

A Newton-Raphson-Based Optimizer-Driven Temporal Convolutional Networks for Birth Rate Prediction in a Small Area

Shengyi Zhou^{1#}, Liang Chen^{2#*}, Wei Han³, Bin Liu⁴

School of Economics and Management, Sichuan Vocational College of Chemical Technology, Luzhou, China^{1, 2, 3}

Universiti Teknologi Malaysia, Iskandar Puteri, Malaysia³

School of Economics and Management, Aba Teachers College, Aba, China⁴

Abstract—For economically developed small geographic regions, population forecasting serves as a vital tool for achieving refined regional management. However, due to relying on the subjective experience of experts, traditional methods for predicting birth rates have shortcomings in accuracy, resulting in unreliable results. To address this limitation, this study introduces deep learning (DL) models into the domain of birth rate prediction. Specifically, a hybrid TCN-Bi-LSTM model is proposed, integrating a Temporal Convolutional Network (TCN) with a Bi-directional Long Short-Term Memory (Bi-LSTM) network to predict birth populations in small regions. The proposed hybrid model effectively leverages the strengths of the TCN and Bi-LSTM to capture both local temporal patterns and long-term hidden dependencies within birth rate time series data. The proposed birth rate prediction model not only incorporates historical data on regional birth rates but also accounts for the influence of factors such as divorce rates, consumption levels, and population size. Furthermore, an enhanced meta-heuristic algorithm is designed to optimize the hyperparameters of the hybrid TCN-Bi-LSTM model, with the aim of increasing its prediction accuracy. The hippopotamus position update strategy was introduced into the Newton-Raphson-Based Optimizer (NRBO), and an improved NRBO (INRBO) algorithm was developed. Finally, the performance of the proposed birth rate prediction model was validated using a dataset from three regions or countries. The prediction results demonstrate that, compared to the other four models, the proposed INRBO-TCN-Bi-LSTM model achieves the best performance, with an average reduction of 95% in training loss.

Keywords—Temporal Convolutional Network; Bi-directional Long Short-Term Memory; prediction model; birth rate; meta-heuristic algorithm

I. INTRODUCTION

For some economically developed small regions, the birth rate is of significant importance in assessing the fiscal sustainability of the social welfare system. At the same time, the number of births plays a crucial role in the rational allocation of public service resources such as education and healthcare [1]. As a result, predicting the birth population or birth rate has long been a major focus for demographers and policymakers. However, birth rates are easily influenced by factors such as economic growth, divorce rates, and regional population size, which pose challenges to the accuracy of birth rate prediction models [2]. An accurate birth rate prediction

model can provide a decision-making basis for the formulation of social welfare, healthcare, and education policies. Traditional population forecasting models often rely heavily on the expertise of demographers, which can lead to limitations in accuracy. With advances in deep learning (DL) and machine learning (ML), recent studies have begun to apply DL and ML models to population prediction and birth rate forecasting tasks [3].

Predicting birth rates is a classic time series forecasting problem. The most popular solutions for solving such problems currently are prediction models based on Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), such as the classic TCN [4] and Long Short-Term Memory (LSTM) networks [5]. The TCN and LSTM networks are capable of capturing local information and hidden information of the long-term dimension in time series data, respectively. However, current deep learning-based birth rate prediction models still rely on manual hyperparameter tuning to improve their accuracy. To address this issue, Huang et al. employed the particle swarm optimization (PSO) algorithm to optimize the hyperparameters of the CNN, with the goal of identifying the optimal hyperparameter configuration for the CNN model [6]. Although CNN and LSTM networks have been widely applied in various prediction problems, DL based birth rate prediction models face the following challenges:

- Birth rate data has obvious low-frequency characteristics and scarcity, and the available sample size is very limited, which is a huge challenge for DL models that require a large amount of data for training.
- There are scale differences and multicollinearity among multiple socio-economic indicators, such as economic growth rate, divorce rate, and population size within the region, which pose difficulties for feature fusion.
- TCN is extremely sensitive to hyperparameters, and the birth rate is strongly influenced by sudden external factors such as policy adjustments and major events. These factors make accurate prediction of birth rates difficult.

To solve the above problems, this study proposes an INBRO-TCN-Bi-LSTM model. The proposed model leverages both TCN and Bi-LSTM architectures to extract different types of key features from birth rate-related time series data. Liao et

[#]Equal contribution.

^{*}Corresponding author.

al. used the output of TCN as the input of LSTM and designed CNN-LSTM for predicting earthquake probability [7]. Differently, the proposed hybrid model employs the output of the Bi-LSTM, carrying rich contextual information, as an enhanced feature input to the fully connected layer of the TCN. This enables the prediction model to incorporate not only the local patterns extracted through TCN convolution but also the long-term bidirectional dependencies captured by the Bi-LSTM. Furthermore, to enhance the prediction accuracy, an improved metaheuristic algorithm was applied to optimize both the hyperparameters and the structure of the TCN-Bi-LSTM model. The overall architecture of the proposed model is illustrated in Fig. 1. The main contributions of this study are summarized as follows:

- A birth rate prediction model based on TCN and Bi-LSTM was proposed. The input of the proposed birth rate model includes not only historical birth rate data but also various socioeconomic indicators such as consumption level, divorce rate, and total population, with the aim of capturing the complex factors influencing birth rates.
- The feature vector of the birth rate prediction results output by the Bi-LSTM network is used as an enhanced

feature input to the TCN fully connected layer. This integration enables the birth rate to better understand the relationship between changes in socio-economic indicators and birth rates.

- The hippopotamus position update strategy has been introduced into the NRBO algorithm, and an improved NRBO (INRBO) algorithm has been designed to optimize the hyperparameters of TCN and Bi-LSTM models.
- Finally, the INRBO-TCN-Bi-LSTM model was validated using birth rate datasets from Singapore, Macau, and Luxembourg. The experimental results show that compared with the other four population prediction models, the proposed model reduces the training loss by an average of 95%.

The remaining content of this study is arranged as follows: In Section II of the manuscript, work related to birth rate prediction is reviewed. Section III introduces the proposed INBRO-TCN-Bi-LSTM model. Section IV presents the application results of the proposed model in the problem of birth rate prediction. Finally, Section V summarizes the entire text.

