Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

How far can routine process records predict final turbidity in industrial beer crossflow filtration? A capability-boundary study on 5000 production batches

Peide Yang

yangpd@xujc.com

Xiamen University Tan Kah Kee College

Han Dong

Xiamen University Tan Kah Kee College

Yuyang Xia

Xiamen University Tan Kah Kee College

Chongyi Zhu

Xiamen University Tan Kah Kee College

Gongquan Weng

Xiamen University Tan Kah Kee College

Songdi Lin

Xiamen University Tan Kah Kee College

Research Article

Keywords: Beer turbidity, crossflow filtration, soft sensor, capability boundary, dose–response, industrial production data

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11487956/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

How far can routine process records predict final turbidity in industrial beer crossflow filtration? A capability-boundary study on 5000 production batches

Peide Yang[1,2,*], Han Dong[1,2], Yuyang Xia[3], Chongyi Zhu[3], Gongquan Weng[3], and Songdi Lin[3]

[1]Xiamen University Tan Kah Kee College, School of Electronic Science & Technology, Xiamen, China

[2]Key Laboratory of Intelligent Manufacturing Equipment and Industrial Internet Technology, Fujian Provincial Universities, Xiamen, China

[3]Xiamen University Tan Kah Kee College, School of Artificial Intelligence & Computer Science, Xiamen, China

Emails: Peide Yang(*Corresponding author), yangpd@xujc.com; Han Dong, donghan@xujcoj.com; Yuyang Xia, 1697493595@qq.com; Chongyi Zhu, 3933909050@qq.com; Gongquan Weng, 3371699827@qq.com; Songdi Lin, 3048668571@qq.com

Abstract

Crossflow filtration is the last separation step before bright beer, and final turbidity is its key quality attribute; yet the most recent systematic review of 188 brewing-AI studies shows that filtration remains largely unstudied. This work presents a systematic capability assessment of turbidity prediction in industrial beer crossflow filtration, based on 5000 production batches with 37 recorded variables (35 features: 17 tank-1, 17 tank-2 and a dual-tank indicator, plus batch number and turbidity). Ten regression model families were evaluated under two feature sets with 5-fold cross-validation, supported by learning-curve analysis, dose–response testing and permutation importance. The best models (PLS and Ridge, 35 features) reached an RMSE of 1.404 NTU and $R^2 = +0.084$, only marginally above the mean baseline, and no non-linear model outperformed the linear ones. The learning curve plateaus at about 2000 batches with a train–test gap below 0.02, showing that the ceiling is a property of the recorded features rather than of sample size. The strongest single predictor was total recovered yeast from tank 1 (Spearman $\rho = -0.128$), with a strictly monotonic dose–response across quintiles (2.016 → 1.505 NTU; Jonckheere–Terpstra $p < 10^{-15}$), in the direction expected from a particle mass balance. Exceedance classification reached AUC 0.64–0.69 but F1 collapsed at high thresholds, supporting weak ranking only. The recorded process variables therefore carry at most 8–10% of the variance in final turbidity; usable soft sensing at this station requires additional measurement points, in particular pre-filtration online turbidity, transmembrane

pressure and flux, and filtration duration and volume.

Keywords: Beer turbidity; crossflow filtration; soft sensor; capability boundary; dose–response; industrial production data

1. Introduction

Turbidity is among the most critical appearance quality attributes of bright beer: the "bright" visual impression directly shapes consumer evaluation and brand consistency (Olšovská et al., 2021; Reid et al., 2024), and beer viscosity and turbidity are strongly related to the content of non-starch polysaccharides (arabinoxylans, β-glucans) (Reid et al., 2024).

Crossflow filtration is the final solid–liquid separation operation before bright beer: the fermentation broth and recovered yeast flow tangentially along the membrane surface, so that solids are retained while the permeate passes through (Chen and Kim, 2006). The final turbidity is jointly determined by the membrane condition, the suspended-solids load of the feed, and the upstream colloidal precursors (protein–polyphenol complexes, β-glucans, residual yeast) (Reid et al., 2024; Delvaux et al., 2000; Jongberg et al., 2020).

Quality control at the plant is currently *post hoc*: samples are drawn after filtration and before filling, and once turbidity exceeds specification the entire batch can no longer be recovered.

If final turbidity could be predicted *before a batch is discharged*, or *during filtration itself*, the plant could adjust the process in advance (membrane replacement, re-balancing of recovered yeast, stream diversion) and convert after-the-fact inspection into advance warning. This is a canonical soft-sensing problem: inferring a quality variable that is difficult to measure online from process variables that are routinely recorded, and one of the core tasks of data-driven intelligent process industries (Shang and You, 2019; Liu and Xie, 2020).

Machine learning has been widely reported in brewing (fermentation modeling, alcohol and flavor prediction, diacetyl warning, etc.). The most recent systematic review (188 studies, 7 databases) explicitly notes that *quality assurance and fermentation* are the most intensively studied stages, while filtration, filling, and production planning are severely under-represented; the filtration step is a blind spot

of brewing AI (Rewak et al., 2026). A parallel systematic review of machine-learning applications in brewing (21 publications) reaches a consistent conclusion at the field level: research on artificial intelligence in brewing lags significantly behind the general trend of artificial intelligence research (Nettesheim et al., 2024).

In the sibling industry, intelligent sensing and AI control have been systematically integrated into a full-chain "Intelligent Oenological System" framework for winemaking (Yang et al., 2026), but that framework likewise does not cover product-quality soft sensing at the filtration stage. Industrial beer filtration itself has been placed under fuzzy/heuristic control for two decades (Campbell et al., 2002), but such controllers regulate the filtration operation itself (e.g., bodyfeed dosing) rather than predicting the turbidity of the discharged batch. Outside brewing, however, data-driven modelling of crossflow filtration is well established, and several of these studies are close neighbours of the present problem. Artificial neural networks were applied to crossflow microfiltration as early as 1997, in both black-box and semi-physical formulations (Piron et al., 1997), and were subsequently extended to the estimation and generalization of crossflow filtration processes (da Silva and Flauzino, 2008). Turbidity also appears in this literature, though as a predictor rather than a target: for pulpy fruit juice the initial turbidity of the feed is a key factor governing cross-flow microfiltration performance (Vaillant et al., 2008). Dynamic soft sensors have further been developed for turbidity in drinking water (Juntunen et al., 2013). A common structural feature separates all of these from the present problem: they model equipment performance (flux, capacity), or they treat turbidity as an input descriptor in a matrix—fruit juice, drinking water—in which the attribute is directly and continuously measurable. None predicts the quality attribute of a discharged beer batch from upstream operating records alone. Our search found no dedicated study on "beer crossflow filtration × turbidity prediction". The eight closest works each miss a key dimension (Table 1).

Table 1 Closest prior work and the gap addressed by this study.

