JSCE (Journal of System and Computer Engineering)
Vol.7, No.4, October 2026, pp. 428-438
ISSN (online): 2723-1240
DOI : https://doi.org/10.61628/jsce.v7i4.2854

Research Article Open Access (CC–BY-SA) ■

Random Survival Forest for Time-to-Maturity Prediction in Black Soldier Fly

Rozalina Amran[1,a,*]; Muhammad Syafaat [2,b]; Andi Taufiqurrahman Akbar [3,c]; Aulia Handayani Tahir[4,d]
[1,2,3,4] *Institut Teknologi Bacharuddin Jusuf Habibie, Jl. Balaikota No.1, Bumi Harapan, Kec. Bacukiki Bar., Kota Parepare 91122, Indonesia*
[a] rozalina954@gmail.com / rozalina24@ith.ac.id ; [b] syaffaat@ith.ac.id ; [c] taufiqstmik201013@gmail.com
** Corresponding author*

Abstract

Black soldier fly (BSF, Hermetia illucens) larvae are widely used to biologically convert organic waste into value-added biomass, but the time required to reach maturity varies with rearing temperature, humidity, and feed intake, which complicates harvest scheduling. This study developed and evaluated a Random Survival Forest (RSF) model to predict the time-to-transition to the adult ('Dewasa') stage of BSF under right-censoring, using temperature, humidity, feed consumption, larval count, and larval weight as covariates, on a monitoring dataset of 8,000 observations (28.1% events, 71.9% censored) over a 30-day rearing period. As groundwork, regression and classification benchmarks were also trained; Random Forest and Gradient Boosting explained most of the variance in total biomass (R^2 = 0.9998) and feed conversion ratio (R^2 = 0.929–0.938) and achieved near-perfect phase classification (accuracy 0.9999), results interpreted cautiously given likely redundancy between predictors and these outcomes. In contrast, the RSF model itself achieved a Harrell and Uno C-index of 0.503, an Integrated Brier Score of 0.195, and a mean time-dependent AUC of 0.503, indicating discrimination and calibration close to chance despite permutation importance detecting statistically distinguishable, if small and partly counterintuitive, contributions from larval weight and humidity. These findings show that, within the observed range of rearing conditions, environmental and feeding covariates carried only marginal information about maturation timing beyond what is captured by elapsed rearing day, and they motivate wider covariate ranges, multistate transition modelling, and additional substrate covariates before an RSF-based decision-support tool for circular organic waste management can be recommended for operational use.

Keywords—Black soldier fly, Machine learning, Random Survival Forest, Right-censoring, Time-to-Transition.

1. Introduction

Black soldier fly (Hermetia illucens, BSF) is a biological agent widely used for organic waste bioconversion, because its larval stage can utilize a broad range of organic substrates as a growth medium. The success of this bioconversion process is closely tied to the biological development of the larvae, and the type and characteristics of the substrate are known to affect developmental time, larval growth, and the overall efficiency of waste processing (Lalander et al., 2019).

Beyond substrate characteristics, environmental factors, particularly temperature, exert a strong influence on the development and survival of H. illucens. Surendra et al., (2020) demonstrated marked differences in BSF development and survival across rearing temperatures,

and Herremans et al., (2023) further showed that both temperature and organic waste type as substrate affect several BSF life-history traits, including larval development and adult emergence.

Because substrate and environmental conditions vary, the time BSF individuals need to reach a given developmental stage is not constant, so predictions based on calendar age alone may fail to capture the underlying biological heterogeneity. Recent studies have applied machine learning to predict biological age and larval weight from combinations of biological, substrate, and rearing variables. Alagappan et al., (2026), for instance, used a partial least squares regression approach to predict the biological age and weight of BSF larvae across several organic waste streams, while Muinde et al., (2023) applied machine learning to identify rearing conditions associated with optimal larval yield. In a similar vein, Belperio et al., (2024) linked diet composition to BSF substrate-utilization and growth outcomes, and Kullan et al., (2024) reviewed BSF-based bioconversion of organic residues as a circular-economy model for the poultry industry. These studies demonstrate the value of predictive approaches for organic-waste-stream management, but neither explicitly models the probability of a developmental-stage transition as a time-to-event problem.

