sustainability

MDPI

Article

Neural Network for AI-Driven Prediction of Larval Protein Yield: Establishing the Protein Conversion Index (PCI) for Sustainable Insect Farming

Claudia L. Vargas-Serna *, Angie N. Pineda-Osorio, Carlos A. Gomez-Velasco, Jose Luis Plaza-Dorado and Claudia I. Ochoa-Martinez

Escuela de Ingeniería de Alimentos, Universidad del Valle, Cali 760001, Colombia; angie.pineda@correounivalle.edu.co (A.N.P.-O.); carlos.gomez.velasco@correounivalle.edu.co (C.A.G.-V.); jose.plaza@correounivalle.edu.co (J.L.P.-D.); claudia.ochoa@correounivalle.edu.co (C.I.O.-M.)
* Correspondence: claudia.vargas@correounivalle.edu.co

Abstract: The predictive capabilities of artificial intelligence for predicting protein yield from larval biomass present valuable advancements for sustainable insect farming, an increasingly relevant alternative protein source. This study develops a neural network model to predict protein conversion efficiency based on the nutritional composition of larval feed. The model utilizes a structured two-layer neural network with four neurons in each hidden layer and one output neuron, employing logistic sigmoid functions in the hidden layers and a linear function in the output layer. Training is performed via Bayesian regularization backpropagation to minimize mean squared error, resulting in a high regression coefficient (R = 0.9973) and a low mean-squared error (MSE = 0.0072401), confirming the precision of the model in estimating protein yields. This AI-driven approach serves as a robust tool for predicting larval protein yields, enhancing resource efficiency and promoting sustainability in insect-based protein production.

Keywords: neural networks; sustainable agriculture; protein yield; insect farming; artificial intelligence

check for updates

Academic Editor: Črtomir Rozman

Received: 15 November 2024
Revised: 16 December 2024
Accepted: 8 January 2025
Published: 16 January 2025

Citation: Vargas-Serna, C.L.; Pineda-Osorio, A.N.; Gomez-Velasco, C.A.; Plaza-Dorado, J.L.; Ochoa-Martinez, C.I. Neural Network for AI-Driven Prediction of Larval Protein Yield: Establishing the Protein Conversion Index (PCI) for Sustainable Insect Farming. *Sustainability* **2025**, *17*, 652. https://doi.org/10.3390/su17020652

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The black soldier fly (*Hermetia illucens*), an insect from the Stratiomyidae family and native to tropical and subtropical regions, has gained attention for its capacity for converting diverse organic waste into protein- and lipid-rich biomass. This process offers a sustainable solution to waste accumulation while aligning with circular economy principles, transforming agro-industrial residues into high-value resources. Although different substrates have been used for larval feeding, not all the meals obtained can be directed toward animal feed. According to European legislation, insects intended for animal feed cannot be reared on feces or human food waste. However, these meals could have other applications, such as the production of oils or biopolymers [1].

Refs. [2,3] represent an eco-friendly and economical approach to waste valorization. In a world where nutrient-dense food waste often contributes to landfill accumulation, repurposing these materials into high-quality proteins and lipids supports both environmental sustainability and food security goals.

The urgency of effective food waste management grows alongside global challenges such as climate change, population growth, and resource scarcity. Efficiently converting food waste into valuable resources addresses essential sustainability goals [4], including

responsible consumption and production. The diet of black soldier fly larvae directly impacts their protein yields and amino acid profiles, making nutrient-rich plant and animal residues ideal for larval feed [5,6]. Organic substrates like meat and coffee residues contain essential nutrients that enhance larval growth and protein content. Meat residues offer proteins and amino acids, while coffee residues supply carbohydrates, minerals, and antioxidants, creating a balanced nutrient profile that supports efficient protein conversion in larvae [7,8].

In recent years, the implementation of artificial intelligence and machine learning have played a significant role in various applications within the food industry, such as distribution, process optimization, crop supervision and management, quality assessments, market pricing [9], the formulation of high-efficiency diets that maximize protein yield sustainably, and in this case, the prediction of protein yield in black soldier fly larvae fed with organic byproducts with specific diet formulations.

Optimizing the protein yield of black soldier fly larvae from various organic wastes, however, requires sophisticated modeling approaches to capture the complex, nonlinear relationships between feed composition and protein production. Artificial intelligence (AI), particularly when using neural networks, offers powerful tools for modeling these interactions [10]. Neural networks are well-suited for handling the intricate dependencies among nutritional components, allowing for accurate predictions of protein yield based on a variety of feed compositions. Training the neural network on data from diverse substrate profiles allows the model to identify patterns and interactions that traditional approaches may overlook, facilitating a data-driven approach to predict feed formulations for maximum protein output. Unlike traditional statistical models, which rely solely on the data, artificial neural networks learn from training data. This enables better modeling of the interactions between the substrate and the larvae, allowing for more accurate predictions. This AI-driven methodology enhances precision in predicting protein yields while promoting more efficient resource utilization in accordance with circular economy principles.

In this study, a neural network model is applied to predict protein yield based on the nutritional composition of larval feed, presenting a novel, data-driven approach to optimizing larval protein production. The model was constructed with input variables, including moisture, protein, carbohydrates, lipids, and ash content, aiming for high predictive accuracy through the systematic evaluation of feed variables. By leveraging artificial intelligence's predictive power, this approach enables precise feed formulations that maximize protein output and promote resource efficiency.

The research is organized into three phases: (1) constructing a comprehensive database of substrate compositions and corresponding protein yields for black soldier fly larvae, (2) training and optimizing a neural network model to predict protein yields based on feed composition variables, and (3) validating the predictions of the model by comparing predicted and observed protein yields across different substrate combinations. This study demonstrates the potential of AI-driven models to enhance sustainable insect farming by optimizing waste-to-protein conversion, highlighting the role of insect farming in circular economy frameworks.

2. Materials and Methods

2.1. Database Construction

A database was built based on the available literature. A systematic search of articles that evaluate the effect of feed on the nutritional quality of black soldier fly larvae in databases such as ScienceDirect, Springer, MDPI, Scielo, and Google Scholar was made. A total of 28 articles containing the nutritional characterization of the larvae and the substrates pointed to 137 different substrates. This database was filtered, removing inconsistent or

incomplete data and retaining only the articles that reported the nutritional characteristics of the feed (dry matter, fat, crude protein, carbohydrates, ash) and the resulting crude protein percentage in the larvae. After filtering, the final database was composed of 105 data points. The data for neural network construction are reported on a wet basis. Table S1 presents the database.

