IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 7, 2019, accepted May 31, 2019, date of publication June 10, 2019, date of current version June 27, 2019.
*Digital Object Identifier 10.1109/ACCESS.2019.2921938*

# Medical Social Media Text Classification Integrating Consumer Health Terminology

**KAN LIU AND LU CHEN**
School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073, China
Corresponding author: Kan Liu (liukan@znufe.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 71573196.

**ABSTRACT** In recent years, advances in technologies, such as machine learning, natural language processing, and automated data processing, have offered potential biomedical and public health applications that use massive data sources, e.g., social media. However, current methods are underutilized for features including consumer health terminology in social media texts. In this paper, we proposed a medical social media text classification (MSMTC) algorithm that integrates consumer health terminology. Classification of text from social media on medical subjects is divided into two sub-tasks: consumer health terminology extraction and text classification. First, text characteristics based on the double channel structure are used for training, and consumer health terminology is subsequently extracted-based using an adversarial network. Then, text classification is implemented based on the extracted consumer health terminology and double channel subtraction method. This paper takes datasets containing patient descriptions from social media as an example. The experimental results show that the algorithm outperforms single channel methods or baseline models, including Convolutional Neural Networks, Long Short-Term Memory Networks, Bi-directional Long Short-Term Memory Networks, Naive Bayesian Model, and Extreme Gradient Boosting.

**INDEX TERMS** Adversarial network, double channel structure, medical social media text classification (MSMTC), terminology.

## I. INTRODUCTION

In recent years, social media has become an inexhaustible source of information on all kinds of topics. According to the latest Pew Research Report, 26% of users have discussed health information and 42% of those users have discussed their personal medical conditions [1]. Moreover, nearly one-third of users changed their behavior based on health information. Simultaneously, with the emergence of some medical social media sites (e.g., www.DXY.com, www.haodf.com, and www.chunyuyisheng.com), some individuals discuss and share health information, as well as make timely prejudgments regarding their own health abnormalities. This has resulted in many medical social media texts, such as user descriptions, inquiries, or comments on medical social media. These texts serve as useful corpora for learning and can be used to discover disease characteristics, locate disease populations, provide early warning of disease transmission, and provide patients with precise treatment. However, because most social media users are not medical professionals, the texts described are composed of informal phrases, and therefore difficulties exist in text mining of medical social media posts.

Text in medical social media mainly contains three types of words. The first category is normalized medical terminology, which refers to some concept or ontology in the medical field, e.g., ''insomnia'' or ''bloating''. The second category is defined as user health terminology by Marshall [2], and refers to the language preferred by non-professionals. This language often includes colloquial expressions of professional terms that express certain health information but lack standardization, such as ''can't sleep'' and ''stomach pain''. The third category is defined as noise words, such as ''because'', ''so'', and ''doctor''. Such words often lack health information and may cause some interference.

Existing medical social media text mining methods typically use normalized medical terminology. The second type of terminology (i.e., user health terminology) is often

The associate editor coordinating the review of this manuscript and approving it for publication was Shirui Pan.

neglected because it is not formal. However, the second category of terms occupies a major part of medical social media texts. Therefore, this non-canonical text must be considered when fully mining medical social media text information [3]. As such, a text classification algorithm that integrates consumer health terminology and effectively improves the performance of medical social media text classification is proposed in this paper. The algorithm outperforms all the baseline models with accuracy reaching 87.65%. This medical social media text classification (MSMTC) algorithm identifies the category to which the disease belongs based on a user's description of the condition, which primarily includes informal terms. The algorithm also assists in the use of medical social media for providing medical care. The main contributions of this paper are:

1) Design a framework with dual-channel and adversarial mechanisms to automatically extract consumer health terminology and use the dual-channel subtraction to perform classification.

2) Compare the performances of single channel method and dual-channel method, and demonstrate that the dual-channel structure can provide a more accurate classification result.

3) Demonstrate that the extraction of consumer health terminology plays an important role in medical social media text classification and that this method can also deal with network language and non-standard language in other fields.

