Scholar Publishing Group
Machine Learning Theory and Practice
https://doi.org/10.38007/ML.2021.020305
ISSN 2790-0983 Vol. 2, Issue 3: 36-43

Prediction Model of Grain Mold Probability Based on Naive Bayes Algorithm

Rajit Mansour*

Case Western Reserve University, USA

**corresponding author*

Keywords: Naive Bayes, Grain Mold, Probabilistic Prediction Model, Prediction Error

Abstract: Grain is by far the largest amount of food stored in the world. The early prediction of grain mold (GM) is mainly to be able to detect it in time before its significant growth activity to avoid causing loss of stored grain. Microorganisms are tiny and not visible to the naked eye in the early stage, if microbial colonies can be observed in the grain pile, these parts of the grain often already have serious deterioration, mold metabolism is more hazardous to produce mycotoxins, which has a more adverse impact on the safe storage of grain, in fact, as long as there is the growth of microorganisms such as mold in the grain, even if the colonies of mold are not visible to the naked eye, the grain quality will have significant changes. In order to ensure the quality of grain output, the best way is to prevent GM. In this regard, this paper suggests a model for predicting the probability of GM based on the Naive Bayes(NB) algorithm, cultivating GM samples, and comparing the prediction errors of these three models with the PSO-LSSVM model and ARIMA model to predict the probability of mold in samples, we know that the prediction error(PE) of the NB model is the smallest, which means that the difference between the predicted value of GM and the real value of the model is small, and the accuracy of using this model for prediction is more reliable. The accuracy of the model is more reliable.

1. Introduction

In recent years, the rapid development of agricultural science, the cultivation of a large number of excellent varieties, the introduction of scientific planting mode, so that the unit area of grain production continues to improve. However, the arable land used for grain cultivation in China is decreasing, and it is difficult to fundamentally solve the increasing contradiction between grain supply and demand just by improving yields, and the imbalance between grain supply and demand will be further aggravated if grain is moldy, so it is necessary to further improve grain quality, reduce grain storage losses and other aspects to achieve grain safety [1-2].

Research on GM detection and prediction has yielded good results. In terms of GM detection, some scholars have applied electronic nose technology self-developed electronic nose to conduct experiments to track the process of GM in storage and monitor GM, and found that it can provide a more accurate response to the occurrence of GM in storage and has high sensitivity in monitoring, and some scholars have combined chemometric methods to process and analyze moldy rice images and constructed a model that can accurately distinguish between rice mold Some scholars have combined chemometric methods to process and analyze moldy rice images and constructed models that can accurately distinguish the types of rice mold [3-4]. In terms of GM probability prediction, many large grain stations judge and control the mold condition by real-time monitoring of grain temperature and humidity as well as example sampling, which cannot accurately reflect the mold activity in a timely manner due to the lag in temperature changes, making the prediction results differ significantly from the true values and leading to serious grain damage [5]. Some grain storage institutions also rely on microbial colony count detection method, but its operation steps are complicated and tedious, time-consuming, poorly sensitive, and require the expertise of the identifier, which makes it difficult to meet the demand for real-time prediction of rice mold condition [6]. Therefore, to ensure grain quality, it is necessary to improve the performance of GM detection techniques and prediction models for mold detection.

In this paper, we first analyze the hazards of GM and introduce several mold detection techniques, then study the detection device of GM, then elaborate the concept of NB and use it to establish the corresponding mold probability prediction model, and finally analyze the mold degree of moldy grain with the increase of storage days, and verify the prediction error of the NB algorithm in this paper by experimentally comparing the mold prediction error of three prediction models The validity of the model is verified by comparing the PE of the three models.

2. Basic Overview

2.1. The Danger of GM and Mildew

Grain grown in the field and exposed to the air will always carry all types of molds present in the soil or air. The rich carbon source, nitrogen source and inorganic salt components in grain can be used in the growth and reproduction of mold, which is a natural medium for mold. When the grain suffers from adverse weather and storage conditions that cause elevated moisture in the grain, the life activity of the mold in the grain will be activated and the grain will begin to mold [7].

