Received August 4, 2018, accepted September 4, 2018, date of publication September 20, 2018, date of current version October 17, 2018.

*Digital Object Identifier 10.1109/ACCESS.2018.2871446*

# Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning

**SUDHA SUBRAMANI[1], HUA WANG[1], HUY QUAN VU[2], AND GANG LI[3], (Senior Member, IEEE)**

[1]Institute for Sustainable Industries and Liveable Cities, Victoria University, Melbourne, VIC 3011, Australia
[2]Centre for Tourism and Regional Opportunities, School of Engineering and Technology, Central Queensland University, Australia
[3]School of Information Technology, Deakin University, Geelong, VIC 3220, Australia

Corresponding author: Sudha Subramani (sudha.subramani1@live.vu.edu.au)

**ABSTRACT** Domestic violence (DV) is a cause of concern due to the threat it poses toward public health and human rights. There is a need for quick identification of the victims of this condition so that DV crisis service (DVCS) can offer necessary support in a timely manner. The availability of social media has allowed DV victims to share their stories and receive support from the community, which opens an opportunity for DVCS to actively approach and support DV victims. However, it is time consuming and inefficient to manually browse through a massive number of available posts. This paper adopts deep learning as an approach for automatic identification of DV victims in critical need. Empirical evidence on a ground truth data set has achieved an accuracy of up to 94%, which outperforms traditional machine-learning techniques. The analysis of informative features helps to identify important words which might indicate critical posts in the classification process. The experimental results are helpful to researchers and practitioners in developing techniques for identifying and supporting DV victims.

**INDEX TERMS** Domestic violence, deep learning, feature extraction, machine learning, social media.

## I. INTRODUCTION

Domestic Violence (DV) involves violent, abusive or intimidating behavior by a partner, or family member to control, dominate or cause fear to other family member(s) [1]. Due to the increasing attention to the high prevalence of DV and its serious consequences on victims' health issues [2], World Health Organization has developed some strategies to prevent and respond to DV [1] that includes: (1) media and advocacy campaigns to raise awareness and the knowledge base, to promote social and economic empowerment of women. (2) Early intervention services for at-risk families and increase access to comprehensive service response to survivors, called Domestic Violence Crisis Service (DVCS). However, victims have barriers to access the formal services [3] due to the victims' demographic factors such as race, beliefs and attitudes, income and fear of negative consequences of help-seeking [4]. This results in underutilization of the specialist support services over the past several years. Hence, the online support of DVCS is promoted for safe advertisement of DV resources, awareness promotion about the need for a compassionate to the victims, resource sharing and buddying between survivors, and non-professional mentoring [3].

In recent years, social-networking platforms (et. Facebook and Twitter) have exploded as a category of online discourse [5], which has shown their important role in the dissemination of supporting information and providing actionable situational knowledge during crises situations [6]. DVCS has been aware of the benefit that the social media platforms can bring to aid their decision and approach to support victims of DV. An issue with the posts shared on social media DVCS is that they are available at large scale, while not all posts are critically important. For examples, the posts $P_{1-4}$ in Table 1 are relevant to DV, but they are mainly for promoting awareness, providing advice, or expression of empathy. Such posts can be treated as ''uncritical'', as they do not describe a situation where a person is in danger or need immediate support. In contrast, the posts $P_{5-7}$ describe ''critical'' situations, where victims may need immediate support from DVCS. The accurate identification of such critical posts are crucially important for DVCS to direct their limited resources to support those in critical need. Manual browsing through a large amount of online posts is time consuming and inefficient to identify critical posts. As such, a tool that can filter the online posts relevant to DV and flag those critical posts is needed.

**TABLE 1.** Examples of DV posts and the corresponding intent labels.

| ID | DV Posts | Context | Label |
|---|---|---|---|
| $P_1$ | "To understand why people stay in abusive relationships, visit the link." | Awareness promotion | Uncritical |
| $P_2$ | "The code of silence is bulls**t" | Personal opinion | Uncritical |
| $P_3$ | "Morning greetings. Enjoy the day." | Greetings | Uncritical |
| $P_4$ | "Rest in Peace, Beautiful Angel." | Expressing empathy | Uncritical |
| $P_5$ | "I hope that this is okay I desperately need help. Please read my story and consider helping me I am desperate thank you very much." | Shared by victim | Critical |
| $P_6$ | "My best friend is fighting for her freedom. My dear friend was brutally beaten by her bf." | Shared by acquaintance | Critical |
| $P_7$ | "A woman who was shot dead by the father of her children on Anzac Day sacrificed her life to save her children." | Shared by media | Critical |

The use of online posts to support decision making in crisis has been investigated in the literature, such as during natural disasters of floods [7] and earthquakes [8]. By far, no attempt has been made to develop techniques for identifying personal crisis due to family disruption or disturbance in case of DV. The identification of critical posts relevant to DV is challenging task. The posts are in form of free text, which is unstructured data. How to represent the textual data for effective identification of critical posts is itself a critical task. It is also unknown, which features might provide an important clue to identify critical posts.

To the best of our knowledge, no prior work has either focused on critical post identification from social media or evaluated Deep Learning and machine learning techniques against different feature extraction methods for DV identification. Hence, this paper aims to provide support for DVCS by introducing an approach to automatically recognize critical posts on social media platforms. Firstly, a benchmark data set of online posts with labels, ''critical'' and ''uncritical'', is constructed. Textual features are then extracted from the unstructured textual data for further processing. Deep Learning, a modern and advanced machine learning architecture, is then applied to construct prediction models for automatic identification of critical posts. We treat the problem of critical post recognition as a binary text classification task, where a post is classified as ''critical'' or ''uncritical'' based on the textual content. We evaluate the performance of an introduced approach against various features for textual data and other traditional machine learning techniques. Analysis of informative features help to identify important words, which can distinguish between critical and uncritical posts. The experiment results and analysis are beneficial to researchers, who are interested in carrying out further research in DV based on online social media data.

The rest of this paper is organized as follows. Section II provides the background on DV, relevant textual features and machine learning techniques. Section III presents an approach for critical post recognition for DV. Section IV provides details on experiments to evaluate our approach with analysis of the results and discussion. Section V concludes the paper and envisages future research directions.

## II. BACKGROUND

### A. DOMESTIC VIOLENCE AND SOCIAL MEDIA

DV is one of the leading causes of injury among women and the problem is pervasive worldwide [9]. The victims usually suffer not only physical abuse, but also sexual, emotional, and verbal abuse. Both informal and formal support for victims in an abusive relationships can play a predominant role in improving safety, physical and mental health outcomes [10], [11]. DVCS were created to provide the combinations of the services, such as crisis hotline, counseling, advocacy, and emergency shelter to DV victims [12]. The victims found that disclosing their situation and receiving emotional support is helpful and consequently improves their mental health [13]. However, the support services are often not utilized effectively by the victims, because they need to actively seek such supports. Many victims have chosen not to contact DVCS groups and disclose their situation due to barriers in social-economic and religious background [14].