## II. RELATED RESEARCH

Although research on medical social media texts is still in its infancy, the research interest in this field has increased rapidly in recent years, especially given the development of new technologies, such as machine learning and natural language processing (NLP). These technologies have led to the emergence of new research-based medical social media text mining, including pharmacovigilance [4] , [5], public health monitoring [6], tracking of disease outbreaks [7], [8], and adverse drug reaction (ADR) mining [9], [10]. These new studies involve mining and analysis of different health content. From public health monitoring to new drug research, as well as disease tracking, medical practice has been transformed from resource allocation to the diagnosis of complex diseases. Most research is based on classification techniques that involve machine learning or deep learning that attempt to solve medical text classification problems using supervised classification techniques [7]. In fact, a large amount of medical terminology, noise and other complex languages in medical social media texts make accurate classification difficult. Among them, the treatment of normative and informal terms in the medical field is currently an active area of research. In line with early biomedical NLP approaches, information mining specific to social media mostly uses lexicon-based techniques that rely on a dictionary [11]–[13], such as the terminology and concept combination of MetaMap [14]. Subsequently, researchers directly normalized the lexicon of social media texts. For example, Han *et al.* [15] proposed a dictionary-based approach that uses context information to generate possible variants and normalization pairs. These are then ranked based on string similarity and used to populate the dictionary. However, this method is prone to high-dimensional sparsity problems. Hassan and Menezes [16] and Yajun *et al.* [17] proposed using random walks in a contextual similarity bipartite graph to obtain the global optimal normalization candidate list. The final normalized dictionary is obtained based on the literal similarity between informal and normative words. However, in medical social media texts, there may be no overlap between consumer health terminology and medical terminology, and therefore the final standardized dictionary cannot be obtained based on literal similarity between informal and normative words. Recently, some researchers became specialized in the standardization of user health terms. As an example, Tutubalina *et al.* [18] proposed using deep recurrent neural networks to normalize consumer health terminology and normative medical concepts. Han *et al.* [19] proposed a hierarchical combined Long Short-Term Memory (LSTM) model to normalize Adverse Drug Reaction (ADR) mentions. Niu *et al* [20] proposed multi-task character-level attentional networks to study the normalization of medical concepts. These studies are based on the unified medical language system (UMLS) [21], which was developed by the National Library of Medicine. UMLS facilitates mapping from a disease mention to a concept within a controlled vocabulary, which can assist deep neural networks in mapping informal phrases to their corresponding concepts in medical ontology. However, this method must mark the user's health term segments according to medical knowledge. the requirements. Labeling is very time-consuming and labor-intensive, and the accuracy is difficult to grasp. Other studies focused on using machine translation models [22], [23]. Intuitively, the machine translation model can translate non-standard language into standardized language; however, training the model requires corpus alignment to produce good results, and there are few aligned corpora in many practical situations.

In summary, methods based on dictionary lookup, string matching, concept mapping, and machine translation do not sufficiently normalize consumer health terminology. The main reason for this includes the lack of similarity between consumer health terminology and medical terminology as compared to semantic similarity; thus, determining mapping between them requires greater consideration of semantics and contextual relationships. In fact, there is a lack of relevant training corpora. Therefore, we propose using an adversarial network [24], [25] for extracting consumer health terminology combined with an attention mechanism [26] to solve the problem of long-distance dependence in text [27]. Two channels containing different terms are designed to classify medical social media texts using word embedding subtraction.

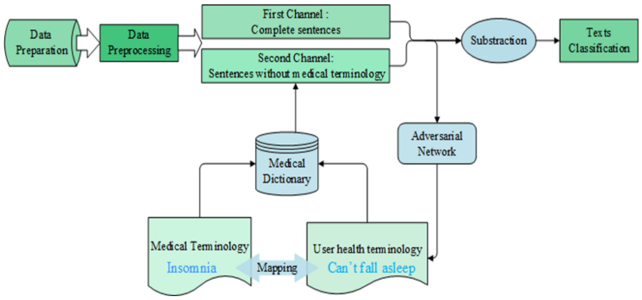


**FIGURE 1.** **Architecture for medical social media text classification.**

## III. MSMTC MODEL

### A. FRAMEWORK

A flowchart of our method is shown in Figure 1. The key step is to use an adversarial network to extract consumer health terminology together with medical terminology to form a dictionary. Words contained in the dictionary are then hidden in the original sentence to form a noise data channel that does not contain medical information. Finally, the outputs from the complete data channel and noise data channel are input to the classification task after subtraction.

Figure 1 displays the process in detail, which consists of the following five components:

(1) Data preparation: Text is mined from medical social media posts. Medical terminology vocabulary was downloaded from Sogou [28] to initialize the medical term dictionary.

(2) Data preprocessing: After obtaining medical social media text, word segmentation was conducted, stop words were removed, and then the whole dataset was divided into training set, test set, and validation set.

(3) Dual channel feature learning: As shown in Figure 1, the first channel refers to inputting a complete sentence, which can be learned directly. The second channel refers to inputting a sentence that hides medical terminology, i.e., medical terminology in the sentence is replaced with ''<pad>''. Both channels will receive and output the coding of the corresponding sentence.

(4) Word extraction with an adversarial network: The results obtained from the previous two channels are sequentially input into the source discriminating task for discrimination. Meanwhile, the output from the first channel is used to classify the medical social media texts. Here, the adversarial network is used for joint training to provide automatic extraction of consumer health terminology and place it into the medical term dictionary.

(5) Text classification based on dual channel subtraction: User health terminology has been added to the medical terminology dictionary. Words contained in the dictionary are then hidden in the original sentence to form the noise data channel. Finally, the subtraction of the outputs from the complete data channel and noise data channel are input to the classification task to obtain the text classification result.

### B. MODEL DETAILS

Our text classification model integrates the consumer health terminology and consists of three parts (Figure 2): a feature learning module (left dual channel structure), an adversarial network for word extraction (module in the bottom right), and text classification module based on dual channel subtraction (module in the upper right). The word extraction module and text classification module have the same classifier structure, which consists of a dense layer and a softmax layer.

#### 1) FEATURE LEARNING MODULE

The purpose of the feature learning module is to obtain a representation of the text. Here, each word is initially mapped into its embedding vector, and positional encoding and attention are added to weight words in each sentence; this yields the final sentence vector. The feature learning process executes the same operation in both channels (complete sentence channel and hidden word channel).

