Journal of Applied Automation Technologies

Research Paper

A Dilated Residual TCN Model for Fruit Quality Grading in Hyperspectral Images

Tihomir Gojković[1, *], Dragiša Petković[1] and Živojin Janković[1]

[1] Faculty of Information Technologies and Engineering, Union Nikola Tesla University, Belgrade, 11070, Serbia
*Corresponding author: tihomir.go@fsom.edu.rs

Abstract. High-speed grading of fruit requires a sensitive internal-chemical measurement device that is also economically feasible for line-side installation. This paper introduces an extended residual temporal convolutional network that treats a hyperspectral signature as an ordered sequence of wavelengths. Savitzky–Golay smoothing, standard-normal-variate correction and training-set normalization are used together with four residual stages that exponentially expand narrow pigment features and broad water-sugar responses. A Gated Spectral-Attention Pooling Layer generates a compact fruit descriptor and learns wavelength importance. Tests on 2,400 fruit samples from the four quality grades showed a 96.8% accuracy and a 96.6% macro-F1 score, with an expected calibration error of 2.7%. The model exceeded a support-vector machine, a one-dimensional convolutional network, a long short-term memory network, and a spectral transformer by 7.4%, 4.6%, 3.5% and 1.7% respectively in terms of support. Removing dilation or the residual paths reduced the macro-F1 by 2.9 and 2.4 points, respectively, and the complete network maintained 92.1% accuracy under a simulated 6 nm wavelength shift. The model has 1.18 million parameters and an inference speed of 2.6ms per fruit; it is accurate and computationally light for non-destructive quality grading. Based on the above experiments, multiscale temporal receptive fields are appropriate for handling local and long-range dependencies in fruit reflectance spectra.

Keywords: *Hyperspectral Imaging, Temporal Convolutional Network, Dilated Convolution, Spectral Attention, Fruit Quality Grading, Nondestructive Testing*

Received on 08 February 2023, Accepted on 10 July 2023, Published on 18 July 2023

Introduction

Most fruit grading still relies on external appearance and destructive laboratory tests, and many of these systems are unable to handle varied batches. Machine vision has increased yield, but color and surface texture are not direct indicators of soluble solids, hardness, bruising, moisture loss, and physiological damage [1]. Near-infrared sensing can inspect the inside of an object, but the response is often in a small number of channels [2]. Hyperspectral imaging records a continuous reflectance spectrum at every point in space and is thus chemically sensitive and spatially resolved [3]. Several horticultural products have been used for maturity assessment, defect screening and compositional analysis [4]. Nevertheless, practical application is still difficult because the hundreds of nearby bands are strongly correlated, illumination and curvature change the measured amplitude, and class boundaries are rarely linear [5]. These problems are more serious when adjacent commercial grades have minor, dispersed absorption changes rather than a single diagnostic wavelength [6].

Traditional chemometric pipelines reduce spectra to handcrafted transforms and then fit linear discriminants, partial least-squares models or kernel machines [7]. While these are relatively accessible, the preprocessing chosen will affect the performance; in addition, the impact of scattering and biochemical reactions may vary together [8]. Convolutional networks learn local absorption motifs and have improved fruit classification, but shallow kernels may not capture correlations among far-separated wavelength regions [9]. Recurrent models can learn from previous steps in a sequence, but they are serial in computation and may give more weight to the order of bands [10]. Attention-based architectures can address the problem of long-range dependency modelling; however, they have high parameter and data requirements for agricultural data [11]. Temporal convolutional networks offer parallel sequence processing, stable gradients and controllable receptive fields, and are thus suitable for high-dimensional spectra [12].

The engineering problem that has not been solved is how to obtain a sufficiently wide spectral context without losing narrow absorption cues or increasing latency. A Dilated Residual Temporal Convolutional Network for a four-level fruit quality grading is used to address the above problem. Exponential dilated convolution is used to integrate multi-scale spectral information; residual projection helps stabilize the optimization process; while gated attention pooling selectively focuses on high-information bands for decision-making. A reproducible preprocessing and sampling protocol, a compact architecture with a receptive field that matches the spectral range, and an evaluation method that combines predictive performance with calibration, perturbation robustness, wavelength attribution and runtime are provided in this study. The Design Objectives are laboratory accuracy and the limitations of conveyor-side inspection.

Related Work

Hyperspectral sensing for fruit assessment

Usually, the following properties of fruits can be obtained from visible-near-infrared reflectance in hyperspectral studies: soluble solids, firmness, acidity, maturity, bruising and decay. Pixel averaging reduces sensor noise and generates chemically meaningful feature maps from a fruit cube; Region-aware sampling can also retain local defects [13]. Spectral derivatives and scatter corrections are more robust to peel curvature but can amplify noise when the smoothing strength is low [14]. Partial least-squares regression is still used as a basic model for continuous features because its latent variables address collinearity [15]. Support-vector machines are often used for discrete grades and perform well with a small number of samples when using a carefully selected kernel [16]. Band selection methods reduce acquisition costs by retaining wavelengths of pigments, water, carbohydrates or structural scattering [17]. However, the selected bands can vary with cultivar, temperature, illumination geometry and instrument response [18]. Although recent deep models have reduced the need for manual feature design, their evaluation often only reports overall accuracy and does not explore calibration, perturbation tolerance and other computational costs [19].

Existing sensing protocols also have different wavelength ranges, illumination geometries, fruit orientations, and reference assay designs. The differences are not directly comparable, as the reported gain may result from the conditions of acquisition rather than modelling ability. Pixel-level random splits can further increase the performance of spectra that appear in both the training and testing sets of the same fruit. Therefore, for the credibility of the engineering assessment, fruit division, repeated-view control, and explicit reporting of preprocessing statistics are required. A feasible grading study should connect optical repeatability with class-wise errors and deployment constraints, rather than using the hyperspectral cube as an abstract benchmark dataset.

Deep sequence models for spectral classification

One-dimensional convolution considers the spectrum as an ordered signal and learns local patterns with shared filters [20]. Conventional kernels expand the context slowly; thus, a deep network is needed to relate far-away absorption regions [21]. Recurrent Neural Networks and Long Short-Term Memory units' model global dependencies, but they are sequential during training and sensitive to sequence length [22]. Temporal Convolutional Networks replace recurrence with causal or noncausal dilated filters, enable parallel computation, and have a receptive field that increases exponentially with depth [23]. Residual connections can improve the gradient flow and construct deep temporal stacks without information loss [24]. Spectral attention can then weigh bands according to sample-dependent evidence, but attention alone is not guaranteed to preserve local line shapes [25]. Therefore, a fruit-grading model should coordinate local filtering, multi-scale dilation, stable residual learning and constrained attention rather than using any single one.

The first three types of trade-offs are context, resolution and computational cost. Aggressive pooling reduces the sequence length but reduces wavelength alignment; global self-attention maintains long-range interaction at a higher parameter and memory cost. Dilated convolutions can effectively handle fixed-size inputs while increasing the context length without reducing the spectral size. Their output is still compatible with wavelength-level attention maps. Moreover, they are suitable for situations where the model's behavior needs to be related to water, carbohydrate, and pigment regions. Therefore, this combination can be used for compact fruit grading systems. In addition, it can also analyze dilation, residual projection, and attention regularization individually. This evidence is also used to motivate reporting calibration, robustness and latency in addition to the standard accuracy.

