IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 6.252
Volume 2, Issue 7, May 2022

Suicide Ideation Detection on Social Media Using a Time-Aware Transformer Model

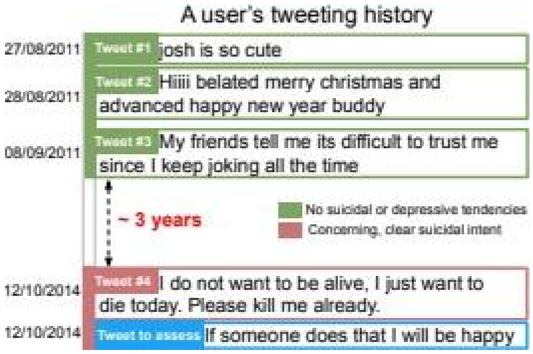
Ms. Sushma A Shirke, Abhishek Bisht, Aman Semwal, Amandeep Rawat, Vijender
Army Institute of Technology, Pune, Maharashtra, India, Pune, Maharashtra, India

Abstract: *The widespread usage of social media allows individuals to express suicidal concernsl therapy settings. Recognizing the development of such ideas is crucial for the Suicide prevention and identifying risk of users is frequently linked to a mental illness. The psychological range of a user's previous social media activity can reveal their mental state changes over time. This project focuses on detecting suicidal intent in tweets in English by adding historical context to linguistic models STATENet is a timeaware model for preliminary suicide risk screening on social media. STATENet outperforms the competition proving the value of emotional and temporal contextual factors in assessing suicide risk. STATENet for detecting suicide is discussed in terms of its, qualitative, practical, and ethical implications.*

Keywords: STATENet

I. INTRODUCTION

Each year, about 800,000 people dies by suicide around the world, with 20 times as many more attempting suicide. Suicide is the second largest cause of mortality among those aged 15 to 29, according to the World Health Organization (WHO), with a 35 percent increase in the US since 1999. . Providing clinical and psychological therapy to those who have suicidal thoughts is dependent on recognising those who are at risk. Unfortunately, 80% of patients do not receive psychiatric care, and almost 60% of those who died by suicide denied having suicidal thoughts to mental health professionals . People with suicide ideation frequently utilise social media, such as Twitter, to disclose their mental status, according to recent studies , with eight out of ten sharing their suicidal thoughts and plans.


A user's tweeting history
27/08/2011
28/08/2011
08/09/2011
~ 3 years
No suicidal or depressive tendencies
Concerning, clear suicidal intent
12/10/2014
12/10/2014

Tweet #1 josh is so cute

Tweet #2 Hiiii belated merry christmas and advanced happy new year buddy

Tweet #3 My friends tell me its difficult to trust me since I keep joking all the time

Tweet #4 I do not want to be alive, I just want to die today. Please kill me already.

Tweet to assess If someone does that I will be happy

Figure 1: A user's most recent tweet does not indicate suicide intent. It's difficult to appropriately measure suicidal risk without seeing the user's most recent past tweet, which demonstrates self-harm tendencies. However, evaluating a user's tweeting history in order without taking into account time gaps between tweets may result in an erroneous portrayal of the user's mental state.

While recent breakthroughs in computational social science have made headway in assessing suicidal risk on social media , studying the linguistic features of tweets is often insufficient for reliable suicidal intent identification. Additional user-level factors, such as tweeting history, can help identify a build-up of negative feelings, which are frequently associated to suicide thoughts. Suicidal ideation can develop weeks, months, or even years before it manifests, and suicidal action might be influenced by previous suicide attempts or thoughts. Figure 1 shows how analysing a user's history and emotion spectrum might provide significant context for estimating suicidal risk in a tweet written by that user. A user's Emotional Historic Context (EHC) throughout time can be indicative of their mental health. Suicidal intent can be detected by modelling temporal user context as a bagof-tweets or sequentially. However, as shown in Figure 1, the impact of different time intervals between tweets is critical for an accurate evaluation. The wide gap between the user's recent tweets that collectively indicate suicidal intent and those three years apart must be modelled. Between successive tweets, such unequal Temporal Tweeting Irregularities (TTI) spanning from seconds to years influence the evaluation of a user's tweet in different ways. Long Short Term Memory (LSTM) networks, for example, presume that posting intervals are uniform, limiting a user's ability to understand their emotional spectrum over different time intervals.

