# Prediction And Analysis Of Population Aging In Eight Ethnic Provinces Based On Machine Learning (SVR)

Hua Chen*

University of North Carolina at Chapel Hill, Department of Mathematics, North Carolina, USA

* Corresponding Author Email: chenhua7@ad.unc.edu

**Abstract.** This paper establishes SVR prediction model of aging coefficient for population over 65 years old in Inner Mongolia, Guangxi, Guizhou, Yunnan, Tibet, Qinghai, Ningxia and Xinjiang, respectively. Three error evaluation indexes, MSE,MAE and MAPE, are selected to test the accuracy of the model. The results show that the model established in this paper has high accuracy. And forecast for the next 9 years the provinces of people aged 65 and aging coefficient changes, the results show that the eight provinces over 65 population and aging coefficient of the overall upward trend, and the model of people aged 65 and forecasts trend of aging coefficient in the next decade, this also verify the rationality of the forecast results.

**Keywords:** Population forecast; Ethnic province; Machine learning; SVR.

## 1. Introduction

Population aging is an important strategic issue in China's economic, social and population development. This problem has been discussed in depth for a long time. The research on the prediction model of population aging in western developed countries started earlier, and the development is relatively stable. As far as China is concerned, China entered the aging society relatively late, starting to enter the aging society in 1999. In 2013, zhang rongyan, ma yanqin and hao shushuang established the GM (1,1) model and the residual grey prediction model by using the grey prediction theory to predict the population aging situation in China from 2012 to 2020.[1-2] In 2013, Xie Xiaojiang predicted nanjing's population status, total population and other factors, providing specific data such as the scale, structure and distribution of future population changes for nanjing's relevant departments to formulate development plans. In 2015, according to the modeling mechanism of GM(1,1) model, ren dawei established a grey prediction model of the number and proportion of the elderly population in China, and proposed to gradually extend the working life and raise the retirement age to solve the problem of labor shortage.[3-4]

For the choice of prediction model, Zhang Zhenhua takes Yantai City as an example and uses the grey prediction model to predict; Zhang Rongyan and others believe that the residual grey prediction model is better than the original grey prediction model in predicting the level of aging; Taking Hunan Province as an example, Chen Yihua and others used RBF neural network model to predict, and proved that its prediction effect was better than linear regression model; Taking Jiangsu Province as an example, Huang Jianyuan made a prediction based on Leslie matrix equation and analyzed its many characteristics; [5-6]The prediction model selected by Chen Guanghui and others is a nonparametric autoregressive model. While proving that it is more effective than ARIMA model, it puts forward relevant policy suggestions on the current situation of aging in China. [7-8]Although a single model can better predict the trend of population aging, different models have their own advantages and disadvantages in prediction. [9-10]In order to give full play to the advantages of a single model in prediction and weaken its disadvantages, each single model can be given a certain weight through certain criteria, and then build a combined prediction model. The combined prediction model has always been the focus of domestic and foreign scholars. Scholars have discussed the applicability of the model for different problems, and proved that the combined prediction model is indeed better than the single model in prediction accuracy.

At present, there is still little research on the population aging prediction of large minority provinces in China. This paper uses SVR machine learning algorithm to predict the population aging of large

minority provinces such as Inner Mongolia, Guangxi, Guizhou, Yunnan, Tibet, Qinghai, Ningxia and Xinjiang, with the prediction time span from 2022 to 2030. The error degree of the prediction results is analyzed, and the analysis results show the accuracy of the prediction. The research results play a guiding role in the population aging policy and the aging development trend of ethnic provinces.

## 2. Data source and data processing

The basic data of this paper are the basic data of the population over 65 years old in Inner Mongolia, Guangxi, Guizhou, Yunnan, Tibet, Qinghai, Ningxia and Xinjiang from 1988 to 2020. The data distribution is shown in the figure below.

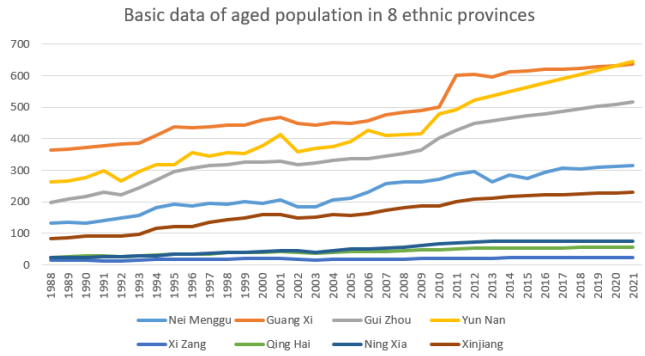



**Figure 1.** Basic data of aged population in 8 ethnic provinces

Before model training, normalize the data. In this paper, mapminmax function is selected for data preprocessing. The formula is as follows:

$$y_i = \frac{x_i - \min\limits_{1<j<n}\{x_j\}}{\max\limits_{1<j<n}\{x_j\} - \min\limits_{1<j<n}\{x_j\}} \tag{1}$$

