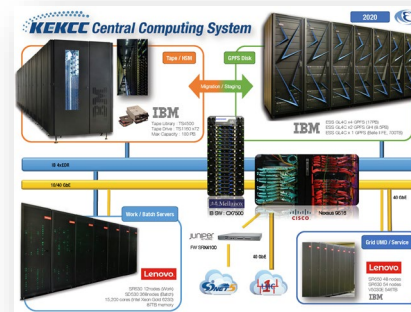


# TYL-FJPPL: Comp\_04

## Evolution of the computing environment for high-energy and astroparticle experiments



### Members (\*Leader, \*\*new member)

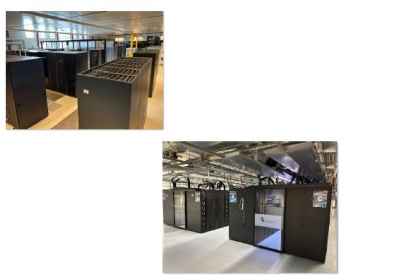
CC-IN2P3: Sebastien GADRAT\*  
F. Hernandez, B. Simon, B. Rigaud, L. Tortay, G. Marchetti\*\*

KEK-CRC: Tomoaki NAKAMURA\*  
T. Kishimoto, G. Iwai, K. Murakami, S. Suzuki, T. Sasaki

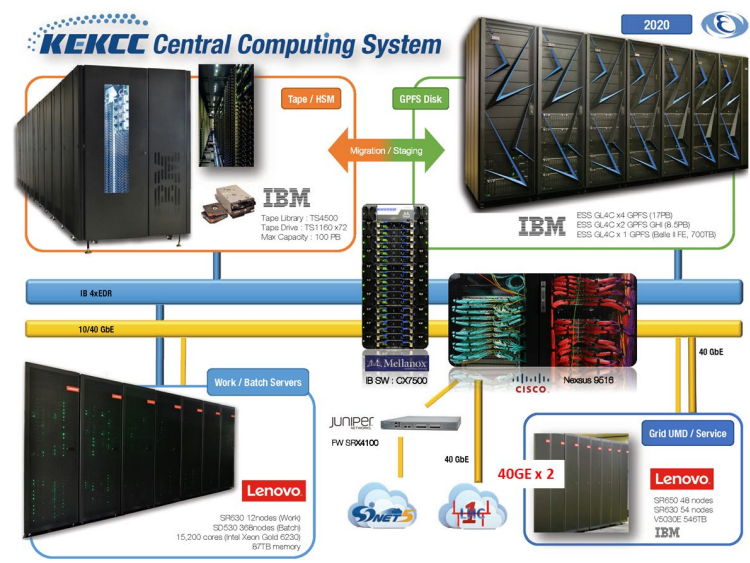
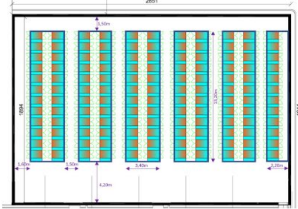
Multiple approaches for introducing brand-new technologies are necessary for the evolution toward the exa-scale computing environment, for example, the construction of more effective data sharing by using high bandwidth network and Cloud technology, application of machine learning for the optimization of the existing computer system, establishment of integrated authentication and authorization infrastructure, under the collaborative work between Computing Center IN2P3 (CC-IN2P3) and KEK Computing Research Center (KEK-CRC).

## Computer rooms

- 2 computer rooms VIL1 and VIL2**
- VIL1**
- Built in 1986
  - 850 m<sup>2</sup>, 1MW IT, 150 racks
  - Old cooling system in the raised floor (PUE=2)
- VIL2**
- Built in 2011
  - 850 m<sup>2</sup>, 3.6 MW IT, 240 racks
  - Modern cooling system with hot corridors (PUE=1.4)
- 2 redundant power sources for the site and a generator for VIL1**



- **New IT room at CC-IN2P3 : VIL3**  
132 racks, 550 m<sup>2</sup>, 2 MW for IT
- **The consortium of companies for the building has been selected, work will start on February 2024**

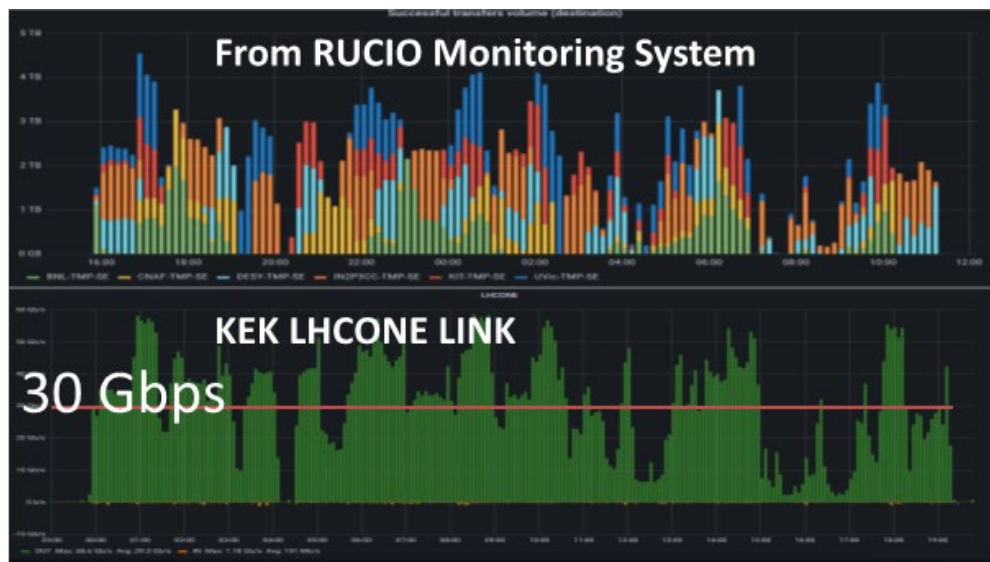


The procurement models for equipment differ between CC-IN2P3 and KEK-CRC. CC-IN2P3 undergoes planned enhancements, whereas at KEK-CRC, all hardware is replaced every 4 to 5 years.

