

RESEARCH

Open Access



Predicting suicidal behavior outcomes: an analysis of key factors and machine learning models

Mohammad Bazrafshan¹ and Kourosh Sayehmiri^{2*}

Abstract

Background Suicidal behaviors, which may lead to death (suicide) or survival (suicide attempt), are influenced by various factors. Identifying the specific risk factors for suicidal behavior mortality is critical for improving prevention strategies and clinical interventions. Predicting the outcomes of suicidal behaviors can help identify individuals at higher risk of death, enabling timely and targeted interventions. This study aimed to determine the critical risk factors associated with suicidal behavior mortality and identify an effective classification model for predicting suicidal behavior outcomes.

Materials and methods This study utilized data recorded in the suicidal behavior registry system of hospitals in Ilam Province. In the first phase, duplicate records were removed, and the data was numerically encoded via Python version 3.11; then, the data was analyzed using chi-square and Fisher's exact tests in SPSS version 22 software to identify the factors influencing suicidal behavior mortality. In the second phase, missing data were removed, and the dataset was standardized. Five binary classification algorithms were utilized, including Random Forest, Logistic Regression, and Decision Trees, with hyperparameters optimized using the area under the receiver operating characteristic curve (AUC) and F1 score metrics. These models were compared based on accuracy, recall, precision, F1 score, and AUC.

Results Among 3833 cases of suicidal behavior in various hospitals in Ilam Province, the results indicated that the method of suicidal behavior ($P < 0.001$), reason for suicidal behavior ($P < 0.001$), age group ($P < 0.001$), education level ($P < 0.001$), marital status ($P = 0.004$), and employment status ($P = 0.042$) were significantly associated with suicide. Variables such as the season of suicidal behavior, gender, father's education, and mother's education were not significantly related to suicidal behavior mortality. Furthermore, the random forest model demonstrated the highest area under the ROC curve (0.79) and the highest classification accuracy and F1 score on both the training data (0.85 and 0.2, respectively) and test data (0.86 and 0.31, respectively) for predicting suicidal behaviors outcomes among the models tested.

Conclusion This study identified key factors such as older age, lower education, divorce or widowhood, employment, physical methods, and socioeconomic issues as significant predictors of suicidal behavior outcomes. A combination of statistical models for feature selection and machine learning algorithms for prediction was used, with Random Forest showing the best performance. This approach highlights the potential of integrating statistical methods with machine learning to improve suicide risk prediction and intervention strategies.

Keywords Suicide, Suicide attempt, Suicidal behavior, Suicide risk factors, Machine learning, Classification algorithms

*Correspondence:
Kourosh Sayehmiri
kourosh86@gmail.com
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/>.

Background

In past literature, varying terminology has been used to describe different facets of suicidal and self-harming actions, which may lead to ambiguity and inconsistency in understanding outcomes. To ensure clarity and precision in this study, we will adopt the following terminology: “Suicide” refers to the act of intentionally ending one’s life, resulting in death. “Suicide Attempt” will denote non-fatal actions where the individual has the intent to die but survives. Lastly, the broader term “Suicidal Behavior” will encompass both fatal and non-fatal suicidal actions, capturing the full spectrum of these behaviors. This approach aims to enhance the consistency of interpretation throughout the article, following recent recommendations in the field [1–4].

Advancements in science and medicine have made many diseases preventable and treatable. However, in recent decades, the leading causes of death have changed, with suicide emerging as a significant global concern. Statistics indicate that nearly one million people worldwide die by suicide each year [5].

In the 19th century, suicidal behaviors were considered crimes in many countries around the world. However, in the 20th century, most of these laws were repealed. The reason for these widespread changes is that suicide and similar behaviors are now more often considered as symptoms of mental disorders and public health problems rather than as crimes or offenses. Nevertheless, in some countries, such as Iran, due to cultural and religious reasons, suicidal behavior, while not legally classified as a crime, still carries significant social stigma. This stigma often leads to underreporting of suicide statistics, resulting in figures that do not reflect the actual reality [6, 7].

Globally, suicide is the second leading cause of death among individuals aged 15 to 29 years (after traffic accidents) and the third leading cause of death in the 15 to 44 age group. The suicide rate varies across different countries, typically measured as the number of suicides per 100,000 people. The lowest rates range from 0 to 4.9. Tragically, in 2015, the vast majority—78%—of all suicides occurred in low- and middle-income countries (LMICs) [5, 8].

In Iran, it was estimated that in 2000, 6.6 per 100,000 people committed suicide. However, this figure has increased to 9.9 per 100,000 people in the last two decades. Additionally, according to a conservative estimate, over 198,000 people in Iran exhibited suicidal behavior in 2016. Thus, expanding knowledge and strategies for suicide prevention is deemed essential in Iranian society [9].

Numerous studies have shown that a wide variety of factors influence suicide and suicide attempts. Some of these factors include age, gender, marital status, seasonal patterns, employment status, education level, and

the educational background of the individual’s parents. Additionally, some studies suggest that these risk factors, alongside factors such as the reason for attempting suicidal behaviors and the method used, can help predict the outcomes of suicidal behaviors [10–13]. Despite advancements, a reliable method for accurately predicting the outcomes of suicidal behaviors or effectively categorizing patients by their risk levels still does not exist [14, 15]. Recent research highlights the growing need for models that not only assess the risk of suicide attempts but also predict the likelihood of death, as such models could significantly improve intervention strategies and potentially save lives by identifying high-risk individuals earlier [16].

In recent years, the use of machine learning models has increased in various fields, including healthcare, due to the absence of certain limitations in traditional classification models. In predicting the risk of suicide, machine learning models have performed better in most cases compared to theory-based models [17, 18].

