iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VI **Month of publication:** June 2026

DOI: https://doi.org/10.22214/ijraset.2026.83791

Prediction of Common Livestock Cattle Farms Diseases Using Machine Learning Techniques

Madhusmita Moharana[1], Dr. Bidush Kumar Sahoo[2]
Department Of Computer Science and Engineering, GIET University, Gunupur, Odisha-765022

Abstract: Stock health management is necessary to ensure animal comfort, boost farm production and ensure the viability of cattle farming systems. Early disease detection is extremely important in order to reduce mortality and to control production costs and allow prompt intervention by vets. The traditional approaches to diagnosis of diseases, on the other hand, involves careful observation and the expert opinion which can be costly, subjective and difficult to access in many rural farm areas. To solve these problems, a system for predicting diseases in cattle using machine learning was created. It is based on a structured database of disease labels and the attributes of diseases related to symptoms obtained from records of cattle in the course of several observation periods. The data was preprocessed in numerous ways, including loading the data, validation, handling missing values, encoding the goal label, feature selection, and exploratory data analysis, which enhanced the data quality and the model's performance. A number of supervised machine learning methods were used to train and classify diseases. Some of these are Decision Tree, Random Forest, K-nearest neighbour, and Naive Bayes. For evaluation, the accuracy, precision, recall, and F1-score metrics were used, to ensure the accuracy of the accuracy in all disease categories. Experiments showed that the Random Forest predictor was the best at making predictions, with a success rate of 95.8%, beating out the other models. The proposed system is fast and accurate to forewarn early diseases in cattle and supports monitoring of the animals' health and informed farm management decisions.
"Index Terms: Cattle Disease Prediction, Livestock Health Monitoring, Machine Learning, Disease Classification, Random Forest, Precision Agriculture".

I. INTRODUCTION

Livestock production plays a significant role in global agricultural production, contributing to food security, economic development, and rural livelihoods. Cattle farming is one of the most significant forms of livestock farming as they yield milk products, meat and other agricultural products for use in the household and for business. Ensuring good health in cattle is paramount to ensuring high productivity, wellbeing and long tenure of the farm. With the increase in demand for livestock products, modern technologies, which can assist in the operation of farms and in monitoring disease, are gaining increasing popularity. Digitalisation has been facilitating the gathering and storage of vast quantities of data pertaining to livestock in recent years within agriculture. This has opened up new possibilities for smart health tracking and predictive analytics [2].

Even though veterinary science is always getting better, managing diseases in cattle is still a big problem for livestock farmers. The traditional method of disease detection involves careful observation and making judgements based on appearance, and seeking advice from experts. This can be time consuming and result in wrong decisions [3]. Such limitations may help diseases to proliferate, reduce efficiency in production, and increase the cost of managing them. In addition, medical assistance in the farm area may be scarce and/or difficult to access, making prompt action difficult [4]. The current way of watching over livestock is mainly to treat them when they are sick rather than preventing them from becoming ill. This demonstrates the need of more effective and data-driven solutions [5]. As the animal systems become more complicated, the use of smart tools is made more important to detect the disease in an early stage and enable the persons to make smart decisions [6].

In order to overcome such issues this research is primarily focused on the development of a smart system for disease prediction in cattle based on data pertaining to the health condition of the animals to assist in locating and categorizing the disease. The main goal is to make disease forecasts more accurate and reliable while relying less on traditional methods based on observations. The suggested framework aims to look at health data related to symptoms and come up with predictive insights that can help farmers and livestock managers make quick decisions about how to take care of animals' health [7]. The framework is also intended to offer a remedy which is extensible and flexible to existing livestock systems in which a vast amount of health data is being generated on a regular basis [8].

The addition is significant as it may help to improve the monitoring process, minimise economic losses and promote sustainable livestock production. Use of predictive tools can assist in preventing the spread of disease and raising the health of the entire herd by catching the signs of disease earlier [9]. Smart analytical methods can also be applied to the management of animals, helping to facilitate precision agriculture and making it easier for farmers to implement technology-driven farming methods. Such enhancements would contribute to increased production, improved resource utilization and sustainable agriculture in the cattle value chain over the long term [10].

