Research Article Open Access

Journal of CardioVascular Insights

DOI: doi.org/10.63721/25 /JCVI0101

Prediction of Hospital Length of Stay Using Artificial Neural Networks

Farnaz Shojaeifar[1*], Morteza Mahmoudzadeh[2] and Shiva Shojaeifar[3]

[1]West Azerbaijan Science and Technology Park- Iran
[2]Islamic Azad University of Tabriz, Iran
[3]Urmia University, Iran

Citation: Farnaz Shojaeifar, Morteza Mahmoudzadeh,Shiva Shojaeifar (2025) Prediction of Hospital Length of Stay Using Artificial Neural Networks. J.of Card Vas Insights 1(1), 01-15. WMJ/JCVI-101

Abstract

With the increasing demands on hospitals and the financial constraints they face, the management of hospital beds has become a complex issue. This study proposes and designs a decision support system for predicting the length of stay (LOS) of specific patients using an artificial neural network (ANN) data mining technique. This system can serve as an effective tool for measuring hospital resource utilization. The target population of this study consists of patients diagnosed with acute myocardial infarction at Seyed al-Shohada Hospital in Urmia, Iran, over a four-year period. A total of 997 records, comprising 32,934 fields, were extracted from medical records and analyzed using MATLAB 2013. After training and testing the network with various hidden layers and learning rates, a three-layer neural network with 10 neurons was constructed. The network was trained using the optimized TRAINBR function, resulting in an acceptable mean error of 5.1 and a correlation coefficient of 0.83, leading to the development of an optimal model. Following successful testing, the findings indicated that the artificial neural network has a strong capability for predicting patient length of stay. It is worth noting that the randomness of the data used for training the network enhances the model's accuracy and robustness. Such a model enables more efficient and effective utilization of human resources and hospital facilities.

***Corresponding author:** Farnaz Shojaeifar, West Azerbaijan Science and Technology Park- Iran.

Submitted: 27.02.2025 Accepted: 05.03.2025 Published: 19.03.2025

Keywords: Length of Stay, Data Mining, Decision Support System, Artificial Neural Network, Myocardial Infarction

Introduction

In recent years, the increasing demands on hospitals, coupled with financial constraints, have made hospital bed management a complex challenge. A model for predicting the length of stay (LOS) for hospitalized patients can serve as an effective tool for measuring hospital resource utilization. Such a model enables early interventions to prevent complications associated with prolonged hospital stays and facilitates more efficient and effective use of human resources and hospital facilities. This, in turn, can significantly enhance hospital resource planning and management. A model that aids in predicting a patient's LOS can be a valuable tool for healthcare providers to (a) plan preventive measures, (b) improve healthcare services, and (c) manage hospital resources more effectively.

Hospital productivity can be significantly compromised in two scenarios: first, if a hospital fails to meet short-term resource demands, such as staffing and equipment needs, and second, if a hospital is over-equipped relative to demand. Both scenarios arise due to significant fluctuations in hospital occupancy, which severely hinder efficient resource allocation and management. By accurately estimating how long patients will stay in the hospital, healthcare facilities can better plan bed management and optimize resource utilization [1, 2].

The length of stay following acute myocardial infarction (AMI) is also considered an indicator of hospital service quality. This metric has improved over time, with the average LOS decreasing from approximately 12 days in 1986–1988 (based on 896 patients) to about 5 days in 2011 (based on 1,330 patients). Studies have shown that older patients, females, and those who experienced post-AMI complications were more likely to have a longer LOS. Additionally, patients with an LOS exceeding 14 days after AMI were nearly twice as likely to experience higher mortality rates within one to three months after discharge compared to those hospitalized for 6–8 days [3, 4].
Therefore, a model capable of predicting the length of stay (LOS) for AMI patients can be highly beneficial for appropriate treatment planning and improving the quality of healthcare services. Data mining methods are now widely used in the healthcare domain for applications such as diagnosis [5] and patient management [6, 7]. Moreover, data mining algorithms have been successfully applied to predict LOS, as demonstrated in studies such as [8, 9].

Data mining refers to the process of discovering previously unknown and useful patterns within a database [10]. It is increasingly being utilized in health and medical fields, such as disease prediction and patient treatment [11]. Extracting relationships, rules, and essential information from data can be challenging due to the nature of the data, the size of databases, and other characteristics. Research on artificial neural networks (ANNs) has demonstrated their powerful capabilities in classification and pattern recognition. Inspired by biological systems, particularly studies of the human brain, artificial neural networks are capable of learning and generalizing from experience. Currently, ANNs are employed for a wide range of tasks across various domains, including business, industry, and science [12].

One of the primary applications of artificial neural networks is prediction, offering an attractive alternative tool for researchers in forecasting. Several distinguishing features make ANNs valuable and appealing for prediction tasks:

- Unlike traditional model-based methods, ANNs are data-driven and self-adaptive. They learn from examples and capture precise functional relationships within the data, even if the underlying relationships are unknown or difficult to describe. Thus, ANNs are well-suited for problems where solutions require knowledge that is hard to define but sufficient data or observations are available [13].
- ANNs are used to perform multivariate analysis, identifying both linear and nonlinear patterns among data variables [14].

Due to their strong performance in prediction, artificial neural networks have become one of the most popular methods in various medical fields, enabling appropriate decision-making [15].
The present study aims to utilize one of the data mining techniques to extract useful knowledge from data related to patients with AMI (Acute Myocardial Infarction) and propose and design a decision support system (DSS) for predicting the Length of Stay (LOS) of these patients. Ultimately, the results of this research can provide effective strategies for better management of treatment for these patients and more efficient utilization of hospital resources. The technique used in this study is the Artificial Neural Network (ANN) algorithm. The target variable is LOS, and a number of input variables are used for prediction.