2.2. Neural Network Training

The study utilized a neural network approach to predict protein yield from larval biomass based on the nutritional composition of the feed mixtures used to cultivate the larvae. Data was extracted from an Excel file, including nutritional parameters such as moisture, protein, carbohydrates, ash, and lipid content, as well as the crude protein measurement of the resulting larval flour. Input variables included moisture, protein, carbohydrates, ash, and lipids, while the output variable was the **Protein conversion index (PCI)**, which quantifies protein conversion efficiency (see Figure 1). The **PCI** is defined as follows:

$$\textbf{PCI} = \frac{\text{g protein in larval flour}}{\text{g dry matter in feed mixture}} \quad (1)$$

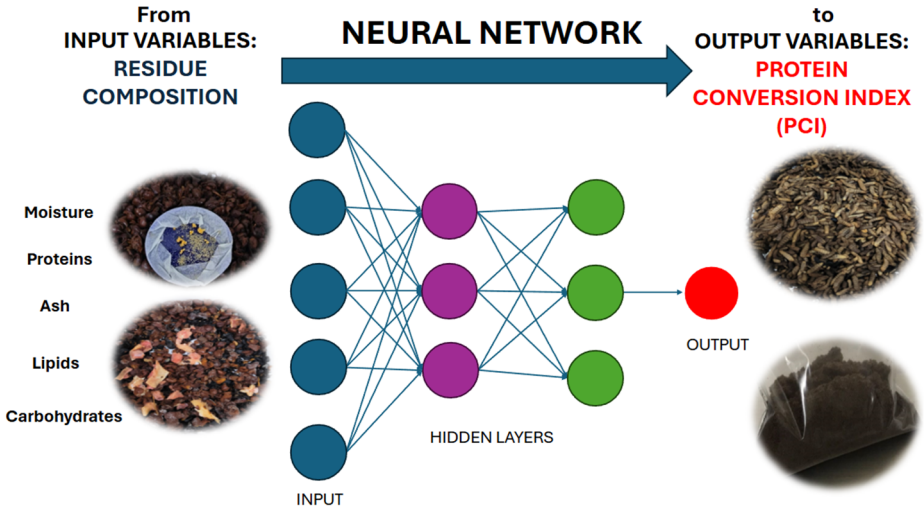


Figure 1. General structure of the neural network construction (research workflow).

The hyperparameters explored included six different hidden layer structures: [2], [3], [2 2], [3 3], [2 2 2], and [3 3 3]. Each configuration was selected to explore varying levels of model complexity to balance the prediction accuracy and computational efficiency. Two training algorithms were tested: Levenberg–Marquardt (*trainlm*) and the Bayesian regularization backpropagation (*trainbr*) were chosen for their respective strengths of convergence speed and capacity to handle complex optimization landscapes. Hidden layers employed either the hyperbolic tangent sigmoid (*tansig*) or the logistic sigmoid (*logsig*) transfer functions, with the output layer applying a linear function (*purelin*).

The dataset was divided into training (70%), validation (15%), and testing (15%) subsets, with indices randomly reorganized to ensure representative sampling and to minimize overfitting. Each combination of hyperparameters was evaluated based on the mean squared error (MSE) and the coefficient of determination (R^2), with a target R^2 of 0.9999. If this threshold was reached, the iteration process stopped; otherwise, the model ran for a maximum of 1000 iterations to identify the best-performing network configuration.

The **mean squared error (MSE)** was calculated as follows:

$$[\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2] \quad (2)$$

where y_t represents the actual values, $\hat{y}_t$ represents the predicted values, and n is the number of data points.

The **correlation coefficient (R)** was calculated to evaluate the linear relationship between the observed and predicted values as follows:

$$R = \frac{\sum_{t=1}^{n} (y_t - \overline{y})(\hat{y}_t - \overline{\hat{y}})}{\sqrt{\sum_{t=1}^{n} (y_t - \overline{y})^2 \sum_{t=1}^{n} (\hat{y}_t - \overline{\hat{y}})^2}} \quad (3)$$

where $\overline{y}$ is the mean of the actual values, $\overline{\hat{y}}$ is the mean of the predicted values, and n is the number of data points.

The **coefficient of determination (R^2)**, which quantifies the proportion of variance in the observed data explained by the model, was calculated as:

$$R^2 = 1 - \frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{n} (y_t - \overline{y})^2} \quad (4)$$

To balance accuracy with model simplicity, the network with the highest R^2 was selected, and a simpler model was also retained if it achieved an R^2 within 1% of the best model while having a lower complexity. Prioritizing simpler models enhances interpretability and reduces computational demands, making the model more practical for potential applications. The simplest model meeting these criteria was chosen.

All neural network modeling and calculations were conducted in MATLAB. The complete code used for data processing, model training, and analysis is provided in the Supplementary Materials (see file *"a0_neural_network.m"*).

The neural network equation, including the normalization and denormalization of data, is presented in a generalized form below, which applies regardless of the number of neurons, layers, or activation functions used:

Normalization of inputs: Each input feature x_i is normalized to the range $[-1, 1]$ using the following:

$$x_i' = 2 \cdot \frac{x_i - x_{\min,i}}{x_{\max,i} - x_{\min,i}} - 1 \quad (5)$$

where x_i is the original value of the i-th input feature, $x_{\min,i}$ and $x_{\max,i}$ are the minimum and maximum values of the i-th input feature across the training data, and x_i' is the normalized input feature.

Generalized form for layer calculations: for a neural network with L layers (where $L - 1$ are hidden layers and the L-th layer is the output layer), the output of each layer l (for $l = 1, 2, \ldots, L$) can be represented as follows:

$$h^{(l)} = f^{(l)}\left(W^{(l)} \cdot h^{(l-1)} + b^{(l)}\right) \quad (6)$$

where

- $h^{(l)}$ is the output vector of layer l;
- $f^{(l)}$ is the activation function applied at layer l (e.g., tansig, logsig, or purelin);
- $W^{(l)}$ is the weight matrix for layer l;
- $b^{(l)}$ is the bias vector for layer l;
- $h^{(0)}$ corresponds to the vector of normalized input features x_i'.

Final output layer (denormalized): after computing the output of the final layer $h^{(L)}$, which represents the normalized network output y', the predicted Protein Conversion Index (PCI) in its original scale is obtained by denormalizing y':

$$\mathbf{PCI} = \frac{(y' + 1)}{2} \cdot (y_{max} - y_{min}) + y_{min} \quad (7)$$

where

- y_{max} and y_{min} are the maximum and minimum values of the PCI in the training data;
- $y' = h^{(L)}$ is the final output of the neural network in the normalized range.

Additionally, a MATLAB GUI tool, createNNPredictionTool, was developed for practical applications, allowing users to input values for moisture, protein, carbohydrates, ash, and lipids to obtain the predicted PCI, the dry matter of the mixture, and the crude protein in flour. This tool facilitates the application of the trained model and visualizes predicted outcomes as compared to real data. The full code for the createNNPredictionTool is available in the Supplementary Materials (see file *"a8_createNNPredictionTool.m"*).

2.3. Sensitivity Analysis Methods

Two complementary sensitivity analysis methods were applied to evaluate the impacts of input variables on the output of the neural network used in this study. These analyses help elucidate the influence of each input feature on the model predictions, highlighting which variables most significantly affect the PCI.

2.3.1. Gradient-Based Sensitivity Analysis

The first method involved a gradient-based sensitivity analysis. This approach calculates the numerical gradient of the network output with respect to each input feature. The gradient provides a direct measure of how variations in input features influence the network's predictions. A small perturbation δ is introduced to each input, and the network output is recorded for both positive and negative perturbations. This allows for an estimation of the sensitivity gradient using the central difference formula.

The numerical gradient (g_i) for each input feature (x_i) is expressed as follows:

$$g_i = \frac{f(x_i + \delta) - f(x_i - \delta)}{2\delta} \quad (8)$$

where f(x) represents the neural network output, and δ is a small perturbation value. The mean gradient for each input feature across all samples was then calculated to summarize its influence on the output as follows:

$$g_i = \frac{1}{N} \sum_{j=1}^{N} g_{ij} \quad (9)$$

This gradient-based analysis highlights the linear sensitivity of the neural network model to small perturbations in each input variable, providing insight into the contribution of each input to the model predictions.

