

# CATTLE DISEASE DIAGNOSIS SYSTEM USING RANDOM FOREST CLASSIFICATION METHOD

Muh. Fadhil Syam<sup>1</sup>, Muhammad Nur Yasir Utomo<sup>2</sup>, Eddy Tungadi<sup>3</sup>

<sup>1,2,3</sup> Department of Informatics and Computer Engineering, Politeknik Negeri Ujung Pandang, Indonesia

---

## Article Info

### Article history:

Received, (15 Mei 2025)

Revised, (1 Juni 2025)

Accepted, (16 Juni 2025)

### Keywords:

Cattle disease diagnosis;

Random forest classification;

Machine learning;

Expert system;

Animal health

---

## ABSTRACT

Cattle farming plays an important role in Indonesia's economy, but its productivity can decline due to livestock health issues. To address this, this study develops a cattle disease diagnosis system based on machine learning using the Random Forest classification method. The system helps farmers identify diseases independently based on input symptoms. The model is built using the Random Forest algorithm, trained on 1745 primary data obtained from the Barru Regency Department of Agriculture. The data undergoes a comprehensive pre-processing stage, including cleaning to remove inconsistencies, One-Hot Encoding for categorical feature transformation, and class balancing using the Synthetic Minority Over-sampling Technique (SMOTE) to ensure fair representation of all disease categories. Model evaluation using a Confusion Matrix demonstrates a high accuracy of 91%, indicating strong predictive performance. Based on the model, a mobile application based on Android is developed to assist farmers in the early detection of cattle diseases.

---

## ABSTRAK

*Peternakan sapi memiliki peran penting dalam perekonomian Indonesia, tetapi produktivitasnya dapat menurun akibat masalah kesehatan ternak. Untuk mengatasi hal ini, penelitian ini mengembangkan sistem diagnosa penyakit sapi berbasis machine learning menggunakan metode klasifikasi Random Forest. Sistem ini membantu peternak mengidentifikasi penyakit secara mandiri berdasarkan gejala yang diinputkan. Model dikembangkan menggunakan algoritma Random Forest dengan 1745 data primer dari Dinas Pertanian Kabupaten Barru. Proses pre-processing mencakup pembersihan data, One-Hot Encoding, dan penyeimbangan kelas menggunakan Synthetic Minority Over-sampling Technique (SMOTE). Evaluasi model menggunakan Confusion Matrix menunjukkan tingkat akurasi yang tinggi, yaitu 91%, yang mengindikasikan kinerja prediksi yang kuat. Berdasarkan model ini, dikembangkan aplikasi mobile berbasis Android untuk membantu peternak dalam deteksi dini penyakit sapi.*

---

## Corresponding Author:

Muh. Fadhil Syam

Department of Informatics and Computer Engineering, Politeknik Negeri Ujung Pandang, Jl. Perintis Kemerdekaan KM. 10, Makassar, Indonesia

Email: fasyamuhammad41@gmail.com

---

## 1. INTRODUCTION

The livestock sector plays a crucial role in the national economy and in meeting the nutritional needs of the Indonesian population. One of the dominant subsectors is cattle farming, which significantly contributes to national meat production. However, the imbalance between demand and supply remains a major challenge. According to the Central Bureau of Statistics (BPS) in 2023, Indonesia's beef consumption reached 816.79 thousand tons (equivalent to 4.07 million cattle), while domestic production could only supply 442.69 thousand tons (2.20 million cattle). This resulted in a deficit of 374.1 thousand tons (1.86 million cattle or approximately 41.41%) [1], highlighting the country's dependence on imports to meet domestic demand.

One of the key factors contributing to the low production is cattle health issues. Diseases such as anthrax, which has a mortality rate of up to 80%, and brucellosis and hemorrhagic septicemia, which can reduce productivity by 60%, pose serious threats to the livestock industry. Additionally, digestive parasite infections can decrease production by up to 90% [2], further exacerbating challenges in the national livestock sector. Therefore, improving livestock health management is crucial in enhancing productivity and sustainability in Indonesian cattle farming.