1.1 Contributions

We propose STATENet, which takes into account a user's emotional historical context as well as temporal tweeting abnormalities. Suicide risk evaluation A neural framework called the Time-Aware TEmporal Network evaluates the existence of suicidal intent on social media (Sec. 3.1). STATENet employs a dual transformer-based architecture to learn the linguistic and emotional cues in tweets, building on the success of transfer learning in Natural Language Processing. In a time-sensitive manner, STATENet learns from the language of the tweet (Section 3.2) to be analyzed, as well as the past emotional spectrum of a user (Sec. 3.3). We show that STATENet greatly outperforms competitor approaches (Sec. 5) in a series of tests (Sec. 4) using real-world data (Sec. 4.1), with an F1 Score of 80%. We use a qualitative analysis to demonstrate practical application (Section 5.4), and we examine the study's ethical implications (Sec. 6).At the very least, we establish the validity of using timeaware emotional temporal context to identify suicide ideation on social media. We concentrate on the junction between NLP and suicide risk assessment by taking a non-intrusive step toward better risk assessment Our work might be viewed as an early screening tool that, hopefully, will become part of a wider infrastructure encompassing psychologists and health care providers.Providers, and social media businesses In practise, STATENet would flag tweets as "at-risk" for suicidality as part of a human-in-the-loop mechanism to aid decision-making about potential intervention.

II. RELATED WORKS

- **Traditional Methods**:- The Suicide Probability Scale , Depression Anxiety Stress Scales-21 , Adult Suicide Ideation Questionnaire, Suicidal Affect Behavior-Cognition Scale , and others have been developed by researchers to assess suicidal risk.While these methods are professional and effective, they rely on participants to complete questionnaires or participate in interviews , excluding suicidal people who are either unable to access these resources or have a low motivation to seek professional help . According to research, conducting a suicide risk assessment can have a negative influence on those who are depressed
- **NLP Methods**: In recent years, social media has shown promise in terms of delivering insights into people's mental health. Twitter is a suitable tool for real-time suicide risk monitoring. User features and online suicide notes were among the first attempts to use social media. Since then, the focus has shifted to textual elements like POS and tense, as well as psycholinguistic lexicons like LIWC .to classify. The use of deep learning for suicidality prediction has increased in shared projects like and CLEF. CNN and LSTM architectures use pre-trained word embeddings to predict suicide risk. Although these text-based algorithms capture the semantic character of posts in isolation, there is no user-associated context that can provide insight into the user's mental state in order to improve prediction power.
- **Contextual Methods**: The best performing model at the CLPsych 2019 shared task for suicidal estimation on Reddit, the dual context BERT, shows the utility of temporal context. The Dual Context BERT employs post-level BERT embeddings that are progressively processed by an attention-based RNN. use an LSTM and fastText-based architecture to model temporal context. These RNN and LSTM-based techniques presume that users' past posts are uniformly spaced in time, which limits the ability of the suicide ideation detection model to learn their relative importance in a time-aware manner. Time-aware sequential models have shown benefits in various clinical tasks, such as patient subtyping, as well as in other areas, such as user activity modeling. More recently, used latent representations of GloVe embeddings of past tweets to model a user's historic mood spectrum. Instead of learning them as sequences, these latent characteristics are aggregated depending on specific functions such as exponential decay and sinusoids. These models presume that suicide ideation follows unique trajectories, which may not generalise well across users and, by aggregating them, lose the context of individual previous tweets.