##### *a: WORD EMBEDDING LAYER*

Each word in the sentence can be represented by a vector. In Figure 2, the first channel learns the embedding vector for complete sentences, the second channel first hides the medical terminology in the sentence (such as ''eyes'' and ''bloodshots''), and then learns word embedding. In Figure 2, $x_c$ is the input of the first channel, $x_u$ is the input of the second channel, and then $e_c \in \mathrm{R}^{l\times d}$ and $e_u \in R^{l\times d}$ are obtained respectively after word embedding, where $l$ is the length of the sentence and $d$ is the dimension of the word vector.

##### *b: POSITIONAL ENCODING*

The sequential structure of the words is an important feature of text data. In order for the model to make use of the sequence, we must inject some information regarding the relative or absolute position of the tokens in the sequence. To this end, the model designs the positional encoding structure after the word embedding layer. In this paper, a trigonometric function is used to encode the positional relationship between words [26]. This encoding method uses the absolute position of a word in a sentence as a variable in a trigonometric function to calculate $\mathrm{PE}_{pos}$, then $\mathrm{PE}_{pos}$ is added to the word's embedding vector. Its purpose is to incorporate relative distance information between words into the word vector. According to the formula proposed by Vaswani *et al.* [26], the formula for updating the word vector is as follows:

$$\mathrm{PE}_{(pos,2i)} = \sin(pos/10000^{2\mathrm{i}/\mathrm{d}_{model}}) \tag{1}$$

$$\mathrm{PE}_{(pos,2i+1)} = \cos(pos/10000^{2\mathrm{i}/\mathrm{d}_{model}}) \tag{2}$$

$$e_{pos} = e_{pos} + \mathrm{PE}_{pos} \tag{3}$$

where $\mathrm{d}_{model}$ is the dimension of the word embedding, $i$ is the $i$th dimension of the word vector, and $pos$ is the position of the word in the sentence. There are two benefits of using positional encoding. First, a trigonometric function has definite periodicity, and the value of the dependent variable

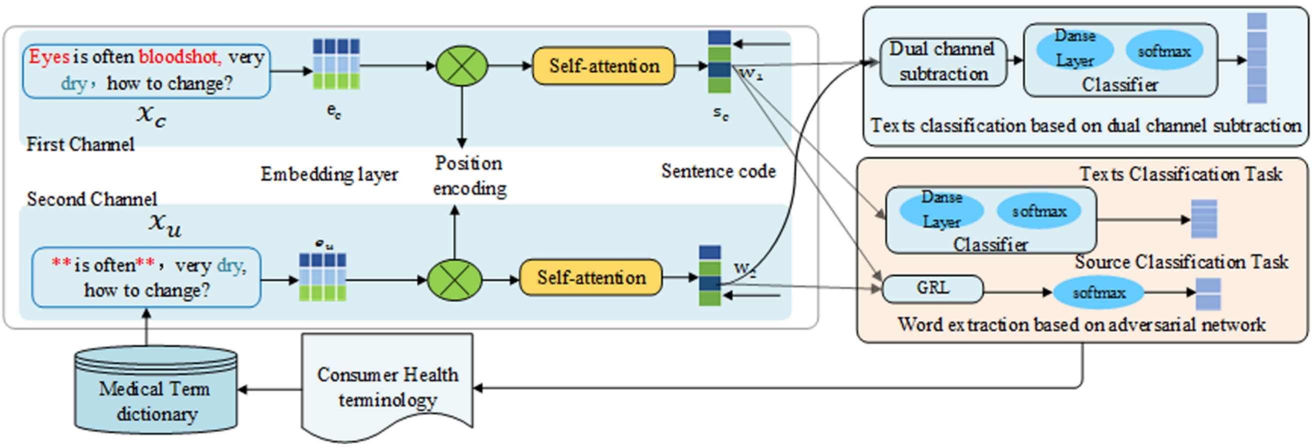


**FIGURE 2. Medical social media text classification frame.**

will appear again. The value will not be limited by the length of the sentence and it may allow the model to extrapolate to sequence lengths that are longer than those encountered during training. Second, the value range of the trigonometric function lies in [−1], [1], and the value can be directly supplied to the word vector.

### c: ATTENTION MECHANISMS

An internal attention mechanism is used to further capture long dependencies within the medical social media texts. Long distance dependence is very common in medical social media texts, e.g., "Hello doctor, I may have inflammation, because I have diarrhea every morning." It is not difficult to find that there is a long-distance dependence between "inflammation" at the beginning of the sentence and "diarrhea" at the end of the sentence. One can infer from "diarrhea" that the term "inflammation" is likely to refer to "enteritis". In addition to capturing the complex interactions and internal dependency between such words, the contribution of context words to sentence semantic information is not uniform for a particular task. Therefore, a self-attention mechanism was added to the model to help focus on words in sentences that have significant impacts on text categorization. Other research in the literature [29] shows that if the word vector of the $i$th word is $e_i$ and the word vector of the $j$th word is $e_j$, the formula for calculating the attention score is as follows:

$$e_{ij} = a\left(e_i, e_j\right) \tag{4}$$

where $e_{ij}$ is the attention score between the $i$th and $j$th words, which indicates the matching quality between $e_i$ and $e_j$. Then, the choice of alignment function a(∗) is based on the specific situation. The alignment function is a fully connected network, in which each word in the sentence and each node of the self-attention layer are fully connected; the interaction distance between any two words is 1. After obtaining the attention score, the entire sentence sequence is encoded as follows:

$$\alpha_{ij} = \sum\nolimits_{i,j=0}^{l-1} softmax(\tanh(w^T\left[e_i; e_j\right] + b)) \tag{5}$$

$$s = \sum\nolimits_{i,j} \alpha_{ij} e_i \tag{6}$$

In equation (5), $l$ is the length of the sentence, $w$ and $b$ are the weights and bias of the fully connected network, respectively, tanh is the activation function, and $\alpha_{ij}$ is the attention score. In equation (6), $s$ is the sentence encoding. The two channels will output $s_c$ and $s_u$ in this layer, respectively.