Acute Myocardial Infarction, a consequence of industrial life, is one of the leading causes of mortality in human societies. In Iran, according to the Ministry of Health and Medical Education, cardiovascular diseases are the primary cause of death. Therefore, the study population in this research consists of patients suffer-

suffering from this prevalent disease. The main objective of this study is to design and implement a decision support system to determine and predict the Length of Stay (LOS) of patients with heart attacks using the Artificial Neural Network technique.

The operational Steps of the Research Include

- Extracting meaningful patterns and relationships between independent and dependent variables (related to LOS) from large datasets of patient records.
- Designing a Decision Support System (DSS) based on Artificial Neural Networks to predict the Length of Stay of heart attack patients.
- Training the system using available data.

All Data Potentially Influencing the Length of Hospital Stay (LOS) in Patients with Heart Attacks were Collected from Patient Records and Utilized in this Study. These Data include:

Demographic Data

- Age
- Gender
- Weight
- Marital status

Laboratory Data

- Cardiac enzymes
- Serum creatinine
- Fasting blood sugar
- Cholesterol levels
- Hemoglobin
- High-density lipoprotein (HDL)
- Low-density lipoprotein (LDL)
- Triglycerides

Medical History

- History of diabetes
- History of hypertension
- Family history of coronary artery disease and hypertension
- History of cardiovascular diseases
- Smoking status
- History of chest pain

Paraclinical Examination Results

- Exercise stress test
- Stress echocardiography
- Nuclear scan
- Electrocardiogram (ECG) changes

Length of Stay (LOS)

- The duration of hospitalization for each patient.

Theoretical Foundations and Research Background

The Concept of Data Mining

In general, data mining refers to the process of analyzing data from various perspectives and summarizing it into useful information. This information can be utilized to increase revenue, reduce costs, or achieve both. Technically, data mining is the process of discovering correlations or patterns within large relational databases across diverse domains [16].

The Data Mining Process

The data mining process consists of three main stages: **Preliminary Exploration, Model Building or Pattern Identification with Validation/Verification, and Deployment.**

- Exploration (Data Preparation)

This stage typically begins with data preparation, which may include data cleaning, data transformation, and selecting subsets of records from a vast number of variables (fields). Depending on the nature of the analytical problem, this stage may require simple predictive models, statistical models, or graphical tools to identify relevant variables and determine the complexity of models for use in the next stage.

- **Model Building and Validation**

This stage involves examining various models and selecting the best one based on its predictive performance. Numerous techniques have been developed to achieve this goal, known as "competitive model evaluation." In this approach, different models are applied to the same dataset, and their performance is compared. The model with the best performance is then selected.

- **Deployment:**

The final stage involves applying the selected model from the previous stage to new data to generate predictions or expected outputs.

A key Distinction Between Data Mining and Statistical Principles Lies in Their Focus.

Data mining is more application-oriented rather than focusing on the underlying nature of phenomena. In other words, data mining is less concerned with identifying relationships between variables [17].

Data Mining Techniques

Data mining employs various techniques such as decision trees, artificial neural networks, association rules, and genetic algorithms to discover patterns [18]. Below, we provide an overview of the artificial neural network (ANN) technique.

Artificial Neural Networks (ANNs)

Artificial neural networks (ANNs), often simply referred to as "neural networks," are analytical techniques modeled after the learning processes of cognitive systems and the neural functions of the brain. These networks are capable of predicting and inferring new concepts (based on specific variables) from other concepts (based on the same or different variables). This process is known as **learning from existing data.**

ANNs are mathematical models inspired by biological systems. They are algorithms designed for optimization and free-form learning, drawing inspiration from research on the nature of the brain. The brain organizes its structural components using units known as neurons, enabling it to perform certain computations much faster than digital computers. In general, a neural network is a machine designed to model the way the brain performs specific tasks or remarkable functions, based on studies by Dr. Simon Haykin.

A neural network is a large parallel distributed processor composed of simple processing units. It has a natural tendency to store experiential knowledge and make it accessible for use [19].

Structure of Artificial Neural Network

Input Layer: This layer receives inputs and sends the input signal to the next layer based on its connection strength with that layer. Hidden Layer: The number of hidden layers and the number of neurons in them can be chosen freely. Hidden layers must be selected carefully to provide the appropriate output. Output Layer: Another group of neurons creates the external world through their outputs. Neuron: A neuron is the smallest unit of an artificial neural network that constitutes the functioning of neural networks. Figure 1 illustrates the structure of an artificial neural network.

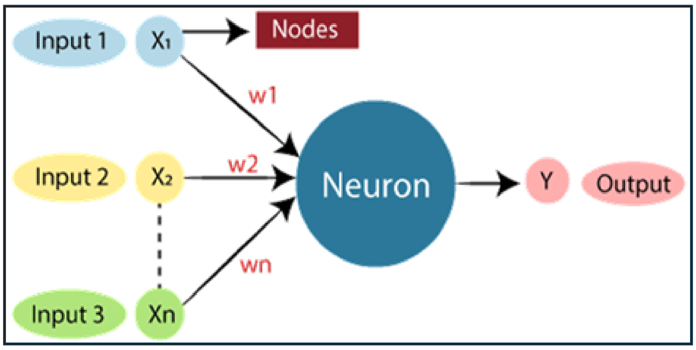


Figure 1: Structure of the Artificial Neural Network.

