Article

A Back Propagation Neural Network Model for Postharvest Blueberry Shelf-Life Prediction Based on Feature Selection and Dung Beetle Optimizer

Runze Zhang [1], Yujie Zhu [1,*], Zhongshen Liu [2,*], Guohong Feng [1], Pengfei Diao [1], Hongen Wang [1], Shenghong Fu [1], Shuo Lv [1] and Chen Zhang [1]

1   College of Engineering and Technology, Northeast Forestry University, Harbin 150040, China;
    cherish9z@nefu.edu.cn (R.Z.); fgh_1980@126.com (G.F.); 17621166595@163.com (H.W.)
2   College of Biopharmaceuticals, Heilongjiang Province Agricultural Engineering Vocational College,
    Harbin 150088, China
*   Correspondence: zhuyujie004@126.com (Y.Z.); swkjlzs@163.com (Z.L.)

Abstract: (1) Background: Traditional kinetic-based shelf-life prediction models have low fitting accuracy and inaccurate prediction results for blueberries. Therefore, this study aimed to develop a blueberry shelf-life prediction method based on a back propagation neural network (BPNN) optimized by the dung beetle optimizer using an elite pool strategy and a Gaussian distribution estimation strategy (GDEDBO); (2) Methods: The "Liberty" blueberry cultivar was used as the research object, and 23 quality indicators, including color parameters, weight loss rate, decay rate, and texture parameters, were measured under storage temperatures of 0, 4, and 25 °C. Based on the maximum relevance minimum redundancy (MRMR) algorithm, seven key influencing factors of shelf life were selected as the input parameters of the model, and then the MRMR-GDEDBO-BPNN prediction model was established; (3) Results: the results showed that the model outperformed the baseline model at all three temperatures, with strong generalization ability, high prediction accuracy, and reliability; and (4) Conclusions: this study provided a theoretical basis for the shelf-life determination of blueberries under different storage temperatures and offered technical support for the prediction of remaining shelf life.

Keywords: blueberry; shelf-life forecast; maximum relevance minimum redundancy (MRMR); gaussian distribution estimation (GDE) strategy

1. Introduction

Blueberries are a nutritious, antioxidant-rich berry fruit with a variety of health benefits that is popular among consumers. "Liberty" is a high-quality blueberry variety developed by Michigan State University in 2003 as a hybrid of "Brigita" and "Eliot" parents. It has firm, flavorful, and storage-resistant fruits that are suitable for fresh consumption or processing. It has high-yielding and disease-resistant plants that are considered to be one of the best blueberry varieties available. Due to its excellent quality and remarkable benefits, it has high competitiveness and economic value in domestic and international markets.

However, due to their delicate skin, blueberries are susceptible to mechanical damage or microbial infestation during transportation and storage, leading to deterioration of their quality and affecting their nutritional value and market value [1,2]. Therefore, it is important to establish an effective shelf-life prediction model for blueberries to optimize the logistics and processing process of blueberries, reduce losses and wastage, ensure their freshness and nutritional value, and enhance their market competitiveness and economic benefits [3,4].

Current methods for food shelf-life prediction fall into two main categories: one is based on kinetic models, which mainly use physicochemical indicators, temperature, and

other factors to establish mathematical equations to describe the laws of food quality change, such as the Arrhenius equation, which is commonly used for fruit shelf life prediction and can describe the relationship between chemical reaction rate constants and temperature well. For example, Wang, H. et al. [5] used the Arrhenius equation to predict the shelf-life of frozen tilapia (*Oreochromis niloticus*) with a relative error of ±10% or less. However, the equation requires measuring complex physicochemical indicators and is sensitive to temperature. Different temperature levels need to be set in the modeling process, which makes the experimental operation complicated [6]. Another type of approach is based on machine learning models, which mainly use data mining and artificial intelligence techniques to extract features and patterns from multidimensional data and build non-linear predictive models, such as multiple linear regression (MLR), support vector machines (SVM), artificial neural networks (ANN), etc. Fan et al. [7] used MLR to predict the effects of ultra-low temperature storage on the shelf life and quality indicators of white fish, and the results showed that ultra-low temperature storage could effectively extend the shelf life of white fish. Although linear models eliminate the dependence of traditional models on temperature, they cannot reflect the non-linear relationships between data. Algorithms such as SVM and ANN can improve prediction accuracy by processing more complex and non-linear data relationships. Back Propagation Neural Network (BPNN) is a special case of ANN, which is now widely used in the shelf-life prediction of agricultural products. Huang et al. [8] used BPNN to establish a shelf-life prediction model of blueberries from the gas perspective, and the prediction results basically met the shelf-life requirements.

However, BPNN makes it difficult to determine the network structure and parameters and has a slow training speed. In order to improve prediction accuracy, some scholars choose to use meta-heuristic methods to optimize the weights and thresholds of BPNN, such as genetic algorithms, ant colony algorithms, etc. [9]. Although the combination of meta-heuristic algorithms and BPNN can effectively improve prediction accuracy, it may still fall below the local optimum, leading to a degradation of model performance. To address this issue, this paper proposes the dung beetle optimizer using the Gaussian distribution estimation strategy (GDEDBO) to optimize the BPNN [10]. The model is improved in three aspects: first, by initializing the dung beetle population with a tent chaos mapping to increase diversity; second, by enhancing the diversity of leaders by introducing an elite pool strategy to improve the exploration performance of the algorithm; and third, by using the Gaussian distribution estimation (GDE) strategy to use dominant population information to adjust the search direction to guide the population to better evolution.

Feature selection is a critical factor for shelf-life prediction accuracy, besides model selection. It can eliminate irrelevant or redundant features, reduce feature dimensionality, improve model accuracy, and shorten runtime [11]. Li et al. [12] screened out highly correlated freshness indicators as input variables for grape shelf-life prediction using Pearson correlation analysis, but the variables were highly intercorrelated, causing feature redundancy and lowering model prediction accuracy. The max-relevance and min-redundancy (MRMR) algorithm maximizes the correlation between the independent variables and the dependent variable while minimizing the correlation among the independent variables [13].

In summary, this paper studied the "Liberty" blueberry variety and measured 23 indicators related to shelf-life under three different storage temperatures (0, 4, and 25 °C), such as color parameters, weight loss rate, decay rate, texture parameters, etc. Then, it used MRMR to select seven optimal input parameters and improved BPNN with GDEDBO to propose the MRMR-GDEDBO-BPNN shelf-life prediction model, which was compared with other benchmark models. This study aimed to provide references for blueberry shelf-life management under different storage temperatures and technical support for a fast and accurate prediction of remaining shelf-life.

2. Materials and Methods

2.1. Materials and Experimental Program

Fresh "Liberty" blueberries were harvested from a blueberry plantation in Dandong City, Liaoning Province, and immediately wrapped in ice packs and transported to the laboratory by air in an insulated box, with the storage conditions of the box set at a refrigerated temperature of 0–4 °C and a relative humidity of 90~95%. After being transported to the laboratory, the blueberries were pre-cooled in the freezer at 10–12 °C for 10–12 h, and then the diseased and rotten fruits were removed. A total of 4200 g of blueberries with basically the same size and shape of fruits and more than 95% maturity were selected as the experimental samples.

The blueberries were divided into three groups, A, B, and C, corresponding to three different storage temperatures: refrigerated (0 °C), refrigerated (4 °C), and room temperature (25 °C), with 1400 g each. Then each group A, B, and C was further divided into seven groups of 200 g each according to their weight (each blueberry weighed ranging from 2.0 to 2.5 g), and 90 blueberries were selected as the experimental samples for each day according to this standard and then placed in a 15 × 10 × 5 cm fresh-keeping box and sealed with PE fresh-keeping film (thickness of 0.02 mm), numbered from 0 to 6. The samples were stored in constant temperature boxes at 0 °C, 4 °C, and 25 °C with an interval of 30 min (to leave time for measuring physicochemical indicators), and the relative humidity was controlled at 90–95%. After sorting the samples, the quality and color parameters of the blueberries at the initial stage of storage (T0) were measured immediately to determine the initial values of the blueberries at zero storage time. The samples were stored for 6 days in total. The sensory evaluation, the TPA tests, and the physicochemical indicator measurement experiments were performed once a day in sequence.

2.2. Measurement of Quality Indicators for Blueberries during Shelf Life

2.2.1. Color Characteristics

This paper used a colorimeter (CR-400 model, Konica Minolta Company, Osaka, Japan) to measure the color of blueberry fruits, which was calibrated with a white board. Two evenly distributed points on the equatorial part of each blueberry fruit were selected for color measurement, and the average of five measurements was taken as the color data of that fruit. In addition, this paper also measured the color of the last discolored part of each blueberry fruit—the berry stem insertion area—in order to evaluate the impact of shelf life.

The color data of blueberry fruits were expressed by the International Commission on Illumination system, namely brightness L, red-greenness a, and yellow-blueness b, which formed a three-dimensional color space. To better describe the color characteristics of blueberry fruits, this paper also calculated the following three parameters: chroma C^*, hue angle h^0, and total color difference ΔE, which were given by Formulas (1)–(3), respectively.

$$C^* = \left(a^{*2} + b^{*2}\right)^{1/2},\tag{1}$$

$$h^0 = \tan^{-1} b^*/a^*,\tag{2}$$

$$\Delta E = \sqrt{(L - L_0)^2 + (a - a_0)^2 + (b - b_0)^2},\tag{3}$$

where C^* is a measure of color intensity; h^0 is an indicator of color hue; and L_0, a_0, and b_0 represent the color parameters at the initial moment, respectively.

2.2.2. Weight Loss Rate

Experimental blueberry samples were determined using an electronic analytical balance at time zero (W_0) and after different storage periods (final weight, W_x) to assess weight

loss and averaged over five repetitions. The results were expressed as a percentage weight loss compared to the initial weights.

$$\alpha = \frac{W_0 - W_x}{W_0} \times 100\%,\tag{4}$$

where W_0 is the mass at storage zero; W_x is the mass after different storage periods; and α is the weight loss rate.

2.2.3. Spoilage Rate

The spoilage rate was determined as follows: the percentage of spoilage (mold, softening, rotting, etc.) of blueberry fruit was calculated from the 90 experimental samples to be measured each day. The formula for this was as follows:

$$\beta = \frac{m_x}{m_0} \times 100\%,\tag{5}$$

where m_0 is the total mass of blueberries; m_x is the mass of decayed fruit; and β is the spoilage rate.

2.2.4. Diameter, Length, and Fruit Shape Index

For each group, 30 fruits were randomly selected and measured by vernier calipers for cross and longitudinal diameters, and the average of 5 replicates was used as input: fruit shape index = longitudinal diameter/cross diameter.

2.2.5. Texture Parameters

A texture meter (TA-XT PLUS type, Stable Micro System, Godalming, UK) was used to measure the texture quality of blueberries in the texture profile analysis (TPA) mode. The parameters were set as follows: pre-test speed, 1 mm/s; downward compression speed, 2 mm/s; interval between compressions, 5 s; deformation rate, 25%; starting force, 5 gf; and rate of rise, same as downward compression [14]. The texture attributes of hardness (peak force in the first compression), elasticity (ratio of sample recovery height in the second compression to deformation amount in the first compression), cohesion (ratio of positive area in the second compression to positive area in the first compression), chewiness (product of hardness, cohesion, and elasticity), and recovery (ratio of energy during ascent to energy during descent in the first compression) were calculated from the force–time curves. For each measurement, thirty fruits were randomly selected and replicated five times to obtain the mean values.

2.2.6. Soluble Solids Content, Titratable Acid Content, and Solid to Acid Ratio

A portable digital refractometer (PLA-3 type, ATAGO, Tokyo, Japan) was used to measure the soluble solids content (SSC) of blueberries. For each sample, one fruit was crushed with a mortar, and the juice was filtered through gauze. Two drops of juice were placed on the prism of the refractometer, and the reading was recorded. This procedure was repeated five times, and the mean value was calculated as the soluble solid content of the sample.

Titratable acidity (TA) was determined by titrating clarified blueberry juice with a 0.1 N NaOH (sodium hydroxide) solution until the pH reached 8.2. The juice was obtained by centrifuging and filtering the crushed fruits. A mixture of 10 mL of juice and 10 mL of distilled water was used for each titration. The volume of NaOH consumed was used to calculate the titratable acidity, expressed in mg/g.

The solid-to-acid ratio (SAR) was calculated as the ratio of soluble solid content to titratable acidity.

2.2.7. PH Values

The pH value of the juice was measured with a digital pH meter after centrifuging the juice at 5000 r/min for 15 min and clarifying the supernatant.

2.2.8. Vitamin C

The vitamin C (VC) content of blueberry fruits was determined by UV spectrophotometry. Five blueberry fruits of uniform size were weighed and homogenized in a mortar with 1% HCl. The homogenate was diluted to 25 mL with distilled water and filtered. A 2 mL aliquot of the filtrate was mixed with 0.2 mL of 10% HCl and diluted to 10 mL with distilled water as the blank control. The absorbance of the sample and the blank control was measured at 245 nm using a UV spectrophotometer. Each experiment was repeated three times, and the average value was taken as the VC content of blueberry fruits. The VC content was calculated from a standard curve prepared with ascorbic acid standard solutions ($C = 15.12\,A + 0.1222$, $R^2 = 0.9833$).

2.2.9. Anthocyanins

The anthocyanin content of blueberry fruits was determined by UV spectrophotometry. Five blueberry fruits of uniform size were weighed and homogenized in a mortar and pestle. The homogenate was transferred to a 10 mL centrifuge tube and centrifuged at 3000 rpm for 10 min. The supernatant was collected, and its absorbance was measured at 546 nm using a UV spectrophotometer. Each experiment was repeated three times, and the average value was taken as the anthocyanin content of blueberry fruits. The anthocyanin content in mg/100 g was calculated from the standard curve prepared with the standard solution of anthocyanin-3-glucoside.

2.2.10. Sensory Evaluation

In order to evaluate the sensory quality of blueberries at different temperatures, sensory evaluation experiments were conducted prior to the determination of physicochemical indexes. Ten trained laboratory team members were selected as evaluators, all of whom were blueberry enthusiasts and consumers and had the ability and experience to identify the quality and taste of blueberries. The attributes of the testers are shown in Table 1 and consist of three main aspects: age, gender, and class. A 9-point scale was used to rate the samples, with 1 being strongly disliked, 5 being fair, and 9 being highly liked. Each sample was evaluated for the following three sensory aspects: appearance, flavor, and taste. Appearance includes glossy rind, weightlessness, degree of decay, etc.; flavor includes scent, sweetness, sourness, etc.; and taste includes brittleness, chewiness, and tightness. The sensory evaluation score sheet is shown in Table 2. Evaluators cleaned their mouths before scoring, tasted each sample in turn, and rinsed their mouths with warm water after tasting. Each sample was scored separately by 10 evaluators, and then the average score was calculated.

Table 1. Attributes of the test personnel.

Number	Age	Gender	Class
1	24	Female	Master's degree
2	35	Male	Lecturer
3	21	Female	Undergraduate
4	30	Male	PhD
5	27	Female	Assistant Researcher
6	32	Male	Associate Professor
7	18	Female	Undergraduate
8	58	Male	Professor
9	23	Female	Master's degree
10	47	Male	Fellow

Table 2. Score criteria for the evaluation of blueberry sensory indicators.

Sensory Indicators		Scoring Criteria	Score
Tier 1 Indicators	**Secondary Indicators**		
Appearance	Glossy rind	Bright, evenly colored fruit with a glossy surface	9
		Bright fruit, largely uniform in color, with a shiny surface	7–8
		Basically uniform color, increased waxy layer, reduced lustre, slightly glossy surface	4–6
		Dull surface of the fruit, uneven color, no luster on the surface	1–3
	Water loss	Normal, smooth surface of the fruit, no wrinkling, no elasticity when pressed lightly by hand, no deformation of the fruit	9
		Slight water loss, no surface wrinkling, elastic when pressed with the hand, slightly wrinkled skin, no deformation of the fruit	7–8
		Moderate water loss; slight surface wrinkling of the fruit; the wrinkling intensifies when lightly pressed with the hand; the fruit is lightly deformed and inelastic	4–6
		Heavy water loss, with obvious surface wrinkling of the fruit, which is severely deformed and inelastic when lightly pressed by hand.	1–3
	Weightlessness	Full fruit with little weight loss	9
		Slight loss of weight; small area of ruffling or crumpling around the fruit abscission site	7–8
		Moderate weight loss; with a medium area of pericarp crinkling or folding around the fruit abscission site	5–6
		Severe weight loss, with large areas of skin crinkling or folding around the fruit abscission site	3–4
		Blueberries almost crumpled into dried blueberries, or whole fruit crumpled to resemble mulberries	1–2
	Degree of decay	Fruit surface are free of mold spots	9
		1–3 mold spots on the surface of the fruit	7–8
		25–50% of the fruit area decayed	4–6
		Fruit surface decay greater than 50% of the fruit area	1–3
Flavor	Scent	Honeyed, richly scented, no bad odor	9
		Slightly fragrant, slightly light odor, no bad smell	7–8
		Has an inconspicuous aroma of fruit and no other odor	5–6
		No fruit aroma, slight other odor	3–4
		No aroma of fruit, bad odor, moldy smell	1–2
	Sweetness and acidity	Intense and sweet	9
		Sweeter	7–8
		Moderate	5–6
		More acidic	3–4
		Very sour	1–2

Table 2. *Cont.*

Sensory Indicators		Scoring Criteria	Score
Tier 1 Indicators	Secondary Indicators		
Taste	Brittleness	Very fluffy	1–2
		Fluffy	3–4
		Moderate	5–6
		Crisp	7–8
		Very crisp	9
	Chewiness	Very poor palatability	1–2
		Poor palatability	3–4
		Moderate	5–6
		Good palatability	7–8
		Very good palatability	9
	Tightness	Very low	1–2
		Relatively low	3–4
		Moderate	5–6
		High	7–8
		Very high	9

2.3. Data Processing

2.3.1. Normalization Process

Normalization is a data pre-processing technique that aims to eliminate the differences in dimension and scale among data so that they have the same range or scale. It can reduce data complexity, redundancy, and noise, enhance data analysis and modeling efficiency and accuracy, and avoid issues such as over- or under-fitting. The normalization formula is given by (6).

$$Y_{norm} = \frac{(y - y_{min})}{(y_{max} - y_{min})},\tag{6}$$

where Y_{norm} denotes the normalized data, y denotes the original data, and y_{min}, and y_{max} denote the minimum and maximum values of the original data, respectively.

2.3.2. The MRMR Algorithm

Feature selection is a crucial task in feature engineering, where the objective is to find the optimal subset of features. Common feature selection techniques include filter, wrapper, and embedded methods. Filtering is a feature selection technique that scores each feature based on its variance or correlation and sets a threshold or number of features to be selected. It can be classified into univariate and multivariate filtering methods. Univariate filtering methods, such as Pearson's correlation coefficient, assess the importance of features based on their individual correlation with the outcome variable and filter out the most relevant features. However, they do not consider the interactions or dependencies among features, and they do not capture non-linear relationships. The MRMR algorithm is a multivariate filtering method that aims to select a subset of features from high-dimensional data with high discriminative power. The basic idea of MRMR is that the selected features should be highly correlated with the outcome variable to ensure their validity and have low redundancy among themselves to ensure their diversity [13]. Unlike Pearson's correlation coefficient, MRMR based on Mutual Information (MI) can capture non-linear relationships. MI is a measure of the dependence between two random variables and can be interpreted mathematically as the degree of uncertainty reduction of one random variable x given another random variable y. It is defined as follows:

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) log \frac{p(x,y)}{p(x)p(y)} dxdy,\tag{7}$$

where X and Y denote variables, $p(x, y)$ denotes the joint probability distribution of X and Y, and $p(x)$ and $p(y)$ denote the marginal probability distributions of X and Y.

The MRMR algorithm consists of two parts: max-relevance and min-redundancy. Max-relevance means that the average mutual information between the selected features and the target variable is maximized. This part ensures that the selected features have a high correlation with the target variable, thus enhancing the validity of feature selection. It is formulated as follows:

$$maxD(S, y), D = \frac{1}{|S|}\sum\nolimits_{x_i \in S} I(x_i; y),$$ (8)

where S denotes the selected feature subset, x_i denotes the feature variable, and y denotes the outcome variable.

However, a single max-relevance criterion may suffer from multicollinearity (redundancy) among the selected features, which may degrade the prediction accuracy of the model. To address this issue, a min-redundancy constraint is added to the max-relevance criterion, i.e., the minimum average mutual information among the selected features, which partly ensures a low level of redundancy among the selected features and thus improves the diversity of feature selection. It is formulated as follows:

$$minR(S), R = \frac{1}{|S|^2}\sum\nolimits_{x_i, x_j \in S} I(x_i; x_j).$$ (9)

Combining the two components gives the MRMR. There are two different ways of combining them: additive and multiplicative. The difference between the two ways is that additive combination tends to balance the weights of the two objectives, while multiplicative combination tends to magnify the differences between the two objectives. In other words, additive combination may be more suitable for situations where the features have low correlation, while multiplicative combination may be more suitable for situations where the features have high correlation. Multiplicative integration was chosen because of the high correlation between features in this paper. Define the operator $\varnothing(D, R)$; maximizing this operator will optimize both D and R. The formula is as follows:

$$max\varnothing(D, R), \varnothing = \frac{D}{R}.$$ (10)

In order to facilitate the solution, incremental search is usually used in practice to find near-optimal features. Assuming that the feature set S_{m-1} is already available, the goal of the search is to find the mth feature from the remaining features $X - S_{m-1}$ by selecting features such that $\varnothing$ is maximized. The incremental search can be expressed as follows:

$$max\Delta, \Delta = max\left[\frac{I(x_j; y)}{\frac{1}{m-1}\sum_{x_i \in S_{m-1}} I(x_j; x_i)}\right],$$ (11)

The specific process of the MRMR algorithm can be expressed as follows:

Step 1: Select the initial feature subset S. Select the feature x_j from the feature set X that has the largest mutual information with the target variable from the optimal feature subset S;

Step 2: Design a search strategy. Use the incremental search strategy to select the features that maximize the value of the objective function x_j into the optimal feature subset S;

Step 3: Output the results. When the value of the objective function is less than the threshold, output the selected feature and stop the search; otherwise, execute the second step.

2.4. Construction of the Blueberry Shelf-Life Prediction Model

2.4.1. The BPNN Model

This paper is based on the BPNN model to construct predictive models for food shelf life. BPNN is a feed-forward supervised learning model with multiple layers of fully connected neurons, including input, hidden, and output layers. Each neuron receives weighted and summed inputs from the previous layer and outputs them to the next layer through a non-linear activation function. The model calculates the error function between the output layer and the target value and adjusts the weights by backpropagating the error signal using the chain rule. The model iterates until the error function is minimized, allowing the network to approximate the desired output. BPNN can capture the non-linear relationships between individual indicators and is widely used in food shelf-life prediction research [15].

2.4.2. The GDEDBO Model

The DBO is a novel swarm intelligence optimization algorithm proposed by Professor Shen Bo's team in 2022, inspired by the ball rolling, dancing, foraging, stealing, and breeding behavior of dung beetles [10]. Simple and effective, the DBO has been successfully applied to several engineering design problems and outperforms other metaheuristic algorithms in the shelf-life prediction of this paper. However, the DBO still suffers from the problems of an imbalance between global and local search capabilities and the tendency to fall into local optimal solutions [9]. To address these problems, the dung beetle optimizer using Gaussian distribution estimation strategy (GDEDBO) is proposed in this paper based on the original algorithm with the following improvement strategy.

Chaotic Mapping

Chaotic mapping is a method that aims to overcome the problem of heuristic algorithms getting trapped in local optima. It is inspired by the chaotic phenomenon that some systems may have unpredictable variables [16]. By introducing unknown variables into a heuristic algorithm and applying the principle of chaotic phenomena, the algorithm's accuracy can be improved, and the global search ability can be enhanced [17]. The tent chaotic mapping can generate a uniform chaotic sequence in the $(0, 1)$ interval, where x_i is the ith particle in the chaotic sequence, as shown in Equation (12):

$$x_{i+1} = \begin{cases} 2x_i, 0 \leq x_i < 0.5 \\ 2(1 - x_i), 0.5 \leq x_i < 1 \end{cases}. \tag{12}$$

Since the chaotic sequences generated by the chaos operator during the iterative process are periodic and more likely to fall into unknown periodic points, the random variable *rand* function can be added to stabilize them. The *tent* operator requires the ability to handle large amounts of data via Bernoulli transformation and yields Equation (13), where $rand(0, 1)$ is a random number in the $[0, 1]$ interval of random numbers:

$$x_{i+1} = \begin{cases} 2x_i + rand(0, 1), 0 \leq x_i < 0.5 \\ 2(1 - x_i) + rand(0, 1), 0.5 \leq x_i < 1 \end{cases}. \tag{13}$$

To generate a more uniform initialization of the population than the original DBO algorithm, which used a random initial population, this paper applied tent chaos mapping. This method produces a relatively uniform distribution of the system that can enhance the diversity of the population and improve the initial solution of the algorithm. Figure 1 shows the distribution of the initialized populations using tent chaos mapping and the random method.
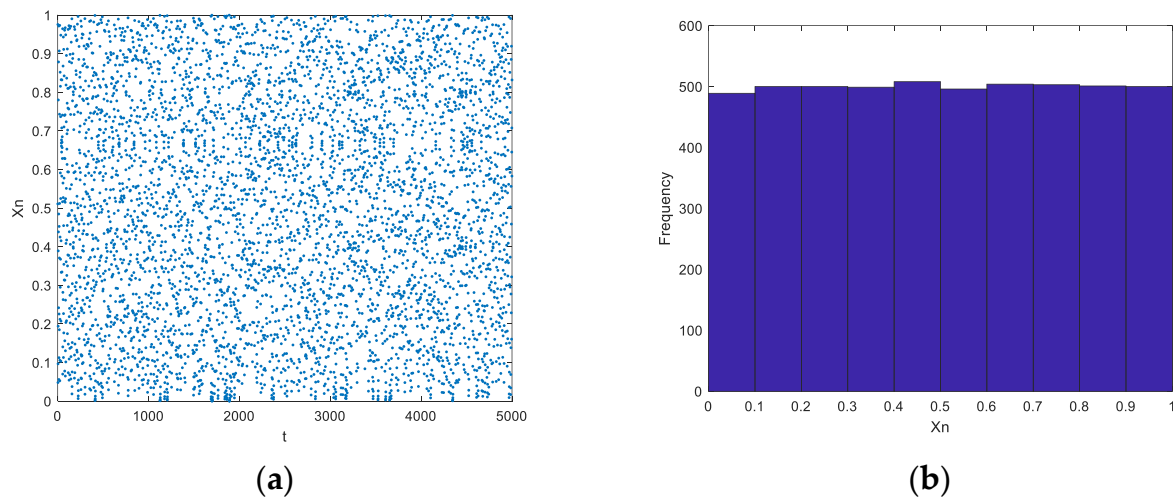
(a)

(b)

Figure 1. Population initialization for the tent chaos mapping: (**a**) scatter plot and (**b**) frequency distribution histogram.

Elite Pool Strategy

The elite pool strategy is a commonly used improvement method for swarm intelligence optimization algorithms that aims to preserve a certain number of the current optimal solutions after each iteration and replace some worse solutions with them in the next iteration, thus retaining excellent solutions and accelerating convergence. This paper defines the rolling dung beetle as the leader and introduces the elite pool strategy into its position update formula to enhance the diversity of the leader. The rolling dung beetle is the first to start searching, and its position update mode is affected by the global worst position (the sun) and obstacles. To avoid local optima and enhance population diversity, this paper proposes an elite pool strategy by referring to the hierachy of the GWO algorithm, in which the top 3 currently optimal individuals and the weighted average position of the leader population are stored, as shown in Figure 2. The 3 optimum individuals can better help the rolling dung beetle determine its rolling direction, while the weighted average position represents the evolutionary trend of the whole dominant population, which is beneficial for algorithm exploration. The leader randomly selects an individual from the elite pool as the rolling direction when updating its position, which to some extent avoids the deficiency of being trapped in a local optimum due to having only one rolling direction. The elite pool strategy is described as follows:

$$X_{\text{Elite}}^t = \left[X_{\text{best1}}^t, X_{\text{best2}}^t, X_{\text{best3}}^t, X_{\text{mean}}^t \right] \tag{14}$$

$$X_{\text{mean}}^t = \sum_{i=1}^{N/2} w_i \times X_t^i \tag{15}$$

$$w_i = \frac{\ln\left(N/2 + 0.5\right) - \ln i}{\sum_{i=1}^{N/2}\left(\ln\left(N/2 + 0.5\right) - \ln i\right)} \tag{16}$$

where X_{best1}^t, X_{best2}^t, and X_{best3}^t are the top three individuals with the best fitness value of the current population, respectively; X_{mean}^t denotes the weighted average position of the rolling dung beetles; N is the population size; X_i^t is the position of the ith rolling dung beetle in generation t; and w_i denotes the weighting coefficients among the rolling dung beetles in descending order of fitness value.
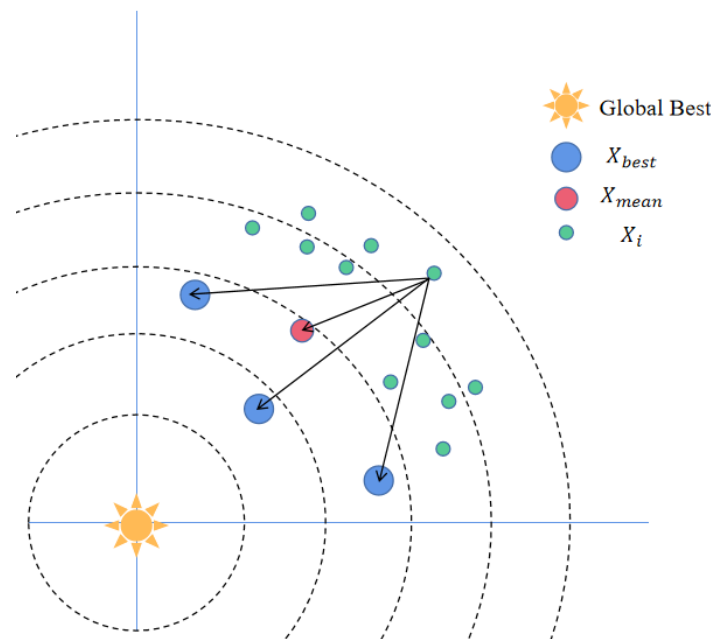
Figure 2. Elite pool strategy diagram.

The adjusted formula for updating the position of the rolling dung beetle is as follows:

$$X_i^{t+1} = \begin{cases} X_i^t - \alpha \cdot \dfrac{X_i^t - X_{\text{Elite}}^t}{\|X_i^t - X_{\text{Elite}}^t\|}, & \text{accessible mode} \\ X_i^t - tan\theta \cdot \dfrac{X_i^t - X_{\text{Elite}}^t}{\|X_i^t - X_{\text{Elite}}^t\|}, & \text{barrier mode} \end{cases}$$
$$X_{\text{Elite}}^t \in \left\{ X_{\text{best1}}^t, X_{\text{best2}}^t, X_{\text{best3}}^t, X_{\text{mean}}^t \right\} \tag{17}$$

where X_{Elite}^t is a randomly selected individual in the elite pool, α denotes the roll speed coefficient, and θ denotes the roll direction.

Gaussian Distribution Estimation Strategy

The Gaussian distribution estimation strategy is a position update method based on probability distribution that simulates information exchange and knowledge sharing among population individuals by using the Gaussian distribution function to improve population diversity and exploration. In this paper, spawning dung beetles and small dung beetles are defined as followers. Spawning dung beetles are generated by rolling dung beetles, whose position updates are attracted by the local optimal position (spawning area) to increase search accuracy; small dung beetles are hatched by spawning dung beetles, whose position updates are attracted by the global optimal position (optimal foraging area) to further improve search accuracy. However, the position updates of the spawning dung beetles and the small dung beetles depend only on information about themselves and adjacent individuals, and a lack of communication with other individuals can easily lead to local optimums converging prematurely. To enhance the optimization ability of the algorithm, this paper improves the follower position formula by using the Gaussian distribution estimation strategy. The formula is as follows:

$$X_i^{(t+1)} = \text{mean} + y, y \sim N(0, \text{Cov}), \tag{18}$$

$$\text{mean} = \frac{\text{elite}_t + X_{\text{mean}}^t + X_i^t}{3} \tag{19}$$

$$\text{Cov}(i) = \frac{1}{N/2} \sum_{i=1}^{N/2} \left(X_i^t - X_{\text{mean}}^t\right) \times \left(X_i^t - X_{\text{mean}}^t\right)^T \tag{20}$$

where Cov is the weighted covariance matrix of the dominant population. Figure 3 shows that by using the Gaussian distribution estimation strategy to adjust the search direction, the algorithm can effectively balance the global and local search abilities and fully utilize the dominant population information to guide the better evolution of the population, thus enhancing the algorithm's search performance.
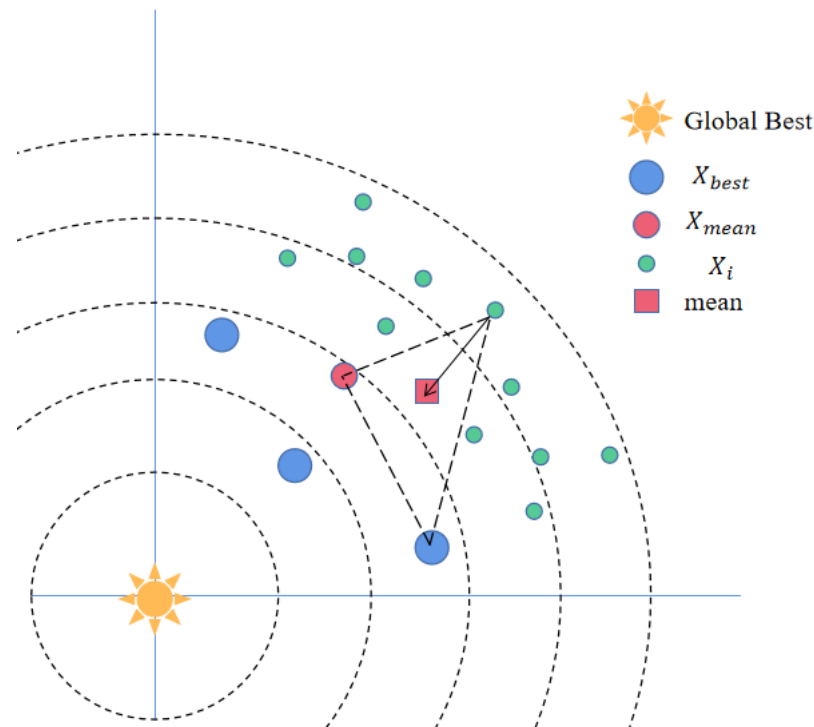

Figure 3. Gaussian distribution estimation diagram.

Finally, a superiority mechanism is applied to preserve the dominant individuals and thus ensure the convergence of the IDBO algorithm, with the following position update formula:

$$X_i^{(t+1)} = \begin{cases} X_i^{(t+1)}, & f\left(X_i^{(t+1)}\right) < f\left(X_i^t\right) \\ X_i^t, & f\left(X_i^{(t+1)}\right) > f\left(X_i^t\right) \end{cases}. \tag{21}$$

2.4.3. The MRMR-GDEDBO-BPNN Model

The MRMR-GDEDBO-BPNN model proposed in this paper consists of four main components, as follows:

(1) Feature selection: The MRMR algorithm was used to select the key factors for predicting the shelf life of blueberries as input. This algorithm reduced the feature dimensionality by filtering out redundant features that could affect the model prediction accuracy;

(2) The GDEDBO model: To address the problems of the original DBO, this paper proposes three improvement strategies. First, tent chaotic mapping was introduced to initialize the dung beetle population and increase its diversity. Second, enhancing the diversity of leaders by introducing an elite pool strategy to improve the exploration performance of the algorithm; and third, using the Gaussian distribution estimation (GDE) strategy to use dominant population information to adjust the search direction to guide the population to better evolution;

(3) The GDEDBO-BPNN model: The GDEDBO-BPNN used the GDEDBO to optimize the parameter weights w and thresholds b of the BPNN and updated the w and b of the BPNN by continuously updating the positions of the dung beetles until the global best position, i.e., the optimal solution, was found;

(4) Forecasting model and evaluation of results: The MRMR-GDEDBO-BPNN model predicted the shelf life of blueberries, using the physicochemical and quality indicators selected by the MRMR algorithm as input parameters and selecting different evaluation indicators to evaluate the prediction results. Figure 4 shows the flow chart of the MRMR-GDEDBO-BPNN model.
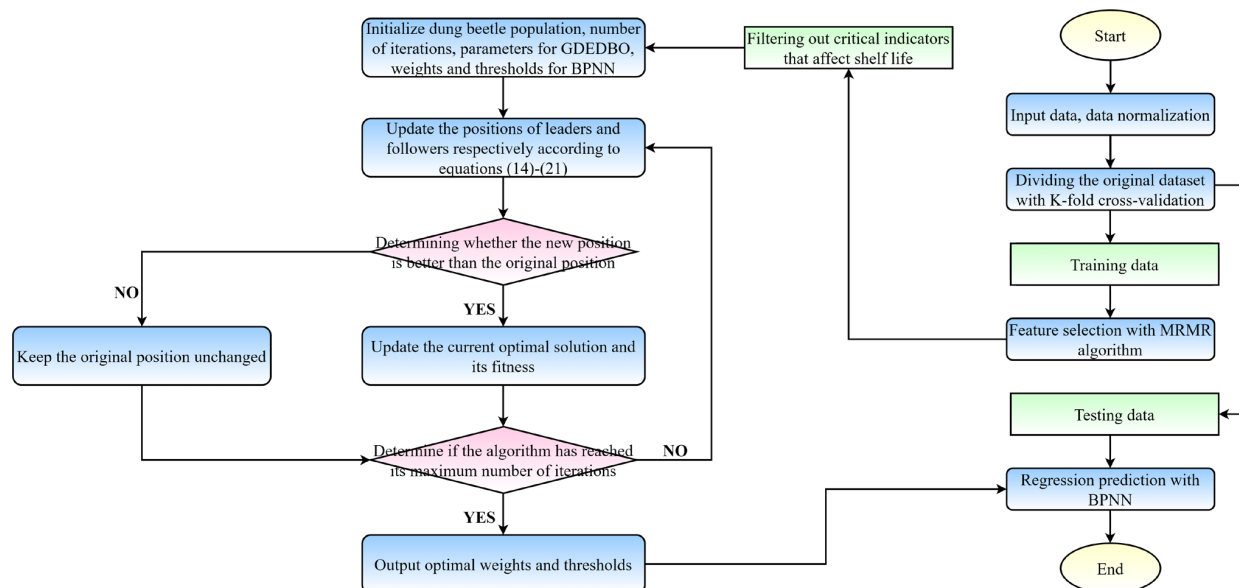

Figure 4. Flow chart of the MRMR-IDBO-BPNN model.

3. Experimental Simulation and Analysis

In order to fully verify the superior performance of the GDEDBO algorithm, this paper selects 14 standard test functions from the IEEE CEC2017 single-objective test function set to test the performance of the algorithm. As shown in Table 3, F1–F8 belong to unimodal test functions, which are used to check the convergence accuracy of the algorithm; F9–F14 belong to multimodal test functions, which are used to test the ability of the algorithm to jump out of the local optimum. The hybrid firefly and particle swarm optimization (HFPSO) [18], the gray wolf optimization algorithm with a hybrid Gaussian distribution estimation strategy (GEDGWO) [19], the Harris Hawk optimization algorithm (HHO) [20], and the original DBO algorithm were also selected as comparison algorithms, with their parameter settings referring to the original literature, as shown in Table 4. HFPSO is an improved particle swarm algorithm with a hybrid firefly algorithm. GEDGWO is a variant of the GWO algorithm incorporating a distribution estimation algorithm. Both of these improved algorithms have demonstrated good performance in their respective literature, while HHO and DBO are the latest proposed swarm intelligence optimization algorithms that have performed well in recent years and have been applied in different disciplines and engineering fields. Therefore, a comparison using these algorithms can improve the validity of the improvement strategies proposed in this paper. To ensure fairness, the number of populations $N = 50$, dimension $D = 30$, and the maximum number of iterations were set uniformly at 500. Each algorithm was run 50 times independently, and the optimal value, mean, and standard deviation of the fitness were recorded to evaluate the exploration performance, convergence accuracy, and stability of the algorithms. The experiments were implemented using MATLAB 2022b on a Windows 11, 64 bit system with an Intel® Core™ i7-11700 2.5 GHz, 16 GB RAM computer. The experimental results are shown in Table 5.

Table 3. Selected standard test functions.

Function	Dim	Range	f_{min}
$f_1(x) = \sum_{i=1}^{n} x_i^2$	30/50/100	$[-100,100]$	0
$f_2(x) = \sum_{i=1}^{n}\|x_i\| + \prod_{i=1}^{n}\|x_i\|$	30/50/100	$[-10,10]$	0
$f_3(x) = \sum_{i=1}^{n}\left(\sum_{j-1}^{i} x_j\right)^2$	30/50/100	$[-100,100]$	0
$f_4(x) = max\{\|x_i\|, 1 \leq i \leq n\}$	30/50/100	$[-100,100]$	0
$f_5(x) = \sum_{i=1}^{n-1}\left[100(x_{i+1} - x_{i2})^2 + (x_i - 1)^2\right]$	30/50/100	$[-30,30]$	0
$f_6(x) = \sum_{i=1}^{n}(x_i + 5)^2$	30/50/100	$[-100,100]$	0
$f_7(x) = x_1^2 + 10^6\sum_{i=2}^{n} x_i^2$	30/50/100	$[-100,100]$	0
$f_8(x) = \sum_{i=1}^{n} x_i^2 + \left(\sum_{i=1}^{n} 0.5ix_i\right)^2 + \left(\sum_{i=1}^{n} 0.5ix_i\right)^4$	30/50/100	$[-5,10]$	0
$f_9(x) = \sum_{i=1}^{n}\left[x_i^2 - 10\cos(2\pi x_i) + 10\right]$	30/50/100	$[-5.12,5.12]$	0
$f_{10}(x) = \sum_{i=1}^{n}\|x_i\sin(x_i) + 0.1x_i\|$	30/50/100	$[-10,10]$	0
$f_{11}(x) = \frac{\pi}{n}\left\{10\sin(\pi y_1) + \sum_{i=1}^{n-1}(y_i - 1)^2\left[1 + 10\sin^2(\pi y_{i+1})\right] + (y_n - 1)^2\right\} + \sum_{i=1}^{n} u(x_i, 10, 100, 4).$ $where\ y_i = 1 + \frac{x_i+1}{4}, for\ all\ i = 1, \cdots, n$ $u(x_i, a, k, m) = \begin{cases} k(x_i - a)^m & x_i > a \\ 0 & -a < x_i < a \\ k(-x_i - a)^m & x_i < -a \end{cases}$	30/50/100	$[-50,50]$	0
$f_{12}(x) = 0.1$ $\left\{\sin^2(3\pi x_1) + \sum_{i=1}^{n}(x_i - 1)^2\left[1 + \sin^2(3\pi x_i + 1)\right] + (x_n - 1)^2\left[1 + \sin^2(2\pi x_n)\right]\right\} +$ $\sum_{i=1}^{n} u(x_i, 5, 100, 4)$	30/50/100	$[-50,50]$	0
$f_{13}(x) = \left[\frac{1}{n-1}\sum_{i=1}^{n-1}\left(\sqrt{s_i} \times \left(\sin(50s_i^{0.2}) + 1\right)\right)\right]^2$ $s_i = \sqrt{x_i^2 + x_{i+1}^2}$	30/50/100	$[-100,100]$	0
$f_{14}(x) = \sin^2(\pi y_1) + \sum_{i=1}^{n-1}(y_i - 1)^2\left[1 + 10\sin^2(\pi y_i + 1)\right]$ $+(y_n - 1)^2\left[1 + \sin^2(2\pi y_n)\right],$ $where\ y_i = 1 + \frac{x_i-1}{4}, for\ all\ i = 1, \cdots, n$	30/50/100	$[-10,10]$	0

Table 4. Algorithm Parameter Settings.

Algorithms	Parameters Settings
HFPSO	$C_1 = C_2 = 1.49445$, $a = 0.2$, $B_0 = 2$ $\gamma = 1$, $\omega_i = 0.9$, $\omega_f = 0.5$
GEDGWO	$r_3 \in (0, 1)$, $r_4 \in (0, 1)$, $r_5 \in (0, 1)$
HHO	$\alpha = 0.01$, $\beta = 1.5$
DBO	$\alpha = \beta = 0.1$, $a = 0.3$, $b = 0.5$
GDEDBO	$\alpha = \beta = 0.1$, $a = 0.3$, $b = 0.5$, $P_{percent} = 0.4 - (t * (0.4 - 0.2)/M);$

3.1. Development and Exploration Capacity Assessment

As can be seen from Table 5, for the unimodal test functions F1 to F8, GDEDBO achieved better solutions on six functions except F5 and F6, indicating that the enhancement strategy can effectively improve the development of the algorithm. GEDGWO ranked first on F5. For the multimodal test functions F9 to F14, GDEDBO ranked first on four functions except F11 and F13, and GEDGWO performed better on F11. For F9, although all the algorithms converge to the theoretical optimum, only DBO and GDEDBO have a mean and standard deviation of 0, indicating that their stability and exploration of the multimodal functions are better than those of the other compared algorithms. It proves that the improvement strategy proposed in this paper can enhance the exploitation and exploration capabilities of the algorithm.

Table 5. Results of standard function tests for each algorithm.

F	Index	HFPSO	GEDGWO	HHO	DBO	GDEDBO
F1	Best	1.67×10^{-95}	2.25×10^{-181}	7.22×10^{-35}	1.81×10^{-115}	3.02×10^{-210}
	Mean	2.55×10^{-83}	4.32×10^{-113}	6.42×10^{-33}	1.31×10^{-99}	9.59×10^{-159}
	STD	1.20×10^{-82}	2.37×10^{-112}	1.69×10^{-32}	5.51×10^{-99}	5.25×10^{-158}
F2	Best	4.85×10^{-62}	1.16×10^{-85}	3.85×10^{-21}	4.97×10^{-62}	2.26×10^{-109}
	Mean	3.92×10^{-54}	6.14×10^{-61}	6.55×10^{-20}	8.26×10^{-53}	4.20×10^{-83}
	STD	1.89×10^{-53}	2.35×10^{-60}	4.77×10^{-20}	3.59×10^{-52}	2.81×10^{-82}
F3	Best	6.29×10^{3}	5.73×10^{-149}	2.27×10^{-11}	1.01×10^{-110}	1.42×10^{-191}
	Mean	2.45×10^{4}	9.90×10^{-45}	9.65×10^{-8}	6.46×10^{-79}	1.22×10^{-115}
	STD	9.33×10^{3}	7.00×10^{-44}	3.94×10^{-7}	4.57×10^{-78}	7.98×10^{-115}
F4	Best	2.99×10^{-6}	4.31×10^{-84}	2.69×10^{-9}	2.57×10^{-57}	1.96×10^{-102}
	Mean	3.70×10^{1}	3.15×10^{-53}	1.88×10^{-8}	7.21×10^{-48}	4.12×10^{-68}
	STD	2.88×10^{1}	2.23×10^{-52}	1.33×10^{-8}	4.23×10^{-47}	2.31×10^{-67}
F5	Best	2.69×10^{1}	1.00×10^{-6}	2.52×10^{1}	2.45×10^{1}	2.44×10^{1}
	Mean	2.75×10^{1}	3.01×10^{-3}	2.67×10^{1}	2.52×10^{1}	2.50×10^{1}
	STD	4.83×10^{-1}	3.32×10^{-3}	7.85×10^{-1}	7.38×10^{-1}	3.28×10^{-1}
F6	Best	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}
	Mean	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}
	STD	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}
F7	Best	2.01×10^{-88}	1.92×10^{-155}	5.23×10^{-29}	4.29×10^{-113}	1.38×10^{-196}
	Mean	9.19×10^{-77}	1.16×10^{-111}	2.88×10^{-27}	1.03×10^{-93}	3.10×10^{-152}
	STD	4.61×10^{-76}	6.14×10^{-111}	4.91×10^{-27}	5.39×10^{-93}	1.30×10^{-151}
F8	Best	2.51×10^{-93}	2.02×10^{-163}	9.38×10^{-37}	2.78×10^{-118}	3.58×10^{-202}
	Mean	2.98×10^{-83}	2.30×10^{-110}	7.05×10^{-35}	2.29×10^{-103}	7.34×10^{-160}
	STD	1.63×10^{-82}	1.26×10^{-109}	1.25×10^{-34}	6.75×10^{-103}	3.86×10^{-159}
F9	Best	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}
	Mean	1.89×10^{-15}	4.61×10^{-1}	3.15×10^{0}	0.00×10^{0}	0.00×10^{0}
	STD	1.04×10^{-14}	2.19×10^{0}	4.36×10^{0}	0.00×10^{0}	0.00×10^{0}
F10	Best	1.60×10^{-60}	1.08×10^{-80}	3.31×10^{-20}	1.02×10^{-61}	8.18×10^{-112}
	Mean	8.12×10^{-1}	1.48×10^{-1}	3.65×10^{-2}	1.36×10^{-46}	6.70×10^{-85}
	STD	4.45×10^{0}	8.11×10^{-1}	1.97×10^{-1}	7.47×10^{-46}	2.51×10^{-84}
F11	Best	1.02×10^{-3}	8.84×10^{-9}	6.30×10^{-1}	2.45×10^{-11}	2.87×10^{-6}
	Mean	1.38×10^{-2}	2.60×10^{-6}	3.25×10^{-2}	3.54×10^{-4}	2.27×10^{-5}
	STD	2.42×10^{-2}	2.49×10^{-6}	2.23×10^{-2}	1.40×10^{-3}	3.04×10^{-5}
F12	Best	2.67×10^{-2}	5.18×10^{-10}	9.78×10^{-2}	1.87×10^{-5}	2.84×10^{-7}
	Mean	2.30×10^{-1}	1.03×10^{-1}	3.90×10^{-1}	1.33×10^{-2}	3.52×10^{-5}
	STD	1.71×10^{-1}	7.80×10^{-2}	2.10×10^{-1}	3.04×10^{-2}	3.61×10^{-5}
F13	Best	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}
	Mean	1.83×10^{-5}	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}
	STD	7.91×10^{-5}	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}
F14	Best	5.81×10^{-2}	4.27×10^{-9}	5.43×10^{-1}	1.60×10^{-4}	2.03×10^{-8}
	Mean	4.71×10^{-1}	1.19×10^{-1}	1.03×10^{0}	1.04×10^{-2}	5.43×10^{-5}
	STD	4.66×10^{-1}	1.14×10^{-1}	2.56×10^{-1}	2.72×10^{-2}	6.87×10^{-5}

3.2. Convergence Curve Analysis

In order to reflect more intuitively the convergence speed of each algorithm and the ability to jump out of the local optimum, the convergence curves of some of the tested functions are shown in Figure 5. As can be seen from Figure 5, GDEDBO has faster convergence speed and convergence accuracy on five functions except F6 and F13. DBO ranks first on F6 and has a faster convergence speed than GDEDBO, which may be due to the elite pooling strategy. GDEDBO ranks second on F13, which proves that the improvement strategy proposed in this paper can effectively improve the convergence performance of the algorithm.
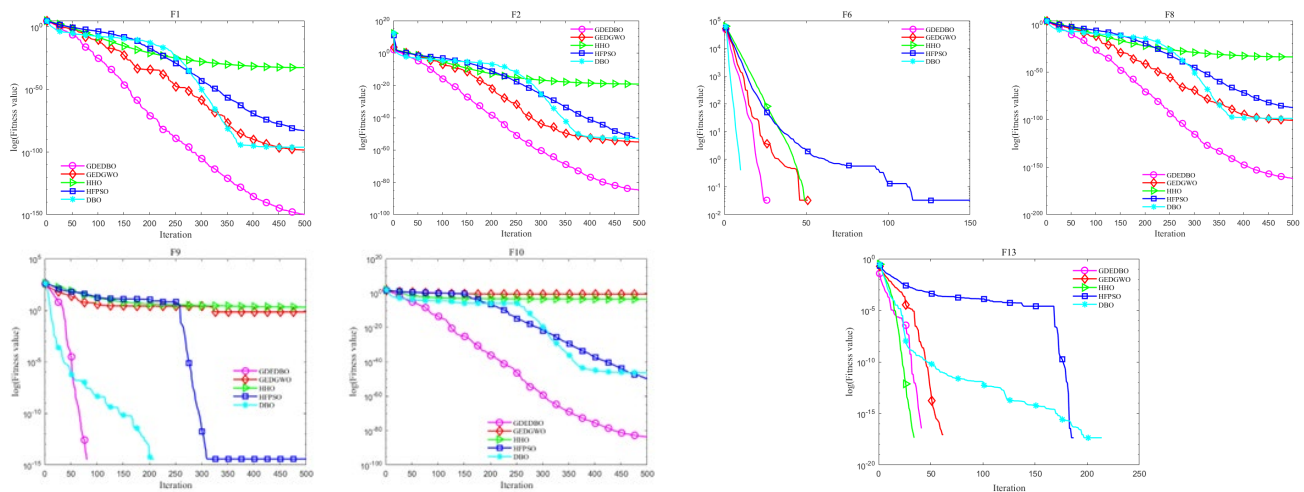
Figure 5. Part of the test function convergence curves.

3.3. Wilcoxon Rank Sum Test

The robustness of the algorithms was evaluated by applying the Wilcoxon rank sum test to compare the overall results of the algorithms. The Wilcoxon rank sum test was conducted at the 5% significance level, and a *p*-value less than 5% indicated a significant difference between the two compared algorithms, while a *p*-value greater than or equal to 5% indicated no significant difference or similar performance in finding the best solution. In this study, five algorithms were used as samples, and each algorithm was run 50 times independently with a population size of $N = 50$ and a dimension of $D = 30$ on 14 standard test functions. The results obtained by the GDEDBO were compared with those obtained by the four other algorithms using the Wilcoxon rank sum test. The *p*-values from the Wilcoxon rank sum test are presented in Table 6. N/A means not applicable, as both merit search results are 0 and there is no data to compare, indicating similar performance. As shown in Table 6, most of the *p*-values are less than 5%, suggesting that the GDEDBO model has a significant difference from the other comparison algorithms.

Table 6. *p*-values of the Wilcoxon rank sum test.

Functions	HFPSO	GEDGWO	HHO	DBO
F1	7.07×10^{-18}	7.07×10^{-18}	7.07×10^{-18}	9.54×10^{-18}
F2	7.07×10^{-18}	1.08×10^{-17}	7.07×10^{-18}	7.07×10^{-18}
F3	7.07×10^{-18}	1.99×10^{-12}	7.07×10^{-18}	7.07×10^{-18}
F4	7.07×10^{-18}	1.32×10^{-14}	7.07×10^{-18}	7.07×10^{-18}
F5	3.02×10^{-11}	3.01×10^{-11}	3.01×10^{-11}	3.01×10^{-11}
F6	N/A	N/A	N/A	N/A
F7	3.01×10^{-11}	3.69×10^{-11}	3.02×10^{-11}	3.02×10^{-11}
F8	3.01×10^{-11}	1.33×10^{-10}	3.02×10^{-11}	3.02×10^{-11}
F9	3.34×10^{-1}	8.15×10^{-2}	1.84×10^{-10}	N/A
F10	3.02×10^{-11}	3.02×10^{-11}	3.02×10^{-11}	3.02×10^{-11}
F11	3.02×10^{-11}	8.48×10^{-9}	3.02×10^{-11}	1.69×10^{-9}
F12	1.96×10^{-10}	8.66×10^{-5}	4.08×10^{-11}	8.89×10^{-10}
F13	1.61×10^{-1}	N/A	N/A	N/A
F14	3.02×10^{-11}	1.22×10^{-2}	3.02×10^{-11}	4.08×10^{-11}
$+/=/-$	11/3/0	12/2/0	12/2/0	11/3/0

The results of the benchmark function test, the convergence curves of some of the tested functions, and the Wilcoxon rank sum test results of each algorithm were combined to assess the performance of the GDEDBO. It was found that the GDEDBO had significantly enhanced local and global capabilities and outperformed comparative optimization

algorithms such as the original DBO and HHO, showing better exploration performance, convergence accuracy, and stability.

4. Model Parameter Settings and Evaluation Criteria

4.1. Model Parameter Settings

4.1.1. The Baseline Model

The performance of the MRMR-GDEDBO-BPNN model proposed here was verified by comparing it with MLR, SVR, ANN, BPNN, MRMR-BPNN, DBO-BPNN, MRMR-DBO-BPNN, GDEDBO-BPNN, and MRMR-GDEDBO-BPNN as the benchmark models. Grid search was used to optimize the parameters of the benchmark models for improved training speed and accuracy. Tables 7 and 8 show the optimal parameter combinations for the SVR and ANN models and for the BPNN model, respectively.

Table 7. Optimal combination of parameters for the benchmark model.

Models	Optimal Combination of Parameters
SVR	{'C': 1, 'epsilon': 0.5, 'kernel': 'linear'}
ANN	{'activation': 'tanh', 'alpha': 0.0001, 'hidden_layer_sizes': (5), 'learning_rate': 'adaptive', 'learning_rate_init': 0.01, 'solver': 'sgd'}

Table 8. Optimal prediction models for the shelf life of blueberries at different storage temperatures and their corresponding topologies and activation functions.

Temperature	Model	Neuron Configuration	Topology	Hidden and Output Activations	Test			
					MAE	MSE	MAPE	R^2
0 °C	BPNN	(6,7)	23–6–7–1	LOGSIG-PURELIN	0.2325	0.0028	0.0981	0.9721
	MRMR-BPNN	6	7–6–1	LOGSIG-PURELIN	0.1174	0.0019	0.1015	0.9775
	DBO-BPNN	(7,9)	23–7–9–1	LOGSIG-TANSIG	0.2072	0.0626	0.0713	0.9827
	MRMR-DBO-BPNN	5	7–5–1	LOGSIG-PURELIN	0.1158	0.0401	0.0351	0.9893
	GDEDBO-BPNN	(8,8)	23–8–8–1	PURELIN-LOGSIG	0.1685	0.0406	0.0561	0.9881
	MRMR-GDEDBO-BPNN	5	7–5–1	LOGSIG-PURELIN	0.0304	0.0014	0.0114	0.9995
4 °C	BPNN	(6,7)	23–6–7–1	LOGSIG-PURELIN	0.2226	0.0035	0.1400	0.9630
	MRMR-BPNN	6	7–6–1	LOGSIG-PURELIN	0.1798	0.0026	0.1229	0.9744
	DBO-BPNN	(8,8)	23–8–8–1	POSLIN-PURELIN	0.2040	0.0611	0.0844	0.9797
	MRMR-DBO-BPNN	5	7–5–1	LOGSIG-TANSIG	0.1183	0.0401	0.0504	0.9837
	GDEDBO-BPNN	(7,9)	23–7–9–1	TANSIG-PURELIN	0.2071	0.0631	0.0573	0.9833
	MRMR-GDEDBO-BPNN	5	7–5–1	LOGSIG-PURELIN	0.0671	0.0077	0.0237	0.9971
25 °C	BPNN	(8,8)	23–8–8–1	LOGSIG-PURELIN	0.2862	0.0049	0.1533	0.9516
	MRMR-BPNN	8	7–7–1	LOGSIG-PURELIN	0.1955	0.0084	0.1393	0.9574
	DBO-BPNN	(9,7)	23–9–7–1	LOGSIG-PURELIN	0.2259	0.0906	0.1073	0.9639
	MRMR-DBO-BPNN	9	7–7–1	LOGSIG-PURELIN	0.1248	0.0548	0.0506	0.9793
	GDEDBO-BPNN	(9,9)	23–8–8–1	POSLIN-PURELIN	0.1313	0.0737	0.0644	0.9724
	MRMR-GDEDBO-BPNN	5	7–5–1	LOGSIG-PURELIN	0.1013	0.0253	0.0301	0.9910

4.1.2. The BPNN Model

Choice of Network Functions

BPNNs commonly employ functions such as activation functions, training functions, learning functions, etc.

