Check for updates

View Article Online
View Journal

Sustainable Food Technology

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: Md. F. Jubayer, M. Hasan, M. K. M. Rabby, Md. M. Hoque, Md. M. Rahman and M. A. R. Sarker, *Sustainable Food Technol.*, 2026, DOI: 10.1039/D6FB00010J.

This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainability spotlight

This study enhances sustainable food manufacturing by introducing a non-destructive, AI-driven framework for real-time starch gelatinization monitoring. Replacing traditional invasive sampling with an automated detection approach, it prevents sample loss and cuts down on raw material waste. Its high accuracy allows for more precise process control, optimizing energy consumption during heating and reducing batch failures. Ultimately, this affordable and accessible technology promotes smarter, resource-efficient manufacturing, aiding the industry's shift to automated, zero-waste processes.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

A novel, non-destructive approach for real-time detection of starch gelatinization using YOLO-based deep learning models

Md. Fahad Jubayer[a,b*], Mahmud Hasan[c], Md Khurram Monir Rabby[d], Md. Mozammel Hoque[e], Md. Masudur Rahman[f], Md. Abdur Rashid Sarker[a*]

[a]*Department of Agricultural Construction & Environmental Engineering, Sylhet Agricultural University, Sylhet 3100, Bangladesh.*

[b]*Department Food Engineering & Technology, Sylhet Agricultural University, Sylhet 3100, Bangladesh.*

[c]*Department of Farm Power & Machinery, Sylhet Agricultural University, Sylhet 3100, Bangladesh.*

[d]*Department of Electrical & Electronic Engineering, Bangladesh University of Engineering & Technology, Bangladesh.*

[e]*Department Food Engineering & Tea Technology, Shahjalal University of Science & Technology, Sylhet 3114, Bangladesh.*

[f]*Department of Pathology, Sylhet Agricultural University, Sylhet 3100, Bangladesh.*

Correspondence:

* fahadbau21@hotmail.com, jubayer.fet@sau.ac.bd (M.F. Jubayer)

** rashidsarker.acee@sau.ac.bd (M.A.R. Sarker)

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Abstract: This research proposes a novel, non-destructive, vision-based approach for real-time detection and confirmation of starch gelatinization using the state-of-the-art YOLO (You Only Look Once) deep learning models to overcome the limitations of traditional methods, which are manual and subjective for industrial control. The primary contribution of this work is twofold: it introduces a novel image-based framework for starch gelatinization detection and, under controlled laboratory conditions, demonstrates the potential of YOLO models as an accurate tool for automated, non-contact process monitoring. A custom dataset was developed by capturing temporal images of a heated potato starch solution, with each frame annotated as "Non-Gelatinized" or "Gelatinized". Four YOLO architectures (v8, v9, v11, v12) were trained and evaluated on the developed dataset. All models demonstrated exceptional and nearly identical performance, achieving maximum precision, recall, and F1-scores (1.0), alongside a high mean Average Precision (mAP@0.5 of 0.995). They also achieved perfect recall (100%) in localizing the reaction zone and converged to very low final losses (around 0.1). While all models excelled, YOLOv8 achieved the highest precision (99.97%) and fastest training time, whereas YOLOv12 showed superior initial learning stability over 20 epochs. The models were successfully validated on a continuous video stream, accurately identifying the gelatinization onset in real-time.

Keywords: Gelatinization detection, image processing, AI application in food, real-time monitoring, vision-based system, non-destructive approach.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

1. Introduction

Starch is a natural, renewable biopolymer that plays a crucial role in food and material sciences. It is the primary glycemic carbohydrate found in cereals, roots, and tubers, and it functions as a key energy source for humans and animals. Structurally, it is a complex polymer composed of α-D-glucose, consisting mainly of two polysaccharides: mostly linear amylose and highly branched amylopectin.[1] It can be extracted from a wide variety of sources, including cereals, seeds, fruits, pulps, tubers, and culms. Conventional sources include cassava, yams, and potatoes, while recent studies have identified novel starches derived from fruit wastes such as avocado, banana, jackfruit, mango, and pineapple.[2,3] Each starch source exhibits distinct molecular architecture and physicochemical characteristics, making starch highly versatile across applications ranging from food formulation to biodegradable materials.[3] In the food industry, starch is used as a thickener, stabilizer, binder, adhesive, and gelling or film-forming agent.[4] Beyond culinary functions, starch is also promising for biodegradable packaging films due to its plasticizing and gelatinization properties.[3] In packaging, starch offers biocompatibility and biodegradability, making it a sustainable alternative to petroleum plastics.[1]

During cooking and industrial processing, starch undergoes gelatinization — a fundamental physicochemical transformation in which native semi-crystalline granules convert into an amorphous form upon heating in water.[4] The gelatinization degree (DG) reflects the extent of this transformation and influences the texture, digestibility, and functional performance of starchy foods.[5] As the temperature rises, hydrogen bonds within the crystalline regions break, permitting water penetration and granule swelling, which ultimately disrupts the molecular order.[6] This transition alters the mechanical, thermal, and rheological properties of starch, directly impacting product texture, stability, and digestibility.[7] Because DG modulates nutritional and sensory characteristics, controlling it is essential to achieve consistent quality in processing and manufacturing.[8] Over- or under-gelatinization leads to undesirable product texture, process variability, and wastage. Consequently, developing rapid, accurate, and real-time techniques for monitoring gelatinization is a major focus of current research. Proper detection of starch gelatinization and gelation behaviors is crucial for manipulating the textural attributes of starch-based food and industrial products.[9]

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Various techniques like polarizing microscopy (PLM), small-angle X-ray scattering (SAXS), X-ray diffraction (XRD), differential scanning calorimetry (DSC), and the rapid visco analyzer (RVA) have been developed to study starch gelatinization.[7] DSC accurately measures temperature and enthalpy changes, but is less reliable for multi-component systems due to overlapping peaks. The amylose–iodine method is simple but inconsistent across sources, while enzymatic hydrolysis is accurate but slow.[5] Most methods are offline and need specialized equipment, which is rarely available in laboratory or industrial settings.[7] Microscopy with hot stage is a popular, simple method for evaluating gelatinization features by monitoring morphological changes in starch granules during heating.[6] Chen et al.[10] used the 'Gun Image Manipulation Program' to track starch granule diameter changes at different amylose contents and temperatures. Since birefringence indicates molecular order, its loss is a reliable gelatinization marker. Niu et al.[11] improved this by using an automated deep learning (DL) pipeline to quantify granule swelling and DG via birefringence. Wu et al.[6] proposed combining artificial neural networks (ANNs), computer vision, and fuzzy logic for real-time characterization. However, these methods are limited by reliance on batch sampling, manual supervision, or high-resolution optical setups, restricting industrial scalability.

All these methods usually involve endpoint or batch measurements, which require withdrawing samples, stopping reactions, or waiting for steady states. They are not ideal for in-line, real-time monitoring in a continuous production environment. Hence, the development of a feasible method of evaluating starch gelatinization using simple operations and equipment would greatly help in controlling the functionalities of starch-based foods or materials.[7] However, some recent studies took part in this challenge. Researchers have recently begun integrating image processing, spectroscopy, and machine learning (ML) for real-time detection of starch gelatinization. Zhong et al.[12] used a Mask R-CNN model on 884 microscopy images to classify granules into four gelatinization stages based on birefringence loss, achieving 96.5% accuracy, and correlating DG with DSC data in less than a second – advancing automated, real-time starch monitoring. Similarly, Chi et al.[7] showed that light transmittance at 620 nm, measured with a simple spectrophotometer, reliably tracks gelatinization, with increased transmittance indicating crystalline order loss. Gelatinization typically begins between 56–58°C, offering a low-cost alternative to complex tools. Techniques like time-resolved NMR and digital image analysis explore swelling kinetics and

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

molecular mobility.[13-15] Advanced sensors like focused beam reflectance measurement (FBRM) have been proposed for real-time particle monitoring.[16] However, these methods often need frequent calibration, are susceptible to fouling, and cannot capture spatial heterogeneity or microscopic changes.

Given the limitations of existing offline and sensor-based methods, there is a pressing need for non-invasive, real-time, and image-driven approaches that can be seamlessly integrated into food production lines. A reliable system should be capable of continuously observing gelatinization phenomena – such as granule swelling, birefringence loss, and translucency changes – without interrupting or damaging the sample.[17] Such a method would not only enhance process control but also serve as a diagnostic tool for assessing physicochemical and digestibility characteristics of starchy products.[5]

Birefringence and color change are two critical visual phenomena associated with starch gelatinization. The semi-crystalline granule structure exhibits birefringence, which diminishes as hydrogen bonds break and crystalline order collapses during heating. The disappearance of birefringence marks the onset of gelatinization.[4] Alongside this optical shift, color evolution provides another real-time visual cue of the progress of gelatinization. Sayar et al.[18] observed an opaque-yellow zone around gelatinized chickpea starch, linking the color change to gelatinization. Lamberts et al.[19] found a strong linear correlation ($r = 0.85$) between color difference (ΔE) and gelatinization in parboiled rice. Taghinezhad et al.[20] reported ΔE increased with gelatinization ($R^2 \approx 0.87$). Abhiram and Amarathunga[21] confirmed optical differences evolve with gelatinization, not after. These results show color change and optical shifts occur during gelatinization, as crystalline granules disrupt and amylose leaches, altering light scattering and absorption, shifting from white to translucent or yellow. Browning or pigment migration happen after complete gelatinization. Monitoring optical cues offers a feasible way for real-time, image-based detection.

In recent years, computer vision and DL have become powerful tools for industrial monitoring, quality control, and autonomous inspection. Particularly in food processing, imaging-based methods utilize visible or near-infrared changes (such as color, opacity, and structural features) to infer physico-chemical alterations.[22] Among DL models, object detection networks like the YOLO (You Only Look Once) family have gained popularity for their real-time inference and ability to

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

detect spatial features in scenes. YOLO models offer a single-pass prediction mechanism (bounding boxes + class scores), which contrasts with multi-stage detectors (e.g., R-CNN variants) and thereby reduces latency and computational overhead.[23] YOLO divides an image into a grid and predicts bounding boxes for each cell. It predicts class probability and confidence score for each box. It also improves accuracy using anchor boxes, predefined boxes of various sizes and aspect ratios, associated with each cell to predict object size and shape.[24]

Within the field of food and ingredient imaging, YOLO-based methods have been used to detect and classify food items, ingredients, and even portions in real time. Recent work employing YOLO algorithms in this area shows promise beyond typical object-detection tasks, including pest and disease detection, vehicle detection, fruit detection, medical imaging, fire and smoke detection, defect detection, and more. Examples include verifying the authenticity of food products,[25] digitizing and managing chemical laboratory operations by recognizing glassware and experimental actions,[26-27] detecting cancer at early stage,[28] enforcing food quality control through automated damage detection,[29] monitoring and classifying microalgae species for environmental protection against harmful algal blooms,[30] improving domestic safety with real-time sensing of cooktops and kitchen objects,[31] quantifying bubbles on electrode surfaces,[32] estimating calorie content of food on plates,[33] and detecting microbial growth on food surfaces.[34] These examples demonstrate YOLO's versatility in handling complex, domain-specific imagery that demands precision, real-time processing, and contextual understanding rather than just basic visual labeling. Although limited exploration of the algorithm's limitations, these studies collectively indicate a trend toward using YOLO as an intelligent perception core for automation, analytics, and scientific research, extending well beyond its original focus on conventional object detection.

The potential benefits of applying YOLO-based models to detect starch gelatinization are significant: a vision-based system could non-invasively monitor gelatinization in real time, enable closed-loop control, and decrease dependence on invasive sensor probes. In this work, we aim to develop and validate YOLO-based (YOLOv8, v9, v11, and v12) DL approaches for real-time detection of potato starch gelatinization under controlled heating conditions. Our approach connects process sensing and computer vision in an innovative way. The key contributions of this work are:

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

1. Building a novel dataset that includes the two stages – before gelatinization/non-gelatinized and gelatinized states of potato starch solution in laboratory settings.
2. A never-tried-before approach of using DL models to capture the phase change of a solution.
3. Reducing the reliance on human observation of the no-to-gelatinized transition of starch solutions during processing.
4. Developing and evaluating trained YOLO models to facilitate future research in the detection of starch gelatinization process in processing and manufacturing operations.

2. Materials and methods

2.1 Problem Formulation

Conventional methods for identifying the gelatinization point and tracking its progress, such as visual observation of bubble formation, viscometry, and DSC, are inherently manual, invasive, and subjective. These techniques have significant limitations for real-time process control because they depend on expert judgment, cause operational delays, and are not easily scalable for industrial automation. As a result, there is a crucial technological gap for a non-contact, rapid, and objective system capable of classifying the state transitions of starch solutions. This research aims to address the following problems:

Problem 1 (Dataset development): To develop a novel, curated image dataset capturing the temporal evolution of potato starch solution from a non-gelatinized to a gelatinized state under controlled laboratory conditions.

Problem 2 (Model development): To develop and adapt state-of-the-art YOLO models to reliably distinguish between the non-gelatinized and gelatinized states of starch from image data.

2.2 Proposed Framework

To address the above challenges, a comprehensive framework was proposed, as illustrated in Fig. 1. The process is explained in sequential phases:

2.2.1 Experimental Setup and Data Acquisition

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

A 5% (w/w) potato starch (Sigma-Aldrich Chemie GmbH, Germany) solution was prepared using deionized water in a 250 mL beaker. The beaker was subsequently placed on a magnetic hotplate stirrer (MS7-H550-S, OniLab, China), and the mixture was heated to 80°C for a duration of 30 minutes with constant stirring maintained at 400 rpm.[35] A light source was placed just behind the beaker to facilitate image capturing. This controlled environment, depicted in Fig. 1(a), ensures consistent and reproducible gelatinization kinetics.

