BMC Medical Informatics and Decision Making

RESEARCH

Open Access

A machine learning-based severity stratification tool for high altitude pulmonary edema

Luobu Gesang[1,2,3*†], Yangzong Suona[1,4*†], Zhuoga Danzeng[1,3], Bai Ci[1,3], Quzhen Gesang[1,3], WangJiu Cidan[4], Qiangba Dingzeng[1,4], Zhuoga Baima[1,4] and Quzhen Zhaxi[4]

Abstract

This study aimed to identify key predictors for the severity of High Altitude Pulmonary Edema (HAPE) to assist clinicians in promptly recognizing severely affected patients in the emergency department, thereby reducing associated mortality rates. Multinomail logistic regression, random forest, and decision tree methods were utilized to determine important predictor variables and evaluate model performance. A total of 508 patients diagnosed with HAPE were included in the study, with 53 variables analyzed. Lung rales, sputum sputuming, heart rate, and oxygen saturation were identified as the most relevant predictors for the LASSO model. Subsequently, Multinomail logistic regression, decision tree, and random forest models were trained and evaluated using these factors on a test set. The random forest model showed the highest performance, with an accuracy of 77.94%, precision of 70.27%, recall of 68.22%, and F1 score of 68.96%, outperforming the other models. Further analysis revealed significant differences in predictive capabilities among the models for HAPE patients at varying severity levels. The random forest model demonstrated high predictive accuracy across all severity levels of HAPE, particularly excelling in identifying severely ill patients with an impressive AUC of 0.86. The study assessed the reliability and effectiveness of the HAPE severity scoring model by validating Multinomail logistic regression and random forest models. This study introduces a valuable screening tool for categorizing the severity of HAPE, aiding healthcare providers in recognizing individuals with severe HAPE, enabling prompt treatment and the formulation of suitable therapeutic approaches.

Objectives

- Develop the first machine learning-based risk stratification model for High-Altitude Pulmonary Edema (HAPE).
- Utilize multiple physiological and clinical indicators to predict disease severity.
- To accurately find critically ill patients to reduce their mortality.

†Luobu Gesang and Yangzong Suona contributed equally to this work and should be considered as co-first authors.

*Correspondence:
Luobu Gesang
KelsangNorbu@hotmail.com
Yangzong Suona
yanglag0106@gmail.com

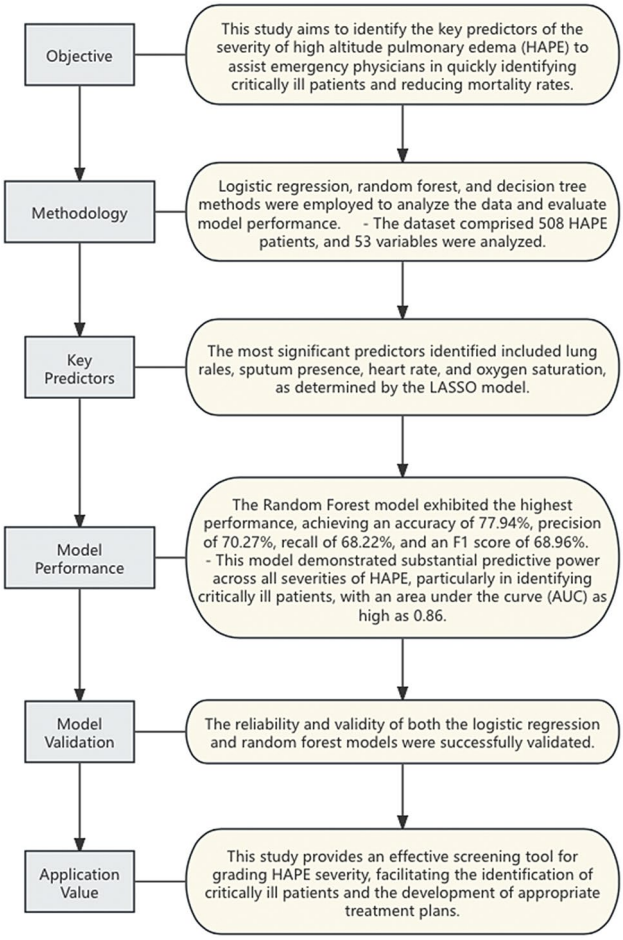
Full list of author information is available at the end of the article

• Compare the performance of various machine learning algorithms, including Multinomail logistic regression, random forest, and decision tree, to identify the optimal model.

Research Gaps

- Evaluate the effectiveness and clinical applicability of the developed model.
- There is a scarcity of machine learning approaches specifically targeting risk stratification for HAPE;
- Current methods for assessing the severity of HAPE are often subjective and lack precision;
- There is an absence of objective and efficient tools for identifying individuals at risk of developing high altitude pulmonary edema.

Graphical Abstract



Keywords High altitude, Hypobaric hypoxia, High altitude pulmonary edema, Acute mountain sickness

Background

High altitude Pulmonary Edema (HAPE) is a life-threatening condition that occurs due to rapid exposure to high altitude environments [6]. The incidence of HAPE has been on the rise with the increasing popularity of travel and outdoor sports, particularly when individuals quickly ascend to high altitudes [3, 8]. According to the United Nations World Tourism Organization (UNWTO), highland tourism accounts for between 9% and 16% of total global international tourism [20]. Thus, early intervention is crucial in reducing mortality rates associated with HAPE as timely treatment can have a significant impact. However, inexperienced physicians may face challenges in promptly diagnosing sever patients, leading to delays in treatment and potentially fatal outcomes. Reports indicate that delayed diagnosis and treatment of severe HAPE can result in mortality rates of 40–50% [2, 15]. Therefore, the development of effective tools for early screening, risk stratification, and treatment based on risk assessment can greatly enhance patient outcomes and reduce deaths caused by treatment delays.

Traditionally, the diagnosis of High Altitude Pulmonary Edema (HAPE) has relied on clinical symptoms and imaging tests. However, assessing the severity of HAPE has often been subjective, lacking accurate screening tools for sever patients. In recent years, advancements in machine learning technology have offered a new approach to evaluating the risk level of HAPE [14]. By integrating various physiological parameters and clinical indicators, machine learning models can provide more precise and efficient risk assessment without adding to the patient's burden [12]. While previous studies have explored machine learning applications in medical diagnosis, research on risk stratification modeling specifically for HAPE patients remains limited [10]. This study aims to address this gap by developing a machine learning-based tool for risk stratification in HAPE. The tool will utilize multiple physiological and clinical indicators to predict disease severity and accurately identify sever patients. Through retrospective analysis of data from a larger sample of patients with High altitude pulmonary edema, three models - Multinomail logistic regression, random forest, and decision tree - were compared to determine the optimal model and assess its effectiveness in real-world clinical settings. This research aims to provide decision support for clinicians, ultimately enhancing the diagnostic and treatment standards, as well as the survival rates, of patients with severe HAPE.