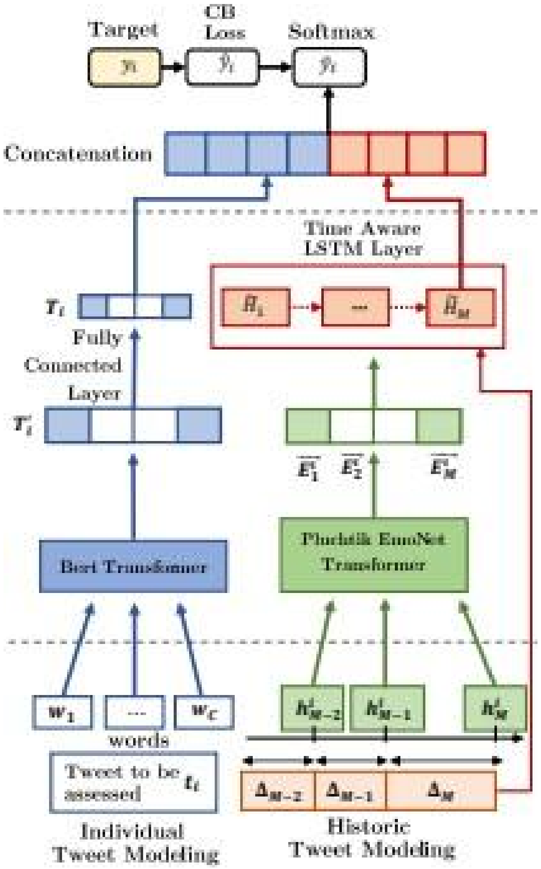


Figure 2: STATENet: Model Architecture

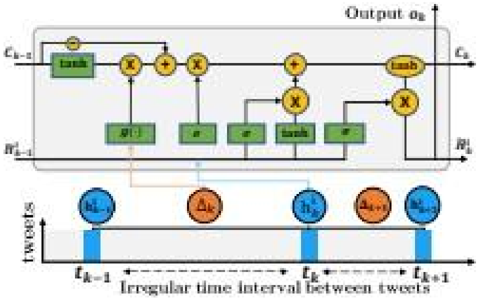
III. METHODOLOGY

3.1 Notations and Problem Formulation

We recognise that representing suicidal intent as a binary classification problem is a significant simplification, hence in this work, we focus on detecting the presence of suicidal ideation inside a tweet utilising a user-level temporal context.

3.2 Encoding the Tweet to be Assessed

Suicidal behaviour is associated with suicidal tweets, according to research. In the past, static word embeddings like GloVe were used to encode tweets for identifying suicidal ideation . Recent research, however, has demonstrated that pre-trained transformer models produce more thorough representations of linguistic characteristics in a tweet . We discovered that SentenceBERT outperforms embeddings utilised in prior works such as FastText, ELMo and others. SentenceBERT's 768-dimensional encoding is used. Formally, T0 i = SentenceBERT(ti) (1) where T0 i 2 R768 is linearly transformed using a dense layer to Ti 2 Rd with dimension d.



3.3 User Historical Emotion Spectrum

Individual Historic Tweet Encoding: Suicide risk can be increased by amplifying emotional components such as emotional reactivity , intensity , and instability . Using this as a foundation, we extract the emotion spectrum of each historic tweet hi. Despite being adept at semantic text modeling, general text encoders struggle to capture the fine-grained emotions represented in social media posts. Plutchik's wheel of emotions is used to capture fine-grained feelings .” This taxonomy proposes three hierarchical groups of eight emotions organised as four opposing dualities.The wheel's core emotions are: Joy - Sadness, Surprise - Anticipation, Anger - Fear, and Trust - Disgust. We produce an encoding that represents the emotional spectrum of a historical tweet, and hence of a user at a historical moment. We fine-tune pre-trained BERT embeddings on the Emonet dataset based on empirical comparisons and the success of transfer learning in NLP . The dataset includes 1,608,233 tweets categorized with 24 emotions according to Plutchik's wheel of emotions. The presence of the principal emotions in the dataset is skewed toward joy, sadness, and fear, with 20.57 percent, 8.85 percent, and 6.13 percent, respectively, with less samples for other emotions. These are categorized using remote supervision and 665 emotion hashtags. This transformer is known as the Plutchik Transformer.” This transformer tokenizes each historical post and inserts the [CLS] token at the start of each one. The aggregate representation of the emotional spectrum is the final hidden state corresponding to this [CLS] token (768-dimensional encoding). The emotion vector (Eik ∈ R768) of each historic tweet hi k is defined as

Ei k = PlutchikTransformer(hi k).

Modeling Historical Tweets Sequentially: The emotional history context of tweets can be utilized to model the author's progressive emotional states. “As a result, recurrent neural networks (RNN), namely LSTMs , are the most natural approaches for encoding and learning from a sequence of a user's previous tweets. However, the time period between uploading previous tweets might range from a few seconds to several years . Such fluctuations can be a significant aspect in assessing a user's emotional states over time. Because LSTM cells assume equally spaced sequences as input, they are unable to predict abnormalities in posting timings of historical tweets. The relative time difference between the user's historical tweets can be used to more correctly model the user's emotions over time. As a result, as shown in Figure 3, we suggest the adoption of a Time-aware LSTM (T-LSTM) , in which the time delay between successive tweets is input to the LSTM cell. The T-LSTM cell thus combines the actual temporal disparities between tweets, as well as the emotional context Eik of each historical tweet. T-LSTM adds memory time decay based on the elapsed time between subsequent elements and weights the short-term memory cell CS k. In theory, the more time that passes between two tweets, the less impact they should have on each other. T-LSTM does this by transforming time into appropriate weights using a monotonically decreasing function of elapsed time”. The T-LSTM incorporates time lapses as follows:

