



# Forecasting Causes of Death in Northern Iraq Using Neural Network

Shahla O. Salih<sup>1</sup> · Antonia Bezenchek<sup>2,3</sup> · Stefania Moramarco<sup>4</sup> ·  
Manuel De Iulii<sup>5</sup> · Denis Stanev<sup>6</sup> · Iuri Fanti<sup>2</sup> · Dario Corsini<sup>2</sup> · Faiq B. Basa<sup>7</sup> ·  
Leonardo Emberti Gialloreti<sup>4</sup>

Received: 25 November 2021 / Accepted: 13 April 2022  
© The Author(s) 2022

## Abstract

The availability of models for predicting future events is essential for enhancing the efficiency of systems. This paper attempts to predict demographic variation by employing multi-layer perceptron network. Here we present the implementation of a system for predicting the number and causes of deaths, for a future 2-year period. The system was built using predictive models and data that is as accurate as possible under the current conditions of the northern Region of Iraq (the Autonomous Region of Kurdistan). Our predictive model is based on quarterly periods, with the intention of providing predictions on the number of deaths, classified by gender, cause of death, age at death, administrative district (governorate), and hospital where the death occurred. The data was collected from birth and death registry bureaus and forensic medicine departments for the years 2009–2020. The python programming language was used to test the designed multi-layer perceptron network with back-propagation training algorithm. With learning rate 0.01 and 500 epochs we were able to obtain good results, as the neural network was able to represent the string, and predict future values well, with a mean squared error of 0.43, and we found that number of deaths is quite stable, with a slight increase.

**Keywords** Forecasting · Neural network · Time series · Mortality · Cause of death · Iraq

## 1 Introduction

In order to seek a better and more sustainable future for all, the sustainable development goals (SDGs), a set of 17 global goals, were drawn up. In 2015, 195 countries agreed that they could change the world for the better. This will

---

✉ Shahla O. Salih  
shahla.salih@univsul.edu.iq

Extended author information available on the last page of the article

be achieved by bringing together their governments, companies, media, higher education and non-governmental institutions to mend the lives of people in their countries by 2030 [1]. Although the SDGs are not legally binding, governments are expected to take charge and set up a national structure to pursue them. Therefore, States will bear the most responsibility for monitoring and reviewing progress made. This requires the collection of qualitative data in a timely manner, so that follow-ups and reviews at a regional level are based on analyzes conducted at the national level, contributing to global follow-up and review. One of the SDGs, goal number 3, is good health and wellbeing; reducing the number of deaths is a part of this goal (goal number 3.9). Increasing birth rates without increasing neonatal and child mortality and, in general, decreasing the mortality rate is considered as a sign of good health and wellbeing of any community. One way to find out where a community is regarding this goal is to use statistical forecasts.

A time series is a set of observations resulting in statistical data ordered by time. The study of time series has become a fertile field, as it is closely related to human life and daily activities [2]. Time series analysis aims to find an appropriate model that simulates and explains the studied behavior, and thus the ability to forecast future values [3]. Forecasting is of great importance in all fields, so researchers and experts have taken an interest in this field, developing many predictive techniques and models and using them to obtain more accurate results. The use of an artificial neural network model has proved to be efficient in several areas.

Artificial intelligence is a branch of computer science and information technology, and can be defined as visualizing the representation, simulation and modeling of human mental abilities by means of a computer [4]. This branch of science has become one of the most active fields in recent times, and it has many and multiple uses, one of which is to predict what will happen in the future, by monitoring the data of a phenomenon or an event for a certain period of time and then predicting what this phenomenon or event will do in the future. The most commonly used neural networks consist of three layers, which have an input layer, a hidden layer(s), and an output layer. The input is fed into the network through the input layer, and the information is processed inside the hidden layer(s), then the output is provided by the output layer. The transmutation components in each layer are called neurons or nodes. The number of neurons and layers hidden within the network depends on the problem, and is determined by the train and error method [5].

Forecasting using artificial neural networks is a modern method that has received wide attention in many fields. It has been widely used because it can explain the behavior of non-linear data, and does not require strict and precise conditions for the purpose of forecasting. In this paper, we forecast deaths in Northern Iraq (Kurdistan Region) using a backpropagation training algorithm to train a network consisting of a multi-layer perceptron neural network for the future planning and knowledge of community characteristics. Also, none of the researches published in the region dealt with all the characteristics of the data studied in this paper. We also look at the strengths and weaknesses of births and deaths registry offices in the region, and attempt to remedy data input problems by suggesting a unified data input strategy.

## 1.1 Literature Review [6–8]

In today's culture, accurate mortality estimates are more vital than ever. Due to the increased societal benefit of more precise death estimates, the field of mortality prediction is expanding and progressing. Several strategies for modeling and predicting mortality have been developed in recent years. The deductive guideline method for estimating deaths for numerous years was the Lee-Carter (LC) mortality projection technique (1992). The age-period mortality model looks at the overall time trend, mortality patterns by age, and age disparities in the magnitude of total time change for a given population over a given time period. The Lee-Carter model has inspired a slew of variants, extensions and alternatives since its inception. Renshaw and Haberman propose a multi-factor variation (2003). Brouhns et al. (2002) estimate the parameters of an LC model using a Poisson distribution and log-likelihood maximization. Renshaw and Haberman's (2006) one-factor LC model was updated to include cohort impact. Young-adult mortality was studied by O'Hare and Li (2012). In his most recent publication, Currie (2016) provides a comprehensive discussion of generalized linear and non-linear mortality models. Han Li Shang and Steven Haberman talk about the recent practice of merging multiple projections rather than just one. Advances in mortality modeling based on machine and deep learning models have been reported in the literature. In the early 1990s, Kramer (1991) reported a neural-network-based generalization of PCA to the problem of non-linear feature extraction in chemical engineering literature. To forecast maternal mortality in a Nigerian region, Abdulkarim and Garko (2015) employed a feed-forward neural network with a particle swarm approach. Puddu and Menotti (2012) extended their approach to rural Italy, showing no difference in performance between multi-layer perceptrons and multiple logistic regressions when predicting 45-year all-cause death. Hainaut (2018) investigated a Neural Network design for mortality analysis with the purpose of extrapolating acceptable non-linearity in observed mortality force. In a multiple population situation, Richman and Wuthrich (2018) employed deep neural networks to determine the parameters in the Lee-Carter model. Deep learning models, particularly recurrent neural networks (RNNs), are gaining in popularity in a wide range of forecasting applications, including mortality prediction.

## 2 Methodology

### 2.1 Data

In most developing countries the burden of recording a birth or a death falls entirely on the family. In many cases it requires significant effort and expense, and can take several weeks. This partly explains why so many births and deaths are not recorded.

Unfortunately, there are still significant gaps in the availability and quality of this important data in many parts of the world. Only about 65% of all births worldwide are recorded, about a third of the world's annual deaths are recorded through civil registration, and up to 80% of deaths occurring outside health facilities are not counted [9].

In all health directorates in the Iraqi governorates, health and vital statistics departments form part of the Health Planning Directorate. These departments receive health-related data from health facilities either on CDs or paper spreadsheets. There are also Registration Bureaus of Births and Deaths (RBBB) in most regions, and they are responsible for registering births and deaths in the surrounding areas [10]. The RBBB records deaths reported either by hospitals or by relatives. When a death occurs in a hospital, a death certificate is issued containing information about the deceased person (age, gender) and also the probable cause of death, which is confirmed by the doctor [11]. This registration system provides fragmented information, however, especially in rural areas. In the Kurdistan Region, many gaps in the recording of deaths still occur, such as:

- the death recording process varies from one region to another, with format, accuracy, completeness and accessibility of data being a challenge when processing local statistics.
- reported data on mortality is inaccurate, since there are no laws forcing relatives to register the death,
- data collection is mostly paper based, therefore important information is lacking, making vital statistics and the evaluation of outcomes difficult to measure efficiently.

We collected data on 62,000 individuals, including information such as: date of birth, date of death, gender, cause of death, place of death, hospital, from the births and deaths registry bureaus and forensic medicine departments in Sulaimanayah, Duhok, Zakho, Aqrah, and Amedi. From the collected data we found that the number one cause of death was Diseases of the circulatory system (I00-I99), and the number two cause of death was Neoplasms (C00-D49); External causes of morbidity and mortality (V00-Y99), and Certain infectious and parasitic diseases (A00-B99) are ranked third and fourth respectively. The number of deaths in the age group 60–79 is the highest among all age groups, confirming the indicator of life expectancy in Iraq at 73.6 years. Male life expectancy is 71.7 years, female life expectancy is 75.6 years [12]. We observed a large number of deaths for the age group 0–1; this aspect must be deeply investigated looking at the complete data records of offices, and will be done in the next phase of our work. Generally speaking, the infant mortality rate in Iraq fell from 83.883 deaths per 1000 live births in 1970 to 22.918 deaths per 1000 live births in 2020 [13].

## 2.2 Methods

In 2015 the University of Rome Tor Vergata (Italy) and the Kurdistan Ministry of Health started cooperating on the implementation of an electronic epidemic monitoring and health surveillance system, called KRG-HIS (Kurdistan Region Governorate Health Information System). This system is currently implemented in 125 primary healthcare centers and hospitals in different Provinces: Duhok, Erbil, Sulaymaniyah and Halabja. The KRG-HIS derives from the DHIS2 (District Health Information

Software) [14] free online software system which is based on ICD-10( International Classification of Disease) coding, and is designed to manage healthcare data by collecting, storing, managing and transmitting electronic medical records of patients, such as diagnoses, vaccinations, births and deaths [15].

Part of the information we get from this system relates to the cause of death. According to WHO (World Health Organization), analyzing why people die is one of the key elements to understanding the health situation of a region, and that is why the system is also concerned with this aspect. We collected data on deaths that are recorded in the RBBDs and Forensic Medicine Departments in different areas. The data we collected were not coded or classified, but were written as text in either Arabic or Kurdish. Each individual record was checked and coded according to the WHO international classification of diseases (ICD-10). After coding we placed our data in different categories.

### 2.3 Aim

The aim of our work is to predict deaths in Kurdistan over the next two years. These predictions are quarterly based. We also aim to predict gender, cause of death, age and governorate. To achieve this, first of all the data were normalized and classified.

The simplest data to classify is gender, with just 2 possible values: Male (M) or Female (F). Age data is simple too, we chose 7 different classes: <1, 1–4, 5–14, 15–34, 35–59, 60–79, 80+. A more complex classification was the cause of death. We chose to use the ICD10 code, and defined the 21 different classes shown in table 1 (Table 1).

At the end of this step, a single line of data has the following fields:

- Quarter of the year (related to deaths), e.g. 01/01/2008 to 04/30/2008;
- Age class, e.g. class 1–4;
- Hospital, e.g. Heevi Children’s Hospital;
- Governorate, e.g. Duhok;
- Cause of death, following the ICD-10 code, e.g. in the range C00–D49.

### 2.4 Statistical Method

The artificial neural network method is a modern knowledge processing method that has become popular, taking on a prominent role worldwide, as it continuously simulates data and non-linear functions to obtain success for a model dealing with the analysis, classification, forecasting, etc. of a phenomenon.

#### 2.4.1 Components of Artificial Neural Networks

An artificial neural network consists of a collection of connected neurons. These neurons work in parallel and at the same time. A neural network consists of three layers: Input Layer, Hidden Layer and Output Layer. It works as a simple system that collects the likely inputs and offers an answer in the form of a numeric value. A

**Table 1** Disease surveillance—Duhok events grouped by ICD10 classes

| Duhok           |       |           |   |  |  |
|-----------------|-------|-----------|---|--|--|
| Number of cases | %     | ICD10 cod | ICD10 chapter   |  |  |
| 28,649          | 6.05  | A00–B99   | Certain infectious and parasitic diseases   |  |  |
| 307             | 0.06  | C00–D48   | Neoplasms   |  |  |
| 6581            | 1.39  | D50–D89   | Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism |  |  |
| 23,496          | 4.96  | E00–E90   | Endocrine, nutritional and metabolic diseases   |  |  |
| 908             | 0.19  | F00–F99   | Mental and behavioral disorders   |  |  |
| 1653            | 0.35  | G00–G99   | Diseases of the nervous system  |  |  |
| 3677            | 0.78  | H00–H59   | Diseases of the eye and adnexa  |  |  |
| 5497            | 1.16  | H60–H95   | Diseases of the ear and mastoid process   |  |  |
| 7639            | 1.61  | I00–I99   | Diseases of the circulatory system  |  |  |
| 175,443         | 37.05 | J00–J99   | Diseases of the respiratory system  |  |  |
| 57,539          | 12.15 | K00–K95   | Diseases of the digestive system  |  |  |
| 19,337          | 4.08  | L00–L99   | Diseases of the skin and subcutaneous tissue  |  |  |
| 24,149          | 5.1   | M00–M99   | Diseases of the musculoskeletal system and connective tissue  |  |  |
| 49,493          | 10.45 | N00–N99   | Diseases of the genitourinary system  |  |  |
| 1125            | 0.24  | O00–O99   | Pregnancy, childbirth and the puerperium  |  |  |
| 147             | 0.03  | P00–P96   | Certain conditions originating in the perinatal period  |  |  |
| 1619            | 0.34  | Q00–Q99   | Congenital malformations, deformations and chromosomal abnormalities                                |  |  |
| 45,880          | 9.69  | R00–R99   | Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified             |  |  |
| 1542            | 0.33  | S00–T98   | Injury, poisoning and certain other consequences of external causes                                 |  |  |
| 3832            | 0.81  | V01–Y99   | External causes of morbidity and mortality  |  |  |
| 14,977          | 3.16  | Z00–Z99   | Factors influencing health status and contact with health services                                  |  |  |
| 473,490         | 100   |           |   |  |  |

neural network that has one input layer and one output layer where the inputs are fed directly to the output via a series of weights is called a single layer neural network, while a network that has one or more hidden layers between the input and output layers is called a multilayer neural network. The system learns through the process of determining the number of neurons in each layer and adjusting the layer of communication weights based on training data.

**2.4.1.1 Supervised Learning** When training the network, the required output is known, to be compared with the network output at each learning cycle, and the calculated error is used in the process of adjusting weights until the correct result is obtained, after which the network does not need to be trained.

**2.4.1.2 Unsupervised Learning** The learning process occurs without predetermining the desired output. In this case, the network is trained to discover the features that are not visible in the dataset used in the training process, and then uses those features to divide the input data into different and closed groups within each group.

In Fig. 1 each of the inputs ( $x_i$ ) features a weight ( $w_i$ ) that represents the strength of that exact connection. The weighted sum of the inputs as well as the bias ( $b$ ) is input to the activation function ( $f(x)$ ) to get the output ( $\hat{x}_i$ ) (Fig. 1) [16]. Activation functions help the network learn complex relationships and data patterns. Bias is used to regulate the output, together with the sum of the weighted inputs to the neuron, shifting the activation function to either the right or left. This mathematical form of this process can be written as follows:

$$\hat{x}_i = f\left(\sum_{i=1}^n w_i x_i + b\right) \tag{1}$$

**2.4.2 Learning Algorithm**

Neural networks can be classified into two categories based on the pattern of connections: feedforward neural networks, in which the spread of the signals entering the network is usually forward. As all the interconnecting lines come in one direction from the input layer to the output layer, the outgoing signals rely upon the incoming signals only. For backpropagation neural networks it is possible to refeed the outgoing signals from the network and divert their direction to become an

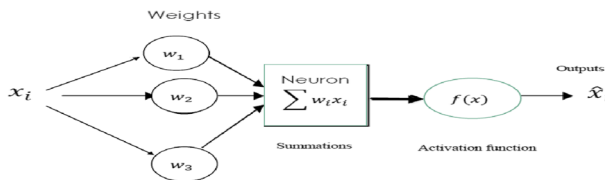


Fig. 1 Single neuron in a neural network

incoming signal, thus the signal leaving any cell depends on the signals entering it, apart from the signals previously leaving it.

- Forward propagation stage: This stage begins with displaying the form of input to the network, where each processing element of the input layer is allocated to one of the components that represents the input, in other words the network operates with a forward feeding system, and there is no adjustment to the weights. The general forward propagation equations are [17]:

$$z^{[l]} = w^{[l]}a^{[l-1]} + b^{[l]} \quad (2)$$

$$a^{[l]} = g^{[l]}(z^{[l]}) \quad (3)$$

where  $z^{[l]}$ , the activation of layer  $l$ ;  $a^{[l-1]}$ , input from layer  $l-1$ ;  $a^{[l]}$ , output of layer  $l$ ;  $g^{[l]}$ , the activation function in layer  $l$ ;  $w^{[l]}$ , weights in layer  $l$ ;  $b^{[l]}$ , bias in layer  $l$ .

The output of the final layer, which is the output of the network, is denoted by  $a^{[L]} = \hat{y}$ .

- Back propagation and Weight adjustment stage: This is when network weights are adjusted, where network outputs are compared during training with the values fed from the outside (or with the target value), and the difference between the two is calculated, then the weights are adjusted, allowing the signal to re-propagate from the output layer to the input layer in reverse during the weight adjustment phase. The process is repeated until the network outputs are the same, correct given value. The standard backpropagation algorithm is the gradient descent algorithm on the cost function:

$$C_0 = (a^{[L]} - a^{[0]})^2 = (x - \hat{y})^2 \quad (4)$$

One important operation used in the backward pass is to calculate derivatives to update each of the weights in the network so that they cause the actual output to be closer to the target output. We then calculate the partial derivative of cost function in respect of the weights (by using the chain rule) with these general equations:

$$\frac{\partial C_0}{\partial w^{[l]}} = \frac{\partial z^{[l]}}{\partial w^{[l]}} * \frac{\partial a^{[l]}}{\partial z^{[l]}} * \frac{\partial C_0}{\partial a^{[l]}} \quad (5)$$

$$\frac{\partial z^{[l]}}{\partial w^{[l]}} = (a^{[l-1]}) \quad (6)$$

$$\frac{\partial a^{[l]}}{\partial z^{[l]}} = g'^{[l]}(z^{[l]}) \quad (7)$$



$$\frac{\partial C_0}{\partial a^{[l]}} = 2(a^{[L]} - a^{[0]}) \quad (8)$$

then, we will use this result to adjust the weights by using the following equation:

$$w_{new}^{[l]} = w^{[l]} - \eta \frac{\partial C_0}{\partial w^{[l]}} \quad (9)$$

where  $\eta$  is the learning rate that scales the size of our weight updates in order to minimize the network's loss function. It is an indicator of the direction in which the weights are updated. This parameter can be fixed or updated adaptively [18]. The learning rate must be updated carefully, because it has a direct impact on the adjusted weights. This process of adjusting weights is applied to the output layer and all hidden layers, and is repeated until the desired result is obtained.

