EXPERT SYSTEM BASED MACHINE LEARNING ANALYSIS FOR DIAGNOSIS AND TREATMENT OF SWINE DISEASES USING DECISION TREE, LOGISTIC REGRESSION AND SUPPORT VECTOR MACHINE ALGORITHMS

Supported by

S. O. Agbo[1], S. E. Samuel[1], D. J. Idoko[2]
[1]Department of Computer Science, Kogi State College of Education, Ankpa, Nigeria.
[2]Department of Computer Science, Joseph Sarwuan Tarka University, Makurdi, Nigeria.
Corresponding author's e-mail address: sunnyreside@gmail.com

Received: September 21, 2022 **Accepted:** November 12, 2022

Abstract:

Late identification of swine diseases results in significant economic loss associated with swine farming. Therefore, it is important to identify predictors of swine health condition. This study presents assessments of swine health using indicators of symptoms characteristics of swine based on machine learning algorithms. Three machine learning classification algorithms (Decision Tree (DT) and Logistic Regression and Support Vector Machine (SVM)) was employed to develop an expert system to capture the knowledge of an expert (endocrinologists) and made it available through a software interface (GUI) for detection, diagnosis and treatment of swine diseases. Comparisons of these algorithms were carried out based on performance factors which include: classification accuracy, precision, sensitivity and specificity. The result showed that Decision Tree produced the highest accuracy of 97.5% as compared to Support Vector Machine which produced 88.8% accuracy, and Logistic Regression which produced 75% accuracy. Hence, Decision Tree is identified as the best performing algorithm for detecting, diagnosing and treatment of swine diseases using a small set of simple identifiable symptoms and clinical measurements at an early stage of animal disease development.

Keywords: Expert System, Machine Learning, Classification Algorithm, Diagnosis, Swine Disease.

Introduction

Over the past several decades a wide variety of information technologies have been deployed within an ever increasing variety of clinical and healthcare settings in an effort to streamline and modernize healthcare delivery for human and animals. Much of this effort has been in response to the limitations of ways that healthcare information has traditionally been collected, retrieved and communicated. For example, the limitations of handwritten paper-based medical records, which have been the predominant form of recording patient and medical information for over a century, have been well documented (Shortliffe & Cimino, 2006). The knowledge of computer science has been aptly applied to make diagnosis process easier and faster. One of such area is the use of expert system to diagnose different diseases (Sajja, *et al.*, 2012).

Typically, an expert system incorporates a knowledge based on accumulated experience and an inference or a set of rules for applying the knowledge base to each particular situation that is described to the program. The system capabilities can be enhanced in additions to the knowledge based on the set of rules, current systems may include machine learning capabilities that allow them to improve their performance based on experience, just as humans do (Gracia, *et al.*, 2008).

Timely disease detection and control when observable symptoms persist is one of the major problems of management, as they are often used in disease prevention, monitoring and control. Consequently, late discovery could lead to total loss to the owner thereby portraying poor managerial principle. In the area of swine rearing and management, there is need for adequate awareness on diseases control and management, accurate and automated analysis of detection, diagnosis, and control.

Diseases in livestock leads to low productivity and when not discovered and treated timely could lead to total loss to the owner. However, the farmer may notice some symptoms but no means of diagnosing them. Early detection of observable symptoms will help the farmer to proffer a reliable and curative means of curtailing the debilitating effects the disease would have had on his production. Disease diagnosis are rapidly becoming computerized or digitalized for faster, convenient, reliable, cost efficient disease management. Many farmers who operates in this digitized form enjoys the opportunity of early and timely diagnosis and control without necessarily consulting a veterinarian which may not be readily available or accessible thereby reducing the risk of total loss of produce (Mulder, *et al.*, 2008). By streamlining and automating the disease diagnosis process, prospective farmers are looking for an application that will help them overcome these challenges, increase productivity and satisfaction.

Swine diseases are contagious viral diseases that affect pigs of all ages, inducing a haemorrhagic fever. It can appear in a variety of forms ranging from peracute, acute, subacute, to chronic and unapparent. It is most often recognized in the acute form with an associated lethality of up to 100 percent (Gallardo *et al*, 2011).

Swine diseases are severe threat to pig production systems. It not only threatens food security and challenges the livelihoods of pig producers and other actors in the supply chain, but may also have major consequences on international trade as a result of trade restrictions (Gallardo *et al.*, 2011). Although African wild suids do not show clinical signs of infection, they are, together with *Ornithodoros* soft ticks, the natural hosts and reservoir of the virus, while domestic pigs are accidental hosts (Haresnape & Mamu, 1986). In domestic pigs, swine diseases is transmitted mainly through direct contact, via the oro-nasal route, through excretions from infected pigs, or from ingestion of pork or other contaminated products containing the virus (e.g. swill, waste, carcasses, etc.) (Gallardo *et al*., 2015). Further transmission pathways are indirect contact through fomites or vector-borne transmission through bites from infected *Ornithodoros* soft ticks, where present. The disease is not a zoonosis, i.e. it does not infect humans (Malogolovkin *et al.*, 2015).