Machine learning (ML) models, such as random forests (RF) and support vector machines (SVM), have shown strong potential in predicting suicidal behaviors by analyzing large healthcare datasets and identifying complex temporal patterns, making them useful in clinical settings. Following RF and SVM, neural networks have successfully analyzed text- and image-based data. However, for demographic and clinical feature-based classification, models like K-nearest neighbors (KNN), logistic regression, and decision trees have outperformed other algorithms in certain studies, showing strong predictive capabilities in suicide risk assessment [19–25]. However, these models face limitations, including the need for high-quality datasets and challenges with interpretability, as some models act like “black boxes” that clinicians struggle to understand. Additionally, ethical concerns about data privacy and reliance on algorithms in sensitive areas like suicide prevention hinder their broader use [15, 26]. Classification algorithms, commonly used in ML, predict categories for data points (e.g., benign vs. malignant tumors) by calculating probabilities [27].

In the present study, we aimed to identify the factors influencing suicide through the application of statistical methods for feature selection, followed by utilizing five binary classification algorithms to predict the outcomes of suicidal behavior. Ultimately, this study aims to propose the best predictive model for the outcome of suicidal behavior by combining statistical methods and machine learning techniques.

Methodology

This study was conducted to identify critical factors influencing suicidal behavior mortality and to develop predictive models using machine learning techniques. The study was carried out in two phases:

1. statistical analysis of demographic and social factors using traditional methods.
2. predictive modeling using machine learning algorithms.

Data collection

The dataset consisted of all recorded suicidal behaviors from hospitals across Ilam Province. In total, data from 8090 suicidal behaviors were included (Before deleting duplications). Given the seasonal distribution of suicidal behaviors, we chose not to split the dataset into training and testing sets randomly. Instead, we utilized a temporal approach, where data from 21st March 2017 to 20th March 2020 (corresponding to the Iranian calendar years 1396 to 1398) were used as the test set, while data from 21st March 2020 to 20th March 2023 (corresponding to the Iranian calendar years 1399 to 1401) served as the training set. This decision was made to preserve the seasonal patterns inherent in the data, as certain variables, such as weather conditions or holidays at the time of the suicidal behavior, could significantly influence suicidal behavior outcomes [12, 28].

By maintaining the temporal integrity of the dataset, we aimed to ensure that the training and test sets reflected similar distributions in terms of seasonal effects. This approach allows for a more accurate assessment of the model's performance on unseen data, reducing the risk of overfitting and enhancing the model's applicability to real-world scenarios.

The following variables were collected for analysis (based on the results and recommendations of previous related studies [29–32]):

- Age group.
- Gender.
- Marital status (never married, married, others (divorce or widowhood)).
- Education (illiterate vs. educated).
- Employment status (employed vs. unemployed).
- Mother's education (illiterate vs. educated).
- Father's education (illiterate vs. educated).
- Season of suicidal behavior (spring, summer, fall, winter).
- Suicidal behavior method (physical vs. chemical).

- Reason for Suicidal behavior (Mental or Physical, family disputes, Socioeconomic).
- Outcome (survived or death).

Data preprocessing

The data underwent several preprocessing steps before statistical and machine learning analyses. Duplicate records were removed using Python scripts, leaving 3833 records in the training set and 3935 in the test set. Missing values were addressed by removing incomplete records. Categorical variables such as gender, marital status, and suicide method were converted into numerical codes to facilitate machine learning algorithms.

Statistical analysis

In the first phase of the study, chi-square or Fisher's exact tests were performed using SPSS software (version 22) to identify statistically significant associations between the selected variables and the outcome of suicidal behavior (survived or death). Fisher's exact test was used for variables where more than 20% of the cells had an expected count of less than 5. Six out of the ten analyzed variables showed significant associations with suicidal behavior mortality, and these were carried forward to the next phase of analysis. It is important to note that the significance level for this phase was set at ($P < 0.05$).

Machine learning analysis

In the second phase, machine learning techniques were applied to predict suicidal behavior outcomes based on the six significant variables from the statistical analysis. We used Python programming language (version 3.11) and Scikit-Learn (version 1.5.0), Pandas (version 2.2.1), Numpy (version 1.26.4), and Matplotlib (version 3.4.3) libraries to design various models for predicting suicidal behavior outcomes. The following steps were performed:

1. Data Standardization: Data were standardized using Python's StandardScaler function to ensure that all features were on the same scale, improving the performance of machine learning algorithms.
2. Algorithm Selection: Five binary classification algorithms were selected based on prior research [19]:

- Random Forest (RF) [21, 33].
- Decision Tree (DT) [22, 34].
- Logistic Regression (LR) [23, 35].
- k-Nearest Neighbors (KNN) [24, 36].
- Support Vector Machine (SVM) [25, 37].

3. Hyperparameter Tuning: Each algorithm underwent hyperparameter tuning using Random Search [38], optimizing for the highest area under the receiver operating characteristic curve (AUC) score in the initial phase [39].
4. Class Weights Optimization: Following the initial hyperparameter tuning, Grid Search was used to determine the optimal class weights due to data imbalance (7479 deaths vs. 289 survivals across train and test sets), with the F1 score as the performance metric [40, 41].
5. Model Evaluation: The models were evaluated on both the training and test sets, with performance compared based on metrics such as AUC, accuracy, precision, recall, and F1 score.

Results

Overall, the first phase of this study examined data from 4013 cases of suicidal behaviors recorded in the Ilam Province registry system during the study period. Among these, 180 duplicate cases were removed, leaving 3833 cases for analysis (130 suicides and 3703 suicide attempts). It should be noted that patients who attempted suicide and their companions often do not have the physical and mental condition to cooperate adequately with the interviewer. Additionally, due to potential errors by the interviewer or data entry personnel, some information about each patient is often incompletely recorded in the registry system. Table 1 details the missing data for key variables recorded in the registry system for the 3833 cases analyzed in the first phase of this study. It should be noted that only cases with complete data for each variable were included in the statistical analysis.

