

Android Based Preliminary Diagnosis of Foot and Mouth Disease in Local Cattle Using Forward Chaining and MobileNetV2

Imam Dwi Arifin¹, Putri Elfa Mas'udia², Sri Wahyuni Dali³

^{1,3} Digital Telecommunication Network Study Program, Department of Electrical Engineering, State Polytechnic of Malang, 65141, Indonesia

² Telecommunication Engineering Study Program, Department of Electrical Engineering, State Polytechnic of Malang, 65141, Indonesia.

imamdwi55555@gmail.com, putri.elfa@polinema.ac.id, sri.wahyuni@polinema.ac.id

Abstract— Foot and Mouth Disease is a highly contagious livestock disease that requires rapid preliminary assessment before veterinary confirmation. This study developed an Android application that combines a Forward Chaining expert system with MobileNetV2 image classification for early assessment of local cattle. The expert system processes seven observable symptoms, whereas the image module classifies mouth and hoof images into four categories: healthy mouth, healthy hoof, diseased mouth, and diseased hoof. The final experiment used 8212 training images, 1528 validation images, and 1278 testing images. The trained model was converted to a five megabyte TensorFlow Lite file for on device inference. The reported overall image classification accuracy was 91.86 percent, with macro precision, recall, and F1 score of 88.77 percent, 89.50 percent, and 89.13 percent. The Forward Chaining module correctly classified 14 of 15 cases, producing 93.3 percent accuracy. All eighteen tested farmer and veterinarian functions operated successfully on Android devices, and nine users produced an average usability score of 89.78 percent. The integrated application can support faster field screening while retaining veterinary consultation as the final clinical authority.

Keywords— Android application, convolutional neural network, expert system, Foot and Mouth Disease, Forward Chaining, MobileNetV2.

I. INTRODUCTION

Foot and Mouth Disease (FMD) is an acute viral infection of cloven hoofed animals and remains a major concern for cattle production because transmission can occur rapidly through direct contact, contaminated equipment, feed, vehicles, aerosols, and human movement. Clinical signs commonly include fever, excessive salivation, vesicles or erosions around the mouth, lesions near the hoof, reduced appetite, lameness, and behavioral changes. The operational consequence for farmers is not limited to animal health: delayed recognition can extend the period in which infected animals remain in contact with a herd and can increase production and treatment losses. Recent cattle health monitoring research has therefore emphasized earlier observation and faster delivery of field information to animal health personnel [1]. Conventional and molecular laboratory procedures remain essential for confirmation, but their equipment, time, and specialist requirements make them unsuitable as the first action for many smallholders [2].

Indonesia experienced renewed FMD outbreaks after a long disease free period, making rapid field screening relevant for local cattle farms. Risk after an outbreak is influenced by animal movement, farm contact, vaccination coverage, and biosecurity practices. In rural settings, however, a veterinarian may not always be present when symptoms are first noticed. Farmers often rely on visual observation and personal experience, which can produce inconsistent decisions when

lesions are mild, lighting is poor, or several symptoms overlap with other conditions. A mobile application cannot replace a laboratory or veterinary examination, but it can structure the initial observation, preserve evidence, and accelerate communication with a professional.

Previous information systems have approached this problem from different directions. Naive Bayes has been used to classify FMD from symptom data [3], while rule based expert systems have supported cattle disease diagnosis and outbreak prevention [4]. Dempster Shafer reasoning has also been applied to represent diagnostic confidence from combinations of symptoms [5]-[8]. Forward Chaining is attractive for field use because each user response activates explicit rules that can be inspected and revised by an expert; a similar inference strategy has shown practical value in symptom based health applications [9]. These methods, however, depend on the accuracy of manual symptom selection and do not directly evaluate lesion appearance.

Image classification provides complementary evidence. Convolutional Neural Networks have been implemented with TensorFlow for disease related visual recognition [6], [7], and comparative studies confirm that CNN architectures can learn discriminative features without manually defining color, texture, or shape descriptors [10]-[13]. A mobile application requires an architecture that balances predictive performance with storage and computation. MobileNetV2 employs depthwise separable convolution, inverted residual blocks, and linear bottlenecks to reduce computation while retaining

representation capacity [8]. TensorFlow also enables a trained model to be converted into a lightweight format that can execute on Android hardware [14]-[15].

The gap addressed in this study is the absence of an integrated workflow that combines symptom reasoning, image based classification, role based mobile access, and remote veterinary follow up in one Android application. The proposed system contributes four elements: first, a Forward Chaining knowledge base for seven observable symptoms; second, a four class MobileNetV2 model for mouth and hoof images; third, on device TensorFlow Lite inference connected to Firebase for user, history, and consultation data; and fourth, quantitative evaluation of classification, expert system accuracy, device compatibility, functional operation, and usability. The objective is to determine whether the integrated application can provide a reliable preliminary assessment that is practical for farmers while keeping the veterinarian responsible for final diagnosis.

II. METHOD

A. Research Workflow

The research was organized as an iterative engineering workflow rather than as a single model training activity. The first stage reviewed FMD characteristics, mobile decision support, expert systems, and image classification to establish the information required by farmers and veterinarians. The second stage translated those requirements into an application architecture, a symptom knowledge base, four image labels, user roles, and a cloud data structure. The third stage curated images and developed the MobileNetV2 classifier, followed by conversion to TensorFlow Lite. The fourth stage integrated local inference, Forward Chaining, Firebase services, and consultation functions into the Android application. The final stage evaluated model performance, rule accuracy, device compatibility, functional behavior, image acquisition routes, and user acceptance.

The workflow included a revision loop. A failed functional test required correction of the corresponding application module, while weak classification results required inspection of labels, image quality, and class representation. This separation was important because a high model score does not guarantee that camera permission, result storage, authentication, or consultation will work correctly on a smartphone. Conversely, a fully functional interface cannot compensate for an image model or knowledge rule that produces an incorrect preliminary assessment. Figure 1 summarizes the stages and the feedback path used during development.

