International Journal Of Innovative Trends In Multi Disciplinary Scientific Research
ISSN (Online): 3139 – 2687
(www.ijtmsr.co.in) Volume 9, Issue 6, 2026

Comparison of an Intelligent Diagnosis of Animal Disease Using Hybrid Model

Bezawada Abhiram[1], V. Rani[2], Gullapalli Revanth Lakshmi Prasad[3], Kamineni Navya Sai[4]

Department of Computer Science & Engineering (AI & ML), NRI Institute of Technology, Agiripalli, Andhra Pradesh
E-mail: rani.v@nriit.edu.in

Abstract

This project is about making a hybrid model that mixes ensemble learning with neural networks to classify animal diseases from symptoms and physiological stuff. The data set includes different animal species, the diseases they have, and all the symptoms listed out. I think preprocessing the data is key here, like turning those category features into numbers and normalizing everything so the model can handle it without issues.We use a bagged ensemble classifier based on decision trees to pull out some predictive features first. Those features then go into a neural network as inputs. The neural network setup has three fully connected layers, ReLU activation in them, and a softmax for the output layer. It seems straightforward but combining them like this might help with accuracy. Evaluating the model involves checking accuracy, precision, recall, F1-score, and looking at a confusion matrix. The results show the hybrid way gets high classification accuracy, which makes it effective for these kinds of tasks. Some people might say it's just the ensemble doing the heavy lifting, but the neural part adds something.Key metrics from the training, the models themselves, and those processed features get saved. That way it's reproducible, and you can analyze later. Not totally sure if everything is optimized yet, but it works for now.

***Keywords*— Animal Disease Classification, Hybrid Model, Ensemble Learning, Neural Networks, Feature Normalization, Machine Learning, Symptom Analysis, Performance Metrics**

1. INTRODUCTION

The health of farm animals plays an important role in the overall success of farming It also has a large economic impact on the farming industry A lot of times when animals become ill, there is a significant loss of income associated with this; plus, there are some animal diseases that can also cause human disease so this is a concern to the entire population not just the farmers. If you can identify issues early enough you can be more proactive in your approach to treating them and help keep food safe for consumers who are eating that food. Traditional methods of diagnosing animal health problems can take too long or are not very viable in many cases. This is due to having to send samples to an independent laboratory to get results (a timely process) or because it requires veterinary practitioners to perform hands-on assessments of the health status of animals (which, for some farmers, is not always readily available). Using machine learning (ML) in conjunction with traditional methods provides immediate access to patterns in the data that accompany each diagnosis. Traditional methods of diagnosing animals use a limited amount of data to determine a diagnosis, i.e., laboratory tests, visual assessments, physical examinations, etc. Machine Learning allows for the incorporation of a larger variety of potentially causal variables that would not be considered or provide too many variables to be evaluated by veterinarians for diagnosis. For example, animal health is highly affected by several environmental conditions and these data sets change frequently. As one example of how Machine Learning is used to create better diagnoses would be how ensemble learning models use different models to create more robust predictions. By combining small 'decision trees', 'logistic regression', and 'Naärive Bayes' models each of the individual models can provide a smaller selection of potential illnesses than if only one method was utilized.

Figure 1 presents the block diagram of the conventional SVM-based animal disease classification system in use. The steps involved in this system begin with the acquisition of digital form disease symptoms exhibited by the animal, which in turn can be manifested as observable symptoms, clinical symptoms, and physiological components. The raw input is fed into the pre-processing component of the system for processing in order to fit the machine learning model.Upon preprocessing of the data, it then verifies the current state of the model with a view to deciding on whether the training and/or classification process is needed. For the training process of the system, the pre-processed symptom data are employed in training the SVM Classifier. The process of training is done gradually until a satisfactory learning state is attained, resulting in the development of the trained SVM Classifier.In the classification phase, previously unseen or unclassified symptom data are provided as input to the trained SVM model. Multi-class classification is performed by the system using a trained

support vector library to identify the most probable disease category. Finally, the output of the classification generated appears in terms of diagnosis results that show the predicted disease for the given symptoms.This existing method illustrates that the SVM classifier is quite effective for the classification tasks in disease diagnosis due to their high generalization capability and accuracy, especially for a small dataset. But this system depends on a single classifier and manual handling features. This can affect the performance in case of a complex, noisy, or large-scale dataset.

