Karadeniz Fen Bilimleri Dergisi, 16(1), 475-496, 2026

Karadeniz Fen Bilimleri Dergisi
The Black Sea Journal of Sciences

DOI: 10.31466/kfbd.1713450
ISSN (Online): 2564-7377
https://dergipark.org.tr/tr/pub/kfbd

Araştırma Makalesi / Research Article

Modelling Adult Obesity in China From 1975 to 2016: A Comparative Analysis of Growth Curves and Time Series Approaches

1975'ten 2016'ya Çin'de Yetişkin Obezitesinin Modellenmesi: Büyüme Eğrileri ve Zaman Serisi Yaklaşımlarının Karşılaştırmalı Analizi

Halil ÇOLAK[1]*

Abstract

This research evaluates four forecasting models—Gompertz, Logistic, Autoregressive Integrated Moving Average (ARIMA), and Long Short-Term Memory (LSTM)—to project adult obesity prevalence in China from 1975 to 2016. Model performance was assessed using R^2, MSE, RMSE, and SSE measures based on World Health Organisation data. The results indicated a continuous and nonlinear increase in obesity, mirroring China's swift urbanisation, economic development, and dietary changes. The Logistic model attained the maximum accuracy ($R^2 = 0.9997$; RMSE = 0.0312), accurately depicting the sigmoidal progression of obesity. The Gompertz model ($R^2 = 0.9961$) effectively represented asymmetric long-term growth, whereas the ARIMA model ($R^2 = 0.8098$) excelled in short-term forecasting but inadequately depicted nonlinear dynamics. The LSTM model ($R^2 = 0.8199$; RMSE = 0.0582) exhibited robust adaptability and temporal learning, as validated by five-fold cross-validation. Research indicates that obesity in China has not yet attained saturation, highlighting the necessity for immediate and evidence-based health treatments. The amalgamation of interpretable growth models with adaptive deep-learning architectures establishes a formidable hybrid framework for predicting chronic disease trends and facilitating evidence-based public health interventions.

Keywords: Adult obesity, Growth curves, Time series, Forecasting, China.

Öz

Bu araştırma, 1975'ten 2016'ya kadar Çin'deki yetişkin obezite yaygınlığını tahmin etmek için dört tahmin modelini değerlendirmektedir: Gompertz, Lojistik, Otoregresif Entegre Hareketli Ortalama (ARIMA) ve Uzun Kısa Süreli Bellek (LSTM). Model performansı, Dünya Sağlık Örgütü verilerine dayanan R^2, MSE, RMSE ve SSE ölçümleri kullanılarak değerlendirilmiştir. Sonuçlar, Çin'in hızlı kentleşmesini, ekonomik gelişimini ve diyet değişikliklerini yansıtan obezitede sürekli ve doğrusal olmayan bir artış olduğunu göstermiştir. Lojistik model, obezitenin sigmoidal ilerlemesini doğru bir şekilde tasvir ederek maksimum doğruluğa ulaşmıştır ($R^2 = 0{,}9997$; RMSE = 0,0312). Gompertz modeli ($R^2 = 0{,}9961$) asimetrik uzun vadeli büyümeyi etkili bir şekilde temsil etmiş, ARIMA modeli ($R^2 = 0{,}8098$) ise kısa vadeli tahminde mükemmel olmuş ancak doğrusal olmayan dinamikleri yetersiz bir şekilde tasvir etmiştir. LSTM modeli ($R^2 = 0{,}8199$; RMSE = 0,0582), beş katlı çapraz doğrulama ile doğrulandığı üzere, güçlü bir uyum ve zamansal öğrenme sergilemiştir. Araştırmalar, Çin'deki obezitenin henüz doygunluğa ulaşmadığını ve acil ve kanıta dayalı sağlık tedavilerine olan ihtiyacı vurgulamaktadır. Yorumlanabilir büyüme modellerinin uyarlanabilir derin öğrenme mimarileriyle birleştirilmesi, kronik hastalık eğilimlerini tahmin etmek ve kanıta dayalı halk sağlığı müdahalelerini kolaylaştırmak için güçlü bir hibrit çerçeve oluşturmaktadır.

Anahtar Kelimeler: Yetişkin obezitesi, Büyüme eğrileri, Zaman serisi, Tahmin, Çin.

[1]Giresun University, Faculty of Sports Science, Giresun, Türkiye

***Sorumlu Yazar/Corresponding Author:** halilcolak1957@gmail.com

Geliş/Received: 03.06.2025 **Kabul/Accepted:** 04.12.2026 **Yayın/Published:** 10.03.2026

Çolak, H. (2026). Modelling Adult Obesity in China From 1975 to 2016: A Comparative Analysis of Growth Curves and Time Series Approaches. *Karadeniz Fen Bilimleri Dergisi*, 16(1), 475-496. https://doi.org/10.31466/kfbd.1713450

1. Introduction

The global incidence of obesity has roughly tripled since 1975, with projections indicating that over 1 billion individuals worldwide would contend with obesity by 2030 (Schreurs et al., 2025). The concerning statistics highlight the immediate necessity for extensive public health measures to combat the escalating obesity epidemic (Das, 2025). Obesity is a chronic, recurrent condition defined by abnormal or excessive body fat and is linked to several problems, including type 2 diabetes and cardiovascular illnesses (Çolak et al., 2003; Çolak & Yiğit, 2017). In summary, obesity-related disorders present a considerable present and future risk to human health (Fu et al., 2025).

A forecasting study indicated that, owing to the persistent obesity epidemic, over 1.31 billion individuals globally are projected to develop diabetes by 2050 (Ong et al., 2023). A separate study indicated that obesity-related cancer cases may increase to 2 million worldwide by 2070, representing 7% of all cancer cases (Soerjomataram & Bray, 2021). Numerous nations, including China, have witnessed consistent rises in obesity rates over recent decades, with projections indicating that by 2030, approximately two-thirds of Chinese adults will be classified as overweight or obese (Wang et al., 2021; Wang et al., 2025b). The data acquired by policymakers and scientists through statistical estimation and artificial intelligence prediction methods is a crucial factor for formulating short- and long-term goals (Çolak & Çolak, 2024).

The Gompertz model, a classical statistical method, is a nonlinear model that necessitates the formulation of the model equation, the selection of suitable initial values for the model parameters, and the application of nonlinear solution techniques that require the development of a solution program (Guo & Wang, 2024; Şenol, 2020a). The Logistic model is a commonly employed strategy within traditional statistical techniques (Şenol, 2020a). Machine learning, a predictive method within artificial intelligence, is gaining popularity daily (Çolak, 2025d; Şenol et al., 2024b). Machine learning is an intricate data-driven methodology that yields precise predictions by discerning the nonlinear relationships between independent variables and the dependent variable, thereby facilitating model development, enhancing computational efficiency, and improving outcome predictions relative to conventional statistical models (Çolak, 2025a; Çolak & Şenol, 2023; Şenol et al., 2024d). The Autoregressive Integrated Moving Average (ARIMA) is the most favoured model among time series methodologies (Çolak, 2025b; Şenol et al., 2024a; Şenol et al., 2025b). Long Short Term Memory (LSTM) is a specialised form of recurrent neural network (RNN) designed to mitigate the vanishing gradients issue and proficiently manage long-term dependencies in sequential data (Miri-Moghaddam et al., 2024).

