SPATIAL-TEMPORAL DEEP LEARNING FOR INDUSTRIAL BIOCONVERSION: PREDICTING PROTEIN YIELD AND HARVEST SCHEDULING

Tom Avikasis Cohen and Anna Brook

Spectroscopy & Remote Sensing Laboratory, School of Environmental Sciences, University of Haifa, Abba Khoushy Ave 199, Haifa, 3498838, Israel

ABSTRACT

Industrial bioconversion using Black Soldier Fly Larvae faces spatial and micro-climatic variances across rearing trays, traditionally evaluated via destructive sampling. This study presents an end-to-end framework utilizing sequential Visible and Near-Infrared hyperspectral (350-1100 nm) and Long-Wave Infrared thermal imaging to non-invasively quantify protein depletion and larval biomass. By extracting baseline-independent protein proxies and segmenting tray matrices into localized 50×50 pixel patches, we map biological heterogeneity across temporal checkpoints. Bypassing rigid Markovian assumptions, a dual-headed CNN-LSTM pipeline processes these sequences. Model 1 classifies the optimal rearing stop-day, achieving 100% aggregated tray-level precision for standard schedules despite localized patch variances (65.7%-68.7%). Model 2 predicts final harvestable protein mass, yielding a minimal mean prediction error of 9.3g (3.7%) in stable cycles, while successfully flagging anomalous biological pathways. This framework provides a scalable, non-destructive roadmap for precision insect farming.

Index Terms **-** Hyperspectral imaging, thermal infrared, convolutional neural networks, long short-term memory, protein prediction.

1. INTRODUCTION

Growing demand for sustainable protein alternatives has driven rapid growth in insect-bioconversion circular economies. Black Soldier Fly Larvae (BSFL) efficiently upcycle organic waste into high-value biomass, yielding 40–50% crude protein by dry weight [1]. However, maintaining optimal feed-to-protein conversion efficiency at industrial scale remains highly complex.

Industrial monitoring currently lacks high-resolution, non-destructive tools. Larval feeding densities and localized metabolic hot-spots cause highly uneven nutrient depletion within a single tray, yet traditional computer-vision approaches average sensory features across the whole tray, obscuring these localized variations and edge-drying effects. Predictive modeling has also relied largely on static regression or restrictive linear structures, such as Hidden Markov Models, which fail to capture the complex, non-linear spatial-temporal dynamics of active larval beds.

This study introduces a multi-modal deep-learning workflow integrating VNIR hyperspectral and LWIR thermal imaging for direct spectral protein quantification. A dual-headed CNN-LSTM architecture maps continuous spatial patches through time to simultaneously classify optimal harvest timing and predict total harvestable protein mass, bypassing destructive sampling.

2. MULTIMODAL SPECTRAL AND THERMAL SENSING METHODOLOGY

2.1. Transformer-Based Dense Matching for Cross-Modal Image Registration

To achieve precise pixel-level co-registration between modalities, an ML-driven cross-modal alignment framework addresses the severe radiometric discrepancy between reflected VNIR radiance and emitted LWIR thermal radiation. The active larval substrate's dynamic, non-rigid topography lacks sharp geometric landmarks, rendering classical keypoint descriptors intractable. To reconcile these domain differences, the framework first performs generative domain translation via a conditional GAN (cGAN) based on Pix2Pix [2]: a multi-band pseudo-RGB synthesis from the VNIR hyperspectral cube serves as the source domain, mapped to a target synthetic pseudo-thermal image matching the true LWIR frame's structural layout. The generator uses a U-Net backbone with skip connections, minimizing a compound adversarial and L1 pixel-distance loss to preserve structural boundaries.

With both datasets rendered in the thermal domain texture space, a detector-free dense matcher based on the Local Feature Transformer (LoFTR) framework [3] is applied. Rather than sparse correspondences from localized image gradients, LoFTR uses self-attention and cross-attention layers in a coarse-to-fine loop to build a dense coordinate match field between the synthetic and true LWIR images, ensuring accurate mapping across smooth, low-texture zones dominated by larval metabolic hotspots. The resulting displacement field is used to calculate a non-rigid geometric transformation via a Thin-Plate Spline (TPS) deformation model, which eliminates localized parallax errors inherent in

standard global homography matrices and guarantees sub-pixel co-registration across the tray matrix.