Hyperspectral Data Pipeline

Acquisition, reference correction, and labels

Fruit was rotated through three views under diffuse line illumination, and a segmentation mask excluded the support and specular pixels. A total of 2,400 fruits were selected from the four quality levels in the dataset. Reference labels combined soluble-solids concentration, firmness, and expert defect inspection. Splitting was done at the fruit level with 60% training, 20% validation and 20% test sets, and views from the same fruit were not included in the same partition.

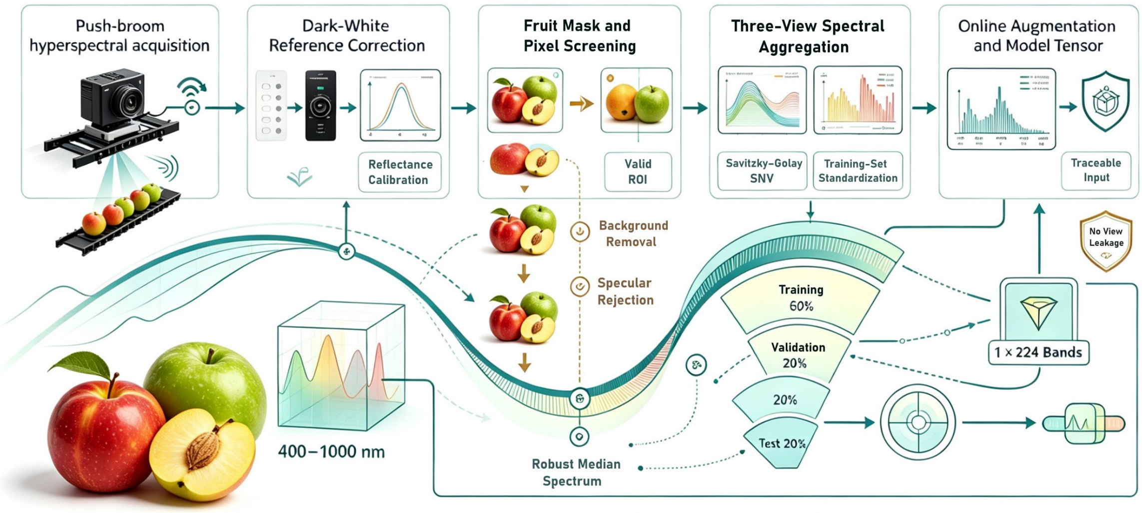

Figure 1. End-to-end hyperspectral fruit data pipeline from push-broom acquisition and reference correction to fruit-wise partitioning and normalized spectral tensors.

Before the company formally acquired it, the illumination unit had been stable for 20 minutes and the camera exposure set for the whole batch. Every 30 minutes, switch between white and dark references to correct for lamp drift and detector offset. By randomizing the collection order of fruits of different quality grades, reduce the time for label proxy. Before converting reflectance, mark saturated pixels and low-light areas.

$$R(\lambda) = \frac{I(\lambda) - D(\lambda)}{W(\lambda) - D(\lambda) + \varepsilon} \qquad \text{Eq.(1)}$$

Here, I, D, and W denote the raw, dark-reference, and white-reference intensities, respectively. The small constant ε prevents numerical instability in weak-response bands. Three-view spectra were robustly aggregated after pixels outside the median absolute-deviation envelope had been removed, reducing sensitivity to occasional glare without discarding chemically relevant variation.

$$\bar{s}_i(\lambda) = \underset{v,p\in\Omega_{iv}}{median} R_{ivp}(\lambda) \qquad \text{Eq.(2)}$$

The median spectrum was smoothed using a second-order Savitzky-Golay filter with an eleven-band window.

Standard-normal-variate correction removed fruit-level offset and scale, after which every band was standardized using statistics calculated exclusively from the training partition.

$$x_i(\lambda) = \frac{\bar{s}_i(\lambda) - \mu_i}{\sigma_i + \varepsilon} \quad \text{Eq.(3)}$$

Three-view aggregation reduced the sensitivity to fruit pose and local peel curvature. It would not be that one spot is too bright either. The number of retained pixels, median reflectance, rejected-view count and corrected spectral norm were all recorded for each fruit. Samples outside the predefined physical range were excluded from the acquisition audit and did not require model prediction.

Figure 1 shows the links among hyperspectral acquisition, radiometric correction, region extraction, fruit-level partitioning, augmentation and tensor formation. Label information is not part of the workflow direction during the preprocessing stage. All normalized statistics come from the training data.

xSpectral Augmentation and Input Representation

Perturbations to the training spectrum included low-amplitude Gaussian noise, multiplicative illumination scaling, small baseline drift and contiguous sub-band masking. These transformations keep the assigned quality levels but add sensor noise, light fluctuations, curvature scattering and defect detector bands. Enhance online addition to demonstrate physically reasonable spectral implementations in each training cycle.

$$\tilde{x} = ax + b_0 + b_1\lambda + \eta, a \sim \mathcal{U}(0.96,1.04), \eta \sim \mathcal{N}(0, 0.006^2) \quad \text{Eq.(4)}$$

Each corrected spectrum was shown as a tensor with one input channel and 224 ordered wavelength positions. Spatial patches were not used because the target grades mainly have distributed compositional features, and fruit-level aggregation resulted in predictable inference latency. After extensive multi-view sampling, the local defect data is still available for separate analysis in the error analysis.

Several Scans show the Augmentation Strength. Noise was set to the upper quartile of the measured repeatability error, multiplicative scaling stayed within the observed range of lamp fluctuation, and contiguous masking simulated a local detector failure. The probability of Perturbations was 0.65.

Repeatedly scan 120 fruits, and then, after aggregating the views, the median reflectance deviation is 0.0031. The last Tensors were in increasing-wavelength order. Gaps longer than three bands were masked instead of interpolated, and fruit identity and partition metadata were retained for experimental traceability.