Random Survival Forest (RSF) offers a relevant approach in this context. RSF is a nonparametric ensemble machine-learning method for time-to-event analysis that can accommodate censored observations and model complex relationships between multiple predictors and an observed outcome (Ishwaran et al., 2008). RSF also allows estimation of the survival function and identification of the relative contribution of each predictor to the survival outcome, capabilities that standard regression or classification models applied to BSF data so far have not provided (Wang et al., 2019). Beyond its original clinical application, RSF and related forest-based survival methods have recently been extended to accelerated, interpretable oblique variants (Jaeger et al., 2024), dynamic prediction from longitudinal biomarkers (Pickett et al., 2021), and ecological time-to-event problems such as forest harvest-timing (Senevirathne et al., 2024), underscoring the flexibility of this approach across domains.

Accordingly, this study is directed toward developing an RSF-based model that predicts the time and probability of BSF development toward specific biological stages, conditioned on rearing conditions and organic-waste-substrate characteristics. Such information can subsequently support decisions on rearing duration and harvest timing, so that the bioconversion of organic waste into value-added biomass proceeds in a more measurable way and supports the principle of circular organic waste management (Alagappan et al., 2026; Lalander et al., 2019; Surendra et al., 2020). Beyond larval biomass itself, BSF rearing also yields a nutrient-rich residue (frass) with recognized value as an organic fertilizer, reinforcing the broader circular-economy rationale for BSF-based organic waste treatment (Gärttling & Schulz, 2022; Herremans et al., 2023).

This manuscript reports (i) a descriptive characterization of a BSF rearing-monitoring dataset spanning nine developmental phases over a 30-day period, (ii) preliminary regression and classification models used as a reference benchmark for the predictive signal carried by the recorded rearing and biological variables, and (iii) the development and evaluation of an RSF model predicting time-to-maturity, including its discrimination (C-index, time-dependent AUC), calibration (Integrated Brier Score), and covariate importance.

Specifically, this study addresses the following research questions: (RQ1) how do organic-waste-substrate characteristics and rearing conditions influence the survival probability and time-to-developmental-transition of H. illucens; (RQ2) can an RSF model accurately predict the time required for H. illucens to reach specific developmental stages while appropriately accounting for censored observations; (RQ3) which biological, environmental, and substrate-related variables are the most influential predictors of H. illucens developmental timing; (RQ4) how accurately does the RSF model predict BSF developmental events based on survival-prediction metrics such as the C-index, Integrated Brier Score, and time-dependent AUC; and (RQ5) how can the predicted developmental probabilities and time-to-stage estimates be translated into operational decisions for BSF-based circular organic waste management. The overall aim is to develop and

evaluate an RSF-based framework for predicting BSF time-to-developmental-transitions, and to assess its potential as a decision-support tool for circular organic waste management; the specific objectives are to (1) characterize BSF survival and developmental-time patterns under varying rearing and substrate conditions, (2) develop an RSF model for predicting time-to-developmental events while accounting for censored observations, (3) identify the most influential predictors of BSF development, (4) evaluate model performance using appropriate survival-analysis metrics, (5) estimate stage-specific survival probabilities and expected developmental timing, and (6) translate the resulting predictions into decision-support information for BSF rearing, harvesting, and organic-waste bioconversion.

2. Method

2.1 Data Description

The dataset used in this preliminary stage consists of 8,000 rearing-monitoring observations recorded from Day 1 to Day 30 of the BSF rearing cycle, covering nine developmental phases: Larval Instar 1–5, Prepupa, Pupa, Early Adult, and Adult. For each observation, the recorded variables were rearing day, number of larvae in the rearing unit, mean larval body weight (mg), daily feed consumption (g), rearing temperature (°C), relative humidity (%), total larval biomass (g), daily growth rate (g/day), and feed conversion ratio (FCR). Table 1 summarizes the descriptive statistics of the numeric variables.

2.2 Preliminary Regression and Classification Benchmarking

Prior to developing the RSF model, two sets of supervised models were trained on the monitoring dataset to identify which variables carry predictive signal for BSF growth and development, and to establish reference error levels for comparison with the forthcoming survival model. First, three regression algorithms Linear Regression, Random Forest Regression, and Gradient Boosting Regression were trained to predict total biomass, growth rate, and FCR from rearing day, temperature, humidity, feed consumption, and larval count, and evaluated using mean absolute error (MAE), root mean squared error (RMSE), and the coefficient of determination (R^2) on a held-out test split. Second, a Random Forest classifier was trained to recognize the developmental phase label from the same predictor set, and its generalization performance was assessed using leave-one-out cross-validation (LOOCV), reporting accuracy, weighted F1-score, and the per-class precision-recall-F1 profile.