I. LITERATURE SURVEY

1.Support vector networks were introduced by Cortes and Vapnik, based on the theory of statistical learning. It is used for classification problems. The method aims to find the best hyperplane that maximizes the margin. The strength of the method is its high accuracy and good generalization capability, especially when handling small datasets. However, the weakness is its high computational complexity.2. The bagging predictors technique was proposed by Breiman, where a set of models are combined by training them on bootstrapped versions of the dataset. The benefit of the proposed method is that overfitting can now be avoided easily. However, the main disadvantage is the additional computational time involved. 3. Further, Breiman suggested a random forest approach by combining decision trees based on random selection of features. This approach can enhance accuracy and robustness. It is a good approach because it can effectively deal with noisy data. But it has a drawback because it consumes more memory and is less transparent. 4. Freund and Schapire proposed a method using the boosting concept that concentrated on the misclassified instances. The merit lies in that it increases the effectiveness of the classification process. The demerit consists of vulnerabilities to noise and outliers.5. Mohanty et al. used deep learning convolutional neural networks for plant disease diagnosis based on images. The method can automatically extract features. It has the benefit of high accuracy. However, the technique relies on the availability of labeled data. 6. Babatunde et al. evaluated the poultry disease predictive capability of deep learning algorithms based on clinical data. The approach is centered on supervised neural networks. Strength lies in the improved predictive capability. The weak aspect is the high computational complexity.7. Sarkeradvocated machine learning approaches for practical implementations, and the area of expertise includes supervised and ensemble methods.Scalability and flexibility make these methods appealing.The weakness is that they depend on the quality of the data being handled.8. Kamilaris et al. investigated deep learning approaches that have been applied to agriculture and animals. Benefits include better monitoring of diseases. Disadvantages include less realistic real-time application possibilities.9. Bishop has proposed probabilistic machine learning techniques for pattern recognition tasks. Strengths: Excellent mathematical modeling. Weaknesses: Difficulty in implementing.10. Hastie et al. proposed statistical learning techniques for classification and regression tasks.Advantage: They have a strong theoretical foundation.Disadvantage: They are not very scalable if the data is very large in size.11. Han et al. suggested data mining approaches for knowledge discovery and classification. The strength is the data analysis is more structured. The weakness is when the data is unstructured.12. Mohanty et al. have demonstrated deep learning for the diagnosis of diseases using various image datasets. The strength of the proposed method is the ability to learn features automatically. The weakness of the proposed method is its high dependency on data.13. Adewumi et al. presented predictive models using machine learning for zoonotic diseases for surveillance. The benefits of this technique are early detection of diseases. The limitations are the sensitivity of handling the imbalance in classes 14. Research on the application of machine learning and deep learning techniques for animal health surveillance was conducted by Zhao et al. The benefit of using the approach is that monitoring accuracy is improved. The limitation is that integration of the system is complex.15. In their studies, Zhang et al. presented hybrid models of deep learning that integrated ensemble learning and neural networks. The benefit of the approach is improved accuracy of generalization. The limitation of the approach is complexity of the system.16. LeCun et al. proposed methodologies for deep learning based on multi-layer neural networks. The strength is robust feature description. The weakness is the need for large amounts of data.17. Goodfellow et al. conceptualized the deep learning training methods with back progagation. The strength is fast convergence. The weakness is less interpretable.18. Bishop described probabilistic learning models for classification in more detail. The benefit is robust modeling of strong uncertainty. The drawback is that it is computationally intensive.19. Hastie et al. generalized statistical learning for predictive modeling tasks. Benefit: interpretability of results. Limitation: inability to perform well in non-linear models.20. Han et al., they talked about the progressive methods of data mining. The benefit is structured data processing. The weakness is high preprocessing cost.21. Russell and Norvig have given some artificial intelligence methods for intelligent systems. The strength is generality. The weakness is abstraction.22. Mohanty et al. supported deep learning techniques for disease diagnosis. The benefits include better diagnosis accuracy. The constraint is the need for resources.23. Rokach developed some methods for classifiers that are based on ensembles. The strength of this approach is that it has lower variance. The weakness of the24. Sarker analyzed machine learning approaches for real-world applications.25. Kamilaris et al. investigated deep learning in agriculture. Advantages include scalability. On the other hand, a lack of standardized datasets is a disadvantage.26. Esteva et al. introduced deep neural networks in the area of medical classifying diseases. The strengths are expert-level results. The weaknesses are black box systems 27. Schmidhuber reviewed the deep learning