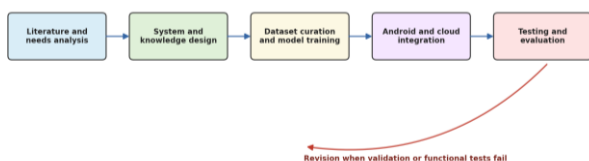


Figure 1. Iterative workflow for knowledge acquisition, model development, mobile integration, and system evaluation.

B. Research Design and Data Acquisition

The study followed a development and experimental evaluation design. Knowledge acquisition was conducted through observation and discussion with a veterinarian to identify clinical signs that could be recognized by farmers without invasive equipment. Seven signs were selected: mouth lesions, fever, reduced appetite, hoof lesions, lameness, weakness or lethargy, and behavioral change. Field photographs of healthy local cattle were collected as primary testing material. Because confirmed FMD cases were not available at the observation site during data collection, diseased image samples were obtained from a secondary dataset and screened for relevance, focus, visibility, and label consistency.

The final experimental split reported in the implementation stage contained 8212 training images, 1528 validation images, and 1278 testing images. Four target classes were used: FMD mouth, healthy mouth, FMD hoof, and healthy hoof. Training and validation images were obtained from the secondary collection, while the testing set combined secondary images with field images. This arrangement allowed the model to learn from a larger labeled collection and then be checked against photographs acquired in the target environment. Table I summarizes the class distribution available for training and validation.

TABLE I
TRAINING AND VALIDATION IMAGE DISTRIBUTION

Class	Training	Validation
FMD mouth	1593	487
Healthy mouth	2656	240
FMD hoof	1596	445
Healthy hoof	2367	356
Total	8212	1528

C. Data Curation and Class Definition

Each label was defined by both anatomical location and health status. The mouth classes contained images of oral tissue, lips, tongue, gums, or surrounding muzzle areas, whereas the hoof classes focused on the foot and coronary region. The FMD labels represented visible lesions or blister like abnormalities associated with the disease, while the healthy labels represented corresponding areas without those signs. Separating mouth and hoof images prevented the classifier from using anatomical location alone as the disease decision. A two class healthy versus diseased formulation could otherwise obtain a misleading result if most diseased photographs showed mouths and most healthy photographs showed hooves.

Image screening removed material that was irrelevant, unreadable, severely blurred, or inconsistent with the intended class. The training distribution was not perfectly balanced: healthy mouth was the largest training class and FMD mouth was the smallest. Validation data showed a different proportion, with FMD mouth and FMD hoof more prominent than in the training set. Figure 2 makes these differences visible. The imbalance was retained from the curated dataset and therefore became an important consideration when interpreting macro

averaged precision, recall, and F1 score rather than relying on accuracy alone.

Primary field images were used in testing to examine whether the model could process photographs captured in the intended environment. This introduced a practical domain difference because secondary images were generally more centered and visually clear, whereas farm images could include shadows, animal movement, irregular distance, and a complex background. The design therefore treated field testing as a generalization check and not merely as another random subset of the same image source.

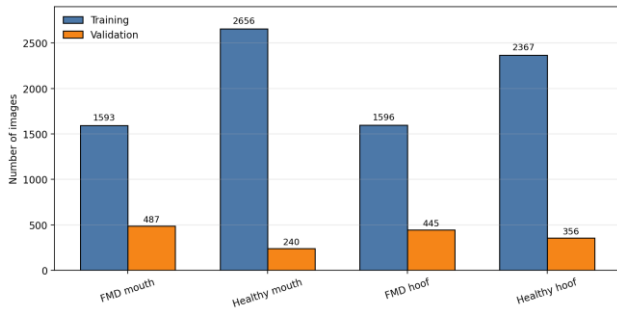


Figure 2. Training and validation image distribution for the four anatomical and health status classes.

D. MobileNetV2 Image Classification

Images were resized to 224 by 224 pixels with three color channels before inference. The classification network used MobileNetV2 as the feature extractor and a dense output layer with four neurons. The implementation report recorded 11169476 trainable parameters and trained the main configuration for 20 epochs. Accuracy and loss were monitored for both training and validation data. A ResNet50 configuration was trained under the same class definition as a comparative baseline. MobileNetV2 was selected for deployment because its mobile oriented architecture reduces computational cost relative to a conventional residual network [8].

After training, the model was converted from the development format into TensorFlow Lite. The resulting file, pmk_model.tflite, had an approximate size of five megabytes and was embedded in the Android project. During prediction, a user can capture a photograph or choose an existing image. The application prepares the bitmap, performs local inference, obtains probabilities for the four classes, and presents the class with the highest probability when it exceeds the application validation threshold. Local inference reduces dependence on network latency; the internet connection is mainly required for authentication, history synchronization, and consultation.

Performance was evaluated with a confusion matrix and class wise accuracy, precision, recall, and F1 score. Precision describes how many images assigned to a class were correct, recall describes how many images from an actual class were recovered, and F1 score balances both quantities. The implementation also compared prediction confidence from gallery images and direct camera images because focus, illumination, distance, and motion blur can differ between these acquisition routes.

The deployed inference sequence consisted of loading the selected bitmap, resizing it to the model input dimensions, arranging the pixel values in the tensor expected by the converted model, executing the interpreter, and mapping the four output values to the class names used by the interface. The application then selected the maximum output and compared it with the 50 percent validation threshold documented in the implementation. A prediction below the threshold was not intended to be treated as a reliable disease decision. When the output passed validation, the class name, confidence, image reference, user identity, and time information could be stored as part of the examination history.

