

Cattle Disease Prediction using Machine Learning: A Systematic Review

Deeksha G Mesta¹, Janani R², Nydile M N³ and Prof. Naveen Kumar M R⁴

¹⁻⁴Information Science and Engineering, Vidyavardhaka College of Engineering, Mysuru, India

Email: mestadeeksha@gmail.com, janani7337@gmail.com, nydilegowda1220@gmail.com, naveen.is@vvce.ac.in

Abstract—The potential benefits of significant livestock health diagnosis is enormous. Animal husbandry being the most important aspect of India's agricultural growth, farmers suffer on a regular basis as a result of their cattle's bad health and lack of qualified veterinary doctors in their region. At a breakneck pace of Big data and Machine Learning, data mining and analysis are implemented widely in the field of animal husbandry because of the fact that disease and bioterrorism constitute a danger to cattle produce and food supply's safety. In this review paper, we will be comparing the different data science techniques.

Keywords—Data Mining, Machine Learning, Cattle Disease.

I. INTRODUCTION

In view of rising demand driven by rising revenue increased health consciousness, and a growing population, the slowing in the dairy business is highly concerning. According to the National Dairy Development Board (NDDB), milk demand would exceed 180 million tonnes by 2022. To meet demand over the next 15 years, an average exponential increase of 5 million tonnes per year is required, which is more than double the average incremental rate achieved during the previous century[1]. If there is a lack of local milk production, India will have to rely on imports from other countries for its dairy needs. As a result, focusing on local dairy expansion is critical.

Identification of symptoms, cattle diseases, and providing proper treatments is becoming a tedious job in the current medical field. In real-time, it is difficult to handle the cattle disease symptoms and disease types as animals can't explain the problems or pain that they are facing. The manual process of identifying the cattle disease and treatment is too complex and time-consuming and also expensive[2]. The existing system is a manual process where doctors diagnose animals, identify diseases, and give treatment. In foreign countries, they use some advanced systems such as the MYCIN expert system, IBM Watson, etc. The systems which use advanced technologies gather the data, retain it in the database and restore it back whenever required. This makes the medical practitioner job easy to treat cattle diseases in an advanced and effective way.

To our knowledge, no work has been done that focuses on giving a thorough classification of Machine Learning techniques in the prediction of cattle disease. Furthermore, none of the research has thoroughly compared and analyzed machine learning algorithms in the prediction of cattle disease based on the benefits and drawbacks of ML-based techniques

This review paper organised as section(i) Related Works, section(ii) Background, section(iii) Comparative Studies, and section(iv) Conclusion.

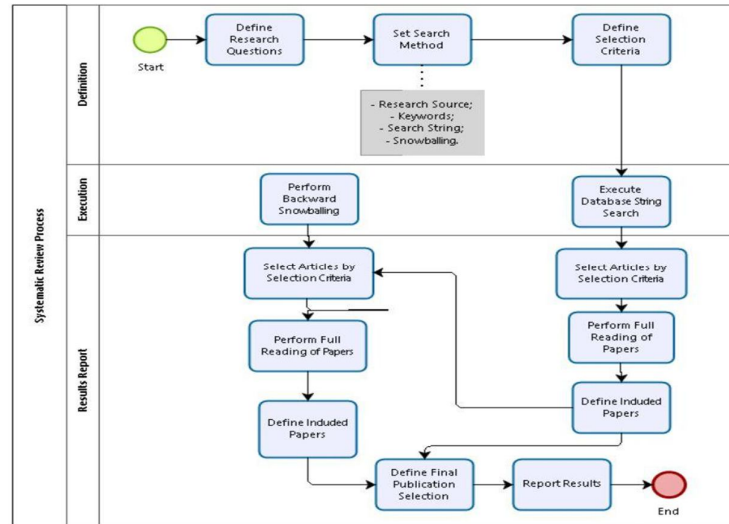


Figure. 1 Systematic Review Process

II. RELATED WORKS

Using Data Mining Techniques and Cattle Health Monitoring System many research papers have been published on disease prediction. Different data mining methods are used for prediction and obtaining dissimilar outcomes for different models.

In their report "Developing Mobile Intelligent System For Cattle Disease Diagnosis and First Aid Action Suggestion," Wiwik Anggraeni, A.Muklason, A.F.Ashari, Wahyu A., and Darminto describe the effort of designing a system, i.e. mobile intelligence, for diagnosing illnesses in cattle and providing first-aid suggestions, which was done using a fuzzy neural network. Another system based on the Fuzzy Neural System was created[1]. In their paper "A Fuzzy Expert System to Diagnose Diseases with Neurological Signs in Domestic Animals," Mahdi Jampour, Mohsen Jampour, Maryam Ashourzadeh, and Mahdi Yaghoobi used the Fuzzy Expert System, which can detect and assess the absence of involvement of any disease with neurological symptoms[2].

Kunja Bihari Swain, Satyasopan Mahato, Meerina patro, Sudeepta Kumar pattnayak in their work on "Cattle health monitoring system using Arduino and LabVIEW for early detection of diseases" they used Arduino UNO, Arduino NANO, Xbee module, and Sensors of various types are used to measure the bodily parameters of cattle[3].

Francisco Gomez, Jeisson Prieto, Juan Galvis, Fausto Moreno, Jimmy Vargas in their work on "Identification of Super-Spreaders of Foot-and-Mouth Disease in the cattle transportation network: The 2018 outbreak case in Cesar (Colombia)" by using cattle transportation network. The centrality of the network may provide information on FMD epidemiology as well as the dynamics of animal movement[4].

"Identification of Acidosis Disease in Cattle Using IoT," by Fatih and Kamil. pH sensor and waterproof temperature sensors were employed with the Internet of Things (IoT) technology. By monitoring and documenting animal diet and behavior on a regular basis, this gadget can detect sickness. Another IoT-based solution has been developed[5]. In their paper "Cattle disease diagnosis using Prediction Technique" Noone Vijay Kishan, Sai Trinath Y, Sandeep Kavalur, Sangamesha V, and Mr. Sumanth Reddy employed IoT and Data Mining. Sensor technologies were used to chart the basic aspects of animal activity, such as heartbeat[6]. The data was aggregated and transmitted to a data mining algorithm to validate the disease. Another IoT-based solution has been developed. Emi Mwaka Katemboh, Raed Abdulla, Vikneswary Jayapal, Sathish Kumar Selvaperumal, Dhakshyani Ratnadurai in their work on "Integrated Animal Health Care Using IoT" used the sensor and the microcontroller that is the Arduino that has been mounted on a LoRa RFM shield and this acts as a node where data is sent to which can predict the disease[7].

