

Article



Forecasting of the Prevalence of Dementia Using the LSTM Neural Network in Taiwan

Stephanie Yang ¹, Hsueh-Chih Chen ^{1,2,3,4,*}, Chih-Hsien Wu ⁵, Meng-Ni Wu ⁶ and Cheng-Hong Yang ^{5,7,8,*}

- ¹ Department of Educational Psychology and Counseling, National Taiwan Normal University, Taipei 106, Taiwan; sfyang@ntnu.edu.tw
- ² Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, Taipei 106, Taiwan
- ³ Chinese Language and Technology Center, National Taiwan Normal University, Taipei 106, Taiwan
- ⁴ MOST AI Biomedical Research Center, Tainan City 701, Taiwan
- ⁵ Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 807, Taiwan; f108152123@nkust.edu.tw
- ⁶ Kaohsiung Medical University Chung-Ho Memorial Hospital, Kaohsiung 80756, Taiwan; mengniwu@kmu.edu.tw
- 7 Ph.D. Program in Biomedical Engineering, Kaohsiung Medical University, Kaohsiung 80708, Taiwan
- ⁸ Drug development and value creation research center, Kaohsiung Medical University, Kaohsiung 80708, Taiwan
- * Correspondence: chcjyh@ntnu.edu.tw (H.-C.C.); chyang@nkust.edu.tw (C.-H.Y.)

Abstract: The World Health Organization has urged countries to prioritize dementia in their public health policies. Dementia poses a tremendous socioeconomic burden, and the accurate prediction of the annual increase in prevalence is essential for establishing strategies to cope with its effects. The present study established a model based on the architecture of the long short-term memory (LSTM) neural network for predicting the number of dementia cases in Taiwan, which considers the effects of age and sex on the prevalence of dementia. The LSTM network is a variant of recurrent neural networks (RNNs), which possesses a special gate structure and avoids the problems in RNNs of gradient explosion, gradient vanishing, and long-term memory failure. A number of patients diagnosed as having dementia from 1997 to 2017 was collected in annual units from a data set extracted from the Health Insurance Database of the Ministry of Health and Welfare in Taiwan. To further verify the validity of the proposed model, the LSTM network was compared with three types of models: statistical models (exponential smoothing (ETS), autoregressive integrated moving average model (ARIMA), trigonometric seasonality, Box-Cox transformation, autoregressive moving average errors, and trend seasonal components model (TBATS)), hybrid models (support vector regression (SVR), particle swarm optimization-based support vector regression (PSOSVR)), and deep learning model (artificial neural networks (ANN)). The mean absolute percentage error (MAPE), root-mean-square error (RMSE), mean absolute error (MAE), and R-squared (R²) were used to evaluate the model performances. The results indicated that the LSTM network has higher prediction accuracy than the three types of models for forecasting the prevalence of dementia in Taiwan.

Keywords: long short-term memory; dementia; prevalence; deep learning

1. Introduction

The elderly population has been increasing sharply worldwide. In March 2018, Taiwan officially became an aged society, with an elderly population accounting for 14% of the total population [1]. Despite over 25 years elapsing before Taiwan progressed from an aging society to an aged society, Taiwan is estimated to become a super-aged society by

Citation: Yang, S.; Chen, H.-C.; Wu, C-H.; Wu, M.-N.; Yang, C.-H. Forecasting of the Prevalence of Dementia using the LSTM Neural Network in Taiwan. *Mathematics* **2021**, *9*, 488. https://doi.org/10.3390/ math9050488

Academic Editor: Alessandro Niccolai

Received: 24 January 2021 Accepted: 23 February 2021 Published: 27 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses /by/4.0/). 2026 because of its fast aging speed [2]. The rapid growth of the elderly population indicates a shift in health care concerns (chronic diseases and conditions) and medical care for older adults as well as emphasis on the importance of long-term care, prevention, and resilience. Especially when facing sudden disastrous events, such as the outbreak of COVID-19 in 2020, a stabilized, resilient, and supportive healthcare system and policies are necessary to a fast-ageing society. Taiwan should involve strategies for alleviating the burden of this growing population on the current health care system. Aside from curing diseases, the goal of health care interventions for the elderly population should be to prevent disability, reduce the occurrence of dementia, and expand the capacity of the current health care system. Based on an accurate proposed time series model, the aim of this study is to provide stakeholders a reference of the changing trend to accommodate this growing need.

The literature indicates that dementia is an acquired, chronic, and progressive cognitive dysfunction in multiple domains, including memory, language, visuospatial, and executive function [3]. Symptoms such as severe memory impairment, disorientation and confusion, mood instability, and behavioral and psychological changes (e.g., hallucinations and delusions) develop as the disease progresses [4]. These symptoms may contribute to the daily social or professional dysfunction of patients, causing emotional exhaustion among caregivers and increasing the socioeconomic burden [5].

Past studies indicated that dementia has a substantial effect on the social care system and societal costs [6]. As an acquired disabling syndrome, characterized by a progressive deterioration in multiple cognitive domains that interferes with daily functioning, several conditions cause dementia symptoms, including Alzheimer's disease, vascular disorders, and Parkinson's disease. The prevalence of dementia among older adults increases as a population ages, significantly affecting the lives of an increasingly large number of older adults globally. Because dementia is the main cause of hospitalization among older adults, the increasing prevalence of dementia places tremendous pressure on the health insurance system. The number of people diagnosed as having dementia worldwide is growing with the increase in the global average living age [7]. The World Alzheimer's Disease Report indicated that the number of people with dementia worldwide exceeded 50 million in 2019 and is expected to increase to 15.2 billion by 2050; moreover, dementia currently incurs a cost of US\$1 trillion per year, and this figure is expected to double by 2030 [7]. Based on the rapid increase in dementia patients and financial demands, dementia is a global problem that requires urgent attention. A time series approach could provide a thorough and accurate understanding of the prevalence rate, which will assist in actions taken to address this disease.

A significant portion of the societal cost of dementia is patient care. Studies reported that the average total costs for the last 5 years of life of patients with dementia are higher than those of patients with heart disease or cancer, among other causes of mortality [8]. In 2015, the overall cost of dementia was approximately US \$818 billion; 40.4% of this cost was attributed to caregivers. Dementia is often associated with disorientation, confusion, mood instability, and behavioral psychological symptoms; care is thus demanding. Care for patients with dementia is generally more time consuming than care for patients with other diseases [9]. The informal caregivers of patients with dementia often develop depression, anxiety, and physical symptoms and even have a relatively high mortality rate [5,10–12]. Therefore, the care of patients with dementia is one of the major sources of socioeconomic burden that should be emphasized in policies on expanding the medical allowance for this population. Suitable social welfare and public health policies necessitate a precise model for predicting the prevalence rate of dementia.

Time series analysis, also known as dynamic series analysis, is a classic statistical method that refers to a sequence formed by arranging various variables according to a time series. Time and data variables play a critical role in time series analysis. Based on irregular changes, various factors in the actual situation can affect the time series forecast. Numerous methods have been proposed in time series-related research. Among the proposed methods, the two most commonly used time series forecasting models are the autoregressive integral moving average (ARIMA) proposed by Box and Jenkins and the exponential smoothing method (ETS) proposed by Brown [13,14]. Although these two methods have been applied in various fields [15,16], there are limitations. In the ARIMA, the future value of the time series and past observations are assumed to satisfy a linear relationship [17]; however, most time series data have a nonlinear relationship, which limits the application scope of the ARIMA. On the other hand, ETS forecasts are based on historical data. If not combined with other methods, the prediction of a nonlinear time series cannot produce satisfactory results. These methods rely heavily on linear assumptions, which use historical data sets, single-variable, or multivariable time series functions to forecast future trends. The lack of nonlinear fitting capabilities may limit the development of these methods.

Machine learning (ML) methods – such as artificial neural networks (ANNs) [18] and support vector regression (SVR) [19]-have demonstrated excellent nonlinear fitting capabilities in demand forecasting. However, improper parameter settings seriously affect the realization of ANN and SVR methods. Studies have strongly indicated that a solution is required to produce suitable hyperparameters [20,21]. Therefore, many hybrid models were proposed to solve the optimization problem, such as PSOSVR, GASVR, DESVR, GSSVR [20,22]. Recurrent neural networks (RNNs) are neural networks developed for time series problems in numerous types of neural network structures. Research has indicated that RNNs outperform traditional neural networks, such as multilayer perceptron machines [23,24]. Cui and Liu used a combination of an RNN and convolutional neural network (CNN) to classify Alzheimer's disease [25]. Maragatham and Devi (2019) established a mental strength failure prediction model based on long short-term memory (LSTM) neural networks [26]. In addition, Lipton et al. (2015) used an LSTM network for the classification and diagnosis of patients in hospital pediatric intensive care units [27]. Wang et al. (2019) also developed a deep learning approach involving the use of longitudinal electronic health records to predict mortality risk for the identification of patients with dementia who may benefit from palliative care [28]. Another study evaluated the role of deep learning models in identifying surgical behaviors and evaluating surgeons' technical performance [29]. These aforementioned studies have demonstrated the successful application of various RNN models to numerous medical prediction tasks through the effective use of the temporal relationship among collected patient data.

This study proposes the use of an RNN structure based on an LSTM network to predict the trends of patients with dementia. Based on our understanding, relatively limited studies have explicitly evaluated the forecast of dementia worldwide to date. Most of the research focuses on predicting patients with dementia and the classification of dementia [30–32], which is substantial progress for physicians. However, for the government or policymakers, the strategic layout and budget of medical care are the issues they are concerned about. With the increasing number of patients with dementia [33], the cost of care is also increasing, which is one of the focuses of this study. Recently, Kingston et al. forecasted the older population's care needs in England over the next 20 years via PACSim model [34]. Ahmadi-Abhari et al. developed a Monte-Carlo Markov model on predicting the number of people living with dementia to 2050 and provided the estimates for the impact of smoking cessation [35]. In addition to no relevant research on the relationship between the number of patients with dementia and nursing cost and policy promotion in Taiwan, the LSTM network with excellent performance in sequence prediction [36], has not been used to predict the number of people with dementia. Therefore, this study aims to establish a prediction model to provide a reference for government budgeting and administration by accurately predicting the number of people with dementia. Comparisons with a series of benchmark models verified the superiority of the proposed model. On the basis of these findings, this paper presents further constructive recommendations to actively support dementia prevention and care, which can considerably improve the health care process for caregivers and society. Section 2 discusses the LSTM architecture and other prediction models. Sections 3 and 4 respectively explain how the LSTM network can be used for regression problems and present the experimental results of a forecasting application.

2. Materials and Methods

2.1. Data Sources and Preprocessing

The sources of the data were extracted from the Health Insurance Database of the Ministry of Health and Welfare in Taiwan, which included the annual number of dementia patients over 60 years old. According to the availability of information from 1997 to 2017, we applied the proposed LSTM method to the abstracted data to determine the trend. The model methodology will be elaborated below. The annual data from 1997 to 2013 was applied as a training set to train the proposed LSTM method. Subsequently, the testing set was used to test the accuracy of the forecast, which consisted of the annual number of dementia patients for 2014–2017.

2.2. LSTM Network

Time series forecasting is emerging as one of the most important branches of data analysis. However, traditional time series forecasting models often result in poor forecasting accuracy, because such methods require large sequence data features [37]. Data collected at fixed time intervals are called time series data; each data point is equally spaced in time. Time series prediction is a method of predicting future trends and patterns of a historical data set using time characteristics. Predicting the number of patients with dementia using input data with a time component and a model that differs from the traditional regression method may be effective.

Figure 1 presents a traditional RNN. The input of the RNN has a sequence length $x = (x_1, ..., x_T)$, which can be processed recursively. When processing each symbol, the RNN maintains an internal hidden state (i.e., *s*). The parameters of this method are the recursive weight matrix *W*, the input weight matrix *U*, and the output weight matrix *V*. The operation of the RNN at each time step *t* can be expressed as

$$h_t = \sigma(U_{x_t} + Ws_{t-1}) \tag{1}$$

where σ is the starting function, and t = 1, 2, ..., T. The output of the RNN is calculated using the formula

$$o_t = \sigma(Vs_t) \tag{2}$$

A one-step-ahead forecast in a time series requires both the previous data and the most recent data. An RNN model has the advantages of a hidden-layer self-feedback mechanism and the ability to avoid long-term dependence problems. However, practical applications still face some difficulties [38].



Figure 1. Schematic diagram of an RNN.

The first LSTM neural network was proposed by Hochreiter and Schmidhuber (1997) and had a targeted design to solve the problem of long-term dependence [39]. The LSTM network memory unit consists of four gates (or units), namely the input gate, output gate, forget gate, and memory unit. The gate that controls the flow of information is displayed in Figure 2. The LSTM network is a variant of RNNs, which have been used for numerous practical situations in fields such as biomedical sciences [40], speech recognition [41], sentiment analysis [42], and image classification [43]. The input gate controls whether the input signal can modify the state of the memory cell. The output gate controls whether the state of other memory units can be modified. The forget gate can choose to forget or remember its previous state. Broadly:

- (1) The input gate allows new information to flow into the network. It has parameters W_i, b_i , where *i* stands for input.
- (2) The memory cell preserves the hidden unit information across time steps. It has parameters W_c , $b_{c'}$ where *c* stands for cell.
- (3) The forget gate allows information, which is no longer pertinent, to be discarded. It has parameters W_f , b_f , where f stands for forget.
- (4) The output gate determines what information should be output to the next neuron and what should be propagated forward as part of the new hidden state. It has parameters W_o and b_o , where o stands for output.

The LSTM network is an effective algorithm for establishing time series models. The basic component of the LSTM network is the memory block, which solves the gradually gradient vanishing problem by storing network parameters for a long period of time. The four gates of the LSTM network are represented by the following formula:

At time t, x_t is the input data of the LSTM unit, h_t is the output of the LSTM unit, h_{t-1} is the output of the LSTM unit at the previous moment, and C_t is the value of the memory unit. The process of the LSTM network can be divided into the following steps.

(1) Calculate the value of the candidate memory unit \tilde{C}_t , where W_c is the weight matrix and b_c is the bias.

$$\tilde{C}_t = \sigma(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{3}$$

(2) Calculate the value of the input gate *I_t*. The input gate controls the update of the current input data to the state value of the memory unit, where *σ* is the sigmoid function, *W_i* is the weight matrix, and *b_i* is the bias.

$$I_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4}$$

(3) Calculate the value of the forget gate F_t . The forget gate controls the update of the historical data to the state value of the memory unit, where W_f is the weight matrix and b_f is the bias.

$$F_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{5}$$

(4) Calculate the value of the current moment memory unit C_t ; C_{t-1} is the state value of the last LSTM unit.

$$C_t = F_t \times C_{t-1} + I_t \times \tilde{C}_t \tag{6}$$

(5) Calculate the value of the output gate O_t . The output gate controls the output of the state value of the memory unit, where W_o is the weight matrix and b_o is the bias.

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{7}$$

(6) Calculate the output of the LSTM unit h_t , where *tanh* is a non-linear activation. It

squashes the permissible amplitude range of the output signal to some finite value. The function is shown as

$$h_t = O_t \times tanh(C_t) \tag{8}$$

$$tanh(x) = \frac{\sin h(x)}{\cos h(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$
(9)



Figure 2. Unit of long short-term memory neural network.

The LSTM network employs three control gates and memory units to save, read, reset, and update long-term information. Due to the common mechanism of the LSTM network's internal parameters, the dimensions of the weight matrix can be set to control the dimension of the output. The LSTM network establishes a long time period between the input and feedback.

An entire univariate or multivariate time series can be used to train LSTM networks. To improve the learning process and effectiveness of the model, smaller subsamples from the original time series were defined in this study. With *T* as the sequence length, the sequences $z^{(t)} = [x_t, x_{t+1}, ..., x_{t+T-1}] \in \mathbb{R}^T$ and $y^{(t)} = x_{t+T-1+q} \in \mathbb{R}$ are the *t*th input and output of the LSTM network, respectively; *q* is a positive integer that indicates the number of steps ahead to be predicted, and *N* is the total number of subsamples which depends on the length of the original time series or the sequence length.

2.3. Statistical Models

Statistical predictions are usually divided into two categories: qualitative and quantitative methods. Time series analysis is a quantitative prediction method that is widely used in mathematical statistics, signal processing, financial prediction, electroencephalography, and other fields and is beneficial to economic and scientific improvement. To objectively present the most robust method with a low error rate, the three following methods were selected for comparison.

2.3.1. ETS (Exponential Smoothing)

Proposed by Brown and Meyer (1961) [44], ETS is a data averaging method that considers three factors: the error, trend, and season. A maximum likelihood estimation is used in ETS to optimize the initial values and parameters, and the optimal exponential smoothing model is then selected. Moreover, the weight of the ETS weighted data decays exponentially. The latest data have a higher weight than older data, and the weight of the older data decreases gradually. The ETS algorithm is a solution that overcomes the limitations of previous exponential smoothing models, but does not provide a convenient forecasting interval calculation method.

2.3.2. ARIMA (Autoregressive Integrated Moving Average)

Proposed by Box and Jenkins (1976), the ARIMA model, also known as the Box–Jenkins model, uses several formed fragments after the time series has passed as the input, and a prediction model is established on the basis of the regression analysis results [13]. This model is frequently used for the prediction of short-term trends in economic areas.

2.3.3. TBATS (Trigonometric Seasonality, Box–Cox Transformation, ARMA Errors, and Trend Seasonal Components Model)

This method, which was proposed by Livera, Hyndman, and Snyder (2011), combines trigonometric seasonality, Box–Cox transformation, ARMA errors, and trend and seasonal components [45]. This approach can be used to analyze and predict whether seasonal data exist based on the exponential smoothing method. The combination of multiple models can achieve more accurate results but also requires more training time, resulting in slower calculation.

2.4. Hybrid Models

2.4.1. SVR (Support Vector Regression)

SVR was proposed by Drucker et al. in 1996 [19]. SVR includes an insensitive loss function and penalty factor to enhance the robustness of SVMs [46,47]. SVR involves the projection of data to a high-dimensional hyperplane and subsequent calculation of the total distance from each point to the hyperplane. The hyperplane with the smallest total distance is identified as the solution. SVR has three hyperparameters: the regularization parameter (*C*), kernel function bandwidth (σ), and ε -insensitive loss function (ε). Changes to these parameters considerably affect the accuracy of SVR prediction. However, the automatic adjustment of the three hyperparameters in SVR remains a challenge in improving the accuracy of SVR prediction.

2.4.2. PSOSVR (Particle Swarm Optimization-Based SVR)

Particle swarm optimization (PSO) was proposed by Kennedy and Eberhart (1995) based on the flight motion of foraging birds [48]. A bird's movement continuously reveals places with food, thereby updating the position of the entire group until the optimal location is finally identified. Defining the stopping criterion and constraining the processing mechanism is crucial in a PSO search space. If the number of iterations of the PSO exceeds a predetermined threshold, the PSO will stop. PSO is an approach to global optimization calculation that can effectively select the optimal combination of internal parameters for an SVR model and improve the model prediction accuracy and generalization ability (Liu et al., 2018). The fitness function is calculated in each optimization process of the PSO to determine the solution for the parameters (C, σ , and ε).

2.5. Deep Learning Model (ARTIFICIAL Neural Network ANN)

ANNs are inspired by the structure of the human nervous system. A neural network is a collection of interrelated 'neurons' in a self-adjusting system. An ANN arithmetically adjusts the weights (free parameters) to meet performance requirements using representative samples. Because of their learning process involves the use of historical data, ANNs demonstrate high effectiveness in complex problems [49]. Back-propagation networks are well-known supervised learning neural network models [50] that consist of an optimization algorithm which combines a backward pass, gradient descent [51], and the chain rule in calculus. The gradient descent method identifies the initial position of the parameter in the steepest downhill direction and updates the parameter position. Slope information is obtained through derivation of the function. The gradient descent method uses cost function to optimize the weights in ANN.

3. Results

3.1. Parameter Settings

Table 1 presents a summary of the data collected from the Department of Statistics of Taiwan's Ministry of Health and Welfare. The number of people with dementia over the age of 60 years registered from 1997 to 2017 was calculated. The data used in this study were grouped by gender and age, and the maximum, minimum, mean, median, first quartile, third quartile, interquartile range (IQR), standard deviation (SD), and coefficient of variation (CV) were calculated for each group. The CV values were used to determine the degree of dispersion of a set of data around the average value, with a larger CV value indicating a higher degree of dispersion. The data indicated that older age was associated with a higher degree of dispersion. The overall number of dementia cases was higher among women than among men in each age group. SVR has three hyperparameters that affect the accuracy of the forecasting task; namely, the tube size of the ε -insensitive loss function (ε), regularization parameter (C), and bandwidth of the kernel function (σ). The parameter settings used in this study are displayed in Table 2 and 3.

Table 1. Descriptive statistics for dementia patients in Taiwan from 1997 to 2017

Years Old	Sex	Min	Max	Mean	Med	Q1	Q3	IQR	SD	COV
60.64	Μ	1003	2092	1392.67	2092	1088	1675	587	322.38	0.23
00~04	F	925	2319	1390.62	2319	1053.5	1690	636.5	385.44	0.28
65~69	Μ	3459	9747	5020.86	4258	3832.5	5324.5	1492	1714.18	0.34
	F	3050	11572	5825.67	3621	3646	6937	3291	2379.33	0.41
70~74	М	5794	11094	8059	5794	7033	9625.5	2592.5	1533.13	0.19
	F	4613	15353	9734.86	4613	5995.5	13676.5	7681	3642.94	0.37
	Μ	5198	17196	11469.9	5198	8578.5	13370.5	4792	3318.71	0.29
/5~/9	F	4644	27538	14564.6	4644	7846	20566.5	12720.5	6973.84	0.48
20.24	Μ	3635	18941	12231.9	3635	6564	17755	11191	5428.23	0.44
00~04	F	3837	33385	15857.2	3837	7379	23432	16053	8964.56	0.57
85 and	М	1959	33696	13619.6	1959	4626	21380	16754	10091	0.74
above	F	2524	49558	18865.5	2524	6831.5	28351	21519.5	13967.5	0.74
Tatal	Μ	25108	94040	52790.7	26079	31968	69794.5	37826.5	21466	0.41
Total	F	25322	140253	67047.5	26008	33068	95259.5	62191.5	35588.5	0.53

Min, minimum; Max, maximum; Med, median; Q1, the first quartile; Q3, the third quartile, IQR, interquartile range; SD, standard deviation; COV, coefficient of variation; M, male; F, female.

Table 2. Parameter settings in PSOSVR

Years Old	ε	С	σ
60~64	0.0625	8	0.5
65~69	0.015625	4	0.25
70~74	0.03125	16	0.5
75~79	0.0078125	32	0.5
80~84	0.000976563	4	0.5
85 and above	0.00390625	4	0.125
Total	0.00390625	8192	0.001953125

ε, ε-insensitive loss function; C, penalty factor; σ, kernel function bandwidth.

Input	LSTM_1	LSTM_	2 Hidden_1	Hidden_	2 Output	Activation function	Learning rate	Epochs
300	250	200	100	50	1	Adam	1E-05	2000

Table 3. Number of neurons and parameter settings for proposed model

3.2. Model Performance

Table 4 presents a comparison of the predicted results of the six time series prediction models and the proposed LSTM model for male patients with dementia. The models were divided into statistical, hybrid, and deep learning models to understand the relationship between each prediction model. The statistical models comprise ETS, ARIMA, and TBATS. The hybrid models comprise SVR and PSOSVR. The deep learning model were represented by ANN. These six models were compared with the prediction results of the proposed LSTM model. Among the statistical models, the model TBATS exhibited favorable performance compared with the models ETS and ARIMA, with respective decreases in mean absolute percentage error (MAPE) of 20% and 21% in TBATS. Hybrid models, PSOSVR also demonstrated favorable performance compared with the SVR and ANN models, with significant decreases in MAPE value of 65% and 16% in PSOSVR.

Comparisons of the predicted results of the six time series prediction models and the proposed LSTM model for female patients with dementia are displayed in Table 5. Similarly, the prediction models were divided into three categories based on whether they used statistical, hybrid, or deep learning models. In the statistical models, the MAPE of TBATS was relatively high compared with the MAPE values of ETS and ARIMA, which were 46% and 24% lower, respectively. This discrepancy is principally related to the prediction model's poor performance in the 65~69 years old group, for which the MAPE value was as high as 30.48%. The inability to modify or iterate the parameters is the major disadvantage of the statistical models. To improve the results, PSOSVR was compared with SVR; the MAPE was reduced by 46% in PSOSVR, and the error percentage was successfully reduced. However, when PSOSVR was compared with the ANN, the MAPE value was 2% higher, which may have been related to PSOSVR not having the yet-identified optimal solution. However, the prediction results were close to those of neural networks, which demonstrates the value of the hybrid models.

Years Old	Criteria	ETS	ARIMA	TBATS	SVR	PSOSVR	ANN	LSTM
	MAE	233.60	248.71	60.78	966.21	138.20	336.99	14.17
60~64	MAPE (%)	16.26	17.02	3.65	52.10	8.89	5.57	0.90
	RMSE	327.64	322.56	76.59	890.08	142.76	87.44	14.66
	MAE	911.40	976.72	653.48	2638.68	995.69	910.94	89.09
65~69	MAPE (%)	10.15	10.99	7.04	26.40	11.96	10.15	1.18
	RMSE	1013.54	1064.55	853.48	2891.80	1210.74	1013.20	84.49
	MAE	331.20	363.88	262.77	637.62	1045.12	179.15	159.58
70~74	MAPE (%)	3.49	3.84	5.08	5.43	2.72	2.00	1.55
	RMSE	391.49	423.19	525.92	749.69	293.18	203.18	189.58
	MAE	809.40	865.17	955.47	2870.72	2618.90	959.79	671.39
75~79	MAPE (%)	4.80	5.19	8.10	15.77	4.46	4.67	2.75
	RMSE	859.95	908.05	1233.16	2911.56	802.91	702.17	483.70
	MAE	520.20	547.82	618.75	5000.10	480.66	2247.28	527.91
80~84	MAPE (%)	2.74	2.87	4.56	3.14	2.72	12.43	2.66
	RMSE	541.61	565.18	1027.30	726.64	612.25	2281.82	526.43
QE and	MAE	2667.20	2768.72	3593.13	4782.60	1727.15	3106.78	1644.49
abovo	MAPE (%)	8.58	8.86	8.36	14.66	7.13	10.57	4.61
above	RMSE	3283.61	3318.90	3557.40	5066.09	2427.37	3137.09	2524.63

Table 4. Predicted results for male dementia patients using different models

Total	MAE	4945.60	5172.84	5532.69	3529.23	3116.40	4116.68	3175.42
	MAPE (%)	6.97	5.78	5.47	5.22	4.19	4.71	3.88
	RMSE	6260.40	6430.58	6940.69	4137.92	4127.25	4388.69	4007.95
Average	MAE	1488.37*	1563.41*	1596.72*	2917.88*	1446.02*	1693.94*	897.44
	MAPE (%)	7.56*	7.79*	6.04*	17.53*	6.01*	7.16*	2.50
	RMSE	1811.18*	1861.86*	2030.65*	2481.97*	1373.78	1687.66*	1118.78

MAE, mean absolute error; MAPE, mean absolute percentage error; RMSE, root mean squared error; Boldface, the best values in each row. ETS, Exponential smoothing; ARIMA, Autoregressive integrated moving average; SVR, Support vector regression; PSOSVR, Particle swarm optimization integrated Support vector regression; TBATS, Trigonometric seasonality Box–Cox transformation ARMA errors Trend Seasonal components; ANN, Artificial neural network; LSTM, Long short-term memory. * p < 0.05, ** p < 0.01, *** p < 0.001.

Table 5. Predicted results for female dementia patients using different models

Years old	Criteria	ETS	ARIMA	TBATS	SVR	PSOSVR	ANN	LSTM
	MAE	128.24	97.49	39.82	577.79	152.94	278.21	6.60
60~64	MAPE (%)	6.35	7.49	6.14	34.04	12.25	15.66	2.33
	RMSE	131.97	175.97	273.88	611.82	229.06	324.11	46.27
	MAE	1215.29	882.27	2927.21	1147.67	1368.60	541.18	429.07
65~69	MAPE (%)	10.84	8.45	30.48	17.24	14.76	5.06	4.42
	RMSE	1203.70	903.06	2734.10	2386.27	1397.15	776.41	634.07
	MAE	614.80	623.69	730.62	2014.93	777.74	754.69	609.49
70~74	MAPE (%)	4.38	4.54	5.49	11.59	5.50	5.33	4.34
	RMSE	766.84	703.18	832.89	2160.61	1020.11	945.34	710.01
	MAE	1321.81	4630.01	1567.16	2139.77	955.08	994.06	893.49
75~79	MAPE (%)	5.03	19.47	6.48	7.45	3.82	4.03	3.61
	RMSE	1409.63	4566.97	1542.58	2181.50	1033.29	1106.36	1014.32
	MAE	2354.40	1763.74	2456.60	3955.26	887.86	1462.82	654.31
80~84	MAPE (%)	7.41	5.23	9.37	10.90	3.43	5.02	2.37
	RMSE	2454.16	1685.58	2547.80	3187.92	1085.27	1645.63	812.76
85 and	MAE	4734.00	3584.35	3496.61	1899.59	8248.37	9350.43	1173.70
ob and	MAPE (%)	10.06	8.83	8.87	4.75	4.12	8.48	2.88
above	RMSE	5341.83	3556.22	3641.71	3128.78	1997.94	3639.94	1678.02
	MAE	9577.40	8370.13	11891.27	3165.95	4105.93	4295.60	2392.24
Total	MAPE (%)	7.66	6.82	9.16	2.70	3.56	2.89	1.99
	RMSE	8952.33	8343.87	10896.10	3733.22	4564.03	3101.85	2764.75
	MAE	2849.42*	2850.24*	3301.33*	2128.71*	2356.65*	2048.20*	879.84
Average	MAPE (%)	7.39*	8.69*	10.86*	12.67*	6.78*	6.64*	3.13
Ũ	RMSE	2894.35*	2847.84*	3209.87*	2322.75*	1618.12*	1648.52*	1094.32

MAE, mean absolute error; MAPE, mean absolute percentage error; RMSE, root mean squared error; Boldface, the best values in each row. ETS, Exponential smoothing; ARIMA, Autoregressive integrated moving average; SVR, Support vector regression; PSOSVR, Particle swarm optimization integrated Support vector regression; TBATS, Trigonometric seasonality Box-Cox transformation ARMA errors Trend Seasonal components; ANN, Artificial neural network; LSTM, Long short-term memory. * p < .05, ** p < .01, *** p < .001.

3.3. Analysis of Individual Data

3.3.1. Patients Aged 60~64 Years

As shown by the blue curves in Figure 3, the slope of the training set (1997–2012) was similar to the slope of the test set (2013–2017), indicating that the trend of the overall patient population was stable. Therefore, the predicted results produced by each model were

relatively similar. The SVR prediction curve deviated from the actual values in Figures 4a and 5a because SVR involved the use of the grid search method for parameter adjustment, meaning that the searched hyperparameters may not have included the optimal solution; thus, the PSOSVR predicted value was closer to the actual value. However, optimization algorithms such as PSO do not guarantee that the output is the global optimum. Therefore, there remains room for improvement in the prediction ability. The predicted value of the LSTM network was closer to the observed value due to network's ability to memorize the overall curve trend and output value control through the gate. Therefore, the LSTM network demonstrated the most favorable prediction accuracy for this age group.



Figure 3. Projected increase in the number of dementia patients in the Taiwan, by age group (1997–2017). (a) male; (b) female.



Figure 4. Prediction results of the number of male dementia patients with different models at different age from 2013 to 2017. (a) 60~64 years old; (b) 65~69 years old; (c) 70~74 years old; (d) 75~79 years old; (e) 80~84 years old; and (f) 85years old and above.



Figure 5. Prediction results of the number of female dementia patients with different models at different age from 2013 to 2017. (a) 60~64 years old; (b) 65~69 years old; (c) 70~74 years old; (d) 75~79 years old; (e) 80~84 years old; and (f) 85years old and above.

3.3.2. Patients Aged 65~69 Years

As indicated by the orange curve in Figure 3, the number of patients steadily increased. Therefore, adapting to the overall trend is not demanding for the algorithms. Both Figures 4b and 5b indicate that the LSTM network was more sensitive to changes in the trend; thus, its performance was predicted to be the most favorable for this age group. Although the curves in Figure 5b suggest that the results obtained with ARIMA and TBATS were in higher agreement with the observed values, the errors of the output values were too large, making access to an accurate prediction impossible.

3.3.3. Patients Aged 70~74 Years

Among the patients aged 70~74 years, the prediction curve of the LSTM network for female patients was slightly behind the trend, and the estimation was affected by the training data. The gray curve in Figure 3b indicates that the number of female patients decreased from 2013 to 2014, which may have caused inconsistency in the test and training

data set trends. Therefore, because of the LSTM network's sensitivity to changes in training data trends, the LSTM network predictions were slightly inferior to those of ANN models, which used neural networks to form direct predictions. PSOSVR and TBATS performed better than the single models.

3.3.4. Patients Aged 75~79 Years

The prevalence of dementia among patients in the age range of 75 to 79 years steadily increased, as illustrated by the yellow curves in Figure 3. This regular growth should have been relatively easy for each model to predict. However, the PSOSVR prediction curve in Figure 4d was completely parallel (and close) to the SVR prediction curve. Even under PSO, SVR failed to identify more suitable parameters, resulting in the use of similar parameters for both PSOSVR and SVR.

3.3.5. Patients Aged 80~84 Years

Male and female patients in this age range exhibited considerably different trends from 2014 to 2017 (see Figure 3). The curves of male patients flattened or even decreased, whereas those of female patients increased significantly. Because of the abnormal downward trend in male patients' test data (Figure 3), obtaining accurate prediction results was difficult for the algorithms. These challenges are reflected in the predictions of SVR, PSO-SVR, ANN, and other statistical models. Due to its memorization ability, the LSTM network produced more stable predictions, consequently achieving the highest prediction accuracy for this age group.

Similar and weak prediction results were obtained by other statistical models for this age group. In the case of a large deviation between the training and test data, both the ANN and LSTM network maintained a certain level of prediction accuracy. The prediction result of the LSTM network was still the most reliable because of its high sensitivity to changes in the training data. Minor trends can be included to improve the training of the model, thereby assisting the LSTM network in achieving even higher prediction accuracy.

3.3.6. Patients Over 85 Years Old

Patients over 85 years of age constituted the largest age group of patients with dementia. As the population aged, the number of patients over the age of 85 years also increased in the test set. SVR, PSOSVR, and other statistical models exhibited difficulty in predicting small trends in the training data set. Even though the prediction results of both the ANN and LSTM network were close to the actual values, the LSTM network reported the exact value in some sections. Overall, compared with other models, the LSTM network produced more accurate predictions of small trends.

Figures 4 and 5 present the predicted numbers of male and female with dementia from 2013 to 2017 obtained by various models (ETS, ARIMA, TBATS, SVR, PSOSVR, ANN, and LSTM network model). The solid black lines represent the number of real patients, and the dashed red lines represent the prediction result of the LSTM model proposed in this study. The experimental results verified that the LSTM network is the model with the lowest average error value in different age blocks and thus the optimal prediction model in this study.

4. Discussion

In this study, we analyzed the patients diagnosed as having dementia from 1997 to 2017 in annual units from a data set extracted from the Health Insurance Database of the Ministry of Health and Welfare in Taiwan. To further verify the validity of the proposed model, the statistical models (ETS, ARIMA, and TBATS), hybrid models (SVR and PSO-SVR), and deep learning models (ANN and LSTM) were compared. Overall, the RMSE

and MAPE demonstrated that LSTM network has superior performance than other existed models. In this section, we discussed the statistical models, hybrid models, and deep learning models.

4.1. Statistical Models: Comparison of the ETS, ARIMA, and TBATS Models

The statistical models used for comparison were the ETS, ARIMA, and TBATS models, as listed in Tables 4 and 5. Both ETS and ARIMA are classic time series forecasting models. However, both have disadvantages for predictions based on data from multiple time periods. If the series has a single root (nonstationary series) or is not adjusted to the appropriate lagging period, then the model cannot achieve a high accuracy. Therefore, complicated preprocessing is required, such as statistical testing of the sequence. Developed from ETS, TBATS is a seasonal model that can predict seasonal time series more effectively. The research results indicate that TBATS has a higher average error for predictions among female patients, which may be because of the instable number of dementia patients; therefore, TBATS did not exhibit sufficient performance improvements.

4.2. Hybrid Models: Comparison of the SVR and PSOSVR Models

Support vector regression (SVR) is a popular choice for prediction and curve fitting for linear and non-linear regression types. Formulated as an optimization problem, SVR can determine the optimal regression model by using the epsilon function, which is mapped to the hyperplane of the solution space. This model has the advantage of adapting to multidimensional tasks and producing suitable predictions for nonlinear data. Therefore, this study used SVR for sequence prediction. The results revealed that the prediction error of SVR was higher than that of statistical models because SVR has three hyperparameters—namely C, σ , and ε . If the hyperparameters are not properly adjusted, the model's predictive ability cannot achieve optimal performance. PSO has been widely used to solve the hyperparameter optimization problem [49,52]. As a result, PSO was used in the present study to adjust the hyperparameters of the SVR by adjusting the parameters to an optimal combination and reduced prediction error. Tables 3 and 4 indicate that the error rate of PSOSVR was considerably lower than that of SVR and was better than most statistical models.

4.3. Deep Learning Models: Comparison of the ANN and LSTM Network Models

ANNs are developed through imitation of the neuron transmission in the human brain. A shallow neural network based on back-propagation has the advantages of efficient training, high accuracy, and suitability for data sets with noise. ANNs demonstrate excellent predictive ability but also have numerous shortcomings, such as the use of multiple hyperparameters, proneness to overfitting, gradient disappearance, gradient explosion, and long-term dependence problems. Numerous neural network models have been used to solve time series forecasting problems. RNNs have received extensive attention [50] because of their internal state and short-term memory. RNNs store a vector for each step, which is especially important when the input data contain short-term correlations. However, because of the vanishing gradient problem, the model has difficulty learning the long-term correlations of the input sequence if the stochastic gradient descent is used to train the model.

In the LSTM network, the special valve structure (gate) can avoid gradient vanishing in a deep network. Furthermore, the memory unit enhances the long-term memory capability and overall prediction efficiency. The results of this study confirm that the LSTM network has the lowest prediction error, with the average MAPE falling between 2.50% and 3.12%, demonstrating its excellent prediction accuracy. This study used *p*-values and R^2 (coefficient of determination) to statistically analyze the prediction results and verify the significance and interpretability of the proposed model. In this study, if the p value is significant, it implies that the difference between the two is obvious, which proves that LSTM has a much lower prediction error than other models. Tables 3 and 4 show that all models exhibited significant differences. The prediction ability of the LSTM network was higher than that of the ETS, ARIMA, TBATS, SVR, PSOSVR, and ANN models. The R² value reflects the proportion of variance of the dependent variable that can be explained by the independent variable, and it is often used for regression models. Higher R² values indicate better explanatory power of the model. Table 6 demonstrates that most of the R² values of the LSTM network were substantially higher than those of the ETS, ARIMA, TBATS, SVR, PSOSVR, and ANN models, which indicates that the LSTM model optimally fit the original data and had the highest explanatory power of all models.

Age	ETS	ARIMA	TBATS	SVR	PSOSVR	ANN	LSTM
60~64	0.7022	0.7854	0.8510	0.9868	0.4920	0.1022	0.9971
65~69	0.8606	0.8627	0.8193	0.8994	0.8812	0.8606	1.0000
70~74	0.9116	0.9111	0.9799	0.9237	0.9116	0.9116	0.9259
75~79	0.8941	0.8953	0.9131	0.3873	0.4873	0.9112	0.9441
80~84	0.7545	0.8330	0.7861	0.6721	0.0881	0.1545	0.9545
85 and	0.7602	0 7757	0.0207	0.8665	0.0272	0.7602	0.0607
above	0.7692	0.7737	0.9297	0.0005	0.9273	0.7692	0.9092
Average	0.6153	0.6272	0.8915	0.6772	0.6312	0.6330	0.9368

Table 6. R² of dementia patients by different models

Boldface, the best values in each row. ETS, Exponential smoothing; ARIMA, Autoregressive integrated moving average; SVR, Support vector regression; PSOSVR, Particle swarm optimization integrated Support vector regression; TBATS, Trigonometric seasonality Box–Cox transformation ARMA errors Trend Seasonal components; ANN, Artificial neural network; LSTM, Long shortterm memory.

4.4. Dementia Prevention and Interventions.

With declining mortality in younger populations, dementia is expected to become one of the greatest global health concerns of the 21st century. Although dementia is not curable, its management and the delay of its manifestation are considered to be theoretically possible. Studies have reported that the course of the disease can be modified with adequate care [53], which supports the focus on the manipulation of modifiable risk factors. In 2017, nine potentially modifiable risk factors were reported, including hypertension, obesity, depression, and low social contact [3]. Three new modifiable risk factors, namely traumatic brain injury, excessive alcohol consumption, and air pollution, were introduced in 2020, with convincing evidence [3]. Approximately 30–50% of dementia cases are attributed to these potentially modifiable risk factors. A reduction of 10–25% of these risk factors can reduce the number of patients with dementia by 1.1 to 3 million worldwide [54]. Furthermore, postponement of the onset of dementia by even 2 years can reduce the burden on public health, society, interpersonal relationships, and the economy [55]. Based on this study's proposed model, policy administrators, medical workers, and stakeholders can implement more effective and extensive societal policies on dementia prevention and care among society. The following suggestions on dementia prevention and care are provided for the aim of a more dementia-friendly society.

Promoting resilience in an aging society is a far-reaching approach to dementia prevention. The maximization of care quality and reduction of dementia incidence should begin at the community level, including through the promotion of dementia awareness and knowledge. According to the UK National Institute of Health and Care Excellence and the US National Institute of Health, social isolation is a potentially modifiable risk factor [56,57]. Aging people may experience loneliness and a lack of social contact and social participation, and the promotion of social engagement opportunities is necessary within the community. Moreover, education and intellectual stimulation alternatives have been demonstrated to enhance cognitive resilience later in life [58]. Therefore, within communities, the establishment of supportive social networks that encourage interaction will alleviate loneliness, hence reduce dementia possibilities.

The cost and burden of dementia care are tremendous and continue to rise as the global population ages. The average total cost incurred by patients with dementia exceeds the total costs of patients with other diseases [8]. Patients with dementia are often elderly people approaching their last years of life; thus, their workforce productivity is naturally weaker. As a group that has relatively low capability of coping with such a household financial crisis, the illness contributes to patients' cognitive and physical burden and hinders the ability of their families to afford future health care [8]. Therefore, especially financially, dementia care often calls for more medical health care support than other illnesses [9]. To actively promote high-quality dementia care, additional medical expenditure on dementia care and prevention is necessary and strongly recommended.

Furthermore, the prevalence of dementia affects not only patients but also their family or the health care workers who must live with these patients and deal with the behavioral and emotional effects of dementia. As mentioned earlier, patients with dementia often also experience disorientation, confusion, mood instability, and behavioral or psychological symptoms. As a result, under high pressure for an extended period of time, studies have reported that informal caregivers of patients with dementia often develop poor mental health, and have a relatively high mortality rate [5,10–12], which results in further socioeconomic problems. Additional dementia-care training is necessary for the development of adequate dementia care, which should also include the emphasis on caregivers' mental and physical health. Not only the quality of care of patients with dementia, but also their family caregivers should be emphasized in future policies; an expand in the medical allowance and societal support for this particular population should be considered.

Prevention is more effective than a cure. Proper social welfare and public health policies necessitate a precise model for predicting the prevalence rate of dementia. The purpose of this study was to examine whether an alternative prediction model, namely the proposed LSTM network, could effectively predict the trends among the population of patients with dementia. The results demonstrate that the proposed model was not only applicable, but also significantly more accurate than the other models. This precise model can successfully predict the prevalence of dementia and can thus aid government administrations in the development of relevant strategies. For example, policymakers can manage the budget allocated to dementia care to reduce its occurrence by implementing legislative changes, developing preventive interventions for younger populations, and providing ongoing education and care for elderly adults and their families. Future research is warranted to investigate the performance of the proposed LSTM network for the prediction of trends in other illnesses.

4.5. Contribution of This Paper

In this study, we analyzed the patients diagnosed as having dementia from 1997 to 2017 and used seven models to forecast the number of patients. This paper's contribution is listed below: (1) analysis of the dementia patient data and figure out the long-term dependency; (2) construction of the LSTM forecasting model; (3) successful forecast of the prevalence of dementia using the LSTM model; (4) provide aid to the government administrations in developing relevant strategies. For example, policymakers can manage the budget allocated to dementia care to reduce its occurrence by implementing legislative changes, developing preventive interventions for younger populations, and providing ongoing education and care for elderly adults and their families. Future research is warranted to investigate the performance of the proposed LSTM network for the prediction of trends in other illnesses Furthermore, the successful application of LSTM in the se-

quence prediction task of this study will significantly improve the prediction of the prevalence of dementia patients if more clinical variables can be analyzed in the future, which achieves the original intention of this study. Prevention is more effective than a cure. Lastly, the LSTM model can also be widely applied in many fields, such as vessel trajectory prediction [59], tidal level forecasting [60], financial market forecasting [61], and realtime crash risk prediction.

5. Conclusions

The accurate prediction of the trends and prevalence of dementia among people of different genders and ages would strongly assist in providing evidence for the development of interventions to prevent or delay dementia onset. The proposed LSTM network demonstrated a higher prediction accuracy compared with ETS, ARIMA, TBATS, SVR, PSOSVR, and ANN models. The prevalence was further analyzed among patients from different gender and age groups to further elucidate the prediction results. Continued effort in the development of advanced prediction models can provide evidence for health care professionals to further improve the care and interventions for people with dementia and their family caregivers. Successful dementia prevention, treatment, and support programs would dramatically reduce the burden on health care systems, individuals, societies, and economies. As the aging population continues to grow, the development of health and social care strategies for patients with dementia using accurate time series models will inevitably be an ongoing process. Being equipped to adequately address dementia will likely be one of the ultimate indicators of societal advancement in the future world.

Author Contributions: Conceptualization, S.Y.; Methodology, H.-C.C. and C.-H.Y.; Software, C.-H.W.; Validation, S.Y., C.-H.W. and M.-N.W.; Formal analysis, S.Y., H.-C.C., and M.-N.W.; Investigation, C.-H.W., M.-N.W., and C.-H.Y.; Resources, C.-H.W.; Data curation, C.-H.W. and M.-N.W.; Writing—original draft preparation, S.Y. and H.-C.C.; Writing—review and editing, S.Y., H.-C.C., M.-N.W., and C.-H.Y.; Visualization, C.-H.W.; Supervision, H.-C.C. and C.-H.Y.; Project administration, C.-H.Y.; Funding acquisition, C.-H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: The funding sources are the Ministry of Science and Technology, Taiwan (under grant no. 108-2221-E-992-031-MY3).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: https://www.mohw.gov.tw/np-126-2.html.

Acknowledgments: We would like to thank the reviewers for their valuable comments, which help us to improve our paper a lot.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Liao, G. MOI: Taiwan officially becomes an aged society with people over 65 years old breaking the 14% mark. *Taiwan News*, April 10, 2018. https://www.taiwannews.com.tw/en/news/3402395
- Strong, 2026. 2. Μ. Taiwan will be a super-aged society bv Taiwan News, 12 February, 2019. https://www.taiwannews.com.tw/en/news/3636704
- Livingston, G.; Huntley, J.; Sommerlad, A.; Ames, D.; Ballard, C.; Banerjee, S.; Brayne, C.; Burns, A.; Cohen-Mansfield, J.; Cooper, C. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet* 2020, 396, 413–446.
- 4. Cerejeira, J.; Lagarto, L.; Mukaetova-Ladinska, E. Behavioral and psychological symptoms of dementia. *Front. Neurol.* **2012**, *3*, 73.
- Allen, A.P.; Curran, E.A.; Duggan, Á.; Cryan, J.F.; Chorcorain, A.N.; Dinan, T.G.; Molloy, D.W.; Kearney, P.M.; Clarke, G. A systematic review of the psychobiological burden of informal caregiving for patients with dementia: Focus on cognitive and biological markers of chronic stress. *Neurosci. Biobehav. Rev.* 2017, 73, 123–164.

- 6. Sabat, S.R. Dementia in developing countries: A tidal wave on the horizon. Lancet 2009, 374, 1805–1806.
- 7. Patterson, C. World Alzheimer Report 2018;; Alzheimer's Disease International: London, UK, 2018.
- Kelley, A.S.; McGarry, K.; Gorges, R.; Skinner, J.S. The burden of health care costs for patients with dementia in the last 5 years of life. Ann. Intern. Med. 2015, 163, 729–736.
- 9. Ory, M.G.; Hoffman, R.R., III; Yee, J.L.; Tennstedt, S.; Schulz, R. Prevalence and impact of caregiving: A detailed comparison between dementia and nondementia caregivers. *Gerontologist* **1999**, *39*, 177–186.
- Baumgarten, M.; Hanley, J.A.; Infante-Rivard, C.; Battista, R.N.; Becker, R.; Gauthier, S. Health of family members caring for elderly persons with dementia: A longitudinal study. *Ann. Intern. Med.* 1994, 120, 126–132.
- 11. Mahoney, R.; Regan, C.; Katona, C.; Livingston, G. Anxiety and depression in family caregivers of people with Alzheimer disease: The LASER-AD study. *Am. J. Geriatr. Psychiatry* **2005**, *13*, 795–801.
- Stall, N.M.; Kim, S.J.; Hardacre, K.A.; Shah, P.S.; Straus, S.E.; Bronskill, S.E.; Lix, L.M.; Bell, C.M.; Rochon, P.A. Association of informal caregiver distress with health outcomes of community-dwelling dementia care recipients: A systematic review. J. Am. Geriatr. Soc. 2019, 67, 609–617.
- 13. Box, G.E.; Jenkins, G.M. Time Series Analysis: Forecasting and Control San Francisco; Holden-Day: San Francisco, CA, USA, 1976.
- 14. Brown, R.G. *Exponential Smoothing for Predicting Demand*; Operations Research, Inst Operations Research Management Sciences: Linthicum, MD, USA, 1957, p. 145.
- 15. Katimon, A.; Shahid, S.; Mohsenipour, M. Modeling water quality and hydrological variables using ARIMA: A case study of Johor River, Malaysia. *Sustain. Water Resour. Manag.* **2018**, *4*, 991–998.
- Smyl, S. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *Int. J. Forecast.* 2020, *36*, 75–85.
- 17. Khashei, M.; Bijari, M.; Ardali, G.A.R. Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs). *Neurocomputing*. **2009**, *72*, 956–967.
- Nasser, I.M.; Abu-Naser, S.S. Predicting Tumor Category Using Artificial Neural Networks. International Journal of Academic Health and Medical Research. 2019, 3, 1-7.
- Drucker, H.; Burges, C.J.; Kaufman, L.; Smola, A.; Vapnik, V. Support vector regression machines. *Adv. Neural Inf. Process. Syst.* 1997, 9, 155–161.
- Liu, H.-H.; Chang, L.-C.; Li, C.-W.; Yang, C.-H. Particle swarm optimization-based support vector regression for tourist arrivals forecasting. *Comput. Intell. Neurosci.* 2018, 2018, 6076475.
- Aljarah, I.; Faris, H.; Mirjalili, S. Optimizing connection weights in neural networks using the whale optimization algorithm. Soft Comput. 2018, 22, 1–15.
- Al-Fugara, A.K.; Ahmadlou, M.; Al-Shabeeb, A.R.; AlAyyash, S.; Al-Amoush, H.; Al-Adamat, R. Spatial mapping of groundwater springs potentiality using grid search-based and genetic algorithm-based support vector regression. *Geocarto Int.* 2020, 1–20, doi:10.1080/10106049.2020.1716396.
- 23. Kawakami, K. Supervised Sequence Labelling with Recurrent Neural Networks. Ph. D. Thesis, Carnegie Mellon University, Pennsylvania, USA, 2008.
- 24. Lipton, Z.C.; Berkowitz, J.; Elkan, C. A critical review of recurrent neural networks for sequence learning. *arXiv* 2015, arXiv:1506.00019.
- Cui, R.; Liu, M.; Initiative, A.S.D.N. RNN-based longitudinal analysis for diagnosis of Alzheimer's disease. Comput. Med. Imaging Graph. 2019, 73, 1–10.
- 26. Maragatham, G.; Devi, S. LSTM model for prediction of heart failure in big data. J. Med. Syst. 2019, 43, 1–13.
- 27. Lipton, Z.C.; Kale, D.C.; Elkan, C.; Wetzel, R. Learning to diagnose with LSTM recurrent neural networks. *arXiv* 2015, arXiv:1511.03677.
- Wang, L.; Sha, L.; Lakin, J.R.; Bynum, J.; Bates, D.W.; Hong, P.; Zhou, L. Development and validation of a deep learning algorithm for mortality prediction in selecting patients with dementia for earlier palliative care interventions. *JAMA Netw. Open* 2019, 2, e196972.
- 29. Khalid, S.; Goldenberg, M.; Grantcharov, T.; Taati, B.; Rudzicz, F. Evaluation of Deep Learning Models for Identifying Surgical Actions and Measuring Performance. *JAMA Netw. Open* **2020**, *3*, e201664.
- Brookmeyer, R.; Abdalla, N.; Kawas, C.H.; Corrada, M.M. Forecasting the prevalence of preclinical and clinical Alzheimer's disease in the United States. *Alzheimer's Dement.* 2018, 14, 121–129.
- 31. Fisher, C.K.; Smith, A.M.; Walsh, J.R. Machine learning for comprehensive forecasting of Alzheimer's Disease progression. *Sci. Rep.* **2019**, *9*, 1–14.
- 32. Aschwanden, D.; Aichele, S.; Ghisletta, P.; Terracciano, A.; Kliegel, M.; Sutin, A.R.; Brown, J.; Allemand, M. Predicting cognitive impairment and dementia: A machine learning approach. *J. Alzheimer's Dis.* **2020**, 1–12, Preprint.
- Kingston, A.; Comas-Herrera, A.; Jagger, C. Forecasting the care needs of the older population in England over the next 20 years: Estimates from the Population Ageing and Care Simulation (PACSim) modelling study. *Lancet Public Health* 2018, 3, e447–e455.
- 34. Ahmadi-Abhari, S.; Bandosz, P.; Shipley, M.J.; Whittaker, H.; Middleton, L.T.; Kivipelto, M.; Brunner, E.; Kivimaki, M. Forecasts for numbers of people living with dementia to 2050 and estimates for impact of smoking cessation: A modelling study in four European countries: Epidemiology/Prevalence, incidence, and outcomes of MCI and dementia. *Alzheimer's Dement.* 2020, 16, e046674.

- 35. Chimmula, V.K.R.; Zhang, L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaossolitons Fractals* **2020**, *135*, 109864.
- Shen, Z.; Zhang, Y.; Lu, J.; Xu, J.; Xiao, G. A novel time series forecasting model with deep learning. *Neurocomputing* 2020, 396, 302–313.
- Bengio, Y.; Simard, P.; Frasconi, P. Learning Long-Term Dependencies with Gradient Descent Is Difficult. *IEEE Trans. Neural Netw.* 1994, 5, 157–166.
- 38. Hochreiter, S.; Schmidhuber, J. Long short-term memory. Neural Comput. 1997, 9, 1735–1780.
- Fagerström, J.; Bång, M.; Wilhelms, D.; Chew, M.S. LiSep LSTM: A Machine Learning Algorithm for Early Detection of Septic Shock. Sci. Rep. 2019, 9, 1–8.
- 40. Graves, A.; Jaitly, N. Towards End-to-End Speech Recognition with Recurrent Neural Networks. In Proceedings of the International Conference on Machine Learning, Beijing, China, June 21-26, **2014**; pp 1764–1772.
- 41. Huang, K.-Y.; Wu, C.-H.; Su, M.-H. Attention-based convolutional neural network and long short-term memory for short-term detection of mood disorders based on elicited speech responses. *Pattern Recognit.* **2019**, *88*, 668–678.
- Nejedly, P.; Kremen, V.; Sladky, V.; Cimbalnik, J.; Klimes, P.; Plesinger, F.; Viscor, I.; Pail, M.; Halamek, J.; Brinkmann, B. Exploiting graphoelements and convolutional neural networks with long short term memory for classification of the human electroencephalogram. *Sci. Rep.* 2019, *9*, 1–9.
- 43. Brown, R.G.; Meyer, R.F. The fundamental theorem of exponential smoothing. Oper. Res. 1961, 9, 673-685.
- 44. De Livera, A.M.; Hyndman, R.J.; Snyder, R.D. Forecasting time series with complex seasonal patterns using exponential smoothing. J. Am. Stat. Assoc. 2011, 106, 1513–1527.
- 45. Vapnik, V.N. The Nature of Statistical Learning Theory. Springer-Verlag: New York, 1995.
- 46. Vapnik, V.N. An overview of statistical learning theory. IEEE Trans. Neural Netw. 1999, 10, 988–999.
- Kennedy, J.; Eberhart, R. Particle Swarm Optimization. In Proceedings of the ICNN'95-International Conference on Neural Networks, Perth, Australia, 27 November–1 December 1995; pp. 1942–1948.
- Moscoso-López, J.-A.; Turias, I.T.; Come, M.; Ruiz-Aguilar, J.; Cerbán, M. Short-term forecasting of intermodal freight using ANNs and SVR: Case of the Port of Algeciras Bay. *Transp. Res. Proceedia* 2016, 18, 108–114.
- 49. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. Nature 1986, 323, 533–536.
- 50. Cauchy, A. Méthode générale pour la résolution des systemes d'équations simultanées. Comp. Rend. Sci. Paris 1847, 25, 536–538.
- 51. Chuang, L.-Y.; Lin, Y.-D.; Chang, H.-W.; Yang, C.-H. An improved PSO algorithm for generating protective SNP barcodes in breast cancer. *PLoS ONE* 2012, *7*, e37018.
- 52. Livingston, G.; Sommerlad, A.; Orgeta, V.; Costafreda, S.G.; Huntley, J.; Ames, D.; Ballard, C.; Banerjee, S.; Burns, A.; Cohen-Mansfield, J. Dementia prevention, intervention, and care. *Lancet* **2017**, *390*, 2673–2734.
- 53. Barnes, D.E.; Yaffe, K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. *Lancet Neurol.* **2011**, *10*, 819–828.
- 54. Brodaty, H.; Breteler, M.M.; DeKosky, S.T.; Dorenlot, P.; Fratiglioni, L.; Hock, C.; Kenigsberg, P.A.; Scheltens, P.; De Strooper, B. The world of dementia beyond 2020. *J. Am. Geriatr. Soc.* **2011**, *59*, 923–927.
- Daviglus, M.L.; Bell, C.C.; Berrettini, W.; Bowen, P.E.; Connolly Jr, E.S.; Cox, N.J.; Dunbar-Jacob, J.M.; Granieri, E.C.; Hunt, G.; McGarry, K. NIH state-of-the-science conference statement: Preventing Alzheimer's disease and cognitive decline. *NIH Consens. State Sci. Statements* 2010, *27*, 1–30.
- 56. Health, N.I.F.; Excellence, C. Dementia, Disability and Frailty in Later Life-Mid-Life Approaches to Delay or Prevent Onset; National Institute for Health and Care Excellence (NICE): Raanana, Israel, 2015.
- Borenstein, A.; Mortimer, J. Alzheimer's Disease: Life Course Perspectives on Risk Reduction; Academic Press: Cambridge, MA, USA, 2016.
- 58. Tang, H.; Yin, Y.; Shen, H. A model for vessel trajectory prediction based on long short-term memory neural network. *J. Mar. Eng. Technol.* **2019**, 1–10, doi:10.1080/20464177.2019.1665258.
- 59. Yang, C.-H.; Wu, C.-H.; Hsieh, C.-M. Long Short-Term Memory Recurrent Neural Network for Tidal Level Forecasting. *IEEE Access* **2020**, *8*, 159389–159401.
- Bukhari, A.H.; Raja, M.A.Z.; Sulaiman, M.; Islam, S.; Shoaib, M.; Kumam, P. Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting. *IEEE Access* 2020, *8*, 71326–71338.
- 61. Li, P.; Abdel-Aty, M.; Yuan, J. Real-time crash risk prediction on arterials based on LSTM-CNN. Accid. Anal. Prev. 2020, 135, 105371.