Forecasting of elderly population in Taiwan using Deep Learning

Stephanie Fu Yang[1],cheng-Hong Yang[2,3] ,Indra Satriawan Sakti[3]

[1]國立成功大學心理學系， [2]台南應用科技大學資管系，[3]國立高雄科技大學電子工程系，

[2]chyang@nkust.edu.tw、[3]C110152154@nkust.edu.tw

Abstract

Taiwan's rapidly aging population poses significant challenges for healthcare and social policy planning. This study conducts a comparative analysis of advanced forecasting models to predict elderly population trends using data from the Department of Household Registration, Ministry of the Interior (M.O.I.), Taiwan (2001–2024). In this paper, a deep learning-based time series prediction method, namely Segment Recurrent Neural Network (segRNN), is proposed. Performance is evaluated using Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The results indicate that the segRNN model outperforms traditional methods, achieving the lowest error rates. These findings provide actionable insights for policymakers to allocate resources efficiently and design sustainable aging-in-place initiatives.

Keywords: Elderly population forecasting; segRNN; Taiwan; demographic policy

Introduction

Population is a foundational element of nation building, while changes in population size and age structure are still more important as key determinants of national development[1]. The impacts of population aging are generally related to family changes, migration of youth to cities, need for support and care, need for attention to health problems, and housing problems[2]. In the face of these dramatic changes in vital rates, deterministic methods are less suitable to capture the uncertainty, while stochastic forecasts have been more successful Lee and Tuljapurkar 1994[15]. Accurate forecasting of elderly population dynamics is thus essential for policymakers to allocate resources efficiently and mitigate future economic and societal burdens. Changes in population demographics, present both opportunities and significant challenges, especially the increasing proportion of elderly individuals, which can reshape the landscape of national development.

Traditional time-series forecasting methods, such a ARIMA(Autoregressive Integrated Moving Average)[4], TBATS(Trigonometric, Box-Cox transform, ARMA errors, Trend, and Seasonal components)[5] have been widely used for time-series forecasting in the past due to their robustness and interpretability. However, these models have multiple shortcomings, such as often struggle to capture complex nonlinear patterns in long-term forecasting, making simple assumptions, and not beingn't to apply to the nonlinear relation, and lower outliers sensitive e.t.c.

To address the limitations, machine learning techniques such a Support Sector Regression (SVR)[6] and deep learning such as a Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), and Segmented Recurrent Neural Network (SegRNN) have been utilized. These models demonstrate an exceptional ability to detect complex temporal patterns while recognizing nonlinear population changes, making them ideal for population forecasting applications. SegRNN, in particular, excels in capturing segmented patterns in data, offering an additional advantage in handling diverse time series data. SegRNN's ability to effectively model different segments of the time series improves its forecasting accuracy, especially in cases with varying patterns over time.

Research Methods

In this study, we utilize data from the Ministry of the Interior (M.O.I.), Taiwan [1], covering the period from January 2001 to February 2025, to explore the effectiveness of deep learning in forecasting Taiwan's elderly population. We compare various models, including ARIMA, TBATS, SVR, LSTM, and GRU. Additionally, we introduce the Segmental Recurrent Neural Network (segRNN)[3], an approach that enhances sequence modeling by incorporating segment-based learning to improve forecasting accuracy. We also identify and remove anomalous data and divide the dataset into training and test sets with a 6:4 ratio. This study aims to provide insights into the strengths and limitations of each method while highlighting the potential of deep learning techniques in demographic forecasting.

