DOI 10.14622/Advances_48_2022_21

Multivariate sensor dataset of an industrial rotogravure printing press (MSDIRPP)

Tobias Enk and Peter Urban

University of Wuppertal, School of Electrical, Information and Media Engineering,
Rainer-Gruenter-Straße 21, D-42119 Wuppertal
E-mails: toenk@uni-wuppertal.de; purban@uni-wuppertal.de

Short abstract

We present a multivariate sensor dataset for machine learning research in context of industrial print application. The dataset contains 7 608 rolls of pre-processed multivariate sensor data of a single production scale rotogravure printing press. The data volume corresponds to 43 181 km of printed cardboard and paper. For each roll, we provide high-resolution sampled inline sensor data, machine condition labels and several meta information. Besides basic information like machine speed the dataset contains web movement data such as web edge and web tension measurements, material measurement like web moisture and print quality data such as register measurements in cross and machine direction for 11 print units. We publish the dataset to provide data researchers a strong baseline dataset for several machine learning applications in industrial printing. The dataset is publicly available under Creative Common Licence CC BY 4.0 International.

Keywords: multivariate time series, industrial rotogravure printing, machine learning, web run, register quality

1. Introduction and background

A lack of access to industrial data is just one of many problems that stand in the way of successfully implementing machine-learning projects for industrial applications. The acquisition and processing of industrial data is a particularly time-consuming and tedious task for data scientists.

In the field of industrial printing, only a few major case studies from practice exist, where a larger amount of production data were gathered to investigate specific research questions. These datasets in most cases are not available to other researchers.

Examples of <u>not</u> publicly available datasets:
Alzghoul, et al. (2009) investigate web breaks and runnability with soft computing techniques using a dataset, which is divided into paper, winding and press related parameters.

Parola, et al. investigate the runnability and web widening in a larger dataset gathered in printing plants (Parola, et al., 2003; Paukku, Parola and Vuorinen, 2004).

Even fewer datasets are publicly accessible in an industrial printing related context, where other researchers can apply, test or develop existing or new machine learning approaches in this field.

Examples of publicly available datasets:
Bob Evans published the “Cylinder Band Data Set”, which is a multivariate categorical dataset for classification task in rotogravure printing application, focussing on banding effects (Evans, 1995).

Chitta Ranjan provides a multivariate time series dataset from a paper manufacturing process for data-driven web break analysis, which is very close to a printing industry (Ranjan, et al., 2018).

Pauline Brum provides a collection of digitized gravure printing samples as an image dataset from a lab-scale rotogravure press for the analysis of hydrodynamic pattern formation (Brumm, et al., 2021).

We provide a large public available multivariate inline measurement dataset collected from a single industrial rotogravure printing press for package printing over two years of production (Enk, 2022). It contains 7608 printed rolls of mainly cardboard material including rich annotations such as meta information and labels of the machine and sensor states. Sensor data were provided as machine speed independent, distance discrete multivariate datasets roll by roll.

The dataset can be used as a baseline dataset for various machine-learning approaches on the gravure printing domain, such as

- Rare event estimation
- Print quality estimation
- Cluster analysis
- Classification tasks
- Time series forecasting
- Feature extraction tasks
- Synthetic data generation
- Pre-training tasks

From a printing domain perspective, potential research topics could be

- Web runnability research
- Web movement research
- Register quality research
- Winding defect research
- Web tension dynamic research

2. Dataset structure

We provide the dataset in a simple structure containing a metadata file 'metadata.csv' with anonymized meta information for each roll and the printed job. The link to each signal file is the roll-id. Accordingly, we named the roll specific sensor files by their roll_id, exemplarily '10001.csv' for roll_id 10001. Figure 1 demonstrates the dataset structure.