In the process of GM, molds secrete various extracellular enzymes to decompose the starch, lipids, cellulose and other substances in the grain, destroying the shell of the grain and consuming the dry matter in the grain; molds in the decomposition of amino acid products, such as thiols, causing the grain to brown and lose its original luster, while the molds multiply to form colored colony forms, accelerating the discoloration of the grain and reducing the market value of the grain. Some molds also produce toxic substances that endanger the liver, kidneys, digestive tract, and reproductive capacity of humans and animals, causing serious safety hazards [8-9].

2.2. GM Detection Techniques

As shown in Figure 1 are four kinds of GM detection techniques, which are described as follows.

(1) Plate counting method

GM is a manifestation of high temperature and high humidity conditions that activate the life activities of mold in grain. The rapid multiplication of mold metabolism consumes nutrients in the grain, and the grain deteriorates. The most obvious feature of GM is the increase in the number of mold. The most intuitive way to understand the mold condition of grain is to determine the number of molds in the grain [10]. Molds are tiny and difficult to count directly with the naked eye, and are generally counted and morphologically analyzed by means of dilution culture as well as microscopic counting. The plate counting method is a detection method based on the total number of mycelia and spores in grain for counting, and this method method has been used for a long time for the study of microbial changes in grain, with good repeatability and feasibility, and is a classical microbial counting method, and the detection results have good foresight for the development of molds in grain [11].

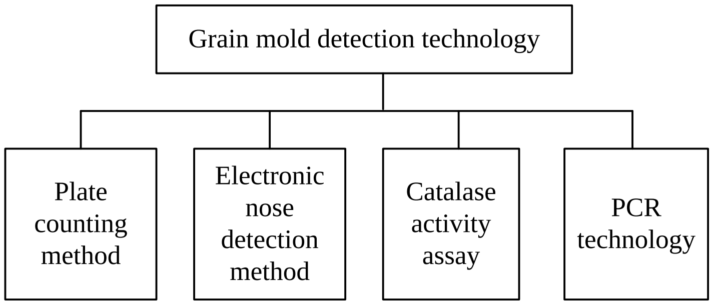

Figure 1. Detection techniques

(2) Electronic nose detection method

The process of grain spoilage and deterioration produces unpleasant and abnormal odors. The human olfactory system can simply and quickly perceive such changes, but the sensitivity of smell is often disturbed by subjective factors (physiological, psychological and environmental, etc.), which affects the accuracy of olfactory detection. Therefore, there is a need for an objective odor detection method. In the last decade or so, efforts have been devoted to the development of bio-olfactory bionic technology, i.e., electronic nose detection [12]. The working principle of the electronic nose is similar to that of the human nose. The volatile substances in the gas enter the sensor array that simulates the olfactory cells in the nose, which will interact with the sensor array and generate electrical signals; the electrical signals are converted into numerical values through an electronic interface into signal processing; and then the recorded data are analyzed for recognition according to a data statistical system that simulates brain functions [13].

(3) Catalase activity assay

Catalase, also known as catalase, is widely present in aerobic organisms. Hydrogen peroxide in organisms is catalyzed by catalase reaction to form oxygen and water, which protects the organism from damage caused by hydrogen peroxide. Molds in grain are also aerobic fungi, and previous studies have found that the peroxidase activity of molds and the number of molds are closely related, and the detection of peroxidase activity of molds can be used to reflect the degree of mold in grain [14].

(4) PCR technology

In order to develop microbial detection technology, improve the early detection of mold using morphological characteristics of the analysis of time-consuming and labor-intensive situation. the emergence of DNA technology for the detection of mold provides a new direction of detection. Mold DNA molecules are the identity characteristics of mold populations, with their own unique gene fragments, through the amplification of the corresponding DNA fragments can accurately identify the species of mold [15].

2.3. GM Detection Device

In the process of moldy grain particle detection, the detection results are easily affected by the light conditions, and when there is insufficient light, the shadows between the grain particles will affect the detection accuracy. Especially when the grain particles are stacked on top of each other, the detection accuracy is even lower, while the manual spreading of particles is very inefficient.

