


RESEARCH

Open Access



Machine learning-based evaluation of prognostic factors for mortality and relapse in patients with acute lymphoblastic leukemia: a comparative simulation study

Zahra Mehrbakhsh^{1,2}, Roghayeh Hassanzadeh^{1,2}, Nasser Behnampour³, Leili Tapak^{1,4*} , Ziba Zarrin⁵, Salman Khazaeli⁶ and Irina Dinu⁷

Abstract

Background Predicting mortality and relapse in children with acute lymphoblastic leukemia (ALL) is crucial for effective treatment and follow-up management. ALL is a common and deadly childhood cancer that often relapses after remission. In this study, we aimed to apply and evaluate machine learning-based models for predicting mortality and relapse in pediatric ALL patients.

Methods This retrospective cohort study was conducted on 161 children aged less than 16 years with ALL. Survival status (dead/alive) and patient experience of relapse (yes/no) were considered as the outcome variables. Ten machine learning (ML) algorithms were used to predict mortality and relapse. The performance of the algorithms was evaluated by cross-validation and reported as mean sensitivity, specificity, accuracy and area under the curve (AUC). Finally, prognostic factors were identified based on the best algorithms.

Results The mean accuracy of the ML algorithms for prediction of patient mortality ranged from 64 to 74% and for prediction of relapse, it varied from 64 to 84% on test data sets. The mean AUC of the ML algorithms for mortality and relapse was above 64%. The most important prognostic factors for predicting both mortality and relapse were identified as age at diagnosis, hemoglobin and platelets. In addition, significant prognostic factors for predicting mortality included clinical side effects such as splenomegaly, hepatomegaly and lymphadenopathy.

Conclusions Our results showed that artificial neural networks and bagging algorithms outperformed other algorithms in predicting mortality, while boosting and random forest algorithms excelled in predicting relapse in ALL patients across all criteria. These results offer significant clinical insights into the prognostic factors for children with ALL, which can inform treatment decisions and improve patient outcomes.

Keywords Acute lymphoblastic leukemia, Classification, Data mining, Machine learning, Relapse, Mortality

*Correspondence:
Leili Tapak
l.tapak@umsha.ac.ir

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



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Children (aged zero to 14) make up about a quarter of the world's population [1]. There are a variety of causes that can contribute to children's deaths, including infectious diseases, noncommunicable diseases, injuries, and difficulties during childbirth [2]. Cancer as a non-communicable disease is the second leading cause of death in children. The most common cancers diagnosed in children under 15 years of age are leukemia, brain tumors and other central nervous system tumors, and lymphomas. Leukemia is the most common childhood malignancy which constitute about 30% of all cancers diagnosed in children under 15 years of age. It is characterized by abnormal growth of immature white blood cells and their substrates in the blood and bone marrow [3]. Leukemia patients can experience death from a variety of causes, including relapse, treatment complications and serious infections [4, 5]. Leukemia was divided into four main types: Acute Lymphoblastic Leukemia (ALL), Acute Myeloid Leukemia (AML), Chronic Lymphocytic Leukemia (CLL) and Chronic Myeloid Leukemia (CML) [3]. Among them, ALL is the most common type of leukemia, accounting for about 80% of cases [6, 7].

Despite the fact that early detection and prompt initiation of treatment can significantly increase the chances of cure and survival for 90% of children, ALL is still one of the leading causes of death in childhood. Also, relapse occurs in 15 to 20% of children with ALL, and cure rates after relapse are much lower [8]. Therefore, it causes not only the death of children, but also high diagnostic and treatment costs for families and health systems [9, 10].

Accurate classification of childhood ALL patients into appropriate risk groups is a critical, but challenging component of treatment management. Early identification of relevant outcomes is critical for tailoring chemotherapeutic treatment and improving patient outcomes, and accurate diagnosis is necessary for selecting the appropriate treatment modality and planning patient care [11]. Recently, machine learning (ML) techniques have been extended to medical applications that enable the detection of complex patterns that can aid in the diagnosis, treatment, outcome prediction, and prevention of disease [12]. These techniques use algorithms that can automatically learn and improve based on experience, without being explicitly programmed [13]. In addition, these methods can account for high-dimensional, complex and nonlinear relationships between prognostic factors and make more accurate predictions in various domains that are not possible with traditional statistical methods [14]. In general, ML methods are nonparametric methods that need no distributional assumptions and can be split into two main categories, (i) supervised learning (ii) unsupervised learning. In supervised learning, the machine learns patterns based on input and output data, while

in unsupervised learning, it discovers patterns without labels. These methods were used for solving problems of classification and regression. In classification, the output variable has class labels, and in regression, the output variable has continuous values [15, 16].

ML algorithms have recently been successfully applied in various medical fields, such as skin cancer classification [17], cancer detection [18], cardiovascular disease prediction [19], COVID-19 mortality prediction [20], traumatic injury mortality prediction [21] and prediction of the transition from pre-diabetes to type 2 diabetes [22]. In cancer care, ML methods can be trained on patient data to identify individuals at high risk of cancer relapse or progression. These algorithms can also help determine the most effective treatment options for each patient by taking into account their individual genetic predispositions and medical history [23].

According to a review study by Cruz and Wishart, ML methods have shown significant potential to improve the accuracy of cancer susceptibility, relapse, and survival prediction, such that their application has led to a 15–20% improvement in cancer prediction accuracy in recent years [24]. This highlights the potential of using ML algorithms to improve cancer care and outcomes through more accurate and personalized predictions for individual patients. In addition, ML methods have been utilized to predict treatment outcomes using gene expression data, and medical images, emphasizing the importance of these methods in predicting cancer [25]. Hence, applying ML methods can assist clinicians make better decision regarding patient care.

In recent years, different types of ML methods have been employed for solving classification problems in medical research, particularly childhood cancer domain, and especially leukemia [26]. Most of the previous related works have widely used gene expression, molecular, blood smear image datasets for diagnosing acute leukemia or classifying acute leukemia subtype, outcome and relapse prediction [27–30]. Moreover, some review articles have also investigated studies that utilized ML methods for leukemia detection using molecular and image data [31, 32]. However, information regarding gene expression, molecular, blood smear image datasets is not available for all settings, especially in lower income countries and the usual medical records include only some demographic and clinical variables. A number of studies have used prediction models to diagnose leukemia and predict the occurrence of death and relapse. Nevertheless, a limited number of studies have utilized ML methods for death mortality and relapse prediction in ALL patients on the basis of laboratory and clinical data. For example, Kashef et al. used several algorithms such as Decision Tree (DT), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Multinomial Linear Regression (MLR),

Gradient Boosting Machine (GBM), Random Forest (RF) and XG-Boost methods to predict the treatment outcome (dead or survived) of pediatric ALL patients [4]. Also, Pan et al. have used several models such as DT, RF, SVM and Logistic Regression (LR) to predict ALL childhood relapse [33]. Given the importance of ALL disease, identifying the prognostic factors that impact treatment outcomes, mortality and relapse rates is critical. From this point of view, this information can be crucial for oncologists and clinicians as it allows them to accurately predict patient outcomes and effectively treat the disease. Therefore, the main objective of this study was to comprehensively compare the performance of different machine learning algorithms for predicting mortality and relapse in patients with ALL, considering clinical and laboratory data. Furthermore, important prognostic factors influencing mortality and relapse in children with ALL were identified.

It should be noted that an important aspect in determining the prognostic factors is their stability in the models selected by cross-validation. Cross-validation with only one sampling is not a reliable foundation for decision-making. Instead, it is recommended to use cross-validation many times and calculate the mean accuracy of each fold. This approach can facilitate variable selection, with the process being iterated until the set of features with the highest mean accuracy is identified as the optimal one. By embracing this methodology, the model can be evaluated with greater precision and the results will exhibit greater stability [34, 35]. Therefore, in this study we investigated the performance of the ML methods according to several criteria and reported the corresponding means and standard deviations over 100 iterations.

Materials and methods

Data collection

The present retrospective cohort study was conducted on 161 children under 16 years of age diagnosed with ALL. These patients were referred to the Taleghani Children's Educational and Therapeutic Hospital in the city of Gorgan, northern Iran, from September 1997 to September 2016 and followed up until June 2021. The study was based on data collected from patients' medical records. This data was carefully selected based on a predefined checklist to ensure inclusion of relevant baseline demographic and clinical information as follows:

Outcomes of interest: In this study, we focused on the information on survival/mortality status (dead/alive) and relapse (yes/no) as outcome variables and associated prognostic factors with ALL.