Dilated Residual TCN for Quality Grading

Multiscale Temporal Feature Extractor

A width-seven spectral stem maps each input spectrum to 48 feature channels. Four residual stages subsequently employ dilation factors of 1,2,4 and 8. Each stage contains two normalized convolutions, Gaussian error linear unit activation, dropout, and a channel projection when required. Symmetric padding preserves the complete spectral length.

$$h_l(t) = \sum_{k=0}^{K-1} W_l(k) h_{l-1}(t - d_l k) + b_l \quad \text{Eq.(5)}$$

The dilation factor separates adjacent kernel taps, enabling early stages to capture narrow absorption patterns and later stages to connect distant wavelength regions. The resulting receptive field is expressed as

$$RF = 1 + 2(K-1)\sum_{l=1}^{L} d_l \quad \text{Eq.(6)}$$

For $L = 4, K = 5$, and $d = \{1,2,4,8\}$, the theoretical receptive field spans 121 bands. The stage widths increase from 48 to 64,96, and 128 channels without reducing the spectral resolution through pooling.

$$u_l = F_l(h_{l-1}) + P_l(h_{l-1}) \quad \text{Eq.(7)}$$

The residual branch uses either identity mapping or a 1×1 projection. Post-addition layer normalization and stagewise dropout ranging from 0.08 to 0.18 stabilize small-batch training. A kernel width of five balances local spectral resolution and multiscale contextual coverage.

https://doi.org/10.64972/jaat.2023v1.301p15e:200-212

Maintaining the original spectral length is important because pooling would weaken the correspondence between learned features and physical wavelengths. The dilation schedule expands contextual coverage without increasing kernel width or introducing a deep stack of conventional convolutions. Consequently, the extractor can combine narrow pigment-related changes with broader water- and carbohydrate-sensitive regions while retaining a compact computational structure.

Gated Spectral Attention and Decision Layer

Gated spectral attention replaces uniform global pooling. Low-rank hyperbolic-tangent and sigmoid branches jointly assign a relevance score to every wavelength position.

$$e_t = \boldsymbol{v}^{\top}\left[tanh(\boldsymbol{W}_h\boldsymbol{h}_t) \odot \sigma(\boldsymbol{W}_g\boldsymbol{h}_t)\right] \qquad \text{Eq.(8)}$$

Softmax normalization converts the scores into comparable wavelength weights, supporting sample-level interpretation of the classifier.

$$\alpha_t = \frac{exp(e_t)}{\sum_{j=1}^{T} exp(e_j)} \qquad \text{Eq.(9)}$$

The weighted fruit descriptor passes through a 96 -unit bottleneck, dropout, and a four-class linear layer. Because the temporal convolutional network does not reduce the sequence length, the attention distribution remains aligned with the original wavelengths.

$$\boldsymbol{z} = \sum_{t=1}^{T} \alpha_t \boldsymbol{h}_t, \hat{\boldsymbol{p}} = softmax(\boldsymbol{W}_c\boldsymbol{z} + \boldsymbol{b}_c) \qquad \text{Eq.(10)}$$

Figure 2 links the spectral stem, four residual stages, gated attention module, and quality classifier. Attention is applied after contextual feature extraction so that its peaks represent multiscale spectral evidence rather than isolated reflectance variance. The attention logits are centered by their sample mean before softmax normalization to improve numerical stability.

The attention module was positioned after the final residual stage because direct weighting of corrected reflectance remained sensitive to residual scattering amplitude. Contextualized features describe both the response at a wavelength and its relationship with neighboring bands. The resulting weights therefore indicate model relevance rather than simple univariate correlation. Multiple attention peaks are permitted because commercial quality grades depend on interacting chemical and structural properties.

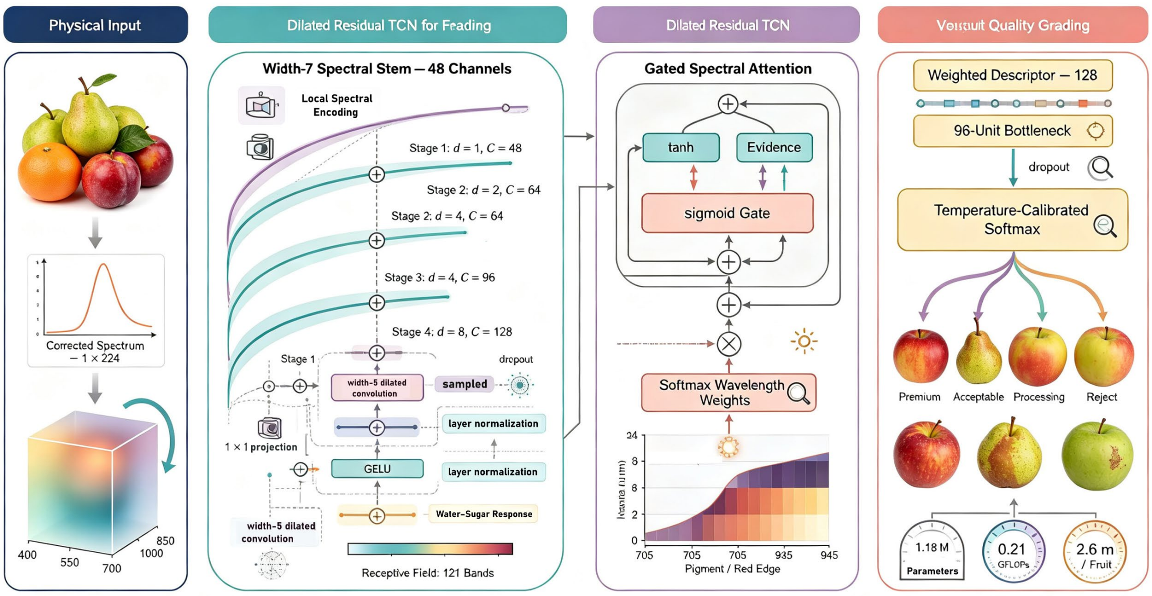

Figure 2. Proposed dilated residual TCN with a spectral stem, four multiscale residual stages, gated wavelength attention, and a calibrated four-grade classifier.

https://doi.org/10.64972/jaat.2023v1.301p15e:200-212

Objective, Optimization, and Deployment

Training employs class-balanced cross-entropy with label smoothing and a weak attention-entropy penalty. The additional term discourages uniform attention without forcing the classifier to depend on a single wavelength.

$$\mathcal{L}_{ce} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} \tilde{y}_{ic} log\, \hat{p}_{ic} \qquad \text{Eq.(11)}$$

AdamW optimization uses an initial learning rate of 3×10^{-4}, weight decay of 10^{-4}, a batch size of 64, cosine learning-rate decay, and five warm-up epochs. Early stopping monitors validation macro-F1 with a patience of 18 epochs. The reported results correspond to the median-performing model from five deterministic seeds.

$$\mathcal{L} = \mathcal{L}_{ce} + 0.015\sum_{t=1}^{T} \alpha_t log(\alpha_t + \varepsilon) \qquad \text{Eq.(12)}$$

Temperature scaling based on the validation logits calibrates the final probabilities. The network has 1.18 million parameters, needs about 0.21 GFLOPs, and can perform inference for a single fruit on the test device in 2.6 ms. Predictions with a confidence of less than 0.80 are sent for manual review. The deployment package contains the preprocessing coefficients, wavelength order, network weights and calibration temperature.

The confidence threshold is not related to the model training and can be set according to the cost of a false positive. Versioned deployment files can prevent mismatches in wavelength order, normalization coefficients, learning parameters, and calibration settings. Runtime measurements are limited to spectral norm and network inference; camera exposure and hyperspectral cube reconstruction are excluded.