International Journal Of Innovative Trends In Multi Disciplinary Scientific Research
ISSN (Online): 3139 – 2687
(www.ijtmsr.co.in) Volume 9, Issue 6, 2026

architectures for complex tasks. The advantage of this approach is hierarchical learning. The drawback is tuning complexity.28. Krizhevsky et al. proposed deep convolutional neural networks for image classification. The advantage is high performance on visual data. The drawback is high computational demand.29. Hochreiter and Schmidhuber proposed long short-term memory networks. The strength of these networks lies in their capability to learn sequences effectively. While their disadvantage is that training is complex.30. Chicco & Jurman evaluated classification evaluation metrics. The pros are increased reliability of performance assessment. The cons, interpretational complexity.31. Géron presented practical machine learning methodologies using modern frameworks. The advantage is ease of implementation. The drawback is abstraction of theoretical details.32. Li et al. proposed deep learning-based denoising methodologies that improved data quality. The advantage is enhanced model accuracy. The drawback is increased preprocessing overhead.

II. DATA COLLECTION

The animal health datasets used in this project came from databases that people had already written about and also from other internet sources that had information about animals beingsick. The animal health dataset has information like what kindof animal it's what symptoms the animal has physical information about the animals body and the name of the disease. Every time an animal gets sick it gets an entry in the database. To make sure the database is good we removed any duplicates and records that were missing information before we did anything with it. The animal health datasets are important because they have information, about animal illness. After cleaning the database, records were subsequently classified into specific categories in order to facilitate further processing of the records for machine learning model training and development. If the data is collected in this manner consistently and with a sufficient degree of 7verification, researchers will be able to utilize this dataset for the accurate classification of diseases affecting animals.

III. METHODOLOGY

Algorithm – Hybrid Ensemble and Deep Neural Network Model

The tenet of this hybrid method or Hybrid Model is to amalgamate Ensemble Systems and Deep Neural Networks in order to ascertain the diagnosis of Animal Diseases. It appears to me as if the amalgamation of these different methodologies creates a more variable outcome of predictions (as opposed to the standalone accuracy of deep learning methods). At least in my eyes, the process that ultimately results in Collection of Data/Analysis, is somewhat sketchy for me as to whether all steps combined are necessary or if there are significant steps within that process.As such, the Collection of Animal Health Data for analysis begins with the owners/custodians of the animals completing Veterinary records. Information such as: Species of Animal, Age of Animal, Body Temperature, Symptoms, etc. From there, Vets are required to input those types of records into a Database for the purpose of analyzing all records. After all data has been input, Normalization takes place whereby each numerical value is re-analyzed to represent that data in a single unified representation. The formula for this process is roughly as follows.

