PRIORIS: Enabling Secure Detection of Suicidal Ideation from Speech using Homomorphic Encryption

Deepika Natarajan * Anders Dalskov † Daniel Kales ‡ Shabnam Khanna §

November 24, 2020

Abstract

Suicidal ideation is a major health concern in the United States, with many millions of people reporting experiencing serious suicidal thoughts each year. Early detection of suicidal thought is critical in preventing suicide attempts and treating affected individuals. Recent research has shown how machine learning can be used to detect suicidal ideation from phone speech data. However, given the very sensitive nature of the data involved in this process (i.e. phone conversations of at-risk persons and prediction results), it is difficult to imagine how such an application could be used in practice. To address this issue, we investigate a privacy-preserving variant of the ideation detection application flow involving homomorphic evaluation of neural networks. We describe multiple realistic use-cases to aid both affected individuals and clinical practitioners that would be enabled as a result of this secure infrastructure. We also give first order performance estimates for homomorphic evaluation of the networks proposed, and discuss various opportunities for further analysis.

1 Introduction

Suicidal ideation, or the state of thinking about or planning a suicide, is a major public health concern in the United States. In 2015 alone, an estimated 9.8 million adults in the US reported having serious suicidal thoughts [AAfBHS]. Moreover, according to the United States Center for Disease Control, the national suicide rate increased by 33 percent between 1999 and 2017 [fDCP]. Early detection of suicidal ideation is critical to prevent suicide attempts and provide treatment for individuals. Yet, in spite of major advances in the fields of medical and psychological science, our ability to predict suicide has remained roughly constant for at least several decades [FRF+17].

Clinical practitioners typically rely on self-report of suicidal thoughts in order to diagnose suicidal patients. However, this method of diagnosis is problematic, since a majority of individuals who die from suicide deny suicide ideation in their last communication about the subject before death [IHM+95, BF04]. Additionally, the current system relies on clinical assessment as a primary means of identifying suicidal ideation. This means that individuals who do not make a habit of regular clinical assessments, for instance, due to concerns about cost of treatment, time constraints, lack of access, feelings of depression/lack of motivation, or social stigma, do not receive adequate diagnosis and treatment [NRTG14, CHE+13].

In order to address some of these inefficiencies, Gideon et al. [GSMP19] proposed a machine learning-based system for detecting suicidal ideation. By determining the emotions present in a subject's natural phone conversations, and noting that individuals with suicidal ideation displayed lower emotional variability than healthy controls, the authors were able to create a machine learning model that could predict the likelihood of suicidal ideation in an individual.

From a security perspective, both the phone call data and the prediction output are extremely sensitive in nature and require complete confidentiality. Any leak of medical data could dramatically affect the patient's well-being, whether through resulting social

*University of Michigan, dnataraj@umich.edu
†Aarhus University, anderspkd@fastmail.com
‡Graz University of Technology, daniel.kales@iaik.tugraz.at
§Centre for Secure Information Technologies (CSIT), Queen's University Belfast, skhanna01@qub.ac.uk

stigma, discrimination by employment or financial institutions, or other types of abuse. The approach taken by Gideon et al. to safeguard user data involves sending encrypted data from the user's phone to a server compliant with U.S. patient privacy laws, which is then able to decrypt and process the user data in the clear. Though this approach may be sufficient for a limited number of users, assuming a heavily safeguarded, small number of private servers used to process the user data, it does not scale well as the number of users increases.

For many applications, a larger user base may signal the need to move from a small private infrastructure to a larger cloud-like environment, where equipment and maintenance costs can be outsourced and/or shared amongst multiple cloud customers. However, many works have shown how seemingly secure cloud-based systems are often easily exploitable by attackers, for example, due to the difficulty of detecting bugs in large cloud operating-systems, or via the use of side-channel attacks [ZJRR12, IGI+16]. Thus, though it solves the problem of scalability, this approach has the potential to violate user security and privacy.

Recently, researchers at Microsoft have demonstrated the feasibility of using Homomorphic Encryption to securely outsource Neural Networks predictions for MNIST and CIFAR datasets [GDL+16, BEGB19, BMMP18]. In this work, we investigate the feasibility of securely detecting suicidal intent from speech data using privacy-preserving homomorphic encryption techniques.