Methods of Knowledge Discovery in Data Mining

Data mining utilizes analytical model building, classification, and prediction of information, providing results through relevant tools. For a data mining algorithm to effectively perform knowledge extraction, a series of preprocessing steps on the data and a series of post-processing steps on the extracted patterns are required. Some techniques used in data mining include: 1. Classification: (Predictive technique) In this method, a sample is categorized into one of several predefined classes. 2. Regression: (Predictive technique) This involves predicting a value of a variable based on other variables. 3. Clustering: (Descriptive technique) This maps a data set to one of several clusters. Clusters are groupings of data sets formed based on the similarity of certain criteria. 4. Association Rule Mining: (Descriptive technique) This expresses dependency relationships between different attributes. 5. Sequence Analysis: This models sequential patterns such as time series [20]. To implement these techniques, there are many algorithms available, among which K-Means, CART, Apriori, Naive Bayes, ANN, Page Rank, C4.5, EM, SVM, and AdaBoost are the most commonly used.

Applications of Data Mining in the Medical Field

In this section, we aim to explore various domains of medical sciences that have been evaluated by data mining algorithms and have successfully achieved desirable outcomes. Additionally, the algorithms employed in each of these fields are discussed.

Identifying Factors Influencing the Onset of Various Cancers and Determining Optimal Treatment Methods

Nasrin Asadi and colleagues, in an article titled "Using Data Mining to Determine Factors Influencing the Onset of Various Cancers and Optimal Treatment Methods in the Cancer Database of Namazi Hospital" [21], sought to examine different types of cancers through data mining techniques and methods to discover dependencies that help identify factors influencing cancer. They then evaluated various treatments using data mining algorithms to propose the best treatment for each type of cancer. To achieve this, the k-means algorithm was used to cluster factors affecting the disease, such as gender, age, family history of the disease, and smoking or alcohol consumption. Additionally, the Apriori algorithm was employed to determine the best treatment among three common methods: chemotherapy, radiotherapy, and surgery.

Prediction and Treatment of Emergency Patients

The emergency department of a hospital is one of its most critical units, as patients in this section often require intensive care and rapid decision-making for treatment. Lin et al. [22], in a study involving patients recorded in the emergency department database of a Taiwanese hospital, aimed to discover the best decisions for treatment, thereby avoiding unnecessary costs resulting from incorrect actions such as surgeries. To accomplish this, they used clustering methods to categorize disease characteristics and symptoms, as well as decisions made by nurses and physicians. They employed the ROSE2 (Rough Sets Data Explorer) software for data mining. For this purpose, a combination of the k-means algorithm, the Self-Organizing Map (SOM) algorithm, and artificial neural networks was used to perform clustering.

Identification and Prediction of Heart Attacks

This study was conducted on 313 data points, and data mining clustering methods were used to identify and predict heart attacks. Clustering is one of the primary tasks in data mining, aiming to group data into meaningful classes (clusters) such that the similarity within a cluster is maximized, and the similarity between different clusters is minimized. In this research, given the presence of cardiac patient data, a combination of genetic algorithms and k-means was used, yielding promising results for better cluster identification and, consequently, the diagnosis and prediction of heart attacks [23].

Knowledge Discovery in Clinical Data Warehouses of Traditional Chinese Medicine

In recent years, Traditional Chinese Medicine (TCM) has been approved as a complementary treatment for many diseases, such as cancer, rheumatoid arthritis, leukemia, and migraines [24]. Since the primary source of empirical knowledge in this field is long-term clinical evidence, including laboratory results, diagnoses, and prescriptions, it encompasses a vast data warehouse that only data mining methods can effectively explore. In a study conducted on 21,111 inpatients and 21,111 outpatients of TCM, clustering methods were employed. Among five clustering methods, such as SVM, neural networks, and decision trees, SVM showed better predictive results for different syndromes in 1,110 clinical epidemiology cases. Additionally, association rules and Complex Network Analysis (CAN) were used to identify useful points for acupuncture and patterns of herbal combinations. Multi-class SVM machine learning techniques were applied to predict the onset of nephropathy, and neural networks were utilized to forecast the role of diagnostic information in the treatment of rheumatoid arthritis.

Predicting the Length of Hospital Stay

Predicting the length of hospital stays enables better management of hospital resources, improved service delivery, and increased patient satisfaction. In a study conducted on 2,113,101 records from a database of children (under 12 years) admitted to 3,111 hospitals in 2011, data mining algorithms were analyzed to build a model for predicting the length of stay for gastrointestinal patients requiring short-term care [25]. To reduce the scale of the large dataset, random samples from the record sets were used, and four strategies were adopted to estimate missing values. Outliers in the features were identified, and an anomaly detection algorithm, based on clustering, was used to detect outlier records. Since many features were irrelevant to the research objectives, 41 out of 112 features that had the most significant impact on the length of stay for gastrointestinal patients were identified, and final model was constructed using these features. Algorithms such as decision trees, neural networks, and Bayesian-based algorithms were used to build the model using training data. The results of the model's accuracy on the test data showed that the C5.0 algorithm had higher accuracy compared to other algorithms. Using this algorithm, the length of stay for gastrointestinal patients requiring short-term care could be predicted with an accuracy of 32%.

Row	Medical Research Domain	Data Mining Method	Data Mining Algorithm	Description
1	Disease Prediction	Clustering, Classification, Association Rules	k-means, Apriori	Identifying factors influencing the development of various types of cancer.
2	Determining Optimal Treatment	Clustering, Classification, Association Rules	k-means, Apriori	Determining the best treatment for cancer among three methods: surgery, chemotherapy, radiotherapy.
3	Analysis of Laboratory Data	Clustering	SAMBA	Studying gene behavior with DNA strands to predict genetic disorders and fatal diseases.
4	Prediction in Emergency Patients	Clustering	Combination of k-means and Self-Organizing Map (SOM)	Making timely and accurate treatment decisions to reduce hospital costs.
5	Predicting Length of Hospital Stay	Clustering, Classification	Decision Tree, Neural Networks, Naïve Bayes, C4.5	Predicting the length of stay for gastrointestinal patients requiring short-term care to reduce hospital costs.
6	Identification and Prediction of Disease Complications	Clustering	Combination of Genetic Algorithm and k-means	Identifying and predicting heart attacks.
7	Traditional Chinese Medicine	Clustering	Support Vector Machine (SVM), Decision Tree, Bayesian Network	Discovering different syndromes.
8	Traditional Chinese Medicine	Association Rules	Complex Network Analysis (CAN)	Identifying optimal acupuncture points and patterns of herbal
9	Traditional Chinese Medicine	Support Vector Machine (SVM)	Clustering	Predicting the onset of diabetic nephropathy.
10	Traditional Chinese Medicine	Clustering	Neural Networks	Treating rheumatoid arthritis.