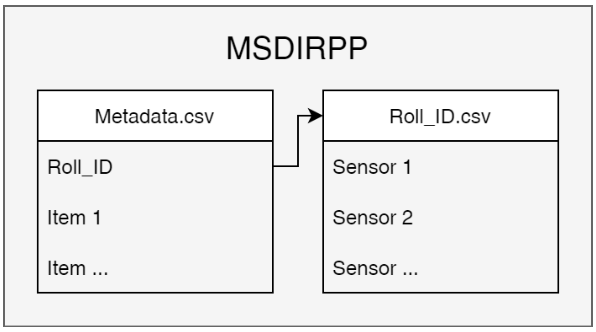

Figure 1: Dataset structure of the MSDIRPP dataset

A list of abbreviations used in MSDIRPP is shown in Table 1.

Table 1: Abbreviations used in MSDIRPP

Abbreviation	Name
CD	Cross Direction
MD	Machine Direction
OS	Operating Side
DS	Drive Side
UW	Unwinder
RW	Rewinder
IPU	Infeed Pull Unit
PU	Print Unit
EU	Embossing Unit
CC	Cross Cut Unit
WGS	Web Guiding System

An overview of columns in the metadata file and the signal files are given in Table 2 and Table 3. Missing values are marked as '−1' in metadata file. In the signal files missing values are marked as 'nan'.

Table 2: Metadata column description

Column name	Unit	Type	Valid within each	Data source	Values
Roll_ID	-	numerical	-	data recorder	[10001, (...), 17608]
SignalFile	-	textual	-	data recorder	[10001.csv, (...), 17608.csv]
TimeStamp	-	temporal	-	data recorder	[value]
Date	-	temporal	-	data recorder	[value]
Start_Time	-	temporal	-	data recorder	[value]
End_Time	-	temporal	-	data recorder	[value]
Roll_Nr	-	categorical	Batch_Nr	extern metadata	[−1, 1, 2, (...)]
Roll_CD_Position	-	categorical	Supplier	extern metadata	[−1, 1, 2, (...)]
Batch_Nr	-	categorical	-	extern metadata	[−1, 1, 2, (...), 246]
Material	-	categorical	-	extern metadata	[−1, C01, C02, C03, GC1, GC2, GZ1]
Web_Width_[mm]	mm	numerical	-	extern metadata	[−1, value]
Grammage_[g/qm]	g/m^2	numerical	-	extern metadata	[−1, value]
Factory	-	categorical	-	extern metadata	[−1, 1, 2, (...), 13]
Supplier	-	categorical	-	extern metadata	[−1, 1, 2, (...), 9]
Converting	-	categorical	-	extern metadata	[−1, 1(Sheet), 2(Roll)]
Order_Nr	-	categorical	-	extern metadata	[−1, 1, 2, (...), 634]
WebGuiding_Ref_Edge	-	categorical	-	extern metadata	[−1, 1(OS), 2(DS), 3(Center)]
Unwinding_Shaft	-	categorical	-	extern metadata	[−1, 1, 2]
Cylinder_ Circumference_[mm]	mm	numerical	-	signal files	[−1, value]
Roll_UW_Core_ Diameter_[mm]	mm	numerical	-	signal files	[−1, value]
Caliper_[mm]	mm	numerical	-	signal files	[−1, value]
Steady_State_Data_[m]	m	numerical	-	signal files	[−1, value]

Column name	Unit	Type	Valid within each	Data source	Values
Real_Run_Length_[m]	m	numerical	-	signal files	[−1, value]
State_Reg_PU02	-	logical	-	signal files	[−1, 0, 1]
State_Reg_PU03	-	logical	-	signal files	[−1, 0, 1]
State_Reg_PU04	-	logical	-	signal files	[−1, 0, 1]
State_Reg_PU05	-	logical	-	signal files	[−1, 0, 1]
State_Reg_PU06	-	logical	-	signal files	[−1, 0, 1]
State_Reg_PU07	-	logical	-	signal files	[−1, 0, 1]
State_Reg_PU08	-	logical	-	signal files	[−1, 0, 1]
State_Reg_PU09	-	logical	-	signal files	[−1, 0, 1]
State_Reg_PU10	-	logical	-	signal files	[−1, 0, 1]
State_Reg_PU11	-	logical	-	signal files	[−1, 0, 1]
State_Reg_PU12	-	logical	-	signal files	[−1, 0, 1]
State_Reg_EU01	-	logical	-	signal files	[−1, 0, 1]
State_Reg_EU02	-	logical	-	signal files	[−1, 0, 1]

