# A Platform for AI-Driven Egg Freshness Assessment from Near Infrared Spectral Data

**Pongtachchai Panachaiboonpipop[1], Manoosak Wongphatcharachai[2], Nutthawan Nonthabenjawan[2], and Kanet Wongravee[3,*]**

*[1]Technology Management and Innopreneurship, Inter-Department of Technopreneurship and Innovation Management, College of Interdisciplinary and Integrative Studies, Chulalongkorn University, Bangkok, 10330, Thailand*
*[2]Biotechnology Research and Innovation Center, CP Innovation Center of Excellence, Charoen Phokphand Group, Thailand*
*[3]Department of Chemistry, Faculty of Science, Chulalongkorn University, Bangkok, 10330, Thailand*
*[*]Corresponding Author E-mail: kanet.w@chula.ac.th, kanet.wongravee@gmail.com*

## ABSTRACT

Egg freshness is critical in industrial processing because variations in freshness can affect processing conditions and product quality, such as in boiled or soft-boiled eggs. Inconsistent freshness within a batch may cause processing defects, leading to product loss, food waste, and increased production costs. Freshness is conventionally quantified by the Haugh unit (HU); however, HU measurement is destructive, time-consuming and unsuitable for high-volume screening. Near infrared (NIR) spectroscopy offers a rapid, non-destructive alternative, but building a reliable calibration requires many preprocessing and model choices that are usually made by hand, applied inconsistently, and difficult to reproduce. This work presents a platform for AI-driven egg freshness assessment that standardizes and automates the complete NIR workflow (spectral ingestion, preprocessing, feature selection and AI model training) and applies it to two complementary tasks: a quantitative regression to predict HU, and a qualitative classification to classify standard grade (grade AA : $HU \geq 70$). We illustrate the platform on 642 hen eggs stored at 25 °C, whose reflectance NIR spectra (1,000-2,500 nm) were acquired non-destructively over 0-10 days of storage. With no manual tuning, the platform automatically found the best preprocessing-wavelength-model combination for each task and showed it clearly outperformed the weaker combinations: the best regression; regularized linear model on the combined top-and-bottom spectra reached an RMSE of 7.0 HU, against up to about 8.7 HU for poorer combinations, and the best classifier; XGBoost about 82% accuracy, against about 71% for the weakest. The egg sample thus serves as a worked example showing that the platform can carry out a complete, quantitative and qualitative, egg-freshness assessment automatically, consistently and reproducibly.

**Keywords:** Egg freshness/ Near infrared spectroscopy/ Artificial intelligence

## 1. INTRODUCTION

Rapid, non-destructive evaluation of egg freshness is increasingly important for both the egg industry and food security. Freshness matters in industrial processing, where it influences the processing conditions required and the quality of products such as boiled and soft-boiled eggs. When freshness varies within a batch, a single set of processing parameters cannot suit every egg, and the resulting defects translate into product loss, food waste and higher production costs. Replacing destructive testing with non-destructive measurement would allow many more eggs to be assessed before processing and so limit these losses on food production. Eggs are among the most widely consumed sources of high-quality animal protein and freshness is one of the most important indicators of their internal quality. Freshness is conventionally quantified by the Haugh unit (HU), defined by Haugh in 1937, which relates the height of the thick albumen to the weight of the egg [1]. Immediately after oviposition, however, the internal quality of the egg begins to decline: water and carbon dioxide are lost through the pores of the shell, the albumen pH rises, the thick albumen liquefies, the air cell enlarges and the yolk index falls [2], so that HU decreases with storage time.

HU determination requires the egg to be broken so that the albumen height can be measured; it is therefore destructive, labor-intensive and applicable only to a small sample drawn from a larger batch. These limitations have motivated the development of non-destructive techniques for estimating freshness on intact eggs.

Spectroscopy in the near infrared region has emerged as a leading non-destructive approach, because NIR light probes the overtone and combination bands of OH, NH and CH bonds and is therefore sensitive to the water and protein changes that accompany egg aging. NIR spectroscopy has been used specifically to predict the Haugh unit (HU), the standard indicator of egg freshness that is otherwise obtained only by breaking the egg to measure albumen height. Comparing acquisition modes, transmission-mode NIR spectra have outperformed reflection for HU prediction because transmitted light probes the internal albumen directly rather than the shell surface, yielding an RMSEP of 4.25 for HU via Si-PLS regression [3]. This capability has been miniaturized for field and online use, with a portable 900-1,700 nm NIR spectrometer combined with support-vector regression predicting HU with a relative error of 7.32% and an RPD of 2.56, with the equatorial mid-shell region identified as the most informative measurement site [4]. Wavelength-selection methods have further refined HU prediction from short-wave NIR spectra: an interval-PLS model combined with Lasso regularization narrowed 26 candidate spectral intervals down to the 10 most informative bands, improving HU prediction accuracy by 7.6% over a full-spectrum model and achieving an RMSEP of 5.161 [5].