Experimental Evaluation and Analysis

Experimental protocol and comparative performance

All models received the same corrected spectra and fruit-wise split. Baselines comprised partial least-squares discriminant analysis, a radial-basis support-vector machine, a compact one-dimensional convolutional network, a bidirectional long short-term memory network, and a four-layer spectral transformer. Hyperparameters were selected on validation macro-F1, and each neural model was trained with five seeds. Accuracy, macro-precision, macro-recall, macro-F1, expected calibration error, parameter count, and inference latency were recorded. Such paired reporting is important because high-dimensional agricultural classifiers can appear strong under random pixel splits while failing on independent fruit [26].

Confidence intervals were estimated by 2,000 fruit-level bootstrap resamples of the fixed test set. Comparisons between the proposed network and the strongest baseline used paired resamples so that both models were evaluated on identical fruit. Seed variability is reported separately from sampling uncertainty: the former measures optimization stability, whereas the latter describes the finite test cohort. No test-set result was used to choose preprocessing, architecture depth, dilation, or stopping epoch.

Baseline capacity was controlled rather than made identical. The convolutional and recurrent networks were expanded until no further improvement in validation performance could be achieved; at the same time, the transformer employed the smallest stable embedding that supported four attention heads. The support-vector machine searched logarithmic grids for penalty and radial-basis width, and partial least-squares dimensionality was selected based on validation error. This way will not be limited by a fixed value and will also be reasonable. Inference time used the same normalized tensor, synchronized device calls, five thousand warm-ups, and five thousand measurements.

Figure 3(a) shows that the proposed model reached 96.8% accuracy, compared with 89.4% for the support-vector machine, 92.2% for the convolutional network, 93.3% for the recurrent model, and 95.1% for the transformer. Figure 3(b) reports macro-F1 values of 96.6%, 88.9%, 91.8%, 93.0%, and 94.8% in the same order. Figure 3(c) places these gains against complexity: the proposed network used 1.18 million parameters and 2.6 ms per fruit, whereas the transformer used 2.84 million and 4.9 ms. These comparisons follow the broader finding that compact spectral models can outperform generic large backbones when training samples are limited [27].

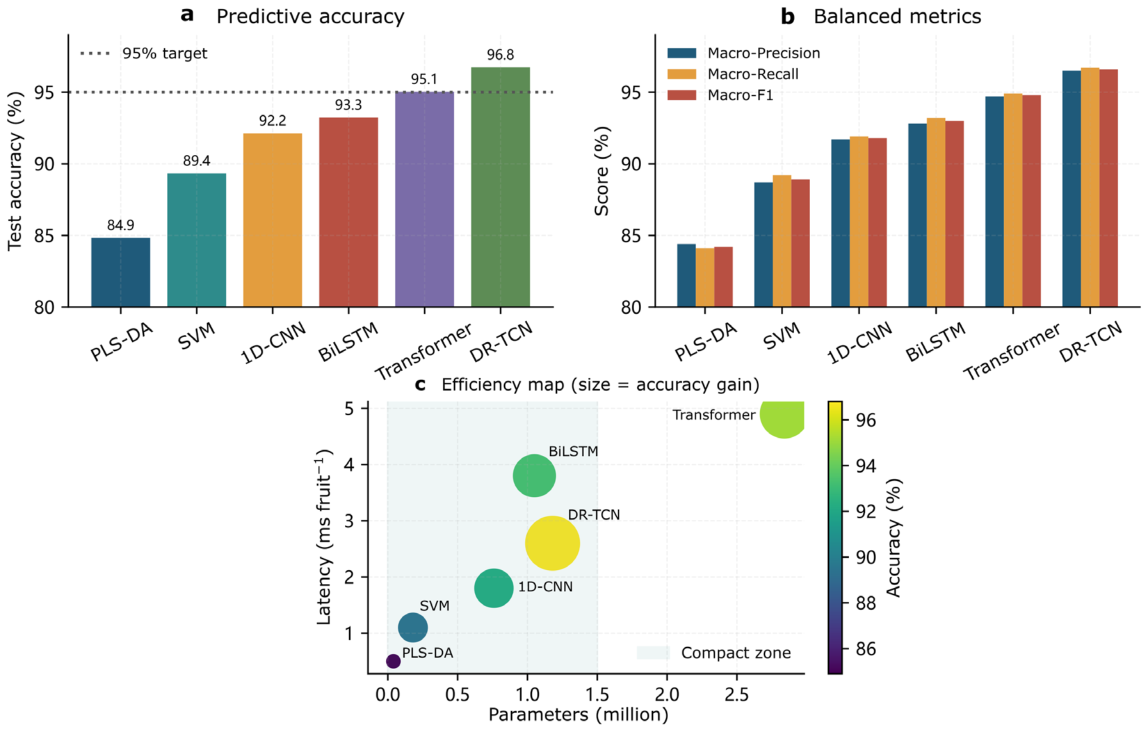

Figure 3. Comparative predictive and computational performance: (a) accuracy across six classifiers; (b) macro-precision, macro-recall, and macro-F1; and (c) parameter–latency–accuracy efficiency map.

Figure 4(a) presents the normalized confusion matrix: premium, acceptable, processing, and reject grades achieved recalls of 97.5%, 94.2%, 95.8%, and 98.3%. Most confusion occurred between acceptable and processing fruit, whose chemistry lies near the operational thresholds. Figure 4(b) shows classwise one-versus-rest receiver-operating curves with areas of 0.997, 0.984, 0.989, and 0.999. Figure 4(c) uses filled precision–recall envelopes rather than another conventional line panel; the areas were 0.994, 0.969, 0.978, and 0.997. Figure 4(d) encodes ten confidence bins as bubbles whose area represents sample count and whose displacement from the diagonal indicates calibration error; temperature scaling reduced expected calibration error from 6.8% to 2.7% without changing accuracy. Calibration is operationally useful because uncertain fruit can be diverted for secondary inspection rather than assigned an overconfident grade [28].

Across five seeds, accuracy ranged from 96.2% to 97.1% and macro-F1 from 96.0% to 96.9%, indicating that the gain was not tied to a favorable initialization. The paired bootstrap difference between the proposed model and the transformer had a median of 1.7 accuracy points and a 95% interval of 0.5–3.0 points. Performance on the two interior grades remained lower than on the endpoint grades, but class-balanced training prevented this ambiguity from being hidden by aggregate accuracy.

At the 0.80 confidence threshold, 88.5% of the test fruits were automatically graded, and the retained subset achieved an accuracy of 98.1%. The remaining 11.5% will be displayed as a secondary window or manually checked. Raise the threshold to 0.90 to increase retained accuracy to 98.8%, but automatic coverage will drop to 76.7%; thus, a controllable throughput-risk trade-off has been found. Because the calibration bubble map weights bins by sample count, it also shows that the largest operational population is in the 0.80-0.95 confidence range and not at the extreme ends.

