DOI: https://doi.org/10.53350/pjmhs22163765

ORIGINAL ARTICLE

Role of Machine Learning in Short-Listing Future Suicidal Candidates

SUMMYIA SADIA[1], FARIHA TARIQ[2], PERVAIZ ZARIF[3], KISHWAR NAHEED[4], ABID RAFIQ[5], HAFIZ MUHAMMAD FAISAL[5]
[1]*Department of Forensic Medicine, Sargodha Medical College, Sargodha-Pakistan*
[2]*Department of Forensic Medicine, King Edward Medical University, Lahore-Pakistan*
[3]*Department of Forensic Medicine, PGMI, Lahore-Pakistan*
[4]*Department of Forensic Medicine, Punjab Medical College Faisalabad-Pakistan*
[5]*Department of CS and IT, University of Sargodha, Sargodha -Pakistan*
Correspondence to: Summyia Sadia, Email: summyiaabid@gmail.com, Cell:+92-333-77745442

ABSTRACT
Suicide is an important issue to address, especially in rural areas. Rural areas are facing unique challenges such as poor health care facilities, lack of awareness, financial constraints and many more for such matters.
Aims: To find the social, educational and medical attributes which may lead a person to deliberate self harm.
Study Design: Retrospective study.
Methodology: Total 100 cases of suicidal attempts taken from DHQ teaching hospital Sargodha from (June to December) 2019. We considered all the suicidal and self harm cases admitted through emergency and medicolegal clinic. Moreover cases less than 9 years of age and autopsy cases were excluded. All the cases were analysed with reference to 10 features (age, gender, locality, education, marital status, duration of stay in hospital, treatment given, prevalence of psychiatric disorder, suicidal attempts, the method used for suicidal attempt).
Statistical analysis: ML models work on numeric data. However the dataset we collected have categorical features except age. The most used method for such purpose is python get dummies function. The get dummies() function is used to convert categorical variable into dummy/indicator variables.
Results: In this study, more preponderance of suicidal attempts at age less than 40 in males which shows the development of more mature attitude with increasing age.
Conclusion: It was concluded that suicide is influenced by many personal factors that cannot be shared publicly on social platforms. However, such information can be used to lower the risk of suicide attempts in rural areas.
Keywords: Machine Learning, Suicide Prevention, Suicide Detection and Ideation.

INTRODUCTION
Suicide is defined as a deliberate act of taking one's own life with prior knowledge and expectation of his actions resulting in fatal outcome[1]. In every 40 seconds, one precious life is taken by suicide globally[2]. According to World Health Organization (WHO), suicide kills about 800,000 people every year[3]. However, each event of suicide is preceded by many failed attempts resulting in self harm not sufficient to cause death of a person. Young people in (15-29) years have been seen attempting suicide as the second major cause of mortality globally, furthermore, the third leading cause of death in late teens(15-19 years)[3].

Poisoning is an important health hazard and one of leading causes of mortality[4-6]. Self-poisoning with pesticides accounts for 14–20% of global suicides, an estimated 110,000–168,000 deaths each year[7], down from an estimated 371,000 in the late 1990s[8]. Ingestion of pesticides, hanging,firearms and stab wounds are few other methods of suicide globally[9,10]. Overdose of over the counter medicines, intake of common household poisons in developing world[11,12]. Level of education, gender, marital status, financial status, personality type, stress of any kind, psychological disorders, area of living and societal impacts influence suicidal behavior[13].

The problem is more severe in rural Asian communities, where a wide range of agricultural highly hazardous pesticides (HHPs) and wheat pills are easily available within the home and from shops[8,14]. In Pakistan, there is an accumulative incidence of increase in both suicide and deliberate self harm at an alarming rate in the recent years[15]. According to WHO, 2.9 % of every 100,000 persons commit suicide in Pakistan making it a total of 5500 deaths in the year 2016[3]. Earlier, the figures were more threatening i.e, 7.5 % per 100,000 in year 2012. Every suicide creates numerous circumstances for family, society and the entire country having lasting effects on lives of those who are left behind the deceased.