Literature reviews contain studies pertaining to obesity. In a study focused on Mexico, Reyes-Sanchez et al. (2023) projected obesity trends for the years 2030 and 2040, employing the Gompertz Model method as one of their analytical approaches. Debele et al. (2023) employed Gompertz Model in their study, which also examined obesity in Ethiopia. Carlsson et al. (2020) analysed mortality and life expectancy in obese individuals utilising Gompertz Model in Sweden. Palloni et al. (2015) did a study in Mexico to determine the correlation between obesity and diabetes with Gompertz Model and Logistic Model. In a study conducted in Türkiye, Çolak and Çolak (2024) forecasted the distribution of pre-obesity by gender until 2030 using artificial neural networks (ANN), ARIMA, and Holt-Winters (HW) methodologies. De Gregory et al. (2018) examined obesity data utilising linear regression, Logistic regression, ANN, and decision tree methodologies. Montanez et al. (2017) predicted obesity utilising gradient boosting (GB), generalised linear model, regression trees, k-nearest neighbours (kNN), support vector machine (SVM), random forest (RF), and multilayer perceptron (MLP) methodologies. Zheng and Ruggiero (2017) employed binary Logistic regression, improved decision tree (IDT), kNN, and ANN to forecast obesity among high school students. Jeon et al. (2023) employed MLP, Light gradient boosting (LGB), RF, SVL, LR, and extreme gradient boosting (XGBoost) techniques to forecast obesity. Ferreras et al. (2023) examined machine learning techniques employed in the prediction of obesity and overweight in the literature. Thamrin et al. (2021) forecasted adult obesity in Indonesia utilising Logistic regression, regression trees, and Naive Bayes approaches. Pang et al. (2019) anticipated risk determinants for paediatric obesity in the United States via the XGBoost methodology. In a study conducted in China, Han et al (2019) examined the correlation between obesity and metabolic syndrome utilising LM for overweight and obese people aged 13 to 19 years. Wang et al. (2022b) assessed the risk of overweight and obesity in Chinese children aged 3-6 years utilising SVM, RF, kNN, GB, XGBoost, LGB, and NB methodologies. Zhang et al. (2024) forecasted the emergence of overweight (pre-obesity phase) among high school kids in China utilising FR, GB, XGBoost, CatBoost, and LR. In a separate study, Lin et al. (2023) employed LR, kNN, ANN, DT, RF, GB, and CatBoost to forecast obesity in overweight persons in China. Du et al. (2024) forecasted obesity rates in China utilising RF, XGBoost, SVM, LGB, DT, GB, MLP, kNN, backpropagation neural network (BPNN), and LR. Wang et al. (2022a) favoured the use of LR, DT, SVM, RF, kNN, GB, XGBoost, LGB, and NB techniques for predicting overweight and obesity in preschool children in China. Zhang et al. (Zhang et al., 2022) identified the risk factors linked to obesity in Chinese students aged 6-14 years utilising kNN DT, SVM, RF, LR, soft voting classifier, hard voting classifier, Bernoulli Naive Bayes, Multi Naive Bayes, Gaussian Naive Bayes, XGBoost, LGB, and GB.

This study's primary innovation is in its revelation of the comparative dynamics of obesity rate changes, utilising growth curves through classical statistical methods and artificial intelligence via machine learning techniques, covering the period from 1975 to 2016 for adult obesity trends in China. This methodology integrates the benefits of conventional statistical techniques with artificial intelligence models to achieve more comprehensive, adaptable, and precise forecasts regarding the trajectory of the obesity pandemic. The derived model-based outcomes serve as an essential reference for formulating health policies and devising strategic strategies to address obesity; the data analysis facilitates the execution of innovative treatments in public health practices. Consequently, the study offers both scientific and practical contributions to enhancing current tactics devised to combat the obesity issue and to formulating solution recommendations.

2. Theoretical Method

2.1. Data Set

The dataset used in this study includes obesity data of adult individuals collected for the period 1975-2016 for China, and the data were provided by World Health Organisation (WHO, 2025) (Table 1).

Table 1. Obesity data for China between 1975 and 2016.

Years	Obesity data (%)	Years	Obesity data (%)
1975	0.4	1996	1.7
1976	0.5	1997	1.8
1977	0.5	1998	2
1978	0.5	1999	2.1
1979	0.5	2000	2.3
1980	0.6	2001	2.5
1981	0.6	2002	2.6
1982	0.6	2003	2.8
1983	0.7	2004	3
1984	0.7	2005	3.2
1985	0.8	2006	3.5
1986	0.8	2007	3.7
1987	0.9	2008	3.9
1988	1	2009	4.2
1989	1	2010	4.5
1990	1.1	2011	4.8
1991	1.2	2012	5.1
1992	1.3	2013	5.5
1993	1.4	2014	5.8
1994	1.5	2015	6.2
1995	1.6	2016	6.6

2.2. Gompertz Model

The Gompertz function serves as an effective mathematical instrument for modelling gradual growth or decline processes seen in the early and late phases of specific time intervals (Abhishek et al., 2025). This model is utilised in various domains, including obesity prediction (Reyes-Sánchez et al., 2023). The Gompertz growth curve belongs to the category of Richards sigmoidal growth models, which includes notable models such as the negative exponential, Logistic, and Bertalanffy models. These models, originally conceived within a deterministic framework, have been expanded to incorporate stochastic effects to reconcile the differences between experimental data and theoretical outcomes (Albano et al., 2020). Equation 1 presents the formula for the modified GM (Şenol, 2020b).

$$y = Ae^{\left(-e\left[\frac{\mu_m e(\lambda - t)}{A}+1\right]\right)} \tag{1}$$

In this context, t is the index associated with the observation year, where t=1 corresponds to 1975, t=2 to 1976, and so forth, culminating in t=42 for the year 2016. In the model, A is the upper limit (asymptotic value) that the pertinent variable (obesity rate in this study) can attain. Additional constants in the exponential function (e.g., λ, μ, etc.) act as coefficients that dictate the beginning and intermediate growth rates, along with the system's inclination to approach saturation. The Gompertz curve is often favoured for modelling processes that exhibit gradual change (growth or decline), particularly during the start and terminal phases (Şenol, 2020a).

This function can illustrate the temporal progression of the phenomenon under investigation (e.g., obesity) within a framework that initially exhibits a gradual growth, subsequently accelerates, and ultimately nears a saturation threshold. Thus, time series data about the percentage of adult obesity from 1975 to 2016 can be systematically analysed, commencing from a defined initial point (t=1), by fitting the Gompertz function to yield predictions concerning future trends and potential upper limits.

2.3. Logistic Model

The Logistic growth process is characterised as a continuously differentiable function, initially used in population studies, and is approximately *S* shaped. Over the past fifty years, this function has garnered significant attention not only in demography but also across various academic disciplines and application domains; it has been utilised as an effective approach for modelling the temporal dynamics of system development (Jin et al., 2025). Research utilising this methodology also encompasses obesity estimations (Palloni et al., 2015). The S-shaped curve has similar stages in explaining the changes in obesity rates over time. Initially, a low obesity prevalence is observed in the population when obesity is new or not widespread; this initial value may be greater than zero if there is an effect of previous cycles of increase. Then, under the influence of epidemiological and lifestyle factors, obesity rates exhibit an almost exponential increase towards the midpoint of the cycle, temporarily showing an increasing trend approaching linear growth. In the final stage, with the introduction of factors such as health policies, awareness and other interventions, the rate of increase slows down and shows a decreasing trend until it reaches a high rate that can be defined as the obesity carrying capacity of the society. The equation of the Logistic growth curve is given in Equation 2 (Şenol, 2020a).

$$y = \frac{A}{[1 + e^{\frac{4\mu_m e(\lambda - t)}{A}} + 2]} \tag{2}$$

A: It denotes the upper limit of growth potential (carrying capacity or asymptotic value). In the realm of obesity research, it denotes the maximum threshold that the obesity prevalence in a population can attain.

μ_m: It is a coefficient that dictates the rate of curve growth (or the steepness of the ascent). As the value escalates, the growth rate in the intermediate range also amplifies.

λ: It is a location parameter that modifies the segment on the time axis where the curve exhibits its maximum growth rate or approaches the inflection point. In other words, it identifies the temporal value at which it encounters the median of the growth process.

t: It represents a temporal variable. By inputting the t value for each time unit (e.g., year) into the model, the anticipated value of the obesity rate (or the growing phenomenon under examination) at that specific moment is computed.

4 and +2: Constants that govern the symmetry of the Logistic curve, the position of the inflection point, and the allocation of the growth rate. These constants provide a more realistic and data-compatible modelling of the transition from the initial low growth rate to the rapid increase phase, culminating in the saturation phase. This Logistic model is effective for comprehending the slow development of a rate over time, such as obesity, and for facilitating future forecasts. The "S"-shaped feature of the model encompasses the complete trajectory, beginning with a gradual ascent, followed by a swift escalation in the medium term, and culminating in a deceleration as it nears the upper limit (carrying capacity). Thus, it can furnish essential information to decision-makers for monitoring major public health issues, such as obesity, and for formulating requisite intervention strategies.

2.4. Autoregressive Integrated Moving Average (ARIMA)

The ARIMA model is a commonly employed technique in time series analysis, facilitating predictions by analysing the influence of historical data on future values (Siddique et al., 2025). This paradigm comprises three fundamental components (Şenol et al., 2024a).

AR (Autoregressive): It delineates the correlation between the historical values of the series and the present values. Autoregressive terms posit that the patterns identified in the time series stem from preceding values. In the AR(p) model, the present value of the series is determined by a linear combination of the preceding p values and an error component.