Barru Regency, particularly Barru District, holds significant potential for cattle farming, with a recorded beef cattle population of 34,095 heads in 2023 [3]. However, most farmers still practice extensive farming systems, where cattle are left to graze in open fields with limited protection facilities. This traditional approach increases the risk of disease transmission due to the lack of structured health management and minimal preventive measures taken by farmers [4].

Since 2015, the Department of Livestock has implemented the ISIKHNAS (National Animal Health Information System) to record livestock disease occurrences. This system enables disease reporting via text messages managed by livestock cadres across various regions. Collected data indicates that common cattle diseases in Barru Regency include scabies, worm infestations, and brucellosis [5]. However, limited veterinary professionals, farmers' lack of awareness of disease symptoms, and high consultation costs hinder early detection and effective disease management [6].

To address these challenges, this study aims to develop a cattle disease diagnosis system using the Random Forest algorithm. This algorithm is selected due to its ability to handle complex data and provide accurate predictions [7], [8]. By implementing this system, farmers are expected to detect cattle diseases more efficiently and promptly, improving overall livestock health management and productivity.

## 2. METHODOLOGY

This research employs a Machine Learning (ML)-based system development approach to develop diagnose cattle diseases system based on android. The system process consists of several key stages, as illustrated in the system flowchart below.

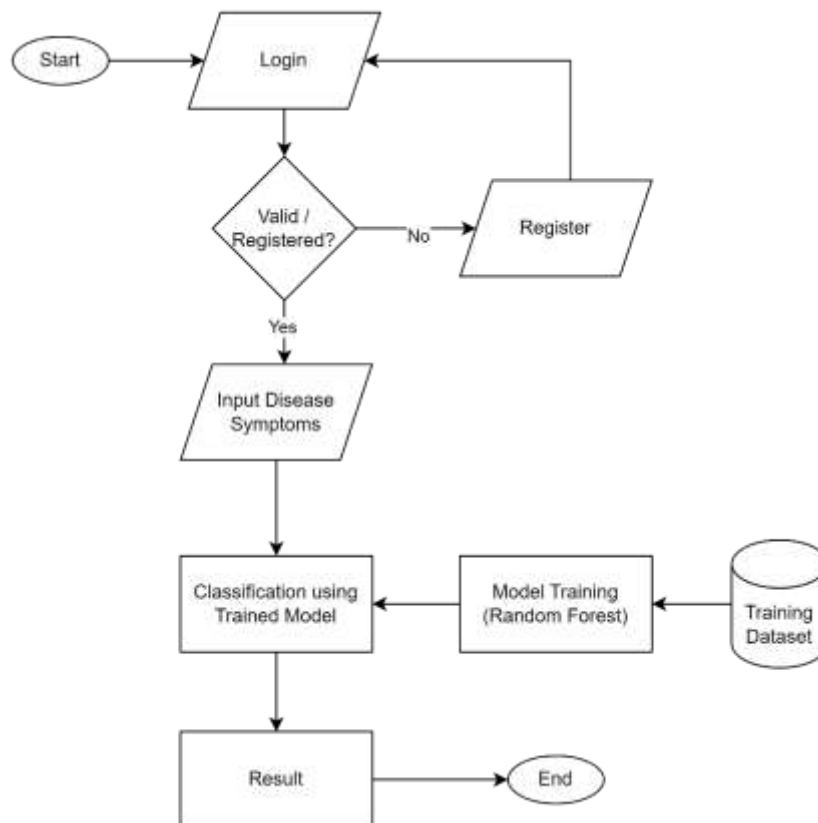


Figure 1. System Architecture

### 2.1. User Access and Authentication

The cattle disease diagnosis system is designed to be accessed by users through a login and registration mechanism. Registered users undergo system identification, while new users must first complete the registration process before accessing the diagnosis features.

## 2.2. Inputting Disease Symptoms

Once authenticated, the user proceeds to input the observed disease symptoms by selecting the relevant options provided by the system. These symptoms serve as the primary data input for the diagnosis process.

## 2.3. Disease Classification Using Machine Learning

The system incorporates a training dataset that has been processed to develop the Random Forest model. Training model in machine learning consist of data pre-processing [9], dataset splitting [10], and the training-evaluation [11], [12]. The system then utilizes the trained Random Forest model to classify the disease based on the symptoms provided. This classification process helps predict the most probable disease affecting the livestock.