Despite this progress, most reported models are developed with a combination of preprocessing and a single algorithm, and the preprocessing steps are frequently chosen without a systematic rationale, even though many preprocessing techniques are available and the choice strongly affects the resulting model. As a result, published pipelines are difficult to reproduce, to compare fairly, or to reuse on new data by non-specialists, a barrier that is particularly acute for deep-learning models, which are sensitive to preprocessing and require a standardized data-handling workflow. There remains a need for an integrated tool that couples systematic preprocessing and AI models. This paper presents such an AI-driven egg freshness assessment platform from NIR spectral data. The platform standardizes spectral ingestion, preprocessing, feature selection and AI model training within a single modular pipeline and evaluates classical and deep-learning models for both quantitative HU regression and qualitative catalog classification. We describe its architecture, demonstrate it on non-destructively acquired NIR spectra of hen eggs, and show that the framework is dataset-agnostic and readily extensible to other NIR food-quality applications.

## 2. METHODOLOGY

### *2.1 Demonstration Dataset: Egg Samples and Storage*

The platform was demonstrated on a non-destructive egg-freshness task. A total of 642 fresh hen eggs were obtained from farms (Biotechnology Research and Innovation Center, CP Innovation Center of Excellence) on the day of lay and stored at 25 °C. The internal structure of a hen egg and the two positions used for spectral measurement are shown in Figure 1: a shell enclosing an air cell at the bottom (blunt end), the thin and thick albumen, and the centrally suspended yolk held in place by the chalazae. To span a range of freshness, separate eggs were sampled at intervals over a storage period of 0-10 days and the Haugh unit of each egg was measured at its sampling time, giving 642 eggs that ranged from fresh (High HU) to aged (low HU), with a mean HU of 70.1 (standard deviation 11.2; range 35.6-93.0); the distribution of HU across the sample.

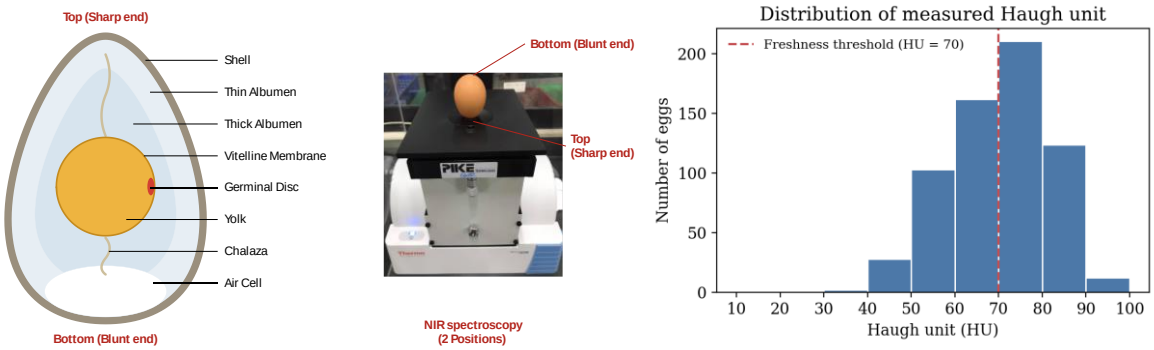


**Figure 1.** (Left) Internal structure of a hen egg (cross-section) detected positions using NIR spectrometer to obtain reflectance NIR spectra. (Right) Distribution of the measured Haugh unit across all 642 eggs (mean 70.1, range 35.6-93.0)

Reflectance NIR spectra (Thermo Scientific Nicolet iS5N) were acquired over the 1,000-2,500 nm range. For each egg, spectra were recorded three times at the top (sharp end) and three times at the bottom (blunt end), as shown in Figure 1. The three repeats at each position were averaged to give one mean spectrum per position. Four measurement configurations were then compared: the top spectrum alone (T), the bottom spectrum alone (B), the point-by-point average of the two (TB-avg), and the two spectra concatenated into a single feature vector (TB-concat).