2.3 Random Survival Forest Framework

BSF development was reframed as a time-to-event problem in which the event of interest was the transition to the adult (‘Dewasa’) stage. Individuals that had not yet reached that stage by the end of the 30-day observation window were treated as right-censored. Of the 8,000 records, 2,249 (28.1%) reached the adult stage within the observation window (an event) and 5,751 (71.9%) were right-censored; the same proportion was preserved in the held-out test set (675 events and 1,725 censored observations out of 2,400). The covariate set comprised rearing temperature, humidity, feed consumption, larval count, and mean larval weight; rearing day was used to define the time axis rather than as a predictor. The RSF model was implemented following Ishwaran et al. (2008), using an ensemble of survival trees grown on bootstrap samples, the log-rank statistic as the splitting rule, and the ensemble cumulative hazard function averaged across trees to obtain individual survival estimates. Model performance was evaluated on the held-out test set using Harrell's and Uno's Concordance Index (Harrell et al., 1996; Uno et al., 2011), the Integrated Brier Score (Berkowitz et al., 2024; Liu et al., 2021), and the time-dependent (cumulative/dynamic) AUC over the days in which events were observed in the test set (day 23–30). Covariate contribution was assessed via permutation importance with 15 repeats per covariate, reported as

the mean change in C-index and its 95% confidence interval, computed via the normal approximation (mean ± 1.96 × SE, where SE = SD/√15) across the repeated permutations. RSF hyperparameters (number of trees, minimum samples required to split an internal node and to form a leaf, and the number of features considered at each split) were selected via five-fold cross-validated grid search on the training set, evaluated over the candidate grid n_estimators ∈ {200, 400}, min_samples_split ∈ {10, 20}, min_samples_leaf ∈ {5, 10, 15}, and max_features ∈ {sqrt, 1.0}, using Harrell's concordance index as the optimization criterion, with the log-rank splitting rule and random-state seed fixed; Table 2 reports the selected configuration and the full search space.

2.4 Tools

All analyses were performed in Python using scikit-learn for the regression and classification benchmarking, and scikit-survival for the RSF model, its evaluation metrics, and the permutation importance procedure (Cardoso et al., 2025; Pölsterl, 2020; Senevirathne et al., 2024).

3. Results And Discussion

3.1 Dataset Characteristics

The monitoring dataset comprised 8,000 observations recorded over a 30-day rearing period and spanning nine developmental phases, from Larval Instar 1 through Adult. Table 1 presents the descriptive statistics of the numeric variables. Total biomass and FCR showed the largest relative variability (coefficient of variation above 90%), reflecting the wide range of larval counts and developmental stages captured across the 30-day window, while temperature (mean 29.5 ± 1.5 °C) and humidity (mean 75.0 ± 5.8%) were comparatively stable, consistent with a controlled rearing environment. This spread supports the premise underlying RQ1: developmental and growth trajectories in this dataset are heterogeneous rather than uniform across rearing days, motivating a modelling approach such as RSF that can represent this heterogeneity explicitly rather than assuming a fixed calendar-age trajectory.

Table 1. Descriptive statistics of the BSF rearing-monitoring dataset

Variable	Mean	SD	Min	Max
Day (day)	15.64	8.74	1.00	30.00
Larval count (individuals)	2766.73	1284.58	500.00	5000.00
Mean larval weight (mg)	141.75	103.24	1.00	299.97
Feed consumption (g)	271.65	131.14	50.01	499.96
Temperature (°C)	29.51	1.46	27.00	32.00
Humidity (%)	75.00	5.83	65.00	85.00
Total biomass (g)	393.58	364.74	0.65	1495.10
Growth rate (g/day)	1.40	2.50	0.02	19.88
FCR	5.97	19.19	0.04	560.74