Methodology

Study population

This retrospective study was conducted at the People's Hospital of the Tibet Autonomous Region (TAR), which accounts for approximately 20% of the total outpatient emergencies and inpatient admissions among the seven municipal level hospitals in TAR. A total of 508 patients diagnosed with High altitude pulmonary edema between January 2014 and April 2022 were included in the study. Inclusion criteria were rapid ascent from the plains to the High altitude and first diagnosis of High altitude pulmonary edema recorded in the electronic medical record (EMR) system. Exclusion criteria were pediatric cases and patients with chronic obstructive pulmonary disease, heart failure, or cancer.

Data collection and processing

Data collection: baseline medical information, including age, gender, ethnicity, body mass index, clinical symptoms (including fever, malaise, loss of appetite, nausea, sleep disorders, dizziness, sputum, chest tightness, shortness of breath, palpitations, blue lips, and sputuming up sputum); vital signs, including temperature, heart rate, SpO2, systolic blood pressure (SBP), and diastolic blood pressure (DBP), demographics (including elevation of residence, elevation of onset of disease, transportation to the High altitude, whether or not they bathed and caught a cold after entering the High altitude, and number of days/time of first symptoms after arrival at the High altitude), and other information (history of smoking, alcohol consumption, hypertension, diabetes mellitus, chronic respiratory disease, and chronic liver disease). Variables with more than 20% missing data were excluded. For variables with less than 20% missing data, the random forest method was used for estimation [21].

In this study, we used a variety of statistical and machine learning methods to predict the severity of High altitude pulmonary edema (HAPE) and to select predictor variables. These methods consisted of the following: (1) data prepossessing, where the dateset was divided into a training set and a test set and split in a ratio of 80:20; (2) feature selection: Multinomail logistic regression models used recursive feature elimination (RFE) to select the most important predictor variables; random forest models used feature importance metrics to identify key features; and decision tree models selected the important variables through the structure of the tree; (3) model Training and Evaluation Multinomail logistic regression, random forest and decision tree models are used to train on the training set; accuracy is calculated from the prediction results on the test set; to evaluate the performance of the models, ROC curves and area under the curve (AUC) are used to measure the performance of each model on different categories.

Statistical analysis

This study used R language (V.3.6) for statistical analysis. General principles: All statistical tests were two-sided, and a P-value of less than or equal to 0.05 was considered to be statistically different from the difference being tested. The distribution of the data will be examined first, and non-normally distributed data will be treated, or special analytical methods will be used. Quantitative indicators will be described by calculating the mean, standard deviation, median, minimum, maximum, percentile, and categorical indicators will be described by the number of cases and percentage of each category. Comparisons between two general groups will be analyzed by appropriate methods depending on the type of indicator, group t-test or Wilcoxon rank test for quantitative data, chi-square test or exact probability method for categorical data, and Wilcoxon rank test or CMH test for hierarchical data. The ROC curves of the Multinomail logistic regression model, random forest model, and decision tree model were compared to determine the differences in model performance in predicting HAPE of different severities. AUC is the key metric used to measure the performance of the model, and the closer the AUC value is to 1, the better the predictive ability of the model.

Clinical diagnostic criteria for HAPE

Clinical diagnosis of high altitude pulmonary edema is based on diagnostic criteria (literature), including the following: (1) recent arrival at a high High altitude (above 2,500 m above sea level) or from a high altitude to a higher altitude area; (2) Symptoms: the presence of dyspnea, sputum, sputum sputum, headache, nausea, fatigue, etc., and in severe cases, resting dyspnea and telangiectasia; signs: central cyanosis, shortness of breath, tachycardia, and rales audible on lung auscultation; (3) X-rays suggesting unilateral or bilateral infiltrative shadows, which are often diffuse, irregularly distributed, or fused to form large, patchy shadows; and CT of the chest suggesting scattered flocculent, nodular, or reticulo-like shadows, and scattered flakes with milled-glass density, in both lungs); (4) Exclude other causes of acute respiratory distress syndrome, cardiogenic pulmonary edema and severe pneumonia [8].

Grading criteria for HAPE

In this study, patients diagnosed with High altitude pulmonary edema were classified as severe by (1) patients who were assessed as sever and admitted to the ICU by a High altitude medicine clinician through the diagnostic criteria after being seen in the emergency room, and (2) patients who had a CT report of the lungs showing: textural changes in one/both lungs with a small amount of flocculation and patchiness were considered to be mild, and flocculation/patchwork/exudation/milled-glass shadows in the middle and lower fields of one or both lungs, or changes in the inner banding of both lungs were considered to be moderate. Flocculent/patchy/exudative/ground glass shadows in both lungs are severe [7].

Results

A total of 336 patients (266 males and 70 females) diagnosed with High altitude pulmonary edema (HAPE) were collected from January 2014 to April 2022 at the People's Hospital of Tibet Autonomous Region. The demographic and clinical profiles of patients with varying degrees of HAPE severity are summarized in Table 1. This table includes details on patient demographics like gender, age, height, weight, BMI, symptoms (e.g., vertigo, sputum, shortness of breath, palpitations, bluish lips, etc.), and physiological measurements (e.g., temperature, heart rate, oxygen saturation, systolic blood pressure, and diastolic blood pressure). The patient distribution across severity levels was as follows: 67 mild cases, 240 moderate cases, and 29 severe cases of HAPE. The median age for patients with mild, moderate, and severe HAPE was 35, 31, and 40 years, respectively. The distribution of mild, moderate, and severe cases was 49, 195, 22 in males and 18, 45, 7 in females. Most basic clinical and vital sign variables did not show significant differences between the training and validation groups (Table 2). However, five variables including nausea symptoms, heart rate (HR), oxygen saturation (SpO2), and CT scans exhibited notable differences among the three severity groups. Figure 1 illustrates the distribution of oxygen saturation and heart rate among patients with varying severity levels, highlighting the significant contrast between mild cases and the other two groups. Noteworthy, gender, age, BMI, and temperature did not exhibit significant differences between the severity groups.