I (Integrated): It encompasses procedures for rendering the series immobile. Numerous real-world time series comprise data with non-constant mean and variance, lacking the characteristic of stationarity. The integration method seeks to render the series stationary by applying a specific level of differencing (d). Thus, the statistical features are maintained in the differenced series across time, facilitating the development of a more reliable model.

MA (Moving Average): It delineates the correlation between the present value and the historical error terms (residuals). The moving average model illustrates how unforeseen shocks or random variations from prior periods influence the present value. In the MA(q) model, the current value of the series is the sum of a linear combination of the error terms from the preceding q periods plus a mean value.

The ARIMA model is represented as ARIMA(p,d,q), where p denotes the order of autoregressive terms, d signifies the order of differencing required for stationarity, and q represents the order of moving average components. Through this structural framework, ARIMA identifies the regularities that arise in the time series by assessing the influence of previous observations alongside the pattern of random error components across time. Consequently, it yields consistent and elucidative outcomes in short- and medium-term projections (Şenol et al., 2024c). The ARIMA model's general equation is presented in Equation 3 (Wang et al., 2025a).

$$Y_t = \phi_1 W_{t-1} + \cdots \phi_p W_{t-p} + e_{t-} \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots \theta_q e_{t-q} \quad (3)$$

In this equation, the time series value at time *t*, Y_t, serves as the fundamental variable of the model μ, representing the mean or constant term of the series, may alternatively be denoted as ϕ_0 or β_0 in certain notations. The AR parameters ϕi elucidate the influence of prior period values (e.g., Y_{t-1}, Y_{t-2}, etc.) on the current period, with the coefficient ϕ_1 serving as the most fundamental illustration of this effect. The value of the series during the previous period *(t-i)*, Y_{t-i}, indicates the historical dependency upon which the AR structure relies. The MA parameters θ_j represent the influence of previous error terms (e.g., ε_{t-1}, ε_{t-2}) on the current value, illustrating the impact of unforeseen variations in the series across time. The error terms are typically presumed to possess a mean of zero and a constant variance, characteristic of white noise. The c term introduces a continuous shift to the overall structure of the model, so enhancing its flexibility, either in conjunction with or independently of μ.

The ARIMA model typically employs a three-stage approach for time series data analysis. Initially, tests are conducted to assess the stationarity of the series; those identified as non-stationary are rendered stationary by applying differences of the requisite order. During the second phase, the parameters of the ARIMA model are estimated. The autoregressive and moving average parameters in the model are identified through the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) graphs; ACF outlines the randomness and correlation framework, whereas PACF facilitates a more nuanced correlation examination. The third and final stage encompasses diagnostic evaluations conducted using the Ljung-Box test. This test assesses the random distribution of residuals and the model's appropriateness. If the model is deemed inadequate, the three processes are reiterated until an acceptable fit is attained (Singh et al., 2025; Singh et al., 2020).

2.5. Long Short-Term Memory (LSTM)

LSTM, in contrast to traditional feedforward neural networks, is a distinct RNN architecture characterised by feedback connections. This structure is essential for capturing long-term dependencies and properly retaining historical information within the model. Due to its capacity to handle both individual data points and sequential data sequences, RNN guarantees the retention of contextual information, particularly in sequential datasets like time series. Consequently, LSTM design can yield more precise and reliable predictions by considering the historical correlations within the data (Hochreiter & Schmidhuber, 1997; Tamilselvi et al., 2024; Xie et al., 2023). The fundamental methodology of LSTM is substituting the neurones in the hidden layer of traditional RNNs with an LSTM architecture that utilises memory cells. The LSTM architecture comprises three gate components designed to efficiently capture long-term dependencies and regulate information flow: the input gate, forget gate, and output gate. The input gate governs the incorporation of new information into the cell state, the forget gate facilitates the removal of superfluous or detrimental information from the prior cell state, and the output gate manages the transmission of information derived from the cell state to subsequent layers. This framework facilitates the efficient management of both short-term and long-term memory, hence enhancing the precision of modelling patterns and dependencies in sequential data, including time series (Huang et al., 2025). Figure 1 illustrates the architecture of the LSTM network.

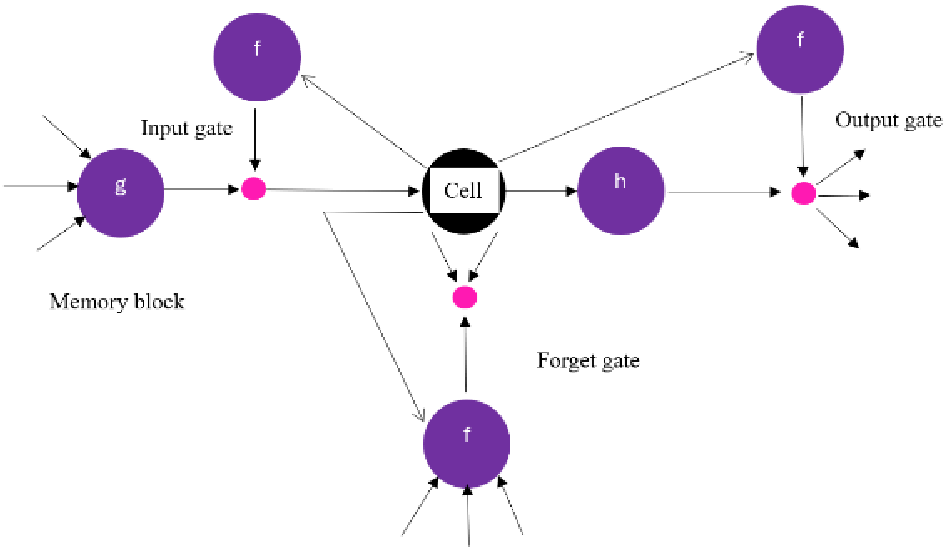

Figure 1. LSTM network structure.

Equations 4-9 contain the general formulas for LSTM (Singh et al., 2024).

$$forget\ gate: f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \quad (4)$$

$$input\ gate: i_t = \sigma(W_i \times [h_{t-1}, x_i] + b_i) \quad (5)$$

$$candidate\ cell\ state: \tilde{C} = tanh(W_c \times [h_{t-1}, x_i] + b_i) \quad (6)$$

$$update\ cell\ state: C_t = f_t \times [C_{t-1} + i_t] + b_i \times \widetilde{C_t} \quad (7)$$

$$output\ gate: O_t = \sigma(W_o \times [h_{t-1}, x_i] + b_o) \quad (8)$$

$$output\ calculation: h_t = O_t \times \tanh\,(C_t) \quad (9)$$

In the equations, the x_t term denotes the input data of the LSTM network architecture. The σ function in the model functions as the sigmoid activation function, whilst the tanh function denotes the hyperbolic tangent activation function. These activation capabilities are essential for regulating the information flow and transformation inside LSTM cells. The parameters wo, w_f, w_c, and w_i are designated as bias terms utilised in the computation of the pertinent gates and cell states. These terms enhance the model's learning flexibility, facilitating a more precise representation of the link between inputs and cell states.

2.6. Model Error Performance Measures

To determine the error performance of the prediction models, mean squared error (MSE), root mean squared error (RMSE), sum of squared errors (SSE), and coefficient of determination (R^2) metrics were preferred (Equations 10-13) (Şenol et al., 2021).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \quad (11)$$

$$SSE = \sum_{i=1}^{n}(Y_i - \hat{y}_i) \quad (12)$$

$$R^2 = 1 - [\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2}] \quad (13)$$

In these equations,

$\hat{y}_i$: The estimated (predicted) value of the relevant observation by the model.

y_i: Represents the actual or expected value.

Y_i: The mean of all observed values yi across the dataset.

n: Represents the total number of observations in the data set examined (Şenol et al., 2025a).

2.7. Establishing Hyperparameters for Time Series Models

This work employed LSTM architecture to predict univariate time series data with great precision (Çolak, 2025c). Prior to initiating the study, preprocessing steps were implemented on the raw dataset to render it appropriate for the LSTM model. Data values were normalised to a range of 0 to 1 utilising MinMaxScaler to enhance the efficacy of neural network training. The lookback value was established at 3 to generate the model's time series input, enabling the model to utilise three preceding time steps to forecast the value at time t+1. The preprocessed dataset was partitioned into 80% for training and 20% for testing, maintaining temporal sequence to accurately evaluate the model's generalisation capability.