In order to quickly and accurately detect moldy corn particles and improve detection accuracy, a closed detection device was studied. The device is made of stainless steel material, in order to prevent the adverse effects of the bottom color on the detection results, the vibratable tray painted in cream color. The light during detection is provided by a circular arrangement of LED natural light beads and UV light beads, which solves the problem of uneven light and insufficient light intensity. Grain particles are placed on the vibratable tray during detection, and the tray is connected to the base of the device with a spring. Two polarizing motors are symmetrically installed under the tray, and when the polarizing motors vibrate, the tray is driven to vibrate so that the grain particles lay flat, solving the problem of grain particles stacking on each other. In order to prevent the grain particles from vibrating out at the edge of the tray, a 5 mm high edge is set around each side of the tray to make the grain particles vibrate and lay flat inside the tray [16-17].

3. GM Probability Prediction Model Construction

3.1. NB Algorithm

The NB algorithm is one of the most widely used classification algorithms, and its classification accuracy and execution efficiency are comparable to those of Decision Tree, K-nearest neighbor (KNN), and ANN, etc. The NB classification model has wide applicability, fast classification speed, and high algorithm stability compared with other classification models, and is often used in real-world environments such as Spam interception, Pattern recognition, and intrusion detection [18]. Pattern recognition and intrusion detection are often used in real-world environments [18]. However, the NB algorithm is limited by the "conditional independence assumption", which ignores the dependencies between attributes, resulting in the failure to reflect the relational nature of the attributes, which undoubtedly contradicts the theorem of universal connection between things in realistic environments and makes it perform poorly in many situations where the conditions cannot be satisfied [19].

3.2. Prediction Model of GM Probability Based on the NB Algorithm

The basic theory of NB is similar to the covariance determinant, but for NB, it has to be based on a basic assumption that each attribute variable has an independent effect on the outcome. The sample has the attributes F = (f1, f2, ... , fn), and we try to find a category C such that the maximum possible value is taken. Since the current goal is to try to have moldy ground grain and mold-free ground grain, the binary class C $\in$ {0, 1} can be introduced. Where 1 means that the grain sample is predicted to be moldy and 0 means that the grain sample is predicted to be mold-free. For such a dichotomous classification problem, the classification result of mold can be determined by comparing two probability values.

$$\frac{P(A=1|F=f1,f2,...,fn)}{P(A=0|F=f1,f2,...,fn)}=\frac{P(A=1)\mathrm{X}_{i=1}^{n}Pi(fi|A=1)}{P(A=0)\mathrm{X}_{i=1}^{n}Pi(fi|A=0)} \tag{1}$$

Taking the logarithm of equation (1), we can obtain.

$$\log\frac{P(A=1|F=f1,f2,...,fn)}{P(A=0|F=f1,f2,...,fn)}=\log\frac{P(A=1)}{P(A=0)}+\sum_{i=1}^{n}\log\frac{Pi(fi|A=1)}{Pi(fi|A=1)} \tag{2}$$

Therefore the sample will be predicted to be 1 (with mold) under the following conditions.

$$\log \frac{P(A=1 \mid F=f1, f2, \ldots, fn)}{P(A=0 \mid F=f1, f2, \ldots, fn)} \geq \theta \qquad (3)$$

The opposite is predicted to be 0 (no mold). Where is the threshold used to compromise sensitivity, which can be obtained by training on the training set to maximize the classifier performance.

4. Analysis of Prediction Results

4.1. GM Degree Detection

The grain was incubated with 16% moisture and mold at 30℃, and the grain test samples with different mold degrees were obtained according to the difference of incubation time. The samples were subjected to NB mold prediction, mold prediction based on ARIMA model and mold prediction based on PSO-LSSVM model, and the GM was detected by using peroxidase activity enzyme assay and moldy grain assay. Shown in Table 1.

Table 1.GM degree detection results

Storage time / d	0	3	6	9	12	15
Catalase activity (U/100g)	62	284	943	1276	1384	1450
Number of moulds/(10^3cfu g^{-1})	0	17	32	235	316	394
Moulded grain (%)	0	0	0.47	1.21	1.83	2.26

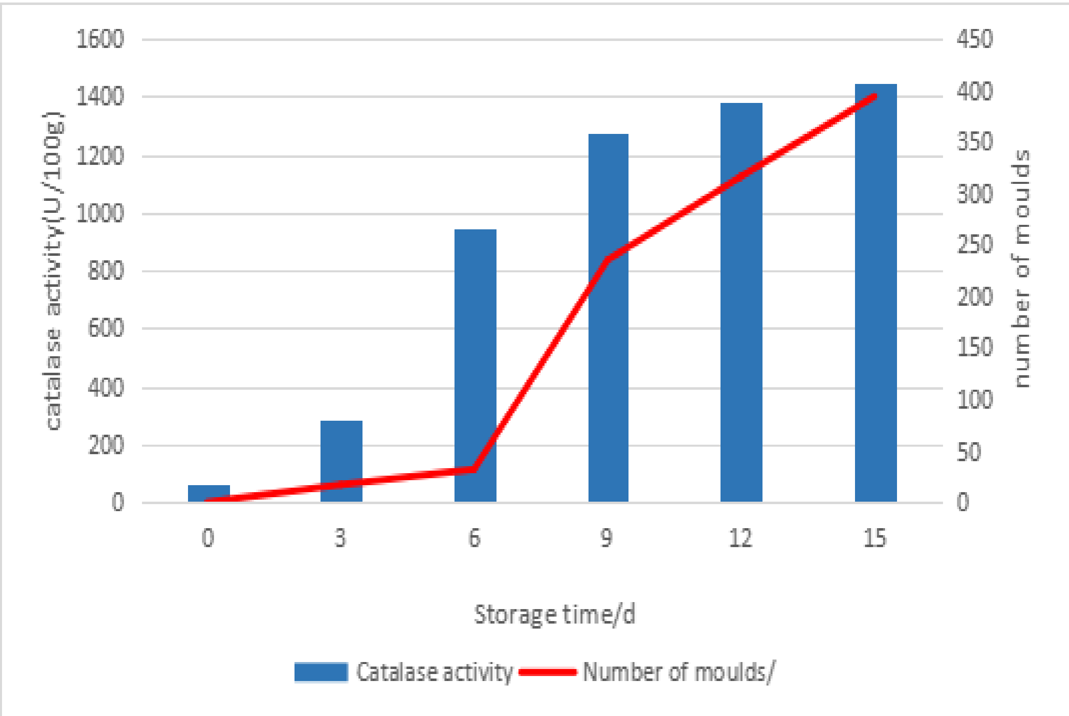

Figure 2. GM rapid detection technology to detect the degree of GM

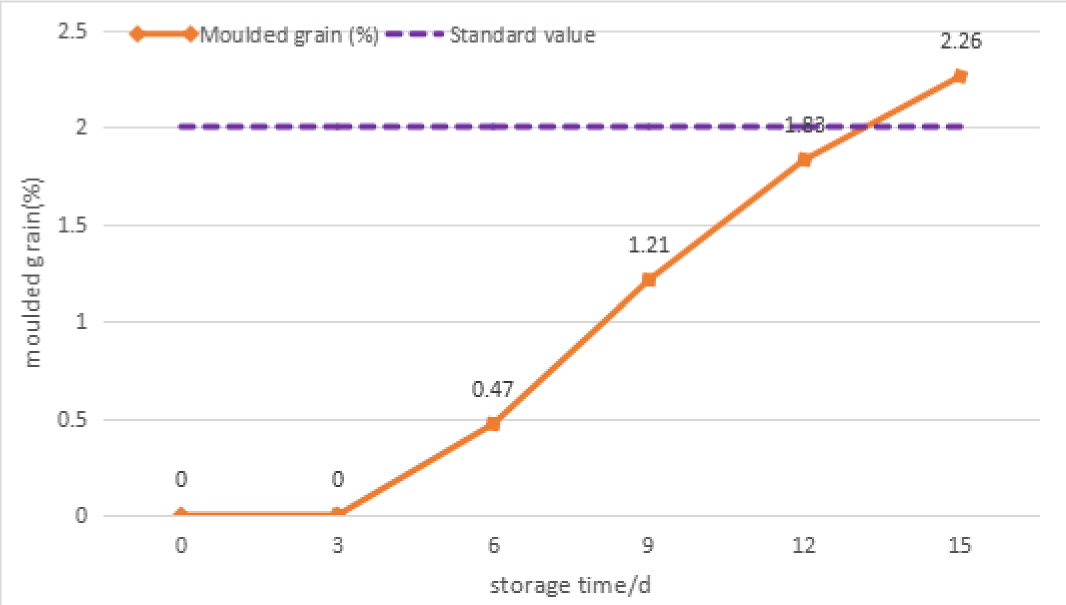

Figure 3. Degree of GM detected by the moldy grain assay

From the curves in Figures 2 and 3, it can be found that with the extension of the mold incubation time (i.e., storage time), the amount of mold, mold peroxidase activity value and mold grain content in the grain increased simultaneously, but the time points for the increase of these detection indicators were different. The peroxidase activity value of the grain increased several times in the 3rd day of storage, the amount of mold in the grain increased significantly in the 6th day of mold storage, and the moldy grain test showed a significant change in the 9th day and exceeded the standard value of moldy grain (2%) in the 15th day. Compared with the amount of carriage and moldy grain test, the rapid detection technology of GM enzyme activity can reflect the abnormal changes of grain earlier and the detection sensitivity is relatively high, respectively. The earlier the moldy state of the grain is detected, the more sufficient time is left to deal with the problem grain, and the subsequent mold treatment measures start earlier, resulting in less economic and quality loss of the grain. In terms of detection sensitivity, enzyme activity rapid detection technology is suitable for controlling the quality of incoming grain.

4.2. Comparison of GM Probability Prediction Models

In order to compare the predicted results with the NB (NB) prediction model, the PSO-LSSVM model and ARIMA model are also used for prediction in this paper. According to the above no change detection experiments can be seen that the GM culture at 30 ℃, the grain cultured out of the class into four samples, Table 1 lists the relative error values of the predicted and true values of different models for the four groups of grain samples.

Table 2. Relative error of prediction of different models of GM probability

	NB	PSO-LSSVM	ARIMA