### *2.2 Reference Freshness Measurement*

After spectral acquisition, each egg was broken and its weight and thick-albumen height were measured with an egg multitester, which computes the Haugh unit automatically. As the egg ages the thick albumen flattens and the albumen height decreases, lowering the HU (Figure 2). The Haugh unit is defined as

$$HU = 100 \times \log_{10}(h - 1.7\, w^{0.37} + 7.6) \qquad (1)$$

Where $h$ is the thick-albumen height (mm) and $w$ is the egg weight (g) [1]. The measured $HU$ served as the single reference (target) variable for all models.

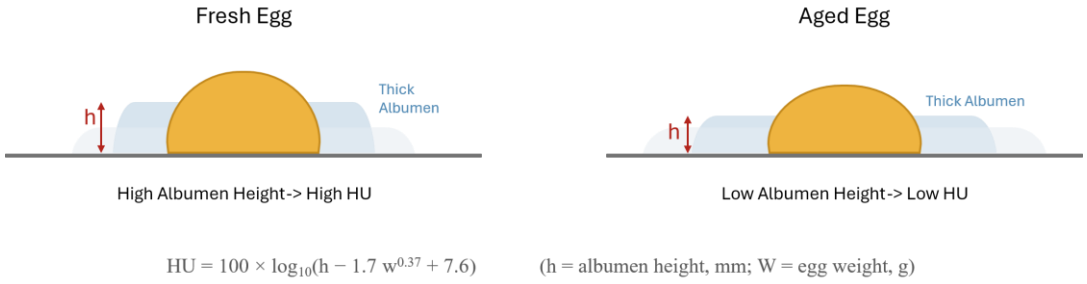


**Figure 2.** The Haugh unit measures the height of the thick albumen of a broken-out egg relative to its weight. Fresh eggs have a tall, firm albumen and a high HU; as the egg ages the albumen flattens and the HU falls

### *2.3 Platform Architecture*

The platform is organized as five interoperable modules (Figure 3): (i) a data-ingestion module that reads spectra and reference HU values from common tabular and instrument formats; (ii) a preprocessing module; (iii) a feature selection module; (iv) a model-training module; and (v) a report and visualization module that reports the predicted HU and the associated performance metrics. The modules communicate through a common data representation, so that the same pipeline can be applied without modification to spectra from different instruments and, indeed, to other food commodities.

