Breed Lineage Prediction of Small Ruminants Using Deep Learning

Mohammad Farizshah Ismail Kamil a,*, Nor Azliana Akmal Jamaludin a, Mohd Rizal Mohd Isa a

Faculty of Defense Science and Technology, National Defense University of Malaysia, Kuala Lumpur, 57000, Malaysia

Abstract—Sheep are a significant food source for humans, besides cattle and poultry. Despite its significance to Malaysian Muslims, who make up approximately 60% of the local population, the sheep supply is limited by the high mortality rate caused by fatal diseases such as foot and mouth disease (FMD) and tetanus. Infected sheep can spread food-borne bacteria, such as Escherichia coli, at various preparation phases, contaminating the meat. The objectives of this study are to identify internal and external factors that influence sheep breed lineage continuity, investigate current practices for collecting and managing data knowledge on sheep breed and hereditary diseases, and propose a sheep breed and disease data knowledge model based on the feedforward artificial neural network (FANN) deep learning method. This study utilized qualitative and quantitative data to obtain in-depth answers to the research questions, which involves collecting all the information required for the system development using the FANN deep learning method. This study found that breeding is the leading data group for tracking each sheep's ADG and BCS. Feed type, sanitization, and medication influence sheep's daily increase and health. Collaboration, worker knowledge, and climate are re...
then, the bacteria have spread around the globe by way of water- and food-borne channels [16]. Shiga toxin-producing E. coli (STEC) is the most lethal of the more than 700 E. coli serotypes identified so far [17].

According to a news article published by the Star newspaper in 2022, Malaysia is currently experiencing a food price hike due to supplier hoarding and the war between Russia and Ukraine. Notably, Russia ranks second in the world regarding agricultural land size. As 60% of Malaysia’s food supply is imported, any consequential event in that region would impact the nation and the rest of the world. Specifically, Malaysia relies heavily on imported mutton (88.8%) and beef (76.4%). At one point, Malaysia was self-sufficient in poultry production [18], but that has now changed due to several factors, such as the closure of many farms and exports to Thailand due to the high demand for chicken caused by swine flu. Sinar Harian has also reported that the sheep and cattle population in Johor state is declining due to diseases and limited farmland in 2020 [19]. Nonetheless, meat demand remains high, with 4,791 carcasses expected in 2020, up from 4,308 in 2019 [19].

Previously, Agrawal et al. [20] employed the transfer learning model to determine the sheep breed by comparing the image of one sheep with hundreds of other sheep photographs. The front layers of a deep neural network (DNN) model are transferred using this technique, which then directs the model to the target domain and uses it as a feature extractor [21]. Similarly, Jwade et al. [22] used GoPro cameras to send video recordings of sheep on a weighing station to identify the breed using the transfer learning model. Furthermore, Salama et al. [23] utilized the Bayesian Optimisation (black-box) approach to compare photographs of sheep faces to a database containing hundreds of images to identify a specific sheep (Barqi). This method was chosen because of the automated procedure with a data-efficient optimization algorithm [24]. Table I presents the findings from earlier investigations on sheep breed identification.

### Table I: Comparative Analysis of Sheep Breed Identification Method

<table>
<thead>
<tr>
<th>Author</th>
<th>Research</th>
<th>Method</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agrawal</td>
<td>Used a camera to send pictures of a single sheep for comparison with hundreds of other images to determine the breed</td>
<td>Transfer Learning Model</td>
<td>27 minutes</td>
</tr>
<tr>
<td>Jwade</td>
<td>Used a variety of GoPro cameras to send videos of a single sheep for comparison with hundreds of other photographs to determine the breed</td>
<td>Transfer Learning Model</td>
<td>27 minutes</td>
</tr>
<tr>
<td>Salama</td>
<td>Compared sheep facial photographs to a database of hundreds of pictures to identify a specific sheep</td>
<td>Bayesian Optimisation</td>
<td>8 minutes</td>
</tr>
</tbody>
</table>

Despite the success reported in the literature, all approaches utilized by the studies involved two lengthy processes: database search and image filtering. Therefore, the current study aimed to establish a method that produces the desired results quickly. Precisely, the objectives of this study are to identify internal and external factors that influence sheep breed lineage continuity, investigate current practices for collecting and managing data knowledge on sheep breed and hereditary diseases, and propose a sheep breed and disease data knowledge model based on the feedforward artificial neural network (FANN) deep learning method [4].

## II. Materials and Method

This study utilized qualitative and quantitative data to obtain in-depth and accurate solutions for the research questions [5]. The mixed-method research (MMR), a technique that uses numerous methods to provide systematic and ethical answers to research questions, was chosen [25]. According to the literature, MMR is useful in describing and clarifying the complexity and difficulties of a study [26].