3.4 Joint Network Optimization

STATENet learns from the language of the tweet to be analysed and the emotional historical spectrum in a time-aware manner to determine the presence of suicide intent in a tweet. To do so, we concatenate Ti and H I k, followed by a dense layer with Rectified Linear Unit (ReLU) (Hahnloser et al., 2000) to generate a prediction vector. Finally, the probability of suicidal intent are output using a softmax function.

Suicidal tweets account for a relatively small part of the data . To solve the issue of class imbalance (in practise, the imbalance is considerably greater in the real world), we train STATENet in conjunction with the Focal Loss. This loss function employs a class-wise re-weighting strategy by introducing a weighting factor that is inversely proportional to the sample count. L is the loss function.

IV. EXPERIMENTS

4.1 Dataset

The Twitter timeline data of users from the Sinha et al. dataset is used (2019). started with a lexicon of 143 suicidal expressions and a collection of Twitter tweets. Their final dataset includes 34,306 tweets after human examination for trivially non-suicidal tweets. Some of these tweets were written by the same person, resulting in a total of 32,558 unique users for whom tweets had to be categorized. We summarize the annotation instructions that two annotators, both Clinical Psychology students, followed when annotating the 34,306 tweets collected:

- Suicidal Intent (SI) Present: Suicide ideation or previous attempts are discussed in a serious and non-joking manner in these posts.
- Suicidal Intent (SI) Absent: Songs, condolence messages, awareness, and news are among the tweets with no evidence of a suicide risk.

It's worth noting that this technique generated suicide risk classifications for individual tweets rather than specific user histories. Under the supervision of a competent clinical psychologist, an adequate inter-annotator agreement was attained with a of 0.72. There are 3984 suicidal tweets in the resulting dataset. Each user's Twitter timeline was compiled. From 2009 to 2019, these timelines cover a ten-year period. With a standard deviation of 789 tweets, the average number of tweets in a user's history is 748 (max 3,200). For users with a large amount of historical tweets, we limit the user history to the last 100 tweets. 4 The average time delay between two consecutive tweets for a person is two days, with a standard deviation of over 24 days between tweets, indicating huge differences amongst users. There were 4070 people who had no history tweets.

Data Preprocessing: By recognizing the named entity and moving all identifiable information such as email addresses, URLs, names, etc., we were able to anonymize the dataset. Then, as in the traditional way, the text is converted to lowercase, punctuation and accents are removed, spaces are removed, and stopwords are removed. Divide the tweets in your dataset by user to avoid duplication of users in trains, validations, or test sets. It uses a 70:10:20 layered split between the three sets to generate 24014, 3431, and 6861 tweets in the train, validation, and test sets, respectively. A user can see multiple tweets, but the associated history depends on the publishing timestamp of the tweet. Historical modeling of each classified tweet should only use historical tweets that are older than the ones that were time stamped.