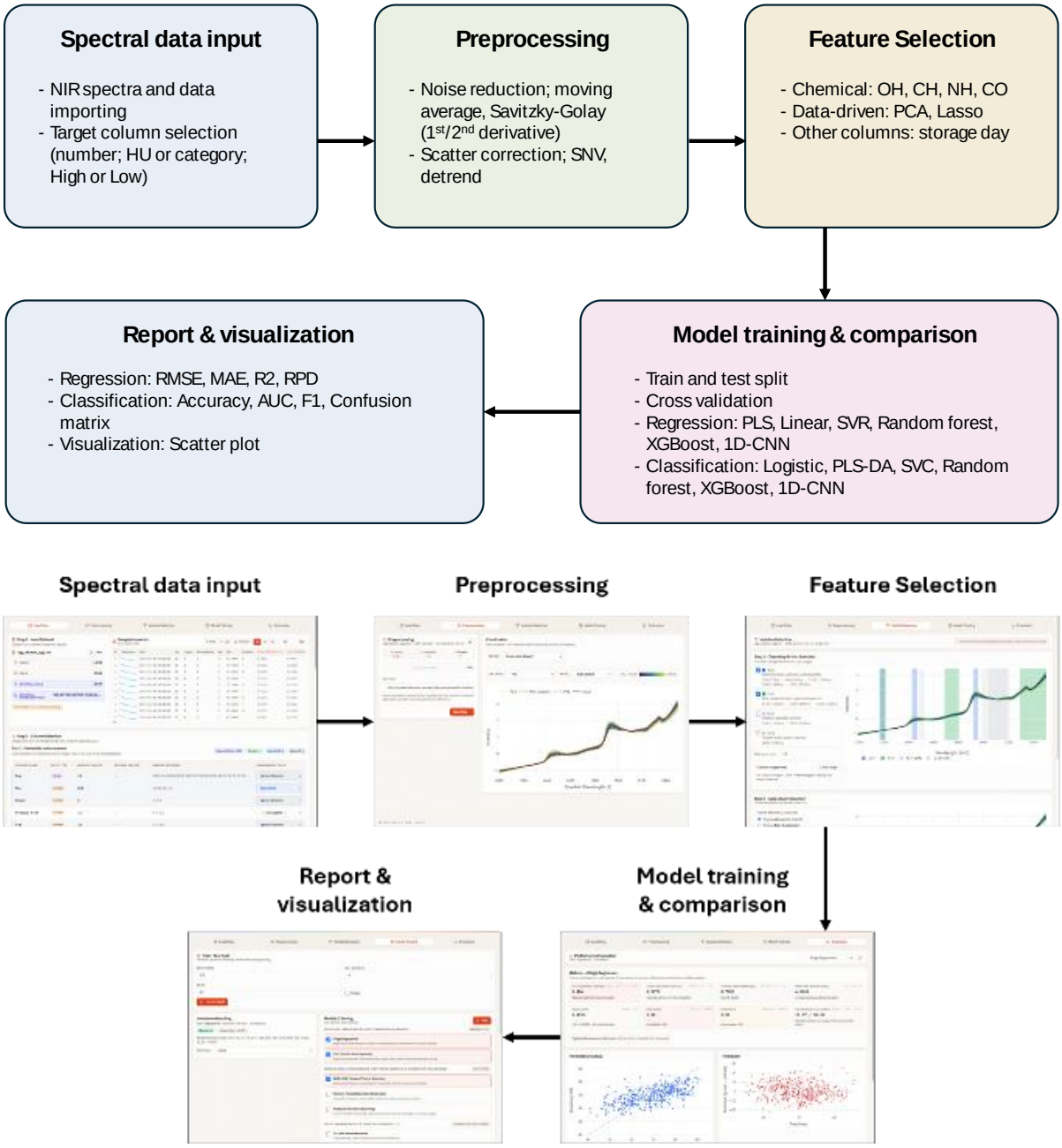


**Figure 3.** (Top) Architecture of the platform for AI-driven egg freshness assessment from NIR spectral data and (Bottom) A sample screen from the platform

### *2.4 Spectral Preprocessing*

The preprocessing engine is organized into two sub-modules. The first, noise reduction, smooths each spectrum with a moving-average filter or with Savitzky-Golay smoothing, optionally taking its first or second derivative to suppress baseline and slope [7]. The second, scatter correction, removes additive and multiplicative scatter effects by standard normal variate (SNV) or detrending [8]. These transformations may be applied singly or in combination, and the combination giving the best cross-validated performance is selected automatically, following recommended practice for systematic preprocessing evaluation [6].

### *2.5 Feature Selection and Modeling*

The platform provides three feature-selection strategies: (i) a chemistry-guided option that keeps the wavelength regions of the O-H, C-H, N-H overtone and combination bands, which carry the water- and protein-related information most relevant to freshness; (ii) a data-driven option based on principal component analysis (PCA), Lasso regression and a univariate linear filter; or (iii) the full spectrum. All models used the NIR spectral features only; no non-spectral inputs, such as storage time, were included. For the quantitative task, regression models were compared: partial least-squares (PLS), ridge and ElasticNet as regularized linear baselines, support-vector regression (SVR) with a radial-basis-function kernel, random forest, XGBoost and a one-dimensional convolutional neural network (1D-CNN). For the qualitative task the corresponding classifiers were used: logistic regression, PLS-DA, support-vector classification, random forest, XGBoost and a 1D-CNN. The 1D-CNN was trained on the full preprocessed spectrum without feature selection. Each of the four measurement configurations (T, B, TB-avg, TB-concat) was run through the same search: the platform evaluated every preprocessing × feature-selection × model combination, tuning hyperparameters by 5-fold cross-validation and selecting the best combination automatically. The data were split 70:30 into training and test sets, stratified by storage day for regression and by class for classification. Because all random operations use a fixed seed, re-running the pipeline reproduces identical metrics.

### *2.6 Model Evaluation*

Two tasks were evaluated. For the quantitative regression of HU, performance on the independent test set was assessed by the root-mean-square error (RMSE), together with the coefficient of determination ($R^2$), and the residual predictive deviation (RPD). For the qualitative classification, each egg was labeled High Freshness or Low Freshness using the acceptance threshold applied by egg-processing plants (High Freshness if HU $\geq$ 70); with this cut-off the sample contained 347 High Freshness and 295 Low Freshness eggs. Classification performance was reported as accuracy, AUC, and F1-score.

## 3. RESULTS AND DISCUSSION

### *3.1 Preprocessing and Spectral Characteristics*

The raw NIR reflectance spectra (Figure 5a) are shown without any noise or scatter correction, in which a pronounced baseline shift between spectra is apparent. Applying a Savitzky-Golay first derivative followed by standard normal variate (SNV) correction removed the multiplicative scatter and baseline offset and sharpened the overlapping absorption features (Figure 5b), which improved the stability of all subsequent models, consistent with previous reports that Savitzky-Golay and SNV are among the most effective preprocessing choices for NIR calibration [7]. Accordingly, the automatically selected pipelines mostly combined a Savitzky-Golay derivative with SNV or detrending and used either the chemistry-guided O-H, C-H, N-H, C-O bands or a data-driven method for wavelength selection.