Normalization of data is to scale the data to a small specific interval. It is often used in the processing of some comparison and evaluation indicators, removing the unit limit of data and converting it into dimensionless pure values, so that indicators of different units or orders of magnitude can be compared and weighted. The most typical is the normalization of data, that is, the data are uniformly mapped to the [0,1] interval. After normalization, the speed of gradient descent to find the optimal solution is accelerated, and normalization may improve the accuracy.

## 3. Introduction of support vector machine regression model

SVR (support vector regression) support vector machine, which is often used in classification and time series prediction problems. According to different problems, it is mainly divided into linear separable and nonlinear separable. In SVM, we need to find a hyperplane with the largest gap, while in SVR, we define a ε， The residual of the data points defining the area within the dotted line is 0, while the distance from the data points (support vectors) outside the dotted line area to the boundary of the dotted line is the residual（ ζ）。 Similar to the linear model, we want these residuals（ ζ）minimum; SVR is to find the best strip area (2 ε Width), and then regress the points outside the region.

The kernel function in the model is very critical. When the variables increase, the dimensions mapped to the high-dimensional space increase exponentially, which may cause a "dimension disaster". At this time, the kernel function works. It converts the features from low dimension to high dimension, but it first calculates on the low dimension, and the actual classification effect is shown on the high dimension. In this way, we can avoid complex calculations in high dimensions and still get the same results. The commonly used kernel functions include polynomial function, kernel Gaussian function and linear function. This paper will choose the kernel function according to the characteristics of the data.

Suppose that the sample set with L training samples is {xi, yi}, and X and y are the column vector and row vector of the input value respectively. Suppose that the linear regression function established in the high-dimensional feature space is as follows.

$$f(x)=w\Phi(x)+b \tag{2}$$

Among them, Φ (x) Define for nonlinear mapping function, ε Linear insensitive loss function in the following formula, f (x) is the predicted value returned by the regression function, and Y is the corresponding real value.

$$L(f(x),y,\varepsilon)=\begin{cases}0 & |y-f(x)|\leqslant\varepsilon\\ |y-f(x)-\varepsilon| & |y-f(x)|>\varepsilon\end{cases} \tag{3}$$

Where C is the penalty factor. The greater C is, the greater the training error is, the greater the penalty of the sample is. E To specify the error requirements of the regression function, ε The smaller the error of the representative regression function is, the smaller the error is.

## 4. Model establishment and prediction

For the population prediction model over 65 years old, we set the time delay to 2, that is, we predict the data of the third year through the data of the past two years. 80% of the original data is trained and 20% is used for testing. The "Gaussian" Gaussian kernel function is the best kernel function of the model.

The training process of SVR model is as follows: the total elapsed time is 20.176 seconds. The total objective function calculation time is 10.3246, the observed objective function value is 0.032586, the estimated objective function value is 0.032514, and the function calculation time is 1.0243.

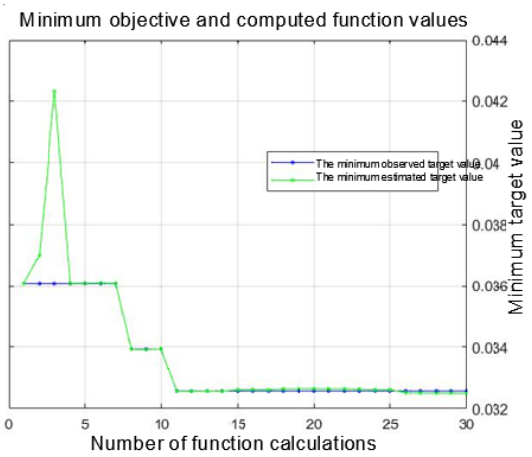



**Figure 2.** Training process of SVR model for population aging prediction in Inner Mongolia

As shown in the figure3 below, the prediction results of SVR model in Inner Mongolia are compared with the real values. It can be seen that the prediction effect of the model is very good, and the population over 65 years old will show an upward trend in the next 10 years, with certain fluctuations.