The model features a two-layer stacked architecture to reflect hierarchical dependencies within the time series. The initial LSTM layer comprises 100 units and is set with return_sequences=True to transmit outputs to the subsequent layer; the second LSTM layer, consisting of 50 units, is tasked with acquiring advanced temporal features and is configured with return_sequences=False to relay only the final state vector to the dense layer. A dropout rate of 0.2 was implemented on the outputs of both LSTM layers to mitigate the network's propensity to overfit. The batch size remained fixed at 1 during the training phase, ensuring that weight updates were tailored to each individual sample due to the sequential characteristics of time series modelling. The model underwent training for 80 epochs, with the MSE as the loss function, a conventional choice for regression tasks, while the Adam optimisation method, known for its rapid convergence and adaptable learning rate, was employed.

A sequential 5-fold K-fold cross-validation was conducted on the training dataset to evaluate the model's training stability. The ultimate performance assessment and model efficacy were determined using R^2, MSE, RMSE, and SSE, which are standard metrics in the literature.

The ARIMA model, which employs historical values from the series for forecasting, was utilised as the principal time series prediction approach in the study. The primary grounds for selecting ARIMA were its robust theoretical framework and its significant alignment with the observed properties of the series (Shumway & Stoffer, 2017). The ARIMA model parameters were meticulously established according to the statistical characteristics of the series. The stationarity test, an essential prerequisite of the model, was conducted utilising the Extended Dickey-Fuller (ADF) test. The initial series was non-stationary, and the p-value (9.49×10^{-19}) along with the ADF statistic (-10.5244), both being below the 1% threshold value (-3.6155), conclusively indicated that the series achieved stationarity following two differencing operations. The d parameter was established at 2. Information criteria were calculated to assess model parsimony and overall goodness of fit. The resulting ARIMA(2,2,1) model attained an AIC of −89.8484 and a BIC of −84.1124, establishing it as the most parsimonious and optimal model among the evaluated choices.

The ACF and PACF graphs of the stationary series with a difference of $d = 2$ were examined visually (Figure 2). In the PACF graph, the truncation of the first substantial lag in the second step established the p parameter at 2; in the ACF graph, the truncation of the first lag in the initial step established the q parameter at 1. The ARIMA model configuration was established as (2, 2, 1).

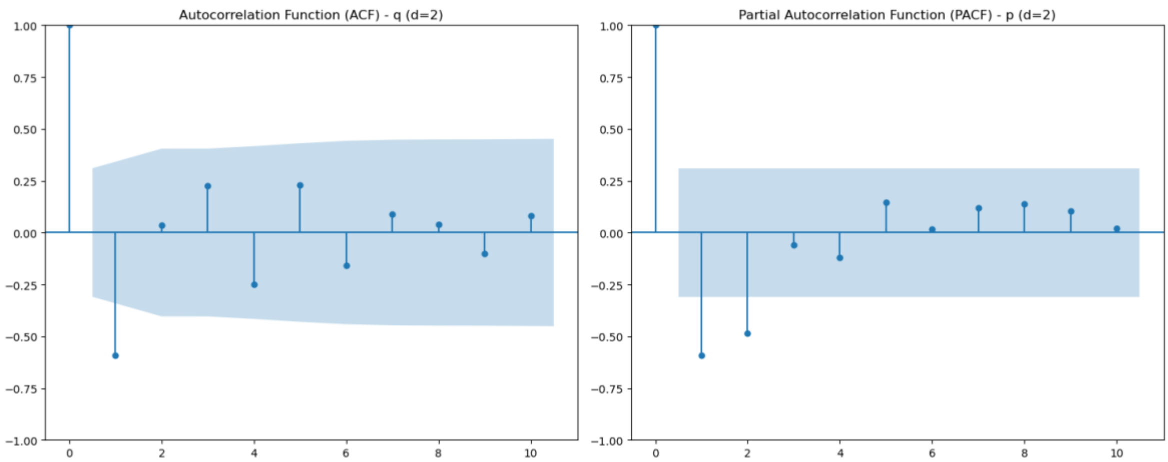

Figure 2. ACF and PACF graphs for the ARIMA model.

2.8. Software and Computational Environment

All computational evaluations were performed on a Windows 11 operating system utilising a machine with an Intel® Core™ i7-11700H CPU, 32 GB RAM, and an NVIDIA RTX 3050 GPU. The Gompertz and Logistic models were executed in MATLAB R2019a, employing its nonlinear regression and curve-fitting capabilities. Analyses for the ARIMA and LSTM models were conducted in Python 3.11 using the Spyder environment, utilising common scientific and machine-learning libraries.

3. Results and Discussion

3.1. Evaluation of Obesity Rate in Chinese Adults

The longitudinal variation in obesity prevalence among Chinese people has been thoroughly documented from 1975 to 2016 (Table 1). Commencing at a minimal rate of 0.4% in 1975, the process experienced modest growth in the subsequent years, attaining 0.6–0.8% during the 1980s. Obesity rates significantly escalated in the 1990s, rising from 1.1% to 2% throughout that period. The growth rate intensified in the early 2000s, fluctuating between 2.3% and 3%, and persisted in its ascent from 2009, culminating at 6.6% in 2016. The statistics indicate a sustained, systematic rise in obesity prevalence in China, potentially attributable to economic, sociocultural, and lifestyle reasons.

3.2. Forecasting the Obesity Percentage in Chinese Adults

The estimations derived from the data in Table 1 utilised the Gompertz model, Logistic model, ARIMA, and LSTM. The black dots in Figure 3 denote the observed values for obesity rates from 1975 to 2016, and the blue curve illustrates the estimations derived from the Gompertz model and Logistic model. The Gompertz model is notable for depicting growth processes that first occur at a low rate, subsequently accelerate, and ultimately converge towards a specific upper limit over time. In this context, the model demonstrates a positioning closely aligned with the data in the initial years, effectively captures the acceleration noted during the 1990s, and progressively decelerates the observed rising trend in the latter portion, approaching potential saturation. The results indicate that the Gompertz model suitably fits obesity data and can estimate temporal rate variations within a comprehensive framework.

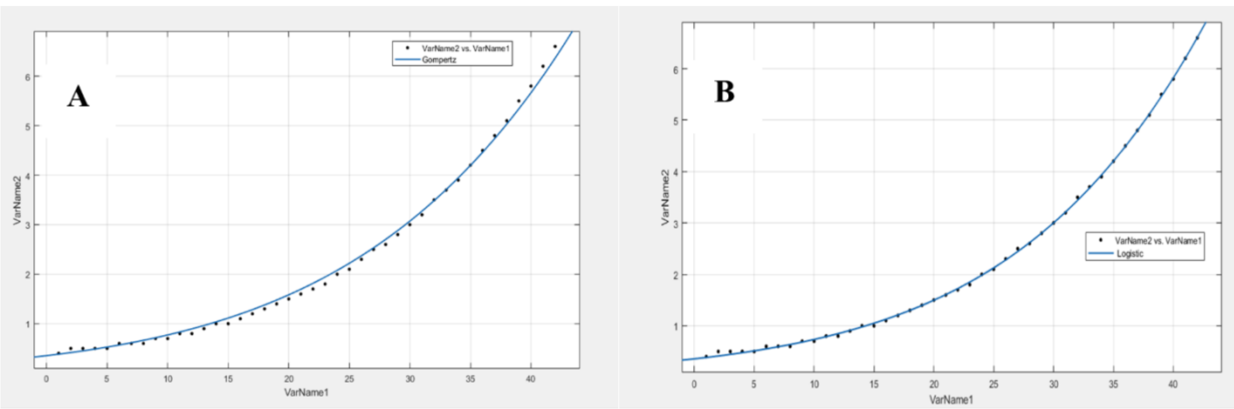

Figure 3. Gompertz growth curve graph (A) and Logistic growth curve graph (B).

The black dots signify the observations of obesity rates from 1975 to 2016, and the blue curve illustrates the values projected by the Logistic model. The *S*-shaped growth, a defining aspect of the Logistic model, indicates that an initial slow increase accelerates with time, gaining velocity throughout the intermediate phase, before subsequently decelerating as it nears saturation levels. The visual analysis reveals that the model accurately reflects the pronounced acceleration witnessed since the 1990s, followed by a deceleration in the latter phase, nearing a saturation point. The Logistic model comprehensively represents the fluctuations in obesity rates, encompassing the beginning, acceleration, and deceleration phases, while effectively forecasting the overarching patterns in the dataset. The numbers in Table 2 present the estimated parameters of the Logistic and Gompertz models, indicating that each model addresses growth dynamics distinctively.

Table 2. Forecasted parameter values of growth curves

Model-Metrics	A	λ	μ
Logistic	46	39	0.8395
Gompertz	1702	114.9	6.212

The A= 46 value in the Logistic model represents a parameter that forecasts the maximum obesity rate, or carrying capacity, at roughly 46%. λ=39.07 is a parameter that modifies the placement of the growth curve along the temporal axis (for instance, the year when the inflection point is near); μ=0.8395 is the coefficient that dictates the growth rate inside the model. The A=1702 value in the Gompertz model does not possess a straightforward interpretation such as 1702%; rather, it represents the highest limit of the Gompertz equation and serves as the scale parameter that defines the curve's shape. λ=114.9 modifies the temporal location and establishes the duration during which the most rapid escalation or transitional phase of the obesity rate occurs. μ=6.212 is the growth coefficient that delineates the gradient of the increase and the acceleration in the central segment of the model.

While both models may demonstrate an *S*-shaped growth, the mathematical distinctions between the Logistic and Gompertz functions result in differing magnitudes and interpretations of their parameters. Consequently, although the A=46 value in the Logistic model can be directly interpreted as the carrying capacity or maximum obesity rate, the A=1702 parameter in the Gompertz model should not be construed similarly; rather, it should be recognised as a factor that influences the model's scale and curve shape.

Figure 4 illustrates the observed and modelled changes in obesity prevalence in China from 1978 to 2016, utilising four modelling methodologies: Gompertz, Logistic, ARIMA, and LSTM, alongside actual historical data. The raw dataset spans from 1975 to 2016; however, the analysis commences in 1978 due to the LSTM model's necessity for a backdated window that uses multiple prior observations as input for sequence learning.

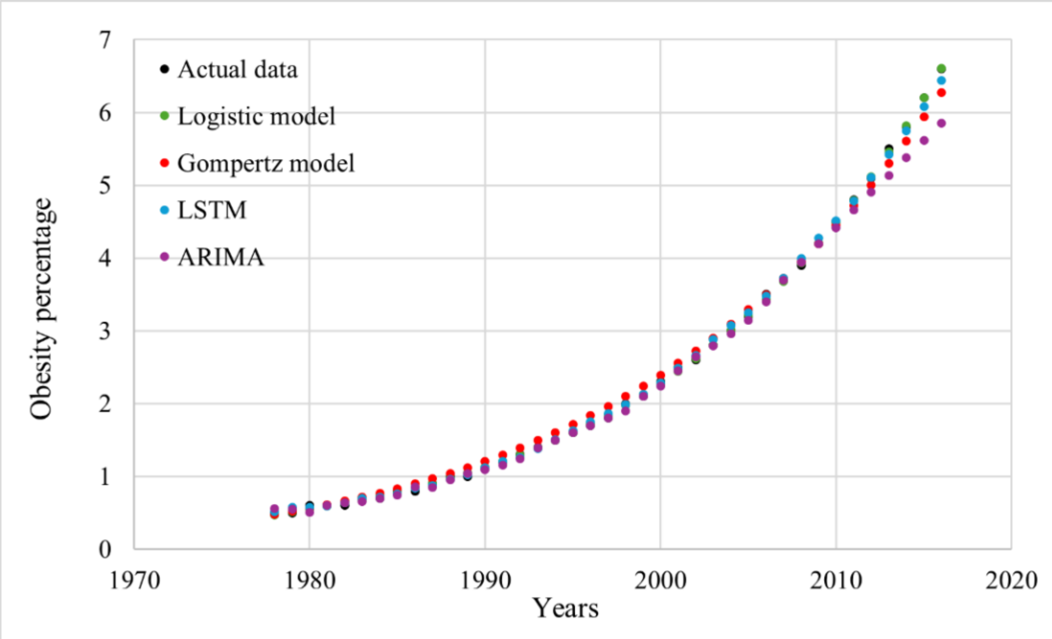

Figure 4. Forecasting of obesity prevalence in China between 1978 and 2016 using Gompertz, Logistic, ARIMA and LSTM models.

The empirical data (black dots) indicate a persistent and rapid rise in obesity prevalence, from under 1% in the late 1970s to exceeding 6% in 2016. This trend illustrates the significant nutritional, social, and lifestyle transformations that have occurred alongside China's swift urbanisation, industrialisation, and culinary westernisation in the last forty years.

The Gompertz model (red) offers a consistent and accurate fit across the observation period, effectively representing the asymmetric growth pattern characteristic of chronic public health issues, especially where prevalence initially rises gradually, accelerates during economic growth and urban lifestyle adoption, and ultimately reaches a plateau. The Gompertz function's adaptability in representing decelerating growth accounts for its excellent alignment with empirical data from 1990 to 2010, a timeframe characterised by rising caloric consumption and diminishing physical activity among China's urban populace.

The Logistic model (green) accurately reflects the overall sigmoidal trajectory but tends to overstate obesity levels post-2010, indicating that the saturation point (carrying capacity) was attained sooner than recorded. This overestimation indicates that, although the prevalence of obesity has risen markedly, China has not yet reached the complete stabilisation phase characteristic of industrialised nations, and public health regulations and awareness initiatives are starting to curtail further escalation.

The ARIMA model (purple) accurately catches short-term fluctuations and efficiently monitors the initial phases (1978–2000). Nonetheless, as a linear time series approach, ARIMA's forecasting efficacy declines when confronted with nonlinear dynamics arising from the intricate interplay of demographic, economic, and behavioural factors influencing obesity.

Conversely, the LSTM model (blue) adeptly captures the nonlinear and memory-dependent characteristics of obesity progression, demonstrating a striking alignment with actual data throughout the whole-time span. The model's architecture, which integrates long-term temporal dependencies, enables the detection of nuanced trends in the acceleration of obesity prevalence amid China's swift urban and economic transition.

The comparison indicates that classical growth models (Gompertz and Logistic) provide interpretability and epidemiological understanding of long-term saturation behaviour, whilst data-driven methods like LSTM excel in forecast accuracy and adaptability. The continuous increase in obesity rates in China underscores the ongoing epidemiological transition towards a chronic illness burden and emphasises the necessity for comprehensive public health interventions, nutritional education, and lifestyle modification initiatives.

This study's findings reveal a constant, nonlinear rise in adult obesity incidence in China from 1978 to 2016, aligning with global patterns that characterise obesity as an enduring and pervasive public health concern. The findings from the Gompertz, Logistic, ARIMA, and LSTM models collectively affirm that the incidence of obesity in China has evolved from an uncommon condition in the late 1970s to a prevalent condition in the 21st century. This development aligns with China's swift economic expansion, urbanisation, and alterations in eating habits and physical activity behaviours (Fardet et al., 2021; Monda et al., 2007; Popkin, 2014b).

The Gompertz and Logistic growth models accurately represented the long-term sigmoidal trajectory of obesity proliferation. The Gompertz model exhibited superior fit throughout the acceleration phase (1990–2010), aligning with its asymmetric structure that resulted in gradual beginning development, succeeded by a more pronounced increase and a subsequent plateau (Tjørve & Tjørve, 2017; Zhao et al., 2020). The Logistic model, while theoretically analogous, frequently overestimated prevalence post-2010 and indicates that China has not yet attained the saturation phase evident in other industrialised nations, like the United States and the United Kingdom (Jaacks et al., 2019). These contradictions underscore that classical growth models are relevant for interpreting long-term trends but may undervalue nonlinear variations and sociocultural disruptions.

The ARIMA model effectively depicts short-term variations in the initial decades when annual fluctuations are relatively minor; nevertheless, its linear assumptions constrain its capacity to account for the behavioural and socioeconomic factors affecting obesity (Tomov et al., 2023). Conversely, the LSTM model exhibited enhanced predicted accuracy, showing virtually perfect concordance with empirical data across the whole-time frame. This outcome underscores the model's efficacy in tackling nonlinear, memory-dependent, and time-delayed processes, which are especially pertinent to chronic public health issues like obesity, characterised by progressive behavioural adaptation and intergenerational transmission (Pham et al., 2016; Weng et al., 2017).

From an epidemiological standpoint, the upward trend indicated by the LSTM model exemplifies the nutritional transition theory, which associates economic development with dietary shifts towards high-calorie, ultra-processed foods and less physical activity (Popkin, 2014a). From 1980 to 2015, China's per capita calorie consumption rose by almost 30%, concurrently accompanied by a decline in physical labour attributable to industrialisation and technical advancement (Sun et al., 2023). The prevalence of obesity has escalated more swiftly in urban regions compared to rural areas, exacerbating socioeconomic disparities in health consequences.

A comparison of statistical and machine learning techniques reveals that data-driven algorithms like LSTM can enhance traditional growth curves by offering adaptable, high-resolution predictions that accommodate the nonlinear characteristics of epidemiological data. Nonetheless, the interpretability of LSTM is constrained in comparison to parametric models like Gompertz and Logistic, which produce epidemiologically significant characteristics like carrying capacity and growth rate. Consequently, the integration of traditional and AI-based methodologies can yield a balanced framework that merges interpretability with predictive efficacy (Şenol & Çolak, 2025).

The findings indicate that China's obesity pandemic has not yet stabilised from a policy standpoint. Should prevailing patterns persist, the nation may encounter a substantial rise in obesity-related comorbidities, such as type 2 diabetes, cardiovascular disease, and certain malignancies (Liu et al., 2022; Zhang et al., 2019)Urgent implementation of effective public health interventions, including taxing sugary beverages, food labelling rules, community-based fitness programmes, and early health education, is essential to mitigate this trend (Teng et al., 2019).

This study illustrates that hybrid forecasting frameworks, integrating classical growth theory and contemporary deep learning, can function as effective decision-support instruments for national health systems. The methodological approach outlined below is applicable to various chronic illnesses characterised by temporal progression, facilitating data-driven policy development and preventive measures.

3.3. Error Performance Analysis Results of Forecasted Models

Table 3 presents the comparative error performance of four prediction models used to model adult obesity prevalence in China. R^2 was evaluated using MSE, RMSE, and SSE.

Table 3. Comparison of error performance among Logistic, Gompertz, ARIMA, and LSTM models for estimating adult obesity prevalence in China.

Model-Metrics	R^2	MSE	RMSE	SSE
Logistic	0.9997	0.000973	0.031185	0.03956
Gompertz	0.9961	0.013363	0.1156	0.5212
ARIMA	0.8098	0.142733	0.3778	1.2846
LSTM	0.8199	0.0041	0.0582	0.0246

The Logistic model exhibited the highest overall accuracy, with an R^2 of 0.9997, an MSE of 0.00097, an RMSE of 0.0312, and an SSE of 0.0396. The results indicate an almost flawless alignment between the modelled and actual data, highlighting the Logistic growth curve's capacity to depict the initial exponential growth phase, subsequent deceleration, and eventual saturation phase of obesity from an epidemiological viewpoint. Consequently, it emerges as a robust model selection for analysing obesity within the framework of long-term behavioural and biological limitations (Pelinovsky et al., 2020).

The Gompertz model yielded commendable results, attaining an R^2 of 0.9961 and an RMSE of 0.1156, albeit with a little elevated residual variance relative to the Logistic model. This aligns with the asymmetric configuration of the Gompertz function, which more accurately depicts the steady increases in the early stages and the slowdown patterns in the latter stages typical of chronic disease dissemination. The elevated MSE (0.0134) and SSE (0.5212) results indicate a potential underestimation of recent accelerated trends in obesity incidence, especially during times of significant socioeconomic change.

Conversely, the ARIMA model had inferior prediction performance, with $R^2 = 0.8098$ and RMSE = 0.3778, signifying considerable residual error and diminished ability to capture the nonlinear dynamics of obesity growth. Although ARIMA adeptly captures short-term variations, its linear autoregressive framework constrains its capacity to represent the intricate behavioural, environmental, and policy-related interactions that affect obesity rates (Özen & Özen, 2025).

The LSTM model, engineered to capture temporal dependencies and nonlinear interactions, exhibited competitive performance ($R^2 = 0.8199$, RMSE = 0.0582, SSE = 0.0246) and markedly reduced total error metrics compared to ARIMA. The MSE value of 0.0041 underscores its proficiency in recognising sequential patterns and maintaining temporal continuity within the sample.

Five-fold cross-validation was conducted to assess the generalisation capability of the LSTM network. The model exhibited variable but strong performance across folds, with R^2 values ranging from 0.4074 to 0.9828 and RMSE values ranging from 0.0224 to 0.0851. The mean fold performance was quantified as $R^2 = 0.8199$, MSE = 0.0041, and SSE = 0.0246, indicating a substantial level of reliability and consistent learning behaviour. Fold 3, with an R^2 of 0.9828, demonstrated near-perfect concordance between predicted and actual values, confirming the network's capacity to accurately describe nonlinear temporal evolution upon achieving optimal weight convergence.

These findings indicate that classical growth curves (Logistic and Gompertz) offer excellent interpretability and a smooth fit for long-term obesity projections, whereas deep learning methods (LSTM) are better suited to the nonlinear dynamics and temporal dependencies present in complex public health data. The ARIMA model, despite its theoretical robustness, is constrained by its linear framework and vulnerability to nonstationarity. Consequently, integrating growth curve interpretability with the adaptability of neural networks may yield the most resilient framework for forecasting chronic health issues, such as obesity, in developing populations.

4. Conclusion

This study conducted a comparative evaluation of four predictive methodologies (Gompertz, Logistic, ARIMA, and LSTM) to model adult obesity trends in China from 1975 to 2016. The findings indicated a consistent, nonlinear rise in obesity rates, mirroring the nation's swift urbanisation, the westernisation of eating patterns, and socioeconomic changes.

- The Logistic model attained the maximum accuracy (R^2 = 0.9997; RMSE = 0.0312), effectively representing the sigmoidal trajectory of obesity progression: initial exponential increase, subsequent acceleration, and final saturation.
- The Gompertz model demonstrated a robust match (R^2 = 0.9961), aligning with China's industrial growth and shifts towards a sedentary lifestyle, especially during the acceleration period from 1990 to 2010.
- The ARIMA model, while suitable for short-term forecasting (R^2 = 0.8098), encountered difficulties with non-linear behavioural and demographic interactions because of its linear structure.
- The LSTM model (R^2 = 0.8199; RMSE = 0.0582) has shown significant adaptability and effectively captured long-term dependencies due to its memory-based architecture. Five-fold validation affirmed the model's strong temporal generalisation, enhancing its capacity for intricate epidemiological forecasting.

These findings underscore the synergistic advantages of classical and AI-based models: growth curves (Logistic, Gompertz) deliver interpretability and epidemiological relevance, whereas deep learning (LSTM) affords flexibility and enhanced nonlinear predictive capability. From a policy standpoint, forecasts suggest that the obesity epidemic in China has not yet attained saturation. In the absence of definitive public health measures, like levies on sugary beverages, enhanced nutrition labelling, and nationwide physical activity programmes, the prevalence of obesity-related comorbidities (type 2 diabetes, cardiovascular illnesses, and malignancies) is expected to rise. The study methodologically introduces a cohesive predictive framework that merges the interpretability of growth curves with the adaptability of deep learning, offering a scalable model for chronic disease prediction. This hybrid model can facilitate evidence-based policymaking, enhance epidemiological surveillance, and direct proactive actions within national and global health systems.

In summary, the integration of statistical growth theory with computational intelligence offers a robust and comprehensible foundation for the analysis and prediction of persistent public health issues, such as obesity, hence enhancing global health management from both scientific and practical viewpoints.

Conflict of Interest Statement

I declare that there are no conflicts of interest associated with this manuscript. The research was carried out independently and objectively, without any financial, personal, or professional relationships that could be perceived as influencing the study’s findings or interpretations.

Funding Statement

This research received no external funding. All analyses and results were obtained using the author’s own resources, without any financial support from public, commercial, or not-for-profit organisations.

References

Abhishek, A., Deb, S. D., Jha, R. K., Sinha, R., & Jha, K. (2025). Ensemble learning using Gompertz function for leukemia classification. *Biomedical Signal Processing and Control*, *100*, 106925.

Albano, G., Giorno, V., Román-Román, P., Román-Román, S., Serrano-Pérez, J. J., & Torres-Ruiz, F. (2020). Inference on an heteroscedastic Gompertz tumor growth model. *Mathematical Biosciences*, *328*, 108428.

Carlsson, L. M., Sjöholm, K., Jacobson, P., Andersson-Assarsson, J. C., Svensson, P.-A., Taube, M., Carlsson, B., & Peltonen, M. (2020). Life expectancy after bariatric surgery in the Swedish obese subjects study. *New England Journal of Medicine*, *383*(16), 1535-1543.