2.3.2. Permutation Sensitivity Analysis

The second method involved permutation sensitivity analysis. This technique evaluates the effect of each input feature by permuting its values randomly while maintaining all other inputs as constant. This process is repeated for each input feature, and the mean squared error (MSE) between the original predictions of the model and the predictions obtained with permuted inputs is computed.

The sensitivity score (S_i) for each input feature (x_i) is defined as follows:

$$S_i = MSE_{permuted} - MSE_{baseline} \quad (10)$$

where $MSE_{baseline}$ denotes the mean squared error of the model with the original inputs, and $MSE_{permuted}$ represents the mean squared error after permuting the values of the input feature (x_i). The sensitivity score quantifies the significance of the feature regarding the model accuracy.

Permutation sensitivity analysis offers a comprehensive method for evaluating the importance of each input variable in a non-linear manner, capturing the effects of disrupting input values and their subsequent impacts on the model performance.

Both sensitivity analysis methods were implemented using MATLAB. The respective codes for the gradient-based and permutation sensitivity analyses are available in the Supplementary Materials (files "*a7_gradientSensitivityAnalysis.m*" and "a*7_permutationSensitivityAnalysis*", respectively). The gradient-based method focuses on local sensitivities by applying small perturbations, while the permutation method provides a more global perspective of variable importance through the evaluation of model performance after disrupting input features.

2.3.3. Validation

The data for model validation was obtained through real experiments that evaluated the protein content in meal derived from larvae fed a mix of meat and coffee residues. The analyzed mixtures contained between 65–75% coffee residue and 25–35% meat residue (a mix of chicken and hen). The substrate was characterized by following the official AOAC [11] methods for protein, lipids, moisture, and ash contents. Protein in the larval meal was measured using the Kjeldahl method.

3. Results

The results are organized into four sections. Section 3.1 describes the construction of the database used for training the neural network, detailing the selection and filtering process for substrates characterized for black soldier fly larvae. Key nutritional components, including moisture, protein, carbohydrates, ash, and lipid content, are emphasized as primary inputs for model training. Section 3.2 provides a comprehensive analysis of the neural network training process, focusing on the selection of network configurations, transfer functions, and training algorithms. This section also evaluates performance using MSE and R^2 to measure predictive accuracy across training, validation, and test sets. Section 3.3 discusses the sensitivity analysis conducted to evaluate the influence of each input variable on model predictions, using gradient-based and permutation sensitivity methods to evaluate the impact of individual nutritional factors on the PCI. Lastly, Section 3.4 presents the experimental validation of the neural network model by comparing the predicted protein conversion index (PCI) values with the observed experimental data, examining the reliability of the model for practical applications in optimizing protein yield in insect farming.

3.1. Training Neural Network

The neural network configuration was trained and evaluated to determine its capability to predict the protein content in black soldier fly larvae flour based on the nutritional compositions of the feed mixtures. Figure 2 presents the regression plot of the initial predictions of the network against the target values, with a close alignment represented by Output $\approx 0.99 \cdot \text{Target} + 0.005$. This alignment indicates a high level of precision in the predictions, suggesting that the initial network captures relevant patterns in feed composition that affect protein yield. Such predictive capability has implications for predicting feed mixtures in order to increase protein yield in insect farming.

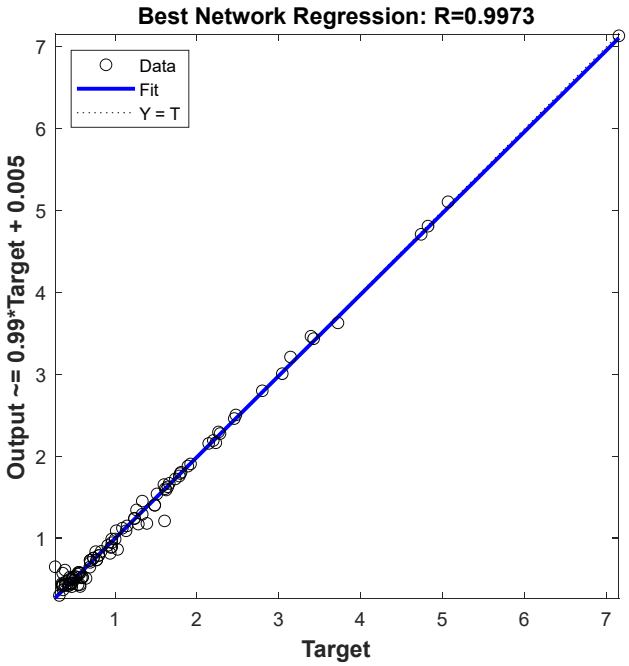


Figure 2. Regression plot of the initial network predictions against the target values.

To achieve this predictive accuracy, several configurations were tested, varying the number of hidden layers and neurons to balance complexity and performance. The model with the highest coefficient of determination (R^2) was selected, while simpler configurations were also considered if their R^2 values were within 1% of the best model. This strategy ensured that the final network achieved high accuracy with minimal computational demands, enhancing interpretability and practical applicability. The selected model, consisting of two hidden layers with four neurons each, demonstrated reliable performance across training, validation, and test sets.

Table 1 provides a summary of the MSE and R^2 for the initial network across the training, validation, and test sets. The low MSE and high R^2 values (R^2 of up to 0.99263 in the test set) indicate strong predictive accuracy, particularly in the test set where the generalization ability is evaluated. These metrics suggest robustness in the predictions of the model of protein content in the new data, a critical aspect for scalability in real-world insect farming applications.

Table 1. Performance summary of initial network.

Dataset	MSE	R^2
Training	0.0083074	0.99461
Validation	0.0029954	0.99306
Test	0.0062904	0.99904
Overall Best	0.0072401	0.99263

Figure 3 presents a comparison of the predicted versus observed values across training, validation, and test sets for the initial network. These plots confirm the reliable performance of the model in each data subset, indicating that it learned relevant patterns for predicting protein yield without compromising generalization.