4.2 Experimental Settings

- **Baseline Methods:** We compare STATENet to two types of baseline methods: tweet level (TL) and user-level (recalls), using macro F1 and recalling for suicidal intent existent (admits) (UL). By combining encoding of a tweet to be evaluated with user level attributes, UL baselines were developed for tweet level evaluation.
- **Random Forest + Tweet features**(Sawhneyet al., 2018b) : Random Forests (RF) with tweet level characteristics such as statistics, LIWC features, n-grams, and POS counts are used in this non-contextual TL method. We reproduce the TL deep Neural Network, which uses CNN to record feature points and LSTMs for tweets encoding.
- **Suicide Detection Model :** UL model that encodes tweets using finetuned FastText embeddings . Historic tweets were concatenated with the tweet to be analyzed after passing them through LSTM + attention in order.
- Contextual CNN (Gaur et al., 2019): Non- sequential UL model using GloVe embeddings for encoding tweets. Bag of tweets were concatenated and fed to a contextual CNN).
- **Exponential Decay (Sinha et al., 2019):** TL model that ensembles GloVe embeddings of previous tweets with the GloVe embedding trained on a BiLSTM + Attention for the tweet to be assessed using an exponential decay function.
- **(Mathur et al., 2020):** Surprise and Episodic Modeling: For historic tweet modeling, a decision level ensemble TL model similar to Exponential Decay is used, but it also includes sinusoidal and white Gaussian noise.
- **At CLPsych 2019, DualContextBert (Matero et al., 2019)** was named the best-performing UL model. BERT is used by Dual ContextBert to encode Reddit postings that are supplied to an attention-based RNN layer. We use all of the user's previous tweets in our implementation.
- **Experimental Setup:** For all models, the highest Macro F1 obtained on the validation set is used to determine hyperparameters. To investigate, we utilize grid search: Dropout 2 0.0, 0.1, 0.5, 2 0.99, 0.999, 0.9999 and 2 1.0, 1.5, 2.0 in class-balanced focal loss, initial learning rateIlr 2 0.01, 0.001, 0.0005, 0.0001 and warm-usteps Sws 2 3, 5, 7 H D = 512, n = 1, = 0.5, = 0.9999, = 2.0, Ilr = 0.0001, Sws = 5 were found to be the best hyperparameters. We use PyTorch 1.5 (Paszke et al., 2019) to develop all methods and optimize them using mini-batch AdamW with a batch size of 256 and Ilr = 0.0001.We train the model for 20 epochs and then use 5 epochs of patience to apply early stopping. On an Nvidia Tesla K80 GPU, the model takes 4,361s to train.

V. RESULTS AND ANALYSIS

5.1 Comparative Performance

“Table 1 shows that STATENet outperforms competitive baselines significantly (p 0.005). For suicidal risk assessment, we evaluate both text-only and temporal contextual models. The non-contextual RF + tweet features and C-LSTM models outperform STATENet and other contextual models. This, we believe, is due to the fact that temporal contextual models provide more insight into the author's previous mental state, resulting in increased predicting power. STATENet and sequential models outperform Contextual CNN, owing to their capacity to train representations from the time dependence in previous tweets, rather than the bag of tweets method used by Contextual CNN.We attribute this to STATENet's time-aware LSTM's ability to detect abnormalities in users' tweeting intervals. This type of time-aware modeling is likely to learn more accurate latent representations of users' emotional history. While exponential decay and episodic modelling are both effective, we find that STATENet outperforms both across the board, notably in terms of recall for the suicidal intent present class. This, we believe, is due to the fact that not every user's emotional historical context follows the same predetermined trajectories that these techniques collect historic tweets on”.

Type of Contextual Modeling	Model	Macro F1+	Recall	Accuracy
None	Random Forest + Tweet features	0,536	0,513	0.548
	C-LSTM	0.588	0.597	0.602
Non Sequential	Contextual CNN	0.729	0.587	0.803
Sequential	Suicide Detection Model (SDM)	0.743	0.755	0.819
	DualContextBert	0.767	0.786	0.823
Specific Temporal Functions	Exponential Decay	0.737	0.759	0.828
	Surprise and Episodic Modeling	0.741	0.762	0.831
Timeaware Sequential	STATENet	0.799*	0.810*	0.831*

Table 1: Means of results obtained over ten different runs

Model component	Macro F1	Recall
Current tweet only	0.731	0.551
Current + Random History (Plutchik)	0.730	0.680*
Current + Sequential History(bert)	0.767*	0.786*
Current + TA History (Plutchik)	0.799*	0.810*

