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.2026.83785

Prediction of Common Livestock Pig Farms Diseases Using Machine Learning

Biswajita Kumar Sabat
Gandhi Institute of Engineering and Technology University, Gunupur, Odisha

Abstract: Livestock diseases have a significant impact on animal health, production, and the economy of rural agricultural communities particularly in areas with intensive pig farming. The early and right diagnosis of disease is crucial to prevent the disease from spreading and to minimize financial losses. However, the traditional tracking method involves manual inspection as well as frequent veterinary examinations, which can result in delayed diagnosis and intervention. To overcome these issues, researchers designed a system which is based on machine learning to predict diseases in pig farms within Odisha, India, to identify common diseases affecting animals. Environmental, management and animal health parameters for a fake dataset that resembles a real farm were gathered in a simulated multi-year period. These parameters were temperature, humidity, rainfall, herd size, vaccination rate, biosecurity score, feed intake, water consumption, average weight and disease history. To make the data better and the model work better, data preparation included cleaning, normalizing, choosing features, and changing the data. Random Forest, Support Vector Machine, Decision Tree and K-Nearest Neighbour were used and compared several techniques of classification. Performance was evaluated by metrics such as accuracy, precision, memory, and F1-score. The Random Forest predictor outperformed the others in the experiments with an accuracy of 94.8%, precision of 94.1%, recall of 95.6% and an F1 score of 94.8%. The suggested framework improves the accuracy of disease risk predictions and helps with timely animal health management by making decisions more quickly.

"Index Terms: Livestock Disease Prediction, Pig Farming, Machine Learning, Precision Livestock Farming, Disease Risk Assessment, Random Forest Classifier".

I. INTRODUCTION

Livestock farming is a significant sector within the agricultural sector around the world as it contributes to food security, employment opportunities in rural regions and economic growth. Animal husbandry is an important livelihood source for people across many developing countries living in rural areas. Pig production has received a considerable amount of interest since pigs are highly fertile, grow rapidly and there exists an increasing demand for pork products. Maintaining good livestock health and minimizing losses due to disease is critical to the long-term viability and success of pig farms. Therefore, disease control is essential for achieving good animal health, improved production and sustainable agricultural production [1, 2, 3].

Despite significant advances in veterinary science and farm management, disease outbreak remains a significant challenge in an animal production system. Most of the traditional methods of monitoring diseases rely on the detection by humans, frequent inspections and frequent veterinary examinations, and may not be adequate in identifying new health issues promptly. In addition, disease is influenced by many interrelated factors including environment, management, nutrition and past disease history. As these factors become more complex, conventional monitoring techniques are less useful and it becomes more difficult to prevent diseases from occurring. Existing methods often have problems with being able to be used on a large scale, staying consistent, and finding risks early on in farming settings that are always changing [4, 5, 6]. To achieve this, the aim of this study is to develop an intelligent prediction tool using farm related information which can assist in the early detection of common diseases in pig farming systems. The proposed methodology will examine environmental, management and animal health issues to determine the risk of disease and assist with decision making. The framework's goal is to give farmers and other agricultural stakeholders reliable and easy-to-reach help by incorporating data-driven analytical tools into the managing of livestock. The contribution is establishing a full predictive platform that will assist in improved health monitoring, awareness of diseases and taking appropriate decisions in real farming settings [7, 8]. This new development is not only relevant for diagnosing diseases, but also for promoting digital agriculture and precision livestock farming technologies. Knowing the risks of disease occurrence may reduce economic losses, improve livestock husbandry and ensure that livestock production systems become viable for longer. Also, finding possible health threats early on can help make the best use of resources, boost the efficiency of farm management, and help keep the food supply stable. As smart technologies become more common in agriculture, it's important to come up with flexible and useful ways to keep track of the health of animals so that we can face future problems and improve modern farming methods [9, 10].

II. LITERATURE REVIEW

Recent improvements in precise livestock farming have made it easier to use ML to keep an eye on the environment and guess what will happen with animals' health in production systems. Peng et al. developed a framework for ML prediction of ammonia emission from pig houses, depending on environmental factors. This shows how data-driven approaches can be used to better handle livestock housing and keep the environment safe [11]. Similarly, Liu et al. proposed some means of determining when infectious diseases will occur in laying hen farms. The example illustrates how ML could be employed to achieve disease surveillance and outbreak forecasting in commercial livestock farms [12]. Although these studies demonstrated the potential of predictive analytics, they only assessed some environmental factors or disease cases, and can therefore not be used for managing the health of all livestock.

Much work has been performed in the field of disease detection and assessment of animal health by using production and behavioural indicators. Ollagnier et al. showed that feeding behavior records could be used to predict when pigs would start biting their tails, which would allow for earlier intervention and better animal care [13]. Taylor explored ML approaches for predicting pig growth in an industrial production system. He stressed how important predictive technologies are for increasing farm output [14]. In the same way, Lee et al. deployed ML to forecast the average daily weight gain for pigs, providing a valuable tool for monitoring growth and optimizing pig production [15]. Some good results were obtained from these inputs, but most of them were related to specific behavioral or productivity outcomes and not to diseases in general.

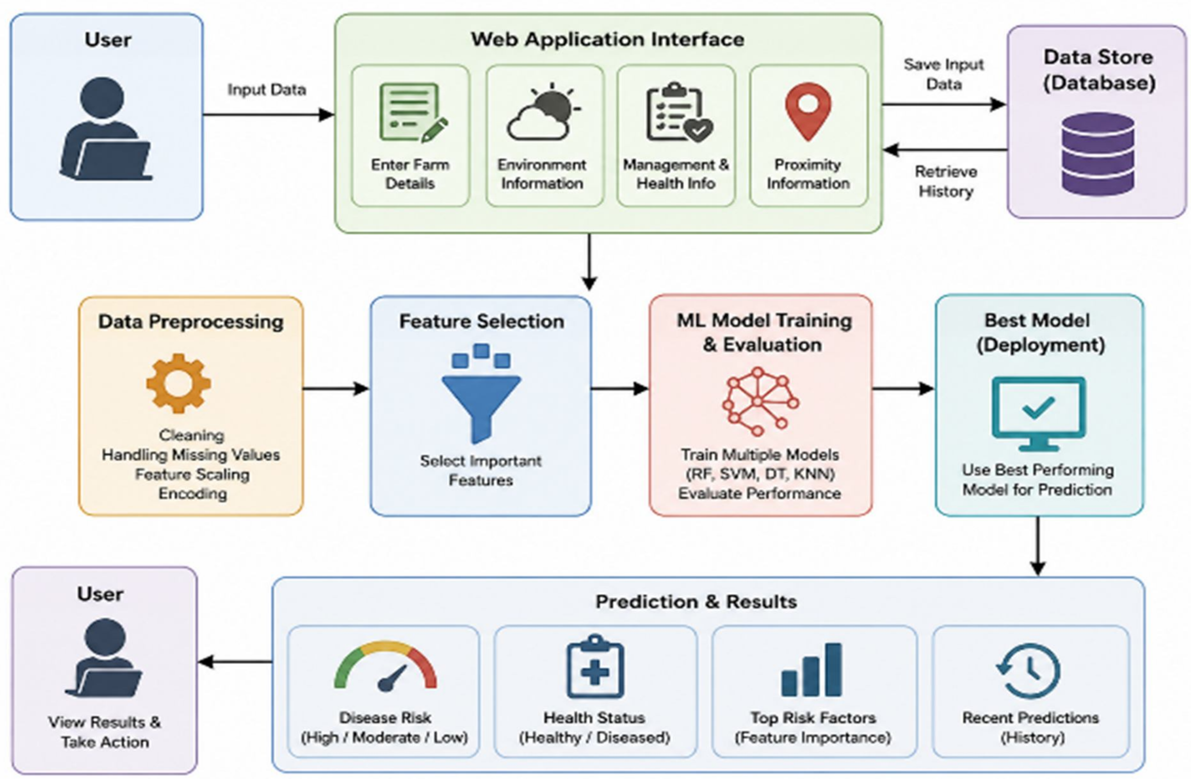
Intelligent analytics have also been looked into as a way to control biosecurity and find strange things. Create interpretable ML models for decision making around animal health, as Sykes et al. did for pig farming and PRRS biosecurity [16]. Park et al. created an anomaly detection system based on DL to keep an eye on livestock farm tools. This resulted in more reliable system and better efficiency of farm management [17]. Ghosh and Mandal also developed a supervised learning system for intelligent classification of pig breeds in another project. This indicates that ML can have several other applications beyond monitoring animal health [18]. Even with these improvements, many of the systems that are already in place only deal with specific problems in farm management and don't offer combined disease risk assessment.

Most of the latest research has been on directly predicting diseases in groups of pigs. Halev et al. showed that it is possible to predict diseases that will happen in pig groups using information from the farms [19]. They did this by employing ML algorithms to make the predictions. Similarly, Kavlak et al. examined the potential of disease diagnosis in pigs based on their feeding patterns and determined they were relatively successful in predicting disease [20]. However, most of the existing approaches are based on relatively small amounts of data, are limited to a specific disease type, or rely on a specific tracking system that may not be readily available in all production contexts.

The literature review shows that ML is becoming more and more important in managing livestock. However, there are still big gaps in making disease prediction frameworks that are easy to use and take into account a lot of different environmental, management, and health factors at the same time. Also, not much thought has gone into making disease risk assessment methods that work well for pig farming in developing agricultural areas. This study aims to solve these problems by creating a single prediction framework that brings together different farm-related factors. This will enhance detection of diseases at an earlier stage, improve monitoring of animal health and improve planning of long-term pig farming operations.

III. MATERIALS AND METHODS

The proposed approach aims to classify the most prevalent diseases affecting pig farms with a ML framework designed to help prevent diseases that impact animal health at an early stage and sustain intelligent farm management. It employs a simulated data set with environmental parameters, farm management parameters, information on livestock health and disease related parameters from the past years to illustrate the actual scenario of pig farming in Odisha. The method has a sequence of activities such as data collection, data preparation, feature selection, model construction, model testing, and prediction. The data is cleaned and normalised in preparation to ensure consistency of data and to make the data ready for the good learning of the model. Feature selection is a special technique to identify significant features, remove irrelevant features and to simplify understanding and computing of the data. Various classification models like RF, SVM, DT and K-NN are used to compare and to make predictions. This method is intended to make reliable disease risk assessments, to effectively monitor livestock health at the right time and to inform of sensible farm decisions.



"Fig.1 Proposed Architecture"

Architecture of the system depicts smart approach to predict animal diseases in pig farms. A web application interface accesses a database (writes and reads data from it). Information on the farm, its environment, management and health problems are entered by the users. The collected data is then preprocessed, including data cleaning, imputation, scaling, and encoding. Then feature selection is used to select important traits. Multiple ML models are trained, tested, and the best performing model is used to make predictions. The system is able to generate disease risk levels, health status assessments, key risk factors, and historical prediction records. This enables the correct management of livestock health and rapid decision making.

A. Dataset Collection

A simulated dataset was used in this research for the prediction of animal diseases which was designed to resemble authentic farm conditions of pigs in Odisha, India. Has a record of the environment, management and health of the animals. The following records are included: temperature, humidity, rainfall, herd numbers, vaccination rate, biosecurity score, water use, feed use, average weight, diseases history, and disease risk labels. A range of simulated farm situations from various years are included in the data set, providing a fair, accurate representation of disease conditions. It can be used to detect disease as it contains many key farm factors. Although there are limited datasets of livestock health in the real world, this facilitates training and testing the model, and generalizing it.

B. Pre-Processing

The proposed approach stipulates a sequence of steps: data collection, data preparation, feature selection, model construction, model testing, and predictions of diseases. All these contribute to the making of correct decisions and livestock disease risk assessments.

1) Data Preprocessing: Procedures for data preparation were used to turn raw data into a shape that could be used for ML analysis. Errors were removed using cleaning processes and feature scales were kept similar and data variation impacts during model training were reduced using normalization processes. Data preprocessing enhanced the quality of data and lessened the effect of patterns that did not have a major significance and may have affected prediction accuracy. Standardizing inputs increased the stability of the model, allowed it to learn faster, and provided it with a foundation to continue with the subsequent steps of analysis.

2) Feature Selection: Variable optimization methods were used to find the most important factors that affect predicting cattle diseases. The selection process eliminated attributes that were not or that didn't contribute much to the picture and retained those that were more predictive. Feature selection helped to reduce the complexity of the data sets and made them easier to comprehend when building models. Narrowing the scope of the analysis to the most relevant environmental, management and livestock health indicators increased the effectiveness of the analysis and resulted in more useful disease risk assessment outcomes.
3) Model Development: Creating a predictive model meant using a number of different classification methods to find links between conditions on the farm and trends of disease occurrence. Different ML models were taught and analyzed to see if they could be used to predict the health of livestock. Comparative evaluation allowed for a systematic comparison of the effectiveness of various classification techniques in making predictions. Having a variety of learning models led to a more reliable analysis and was also used to identify a good prediction mechanism, which could lead to accurate livestock health assessments.
4) Model Evaluation: A performance review was done to see how well predictive models could classify the health conditions of livestock. The methods used for evaluation were standard performance measures which measured the quality of the disease classification and prediction, and the ability to identify diseases. Assessment of the object was carried out through comparative assessment, which enabled us to select an objective model and to gain a structured knowledge of the general nature of predictions. During the review, it was ensured that the selected model continued to function effectively and perform well in terms of disease risk assessment.
5) Disease Prediction And Visualization: Prediction making and result showing were integrated together to create a friendly and easy-to-understand disease assessment environment which can be applied in reality. Health of the animals were estimated using information provided on farms and disease risk output generated. Making predictions easier to understand made their use easier and more accessible for people to understand health issues and other factors that impacted them. It presented in an organized manner the results of an analysis and made decisions easier and more useful when managing livestock in situations.

C. *Algorithms*

1) Random Forest (RF): The ensemble-based classification technique was used as a random forest to enhance overall classification accuracy and make the classification more accurate in the assessment of livestock diseases. The algorithm consists of constructing multiple decision structures and summing up the outcomes to obtain an overall estimate. This type of group learning keeps the system less sensitive to individual models, and leads to more stable predictions with a broad spectrum of input situations. It is also more robust and generalizable since it can manage complex relationships between environmental, management and health-related factors. The ensemble process also mitigates the danger of overfitting, and aids decision making in assessing disease risk when multiple decisions are to be made.
2) Support Vector Machine (SVM): To identify decision boundaries which were able to separate different health conditions better, SVM was employed as a classification method. The algorithm performs the separation in the best area which would maintain the stability of the predictions and also facilitate distinguishing between classes. The method of learning facilitates dealing with interactions among several dimensions of features and contributes to the accuracy of classification. SVM is particularly useful in structured prediction problems, where the generalization capability is very important. It also aids in extracting meaningful patterns from a plethora of factors that influence livestock disease evaluation.
3) Decision Tree (DT): An easy to understand method of grouping things into categories, which was added to DT to study the relations between the various factors which influence the health of animals. The algorithm predicts by segmenting the data into paths of decisions, indicating possible outcomes according to conditions. This hierarchical decision making process facilitates proper analysis and also helps in grasping the factors influencing the behavior of predictions. It has a well-structured learning mechanism that is effective in categorizing objects, and the decision logic is understandable. DT is used to clarify analyses and facilitate understanding of disease-related situations in places where animals are being monitored.
4) K-Nearest Neighbour (KNN): KNN was implemented as a similarity-based classification technique to determine the health of animals based on comparing patterns. The algorithm makes predictions by looking for nearby observations that are similar and using the results to assign predictions that fit together. This is an instance-based learning behavior and is used to make decisions in a flexible way without complicated model assumptions. KNN assists to view the relationship among the health and natural indicators within an area, yet still switch the sorting. That's because its neighbourhood based prediction process generates patterns and offers a new perspective on circumstances that could trigger disease.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

IV. EXPERIMENTAL RESULTS

Accuracy: The accuracy of a test is how well it is able to differentiate between individuals who are sick and those who are healthy. In order to see whether a test is accurate or not, it is necessary to determine what percentage of the cases are true positive and what percentage are true negative. In terms of math, this can be written as

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

Precision: Precision is the percentage of cases or samples that were correctly classified over the ones that were correctly classified as positive. So, this is how to determine the precision:

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

Recall: Recall is a measure used for evaluating the performance of a ML model by measuring its ability to identify all the relevant examples of a given class. It demonstrates the ability of a model to represent examples of a class. It is the ratio of correctly predicted positive observations to the total value of positive observations.

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

F1-Score: F1 Score is the measurement of the accuracy of a ML model. It simply adds the scores of a model on its accuracy and on its recall. The accuracy measure is the number of times a model was correct in predicting the response variable throughout the data set.

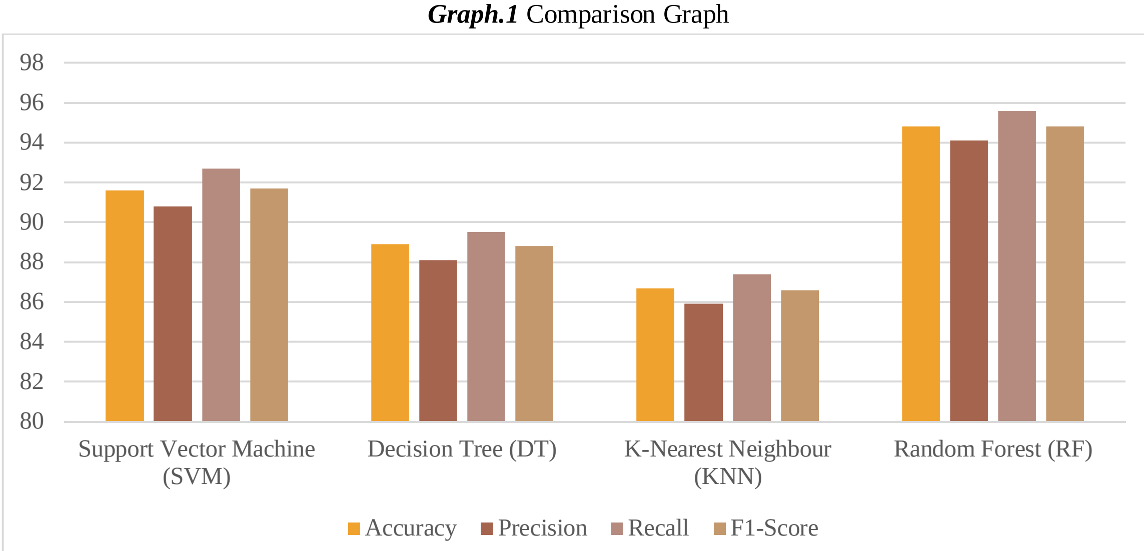
$$F1\,Score = 2 * \frac{Recall \text{ X } Precision}{Recall + Precision} * 100(1)$$

"Table.1 Performance Evaluation Table"

Algorithm	Accuracy	Precision	Recall	F1-Score
Support Vector Machine (SVM)	91.6	90.8	92.7	91.7
Decision Tree (DT)	88.9	88.1	89.5	88.8
K-Nearest Neighbour (KNN)	86.7	85.9	87.4	86.6
Random Forest (RF)	94.8	94.1	95.6	94.8

In Table 1, you can see how well different machine learning classifiers predict diseases in cattle. RF achieved the highest accuracy, precision, recall and F1-scores among all the models, demonstrating its strong performance in predicting outcomes and making reliable predictions.

Graph.1 Comparison Graph



Graph 1 depicts the performance of SVM, DT, KNN and RF models with regards to various metrics such as accuracy, precision, recall and F1 score. RF always achieved the top accuracy, making it a better predictor.

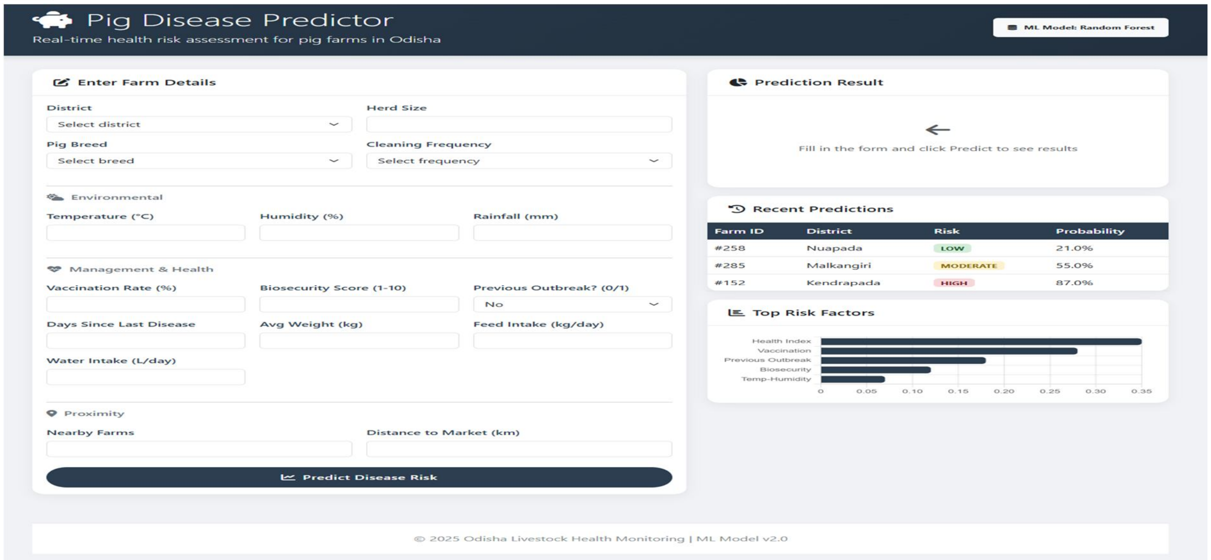


Fig.2 System Interface for Pig Disease Prediction

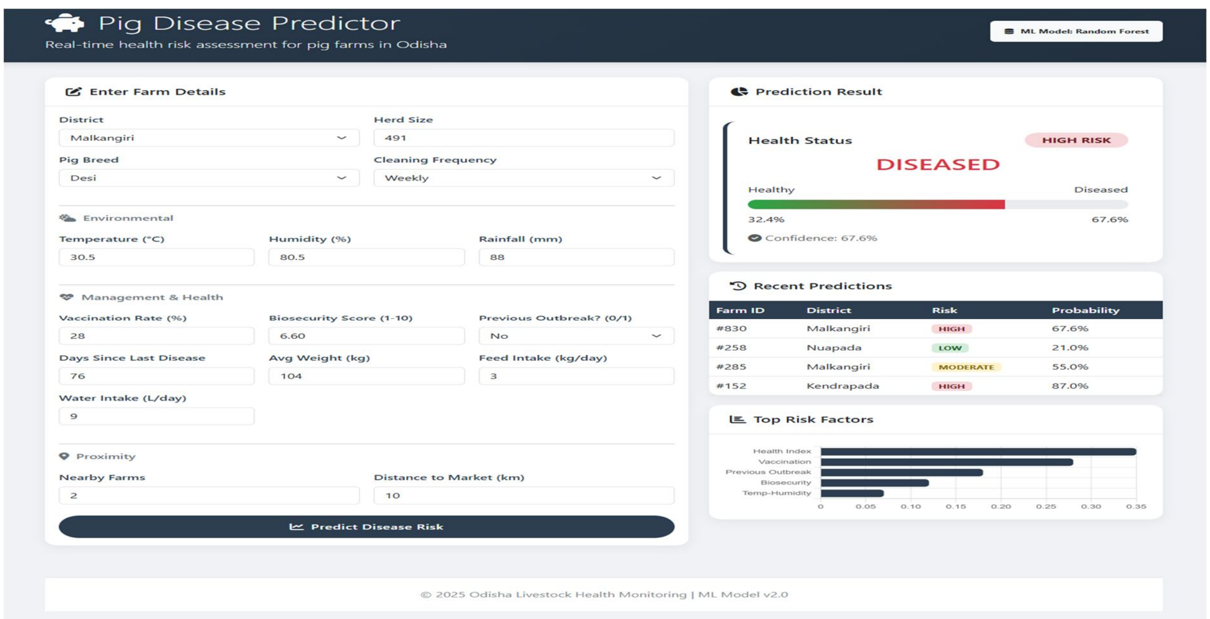


Fig.3 Input and Output Interface of Pig Disease Prediction System

V. CONCLUSION

This system was primarily aimed at providing a more convenient way for animal health monitoring, by establishing a framework for disease prediction in pig farms through ML. The main objective was to help find disease risks earlier and rely less on traditional monitoring methods based on observations. The method provided a structured approach to assess disease risk, and assisted livestock managers to make informed decisions based on environmental conditions, indicators of farm management, information regarding livestock health and historical disease-related factors. Several models of disease classification were compared to determine which one is best to predict disease. The total performance of the model: RF was highest with an F1 score of 94.8%, and was the best model for classifying things. Its accuracy was 94.8%, its precision was 94.1%, its recall was 95.6%, and its F1-score was 94.8%. The use of feature selection boosted the quality of predictions and made them more understandable due to focusing on factors that are relevant to the occurrence of disease. In summary, the application demonstrates the potential of ML to help diagnose the health of animals and provides a reliable and scalable tool for farm timely monitoring and sustainable farming practices.

Future improvements to make it easier to use and to add more features to livestock disease monitoring tools are possible. The framework can be expanded to include additional monitoring tools that gather data from farms in real-time to enable continuous assessment of health. Use of larger datasets specific to the area may be easier to adapt to different farming and livestock situations. Other factors of the environment, nutrition, and health could be incorporated to obtain a better picture of the diseases. The system could be expanded to include other livestock and forecasting specific to the disease. Use of enhanced visualization and decision making tools will further make it easier to use and aid in livestock control.

REFERENCES

[1] Yamsakul, P., Yano, T., Junchum, K., Anukool, W., & Kittiwan, N. (2025). Machine Learning-Based detection of pig coughs and their association with respiratory diseases in fattening pigs. Veterinary Sciences, 12(9), 818.
[2] Kavlak, A. T., Pastell, M., & Uimari, P. (2023). Disease detection in pigs based on feeding behaviour traits using machine learning. biosystems engineering, 226, 132-143.
[3] Halev, A., Martínez-López, B., Clavijo, M., Gonzalez-Crespo, C., Kim, J., Huang, C., ... & Liu, X. (2023). Infection prediction in swine populations with machine learning. Scientific Reports, 13(1), 17738.
[4] Paploski, I. A. D., Bhojwani, R. K., Sanhueza, J. M., Corzo, C. A., & VanderWaal, K. (2021). Forecasting viral disease outbreaks at the farm-level for commercial sow farms in the US. Preventive Veterinary Medicine, 196, 105449.
[5] Valdes-Donoso, P., VanderWaal, K., Jarvis, L. S., Wayne, S. R., & Perez, A. M. (2017). Using machine learning to predict swine movements within a regional program to improve control of infectious diseases in the US. Frontiers in veterinary science, 4, 2.
[6] Bakoev, S., Getmantseva, L., Kolosova, M., Kostyunina, O., Chartier, D. R., & Tatarinova, T. V. (2020). PigLeg: prediction of swine phenotype using machine learning. PeerJ, 8, e8764.
[7] Zhang, S., Su, Q., & Chen, Q. (2021). Application of machine learning in animal disease analysis and prediction. Current Bioinformatics, 16(7), 972-982.
[8] Nadeem, R. M., Bajwa, M. T. T., Mahmood, M., Saleem, R. M., & Maqbool, M. N. (2024). Machine learning-based prediction of african swine fever (ASF) in pigs. VFAST Transactions on Software Engineering, 12(3), 199-216.
[9] Suresh, K. P., Barman, N. N., Bari, T., Jagadish, D., Sushma, B., Darshan, H. V., ... & Deka, A. (2023). Application of machine learning models for risk estimation and risk prediction of classical swine fever in Assam, India. Virusdisease, 34(4), 514-525.
[10] Silva, G. S., Machado, G., Baker, K. L., Holtkamp, D. J., & Linhares, D. C. (2019). Machine-learning algorithms to identify key biosecurity practices and factors associated with breeding herds reporting PRRS outbreak. Preventive Veterinary Medicine, 171, 104749.
[11] Peng, S., Zhu, J., Liu, Z., Hu, B., Wang, M., & Pu, S. (2022). Prediction of ammonia concentration in a pig house based on machine learning models and environmental parameters. Animals, 13(1), 165.
[12] Liu, Y., Zhuang, Y., Yu, L., Li, Q., Zhao, C., Meng, R., ... & Guo, X. (2023). A machine learning framework based on extreme gradient boosting to predict the occurrence and development of infectious diseases in laying hen farms, taking H9N2 as an example. Animals, 13(9), 1494.
[13] Ollagnier, C., Kasper, C., Wallenbeck, A., Keeling, L., Bee, G., & Bigdeli, S. A. (2023). Machine learning algorithms can predict tail biting outbreaks in pigs using feeding behaviour records. PloS one, 18(1), e0252002.
[14] Taylor, C. (2024). Machine learning strategies for the forecasting of pig growth in industry (Doctoral dissertation, Newcastle University).
[15] Lee, W., Han, K. H., Kim, H. T., Choi, H., Ham, Y., & Ban, T. W. (2019). Prediction of average daily gain of swine based on machine learning. Journal of Intelligent & Fuzzy Systems, 36(2), 923-933.
[16] Sykes, A. L., Silva, G. S., Holtkamp, D. J., Mauch, B. W., Osemeke, O., Linhares, D. C., & Machado, G. (2022). Interpretable machine learning applied to on-farm biosecurity and porcine reproductive and respiratory syndrome virus. Transboundary and emerging diseases, 69(4), e916-e930.
[17] Park, H., Park, D., & Kim, S. (2021). Anomaly detection of operating equipment in livestock farms using deep learning techniques. Electronics, 10(16), 1958.
[18] Ghosh, P., & Mandal, S. N. (2022). PigB: Intelligent pig breeds classification using supervised machine learning algorithms. International Journal of Artificial Intelligence and Soft Computing, 7(3), 242-266.
[19] Halev, A., Martínez-López, B., Clavijo, M., Gonzalez-Crespo, C., Kim, J., Huang, C., ... & Liu, X. (2023). Infection prediction in swine populations with machine learning. Scientific Reports, 13(1), 17738.
[20] Kavlak, A. T., Pastell, M., & Uimari, P. (2023). Disease detection in pigs based on feeding behaviour traits using machine learning. biosystems engineering, 226, 132-143.