Any country with a pig sector is at risk of swine disease since there is currently no most effective vaccine or treatment for some swine diseases, the best strategy against swine disease for countries or pig farms that are still free of the disease is preventing the entry of the virus and early detection of the virus (OIE, 2013; OIE 2016).

Swine diseases are highly contagious and economically significant viral diseases of pigs. The severity of the illness varies with the strains of the virus, the age of the pig, and the immune status of the herd. Acute infections, which are caused by highly virulent isolates and have a high mortality rate in naive herds, are more likely to be diagnosed rapidly (Quembo *et al*., 2014). However, infections with less virulent isolates can be more difficult to recognize, particularly in older pigs. The range of clinical signs and its clinical similarity to other diseases can make swine diseases challenging to diagnose. For these reasons, early detection of observable symptoms will help the farmer to proffer a reliable and curative means of curtailing the debilitating effects the disease would have had on his production

As a result, this study develops an expert system that detect and predicts swine diseases using machine learning techniques (Logistic Regression, Support Vector Machine and Decision Tree algorithms); and to evaluate the performances of these machine learning algorithms.

Several related research works have been carried out on deploying machine learning algorithms for animal disease diagnosis.

Jyoti and Ujma (2011) worked on swine disease identification system using Genetic algorithm for the prediction of swine diseases. The study also compared various machine learning techniques for the identification of swine diseases. The result showed a positive prediction of presence of swine disease with an estimated accuracy of 0.73%. However, there is need for improvement on the outcome. Also, Shouman *et al.* (2012) in their work adopted KNN algorithm for prediction of swine diseases. The efficiency of the algorithm was verified for the prediction of heart diseases. The result and performance of the KNN algorithm was tabulated and presented for better understanding of the model. Furthermore, Kumari and Pandey (2017) developed a model for the identification of swine disease using Decision Trees and Naïve Bayes techniques. The results generated revealed that the developed model predicted the presence of swine disease. Though, there is need for improvement on the performance and accuracy of the model.

Ankita *et al*. (2015) worked on swine flu prediction using KNN classifier and Euclidean distance. Parameters such as effected area in the country and swine flu symptoms were considered. The model produced significant result as it identified the highest rate of flu predictions in a particular area. But, future work is required on the accuracy of the model. Similarly, Harshavardhan *et al*. (2011) worked on the identification and the spread of Influenza using KNN algorithm and Naïve Bayes classifier techniques. The aim is to predict death rate and recovery rates from flu. The result showed great improvement in predicting death rate and recovery rate from influenza infection. However, improvement is needed on the accuracy of the model through availability of adequate dataset.

Materials and Methods

Dataset

Swine Disease Dataset was used to train and test models for diagnosis of swine diseases. This dataset consists of 10 independent variables as symptoms and one dependent variable as the class label that is used to diagnose heart disease. The symptoms' names are fever, loss of appetite, dull or depressed, red to purple skin lesions, respiratory distress, vomiting, diarrhea, bloody diarrhea, sudden death, and abortion. The class label has two values which are: 0 which represents the absence of symptom; while the value 1 represents the presence of symptom. Table 1 illustrates the complete information about the symptoms.

Table 1: Symptoms Name and their Class Value

S/N	Symptoms	Description
	Fever	1 – positive 0 - negative
	Loss of appetite	1 – positive 0 - negative
	Dull or depressed	1 – positive 0 - negative
	Red to purple skin lesions	1 – positive 0 - negative
	Respiratory distress	1 – positive 0 - negative
	Vomiting	1 – positive 0 - negative
	Diarrhea	1 – positive 0 - negative
	Bloody diarrhea	1 – positive 0 - negative
	sudden death	1 – positive 0 - negative
	Abortion	1 – positive 0 - negative

Proposed system of predicting swine disease

This proposed system includes five stages as follows:

i. Loading swine disease dataset
ii. Data pre- processing
iii. Cross-validation and hyper parameter tuning
iv. Classifiers
v. Evaluating classifiers.

Figure 1 below illustrates the architecture of the swine diseases diagnosing system.