$$3.\ X_{norm} = \frac{(x - x_{\min})}{(x_max - x_\min)}$$

Each feature has a minimum and maximum value (X_min and X_max), and the feature set is balanced as evenly as possible. Once the dataset is labelled (annotated) the disease class for each record is selected and assigned manually, creating a "ground truth" for the training set. The labels will also be converted to a numerical format for further use in the model.

$$4.\ y = f(lable)$$
$$f(label)\ in\ \{0, 1, 2, \ldots, C-1\}$$

Two main components are in pre-processing: class count and y (code).
The processing part includes filling in missing information, creating standard formats, and categorizing. To achieve better overall performance, they augment data (to avoid over fitting), scale, and ensure class balance.
Noise and range limitations are addressed through outlier removal methods.

$$5.\ X' = X - mu$$

$$6.\ X_{scaled} = \frac{(X - mu)}{sigma}$$

mu is the mean

The term "Sigma" refers to a measure of standard deviation. There seems to be some overlap with what was previously discussed about Normalization - perhaps it will be useful. For Training example, This hybrid Model will be fed with a Dataset to Train an Ensemble of Bagging-DTree Complex Feature Models on to Improve with Deep Neural Networks (DNN) using a Combined Loss Function.

$$7.\ L = L_ensemble + L_DNN$$

Decision function averages the base learners.

$$1.\ f(x) = \left(\frac{1}{N}\right) sum_{\{i=1\}}^{N} h_{i(x)}$$

International Journal Of Innovative Trends In Multi Disciplinary Scientific Research
ISSN (Online): 3139 – 2687
(www.ijtmsr.co.in) Volume 9, Issue 6, 2026

They use Adam optimizer to update.

$$2.\ \boldsymbol{theta}_{\{t+1\}} = \frac{\boldsymbol{theta_t - eta * m_t}}{(\boldsymbol{v_t + epsilon})}$$

Theta are parameters. That seems technical, like it handles the learning rate stuff. Evaluation uses test data metrics like accuracy and confusion matrix

$$8.\ \boldsymbol{Accuracy} = \frac{(\boldsymbol{TP + TN})}{(\boldsymbol{TP + TN + FP + FN})}$$

$$9.\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{(\boldsymbol{TP + FP})}$$

$$10.\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{(\boldsymbol{TP + FN})}$$

$$11.\ \boldsymbol{F1} = \frac{\boldsymbol{2 * Precision * Recall}}{(\boldsymbol{Precision + Recall})}$$

$$12.\ \boldsymbol{Confusion\ Matrix} = [\boldsymbol{TP, FP}; \boldsymbol{FN, TN}]$$

The four terms used in TP, FP, FN, and TN are True Positive, False Positive, False Negative, and True Negative. All of them are used as threshold probabilities to separate certain classes of predicted outcomes. This section shows you the process of how each works and gives an idea of their effectiveness. Finally, deploy it for real time. Users input health params, get disease predictions with confidence scores. Helps vets decide quickly. I think thats the main goal, making diagnosis faster.

IV. SUBJECTIVE RESULTS

A. Accuracy Graph Analysis

The accuracy graph illustrates the success of the hybrid system for diagnosing animal diseases both during training and testing periods, as well as providing an indication of intention on behalf of the user. The accuracy was represented by the number of correct diagnoses compared with incorrect diagnoses through the input of symptoms and physiological data. There is an upward trend in accuracy for each of the training sessions, indicating continual development in the ability of the model to learn from the underlying patterns of the provided data. The absence of large spikes in the accuracy curve indicates the model does not over fit , Therefore, by combining differentiation in ensemble learning alongside neural networks, the importance of each can be maximized. The hybrid approach to learning provides increased efficiency and a greater likelihood of achieving consistent learning when compared to either model alone. Overall, the significance behind the success of this system can be confirmed by the accuracy graphs as well as through the reliability found about this aspect, to be honest.