Table 2. Random Survival Forest hyperparameter configuration

Hyperparameter	Candidate Values Tested	Value	Determination Method
n_estimators	[200, 400]	200	GridSearchCV, 5-fold CV (optimization criterion: Harrell's C-index)
min_samples_split	[10, 20]	10	GridSearchCV, 5-fold CV (optimization criterion: Harrell's C-index)

min_samples_leaf	[5, 10, 15]	5	GridSearchCV, 5-fold CV (optimization criterion: Harrell's C-index)
max_features	[sqrt, 1.0]	sqrt	GridSearchCV, 5-fold CV (optimization criterion: Harrell's C-index)
max_depth	–	None (unrestricted)	Default (not included in tuning grid)
bootstrap	–	True	Default (scikit-survival default)
random_state (seed)	–	42	Fixed for reproducibility
Splitting rule	–	Log-rank statistic	Fixed (Ishwaran et al., 2008)

3.2 Preliminary Regression Models

Table 3 summarizes the performance of the three regression algorithms for the three continuous targets. Random Forest achieved the lowest error for total biomass (MAE = 3.31 g, R^2 = 0.9998), clearly outperforming Linear Regression (R^2 = 0.8658) and Gradient Boosting (R^2 = 0.9988). For growth rate, all three algorithms performed considerably worse, with Random Forest again the best model but explaining only 61.0% of the variance (R^2 = 0.610), and for FCR, Gradient Boosting was marginally superior to Random Forest (R^2 = 0.9375 versus 0.9287), both far exceeding Linear Regression (R^2 = 0.210).

Table 3. Performance of regression models for total biomass, growth rate, and FCR

Target	Model	MAE	RMSE	R^2
Total biomass (g)	Linear Regression	102.75	131.73	0.8658
	Random Forest	3.31	5.08	0.9998*
	Gradient Boosting	9.12	12.38	0.9988
Growth rate (g/day)	Linear Regression	1.1677	2.0241	0.2915
	Random Forest	0.7078	1.5017	0.6100*
	Gradient Boosting	0.7111	1.5178	0.6016
FCR	Linear Regression	7.4897	15.7724	0.2100
	Random Forest	0.7771	4.7387	0.9287
	Gradient Boosting	1.1121	4.4367	0.9375*

**Best-performing model for the corresponding target.*

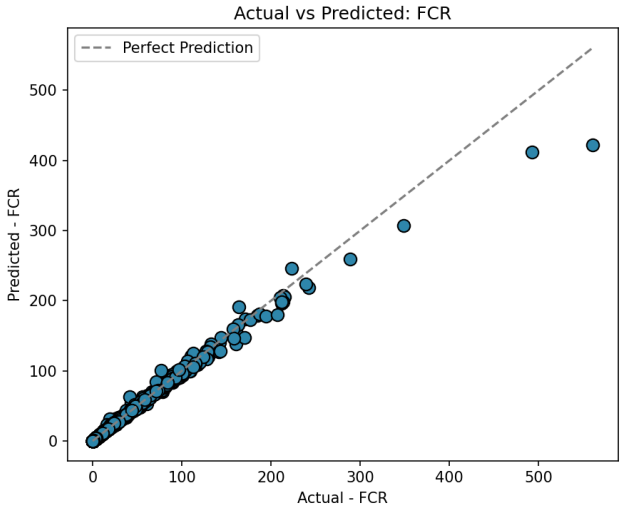


Figure 1. Actual versus predicted FCR (best-performing model, Gradient Boosting). Points close to the dashed line indicate accurate prediction; the largest observed FCR values are visibly underpredicted.

3.3 Preliminary Phase Classification

The Random Forest classifier trained to recognize the developmental phase achieved a LOOCV accuracy and weighted F1-score of 0.9999, with precision, recall, and F1-score at or near 1.00 for all nine phase classes (Table 4). While this indicates that the recorded variables are highly informative of developmental phase, an accuracy this close to perfect on a nine-class problem most plausibly reflects a strong, near-deterministic association between phase labels and one or more predictors, such as rearing day or larval weight bins used when the monitoring records were compiled, rather than a fully independent validation of a generalizable classifier. This result should therefore be read as an upper-bound feasibility check—it shows that phase can, in principle, be inferred from the recorded rearing variables—rather than as a claim of production-ready classification performance; confirmation on an independent rearing batch is recommended before the result is relied upon operationally.

Table 4. Classification performance for developmental phase (Random Forest, LOOCV).

