

SummerSOC, 21 June 2025

Engineering Digital Twins of Critical Infrastructures

Victoria Degeler



Digital Twins in Industries & Critical Infrastructures



Tello, A., & Degeler, V. (2022). Digital Twins: An enabler for digital transformation. In Digital Transformation: A Guide for Managers (pp. 176–203). Groningen Digital Business Centre (GDBC), University of Groningen. doi: 10.5281/zenodo.7647493

Ň

Challenges



×X×

DiTEC in a nutshell

Digital Twin for Evolutionary Changes in Water Networks





Š

DiTEC in a nutshell







DiTEC in a nutshell



- Network conditions estimated based on minimal external sensor information
- Changes in the network are reflected in a DT in real-time
- Deep Learning for state estimation









DiTEC in a nutshell



- Network conditions estimated based on minimal external sensor information
- Changes in the network are reflected in a DT in real-time
- Deep Learning for state estimation

Evolutionary Changes in

- Discrepancy detection: leaks, sensor issues, valves
- Alerts and monitoring system
- Reasoning on sensor data
- Logical explanations of the causes of changes

Water Networks





DiTEC in a nutshell



- Network conditions estimated based on minimal external sensor information
- Changes in the network are reflected in a DT in real-time
- Deep Learning for state estimation

Evolutionary Changes in

- Discrepancy detection: leaks, sensor issues, valves
- Alerts and monitoring system
- Reasoning on sensor data
- Logical explanations of the causes of changes

Water Networks

• Water Distribution Network in Oosterbeek, Gelderland





Oosterbeek, Gelderland

Š

Tank T

Problems: Partially Observable Data

Real Sensor Coverage

High installation & maintenance costs

Ideal Sensor Coverage



Š

Problems: Partially Observable Data

Ideal Sensor Coverage Real Sensor Coverage High installation & maintenance costs Tank T

University of Amsterdam

Š

Network state reconstruction: Hydraulic Models



Physical simulation (needs to know correct parameters)

Real network state (cannot be observed directly!)



University of Amsterdam

Ň

Parameters



Ň

Network state reconstruction



Physical simulation (needs to know correct parameters)

Real network state (cannot be observed directly!)





Deep Learning Reconstruction Model (needs to be trained)

×××

Dataset creation



Model architecture - GatRes



H. Truong, A. Tello, A. Lazovik, V. Degeler (2024) Graph Neural Networks for Pressure Estimation in Water Distribution Systems, Water Resources Research, 60, e2023WR036741. doi:10.1029/2023WR036741



University of Amsterdam

Model Training







Ň

Results: Oosterbeek

Table 5. Model comparison in the noisy test performed on 24-hour OosterbeekWDN at 95% masking rate.

Model		$\mathrm{MAE}(\downarrow)$	$MAPE(\downarrow)$	$NSE(\uparrow)$	$Acc(@0.1)(\uparrow)$
GCNii (M. Chen et al., 2020)	0.65	$6.696{\scriptstyle \pm 0.0838}$	$0.2484 {\pm} 0.0552$	-0.1064 ± 0.0266	$36.02{\scriptstyle\pm0.4684}$
GAT (Veličković et al., 2018)	0.35	$4.397{\scriptstyle\pm0.3052}$	$0.2112 {\pm} 0.0767$	0.1490 ± 0.1153	$66.98{\pm}1.6290$
GraphConvWat (Hajgató et al., 2021) GraphConvWat-tuned	0.92 0.23	$\substack{3.611 \pm 0.1234 \\ 2.347 \pm 0.0252}$	$\substack{0.1551 \pm 0.0376 \\ 0.0963 \pm 0.0363}$	$\begin{array}{c} 0.5877 {\scriptstyle \pm 0.0370} \\ 0.749 {\scriptstyle \pm 0.0086} \end{array}$	$\begin{array}{c} 62.99{\scriptstyle\pm1.1600} \\ 81.09{\scriptstyle\pm0.3877} \end{array}$
mGCN (Ashraf et al., 2023)	2.48	$2.188{\scriptstyle\pm0.0558}$	$0.0948 {\pm} 0.0155$	$0.6993 {\pm} 0.0213$	$82.83 {\pm} 0.4199$
GATRes-small (ours) GATRes-large (ours)	$\begin{array}{c} 0.66\\ 1.67\end{array}$	$\frac{1.964 \pm 0.0301}{2.115 \pm 0.0503}$	$\begin{array}{c} 0.0802 {\pm} 0.0458 \\ \textbf{0.0799} {\pm} 0.0207 \end{array}$	$\begin{array}{c} \textbf{0.778} {\scriptstyle \pm 0.0113} \\ 0.7417 {\scriptstyle \pm 0.0140} \end{array}$	$\frac{86.56 \pm 0.2826}{83.43 \pm 0.5044}$