Below are some additional descriptions of the column names:

- Roll_Nr: sequential number of produced roll (for each batch)
- Roll_CD_Position: CD position on original mother roll (supplier specific)
- Batch_Nr: categorical material batch number
- Converting: converting to sheets or roll after printing
- WebGuiding_Ref_Edge: reference edge using as input for web guiding
- Unwinding_Shaft: used shaft to unwind the roll
- Cylinder_Circumference_[mm]: printing cylinder circumference in mm
- Roll_UW_Core_Diameter_[mm]: diameter of roll core in mm
- Steady_State_Data_[m]: total length of steady state production
- Real_Run_Length_[m]: total unwound material length in m

Table 3: Signal file column description

Column name	x/y	Type	Unit	Description
Length	x	numerical	m	processed material length
[state_roll_inner_part]_L	y	logical	-	roll length minus first and last 200 m
[state_tension_UW]_L	y	logical	-	condition web tension at UW
[state_tension_IPU]_L	y	logical	-	condition web tension at IPU
[state_tension_PU7]_L	y	logical	-	condition web tension at PU7
[state_tension_PU12]_L	y	logical	-	condition web tension at PU12
[state_tension_EU1]_L	y	logical	-	condition web tension at EU1
[state_tension_CC]_L	y	logical	-	condition web tension at CC
[state_machine_speed]_L	y	logical	-	condition machine speed
[state_waste]_L	y	logical	-	unsaleble products marked by machine operator
[state_machine]_L	y	logical	-	steady state machie condition
[state_register_PU2]_L	y	logical	-	activity of register measurement at PU2
...	y	logical	-	activity of register measurement at PU3-12
[state_register_EU1]_L	y	logical	-	activity of register measurement at EU1

Column name	x/y	Type	Unit	Description
[state_register_EU2]_L	y	logical	-	activity of register measurement at EU2
[Web_Speed]_L	y	numerical	m/min	web speed
[WebGuiding_Motorposition]_L	y	numerical	-	current position of WGS
[WebGuiding_Center]_L	y	numerical	mm	center position of web
[WebGuiding_Edge1]_L	y	numerical	mm	web edge measured at WGS (OS)
[WebGuiding_Edge2]_L	y	numerical	mm	web edge measured at WGS (DS)
[Roll_UW_Rest]_L	y	numerical	m	rest material length on core at UW
[Roll_UW_Diameter]_L	y	numerical	mm	roll diameter at UW
[Caliper]_L	y	numerical	mm	material thickness/caliper
[Reg_PU2_CD]_L	y	numerical	mm	CD misregister at PU2
...	y	numerical	mm	CD misregister at PU3-12
[Reg_EU1_CD]_L	y	numerical	mm	CD misregister at EU1
[Reg_EU2_CD]_L	y	numerical	mm	CD misregister at EU2
[Reg_PU2_MD]_L	y	numerical	mm	MD misregister at PU2
...	y	numerical	mm	MD misregister at PU3-12
[Reg_EU1_MD]_L	y	numerical	mm	MD misregister at EU1
[Reg_EU2_MD]_L	y	numerical	mm	MD misregister at EU2
[WebEdge_UW]_L	y	numerical	mm	web edge measured at UW (DS)
[WebEdge_PU2]_L	y	numerical	mm	web edge measured at PU2 (DS)
[WebEdge_PU7]_L	y	numerical	mm	web edge measured at PU7 (DS)
[WebEdge_PU10]_L	y	numerical	mm	web edge measured at PU10 (DS)
[WebEdge_PU11]_L	y	numerical	mm	web edge measured at PU11 (DS)
[Dancer_UW]_L	y	numerical	%	dancer movement after UW
[Dancer_RW]_L	y	numerical	%	dancer movement before RW
[WebTension_UW]_L	y	numerical	N	web tension after UW
[WebTension_InfeedPullUnit]_L	y	numerical	N	web tension at IPU
[WebTension_PU7]_L	y	numerical	N	web tension at PU7
[WebTension_PU12]_L	y	numerical	N	web tension at PU12
[WebTension_EU1]_L	y	numerical	N	web tension at EU1
[WebTension_CC]_L	y	numerical	N	web tension at CC
[WebMoisture]_L	y	numerical	%	relative web moisture direct after UW