II. LITERATURE REVIEW

The recent advancements in ML have significantly contributed to the management of cattle health with the ability to monitor and forecast diseases using data. Cockburn discussed the increasing use of ML in dairy farm management, and the potential benefits of using ML to monitor animal health, productivity, and farm efficiency [11]. In the same vein, Swapna et al. discussed the significance of AI and ML in identifying and early detection of cow diseases. This demonstrates the potential of intelligent systems for enhancing the health of animals and their productivity [12]. Mia et al. thoroughly examined all the uses of ML in the livestock business. They discussed various predictive and analytical approaches used for disease detection and farm decision making systems [13]. These studies demonstrated that intelligent technologies can be used to manage livestock, but they were generally for other purposes rather than developing a model for predicting diseases which can be applied to all kinds of cattle diseases.

There's a lot of research that's been done on jobs such as predicting and forecasting diseases. Afshari Safavi explored the potential of employing ML methods to detect the presence of LSD and determine whether meteorological and geospatial factors are associated with its presence. This demonstrated the potential of environmental factors as a tool to aid disease forecasting [14]. Niloy et al. looked into finding diseases in livestock based on their symptoms and showed how well ML can work at finding diseases from the symptoms that farmers and livestock managers report [15]. These approaches were found to work in some cases, but many methods could only apply to certain diseases and/or datasets. That was not very helpful when it came to adapting to the many different animal health issues that real-life farmers encounter.

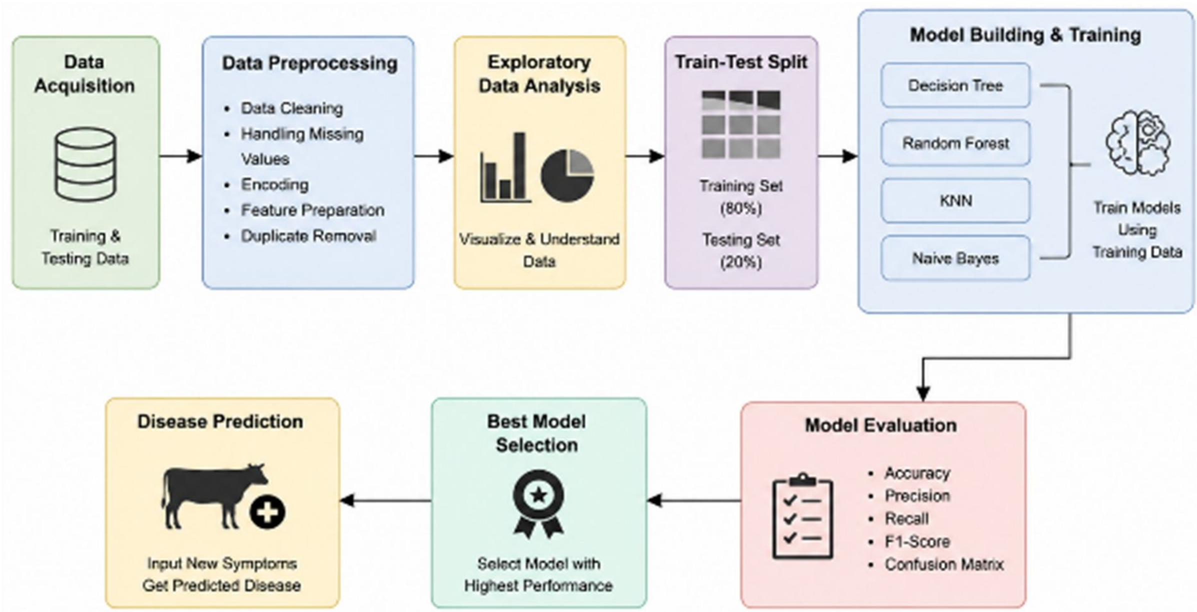
Other health and disease experts have considered other sources of data and health measures in determining the impact of disease. Using milk mid-IR spectroscopy data, Contla Hernández et al. used ML classification methods to identify what health issues the grazing dairy cows had [16]. Nayeri et al. examined the potential of applying commonly used statistical approaches and ML algorithms to enhance animal health and breeding. They noted the increasing trend of data-based livestock analytics [17]. Hyde et al. developed an automated approach to predict pattern of mastitis infections in dairy herds and demonstrated how disease control strategies can be aided by predictive modeling [18]. Although these studies were useful, they focused on specific and narrow biological indicators or a single health issue. This meant that they couldn't be used for regular disease tracking in cattle.