## 2.4. Displaying Diagnosis Results

After classification, the system displays the diagnosis results to the user through API as shown in Figure 2. The results provide insights into the possible disease, enabling farmers to take appropriate action.

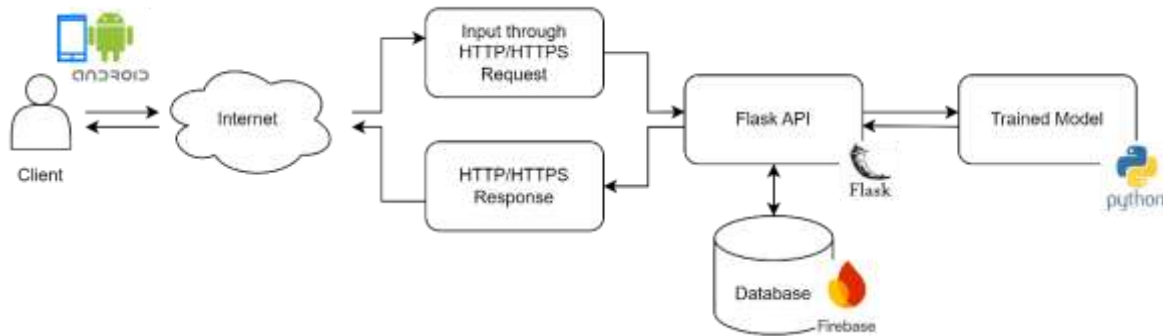


Figure 2. Infrastructure Design

Through this system, farmers can detect cattle diseases more efficiently, allowing for timely and appropriate interventions. By leveraging Machine Learning technology, this system contributes to improving livestock health management in Barru Regency.

## 3. RESULTS AND DISCUSSION

This research resulted in a model and a mobile-based information system for diagnosing and managing cattle, assisting farmers in organizing and diagnosing cattle diseases. The system helps accelerate cattle health management and improve livestock organization.

### 3.1. Cattle Disease Diagnosis Model

The Random Forest algorithm is a classification method that utilizes a combination of multiple decision trees to enhance prediction accuracy. This algorithm is used in the classification process of disease symptom data in animals by analyzing patterns from previously trained data. The implementation phase involves transforming the designed system into an application, where the Random Forest model processes user input data and automatically provides diagnostic results. Here are steps of model training conducted:

#### 3.1.1. Dataset Collection

The dataset was obtained from the Agriculture and Livestock Office of Barru Regency. The collected data is in CSV format, consisting of 11 columns and 3,392 records. This dataset represents a collection of animal disease cases over the past five years. The dataset consist of The example of the dataset is shown in Figure 3.



```

↳ <class 'pandas.core.frame.DataFrame'>
RangeIndex: 1745 entries, 0 to 1744
Data columns (total 27 columns):
#   Column                                     Non-Null Count  Dtype
---  ---
0   no                                           1745 non-null   int64
1   anorexia                                     1745 non-null   object
2   bulu kusam                                  1745 non-null   int64
3   kekurusan                                    1745 non-null   int64
4   lemah                                        1745 non-null   int64
5   mencret                                      1745 non-null   int64
6   gatal                                        1745 non-null   int64
7   leleran nanah vulva                         1745 non-null   int64
8   hidung beringus                             1745 non-null   int64
9   demam                                        1745 non-null   int64
10  pincang                                      1745 non-null   int64
11  lumpuh                                       1745 non-null   int64
12  bulu rontok                                  1745 non-null   int64
13  radang mata                                  1745 non-null   int64
14  luka berdarah                               1745 non-null   int64
15  mencret berdarah                           1745 non-null   int64
16  ovarium kecil keras                         1745 non-null   int64
17  berahi tenang                              1745 non-null   int64
18  tidak ada ovulasi                          1745 non-null   int64
19  ovarium kenyal dan licin                   1745 non-null   int64
20  anestrus                                    1745 non-null   int64
21  corpus luteum tidak sempurna               1745 non-null   int64
22  daun telinga berkopeng                    1745 non-null   int64
23  partus                                       1745 non-null   int64
24  fases kering keras                         1745 non-null   int64
25  rahim bernanah                             1745 non-null   int64
26  Diagnosa                                    1745 non-null   object
dtypes: int64(25), object(2)
memory usage: 368.2+ KB
    
```

Figure 5. Feture Encoding Rule

An example of the encoding process results is shown in Table 1.

