

Machine Learning Image Classification model to Identify Cattle in Kenya

Benard Onyango¹, Obadiah Musau², Kennedy Ondimu³

Technical university of Mombasa, Kenya

DOI: <https://dx.doi.org/10.51244/IJRSI.2025.1208004131>

Received: 12 Sep 2024; Accepted: 19 Sep 2025; Published: 25 October 2025

ABSTRACT

Classifying cattle using muzzle images is an emerging technology in livestock management for recognition and classification. This study used Convolutional Neural Networks (CNN) algorithm to uniquely identify cattle by using their muzzle patterns which are unique to every single cattle. The study used a dataset of 4,923 muzzle images of different cattle breeds which were pre-processed to improve the dataset's performance and reduce overfitting. The Convolution Neural Network used several convolutional layers to capture muzzle patterns, pooling and dense layers to differentiate breeds. Adam optimizer and categorical cross-entropy loss were employed for model training. The results revealed high accuracy, verifying muzzle images as an effective biometric method for cattle identification. Transfer learning via pre-trained models positively impacted model accuracy and generalization. The technology can be integrated into livestock management and breeding programs, as well as agricultural and farming systems.

Keywords: Biometric Identification, Muzzle Images, Convolutional Neural Networks, Keras Framework.

INTRODUCTION

Cattle rustling refers to the unlawful theft of cattle, often involving violent and destructive methods. This practice not only result in loss of livestock but also leads to so many deaths, destruction of property, and displacement of families. It poses a serious threat to peace and security along transnational borders, undermining governance and stability. The root causes of cattle rustling are deeply rooted in social, cultural, economic, political, and historical disputes. According to a 2024 report by the Kenya Institute of Public Policy Research and Analysis (KIPPRA), Kenya's Parliamentary Committee on Administration and National Security highlights several drivers of banditry. These include competition for scarce resources such as water and pasture, widespread access to firearms, cycles of retaliatory attacks, political mobilization, land and border disputes, as well as neglect and underdevelopment in vulnerable regions. According to the report, areas most affected by cattle rustling include Turkana, Baringo, West Pokot, Elgeyo Marakwet, Samburu, and Laikipia. Data from the National Crime Research Centre (2020) revealed alarming rates of cattle rustling, with Laikipia at 82%, Turkana at 70%, Samburu at 46%, and Elgeyo Marakwet at 45.5%, compared to a national average of 37.2%.

Several Machine learning studies aimed at solving animal identification problems have been conducted. A study by Hang and Basanta (2019) on developing a model for identifying bird species from their images. The model processed bird images as input and classified them into specific species using convolutional neural networks (CNNs) alongside other machine learning techniques.

While the study successfully demonstrated the effectiveness of CNNs for image-based bird classification, its approach encountered limitations when distinguishing between highly similar or nearly identical objects.

R.W. Bello et al. (2020) developed a machine-learning model for identifying Cattle using their unique body patterns. The model accepts image inputs and output classifications by categorizing the cattle into specific groups using convolutional neural networks. While the study demonstrated the power of machine learning

models for classifying cattle using their unique body patterns, its approach suffered limitations in scenarios where objects are highly similar or nearly identical.

In the study by Alharbi et al. (2019), the researchers aimed to classify animals as either predators or pests based on their biometric features. The model accepts biometric inputs and categorizes inputs accordingly, utilizing machine learning techniques such as Support Vector Machines (SVM) and Multi-Layer Perceptron (MLP). However, this approach faced limitations in cases where objects were highly similar or nearly identical, and situations where biometric inputs can be open to manipulation.

According to a study by Crall et al. (2013), a machine-learning model was developed to classify animal species using a labelled dataset, with coat colour as the primary input. The model applied the hot-porter algorithm. While this model proved effective in species identification based on color patterns, it was however unable to accurately classify animal species in cases where coat color was altered by scratches from vegetation or injuries from fights.

In a study by Parham and Stewart (2016) to assist in tracking animals in the wild, the researcher developed a model that identifies zebras from images. The model employed Naive Bayes nearest neighbor algorithm. However, the model faced limitations in cases where the zebra's coat color has been altered by scratches from vegetation, injuries sustained during fights, or deliberate tampering.

Many existing studies on the use of machine learning algorithms demonstrate the power of machine learning models for classifying cattle using their unique body patterns. However, most of these studies highlight challenges in accurately classifying objects that are highly similar or nearly identical especially where the researcher relies heavily on biometric features that are open to manipulation. Again, most studies have shown limited or no use of muzzle images for cattle identification despite their uniqueness thus representing a significant research gap.

Studies have also shown that external factors such as vegetation scratches, injuries, or deliberate tampering can modify animal's distinguishing features, further limiting the reliability of these methods. These gaps underscore the need for alternative identification techniques that enhance classification accuracy, minimize susceptibility to manipulation, and remain effective under varying environmental conditions.

METHODS

Dataset Collection

A dataset of cattle muzzle images was assembled, including various breeds to ensure diversity. To maintain uniformity, all images went through pre-processing, standardizing resolution and aspect ratio. Additionally, data augmentation techniques including image rotation, flipping, and contrast adjustments were implemented to strengthen the model's robustness. The dataset used in this study was sourced from Kaggle.com and it is made up of 4,923 muzzle images of 268 beef cattle. Images were captured using a mirrorless digital camera equipped with a 70–300 mm F4-5.6 lens. The dataset spans three cattle breeds: Angus, Angus x Hereford, and Continental x British cross. All images were obtained under a Creative Commons 4.0 license and were taken at a right angle from a distance of one meter, minimizing disturbance to the animals. Only clear, properly cropped muzzle images were included in the final dataset.

Data Pre-processing

Data pre-processing techniques were applied on the raw data to transform it into a clean, usable format to enhance model performance, it included the following:

- **Resizing:** Images were resized to 244 by 244 pixels using Keras's `tf.image.resize()` function to ensure compatibility with CNN input dimensions.
- **Normalization and Augmentation:** Image pre-processing involved normalization, where pixel values were scaled to a consistent range to stabilize and accelerate training, and augmentation, where transformations

such as rotation and flipping were applied to increase data variability and enhance the model's ability to generalize.

Model Development

A CNN architecture using Keras was used. The model consists of convolutional layers for feature extraction, pooling layers for dimension reduction, and fully connected layers for classification. Transfer learning with pre-trained models was utilized to improve model performance.

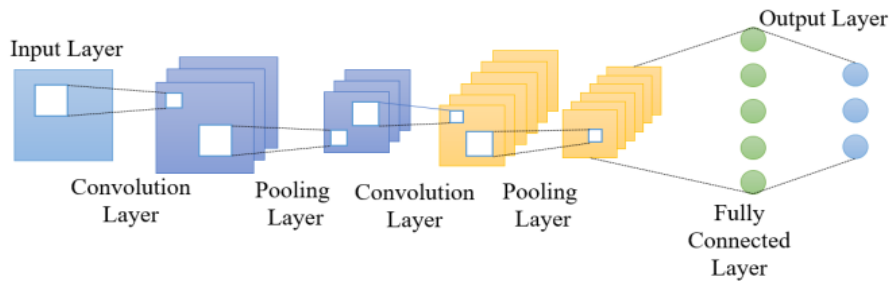


Figure 1: Basic Architecture of a Convolutional Neural Network (Adapted from Research Gate, 2025).

Training and Validation

Dataset was split into training, validation, and testing subsets. The developed model was trained using categorical cross-entropy loss and the Adam optimizer. Early stopping and learning rate scheduling was employed to prevent overfitting. 80% of the images were used for training, while the remaining 20% was split equally into validation and test sets. To split a dataset in Keras (or TensorFlow) train_test_split script was used from the sklearn. model selection module. 10% of the images were utilized for model validation during training, and the final 10% were employed as an independent test set to evaluate the model's performance in real-world situations

Deployment

The trained model was then converted to TFLite format so that it could be used on the mobile device. An Android application was created in Java and Kotlin and integrated with the TensorFlow Lite model for real-time classification. Heroku cloud platform was used to provide computing services that enable users to build, deploy, manage, and scale applications and services via the internet.

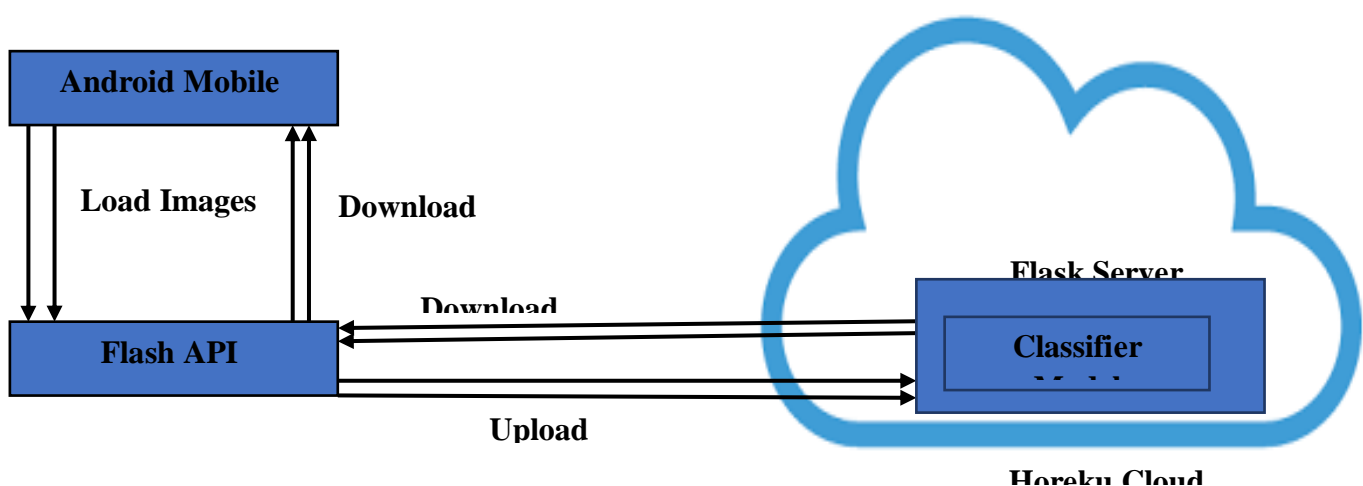


Figure 2: Model Deployment Diagram

Model Evaluation

Model evaluation involves using different metrics to check a model's performance. This helps to identify its strengths and weaknesses. This step is important for figuring out how effective a model is during the early research stages and for ongoing monitoring. To evaluate the performance and validity of the classifier model, 800-labelled images were used for training. The output was then organized into a table, and the results were examined to evaluate its classification performance. The evaluation matrix compared the actual target values with the model's predictions. The model showed strong performance in accurately identifying the images, as shown below.

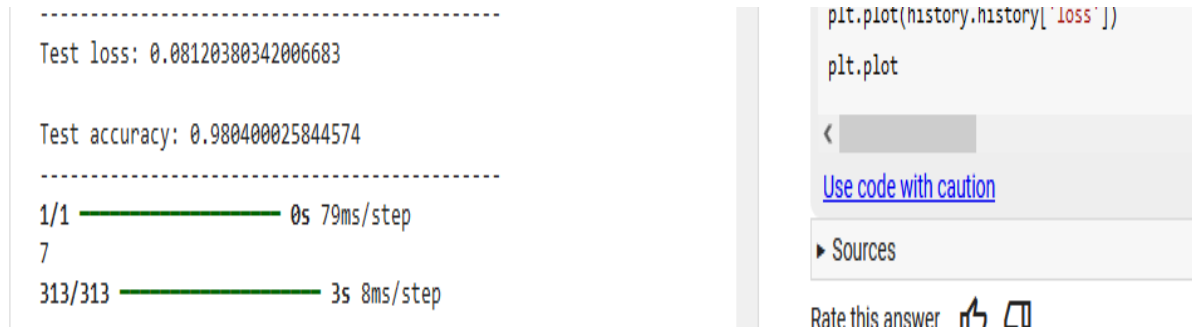


Figure 3: Model Evaluation Diagram

The model was evaluated based on accuracy, precision, recall, and F1-score. A comparative analysis was conducted with existing identification methods. The application's performance was the assessed in real-world situations, including changing lighting conditions and image quality. The model achieved high accuracy and showed strong performance in classifying muzzle images. The study used detailed performance metrics and visual results, including graphs and confusion matrices, to confirm its effectiveness. The model's prototype was tested through a series of experiments to measure its accuracy in identifying individual cattle. The test dataset represented real-world conditions and was used to evaluate the model's performance on data it had never seen before. The model achieved a test accuracy of 98%, with 98% of test images correctly labelled. This highlights its strength and reliability in identifying cattle based on their muzzle patterns.

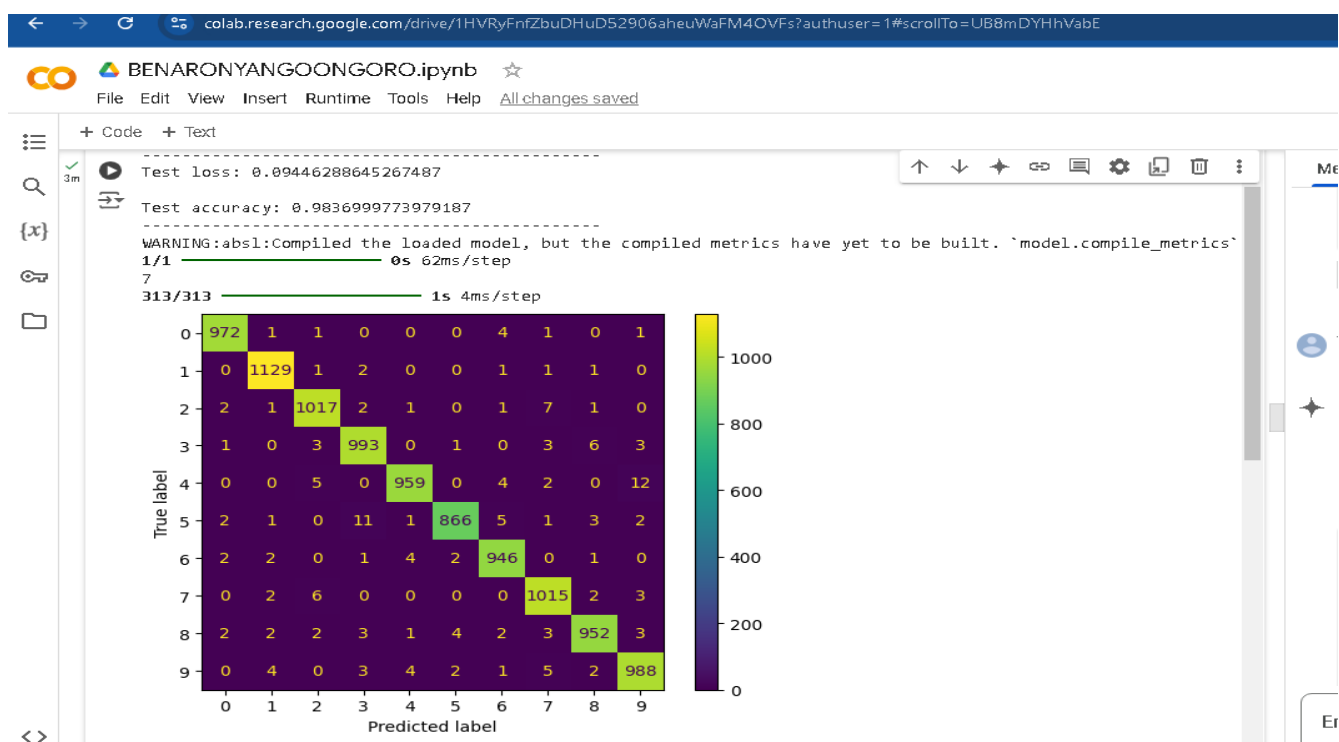


Figure 4: Diagram of Confusion Matrix

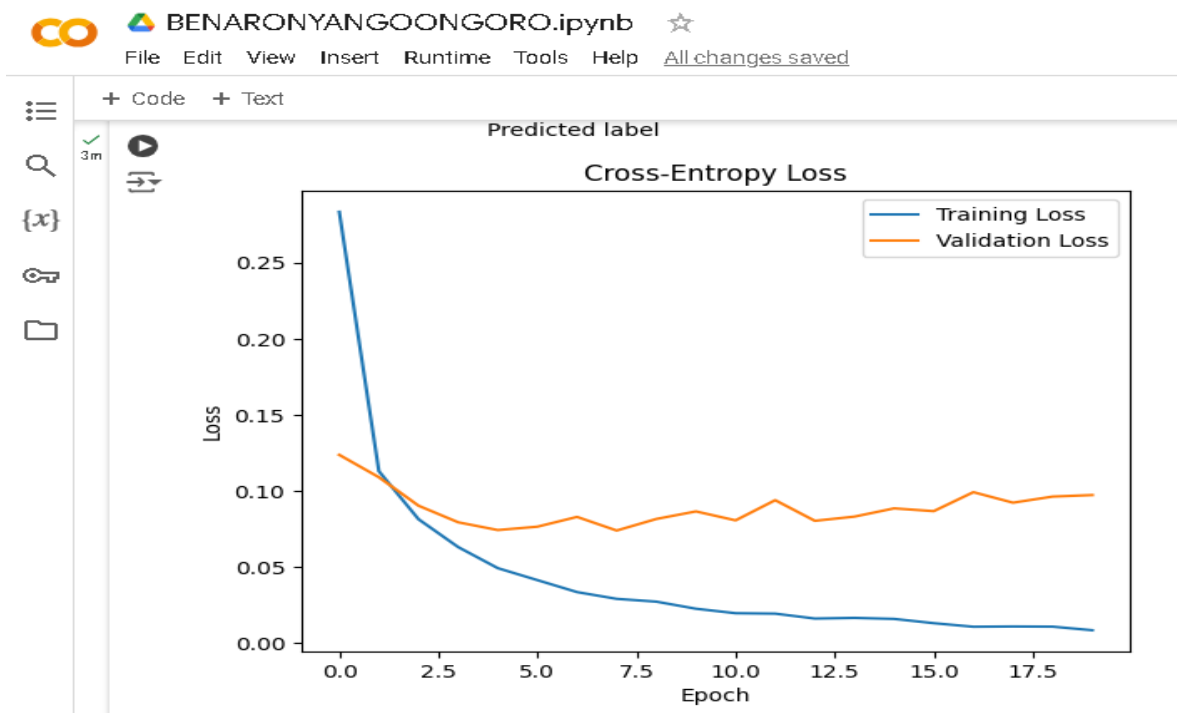


Figure 5: Cross Entropy Loss Diagram

RESULTS

The prototype model achieved a test accuracy of 98 % in identifying individual cattle from muzzle images, demonstrating strong robustness and effectiveness. The use of data augmentation, normalization, and transfer learning likely contributed to this high performance by enabling the model to generalize effectively across unseen images.

DISCUSSION

This accuracy aligns with findings in related research where convolutional neural networks (CNNs) have shown superior performance in visual classification tasks. The integration of the model into an Android platform enhances its practicality and accessibility, particularly in regions affected by cattle rustling, offering significant value for livestock management and security. The results further suggest that muzzle patterns provide a highly reliable biometric trait for cattle identification, comparable to human fingerprint recognition. However, further testing under real-world farm and market conditions is necessary to validate the system's reliability beyond controlled datasets. Beyond technical performance, the deployment of such systems in rural communities carries important ethical, economic, and practical implications. Ethically, questions of data privacy and ownership must be addressed, as farmers may be concerned about how biometric records of their cattle are stored, who controls access, and whether such data could be misused for surveillance or commercial exploitation. Economically, while the technology promises to reduce financial losses from cattle theft and disputes over ownership, the costs of smartphones, internet access, and system maintenance may be prohibitive for smallholder farmers unless subsidized or supported through cooperative models. Practically, farmer adoption hinges on the usability of the application, the availability of training, and the degree of trust in digital tools, especially in communities where technological literacy may be limited. Ensuring equitable access, transparent governance of data, and affordable implementation strategies will therefore be critical to realizing the potential of this system for improving livestock security and management in rural settings.