- New IT room will be available in CC-IN2P3 in 2024
- New KEK CC will be lunch in 2024



The computing centers at CC-N2P3 and KEK-CRC support many high-energy experiments and cosmic ray experiments. Belle II, LHC (ATLAS, ALICE), and T2K are common projects supported by their respective computing centers, requiring coordinated use of computational resources.



## Belle II on the WLCG Dashboard for DC24



- WLCG Data Challenge aims to demonstrate readiness for expected HL-LHC data rates, coordinated by WLCG.
- Belle II DC24: Raw data replication from KEK to Belle II RAW data centers: BNL, UVic, CNAF, DESY, KIT, and **CC-IN2P3**.
- 260 TB / 50K files of pseudo-RAW data have been transferred in 20 hours at an average of 30 Gbps from KEK to RDCs. Belle II HL scenario: 40 TB /day (3.7 Gbps).
- Transfer protocol: Only HTTPS protocol in use, no GridFTP anymore. This is a mandated step for migrating to OIDC (Open ID Connect).
- Most of network traffic load among RDCs is via IPv6.
- This kinds of activities are beneficial in identifying and resolving bottlenecks related to networks, storage systems, and transfer methods.

## FJPL – Japan-France workshop on computing technologies

30 Jan 2024, 09:10 → 31 Jan 2024, 22:50 Europe/Paris

202 (CC-IN2P3)

Sébastien Gadrat (CC-IN2P3)

**Description** The goal of this workshop is to explore relevant technologies, exchange experience and share ideas among experts of both Japan and France organisations in several scientific domains.

This is the 7th edition of this workshop, which is organized annually in the framework and with the sponsorship of the [France-Japan Particle Physics Laboratory](#). The agendas of previous editions are available:

2024: <https://indico.in2p3.fr/event/31887/>

2023: <https://indico.in2p3.fr/event/28953/>

2019: <https://indico.in2p3.fr/event/19919/>

2018: <https://indico.in2p3.fr/event/16922/>

2017: <https://indico.in2p3.fr/event/14157/>

2016: <https://indico.in2p3.fr/event/12701/>

2015: <https://indico.in2p3.fr/event/11289/>

- 11 and 2 people attended the workshop from CC-IN2P3 and KEK-CRC, respectively.
- 11 presentations presented the status of our respective computing centers, and usage, as well as new methods and tools to face problems and scientific challenges.
- The presentations include a lot of topics, for example, status report and future plans on each computing centers, data analysis method by using deep learning, testing method of batch job system and so on.

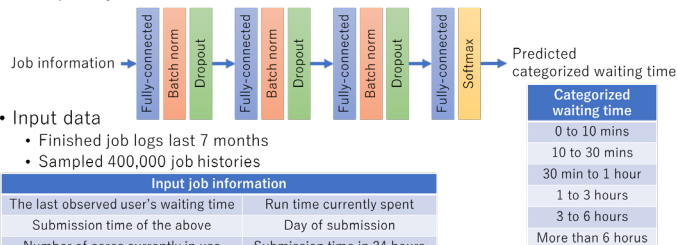


## Application and development of machine learning for computer system

### Waiting time estimation based on simple Deep Learning at KEK-CRC

### Command line tool for end user

- Set up Fully-Connected Neural Network model



- Input data
  - Finished job logs last 7 months
  - Sampled 400,000 job histories

Input job information	
The last observed user's waiting time	Run time currently spent
Submission time of the above	Day of submission
Number of cores currently in use	Submission time in 24 hours
Current CPU utilization efficiency	Queue
Number of current waiting jobs	

Categorized waiting time
0 to 10 mins
10 to 30 mins
30 min to 1 hour
1 to 3 hours
3 to 6 hours
More than 6 hours

- The last observed user's waiting time
- Submission time of the above
- Run time currently spent
- Day of submission
- Number of cores currently in use
- Number of current waiting jobs
- Submission time in 24 hours

W. Takase (KEK-CRC)

```

takase — 53x5
[takase@cw14 ~]$ bpredict -q 1 -u takase
Expected waiting time: 0~599 sec (prob. 92.1123 %)
[takase@cw14 ~]$
    
```

- The effectiveness of the simple Deep Learning for the estimation of **waiting time after the submission of user's jobs to the computing clusters** had been studied.
- A command line tool to provide the current estimation for users was also developed at **KEK-CRC**.
- We confirmed the applicability of the tools developed even for the batch job system utilized in **CC-IN2P3**.
- A detailed study for the waiting time estimation by utilizing several Machine Learning algorithms (not DL) had been done by the members in CC-IN2P3 inspired by the study at KEK-CRC.
- The result was published in 2021; <https://doi.org/10.6084/m9.figshare.13913912.v2>.

### Learning-based Approaches to Estimate Job Wait Time at CC-IN2P3

#### Parameters of jobs/system features

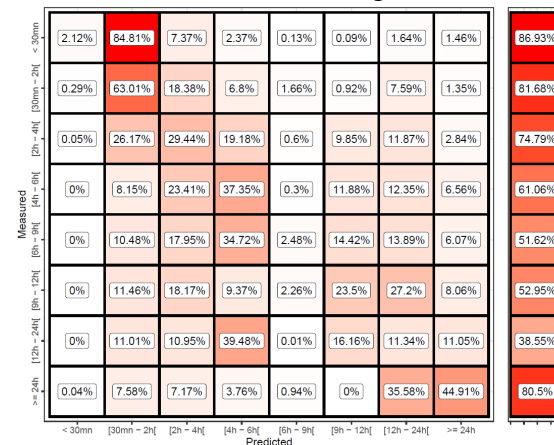
	Name
Who	Job Owner
	User Group
What	Requested Number of Slots
	Requested Storage Subsystem
	Requested Time
	Requested Memory
When	Submission Period
Where	Queue Name
System Configuration	Share Consumption
	RQS Consumption
System State	PCA Components