H. Truong, A. Tello, A. Lazovik, V. Degeler (2024) Graph Neural Networks for Pressure Estimation in Water Distribution Systems, Water Resources Research, 60, e2023WR036741. doi:10.1029/2023WR036741

Ň

Results: Oosterbeek



University of Amsterdam

Results: Oosterbeek



gatres-small
hydraulic model
raw sensor



DiTEC dataset

Why do we need new data?

- Publicly available WDN networks are provided as a collection of asset configuration files
- Need to run simulations for full data generation
- Existing data rarely include time-dependent patterns
 - Privacy and safety concerns
 - Few patterns reused many times on several nodes in the network
 - Limits the variety of the data and robustness of models to uncertainties

LeakDB 2018

Leakage Diagnosis Benchmark

the only "ready-to-use" published dataset contains a small Hanoi network of 32 nodes and 4 adjustable parameters

₽ master + ₽	1 Branch 🛇 0 Tags	Q, Go to file	<> Code +	About	
🐣 Mariosmsk mv	LICENSE.md	16c7eb9 - 3 weeks	igo 🕚 27 Commits	LeakDB (Leakage Diagnosis Benchmark) is a realistic leakage dataset for water	
BENCHMARK	fix link datase	rt in bench	3 weeks ago	distribution networks. The dataset is comprised of a large number of artificial	
CCWI-WDSA201	8 Create readD	ata.py	last year	created but realistic leakage scenarios, or different water distribution petworks	
	First Commit		6 years ago	under varying conditions. A scoring	
LICENSE.md	mv LICENSE.n	nd	3 weeks ago	algorithm in MATLAB code is provided to evaluate the results of diff	
C README.md	fix kios logo 8	8. add zenodo link.	3 weeks ago	benchmark dataset leakage	
Kaĩa				Activity Custom properties A 35 store	
K Q 10.5201/am LeakDB	S 15.				
LeakDB (Leakag comprised of a networks, unde different algorit	e Diagnosis Benchmark) is a realistic leakage arge number of artificially created but realist varying conditions. A scoring algorithm in M hms.	dataset for water distribution networks. Ti iic leakage scenarios, on different water dis IATLAB code is provided to evaluate the re	ne dataset is tribution sults of	Activity Custom properties 35 stars 0 7 watching ¥ 18 forks Report repository Releases No releases published Packages No packages published	
LeakDB (Leakag comprised of a networks, unde different algorit How to	e Diagnosis Benchmark) is a realistic leakage large number of artificially created but realist 'varying conditions. A scoring algorithm in N hms. run a use example of	dataset for water distribution networks. Ti tic leakage scenarios, on different water di IATLAB code is provided to evaluate the re the dataset	ne dataset is tribution sults of		
LeakDB (Leakag comprised of a networks, unde different algorit How to Download the fi interface in Mat number of scen	e Diagnosis Benchmark) is a realistic leakage large number of artificially created but realist 'varying conditions. A scoring algorithm in M hms. run a use example of sider LeakD8, set the run path in Matlab to C lab where the user can run and evaluate thej arios on the Hanoi benchmark network.	dataset for water distribution networks. Ti tic leakage scenarios, on different water dis MATLAB code is provided to evaluate the re the dataset CCWI-WDSA2018, and run main.m. This pro provided leakage detection algorithms on	ie dataset is tribution sults of vides a user a limited		