Varun Garg, Kritika Garg in their work on "Early-Stage Disease Detection Platform in Cattles" used Neural Network which is used to provide economic medical solutions[8].

Lauren Wisnieski, Steven Pierce, Bo Norby, Tyler Becker, Jeff Gandy, and Lorraine Sordillo employed Support Vector Machine in their project "Cattle Disease Auxiliary Diagnosis and Treatment System Based on Data Analysis and Mining." They used electronic data and data analysis technology to construct an intelligent cow sickness detection system[9]. Another solution to the problem was established by Daksh Ashar, Amit Kanojia, and Rahul Parihar in their work on "Livestock Sickness Prediction System," which employed Support Vector Classifier to predict the data and inform the livestock owner if the predicted disease may cause sudden death[10]. In this paper S. Shahinfar, D. Page, J. Guenther, V. Cabrera, P. Fricke, and K. Weigel, "Prediction of insemination outcomes in Holstein dairy cattle using alternative machine learning algorithms," have addressed reproduction in cattle by developing tool to make decisions on cattle breeding. The Random forest algorithm has given better results than naïve Bayes, Bayesian network and decision tree algorithms. The parameters which they have considered are gain ratio (GR) Herd-year-month (HYM) and Information gain (IG). They have achieved an accuracy of 75.6 and 73.6 in primiparous and multiparous respectively[11].

In this paper A. D. Sunny, S. Kulshreshtha, S. Singh, M. Ba, and H. Sarojadevi, "Disease Diagnosis System By Exploring Machine Learning Algorithms," have developed a system which diagnose the cattle disease based on the input text data. The authors have compared different machine learning algorithms like naïve bayes, k-nearest neighbor and aprior algorithms where Aprior algorithm has yielded good result and naïve bayes are also good depending upon the dataset which they have considered[12].

In this paper L. Wan, W. Bao, A. Disease, D. Expert, S. Based, and T. Ifip, "Animal Disease Diagnoses Expert System have proved the practicality of Support vector machine (SVM) which can be used to livestock breeding of cattle and developed the expert system that can diagnose the diseases. The experiment carried out on a small dataset which yielded the good result which is the limitation of the work[13].

In this article D. Wu *et al.*, "Lameness detection of dairy cows based on the YOLOv3 deep learning algorithm and a relative step size characteristic vector," have proposed yolov3 deep learning algorithm and relative step size characteristics to identify the lameness detection in cows. Lameness detection was done using the videos where LSTM, SVM, KNN and DTC are used to classify the lame and non-lame legs of the cow. The LSTM outperformed all other algorithms with 98.57 accuracy. The yolov3 model is used to identify the target position of the legs in videos. The authors have proposed the extension of the work can be used for automatic detection of lameness detection in cows. Lameness can increase the risk of premature elimination of dairy cows, and can cause huge economic losses. Data on the acceleration when cows walked were collected by acceleration sensors fixed to the hind legs, and gait analysis was carried out[14]. In order to realise the detection and classification of lame cows and nonlame cows, a method based on the YOLOv3 deep learning algorithm and a relative step size characteristic vector is proposed in this study. Intelligent lameness detection is important for improving cow welfare. A method based on YOLOv3 deep learning algorithm and relative step size characteristic vector is proposed to classify lame and non-lame cows. Relative step sizes of cow's front and rear legs were calculated based on leg coordinates, and the relative step size characteristic vector was constructed. Finally, a trained Long Short-Term Memory (LSTM) classification model was used to classify lame and non-lame cows based on the characteristic vector.

In this paper, H. Sadeghi, H. S. Braun, B. Panti, G. Opsomer, and O. B. Pascottini, "Validation of a deep learning-based image analysis system to diagnose subclinical endometritis in dairy cows," have assessed the polymorph nuclear leukocyte (PMN) proportion of endometrial samples using deep learning. When compared to naphthol and Diff-Quick the deep learning algorithms have yielded good results. In endometrial cytology, the deep learning algorithms are reliable and deduce the bias in subclinical endometriosis[15].

In this article, V. Sturm, D. Efrosinin, M. Öhlschuster, E. Gusterer, M. Drillich, and M. Iwersen, "Combination of sensor data and health monitoring for early detection of subclinical Ketosis in dairy cows," have proposed an algorithm that can predict the occurrence of subclinical ketosis in cows. The flexible algorithm uses TSC (TIME SERIES CLASSIFIER) with a traditional classification algorithm to classify the data from the sensors. The results show that the method used for automatic detection of diseases in dairy cows performed good. Future elaborated machine learning algorithms can be used to early detect subclinical ketosis[16].

In this paper, N. Misk, T. Misk, A. El-Khamary, and M. Semeika, "A retrospective study of surgical affections of mammary glands in cattle and buffaloes and their management in the field," have used non deep learning algorithm like random forests and support vector machines to classify the data. The experiment was conducted on image data where random forest achieved the accuracy of 95.20%. Since data is the challenge to train the deep models and it consumes more computation the authors have adopted non- deep models for both classifying and training[17].

In this article, A. Hidalgo, F. Zouari, H. Knijn, and S. van der Beek, "Prediction of postpartum diseases of dairy cattle using machine learning," have used machine learning algorithm to predict occurrence of postpartum

diseases by using prepartum data. In this study they have used random forest algorithm to predict the postpartum diseases. The data set were real time data taken from Dutch farms from 2009 to 2016. The algorithm outperformed other state of art algorithms in machine learning[18].

In this article, “Algorithms for Detecting Cattle Diseases at Early Stages and for Making Diagnoses and Related Recommendations” have developed an algorithm to identify infectious, parasitic, and nervous diseases in large horned cattle. They have taken a data from 5 farms. They wrote the code in a software package “Matlab R2019b”[19].