2.2.2 Temporal Image Capturing

To document the gelatinization process, a high-resolution digital camera (Canon EOS M50, 15-45mm lens) was employed to capture images of the starch solution at fixed 5-second intervals from the initiation of heating. This procedure yielded a temporal image series chronicling the solution's visual evolution from an initial non-gelatinized state, through the gelatinization transition, to a final fully gelatinized state.

2.2.3 Dataset Construction and Annotation

The captured image sequence was transferred to a computing system for curation. A novel dataset was constructed by manually annotating each image frame with one of two class labels: Non-Gelatinized or Gelatinized. This annotated dataset serves as the fundamental ground truth for supervised model training.

2.2.4 Model Development and Comparative Analysis

The annotated dataset was partitioned and used to train four advanced YOLO object detection architectures: YOLOv8, v9, v11, and v12. The objective of this training phase, conceptualized in Fig. 1(b,c), was for each model to autonomously learn the discriminative visual features that characterize the two distinct states of the starch solution.

2.2.5 Performance Evaluation and Cross-Validation

The performance of the trained models was quantitatively compared using standard detection metrics, including precision, recall, and mean average precision (mAP). To assess model robustness and generalizability, the top-performing model was further validated on an independent, unseen dataset featuring corn starch gelatinization, captured via a smartphone-based

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

video system to emulate a practical monitoring scenario. The model's successful application in classifying a frame from this new data is shown in Fig. 1(d,e).

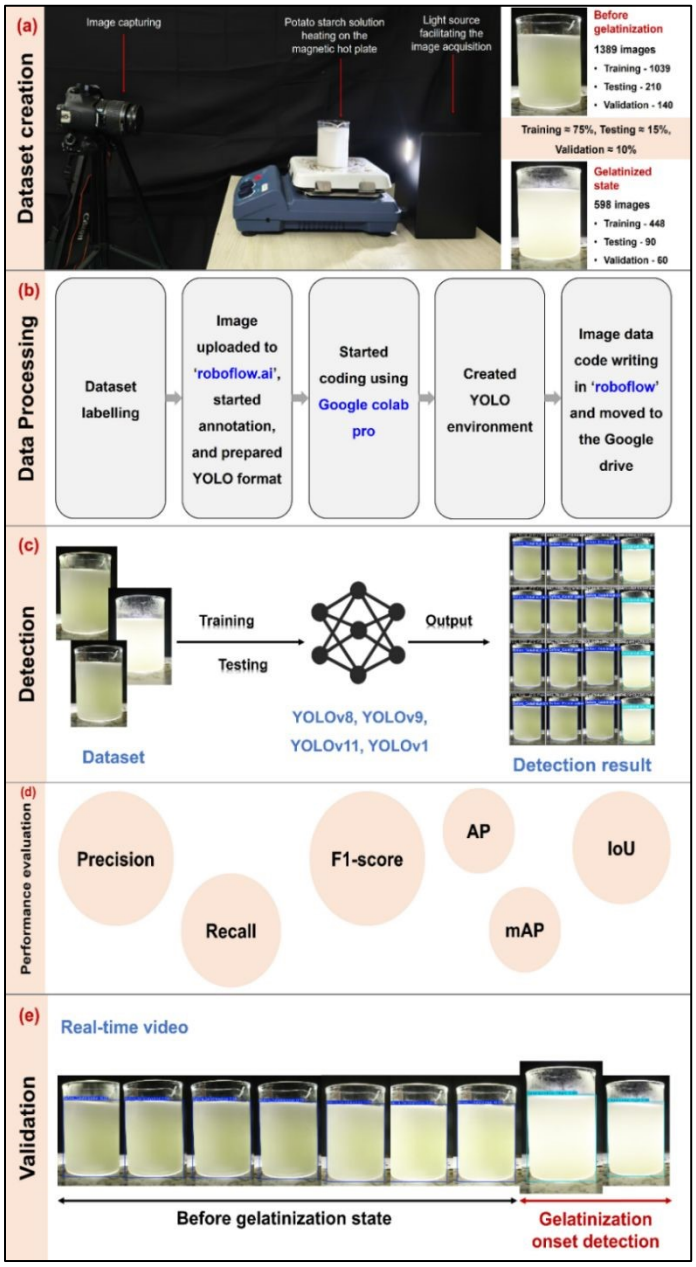


Fig. 1 Framework of the proposed study for the determination of starch gelatinization using the YOLO models. a) Laboratory environment for the creation of the dataset and the dataset details, b) data processing workflow, c) real-time detection using the dataset, d) performance evaluation of the used YOLO models, and e) process validation through real-time video monitoring employing the YOLO models.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

2.3 Image-Based Dataset Development

The research complied with all relevant regulations and protocols during the experiments. Initially, 10 g of potato starch powder was weighed and dissolved in 200 mL of deionized water (5% starch solution) in a 250 mL glass beaker. The 5% starch solution was placed on the magnetic hotplate, 400 rpm was set, and heating was started. From the first minute, the images were captured using a pre-arranged set-up. The resolution was 2656×3984 pixels (f/7.1, 1/125s, ISO 3200, 45mm). The camera was held steady on a tripod. The images were categorized into two classes: 1. From starting to gelatinization (before gelatinization), 2. Gelatinization onset state (Fig. 2). A total of 1987 images were collected to create a dataset for this study. Of these, 300 images (about 15%) were randomly selected to form a test dataset for model evaluation, and 200 images (about 10%) were selected for a validation set. The remaining 1487 (about 75%) images were used to prepare the training set. To prevent data leakage and ensure model generalizability, image acquisition timing was carefully controlled, and all frames were systematically organized into distinct folders by experimental run. The division into train, validation, and test sets (75%/15%/10%) was performed at the run level, ensuring that all frames from a given heating process were allocated to a single subset. This ensures that temporally adjacent or process-correlated frames never appear across different splits, preserving statistical independence and preventing performance inflation. Additionally, the training pipeline employed online data augmentation, including mosaic and mixup techniques, which synthesized new training samples by combining image fragments. This expanded the feature space for the 'Gelatinization_Start' class, ensuring that the final detection performance remained unbiased. Since the dataset images were not uniform, a preprocessing step was performed to normalize and resize all images to a consistent resolution of 640 × 640 pixels. For manual annotation, the Roboflow.ai platform was used to label all images in the training set. The Smart Polygon tool was employed to annotate the affected areas, and each region was assigned its respective class. After completion, Roboflow automatically generated and exported the annotated dataset in YOLO-compatible format, including the image files, class labels, and corresponding annotation text files. The dataset is accessible for viewing, and the complete information is available in the supplementary data section.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

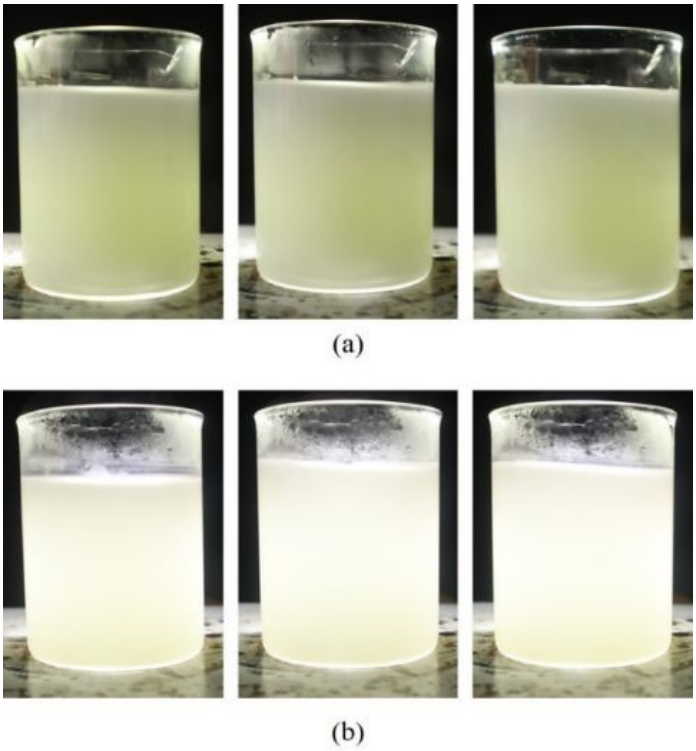


Fig. 2 5% starch solution in their two different stages, which are also categories of the study's dataset class: a) Before gelatinization and b) Gelatinization onset state.

2.4 Model Training

The YOLO models (YOLOv8, v9, v11, and v12) were trained using the Ultralytics YOLO framework in Google Colab. To ensure reproducibility, we fixed the random seed for training at seed = 0 for all models. Data augmentation was applied during training and included random scaling, cropping, flipping, and color jitter. The optimizer used for training was SGD by default, with an initial learning rate of 0.01 and weight decay set to 0.005. For inference, we set the confidence threshold to 0.25 and the Intersection over Union (IoU) threshold to 0.7 to filter out weak predictions and ensure higher localization accuracy. Each version was installed directly through the Ultralytics package, removing the need for manual cloning from GitHub as in older versions like YOLOv7. The installation automatically configured the necessary directory structure, including folders for datasets, model configurations, and weight files. After installation, model training was initiated using the yolo train command, which stores the training outputs (weights, logs, and results) in a new experiment folder at dataset/runs/detect/train5/weights/last.pt. The training used pre-trained weights supplied by Ultralytics, and parameters were configured based

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

on the dataset configuration file (data.yaml). The paths for training and validation data, the number of classes, and the class names were specified in this YAML file.

Training configuration details are as follows:

- Image size: 640 × 640
- Batch size: 32
- Epochs: 20
- Data configuration: data.yaml (contains paths, class count, and class names)
- Model configuration: selected from YOLOv8–YOLOv12 versions(yolov8s.pt, yolov9s.pt, yolo11s.pt, yolo12s.pt)

During the training process, data was collected, the loss was examined, and the model weights were saved at each epoch using the Tensorboard visualization tool. The following desktop computer specifications (Table 1) were utilized for training and testing with the PyTorch DL framework.

Table 1. Hardware and Software details used for the study

Processor	Intel Core i7-13620H 13 Gen
Operating system	Windows 11 64-bit operating system
RAM	16GB 2x DDR5 SO-DIMM 5600MHz
Programming language	Python 3.12.12
GPU	NVIDIA A100-SXM4-40GB, 40507MiB
Software	Ultralytics YOLO (8.3.214–v12), PyTorch 2.8.0 + CUDA 12.6 + cuDNN 9.6, OpenCV 4.5, Python 3.12.12, TensorBoard, Roboflow, Google Colab, and Visual Studio Code

2.5 Model Development

The starch gelatinization was modeled as a visual state-change detection task: subtle, time-dependent shifts in hue, saturation, and micro-texture within a designated beaker region. All

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

YOLO variants were employed in single-stage detection mode to (i) locate the reaction area (bounding box around the working fluid and meniscus) and (ii) determine its state ('pre-gelatinized' vs. 'gelatinized'). The selection of YOLO-based architectures over standard CNN or color thresholding was driven by three primary factors: spatial heterogeneity, robustness to environmental noise, and industrial scalability. While the transition of a starch solution involves a color shift, this reaction is often non-uniform (heterogeneous), with gelatinization occurring at different rates within the vessel due to localized heat distribution. Unlike a standard CNN, which classifies an entire image and can be misled by background pixels, YOLO architectures perform simultaneous localization and classification.[23] By dynamically 'bounding' the reaction zone, the models ignore irrelevant spatial noise, ensuring that the classification score is derived exclusively from the fluid itself. The following subsections describe the architectures and mechanisms of YOLOv8, v9, v11, and v12, focusing specifically on their features relevant to this precise visual recognition challenge.

2.5.1 YOLOv8

YOLOv8, developed by Ultralytics on 10 January 2023, represents a state-of-the-art single-stage object detection model optimized for real-time performance.[36] Its architecture (Fig. 3) comprises a CSPDarknet-53 backbone, a Path Aggregation Network–Feature Pyramid Network (PAN-FPN) neck, and a decoupled detection head.[37] This variant eliminates the need for pre-defined anchor boxes, simplifying the training pipeline and reducing model complexity. For the task of color-shift detection, where solution containers may be present at various scales and orientations, this anchor-free approach improves the model's flexibility to localize regions of interest without anchor-induced bias. The YOLOv8 model uses a new anchor alignment metric. The anchor alignment metric is computed by multiplying the class score by the IoU between the predicted and ground truth frames.[38] Let the input image $I \in \mathbb{R}^{H\times W\times 3}$ represent the captured solution. The backbone extracts multi-scale feature maps through successive convolutional and activation operations defined as:

$$F_l = \sigma\,(W_l * F_{l-1} + b_l), \quad l = 1,2,\ldots,L \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Where, * denotes convolution operation, W_l and b_l are layer weights and biases, and $\sigma(\cdot)$ represents the SiLU activation function. The CSPDarknet53 backbone partitions gradient flow for efficient

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

feature reuse, while the SPPF (Spatial Pyramid Pooling–Fast) module enhances receptive field aggregation.

Bounding box prediction follows an anchor-free formulation:

$$B\hat{} = (\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{w}}, \hat{\mathbf{h}}) = \text{Sigmoid}\,(\mathbf{P}) \odot \mathbf{S} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Where P is the raw output vector and S represents the stride scale.

