Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 5, October 2025, pp. 3756~3764
ISSN: 2302-9285, DOI: 10.11591/eei.v14i5.9837

Explainable artificial intelligence for multiclass prediction model of suicide attempt

Noratikah Nordin[1], Mohd Halim Mohd Noor[2], Zurinahni Zainol[3], Chan Lai Fong[4], Ryna Imma Buji[5]
[1]School of Communication, Universiti Sains Malaysia, Penang, Malaysia
[2]School of Computer Sciences, Universiti Sains Malaysia, Penang, Malaysia
[3]School of Computing and Informatics, Albukhary International University, Alor Setar, Malaysia
[4]Department of Psychiatry, Faculty of Medicine, National University of Malaysia, Cheras, Malaysia
[5]Department of Psychiatry, Hospital Mesra Bukit Padang, Kota Kinabalu, Malaysia

Article Info

Article history:

Received Jan 3, 2025
Revised Aug 21, 2025
Accepted Sep 1, 2025

Keywords:

Explainable artificial intelligence
Health informatics
Machine learning
Multiclass prediction
Suicide attempt

ABSTRACT

Suicide attempt prediction is a challenging classification problem that involves a variety of risk factors in individuals with various medical conditions. Accurate risk stratification prediction is hampered by the absence of reasons for those who have attempted suicide and developing prediction model is challenged to be explained. Therefore, this work aimed to develop a multiclass prediction model for suicide attempts and to use Shapley additive explanations (SHAP), an explainable artificial intelligence (XAI) method to analyze the prediction model for suicide attempts in explaining the decision of the model. The prediction model is trained using machine learning approaches, random forest (RF) and gradient boosting (GB), on a clinical dataset of patients with chronic diseases. GB demonstrated higher accuracy with 0.81 than RF with 0.78 for multiclass classification results (no risk, low risk, moderate risk, and high risk). By analyzing the SHAP explanations, clinicians can gain valuable insights into the factors contributing to suicide attempt predictions in patients with chronic diseases. This enhanced understanding can facilitate more informed and appropriate treatment decisions, potentially leading to improved patient outcomes and targeted interventions.

Corresponding Author:

Noratikah Nordin
School of Communication, Universiti Sains Malaysia
11800, Penang, Malaysia
Email: noratikahnordin@usm.my

1. INTRODUCTION

The global rise in suicide rates, which accounts for nearly one million deaths annually, highlights its status as a critical and escalating public health crisis [1]. A growing number of suicide cases are linked to a combination of risk factors, including stressful life events, economic problems, and mental health problems. Therefore, it has recently become important and critical to accurately assess individual suicide risk [2], [3]. In addition, understanding and analyzing risk factors are a critical component of suicidal behavior in an individual.

Suicidal behavior is defined by the study from Leo *et al.* [4] as behaviors that intentionally or unintentionally harm himself or herself; it includes suicide deaths, suicidal ideation, suicide attempts, and self-harm. There are multifactorial risk factors that affect suicidal behavior such as mental disorders, history of suicide attempts, financial problems, social problems, and substance abuse [3]. Assessing individual suicide risk based on predictions is important for clinicians when making decisions. Prediction of suicide attempts currently based on manual risk assessments such as the Columbia suicide severity rating scale (C-SSRS), the SADS person scale, and the Beck's scale for suicide ideation (BSSI) has been slow to improve and is not sufficiently reliable for predicting suicide attempts [5], [6]. Research by Belsher *et al.* [7] found that the development of new scales for suicide attempts is costly and time-consuming.

Recently, machine learning techniques have been explored for predicting suicide attempts in several studies [8]-[10] using decision trees, support vector machine and logistic regression. The main task in predicting suicide attempts is to classify individuals with suicide attempts (suicide attempter) and individuals without suicide attempts (non-suicide attempter). In simple terms, this is a binary classification task. Binary classification models can be effective in predicting suicide attempts, but there is a limitation that they oversimplify the complex phenomenon due to the influence of multiple risk factors, lead to inaccurate predictions of patients with suicide attempts [11]. Besides that, in the context of predicting suicide attempts, the results evaluated based on positive and negative screening are less beneficial and not precise for further management and treatment of suicidal behavior [12]. This is because treatment based on positive (yes) or negative (no) screening might lead to inefficient decision making and depends only on the risk factors that the individual presents. Therefore, identifying and monitoring individuals at varying risk of suicide helps clinicians to make accurate classification (low risk, moderate risk, and high risk).