Methodology

This study is descriptive in nature and employs data mining techniques. It examines all adult patients admitted to the specialized Seyed al-Shohada Hospital in Urmia over a 4-year period due to acute myocardial infarction (AMI). The data related to these patients were recorded in the hospital's archival records. Cases with incomplete information were excluded, and those with duplicate, incomplete, or ambiguous data were also removed from the study.

Data Collection Tools

The data collection tools in this study included the documents available in the medical records of the admitted patients. A medical record is a documented collection of facts about a patient's health status, including medical history, physical examinations, tests and investigations, diagnoses, treatment plans, and outcomes of the implemented treatments, as well as admission and discharge summaries. To collect the required information for this study, the following documents were obtained: discharge summaries, medical history, disease progression, physician orders, nursing reports, vital signs charts, vital signs monitoring, and laboratory reports.

Variables Examined

The Variables Examined for Each Patient in this Study Included

- **Demographic Data:** Weight, age, gender, marital status.
- **Clinical Data:** Pulse, blood group, hemoglobin level, white blood cell count, platelet count, fasting blood sugar, cholesterol, triglycerides, low-density lipoprotein (LDL), high-density lipoprotein (HDL), creatinine, blood urea nitrogen (BUN), creatine phosphokinase (CPK), prothrombin time (PT), activated partial thromboplastin time (aPTT), troponin.
- **Medical History:** Family history of heart attack, previous heart attack, diabetes, hypertension, smoking history, medication history.
- **Clinical Presentation:** Time elapsed from symptom onset, high blood pressure, low blood pressure, use of streptokinase, length of hospital stay, and electrocardiogram (ECG) changes.

Data Preparation

The collected data from medical records were stored in Excel format. All stages of data preparation, including cleaning, removing duplicates, correcting typographical errors, and integrating data in terms of data scale, standardizing fields, and sorting records, were performed. Missing data were handled using the mean value of the available data, and new fields were created to prepare the data for building a dedicated data warehouse for acute myocardial infarction patients.

The first step was the creation of a data warehouse, where data from 4 consecutive years of patient records were extracted and entered as records into Excel. For data analysis and processing, 997 records, comprising 32,934 fields, were initially analyzed using descriptive statistics (frequency, percentage, and mean) through MATLAB 2013. Subsequently, artificial neural network techniques were applied to identify patterns, models, and potential rules governing the observed patterns.

Predictive Modeling

One of the critical factors in the complexity of predictive models generated by machine learning algorithms is the number of predictor variables. Some researchers reduce the number of predictor variables to avoid model complexity, using only the most important variables. However, since predictor variables vary in type and each plays a different role in predicting outcomes, it is preferable to use all of them in generating predictive models. In this study, 35 quantitative and qualitative variables were coded based on their values and examined.

Data Analysis and Findings

Description of Variables

The total number of initial patient records admitted to Seyed al-Shohada Hospital in Urmia due to heart attack was 1,100. A total of 39 fields were completed as patient characteristics for these 1,100 individuals. After the data cleaning process, 997 records across 35 fields were finalized as the dataset for analysis. A portion of this data is presented in Table 2.

Table 2: Values Corresponding to Patient Medical History Parameters

Parameter	Values
Age	(18-35): 1, (36-50): 2, (51-80): 3, (>80): 0
Gender	Male: 1, Female: 0
History of Diabetes	Positive: 1, Negative: 0
Family History of MI	Positive: 1, Negative: 0
Previous History of MI	Positive: 1, Negative: 0
History of Smoking	Positive: 1, Negative: 0
History of Hypertension	Positive: 1, Negative: 0
Length of Hospital Stay	(0-5 days): 2, (6-9 days): 3, (>10 days): 1

Visualization Results

Some statistical analyses of the data revealed that the patients admitted were aged between 19 and 100 years, with an average age of 60 years. Figure 2 illustrates the age distribution of the patients. Their average weight was 70 kg. According to the findings, the average heart rate of the patients was 73, and the average time from symptom onset to hospital admission was 11 days. Additionally, 73% of the patients were male, while the remaining were female (Figure 3).

The most common blood type was A+ (39%), followed by AB- (27%) and B+ (16%). The blood type AB- had the lowest frequency among the recorded data (Figure 4).

Furthermore, the results indicated that 45.3% of the studied patients had a history of smoking. A history of hypertension and diabetes was reported in 39% and 21% of the patients, respectively. 35% of the patients had a previous history of heart attack, and 12% had a family history of heart attack in their first-degree relatives (Figure 5). Additionally, the results showed that 47% of the patients with a history of heart attack had a history of medication use.