Tabel 1. Example of Encoding Result

	<i>Gejala 1</i>	<i>Gejala 1</i>	<i>Gejala 1</i>	<i>Gejala 1</i>	<i>Gejala 1</i>	Label
Original data	anorexia	demam	kekurusan	lemah	mencret	Coccidiosis
Encoding	1	9	3	4	5	Coccidiosis

### 3.1.3. Data Balancing using SMOTE

Hasil pre-processing data menghasilkan jumlah data yang tidak seimbang pada tiap label yang ada sehingga perlu dilakukan data balancing [13]. Pada penelitian ini, data balancing dilakukan dengan SMOTE. Hasil balancing data diperlihatkan pada Table 2.

Tabel 2. Result of Data Balancing using SMOTE

Label	Number of Data	
	Original	SMOTE
Avitaminosis	399	900
Bovine Ephemeral Fever	97	900
Cacingan	900	900
Scabies	35	900
Coccidiosis	71	900
Hipofungsi Ovari	71	900
Corpus Luteum Persisten	54	900
Baliziekte	33	900
Endometritis	24	900
Thelaziasis	61	900

### 3.1.4. Model Training

The model is trained using randomized data, which is split into two parts: 75% for training and the remaining 25% for testing [14]. The model development involves the Random Forest algorithm from the Scikit-Learn library for classification and GridSearchCV for parameter tuning in the model creation process.

### 3.1.5. Model Evaluation

At this stage, the evaluation is conducted to assess the performance of the developed model. The evaluation uses a Confusion Matrix, and the results are shown in Figure 6.

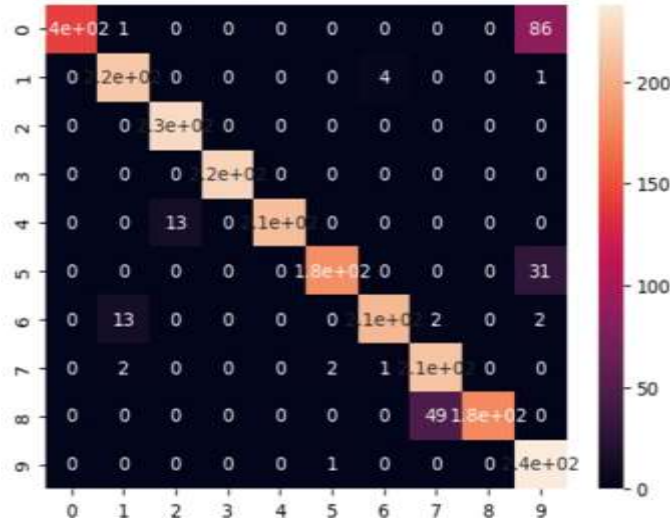


Figure 6. Feture Encoding Rule

Based on the Confusion Matrix results, the accuracy, precision, recall, F1-score, and support values can be determined [15], as shown in Figure 7.

	precision	recall	f1-score	support
Baliziekte	1.00	0.62	0.77	231
Bovine Ephemeral Fever	0.93	0.98	0.95	221
Corpus Luteum Persisten	0.95	1.00	0.97	227
Endometritis	1.00	1.00	1.00	224
Hipofungsi ovari	1.00	0.94	0.97	223
Thelaziasis	0.98	0.86	0.92	214
avitaminosis	0.98	0.92	0.95	224
caingan	0.81	0.98	0.88	218
coccidiosis	1.00	0.79	0.88	229
scabies	0.66	1.00	0.80	239
accuracy			0.91	2250
macro avg	0.93	0.91	0.91	2250
weighted avg	0.93	0.91	0.91	2250

Figure 7. Accuracy of the Model

As shown in Figure 7, the accuracy of the proposed model reaches 91%. This model will then be used as the classification engine for the Android application.