Machine Learning Classifiers:
A. K-Nearest Neighbor: K-Nearest Neighbor (KNN) has been used in predicting medical diagnosis, suicide risk and suicide detection[16,17] [23]–[27]. KNN originally formulated by Fix and Hodges (1951), and it is a lazy learning, non-parametric, machine learning algorithm that works by classifying instances based on their similarity[18]. Lazy learning quotes to the method of generalization for the training data. KNN falls under the supervised learning variant of machine learning[19] where the system is trained by known examples. The system examines a set of data consisting of observed values and classifications, and uses that data to predict classifications for a new set of observed values.
B. Logistic Regression: Logistic Regression (LR) is a most popular classifier used for various medical research purposes[20,21]. It is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes. In simple words, the dependent variable is binary in nature having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no)[22,23].
C. Decision Tree: Decision trees are one of the first inherently non-linear machine learning techniques. It generates an approximate solution via greedy, top-down, recursive partitioning. Another advantage of decision trees is that they can easily deal with categorical variables. Moreover DT is recommended when there are predetermined set of attributes, the response is discrete and disjunctive and graphical results are required[24]. There are three components of DT: decision nodes, branches, and leaves. The direction begins at the node and extends to the leaf, which connects the features. The tree is a disjunction of these connections and these disjunctions separate the branch population into sets with the same likelihood of events. At each stage, the disjunctions cause the highest possible predictive power. The graphical feature presentation makes ease of interpretation and allowing to different alternative[24].
Objectives: To find the social, educational and medical attributes which may lead a person to deliberate self harm (DSH).

METHODOLOGY
Present retrospective study enrolled 100 cases of suicidal attempts taken from DHQ teaching hospital Sargodha from (June to December) 2019. We considered all the suicidal and self harm cases admitted through emergency and medicolegal clinic. Moreover cases less than 9 years of age and autopsy cases were excluded. All the cases were analysed with reference to 10 features (age, gender, locality, education, marital status, duration

of stay in hospital, treatment given, prevalence of psychiatric disorder, suicidal attempts, the method used for suicidal attempt). All this information was recorded on Performa.

Statistical Analysis: Pre-processing is first step to ensure ML model work properly on the dataset as expected. ML models work on numeric data. However the dataset we collected have categorical features except age. All features such as education, marital status, presence of psychiatric disorder, location, marital status and gender which we used to feed the network are converted to binary format. The most used method for such purpose is python get dummies function. The get dummies () function is used to convert categorical variable into dummy/indicator variables.

B. Train Test Split: For any machine learning model to produce good result is to split the data in (70-30) ratio. The 70 % of data used for training the model and remaining 30 % used for testing the accuracy of model on unknown instances.

RESULTS

The parameters used during training of classifiers were listed in Table-1.

Table-1: Parameters for training of classifiers

Model	Parameters
Logistic Regression	LogisticRegression(C=1.0, class weight=None, dual=False, fit intercept=True, intercept scaling=1, l1 _ratio=None, max iter=100, multi class=auto, n jobs=None, penalty=l2, random state=None, solver=lbfgs, tol=0.0001, verbose=0, warm start=False)
KNN	KNeighborsClassifier(algorithm=auto, leaf size=30, metric=minkowski, metric params=None, n jobs=None, n _neighbors=3, p=2, weights=uniform)
Decision Tree	DecisionTreeClassifier(ccp alpha=0.0, class weight=None, criterion=gini, max depth=3, max features=None, max leaf nodes=None, min impurity decrease=0.0, min impurity split=None, min samples leaf=1, min samples _split=2, min weight _fraction leaf=0.0, presort=deprecated, random state=1, splitter=best)

The accuracy of three classifiers was depicted in Table-2 and was presented as percentage.

Table 2: Knn, Naive Bayes And Logistic Regression Accuracy

Model	Accuracy (%)
KNN	91.30
Decision Tree	91.30
Logistic Regression	88.40

The study showed more preponderance of suicidal attempts at age less than 40 in males which shows the development of more mature attitude with increasing age. Most of the suicide attempts are above the age of 30. Furthermore, lack of an education for most of the respondents was seen in Figure-1.

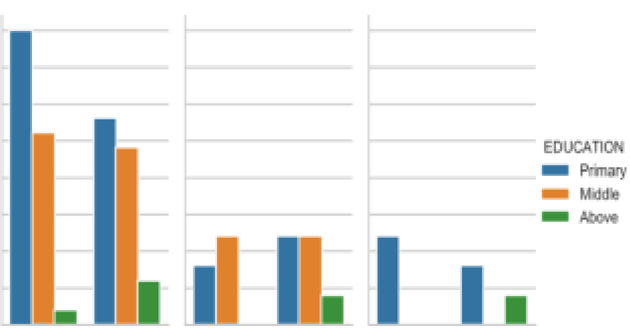

Fig 1: Educational status of respondants

The source of intake (severe) has been taken mostly by females (Figure 2&3). We can infer from the results, the respondents who have committed by marital status.

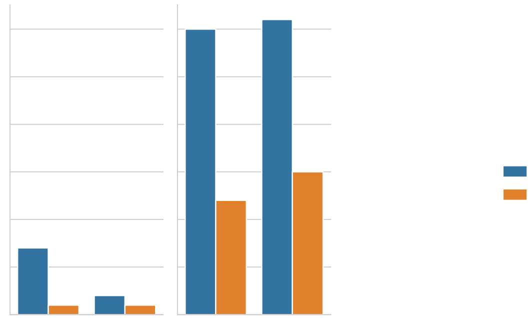
Fig 2: Mode of intake based on education marital status

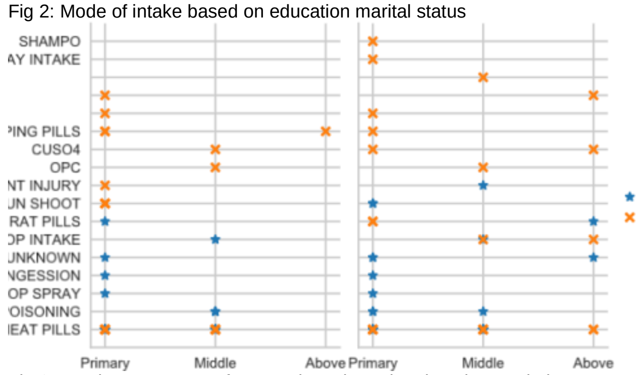

Fig 3: Previous attempts of respondents based on location marital status

On x-axis we can see, the females with presence of psychiatric disorder converge more towards severely toxic intake and they are mostly from Rural areas seen in Figure 4.

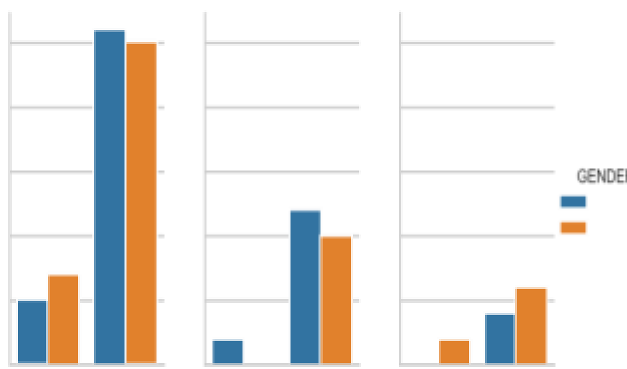

Fig 4: Distribution of respondents in takes by presence psychiatric disorder

DISCUSSION

Moreover, most of suicides attempts are committed around (30-40) years of age for married respondents, which can be related to more demanding marital responsibilities, pressure of kids,financial sufferings etc. Unmarried women usually tried self harm at early age(teen and early adulthood) however ,there are lesser incidents in elderly women due to their commitments and involvements in the family. Results were in line with many previous studies done[11-14].We can see from Figure-1 that education level of the individuals mostly primary and middle and they belong to rural areas side-by city Sargodha, and used (Mild) intake method for committing suicide such as wheat pills, shampo and rat pills which are easily accessible in rural areas as shown by previous studies too[7-9]. Moreover, from Figure 2 and 3 mostly married had attempted (Mild) intake method, and they had previous attempt history.

The earlier and rigorous treatment modalities (antidote, gastric lavage and symptomatic medicines) are related to better survival rates and lesser fatalities. Similarly, longer hospital request. This recommends more strict strategies for hospital admissions and follow-ups in suicidal cases. A strong database setting can help the medical teams in keeping them in loop for future surveillance.

Our study also showed the relation between level and method of committing suicide similar to one previous study[20]. We collected data from DHQ teaching hospital, hence most of respondents belong to lower class. We can see females with primary or lower level education took wheat-pills and organophosphorous compounds for committing self harm, while those who are educated middle and above took alternate methods like CuSO4, shampoo, acids and drugs. Similarly those committing suicide for the first time used cheap and easily available drugs like wheat pills and pesticides while the experienced ones used the methods with less easy recognition and more lethal effects like acid ingestion,copper sulphate or other household poisons[22].

Limitations: Our study had limitations like financial constraints, lack of resources, genetic workup and short duration of study.

CONCLUSION

It was concluded that suicide is influenced by many personal factors that cannot be shared publicly on social platforms. However, such information can be used to lower the risk of suicide attempts in rural areas. We used machine learning classifiers to classify attempts into two categories, 1: (the first attempts), 2: (twice and more). We believe such intelligence can be used by giving a proper rehabilitation and counselling at an appropriate time soon after the first incidence.

Authors' Contribution: SS&FT: Conceptualized the study, analyzed the data, and formulated the initial draft.

PZ&KN: Contributed to the proof reading.

AR&HMF: Collected data.

Acknowledgements: I am thankful to Allah and all my colleagues for their help.

REFERENCES

1. D. De Leo, S. Burgis, J. M. Bertolote, A. J. Kerkhof, and U. Bille-Brahe, "Definitions of suicidal behavior: Lessons learned from the who/euro multicentre study," Crisis, vol. 27, no. 1, pp. 4–15, 2006.
2. F. Judd, A.-M. Cooper, C. Fraser, and J. Davis, "Rural suicide—people or place effects?," Australian and New Zealand Journal of Psychiatry, vol. 40, no. 3, pp. 208–216, 2006.
3. W. H. Organization, World health statistics 2016: monitoring health for the SDGs sustainable development goals. World Health Organization, 2016.
4. W. H. Organization et al., "Who mortality database. geneva: Who," World Health Org, 2003.
5. W. H. Organization, The World Health Report 2001: Mental health: new understanding, new hope. World Health Organization, 2001.
6. J. M. Bertolote and A. Fleischmann, "A global perspective in the epidemiology of suicide," Suicidologi, vol. 7, no. 2, 2002.
7. M. Eddleston, "Patterns and problems of deliberate self-poisoning in the developing world," Qjm, vol. 93, no. 11, pp. 715–731, 2000.
   Bose, C. Sandal Sejbaek, P. Suganthy, V. Raghava, R. Alex, J. Muliyil, and F. Konradsen, "Self-harm and self-poisoning in southern india: choice of poisoning agents and treatment," Tropical Medicine & International Health, vol. 14, no. 7, pp. 761–765, 2009.
8. D. Owens, C. Wood, D. C. Greenwood, T. Hughes, and M. Dennis, "Mortality and suicide after non-fatal self-poisoning: 16-year outcome study," The British Journal of Psychiatry, vol. 187, no. 5, pp. 470–475, 2005.
9. S. Mahmood, "The pakistan penal code (xlv of 1880), vol. ii, sections 300–374," Legal Research Centre: Lahore, 1989.
   T. Cheng and C.-S. Lee, "Suicide in asia and the far east," The international handbook of suicide and attempted suicide, pp. 29–48, 2000.
10. T. S. Ali, G. Arnad´ ottir, and A. Kulane, "Dowry practices and their´ negative consequences from a female perspective in karachi, pakistan—a qualitative study," Health, vol. 5, no. 7D, p. 84, 2013.
11. R. E. Gearing and D. Lizardi, "Religion and suicide," Journal of religion and health, vol. 48, no. 3, pp. 332–341, 2009.
12. M. Eddleston, A. Karunaratne, M. Weerakoon, S. Kumarasinghe, M. Rajapakshe, M. Rezvi Sheriff, N. A. Buckley, and D. Gunnell, "Choice of poison for intentional self-poisoning in rural sri lanka," Clinical Toxicology, vol. 44, no. 3, pp. 283–286, 2006.
13. M. M. Khan and M. Prince, "Beyond rates: the tragedy of suicide in pakistan," Tropical Doctor, vol. 33, no. 2, pp. 67–69, 2003.
14. M. Adinkrah, "Anti-suicide laws in nine african countries: Criminalization, prosecution and penalization.," African Journal of Criminology & Justice Studies, vol. 9, no. 1, 2016.
15. M. R. Phillips, G. Yang, Y. Zhang, L. Wang, H. Ji, and M. Zhou, "Risk factors for suicide in china: a national case-control psychological autopsy study," The Lancet, vol. 360, no. 9347, pp. 1728–1736, 2002.
16. B. H. Wagenaar, A. K. Hagaman, B. N. Kaiser, K. E. McLean, and B. A. Kohrt, "Depression, suicidal ideation, and associated factors: a cross-sectional study in rural haiti," BMC psychiatry, vol. 12, no. 1, p. 149, 2012.
    Karunarathne, D. Gunnell, F. Konradsen, and M. Eddleston, "How many premature deaths from pesticide suicide have occurred since the agricultural green revolution?," Clinical toxicology, vol. 58, no. 4, pp. 227–232, 2020.
17. S. Shazia, A. H. Wazir, H. ur Rashid, and Z. H. Khalil, "Three years analysis of poisoning cases of forensic medicine toxicological laboratory, khyber medical college, peshawar," Journal of Ayub Medical College Abbottabad, vol. 32, no. 2, 2020.
18. M. D. Singleton, L. M. Frey, A. Webb, and J. Cerel, "Public health surveillance of youth suicide attempts: challenges and opportunities," Suicide and Life-Threatening Behavior, vol. 50, no. 1, pp. 42–55, 2020.
19. J. M. Twenge and G. N. Martin, "Gender differences in associations between digital media use and psychological well-being: evidence from three large datasets," Journal of Adolescence, vol. 79, pp. 91–102, 2020.
20. S. A. Medjahed, T. A. Saadi, and A. Benyettou, "Breast cancer diagnosis by using k-nearest neighbor with different distances and classification rules," International Journal of Computer Applications, vol. 62, no. 1, 2013.
21. L. E. Peterson, "K-nearest neighbor," Scholarpedia, vol. 4, no. 2,p. 1883, 2009.