The availability of social media has challenged the notions of violence as private one [15]. Social media has been used as a tool to prevent violence by raising awareness through sharing knowledge and bringing stories to the public [16], [17]. The advantages of social media in information dissemination has been utilized for several applications such as crisis preparation, response and recovery [5] during disasters such as flooding [7], earthquake [18], tsunami [19]. However, the potential benefits of social media in identifying and providing instant support for DV victims, who are in critical need, have not been realized.

### B. AUTOMATIC TEXT CLASSIFICATION

Automatic labeling online posts as ''critical'' or ''uncritical'' is basically a text classification problem. Machine learning is a popular computational approach, which has been adopted into various applications for automatic text classification. Examples of such applications include sentiment analysis [20], topic modeling [21]–[25], opinion mining [26], predicting cyber bullying and online harassment [27], [28], and crisis response and management during natural disasters [7], [18], [19].

There are two main tasks in text classification, *textual feature extraction* and *label prediction*. The purpose of feature extraction is to extract significant vocabulary items from the textual data and represent them in suitable format, that is required by machine learning algorithms for the further analysis. Examples of popular textual features include word n-grams [29], bag-of-opinions [30], syntactic relations [31], sentiment lexicon features [32], Bag-of-Words (BoW) and TF-IDF [33], [34]. Some studies place extra efforts to extract additional information such as user profile features [35], semantic features [28], psycholinguistic features [36], and profanity word occurrences [37].

The label prediction task usually involves the training of learning models on the ground truth labeled data set and then applying the trained model to classify unlabeled new data. Some popularly used machine learning techniques are Naive Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), Decision Tree(DT), k-Nearest Neighbor (KNN) [38], [39]. They have been applied successfully in emergency situational awareness during natural disasters and crisis response [40], [41].

In the context of text classification, the performance of the aforementioned classifiers often rely on the quality of the features extracted from textual data. Popular features such as TF-IDF and BoW model are usually ineffective due to data sparsity and non-semantic representation [34]. They ignore the word ordering and lose the semantic coherence information. For example, the terms ''physical violence'', ''physical abuse'' and ''physical assault'' would be treated as separate features, although they share similar meaning in the context of DV. Traditional textual features are neither scalable nor generalize well across different domains. Alternative approaches, that are robust in capturing the semantic coherence of related words, is needed such as the case of Deep Learning discussed in the next section.

### C. APPLICATIONS OF DEEP LEARNING

Deep Learning is a relatively new branch of machine learning, whose advantage is the ability to automatically extract intermediate feature representations of raw textual data by building a hierarchical structure [42]. Deep Learning has been applied in various Natural Language Processing (NLP) applications, such as sentence modeling [43], text classification [44], and topic categorization [45]. Deep Learning also plays a tremendous role in various real-time applications using online social media data, which include the detection of cyber-bullying and online harassment [46], [47], disaster response and management [7], [48], online medical guidance and health prediction [49]–[52] and massive open online courses forums [53].

There are two primary Deep Learning architectures, Convolutional Neural Networks (CNNs) [44] and Recurrent Neural Networks (RNNs) [54]. Both models take input as the embedding of words in the text sequence, and generate the real-valued and continuous feature vector for the words. CNNs have been applied in sentence-level sentiment classification and question classification [43], [44], which show advanced performance over traditional machine learning techniques (SVM, MaxEnts). Similarly, RNNs are implemented to model the text sequence and achieved improved performance for multi-task learning [55]. The improved version of RNNs such as Long Short-Term Memory networks (LSTMs) [56], Gated Recurrent Units (GRUs) [57], and Bidirectional LSTMs (BLSTMs) [58] are widely used in NLP applications due to their long-range dependencies and storing historical information over time.

Deep learning techniques have applied to detect hateful vs non-hateful speech posts regarding racism and sexism [46], [47], informative vs non-informative posts in natural disaster response application [7], [48]. However, no attempt has been made to investigate the potential of Deep Learning in applications of DV context. This paper aims to address the challenges in domestic violence crisis identification by adopting Deep Learning techniques for the classification problem of ''critical'' and ''uncritical'' online posts.

## III. METHODOLOGY

This section presents our approach to critical post identification, which consists of five stages: 1) Data Extraction: 2) Data Labeling; 3) Feature Extraction; 4) Model Construction; 5) Performance Evaluation. Their details are described in the following subsections.

### A. DATA EXTRACTION

Our approach is designed for identifying online critical posts, thus the main source for data extraction is social media platforms. We use Facebook as an example to describe the data extraction process, since Facebook is one popular social media with around 2 billion users worldwide and ranked first among the top 15 social networking sites [59]. According to [60], emotional support in online context complements the emotional support received in off-line contexts. Facebook users are benefited by receiving support-based needs (emotional and informational support), due to the ease of sharing with the wide range of people through DVCS. Thus, we collected the posts from pages, that discuss the range of DV issues, through Facebook Graph API[1] with the search term of ''Domestic Violence and Domestic Abuse''. Considering the ethical concern, we collected from the open pages rather than closed and secret pages. The benefit of Facebook Graph API is that researcher can develop applications to detect new posts about DV in real-time, which can support DVCS in quickly identifying DV victims. Please be noted that only publicly available data on Facebook are extracted, which comply with

[1] https://developers.facebook.com/docs/graph-api

the privacy policy of Facebook. The identity of individuals included in the collected data set are not disclosed in this paper.

### B. DATA LABELING

Our next stage is to label the collected posts as "critical" or "uncritical" to construct a benchmark data for evaluating the proposed approach. The posts were manually examined by human scorers independently. If the content of a post was found to describe a critical situation or a situation where a victim indicates the need for help (eg. posts $P_{5-7}$ in Table 1), the post is labeled as "critical". Otherwise, it is labeled as "uncritical". Posts that contain only hyper-links are treated as irrelevant and discarded from further processing. There is some borderline posts, which may be perceived differently by different human scorers. For example, the post *"I've been there as DV victim. I conquered, I lost my child, I rose up, I walked away, I won that battle scar."* can be treated as "critical", because it implies that victim needs emotional support, as the child was lost. Other scorer may perceive this post as "uncritical", because he/she reasons that the victim has already battled the situation and that post conveys an aspiration message to stay strong rather than describing a critical situation. The limited context from these posts makes it difficult to interpret fully, and may causes discrepancy in human annotation. As such, only posts that were scored with the same label by all scorers are kept in a benchmark data set for evaluating the Deep Learning algorithms in the later stages.

### C. FEATURE EXTRACTION

The next stage is to extract features to mathematically describe the characteristics of the data set based on Word2Vec model. The vectors are learnt in such a way that words have similar meanings will have nearby representations in the vector space. Thus, this model overcomes the limitations of the traditional text feature representation techniques such as non-semantic representation and data sparsity. Word2Vec model is the more expressive text representation form, where the relationship between words are highly preserved.