Phase	Precision	Recall	F1-score	Support
Adult	1.00	1.00	1.00	2249
Early Adult	1.00	1.00	1.00	513
Larval Instar 1	1.00	1.00	1.00	303
Larval Instar 2	1.00	1.00	1.00	263
Larval Instar 3	1.00	1.00	1.00	515
Larval Instar 4	1.00	1.00	1.00	509
Larval Instar 5	1.00	1.00	1.00	756
Prepupa	1.00	1.00	1.00	1072
Pupa	1.00	1.00	1.00	1820
Accuracy / Weighted avg.	1.00	1.00	1.00	8000

3.4 From Preliminary Models to RSF Covariate Selection

Taken together, the preliminary models point to temperature, humidity, and feed consumption as covariates that contribute genuine, non-trivial predictive signal (evidenced by the growth-rate and FCR results), while variables that are structurally linked to the outcome (larval count and weight for total biomass; day/weight for phase) were retained cautiously as RSF covariates to avoid encoding a tautological relationship into the survival model. These findings, together with prior evidence that temperature and substrate strongly shape BSF life-history traits (Shumo et al., 2019; Tomberlin et al., 2009), guided the covariate set used in the RSF model reported in Section 3.5: rearing day was used to define the time axis rather than as a predictor, and temperature, humidity, feed consumption, larval count, and larval weight were retained as candidate covariates.

3.5 Random Survival Forest: Time-to-Maturity Prediction

Of the 8,000 monitored individuals, 2,249 (28.1%) reached the adult ('Dewasa') stage within the 30-day observation window and were treated as events, while 5,751 (71.9%) had not yet matured by day 30 and were treated as right-censored; the same proportions were preserved in the 2,400-observation held-out test set (675 events, 1,725 censored). Table 5 summarizes the RSF model's discrimination and calibration performance on the test set.

Table 5. Random Survival Forest performance for predicting time-to-maturity

Metric	Value
C-index (Harrell)	0.5031
C-index (Uno, IPCW-weighted)	0.5031
Integrated Brier Score (IBS)	0.1953
Mean time-dependent AUC (day 23–30)	0.5029

Both C-index variants and the mean time-dependent AUC were close to 0.50, the value expected from an uninformative, chance-level ranking, and the Integrated Brier Score (0.1953) was close to the reference value obtained by a naive, covariate-free (Kaplan–Meier) estimate at this event rate. In other words, once elapsed rearing day is accounted for, the RSF model built on temperature, humidity, feed consumption, larval count, and larval weight discriminated individuals reaching maturity earlier from those maturing later, and calibrated their event probabilities, no better than chance within the range of conditions observed in this dataset.

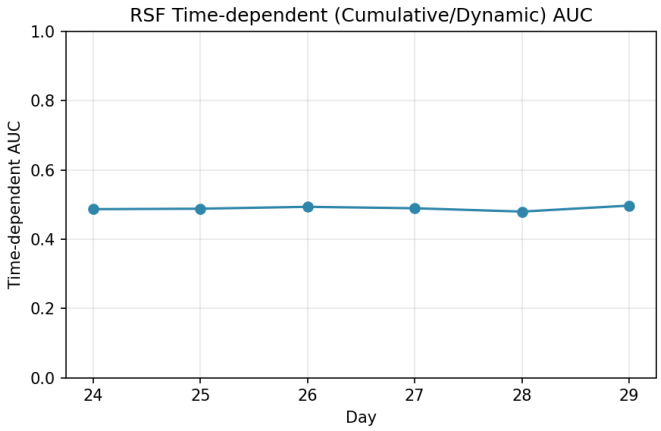


Figure 2. Time-dependent (cumulative/dynamic) AUC of the RSF model, days 23–30 of the test set, the window in which adult-stage events occurred.

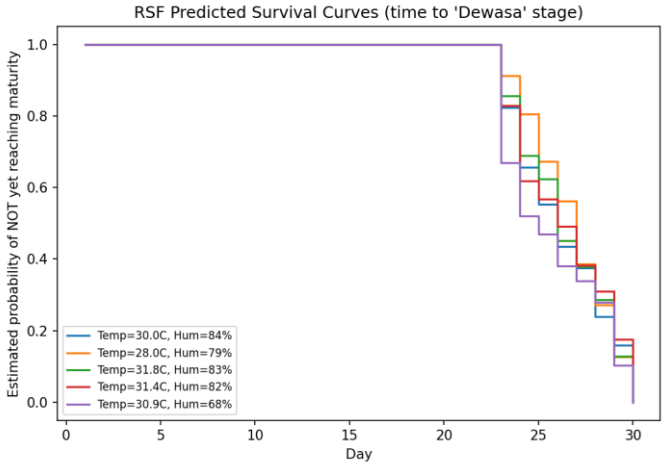


Figure 3. RSF-predicted survival curves (probability of not yet reaching the adult stage) for five representative temperature–humidity profiles