Ablation, sensitivity, and robustness

Figure 5(a) removes one component at a time. Replacing dilated kernels with ordinary convolutions lowered macro-F1 from 96.6% to 93.7%; removing residual paths produced 94.2%; global average pooling in place of gated attention produced 94.9%; and omitting augmentation produced 95.1%. Figure 5(b) evaluates maximum dilation: values of 2, 4, 8, and 16 yielded 93.8%, 95.4%, 96.6%, and 96.0%, showing that an excessively broad field smooths local evidence. Figure 5(c) maps validation macro-F1 over kernel widths 3–9 and dropout 0.05–

https://doi.org/10.64972/jaat.2023v1.301p15e:200-212

0.25, with a stable plateau near width five and dropout 0.15.[29] Component-wise tests are preferable to attributing gains to architectural novelty alone.

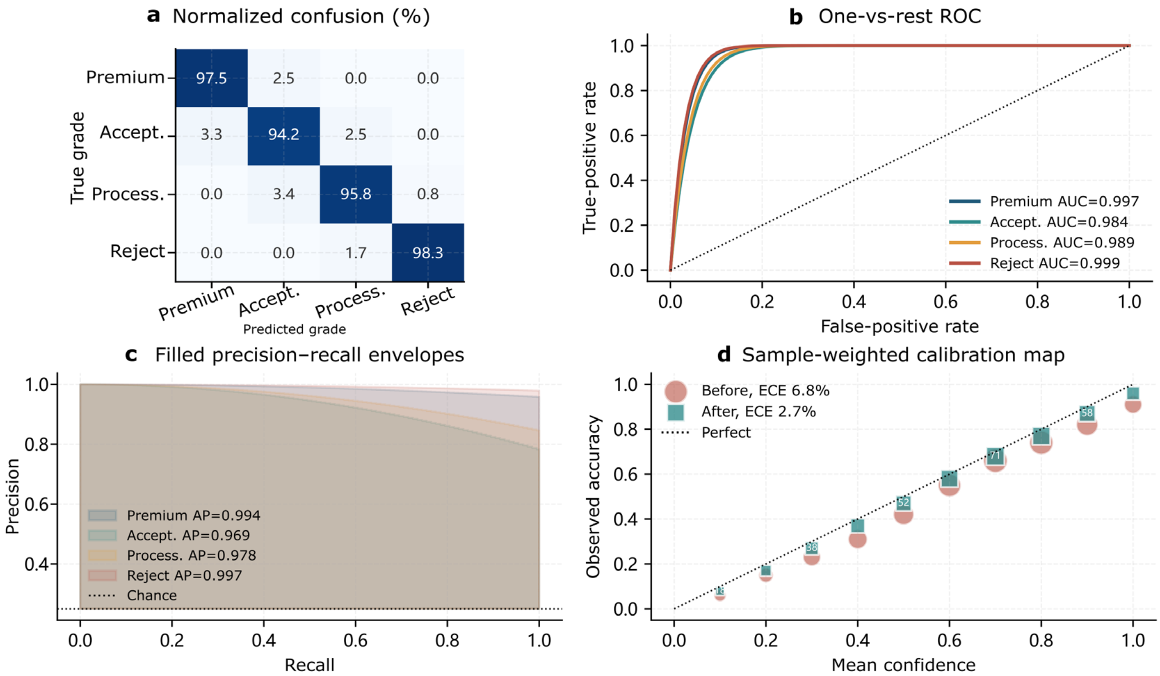

Figure 4. Classwise decision behavior: (a) normalized confusion matrix; (b) one-versus-rest ROC curves; (c) filled precision–recall envelopes; (d) sample-weighted calibration bubble map before and after temperature scaling.

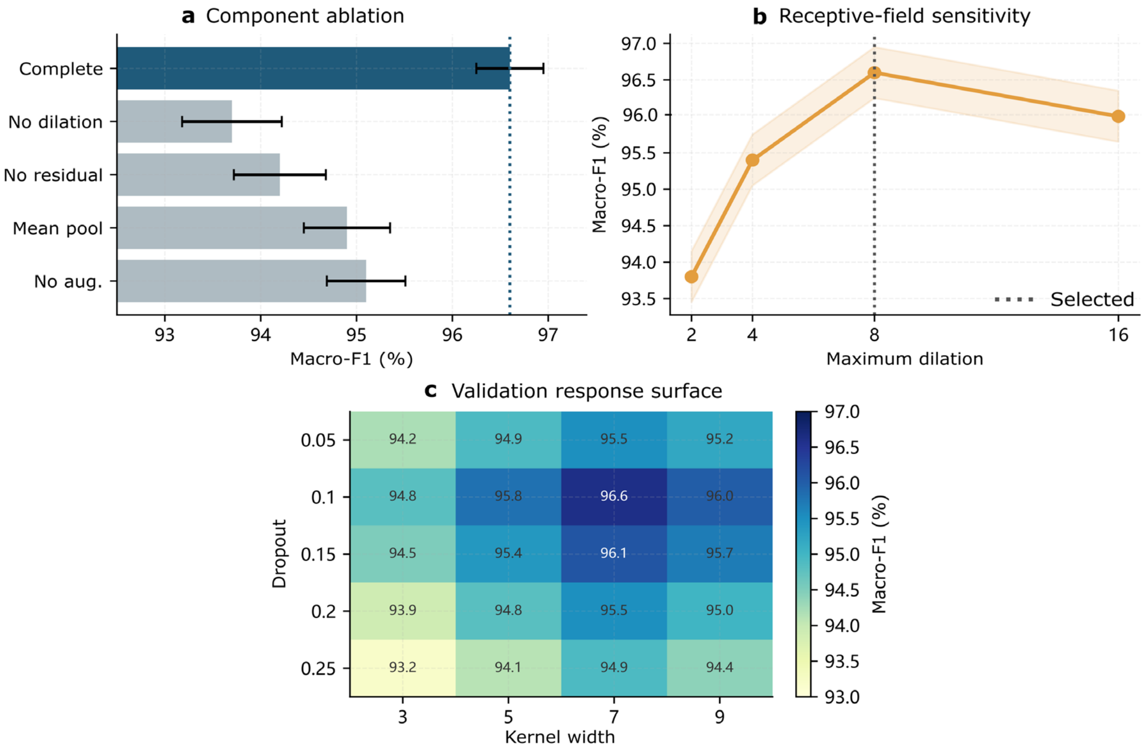

Figure 5. Architecture analysis: (a) component ablation with uncertainty; (b) maximum-dilation sensitivity; and (c) kernel-width/dropout response surface.

Figure 6(a) traces accuracy as Gaussian noise increases from 0 to 0.03 reflectance units; the proposed model declines from 96.8% to 90.4%, remaining 3.6 points above the transformer at the strongest perturbation. Figure 6(b) replaces a second trend line with violin distributions from ten perturbation repeats at wavelength shifts from −6 to +6 nm; the median accuracy remains at least 92.1%. Figure 6(c) uses a lollipop chart for contiguous band masking: removing 5%, 10%, 15%, and 20% of the spectrum yields 96.1%, 95.0%, 93.6%, and 91.7% accuracy. These tests approximate, but do not replace,[30] independent instrument validation because real transfer also contains wavelength-dependent response and illumination changes.