#### Accuracy of the different classification algorithms

Method	Correct	Adjacent	Combined
Naive Bayes	44.57%	19.55%	64.12%
Ada Boost (Depth 11)	44.81%	29.59%	74.40%
Decision Tree (Depth 9)	49.73%	19.55%	75.19%
Bagging (Depth 9)	47.08%	28.28%	75.36%
Reg. + Clas. (Depth 8)	22.26%	54.65%	76.92%

F. Suter and L. Gombert (CC-INP3)

#### Confusion matrix of Reg. + Clas.

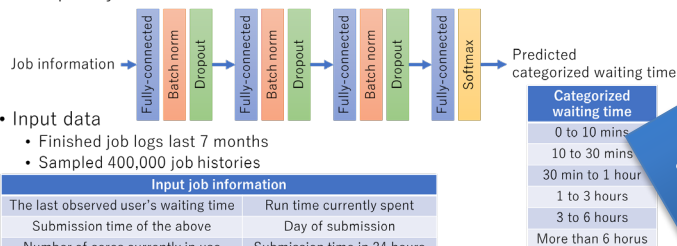


## Application and development of machine learning for computer system

### Waiting time estimation based on simple Deep Learning at KEK-CRC

### Command line tool for end user

- Set up Fully-Connected Neural Network model



- Input data
  - Finished job logs last 7 months
  - Sampled 400,000 job histories

Input job information	
The last observed user's waiting time	Run time currently spent
Submission time of the above	Day of submission
Number of cores currently in use	Submission time in 24 hours
Current CPU utilization efficiency	Queue
Number of current waiting jobs	

2021

- The last observed user's waiting time
- Submission time of the above
- Run time currently spent
- Day of submission
- Number of cores currently in use
- Number of current waiting jobs
- Submission time in 24 hours

W. Takase (KEK-CRC)

```

takase@cw14 ~]$ bpredict -q 1 -u takase
Expected waiting time: 0~599 sec (prob. 92.1123 %)
takase@cw14 ~]$
    
```

- The effectiveness of the simple Deep Learning for the estimation of **waiting time after the submission of user's jobs to the computing clusters** had been studied.
- A command line tool to provide the current estimation for users was also developed at **KEK-CRC**.
- We confirmed the applicability of the tools developed even for the batch job system utilized in **CC-IN2P3**.
- A detailed study for the waiting time estimation by utilizing several Machine Learning algorithms (not DL) had been done by the members in CC-IN2P3 inspired by the study at KEK-CRC.
- The result was published in 2021; <https://doi.org/10.6084/m9.figshare.13913912.v2>.

### Learning-based Approaches to Estimate Job Wait Time at CC-IN2P3

Parameters of jobs/system features

	Name
Who	Job Owner
	User Group
What	Requested Number of Slots
	Requested Storage Subsystem
	Requested Time
	Requested Memory
When	Submission Period
Where	Queue Name
System Configuration	Share Consumption
	RQS Consumption
System State	PCA Components

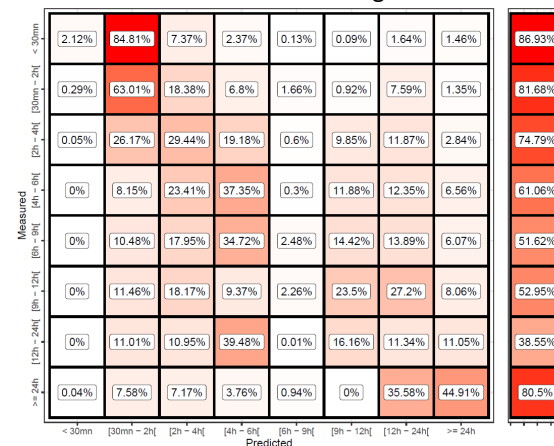
Accuracy of the different classification algorithms

Method	Correct	Adjacent	Combined
Naive Bayes	44.57%	19.57%	34.12%
Ada Boost (Depth 11)	44.81%	19.57%	34.40%
Decision Tree (Depth 9)	40.00%	19.57%	30.00%
Bagging (Depth 9)	40.00%	19.57%	30.00%
Reg. + Clas. (Depth 8)	20.00%	19.57%	76.92%

2022

F. Suter and L. Gumbert (CC-INP3)

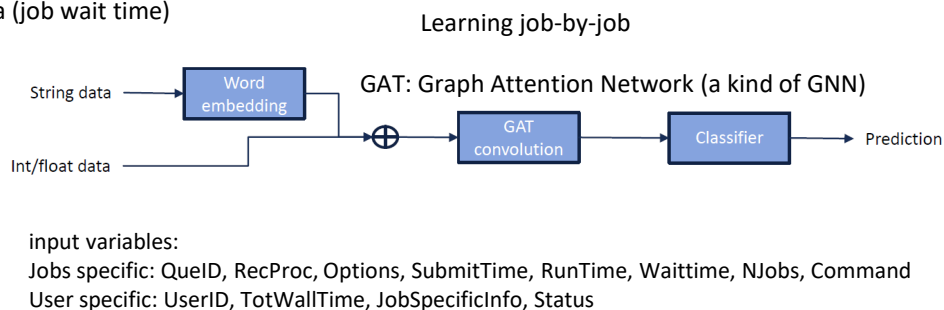
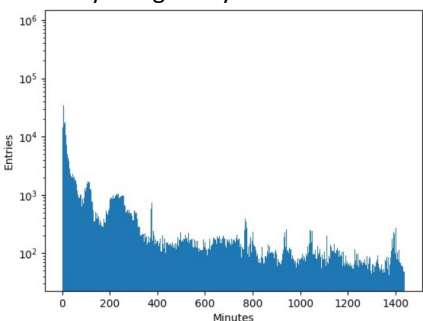
Confusion matrix of Reg. + Clas.



- We had started a further strict investigation on the application of Deep Learning for accounting information by introducing Graph Neural Network at KEK-CRC. It is to identify the critical component by taking the relation of individual jobs into account.
- We tested the equivalent method by using Open Data collected by the different computing facilities in order to find the commonality among the different scheduling system, use case, and workflow.
- Preliminary results were obtained in 2022, and many things remained to optimize the method, for example, scanning hyper-parameters, selection of metrics, and so on.