## 3.2. Implementation on Android

To ensure that the disease classification model can be easily used by cattle farmers, an Android application was developed. The results of the Android application are shown in Figure 8.

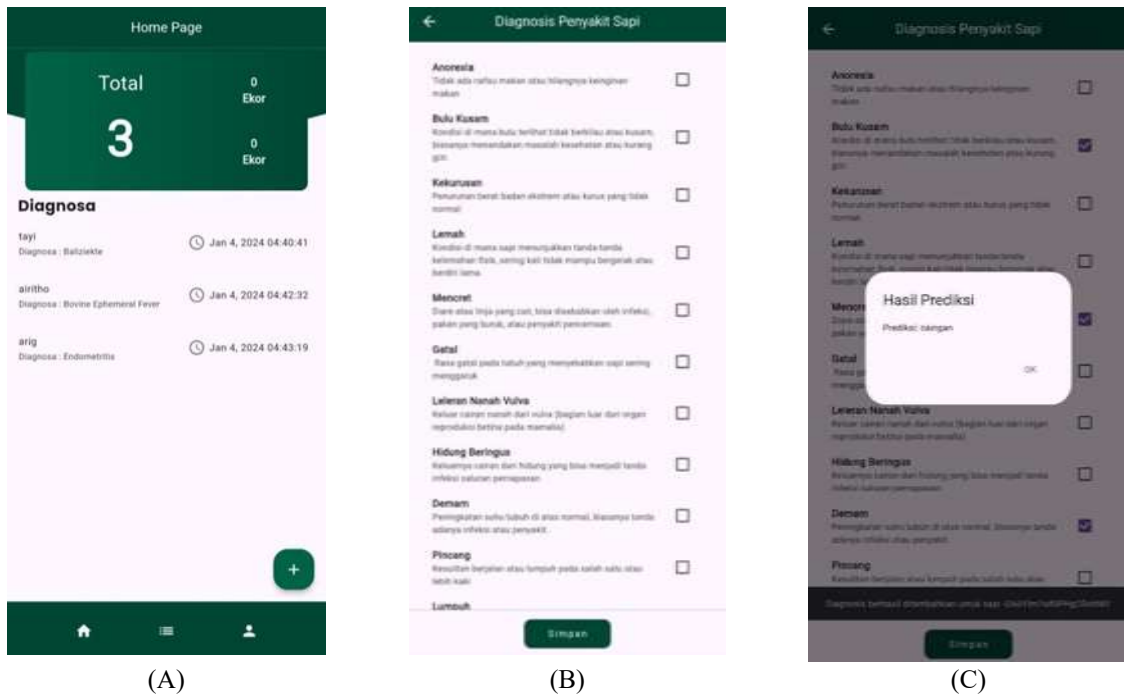


Figure 8. Display of (a) the home page, (b) disease symptom input, and (c) prediction results

The developed Android application is designed to be user-friendly for both cattle farmers and veterinarians, making it a useful tool for early disease detection in livestock.

#### 4. CONCLUSION

The developed cattle disease diagnosis system successfully leverages machine learning, specifically the Random Forest classification method, to assist farmers in identifying livestock diseases based on input symptoms. Using 1,745 primary data points collected directly from the Barru Regency Department of Animal Husbandry, the model underwent comprehensive pre-processing, including data cleaning, One-Hot Encoding, and class balancing with SMOTE. Evaluation using a Confusion Matrix demonstrated a high accuracy of 91%, confirming the model's reliability in disease classification. To enhance accessibility, the model was integrated into a user-friendly Android application, enabling farmers to perform early disease detection efficiently. This system has the potential to improve cattle health management and support sustainable livestock farming in Indonesia.

