HHS Public Access

Author manuscript

Am J Vet Res. Author manuscript; available in PMC 2025 March 31.

Published in final edited form as:

Am J Vet Res. 2025 March 01; 86(Suppl 1): S27–S37. doi:10.2460/ajvr.24.10.0305.

Toward a rapid, sensitive, user-friendly, field-deployable artificial intelligence tool for enhancing African swine fever diagnosis and reporting

Aliva Bakshi, MTech, MSc[1], **Jake Stetson, BS**[1], **Lihua Wang, PhD**[2], **Jishu Shi, PhD**[2], **Doina Caragea, PhD**[1], **Laura C. Miller, PhD**[3,*]

[1]Department of Computer Science, Carl R. Ice College of Engineering, Kansas State University, Manhattan, KS

[2]Department of Anatomy and Physiology, College of Veterinary Medicine, Kansas State University, Manhattan, KS

[3]Department of Diagnostic Medicine/Pathobiology, College of Veterinary Medicine, Kansas State University, Manhattan, KS

Abstract

Objective—African swine fever (ASF) is a lethal and highly contagious transboundary animal disease with the potential for rapid international spread. Lateral flow assays (LFAs) are sometimes hard to read by the inexperienced user, mainly due to the LFA sensitivity and reading ambiguities. Our objective was to develop and implement an AI-powered tool to enhance the accuracy of LFA reading, thereby improving rapid and early detection for ASF diagnosis and reporting.

Methods—Here, we focus on the development of a deep learning–assisted, smartphone-based AI diagnostic tool to provide accurate decisions with higher sensitivity. The tool employs state-of-the-art You Only Look Once (YOLO) models for image classification. The YOLO models were trained and evaluated using a dataset consisting of images where the lateral flow assays are manually labeled as positives or negatives. A prototype JavaScript website application for ASF reporting and visualization was created in Azure. The application maintains the distribution of the positive predictions on a map as the positive cases are submitted by users.

Results—The performance of the models is evaluated using standard evaluation metrics for classification tasks, specifically accuracy, precision, recall, sensitivity, specificity, and F1 measure. We acquired 86.3 ± 7.9% average accuracy, 96.3 ± 2.04% average precision, 79 ± 13.20% average recall, and an average F1 score of 0.87 ± 0.088 across 3 different train/development/test splits of the datasets. Submitting a positive result of the deep learning model updates a map with a location marker for positive results.

*Corresponding author: Dr. Miller (lauramiller@ksu.edu).

Disclosures

The authors have nothing to disclose. No AI-assisted technologies were used in the composition of this manuscript.

Conclusions—Combining clinical data learning and 2-step algorithms enables a point-of-need assay with higher accuracy.

Clinical Relevance—A rapid, sensitive, user-friendly, and deployable deep learning tool was developed for classifying LFA test images to enhance diagnosis and reporting, particularly in settings with limited laboratory resources.

Keywords

deep learning; predictive diagnostics; AI; African swine fever (ASF); smartphone mobile app

African swine fever (ASF) is a highly contagious viral disease that affects domestic and wild pigs, with no vaccine or cure.[1–3] Early diagnosis is vital to prevent its rapid spread and massive pig fatalities and to minimize economic losses, trade restrictions, and food security issues. Misdiagnosis or delays can worsen outbreaks, severely impacting farmers' livelihoods and disrupting global meat supply chains. Hence, accurate diagnosis of ASF is crucial to protect animal health and agricultural economies. Traditionally, clinical observations, laboratory tests, or postmortem examinations are widely used for diagnosing ASF. These methods are crucial for accurate identification but require specialized equipment, expertise, and time, and thus reliance only on such methods could lead to delays in detecting outbreaks. Furthermore, the clinical signs of ASF can be similar to those of other diseases, a common cause for misdiagnosis. Together, these factors can hinder effective outbreak detection and control. This is why rapid and field-based diagnostic alternatives for ASF are extremely important and greatly needed for controlling outbreaks.

Several recent studies have focused on this task. For example, Chen et al[4] developed a portable, sample-to-answer device that allows for ASF detection as needed in less than 30 minutes. When evaluated at 6 different farms and slaughter facilities, the devices achieved high accuracy in terms of ASF diagnosis. A comparison of the device's predictions with lab-based reference quantitative PCR tests showed 92.2% accuracy for positive tests and 93.6% for negative tests. Tilocca et al[5] focused on identifying ASF virus (ASFV) epitopes likely to elicit antibody production from ASFV structural proteins using bioinformatic tools. The predicted epitome sequences were used to generate a list of potential biomarkers that can serve as targets for diagnostic tests. Such biomarkers can be used in the development of rapid tests for detecting ASF at the point of need on a farm. In another study, Zhang et al[6] used the P30 protein as a target for detecting ASF infection. This protein is known to be expressed early during the infection and has high antigenic properties.[7] By targeting this protein, a recombinant protein-based colloidal gold immunochromatography assay was developed as a potential tool for fast ASF detection. In a related study, Aira et al[8] used the P72 protein as another highly antigenic target for ASF infections to develop both an antigen detection test based on a recombinant antibody against the P72 protein as well as an antibody detection test based on a recombinant form of this protein. By integrating the antigen and antibody detection tests, the authors further developed a combined lateral flow assay (LFA). Lateral flow assays represent disposable, fast, inexpensive, convenient, and easy-to-use pen-side tests for clinical diagnosis. Similar to the abovementioned studies, Wang et al[9] developed 2 LFA tests by targeting proteins P30 and P72, respectively. Evaluation of these tests suggested that they can be used as rapid, sensitive, user-friendly,

and field-deployable tools for ASF control, particularly when the laboratory resources are limited.

Although LFA is a quick diagnosis tool for ASF, it lacks reliability and accuracy compared to traditional laboratory assays; moreover, untrained individuals analyze most LFAs with naked-eye detection, which inevitably limits the accuracy, especially for low virus titers. This is where technologies like AI can be integrated to make an efficient diagnosis system, particularly deep learning and computer vision that handles image data.

Artificial intelligence-[10,11] and machine learning (ML)-assisted[12] approaches in veterinary medicine can improve the reliability and accuracy of various sensors, making their performance more comparable to human performance. These technologies can lead to accurate and consistent predictions at fast speeds. Various AI and ML techniques have been explored for diagnosing ASF. For example, based on the observation that animals affected by ASF have clinical signs like fever, weakness, and subsequent deceleration in activities, Fernandez-Carrion et al[13] utilized video monitoring for early ASF detection in pigs. The movements were recorded and digitally processed using the optical flow algorithm based on the Horn-Schunck methodology.[14]

Using the k-means algorithm[15] and a support vector machine classifier,[16] the approach predicted significant changes in an animal's motion after the animal was experimentally infected with ASFV. In a similar study, Fernandez-Carrion et al[17] applied deep learning and computer vision models, specifically AlexNet[18] and Kalman filters,[19] to track and compute animal motion in real time, with a focus on Eurasian wild boar. The experiments showed a negative correlation between deceleration in motion and fever caused by ASF. The results of these studies suggest that ML-based animal motion monitoring systems can be helpful for both farmers and animal health services to detect early signs of ASF and provide a potential solution toward nonintrusive, real-time ASF diagnosis.