In this article, Y. Ali, A. Farooq, T. M. Alam, M. S. Farooq, M. J. Awan, and T. I. Baig, “Detection of Schistosomiasis Factors Using Association Rule Mining” have developed detection technique by using Apriori Algorithm to detect the Schistosomiasis disease in an early stage. They collected dataset from the record of 4136 (2674 male and 1462 female) patients, diagnosis with advance disease in them[20] and another article based on Apriori algorithm, M. Shamsul Arifin, M. Golam Kibria, A. Firoze, M. Ashraful Amini, and H. Yan, “Detection and Prediction of Frequent Diseases in India through Association Technique using Apriori Algorithm and Random Forest Regression” has developed a technique to identify the frequency of the data[21] and there is another article which is based on clustering method, “Prediction of Heart Disease by Clustering and Classification Techniques Prediction of Heart Disease by Clustering and Classification Techniques” have developed an system from which the heart disease can be predicted[22].

In this article, A. A. Elngar, R. Kumar, A. Hayat, and P. Churi, “Intelligent System for Skin Disease Prediction using Machine Learning” have developed a system for detecting the skin disease and provide treatment for the same. They have used CNN and SVM algorithm for detecting the skin disorders[23] and there is another prediction system article, M. Shamsul Arifin, M. Golam Kibria, A. Firoze, M. Ashraful Amini, and H. Yan, “Dermatological disease diagnosis using color-skin images” where they used backpropagation neural network to detect the skin diseases in an early stage[24].

In this article, N. Wagner *et al*, “Machine learning to detect behavioural anomalies in dairy cows under subacute ruminal acidosis” to detect abnormal behaviour in dairy cows suffering from subacute ruminal acidosis. They used K Nearest Neighbours for Regression (KNNR); Decision Tree for Regression (DTR); MultiLayer Perceptron (MLP); Long Short-Term Memory (LSTM) algorithms[25].

In this article, W. Xu *et al*, “Prediction of metabolic status of dairy cows in early lactation with on-farm cow data and machine learning algorithms” here based on plasma metabolites and metabolic hormones, dairy cows might be grouped according to their metabolic condition. They used Decision Tree, Naive Bayes, Bayesian Network, Support Vector Machine, ANN, Bootstrap Aggregation, Random Forest, and K-Nearest Neighbor[26].

“Identifying Health Status in Grazing Dairy Cows from Milk Mid-Infrared Spectroscopy by Using Machine Learning Methods”, B. C. Hernández, N. Lopez-Villalobos, and M. Vignes by have developed a technique in which they can identify the health status of cows. They have used Random forest, support vector machine, neural network, convolutional neural network[27].

“Assessing machine learning techniques in forecasting lumpy skin disease occurrence based on meteorological and geospatial features” E. Afshari Safavi, using ANN may be used to accurately predict the occurrence of LSDV infection using geographical and meteorological data[28].

“MasPA: A Machine Learning Application to Predict Risk of Mastitis in Cattle from AMS Sensor Data” N. Abdul Ghafoor and B. Sitkowska, used random forest algorithm can forecast the likelihood of mastitis in cattle based on their udder's inhale and exhale limitations and temperature[29].

In article, U I Bajwa*, S Alam, N ul Haq, N I Ratyal and M Waqas Anwar, V, “Skin Disease Classification using Neural Network” main goal of their research is to create an automated method for skin disease categorization that can assist a general practitioner in detecting and classifying lesions. They used ANN for classification[30].

In machine learning, a dataset is essentially a collection of data bits that can be analyzed and predicted by a computer as a single entity. This implies that the data gathered should be homogeneous and understandable to a machine that does not see data in the same manner that people do. After gathering the data, it's crucial to clean and complete it, as well as label it with significant computer-readable tags. Additionally, a useful dataset should meet specific quality and quantity requirements. The dataset must be relevant and well-balanced for a smooth and quick training experience.

III. BACKGROUND

A) MACHINE LEARNING

Machine learning is classified as a multidisciplinary branch of computer science. The main goal of machine learning is to find undiscovered facts, hidden insights, correlations, patterns, and knowledge in massive data sets

that are difficult to find and identify using traditional statistical approaches; the design and development of data-learning systems. The main steps involved in machine learning are preprocessing, modeling, and post-processing of data.

Machine learning is classified into three categories:

Supervised Machine learning:-The algorithms that use externally given cases to build a broad hypothesis and make use of it in the future for predictions are called supervised machine learning algorithms. One of the distinguishing features is their capacity to use annotated data. SML algorithms use class labels for the classification process. There is a wide range of supervised algorithms some of them are: Linear Classifiers, Support Vector Machine Logistic Regression, Naïve Bayes Classifier, Perceptron, Quadratic Classifiers, Decision Tree, K-Means Clustering, Boosting, Random Forest (RF); Bayesian Networks, Neural Networks and so on

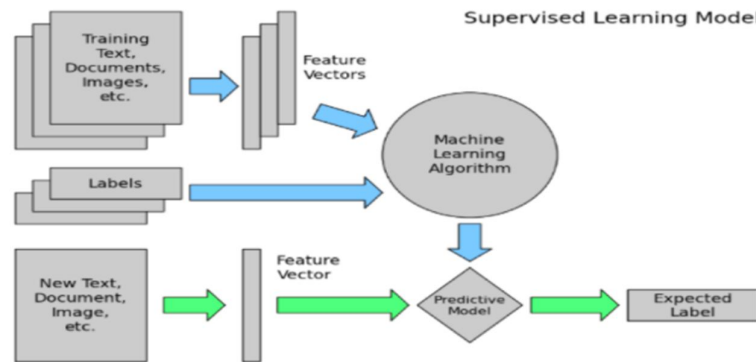


Figure. 2 Supervised machine learning model

Support Vector Machine: SVM being a supervised machine learning algorithm was formulated by Vapnik et al. in 1995. Support Vector Machine (SVM) models are similar to traditional multilayer perceptron neural networks. The concept of “margin”(separates two different classes) is used in SVM. The margin is maximized to produce the biggest possible space between the separating hyperplane and instances and thus reduces the predicted generalization error. SVM first calculates the low dimensional space. Then the kernel function is used to map high dimensional feature space and input space. Finally, in the high dimensional feature space, the best-classified hyperplane is built[9].