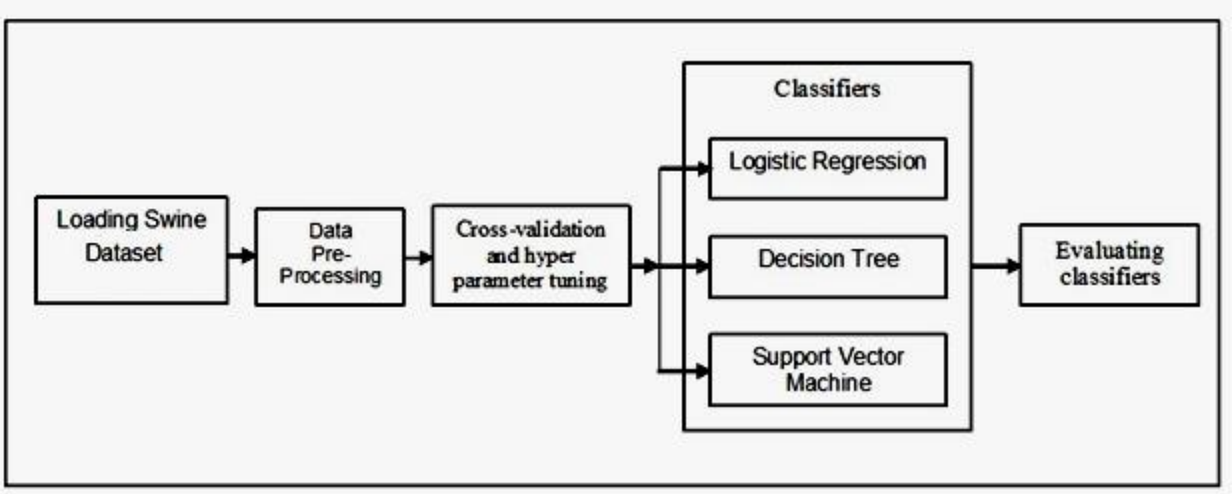


Figure 1: The architecture of the swine disease prediction system.

In data pre-processing stage, several steps were applied: categorical features was converted into numerical data; and imbalanced data was handled using random resample techniques (that is, data with an unbalanced ratio of values for each class label).

Also, in applying cross-validation and hyper-parameter tuning, a set of values was defined for each hyper-parameter for each class. Then, grid search method was applied to test each value and select the best values that achieve the best performance. Using K-Fold Cross-Validation (CV), the dataset was divided within k equal size of folds. K-1 groups were applied for the training, and the remaining part was utilized to evaluate the models. In this work, k = 10 was applied. In the 10 - fold CV process, 10% of data was used to test the models, and 90% was used to train the models.

Furthermore, for evaluating the performance of the models, confusion matrix was used to calculate accuracy, precision, and recall. Confusion matrix describes the performance of a model on a set of test data. It gives two types of correct predictions and two types of incorrect predictions for the classifier (Haq et al., 2018). Table 2 shows the confusion matrix. TP is the predicted output as true positive, TN is the predicted output as true negative, FP is the predicted output as false positive, and FN is the predicted output as a false negative. The accuracy, precision, and recall are defined in Table 2.

Table 2: Confusion Matrix

	Predicted Class 0	**Predicted Class 1**
Actual Class 0	TP	FN
Actual Class 1	FP	TN

a. Accuracy shows the performance of the classification system as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

b. Precision is the total number of correctly classified positive divide on the total number of predicted positive examples. The equation of the precision is given as follows

$$Prescision = \frac{TP}{TP + FP}$$

c. Recall: The equation of the recall is given as follows:

$$Recall = \frac{TP}{TP + FN}$$

System Design

Unified Modelling Language (UML) method was used for modelling and designing diagrams to offer a clear picture of the system developed. The study employed two different UML diagrams (Use case diagram and Activity diagram) for its design.