Group 1	0.72	1.25	2.88
Group 2	0.46	0.93	1.75
Group 3	0.69	1.37	2.43
Group 4	0.85	1.41	3.67

According to Table 2, the PE of the NB model is less than 1 for all four groups of grain samples mold probability, and the ARIMA model is the largest prediction error among the three models.

This verifies that the NB model has the highest prediction accuracy and is more suitable for GM probability prediction, so that grain storage can be done in advance by analyzing the GM concept rate.

5. Conclusion

GM will bring huge losses to the agricultural economy, and major grain storage institutions should pay attention to grain quality inspection, timely detection of moldy grain, and good measures to prevent GM. This paper establishes a GM probability prediction model to predict the mold rate, and analyzes the prediction effect of the NB model in this regard, and compares it with the PE of other two models, and gets that the NB algorithm proposed has better prediction accuracy, so we hope to apply this model to GM prediction, which can help reduce grain loss and improve grain taste and quality.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Kristin McCullough, Tatiana Dmitrieva, Nader Ebrahimi: New approximate Bayesian computation algorithm for censored data. Comput. Stat. 37(3): 1369-1397 (2021). https://doi.org/10.1007/s00180-021-01167-3

[2] Ciniro Aparecido Leite Nametala, Wandry R. Faria, Benvindo R. Pereira Junior: On the performance of the Bayesian optimization algorithm with combined scenarios of search algorithms and scoring metrics. Genet. Program. Evolvable Mach. 23(2): 193-223 (2021).

[3] Vicente P. Soloviev, Pedro Larranaga, Concha Bielza: Estimation of distribution algorithms using Gaussian Bayesian networks to solve industrial optimization problems constrained by environment variables. J. Comb. Optim.44(2): 1077-1098 (2021).

[4] Jakeoung Koo, Abderrahim Halimi, Stephen McL aughlin: A Bayesian Based Deep Unrolling Algorithm for Single-Photon Lidar Systems. IEEEJ. Sel. Top. Signal Process. 16(4): 762-774 (2021).