3. Materials and methods

3.1 Printing press

Our dataset was gathered from a 178 m long production scale roll-to-roll mechanical line shaft rotogravure press with 11 printing units (PU) and two embossing units (EU). The press consists of an unwinder (UW), a tension control system, a web guiding system (WGS) for compensation of lateral web displacements, followed by the infeed pull unit (IPU). In the 134 m long printing section are 11 equal PUs. The press can optionally convert printed web to roll or sheet. Solvent-based inks are thermal dried directly after each PU. Misregister is measured once per print length (once per printing cylinder turn) by observing printed triangle marks. The first printed colour is used as a reference for misregister calculation. These displacement values are then used for register control, which is realised by adjusting the print cylinders in CD and changing the web length in MD before each PU. The press is equipped with several sensors see Figure 2.

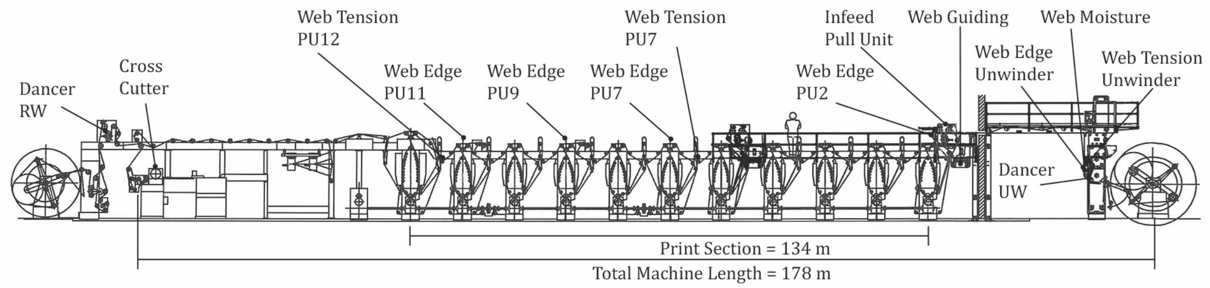

Figure 2: Production scale rotogravure printing press with some sensor positions

3.2 Data acquisition

All measurements are recorded by a process data acquisition system with a sampling rate of 100 Hz, independent of the real sensor specific sampling rate. Therefore, we fill missing values with the last recorded measurement value. Details of all inline measurements are given in Table 4.

Table 4: Inline measurement details

Nr	Measurement	Unit	MD Pos.	CD Pos.	Objective	Unit	Resolution	Sampling	Sensor
1	Web Moisture	UW	0 m	center	rel. moisture	%	-	> 100 Hz	infrared
2	Web Speed	Drive	0 m	-	speed	m/min	-	> 100 Hz	encoder
3	Roll Diameter	UW	0 m	center	distance	mm	-	> 100 Hz	laser distance
4	Web Tension	UW	2 m	CD	force	N	-	> 100 Hz	strain gauge
5	Web Tension	IPU	19 m	CD	force	N	-	> 100 Hz	strain gauge
6	Web Tension	PU7	78 m	CD	force	N	-	> 100 Hz	strain gauge
7	Web Tension	PU12	137 m	CD	force	N	-	> 100 Hz	strain gauge
8	Web Tension	EU	154 m	CD	force	N	-	> 100 Hz	strain gauge
9	Web Tension	CC	178 m	CD	force	N	-	> 100 Hz	strain gauge
10	Dancer Movement	UW	2 m	CD	movement	%	-	> 100 Hz	potentiometer
11	Dancer Movement	RW	170 m	CD	movement	%	-	> 100 Hz	potentiometer
12	Web Edge	UW	2 m	DS	position	mm	0.005 mm	500 Hz	CCD
13	Web Edge 1 + 2	WGS	18 m	OS + DS	position	mm	0.001 mm	500 Hz	infrared
14	Web Edge	PU2	20 m	DS	position	mm	0.005 mm	500 Hz	CCD
15	Web Edge	PU7	88 m	DS	position	mm	0.005 mm	500 Hz	CCD
16	Web Edge	PU10	112 m	DS	position	mm	0.005 mm	500 Hz	CCD
17	Web Edge	PU11	135 m	DS	position	mm	0.005 mm	500 Hz	CCD
18, 19	Register CD + MD	PU2	23 m	OS	misregister	mm	0.005 mm	1/cyl. turn	contrast
20, 21	Register CD + MD	PU3	35 m	OS	misregister	mm	0.005 mm	1/cyl. turn	contrast