Çolak, E. (2025). *Türkiye’nin coğrafi bölgelerine göre sığır gübresi kaynaklı biyogaz potansiyelinin makine öğrenmesi ile 2035 yılına kadar tahmin edilmesi*

Çolak, H. (2025a). Forecasting the number of physical education teachers to graduate in Türkiye by 2030 utilizing artificial neural networks and time series analysis. *Journal of Physical Education Sports Health and Effort*, *1*, 24-40.

Çolak, H. (2025b). Future Projections of Elderly Obesity in the United States Using Time Series Models. *Obesity Medicine*, 100627.

Çolak, H. (2025c). Prediction the number of coaching certificates in Türkiye with machine learning: a comparative study of artificial neural networks and adaptive neuro-fuzzy inference system. *Journal of Physical Education Sports Health and Effort*, *1*, 41-63.

Çolak, H., & Çolak, E. (2024). Türkiye'de Cinsiyete göre Obezite Öncesi Yüzdelik Dağılımının Yapay Sinir Ağı ve Zaman Serileri ile Tahmini. *Karadeniz Fen Bilimleri Dergisi*, *14*(3), 1340-1359.

Çolak, H., Kale, R., & Cihan, H. (2003). Yoğunlaştırılmış Yürüyüş ve Jogging Programının Yüksek Dansiteli Liporotein (HDL) Ve Düşük Dansiteli Lipoproteinler (LDL) Üzerine Olan Etkisi. *Spormetre Beden Eğitimi ve Spor Bilimleri Dergisi*, *1*(1), 69-76.

Çolak, H., & Şenol, H. (2023). Türkiye'nin lisansli sporcu sayısının yapay sinir ağları ile 2030 yılına kadar tahmini.

Çolak, H., & Yiğit, Z. (2017). Investigating the changes on body composition in women regularly exercise. *Journal of Current Researches on Health Sector*, *7*(2), 69-78.

Das, S. (2025). Association and inequality between socioeconomic status and the prevalence of overweight and obesity among adults aged 18 and older in Bangladesh. *Obesity Medicine*, *54*, 100595.

Debele, G. R., Kuse, S. A., Kefeni, B. T., Geda, A., Jifar, W. W., Kitila, K. M., & Hajure, M. (2023). Why too soon? Predictors of time to diabetic peripheral neuropathy among newly diagnosed diabetes mellitus patients: a multicenter follow-up study at health-care setting of Ethiopia. *Archives of Public Health*, *81*(1), 186.

DeGregory, K. W., Kuiper, P., DeSilvio, T., Pleuss, J. D., Miller, R., Roginski, J. W., Fisher, C. B., Harness, D., Viswanath, S., & Heymsfield, S. B. (2018). A review of machine learning in obesity. *Obesity reviews*, *19*(5), 668-685.

Du, J., Yang, S., Zeng, Y., Ye, C., Chang, X., & Wu, S. (2024). Visualization obesity risk prediction system based on machine learning. *Scientific Reports*, *14*(1), 22424.

Fardet, A., Aubrun, K., & Rock, E. (2021). Nutrition transition and chronic diseases in China (1990–2019): Industrially processed and animal calories rather than nutrients and total calories as potential determinants of the health impact. *Public Health Nutrition*, *24*(16), 5561-5575.

Ferreras, A., Sumalla-Cano, S., Martínez-Licort, R., Elío, I., Tutusaus, K., Prola, T., Vidal-Mazón, J. L., Sahelices, B., & de la Torre Díez, I. (2023). Systematic review of machine learning applied to the prediction of obesity and overweight. *Journal of medical systems*, *47*(1), 8.

Fu, L., Han, X., Wang, Y., & Hu, Y.-Q. (2025). Identifying genetic overlaps in obesity and metabolic disorders unlocking unique and shared mechanistic insights. *Free Radical Biology and Medicine*, *231*, 80-93.

Guo, X., & Wang, J. (2024). The Gompertz model for biohydrogen production kinetics: origin, application and solving methods. *International Journal of Hydrogen Energy*, *88*, 242-250.

Han, Y., Ma, Y., Liu, Y., Zhao, Z., Zhen, S., Yang, X., Xu, Z., & Wen, D. (2019). Plasma cholinesterase is associated with Chinese adolescent overweight or obesity and metabolic syndrome prediction. *Diabetes, metabolic syndrome and obesity: targets and therapy*, 685-702.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

Huang, J., Hong, Y., & Sun, D. (2025). Urban flood depth prediction using an improved LSTM model incorporating precipitation forecasting. *Natural Hazards*, 1-22.

Jaacks, L. M., Vandevijvere, S., Pan, A., McGowan, C. J., Wallace, C., Imamura, F., Mozaffarian, D., Swinburn, B., & Ezzati, M. (2019). The obesity transition: stages of the global epidemic. *The lancet Diabetes & endocrinology*, *7*(3), 231-240.

Jeon, J., Lee, S., & Oh, C. (2023). Age-specific risk factors for the prediction of obesity using a machine learning approach. *Frontiers in Public Health*, *10*, 998782.

Jin, G., Ni, X., Wei, K., Zhao, J., Zhang, H., & Jia, L. (2025). Will the technological singularity come soon? Modeling the dynamics of artificial intelligence development via multi-logistic growth process. *Physica A: Statistical Mechanics and its Applications*, *664*, 130450.

Lin, W., Shi, S., Huang, H., Wen, J., & Chen, G. (2023). Predicting risk of obesity in overweight adults using interpretable machine learning algorithms. *Frontiers in Endocrinology*, *14*, 1292167.

Liu, T., Yang, Z., Li, F., Li, Y., Liu, X., Wu, Z., & Cao, C. (2022). Tendency analysis of incidence and mortality of mechanical injuries among Chinese residents from 1990 to 2019. *Zhonghua wei zhong bing ji jiu yi xue*, *34*(10), 1082-1087.

Miri-Moghaddam, E., Bizhaem, S. K., Moezzifar, Z., & Salmani, F. (2024). Long-term prediction of Iranian blood product supply using LSTM: a 5-year forecast. *BMC Medical Informatics and Decision Making*, *24*(1), 213.

Monda, K. L., Gordon-Larsen, P., Stevens, J., & Popkin, B. M. (2007). China's transition: the effect of rapid urbanization on adult occupational physical activity. *Social science & medicine*, *64*(4), 858-870.

Montañez, C. A. C., Fergus, P., Hussain, A., Al-Jumeily, D., Abdulaimma, B., Hind, J., & Radi, N. (2017). Machine learning approaches for the prediction of obesity using publicly available genetic profiles. 2017 International Joint Conference on Neural Networks (IJCNN),

Ong, K. L., Stafford, L. K., McLaughlin, S. A., Boyko, E. J., Vollset, S. E., Smith, A. E., Dalton, B. E., Duprey, J., Cruz, J. A., & Hagins, H. (2023). Global, regional, and national burden of diabetes from 1990 to

2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. *The Lancet*, *402*(10397), 203-234.

Özen, H., & Özen, D. (2025). Trends and Forecasts of Overweight Prevalence in Türkiye: A Time Series Approach Using ARIMA Models. *Acta Medica Nicomedia*, *8*(1), 52-58.

Palloni, A., Beltrán-Sánchez, H., Novak, B., Pinto, G., & Wong, R. (2015). Adult obesity, disease and longevity in Mexico. *salud pública de méxico*, *57*, s22-s30.

Pang, X., Forrest, C. B., Le-Scherban, F., & Masino, A. J. (2019). Understanding early childhood obesity via interpretation of machine learning model predictions. 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA),

Pelinovsky, E., Kurkin, A., Kurkina, O., Kokoulina, M., & Epifanova, A. (2020). Logistic equation and COVID-19. *Chaos, Solitons & Fractals*, *140*, 110241.

Pham, T., Tran, T., Phung, D., & Venkatesh, S. (2016). Deepcare: A deep dynamic memory model for predictive medicine. Pacific-Asia conference on knowledge discovery and data mining,

Popkin, B. M. (2014a). Synthesis and implications: C hina's nutrition transition in the context of changes across other low-and middle-income countries. *Obesity reviews*, *15*, 60-67.

Popkin, B. M. (2014b). Synthesis and implications: China's nutrition transition in the context of changes across other low-and middle-income countries. *Obesity reviews*, *15*, 60-67.

Reyes-Sánchez, F., Basto-Abreu, A., Torres-Álvarez, R., Carnalla-Cortés, M., Reyes-García, A., Swinburn, B., Meza, R., Rivera, J. A., Popkin, B., & Barientos-Gutiérrez, T. (2023). Caloric reductions needed to achieve obesity goals in Mexico for 2030 and 2040: A modeling study. *PLoS Medicine*, *20*(6), e1004248.