Overview. We begin by describing an end-to-end flow for suicidal ideation detection, first shown to be effective in [GSMP19]. We then describe a privacy-preserving cyclical system of evaluations to further improve suicidal ideation detection and treatment. We discuss at a high level the trade-offs that would need to be considered for this implementation and give a first-order approximation of the performance of our proposed approach, using [BEGB19] as a reference for homomorphic operation performance. Finally, we discuss extensions of this work and identify several interesting areas for future analysis.

2 Suicide Ideation Detection

Researchers at the University of Michigan have demonstrated the feasibility of detecting suicide ideation from natural phone call speech data using their PRIORI smartphone application [GSMP19]. In their analysis, the researchers make use of two main neural networks: a convolutional neural network (CNN), consisting of a feature encoder and an emotion classifier, and a dense neural network (DNN). The CNN and DNN are shown in Figure 1 and Figure 2, respectively. We use their process as a basis for our proposed application, and describe key components of their approach below.

2.1 Dataset

The Ecological Measurement of Affect, Speech, and Suicide (EMASS) dataset is a collection of natural phone conversations and regular reports of emotion, mood, and suicidal thoughts. Specifically, it consists of over 400 hours of phone conversations recorded from 43 different participants, including healthy control samples and patients who experience suicidal ideation. The calls were recorded by the PRIORI phone application over the course of 8 weeks. The authors of [GSMP19] use the EMASS dataset to train and test their models. The collection of this dataset is still ongoing; however, the authors plan to publish the extracted EMASS dataset features upon completion of the study.

2.2 Application

The PRIORI application, as described in [GSMP19], utilizes neural networks to perform various operations. For convenience, we give the steps of the application flow below (for inference only):

1. Use the PRIORI phone application to save user-side call audio.
2. Use an algorithm for speech activity detection (such as the COMBO-SAD algorithm [SH13]) to extract short segments of uninterrupted speech from a call.

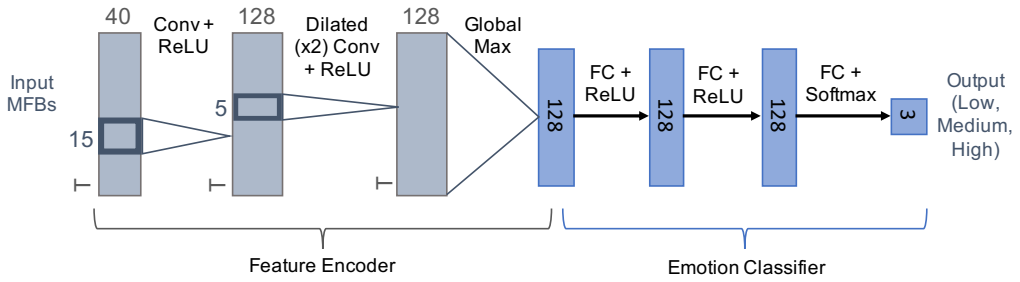

Figure 1: MADDoG Convolutional Neural Network, which consists of a Feature Encoder (left) and an Emotion Classifier (right). Figure adapted from [GMM19].

3. Divide each segment into overlapping frames and extract a 40-dimensional log Mel-filter bank (MFB) spectrum. This will result in a matrix of 40-by-t_i, where t_i is the number of frames in segment i.

4. Pad the above matrix with enough zero vectors to get a 40-by-T matrix, where T is the maximum number of frames in a segment for all segments in the training set.

5. Feed the 40-by-T matrix into two separate MADDoG Feature Encoders to get "segment-level" representations of the data.

6. Feed the outputs from the MADDoG Feature Encoder into two separate MADDoG Emotion Classifiers (one for valence and one for activation). Each classifier output will be a 3-element vector denoting "low", "medium", or "high" for valence or activation, resulting in a 6-element result. See Figure 1 for more details.

7. Repeat the above two steps for all segments in a call. This will result in a 6×N matrix, where N is the number of segments in a call.

8. Take 31 statistics (including mean, standard deviation, skewness, kurtosis, min, max, range, and statistics from performing a linear regression) across each row in the matrix from the previous step. This will result in a final feature vector of $6 \times 31 = 186$ elements per call.