This study was designed by combining quantitative and qualitative data at different levels (data collection, analysis, reporting, interpretation, integration) in numerous ways. The quantitative and qualitative data collection and analysis were conducted using the exploratory sequential design (see Figure 1).

![Fig. 1 Exploratory sequential mixed-method research](image-url)

The qualitative analysis involved collecting all the information required for the quantitative data analysis, which would focus on the system development via the feedforward artificial neural network (FANN) deep learning method. This paper covers the first part of the study, the qualitative data analysis, where the related data was collected while bridging the gap between farmers and buyers. This section details the participant selection, procedure, and instruments used in this study.

### A. Participants

Several farms were selected for this study, including XX, YY, and ZZ. The XX farm, a pilot farm under the East Coast Economic Region (ECER), is the largest in Terengganu. Meanwhile, YY Livestock was chosen because the farm is registered as a Research Lab for Universiti Putra Malaysia (UPM) in the northern region. Lastly, ZZ Farm was recommended by the university as one of the fastest-growing farms in Putrajaya.

The XX farm is the largest among all three farms, with approximately 20-25 acres of land that houses ± 11,000 goats in 11 pens. Workers record the daily data in a book, and later digitized by an IT Exec via a web-based form. Meanwhile, the YY farm has a land size of 3 acres with 665 goats in two pens and 200 chickens in another. The workers...
record the farm data daily on paper and in books. In addition, workers record data on paper sheets and books that are collected, which will be input into an Excel sheet (Microsoft, USA) and Google Drive (Google, USA). Finally, the ZZ farm is the smallest of three farms, with 1 acre of land and one pen that houses 200 sheep. This farm never recorded a data.

B. Procedure

The data were obtained via interviews, and data management systems of the respective farms were analyzed. First, two volunteers from UPM with different educational backgrounds pre-tested the interview guide to ensure that the language and questions were appropriate. This qualitative study was performed under COREQ (Consolidated Criteria for Reporting Qualitative Research) [27].

Farmhands participated in one-on-one, semi-structured interviews that largely comprised open-ended questions. The authors developed the interview guide, which contained 45 open-ended questions. The interview guide was reviewed by UPM agriculture experts prior to administration. Each participant understood all interview questions in the Malay language and dialect. Participants' understanding of small ruminant farming, the sector, breeding, nursing, and damage control was evaluated. Furthermore, their knowledge concerning livestock breed selection, breeding techniques, and sheep and goat feed was assessed.

1) Pre-Interview: First, the appointment was set between the researcher and the farm's contact person before the interview. The interview details were then provided to the participants by one of the researchers. The interview began with an introduction by the visiting researcher, including their university identification.

2) Interview: The interviews were conducted on-site at XX farm in Bandar Permaisuri, Terengganu, YY farm in Yan, Kedah, and ZZ farm in Putrajaya, Kuala Lumpur.

3) Farm Management System Overview: The XX farm hired a team of IT executives to digitize the daily records collected by farm workers in the assigned books. The data obtained include mating, birth, death, daily gain, daily feed, plotting, carcass sales, and sliced meat sales. The YY farm owners and workers are IT literate; thus, no hiring was done for record keeping. Similar to XX farm, the workers' records are required to record the farm's daily activities in books before the information is digitized in Excel sheets and stored in Google Drive. The ZZ farm does not collect data and performs its tasks based on a fixed routine.

C. Instrument

The interview questions were divided into nine sections as follows: Section I: education level of the interviewee, Section II: farm details (location and size), Section III: livestock (animal type, breed and population size), Section IV: farm record-keeping practices, Section V: breeding methods, Section VI: livestock feed types, Section VII: medication and sanitation, Section VIII: livestock diseases, and Section IX: association between the weather and animals' general well-being at the farm (see Table II).