Nr	Measurement	Unit	MD Pos.	CD Pos.	Objective	Unit	Resolution	Sampling	Sensor
22, 23	Register CD + MD	PU4	46 m	OS	misregister	mm	0.005 mm	1/cyl. turn	contrast
24, 25	Register CD + MD	PU5	59 m	OS	misregister	mm	0.005 mm	1/cyl. turn	contrast
26, 27	Register CD + MD	PU6	70 m	OS	misregister	mm	0.005 mm	1/cyl. turn	contrast
28, 29	Register CD + MD	PU7	81 m	OS	misregister	mm	0.005 mm	1/cyl. turn	contrast
30, 31	Register CD + MD	PU8	93 m	OS	misregister	mm	0.005 mm	1/cyl. turn	contrast
32, 33	Register CD + MD	PU9	105 m	OS	misregister	mm	0.005 mm	1/cyl. turn	contrast
34, 35	Register CD + MD	PU10	117 m	OS	misregister	mm	0.005 mm	1/cyl. turn	contrast
36, 37	Register CD + MD	PU11	129 m	OS	misregister	mm	0.005 mm	1/cyl. turn	contrast
38, 39	Register CD + MD	PU12	141 m	OS	misregister	mm	0.005 mm	1/cyl. turn	contrast
40, 41	Register CD + MD	EU1	157 m	OS	misregister	mm	0.005 mm	1/cyl. turn	contrast
42, 43	Register CD + MD	EU2	157 m	OS	misregister	mm	0.005 mm	1/cyl. turn	contrast

3.3 Pre-processing

Due to the special nature of sequentially ordered sub-processes in a web-processing system such as a printing press, process and quality measurement data are generated at different points in the press with different time stamps. To enable locally resolved analysis of all process and quality data over the web material independent of its original process speed, we pre-process the collected time series in three steps.

1. The continuous data stream coming from the data acquisition system is truncated, when a roll is changed at the unwinder and accumulated into a single multivariate time series raw data file containing all inline measurements for each roll.
2. For a web speed independent analysis, the signal data $x_i(t)$ of each sensor s_i were aligned via their specific time constant τ_i, according to the sensors position in the printing machine l_i relative to the unwinder (UW) and the web speed v (see Figure 3). Where $\hat{t}_i$ is the sensor specific aligned time axis. The individual web stretch is not considered.

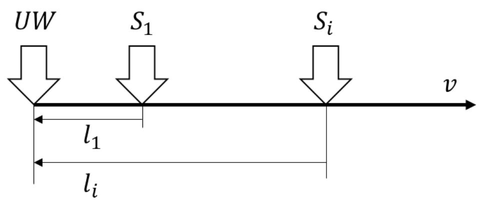

Figure 3: Time axis alignment relative to unwinder in printing machine

$$\tau_i = \frac{l_i}{v} \quad [1]$$

$$\hat{t}_i = t_i - \tau_i \quad [2]$$

3. Time series are resampled to a new uniform distance discrete base. We choose a sampling rate of 10 m^{-1} as a trade-off between data volume and resolution. By considering the Nyquist-Shannon sampling theorem, we think it is a sufficient value, because most relevant process fluctuations are considered to be slower than 5 m^{-1}, which corresponds to a wavelength of 0.2 m.