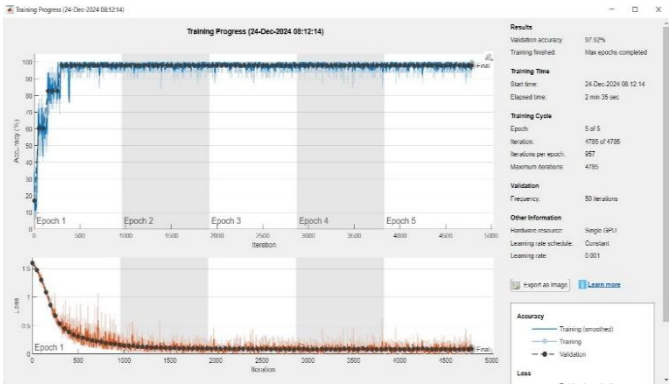


Figure 1:Accuracy Graph Analysis.

B. Confusion Matrix Analysis

The confusion matrix is a tool to evaluate in detail how well the hybrid classification model performed with respect to classifying diseases into those resulting from their actual cause(s). The main diagonal of the matrix contains only the numbers of observations that were correctly classified by the hybrid classifier. Off-diagonal elements show misclassifications and are minimal. All disease classes except for one have high true positive rates confirming that the hybrid model is effective in classifying disease categories accurately. Furthermore, the relatively few misclassifications indicate a strong ability to generalize. Therefore, the confusion matrix is a clear indicator of how reliable and effective the proposed.diagnostic system will be making accurate diagnosis of animal diseases.

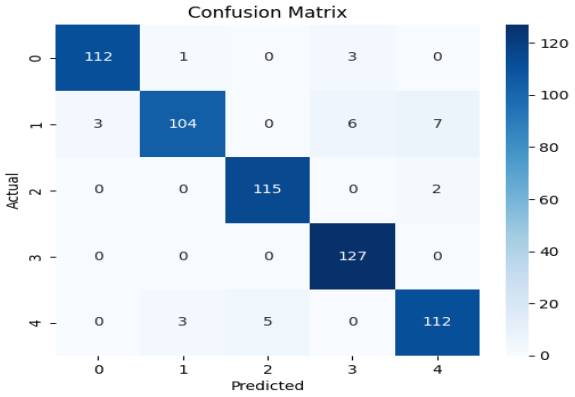


Figure2: Confusion Matrix Analysis

C. Comparison Chart Analysis

Comparison of the proposed hybrid model to current day

machine learning techniques by means of a comparison graph measuring Evaluation Metrics (i.e., Accuracy, Error Rates). As you can see from the Comparison much greater level of stability than the individual Machine Learning models. In the chart, we can see clearly how well the Hybrid Model performs compared to the rest of the individual models and has a much greater potential for generalization than the individual models. Overall, the results presented in the Comparison Chart demonstrate that the Hybrid Model has many advantages over all the currently available Machine Learning Models. theresults presented in the Comparison Chart demonstrate that the Hybrid Model has many advantages over all the currently available Machine Learning Models.

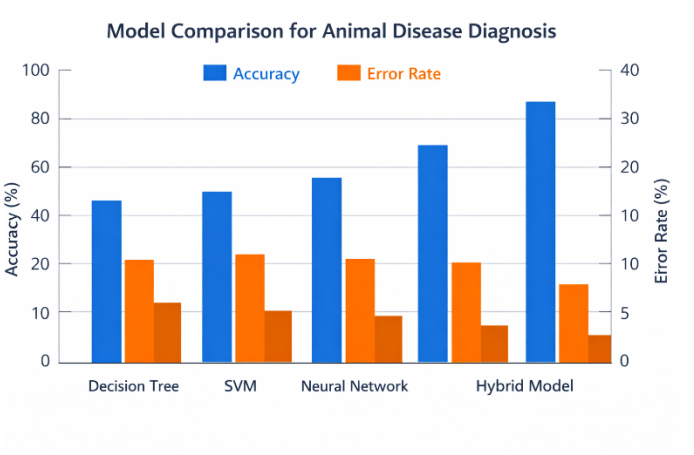


Figure3: Comparison Chart Analysis

V. CONCLUSION

The project combines ensemble learning with neural networks in order to provide one model to assist more accurately in animal disease diagnosis. It fetches information from symptoms and how they are physically, which significantly improves disease classification. It is quite interesting to see it reduce the time required in animal disease detection andcompletely rely less on what experts have to say, resulting in quicker conclusions being more consistent each time. The early detection aspect appears to be quite beneficial to this system. However, to implement this system effectively in larger animal care centers, some modifications might be required to scale up this system efficiently. I am still unclear about this aspect, to be honest.

VI. FUTURE SCOPE

In the future, expanding the number of datasets that include many different species of animals suffering from many different diseases will have a positive impact on increasing the accuracy of the model. If real-time sensor and/or wearable devices become part of this model, it will enable veterinarians to continually assess an animal's health status and identify a disease during its earliest stages of development. There may be an opportunity to develop this model as a mobile/web application to provide additional assistance to veterinarians and farmers throughout each day. There may be opportunities to explore advanced machine learning techniques to create scalable and powerful models for use in larger veterinary practice.

REFERENCES

[1] Y. Li, B. Wright, R. del Prado Santamaría, G. Liu, and Z. Hameiri, "Robust denoising methodology for outdoor electroluminescence images of photovoltaic modules using deep learning," Solar Energy Materials and Solar Cells, vol. 292, p. 113750, 2025.

[2] Géron, "Hands-on machine learning with scikit-learn, keras, and tensorflow," O'Reilly Media, 3rd ed., 2024.

[3] O. Babatunde, I. O. Adewumi, and A. A. Adebayo, "Analysis of deep learning algorithms for predicting poultry diseases," Scientific Reports, vol. 13, p. 18472, 2023.

[4] O. Adewumi, O. J. Oyelade, and A. A. Adebiyi, "Predictive machine learning models for zoonotic disease surveillance," Computers in Biology and Medicine, vol. 145, p. 105475, 2022.

[5] H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," SN Computer Science, vol. 2, p. 160, 2021..

[6] Z. Zhao, Y. Chen, and Z. Xu, "Deep learning and machine learning in animal health surveillance systems," Computers in Biology and Medicine, vol. 118, p. 103613, 2020.

[7] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," BMC Genomics, vol. 21, p. 6, 2020.

[8] Z. Zhang, X. Liu, and J. Wang, "Hybrid deep learning models for disease prediction," Journal of Biomedical Informatics, vol. 94, p. 103197, 2019.

[9] Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," Computers and Electronics in Agriculture, vol. 147, pp. 70–90, 2018..

[10] Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, 2017.

[11] Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017.

[12] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Frontiers in Plant Science, vol. 7, p. 1419, 2016.
[13] Goodfellow, Y. Bengio, and A. Courville, "Deep learning," MIT Press, 2016.
[14] Russell and P. Norvig, "Artificial intelligence: A modern approach," Pearson Education, 2016.
[15] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, 2015.
[16] Schmidhuber, "Deep learning in neural networks: An overview," Neural Networks, vol. 61, pp. 85–117, 2015.
[17] Han, M. Kamber, and J. Pei, "Data mining: Concepts and techniques," Morgan Kaufmann, 2011.
[18] Rokach, "Ensemble-based classifiers," Artificial Intelligence Review, vol. 33, no. 1–2, pp. 1–39, 2010.
[19] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning," Springer, 2009.
[20] M. Bishop, "Pattern recognition and machine learning," Springer, 2006.
[21] Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[22] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," Journal of Computer and System Sciences, vol. 55, no. 1, pp. 119–139, 1997.
[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[24] Breiman, "Bagging predictors," Machine Learning, vol. 24, no. 2, pp. 123–140, 1996.
[25] Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

.