9. Feed the 186-dimensional vector into five separate DNNs, where each DNN is trained to classify one of the following emotions: Guilt, Hopelessness, Anger at Others, Anger at Self, and Irritability. Each DNN consists of four hidden layers with widths of 1024, 512, 256, and 256, respectively. The activation function of the hidden layers is a RReLu (Randomized Leaky ReLu), which corresponds to a LReLu (Leaky ReLu) during model evaluation. The final layer uses a sigmoid activation function and outputs a 3-element vector, denoting a rating of 0, 0.5, or 1. These ratings correspond to ratings on a Likert scale of 1-5 for emotion intensity.

10. Repeat the above steps for a set of calls. For each of the 5 emotions, calculate the (within-subject) standard deviation. This is representative of emotion variability across a set of calls.

11. Take the average of the 5 standard deviation values found in the previous step. Use the result as a measure of the average emotion variability over a set of calls for a particular individual.

12. Use either a threshold or a linear classifier to determine whether the output from the previous step indicates suicidal ideation.

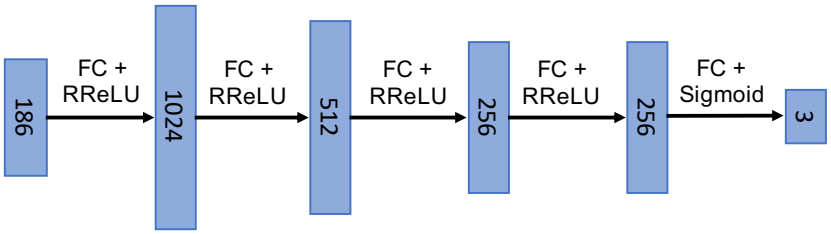

Figure 2: Dense Neural Network used for Emotion Identification, as described in [GSMP19].

3 Use cases

As mentioned previously, the PRIORI application-based system described in [GSMP19] sends raw speech audio to a remote server to be processed in the clear. This speech data and the resulting network prediction constitute highly sensitive information, especially since the PRIORI application records all (user-side) natural phone call conversations during day-to-day life.

We propose modifications to the PRIORI application flow that would allow for a more secure approach to suicide ideation detection. Namely, our approach would protect all user-created data as well as the result of the suicide ideation prediction from cloud adversaries. We call this approach "PRIORIS" to refer to a secure version of the PRIORI approach. In this approach, the ideation detection flow would be segmented as follows:

1. Audio recording, speech activity detection, and MFB extraction
2. Evaluation of MADDoG Feature Encoders and Emotion Classifiers
3. Calculation of statistics across result
4. Evaluation of Emotion DNNs
5. Calculation of standard deviation, average, and threshold/linear classifier

We envision that steps 1, 3, and 5 would be computed in-the-clear on the local smartphone device, while steps 2 and 4 (i.e. the neural networks) would be calculated homomorphically in the cloud. This would add an additional homomorphic encryption step and data communication step between steps (1,2) and (3,4), as well as an additional homomorphic decryption step and data communication step between steps (2,3) and (4,5).

Preserving the privacy of this speech data could persuade more people to use emotion detection recognition technology outside the context of clinical studies. Consequently, the security guarantees afforded to the application flow by our proposed modifications could render a variety of new opportunities for secure deployment. We identify three such use cases and describe them below:

Use-Case 1: Secure Detection and Response

In this scenario, the goal of the application would be to understand and respond to the mental health status of an individual. For example, when the application predicts that a user is experiencing suicidal inclinations, it could alert the user and recommend a clinical visit, potentially even displaying locations, hours of operation, and/or open appointment slots for nearby clinics. In more extreme cases (e.g. when the prediction of suicidal ideation is strong), the application could display the phone numbers of suicide prevention hotlines or even immediately connect an individual to a hotline volunteer or trained professional.

Use-Case 2: Secure Clinical Assessment Assistance

The concept of using speech patterns to identify mood disorders is not new; clinicians typically consider speech factors such intonation, conversation dominance, and voice level (e.g. quiet, loud) when diagnosing patients with mood disorders [Guy76, YBZM78]. Figure 3 (Block 2) shows how the application could be used to augment the capabilities of clinicians to understand the mental health of their patients from speech-level information.

Importantly, this application would allow professionals to take into account predictions made over a larger group of people. In the case that the application prediction matches that of the clinician, this could help a clinician be more confident in their diagnosis. In the case where the prediction differs, the clinician could recommend a follow-up screening. In the controlled setting of an in-person appointment, special recording equipment can be used instead of the patient's phone application.