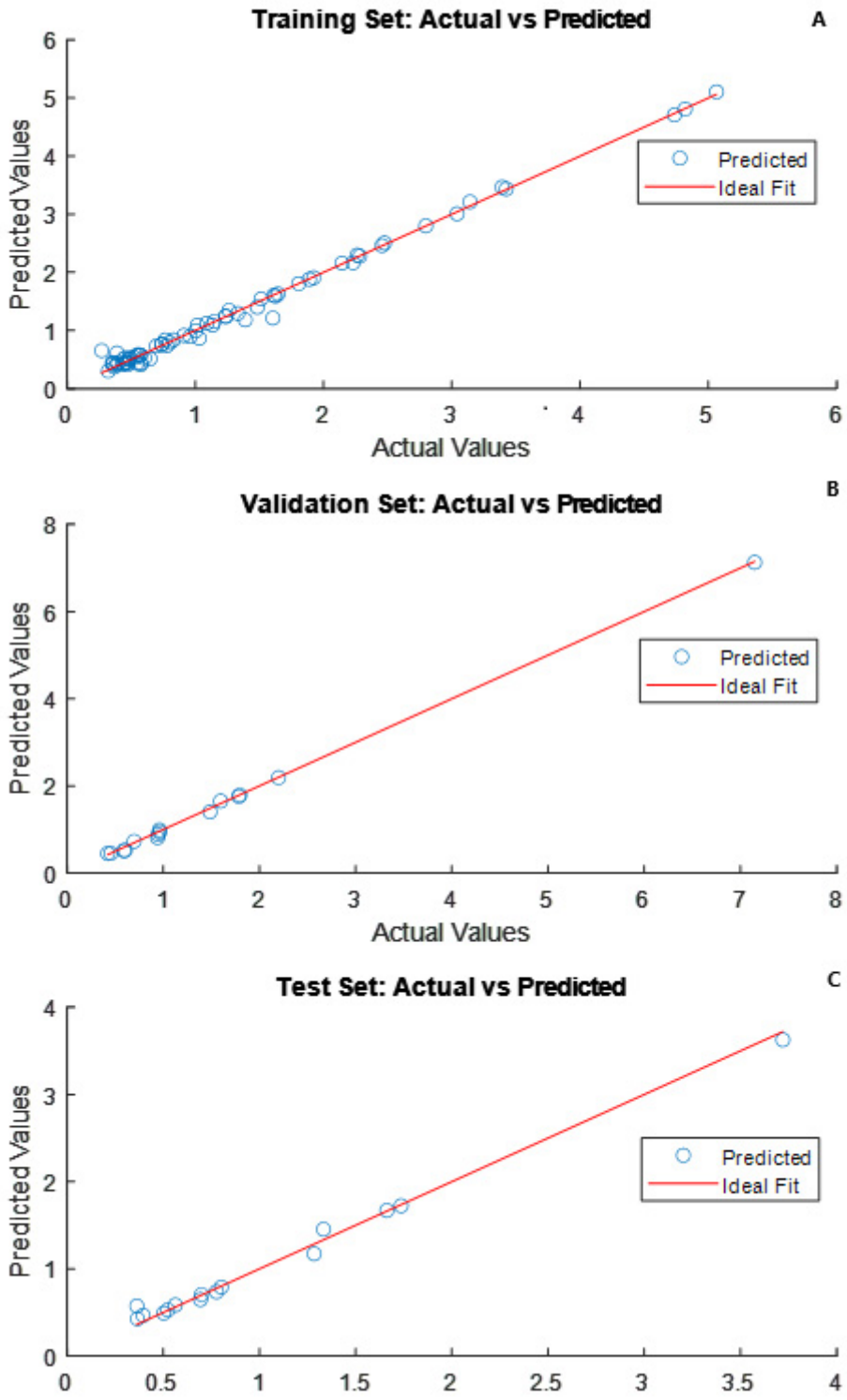


Figure 3. Predicted versus observed values in the initial network for (**A**) training, (**B**) validation, and (**C**) test sets.

The architecture of the neural network (see Figure 4), featuring two layers with four neurons in each hidden layer and one neuron in the output layer, demonstrates that a moderately complex structure is adequate to achieve high accuracy. This network architecture, combined with the *logsig* activation function in the hidden layers and *purelin* in the output layer, enables a non-linear transformation of input data, allowing the model to capture intricate relationships between feed composition variables. The Bayesian regularization algorithm (*trainbr*) was used for training, which helps reduce overfitting by

penalizing complex models, enhancing the applicability of the model in potentially noisy data conditions.

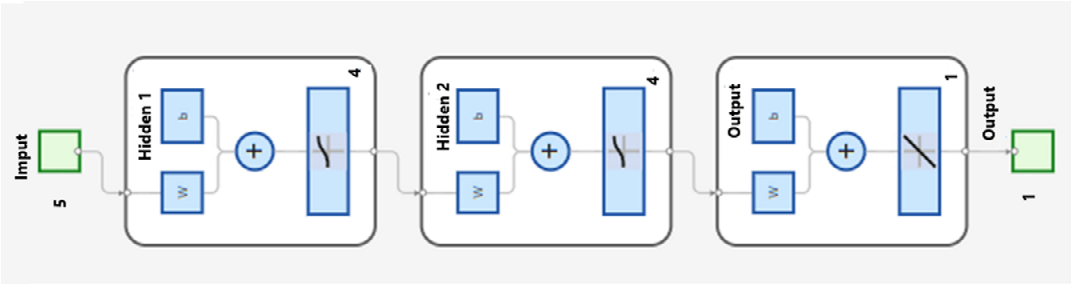


Figure 4. Network configuration.

Tables S2–S5 provides specific weights and biases for each layer in the initial network. These values illustrate how the network processes the input variables for accurate prediction. The compact symbolic equation representing the neural network is as follows:

$$\text{Output} = \text{purelin}(W3 \cdot (\text{logsig}(W2 \cdot (\text{logsig}(W1 \cdot x + b1)) + b2)) + b3) \quad (11)$$

3.2. Neural Network Prediction Tool

To facilitate practical application, a Neural Network Prediction Tool (Figure 5) was developed. This tool accepts user input for values such as moisture, protein, carbohydrates, ash, and lipid content, generating predictions for the PCI, dry matter content, and crude protein content in the resulting flour.

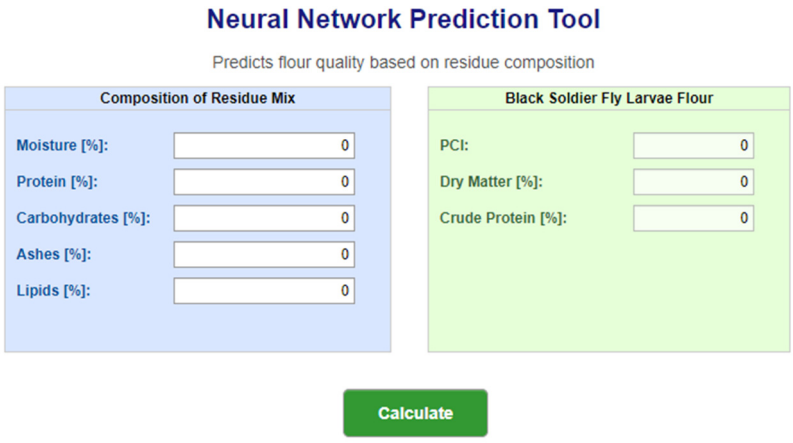


Figure 5. Interface of the Neural Network Prediction Tool for the real-time prediction of protein conversion efficiency.

The use of a neural network model within this tool supports real-time applications, particularly for users requiring rapid, accurate predictions without extensive computational resources. This predictive tool provides a user-friendly interface, integrating input fields for feed composition and output fields for predicted results. Its design supports efficient decision-making in feed formulation and production practices.

3.3. Sensitivity Analisys

The sensitivity analysis aimed to examine the influence of each input variable on the predictions of the PCI generated by the neural network model. Two complementary methods, gradient-based sensitivity analysis and permutation sensitivity analysis, were used to evaluate the significance of each input feature.

The gradient-based sensitivity analysis involved calculating the numerical gradient of the output of the neural network with respect to each input feature. This approach introduced small perturbations to each input variable to observe changes in the predictions of the model, with the central difference formula used to derive gradient values. For each input, positive and negative perturbations were applied independently, and the average sensitivity was calculated across all samples.

The results, presented in Figure 6, reveal that Input Feature 1 (moisture) had the highest mean gradient, approximately 0.07, indicating a substantial influence on the PCI predictions. Input Feature 2 (protein) demonstrated a lesser yet notable effect. Conversely, Input Features 3, 4, and 5 (carbohydrates, lipids, and ashes, respectively) exhibited minimal mean gradients, suggesting a limited impact on the output of the model. These findings indicate that the neural network model depend primarily on Input Feature 1 with Input Feature 2 also contributing, though to a lesser degree.