Great interest has been shown also in the potential of new technologies to aid livestock health. A smart system has been suggested by Swain et al. to manage the livestock using IOT and ML to make diagnosis easier of health issues in cattle and prediction of disease in real-time [19]. Recently, Ahmed et al. reviewed ML techniques for the prediction of diseases in cattle and demonstrated that the computer models can be used to evaluate and take decisions regarding cattle health [20]. However, there are challenges with generalizing models, building a disease classification system from symptoms, making information accessible for farmers, and building trustworthy systems capable of managing a variety of diseases through structured information on animal health.

There is good evidence that there is a need for further research in the area of developing disease prediction systems that are comprehensive and accurate for cattle that can utilize information from a variety of symptoms and still be able to predict diseases well across a range of disease classes. This study is thus aimed at providing a unified disease prediction system which will facilitate early detection of diseases, more reliable disease classification and at assisting individuals in decision making regarding the care of livestock health in a modern cattle farm.

III. MATERIALS AND METHODS

The proposed approach is designed to facilitate prediction of diseases in the cattle through a structured data set of symptoms and ML to develop a classification framework. The data consists of training and test data with a list of symptoms and the corresponding disease labels. The aim variable is the disease category that has to be predicted. The method starts with collecting and validating data. After that, it performs preparatory actions such as target label encoding, feature selection, exploratory analysis, and so on, to obtain consistent data for model construction. The processed data is divided into a training group and a testing group to facilitate training and testing. A number of supervised ML methods are used to look at the connections between symptoms and classify diseases into more than one group. They are DT, RF, KNN and NB. The Accuracy, Precision, Recall, and F1-score are used to rate the performance of the model. The aim is to improve the speed of diagnosis of diseases and to standardise the health rating of cattle.



"Fig.1 Proposed Architecture"

Data is the first step toward developing a system to predict diseases in cattle. Health records of cattle are used to collect training and testing data. Preprocessing includes cleaning the data, dealing with missing values, encoding categorical traits, feature preparation, and getting rid of duplicates. Next, EDA is employed to discover trends and patterns in the data and to understand its distribution. The information that was processed is split into two sets: training (80%) and testing (20%). Different metrics such as accuracy, precision, recall, F1 score and confusion matrix are used for training and testing of different ML models. The model which performs the best is selected and applied to predict the diseases of new cow symptoms.

A. Dataset Collection

The data for this study was obtained from a public data base of livestock diseases. It consists of structured information about cattle, created for specific purposes, such as disease forecasting.It's comprised of organised information concerning cattle, developed for specific purposes, such as disease forecasting. There are numerous samples from which cattle with varying health states can be selected from the dataset. Input features (attributes) are those that are related to the symptom and target labels (classes) are disease categories. A wide range of disease cases have been collected, and sufficient variety for classification analysis. It is well structured, has many symptoms and is fairly split into the different types of diseases that make it good to use ML to predict. The dataset can also be used in the development of a useful framework to monitor the health of cattle that enables accurate model training, evaluation and comparison of model performance.

B. Pre-Processing

During the preprocessing step, raw health information for cattle were changed into a format that is structured and ready for ML. It consisted of data validation, goal encoding, feature preparation, exploratory analysis and data partition to enhance the data quality, enhance predictive accuracy and assist in disease classification.

1) Dataset Acquisition and Validation: The information was collected from structured cattle health records that included information on the name of the disease and the attributes of the symptoms. Before the analytical processing, the data integrity, attribute consistency, and class representation were checked with the help of initial validation methods. This ensures that the data obtained represents a good representation of the underlying diseases, and provides a good foundation for prediction. A validated data set ensures data accuracy, minimisation of errors and facilitates the development of robust disease classification models for successful prediction.
2) Target Label Encoding: The labels for each disease group were translated into a standard set of numbers, which were then passed on as target labels for ML and classification. Since the goal values have to be consistent and understandable to the predictive models, categorical disease labels have been converted to the formats that can be understood by the machines,

maintaining its difference between classes. This process facilitates rapid learning of the trends of diseases and classification of diseases more accurately across multiple disease groups. With proper target encoding, models will be more compatible, faster to compute and better to be able to predict during the training and review phase.

