Research on Health Early Warning and Decision Support Model of Smart Breeding Based on Machine Learning Algorithm, Internet of Things and AI

Runbo Zhang

The High School Affiliated to Renmin University of China, Beijing, 100080, China

Zhangrunbo1223@163.com

Keywords: Smart Farming; Machine Learning Algorithm; Internet of Things; Health Early Warning Model; Decision Support Model

Abstract: This article focuses on the field of smart farming, and is committed to building a health early warning and decision support model based on ML (Machine learning) algorithm, IoT (Internet of Things) and AI (Artificial intelligence) to cope with the dilemma of traditional farming mode. Through IoT technology, multi-source big data such as breeding environment and animal signs are collected comprehensively, and the data are preprocessed by statistical analysis and feature engineering. Then, the health early warning model is built with the help of random forest algorithm, and the decision support model is built with decision tree algorithm as the core and multi-source information. Verified by the actual data of several farms, the accuracy of the health early warning model is 90.5% and the recall rate is 88.2%. The decision support model can reduce the average feed cost of farms by 10% and increase the breeding income by 15%. The research shows that the fusion model is highly accurate and practical in smart farming, which can provide effective decision support for aquaculture practitioners and help the smart farming industry to achieve efficient and scientific development.

1. Introduction

Under the background of the continuous growth of global population and the steady improvement of people's living standards, the demand for livestock products such as meat and eggs has shown a sharp rise [1]. The traditional farming model has been unable to meet the needs of the current market due to many disadvantages such as low efficiency, serious waste of resources and difficulty in accurately controlling the farming environment and animal health [2]. As a new type of breeding mode that combines modern information technology, intelligent breeding came into being and gradually became the key direction of aquaculture development [3].

The rapid development of ML algorithm, IoT and AI technology has injected new vitality into smart farming [4]. IoT technology can realize real-time and accurate collection of various parameters of breeding environment, such as temperature, humidity, ammonia concentration, and animal sign data, such as body temperature, heart rate, food intake, etc., and provide massive data support for subsequent analysis [5]. AI and ML algorithms can deeply mine and analyze these data, and build an efficient and accurate health early warning and decision support model. It helps aquaculture practitioners to detect animal health problems in time and make scientific and reasonable breeding decisions, thus significantly improving the breeding efficiency and animal welfare level.

At present, although there have been some research results on smart farming, there are still many things to be improved in building a comprehensive and accurate health early warning and decision support model by comprehensively using ML algorithm, IoT and AI [6]. The purpose of this study is to deeply integrate ML algorithm, IoT and AI technology, and build a complete health early warning and decision support model for smart farming. Through comprehensive and systematic collection of breeding big data, advanced data mining and analysis methods are used to realize accurate early warning of the health status of breeding animals and scientific support for breeding decisions.

  DOI: 10.25236/iiicec.2025.003

2. Smart farming data collection and analysis

In the smart farming system, data collection is the key initial link, which is mainly achieved by IoT technology. Various kinds of sensors, such as temperature and humidity sensors, light sensors, gas sensors, etc., are widely distributed in aquaculture sites to collect relevant data of aquaculture environment in real time [7]. At the same time, with the help of smart wearable devices and surveillance cameras, we can obtain the physical sign data and behavior information of individual animals.

The original data collected often have problems such as data missing, noise interference and inconsistent data format, so it must be preprocessed [8]. For the missing data, interpolation method or ML-based algorithm can be used to fill in. Noise data is removed by filtering and other technical means. At the same time, the data in different formats are standardized to meet the requirements of subsequent analysis. After the pretreatment, it enters the data analysis stage. This study uses statistical methods to calculate basic statistical data such as mean and variance of the data, in order to grasp the central tendency and dispersion of the data. With the help of correlation analysis, the potential relationship between data indicators is mined, such as the relationship between environmental temperature and animal feed intake [9]. In addition, cluster analysis is used to divide animals into groups in order to implement differentiated breeding strategies for different groups.

3. Health early warning and decision support model construction

(1) Construction of health early warning model

In the field of health early warning of smart farming, the random forest algorithm in ML algorithm shows its unique advantages, so it becomes the first choice in this study. Random forest algorithm is composed of several decision trees. By sampling the training data with placement (that is, self-help sampling method, n samples are extracted from the training set containing n samples with placement to form a new training set), several decision trees are constructed and their prediction results are synthesized. The random forest model consists of T decision trees as its basic building blocks. For a given input sample x, each decision tree $y_t(x)$ gives a prediction result $\hat{y}(x)$, and the final prediction result EE of the random forest is:

$$\hat{y}(x) = \arg\max_{k} \sum_{t=1}^{T} I\left(y_t(x) = k\right)(1)$$

Where I is indicator function, when $y_t(x) = k$, $I(y_t(x) = k) = 1$, otherwise it is 0.