The choice of the optimizer affects both the speed of convergence and whether this actually occurs. Several alternatives to the usual gradient descent algorithms have been developed. Fit optimization algorithms such as Adam or RMS Prop perform well in the first part of the training, but generalization have been found to be poor in later stages compared to random derivative regression.

Before beginning training, we have to choose the appropriate size of the network. This is the most difficult step in the design of neural networks. In addition to the numerous options available for the activation function for each layer, there is the matter of selecting an acceptable number of layers and the choice of the optimal number of nodes in each of those layers. Choosing an improper network size leads to unacceptable results. The trial and error method is the easiest way of selecting the network size. The designer should try a variety of networks and choose the best one.

In this paper, the goal of the designed neural network was to have in each single row of the dataset the forecast number of deaths in a single governorate, with a specific age, gender, hospital and class of ICD-10 code. The neural network developed is a multilayer perceptron neural network that consists of one input layer with five nodes, three hidden layers with thirty nodes, and one output layer with one node. The activation function used is scaled exponential linear units (SELU). A 'normal' kernel initializer is used as a weight initializer. As a computing loss function we used mean squared error (MSE), and as an optimizer function Root Mean Square Propagation (RMSprop) with learning rate 0.01 and 500 epochs used for training. Loss function is a measure of how good the prediction model performs in terms of the ability to predict the expected outcome. The RMSprop optimizer restrains oscillations in the vertical direction.

Therefore, our learning rate can increase and our algorithm can take larger steps in the faster-converging horizontal direction.

The SELU activation function was chosen because it has an excellent self-normalization quality, there is no fear of vanishing gradients, and it learns faster and better than other activation functions [19], in part due to the pattern of the dataset, equivalent to SELU; furthermore, it gives more accurate results with minimum MSE.

This designed network produced good forecasting results, as shown in the results section. We should not forget that with more data we can obtain better and more accurate forecasting results. All these characteristics are valid for our datasets, and it is not necessary to get the same result for a different dataset.

In contrast to other machine learning methods, neural networks require substantially larger datasets for training. They also require a lot of computing power to be trained. A problem arises when either the dataset or the scale of the neural network has become too large. When the amount of data or the number of layers and neurons in a neural network grows, it usually does not scale well. This behavior is caused by a number of factors:

Non-linear activation functions are common in neural networks. The sequence in which we feed training data into a neural network has an impact on the outcome. The error function of neural networks frequently ends up in local minima. The only method to get around this is to train the network numerous times on different batches of the training dataset, however this requires multiplying training efforts several times.

Another issue is the weights of the network. The weight matrices scale faster than a linear model as the neural network grows in size. Finally, not all neural network topologies are created equal. Different architectures may tackle the same problem with a similar precision despite using vastly different amounts of processing. This means that we frequently use trial and error to find the most successful architecture, but we have no theoretical understanding of why it works.

Lastly, we should emphasize that neural networks are often closed systems. While they can learn abstract representations of a dataset, human analysts find it difficult to interpret these representations. This means that, while neural networks can make correct predictions in theory, we're unlikely to gain insights into the structure of a dataset through them [20].

### 3 Results

We can divide the results of this paper into the following sections:

#### 3.1 Data Processing

After collecting 62,000 individuals from RBBD we coded the causes of death according to the WHO classification (ICD-10). Table 2 shows a sample of the prepared data that we used in this study (Table 2).

#### 3.2 Designing the Network

We used Python programming language to design the neural network, starting with the following steps:

- Created a csv file for every class (governorates, hospital, age, gender).
- Created the csv file relating to ICD-10 codes;
- Developed a script to add the specific ICD-10 class.
- Developed a script to take a single row from the starting dataset and increase the death counter of the specific row in the new dataset.
- Developed a script to create a csv file containing data for forecasting. This script had two goals, the first was to divide the dataset into true and false examples, true was the rows with the number of deaths  $> 1$ , false the remaining rows, the second to put the data in a common format for the forecast data.
- All data was normalized between 0 and 1, for false, true and forecast data.
- Validation was developed to set validation and training (called test in the script) for the neural network, the concept was to take false and true rows, shuffle them and take 80% of values as a training set and 20% as a validation set.
- Converted forecast data to a JavaScript Object Notation (JSON) file;
- Converted elements of the forecast data from numbers to strings, to gain a better explanation of results.

The exact number of records used as a training set was 1,819,272, for the validation set the number was 363,856. The forecast records numbered 279,888. The best designed network for our forecasts consisted of one input layer with five nodes, three hidden layers with thirty nodes and one output layer with one node.

With learning rate 0.01 and 500 epochs, we get the forecasting result with loss function Mean Squared Error=0.43 and Mean Absolute Error=0.04, Root Mean Squared Error (RMSE)=0.65, Akaike information criterion (AIC)=3.6879, and Bayesian information criterion (BIC)=9.3469.

### 3.3 Forecasting Results

From table 3, we can conclude that the number one cause of death will be Diseases of the circulatory system (I00-I99), and the number two causes of death is Neoplasms(C00-D49), then External causes of morbidity and mortality (V00-Y99), Certain infectious and parasitic diseases (A00-B99) are ranked third and fourth respectively. The last cause of death is Diseases of the ear and mastoid process which is coded as H60-H95 (Table 3).

Table 4 gives the forecast number of deaths by gender (Table 4). Table 5 shows the forecast number of deaths by age, we conclude that the number of deaths increases in the elderly population, as the number of deaths in the 80+ age group is higher than all other groups (Table 5).

The table 6 shows the forecast number of deaths in the 20 hospitals with the highest number of deaths. It starts with Azadi Teaching Hospital, with 2496 deaths (Table 6).

**Table 2** Part of the coded data

| Gender | Date of birth | Date of death | Hospital                           | Place of death       | Hospital location | Governorate  | Cause of death                                | Code  |
|--------|---------------|---------------|------------------------------------|----------------------|-------------------|--------------|---|-------|
| Male   | 30/3/2008     | 29/12/2008    | Heevi children's hospital          | Inside the hospital  | Duhok             | Duhok        | Pneumonia, unspecified                        | J18.9 |
| Male   | 23/12/2008    | 27/12/2008    | Heevi children's hospital          | Inside the hospital  | Duhok             | Duhok        | Sepsis, unspecified                           | A41.9 |
| Female | 1963/07/01    | 2013/04/24    | Hiwa hospital                      | Inside the hospital  | Sulaimaneyah      | Sulaimaneyah | Leukemia, unspecified                         | C95.9 |
| Male   | 1/1/2009      | 3/1/2009      | Heevi children's hospital          | Inside the hospital  | Duhok             | Duhok        | Sepsis, unspecified                           | A41.9 |
| Male   | 1939/07/01    | 2013/05/25    | General surgical teaching hospital | Inside the hospital  | Sulaimaneyah      | Bagdad       | Acute coronary syndrome                       | I24.9 |
| Male   | 1987          | 26/3/2009     | Duhok forensic medicine            | Inside the hospital  | Duhok             | Duhok        | Shotgun, undetermined intent                  | Y23   |
| Female | 1/4/2008      | 15/1/2009     | Heevi children's hospital          | Inside the hospital  | Duhok             | Duhok        | Other specified disorder of kidney            | N28.8 |
| Male   | 1932          | 39.816        | Emergency Hospital                 | Inside the hospital  | Duhok             | Duhok        | Neoplasm of brain                             | D43.2 |
| Male   | 1945/05/01    | 2013/04/04    | Shorsh Hospital                    | Inside the hospital  | Sulaimaneyah      | Sulaimaneyah | Stroke  | I64   |
| Male   | 1946/07/01    | 2013/04/19    | Dr. Hemn Hospital                  | Inside the hospital  | Sulaimaneyah      | Sulaimaneyah | Chronic renal failure                         | N18.9 |
| Male   | 1934          | 25/2/2009     | Aqrah hospital                     | Outside the hospital | Aqrah             | Duhok        | Acute coronary syndrome                       | I24.9 |
| Female | 1/4/2008      | 31/12/2008    | Heevi children's hospital          | Inside the hospital  | Duhok             | Duhok        | Congenital malformation of heart, unspecified | Q24.9 |
| Male   | 1/7/1945      | 25/3/2009     | Zakho general hospital             | Outside the hospital | Zakho             | Duhok        | Stroke  | I64   |
| Male   | 1980          | 14/3/2010     | Sulaimaneyah forensic medicine     | Outside the hospital | Sulaimaneyah      | Sulaimaneyah | Effect of electric current                    | T75.4 |
| Male   | 2013/01/23    | 2013/03/02    | Maternity Hospital                 | Inside the hospital  | Kalar             | Sulaimaneyah | Sepsis, unspecified                           | A41.9 |

**Table 3** Forecast number of deaths classified by cause of death (1/1/2021–30/9/2021)

| Cause of death<br>ICD10 | Total forecasted number of death |                        |                        |                          |                        |                        |                        |                          |  |  |
|-------------------------|----------------------------------|------------------------|------------------------|--------------------------|------------------------|------------------------|------------------------|--------------------------|--|--|
|                         | 1/1/2021–<br>31/3/2021           | 1/4/2021–<br>30/6/2021 | 1/7/2021–<br>30/9/2021 | 1/10/2021–<br>31/12/2021 | 1/1/2022–<br>31/3/2022 | 1/4/2022–<br>30/6/2022 | 1/7/2022–<br>30/9/2022 | 1/10/2022–<br>31/12/2022 |  |  |
| A00-B99                 | 134                              | 84                     | 89                     | 89                       | 129                    | 83                     | 88                     | 88                       |  |  |
| C00-D49                 | 229                              | 199                    | 249                    | 234                      | 219                    | 193                    | 238                    | 226                      |  |  |
| D50-D89                 | 6                                | 4                      | 3                      | 8                        | 6                      | 4                      | 3                      | 8                        |  |  |
| E00-E89                 | 17                               | 10                     | 10                     | 10                       | 16                     | 10                     | 11                     | 11                       |  |  |
| F01-F99                 | 0                                | 0                      | 0                      | 0                        | 0                      | 0                      | 0                      | 0                        |  |  |
| G00-G99                 | 41                               | 27                     | 32                     | 34                       | 39                     | 26                     | 31                     | 33                       |  |  |
| H00-H59                 | 0                                | 0                      | 0                      | 0                        | 0                      | 0                      | 0                      | 0                        |  |  |
| H60-H95                 | 0                                | 0                      | 0                      | 0                        | 0                      | 0                      | 0                      | 0                        |  |  |
| I00-I99                 | 845                              | 679                    | 678                    | 788                      | 812                    | 654                    | 652                    | 763                      |  |  |
| J00-J99                 | 128                              | 60                     | 64                     | 75                       | 123                    | 60                     | 62                     | 75                       |  |  |
| K00-K95                 | 38                               | 38                     | 32                     | 31                       | 36                     | 37                     | 32                     | 30                       |  |  |
| L00-L99                 | 0                                | 0                      | 0                      | 0                        | 0                      | 0                      | 0                      | 0                        |  |  |
| M00-M99                 | 0                                | 1                      | 0                      | 0                        | 0                      | 1                      | 0                      | 0                        |  |  |
| N00-N99                 | 115                              | 83                     | 85                     | 103                      | 110                    | 81                     | 83                     | 99                       |  |  |
| O00-O9A                 | 1                                | 0                      | 0                      | 0                        | 1                      | 0                      | 0                      | 0                        |  |  |
| P00-P96                 | 73                               | 66                     | 61                     | 56                       | 74                     | 68                     | 64                     | 60                       |  |  |
| Q00-Q99                 | 39                               | 36                     | 29                     | 36                       | 37                     | 35                     | 28                     | 35                       |  |  |
| R00-R99                 | 40                               | 29                     | 37                     | 36                       | 40                     | 29                     | 36                     | 36                       |  |  |
| S00-T88                 | 37                               | 51                     | 42                     | 37                       | 38                     | 50                     | 43                     | 38                       |  |  |
| V00-Y99                 | 100                              | 107                    | 118                    | 97                       | 100                    | 108                    | 120                    | 100                      |  |  |
| Z00-Z99                 | 0                                | 0                      | 0                      | 1                        | 0                      | 0                      | 0                      | 1                        |  |  |

**Table 4** Forecast number of deaths classified by gender (1/1/2021–31/12/2022)

| Date                 | Gender | Deaths | Date                 | Gender | Deaths |
|----------------------|--------|--------|----------------------|--------|--------|
| 1/1/2021–31/3/2021   | M      | 930    | 1/1/2022–31/3/2022   | M      | 941    |
| 1/1/2021–31/3/2021   | F      | 745    | 1/1/2022–31/3/2022   | F      | 755    |
| 1/4/2021–30/6/2021   | M      | 933    | 1/4/2022–30/6/2022   | M      | 943    |
| 1/4/2021–30/6/2021   | F      | 748    | 1/4/2022–30/6/2022   | F      | 757    |
| 1/7/2021–30/9/2021   | M      | 935    | 1/7/2022–30/9/2022   | M      | 946    |
| 1/7/2021–30/9/2021   | F      | 750    | 1/7/2022–30/9/2022   | F      | 760    |
| 1/10/2021–31/12/2021 | M      | 938    | 1/10/2022–31/12/2022 | M      | 948    |
| 1/10/2021–31/12/2021 | F      | 752    | 1/10/2022–31/12/2022 | F      | 762    |

In table 7, we noted that the total number of deaths from Sulaimaneyah governorate are greater than other governorates, this is because the number of residents in Sulaimaneyah is greater compared to the other governorates. According to population projection by governorate and region for 2019 published by the Iraqi central statistical organization the total number of Sulaimaneyah population is 2,219,194, while the Duhok population totals 1,326,562 [21] (Table 7).

The results shown in these tables are part of forecast results that include only one category in each table, since we cannot show all the categories in one table. Also, some tables are so long they could not be presented in this paper. We are working on implementing a web application to show all the results. In the WebApp users can view all the results of the forecast together or the results of a specific category. Users can for example view the total forecast number of deaths in the age class 5–14 in the period 1/1/2021–31/3/2021, and in this specific category we can determine the proportion of males and females, proportions of cause of death, places of death and governorate.

Table 8 shows the total forecast number of deaths by quarterly period between 1/1/2021 and 31/12/2022. From the above table we can see a quite stable situation, with a slight increase in the number of deaths. This is not a bad indicator in view of the continuous increase in the population [22] (Table 8).

**Table 5** Forecast number of deaths classified by age group (1/1/2021–30/9/2021)

| Age group | Total forecasted number of death |                        |                        |                          |                        |                        |                        |                          |
|-----------|----------------------------------|------------------------|------------------------|--------------------------|------------------------|------------------------|------------------------|--------------------------|
|           | 1/1/2021–<br>31/3/2021           | 1/4/2021–<br>30/6/2021 | 1/7/2021–<br>30/9/2021 | 1/10/2021–<br>31/12/2021 | 1/1/2022–<br>31/3/2022 | 1/4/2022–<br>30/6/2022 | 1/7/2022–<br>30/9/2022 | 1/10/2022–<br>31/12/2022 |
| 0–1       | 191                              | 192                    | 192                    | 193                      | 194                    | 194                    | 195                    | 196                      |
| 2–4       | 207                              | 207                    | 208                    | 209                      | 209                    | 210                    | 211                    | 211                      |
| 5–14      | 222                              | 223                    | 224                    | 225                      | 225                    | 226                    | 227                    | 227                      |
| 15–34     | 239                              | 240                    | 240                    | 241                      | 242                    | 242                    | 243                    | 244                      |
| 35–59     | 255                              | 256                    | 257                    | 258                      | 258                    | 259                    | 260                    | 261                      |
| 60–79     | 272                              | 273                    | 274                    | 274                      | 275                    | 276                    | 277                    | 277                      |
| 80+       | 289                              | 290                    | 290                    | 291                      | 292                    | 293                    | 293                    | 294                      |

## 4 Conclusion and Discussion

Setting up a good forecasting model is important because of the impact of its results on the various processes of social and economic planning of a country. In this paper we chose artificial neural networks to forecast mortality in Kurdistan Region for the next two years because their results are more accurate and efficient in forecasting than traditional statistical methods. The best designed network for our data forecasts was one input layer with five nodes, three hidden layers with thirty nodes and one output layer with one node. We approved forecasting results with loss function Mean Squared Error=0.43, Mean Absolute Error=0.04.

From the results of running our network we observe that the number one cause of death is Diseases of the circulatory system, while the number two cause of death is neoplasms. The number of deaths among males is greater than the number of females, and the number of deaths increases in the elderly population, as the number of deaths in the 80+ age group is higher than all other groups. For the total number of deaths, we observed a quite stable situation, with a slight increase in the number of deaths, not a bad indicator in view of the continuous increase in the population.

The forecast number of deaths by gender, age, hospital, governorate and cause of death classified according to the ICD10 classification could not be shown in one table, therefore we are working on implementing a web application to show all the forecast results. In the web application we can find the results of forecasting in any specific category.

**Table 6** Forecast number of deaths by hospital

| Hospital                                      | 2021                    |          |          |            |          | 2022     |          |            |            |      | Total |
|---|-------------------------|----------|----------|------------|----------|----------|----------|------------|------------|------|-------|
|   | 1/1-31/3                | 1/4-30/6 | 1/7-30/9 | 1/10-31/12 | 1/1-31/3 | 1/4-30/6 | 1/7-30/9 | 1/10-31/12 | 1/10-31/12 |      |       |
|   | Azadi Teaching Hospital | 413      | 297      | 260        | 271      | 427      | 302      | 260        | 265        | 2496 |       |
| Zakho emergency hospital                      | 207                     | 200      | 212      | 229        | 223      | 216      | 229      | 247        | 1763       |      |       |
| Shar hospital/Suli                            | 206                     | 167      | 191      | 231        | 227      | 184      | 210      | 254        | 1670       |      |       |
| Duhok forensic medicine                       | 157                     | 152      | 174      | 152        | 158      | 152      | 174      | 151        | 1270       |      |       |
| Heevi Pediatrics Hospital                     | 128                     | 97       | 103      | 101        | 132      | 99       | 106      | 101        | 869        |      |       |
| Maternity hospital—Duhok                      | 93                      | 90       | 84       | 81         | 100      | 96       | 90       | 86         | 719        |      |       |
| Hiwa hospital/Suli                            | 83                      | 83       | 87       | 94         | 88       | 88       | 92       | 100        | 716        |      |       |
| Emergency Hospital—Duhok                      | 50                      | 42       | 64       | 94         | 51       | 43       | 67       | 99         | 510        |      |       |
| Gullan hospital—Aqrah                         | 50                      | 42       | 40       | 48         | 53       | 45       | 42       | 51         | 371        |      |       |
| Special Cardiac hospital/Suli                 | 46                      | 41       | 39       | 46         | 44       | 39       | 38       | 44         | 337        |      |       |
| Aqrah hospital                                | 68                      | 29       | 28       | 31         | 70       | 29       | 27       | 30         | 312        |      |       |
| Shorsh hospital/Suli                          | 44                      | 33       | 25       | 33         | 48       | 35       | 27       | 35         | 280        |      |       |
| Dr.Hemn hospital/Suli                         | 57                      | 25       | 34       | 37         | 49       | 16       | 26       | 27         | 272        |      |       |
| General surgical teaching hospital\Suli       | 56                      | 30       | 24       | 9          | 53       | 26       | 21       | 0          | 219        |      |       |
| Maternity hospital—Zakho                      | 18                      | 18       | 19       | 18         | 19       | 19       | 20       | 19         | 150        |      |       |
| Faruq medical city/Suli                       | 13                      | 16       | 18       | 18         | 14       | 18       | 20       | 20         | 137        |      |       |
| Aqrah forensic medicine                       | 14                      | 9        | 18       | 11         | 15       | 9        | 18       | 10         | 104        |      |       |
| Dr.Jamal ahmad rashid pediatric hospital\Suli | 17                      | 7        | 11       | 13         | 17       | 6        | 11       | 12         | 94         |      |       |
| Zakho forensic medicine                       | 8                       | 13       | 9        | 12         | 8        | 15       | 10       | 13         | 88         |      |       |
| Zakho general hospital                        | 16                      | 8        | 14       | 20         | 10       | 4        | 8        | 5          | 85         |      |       |



**Table 7** Forecast number of deaths classified by residence of the deceased (Governorate) (1/1/2021–31/12/2022)

| Place of residence | Total forecast number of deaths |                    |                    |                      |                    |                    |                    |                      |
|--------------------|---------------------------------|--------------------|--------------------|----------------------|--------------------|--------------------|--------------------|----------------------|
|                    | 1/1/2021–31/3/2021              | 1/4/2021–30/6/2021 | 1/7/2021–30/9/2021 | 1/10/2021–31/12/2021 | 1/1/2022–31/3/2022 | 1/4/2022–30/6/2022 | 1/7/2022–30/9/2022 | 1/10/2022–31/12/2022 |
| –                  | 19                              | 19                 | 19                 | 19                   | 19                 | 19                 | 19                 | 19                   |
| Sulaimaneyah       | 1006                            | 1009               | 1012               | 1015                 | 1018               | 1021               | 1024               | 1027                 |
| Duhok              | 424                             | 425                | 426                | 428                  | 429                | 430                | 431                | 432                  |
| Halabja            | 101                             | 101                | 101                | 102                  | 102                | 102                | 103                | 103                  |
| Al-Anbar           | 31                              | 31                 | 31                 | 31                   | 31                 | 31                 | 31                 | 31                   |
| Bagdad             | 14                              | 14                 | 14                 | 14                   | 14                 | 14                 | 14                 | 14                   |
| Diyala             | 28                              | 28                 | 28                 | 28                   | 28                 | 28                 | 28                 | 28                   |
| Germany            | 12                              | 12                 | 12                 | 12                   | 13                 | 13                 | 13                 | 13                   |
| Russia             | 12                              | 12                 | 12                 | 12                   | 12                 | 12                 | 12                 | 12                   |
| Kirkuk             | 10                              | 10                 | 10                 | 10                   | 11                 | 11                 | 11                 | 11                   |
| Virginia           | 10                              | 10                 | 10                 | 10                   | 10                 | 10                 | 10                 | 10                   |
| Mithan             | 9                               | 9                  | 9                  | 9                    | 9                  | 9                  | 10                 | 10                   |

**Table 8** Total forecast number of deaths

| Quarter              | Forecast number of deaths |
|----------------------|---------------------------|
| 1/1/2021–31/3/2021   | 1674                      |
| 1/4/2021–30/6/2021   | 1680                      |
| 1/7/2021–30/9/2021   | 1685                      |
| 1/10/2021–31/12/2021 | 1690                      |
| 1/1/2022–31/3/2022   | 1694                      |
| 1/4/2022–30/6/2022   | 1699                      |
| 1/7/2022–30/9/2022   | 1705                      |
| 1/10/2022–31/12/2022 | 1711                      |

List of abbreviations

| Abbreviation | Meaning  |
|--------------|--|
| SDGs         | Sustainable development goals                          |
| LC           | Lee-carter   |
| RNNs         | Recurrent neural networks                              |
| RBBD         | Registration Bureaus of Births and Deaths              |
| KRG-HIS      | Kurdistan Region Governorate Health Information System |
| DHIS         | District Health Information Software                   |
| ICD          | International Classification of Disease                |
| WHO          | World Health Organization                              |

## List of abbreviations

| Abbreviation | Meaning                         |
|--------------|---------------------------------|
| SELU         | Scaled Exponential Linear units |
| MSE          | Mean Squared Error              |
| RMSprop      | Root Mean Square Propagation    |
| M            | Male                            |
| F            | Female                          |
| JSON         | JavaScript Object Notation      |
| RMSE         | Root Mean Squared Error         |
| AIC          | Akaike information criterion    |
| BIC          | Bayesian information criterion  |

When we started collecting data for this research, we discovered that most information was paper-based and incomplete. To overcome this problem, we recommend using the KRG-HIS program in all health facilities to collect more correct daily information about health situations. Since its use, the health information system has begun to provide high-quality data, which is essential for planning, policy implementation and monitoring of health outcomes and services. Therefore, if this system is applied in all hospitals and health centers in the whole Kurdistan Region, we can obtain more reliable data and provide timely statistics on the health situation of the region, and (in our case) obtain more accurate forecasts. KRG-HIS is a system that can fill in the gaps for obtaining reliable and timely information on recorded diseases and deaths among the population.

We recommend that statisticians in local health departments use modern forecasting and estimating concepts and technologies, and keep up with the latest developments concerning these methods, as they are of great importance for the health services provided to society.

**Funding** No funding was received for conducting this study.

**Availability of data and material** The datasets analyzed during the current study are available from KRG-HIS platform on reasonable request.

**Code availability** Python program.

## Declarations

**Conflicts of interest** There is no conflict of interest in this article.

**Ethics approval** The study has been granted ethics approval from the Scientific Committee of the Statistics and Informatics Department at the University of Sulaimani/ Iraq (reference number 5, on the date 26/10/2021).

**Consent to participate** Not applicable.


**Consent for publication** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. UNDP: Sustainable development goals [Internet]. [https://www.undp.org/content/dam/undp/library/corporate/brochure/SDGs\\_Booklet\\_Web\\_En.pdf](https://www.undp.org/content/dam/undp/library/corporate/brochure/SDGs_Booklet_Web_En.pdf) (2015)
2. Jaha, A.A., Emasallati, A.: Using neural networks to predict time series using backpropagation training. ResearchGate (2019)
3. Hamoodat, A.A.: Comparison between exponential Smoothing model and Intervention method on international prices of barley. *Tikrit J. Pure Sci.* **18**(1), 255–262 (2013)
4. Norvig, S.J.R., Zhang, P.: *Artificial Intelligence A Modern Approach*. Pearson Education Inc, London (2010)
5. Raju, M.M., Srivastava, R.K., Bisht, D.C.S., et al.: Development of artificial neural-network-based models for the simulation of spring discharge. *Adv. Artif. Intell.* **2011**, 1–11 (2011)
6. He, L., Huang, F., Shi, J., et al.: Mortality forecasting using factor models: time-varying or time-invariant factor loadings? *Insur. Math. Econ.* **98**, 14–34 (2021)
7. Marino, M., Levantesi, S., Nigri, A.: Deepening Lee–Carter for longevity projections with uncertainty estimation. <http://arxiv.org/abs/2103.10535> (2021)
8. Hainaut, D.: A neural-network analyzer for mortality forecast. *ASTIN Bull.* **48**, 481–508 (2018)
9. Mikkelsen, L., Lopez, A., Phillips, D.: Why birth and death registration really are “vital” statistics for development [Internet]. Human Development Reports. <http://hdr.undp.org/en/content/why-birth-and-death-registration-really-are-“vital”-statistics-development> (2015)
10. Ministry of Health, MOHE, MOP/CSO, MOD W: Iraqi health information system review and assessment [Internet]. [https://applications.emro.who.int/dsaf/libcat/EMROPD\\_110.pdf?ua=1](https://applications.emro.who.int/dsaf/libcat/EMROPD_110.pdf?ua=1) (2011)
11. Salih, S.O., Moramarco, S., Di, G.D., et al.: Ten-year mortality trends and natural causes of death in the Iraqi Kurdistan. *Open Public Health J.* **14**, 3–11 (2021)
12. Iraq Government: First national voluntary review on sustainable development goals 2019 (2019)
13. Iraq infant mortality 1950–2020 [Internet]: Macrotrends. <https://www.macrotrends.net/countries/IRQ/iraq/infant-mortality-rate> (2020)
14. DHIS2, global collaboration led by UiO [Internet]: <https://dhis2.org/about/> (2020)
15. Emberti Gialloreti, L., Basa, F.B., Moramarco, S., et al.: Supporting Iraqi Kurdistan health authorities in post-conflict recovery: the development of a health monitoring system. *Front. Public Health* **8**, 1–13 (2020)
16. Tan, C.: Financial time series forecasting using improved wavelet neural network master's thesis. In: Exchange organizational behavior teaching journal [Internet]. [http://www.daimi.au.dk/~cstorm/students/Chong\\_Jul2009.pdf](http://www.daimi.au.dk/~cstorm/students/Chong_Jul2009.pdf) (2009)
17. Rojas, R.: *Neural Networks A Systematic Introduction*. Springer, Berlin (1996)
18. Krittanawong, C., Johnson, K.W., Rosenson, R.S., et al.: Deep learning for cardiovascularmedicine: a practical primer. *Eur. Heart J.* **40**, 2058–2069C (2019)
19. Nguyen, A., Pham, K., Ngo, D., et al.: An analysis of state-of-the-art activation functions for supervised deep neural network. In: Proceedings of 2021 international conference on system science and engineering, ICSSE 2021, pp. 215–220 (2021)
20. Mijwil, M.M.: Artificial neural networks advantages and disadvantages. LinkedIn [Internet], pp. 1–2. <https://www.researchgate.net/publication/323665827> (2018)
21. Central statistical organization Iraq: Annual statistical abstract [Internet]. <http://www.cosit.gov.iq/ar/2018-2019> (2019)
22. Central statistical organization of Iraq: Main indicators, population indicators. [Internet]. <http://cosit.gov.iq/ar/62-indicators?start=20> (2015)

## Authors and Affiliations

**Shahla O. Salih<sup>1</sup>**  · **Antonia Bezenchek<sup>2,3</sup>** · **Stefania Moramarco<sup>4</sup>** ·  
**Manuel De Iuliis<sup>5</sup>** · **Denis Stanev<sup>6</sup>** · **Iuri Fanti<sup>2</sup>** · **Dario Corsini<sup>2</sup>** · **Faiq B. Basa<sup>7</sup>** ·  
**Leonardo Emberti Gialloreti<sup>4</sup>**

<sup>1</sup> Department of Statistics and Informatics, University of Sulaimani, Sulaimani, Iraq

<sup>2</sup> Informa-PRO, Rome, Italy

<sup>3</sup> EuResist Network, Rome, Italy

<sup>4</sup> Department of Biomedicine and Prevention, University of Rome Tor Vergata, Rome, Italy

<sup>5</sup> Medyx, Rome, Italy

<sup>6</sup> Department of Physics, Sapienza University of Rome, Rome, Italy

<sup>7</sup> Rizgary Teaching Hospital, Erbil, Iraq