After preprocessing, four regions in which the spectra separate most clearly become apparent: 1,410-1,520, 1,910-1,970, 2,140-2,200 and 2,320-2,365 nm (Figure 5b). The 1,410-1,520 and 2,140-2,200 nm regions correspond to N-H overtone and combination bands associated with protein, the 1,910-1,970 nm region to the O-H combination band of water, and the 2,320-2,365 nm region to C-O vibrations associated with the CaCO3 of the eggshell. As the egg ages the albumen loses water through the shell and its ovalbumin slowly converts to s-ovalbumin, so the intensities in these regions correlate with freshness. Because the bands overlap and sit on a sloping, scatter-dominated baseline, the freshness information is spread across many wavelengths rather than confined to a single peak.

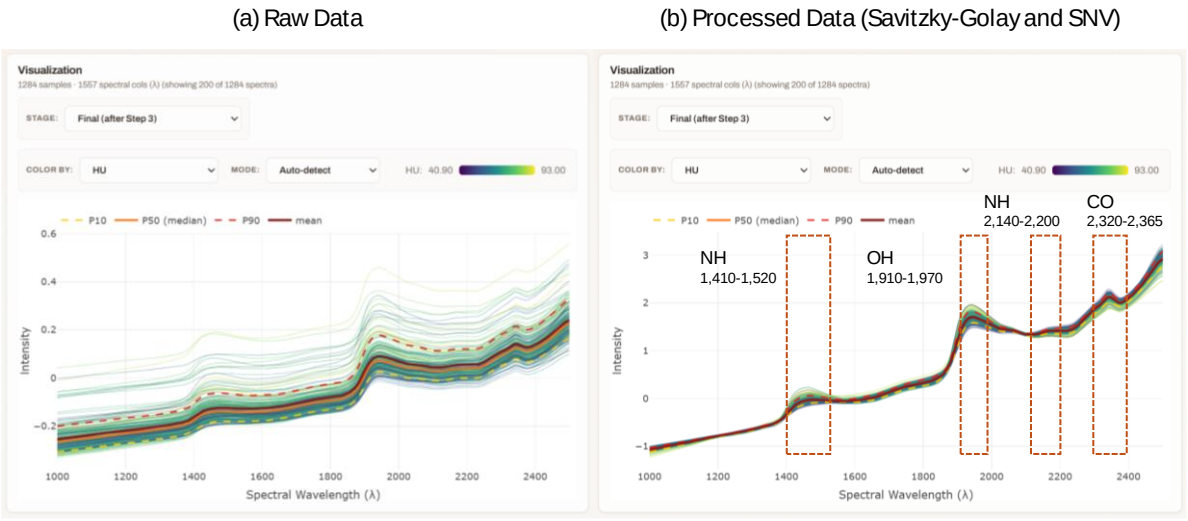


**Figure 5.** (a) Raw NIR reflectance spectra of hen eggs and (b) the same spectra after Savitzky-Golay and SNV preprocessing

### *3.2 Feature Selection and Model Comparison*

For each algorithm the platform kept the preprocessing and feature-selection combination with the lowest cross-validated error; Tables 1 and 2 list these settings and the resulting test performance. Wavelength selection mattered more than the exact preprocessing: a small set of principal components or the chemistry-guided O-H, C-H, N-H, C-O bands were selected for most models, confirming that the freshness signal is spread across many wavelengths rather than sitting in a few. Combining the two measurement ends helped both tasks: the concatenated top+bottom spectrum gave the best regression and the averaged top+bottom spectrum the best classification.

The best regression model was a regularized linear model (ridge regression) on the concatenated top+bottom spectrum with RMSE = 7.0 HU ($R^2$ = 0.56, RPD = 1.5); the other linear models were within about 0.6 HU, the 1D-CNN gave mixed results, and the tree-based models (random forest and XGBoost) were consistently worse. The best classifier was XGBoost on the averaged top+bottom spectrum, at 82.4% accuracy (AUC = 0.87, F1 = 0.84; confusion matrix in Figure 6b). In both tasks the best conventional model outperformed the deep-learning model, as expected when more than 1,500 highly correlated wavelengths are modeled from only 642 eggs. The value of the demonstration is not the winning model itself but that the platform found and reported it automatically, so the comparison can be reproduced without manual tuning.