Input: Consider the training dataset T with multiple symptoms, where symptoms of diseases and category of diseases are denoted by x_i and y_i respectively. $f(x)$ is the decision function which produces the optimal classification plane is shown below

Neural Networks: Neural Networks (NN) can execute many classifications and/or regression tasks at the same time, despite the fact that most networks can only do one. In the great majority of situations, there will be a single output variable for the network. However, with respect to many state classification problems, it equates to the number of output units[8].

$$f(x) = \text{sgn} \left(\sum_{i=1}^n a_i^* y_i (x \cdot x_i) + b^* \right)$$

$$b^* = y_j - \sum_{i=1}^l a_i^* y_i (x \cdot x_i)$$

Figure. 3 Neural Network Equation

Fuzzy Neural Networks: A simple fuzzy neural network produces a Gaussian function by narrowing down the set of input vectors, which is then correlated to the target label. Secondly, in order to obtain a fuzzy truth, the previously found gaussian function is used, hence classifying the unknown feature vector. Thirdly, to find the class of input feature vectors, fuzzy truths are maximized [3].

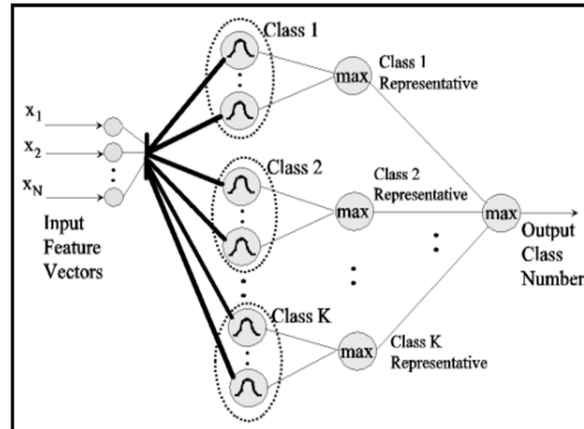


Figure. 4 Simple Fuzzy Neural Network

Naive Bayes Classifier: Naive Bayes Classifier is a classification method based on Bayes' Theorem and the condition of predictor independence. A Naive Bayes classifier, in simple terms, asserts that the existence of one feature in a class is independent of the presence of any other feature. The Naive Bayes model is simple to construct and is exceptionally effective for huge datasets. Naive Bayes is recognized to outperform even the most advanced classification systems due to its simplicity.

$$P(c | x) = \frac{P(x | c) P(c)}{P(x)}$$

Likelihood
Class Prior Probability
Posterior Probability
Predictor Prior Probability

$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \dots \times P(x_n | c) \times P(c)$$

Figure. 5 Naive Bayes Equation

Random Forest Classifier: A random forest is a meta estimator that uses averaging to increase predicted accuracy and control over-fitting by fitting a number of decision tree classifiers on various sub-samples of the dataset. If bootstrap=True (default), the sub-sample size is regulated by the max samples argument; otherwise, the entire dataset is utilized to create each tree.

Random Forest Classifier

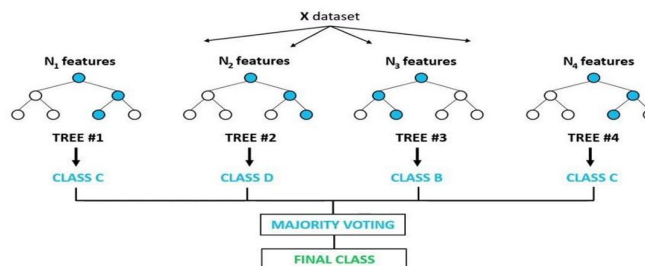


Figure. 6 Random Forest Classifier

Decision Tree Classifier: Decision Tree is a supervised learning technique that may be used to solve both classification and regression problems, however, it is most commonly used to solve classification issues. Internal nodes represent dataset attributes, branches represent decision rules, and each leaf node provides the result in this tree-structured classifier. The Decision Node and the Leaf Node are the two nodes of a Decision tree. Leaf nodes are the output of those decisions and do not contain any more branches, whereas Decision nodes are used to make any decision and have several branches. The decisions or tests are made based on the characteristics of the given dataset.

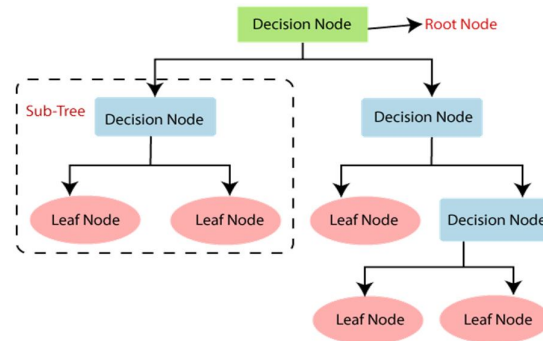


Figure. 7 Decision Tree Classifier

B) Unsupervised Learning

This Learning technique will not make use of training datasets to build models, but instead, find the hidden patterns in the given dataset. In this technique, class labels are not present. Unsupervised techniques cannot be applied to classification or regression tasks. It covers the underlying patterns based on similarities to display the dataset in a compact fashion. Different algorithms that come under unsupervised methodologies are “eclat algorithm”, “FP Growth algorithm”, “AIT algorithm” and many more.

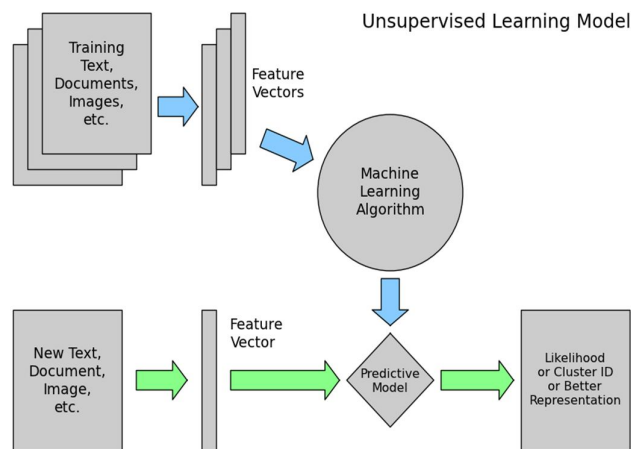


Figure. 8 Unsupervised learning model

C) Reinforcement Learning

Machine Learning includes the field of reinforcement learning. It's all about taking the right steps to maximize your reward in a given situation. It is used by a variety of software and computers to determine the best feasible action or path in a given situation. Reinforcement learning differs from supervised learning in that supervised learning includes the answer key, allowing the model to be trained with the correct answer, whereas reinforcement learning does not include an answer and instead relies on the reinforcement agent to decide what to do to complete the task. It is obligated to learn from its experience in the absence of a training dataset.