Multiclass classification is a classification problem in which the goal is to predict to which of several classes a new data point belongs. In general, there are three or more possible classes in multiclass classification [13]. In the context of suicide attempt prediction, multiclass classification is defined as a classification problem for classifying individuals with suicide attempt that includes no risk, low risk, moderate risk, and high risk. Multiclass classification is a challenging, but also a very important problem in practice to predict suicide attempt to improve decision-making. Thus, there is a need for a multiclass prediction model to predict and analyze suicide attempts in the context of the complex occurrence of risk factors (suicide risk) with multiclass classification [6]. Multiclass classification is a suitable approach for predicting suicide attempts, as individuals may be at low, medium or high risk based on different risk factors, and can enable personalized interventions and support [14].

Explainable artificial intelligence (XAI) provides methods for making artificial intelligence (AI) predictions more transparent and understandable which has been demonstrated in numerous studies [15], [16]. XAI techniques include Anchors, local interpretable model-agnostic explanations (LIME), Shapley additive explanations (SHAP), partial dependence plots (PDP), and counterfactual explanations that have been used in different problems and application. In healthcare, XAI has been used in the analysis of patients with cardiovascular disease to explain the associated risk factors. However, there is lack of study that addresses the analysis of the healthcare prediction using XAI in the context of patients with chronic diseases. In fact, there is a lack of studies using XAI as a method to analyze and explain suicide attempts in detail in the context of populations and individual cases [17]. The XAI method in the context of suicide prediction is crucial for clinical interpretability and building trust in the prediction model for clinical decision making. SHAP method contribute to clinical interpretability in suicide attempt prediction by providing clear explanations, identifying key risk factors, supporting decision making and building confidence in the prediction model. By providing insight into feature importance, causal relationships and the rationale for model predictions, SHAP can help improve the effectiveness of XAI in predicting suicide attempts.

Therefore, the main contribution of this study is to develop a multiclass prediction model for suicide attempts and utilize the XAI method, SHAP, to analyze and explain predictions about individuals for clinical decision making and clinical interpretability. The novelty of the study is that it focuses on multiclass classification in predicting suicide attempts, which has not been explored and investigated in previous studies that are usually categorized into binary classification only. Risk stratification with different classes of suicide attempts is important for clinicians to create personalized treatment plans and make accurate decisions. Understanding suicidal risk for clinical interpretability is necessary for clinicians to understand the risk factors underlying a model's predictions in order to make informed decisions and provide effective interventions. Assist from the XAI method, specifically SHAP, can help build trust and transparency between clinicians, and facilitate its use in clinical practice.

This paper is organized as follows: section 2 discusses the framework for the multiclass prediction model based on the methodology and section 3 presents the results and analysis of the prediction model. Finally, section 4 summarizes the conclusions with future improvements.

2. METHOD

Figure 1 shows the framework of a multiclass prediction model with XAI for analyzing suicide risk in individuals with a suicide attempt and without a suicide attempt. The framework consists of five stages: i) data collection, ii) data pre-processing, iii) model development, iv) model evaluation, and v) clinical decision making. Each of the stages is discussed in details in order to develop a multiclass prediction model for suicide attempt prediction using the XAI method.

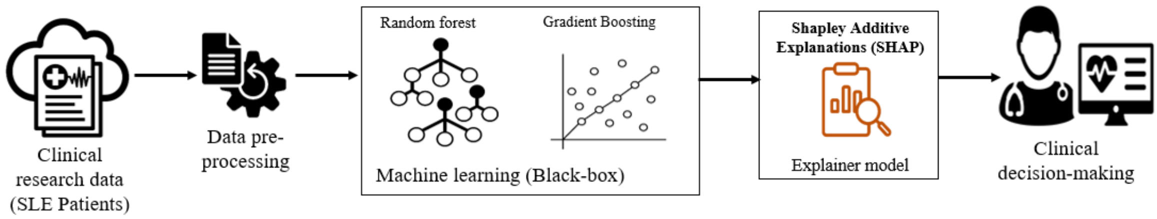

Figure 1. A multiclass prediction model for suicide attempt using the SHAP method

2.1. Data collection

The study focuses on a specific chronic disease that leads to suicide attempts, namely systemic lupus erythematosus (SLE). SLE is a chronic and complex multi-organ disorder categorized by autoimmune-mediated and systemic inflammation. Research by Karassa [18] found that the vulnerability of SLE patients is higher risk of suicide compared to other chronic diseases.