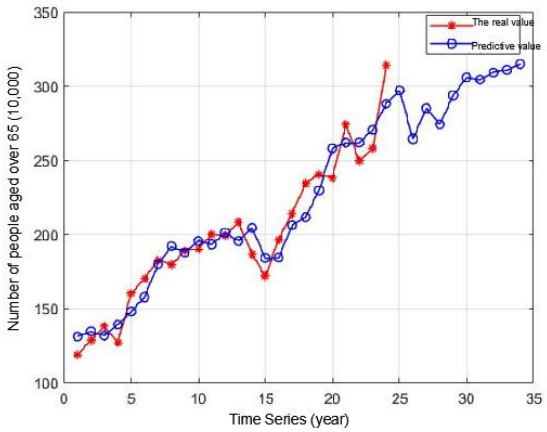



**Figure 3.** Inner Mongolia population aging prediction SVR model predicted value vs real value

Similarly, the training process of SVR model for aging population prediction in Guangxi is as follows: the total elapsed time is 20.5863 seconds. The total objective function calculation time is 11.172, the observed objective function value is 0.049721, the estimated objective function value is 0.047274, and the function calculation time is 0.026987.

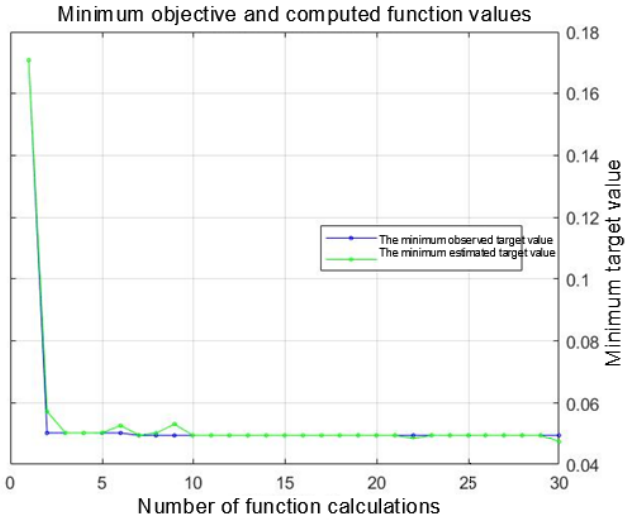



**Figure 4.** Training process of SVR model for population aging prediction in Guang Xi

The following figure5 shows the comparison between the prediction results of SVR model of aging population in Guangxi and the real value. It can be seen that the prediction effect of the model is very good, and the population over 65 years old will show a slow growth trend in the next 10 years.

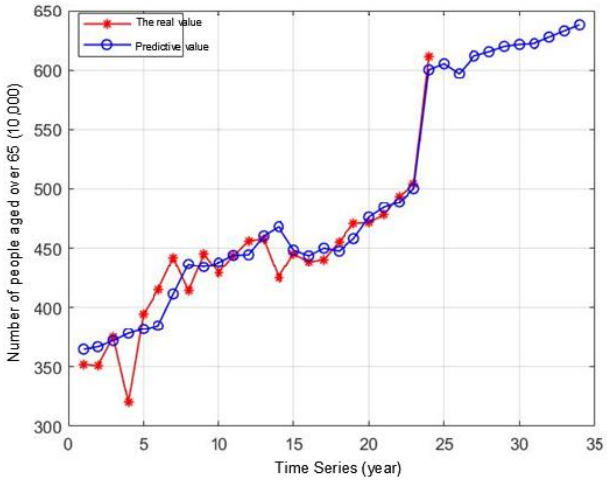



**Figure 5.** Guang Xi population aging prediction SVR model predicted value vs real value

The training process of SVR model for aging population prediction in Guizhou Province is as follows: the total elapsed time is 21.4768. The total objective function calculation time is 11.4249, the observed objective function value is 0.018818, the estimated objective function value is 0.018782, and the function calculation time is 0.025602.

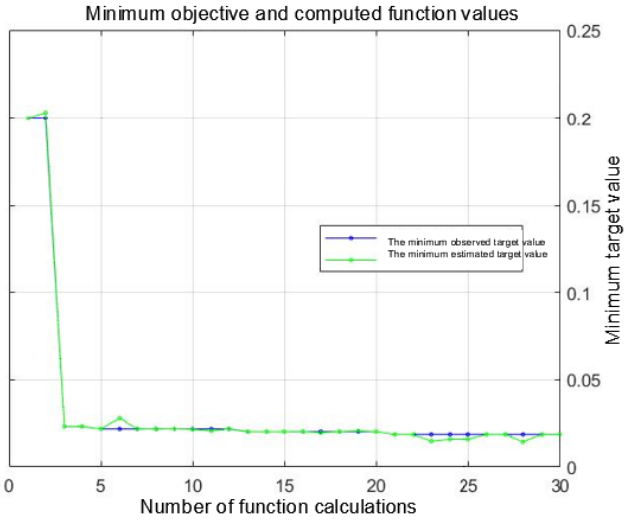



**Figure 6.** Training process of SVR model for population aging prediction in Gui Zhou

As shown in the figure7 below, the prediction results of the SVR model for the prediction of the aging population in Guizhou Province are compared with the real values. It can be seen that the prediction effect of the model is very good, and the population over 65 years old will show a straight-line upward trend in the next 10 years.