### 2) WORD EXTRACTION MODULE BASED ON AN ADVERSARIAL NETWORK

The two sentence encodings, $s_c$ and $s_u$, obtained from the feature learning module are input into the word extraction module for extracting consumer health terminology. Here, we draw on the research ideas of Li *et al.* [30], [31] and propose the use of an adversarial network to automatically extract consumer health terminology.

### a: ADVERSARIAL NETWORKS

The use of an adversarial network refers to adversarial learning of a dual channel structure for learning features in text, as well as a source discriminant structure that is used to discriminate from which channel the data originated. The details of the process are as follows. First, $s_c$ is used for text classification and must provide accurate classification results in order to increase the weight of medical terminology and consumer health terminology in $s_c$. Then, $s_c$ and $s_u$ are input into the source discriminant structure to discriminate from which channel the data originated. This process will produce a judgment error that will pass through the gradient reversal layer during backpropagation. This layer uses the reversal gradient for weight updates in the dual channel. This adversarial training process is conducted repeatedly to increase the weight of consumer health terminology in $s_u$. Finally, the model will output words with the highest weight in $x_u$. The objective function in the model is:

$$L^{total} = L^{c} + L^{datasource} + \rho L^{reg} \tag{7}$$

where $\rho$ is a regularization parameter that balances the regularization term $L^{reg}$ and other related terms. $L^{reg}$ is responsible of preventing overfitting by imposing L1 regularization on the parameters for the word extraction module and text classification module. Here, L1 regularization is used to help to generate a sparse weight matrix, which can be used to select word features in each sentence. $L^{c}$ and $L^{datasource}$ represent the loss of text classification and the loss of source adversarial loss, respectively, which are both cross-entropy costs.

#### b: GRADIENT REVERSAL LAYER

In the source discriminant task, we mentioned the gradient reversal layer. The model adds a gradient reversal layer (GRL) before the softmax layer to achieve gradient reversal during backpropagation. Here, during forward propagation, the gradient reversal layer acts as an identification function [32] $Q_\lambda(x)$. During backpropagation, the gradient $I$ is obtained by discriminating loss and is subsequently multiplied by a predefined coefficient -λ. Then, $-\lambda I$ subsequently propagates to the previous layer. The two propagation processes are expressed as follows:

$$\text{Forward propagation}: Q_\lambda(x) = x \tag{8}$$

$$\text{Back propagation}: \frac{\partial Q_\lambda(x)}{\partial x} = -\lambda I \tag{9}$$

where $x$ can be $s_c$ or $s_u$.

Now we have completed the extraction of consumer health terminology. In medical social media texts, consumer health terminology can provide certain symptom information, but it is difficult to extract for three reasons. First, consumer health terminology is not included in medical terminology dictionaries, thus it is difficult to extract directly by searching a dictionary. Second, there is almost no overlap between the corresponding terminology, and therefore it is difficult to extract using string matching technology. Third, existing keyword extraction technologies do not produce the desired effect due to the lack of sufficient hard constraints and corresponding labeled data to evaluate and provide feedback.

### 3) TEXT CLASSIFICATION MODULE BASED ON DUAL CHANNEL SUBTRACTION

In order to accurately capture semantics of medical social media text, we need to make further consideration. For example, "Baby doesn't say hungry or wants to eat", if you only rely on consumer health terminology, such as "hunger" and "want to eat", the semantics of the sentence will be mapped to the medical concept ontology of "hunger" due to the use of the word "hunger". However, the semantics of this sentence is the medical concept "loss of appetite." Therefore, a dual channel subtraction structure is proposed for simultaneously using consumer health terminology and fully learning the context information of the sentences to more accurately classify medical social media texts.

The dual channel subtraction structure can accurately learn sentence semantic coding. In Figure 2, $s_c$ is a sentence code obtained by learning the complete sentence in the first channel, and $s_u$ is a sentence code obtained by learning the sequence after hiding words in the second channel that are found in the medical term dictionary. Due to the previous full extraction of consumer health terminology, the remaining words are primarily noise words without health information. The $s_u$ obtained by feature learning of these noise words is defined as noise coding. The obtained $s_c$ and $s_u$ are multiplied by a weight in order to adjust the importance dimension, and noise information is subsequently eliminated by subtracting the two. This yields more accurate sentence coding, which is represented by $s_{new}$ as defined as follows:

$$s_{new} = w_1 s_c - w_2 s_u, \tag{10}$$

where $w_1$ and $w_2$ are parameters that can be optimized in the network. $s_{new}$ will be used for classification of medical social media text.

