

Research Article

Prediction of Bronchopneumonia Inpatients' Total Hospitalization Expenses Based on BP Neural Network and Support Vector Machine Models

Cuiyun Wu , Dahui Zha , and Hong Gao 

The Third People's Hospital of Hefei, Hefei, Anhui 230022, China

Correspondence should be addressed to Cuiyun Wu; wucuiyun3307@163.com and Hong Gao; gaohong_1971@126.com

Received 8 March 2022; Revised 13 April 2022; Accepted 5 May 2022; Published 18 May 2022

Academic Editor: Yao Chen

Copyright © 2022 Cuiyun Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. BP neural network (BPNN) model and support vector machine (SVM) model were used to predict the total hospitalization expenses of patients with bronchopneumonia. **Methods.** A total of 355 patients with bronchopneumonia from January 2018 to December 2020 were collected and sorted out. The data set was randomly divided into a training set ($n = 249$) and a test set ($n = 106$) according to 7:3. The BPNN model and SVM model were constructed to analyze the predictors of total hospitalization expenses. The effectiveness was compared between these two prediction models. **Results.** The top three influencing factors and their importance for predicting total hospitalization cost by the BPNN model were hospitalization days (0.477), age (0.154), and discharge department (0.083). The top 3 factors predicted by the SVM model were hospitalization days (0.215), age (0.196), and marital status (0.172). The area under the curve of these two models is 0.838 (95% CI: 0.755–0.921) and 0.889 (95% CI: 0.819–0.959), respectively. **Conclusion.** Both the BPNN model and SVM model can predict the total hospitalization expenses of patients with bronchopneumonia, but the prediction effect of the SVM model is better than the BPNN model.

1. Introduction

Bronchopneumonia (also known as lobular pneumonia) is one of the most common respiratory infections [1], with an incidence rate of more than 20% [2]. Bronchopneumonia is a severe disease that threatens people's health in China. It is also a disease that accounts for a large proportion of the spectrum of hospitalized infections [3, 4]. Bronchopneumonia is often caused by bacteria, viruses, molds, mycoplasma pneumonia, and other pathogens. It can also be "a mixed infection" by viruses and bacteria [5]. After the onset, the inflammation of lung tissue thickens the respiratory membrane and blocks the lower respiratory tract, causing dysfunction of ventilation and ventilation. The clinical manifestations are fever, cough, and shortness of breath [2].

The aging stage with the highest incidence of bronchopneumonia among children is 5~9 years old, and the onset age of patients gradually decreases [6]. Once infected, it will affect patients' quality of life and bring a certain economic

burden to families. In addition, the disease can cause pressure on the national medical insurance fund [7, 8]. Therefore, strengthening the cost research of bronchopneumonia and formulating effective intervention measures can reduce the economic burden on patients and medical insurance [9].

Data mining is a process that combines artificial intelligence and database technology to extract potentially valuable information from a large number of complex and fuzzy data [10–14]. The application of artificial intelligence in the medical field is gradually maturing. BP neural network (BPNN) model [15] and support vector machine (SVM) model [16] have no special requirements for data distribution and have certain fault tolerance. In addition, they are widely used in dealing with complex relationships between data and can seek the optimal solution under the current information [17]. Thus, this study used these two models to predict the total hospitalization cost of patients with bronchopneumonia and compared the prediction efficiency of the two models.

2. Methods

2.1. General Information. A total of 355 patients with bronchopneumonia who were mainly diagnosed as discharged from the first page of medical records from a grade III class hospital in Anhui province from January 2018 to December 2020 were collected. Inclusion criteria: (1) inpatients; (2) the diagnosis was bronchopneumonia. Exclusion criteria: (1) length of stay was 1 day; (2) it costs more than 40,000 yuan.

2.2. Research Indicators. The preliminary included research indicators include medical payment method, hospitalization times, sex, age, nationality, occupation, marital status, admission way, admission situation, whether to change majors, discharge departments, actual hospitalization days, whether to implement clinical pathway management, whether to complete clinical pathway, whether to have complications, whether to be critically ill or seriously ill during hospitalization, whether to meet the outpatient discharge diagnosis, whether to meet the admission and discharge diagnosis, admission condition, and whether to merge. The dependent variable is the total hospitalization expenses.