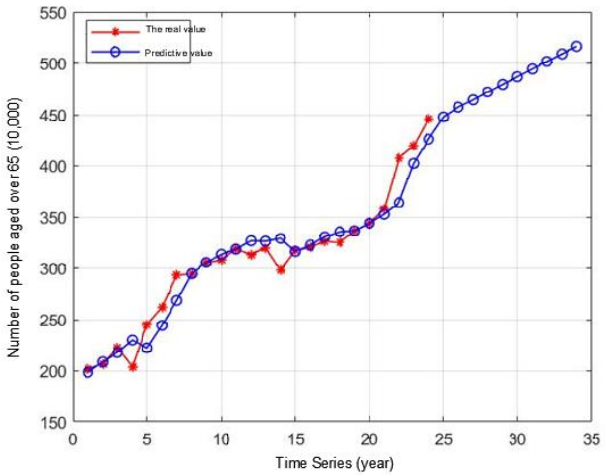



**Figure 7.** Gui Zhou population aging prediction SVR model predicted value vs real value

The total elapsed time of SVR model prediction process of aging population in Yunnan Province: 26.7563 seconds, the total objective function calculation time is 13.869, the observed objective function value is 0.028704, the estimated objective function value is 0.02854, and the function calculation time is 0.026922.

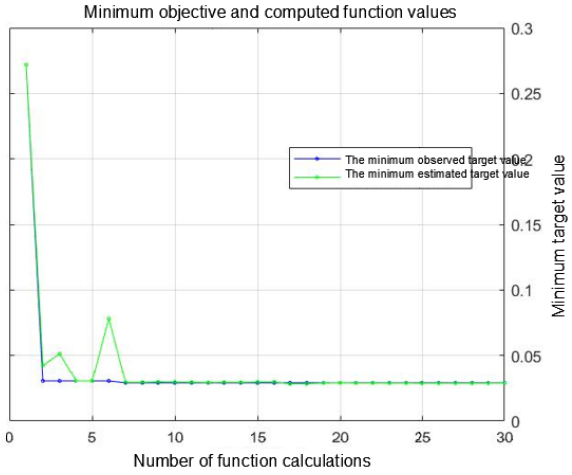



**Figure 8.** Training process of SVR model for population aging prediction in Yun Nan

As shown in the figure9 below, the prediction results of the SVR model for population prediction in Yunnan Province are compared with the real values. It can be seen that the prediction effect of the model is very good, and the population over 65 years old will rise rapidly in the next 10 years.

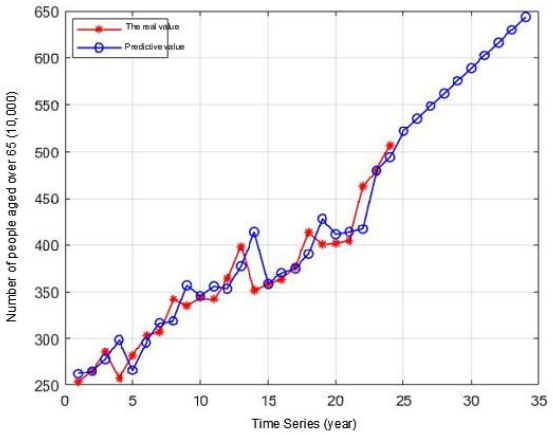



**Figure 9.** Yun Nan population aging prediction SVR model predicted value vs real value

As shown in the figure below, the total elapsed time of the prediction process of SVR model for Tibet population prediction is 25.5746 seconds, the total objective function calculation time is 12.3189, the observed objective function value is 0.14594, the estimated objective function value is 0.14699, and the function calculation time = 0.030524.

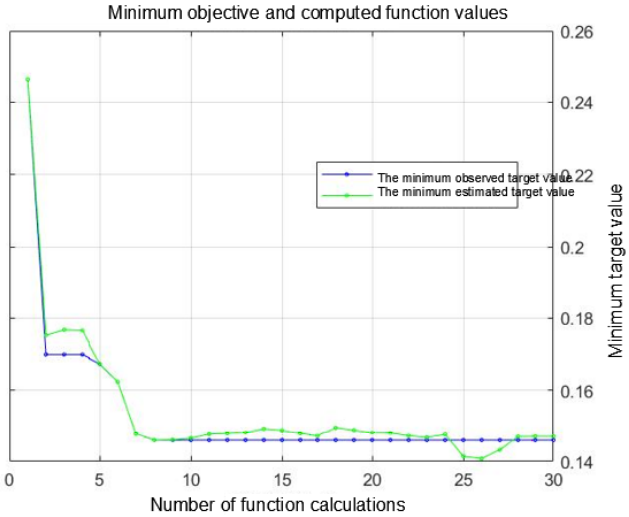



**Figure 10.** Training process of SVR model for population aging prediction in Tibet

As shown in the figure below, the prediction results of SVR model for population prediction in Tibet are compared with the real values. It can be seen that the prediction effect of the model is very good, and the population over 65 years old will show a slow downward trend in the next 10 years

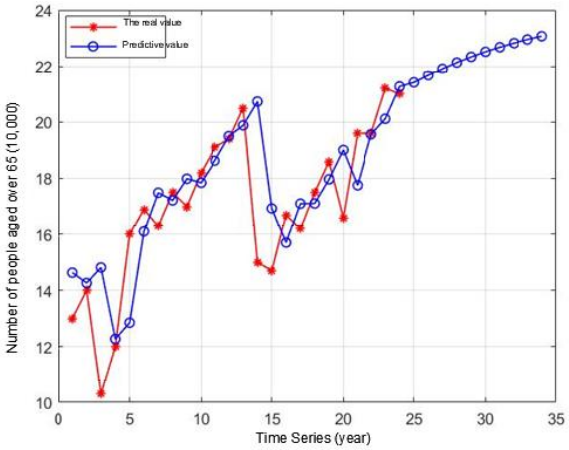



**Figure 11.** Tibet population aging prediction SVR model predicted value vs real value

The total elapsed time of Qinghai population prediction SVR model is 27.8745 seconds. The total objective function calculation time is 15.4199, the observed objective function value is 0.03715, the estimated objective function value is 0.037145, and the function calculation time is 1.9551.

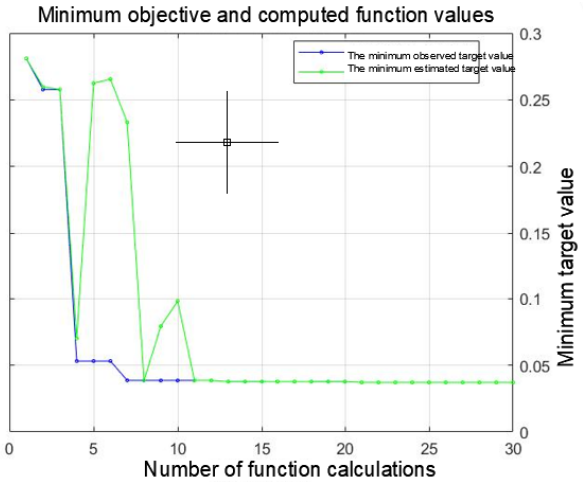



**Figure 12**. Training process of SVR model for population aging prediction in Qing Hai

The following figure13 shows the comparison between the prediction results of the SVR model of Qinghai population prediction and the real value. It can be seen that the prediction effect of the model is very good, and the population over 65 years old will have a steady upward trend in the next 10 years.

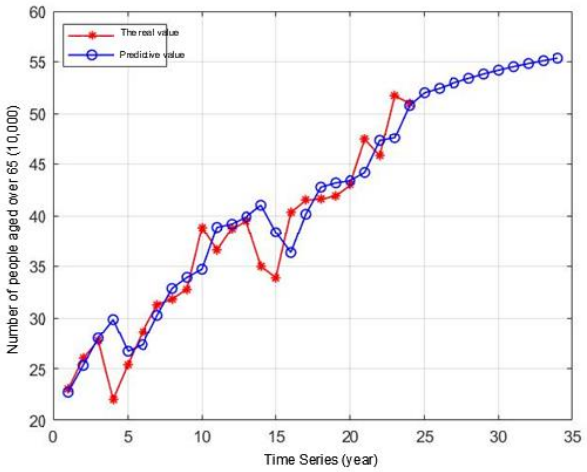



**Figure 13.** Qing Hai population aging prediction SVR model predicted value vs real value

The total elapsed time of the SVR model for Ningxia population prediction is 32.235 seconds, the total objective function calculation time is 18.0276, the observed objective function value is 0.015149, the estimated objective function value is 0.01508, and the function calculation time is 0.029552.