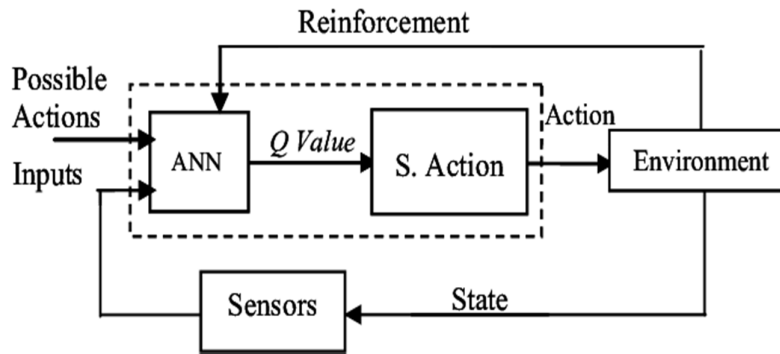


Figure. 9 Reinforcement Learning

IV. COMPARATIVE STUDY

Author	Year	Methodology	Dataset	Description	Limitation
M. Jampour, M. Jampour, M. Ashourzadeh, and M. Yaghoobi	2011	Fuzzy Expert System	20 test board of animal	The fuzzy logic system method is used to determine and compute the absence of participation of each probable disease with neurological symptoms	Used only to Diagnose Diseases with Neurological Signs
M. Shamsul Arifin, M. Golam Kibria, A. Firoze, M. Ashraful Amini, and H. Yan	2012	Backpropagation Neural Network	Image dataset	The system uses colour image processing techniques, k-means clustering, and colour gradient approaches to identify the sick skin in terms of machine intervention. The system uses feedforward backpropagation artificial neural networks to classify diseases.	This system gave best results for image dataset only.
W. Anggraeni, A. Muklason, A. F. Ashari, A. Wahyu, and Darminto	2013	Fuzzy Neural Network	250 medical records for three different cattle diseases i.e. Mastitis, foot and mouth disease, and Septicemia Epizootic.	The purpose of this study is to demonstrate the work that has gone into establishing a mobile intelligent system for diagnosing cattle ailments and suggesting first-aid actions	Android app is developed, visualization problem on training data-sets.
S. Shahinfar, D. Page, J. Guenther, V. Cabrera, P. Fricke, and K. Weigel	2014	Random forest algorithm	26 Wisconsin dairy farms that were enrolled in the Alta Genetics.	It is very difficult to predict the fertilization results of lactating cows. we can use information about health, breeding history, production level, and other environmental characteristics to identify a very fertile subgroup of heads.	The outcome for individual lactating dairy cows is extremely difficult, information regarding health, reproductive history, production level, and other environmental features can be used to identify highly fertile subsets of cows
L. Wan, W. Bao, A. Disease, D.	2014	Support Vector Machine		Supported vector machines (SVMs) can be used for breeding and developing expert systems	The difficulty of the rapid diagnosis processed because of the complexity

Expert, S. Based, and T. Ifip				that can diagnose diseases. Testing conducted on a small data set for precise results is the limit of the work.	and diversity of the symptoms of animal diseases.
V. Garg, K. Garg	2016	Neural Network	Differing types of skin diseases from color Image	Machine learning was applied to detect cow ailments, resulting in a cost-effective medical treatment	Achieved accuracy is more only for skin related diseases.
K. B. Swain, S. Mahato, M. Patro, S. K. Pattanayak	2017	Arduino UNO, Arduino NANO, Xbee module	Data from sensors	Designed a system that allows farmers to monitor and compare the current health parameters of their livestock to the conventional healthy reference parameters	Only used to monitor the cattle health condition.
F. Gomez, J. Prieto, J. Galvis, F. Moreno, J. Vargas	2018	Cattle Transportation Network (ranking nodes)		The cow transportation network is computed using several centrality measures, each of which may provide useful information about the animal movement	Only used to monitor the “foot-and-mouth disease”. Small data-sets used
A. D. Sunny, S. Kulshreshtha, S. Singh, M. Ba, and H. Sarojadevi	2018	Naive Bayes and Apriori	Disease-Symptom Knowledge data set prepared by New York Presbyterian Hospital of patients admitted during 2004	The authors compared different machine learning algorithms like naive algorithm, neighbor algorithms, and aprior where the Aprior algorithm performed well and the naive algorithms were also good based on the data set they studied. consider.	The disease being predicted based on the input symptoms, using multiple algorithms.
A. Hidalgo, F. Zouari, H. Knijn, and S. van der Beek	2018	Random forest	The data set were real time data taken from Dutch farms from 2009 to 2016.	Predict the occurrence of postpartum diseases using prenatal data. In this study, they used a random forest algorithm to predict postpartum morbidity. This algorithm has performed better than other modern algorithms in machine learning.	We envisage that big data is out there to stay and it has different applications.
F. Basciftci, K. A. Gunduz	2019	Internet of Things (IoT) pH sensor, Waterproof temperature sensors	Data from sensors	Regular monitoring and documenting of animal nutrition and daily behavior can help to identify the condition	Only used to monitor one disease “Acidosis Disease”. Sensors used for monitoring leads to less accurate results.
Y. Ali, A. Farooq, T. M. Alam, M. S. Farooq, M. J. Awan, and T. I. Baig	2019	Apriori Algorithm	Data from 4136 patients were diagnosed with advanced schistosomiasis illness (2674 males and 1462 females).	The major goal of this study is to investigate the disease in depth, assess connections between disease factors, and analyse the disease using association rules in schistosomiasis patients.	This system disease detection was only for the human beings they didn't concentrate on animals.
Singh, Reetu Rajesh, E.	2019	K-Means Clustering	clinical dataset	Goal was to predict the heart diseases by clustering and classifying the data.	The accuracy of clustering and classification is reduced when some data remain un-clustered or wrongly clustered.
W. Xu <i>et al.</i>	2019	Random Forest	collected from 2	By using machine learning	Accuracy was low.