Feature Selection and Machine Learning

Autoregressive Integrated Moving Average(ARIMA)

ARIMA mode was formulated in 1976 by Box and Jenkins[4] and id also known as the Box-Jenkins model. This model using statistic for analysis and using prediction in time series. This Model devided into three parts, a combination of the autoregressionAR(p),Moving average MA(q), and differencing degree(d)[7].The formula for the Arima will explain in equation (1):

$$y_t = \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{1}$$

Where y_t represents observation value at t th time, ε_t represent the noise and Gaussian distributed.$\emptyset_i (i = 1,2, \dots, p)$ is an autoregressive (AR) coefficient, and $\theta_j (j = 1,2, \dots, j)$ is the Moving average(MA). The integers (p) and (q) are referred to as the model orders. The ARIMA model is denoted as ARIMA(p, d, q). One of the most important aspects of ARIMA modeling is identifying the appropriate model order (p, q).

Trigonometric, Box-Cox transform, ARMA errors, Trend, and Seasonal components (TBATS)

TBATS is an advanced forecasting method that extends the classical BATS(Box-Cox, ARMA, Trend, and Seasonal) model. It is designed to handle complex time series data with multiple seasonality and non linear trends[8]. TBATS combines several componets, making it suitable for forecasting time series with intricate seasonal patterns, long range trends, and non-linear effects

The general component of TBATS explain in below:

- Trigonometric Seasonality : TBATS uses **trigonometric functions** (such as sine and cosine) to handle **multiple seasonalities** in the data, which is useful for capturing seasonal patterns at different frequencies (e.g., weekly, monthly, yearly).

-Box-Cox Transformation : This transformation is used to stabilize variance and make the data more suitable for modeling, especially when the data has heteroscedasticity (non-constant variance over time).

- ARMA(AutoRegressive Moving Average): The model includes an ARMA component to capture autocorrelation in the residuals.

-Trend : TBATS models both local trends (changes in the series over time) and non-linear trends.

- Seasonality : TBATS can handle multiple seasonalities in the data, which is useful when the data exhibits seasonal patterns at different frequencies.

Support Vector Regression (SVR)

SVR wa proposed by Vapnik et al. in 1997[9] for classification problems. Support Vector Machine was later adapted to Support Vector Regression (SVR) by using a new type of loss function called ε-insensitive loss function which is used to penalize data as long as they are greater than ε [9]. SVR is a non-linear kernel-based regression method that provides the best regression hyperplane with smallest structural risk in a high-dimensional feature space [10].

Long Short Term Memory (LSTM)

LSTM mode was formulated in 1997 by Sepp Hochreiter and Jürgen Schmidhuber[12]. LSTM is a type of artificial neural network that belongs to the family of Recurrent Neural Networks (RNNs). This model was designed to overcome the vanishing gradient problem that traditional RNNs face, especially when learning long-term dependencies in