Figure 3 illustrates this limited discrimination directly: the predicted survival curves for five representative temperature–humidity profiles (27.8–31.8°C; 68–84% humidity) are nearly superimposed across the full 30-day window, with only minor separation appearing after day 23. This indicates that, within the narrow range of rearing conditions captured by this dataset, the model predicts essentially the same maturation trajectory regardless of the specific temperature or humidity profile, consistent with the near-chance C-index and AUC in Table 5.

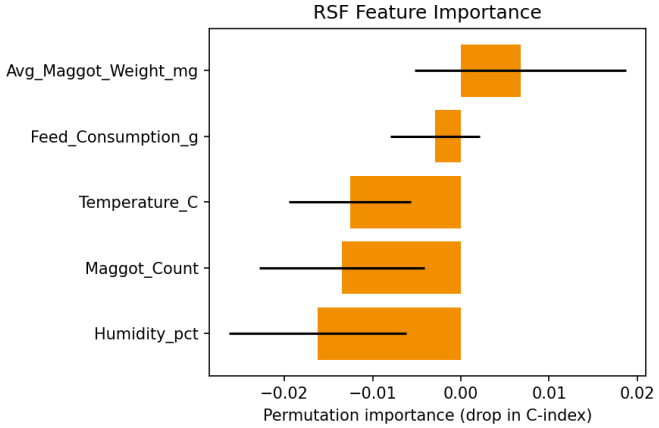


Figure 4. Permutation importance of RSF covariates, expressed as the mean change in C-index when each covariate is permuted (error bars: 95% interval over repeated permutations).

Table 6 reports the numerical permutation importance values underlying Figure 4. Permuting larval weight produced a mean increase in C-index of 0.0091 (95% CI: 0.0025 to 0.0157), and permuting humidity produced a mean change of −0.0094 (95% CI: −0.0129 to −0.0059); both intervals excluded zero, indicating a statistically detectable, non-random contribution from these two covariates. The intervals for temperature (0.0005; 95% CI: −0.0039 to 0.0049), feed consumption (−0.0002; 95% CI: −0.0036 to 0.0032), and larval count (−0.0025; 95% CI: −0.0071 to 0.0020) crossed zero and were therefore not distinguishable from noise. The negative sign for humidity is counterintuitive — permuting this covariate slightly improved rather than degraded discrimination — a pattern more consistent with sampling noise interacting with this specific tuned configuration than with a genuine protective effect of humidity on maturation timing, and it should not be read as evidence that humidity is biologically uninformative. It is important to reconcile these results with the near-chance overall C-index in Table 5: the permutation effects, while statistically detectable, were small in magnitude ($|\Delta$ C-index$| \leq 0.01$), so even the covariates

identified as informative moved the model only marginally above an essentially uninformative baseline. This pattern is consistent with the near-deterministic phase-versus-day relationship already noted in Section 3.3: if elapsed rearing day accounts for most of the variation in developmental stage, only a small residual signal remains for the rearing covariates to explain, which is what the RSF results in this section show directly for the specific transition to the adult stage.

Table 6. Permutation importance of RSF covariates: mean change in C-index and 95% confidence interval over 15 repeated permutations per covariate (CI via normal approximation)

Covariate	Mean ΔC-index	SD	95% CI Lower	95% CI Upper	Significant
Larval weight	0.0091	0.0130	0.0025	0.0157	Yes
Humidity	−0.0094	0.0069	−0.0129	−0.0059	Yes
Temperature	0.0005	0.0087	−0.0039	0.0049	No
Feed consumption	−0.0002	0.0067	−0.0036	0.0032	No
Larval count	−0.0025	0.0090	−0.0071	0.0020	No

Several factors likely limit discrimination in the current analysis and should be addressed before an operational RSF-based tool is developed. First, the covariates were recorded within a narrow range (temperature 27–32°C; humidity 65–85%) that may be too restricted to reveal the sharper temperature effects previously reported over wider ranges (Shumo et al., 2019; Tomberlin et al., 2009). Second, the analysis modeled only the final transition to the adult stage; a multistate formulation that models each instar-to-instar and instar-to-prepupa-to-pupa-to-adult transition individually may recover the stage-specific signal that a single terminal event, aggregated across the whole rearing period, can dilute. Third, substrate composition, a factor with well-documented influence on BSF development (Alagappan et al., 2026; Lalander et al., 2019), was not available as a covariate in this dataset and is a priority addition for the next iteration of the model.