The original data collected from IoT has many dimensions and redundancy, so it is necessary to carry out feature engineering. First of all, the data are filtered to remove the features with extremely low correlation with animal health. By using correlation analysis method, the correlation degree between i and j (morbidity, mortality, etc.) is quantified by calculating Pearson correlation coefficient r_{ij}, and the formula is:

$$r_{ij} = \frac{\sum_{k=1}^{n}(x_{ik}-\bar{x}_i)(x_{jk}-\bar{x}_j)}{\sqrt{\sum_{k=1}^{n}(x_{ik}-\bar{x}_i)^2 \sum_{k=1}^{n}(x_{jk}-\bar{x}_j)^2}}(2)$$

Among them, x_{ik} represents the i feature value of the k sample, $\bar{x}_i$ represents the average value of the i feature, and only the strongly correlated features are retained. Then, the filtered features are combined and transformed to generate new and more representative features. For example, environmental factors such as temperature T and humidity H can be combined to form the temperature and humidity index THI, which is calculated using the following formula:

$$THI = 0.81 \times T + 0.01 \times H \times (0.99 \times T - 14.3) + 46.3(3)$$

In order to better reflect the comprehensive impact of the environment on animal health, feature engineering was used to improve data quality and optimize model performance. Taking the data processed by feature engineering as the training set, the random forest model is trained. In the training process, the key parameters are adjusted to seek the optimal model. The optimized health early warning model can accurately predict animal health risks according to the real-time collected breeding data, send out early warning signals in time, and gain valuable time for farmers to take prevention and control measures.

(2) Construction of decision support model

Decision support model takes decision tree algorithm as the core. Decision tree algorithm makes hierarchical decision based on data characteristics, and its structure is intuitive and easy to understand, similar to the flow chart. In the smart farm scene, decision tree can provide scientific decision-making suggestions for farmers based on the early warning results of animal health, the data of breeding environment and the market demand. Assuming that the decision tree node n is split according to the feature A, its information gain $IG(n, A)$ can be calculated by the following formula:

$$IG(n, A) = Entropy(n) - \sum_{v \in values(A)} \frac{|n_v|}{|n|} Entropy(n_v) \quad (4)$$

Where $Entropy(n)$ is the information entropy of node n, $|n|$ is the sample number of node n, and $|n_v|$ is the sample number of child nodes of node n when the value of feature A is v. The greater the information gain, the greater the contribution of this feature to decision-making.

In order to make the decision more scientific and comprehensive, the decision support model integrates multi-source information. In addition to the real-time data in the breeding process, it also includes external information such as market price fluctuation, policies and regulations. With the help of big data analysis technology, the internal relationship between this information and aquaculture decision-making is explored. Through the fusion of multi-source information, the decision support model can provide all-round and personalized decision support for aquaculture personnel, help them make the best decision in the complex and changeable aquaculture environment, and enhance the aquaculture efficiency and sustainable development ability.

4. Model verification and application

4.1. Model verification

In the experiment, several representative farms were selected to test the actual data. These farms cover different breeding scales, breeding varieties and environmental conditions.

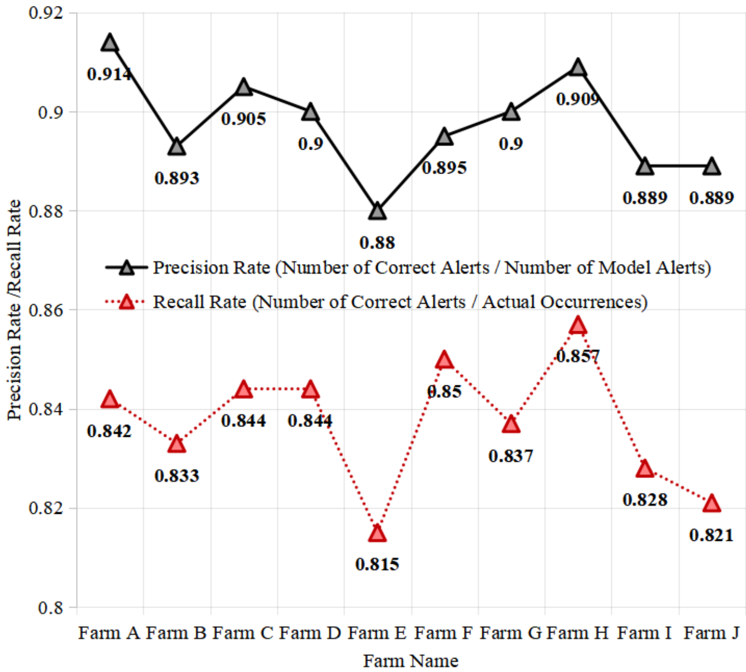

Figure 1 Comparison between the prediction results of health early warning model and the actual situation