**Table 1.** Quantitative HU regression: for every measurement configuration (T, B, TB-avg, TB-concat) and every algorithm, the automatically selected best preprocessing and wavelength selection and the resulting test performance (held-out 30% of eggs, n = 193 per configuration)

| Configuration | Model | Preprocessing | Wavelengths | RMSE | $R^2$ | RPD |
|---|---|---|---|---|---|---|
| Top (T) | PLS | SG 1st-deriv + SNV | Chemical bonds | 7.81 | 0.45 | 1.35 |
| Top (T) | Ridge | MA(3) + SNV | Chemical bonds | 7.62 | 0.48 | 1.39 |
| Top (T) | ElasticNet | SG 1st-deriv + SNV+detrend | Chemical bonds | 7.74 | 0.46 | 1.36 |
| Top (T) | SVR | SG 1st-deriv + SNV | PCA | 8.08 | 0.41 | 1.31 |
| Top (T) | Random forest | SG 1st-deriv + SNV | full spectrum | 8.64 | 0.33 | 1.22 |
| Top (T) | XGBoost | SG 1st-deriv + SNV | full spectrum | 8.59 | 0.33 | 1.23 |

**Table 1.** Quantitative HU regression: for every measurement configuration (T, B, TB-avg, TB-concat) and every algorithm, the automatically selected best preprocessing and wavelength selection and the resulting test performance (held-out 30% of eggs, n = 193 per configuration) (cont)

| Configuration | Model | Preprocessing | Wavelengths | RMSE | $R^2$ | RPD |
|---|---|---|---|---|---|---|
| Top (T) | 1D-CNN | SG 2nd-deriv + SNV+detrend | full spectrum | 7.57 | 0.48 | 1.39 |
| Bottom (B) | PLS | SG 1st-deriv + SNV+detrend | Chemical bonds | 7.61 | 0.48 | 1.39 |
| Bottom (B) | Ridge | SNV+detrend | Chemical bonds | 7.34 | 0.51 | 1.44 |
| Bottom (B) | ElasticNet | SG 1st-deriv + SNV+detrend | full spectrum | 7.52 | 0.49 | 1.40 |
| Bottom (B) | SVR | SG + SNV+detrend | PCA | 7.78 | 0.45 | 1.36 |
| Bottom (B) | Random forest | SNV+detrend | PCA | 8.42 | 0.36 | 1.25 |
| Bottom (B) | XGBoost | SG 2nd-deriv + detrend | Chemical bonds | 8.43 | 0.36 | 1.25 |
| Bottom (B) | 1D-CNN | SG 2nd-deriv + SNV | full spectrum | 8.40 | 0.36 | 1.26 |
| TB-avg | PLS | SG 1st-deriv + SNV+detrend | Chemical bonds | 7.68 | 0.47 | 1.38 |
| TB-avg | Ridge | MA(3) + SNV | Chemical bonds | 7.30 | 0.52 | 1.45 |
| TB-avg | ElasticNet | SG 1st-deriv + SNV+detrend | full spectrum | 7.56 | 0.48 | 1.40 |
| TB-avg | SVR | SG 1st-deriv + SNV | PCA | 7.98 | 0.43 | 1.32 |
| TB-avg | Random forest | SG 2nd-deriv + detrend | Chemical bonds | 8.03 | 0.42 | 1.31 |
| TB-avg | XGBoost | SG 2nd-deriv | Lasso | 8.27 | 0.38 | 1.28 |
| TB-avg | 1D-CNN | SG 1st-deriv + SNV+detrend | full spectrum | 8.53 | 0.34 | 1.24 |
| TB-concat | PLS | SG 1st-deriv + SNV+detrend | full spectrum | 7.60 | 0.48 | 1.39 |
| TB-concat | Ridge | SNV+detrend | Chemical bonds | 7.01 | 0.56 | 1.51 |
| TB-concat | ElasticNet | SG 1st-deriv + detrend | full spectrum | 7.41 | 0.51 | 1.43 |
| TB-concat | SVR | SG 1st-deriv + SNV | PCA | 7.93 | 0.43 | 1.33 |
| TB-concat | Random forest | SG 1st-deriv | full spectrum | 8.57 | 0.34 | 1.23 |
| TB-concat | XGBoost | SG 1st-deriv + SNV | PCA | 8.74 | 0.31 | 1.21 |
| TB-concat | 1D-CNN | SG 2nd-deriv + detrend | full spectrum | 7.52 | 0.49 | 1.40 |