4. Conclusions

This study developed and evaluated a Random Survival Forest model to predict time-to-maturity in the black soldier fly, supported by preliminary regression and classification benchmarking. Descriptive analysis of 8,000 rearing-monitoring observations across nine developmental phases confirmed substantial variability in growth and developmental timing, supporting the premise that calendar age alone is insufficient to characterize BSF development. Preliminary regression benchmarking showed that Random Forest and Gradient Boosting outperformed Linear Regression for total biomass, growth rate, and feed conversion ratio, although the near-perfect accuracy obtained for total biomass and for developmental-phase classification most likely reflects structural redundancy between predictors and outcomes rather than a fully validated predictive capability. The RSF model built on temperature, humidity, feed consumption, larval count, and larval weight to predict time-to-adult-stage achieved a C-index and mean time-dependent AUC close to 0.50 and an Integrated Brier Score of 0.195, indicating discrimination and calibration close to chance despite larval weight and humidity showing statistically detectable, if small and partly counterintuitive, permutation-importance effects. This negative result is itself informative: within the narrow range of rearing conditions captured here, environmental and feeding covariates added little predictive value beyond elapsed rearing day for the specific transition to maturity, and the current model is not yet ready to support operational rearing or harvest-timing decisions. Future work should evaluate covariates over a wider range of temperature and humidity, model each developmental transition individually within a multistate framework rather than a single terminal event, incorporate substrate composition as a covariate, and validate the resulting model on an independent rearing batch before it is translated into decision-support guidance for black-soldier-fly-based organic waste bioconversion.

Acknowledgements

The authors would like to express their sincere gratitude to the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia (Kemendiktisaintek) through the Directorate of Research and Community Service (DPPM) under the BIMA Research Grant Program 2026 for the financial support provided for this research entitled "Smart Farm Maggot Black Soldier Fly Using IoT-Machine Learning Technology for Organic Waste Reduction and Strengthening the National Circular Economy". The authors also acknowledge all parties who contributed to the successful implementation of this research through academic, technical, and institutional support.

References

Alagappan, S., Hoffman, L., & Cozzolino, D. (2026). Modelling black soldier fly larvae growth performance based on chemical composition and rearing conditions---A machine learning approach. *Waste and Biomass Valorization*. https://doi.org/10.1007/s12649-026-03541-y

Belperio, S., Cattaneo, A., Nannoni, E., Sardi, L., Martelli, G., Dabbou, S., & Meneguz, M. (2024). Assessing substrate utilization and bioconversion efficiency of black soldier fly (Hermetia illucens) larvae: Effect of diet composition on growth and development temperature. *Animals*, *14*(9), 1340. https://doi.org/10.3390/ani14091340

Berkowitz, M., Altman, R. M., & Loughin, T. M. (2024). Random forests for survival data: Which methods work best and under what conditions? *The International Journal of Biostatistics*, *20*(2), 315–345. https://doi.org/10.1515/ijb-2023-0056

Cardoso, L., Angelo, S. A., Bonilha, Y. P. G., Maia, F., Ribeiro, A., Curado, M., Fernandes, G., Parro, V., Cipparrone, F., Filho, A. C., & Toporcov, T. (2025). Methodology for Comparing Machine Learning Algorithms for Survival Analysis. *ArXiv*, *abs/2510.24473*. https://doi.org/10.48550/arxiv.2510.24473

Gärttling, D., & Schulz, H. (2022). Compilation of black soldier fly frass analyses. *Journal of Soil Science and Plant Nutrition*, *22*, 937–943. https://doi.org/10.1007/s42729-021-00703-w

Harrell, F. E., Lee, K. L., & Mark, D. B. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, *15*(4), 361–387.