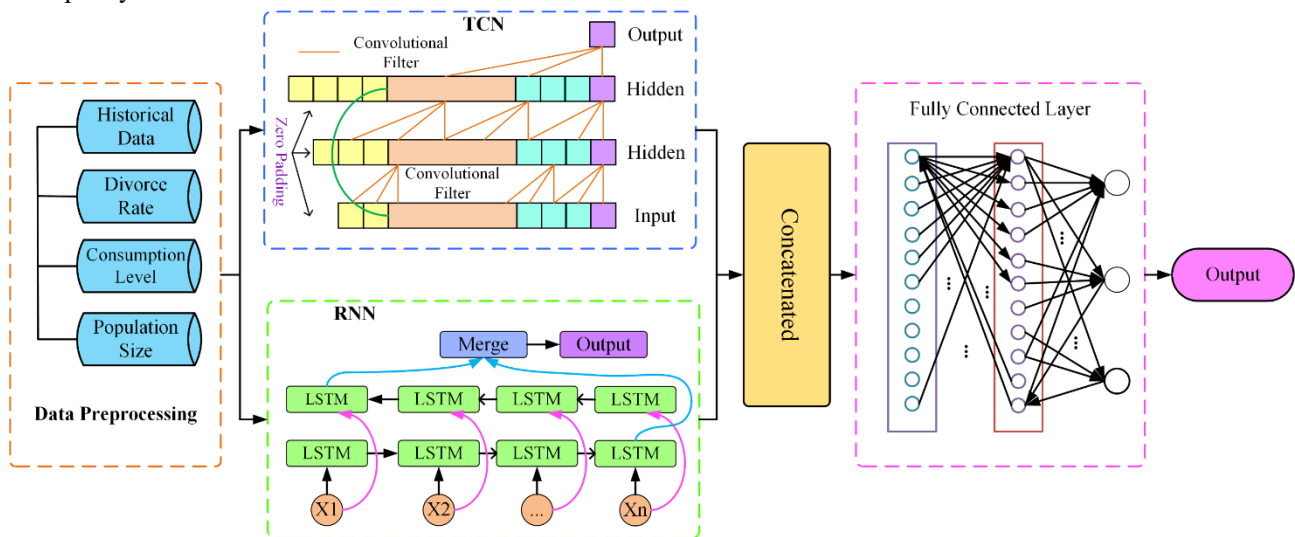


Fig. 1. The structure of the TCN-Bi-LSTM model.

II. RELATED WORK

A. The Prediction Models Based on Deep Learning

For learning based birth rate prediction methods, some studies have explored the application of LSTM in birth rate prediction problems. Alemayehu et al. used a DL model to predict the birth rate trend in Ethiopia [8]. This study is based on population data from Ethiopia from 2000 to 2019, and evaluates the applicability of these methods in the field of birth population prediction by comparing the predictive performance of different models. Tziritidou-Chatzopoulou et al. conducted research on birth rate prediction in Scotland and designed an adaptive machine learning algorithm [9]. The adaptive birth rate prediction framework developed in this study can continuously optimize prediction accuracy by dynamically

adjusting model parameters, thereby capturing subtle changes in socio-economic and environmental factors and the complex relationship between reproductive behaviors.

In response to the challenges of large errors and slow inference speed in traditional autoregressive models, Weber et al. developed an improved RNN architecture for multi-output prediction tasks of unmanned aerial vehicle angular velocity. The improved RNN framework has a dual-layer prediction structure, and experimental results show that the dual-layer prediction structure effectively improves the accuracy of UAV flight attitude prediction [10]. Regarding the dynamic frequency prediction problem in wireless communication networks, Zhang et al. designed a hybrid RNN model based on Graph Convolutional Network (GCN), RNN, and attention mechanism [11]. Unlike classical RNN models, the hybrid

RNN models utilize GCN to capture correlations between predicted objects and dynamically adjust the weights of output hidden states through attention mechanisms. Finally, this study validated the improved hybrid RNN architecture based on real wireless communication spectrum prediction tasks. Ghazi et al. proposed an autoregressive RNN model for predicting time series with irregular time intervals. The improved RNN model introduces an adaptive time embedding architecture, which enables accurate prediction of irregular time interval data [12].

Outliers have a significant impact on traditional RNN prediction frameworks, and the occurrence of outliers in birth rates due to policy adjustments or unexpected events can lead to a decrease in the accuracy of RNN prediction results. Therefore, improving the prediction accuracy of RNN models is a new problem faced by learning-based birth rate prediction models. Kim et al. proposed a prediction framework based on an improved RNN by integrating RNN and statistical methods [13]. The results indicate that the framework effectively reduces the interference of outliers on prediction accuracy by integrating robust statistical methods with LSTM networks. In [14], the authors proposed a reinforcement learning based RNN architecture and successfully applied it to the problem of predicting public opinion information diffusion. The experimental results show that compared with traditional RNN frameworks, the reinforcement learning based RNN architecture has higher accuracy in predicting explosive propagation events. In [15], the authors used Bi-LSTM for predicting the remaining life of aircraft engines. This study effectively fused the forward and backward temporal features of sensor data by stacking a Bi-LSTM and introducing a Gate Recurrent Unit (GRU). The experimental results indicate that the proposed prediction framework is more robust in handling high-dimensional, noisy sensor data. In [16], the authors apply the RNN model to user behavior analysis tasks and verify the potential of RNN in user behavior analysis.

In recent years, with the advancement of deep learning technology, the practical applications of CNN in fields such as healthcare, industry, and biology have been expanded. For example, Yousif et al. proposed a quantum CNN (QCNN) framework for medical image classification tasks, aiming to solve the problem of long training time for traditional CNNs in medical image processing tasks [17]. This study designed an improved strategy based on quantum superposition and entanglement characteristics. The experimental results showed that the improved strategy improved the feature extraction efficiency of the traditional CNN and reduced the training time of the model. In addition, Chen et al. proposed an improved multimodal fusion CNN framework for mechanical fault diagnosis based on multi-source sensor data [18]. This improved framework can dynamically adjust the weights of different modalities, aiming to alleviate the modal conflict problem caused by traditional hard label training. In [19], the authors combine deep learning techniques with time series discretization techniques to address the challenge of manually labeling a large number of classification features in traditional classification tasks.

In [20], the authors also designed a machine learning prediction framework aimed at using interpretable machine learning methods to predict financial risks of enterprises. In

[21], the authors propose a hybrid optimization method combining machine learning and swarm intelligence algorithms for antenna design problems. In this study, machine learning models were used to learn the features of the search space and predict the potential positions of candidate solutions, thereby guiding swarm intelligence optimization algorithms to converge faster to the target area and solving antenna optimization problems containing continuous variables.

