Proceedings of the 3rd International Conference on Environmental Geoscience and Earth Ecology
DOI: 10.54254/2753-8818/94/2025.21327

Survival Prediction of Agricultural Trees by Optimizing Long and Short Term Memory Networks Based on Multi-head Attention Mechanism

Yifan Dai[1,a], Jie Zhou[1,a], Xingxing Bai[1,b], Jiangjing Hua[1], Haiyan Sun[1,c*]

[1]Economics and Management, Yunnan Agricultural University, Kunming, Yunnan, 650051, China
a. 1021466464@qq.com, b. 1761682182@qq.com, c. 37398691@qq.com
**corresponding author*

Abstract: In this study, long short-term memory network (LSTM) model was optimized based on multi-head attention mechanism to effectively predict the survival of agricultural trees. Based on the in-depth analysis of a large number of agricultural tree survival data, the corresponding prediction model was constructed. In the training phase, we analyzed the confusion matrix of the training set, and the results revealed that the accuracy of the model in predicting the survival of agricultural trees was as high as 99.74%. In addition, the performance on the independent test set is also very good, with an accuracy of 99.40%, although there is a slight decrease compared to the training set (0.34%), but both maintain a high accuracy level of more than 99%. This shows that the model has very high prediction accuracy and maintains good generalization ability across different data sets. Further, by drawing the ROC curve and calculating the AUC (area under the curve), the result is 0.9857, which fully reflects the strong ability of the model in the tree survival prediction task. An AUC value higher than 0.9 indicates that the model has a very low error rate on the classification task, which ensures the reliability of the prediction results. This study shows that the LSTM optimization model based on multi-head attention mechanism can provide a high-precision tree survival prediction tool for agricultural management, so as to help farmers make more effective decisions. This result not only has theoretical value, but also has important significance for the growth management and sustainable development of trees in practical application. Through this research, we expect to be able to contribute to the development of smart agriculture.

Keywords: Multi-head attention mechanism, Long and short term memory network, Agricultural tree survival prediction.

1. Introduction

Predicting the survival of agricultural trees is a key task in modern forestry and agricultural management. With the increasingly serious global climate change, planting patterns and crop growth are greatly affected, and how to effectively assess and predict the survivability of trees is particularly important [1]. The healthy growth of trees is not only related to the balance of ecosystems, but also affects the economic benefits of agriculture, soil conservation and biodiversity. Therefore, it is of

DOI: 10.54254/2753-8818/94/2025.21327

great significance to study the living condition of trees and its influencing factors for sustainable agricultural development and ecological protection.

In practice, although plant growth is affected by a variety of environmental and biological factors, such as climate, soil characteristics, pests and diseases, and management practices, traditional forecasting methods often rely on qualitative analysis and simple linear models, which are inadequate when dealing with complex ecosystems. Therefore, researchers urgently need to find more accurate and efficient forecasting tools to meet this challenge [2].

In recent years, machine learning, as a powerful data analysis technology, has gradually received attention in the agricultural field. Machine learning can automatically extract features and perform classification or regression through learning and pattern recognition of a large amount of data, thus playing an important role in tree survival prediction [3]. Compared to traditional methods, machine learning algorithms have the advantage of handling high-dimensional data and nonlinear relationships, which can help researchers capture potentially complex patterns and interactions. Machine learning algorithms can quickly and efficiently analyze multi-dimensional data from different sources, including climate data, soil properties, tree growth parameters, and historical growth records. The integrated use of these data can improve the accuracy of survival predictions. For example, algorithms such as random forest, support vector machine and neural network are widely used in the modeling of tree survival prediction, and the factors affecting tree survival can be comprehensively understood and quantified by building complex prediction models [4].

Machine learning enables self-optimization and adjustment of the model. With the continuous collection of new data, the model can be further trained and verified by incremental learning. This flexibility allows the model to adapt more quickly to environmental changes and maintain high prediction accuracy [5]. For example, when new climate patterns or soil properties emerge, machine learning models can quickly integrate this information to more accurately assess the survivability of trees and reduce the risk of miscalculation. Predictive models based on machine learning can conduct risk assessment and help agricultural managers develop more scientific management strategies. By identifying potential risk factors, the model can provide targeted recommendations for tree planting and management, including suitable variety selection, cultivation technology optimization and conservation management measures, so as to improve tree survival and economic benefits. This paper optimizes the long and short term memory network based on multi-head attention mechanism to predict the survival of agricultural trees.