The dual channel subtraction method differs from the traditional method of directly learning terminology sequences for text classification. The method of directly learning a terminology sequence can be regarded as single-channel learning. This method tends to lose some of the context information and is prone to interpreting sentences out of context, as many words have different semantics when expressed in different contexts. However, the dual channel subtraction structure completely considers all contextual information in medical social media text. According to the principle of semantic synthesis in computational linguistics [34], the semantics of a sentence are composed of sub-components, e.g., semantics of words and phrases according to different rules. Dual channel subtraction is used to obtain the context information of the sentence based on the words, phrases, and combination rules of the entire sentence. In other words, the dual channel structure obtains the complete information and noise information of the entire sentence respectively from two channels at the same time, and then subtracts the two to obtain the correct semantic information. From the perspective of the semantic synthesis principle, this structure can minimize the loss of sentence structure information.

## IV. EXPERIMENTS

### A. DATASET

Patient description text on medical social media is taken as an example and the patient's required medical treatment is treated as a classification problem. According to the description of the condition, the algorithm presented in this paper can automatically assign the description to a corresponding medical department so as to provide more accurate consultation or medical guidance for the patient. In total, 19,824 patient description texts were taken from

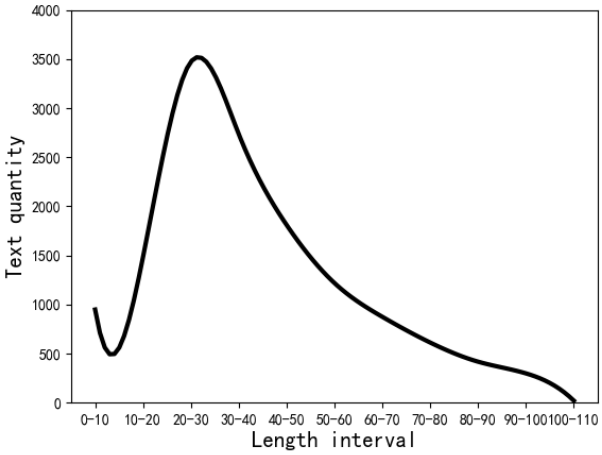


**FIGURE 3.** Length distribution of dataset text.

**TABLE 1.** Dataset composition.

| Datasets | Training set | Test set | Validation set |
|---|---|---|---|
| Pediatrics | 3000 | 500 | 500 |
| Otorhinolaryngology | 3000 | 500 | 500 |
| Internal medicine | 3000 | 500 | 500 |
| Surgery | 3000 | 500 | 500 |
| Gynecology | 3000 | 500 | 324 |

*DingXiangyisheng*'s[1] question and answer module,[1] and the department labels for the doctor who answered the question were used as the department label (e.g., gynecology, pediatrics, otorhinolaryngology, internal medicine and surgery) corresponding to the patient's description. The minimum text length in the dataset is five characters, the maximum length is 3490 characters; the length distribution of the text is shown in Figure 3. Table 1 shows the composition of the dataset.

### B. EXPERIMENTAL SETTINGS

We conducted the experiments and evaluatation based on Google's TensorFlow deep learning framework. There were three sets of experiments designed to evaluate the performance of the model. First, the performance comparison between the MSMTC and baseline methods which are widely used for text classification, such as Convolutional Neural Networks(CNN), LSTM, Bi-directional Long Short-Term Memory (BiLSTM), eXtreme Gradient Boosting (Xgboost), and Naive Bayesian Model (NBM). Second, the performance comparison between dual-channel subtraction and the single-channel learning strategy. Third, the performance comparison among different values of parameters which are the number of times about words extraction and the dimension of the word embedding.

#### 1) IMPLEMENTATION DETAILS

For training the model, the parameters in the experiment were defined as follows. The weights in the networks

[1] https://ask.dxy.com/

**TABLE 2.** Experimental results from different models.

| Models | Accuracy | Precision | Recall | $F_1$ −measure |
|---|---|---|---|---|
| CNN | 86.28% | 87.40% | 86.28% | 85.57% |
| LSTM | 85.74% | 86.84% | 85.94% | 85.80% |
| BiLSTM | 86.56% | 87.17% | 86.56% | 86.54% |
| NVB | 83.84% | 85.80% | 83.80% | 82.79% |
| Xgboost | 77.88% | 82.84% | 78.00% | 77.20% |
| MSMTC | **87.65%** | **87.95%** | **87.39%** | **87.30%** |

were randomly initialized from a uniform distribution U $[-0.01, 0.01]$. The length of the input text was set to 100. Sentences less than 100 characters in length were filled with "<pad>". The size of cells in the fully connected layer was set to 256. The learning rate was $10^{-3}$ and the regularization weight $\rho$ was set to 0.005. The adaptation rate in the gradient reversal layer was set to 1. Regarding training, the model was optimized with stochastic gradient descent over shuffled mini-batches with batch size $bs = 64$. In addition to regularization, the experiment used two other methods to control overfitting. First, the keep rate in Dropout is set to 0.5 [35]. Second, performing early stopping on the validation set during the training process if the accuracy of validation set doesn't improved after training more than 1000 rounds.

#### 2) EVALUATION METRICS

The experiment used the common evaluation metrics accuracy (Accuracy), precision (Precision), recall rate (Recall), and $F_1$-measure ($F_1$-measure) to evaluate the performance of the medical social media text classification model. These indicators are calculated with the following equations:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{11}$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \tag{12}$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \tag{13}$$

$$F_1 = \frac{2 \times precision \times recall}{precision + recall} \times 100\% \tag{14}$$

where TP indicates the samples and prediction are positive, FN indicates the samples are positive but the prediction is negative, TN indicates the samples are negative, and the prediction is negative, and FP indicates the samples are negative, but the prediction is positive.