2.2. Thermal-Gravimetric Calibration and Spatial Moisture Mapping

High-resolution LWIR imaging [4] mapped surface temperature distributions, capturing substrate evaporative cooling rates and localized metabolic heat fluxes from active larval clusters. To replace destructive gravimetric moisture assessments with continuous, non-invasive monitoring, a physically grounded thermal-gravimetric calibration framework was developed based on surface energy balance dynamics. Calibration datasets spanning a 0%–62% operational window were established under fixed sensor geometries, subjecting a 1600 g dry homogeneous baseline diet to an incremental wetting protocol of discrete 480 ml water steps. True gravimetric moisture was computed via $M(\%)=[(W_{wet}-W_{dry})/W_{wet}]x100$, providing a robust structural and thermal baseline to contextualize the hyperspectral protein signals.

Raw thermal data underwent emissivity and ambient temperature corrections prior to global normalization. To decouple moisture-induced evaporative cooling from the escalating biological heat fluxes of maturing larvae, a dynamic, stage-dependent intensity normalization was implemented based on the Day of Larval development (DOL). Within the co-registered region of interest, intra-tray spatial variations were segmented using a mean-based adaptive threshold, $mask_T(x,y)=1$ if $T(x,y)<\mu_T$ else 0, where μ_T is the mean thermal intensity, isolating high-moisture cooling regions from localized drying zones, while a complementary inverse mask extracted high-temperature metabolic hotspots. This defined strict operational constraints: below ~56% moisture the substrate kept a granular topography optimal for stable thermal contrast and noise-free VNIR collection, whereas approaching the 62% saturation boundary triggered a semi-liquid transition with severe VNIR water-absorption noise and high thermal inertia, dictating the system's functional deployment window.

2.3. Spectral-Based Protein Quantification (VNIR)

Hyperspectral measurements are captured across the 350–1100 nm Visible and Near-Infrared (VNIR) spectrum [5]. This specific region is highly sensitive to the overtone and combination bands of fundamental molecular vibrations, specifically the N-H, C-H, and O-H bonds that characterize crude protein and amino acid structures.

To isolate protein signatures from the highly variable wet-substrate background, raw digital numbers are first converted to relative reflectance using a calibrated Spectralon white reference. Advanced spectral feature engineering is then applied: continuum removal (CR), first-order derivative analysis (1stDA), and spectral area under polygon (SAUP). CR normalizes the reflectance spectra to a common baseline, isolating absorption troughs associated with protein concentration; 1stDA removes baseline drift from varying lighting and physical scattering across the uneven tray topography. SAUP metrics serve as a direct quantitative proxy for both residual dietary protein in the substrate and crude protein accumulated within the larval bodies.

To validate larval protein (940-960 nm) accumulation without interference from the underlying background diet (710-830 nm), the spectral data were filtered using the complementary inverse thermal mask, which isolates high-temperature metabolic hotspots and segments active larval biomass from the surrounding substrate. Within these thermally isolated zones, protein quantification targeted specific VNIR absorption features at the 940 nm, 950 nm, and 960 nm bandwidths correlating with the molecular overtone vibrations of crude protein.

To validate this spectral proxy, a controlled laboratory calibration used standardized, gravimetrically prepared mixtures of extracted BSFL protein and starch graduating from 0% to 100% protein concentration, with the integrated spectral absorption area within these bandwidths calculated for each mixture. Results showed a highly robust correlation between true protein concentration and the spectral absorption metric, scaling from a baseline absorption area of ~0.57 in pure starch (0% protein) to a peak of 31.85 in 100% pure larval protein. Intersecting the dynamic, DOL-adjusted metabolic thermal masks with this calibrated 940–960 nm absorption model lets the system decouple true larval protein signatures from the variable, moisture-dependent substrate, establishing an accurate, non-invasive baseline for absolute protein yield prediction.

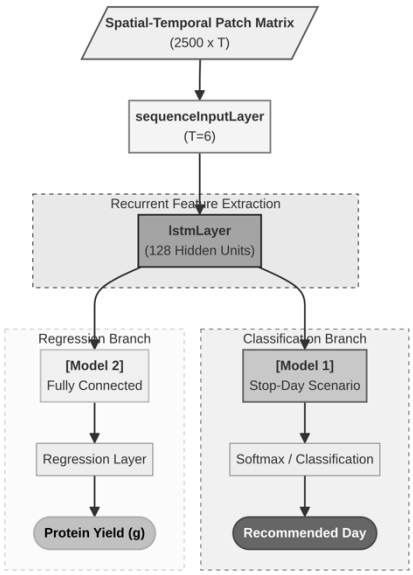
2.4. CNN-LSTM model architectures

The architecture is an end-to-end recurrent framework capable of modeling spatial-temporal dependencies from sequence inputs. The underlying assumption is that, over time, the larvae consume the diet, leading to a decrease in the concentration of available dietary protein in the tray, while the protein content of the larvae increases.