3.4 Machine condition labeling

For statistical evaluations, it can be useful to distinguish steady state machine condition data from unsteady production data. For this reason, we derive the boolean label signals 'state_machine' and 'state_register' from sensor signals. Where 'state_machine' marks steady state machine condition data and 'state_register' the activity of PUs. We use the web speed and web tension data as indicator signals to determine the machine condition. Both has to be in a specific value range, which indicates the machine runs in a stable and continuous production mode and not in a transient or setup mode. Further, we use standard deviation thresholds for indication of process parameter changes or sensor activity. In Table 5, all used thresholds are given to check machine condition or activity of indicator signals.

Table 5: Machine condition labelling thresholds

Activity	min value	max value	std dev threshold
check web speed steady state	50	800	10.00
check web tension steady state	100	3 000	100.00
check register activity	-	-	0.01

Condition state labels were calculated with a moving window of 200 m width and a movement of 0.1 m. To consider transition periods before and after a significant process change we keep the label centred over the signal by shifting the calculated label signal back in time with the half of the window width. Further the boolean indicator signal 'state_roll_inner_part' marks the entire roll except the first and last 200 m to ignore effects from a roll change. The final 'state_machine' label signal is the intersection of all indicator label signals. Steady state machine condition labelled data take account for approximately 87 % of the total dataset. The boolean 'state_waste' label is a manual signal, which is generated by the machine operator directly at the printing press to mark unsaleable products.

3.5 Data exploration

Our dataset contains 7 608 rolls of printed cardboard and paper, which corresponds to 43 181 km of printed web material. The dataset contains mainly multi-ply and coated cardboard materials in a grammage range of 210 g/m^2 – 240 g/m^2, next to some more lightweight materials, see Table 6.

Table 6: Material information

Material	Description	Rolls	Total run length [km]	Steady state [km]	Roll length median [m]	Grammage median [g/m²]	Web width median [mm]	Caliper median [mm]
GC1	Coated FBB*, white back	3 879	20 458	18 065	5 352	220	743	0.316
GC2	Coated FBB*, bright back	44	228	201	5 218	210	705	0.330
GZ1	Coated SBB*	3 151	18 300	15 720	5 654	240	835	0.291
C01	Custom specific	12	172	165	14 290	70	757	0.052
C02	Custom specific	135	2 174	2 031	16 387	80	1 015	0.073
C03	Custom specific	29	222	208	7 707	−1	−1	0.148
−1	no information	358	1 627	1 328	4 933	−1	−1	0.336
Total		7 608	43 181	37 718				

* according to DIN 19303

The distributions of some dataset aspects are shown in the following plots. Figure 4 shows the dominance of some material suppliers. Figure 5 shows the preferred use of late PUs for jobs with less than 11 colors. There is no PU1 in the printing machine, so first possible printing unit is PU2. This means first register measurement can be made in PU3.