The clinical research data of SLE patients, consisting of 130 patients, were used to predict suicide attempts. The SLE dataset is based on the study by Buji *et al.* [19] and was collected from June to September 2015 at Universiti Kebangsaan Malaysia Medical Centre, Malaysia. Due to the scarcity and confidentiality of suicidal behavior dataset, this study focuses only on the patients in Malaysia. The original SLE dataset consists of 23 features including demographic information (gender, religion, ethnicity, marital status, and employment status), and psychosocial measures (patient health questionnaire-9 (PHQ9), SELENA-SLE disease activity index). The risk level of a suicide attempt is based on the C-SSRS which is categorized into four classifications: i) no risk, ii) low risk, iii) moderate risk, and iv) high risk. The four classifications of risk level of suicide attempt are evaluated by medical experts and clinicians.

2.2. Data pre-processing

The SLE dataset consists of socio-demographic information, psychosocial information, and assessment scores are structured in a table form. There are no missing values in the dataset. Scaling of the data is performed using the standardization technique and feature selection is completed to select the most important features for developing a multiclass prediction model. Feature selection is performed using recursive feature elimination (RFE). Based on RFE, a set of ten features is selected and retrieved from both machine learning models [8]. Table 1 shows the total of ten features identified in the dataset after feature selection and one output feature (risk level of suicide attempt: no risk, low risk, moderate risk, and high risk).

Table 1. Features in the clinical dataset

Features	Values
SLEDAI scores	0–19
Depressive disorder	0–no and 1-yes
History of suicide attempt	0–no and 1-yes
Past psychiatric treatment	0–no and 1-yes
Little interest (PHQ9)	0–Not at all, 1–Several days, 2–More than half the days, and 3–Nearly everyday
Feeling down (PHQ9)	0–Not at all, 1–Several days, 2–More than half the days, and 3–Nearly everyday
Sleeping trouble (PHQ9)	0–Not at all, 1–Several days, 2–More than half the days, and 3–Nearly everyday
Trouble concentrating (PHQ9)	0–Not at all, 1–Several days, 2–More than half the days, and 3–Nearly everyday
Feeling tired (PHQ9)	0–Not at all, 1–Several days, 2–More than half the days, and 3–Nearly everyday
Thoughts of self-harm (PHQ9)	0–No risk, 1–Low risk, 2–Moderate risk, and 3–High risk
Risk level of suicide attempt (Output)	0–No risk, 1–Low risk, 2–Moderate risk, and 3–High risk

The table provides an overview of the features used to develop the prediction model, which consist of the SLEDAI score (the clinical assessment scale for SLE disease), depressive disorder, history of suicide attempt, and past psychiatry treatment, which indicate the presence (1) or absence (0) of the condition in an individual. Based on feature selection also, five features were selected from the PHQ-9 to develop the prediction model, which are little interest, feeling down, sleeping trouble, trouble concentrating, feeling tired, and thought of self-harm. For each question, individuals are asked to indicate whether they have been bothered by the problem in the last two weeks. The measure is based on the frequency of experiencing these features with values ranging from not at all (0) to several days (1), more than half the days (2), and nearly everyday (3). All of these features are useful indicators of suicide risk to develop a reliable and accurate prediction model. The risk level of suicide attempt is the target feature of the prediction model. It classifies the overall risk of suicide attempt into four categories: no risk (0), low risk (1), moderate risk (2), and high risk (3).

2.3. Model development

In developing a multiclass prediction model for suicide attempt, two ensemble tree-based techniques are used: random forest (RF) and gradient boosting (GB). These two ensemble tree-based techniques are widely used in the medical domain for classification tasks [11], [20]. Both ensemble techniques are suitable for suicide attempt prediction as they perform well on tabular data, specifically clinical data, can handle different feature types (numerical and categorical) and are compatible with SHAP. Easy to implement and low computational effort are advantages when training the models. To avoid overfitting due to a small sample size, stratified three-fold cross-validation was employed and optimal hyperparameters were selected to improve the performance of the predictions.

The hyperparameters chosen for RF are the criterion, the maximum depth of the tree and the number of estimators (trees), and the hyperparameters for GB are the loss function, the learning rate and the number of estimators. These hyperparameters were chosen because they are important for the prediction model: the criterion for splitting the nodes in a decision tree, the depth of the tree to prevent overfitting by controlling the complexity of the model, and the number of decision trees, which affects the accuracy and stability of the model. Therefore, the chosen hyperparameters are the best combination to find the best configuration for this multiclass prediction model. The optimal hyperparameter settings are performed using GridSearchCV. The best parameters for RF are criterion (entropy), maximum depth of trees (4), number of estimators (200), and the best parameters for GB are loss (deviance), learning rate (0.01), and number of estimators (350).