E. Forward Chaining Expert System

The expert system begins with facts supplied by the user and moves forward through rules until a status is satisfied. Each symptom is stored as a Boolean fact. Five output states were defined: confirmed indication of FMD in the rule base, suspected FMD, suspected ordinary fever, suspected ordinary illness, and healthy. The term confirmed indication refers to the internal rule outcome and not to laboratory confirmation. A healthy result is generated when none of the seven symptoms is selected.

The rule base was constructed from veterinarian input and documented clinical characteristics. The most complete FMD rule requires all seven symptoms. Suspected FMD is produced by combinations dominated by mouth lesions, fever, reduced appetite, and hoof lesions. Weakness and behavioral change without the principal lesion pattern lead to non FMD illness categories. Representative rules are shown in Table II. The rule engine was tested by entering symptom combinations and comparing the resulting code with the expected expert decision.

TABLE II
REPRESENTATIVE FORWARD CHAINING RULES

Condition	Output
G001 and G002 and G003 and G004 and G005 and G006 and G007	K001 FMD indication
G001 and G002 and G003 and G004	K002 suspected FMD
G001 and G006 and G007	K003 ordinary fever
G006 or G007	K004 ordinary illness
No symptom selected	K005 healthy

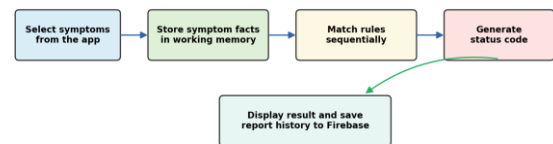


Figure 3. Data driven inference from selected symptoms to a preliminary status and stored result.

A traceable example illustrates the reasoning mechanism. When G001, G002, G003, and G004 are true, the working memory contains evidence of a mouth lesion, fever, reduced appetite, and a hoof lesion. The engine scans the rule conditions, matches the combination assigned to K002, and produces a suspected FMD status. If all seven facts are present, the more complete K001 rule is satisfied. When only lethargy or

behavioral change is present, the facts are insufficient for the lesion dominated FMD rules and the engine directs the case to an ordinary illness status. This traceability allows a veterinarian or developer to identify exactly which symptom pattern generated the displayed result.

Rule order and symptom completeness are important. The application must evaluate the more specific combinations before a broader rule that contains only one or two facts; otherwise a partial conclusion could be produced too early. The implementation also distinguishes the internal K001 wording from clinical confirmation. Even when the most complete rule fires, the output remains a preliminary indication that should trigger isolation precautions and professional review rather than an autonomous final diagnosis.

F. Android and Cloud Integration

The application was developed in Android Studio using separate access paths for farmers, veterinarians, and administrators. Android was selected because it supports portable image acquisition and can host the TensorFlow Lite model directly; Android based educational and service applications have also demonstrated broad implementation flexibility [12]. The farmer interface includes registration, login, articles, camera or gallery selection, image prediction, expert system questions, result storage, history, doctor selection, and consultation chat. The veterinarian interface receives reports, reviews histories, communicates with farmers, and views summary statistics. The administrator manages accounts, reports, payments, and system status.

Firebase Authentication, Realtime Database, and Cloud Firestore provide identity management and synchronized storage. Similar mobile health applications use Firebase to reduce backend development overhead and support near real time client updates [15]. A payment gateway is included for optional paid consultation, but the diagnostic modules can be used independently. Figure 4 presents the implemented logical architecture.

Data were separated according to function. Authentication records controlled access, profile records identified the user role, prediction records stored image classification results, symptom records preserved the expert system input and output, consultation records connected a farmer with a veterinarian, and transaction records supported the optional consultation payment process. This organization allowed the image model to remain local while the evidence and communication history were synchronized. The veterinarian could therefore review the farmer supplied image, symptom result, and previous records without requiring the model itself to execute on the veterinarian account.

G. Testing Procedure

Five groups of tests were conducted. First, the image model was evaluated on the testing set and compared with ResNet50. Second, the Forward Chaining module was tested with 15 cases representing four nonhealthy rule codes; the healthy code was deterministic when all symptoms were absent. Third, installation and functional tests were performed on four smartphones using Android 10, Android 12, and Android 15.

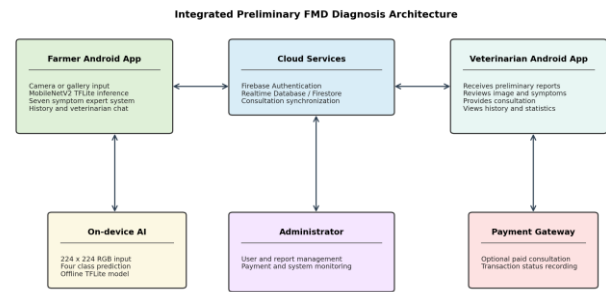


Figure 4. Integrated Android, on device inference, cloud storage, and consultation architecture.

Thirteen farmer functions and five veterinarian functions were checked. Fourth, predictions produced from gallery files and direct camera acquisition were compared with development environment results. Fifth, usability was measured with five Likert scale questions completed by nine users. The questions covered ease of use, interface attractiveness, operational smoothness, feature function, and confidence in the diagnostic output.

An infrared mobile thermometer was also examined as supporting information. Smartphone connected infrared thermometers can differ from conventional noncontact instruments and should be interpreted according to measurement site and procedure [11]. In the field trial, head surface readings were lower than normal rectal temperature ranges; therefore, temperature was not used as the principal decision variable. The final preliminary assessment relied primarily on symptoms and image analysis.

H. Evaluation Measures

The confusion matrix was used to identify not only how many predictions were correct, but also which classes were exchanged. For each class, precision was calculated from correct positive predictions relative to all images predicted as that class, recall from correct positive predictions relative to all actual images in that class, and F1 score from the harmonic balance of precision and recall. Macro values were obtained by averaging the four class results so that a large class could not dominate the evaluation. The article reports the overall confusion matrix accuracy and the class wise mean accuracy separately because they answer different questions.