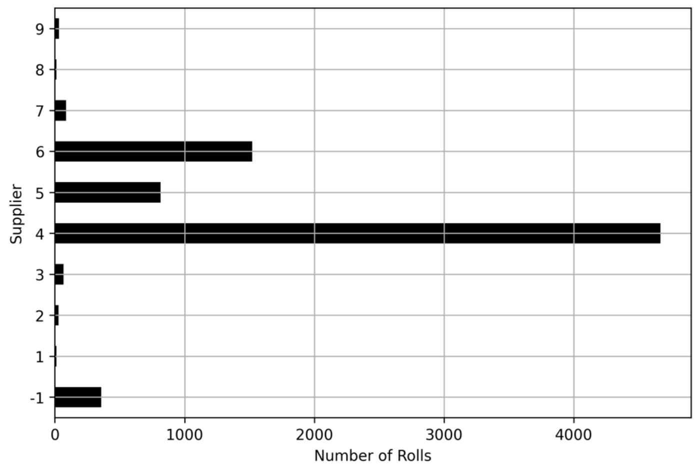

Figure 4: Number of rolls per suppliers

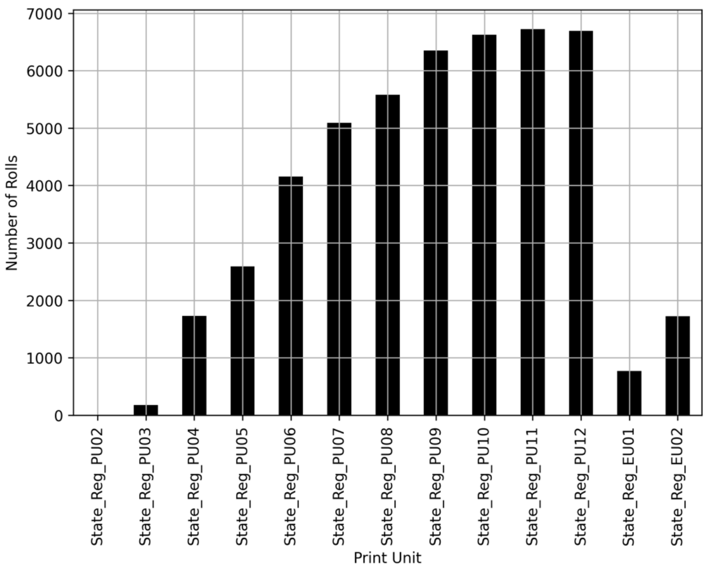

Figure 5: Frequency of active register measurements along the printing press

We would like to point out, that our dataset shows only a partial section of the entire complexity of the printing process of the studied rotogravure press. For example, we do not collect information about changes such as control, ink or dryer parameters that the press operator may have made. Further, one could gather many other potential interesting inline measurement data. However, our dataset provides a strong baseline dataset for various machine learning tasks in context of industrial print application.

4. License and accessibility

The dataset is available under Creative Commons public licence CC BY 4.0 International and can be accessed via DOI: https://doi.org/10.57899/4yjq-h434.

Acknowledgements

The author would like to express extraordinary gratitude to the anonymous industry partners for their great support of this research.

References

Alzghoul, A., Verikas, A., Hållander, M., Bacauskiene, M. and Gelzinis, A., 2009. Screening paper runnability in a web-offset pressroom by data mining. In: P. Perner, ed. *Advances in Data Mining: Applications and Theoretical Aspects: 9th Industrial Conference, ICDM 2009*. Leipzig, Germany, 20–22 July, 2009. Berlin, Heidelberg: Springer, pp. 161–175.

Brumm, P., Weber, T., Lindner, N., Sauer, H.M. and Doersam, E., 2021. *Collection of digitized gravure printed samples for the analysis of hydrodynamic pattern formation*. Technical University of Darmstadt. https://doi.org/10.48328/tudatalib-528.2.

Enk, T., 2022. *Multivariate sensor dataset of an industrial rotogravure printing press (MSDIRPP).* University of Wuppertal. https://doi.org/10.57899/4yjq-h434.

Evans, B., 1995. *Cylinder bands data set*. UCI Machine Learning Repository. [online] Available at: <https://archive.ics.uci.edu/ml/datasets/Cylinder+Bands> [Accessed 18 February 2022].

Parola, M., Kaljunen, T., Beletski, N. and Paukku, J., 2003. Analysing printing press runnability by data mining. In: *Proceedings of the 55th Annual Technology Conference of the Technical Association of the Graphic Arts.* Montreal, Canada, 6–9 April 2003. TAGA, pp. 59–60.

Paukku, J,. Parola, M. and Vuorinen, S., 2004. Web widening and lateral movements of paper web in a printing press. In: *International Printing and Graphic Arts Conference.* Vancouver, BC, Canada, October 2004. IPGAC Preprints, pp. 245–250.

Ranjan, C., Reddy, M., Mustonen, M., Paynabar, K. and Pourak, K., 2018. Dataset: rare event classification in multivariate time series. *arXiv.* [online] Available at: <https://arxiv.org/pdf/1809.10717> [Accessed February 2022].