The loss function, L_{v8}, is a composite of three components, guiding the model to localize and classify the gelatinization state accurately:

$$L_{v8}=\lambda_{box}L_{CIoU}+\lambda_{cls}L_{BCE}+\lambda_{dfl}L_{DFL} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Where L_{CIoU} is the complete IoU loss, which regresses the bounding box coordinates by considering overlap, center distance, and aspect ratio,[39] L_{BCE} is the Binary Cross-Entropy (BCE) loss for class probability distribution, and L_{DFL} is the Distribution Focal Loss, which focuses on learning a more accurate bounding box distribution by sharpening the probabilities around the target values.[40]

In color change detection, YOLOv8's high-resolution PAN-FPN improves sensitivity to small pixel-level differences, enabling it to detect subtle color transitions within the solution area. The C2f modules and PANet neck are highly effective at capturing the subtle color gradients and textures associated with starch gelatinization.[38] The decoupled head and high-resolution PAN/FPN features assist in distinguishing localization from class prediction: the regressor precisely fits the boundary of the beaker or fluid, while the classifier focuses on chroma or texture cues within that area. TAL decreases misclassification when color changes are subtle (borderline positives), enhancing early detection of slight hue variations.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

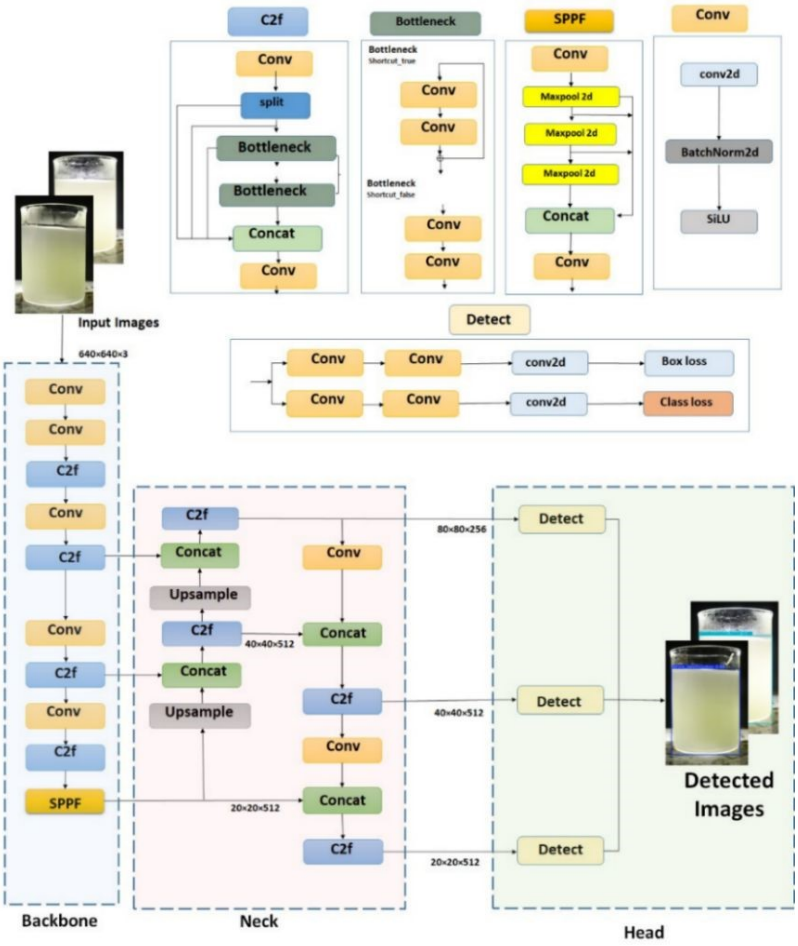


Fig. 3 YOLOv8 architecture for the detection of starch gelatinization.

2.5.2 YOLOv9

YOLOv9 introduces the Programmable Gradient Information (PGI) flow and Generalized Efficient Layer Aggregation Network (GELAN), offering adaptive gradient propagation and structural optimization (Fig. 4). The model enhances small-object detection using hybrid convolutional and transformer modules, which improve the representation of subtle hue gradients.[41] The main innovation of YOLOv9 is the PGI, which produces reliable gradients to update the weights of shallow layers, ensuring that features crucial for the final task are preserved from input to output. This is achieved through an auxiliary reversible branch that provides extra supervision. The GELAN backbone replaces YOLOv8's CSPDarknet with CSP-ELAN modules, enabling flexible computational blocks for optimized parameter utilization. It aggregates layers

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

via planned gradient paths, retaining complete input information across depths. For feature extraction, convolutional operations follow $F_{out} = \Sigma\ W_i * F_{in} + b_i$, with CSP splitting to reduce redundancy. The PAN-FPN neck facilitates multi-scale fusion, benefiting from GELAN's optimized inputs. The anchor-free decoupled head predicts bounding boxes and confidence scores using CIoU for regression and BCE/Focal loss for objectness and classification.[41]

The PGI mechanism ensures that the input information, such as the specific hue and saturation values indicating the onset of gelatinization, is programmable and can be propagated deeply without loss. This is formulated by introducing an auxiliary loss that guides the feature extraction process:

$$L_{v9} = L_{\text{main}} + \alpha L_{\text{aux}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

Where L_{main} is the standard YOLO detection loss (e.g., CIoU and classification loss) from the main branch, and L_{aux} is the auxiliary loss from the reversible branch that reinforces gradient flow.

In the context of starch gelatinization, the initial color change is a key visual indicator. YOLOv9's PGI is highly effective at transmitting the specific hue and saturation values that signal the onset of gelatinization throughout the network without sacrificing quality. This creates a model that more accurately captures the entire gelatinization process, improving detection accuracy, especially in complex cases with subtle or early color changes.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

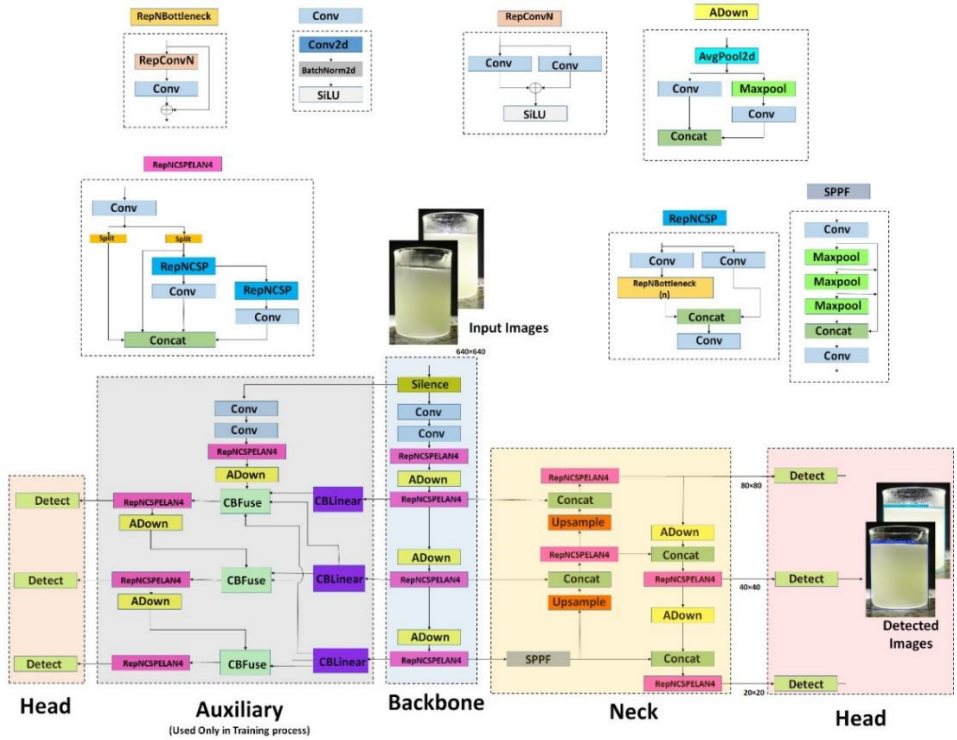


Fig. 9 YOLOv9 architecture for the detection of starch gelatinization.

2.5.3 YOLOv11

YOLOv11, the latest from Ultralytics, refines YOLOv9's design with enhanced backbone and neck architectures (Fig. 5). It enhances computational efficiency and feature abstraction using the Hybrid Transformer-CNN backbone and an Improved Bi-Directional Feature Pyramid Network (Bi-FPN). The model integrates Global Context Attention (GCA), enabling it to adaptively weight color and texture information – a vital factor for distinguishing subtle spectral shifts in chemical or biological solutions.[42] The backbone builds on GELAN principles but incorporates optimized modules for efficiency, such as refined CSP-ELAN variants, and emphasizes hierarchical feature maps via convolutions: $y = W * x + b$. The neck features improved PAN/FPN fusion, with adaptations for better multi-scale integration and reduced computational cost.[43] The attention mechanism can be represented as applying weights α_i to feature maps F_i:

$$F_{att} = \Sigma\, \alpha_i . F_i \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

Where α_i is computed to emphasize informative spatial regions and feature channels. This allows the model to focus on the specific area of the beaker where the reaction occurs and on the color channels most indicative of change. Head uses a refined anchor-free decoupled head, similar to YOLOv8, but with optimizations for faster computation and better feature alignment. The loss function incorporates a variant of focal loss to handle class imbalance and is defined as:

$$L_{v11} = \lambda_{\text{box}} L_{\text{CIoU}} + \lambda_{\text{cls}} L_{\text{focal}} + \lambda_{\text{dfl}} L_{\text{DFL}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

Here, L_{focal} addresses the potential imbalance between easy background samples and the more challenging foreground (gelatinized starch) samples, ensuring the model remains focused on hard examples.[44] For color change detection in starch solutions, the enhanced feature extraction bolsters sensitivity to subtle variations, with multi-scale capabilities ensuring robust identification of gelatinized areas amid fluid dynamics.[45]

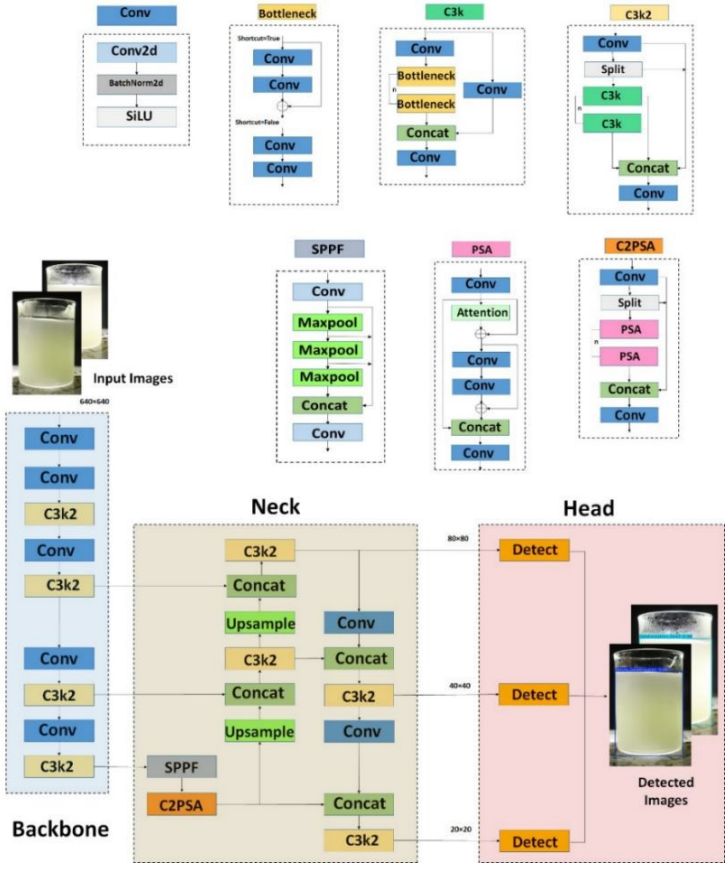


Fig. 5 YOLOv11 architecture for the detection of starch gelatinization.

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

2.5.4 YOLOv12

YOLOv12 represents the latest evolution, featuring Dynamic Tokenized Vision Transformers (DTVT) and an Adaptive Query Head (AQH) for end-to-end detection (Fig. 6). It unifies local-global feature learning, offering superior generalization for fine-grained visual cues like color intensity variation in real-time solution monitoring.[46] The R-ELAN backbone fuses layers with residual connections, using 7×7 separable convolutions for spatial context: $F_{out} = \Sigma\ W_i * F_{in} + b_i$, reducing parameters while capturing intricate details. The neck employs area attention with FlashAttention, segmenting features for focused refinement: Attention (Q,K,V) = softmax $(QK^T/\sqrt{d_k})V$, allowing the model to establish global relationships between pixels, and understand the context of the entire solution. The head employs a task-specific context module further to separate the feature spaces for classification and regression tasks.[47]

YOLOv12 uses attention for real-time object detection, bridging the gap between CNNs and attention-based models. Unlike earlier versions that relied mainly on CNNs, YOLOv12 adds attention without losing speed. This is achieved through three main architectural improvements: the A^2 Module, R-ELAN, and updates to the overall model structure, including Flash Attention and reduced computational overhead in the multi-layer perceptron (MLP).[42] A significant advancement in YOLOv12 is the use of a dynamic label assignment strategy which dynamically selects positive samples during training based on both classification and regression quality. This leads to better-aligned training targets. The overall loss function is a carefully balanced combination:

$$L_{v12} = \lambda_{box}L_{EIoU} + \lambda_{cls}L_{varifocal} + \lambda_{dfl}L_{DFL} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

Where L_{EIOU} is the Enhanced IoU loss that directly minimizes the disparity in width and height, and $L_{varifocal}$ is used to train a dense objectness predictor with an IOU-aware classification score. Unlike YOLOv11's general enhancements, YOLOv12's attention mechanisms and separable convolutions provide 25% parameter reduction with higher accuracy.[48] In detecting starch gelatinization, the area attention focuses on color-changing regions, improving robustness to occlusions and variations, while R-ELAN ensures precise feature capture for accurate real-time analysis. The hybridized structure allows YOLOv12 to detect minute color deviations even in

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

reflections or partially transparent liquids, offering real-time adaptability with minimal false positives.

