C3/23/047 - Retrokit - Optimising Machine Retrofitting with Retrokit: A Modular Machine Learning Approach for Edge-Based Condition Monitoring

A Data Management Plan created using DMPonline.be

Creator: Dries Vanoost

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Principal Investigator: n.n. n.n., Mathias Verbeke, Dries Vanoost, Jonas Lannoo https://orcid.org/0000-0002-6390-3587

Grant number / URL: C3/23/047

ID: 204647

Start date: 01-03-2024

End date: 28-02-2026

Project abstract:

An increasing number of companies is transitioning to Industry 4.0, yet facing an aging fleet of machines. Retrofitting is a cost-effective and non-invasive strategy to make legacy industrial machinery compliant with Industry 4.0 standards by the addition of additional technological features. The M-Group at KU Leuven Bruges Campus has created a condition-monitoring methodology using unsupervised anomaly detection with machine learning for retrofitting applications. This resulting model is about 50% smaller than the state of the art, yet as performant in terms of fault detection accuracy. This reduced size is an essential advantage when compared to competing approaches, as it enables to perform the actual fault detection on Micro-Controller Units (MCU) at the edge. This project aims to develop the Retrokit, a flexible and modular condition monitoring system that encapsulates the resulting methodology in an embedded device. As such, it will include generic components for data collection, processing, and classification with a focus on a minimally invasive implementation through analysis of electrical signal data. The objective is to arrive a modular design that can be connected to fit the specific application. The machine learning models will be designed for multi-scale distributed monitoring, covering various levels of the system being studied, from component to system level. For the hardware-related expertise related to embedded devices, M-Group will closely collaborate with the IoT Lab from VIVES University of Applied Sciences.

Last modified: 09-04-2024

C3/23/047 - Retrokit - Optimising Machine Retrofitting with Retrokit: A Modular Machine Learning Approach for Edge-Based Condition Monitoring

Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		Indicate: ***N**(ew data) or* ***E**(xisting data)*	Indicate: **D**(igital) or **P**(hysical)	Indicate: **A**udiovisual **I**mages **S**ound **N**umerical **T**extual **M**odel **SO**ftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
Case Western Reserve University bearing fault datase	Faulty ball bearing acceleration dataset using motor setup of Case Western Reserve University	Reuse existing data	Digital	Experimental	.mat	<1GB	
Magnetic Equivalent Circuit model motor simulation dataset	Motor state data generated using the Magnetic Equivalent Circuit induction motor model and rotor suspension model Python scripts (IMMEC)	Generate new data	Digital	Simulation data	.npy; Python dictionaries for metadata; .csv	<1TB	
KU Leuven Bruges FMEC lab 1.1 kW motor dataset	1.1 kW induction motor setup with various fault options. Current and voltage measurements, potentially added with acceleration, rotor positional measurements in the future	Reuse existing and generate new data	Digital	Experimental	.csv; .xlsx	<100GB	
Magnetic Equivalent Circuit model	IMMEC: Induction motor magnetic equivalent circuit model Python script with rotor suspension model	Reuse existing and generate new data	Digital	Software	Computational Script	<100MB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

The data named 'Case Western Reserve University bearing fault dataset' can be found in: https://engineering.case.edu/bearingdatacenter/welcome. (Smith WA and Randall RB (2015) Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. Mechanical systems and signal processing 64: 100-131.)

The software named 'IMMEC' can be found in: https://gitlab.kuleuven.be/m-group-campus-brugge/wavecore/phd-philip-desenfans/immec.

The data named 'KU Leuven Bruges FMEC lab 1.1 kWmotor datase' is currently in preparation for publication. The DMP will be updated when the source repository is made publicly available.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- No

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or Data transfer agreements, Research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, codebook.tsv etc. where this information is recorded).

Measurement datasets will be accompanied by metadata which is done in-document for .xlsx-stored measurements or using linked guidance documents which accompany the raw data stored in .csv. Simulation datasets are stored as Python dictionaries for small datasets with (1) metadata added as key names or as metadata string values, (2) as .csv or (3) .npy files with accompanying text-based documents for metadata description. Scripts are explained thoroughly using in-script annotation. Furthermore, notebooks are accompanied by added text segments, including LaTex symbols, linking implementation to theory.

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- Yes

For active data storage, the KU Leuven research storage platform ManGO will be used. There, metadata standards are defined per data type to ensure adequate data description. For measurement data this includes: setup specifications, measurement device specifications, fault state specifications, time of capture and person/people responsible. For simulation data, this includes: simulation setup specifications, hardware specifications, fault state specifications, time of simulation and person/people responsible. Lastly, for scripts and models, this includes: code annotation, theory referencing at the beginning of the script, accompanied by a demo notebook with further explanation.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- ManGO
- OneDrive (KU Leuven)
- Sharepoint online

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

A repository on the shared research platform ManGO is moderated by the M-group. At this point, only four people have access to the data. Further access has to be given by repository manager Dries Vanoost. Inside ManGO, several groups will be defined throughout the project such as read-only groups for students, where access to files can be given on a file-to-file basis or folder-to-folder basis. Additionaly, individual files are shared using OneDrive for business, where the generated links restrict access only to the intended receiver.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

At this point in time, the used 1 TB of data storage on ManGO is free of charge.

Data Preservation after the end of the Research Project

Which data will be retained for 10 years (or longer, in agreement with other retention policies that are applicable) after the end of the project?

In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

- All data will be preserved for 10 years according to KU Leuven RDM policy

Where will these data be archived (stored and curated for the long-term)?

- KU Leuven RDR

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

Depending on the amount of data needed for long-term storage, the 50GB offered free of charge by KUL RDR may not be sufficient. In this case, the RDM helpdesk will be contacted to know the cost for storage increase. The cost of the earlier stated 2TB upper limit can be estimated from the ManGO cost sheet as 70 euros per year.

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?
Please explain per dataset or data type which data will be made available.

- Yes, as restricted data (upon approval, or institutional access only)

If access is restricted, please specify who will be able to access the data and under what conditions.

The Retrokit consortium will grant access to the ManGO project file system to individuals who require the data or models for academic purposes, including students, supervisors, and colleagues with whom we collaborate or on request related to academic output. Unless write permissions to data is necessary, the person will be granted read-only access. Access may be limited to a specific subfolder as per the requirement of the intended use.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

Please explain per dataset or data type where appropriate.

- No

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- KU Leuven RDR (Research Data Repository)

When will the data be made available?

- Upon publication of research results

Which data usage licenses are you going to provide?

If none, please explain why.

- GNU GPL-3.0 (code)
- CC-BY 4.0 (data)

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- Yes, a PID will be added upon deposit in a data repository

What are the expected costs for data sharing? How will these costs be covered?

The costs for data sharing in KUL RDR are free of charge up to 50GB.

Responsibilities

Who will manage data documentation and metadata during the research project?

The Retrokit consortium. Those who upload new data are initially responsible for providing data documentation of their data.

Who will manage data storage and backup during the research project?

The Retrokit consortium

Who will manage data preservation and sharing?

The Retrokit consortium

Who will update and implement this DMP?

The Retrokit consortium