B. Meta-Heuristic Optimization Algorithm

As an important technology in the field of intelligent computing, metaheuristic optimization algorithms are applied to solve problems such as robot allocation, antenna design, delivery path planning, and hyperparameter optimization of machine learning models. Akopov designed an improved genetic algorithm (GA) aimed at solving optimization problems in trade interaction problems [22]. Bouali and Alamri conducted research on the modeling problem of photovoltaic systems, aiming to extract the parameters of diodes in photovoltaic cells based on flood algorithms [23]. In [24], the authors designed a starfish optimization algorithm (SFOA) to solve the symmetric traveling salesman problem (TSP). In this study, a discrete SFOA algorithm was designed by defining continuous operators for starfish movement, foraging, and reproduction, providing a new and effective metaheuristic method for solving TSP.

Abdulaziz et al. designed optimization algorithms for collaborative optimization of smart grids to optimize the configuration of components such as electric vehicles (EVs) and distributed static synchronous compensators. In this study, the Hippopotamus Optimization (HO) Algorithm was used to optimize the configuration of these components, aiming to improve the energy efficiency, voltage stability, and economic efficiency of the power grid operation [25]. In [26], the authors propose a novel metaheuristic algorithm, Newton-Raphson-Based Optimizer (NRBO), and the performance of the NRBO algorithm was validated on a standardized optimization test set. Overall, the above research provides a foundation for optimizing the hyperparameters of the TCN-Bi-LSTM model using metaheuristic algorithms.

III. PROPOSED SYSTEM OVERVIEW

Fig. 1 shows the architecture of the proposed hybrid TCN-Bi-LSTM model. The proposed framework takes the historical data of the population born in the region, consumption level, divorce rate, and total population as input features for the mixed TCN-Bi-LSTM model. The historical birth rate within the region is the core time series data and the basic data for birth rate prediction. The consumption level within a region represents the economic situation within the region, and this indicator is usually related to the ability and willingness of families to raise children. The divorce rate represents the stability of the family structure and is one of the important social factors affecting family fertility decisions. The total population provides macro demographic background data.

During the training process of the hybrid TCN-Bi-LSTM model, the input data is simultaneously fed into two parallel neural network modules. Among them, the TCN module uses a series of causal convolutions and dilation convolutions to

efficiently capture local feature patterns and long-term trends in time series. The bidirectional structure of the Bi-LSTM module enables it to fully understand the meaning of each time point in its complete context, thereby capturing long-term dependencies. After obtaining the outputs of TCN and Bi-LSTM, the proposed framework concatenates the output feature vectors V_{TCN} and $V_{Bi-LSTM}$ of TCN and Bi-LSTM at the feature fusion layer. The fused joint feature vector $V_{all,output}$ is fed into multiple fully connected layers to learn the nonlinear relationship between TCN features and Bi-LSTM features, and ultimately mapped to the birth rate prediction target.

A. The TCN-Bi-LSTM Framework

The residual blocks of the TCN model can improve the training efficiency of TCN. In this study, the residual block of the TCN model used Parametric Rectified Linear Unit (PReLU) as the activation function, aiming to improve the prediction accuracy and expressive ability of the TCN model. The definition of the PReLU activation function is as follows [Eq. (1)]:

$$PReLU(x) = \begin{cases} x, & \text{if } x \geq 0 \\ a \times x, & \text{if } x < 0 \end{cases} \quad (1)$$

where, a is the slope on the negative interval. x is the input value.

The proposed hybrid TCN-Bi-LSTM model requires multi-layer convolution operations of the TCN module on the input during the training process. The output of the last convolutional layer of the TCN module in the birth rate prediction model is a sequence V_{TCN} . The definition of sequence V_{TCN} is as follows [Eq. (2)]:

$$\begin{aligned} V_{TCN} &= f_{TCN_Conv}(\vec{x}) \\ &= PReLU(DilatedCasual_Conv(\vec{x}_{t-1}) + \vec{d}) + \vec{x}_t \end{aligned} \quad (2)$$

where, $\vec{x}_t$ is the input variable sequence of time step t . $DilatedCasual_Conv(\cdot)$ is a dilated and causal convolution calculation operation that relies on historical data related to birth rates. $\vec{d}$ is the output layer bias.

In Eq. (3), expansion and causal convolution calculation operations are defined.

$$DilatedCasual_Conv(\vec{x}_{t-1}) = \sum_p w_p \times \vec{x}_{t-\alpha} \quad (3)$$

where, $\{w_p, \forall p \in P\}$ is a filter. P is the filter size. α is a constant.

In the Bi-LSTM module, the input vector is processed by both forward LSTM and backward LSTM simultaneously. After processing the entire input time series, the last hidden state V_{h_t} of the forward LSTM and the last hidden state $V_{h'_t}$ of the backward LSTM are generated. The definitions of hidden states V_{h_t} and $V_{h'_t}$ are as follows [Eq. (4) and Eq. (5)]:

$$V_{h_t} = Q_t \odot \tanh(L_t) \quad (4)$$

$$V_{h'_t} = Q'_t \odot \tanh(L'_t) \quad (5)$$

where, L_t and L'_t are the cell states of forward LSTM and backward LSTM at the t -th time step, respectively. Q_t and Q'_t are the output vectors of the forward LSTM and backward LSTM at the output gate, respectively.

After the input vector is processed through multiple layers of residual blocks and convolutional blocks, the output of the last time step is generated by the TCN module, and a feature vector V_{TCN} that combines local and long-term trends is obtained. For the Bi-LSTM module, the feature vectors V_{h_t} and $V_{h'_t}$ of the hidden state of the forward LSTM and backward LSTM are generated at the last time step, respectively. The final output vector $V_{Bi-LSTM}$ of Bi-LSTM is obtained by concatenating two hidden states V_{h_t} and $V_{h'_t}$. The output vector $V_{Bi-LSTM}$ includes the complete bidirectional contextual information of the input sequence.