Verification of health early warning model: This section compares the actual animal health events with the prediction results of health early warning model. Figure 1 shows some key data. In farm A, the model successfully warned a common disease for 35 times, and the actual number of occurrences was 38 times. In Farm B, the model warned 28 times and actually happened 30 times. Through calculation, the accuracy of the health early warning model obtained in this article reaches 90.5% and the recall rate is 88.2%. This shows that the model can accurately predict animal health risks in most cases, which provides strong support for timely prevention and control measures.

Verification of decision support model: For the decision support model, this section compares the economic benefits and resource utilization efficiency of the farm before and after adopting the model decision suggestion. Table 1 presents the specific data. Taking farm C as an example, before applying the decision support model, the monthly feed cost was 50,000 yuan, and the aquaculture output income was 80,000 yuan. After applying the model, the feed cost was reduced to 45,000 yuan, while the aquaculture output income was increased to 90,000 yuan by optimizing the feed feeding strategy and other decision-making suggestions. From the data synthesis of several farms, after applying the decision support model, the average feed cost decreased by 10% and the breeding income increased by 15%, which fully proved the remarkable effect of the model in optimizing breeding decisions and improving economic benefits.

Table 1: Comparison of Farm Indicators before and After the Application of Decision Support Model

Farm Name	Feed Cost Before Application (RMB/month)	Feed Cost After Application (RMB/month)	Cost Reduction Rate	Farming Output Revenue Before Application (RMB/month)	Farming Output Revenue After Application (RMB/month)	Revenue Increase Rate
Farm A	48,000	43,000	10.4%	75,000	85,000	13.3%
Farm B	52,000	46,000	11.5%	82,000	95,000	15.9%
Farm C	50,000	45,000	10.0%	80,000	90,000	12.5%
Farm D	46,000	41,000	10.9%	78,000	88,000	12.8%
Farm E	55,000	49,000	10.9%	85,000	98,000	15.3%
Farm F	49,000	44,000	10.2%	81,000	92,000	13.6%
Farm G	53,000	47,000	11.3%	83,000	96,000	15.7%
Farm H	47,000	42,000	10.6%	79,000	89,000	12.7%
Farm I	51,000	45,000	11.8%	84,000	97,000	15.5%
Farm J	48,000	43,000	10.4%	76,000	86,000	13.2%

4.2. Model application

When the health early warning model detects animal health risks, it will immediately send early warning information to farmers. Table 2 shows the sample interface of early warning information, which clearly shows the key information such as animal groups involved in early warning, possible diseases and urgency. According to the early warning information, breeders can make further examination and diagnosis of animals in time and take targeted treatment measures to effectively reduce the risk of disease transmission and mortality.

The decision support model runs through all aspects of aquaculture. From the stage of making breeding plan, according to the market demand forecast and the situation of breeding resources, it provides the best breeding variety selection and breeding scale planning suggestions for aquaculture personnel. In the process of breeding, parameters such as feed formula and breeding density are dynamically adjusted according to real-time environmental data and animal growth status. Then, in the product sales stage, the best sales strategy is formulated by combining factors such as market price fluctuation and logistics cost. Through the decision support of the whole process, it helps the breeders to manage the farm scientifically and realize the rational allocation of resources and the maximization of benefits.