Closest prior work	**Setting**	**Predicted target**	**Difference from this study**
Brewing-AI review (Rewak et al., 2026)	Full brewing chain	QA / fermentation	Filtration systematically absent
Fuzzy filtration control (Campbell et	Beer (DE) filtration	Filtration performance	Heuristic control, not quality

Closest prior work	Setting	Predicted target	Difference from this study
al., 2002)		(bodyfeed dosing)	prediction
ANN crossflow MF (Piron et al., 1997)	Crossflow microfiltration	Permeate flux	Black-box/semi-physical flux model; not beer
NN filtration estimation (da Silva and Flauzino, 2008)	Crossflow filtration	Process estimation	Not a product-quality target
Initial turbidity, MF (Vaillant et al., 2008)	Fruit-juice crossflow MF	Filtration performance	Turbidity as input, not target
Dynamic turbidity SS (Juntunen et al., 2013)	Drinking water	Turbidity	Other matrix; online sensors available
DCF digital twin (Heusel et al., 2024)	Crossflow filtration	Flux & capacity	Grape juice; non-quality attribute
Membrane-flux ANN/CNN (Chen and Kim, 2006; Yogarathinam et al., 2021)	Membrane filtration	Equipment performance	Not product quality; not beer

Machine learning has already achieved accurate dynamic prediction of other food quality attributes, such as pork pH and meat colour ($R^2 > 0.7$; Yang et al., 2025), which suggests that the limiting factor is not the modelling of food data in general but the variables recorded at the filtration stage. Product-quality (turbidity) soft sensing in beer crossflow filtration therefore remains an open gap.

This study addresses three questions:

- **RQ1**: How much predictive power do the recorded process variables (yeast generation, recovery volumes, temperature, recovery-event sequences) carry for final turbidity, and where is the ceiling?
- **RQ2**: Which recorded variables carry signal, and is it consistent with the physical expectation of a particle mass balance?
- **RQ3**: Can the recorded variables support an industrially usable exceedance warning?

This study makes three contributions. First, it provides the first systematic capability-boundary assessment of this process step, comparing ten model families under two feature sets on 5000 batches with 5-fold cross-validation and multi-seed robustness checks. Second, learning-curve analysis shows that the low R^2 ceiling

originates in the recorded features rather than in sample size, ruling out the explanation that additional data alone would suffice. Third, it documents a strictly monotonic dose–response for tank-1 total recovered yeast (quintile means from 2.016 to 1.505 NTU) that is consistent with a particle mass balance, and identifies the measurement points needed to break through the boundary.

2. Materials and Methods

This section describes the data source and its quality diagnostics, the two feature sets, and the model and evaluation protocol.

2.1 Data source

The data come from historical records of the crossflow filtration section of a commercial brewery (5000 batches, 37 fields each), with features already pre-aggregated on the database side:

- **Tank-1 static parameters (5 dims)**: yeast generation, 0-h recovery, end-of-conversion (EOC) recovery, pre-filtration recovery, and tank temperature at crossflow start;
- **Tank-1 recovery-event sequence statistics (12 dims)**: event count, total recovered volume, per-event mean/max/min, storage mean/max/min, interval mean/max/total, and mean yeast generation;
- **Tank-2 features (17 dims, same structure)**: present only for dual-tank batches; for single-tank batches they are structurally missing (zero-filled) and distinguished by the dual-tank flag;
- **Dual-tank flag (1 dim)**;
- **Target**: final turbidity in NTU.

Table 2 summarizes the data (turbidity 1.713 ± 1.468 NTU, range 0.284–10.373, right-skewed with skewness 1.85; 38.7% of batches below 1 NTU and 3.8% above 5 NTU; 760 dual-tank batches, 15.2%). Fig. 1 shows the distribution, the tank-1 correlation structure, and the features most strongly correlated with turbidity.

Table 2 Data description statistics (n = 5000 batches).

Statistic	Value

Statistic	Value
Samples (batches)	5000
Raw fields	37 (tank-1: 17, tank-2: 17, dual-tank flag, turbidity, batch no.)
Single-tank / dual-tank batches	4240 / 760 (15.2%)
Turbidity, mean ± std	1.713 ± 1.468 NTU
Turbidity range / P50 / P95	0.284–10.373 / 1.32 / 4.55 NTU
Turbidity < 1 NTU / > 5 NTU	38.7% / 3.8%
Skewness / kurtosis	1.85 / 4.98
Yeast generation / temperature range	3–9 generations / 3.11–15.49 °C
Tank-1 feature missingness	0

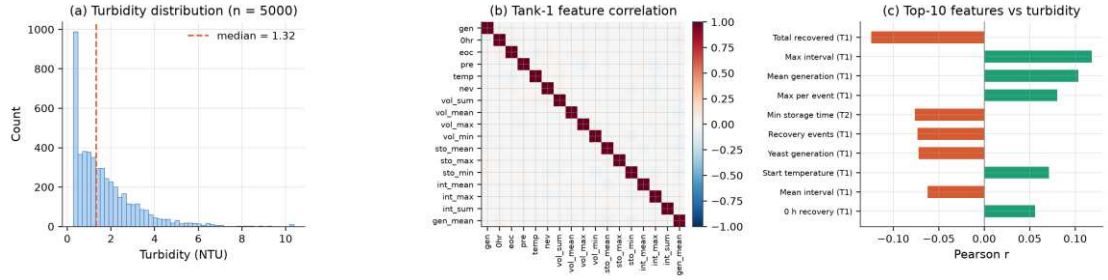


Fig. 1 Data overview. (a) Distribution of final turbidity (n = 5000; median 1.32 NTU): right-skewed (skewness 1.85), with 38.7% of batches below 1 NTU and 3.8% above 5 NTU. (b) Pairwise correlation matrix of the 17 tank-1 features; maximum |r| = 0.039 (near-orthogonal). (c) The ten features most strongly correlated with final turbidity (Pearson r); no feature exceeds |r| = 0.124

2.2 Data quality diagnostics and the chronological limitation

Diagnostics revealed that the batch number field in this dataset is a *row index rather than production chronology* (the export was shuffled):

- *Evidence 1*: yeast generation jumps erratically along batch number (4, 6, 7, 6, 4, 7, 4, 4, 9, 3, 3, 5, …). Yeast propagation is a slow variable; adjacent jumps are physically impossible;
- *Evidence 2*: lag-1 autocorrelations along batch number are −0.017 (yeast generation), −0.003 (temperature) and +0.002 (turbidity), whereas genuine production chronology would give these slow variables substantial positive autocorrelation;
- *Evidence 3*: temperature oscillates between 14.58 → 13.04 → 3.34 → 12.89 °C

along batch number, which is physically impossible.

Consequences and handling. Lag features and rolling-origin chronological cross-validation are both inapplicable to this dataset. All model evaluation in this study therefore uses *random 5-fold cross-validation with fixed seeds*, and Sect. 2.4 tests the robustness of the conclusions across five seeds.

The choice of evaluation strategy itself affects model performance and interpretability: clustered cross-validation in feature or geographical space and naive random cross-validation can yield markedly different conclusions (Sweet et al., 2023). Information leakage between training and test sets (duplicate samples, raw records prior to aggregation) likewise inflates performance metrics; MCC inflation of 0.07–0.43 has been reported (Tampu et al., 2022).