2. Data set sources and data analysis

Using a publicly available dataset, the team conducted a factor block design field experiment involving 3,024 seedlings from 4 tree species, 7 soil sources (sterile isospecies, live isospecies, and 5 xenospecies) and forest understory light gradients (low, medium, and high). Over the course of a growing season, the team monitored seedling survival twice a week and randomly selected a subset of seedlings over three weeks to measure mycorrhizal colonization and phenolic, lignin, and NSC measurements. The data set recorded a number of survival environmental indicators of trees, including core, soil, adult worms, sterelessness, homology, fungi, soil fungi, AMF, EMF, phenolic substances, acid equivalent calculation, NSC, lignin, time and the final survival state Alive. Some data were selected for display [6], as shown in Table 1.

Table 1: Selected data sets.

Species	Core	Soil	Myco	AMF	Phenolics	Lignin	NSC	Time	Event	Alive
1	2017	1	1	22	-0.56	13.86	12.15	14	1	2
2	2017	2	2	15.82	5.19	20.52	19.29	115.5	0	1

DOI: 10.54254/2753-8818/94/2025.21327

Table 1: (continued).

3	2017	1	2	24.45	3.36	24.74	15.01	63	1	2
1	2016	1	1	22.23	-0.71	14.29	12.36	14	1	2
1	2017	1	1	21.15	-0.58	10.85	11.2	14	1	2
4	2016	3	1	35.29	0.3	10.8	13.79	24.5	1	2
2	2016	4	2	24	5.11	18.82	22.51	24.5	0	2
3	2017	5	2	4	3.43	25.22	14.81	24.5	0	2
3	2016	6	2	28.74	3.83	26.65	14.65	115.5	0	1
1	2016	4	1	14.16	-0.05	13.3	12.16	24.5	1	2
3	2017	1	2	24.45	3.36	24.74	15.01	115.5	0	1
2	2017	6	2	19.68	5.1	18.16	22.56	73.5	1	2
2	2016	7	2	11.14	5.05	21.93	17.75	73.5	1	2
2	2017	3	2	11.45	4.79	17.83	23.72	59.5	1	2

3. Method

3.1. Multi-head attention mechanism

The multi-head attention mechanism was originally proposed in the Transformer model, which is an important sequence modeling method in the field of deep learning and is widely used in tasks such as natural language processing and computer vision. The core idea is that through multiple sets of parallel attention mechanisms, the model can learn in parallel in different representation subspaces of input sequences, capturing the diversity and fine-grained characteristics of information. Traditional single-headed attention mechanisms may lead to insufficient importance of information when processing long sequences. Therefore, multi-headed attention mechanisms improve the expressiveness and flexibility of the model by focusing on different parts of the input in parallel [7]. The structure of the multi-head attention mechanism is shown in Figure 1.