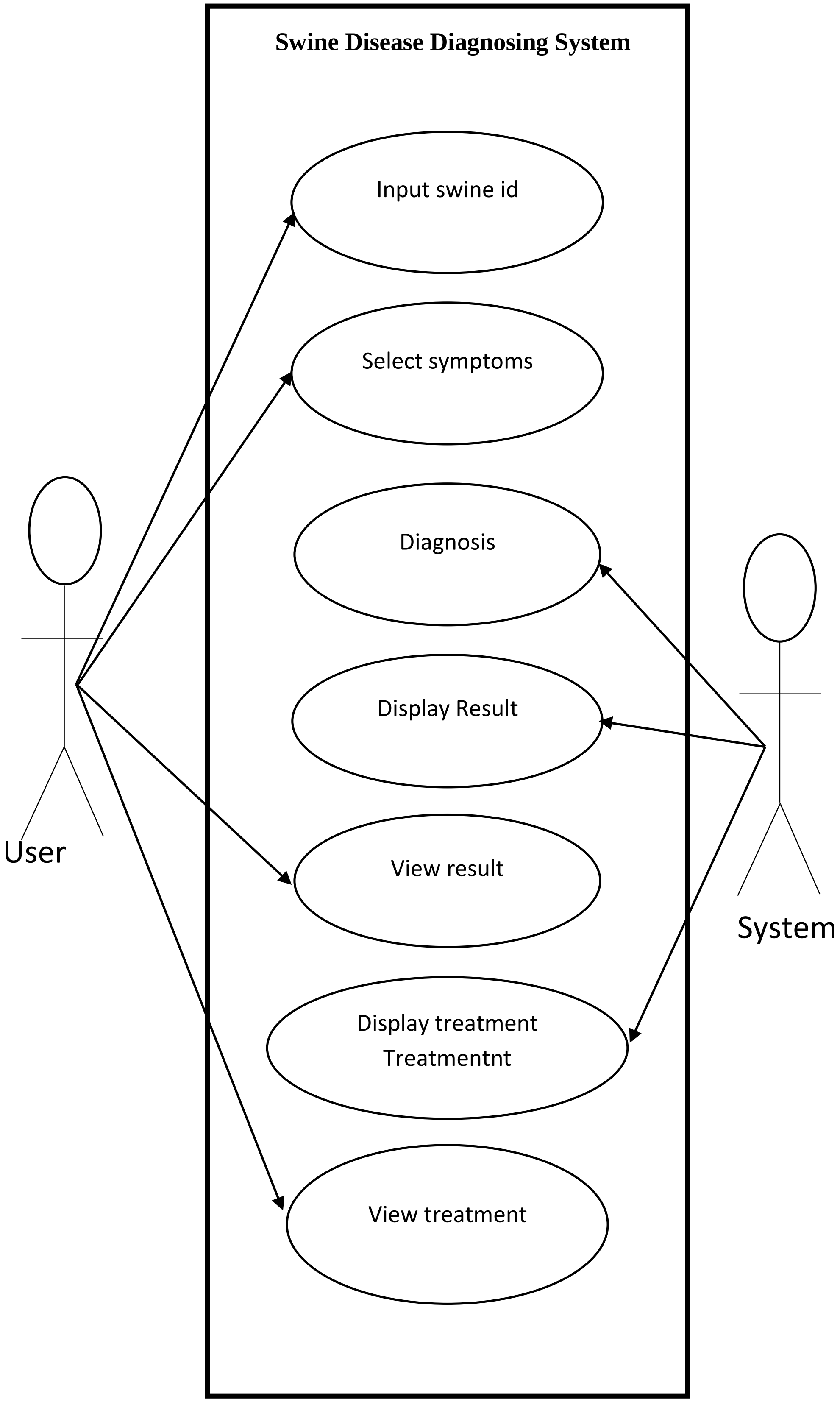


Figure 2: Swine Disease Diagnosis System use case diagram

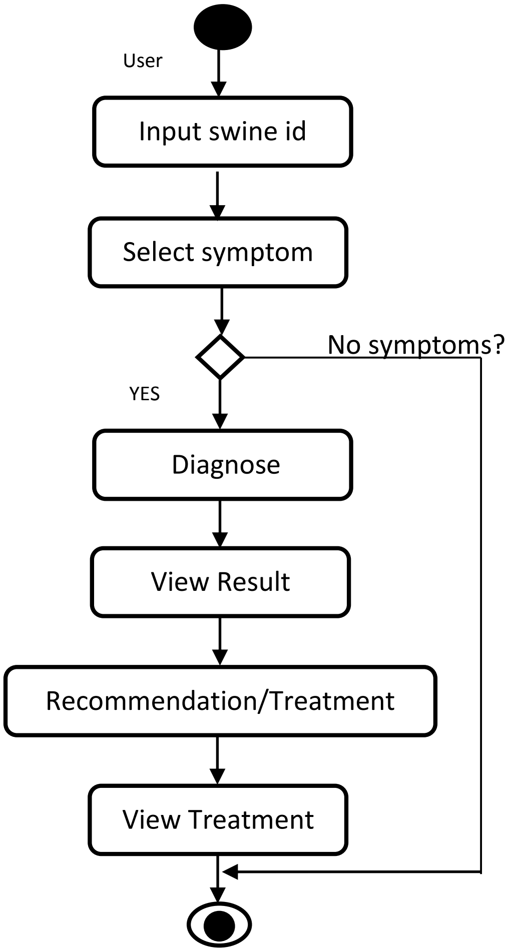


Figure 3: Swine Disease Diagnosis activity diagram.

Input Design

Input design facilitates the entry of data into the system. It involves the selection of best strategy for capturing data into the computer system as accurate as possible without error. The following are the input design specification for this study.

Table 3: Input Design of the Proposed System

SN	Field Name	Data Type	Description
1.	Swine Id	String	This field collects the Id of swine to be diagnosed
2.	Fever	Integer	This field collects the class label of fever symptom
3.	Loss of appetite	Integer	This field collects the class label of fever symptom
4.	Dull or depressed	Integer	This field collects the class label of Loss of appetite symptom
5.	Red to purple skin lesions	Integer	This field collects the class label of Red to purple skin lesions symptom
6.	Respiratory distress	Integer	This field collects the class label of Respiratory distress symptom
7.	Vomiting	Integer	This field collects the class label of vomiting symptom
8.	Diarrhea	Integer	This field collects the class label of Diarrhea symptom
9.	Bloody diarrhea	Integer	This field collects the class label of Bloody diarrhea symptom
10.	sudden death	Integer	This field collects the class label of sudden death symptom
11.	Abortion	Integer	This field collects the class label of Abortion symptom

In this study, Python was used as the front end because of its support to web applications, and is suitable for database management system that handless creation, storage, manipulation and data retrieval.

Python is a multi-paradigm programming language because it can be used in Web development, Artificial Intelligence, Machine Learning, Data science and networking.

Results and Discussion

System Implementation

i. *Index Window:* It is the first dialogue window of the designed software. It contains a start button which is linked to another window of the system.

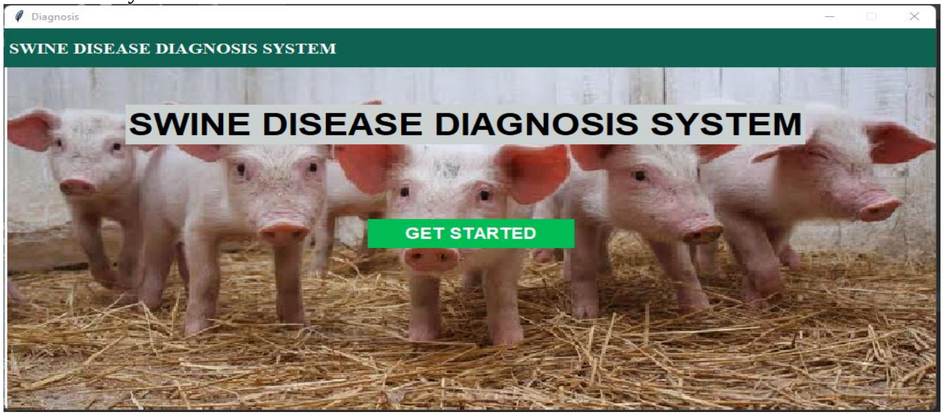


Figure 3: Index window of the Swine Disease Diagnosis System

ii. *Symptoms Window:* This window has several input fields to collect symptoms of swine. The collected symptoms are analyzed to diagnose for the presence of swine disease.

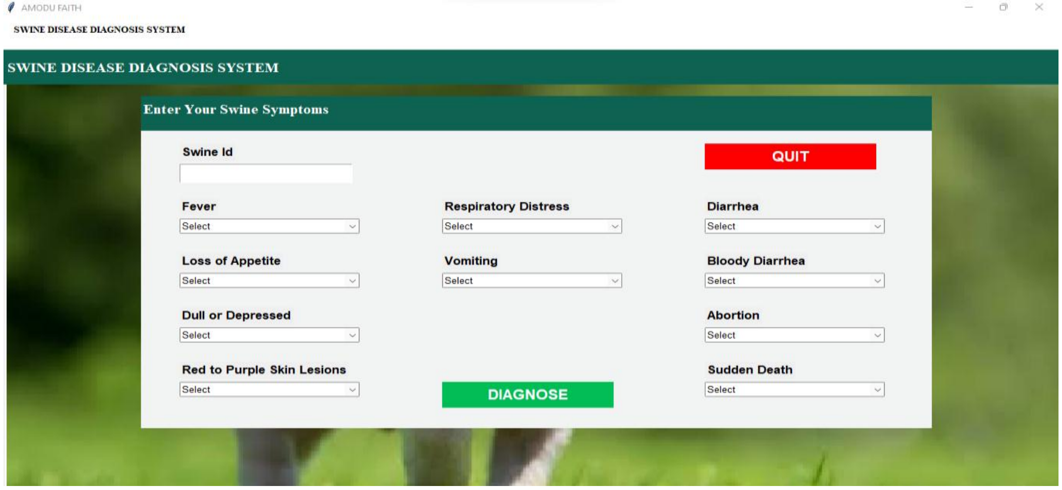


Figure 4: Symptoms window of the Swine Disease Diagnosis System

iii. *Diagnosis Result window:* This windows shows diagnosis report and treatment recommendation for the diagnosed swine disease.

...::Swine Disease Diagnosis Result::...

SWINE DIAGNOSED DISEASE

['SWN1212 Diagnosed Disease : Pasteurellosis']

TREATMENT RECOMMENDATION

['Treatment Recommendation : Clavulanic Acid']

Figure 5: Pasteurellosis Diagnosis and Treatment Recommendation Output

...::Swine Disease Diagnosis Result::...

SWINE DIAGNOSED DISEASE

['SWN01 Diagnosed Disease : African swine fever (AsF)']

TREATMENT RECOMMENDATION

['Treatment Recommendation : Oseltamivir']

Figure 6: African Swine Fever Diagnosis and Treatment Recommendation Output

...::Swine Disease Diagnosis Result::...

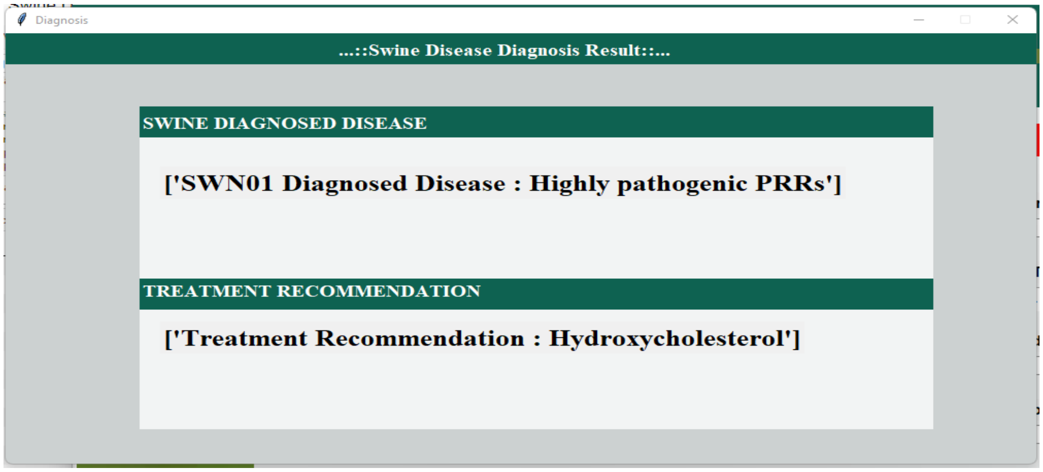
SWINE DIAGNOSED DISEASE

['SWN01 Diagnosed Disease : Classical swine fever (CsF)']

TREATMENT RECOMMENDATION

['Treatment Recommendation : Oxytotracycline']

Figure 7: Classical Swine Fever Diagnosis and Treatment Recommendation Output



...::Swine Disease Diagnosis Result::...

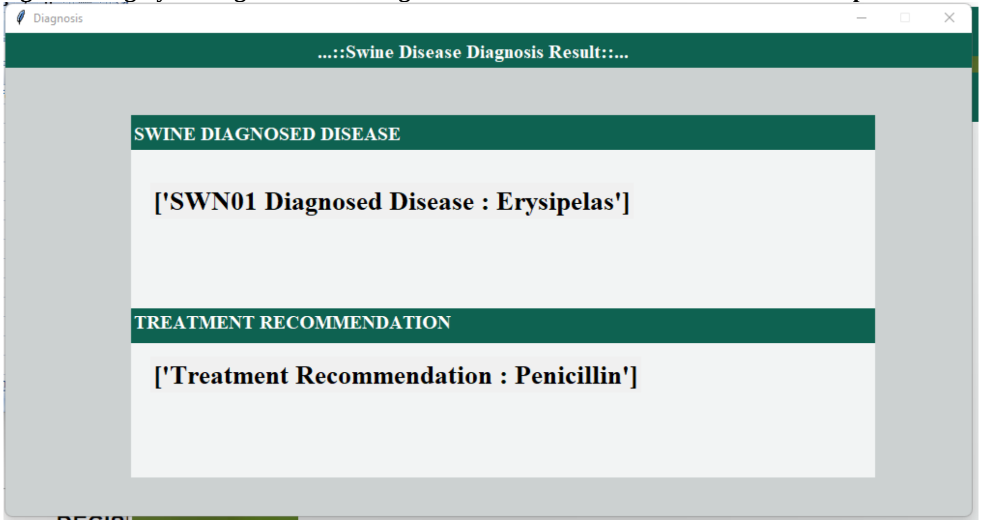
SWINE DIAGNOSED DISEASE

['SWN01 Diagnosed Disease : Highly pathogenic PRRs']

TREATMENT RECOMMENDATION

['Treatment Recommendation : Hydroxycholesterol']

Figure 8: Highly Pathogenic PRRS Diagnosis and Treatment Recommendation Output



...::Swine Disease Diagnosis Result::...

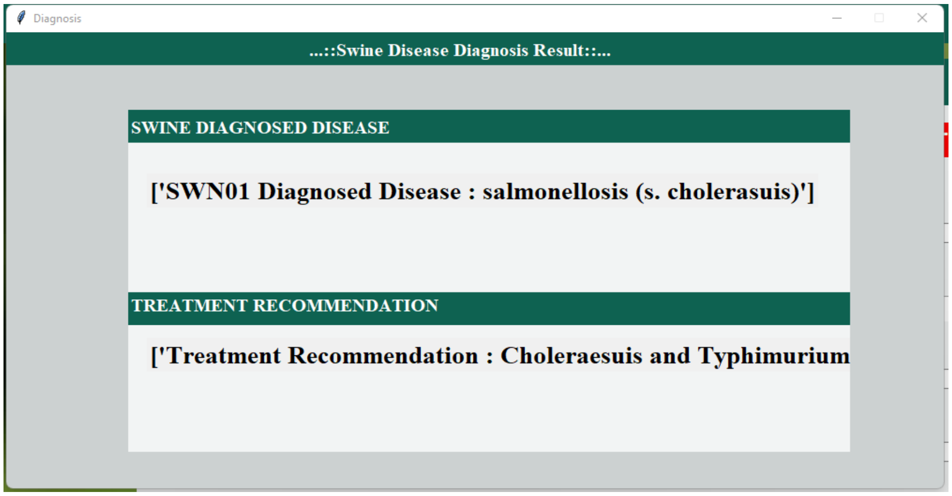
SWINE DIAGNOSED DISEASE

['SWN01 Diagnosed Disease : Erysipelas']

TREATMENT RECOMMENDATION

['Treatment Recommendation : Penicillin']

Figure 9: Erysipelas Diagnosis and Treatment Recommendation Output



...::Swine Disease Diagnosis Result::...

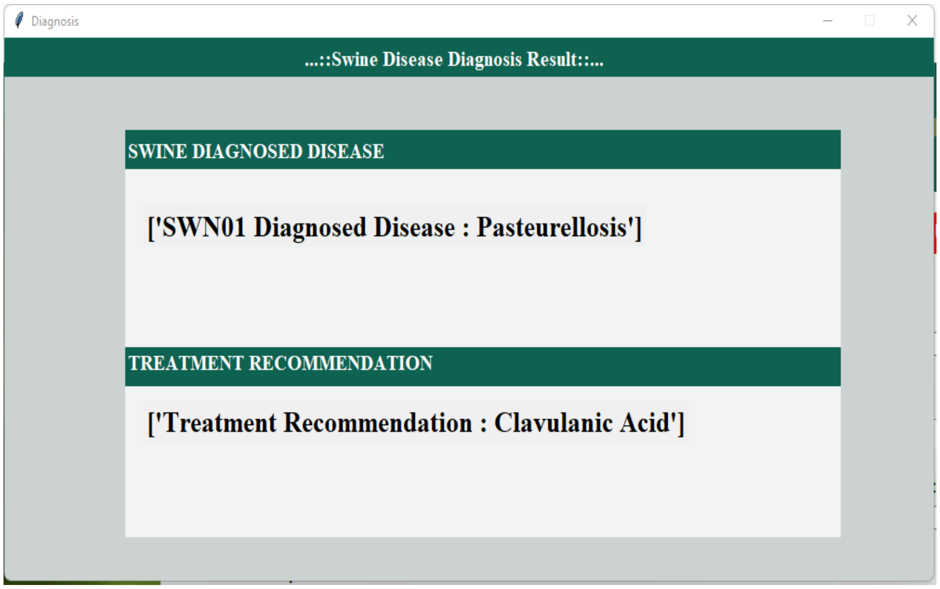
SWINE DIAGNOSED DISEASE

['SWN01 Diagnosed Disease : salmonellosis (s. cholerasuis)']

TREATMENT RECOMMENDATION

['Treatment Recommendation : Choleraesuis and Typhimurium

Figure 10: Salmonellosis Diagnosis and Treatment Recommendation Output



['SWN01 Diagnosed Disease : Pasteurellosis']

['Treatment Recommendation : Clavulanic Acid']

Figure 11: Pseudorabies Diagnosis and Treatment Recommendation Output

System Testing

During testing, the software was executed with set of test cases and the output of programs for the test cases was evaluated to determine if the program is performing as specified. The results of the test carried out are shown in the Table 4.

Table 4: System Test Results

SN	Test Cases	Test Objectives	Result	Reference
1.	T_1	To test if the system can predict the presence of African Swine Fever	During system testing, the system was able to predict the presence of African Swine Diseases.	Figure
2.	T_2	To test if the system can predict the presence of Classical Swine Fever	During system testing, the system was able to predict the presence of Classical Swine Fever	Figure
3.	T_3	To test if the system can predict the presence of Pseudorabies	During system testing, the system was able to predict the presence of Pseudorabies	Figure 15
4.	T_4	To test if the system can predict the presence of Highly Pathogenic PRRS	During system testing, the system was able to predict the presence of Pathogenic PRRS	Figure 12
5.	T_5	To test if the system can predict the presence of Erysipelas	During system testing, the system was able to predict the presence of Erysipelas	Figure 13
6.	T_6	To test if the system can predict the presence of Salmonellosis	During system testing, the system was able to predict the presence of Salmonellosis	Figure 14
7.	T_7	To test if the system can predict the presence of Pasteurellosis	During system testing, the system was able to predict the presence of Pasteurellosis	Figure 9
8.	T_8	To test if the system can make possible treatment recommendation	During system testing, the system was able to make possible treatment recommendation	Figure 10-15

Based on the result gotten from the test carried out, the system was able to diagnose the presence of swine diseases from the given dataset. Results gotten for each test case are described as follows:

T_1 – System test result showed that the system was able to predict the presence of African Swine Disease.
T_2 – System test result showed that the system was able to predict the presence of Classical Swine Disease.
T_3 – System test result showed that the system was able to predict the presence of Pseudorabies.
T_4 – System test result showed that the system was able to predict the presence of Pathogenic PRRS.
T_5 – System test result showed that the system was able to predict the presence of Erysipelas.
T_6 – System test result showed that the system was able to predict the presence of Salmonellosis.

T_7 – System test result showed that the system was able to predict the presence of Pasteurellosis.

T_8 – System test result showed that the system was able to predict possible treatment recommendation.

Furthermore, three machine learning algorithms are compared (support vector machine, decision tree and logistic regression) in diagnosing the presence of swine diseases on the basis of the input dataset. The performance analysis of these algorithms are presented Table 5 with Accuracy, Precision, Sensitivity and Specificity of each algorithm observed.

Table 5: Performance Analysis Table

Algorithm	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**
Decision Tree	97.5%	100%	94.7%	100%
Logistic Regression	75%	0.00%	0.00%	100%
Support Vector Machine	88.8%	100%	87.5%	100%

From Table 5 Decision Tree (97.5%) showed the highest accuracy as compared to Support Vector Machine (88.8%) and Logistic Regression (75%). Also, Decision Tree (100%) and Support Vector Machine (100%) showed the best precision as compared to Logistic Regression (0.00%). In the same vain, Decision Tree (94.7%) showed the highest sensitivity as compared to Support Vector Machine (87.5%) and Logistic Regression (0.00%). However, the three algorithms showed the same specificity at 100% as observed.