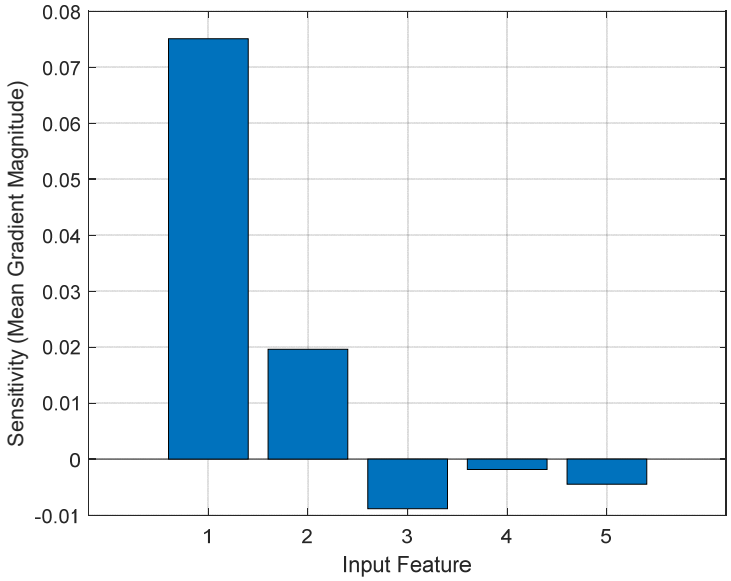


Figure 6. Sensitivity of the network output to each input feature (gradient method).

The permutation sensitivity analysis examined the impact of each input variable by randomizing its values across samples while other inputs remained constant. This process allowed for the measurement of changes in the mean MSE between the baseline predictions and the predictions generated with permuted inputs, thereby quantifying the significance of each input feature.

As shown in Figure 7, permuting Input Feature 1 (moisture) produced the largest increase in MSE, approximately 3.9, highlighting its critical role in the accuracy of the PCI predictions. Input Feature 3 (carbohydrates) also had a considerable influence, as evidenced by a moderate increase in MSE. The remaining input features—2, 4, and 5—induced minimal changes in MSE, confirming their lesser impacts on model predictions. This analysis reinforces the centrality of Input Feature 1, with Input Feature 3 also contributing to the prediction accuracy.

Both the gradient-based and permutation sensitivity analyses indicate a consistent hierarchy of influence among the input features, with Input Feature 1 emerging as the most impactful variable for PCI prediction. Input Feature 3, although less influential, also contributes meaningfully to the accuracy of the model

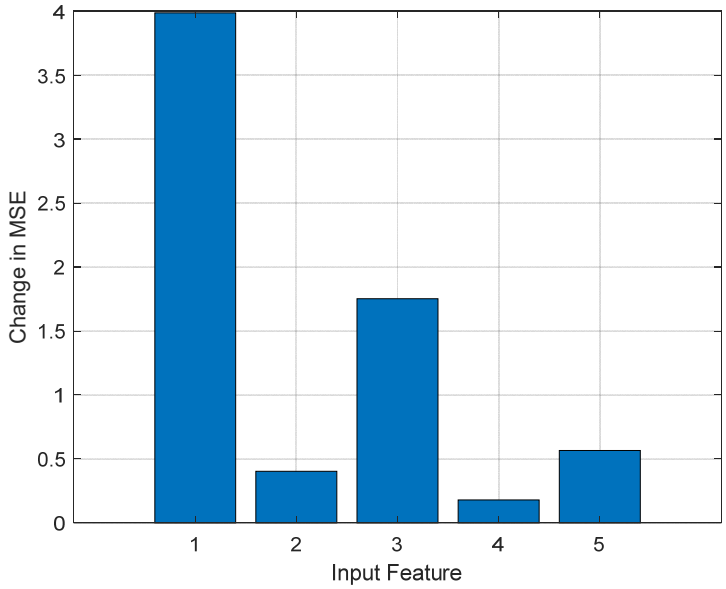


Figure 7. Permutation sensitivity analysis for each input feature.

3.4. Validation Results

The validation of the neural network model was explored by comparing experimental values of the PCI and protein content with their predicted equivalents. Table 2 presents the observed experimental PCI and protein values together with the predicted values generated by the model.

Table 2. Comparison of the experimental and predicted values for the protein conversion index (PCI) and protein content in black soldier fly larvae under different feed compositions.

Moisture (%)	Protein (%)	Carb. (%)	Ashes (%)	Lipids (%)	Predicted PCI	Exp. PCI	Exp. Protein (%)	Predicted Protein (%)	Relative Error
77.63	5.1608	15.091	0.36463	1.7538	1.7052	1.4256	31.89	38.146	19.617
76.14	6.5186	14.452	0.36983	2.5196	1.5989	1.6685	39.81	38.151	4.168
76.89	5.8214	14.795	0.36745	2.1261	1.5802	1.6257	37.57	36.52	2.796
77.26	5.4872	14.949	0.36611	1.9374	1.6292	1.4222	32.34	37.048	14.558
76.52	6.1658	14.626	0.36864	2.3198	1.5678	1.3194	30.98	36.812	18.826
76.14	6.5186	14.452	0.36983	2.5196	1.5989	1.6685	39.81	38.151	4.168
76.89	5.8214	14.795	0.36745	2.1261	1.5802	1.5257	35.26	36.52	3.5721
76.89	5.8214	14.795	0.36745	2.1261	1.5802	1.6811	38.85	36.52	5.9986

The validation results for the neural network model, presented in Table 2, indicate good accuracy in predicting the PCI and protein contents in black soldier fly larvae. The predicted PCI values ranged from 1.5678 to 1.7052, closely approximating the experimental values, which spanned from 1.3194 to 1.6811. This agreement reflects the capacity of the model to capture key patterns in protein conversion efficiency based on feed composition, supporting reliable performance across various feed conditions.

Similarly, the protein content predictions achieved a high level of accuracy. The predicted values ranged from 36.52 to 38.151, aligning well with experimental values between 30.98 and 39.81. Relative errors as low as 2.796% and 3.5721% reflect the reliability of the model's performance in predicting protein yields for specific feed compositions, confirming its utility as a predictive tool for enhancing protein production projections in insect farming applications.

4. Discussion

This study highlights the potential of artificial intelligence, particularly neural network models, to advance sustainable insect farming through the accurate prediction and optimization of protein yield from larval biomass. This model represents a practical application of circular economy principles, addressing critical issues in food production and organic waste valorization. This approach supports the diversification of protein sources by converting organic byproducts into high-protein biomass, contributing to circular economy objectives by reducing reliance on conventional feed resources and promoting more sustainable food systems.

The neural network demonstrated high predictive accuracy, with a $R^2 = 0.99263$ and an MSE of 0.0072401. This precision reflects the ability of the model to capture complex relationships between substrate composition and protein yield, enabling reliable predictions for implementation within insect farming systems. Previous research conducted by Cruz-Tirado et al. [12] demonstrated that models like the Partial Least Squares Regression (PLSR) and the Support Vector Machine Regression (SVMR) are capable of predicting protein content in black soldier fly larvae, achieving R^2 values of up to 0.847. These findings support the possibility of predicting protein content in insect larvae, providing a strong foundation for the developed neural network model, which achieved superior predictive performance. Such accuracy could facilitate the optimization of feed formulations, allowing for targeted adjustments that maximize protein output. While the neural network approach provided robust predictive accuracy, comparative analyses with other machine learning methods, such as PLSR and SVMR, were not performed directly in this study due to constraints in data availability and computational resources. Neural networks were prioritized because of their ability to model intricate interactions between feed composition variables and protein yields more effectively. Future work will incorporate these comparative analyses to comprehensively evaluate the performance of the model and its robustness. This will ensure that the most effective predictive tool is identified and further validate the applicability of the model in sustainable insect farming.

Several potential sources of error may influence the accuracy and reliability of the predictions of the model. Data quality issues represent a significant factor, including variability in the substrate composition, measurement inaccuracies in the feed's nutrient content, and inconsistencies in experimental conditions. Such variability can introduce noise that affects the model training and validation outcomes. Additionally, the dataset size and diversity may limit the ability of the model to generalize to new conditions. Expanding the dataset with more observations could enhance the robustness of the model and reduce the impact of data variability.

The model's limitations also need to be considered, particularly the constrained ability of the neural network to extrapolate beyond the range of the training data. However, it is worth noting that the model was validated with experimental data, demonstrating high accuracy within the tested range. This experimental validation, along with the internal validation during the training process (using 15% of the data for validation), provides confidence in the reliability of the model for the feed compositions evaluated in this study. Despite this, the performance of the model for extreme or unseen feed compositions remains uncertain. Regularization techniques and additional validation with unseen datasets could further mitigate this limitation and enhance generalizability.

Environmental factors, such as variations in temperature, humidity, and larval growth conditions, can influence the protein yield and may affect model predictions. These factors introduce additional complexity and should be accounted for in future studies by integrating environmental parameters into the model or conducting experiments under controlled conditions to isolate their effects.

Furthermore, the model's robustness under varying operating conditions has not been fully explored. Testing the model across different scenarios, such as varying feedstock types and operational scales, is essential to ensure its applicability and reliability in real-world contexts. Future work should include sensitivity analyses and external validations to assess how well the model adapts to diverse conditions and to identify potential failure points.

The accuracy of the model protein yield predictions facilitates the development of more sustainable and resource-efficient practices. This accuracy enables targeted feed formulations that optimize resource use and minimize waste. These improvements have broader implications for waste reduction and responsible resource management, aligning with global sustainability goals.

This optimization enhances resource efficiency and supports waste reduction, directly contributing to the United Nations' Sustainable Development Goals (SDGs), particularly SDG 12 (Responsible Consumption and Production) and SDG 2 (Zero Hunger), enabling sustainable food production and reducing dependence on unsustainable inputs in feed formulations.

Additionally, the support model for optimized protein conversion from organic waste aligns with SDG 13 (Climate Action), offering an alternative protein source involving potentially lower greenhouse gas emissions and a reduced environmental footprint as compared to traditional livestock farming. The use of black soldier fly larvae to upcycle agricultural and food waste into protein-rich biomass addresses waste management and resource efficiency challenges in an innovative way, demonstrating how insect farming contributes to a resilient and sustainable food system. Albalawneh et al. [13] corroborates the prior statement, emphasizing that the utilization of black soldier fly larvae's capacity to effectively reduce diverse types of waste is also crucial in sustainable waste management and administrative protocols. This approach reflects a systems-based perspective on circular economy principles, where resources are continually repurposed to minimize waste and resource depletion.

The sensitivity analysis conducted in this study revealed that the moisture and protein contents in feed mixtures significantly influence protein yield predictions, with the carbohydrate content, although less influential, contributing meaningfully. The literature confirms these results, as black soldier fly larvae's development and bioconversion is highly affected by dietary protein (improves) and carbohydrate (delay) contents [14–16]; this outcome helps to highlight specific areas for feed formulation optimization. Through the recognition of these critical variables, this study offers actionable insights for guiding feedstock adjustments, enabling insect farming systems to achieve greater efficiency with minimal resource use. This targeted approach to feed formulation enhances protein yield and supports sustainable resource management in line with SDG 12, emphasizing responsible resource use and waste reduction.

Moreover, the development of a neural network prediction tool in this study illustrates the potential of digital technologies for advancing sustainable practices within food production. This tool enables real-time predictions of protein conversion efficiency based on feed composition inputs, facilitating prompt adjustments to optimize production outcomes. Adoption of such AI-driven tools aligns with Industry 4.0 principles, promoting data-driven decision-making, optimization, scalability, possible automation, and process control in food systems. This innovation supports SDG 9 (Industry, Innovation, and Infrastructure), incorporating advanced digital infrastructure and AI into agricultural practices and driving technological advancement in the food sector.

The broader implications of this research extend beyond insect farming, offering a scalable model adaptable to various bio-waste valorization applications within a circular economy framework. The successful demonstration of this approach to convert organic waste into high-value protein has relevance for multiple industries, including animal feed,

aquaculture, and human food production, where interest in sustainable alternative protein sources is growing. This model could be applied to other organic byproducts, helping to divert waste from landfills, reduce environmental pollution, and support a more sustainable and circular food system.

While the neural network model presented in this study demonstrates significant potential, further research is needed to evaluate its performance across a broader range of organic substrates. Expanding the dataset to include varied feed compositions would improve the model's robustness, allowing for applications in diverse environmental and operational contexts. Integrating hybrid modeling approaches that combine neural networks with other AI techniques, such as ensemble learning or support vector machines, could further enhance prediction accuracy and yield more detailed insights into interactions between feed components and protein yield. This research aligns with SDG 9 by using innovation and technological progress in sustainable food production.

The parallel exploration of the socioeconomic impact of scaling up AI-driven insect farming in developing economies could reveal further benefits aligned with SDG 8 (Decent Work and Economic Growth). This approach could open new opportunities in agriculture and food production, fostering job creation and supporting economic resilience, especially in communities where sustainable food production methods are urgently needed.

5. Conclusions

This study demonstrates the potential of artificial intelligence, specifically neural networks, as a valuable tool for optimizing protein production from black soldier fly larvae. The model developed here achieves high predictive accuracy in estimating protein yields based on substrate composition, offering a practical approach to enhance the efficiency and sustainability of insect farming. This AI-driven method supports a circular economy framework by transforming organic waste into high-quality protein sources, reducing dependence on conventional feedstocks and minimizing the environmental impact. These findings reveal the broader applicability of AI in advancing sustainable protein resources and positioning insect farming as a viable component of future food systems focused on resource efficiency and ecological balance, offering a feasible business model that could lead to broader economic benefits and providing communities with a sustainable livelihood while contributing to environmental sustainability.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/su17020652/s1, Table S1: Database; Table S2: Layer 1 weights (W1); Table S3: Layer 2 weights (W2); Table S4: Output layer (W3); and Table S5: Biases for each layer [17–38].