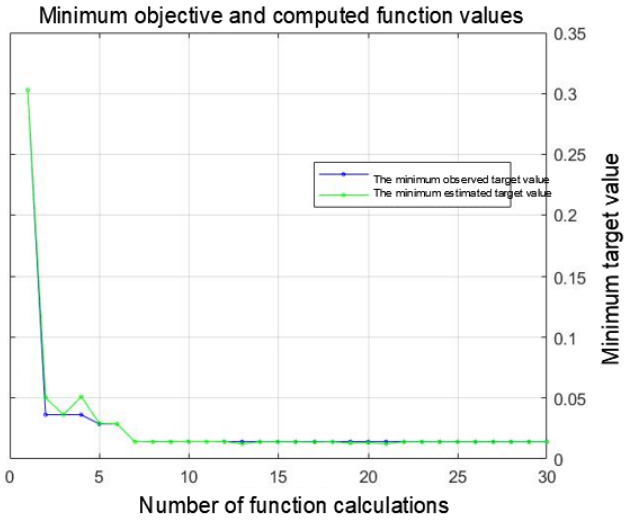



**Figure 14.** Training process of SVR model for population aging prediction in Ning Xia

As shown in the figure15 below, the prediction result of SVR model for Ningxia population prediction is compared with the real value. It can be seen that the prediction effect of the model is very good, and the population over 65 years old will show an upward trend in the next four years, and then gradually flatten out.

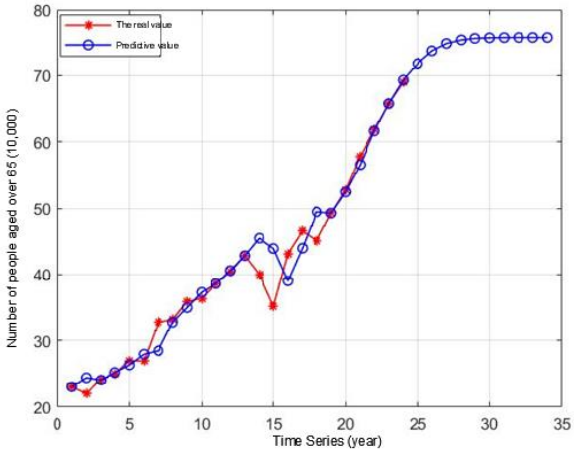



**Figure 15.** Ning Xia population aging prediction SVR model predicted value vs real value

The total elapsed time of Xinjiang population prediction SVR model is 18.0017 seconds, the total objective function calculation time is 6.9164, the observed objective function value is 0.027645, the estimated objective function value is 0.028208, and the function calculation time is 0.025656.

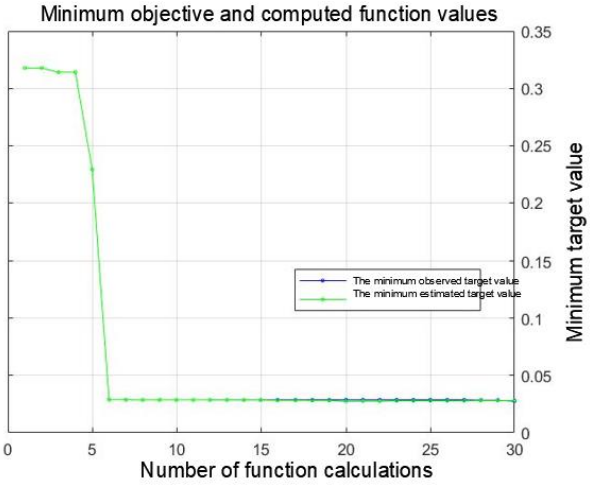



**Figure 16.** Training process of SVR model for population aging prediction in Xin Jiang

As shown in the figure16 below, the prediction results of the SVR model for population prediction in Xin Jiang are compared with the real values. It can be seen that the prediction effect of the model is very good, and the population over 65 years old will increase in the next 10 years.

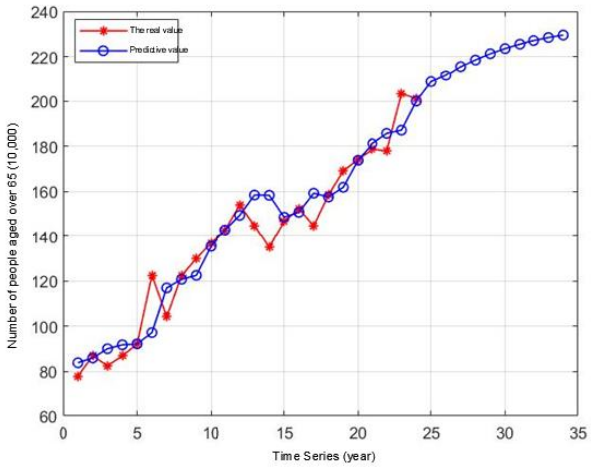



**Figure 17.** Xin Jiang population aging prediction SVR model predicted value vs real value

## 5. SVR model validation and error analysis

After the model training, this paper selects three indicators to evaluate the accuracy of the prediction model. In this paper, they are MSE, Mae and MAPE. The smaller their values are, the smaller the error between the predicted value and the real value is, and the higher the accuracy of the model is. The calculation formula is as follows:

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y})^2 \tag{4}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}| \tag{5}$$

$$MAPE = \frac{100\%}{m}\sum_{i=1}^{m}\left|\frac{y_i - \hat{y}}{y_i}\right| \tag{6}$$

Where is the actual value, the predicted value and the number of samples.

The following table1 shows the error index results of the eight ethnic provinces studied under the SVR prediction model.

**Table 1.** The error index results of the eight ethnic provinces

| Province | MAE | MPE | RMSE |
|---|---|---|---|
| Nei Menggu | 11.29 | 5.74% | 12.7789 |
| Guang Xi | 17.526 | 4.02% | 28.0593 |
| Gui Zhou | 10.92 | 3.62% | 15.9905 |
| Yun Nan | 15.9982 | 4.58% | 22.0509 |
| Xi Zang | 1.2985 | 8.70% | 1.9098 |
| Qing Hai | 2.075 | 6.16% | 2.8686 |
| Ning Xia | 1.6022 | 4.39% | 2.7223 |
| Xin Jiang | 6.8154 | 5.25% | 9.8574 |

## 6. Conclusion

Finally, the prediction results of the population over 65 years old (10000 people) in the next nine years in the eight ethnic provinces are shown in the table below.

**Table 2.** The prediction results of the population over 65 years old (10000 people) in the next nine years

| **Year** | Inner Mongolia | **Guangxi** | **Guizhou** | **Yunnan** | **Tibet** | **Qinghai** | **Ningxia** | **Xinjiang** |
|---|---|---|---|---|---|---|---|---|
| 2022 | 264.49 | 596.86 | 457.33 | 535.21 | 21.66 | 52.43 | 73.74 | 211.42 |
| 2023 | 285.12 | 611.64 | 464.36 | 548.66 | 21.91 | 52.96 | 74.83 | 215.53 |
| 2024 | 274.35 | 615.17 | 472.13 | 562.14 | 22.14 | 53.40 | 75.36 | 218.48 |
| 2025 | 293.88 | 619.62 | 479.60 | 575.66 | 22.34 | 53.82 | 75.59 | 221.26 |
| 2026 | 305.88 | 621.25 | 487.09 | 589.22 | 22.52 | 54.19 | 75.67 | 223.53 |
| 2027 | 304.45 | 622.04 | 494.51 | 602.83 | 22.68 | 54.53 | 75.71 | 225.48 |
| 2028 | 309.28 | 627.31 | 501.88 | 616.47 | 22.83 | 54.83 | 75.72 | 227.09 |
| 2029 | 311.00 | 632.41 | 509.21 | 630.16 | 22.96 | 55.11 | 75.72 | 228.44 |
| 2030 | 315.00 | 637.44 | 516.48 | 643.90 | 23.07 | 55.35 | 75.72 | 229.54 |

It can be seen from the prediction of the eight ethnic provinces in the next 10 years that the number of people over the age of 65 in other provinces, except Ningxia Province, is on the rise. These prediction results are helpful to the formulation of pension measures by relevant departments. In addition, this paper also predicts the population aging coefficient through the same method, and the results are shown in the table below.

**Table 3.** The prediction results of the population aging coefficient in the next nine years

| **Year** | Inner Mongolia | **Guangxi** | **Guizhou** | **Yunnan** | **Tibet** | **Qinghai** | **Ningxia** | **Xinjiang** |
|---|---|---|---|---|---|---|---|---|
| 2022 | 13.10 | 13.81 | 11.73 | 11.09 | 5.71 | 8.78 | 10.37 | 8.10 |
| 2023 | 13.28 | 15.36 | 11.80 | 11.27 | 6.06 | 8.83 | 10.75 | 8.24 |
| 2024 | 13.47 | 17.80 | 11.86 | 11.45 | 5.66 | 8.88 | 11.13 | 8.34 |
| 2025 | 13.65 | 20.00 | 11.92 | 11.63 | 5.89 | 8.92 | 11.52 | 8.44 |
| 2026 | 13.83 | 20.57 | 11.98 | 11.81 | 5.75 | 8.97 | 11.93 | 8.54 |
| 2027 | 14.00 | 20.02 | 12.04 | 12.00 | 6.13 | 9.00 | 12.36 | 8.63 |
| 2028 | 14.17 | 19.87 | 12.09 | 12.18 | 5.67 | 9.04 | 12.80 | 8.72 |
| 2029 | 14.34 | 20.05 | 12.14 | 12.37 | 5.82 | 9.07 | 13.26 | 8.80 |
| 2030 | 14.51 | 20.09 | 12.18 | 12.56 | 5.78 | 9.10 | 13.74 | 8.88 |

According to the prediction results of the aging coefficient of the eight ethnic provinces, it can be found that the overall aging coefficient of Tibet will show an upward trend in the next ten years, which is consistent with the change trend of the population over 65 predicted above, which once again verifies the rationality of the prediction results of the SVR model established in this paper.

To sum up, for the eight ethnic provinces of Inner Mongolia, Guangxi, Guizhou, Yunnan, Tibet, Qinghai, Ningxia and Xinjiang, SCR prediction model and aging coefficient SVR prediction model for the population over 65 years old are established respectively. MSE, Mae and MAPE are selected to test the accuracy of the model. The results show that the accuracy of the model established in this paper is high, and the changes of population over 65 years old and aging coefficient in each province in the next 10 years are predicted, The results show that the population over 65 years old and the aging coefficient in the eight provinces have an overall upward trend, and the prediction trend of the model for the population over 65 years old and the aging coefficient in the next ten years is consistent, which also verifies the rationality of the prediction results.

## References

[1] Wang Chenming,Guan Xin,Bai Yansen,Feng Yue,Wei Wei,Li Hang,Li Guyanan,Meng Hua,Li Mengying,Jie Jiali,Fu Ming,Wu Xiulong,He Meian,Zhang Xiaomin,Yang Handong,Lu Yanjun,Guo Huan. A machine learning-based biological aging prediction and its associations with healthy lifestyles: the Dongfeng-Tongji cohort.[J]. Annals of the New York Academy of Sciences,2021,1507(1).

[2] Helen-Maria Dounavi,Yiorgos Sfikas,Yiorgos Tsiatouhas. Aging Prediction and Tolerance for the SRAM Memory Cell and Sense Amplifier[J]. Journal of Electronic Testing,2021(prepublish).

[3] Meng Haining,Shi Yuekai,Qu Yilin,Li Junhuai,Liu Jianjun. ARIMA-Based Aging Prediction Method for Cloud Server System[J]. IOP Conference Series: Materials Science and Engineering,2021,1043(2).

[4] Xin-Ke Zhang,Yi-Bin Lan,Yue Huang,Xu Zhao,Chang-Qing Duan. Targeted metabolomics of anthocyanin derivatives during prolonged wine aging: Evolution, color contribution and aging prediction[J]. Food Chemistry,2021,339.

[5] Sukanta Dey,Sukumar Nandi,Gaurav Trivedi. Machine Learning Approach for Fast Electromigration Aware Aging Prediction in Incremental Design of Large Scale On-chip Power Grid Network[J]. ACM Transactions on Design Automation of Electronic Systems,2020,25(5).

[6] Information Technology - Industrial Informatics; New Industrial Informatics Findings from University of Warwick Discussed (Gaussian Process Regression With Automatic Relevance Determination Kernel for Calendar Aging Prediction of Lithium-ion Batteries)[J]. Journal of Technology & Science,2020.

[7] Kailong Liu,Yi Li,Xiaosong Hu,Mattin Lucu,Widanalage Dhammika Widanage. Gaussian Process Regression With Automatic Relevance Determination Kernel for Calendar Aging Prediction of Lithium-Ion Batteries[J]. IEEE Transactions on Industrial Informatics,2020,16(6).

[8] Agne Laucyte-Cibulskiene,Ligita Ryliskyte,Milda Kovaite,Laurynas Rimsevicius,Laura Balkeviciene,Jolita Badariene,Aleksandras Laucevicius. P72 Vascular Aging Prediction in High Cardiovascular Risk Cohort[J]. Artery Research,2020,25(S1).

[9] Computers - Computer-Aided Design; Findings from University of Tehran Has Provided New Data on Computer-Aided Design (Instruction-level Nbti Stress Estimation and Its Application In Runtime Aging Prediction for Embedded Processors)[J]. Computer Weekly News,2019.

[10] Guillaume Lahiner,James Zappata,Jeremy Cure,Nicolas Richard,Mehdi Djafari-Rouhani,Alain Estève,Carole Rossi. A redox reaction model for self-heating and aging prediction of Al/CuO multilayers[J]. Combustion Theory and Modelling,2019,23(4).