More specifically, Word2Vec takes a textual data as input and each word in the vocabulary is projected as a low dimensional, real-valued and continuous vector, also known as word embedding [61] in the high dimensional space. Suppose an input post is denoted as $\mathtt{P} = \{x_1, x_2 ... x_n\}$, where $x_i$ is an individual word token in the post $\mathtt{P}$. We initially transform it into a feature space by mapping each word token $x_i \in \mathtt{P}$ to an index of embedding matrix $L$. Thus, the word embedding matrix is represented as $L_x \in \mathbb{R}^{D\times|V|}$, where $D$ is the dimensional word vector and $|V|$ is vocabulary size. $L$ can be randomly initialized from a uniform distribution or pre-trained from text corpus with embedding learning algorithms [62], [63]. In simple terms, the mathematical equation represents that the embedding matrix $L$ to be built for each index of the unique tokens in the vocabulary set. We used the latter strategy to make better use of semantic and grammatical associations of words, that is already pre-trained on large external corpus such as Google's Word2Vec [62] and Twitter's crawl of GloVe [63] for our intent classification task.

### D. MODEL CONSTRUCTION

This stage constructs the prediction model for critical post recognition. We adopt five Deep Learning models for our task, namely:

- CNNs: We adopt the CNNs architecture as described in [45] and used for our approach. Its first layer is called the embedding layer, which extracts the most informative n-grams features and stores the word embeddings for each word. Convolutional layer of CNNs has varying number of computation units, with each unit represents an n-gram (also known as region size) from the input text. Suppose the vocabulary includes $V$ = *'hope', 'I', 'was', 'abused', 'love'*, there is a post $P$ = *"I was abused"*. In case, the region size is set to 1. The post $P$ is represented as word embedding features [0 1 0 0 0 | 0 0 1 0 0 | 0 0 0 1 0], which is equivalent to Unigram approach. If the region size is set to 2, the post $P$ is represented as [0 1 1 0 0 | 0 0 1 1 0] for the pairs of words *'I was'* and *'was abused'*. This is equivalent to the Bigram approach. Given the variable sizes of the convolutional layer outputs, the pooling layer transforms the previous convolutional representation into a higher level of abstract view and produce fixed size output. Finally, the dense layer takes the combinations of produced feature vectors as input and makes prediction for corresponding post. When the consecutive words are given as input, CNNs can learn the embedding of text regions internally, which captures the semantic coherence information in the text.
- RNNs: The RNNs architectures described in [54] is adopted into our approach. RNNs handle a variable-length sequence input by having loops called recurrent hidden state, which captures the information from previous states. At each time stamp, it receives an input and updates the hidden state. The advantage of RNNs is that the hidden state integrates information over previous time stamps.
- LSTMs, GRUs and BLSTMs: LSTMs [56], GRUs [57] and BLSTMs [58] are improved version of RNNs. The core idea behind LSTMs are memory units, which maintain historical information over time, and the non-linear gating units regulating the information flow. GRUs are basically, an LSTMs with two gates, whereas LSTMs has three gates. GRUs merges the input and forget gates into one unit, named as "update gate". BLSTMs consists of two LSTMs, that integrates the long periods of contextual information from both forward and backward directions at a specific time frame. This enables the hidden state to store both the historical and future information. Thus LSTMs, GRUs and BLSTMs are the state-of-the-art semantic composition models for the

text classification task and learn long-term dependencies between the words in a sequence, without keeping redundant information.

The models are trained on feature sets extracted from the constructed data set, so that they can be used to predict the posts as critical or uncritical. In order to examine and compare the prediction performance of the models, we adopt several evaluation measures as presented in the next subsection.

### E. PERFORMANCE EVALUATION

The last stage is to evaluate the performance of the proposed approach to identifying critical posts in relevant to DV. We adopt Precision, Recall, F-Measure, and Accuracy as evaluation metrics of our classifier. These metrics have been used widely in various works to evaluate classifier performance [7], [47], [48], which is suitable for our problem of critical post identification. Since, only one data set is constructed for critical and uncritical posts identification, we adopt $k$-fold cross validation approach for the evaluation. The collected data set is randomly divided into $k$ partitions, where one partition is reserved as test set while the others are combined into a training set. The procedure is repeated $k$ times for different test sets, whose results are averaged to indicate an overall performance.

## IV. EXPERIMENT AND ANALYSIS

### A. EXPERIMENT DESIGN

We start with data collection, where online posts are extracted from Facebook user pages with the keywords ''domestic violence'' and ''domestic abuse'' using its Graph API. A large number of posts and comments were returned. The next step was to label the posts as ''critical'' and ''uncritical'' to construct a benchmark data set for evaluating the performance of the proposed approach. Since, the labeling process was done manually which is time consuming; we randomly selected a subset of the returned Facebook posts for benchmark data construction. We excluded the posts containing only hyperlinks or having less than three words, as they are unlikely to describe a DV situation. The remaining posts are labeled by three research students, under the supervision of a consultant psychiatrist dealing with DV and gender related issues in psychiatric illness, anxiety and depressive illness. The involvement of the domain expert is necessary to ensure the quality of the labeled data set. We used Kappa coefficient [64] to validate the inter-rater reliability of the human scorers. The achieved degree of agreement was reasonably high at 0.85. Only posts that have consistent labels by all scorers were included in the final data set. We arrive with 750 posts with label ''critical'' and 1310 posts with label ''uncritical''. This is a data set with considerable size, considering no previous work on identifying DV victims in critical needs from social media data was carried out.

Several experiments were performed to evaluate the performance of the introduced approach using Deep learning, namely:

(a) *Accuracy Evaluation:* We evaluate the performance of five Deep Learning models, CNNs, RNNs, LSTMs, GRUs and BLSTMs on the constructed benchmark data set. Additional experiments using traditional machine learning techniques, NB, SVM, RF, LR, and DT are also carried out for comparison purpose. We compared the performance of classifiers using various evaluation metrics such as Precision, Recall, F-Measure and Accuracy.

(b) *Hyper-parameters Evaluation:* The performance of Deep Learning models can be influenced by their associated hyper-parameters, such as pre-trained word embeddings, selection of optimizer, dropout rate, number of recurrent units, and number of LSTM memory units or convolution filters. Thus, we carried out experiments with various hyper-parameters to examine their influence to the classification performance. Since, training and tuning a neural network can be time consuming [65], [66], some of the most important parameters, based on the study by Reimers and Gurevych [67] were selected for evaluation.

(c) *Semantic Coherence Analysis:* We first examine some important words that may help to distinguish posts belonging to different classes. Then, we examine the semantic composition of the textual features generated by word embedding. The analysis demonstrates the ability to capture semantic meanings between words results in better prediction performance for the Deep Learning models.

In the above experiment, the features for Deep Learning model was extracted using pre-trained Word2Vec models. We used the pre-trained models on two different data sets, Google News [62] and general Twitter posts [63], to examine the robustness of the algorithms. Word2Vec features trained on Google news includes 300 dimensional vectors for a vocabulary of 3 million words and phrases that trained on roughly 100 billion words. Word2Vec features trained on Twitter posts includes 300 dimensional vectors for a vocabulary set of 2.2 million words and phrases that trained on roughly 840 billion words. Thus, for both feature sets, each word is represented by a vector of word embedding containing $D = 300$ dimensions. The first layer of the models is the embedding layer that computes the index mapping for all the words in the vocabulary, and then convert into dense vectors of fixed size by parsing the pre-trained embedding. The next layers contain 128 memory cells, which is popularity used in various applications [67]. The models were trained up to 50 epochs and implemented using Keras [68].

For the traditional class models, we used TF-IDF and BoW features, because they have been widely used in various text classification applications. We considered 3 different cases of preprocessing for these features, which include (a) stop-words removal only; (b) stemming only; (c) both stop-words removal and stemming, because, the traditional machine learning techniques may produce different results with different settings. Average numbers of words in a post

**TABLE 2.** Accuracy of machine learning classifiers with different word settings.

| Classifiers | Stop-words removal only | Stemming only | Both stop-words removal and stemming |
|---|---|---|---|
| NB+TF.IDF | 89.65 | 91.66 | 89.55 |
| SVM+TF.IDF | 91.42 | 92.20 | 91.61 |
| RF+TF.IDF | 86.85 | 89.24 | 87.44 |
| LR+TF.IDF | 88.08 | 90.74 | 89.26 |
| DT+TF.IDF | 86.27 | 88.58 | 86.07 |
| NB+BoW | 86.12 | 87.40 | 84.36 |
| SVM+BoW | 89.60 | 90.82 | 89.75 |
| RF+BoW | 80.19 | 84.97 | 82.44 |
| LR+BoW | 89.26 | 90.72 | 89.94 |
| DT+BoW | 86.61 | 89.83 | 85.53 |

**TABLE 3.** Evaluation metrics of classification models.

| Model | Precision | Recall | F-Measure | Accuracy |
|---|---|---|---|---|
| CNNs+Word2Vec | 92.40 | 92.40 | 92.40 | 92.30 |
| RNNs+Word2Vec | 76.90 | 77.20 | 77.05 | 77.09 |
| LSTMs+Word2Vec | 93.30 | 93.30 | 93.30 | 93.08 |
| GRUs+Word2Vec | 92.70 | 92.70 | 92.70 | 92.64 |
| BLSTMs+Word2Vec | 92.80 | 92.80 | 92.80 | 92.54 |
| CNNs+GloVe | 93.90 | 93.90 | 93.90 | 93.82 |
| RNNs+GloVe | 86.00 | 85.70 | 85.85 | 85.72 |
| LSTMs+GloVe | 94.10 | 94.10 | 94.10 | 94.02 |
| GRUs+GloVe | 94.50 | 94.50 | 94.50 | 94.26 |
| BLSTMs+GloVe | 94.40 | 94.40 | 94.40 | 94.16 |
| NB+TF-IDF | 91.80 | 89.60 | 90.70 | 91.66 |
| SVM+TF-IDF | 92.90 | 91.10 | 92.00 | 92.20 |
| RF+TF-IDF | 90.60 | 87.20 | 88.90 | 89.24 |
| LR+TF-IDF | 90.80 | 89.80 | 90.30 | 90.74 |
| DT+TF-IDF | 89.15 | 87.90 | 88.53 | 88.58 |
| NB+BoW | 87.90 | 86.50 | 87.20 | 87.40 |
| SVM+BoW | 91.80 | 89.80 | 90.80 | 90.82 |
| RF+BoW | 85.00 | 83.80 | 84.40 | 84.97 |
| LR+BoW | 90.80 | 89.80 | 90.30 | 90.72 |
| DT+BoW | 89.80 | 88.80 | 89.30 | 89.83 |

before pre-processing, after stop-words removal and after stemming are 155, 71 and 151 respectively. For Deep Learning models, the pre-processing is not carried out, because Deep Learning models process the sequence of words in the order they appear. Stop-words might hold valuable information that could be leveraged. Words are preserved in their original form without stemming, as they can represent different context (e.g., the words 'abusive', 'abuser', 'abuse' are context dependent). We use default parameters settings as in WEKA to evaluate the traditional classifiers. For the Deep Learning models, Nadam optimizer is used. Batch size was set to 32 posts, as the dataset size was moderate. Relu activation function and recurrent units set to 128 was used.

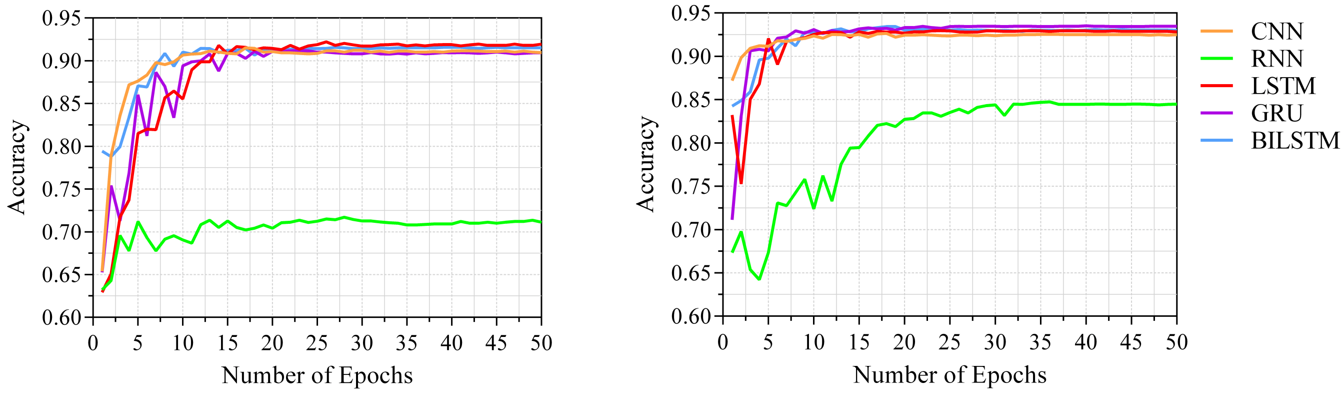
### B. PREDICTION PERFORMANCE EVALUATION

The machine learning algorithms were applied to the constructed data set. Since, a single data set were constructed for evaluation, we partitioned the data into training and test sets following 10-fold cross validation approach to measure the performance of the algorithms. We first evaluated the traditional classifiers with different word settings to identify the best setting for comparing with the deep learning classifiers. Due to space limitation, only overall accuracy is shown in Table 2. The results indicate that the traditional classifiers achieved the best performance with stemming only setting. In the context of DV crisis identification, some stop-words could be helpful to distinguish critical and non-critical posts. We used the stemming only setting for traditional classifiers to compare with other Deep Learning techniques. Evaluation metrics, precisions, recall, and accuracy were computed, as shown in Table 3. In general, Deep Learning models, except for RNNs, achieved better performance than traditional machine learning techniques, as indicated by higher evaluation metrics, that showed lower performance. RNNs achieved lower performance among all Deep Learning Models and probably due to the problem of vanishing gradients. Given a long sequence, information of initial sequence fades away as the new sequences are fed into the networks of RNNs. Nevertheless, such limitation of RNNs seems to be overcome by its later versions LSTMs, GRUs and BLSTMs. These models can capture long term dependencies efficiently, which is suitable for dealing with sequential textual data. The Deep Learning models appear to achieve better performance with GloVe than with Word2Vec. With Word2Vec embedding, LSTMs achieved best performance of 93.08%. With the GloVe embedding, LSTMs, GRUs and BLSTMs achieved relatively similar accuracy of more than 94%, which is better than all other algorithms.

### C. HYPER-PARAMETERS EVALUATION

We first evaluated the performance of the Deep Learning models with respect to training epochs. Ideally, the more training epochs would result in well-trained and stable models. However, Deep Learning models often take a long time to run. Setting high number of training epochs would result in significant and unnecessary costs. Figure 1 shows the accuracy of the Deep Learning models on the two feature sets (Word2Vec and GloVe) with respect to various training epochs. The models appear to converge faster on GloVe features set than on Word2Vec feature set. With Word2Vec embedding (Figure 1a), the accuracy of Deep Learning models fluctuated at the beginning and then become stable at their performance after 30 epochs on average. With GloVe embedding (Figure 1b), most models become stable after 20 to 23 epochs, except RNNs. Thus, Deep Learning models attained the optimal accuracy and consistency in learning rate, in minimal training epochs with respect to GLoVe embedding.

Next, we evaluated the performance of Deep Learning models with different hype-parameters settings, including optimizer, batch size, number of recurrent units, and




(a) (b)

**FIGURE 1. Accuracy of Deep Learning models at different epoches. (a) Word2Vec. (b) GloVe.**

activation function. We focused on evaluating GRUs and LSTMs, as they achieved highest performance as shown in the previous sections. The accuracies with 10-fold cross validation are shown in Table 4.

Among the optimizers, SGD is quite sensitive with the learning rate and it failed in many instances to converge. On the other hand, Nadam, RMSProp and Adam produced stable results of more than 91%. With respect to batch size, the mini-batch size of 1 produced poor accuracy. However, the algorithm achieved relatively good performance for batch size of 8 or more. Higher batch size value does not increase the performance of the models. Very big batch size of 256 seems to slightly decrease the conformance. The algorithm was also evaluated with different activation functions, including relu, sigmoid, softmax and softplus. The choice of activation function does not influence the performance of the algorithms as indicated by similar accuracies for both algorithms. Similarly, the number of recurrent units does not have any influence on their performance. Even though, the standard setting of 128 recurrent units appear to result in slightly better performance than other settings.

**TABLE 4. Accuracy of GRUs and LSTMs with different parameters settings.**

| Hyper-parameters | Variants | GRUs Acc | LSTMs Acc |
|---|---|---|---|
| Optimizer | Nadam | 94.26 | 93.08 |
| | RMSProp | 93.57 | 92.89 |
| | SGD | 71.88 | 79.20 |
| | Adam | 91.33 | 92.99 |
| Batch_Size | 1 | 55.72 | 54.55 |
| | 8 | 94.21 | 93.05 |
| | 32 | 94.26 | 93.08 |
| | 256 | 93.13 | 92.88 |
| Activation Function | relu | 94.26 | 93.08 |
| | softmax | 93.53 | 92.59 |
| | sigmoid | 94.06 | 92.64 |
| | softplus | 94.11 | 92.89 |
| No. of. Rec Units | 20 | 93.58 | 92.35 |
| | 40 | 93.87 | 92.84 |
| | 64 | 93.77 | 92.88 |
| | 128 | 94.26 | 93.08 |
| | 256 | 94.21 | 92.54 |

### *D. SEMANTIC COHERENCE ANALYSIS*

This section first examines the data sets to identify important words that helps distinguish critical from uncritical posts. We computed the support of each word in their corresponding class, which reflects their likelihood of occurrence. The difference in the supports of each word between two classes are computed, and the words having highest differences are reported in Table 5. Z-test with $p - value \leq 0.05$ were performed to verify statistical significant of the difference.

In the DV corpus, stop-words such as *linguistic dimensions (I, she, he, my, him)* and *time oriented tenses (was, is, were)* are more associated with critical posts. We may understand that, when the victims or survivors post about their abusive experience, they use more past tense *(was, were)*. *"She, he"* notions are often used to refer the abusive partners. *"I, me, my"* are often used by the victims to express their sufferings. Example posts are: *(1) I am a survivor of DV and rape. I really need help right now. I was in a relationship with a man for 8 years. (2) He was cheating on her. When she confronted him, he hurts her. He is just an evil and greedy man.*

Besides, the words *year*, *abuse*, *time"* are most likely to occurred in critical class than uncritical class with large differences. Those words usually appear in critical post, when victim made a post online to seek help from DVCS groups. The post content usually mentions about the victim was in an *abusive* relationship for the number of *years*, and when the last *time* the violence has happened. An example of such post is: *(He abused me for 5 years and each time he does something to scare me)*. Many posts mentioned about the context of the abusive incident, which is with the presence of their child/children, and sometimes, the child is also a victim. Thus, many critical posts contained the word *"child"*.

**TABLE 5.** Words have significant difference of occurrence likelihood between classes.

| Word | Critical | Uncritical | Difference | z-score | p-value |
|---|---|---|---|---|---|
| He | 0.62 | 0.02 | 0.60 | 30.662 | 0.000 |
| My | 0.74 | 0.18 | 0.56 | 25.018 | 0.000 |
| I | 0.75 | 0.27 | 0.48 | 20.950 | 0.000 |
| She | 0.43 | 0.02 | 0.41 | 23.592 | 0.000 |
| Him | 0.39 | 0.01 | 0.38 | 23.367 | 0.000 |
| We | 0.35 | 0.18 | 0.17 | 8.494 | 0.000 |
| Was | 0.67 | 0.04 | 0.63 | 30.654 | 0.000 |
| Is | 0.70 | 0.37 | 0.33 | 14.469 | 0.000 |
| Were | 0.24 | 0.01 | 0.23 | 16.881 | 0.000 |
| Will | 0.36 | 0.16 | 0.20 | 10.177 | 0.000 |
| Year | 0.51 | 0.05 | 0.46 | 24.294 | 0.000 |
| Abuse | 0.46 | 0.12 | 0.34 | 16.910 | 0.000 |
| Time | 0.40 | 0.09 | 0.31 | 16.998 | 0.000 |
| Life | 0.39 | 0.08 | 0.31 | 17.055 | 0.000 |
| Child | 0.35 | 0.07 | 0.29 | 16.441 | 0.000 |
| Friend | 0.35 | 0.06 | 0.29 | 16.654 | 0.000 |
| Husband | 0.23 | 0.01 | 0.23 | 17.173 | 0.000 |
| Night | 0.24 | 0.01 | 0.22 | 16.634 | 0.000 |
| Leave | 0.24 | 0.03 | 0.21 | 15.332 | 0.000 |
| Kill | 0.22 | 0.01 | 0.21 | 16.112 | 0.000 |
| Story | 0.29 | 0.09 | 0.20 | 11.664 | 0.000 |
| Love | 0.29 | 0.12 | 0.17 | 9.820 | 0.000 |
| Police | 0.17 | 0.01 | 0.16 | 13.673 | 0.000 |
| Woman | 0.17 | 0.03 | 0.14 | 10.892 | 0.000 |
| Survivor | 0.24 | 0.11 | 0.13 | 7.897 | 0.000 |
| Court | 0.14 | 0.01 | 0.13 | 12.033 | 0.000 |
| Control | 0.14 | 0.01 | 0.12 | 11.676 | 0.000 |
| Fear | 0.14 | 0.01 | 0.12 | 11.287 | 0.000 |
| Domestic | 0.38 | 0.26 | 0.12 | 5.631 | 0.000 |
| Victim | 0.19 | 0.08 | 0.11 | 7.340 | 0.000 |
| Violence | 0.38 | 0.28 | 0.10 | 4.759 | 0.000 |