Additionally, another issue related to the data of this study was the discrepancy in the mortality statistics from suicide between the hospital registry system and

the forensic medicine data (Fig. 1). This discrepancy occurred because many suicide victims died before reaching the hospital. As a result, these cases were transferred directly to forensic medicine without being recorded in the hospital registry. However, due to the lack of individual-level data in the information provided by the forensic department, the hospital registry system remains the most reliable option for analyzing suicide-related data.

More than 20% of the cells had expected counts less than 5 for age group and education variables, so Fisher's exact test was employed. For the other variables, the Chi-square test was applied. The results of chi-square and Fisher's exact test are shown in Table 2. This table shows that in the marital status variable, the category "others" included individuals who were either divorced or widowed. Regarding the method of suicidal behavior, methods were categorized as either 'chemical' or 'physical.' Chemical methods included poisoning by substances such as drugs, alcohol, narcotics, or pesticides. Physical methods included actions like hanging, shooting, stabbing, self-immolation, drowning, and falling.

According to this table, Suicide was significantly associated with age group ($P < 0.001$), education level ($P < 0.001$), marital status ($P = 0.004$), employment status ($P = 0.042$), suicide method ($P < 0.001$), and the reason for the suicide attempt ($P < 0.001$). However, no significant association was found with the season of the suicide attempt ($P = 0.62$), gender ($P = 0.19$), father's education level ($P = 0.421$), or mother's education level ($P = 0.515$).

After identifying significant variables using the Chi-square and Fisher's exact tests, these variables were retained for modeling with machine learning classification methods, and the remaining variables were discarded. Table 3 shows the first phase's univariate binary logistic regression results on significant variables.

Table 1 Frequency of complete and Missing Data for each variable in train dataset

Variable	Complete		Missing	
	Frequency	Percentage	Frequency	Percentage
Age group	3833	100.0%	0	0.0%
Gender	3833	100.0%	0	0.0%
Marital status	3794	99.0%	39	1.0%
Season	3833	100.0%	0	0.0%
Suicidal behavior method	3691	96.3%	142	3.7%
Reason for suicidal behavior	1470	38.4%	2363	61.6%
Education	3467	90.5%	366	9.5%
Employment status	3550	92.6%	283	7.4%
Father's education	610	15.9%	3223	84.1%
Mother's education	600	15.7%	3233	84.3%

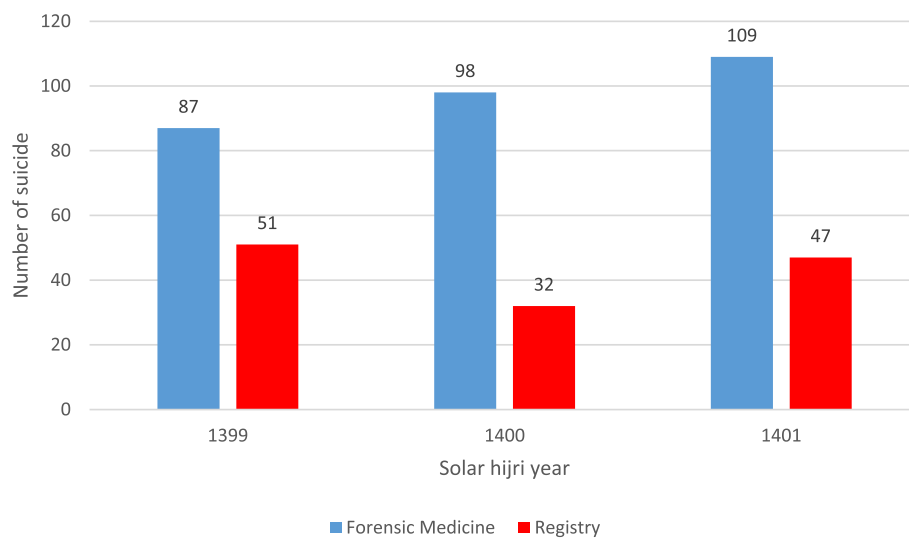


Fig. 1 Comparison of Suicide Mortality Statistics in the Hospital Registry and Forensic Medicine

The Receiver Operating Characteristic (ROC) curve and the area under it (AUC) are two metrics that can be used to compare the quality of classification algorithms. In the first step, we utilized the AUC as a score for tuning the hyperparameters. This value was highest in the Random Forest algorithm (0.79) (Fig. 2).

Next, various indices for comparing classification models are examined. The first index to be reviewed is classification accuracy. The Random Forest algorithm performed best in training and test data sets, with a value of 0.85 for the training data and 0.86 for the test data (Fig. 3).

The other index was classification precision. The value of this index in the training data for all algorithms ranged from 0.11 to 0.12. However, in the test data, the highest value belonged to the Random Forest algorithm (0.21) (Fig. 4).

Another index to be reviewed is recall. The highest value of this index in training data was related to the logistic regression algorithm (0.57), and in the test data, it was related to the decision tree algorithm (0.61) (Fig. 5).

In cases like the present study, where the two previous indices show conflicting results, the F1 score, which is the geometric mean of these two indices, can be helpful. The random forest algorithm had the highest F1 score in training and test data sets (0.2 and 0.31, respectively) (Fig. 6).

Discussion

This study examined the most significant factors influencing the suicide rate in Ilam Province. Globally, approximately 800,000 individuals die by suicide each year, which constitutes 1.5% of all deaths. Suicide is the

tenth leading cause of death in North America and the primary cause of death for individuals aged 15 to 24 worldwide [14]. Most individuals who survive a suicide attempt do not die from subsequent suicidal behaviors. This suggests that preventing methods leading to death could save many lives [42, 43].