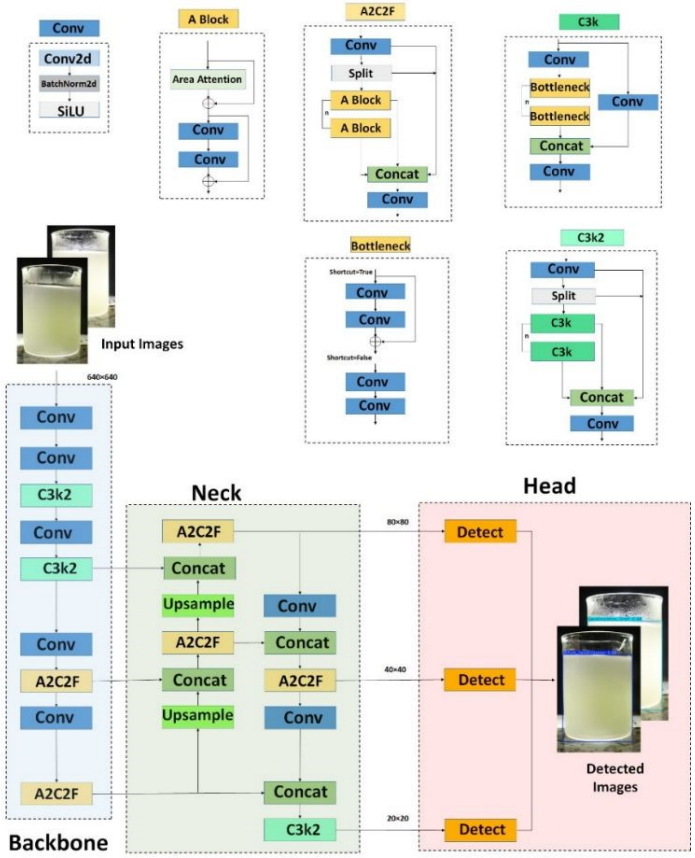


Fig. 6 YOLOv12 architecture for the detection of starch gelatinization.

2.6 Evaluation Indicators

The experimental findings described four distinct outcomes: true positive (TP) indicating the accurate detection of individually labeled starch solution; false positive (FP) denoting an object wrongly identified as the starch solution; true negative (TN) representing negative samples correctly predicted as such by the system; and false negative (FN) indicating the solution either missed or undetected.[25] For evaluating the performance of the model, several evaluation metrics are commonly used to gauge its accuracy and effectiveness, namely, precision, recall, accuracy, F1-score, and confusion matrix. Key statistical metrics are computed using the following equations.

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

$$Precision = TP/(TP + FP) \quad \ldots\ldots\ldots\ldots\ldots\ldots (8)$$

$$Recall = TP/(TP + FN) \quad \ldots\ldots\ldots\ldots\ldots\ldots (9)$$

$$AP = 1/11 * \sum r \in (0,0.1,0.2, \ldots..1)\ pinterp(r) \quad \ldots\ldots\ldots\ldots\ldots\ldots (10)$$

$$F1 - score = (2) / ((1/precision) + (1/recall)) \quad \ldots\ldots\ldots\ldots\ldots\ldots (11)$$

3. Experimental results

3.1 Image and Label Database

Fig. 7 shows the distribution and characteristics of the labeled data in the starch gelatinization dataset, which is essential for developing and evaluating YOLO models. Labeling was performed to annotate the ground-truth bounding boxes in the starch solution, which contained part of the beaker, from the images. Fig. 7a shows the total number of instances in the starch gelatinization dataset, which consists mainly of images of starch solutions in beakers. The dataset seems to have a fairly distributed set of object centers across the image, with no strong clustering observed (Fig. 7c). This indicates a balanced distribution of object locations, which is beneficial for the generalization of YOLO models. Fig. 7d displays the locations and sizes of labels within the dataset images, revealing an uneven distribution skewed toward the middle left corner, likely due to the limited dataset size. The dataset, with accurately and evenly distributed annotated bounding boxes (Fig. 7b), presents size statistics of the image borders. Anchor boxes of different sizes, created by a clustering algorithm based on ground truth boxes,[49] ensure that all initial anchor box sizes used by the YOLO algorithm match the part of the starch solution within the full image. The distribution of bounding boxes plays an essential role in how well YOLO models detect objects, as YOLO relies on anchor boxes and their placement relative to the true bounding boxes.[25] These ground truth boxes, surrounding the solution part in the beaker, help YOLO models learn to recognize similar items in new images. Notably, all bounding boxes in Fig. 7b are positioned just outside the image center, indicating a regular spread of the starch solution and consistent physical and optical properties. While the bounding boxes encompass the visible fluid region, the YOLO algorithms use these coordinates to perform spatial feature gating. Unlike a whole-image classifier that produces a single feature map for the entire frame, the YOLO architecture uses bounding box regression to isolate the region of interest (ROI).[36,38] This enables the model to learn localized

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

boundary coordinates that distinguish the meniscus and the fluid-glass interface from the surrounding hotplate and background. The analysis indicates that the dataset has a fairly uniform distribution regarding object sizes and locations, which should support YOLO's ability to detect objects effectively. However, the class imbalance (as shown in Fig. 7a) could impact detection performance for the minority class "Gelatinization_Start". The newer YOLO models (v8, v9, v11, v12) are expected to better handle the data distribution, especially with their improved features designed to enhance detection accuracy and efficiency.[50] In this study, the detection target was based on the color change in the same area throughout the process. That is why class imbalance did not affect the final detection. Another reason for the smaller dataset in the gelatinized state is that the time of image capturing was shorter than in the pre-gelatinization state.

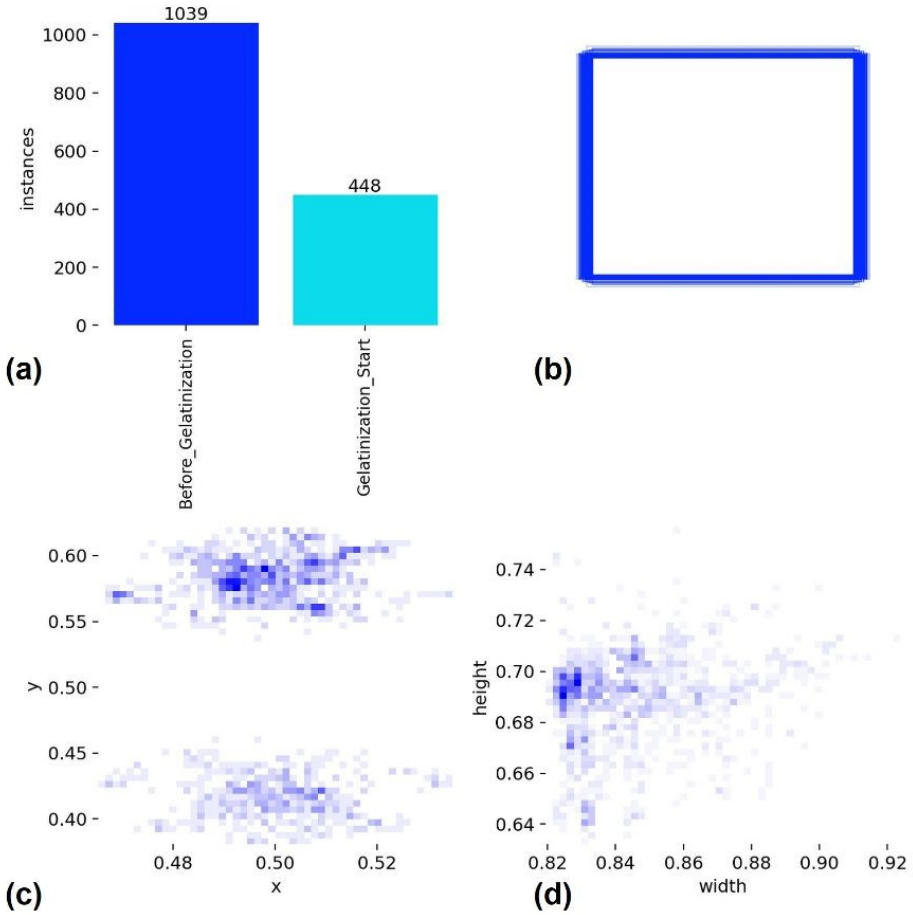


Fig. 7 Labels and label distribution of the YOLO models (YOLOv9, v10, v11, and v12). a) Number and class of labels in the starch gelatinization dataset, b) ground truth boxes, c) location distribution

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

of the dataset target center point, where the x-axis represents the horizontal position ratio and the y-axis represents the vertical position ratio, d) distribution of dataset target size.

3.2 Training Dynamics and Convergence

Fig. 8a, 8b, and 8c show that training is stable and effective for all models, with YOLOv9 exhibiting slightly higher initial losses but converging well. YOLOv8 displays the highest initial loss but drops sharply across epochs, indicating fast convergence. YOLOv12 has the lowest starting box loss, demonstrating the best performance at the outset (Fig. 8a). All models end with a low loss of around 0.1. YOLOv12 maintains a lower and more consistent box loss than the other models, confirming its robustness in generalizing to unseen data (Fig. 8c). All models start with high loss and converge very closely to a very low value, approximately 0.05 by epoch 16, indicating consistent classification learning. YOLOv12 performs reliably, with its box loss remaining lower than that of other models (Fig. 8e). The ability of YOLOv12 to start with a lower box loss and sustain it indicates better initial feature learning and quicker convergence. YOLOv8, although slower at first, catches up as the training advances.

The object detection loss measures how well the model detects objects across the dataset. Fig. 8b shows that YOLOv8 initially has the highest loss but improves steadily, while YOLOv12 maintains a significant advantage with a smoother loss decline. Validation loss is generally low, indicating good generalization. All models exhibit low, similar final losses around 0.1, confirming excellent localization on unseen data. YOLOv12 consistently maintains a lower object detection loss compared to the others, especially YOLOv8, which still shows higher fluctuations (Fig. 8d). The loss remains very low (< 5) and slightly noisy. All models achieve a very low final validation classification loss, suggesting minimal overfitting in class separation. The validation loss for bounding box refinement is low and similar across all models (around 1.0), confirming effective boundary learning. Some high spikes occur in the early epochs (2-8) (Fig. 8f). YOLOv12 consistently outperforms other versions on object detection loss, suggesting it learns object-specific features and generalizes them to unseen data. The fluctuations in YOLOv8's curve might suggest that its learning rate or model configuration could be further optimized to improve stability.[51]

All models reach maximum precision (1.0) very quickly (by epoch 5-8), confirming the near-zero false positive rate observed in the precision-confidence curves (Fig. 8g). YOLOv12 attains the

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

highest Intersection over Union (IoU) early in training and sustains it, while YOLOv8 begins with a lower IoU but improves significantly, though not as much as YOLOv12. All models achieve maximum recall 1.0 rapidly (by epoch 5-8), confirming their ability to identify all true positives, as shown in the recall-confidence curves (Fig. 8i). Fig. 8h illustrates a similar trend, with YOLOv12 consistently outperforming other models, especially YOLOv9 and YOLOv8. All models quickly reach maximum mAP@0.5 (1.0), which is the gold standard for object detection and confirms the previously shown (0.995) map. High accuracy is maintained even at strict IoU thresholds (Fig. 8j). YOLOv8 and YOLOv9 show the most fluctuation in the loss and precision metrics, suggesting that their configurations may need additional tuning for optimal performance.

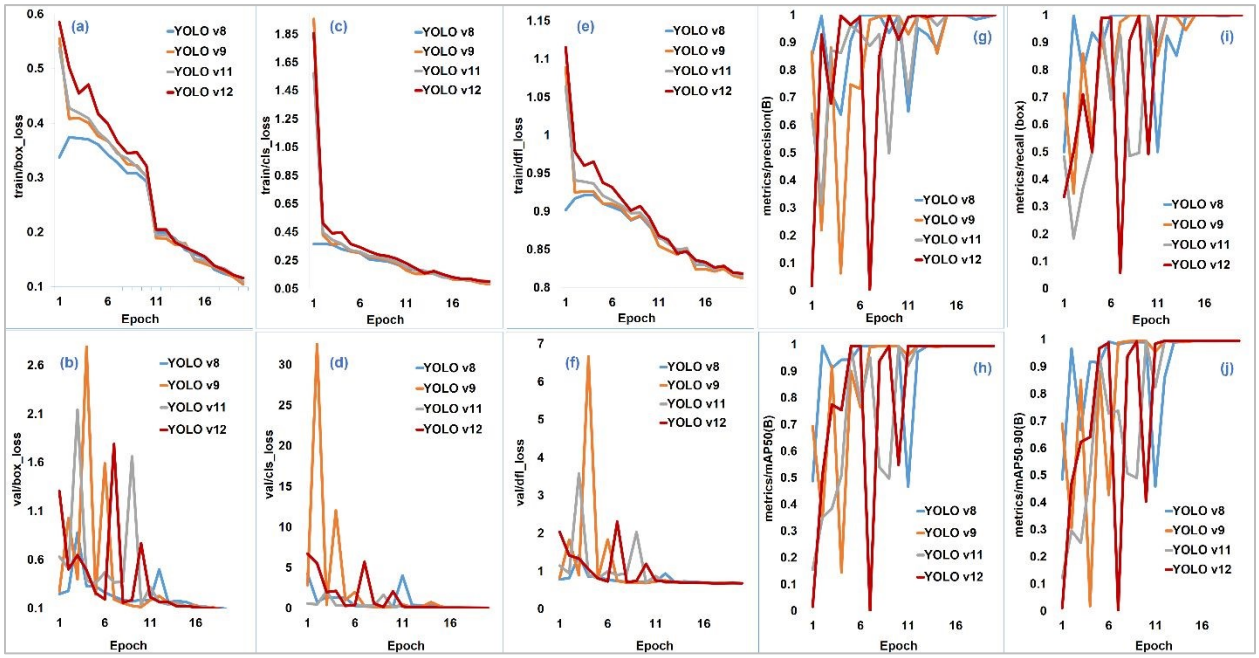


Fig. 8 (a-j) Performance analysis graphs and magnitudes of the YOLO models (YOLOv8, v9, v11, and v12) varying with 20 epochs during the training and validation of the dataset.

3.3 Measure of Performance

Supplementary curves were added to evaluate the model's performance, as shown in Fig. 9-12 and Fig. 13.