Fig. 2 shows the structure of the feature fusion layer designed in this study. In the feature fusion layer, the output vector V_{TCN} of the TCN module and the output vector $V_{Bi-LSTM}$ of the Bi-LSTM module are concatenated along the feature dimension of vector V_{TCN} to generate a more informative joint feature vector $V_{all,output}$. The definition of $V_{all,output}$ is as follows [Eq. (6)]:

$$V_{all,output} = \text{Concatenate}([V_{TCN}, V_{Bi-LSTM}]) \quad (6)$$

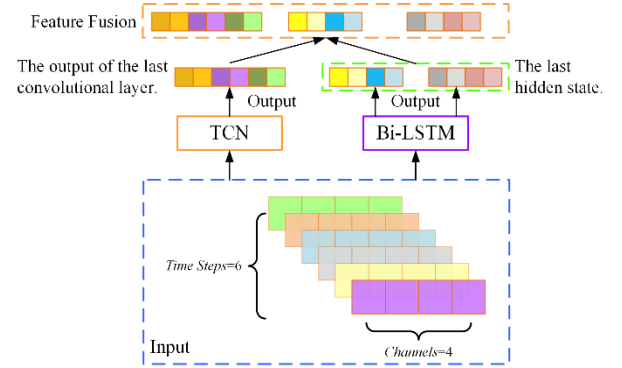


Fig. 2. The structure of the feature vector fusion layer.

After the joint feature vector $V_{all,output}$ is generated in the feature fusion layer, $V_{all,output}$ is input into the fully connected layer of the TCN-Bi-LSTM model to weight and combine all the features in $V_{all,output}$, and calculate the final predicted birth rate. The calculation method for the fully connected layer of TCN-Bi-LSTM is defined in Eq. (7):

$$\hat{y}_{PeoRate} = \sigma(w \times V_{all,output} + d') \quad (7)$$

where, w is the weight matrix of the fully connected layer. d' is the bias vector of the fully connected layer. $\sigma(\cdot)$ is the

activation function. $\hat{y}_{PeoRate}$ is the output variable of the proposed TCN-Bi-LSTM model.

To demonstrate the performance of the designed hybrid TCN-Bi-LSTM framework, the loss function L_{Loss_F} of the TCN-Bi-LSTM model is defined as follows [Eq. (8)]:

$$L_{Loss_F} = \alpha \cdot \frac{1}{G} \sum_{g=1}^G (y_g - \hat{y}_g)^2 + \beta \cdot \frac{1}{G-1} \sum_{g=1}^{G-1} (\hat{y}_{g+1} - \hat{y}_g)^2 \quad (8)$$

where, α and β are weight coefficients. G is the size of the test set, $\forall g \in G$. y_g is the true value, $\hat{y}_g$ is the predicted value.

In addition, Root Mean Square Error (RMSE) is also used to evaluate the performance of the proposed hybrid TCN-Bi-LSTM model. The definition of RMSE is as follows [Eq. (9)]:

$$RMSE = \sqrt{\frac{1}{G} \sum_{g=1}^G (y_g - \hat{y}_g)^2} \quad (9)$$

B. The Improved NRBO (INRBO) Algorithm

Based on the work [26], the hyperparameter optimization module designed in this study using the INRBO algorithm is as follows.

1) *Initialize the population*: In the initialization stage of INRBO, the convolution kernel size, expansion factor, number of residual blocks, regularization parameters, and initial learning rate of TCN and Bi-LSTM are mapped to the decision vectors of the INRBO algorithm. Further, generate a decision matrix based on the population size I_{max} . Fig. 3 shows the mapping relationship between the hyperparameters of TCN and Bi-LSTM and decision vectors, as well as decision matrices.

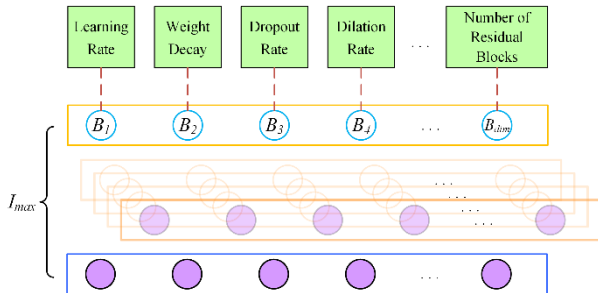


Fig. 3. The mapping relationship between the hyperparameters of TCN and Bi-LSTM and decision vectors.

The decision matrix DM_INRBO is defined in Eq. (10):

$$DM_INRBO = \begin{bmatrix} B_1^1 & B_2^1 & \dots & B_{dim}^1 \\ B_1^2 & B_2^2 & \dots & B_{dim}^2 \\ \vdots & \vdots & \vdots & \vdots \\ B_1^{I_{max}} & B_2^{I_{max}} & \dots & B_{dim}^{I_{max}} \end{bmatrix} \quad (10)$$

where, B_j^i is the decision vector. dim is the dimension of the vector. I_{max} is the size of the population.

2) *Calculate the value of the objective function*: Calculate the objective function for each individual based on the decision matrix DM_INRBO in each iteration. The INRBO algorithm proposed in this study aims to optimize the parameters of TCN and LSTM, so the objective function f_{INRBO} is mainly determined by the accuracy of prediction. In this study, the objective function is composed of two parts, the loss function L_{Loss_F} and the RMSE, which are defined as follows [Eq. (11)]:

$$f_{INRBO} = \lambda_1 \times RMSE + \lambda_2 \times L_{Loss_F} \quad (11)$$

where, λ_1 and λ_2 are weight coefficients, respectively.

3) *Newton-Raphson search strategy*: In the process of updating the decision matrix in INRBO, the position of the optimal solution is also updated. Therefore, each iteration of INRBO generates a new set of hyperparameters for TCN and Bi-LSTM. The process of updating the INRBO decision matrix is as follows. Firstly, generate a random number H within the $[-1,1]$ interval. If $H > 0$, execute the Newton-Raphson search strategy. The definition of Newton-Raphson search strategy is as follows [Eq. (12)]:

$$B_j^i_{new} = B_j^i - \phi \times (B_j^i - B_2^i) \quad (12)$$

where, $B_j^i_{new}$ is the updated decision variable. B_1^i and B_2^i are the decision variables for the optimal and suboptimal solutions generated during the current iteration process, respectively. ϕ is the Newton-Raphson search coefficients.

4) *Hippopotamus foraging search strategy*: If $H \leq 0$, execute the hippopotamus foraging search strategy. The definition of the hippopotamus foraging search strategy is as follows [Eq. (13)]:

$$B_j^i_{new} = B_j^i + rand \times |B_rand_j^i - B_j^i| \quad (13)$$

where, $rand$ is a random number on the $[0,1]$ interval. $B_rand_j^i$ is a decision variable randomly selected from the decision matrix.