3) Feature Selection and Preparation: To identify the significant symptom attributes which aid greatly in predicting the disease in the cattle, feature preparation and feature selection were performed. To improve the quality of the dataset and make model learning simpler, information that wasn't needed or was repeated was taken out. The selected symptom variables were classified as predictive variables which would express the disease-related characters well. This step will make the analysis faster, more effective at identifying patterns and easier to understand by ensuring that decisions on classification are made using health indicators that are meaningful and useful.
4) Exploratory Data Analysis: Exploratory Data Analysis was employed to examine the characteristics of the data, to ascertain the distribution of symptoms and to uncover associations between health variables and disease categories. Statistics and visual interpretation were used before constructing the model to obtain meaningful information related to the structure of the data. This way of analyzing helps find possible problems, how classes are distributed, and how features behave. It is important to learn these patterns to prepare data more effectively, to make more meaningful modeling decisions, and to improve the overall accuracy of the disease prediction results.

C. Training and Testing

The preprocessed dataset underwent exploratory analysis, and then was divided into a training dataset and a testing dataset for more objective modeling and evaluation. The training part was employed to learn patterns associated with diseases while the testing part provided an independent assessment of the predictions. The approach of CPP reduces bias in the assessment of a model and allows reliable assessment of its generalization capacity. Correct separation of data is required to ensure that performance estimates are correct in an unprecedented scenario.

D. Algorithms

1) Decision Tree: DT is a supervised classification technique that is used for making decision in a hierarchical manner based on feature conditions. It then makes predictions. It provides classification results that are easily interpretable, facilitates pattern recognition, contributes to the analytical clarity, and performs well as a baseline model for multiclass prediction.
2) Random Forest: RF is a multi-classifier technique that enhances the classification accuracy by aggregating the predictions of different decision structures. It is more powerful because of its combined decision making process, has less spread in the predictions and is more generalizing. This will ensure its reliability and consistency in case of complicated multiclass classification problem.
3) K-Nearest Neighbor (KNN): An instance-based classification technique that predicts the outcome based on the similarity between new events and patterns previously learned is KNN. It has a neighborhood-based approach that allows for flexible pattern recognition, effectively provides multiclass classification capabilities, and provides an ability to compare and assess the behavior that is predicted.
4) Naive Bayes: NB is a probabilistic classification algorithm which computes the likelihood that an object belongs to a particular class based on the observed properties of the object. It is a good method for baseline evaluation and classification analysis as it can be computed quickly, predictions are stable, and it is able to deal with problems having more than one class.

IV. EXPERIMENTAL RESULTS

1) Accuracy: The measure of how well a test can distinguish healthy people from sick people is the test's accuracy. In order to have an impression of the accuracy of a test, we need to determine the percentage of true positives and true negatives. As for mathematics, it can be expressed as

$$Accuracy = \frac{\text{TP + TN}}{\text{TP + FP + TN + FN}} \quad (1)$$

2) Precision: Precision is the percentage of correctly classified cases or samples divided by the number of cases/correctly classified cases that were classified as positive. Therefore, here is the procedure to determine the precision:

$$Precision = \frac{\text{True Positive}}{\text{True Positive + False Positive}} \quad (2)$$

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

3) Recall: Recall is a measure of how well a machine learning model can identify all the significant instances of a given class. It demonstrates the ability of a model to represent examples of a particular class. It is the ratio of the number of correctly predicted positive observations to the number of actual positive observations.

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} (3)$$

4) F1-Score: F1 score is a metric to evaluate the accuracy of a ML model. Computes the sum of a model's accuracy and recall scores. The accuracy number is the number of times a model correctly predicted an answer in the entire data set.

