Public Health and Informatics
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210280

Developing a Prognostic Model to Predict Mortality in Patients with Acute Bacterial Meningitis

Atiehsadat MIRKHANI[a,1] Arash ROSHANPOOR[b], Omid POURNIK[c], Hamed HADDADI[d], Jamal MIRZAEI[e] and Farzad KAVEH[f]
[a] *Biomedical Engineering faculty, Amirkabir University of Technology, Iran*
[b] *Departement of Computer Science, Sama Technical and Vocational Training College, Tehran Branch, Islamic Azad University, Iran*
[c] *Department of Medical Informatics, Mashhad University of Medical Sciences, Mashhad, Iran*
[d] *Dyson School of Engineering, Faculty of Engineering, Imperial College London, UK*
[e] *Infectious Disease Department, Shahid Beheshti University of Medical Sciences, Iran*
[f] *Center for Disease Control and Prevention, Tehran, Iran*

Abstract. Bacterial meningitis is one of the harmful and deadly infectious diseases, and any delay in its treatment will lead to death. In this paper, a prognostic model was developed to predict the risk of death amongst probable cases of bacterial meningitis. Our prognostic model was developed using a decision tree algorithm on the national meningitis registry of the Iranian Center for Disease and Prevention (ICDCP) containing 3,923 records of meningitis suspected cases in 2018-2019. The most important features have been selected for the model construction. This model can predict the mortality risk for the meningitis probable cases with 78% accuracy, 84% sensitivity, and 73% specificity. The identified variables in prognosis the death included age and CSF protein level. CSF protein level (mg/dl) <= 65 versus > 65 provided the first branch of our decision tree. The highest mortality risk (85.8%) was seen in the patients >65 CSF protein level with 30 years < of age. For the patients <=30 year of age with CSF protein level >137 (mg/dl), the mortality risk was 60%. The prognostic factors identified in the present study draw the attention of clinicians to provide early specific measures, such as the admission of patients with a higher risk of death to intensive care units (ICU). It could also provide a helpful risk score tool in decision-making in the early phases of admission in pandemics, decrease mortality rate and improve public health operations efficiently in infectious diseases.

Keywords. Prognostic model, acute bacterial meningitis, decision tree, crisp-dm, infectious disease pandemics

1. Introduction

Bacterial meningitis is a hazardous disease in the world, and its incidence is estimated to be around 20 cases per 100,000 people annually, namely 1.2 million patients. In Iran,

[1] Corresponding Author, Atiehsadat Mirkhani, Faculty of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran, Email: mirkhani_th@aut.ac.ir

more than 8,000 new suspected cases of meningitis are annually reported to ICDCP through health care centres and hospitals [1], which about 0.01% population of Iran. Using new methods such as prognostic models developed by machine learning algorithms could provide a helpful risk score tool in decision-making in the early phases of patients' admission and allocate high-care facilities or wards to them. Prognostic models can forecast the risk of future events in patients with a given disease and classify them by these risks [2]. In this study, we developed a prognostic model by decision tree algorithm on bacterial meningitis surveillance data in Iran to determine the probability of death among bacterial meningitis probable cases. In the methodology section, data understanding, data preparation, and the way of developing the model are described in detail. In Section 3, the prediction results of our model and the confusion matrix are described. In Section 4, the discussion of the research is presented. Finally, the summary of the findings and future research direction are discussed in Section 5.

2. Method

This study has been conducted using the process-based standard Cross-Industry Standard Process for Data Mining (CRISP-DM) [3]. It was performed on the Iranian Meningitis Registry that includes 26 discrete and continuous variables and 3,923 records of suspected meningitis cases in 2018-2019. In the business understanding phase, the diagnosis and treatment of bacterial meningitis were studied. The data distribution was considered in the data understanding phase. Due to the goal of our study, our target variable was the patient's discharge status. In the data preparation stage, missing values were imputed by classification and regression trees (CART) algorithm; also, outliers and extremes were replaced using Coerce method [4]. The software used in this phase included SPSS Modeler 14.2 and MS Excel 2013. According to an imbalance in the data set, the minority class (e.g. the patients who died of acute bacterial meningitis) in comparison with the majority class (e.g. the patients who survived the disease), synthetic minority over-sampling technique (SMOTE) [5, 6] was used to reach a more balanced number of cases in each class. In the next step, feature selection by principal component analysis (PCA) [7] was used to select appropriate variables to enter into the model. The variables have shown in Tables 1 and 2, respectively.

Table 1. Continuous Variables for Modeling

Variable	Range	Mean	Standard Deviation
Age (year)	1>=versus 68<	4,476	6,512
WBC count in CSF (cell/mm3)	0-34,320	1,046	3,296
Polymorphonuclear (PMN) percentage of WBCs	0-100	26	33
WBC count in CSF (cell/mm3)	0-34,320	1,046	3,296
Lymphocyte percentage of WBCs	0-100	26	32
CSF Protein Level (mg/dl)	0-730	74	98
CSF Glucose Level (mg/dl)	0-148	59	24

Table 2. Discrete variables for modeling

Variable	Frequency (Percent)
Gender	Male= 2,350 (60%) Female= 1,573 (40%)
CSF Appearance	Bloody = 170 (4.33%) Clear = 2,993 (76.29%)

	Turbid = 591 (15.08%) Unknown = 169 (4.3%)
Probable Diagnosis	No=1,375 (35%) Yes=2,548 (65%)

The selected attributes were entered into the modeling phase. Considering the goal of the study as patients' discharge status, the decision tree algorithm was used for the modeling by KNIME (Konstanz Information Miner, version 2.10.0) [8] (Fig. 1). The reason for selecting this algorithm was the strong ability to handle both continuous and discrete variables, and also its capability to provide the outcome as rules [9]. In this study, 10-fold cross-validation was performed in order to assess how the result of our analysis would generalize to an independent dataset [8, 10].

3. Result

In this study, the discharge status was defined as death (positive event) and alive (negative event), respectively. Table 3 shows the performance of the model in a way that each column represents the actual class instances while each row represents the instances in the predicted class [11]. After that, the results were compared based on accuracy, sensitivity, and specificity to evaluate the performance of the model. The model achieved an accuracy of 78% with a sensitivity of 84% and specificity of 73%. The area under the receiver operating characteristic (ROC) curve was used to measure the quality of the model [12]. The area under the curve (AUC) was 85%, which shows good discrimination (Fig. 2).

Table 3–Confusion matrix of outcome / prediction (Result)

	Actual Outcome	
Predicted Outcome	Death	Alive
Death	1,067	347
Alive	207	927

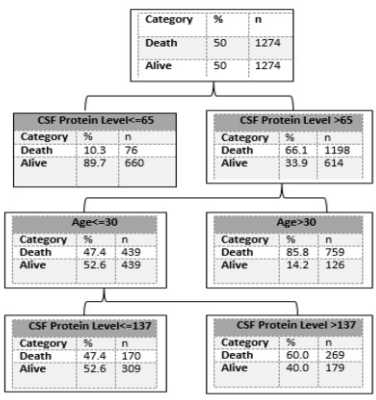

Figure 1. The Decision Tree with the training dataset predicts mortality

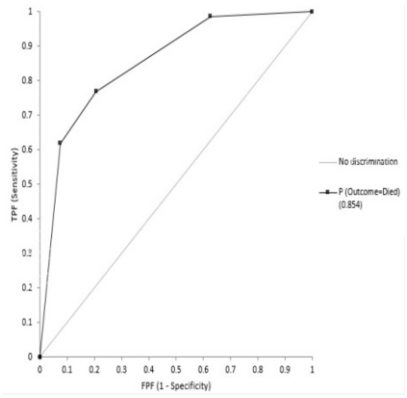

Figure 2. ROC curve and AUC of the model

4. Discussion

The purpose of this study was to develop a prognostic model that could assist clinicians in predicting the patients' discharge status. Figure 1 demonstrated a decision tree model to predict the risk of mortality among probable cases of acute bacterial meningitis in Iran. This model established distinct cut-off points of CSF protein level (mg/dl) <= 65 versus > 65 and <= 137 versus > 137 to prognosticate the death among the probable individuals. This is consistent with the previous studies that have established a high CSF protein level as a predictor of mortality. Previous studies have identified high CSF protein levels as an important risk factor; however, they did not report any cut-off point in the patients with different age groups [11]. The range of 100–500 mg/dl for CSF protein concentration is emphasized by the acute bacterial meningitis guidelines [8]. The second predictor established by the model is age <=30 versus > 30 (year). Although age in the former studies was categorized differently (children and adults), it has been associated with higher mortality in those groups [13]. The present study identified a strong association between CSF protein level and age as significant predictors of death probability among 2,548 probable meningitis cases in Iran. These two predictors identified the new cut-off points that were acquired from laboratory findings at the time of admission to the hospital for making the prognosis of the death probability among the probable. If there could be CSF analysis possibility, this model can be used for probable cases at primary care centers. Since no previous study has been performed in this manner, the results deserve significant consideration about probable cases in the matter of assessing the mortality. The final variables included in our model have been determined to be reliable predictors of the death probability in a large heterogeneous group of probable cases (different ages from the variety of different areas), which supports the generalizability of our results. Identification of prognostic factors in the early phases could minimize the occurrence of undesirable events. In addition, the results from this model can be applied in clinical decision support systems (CDSS) as computerized clinical guidelines.

Our study has some important limitations: The model was developed from the data of the national meningitis registry of Iran, which included poor and inappropriate data; as a result, there were a large number of deficiencies. The data were also gathered from across Iran; thus, we could not perform a validation on a dataset from the other parts of Iran. Further research is needed to validate the model in practice by suspected or probable cases from other countries in order to compare their results with ours.

5. Conclusion

The model was presented in this study may help physicians to prognosticate the mortality among the probable patients with acute bacterial meningitis. This model can be applied in clinical decision support systems (CDSS) as computerized clinical guidelines to draw the attention of clinicians to provide early specific measures, such as the admission of patients with a higher risk of death to intensive care units (ICU) [14]. In this way, special attention should be given to patients aged over 30 years and their CSF protein levels greater than 137.

References

[1] Esteghamati A, Askari F, and Goudarzi A. Guildline of Meningit Surveillance, ed. M.M. Gouya2006, Tehran: Iran Center for Disease Control and Prevention; 2006.
[2] Altman DG, et al. Prognosis and prognostic research: validating a prognostic model. BMJ. 2009;338.
[3] IBM, IBM SPSS Modeler CRISP-DM Guide, 2011.
[4] InternationalBusinessMachinesCorporation(IBM), Clementine 12.0, 2008.
[5] Chawla NV, et al. SMOTE: synthetic minority over-sampling technique. Journal Of Artificial Intelligence Research. 2002;16(1):321-357.
[6] Fernandez A, et al. SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. Journal of Artificial Intelligence Research. 2018;61.
[7] Oppel S, Strobl C and Huettmann F, Alternative Methods to Quantify Variable Importance in Ecology, Department of Statistics University of Munich. 2009.
[8] Yadav S, and Shukla S. Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. in 2016 IEEE 6th International Conference on Advanced Computing (IACC). 2016.
[9] Ramezankhani A, et al. Applying decision tree for identification of a low risk population for type 2 diabetes. Tehran Lipid and Glucose Study. Diabetes Research and Clinical Practice. 105(3):391-398.
[10] Mallett S, et al., Reporting performance of prognostic models in cancer: a review. BMC Medicine, 2010; 8(1):1-11.
[11] Jonge RC, et al. Predicting sequelae and death after bacterial meningitis in childhood: A systematic review of prognostic studies. BMC Infectious Diseases. 2010;10(1):1-14.
[12] Martínez-Camblor P. Area under the ROC curve comparison in the presence of missing data. Journal of the Korean Statistical Society. 2013;42(4):431-442.
[13] de Fatima Magalhaes Acioly Mendizabal M, et al. Prognostic indicators in bacterial meningitis: a case-control study. Braz J Infect Dis. 2013;17(5):538-44.
[14] WorldHealthOrganization. The 10 Essential Public Health Operations. Available at: https://www.euro.who.int/en/health-topics/Health-systems/public-health-services/policy/the-10-essential-public-health-operations, Accessed 2021.