The word *"husband"* appears more in critical posts as male partner violence is predominant. Many posts mentioned that *husband* is abusive. Similarly, the word *"friend"* is used often in critical posts. Some posts mentioned that the victims called friend for help when the violence occurred, or sometimes male friend is mentioned as the abusive person in the posts. The words *"night"* describes the time of abusive incident, which is usually represented at night times. The abuse is either physical or sexual assault.

The words *"leave, love, control, fear, kill"* often occur in posts that represent the emotion of the victims and explaining the reason they want to stay or leave the relationship. An example of such posts is: *(I live in fear every night, that he will kill me and finally I decided to leave him)*. When the victim seeks legal support or guidance in critical situation, the words *"police* and *court"* usually occur in the posts.

We noticed that the words *"domestic, violence, abuse"* have high support in both critical class and uncritical class. Because, these terms *domestic violence* and *domestic abuse* are commonly used in difference context in relation to

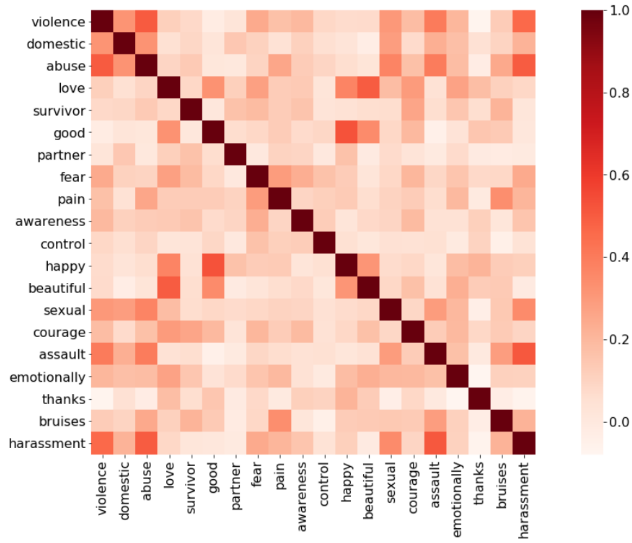



**FIGURE 2.** Correlation of sample words.

domestic violence. They are often used in uncritical posts to create awareness messages such as *(lets spread the word on domestic violence against women, please share this page with your friends.)*

Although, the words presented in Table 5, highlighted some difference between critical and uncritical posts, solely relying on term frequency may not be effective in automatic classification of the posts. Because, some words are often used to gather and share similar meaning such as *domestic*, *violence*, *abuse*. The classification model should account for their semantic relationships rather than treating them as separate words as in the traditional features of Bag of Words and TF-IDF.

Fortunately, the word embedding features used in Deep Learning could be able to address this issue. Note that, each word is represented by a vector feature of 300-dimensions that captures its semantic meaning. Words with similar meaning would have similar vector features. In other words, vector features of similar words are highly correlated with each other. As a demonstration, we visualize the correlation between the embedding vector features for some sample words using a heat map in Figure 2. We can see that, there is a strong correlation between the word *abuse* and words *violence*, *harassment* or *assault*. There is a low correlation between words having difference meanings, such as *love* versus *assault*, *bruises* or *pain*. The word embedding features could be able to account for such relationship, which explain the higher performance of Deep Learning models in comparison with the traditional models.

## V. CONCLUSION

In this paper, we presented an approach for critical post identification using Deep Learning. The contributions of this work are: (1) A benchmark dataset was constructed from Facebook posts made by DV victims, with labels for critical and uncritical posts; (2) We evaluated the performance of the various Deep Learning models in comparison with other traditional methods and with different parameter settings on DV critical post identification task. Due to the use of word embedding features, Deep Learning models (except for RNNs) achieved better performance than traditional models. The best setting for critical post identification of DV dataset is GloVe word embedding and GRUs model, with the Nadam optimizer and batch size of 32. Although, GRUs achieved the highest prediction rates in our experiments, other models CNNs, LSTMs and BLSTMs also achieved relatively high performance. Thus, Deep Learning models were demonstrated as promising to be adopted for developing practical solutions to identify the critical posts to support DV victims in critical needs. The analysis of the word occurrences also highlighted some context when and where DV take place. Future work, can consider classifying the posts into different DV context so that better detection of critical posts can be achieved and appropriate corresponding support can be provided to DV victims.

Despite the achieved results and findings, our work has several limitations. Namely, the data set used in the experiments was not at a big scale due to the labor-intensive job of manually labeling the posts. We currently recognized the critical post identification was mainly evaluated for posts from Facebook. Other social media platforms such as Twitter and Reddit can be considered in the future studies. Application for real-time critical post identification can be considered in the future so that instance support to DV victims can be provided. A novel algorithm for feature extraction or Deep Learning technique was not proposed in this paper. Because, our primary focus is to evaluate the existing state of the art features and Deep Learning algorithms on the new research problem of critical and uncritical post identification. Nevertheless, the results and findings are valuable in guiding the future works on DV crisis identification.

## REFERENCES

[1] World Health Organization. (2012) *Understanding and Addressing Violence Against Women: Sexual Violence*. [Online]. Available: http://apps.who.int/iris/bitstream/10665/77433/1/WHO_RHR_12.35_eng.pdf

[2] C. García-Moreno, C. Pallitto, K. Devries, H. Stöckl, C. Watts, and N. Abrahams, "Global and regional estimates of violence against women: Prevalence and health effects of intimate partner violence and non-partner sexual violence," World Health Org., Geneva, Switzerland, Tech. Rep., 2013.

[3] M. A. Evans and G. S. Feder, "Help-seeking amongst women survivors of domestic violence: A qualitative study of pathways towards formal and informal support," *Health Expectations*, vol. 19, no. 1, pp. 62–73, 2016.

[4] K. R. Henning and L. M. Klesges, "Utilization of counseling and supportive services by female victims of domestic abuse," *Violence Victims*, vol. 17, no. 5, pp. 623–636, 2002.

[5] H. Gao, G. Barbier, and R. Goolsby, "Harnessing the crowdsourcing power of social media for disaster relief," *IEEE Intell. Syst.*, vol. 26, no. 3, pp. 10–14, May/Jun. 2011.

[6] T. Flew, A. Bruns, J. Burgess, O. Ben-Harush, E. Potter, and J. Newton, "Support frameworks for the use of social media by emergency management organisations," Digi. Media Res. Centre, Centre Emergency Disaster Manage., Queensland Univ. Technol., Brisbane, QLD, Australia, Tech. Rep., 2015.

[7] C. Caragea, A. Silvescu, and A. H. Tapia, "Identifying informative messages in disaster events using convolutional neural networks," in *Proc. Int. Conf. Inf. Syst. Crisis Response Manage.*, 2016, pp. 1–6.

[8] D. T. Nguyen, S. Joty, M. Imran, H. Sajjad, and P. Mitra, "Applications of online deep learning for crisis response using social media information," in *Proc. 4th Int. Workshop Social Web Disaster Manage.*, Indianapolis, IN, USA, 2016, pp. 1–6.

[9] D. Houry *et al.*, "Does screening in the emergency department hurt or help victims of intimate partner violence?" *Ann. Emergency Med.*, vol. 51, no. 4, pp. 433–442, 2008.

[10] L. E. Rose and J. Campbell, "The role of social support and family relationships in women's responses to battering," *Health Care Women Int.*, vol. 21, no. 1, pp. 27–39, 2000.

[11] B. Liang, L. Goodman, P. Tummala-Narra, and S. Weintraub, "A theoretical framework for understanding help-seeking processes among survivors of intimate partner violence," *Amer. J. Community Psychol.*, vol. 36, nos. 1–2, pp. 71–84, 2005.

[12] L. Bennett, S. Riger, P. Schewe, A. Howard, and S. Wasco, "Effectiveness of hotline, advocacy, counseling, and shelter services for victims of domestic violence: A statewide evaluation," *J. Interpersonal Violence*, vol. 19, no. 7, pp. 815–829, 2004.

[13] K. M. Sylaska and K. M. Edwards, "Disclosure of intimate partner violence to informal social support network members: A review of the literature," *Trauma, Violence, Abuse*, vol. 15, no. 1, pp. 3–21, 2014.

[14] A. Kulwicki, B. Aswad, T. Carmona, and S. Ballout, "Barriers in the utilization of domestic violence services among arab immigrant women: Perceptions of professionals, service providers & community leaders," *J. Family Violence*, vol. 25, no. 8, pp. 727–735, 2010.

[15] C. Liou. *Using Social Media for the Prevention of Violence Against Women*. [Online]. Available: http://www.partners4prevention.org/sites/default/files/resources/socialmedia_final.pdf

[16] R. Clark, "'Hope in a hashtag': The discursive activism of #WhyIStayed," *Feminist Media Stud.*, vol. 16, no. 5, pp. 788–804, 2016.

[17] S. Subramani, H. Wang, M. R. Islam, A. Ulhaq, and M. O'Connor, "Child abuse and domestic abuse: Content and feature analysis from social media disclosures," in *Proc. Australas. Database Conf.* Cham, Switzerland: Springer, 2018, pp. 174–185.

[18] M. Imran, C. Castillo, J. Lucas, P. Meier, and S. Vieweg, "AIDR: Artificial intelligence for disaster response," in *Proc. 23rd Int. Conf. World Wide Web*, 2014, pp. 159–162.

[19] A. Acar and Y. Muraki, "Twitter for crisis communication: Lessons learned from Japan's tsunami disaster," *Int. J. Web Based Commun.*, vol. 7, no. 3, pp. 392–402, Jul. 2011.

[20] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in *Proc. LREc*, vol. 10, 2010, pp. 1320–1326.

[21] H. Jiang, R. Zhou, L. Zhang, H. Wang, and Y. Zhang, "A topic model based on Poisson decomposition," in *Proc. ACM Conf. Inf. Knowl. Manage.*, 2017, pp. 1489–1498.

[22] J. Huang, M. Peng, H. Wang, J. Cao, W. Gao, and X. Zhang, "A probabilistic method for emerging topic tracking in microblog stream," *World Wide Web*, vol. 20, no. 2, pp. 325–350, 2017.

[23] M. Peng, Q. Xie, H. Wang, Y. Zhang, and G. Tian, "Bayesian sparse topical coding," *IEEE Trans. Knowl. Data Eng.*, to be published, doi: 10.1109/TKDE.2018.2847707.

[24] M. Peng *et al.*, "Mining event-oriented topics in microblog stream with unsupervised multi-view hierarchical embedding," *ACM Trans. Knowl. Discovery Data*, vol. 12, no. 3, 2018, Art. no. 38.

[25] M. Peng *et al.*, "Neural sparse topical coding," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, vol. 1, 2018, pp. 2332–2340.

[26] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, nos. 1–2, pp. 1–135, 2008.

[27] M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong, "Improving cyberbullying detection with user context," in *Proc. Eur. Conf. Inf. Retr. (ECIR)*. Berlin, Germany: Springer, 2013, pp. 693–696.

[28] M. Dadvar, D. Trieschnigg, and F. de Jong, "Experts and machines against bullies: A hybrid approach to detect cyberbullies," in *Proc. Can. Conf. Artif. Intell.* Cham, Switzerland: Springer, 2014, pp. 275–281.

[29] S. Wang and C. D. Manning, "Baselines and bigrams: Simple, good sentiment and topic classification," in *Proc. 50th Annu. Meeting Assoc. Comput. Linguistics*, vol. 2, 2012, pp. 90–94.

[30] L. Qu, G. Ifrim, and G. Weikum, "The bag-of-opinions method for review rating prediction from sparse text patterns," in *Proc. 23rd Int. Conf. Comput. Linguistics*, 2010, pp. 913–921.