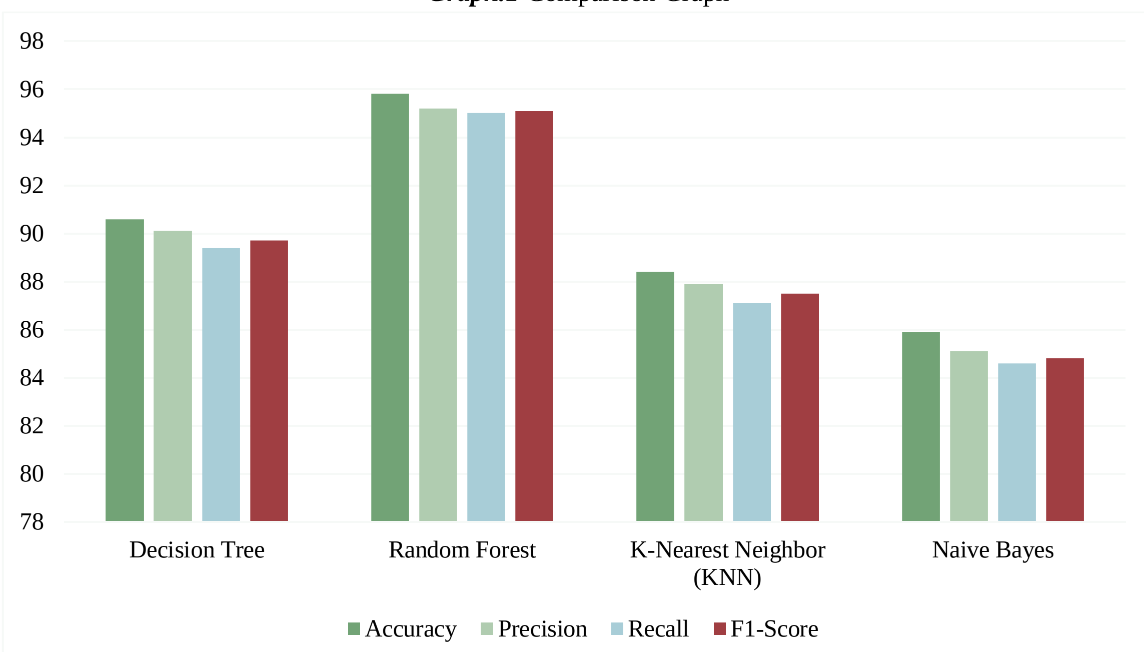
$$F1\, Score = 2 * \frac{Recall \times Precision}{Recall + Precision} * 100(1)$$

Table.1 Performance Evaluation Table

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	90.6	90.1	89.4	89.7
Random Forest	95.8	95.2	95.0	95.1
K-Nearest Neighbor (KNN)	88.4	87.9	87.1	87.5
Naive Bayes	85.9	85.1	84.6	84.8

Table 1 shows how well four ML models compare in terms of performance. RF model is performing best among all the models with 95.8% accuracy, which is better in terms of predictive power, robustness and classification consistency than other three models - Decision tree, KNN and Naive Bayes.

Graph.1 Comparison Graph



The results of comparing the ML models using Accuracy, Precision, Recall, and F1-Score measures are shown in Graph 1. Random forest was the top performer in every one of the algorithms analyzed, as it was the most successful at both predictions and classification.

V. CONCLUSION

In conclusion, the developed method for predicting cattle diseases was made to help with managing the health of livestock by using ML to accurately and quickly spot disease conditions. Structured cattle health records with disease labels and symptom attributes were used in the framework. This enabled systematic studies of disease-related patterns for multiclass classification. To make the best predictions, several algorithms of guided learning were used and compared. These included DT, RF, KNN and NB. The performance evaluation indicated that the predictor with the highest accuracy in the prediction was the RF predictor, which achieved 95.8% accuracy, high Precision, Recall and F1-score.

A significant enhancement was achieved by preprocessing the data completely, creating the features and comparing the models. This allowed for a more reliable classification and higher overall accuracy of the predictions. The framework created converts health data relating to symptoms into actionable forecasts about disease, providing a dependable and structured approach for assessing disease in cattle. The system, which can detect disease early and make sound decisions, enhances the monitoring of livestock health, the management of farms, and reduces reliance on traditional assessment methods that rely on observation. Overall, the results show that ML is a good way to provide reliable analytical help for modern cattle farming settings.

The resulting system for predicting cattle disease is a stepping stone towards data-driven approaches to evaluate animal health in more practical applications. Future improvements may include more extensive and diverse information regarding the health of the cattle, for easier application in various farming situations. Other health-related inputs may be included besides symptoms, to provide a more comprehensive evaluation of disease. There could also be an integration of digital records and real-time monitoring of animals in the system to enhance its decision-making power. Bettering predictive models and analytical methods can lead to more dependable, scalable, and useful uses for managing the health of cattle in modern livestock settings.

REFERENCES

[1] Ahmed, M., Javaid, S., & Saepudin, S. (2025). Cattle Disease Prediction Using Machine Learning Algorithms. Engineering Proceedings, 107(1), 85.