In another line of work, Liang et al[20] utilized ASF outbreak data together with meteorological data from the WorldClim database and developed an ASF outbreak prediction model using feature selection methods with a random forest classifier.[21] The model achieved accuracy between 76.02% and 84.65% on an independent test set, and the feature analysis suggested that the outbreak was significantly related to precipitation. A global online system for ASF prediction was also developed using this model. In a similar study, Andraud et al[22] proposed a statistical framework based on random forest[21] to analyze the spatiotemporal features of an ASF epidemic in Romania. The landscape, particularly forest and wetland area, waterways, and human activity, were recognized as the main risk factor for the epidemic. The study emphasized the need for strict biosecurity measures on farms and in transportation. In a recent review, Lububu and Kabaso[23] examined both ML approaches and laboratory tests for ASF diagnosis and highlighted their strengths and limitations. They suggested that ML approaches can complement rapid laboratory tests as they can provide insights into the causal relationship between viruses and clinical signs. Furthermore, they emphasized the need for better, more accurate ML models in order to be able to deploy these techniques into a trustworthy, real-time system.

While significant progress has been made in terms of using AI and ML for ASF diagnosis, these technologies have not been used together with LFA test images to eliminate the need for naked-eye interpretation of the LFA results. The integration of LFA and deep learning models for image analysis has the potential to lead to accurate and efficient diagnosis systems. Moreover, if such a process can be deployed on smartphones, it can expedite the diagnosis process even more, making it more user friendly at the same time. Smartphone-based diagnostics can be considered a potential candidate for pen-side testing because it is easy to acquire digitalized images from a smartphone for user-friendly diagnostics. Since the number of current smartphone users is approximately 7 billion, meaning > 80% of the world's population owns smartphones, the assay using a smartphone can provide accessibility and affordability.

Smartphone-based, AI-assisted assays have been used in HIV rapid tests[24] and SARS-CoV-2 detection.[25–28] In this paper, we present deep learning–assisted, smartphone-based AI-ASF-LFA diagnostics as a model system to provide accurate ASF decisions with higher sensitivity. Combining clinical data learning and computer vision algorithms enables a pen-side assay with higher accuracy.

As a main contribution of this work, the deep leaning/AI smartphone application (app) is intended to enhance the interpretation of the LFA testing and reporting based on image analysis and thus prevent inconsistent or erroneous interpretations while minimizing the efforts of domain experts. It can be seen as a tool for a rapid and accurate first-stage screening of the LFA testing images. However, subsequent screening and validation of the positive cases would need to be performed by experts before any important decisions, such as animal movement or euthanasia of the animals, would be performed. In that respect, the tool is better suited for countries with existing cases of the disease instead of previously disease-free countries.

As a secondary contribution, we developed a mapping app that can be used to visualize emerging positive cases. The mapping app can also be used to collect images corresponding to positive cases in a specific country or region. However, the map is not updated with controlled cases and thus cannot be used to reflect only active cases. It could still be useful in terms of preventive measures in the neighborhood of outbreaks to avoid further spread.

Methods

This study was a retrospective analysis of historical data[9] and did not involve any animal research. Therefore, no IACUC approval was required. We used a dataset consisting of lateral flow test (LFT) images to train You Only Look Once (YOLO)[29] models for classifying the test images as positive or negative. You Only Look Once is a family of advanced object detection and classification models used in computer vision. We focused on 2 classification models in the YOLO family,[29] YOLO, version 8 (YOLOv8),[30] and YOLO, version 11 (YOLOv11),[31] as the YOLO models are designed with speed, accuracy, and ease of use in mind and are thus ideal candidates for running in the backend on a smartphone app. We describe the dataset used, the YOLO models, and the workflow for training and evaluating the models in what follows. Furthermore, we also provide an overview of the

smartphone app that can be used pen-side and its interaction with a website that displays the results on a map in real time.

Dataset

Lateral flow assay was performed using Rapid ASFV Ag (p30) and Rapid ASFV Ag (p72) test kits as developed and described by Wang et al.[9] Briefly, ASFV-negative pig sera samples were obtained from pigs inoculated with PBS (pH 7.4; Thermo Scientific), and ASFV-infected pig sera samples were collected from pigs infected with the ASFV VNUA-ASFV-05 L1 strain at various time points: day 7, day 14, and the day of euthanasia. Samples were brought to room temperature (15 to 30 °C) before testing. A 120-μL aliquot of each sample was applied to the sample pad. The sample flowed through the result window by capillary action. The results were interpreted 15 minutes after the sample addition by 2 experienced technicians (annotators). To minimize bias and capture individual interpretations, annotators were instructed not to discuss or adjudicate their responses. A positive test was indicated by the presence of both a red/pink test line and a red/pink control line. A negative test showed only a red/pink control line. The control line was always required to be visible; otherwise, the test was considered invalid and repeated with a new LFA cassette. All photographs were taken using an iPad 10.9-inch 10th Generation (Apple Inc). The technicians had 100% agreement. However, when comparing their manually annotated labels with the real-time PCR as ground truth labels, “Both assays demonstrated high sensitivity (84.21% for p30_LFA and 100% for p72_LFA) with experimentally ASFV-infected pig sera. Specificity was 100% for both LFAs using a panel of PBS-inoculated domestic pig sera. Excellent specificity was also shown for field domestic pig sera (100% for p30_LFA and 93% for p72_LFA) and feral pig sera (100% for both LFAs).”[9]

The final dataset used to develop our models consists of 897 positive LFT images and 766 negative LFT images of combined Rapid ASFV Ag (p30), Rapid ASFV Ag (p72), and Rapid ASFV Ab (p30) tests performed using serum samples collected from 30 subjects on days post challenge (DPC) 0, DPC 3, and last day of DPC. The dataset was split into train, validation, and test subsets used for training our models, fine-tuning hyperparameters, and evaluating the models, respectively. Specifically, we used images of 20 subjects as training a subset, images of 3 subjects as a validation subset, and images of 7 subjects as a test subset. To evaluate model robustness, we constructed 3 different train/validation/test splits based on different combinations of 20, 3, and 7 subjects in those subsets. The 3 splits are denoted by LFA-1, LFA-2, and LFA-3. Statistics of the data for the P30 tests, P72 tests, and the combination of P30 and P72 tests for 3 the splits (LFA-1, LFA-2, and LFA-3) and the train/validation/test subsets are shown in Table 1.