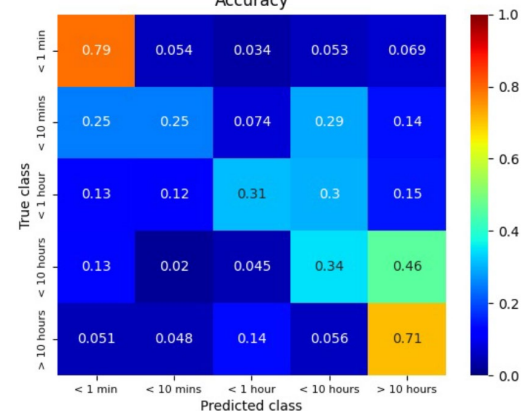
## Applying GAT for individual job information (KEK scheduler: LSF)

Study using one-year measured data (job wait time)



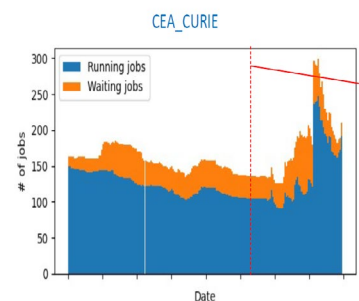
T. Kishimoto (KEK-CRC)

Confusion matrix (KEK): **Work in Progress**



## Testing GNN for the data in different computing facilities (Open Data)

Job information is extracted as table data by taking snapshots

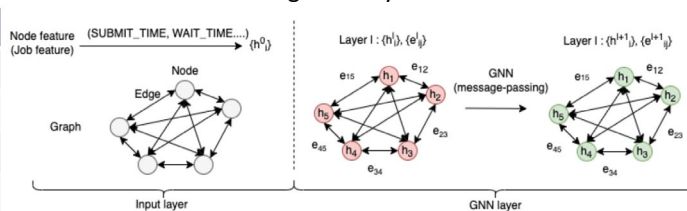


snapshot

Job id	Status	Submit	...	Cores
0	running	xxx	...	6
1	running	yyy	...	128
...	...	...	...	...
1234	waiting	zzz	...	64

- > 1 snapshot = 1 data in deep learning
- > Snapshots are created by 10 minutes step

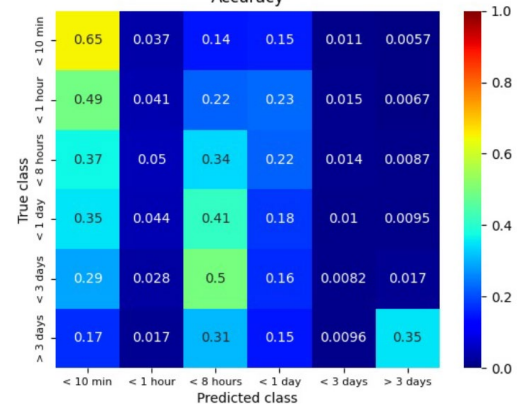
Learning table-by-table



← Open data example (CEA-CURIE)

Tomoaki Nakamura, KEK-CRC

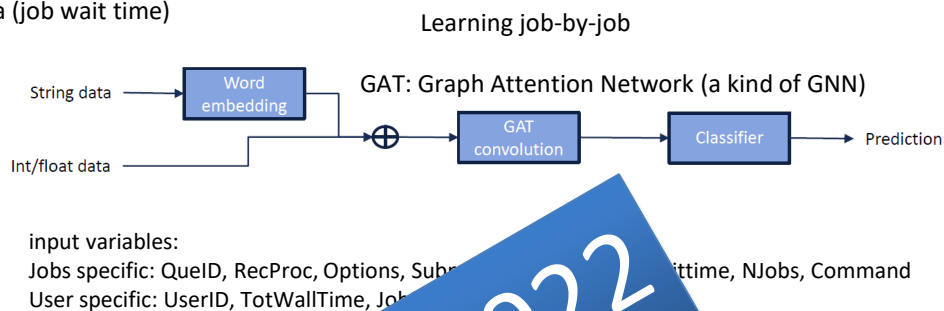
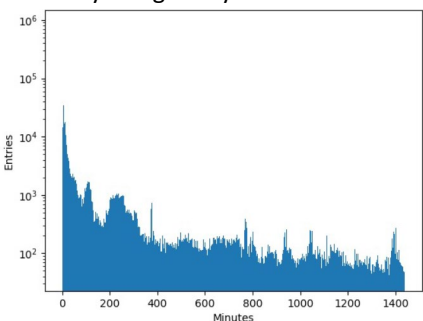
Confusion matrix (CURIE): **Work in Progress**



- We had started a further strict investigation on the application of Deep Learning for accounting information by introducing Graph Neural Network at KEK-CRC. It is to identify the critical component by taking the relation of individual jobs into account.
- We tested the equivalent method by using Open Data collected by the different computing facilities in order to find the commonality among the different scheduling system, use case, and workflow.
- Preliminary results were obtained in 2022, and many things remained to optimize the method, for example, scanning hyper-parameters, selection of metrics, and so on.

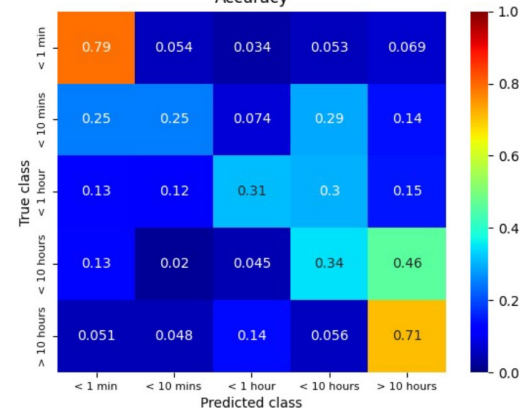
## Applying GAT for individual job information (KEK scheduler: LSF)

Study using one-year measured data (job wait time)