[5] Vahid Rafe, Somayeh Mohammady, Erik Cuevas: Using Bayesian optimization algorithm for model-based integration testing. Soft Comput.26(7): 3503-3525 (2021). https://doi.org/10.1007/s00500-021-06476-9

[6] Raymond Leung, Alexander Lowe, Anna Chlingaryan, Arman Melkumyan, John N. Zigman: Bayesian Surface Warping Approach for Rectifying Geological Boundaries Using Displacement Likelihood and Evidence from Geochemical Assays. ACM Trans. Spatial Algorithms Syst. 8(1): 1-23 (2021). https://doi.org/10.1145/3476979

[7] Thomas Callens, Tuur van der Have, Sam Van Rossom, Joris De Schutter, Erwin Aertbelien: A Framework for Recognition and Prediction of Human Motions in Human-Robot Collaboration Using Probabilistic Motion Models. IEEE Robotics Autom. Lett. 5(4): 5151-5158 (2020). https://doi.org/10.1109/LRA.2020.3005892

[8] Alberto Castellini, Federico Bianchi, Alessandro Farinelli: Generation and interpretation of parsimonious predictive models for load forecasting in smart heating networks. Appl. Intell. 52(9): 9621-9637 (2021). https://doi.org/10.1007/s10489-021-02949-4

[9] Oscar Serradilla, Ekhi Zugasti, Jon Rodriguez, Urko Zurutuza: Deep learning models for predictive maintenance: a survey, comparison, challenges and prospects. Appl. Intell. 52(10): 10934-10964 (2021). https://doi.org/10.1007/s10489-021-03004-y

[10] Anne-Christin Hauschild, Marta Lemanczyk, Julian. Matschinske, Tobias Frisch, Olga 1. Zolotareva, Andreas Holzinger, Jan Baumbach, Dominik Heider: Federated Random Forests can improve local performance of predictive models for various healthcare applications. Bioinform. 38(8): 2278-2286 (2021). https://doi.org/10.1093/bioinformatics/btac065

[11] Faygal Hamdi, Abderaouf Khalfi: Predictive density criterion for SETAR models. Commun. Stat. Simul. Comput. 51(2): 443-459 (2021). https://doi.org/10.1080/03610918.2019.1653915

[12] Eleonora Proia, Alessio Ragno, Lorenzo Antonini, Manuela Sabatino, Milan Mladenovic, Roberto Capobianco, Rino Ragno: Ligand-based and structure-based studies to develop predictive models for SARS-CoV-2 main protease inhibitors through the 3d-qsar.com portal. J. Comput. Aided Mol.Des.36(7): 483-505 (2021)

[13] Cinzia Giannetti, Aniekan Essien: Towards scalable and reusable predictive models for cyber twins in manufacturing systems. J. Intell. Manuf. 33(2): 441-455 (2021). https://doi.org/10.1007/s10845-021-01804-0

[14] Williams Rizzi, Chiara Di Francescomarino, Chiara Ghidini, Fabrizio Maria Maggi: How do I update my model? On the resilience of Predictive Process Monitoring models to change. Knowl. Inf. Syst. 64(5): 1385-1416 (2021).

[15] Maxime De Bois, Mounim A. EI-Yacoubi, Mehdi Ammi: GLYFE: review and benchmark of personalized glucose predictive models in type 1 diabetes. Medical Biol. Eng. Comput.60(1): 1-17 (2021). https://doi.org/10.1007/s11517-021-02437-4

[16] Ebru Akis, Gural Guven, Bahram Lotfi Sadigh: Predictive models for mechanical properties of expanded polystyrene (EPS) geofoam using regression analysis and artificial neural networks. Neural Comput. Appl. 34(13):10845-10884 (2021).

[17] Krisztian Mark Balla, Christian Schou, Jan Dimon Bendtsen, Carlos Ocampo-Martinez, Carsten Skovmose Kallesoe: A Nonlinear Predictive Control Approach for Urban Drainage Networks Using Data-Driven Models and Moving Horizon Estimation. IEEE Trans. Control. Syst. Technol. 30(5):21 47-2162 (2021). https://doi.org/10.1109/TCST.2021.3137712

[18] Daehee Kwon, Eunseop Yeom: Shape evaluation of highly overlapped powder grains using U-Net-based deep learning segmentation network. J. Vis. 24(5): 931-942 (2021). https://doi.org/10.1007/s12650-021-00748-0

[19] Petra Perner: Detection of Hygiene-Relevant Parameters from Cereal Grains such as Mycotoxinbased on Microscopic lmaging, Intelligent lmage Processing and Data Mining. Trans. Mass Data Anal. Images Signals 12(1): 13-21 (2021).