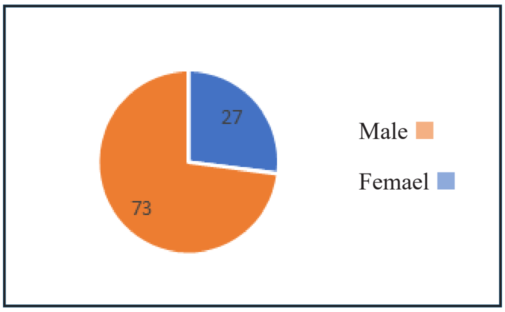


Figure 3: Percentage Distribution of Patients' Gender

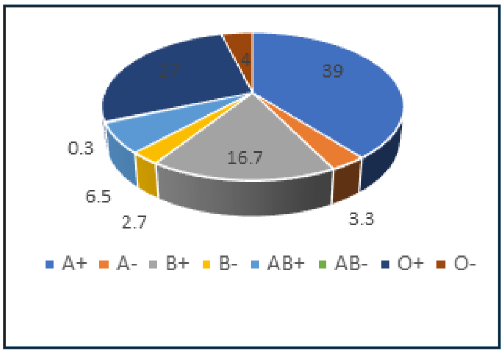


Figure 4: Percentage Distribution of Patients' Blood Groups.

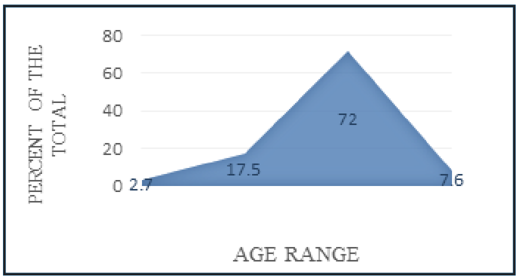


Figure 2: Percentage Distribution of Patients' Age Group.

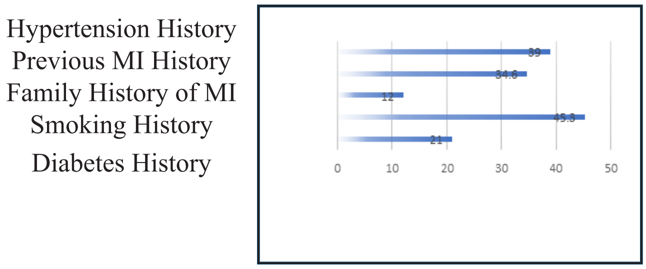


Figure 5: Percentage Distribution of Patients' Medical History Status.

Analysis of Hospital Stay Duration

Based on the coding scheme provided in Table 2 for the length of hospital stay, the distribution of patients across the classes is as follows:

- Class 1 (0-5 days): More than 57% of patients.
- Class 2 (6-9 days): 24% of patients.
- Class 3 (>10 days): The remaining patients.

This indicates that **more than 57% of patients** were discharged from the hospital within **1 to 5 days** of admission **(Figure 6).**

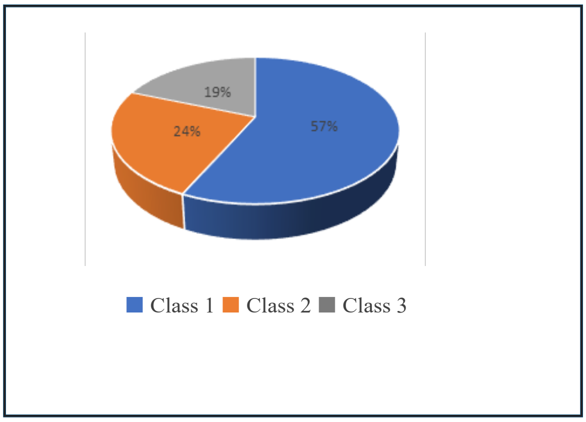


Figure 6: Percentage Distribution of Patient Length of Stay (LOS).

Findings on the Role of Patient History in Hospital Stay Duration

In another part of the research findings, the role of patient history—including smoking, hypertension, diabetes, previous MI (myocardial infarction), and family history of MI—in the length of hospital stay was examined. The results are as follows:

Smoking History

- All patients without a history of smoking were discharged within 1 to 5 days.
- 94% of patients with a history of smoking were hospitalized for more than 6 days.

Hypertension History

- 93% of patients without a history of hypertension were hospitalized for less than 5 days.
- All patients with a history of hypertension were hospitalized for more than 6 days.

Diabetes History

- All patients with a history of diabetes were hospitalized for more than 6 days.

Previous MI History

- 87% of patients with a history of previous MI were hospitalized for 1 to 5 days.
- All patients without a history of previous MI were hospitalized for more than 6 days.

Family History of MI

- 97.5% of patients with a family history of MI were hospitalized for more than 10 days.

The related results are summarized in Table 3.

Table 3: Distribution of Patients Across Hospital Stay Classes Based on Medical History

Variable	Value	Class 1		Class 2		Class 3	
		Count	%	Count	%	Count	%
Smoking History	0	545	100%	0	0%	0	0%
	1	26	6%	302	67%	124	27%
Hypertension History	0	568	93%	38	7%	0	0%
	1	0	0%	265	68%	123	32%
Diabetes History	0	566	73%	214	27%	0	0%
	1	0	0%	88	42%	121	58%
Previous MI History	0	566	87%	85	13%	0	0%
	1	0	0%	217	62%	132	38%
Family History of MI	0	568	64.50%	302	35.50%	0	0%
	1	0	0%	3	2.50%	117	97.50%

The Process of Selecting an Appropriate Neural Network for Predicting Patient Length of Stay

All artificial neural networks (ANNs) are based on the concepts of neurons, connections, and transfer functions. The primary differences among them arise from various learning rules and how these rules modify the network's topology. Essentially, the applications of these networks include modeling, identification, classification, pattern recognition, optimization, control, industrial applications, communications, and signal processing [26]. The ability to "learn" the relationship between inputs and outputs is one of the most fundamental advantages of neural networks, making them highly attractive. This learning capability makes neural networks suitable for problems with unknown and nonlinear structures, such as pattern recognition, medical diagnosis, time series prediction, and other similar tasks [27].