Ň

DiTEC initial dataset

Search records Q Communities My dashboard	+- L vdegeler
Published May 27, 2024 Version v2	510 474
Large-Scale Multipurpose Benchmark Datasets For Assessing Data-Driven Deep Learning Approaches For Water Distribution Networks	VIEWS DOWNLOADS Show more details
Tello, Andrés (Researcher) ¹ ; Truong, Huy (Researcher) ¹ ; Lazovik, Alexander (Supervisor) ¹ ;	Versions
Degeler, Victoria (Supervisor) ² 💿 Show affiliations	Version v2 May 27, 2024 10.5281/zenodo.11353195
Andres Tello*, Huy Truong*, Alexander Lazovik, Victoria Degeler. Large-Scale Multipurpose Benchmark Datasets For Assessing Data-Driven Deep Learning Approaches For Water Distribution Networks. Engineering Proceedings. 2024; 69(1):50. https://doi.org/10.3390/engproc2024069050	Version v1 Apr 15, 2024 10.5281/zenodo.10974087 View all 2 versions
(*) Both authors contributed equally.	Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.10974086. This DOI represents all versions, and will always
Update	resolve to the latest one. Read more.
(04/09/2024): Citation is updated. We have added headers for CSVs and auxiliary data (duration time, edge list, ordered names) in the configuration file (JSON format). As such, corresponding INP files can be omitted when working with this version. The EXN network has been included in this version, so the total number of processed networks is 11. For more details, please read ZENODO_README.md.	External resources Indexed in
Contact	
For dataset-related questions: Huy Truong	
For data acquisition: Andres Tello	Details
If you use this dataset, please cite:	

A. Tello, H. Truong, A. Lazovik, V. Degeler (2024) Large-Scale Multipurpose Benchmark Datasets For Assessing Data-Driven Deep Learning Approaches For Water Distribution Networks. Engineering Proceedings. 2024; 69(1):50. doi: 10.3390/engproc2024069050; Dataset: 10.5281/zenodo.10974086

DiTEC-WDN Large-scale Dataset

• To facilitate research into data driven WDN models on a wide range of tasks, such as:

 $N \times$

• foundation models

UNIVERSITY OF AMSTERDAM

- surrogate modelling
- state estimation
- demand forecasting
- 36 publicly available networks
- 228 million of network state snapshots
- 38 hydraulic parameters are considered



correlation between scenarios in the data generated in LeakDB⁸. The right figure (b) shows the correlation between the scenarios in our dataset. Both matrices include all 1,000 scenarios, each containing 1-year of demand data. The low correlation between scenarios in our dataset shows the diversity of the data, contrary to the similarity observed across LeakDB scenarios.

Ň

DiTEC-WDN dataset

cuongth95 Huy github-actions[bot]

📩 atello

Languages

• Python 100.0%

DITEC_WDN_dataset Public		🖒 Edit Pins 👻 💿 Unwatch	2 ▼ 💡 Fork 0 ▼ 🖧 Star 2			
🐉 main 👻 🐉 2 Branches 🛇 0 Tags	Q Go to file	t Add file + Code +	About			
🧟 cuongth95 docs: Update README.md 🗸		2c430dc · 3 weeks ago 🕥 16 Commits	This repository contains parameter generation, simulation, and encapsulation			
igithub/workflows docs: ci for wiki		4 months ago	free to use it on your "private" WDN!			
📋 ditec_wdn_dataset	refactor: gida to dwd	last month	huggingface.co/datasets/rugds/ditec-w/			
docs	docs: change title, remove GiDa	last month	machine-learning dataset datasets			
🗋 .gitignore	refactor: gida to dwd	last month	h dataset-generation graph-dataset water-distribution-networks			
	Create LICENSE	last month	time-series-forecasting graph-datasets			
🖺 README.md	docs: Update README.md	3 weeks ago	water-network-distribution			
🗋 main.py	refactor: gida to dwd	last month	DD Davidere			
🗋 mkdocs.yml	docs: fix typo	last month	▲ MIT license			
profiler_report.json	refactor: gida to dwd	last month	- Activity			
pyproject.toml	build: fix lib name	last month	 Custom properties 2 stars 			
🗋 requirements.txt	feat:complete clenup	4 months ago	② 2 watching			
			약 0 forks			
C README MIT license		Ø∷≡	Report repository			
			Contributors 3			

DITEC-WDN

This is the official repository for the paper:

DITEC-WDN: A Large-Scale Dataset of Hydraulic Scenarios across Multiple Water Distribution Networks.