Forward Chaining accuracy was calculated as the number of rule outputs equal to the expected expert result divided by the 15 cases. The tested case distribution comprised one K001 case, four K002 cases, three K003 cases, and seven K004 cases. K005 was not assigned a separate test combination because it is generated deterministically when no symptom is selected. Functional performance was reported as passed checks rather than as a prediction metric, while usability was calculated by dividing the weighted Likert score by the maximum possible score for nine respondents.

TABLE III
EXPERT SYSTEM TEST CASE DISTRIBUTION

Code	Cases	Interpretation
K001	1	Complete FMD indication rule
K002	4	Suspected FMD combinations
K003	3	Ordinary fever combinations
K004	7	Ordinary illness combinations
K005	0	No symptom selected; deterministic

III. RESULTS AND DISCUSSION

I. Image Classification Performance

The MobileNetV2 model correctly recognized most mouth and hoof images. The study reported an overall testing accuracy of 91.86 percent. Macro precision, recall, and F1 score were 88.77 percent, 89.50 percent, and 89.13 percent. The normalized confusion pattern in Figure 5 shows a clear difference between mouth and hoof categories. FMD mouth and healthy mouth reached recalls of 97.47 percent and 98.59 percent, whereas FMD hoof and healthy hoof reached 78.92 percent and 83.03 percent. This pattern indicates that oral lesions were more consistently represented in the available images than hoof lesions.

The largest confusion occurred between FMD hoof and healthy hoof. Approximately 21.08 percent of FMD hoof images were assigned to healthy hoof, while 16.97 percent of healthy hoof images were assigned to FMD hoof. Hoof photographs can contain dirt, shadows, variable viewing angles, and weak lesion boundaries. These factors create visual similarities that are less dominant in close mouth photographs. The result supports additional data collection focused on early hoof lesions, different cattle breeds, varied ground conditions, and controlled close range acquisition.

Table IV reports the class wise metrics. The very high precision for FMD mouth means that a mouth image classified as diseased was rarely a false positive in the test set. In contrast, the lower precision and recall for FMD hoof indicate a need for more balanced field data. The class wise mean accuracy of 95.81 percent reported in the thesis is a one versus rest average and is different from the global confusion matrix accuracy of 91.86 percent; both values are retained here according to their respective definitions.

TABLE IV
CLASS WISE PERFORMANCE OF MOBILENETV2

Class	Accuracy	Precision	Recall	F1
FMD hoof	92.80	76.67	78.92	77.78
FMD mouth	98.59	99.48	97.47	98.47
Healthy mouth	99.06	95.89	98.59	97.22
Healthy hoof	92.80	83.03	83.03	83.03
Mean	95.81	88.77	89.50	89.13

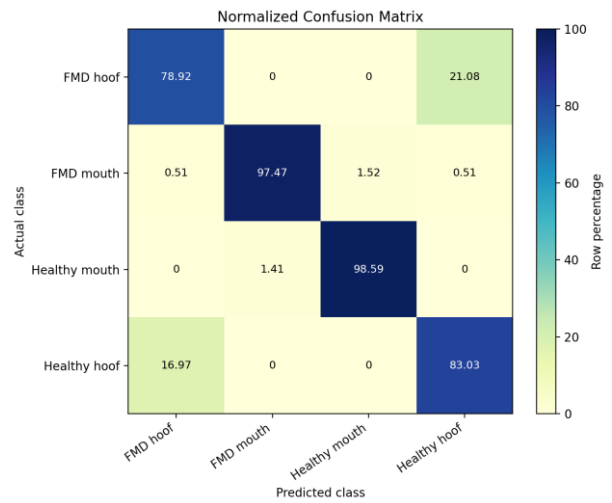


Figure 5. Normalized confusion matrix for the four cattle image classes.

J. Model Selection and Mobile Deployment

The comparison experiment showed that MobileNetV2 and ResNet50 converged to similar final training and validation accuracy, approximately 0.82 and 0.85. MobileNetV2 started with higher training accuracy, while ResNet50 started with higher validation accuracy. Across the reported evaluation metrics, MobileNetV2 achieved a class wise mean accuracy of 95.81 percent, compared with 94.96 percent for ResNet50. It also exceeded ResNet50 in precision, recall, and F1 score by 1.56, 1.46, and 1.51 percentage points, respectively.

The difference is modest in predictive terms, but deployment cost is important for an application intended for ordinary smartphones. MobileNetV2 was designed for resource constrained devices through inverted residuals and efficient depthwise operations [8]. Converting the selected model to a five megabyte TensorFlow Lite file enabled offline classification without transmitting every image to a server. This reduces network dependence and avoids additional upload delay during the initial screening step. Table V summarizes the reported comparison.

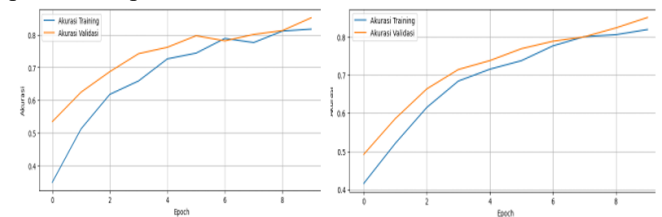


Figure 6. Training and validation accuracy trends for ResNet50 and MobileNetV2 during the comparison experiment.

The learning curves in Figure 6 support the numerical comparison. ResNet50 rose rapidly from a lower training starting point and obtained a higher initial validation score. MobileNetV2 began with higher training accuracy and then improved more gradually. By the last comparison epoch, both training curves were near 0.82 and both validation curves were near 0.85. The absence of a large final gap between training and validation suggests that neither model displayed severe overfitting in this limited experiment. Model selection

therefore depended not only on final validation accuracy but also on class wise metrics and the need for efficient smartphone execution.