#### REFERENCES

- [1] Direktorat Statistik Peternakan Perikanan dan Kehutanan, "Peternakan Dalam Angka 2023," Badan Pusat Statistik (BPS) Indonesia, Jakarta, Dec. 2023.
- [2] C. Rifqi, "Sistem Pakar Pengenalan Penyakit pada Hewan Ternak Sapi Menggunakan Metode Certanty Factor Berbasis Web," 2019. Accessed: Apr. 01, 2025. [Online]. Available: <https://eprints.uty.ac.id/3585/1/Naskah%20Publikasi-Choirur-312511146.pdf>
- [3] Badan Pusat Statistik (BPS) Kabupaten Barru, "Kabupaten Barru Dalam Angka 2024 (Barru Regency in Figures 2024)," Barru, Feb. 2024.
- [4] M. Nuriski, A. Wicaksono, and C. Basri, "Distribusi Skabies pada Peternakan Sapi Potong di Kabupaten Barru Provinsi Sulawesi Selatan," *Jurnal Ilmu Peternakan dan Veteriner Tropis (Journal of Tropical Animal and Veterinary Science)*, vol. 10, no. 2, p. 159, Sep. 2020, doi: 10.46549/jipvet.v10i2.97.
- [5] A. Syah, C. Basri, and A. Wicaksono, "Incidence of Surra Disease on Beef Cattle in Barru Regency, South Sulawesi Province in 2015-2017," *Jurnal Medik Veteriner*, vol. 3, no. 2, pp. 145–153, Oct. 2020, doi: 10.20473/jmv.vol3.iss2.2020.145-153.
- [6] M. M. Biesheuvel *et al.*, "Understanding Farmers' Behavior and Their Decision-Making Process in the Context of Cattle Diseases: A Review of Theories and Approaches," Dec. 02, 2021, *Frontiers Media S.A.* doi: 10.3389/fvets.2021.687699.
- [7] M. Bovo, M. Agrusti, S. Benni, D. Torreggiani, and P. Tassinari, "Random forest modelling of milk yield of dairy cows under heat stress conditions," *Animals*, vol. 11, no. 5, 2021, doi: 10.3390/ani11051305.
- [8] D. T. Muhamediyeva, L. U. Safarova, and R. F. Ruzikulov, "Expert systems for diagnostics of infectious diseases in cattle," in *BIO Web of Conferences*, EDP Sciences, Nov. 2023. doi: 10.1051/bioconf/20237101073.
- [9] M. N. Y. Utomo, A. E. Permanasari, E. Tungadi, and I. Syamsuddin, "Determining single tuition fee of higher education in Indonesia: A comparative analysis of data mining classification algorithms," in *Proceedings of 2017 4th International Conference on New Media Studies, CONMEDIA 2017*, 2017, pp. 113–117. doi: 10.1109/CONMEDIA.2017.8266041.
- [10] Maemunah, M. N. Y. Utomo, and R. Nur, "Development of an Agricultural Department Application to Predict Small Chili Prices," *Jurnal Teknologi Elektroika*, vol. 21, no. 2, 2024.
- [11] M. N. Y. Utomo and F. Violita, "Face Mask Wearing Detection Using Support Vector Machine (SVM)," *IJID (International Journal on Informatics for Development)*, vol. 10, no. 2, pp. 72–81, 2021, doi: 10.14421/ijid.2021.3038.

- [12] M. N. Y. Utomo, T. Indrabulan, S. J. Juanda, and A. Rahman, "Image processing for corn quality classification using convolutional neural network (CNN) method," in *AIP Conference Proceedings*, American Institute of Physics, Jul. 2024. doi: 10.1063/5.0221041.
- [13] S. Suparyati, Emma Utami, and Alva Hendi Muhammad, "Applying Different Resampling Strategies In Random Forest Algorithm To Predict Lumpy Skin Disease," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 4, pp. 555–562, Aug. 2022, doi: 10.29207/resti.v6i4.4147.
- [14] G. Rai, A. Hussain, A. Kumar, and R. Nijhawan, "A Deep Learning Approach to Detect Lumpy Skin Disease in Cows," in *Computer Networks, Big Data and IoT*, Springer, Jun. 2021. doi: [https://doi.org/10.1007/978-981-16-0965-7\\_30](https://doi.org/10.1007/978-981-16-0965-7_30).
- [15] B. Lake, F. Getahun, and F. T. Teshome, "Application of Artificial Intelligence Algorithm in Image Processing for Cattle Disease Diagnosis," *Journal of Intelligent Learning Systems and Applications*, vol. 14, no. 04, pp. 71–88, 2022, doi: 10.4236/jilsa.2022.144006.