Image classification using YOLO models

Image classification is a task where an ML model (specifically, a computer vision model) assigns a label to an image based on its content. Modern computer vision models are based on deep learning, in particular convolutional neural networks (CNNs),[32] and, more recently, transformer networks,[33] which perform feature representation learning. The feature representations can be used for a variety of tasks, including image classification into 1 of several predefined categories (or classes). In binary image classification, an image is

classified into 1 of 2 distinct classes. In our case, LFT images are classified as positive or negative. We train models based on labeled LFT images (ie, images for which the ground truth label is known) and subsequently use the models to classify new LFT test images as positive or negative.

There are many pretrained CNN and transformer-based models that can be leveraged for image classification through transfer learning.[34] Such pretrained models can be fine-tuned on various downstream tasks with a relatively small number of labeled images. In this study, we focus on YOLOv8[30] and YOLOv11[31] models due to their versatility, speed, and accuracy, which enabled the use of these models on a smartphone. The YOLO models were originally designed for real-time object detection, and they traded off accuracy for model size and speed.[29] However, recent version of these models have produced state-of-the-art accuracy in object detection, and the YOLOv8 and YOLOv11 models have expanded their capabilities to other tasks, including image classification, given that the backbone network (usually a CNN network, such as variants of CSPDarknet-53) has improved significantly over the years. This allows us to use these YOLO models for LFT image classification. A brief overview of these models is provided below.

YOLOv8, is a modification on YOLOv5 model, developed by Ultralytics (https://www.ultralytics.com). It has a modular architecture consisting of a *backbone* that extracts features, a *neck* that connects the backbone to the head and performs multiscale feature fusion, and a *head* that makes predictions. The backbone of YOLOv8 is an adaptation of the CSPDarknet53[30] network, where several architectural innovations and optimizations are introduced to ensure the learning of better, more complex features while maintaining a balance between accuracy and speed. For classification tasks, the backbone network is pretrained on the ImageNet dataset,[35] and the head consists of fully connected layers followed by a softmax activation, which outputs a probability distribution over classes. To fine-tune the network for a particular classification task, the final layer has to be changed according to the classes of interest. Like other YOLO models, YOLOv8 has several variants with smaller or larger number of parameters, including yolov8n-cls.pt (nano), yolov8s-cls.pt (small), yolov8m-cls.pt (medium), yolov8l-cls.pt (large), and yolov8x-cls.pt (extra large). In our experiments, we have used yolov8x-cls.pt weights as the base model, which was further fine-tuned using our dataset. Yolov8x-cls.pt has been pretrained with the architecture specified in the corresponding yolov8x-cls.yaml. We chose yolov8x-cls as our base model due to its best performance.

Another model, YOLOv11 is the most recent development from the YOLO family, which is an improvement on YOLOv8 and was released on September 30, 2024. Like YOLOv8, YOLOv11 can also be effectively used for image classification tasks.[31] The architecture of YOLOv11's backbone is similar to the CSPDarknet53 architecture in YOLOv8, but it has been further optimized for feature extraction and speed by leveraging innovations from prior YOLO versions, including YOLOv8, YOLO version 9,[36] and YOLO version 10.[37] In the original benchmark experiments by Ultralytics, YOLOv11 was shown to classify images at greater accuracy with a lower number of parameters and minimal computational overhead.[31] Given its smaller size and improved speed, YOLOv11 can be deployed on a variety of devices, including edge devices, such as smartphones, cloud platforms,

and servers equipped with NVIDIA graphics processing units. This makes YOLOv11 a preferred model for real-time and large-scale apps. Similar to YOLOv8, YOLOv11 also has several variants, including yolov19n-cls.pt (nano), yolov11s-cls.pt (small), yolov11m-cls.pt (medium), yolov11l-cls.pt (large), and yolov11x-cls.pt (extra large). With the largest architecture, YOLOv11x-cls.pt has the best performance on ImageNet. This version of the model has been pretrained using the architecture in YOLOv11x-cls.yaml. We have used the pretrained weights of the YOLOv11x-cls.pt model and further fine-tuned this model using our training images.

We trained each model for 3,000 iterations with a custom patience of 300. This means that the training process continued for 300 iterations even when there were no more changes in accuracy. Image sizes were fixed to 412 × 412 for all images in the dataset and for both of the models. All of the training and evaluation were conducted on Nvidia A40 graphics processing units with 40GB memory.

Performance metrics

The performance of the trained models is measured using standard evaluation metrics for classification tasks, including accuracy, precision, recall (also known as sensitivity or the true positive [TP] rate [TPR]), specificity (also known as the true negative [TN] rate), and F1 measure. We also used receiver operating characteristic (ROC) curves and the area under the curve (AUC) metric to analyze the performance of our models. These metrics are defined in terms of TP, TN, false positive (FP), and false negative (FN), where TP represents the number of correct positive predictions, TN represents the number of correct negative predictions, FP represents the number of incorrect positive predictions, and FN represents the number of incorrect negative predictions. Together, these numbers define a confusion matrix between model predictions and ground truth. Given these numbers, accuracy can be defined as the ratio of correct predictions by the model over the total number of instances in the dataset, specifically: accuracy = (TP + TN)/(TP + FP + TN + FN).

Precision measures the proportion of TP predictions out of all positive predictions made by the model (both TP and FP), whereas recall (sensitivity or TPR) measures the proportion of TP that the model correctly identifies out of the total number of TPs. Specificity measures the proportion of TN predictions out of all negative cases (TN and FP; ie, how well the model identifies cases without disease). Precision and recall/sensitivity/TPR are defined as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}},$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}},$$

and specificity/TN rate is defined as Specificity = TN/(TN + FP).

Finally, the F1 measure is computed as the harmonic mean of precision and recall, and it is defined as: F1 = 2·precision·recall/(precision + recall).

The F1 measure captures the trade-off between precision and recall, and it is high when both are high.

An ROC curve plots the TPR against the FP rate (FPR) for various threshold values on the prediction probability score, where TPR is calculated as TP/(TP + FN), and FPR is calculated as FP/(FP + TN). Different thresholds will result in different TPR and FPR values, and the ROC curve shows how these values change across thresholds. The AUC value quantifies the level of separability between classes. A higher AUC value indicates better model performance.

Workflow for model training and evaluation

Figure 1 shows the workflow that we follow to train YOLOv8 and YOLOv11 models. Lateral flow test images are first split into 3 subsets, specifically train, development, and test subsets. You Only Look Once, version 8, and YOLOv11 models are trained on the LFT image classification task using the training data, whereas hyperparameters are fine-tuned using the development data. Finally, the test data are used to assess the model performance on unseen data (ie, data that the model has not seen during training).

Artificial intelligence–ASF-LFA tool: clinical test interpretations and outcome visualization with an AI-based tool

Figure 2 shows the workflow of the AI-ASF-LFA tool that enables a sample-to-answer platform for ASF with the aid of deep learning–assisted determination. From inside a smartphone app, users can take a picture of an LFA test image or choose a picture from the library and then submit the image to the YOLO model for analysis. The model outputs the result as positive or negative together with the model confidence for the prediction made, thus helping to reduce reading ambiguities. Subsequently, the model, through the smartphone app, interacts with a website mapping app that displays positive results on a map by using the location information of the user (or a location provided by the user) together with a timestamp of the positive case report. Positive cases accumulate on the map for various users to keep track of the spread of the disease by time/day.

Together, the smartphone-based AI tool and the website mapping app provide a great opportunity to meet the *real-time connectivity*, *ease of specimen collection*, *affordable*, *sensitive*, *specific*, *user friendly*, *rapid and robust*, *equipment free*, and *deliverable to end users* criteria,[38] which are the new criteria for digital connectivity.

Results

This section presents the evaluation and performance analysis of our models. Table 2 shows the performance metrics for both YOLOv8 and YOLOv11 models for the 3 dataset splits (LFA-1, LFA-2, and LFA-3, respectively) as well as average and SD results over the 3 splits. The performance metrics are measured on test data, which is unknown to our trained models. As can be seen from the table, both models had the best performance on the LFA-3 split and the worst performance on the LFA-1 split, suggesting that more training data may be needed to ensure more robust models to data variations. By analyzing the average and SD of the models over the 3 splits, we can conclude that the YOLOv8 performs better than the

YOLOv11 on our LFA test classification task. While this may seem counterintuitive given the YOLOv11's state-of-the-art results originally reported by Ultralytics on benchmark datasets, most probably the original YOLOv11 model was significantly fine-tuned to identify the best-performing model hyperparameters. However, there is no guarantee that the model will produce the best results on any custom datasets even with fine-tuning as suggested by our experiments and other related works. For example, Sharma et al[39] compared Faster Region-Convolutional Neural Network (Faster R-CNN), YOLOv8, YOLO, version 9, YOLO, version 10, and YOLOv11 models for object detection/classification. In their experiments, YOLOv8 had better performance than YOLOv11 in most metrics but YOLOv11 had better inference speed. The authors concluded that "the YOLOv11 model offers an excellent balance between speed and accuracy."

Figure 3 shows confusion matrices for the test subsets of each of the LFA-1, LFA-2, and LFA-3 splits as well as ROC curves and the corresponding AUC values for each of the 2 models trained. Specifically, panels on the left show the confusion matrices and the ROC curves/AUC values for the YOLOv8 model, whereas panels on the right show the confusion matrices and the ROC curves/AUC values for the YOLOv11 model. We can see that in general, the models make more FN mistakes as compared to FP mistakes, which is ideal with respect to minimizing the human effort required to validate positive cases before any significant action, such as animal movement or animal euthanasia, would be taken. In such scenarios, the YOLO models would only be used for the initial screening of potential positive cases, thus reducing the amount of data that the experts may need to validate. In practice, this is feasible in countries where the disease is more prevalent. However, it could still cause delays in previous disease-free countries despite having a small number of FPs.

Figure 4 shows some sample LFA test images predicted by our models. For each image, "Ground Truth" indicates the original label manually assigned by domain experts; "YOLOv8 → label" and "YOLOv11 → label" show the predicted class for each model, with confidence score in brackets. Both YOLOv8 and YOLOv11 correctly classified several clearly positive images, achieving 100% confidence. Even with a faint second line, both models correctly identified a positive result with high confidence (99%). In another instance with a faint line, both models correctly classified the image as positive, but YOLOv8 exhibited higher confidence (100%) than YOLOv11 (92%). However, the models also exhibited discrepancies. For example, YOLOv8 correctly classified an image as positive (91% confidence), while YOLOv11 incorrectly classified it as negative (82% confidence). Conversely, YOLOv8 misclassified a positive image as negative (96% confidence), while YOLOv11 correctly classified it as positive (62% confidence). Similar discrepancies are shown in other images that demonstrate the models' ability to classify LFA images with high confidence, particularly when both lines are clear. Furthermore, the models generally exhibit lower confidence on images with faint, difficult-to-recognize lines, suggesting good calibration between performance and confidence. The errors primarily occur with these ambiguous images. Overall, this analysis supports the conclusion that YOLOv8 generally outperforms YOLOv11 in both accuracy and confidence. Finally, one image highlights a potential annotation error: while human annotators labeled the image as positive, both models correctly predicted negative (YOLOv8: 100% confidence; YOLOv11: 73% confidence), suggesting the models' potential to identify inconsistencies in manual labeling.

As seen in Table 2, sensitivity and specificity of our best YOLOv8 model are 73% and 91%, respectively, suggesting a relatively modest positive predicting power of the model. While we do not evaluate the YOLO models in the same fashion based on the type of test and the type of pig sera, and a direct comparison is not possible, our sensitivity and specificity results (specifically, 73% and 91% for the YOLOv8 model) are very promising given the sensitivity and specificity of the test according to the human annotations. As reflected by our own error analysis and also the results from Wang et al,[9] ground truth annotation based on PCR may not always agree with what can be observed in the LFA testing images, and thus our dataset can be seen as noisy and rather small according to deep learning standards. With a larger dataset, ideally annotated by domain experts based on images, we expect that specificity and sensitivity, and thus the positive predictive power of the model, can be further improved, leading to models that can potentially be used in practice to inform ASF decision makers, especially in countries with a high prevalence of ASF disease. However, as is the case with many emerging AI technologies, actions such as movement of the animal would not happen without additional testing and validation by experts. In that respect, the deep learning model would help filter out irrelevant/negative cases in a consistent manner while allowing the ASF experts and decision makers to focus on potentially positive cases.

Discussion

Currently, ASF is listed as a legally notifiable disease that must be reported to the World Organization for Animal Health. The economic losses to the global pig industry have been insurmountable since the outbreak of ASF. Control and eradication of ASF are very critical during the current pandemic. Until an ASF vaccine with high immunoprotective potential is available, the prevention, control, and eradication of ASF is based on the implementation of appropriate surveillance and strict biosecurity measures. The success of surveillance activities depends on the availability of the most appropriate, sensitive, and rapid diagnosis. Toward this goal, as a main contribution of this work, we developed an AI tool (based on YOLO models) for enhancing ASF diagnosis and reporting based on image analysis. The classifier can be used as a smartphone app. The validation of the YOLO models was performed using ground truth annotations of the LFA test images in terms of standard metrics in ML.

As a second contribution of this work, we also developed a proof-of-concept website mapping app. Its main purpose is to keep track of reported positive cases and to collect the corresponding images for further validation. The granularity is currently at the city level, but it can be customized as needed. If multiple positive cases are reported at the same location, a counter is used to keep track of the number of cases at the same location, and that location would be given a brighter color to account for the larger number of cases. Potentially, an admin could “approve”/validate the positive cases before allowing them to be displayed on the map for enhanced accuracy. The day when a positive case was reported can be recorded, and that information can be used to see reported cases by day, week, month, etc. Given that the map is populated based on positive cases as initially identified by the YOLO models based on LFA test image classification, the map would not reflect cases when the animal is euthanized. In that respect, the map provides spatial and temporal information about the occurrence of the disease but does not reflect the actions taken to control the disease. Thus,

it would be useful for the visualization of emerging positive cases by the general public and decision makers, who could potentially help prevent new cases in neighboring regions by using extra measures to prevent spread.

The development of the AI-ASF-LFA app represents a significant advancement in field-deployable point-of-need testing. This novel approach offers several practical advantages over existing commercial apps, such as the MatMaCorp ASFV Detection System.[40] By eliminating the need for human interpretation and combining clinical data learning with classification models, AI-ASF-LFA provides higher-accuracy test results within a short span of 15 minutes. Given the widespread use of smartphones, the integration of AI-ASF-LFA with these devices has the potential to improve accessibility and affordability as well as immediate reporting.

The ability of AI-ASF-LFA to continuously learn from additional clinical data and increase its diagnostic accuracy further strengthens its value as a reliable and adaptable tool for diagnosis. This achievement has the potential to significantly impact diagnostics by providing World Organization for Animal Health members with the ability to make rapid decisions and share the relevant epidemiological information. Furthermore, the AI-ASF-LFA has the potential to be used in a differentiating infected from vaccinated animals[41] strategy as planned by VACDIVA (https://vacdiva.eu/). However, several limitations should be acknowledged. The current study does not explicitly demonstrate an increase in sensitivity of the LFA test through the use of the smartphone-based tool. To directly assess this, future studies should incorporate a larger number of samples near the limit of detection of the LFA test and compare the AI model's predictions with those of experienced human evaluators and with gold-standard laboratory assays, such as PCR or ELISA. This study did not explicitly evaluate the model's robustness to variations in image acquisition (eg, camera settings and device used). Furthermore, the model's generalizability to different populations, field settings, and LFA test types remains to be determined. These factors may require further investigation and potential adjustments to the model for broader applicability.

In conclusion, AI-ASF-LFA offers a promising solution for improving the accuracy, efficiency, and accessibility of field-deployable point-of-need testing.

Acknowledgments

The authors would like to thank the Veterinary Diagnostic Laboratory staff, Comparative Medicine staff, and staff of the Biosecurity Research Institute at Kansas State University for their technical help in the lateral flow assays.

Funding

This study was supported by the K-State College of Veterinary Medicine SMILE Intramural Grant Program. Research reported in this publication was also partially supported by the Cognitive and Neurobiological Approaches to Plasticity Center of Biomedical Research Excellence of the NIH under grant No. P20GM113109. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

1. Dixon LK, Sun H, Roberts HJAR. African swine fever. Antiviral Res. 2019;165:34–41. doi:10.1016/j.antiviral.2019.02.018 [PubMed: 30836106]

2. Wu K, Liu J, Wang L, et al. Current state of global African swine fever vaccine development under the prevalence and transmission of ASF in China. Vaccines (Basel). 2020;8(3):531. doi:10.3390/vaccines8030531 [PubMed: 32942741]
3. Rock DL. Challenges for African swine fever vaccine development—"… perhaps the end of the beginning." Vet Microbiol. 2017;206:52–58. doi:10.1016/j.vetmic.2016.10.003 [PubMed: 27756505]
4. Chen L, Wen K, Chen FE, et al. Portable magnetofluidic device for point-of-need detection of African swine fever. Anal Chem. 2021;93(31):10940–10946. doi:10.1021/acs.analchem.1c01814 [PubMed: 34319068]
5. Tilocca B, Greco V, Soggiu A, et al. Multiepitope array as the key for African swine fever diagnosis. Vet Immunol Immunopathol. 2023;257:110548. doi:10.1016/j.vetimm.2023.110548 [PubMed: 36736103]
6. Zhang M, Lingling G, Xinxin X, Kuang H, Chuanlai X, Liqiang L. An ultrasensitive strip sensor for rapid detection of African swine fever virus. Chin J Anal Chem. 2024;52(7):100416. doi:10.1016/j.cjac.2024.100416
7. Gomez-Puertas P, Rodrıguez F, Oviedo JM, Brun A, Alonso C, Escribano JM. The African swine fever virus proteins p54 and p30 are involved in two distinct steps of virus attachment and both contribute to the antibody-mediated protective immune response. Virology. 1998;243(2):461–471. doi:10.1006/viro.1998.9068 [PubMed: 9568043]
8. Aira C, Gonzalez-Garcıa G, Martınez-Cano J, et al. Simultaneous detection of antigen and antibodies of African swine fever in a novel combo lateral flow assay. Vaccines. 2024;12(3):307. doi:10.3390/vaccines12030307 [PubMed: 38543941]
9. Wang L, Kim J, Kang H, et al. Development and evaluation of two rapid lateral flow assays for on-site detection of African swine fever virus. Front Microbiol. 2024;15:1429808. doi:10.3389/fmicb.2024.1429808 [PubMed: 39268541]
10. Appleby RB, Basran PS. Artificial intelligence in veterinary medicine. J Am Vet Med Assoc. 2022;260(8):819–824. doi:10.2460/javma.22.03.0093 [PubMed: 35358065]
11. Pereira AI, Franco-Gonccalo P, Leite P, et al. Artificial intelligence in veterinary imaging: an overview. Vet Sci. 2023;10(5):320. doi:10.3390/vetsci10050320 [PubMed: 37235403]
12. Cihan P, Gokce E, Kalipsiz O. A review of machine learning applications in veterinary field. Kafkas Universitesi Veteriner Fakultesi Dergisi. 2017;23(4):673–680.
13. Fernandez-Carrion E, Martınez-Aviles M, Ivorra B, Martınez-Lopez B, Ramos AM, Sanchez-Vizcaıno JM. Motion-based video monitoring for early detection of live-stock diseases: the case of African swine fever. PloS One. 2017;12(9):e0183793. doi:10.1371/journal.pone.0183793 [PubMed: 28877181]
14. Horn BKP, Schunck BG. Determining optical flow. Artif Intell. 1981;17(13):185–203. doi:10.1016/0004-3702(81)90024-2
15. Lloyd S Least squares quantization in PCM. IEEE Trans Inf Theory. 1982;28(2):129–137. doi:10.1109/TIT.1982.1056489
16. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intell Sys Appl. 1998;13(4):18–28. doi:10.1109/5254.708428
17. Fernandez-Carrion E, Barasona JA, Sanchez A, Jurado C, Cadenas-Fernandez E, Sanchez-Vizcaıno JM. Computer vision applied to detect lethargy through animal motion monitoring: a trial on African swine fever in wild boar. Animals (Basel). 2020;10(12):2241. doi:10.3390/ani10122241 [PubMed: 33260362]
18. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012;25(2):84–90.
19. Kalman RE. A new approach to linear filtering and prediction problems. J Basic Eng. 1960;82(series D):35–45.
20. Liang R, Lu Y, Qu X, et al. Prediction for global African swine fever outbreaks based on a combination of random forest algorithms and meteorological data. Transbound Emerg Dis. 2020;67(2):935–946. doi:10.1111/tbed.13424 [PubMed: 31738013]
21. Breiman L Random forests. Mach Learn. 2001;45(1):5–32. doi:10.1023/A:1010933404324

22. Andraud M, Bougeard S, Chesnoiu T, Rose N. Spatiotemporal clustering and random forest models to identify risk factors of African swine fever outbreak in Romania in 2018–2019. Sci Rep. 2021;11(1):2098. doi:10.1038/s41598-021-81329-x [PubMed: 33483559]
23. Lububu S, Kabaso B. A systematic literature review on machine learning and laboratory techniques for the diagnosis of African swine fever (ASF). In Proceedings of the 2023 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems. icABCD, Durban, South Africa. 2023. doi:10.1109/icABCD59051.2023.10220551
24. Turbe V, Herbst C, Mngomezulu T, et al. Deep learning of HIV field-based rapid tests. Nat Med. 2021;27(7):1165–1170. doi:10.1038/s41591-021-01384-9 [PubMed: 34140702]
25. Fozouni P, Son S, Leon Derby MD, et al. Amplification-free detection of SARS-CoV-2 with CRISPR-Cas13a and mobile phone microscopy. Cell. 2021;184(2):323–333. doi:10.1016/j.cell.2020.12.001 [PubMed: 33306959]
26. Mendels DA, Dortet L, Emeraud C, et al. Using artificial intelligence to improve COVID-19 rapid diagnostic test result interpretation. Proc Nat Acad Sci. 2021;118(12):e2019893118. doi:10.1073/pnas.2019893118 [PubMed: 33674422]
27. Ning B, Yu T, Zhang S, et al. A smartphone-read ultrasensitive and quantitative saliva test for COVID-19. Sci Adv. 2021;7(2):eabe3703. doi:10.1126/sciadv.abe3703
28. Zhang T, Deng R, Wang Y, et al. A paper-based assay for the colorimetric detection of SARS-CoV-2 variants at single-nucleotide resolution. Nat Biomed Eng. 2022;6(8):957–967. doi:10.1038/s41551-022-00907-0 [PubMed: 35835993]
29. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition. CVPR, Las Vegas, NV. 2016;779–788. doi:10.1109/CVPR.2016.91
30. Jocher G, Chaurasia A, Qiu J. Ultralytics YOLOv8. 2023. Accessed July 1, 2025. https://github.com/ultralytics/ultralytics
31. Jocher G, Qiu J. Ultralytics YOLO11. 2024. Accessed October 1, 2024. https://github.com/ultralytics/ultralytics
32. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–444. doi:10.1038/nature14539 [PubMed: 26017442]
33. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale arXiv preprint arXiv: 2010.11929. 2020.
34. Shin HC, Roth HR, Gao M, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging. 2016;35(5):1285–1298. doi:10.1109/TMI.2016.2528162 [PubMed: 26886976]
35. Deng J, Dong W, Socher R, Li LJ, Li K, Li FF. Imagenet: A Large-Scale Hierarchical Image Database. Presented at: 2009 IEEE Conference on Computer Vision and Pattern Recognition; June 20–25, 2009; Miami, FL.
36. Wang CY, Yeh IH, Mark Liao HY. Yolov9: learning what you want to learn using programmable gradient information. In Leonardis A, Ricci E, Roth S, Russakovsky O, Sattler T, Varol G. (eds) Computer Vision – ECCV 2024. Lecture Notes in Computer Science, vol 15089. Springer, Cham. 2024. doi:10.1007/978-3-031-72751-1_1
37. Wang A, Chen H, Liu L, et al. Yolov10: real-time end-to-end object detection. ArXiv, abs/2405.14458. 2024.
38. Land KJ, Boeras DI, Chen XS, Ramsay AR, Peeling RW. REASSURED diagnostics to inform disease control strategies, strengthen health systems and improve patient outcomes. Nat Microbiol. 2019;4(1):46–54. doi:10.1038/s41564-018-0295-3 [PubMed: 30546093]
39. Sharma A, Kumar V, Longchamps L. Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and faster R-CNN models for detection of multiple weed species. Smart Agric Technol. 2024;9:100648. doi:10.1016/j.atech.2024.100648
40. Zurita M, Martignette L, Barrera J, et al. Detection of African swine fever virus utilizing the portable MatMaCorp ASF detection system. Transbound Emerg Dis. 2022;69(5):2600–2608. doi:10.1111/tbed.14411 [PubMed: 34871471]
41. Vu H, McVey DS. Recent progress on gene-deleted live-attenuated African swine fever virus vaccines. NPJ Vaccines. 2024;9(1):60. doi:10.1038/s41541-024-00845-9 [PubMed: 38480758]

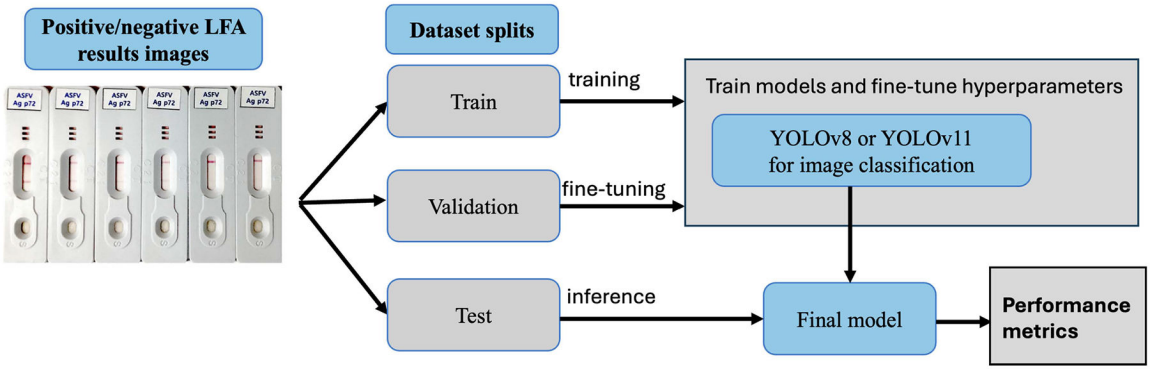

Figure 1—.

Workflow for model training and evaluation. LFA = Lateral flow assay. YOLOv8 = You Only Look Once, version 8. YOLOv11 = You Only Look Once, version 11.

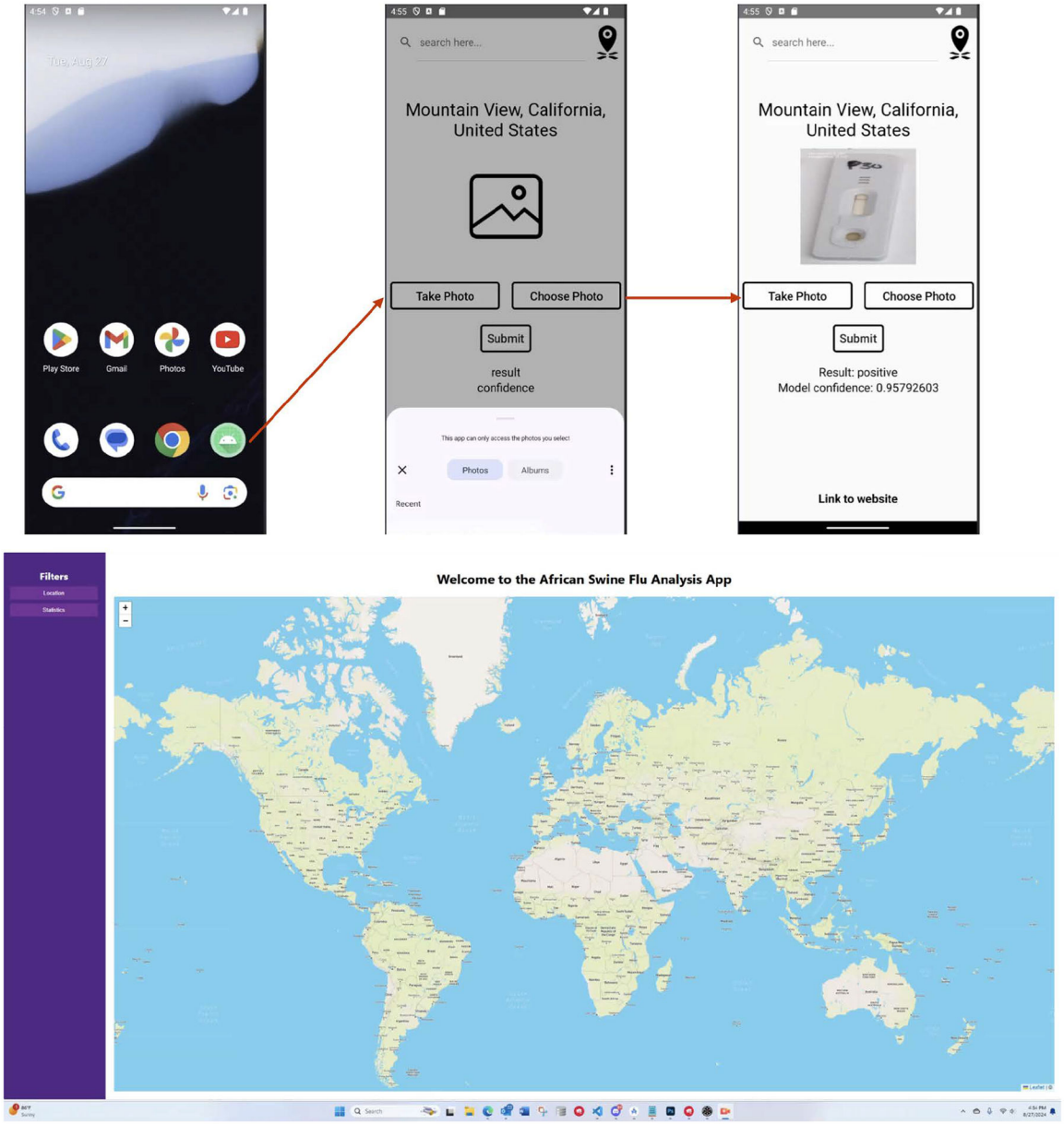

Figure 2—.
Workflow of the smartphone application.

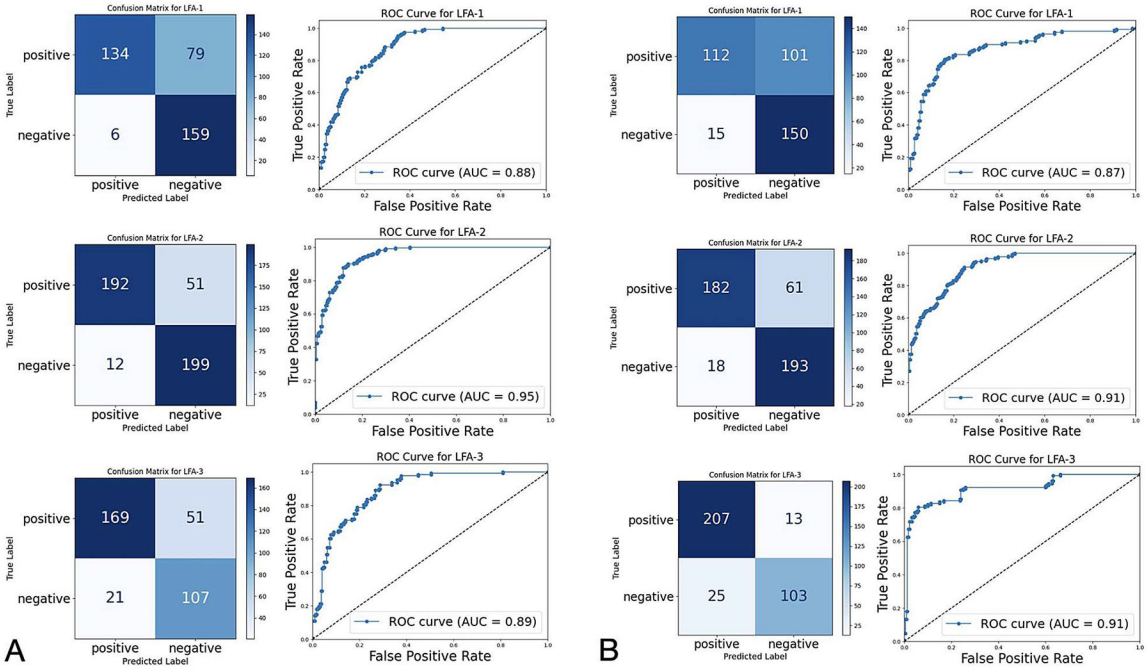

Figure 3—.

Performance analysis of YOLOv8 (A) and YOLOv11 (B). For each model, the column on the left shows confusion matrices on the test data for the 3 splits considered in our experiments (LFA-1, LFA-2, and LFA-3), whereas the column on the right shows the receiver operating characteristic curves (ROCs) for the 3 splits and their corresponding area under the curve (AUC) values.

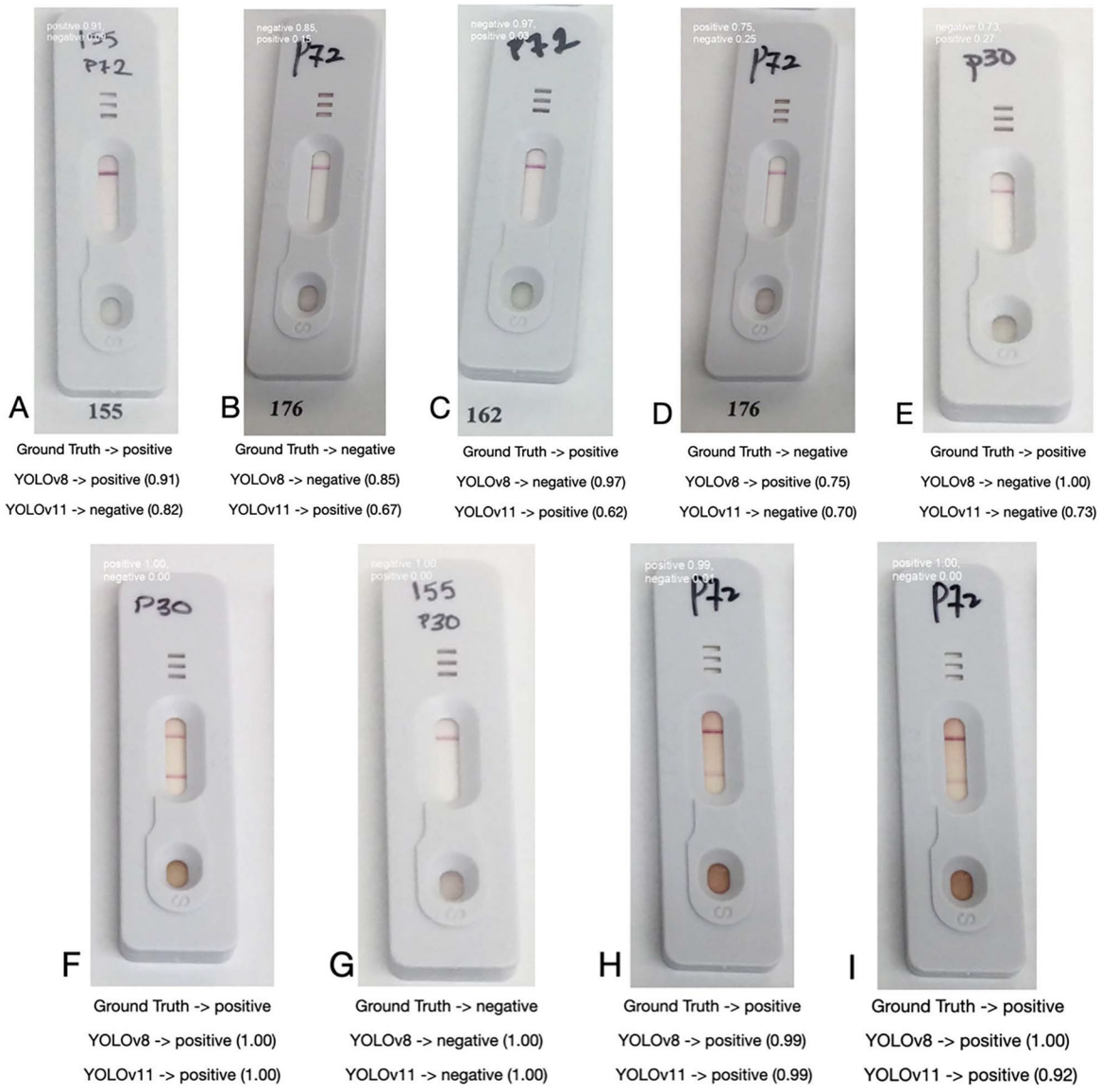

Figure 4—.

Example of images classified by YOLOv8 and YOLOv11. Ground truth outcome is shown right below an image, followed by the model prediction together with model confidence in parentheses.

Table 1—

Dataset statistics for 3 splits (lateral flow assay [LFA]-1, LFA-2, and LFA-3) of the data into train (20 subjects), development (3 subjects), and test (7 subjects) subsets.

		Train		Test		Development				
LFA type	**Split**	**Negative**	**Positive**	**Negative**	**Positive**	**Negative**	**Positive**	**Total negative in whole dataset**	**Total positive in whole dataset**	**Total**
P30	LFA-1	363	143	111	78	88	0	562	221	783
	LFA-2	286	170	176	51	100	0	562	221	783
	LFA-3	438	123	91	83	33	15	562	221	783
P72	LFA-1	104	403	54	135	43	45	201	583	784
	LFA-2	144	312	35	192	22	79	201	583	784
	LFA-3	164	398	37	137	0	48	201	583	784
P30 and P72	LFA-1	467	546	165	213	131	45	763	804	1,567
	LFA-2	430	482	211	243	122	79	763	804	1,567
	LFA-3	602	521	128	220	33	63	763	804	1,567

Statistics are shown separately for P30 and P72 tests and also for the combined P30 and P72 tests.

Am J Vet Res. Author manuscript; available in PMC 2025 March 31.

Table 2—

Performance metrics for the 3 splits (LFA-1, LFA-2, LFA-3) as well as average (AVG) and 95% CIs over the 3 splits.

Dataset split	Model	Accuracy	Precision	Recall/sensitivity	F1 score	AUC	Specificity
LFA-1	YOLOv8	77.5%	95.7%	62.9%	0.76	0.88	96.4%
LFA-1	YOLOv11	69.3%	88.2%	52.6%	0.66	0.87	90.6%
LFA-2	YOLOv8	86.1%	94.1%	79.0%	0.86	0.95	94.3%
LFA-2	YOLOv11	82.6%	91.0%	74.9%	0.82	0.91	91.5%
LFA-3	YOLOv8	79.3%	88.9%	76.8%	0.82	0.89	83.6%
LFA-3	YOLOv11	89.1%	89.2%	94.1%	0.91	0.91	80.5%
AVG and 95% CI	YOLOv8	80.9 ± 5.13%	92.9 ± 4.02%	72.9 ± 9.88%	0.81 ± 0.06	0.91 ± 0.04	91.4 ± 7.77%
AVG and 95% CI	YOLOv11	80.3 ± 11.42%	89.5 ± 1.61%	74.0 ± 23.33%	0.80 ± 0.14	0.90 ± 0.03	87.5 ± 6.91%

AUC = Area under the curve. YOLO = You Only Look Once.