Prognostic factors / Predictors: To predict the outcome variables, the values of 15 prognostic factors were recorded during the follow-up period and divided into three parts: demographic characteristics, laboratory information, and clinical side effects. Table 1 provides an overview of all characteristics included in the dataset with their types and values. The process of modelling is also illustrated in Fig. 1.

Machine learning methods

Logistic regression

LR is an important and common method used for categorical responses. This statistical technique predicts response values without assuming a normal distribution for the response or the predictor variables. LR is a generalized linear model (GLM) approach for modeling binary responses. In this method, the logit of the conditional probability of the dependent variable (death/ alive

Table 1 Specification of the features considered in the collected dataset

Variable	Type	Value
Demographic Characteristics		
Gender	Binary	"1" = Girl, "0" = Boy
Age at diagnosis time (year)	Numeric	Min = 0.6, Max = 15.9
Laboratory Information		
White Blood Cells count at diagnosis (cells/mL)	Numeric	Min = 100, Max = 780,000
Hemoglobin (g/dL)	Numeric	Min = 2.3, Max = 14.2
Platelets (cells/mL)	Numeric	Min = 2000, Max = 873,000
Clinical Side Effects		
Central Nervous System involvement	Binary	"1" = Yes, "0" = NO
Mediastinal Tumor	Binary	"1" = Yes, "0" = NO
Splenomegaly	Binary	"1" = Yes, "0" = NO
lymphadenopathy	Binary	"1" = Yes, "0" = NO
Hepatomegaly	Binary	"1" = Yes, "0" = NO
Rheumatoid arthritis signs	Binary	"1" = Yes, "0" = NO
Relapse	Binary	"1" = Yes, "0" = NO
Survival Status	Binary	"1" = Dead, "0" = Alive
Time to death (month)	Numeric	Min = 0.1, Max = 202.9

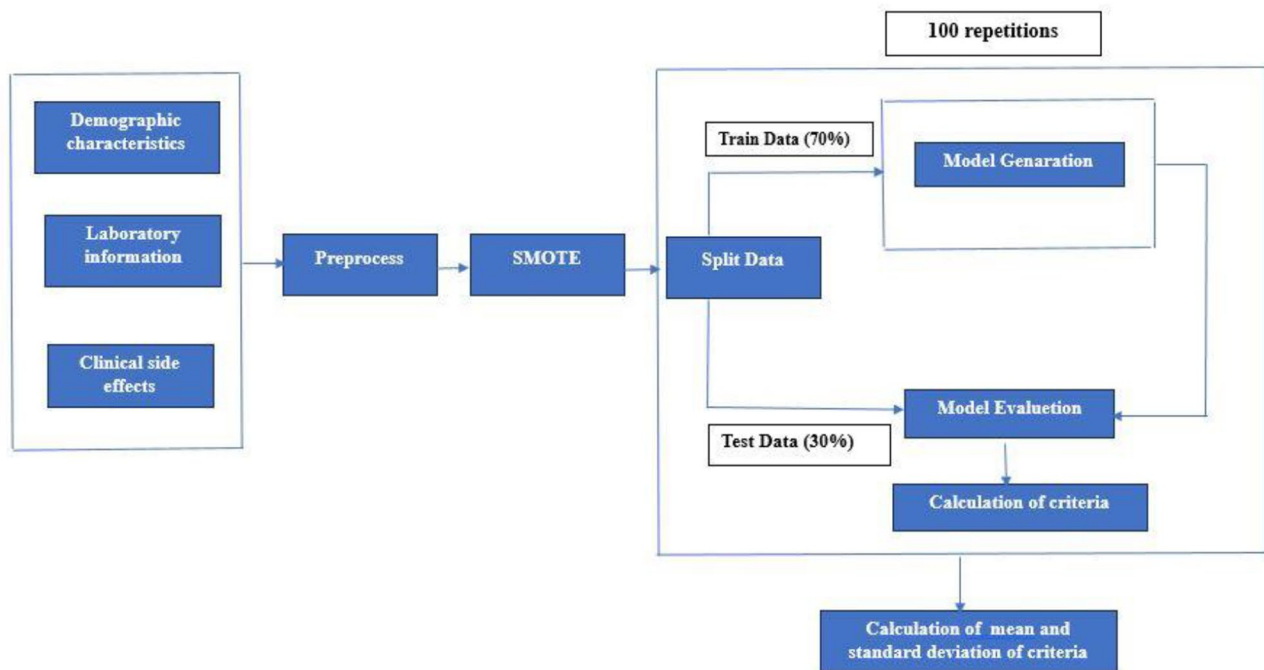


Fig. 1 Classification model building process

for survival status and yes/no for relapse) is formulated as a linear function of the independent variables [36].

Decision tree

DT algorithms are one of the most common non-parametric algorithms used in classification. This method introduces a classification pattern for the observations that has a simple and understandable structure for the decision-making process [20, 37]. The DT is a simple and powerful method for classifying a set of data into distinct and homogeneous categories, which has a tree-like graph. This tree is formed by a set of questions, where each question represents a predictor variable.

The DT has two approaches, a classification tree if the dependent variable is categorical, and a regression tree if the dependent variable is continuous [37]. The DT consists of three main components: a root, internal nodes and external nodes (leaves). In the process of building a tree, a predictor variable is first selected as the root and it is divided into several internal nodes according to a number of characteristics [20, 38]. Classification algorithms aim to find the optimal partition among all possible partitions based on various criteria. Tree algorithms minimize the heterogeneity in the nodes, which can be measured using impurity criteria, such as the widely used Gini index. This process is repeated until the dataset is divided into a number of unique groups [39].

Random forest

The RF, which belongs to the family of ensemble methods, is a popular supervised learning algorithm used for classification and regression problems [14].

The RF algorithm generates a large number of trees based on the re-sampling of the training data and averages the results of these trees to predict an outcome. This technique controls over-fitting and improves accuracy [40]. The classification error rate of the RF, the so-called Out-of-Bag (OOB) error, is estimated by considering all excluded samples by bootstrapping samples [14]. The RF can be used as a variable selection approach for the identification of informative variables. A variable's importance measures the relationship between a variable and the classification outcome. Mean Decrease Gini and Mean Decrease Accuracy can be used to find the most important predictors. This can lead to more accurate predictions of binary outcomes [14, 41].

Support vector machine

The SVM is a widely used supervised ML algorithm for classification and regression that separates binary labeled training data using a hyperplane in a high-dimensional space. It has an excellent ability to solve non-linear and high-dimensional problems and provides efficient solutions to classification problems without making assumptions about the data distribution. However, to achieve optimal results, SVM requires careful parameter selection, particularly with regard to kernel function choice.

SVM is preferred due to its significant accuracy and low computational requirements [40, 42].

Least square support vector machine

The Least square support vector machine (LS-SVM), an alternative to SVM, has been developed to address a significant limitation of SVM. Specifically, while SVM is capable of accurately approximating non-linear relationships between input and output variables, it requires the solution of a large quadratic programming problem, which can be computationally burdensome. In contrast, LS-SVM solves linear equations instead of quadratic programming problems, thereby reducing the complexity of the optimization process. Consequently, LS-SVM is a valuable tool for addressing problems related to non-linear classification and regression [42, 43].

Artificial neural network

The Artificial Neural Network (ANN) is a machine learning technique that is inspired by the neurons of the human brain. The ANN is commonly utilized for classification and pattern recognition assignments. The ANN has the ability to identify patterns and relationships in data, and it is capable of learning from experience. A multilayer feed-forward ANN comprises an input layer, one or more hidden layers, and an output layer. Neurons in adjacent layers are fully connected and are given weights associated with their connections. Information flows unidirectionally from input to output through the hidden layers. Activation functions facilitate complex non-linear mappings between input and output [44, 45]. The most important prognostic factors can be found in neural networks using Garson's algorithm [46].

Naïve bayes

The Naïve bayes (NB) algorithm, which utilizes Bayes' theorem and assumes strong independence between predictor variables, is a simple algorithm for quickly categorizing samples. This is achieved by calculating the probability that an object belongs to a specific category based on both the prior and posterior probabilities. An important advantage of the NB method is its ease of implementation, strong performance, and its ability to generate probabilistic predictions with minimal training data. Moreover, it is robust to correlated variables and can produce reliable results even when the independence assumption is not met. However, it can be computationally expensive for models with many variables [44, 45].

Bagging

Bagging is an ensemble learning technique that combines bootstrapping and aggregation methods. It selects B bootstrap samples from the training set and reduces noisy observations. As a result, the generated classifiers

show superior performance compared to the original set. Consequently, bagging is a valuable tool for developing classifiers that can handle noisy observations. These B-classifiers leads to better performance than using individual classifiers [44, 47].

Boosting

Boosting is a powerful ensemble learning technique that combines multiple weak models or base models to create a stronger model and make more accurate predictions. In other words, a set of weak models is first trained using modified versions of the original data. Modifications are made to the weights or features of the data to highlight examples that were misclassified by the previous models or where correct classification was difficult. Then, by combining the predictions of these weak models, a strong model is created that can make more accurate predictions than any single weak model. Boosting algorithm is effective in processing complex data sets, dealing with noisy or missing data, and reducing the impact of individual models' weaknesses. AdaBoost, Gradient Boost, and XG-Boost are three popular boosting algorithms. In this study, we used AdaBoost to predict relapse and survival [48].

Linear discriminant analysis

LDA employs multiple independent variables to classify observations into predetermined groups. Among the various forms of discriminant analysis, LDA is a widely utilized approach that uses a linear amalgam of independent variables to optimize the intergroup ratio to intragroup changes in discriminant scores [49]. LDA uses the conditional probability of predictors given the outcome class to solve the problem. This approach serves to minimize the dispersion between cases in the same category and to maximize the dispersion between cases in dissimilar categories [49, 50].

Performance Criteria

The discrimination power of ML methods was evaluated using several criteria, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy and area under the curve (AUC), and calculated as follows:

$$\begin{aligned} \text{sensitivity} &= \frac{TP}{TP+FN}, \quad \text{specificity} = \frac{TN}{TN+FP}, \\ \text{PPV} &= \frac{TP}{TP+FP}, \quad \text{NPV} = \frac{TN}{TN+FN}, \\ \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN}. \end{aligned}$$

where TP (true positive) stands for dead/relapsed pediatric patients with leukemia that were correctly diagnosed as dead/relapsed, TN (true negative) stands for alive/non-relapsed pediatric patients with leukemia that were correctly identified as alive/non-relapsed, FP (false

positive) stands for alive/non-relapsed pediatric patients with leukemia that were incorrectly identified as dead/relapsed, and FN (false negative) stands for dead/relapsed pediatric patients with leukemia incorrectly identified as alive/non-relapsed [51].

The performance of each method was evaluated using cross-validation, in which the data set was randomly divided into a test set (30%) and a training set (70%). This procedure was repeated 100 times, and the mean and standard deviation of the evaluation criteria were calculated. In order to prevent over-fitting, the tuning parameters in the ML algorithms were selected by 5-fold cross-validation. The optimal values of the hyper-parameters selected for each of the ML models are reported in Table 2.

Software Packages

In the present study, all analyses of ML methods were executed using R software version 4.1.1, with the following packages: “e1071” for SVM; “kernlab” for LS-SVM; “adabag” for bagging and boosting; “nnet” for NN; “naivebayes” for NB; “MASS” for LDA “randomForest” for RF and variable importance (VIMP) in the RF and drawing a partial plot; “rpart” for DT; “caret” determining tune parameters and “DMwR” for balancing the dataset.

Table 2 The tuning parameter values of machine learning algorithms

Methods	Hyperparameters	Value for death outcome	Value for relapse outcome
ANN	size	5	5
	weight decay	0.1	0.1
SVM	gamma	0.061	0.074
	cost	1	1
LS-SVM	kernel	radial	radial
	tau	0.0625	0.0625
	sigma	0.099	0.131
NB	kernel	rbfdot	rbfdot
	laplace	0	0
RF	use kernel	TRUE	TRUE
	adjust	1	1
DT	mtry	7	6
	ntree	1000	500
DT	minsplit	6	6
	minbucket	2	2
	cp	0.02	0.01
	maxdepth	4	8
Bagging	mfinal	100	50
	maxdepth	3	3
Boosting	mfinal	100	100
	coeflearn	Breiman	Freund
	maxdepth	3	3

Results

Data description

Of the 161 participating children with ALL in this study, 104 (64.6%) children were alive and 57 (35.4%) were dead. In addition, more than 90% of the children did not relapse. In this study, we used two variables as outcome variables, both containing two classes (alive or dead for survival status and yes or no for relapse experience).

The demographic, clinical and laboratory characteristics of all study participants are listed in Table 3. According to the results of Table 3, the majority of patients were male (57.1%), without central nervous system involvement (96.9%), without mediastinal tumor (91.3%), with rheumatoid arthritis signs (54%), without hepatomegaly (54.7%), with splenomegaly involvement (52.8%), and with lymphadenopathy (56.5%). The mean age of patients at diagnosis was 5.77±3.68 years with a range of 0.6 to 15.9 years. Also, the mean follow-up time of patients was 68.57 months with a minimum and maximum of 0.1 and 202.9 months, respectively. In addition, Table 3 shows the distributions of the characteristics of the patients who were randomly divided into two groups (train 70% and test 30%). As can be seen, there were no significant differences between the training and test data sets: for all variables *p*-values>0.05, except for central nervous system involvement (*p*-value=0.03).

As shown in Table 3, about 35% of children with ALL were classified as dead (minority class) and 65% of them as alive (majority class). In fact, the number of children who survived (104) was slightly higher than the number of children who died [57]. The imbalance ratio (IR) was 1.82, which means that for every sample of the minority class (dead) there were 1.82 samples of the majority class (alive). This indicates that the distribution of the samples between the two categories was slightly imbalanced. In terms of relapse, the majority of children did not relapse and the minority did relapse. The ratio of children who did not relapse to children who did relapse is 1:9 (IR=9). This shows an extreme imbalance between the two classes. Therefore, the SMOTE technique was used to correct an imbalanced problem in the main data sets. It should be noted that the ratio of dead to live samples and the ratio of relapsed to non-relapsed samples was considered to be 1:1 after SMOTE was used for both train and test data.

The performance of the ten ML algorithms for predicting mortality and relapse in children with ALL on the balanced dataset in terms of sensitivity, specificity, PPV, NPV, accuracy and AUC are reported in Tables 4 and 5, respectively.

Performance of the ML algorithms in predicting mortality

The results in Table 4 shown that the mean sensitivity in the test sets was above 70% for all algorithms except LR

Table 3 Description of the characteristics of children with acute lymphoblastic leukemia

Variable	Total		leukemia				P-value
	N	%	Train set		Test set		
	N	%	N	%	N	%	
All	161	100	112	69.6	49	30.4	
Sex							
Female	69	42.9	46	41.1	23	46.9	0.489
Male	92	57.1	66	58.9	26	53.1	
Survival status							
Alive	104	64.6	69	61.6	35	71.4	0.231
Dead	57	35.4	43	38.4	14	28.6	
Relapse							
No	145	90.1	100	89.3	45	91.8	0.619
Yes	16	9.9	12	10.7	4	8.2	
CNS							
No	156	96.9	111	99.1	45	91.8	0.030
Yes	5	3.1	1	0.9	4	8.2	
Mediastinal Tumor							
No	147	91.3	102	91.1	45	91.8	0.874
Yes	14	8.7	10	8.9	4	8.2	
RA							
No	74	46	50	44.6	24	49	0.611
Yes	87	54	62	55.4	25	51	
Hepatomegaly							
No	88	54.7	59	52.7	29	59.2	0.446
Yes	73	45.3	53	47.3	20	40.8	
Splenomegaly							
No	76	47.2	51	45.5	25	51	0.521
Yes	85	52.8	61	54.5	24	49	
lymphadenopathy							
No	70	43.5	49	43.8	21	42.9	0.916
Yes	91	56.5	63	56.3	28	57.1	
Age(year): Mean ± SD	5.77 ± 3.68		5.94 ± 3.82		5.40 ± 3.34		0.675
Hemoglobin	7.33 ± 2.32		7.52 ± 2.47		6.89 ± 1.89		0.114
WBC	37208.69 ± 77582.98		39787.76 ± 87369.78		31313.82 ± 48583.08		0.464
Time	68.57 ± 47.73		67.07 ± 49.60		72.01 ± 43.45		0.485
platelets	89809.32 ± 127372.21		91983.04 ± 136193.17		84840.82 ± 105589.39		0.851

CNS: Central Nervous Involvement, RA: Rheumatoid arthritis signs, WBC: White Blood Cell

and LDA. The mean specificity for the three algorithms, namely LS-SVM, boosting and RF was below (50%) for the test sets and between 66 and 86% for the other models. The mean PPV of the algorithms ranged from 0.59 to 0.82, with the lowest value for LS-SVM and the highest for LDA. Also, the mean NPV performance of all algorithms was greater than (65%) for the test sets. In addition, the mean accuracy and AUC for four algorithms, namely SVM, ANN, NB and bagging were similar (74%) for the test sets. The mean accuracy and AUC for the ML algorithms for the test sets ranged from 0.64 to 0.74, with the lowest values belonging to LS-SVM. In general, the performance of all ML methods for the prediction of mortality in children with ALL was approximately the same in terms of accuracy and AUC. In summary, the results shown that ANN and bagging outperformed the

other ML algorithms in all criteria, especially with a sensitivity of over 80%.

Performance of the ML algorithms in predicting relapse

As can be seen from the results presented in Table 5, the mean sensitivity of three of the ML algorithms, namely LS-SVM, bagging and boosting, was above 90%, and of the remaining algorithms, RF had the lowest mean sensitivity (64%). The mean specificity of all algorithms was below (70%) for the test sets, except for the RF algorithms, which achieved 96%. In addition, three ML methods, namely NB, LR and LDA had the lowest specificity (59%) for the test sets among all ML algorithms. The mean PPV of the algorithms was moderate and varied from 0.63 to 0.77 for the test sets, with LR and LDA having the lowest, and SVM and RF having the highest.

Table 4 Performance criteria of machine learning algorithms for predicting mortality of children with acute lymphoblastic leukemia

Dataset	Method	Sensitivity		Specificity		PPV		NPV		Accuracy		AUC		
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Train	SVM	0.73	0.050	0.94	0.036	0.93	0.034	0.78	0.024	0.84	0.017	0.84	0.018	
	LS-SVM	0.90	0.005	0.48	0.097	0.66	0.043	0.98	0.010	0.74	0.047	0.74	0.048	
	ANN	0.83	0.021	0.97	0.015	0.97	0.017	0.86	0.015	0.90	0.012	0.90	0.013	
	NB	0.77	0.019	0.91	0.028	0.89	0.026	0.80	0.013	0.84	0.014	0.84	0.014	
	Bagging	0.88	0.027	0.98	0.015	0.98	0.015	0.90	0.018	0.93	0.012	0.93	0.013	
	Boosting	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	
	DT	0.74	0.060	0.88	0.063	0.86	0.055	0.78	0.032	0.81	0.019	0.81	0.019	
	RF	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	
	LR	0.65	0.030	0.89	0.027	0.85	0.024	0.72	0.018	0.77	0.016	0.77	0.015	
	LDA	0.61	0.027	0.91	0.028	0.87	0.028	0.70	0.015	0.76	0.014	0.76	0.014	
	Test	SVM	0.71	0.063	0.77	0.093	0.76	0.078	0.72	0.059	0.74	0.041	0.74	0.040
		LS-SVM	0.98	0.015	0.30	0.098	0.59	0.060	0.92	0.047	0.64	0.061	0.64	0.047
		ANN	0.80	0.034	0.68	0.102	0.72	0.078	0.77	0.047	0.74	0.049	0.74	0.048
		NB	0.76	0.039	0.73	0.092	0.74	0.074	0.75	0.054	0.74	0.043	0.74	0.044
Bagging		0.82	0.047	0.66	0.095	0.71	0.068	0.78	0.063	0.74	0.044	0.74	0.045	
Boosting		0.95	0.019	0.43	0.092	0.63	0.060	0.90	0.045	0.69	0.053	0.69	0.045	
DT		0.72	0.077	0.73	0.109	0.74	0.075	0.73	0.060	0.73	0.042	0.73	0.043	
RF		0.93	0.026	0.46	0.095	0.64	0.060	0.87	0.054	0.70	0.051	0.70	0.045	
LR		0.65	0.046	0.83	0.071	0.79	0.072	0.70	0.050	0.73	0.032	0.74	0.031	
LDA		0.60	0.043	0.86	0.067	0.82	0.078	0.68	0.049	0.73	0.033	0.73	0.032	

However, the values of the mean NPV of the algorithms ranged from 0.66 to 0.96, with the LR algorithm having the lowest value and the boosting algorithm having the highest value. In addition, the mean accuracy of all algorithms for the test sets was greater than (70%), except for the two algorithms LR (64%) and LDA (65%). The AUC of the ML algorithms ranged from 0.64 to 0.84 for the test sets, with the lowest values obtained by LR and LDA, and the highest by RF and boosting. In general, of all ML methods, boosting and RF performed well in terms of accuracy and AUC criteria, whereas LR and LDA performed poorly.

The relative importance of variables in predicting mortality and relapse

The results indicated that ANN and bagging were the two best classifiers to predict mortality, while RF and boosting were the two best classifiers to predict relapse. Therefore, for a more detailed investigation, the relative importance of the prognostic factors was extracted based on these algorithms.

Hence, the relative importance of each prognostic factors in the prediction of mortality was calculated by ANN and bagging based on the Garson algorithm and the Gini index, respectively (see Fig. 2). As shown in Fig. 2, WBC was identified as the most important and influential variable affecting patient mortality based on the two algorithms. In addition, using ANN method, three variables of clinical side effects, namely splenomegaly, hepatomegaly and lymphadenopathy were found to be important variables for predicting mortality of children. Nevertheless, age, hemoglobin and platelets were detected as the most critical and important factors using the bagging algorithm.

As well, the relative importance of each prognostic factors in predicting relapse was calculated using RF and boosting based on the Gini index (see Fig. 3). As illustrated in Fig. 3, the two algorithms overlap in determining the significance of the prognostic factors. According to Fig. 3, based on the Gini index, age at diagnosis and three laboratory factors such as WBC, hemoglobin and platelets were identified as the four most important and influential prognostic factors to predict relapse in both algorithms. Interestingly, age at diagnosis was identified as the first and second most important prognostic factors in RF and boosting, respectively.

The partial dependence plots for the four most influential prognostic factors affecting the predicted mortality and relapse probabilities in children with ALL based on a bagging classifier and a random forest classifier are shown in Figs. 4 and 5.

As can be seen, the non-linear relationships between the importance predictors and the probability of mortality and relapse were evident. For example, the probability

of death in children was high at the beginning and then decreased until the age of eight, while it increased again in children over the age of eight until the age of 14 and then decreased again (see Fig. 4). There was an increasing relationship between relapse and hemoglobin (g/dL) until reaching 6.5 and then the relationship was declined; such that the probability of relapse decreased as the hemoglobin (g/dL) increases up to 11, followed by an increasing trend and then decreasing trend (see Fig. 5). Also, the probability of relapse increased for children under the age of five, but decreased for children over the age of five.

Discussion

In the present study, we aimed to demonstrate the potential of different ML algorithms to predict mortality and relapse in children aged seven months to sixteen years treated at Taleghani Hospital in Gorgan, north of Iran. Given the significantly increased risk of mortality associated with childhood ALL relapse after treatment, accurate prediction of treatment outcomes is essential for the development of effective treatment plans. However, there is currently no clinical screening tool that can predict mortality and relapse with high accuracy [33]. In this context, the eight machine learning models LR, DT, RF, SVM, LS-SVM, LDA, ANN, NB, bagging and boosting as well as the two classical methods LR and LDA were applied and their performance in predicting survival and relapse in children with ALL was compared. The findings indicated that ANN and bagging were effective in predicting mortality, while boosting and RF algorithms excelled in predicting relapse.

Based on the accuracy measures obtained from our experiments over 100 repetitions, our findings indicated that all the classification algorithms performed similarly in their ability to predict mortality of the patients, with results ranging from 64 to 74%. Similarly, for relapse classification, the algorithms demonstrated accuracy results that varied from 64 to 84% on test data sets (over 100 repetitions). AUC of ML algorithms for mortality and relapse were above 64%. The ANN and bagging outperformed the other ML algorithms in all criteria, especially with a sensitivity of over 80% and boosting and RF performed well in terms of accuracy and AUC criteria.

To the best of our current understanding, the vast majority of previous work in this area has relied predominantly on gene expression, molecular and blood smear image datasets to diagnose acute leukemia, determine survival status and predict relapse [27–30]. Nevertheless, a limited number of studies have used ML algorithms to predict mortality and relapse using laboratory and clinical data. For instance, Pan et al. used four ML algorithms based on clinical prognostic factors to predict relapse in ALL patients, with model accuracy ranging from 79 to 83% and AUC ranging from 79 to 90%. Based on the

Table 5 Performance criteria of machine learning algorithms for predicting relapse of childhood acute lymphoblastic leukemia

Dataset	Method	Sensitivity		Specificity		PPV		NPV		Accuracy		AUC		
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Train	SVM	0.90	0.033	0.84	0.046	0.85	0.031	0.90	0.026	0.87	0.017	0.87	0.018	
	LS-SVM	0.98	0.011	0.76	0.082	0.81	0.054	0.98	0.015	0.87	0.043	0.87	0.044	
	ANN	0.89	0.044	0.88	0.032	0.88	0.029	0.89	0.037	0.89	0.031	0.88	0.031	
	NB	0.92	0.028	0.67	0.047	0.73	0.026	0.90	0.028	0.79	0.023	0.79	0.022	
	Bagging	0.97	0.019	0.93	0.024	0.93	0.020	0.97	0.017	0.95	0.014	0.95	0.014	
	Boosting	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	
	DT	0.94	0.031	0.88	0.048	0.89	0.039	0.94	0.027	0.91	0.023	0.91	0.022	
	RF	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	
	LR	0.72	0.048	0.67	0.043	0.69	0.029	0.70	0.030	0.69	0.028	0.69	0.028	
	LDA	0.74	0.058	0.67	0.049	0.69	0.025	0.73	0.033	0.71	0.026	0.71	0.025	
	Test	SVM	0.87	0.052	0.68	0.094	0.74	0.069	0.83	0.068	0.77	0.046	0.77	0.045
		LS-SVM	0.96	0.030	0.62	0.098	0.72	0.063	0.93	0.042	0.79	0.051	0.79	0.049
		ANN	0.82	0.060	0.65	0.094	0.70	0.075	0.78	0.073	0.73	0.053	0.74	0.052
		NB	0.88	0.052	0.59	0.086	0.69	0.063	0.82	0.068	0.74	0.045	0.74	0.043
Bagging		0.91	0.041	0.66	0.091	0.73	0.062	0.87	0.056	0.79	0.044	0.79	0.045	
Boosting		0.97	0.019	0.71	0.081	0.77	0.058	0.96	0.029	0.84	0.042	0.84	0.041	
DT		0.86	0.060	0.63	0.088	0.70	0.062	0.82	0.073	0.74	0.045	0.74	0.044	
RF		0.64	0.083	0.96	0.025	0.73	0.070	0.94	0.037	0.80	0.050	0.81	0.083	
LR		0.70	0.067	0.59	0.096	0.63	0.073	0.66	0.073	0.64	0.047	0.64	0.046	
LDA		0.71	0.076	0.59	0.087	0.64	0.064	0.67	0.079	0.65	0.040	0.65	0.039	

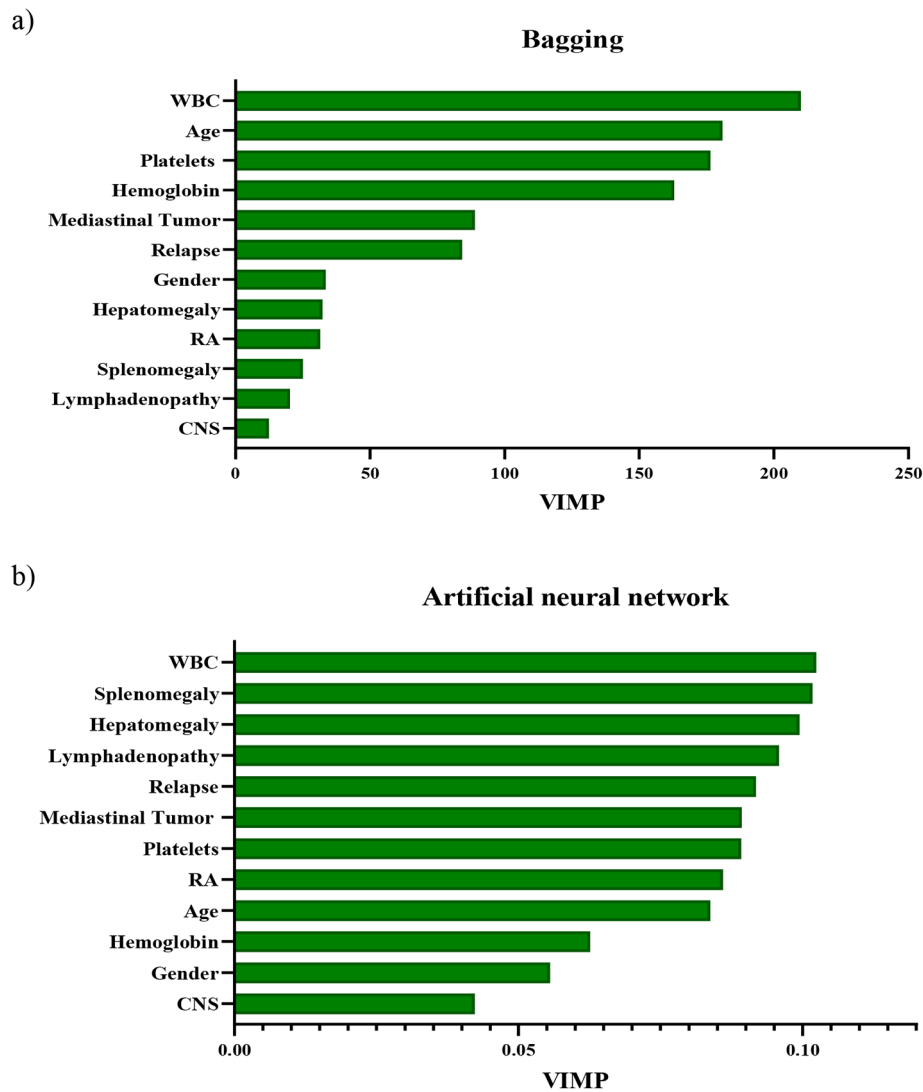


Fig. 2 Variable importance (VIMP) for predicting mortality of childhood acute lymphoblastic leukemia based on Gini Index and Garson algorithm using two machine learning algorithms: (a) bagging and (b) artificial neural network
 WBC: White Blood Cell, RA: Rheumatoid arthritis signs, CNS: Central Nervous Involvement

evaluation criteria, the RF was identified as the best algorithm for predicting relapse in patients in their study [33]. In another study, Kashfi et al. used seven ML algorithms to predict survival in children with ALL. SVM with an accuracy of 94.90% (95% CI: 88.49–98.32) and XG-Boost with an accuracy of 88.5% (95% CI: 82.3–94.0) were identified as the superior algorithms compared to classifiers in their study [4]. Feature selection plays a vital role in ML, as irrelevant features can result in lower accuracy, reduced interpretability, and overfitting in classification analysis. Additionally, it is crucial for clinicians to identify the most predictive prognostic factors for treatment outcomes in order to personalize treatment plans [33]. However, the selection of prognostic factors could lead to bias. To objectively identify the strongest predictors

of mortality and relapse, we used the cross-validation method.

Our research findings revealed that WBC, age at diagnosis, hemoglobin, platelets, splenomegaly, hepatomegaly and lymphadenopathy emerged as the most significant prognostic factors for predicting mortality. Furthermore, age at diagnosis, hemoglobin, WBC and platelets were recognized as the critical variables for predicting relapse according to the two most effective algorithms identified. Consistent with our findings, in the study conducted by Pan et al., age, WBC, hemoglobin, and platelets were recognized as important predictors for relapse by RF [33]. Also, in a retrospective study by Bhojwani et al. age at diagnosis, WBC, platelets, and hemoglobin levels were significant predictors of both overall survival and relapse.

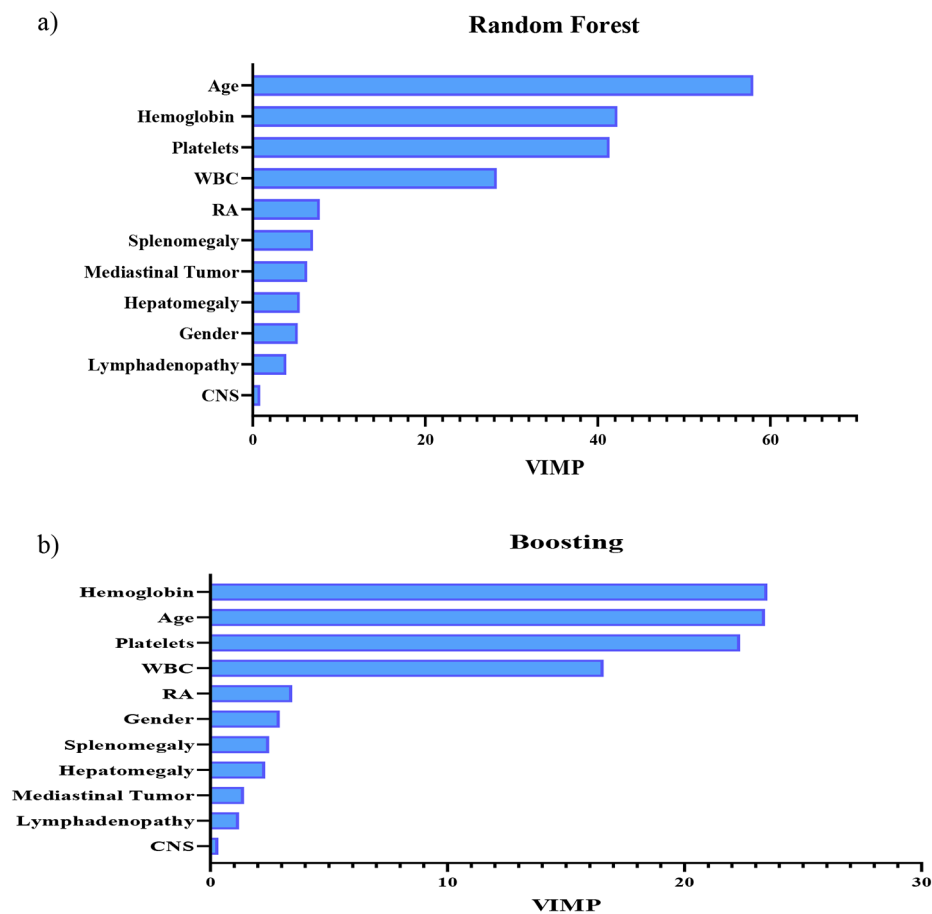


Fig. 3 Variable importance (VIMP) for predicting relapse of childhood acute lymphoblastic leukemia based on Gini Index using two machine learning algorithms: (a) random forest and (b) boosting

WBC: White Blood Cell, RA: Rheumatoid arthritis signs, CNS: Central Nervous Involvement

The study also found that the presence of central nervous system involvement at diagnosis and early response to treatment were significant prognostic factors of relapse, while sex and race were not significant prognostic factors of either overall survival or relapse [52]. In another study by Hunger et al., age, sex, WBC, and immunophenotype were significant prognostic factors of event free survival, while age, sex and WBC were significant prognostic factors of overall survival. Contrary with us, in this study platelets and hemoglobin were not found to be significant prognostic factors of survival [53].

WBC, platelets, and hemoglobin are significant prognostic factors of both mortality and relapse in children with ALL because they are indicators of the extent of bone marrow infiltration by leukemic cells and the degree of bone marrow suppression caused by the disease and its treatment. Higher WBC and lower platelets and hemoglobin levels are associated with more advanced and aggressive disease, which can lead to a higher risk of relapse and poorer overall survival. Therefore, monitoring these hematological parameters can help clinicians

to identify patients at higher risk of relapse and poorer overall survival and may also guide treatment decisions to optimize patient outcomes [53–55].

One of the strengths of this study is the inclusion of 161 subjects, which appears to be a sufficient sample size for evaluating the performance of different ML methods. Although there are no universal recommendations for the optimal sample size for ML methods, Rajput et al. have shown that increasing the sample size improves both effect size and model accuracy. This means that models with larger data sets can detect patterns more effectively. Furthermore, their study showed that the variability of effect size and model accuracy decreases with a sample size above 100. Therefore, our sample size of 161 individuals is likely sufficient for robust analysis and reliable results [56].

Another strength of the present study is that we repeated the ML algorithms 100 times, which provides better insight into the performance of different algorithms in predicting patient mortality and relapse of ALL in children. Additionally, this study includes partial

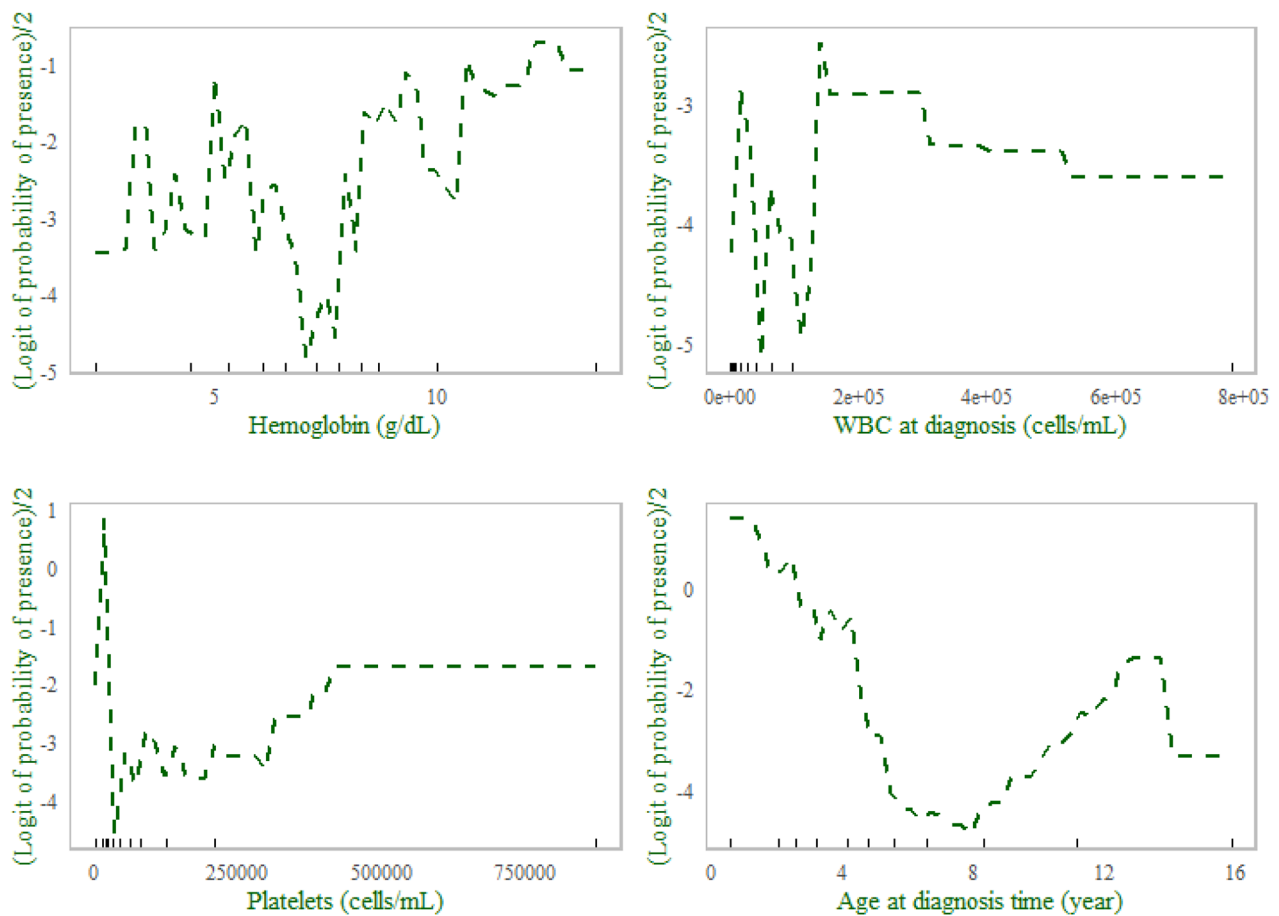


Fig. 4 Partial dependence plots for the four most influential variables on mortality in acute lymphoblastic leukemia data based on a bagging classifier

dependence plots to investigate non-linear, monotonic, or more complex relationships between outcomes (mortality and relapse) and prognostic factors. By identifying the most important prognostic factors for relapse and mortality and depicting non-linear relationships of risk factors, clinicians can better tailor treatment protocols and monitoring strategies to optimize patient outcomes and reduce the risk of relapse or death. However, one limitation of this study is its retrospective design, as the data was obtained from the patients’ medical records. This approach makes the analysis susceptible to possible bias in the estimates of measures such as sensitivity. In addition, some prognostic factors that may need to be considered in future prediction models for these patients were not recorded in the medical records. It is suggested that a future prospective cohort study be designed to include prognostic factors that may impact patient survival and relapse. In addition, the study was conducted at a single center, which could limit the generalizability of the results to other centers. It is therefore suggested that a multicenter study should be conducted for further investigation.

Recently, fuzzy-based clustering methods have gained attention for identifying important prognostic factors related to disease outcome. It is suggested that future research should explore the integration of clustering techniques with machine learning algorithms to predict survival and relapse in children with ALL. These clustering techniques help to identify subgroups of patients with similar prognostic factors. By using these models, researchers can uncover hidden patterns in the data, ultimately leading to more personalized treatment strategies. In addition, clustering approaches have the potential to improve the accuracy and reliability of predictions, thereby contributing to more effective management of ALL [57].

Conclusion

The focus of this study was to evaluate the performance of eight machine learning techniques and two classical methods in predicting mortality and relapse in patients with ALL. Our results showed that ANN and the bagging method were the best algorithms for predicting mortality

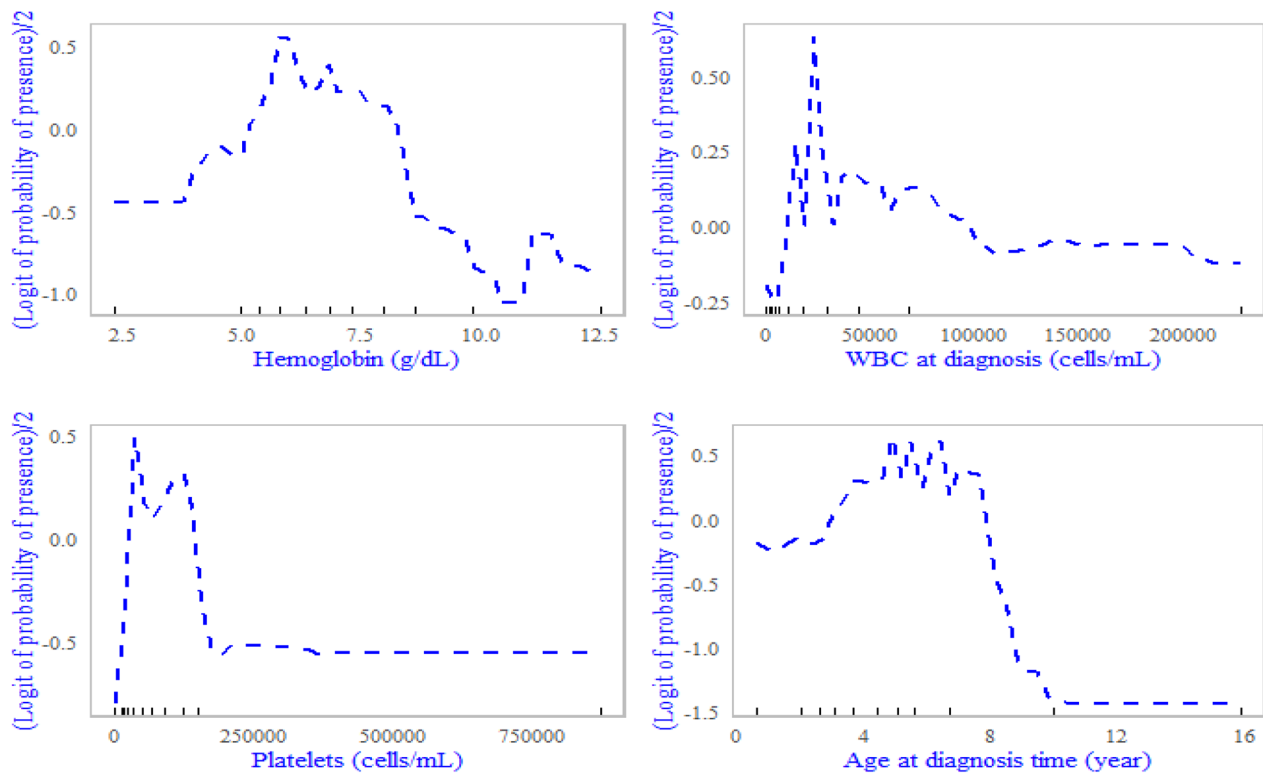


Fig. 5 Partial dependence plots for the four most influential variables on experiencing a relapse in acute lymphoblastic leukemia data based on a random forest classifier

while boosting and RF were the most effective algorithms for predicting relapse in ALL patients in this study.

In conclusion, these findings have significant clinical implications, as they provide valuable insights into the prognostic factors associated with mortality and relapse in children with ALL. This information can help inform treatment decisions and ultimately enhance patient outcomes.

Abbreviations

ALL	Acute Lymphocytic Leukemia
AML	Acute Myeloid Leukemia
ANN	Artificial Neural Network
AUC	Area Under Curve
CLL	Chronic Lymphocytic Leukemia
CML	Chronic Myeloid Leukemia
CNS	Central Nervous Involvement
DT	Decision Tree
FN	False Negative
FP	False Positive
GBM	Gradient Boosting Machine
GLM	Generalized linear model
LDA	Linear Discriminant Analysis
LR	Logistic Regression
LS-SVM	Least square support vector machine
ML	Machine learning
MLR	Multinomial Linear Regression
NB	Naïve bayes
NPV	Negative Predictive Value
OOB	Out-of-Bag error
PPV	Positive Predictive Value
RA	Rheumatoid Arthritis Signs

RF	Random Forest
SD	Standard Deviation
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
WBC	White Blood Cells

Acknowledgements

We would like to appreciate the Vice-chancellor of Education of the Hamadan University of Medical Science for technical support for their approval and support of this work.

Author contributions

LT, ZM and RH conceived the research topic, explored that idea and drafted the manuscript. LT, ZM, RH and ZZ performed the statistical analysis. NB provided the data and participated in data analysis and writing. SK, NB, ZZ and ID participated in the interpretations and drafting of the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported and approved by Hamadan University of Medical Sciences (Grant NO: 1402121510951). The funding body had no role in the design of the study and collection as well as in writing the manuscript.

Data availability

The datasets generated during and analyzed during the current study are not publicly available due to the Hamadan University of Medical Science restrictions on public sharing data, but are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

This study was submitted to and approved by the Ethical Committee of Hamadan University of Medical Science (Ethical code: IR.UMSHA.REC.1402.748). Informed consent was obtained from a parent or legal guardian for being included in the study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran

²Student Research Committee, Hamadan University of Medical Sciences, Hamadan, Iran

³Department of Biostatistics and Epidemiology, School of Health, Golestan University of Medical Sciences, Gorgan, Iran

⁴Modeling of Noncommunicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Iran

⁵Department of Photogrammetry and Remote Sensing, K.N. Toosi University of Technology, Tehran, Iran

⁶Health Sciences Research Center, Health Sciences & Technology Research Institute, Hamadan University of Medical Science, Hamadan, Iran

⁷School of Public Health, University of Alberta, Edmonton, Canada

Received: 6 January 2024 / Accepted: 21 August 2024

Published online: 16 September 2024

References

- World Population Prospects 2023. <https://population.un.org/wpp>.
- World Health Organization 2023. <https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/child-mortality-and-causes-of-death>.
- Belson M, Kingsley B, Holmes A. Risk factors for acute leukemia in children: a review. *Environ Health Perspect*. 2007;115(1):138–45.
- Kashef A, Khatibi T, Mehrvar A. Treatment outcome classification of pediatric acute lymphoblastic leukemia patients with clinical and medical data using machine learning: a case study at MAHAK hospital. *Inf Med Unlocked*. 2020;20:100399.
- Torres-Flores J, Espinoza-Zamora R, Garcia-Mendez J, Cervera-Ceballos E, Sosa-Espinoza A, Zapata-Canto N. Treatment-related mortality from infectious complications in an acute leukemia clinic. *J Hematol*. 2020;9(4):123.
- Kaplan JA. Leukemia in children. *Pediatr Rev*. 2019;40(7):319–31.
- Torres-Roman JS, Valcarcel B, Guerra-Canchari P, Santos CAD, Barbosa IR, La Vecchia C, et al. Leukemia mortality in children from Latin America: trends and predictions to 2030. *BMC Pediatr*. 2020;20(1):1–9.
- Nguyen K, Devidas M, Cheng S-C, La M, Raetz EA, Carroll WL, et al. Factors influencing survival after relapse from acute lymphoblastic leukemia: a children's oncology group study. *Leukemia*. 2008;22(12):2142–50.
- Zawitkowska J, Lejman M, Romiszewski M, Matysiak M, Ćwiklińska M, Balwier W, et al. Results of two consecutive treatment protocols in Polish children with acute lymphoblastic leukemia. *Sci Rep*. 2020;10(1):1–9.
- Conneely SE, Stevens AM. Acute myeloid leukemia in children: emerging paradigms in genetics and new approaches to therapy. *Curr Oncol Rep*. 2021;23:1–13.
- Jerez-Aragonés JM, Gómez-Ruiz JA, Ramos-Jiménez G, Muñoz-Pérez J, Alba-Conejo E. A combined neural network and decision trees model for prognosis of breast cancer relapse. *Artif Intell Med*. 2003;27(1):45–63.
- Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380(14):1347–58.
- Janiesch C, Zszech P, Heinrich K. Machine learning and deep learning. *Electron Markets*. 2021;31(3):685–95.
- Farhadian M, Torkaman S, Mojarad F. Random forest algorithm to identify factors associated with sports-related dental injuries in 6 to 13-year-old athlete children in Hamadan, Iran-2018-a cross-sectional study. *BMC Sports Sci Med Rehabilitation*. 2020;12:1–9.
- Soofi AA, Awan A. Classification techniques in machine learning: applications and issues. *J Basic Appl Sci*. 2017;13:459–65.
- Wu W-T, Li Y-J, Feng A-Z, Li L, Huang T, Xu A-D, et al. Data mining in clinical big data: the frequently used databases, steps, and methodological models. *Military Med Res*. 2021;8:1–12.
- Esteve A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115–8.
- Karmakar R, Chatterjee S, Das AK, Mandal A. BCPUML: breast cancer prediction using machine learning approach—A performance analysis. *SN Comput Sci*. 2023;4(4):377.
- Chang V, Bhavani VR, Xu AQ, Hossain M. An artificial intelligence model for heart disease detection using machine learning algorithms. *Healthc Analytics*. 2022;2:100016.
- Moslehi S, Rabiei N, Soltanian AR, Mamani M. Application of machine learning models based on decision trees in classifying the factors affecting mortality of COVID-19 patients in Hamadan, Iran. *BMC Med Inf Decis Mak*. 2022;22(1):192.
- Hassanzadeh R, Farhadian M, Rafeiemehr H. Hospital mortality prediction in traumatic injuries patients: comparing different SMOTE-based machine learning algorithms. *BMC Med Res Methodol*. 2023;23(1):1–15.
- Anderson JP, Parikh JR, Shenfeld DK, Ivanov V, Marks C, Church BW, et al. Reverse engineering and evaluation of prediction models for progression to type 2 diabetes: an application of machine learning using electronic health records. *J Diabetes Sci Technol*. 2016;10(1):6–18.
- Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J*. 2015;13:8–17.
- Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform*. 2006;2:117693510600200030.
- Yeoh E-J, Ross ME, Shurtleff SA, Williams WK, Patel D, Mahfouz R, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell*. 2002;1(2):133–43.
- Salah HT, Muhsen IN, Salama ME, Owaidah T, Hashmi SK. Machine learning applications in the diagnosis of leukemia: current trends and future directions. *Int J Lab Hematol*. 2019;41(6):717–25.
- Ross ME, Zhou X, Song G, Shurtleff SA, Girtman K, Williams WK, et al. Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood*. 2003;102(8):2951–9.
- Willenbrock H, Juncker A, Schmiegelow K, Knudsen S, Ryder L. Prediction of immunophenotype, treatment response, and relapse in childhood acute lymphoblastic leukemia using DNA microarrays. *Leukemia*. 2004;18(7):1270–7.
- Mohapatra S, Patra D, Satpathi S, editors. Image analysis of blood microscopic images for acute leukemia detection. 2010 international conference on industrial electronics, control and robotics; 2010: IEEE.
- Tran V-N, Ismail W, Hassan R, Yoshitaka A, editors. An automated method for the nuclei and cytoplasm of acute myeloid leukemia detection in blood smear images. 2016 World Automation Congress (WAC); 2016: IEEE.
- Eckardt J-N, Bornhäuser M, Wendt K, Middeke JM. Application of machine learning in the management of acute myeloid leukemia: current practice and future prospects. *Blood Adv*. 2020;4(23):6077–85.
- Ghaderzadeh M, Asadi F, Hosseini A, Bashash D, Abolghasemi H, Roshanpour A. Machine learning in detection and classification of leukemia using smear blood images: a systematic review. *Sci Program*. 2021;2021:1–14.
- Pan L, Liu G, Lin F, Zhong S, Xia H, Sun X, et al. Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia. *Sci Rep*. 2017;7(1):1–9.
- Ramezan A, Warner CA, Maxwell TE. Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. *Remote Sens*. 2019;11(2):185.
- Tougui I, Jilbab A, El Mhamdi J. Impact of the choice of cross-validation techniques on the results of machine learning-based diagnostic applications. *Healthc Inf Res*. 2021;27(3):189–99.
- Agresti A, Kateri M. *Categorical data analysis*. Springer; 2011.
- Lee SK. On classification and regression trees for multiple responses and its application. *J Classif*. 2006;23(1):123–41.
- Najafi-Ghobadi S, Najafi-Ghobadi K, Tapak L, Aghaei A. Application of data mining techniques and logistic regression to model drug use transition to injection: a case study in drug use treatment centers in Kermanshah Province, Iran. *Subst Abuse Treat Prev Policy*. 2019;14(1):1–11.

39. Buntine W, Niblett T. A further comparison of splitting rules for decision-tree induction. *Mach Learn.* 1992;8:75–85.
40. Najafi-Vosough R, Faradmal J, Hosseini SK, Moghimbeigi A, Mahjub H. Predicting hospital readmission in heart failure patients in Iran: a comparison of various machine learning methods. *Healthc Inf Res.* 2021;27(4):307–14.
41. Breiman L. Random forests. *Mach Learn.* 2001;45:5–32.
42. Suykens JA, De Brabanter J, Lukas L, Vandewalle J. Weighted least squares support vector machines: robustness and sparse approximation. *Neurocomputing.* 2002;48(1–4):85–105.
43. Singh S, Parmar KS, Makkhan SJS, Kaur J, Peshoria S, Kumar J. Study of ARIMA and least square support vector machine (LS-SVM) models for the prediction of SARS-CoV-2 confirmed cases in the most affected countries. *Chaos Solitons Fractals.* 2020;139:110086.
44. Hastie T, Tibshirani R, Friedman JH, Friedman JH. *The elements of statistical learning: data mining, inference, and prediction.* Springer; 2009.
45. Ray S, editor. *A quick review of machine learning algorithms.* 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon); 2019: IEEE.
46. Garson DG. *Interpreting neural network connection weights.* 1991.
47. Tapak L, Shirmohammadi-Khorram N, Amini P, Alafchi B, Hamidi O, Poorolajal J. Prediction of survival and metastasis in breast cancer patients using machine learning classifiers. *Clin Epidemiol Global Health.* 2019;7(3):293–9.
48. Mayr A, Binder H, Gefeller O, Schmid M. The evolution of boosting algorithms. *Methods Inf Med.* 2014;53(06):419–27.
49. Shariatnia S, Ziaratban M, Rajabi A, Salehi A, Abdi Zarrini K, Vakili M. Modeling the diagnosis of coronary artery disease by discriminant analysis and logistic regression: a cross-sectional study. *BMC Med Inf Decis Mak.* 2022;22(1):85.
50. Izenman AJ. *Linear discriminant analysis. Modern multivariate statistical techniques: regression, classification, and manifold learning.* Springer; 2013. pp. 237–80.
51. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med.* 2013;4(2):627.
52. Bhojwani D, Kang H, Menezes RX, Yang W, Sather H, Moskowitz NP, et al. Gene expression signatures predictive of early response and outcome in high-risk childhood acute lymphoblastic leukemia: a children's oncology group study. *J Clin Oncol.* 2008;26(27):4376.
53. Hunger SP, Lu X, Devidas M, Camitta BM, Gaynon PS, Winick NJ, et al. Improved survival for children and adolescents with acute lymphoblastic leukemia between 1990 and 2005: a report from the children's oncology group. *J Clin Oncol.* 2012;30(14):1663.
54. Schultz KR, Pullen DJ, Sather HN, Shuster JJ, Devidas M, Borowitz MJ, et al. Risk-and response-based classification of childhood B-precursor acute lymphoblastic leukemia: a combined analysis of prognostic markers from the pediatric oncology group (POG) and children's cancer group (CCG). *Blood.* 2007;109(3):926–35.
55. Pui C-H, Carroll WL, Meshinchi S, Arcenci RJ. Biology, risk stratification, and therapy of pediatric acute leukemias: an update. *J Clin Oncol.* 2011;29(5):551.
56. Rajput D, Wang W-J, Chen C-C. Evaluation of a decided sample size in machine learning applications. *BMC Bioinformatics.* 2023;24(1):48.
57. Yang Y, Su X, Zhao B, Li G, Hu P, Zhang J et al. Fuzzy-based deep attributed graph clustering. *IEEE Trans Fuzzy Syst.* 2023.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.