Table 2: Ablation Study over STATENet

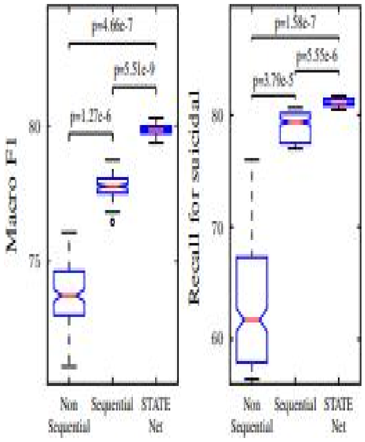
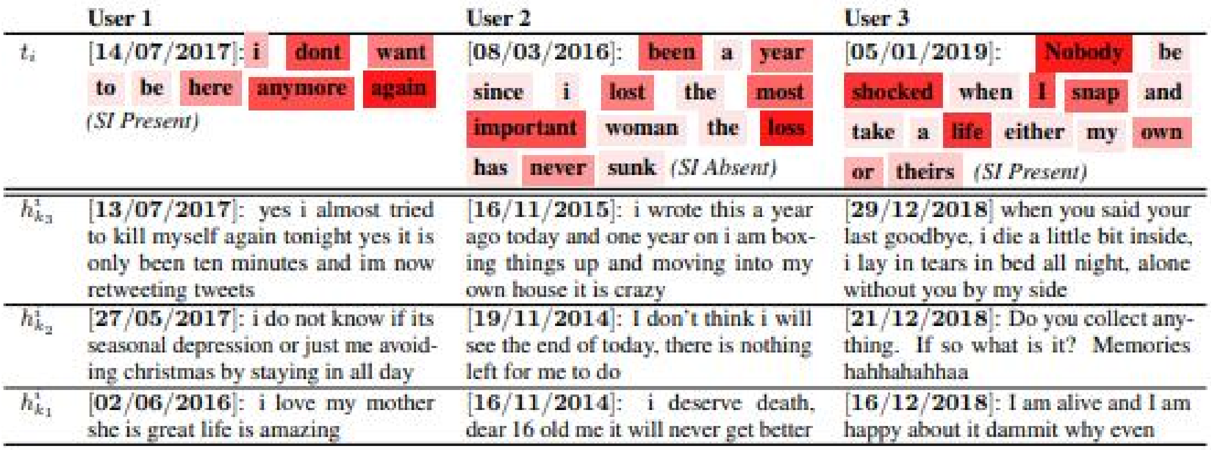


Figure 4: Confidence intervals for evaluation metrics of temporal variations over 10 different runs and data splits

5.2 Ablation Study



	User 1	User 2	User 3
t_i	[14/07/2017]: i dont want to be here anymore again *(SI Present)*	[08/03/2016]: been a year since i lost the most important woman the loss has never sunk *(SI Absent)*	[05/01/2019]: Nobody be shocked when I snap and take a life either my own or theirs *(SI Present)*
$h^i_{k_3}$	[13/07/2017]: yes i almost tried to kill myself again tonight yes it is only been ten minutes and im now retweeting tweets	[16/11/2015]: i wrote this a year ago today and one year on i am boxing things up and moving into my own house it is crazy	[29/12/2018] when you said your last goodbye, i die a little bit inside, i lay in tears in bed all night, alone without you by my side
$h^i_{k_2}$	[27/05/2017]: i do not know if its seasonal depression or just me avoiding christmas by staying in all day	[19/11/2014]: I don't think i will see the end of today, there is nothing left for me to do	[21/12/2018]: Do you collect anything. If so what is it? Memories hahhahahhaa
$h^i_{k_1}$	[02/06/2016]: i love my mother she is great life is amazing	[16/11/2014]: i deserve death, dear 16 old me it will never get better	[16/12/2018]: I am alive and I am happy about it dammit why even

Table 3: Tweet to be assessed (t_i) and historic tweets ($h^i_{k_1}$, $h^i_{k_2}$ and $h^i_{k_3}$ are chronologically ordered) of three users along with tweet timestamp information. We also show visualized self-attention (averaged over all 12 Sentence-BERT attention heads) per token. Darker intensity of the red color denotes higher attention weights.

We conduct ablation research (Table 2) with various configurations to test EHC and TTI. The model's effectiveness plummets when historical tweets aren't taken into account. We believe that adding historical tweets, even in a random order, gives more contextual cues about the user, resulting in better performance. "The Plutchik Transformer variant of Current + Sequential History outperforms its BERT equivalent , according to our findings. This can be ascribed to the Plutchik Transformer's ability to record a user's EHC. STATENet models the Current Tweet and EHC in real time, addressing the constraint of prior models that assumed equal time intervals between posts.This study confirms the ability of linguistic-only non-contextual models to detect suicidal intent. This is particularly intriguing because assessment can still be undertaken to some extent for individuals who have no prior history."