The activation function is a key component of a neural network that determines how the neurons generate outputs from the inputs. Activation functions enable neural networks to model nonlinearly, allowing them to approximate complex data and functions [21]. The activation functions frequently used in BPNNs are the logistic function LOGSIG, the hyperbolic tangent function TANSIG, the linear function PURELIN, and the rectified linear unit POSLIN. LOGSIG and TANSIG are often applied as activation functions for the input layer, while PURELIN is often applied as the activation function for the output layer. This paper determines the activation function that minimizes the model error by trial and error

and presents the results in Table 8. It shows that the activation function combination that minimizes the error of the MRMR-GDEDBO-BPNN model at three temperatures, 0, 4, and 25 °C, is LOGSIG-PURELIN.

The training functions of BPNN are functions used to adjust the network parameters, usually based on the gradient descent method or its variants, aiming to minimize the network error. Commonly used training functions include trainlm, trainrp, trainscg, trainbfg, and traingdx. Among them, trainlm is an algorithm that combines gradient descent and Newton's method, which has the advantages of fast convergence and high accuracy, so this paper uses trainlm as the training function.

The common learning functions of BPNN include learngd (gradient descent method) and learngdm (gradient descent method with momentum). The learngd function only adjusts the weights and bias according to the gradient direction and learning rate, which are slow to converge and prone to falling into local optima. The learngdm function increases the momentum constant to effectively overcome the drawbacks of the gradient descent method, so this paper uses learngdm as the learning function, and its formula is shown in (22):

$$\Delta W = \alpha \Delta W_{prev} - (1 - \alpha)\eta \frac{\partial E}{\partial W}, \tag{22}$$

where ΔW denotes the amount of change in weight or bias; α denotes the momentum constant, between 0 and 1; ΔW_{prev} denotes the amount of change in the previous weight or bias; and η denotes the learning rate; $\frac{\partial E}{\partial W}$ denotes the direction of the gradient.

Selection of Topologies

The topology of the BPNN model is formed by the number of neurons in each layer and the connections between two adjacent layers. The topology affects the complexity and expressiveness of the neural network, which in turn affects its performance and convergence speed [22]. Therefore, selecting a suitable topology is a critical step in neural network design. The optimal number of neurons in the hidden layer should avoid underfitting and overfitting. Underfitting occurs when the network has insufficient neurons in the hidden layer to capture the complex features of the data. Overfitting or gradient vanishing occurs when the network has excessive neurons in the hidden layer that fit too closely to the training data [23]. The number of neurons in the hidden layer is not fixed but varies according to the complexity of the problem and the size of the data. Choosing an appropriate number of neurons in the hidden layer is essential for improving the generalization and prediction abilities of the model.

An empirical formula (Equation (23)) was proposed to estimate the number of neurons in the hidden layer, which ranged from 4 to 12. The optimal number of neurons for different models with the minimum error was obtained by grid search and manual parameter tuning, as shown in Table 8. Table 8 shows that the MRMR-GDEDBO-BPNN model had the lowest error with 5 neurons in the hidden layer (single layer) at all three temperatures of 0, 4, and 25 °C.

$$N_h = \sqrt{(N_i + N_o)} + \alpha, \alpha \in [1, 10], \tag{23}$$

where N_h, $+ + N_i$, and N_o represent the number of neurons in the hidden layer, input layer, and output layer, respectively, and α is a modulation constant between 1 and 10.

The hidden layer of a neural network allows it to handle non-linearly separable data, which cannot be represented by a neural network without hidden layers [22]. The number of hidden layers influences the representation and adaptation of a neural network. More hidden layers generally reduce error but also increase complexity, training difficulty, and the risk of overfitting. This paper used double hidden layer and single hidden layer neural network models for 23 input variables without feature selection and 7 input variables after MRMR filtering, respectively. Different combinations of activation functions were also tested to determine the optimal network topology.

Table 8 shows that the best topology for the GDEDBO-BPNN model with the lowest error without feature selection was 23−8−8−1 at 0, 4, and 25 °C. After MRMR selection, the best topology for the MRMR-GDEDBO-BPNN model with the lowest error at all three temperatures was 7−5−1 (as shown in Figure 6).
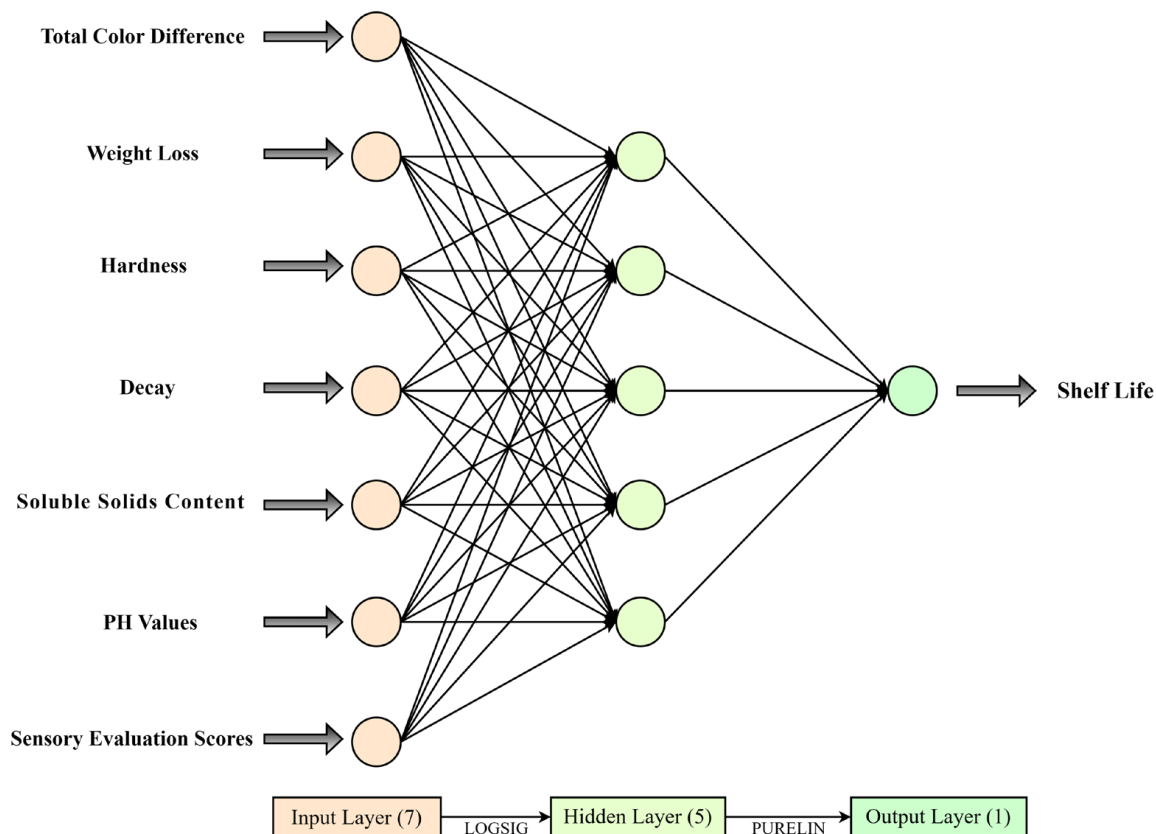

Figure 6. Schematic representation of the optimal topology of the MRMR-GDEDBO-BPNN model for predicting the shelf life of blueberries at different storage temperatures.

Network Training

The network training parameters included an upper limit of 100 iterations, a performance target error of 0.0001, a population size of 50, a learning rate of 0.01, and a momentum constant with a default value of 0.9.

4.2. Model Evaluation

4.2.1. K-Fold Cross-Validation

K-fold cross-validation is a widely used technique for model evaluation and parameter tuning. It can reduce the variability of data splitting and enhance the generalization ability of the model [24]. This paper used a 10-fold cross-validation method to build the model, i.e., the data was divided into 10 subsets. The cross-validation was repeated in 10 rounds. In each round, the data were randomly split into a training set and a test set. The model was trained on the training set and validated on the test set. The model performance metrics (e.g., MSE values) were averaged over 10 rounds as an estimate of the model's prediction accuracy. The optimal results are shown in Table 8, where the neuron configuration that minimized the model error at 0 °C for the BPNN model, for example, was a two-layer hidden layer of (6,7).

4.2.2. Evaluation Indicators

Statistical error is widely used to assess the predictive performance of a model. Common regression evaluation metrics include mean absolute error (*MAE*), mean squared

error (*MSE*), mean absolute percentage error (*MAPE*), and coefficient of determination (*R²*). *MAE* indicates the difference between the optimal value of the algorithm and the theoretical optimal value and reflects the exploration capability and convergence accuracy of the algorithm. *MAPE* measures the relative error between predicted and true values and is expressed as a percentage. It is useful for comparing models with data at different scales. *R²* measures how well the model fits the data; as it approaches 1, the better the model fits, and conversely, as it approaches 0, the worse the fit. The formula is as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|Y_i - Z_i|, \tag{24}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Y_i - Z_i)^2, \tag{25}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{Y_i - Z_i}{Y_i}\right| \times 100\%, \tag{26}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - Z_i)^2}{\sum_{i=1}^{n}\left(Y_i - Y_i\right)^2}, \tag{27}$$

where Y_i and Z_i denote the true and predicted values, respectively.

5. Results

The critical factors affecting the shelf life of blueberries filtered by the MRMR algorithm in this paper were: total color difference, weight loss, hardness, decay rate, soluble solids content, pH, and sensory evaluation score. Figures 7–11 show the trends of these seven quality indicators with storage time at different storage temperatures.
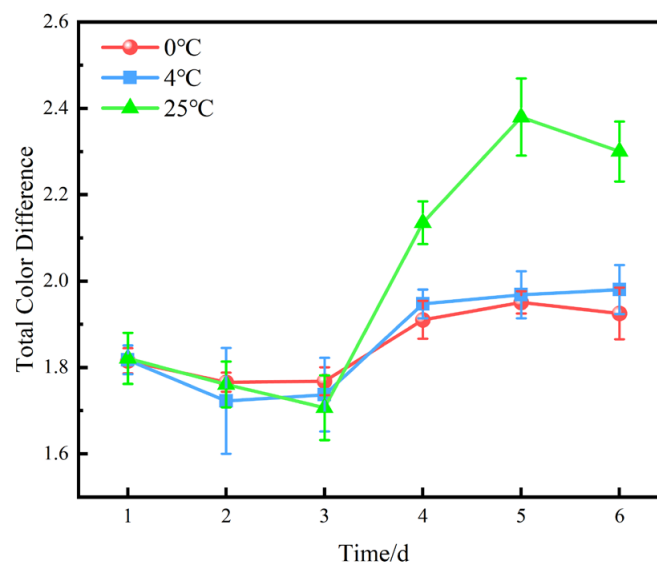
Figure 7. Variations in total color difference in blueberries at different storage temperatures.
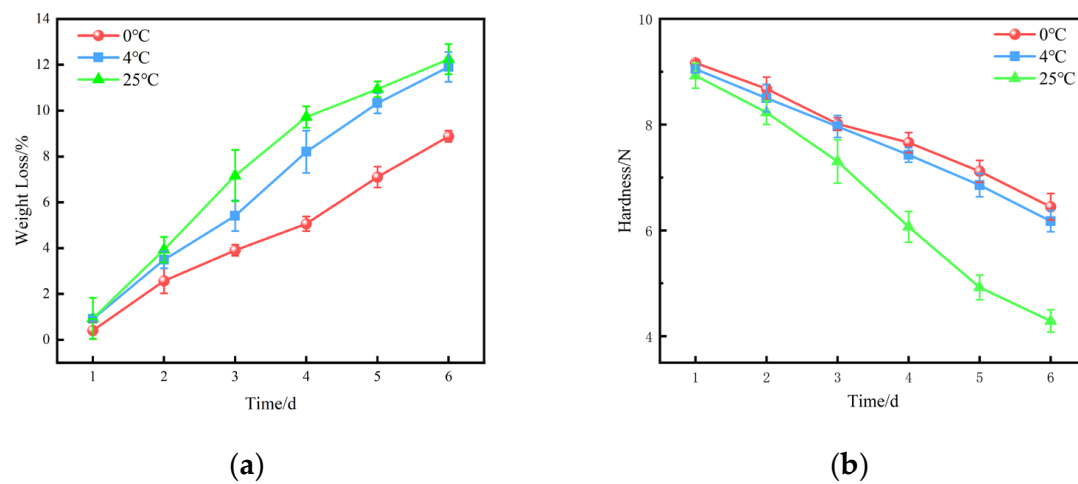
Figure 8. Changes in weight loss and hardness of blueberries at different storage temperatures: (**a**) weight loss and (**b**) hardness.
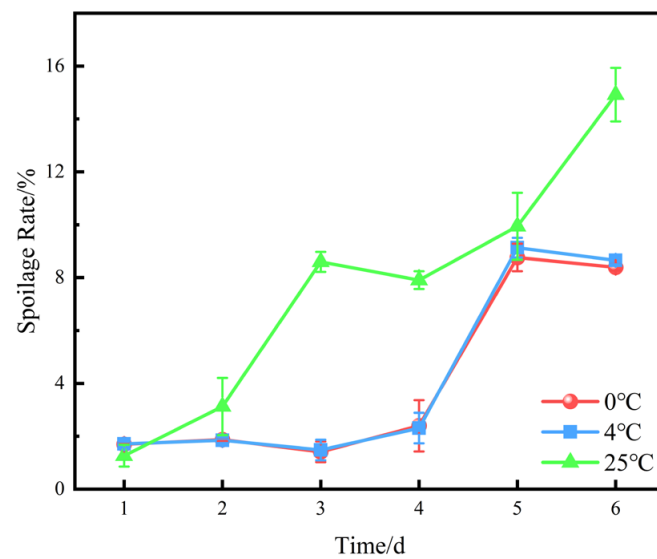
Figure 9. Variations in spoilage rate of blueberries at different storage temperatures.
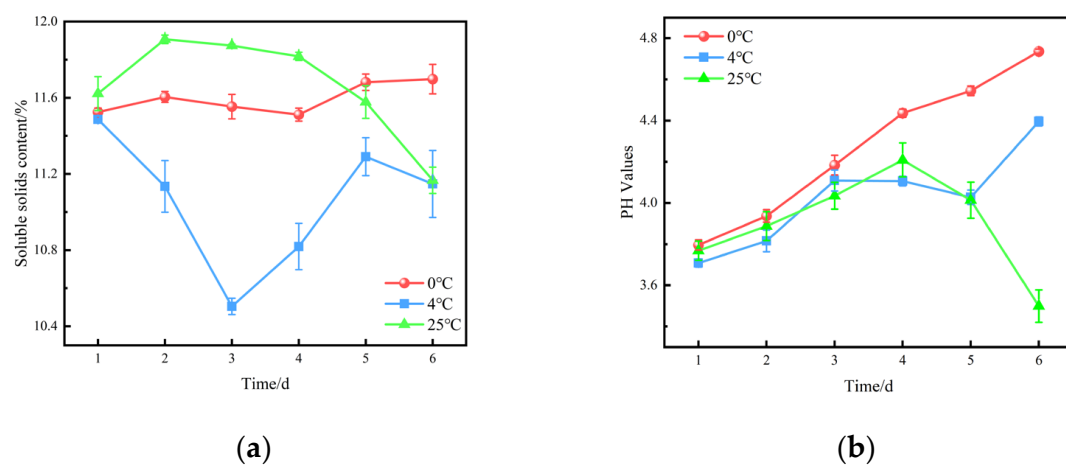
Figure 10. Changes in soluble solids and pH of blueberries at different storage temperatures: (**a**) soluble solids content and (**b**) pH values.
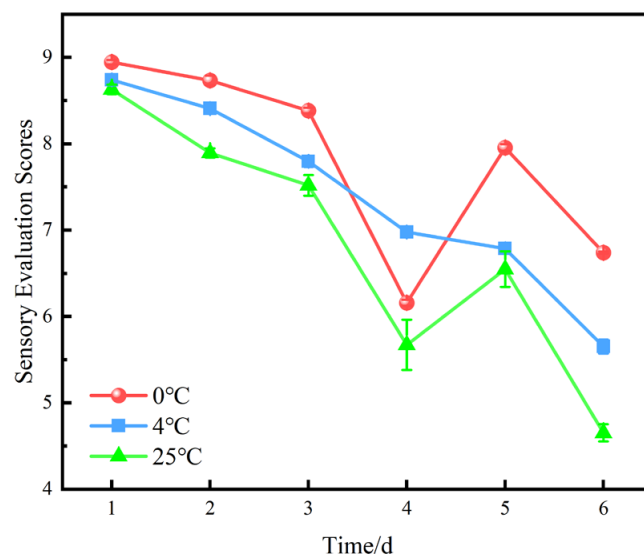
Figure 11. Changes in sensory evaluation scores for blueberries at different storage temperatures.