This repository contains configuration optimization, scenario generation, and encapsulation code for the DITEC-WDN dataset.

This is useful for individuals or organizations to generate scenarios on their own private Water Distribution Networks.

Those interested in the data can directly refer to the dataset.

Tutorial

Access the wiki at https://ditec-project.github.io/DiTEC_WDN_dataset for more details.

Repo map

😕 Hugging Face 🛛 🔍 Search models, datasets, users...

Models Datasets Spaces Posts Docs Determine Pricing $\sim \equiv$

As you have a valid rug.nl email, you are invited to join GroNLP. Join

📕 Datasets: 💈 rugds / ditec-wdn 🗇 🗢 like 1 🛛 Following 💄 Distributed Systems... 9

Tasks: 📽 Graph Machine Learning 📴 Time Series Forecasting Modalities: 🔲 Tabular Languages: 📵 English Size: 1B - 10B ArXiv: 🗅 arxiv:2503.17167

Tags: water graph distribution network benchmark time-series +1 License: 🏛 cc-by-nc-4.0

💗 Dataset card 🛛 🖽 Data Studio 🐳 📲 Files and versions 🥚 Community 🛐 🔅 Settings

Subset (73) 19PipeSystem_4GB_1Y_link · 35M rows			Split (1) train · 35M rows				~		
Search is not av	ailabl	e for this dataset							
cenario_id loat64	¢	time_id \$	1 float64	0	2 float64	\$	3 float64	\$	4 float64
	999	0 8.76k	-2.95 8	.08	-2.22	285	-1.33	32.9	-1.14
	Θ	Θ	0.165	341		0.104363		0.058629	
	1	Θ	0.166	647		0.105834		0.058567	
	2	0	0.164	107		0.105458		0.058925	
	3	Θ	0.163	892		0.105394		0.058197	
	4	0	0.167	576		0.104086		0.057941	
	5	0	0.173	205		0.104605		0.058093	
	6	0	0.176	138		0.105472		0.057897	
	7	0	0.15	632		105153		0.05787	

Repository: Paper: ditec-project.github.io github.com arxiv.org Update later! Huy Truong

Dismiss this message

8,410

1.74 TB

Number of rows (First 5GB per split): 918,428,160

Downloads last month

∠ Edit dataset card

View full history

Estimated number of rows: 1,648,037,535

Dataset Card for DiTEC-WDN

Dataset Summary

DITEC-WDN consists of 36 Water Distribution Networks (WDNs). Each network has unique 1,000 scenarios with distinct characteristics. Scenario represents a timeseries of undirected shared-topology graphs, referred to as states or snapshots. In terms of graph-ml, it can be seen as a spatiotemporal graph where nodes and edges are multivariate time series.

H. Truong, A. Tello, A. Lazovik, V. Degeler (2025) DiTEC-WDN: A Large-Scale Dataset of Hydraulic Scenarios across Multiple Water Distribution Networks, arxiv preprint: 10.48550/arXiv.2503.17167

Ŵ

Training foundation models



Ŵ

Recreating the pipeline in a wastewater domain



Ň

GatRes results in Heusden (NL)



Š

Adaptive Model Selection



M. H. Nejad Yousefi, V. Degeler and A. Lazovik, "Self-Adaptive Service Selection for Machine Learning Continuous Delivery," 2024 IEEE International Conference on Web Services (ICWS), Shenzhen, China, 2024, pp. 1048-1056, doi: 10.1109/ICWS62655.2024.00123.

Ň

Adaptive Model Selection



Ň

Adaptive Model Selection





PJM Hourly Energy Consumption dataset, Kaggle

Adaptive Model Selection

- Production environment
 - Runs the best available module
 - Given the highest priority in resources
- What-If environment
 - Operates on ephemeral computing environments with fewer resources
 - A sampling sensor data is given
 - Tests a promising sub-set of modules against new conditions.
 - Learning-Based Selection (LBS) is employed to decide which modules are promising.
- Two strategies are employed:
 - Rolling Average
 - KEEP (<u>Keeping Errors down with Enhanced Persistence</u>)





Explainable Sensor Data

Interpret raw sensor data

- Perform profiling of system's behavior
- Explain patterns in a clear manner (e.g. with logical notation)
- Create meaningful rules that describe system's behavior

With a straightforward approach, logical rules would contain:

- Raw sensors and their measurements
- Supporting entities of the system
- Features from the original ML approach This is not very explainable!