After developing a multiclass prediction model, analysis of the prediction is performed using XAI. In this study, a post-hoc explainability approach is used to explain the prediction of suicide attempt in patients with chronic disease. SHAP is a post-hoc explainability approach that gives each input feature an important value for a given prediction [21]. SHAP is based on cooperative game theory principles and the important value is calculated using the Shapley value to determine the contribution of the players in the game to the final game outcome.

2.4. Model evaluation

The performance of the multiclass prediction model is evaluated in terms of accuracy, precision, specificity, and sensitivity. The confusion matrix is used in multiclass classification to measure the metrics and visualize the goodness of machine learning techniques. The measurement of the confusion matrix is taken from the study by Sokolova and Lapalme [22]. Classification accuracy is used to identify the ratio of correct predictions to the number of predictions, which is normally known as how often the classifier (model) is correct, while the precision is used to quantify the number of positive class predictions that actually belong to the positive class. Specificity is the proportion of individuals who test negative among all those who actually do not have the possibility of suicide attempt, while sensitivity is also known as recall which is the proportion of individuals who test positive among all those who are actually having the possibility of suicide attempt.

3. RESULTS AND DISCUSSION

We compare RF and GB as machine learning techniques for classifying suicide attempts. Three-fold cross-validation was performed to evaluate the prediction model. The performance results of the two machine learning models were evaluated in terms of accuracy, precision, specificity, and sensitivity, as shown in Table 2. These four measures are commonly used for the prediction of suicide attempts.

Table 2. Performance results of machine learning models

Performance results	RF	GB
Accuracy	0.78	0.81
Precision	0.75	0.80
Specificity	0.70	0.75
Sensitivity (recall)	0.78	0.80

Table 2 shows the performance of machine learning models for predicting suicide attempts in multiclass classification. Based on two models, GB revealed a good performance in the context of accuracy, precision, specificity, and sensitivity. As can be seen in Table 2, GB shows the highest accuracy (accuracy–0.81) compared to RF (accuracy–0.78). Precision results also shows that GB has a higher precision with 0.80, than RF with 0.75. In addition, GB shows a higher specificity (0.75) and higher sensitivity (0.80) compared to RF. Development of multiclass prediction model able to classify suicide attempt with a good performance using RF and GB. However, these models are known to be black boxes (difficult to interpret and explain). Therefore, SHAP is integrated into the model (GB) for understanding the prediction of suicide attempts in patients with SLE chronic disease. Using the Shapley value, feature importance has been generated and illustrated in Figure 2.