**Table 2.** Qualitative High-Freshness/Low-Freshness classification (High Freshness if HU ≥ 70): for every measurement configuration and algorithm, the automatically selected best preprocessing and wavelength selection

| Configuration | Model | Preprocessing | Wavelengths | Accuracy | AUC | F1 |
|---|---|---|---|---|---|---|
| Top (T) | Logistic | SNV | Chemical bonds | 0.762 | 0.872 | 0.758 |
| Top (T) | PLS-DA | SG 1st-deriv + SNV | PCA | 0.767 | 0.830 | 0.783 |
| Top (T) | SVC | SG 1st-deriv + detrend | PCA | 0.756 | 0.835 | 0.773 |
| Top (T) | Random forest | SG + SNV | PCA | 0.777 | 0.818 | 0.784 |

**Table 2.** Qualitative High-Freshness/Low-Freshness classification (High Freshness if HU ≥ 70): for every measurement configuration and algorithm, the automatically selected best preprocessing and wavelength selection (cont)

| Configuration | Model | Preprocessing | Wavelengths | Accuracy | AUC | F1 |
|---|---|---|---|---|---|---|
| Top (T) | XGBoost | SG 1st-deriv + detrend | full spectrum | 0.751 | 0.796 | 0.765 |
| Top (T) | 1D-CNN | SG 1st-deriv + detrend | full spectrum | 0.782 | 0.820 | 0.790 |
| Bottom (B) | Logistic | SG 1st-deriv + SNV+detrend | full spectrum | 0.777 | 0.846 | 0.786 |
| Bottom (B) | PLS-DA | SG 1st-deriv + SNV+detrend | full spectrum | 0.746 | 0.823 | 0.763 |
| Bottom (B) | SVC | SG + detrend | PCA | 0.782 | 0.842 | 0.794 |
| Bottom (B) | Random forest | SG 2nd-deriv + SNV | full spectrum | 0.782 | 0.818 | 0.794 |
| Bottom (B) | XGBoost | SG 2nd-deriv | Chemical bonds | 0.808 | 0.853 | 0.818 |
| Bottom (B) | 1D-CNN | SG 1st-deriv + SNV | full spectrum | 0.772 | 0.859 | 0.773 |
| TB-avg | Logistic | SG + SNV+detrend | PCA | 0.767 | 0.850 | 0.769 |
| TB-avg | PLS-DA | SG 1st-deriv + SNV | Lasso | 0.782 | 0.867 | 0.798 |
| TB-avg | SVC | SG 1st-deriv + SNV | PCA | 0.772 | 0.851 | 0.768 |
| TB-avg | Random forest | MA(3) + SNV | PCA | 0.751 | 0.846 | 0.750 |
| TB-avg | XGBoost | SG 2nd-deriv | Chemical bonds | 0.824 | 0.867 | 0.837 |
| TB-avg | 1D-CNN | SNV | full spectrum | 0.710 | 0.794 | 0.723 |
| TB-concat | Logistic | SG 1st-deriv + SNV | Chemical bonds | 0.808 | 0.873 | 0.810 |
| TB-concat | PLS-DA | SG 1st-deriv + SNV | Chemical bonds | 0.803 | 0.871 | 0.810 |
| TB-concat | SVC | SG | PCA | 0.751 | 0.824 | 0.771 |
| TB-concat | Random forest | SG 1st-deriv + SNV | PCA | 0.756 | 0.838 | 0.771 |
| TB-concat | XGBoost | SG 2nd-deriv | Lasso | 0.798 | 0.860 | 0.812 |
| TB-concat | 1D-CNN | SG 1st-deriv + detrend | full spectrum | 0.756 | 0.825 | 0.777 |

The predicted-versus-measured plot (Figure 6a) shows the model captures the overall trend, with points clustering along the diagonal, but the fitted line is flatter than the 1:1 ideal: the model reads the freshest eggs a little low and the most-aged eggs a little high, pulling its predictions toward the middle of the range. This regression-to-the-mean effect is expected here because the HU distribution (Figure 2) is centered near the 70 HU mean and thins out toward both extremes - relatively few eggs are very fresh (only about 21% above 80 HU) or very aged (few below 50 HU) - so the model has little data at either end and leans toward the common mid-HU values.