Instead: We can use a system ontology.



Ontologies and Knowledge Graphs

Ontology: An ontology provides a formal machine-processable and human-understandable conceptualization of a given domain.

Knowledge Graph: is the data with the ontology as underlying structure.



Ontologies in Digital Twins







Karabulut, E., Pileggi, S. F., Groth, P., & Degeler, V. (2024). Ontologies in Digital Twins: A Systematic Literature Review. Future Generation Computer Systems, vol. 153, pp. 442-456. doi:10.1016/j.future.2023.12.013

Reference DT Architecture



Ranges between task-specific terms to domain-independent application terms

Generic DT concepts, real or abstract/derived digital terms, assets and operations

Protocols, access parameters

Physical entities, actions and processes

Karabulut, E., Pileggi, S. F., Groth, P., & Degeler, V. (2024). Ontologies in Digital Twins: A Systematic Literature Review. Future Generation Computer Systems, vol. 153, pp. 442-456. doi:10.1016/j.future.2023.12.013

Ŵ

Ontologies in Digital Twins



Karabulut, E., Pileggi, S. F., Groth, P., & Degeler, V. (2024). Ontologies in Digital Twins: A Systematic Literature Review. Future Generation Computer Systems, vol. 153, pp. 442-456. doi:10.1016/j.future.2023.12.013 Ň

Ontology Learning with LLMs

Traditional manual methodologies for Ontology Construction are investigated

Ontology Methodologies	Time consuming	Consistency	Structure	Descriptions of Steps	Reusability	Error Avoidance Techniques	Strategies for Identifying Concepts	Interoperability
Development 101	High	Yes	Yes	Some details	Yes	No	Developer's consent	No
METHODOLOGY	High	Yes	Yes	Sufficient details	Yes	No	Middle out strategy	No
KBSI IDEF5	High	Yes	Yes	Sufficient details	Yes	Yes	Not clear	No
KACTUS	Low	Yes	No	Insufficient details	Yes	No	Top-down strategy	No
Cyc	Low	Yes	No	Insufficient details	Yes	No	Not clear	No
ONIONS	High	Yes	Yes	Insufficient details	No	No	Not clear	Yes
SENSUS	Moderate	Yes	Yes	Some details	Yes	No	Bottom-up	Yes
UPON	Moderate	Yes	Yes	Sufficient details	Yes	Yes	Middle-out strategy	No
TOVE	Moderate	Yes	Yes	Some details	Yes	No	User-driven	No
Uschold & King	High	Yes	Yes	Some details	Yes	No	Bottom-up	No
On-To-Knowledge	Moderate	Yes	Yes	Some details	No	Yes	Middle out strategy	No
Gruninger & Fox	Moderate	Yes	Yes	Insufficient details	Yes	No	User-driven	No
CommonKADS	Low	Yes	No	Insufficient details	Yes	No	Not clear	No

UPON Life Ontology Construction



↑ Six Steps to Light Ontology Building

UPON Lite is organized as a sequence of :

Step 1. Domain terminology. The domain

Step 2. Domain glossary. The terms of the

Step 3. Taxonomy.¹¹ Domain terms organi

Step 4. Predication.¹⁷ Terms representing

Step 5. Parthood (meronymy).⁹ Complex (

Step 6. Ontology. This last step produces t

Antonio De Nicola and Michele Missiko. 2016. A lightweight methodology for rapid ontology engineering. Communications of ACM 59, 3 (2016), 79–86.

Retrieval-Augmented Generation (RAG)

. The Network Model

nis chapter discusses how EPANET models the physical objects that constitute a distribution system a ell as its operational parameters. Details about how this information is entered into the program are resented in later chapters. An overview is also given on the computational methods that EPANET uses mulate hydraulic and water quality transport behavior.

.1. Physical Components

EPANET models a water distribution system as a collection of links connected to nodes. The links represent pipes, pumps, and control valves. The nodes represent junctions, tanks, and reservoirs. Fig. 3.1 below illustrates how these objects can be connected to one another to for a network.



Fig. 3.1 Physical Components in a Water Distribution System.

Inctions

Junctions are points in the network where links join together and where water enters or leave

Retrieval-Augmented Generation (RAG) incorporates external documents into the context of a query to mitigate LLM hallucinations

Input document

The Network Model from EPANET



ŝ

WDN Knowledge Graph

ർ

🚼 wdn-knowledge-graph Public

🐉 main 🔹 🐉 1 Branch 🛇 0 Tags	Q Go to file t Add file *	<> Code •			
network and the readme file end of the readme	1f1b03f · 3 weeks ago	🕙 20 Commits			
sample_data	fix bug in pipe node insertion to knowledge graph	3 weeks ago			
tests	create the first version of knowledge graph construction Pyt	last month			
wdn_knowledge_graph	convert kg properties to python types when creating a netw	3 weeks ago			
🕒 .gitignore	add .idea to gitignore	last month			
LICENSE	create the first version of knowledge graph construction Pyt	last month			
MANIFEST.in	recursively include sample data folder content and the sourc	last month			
README.md	ADME.md update the readme file				
requirements.txt	create the first version of knowledge graph construction Pyt	last month			
🗋 setup.py	update the readme file	3 weeks ago			
🗋 wdn_ontology.ttl	add copyright statement to the wdn ontology	3 weeks ago			

README AT MIT license

wdn-knowledge-graph

wdn-knowledge-graph is a Python package designed to convert water distribution network (WDN) data from .inp files (the standard format for the EPANET tool) into RDF knowledge graphs in the .tt1 format using a WDN ontology we created.

The WDN ontology in wdn_ontology.ttl captures the physical components of a WDN, inspired from the EPANET tool.

This package also provides the possibility of converting the final knowledge graph to the popular NetworkX format for further processing, making it easier to analyze and manipulate the water distribution network data.

Table of Contents

- Introduction
- Installation
- Usage

About A Python package to create a knowledge graph for water distribution networks Readme MIT license - Activity Custom properties ☆ 3 stars ② 2 watching & 0 forks Report repository Releases No releases published

Create a new release

Packages

0 :=

🖍 Edit Pins 👻 💿 Unwatch 2 👻 😵 Fork 0 💌 🏠 Star 3

No packages published Publish your first package

Languages

• Python 100.0%

Suggested workflows Based on your tech stack

> Python Package using Configure Anaconda Create and test a Python package on multiple Python versions using Anaconda for package management.

Python application Configure

Create and test a Python application.

dj Django Configure Build and Test a Django Project

Converts EPANET *.inp files into RDF (Resource Description Framework), to work with semantic graph data

Sample output from Pipe_1 of LeakDB



Rule learning from sensor data: finding missing connection pieces with knowledge graphs



Automated environmental rules extraction

No semantics:

if *sensor1* measures *V1*, then *sensor2* measures *V2*

Semantically-enriched:

if a *flow sensor* placed in a *pipe P1* with *length > L1* measures *V1*, then a *pressure sensor* that is placed in a *junction J1* connected to *P1* measures *V2*

Figure 3 Overview of the SOSA classes and properties (observation perspective,

Ŵ

Semantic Rule Learning



Example:

$$f_1 = \{a, b\}, f_2 = \{c, d, e\}, threshold = 0.8$$

$$[1, 0, 0.33, 0.33, 0.33] \quad f_1(a) \to f_2(x)? \quad \text{Trained} \quad \text{Autoencoder} \quad [0.8, 0.2, 0.9, 0.04, 0.06] \quad f_1(a) \to f_2(c)$$

E. Karabulut, V. Degeler, P. Groth "Neurosymbolic Association Rule Mining from Tabular Data", preprint: arXiv:2504.19354

Downstream classification tasks

Table 5: Rule-based interpretable ML models with Aerial+ achieve higher or comparable accuracy with significantly lower execution time. Bold indicates the highest performance.