5.3 Temporal Analysis

According to the EHC, the language of the tweet should be evaluated in conjunction with historical context to better understand the user's emotional state over time. To investigate the significance of the order and temporal dependency of historical tweets, we first employ a nonsequential, bag of tweets-like version. "The Plutchik transformer-based encodings are fed into a Contextual CNN. We find that the bag of tweets technique outperforms the Contextual CNN baseline, most likely due to the transformer-based encoding as opposed to the baseline's static GloVe embeddings. Over 10 runs, the non-sequential technique significantly underperforms temporal versions. To investigate EHC and TTI further, we first feed previous tweets in sequential order (Sequential Model) to a conventional LSTM, and then we factor TTI in STATENet using T-LSTM. Figure 4 indicates that STATENet outperforms the sequential but non-time-aware counterpart and has the least fluctuation in performance throughout ten runs (p 0.005). We believe that the performance difference between the Sequential Model and STATENet is due to the temporal dependency of prior tweets on the elapsed time between succeeding tweets.

5.4 Qualitative Analysis

We examine certain scenarios where STATENet performs effectively to gain a more detailed understanding and to help interpretability. Through error analysis, we also highlight STATENet's shortcomings. Table 3 and Figure 5 present a qualitative analysis of three fascinating situations. We can observe that the tweet to be evaluated for User 1 lacks explicit suicide intent and may not be sufficient to estimate suicidal risk. However, as shown by Pluchtik's emotional strength in Figure 5a, the time model correctly labels tweets because it experiences an accumulation of sadness in previous tweets. If the user's current tweet doesn't make sense, the time model can collect more information by looking at the user's past activity. Time patterns are often different and posting frequencies are very different. The problem arises because these TTIs rely only on the sequence of previous tweets, not the actual time gap. For example, User 2's first tweet showed sadness and suicide intent, while User's latest past tweets (h2, k3) showed joy (Figure 5b). LSTM-based models accumulate melancholy, thereby suggesting a past of suicide. STATENet, on the other hand, can be learned from the various time-lapses and their relative importance in the context of suicidal ideation. However, I have found some situations where all models fail. User 3's recent tweets lack a powerful semantic marker of suicidal ideation. In addition, past tweets have no recognizable emotional tendencies (Figure 5c). Such scenarios show the difficulty of assessing the risk of suicide. Figure 5 further shows that the emotional intensity distribution of the plutic learned for the user is biased towards joy (positive) and depression (negative) (negative). Despite the fact that the very fine-grained emotional context captured by Plutchik Transformer improves the performance of STATENet, it does not improve the more general audio quality captured by BERT. Future research goals will further investigate the effects of emotional particle size.

VI. DISCUSSION

Ethical Considerations

The majority of the work presented in our debate raises serious ethical concerns. We discuss the trade-off between privacy and effectiveness . While data is critical to the success of models like STATENet, we must operate within the bounds of appropriate privacy standards to avoid coercion and intrusive treatment. To that purpose, we use publicly available Twitter data in a non-intrusive and purely observational approach . Although each user's informed consent was not sought since it could be construed as coercive, automatic de-identification of the dataset was carried out to limit the danger of identifying data being included in the raw data.To safeguard individual privacy, all tweets provided as samples in Figure 1 and Section 5.4 have been paraphrased using mild disguise method . Annotated user data is stored separately from raw user data on secure servers connected exclusively via anonymous IDs . STATENet's assessments, such as Samaritan's Radar, are sensitive and should only be shared with the right people . Our research makes no suicide-related diagnostic claims. We observe the social media posts from a distance and do not interfere with the user's experience in any way.

Limitations

We realise that suicidality research is subjective (Keilp et al., 2012), and that individual interpretations of the data given may differ. Because of the contextual nature of language, the data under study may be subject to demographic, annotator, and medium-specific biases (Hovy and Spruit, 2016). We recognise that suicide risk exists on a spectrum, and that using binary labels to categorise it could lead to erroneous risk assessments (Bryan and Rudd, 2006).

Practical Implications:

We propose a neural architecture for preliminary screening of at-risk individuals on social media with STATENet to aid in the prioritisation of clinical resources. Our work observes Twitter without interfering with the user experience in any way. STATENet should be part of a distributed human-in-the-loop system for finer risk interpretation (de Andrade et al., 2018). We deal with tweet level annotations rather than the more subjective and difficult to scale user-level annotations to emphasise STATENet's practical relevance. Although we focus on tweet-level prediction, STATENet can also be used to assess user-level suicide risk due to its dual text and historic modelling components.