5.1. Trends in Quality Indicators of Blueberries at Different Storage Temperatures

5.1.1. Total Color Difference

Anthocyanins are the main pigments that determine the color of blueberry fruit, which is important for both aesthetic and nutritional value. However, anthocyanins are susceptible to degradation during storage due to various factors, such as condensation with proanthocyanidins, enzymatic oxidation by peroxidases and polyphenol oxidases, and binding to macromolecules (e.g., proteins and cell wall polysaccharides) to form precipitates [25]. Anthocyanin degradation can result in a color change from blue-purple to red, which may reduce consumer acceptance. Therefore, it is of interest to investigate the color characteristics of blueberry fruit stored at different temperatures and to develop effective preservation strategies to retain anthocyanins as much as possible.

Total color difference (ΔE) is a common colorimetric parameter to describe the color variation of fruits. A ΔE value less than 1.0 indicates an insignificant color difference; a value less than 2.0 indicates a detectable color difference by trained observers; and a value above 3.5 indicates a noticeable color difference by average consumers [26]. As shown in Figure 7, ΔE increased after 3 d of storage at 0, 4, and 25 °C, with a similar trend at 0 °C and 4 °C and a larger increase at 25 °C. This may be attributed to the higher anthocyanin retention at low temperature storage [27]. However, this color change was not statistically significant and was not easily perceived by the average consumer (maximum $\Delta E < 2.5$), as shown in Figure 7.

5.1.2. Weight Loss and Hardness

Figure 8 shows the effects of storage time and temperature on weight loss and fruit hardness in blueberries. During storage, blueberries undergo respiration and transpiration processes that result in moisture loss, macromolecule degradation, mass reduction, weight loss increase, and fruit hardness decrease. The hardness of blueberries stored at 0, 4, and 25 °C decreased by 32.68%, 34.53%, and 55.36%, respectively, after 6 days of storage, indicating that lower storage temperatures could better preserve the fruit's firmness and that 0 °C was more effective than 4 °C in delaying fruit softening.

5.1.3. Spoilage Rate

Figure 9 shows the spoilage rates of blueberries stored at 0 °C, 4 °C, and 25 °C. At 0 °C and 4 °C, the spoilage rates were similar and remained stable for the first 3 days, followed by a sharp increase after day 4. At 25 °C, the spoilage rate increased rapidly from day 2 and reached a plateau for the last 3 days. This could be attributed to the effects of temperature and confined space on microbial growth and metabolic activity. The spoilage rates at day 6

were 7.66%, 7.92%, and 16.44% for 0 °C, 4 °C, and 25 °C, respectively. These results indicate that low temperatures can delay spoilage and extend the shelf life of blueberries, and that 0 °C is more effective than 4 °C in inhibiting spoilage.

5.1.4. Soluble Solids Content and pH Values

Figure 10 shows that the SSC of blueberries increased significantly after 2 days of storage at 0 °C and 4 °C. This could be attributed to the degradation of polysaccharides and the conversion of organic acids, which enhanced the sweetness of the fruit. At 25 °C, the SSC increased rapidly after 3 days of storage, which might reflect the different ripening stages of the fruit at harvest [28]. Fluctuations in SSC and pH values were observed at all storage temperatures, especially at 25 °C. These fluctuations could be related to changes in fruit metabolism induced by respiration, which is more intense at higher temperatures. These results indicate that low-temperature storage can extend the shelf life of blueberries by reducing their respiration rate.

5.1.5. Sensory Evaluation Scores

Sensory evaluation is a subjective measure of human perception of changes in blueberry quality. Figure 11 shows that the overall sensory score of blueberries was the lowest at 25 °C. This was consistent with Table 2, which showed that blueberries stored at 25 °C for 6 days had a dull and lusterless appearance, severe quality loss, and a 50% decay area. The sensory scores of blueberries were higher at 0 °C and 4 °C, with the highest score at 0 °C.

5.2. Pearson's Correlation Analysis

The quality attributes of blueberries changed significantly during storage, and these changes were correlated with each other and affected the shelf life of the fruit. Pearson correlation analysis was performed on the seven key factors selected by the MRMR algorithm and the storage time, as shown in Figure 12 (sorted by hclust hierarchical clustering order). Figure 12 shows that there was no significant correlation ($p < 0.01$) between soluble solids content and other factors at 4 °C or between pH value and other factors at 25 °C. At all storage temperatures, weight loss and firmness had the highest correlation with shelf life. The correlation analysis results indicate that the shelf-life prediction model based on quality indicators and sensory evaluation scores of blueberries is reliable.

5.3. Model Performance Comparison Analysis

The prediction performance of each model on the test set is shown in Table 9, from which it can be seen that (taking 0 °C as an example):

(1) Compared with the benchmark model, the MRMR-GDEDBO-BPNN model proposed in this paper performed the best, with its MAE, MSE, MAPE, and R^2 of 0.0304, 0.0014, 1.1379%, and 0.9995, respectively, significantly outperforming the other models;

(2) Compared with the MLR, ANN, and SVR models, the BPNN model reduced the MAE, MSE, and MAPE values by 56.13–60.02%, 99.54–99.60%, and 35.68–51.71%, respectively. Compared with MRMR-MLR, MRMR-ANN, and MRMR-SVR models, the MRMR-BPNN model reduced the MAE, MSE, and MAPE values by 74.83–78.25%, 99.62–99.69%, and 39.47–43.38%, respectively. These results indicate that the BPNN model had better prediction performance than other models, regardless of feature selection, indicating that the BPNN model has a stronger fitting and generalization ability for the dataset in this paper. The possible reasons are as follows: The MLR model may fail to capture the nonlinear features and complex relationships of the data, leading to large prediction errors [29]. The BPNN model can fit the data more accurately than the ANN model because the BPNN model uses the structure of a multilayer perceptron to learn the high-order features and complex relationships of the data, while the ANN model only uses the structure of a single-layer perceptron. The SVR model is a regression model based on support vector machines, which seeks

the optimal hyperplane by maximizing the margin, but the data may not be linearly separable in reality, and the SVR model may not capture the true pattern of the data, resulting in large prediction errors [30];

(3) Compared with the conventional BPNN model, the DBO-BPNN model decreased the MAE and MAPE values by 10.88% and 29.81%, respectively, and increased the R^2 value by 1.09%. This indicates that optimizing the weights and thresholds of BPNN with DBO can significantly improve the prediction performance of the neural network. Compared with MRMR-BPNN, the MAE, MSE, and MAPE values of the MRMR-GDEDBO-BPNN model decreased by 74.11%, 26.32%, and 88.40%, respectively, and increased the R^2 value by 2.25%, further demonstrating the importance of optimizing the weights and thresholds of BPNN;

(4) Compared with DBO-BPNN, the GDEDBO-BPNN model decreased the MAE, MSE, and MAPE values by 18.68%, 35.14%, and 21.25%, respectively, indicating that the prediction results of the GDEDBO-BPNN model were closer to the actual values and verifying the effectiveness of the improved strategy proposed in this study;

(5) Compared with MLR, ANN, SVR, and BPNN, after feature selection, their MAE, MSE, and MAPE values decreased by 7.20–49.51%, 4.70–32.14%, and 3.34–20.61%, respectively. Compared with GDEDBO-BPNN, the MAE, MSE, and MAPE values of the MRMR-GDEDBO-BPNN model after feature selection decreased by 81.96%, 96.55%, and 79.72%, respectively, and the R^2 value increased by 1.15%. This indicates that feature selection can improve the prediction accuracy of the model by effectively reducing redundant features.
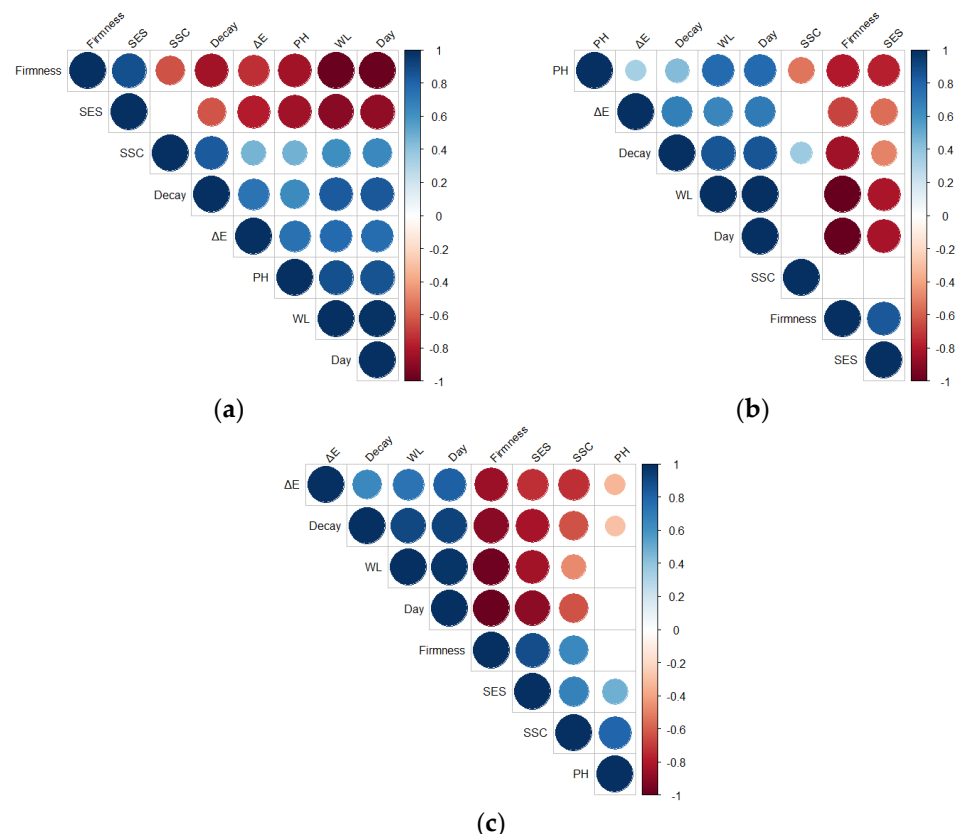

Figure 12. Matrix for correlation analysis of quality indicators of blueberries at different storage temperatures ($p < 0.01$): (**a**) 0 °C; (**b**) 4 °C; and (**c**) 25 °C; the size and color of the circles were positively correlated with the correlation coefficient, i.e., larger circles and darker colors indicated higher correlation coefficients, and vice versa.

Table 9. Comparison of the performance of different benchmark models for predicting the shelf life of blueberries at different storage temperatures.

Model	25 °C				4 °C				0 °C			
	MAE	MSE	MAPE	R^2	MAE	MSE	MAPE	R^2	MAE	MSE	MAPE	R^2
MLR	0.6088	0.7353	22.3812%	0.8588	0.6452	0.6717	21.0512%	0.8602	0.5816	0.7081	21.0245%	0.8716
ANN	0.5807	0.6420	19.2673%	0.8607	0.5650	0.6263	17.6226%	0.8777	0.5493	0.6106	17.5959%	0.8947
SVR	0.5738	0.6589	18.4095%	0.8771	0.5519	0.637	16.1159%	0.8929	0.5300	0.6151	17.3324%	0.9087
MRMR-MLR	0.5673	0.6505	18.8400%	0.8689	0.5585	0.6317	16.8693%	0.8853	0.5397	0.6129	16.6905%	0.9017
MRMR-ANN	0.5507	0.6420	16.2673%	0.8807	0.5071	0.6037	16.2406%	0.8851	0.5027	0.5819	16.2139%	0.8895
MRMR-SVR	0.5349	0.5654	18.0642%	0.8904	0.5007	0.5312	17.3591%	0.9013	0.4665	0.4970	15.7850%	0.9122
BPNN	0.2862	0.0049	15.3267%	0.9516	0.2226	0.0035	13.9958%	0.9630	0.2325	0.0028	10.1526%	0.9721
MRMR-BPNN	0.1955	0.0084	13.9324%	0.9574	0.1798	0.0026	12.2877%	0.9744	0.1174	0.0019	9.8137%	0.9775
DBO-BPNN	0.2259	0.0906	10.7300%	0.9639	0.2040	0.0611	8.4364%	0.9797	0.2072	0.0626	7.1264%	0.9827
MRMR-DBO-BPNN	0.1248	0.0548	5.0646%	0.9793	0.1183	0.0401	5.0379%	0.9837	0.1158	0.0401	3.5118%	0.9893
GDEDBO-BPNN	0.1313	0.0737	6.4364%	0.9724	0.2071	0.0631	5.7313%	0.9833	0.1685	0.0406	5.6121%	0.9881
MRMR-GDEDBO-BPNN	0.1013	0.0253	3.0083%	0.9910	0.0671	0.0077	2.3654%	0.9971	0.0304	0.0014	1.1379%	0.9995

The network performance and regression analysis graphs of the MRMR-GDEDBO-BPNN configuration after training during storage at 0, 4, and 25 °C are shown in Figure 13a–c, respectively. As shown in Figure 13, the best validation performance at 0 °C and 4 °C was achieved at the 32nd and 17th iterations, respectively, with MSE values very close to 0. At 25 °C, although the number of iterations was less than at 0 °C, the error was also larger, with an MSE value of approximately 0.0004. The regression analysis graphs reflect the fit between the predicted values and the actual values of the data. The *R* values of all four regression lines were very close to one at all three temperatures, indicating a very good consistency between the actual values and the predicted values, which ensured the good generalization ability of the improved neural network model proposed in this study.