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Education level</td>
<td>To identify the correlation between the farm owner and/ or worker's education level with daily activity.</td>
</tr>
<tr>
<td>II</td>
<td>Farm details</td>
<td>To understand the connection between farm size, location, and initial capital with the farm growth over the years.</td>
</tr>
<tr>
<td>III</td>
<td>Livestock</td>
<td>To understand if animal type, breed and population contribute to farm revenue fluctuations.</td>
</tr>
<tr>
<td>IV</td>
<td>Management &amp; recording</td>
<td>To identify the link between Section I with farm management and record keeping.</td>
</tr>
<tr>
<td>V</td>
<td>Breeding method</td>
<td>To understand the connection and impact of pure-breeding and cross-breeding methods on animals' general health and productivity.</td>
</tr>
<tr>
<td>VI</td>
<td>Feed</td>
<td>To understand if feed varieties influence livestock growth and gains. Feed source is also considered because this factor could explain the connection between farms, agencies and suppliers.</td>
</tr>
<tr>
<td>VII</td>
<td>Medication &amp; sanitization</td>
<td>To understand the farm practices in maintaining a safe environment and healthy livestock. The standard operating procedure (SOP) in treating sick animals is also observed to learn about the administered treatments and outbreak prevention.</td>
</tr>
<tr>
<td>VIII</td>
<td>Disease &amp; illness</td>
<td>To investigate whether the farm has experienced an outbreak and the measures taken to contain the situation.</td>
</tr>
<tr>
<td>IX</td>
<td>Weather</td>
<td>To evaluate the role of weather on the livestock's overall well-being.</td>
</tr>
</tbody>
</table>

D. Qualitative Data and Analysis

The underlying data for the initial analysis were obtained from existing databases that was built and maintained XX farm employees. Missing data and outliers in the dataset were examined by using the "outlier labelling rule". Outliers are values that fall beyond the calculated range [28]. Subsequently, the data were assessed using the Multilayer Perceptron Classifier (MPC) [29] algorithm and PHP's built-in Fast Artificial Neural Network (FANN). The Feedforward Artificial Neural Network, which maps sets of data onto a collection of suitable outputs, includes FANN and MPC algorithms. The outcome was then checked for multiple linear regression using statistical software, such as JAMOVI. This program allows users to modify the R syntax to produce more precise charts and graphs.

III. RESULTS AND DISCUSSION

This section is divided into two subsections: 1) the XX, YY and ZZ farm interview analysis, and 2) the management system analysis of the respective farms.

A. Interview Analysis

1) Farm Owner: The state of Terengganu owns the XX farm; thus, there was no information on the owner’s education level. Meanwhile, the YY farm owner is an Animal Science graduate from UPM, and the ZZ farm owner is an SPM holder.
2) Participants: Generally, farm owners are required to register their farms with the Veterinary Department Malaysia before they are allowed to operate legally. The local authorities verified the three farms involved in this study. The XX farm is located within a farming area around Bandar Permaisuri, Setiu, and Terengganu. The farm was set up in 2016, 20 - 25 acres wide, with most areas planted with Napier grass. However, only a small farm section is high enough to construct pens; other areas are submerged in floods during the monsoon season. Furthermore, limited information is publicly available as XX farm is state property. The YY farm is located at the base of Gunung Jerai in Yan, Kedah. The farm is 3 acres wide and belongs to a family of farmers. The farm was established in 2016 with only eight cows and a capital of RM100,000. In addition, the first goat pen was built in 2017, and the chicken coop in 2018. Notably, the farm did not acquire any bank loans as of the date of the interview. The ZZ farm site is in Putrajaya, specifically at the final bend of Desa Pinggiran Putra. This small farm is approximately 1 acre in size and is located at a water reservoir between two hills. Moreover, the farm was set up in 2016 with a capital of RM50,000 and only one pen to house a few sheep.

3) Livestock: The XX farm specialises in Boer goat farming. At the time of the visit, approximately 11,000 goats were housed in 11 pens, with a 1,000 goats/pen capacity. The farm sells an average of 3,000 goat carcasses annually to the state government. Furthermore, the farm sells processed sheep meat, fresh Napier and mushrooms to increase revenue. When the interview was conducted, three major livestock were found in YY farm, including 200 chickens, 665 goats, and four cows. The main goat breeds on the farm are Boer, Saanen and Shami, and small groups of Santa Ines. The YY farm goats are provided with fresh Napier and Gro Bio pellets. Moreover, the Napier grass is homegrown, thus, allowing access to fresh and dried feed supply all year round. In ZZ farm, sheep are given fresh Napier and Gro Bio pellets. Napier grass is grown on-site for farm supply. In addition, the farm purchases Napier from nearby farms due to the lack of ample space for Napier cultivation.

4) Breed: The XX farm practices the pure-breeding method, as the only breed available is the Boer goat. Besides their lucrative value, the Boer breed was preferred because of their resilience to extreme weather and diseases. The YY farm practices pure and cross-breeding. Saanen is preserved as purebred to maintain their milk quality, while Boers are for meat quality. Other breeds are cross-bred with Shami from Syria to improve their meat volume and resistance to harsh weather and diseases. The ZZ farm practices the pure-breeding method for a handful of sheep, while the rest are traded within short intervals. Furthermore, the sheep are separated according to their breeds, such as Black Belly and Santa Ines.

