BMJ Health & Care Informatics

AI enabled suicide prediction tools: a qualitative narrative review

Daniel D'Hotman,[1] Erwin Loh [2,3]

To cite: D'Hotman D, Loh E. AI enabled suicide prediction tools: a qualitative narrative review. *BMJ Health Care Inform* 2020;**27**:e100175. doi:10.1136/bmjhci-2020-100175

Received 14 May 2020
Revised 19 August 2020
Accepted 21 August 2020

[1]Oxford Uehiro Centre for Practical Ethics, University of Oxford, Oxford, United Kingdom
[2]Monash Centre for Health Research and Implementation, Monash University, Clayton, Victoria, Australia
[3]Group Chief Medical Officer, St Vincent's Health Australia Ltd, East Melbourne, Victoria, Australia

Correspondence to
Dr Daniel D'Hotman; daniel.dhotman@philosophy.ox.ac.uk

ABSTRACT

Background: Suicide poses a significant health burden worldwide. In many cases, people at risk of suicide do not engage with their doctor or community due to concerns about stigmatisation and forced medical treatment; worse still, people with mental illness (who form a majority of people who die from suicide) may have poor insight into their mental state, and not self-identify as being at risk. These issues are exacerbated by the fact that doctors have difficulty in identifying those at risk of suicide when they do present to medical services. Advances in artificial intelligence (AI) present opportunities for the development of novel tools for predicting suicide.

Method: We searched Google Scholar and PubMed for articles relating to suicide prediction using artificial intelligence from 2017 onwards.

Conclusions: This paper presents a qualitative narrative review of research focusing on two categories of suicide prediction tools: medical suicide prediction and social suicide prediction. Initial evidence is promising: AI-driven suicide prediction could improve our capacity to identify those at risk of suicide, and, potentially, save lives. Medical suicide prediction may be relatively uncontroversial when it pays respect to ethical and legal principles; however, further research is required to determine the validity of these tools in different contexts. Social suicide prediction offers an exciting opportunity to help identify suicide risk among those who do not engage with traditional health services. Yet, efforts by private companies such as Facebook to use online data for suicide prediction should be the subject of independent review and oversight to confirm safety, effectiveness and ethical permissibility.

INTRODUCTION

Suicide poses a significant health burden worldwide. The WHO estimates that the 2016 suicide rate was 10.6 suicides per 100 000 persons, with 80% of suicides occurring in low-income and middle-income countries.[1] In many cases, people at risk of suicide do not engage with their doctor or community due to concerns about stigmatisation and forced medical treatment; worse still, people with mental illness (who form a majority of people who die from suicide) may have poor insight into their mental state, and not self-identify as being at risk. These issues are exacerbated by the fact that doctors have difficulty in identifying those at risk of suicide when they do present to medical services.

In an attempt to reduce the impact of suicide, there is increased interest in using artificial intelligence (AI), data science and other analytical techniques to improve suicide prediction and risk identification. Broadly, these tools fall under two categories.

- *Medical suicide prediction tools:* researchers and doctors using AI techniques such as natural language processing and machine learning, among others, to determine patterns of information and behaviour that indicate suicide risk, using data from electronic medical records, hospital records and potentially other government data sources. Most typically, these tools would be used in a hospital setting or general practitioner surgery to provide 'decision support' for doctors when determining a patient's suicide risk.
- *Social suicide prediction tools:* AI and data tools that leverage information from social media and browsing habits to determine suicide risk—for example, Facebook, Google and Apple using data from platforms to determine which users are at risk of suicide, and deploying appropriate interventions, such as free information and counselling services.

METHODOLOGY

This paper discusses the reasoning behind efforts to use AI predict suicide, and examines emerging literature surrounding medical and social suicide prediction tools. The authors have specifically restricted this review to recent research in AI published in peer reviewed medical journals since 2017. The time period is chosen due to the significant growth and improvements in AI technology in recent years, which means that results in older studies may no longer be applicable. Where recent papers published after 2017 were not available, earlier papers have been included to demonstrate particular use cases.

A search was conducted using Google Scholar and PubMed, using selection criteria including keywords artificial intelligence, machine learning, deep learning, artificial

neural networks and algorithms relating to suicide prediction, suicide ideation and suicide risk factors. Non-academic articles relating to social suicide prediction efforts currently underway in the private sector by 'big tech' (Google and Facebook), as well as smaller organisations, were analysed through search engines.

This review is not intended to be a systematic review, but rather a qualitative narrative review. We have restricted the studies featured to those that represent promising opportunities for future research in this emerging and rapidly changing area of psychiatry. This judgement is based on DD's research of this topic area and EL's experience and expertise as a specialist medical administrator in both academia and practice.

The analysis aims to inform medical professionals of AI's potential future use in suicide prediction. We note that these tools have a number of ethical and policy implications; these issues will be discussed in separate papers.

Limitations and areas of uncertainty

Many of the studies included in this paper are not necessarily generalisable to other geographies or demographics—this would require additional exploration and research. In a similar vein, the relationships outlined by these studies are only relevant to the data sets applied within said research; that is, the specific data sets used by the study reported. As such, it would be imprudent to infer that the results of the studies detailed in this paper indicate clinical applicability on their own. Rather, they offer guidance for promising avenues of further research with larger and more diverse data sets in specific patient populations, and/or by modifying the algorithmic methodology outlined to further improve accuracy. Finally, many studies use different AI techniques to analyse data or statistical methods for reporting data, which poses some limits on comparison of results between studies.

A note on units of measurement and definitions

Studies that examine AI suicide prediction models use different units of measurement when reporting results. Given this is a nascent area of research, it is not always possible to find studies that share the same units of measurement for comparison. Definitions are included here to provide context to the reader, with explanations relevant to their use within this paper.

- AUC (area under the receiver operator characteristics curve): AUC assesses chance-level discriminative accuracy. AUC takes into account both true positives and true negatives. An AUC of 1.0 equates to a model with perfect discriminative accuracy, while an AUC of 0.5 means that the model produces results with the same accuracy as chance.
- Accuracy: measured by comparing the computed result (positive or negative) against its true value.
- Precision (also known as positive predictive value): precision reflects the proportion of positive results in a model that are true positives.
- Sensitivity (also known as recall): the proportion of total true positives that were registered as positive by the model.