Herremans, M., Wylant, B., Shumo, M., Shoo, A., James, M., Ngalesoni, F., Ndjovu, A., Mbuligwe, S., & Yhedgo, M. (2023). Advancing circular economy principles through wild black soldier flies. *AIMS Environmental Science*, *10*(6), 868–893. https://doi.org/10.3934/environsci.2023047

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *The Annals of Applied Statistics*, *2*(3), 841–860. https://doi.org/10.1214/08-AOAS169

Jaeger, B. C., Welden, S., Lenoir, K., Speiser, J. L., Segar, M. W., Pandey, A., & Pajewski, N. M. (2024). Accelerated and interpretable oblique random survival forests. *Journal of Computational and Graphical Statistics*, *33*(1), 192–207. https://doi.org/10.1080/10618600.2023.2231048

Kullan, A. R. K., Suresh, A., Choi, H. L., Neumann, E. G., & Hassan, F. (2024). Bioconversion of poultry litter into insect meal and organic frasstilizer using black soldier fly larvae as a circular economy model for the poultry industry: A review. *Insects*, *16*(1), 12. https://doi.org/10.3390/insects16010012

Lalander, C., Diener, S., Zurbrügg, C., & Vinnerås, B. (2019). Effects of feedstock on larval development and process efficiency in waste treatment with black soldier fly (Hermetia illucens). *Journal of Cleaner Production*, *208*, 211–219. https://doi.org/10.1016/j.jclepro.2018.10.017

Liu, Y., Zhou, S., Wei, H., & An, S. (2021). A comparative study of forest methods for time-to-event data: Variable selection and predictive performance. *BMC Medical Research Methodology*, *21*, 193. https://doi.org/10.1186/s12874-021-01386-8

Muinde, J., Tanga, C. M., Olukuru, J., Odhiambo, C., Tonnang, H. E. Z., & Senagi, K. (2023). Application of machine learning techniques to discern optimal rearing conditions for

improved black soldier fly farming. *Insects*, *14*(5), 479. https://doi.org/10.3390/insects14050479

Pickett, K. L., Suresh, K., Campbell, K. R., Davis, S., & Juarez-Colunga, E. (2021). Random survival forests for dynamic predictions of a time-to-event outcome using a longitudinal biomarker. *BMC Medical Research Methodology*, *21*, 216. https://doi.org/10.1186/s12874-021-01375-x

Pölsterl, S. (2020). scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn. *J. Mach. Learn. Res.*, *21*, 212:1-212:6.

Senevirathne, D. M., Yang, S.-I., Brandeis, C., & Hodges, D. G. (2024). Predicting time-to-harvest in mixed-species forests using a random survival forest algorithm. *Forest Ecosystems*, *11*, 100236. https://doi.org/10.1016/j.fecs.2024.100236

Shumo, M., Khamis, F. M., Tanga, C. M., Fiaboe, K. K. M., Subramanian, S., Ekesi, S., van Huis, A., & Borgemeister, C. (2019). Influence of temperature on selected life-history traits of black soldier fly (Hermetia illucens) reared on two common urban organic waste streams in Kenya. *Animals*, *9*(3), 79. https://doi.org/10.3390/ani9030079

Surendra, K. C., Tomberlin, J. K., van Huis, A., Cammack, J. A., Heckmann, L.-H. L., & Khanal, S. K. (2020). Rethinking organic wastes bioconversion: Evaluating the potential of the black soldier fly (Hermetia illucens (L.)) (Diptera: Stratiomyidae) (BSF). *Waste Management*, *117*, 58–80. https://doi.org/10.1016/j.wasman.2020.07.050

Tomberlin, J. K., Adler, P. H., & Myers, H. M. (2009). Development of the black soldier fly (Diptera: Stratiomyidae) in relation to temperature. *Environmental Entomology*, *38*(3), 930–934. https://doi.org/10.1603/022.038.0347

Uno, H., Cai, T., Pencina, M. J., D'Agostino, R. B., & Wei, L. J. (2011). On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine*, *30*(10), 1105–1117. https://doi.org/10.1002/sim.4154

Wang, P., Li, Y., & Reddy, C. K. (2019). Machine learning for survival analysis: A survey. *ACM Computing Surveys*, *51*(6), 1–36. https://doi.org/10.1145/3214306