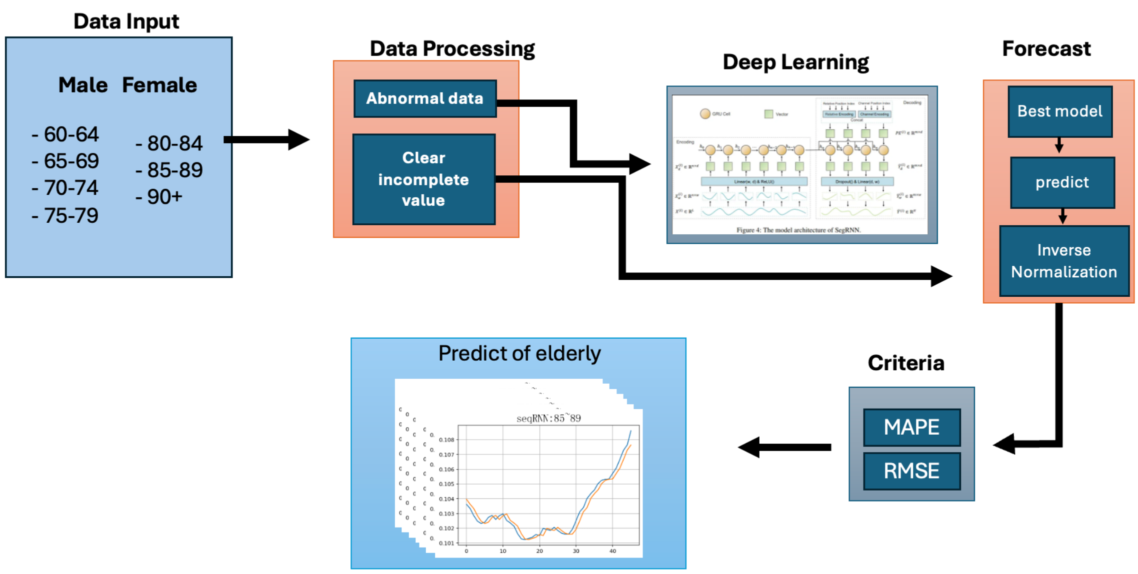


Figure 1. Flowchart of the elderly population forecasting.

sequential data. LSTM is widely used for tasks involving time series forecasting, language translation, speech recognition, and other sequence-based tasks. The general component of LSTM explain in below:

- Memory Cell: Stores information over time, allowing the network to maintain long-term dependencies.
- Input Gates : Control how much new information should be stored in the memory cel
- Forget Gate : Decides how much of the previous memory should be forgotten or discarded
- Output Gate : Determines what the final output of the LSTM should be, based on the information stored in the memory cell.

Gradient Recurrent Unit (GRU)

GRU was introduced by Kyunghyun Cho et al. in 2014 as a variant of Recurrent Neural Networks (RNNs)[13]. Like LSTMs, GRUs were designed to address the vanishing gradient problem in traditional RNNs, making them effective for learning long-term dependencies in sequential data. However, GRUs have a simpler structure compared to LSTMs, making them computationally more efficient while still achieving comparable performance in many sequence-based tasks, such as time series forecasting, language translation, and speech recognition.

The general components of a GRU are explained below:

- Update Gate: Controls how much of the past information should be carried forward to the next time step.
- Reset Gate: Determines how much of the previous memory should be forgotten.
- Hidden State: Combines the previous hidden state and new candidate information to update the current state dynamically.

Segmental Recurrent RNN(segRNN)

SEG-RNN is a model designed for time series data processing by dividing the sequence into segments. This approach is particularly effective in handling missing values and irregular patterns, outperforming conventional RNN-based methods in such scenarios.

The foundation of SEG-RNN consists of several stages. First, the time series is divided into multiple segments based on specific rules, such as fixed length or patterns within the data. Each segment is then processed using a Recurrent Neural Network (RNN), which can be either LSTM or GRU, to capture temporal patterns within each segment. The results from each segment are subsequently combined using an aggregation mechanism, such as mean pooling, attention, or concatenation, to generate the final prediction.In its implementation, I use the Pypots library, which is specifically designed to handle time series data with missing values. Pypots provides various preprocessing methods, RNN-based models, and segmentation mechanisms to process data more efficiently. The Pypots implementation I use is based on references from paper [14], which serves as the foundation for selecting the architecture and methods applied in the SEG-RNN model.

Evaluation Metrics

The performance evaluation metrics employed in this study provided complete assessment of three machine learning regressors through multiple variables. Two assessment metrics evaluate prediction accuracy: RMSE that measures the total error magnitude as shown in formula (2) and MAPE that determines prediction accuracy based on error percentage expressed in formula (3).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - f_i)^2} \qquad (2)$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - f_i}{y_i}\right| \times 100\% \qquad (3)$$

Research Results

Dataset

The demographic data obtained for this analysis from Taiwan's Ministry of Home Affairs official data portal located at https://www.ris.gov.tw/app/portal/346. The data population dataset from Taiwan includes groupings of citizens across 5-year age categories beginning with 60-64 years old up to 90 years and over (90+). The analyzed data extends from January 2001 through February 2025. The data separates its information between male and female gender groups. The data provides definition-level statistical data used to conduct population studies across Taiwan.