We note that there may be differences between the ability of the network to predict suicidal ideation from structured speech (i.e. question-response, clinical assessment) versus natural speech (i.e. phone calls to friends and family). In this case, the model used for clinical assessment would have to be trained differently from the model used in use-cases 1 and 3. This will need to be investigated in future work.

Use-Case 3: Secure Treatment Evaluation

In many cases, suicidal ideation can be linked to a mental health disorder which can be treated [Brå18]. As is the case with many mental health disorders, the treatment procedure is usually a very iterative process [JBC+10]. For example, this may take the form of trying a certain dose of a medication for a month, re-evaluating symptoms at another clinical visit, adjusting the medication dose, re-evaluating symptoms at a clinic again, one month later, etc. Additionally, psychotherapy may be used to a varying degree before the best therapy schedule is determined.

As mentioned earlier, relying on patients to self-report suicidal ideation is problematic, as a majority of people who die from suicide deny suicide ideation in their last communication before death [BFJ03]. In addition to outright denial, patients may simply be unable to detect suicidal behavior in themselves, simply because they have not been trained to do so. Moreover, a patient may rely on memory alone to describe how their mental health has been affected as a result of the particular treatment iteration. This reliance on patient memory is problematic, as it is unrealistic to expect an individual to remember every detail of their mood changes across extended periods of time.

To help solve the above problems, the application could be used to track effectiveness of treatment over time, for example, by recording the suicidal ideation prediction values over the course of a month and plotting the change in values on a graph, as shown in Figure 3 (Block 3). Downward trends could be interpreted as relative ineffectiveness of a treatment iteration, while upward trends could be interpreted as relative effectiveness of the iteration. The clinician could evaluate effectiveness of treatment without having to rely solely on patient recollection at the time of visit. As in use-case 1, the application may also respond with helplines or clinic availability for particularly strong predictions of suicidal inclinations.

We note that the above use cases could be proposed without any notion of security. However, we argue that the data input and application output constitute extremely sensitive information, and users would not use the application without robust security guarantees. Thus, the use cases we discuss are only possible at scale with the type of protection offered by the secure network evaluation we propose.

We also note that the above use cases are related. An individual may initially use the application for a preliminary diagnosis to decide whether they should seek further evaluation from a clinician (use-case 1). During the clinical visit, the clinician may use the application to confirm their diagnosis, utilizing the results from case 1 where helpful (use-case 2). If diagnosed with a mental illness associated with suicide ideation, the individual would use the application to monitor the effectiveness of the initial treatment plan (use-case 3). The next clinical appointment would involve use-case 2 followed by use-case 3 once again. In this way, the application could be used in a cyclical manner to enable a more accurate, effective, and efficient treatment process.

4 Network Training

This work mainly focuses on homomorphic evaluation of the described networks. Accordingly, we assume that the models referenced are trained beforehand. Nevertheless, we wish to devote some discussion as to how such models could be obtained in practice.

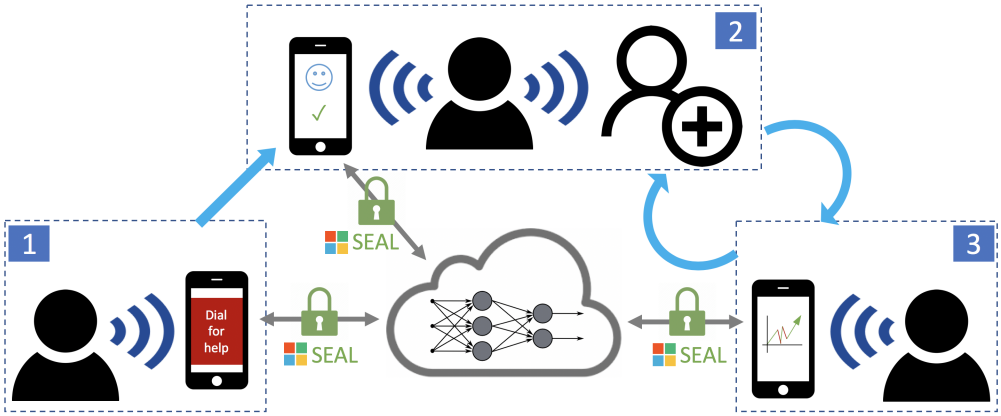

Figure 3: Use-cases for proposed secure suicide ideation detection application. Each block number corresponds to the use-case of the same number: 1) Suicide hotline connection upon detection of suicidal ideation, 2) Validation of clinical assessment, 3) Monitoring of treatment effectiveness over time. Use cases may be related to each other as depicted by (light blue) arrows between blocks. Arrows between blocks and cloud denote HE-based data encryption and model evaluation, using Microsoft SEAL library as an example HE infrastructure.

The authors of [GSMP19] have already demonstrated how useful models can be generated using the EMASS dataset, which we summarized in Section 2. The models they were able to train using the EMASS dataset have proven successful at using natural phone conversations to distinguish healthy controls from suicidal individuals, achieving an AUC (Area Under the Curve) of 0.79. Datasets such as EMASS could therefore be used to build initial networks.

For optimal performance, it is likely that much more data would need to be collected in order to further train the initial models. We imagine that successful deployment of this application would encourage enough users to volunteer their data for network training. However, the inputs and outputs of the networks described above contain highly sensitive data; thus, it may not be plausible that enough users would be willing to volunteer this information. Moreover, it is possible that selecting volunteers in this manner would significantly skew the set of training data such that it no longer resembles testing data (for example, users may only allow evaluation of more "benign" calls, such as those made to customer service lines, rather than calls they make to family and friends).

Ideally, we would like to collect enough useful data from users while protecting user privacy. In order to ensure patient privacy during the training process, a variety of approaches could be taken. Federated learning, for example, which has been popularized in recent years by Google, could allow models to be trained locally and combined later in a privacy-preserving manner [SMK+17]. Homomorphic training could also be used to preserve patient privacy. We note, however, that while some works show homomorphic training as possible, other works report the technique as practically infeasible [NRPH19]. Future work would therefore require a much deeper analysis of this component.

5 Homomorphic Network Evaluation

The PRIORI application flow involves the use of two types of neural networks: a CNN (which consists of a Feature Encoder and an Emotion Classifier) and a DNN (used to identify emotions). We now wish to analyze the amenability of each of these networks to homomorphic evaluation.

We follow the approach of CryptoNets [GDL+16], which first described how to homomorphically process each layer of a CNN used to classify MNIST images. Specifically, we approximate the ReLU activation functions with square activation functions (i.e. a low-degree polynomial), replace "pool" layers with "scaled pool" layers, and do not homomorphically

evaluate any final sigmoid activation layers. We also make two further modifications: 1) we do not homomorphically evaluate any final softmax layers, since, like the sigmoid layers, these are necessary for training but not required for evaluation, and 2) we replace RReLU activations with ReLU activations (which we approximate with square activations), since these operations are similar given limited RReLU leakage [XWCL15]. We also model the dilated convolution layer the same way as a convolution layer, since both are implemented as a weighted sum. A more detailed description of the RReLU approximation is given in the next section.

Tables 1 and 2 give the modified layers for the described CNN and DNN, respectively, as well as their per-layer homomorphic evaluation runtimes. We obtain these estimates through simple scaling of the execution times of similar layers used in CryptoNets 3.2 [BEGB19], which uses the BFV encryption scheme. The CryptoNets 3.2 numbers were obtained from running the CrytoNets application on a single Intel Xeon E5-1620 CPU at 3.5GHz, with 16GB of RAM and Windows operating system. Note that these times assume that model weights and bias values are unencrypted. We set the CNN Feature Encoder dimension T to 600, which corresponds to a 6 second average segment length and 10ms frame shift length for MFB extraction.

Using the first order approximation, we obtain full network evaluation time estimates of 16777.178 seconds and 194.656 seconds for the CNNs and DNNs, respectively. The authors of [KPS+14b] use a dataset similar to the EMASS dataset for monitoring mood from speech data and report that the phone calls made consists of 24.3 +/- 46.6 segments on average. Assuming a similar 24 segments per call, sequential application of the CNN should take approximately 402652.272 seconds (111.848 hours) per call. Sequential application of the DNN should take approximately 194.656 seconds (3.244 minutes) per call. We stress that these estimates do not include any batching, pipelining, or parallelization techniques, each of which are expected to provide significant performance benefits (up to multiple orders of magnitude). Simply processing each of the 24 segments in parallel, for example, would result in only 4.66 hours per call for application of the CNN.

HE Layer	Time Estimate (sec)
Conv.	11247.291
Square Activation	886.156
Dilated Conv.	3749.097
Square Activation	886.156
Scaled Max Pool	8.478
FC	3.072
Square Activation	1.477
FC	3.072
Square Activation	1.477
FC	0.072
Total	6476.331

Table 1: Layers proposed for homomorphic evaluation of MADDoG Feature Encoder and Emotion Classifier CNN and corresponding first order approximations of evaluation times. Results were obtained through simple scaling of execution times for similar layers used by the CryptoNets v3.2 MNIST CNN [BEGB19], and assume $T = 600$.

Activation Functions As stated above, we follow the approach of CryptoNets and replace ReLU with square activation functions. Although such low-order polynomials can be used to approximate ReLU activations, there are places in which the functions differ significantly. It is therefore vital to empirically evaluate whether a such an approximation still achieves accurate results. The CryptoNets work achieved an accuracy of 99 percent using this square activation approximation of ReLU (for an MNIST classification network). Therefore, it is plausible that this approximation could be used successfully in the networks we describe as well.

The case of RReLU, however, is more complicated. As noted above, we chose to replace RReLU with regular ReLU activations (which we then approximate with square). When

HE Layer	Time Estimate (sec)
FC	35.712
Square Activation	11.815
FC	98.305
Square Activation	5.908
FC	24.576
Square Activation	2.954
FC	12.288
Square Activation	2.954
FC	0.144
Total	194.656

Table 2: Layers proposed for homomorphic evaluation of Emotion Detection DNN and corresponding first order approximations of evaluation times. Results were obtained through simple scaling of execution times for similar layers used by the CryptoNets v3.2 MNIST CNN [BEGB19].

the choice of leakage is small, the two functions are similar (i.e. $\mathsf{RReLU}_\alpha(x) = \max(\alpha x, x) \approx \max(0, x) = \mathsf{ReLU}(x)$ for small α.) The authors of [GSMP19] do not specify the particular α they use, though they do refer to [XWCL15] (which explores small α values, between 0.01 and 0.2) as motivation for the choice of activation function . Nonetheless, the authors of [GSMP19] do not compare the accuracy they achieved with RReLU to accuracy possible with ReLU. This should be explored in future work.

Finally, we note that while the proposed approximation could still yield accurate results, it may render training the network more difficult. As noted in [GDL$^+$16], in particular, the derivative of x^2 is not bounded. Although this may result in strange behavior during gradient descent, the authors of [GDL$^+$16] have successfully combated this issue by adding extra convolution layers without activation layers to prevent overfitting. The effect of this approximation on network training should be assessed in future work.

6 Extensions and Future Work

In the previous sections, we described PRIORIS application for secure suicide ideation detection in the context of three main use-cases. In this section, we describe some additional extensions to the application and opportunities for future work.

Adaptation of application to other types of mood disorder detection. Suicide ideation is highly related to mood disorders, which in turn often result in altered speech patterns. This suggests that speech data may be used to identify other types of mood disorders. In fact, this type of analysis has already been shown useful for detecting disorders such as bipolar disorder, depression, and post-traumatic stress disorder [KPS$^+$14a, MBQ$^+$19]. Future work could analyze other types of mood disorder detection for their amenability to FHE-based private-preserving machine learning. Generalizing further, it may even be useful to construct a general mood detection and tracking application in combination with therapeutic smartphone applications such as those commonly used for mindfulness and relaxation.

Determining Optimal Intervention. To improve the usefulness of the proposed system, it would be beneficial to identify exact moments when medical intervention is required. The authors of [GMA$^+$19] have already explored this question in the context of bipolar disorder. Their method involves using an initial data collection period to establish a baseline emotion level. The authors then use anomaly detection techniques to compare subsequent user behavior relative to this baseline in order to determine optimal medical intervention points. This type of investigation and fine-tuning could be extended to the context of suicide ideation detection and treatment. This is particularly significant for uses cases involving smartphone monitoring of medical symptoms, as these devices have the potential to

provide intervention close to the time of need (see: for example, suicide prevention hotline connection in use-case 1 above).

Choice of FHE Scheme. We perform analysis of the layer execution times with respect to the BGV homomorphic encryption scheme, since this is the scheme used by the CryptoNets 3.2 MNIST network. Microsoft SEAL also implements the CKKS scheme, which differs from BGV in its ability to efficiently compute approximate computations on real-valued data. CKKS is thus particularly amenable to machine learning use cases, as neural networks typically make use of approximate values. Future work should investigate the performance of the using the BGV scheme relative to using the CKKS scheme for the proposed application.

Analysis of detection segmentation flow. In section 2.1, we proposed an initial segmentation of the application flow with respect to computation execution location. However, this initial segmentation is based on intuition. Future work should analyze the trade-off between executing each step of the full procedure either locally in the clear or homomorphically on the server, including the resulting impact on overall performance, energy, and storage requirements of the application.

The modifications we propose as a result of this analysis, as well as the choice of FHE scheme, will likely result in variable model accuracy. However, any loss in accuracy could potentially be offset by increasing the amount of available training data. Our proposed application places a strong emphasis on user privacy guarantees, and would thus encourage more widespread adoption of the PRIORIS application. This, in turn, could increase the amount of training data available (e.g. by informing more users about the study, some of whom may volunteer to have their data added to the training set, or via the secure training techniques mentioned in section 4) and render the system robust enough to be integrated into everyday mood health monitoring and clinical assessment.

7 Conclusion

In this work, we propose a privacy-preserving based suicidal ideation detection flow. We describe how homomorphic encryption could be used for secure inference of networks previously shown useful for detecting suicidal ideation from phone speech data. We also describe multiple use-cases that are enabled as a result of the proposed security mechanisms. Finally, we give first-order approximations of homomorphic evaluation runtimes for the models used in our application, and describe several directions for future research.

8 Acknowledgements

This work was created in collaboration with researchers at Microsoft. The authors would like to thank the organizers of Microsoft's Private AI Bootcamp for their helpful feedback and reviews. The authors would also like to thank researchers in the Computational Human Artificial Intelligence (CHAI) Lab at the University of Michigan for their helpful discussions.

References

[AAfBHS] Substance Abuse, Mental Health Services Administration, and Center for Behavioral Health Statistics. Suicidal thoughts and behavior among adults:results from the 2015 national survey on drug use and health.

[BEGB19] Alon Brutzkus, Oren Elisha, and Ran Gilad-Bachrach. Low latency privacy preserving inference. In *International Conference on Machine Learning*, 2019.

[BF04] Katie A Busch and Jan Fawcett. A fine-grained study of inpatients who commit suicide. *Psychiatric Annals*, 34(5):357–364, 2004.

[BFJ03] Katie A Busch, Jan Fawcett, and Douglas G Jacobs. Clinical correlates of inpatient suicide. *The Journal of clinical psychiatry*, 2003.

[BMMP18] Florian Bourse, Michele Minelli, Matthias Minihold, and Pascal Paillier. Fast homomorphic evaluation of deep discretized neural networks. In *CRYPTO (3)*, volume 10993 of *Lecture Notes in Computer Science*, pages 483–512. Springer, 2018.

[Brå18] Louise Brådvik. Suicide risk and mental disorders, 2018.

[CHE+13] Ewa K Czyz, Adam G Horwitz, Daniel Eisenberg, Anne Kramer, and Cheryl A King. Self-reported barriers to professional help seeking among college students at elevated risk for suicide, 2013.

[fDCP] Centers for Disease Control and Prevention. Webbased injury statistics query and reporting system (WISQARS). Online, accessed 2020-01-21. Available at URL: `https://www.cdc.gov/injury/wisqars/index.html`.

[FRF+17] Joseph C Franklin, Jessica D Ribeiro, Kathryn R Fox, Kate H Bentley, Evan M Kleiman, Xieyining Huang, Katherine M Musacchio, Adam C Jaroszewski, Bernard P Chang, and Matthew K Nock. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychological bulletin*, 143(2):187, 2017.

[GDL+16] Ran Gilad-Bachrach, Nathan Dowlin, Kim Laine, Kristin E. Lauter, Michael Naehrig, and John Wernsing. CryptoNets: Applying neural networks to encrypted data with high throughput and accuracy. In *ICML*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 201–210. JMLR.org, 2016.