In each step of creating a neural network, the desired parameters, along with the input and output datasets, are defined. The inputs are propagated through each layer of the network to produce the outputs. Then, the network undergoes training. During this process, the errors between the outputs and the correct responses are propagated back through the network, and the connection weights are individually adjusted to minimize these errors. After a significant number of training patterns have been propagated through the network multiple times, the mapping function is trained within an acceptable error margin. The goal of the training process is to adjust the connections between weights and bias nodes so that, given a set of input data, the expected output from the neural network can be achieved.

As mentioned earlier, during the training of the network, its properties and details must be defined before the training begins. The first step in estimating the network's properties is determining the number of processing layers in the network. As a minimum, the network must have two layers: an input layer and an output layer. Multilayer feedforward neural networks often include one or more hidden layers as well. The number of neurons in each layer and the training function are other parameters adjusted to improve the performance of the neural network in this study. The default fixed parameters and the defined properties and details of the neural networks created in this study are provided in Tables 4 and 5.

Table 4: Default Fixed Parameters in the Created Neural Networks

Network Type	Performance Function	Transfer Function
Feed-Forward Backpropagation	Mean Squared Error (MSE)	Sigmoid (TANSIG)

Table 5: Defined Properties and Details of the Created Neural Networks

Network Name	Training Function	Number of Layers	Number of Neurons in Each Layer
Network1	TRAINLM	2	10
Network2	TRAINR	2	10
Network3	TRAINLM	3	30
Network4	TRAINBR	3	30
Network5	TRAINBR	2	20
Network6	TRAINBR	3	10

Artificial Neural Network Performance

The statistical metrics used in this study to evaluate the models are the correlation coefficient (R) and the mean squared error (MSE). The results are assessed and presented based on the highest efficiency coefficient (R) and the lowest error (MSE). The results of analyzing the input data to the network using the training functions LM, BR, and R are provided in Table 6.

Table 6: Values of Statistical Metrics for Inputs and Training Functions

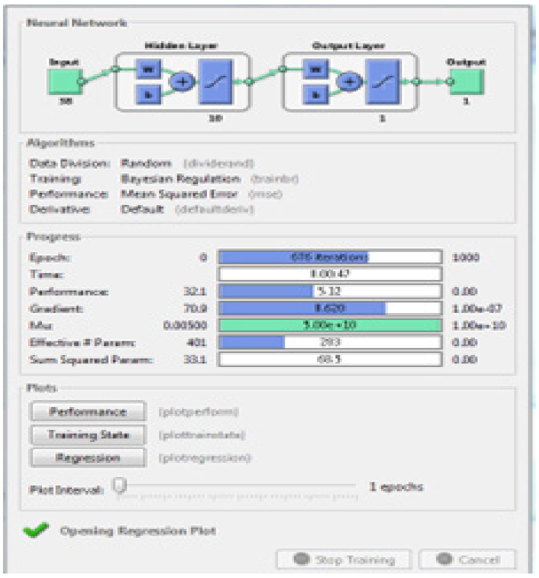
Network Name	Epoch	Correlation Coefficient (R)	Mean Squared Error (MSE)
Network1	8	0.30	14.70
Network2	8	0.32	18.90
Network3	8	0.50	19.54
Network4	295	0.29	15.20
Network5	1000	0.35	14.05
Network6	676	0.83	5.10

Explanation

- Epoch: The number of times the entire dataset is used to train the neural network.
- Correlation Coefficient (R): Indicates the correlation between predicted and actual values. A value closer to 1 indicates better performance.
- Mean Squared Error (MSE): A measure of prediction error. The lower this value, the higher the model's accuracy.

Based on the results in the table, Network6 with a correlation coefficient of 0.83 and a mean squared error of 5.10 demonstrates the best performance among the evaluated networks.

As evident from the table, the minimum error is 5.1, corresponding to the BR function, while the maximum error is 19.54, corresponding to the LM function. The simulated results are in very close agreement with the actual observations, indicating that the network performs exceptionally well. The characteristics and results related to Network 6 (the optimized network) are presented in Figures 7, 8, and 9, as well as in Table 7.



Network in MATLAB 2013

Figure 7: Structure of Network 6 in MATLAB 2013.

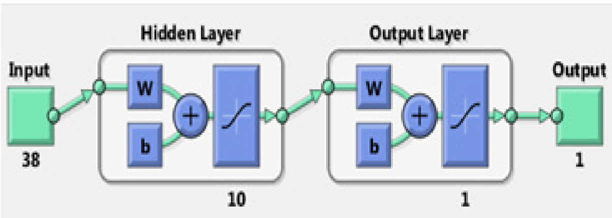


Figure 8: Training Window of the Optimized Network in MATLAB 2013.

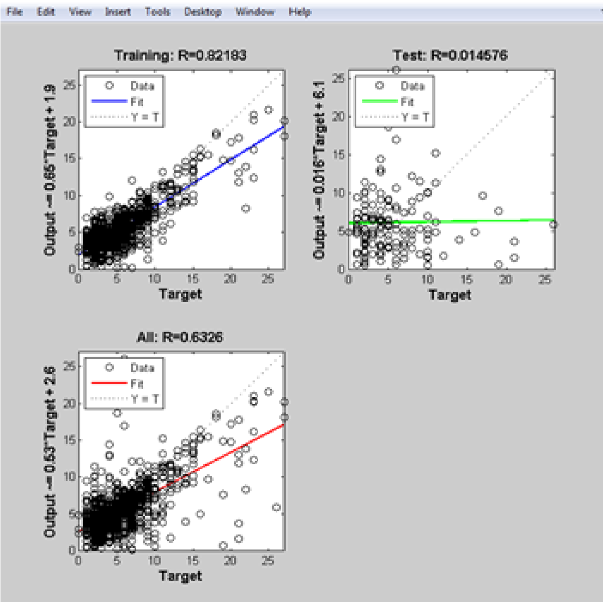


Figure 9: Regression Plot Window Between Target and Output Data.

As the results in Table 7 indicate, the BR learning algorithm was utilized as the best algorithm during the training phase. The sigmoid activation function in the perceptron neural network demonstrated the best performance in the hidden and output layers. Accordingly, the best performance was observed in Network 6, which achieved a correlation coefficient of 0.82 between observed and simulated data and, compared to other networks, had the lowest error with a mean squared error of 5.1. Thus, based on the conducted evaluations, Network 6 with the mentioned specifications is introduced as the optimal network for predicting patient hospitalization duration.

Table 7: Comparison of Target Data and Output from the Defined Model

Row	Input Value (Actual Value)	Network Output (Predicted Value)	Row	Input Value (Actual Value)	Network Output (Predicted Value)
1	17	15.492	14	7	7.864
2	6	6.637	15	7	6.845
3	9	8.830	16	10	9.962
4	6	6.728	17	7	5.303
5	3	2.481	18	3	3.061
6	2	1.763	19	9	9.571
7	2	1.465	20	3	2.908
8	5	5.052	21	6	6.268
9	3	2.606	22	13	10.83
10	5	4.878	23	5	4.247
11	1	1.362	24	4	5.016
12	5	5.087	25	6	5.543
13	9	7.990			

To further evaluate the accuracy of the obtained model, data related to 25 patients were selected using a numbering method and the RANDBETWEEN function in Excel, which generates random integers. These data were then tested in the model derived from this study, and the results are presented in Table 8. Additionally, Figures 10 and 11 respectively display the linear and scatter plots comparing the observed and simulated data for the 25 patient cases.

Table 8: Comparison of Observed and Simulated Data from the Proposed Model

Network Name	Network Type	Performance Function	Transfer Function	Training Function	Number of Layers	Number of Neurons
Network 6	Backpropagation	Mean Squared Error	Sigmoid	TRAINBR	3	10

Figure 10: Linear Plot Comparing Observed and Simulated Data for the LOS Index.

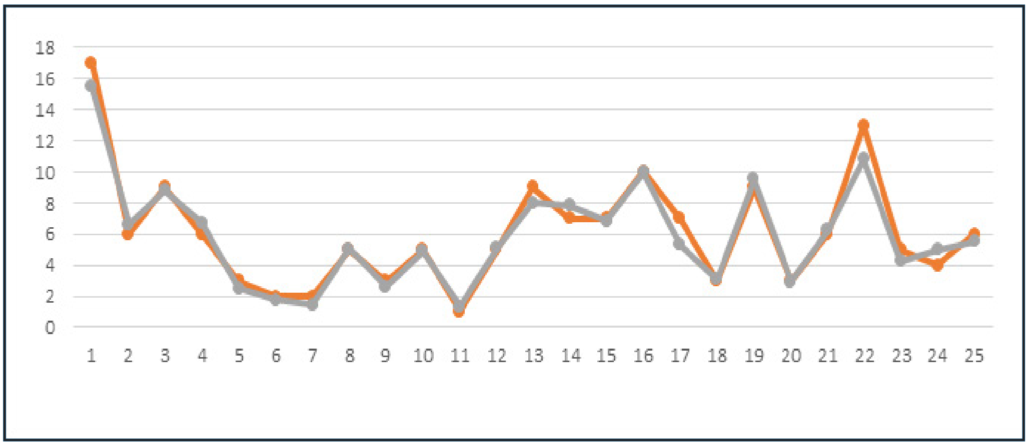
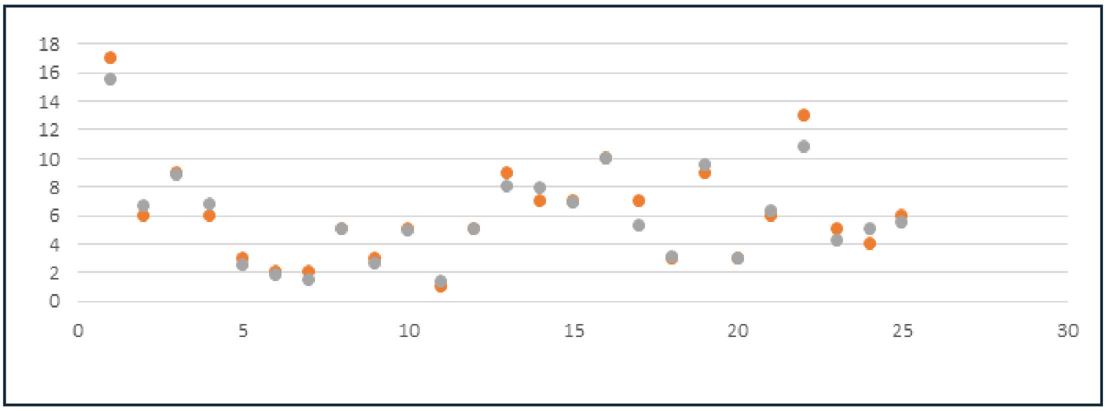


Figure 11: Dot Plot Comparing Observed and Simulated Data of the LOS Index.



Discussion and Conclusion

Artificial neural networks (ANNs) are capable of analyzing and simulating phenomena in nonlinear and uncertain systems where the relationships between components and system parameters are not well understood or easily describable. Given the importance of predicting patient hospitalization duration in hospital management, the results of this study demonstrate the successful application of intelligent models in predicting complex parameters. The high correlation coefficient and low error in the optimized network modeled in this research confirm the reliability of artificial neural networks.

One of the key evaluation metrics for artificial neural networks is the correlation coefficient, which indicates the degree of correlation between the model's predicted results and the actual observed data. In the proposed model of this study, the correlation coefficient is 0.82, suggesting that multilayer perceptron networks are a suitable and reliable model for predicting the Length of Stay (LOS) index. The best-performing artificial neural network model in this study is a 3-layer perceptron model (one output layer and two hidden layers) with 10 neurons.

The proposed neural network can assist hospital administrators in predicting patient hospitalization duration with high accuracy and without requiring excessive time. This enables them to leverage the benefits of such predictions in hospital management.

However, like all research studies, this study faced certain limitations. These included the lengthy process of obtaining permissions to access patient data, the lack of electronic records or a comprehensive database of patient information, inconsistencies in medical records, and the limited availability of equivalent research studies. To address these limitations, future research is recommended to identify and utilize new data fields in databases and explore methods such as neuro-fuzzy and neuro-fuzzy genetic approaches.

In conclusion, this study highlights the effectiveness of artificial neural networks, particularly multilayer perceptrons, in predicting patient hospitalization duration. The proposed model offers a reliable tool for hospital management, enabling more efficient resource allocation and improved decision-making processes. Future research should focus on overcoming the identified limitations and exploring advanced hybrid methods to further enhance prediction accuracy.

References

1. Liu P, Lei L, Yin J, Zhang W, Naijun W, et al. (2012) Healthcare data mining: Predicting hospital length of stay (PHLOS). International Journal of Knowledge Discovery in Bioinformatics 44-66.
2. Gustafson D (2002) Evaluation of quality improvement programmes. Quality and Safety in Health Care 11: 270-275.
3. Kononenko I, Bratko I, Kukar M (1997) Application of machine learning to medical diagnosis. In Machine Learning and Data Mining: Methods and Applications Wiley 389-408.
4. Harper P (2005) A review and comparison of classification algorithms for medical decision making. Health Policy 71: 315-331.
5. Ceglowski A, Churilov L, Wassertheil J (2005) Knowledge discovery through mining emergency

department data. Proceedings of the 38th Hawaii International Conference on System Sciences 1-9.
6. Harper P (2002) A framework for operational modelling of hospital resources. Health Care Management Science 5: 165-173.
7. Isken M, Rajagopalan B (2002) Data mining to support simulation modeling of patient flow in hospitals. Journal of Medical Systems 26: 179-197.
8. Vahidi R, Kushavar H, Khodayari R (2006) Factors affecting coronary artery patients' hospital length of stay in Tabriz Madani Hospital. Journal of Health Administration 9: 63-68.
9. Ahmadi M (2013) Use of data mining techniques to determine and predict length of stay of cardiac patients. Healthcare Informatics Research 19: 121-129.
10. Jilani TA, Yasin H, Yasin M, Ardil C (2009) Acute coronary syndrome prediction using data mining techniques. International Journal of Information and Mathematical Sciences 5: 295-299.
11. LiuP, Lei L, Yin J, Zhang W, Naijun W, et al. (2006) Healthcare data mining: Predicting inpatient length of stay. Proceedings of the 3rd International IEEE Conference on Intelligent Systems 832-837.
12. Widrow B, Rumelhart DE, Lehr MA (1994) Neural networks: Applications in industry, business, and science. Communications of the ACM 37: 93-105.
13. White H (1989) Learning in artificial neural networks. Statistical Science 4: 425-464.
14. Kudyba S, Gregorio T (2010) Identifying factors that impact patient length of stay metrics for healthcare providers with advanced analytics. Health Informatics Journal 16: 235-245.
15. Rani KU (2011) Analysis of heart diseases dataset using neural network approach. International Journal of Data Mining and Knowledge Management Process 1: 1-8.
16. Bakhtiari H (2011) Determining the pattern and distribution of mortality using data mining methods from 2004 to 2010 in Fars Province. Thesis Submitted for the Degree of MPH, Shiraz University of Medical Sciences.
17. Park K (2005) Park's Textbook of Preventive and Social Medicine. Banarsidas Bhanot Publishers https://milonm28.wordpress.com/wp-content/uploads/2017/08/parks-preventive-social-medicine-23rd-ed.pdf.
18. Chen H, Fuller SS, Friedman C, Hersh W (2005) Knowledge Management and Data Mining in Biomedicine. Springer 8.
19. Sharma IK (2012) Use of data mining and neural network in the medical industry. International Journal of Current Development in Artificial Intelligence 3: 1-8.
20. Salalri M, Adibnia F (2010) The top 10 algorithms for data mining. 13th Iranian Student Conference on Electronic Engineering
21. Asadi N, Sadrodini M (2010) Employing data mining to identify cancer risk factors and determine the optimal treatment in Namazi Hospital cancer database. 16th Annual National Conference of the Computer Society of Iran, Sharif University
22. Lin WT, Wu YC, Zheng JS, Chen MY, Zhang R (2011) Analysis by data mining in the emergency medicine triage database at a Taiwanese regional hospital. Expert Systems with Applications 38: 11078-11084.
23. Dehghani T, Afshari Saleh M, Khalilzadeh M (2011) A genetic K-means clustering algorithm for heart disease data. 5th Conference on Data Mining of Iran, Amirkabir University.
24. Zhou X, Chen S, Liu B, Zhang R, Wang Y, et al. (2010) Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support. Artificial Intelligence in Medicine 48: 139-152.
25. Oliyaei A, Salmasi N (2011) An efficient model to predict the duration of hospitalization of digestive system patients. 5th Conference on Data Mining of Iran, Amirkabir University.
26. Du KL, Swamy MNS (2006) Neural Networks in a Softcomputing Framework. Springer.
27. Gopalakrishnan K (2010) Effect of training algorithms on neural networks aided pavement diagnosis. International Journal of Engineering Science and Technology 2: 83-92.