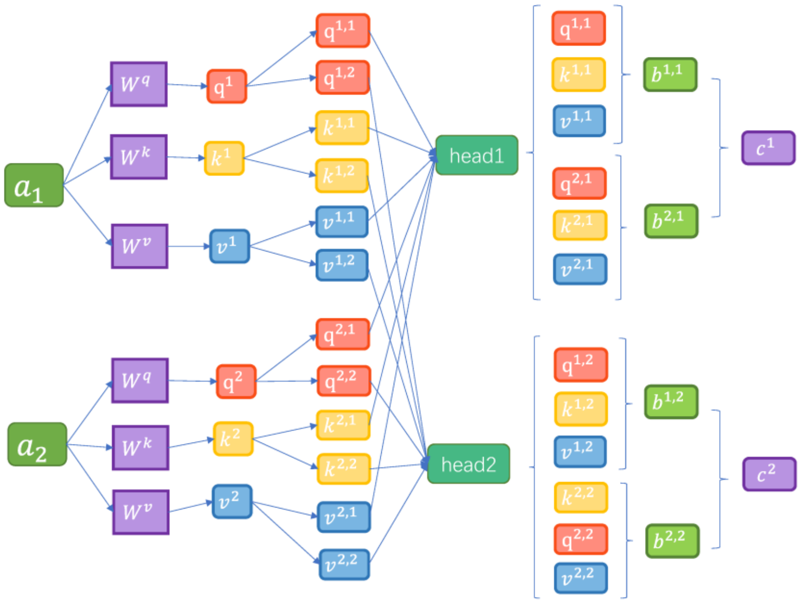

Figure 1: The structure of the multi-head attention mechanism.

DOI: 10.54254/2753-8818/94/2025.21327

The working process of the multi-head attention mechanism can be broken down into several steps. First, for each element in the input sequence, a linear transformation generates three vectors: Query, Key, and Value. The similarity score between the query and all keys is then calculated to determine the relationship of the query to the individual input positions [8]. By soft normalization of the similarity score (i.e., using the Softmax function), we can obtain a weighted coefficient for each value, achieve a weighted summation, and then obtain the corresponding attention vector. However, single-headed attention can only focus on a specific aspect of the input sequence, while multi-headed attention mechanisms can provide a richer representation by repeating this process many times (i.e., multiple heads), each of which can learn different patterns of attention and information, thus forming a Mosaic of multiple attention vectors.

Finally, the output of all the heads is spliced and the final result is obtained by linear transformation. This process not only improves the model's ability to analyze and learn from multiple perspectives, but also increases the ability to represent information abstractly, making the model better at understanding context and capturing long-distance dependencies. The introduction of multi-head attention mechanism solves many limitations in sequence modeling, and Transformer and its derivative models can achieve higher performance when dealing with complex tasks, becoming one of the mainstream methods in natural language processing and computer vision.

3.2. Long short-term memory network

Long short-term memory network (LSTM) is a special type of recurrent neural network (RNN) designed to solve the problems of gradient disappearance and gradient explosion encountered by traditional RNN when dealing with long sequences. LSTM enables the model to capture long distance dependencies by introducing a gating mechanism that effectively maintains and adjusts long term memories. The LSTM structure consists of a number of connected memory cells, each containing a cell state and three main gates: input gate, forget gate, and output gate. By controlling the inflow, retention and outflow of information, these gating mechanisms solve the memory limitation of RNN for long sequences of information to a certain extent [9]. The structure diagram of LSTM is shown in Figure 2.

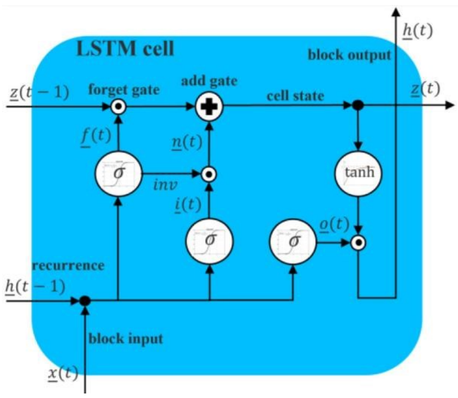

Figure 2: The structure diagram of LSTM.

At each time step of the LSTM, the value of the forget gate is first calculated based on the current input and the hidden state of the previous moment, which determines what information in the previous memory unit state should be forgotten. Specifically, it uses a Sigmoid activation function to map the previous hidden state and the current input to a value between 0 and 1, where 0 means "completely forgotten" and 1 means "completely retained." Then, the influence of the current input on the memory is controlled through the input gate, which is also generated by the Sigmoid activation function to

DOI: 10.54254/2753-8818/94/2025.21327

determine which new information will be added to the memory unit by combining the current input and the state of the previous memory unit [10]. Finally, the output gate determines the output of the next hidden state based on the current input and the previous hidden state, which will contain both the state information of the current memory unit and the prediction of the next state.

3.3. Improved long short-term memory network based on multi-head attention mechanism

The structure of the LSTM network will be changed after the integration of multiple attention mechanisms. In particular, the calculation of the output vector of the traditional LSTM unit will be integrated into the calculation of attention to form a new composite structure. At each time step, in addition to generating outputs of hidden states and memory units, attention mechanisms dynamically adjust these outputs based on the relevance of information inputs. First, for each element of the input sequence, the distribution of relationships with other elements can be obtained through the transformation of Query, Key and Value. This distribution of relationships allows the model to know to what extent it needs to focus on where in the sequence at each time step, thereby improving the effectiveness of information flow.

In a multi-head attention mechanism, multiple heads can calculate different attention weights in parallel, so that each head can focus on a different aspect of the data. This parallel mechanism enables LSTM to obtain other important context information while processing sequences, which effectively solves the problem of information loss in long sequence modeling. Specifically, the weights calculated by the attention mechanism can be directly combined with the hidden state of the LSTM, thus adjusting the expressive power of the LSTM in real time, making the model dynamically change its focus of attention according to the context in a short period of time. In this way, LSTM not only combines the information of memory units, but also integrates the dynamic selection ability of global information, thus improving the processing ability of complex sequence tasks.

4. Result

In terms of experimental parameter setting, the number of LSTM units is 128, the number of LSTM layers is 2, the activation function is tanh, the forgetting gate activation function is sigmoid, the number of multi-head attention mechanism heads is 4, the activation function is softmax, the optimizer is Adam, the learning rate is 0.001, and the loss function is cross entropy loss function. The training batch size was 64 and the number of training iterations was 150 EPOCs.

In terms of hardware Settings, the processor is Intel Core i7-9700K, the memory is 16GB, the graphics card is NVIDIA GeForce RTX 2080, the storage is 256GB SSD+1TB HDD, and the operating system is Windows 10.

The agricultural tree growth data set was introduced, the training set and test set were divided according to the ratio of 7:3, and the model was introduced for training. Output the confusion matrix of the predictions of the training set and the test set, as shown in Figure 3 and figure 4.

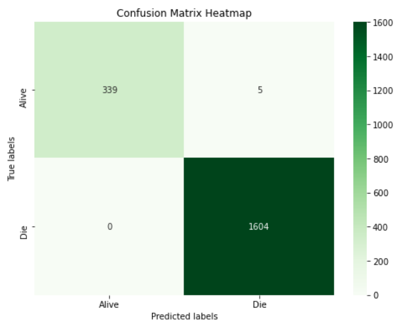

Figure 3: Confusion matrix of training set.

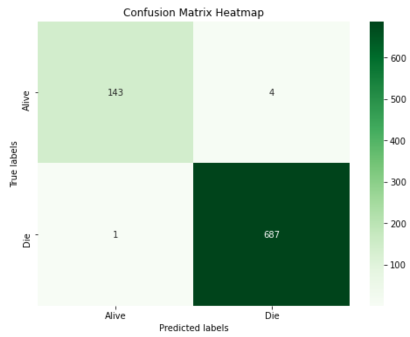

Figure 4: Confusion matrix of test set.

From the confusion matrix of the training set, it can be seen that the accuracy of this model in predicting the survival of agricultural trees is 99.74% and 99.40% in the test set. The accuracy of the test set is 0.34% lower than that of the training set, and the accuracy of both the training set and the test set is more than 99%, indicating that this model can predict the survival or death of trees very accurately. And the model also shows a very good effect on the test set, indicating that the model has a good generalization ability.

The ROC change curve of the test set is output, as shown in the figure. According to the ROC curve, the AUC predicted by the model was 0.9857, indicating that the model achieved a very good effect on tree survival prediction.

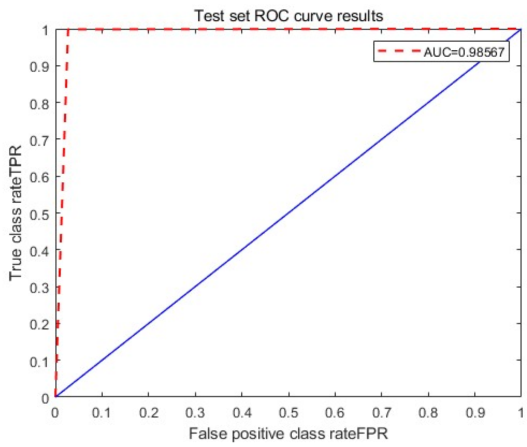

Figure 5: ROC curve.

A number of indicators were used to evaluate the prediction effect of the model, and the evaluation results were shown in Table 2.

Table 2: Model evaluation parameter.

Parameter	Value
Polygon areaPAM	0.98
Classification accuracy	0.99
sensitivity	1
specificity	0.97
Area under the curveAUC	0.99
Kappa coefficient	0.99
F measure	1

5. Conclusion

In this paper, an optimization model of long short-term memory network (LSTM) based on multi-head attention mechanism is proposed to improve the ability to predict the survival of agricultural trees. In the process of model training, we used large-scale data sets and performed a series of tuning and validation of the model. By analyzing the confusion matrix of the training set, the results show that the accuracy of this model in predicting the survival of agricultural trees reaches 99.74%. This high accuracy indicates that the model has a strong ability to distinguish between the survival and death of trees, and can basically achieve perfect classification. It is worth noting that the accuracy of the test set also reached 99.40%, although compared with the accuracy of the training set decreased slightly (0.34%), but still at a high level of more than 99%, fully indicating that the model has good prediction ability and strong application value.

To further analyze the predictive power of the model, we used the ROC curve to evaluate the performance of the model and calculated the AUC (area under the curve) to be 0.9857. This value further validates the superior performance of the model, indicating that the model not only has high accuracy, but also shows good discriminative ability in the task of predicting tree survival. The results of ROC curve and AUC value show that the error rate of the model is very low when dealing with classification problems, which is helpful to provide more reliable decision support in practical applications.

In summary, the LSTM optimization model based on multi-head attention mechanism constructed in this study, with its excellent accuracy and good generalization ability, provides a new idea and method for improving the accuracy of agricultural tree survival prediction.

References

[1] Zhou, Z., et al. "Improving the classification accuracy of fishes and invertebrates using residual convolutional neural networks." ICES Journal of Marine Science 80.5 (2023): 1256-1266.
*[2] Nandika, Muhammad Rizki, et al. "Assessing the Shallow Water Habitat Map** Extracted from High-Resolution Satellite Image with Multi Classification Algorithms." Geomatics and Environmental Engineering 17.2 (2023).*
[3] Lopez-Vazquez, Vanesa, et al. "Deep learning based deep-sea automatic image enhancement and animal species classification." Journal of Big Data 10.1 (2023): 37.
[4] Ahmad, Usman, et al. "Large Scale Fish Images Classification and Localization using Transfer Learning and Localization Aware CNN Architecture." Comput. Syst. Sci. Eng. 45.2 (2023): 2125-2140.
[5] Fan, Jianchao, and Chuan Liu. "Multitask gans for oil spill classification and semantic segmentation based on sar images." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 16 (2023): 2532-2546.
[6] Aziz, Rabia Musheer, et al. "CO-WOA: novel optimization approach for deep learning classification of fish image." Chemistry & Biodiversity 20.8 (2023): e202201123.
[7] Danqing Ma, , Shaojie Li, Bo Dang, Hengyi Zang, Xinqi Dong. "Fostc3net:A Lightweight YOLOv5 Based On the Network Structure Optimization." (2024).
[8] Zongqing Qi, , Danqing Ma, Jingyu Xu, Ao Xiang, Hedi Qu. "Improved YOLOv5 Based on Attention Mechanism and FasterNet for Foreign ObjectDetection on Railway and Airway tracks." (2024).
[9] Lu, Liangdong, Jia Xu, and Jiuchang Wei. "Understanding the effects of the textual complexity on government communication: Insights from China's online public service platform. "Telematics and Informatics 83 (2023): 102028.
[10] Kumari, Rina, et al. "Identifying multimodal misinformation leveraging novelty detection and emotion recognition." Journal of Intelligent Information Systems 61.3 (2023): 673-694.