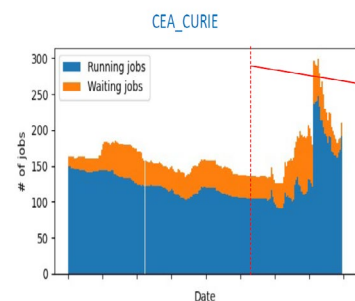
2022

Confusion matrix (KEK): **Work in Progress**



## Testing GNN for the data in different computing facilities (Open Data)

Job information is extracted as table data by taking snapshots

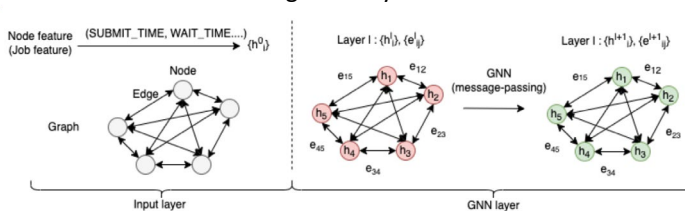


snapshot

Job id	Status	Submit	...	Cores
0	running	xxx	...	6
1	running	yyy	...	128
...	...	...	...	...
1234	waiting	zzz	...	64

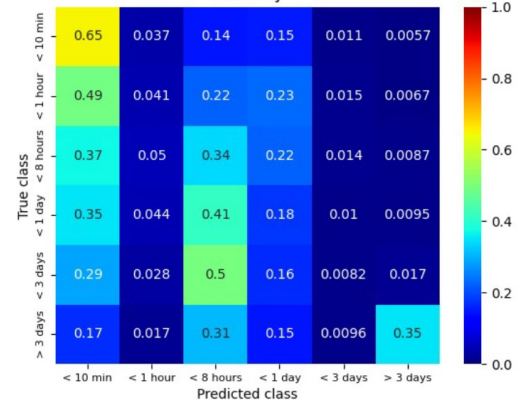
- > 1 snapshot = 1 data in deep learning
- > Snapshots are created by 10 minutes step

Learning table-by-table



← Open data example (CEA-CURIE)

Confusion matrix (CURIE): **Work in Progress**



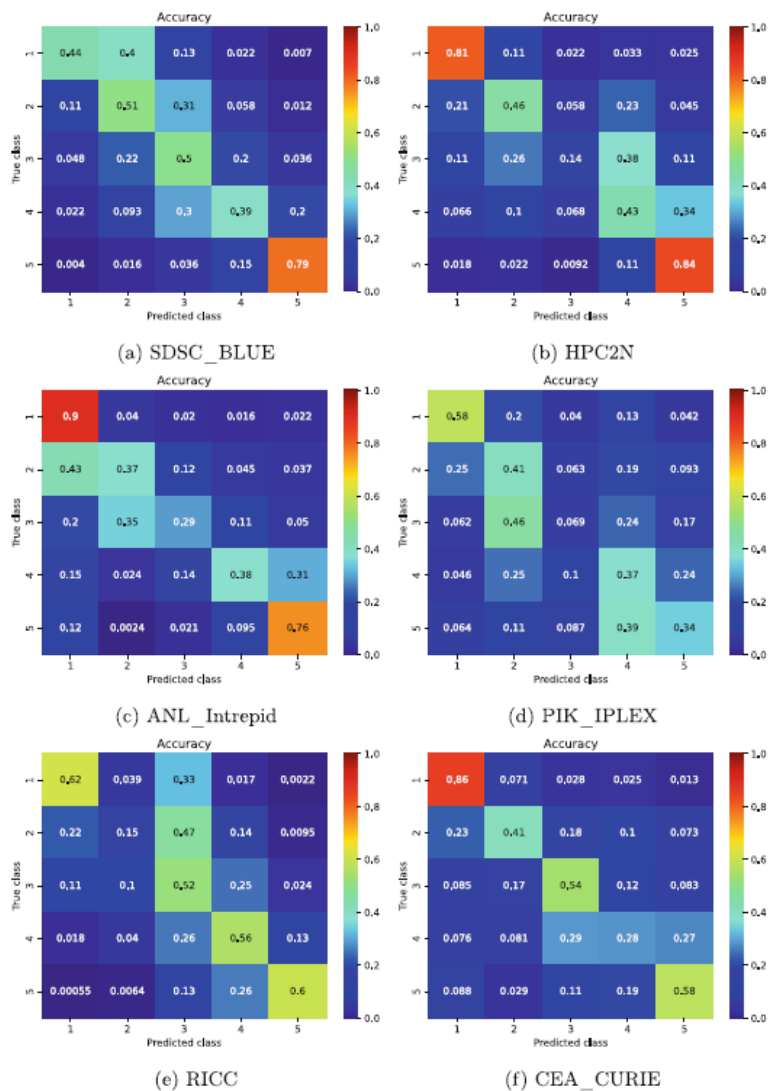


Fig. 4. Confusion matrix of test data for each dataset. The accuracies were obtained by averaging 20 runs with different initial values.