Figure 14a–c show the predicted values of the three neural network models BP, GDEDBO-BP, and MRMR-GDEDBO-BP for the shelf life of blueberries at three storage temperatures of 0, 4, and 25 °C compared with the actual values, respectively (the prediction results are shown in Table 10. As shown in Figure 14, compared with the original BP neural network, the predicted values of the BP neural network model after MRMR feature selection and GDEDBO optimization were closer to the actual values and almost completely coincided with the actual value curve, proving the superiority of the MRMR-GDEDBO-BPNN model proposed in this paper in predicting blueberry shelf life.

(**a**)
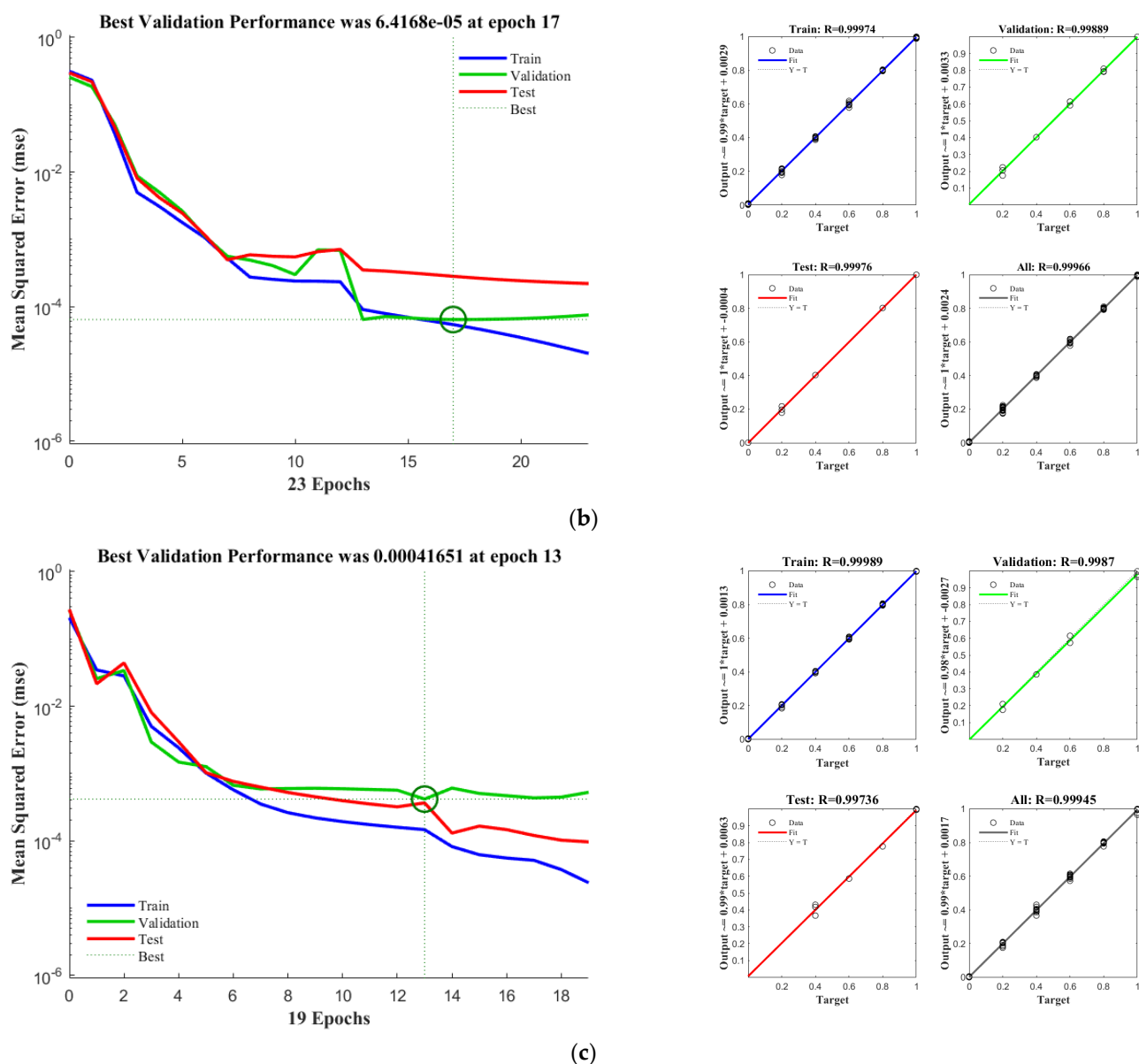
Figure 13. *Cont*.

(b)

(c)
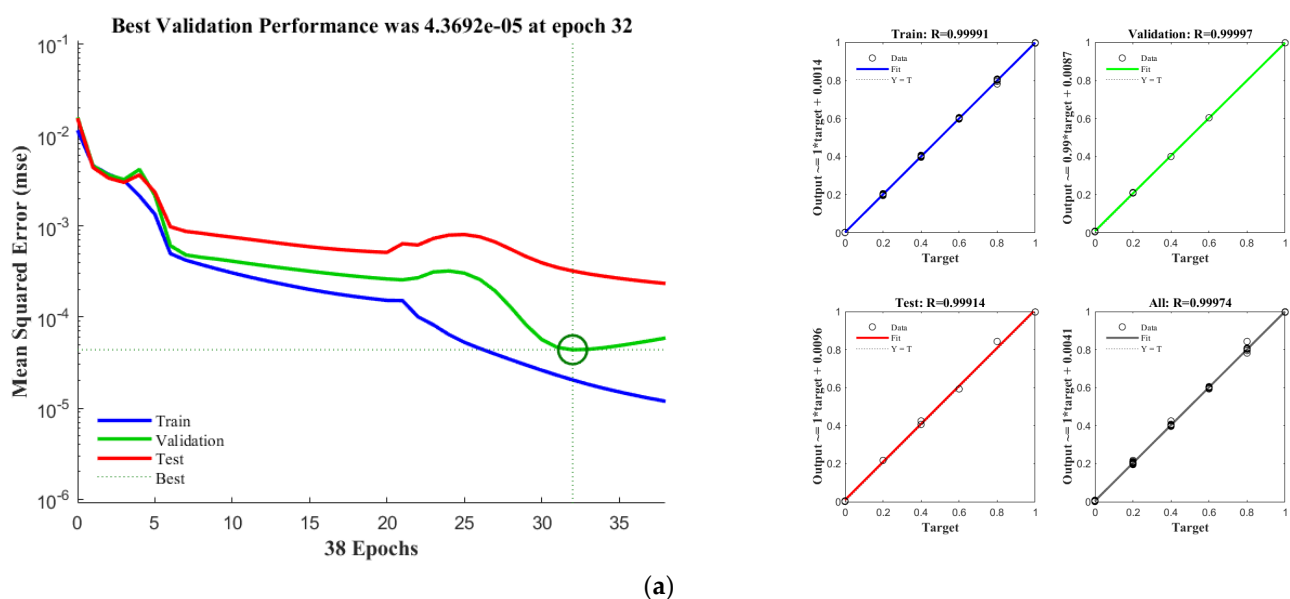
Figure 13. Network performance plots and regression analysis of the MRMR-IDBO-BPNN model for predicting the shelf life of blueberries at different storage temperatures: (**a**) 0 °C; (**b**) 4 °C; and (**c**) 25 °C.

Table 10. Shelf-life prediction of blueberries at different storage temperatures and baseline models.

Actual Shelf-Life/d	Predicted Shelf-Life/d	0 °C		4 °C			25 °C		
	BP	GDEDBO-BP	MRMR-GDEDBO-BP	BP	GDEDBO-BP	MRMR-GDEDBO-BP	BP	GDEDBO-BP	MRMR-GDEDBO-BP
1	1.1285	1.0238	1.0049	1.2882	1.0759	1.0095	1.3616	1.3128	1.0112
2	1.9468	2.0151	1.9926	1.8944	2.1021	1.9359	1.8428	2.1134	2.0654
3	3.2360	3.1095	3.0223	3.0919	2.9635	3.0429	2.8186	2.9526	3.0092
4	4.0608	3.9774	3.9812	4.1867	3.8431	4.0615	4.1677	3.9444	3.9759
5	5.2478	5.1979	4.9893	4.8434	5.1551	5.0225	5.2524	5.1477	5.1123
6	6.2575	5.9098	5.9638	5.7328	5.9729	5.9884	5.8174	5.8207	5.9867
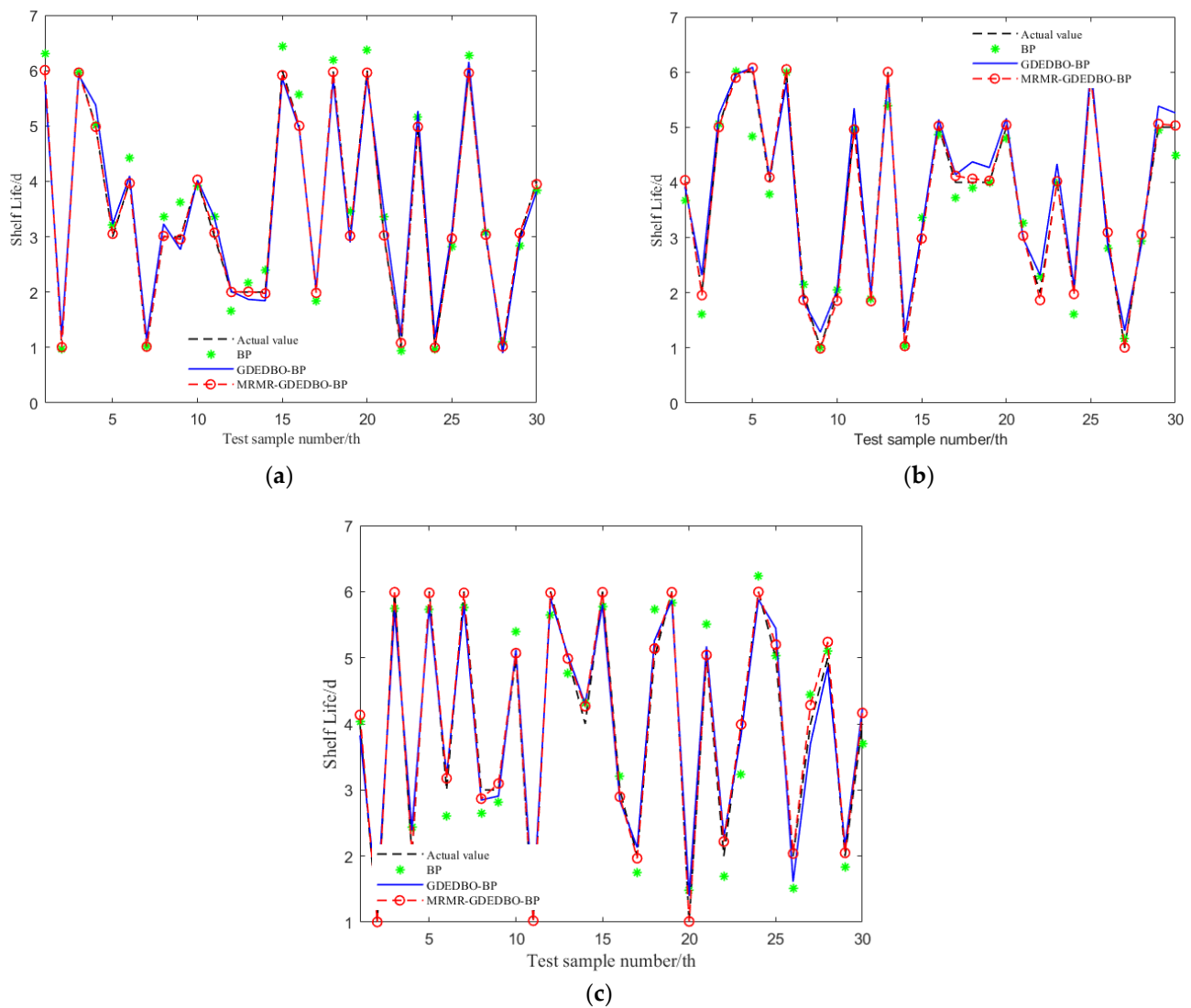
Figure 14. Comparison of predicted and actual values of blueberry shelf life at different storage temperatures by different benchmark models: (**a**) 0 °C; (**b**) 4 °C; and (**c**) 25 °C.

As shown in Table 10, the average error and prediction accuracy of the three neural network models, BP, GDEDBO-BP, and MRMR-GDEDBO-BP, at 0, 4, and 25 °C were 5.69% and 94.31%, 2.14% and 97.86%, 0.48% and 99.52%, 8.24% and 91.76%, 3.56% and 96.44%, 1.29% and 98.71%, 10.39% and 89.61%, 7.64% and 92.36%, and 2.96% and 97.04%, respectively.

The comparison of the prediction results of three neural network models shows that:

(1) The prediction accuracy of the MRMR-GDEDBO-BP model for blueberry shelf life was the highest, with a prediction error within 3.27%, meeting the needs of shelf-life prediction;

(2) The prediction error increased with the increase in blueberry storage temperature, which may be due to the fact that the higher the temperature, the stronger the respiration and transpiration of blueberries, resulting in quality decline and shelf-life reduction acceleration, thus increasing the model error. It can be concluded that 0 °C is the best storage temperature for blueberries during their shelf life.