		and SVM	earlier studies with protocol number 2010026 (study 1; van Kneegsel et al., 2014; Chen et al., 2016) and 2014125 (study 2; van Hoeij et al., 2017). The	algorithm, Predicting of metabolic status of dairy cows in early lactation with on-farm cow data.	
L. Niu, C. Yang, Y. Du, L. Qin, B. Li	2020	Support Vector Machine (SVM)	Multi-source cow electronic medical record data	To create the intelligent diagnosis system for cattle diseases, a huge quantity of multi-source cow electronic medical record data was collected and data analysis and mining technologies was applied	Large dataset is required to give accurate results. Only when the symptoms are less accuracy is more.
D. Wu <i>et al.</i>	2020	Support Vector Machine	Walking crawl acceleration data is collected by accelerationSensors were attached to the hind legs	The lameness detection was performed using videos in which LSTM, SVM, KNN and DTC were used to classify lame and non-crippled cows. The yolov3 model is used to determine the target position of the foot in the video. They propose that an extension of the work could be used to automatically detect lameness in cows.	The algorithm proposed in this study could effectively overcome different forms of interference such as occlusion, light change and near-view objects, and can achieve rapid and accurate detection of leg targets.
V. Sturm, D. Efrosinin, M. Öhlschuster, E. Gusterer, M. Drillich, and M. Iwersen	2020	Naive Bayes	Dataset approved from University of Veterinary Medicine Vienna, Austria and Slovakian Regional Veterinary Food Administration.	This may predict the occurrence of subclinical ketosis in cows. The flexible algorithm uses TSC with traditional classification algorithm to classify sensor data. The results show that the automatic disease detection method on dairy cows gives good results. In the future, developed machine learning algorithms can be used for early detection of subclinical ketosis.	Early detection of SCK is not possible.
M. Bhanu Sridhar P. Aiswarya L. Kavitha	2020	Apriori Algorithm and Random Forest Regression	National Health Profile by the Central Bureau of Health Intelligence, Government of India, along with the collaboration of World Health Organization (WHO)	Analysing the data to predict the diseases.	Dataset they have taken was very less.
N. Wagner <i>et al.</i>	2020	K Nearest Neighbours for Regression, Decision Tree for Regression, MultiLayer Perceptron, Long Short-Term Memory	Clinical Dataset	Using algorithms they have analyzed the disease behavior characteristics.	Less accuracy
U I Baiwa*, S	2020	ANN	588 images containing 6907	Goal of this automated system is to classify the disease in skin.	They didn't analyze of different skin tones on the

Alam, N u Haq, N I Ratval and M W Anwar			lesion regions		performance of lesions detection and classification system
N. V. Kishan, S. T. Y, S. Kavalur	2021	Random forest Classifier, Decision Tree Classifier, Naïve Bayes Classifier	95 symptoms of cow	The basic characteristics of animal activity, such as heartbeat, were charted using sensor technologies. To validate the disease, the data was compiled and fed into a data mining model	Doesn't predicts the relationship between symptoms, disease and treatments.
D. Ashar, A. Kanojia, R. Parihar, S. Kudoo	2021	Support Vector Classifier	Animal Dataset	A multi-class classification algorithm was utilised to provide preventative actions based on the disease forecast, as well as to inform the livestock owner if the disease indicated could result in a sudden death	Accuracy achieved was less. Time consuming as the prediction is made for goat, sheep and cow using a single model.
D. Y. Pavkin, A. S. Dorokhov, F. E. Vladimirov, I. M. Dovlatov, and K. S. Lyalin	2021		5 large farms: "Ecocorm" Ltd. (Voronezh region, Russia), "Research Center of Traditional Technologies" Ltd. (Ivanovo region, Russia), Agricultural enterprise "Novomarkovskoye" Ltd. (Voronezh region, Russia), Agricultural enterprise "Voshchazhnikovo" Ltd. (Yaroslavl region, Russia)	They looked at the relationships between the temperature and pH of the cows' rumens, their locomotor activity, and environmental factors including air temperature and relative humidity in the cowsheds. After that, they created graphs of functionalities that are intertwined for the first time, algorithms were compiled as a result of the research. Detecting viral, parasitic, and neurological illnesses was possible.	They concentrated only on the cattle not on the other animals.
A. A. Elngar, R. Kumar, A. Hayat, and P. Churi	2021	SVM and CNN	3000 images which collected from a lot of sources like Beni-Suef University Hospital, Cairo University Hospital	To detect the skin disease at early stage they have developed a prediction system from which they will be able to get the treatment for the particular skin disease.	This was helpful for only the image data.
M. Shamsul Arifin, M. Golam Kibria, A. Firoze, M. Ashraful Amini, and H. Yan	2021	Apriori algorithm and Random Forest Classifier	National Health Profile by the Central Bureau of Health Intelligence, Government of India, along with the collaboration of World Health Organization (WHO), various Central Ministries and all	a strategy for extracting interesting patterns from health data using the Association Method.	This work can be expanded to include pandemics that are unforeseen and predictions that can help a country prepare for the invisible adversary.

			the state/union territory health departments		
B. C. Hernández, N. Lopez-Villalobos, and M. Vignes	2021	random forest, support vector machine, neural network (NN), convolutional neural network and	1909 milk samples	Identifying the health status of the cows.	Less number of dataset
N. Abdul Ghafoor and B. Sitkowska	2021	Random Forest	6600 entries (three entries per cattle) for cattle with 15 attributes	To predict the risk of mastitis in cattle	Less accuracy
H. Sadeghi, H. S. Braun, B. Panti, G. Opsomer, and O. B. Pascottini	2022	Oculyze MUH system	A large dataset of cells of the different classes has been used to obtain the retrained model	This system evaluates the polymorphonuclear leukocyte (PMN) percentage of endometrial samples using deep learning. Compared to naphthol and DiffQuick, the deep learning algorithms perform well. In endometrial cytology, deep learning algorithms are reliable and inferentially biased in subclinical endometriosis.	To improve the performance and robustness of the Oculyze MUH, an expansion of the dataset to train the system with an extensive set of images is warranted.
N. Misk, T. Misk, A. El-Khamary, and M. Semeika	2022	Random Forest and Support Vector Machine	Image data	The experiment was conducted on image data. Since the data is challenging to train deep models and it is more computationally intensive, the authors used non-deep models for classification and training.	The results of this study demonstrate that treatment of most teat and udder affections is feasible in field settings and veterinarians can expect satisfactory results after interventions.
E. Afshari Safavi	2022	ANN	LSDV infection data	Using geospatial and climatic parameters, ANN can be used to predict the occurrence of LSDV infection with great precision. In locations where the occurrence of LSDV infection is a high risk, using the predicting capacity of these methodologies could be very useful in implementing screening and awareness initiatives, as well as taking preventive measures like vaccination. Keywords	Difficult to predict for larger dataset.