In this study all preprocessing (standardization, imputation) is performed *within folds*, and the aggregated statistics are already database-side products, so leakage paths are excluded by protocol. This data limitation is explicitly declared in the limitations (Sect. 3.12). Should the plant supplement a crossflow start time field, the dataset could be upgraded to chronological evaluation with batch-state memory features.

2.3 Feature sets

- **Feature set A (tank 1, 17 dims)**: records of the first yeast tank only;
- **Feature set B (tank 1 + tank 2 + dual-tank, 35 dims)**: adds the second tank's 17 dims and the dual-tank flag.

The 17 tank-1 features are near-orthogonal (maximum pairwise $|r| = 0.039$, maximum VIF 1.01), so no decorrelation was applied and coefficient instability from collinearity is not a concern.

2.4 Models and evaluation protocol

Models (ten regression families plus a mean baseline): Ridge, Lasso, ElasticNet, PLS (partial least squares, a classical multivariate calibration method founded by Wold and preferred over ordinary multiple regression under collinearity and small samples; Geladi and Kowalski, 1986), SVR-RBF, KNN, RandomForest, GBM, XGBoost, and MLP. Linear, kernel, distance and neural models go through a

unified StandardScaler pipeline.

Table 3 reports the hyperparameter settings. All values were fixed *a priori* and held constant across both feature sets, all folds and all seeds, so that every family is compared under one fully reproducible protocol; no exhaustive grid search was performed, because the objective is an upper-bound assessment of what the recorded variables can support rather than optimal tuning of any single model. The values follow standard defaults together with preliminary runs, and all remaining arguments take library defaults.

Table 3 Hyperparameter settings of the ten regression families (fixed a priori; identical for both feature sets, all folds and all seeds).

Model	**Hyperparameters (value used)**
Ridge	α = 10.0
Lasso	α = 0.05, max_iter = 50,000
ElasticNet	α = 0.05, l1_ratio = 0.5, max_iter = 50,000
PLS	n_components = 8
SVR-RBF	C = 1.0, ε = 0.1
KNN	k = 25
RandomForest	n_estimators = 500, min_samples_leaf = 5, max_features = 0.5
GBM	n_estimators = 400, learning_rate = 0.03, max_depth = 3, subsample = 0.8
XGBoost	n_estimators = 800, learning_rate = 0.03, max_depth = 4, subsample = 0.8, colsample_bytree = 0.8, λ = 1.0, min_child_weight = 5
MLP	hidden layers (64, 32), max_iter = 3000, early stopping enabled

Evaluation protocol:

1. Main results: random 5-fold CV (fixed seed = 42), reporting fold-wise mean ± std of RMSE / MAE / MedAE / R^2;
2. Robustness: Ridge / RF / XGBoost × 5 seeds (0, 1, 42, 123, 2024) × 5 folds = 25 runs;
3. Learning curves: fixed 1000-batch test block, training sizes 250 → 4000, reporting train/test R^2;
4. Metrics reported in both the original space (NTU) and the log1p space (for the right-skewed long tail).

Statistical methods: Spearman correlation, Jonckheere–Terpstra trend test

(monotonicity of the dose–response), linear regression slope, Welch t-test and Mann–Whitney U (dual-tank effect, with Cohen's d and Cliff's delta effect sizes), VIF (collinearity), and permutation importance (held-out 30%, n_repeats = 5).

3. Results and Discussion

We first compare the model families under both feature sets, then present the learning curve, the feature-level signal diagnostics, the dose–response of the strongest variable, the collinearity check, the classification performance, and the dual-tank effect.

3.1 Model comparison: the predictive ceiling of process variables is about 8%

The main results are reported in Table 4 and Fig. 2, from which three observations follow.

1. **All models fall within a narrow band**: excluding the mean baseline, R^2 spans +0.029 to +0.069 under feature set A and +0.021 to +0.084 under feature set B, with SVR-RBF under set A the only negative value (−0.0216); the best PLS/Ridge reach only +0.084, a 4.3% RMSE improvement over the mean baseline ($R^2 \approx 0$);
2. **No non-linear model (RF, GBM, XGBoost, MLP) outperforms the linear models. The records therefore contain no exploitable non-linear structure, and the limitation does not lie in model capacity.**
3. Adding tank-2 features brings a small but directionally consistent improvement (RMSE 1.4158 → 1.4044).

Multi-seed robustness (Table 5) confirms that the conclusions are insensitive to the partition: Ridge 0.0845 ± 0.0137, RF 0.0697 ± 0.0128, XGBoost 0.0505 ± 0.0204 (25 runs).

Table 4 Regression model comparison (5-fold CV, fold-wise mean; R^2 std in parentheses). The mean baseline is identical for both feature sets. Best results in bold.

Model	RMSE	R^2	R^2 (std)
Feature set A: tank 1 (17 features)			
Mean	1.4681	−0.0009	0.0291
Ridge	1.4158	+0.0688	0.0239

Model	RMSE	R^2	R^2 (std)
Lasso	1.4268	+0.0545	0.0119
ElasticNet	1.4193	+0.0643	0.0166
PLS	1.4158	+0.0688	0.0240
SVR-RBF	1.4828	−0.0216	0.0285
KNN	1.4461	+0.0288	0.0162
RandomForest	1.4178	+0.0663	0.0215
GBM	1.4239	+0.0581	0.0243
XGBoost	1.4401	+0.0367	0.0241
MLP	1.4338	+0.0450	0.0276
Feature set B: tank 1 + tank 2 + dual-tank (35 features)			
Mean	1.4681	−0.0009	0.0291
PLS	**1.4044**	**+0.0836**	0.0230
Ridge	**1.4045**	**+0.0835**	0.0227
ElasticNet	1.4116	+0.0745	0.0141
GBM	1.4115	+0.0742	0.0249
RandomForest	1.4134	+0.0720	0.0208
Lasso	1.4229	+0.0597	0.0098
XGBoost	1.4266	+0.0545	0.0227
MLP	1.4273	+0.0537	0.0207
KNN	1.4385	+0.0389	0.0205
SVR-RBF	1.4517	+0.0208	0.0285

Table 5 Multi-seed robustness (5 seeds × 5 folds = 25 runs; mean ± std).

Model	RMSE	R^2
Ridge	1.4036 ± 0.0398	+0.0845 ± 0.0137
RandomForest	1.4149 ± 0.0364	+0.0697 ± 0.0128
XGBoost	1.4293 ± 0.0361	+0.0505 ± 0.0204