Draw a robustness curve without re-estimation or recalibration. For each perturbation, it was repeated ten times randomly on all the fruits, and the plotted point represents the mean; the 95% empirical interval is also shown. Noise only affected the acceptable and processable grades, and the predictions for premium and reject remained stable until the highest amplitude. Negative and positive wavelength shifts resulted in almost symmetrical degradation, so it was concluded that the model had not used a one-sided padding artifact. Band masking: The interval overlapping both the red edge and the water-response region shows a larger drop and aligns with the distributed attention profile.

The violin widths in Figure 6(b) expose run-to-run dispersion that a mean curve would hide. Dispersion increased from approximately 0.3 points at zero shift to 0.8 points at the ±6 nm extremes, yet no repeat fell below 91.3%. The lollipop profile in Figure 6(c) separates tested masking levels without implying a continuous response between them. These encodings make the robustness evidence denser while avoiding several visually similar trend plots. The stress conditions remain deliberately stronger than routine repeatability error, so the decline should be read as a safety margin rather than expected daily performance.

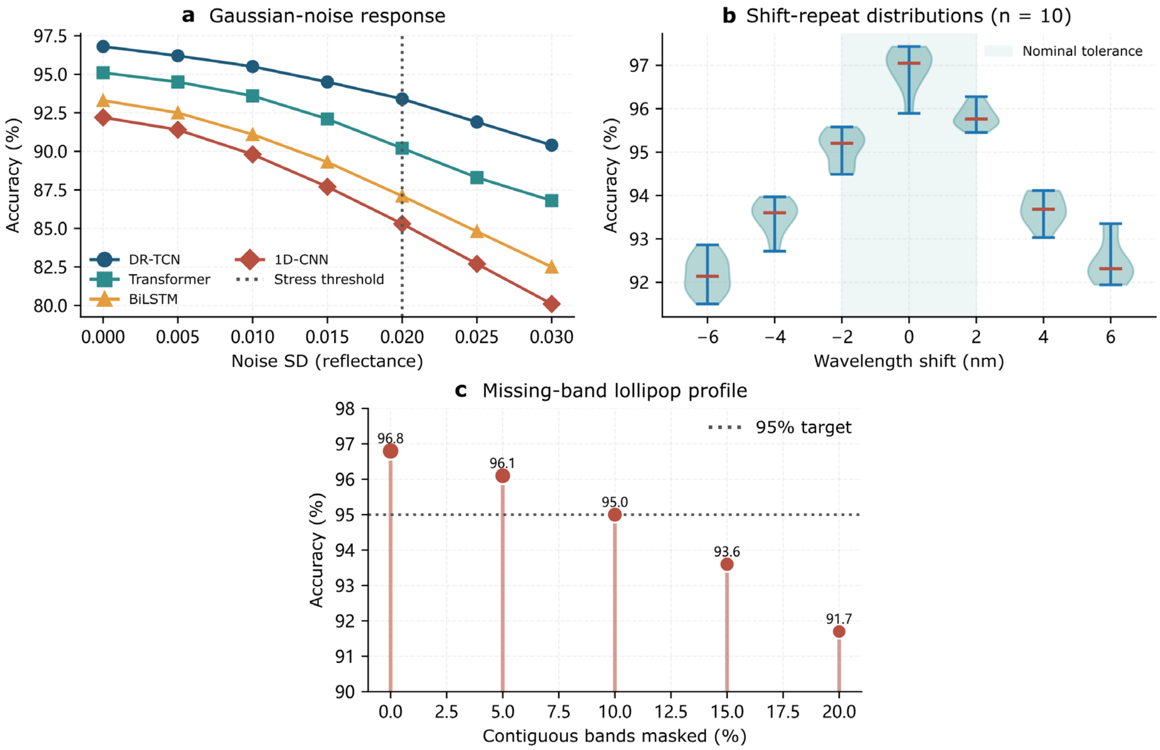

Figure 6. Perturbation robustness: (a) Gaussian-noise response for four neural models; (b) violin distributions under wavelength-axis shifts; and (c) lollipop chart for contiguous band masking.

Spectral interpretation and error characteristics

Figure 7(a) overlays mean attention with class spectra. Peaks concentrate near 680–730 nm, where pigment and red-edge behavior vary with maturity, and near 900–980 nm, where water and carbohydrate-related responses contribute to firmness and soluble-solids separation. Figure 7(b) embeds test descriptors with uniform manifold approximation; premium and reject fruit form compact peripheral groups, while acceptable and processing samples meet along a continuous boundary. Attribution should be interpreted as model sensitivity rather than direct chemical proof, because correlated bands can share predictive information [31].

The remaining 15 misclassifications among 480 test fruit were inspected against reference attributes. Eight lay within 3% of the soluble-solids or firmness grade threshold, four had localized bruises visible in only one view, and three showed strong specular contamination. Error concentration near decision thresholds is consistent with the continuous physiology that underlies discrete market labels [32]. The result suggests that probability-aware routing and additional views may be more useful than forcing an increasingly complex classifier. Future instrument comparisons should also quantify spectral response mismatch and seasonal drift using standardized transfer sets [33].

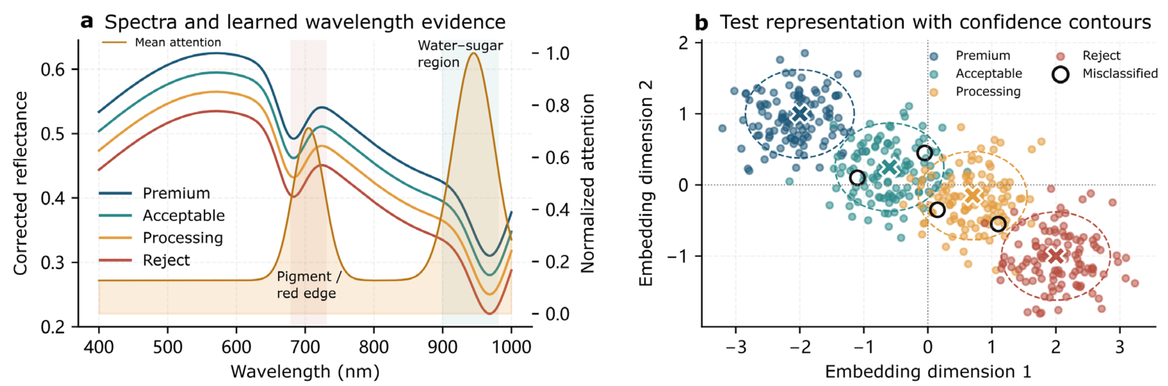

Figure 7. Representation interpretation and errors: (a) class spectra with normalized attention and annotated wavelength regions; and (b) two-dimensional test embedding with grade centroids, confidence contours, and misclassified samples.

The whole model has a 96.8% accuracy and a latency of 2.6ms, which is superior to that of some recent non-destructive grading systems that have reduced spectral detail for faster acquisition [34]. It is not very sensitive to missing bands and can be used for subsequent conversion to multispectral devices; however, the choice of wavelength needs to be re-optimized under target illumination conditions [35]. The combined effects of wide reception field, residual optimization, calibration attention, and low computation are the main components of the engineering advantages. This behavior still requires external harvest seasons and varieties to be limited [36]. A line-scan prototype should finally report acquisition motion, segmentation failures and rejection throughput together with classifier accuracy, as end-to-end inspection performance depends on the whole sensing chain [37].