Schreurs, L., Bowman-Busato, J., Gies, I., Calabrò, M., Morovic, M. L., Silva-Nunes, J., Woodward, E., & De Cock, D. (2025). Needs and challenges of the payor and regulatory community towards integrating obesity into a chronic care framework. *Public Health*, *241*, 151-157.

Shumway, R. H., & Stoffer, D. S. (2017). ARIMA models. In *Time series analysis and its applications: with R examples* (pp. 75-163). Springer.

Siddique, M. A. B., Mahalder, B., Haque, M. M., & Ahammad, A. S. (2025). Forecasting air temperature and rainfall in Mymensingh, Bangladesh with ARIMA: Implications for aquaculture management. *Egyptian Journal of Aquatic Research*.

Singh, S., Parmar, K. S., & Kumar, J. (2025). Development of multi-forecasting model using Monte Carlo simulation coupled with wavelet denoising-ARIMA model. *Mathematics and Computers in Simulation*, *230*, 517-540.

Singh, S., Parmar, K. S., Makkhan, S. J. S., Kaur, J., Peshoria, S., & Kumar, J. (2020). Study of ARIMA and least square support vector machine (LS-SVM) models for the prediction of SARS-CoV-2 confirmed cases in the most affected countries. *Chaos, Solitons & Fractals*, *139*, 110086.

Singh, U., Saurabh, K., Trehan, N., Vyas, R., & Vyas, O. (2024). GA-LSTM: Performance Optimization of LSTM driven Time Series Forecasting. *Computational Economics*, 1-36.

Soerjomataram, I., & Bray, F. (2021). Planning for tomorrow: global cancer incidence and the role of prevention 2020–2070. *Nature reviews Clinical oncology*, *18*(10), 663-672.

Sun, H., Lv, W., & Lee, J. (2023). Effect of Tai Chi exercise on healthcare and quality of life of middle-aged and elderly patients with hypertension. *International Journal of Healthcare Information Systems and Informatics (IJHISI)*, *18*(1), 1-12.

Şenol, H. (2020a). Anaerobic digestion of hazelnut (Corylus colurna) husks after alkaline pretreatment and determination of new important points in Logistic model curves. *Bioresource technology*, *300*, 122660.

Şenol, H. (2020b). Identification of new critical points for logistics model in cumulative methane yield curves after co-digestion of apple pulp and chicken manure with sulphuric acid pretreatment and a new modelling study. *International Journal of Energy Research*, *44*(7), 6078-6087.

Şenol, H., Çakır, İ. T., Bianco, F., & Görgün, E. (2024). Improved methane production from ultrasonically-pretreated secondary sedimentation tank sludge and new model proposal: Time series (ARIMA). *Bioresource technology*, *391*, 129866.

Şenol, H., Çolak, E., & Başer, V. (2025). Forecasting biogas potential in Türkiye's Central Anatolia region with artificial neural networks and geographical information system-based analysis. *Engineering Applications of Artificial Intelligence*, *160*, 111843.

Şenol, H., Çolak, E., Elibol, E. A., Hassaan, M. A., & El Nemr, A. (2024). Optimisation of biochar dose in anaerobic co-digestion of green algae and cattle manure using artificial neural networks and response surface methodology. *Chemical Engineering Journal*, *493*, 152750.

Şenol, H., Çolak, E., & Oda, V. (2024). Forecasting of biogas potential using artificial neural networks and time series models for Türkiye to 2035. *Energy*, *302*, 131949.

Şenol, H., & Çolak, H. (2025). Forecasting childhood and adolescent obesity prevalence in Türkiye through 2030: A novel application of artificial neural networks for public health planning. *Engineering Applications of Artificial Intelligence*, *159*, 111720.

Şenol, H., Çolak, H., & Çolak, E. (2024). Analysis of the Total Number of Licensed Athletes Using Artificial Neural Networks for the Future Years in Türkiye. *Karadeniz Fen Bilimleri Dergisi*, *14*(4), 2153-2171.

Şenol, H., Dereli, M. A., & Özbilgin, F. (2021). Investigation of the distribution of bovine manure-based biomethane potential using an artificial neural network in Turkey to 2030. *Renewable and Sustainable Energy Reviews*, *149*, 111338.

Şenol, H., Elibol, E. A., Bianco, F., & Race, M. (2025). Enhancing methane production from pistachio skin via optimized hydrothermal-alkaline pretreatment and Autoregressive Integrated Moving Average modeling. *Journal of Environmental Management*, *380*, 125121.

Tamilselvi, C., Paul, R. K., Yeasin, M., & Paul, A. (2024). Novel wavelet-LSTM approach for time series prediction. *Neural computing and applications*, 1-10.

Teng, A. M., Jones, A. C., Mizdrak, A., Signal, L., Genç, M., & Wilson, N. (2019). Impact of sugar-sweetened beverage taxes on purchases and dietary intake: Systematic review and meta-analysis. *Obesity reviews*, *20*(9), 1187-1204.

Thamrin, S. A., Arsyad, D. S., Kuswanto, H., Lawi, A., & Nasir, S. (2021). Predicting obesity in adults using machine learning techniques: an analysis of Indonesian basic health research 2018. *Frontiers in nutrition*, *8*, 669155.

Tjørve, K. M., & Tjørve, E. (2017). The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family. *PloS one*, *12*(6), e0178691.

Tomov, L., Chervenkov, L., Miteva, D. G., Batselova, H., & Velikova, T. (2023). Applications of time series analysis in epidemiology: Literature review and our experience during COVID-19 pandemic. *World Journal of Clinical Cases*, *11*(29), 6974.

Wang, L., Chen, L., Jin, S., & Li, C. (2025). Forecasting the Green Behaviour Level of Chinese Enterprises: A Conjoined Application of the Autoregressive Integrated Moving Average (ARIMA) Model and Multi-Scenario Simulation. *Technology in Society*, 102825.

Wang, Q., Yang, M., Pang, B., Xue, M., Zhang, Y., Deng, X., Zhang, Z., & Niu, W. (2022). Machine/Deep Learning-based Approaches to Predict Overweight or Obesity in Chinese Preschool-Aged Children.

Wang, Q., Yang, M., Pang, B., Xue, M., Zhang, Y., Zhang, Z., & Niu, W. (2022). Predicting risk of overweight or obesity in Chinese preschool-aged children using artificial intelligence techniques. *Endocrine*, *77*(1), 63-72.

Wang, Y., Zhao, L., Gao, L., Pan, A., & Xue, H. (2021). Health policy and public health implications of obesity in China. *The lancet Diabetes & endocrinology*, *9*(7), 446-461.

Wang, Z., Evans, M. F., Wang, H., & Wang, S. (2025). Childhood obesity and air pollution: Evidence from China. *Economics Letters*, 112181.

Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PloS one*, *12*(4), e0174944.

WHO. (2025). World Health Organization https://www.who.int/.

Xie, T., Dai, Z., Zhu, X., Chen, B., & Ran, C. (2023). LSTM-based short-term ionospheric TEC forecast model and positioning accuracy analysis. *GPS Solutions*, *27*(2), 66.

Zhang, L., Wang, Z., Wang, X., Chen, Z., Shao, L., Tian, Y., Dong, Y., Zheng, C., Li, S., & Zhu, M. (2019). Prevalence of abdominal obesity in China: results from a cross-sectional study of nearly half a million participants. *Obesity*, *27*(11), 1898-1905.

Zhang, Y., Wang, Q., Xue, M., Pang, B., Yang, M., Zhang, Z., & Niu, W. (2022). Identifying factors associated with central obesity in school students using artificial intelligence techniques. *Frontiers in Pediatrics*, *10*, 1060270.

Zhang, Z., Peng, W., Sun, S., Ma, J., Sun, Y., & Zhang, F. (2024). Predicting the onset of overweight in Chinese high school students: a machine-learning approach in a one-year prospective cohort study. *Endocrine*, *86*(2), 600-611.

Zhao, S., Musa, S. S., Fu, H., He, D., & Qin, J. (2020). Large-scale Lassa fever outbreaks in Nigeria: quantifying the association between disease reproduction number and local rainfall. *Epidemiology & Infection*, *148*, e4.

Zheng, Z., & Ruggiero, K. (2017). Using machine learning to predict obesity in high school students. 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM),