Data Classification and Prediction of Type and Outcome of Treatment of Nosocomial Infection Patients: Using Neural Network Algorithm

Amin Farahbakhsh
Bu-Ali Sina University

Hamidreza Dezfoulian (Hrdezfoolian@basu.ac.ir)
Bu-Ali Sina University

Salman Khazaee
Hamadan University of Medical Sciences

Research Article

Keywords: Nosocomial Infection, Data Mining, Neural Network

Posted Date: August 2nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-753543/v1

Abstract

Background

In general, a predictive system for the diagnosis and prediction of nosocomial infections can help improve the quality of services in health centers and reduce the associated costs. Therefore, the present study was designed to predict the type and outcome of treatment of nosocomial infection patients using neural network algorithm.

Methods

The study is a cross-sectional one based on the registered data done on the nosocomial infection data of selected hospitals of Hamadan Province from March 2017 to March 2019, including 5,680 cases of nosocomial infections. According to patients' information registered in the Nosocomial Infection Registration System, nine criteria including the patients’ basic clinical information including name of the hospital and ward, infection-causing tool, organism, time from hospitalization to infection, length of hospital stay, age, gender, and weight were used as the explanatory criteria, and two criteria of type of infection and the final outcome of treatment, together with the actual available results, were used as the intended targets for performing the necessary analyses in data mining. In this study, neural network algorithm was implemented using IBM SPSS Modeler Software, version 18.0.

Results

The study results showed that length of hospital stay and hospitalization ward were the most influencing factors on the incidence of nosocomial infection. This network is multilayer and has two output layers and 11 hidden layers as well as 9 input layers. The results of the algorithm, with an accuracy of 91.8%, have had good adjustment with the actual results.

Conclusions

The results obtained from the model can help take an accurate step towards prediction of the likelihood of nosocomial infection and its outcome. It also identifies and prioritizes the factors influencing the incidence of infection.

Background

Infections acquired within 48 to 72 hours from admission to the hospital are usually considered as nosocomial infections, if the patient had not been in the incubation period [1, 2]. Nosocomial infection manifestations may occur during the hospitalization period or after the discharge [3]. According to the statistics, more than two million people worldwide suffer from nosocomial infections annually [3]. The

World Health Organization (WHO) estimates that at least 15% of hospitalized patients develop nosocomial infections [4]. The rate of nosocomial infections in developed industrialized countries is about 5–10% and in developing countries, it is 20–25% [5]. Studies done around the world have reported a prevalence of nosocomial infections ranging from 4 to 47% [8 – 5,3]. Nosocomial infections significantly increase hospital costs by imposing extra medicine use, especially antibiotics, and increasing the length of hospital stay [9]. Evidence shows that by employing effective methods in relation to the nosocomial infection care system, the incidence of nosocomial infection will be reduced by one-third [10].

Today, data are the most important asset of organizations, including health organizations, and proper data collection and analysis is a crucial factor in the success of health organizations. The presence of these data is useful only when they can be used and data mining makes the use of data possible [11]. In fact, data mining is an advanced form of decision support and generates patterns, processes, and rules without any need for the user to ask questions [12].

The most important reason for using data mining is to increase the volume of existing and future data of the institutions and societies that require higher processing than human mental capacity and traditional approaches [13].

However, one of the most important and sensitive areas in the health realm is the problem of nosocomial infection which is of very sensitive physical, social and economic dimensions. Two million people on average face nosocomial infections in the United States annually, and this has led to many researches in this area. Obenshain in 2004, using data mining laws on patient blood cultures and clinical data obtained from the laboratory information system, has identified interesting new patterns. In this study, it has been shown that infection control with data mining system is more sensitive and precise than traditional infection control system [14].

In 2016, Silva et al. conducted a study based on hospital data from internal wards of Mexico hospitals over a five-year period. By using the powerful CDSS system, they could predict internal infections and reduce the incidence of nosocomial infections, decrease the costs, improve the health system, and increase patient safety [15].

In general, the availability of a system for predicting nosocomial infections can help improve the quality of the services at the health centers and reduce costs. Accurate and timely analysis of available information helps prevent the infections to a large extent and provide knowledge to detect and predict infections [16]. Therefore, the present study is designed to predict the type and outcome of treatment of nosocomial infection patients using neural network algorithm.

Methods

The present study is cross-sectional based on the registered data, done on the hospital infection data of selected hospitals of Hamadan Province including Alimoradian Hospital in Nahavand, Valiasr Hospital in

Tuyserkan, Ghaem Hospital in Asadabad, Valiasr Hospital in Razan, Imam Reza Hospital in Kabudarahang, Imam Hossein and Mehr Hospitals in Malayer, Be’sat Hospital, Shahid Beheshti Hospital, Fatemiyeh Hospital, Farshchian Hospital, and Sina Hospital in Hamadan City from March 2017 to March 2019, which included 5,680 cases. Nosocomial infections data of the hospitals are regisyered daily by a trained nosocomial infection expert in the Nosocomial Infection Software and is sent quarterly to the Deputy of Health. The data collection tool in this study was the output of Nosocomial Infection Registration Software that can be accessed as an excel file.

This software contains information such as hospital name, ward, infection type, date of infection incidence, date of hospitalization, gender and age of the patient, time from hospitalization to infection, and types of organism causing the infection. To register the information in the software, some considerations should be taken into account; for example, if more than one organism is reported, organisms that have had a larger number of colonies as well as a significant share of infection are registered. Besides, the criterion for the record in the system is the infection, not the patient, so a separate file is created for each infection, and one patient may be registered in the system several times due to the incidence of different infections during one hospitalization.

Considering the purpose of the present study, which is the classification and prediction of the type of micro-organism causing the infection as well as the nosocomial infection treatment outcome according to the patient's condition, a suitable and high-precision method should be selected for obtaining the results. One of the best methods for data prediction and classification in data mining is neural network algorithm whose procedure is derived from human brain neurons. This algorithm first receives training based on the existing data and then predicts and classifies new data based on the received training.

In the neural network method, the result is obtained by experiment and iteration, i.e. at first one or more subjects are selected as the target and then a set of data is iteratively tested with the predetermined target results until the algorithm achieves the desired learning pattern. As a result, the data analysis method is experimental. Overall, 80% of the available data were designated as training data and 20% as experimental data. To evaluate the efficiency of the applied methods, the model accuracy criterion or percentage of correct responses was used.

The neural network entries were coded medical information extracted from the patient’s past diseases history. In the present study, based on the patient information registered in the system, nine criteria involving the patient's basic clinical information, including name of the hospital and ward, infection-causing tool, organism, time from hospitalization to infection, length of hospital stay, age, gender, and weight were used as the explanatory criteria and two criteria of type of infection and the final outcome of treatment which is either death or discharge, together with the actual available results, were used as the intended targets for performing the necessary analyses in data mining.

Intelligent neural networks can discover the hidden scientific aspects and hidden relationships behind the data by analyzing the data and examining the relationship between the actual inputs and outputs. In this

study, neural network algorithm was implemented using IBM SPSS Modeler Software, version 18.0. Significance level of relations was considered less than 5% in all statistical stages.

Results

Prediction using artificial neural network of the 9 parameters, which are also depicted in Fig. 1, is also used as an input for network training. The neurons in the input layer are the clinical features of the patient. For network training, after introducing the input data, the output values for the inputs are introduced to the system and network training begins. After training the network using real-world inputs and outputs, the neural network can be used to get results for new inputs.

Each input has a real output and in the neural network, each output is a linear function of its input. In artificial neural networks, linear functions have parameters that can be adjusted to achieve the minimum error in training. The network can be trained by data from a group of subjects, and this can be done until the best value of the parameters is found to get the least error in the network [17].

Supervised Learning and Feed-Forward Neural Network with Back Propagation of Error Algorithm was used to predict the type of nosocomial infection and the outcome of the treatment. Neurons can use different stimulus functions to generate output, the most common of which are sigmoid logarithm functions, the sigmoid tangent, and the linear stimulus function.

The stimulus function used in this study is of the sigmoidal tangent type.

In the present study, among the different training methods, Back Propagation of Error Method, the Levenberg-Marquardt Algorithm, has been used due to the faster convergence in the training of medium-sized networks. The Back Propagation of Error Algorithm changes the network weights and bias values to reduce the performance function more rapidly [18]. The purpose is to predict the type of infection and the outcome of this type of infection, i.e. discharge or death. Hence, there will be two neurons for the output layer.

The network is trained with different values and the error rate was measured at the end of each iteration and the best responses were extracted. One of the important and difficult problems is determining the optimal number of hidden layers and the number of nodes. Trial and error method is a way to determine the best network structure. To find the number of nodes in the hidden layer and to have the minimum error, an algorithm was proposed by Hirose et al that through the change in the number of hidden nodes and calculation of the minimum error iterates until the iteration in which the mean squared error (MSE) which is less than determined value is found.

As shown in Fig. 2, the model is multilayered and the appropriate number of hidden nodes of the network is assumed to be 11 which has the lowest error rate of 8.2%. In addition, the accuracy of the model in this case is 76.9% that indicates the optimal performance for prediction using artificial neural network. After selecting the best architecture and building the network with 11 neurons in the hidden layer, the process

of training mode and testing the model began. After preparing the data, an artificial neural network model was developed which was of MLP neural network type.

After network training, the system error was determined via the specified output given by the user (actual output) and the output specified by the model. Figures 3 and 4 illustrate the degree of consistency and inconsistency between the actual output and the model output, respectively.

Figure 4. Part of the results of the table of the adjustment of the prediction with reality for the target of the infection code

Since each of the indicated criterion has a specific impact with a certain percentage, each of these criteria is identified with a degree of importance in this model, which is significant in the results of the model. The degree of importance determined by the model is illustrated in Fig. 5. As shown in the figure, the most important criterion is the length of hospital stay, followed by a slight difference by the name of the ward and tool, and the least significant is the gender criterion.

Based on the results obtained for the analysis of the results of all data, including the experimental data and test data - and not merely the test data-, were compared with their actual results to compare the adjustment percentage of the results of the algorithm and the actual results. The results of this adjustment for both of the intended targets can be observed in Table 1.

Table 1
Analysis of obtained results and actual results

Comparison with the target of the result				
	Training Data		Test Data	
Correct	4157	91.48%	1027	90.48%
Incorrect	387	8.52%	108	9.52%
Total	4544		1135	
Comparison with the target of infection code				
	Training Data		Test Data	
Correct	2833	62.35%	686	60.44%
Incorrect	1711	37.65%	449	39.56%
Total	4544		1135	

As shown in the table, the rate of adjustment in the first target, which is the result of the infection, is very high, but in the second target, although this result has a high percentage, it is not as large as the first target. This is due to two factors. The first target only has two modes, namely, death or discharge, but in the second target, the high incidence of infections has increased the number of probable modes; thus, the

adjustment of the results would have a lower quality. There are 43 types of infection codes in this type of data. The second reason is related to the data malfunction, which despite the pre-processing and fundamental correction of the data, still has an impact on the results.

In the first target characteristic, all records are registered and identified and there are no defects but in the second target characteristic, many records are not registered, meaning that the type of infection is not identified, which greatly affects the results obtained from the model. It is clear that the above two problems must be resolved to achieve higher quality results in the second target function, i.e. the type of infection. The first problem decreases sharply as the number of data increases and the second problem ends with the correct data registration.

Discussion And Conclusions

Nosocomial infection is one of the most important problems in the medical field that in addition to the adverse effects on patients with infection such as prolonged treatment and the patient's death has negative effects on the health and economy of society. Given the problems caused by nosocomial infections for patients, the patient's awareness of infection likelihood in terms of the clinical and location conditions present in the ward and hospital as well as the length of hospital stay is of importance. Consequently, if, based on the input data of a prediction, expert physicians be able to be aware of the likelihood of the type of infection that a patient may acquire, and predict the percentage of the probability of death, treatment, and discharge, they can be of great help in preventing infection as well as in creating favorable conditions to prevent the death of the patient [19].

Data mining technique in health care settings can be used in health insurance issues to detect fraud and abuse in health policy making, in order to use the best and most effective treatment method [20]. Due to the importance of this subject, this study seeks to provide an algorithm for classification and prediction of data in databases of medical and health care organizations and among available methods, the Artificial Neural Network method is used.

This network is multilayer and has two output layers and 11 hidden layers as well as 9 input layers. The results of the algorithm, with an accuracy of 91.8%, have had good adjustment with the actual results. The study results showed that the length of hospital stay is the most influencing factor on the incidence of nosocomial infection. Further, the hospitalization ward has the second most important role in the incidence of infection and the hospital has the fifth rank. Apart from the physicians, hospital managers and authorities can also use these infection-related criteria to improve general conditions of the hospital. In regards with the charts showing the impact of hospital type and ward type on the incidence of nosocomial infection, government and private sectors can provide incentive and punitive plans for hospital management. Besides, by focusing on the intended hospitals and wards, the conditions causing the incidence of infection can be reduced.

Two million people acquire nosocomial infections in the United States annually; therefore, big focus has been put on identifying these patients. In Alabama, for example, by using data mining laws and

relationships on patient blood cultures and clinical data obtained from the laboratory information system, new and interesting patterns are specified and patterns examined by infection control experts are prepared on a monthly basis. The makers of this system have found that promotion of the nosocomial infection control by a data mining system is more sensitive than the traditional nosocomial infection control system [14]. The results of the study by Karim et al., in Mashhad using data mining techniques to determine factors related to the length of hospital stay showed that hospital stay, referral reason and age variables are the most important factors associated with the increase of the length of hospital stay [21]. In their study, Azizi et al., using data mining techniques, identified body surface area (BSA) burn, degree of burn, and length of hospital stay as three factors influencing mortality of burn patients [22]. Obviously, such findings can be useful for planning and optimal allocation of resources and reducing medical costs in health care centers [19].

The important issue of nosocomial infection and its prevalence, as well as the registering and storage of infection-related data, have provided a good basis for future research. As a suggestion, classification and evaluation of the effects of infection in the two sensitive groups of infants and the elderly can be very important. In addition, considering the importance of the duration of treatment and its effects, the length of hospital stay related to nosocomial infection and its influencing factors can be evaluated and predicted.

Given that the main basis of this type of research is the existence of actual data on the subject in question, there are many limitations in the collection and availability of relevant data in these types of researches from which the present study is not excluded. Data mining methods also need data that are more sensitive from the point of view of accuracy to do the analysis. This research has spent a great deal of time to correct and adjust the data in order to use them.

Declarations

Ethics approval and consent to participate

Institutional review board approval was obtained from the ethics committees of Hamadan University of Medical Sciences.

Consent for publication

Written and informed consent was obtained from the patients.

Availability of data and materials

The data that support the findings of the study are available from the corresponding author in SPSS form upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was funded by the Hamadan University of Medical Sciences, Iran.

Authors' contributions

AF, HD, and SK developed the original idea and the protocol, abstracted, and prepared the manuscript. AF and SK participated in the study design and analyzed the data. AF, HD, and SK contributed to the data gathering. All authors read and approved the final manuscript.

Acknowledgments

This article is based on the results of the HSR research project with the Ethics Committee specific ID of IR.UMSHA.REC.1398.898 approved by the Research and Technology Department of Hamadan University of Medical Sciences whose authors must acknowledge the financial support of this department.

Conflicts of Interest

The authors declare that they have no competing interests.

References

1. Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J. Harrison's Principle of Internal Medicine .Mc Graw - Hill: New York; 2012.

2. Goldman L, Schafer AI .Goldman ís Cecil Medicine. Elsevierís Company: USA ;2 0 1

3. Reed D, Kemmerly SA. Infection control and prevention: a review of hospital-acquired infections and the economic implications. The Ochsner Journal. 2009;9(1):27-31.

4. Sydnor ER, Perl TM. Hospital epidemiology and infection control in acute-care settings. Clinical microbiology reviews. 2011;24(1):141-73.

5. Jroundi I, Khoudri I, Azzouzi A, Zeggwagh AA, Benbrahim NF, Hassouni F, et al. Prevalence of hospital-acquired infection in a Moroccan university hospital. American journal of infection control. 2007;35(6):412-6.

6. Lee M, Chiu C, Chow V, Lam R, Lai R. Prevalence of hospital infection and antibiotic use at a university medical center in Hong Kong. Journal of Hospital Infection. 2007;65(4):341-

7. Danchaivijitr S, Judaeng T, Sripalakij S, Naksawas K, Plipat T. Prevalence of nosocomial infection in Thailand 2006. Journal of the Medical Association of Thailand. 2007;90(8):1524.

8. Kallel H, Bahoul M, Ksibi H, Dammak H, Chelly H, Hamida CB, et al. Prevalence of hospital-acquired infection in a Tunisian hospital. Journal of Hospital Infection. 2005;59(4):343-7.

9. Khan HA, Ahmad A, Mehboob R. Nosocomial infections and their control strategies. Asian pacific journal of tropical biomedicine. 2015;5(7):509-14.

10. Haley RW, Culver DH, White JW, Morgan WM, Emori TG, Munn vP, et al. The efficacy oe infection surveillance and control programs in preventing nosocomial infections in us hospitals. American journal of epidemiology. 1985;121(2):182-205.

11. Jahanbakhsh M, Tavakoli N, Mokhtari H. Challenges of EHR implementation and related guidelines in Isfahan. Procedia Computer Science. 2011;3:1199-204.

12. Wager KA, Lee FW, Glaser JP. Health care information systems: a practical approach for health care management: John Wiley & Sons; 2017.

13. Berka P. Data Mining and Medical Knowledge Management: Cases and Applications: Cases and Applications: IGI Global; 2009.

14. Obenshain MK. Application of data mining techniques to healthcare data. Infection Control & Hospital Epidemiology. 2004;25(8):690-5.

15. Silva E, Cardoso L, Faria R, Santos MF. Nosocomial infection prediction using data mining technologies. Applying Business Intelligence to Clinical and Healthcare Organizations: IGI Global; 2016. p. 188-207.

16. Santos RS, Vaz TA, Santos RP, de Oliveira JMP, editors. Big Data Analytics in a Public General Hospital. International Workshop on Machine Learning, Optimization, and Big Data; 2016: Springer.

17. Hirose Y, Yamashita K, Hijiya S. Back-propagation algorithm which varies the number of hidden units. Neural Networks. 1991;4(1):61-6.

18. Sapra R, Mehrotra S, Nundy S. Artificial Neural Networks: Prediction of mortality/survival in gastroenterology. Current Medicine Research and Practice. 2015;5(3):119-29.

19. Silva E, Cardoso L, Portela F, Abelha A, Santos MF, Machado J. Predicting nosocomial infection by using data mining technologies. New Contributions in Information Systems and Technologies: Springer; 2015. p. 189-98.

20. Paramasivam V, Yee TS, Dhillon SK, Sidhu AS. A methodological review of data mining techniques in predictive medicine: An application in hemodynamic prediction for abdominal aortic aneurysm disease. Biocybernetics and Biomedical Engineering. 2014;34(3):139-45.

21. Karim H, Etminani K, Tara SM, Mardani M. Identifying Factors Associated with Length of Hospital Stay Using Decision Tree. Journal of Health Adminstration. 2015;18(61):57-68.

22. Azizi A, Zarei J, Nabovati E, Vakili-Arki H, Abbasi E, Razavi A. Determining of the Factors Affecting Mortality in Burn Patients Using a Decision Tree Data Mining Algorithm Journal of Health Adminstration. 2013;16(54):34-45.

Figures

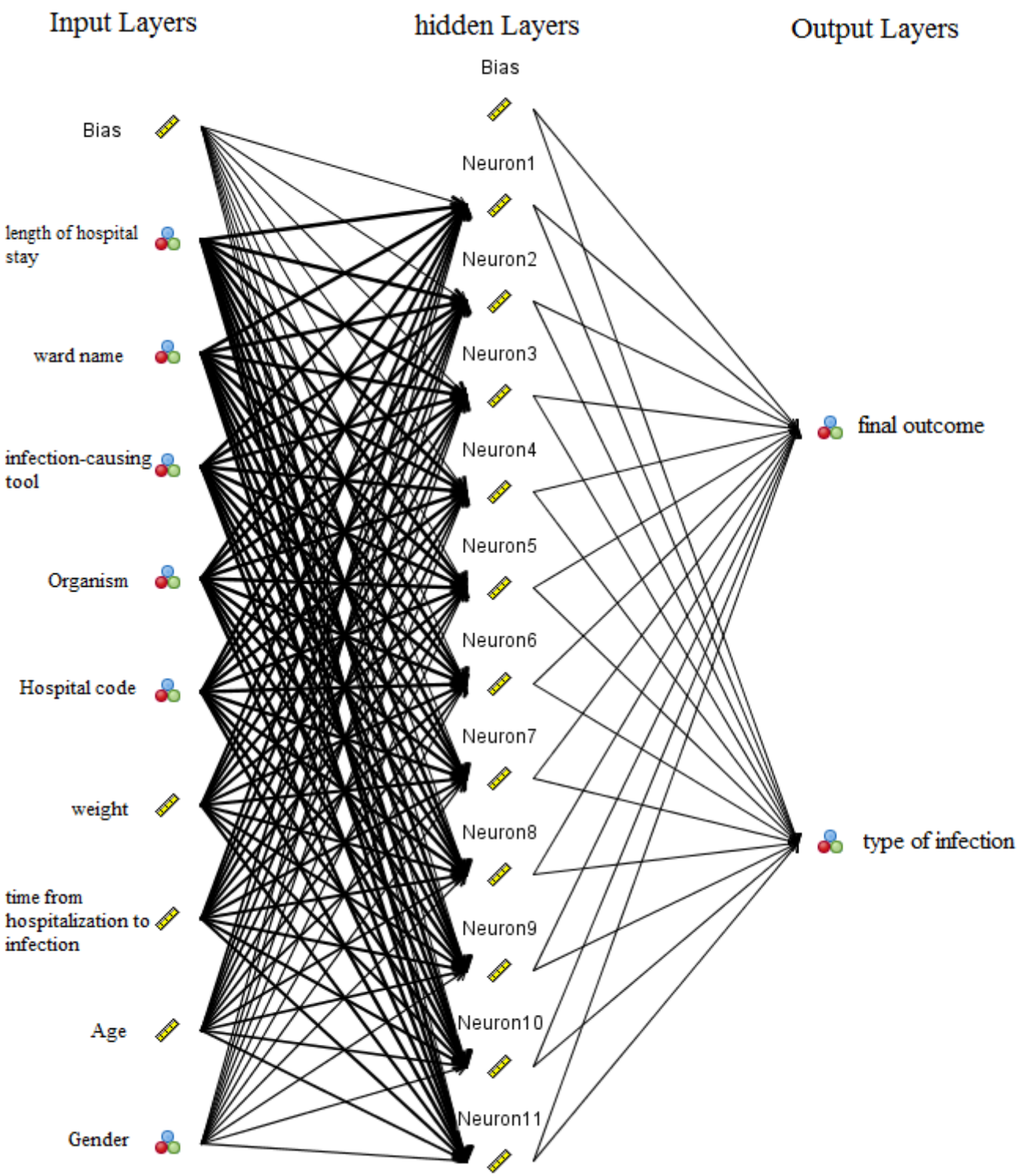

Figure 1

Three layer neural network architecture of the research

Model Summary

Targets	Final outcome Infection code
Model	Multilayer Perceptron
Stopping Rule Used	Error cannot be further decreased
Hidden Layer 1 Neurons	11

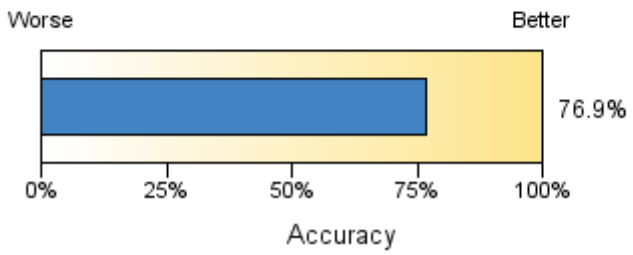

Figure 2

Structure of the accuracy of the model

Classification for Final outcome

Overall Percent Correct = 91.8%

Observed	Predicted		
	-	death	discharge
-	100.0%	0.0%	0.0%
death	0.0%	76.4%	23.6%
discharge	0.0%	8.3%	91.6%

Row Percent
100.00
80.00
60.00
40.00
20.00
0.00

Figure 3

Table of the adjustment of the prediction with reality for the target of the final outcome (death or discharge)

Observed	BJ-BONE	BSI-LCBI	BSI-MBI-LCBI	CNS-MEN	CVS-VASC	EENT-CONJ	EENT-EAR	EENT-EYE	EENT-UR	GI-GE	GI-IAB	LRI-LUNG	PNEU-PNU1
BJ-BONE	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
BSI-LCBI	0.0%	46.9%	0.0%	0.0%	0.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	16.8%
BSI-MBI-LCBI	0.0%	31.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	37.5%
CNS-MEN	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	50.0%
CVS-VASC	0.0%	4.0%	0.0%	0.0%	39.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	24.8%
EENT-CONJ	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%
EENT-EAR	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
EENT-EYE	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%
EENT-UR	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
GI-GE	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	41.7%
GI-IAB	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LRI-LUNG	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
PNEU-PNU1	0.0%	1.3%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	84.3%
PNEU-PNU2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	74.1%
PNEU-PNU3	0.0%	28.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	71.4%

Figure 4

Part of the results of the table of the adjustment of the prediction with reality for the target of the infection code

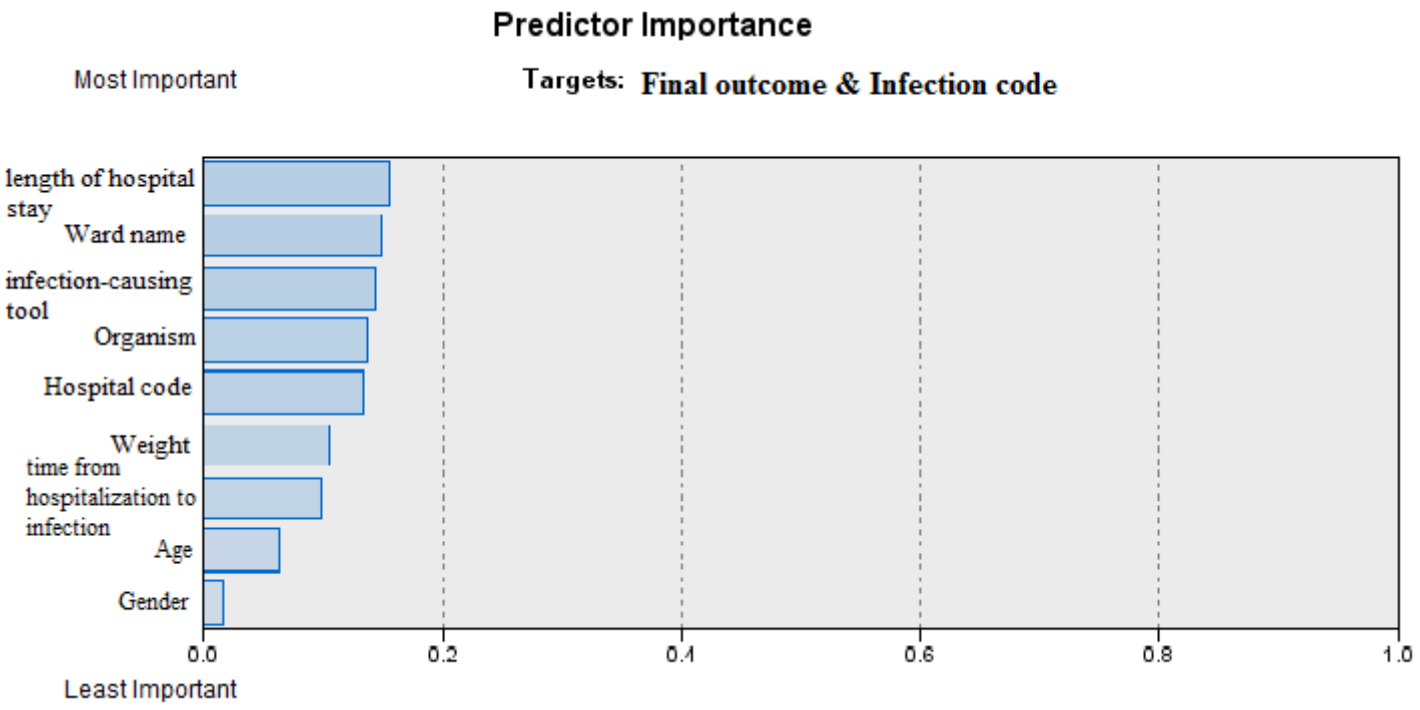

Figure 5

The importance degree of the criteria derived from the mod