IV. RESULTS AND DISCUSSION

In order to evaluate the application effectiveness of the proposed INRBO-based hybrid TCN-Bi-LSTM (INRBO-TCN-Bi-LSTM) framework of birth rate prediction in a small area, the proposed INRBO-TCN-Bi-LSTM framework was validated based on historical data from three countries or regions: Singapore, Macau, and Luxembourg. Fig. 4 displays historical data on birth rates in Singapore, Macau, and Luxembourg. To demonstrate the benefits of the INRBO-TCN-Bi-LSTM framework in predicting birth rates, the TCN-Bi-LSTM model, TCN model, Bi-LSTM model, and GRU model were trained separately to compare with the INRBO-TCN-Bi-LSTM model. Table I shows the relevant parameters of the above models during the training process [27]-[28]. A training set was generated using birth rate data from Singapore, Macau, and Luxembourg for the period 1959 to 2011. Three test sets

were formed based on birth rate data from Singapore, Macau, and Luxembourg from 2012 to 2023, and each test set was tested separately to validate the performance of the birth rate prediction model. In [29], the authors proposed using a Monte Carlo-based machine learning model to predict birth rates and evaluated the model based on the difference between the predicted birth rate values and the true values. Differently, in this study, RMSE, sum of squared errors (SSE), mean squared error (MSE), and mean absolute percentage error (MAPE) were used to evaluate the birth rate prediction model.

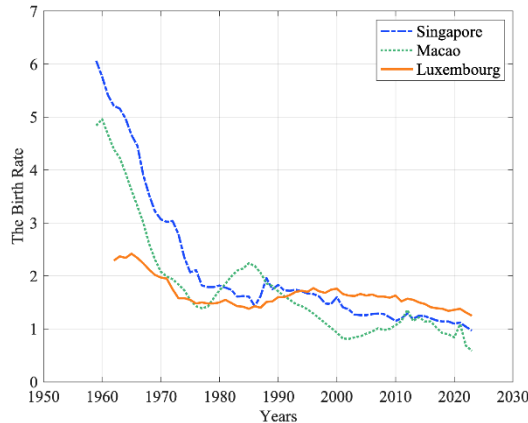


Fig. 4. The historical data on birth rates in Singapore, Macau, and Luxembourg.

TABLE I. TRAINING PARAMETERS FOR BIRTH RATE PREDICTION MODELS

Parameter	Value
Training Epochs	50
Batch Size	30
Time Steps	6
Features	4

In addition, to demonstrate the optimization effect of INBRO on the hyperparameters of the TCN-Bi-LSTM framework, GA [22], SFOA [24], HO [25], and NRBO [26] algorithms were compared with the INRBO algorithm. Table II shows the fitness function values for optimizing the TCN-Bi-LSTM framework using different metaheuristic algorithms.

TABLE II. THE FITNESS FUNCTION VALUES OF DIFFERENT METAHEURISTIC ALGORITHMS

Algorithm	Population Size	Individual Dimension	Fitness Function
GA	70	5	0.0536
SFOA	70	5	0.0497
HO	70	5	0.0517
NRBO	70	5	0.0451
INRBO	70	5	0.0433

From Table II, it can be seen that the fitness function value corresponding to INRBO in optimizing the TCN-Bi-LSTM model is 0.0433. Compared with GA, SFOA, HO, and NRBO algorithms, INRBO has the smallest fitness function value.

This indicates that the hippopotamus serving search strategy effectively improves the convergence accuracy of the NRBO algorithm.

To demonstrate the robustness of the proposed birth rate prediction model, the INRBO-TCN-Bi-LSTM model, TCN-Bi-LSTM model, TCN model, Bi-LSTM model, and GRU model were independently trained 15 times, with each training session lasting 50 epochs. Fig. 5 shows the loss function curves of different birth rate prediction models. Table III shows the training losses of different birth rate prediction models.

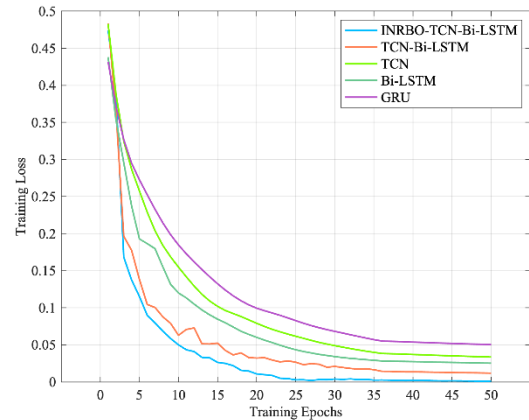


Fig. 5. The loss function curves of different birth rate prediction models.

TABLE III. THE TRAINING LOSSES OF DIFFERENT BIRTH RATE PREDICTION MODELS

Predictive Model	Training Epochs	Training Loss
INRBO-TCN-Bi-LSTM	50	0.001149
TCN-Bi-LSTM	50	0.011808
TCN	50	0.033568
Bi-LSTM	50	0.025266
GRU	50	0.050274

From Table III, it can be seen that the GRU model has the highest training loss, indicating that GRU has the weakest ability to capture temporal dependencies in birth rate prediction problems. TCN and Bi-LSTM, as more complex structures, perform better than GRU models. The training loss of the TCN-Bi-LSTM model is 0.011808, which is one order of magnitude lower than the classical TCN and Bi-LSTM models. The training loss of the INRBO-TCN-Bi-LSTM model is 0.001149, which is one order of magnitude lower than the second-ranked TCN-Bi-LSTM model and outperforms all other models. Therefore, the INRBO algorithm effectively optimized the parameters of the TCN-Bi-LSTM hybrid model during the training process. In summary, compared to the TCN-Bi-LSTM model, INRBO-TCN-Bi-LSTM reduces training loss by 90.3%. Compared to the GRU model, the training loss of INRBO-TCN-Bi-LSTM has been reduced by approximately 97.7%.

Fig. 6 shows the birth rate prediction results of the INRBO-TCN-Bi-LSTM model. Fig. 7 shows the birth rate prediction results of the TCN model. Fig. 8 shows the birth rate prediction results of the Bi-LSTM model. Table IV shows the RMSE,

SSE, MSE, and MAPE indicators of different birth rate prediction models. Fig. 9 shows the MAPE indicators of different birth rate prediction models in three tests. In addition, Table V shows the birth rate prediction values of INRBO-TCN-Bi-LSTM. Table VI shows the birth rate prediction values of TCN model. Table VII shows the birth rate prediction values of Bi-LSTM mode.

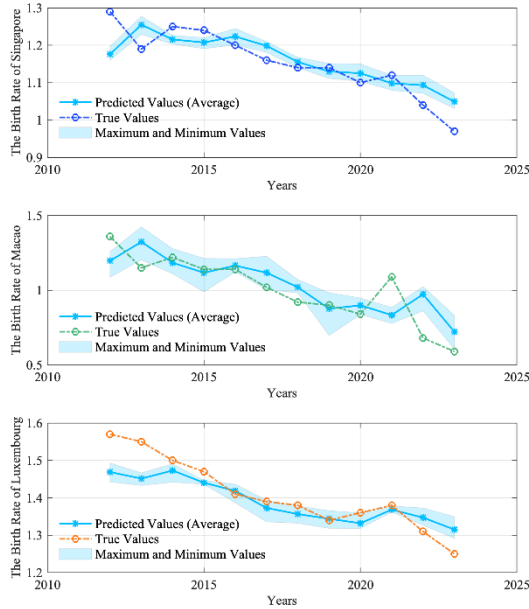


Fig. 6. The birth rate prediction results of the INRBO-TCN-Bi-LSTM model.

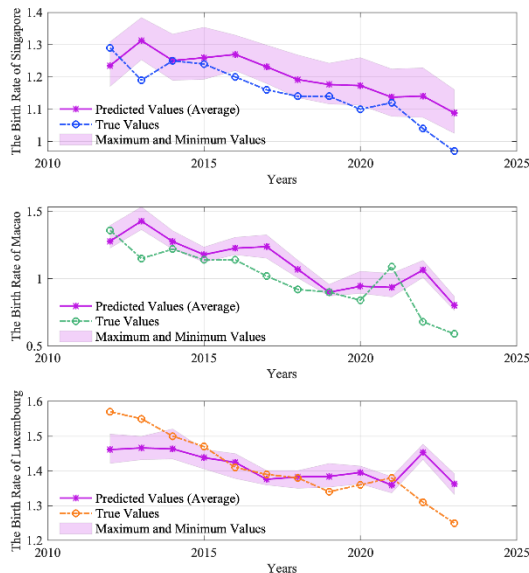


Fig. 7. The birth rate prediction results of the TCN model.

From Table IV, it can be seen that in the three tests, compared with the other four models, the INRBO-TCN-Bi-LSTM model performed the best in the RMSE metric, while the classical GRU model performed the worst. When predicting Singapore's birth rate, the INRBO-TCN-Bi-LSTM model has the lowest RMSE of 0.0722 and MAPE of 5.47%. Compared with the other four models, the INBRO-TCN-RNN

model achieved the highest prediction accuracy. In addition, compared with the RMSE index of the TCN-Bi-LSTM model, the RMSE of the INRBO-TCN-Bi-LSTM model decreased by 15.56%.

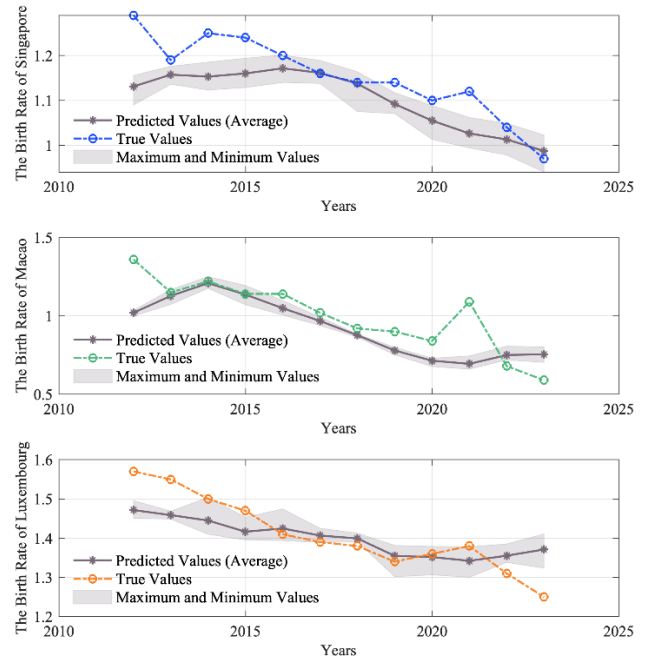


Fig. 8. The birth rate prediction results of the Bi-LSTM model.

TABLE IV. THE INDICATORS OF DIFFERENT BIRTH RATE PREDICTION MODELS

Predictive Model	Regions or Country	RMSE	SSE	MSE	MAPE
INRBO-TCN-Bi-LSTM	Singapore	0.0722	0.0722	0.0052	5.47%
	Macao	0.1442	0.2494	0.0207	12.51%
	Luxembourg	0.0502	0.0302	0.00252	2.84%
TCN-Bi-LSTM	Singapore	0.0855	0.0877	0.0073	6.03%
	Macao	0.1646	0.3312	0.0471	13.60%
	Luxembourg	0.0537	0.0346	0.0155	3.17%
TCN	Singapore	0.0898	0.0967	0.0081	7.44%
	Macao	0.1952	0.4574	0.0381	16.35%
	Luxembourg	0.0579	0.0403	0.00336	3.38%
Bi-LSTM	Singapore	0.1062	0.1354	0.0112	6.29%
	Macao	0.1488	0.2658	0.0221	11.48%
	Luxembourg	0.0619	0.0460	0.00383	3.43%
GRU	Singapore	0.2030	0.1675	0.0140	7.34%
	Macao	0.2030	0.4943	0.0412	19.79%
	Luxembourg	0.0710	0.0605	0.0050	4.92%

Among the three tests, all five birth rate prediction models had the highest error values when predicting the birth rate in Macau. The MAPE of INRBO-TCN-Bi-LSTM is 12.51%, which is better than the TCN-Bi-LSTM, TCN, and GRU models. The performance of Bi-LSTM is better than that of the more complex TCN-Bi-LSTM, indicating that the Bi-LSTM model may have overfitting. When predicting the birth rate in

Luxembourg, the MAPE of the INRBO-TCN-Bi-LSTM model is 2.84%. Compared with the TCN-Bi-LSTM model, the

INRBO-TCN-Bi-LSTM model has decreased RMSE and MAPE by 6.52% and 10.42%, respectively.

TABLE V. THE BIRTH RATE PREDICTION VALUES IN SINGAPORE

Years	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Prediction Models	The Birth Rate											
True Values	1.29	1.19	1.25	1.24	1.20	1.16	1.14	1.14	1.10	1.12	1.04	0.97
INRBO-TCN-Bi-LSTM	1.18	1.25	1.22	1.21	1.22	1.20	1.16	1.13	1.13	1.10	1.09	1.05
TCN	1.24	1.31	1.25	1.26	1.27	1.23	1.19	1.18	1.17	1.14	1.14	1.09
Bi-LSTM	1.13	1.16	1.15	1.16	1.17	1.16	1.14	1.09	1.05	1.03	1.01	0.99

TABLE VI. THE BIRTH RATE PREDICTION VALUES IN MACAU

Years	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Prediction Models	The Birth Rate											
True Values	1.36	1.15	1.22	1.14	1.14	1.02	0.92	0.90	0.84	1.09	0.68	0.59
INRBO-TCN-Bi-LSTM	1.20	1.32	1.18	1.12	1.16	1.12	1.02	0.88	0.90	0.83	0.97	0.72
TCN	1.28	1.43	1.28	1.18	1.23	1.24	1.07	0.90	0.94	0.94	1.06	0.80
Bi-LSTM	1.02	1.13	1.21	1.14	1.05	0.97	0.88	0.78	0.71	0.69	0.75	0.76

TABLE VII. THE BIRTH RATE PREDICTION VALUES IN LUXEMBOURG

Years	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Prediction Models	The Birth Rate											
True Values	1.57	1.55	1.5	1.47	1.41	1.39	1.38	1.34	1.36	1.38	1.31	1.25
INRBO-TCN-Bi-LSTM	1.47	1.45	1.47	1.44	1.42	1.37	1.36	1.34	1.33	1.37	1.35	1.32
TCN	1.46	1.47	1.46	1.44	1.42	1.38	1.38	1.38	1.40	1.36	1.45	1.36
Bi-LSTM	1.47	1.46	1.44	1.42	1.42	1.41	1.40	1.35	1.35	1.34	1.36	1.37

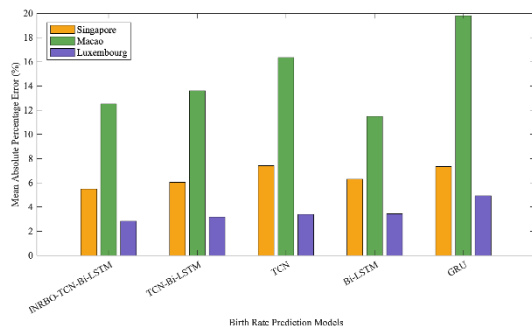


Fig. 9. The MAPE indicators of different birth rate prediction models in three tests.

V. CONCLUSION

This study constructs a birth rate prediction model for small regions based on the TCN-Bi-LSTM hybrid architecture. In addition, the proposed birth rate prediction model introduces three input vectors, including regional consumption level, population size, and divorce rate, aiming to improve the adaptability of the prediction model to policy or economic fluctuations, and make the birth rate prediction results closer to the trend of real birth rates. At the same time, an INRBO algorithm was designed to optimize the hyperparameters of the TCN-Bi-LSTM model, aiming to improve the prediction

accuracy of the birth rate prediction model. The results showed that in the three tests, the INRBO-TCN-Bi-LSTM model showed an average reduction of 11.49% and 9.24% in RMSE and MAPE, respectively, compared to the TCN-Bi-LSTM model. In summary, the proposed INRBO-TCN-Bi-LSTM model has improved the accuracy of birth rate prediction within a small area. The proposed INRBO-TCN-Bi-LSTM model has limitations regarding interpretability and a lack of validation of its generalization capability across diverse time-series forecasting tasks. Future work will focus on addressing these limitations. To this end, we will enhance model interpretability using methods such as feature importance analysis and attention weight visualization. We will also systematically evaluate the model's predictive performance on diverse time series data to verify its robustness.

ACKNOWLEDGMENT

This work was supported by the Luzhou Municipal Research Center for Population and Development.

REFERENCES

- [1] M. Kuroki, "Does free pre-kindergarten increase birth rates? Preliminary evidence from Vermont," *Research in Economics*, vol. 79, no. 1, Art. no. 101040, 2025. <https://doi.org/10.1016/j.rie.2025.101040>
- [2] J. Chen, J. Chi, D. Smith, and M. K. Yuen, "How does digital finance impact birth rates: Evidence from China," *Economic Analysis and*