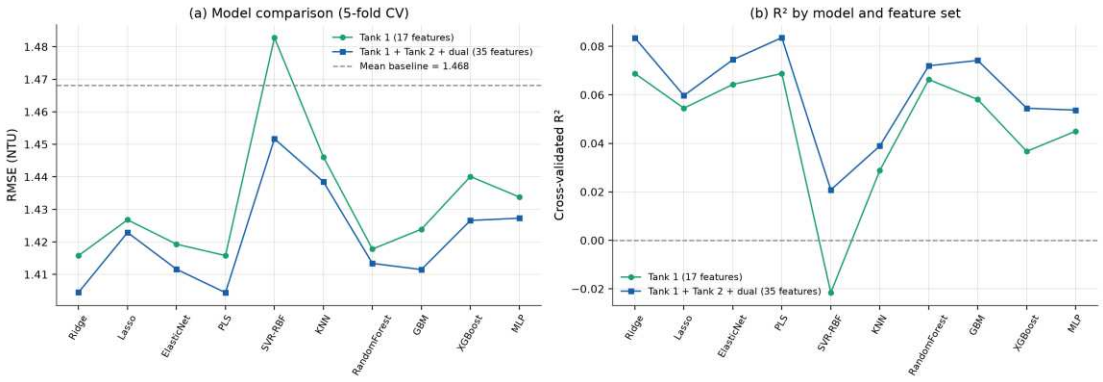


Fig. 2 Model comparison under 5-fold cross-validation (mean ± std). (a) RMSE and (b) cross-validated R^2 for ten model families on feature set A (tank 1, 17 features) and feature set B (tank 1 + tank 2 + dual-tank flag, 35 features). The mean baseline (1.468 NTU, $R^2 \approx 0$) is shown as a dashed line; the best models reach $R^2 = +0.084$, and no non-linear model outperforms the linear ones

3.2 Learning curves: the ceiling is in the features, not in the sample size

The learning curves in Table 6 and Fig. 3 address this question directly.

Test R^2 climbs to 0.082 at n = 1000 and plateaus at ≈ 0.10 from n = 2000 onward; increasing the training set to 4000 batches only reaches 0.103. Meanwhile train R^2 decreases monotonically from 0.265 to 0.095, and the train–test gap narrows from 0.271 at n = 250 to below 0.02 once n ≥ 2000 (0.0040, 0.0030 and −0.0082 at n = 2000, 3000 and 4000, respectively). These results rule out both overfitting, since the gap narrows rather than widens as the data grow, and insufficient sample size, since a fourfold increase in data yields almost no gain. The ceiling is therefore a property of the recorded features.

Table 6 Learning curve (Ridge, 35 features, fixed 1000-batch test block).

Training samples	Train R^2	Test R^2
250	+0.2654	−0.0051
500	+0.2065	+0.0546
1000	+0.1532	+0.0822
2000	+0.1062	+0.1022
3000	+0.1027	+0.0997
4000	+0.0945	+0.1027

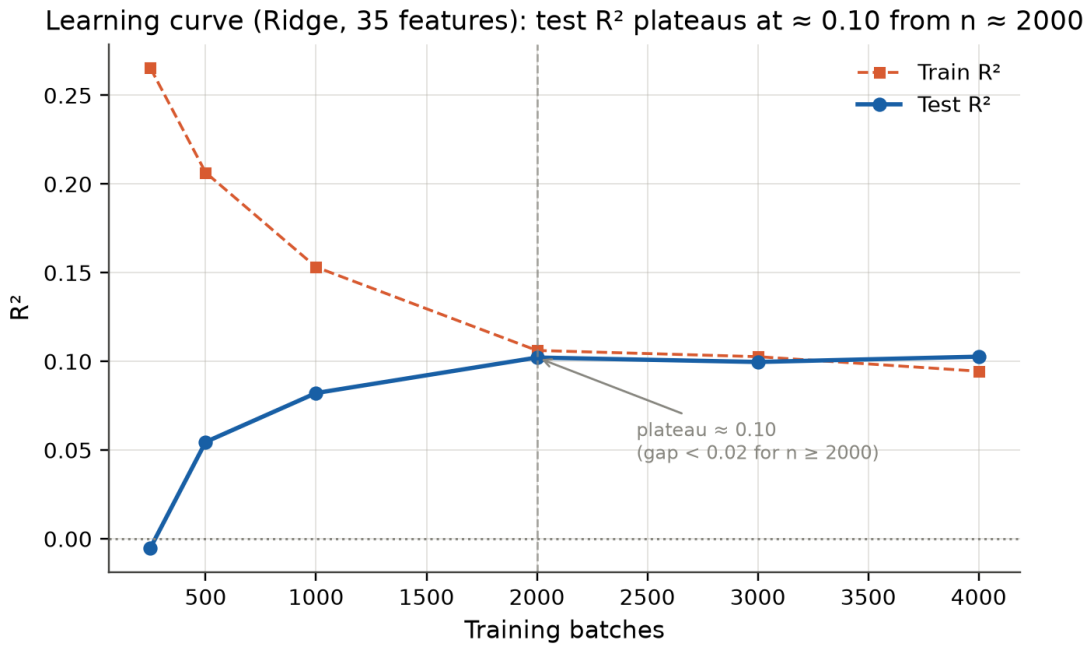


Fig. 3 Learning curve (Ridge, 35 features; fixed 1000-batch test block). Test R^2 reaches 0.082 at n = 1000 and plateaus at ≈ 0.10 from n = 2000, while train R^2 declines monotonically; the train–test gap falls below 0.02 for n ≥ 2000, excluding both overfitting and insufficient-sample explanations

3.3 Signal diagnostics: who carries information, and how much

Table 7 lists the features most strongly correlated with turbidity (Figs. 1c and 4). Across all 35 dims, **no feature exceeds |r| = 0.124**. Although the large sample makes p-values tiny, all effect sizes are weak; the highest permutation importance is 0.035, i.e., removing the most informative single variable decreases held-out R^2 by at most 0.035. For the strongest feature (tank-1 total recovery), the corresponding Spearman coefficient is $\rho = -0.128$, consistent with the Pearson value reported above.

Table 7 The eight features most strongly correlated with turbidity (permutation importance: decrease in R^2, held-out 30%).

Feature	Pearson r	p-value	Perm. importance (ΔR^2)
Total recovered (T1)	−0.124	1.4×10^{-18}	0.0292
Max interval (T1)	+0.118	5.0×10^{-17}	0.0347
Mean generation (T1)	+0.104	2.0×10^{-13}	0.0291
Max per event (T1)	+0.080	1.3×10^{-8}	0.0088
Min storage time	−0.076	6.4×10^{-8}	0.0054

Feature	Pearson r	p-value	Perm. importance (ΔR^2)
(T2)			
Recovery events (T1)	−0.073	2.0×10^{-7}	0.0051
Yeast generation (T1)	−0.072	3.3×10^{-7}	0.0052
Start temperature (T1)	+0.071	4.5×10^{-7}	0.0102