THE USE OF AI IN SUICIDE PREDICTION

While it is impossible to completely eliminate suicide, it should be possible to improve prediction and prevention through better analytical tools. Yet, prediction of suicide risk continues to present a challenge for traditional epidemiological studies and doctors. This is due to the complex factors that underpin suicide and the difficulties around identification of a small number of individuals in a large group with similar risk factors. A landmark meta-analysis by Franklin *et al* spanning 365 studies over 50 years found that prediction of suicide was only slightly better than chance for all outcomes, and that this predictive ability has not improved across 50 years of research.[2] Prediction by doctors is made more difficult by the fact that many people who die from suicide never disclose suicidal thoughts to their doctor.[3 4] People with suicidal thoughts also feel afraid to discuss these thoughts with friends and family because of fear they might be judged, hospitalised or medicated.

Despite these difficulties, a recent longitudinal study found that 83% of people that die from suicide have contact with health services in the year prior to their death, and 45% have contact in the month prior.[5] This suggests a significant opportunity to use medical prediction tools to assist doctors in predicting suicide risk when these patients present. Franklin *et al* actually recommended that such prediction tools should shift away from a focus on risk factors, and instead leverage machine learning algorithms and data science to predict suicide risk using novel analytical techniques.[2]

There is an emerging body of evidence suggesting that AI and data science may be effective tools in predicting and preventing suicide. Two potential use cases have been suggested: medical suicide prediction and social suicide prediction. Medical suicide prediction involves AI being deployed as a real time decision support tool to assist clinicians in identifying patients at risk of suicide. Social suicide prevention involves analysis of behaviour on social media, smartphone applications and other online sources to determine those at risk of suicide. Each of these examples will be discussed in turn; they both present different opportunities and challenges. Additionally, current use cases are listed to demonstrate possible methods of implementation.

AI for medical suicide prediction

With the proliferation of electronic medical records (EMRs), there is now a wealth of health data available. When linked with other data sources, analysis of these complex sets of information (known colloquially as 'big data') can provide a snapshot of the biological, social and psychological state of a person at one time. Machines can learn to detect patterns, which are indecipherable using traditional forms of biostatistics, by processing big data

through layered mathematical models (AI algorithms). Algorithms can be designed to correct and learn from mistakes (training) to add the accuracy of an AI predictive model confidence—this is called machine learning.[6] As such, AI—and machine learning more specifically—is well positioned to address the challenge of navigating big data for suicide prediction.

Results across multiple studies indicate that AI consistently outperforms doctors at predicting suicide completion and suicide attempts, highlighting the promise of AI-based medical suicide prediction. Research suggests a promising clinical application for AI in identifying risk of suicide completion. Kessler *et al* used machine learning protocols (Naive Bayes, random forests, support vector regression) to predict suicide completion among military veterans within 26 weeks following an outpatient mental health visit. The study demonstrated an AUC of 0.72 for those with prior hospitalisation for psychiatric issues, 0.61 for those without hospitalisation and 0.66 when both samples were combined. Relevant characteristics of hospitalisation and previous outpatient visits included suicidality, depression, bipolar disorder and non-affective psychosis. Interestingly, AUC improved to 0.75 when predicting suicide death within 5 weeks of the outpatient visit.[7]

In 2018, a study by Del Pozo-Banos *et al* used artificial neural networks (a type of machine learning technique) to analyse routinely collected information in EMRs to assess suicide risk in patients attending health services for *any* reason.[8] Using only EMR and hospital data in the 5 years prior to a patient dying by suicide, the model accurately matched control patients and suicide cases (that is, whether patients committed suicide or not) with an accuracy over 73%. The authors noted that more complex models incorporating more data points would likely yield better results, and such a model will be built in the next stage of experimentation.

AI has achieved high accuracy when predicting suicide attempts, too. By applying machine learning to EHRs, Walsh *et al* created machine-learning algorithms (random forest and logistic regression) that achieved AUC values of 0.80-0.84when predicting whether a suicide attempt was likely to occur within the next 2 years and within the next week, respectively . Depression with psychosis, schizophrenia and prior suicide attempt were classified as important predictors in long and short term prediction.[9] Ryu *et al* used a machine learning technique (random forest) to predict suicide attempts among those with suicidal ideation. The prediction model achieved strong results, with an AUC of 0.947 and accuracy of 88.9%.[10] It is important to note that the clinical applicability of these tools in the real world remains unproven; however, initial results are extremely promising.[11]

Results across multiple studies indicate that AI consistently outperforms doctors at predicting suicide completion and suicide attempts, highlighting the promise of AI-based medical suicide prediction. One could imagine a future where initial screening tools such as those proposed by Del Pozo-Banos *et al* and Walsh *et al*/Ryu *et al* are combined to give an extremely accurate picture of an individual's suicide risk.[8 9] In turn, this could be used to inform treatment options for high risk patients.

The Department of Veteran's Affairs in the USA is putting medical suicide prediction into practice. Rates of suicide among US military veterans are 1.5 times greater than those who have not served, even when adjusting for age and gender. In an effort to close this gap, the Recovery Engagement and Coordination for Health—Veterans Enhanced Treatment (REACH VET) programme uses an AI to examine millions of records on medications, treatment, traumatic events, overall health and other information. It then identifies veterans most at risk of suicide. Initial results have been impressive: those classified by the algorithm in the top 0.1% of risk were 15 times more likely to complete suicide in the next year, and 81 times more likely to attempt suicide in the next year, than the average veteran. Following risk assessment, clinicians then establish contact with at-risk veterans to offer resources and support, as well as an optional psychological consult. In the first year since implementing the programme, there were 250 less suicides (a 4% reduction) than what would have been expected from previous rates. While it is difficult to tell whether the REACH VET programme specifically contributed to this reduction, the Department has commissioned an independent evaluation of the programme's effectiveness, and will look to expand the use of predictive analytics and share risk data to improve the AI's modelling in coming years.[12 13]

An important question is what should be done when individuals are identified as being at risk of suicide. For example, hospitalisation may be the right step for some, but could cause more harm than good in other patients. Furthermore, forcibly detaining patients in a hospital or other medical setting could cause significant psychological stress and potentially hasten future suicide attempts. Identifying which types of treatment should be used for which patients is a valuable area of future research.