5) Fodder: Feed greatly influences livestock growth and gains. Animals must be fed various feeds during different seasons and times of day to ensure optimum productivity. Livestock that depend solely on grass will have a different yield than those fed with only dry hay or pellets. Fresh and pickled Napier grass or silage is the feed of choice for goats in the XX farm. The grass is cultivated on-site for farm use, and the remaining fodder is sold to other farms. Additionally, the abundant farmland allows XX farm to supply the grass to various locations. The YY farm goats are provided with various fodder, such as fresh Napier, Napier hay, pellets (Grobio Industry Sdn Bhd, Malaysia), silage and Bokashi. Moreover, the Napier grass is homegrown, thus, allowing access to fresh and dried feed supply all year round. In ZZ farm, sheep are given fresh Napier and Gro Bio pellets. Napier grass is grown on-site for farm supply. In addition, the farm purchases Napier from nearby farms due to the lack of ample space for Napier cultivation.

6) Medication and Sanitisation: The XX farm medicated their animals as scheduled by the Veterinary Department. In addition, each pen is cleaned with water every morning, and each plot is supplied with salt blocks as a mineral supply for the goats. Meanwhile, the animal houses are sprayed daily with EM (Effective Microorganism) mixed with molasses at the YY farm. Specifically, the spraying is done around the farm, on the goats and their faeces. This measure is essential to eliminate undesirable odours from animal meat and milk. Furthermore, each goat will be medicated periodically as the Veterinary department suggests. When an animal is ill, it will be separated from the flock and placed in a different plot while awaiting inspection and treatment from the veterinary officers. The medications are usually administered via subcutaneous injection. Apart from medicine, each plot is supplied with mineral salt blocks for the goat to lick throughout the day. As for ZZ farm, each sheep is given scheduled medication as the Veterinary department recommends.

7) Disease and Illness: Since the farms were established, their livestock has been generally healthy without any outbreaks when supplied with ample feed. Therefore, the mortality recorded is from natural causes and none from fatal diseases. An animal will be slaughtered immediately if it is severely ill and beyond treatment.

8) Weather: All three farms reported that the weather highly affects their goat and sheep health. A common observation among young animals includes catching a cold and eventually dying of a cold or a stuffed nose during the cold season.
B. Farm Management Analysis

The XX farm has 16 active workers and university interns on site. Daily data such as average daily gain (ADG) is recorded in a book before being digitized by the farm’s IT executive. An on-site office was set up with electricity and an internet connection, and the data was stored on an online MySQL database. The YY farm workforce consists of two main workers and three general helpers. The main workers are UPM graduates, while the helpers are SPM holders. The main workers and three general helpers. The main workers later collect the book and input it into Excel sheets. Subsequently, the digitized data are uploaded to Google Drive when the workers go into town to access a proper internet connection.

The two owners of the ZZ farm are in charge of their establishment, but no data has ever been recorded. The daily activities are performed as the Veterinary department recommends, including the feeding and medication schedule. Despite using separate databases and Excel sheets, the bulk of the data from XX and YY farms are summarized in a master sheet and table at specific time points. These changes indicated the shift in staffing when a previous worker resigns and a new worker is replaced without proper handover.

Generally, the daily data collection for XX comprised ADG, birth, death, daily plot population count, and mating. Furthermore, essential data such as pen, plot, date, tag number, ram tag number, ewe tag number, category, gender, weight, comments, and Body Composition Score (BCS) are recorded [6] (see Table IV).

<table>
<thead>
<tr>
<th>BCS</th>
<th>Details</th>
</tr>
</thead>
</table>
| 1   | a. Transverse and spinous processes are distinct and pointed.  
     | b. Absence of fat cover and shallow loin muscle.  
     | c. Obvious ribs. |
| 2   | a. The presence of sharp and noticeable spines.  
     | b. Muscles have little fat covering, although large.  
     | c. Transverse processes' ends are rounded and have a smooth finish.  
     | d. Some ribs are evident.  
     | e. There is some fat cover, but not much.  
     | f. Ribs are still felt. |
| 3   | a. Spinal processes are rounded and smooth.  
     | b. Adequate fat covers the muscles.  
     | c. Firm pressure can be used to feel the transverse processes.  
     | d. A uniform coating of fat covers the ribs, making them hardly visible.  
     | e. Pressure is used to feel the spaces between the ribs. |
| 4   | a. The spinous processes can only be found by using some pressure.  
     | b. The presence of full muscles and dense fat cover.  
     | c. Transverse processes are not perceptible.  
     | d. Ribs are not seen. |
| 5   | a. Although it is impossible to feel the spine and transverse processes, a dimple can be observed over the spine.  
     | b. The muscles are exceedingly full and covered in thick fat.  
     | c. Ribs are hidden by excessive fat and are not apparent. |

Daily data collection in YY farm focused on milking, birth, death and slaughtering. In addition, important information such as date, time, tag number, age, ram tag number, ewe tag number, plot, fur color, breed, health/physical condition, and comments are recorded.

The datasheet in both farms has a comments column for workers to additional data that does not belong in any other columns or categories. Consequently, the IT executives can introduce new data tables or sheets into the system for recurring data that may be important for farm management.

C. Proposed Framework on Small Ruminant Breed Lineage Prediction

Based on the findings of this qualitative study, a framework for small ruminant breed lineage prediction was proposed and developed. This framework enables users to input daily data using various formats in their native tongues and jargon. In addition, the proposed entry method is from a mobile application to an online database (see Figure 2).

![Fig. 2 Proposed framework](image)

The proposed framework collects daily input from participating farms in the form of a non-linear and non-sequential dataset, categorized into breeding, fodder, medicine and sanitation. Breeding is the ultimate data because each sheep's ADG and BCS is tracked. Feed type, such as fresh grass, dry hay, pellets, fermented corn, fodder or feed, strongly influences sheep’s daily increase [7]. Furthermore, the ADG data are affected by sheep health [30]. Cleaning the plots with EM spray is one of the best methods for sanitation. Meanwhile, an ill sheep is only medicated upon detection.

Collaboration, workers, and climate are recognized as external factors that potentially influence sheep's daily increase. For instance, the knowledge of farm handlers is increased significantly through a voluntary partnership with colleges or governing authorities, such as the Veterinary department. Furthermore, daily farm operations and livestock management are substantially improved because of the collaboration. Meanwhile, farm productivity is enhanced when the workers apply their prior knowledge and experience as university graduates from related fields, such as agriculture. Finally, sheep's daily gains are impacted by climate changes because animals often fall ill and die from the cold during the rainy season.

Table V summarizes the data collected from this study before further analysis. For example, the demographics of
the sheep and goat population of the YY farm are illustrated in a simplified form of the master data. The F1 Shami sheep are pure Shami breed, while F2 Shami has been cross-bred with another breed.

### Table V: Demographics of YY Farm

<table>
<thead>
<tr>
<th>Breed</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boer Cross</td>
<td>205</td>
</tr>
<tr>
<td>F1 Shami</td>
<td>82</td>
</tr>
<tr>
<td>F2 Shami</td>
<td>2</td>
</tr>
<tr>
<td>Jamnapari Cross</td>
<td>1</td>
</tr>
<tr>
<td>Kampung</td>
<td>11</td>
</tr>
<tr>
<td>Saanen Cross</td>
<td>16</td>
</tr>
<tr>
<td>Shami</td>
<td>40</td>
</tr>
<tr>
<td>Toggenburgh Cross</td>
<td>1</td>
</tr>
</tbody>
</table>

The interviews conducted in this study were analyzed in the qualitative data and analysis phase. The findings indicated that section I correlated with Section IV, where farm owners and/or workers with higher education levels understood the importance of recording information for future reference and to improve farm operation and management. Record keeping enables farmers to identify the existing problems and enhance areas within their control. In addition, Sections II and III suggested that population size determines the farm profits and losses. A small farm that experiences disease outbreaks and deaths will be significantly impacted by the loss of revenue than larger farms because the latter could absorb the losses for longer periods and eventually recover.

Sections V and VI are closely related because farms usually perform pure or crossbreeding to maintain animals’ overall vigor and productivity. Livestock also requires proper nutrition for good health and growth to achieve the desired size in time. Finally, sections VII, VIII, and IX are correlated, where proper medication, isolation and sanitization, and climate control contribute to optimum health conditions for animals to remain healthy.

### IV. Conclusion

This study proposes deep learning models to identify and predict breed lineage and inherited diseases in sheep. Farmers can produce high-quality output, such as sheep meat and milk, by utilizing this method. Problems such as food security, disease-free produce, and high seasonal demand could be addressed effectively. Additionally, this method saves time in identifying the breed lineage of ruminants. The interview analysis also suggested attributes that could contribute to detecting breed lineages, including breed, category, ADG, and BCS. This information is vital as input data to train the FANN algorithm. Therefore, it is recommended that future research adopt this method for other farmed animals.

### References