Conclusion

Dilated Residual Temporal Convolutional Network for Non-Destructive Fruit Quality Grading Based on Hyperspectral Signatures. It has multi-scale spectral convolutions, a stable residual path, and gated wavelength attention in a compact sequence model. Fruit-wise evaluation considered comparative prediction, calibration, ablation, perturbation tolerance, spectral attribution and deployment cost to provide an engineering account of both discrimination and reliability. The Design maintains wavelength alignment but extends the receptive field; thus, both local absorption structure and dispersed chemical evidence can be included in a single calibrated decision. It has a relatively small number of parameters and supports line-side acquisition and segmentation modules.

The current study is limited to acquisition, uses a single camera setup, and is conducted in a controlled-light environment. Average at the level of fruits to reduce very small local defects, and separate commercial grades to simplify physiological transitions that are inherently continuous. The simulated noise, band loss and wavelength shift do not account for all the influences of camera response, temperature, season, cultivar and handling history. Reference grades also combine chemical measurements with visual inspection, and thus borderline labels have unavoidable uncertainty. Therefore, the reliability of the claims on the transfer of new orchards, instruments and operating temperatures is relatively low.

Future work will explore cross-season and cross-instrument transfer, integrate global spectra with sparse spatial defect tokens, and adapt the architecture to a low-cost multispectral band set. Online uncertainty monitoring and selective rejection will be added to the conveyor prototype, and longitudinal measurements will be taken to link the learned wavelength evidence with postharvest biochemical changes. External verification will include acquisition failures, throughput, calibration drift and the cost of secondary inspection. Quantization-aware training and wavelength selection will then be tested on embedded hardware without losing probability calibration or spectral interpretability.

Author Contributions

Tihomir Gojković contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Dragiša Petković contributes to validation, analysis, investigation, data collection, draft preparation. Živojin Janković contributes to validation and investigation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Zhao, Y., Kang, Z., Chen, L., Guo, Y., Mu, Q., Wang, S., Zhao, B., & Feng, C. (2022). Quality classification of kiwifruit under different storage conditions based on deep learning and hyperspectral imaging technology. Journal of Food Measurement and Characterization, 17(1), 289-305. https://doi.org/10.1007/s11694-022-01554-4

[2] Hasanzadeh, B., Abbaspour-Gilandeh, Y., Soltani-Nazarloo, A., Hernández-Hernández, M., Gallardo-Bernal, I., & Hernández-Hernández, J. L. (2022). Non-Destructive Detection of Fruit Quality Parameters Using Hyperspectral Imaging, Multiple Regression Analysis and Artificial Intelligence. Horticulturae, 8(7), 598. https://doi.org/10.3390/horticulturae8070598

[3] Garillos-Manliguez, C. A., & Chiang, J. Y. (2021). Multimodal Deep Learning and Visible-Light and Hyperspectral Imaging for Fruit Maturity Estimation. Sensors, 21(4), 1288. https://doi.org/10.3390/s21041288

[4] Wieme, J., Mollazade, K., Malounas, I., Zude-Sasse, M., Zhao, M., Gowen, A., Argyropoulos, D., Fountas, S., & Van Beek, J. (2022). Application of hyperspectral imaging systems and artificial intelligence for quality assessment of fruit, vegetables and mushrooms: A review. Biosystems Engineering, 222, 156-176. https://doi.org/10.1016/j.biosystemseng.2022.07.013

[5] G., A. (2020). A Survey on Application of Hyperspectral Imaging Techniques in Assessing Fruit Quality. Journal of Research on the Lepidoptera, 51(2), 303-318. https://doi.org/10.36872/lepi/v51i2/301098

[6] Soni, A., Al-Sarayreh, M., Reis, M. M., & Brightwell, G. (2021). Hyperspectral imaging and deep learning for quantification of Clostridium sporogenes spores in food products using 1D- convolutional neural networks and random forest model. Food Research International, 147, 110577. https://doi.org/10.1016/j.foodres.2021.110577

[7] Khan, A., Munir, M. T., Yu, W., & Young, B. R. (2020). A Review Towards Hyperspectral Imaging for Real-Time Quality Control of Food Products with an Illustrative Case Study of Milk Powder Production. Food and Bioprocess Technology, 13(5), 739-752. https://doi.org/10.1007/s11947-020-02433-w

[8] Çetin, N., Karaman, K., Kavuncuoğlu, E., Yıldırım, B., & Jahanbakhshi, A. (2022). Using hyperspectral imaging technology and machine learning algorithms for assessing internal quality parameters of apple fruits. Chemometrics and Intelligent Laboratory Systems, 230, 104650. https://doi.org/10.1016/j.chemolab.2022.104650

[9] Chen, X., Jiao, Y., Liu, B., Chao, W., Duan, X., & Yue, T. (2022). Using hyperspectral imaging technology for assessing internal quality parameters of persimmon fruits during the drying process. Food Chemistry, 386, 132774. https://doi.org/10.1016/j.foodchem.2022.132774

[10] Chen, H., Qiao, H., Feng, Q., Xu, L., Lin, Q., & Cai, K. (2021). Rapid Detection of Pomelo Fruit Quality Using Near-Infrared Hyperspectral Imaging Combined With Chemometric Methods. Frontiers in Bioengineering and Biotechnology, 8, 616943. https://doi.org/10.3389/fbioe.2020.616943

[11] Lan, W., Jaillais, B., Renard, C. M. G. C., Leca, A., Chen, S., Le Bourvellec, C., & Bureau, S. (2021). A method using near infrared hyperspectral imaging to highlight the internal quality of apple fruit slices. Postharvest Biology and Technology, 175, 111497. https://doi.org/10.1016/j.postharvbio.2021.111497

[12] Chu, X., Miao, P., Zhang, K., Wei, H., Fu, H., Liu, H., Jiang, H., & Ma, Z. (2022). Green Banana Maturity Classification and Quality Evaluation Using Hyperspectral Imaging. Agriculture, 12(4), 530. https://doi.org/10.3390/agriculture12040530

[13] Wang, B., He, J., Zhang, S., & Li, L. (2021). Nondestructive prediction and visualization of total flavonoids content in Cerasus Humilis fruit during storage periods based on hyperspectral imaging technique. Journal of Food Process Engineering, 44(10), e13807. https://doi.org/10.1111/jfpe.13807

[14] Jiang, H., Jiang, X., Ru, Y., Chen, Q., Li, X., Xu, L., Zhou, H., & Shi, M. (2022). Rapid and non-destructive detection of natural mildew degree of postharvest Camellia oleifera fruit based on hyperspectral imaging. Infrared Physics & Technology, 123, 104169. https://doi.org/10.1016/j.infrared.2022.104169