3.3.1 Precision-Recall Curves

The precision-recall curves illustrate the trade-off between precision and recall as the threshold varies. The models demonstrate high precision even at lower recall levels, indicating strong

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

specificity in object detection. All four YOLO versions achieve nearly 100% precision across almost the entire range of confidence and recall (Fig. 9). The very high precision indicates that the models produce almost no false alarms in this task. The precision-recall curves reach a recall of 1.0 while maintaining maximum precision, showing that the models successfully detected nearly all actual instances of "Before_Gelatinization" and "Gelatinization_Start." mAP@0.5 of 0.995 across all classes confirms the exceptional, near-perfect balance between precision and recall. YOLOv8 and YOLOv9 have almost identical curves, both displaying high precision even at very low recall values. YOLOv11 and YOLOv12 exhibit a slightly more gradual decline in precision as recall increases, indicating a minor trade-off between recall and precision compared to YOLOv8 and YOLOv9 (Fig. 9).

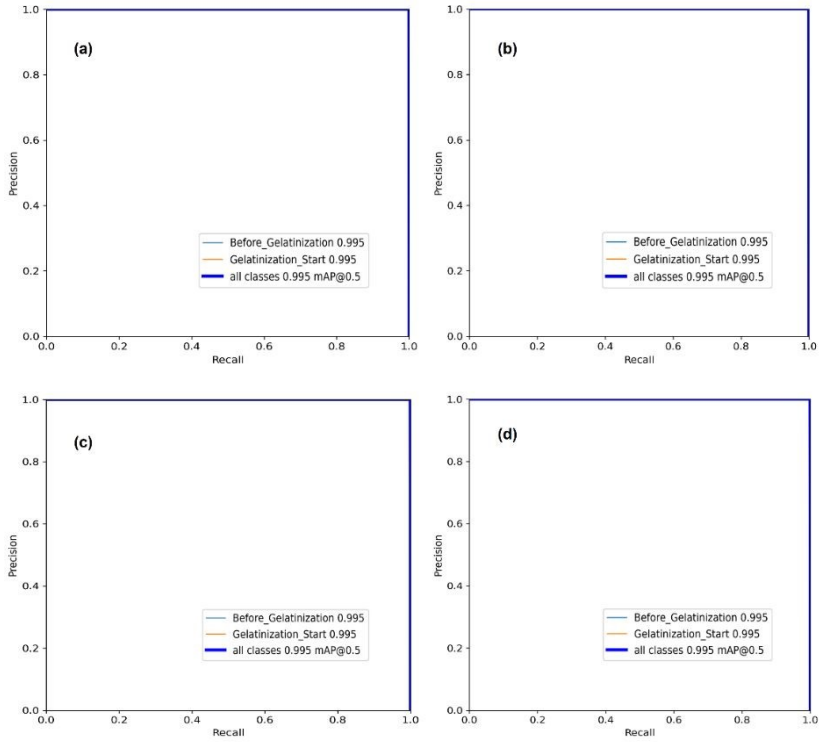


Fig. 9 Performance evaluation (precision-recall curves) of the YOLO models [(a) YOLOv8, (b) YOLOv9, (c) YOLOv11, and (d) YOLOv12] in the detection of starch gelatinization.

3.3.2 Precision-Confidence Curves

The precision-confidence curves provide insights into the models' ability to minimize false positives across different confidence levels. Precision improves as the confidence threshold

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript

increases, indicating fewer false positives are identified. Fig. 10 shows that all YOLO models display nearly perfect precision, approaching the 1.0 mark across the entire confidence spectrum. This signifies an extremely low false positive rate for the gelatinization classification task. The models are very dependable; even at lower confidence levels, their predictions are likely accurate. Their performance remains highly consistent and is not markedly affected by the specific Confidence Threshold set for deployment. Visually, the curves for YOLOv8, v9, v11, and v12 are almost indistinguishable, showing no clear advantage in precision for the newer versions on this particular dataset. YOLOv8 exhibits the highest precision at the highest confidence level, closely followed by YOLOv9, YOLOv11, and YOLOv12. This indicates that YOLOv8 is the most consistent at avoiding false positives at the highest confidence level. YOLOv12 shows slightly greater variability in precision as confidence increases compared to the other models, suggesting more inconsistent performance in avoiding false positives.

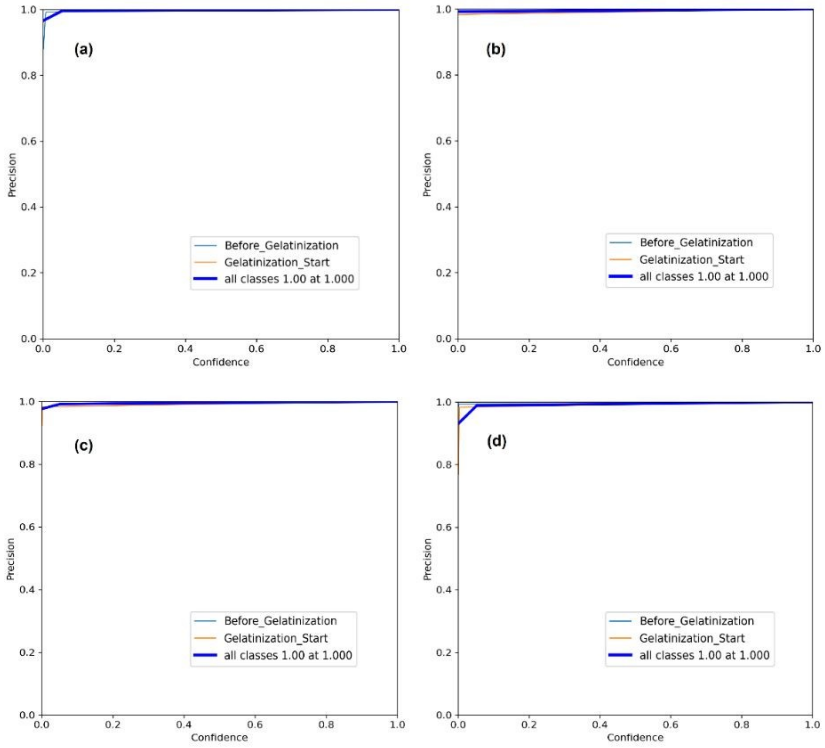


Fig. 10 Performance evaluation (precision-confidence curves) of the YOLO models [(a) YOLOv8, (b) YOLOv9, (c) YOLOv11, and (d) YOLOv12] in the detection of starch gelatinization.

3.3.3 F1-Confidence Curves

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

The F1-confidence curves (Fig. 11) illustrate the trade-off between precision and recall. The curves for all models are again quite similar, showing a sharp rise in F1-scores at low confidence thresholds, followed by stabilization as confidence increases. The F1-score reaches high levels very quickly but subsequently plateaus. The similarity of the curves across all models indicates that, once a certain confidence threshold is reached, the models effectively balance recall and precision. All four YOLO models achieve a maximum F1 score of 1.00. This confirms that they get a perfect balance of precision and recall at their optimal operating points. The F1 curve stays near F1 ~ 1.0 over a wide range of confidence thresholds (from close to 0.0 up to 0.9). This shows the models are highly robust and reliable. The point where the F1-score peaks at 1.00 occurs at a high confidence threshold for each model: (a) YOLOv8: optimal threshold is 0.913; (b) YOLOv9: optimal threshold is 0.921; (c) YOLOv11: optimal threshold is 0.916; (d) YOLOv12: optimal threshold is 0.916 (Fig. 11). All four successive YOLO versions display virtually no practical difference in their overall accuracy (F1-score) for this specific classification task. YOLOv9 appears to achieve the highest F1-score at the point where the curve stabilizes, indicating a slightly better balance between precision and recall than the other models. YOLOv11 and YOLOv12 show marginally lower but still competitive F1-scores, meaning they maintain a good balance of precision and recall but are somewhat less effective than YOLOv9 in the mid to high confidence ranges.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

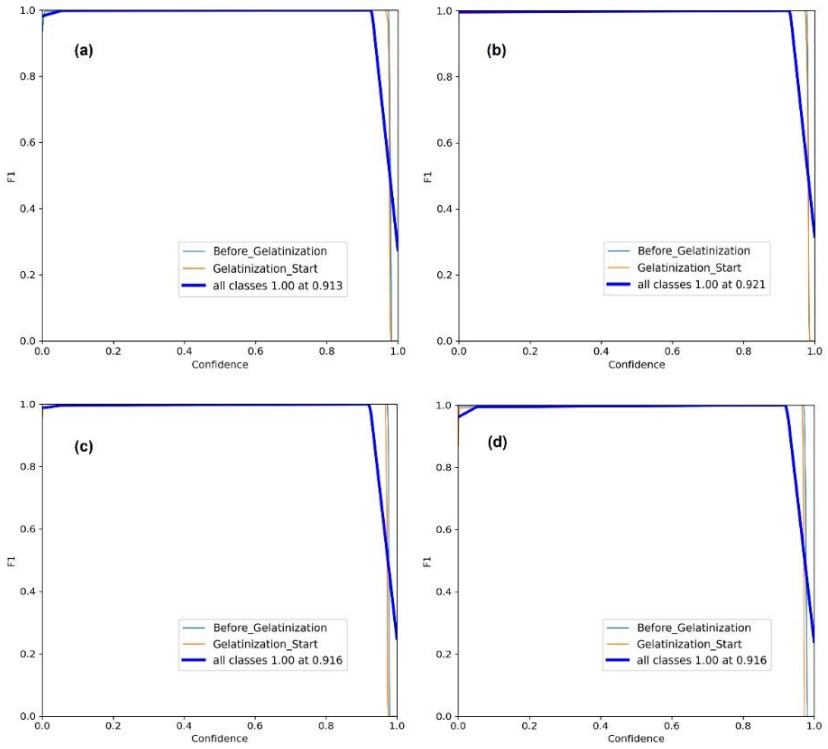


Fig. 11 Performance evaluation (F1-confidence curves) of the YOLO models [(a) YOLOv8, (b) YOLOv9, (c) YOLOv11, and (d) YOLOv12] in the detection of starch gelatinization.

3.3.4 Recall-Confidence Curves

The recall-confidence curves (Fig. 12) illustrate the recall values (sensitivity) at different confidence thresholds. The curves for all the YOLO models are quite steep, indicating that the models are very confident in their predictions as the confidence level rises. A high recall indicates that the models detect most of the relevant objects. Across all models, the recall stays high at low confidence thresholds but drops sharply as confidence increases. This suggests that the models initially identify many objects, but as the confidence requirement grows, only the most confidently detected objects are retained. All models maintain 100% recall when the confidence threshold is set at 0.00. This means the models successfully detect all true instances of "Before_Gelatinization" and "Gelatinization_Start" when considering all predictions. Recall drops sharply only at very high thresholds, which is expected because the model begins filtering out slightly uncertain but correct predictions. Both YOLOv8 and v9 exhibit a similar trend with relatively high recall values, indicating strong performance in detecting starch gelatinization before and after it begins. The recall curves for YOLOv11 and v12 also follow the same pattern but appear to decline slightly

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

more slowly, suggesting better overall sensitivity in detecting starch gelatinization events across various confidence thresholds.

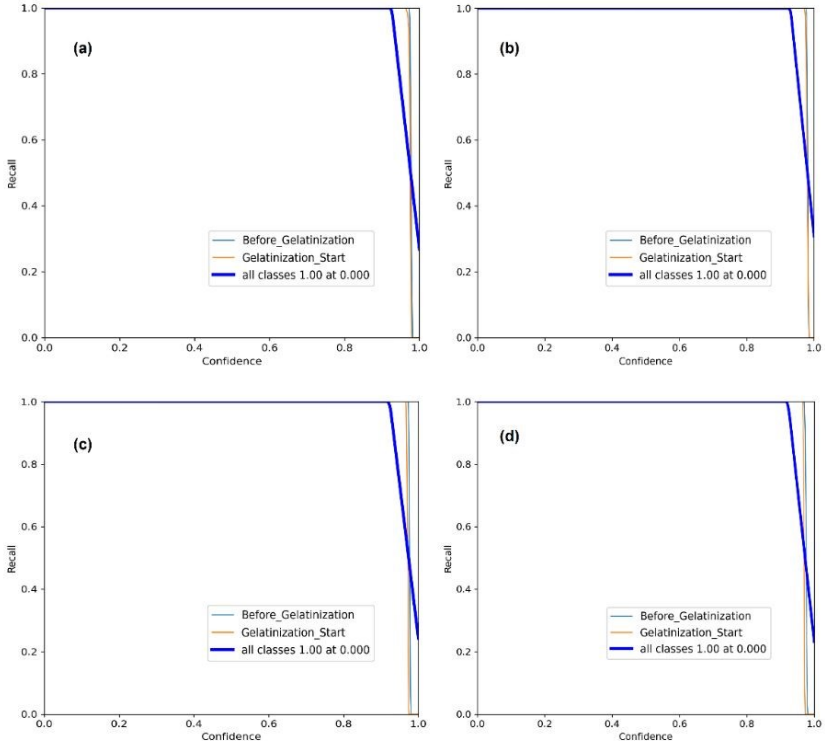


Fig. 12 Performance evaluation (recall-confidence curves) of the YOLO models [(a) YOLOv8, (b) YOLOv9, (c) YOLOv11, and (d) YOLOv12] in the detection of starch gelatinization.

3.4 Training Time

As the epoch count rises (0-20), the training time for all models increases gradually, which is expected since training involves multiple passes through the data. However, there are differences in how each model scales with increasing epochs. YOLOv9 consistently requires the longest training time per epoch, taking up to 312 seconds by epoch 20. YOLOv8, YOLOv11, and YOLOv12 generally have significantly faster training times than YOLOv9, requiring 208, 216, and 245 seconds respectively by epoch 20 (Fig. 13). YOLOv8 has the shortest training time at 207.52 seconds, showing that it is the most efficient in terms of training duration. YOLOv9 takes the longest at 312.44 seconds, which is significantly higher than the others, suggesting that the increased complexity in YOLOv9 might lead to longer training times. YOLOv11 and YOLOv12

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

take 215.03 and 244.10 seconds, respectively, placing them between YOLOv8 and YOLOv9 in efficiency (Fig. 13).