VII. Conclusion

We present STATENet in response to the increasing usage of social media for displaying suicidal ideation as opposed to normal therapeutic treatment . STATENet models the time aware emotional context of users through historical tweets for more accurate suicide risk prediction on social media, based on psychology studies on assessing a user's temporal emotional spectrum. In the future, we intend to investigate the influence of varied amounts of historical background on a user.We demonstrate STATENet's utility as a prototype approach for detecting suicidality in tweets. We give a qualitative analysis to help you better grasp STATENet. We hope that this effort will contribute to a bigger human-in-the-loop infrastructure for assessing potentially problematic suicide-related social media messages. Our future objective is to use priority-based suicide risk assessment to rate tweets for suicidal risk rather than classify them. Furthermore, we would like to measure the impact of increasing degrees of granularity in learning emotional aspects from tweets on STATENet's performance in the future.

REFERENCES

[1]. Lyle Ungar and Muhammad Abdul-Mageed 2017. EmoNet is a network that uses gated recurrent neural networks to identify fine-grained emotions. Pages 718–728, in Proceedings of the Association for Computational Linguistics' 55th Annual Meeting (Volume 1: Long Papers), Vancouver, Canada. The Association for Computational Linguistics (ACL) is a non-profit organisation dedicated to the study of language

[2]. Dave Pawson, Dan Muriello, Lizzy Donahue, and Jennifer Guadagno. Norberto Nuno Gomes de Andrade, Dave Pawson, Dan Muriello, Lizzy Donahue, and Jennifer Guadagno. 2018. Suicide prevention on Facebook combines ethics and artificial intelligence. 669–684 in Philosophy & Technology, vol. 31, no. 4, pp. 669–684.

[3]. The suicide probability scale: Norms and Factor Structure, Courtney Bagge and Augustine Osman, 1998. 83(2):637–638 in Psychological Reports.

[4]. Cao Xiao, Xi Zhang, Fei Wang, Anil KJain, and Jiayu Zhou. 2017. Inci M Baytas, Cao Xiao, Xi Zhang, Fei Wang, Anil KJain, and Jiayu Zhou. Time-aware lstm networks for patient subtyping. Pages 65–74 in Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining, 23rd ACM SIGKDD international conference on knowledge discovery and data mining.

[5]. Mark Dredze, Adrian Benton, and Glen Coppersmith. 2017. Protocols for doing ethical research in the field of social media health. Valencia, Spain, pages 94–102 in Proceedings of the First ACL Workshop on Ethics in Natural Language Processing. The Association for Computational Linguistics (ACL) is a non-profit organisation dedicated to the study of language.

[6]. Carl Lee Hanson, Scott R Braithwaite, Christophe Giraud-Carrier, Josh West, Michael D Barnes, and Christophe Giraud-Carrier 2016. Using Twitter data to test machine learning algorithms against recognised measures of suicidality. 3(2):e21 in JMIR Mental Health.

[7]. Tineke Broer. 2020. Is technology the key to our future success? In relation to digitalized suicide prevention, I'm looking into the duty to report and subjectification processes. Information, vol. 11, no. 3, p. 170.

[8]. Amy Bruckman, "Studying the Amateur Artist: A Perspective on Disguising Data Collected in Human Subjects Research on the Internet," in Amy Bruckman, "Studying the Amateur Artist: A Perspective on Disguising Data Collected in Human Subjects Research on the Internet," Ethics and Information Technology, vol. 4, no. 3, pp. 217–231.

[9]. M David Rudd and Craig J Bryan. 2006. Suicide risk assessment has progressed. 185–200 in Journal of Clinical Psychology, Vol. 62, No. 2.

[10]. Cantor, Alan B., 1996. Cohen's kappa sample size calculations. 1(2):150 in Psychological Methods.

[11]. Munmun De Choudhury, Michael L. Birnbaum, Eric D. Caine, Vincent M. B. Silenzio, and Stevie Chancellor 2019. Inferring mental health statuses from social media: a taxonomy of ethical tensions. FAT* '19, New York, NY, USA, page 79–88 in Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* '19, page 79–88 in Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* '19, page 79–88 in Proceedings of the Conference on Fairness The Association for Computing Machinery is an organisation dedicated to the advancement of computing technology.

[12]. Craig Harman, Glen Coppersmith, and Mark Dredze 2014. Twitter is being used to quantify mental health signs. From linguistic signal to clinical reality: Proceedings of the workshop on computational linguistics and clinical psychology, pages 51–60.

DOI: 10.48175/568