Predicting the Lifespan of Industrial Printheads with Survival Analysis

Dan Parii Evelyne Janssen Guangzhi Tang Charis Kouzinopoulos Marcin Pietrasik

Department of Advanced Computing Sciences
Maastricht University, Netherlands

May 14, 2025

Maastricht University Canon CANON PRODUCTION PRINTING

Overview

Context

- This research was conducted in collaboration with **Canon Production Printing (CPP)**, which specializes in **industrial-scale printers** used for commercial and production-grade applications.
- A core component of these systems is the **printhead**, responsible for jetting toner or ink.
- Ensuring accurate lifespan prediction of these components is crucial for **maintenance scheduling**, **cost reduction**, and **customer satisfaction**.

Lifespan Prediction

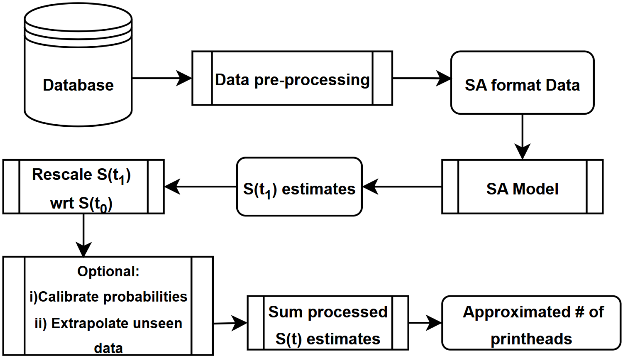
Definition

Lifespan prediction involves estimating the amount of time a component or system will continue to function before it fails.

Why it's challenging:

- Failures may occur unpredictably or after long operational periods.
- Often only **censored** information is available (the unit hasn't failed yet).
- Usage data can be noisy, sparse, or irregular.

Objective



- Improve the accuracy of **printhead lifespan prediction** using modern survival analysis techniques.

Dataset

- Focus on one specific printhead model (older, lots of data).
- **Timeframe:** 2008 - 2024
- Data sources:
  - **Printer metadata**: installation date, position, managed color.
  - **Operational logs**: warm hours (when the printer is on, even if not printing), toner volume.
- Considerations:
  - Heavy **censoring rate: over 70%** of printheads had not failed at time of observation.
  - Missing nozzle-level data due to pre-logging architecture.

Data Cleaning

- **Motivation:** Raw operational data included outliers, inconsistencies, and noise due to diverse printer environments and incomplete logging.
- **Domain knowledge informed cleaning:** Rules and cutoffs were defined in collaboration with Canon Production Printing domain experts.
  - Implausibly high usage statistics
  - Usage $>$ **12 hours/day** (considered abnormal)
  - Devices in storage $>$ **1.5 years** before installation
  - **Dead-on-arrival units** (immediate failures after installation)

Survival Analysis Overview

Goal: Estimate the time until a printhead fails or number of failures.

Each instance i is represented as a triplet (X_i, y_i, δ_i), where:

- X_i: feature vector (e.g., warm hours, toner volume)
- y_i: survival time (T_i)
- δ_i: event indicator ($1 =$ failure, $0 =$ censored)

Survival Function $S(t)$ represents the probability that a printhead survives beyond time t:

$$S(t) = \Pr(T \geq t)$$

To estimate the number of failing printheads, each instance is treated as a Bernoulli variable where p_i is the model-estimated failure probability for unit i:

$$E(X) = \sum_{i=1}^{n} p_i$$

Survival Analysis Methods

Models Implemented:

- Kaplan-Meier (KM)
- Random Survival Forest (RSF)
- Gradient Boosting (CBoost)
- Cox Proportional Hazards (CoxPH)
- Weibull Accelerated Time-Failure Model (ATF)

Purpose: To compare models under consistent preprocessing and evaluation settings.

Non-parametric

$\updownarrow$

Parametric

Evaluation Procedure

Evaluation Setup:

- A prediction window is a specific range of time ($[t_0; t_1]$), where:
  - t_0: maximum date of data logging per printhead (used for training)
  - t_1: future date for which failures are predicted (used for evaluation)
- We extracted six windows, each one year long and spaced six months apart, from:

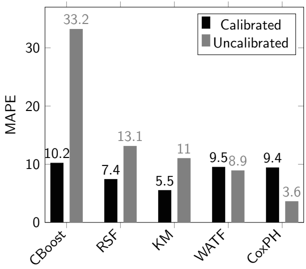
$$[t_0 = \text{May 2021};\ t_1 = \text{May 2022}] \text{ to } [t_0 = \text{Nov 2023};\ t_1 = \text{Nov 2024}]$$

Models were evaluated using:

- Concordance Index (CI)
- Integrated Brier Score (IBS)
- Mean Absolute Percentage Error (MAPE)

Results and Discussion

- **CoxPH (uncalibrated)** had the lowest MAPE: **3.6%**.
- **KM (calibrated)** also performed well: **5.5% MAPE**.
- **Calibration** notably improved CBoost and RSF.
- **MAPE** proved more meaningful for failure prediction than CI or IBS.



Model	CI	IBS
CBoost	**0.818**	**0.077**
RSF	0.807	0.096
ATF	0.79	0.091
CoxPH	0.774	0.094
KM	N/A	0.2
Random Estimator	0.5	0.25

Conclusion and Future Work

Conclusion:

- Survival analysis effectively models printhead lifespan.
- **CoxPH (uncalibrated)** and **KM (calibrated)** gave the most accurate predictions.
- **MAPE** proved more practical than CI or IBS for failure estimation.

Future Work:

- Apply models to additional printheads.
- Integrate richer sensor data (e.g., **nozzle-level logs**).
- Explore deep learning and calibration curves.

Thank You!

Questions or Comments?

This work was conducted in collaboration with Canon Production Printing.