Table 1. demographic of eldery population

age	gender	mean	median	Min	max	Sd
60-64	Female	638164	665344	391428	916096	200049
	Male	600130	629581	371990	849692	184253
65-69	Female	511124	400200	326655	834822	171772
	Male	466084	364463	322139	748276	152416
70-74	Female	384472	359360	257157	695201	111000
	Male	349002	316861	280263	597433	084409
75-79	Female	276424	281742	172922	397104	056479
	Male	251257	253588	206936	321185	019941
80-84	Female	191062	191965	094878	282122	056872

	Male	166531	174436	096146	203421	028010
85-89	Female	103390	100955	045658	178015	040285
	Male	084936	099206	038958	114885	025939
90+	Female	050687	045690	017404	098953	025121
	Male	039286	037262	011532	067796	018661

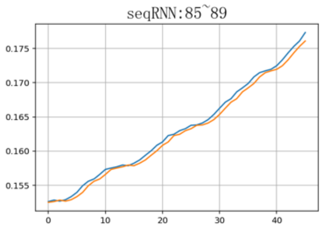
Comparison of Different Algorithms in Predicting Scoring Ability in

In this study, we employed various forecasting approaches including ARIMA, TBATS, SVR, LSTM, GRU, and segRNN to predict and analyze the data. A series of evaluation metrics, specifically RMSE and MAPE, were used to compare the performance of these models (Table 2). The results demonstrated that the segRNN model achieved the lowest RMSE and MAPE values across all datasets, indicating superior forecasting accuracy and overall performance. Further validation through residual analysis (Fig. 2) confirmed that the segRNN model consistently exhibited the most favorable error distribution, underscoring its superior predictive capability among the evaluated models.

Table 2. Comparison of the value of RMSE and MAPE between the segRNN model and some traditional models

age	Gender	metric	ARIMA	TBATS	SVR	LSTM	GRU	segRNN
60-64	Male	MAPE	0.236	0.886	28.715	1.97	5.643	**0.128**
		RMSE	0.0025	0.0082	0.239	0.017	0.0472	**0.0014**
	Female	MAPE	0.242	1.03	28.71	0.741	3.07	**0.149**
		RMSE	0.0027	0.01	0.239	0.0069	0.028	**0.0017**
65-69	Male	MAPE	0.426	0.784	29.31	9.167	0.38	**0.217**
		RMSE	0.0042	0.007	0.211	0.0688	0.0029	**0.0017**
	Female	MAPE	0.369	1.123	29.31	3.988	28.35	**0.242**
		RMSE	0.0039	0.011	0.211	0.039	0.227	**0.002**
70-74	Male	MAPE	0.743	15.866	31.33	7.097	35.67	**0.662**
		RMSE	0.0054	0.111	0.173	0.042	0.192	**0.0035**
	Female	MAPE	2.39	12.47	31.33	14.48	5.415	**0.703**
		RMSE	0.092	0.102	0.173	0.105	0.036	**0.0043**
75-79	Male	MAPE	1.471	16.869	12.43	0.844	7.689	**0.529**
		RMSE	0.006	0.06	0.041	0.0041	0.0277	**0.0019**
	Female	MAPE	1.135	131	12.43	1.774	49.52	**0.501**
		RMSE	0.005	0.063	0.041	0.01	0.178	**0.0023**
80-84	Male	MAPE	0.574	4.886	26.863	0.981	0.597	**0.183**
		RMSE	0.0014	0.011	0.05	0.0021	1.225	**0.0004**
	Female	MAPE	0.499	4.188	26.86	1.57	49.7	**0.165**
		RMSE	0.0017	0.013	0.053	0.0043	0.136	**0.0005**
85-89	Male	MAPE	1.02	5.09	25.47	0.306	50.895	**0.278**
		RMSE	0.0014	0.007	0.026	0.0003	21.679	**0.0003**
	Female	MAPE	0.654	1.606	25.47	7.43	29.64	**0.334**
		RMSE	0.0012	0.002	0.026	0.013	0.048	**0.0006**
90+	Male	MAPE	1.111	12.23	39.079	0.669	1.887	**0.392**
		RMSE	0.0009	0.0094	0.025	0.0004	1.37	**0.0002**
	Female	MAPE	0.876	6.503	39.079	2.6	11.77	**0.4126**
		RMSE	0.001	0.007	0.0252	0.0023	0.0108	**0.0003**
average	Male	MAPE	0.7972	8.087	27.599	3.004	14.680	**0.3412**

