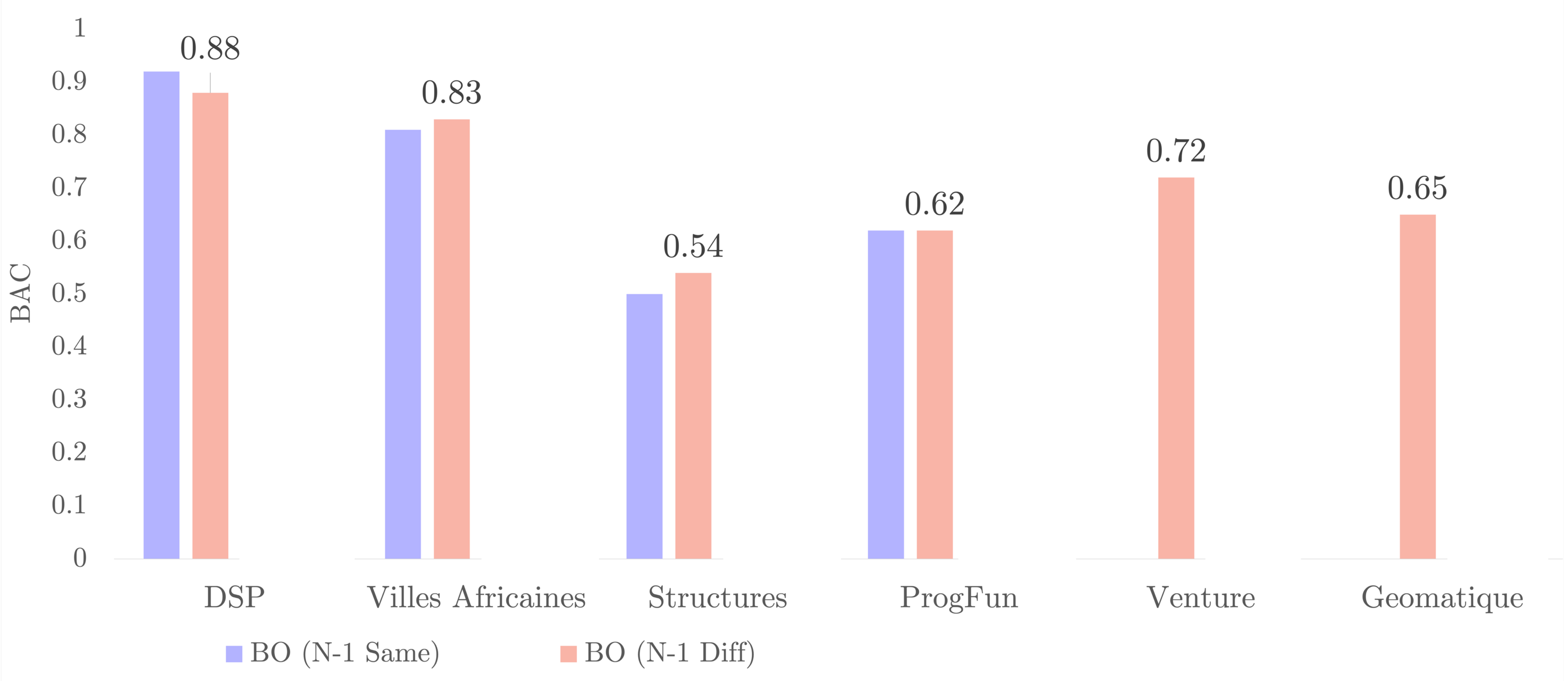
Is a meta learning model trained on a **combination of behavior and course metadata** information more transferable?



RQ2: Meta Models

RESULTS

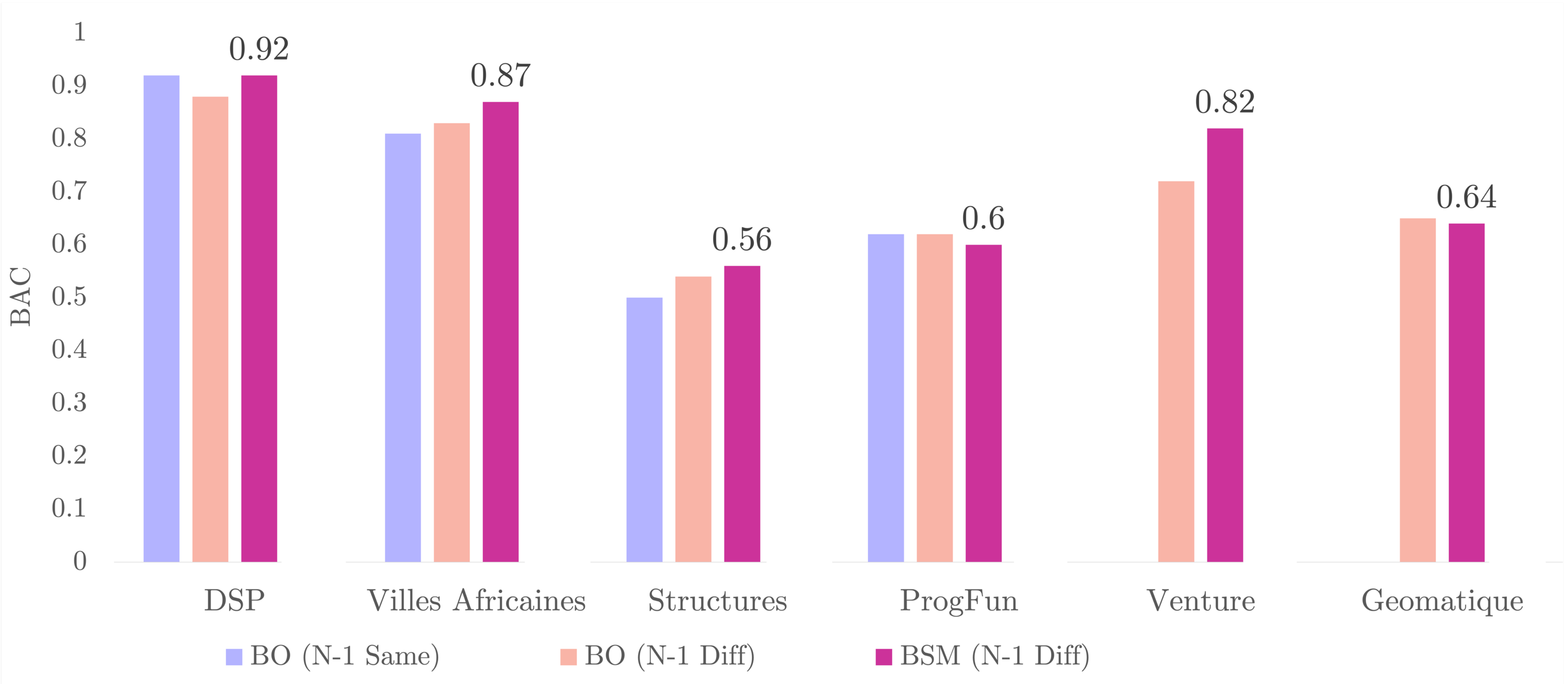
Is a meta learning model trained on a **combination of behavior and course metadata** information more transferable?



RQ2: Meta Models

RESULTS

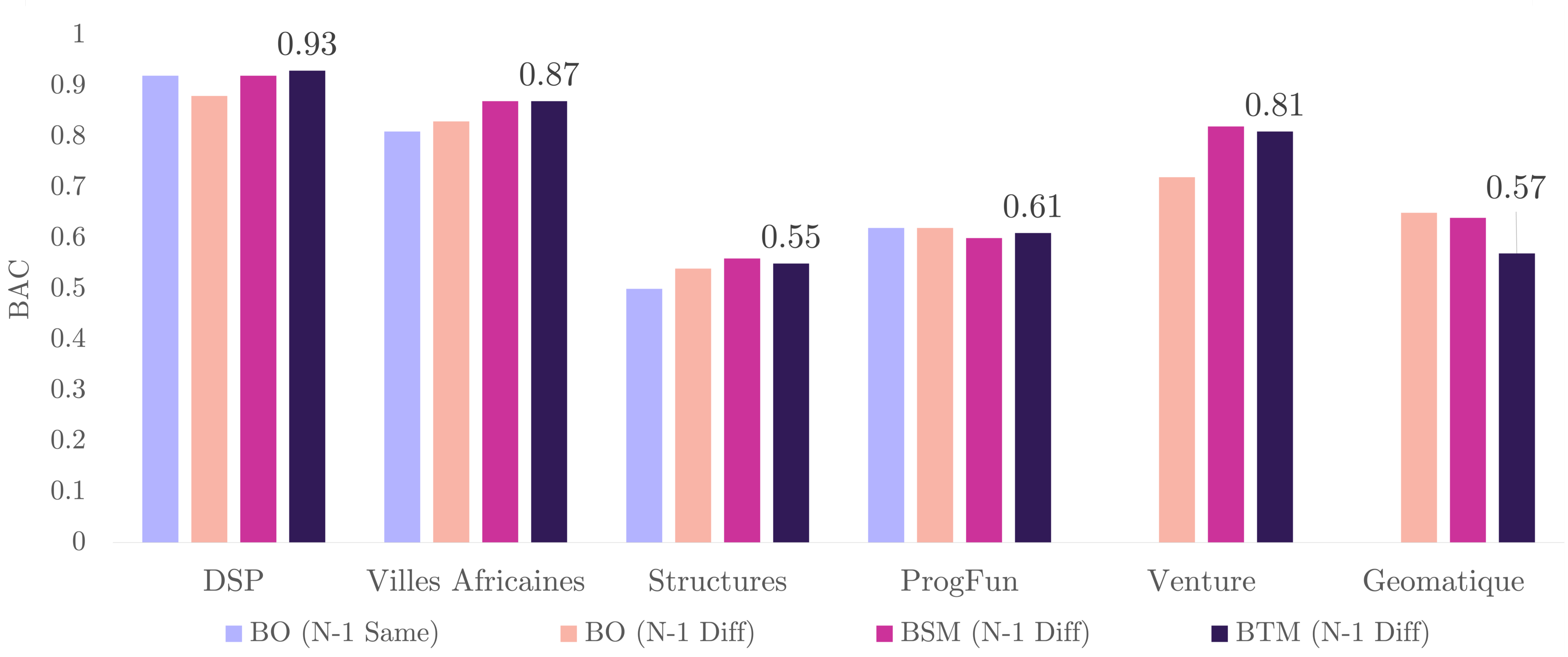
Is a meta learning model trained on a **combination of behavior and course metadata** information more transferable?



RQ2: Meta Models

RESULTS

Is a meta learning model trained on a **combination of behavior and course metadata** information more transferable?

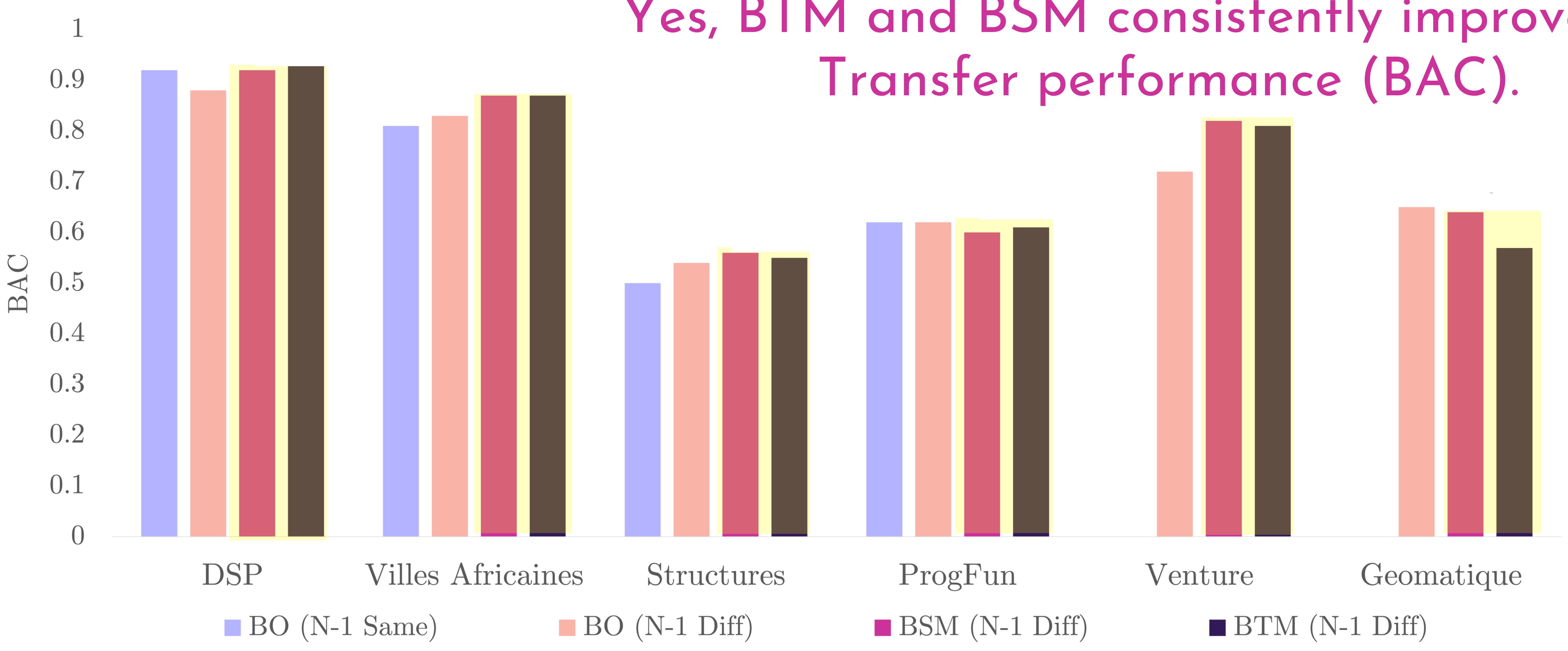


RQ2: Meta Models

RESULTS

Is a meta learning model trained on a **combination of behavior and course metadata** information more transferable?

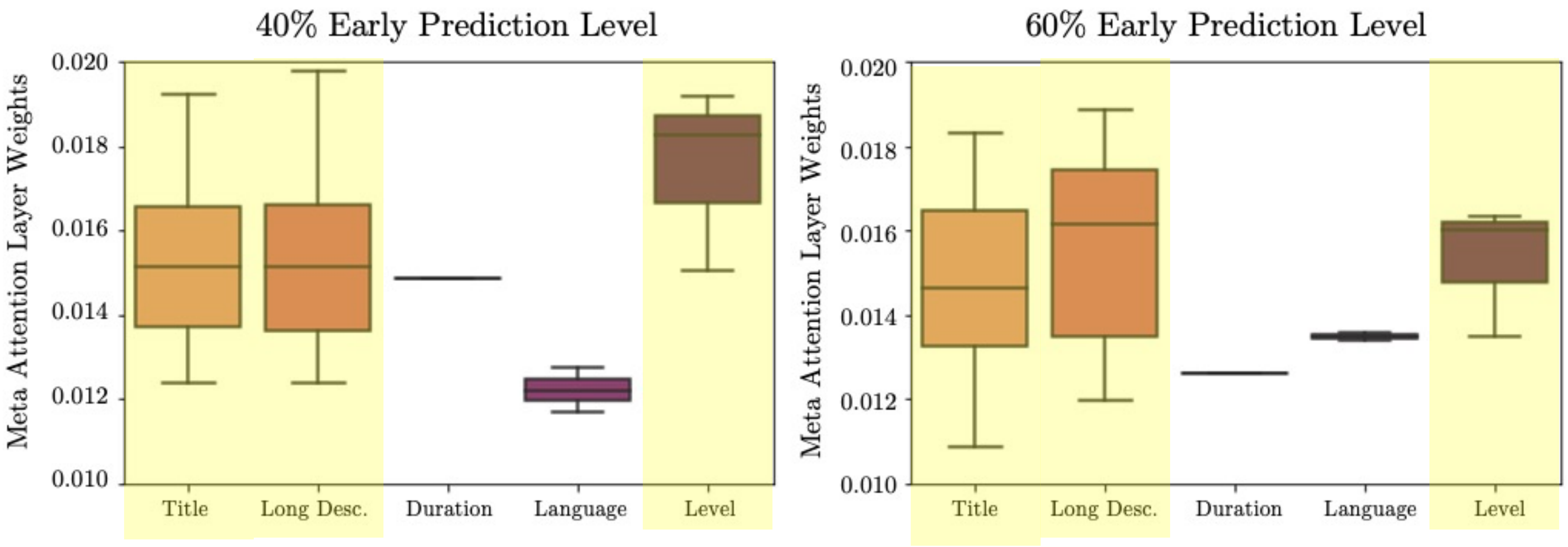
Yes, BTM and BSM consistently improve Avg. Transfer performance (BAC).



RQ2: Attention Layers

RESULTS

Which **meta features** are important?

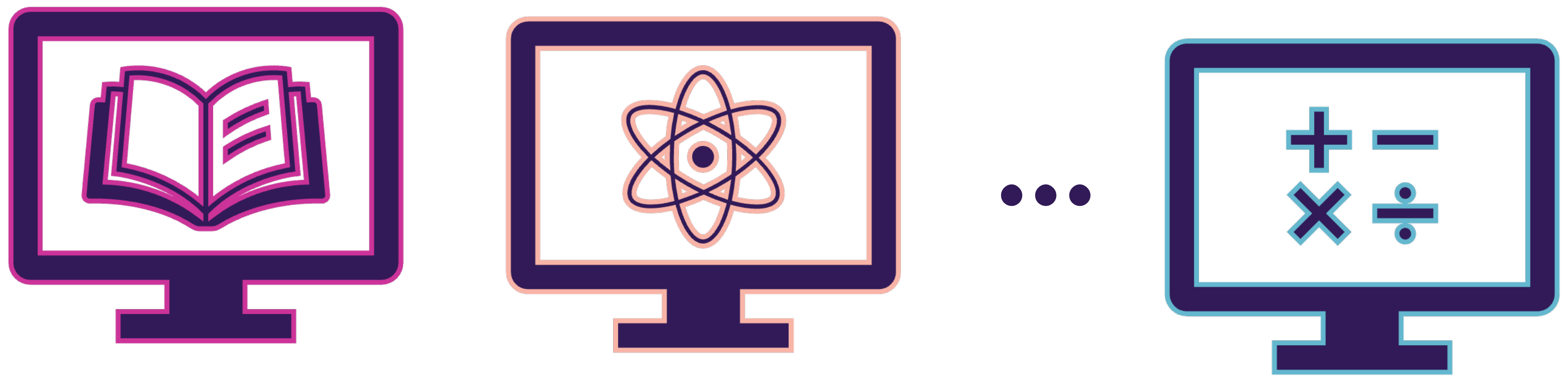


Title, Level and Long Description get a lot of attention.
Behavior and meta features have similar levels of attention.

RQ3: Fine-Tuning

RESULTS

Can fine-tuning a combined model on past iterations of an unseen course lead to better transferable models?



Train on 25 courses
(behavior + meta)

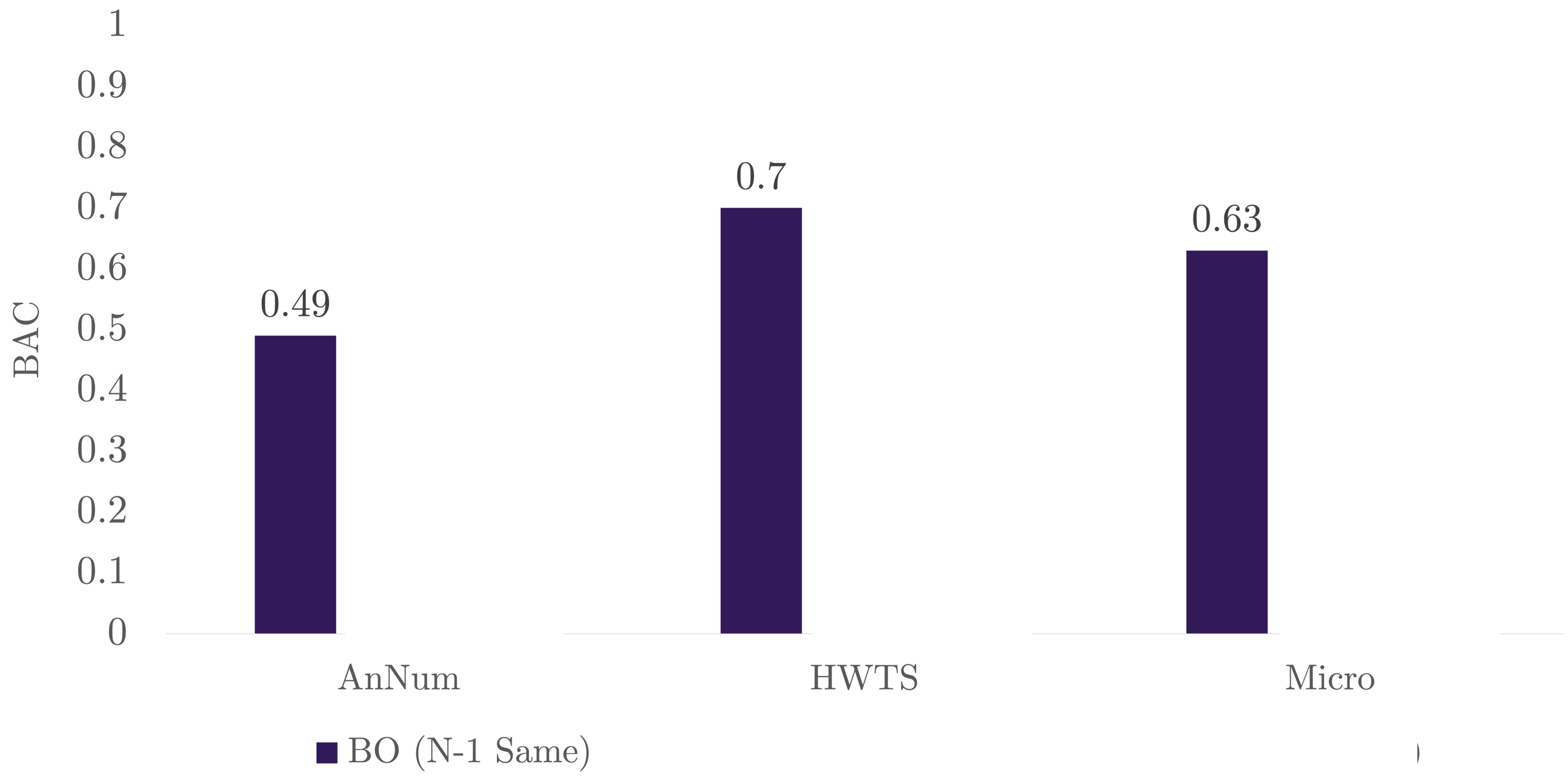
Fine-tune + Transfer on 3 courses

Diverse in: # students, pass-ratio, # quizzes!

RQ3: Fine-tuning

RESULTS

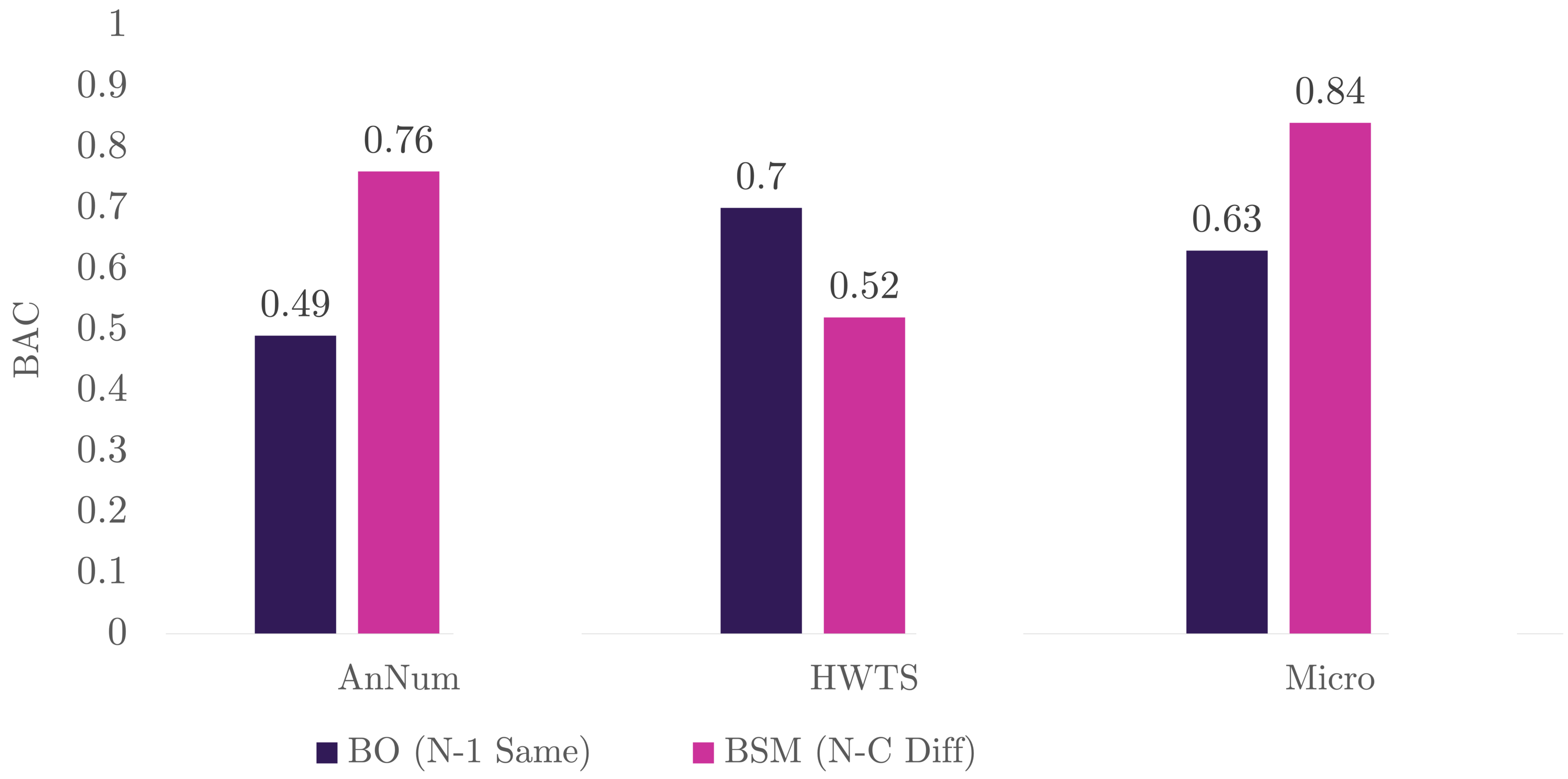
Can fine-tuning a combined model on past iterations of an unseen course lead to better transferable models?



RQ3: Fine-tuning

RESULTS

Can fine-tuning a combined model on past iterations of an unseen course lead to better transferable models?



RQ3: Fine-tuning

RESULTS

Can fine-tuning a combined model on past iterations of an unseen course lead to better transferable models?

When the past iterations are very similar (in student population and course structure) to the current iteration, FT helps. If not, it hurts.

Implications

DISCUSSION

1) Training on a large model is very helpful!

2) Combining behavior + meta features using attention leads to the best performance.

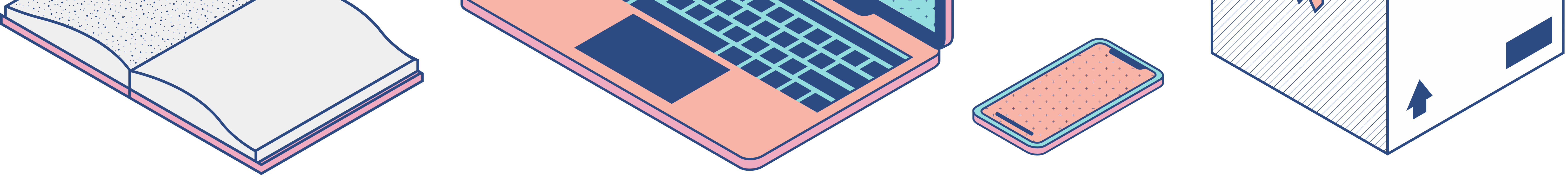
3) Level, Title, and Duration meta features are very important.

4) When previous iterations are available and similar to the ongoing course, fine-tuning helps better transfer.

Extensions

FUTURE WORK

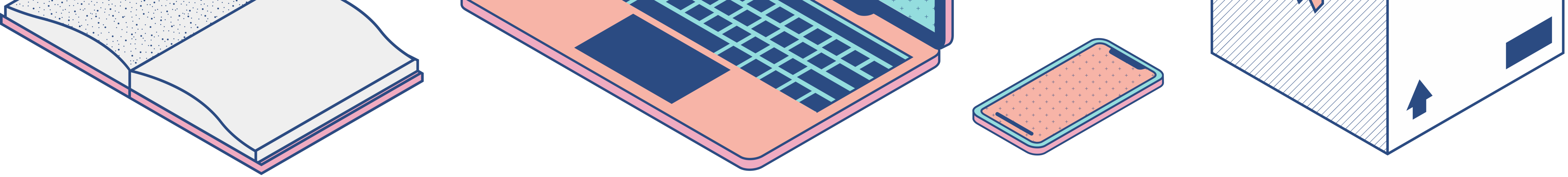
- Extend dataset to other universities and world regions.
- Extend to different modalities (flipped classrooms and blended courses) to see if transferability is modality-agnostic.
- Use latent feature representations from autoencoders.



Main Takeaways

META TRANSFER LEARNING FOR EARLY SUCCESS PREDICTION IN MOOCS

Large models combining **interaction data and meta information** have comparable or better performance than models which have access to previous iterations of the course.



Main Takeaways

META TRANSFER LEARNING FOR EARLY SUCCESS PREDICTION IN MOOCS

epfl-ml4ed/
meta-transfer-learning

Using our models, educators can
warm-start predictions
for their ongoing or small courses!

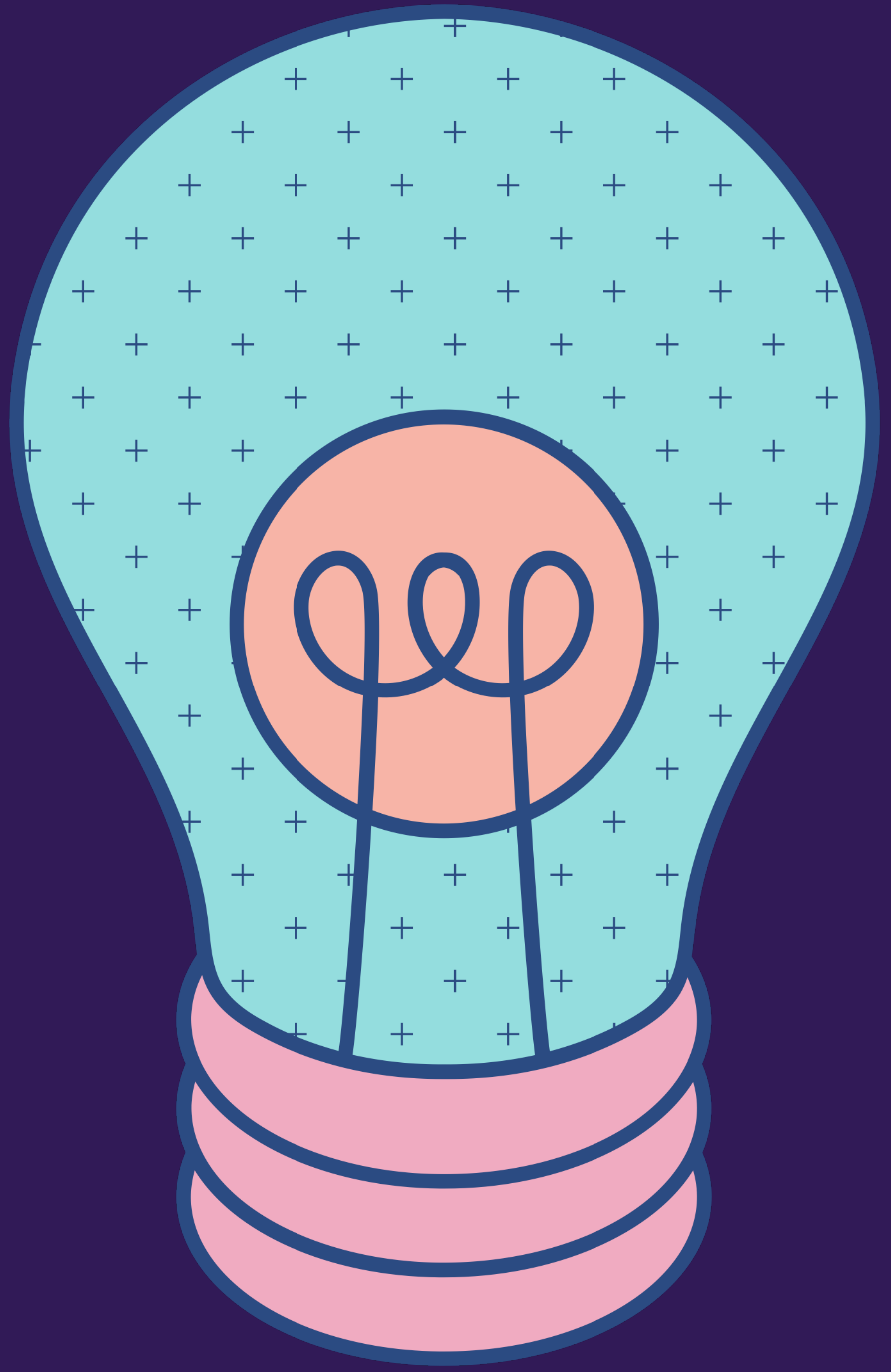
Thank you!

META TRANSFER LEARNING FOR STUDENT SUCCESS PREDICTION IN MOOCS

Questions?

META TRANSFER LEARNING FOR STUDENT SUCCESS PREDICTION IN MOOCS

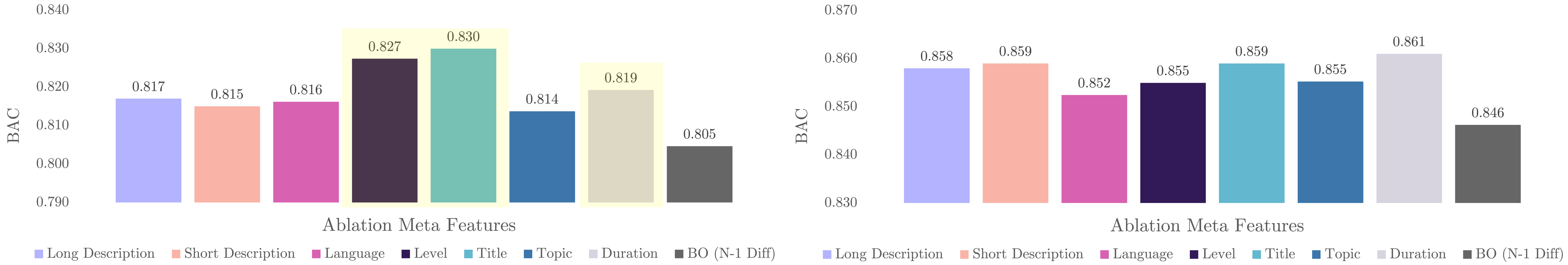
Vinitra Swamy
epfl-ml4ed/meta-transfer-learning
vinitra.swamy@epfl.ch



RQ2: Ablation Study

RESULTS

Which **meta features** are important?

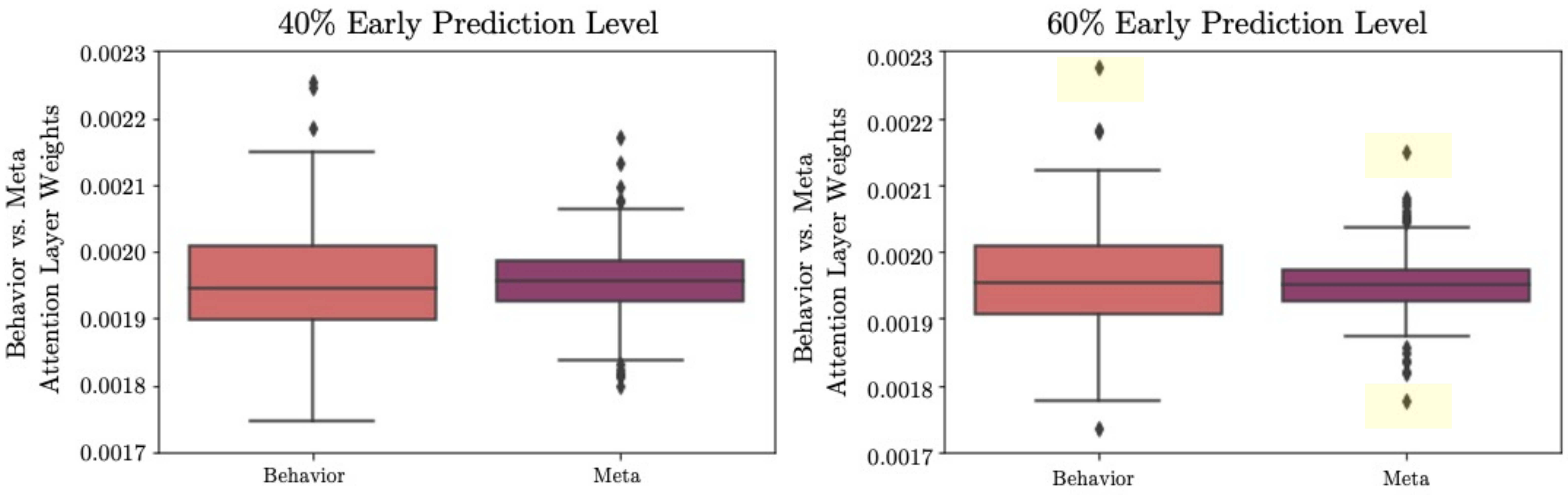


Level, Title, and Duration meta features are important for very early predictions.

RQ2: Attention Layers

RESULTS

Are meta features important?



More variance in importance of behavior and meta features in later predictions.

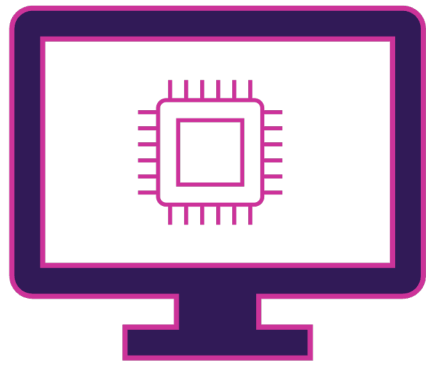
Data Collection

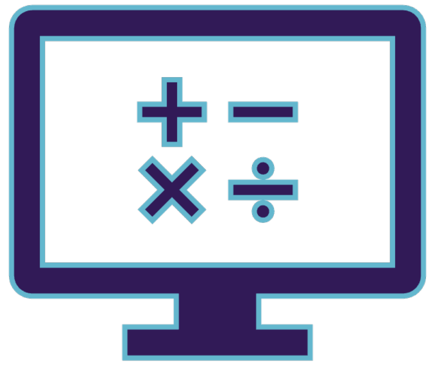
METHODOLOGY

	Digital Signal Processing	Villes Africaines	Structures	Functional Programming	Venture	Geomatique
					--- only 1 iteration ---	
Lang.	English	En/Fr	French	French	English	French
Level	MSc	BSc/Prop	BSc	BSc	BSc	BSc
Students	15k	13k	350	19k	7k	450

Data Collection

METHODOLOGY





French BSc

	Microcontrôleurs	Analyse Numerique	HWTS
Students	4k	1.5k	2.5k
Quizzes	18	36	10
Weeks	10	9	5

Dataset

METHODOLOGY

Title	Identifier	Iterations[1] Trn	Iterations[1] Trs	Topic[2]	Level[3][8]	Language[4]	No. Weeks[5]	No. Students[3] Trn	No. Students[3] Trs	Passing Rate[6] [%] Trn	Passing Rate[6] [%] Trs	No. Quizzes[7]
Comprendre les Microcontrôleurs	Micro	4	0	Eng	BSc	French	10	3,974	-	26.9	-	18
Analyse Numérique	AnNum	3	0	Math	BSc	French	9	1,471	-	51.5	-	36
Household Water Treatment and Storage	HWTS	2	0	NS	BSc	French	5	2,438	-	47.2	-	10
Programmation Orientée Objet	OOP	1	0	CS	Prop	French	10	797	-	38.1	-	10
Programmation en C++	InitProgC++	1	0	CS	Prop	English	8	728	-	63.3	-	13
Digital Signal Processing	DSP	4	1	CS	MSc	English	10	11,483	4,012	22.6	23.1	38
Villes Africaines	Villes Africaines	2	1	SS	BSc/Prop[9]	En/Fr[9]	12	7,888	5,643	6.3	9.9	18
L'Art des Structures I	Structures	2	1	Arch	BSc	French	10	278	95	57.7	66.3	6
Functional Programming	ProgFun	1	1	CS	BSc	French	7	11,151	7,880	50.72	81.33	3
Launching New Ventures	Venture	0	1	Bus	BSc	English	7	-	6,673	-	1.4	13
Éléments de Géomatique	Geomatique	0	1	Math	BSc	French	11	-	452	-	45.1	27

[1] **Set abbrev.** *Trn*: training; *Trs*: transfer.
[2] **Topic abbrev.** *Eng*: Engineering; *Math*: Mathematics; *NS*: Natural Science; *CS*: Computer Science; *SS*: Social Science; *Arch*: Architecture; *Bus*: Economics and Business.
[3] The values are computed after removing early-dropout students.[4]**Level** is chosen by majority label in *Trs* or *Trn*. [5]**Language** is chosen by majority label in *Trs* or *Trn*.
[6] **Passing Rate** is averaged over the courses in *Trs* or *Trn* weighted by number of students. [7]**No. Quizzes** is averaged over the courses in *Trs* or *Trn*.
[8] **Level abbrev.** *Prop*: Propedeutic / Other; *BSc*: Bachelor; *MSc*: Master. [9] For *Villes Africaines*, the / operator represents characteristics of courses in *Trn* / *Trs*.

Data Collection

METHODOLOGY

High Pass Ratio → Low Pass Ratio

Functional Programming

Structures

HWTS

Analyse Numerique

Geomatique

Villes Africaines

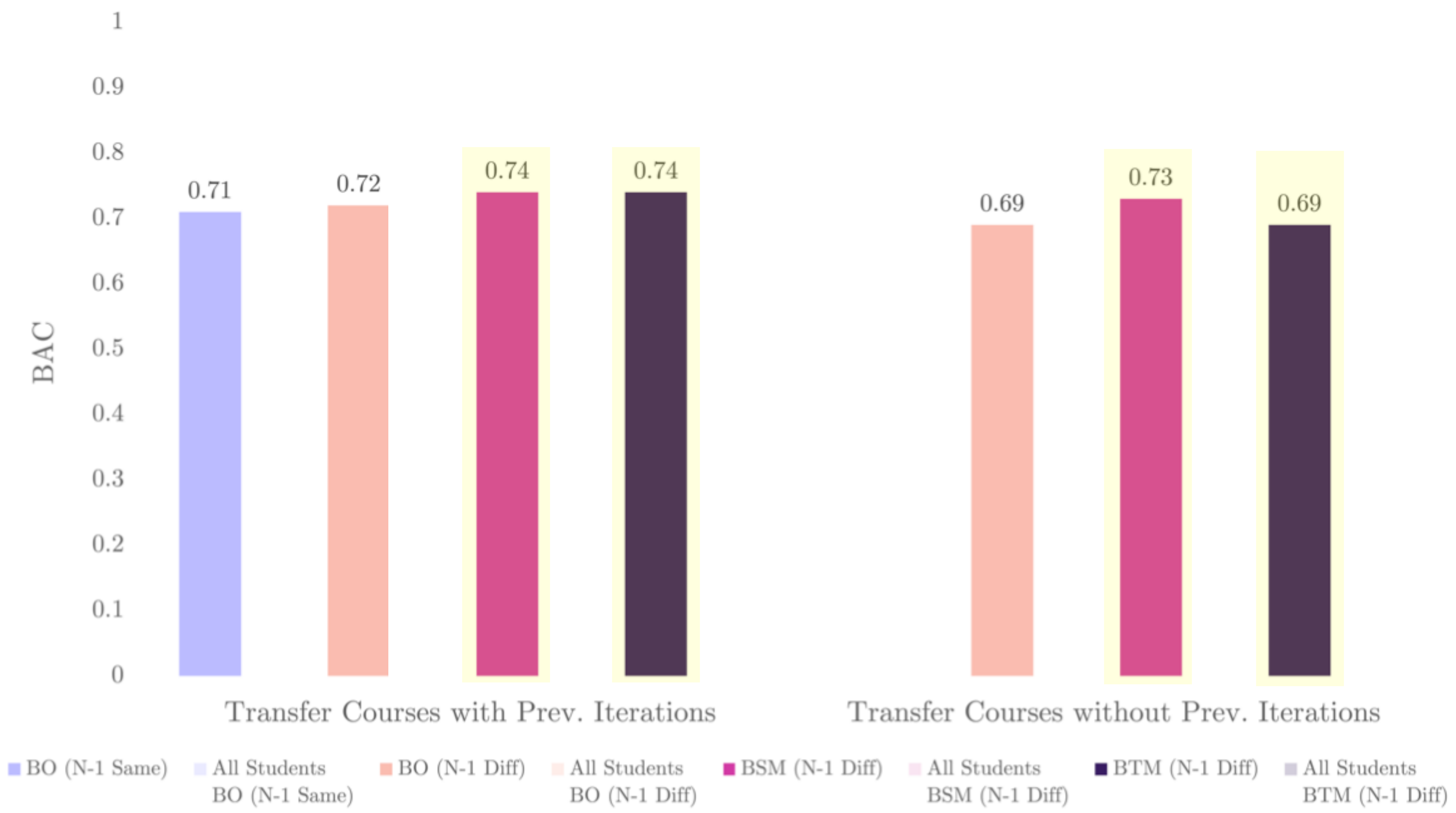
Venture

Digital Signal Processing

Micro

RQ2: Meta Models

RESULTS



Is a meta learning model trained on a **combination of behavior and course metadata** information more transferable?

On average, meta models beat prev. iterations models.