[15] Jiang, H., Hu, Y., Jiang, X., & Zhou, H. (2022). Maturity Stage Discrimination of Camellia oleifera Fruit Using Visible and Near-Infrared Hyperspectral Imaging. Molecules, 27(19), 6318. https://doi.org/10.3390/molecules27196318

[16] Yang, L., Gao, H., Meng, L., Fu, X., Du, X., Wu, D., & Huang, L. (2021). Nondestructive measurement of pectin polysaccharides using hyperspectral imaging in mulberry fruit. Food Chemistry, 334, 127614. https://doi.org/10.1016/j.foodchem.2020.127614

[17] Wang, F., Zhao, C., Yang, H., Jiang, H., Li, L., & Yang, G. (2022). Non-destructive and in-site estimation of apple quality and maturity by hyperspectral imaging. Computers and Electronics in Agriculture, 195, 106843. https://doi.org/10.1016/j.compag.2022.106843

[18] Ndisya, J., Gitau, A., Mbuge, D., Arefi, A., Bădulescu, L., Pawelzik, E., Hensel, O., & Sturm, B. (2021). Vis-NIR Hyperspectral Imaging for Online Quality Evaluation during Food Processing: A Case Study of Hot Air Drying of Purple-Speckled Cocoyam (Colocasia esculenta (L.) Schott). Processes, 9(10), 1804. https://doi.org/10.3390/pr9101804

[19] Kang, R., Park, B., Ouyang, Q., & Ren, N. (2021). Rapid identification of foodborne bacteria with hyperspectral microscopic imaging and artificial intelligence classification algorithms. Food Control, 130, 108379. https://doi.org/10.1016/j.foodcont.2021.108379

[20] Aviara, N. A., Liberty, J. T., Olatunbosun, O. S., Shoyombo, H. A., & Oyeniyi, S. K. (2022). Potential application of hyperspectral imaging in food grain quality inspection, evaluation and control during bulk storage. Journal of Agriculture and Food Research, 8, 100288. https://doi.org/10.1016/j.jafr.2022.100288

[21] Park, B., Shin, T. S., Cho, J. S., Lim, J. H., & Park, K. J. (2021). Characterizing Hyperspectral Microscope Imagery for Classification of Blueberry Firmness with Deep Learning Methods. Agronomy, 12(1), 85. https://doi.org/10.3390/agronomy12010085

[22] Mohan, A., & Meenakshi Sundaram, V. (2020). V3O2: hybrid deep learning model for hyperspectral image classification using vanilla-3D and octave-2D convolution. Journal of Real-Time Image Processing, 18(5), 1681-1695. https://doi.org/10.1007/s11554-020-00966-z

[23] Saha, D., & Manickavasagan, A. (2021). Machine learning techniques for analysis of hyperspectral images to determine quality of food products: A review. Current Research in Food Science, 4, 28-44. https://doi.org/10.1016/j.crfs.2021.01.002

[24] Xuan, G., Gao, C., & Shao, Y. (2022). Spectral and image analysis of hyperspectral data for internal and external quality assessment of peach fruit. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 272, 121016. https://doi.org/10.1016/j.saa.2022.121016

[25] Sun, Y., Liu, B., Yu, X., Yu, A., Xue, Z., & Gao, K. (2022). Resolution reconstruction classification: fully octave convolution network with pyramid attention mechanism for hyperspectral image classification. International Journal of Remote Sensing, 43(6), 2076-2105. https://doi.org/10.1080/01431161.2022.2054299

[26] Jing, H., Wang, Y., Du, Z., & Zhang, F. (2022). Hyperspectral Image Classification with a Multiscale Fusion-Evolution Graph Convolutional Network Based on a Feature-Spatial Attention Mechanism. Remote Sensing, 14(11), 2653. https://doi.org/10.3390/rs14112653

[27] Mou, L., & Zhu, X. X. (2020). Learning to Pay Attention on Spectral Domain: A Spectral Attention Module-Based Convolutional Network for Hyperspectral Image Classification. IEEE Transactions on Geoscience and Remote Sensing, 58(1), 110-122. https://doi.org/10.1109/tgrs.2019.2933609

[28] Pu, C., Huang, H., & Yang, L. (2021). An attention-driven convolutional neural network-based multi-level spectral–spatial feature learning for hyperspectral image classification. Expert Systems with Applications, 185, 115663. https://doi.org/10.1016/j.eswa.2021.115663

[29] Liu, N., Zhang, B., Ma, Q., Zhu, Q., & Liu, X. (2021). Stack Attention-Pruning Aggregates Multiscale Graph Convolution Networks for Hyperspectral Remote Sensing Image Classification. IEEE Access, 9, 44974-44988. https://doi.org/10.1109/access.2021.3061489

[30] Pan, H. Z., Liu, M. Q., Ge, H. M., & Yuan, Q. (2022). Learnable three-dimensional Gabor convolutional network with global affinity attention for hyperspectral image classification. Chinese Physics B, 31(12), 120701. https://doi.org/10.1088/1674-1056/ac8cd7

[31] Dong, Y., Liu, Q., Du, B., & Zhang, L. (2022). Weighted Feature Fusion of Convolutional Neural Network and Graph Attention Network for Hyperspectral Image Classification. IEEE Transactions on Image Processing, 31, 1559-1572. https://doi.org/10.1109/tip.2022.3144017

[32] Zhang, T., Shi, C., Liao, D., & Wang, L. (2021). A Spectral Spatial Attention Fusion with Deformable Convolutional Residual Network for Hyperspectral Image Classification. Remote Sensing, 13(18), 3590. https://doi.org/10.3390/rs13183590

[33] Li, W., Chen, H., Liu, Q., Liu, H., Wang, Y., & Gui, G. (2022). Attention Mechanism and Depthwise Separable Convolution Aided 3DCNN for Hyperspectral Remote Sensing Image Classification. Remote Sensing, 14(9), 2215. https://doi.org/10.3390/rs14092215

[34] Shi, H., Cao, G., Zhang, Y., Ge, Z., Liu, Y., & Fu, P. (2022). H2A2Net: A Hybrid Convolution and Hybrid Resolution Network with Double Attention for Hyperspectral Image Classification. Remote Sensing, 14(17), 4235. https://doi.org/10.3390/rs14174235

[35] Huang, H., Luo, L., & Pu, C. (2022). Self-Supervised Convolutional Neural Network via Spectral Attention Module for Hyperspectral Image Classification. IEEE Geoscience and Remote Sensing Letters, 19, 1-5. https://doi.org/10.1109/lgrs.2022.3141870

[36] Ahmed, M. R., Reed, D. D., Young, J. M., Eshkabilov, S., Berg, E. P., & Sun, X. (2021). Beef Quality Grade Classification Based on Intramuscular Fat Content Using Hyperspectral Imaging Technology. Applied Sciences, 11(10), 4588. https://doi.org/10.3390/app11104588

[37] Jiang, G., Sun, Y., & Liu, B. (2021). A fully convolutional network with channel and spatial attention for hyperspectral image classification. Remote Sensing Letters, 12(12), 1238-1249. https://doi.org/10.1080/2150704x.2021.1978582