## V. RESULTS AND ANALYSIS

### A. COMPARISON BETWEEN THE MSMTC AND BASELINES

A comparison between the performance of the MSMTC and the baseline models (CNN, LSTM, BiLSTM, NBM, and Xgboost) is shown in Table 2.

Table 2 shows that MSMTC outperforms all the baseline models with accuracy reaching 87.65%. BiLSTM is the best baseline model with accuracy reaching 86.56%, and the Precision, Recall and $F_1$-measure reached 87.17%, 86.56%,

**TABLE 3. Dual-channel and single-channel experimental results.**

| Methods | Accuracy | Precision | Recall | $F_1$ −measure |
|---|---|---|---|---|
| Single-channel | 86.73% | 86.99% | 86.61% | 86.42% |
| Dual-channel subtraction | **87.65%** | **87.95%** | **87.39%** | **87.30%** |

and 86.54%, respectively. This was followed by CNN and LSTM, whose accuracies reached 86.28% and 85.74%, respectively. Compared to the baseline models (e.g., Xgboost and NVB), our model is 4.01% higher in terms of accuracy than NVB and 9.97% higher than Xgboost.

BiLSTM and LSTM exhibit better performances, probably because BiLSTM, LSTM, and MSMTC can solve the long-distance dependence problem in sentences. Here, the gating mechanism in BiLSTM and LSTM allows information to pass selectively in order to remove or increase information to be input to neurons. The classification performance of convolutional neural networks is relatively good because the convolution operator can extract a variety of local features. However, the network implements a feature extraction process that ranges from local to global, i.e., words that are far apart can only meet and interact at higher nodes, which may cause information loss. The above three baseline models are all based on using a word vector for text classification without processing the consumer health terminology. But the MSMTC emphasizes the use of consumer health terminology, and the results show that the classification results from MSMTC are the most accurate. MSMTC is also superior to Xgboost and NBM for two reasons. First, these two models learn texts by using high-dimensional discrete features. Second, NBM theoretically introduces the assumption that the features are independent of each other. The correlation between the word features of the text is strong, therefore the classification performance of NBM is not outstanding. This also explains why the deep neural network is superior to the traditional text classification model such as NBM and Xgboost.

### *B. COMPARISON BETWEEN THE DUAL-CHANNEL METHOD AND SINGE-CHANNEL LEARNING METHOD*

The method of directly learning the extracted terminology for text classification is regarded as a single-channel learning method. The following experiments compare the classification performance of the dual-channel method proposed in this paper and the single-channel method. The results are shown in Table 3.

The dual channel subtraction method uses weighted dual channel sentence codes that are subtracted and then input to the classifier, while the single-channel learning strategy directly inputs the code learned from the extracted terminology sequence to the classifier. Table 3 shows that the Accuracy, Precision, Recall, and $F_1$-measure in the dual-channel method are higher than in the single-channel method, which reflects the importance of dual-channel

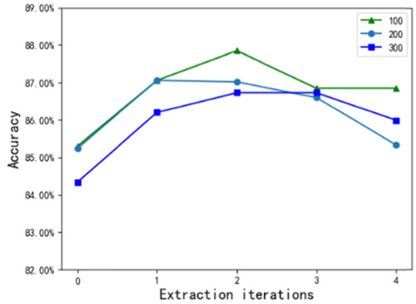


**FIGURE 4. Impact of related parameter settings on classification results.**

subtraction. First, dual channel code subtraction can fully learn context information and can be used to remove noise. Second, dual-channel subtraction can compensate for defects in the word segmentation tool because some phrases in the medical social media text cannot be completely extracted. For example, the phrase ''not comfortable'' may only be extracted as ''comfortable''. At this time, if the extracted words are directly learned and used for text classification, the performance of text classification may be affected by the lack of context information. Finally, the adversarial and self-attention mechanisms also improve the results. Therefore, the dual-channel subtraction text classification strategy provides better classification performance with medical social media text classification tasks that contain many colloquial phrases.

### *C. IMPACT OF RELATED PARAMETER SETTINGS ON CLASSIFICATION RESULTS*

The number of dimensions in the word vector were set to 100, 200, and 300 in succession. The number of consumer health terminology extraction iterations were set to 1, 2, 3, and 4 for each number of dimensions in the word vector. Text classification accuracy can be found in Figure 4.

In Figure 4, the number 0 indicates the direct use of the medical terminology dictionary. As can be seen from Figure 4, the classification performance of the model is generally better when the number of word vector dimensions is 100. When the consumer health terminology is extracted twice, the classification accuracy rate is generally the highest, and the blue line reaches a peak. As the number of iterations continues to increase, the classification accuracy decreases but is always higher than the results obtained from using the medical terminology dictionary alone. Therefore, it is not difficult to find that using consumer health terminology with medical terminology can improve the experimental results, and extraction of consumer health terminology plays an important role in text classification. However, as the number of iterations increases, a small number of irrelevant words in some texts are also extracted, which decreases the classification accuracy.

In order to provide some context regarding the extracted consumer health terminology, certain consumer health terminology and medical concept terminology are listed

**TABLE 4. Consumer health terminology and medical terminology.**

| Medical terminology | Consumer health terminology |
|---|---|
| Helium, cancerous fever, euthanasia, leave aside, pitting edema, x-protein, diphtheritis, hemiplegia, tinea cruris, transplant rejection, bone transplant status, osseous cryptococcosis, pollinosis, chronic glomerulonephritis, rapidly progressive, Paraneoplastic encephalitis, Petit mal status, Resting tremor, Seizure, Sleep disorder, Spina bifida, Spinal epidural abscess, West syndrome, Viral meningitis, Vasculitis, Tremor, Toe-walking gait, Restless legs, Occipital lobe epilepsy, Sleep-related bruxism | low fever, yellow, heat, pain, urine volume, hands and feet, can't sleep, stomachache, swell, mold, hot and cold, antipyretic, in the palace, chest pain, swallowing, allergic, saliva, pathological, comfortable, do B-ultrasound, obesity, epithelium, allergic reaction, eat less and eat more, hair loss, scorpion pain, red, belly, pain, heartbeat, fever, body, cold, belly pain, pressure, Diet therapy, intraocular pressure, breast-feeding, psychological factors, not too good, high fever, Overeating, uncomfortable |

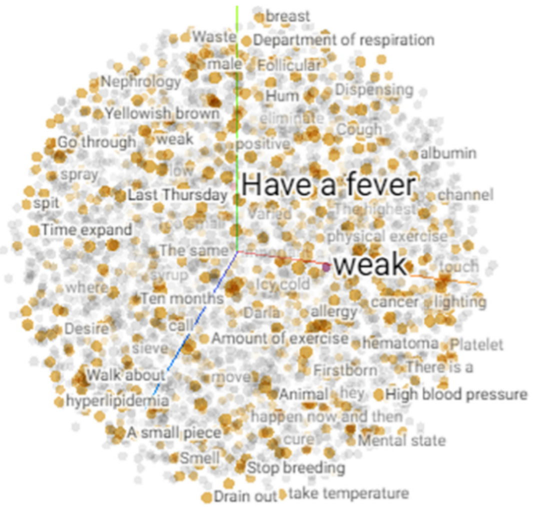


**FIGURE 5. T -SNE visualization of word vectors.**

in Table 4 for comparison. We further used the Tensorboard projector tool to visualize the word vector, as shown in Figure 5.

The medical terminology in Table 4 refers to professional medical ontology, while the consumer health terminology consists of mostly colloquial expressions, such as ''Have a fever'' and ''weak'' in the right column. These words do have a disease-related meaning in a specific context. In Figure 5, T-SNE is used as a data dimensionality reduction and visualization technique to project the word vector into a 3-dimensional space for observation. One can see that words such as ''weak'', ''take temperature'' and ''Cough'' appear near the word ''Have a fever''. This shows that consumer health terminology does have certain semantic information.

## VI. CONCLUSION

A medical social media text classification algorithm that incorporates consumer health terminology is proposed in this paper. This algorithm addresses the insufficient use of consumer health terminology features in medical social media text classification. The dual channel and adversarial mechanisms are used to automatically extract consumer health terminology, and dictionary-related medical terminology and consumer health terminology are used with dual-channel subtraction to perform classification. Various baseline models and the proposed method were compared, as well as single channel method and dual channel method. The experimental results demonstrate that our model outperforms the baseline models and shows the dual-channel structure designed in this paper provides a more accurate classification than directly learning sequences of words. We also compared the results produced with different parameter values, including the number of iterations and the dimensions of word embedding, and find that the extraction of consumer health terminology plays an important role in text classification. Extracting consumer health terminology improves the accuracy of the model and provides strong interpretability. The method proposed in this paper can automatically extract words containing health information but lack normative expressions from original medical social media texts. which attempts to deal with network language and non-standard language. The model presented here can provide more accurate medical guidance and assign patients to various medical departments. Further research can be extended to a greater range of medical social media text classification applications, such as disease severity classification and adverse drug reaction classification. It is also possible to explore the mapping of normative and non-standard terms in text classification.

## REFERENCES

[1] J. Poushter, C. Bishop, and H. Chwe, ''Social media use continues to rise in developing countries but plateaus across developed ones,'' Res. Center, Washington, Pew Internet Amer. Life Project, 2018. [Online]. Available: http://www.pewglobal.org/wp-content/uploads/sites/2/2018/06/Pew-Research-Center_Global-Tech-Social-MediaUse_2018.06.19.pdf

[2] D. P. Marshall, ''Bridging the terminology gap between health care professionals and patients with the Consumer Health Terminology (CHT),'' in *Proc. AMIA Symp. Amer. Med. Inform. Assoc.*, Apr. 2000, p. 1082.

[3] N. Limsopatham and C. Nigel, ''Normalising medical concepts in social media texts by learning semantic representation,'' in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, Berlin, Germany, Aug. 2016, vol. 1, pp. 1014–1023.

[4] A. Sarker, R. Ginn, A. Nikfarjam, K. O'Connor, K. Smith, S. Jayaraman, T. Upadhaya, and G. Gonzalez, ''Utilizing social media data for pharmacovigilance: A review,'' *J. Biomed. Informat.*, vol. 54, pp. 202–212, Apr. 2015.

[5] X. Liu and H. Chen, ''A research framework for pharmacovigilance in health social media: Identification and evaluation of patient adverse drug event reports,'' *J. Biomed. Inf.*, vol. 58, pp. 268–279, Dec. 2015.