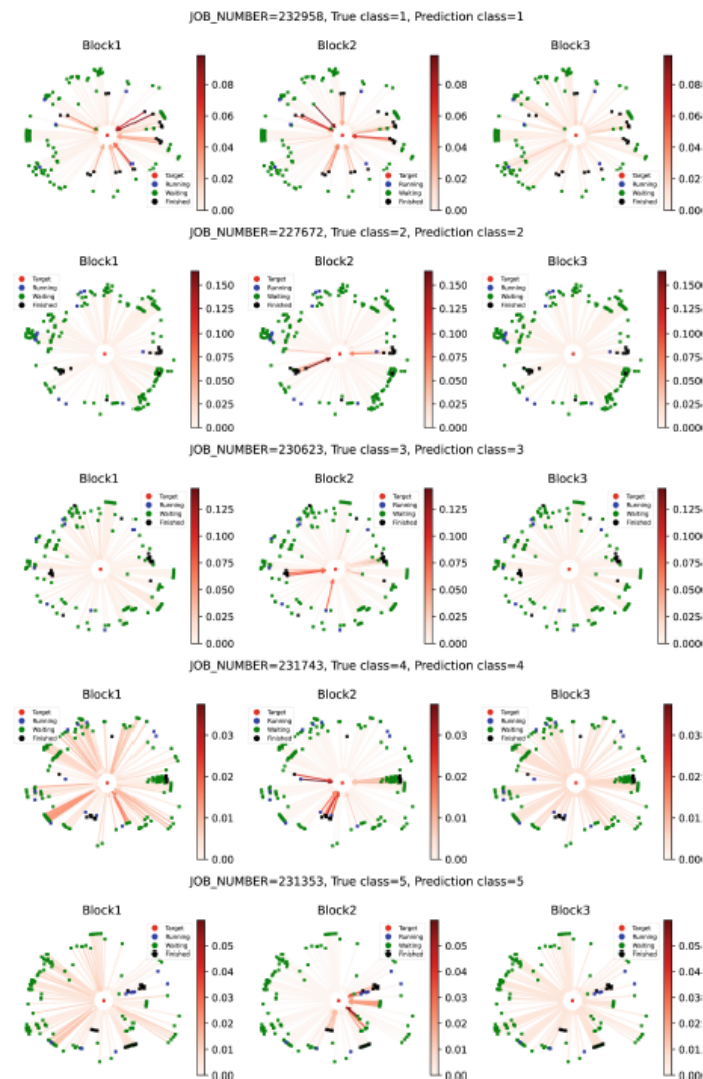
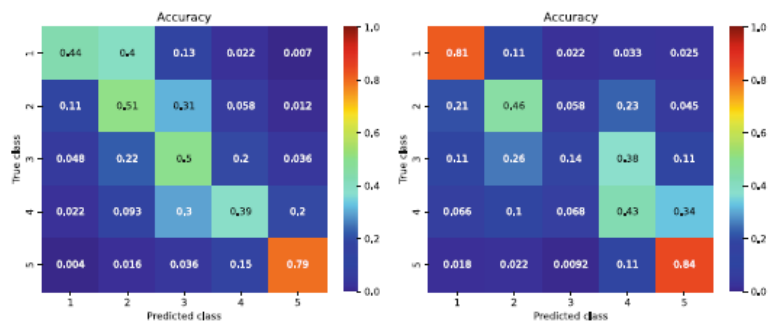
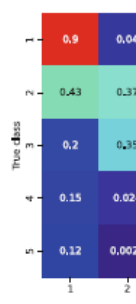


Fig. 7. Attention weights for 5 selected snapshots in the SDSC\_BLUE dataset. The 5 snapshots, one from each prediction class, are shown from top to bottom. The attention weights for 3 blocks in the feature module are shown from left to right for each snapshot. The snapshots were selected from the test data. (Color figure online)



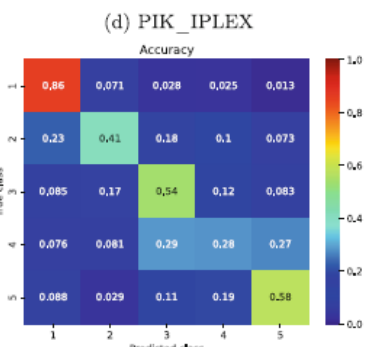
(a) SDSC\_BLUE

(b) HPC2N



(c) ANL\_Intrepid

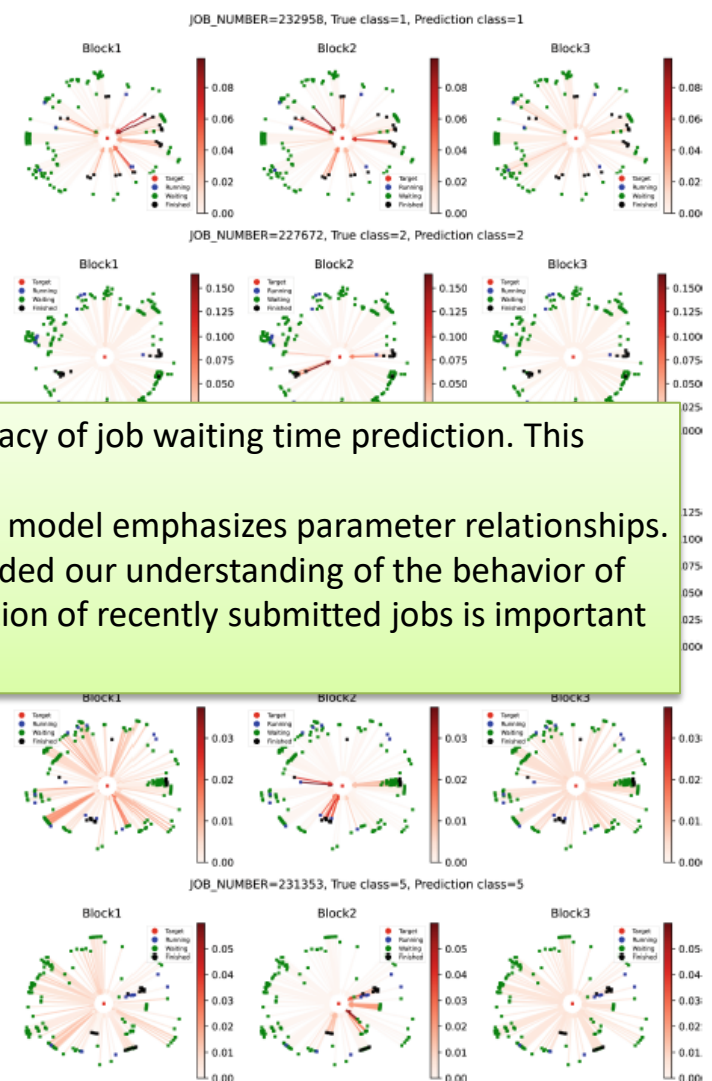
- Further improvements have increased the accuracy of job waiting time prediction. This trend is similar for each HPC data center.
- Deep learning with the graph attention network model emphasizes parameter relationships.
- The visualization of the attention weights expanded our understanding of the behavior of the DL model, which indicated that the information of recently submitted jobs is important for prediction.