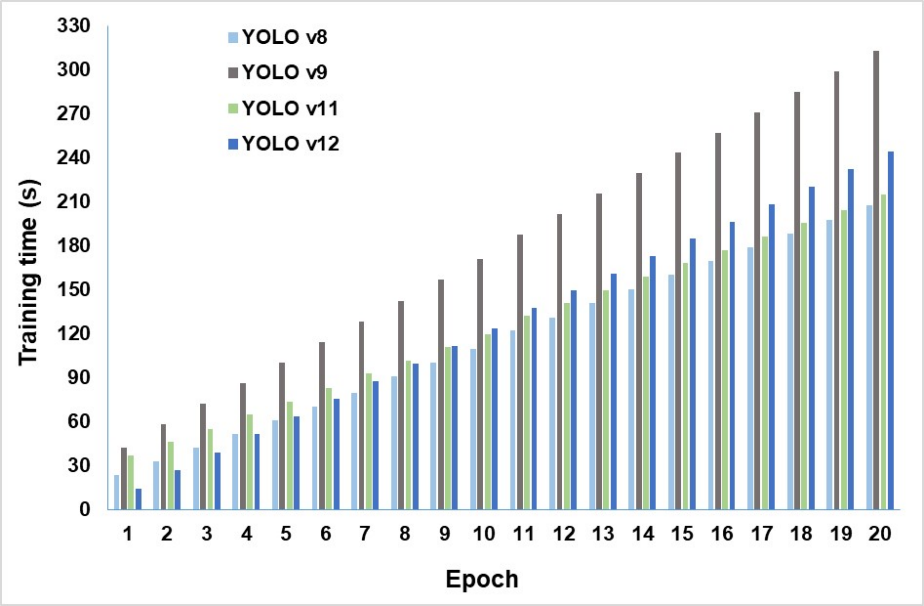


Fig. 13 Comparison of training times for the YOLO models (YOLOv8, v9, v11, and v12) in the detection of starch gelatinization state with 20 epoch.

3.5 Confusion Matrix

The illustration in Fig. 14 shows a normalized confusion matrix, comparing actual classifications with predicted ones. It highlights cases where the model finds it difficult to classify or distinguish between classes accurately. The YOLO models made zero classification errors between the two main classes; all instances of gelatinization stages were correctly labelled. This matrix is a 2 × 2 representation, with one axis displaying actual (true) values and the other showing predicted values by the model. Ideally, a diagonal line of 1.00 would run from the upper left to the lower right, indicating perfect predictions. For instance, in the analysis of gelatinization starting point (Fig. 14), 100% of the time, the gelatinization starting point was correctly identified. The correct classification percentages for the starch gelatinization detection task according to the model are as follows:

- Before gelatinization state of the potato starch solution: 100%
- Gelatinization starting/gelatinized state of the potato starch solution: 100%

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

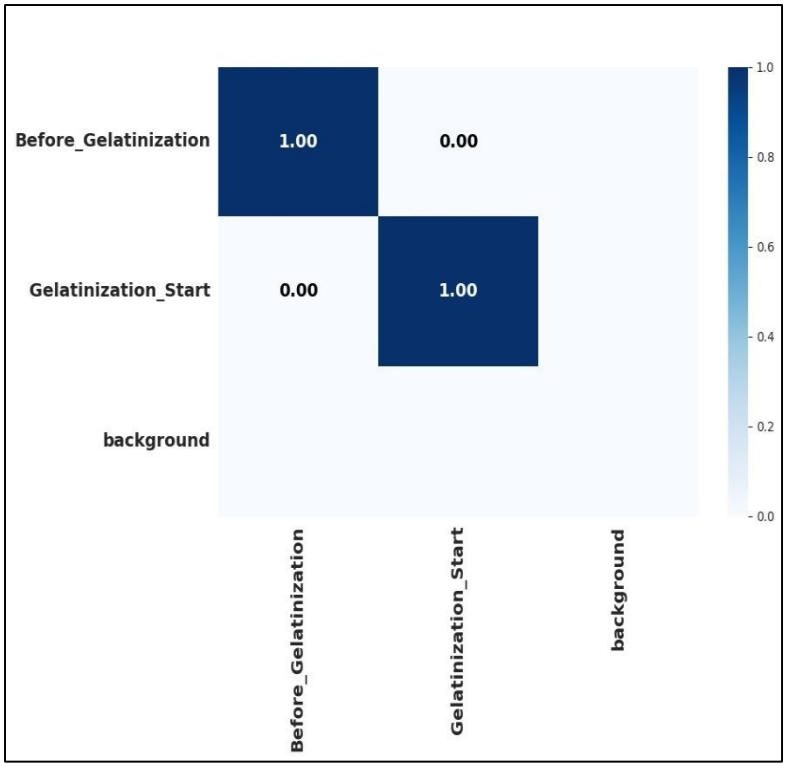


Fig. 14 Confusion matrix normalized for all YOLO models (YOLOv8, v9, v11, and v12).

3.6 Model Comparison (Box Prediction)

The YOLO model accurately identifies bounding boxes for the target objects in the starch gelatinization detection task (Table 2). The models demonstrate very high precision, with YOLOv8 slightly leading at 99.97%. Although small, these differences in precision suggest that YOLOv8 has a slight advantage in predicting correct boxes compared to the others. The performance remains remarkably consistent across all versions, indicating the stability of the YOLO architecture in this specific object detection task. Recall measures the model's ability to identify all relevant boxes, indicating how well it detects every object in the image. All four models achieve a perfect recall rate of 100%, correctly identifying every relevant box. This shows that the models are highly effective at detecting starch gelatinization features without overlooking any. The slight differences in F1-score between YOLOv9, YOLOv11, and YOLOv12 indicate that all three versions perform almost identically, with YOLOv12 experiencing a very small drop (99.96%). This highlights that the models maintain a strong balance between detecting relevant objects and minimizing false positives, with slight performance variation. The mAP metric indicates that there is no significant difference in performance across object categories. While all

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

models show consistent results in terms of precision, recall, and F1-score, they also achieve the same mAP score, further supporting the idea that the architectures are similarly optimized for this specific task. The consistent performance of all YOLO models across evaluation indicators suggests that the recent model versions (YOLOv9, YOLOv11, and YOLOv12) may offer enhanced capabilities for other, more complex tasks. Still, for this specific application, the improvements in newer versions are marginal.

Table 2. Comparison of evaluation indicators for box prediction features among the YOLO models used for starch gelatinization detection after 20 epochs and 47 iterations.

Evaluation indicators	YOLO models employed			
	YOLOv8	YOLOv9	YOLOv11	YOLOv12
Precision (%)	99.97	99.95	99.95	99.93
Recall (%)	100	100	100	100
F1-score (%)	99.98	99.97	99.97	99.96
mAP (%)	99.50	99.50	99.50	99.50
Training time (s)	207.52	312.44	215.03	244.10

4. Real-time application: visualization and validation

4.1 Visualization and Discussion

The visualization outcome of the starch gelatinization onset/state detection using the YOLO algorithms (YOLOv8, v9, v11, and v12) are shown in Fig. 15. The bounding boxes around the starch solution are usually used to detect the object that was at the non-gelatinized and gelatinized stages throughout the entire process. Each bounding box includes a probability value representing the likelihood of a specific item class within that region.[52] In this study, the bounding boxes effectively detect various stages of the starch gelatinization process. In this study, four versions of YOLO models are used, namely YOLOv8, YOLOv9, YOLOv11, and YOLOv12.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

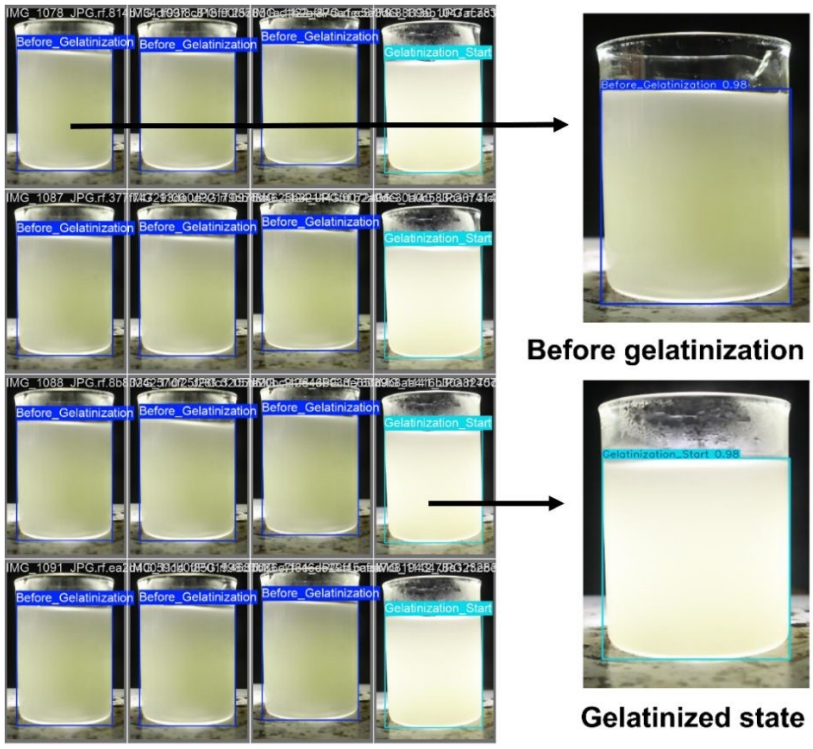


Fig. 15 The detection of the non-gelatinized and gelatinized states of the starch solution using the YOLO models. 'Before gelatinization' shows images with bounding boxes when the starch solution is heating on a magnetic hot plate in its normal state. 'Gelatinized state' shows images with bounding boxes when the starch solution begins to gelatinize.

This study successfully developed and validated a novel, image-based approach for real-time detection of starch gelatinization onset using four versions of the YOLO DL models (YOLOv8, v9, v11, and v12). The primary objective was to assess the feasibility of using these object detection models to identify visual changes indicative of the onset of gelatinization in a heated potato starch solution. Our key findings show that all YOLO models achieved outstanding performance, with nearly perfect precision, recall, F1-scores, and a mAP@0.5 of 0.995. The models accurately classified the "Before_Gelatinization" and "Gelatinization_Start" states with 100% accuracy on the test set, as confirmed by the confusion matrices. While all models performed excellently, YOLOv8 had a slight edge in precision (99.97%), whereas YOLOv12 showed the most stable and efficient learning curves with lower initial losses.

The impressive performance of all YOLO variants can be credited to their ability to learn and generalize subtle visual cues, specifically, shifts in hue, saturation, and texture, which signal the

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

onset of gelatinization. The breakdown of the semi-crystalline granule structure and the subsequent leaching of amylose cause changes in light scattering and absorption, leading to a visible shift from an opaque, white suspension to a more translucent, yellowish gel.[4] The YOLO models, especially with their advanced feature extraction backbones and multi-scale fusion necks (e.g., PAN-FPN, GELAN), proved highly sensitive to these pixel-level changes.[53]

The slight performance differences among the models, though minor in this case, reveal their architectural details. YOLOv8's top-tier accuracy indicates that its anchor-free, decoupled head and streamlined CSPDarknet backbone are very well-suited for this specific, limited detection task, where the object (the beaker's reaction zone) is always in the same spot. The better learning stability and lower initial loss of YOLOv12 can be seen as advantages of its hybrid transformer-CNN architecture and dynamic label assignment, which probably helped it learn more robust features from the start.[46] The fact that the more complex models (v9, v11, v12) didn't significantly outperform YOLOv8 suggests the visual task, while important, may not be complicated enough to require advanced features like programmable gradient information (PGI) or global context attention for a significant performance boost. Another possibility is that the high-quality, pre-processed dataset and the clear visual difference between the two states allowed the models to learn the task almost perfectly, leaving little room for measurable improvement. However, training YOLO models on such a dataset may simplify more complex laboratory processes.

Despite the near-perfect precision, recall, and F1-score values obtained in this study, these results should be interpreted as proof-of-concept performance under controlled laboratory conditions rather than definitive evidence of universal generalizability. The setup used a fixed camera, consistent beaker placement, controlled lighting, and a stable background. As a result, models may have learned not only visual cues of gelatinization, such as changes in opacity, hue, saturation, and texture, but also context-specific features, including background contrast, hotplate shape, reflections, meniscus, shadows, and lighting. This is especially true since the object location was consistent and the ROI was fixed. Therefore, high metrics may reflect overfitting to the laboratory environment rather than generalizable features. However, several measures were implemented to assess its generalizability. First, we carefully tracked the training and validation loss curves throughout the optimization process; both curves showed steady declines and converged without signs of divergence, indicating no evidence of overfitting during training. Second, we used the

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

trained YOLO model for real-time inference on live video feeds of previously unseen heating processes, where it consistently demonstrated accurate and stable detection in the same controlled environment.

Our findings strongly align with and expand on the emerging trend of integrating computer vision and deep learning into food science analytics. Previous studies have established the foundation by using deep learning for detailed analysis, such as Zhong et al.,[12] who used Mask R-CNN to segment and classify individual starch granules based on birefringence loss. Our study differs by operating at a larger scale, showing that complex physicochemical transitions can be monitored without high-resolution microscopy, using simple camera setups to observe bulk solution properties. This method shows strong potential for scalability to industrial applications, but further validation across diverse starch sources, process conditions, and environments is needed before broader industrial deployment.