- Policy, vol. 86, pp. 1945–1965, 2025. <https://doi.org/10.1016/j.eap.2025.05.030>
- [3] Wang, Chian-Yue, and Shin-Jye Lee. "Regional Population Forecast and Analysis Based on Machine Learning Strategy" *Entropy*, vol. 23, no. 6: pp. 656, 2021. <https://doi.org/10.3390/e23060656>
- [4] T M Swathy, K.Ruth Isabels, A. Sindhiya Rebecca, Venubabu Rachapudi, Yousef A.Baker El-Ebiary, Shobana Gorintla and Elangovan Muniyandy, "Game Theory-Optimized Attention-Based Temporal Graph Convolutional Network for Spatiotemporal Forecasting of Sea Level Rise" *International Journal of Advanced Computer Science and Applications (IJACSA)*, 16 (8), 2025. <http://dx.doi.org/10.14569/IJACSA.2025.0160857>
- [5] Prasanthi Yavanamandha and D. S. Rao, "Convolutional Neural Network and Bidirectional Long Short-Term Memory for Personalized Treatment Analysis Using Electronic Health Records" *International Journal of Advanced Computer Science and Applications (IJACSA)*, 16 (1), 2025. <http://dx.doi.org/10.14569/IJACSA.2025.0160136>
- [6] C. Huang, Z. Zhang and L. Wang, "PSOPruner: PSO-Based Deep Convolutional Neural Network Pruning Method for PV Module Defects Classification," *IEEE Journal of Photovoltaics*, vol. 12, no. 6, pp. 1550–1558, Nov. 2022. doi: 10.1109/JPHOTOV.2022.3195099.
- [7] L. Liao, H. Zhou, J. Liu, J. Zhou, D. Zhang and J. Wang, "A PSO-CNN-LSTM Model for Seismic Facies Analysis: Methodology and Applications," *IEEE Access*, vol. 13, pp. 84094–84111, 2025. doi: 10.1109/ACCESS.2025.3566005.
- [8] M. A. Alemayehu et al., "Forecasting birth trends in Ethiopia using time-series and machine-learning models: a secondary data analysis of EDHS surveys (2000–2019)," *BMJ Open*, vol. 15, no. 7, p. e101006, 2025. doi: 10.1136/bmjopen-2025-101006
- [9] Tzitiroidou-Chatzopoulou, Maria, Georgia Zournatzidou, and Michael Kourakos. "Predicting Future Birth Rates with the Use of an Adaptive Machine Learning Algorithm: A Forecasting Experiment for Scotland" *International Journal of Environmental Research and Public Health*, vol. 21, no. 7, pp. 841, 2024. <https://doi.org/10.3390/ijerph21070841>
- [10] D. O. M. Weber, C. Gühmann and T. Seel, "FranSys—A Fast Non-Autoregressive Recurrent Neural Network for Multi-Step Ahead Prediction," *IEEE Access*, vol. 12, pp. 145130–145147, 2024. doi: 10.1109/ACCESS.2024.3473014.
- [11] X. Zhang et al., "A-GCRNN: Attention Graph Convolution Recurrent Neural Network for Multi-Band Spectrum Prediction," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 2, pp. 2978–2982, 2024. doi: 10.1109/TVT.2023.3315450.
- [12] M. M. Ghazi, L. Sørensen, S. Ourselin and M. Nielsen, "CARRNN: A Continuous Autoregressive Recurrent Neural Network for Deep Representation Learning From Sporadic Temporal Data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 1, pp. 792–802, 2024. doi: 10.1109/TNNLS.2022.3177366.
- [13] Y. Kim and K. Park, "Outlier-Aware Demand Prediction Using Recurrent Neural Network-Based Models and Statistical Approach," *IEEE Access*, vol. 11, pp. 129285–129299, 2023, doi: 10.1109/ACCESS.2023.3333030.
- [14] C. Yang et al., "Full-Scale Information Diffusion Prediction With Reinforced Recurrent Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 5, pp. 2271–2283, May 2023, doi: 10.1109/TNNLS.2021.3106156.
- [15] K. Hu, Y. Cheng, J. Wu, H. Zhu and X. Shao, "Deep Bidirectional Recurrent Neural Networks Ensemble for Remaining Useful Life Prediction of Aircraft Engine," *IEEE Transactions on Cybernetics*, vol. 53, no. 4, pp. 2531–2543, April 2023, doi: 10.1109/TCYB.2021.3124838.
- [16] Keling Bi, "Internet of Things User Behavior Analysis Model Based on Improved RNN" *International Journal of Advanced Computer Science and Applications (IJACSA)*, 15 (11), 2024. <http://dx.doi.org/10.14569/IJACSA.2024.0151120>
- [17] M. Yousif, B. Al-Khateeb and B. Garcia-Zapirain, "A New Quantum Circuits of Quantum Convolutional Neural Network for X-Ray Images Classification," *IEEE Access*, vol. 12, pp. 65660–65671, 2024, doi: 10.1109/ACCESS.2024.3396411.
- [18] L. Chen, K. An, D. Huang, X. Wang, M. Xia and S. Lu, "Noise-Boosted Convolutional Neural Network for Edge-Based Motor Fault Diagnosis With Limited Samples," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 9, pp. 9491–9502, Sept. 2023, doi: 10.1109/TII.2022.3228902.
- [19] M. H. Tahan, M. Ghasemzadeh and S. Asadi, "A Novel Embedded Discretization-Based Deep Learning Architecture for Multivariate Time Series Classification," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 4, pp. 5976–5984, April 2023, doi: 10.1109/TII.2022.3188839.
- [20] Wang, Y., Wei, W., Liu, Z., Liu, J., Lv, Y., Li, X., "Interpretable Machine Learning Framework for Corporate Financialization Prediction: A SHAP-Based Analysis of High-Dimensional Data," *Mathematics*, vol. 13, pp. 2526, 2025. <https://doi.org/10.3390/math13152526>
- [21] K. Fu and K. W. Leung, "Machine-Learning-Assisted Swarm Intelligence Algorithm for Antenna Optimization With Mixed Continuous and Binary Variables," *IEEE Transactions on Antennas and Propagation*, vol. 73, no. 3, pp. 1662–1673, March 2025, doi: 10.1109/TAP.2024.3510114.
- [22] A. S. Akopov, "MBHGA: A Matrix-Based Hybrid Genetic Algorithm for Solving an Agent-Based Model of Controlled Trade Interactions," *IEEE Access*, vol. 13, pp. 26843–26863, 2025, doi: 10.1109/ACCESS.2025.3539460.
- [23] Bouali, Y., Alamri, B., "Parameter Extraction for Photovoltaic Models with Flood-Algorithm-Based Optimization," *Mathematics*, vol. 13, pp. 19, 2025, <https://doi.org/10.3390/math13010019>
- [24] M. Aktaş and F. Kiliç, "Discrete Starfish Optimization Algorithm for Symmetric Travelling Salesman Problem," *IEEE Access*, vol. 13, pp. 102675–102687, 2025, doi: 10.1109/ACCESS.2025.3579248.
- [25] M. A. Abdelaziz, A. A. Ali, R. A. Swief, and R. Elazab, "Optimizing energy-efficient grid performance: integrating electric vehicles, DSTATCOM, and renewable sources using the Hippopotamus Optimization Algorithm," *Scientific Reports*, vol. 14, no. 1, pp. 28974–15, 2024, doi: 10.1038/s41598-024-79381-4
- [26] R. Sowmya, M. Premkumar, and P. Jangir, "Newton-Raphson-based optimizer: A new population-based metaheuristic algorithm for continuous optimization problems," *Engineering Applications of Artificial Intelligence*, vol. 128, pp. 107532, 2024, doi: 10.1016/j.engappai.2023.107532
- [27] K. Ohtomo, R. Hamakawa, M. Iisaka and M. Iwahashi, "AM-Bi-LSTM: Adaptive Multi-Modal Bi-LSTM for Sequential Recommendation," in *IEEE Access*, vol. 12, pp. 12720–12733, 2024, doi: 10.1109/ACCESS.2024.3355548.
- [28] V. M. Reddy M, P. S. Pandian and K. P A, "Multiclass Gait Phase Classification From the Temporal Convolutional Network of Wireless Surface Electromyography Measurements," in *IEEE Sensors Letters*, vol. 8, no. 10, pp. 1–4, Oct. 2024, Art no. 7004904, doi: 10.1109/LSSENS.2024.3453558.
- [29] Vanella, P., Hassenstein, M.J., "Stochastic Forecasting of Regional Age-Specific Fertility Rates: An Outlook for German NUTS-3 Regions," *Mathematics*, vol. 12, pp. 25, 2024, <https://doi.org/10.3390/math12010025>