	RMSE	0.0031	0.030	0.1092	0.0192	3.506	**0.0013**
Female	MAPE	0.8807	22.56	27.598	4.6547	25.352	**0.3580**
	RMSE	0.015	0.0297	0.1097	0.0257	0.0948	**0.0016**



(a)

(b)

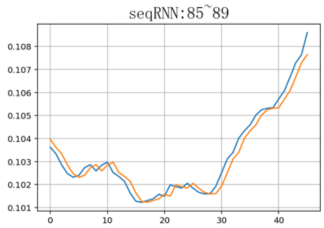


Figure 2.a) compare time step and scales of male dataset
b) combare time step and scales of female dataset

The graph above represents the predictive performance of the SegRNN model. The blue line represents the actual data, while the orange line corresponds to the predicted values. The x-axis denotes the time steps, and the y-axis represents the scaled values of the dataset. The two curves closely follow the same trend, indicating that the model effectively captures the underlying pattern in the data.

To evaluate the model's accuracy, Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) were calculated. A lower MAPE suggests that the model's predictions are close to actual values in percentage terms, while a lower RMSE indicates minimal deviation from the actual data. The results show that SegRNN demonstrates a good fit, with only slight deviations in certain sections.

However, the similarity between the predicted and actual values raises the possibility of overfitting, where the model learns the training data too well but may not generalize effectively to unseen data. To further improve the model, adjustments such as modifying the dropout rate[11], increasing the training data size, or comparing with other deep learning models like LSTM could be explored. The SegRNN model shows promising results in forecasting, but further evaluation is necessary to ensure its robustness across different datasets.

Discussion

Result of Elderly Population Forecasting Using Time Series Data

In the context of elderly population prediction using time series data in Taiwan, various machine learning models were tested to determine which would provide the most accurate and reliable forecast. Among these models, SegRNN (Segmented Recurrent Neural Network) stands out as the most effective, outperforming other algorithms, especially traditional models such as GRU (Gated Recurrent Units) and LSTM (Long Short-Term Memory). evaluating the performance, SegRNN achieved the lowest metrics overall, with an average of RMSE male is 0.3412. This highlights the superior predictive capability of SegRNN in comparison to the other models tested. Specifically, SegRNN has demonstrated a remarkable ability to handle time series data with high accuracy, which is essential when making predictions about the population of elderly individuals in Taiwan.

Furthermore, SegRNN showed remarkable stability in its predictions, as indicated by the values of both RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error), all of which consistently fell below 1 across all testing scenarios. This level of stability is critical in forecasting, especially for time-sensitive applications where precision and consistency are essential for effective decision-making.

The stability of SegRNN can be attributed to its advanced segmentation techniques, which allow the model to capture long-term dependencies in the data more effectively than traditional models. This characteristic of SegRNN ensures that predictions remain reliable, even in the face of fluctuating trends in the data.

Based on these results, it is strongly recommended to use SegRNN for forecasting similar datasets or related time series prediction tasks. The combination of superior accuracy, stability, and efficiency makes SegRNN a compelling choice for time series analysis in fields such as demographic forecasting, resource allocation for elderly care, and policy planning.