V. PROPOSED METHOD

Disease remains the major cause of death. Several cattle have died as a result of an illness for which the proper treatment is unknown. So we did a comparison research and discovered that in many research papers they have used a different algorithm such as Aprior. As a result, we'll use a number of unsupervised algorithms to provide accurate results. The proposed system takes a symptom as input and predicts the disease as well as therapy options. We used a few dataset to try to predict illness. The eclat algorithm was found to be 0.38 millisecond efficient and 86.2%.

VI. PRILIMINARY RESULT

Cattle disease can be diagnosed using machine learning algorithms, which aids veterinary doctors in providing therapy. We must use the algorithm on the training dataset. As we discovered throughout our comparison

research, majority of them used a supervised technique to predict sickness. As a result, we will attempt to build the project by applying an unsupervised learning technique

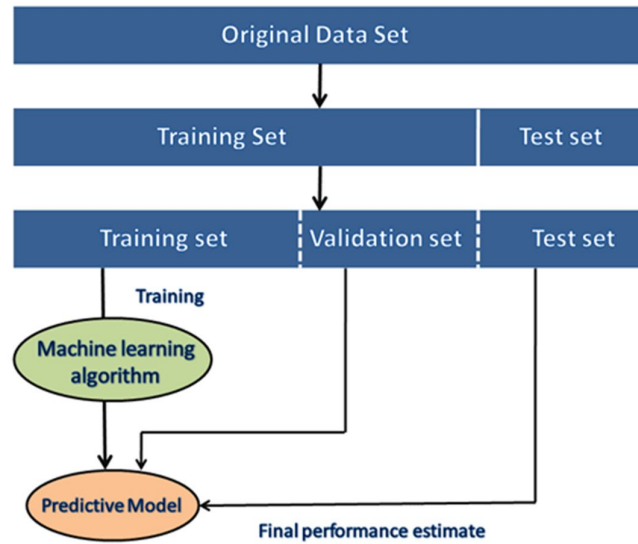


Figure. 10 Dataset Classification Flowchart

VII. CONCLUSION

Several cattle diseases result in a decrease in milk yield. Many researches were conducted to predict these diseases with the help of machine learning models. But the performance of the technique depends upon the missing and noisy data in the dataset. Accuracy and confusion matrix relies upon the above-mentioned factors which can be avoided by using accurate statistical approaches to clean the dataset. We have reviewed the drawbacks and advantages of each algorithm in the comparative study section. According to the survey Support Vector Machines produced more accurate results and were found useful in many areas of animal husbandry when compared to neural networks and IOT. The performance of algorithms vary greatly between different datasets because of their data source, techniques used to preprocess the data and selection the parameters. But they can use any algorithm which is suitable for their dataset. However, there is still scope in the veterinary field for experts to work. We will be implementing a project where we will be comparing our algorithm with Aprior algorithm.

REFERENCES

- [1] W. Anggraeni, A. Muklason, A. F. Ashari, A. Wahyu, and Darminto, "Developing mobile intelligent system for cattle disease diagnosis and first aid action suggestion," Proc. - 2013 7th Int. Conf. Complex, Intelligent, Softw. Intensive Syst. CISIS 2013, pp. 117–121, 2013, doi: 10.1109/CISIS.2013.27.
- [2] M. Jampour, M. Jampour, M. Ashourzadeh, and M. Yaghoobi, "A fuzzy expert system to diagnose diseases with neurological signs in domestic animal," Proc. - 2011 8th Int. Conf. Inf. Technol. New Gener. ITNG 2011, pp. 1021–1024, 2011, doi: 10.1109/ITNG.2011.217.
- [3] K. B. Swain, S. Mahato, M. Patro, and S. K. Pattanayak, "Cattle health monitoring system using Arduino and LabVIEW for early detection of diseases," Proc. 2017 3rd IEEE Int. Conf. Sensing, Signal Process. Secur. ICSSS 2017, no. May, pp. 79–82, 2017, doi: 10.1109/SSPS.2017.8071569.
- [4] F. Gomez, J. Prieto, J. Galvis, F. Moreno, and J. Vargas, "Identification of Super-Spreaders of Foot-And-Mouth Disease in the cattle transportation network: The 2018 outbreak case in Cesar (Colombia)," Proc. 2019 IEEE World Conf. Complex Syst. WCCS 2019, pp. 5–10, 2019, doi: 10.1109/ICoCS.2019.8930765.
- [5] F. Basciftci and K. A. Gunduz, "Identification of Acidosis Disease in Cattle Using IoT," UBMK 2019 - Proceedings, 4th Int. Conf. Comput. Sci. Eng., pp. 58–62, 2019, doi: 10.1109/UBMK.2019.8907035.