CONCLUSION

This study successfully developed a cattle identification model using muzzle images and convolutional neural networks (CNNs), effectively addressing the challenge of reliably distinguishing individual animals in regions prone to cattle rustling. The system achieved high prediction accuracy, making it practical even in situations

where other models fail to classify animals accurately due to altered features such as coat color changes from vegetation scratches or injuries. The findings affirm CNNs as a powerful tool for biometric-based cattle identification.

RECOMMENDATIONS

Future research should focus on expanding the dataset to capture greater variability, improving error detection mechanisms to minimize misclassification, and exploring additional biometric traits to further enhance the model's reliability and applicability in diverse real-world environments.

REFERENCES

1. KIPPRA. (2024, January 8). Banditry and lawlessness in arid and semi-arid lands of Kenya: Which way out? Retrieved from <https://kippra.or.ke/banditry-and-lawlessness-in-arid-and-semi-arid-lands-of-kenya-which-way-out/>
2. Alsaadi, E. M. T. A., & El Abbadi, N. K. (2020). An Automated Classification of Mammals and Reptiles Animal Classes Using Deep Learning. *Iraqi Journal of Science*, 2361–2370.
3. Bello, R., & Abubakar, S. (2019). Development of a software package for cattle identification in Nigeria. *Journal of Applied Sciences and Environmental Management*, 23(10), 1825–1828.
4. Crall, J. P., Stewart, C. V., Berger-Wolf, T. Y., Rubenstein, D. I., & Sundaresan, S. R. (2013). Hotspotter—Patterned species instance recognition, 230–237.
5. Huang, Y.-P., & Basanta, H. (2019). Bird Image Retrieval and Recognition Using a Deep Learning Platform. *IEEE Access*, 7, 66980–66989. <https://doi.org/10.1109/ACCESS.2019.2918274>
6. Parham, J., & Stewart, C. (2016). Detecting plains and Grevy's Zebras in the real world. 1–9.