

Chapter 3

Machine Learning for Suicide Prediction and Prevention: Advances, Challenges, and Future Directions



Shirley B. Wang, Walter Dempsey, and Matthew K. Nock

In the 50 years from 1965 to 2015, researchers published over 350 papers examining variables that might enhance the prediction of youth suicidal thoughts and behaviors (STBs). Unfortunately, a meta-analysis of this work found predictive accuracy has *not* increased over time, but rather, it has remained just slightly above chance for all outcomes (Franklin et al., 2017). One possible explanation is that the vast majority of studies have focused on single risk factors from the same few domains (e.g., mental health) combined in simple ways (e.g., multiple linear regression) across extended timeframes (e.g., >10 years). To address these limitations, researchers recently have turned to novel machine learning methods, which can model high-dimensional datasets with potentially complex nonlinear relationships among risk factors and outcomes. These studies have so far demonstrated superior performance of machine learning compared to traditional statistical methods (Linthicum et al., 2019). For instance, machine learning models have provided high accuracy in predicting suicide attempts in large, nationally representative surveys (García de la Garza et al., 2021), US Army soldiers (Kessler et al., 2017), and patients hospitalized for suicidal thoughts and behaviors (Wang et al., 2021). However, several outstanding questions remain regarding how to best build and implement machine learning models to guide clinical decision-making. In this chapter, we discuss key challenges at each step of the research process to provide recommendations for researchers, clinicians, and policy makers interested in machine learning for youth suicide prevention. Of note, we focus on broad, higher-level concepts throughout this chapter, rather than technical aspects of

S. B. Wang (✉) · M. K. Nock

Department of Psychology, Harvard University, Cambridge, MA, USA
e-mail: shirleywang@g.harvard.edu; nock@wjh.harvard.edu

W. Dempsey

Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA
e-mail: wdem@umich.edu

© The Author(s) 2022

J. P. Ackerman, L. M. Horowitz (eds.), *Youth Suicide Prevention and Intervention*, SpringerBriefs in Psychology,
https://doi.org/10.1007/978-3-031-06127-1_3

implementation and analysis, and direct interested readers to recent tutorials and textbooks for greater technical detail (Dwyer et al., 2018; Kuhn & Johnson, 2013).

Important Questions and Challenges

Data Collection

How researchers collect data influences the effectiveness of STB prediction. Choices made during data collection can significantly impact a model's accuracy. For instance, models using predictors that are *causes* of the outcome may be more deployable in other sites than models with predictors that are *effects* of the outcome (Piccininni et al., 2020), though model adjustments also remain important if site populations are very different from one another. In addition to predictor selection, researchers should carefully consider the timeframes of interest. Most existing youth STB prediction models have considered long follow-up periods (an average of 7.9 years for adolescents; Franklin et al., 2017), which do not reflect the timeframe of greatest clinical interest (i.e., risk of a patient attempting suicide in the next few days, weeks, or months), especially during periods of rapid emotional and cognitive development. Recent research harnessing advances in smartphone and wearable biosensor technology has enabled shorter-term risk prediction during these critical time periods (e.g., following psychiatric hospitalization) (Wang et al., 2021) demonstrating that despite the time- and effort-intensive nature of real-time monitoring studies, they can provide important data for STB prediction in high-risk time periods.

Model Building

Numerous machine learning algorithms have been applied in STB prediction, including regularized regression (e.g., elastic net), random forests, neural networks, and naive Bayes classifiers. A full review of these models is beyond the scope of this chapter, and we encourage readers to consult excellent reviews (Dwyer et al., 2018) and textbooks (James et al., 2013) for greater technical detail. It is worth noting that each approach has benefits and drawbacks, with complex nonlinear methods (e.g., random forests, neural networks) typically requiring more data to perform well and yielding higher prediction accuracy at the cost of lower interpretability, and vice versa for simpler linear methods (e.g., regularized regression). When choosing an algorithm, researchers should consider their ultimate goals, which could be (1) to maximize accuracy, (2) to interpret the logic of how each variable contributes to the prediction of outcomes, or (3) to identify potential targets for prevention and intervention efforts.

Once an algorithm has been selected, an important question is whether to consider missing data as a predictor in the model. Such an informative missingness approach has the potential to pick up on key contributors to suicide risk. For instance, in a sample of nearly 4,000 US Army soldiers, nonresponse to a question about suicidal thinking emerged as a particularly strong predictor of future suicide attempts (Nock et al., 2018). However, researchers should proceed with caution when using missingness as a predictor in machine learning models, as changes to study design would lead to changes in missing data patterns, and some evidence suggests it may also introduce bias into models that generalize poorly to new data (van Smeden et al., 2020). Following these decisions, researchers should split data into training and test datasets to reduce likelihood of overfitting and evaluate accuracy with multiple metrics for a complete understanding of model performance.

Model Implementation and Translation

As our ability to refine predictive models improves, they can be implemented in settings where youth with elevated suicide risk are most likely to present, such as healthcare settings. How models are best implemented is discussed here. Broadly speaking, there are three options. The first involves building a model and applying this exact model to new sites. This often is used for other health outcomes, such as eye diseases, cardiac abnormalities, and cancer (Ngiam & Khor, 2019). Benefits of this approach include faster implementation and model dissemination, while drawbacks include less tailoring to site characteristics that could influence predictive accuracy (e.g., population health status, prescribing patterns, billing code assignments). Another approach involves using the same modeling *approach* but training a new model at each new site. Across five US healthcare systems, a recent study using this approach found remarkably consistent accuracy for predicting suicide attempts (Barak-Corren et al., 2020). The third option offers a compromise: rather than build entirely new models or implement identical models across sites, researchers could use existing models to update models for new populations. This could involve shrinkage of a new model toward existing models or using information from previous models as priors at new sites.

A related concern in implementing machine learning models involves temporal drift. For example, it is unknown if a model built in 2020 would show similar accuracy in the same population in 2030. This challenge is perhaps best exemplified by the current worldwide COVID-19 pandemic. Many models built prior to COVID-19 may fail to adequately capture the importance and magnitude of current strong predictors of STBs, such as feelings of isolation (Fortgang et al., 2021). Thus, even after models are implemented clinically, they should be continually updated based on newly available data.

Using Models to Guide Clinical Decision-Making

Healthcare providers must also consider how to integrate information from machine learning models into their decision-making. A critical concern when working with high-risk patients is forecasting risk of suicide to make decisions about clinical care and need for hospitalization. The goal of building and implementing STB prediction models is not to replace clinical judgment, but rather to guide, support, and augment clinical decision-making. For instance, when faced with a decision about whether to hospitalize or discharge a patient who may be at risk for suicide, clinicians could consult predictions from a machine learning model, just as they may consult other members of the clinical care team.

However, two concerns that undermine use of models currently are the high rate of false positives *and* false negatives in STB prediction models to date. The problem of false positives has been noted as early as the 1980s (Pokorny, 1983) and continues to present challenges with integrating machine learning into clinical decision-making today. As psychiatric hospitalization is often the first-line intervention for individuals at imminent suicide risk, high false positive rates could risk unnecessary hospitalization for thousands of patients erroneously predicted to be at acute suicide risk annually. When hospitalization occurs in the absence of clinical need, this can have serious iatrogenic effects via increased distress, stigma, trauma (e.g., witnessing threatening/violent behavior from patients or staff), coercion, and loss of autonomy, particularly for involuntary hospitalizations (Ward-Ciesielski & Rizvi, 2020). Many hospitals are already overburdened, and false positives may compromise a hospital's ability to meet the needs of true positive cases. Failing to detect acute suicide risk when it exists (e.g., false negatives) is also highly concerning as they represent missed opportunities for timely and potentially lifesaving intervention. In light of these potentials for harm, machine learning models should be used to augment, not replace, clinical decision-making.

Ethics of Machine Learning for Youth Suicide Prediction

Accurate prediction of youth suicide is only useful insofar as there are effective STB prevention strategies. Unfortunately, we currently lack strong and universally effective interventions (Fox et al., 2020), and the common intervention of hospitalization has serious potential harms, including high suicide risk post-discharge. Crucially, we do not know if psychiatric hospitalization helps more people than it harms nor the precise effectiveness of hospitalization in preventing suicide (Large & Kapur, 2018). Thus, alongside research optimizing machine learning algorithms for STB prediction, there is a critical need to develop and disseminate effective and scalable STB interventions, particularly for youth (see Thomas et al., Chap. 15, this volume; Zullo et al., Chap. 8, this volume).

Regardless of these limitations, researchers and clinicians must know how to respond if a child or adolescent is predicted to be at high suicide risk. A recent Delphi study by Nock et al. (2021) including scientists, clinicians, ethicists, legal experts, and individuals with lived experience provided a consensus statement that individuals identified in a research context to be at high risk for suicide should (1) be contacted as soon as possible (including contact with parents), (2) receive an individualized safety plan, (3) receive additional risk assessment, and (4) receive personalized outreach rather than automated contact. Importantly, many experts discouraged calling 911 as a standard response, as police contact can result in elevated rates of physical force, trauma, and death, particularly for racial or ethnic minorities (Nock et al., 2021). We also note that simply contacting people predicted to be at high suicide risk is itself an intervention, the effects of which are unknown and worth investigating. Although this Delphi study was conducted in the context of real-time monitoring research studies, many principles may apply to ethical concerns of machine learning risk predictions. We encourage researchers, clinicians, and policy makers to continually update best-practice guidelines over time as more data and considerations become available.

Future Directions

In this chapter, we have outlined critical unanswered questions at every stage of the process from building to implementing machine learning models for youth suicide prevention. Clearly, there is much work to be done, and we believe that expertise is needed from multiple domains and perspectives, including psychology, psychiatry, and clinical practitioners, in addition to computer scientists, statisticians, ethicists, and those with lived experience. Collaborative science is essential for making meaningful progress especially in the challenging arena of predicting suicide risk.

In addition to data-driven machine learning methods, we also note the importance of strong theory in advancing STB prediction and prevention. Although there are many influential suicide theories, these have all been instantiated *verbally*, which renders them underspecified due to the inherent imprecision of language. Formalizing theories using mathematical and computational modeling can advance the prediction and prevention of suicide by identifying factors *causally* associated with STBs and potential targets for intervention (which can also be simulated to understand if, how, and why a treatment may be effective for reducing suicide risk).

Both theory- and data-driven computational work are crucial for youth STB prevention. Machine learning has revolutionized many fields of medicine over the past decade. To make similar progress, we need a better understanding of the causes of STBs, the effect of model predictions on clinical decision-making, external validation of models, best-practice ethical guidelines, and effective and scalable interventions. In addition, greater funding for suicide research is crucial for driving innovation and exploring the challenges described above. Whereas increased federal funding has led to declines in other leading causes of death (e.g., tuberculosis)

over the past century, funding for suicide research has lagged far behind, and the suicide rate today is nearly identical to what it was 100 years ago (Fortgang & Nock, 2021). Increased funding and policy to support continued research in prediction of youth suicide can provide critical information to inform the development and implementation of machine learning models to meaningfully reduce suicide in youth.

Funding Details Shirley Wang is supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1745303 and the National Institute of Mental Health under Grant F31MH125495. Walter Dempsey is supported by the National Institute of Drug Abuse under Grants R01DA039901 and P50DA054039.

References

- Barak-Corren, Y., Castro, V. M., Nock, M. K., Mandl, K. D., Madsen, E. M., Seiger, A., Adams, W. G., Applegate, R. J., Bernstam, E. V., Klann, J. G., McCarthy, E. P., Murphy, S. N., Natter, M., Ostasiewski, B., Patibandla, N., Rosenthal, G. E., Silva, G. S., Wei, K., Weber, G. M., ... Smoller, J. W. (2020). Validation of an electronic health record–based suicide risk prediction modeling approach across multiple health care systems. *JAMA Network Open*, 3(3), e201262. <https://doi.org/10.1001/jamanetworkopen.2020.1262>
- Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology*, 14(1), 91–118. <https://doi.org/10.1146/annurev-clinpsy-032816-045037>
- Fortgang, R. G., & Nock, M. K. (2021). Ringing the alarm on suicide prevention: A call to action. *Psychiatry*, 84(2), 192–195. <https://doi.org/10.1080/00332747.2021.1907871>
- Fortgang, R. G., Wang, S. B., Millner, A. J., Reid-Russell, A., Beukenhorst, A. L., Kleiman, E. M., Bentley, K. H., Zuromski, K. L., Al-Suwaidi, M., Bird, S. A., Buonopane, R., DeMarco, D., Haim, A., Joyce, V. W., Kastman, E. K., Kilbury, E., Lee, H.-I. S., Mair, P., Nash, C. C., ... Nock, M. K. (2021). Increase in suicidal thinking during COVID-19. *Clinical Psychological Science*, 9, 482–488. <https://doi.org/10.1177/2167702621993857>
- Fox, K. R., Huang, X., Guzmán, E. M., Funsch, K. M., Cha, C. B., Ribeiro, J. D., & Franklin, J. C. (2020). Interventions for suicide and self-injury: A meta-analysis of randomized controlled trials across nearly 50 years of research. *Psychological Bulletin*. <https://doi.org/10.1037/bul0000305>
- Franklin, J. C., Ribeiro, J. D., Fox, K. R., Bentley, K. H., Kleiman, E. M., Huang, X., Musacchio, K. M., Jaroszewski, A. C., Chang, B. P., & Nock, M. K. (2017). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin*, 143(2), 187–232. <https://doi.org/10.1037/bul0000084>
- García de la Garza, Á., Blanco, C., Olfson, M., & Wall, M. M. (2021). Identification of suicide attempt risk factors in a national US survey using machine learning. *JAMA Psychiatry*, 78(4), 398–406. <https://doi.org/10.1001/jamapsychiatry.2020.4165>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 103). Springer. <https://doi.org/10.1007/978-1-4614-7138-7>
- Kessler, R. C., Stein, M. B., Petukhova, M. V., Bliese, P., Bossarte, R. M., Bromet, E. J., Fullerton, C. S., Gilman, S. E., Ivany, C., Lewandowski-Romps, L., Bell, A. M., Naifeh, J. A., Nock, M. K., Reis, B. Y., Rosellini, A. J., Sampson, N. A., Zaslavsky, A. M., & Ursano, R. J. (2017). Predicting suicides after outpatient mental health visits in the army study to assess risk and resilience in servicemembers (Army STARRS). *Molecular Psychiatry*, 22(4), 544–551. <https://doi.org/10.1038/mp.2016.110>

- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer-Verlag. <https://doi.org/10.1007/978-1-4614-6849-3>
- Large, M. M., & Kapur, N. (2018). Psychiatric hospitalisation and the risk of suicide. *The British Journal of Psychiatry*, 212(5), 269–273. <https://doi.org/10.1192/bjp.2018.22>
- Linthicum, K. P., Schafer, K. M., & Ribeiro, J. D. (2019). Machine learning in suicide science: Applications and ethics. *Behavioral Sciences & the Law*, 37(3), 214–222. <https://doi.org/10.1002/bsl.2392>
- Ngiam, K. Y., & Khor, I. W. (2019). Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, 20(5), e262–e273. [https://doi.org/10.1016/S1470-2045\(19\)30149-4](https://doi.org/10.1016/S1470-2045(19)30149-4)
- Nock, M. K., Millner, A. J., Joiner, T. E., Gutierrez, P. M., Han, G., Hwang, I., King, A., Naifeh, J. A., Sampson, N. A., Zaslavsky, A. M., Stein, M. B., Ursano, R. J., & Kessler, R. C. (2018). Risk factors for the transition from suicide ideation to suicide attempt: Results from the army study to assess risk and resilience in servicemembers (Army STARRS). *Journal of Abnormal Psychology*, 127(2), 139–149. <https://doi.org/10.1037/abn0000317>
- Nock, M. K., Kleiman, E. M., Abraham, M., Bentley, K. H., Brent, D. A., Buonopane, R. J., Castro-Ramirez, F., Cha, C. B., Dempsey, W., Draper, J., Glenn, C. R., Harkavy-Friedman, J., Hollander, M. R., Huffman, J. C., Lee, H. I. S., Millner, A. J., Mou, D., Onnela, J.-P., Picard, R. W., ... Pearson, J. L. (2021). Consensus statement on ethical & safety practices for conducting digital monitoring studies with people at risk of suicide and related behaviors. *Psychiatric Research and Clinical Practice*, 3(2), 57–66. <https://doi.org/10.1176/appi.prcp.20200029>
- Piccininni, M., Konigorski, S., Rohmann, J. L., & Kurth, T. (2020). Directed acyclic graphs and causal thinking in clinical risk prediction modeling. *BMC Medical Research Methodology*, 20(1), 179. <https://doi.org/10.1186/s12874-020-01058-z>
- Pokorny, A. D. (1983). Prediction of suicide in psychiatric patients: Report of a prospective study. *Archives of General Psychiatry*, 40(3), 249. <https://doi.org/10.1001/archpsyc.1983.01790030019002>
- van Smeden, M., Groenwold, R. H. H., & Moons, K. G. M. (2020). A cautionary note on the use of the missing indicator method for handling missing data in prediction research. *Journal of Clinical Epidemiology*, 125, 188–190. <https://doi.org/10.1016/j.jclinepi.2020.06.007>
- Wang, S. B., Coppersmith, D. D. L., Kleiman, E. M., Bentley, K. H., Millner, A. J., Fortgang, R., Mair, P., Dempsey, W., Huffman, J. C., & Nock, M. K. (2021). A pilot study using frequent inpatient assessments of suicidal thinking to predict short-term postdischarge suicidal behavior. *JAMA Network Open*, 4(3), e210591. <https://doi.org/10.1001/jamanetworkopen.2021.0591>
- Ward-Ciesielski, E. F., & Rizvi, S. L. (2020). The potential iatrogenic effects of psychiatric hospitalization for suicidal behavior: A critical review and recommendations for research. *Clinical Psychology: Science and Practice*, e12332. <https://doi.org/10.1111/cpsp.12332>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, 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 license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license 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.