Dataset	Algorithm	# Rules or Items	Accuracy	Exec. Time (s)
		Exhaustive Aerial+	Exhaustive Aerial+	Exhaustive Aerial+
Congressional	CBA	3437 1495	91.91 92.66	0.34 0.14
Voting	BRL	2547 57	96.97 96.97	15.37 9.69
Records	CORELS	4553 61	96.97 96.97	3.04 0.17
Mushroom	CBA	27800 2785	99.82 99.82	1.75 1.30
	BRL	5093 493	99.87 99.82	244 167
	CORELS	23271 335	90.14 99.04	61 2
Breast Cancer	CBA BRL CORELS	695 601 2047 290 2047 369	66.42 71.13 71.13 71.46 73.69 75.82	0.08 0.28 16.82 14.5 1.42 0.40
Chess	CBA	49775 34490	94.02 93.86	24.31 6.24
	BRL	19312 1518	96.21 95.93	321 119
	CORELS	37104 837	81.1 93.71	106 3.87
Spambase	CBA	125223 33418	84.5 85.42	23.87 7.56
	BRL	37626 5190	72.78 84.93	1169 431
	CORELS	275003 1409	85.37 87.28	1258 5.23



Š

Effect of adding semantic information on rules

Table 5. Comparison of ARM on sensor data with semantics (w-s, our pipeline) and without (wo-s), showing a significant increase in support and rule coverage (cov.) with semantics (FP-G = FP-Growth, Conf = Confidence).

	# Rules	Support	Cov.	Conf.
	w-s wo-s	w-s wo-s	w-s wo-s	w-s wo-s
		LeakDB		
FP-G	103K 9K	0.41 0.19	0.43 0.2	0.95 0.97
Aerial	554 2.5K	0.54 0.25	0.59 0.3	0.91 0.87
		L-Town		
FP-G	25K 5K	0.86 0.36	0.9 0.38	0.96 0.96
Aerial	1K 2.5K	0.59 0.39	0.65 0.45	0.91 0.88
		LBNL		
FP-G	7K 2K	0.84 0.73	0.85 0.75	0.98 0.99
Aerial	73 258	0.74 0.65	0.74 0.66	1.0 0.99

Ŵ

Digital Twin Architecture



V. Degeler, M. Hadadian, E. Karabulut, A. Lazovik, H. van het Loo, A. Tello, H. Truong (2024) DiTEC: Digital Twin for Evolutionary Changes in Water Distribution Networks. Leveraging Applications of Formal Methods, Verification and Validation. Application Areas. ISoLA 2024. Lecture Notes in Computer Science, vol 15223. Springer. doi: <u>10.1007/978-3-031-75390-9_5</u>



Q&A

- V. Degeler, M. Hadadian, E. Karabulut, A. Lazovik, H. van het Loo, A. Tello, H. Truong (2024) DiTEC: Digital Twin for Evolutionary Changes in Water Distribution Networks. Leveraging Applications of Formal Methods, Verification and Validation. Application Areas. ISoLA 2024. Lecture Notes in Computer Science, vol 15223. Springer. doi: 10.1007/978-3-031-75390-9_5
- H. Truong, A. Tello, A. Lazovik, V. Degeler (2024) Graph Neural Networks for Pressure Estimation in Water Distribution Systems, Water Resources Research, 60, e2023WR036741. doi: <u>10.1029/2023WR036741</u>
- H. Truong, A. Tello, A. Lazovik, V. Degeler (2025) DiTEC-WDN: A Large-Scale Dataset of Hydraulic Scenarios across Multiple Water Distribution Networks, arxiv preprint: <u>10.48550/arXiv.2503.17167</u>
- Karabulut, E., Pileggi, S. F., Groth, P., & Degeler, V. (2024). Ontologies in Digital Twins: A Systematic Literature Review. Future Generation Computer Systems, vol. 153, pp. 442-456. doi: <u>10.1016/j.future.2023.12.013</u>
- E. Karabulut, P. Groth, V. Degeler (2025) "Neurosymbolic Association Rule Mining from Tabular Data", preprint: <u>10.48550/arXiv.2504.19354</u>
- M. H. Nejad Yousefi, V. Degeler and A. Lazovik, "Self-Adaptive Service Selection for Machine Learning Continuous Delivery," 2024 IEEE International Conference on Web Services (ICWS), Shenzhen, China, 2024, pp. 1048-1056, doi: <u>10.1109/ICWS62655.2024.00123</u>