Among the three groups of patients (mild, moderate, and severe), symptomatic nausea was reported in 18% of mild patients, 19% of moderate patients, and 41% of severe patients. The incidence of nausea was significantly higher in the severe group compared to the mild and moderate groups, with a statistically significant difference between the three groups ($p = 0.017$). Regarding SpO2 levels, the median SpO2 was 80 (IQR: 74–86) in mild patients, 74 (IQR: 65–84) in moderate patients, and 74 (IQR: 63–85) in severe patients. SpO2 was significantly higher in the mild group than in the moderate and severe groups ($p = 0.007$). The median HR was 90 (IQR: 80–100) in mild patients, 100 (IQR: 87–119) in moderate patients, and 100 (IQR: 92–115) in severe patients. HR was significantly lower in the mild group compared to the moderate and severe groups, with a significant difference among the three groups ($p < 0.001$). The prevalence of a large number of rales was significantly higher in the severe group (76%) than in the mild and moderate groups ($p = 2.2E\text{-}16$). Minor rales were most prevalent in the moderate group (52%), moderate rales in the moderate group (25%), and major rales in the severe group (76%), showing a significant difference between the three groups ($p = 2.2E\text{-}16$). In terms of CT findings, banding/patchiness/flocculation abnormalities in the lower fields of both lungs were most common in the mild group (75%), banding/flocculation in the lower fields of both lungs were most common in the moderate group (87%), and banding/flocculation in the lower fields of both lungs were most common in the severe group (69%). Abnormalities on CT scans were significantly higher in the severe group compared to the mild and moderate groups, with a significant difference between the three groups ($p < 0.001$).

Predictor selection and modeling

The preliminary data analysis revealed that the target variable involves three categories. To identify the most significant predictors, feature selection was initially performed using the LASSO model. The LASSO model, a regularized linear model, helps prevent overfitting by incorporating an L1 regularization term and identifies the most crucial predictors. Ultimately, the LASSO model identified four key predictors: lungrales, sputum, heart rate (HR), and SpO2. The visualization of feature

Table 1 Clinical characteristics of patients with mild, moderate and severe HAPE

Characteristic	Mild, $N=67$[1]	Moderate, $N=240$[1]	Severe, $N=29$[1]	p-value[2]
Sex				0.3
Female	18(27%)	45(19%)	7(24%)	
Male	49(73%)	195(81%)	22(76%)	
Age	35 (27, 42)	31 (26, 42)	40 (27, 44)	0.089
Height	169 (165, 172)	170 (165, 173)	169 (163, 171)	0.3
Weight	74 (64, 77)	74 (64, 77)	75 (69, 78)	0.5
BMI	24.9 (22.9, 26.3)	24.8 (23.0, 26.4)	25.5 (23.8, 26.4)	0.6
Fatigue	15 (22%)	60 (25%)	7 (24%)	>0.9
Loss of taste	15 (22%)	56 (23%)	6 (21%)	>0.9
Nausea	12 (18%)	46 (19%)	12 (41%)	0.017
Sleep disorder	17 (25%)	47 (20%)	8 (28%)	0.4
Dizzy	37 (55%)	155 (65%)	21 (72%)	0.2
Cough	57 (85%)	205 (85%)	25 (86%)	>0.9
Short breath	34 (51%)	125 (52%)	16 (55%)	>0.9
Dyspepsia	17 (25%)	54 (23%)	7 (24%)	0.9
Palpitations	12 (18%)	46 (19%)	4 (14%)	0.8
Bluish lips	11 (16%)	36 (15%)	2 (6.9%)	0.5
Sputum				>0.9
No	62 (93%)	219 (91%)	27 (93%)	
Yes	5 (7.5%)	20 (8.3%)	2 (6.9%)	
Lung reals				2.2E-16
None	11 (16%)	55 (23%)	2 (6.9%)	
Unilateral or bilateral lung floor rales (small amount)	41 (61%)	124 (52%)	3 (10%)	
Unilateral or bilateral middle and lower lobes of the lungs (moderate amount)	15 (22%)	60 (25%)	2 (6.9%)	
Large amount of rales (full lungs)	0 (0%)	1 (0.4%)	22 (76%)	
Temperature	36.50 (36.30, 36.70)	36.50 (36.40, 36.80)	36.50 (36.50, 36.80)	0.4
Heart rate	90 (80, 100)	100 (87, 119)	100 (92, 115)	<0.001
SpO2	80 (74, 86)	74 (65, 84)	74 (63, 85)	0.007
SBP	122 (117, 133)	120 (111, 130)	120 (117, 130)	0.3
DBP	80 (74, 88)	80 (70, 87)	77 (70, 83)	0.6
CT				<0.001
Normal	7 (10%)	2 (0.8%)	0 (0%)	
Double lung/texture	10 (15%)	8 (3.3%)	2 (6.9%)	
Bipulmonary lower field inner bands/patchy/bipulmonary lower field inner bands/flocculent	50 (75%)	21 (8.8%)	7 (24%)	
Others	0 (0%)	209 (87%)	20 (69%)	

[1]Median (IQR); n (%) 2 Kruskal-Wallis rank sum test; 2 Pearson's Chi-squared test; Fisher's exact test

coefficients in Fig. 2 indicated that lungrales and HR had positive coefficients, implying that an increase in these variables raised the predictive probability of the target variable. Conversely, sputum and SpO2 had negative coefficients. Subsequently, the decision tree and random forest models were trained using these four predictors identified by LASSO, and feature importance scores were evaluated. In the decision tree model, HR, SpO2, lungrales, and sputum were ranked in descending order of importance. Figure 3 illustrates that the variable importance analysis of the random forest and decision tree models reveals four key predictors that consistently exhibit high significance. This analysis excludes variables such as age, BMI, and blood pressure, as well as other clinically relevant factors that may be impractical to collect during emergency screenings of critically ill patients, particularly in remote plateau regions.