[GMA+19] John Gideon, Katie Matton, Steve Anderau, Melvin G McInnis, and Emily Mower Provost. When to intervene: Detecting abnormal mood using everyday smartphone conversations, 2019.

[GMM19] J. Gideon, M. McInnis, and E. Mower Provost. Improving cross-corpus speech emotion recognition with adversarial discriminative domain generalization (ADDoG). *IEEE Transactions on Affective Computing*, 2019.

[GSMP19] John Gideon, Heather T Schatten, Melvin G McInnis, and Emily Mower Provost. Emotion recognition from natural phone conversations in individuals with and without recent suicidal ideation. In *The 20th Annual Conference of the International Speech Communication Association INTERSPEECH 2019*, 2019.

[Guy76] William Guy. *ECDEU assessment manual for psychopharmacology*. US Department of Health, Education, and Welfare, Public Health Service . . . , 1976.

[IGI+16] Mehmet Sinan Inci, Berk Gulmezoglu, Gorka Irazoqui, Thomas Eisenbarth, and Berk Sunar. Cache attacks enable bulk key recovery on the cloud. In *International Conference on Cryptographic Hardware and Embedded Systems*, pages 368–388. Springer, 2016.

[IHM+95] Erkki T Isometsä, Martti E Heikkinen, Mauri J Marttunen, Markus M Henriksson, Hillevi M Aro, and Jouko K Lönnqvist. The last appointment before suicide: is suicide intent communicated? *The American journal of psychiatry*, 1995.

[JBC+10] Douglas G Jacobs, Ross J Baldessarini, Yeates Conwell, Jan A Fawcett, Leslie Horton, Herbert Meltzer, Cynthia R Pfeffer, and Robert I Simon. Assessment and treatment of patients with suicidal behaviors. 2010.

[KPS+14a] Z. N. Karam, E. M. Provost, S. Singh, J. Montgomery, C. Archer, G. Harrington, and M. G. Mcinnis. Ecologically valid long-term mood monitoring of individuals with bipolar disorder using speech. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4858–4862, May 2014.

[KPS+14b] Zahi N Karam, Emily Mower Provost, Satinder Singh, Jennifer Montgomery, Christopher Archer, Gloria Harrington, and Melvin G Mcinnis. Ecologically valid long-term mood monitoring of individuals with bipolar disorder using speech. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4858–4862. IEEE, 2014.

[MBQ+19] Charles R. Marmar, Adam D. Brown, Meng Qian, Eugene Laska, Carole Siegel, Meng Li, Duna Abu-Amara, Andreas Tsiartas, Colleen Richey, Jennifer Smith, Bruce Knoth, and Dimitra Vergyri. Speech-based markers for posttraumatic stress disorder in us veterans. *Depression and Anxiety*, 36(7):607–616, 2019.

[NRPH19] Karthik Nandakumar, Nalini K. Ratha, Sharath Pankanti, and Shai Halevi. Towards deep neural network training on encrypted data. In *CVPR Workshops*, page 0. Computer Vision Foundation / IEEE, 2019.

[NRTG14] Thomas Niederkrotenthaler, Daniel J Reidenberg, Benedikt Till, and Madelyn S Gould. Increasing help-seeking and referrals for individuals at risk for suicide by decreasing stigma: The role of mass media. *American journal of preventive medicine*, 47(3):S235–S243, 2014.

[SH13] S. O. Sadjadi and J. H. L. Hansen. Unsupervised speech activity detection using voicing measures and perceptual spectral flux. *IEEE Signal Processing Letters*, 20(3):197–200, March 2013.

[SMK+17] Aaron Segal, Antonio Marcedone, Benjamin Kreuter, Daniel Ramage, H. Brendan McMahan, Karn Seth, Keith Bonawitz, Sarvar Patel, and Vladimir Ivanov. Practical secure aggregation for privacy-preserving machine learning. In *CCS*, 2017.

[XWCL15] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network. *CoRR*, abs/1505.00853, 2015.

[YBZM78] Robert C Young, Jeffery T Biggs, Veronika E Ziegler, and Dolores A Meyer. A rating scale for mania: reliability, validity and sensitivity. *The British journal of psychiatry*, 133(5):429–435, 1978.

[ZJRR12] Yinqian Zhang, Ari Juels, Michael K Reiter, and Thomas Ristenpart. Cross-vm side channels and their use to extract private keys. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 305–316, 2012.