- [6] N. V. Kishan, S. T. Y, and S. Kavalur, "Cattle disease identification using Prediction Techniques," no. 3, pp. 2764–2770, 2021.
- [7] E. M. Katemboh, R. Abdulla, V. Jayapal, S. K. Selvaperumal, and D. Ratnadurai, "Integrated animal health care using IoT," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 1, pp. 42–56, 2020.
- [8] V. Garg and K. Garg, "Early Stage Disease Detection Platform in," vol. 3, no. 11, pp. 13–15, 2016.
- [9] L. Niu, C. Yang, Y. Du, L. Qin, and B. Li, "Cattle disease auxiliary diagnosis and treatment system based on data analysis and mining," 2020 5th Int. Conf. Comput. Commun. Syst. ICCCS 2020, pp. 24–27, 2020, doi: 10.1109/ICCCS49078.2020.9118590.
- [10] D. Ashar, A. Kanojia, R. Parihar, and S. Kudoo, "Livestock Disease Prediction System," *VIVA-Tech Int. J. Res. Innov.*, vol. 1, no. 4, pp. 1–3, 2021.
- [11] S. Shahinfar, D. Page, J. Guenther, V. Cabrera, P. Fricke, and K. Weigel, "Prediction of insemination outcomes in Holstein dairy cattle using alternative machine learning algorithms," *J. Dairy Sci.*, vol. 97, no. 2, pp. 731–742, 2014, doi: 10.3168/jds.2013-6693.
- [12] A. D. Sunny, S. Kulshreshtha, S. Singh, M. Ba, and H. Sarojadevi, "Disease Diagnosis System By Exploring Machine Learning Algorithms," *Int. J. Innov. Eng. Technol.*, vol. 10, no. 2, pp. 14–21, 2018.
- [13] L. Wan, W. Bao, A. Disease, D. Expert, S. Based, and T. Ifip, "Animal Disease Diagnoses Expert System Based on Long Wan , Wenxing Bao To cite this version : HAL Id : hal-01055402," pp. 0–8, 2014.
- [14] D. Wu et al., "Lameness detection of dairy cows based on the YOLOv3 deep learning algorithm and a relative step size characteristic vector," *Biosyst. Eng.*, vol. 189, pp. 150–163, 2020, doi: 10.1016/j.biosystemseng.2019.11.017.
- [15] H. Sadeghi, H. S. Braun, B. Panti, G. Opsomer, and O. B. Pascottini, "Validation of a deep learning-based image analysis system to diagnose subclinical endometritis in dairy cows," *PLoS One*, vol. 17, no. 1 January, pp. 1–14, 2022, doi: 10.1371/journal.pone.0263409.
- [16] V. Sturm, D. Efronin, M. Öhlschuster, E. Gusterer, M. Drillich, and M. Iwersen, "Combination of sensor data and health monitoring for early detection of subclinical Ketosis in dairy cows," *Sensors (Switzerland)*, vol. 20, no. 5, pp. 1–19, 2020, doi: 10.3390/s20051484.
- [17] N. Misk, T. Misk, A. El-Khamary, and M. Semeika, "A retrospective study of surgical affections of mammary glands in cattle and buffaloes and their management in the field," *J. Vet. Med. Sci.*, vol. 80, no. 10, pp. 1576–1583, 2018, doi: 10.1292/jvms.17-0586.
- [18] A. Hidalgo, F. Zouari, H. Knijn, and S. van der Beek, "Prediction of postpartum diseases of dairy cattle using machine learning," *Proc. World Congr. Genet. Appl. to Livest. Prod.*, vol. 11, no. February, p. 104, 2018, [Online]. Available: <http://www.wcgalp.org/proceedings/2018/prediction-postpartum-diseases-dairy-cattle-using-machine-learning>.
- [19] D. Y. Pavkin, A. S. Dorokhov, F. E. Vladimirov, I. M. Dovlatov, and K. S. Lyalin, "Algorithms for detecting cattle diseases at early stages and for making diagnoses and related recommendations," *Appl. Sci.*, vol. 11, no. 23, 2021, doi: 10.3390/app112311148.
- [20] Y. Ali, A. Farooq, T. M. Alam, M. S. Farooq, M. J. Awan, and T. I. Baig, "Detection of schistosomiasis factors using association rule mining," *IEEE Access*, vol. 7, pp. 186108–186114, 2019, doi: 10.1109/ACCESS.2019.2956020.
- [21] M. Bhanu Sridhar, P. Aiswarya, and L. Kavitha, "Detection and Prediction of Frequent Diseases in India through Association Technique using Apriori Algorithm and Random Forest Regression," *Int. J. Eng. Res.*, vol. V9, no. 03, pp. 386–393, 2020, doi: 10.17577/ijertv9is030427.
- [22] R. Singh and E. Rajesh, "Prediction of Heart Disease by Clustering and Classification Techniques," *Int. J. Comput. Sci. Eng.*, vol. 7, no. 5, pp. 861–866, 2019, doi: 10.26438/ijcse/v7i5.861866.
- [23] A. A. Elngar, R. Kumar, A. Hayat, and P. Churi, "Intelligent System for Skin Disease Prediction using Machine Learning," *J. Phys. Conf. Ser.*, vol. 1998, no. 1, 2021, doi: 10.1088/1742-6596/1998/1/012037.
- [24] M. Shamsul Arifin, M. Golam Kibria, A. Firoze, M. Ashraf Amini, and H. Yan, "Dermatological disease diagnosis using color-skin images," *Proc. - Int. Conf. Mach. Learn. Cybern.*, vol. 5, pp. 1675–1680, 2012, doi: 10.1109/ICMLC.2012.6359626.
- [25] N. Wagner et al., "Machine learning to detect behavioural anomalies in dairy cows under subacute ruminal acidosis," *Comput. Electron. Agric.*, vol. 170, p. 105233, 2020, doi: 10.1016/j.compag.2020.105233.
- [26] W. Xu et al., "Prediction of metabolic status of dairy cows in early lactation with on-farm cow data and machine learning algorithms," *J. Dairy Sci.*, vol. 102, no. 11, pp. 10186–10201, 2019, doi: 10.3168/jds.2018-15791.
- [27] B. C. Hernández, N. Lopez-Villalobos, and M. Vignes, "Identifying health status in grazing dairy cows from milk mid-infrared spectroscopy by using machine learning methods," *Animals*, vol. 11, no. 8, 2021, doi: 10.3390/ani11082154.
- [28] E. Afshari Safavi, "Assessing machine learning techniques in forecasting lumpy skin disease occurrence based on meteorological and geospatial features," *Trop. Anim. Health Prod.*, vol. 54, no. 1, 2022, doi: 10.1007/s11250-022-03073-2.
- [29] N. Abdul Ghafoor and B. Sitkowska, "MasPA: A Machine Learning Application to Predict Risk of Mastitis in Cattle from AMS Sensor Data," *AgriEngineering*, vol. 3, no. 3, pp. 575–584, 2021, doi: 10.3390/agriengineering3030037.
- [30] U I Bajwa*, S Alam, N u Haq, N I Ratyal and M W Anwar, "Skin Disease Classification using Neural Network" Volume 16 , Issue 6 , 2020.