6. Discussion

Shelf life is a complex trait that is greatly influenced by genotype and variety. This study used 'Liberty' blueberry as the research object, which was obtained by crossing 'Brigita' and 'Eliot'. Both 'Brigita' and 'Eliot' are northern highbush blueberry varieties, which have characteristics such as cold resistance, large and sweet fruits, and a long

maturation period [31]. 'Liberty' blueberries inherited their excellent characteristics and had advantages such as a tall tree body, uniform fruit size, and strong resistance to diseases and pests [25]. These characteristics make 'Liberty' blueberries suitable for long-term storage and transportation and provide a good basis for shelf-life prediction. However, the shelf-life prediction problem of 'Liberty' blueberries has not been well studied before. Therefore, this paper proposed a novel shelf-life prediction model based on MRMR feature selection and the GDEDBO-BPNN algorithm and achieved high prediction accuracy at three temperatures of 0, 4, and 25 °C. The proposed model outperformed other models or methods reported for different crops in previous studies.

The impact of different input indicators on the model's prediction accuracy may vary. Figure 12 shows the correlation analysis results of different indicators and shelf-life at 0, 4, and 25 °C. Weight loss and firmness had the highest correlations with shelf life at these temperatures, respectively. Thus, these two indicators could predict the shelf life of blueberries more accurately than others. This agreed with Huang et al., who built blueberry freshness prediction models based on gas information and machine learning algorithms (BP, RBF, SVM, and ELM) at different temperatures and reported that SVM achieved the highest prediction accuracy of 94.01% [8]. However, the proposed model in this paper had a higher prediction accuracy of 96.73%, which might be attributed to the genotype differences that affected the quality change patterns and shelf-life of blueberries and consequently the model's performance.

Owoyemi et al. applied four models (MLR, SVR, RF, and XGBoost) to predict the shelf-life of 'Rustenburg' navel oranges at different temperatures and found that XGBoost performed the best with RMSE and R^2 of 0.195 and 0.914, respectively, on the whole dataset [30]. Zhang et al. used two methods (PLS and ANN) to establish the shelf-life prediction models of postharvest apples at 4 °C and 20 °C and found that ANN outperformed PLS with the optimal RMSE and R^2 of 3.51 and 0.991, respectively, under multiple indicators [6]. However, the proposed model in this paper had a higher prediction accuracy and achieved the optimal RMSE and R^2 of 0.037 and 0.999, respectively, on the test set. The possible reasons for this are as follows:

(1) Species differences. The quality change patterns and shelf life of fruits are influenced by their species characteristics, such as moisture content, antioxidant capacity, cell wall structure, sugar-acid ratio, etc. Generally speaking, fruits with higher moisture content, stronger antioxidant capacity, more stable cell wall structure, and a moderate sugar-acid ratio have better quality and a longer shelf life. According to previous studies, 'Liberty' blueberry is a high-quality variety with high moisture content (about 85%), strong antioxidant capacity (about 13.8 mmol/100 g), stable cell wall structure (about 0.5%), and a moderate sugar-acid ratio (about 15.6) [25]. However, 'Rustenburg' navel orange is a low-quality variety with low moisture content (about 60%), weak antioxidant capacity (about 2.4 mmol/100 g), unstable cell wall structure (about 1.2%), and a high sugar-acid ratio (about 25.4) [30];

(2) Improper feature selection. Owoyemi et al. did not perform feature selection and might have redundant features. Zhang et al. used PLS to select features, but this method only considered the linear correlation between features and responses without considering the redundancy and nonlinearity among features. However, the MRMR algorithm used in this paper could simultaneously consider relevance, redundancy, and nonlinearity among features. Therefore, it obtained higher prediction accuracy;

(3) Limitations of the models themselves. Owoyemi et al.'s XGBoost model had problems such as sensitivity to noise and outliers, overfitting, tedious hyperparameter tuning, etc., which affected its generalization ability. Zhang et al. used PLS to build the quality change and shelf-life prediction models of apples, but PLS assumed a linear relationship between features and responses, while there might be a nonlinear relationship in reality. They also used ANN to optimize the parameters of the PLS model, but ANN was prone to falling into local optima and was sensitive to parameter selection. However, the BPNN model used in this paper could handle nonlinear and

high-dimensional data and used the GDEDBO algorithm to optimize its parameters to avoid falling into local optima and minimize prediction error. Therefore, it obtained more satisfactory prediction results.

Although the MRMR-GDEDBO-BPNN model proposed in this paper showed good predictive performance for the shelf life prediction of blueberries, there are still some aspects that deserve further discussion. For example, feature selection using the MRMR algorithm is independent of the prediction model, and there is no guarantee that the selected features are necessarily optimal in a given prediction model. In future research, attempts can be made to combine the MRMR algorithm with the embedding method to incorporate the prediction model into the feature selection process.

7. Conclusions

To address the problems of complex and multiple influencing factors on blueberry shelf-life prediction, the low training speed of traditional BPNN, and the possible local optimum of the original DBO algorithm, this paper proposes a blueberry shelf-life prediction model based on MRMR-GDEDBO-BPNN. The main conclusions of this paper are as follows:

(1) Feature selection helps improve the prediction accuracy of shelf-life models;

(2) Optimizing the weights and thresholds of the BPNN with DBO helps enhance the prediction performance of the neural network;

(3) To overcome the limitations of the original DBO algorithm, this paper proposes the GDEDBO model. It uses tent chaotic mapping to initialize the population and increase its diversity. It enhances the diversity of leaders by introducing an elite pool strategy to improve the exploration performance of the algorithm. Moreover, it adjusts the search direction by means of a Gaussian distribution estimation strategy, effectively coordinating the algorithm's global and local search capabilities and strengthening the algorithm's late-stage optimization search capability. The GDEDBO is evaluated with four meta-heuristic algorithms, including DBO, on 14 benchmark functions. The results show that GDEDBO exhibits good outperformance for both unimodal and multimodal functions, verifying the effectiveness of the improved strategy and being highly competitive with other meta-heuristic algorithms;

(4) This paper proposed the MRMR-GDEDBO-BPNN model, which uses seven critical factors (hardness, weight loss rate, spoilage rate, total color difference, soluble solids content, pH values, and sensory evaluation scores) filtered by the MRMR algorithm as input variables to establish shelf-life prediction models at three temperatures of 0 °C, 4 °C, and 25 °C, respectively. The results show that compared with the original BPNN, the values of MAE, MSE, and MAPE of the MRMR-GDEDBO-BPNN model are reduced by 86.92%, 50%, and 88.79%, respectively, and the value of R^2 is increased by 2.82%. This indicates that the BPNN model with feature selection by the MRMR algorithm and optimization by the GDEDBO model significantly improved the shelf-life prediction accuracy;

(5) This paper compares the optimal prediction models of shelf life at three storage temperatures, as shown in Table 9. It can be seen from Table 9 that the optimal prediction models at three temperatures are all MRMR-GDEDBO-BPNN models, whose optimal topological structures are all 7–5−1, and the activation function combinations are all LOGSIG-PURELIN. By comparing the prediction accuracy of models at three different temperatures, it can be concluded that the prediction accuracy of shelf-life models at 0 °C and 4 °C is higher than that at 25 °C, and 0 °C is higher than 4 °C. This may be due to the fact that the higher the temperature, the stronger the respiration and transpiration of blueberries, which leads to a faster rate of quality decline and shelf-life shortening, thus increasing the error of the prediction model. It can be concluded that 0 °C is the optimum storage temperature for blueberries during their shelf life;

(6) This paper compares the shelf life at three storage temperatures with a sensory evaluation score of six as the end point of shelf life. The results are shown in Figure 11.

It can be seen from Figure 11 that the shelf lives of this variety of blueberries at 0 °C, 4 °C, and 25 °C are 5.99 days, 5.35 days, and 3.67 days, respectively.

Shelf life is a complex trait that is greatly influenced by genotype and variety. Although the model proposed in this paper achieved high prediction accuracy at three temperatures of 0, 4, and 25 °C, we also recognize that our study has some limitations and shortcomings, such as using only a single Liberty variety of blueberries as the research object, which may cause variety bias by not considering the shelf-life prediction problem of other crops (such as fruits, vegetables, etc.). Therefore, we propose some future directions for further research:

(1) To train and test the proposed shelf-life prediction model on multiple varieties or genotypes of blueberries, to verify its applicability in evaluating other varieties or genotypes, and to compare its prediction performance with other models or methods;

(2) To extend the proposed shelf-life prediction model to predict the shelf-life of other crops (such as fruits, vegetables, etc.) and explore their similarities and differences among different crops.

References

1. Zhao, Y.; Zhu, X.; Hou, Y.; Pan, Y.; Shi, L.; Li, X. Effects of harvest maturity stage on postharvest quality of winter jujube (*Zizyphus jujuba* Mill. cv. Dongzao) fruit during cold storage. *Sci. Hortic.* **2021**, *277*, 109778. [CrossRef]
2. Matrose, N.A.; Obikeze, K.; Belay, Z.A.; Caleb, O.J. Plant extracts and other natural compounds as alternatives for post-harvest management of fruit fungal pathogens: A review. *Food Biosci.* **2021**, *41*, 100840. [CrossRef]
3. Han, J.W.; Zuo, M.; Zhu, W.Y.; Zuo, J.H.; Lü, E.L.; Yang, X.T. A comprehensive review of cold chain logistics for fresh agricultural products: Current status, challenges, and future trends. *Trends Food Sci. Technol.* **2021**, *109*, 536–551. [CrossRef]
4. Mesías, F.; Martín, A.; Hern'andez, A. Consumers' growing appetite for natural foods: Perceptions towards the use of natural preservatives in fresh fruit. *Food Res. Int.* **2021**, *150*, 110749. [CrossRef]
5. Wang, H.; Zheng, Y.; Shi, W. Comparison of Arrhenius model and artificial neuronal network for predicting quality changes of frozen tilapia (*Oreochromis niloticus*). *Food Chem.* **2022**, *372*, 131268. [CrossRef]
6. Zhang, Y.; Zhu, D.; Ren, X. Quality changes and shelf-life prediction model of postharvest apples using partial least squares and artificial neural network analysis. *Food Chem.* **2022**, *394*, 133526. [CrossRef]
7. Fan, X.; Jin, Z.; Liu, Y. Effects of super-chilling storage on shelf-life and quality indicators of Coregonus peled based on proteomics analysis. *Food Res. Int.* **2021**, *143*, 110229. [CrossRef]
8. Huang, W.; Wang, X.; Zhang, J. Improvement of blueberry freshness prediction based on machine learning and multi-source sensing in the Cold Chain Logistics. *Food Control* **2023**, *145*, 109496. [CrossRef]
9. Zhang, R.; Zhu, Y. Predicting the Mechanical Properties of Heat-Treated Woods Using Optimization-Algorithm-Based BPNN. *Forests* **2023**, *14*, 935. [CrossRef]
10. Xue, J.; Shen, B. Dung beetle optimizer: A new meta-heuristic algorithm for global optimization. *J. Supercomput.* **2022**, *79*, 7305–7336. [CrossRef]
11. Obsie, E.Y.; Qu, H.; Drummond, F. Wild blueberry yield prediction using a combination of computer simulation and machine learning algorithms. *Comput. Electron. Agric.* **2020**, *178*, 105778. [CrossRef]
12. Li, Y.; Chu, X.; Fu, Z. Shelf-life prediction model of postharvest table grape using optimized radial basis function (RBF) Neural Network. *Br. Food J.* **2019**, *121*, 2919–2936. [CrossRef]
13. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [CrossRef] [PubMed]

14. Rivera, S.; Giongo, L.; Cappai, F. Blueberry firmness—A review of the textural and mechanical properties used in quality evaluations. *Postharvest Biol. Technol.* **2022**, *192*, 112016. [CrossRef]

15. Yu, S.; Lan, H.; Li, X. Prediction method of shelf life of damaged Korla fragrant pears. *J. Food Process Eng.* **2021**, *44*, e13902. [CrossRef]

16. Lian, S.; Sun, J.; Wang, Z. A block cipher based on a suitable use of the chaotic standard map. *Chaos Solitons Fractals* **2005**, *26*, 117–129. [CrossRef]

17. Wu, Q. A self-adaptive embedded chaotic particle swarm optimization for parameters selection of Wv-SVM. *Expert Syst. Appl.* **2011**, *38*, 184–192. [CrossRef]

18. Ibrahim, B.A. A hybrid firefly and particle swarm optimization algorithm for computationally expensive numerical problems. *Appl. Soft Comput.* **2018**, *6*, 232–249.

19. Wang, X.F.; Zhao, H.; Han, T. A grey wolf optimizer using Gaussian estimation of distribution and its application in the multi-UAV multi-target urban tracking problem. *Appl. Soft Comput.* **2019**, *78*, 240–260. [CrossRef]

20. Heidari, A.A.; Mirjalili, S.; Faris, H. Harris hawks optimization: Algorithm and applications. *Future Gener. Comput. Syst.* **2019**, *97*, 849–872. [CrossRef]

21. Greff, K.; Srivastava, R.K.; Koutnik, J. LSTM: A Search Space Odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 2222–2232. [CrossRef]

22. Reed, R.; Marks, R.J., II. *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks*; MIT Press: Cambridge, UK, 1999.

23. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]

24. Wong, T.T.; Yeh, P.Y. Reliable Accuracy Estimates from k-Fold Cross Validation. *IEEE Trans. Knowl. Data Eng.* **2020**, *32*, 1586–1594. [CrossRef]

25. Dragišić Maksimović, J.; Milivojević, J.; Djekić, I.; Radivojević, D.; Veberič, R.; Mikulič Petkovšek, M. Changes in quality characteristics of fresh blueberries: Combined effect of cultivar and storage conditions. *J. Food Compos. Anal.* **2022**, *111*, 104597. [CrossRef]

26. Mokrzycki, W.S.; Tatol, M. Colour difference ΔE—A survey. *Mach. Graph. Vis.* **2011**, *20*, 383–412.

27. Singh, B.; Suri, K.; Shevkani, K. Enzymatic Browning of Fruit and Vegetables: A Review. In *Enzymes in Food Technology*; Springer: Singapore, 2018; pp. 63–78. [CrossRef]

28. Ates, U.; Islam, A.; Ozturk, B. Changes in Quality Traits and Phytochemical Components of Blueberry (*Vaccinium Corymbosum* Cv. Bluecrop) Fruit in Response to Postharvest Aloe Vera Treatment. *Int. J. Fruit Sci.* **2022**, *22*, 303–316. [CrossRef]

29. Ktenioudaki, A.; O'Donnell, C.P.; Emond, J.P. Blueberry Supply Chain: Critical steps impacting fruit quality and application of a boosted regression tree model to predict weight loss. *Postharvest Biol. Technol.* **2021**, *179*, 111590. [CrossRef]

30. Owoyemi, A.; Porat, R.; Lichtern, A. Large-scale, high-throughput phenotyping of the postharvest storage performance of 'rustenburg' navel oranges and the development of shelf-life prediction models. *Foods* **2022**, *11*, 1840. [CrossRef]

31. Piechowiak, T.; Grzelak-Błaszczyk, K.; Sójka, M.; Skóra, B.; Balawejder, M. Quality and antioxidant activity of highbush blueberry fruit coated with starch-based and gelatine-based film enriched with Cinnamon Oil. *Food Control* **2022**, *138*, 109015. [CrossRef]