Model objectives: To identify an optimal rearing termination point based on the spatial distribution patterns of available dietary protein within the tray. To predict the total protein yield (in grams) harvested from each tray.

To account for spatial variability, analysis is performed at the pixel level and by aggregating pixels into square blocks, further grouped into three tray regions: corners, edges, and center.

Each 50 × 50 patch is flattened into a 2,500-dimensional continuous signal vector evaluated across a sequence of up to 6 distinct developmental check-points.



Given that dietary protein depletion exhibits complex, non-linear trends, the recurrent architecture of the LSTM [6] is uniquely suited for this task due to its internal memory cell structure, which successfully retains temporal patterns and identifies continuous variations across successive rearing days.

The prepared spatial-temporal matrices are routed through two network configurations handling discrete operational scheduling and continuous yield estimation, respectively. Both architectures share a recurrent backbone: a sequenceInputLayer(6) ingesting the temporal sequence, followed by an lstmLayer(128) learning the localized space-time dynamics.

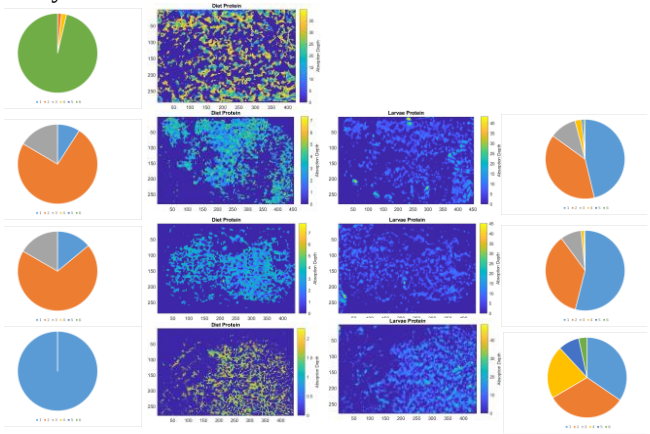
Model 1 frames harvest readiness as a temporal classification problem: the 2500 × 6 temporal matrix passes through the sequence input and 128-unit LSTM layers to capture non-linear trends over time, and a final Softmax layer assigns patch sequences to one of three targeted industrial scenarios:

Scenario 1: Recommended harvest at the end of Day 3.
Scenario 2: Recommended harvest at the end of Day 4.
Scenario 3: Recommended harvest continuing into Day 5.

Model 2 uses a structurally parallel front-end but replaces the classification head with a regression bottleneck: after the 128-unit LSTM feature-learning phase, fully connected layers contract the spatial dimensions to a continuous regression output estimating final harvestable mass, defined via validated baseline measurements: Larval Protein Mass (g) is equal to Total Biomass * Dry Matter * Crude Protein * 4.43.

3. RESULTS

To translate continuous multi-modal spectral measurements into operationally actionable metrics for downstream predictive modeling, calculated protein content was discretized into six semi-quantitative classes, governed by fixed mathematical thresholds on the integrated area under the peak that simultaneously track dietary protein depletion (710-830 nm) and larval protein accumulation (940-960 nm). Beyond serving as stabilized, noise-tolerant labels for the spatial-temporal classification pipeline, this six-class categorization illustrates the dynamic spatial-temporal interactions (Fig.2) between substrate protein depletion and larval protein accumulation across the bioconversion lifecycle.



3.1. Rearing Stop-Day Classification (Model 1)

The classification module of the dual LSTM architecture, tasked with identifying the optimal conclusion of the bioconversion cycle, achieved an overall patch-level prediction accuracy of 67.97%. The model demonstrated a highly robust capability to correctly identify the standard optimal endpoints (Day 3 and Day 4, representing Scenarios 1 and 2, respectively). Discrepancies in the patch-level classification primarily occurred during extended rearing cycles; in instances where the actual rearing concluded on Day 5, the model frequently recommended an earlier Day 4 harvest. This conservative bias was largely attributed to a structural data imbalance during the experimental phase, where Day 5 biological continuations were significantly underrepresented compared to the standard Day 3 and Day 4 endpoints.

The confusion matrix reveals high robustness in identifying standard production endpoints, but severe bias when handling extended cycles:

1. True Class 1 (Day 3 Harvest): The model demonstrated excellent sensitivity, correctly classifying 474 patches as Day 3. Only a negligible

fraction (6 patches) were misclassified as Day 4, with zero patches misclassified as Day 5.

2. True Class 2 (Day 4 Harvest): The model accurately recognized 615 patches as Day 4. Discrepancies were minimal, with 21 patches prematurely predicted as Day 3 and 4 patches delayed to Day 5.
3. True Class 3 (Day 5 Harvest): The primary source of the model's overall error originated here. Out of 600 actual Day 5 patches, only 80 were correctly classified. The model overwhelmingly misclassified 500 of these patches as Day 4, alongside 20 patches as Day 3.

This exact error distribution pattern remained consistent when breaking down the predictions by spatial regions (tray edges, corners, and centers), highlighting a systematic classification behavior rather than a spatial artifact.

When the individual patch classifications were aggregated into a holistic majority vote to make a single operational decision per rearing tray, the polarity of the model's performance became even more absolute:

1. Day 3 and Day 4: The aggregated matrix showed 100% accuracy for standard operations. All 13 physical Day 3 trays were correctly predicted as Day 3, and all 16 physical Day 4 trays were strictly predicted as Day 4.
2. Day 5: The model completely failed to identify extended rearing cycles at the tray level. All 14 actual Day 5 trays were incorrectly recommended for an earlier Day 4 harvest.

As established in the experimental analysis, this heavy bias toward Day 4 in the confusion matrix is not a fundamental algorithmic failure, but rather a direct reflection of structural data imbalance during the model's training phase, where extended Day 5 biological cycles were significantly underrepresented.

3.2. Spatial-Temporal Protein Yield Prediction (Model 2)

The evaluation of the regression module explicitly highlighted a critical interaction between temporal harvest scheduling and mass prediction accuracy. The absolute error in the predicted harvestable larval protein was directly coupled to the agreement between the model's recommended stop-day and the actual physical execution of the harvest.

3.2.1. High-Synchrony Cycles

Consistent with earlier models that rely on spatial patterns, the system's accuracy heavily depends on the physical location within the tray. Center Accuracy: The model is most accurate when predicting based on the center of the tray, producing the smallest errors across all scenarios. Boundary Inaccuracies: The error increases towards the boundaries; the edges show higher errors than the center, and the corners exhibit the most significant, highest prediction errors.

The total duration of the fattening period drastically affects the model's capability to predict the final protein yield. Standard Cycles (Days 3 & 4): When the rearing concludes in Scenarios 1 or 2 (harvesting on Day 3 or Day 4, respectively), the protein quantity prediction error is relatively small and manageable. Extended Cycles (Day 5): When the process extends to Scenario 3 (harvesting on Day 5), the error in the predicted protein quantity grows significantly higher compared to the earlier harvest days.

The accompanying chart MAE by region and feeding duration explicitly quantifies these trends: Day 3 Harvest: Center error is ~13.29 g, Edge is ~21.44 g, and Corner is ~67.97 g. Day 4 Harvest: Center error is ~47.53 g, Edge is ~59.06 g, and Corner is ~84.47 g. Day 5 Harvest: The errors spike considerably, with the Center at ~102.40 g, Edge at ~104.96 g, and Corner reaching the maximum at ~144.77 g.

When the model's predicted optimal harvest timeline aligned perfectly with the actual operational rearing schedule, the spatial-temporal regression framework demonstrated exceptional precision. In stable Day 4 developmental cycles, where both the actual and predicted stop days were synchronously identified as Day 4, the model tracked the true final protein yields with remarkable accuracy. Evaluating consecutive validation trays yielded actual physical extractions of 244 g, 252 g, 256 g, and 248 g, which were precisely predicted by the model as 255 g, 267 g, 261 g, and 252 g, respectively. Across these synchronized cycles, the framework achieved a minimal mean absolute error of just 9.3 grams per tray, corresponding to a highly reliable 3.7% relative error rate.

3.2.2. Asynchronous Cycles

Conversely, the prediction error magnified significantly when there was a temporal disagreement between the system's recommended optimal endpoint and the actual physical harvest execution. In sub-optimal batches where trays were prematurely harvested on Day 3, the model's spatial-temporal analysis actively recognized that the dietary substrate was not fully depleted. Consequently, the model correctly recommended extending the fattening period to Day 4 to maximize yield potential. Because the physical harvest was aborted early on Day 3, the model's Day 4 projected yield severely mismatched the immature physical extraction. Actual stunted yields of 187 g, 192 g, 175 g, and 223 g were mathematically projected forward by the model to mature at 225 g, 232 g, 221 g, and 265 g, respectively. This temporal misalignment generated a substantially elevated mean prediction error of 41.2 grams per tray, a 16.3% deviation.

Ultimately, these results confirm that the system acts not only as a high-precision mass estimator during standard operations (3.7% error), but also as an active biological

anomaly detector, where a sudden spike in yield discrepancy mathematically flags that the rearing protocol is fundamentally out of sync with the optimal substrate depletion timeline.

4. DISCUSSION

The proposed dual-modal imaging and deep learning framework demonstrates significant potential for advancing precision insect agriculture by addressing the core limitations of conventional monitoring. By successfully replacing destructive, localized gravimetric sampling with continuous, non-invasive hyperspectral and thermal imaging, the system provides unprecedented spatial- BSFL bioconversion process. The integration of dynamic, stage-dependent thermal masking alongside localized 940-960 nm near-infrared absorption calculations effectively isolated larval biomass from the heterogeneous substrate. This decoupling overcame a major bottleneck in optical monitoring of active, moisture-variable biological systems.

A fundamental finding is the interplay between harvest-timing synchrony and mass-prediction accuracy: when rearing cycles align with the model's optimal timeline (Days 3 and 4), the regression module is a highly reliable mass estimator (3.7% relative error), whereas the magnified errors under asynchronous cycles reveal the framework's secondary, critical utility as an active biological anomaly detector – a spike in prediction error flags that the rearing protocol or substrate depletion rate has deviated from the optimal bioconversion trajectory.

Despite these successes, limitations remain regarding the classification module's capability to generalize across extended rearing cycles. The algorithmic bias toward recommending a Day 4 harvest for actual Day 5 biological profiles underscores a vulnerability to structural data imbalances inherent in the training phase, where extended biological cycles were underrepresented. Future work must address this limitation by expanding the longitudinal dataset and employing targeted data augmentation techniques, such as generative adversarial synthetic oversampling of extended rearing scenarios, to ensure robust temporal classification across the complete spectrum of developmental timelines.

5. CONCLUSION

This study presents a comprehensive, end-to-end framework for the non-destructive monitoring of BSFL bioconversion, using VNIR hyperspectral data and LWIR thermal imaging. By accurately co-registering cross-modal data and translating localized substrate variations into a spatial-temporal sequence, the deployed dual CNN-LSTM architecture successfully models the complex, non-linear dynamics of dietary protein depletion and larval biomass accumulation. The methodology precisely dictates operational constraints, such as the 62% moisture saturation limit for viable hyperspectral acquisition, while utilizing dynamic thermal masking to extract highly accurate protein signatures. Ultimately, the system accurately predicts both the optimal production stop-day and the final harvestable protein mass, bridging the gap between raw optical sensing and actionable industrial forecasting. This predictive framework transitions insect rearing from reactive, manual estimations to proactive algorithmic management, establishing a highly scalable, automated roadmap for precision bioconversion facilities.

ACKNOWLEDGEMENT

This work has been conducted as part of the Israeli BSF (black soldier fly) consortium and supported by the Israel Innovation Authority (grant # 79630). The authors express their deepest gratitude to the Dr Moshe Elbaz and team members from BINGE ORGANICS LTD for data curation and validation, and to the project referents for their invaluable guidance, professional support, and dedication throughout the execution of this research.

REFERENCES

[1] Makkar, H.P., Tran, G., Heuzé, V. and Ankers, P., 2014. State-of-the-art on use of insects as animal feed. *Animal feed science and technology*, *197*, pp.1-33.

[2] Isola, P., Zhu, J.Y., Zhou, T. and Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).

[3] Sun, J., Shen, Z., Wang, Y., Bao, H. and Zhou, X., 2021. LoFTR: Detector-free local feature matching with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8922-8931).

[4] Jones, H.G., 2004. Application of thermal imaging and infrared sensing in plant physiology and ecophysiology. In *Advances in botanical research* (Vol. 41, pp. 107-163). Academic Press.

[5] ElMasry, G., Sun, D.W. and Allen, P., 2012. Near-infrared hyperspectral imaging for predicting colour, pH and tenderness of fresh beef. *Journal of Food Engineering*, *110*(1), pp.127-140.

[6] Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. *Neural computation*, *9*(8), pp.1735-1780.