[2] Zhou, X., Xu, C., Wang, H., Xu, W., Zhao, Z., Chen, M., ... & Huang, B. (2022). The early prediction of common disorders in dairy cows monitored by automatic systems with machine learning algorithms. Animals, 12(10), 1251.

[3] Punyapornwithaya, V., Klaharn, K., Arjkumpa, O., & Sansamur, C. (2022). Exploring the predictive capability of machine learning models in identifying foot and mouth disease outbreak occurrences in cattle farms in an endemic setting of Thailand. Preventive veterinary medicine, 207, 105706.

[4] Lasser, J., Matzhold, C., Egger-Danner, C., Fuerst-Waltl, B., Steininger, F., Wittek, T., & Klimek, P. (2021). Integrating diverse data sources to predict disease risk in dairy cattle—a machine learning approach. Journal of Animal Science, 99(11), skab294.

[5] Mahmud, M. S., Zahid, A., Das, A. K., Muzammil, M., & Khan, M. U. (2021). A systematic literature review on deep learning applications for precision cattle farming. Computers and Electronics in Agriculture, 187, 106313.

[6] Zhang, S., Su, Q., & Chen, Q. (2021). Application of machine learning in animal disease analysis and prediction. Current Bioinformatics, 16(7), 972-982.

[7] Rao, A., Monika, H. R., Rakshitha, B. C., & Thaseen, S. (2023). Cattle disease prediction using artificial intelligence. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 11(IV), 2184.

[8] Swain, S., Pattnayak, B. K., Mohanty, M. N., Jayasingh, S. K., Patra, K. J., & Panda, C. (2024). Smart livestock management: integrating IoT for cattle health diagnosis and disease prediction through machine learning. Indonesian Journal of Electrical Engineering and Computer Science, 34(2), 1192-1203.

[9] Neupane, R., Aryal, A., Haeussermann, A., Hartung, E., Pinedo, P., & Paudyal, S. (2024). Evaluating machine learning algorithms to predict lameness in dairy cattle. Plos one, 19(7), e0301167.

[10] Grzesiak, W., Zaborski, D., Pluciński, M., Jędrzejczak-Silicka, M., Pilarczyk, R., & Sablik, P. (2025). The use of selected machine learning methods in dairy cattle farming: A review. Animals, 15(14), 2033.

[11] Cockburn, M. (2020). Application and prospective discussion of machine learning for the management of dairy farms. Animals, 10(9), 1690.

[12] Swapna, P., Geetha, M., Nuthana, B., & Rohan, R. (2024). Using AI and machine learning for early detection and management of cattle diseases to improve livestock health and productivity. Int. Res. J. Educ. Technol, 6, 1815-1822.

[13] Mia, N., Sarker, T., Halim, M. A., Alam, A. M. M. N., Ali, M. S., Rahman, M. M., & Hashem, M. A. (2025). Machine learning overview and its application in the livestock industry. Meat Research, 5(1).

[14] Afshari Safavi, E. (2022). Assessing machine learning techniques in forecasting lumpy skin disease occurrence based on meteorological and geospatial features. Tropical Animal Health and Production, 54(1), 55.

[15] Niloy, M. A., Bhowmik, T., Abedin, J., Ferdous, S. J., & Jahan, I. (2024). Exploring machine learning techniques for symptom-based detection of livestock diseases (Doctoral dissertation, Brac University).