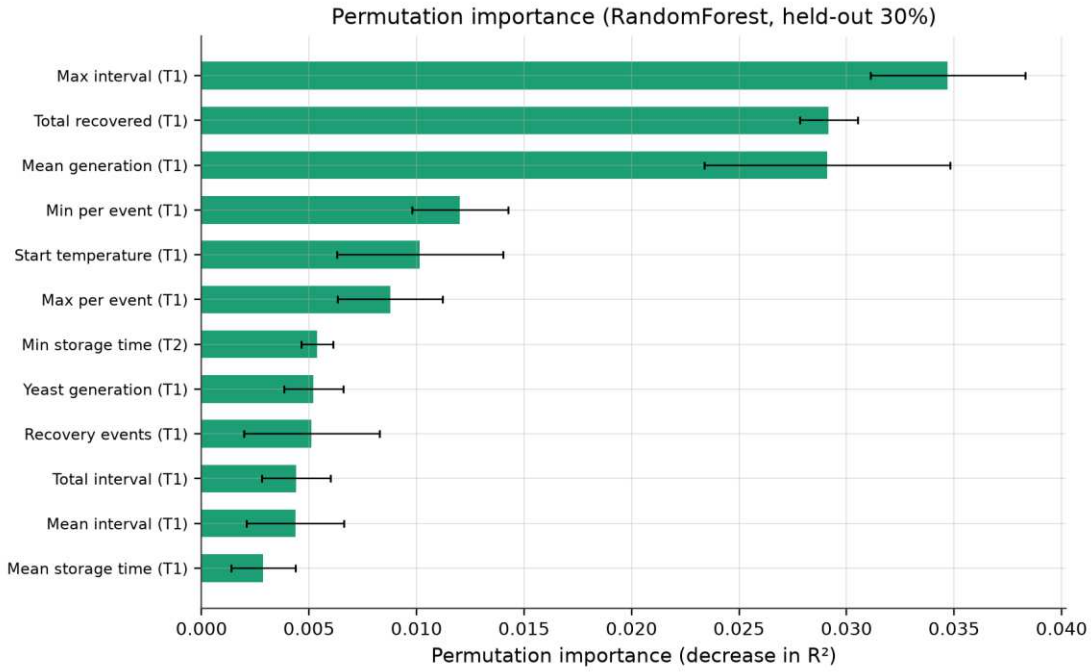


Fig. 4 Permutation importance (RandomForest, held-out 30%, n_repeats = 5; error bars: std over repeats). The strongest features decrease R^2 by at most 0.035 when permuted

3.4 Dose–response: positive evidence in the mass-balance direction

Total recovered yeast from tank 1 is not only the strongest single feature but also shows a **strictly monotonic dose–response** (Table 8, Fig. 5). Spearman $\rho = -0.128$ ($p = 1.1\times10^{-19}$); Jonckheere–Terpstra trend test $p < 10^{-15}$; linear slope −0.00106 NTU per unit ($p = 1.4\times10^{-18}$). **From Q1 to Q5 turbidity drops by 0.511 NTU (about 25%).**

Table 8 Turbidity by quintile of tank-1 total recovered yeast (n = 1000 per quintile).

Quintile	Total-recovered range	Mean ± std	Median	95% CI
Q1 (lowest)	72.0–191.1	2.016 ± 1.657	1.612	±0.103

Quintile	Total-recovered range	Mean ± std	Median	95% CI
Q2	191.1–310.3	1.784 ± 1.491	1.411	±0.092
Q3	310.4–429.5	1.682 ± 1.426	1.310	±0.088
Q4	429.6–548.7	1.577 ± 1.388	1.166	±0.086
Q5 (highest)	548.9–667.9	1.505 ± 1.304	1.134	±0.081

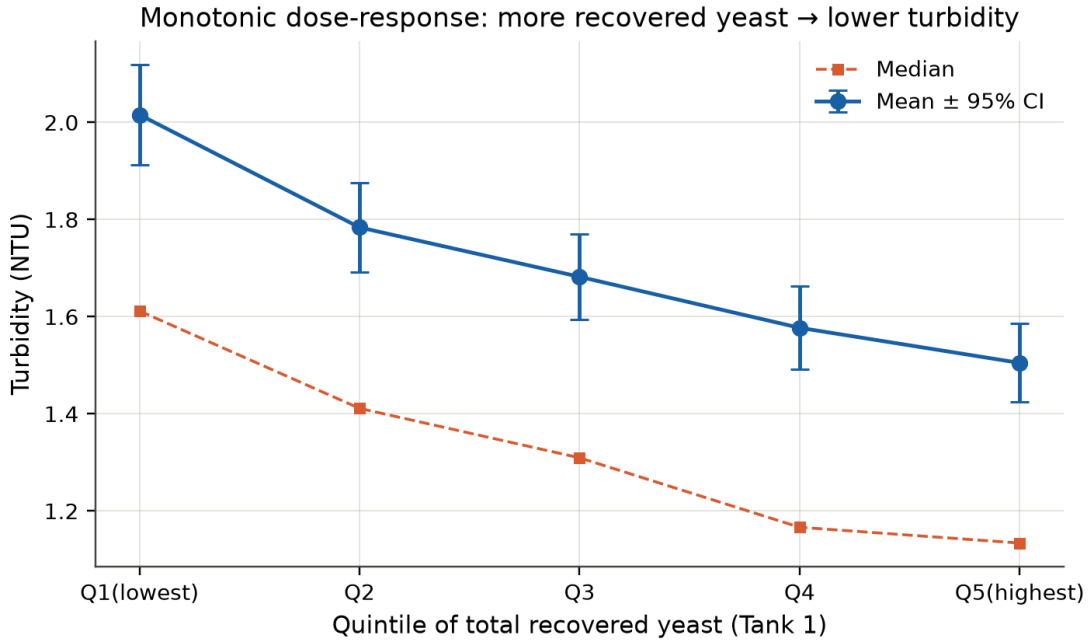


Fig. 5 Monotonic dose–response between tank-1 total recovered yeast and final turbidity (quintile means ± 95% CI and medians). More recovered yeast is associated with lower turbidity, consistent with a particle mass-balance mechanism

The direction of this association is physically plausible: when more yeast is recovered, fewer residual suspended solids enter the filtration stage with the beer, so a lower feed-solids load should yield lower final turbidity, as expected from a particle mass balance.

Experimental studies have confirmed that membrane retention of haze-active proteins and suspended particles directly determines the turbidity of the permeate (Yazdanshenas et al., 2010), and the cake/blocking filtration laws provide the quantitative link between particle load, fouling resistance, and retention efficiency (Iritani and Katagiri, 2016).

Although statistically robust, this relationship accounts for only about 1.5% of the total variance ($r^2 \approx 0.015$), far from enough to support a usable prediction.

3.5 Collinearity: the features are near-orthogonal

For the 17 tank-1 features, the maximum pairwise |r| is 0.039 (mean generation vs. 0-h recovery), zero feature pairs have |r| > 0.3, and VIF is at most 1.01 (mean 1.00). This means the low R^2 ceiling is *not* caused by collinearity: linear models are not missing any explainable correlation structure.

3.6 Classification: weak ranking only, no early warning

Table 9 gives the binary "exceedance vs. not" classification results (Fig. 6). The best AUC per threshold stays in the weak 0.64–0.69 band, and F1 collapses as the threshold rises (0.064 at > 3 NTU). The recorded variables support weak ranking only, not an actionable industrial warning. Because the reported values are the best across the twenty model–feature-set combinations at each threshold, they are optimistic; the RandomForest ROC curves for the 2.0 NTU threshold in Fig. 6 illustrate the model-level performance (AUC 0.630–0.640).

Table 9 Classification task (5-fold CV; best over models and feature sets).

Threshold	Positive rate	Best AUC	Best F1	Note
> 1.0 NTU	61.3%	0.641	0.755	Majority-class prediction gives high F1
> 2.0 NTU	31.5%	0.651	0.305	Weak ranking ability
> 3.0 NTU	15.8%	0.646	0.064	F1 collapse
> 5.0 NTU	3.8%	0.693	0.011	Accuracy inflated by imbalance; F1 ≈ 0

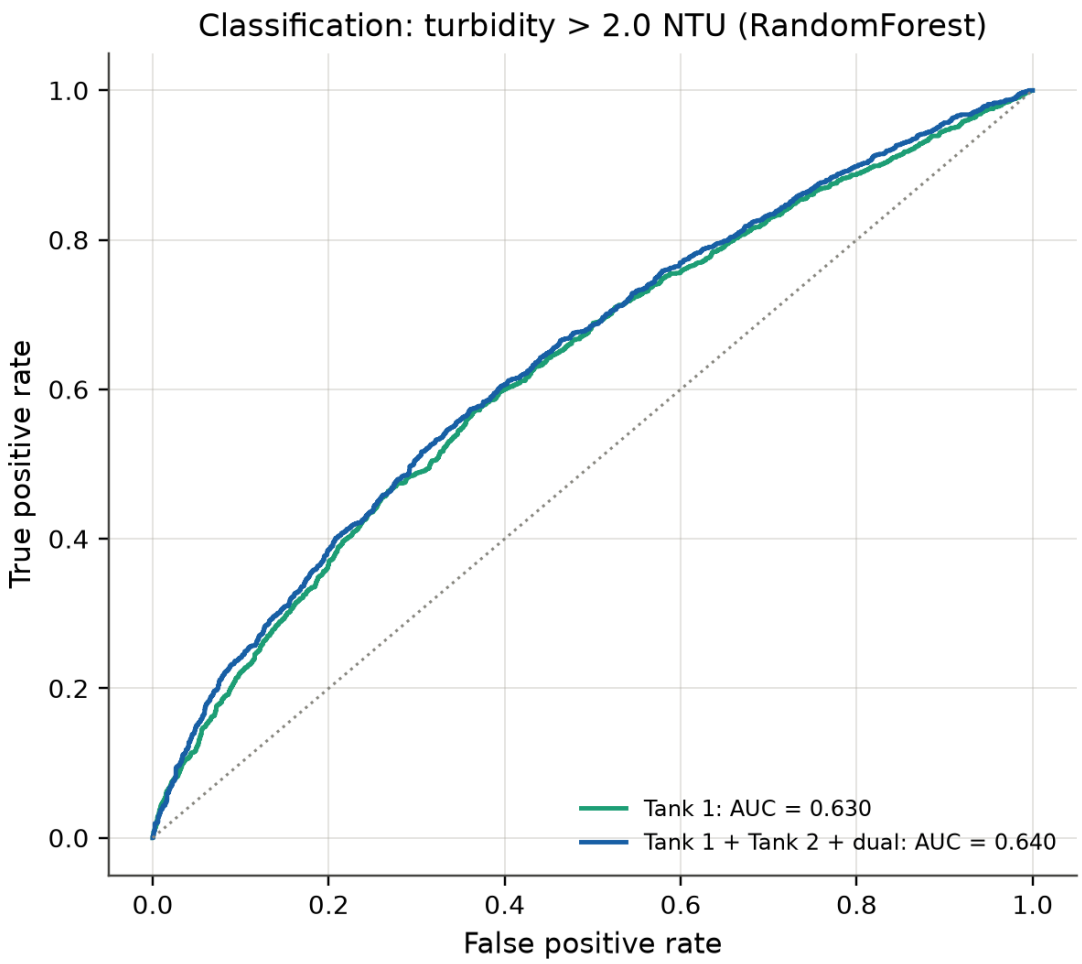


Fig. 6 ROC curves for exceedance of 2.0 NTU (RandomForest): tank-1 features (AUC = 0.630) and tank 1 + tank 2 + dual-tank flag (AUC = 0.640); the dotted diagonal marks the random classifier

3.7 Dual-tank effect: statistically significant but practically limited

There are 760 dual-tank batches with turbidity 1.573 ± 1.389 NTU and 4240 single-tank batches with 1.738 ± 1.480 NTU (Fig. 7). Welch t = 2.974 (p = 0.0030), Mann–Whitney p = 0.0008, but **Cohen's d = 0.112 and Cliff's delta = 0.076; the effect size is negligible**. Exceedance proportions are nearly identical (> 2 NTU: 29.2% vs. 31.9%). Dual-tank batches do exhibit slightly lower turbidity, but the difference carries no engineering significance.

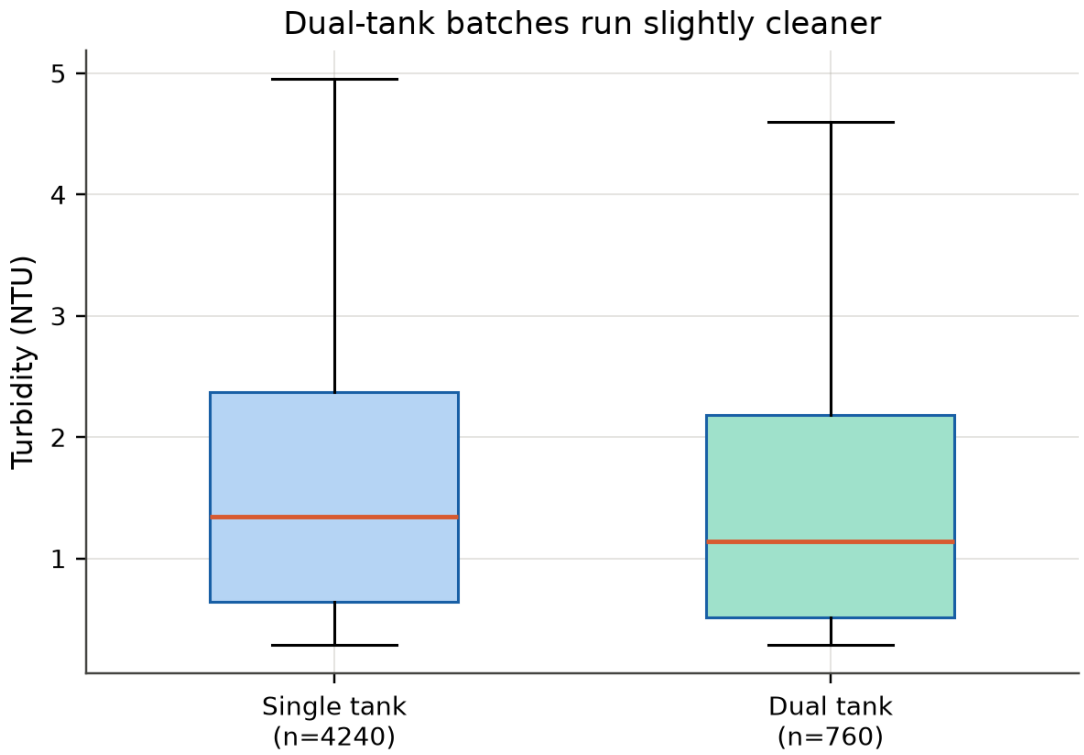


Fig. 7 Final turbidity of single-tank (n = 4240) and dual-tank (n = 760) batches (box: interquartile range; line: median; whiskers: range). The difference is statistically significant but practically negligible (Cohen's d = 0.11)

3.8 Why process variables cannot predict turbidity

Turbidity is an *optical/colloidal* attribute. Its determinants, the colloidal precursors (protein–polyphenol complexes, β-glucans, residual yeast), are formed during fermentation and storage, before filtration (Reid et al., 2024; Delvaux et al., 2000; Jongberg et al., 2020; Królak et al., 2023).

The process variables recorded at the filtration stage (yeast generation, recovery volumes, temperature, recovery-event sequences) reflect *operating conditions*, not the *colloidal load of the feed*. This mismatch between measurement granularity and causal hierarchy is the root cause of the $R^2 \leq 0.084$ ceiling.

The limitation is not attributable to the choice of model: Sect. 3.1 showed that non-linear models perform no better, Sect. 3.5 excluded collinearity, and Sect. 3.2 excluded sample size. This is consistent with a recent critical assessment of AI in food processing, which identifies inconsistent external validation, limited physical credibility and insufficient industrial readiness, rather than model capacity, as the

main bottlenecks (Alaei et al., 2026).

3.9 Mechanistic meaning and quantitative boundary of the dose–response

The monotonic dose–response of tank-1 total recovery (Sect. 3.4) is the most physically coherent result in this study: it runs in the mass-balance direction and is stable across models and feature sets (Pearson r = −0.124, permutation importance 0.029, quintile difference 0.511 NTU).

From the filtration-mechanism perspective, the feed suspended-solids load affects both flux decline and permeate turbidity through cake growth and membrane pore blocking (Chen and Kim, 2006; Madsen et al., 2022; Neuber et al., 2024), so the chain from total recovery to turbidity is physically coherent.

At the same time it delimits the boundary: **the single strongest variable explains only about 1.5% of the variance**. This suggests that follow-up work should construct features along the "feed suspended-solids load" direction (e.g., combinations of total recovery with event structure, or per-volume recovery intensity) rather than continue stacking operating parameters.

3.10 Relation to the literature

To the best of our knowledge, this study is the first systematic assessment of this process step. Consistent with the brewing-AI review's identification of a "filtration blind spot" (Rewak et al., 2026), we find that this blind spot reflects not merely a small number of studies but a genuine insufficiency of the existing records. This negative result is nevertheless informative: it can spare other researchers from repeating similar modelling attempts on comparable records, and it specifies which measurements are needed to make the step predictable.

This conclusion also echoes recent reflection on validation quality and industrial readiness in the field (Alaei et al., 2026): until the information content of the recorded variables at a given process step is established, discussion of model sophistication is premature.

3.11 Industrial implications: what it takes to break the boundary

Breaking through the $R^2 \approx 0.084$ ceiling requires directly measuring the physical

quantities that determine turbidity. In priority order:

1. **Pre-filtration online turbidity, a direct measure of the feed colloidal load, with the largest expected gain;** the link between colloidal load and final turbidity is well supported by physico-chemical evidence (Reid et al., 2024; Yazdanshenas et al., 2010);
2. **Transmembrane pressure (TMP) and flux, direct measures of membrane condition and filtration dynamics; the membrane-filtration field has shown** with ANN and CNN that TMP/flux sequences can be modeled with high accuracy (Chen and Kim, 2006; Yogarathinam et al., 2021; Madsen et al., 2022), and data-driven digital twins have already achieved real-time optimized control of ultrafiltration (Møller et al., 2022);
3. **Filtration duration and filtered volume, measures of filtration intensity that are entirely absent from the current records; time is a core variable** of cake growth and the blocking filtration laws (Iritani and Katagiri, 2016; Neuber et al., 2024);
4. The crossflow start time field (restores chronological information so that batch-state memory features become usable, and chronological/grouped cross-validation can replace random CV, because the evaluation strategy itself changes conclusions (Sweet et al., 2023)).

A parallel route is spectroscopic process analytical technology (PAT): near-infrared spectroscopy combined with machine learning has already enabled real-time monitoring of key parameters in diverse food fermentation processes (Alamprese and Grassi, 2025), and industrial-grade soft sensors have been deployed in biomanufacturing (in-line viable cell density prediction, $R^2 = 0.92$, significantly outperforming the PLS baseline) (Suman et al., 2025).

The spectroscopic route therefore complements the direct physical measurements discussed above and is a medium-term option for automating the operation of this process step.

3.12 Limitations

1. **The data were shuffled: no production chronology, hence no chronological CV and no lag features (Sect. 2.2). Because random cross-validation**

generally yields optimistic estimates when chronological order is informative, the reported figures should be regarded as an upper bound on what prospective evaluation would yield;

2. **Single-plant data: external validity awaits validation on other production lines; inconsistent external validation** is one of the core criticisms of AI in food processing (Alaei et al., 2026);
3. **Coarse record granularity**: all fields are aggregated statistics; no raw time-series curves;
4. **No added measurement points**: see Sect. 3.11;
5. Turbidity is an off-line laboratory measurement whose own measurement error is not modeled.

4. Conclusions

A systematic assessment of 5000 industrial production batches shows that the predictive ceiling of the *current* process records for final turbidity is about $R^2 = 0.08$, and that this ceiling originates in feature properties rather than sample size or model capacity.

The only robust signal in the records is the monotonic dose–response of tank-1 total recovery ($\Delta = 0.511$ NTU, mass-balance direction), which explains only about 1.5% of the variance—insufficient to support usable prediction or warning.

To make this process step amenable to soft sensing (Shang and You, 2019; Liu and Xie, 2020), three classes of measurement points are needed (pre-filtration online turbidity, transmembrane pressure and flux, and filtration duration and volume), and the production-chronology field must be restored.

Acknowledgements

The authors thank the brewery staff for providing access to the production records and for their assistance with data interpretation.

Declarations

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Data availability

The aggregated dataset, the data dictionary and the analysis scripts that support the findings of this study are available from the corresponding author upon reasonable request. The raw production records are proprietary to the brewery and are not publicly available.

Authors' contributions

Peide Yang and Han Dong conceived the study, designed the methodology, supervised the project, and reviewed and edited the manuscript. Yuyang Xia, Chongyi Zhu, Gongquan Weng and Songdi Lin contributed equally to the investigation: they performed the experiments, collected and organised the data, and contributed to the formal analysis. All authors read and approved the final manuscript.

Ethics approval

Not applicable. This study did not involve human participants or animals.

Consent to participate

Not applicable.

Consent to publish

Not applicable.

References

- Alaei B, Chayjan RA, Martynenko AI (2026) Trustworthy and physics-informed artificial intelligence for food processing operations: validation, control relevance, and industrial readiness. *Trends in Food Science & Technology* 178:106132. https://doi.org/10.1016/j.tifs.2026.106132
- Alamprese C, Grassi S (2025) Food fermentations: NIR spectroscopy as a tool for process analytical technology. *Advances in Food and Nutrition Research* 115:391–430. https://doi.org/10.1016/bs.afnr.2025.06.002
- Campbell D, Lees MJ, Rogers PG, Glastonbury R, Pecar M (2002) Process optimisation of beer filtration using fuzzy control. In: *Proceedings of the 6th IASTED International Conference on Artificial Intelligence and Soft Computing*, ACTA Press, Anaheim, pp. 636–640
- Chen H, Kim AS (2006) Prediction of permeate flux decline in crossflow membrane filtration of colloidal suspension: a radial basis function neural network approach. *Desalination* 192(1–3):415–428. https://doi.org/10.1016/j.desal.2005.07.045
- da Silva IV, Flauzino RO (2008) An approach based on neural networks for estimation and generalization of crossflow filtration processes. *Applied Soft Computing* 8(1):590–598. https://doi.org/10.1016/j.asoc.2007.03.008
- Delvaux FR, Delvaux J, Delcour JA (2000) Characterisation of the colloidal haze in commercial and pilot scale Belgian white beers. *Journal of the Institute of Brewing* 106(4):221–228. https://doi.org/10.1002/j.2050-0416.2000.tb00061.x
- Geladi P, Kowalski BR (1986) Partial least-squares regression: a tutorial. *Analytica Chimica Acta* 185:1–17. https://doi.org/10.1016/0003-2670(86)80028-9
- Heusel M, Grim G, Rauhut J, Franzreb M (2024) Regression metamodel-based digital twin for an industrial dynamic crossflow filtration process. *Bioengineering* 11(3):212. https://doi.org/10.3390/bioengineering11030212
- Iritani E, Katagiri N (2016) Developments of blocking filtration model in membrane filtration. *KONA Powder and Particle Journal* 33:179–202. https://doi.org/10.14356/kona.2016024
- Jongberg S, Andersen ML, Lund MN (2020) Characterisation of protein–polyphenol interactions in beer during forced aging. *Journal of the Institute of Brewing* 126:371–381. https://doi.org/10.1002/jib.623
- Juntunen P, Liukkonen M, Lehtola M, Hiltunen Y (2013) Dynamic soft sensors for detecting factors affecting turbidity in drinking water. *Journal of Hydroinformatics* 15(2):416–426. https://doi.org/10.2166/hydro.2012.052
- Królak K, Kobus K, Kordialik-Bogacka E (2023) Effects on beer colloidal stability of full-scale brewing with adjuncts, enzymes, and finings. *European Food Research and Technology* 249(1):47–53. https://doi.org/10.1007/s00217-022-04131-7
- Liu Y-Q, Xie M (2020) Rebooting data-driven soft-sensors in process industries: a review of kernel methods. *Journal of Process Control* 89:58–73.

https://doi.org/10.1016/j.jprocont.2020.03.012

- Madsen E, Kaiser J, Krühne U, Pinelo M (2022) Single pass tangential flow filtration: critical operational variables, fouling, and main current applications. *Separation and Purification Technology* 291:120949. https://doi.org/10.1016/j.seppur.2022.120949
- Møller JK, Goranović G, Brath P, Madsen H (2022) A data-driven digital twin for water ultrafiltration. *Communications Engineering* 1:23. https://doi.org/10.1038/s44172-022-00023-6
- Nettesheim P, Burggräf P, Steinberg F (2024) Applications of machine learning in the brewing process: a systematic review. *Discover Artificial Intelligence* 4(1):80. https://doi.org/10.1007/s44163-024-00177-6
- Neuber D, Buchwald T, Ditscherlein R, Friebel JM, Peuker UA (2024) Analytical characterization of filtration laws in filter aid filtration. *Separation and Purification Technology* 343:127138. https://doi.org/10.1016/j.seppur.2024.127138
- Olšovská J, Kyselová L, Kubizniaková P, Slabý M (2021) Non-microbiological turbidity of beer: part 1—semireview. *Kvasny Prumysl* 67(4):484–497. https://doi.org/10.18832/kp2021.67.484
- Piron E, Latrille E, René F (1997) Application of artificial neural networks for crossflow microfiltration modelling: `black-box' and semi-physical approaches. *Computers & Chemical Engineering* 21(9):1021–1030. https://doi.org/10.1016/S0098-1354(96)00332-8
- Reid JESJ, Yakubov GE, Lawrence SJ (2024) Non-starch polysaccharides in beer and brewing: a review of their occurrence and significance. *Critical Reviews in Food Science and Nutrition* 64(3):837–851. https://doi.org/10.1080/10408398.2022.2109585
- Rewak K, Zygadło K, Jackowski M (2026) Artificial intelligence in beer brewing: a comprehensive review. *Beverages* 12(10):118. https://doi.org/10.3390/beverages12100118
- Shang C, You F (2019) Data analytics and machine learning for smart process manufacturing: recent advances and perspectives in the big data era. *Engineering* 5(6):1010–1016. https://doi.org/10.1016/j.eng.2019.01.019
- Suman S, Murr M, Crowe J, Holt S, Morris J, Yongky A, McElearney K, Bolton GR (2025) In-line prediction of viability and viable cell density through machine learning-based soft sensor modeling and an integrated systems approach: an industrially relevant PAT case study. *Biotechnology Progress* 41(3):e3520. https://doi.org/10.1002/btpr.3520
- Sweet L-B, Müller C, Anand M, Zscheischler J (2023) Cross-validation strategy impacts the performance and interpretation of machine learning models. *Artificial Intelligence for the Earth Systems* 2(4):e230026. https://doi.org/10.1175/AIES-D-23-0026.1
- Tampu AE, Eklund A, Haj-Hosseini N (2022) Inflation of test accuracy due to data leakage in deep learning-based classification of OCT images. *Scientific Data* 9:571. https://doi.org/10.1038/s41597-022-01618-6
- Vaillant F, Pérez A, Acosta O, Dornier M (2008) Turbidity of pulpy fruit juice: a key factor for predicting cross-flow microfiltration performance. *Journal of Membrane Science* 325(1):404–412. https://doi.org/10.1016/j.memsci.2008.08.003

- Yang G, Ma X, Chen X, Liu Q, Tao Y-S, Bai X (2026) Toward an Intelligent Oenological System (IOS): from automated sensing to active control in winemaking. *Comprehensive Reviews in Food Science and Food Safety* 25(1):e70375. https://doi.org/10.1111/1541-4337.70375
- Yang X-D, Zhu L-Y, Jiang W, Yang Y-T, Niu L-L, Zhao Y et al (2025) Artificial intelligence-driven food quality prediction: applying machine learning ensemble models for dynamic forecasting of pork pH and meat color changes. *Food Chemistry* 492:145383. https://doi.org/10.1016/j.foodchem.2025.145383
- Yazdanshenas M, Tabatabaei Nejad SAR, Soltanieh M, Tavakkoli A, Babaluo AA, Fillaudeau L (2010) Dead-end microfiltration of rough nonalcoholic beer by different polymeric membranes. *Journal of the American Society of Brewing Chemists* 68(2):83–88. https://doi.org/10.1094/asbcj-2010-0315-01
- Yogarathinam L, Velswamy K, Gangasalam A, Ismail AF, Goh P, Narayanan A, Abdullah MS (2021) Performance evaluation of whey flux in dead-end and cross-flow modes via convolutional neural networks. *Journal of Environmental Management* 301:113872. https://doi.org/10.1016/j.jenvman.2021.113872