TABLE V
MOBILENETV2 AND RESNET50 COMPARISON

Model	Accuracy	Precision	Recall	F1
MobileNetV2	95.81	88.77	89.50	89.13
ResNet50	94.96	87.21	88.04	87.62

K. Application and Expert System Evaluation

The Forward Chaining module produced 14 correct decisions from 15 test cases, corresponding to 93.3 percent accuracy. The rule base therefore provided a stable second source of evidence when users supplied the expected symptom combinations. This result is consistent with the transparent nature of rule based systems: when the fact pattern matches a documented rule, the conclusion can be reproduced and explained. It improves upon symptom only approaches by being paired with image evidence, but it remains sensitive to incorrect user observations. The application consequently presents the output as a preliminary status and provides a path to veterinary consultation.

Four smartphones successfully installed and opened the application. The devices represented Android 10, Android 12, and Android 15 with memory capacities from three to six gigabytes. All 13 farmer functions and all five veterinarian functions completed successfully, giving 18 successful functional checks. Tested functions included authentication, camera permission, image selection, prediction validation, symptom input, result storage, history deletion, doctor consultation, and chat. The interface used for selecting a source image and entering the prediction flow is illustrated in Figure 8.

Gallery images produced confidence values from 60 to 72 percent, whereas direct camera images produced values from 53 to 57 percent. The difference across paired trials ranged from 7 to 17 percentage points. Nevertheless, the predicted classes remained consistent with the reference for the tested FMD mouth and FMD hoof images, and the probability difference between the Android implementation and the development environment ranged from 0.67 to 5.67 percentage points. The lower camera confidence is plausibly related to variable lighting, distance, viewing angle, background complexity, and motion blur. The result suggests that the application should provide acquisition guidance, such as adequate lighting, a stable close distance, and a clear view of the lesion area.

TABLE VI
INSTALLATION AND FIELD PREDICTION DEVICES

Device	Android	Memory	Field Result
Redmi 8	10	3 GB	Healthy mouth, 54%
Vivo Y16	12	3 GB	Healthy mouth, 57%
Realme C71	15	6 GB	Healthy mouth, 51%
Vivo Y12s 2021	10	3 GB	Healthy mouth, 50%

The four field photographs used in the installation test were all labeled healthy mouth and were correctly classified with

confidence values from 50 to 57 percent. These values were only slightly above the application threshold. The result demonstrates compatibility across the tested operating system versions, but it also shows that a correct class label can be accompanied by modest confidence when the mouth is not fully facing the camera, shadows are present, or the cattle pen background occupies a substantial part of the frame. Device compatibility should therefore not be interpreted as evidence that every camera condition is equally suitable for diagnosis.

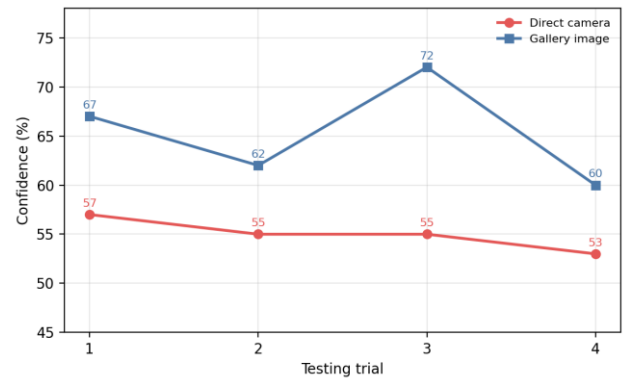


Figure 7. Confidence comparison between gallery uploads and direct smartphone camera images in four paired trials.

Figure 7 shows the paired acquisition results in detail. Gallery confidence values were 67, 62, 72, and 60 percent, while direct camera values were 57, 55, 55, and 53 percent. The corresponding differences were 10, 7, 17, and 7 percentage points, with mean confidence of 65.25 percent for gallery images and 55.00 percent for camera images. The third trial produced the largest gap, indicating that acquisition variability can affect confidence even when the predicted class remains unchanged. A future interface should display a framing guide, warn about blur, and ask the user to retake an image when the target area is too small or poorly illuminated.

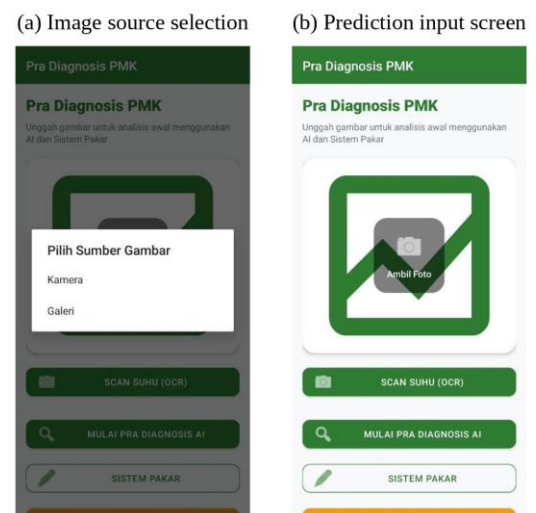


Figure 8. Android interface for selecting an image source and starting preliminary diagnosis.

L. Usability and Practical Implications

Nine respondents completed the five item usability questionnaire. The overall mean was 89.78 percent, which falls in the very positive category used in the study. Interface attractiveness received the highest score at 95.56 percent, followed by diagnostic output at 93.33 percent and operational smoothness at 91.11 percent. Feature function scored 86.67 percent, while ease of use received the lowest score at 82.22 percent. The distribution is shown in Figure 9.