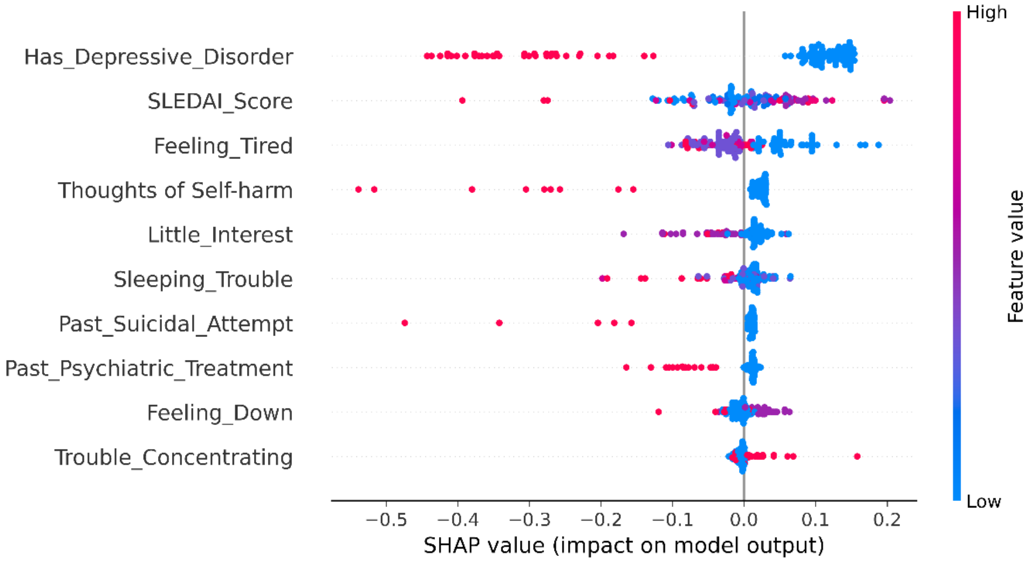

Figure 2. Important features of a multiclass prediction model

Figure 2 shows the importance of features in predicting suicide attempt in SLE patients based on the Shapley value. As can be seen in Figure 2, the feature values are represented from high to low values by red color and blue color respectively, indicating that the feature has strong and weak importance in predicting suicide attempt. Based on Figure 2, the presence of depressive disorder is the strongest and most significant feature in predicting suicide attempt with a high SHAP value, followed by SELENA-SLE score, feeling tired, thoughts of self-harm and little interest. Interestingly, three of the five features that have high importance in predicting suicide attempt are from the PHQ9 assessment. This shows that the items and features in the risk assessment tools can be used to predict suicide attempt rather than using demographic information only [23].

In addition to the importance of features contributing to the prediction of suicide attempt at population level in patients with SLE chronic disease, SHAP is able to provide analysis and explanations at the individual level. In the context of SHAP, this is referred to local explanations. In complex machine learning models, local explanations provide a local understanding of how and why a particular prediction was made. Using SHAP, the analysis and understanding of suicide attempt prediction are represented in a force plot, as illustrated in Figure 3.

Figure 3. Probability of a patient with a high risk level of suicide attempts

Figure 3 shows that a patient with SLE has a probability of 0.84 of being at high risk for suicide attempt. Based on the force plot, the features that increase prediction higher are represented by red arrows, while the features that decrease prediction lower are represented by blue arrows. As can be seen in Figure 3, the red arrow indicates that the patient has a depressive disorder and has sleeping trouble nearly everyday, which means a high risk for suicide attempt, while blue arrow represents by no feeling tired. This is an example of a patient with SLE at high risk for suicide attempt. As compared to Figure 4, the probability of SLE patient is at low risk for suicide attempt.

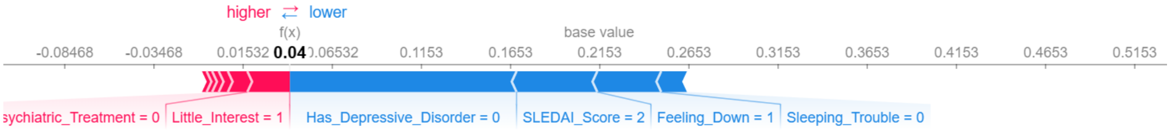

Figure 4. Probability of a patient with a low risk level of suicide attempts

Figure 4 shows that a patient with SLE has a low risk of suicide attempt with a probability of 0.04. This is because the feature/risk factors that contribute to a low prediction are the absence of a depressive disorder, a low SLEDAI score, and no sleep problems. Based on Figures 3 and 4, SHAP explains the prediction of suicide attempt comprehensively and individually. Based on the multiclass prediction model with XAI, the understanding of suicidal risk for suicide attempt can improve decision-making efficiently and effectively [17]. Therefore, the prediction model for suicide attempt can be analyzed successfully using an XAI method. This is important for clinicians to know and understand the features that contribute most to suicide risk in individual cases and also in the population. Integrating the prediction model into SHAP will increase the reliability and trustworthiness of predictions among clinicians in the medical domain, especially for clinicians and psychiatrists [24].

It has been shown that depressive disorders or depression in the SLE population are an important predictor of a high risk of suicide attempt. This is consistent with the results of previous studies using statistical analyses [7]. In the study by Macalli *et al.* [25], the group of students with depressive symptoms was found to be at high risk for suicidality. However, this study found that SLE patients with depressive disorders were most likely to attempt suicide in the future, showing that it is not only students who are prone to suicide attempts. In addition, the SLE patient with a higher SELENA-SLEDAI score has an impact on the prediction model for distinguishing the risk level for suicidal behavior. Therefore, clinicians are advised to assess suicide risk in patients with severe SLE disease activity. A multiclass prediction model showed comparable and better overall prediction performance when using multiclass classification problems to predict high risk for suicidal behavior.

The difficulty of the small sample size and the limited features available is the limitation of this study. A small sample size impairs the ability of the model to learn robust patterns and generalize to the broader population of individuals with chronic diseases at risk of suicide. As suicide risk is highly heterogeneous, a small sample might not adequately represent all relevant subgroups, and further acquiring a larger dataset would improve the model's ability to generalize diverse populations as well as represent more robust patterns of individuals with chronic disease. Limited features availability influenced the factors beyond standard clinical dataset. Enrichment with more diverse features, including social support, economic status, treatment response, and genetic biomarker data, will be explored in future research. In this study, the prediction model was developed using RF and GB only. Comparison with other models is limited as the study uses tabulated clinical data, which is often useful and appropriate for suicide attempt prediction model. The study focuses on the use of SHAP to explain the model's decisions that is integrated with SHAP, and both models are compatible with SHAP. This study is a first step in establishing the feasibility of using XAI for suicide prediction, with the intention that future work can extend the comparative analysis to more complex models. In fact, only SHAP was used as an XAI in this study to understand the prediction model. Other XAI such as LIME and Anchors, could be investigated with a larger clinical dataset to improve healthcare, especially in the context of suicidality.

4. CONCLUSION

In this study, we develop a multiclass prediction model for suicide attempt in patients with SLE chronic disease using an XAI method (SHAP). We found that GB achieved higher accuracy (0.81) in predicting suicide attempt compared to RF. The multiclass prediction model can classify four risk levels of suicide attempt (high risk, moderate risk, low risk, and no risk). With the different risk levels, clinicians can make a better decision. This study contributes to explain the model's decision which shows that depressive disorders are the most important features influencing individuals at high risk for suicide attempts. The clinical implications of multiclass prediction models can contribute to personalized treatment plans that allow clinicians to tailor treatment plans to individual patients and also allow them to target specific interventions to a particular group of suicide attempt patients. Indeed, multiclass prediction models can help diagnose suicide attempt patients with multiple possible outcomes (treatment options), leading to more accurate and timely treatment decisions. This study is a guide and basis for clinicians in identifying, classifying, and predicting individuals with suicide attempt for further treatment and management. Future work should focus on diverse patient populations, integrating other clinical data and developing a user-friendly clinical tool for real-time assessment and decision support.

ACKNOWLEDGMENTS

The authors would like to thank all parties that support and involve directly or indirectly into this research, especially Universiti Sains Malaysia, Pulau Pinang and Universiti Kebangsaan Malaysia Medical Centre, Kuala Lumpur. This manuscript is an upgraded version of a conference paper presented at the 2nd International Conference on Emerging Technologies and Sustainability in Kota Tangerang Selatan, Indonesia, from 31 August to 1 September 2024.

FUNDING INFORMATION

The authors also wish to express their gratitude to the Research Creativity and Management Office (RCMO), Universiti Sains Malaysia. This work was supported by a Universiti Sains Malaysia, Short Term Grant with Project No: R501-LR-RND002-0000001713-0000.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Noratikah Nordin	✓	✓	✓		✓	✓			✓	✓	✓			✓
Mohd Halim Mohd Noor		✓	✓			✓				✓	✓	✓		
Zurinahni Zainol	✓			✓						✓	✓	✓	✓	
Chan Lai Fong				✓	✓	✓	✓	✓		✓		✓		
Ryna Imma Buji				✓	✓		✓	✓		✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study as the data is collected from the article, doi: 10.1177/0961203317742711.

REFERENCES

[1] World Health Organization, *Live life: an implementation guide for suicide prevention in countries*. Geneva: World Health Organization, 2021, [Online]. Available: https://apps.who.int/iris/handle/10665/341726. (accessed: Sep. 01, 2022).
[2] E. D. Boudreaux *et al.*, "Applying Machine Learning Approaches to Suicide Prediction Using Healthcare Data: Overview and

Future Directions," *Front. Psychiatr.*, vol. 12, pp. 1-7, Aug. 2021, doi: 10.3389/fpsyt.2021.707916.
[3] J. C. Franklin *et al.*, "Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research.," *Psychol. Bull.*, vol. 143, no. 2, pp. 187–232, 2017, doi: 10.1037/bul0000084.
[4] D. D. Leo *et al.*, "International study of definitions of English-language terms for suicidal behaviours: a survey exploring preferred terminology," *BMJ Open*, vol. 11, no. 2, pp. 1-10, Feb. 2021, doi: 10.1136/bmjopen-2020-043409.
[5] B. Runeson, J. Odeberg, A. Pettersson, T. Edbom, I. J. Adamsson, and M. Waern, "Instruments for the assessment of suicide risk: A systematic review evaluating the certainty of the evidence," *PLOS ONE*, vol. 12, no. 7, pp. 1-13, Jul. 2017, doi: 10.1371/journal.pone.0180292.
[6] S. Velupillai *et al.*, "Risk Assessment Tools and Data-Driven Approaches for Predicting and Preventing Suicidal Behavior," *Front. Psychiatry*, vol. 10, pp. 1-8, Feb. 2019, doi: 10.3389/fpsyt.2019.00036.
[7] B. E. Belsher *et al.*, "Prediction Models for Suicide Attempts and Deaths: A Systematic Review and Simulation," *JAMA Psychiatr.*, vol. 76, no. 6, pp. 642-651, Jun. 2019, doi: 10.1001/jamapsychiatry.2019.0174.
[8] N. Nordin, Z. Zainol, M. H. M. Noor, and C. L. Fong, "A comparative study of machine learning techniques for suicide attempts predictive model," *Health Informatics J.*, vol. 27, no. 1, pp. 1-2, Jan. 2021, doi: 10.1177/1460458221989395.
[9] C. G. Walsh, J. D. Ribeiro, and J. C. Franklin, "Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning," *J. Child Psychol. Psychiatry*, vol. 59, no. 12, pp. 1261–1270, Dec. 2018, doi: 10.1111/jcpp.12916.
[10] T. A. Burke, R. Jacobucci, B. A. Ammerman, L. B. Alloy, and G. Diamond, "Using machine learning to classify suicide attempt history among youth in medical care settings," *J. Affect. Disord.*, vol. 268, pp. 206–214, May. 2020, doi: 10.1016/j.jad.2020.02.048.
[11] T. A. Burke, B. A. Ammerman, and R. Jacobucci, "The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: A systematic review," *J. Affect. Disord.*, vol. 245, pp. 869–884, Feb. 2019, doi: 10.1016/j.jad.2018.11.073.
[12] R. C. O'Connor and G. Portzky, "Looking to the Future: A Synthesis of New Developments and Challenges in Suicide Research and Prevention," *Front. Psychol.*, vol. 9, pp. 1-14, Nov. 2018, doi: 10.3389/fpsyg.2018.02139.
[13] A. Mustaqeem, S. M. Anwar, and M. Majid, "Multiclass Classification of Cardiac Arrhythmia Using Improved Feature Selection and SVM Invariants," *Comput. Math. Methods Med.*, vol. 2018, pp. 1–10, 2018, doi: 10.1155/2018/7310496.
[14] A. Thieme, D. Belgrave, and G. Doherty, "Machine Learning in Mental Health: A Systematic Review of the HCI Literature to Support the Development of Effective and Implementable ML Systems," *ACM Trans. Comput.-Hum. Interact.*, vol. 27, no. 5, pp. 1–53, Oct. 2020, doi: 10.1145/3398069.
[15] J. Amann, A. Blasimme, E. Vayena, D. Frey, and V. I. Madai, "Explainability for artificial intelligence in healthcare: a multidisciplinary perspective," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, pp. 1-9, Dec. 2020, doi: 10.1186/s12911-020-01332-6.
[16] T. A. A. Abdullah, M. S. M. Zahid, and W. Ali, "A Review of Interpretable ML in Healthcare: Taxonomy, Applications, Challenges, and Future Directions," *Symmetry*, vol. 13, no. 12, pp. 1-28, Dec. 2021, doi: 10.3390/sym13122439.
[17] N. Nordin, Z. Zainol, M. H. M. Noor, and L. F. Chan, "An explainable predictive model for suicide attempt risk using an ensemble learning and Shapley Additive Explanations (SHAP) approach," *Asian J. Psychiatry*, vol. 79, p. 103316, Jan. 2023, doi: 10.1016/j.ajp.2022.103316.
[18] F. B. Karassa, "Suicide attempts in patients with systemic lupus erythematosus," *Ann. Rheum. Dis.*, vol. 62, no. 1, pp. 58–60, Jan. 2003, doi: 10.1136/ard.62.1.58.
[19] R. I. Buji *et al.*, "Suicidal ideation in systemic lupus erythematosus: *NR2A* gene polymorphism, clinical and psychosocial factors," *Lupus*, vol. 27, no. 5, pp. 744–752, Apr. 2018, doi: 10.1177/0961203317742711.
[20] B. F. Darst, K. C. Malecki, and C. D. Engelman, "Using recursive feature elimination in random forest to account for correlated variables in high dimensional data," *BMC Genet.*, vol. 19, no. Suppl 1, pp. 1-6, Sep. 2018, doi: 10.1186/s12863-018-0633-8.
[21] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *arXiv*, 2017, doi: 10.48550/arXiv.1705.07874.
[22] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, Jul. 2009, doi: 10.1016/j.ipm.2009.03.002.
[23] S. Kim, H.-K. Lee, and K. Lee, "Which PHQ-9 Items Can Effectively Screen for Suicide? Machine Learning Approaches," *Int. J. Environ. Res. Public. Health*, vol. 18, no. 7, pp. 1-10, Mar. 2021, doi: 10.3390/ijerph18073339.
[24] R. C. Kessler, R. M. Bossarte, A. Luedtke, A. M. Zaslavsky, and J. R. Zubizarreta, "Suicide prediction models: a critical review of recent research with recommendations for the way forward," *Mol. Psychiatry*, vol. 25, no. 1, pp. 168–179, Jan. 2020, doi: 10.1038/s41380-019-0531-0.
[25] M. Macalli *et al.*, "A machine learning approach for predicting suicidal thoughts and behaviours among college students," *Sci. Rep.*, vol. 11, no. 1, pp. 1-8, Dec. 2021, doi: 10.1038/s41598-021-90728-z.

BIOGRAPHIES OF AUTHORS

Noratikah Nordin received a Bachelor Degree in Information Studies in 2014 and a Master degree of Science in Knowledge Management in 2016, both from Universiti Teknologi MARA (UiTM), Malaysia, and a Ph.D. in Computer Science from Universiti Sains Malaysia, Malaysia in 2024. She is currently a lecturer at the School of Communication, Universiti Sains Malaysia, teaching in the Digital Communication Program. Her research interests include data mining, machine learning, health informatics, and digital media analytics. She has published several papers in international journals and conferences. She can be contacted at email: noratikahnordin@usm.my.

Mohd Halim Mohd Noor is an Associate Professor at the School of Computer Sciences, Universiti Sains Malaysia. He received a Bachelor degree in Computer and Information Engineering from the International Islamic University Malaysia (UIAM), Malaysia, in 2004, the Master of Science in Electrical and Electronic Engineering from Universiti Sains Malaysia, Malaysia, in 2009, and a Ph.D. in Computer Systems Engineering from the University of Auckland, New Zealand. His research interests include machine learning, deep learning, computer vision, and pervasive computing. He currently focuses on problems in human activity recognition, including segmentation, feature learning, and prediction. In addition, he has published numerous papers in international journals, conferences and book chapters. He can be contacted at email: halimnoor@usm.my.

Zurinahni Zainol received a Bachelor's degree in Computer Science from Universiti Kebangsaan Malaysia (UKM)/Universiti Teknologi Mara (UiTM), a Master's Degree in Artificial Intelligence from Universiti Sains Malaysia (USM) and a Ph.D. in Computer Science from University of Hull, United Kingdom in 2012. She was previously an Associate Professor and served as a Deputy Dean for Academic, Career and International Affairs at the School of Computer Sciences, Universiti Sains Malaysia (USM). She is currently a Professor at the School of Computing and Informatics, Albukhari International University, Kedah. She has published in international journals, conferences, and book chapters. Her research interests include data modeling, XML database schema, optimization algorithm, and information retrieval in multidisciplinary fields, such as medical, health, biological, and education data. She can be contacted at email: zurinahni.zainol@aiu.edu.my.

Chan Lai Fong is currently Professor of Psychiatry and Consultant Psychiatrist at the National University of Malaysia (UKM). She trained in psychiatry at the National University of Malaysia and completed a Clinical Fellowship in Mood and Anxiety Disorders at the University of Toronto, followed by a Master of Science in Affective Neuroscience at Maastricht University. She was awarded the 2017 De Leo Fund Award by the International Association of Suicide Prevention (IASP) for outstanding research on suicidal behaviours carried out in developing countries. She current areas of research focus include suicide prevention among healthcare workers, media safe messaging and pesticide suicide prevention. She is the Malaysian representative of IASP, a member of the Malaysian National Technical Working Group on Suicide Prevention and the scientific committees of the IASP World Congresses from 2017 until 2021. She can be contacted at email: laifchan@hctm.ukm.edu.my.

Ryna Imma Buji received a Bachelor of Medicine and Bachelor of Surgery (MBBS) from the University of Malaya in 2005. After completing her Advanced Masters in Psychiatry at the National University of Malaysia (UKM) in 2016, she worked as a clinical psychiatrist at Hospital Mesra Bukit Padang, Kota Kinabalu, Sabah. She has over 15 years of experience as a medical expert in the field of psychiatry. She has a special interest in child and adolescent psychiatry and has conducted several programs for parents with special needs. Currently, she has been appointed as the Deputy Medical Director of Hospital Mesra Bukit Padang, Sabah. She can be contacted at email: rynaimma@gmail.com.