Model validation and scoring tools

Table 3 Illustrates the training, evaluation, and comparison of multinomail multinomail logistic regression, decision tree, and random forest models using four filtered predictors on the test set. The random forest model exhibited the highest accuracy at 77.94%, along with superior precision, recall, and F1 score compared to the other models. The final set of predictors, including lung

Table 2 Clinical characteristics of the training set validation set

Characteristic	test, N = 100[1]	train, N = 236[1]	p-value[2]
Sex	76 (76%)	190 (81%)	0.4
Age	37 (26, 44)	31 (26, 42)	0.095
Height	169 (165, 172)	170 (165, 173)	0.056
Weight	72 (62, 78)	74 (65, 77)	0.4
BMI	24.6 (22.7, 26.2)	25.0 (23.1, 26.4)	0.4
Fatigue	20 (20%)	62 (26%)	0.2
Loss of taste	24 (24%)	53 (22%)	0.8
Nausea	71 (71%)	194 (82%)	0.03
Sleep disorder	20 (20%)	52 (22%)	0.7
Dizzy	70 (70%)	143 (61%)	0.1
Cough	83 (83%)	204 (86%)	0.4
Short breath	52 (52%)	123 (52%)	> 0.9
Dyspnea	25 (25%)	53 (22%)	0.6
Palpitations	21 (21%)	41 (17%)	0.4
Bluishlips	19 (19%)	30 (13%)	0.14
Sputum			0.7
no	94 (94%)	214 (91%)	
yes	6 (6.0%)	21 (8.9%)	
Lungrales			0.4
None	16 (16%)	52 (22%)	
Unilateral or bilateral lung floor rales (small amount)	55 (55%)	113 (48%)	
Unilateral or bilateral middle and lower lobes of the lungs (moderate amount)	24 (24%)	53 (22%)	
Large amount of rales (full lungs)	5 (5.0%)	18 (7.6%)	
SPO2	75 (65, 84)	76 (68, 85)	0.4
SBP	120 (115, 132)	120 (111, 130)	0.3
DBP	80 (70, 86)	80 (70, 88)	> 0.9
ct			0.5
Normal	1 (1.0%)	8 (3.4%)	
Double lung/texture	7 (7.0%)	13 (5.5%)	
Bipulmonary lower field inner bands/patchy/bipulmonary lower field inner bands/flocculent	26 (26%)	52 (22%)	
Others	66 (66%)	163 (69%)	
Group			> 0.9
Mild	20 (20%)	47 (20%)	
Moderate	72 (72%)	168 (71%)	
Sever	8 (8.0%)	21 (8.9%)	

[1] n (%); Median (IQR); 2 Pearson's Chi-squared test; Wilcoxon rank sum test; Fisher's exact test

Rales, sputum (sputum), heart rate (HR), and oxygen saturation (SpO2), was determined through this process. Following model comparison and evaluation in Fig. 4, the ROC curves of the random forest model and the decision tree model exhibit distinct performances on the training and validation sets. Figure 4B, which depicts the training set of the random forest model, indicates an average AUC of 0.95, with AUCs for each category being notably high (0.93 for class 1, 0.93 for class 2, and 0.99 for class 3). This suggests that the model demonstrates excellent classification performance on the training set. In contrast, the average AUC for Fig. 4C, representing the training set of the decision tree model, is 0.93, slightly lower than that of the random forest model. The AUCs for class 1 and class 2 are 0.91 and 0.90, respectively, while class 3 achieves an AUC of 0.97, indicating a somewhat weaker classification performance in certain categories. On the validation set, Fig. 4E (the validation set for the random forest model) records an average AUC of 0.90, reflecting good classification performance despite a slight decrease compared to the training set. The AUCs for class 1, class 2, and class 3 are 0.93, 0.92, and 0.86, respectively, demonstrating commendable generalization ability. Conversely, the average AUC for Fig. 4F (the validation set of the decision tree model) is 0.81, which represents a significant decline compared to the training set. The AUCs for class 1 and class 2 are both 0.78, while class 3 achieves an AUC of 0.87, indicating a potential tendency towards overfitting. Overall, the random forest model outperforms the decision tree model in both the training and validation sets, particularly in terms of classification performance and generalization ability, highlighting the stronger advantages of the random forest model.

Figure 4 illustrates that the Multinomail logistic regression model exhibited better stability and performance in predicting HAPE at varying severity levels within the validation set. The AUC values for mild, moderate, and severe cases were 0.86, 0.92, and 0.89, respectively. The confusion matrix of the random forest model demonstrated balanced performance across all categories, with higher prediction accuracy for mild, moderate, and severe cases. This model effectively differentiated between the different severity levels, particularly excelling in predicting moderate and severe cases. On the other hand, the decision tree model showed higher accuracy in predicting mild and moderate cases, but slightly lower accuracy in predicting severe cases compared to the random forest model. Despite being slightly less accurate than the Random Forest model, the decision tree model remained effective in predicting mild and moderate cases. In summary, both Multinomail logistic regression and random forest models consistently performed well in predicting HAPE severity, with high AUC values. The random forest model excelled in maintaining balance across all severity categories, while the decision tree model was more adept at distinguishing between mild and moderate cases, albeit with slightly less effectiveness in predicting severe cases. In conclusion, both Multinomail logistic regression and random forest models prove to be valuable tools for predicting HAPE severity.

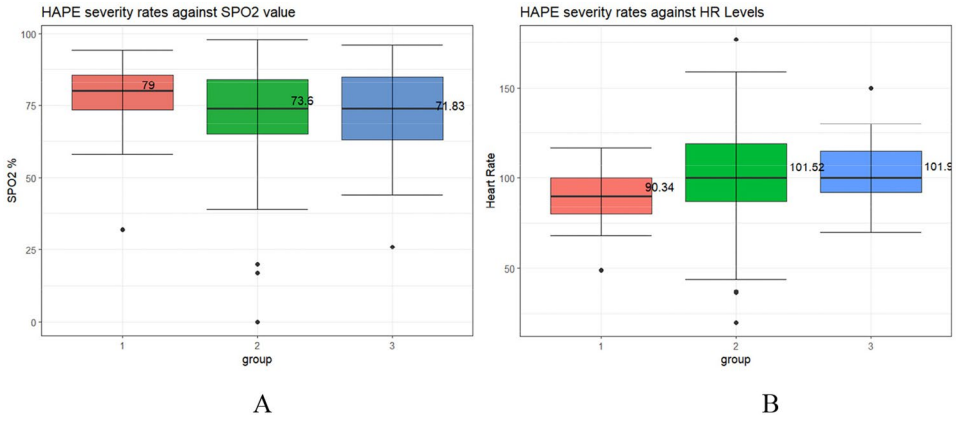


Fig. 1 (**A**) Level of heart rate across HAPE risk stratification (**B**) Level of SpO2 across HAPE risk stratification

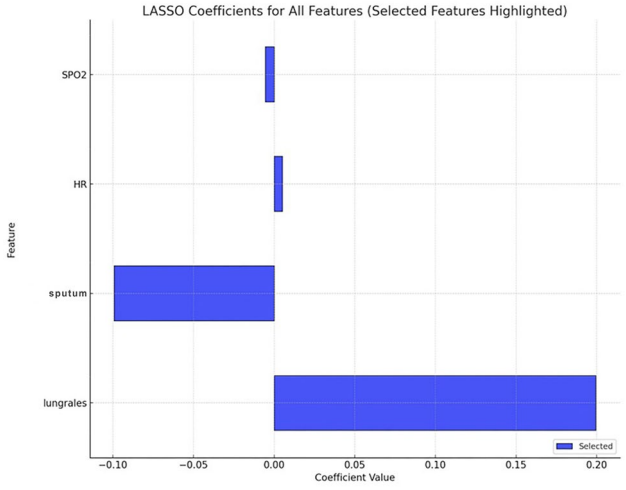


Fig. 2 LASSO coefficients for all features

Table 3 Different model performance evaluation accuracy, precision, recall and F1 score

	Multinomail logistic regression (feature selection)	Decision Tree (feature selection)	Random Forest (feature selection)
Accuracy	0.720588	0.735294	0.779412
Precision	0.535088	0.615	0.702655
Recall	0.571111	0.662222	0.682222
F1 score	0.551921	0.632641	0.689552

Clinical benefit curves for the three models (Multinomail logistic regression, decision tree, and random forest) Fig. 5 shows that the random forest model has a higher net benefit than the other models for most threshold ranges. The Multinomail logistic regression and decision tree models perform better in the low threshold range (0.1–0.3), indicating that the models have good clinical utility at these thresholds.The red dashed line (Treat All) and green dashed line (Treat None) for the Treat All and Treat None strategies represent the net benefit of the two

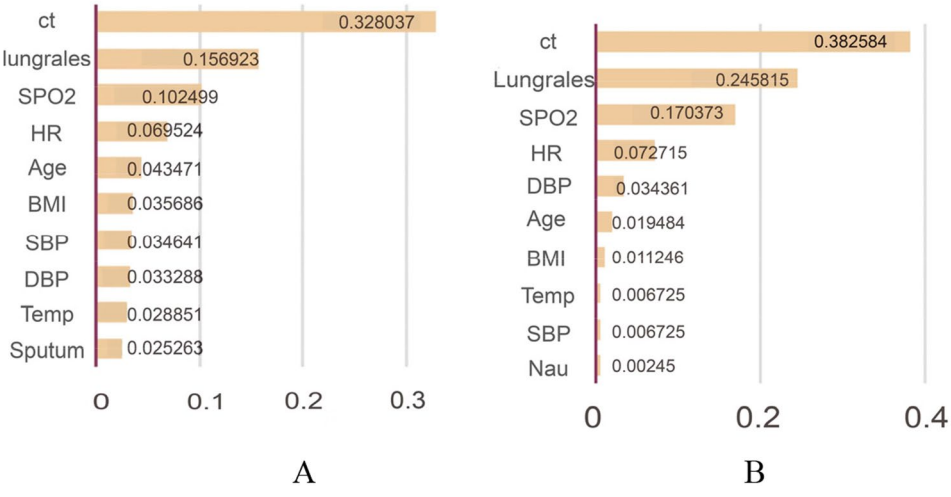


Fig. 3 (**A**) Random forest and (**B**) Decision tree importance score

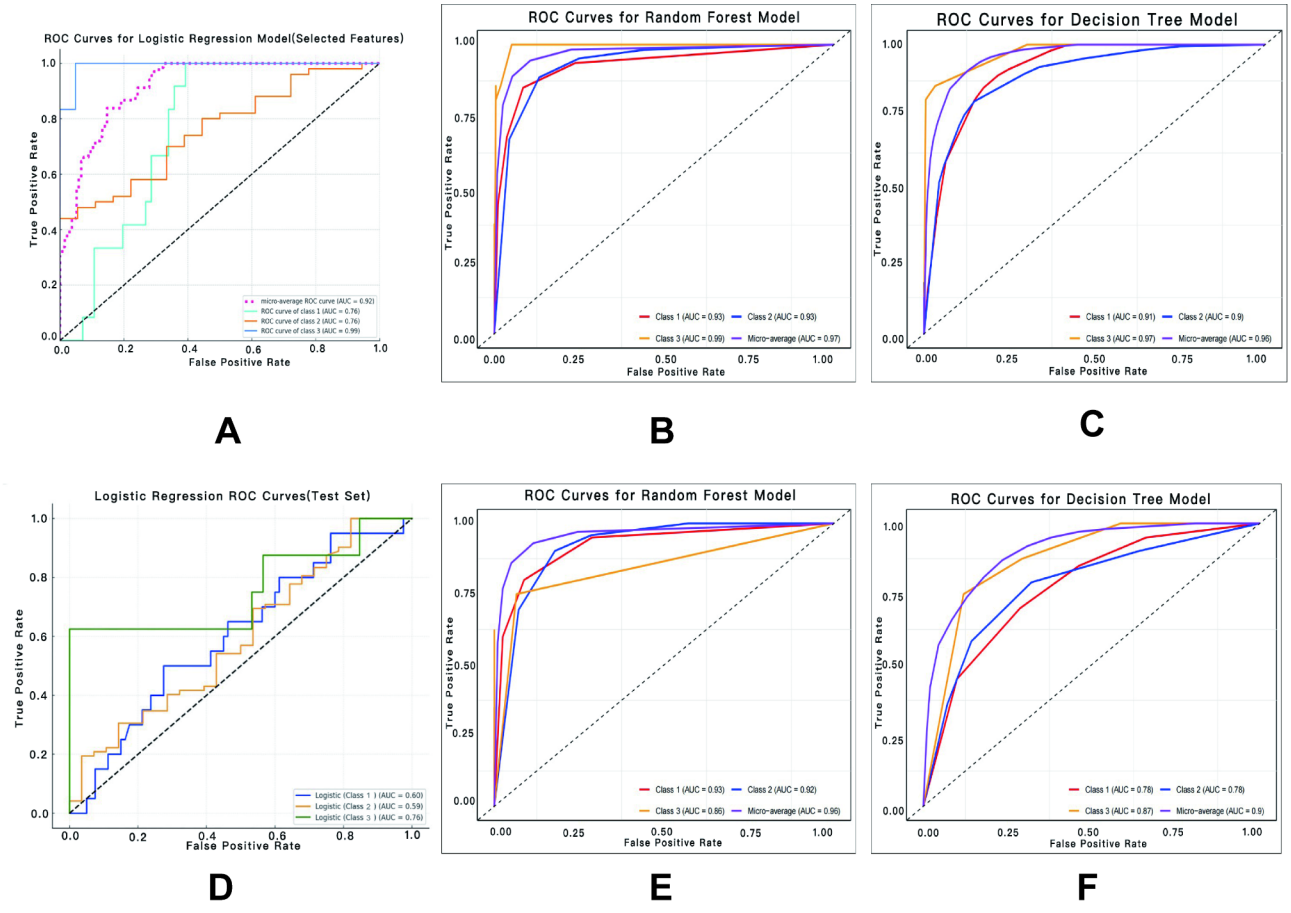


Fig. 4 (**A, B, C**) for training set; (**D, E, F**) for validation set; (**A, D**) for ROC curve of multinomail logistic regression; (**B, E**) for ROC curve of random forest model; (**C, F**) for ROC curve of decision tree model

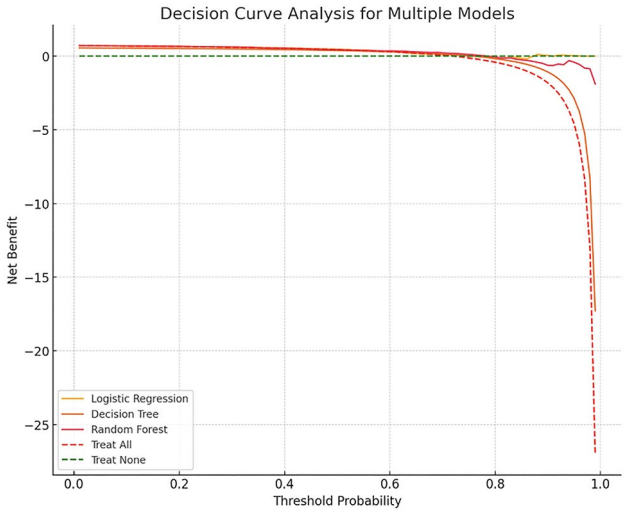


Fig. 5 Yellow curve: net benefit of the Multinomail logistic regression model. Orange curve: net benefit of the decision tree model. Green curve: net benefit of the random forest model. Red dashed line: net benefit assuming all patients receive treatment (Treat All). Green dashed line: net benefit assuming all patients receive no treatment (Treat None)

extreme strategies. The Random Forest model maintains a higher net benefit at different thresholds. The multinomial logistic regression model performs well within the lower threshold range (0.1–0.3). However, since the AUC during the validation evaluation was lower than that of the other two machine learning models, it was not considered for further analysis. The decision tree model's net benefit curve demonstrates superior performance in the lower threshold range. The final scores for each predictor, as presented in Table 4, were derived from the Random Forest algorithm. Scores below 30 points were classified as mild disease (level 1), scores between 30 and 49 points were classified as moderate disease (level 2), and scores of 50 points and above were classified as severe disease (level 3).

SHAP analysis results

In our analysis, we employed SHAP (Shapley Additive Explanations) values to interpret the feature importance of the variables in our model. The SHAP values offer a quantitative measure of the effect each feature has on the predictions, emphasizing both the magnitude and direction of their influence. The figure presented below

Table 4 Random forest scoring form

		Score		Score		Score		Score
Lungrales	Normal	0	Small amount of crackles in one or both lung bases	9	Medium amount of crackles in unilateral or bilateral lower and middle lobes of the lungs	18	Lots of crackles in full lungs	27
Sputuming sputum	None	7	Infrequent	7	Frequent	13		
HR	20–59	0	60–99	8	100–139	16	140–177	24
SpO_2	0–24	0	25–49	7	50–74	14	75–98	21

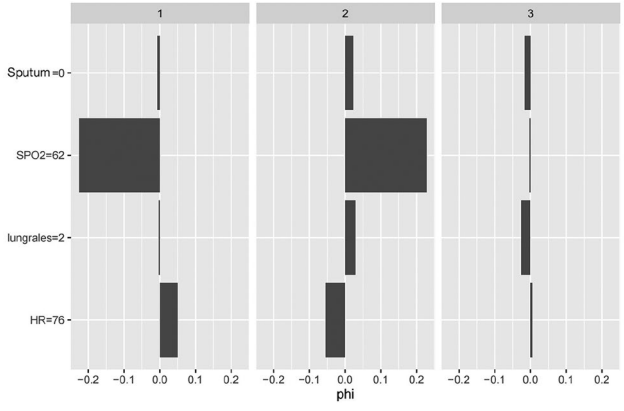


Fig. 6 SHAP analysis in HAPE severity prediction

illustrates the SHAP values for each feature, with variables such as 'SPO2', 'HR', and 'lungales' demonstrating significant variation in their contributions across different instances within the dataset. Figure 6.

Discussion

The study aimed to develop a machine-learning based risk stratification tool for High Altitude Pulmonary Edema (HAPE) to assist clinicians in accurately screening severe patients in emergency care. Key predictors identified include lung rales, sputum production, heart rate, and blood oxygen saturation. Modeling was conducted to assess the stability and performance of Multinomial logistic regression and random forest models in determining HAPE severity. The goal is to stratify severity levels, accurately screen severe patients diagnosed with HAPE, and facilitate prompt and effective treatment. In the process of predictor selection, we excluded CT (Computed Tomography) imaging as a potential predictor. This decision was based on several practical and clinical considerations. Firstly, CT scans are often unavailable in many high-altitude regions with limited medical resources. These areas typically lack the advanced imaging equipment required for CT examinations, making it impractical to rely on CT as a routine diagnostic tool in such settings [16]. Secondly, acute high-altitude pulmonary edema (HAPE) is characterized by its rapid onset and progression. In emergency situations, transporting patients to larger hospitals for CT scans can significantly delay diagnosis and treatment, potentially exacerbating the condition [1]. Therefore, considering the accessibility and urgency of care in high-altitude environments, we chose to exclude CT from our list of predictors to ensure the practicality and feasibility of our risk stratification tool. Despite the exclusion of CT, our study successfully identified several clinically relevant predictors that can be readily assessed in emergency settings. These predictors, including lung rales, sputum production, heart rate, and blood oxygen saturation, provide a robust foundation for our machine-learning models. Our findings highlight the importance of these easily accessible clinical signs and measurements in accurately stratifying HAPE severity. In conclusion, our machine-learning based risk stratification tool offers a practical and effective solution for identifying severe cases of HAPE in high-altitude regions. By focusing on readily available clinical data, we aim to facilitate prompt and accurate diagnosis, ultimately improving patient outcomes in emergency care settings.

Four predictors on high altitude pulmonary edema to predict disease severity

Pulmonary rales are a common clinical manifestation in patients with High Altitude Pulmonary Edema (HAPE), and their severity is closely linked to disease progression. Research indicates that the presence and intensity of pulmonary rales in patients vary significantly based on the severity of HAPE [15, 17]. Our study identified lung rales as a crucial predictor and a key feature utilize both the LASSO model and the random forest model. This finding suggests that lung rales are not only a typical symptom of HAPE but also exhibit high sensitivity and specificity in predicting disease severity. sputuming up sputum is another significant clinical symptom in HAPE patients, with its presence or absence reflecting the extent of damage to the respiratory tract [1]. Our study revealed significant differences in the incidence of sputum production among HAPE patients of varying severity, particularly higher in those with severe HAPE. This finding aligns with previous studies, indicating that sputum production can serve as a reliable predictor of HAPE severity. Heart rate, as a crucial physiological parameter reflecting the

condition of the cardiovascular system, tends to be notably elevated in patients with HAPE, especially in critical conditions. Our modeling analysis further confirmed the significance of heart rate in predicting HAPE [5, 19]. Heart rate consistently emerged as a crucial predictor in both Random Forest and Decision Tree models, as indicated by feature importance rankings. This underscores the value of monitoring heart rate for gaining insights into a patient's condition and facilitating early detection of disease progression. Additionally, oxygen saturation serves as a vital indicator for assessing hypoxia and oxygen levels in the body [2, 9, 11, 18]. Our findings highlight a notable decline in SpO2 as a key characteristic among patients with severe HAPE. Both Random Forest and Multinomail logistic regression models highlighted the importance of SpO2 as a predictor, aligning with existing literature. Furthermore, we conducted stratification of SpO2 levels among HAPE patients with varying severities.

Models on grading the severity of high altitude pulmonary edema disease

In this study, three machine learning models, Multinomail logistic regression, random forest, and decision tree, were selected for predicting the severity of high-altitude pulmonary edema (HAPE). Each model possesses distinct strengths and weaknesses, performing variably in different scenarios. Multinomail logistic regression models are valued for their simplicity and interpretability, commonly used in medical prediction research due to their explicit coefficients, which elucidate the impact of individual predictors [14]. The Multinomail logistic regression model utilized in this study demonstrated strong performance in predicting HAPE severity, achieving a high area under the curve (AUC) value of 0.82. However, Multinomail logistic regression models have limitations in handling nonlinear relationships and complex interaction effects. On the other hand, the random forest model, an ensemble learning method that combines multiple decision trees, showed superior performance in HAPE severity prediction in this study. It delivered higher accuracy, precision, recall, and F1 score compared to the other models.The random forest model demonstrated exceptional capability in managing a large number of features and effectively handling interactions and nonlinear relationships among them, resulting in the highest AUC value of 0.86 in this study, particularly in the prediction of moderate and severe HAPE patients. Decision tree models, operating through a tree structure, offer robust interpretive and visualization features. While the performance of the decision tree model in this study was not as strong as that of the random forest model, it still outperformed the Multinomail logistic regression model. The decision tree model provides a clear decision path for each predictor, aiding clinicians in understanding the specific contribution of each factor to the prediction. However, a single decision tree is susceptible to overfitting and does not perform as well as the random forest model. Furthermore, incorporating SHAP values into our model has yielded a nuanced understanding of feature importance [4, 13].

SHAP values stand out from conventional feature importance metrics by providing a detailed and interpretable breakdown of each feature's contribution to individual predictions. For instance, the SHAP analysis of 'SPO2' consistently reveals its significant impact on the model's output, indicating that reduced oxygen saturation levels are likely to substantially affect patient outcomes. This insight not only enhances the transparency of our model but also has practical implications for guiding clinical decision-making, especially in scenarios where comprehending the rationale behind each prediction is essential.

Limitations and future of the study

This study was conducted at a single healthcare facility, which limits the generalizability of the results. The retrospective study design may have led to incomplete or inaccurate data recording, potentially affecting the accuracy of the findings. Retrospective studies inherently face challenges in controlling for confounding variables, with the possibility of unmeasured factors influencing outcomes. The assessment of rales is largely dependent on the subjective judgment of the clinician's experience, and lacks quantification. Furthermore, external validation in independent samples or different centers is absent, highlighting the need for further verification in larger studies and diverse settings to ensure reliability. While the patients included in this study underwent appropriate laboratory tests, the focus was on stratifying high-altitude pulmonary edema (HAPE) with acute onset based on symptoms and signs in a timely manner. Laboratory indicators were not considered in this study, and it is recommended that future research on hospitalized patients with HAPE explore this aspect more thoroughly. Additionally, due to the unbalanced dataset generated during real-world sample collection, the second category of the target variable contained significantly more patients than the other two categories. To improve the results in future studies, it is crucial to either increase the sample size or implement additional reweighting methods.

In future research, we will focus on the following areas: (1) Prospective Studies: We aim to conduct prospective studies to address the limitations of retrospective data, thereby ensuring more accurate and comprehensive data collection. (2) Incorporation of Laboratory Data: Future studies will include laboratory indicators such as arterial blood gases, complete blood counts, and inflammatory markers to enhance the predictive accuracy of the model.

(3) Development of a Clinical Decision Support System: We will integrate the validated model into a clinical decision support system to provide clinicians with real-time risk assessments and treatment recommendations.

Conclusion

This study further confirmed that lungrales, heart rate, oxygen saturation, and sputum sputum are key predictors for urgently determining the severity of High Altitude Pulmonary Edema (HAPE) by systematically analyzing the data of HAPE patients. A model for identifying the severity of HAPE was established and validated for the first time, providing a scientific basis for early diagnosis and intervention, as well as a valuable reference tool for clinicians. Future studies could delve into the specific mechanisms of these factors to enhance the diagnosis and treatment of HAPE, ultimately reducing mortality rates attributed to diagnostic delays.

Acknowledgements
Not applicable.

Author contributions
L.G designed the study, provided scientific guidance, made modifications to the manuscript, and approved the report. Y.S conducted data analysis, wrote the initial draft of the manuscript, and served as the guarantor for the overall content. Z.D and W.C assisted with data analysis and provided clinical background knowledge guidance. Z.B performed the proofreading of the chart format, Q.G, Q.D, and Q.Z assisted with data collection and were actively involved in the preparation.

Funding
This study was supported by the Science and Technology Department of the Tibet Support program (NO. XZ2021JR0004G). The funders had no role in the study design, data collection, data analysis, and interpretation, as well as the writing of the manuscript.

Data availability
Data used in this study are available upon reasonable request. To access the data, please contact the corresponding author at KelsangNorbu@hotmail.com. When requesting the data, kindly include a brief description of your research purpose and how you plan to use the data. Requests will be reviewed to ensure alignment with our data sharing policy and ethical considerations. Approval for data access is required, with a focus on legitimate research purposes and respect for privacy and confidentiality. We support open and collaborative scientific efforts and thank you for your interest in utilizing our data for research purposes.

Declarations

Ethics approval and consent to participate
This study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. The research utilized retrospective data collected from the hospital's inpatient system, including anonymized electronic medical records from previous years. As the study involved no direct participant involvement and used anonymized data, informed consent was deemed unnecessary. This determination was made in accordance with the national regulations outlined in the Measures for Ethical Review of Biomedical Research Involving Human Beings (2016), which stipulate that informed consent may be waived under certain conditions, such as when the research involves anonymized data and poses no risk to subjects. The study was approved by the Medical Ethics Committee of Tibet Autonomous Region People's Hospital (Ethical Approval No. ME-TBHP-22-KJ-064). The committee determined that the research content does not constitute harm or risk to subjects and that the confidentiality of the data is maximized to ensure the privacy of the relevant observed persons.

Consent for publication
Not applicable.

Relevant guidelines and regulations
This study adhered to all relevant guidelines and regulations for medical research involving human data, including the protection of privacy and confidentiality of personal information.

Competing interests
The authors declare no competing interests.

Author details
[1]High Altitude Medical Research Institute of Tibet Autonomous Region, 18 Linkuo North Road, Lhasa 850000, China
[2]Key Laboratory of Transitional Medicine for Human Adaptation to the High-Altitude of Tibet Autonomous Region, Tibet Autonomous Region People's Hospital, Lhasa 850000, China
[3]Tibet Autonomous Region Clinical Research Center for High Altitude Diseases, Tibet Autonomous Region People's Hospital, Lhasa 850000, China
[4]Tibet Autonomous Region People's Hospital, Lhasa 850000, China

Received: 23 October 2024 / Accepted: 1 April 2025
Published online: 18 April 2025

References

1. Bärtsch P, Swenson ER. Acute high-altitude illnesses. N Engl J Med. 2013;368(24):2294–302. https://doi.org/10.1056/NEJMcp1214870.
2. Berger MM, Sareban M, Schiefer LM, Swenson KE, Treff F, Schäfer L, Schmidt P, Schimke MM, Paar M, Niebauer J, Cogo A, Kriemler S, Schwery S, Pickerodt PA, Mayer B, Bärtsch P, Swenson ER. Effects of Acetazolamide on pulmonary artery pressure and prevention of high-altitude pulmonary edema after rapid active ascent to 4,559 m. J Appl Physiol (1985). 2022;132(6):1361–9. https://doi.org/10.1152/japplphysiol.00806.2021.
3. Burtscher M, Hefti U, Hefti JP. High-altitude illnesses: old stories and new insights into the pathophysiology, treatment and prevention. Sports Med Health Sci. 2021;3(2):59–69.
4. Duan J, Li H, Ma X, Zhang H, Lasky R, Monaghan CK, Chaudhuri S, Usvyat LA, Gu M, Guo W, Kotanko P, Wang Y. Predicting SARS-CoV-2 infection among Hemodialysis patients using multimodal data. Front Nephrol. 2023;3:1179342. https://doi.org/10.3389/fneph.2023.1179342.
5. Hackett PH, Rennie D, Levine HD. The incidence, importance, and prophylaxis of acute mountain sickness. Lancet. 1976;2(7996):1149–55. https://doi.org/10.1016/s0140-6736(76)91677-9.
6. Jensen JD, Vincent AL. High altitude pulmonary edema. StatPearls. StatPearls Publishing LLC; 2024. StatPearls Publishing Copyright © 2024.
7. Luks AM, Ainslie PN, Lawley JS, Roach RC, Simonson TS. High altitude medicine and physiology (6th ed.) 2021. https://doi.org/10.1201/9780429444333.
8. Luks AM, Hackett PH. Medical conditions and high-altitude travel. N Engl J Med. 2022;386(4):364–73.
9. Luks AM, McIntosh SE, Grissom CK, Auerbach PS, Rodway GW, Schoene RB, Zafren K, Hackett PH. Wilderness medical society practice guidelines for the prevention and treatment of acute altitude illness: 2014 update. Wilderness Environ Med. 2014;25(4suppl):S4–14.
10 Luo C, Zhu Y, Zhu Z, Li R, Chen G, Wang Z. A machine learning-based risk stratification tool for in-hospital mortality of intensive care unit patients with heart failure. J Transl Med. 2022;20(1):136. https://doi.org/10.1186/s12967-022-03340-8.
11. Maa EH. Hypobaric hypoxic cerebral insults: the neurological consequences of going higher. NeuroRehabilitation. 2010;26(1):73–84. https://doi.org/10.3233/nre-2010-0537.
12. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform. 2018;19(6):1236–46.
13. Rai T, Shen Y, He J, Mahmud M, Brown DJ, Kaur J, O'Dowd E, Baldwin DR, Hubbard R. (2024). Understanding Feature Importance of Prediction Models Based on Lung Cancer Primary Care Data. In 2024 International Joint

Conference on Neural Networks (IJCNN) (pp. 1–8). IEEE. https://doi.org/10.1109/IJCNN60899.2024.10650819
14. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med. 2019;380(14):1347–58.
15. Richalet JP, Larmignat P, Poitrine E, Letournel M, Canouï-Poitrine F. Physiological risk factors for severe high-altitude illness: a prospective cohort study. Am J Respir Crit Care Med. 2012;185(2):192–8. https://doi.org/10.1164/rccm.201108-1396OC.
16. Roach R, Hackett P. Frontiers in high-altitude medicine. In: Hornbein TF, Schoene RB, editors. High altitude: an exploration of human adaptation. Marcel Dekker; 2001. pp. 363–76.
17. Singh I, Kapila CC, Khanna PK, Nanda RB, Rao BD. HIGH-ALTITUDE PULMONARY OEDEMA. Lancet. 1965;1(7379):229–34. https://doi.org/10.1016/s0140-6736(65)91520-5.
18. Suona Y, Gesang L, Danzeng Z, Ci B, Zhaxi Q, Huang J, Zhang R. Predictive model for estimating the risk of high-altitude pulmonary edema: a single-centre retrospective outcome-reporting study. BMJ Open. 2023;13(11):e074161.
19. Swenson ER. (2011). High-altitude pulmonary edema. Textbook Pulmonary Vascular Disease, 871–88.
20. UNWTO. (2024). Mountain Tourism. UNWTO. https://www.unwto.org/mountain-tourism
21. Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. Ann Transl Med. 2016;4(2):30. https://doi.org/10.3978/j.issn.2305-5839.2015.12.63.

Publisher's note