The relatively lower ease score indicates that first time users may need clearer instructions for image capture, symptom selection, and the difference between preliminary screening and definitive diagnosis. This is particularly important because confidence values are not probabilities of disease prevalence; they represent the model preference among the four trained classes. Explanatory text should therefore avoid overstating certainty and should recommend veterinary review when FMD is suspected.

The integrated design has practical advantages over an isolated classifier or a standalone expert system. Image inference supplies visual evidence, the rule engine structures clinical observations, Firebase preserves a time stamped history, and consultation tools connect the farmer with a veterinarian. At the same time, the evaluation has limitations. Diseased images were dominated by secondary data, the primary field sample was small and mainly healthy, the expert system test contained only 15 cases, and usability involved nine respondents. External validation with confirmed clinical cases, multiple farms, and independent veterinarians is required before the application can be considered for operational disease surveillance.

The complementary nature of the two inference paths is central to the design. An image can reveal a visible lesion, but cannot reliably represent fever, appetite, gait, or behavioral change. The expert system can process those clinical observations, but depends on the farmer selecting symptoms accurately. Presenting both outputs gives the veterinarian more context than either module alone. A disagreement should not be hidden by automatically averaging scores; it should instead be displayed as a reason to repeat image acquisition, review the selected symptoms, and request professional assessment.

From a disease control perspective, the practical benefit is the reduction of communication delay. A farmer can preserve the visual evidence and structured symptoms immediately, while the veterinarian can inspect the record remotely and determine whether an on site examination or laboratory confirmation is warranted. The application must nevertheless protect against overconfidence. The model recognizes patterns within four trained image classes and does not exclude other diseases, trauma, dirt, or non FMD causes of oral and hoof abnormalities. Clear warnings, secure data handling, and a referral pathway are therefore part of system quality rather than optional interface additions.

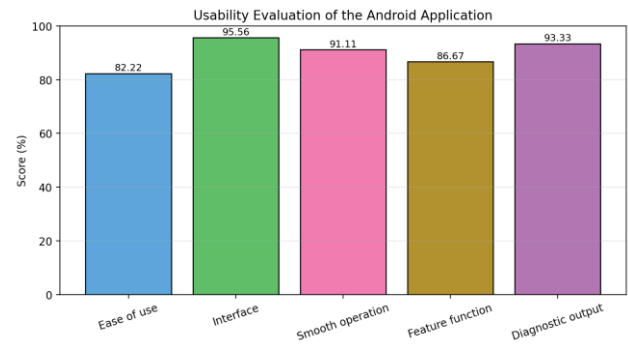


Figure 9. Usability scores obtained from nine respondents.

TABLE VII
SUMMARY OF SYSTEM VALIDATION RESULTS

Test	Result	Interpretation
Image classification	91.86% overall accuracy	Strong mouth classification; hoof confusion remains
Expert system	14 of 15 correct	93.3% rule based accuracy
Installation	4 of 4 devices	Android 10, 12, and 15 supported
Functional checks	18 of 18 passed	Farmer and veterinarian flows operational
Usability	89.78% mean	Positive acceptance among nine users
Camera vs gallery	7 to 17 point gap	Acquisition quality affects confidence

M. Deployment Implications and Error Sources

The combined evaluation indicates that the application is strongest when it is used as a structured screening tool rather than as an autonomous diagnostic engine. The image classifier achieved 91.86 percent accuracy, the expert system reached 93.3 percent accuracy, and all functional tests passed. These results show that the architecture is technically workable for preliminary field use. Nevertheless, the error pattern remains meaningful. Misclassification was concentrated in the hoof classes, where healthy hoof and FMD hoof images shared more visual similarity than the mouth classes. This suggests that lesion texture, dirt, angle, and background complexity affect discriminative power more strongly for hoof images than for oral lesions.

Field acquisition conditions also explain the difference between gallery images and direct camera images. Gallery images tended to be more carefully framed and stable before upload, while direct camera capture was more sensitive to blur, lighting variation, subject movement, and inconsistent distance. As a result, gallery confidence values ranged from 60 to 72 percent, whereas direct camera confidence values ranged from 53 to 57 percent. In practical deployment, this gap means that the application should guide the user to repeat image capture whenever focus is weak, the lesion area occupies only a small part of the frame, or illumination is uneven. Such guidance can improve reliability without changing the trained model.

Figure 10 summarizes the major validation results, and Table VIII converts the findings into operational recommendations for farmers and veterinarians. The figure shows that system quality is not represented by one metric alone. A useful field application must simultaneously maintain acceptable image classification, interpretable symptom reasoning, successful device execution, and satisfactory user

experience. The table emphasizes that the application is most effective when farmers treat the result as an early alert, capture a clear image, complete symptom selection honestly, and immediately contact a veterinarian when an FMD indication appears. These recommendations reduce the risk that a technically correct prediction is misunderstood or acted on without professional review.

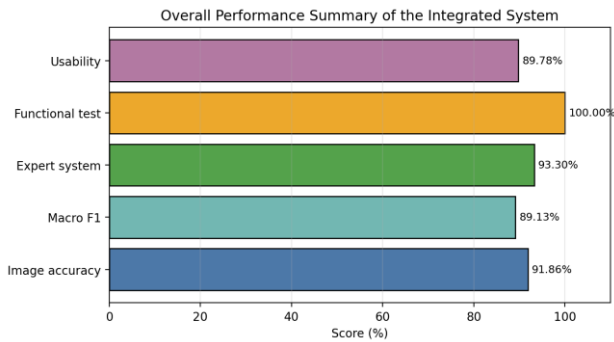


Figure 10. Overall performance summary combining image classification, rule based inference, functional operation, and usability evaluation.