Our research found that the highest likelihood of suicide occurred in individuals over 55 years. A significant association was observed between age and suicide mortality. These findings align with studies by Beghi M et al. [44] and Chen IM et al. [45], which identified high age as a critical risk factor for suicide. This may be related to comorbid conditions and fewer attempts involving non-lethal, attention-seeking methods among older adults. Interventions focused on enhancing social support and providing targeted mental health services for older adults could help reduce suicide rates in this vulnerable population.

Haghparast-Bidgoli H et al. [46] reported that an increase in the average educational level of populations in Iranian provinces significantly reduced suicide rates. Similarly, our study found the highest risk of suicidal behavior deaths in illiterate individuals. A significant relationship was observed between education and suicide, consistent with the study mentioned above and other research by Favril L et al. [30] and Denney JT et al. [47]. Implementing educational programs focused on mental health awareness and coping strategies could be beneficial in reducing suicide risk among illiterate populations.

Yoshimasu K et al. [48] described a minimal relationship between marital status and suicidal behavior mortality. However, our study, along with Haghparast-Bidgoli H et al. [46], demonstrated a significant association between

Table 2 Frequency of variables and Chi-Square/Fisher's exact test results

	Suicide Attempt	Suicide	Total	P-value
Age group				0.001>
Under 15 Years	129	4	133	
15 to 34 Years	2751	78	2829	
35 to 54 Years	709	33	742	
Above 54 Years	114	15	129	
Education				0.001>
Illiterate	94	12	106	
Educated	3266	95	3361	
Marital status				0.004
Never married	2094	60	2154	
Married	1541	64	1605	
Others	31	4	35	
Season				0.62
Spring	1070	36	1106	
Summer	931	36	967	
Autumn	781	30	811	
Winter	917	26	943	
Gender				0.19
Male	1806	71	1877	
Female	1897	59	1956	
Employment status				0.042
Unemployed	2683	82	2765	
Employed	750	35	785	
Father's education				0.421
Illiterate	300	12	312	
Educated	290	8	298	
Mother's education				0.515
Illiterate	331	11	342	
Educated	252	6	258	
Suicidal behavior method				0.001>
Chemical	3427	82	3509	
Physical	148	34	182	
Reason for suicidal behavior				0.001>
Mental disorders or physical illness	550	23	573	
Family disputes	794	27	821	
Socioeconomic problems	66	10	76	

divorce or widowhood and suicidal behavior mortality. Denney JT et al. [47] also found a link between divorce and suicide mortality in men, with no similar relationship observed in women. These differences may be rooted in the cultural context and the stigma associated with divorce in Iranian society. Interventions aimed at reducing the stigma surrounding divorce and providing psychological support for divorced or widowed individuals could help lower suicide in these vulnerable groups.

Although our study did not find a significant association between suicide and different seasons of the year, the highest number of suicides occurred in the spring and summer months. Previous studies by Villeneuve PJ et al. [28] and Likhvar V et al. [12] showed a correlation between warmer seasons and increased suicide. Likhvar V et al. [12] also noted an inverse relationship between holidays and suicide mortality rates.

Wu Y et al. [49] and Chen IM et al. [45] reported that suicide is higher in men than in women. Similarly, our

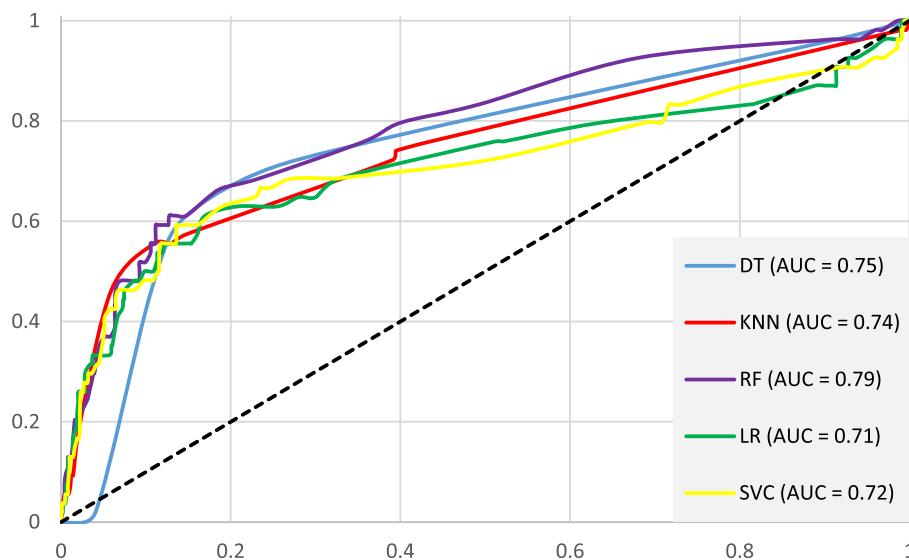
Table 3 Results of Univariate Logistic regression on significant variables in the First Phase

	Odds Ratio	CI 95%		P-Value
		Lower	Upper	
Marital status				
Never married	Reference			
Married	1.45	1.01	2.07	0.042
Others	4.5	1.54	13.16	0.006
Education				
Educated	Reference			
Illiterate	4.389	2.327	8.278	0.001>
Age group				
Under 15 Years	Reference			
15 to 34 Years	0.91	0.33	2.54	0.863
35 to 54 Years	1.5	0.52	4.31	0.45
Above 55 Years	4.24	1.37	13.15	0.012
Employment status				
Unemployed	Reference			
Employed	1.53	1.02	2.29	0.04
Reason for suicide				
Family disputes	Reference			
Mental disorder or physical illness	1.23	0.7	2.17	0.474
Socioeconomic Problems	4.46	2.07	9.6	0.001>
Suicidal behavior method				
Chemical	Reference			
Physical	9.6	6.23	14.79	0.001>

study found that suicide and the probability of suicidal behavior mortality were higher in men compared to women. However, a significant relationship between gender and suicide was not observed. Denney JT et al. [47] assessed the impact of various factors on suicide mortality by gender, revealing differences in risk factors between men and women, which may underscore the importance of gender in determining suicide-related factors.