(d) PIK\_IPLEX

(e) RICC

(f) CEA\_CURIE



**Fig. 7.** Attention weights for 5 selected snapshots in the SDSC\_BLUE dataset. The 5 snapshots, one from each prediction class, are shown from top to bottom. The attention weights for 3 blocks in the feature module are shown from left to right for each snapshot. The snapshots were selected from the test data. (Color figure online)

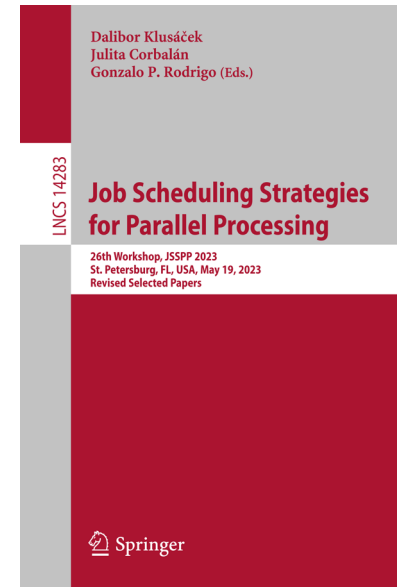
**Fig. 4.** Confusion matrix of test data for each dataset. The accuracies were obtained by averaging 20 runs with different initial values.

2023

## An Efficient Approach Based on Graph Neural Networks for Predicting Wait Time in Job Schedulers

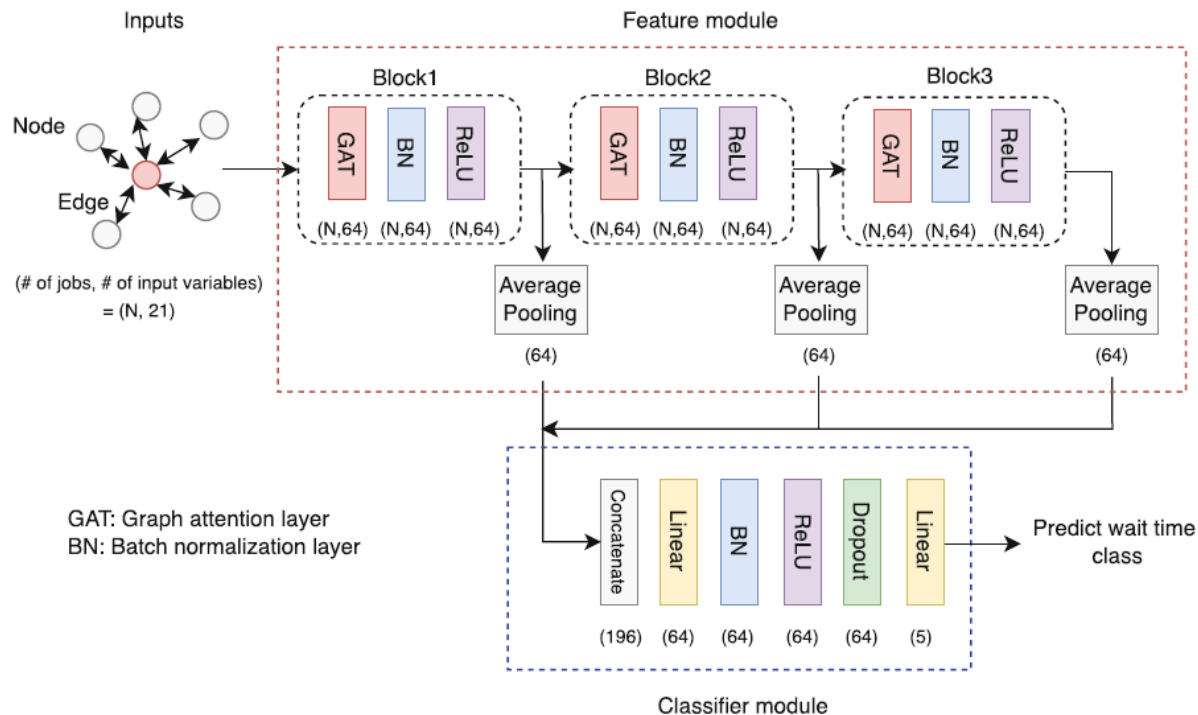
Tomoe Kishimoto<sup>(✉)</sup> and Tomoaki Nakamura

Computing Research Center, High Energy Accelerator Research Organization, KEK,  
Oho 1-1, Tsukuba, Japan  
tomoe.kishimoto@kek.jp



- A graph neural network architecture of deep learning, was employed in this study.
- The challenge lies in the dynamic nature of scheduling systems affected by numerous uncertainties.
- To tackle this, the study employs a graph neural network architecture within deep learning, a novel method for processing job information.
- Experimental results using real historical logs demonstrated that the proposed deep learning model outperformed boosted decision tree and multi-layer perceptron models by 0.3–7.9% in prediction accuracy.
- Extensive analysis, including visualization of attention weights in the graph neural network, enhanced understanding of the model's behavior and improves explainability of experimental results.
- We presented the results of the study at 26th Workshop on Job Scheduling Strategies for Parallel Processing (JSSPP 2023, <https://jsspp.org/>) in conjunction with the 37th IEEE International Parallel and Distributed Processing Symposium (IPDPS2023) on May 19th, 2023, and **a full paper was finally published on September 15th, 2023.**

- According to that study, which we developed several years, we will try to apply our learning model to the accounting information of batch job schedulers within several Grid sites using these research results to identify commonalities that are independent of facilities.
- We are reaching out to several sites that provide computing resources for LHC experiment and Belle II for collaboration, not just CC-IN2P3 and KEK-CRC.



**Fig. 3.** Overview of proposed DL model. The numbers in the brackets indicate output shapes.

- We have been using a Grid authentication infrastructure based on X.509 certificates for over 20 years.
- However, identity and authentication federation using Grid Certification Authorities is undergoing a period of transformation, transitioning from associating IDs obtained from respective institutions to accessing computing resources via access tokens rather than certificates.
- In fact, in Europe, there is the EduGAIN framework, and in Japan, there is the GakuNin framework for authentication federation. These frameworks allow mutual collaboration between them.
- In this situation, we will verify whether we can improve the usability of the IT services provided by CC-IN2P3 and KEK-CRC by utilizing authentication federation, thereby enhancing usability for both parties.

## European ID Federation



## Japanese ID Federation

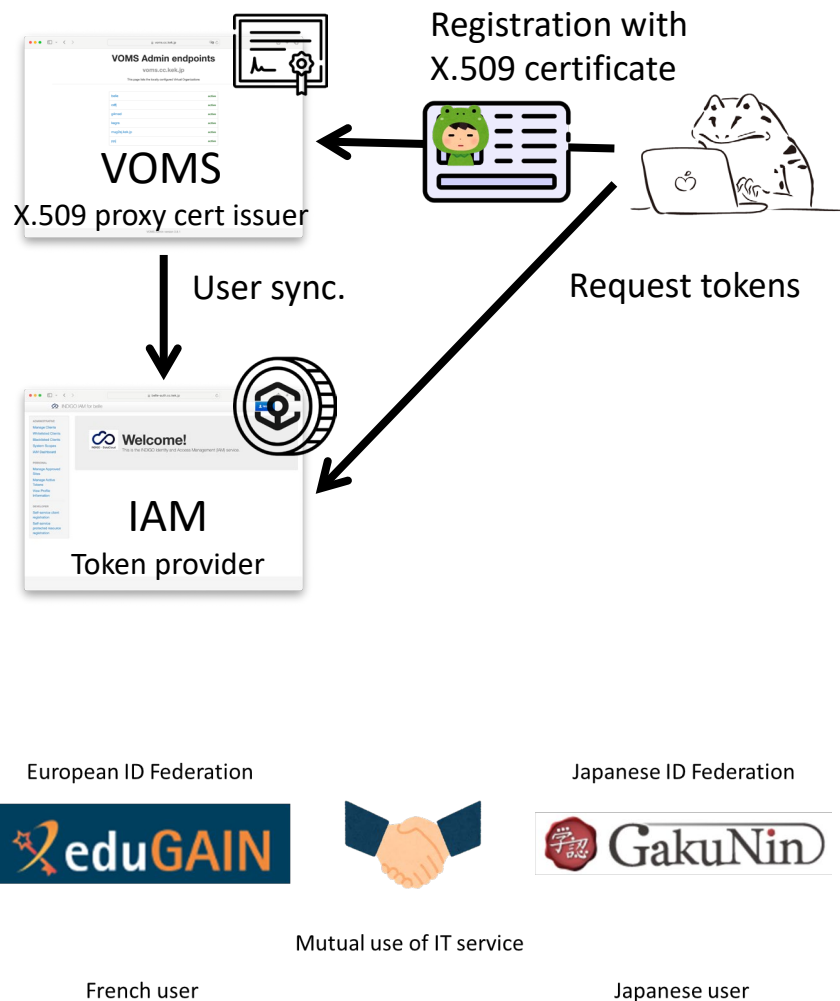


Mutual use of IT service

French user

Japanese user

- IAM (Identity and Access Management) instances have been deployed for Belle II (with limited users).
- Currently, user information is synchronized with VOMS (Virtual Organization Management System).
- Need to establish a registration procedure for IAM without X.509 user certificate after terminating VOMS.
- We think that ID federation is essential because there are many KEK collaborators.
- GakuNin based on Shibboleth is a Japanese academic ID federation.
- GakuNin can federate with eduGAIN.
- We try to deploy a GakuNin IdP based on a trusted DB within KEK. 2024 summer is target.



## Current:

- Can access to Storage Element by tokens.
- Computing Element is not fully available yet.

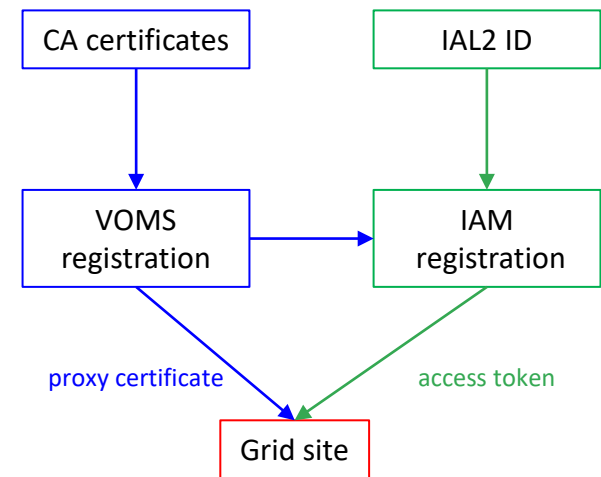
## Future:

- End of X.509 certificate authentication infra.
- proxy certificate will be replaced by token.
- End of VOMS.

## Other background:

- ID Federation is one of the requirements for Research Data Management from Japanese government.
  - Data integrity
  - Open science
  - Open data access
- Identity Assurance Level 2 must be referred to the F2F authentication or human resource DB in each institute (equal level of IGTF).

## Grid Authentication



- COMP\_03: Computing platforms for future experiments
  - FY2015 - FY2018
    - FR leader: Fabio Hernandez
    - JP leader: Tomoaki Nakamura
  
- COMP\_04: Title: Evolution of the computing environment for high-energy and astroparticle experiments
  - FY2019
    - FR leader: Fabio Hernandez
    - JP leader: Tomoaki Nakamura
  - FY2020 - FY2021
    - FR leader: Renaud Vernet
    - JP leader: Tomoaki Nakamura
  - FY2022 - FY2024
    - FR: leader: Sebastien Gadrat
    - JP leader: Tomoaki Nakamura (*final year as a leader*)
  
- COMP\_NEW
  - FY2025
    - FR: leader: Sebastien Gadrat
    - JP leader: Tomoe Kishimoto

## Impact of the Pandemic: 2020 - 2022

- This project has taken a little longer due to the pandemic.
- This year will be my last year as leader of the project.
- We have agreed to start a new computing project in FY2025 with Tomoe Kishimoto as the leader of the Japanese side.