Conclusion

Classification is a major data mining technique primarily used in healthcare sectors for medical diagnosis and predicting diseases. Recent advances in Artificial intelligence (Al) has provided more sophisticated methods and techniques for solving tasks previously difficult to handle within the medical domains. In this study, a combination of various machine learning algorithms (Decision Tree (DT) and Logistic Regression and Support Vector Machine (SVM)) was used for detection, diagnosis and treatment of swine diseases. Comparisons of these algorithms were carried out based on certain performance criteria. The GUI contains input fields for all attributes in the dataset. The system proffer treatment recommendation based on the trained model after detecting and predicting a disease. Thus, this work demonstrates the utility and relative ease of using machine learning algorithms to assess the state of swine health based on symptoms characteristics. Hence, the developed system is a useful tool for pig farmers and veterinarians in the early diagnosis and treatment of swine diseases.

Conflict of Interest

The authors declare no conflict of interest, financial or otherwise.

References

Borkar, A. R. & Deshmukh, P. R. (2015). Naïve Bayes classifier for prediction of swine flu disease. *International Journal of Computer Sciences and Engineering,* 5(4), 120-123.

Gallardo, C., Nieto, R., Soler, A., Pelayo, V., Fernández-Pinero, J., Markowska-Daniel, I., Pridotkas, G., Nurmoja, I., Granta, R., Simón, A., Pérez, C., Martín, E., Fernández-Pacheco, P. & Arias, M. (2015). Assessment of African swine fever diagnostic techniques as a response to the epidemic outbreaks in Eastern European Union countries: How to improve surveillance and control programs. *Journal of Clinical Microbiology,* 53(8), 2555-65. doi: 10.1128/JCM.00857-15.

Gallardo, C., Okoth, E., Pelayo, V., Anchuelo, R., Martín, E., Simón, A., Llorente, A., Nieto, R., Soler, A., Martín, R., Arias, M. & Bishop, R.P. (2011). African swine fever viruses with two different genotypes, both of which occur in domestic pigs, are associated with ticks and adult warthogs, respectively, at a single geographical site. *Journal of General Virology.* 92(2), 432-44. doi: 10.1099/vir.0.025874-0.

Garcia, B., Pham, H., & Triantraphyllou, E., (2008). Prediction or diabetes by employing a new data mining approach which balances fitting and generation. *Computer and info. Science.* 11-26.

Haque, M., Sartelli, M., Mckimm, J. & Bakar, M. A. (2018). Health care-associated infections – an overview. *Infection and drug resistance,* 11, 2321.

Haresnape, J. & Mamu, F.D. 1986. The distribution of ticks of the *Ornithodoros moubata* complex (Ixodoidea: Argasidae) in Malawi, and its relation to African swine feverepizootiology. *Journal of Hygiene, Cambridge* 96 (3), 535–544.

Harshavardhan, A., Gandhe, A., Lazarus, R., Yu, S. & Liu, B. (2011). Twitter improves seasonal influenza prediction. HEALTHINF.

Jyoti, D. S. & Ujma, S. A. (2011). Predictive data mining for swine medical diagnosis: An overview of swine disease prediction. *Int. J. Com. Appl,* 43-48.

Kumari, R., & Pandey, R. (2017). An efficient resource scheduling algorithm using knapsack. *International Journal of Computer Science Engineering and Information Technology Research,* 7(2), 25-28.

Malogolovkin, A., Burmakina, G., Titov, I., Sereda, A., Gogin, A., Baryshnikova, E. & Kol-basov, D. (2015). Comparative analysis of African swine fever virus genotypes and serogroups. *Emerg Infect Dis.* 21(2):312-5. doi: 10.3201/eid2102.140649.

Mangesh, J., Shinde, S. & Pawar, S. (2015). Comparative study of decision tree algorithm and Naive Bayes classifier for swine flu prediction. *Int JRes in Eng. and Tech,* 4(6), 45-50.

Mulder, A., Mebus, C., & Arias, M. (2008). Survival of Several Porcine Viruses in Different Spanish Dry Cured Meat Products. *Food Chemistry,* 5(9), 555-559.

OIE African swine fever Technical Disease Card. 2013. Available at http://www.oie.int/ fileadmin/Home/eng/Animal_Health_in_the_World/docs/pdf/Disease_cards/AFRICAN_ SWINE_ FEVER.pdf

OIE Manual of Diagnostic Tests and Vaccines for Terrestrial Animals. 2016. Available at:http://www.oie.int/en/international-standard-setting/terrestrial-manual/access-online/

Quembo, C.J., Jori, F., Heath, L., Pérez-Sánchez, R. & Vosloo, W. (2014). Investigation into the epidemiology of African swine fever virus at the wildlife-domestic interface of the Gorongosa National Park, central Mozambique. *Transboundary and Emerging Diseases*. (e-pub and Online print)

Sajja, A., Feigenbaum, E., & McCorduck, P. (2012). The Fifth Generation. Addison - Wesley. ISBN 978-0451152640

Shortlife, E. H., & Cimino, J. J. (2006). Biomedical informatics' Computer applications in health care and biomedicine. New York, NY: Springer.

Shouman Mai, Tim Turner, & Rob Stocker (2012). Applying K-Nearest Neighbour in Diagnosing Swine Disease. *Int JInf and Edu Tech,* 2(3), 220-223.