TABLE VIII
OPERATIONAL RECOMMENDATIONS FOR FIELD SCREENING

Aspect	Observed issue	Recommended action
Image capture	Blur, shadow, or lesion too small	Retake the image under brighter light and fill more of the frame with the lesion area
Prediction confidence	Camera confidence lower than gallery confidence	Allow users to review the image first and recapture when focus is poor
Symptom selection	Users may omit relevant symptoms	Provide short symptom descriptions and veterinarian oriented notes in the interface
Decision making	Preliminary result may be mistaken for final diagnosis	Display warnings that laboratory or veterinary confirmation is still required
Record keeping	Loss of historical evidence may slow follow up	Store time stamped images, symptoms, and consultation responses in Firebase history

N. Detailed Quantitative Findings

The class wise comparison provides a more specific interpretation than the overall accuracy alone. Figure 11 shows that FMD mouth and healthy mouth produced precision, recall, and F1 scores above 95 percent, while both hoof categories remained below 84 percent for the same metrics. FMD hoof obtained the lowest F1 score at 77.78 percent. This result confirms that the principal improvement target is not the entire model uniformly, but the representation of hoof conditions. Additional data should emphasize early lesions, variable hoof cleanliness, partial occlusion, different floor surfaces, and consistent close range framing so that the model can separate disease patterns from environmental texture.

The study also calculated average confidence from 12 secondary test images, with three images for each class. The highest mean was 72.67 percent for FMD mouth, followed by 60.00 percent for healthy mouth, 58.67 percent for healthy hoof, and 57.00 percent for FMD hoof. These confidence values should not be interpreted as class accuracy. They describe the strength of the selected output for a small reference sample,

whereas accuracy measures whether the selected class was correct across the evaluation set. Table IX is therefore presented as an implementation reference for comparing Android outputs, not as a replacement for the confusion matrix and class wise metrics.

The application used a 50 percent validation threshold before a result was accepted by the interface. The four field images classified as healthy mouth produced confidence values between 50 and 57 percent, showing that correct predictions can remain close to the threshold under farm conditions. A fixed threshold protects against very weak outputs, but it does not identify blur, poor framing, or an unfamiliar disease. A more robust future implementation should combine the threshold with image quality checks and provide a retake instruction when the region of interest is unclear.

Functional validation was divided by user role. All 13 farmer functions and all five veterinarian functions passed, corresponding to 100 percent completion for the 18 tested scenarios. The farmer tests covered registration, authentication, camera permission, image selection, prediction, symptom entry, storage, history, consultation, and chat. The veterinarian tests covered access to reports, prediction related screens, communication, and history management. Table X summarizes this coverage. Functional success confirms that the workflow can be executed, but it does not remove the need for larger clinical validation of the prediction modules.

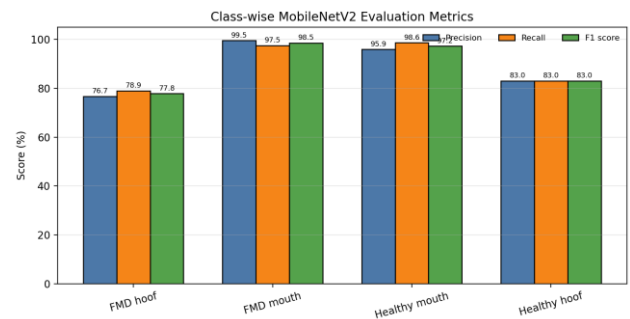


Figure 11. Class wise precision, recall, and F1 score of the MobileNetV2 classifier.

TABLE IX
MEAN CONFIDENCE OF THE SECONDARY IMAGE REFERENCE SET

Class	Images	Mean Confidence
FMD mouth	3	72.67%
FMD hoof	3	57.00%
Healthy mouth	3	60.00%
Healthy hoof	3	58.67%

TABLE X
ROLE BASED FUNCTIONAL TEST COVERAGE

User Role	Tested	Passed	Success
Farmer	13	13	100%
Veterinarian	5	5	100%
Total	18	18	100%

O. Representative Cattle Image Trial Results

Representative color images from the cattle trials are presented to connect the numerical evaluation with the actual

visual material processed by the application. The primary subset consisted of five field photographs of healthy local cattle mouths. All five images were classified as healthy mouth, with confidence values of 72, 61, 52, 54, and 72 percent. The mean confidence was 62.20 percent and the range was 52 to 72 percent. These results show that the application could recognize healthy mouth patterns under real farm backgrounds, although confidence decreased when the mouth region was less centered or partially affected by shadows.

The secondary diseased subset contained five images, comprising four FMD mouth samples and one FMD hoof sample. The corresponding confidence values were 60, 54, 63, 53, and 50 percent, producing a mean of 56.00 percent and a range of 50 to 63 percent. Every image in this representative subset was assigned to the same class as its documented condition. Figure 12 displays four examples in full color: two primary healthy mouth images, one FMD mouth image, and one FMD hoof image. Green borders indicate healthy results and red borders indicate FMD results.

The combined ten image subset therefore produced ten matching class decisions. This result should be interpreted as an illustrative application trial rather than a replacement for the full 1278 image test set and confusion matrix. The relatively modest confidence values, particularly for diseased hoof images, support the earlier finding that acquisition quality and hoof class variability remain important limitations. Clear framing, adequate illumination, and a close view of the anatomical region are required before the result is submitted for veterinary review.



Figure 12. Color examples of cattle image trials: primary healthy mouth images and secondary FMD mouth and hoof images with the corresponding predictions.

TABLE XI
REPRESENTATIVE CATTLE IMAGE TRIAL SUMMARY

Subset	Images	Correct	Confidence Range	Mean
Primary healthy mouth	5	5	52-72%	62.20%

Subset	Images	Correct	Confidence Range	Mean
Secondary FMD mouth/hoof	5	5	50-63%	56.00%
Combined representative set	10	10	50-72%	59.10%

IV. CONCLUSION

This study produced an Android based preliminary FMD assessment system that combines MobileNetV2 image classification, a Forward Chaining expert system, Firebase data services, and veterinary consultation. The reported image classification accuracy was 91.86 percent, with macro precision of 88.77 percent, recall of 89.50 percent, and F1 score of 89.13 percent. The symptom module achieved 93.3 percent accuracy on 15 cases, all 18 tested farmer and veterinarian functions operated successfully, and the usability evaluation reached 89.78 percent. MobileNetV2 was selected over ResNet50 because it provided slightly better reported metrics and could be converted into a five megabyte TensorFlow Lite model for on device operation. A representative ten image cattle trial also produced matching predictions for five primary healthy mouth images and five secondary FMD mouth or hoof images, while the full model evaluation remained the basis for the reported accuracy. The application is suitable as a screening and communication aid, not as a replacement for veterinary or laboratory confirmation. Future work should correct class imbalance in hoof images, enlarge the primary dataset with clinically confirmed cases, validate rules with multiple veterinarians, evaluate inference time and network quality on more devices, and conduct multi farm prospective testing.

ACKNOWLEDGEMENTS

The authors thank the Digital Telecommunication Network Study Program, State Polytechnic of Malang, the supervising lecturers, veterinarian Hadi Prayitno, and the farmers who supported field observation and application testing.

REFERENCES

- [1] H. Shahab, M. Iqbal, A. Sohaib, A. ur Rehman, A. Bermak, and K. Munir, "Design and implementation of an IoT based monitoring system for early detection of lumpy skin disease in cattle," *Smart Agricultural Technology*, vol. 9, p. 100609, 2024, doi: 10.1016/j.atech.2024.100609.
- [2] N. Longjam, R. Deb, A. K. Sarmah, T. Tayo, V. B. Awachat, and V. K. Saxena, "A brief review on diagnosis of Foot and Mouth Disease of livestock: Conventional to molecular tools," *Veterinary Medicine International*, vol. 2011, p. 905768, 2011, doi: 10.4061/2011/905768.
- [3] A. U. Bani and A. Asruddin, "Pendeteksian Penyakit Mulut dan Kuku pada sapi dengan menerapkan metode Naive Bayes," *Journal of Computer System and Informatics*, vol. 3, no. 4, pp. 264-268, 2022, doi: 10.47065/josyc.v3i4.1934.
- [4] M. R. Zamroni, Q. C. K. N. S., and A. Wahyudi, "Sistem pakar diagnosa penyakit sapi sebagai upaya pencegahan penyebaran wabah PMK di Lamongan," *Jurnal Ilmiah Informatika*, vol. 10, no. 2, pp. 145-152, 2022, doi: 10.33884/jif.v10i02.6373.
- [5] M. A. S. Padilla Savira and Novriyenni, "Diagnosis of cattle disease as an effort to prevent the spread of FMD using the Dempster Shafer method in Langkat district," *Journal of Information Technology*, vol. 2, pp. 85-94, 2023.
- [6] I. W. Prastika and E. Zuliarso, "Deteksi penyakit kulit wajah menggunakan TensorFlow dengan metode Convolutional Neural Network," *Jurnal Manajemen Informatika dan Sistem Informasi*, vol. 4, no. 2, pp. 84-91, 2021, doi: 10.36595/misi.v4i2.418.

- [7] A. J. Rozaqi, A. Sunyoto, and M. R. Arief, "Deteksi penyakit pada daun kentang menggunakan pengolahan citra dengan metode Convolutional Neural Network," *Creative Information Technology Journal*, vol. 8, no. 1, pp. 22-31, 2021, doi: 10.24076/citec.2021v8i1.263.
- [8] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4510-4520, 2018, doi: 10.1109/CVPR.2018.00474.
- [9] M. Sari, S. Defit, and G. W. Nurcahyo, "Sistem pakar deteksi penyakit pada anak menggunakan metode Forward Chaining," *Jurnal Sistim Informasi dan Teknologi*, vol. 2, pp. 130-135, 2020, doi: 10.37034/jsisfotek.v2i4.34.
- [10] A. Widyadhari, C. Basri, and E. Sudarnika, "High Risk Period kasus Penyakit Mulut dan Kuku pasca wabah pada peternakan sapi perah di Kabupaten Malang," *Jurnal Sain Veteriner*, vol. 42, no. 3, 2024, doi: 10.22146/jsv.91219.
- [11] M. I. Nepesov, D. Kilinc, and R. G. Sezer Yamanel, "Comparison of infrared smart mobile phone thermometer to non contact forehead infrared thermometer in pediatric emergency room patients," *Iranian Journal of Pediatrics*, vol. 34, no. 5, 2024, doi: 10.5812/ijp-144744.
- [12] M. A. Al Hawari Nasution, S. Siswanto, and E. Suryana, "Rancangan media pembelajaran berupa aplikasi Augmented Reality berbasis Android," *Jurnal Media Infotama*, vol. 19, no. 2, pp. 528-537, 2023, doi: 10.37676/jmi.v19i2.4771.
- [13] Y. N. Yenusi, S. Trihandaru, and A. Setiawan, "Comparison of Convolutional Neural Network models in face classification of Papuan and other ethnicities," *Jurnal Sains dan Teknologi*, vol. 12, no. 1, pp. 261-268, 2023, doi: 10.23887/jstundiksha.v12i1.46861.
- [14] M. Ihsan, R. K. Niswatin, and D. Swanjaya, "Deteksi ekspresi wajah menggunakan TensorFlow," *Joutica*, vol. 6, no. 1, p. 428, 2021, doi: 10.30736/jti.v6i1.554.
- [15] A. Fiorenza and H. Tolle, "Pengembangan aplikasi mobile sebagai media edukasi kesehatan gigi menggunakan teknologi Firebase serta metode prototyping," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 1, pp. 258-266, 2023.