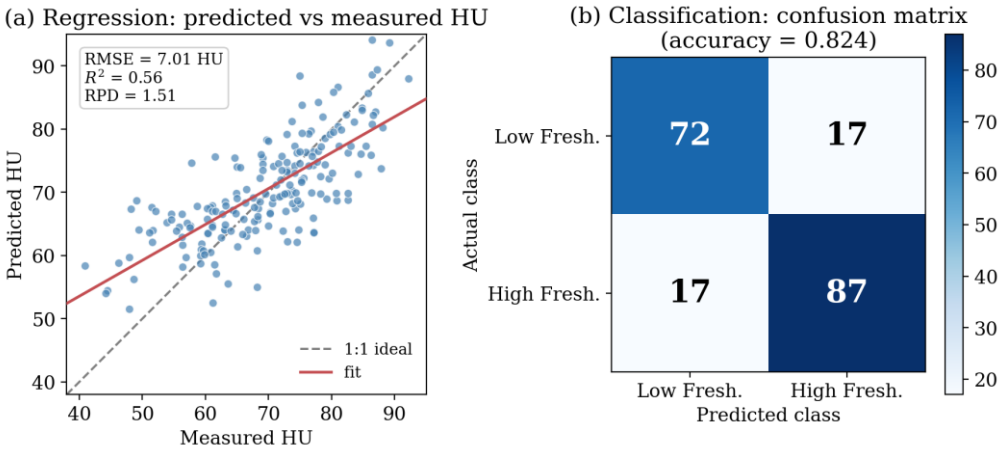


**Figure 6.** (a) Predicted versus measured Haugh unit for the best regression model (ridge regression on the concatenated top+bottom spectrum; RMSE = 7.0 HU). (b) Confusion matrix on the held-out test set for the best classifier (XGBoost on the averaged top+bottom spectrum; accuracy 82.4%)

## 4. CONCLUSIONS

We have built a platform for AI-driven egg freshness assessment that automates the complete NIR workflow; spectral ingestion, preprocessing, wavelength selection and AI model training and applied it to two complementary tasks: a quantitative task, predicting the Haugh unit, and a qualitative task, classifying each egg as High or Low Freshness against the HU ≥ 70 acceptance threshold used by egg-processing plants. For a given set of eggs it searches for the preprocessing, wavelength-selection and model combinations, selects the best one automatically by cross-validation, and records every choice so the analysis can be reproduced. On this sample of 642 hen eggs, the best regression reached an RMSE of 7.0 HU, against up to about 8.7 HU for poorer combinations, and the best classifier 82.4% accuracy, against about 71% for the weakest; in both tasks the best conventional model outperformed the deep-learning model, consistent with the modest sample size. The value of the platform lies not in any single accuracy figure but in how it is obtained: by searching the whole space it reliably settles on a strong model and avoids the poor combinations that an arbitrary or manual, one-off choice can easily pick, without expert tuning. Because it works non-destructively on eggs, the platform can replace destructive HU testing and allow far more eggs to be assessed before processing, helping to detect inconsistent freshness before it causes processing defects, product loss and food waste, and thereby contributing to food security.

## Acknowledgment

We would like to express our sincere gratitude to the Biotechnology Research and Innovation Center, the CP Innovation Center of Excellence, and the Charoen Pokphand Group, Thailand, for their support in providing access to the farm facilities, egg samples, and Haugh unit (HU) measurements. We also gratefully acknowledge the CUTIP program and the Department of Chemistry, Faculty of Science, Chulalongkorn University, for their academic guidance and support throughout this work.

## References

[1] Haugh, R. R. (1937). The Haugh unit for measuring egg quality. *U.S. Egg & Poultry Magazine*, *43*, 552–555, 572–573.
[2] Giunchi, A., Berardinelli, A., Ragni, L., Fabbri, A., & Silaghi, F. A. (2008). Non-destructive freshness assessment of shell eggs using FT-NIR spectroscopy. *Journal of Food Engineering*, *89*(2), 142–148.
[3] Wang, F., Lin, H., Xu, P., Bi, X., & Sun, L. (2021). Egg freshness evaluation using transmission and reflection of NIR spectroscopy coupled multivariate analysis. *Foods*, *10*(9), 2176.

[4] Cruz-Tirado, J. P., Medeiros, M. L. da S., & Barbin, D. F. (2021). On-line monitoring of egg freshness using a portable NIR spectrometer in tandem with machine learning. *Journal of Food Engineering*, *306*, 110643.
[5] Yuan, L., Fu, X., Yang, X., Chen, X., Huang, G., Chen, X., Shi, W., & Li, L. (2023). Non-destructive measurement of egg's Haugh unit by Vis-NIR with iPLS-Lasso selection. *Foods*, *12*(1), 184.
[6] Rinnan, Å., van den Berg, F., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, *28*(10), 1201–1222.
[7] Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, *36*(8), 1627–1639.
[8] Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, *43*(5), 772–777.