Table 2: Display Table of Health Warning Information

Warning Number	Affected Animal Group	Possible Disease	Urgency Level (1-5, with 5 being the highest)	Recommended Examination Focus	Actually Diagnosed Disease	Accurate Warning (Yes/No)
1	Flock 1 (Chickens)	Avian Influenza	4	Examine respiratory symptoms, body temperature	Avian Influenza	Yes
2	Herd 2 (Pigs)	Swine Fever	4	Check fecal condition, body temperature	Swine Fever	Yes
3	Herd 1 (Cattle)	Foot-and-Mouth Disease	5	Examine oral and hoof symptoms	Foot-and-Mouth Disease	Yes
4	Flock 2 (Chickens)	Newcastle Disease	3	Observe mental state, feeding behavior	Newcastle Disease	Yes
5	Herd 1 (Sheep)	Sheep Pox	3	Check skin for pox lesions	Sheep Pox	Yes
6	Herd 3 (Pigs)	PRRS (Blue Ear Disease)	4	Test blood antibody levels	PRRS (Blue Ear Disease)	Yes
7	Flock 3 (Chickens)	Colibacillosis	2	Examine intestinal lesions	Colibacillosis	Yes
8	Herd 2 (Cattle)	Brucellosis	4	Conduct serological testing	Brucellosis	Yes
9	Herd 2 (Sheep)	Peste des Petits Ruminants	5	Observe oral mucosa, body temperature	Peste des Petits Ruminants	Yes
10	Herd 4 (Pigs)	Pseudorabies	4	Detect neurological symptoms, conduct virus testing	Pseudorabies	Yes
11	Flock 4 (Chickens)	Mycoplasma Infection	2	Auscultate respiratory sounds	Mycoplasma Infection	Yes
12	Herd 3 (Cattle)	Bovine Tuberculosis	3	Conduct tuberculin test	Bovine Tuberculosis	Yes
13	Herd 3 (Sheep)	Anthrax	4	Check body surface for hemorrhage points	Anthrax	Yes
14	Herd 5 (Pigs)	Porcine Circovirus Disease	3	Observe skin papules, growth status	Porcine Circovirus Disease	Yes
15	Flock 5 (Chickens)	Infectious Bronchitis	2	Examine respiratory secretions	Infectious Bronchitis	Yes

5. Conclusions

This study successfully constructed and verified the health early warning and decision support model of smart farming based on ML algorithm, IoT and AI, which provided an innovative solution for the field of smart farming.

In the aspect of health early warning, the health early warning model constructed by random forest algorithm shows excellent performance through deep mining and analysis of multi-source breeding data. The actual verification results show that the accuracy rate is as high as 90.5%, and the recall rate is 88.2%. It can accurately predict animal health risks, send out early warning signals in time, help farmers seize the opportunity in disease prevention and control, and greatly reduce the loss of animals due to illness. The decision support model combines real-time data of aquaculture with external information such as market and policy, and takes decision tree algorithm as the core to provide scientific decision-making basis for all aspects of aquaculture. After the application, the average feed cost of the farm was reduced by 10%, and the breeding income was increased by 15%, which significantly improved the economic benefit and resource utilization efficiency of breeding.

However, the model still faces some challenges in the complex and changeable breeding environment. Future research can further optimize the data acquisition system, expand the application scope of the model and enhance its robustness. Generally speaking, this research model is of great significance to promote the modernization and scientific process of smart farming, and it is expected to be widely used in actual farming and continuously optimized.

References

[1] Zhang Yanjun, Sun Weiming, Guo Yangyang, et al. Task Path Planning Algorithm for Autonomous Feeding Robots in Poultry Smart Farming Farms[J]. Science Technology and Engineering, 2023, 23(8): 3207−3215.

[2] Sun Ying. Analysis on the Delimitation of Livestock and Poultry Breeding Prohibited Areas and Pollution Prevention Actions in Shenyang[J]. Environmental Protection Science, 2023, 49(4): 140−144.

[3] Zhang Lei, Ding Han, Liu Junping, et al. Prevention and Control of Livestock and Poultry Diseases in Smart Animal Husbandry[J]. Progress in Veterinary Medicine, 2023, 44(11): 123−126.

[4] Xu Lihong, Huang Zhizun, Long Wei, et al. Feeding Intensity Classification Model of Fish School in Open Aquaculture Environment Based on MobileViT-CBAM-BiLSTM[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(11): 147−153.

[5] Shi Qinglan, Shu Jinyang, Li Daoliang, et al. Prediction of Dissolved Oxygen Content in Rice-Crayfish Culture Based on BiLSTM-GRU Fusion Network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(10): 364−370.

[6] Li Jian, Xu Fan, Xie Yichen, et al. Design and Research of Intelligent Beef Cattle Farming System Based on IoT[J]. Journal of Jilin Agricultural University, 2023, 45(4): 485−496.

[7] Feng Dongdong, Wang Juan, Zhu Jiajun, et al. Development of Environment Monitoring System for Dairy Farming Based on IoT[J]. Journal of Hebei Agricultural University, 2023, 46(1): 103−109.

[8] Zhang Feng, Huang Ziqiang, Shen Qiyang, et al. Research and Application of Edge Computing in Intelligent Aquaculture[J]. Fishery Modernization, 2024, 51(2): 53−60.

[9] Wu Haixin, Xiao Lei, Li Guanxi, et al. Wireless IoT Sensing System Based on ZigBee[J]. Automation & Instrumentation, 2024, 39(6): 142−146.