Conclusion

segRNN stands out as the best-performing model for this task, offering a combination of high accuracy, stability, and efficiency. The results suggest that future research and real-world implementations should consider SegRNN as a preferred model for time series forecasting, especially in applications involving demographic and social planning. Further studies could explore optimizing SegRNN for different datasets, integrating it with hybrid approaches, or testing its generalizability in other forecasting domains. By leveraging SegRNN's capabilities, researchers and policymakers can make more informed decisions based on reliable, data-driven insights.

Refference

1. 內政部移民署, "全國人口統計資料," 內政部移民署, 2023. [線上]. Available: https://www.ris.gov.tw/app/portal/346. [存取日期: 2025 年 3 月 31 日].
2. Z. Ismail, W. I. W. Ahmad, S. H. Hamjah, and I. K. Astina, "The Impact of Population Ageing: A Review," *Iran J. Public Health*, vol. 50, no. 12, pp. 2451-2460, Dec. 2021, doi: 10.18502/ijph.v50i12.7927.
3. L. Kong, C. Dyer, and N. A. Smith, "Segmental Recurrent Neural Networks," *ArXiv*, 2015. [Online]. Available: https://arxiv.org/abs/1511.06018.
4. G. M. Jenkins and G. E. P. Box, *Time Series Analysis: Forecasting and Control*, 1976.
5. A. M. De Livera, R. J. Hyndman, and R. D. Snyder, "Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing," *J. Am. Stat. Assoc.*, vol. 106, no. 496, pp. 1513–1527, 2011, doi: 10.1198/jasa.2011.tm09771.
6. M. Awad and R. Khanna, "Support Vector Regression," in *Efficient Learning Machines*, Berkeley, CA, USA: Apress, 2015, doi: 10.1007/978-1-4302-5990-9_4.
7. R. D. Lee and S. Tuljapurkar, "Stochastic Population Forecasts for the United States: Beyond High, Medium, and Low," *J. Amer. Stat. Assoc.*, vol. 89, no. 428, pp. 1175-1189, 1994. doi: 10.1080/01621459.1994.10476857.
8. L. Schmid, M. Roidl, and M. Pauly, "Comparing statistical and machine learning methods for time series forecasting in data-driven logistics–A simulation study," *arXiv preprint*, arXiv:2303.07139, Jun. 2024. [Online]. Available: https://arxiv.org/pdf/2303.07139.
9. V. Vapnik, S. E. Golowich, and A. J. Smola, "Support vector method for function approximation, regression estimation and signal processing," in *Advances in Neural Information Processing Systems*, Denver, CO, USA, Dec. 1996, pp. 281–287.
10. M. Ghanbari and H. Arian, "Forecasting Stock Market with Support Vector Regression and Butterfly Optimization Algorithm," *arXiv preprint*, arXiv:1905.11462, May 2019. [Online]. Available: https://arxiv.org/pdf/1905.11462.
11. N. Srivastava, *Improving Neural Networks with Dropout*, M.S. thesis, Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, 2013. [Online]. Available: https://www.cs.toronto.edu/~nitish/msc_thesis.pdf.
12. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
13. R. Dey and F. M. Salem, "Gate-variants of Gated Recurrent Unit (GRU) neural networks," in *Proc. 2017 IEEE 60th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Boston, MA, USA, 2017, pp. 1597-1600, doi: 10.1109/MWSCAS.2017.8053243.
14. S. Lin, W. Lin, W. Wu, F. Zhao, R. Mo, and H. Zhang, "SegRNN: Segment Recurrent Neural Network for Long-Term Time Series Forecasting," *ArXiv*, 2023. [Online]. Available: https://arxiv.org/abs/2308.11200.

15. R. D. Lee and S. Tuljapurkar, "Stochastic Population Forecasts for the United States: Beyond High, Medium, and Low," *J. Amer. Stat. Assoc.*, vol. 89, no. 428, pp. 1175-1189, 1994. doi: 10.1080/01621459.1994.10476857.