Author Contributions: Conceptualization, C.L.V.-S., A.N.P.-O., C.A.G.-V., J.L.P.-D. and C.I.O.-M.; methodology, C.L.V.-S., A.N.P.-O., C.A.G.-V. and J.L.P.-D.; software, C.A.G.-V. and J.L.P.-D.; validation, C.L.V.-S., A.N.P.-O., C.A.G.-V. and J.L.P.-D.; formal analysis, C.L.V.-S., A.N.P.-O., C.A.G.-V., J.L.P.-D. and C.I.O.-M.; investigation, C.L.V.-S. and A.N.P.-O.; resources, C.L.V.-S., A.N.P.-O., J.L.P.-D. and C.I.O.-M.; data curation, C.L.V.-S.; writing—original draft preparation, C.L.V.-S., A.N.P.-O., C.A.G.-V. and J.L.P.-D.; writing—review and editing, J.L.P.-D.; visualization, C.L.V.-S. and J.L.P.-D.; supervision, J.L.P.-D. and C.I.O.-M.; project administration, C.L.V.-S.; funding acquisition, C.L.V.-S., A.N.P.-O., J.L.P.-D. and C.I.O.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Orchids Women in Science Program 935 of the Ministry of Science, Technology, and Innovation of Colombia.

Institutional Review Board Statement: The animal study protocol was approved by Ethics Committee of Universidad de Valle (protocol code 010 and date of approval 22 May 2023).

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors thank the Orchids Women in Science Program of the Ministry of Science, Technology, and Innovation for their support in the development of this project.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. *Reglamento (UE) 2017/893*; European Union: Brussels, Belgium, 2017; pp. 92–116.
2. Siddiqui, S.A.; Ristow, B.; Rahayu, T.; Putra, N.S.; Widya Yuwono, N.; Nisa', K.; Mategeko, B.; Smetana, S.; Saki, M.; Nawaz, A.; et al. Black Soldier Fly Larvae (BSFL) and Their Affinity for Organic Waste Processing. *Waste Manag.* **2022**, *140*, 1–13. [CrossRef]
3. Bermúdez-Serrano, I.M.; Sánchez-Velázquez, O.A. Comprehensive Utilization of the Black Soldier Fly: Bioconversion, Sustainability, and Emerging Challenges. *Sci. Agropecu.* **2023**, *14*, 571–590. [CrossRef]
4. United Nations. *The 2030 Agenda and the Sustainable Development Goals An Opportunity for Latin America and the Caribbean Goals, Targets and Global Indicators*; United Nations: San Francisco, CA, USA, 2018.
5. Taufek, N.M.; Mohamad Zulkifli, N.F.N.; Hamizah, A.N. Upcycling of Food Waste Generated from the Fresh Market by Utilising Black Soldier Fly Larvae: Influence on Growth, Bioconversion, and Nutritional Composition. *J. Environ. Manag.* **2024**, *349*, 119467. [CrossRef] [PubMed]
6. Laksanawimol, P.; Anukun, P.; Thancharoen, A. Use of Different Dry Materials to Control the Moisture in a Black Soldier Fly (*Hermetia illucens*) Rearing Substrate. *PeerJ* **2024**, *12*, e17129. [CrossRef]
7. Gemechu, F.G. Embracing Nutritional Qualities, Biological Activities and Technological Properties of Coffee Byproducts in Functional Food Formulation. *Trends Food Sci. Technol.* **2020**, *104*, 235–261. [CrossRef]
8. Cañas, S.; Rebollo-Hernanz, M.; Cano-Muñoz, P.; Aguilera, Y.; Benítez, V.; Braojos, C.; Gila-Díaz, A.; Rodríguez-Rodríguez, P.; Cobeta, I.M.; Pablo, Á.L.L.d.; et al. Critical Evaluation of Coffee Pulp as an Innovative Antioxidant Dietary Fiber Ingredient: Nutritional Value, Functional Properties, and Acute and Sub-Chronic Toxicity. *Proceedings* **2020**, *70*, 65. [CrossRef]
9. Przybył, K.; Gawrysiak-Witulska, M.; Bielska, P.; Rusinek, R.; Gancarz, M.; Dobrzański, B.; Siger, A. Application of Machine Learning to Assess the Quality of Food Products—Case Study: Coffee Bean. *Appl. Sci.* **2023**, *13*, 10786. [CrossRef]
10. Sheikh, A.; Simske, S.J.; Chong, E.K.P. Evaluating Artificial Intelligence Models for Resource Allocation in Circular Economy Digital Marketplace. *Sustainability* **2024**, *16*, 10601. [CrossRef]
11. *AOAC Official Methods of Analysis*; AOAC: Rockville, MD, USA, 2000.
12. Cruz-Tirado, J.P.; Amigo, J.M.; Barbin, D.F. Determination of Protein Content in Single Black Fly Soldier (*Hermetia illucens* L.) Larvae by near Infrared Hyperspectral Imaging (NIR-HSI) and Chemometrics. *Food Control* **2023**, *143*, 109266. [CrossRef]
13. Albalawneh, A.; Hasan, H.; Alarsan, S.F.; Diab, M.; Znaimah, S.A.; Sweity, A.; Aladwan, M.M.; Sharman, B.; Alalwan, A.M.; AlBalawnah, Y.; et al. Evaluating the Influence of Nutrient-Rich Substrates on the Growth and Waste Reduction Efficiency of Black Soldier Fly Larvae. *Sustainability* **2024**, *16*, 9730. [CrossRef]
14. Cattaneo, A.; Belperio, S.; Sardi, L.; Martelli, G.; Nannoni, E.; Meneguz, M.; Dabbou, S. A First Step Towards Black Soldier Fly Larvae (Diptera: Stratiomyidae) Welfare by Considering Dietary Regimes (Part I). *Insects* **2024**, *15*, 817. [CrossRef]
15. Peguero, D.A.; Gold, M.; Vandeweyer, D.; Zurbrügg, C.; Mathys, A. A Review of Pretreatment Methods to Improve Agri-Food Waste Bioconversion by Black Soldier Fly Larvae. *Front. Sustain. Food Syst.* **2022**, *5*, 745894. [CrossRef]
16. Nguyen, T.T.X.; Tomberlin, J.K.; Vanlaerhoven, S. Influence of Resources on *Hermetia illucens* (Diptera: Stratiomyidae) Larval Development. *J. Med. Entomol.* **2013**, *50*, 898–906. [CrossRef]
17. Hosseindoust, A.; Ha, S.H.; Mun, J.Y.; Kim, J.S. A Metanalysis to Evaluate the Effects of Substrate Sources on the Nutritional Performance of Black Soldier Fly Larvae: Implications for Sustainable Poultry Feed. *Poult. Sci.* **2024**, *103*, 103299. [CrossRef]
18. Lalander, C.; Diener, S.; Zurbrügg, C.; Vinnerås, B. Effects of Feedstock on Larval Development and Process Efficiency in Waste Treatment with Black Soldier Fly (*Hermetia illucens*). *J. Clean. Prod.* **2019**, *208*, 211–219. [CrossRef]
19. Eggink, K.M.; Lund, I.; Pedersen, P.B.; Hansen, B.W.; Dalsgaard, J. Biowaste and By-Products as Rearing Substrates for Black Soldier Fly (*Hermetia illucens*) Larvae: Effects on Larval Body Composition and Performance. *PLoS ONE* **2022**, *17*, e0275213. [CrossRef] [PubMed]
20. Arena, R.; Manuguerra, S.; Curcuraci, E.; Cusimano, M.; Lo Monaco, D.; Di Bella, C.; Santulli, A.; Messina, C.M. Fisheries and Aquaculture By-Products Modulate Growth, Body Composition, and Omega-3 Polyunsaturated Fatty Acid Content in Black Soldier Fly (*Hermetia illucens*) Larvae. *Front. Anim. Sci.* **2023**, *4*, 1204767. [CrossRef]
21. Nyakeri, E.M.; Ogola, H.J.; Ayieko, M.A.; Amimo, F.A. An Open System for Farming Black Soldier Fly Larvae as a Source of Proteins for Smallscale Poultry and Fish Production. *J. Insects Food Feed.* **2017**, *3*, 51–56. [CrossRef]
22. Hosseindoust, A.; Ha, S.H.; Mun, J.Y.; Kim, J.S. Quality Characteristics of Black Soldier Flies Produced by Different Substrates. *Insects* **2023**, *14*, 500. [CrossRef] [PubMed]

23. Jucker, C.; Erba, D.; Leonardi, M.G.; Lupi, D.; Savoldelli, S. Assessment of Vegetable and Fruit Substrates as Potential Rearing Media for *Hermetia illucens* (Diptera: Stratiomyidae) Larvae. *Environ. Entomol.* **2017**, *46*, 1415–1423. [CrossRef] [PubMed]
24. Adebayo, H.A.; Kemabonta, K.A.; Ogbogu, S.S.; Elechi, M.C.; Obe, M.T. Comparative Assessment of Developmental Parameters, Proximate Analysis and Mineral Compositions of Black Soldier Fly (*Hermetia illucens*) Prepupae Reared on Organic Waste Substrates. *Int. J. Trop. Insect Sci.* **2021**, *41*, 1953–1959. [CrossRef]
25. Tschirner, M.; Simon, A. Influence of Different Growing Substrates and Processing on the Nutrient Composition of Black Soldier Fly Larvae Destined for Animal Feed. *J. Insects Food Feed.* **2015**, *1*, 249–259. [CrossRef]
26. Naser El Deen, S.; van Rozen, K.; Elissen, H.; van Wikselaar, P.; Fodor, I.; van der Weide, R.; Hoek-van den Hil, E.F.; Rezaei Far, A.; Veldkamp, T. Bioconversion of Different Waste Streams of Animal and Vegetal Origin and Manure by Black Soldier Fly Larvae *Hermetia illucens* L. (Diptera: Stratiomyidae). *Insects* **2023**, *14*, 204. [CrossRef] [PubMed]
27. Meneguz, M.; Schiavone, A.; Gai, F.; Dama, A.; Lussiana, C.; Renna, M.; Gasco, L. Effect of Rearing Substrate on Growth Performance, Waste Reduction Efficiency and Chemical Composition of Black Soldier Fly (*Hermetia illucens*) Larvae. *J. Sci. Food Agric.* **2018**, *98*, 5776–5784. [CrossRef]
28. Ebeneezar, S.; Linga Prabu, D.; Tejpal, C.S.; Jeena, N.S.; Summaya, R.; Chandrasekar, S.; Sayooj, P.; Vijayagopal, P. Nutritional Evaluation, Bioconversion Performance and Phylogenetic Assessment of Black Soldier Fly (*Hermetia illucens*, Linn. 1758) Larvae Valorized from Food Waste. *Environ. Technol. Innov.* **2021**, *23*, 101783. [CrossRef]
29. Fitriana, E.L.; Jayanegara, A.; Astuti, D.A.; Laconi, E.B. Growth Performance and Nutrient Composition of Black Soldier Fly Larvae Reared on Solid-State Fermentation Substrates with Various White Rot Fungi. *Biodiversitas* **2022**, *23*, 4894–4905. [CrossRef]
30. Lopes, I.G.; Lalander, C.; Vidotti, R.M.; Vinnerås, B. Using *Hermetia illucens* Larvae to Process Biowaste from Aquaculture Production. *J. Clean. Prod.* **2020**, *251*, 119753. [CrossRef]
31. Opoku, O.; Hamidu, J.A.; Adjei-Mensah, B.; Quaye, B.; Sasu, P.; Benante, V.; Amankrah, M.A.; Donkoh, A.; Atuahene, C.C. Alternative Organic Wastes as Substrates and Its Impact on Growth, Proximate Chemical and Mineral Composition of Black Soldier Fly (*Hermetia illucens*) Larvae for Subsistent Animal Feed. *J. Appl. Anim. Res.* **2023**, *51*, 564–572. [CrossRef]
32. Jing Lim, J.; Seng Liew, C.; Raksasat, R.; Merican, Z.M.A.; Kiatkittipong, K.; Alaaeldin Abdelfattah, E.; Mohamad, M.; Bashir, M.J.K.; Karabo Obed Ntwampe, S.; Lim, J.W. Cellulase Pretreated Palm Decanter Cake for Feeding of Black Soldier Fly Larvae in Triggering Bioaccumulation of Protein and Lipid into Biodiesel Productions. *Sustain. Energy Technol. Assess.* **2022**, *53*, 102485. [CrossRef]
33. Montevecchi, G.; Macavei, L.I.; Zanelli, E.; Benassi, G.; Pinotti, G.; D'Arco, S.; Buffagni, S.; Masino, F.; Maistrello, L.; Antonelli, A. Seasonal Variability of the HO.RE.CA. Food Leftovers Employed as a Feeding Substrate for Black Soldier Fly (*Hermetia illucens* L.) Larvae and Effects on the Rearing Performance. *Sustain. Chem. Pharm.* **2023**, *33*, 101061. [CrossRef]
34. Spranghers, T.; Ottoboni, M.; Klootwijk, C.; Ovyn, A.; Deboosere, S.; De Meulenaer, B.; Michiels, J.; Eeckhout, M.; De Clercq, P.; De Smet, S. Nutritional Composition of Black Soldier Fly (*Hermetia illucens*) Prepupae Reared on Different Organic Waste Substrates. *J. Sci. Food Agric.* **2017**, *97*, 2594–2600. [CrossRef] [PubMed]
35. Bava, L.; Jucker, C.; Gislon, G.; Lupi, D.; Savoldelli, S.; Zucali, M.; Colombini, S. Rearing of *Hermetia illucens* on Different Organic By-Products: Influence on Growth, Waste Reduction, and Environmental Impact. *Animals* **2019**, *9*, 289. [CrossRef]
36. Chia, S.Y.; Tanga, C.M.; Osuga, I.M.; Mohamed, S.A.; Khamis, F.M.; Salifu, D.; Sevgan, S.; Fiaboe, K.K.M.; Niassy, S.; van Loon, J.J.A.; et al. Effects of Waste Stream Combinations from Brewing Industry on Performance of Black Soldier Fly, *Hermetia illucens* (Diptera: Stratiomyidae). *PeerJ* **2018**, *2018*, e5885. [CrossRef] [PubMed]
37. Chia, S.Y.; Tanga, C.M.; Osuga, I.M.; Cheseto, X.; Ekesi, S.; Dicke, M.; van Loon, J.J.A. Nutritional Composition of Black Soldier Fly Larvae Feeding on Agro-Industrial by-Products. *Entomol. Exp. Appl.* **2020**, *168*, 472–481. [CrossRef]
38. Palma, L.; Fernández-Bayo, J.; Putri, F.; VanderGheynst, J.S. Almond By-Product Composition Impacts the Rearing of Black Soldier Fly Larvae and Quality of the Spent Substrate as a Soil Amendment. *J. Sci. Food Agric.* **2020**, *100*, 4618–4626. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.