2.3. Partition of Data Set. Since the model construction of deep learning depends on the training of a large amount of data, it has the problem of uneven data distribution. Therefore, it is necessary to preprocess the data set. To prevent overfitting, this study verified the included data by a 10-fold crossover method. That is to say, it is divided according to the ratio of 7:3 to form a training set and a test set [18]. Among the patients, 70% of the data sets were used for the training set ($n = 249$) and 30% for the test set ($n = 106$).

2.4. Construction of Prediction Model

2.4.1. BP Neural Network Model. Total hospitalization expenses were used as the output variable, and statistically significant variables in univariate analysis were used as input variables. The hidden layer activation function is the hyperbolic tangent function, and the output layer activation function is the identity function. The data set is divided into the training set and test set, and the prediction model and BPNN model are constructed, respectively. The accuracy of the network will be calculated based on the verification set, where the relative error is the proportion of the sum of squares of the residuals and mean deviations to the sum of squares of the dependent variables. The prediction accuracy is 1-relative error [19]. After the network training is completed, the importance of each input variable to the prediction of the target variable is judged to reflect the relative effect of the input variable. The specific process is shown in Figure 1.

First, the algorithm is propagated forward. Calculate the output values of each neuron in the hidden layer and the output layer:

$$W_j = \sum_i w_{ji} u_i + b_j, \quad (1)$$

$$u_j = f(W_j). \quad (2)$$

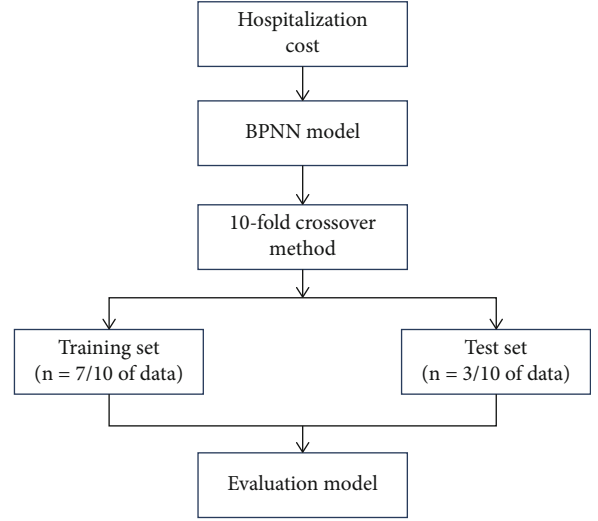


FIGURE 1: BP neural network construction flow chart.

Then, back propagation is carried out to calculate the error of each hidden layer neuron:

$$\delta_j = u_j(1 - u_j) \sum_i w_{ij} \delta_i. \quad (3)$$

δ_j is the sum of error information of all neurons in layer $j + 1$.

Finally, the weights of neurons are updated:

$$w_{ji} = w_{ji} - \Delta w_{ji}. \quad (4)$$

2.4.2. Support Vector Machine Model. In this study, the total hospitalization cost was a continuous variable, and the dependent variable should be discretized before the SVM fitting. The main parameters of the SVM model include penalty coefficient C and kernel function parameter σ . The selection of parameters in the SVM algorithm is very important to the learning performance of SVM. Reasonable parameter values can make SVM have higher training accuracy and stronger generalization ability. Therefore, this study will first screen out the optimal combination of C and σ parameters and establish the SVM model under the optimal combination of C and σ parameters. To select the optimal combination of C and σ parameters, the data is first normalized. Then, the data were input into the SVM model for verification to screen out the optimal combination. If the verification is inconsistent, the parameters need to be updated for verification again until the optimal combination is screened out. The SVM model building process is shown in Figure 2.

2.5. Statistical Analysis. One-way analysis was conducted on the relationship between the included research indicators and the total hospitalization expenses. Then, according to the results of the one-way analysis, significant variables are included in the BPNN model and SVM model as independent variables. After the training of the built models, the

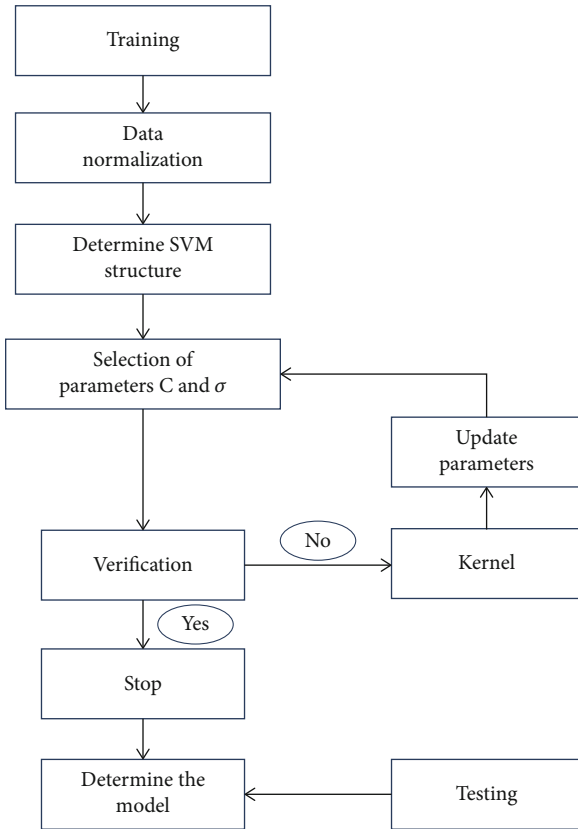


FIGURE 2: SVM model training process chart.

feature scores of associated predictors are screened out by machine learning.

3. Results

3.1. Analysis Results of Research Indicators. Univariate analysis was performed on the variables initially included. According to the results of medical payment method, hospitalization times, age, marital status, admission situation, critical illness during hospitalization, meet admission, and discharge, combined with other diagnosis, discharge departments, and receive surgical treatment have statistical significance ($P < 0.05$). Details are shown in Table 1.

However, gender, ethnicity, occupation, admission route, transfer department, complications, and discharge mode had no statistical significance on the total hospitalization cost ($P > 0.05$) (Table 2).

3.2. Results of Scoring Important Features in BPNN Algorithm Model. The results of BPNN model analysis show that the top three research indicators related to the total hospitalization expenses are hospitalization days (0.477), age (0.154), and discharge department (0.083). The characteristic scores of other research indicators are low (Figure 3).

3.3. Score Results of Important Features in SVM Algorithm Model. The results of the SVM algorithm model analysis show that the top three research indicators related to the total hospitalization expenses are hospitalization days

TABLE 1: Analysis results related to total hospitalization expenses of patients with bronchopneumonia.

Research indicators	χ^2/F	P
Medical payment method	151.359	0.023
Hospitalization times	0.652	0.038
Age	0.166	<0.001
Marital status	-18.791	<0.001
Admission situation	35.515	0.036
Critical illness during hospitalization	-11.746	0.027
Meet admission and discharge	-4.290	0.041
Combined with other diagnoses	-12.210	0.026
Discharge departments	421.631	0.048
Receive surgical treatment	-4.256	0.015

TABLE 2: Analysis results unrelated to the total hospitalization expenses of patients with bronchopneumonia.

Research indicators	χ^2/F	P
Gender	-0.314	0.814
Ethnicity	-0.325	0.792
Occupation	3.540	0.357
Admission route	3.451	0.069
Transfer department	4.259	0.071
Complications	-2.452	0.216
Discharge mode	2.421	0.255

(0.215), age (0.196), and marital status (0.172). The characteristic scores of other research indicators are low (Figure 4). As age may cause confounding of marital status, stratified analysis was conducted on age (≤ 25 years old and > 25 years old) and marital status. The results showed that if age was controlled, the correlation between marital status and total hospitalization cost was not statistically significant ($P > 0.05$).

3.4. Distinction between BPNN Model and SVM Model. The area under the curve (AUC) of the BPNN model is 0.838 (95% CI: 0.755~0.921), which meets the prediction accuracy requirements. In comparison, the AUC of the SVM model is 0.889 (95% CI: 0.819~0.959) (Figure 5). The two prediction models have obtained a good prediction effect. However, the prediction efficiency of the SVM model is higher than the BPNN model.

4. Discussion

Bronchopneumonia is an infectious disease with a high incidence in China, especially among children. Its clinical manifestations are fever, cough, and shortness of breath, which affect the normal life [20]. It will not only affect the quality of life of patients but also bring a certain economic burden to families and pressure to the national medical insurance fund. Symptomatic treatment is a common intervention method with bronchopneumonia, which can effectively

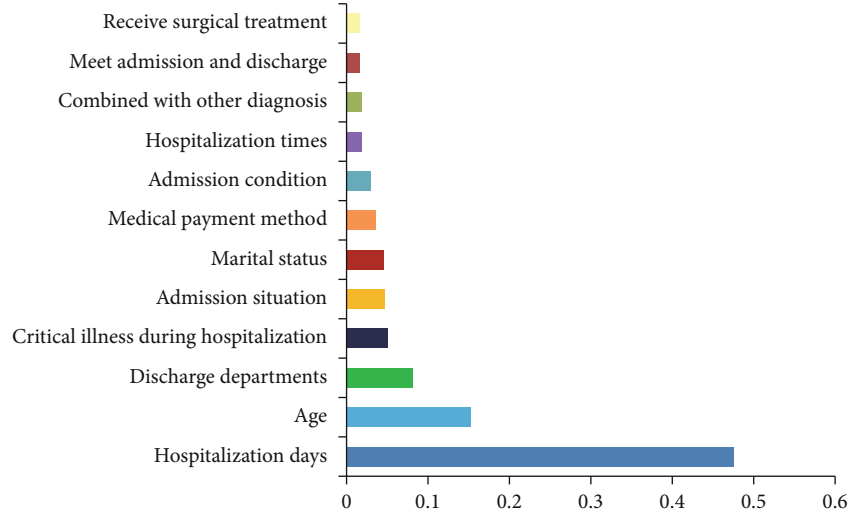


FIGURE 3: Score of important features in BPNN algorithm model.

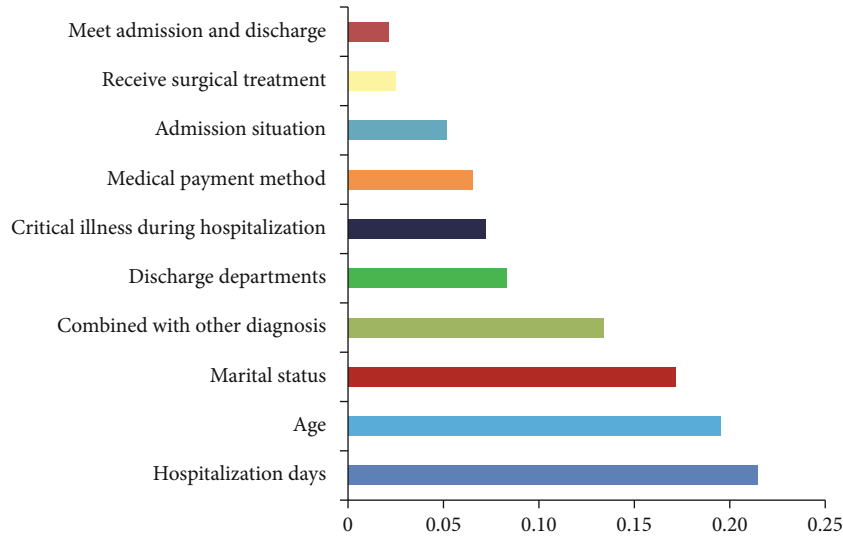


FIGURE 4: Score of important features in SVM algorithm model.

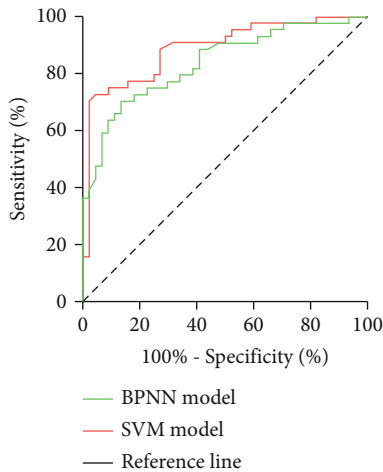


FIGURE 5: ROC curve of BPNN model and SVM model.

improve the symptoms and better control the development of the disease [21]. However, due to the younger age of patients, poor treatment compliance, and strong stress reaction, the hospitalization expenses are increased. Therefore, it is of specific clinical significance to predict the related indexes of total hospitalization expenses of patients with bronchopneumonia.

At present, the research on hospitalization expenses mainly includes traditional statistical methods, improved statistical methods, and machine learning methods [22, 23]. Traditional statistical methods have strict requirements on data, such as data normal and independent. Although nonparametric methods have no strict requirements on data characteristics, their efficiency is reduced because they do not use sample information to the maximum extent [24]. The improved statistical method combines other theories based on traditional methods and overcomes the inevitable

defects of traditional methods to a certain extent. However, for some complex data, such as hierarchical data, subdepartment data, and doctor data within the hospital, the improved method is more complicated to calculate [25]. Before machine learning, we need to conduct a careful and in-depth preanalysis of the included data set. Otherwise, the results may be misleading. As a grey-box method, data mining can get correct results as long as researchers correctly master the input format of data and the way of reading the results, so it has perfect practicability [26].

In this study, the BPNN model and SVM model were used to analyze the total hospitalization cost of patients with bronchopneumonia, and good prediction results were obtained. The analysis of influencing factors pointed out that the length of stay and discharge department are two significant factors affecting the cost, which has practical guiding significance. The length of hospitalization is related to the severity of the disease and the effect of treatment, so it is necessary to improve the accuracy of treatment and rational use of antibacterial drugs in clinical practice. Different antibiotics can be selected for different patients at the beginning of admission according to their sputum culture results [27, 28]. For example, for older patients who have been exposed to antibiotics for a long time and higher-grade antibiotics can be selected to improve the treatment effect. For those with good physical quality and sensitivity to antibiotics, low-grade antibiotics can be selected appropriately [29]. The different treatment methods and medication habits of doctors in different departments lead to the difference in total hospitalization expenses [30]. Therefore, it is recommended that doctors select appropriate treatment plans for patients. At the same time, the patient's age is also a major factor affecting hospitalization expenses. This is mainly because older patients have more comorbidities, relatively poor resistance, and low sensitivity to drugs, so there are relatively more drugs in the treatment process, resulting in prolonged hospitalization days and increased hospitalization expenses [31]. The results of the SVM model in this study showed that marital status was a major factor affecting the total hospitalization cost. Still, there was no statistical significance after stratified analysis and control of confounding factors.

The AUC of the prediction models was also compared in our study. The results show that the AUC of these two models is 0.838 (95% CI: 0.755~0.921) and 0.889 (95% CI: 0.819~0.959), respectively. It further shows that the prediction effect of SVM is better than the BPNN model. The reason may be that the SVM pursues optimal solutions under existing information and can perfectly solve high and local extremum problems [32]. The SVM overcomes the defects of the BPNN method, such as the difficulty in determining the reasonable structure and the existence of local optimum, especially for the data with dependent variables as classification variables, and has been effectively used in practice [17].

There are several limitations to our study. First, the specimens included were too small and from the same region. In addition, the incompleteness of the included predictors may directly affect the prediction results after deep learning. In future research, the sample size and prediction factors will be increased.

5. Conclusion

In conclusion, BPNN and SVM prediction models can effectively predict the total cost of hospitalized patients, and the most critical factor affecting the total cost of hospitalization is the length of stay. Therefore, shortening the length of stay may minimize the financial burden of patients.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

All authors declare no conflicts of interest in this paper.

Acknowledgments

The project is supported by the key scientific research project of Hefei Health Commission in 2020 (hwk2020zd006).

References

- [1] D. Rutenberg, M. Venner, and S. Giguère, "Efficacy of tula-thromycin for the treatment of foals with mild to moderate bronchopneumonia," *Journal of Veterinary Internal Medicine*, vol. 31, no. 3, pp. 901–906, 2017.
- [2] Y. Wang, Y. Sun, H. Zhang, X. Yang, and X. Song, "Comprehensive analysis of the diagnosis and treatment of tracheo-bronchial foreign bodies in children," *Ear, Nose, & Throat Journal*, vol. 100, article 1455613211023019, 2021.
- [3] C. You, G. Ran, X. Wu et al., "High immunoglobulin e level is associated with increased readmission in children with bronchopneumonia," *Therapeutic Advances in Respiratory Disease*, vol. 13, article 1753466619879832, 2019.
- [4] X. Han, Y. Yang, Q. Zhu, X. Wang, and W. Huang, "Clinical value of atomization therapy in children with bronchopneumonia," *Minerva Pediatr (Torino)*, vol. 74, no. 1, pp. 94–96, 2022.
- [5] L. Lindström, F. A. Tauni, and K. Vargmar, "Bronchopneumonia in Swedish lambs: a study of pathological changes and bacteriological agents," *Acta Veterinaria Scandinavica*, vol. 60, no. 1, p. 54, 2018.
- [6] J. Ye, H. Ye, M. Wang, and Y. Zhao, "Total serum il-6 and tnf-c levels in children with bronchopneumonia following treatment with methylprednisolone in combination with azithromycin," *American Journal of Translational Research*, vol. 13, no. 8, pp. 9458–9464, 2021.
- [7] D. A. Aziz, A. G. Billoo, A. Qureshi, M. Khalid, and S. Kirmani, "Clinical and laboratory profile of children with cystic fibrosis: experience of a tertiary care center in Pakistan," *Pak J Med Sci*, vol. 33, no. 3, pp. 554–559, 2017.
- [8] Z. Shen, Y. Zhang, H. Li, and L. Du, "Rapid typing diagnosis and clinical analysis of subtypes a and b of human respiratory syncytial virus in children," *Virology Journal*, vol. 19, no. 1, p. 15, 2022.
- [9] A. Buja, A. Bardin, G. Grotto et al., "How different combinations of comorbidities affect healthcare use by elderly patients with obstructive lung disease," *NPJ Prim Care Respir Med*, vol. 31, no. 1, p. 30, 2021.

- [10] K. C. Chen, H. R. Yu, W. S. Chen et al., "Diagnosis of common pulmonary diseases in children by X-ray images and deep learning," *Scientific Reports*, vol. 10, no. 1, p. 17374, 2020.
- [11] S. Schalekamp, W. M. Klein, and K. G. van Leeuwen, "Current and emerging artificial intelligence applications in chest imaging: a pediatric perspective," *Pediatric Radiology*, vol. no, pp. 1–11, 2021.
- [12] Y. Ye, J. Shi, D. Zhu, L. Su, J. Huang, and Y. Huang, "Management of medical and health big data based on integrated learning-based health care system: a review and comparative analysis," *Computer Methods and Programs in Biomedicine*, vol. 209, p. 106293, 2021.
- [13] D. Wang, S. Fong, R. K. Wong, S. Mohammed, J. Fiaidhi, and K. K. Wong, "Robust high-dimensional bioinformatics data streams mining by odr-iovfdt," *Scientific Reports*, vol. 7, p. 43167, 2017.
- [14] L. Y. Chuang, C. H. Yang, J. H. Tsai, and C. H. Yang, "Operon prediction using chaos embedded particle swarm optimization," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 10, no. 5, pp. 1299–1309, 2013.
- [15] D. Zhao, M. Chen, K. Shi, M. Ma, Y. Huang, and J. Shen, "A long short-term memory-fully connected (lstm-fc) neural network for predicting the incidence of bronchopneumonia in children," *Environmental Science and Pollution Research International*, vol. 28, no. 40, pp. 56892–56905, 2021.
- [16] A. Nedaie and A. A. Najafi, "Support vector machine with Dirichlet feature mapping," *Neural Networks*, vol. 98, pp. 87–101, 2018.
- [17] C. Bao, Y. Pu, and Y. Zhang, "Fractional-order deep backpropagation neural network," *Comput Intell Neurosci*, vol. 2018, article 7361628, 2018.
- [18] E. Gong, J. M. Pauly, M. Wintermark, and G. Zaharchuk, "Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI," *Journal of Magnetic Resonance Imaging*, vol. 48, no. 2, pp. 330–340, 2018.
- [19] S. K. Tian, N. Dai, L. L. Li, W. W. Li, Y. C. Sun, and X. S. Cheng, "Three-dimensional mandibular motion trajectory-tracking system based on bp neural network," *Mathematical Biosciences and Engineering*, vol. 17, no. 5, pp. 5709–5726, 2020.
- [20] J. E. Jeong, J. E. Soh, J. H. Kwak et al., "Increased procalcitonin level is a risk factor for prolonged fever in children with mycoplasma pneumonia," *Korean Journal of Pediatrics*, vol. 61, no. 8, pp. 258–263, 2018.
- [21] X. Liu and J. Meng, "Luteolin alleviates LPS-induced bronchopneumonia injury *in vitro* and *in vivo* by down-regulating microRNA-132 expression," *Biomedicine & Pharmacotherapy*, vol. 106, pp. 1641–1649, 2018.
- [22] S. Lee and H. Lim, "Review of statistical methods for survival analysis using genomic data," *Genomics Inform*, vol. 17, no. 4, article e41, 2019.
- [23] Z. R. Zhou, W. W. Wang, Y. Li et al., "In-depth mining of clinical data: the construction of clinical prediction model with r," *Ann Transl Med*, vol. 7, no. 23, p. 796, 2019.
- [24] L. N. Grendas, L. Chiapella, D. E. Rodante, and F. M. Daray, "Comparison of traditional model-based statistical methods with machine learning for the prediction of suicide behaviour," *Journal of Psychiatric Research*, vol. 145, pp. 85–91, 2021.
- [25] X. Qiu, J. Gao, J. Yang, J. Hu, W. Hu, L. Kong et al., "A comparison study of machine learning (random survival forest) and classic statistic (cox proportional hazards) for predicting progression in high-grade glioma after proton and carbon ion radiotherapy," *Frontiers in Oncology*, vol. 10, p. 551420, 2020.
- [26] A. K. Verma, S. Pal, and S. Kumar, "Prediction of skin disease using ensemble data mining techniques and feature selection method—a comparative study," *Applied Biochemistry and Biotechnology*, vol. 190, no. 2, pp. 341–359, 2020.
- [27] S. L. Rolsma, D. A. Rankin, Z. Haddadin et al., "Assessing the epidemiology and seasonality of influenza among children under two hospitalized in Amman, Jordan, 2010–2013," *Influenza and Other Respiratory Viruses*, vol. 15, no. 2, pp. 284–292, 2021.
- [28] L. Zhang, M. Lai, T. Ai et al., "Analysis of mycoplasma pneumoniae infection among children with respiratory tract infections in hospital in Chengdu from 2014 to 2020," *Transl Pediatr*, vol. 10, no. 4, pp. 990–997, 2021.
- [29] S. T. Kudagammana, R. R. Karunaratne, T. S. Munasinghe, and H. Kudagammana, "Community acquired paediatric pneumonia; experience from a pneumococcal vaccine-naive population," *Pneumonia (Nathan)*, vol. 12, no. 1, p. 8, 2020.
- [30] M. Wetzig, M. Venner, and S. Giguère, "Efficacy of the combination of doxycycline and azithromycin for the treatment of foals with mild to moderate bronchopneumonia," *Equine Veterinary Journal*, vol. 52, no. 4, pp. 613–619, 2020.
- [31] Y. Lu, Y. Wang, C. Hao et al., "Clinical characteristics of pneumonia caused by mycoplasma pneumoniae in children of different ages," *International Journal of Clinical and Experimental Pathology*, vol. 11, no. 2, pp. 855–861, 2018.
- [32] S. Huang, N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang, and W. Xu, "Applications of support vector machine (svm) learning in cancer genomics," *Cancer Genomics Proteomics*, vol. 15, no. 1, pp. 41–51, 2018.