AI for social suicide prediction

A growing number of researchers and technology companies are using AI to monitor suicide risk through online activity. This builds on emerging evidence that language patterns on social media and methods of smart phone use can indicate psychiatric issues.[14]

A large number of studies have demonstrated the potential efficacy of applying social media to predicting suicide risk.[15–22] In most cases, natural language processing is used to analyse the online activity of users on social media platforms for suicidal behaviours (such as a mention of suicide attempt, suicidal ideation, or discussion of suicidal themes). This may be combined with machine learning techniques to compare and contrast findings across and within platforms—for example, to determine patterns of behaviour and how this may relate to risk severity.

In the vast majority of studies examining the use of AI to predict suicidal behaviours on social media, it is not

possible to verify against 'ground truth'. That is, it is not possible to use medical records to determine whether an individual posting on social media has actually experienced what they are describing on the platform. Where verification against medical records is not possible, medical professionals with expertise in suicide can verify the likely veracity of user claims. This is, of course, based on their subjective, professional judgement. Some higher quality studies only include cases where there is unanimous agreement by medical professionals that the individual in question is legitimately at risk of suicidal behaviour; these cases are then included in a 'gold standard' sample to assess an AI model's predictive power. One example is Gaur *et al*, where Reddit posts were examined for uses of suicidal language to determine suicide risk. Different clinical classification schemes were compared against machine learning techniques, including random forest and convolutional neural networks. Convolutional neural networks was the strongest performer, achieving an overall precision of 70%—40% better than baseline approaches that only applied medical classification systems.[23]

A landmark study published in Biomedical Informatics Insights by Coppersmith *et al* combined many of the insights of previous studies in this area. Coppersmith *et al* applied natural language processing and supervised and unsupervised machine learning methods to social media data from a variety of sources (eg, Facebook, Twitter, Instagram, Reddit, Tumblr, Strava and Fitbit, among others)—for which they were granted permission by test subjects—in order to determine the risk of attempted suicide. AUC was 0.89–0.93 for time periods ranging from 1 month to 6 months in length.[24] As outlined by Coppersmith *et al,* i if a false alarm rate of 1%-2% is assumed, this model may be up to 10 times more accurate at *correctly predicting suicide attempts* when compared with clinician averages (40%–60% vs 4%–6%).[25 24 26]

Coppersmith *et al* cautioned that these results focused on an 18–24 age group of mostly American women, so may not be generalised to other demographics, cultures or norms. For example, stigma in different communities may influence whether people post about suicide on social media. Nonetheless, initial evidence suggests comparable results for men and lesbian, gay, bisexual, transgender, intersex, and questioning (LBGTIQ) people—albeit in a small sample size.More broadly, the model's high accuracy in determining suicide risk, with access *only* to social media data, suggests a promising avenue for further research.

It is worth noting that the analytical tools deployed by Coppersmith *et al* are likely to be far less advanced and granular than those being undertaken by Facebook, Google, Twitter and other technology companies (which will be outlined later in this paper). This is on account of these firms having access to rich troves of online user data and cutting-edge analytical techniques. Given that these companies have not provided their results or techniques for independent evaluation (as will be discussed), it is not possible to draw further inferences. Yet, as more data becomes available through public forums, and algorithms for analysing this data advance, social suicide prediction is likely to yield significantly more accurate and clinically useful results than those described by Coppersmith *et al.*

Further, and finally, it is worth noting that the analytical power of such tools could be leveraged to enhance medical suicide prediction efforts—all that is required is that at-risk patients provide consent to access social media data. Padrez *et al* demonstrated the feasibility of such an approach; when asked in a hospital emergency setting, 37% of 2717 Facebook and/or Twitter users consented to share both their health record and social media data for the purpose of data linkage.[27] Patient sensitivity around suicide and mental illness information may mean lower rates of consent in this cohort. However, the potential clinical usefulness of combining medical and social suicide prediction tools means that this topic deserves future research and consideration.

AI-DRIVEN PREDICTION RELATING TO SUICIDE RISK FACTORS

Suicidal ideation

A study by Lin *et al* examined the effectivities of machine learning techniques in detecting suicidal ideation based on six psychological stressors in EMRs. This study of Taiwanese military men and women used machine learning techniques, including logistic regression, decision trees, random forest, gradient boosting regression tree, support vector machine and multilayer perceptron; all machine learning methods achieved accuracies over 98% in predicting suicidal ideation. When compared with conventional clinical criterion for assessing the presence of suicidal ideation, the algorithms improved sensitivity by more than 35% and precision by 65%.[28] In another study, researchers used a machine-learning algorithm (Naive Bayes) to identify those at risk of suicide ideation with 91% accuracy, based on their altered functional MRI neural signatures of death-related and life-related concepts.[29]

Turning to social media, Tadesse *et al* outlined a number of machine learning approaches for identifying suicidal ideation on Reddit using convolutional linguistics. One model, using a combination of long short-term memory and convolutional neural networks, achieved an accuracy and precision of 93% in identifying users with suicidal ideation.[30] Ji *et al* found comparable results, demonstrating that machine learning techniques could leverage statistical, linguistic, word embedding and topic features to achieve 90% accuracy in identifying suicide ideation on Reddit and Twitter.[31]

Despite these promising results, the utility of identifying suicidal ideation may be limited due to low positive predictive value and modest sensitivity for suicide attempts. This is due to the low incidence of suicide attempts when compared with incidence of suicidal ideation.[26 32] Saying this, such tools may still be useful. Many patients with suicidal ideation may not be willing

to disclose this fact to their doctors, meaning that tools which can predict suicidal ideation based on psychological stressors could be valuable to medical practitioners, particularly when dealing with high risk populations such as military personnel—in turn, these tools could be combined with those which predict suicide attempts and completed suicide, as outlined above, to increase their clinical applicability. In this fashion, algorithms that deliver advantages in precisely identifying those who may be at risk of suicidal ideation could help to provide targeted care to more patients who are in need, and with subsequent benefits for the efficient allocation of scarce medical resources.

Mental illness + AI: prediction and diagnosis

While only a small fraction of those with mental illness die from suicide, more than 80% of people who die from suicide are thought to have mental illness. Risk increases for patients with multiple comorbid mental illnesses.[26] As a result, there is clinical interest in better understanding the risk of mental illnesses in patients who may also be at risk of suicide (prediction) and correctly identifying mental illness when it is present (diagnosis).

One of the limitations of current psychiatric diagnosis of mental illness is that many conditions overlap with each other—at least 50% of patients receive more than one psychiatric diagnosis.[33] AI prediction tools in medical settings could provide better diagnostic clarity, thus improving treatment efficacy in patients and reducing the impact of unnecessary side effects. As such, many researchers are excited by the potential of AI to improve access to mental health services and drive down the cost of diagnosis—particularly in rural/remote and low-income settings.[34 35]

A full analysis of the opportunities for using AI to predict mental illness is beyond the scope of this paper. However, examples of AI's potential to predict and diagnose mental illness include the following.

Depression, anxiety and mood disturbances

- MIT researchers built an AI model able to identify a depressed individual based on speaking patterns—depressed people tend to have a lower range and pitch of their voice, with more pauses, starts and stops between their words.[36]
- A study by Zhao *et al* demonstrated that a trained AI (using linear regression, epsilon support vector regression and gaussian processes) could identify patients with anxiety and depression in real time based on their walking style. Remarkably, the algorithm was also able to determine the severity of their illness.[37]
- Harvard researchers Andrew Reece and Christopher Danforth applied a machine learning tool (logistic regression) to nearly 44 000 Instagram photos from 166 individuals to successfully identify markers of depression with 70% accuracy, which is markedly superior to success rates by unassisted GPs (just over 50%).[38]
- Xu *et al* constructed a multitask deep learning model that accurately predicted the onset of depressive disorder for elderly individuals by capturing 22 years of longitudinal household survey data on depressive risk factors; this model outperformed existing regression models for predicting depression.[39]

Schizophrenia

- A study by Kalmady *et al*, published in Nature, demonstrated that a machine learning model could correctly diagnose schizophrenia with 87% accuracy (chance accuracy of 53%), based on alterations in brain activity on functional MRI imaging.[40]

Post-traumatic stress disorder (PTSD)

- A Danish prospective study used machine learning to analyse risk indicators and forecast long term post-traumatic stress responses among a cohort of Danish soldiers; after following the soldiers for 6 years, the algorithm had demonstrated an AUC of 0.84 in pre-deployment screening and 0.88 in post-deployment screening. The authors noted the potentially significant benefits of such technology in identifying high-risk soldiers early to improve treatment and reduce long public health costs.[41]

A Danish prospective study used machine learning to analyse risk indicators and forecast long term post-traumatic stress responses among a cohort of Danish soldiers; after following the soldiers for 6 years, the algorithm had demonstrated an AUC of 0.84 in pre-deployment screening and 0.88 in post-deployment screening. The authors noted the potentially significant benefits of such technology in identifying high-risk soldiers early to improve treatment and reduce long public health costs.[41]

Predicting mental illness from social media data

Research into the use of social media data to aid diagnosis of mental illnesses has also been promising. Social media data has been found to contain predictive signals for a variety of conditions, including: major depressive disorder,[42 43] PTSD,[44–47] schizophrenia,[48] eating disorders,[49 50] bipolar affective disorder,[51] borderline personality disorder[52] and others.[53] Further research is required to demonstrate the effectiveness of these tools in different contexts, cultures and settings. However, it is clear that these tools have the potential to act as a useful adjunct to prediction and diagnosis of mental illness in medical settings—particularly in relation to determining suicide risk—as well as creating a rich and powerful data set to inform mental health resourcing by policy makers.

Combining analytical insights from different mental illnesses

Models that combine information on different mental illnesses could generate more accurate results than those focussed on one type of mental illness—this is called multitask learning (MTL). MTL involves applying the learnings from different but related tasks (in this case, predicting different mental illnesses using AI) to improve the accuracy of each individual prediction. This

is hypothesised to be effective because of the close relationship and overlap between risk factors/demographic factors for these mental illnesses, as well as the likelihood of comorbidity.[54]

Benton *et al* examined the potential of MTL algorithms to predict the risk of various mental illnesses. When compared against self-stated presence of illness (as determined by a human annotator on Twitter), the MTL model achieved AUC of 0.70 for all mental illnesses analysed—anxiety, depression, eating disorder, panic attacks, schizophrenia, bipolar disorder and PTSD. Predictions for less common conditions (eg, PTSD and bipolar) became more accurate when models were forced to also predict comorbid conditions for which there was more data (such as depression and anxiety).[54] This demonstrates the potential for using MTL models to predict less common mental illnesses, many of which are also direct risk factors for suicidality. As models are able to accommodate greater amounts of related mental health information, they are likely to see significant gains in their predictive power.

Suicide amongst adolescents

A number of promising studies have analysed the potential for AI to predict suicide attempts among adolescents. Jung *et al* applied machine learning algorithms to a nationally representative sample of nearly 60 000 Korean adolescents to determine risk of suicide via history of suicide attempts/ideation. Taking into account 26 predictors of suicide risk, five different models (logistic regression, random forest, support vector machine, artificial neural network and extreme gradient boosting) achieved an accuracy between 77.5% and 79%.[55] Walsh *et al* conducted a retrospective cohort study of 33 000 adolescents to predict suicide attempts; random forests achieved AUC values >0.80 across time frames that ranged from prediction windows of 7 days to 2 years.[56] Finally, Bhat and Goldman-Mellor used deep neural networks to predict suicide attempts among Californian adolescents using a sample of over 500 000 medical records. The strongest performing model of the experiment achieved a sensitivity of 70%, specificity of 98% and AUC of 0.958.[57]

Non-suicidal self-injury and self-harm

Non-suicidal self-injury (NSSI) is defined as deliberate direct destruction or alteration of body tissue without conscious suicidal intent.[58] Deliberate self-harm is an encompassing term for self-injurious behaviour, both with and without suicidal intent, that has a non-fatal outcome.[59] NSSI has been linked to increased risk of severe self-harm and suicide attempts.[60]Ammerman *et al* used lasso regression (a type of regression with regularisation) and random forests to analyse NSSI patterns among 712 undergraduate students. Findings demonstrated that suicide plans and depression, both risk factors for suicide, were significant predictors of lifetime NSSI risk.[61] Using a sample of 359 undergraduate students with a history of NSSI, Burke *et al* attempted to determine which NSSI factors were most salient to suicide risk.[62] Three machine learning techniques (elastic net regression, decision trees, random forests) were used to determine that motivations, method lethality and scarring are likely the most important factors in ascertaining suicide risk. Further research is required to analyse the replicability of these results with larger sample sizes and across different geographies and age groups.

Physical illness

The presence of physical illness has been found to contribute to suicide risk.[63] A study by Karmakar *et al* aimed to quantify the impact of a history of physical illness on suicide risk by using machine learning techniques to analyse EMR data of 7399 mental health patients with a history of physical illness. The best performing machine learning model combined data across all time periods to significantly outperform clinical baseline risk assessment in predicting suicide risk (AUC of 0.71 vs AUC of 0.56).[64] This infers that AI suicide prediction tools are likely to be more effective when history of physical illness is taken into account.

SUICIDE PREDICTION USING WEARABLES DATA

A type of suicide prediction yet to be discussed is the potential to combine wearables data with social media data to determine suicide risk in real time. This may include combining health data on sleep, nutrition, stress, heart rate and other biomedical indicators from personal health apps and social media. Personal health apps compile information from wearables such as the Apple Watch and Fitbit, among others.

In a first-generation British study, Haines-Delmont *et al* created a smartphone app that linked Fitbit, Apple Health kit and Facebook to collect information on sleep behaviour, mood, step frequency and count, and technology engagement. Despite a small sample size (66 patients from acute mental health inpatient wards), this study demonstrated a technically feasible pathway to use machine learning models to assess suicide risk among inpatients by leveraging information from mobile devices.[65] Such tools could support clinical judgement making in inpatient settings; however, further research using larger data sets is required to determine veracity.

SOCIAL SUICIDE PREDICTION IN THE PRIVATE SECTOR

The prevalence of suicide in conjunction with the difficulty in identifying those in need of support has led to development of social suicide prediction effort by companies that accumulate user data.

Facebook has one of the most public social suicide prediction programmes. Various types of prevention tools have been available on the platform for more than 10 years. In November 2017, in response to users' live streaming suicide attempts, Facebook stepped up its efforts, rolling out a detailed prediction and prevention

programme. One arm of the programme involves users reporting posts of concern, which are then reviewed by a human member of Facebook's community support team. In efforts to improve the accuracy and efficiency of the project, Facebook later developed a machine learning tool that uses machine learning (random forest technique) to determine the risk profile of users by scanning posts and live videos for threats of suicide and self-harm, alerting the team of human reviewers to suspect posts.[66] Facebook claims that this AI supported prediction tool is more accurate than human reports. However, as of February 2020, no data has been provided to authenticate this claim.

If reviewers are concerned about suicidal intent, the user in question is provided with free information about support services the next time they log on to Facebook, including country-based support hotlines, online chat resources, and tips and suggestions. Facebook points out that use of these services is completely optional. In rare instances, reviewers contact emergency services who respond using geolocation data from Facebook to assist users who may be at immediate risk to themselves.[67] In its first month of operation, Facebook claimed that its AI helped connect first responders with 100 people at immediate risk; in late 2018, this number exceeded 3500.[68 69] However, no further data were published on the outcomes of these cases, or the programme's effectiveness more broadly.[3] Facebook has also developed a photo identification AI tool for Instagram to assist these efforts; yet little further information has been published on this tool.[68]

Facebook explains that it developed these tools in collaboration with mental health organisations such as Forefront Suicide Prevention and National Suicide Prevention Lifeline, as well as receiving contributions from members of the public with experience of suicide.[68] Notably absent from this list is an independent review of results and/or methodology by academics, a human research ethics approval process, or input from expert medical organisations, such as the American Psychiatric Association, and government regulators. Facebook claims that it has considered user privacy in the creation of its suicide prediction tool by not allowing the AI to train on information that is published under 'only me' posts, not taking into account demographic data about an individual, and not alerting friends or networks to an individual's suicidal intent. It is worth noting once again that we have to take Facebook's claims on their word, given a lack of access to data by outside researchers. Facebook also points out that it has engaged members of the public about the technical details and deployment of its suicide prediction tool, including in a scientific publication by its Global Head of Privacy and Public Policy,[68] and a variety of articles on its platform.[70–72]

In the USA, any Google search for clinical depression symptoms launches a knowledge panel and private screening test for depression, along with educational and referral mechanisms. Google states that this data is de-identified and may be used to generate a digital fingerprint of depression that could aid further research; however, it has refused to release any details of its algorithms.[73] Google probably already uses AI to monitor videos posted by users on its video-sharing platform YouTube.[74] Siri (Apple), Google Assistant, Alexa (Amazon) and Cortana (Microsoft) all have features which direct people to suicide prevention resources based on trigger words and phrases.[68] Other companies and services active in suicide prediction and prevention include the following.

- Radar—an app developed by the UK non-profit Samaritans which alerted users to when a friend or contact exhibited signs of suicide risk using an AI algorithm on Twitter. The Radar app created significant controversy due to community concerns that a nefarious actor could use it to profile suicidal risk of subjects, regardless of their prior relationship.[68]
- Crisis Text Line—a non-profit providing text message crisis support across the USA, Canada, South Africa and Ireland. Crisis Text Line uses machine learning algorithms to help researchers and counsellors determine when a social media post is indicative of a real suicidal threat, rather than just a joke or expression of emotion. With AI analysing more than 54 million messages, counsellors can usually determine within three messages whether they should alert emergency services based on key words and phrases. For example, those individuals who use words such as 'ibuprofen' or 'Advil' are 14 times more likely to need emergency services than a person using the word 'suicide'. Similarly, a person using a crying face emoticon is 11 times more likely to need emergency services than a person using the word 'suicide.' Crisis Text Line has partnered with Facebook, YouTube, Kik and a number of universities to provide crisis counselling to people in need.[24 75]
- Trevor Project—a similar organisation to Crisis Text Line, Trevor Project works with Google to incorporate machine learning into its text-based counselling service for LBGTIQ young people, so that counsellors can more quickly determine the risk profile of those contacting the service.[76]

Some mental health professionals have encouraged the development and use of these tools as a means to reduce the number of people who attempt suicide. For example, Facebook originally began building its suicide prediction and prevention tools after being approached from suicide prevention experts and non-profits that are active in the space.[68] These tools may be a particularly promising mechanism of engaging young people, a vulnerable group who are more likely to reach out for help through social media than to see a therapist or call a crises hot line.[14] That is, the data available to these tech giants and the ubiquitous nature of their platforms, particularly among young people, offers an invaluable opportunity to identify at-risk individuals who may not otherwise engage with health services.

However, in contrast to the peer reviewed papers described earlier in this paper, a number of concerns

have been raised about these tools, including: a lack of independent review to assess efficacy, poor transparency about methodology, storage of sensitive medical data and a lack of ethical oversight.[3 69]

EXAMPLES OF POPULATION-WIDE INITIATIVES

Other initiatives are being developed to inform suicide prevention efforts at a population level. The benefit of these initiatives is that they do not require the identification of individuals; rather, they rely on insights of population data to inform provision of health resources for suicide intervention. Two examples of such initiatives are currently underway.

1. The Canadian Government, through Public Health Canada, has signed a contract with Ottawa-based AI company Advanced Symbolics to identify suicide-related behaviour and monitor discussions about suicide. The aim of the project is to determine suicide-hotspots and inform government allocation of resources to high-risk areas. Data will be de-identified. Interestingly, Advanced Symbolics' technology is best known for correctly predicting the result of the 2016 US election and Brexit referendum.[77]
2. In Australia, in May 2019, Melbourne-based research centre Turning Point was awarded a $A1.21 million grant from Google's non-profit arm to establish a world-first suicide surveillance system, along with Monash University and Eastern Health. The system will use AI techniques to code suicide-related ambulance data, and in doing so, identify geographic trends and hotspots to help inform public health policy and intervention. Successful applicants to Google's programme, such as Turning Point, also receive coaching and consulting services from Google's AI experts.[78]

These case studies present interesting examples of how countries could leverage capabilities within the private sector and non-profit organisations to develop analytical tools that inform broader suicide prevention efforts. Similar projects could be funded in areas of strategic importance (such as Indigenous, rural and LBGTIQ mental health). Governments could also consider providing access to de-identified health data to assist organisations and academics to increase the analytical power of similar research efforts. These examples seem *prima facie* to be ethically permissible, given that the data is de-identified and the results of the research could result in clear benefits in terms of suicide prediction and prevention.

CONCLUSION

Advances in AI present opportunities for the development of novel tools for predicting suicide. This paper has provided an overview of research focusing on two broad categories: medical suicide prediction tools and social suicide prediction tools. Furthermore, this paper analysed AI's potential to predict suicidal ideation and mental illness, as well as the implications of physical illness, age (adolescents) and selfharm in AI driven suicide prediction.

Evidence suggests that medical and social suicide prediction tools could improve our capacity to identify those at risk of suicide, and, potentially, save lives. However, further research is required to determine the validity and ethics of using these tools in different contexts. Population-wide suicide prediction is likely to offer an ethical and useful application of AI, aiding policy makers and medical professionals in better allocating healthcare resources. Efforts by private companies to use online data for suicide prediction must be closely monitored by the scientific community; this paper suggests that these efforts should be subject to independent review and ethical oversight to confirm safety, effectiveness and permissibility.

Twitter Erwin Loh @erwinloh

Contributors DD planned and wrote the original draft of this paper. EL provided feedback on this draft and contributed to the final version of the paper.

Funding DD's DPhil is funded on a Rhodes Scholarship.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available.

ORCID iD

Erwin Loh http://orcid.org/0000-0001-7157-0826

REFERENCES

1 Fazel S, Runeson B, Ropper AH. Suicide. *N Engl J Med* 2020;382:266–74.
2 Franklin JC, Ribeiro JD, Fox KR, *et al*. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol Bull* 2017;143:187–232.
3 Marks M. Artificial intelligence based suicide prediction. *Yale J Health Policy Law Ethics* 2019.
4 Sheehan L, Dubke R, Corrigan PW. The specificity of public stigma: a comparison of suicide and depression-related stigma. *Psychiatry Res* 2017;256:40–5.
5 Ahmedani BK, Simon GE, Stewart C, *et al*. Health care contacts in the year before suicide death. *J Gen Intern Med* 2014;29:870–7.
6 Miller DD, Brown EW. Artificial intelligence in medical practice: the question to the answer? *Am J Med* 2018;131:129–33.
7 Kessler RC, Stein MB, Petukhova MV, *et al*. Predicting suicides after outpatient mental health visits in the army study to assess risk and resilience in Servicemembers (army STARRS). *Mol Psychiatry* 2017;22:544–51.
8 DelPozo-Banos M, John A, Petkov N, *et al*. Using neural networks with routine health records to identify suicide risk: feasibility study. *JMIR Ment Health* 2018;5:e10144.
9 Walsh CG, Ribeiro JD, Franklin JC. Predicting risk of suicide attempts over time through machine learning. *Clin Psychol Sci* 2017;5:457–69.
10 Ryu S, Lee H, Lee D-K, *et al*. Detection of suicide attempters among suicide ideators using machine learning. *Psychiatry Investig* 2019;16:588–93.
11 Loh E. Medicine and the rise of the robots: a qualitative review of recent advances of artificial intelligence in health. *Leader* 2018;2:59–63.

12 VA Releases National Suicide Data Report for 2005-2016Us department of Veterens Affairs, 2018. https://www.va.gov/opa/pressrel/includes/viewPDF.cfm?id=5114 [Accessed 18 September 2020].
13 Department of Veterans Affairs. REACH VET: recovery engagement and coordination for health - veterans enhanced treatment, predictive analytics for suicide prevention, 2017. Available: https://www.dspo.mil/Portals/113/Documents/2017%20Conference/Presentations/REACH%20VET%20August%20for%20DoD.VA%20Conf.pptx?ver=2017-08-10-132612-030
14 Reardon S. AI algorithms to prevent suicide gain traction. *Nature* 2017;64.
15 Coppersmith G, Leary R, Whyne E, *et al*. *Quantifying suicidal ideation via language usage on social media*. Joint Statistics Meetings Proceedings, Statistical Computing Section, JSM, 2015.
16 Kumar M, Dredze M, Coppersmith G. Detecting changes in suicide content manifested in social media following celebrity suicides. *Proceedings of the 26th ACM conference on Hypertext & Social Media*, 2015:85–94.
17 O'Dea B, Wan S, Batterham PJ, *et al*. Detecting suicidality on Twitter. *Internet Interv* 2015;2:183–8.
18 De Choudhury M, Kiciman E, Dredze M. Discovering shifts to suicidal ideation from mental health content in social media. *Proceedings of the 2016 chi conference on human factors in computing systems*, 2016:2098–110.
19 Bryan CJ, Butner JE, Sinclair S, *et al*. Predictors of emerging suicide death among military personnel on social media networks. *Suicide Life Threat Behav* 2018;48:413–30.
20 Coppersmith G, Ngo K, Leary R, *et al*. Exploratory analysis of social media prior to a suicide attempt. In: *Proceedings of the third workshop on computational linguistics and clinical psychology*, 2016: 106–17.
21 Dinakar K, Chen J, Lieberman H. Mixed-initiative real-time topic modeling & visualization for crisis counseling. *Proceedings of the 20th International Conference on intelligent user interfaces*, 2015:417–26.
22 Pestian JP, Sorter M, Connolly B, *et al*. A machine learning approach to identifying the thought markers of suicidal subjects: a prospective multicenter trial. *Suicide Life Threat Behav* 2017;47:112–21.
23 Gaur M, Alambo A, Sain JP, *et al*. *Knowledge-aware assessment of severity of suicide risk for early intervention*, 2019: 514–25.
24 Coppersmith G, Leary R, Crutchley P, *et al*. Natural language processing of social media as screening for suicide risk. *Biomed Inform Insights* 2018;10:1178222618792860.
25 Franklin JC, Ribeiro JD, Fox KR, *et al*. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol Bull* 2017;143:187–232.
26 Nock MK, Borges G, Bromet EJ, *et al*. Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. *Br J Psychiatry* 2008;192:98–105.
27 Padrez KA, Ungar L, Schwartz HA, *et al*. Linking social media and medical record data: a study of adults presenting to an academic, urban emergency department. *BMJ Qual Saf* 2016;25:414–23.
28 Lin G-M, Nagamine M, Yang S-N, *et al*. Machine learning based suicide ideation prediction for military personnel. *IEEE J Biomed Health Inform* 2020;24:1907–16.
29 Just MA, Pan L, Cherkassky VL, *et al*. Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth. *Nat Hum Behav* 2017;1:911–9.
30 Tadesse MM, Lin H, Xu B, *et al*. Detection of suicide ideation in social media forums using deep learning. *Algorithms* 2020;13:7.
31 Ji S, Yu CP, Fung S-fu, *et al*. Supervised learning for suicidal ideation detection in online user content. *Complexity* 2018;2018:1–10.
32 McHugh CM, Corderoy A, Ryan CJ, *et al*. Association between suicidal ideation and suicide: meta-analyses of odds ratios, sensitivity, specificity and positive predictive value. *BJPsych Open* 2019;5:e18.
33 Psychology Today. Can artificial intelligence improve psychiatric diagnosis? 2018. Available: https://www.psychologytoday.com/blog/psychiatry-the-people/201802/can-artificial-intelligence-improve-psychiatric-diagnosis [Accessed 22 Aug 2019].
34 Tiffin PA, Paton LW. Rise of the machines? machine learning approaches and mental health: opportunities and challenges. *Br J Psychiatry* 2018;213:509–10.
35 Lovejoy CA, Buch V, Maruthappu M. Technology and mental health: the role of artificial intelligence. *Eur Psychiatry* 2019;55:1–3.
36 News MIT. Model can more naturally detect depression in conversations, 2018. Available: http://news.mit.edu/2018/neural-network-model-detect-depression-conversations-0830 [Accessed 22 Aug 2019].
37 Zhao N, Zhang Z, Wang Y, *et al*. See your mental state from your walk: recognizing anxiety and depression through Kinect-recorded gait data. *PLoS One* 2019;14:e0216591.
38 Reece AG, Danforth CM. Instagram photos reveal predictive markers of depression. *EPJ Data Sci* 2017;6:15.
39 Xu Z, Zhang Q, Li W, *et al*. Individualized prediction of depressive disorder in the elderly: a multitask deep learning approach. *Int J Med Inform* 2019;132:103973.
40 Kalmady SV, Greiner R, Agrawal R, *et al*. Towards artificial intelligence in mental health by improving schizophrenia prediction with multiple brain parcellation ensemble-learning. *NPJ Schizophr* 2019;5:2.
41 Karstoft K-I, Statnikov A, Andersen SB, *et al*. Early identification of posttraumatic stress following military deployment: application of machine learning methods to a prospective study of Danish soldiers. *J Affect Disord* 2015;184:170–5.
42 Chung C, Pennebaker JW. The psychological functions of function words. *Soc Commun* 2007;1:343–59.
43 De Choudhury M, Gamon M, Counts S. Predicting depression via social media. *Seventh International AAAI conference on weblogs and social media*, 2013.
44 Coppersmith G, Harman C, Dredze M. Measuring post traumatic stress disorder in Twitter. *Eighth international AAAI conference on weblogs and social media*, 2014.
45 Yates A, Cohan A, Goharian N. Depression and self-harm risk assessment in online forums. *ArXiv Prepr* 2017.
46 Resnik P, Armstrong W, Claudino L, *et al*. The University of Maryland CLPsych 2015 shared task system. In: *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, 2015: 54–60.
47 Pedersen T. Screening Twitter users for depression and PTSD with lexical decision lists. In: *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, 2015: 46–53.
48 Mitchell M, Hollingshead K, Coppersmith G. Quantifying the language of schizophrenia in social media. In: *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, 2015: 11–20.
49 Walker M, Thornton L, De Choudhury M, *et al*. Facebook use and disordered eating in college-aged women. *J Adolesc Health* 2015;57:157–63.
50 Chancellor S, Mitra T, De Choudhury M. Recovery amid pro-anorexia: Analysis of recovery in social media. In: *Proceedings of the 2016 chi conference on human factors in computing systems*. ACM, 2016: 2111–23.
51 Coppersmith G, Dredze M, Harman C. Quantifying mental health signals in Twitter. In: *Proceedings of the workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, 2014: 51–60.
52 Loveys K, Crutchley P, Wyatt E, *et al*. Small but mighty: affective micropatterns for quantifying mental health from social media language. In: *Proceedings of the fourth workshop on computational linguistics and clinical Psychology—From linguistic signal to clinical reality*, 2017: 85–95.
53 Coppersmith G, Dredze M, Harman C, *et al*. From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. In: *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, 2015: 1–10.
54 Benton A, Mitchell M, Hovy D. *Multitask learning for mental health conditions with limited social media data*, 2017: 152–62.
55 Jung JS, Park SJ, Kim EY, *et al*. Prediction models for high risk of suicide in Korean adolescents using machine learning techniques. *PLoS One* 2019;14:e0217639.
56 Walsh CG, Ribeiro JD, Franklin JC. Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning. *J Child Psychol Psychiatry* 2018;59:1261–70.
57 Bhat HS, Goldman-Mellor SJ. Predicting adolescent suicide attempts with neural networks. *ArXiv Prepr* 2017.
58 Nock MK. Why do people hurt themselves? new insights into the nature and functions of self-injury. *Curr Dir Psychol Sci* 2009;18:78–83.
59 Lim K-S, Wong CH, McIntyre RS, *et al*. Global lifetime and 12-month prevalence of suicidal behavior, deliberate self-harm and non-suicidal self-injury in children and adolescents between 1989 and 2018: a meta-analysis. *Int J Environ Res Public Health* 2019;16:4581.
60 Ammerman BA, Jacobucci R, Kleiman EM, *et al*. The relationship between nonsuicidal self-injury age of onset and severity of self-harm. *Suicide Life Threat Behav* 2018;48:31–7.

61 Ammerman BA, Jacobucci R, McCloskey MS. Using exploratory data mining to identify important correlates of nonsuicidal self-injury frequency. *Psychol Violence* 2018;8:515–25.
62 Burke TA, Jacobucci R, Ammerman BA, *et al*. Identifying the relative importance of non-suicidal self-injury features in classifying suicidal ideation, plans, and behavior using exploratory data mining. *Psychiatry Res* 2018;262:175–83.
63 Abraham N, Buvanaswari P, Rathakrishnan R, *et al*. A meta-analysis of the rates of suicide ideation, attempts and deaths in people with epilepsy. *Int J Environ Res Public Health* 2019;16:1451.
64 Karmakar C, Luo W, Tran T, *et al*. Predicting risk of suicide attempt using history of physical illnesses from electronic medical records. *JMIR Ment Health* 2016;3:e19.
65 Haines-Delmont A, Chahal G, Bruen AJ, *et al*. Testing suicide risk prediction algorithms using phone measurements with patients in acute mental health settings: feasibility study. *JMIR Mhealth Uhealth* 2020;8:e15901.
66 Facebook Engineering. Under the hood: suicide prevention tools powered by AI, 2018. Available: https://engineering.fb.com/ml-applications/under-the-hood-suicide-prevention-tools-powered-by-ai/ [Accessed 20 Aug 2019].
67 Singer N. *Screening for suicide risk, facebook takes on tricky public health role*. N Y Times, 2018.
68 Gomes de Andrade NN, Pawson D, Muriello D, *et al*. Ethics and artificial intelligence: suicide prevention on facebook. *Philos Technol* 2018;31:669–84.
69 Barnett I, Torous J, Ethics TJ. Ethics, transparency, and public health at the intersection of innovation and Facebook's suicide prevention efforts. *Ann Intern Med* 2019;170:565–6.
70 Facebook Newsroom. Building a safer community with new suicide prevention tools, 2017. Available: https://newsroom.fb.com/news/2017/03/building-a-safer-community-with-new-suicide-prevention-tools/ [Accessed 20 Aug 2019].
71 Facebook Newsroom. Getting our community help in real time, 2017. Available: https://newsroom.fb.com/news/2017/11/getting-our-community-help-in-real-time/ [Accessed 20 Aug 2019].
72 Facebook Newsroom. How facebook AI helps suicide prevention, 2018. Available: https://newsroom.fb.com/news/2018/09/inside-feed-suicide-prevention-and-ai/ [Accessed 20 Aug 2019].
73 Chae L. How search engines are failing suicidal users, 2018. Available: https://www.fastcompany.com/90230313/how-search-engines-are-failing-suicidal-users [Accessed 20 Aug 2019].
74 Matsakis L. A window into how youtube trains AI to moderate videos, 2018. Available: https://www.wired.com/story/youtube-mechanical-turk-content-moderation-ai/ [Accessed 20 Aug 2019].
75 Nguyen C. This text-message hotline can predict your risk of depression or stress, 2016. Available: https://www.businessinsider.com.au/crisis-text-line-is-gathering-data-about-depression-stress-2016-6 [Accessed 20 Aug 2019].
76 Fussell S. The AI that could help curb youth suicide, 2019. Available: https://www.theatlantic.com/technology/archive/2019/07/google-partners-lgbt-suicide-prevention-nonprofit/593821/ [Accessed 20 Aug 2019].
77 Vogel L. AI opens new frontier for suicide prevention. *CMAJ* 2018;190:E119.
78 Turning Point. Google grants eastern health's turning point $1.2 million to establish world-first suicide monitoring system, 2019. Available: https://www.turningpoint.org.au/about-us/news/google-grant-turning-point-suicide-surveilance [Accessed 20 Aug 2019].