Furthermore, our results support the findings of Chi et al.[7] who identified a link between optical properties (light transmittance) and gelatinization. We expand on this by moving beyond spectrophotometry to a spatial-vision-based method that can detect heterogeneity within the sample. The successful use of YOLO models also aligns with their growing application in complex food quality control tasks, such as real-time sensing of kitchen objects[31] and automated damage detection in food products.[29] Our work demonstrates the versatility of these models, broadening their use from traditional object detection to the more detailed area of real-time process-state monitoring. However, our study differs from some sensor-based approaches[16] that require physical probes. While those methods provide specific particle-size data, our vision-based approach is completely non-invasive, reducing the risk of sensor fouling and process interruptions. This makes our method a useful alternative or complement for environments where non-invasiveness and spatial information are important. Additionally, the YOLO models' performance in this study aligns with recent advances in object detection for specialized tasks. YOLO can classify complex hyperspectral data, highlighting its ability to extract features from subtle visual cues.[54] Integrating these models into real-time monitoring supports Industry 4.0. Automation in assembly lines reduces human error and improves consistency.[55] Applying this to starch gelatinization, our system replaces subjective manual observation with an automated, objective monitoring tool, bridging traditional food science and modern manufacturing.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

4.2 Validation

To move from image-based analysis to an actual real-time application, the trained YOLO models were tested using continuous video footage of the gelatinization process. This crucial step evaluated the model's robustness and its suitability for deployment in dynamic, real-world situations. For validation, instead of capturing still images at intervals, a continuous video stream was recorded throughout the entire process using the same camera, from the initial suspension of starch powder in deionized water to complete gelatinization. This video, recorded at 30 frames per second (fps), served as the unseen, dynamic data source for validation. The pre-trained YOLO models (YOLOv8, v9, v11, and v12) were then used to process this video stream. The video was fed into the model frame by frame, simulating a real-time monitoring setup. For each frame, the model performed inference, generating bounding boxes around the reaction zone and classifying its state as either "Before_Gelatinization" or "Gelatinization_Start." The timestamp of the frame where the model's classification first consistently switched to "Gelatinization_Start" was recorded as the detected gelatinization onset point (Fig. 16).

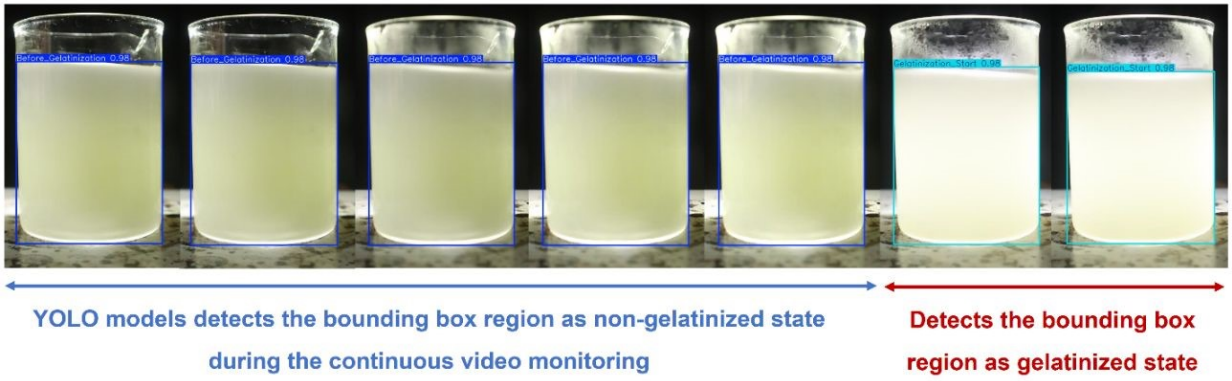


Fig. 16 The validation process for the YOLO models in detecting the starch gelatinization point during continuous video monitoring.

The results of this video validation were highly successful. All four YOLO models consistently and accurately identified the onset of gelatinization in the video stream. The transition in classification occurred within a narrow window of a few seconds, visually aligning with the point at which the solution began to show increased translucency and viscosity. No flickering was observed between classes before or after the transition, indicating strong model confidence and stability in its predictions. This shows that the features learned from the static image dataset

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

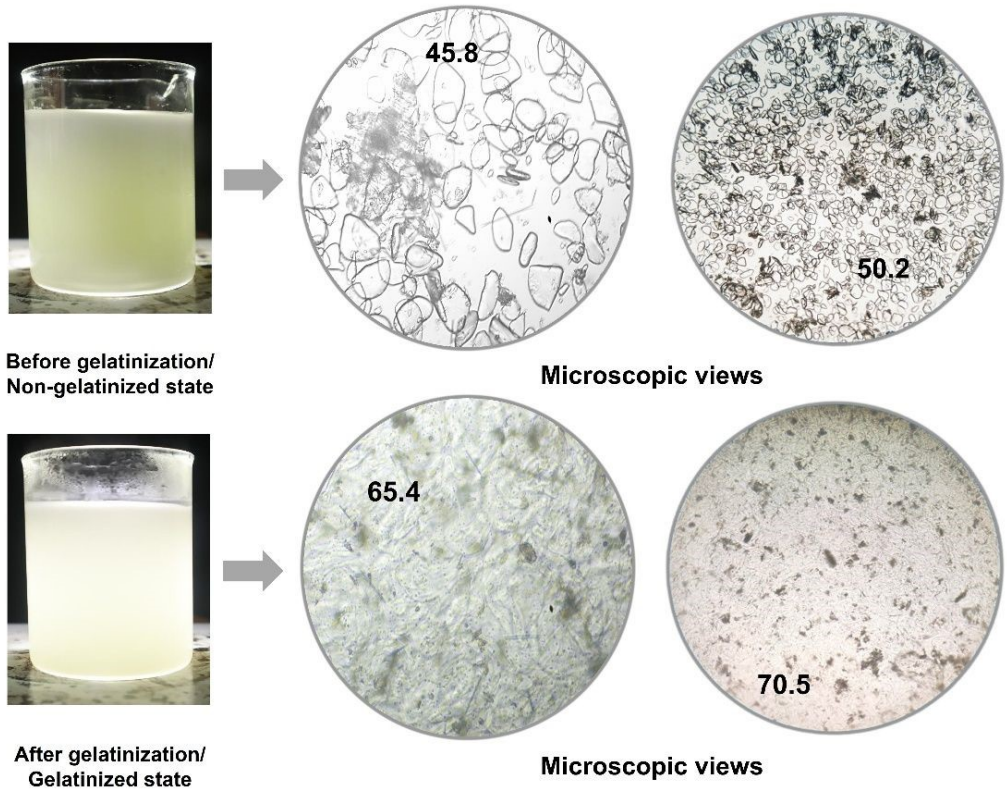
effectively generalized to the temporal domain of a video, capturing the key visual event without being misled by minor frame-to-frame variations or motion artifacts. This video-based validation overcomes a key limitation of many offline methods and confirms the practical viability of our approach. It demonstrates that the system can analyze not only pre-selected still images but also operate on a continuous feed, which is essential for in-line process control.

In addition, an additional validation study was conducted to confirm that the model detects true physicochemical gelatinization rather than a subjective visual transition (Fig. 17). The heated starch solution was examined by taking a drop from it under the microscope (B-159, Optika, Italy) at 10x magnification four times, two from when the bounding box was presented 'Before_Gelatinization' indication and two when the bounding box was presented with 'Gelatinization_Start' indication during the video validation phase. The microscopic images (Fig. 17) show clear structural differences. In the non-gelatinized sample, starch granules remain discrete, intact, and well defined.[56] The granules appear separated from one another, indicating that the native granular organization is largely preserved. In contrast, the gelatinized sample exhibits extensive swelling, distortion, and loss of individual granule identity.[54] The micrograph is dominated by a continuous amorphous matrix with blurred boundaries and fused granular remnants, indicating water absorption, granule rupture, and leaching of starch polymers. These changes are characteristic of starch gelatinization, during which the native semicrystalline granule structure is disrupted as found in several studies.[56-58]

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.



781

782 **Fig. 17** Microscopic images of starch solution in the non-gelatinized and gelatinized states. The
783 numbers inside the microscopic images represent the corresponding heating temperature (°C).

784 5. Limitations

785 While the results are highly promising, several limitations must be acknowledged. First, the study
786 was conducted under highly controlled laboratory conditions with a single starch type (potato) and
787 a fixed concentration (5%). The performance of the models on other starch sources (e.g., corn,
788 wheat) with different gelatinization kinetics and visual characteristics, or under varying
789 concentrations and stirring rates, remains to be fully validated. Second, the "ground truth" for
790 gelatinization onset was based on a standardized heating protocol rather than simultaneous
791 validation with a reference method like DSC for each experiment. Although the protocol is well-
792 established,[35] this introduces a potential source of systematic error. Third, the dataset, while
793 sufficient for this proof-of-concept, was relatively small and exhibited a class imbalance. Although
794 the models handled this expertly, a larger and more balanced dataset encompassing a wider range
795 of environmental and compositional variability would enhance model robustness and

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

generalizability. Finally, the study used a fixed imaging setup, including a fixed camera position, consistent beaker placement, controlled illumination, and a relatively uniform background. Although this setup enabled reproducible image acquisition, it also increased the likelihood that the models learned laboratory-specific visual features alongside gelatinization-related physicochemical cues. Therefore, the near-perfect precision, recall, and F1-score values should not be interpreted as a complete elimination of overfitting risk. Instead, they indicate strong performance within the experimental domain tested in this study.

6. Conclusions

This research successfully developed and validated a novel, vision-based framework for real-time detection of starch gelatinization using state-of-the-art YOLO deep learning models. The key finding is that all evaluated YOLO architectures (YOLOv8, v9, v11, and v12) exhibited exceptional and almost identical performance in this task, achieving near-perfect precision, recall, and F1-scores (1.0), along with a high mean Average Precision (mAP@0.5 of 0.995). The models demonstrated the ability to learn subtle visual cues, specifically changes in hue, saturation, and texture, related to the phase transition from a non-gelatinized to a gelatinized state in a heated potato starch solution. Validated successfully on a continuous video stream, this approach proves practical for real-time, non-invasive monitoring, accurately identifying the start of gelatinization without interrupting the process. This study conclusively shows that YOLO-based models provide a highly accurate, automated alternative to subjective, manual methods, thereby reducing reliance on human observation for process control. Additionally, it bridges the gap between process sensing and computer vision, demonstrating the versatility of YOLO models for complex state-monitoring tasks in food science and industrial automation.

Data Availability

The dataset is publicly available at Kaggle:

https://www.kaggle.com/datasets/629a773e6383cd7b450a6f69e8c0cc0e32ec29e702bb8c1d0f8e8e7a6ae7fbe1

Acknowledgement

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

This work is supported by the 'Sylhet Agricultural Univeristy Research System (SAURES) & University Grants Commission (UGC) of Bangladesh research grant 2024-25' (Grant id.: SAURES-UGC-2024-25-LT03-AET-103) and PhD research fellowship provided by SAURES through UGC.

Author contributions

Md. Fahad Jubayer: Conceptualization, methodology, formal analysis, investigation, resources, software, visualization, validation, writing - original draft, writing - reviewing and editing; Mahmud Hasan and Md Khurram Monir Rabby: Methodology, investigation, visualization, software, writing - original draft, writing - reviewing and editing; Md. Mozammel Hoque and Md. Masudur Rahman: Resources, validation, writing - original draft, writing - reviewing and editing; Md. Abdur Rashid Sarker: Conceptualization, methodology, validation, writing - original draft, writing - reviewing and editing, funding acquisition, supervision.

Data Availability

Data will be made available upon reasonable request to the corresponding authors.

Declaration of Competing Interest

The authors declare that they have no competing interests.

References

1 J. Wang, X. Xu, B. Cui, B. Wang, and A.M. Abd El-Aty, Changes in the properties of the corn starch glycerol film in a time-dependent manner during gelatinization, *Food Chem*., 2024, **458**, 140183. https://doi.org/10.1016/j.foodchem.2024.140183

2 D. H. Kringel, A. R. G. Dias, E.D.R. Zavareze and E.A. Gandra, Fruit wastes as promising sources of starch: Extraction, properties, and applications, *Starch-Stärk*, 2020, **72(3-4)**, 1900200. https://doi.org/10.1002/star.201900200

3 R. Naveen and M. Loganathan, Role of varieties of starch in the development of edible films—A review, *Starch-Stärk*, 2024, **76(11-12)**, 2300138. https://doi.org/10.1002/star.202300138

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

4 X. Yan, D.J. McClements, S. Luo, C. Liu and J. Ye, Recent advances in the impact of gelatinization degree on starch: Structure, properties and applications, *Carbohydr. Polym.*, 2024, **340**, 122273. https://doi.org/10.1016/j.carbpol.2024.122273

5 K. Liu and Q. Liu, Enzymatic determination of total starch and degree of starch gelatinization in various products, *Food Hydrocoll.*, 2020, **103**, 105639. https://doi.org/10.1016/j.foodhyd.2019.105639

6 W. Wu, J. Tao, P. Zhu, H. Liu, Q. Du, J. Xiao ... and S. Zhang, A new characterization methodology for starch gelatinization, *Int. J. Biol. Macromol.*, 2019, **125**, 1140-1146. https://doi.org/10.1016/j.ijbiomac.2018.12.180

7 C. Chi, Y. Zou, X. Peng, Y. Yang, B. Chen, Y. He ... and L. Weng, Measurement of starch gelatinization using a spectrophotometer, *Food Hydrocoll.*, 2023, **144**, 108956. https://doi.org/10.1016/j.foodhyd.2023.108956

8 C. Li, Recent progress in understanding starch gelatinization-An important property determining food quality, *Carbohydr. Polym.*, 2022, **293**, 119735. https://doi.org/10.1016/j.carbpol.2022.119735

9 X. Wang, S. Liu and Y. Ai, Gelation mechanisms of granular and non-granular starches with variations in molecular structures, *Food Hydrocoll.*, 2022, **129**, 107658. https://doi.org/10.1016/j.foodhyd.2022.107658

10 P. Chen, L. Yu, T. Kealy, L. Chen and L. Li, Phase transition of starch granules observed by microscope under shearless and shear conditions, *Carbohydr. Polym.*, 2007, **68(3)**, 495-501. https://doi.org/10.1016/j.carbpol.2006.11.002

11 Y. Niu, Y. Zheng, X. Fu, D. Zeng and H. Liu, A novel characterization of starch gelatinization using microscopy observation with deep learning methodology, *J. Food Eng.*, 2022, **327**, 111057. https://doi.org/10.1016/j.jfoodeng.2022.111057

12 G. Zhong, Y. Liu, S. Zhang, J. Liao, Y. Wang, D. Zeng and H. Liu, Efficient and rapid assessment of starch gelatinization through intelligent methodologies, *Int. J. Biol. Macromol.*, 2025, **309**, 142954. https://doi.org/10.1016/j.ijbiomac.2025.142954

13 Q. Li, H. Li and Q. Gao, The influence of different sugars on corn starch gelatinization process with digital image analysis method, *Food Hydrocoll.*, 2015, **43**, 803-811. https://doi.org/10.1016/j.foodhyd.2014.08.012

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

14 Q. Li, Q. Xie, S. Yu and Q. Gao, Application of digital image analysis method to study the gelatinization process of starch/sodium chloride solution systems, *Food Hydrocoll.*, 2014, **35**, 392-402. https://doi.org/10.1016/j.foodhyd.2013.06.017

15 Y. Wang, Y. Ma, X. Gao, Z. Wang and S. Zhang, Insights into the gelatinization of potato starch by in situ 1HNMR, *RSC Adv.*, 2022, 12(6), 3335-3342. https://doi.org/10.1039/D1RA08181K

16 F. Li, L. Zhang, H. Liu, F. Wang, J. Zhao, Z. Ke, L. Liu, Z. Hu and W. Huang, Focused beam reflectance measurement (FBRM) for determination of starch granule diameter distribution and monitoring granule change and gelatinization degree during the gelatinization process, *LWT - Food Sci. Technol.*, 2025, **232**, 118442. https://doi.org/10.1016/j.lwt.2025.118442

17 M Schirmer, M. Jekle, and T. Becker, Starch gelatinization and its complexity for analysis. *Starch-Stärk*, 2015, **67(1-2)**, 30-41. https://doi.org/10.1002/star.201400071

18 S. Sayar, M. Turhan and H. Köksel, Application of unreacted-core model to in situ gelatinization of chickpea starch, *J. Food Eng.*, 2003, **60(4)**, 349-356. https://doi.org/10.1016/S0260-8774(03)00057-8

19 L. Lamberts, E. De Bie, V. Derycke, W. S. Veraverbeke, W. De Man and J. A. Delcour, (2006). Effect of processing conditions on color change of brown and milled parboiled rice, *Cereal Chem.*, 2006, *83(1)*, 80-85. https://doi.org/10.1094/CC-83-0080

20 E. Taghinezhad, M. H. Khoshtaghaza, S. Minaei, T. Suzuki and T. Brenner, Relationship between degree of starch gelatinization and quality attributes of parboiled rice during steaming, *Rice Sci.*, 2016, **23(6)**, 339-344. https://doi.org/10.1016/j.rsci.2016.06.007

21 G. Abhiram and K. S. P. Amarathunga, Effects of far-infrared radiation on the gelatinized rice starch granules, *Drying Technol.*, 2024, **42(1)**, 114-124. https://doi.org/10.1080/07373937.2023.2272179

22 L. Zhu, P. Spachos, E. Pensini and K. N. Plataniotis, Deep learning and machine vision for food processing: A survey, *Curr. Res. Food Sci.*, 2021, **4**, 233-249. https://doi.org/10.1016/j.crfs.2021.03.009

23 M. Hasan, M.K.M. Rabby, I. Jahan, M.J.A. Soeb and M.F. Jubayer, The Evolution and Advancement of YOLO Algorithms in Object Detection: From Real-Time Breakthroughs to Modern Architectures, *Preprints*, 2025. https://www.preprints.org/manuscript/202510.2019

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

24 M. J. A. Soeb, M. F. Jubayer, T. A. Tarin, M. R. Al Mamun, F. M. Ruhad, A. Parven ... and I. M. Meftaul, Tea leaf disease detection and identification based on YOLOv7 (YOLO-T), *Sci. Rep*., 2023, 13(1), 6078. https://doi.org/10.1038/s41598-023-33270-4

25 M. F. Jubayer, F. M. Ruhad, M. S. Kayshar, Z. Rizve, M. J. Alam Soeb, S. Izlal and I. Md Meftaul, Detection and Identification of Honey Pollens by YOLOv7: A Novel Framework toward Honey Authenticity, *ACS Agric. Sci. Technol.*, 2024, **4(7)**, 747-758. https://doi.org/10.1021/acsagscitech.4c00220

26 X. Cheng, S. Zhu, Z. Wang, C. Wang, X. Chen, Q. Zhu and L. Xie, Intelligent vision for the detection of chemistry glassware toward AI robotic chemists. *Artif. Intell. Chem.*, 2023, **1(2)**, 100016. https://doi.org/10.1016/j.aichem.2023.100016

27 R. Sasaki, M. Fujinami and H. Nakai, Application of object detection and action recognition toward automated recognition of chemical experiments. *Digit. Discov.*, 2024, **3(12)**, 2458-2464. https://doi.org/10.1039/D4DD00015C

28 C.K. Chou, R. Karmakar, Y.M. Tsao, L.W. Jie, A. Mukundan, C.W. Huang, T.H. Chen, C.Y. Ko and H.C. Wang, Evaluation of spectrum-aided visual enhancer (SAVE) in esophageal cancer detection using YOLO frameworks. *Diagnostics*, 2024, **14(11),** 1129. https://doi.org/10.3390/diagnostics14111129

29 A. Dhelia and S. Chordia, YOLO-based food damage detection: an automated approach for quality control in food industry. In *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)* (pp. 1444-1449), 2024, IEEE. https://doi.org/10.1109/I-SMAC61858.2024.10714664

30 J. Dong, J. Wang, H. Lin and W. Liu, M-YOLOv8s: Classification and Identification of Different Microalgae Species Based on the Improved YOLO v8s Model for Prevention of Harmful Algal Blooms, *ACS ES&T Water*, 2024, **5(1)**, 329-340. https://doi.org/10.1021/acsestwater.4c00853

31 I. Azurmendi, E. Zulueta, J. M. Lopez-Guede, J. Azkarate and M. González, Cooktop sensing based on a YOLO object detection algorithm, *Sens.*, 2023, **23(5)**, 2780. https://doi.org/10.3390/s23052780

32 B. Wang, H. Lv, X. Wang, M. Hao, D. Kirk, D. Guay ... and Z. Ruan, Quantifying bubble-induced diffusion resistance through real-time SAM-assisted YOLO high density bubble

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

detection algorithm, *Chem. Eng. J.*, 2025, **512**, 162422. https://doi.org/10.1016/j.cej.2025.162422

33 F. Romadhon, F. Rahutomo, J. Hariyono, S. Sutrisno, M. E. Sulistyo, M. H. Ibrahim and S. Pramono, Food image detection system and calorie content estimation using yolo to control calorie intake in the body, In *E3S Web of Conferences* (Vol. 465, p. 02057), 2023, EDP Sciences. https://doi.org/10.1051/e3sconf/202346502057

34 F. Jubayer, J. A. Soeb, A. N. Mojumder, M. K. Paul, P. Barua, S. Kayshar ... and A. Islam, Detection of mold on the food surface using YOLOv5, *Curr. Res. Food Sci.*, 2021, **4**, 724-728. https://doi.org/10.1016/j.crfs.2021.10.003

35 I. u Nisa, B. A. Ashwar, A. Shah, A. Gani, A. Gani and F. A. Masoodi, Development of potato starch based active packaging films loaded with antioxidants and its effect on shelf life of beef, *J. Food Sci. Technol.*, 2015, **52(11)**, 7245-7253. https://doi.org/10.1007/s13197-015-1859-3

36 G. Jocher, A. Chaurasia and J. Qiu, YOLO by Ultralytics, 2023. Available at: https://github.com/ultralytics/ultralytics (Accessed on 25 September 2025)

37 G. Wang, Y. Chen, P. An, H. Hong, J. Hu and T. Huang, UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios, *Sens.*, 2023, 23(16), 7190. https://doi.org/10.3390/s23167190

38 B. Xiao, M. Nguyen and W. Q. Yan, Fruit ripeness identification using YOLOv8 model, *Multimedia Tools Appl.*, 2024, **83(9)**, 28039-28056. https://doi.org/10.1007/s11042-023-16570-9

39 Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye and D. Ren, Distance-IoU loss: Faster and better learning for bounding box regression, In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 07, pp. 12993-13000), 2020. https://doi.org/10.1609/aaai.v34i07.6999

40 X. Li, W. Wang, L. Wu, S. Chen, X. Hu, J. Li ... and J. Yang, Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection, *Adv. Neural Inf. Process. Syst.*, 2020, **33**, 21002-21012.

41 C. Y. Wang, I. H. Yeh and H. Y. Mark Liao, YOLOv9: Learning what you want to learn using programmable gradient information, In *European conference on computer vision* (pp. 1-21), 2024, Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72751-1_1

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

42 R. Khanam and M. Hussain, YOLOv11: An overview of the key architectural enhancements. arXiv preprint arXiv:2410, 2024, 17725. https://doi.org/10.48550/arXiv.2410.17725

43 Ultralytics, "YOLO11 Documentation," Ultralytics YOLO Docs, 2024. [Online]. Available: https://docs.ultralytics.com/models/yolo11/

44 T. Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, Focal loss for dense object detection, In *Proceedings of the IEEE international conference on computer vision* (pp. 2980-2988), 2017. https://doi.org/10.48550/arXiv.1708.02002

45 N. Jegham, C. Y. Koh, M. Abdelatti and A. Hendawi, YOLO evolution: A comprehensive benchmark and architectural review of YOLOv12, YOLO11, and their previous versions, arXiv preprint arXiv:2411.00201, 2024. https://doi.org/10.48550/arXiv.2411.00201

46 Y. Tian, Q. Ye and D. Doermann, YOLOv12: Attention-centric real-time object detectors. arXiv preprint arXiv:2502.12524, 2025. https://doi.org/10.48550/arXiv.2502.12524

47 R. Sapkota, M. Flores-Calero, R. Qureshi, C. Badgujar, U. Nepal, A. Poulose ... and M. Karkee, YOLO advances to its genesis: a decadal and comprehensive review of the You Only Look Once (YOLO) series, *Artif. Intell. Rev.*, 2025, **58(9)**, 274. https://doi.org/10.1007/s10462-025-11253-3

48 M. A. R. Alif and M. Hussain, YOLOv12: A breakdown of the key architectural features, arXiv preprint arXiv:2502.14740, 2025. https://doi.org/10.48550/arXiv.2502.14740

49 Z. Hong, T. Yang, X. Tong, Y. Zhang, S. Jiang, R. Zhou, ... and S. Liu (2021). Multi-scale ship detection from SAR and optical imagery via a more accurate YOLOv3. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 6083-6101. https://doi.org/10.1109/JSTARS.2021.3087555

50 B. Gašparović, G. Mauša, J. Rukavina and J. Lerga, Comparative Analysis of YOLOv7 with Modified Mosaic Augmentation against YOLOv8-11 for Object Detection in Unbalanced Datasets, In *2025 10th International Conference on Smart and Sustainable Technologies (SpliTech)* (pp. 1-4), 2025, IEEE. https://doi.org/10.23919/SpliTech65624.2025.11091726

51 Y. Shen, Z. Yang, Z. Khan, H. Liu, W. Chen and S. Duan, Optimization of improved YOLOv8 for precision tomato leaf disease detection in sustainable agriculture. *Sens.*, 2025, **25(5)**, 1398. https://doi.org/10.3390/s25051398

52 D. Roblek, C. Szegedy and J. S. Jurewicz, U.S. Patent No. 10,467,493. Washington, DC: U.S. Patent and Trademark Office, 2019.

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

53 K. Li, X. Wei, Q. Wang and W. Zhang, Research on Strawberry Visual Recognition and 3D Localization Based on Lightweight RAFS-YOLO and RGB-D Camera. *Agric.*, 2025, **15(21)**, 2212. https://doi.org/10.3390/agriculture15212212

54 H.Y. Huang, Y.P. Hsiao, A. Mukundan, Y.M. Tsao, W.Y. Chang and H.C. Wang, Classification of skin cancer using novel hyperspectral imaging engineering via YOLOv5. *J. Clin. Med.,* 2023, **12(3)**, 1134. https://doi.org/10.3390/jcm12031134

55 A. Mukundan, R. Karmakar, D. Gupta and H.C. Wang, Deep Learning-Based Toolkit Inspection: Object Detection and Segmentation in Assembly Lines. *Comput Mater Contin.*, 2026, **86(1)**, 1-23. http://dx.doi.org/10.32604/cmc.2025.069646

56 J. Tao, J. Huang, L. Yu, Z. Li, H. Liu, B. Yuan and D. Zeng, A new methodology combining microscopy observation with Artificial Neural Networks for the study of starch gelatinization. *Food Hydrocoll*., 2018, **74**, 151-158. https://doi.org/10.1016/j.foodhyd.2017.07.037

57 C. Cai, L. Zhao, J. Huang, Y. Chen and C. Wei, C, Morphology, structure and gelatinization properties of heterogeneous starch granules from high-amylose maize. *Carbohydr. Polym*., 2014, **102**, 606-614. https://doi.org/10.1016/j.carbpol.2013.12.010

58 J. Brunnschweiler, D. Luethi, S. Handschin, Z. Farah, F. Escher and B. Conde-Petit, Isolation, physicochemical characterization and application of yam (*Dioscorea* spp.) starch as thickening and gelling agent. *Starch-Stärke*, 2005, **57(3-4)**, 107-117. https://doi.org/10.1002/star.200400327

Sustainable Food Technology Accepted Manuscript

View Article Online
DOI: 10.1039/D6FB00010J

Data availability statement

Data will be made available upon reasonable request to the corresponding authors.

Open Access Article. Published on 18 May 2026. Downloaded on 5/21/2026 9:30:34 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Sustainable Food Technology Accepted Manuscript