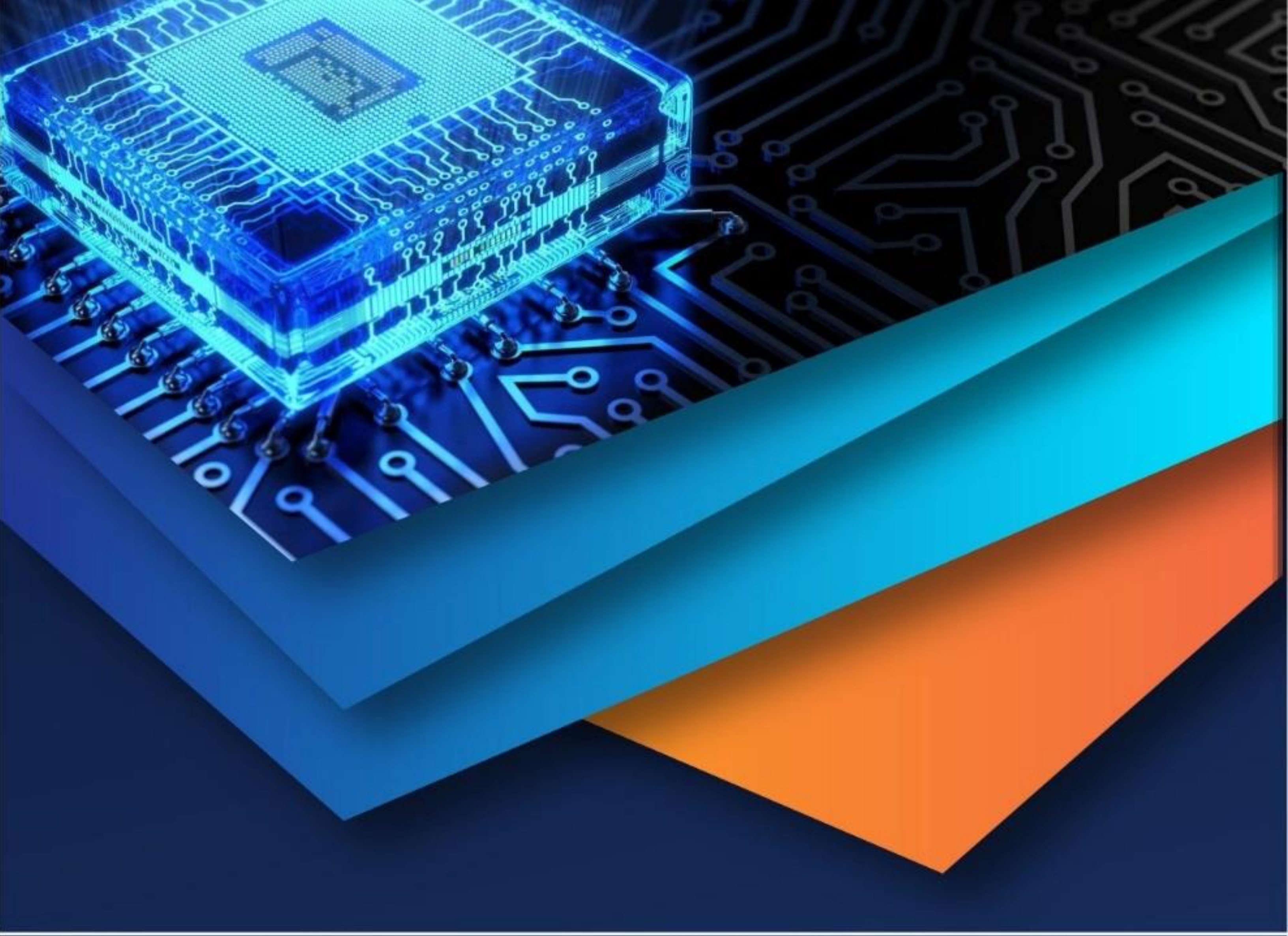
[16] Contla Hernández, B., Lopez-Villalobos, N., & Vignes, M. (2021). Identifying health status in grazing dairy cows from milk mid-infrared spectroscopy by using machine learning methods. Animals, 11(8), 2154.

[17] Nayeri, S., Sargolzaei, M., & Tulpan, D. (2019). A review of traditional and machine learning methods applied to animal breeding. Animal health research reviews, 20(1), 31-46.

[18] Hyde, R. M., Down, P. M., Bradley, A. J., Breen, J. E., Hudson, C., Leach, K. A., & Green, M. J. (2020). Automated prediction of mastitis infection patterns in dairy herds using machine learning. Scientific reports, 10(1), 4289.

[19] Swain, S., Pattnayak, B. K., Mohanty, M. N., Jayasingh, S. K., Patra, K. J., & Panda, C. (2024). Smart livestock management: integrating IoT for cattle health diagnosis and disease prediction through machine learning. Indonesian Journal of Electrical Engineering and Computer Science, 34(2), 1192-1203.

[20] Ahmed, M., Javaid, S., & Saepudin, S. (2025). Cattle Disease Prediction Using Machine Learning Algorithms. Engineering Proceedings, 107(1), 85.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY