Transactions on Machine Intelligence 4(2) (2021) 63-69

ISSN Online: 2821-1693

Transactions on Machine Intelligence

Journal Homepage: https://tmachineintelligence.ir/index.php/TMI

# Predicting Survival of Leukemia Patients Using a Support Vector Machine Based on the Bowerbird Algorithm

A. Marousi[1,*], F. Marousi[2,3]

[1] Department of Computer Engineering, Khavaran Non-Profit Higher Education Institution, Mashhad, Iran
[2] Department of Computer Engineering, Gonabad Higher Education Complex, Gonabad, Iran
[3] Computer Instructor, Khorasan Razavi Department of Education

**ARTICLE INFO**

Article History:
Received 160 March 2021
Received in revised form 13 May 2021
Accepted 4 June 2021
Available online 5 June 2021

Keywords:
Cancer, Leukemia, Bowerbird Algorithm, Support Vector Machine

## ABSTRACT

Unfortunately, research has shown that cancer incidence has been increasing in recent years. Leukemia is a type of blood cancer caused by an increase in the number of white blood cells. Generally, any type of blood cancer is extremely dangerous, and in most cases, there is no cure. Acute myeloid leukemia (AML) is a common and fatal type of this cancer. Predicting survival after diagnosis is one of the key indicators for evaluating treatment methods, which is the focus of the present study. In this research, we used a combination of Support Vector Machine (SVM) with the Bowerbird Algorithm to analyze survival status and predict mortality in patients. The data used pertains to patient information from Seyyed al-Shohada Hospital in Isfahan, with 197 samples and 9 features. MATLAB software was used to run the programs. Evaluation was based on diagnostic indices including sensitivity, specificity, and accuracy. The proposed SVM based on the Bowerbird Algorithm achieved a performance of 69.57% accuracy, 75.52% sensitivity, and 64.48% specificity, outperforming the combination of SVM with other optimization algorithms such as Cuckoo Search, Harmony Search, and Firefly Algorithm. Therefore, the proposed method is a promising tool for predicting survival in leukemia patients with improved diagnostic accuracy.

## 1. INTRODUCTION

Leukemia is a type of cancer that originates in the bone marrow and typically grows rapidly, ranking fifth in mortality worldwide and second in Iran. It is divided into three main categories: leukemia, lymphoma, and myeloma. Hence, predicting the survival of these patients is crucial [1, 2].

Research indicates that in Iran, 8% of all cancer patients are diagnosed with leukemia. In chronic diseases such as cancer, the goal of treatment is to increase the survival time of the patients. Therefore, factors affecting the survival time of leukemia patients are highly significant [3]. Various techniques are used for cancer classification, including Support Vector Machines, Artificial Neural Networks, and others. This article employs the Support Vector Machine (SVM) algorithm. SVM is a relatively new method that has shown good performance compared to older methods

---

[*] Corresponding Author: akrammaroosi@gmail.com
Department of Computer Engineering, Khavaran Non-Profit Higher Education Institution, Mashhad, Iran

http://dx.doi.org/10.47176/TMI.2021.63

for classification in recent years [4]. In some studies, its performance is superior to other classifiers, but in high-dimensional datasets, its performance declines due to complex calculations. Hence, heuristic algorithms are used to optimize SVM parameters to improve its accuracy [5]. In this article, we used the Bowerbird Algorithm (SBO) to enhance the accuracy of the SVM for predicting the survival of leukemia patients. Additionally, we compared the proposed method with other evolutionary algorithms such as Cuckoo Optimization, Genetic Algorithm, Firefly Algorithm, and Harmony Search Algorithm, and analyzed the results.

## 2. LITERATURE REVIEW

In recent years, extensive research has been conducted using Support Vector Machine (SVM) classification methods to reduce errors in disease identification and diagnosis. Additionally, studies conducted on predicting the survival of cancer patients have primarily employed regression methods and neural networks.

Karimi et al. [6] used SVM to classify red blood cells as normal or abnormal to diagnose anemia. The proposed classifier algorithm achieved very high accuracy rates based on validation metrics of sensitivity, specificity, and kappa.

In [7], a combination of SVM based on the Monarch Butterfly-Taylor algorithm was proposed for leukemia diagnosis. The proposed algorithm performed the SVM training process. However, the Taylor-based SVM achieved better performance in terms of accuracy, sensitivity, and specificity.

Another study attempted to classify leukemia images using shape features, comparing SVM classification methods with K-Nearest Neighbors (KNN) algorithms. Based on the results from comparing the average accuracy and training time in both methods, the accuracy of the SVM method was higher than that of KNN [8].

Hosseini Tashnizi and colleagues in 2013 compared Cox regression models and artificial neural networks in predicting the survival of acute leukemia patients. The results of this study showed that artificial neural network models, compared to Cox regression models, which assume proportional hazards for each variable, provide a suitable replacement with higher accuracy for predicting the survival of cancer patients [9].

Rawat et al. [10] introduced an SVM based on a genetic algorithm for classifying myeloblast cells, acute lymphoblastic cells, and blast subgroups. This method achieved better accuracy but was not executable for large database sizes.

Studies indicate that previous research on predicting the survival of cancer patients has primarily used Cox regression methods and neural networks, or SVM with a genetic algorithm. Considering the limitations of the Cox model, as well as the high accuracy and reliability of SVM and its flexibility in combination with other existing algorithms, we decided to use SVM models based on the Bowerbird Algorithm (SBO-SVM) in this study for predicting the survival of leukemia patients.

## 3. METHODOLOGY

### 3.1. Statistical Population

In a study, information was collected on 197 acute leukemia patients admitted with a definitive diagnosis of acute leukemia at Seyyed al-Shohada Hospital (Omidi) in Isfahan from 2006 to 2009. The variables included age, residence, gender, occupation, blood type, type of leukemia, white blood cell count (WBC), hemoglobin level, date of cancer diagnosis, and final status (alive or deceased) [9]. For implementing the results, SVM, a combination of SVM with the Bowerbird Algorithm, was compared with Firefly, Harmony Search, and Genetic Algorithms using MATLAB-written code.

### 3.2. Data Analysis

Out of the 197 acute leukemia patients in this study, 102 (51.8%) were diagnosed with Acute Lymphoblastic Leukemia (ALL), and 95 (48.2%) with Acute Myeloid Leukemia (AML). There were 116 (59%) males and the rest were females; 112 (57%) resided in urban areas, and 85 (43%) in rural areas. Blood type distribution included 89 (45.2%) with type O, 55 (28%) with type A, 45 (22.8%) with type B, and the remaining with type AB. The mean age for AML patients was 42.1 weeks, and for ALL patients, it was 23.1 weeks.

**Table 1.** Frequency percentage of characteristics of acute leukemia patients

| Feature | Value | Number (Person) | Percentage |
|---|---|---|---|
| Residence Type | Urban | 112 | 56.90% |
| | Rural | 85 | 43.10% |
| Gender | Male | 116 | 58.90% |
| | Female | 81 | 41.10% |
| Job | 0 | 70 | 35.50% |
| | 1 | 48 | 24.40% |
| | 2 | 33 | 1608% |
| | 3 | 22 | 11.20% |
| | 4 | 23 | 12.20% |
| Blood Type | A | 55 | 27.90% |
| | B | 45 | 22.80% |
| | O | 89 | 45.20% |
| | AB | 8 | 4.10% |
| Type of Leukemia | AML | 102 | 51.80% |
| | ALL | 95 | 48.20% |
| Final Status | Live | 104 | 52.80% |
| | Dead | 93 | 47.20% |

**Table 2.** Descriptive statistics of features of acute leukemia patients

| Feature | Min | Max | Average |
|---|---|---|---|
| Age | 0.5 | 80 | 33.098 |
| WBC | 0 | 1600 | 42.5889 |
| PLT | 0 | 97000 | 588.1909 |
| HB | 1.4 | 81 | 8.6932 |

### 3.3. Support Vector Machine

The Support Vector Machine (SVM) algorithm is a classification technique applicable to a wide variety of datasets. It is a relatively new method that has shown significant performance improvements over older classification methods in recent years. In SVM, datasets can be separated linearly or non-linearly [5].

When data can be separated linearly, SVM finds the optimal hyperplane with maximum margin by solving an optimization problem based on the training dataset, as represented by Equation (1).

$$\underset{x,b}{Min}\frac{1}{2}\|w\|, Y_i(w^t x_i + b) - 1 \geq 0;\ i = 1,\dots,L \tag{1}$$

where n is the number of samples, $x_i$ is the feature vector of the i-th sample, and $y_i$ is the output for the i -th sample. A critical assumption in SVM is that data is linearly separable, which is often not the case in practice. In 1995, Cortes and Vapnik introduced non-negative variables $\varepsilon_i \geq 0$ to account for errors for each vector. Consequently, the optimization problem transforms into finding w such that Equation (2) is minimized [11].

$$\begin{gathered} Min\ \frac{1}{2}\|w\| + C\sum_i \varepsilon_i \\ St: y_i\left(< wx_i > +b\right) \geq 1 - \varepsilon_i . \varepsilon_i \geq 0 \forall i \end{gathered} \tag{2}$$

Here, C is a positive constant for regularizing the margin, balancing the trade-off between maximizing the margin and minimizing classification error. When data cannot be separated linearly, and classes overlap, a non-linear transformation can map the data to a higher-dimensional space to reduce class overlap [12]. After mapping to a higher dimension, the optimal hyperplane is found using the previous equations by replacing $x_i$ with $\varphi(x_i)$. The radial basis function (RBF) kernel, shown in Equation (3), has demonstrated better classification performance than other kernels [13]. In this study, we used the RBF kernel.

$$k\left(x_i, x_j\right) = \exp\left(-\left|x_i - x_j\right|^2 \big/ 2\sigma^2\right) \tag{3}$$

To validate the proposed algorithm, we evaluated three metrics: Sensitivity, Specificity, and Accuracy, as given by Equations (4), (5), and (6).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

Additionally, we used the Bowerbird Optimization Algorithm to select the best parameters for the SVM algorithm to achieve optimal classification results.

### 3.4. Bowerbird Algorithm

The Bowerbird Algorithm is a novel metaheuristic algorithm introduced in 2017 [14]. It mimics the behavior of male bowerbirds constructing nests to attract females. Females judge males based on the nest decorations and the males' display behaviors. Male bowerbirds use both instinct and imitation of other males for nest building.

Both male and female bowerbirds select a bower based on an assigned probability, calculated by Equation (7). The fitness value $F_i$ for the i-th solution is obtained from Equation (8) [14].

$$prob_i = \frac{fit_i}{\sum_{n=1}^{NB} fit_n} \tag{7}$$

$$fit_i = \begin{cases} \frac{1}{1+f(x_i)}, f(x_i) \geq 0 \\ 1+\left|f(x_i)\right|, f(x_i) < 0 \end{cases} \tag{8}$$

where $f(x_i)$ is the cost function value of the i-th position, and NB is the number of bowers. Changes in each bower's position are determined by the probability in each cycle of the algorithm, calculated by Equation (9).

$$x_{ik}^{new} = x_{ik}^{old} + \lambda_k \left( \left( \frac{x_{ik} + x_{elite}}{2} \right) - x_{ik}^{old} \right) \tag{9}$$

where $x_{ik}$ is the k-th member of the $x_i$ , and j is chosen randomly based on the probabilities obtained. The elite position, which is the bower with the highest fitness in the current iteration, determines the step size calculated by Equation (10).

$$\lambda_k = \frac{\alpha}{1+p_j} \tag{10}$$

Here, $\alpha$ is a constant, and $p_j$ is the probability obtained from Equation (7) for the elite bower.

## 4. EVALUATION

Timely and rapid diagnosis plays a significant role in treating and reducing the risk of mortality. This study aims to assist physicians in diagnosing leukemia using predictive systems combining SVM with evolutionary algorithms. SVM is regarded as one of the best methods for disease diagnosis in predictive domains. It is a supervised learning method that maximizes the distance to the closest point from both classes to achieve better classification performance on test data. We compared SVM combined with optimization algorithms such as Cuckoo Optimization Algorithm (COA), Bowerbird Optimization Algorithm (SBO), Firefly Algorithm (FA), and Harmony Search Algorithm (HS) using the same training method. After obtaining and recording the results in Table 3, we concluded that the combination with the Bowerbird Algorithm provides significantly better and more accurate results than other algorithms.

**Table 3.** Simulation results comparing different methods

| Method | Accuracy | Sensitivity | Specificity |
|---|---|---|---|
| **SVM-SBO** | **0.6957** | **0.7552** | **0.6448** |
| **SVM** | 0.6513 | 0.6210 | 0.6760 |
| **HS** | 0.6359 | 0.7280 | 0.5667 |
| **GA** | 0.6889 | 0.7163 | 0.6634 |
| **FA** | 0.6598 | 0.7280 | 0.6004 |
| **COA** | 0.6359 | 0.6773 | 0.5843 |

## 5. DISCUSSION AND CONCLUSION

In this study, we employed the Support Vector Machine (SVM) classification method combined with the Bowerbird Optimization Algorithm (SBO-SVM) to analyze survival status and predict mortality in patients. To optimize the SVM parameters, the novel Bowerbird Optimization Algorithm was utilized. Initially, the input data was tested using SVM alone, and subsequently, to enhance its performance, SVM was combined with the mentioned algorithms. A comparison of the obtained results revealed that the SVM based on the Bowerbird Algorithm, using accuracy, sensitivity, and specificity metrics, achieved superior results with values of 69.57%, 75.52%, and 64.48%, respectively, compared to other methods.

## Transparency Statement

The data supporting this study are available upon reasonable request to the corresponding author, subject to ethical and confidentiality considerations.

## Acknowledgments

We would like to express our gratitude to all individuals who contributed to this project.

## Declaration of Interest

The authors declare that they have no competing interests.

## Funding

This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

## REFERENCES

[1] Hajizadeh, E., & Haji Fathali, A. (2020). Mixture and non-mixture cured models in survival analysis of leukemia patients: A cohort study. Studies in Medical Sciences, 31(10), 802-812.

[2] Cancer Treatment Centers of America. (2018). Blood cancers types and treatment options. Retrieved from https://www.cancercenter.com/blood-cancers

[3] Najafi, R., Olfatifar, M., Razi, M., & Amiri, F. (2020). Analyzing the survival of patients with breast cancer with a multistate model. Iranian Quarterly Journal of Breast Disease, 13(2), 19-27. https://doi.org/10.30699/ijbd.13.2.19

[4] Safdarian, N., & Yousefian Dezfoulinejad, S. (2020). Mammographic image processing for classification of breast cancer masses by using support vector machine method and grasshopper optimization algorithm. Journal of Arak University of Medical Sciences, 23(2), 246-263. https://doi.org/10.32598/JAMS.23.2.4672.1

[5] Fradkin, D., & Muchnik, I. (2006). Support vector machines for classification. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 70, 13-20.

[6] Akrimi, J. A., Suliman, A., George, L. E., & Ahmad, A. R. (2014). Classification of red blood cells using support vector machine. In 2014 International Conference on Information Technology and Multimedia (ICIMU) (pp. 265-269). IEEE. https://doi.org/10.1109/ICIMU.2014.7066642

[7] Bai, G. M., & Venkadesh, P. (2021). Taylor-monarch butterfly optimization-based support vector machine for acute lymphoblastic leukemia classification with blood smear microscopic images. Journal of Mechanics in Medicine and Biology, 21(06), 2150041. https://doi.org/10.1142/S021951942150041X

[8] Jusman, Y., Hasanah, A. N., Purwanto, K., Kanafiah, S. N. A. M., Riyadi, S., Hassan, R., & Mohamed, Z. (2021). Comparison between support vector machine and K-nearest neighbor algorithms for leukemia images classification using shape features. In 2021 25th International Computer Science and Engineering Conference (ICSEC) (pp. 70-74). IEEE. https://doi.org/10.1109/ICSEC53205.2021.9684585

[9] Hosseini Tashnizi, S., Tehrani, M., & Toosali Farhi, M. (2013). Comparison of Cox regression model and artificial neural network in predicting survival of leukemia patients. Iranian Journal of Epidemiology, 9(2), 71-92.

[10] Rawat, J., Singh, A., Bhadauria, H. S., Virmani, J., & Devgun, J. S. (2017). Computer-assisted classification framework for prediction of acute lymphoblastic and acute myeloblastic leukemia. Biocybernetics and Biomedical Engineering, 37(4), 637-654. https://doi.org/10.1016/j.bbe.2017.07.003

[11] Hosseinzadeh Kashan, A., & Garoosi, F. (2020). Hybrid model binary ant colony algorithm and support vector machine (BACO-SVM) for feature selection and classification of bank customers with case study. Financial Management Strategy, 8(2), 71-92.

[12] Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 2(2), 121-167. https://doi.org/10.1023/A:1009715923555

[13] Hosseinzadeh Kashan, A., & Garoosi, F. (2020). Hybrid model binary ant colony algorithm and support vector machine (BACO-SVM) for feature selection and classification of bank customers with case study. Financial Management Strategy, 8(2), 71-92.

[14] Moosavi, S. H. S., & Bardsiri, V. K. (2017). Satin bowerbird optimizer: A new optimization algorithm to optimize ANFIS for software development effort estimation. Engineering Applications of Artificial Intelligence, 60, 1-15. https://doi.org/10.1016/j.engappai.2017.01.006