[31] R. Xia and C. Zong, "Exploring the use of word relation features for sentiment classification," in *Proc. 23rd Int. Conf. Comput. Linguistics, Posters*, 2010, pp. 1336–1344.
[32] S. Kiritchenko, X. Zhu, and S. M. Mohammad, "Sentiment analysis of short informal texts," *J. Artif. Intell. Res.*, vol. 50, pp. 723–762, Aug. 2014.
[33] G. Salton, A. Wong, and C. S. Yang, "A vector space model for automatic indexing," *Commun. ACM*, vol. 18, no. 11, pp. 613–620, 1975.
[34] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Inf. Process. Manage.*, vol. 24, no. 5, pp. 513–523, 1988.
[35] M. Dadvar, F. de Jong, R. Ordelman, and R. Trieschnigg, "Improved cyberbullying detection using gender information," in *Proc. 12th Dutch-Belgian Inf. Retr. Workshop*, 2012, pp. 23–25.
[36] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of LIWC2015," Univ. Texas at Austin, Austin, TX, USA, Tech. Rep., 2015.
[37] Y. Chen, Y. Zhou, S. Zhu, and H. Xu, "Detecting offensive language in social media to protect adolescent online safety," in *Proc. Int. Conf. Privacy, Secur., Risk Trust (PASSAT), Int. Conf. Social Comput. (SocialCom)*, Sep. 2012, pp. 71–80.
[38] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014.
[39] J. Khairnar and M. Kinikar, "Machine learning algorithms for opinion mining and sentiment classification," *Int. J. Sci. Res. Pub.*, vol. 3, no. 6, pp. 724–796, 2013.
[40] S. Verma *et al.*, "Natural language processing to the rescue? Extracting 'situational awareness' tweets during mass emergency," in *Proc. ICWSM*, 2011, pp. 385–392.
[41] M. A. Cameron, R. Power, B. Robinson, and J. Yin, "Emergency situation awareness from twitter for crisis management," in *Proc. 21st Int. Conf. World Wide Web*, 2012, pp. 695–698.
[42] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[43] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, 2014, pp. 655–665.
[44] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Doha, Qatar, 2014, pp. 1746–1751.
[45] R. Johnson and T. Zhang, "Effective use of word order for text categorization with convolutional neural networks," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, Denver, CO, USA, 2014, pp. 103–112.
[46] B. Gambäck and U. K. Sikdar, "Using convolutional neural networks to classify hate-speech," in *Proc. 1st Workshop Abusive Lang. Online*, 2017, pp. 85–90.
[47] P. Badjatiya, S. Gupta, M. Gupta, and V. Varma, "Deep learning for hate speech detection in tweets," in *Proc. 26th Int. Conf. World Wide Web Companion*, Perth, WA, Australia, 2017, pp. 759–760.
[48] D. T. Nguyen, K. A. Al Mannai, S. R. Joty, H. Sajjad, M. Imran, and P. Mitra, "Robust classification of crisis-related data on social networks using convolutional neural networks," in *Proc. ICWSM*, 2017, pp. 632–635.
[49] C. Che, C. Xiao, J. Liang, B. Jin, J. Zho, and F. Wang, "An RNN architecture with dynamic temporal matching for personalized predictions of parkinson's disease," in *Proc. SIAM Int. Conf. Data Mining*. Philadelphia, PA, USA: SIAM, 2017, pp. 198–206.
[50] C. Yao *et al.*, "A convolutional neural network model for online medical guidance," *IEEE Access*, vol. 4, pp. 4094–4103, 2016.
[51] L. Guo, B. Jin, R. Yu, C. Yao, C. Sun, and D. Huang, "Multi-label classification methods for green computing and application for mobile medical recommendations," *IEEE Access*, vol. 4, pp. 3201–3209, 2016.
[52] B. Jin, C. Che, Z. Liu, S. Zhang, X. Yin, and X. Wei, "Predicting the risk of heart failure with EHR sequential data modeling," *IEEE Access*, vol. 6, pp. 9256–9261, 2018.
[53] X. Wei, H. Lin, L. Yang, and Y. Yu, "A convolution-LSTM-based deep neural network for cross-domain MOOC forum post classification," *Information*, vol. 8, no. 3, p. 92, 2017.
[54] I. Sutskever, J. Martens, and G. E. Hinton, "Generating text with recurrent neural networks," in *Proc. 28th Int. Conf. Mach. Learn. (ICML)*, Stockholm, Sweden, 2011, pp. 1017–1024.
[55] P. Liu, X. Qiu, and X. Huang, "Recurrent neural network for text classification with multi-task learning," in *Proc. 25th Int. Joint Conf. Artif. Intell.*, New York, NY, USA, 2016, pp. 2873–2879.
[56] A. Graves. (2013). "Generating sequences with recurrent neural networks." [Online]. Available: https://arxiv.org/abs/1308.0850
[57] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," in *Proc. 8th Workshop Syntax, Semantics Struct. Stat. Transl.*, Doha, Qatar, 2014, pp. 103–111.
[58] A. Graves, N. Jaitly, and A.-R. Mohamed, "Hybrid speech recognition with deep bidirectional LSTM," in *Proc. IEEE Workshop Autom. Speech Recognit. Understand. (ASRU)*, Dec. 2013, pp. 273–278.
[59] P. Kallas. *Top 15 Most Popular Social Networking Sites and Apps*. [Online]. Available: https://www.dreamgrow.com/top-15-most-popular-social-networking-sites/
[60] S. Trepte, T. Dienlin, and L. Reinecke, "Influence of social support received in online and offline contexts on satisfaction with social support and satisfaction with life: A longitudinal study," *Media Psychol.*, vol. 18, no. 1, pp. 74–105, 2015.
[61] T. Mikolov, W.-T. Yih, and G. Zweig, "Linguistic regularities in continuous SpaceWord representations," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2013, pp. 746–751.
[62] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *Proc. Int. Conf. Learn. Represent. Workshop Track*, Phoenix, AZ, USA, 2013, pp. 1301–3781.
[63] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors forword representation," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Doha, Qatar, 2014, pp. 1532–1543.
[64] M. L. McHugh, "Interrater reliability: The kappa statistic," *Biochemia Med.*, vol. 22, no. 3, pp. 276–282, 2012.
[65] F. Hutter, H. Hoos, and K. Leyton-Brown, "An efficient approach for assessing hyperparameter importance," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 754–762.
[66] F. Hutter, J. Lücke, and L. Schmidt-Thieme, "Beyond manual tuning of hyperparameters," *KI-Künstliche Intell.*, vol. 29, no. 4, pp. 329–337, 2015.
[67] N. Reimers and I. Gurevych. (2017). "Optimal hyperparameters for deep LSTM-networks for sequence labeling tasks." [Online]. Available: https://arxiv.org/abs/1707.06799
[68] F. Chollet *et al.* (2015). *Keras*. [Online]. Available: https://github.com/fchollet/keras

**SUDHA SUBRAMANI** received the B.E. and M.E. degrees in computer science from Anna University, India, in 2010 and 2012, respectively. She is currently pursuing the Ph.D. degree with Victoria University. From December 2016 to March 2017, she did summer internship with Robert Bosch, Japan, and was involved in a project to develop disease assistance tool for disease management in tomato green houses. She is currently a Research Ambassador with Victoria University. Her research interests are social media data analytics, data mining, machine learning, deep learning, and text mining. She was a recipient of the Victoria University International Postgraduate Research Scholarship. She received the Gold Medal from the Sona College of Technology, for her academic excellence in the master's degree.

**HUA WANG** received the Ph.D. degree from the University of Southern Queensland, Australia. He was a Professor with the University of Southern Queensland. He joined Victoria University, where he is currently a Full Time Professor. He has more than 10 years teaching and working experience in applied informatics at both enterprise and university. He has expertise in electronic commerce, business process modeling, and enterprise architecture. Six Ph.D. students have graduated under his principal supervision. He received three Australian Research Council Discovery grants, as a Chief Investigator, since 2006. He has authored 200 peer-reviewed scholar papers.

**HUY QUAN VU** received the Ph.D. degree from the School of Information Technology, Deakin University, Australia. He was a Research Fellow with the Center for Applied Informatics, Victoria University. He is currently a Lecturer in ICT with Central Queensland University. His research interests include data mining and social media analytics applications in tourism management.

**GANG LI** (M'11–SM'13) received the Ph.D. degree in computer science from Deakin University, Australia, in 2005. He is currently an Associate Professor with the School of Information Technology, Deakin University. His research interests are in the areas of data privacy, data science, machine learning, and business intelligence.