Yoshimasu K et al. [48] found a minimal relationship between social factors such as marital status and employment status and suicide. In contrast, our study showed the highest likelihood of suicidal behavior deaths among employed persons, and a significant relationship was found between employment and suicide mortality. Moreover, studies by Graetz N et al. [50], Denney JT et al. [47], Haghparast-Bidgoli H et al. [46], and Favril L et al. [30] identified unemployment as a risk factor for suicide. The discrepancy between our findings and previous research may stem from specific occupational stressors or pressures unique to employed individuals in our study population, potentially exacerbated by cultural or economic factors in the region. To mitigate suicide risk among employed individuals, introducing workplace mental health programs, stress management workshops, and providing access to psychological support services could be effective measures.

Vidal-Ribas et al. [13] found a direct relationship between low parental education levels and suicide mortality. In our study, although a significant relationship between parental education and suicide mortality was not observed, the highest rates of suicide were found

**Fig. 2** ROC Curves of the Classification Algorithms

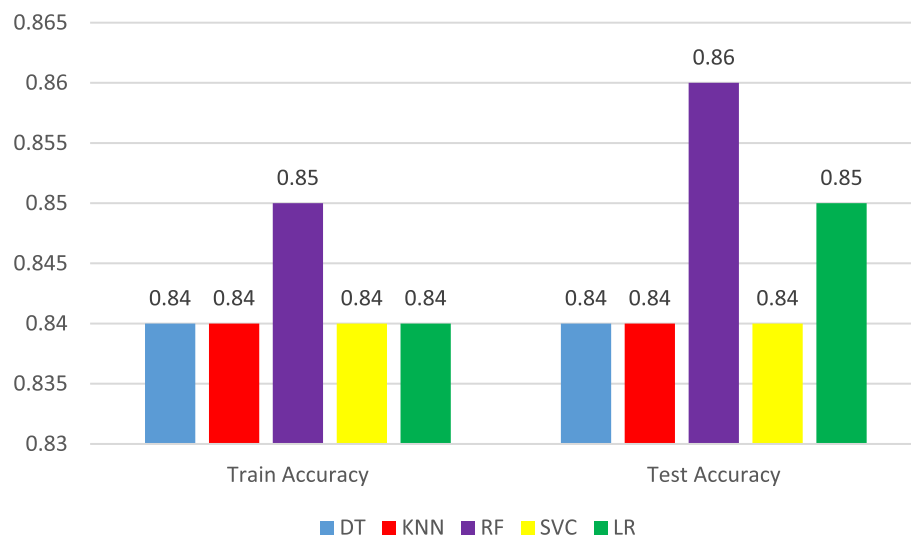


Fig. 3 Comparison of the Accuracies of the Classification Algorithms

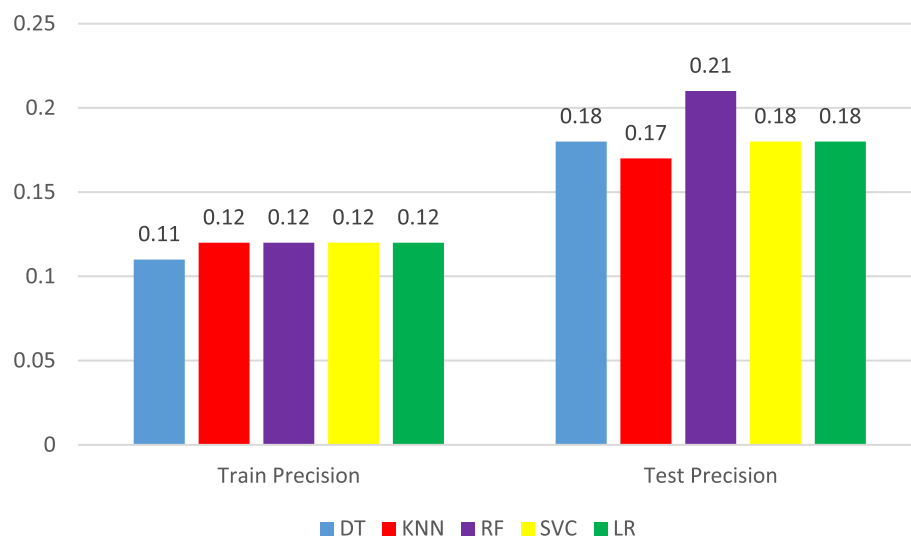


Fig. 4 Comparison of the Precisions of Classification Algorithms

among individuals with parents who had no education, consistent with the previous study.

Cai Z et al. [51] identified firearms, hanging, and drowning as the most lethal suicide methods. Elnour AA and Harrison J [52] similarly ranked firearms and hanging as the most fatal methods. Although our study did not analyze suicide methods separately, it compared physical and chemical methods and found that physical methods resulted in higher mortality rates, aligning with previous research. Implementing stricter regulations on access to highly lethal means, such as firearms and tools for hanging, could help reduce suicide rates.

Favril L et al. [30] reported that psychiatric disorders, physical illnesses, and socioeconomic factors significantly impact suicide rates. Our study also found a significant relationship between the cause of suicide attempts and mortality rates, with socioeconomic reasons being the most strongly associated with higher mortality. This finding is consistent with the study mentioned above. Addressing socioeconomic inequalities through targeted support programs, financial assistance, and access to mental health care for individuals facing economic distress may help reduce suicide rates.

Our findings align with theories emphasizing social isolation, like those of Durkheim and Joiner. Durkheim



Fig. 5 Comparison of the Recalls of Classification Algorithms

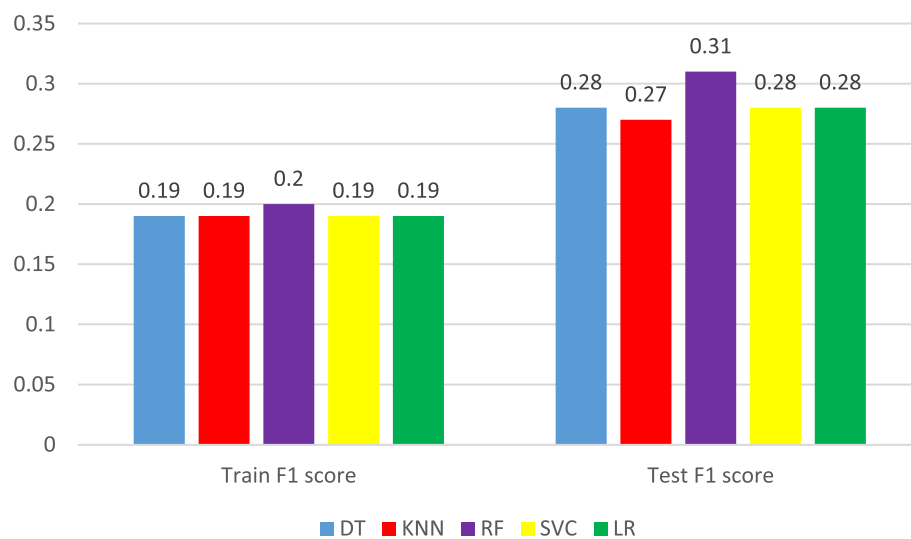


Fig. 6 Comparison of the F1 Scores of Classification Algorithms

highlights how a lack of social integration increases suicide risk [53], which corresponds with our results showing higher mortality among divorced, widowed, and socioeconomically distressed individuals. Similarly, Joiner's interpersonal theory posits that perceived burdensomeness and acquired capability for self-harm contribute to suicide [54]. Our findings reflect this, where employment and physical methods were linked to higher mortality, suggesting increased risk tolerance.

The random forest algorithm performed best in the study's second phase, achieving the highest AUC (0.79) and F1 score (0.31 on the test set). This aligns with previous research, such as that by Pigi A. et al. [19], Amini

P. et al. [55], and Ehtemam H et al. [15], highlighting the effectiveness of machine learning methods like Random Forest and SVM in handling complex, multidimensional data related to suicide risk prediction. Notably, the positive predictive values in the different models used in this research ranged from 0.17 to 0.21, which is significantly higher compared to the findings of Belsher BE et al. [56], where this index was reported to be less than 0.01 in most previous studies in the field of suicide. This finding is also consistent with the study of Schafer KM et al. [18].

Additionally, although Logistic Regression had the highest recall in the training set and Decision Tree had

the highest recall in the test set, Random Forest consistently outperformed the other algorithms regarding accuracy and precision, reaffirming its robustness in balancing false positives and false negatives. These results suggest that machine learning models, particularly ensemble methods, can effectively enhance the prediction of suicide outcomes when combined with traditional statistical methods. However, further refinement of models and inclusion of additional contextual data, such as detailed psychiatric histories or socioeconomic status, could improve predictive performance. Moreover, real-world implementation of these models in clinical settings may help professionals assess suicide risk more effectively, enabling timely interventions and potentially reducing suicidal behavior mortality.

A critical strength of the study lies in the relatively large dataset and the application of machine learning models, an innovative approach underutilized in this field. However, several limitations must be acknowledged. The high proportion of missing data, reliance on self-reported information from patients or their families, the lack of data on individuals who did not present to hospitals, and the regional nature of the dataset, which may limit generalizability, are essential constraints that should be considered when interpreting the results.

Conclusion

In conclusion, this study successfully identified significant demographic and social factors associated with suicidal behavior mortality in Ilam Province. By integrating statistical analysis and machine learning techniques, we highlighted vital patterns and risk factors, providing a comprehensive understanding of the contributors to suicide.

Future research and policy efforts should focus on validating the effectiveness of these models using more extensive datasets and additional variables. Exploring the potential of other algorithms, including ensemble models, will be essential in improving prediction accuracy. Furthermore, implementing targeted interventions for high-risk groups—for example, older adults, illiterate individuals, and those experiencing socioeconomic distress—can be crucial.

Additionally, the application of predictive algorithms could play a vital role in the early detection and prevention of suicide, opening promising avenues for future public health strategies.

Acknowledgements

The Ilam University of Medical Science has supported this study. The manuscript is extracted from the first author's professional doctorate thesis.

Clinical trial number

Not applicable.

Authors' contributions

Bazrafshan M. cleaned and analyzed data and wrote the manuscript. Sayehmiri K. assisted in the study design and data collection and supervised the procedures. Both authors read and approved the final manuscript.

Authors' information

Bazrafshan M. is a general physician and an Ilam University of Medical Sciences graduate. He presented the idea of this project due to his activity in Python programming and his interest in expanding the use of machine learning in medicine. Sayehmiri K. is a professor of biostatistics and is currently working in the Department of Statistics, Faculty of Health, Ilam University of Medical Sciences. He has a long history of research in various fields, including psychiatry and especially suicide.

Funding

This study had no funding.

Data availability

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

The ethics committee of Ilam University of Medical Sciences approved the thesis from which this study was extracted. The ethics number was "IR.MEDI-LAM.REC.1401.179".

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Medical Doctor, Faculty of Medicine, Ilam University of Medical Sciences, Ilam, Iran. ²Department of Biostatistics, Faculty of Health, Ilam University of Medical Sciences, Ilam, Iran.

Received: 29 August 2024 Accepted: 8 November 2024

Published online: 21 November 2024

References

- O'Connor E, Gaynes B, Burda BU, Williams C, Whitlock EP. Screening for suicide risk in primary care: A systematic evidence review for the US Preventive Services Task Force. 2013. PMID: 23678511.
- Carrasco-Barrios MT, Huertas P, Martín P, Martín C, Castillejos MC, Petkari E, et al. Determinants of suicidality in the European general population: a systematic review and meta-analysis. *Int J Environ Res Public Health*. 2020;17(11):4115.
- Hill NT, Robinson J, Pirkis J, Andriessen K, Krysinaka K, Payne A, et al. Association of suicidal behavior with exposure to suicide and suicide attempt: a systematic review and multilevel meta-analysis. *PLoS Med*. 2020;17(3):e1003074.
- Lovero KL, Dos Santos PF, Come AX, Wainberg ML, Oquendo MA. Suicide in global mental health. *Curr Psychiatry Rep*. 2023;25(6):255–62.
- Organization WH. Suicide worldwide in 2019: global health estimates. 2021. ISBN: 9789240026643
- Wu KC-C, Cai Z, Chang Q, Chang S-S, Yip PSF, Chen Y-Y. Criminalisation of suicide and suicide rates: an ecological study of 171 countries in the world. *BMJ open*. 2022;12(2):e0049425.

7. Mirdamadi M. How does the death conscious culture of Iran affect experiences of depression? *Cult Med Psychiatry*. 2019;43(1):56–76.
8. Bachmann S. Epidemiology of suicide and the psychiatric perspective. *Int J Environ Res Public Health*. 2018;15(7):1425.
9. Kiani Chalmardi A, Rashid S, Honarmand P, Tamook F. A structural test of the interpersonal theory of suicide model in students. *Contemporary psychology*. Biannual J Iran Psychol Association. 2018;13(1):50–61.
10. Kim S, Park J, Lee H, Lee H, Woo S, Kwon R et al. Global public concern of childhood and adolescence suicide: a new perspective and new strategies for suicide prevention in the post-pandemic era. *World J Pediatr*. 2024;20(9):872–900.
11. Qin P, Syeda S, Canetto SS, Arya V, Liu B, Menon V, et al. Midlife suicide: a systematic review and meta-analysis of socioeconomic, psychiatric and physical health risk factors. *J Psychiatr Res*. 2022;154:233–41.
12. Likhvar V, Honda Y, Ono M. Relation between temperature and suicide mortality in Japan in the presence of other confounding factors using time-series analysis with a semiparametric approach. *Environ Health Prev Med*. 2011;16:36–43.
13. Vidal-Ribas P, Govender T, Sundaram R, Perlis RH, Gilman SE. Prenatal origins of suicide mortality: a prospective cohort study in the United States. *Translational Psychiatry*. 2022;12(1):14.
14. Ropper AH, Seena Fazel MD, Runeson B. MD Ph D N Engl J Med. 2020;382:266–74.
15. Ehtemam H, Sadeghi Eshahlani S, Sanaei A, Ghaemi MM, Hajesmaeel-Gohari S, Rahimisaleh R, et al. Role of machine learning algorithms in suicide risk prediction: a systematic review-meta analysis of clinical studies. *BMC Med Inf Decis Mak*. 2024;24(1):138.
16. Seyed-salehi A, Fazel S. Suicide risk assessment tools and prediction models: new evidence, methodological innovations, outdated criticisms. *BMJ Ment Health*. 2024;27:e300990.
17. Lee W, Lee J, Woo S-I, Choi SH, Bae J-W, Jung S, et al. Machine learning enhances the performance of short and long-term mortality prediction model in non-ST-segment elevation myocardial infarction. *Sci Rep*. 2021;11(1):12886.
18. Schafer KM, Kennedy G, Gallyer A, Resnik P. A direct comparison of theory-driven and machine learning prediction of suicide: a meta-analysis. *PLoS ONE*. 2021;16(4):e0249833.
19. Piloni A, Delvecchio G, Turtulici N, Madonna D, Pietrini P, Cecchetti L, et al. Machine learning and the prediction of suicide in psychiatric populations: a systematic review. *Translational Psychiatry*. 2024;14(1):140.
20. Nordin N, Zainol Z, Noor MMH, Chan LF. Suicidal behaviour prediction models using machine learning techniques: a systematic review. *Artif Intell Med*. 2022;132:102395.
21. Saravanan N, Moheshkumar G, Shaid VM, Purushothman S, Sanjai VG, editors. Accurate Prediction and Detection of Suicidal Risk using Random Forest Algorithm. 2024 4th International Conference on Pervasive Computing and Social Networking (ICPCSN); 2024. p. 287–92. <https://doi.org/10.1109/ICPCSN62568.2024.00053>.
22. Bae S-M. The prediction model of suicidal thoughts in Korean adults using decision tree analysis: a nationwide cross-sectional study. *PLoS ONE*. 2019;14(10):e0223220.
23. Su C, Aseltine R, Doshi R, Chen K, Rogers SC, Wang F. Machine learning for suicide risk prediction in children and adolescents with electronic health records. *Translational Psychiatry*. 2020;10(1):413.
24. Aslan H, Yilmaz AB, Jeong N, Lee S, Choi C, editors. Prediction of number of suicidal people based on KNN. 2022 International Conference on Electronics, Information, and Communication (ICEIC); 2022. p. 1–4. <https://doi.org/10.1109/ICEIC54506.2022.9748557>.
25. Indrawan G, Sudiarsa I, Agustini K, Sariyasa S. Smooth support vector machine for suicide-related behaviours prediction. *Int J Electr Comput Eng*. 2018;8(5):3399.
26. Boudreaux ED, Rundensteiner E, Liu F, Wang B, Larkin C, Agu E, et al. Applying machine learning approaches to suicide prediction using healthcare data: overview and future directions. *Front Psychiatry*. 2021;12:707916.
27. Sidey-Gibbons JA, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol*. 2019;19:1–18.
28. Villeneuve PJ, Huynh D, Lavigne É, Colman I, Anisman H, Peters C, et al. Daily changes in ambient air pollution concentrations and temperature and suicide mortality in Canada: findings from a national time-stratified case-crossover study. *Environ Res*. 2023;223:115477.
29. Nie J, O'Neil A, Liao B, Lu C, Aune D, Wang Y. Risk factors for completed suicide in the general population: a prospective cohort study of 242, 952 people. *J Affect Disord*. 2021;282:707–11.
30. Favril L, Yu R, Geddes JR, Fazel S. Individual-level risk factors for suicide mortality in the general population: an umbrella review. *Lancet Public Health*. 2023;8(11):e868–77.
31. Favril L, Yu R, Uyar A, Sharpe M, Fazel S. Risk factors for suicide in adults: systematic review and meta-analysis of psychological autopsy studies. *BMJ Ment Health*. 2022;25(4):148–55.
32. Chau K, Kabuth B, Chau N. Gender and family disparities in suicide attempt and role of socioeconomic, School, and Health-related difficulties in early adolescence. *Biomed Res Int*. 2014;2014(1):314521.
33. Breiman L. Random forests. *Mach Learn*. 2001;45:5–32.
34. Quinlan JR. Induction of decision trees. *Mach Learn*. 1986;1:81–106.
35. Stoltzfus JC. Logistic regression: a brief primer. *Acad Emerg Med*. 2011;18(10):1099–104.
36. Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. *Am Stat*. 1992;46(3):175–85.
37. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20:273–97.
38. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. *J Mach Learn Res*. 2012;13:281–305.
39. Ling CX, Huang J, Zhang H, editors. AUC: a better measure than accuracy in comparing learning algorithms. *Advances in Artificial Intelligence: 16th Conference of the Canadian Society for Computational Studies of Intelligence, AI 2003, Halifax, Canada, June 11–13, 2003, Proceedings 16*; 2003. Springer.
40. Bakirarar B, Elhan AH. Class weighting technique to deal with Imbalanced Class Problem in Machine Learning: Methodological Research. *Türkiye Klinikleri Biyoistatistik*. 2023;15(1):19–29.
41. Salmi M, Atif D, Oliva D, Abraham A, Ventura S. Handling imbalanced medical datasets: review of a decade of research. *Artif Intell Rev*. 2024;57(10):273.
42. Daigle MS. Suicide prevention through means restriction: assessing the risk of substitution: a critical review and synthesis. *Accid Anal Prev*. 2005;37(4):625–32.
43. Yip PS, Caine E, Yousuf S, Chang S-S, Wu KC-C, Chen Y-Y. Means restriction for suicide prevention. *Lancet*. 2012;379(9834):2393–9.
44. Beghi M, Rosenbaum JF, Cerri C, Cornaggia CM. Risk factors for fatal and nonfatal repetition of suicide attempts: a literature review. *Neuropsychiatr Dis Treat*. 2013;8:1725–36.
45. Chen I-M, Liao S-C, Lee M-B, Wu C-Y, Lin P-H, Chen WJ. Risk factors of suicide mortality among multiple attempters: a national registry study in Taiwan. *J Formos Med Assoc*. 2016;115(5):364–71.
46. Haghparast-Bidgoli H, Rinaldi G, Shahnavaz H, Bouraghi H, Kiadaliri AA. Socio-demographic and economics factors associated with suicide mortality in Iran, 2001–2010: application of a decomposition model. *Int J Equity Health*. 2018;17(1):1–7.
47. Denney JT, Rogers RG, Krueger PM, Wadsworth T. Adult suicide mortality in the United States: marital status, family size, socioeconomic status, and differences by sex. *Soc Sci Q*. 2009;90(5):1167–85.
48. Yoshimasu K, Kiyohara C, Miyashita K, Hygiene SRGJS. Suicidal risk factors and completed suicide: meta-analyses based on psychological autopsy studies. *Environ Health Prev Med*. 2008;13:243–56.
49. Wu Y, Schwebel DC, Huang Y, Ning P, Cheng P, Hu G. Sex-specific and age-specific suicide mortality by method in 58 countries between 2000 and 2015. *Injury prevention*. 2021;27(1):61–70.
50. Graetz N, Preston SH, Peele M, Elo IT. Ecological factors associated with suicide mortality among non-hispanic whites. *BMC Public Health*. 2020;20:1–12.
51. Cai Z, Junus A, Chang Q, Yip PS. The lethality of suicide methods: a systematic review and meta-analysis. *J Affect Disord*. 2022;300:121–9.
52. Elnour AA, Harrison J. Lethality of suicide methods. *Inj Prev*. 2008;14(1):39–45.

53. Durkheim E. Suicide: a study in sociology. Routledge; 2005.
54. Joiner TE. Why people die by suicide. Harvard University Press; 2005.
55. Amini P, Ahmadiania H, Poorolajal J, Amiri MM. Evaluating the high risk groups for suicide: a comparison of logistic regression, support vector machine, decision tree and artificial neural network. *Iran J Public Health*. 2016;45(9):1179.
56. Belsher BE, Smolenski DJ, Pruitt LD, Bush NE, Beech EH, Workman DE, et al. Prediction models for suicide attempts and deaths: a systematic review and simulation. *JAMA Psychiatry*. 2019;76(6):642–51.

Publisher's Note

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