[6] M. J. Paul and D. Mark, ''You are what you tweet: Analyzing Twitter for public health,'' in *Proc. 5th Int. AAAI Conf. Weblogs Social Media*, Barcelona, Spain, Jul. 2011, pp. 1–8.

[7] E. Aramaki, M. Sachiko, and M. Mizuki, ''Twitter catches the flu: Detecting influenza epidemics using Twitter,'' in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Edinburgh, U.K., Jul. 2011, pp. 1568–1576.

[8] L. Mollema, A. H. Irene, B. Emma, C. Rutger, M. De Hester, P. Theo, K. Gerjo, R. Robert, and D. Enny, ''Disease detection or public opinion reflection? Content analysis of tweets, other social media, and online newspapers during the measles outbreak in The Netherlands in 2013,'' *J. Med. Internet Res.*, vol. 17, no. 5, p. e128, 2015.

[9] V. Plachouras, L. L. Jochen, and G. G. Andrew, ''Quantifying self-reported adverse drug events on Twitter: Signal and topic analysis,'' in *Proc. 7th Int. Conf. Social Media Soc.*, New York, NY, USA, Jul. 2016, pp. 1–10.

[10] A. Patki, S. Abeed, P. Pranoti, N. Azadeh, G. Rachel, K. O'Connor, S. Smith, and G. Graciela, "Mining adverse drug reaction signals from social media: Going beyond extraction," in *Proc. BioLinkSig*, Boston, MA, USA, Jun. 2014, pp. 1–8.
[11] R. Leaman, L. Wojtulewicz, R. Sullivan, A. Skariah, J. Yang, and G. Gonzalez, "Towards Internet-age pharmacovigilance: Extracting adverse drug reactions from user posts to health-related social networks," in *Proc. Workshop Biomed. Natural Lang. Process.*, Uppsala, Sweden, Jul. 2010, pp. 117–125.
[12] M. T. Wiley, C. Jin, V. Hristidis, and K. M. Esterling, "Pharmaceutical drugs chatter on online social networks," *J. Biomed. Inform.*, vol. 49, no. 14, pp. 245–254, 2014.
[13] K. Denecke, "Information extraction from medical social media," in *Health Web Science*. Cham, Switzerland: Springer, 2015, pp. 61–73. doi: 10.1007/978-3-319-20582-3_8.
[14] A. R. Aronson and F.-M. Lang, "An overview of MetaMap: Historical perspective and recent advances," *J. Amer. Med. Inform. Assoc.*, vol. 17, no. 3, pp. 229–236, 2010.
[15] B. Han, P. Cook, and T. Baldwin, "Automatically constructing a normalisation dictionary for microblogs," in *Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn.*, Jeju Island, South Korea, Jul. 2012, pp. 421–432.
[16] H. Hassan and A. Menezes, "Social text normalization using contextual graph random walks," in *Proc. 51st Annu. Meeting Assoc. Comput. Linguistics*, Sofia, Bulgaria, Aug. 2013, pp. 1577–1586.
[17] S. Yajun, Y. U. Zhonghua, and C. Li, "An improving method for social media text normalization," *J. Chin. Inf. Process.*, vol. 29, no. 5, pp. 104–112, 2015.
[18] E. Tutubalina, Z. Miftahutdinov, S. Nikolenko, and V. Malykh, "Medical concept normalization in social media posts with recurrent neural networks," *J. Biomed. Inform.*, vol. 84, no. 2, pp. 93–102, Aug. 2018.
[19] S. Han, T. Tran, R. Anthony, and R. Kavuluru, "Team UKNLP: Detecting ADRs, classifying medication intake messages, and normalizing ADR mentions on Twitter," in *Proc. 2nd Social Media Mining Health Res. Appl. Workshop*, Washington, DC, USA, Nov. 2017, pp. 49–53.
[20] J. Niu, Y. Yang, S. Zhang, Z. Sun, and W. Zhang, "Multi-task character-level attentional networks for medical concept normalization," *Neural Process. Lett.*, vol. 49, no. 3, pp. 1239–1256, Jun. 2019.
[21] (May 6, 2019). *Unified Medical Language System*. Accessed: Oct. 11, 2018. [Online]. Available: http://www.nlm.nih.gov/research/umls/
[22] F. Liu, F. Weng, and X. Jiang, "A broad-coverage normalization system for social media language," in *Proc. 50th Annu. Meeting Assoc. Comput. Linguistics, Long Papers*, Jeju, South Korea, Jul. 2012, pp. 1035–1044.
[23] N. Limsopatham and N. Collier, "Adapting phrase-based machine translation to normalise medical terms in social media messages," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Lisbon, Portugal, Sep. 2015, pp. 1675–1680.
[24] I. Goodfellow, P.-A. Jean, M. Mehdi, X. Bing, W.-F. David, O. Sherjil, C. Aaron, and B. Yoshua, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[25] Y. Ganin, U. Evgeniya, A. Hana, G. Pascal, L. Hugo, L. François, M. Mario, and L. Victor, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2030–2096, May 2015.
[26] A. Vaswani, S. Noam, P. Niki, U. Jakob, J. Llion, G. N. Aidan, K. Łukasz, and P. Illia, "Attention is all you need," in *Proc. 31st Conf. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 5998–6008.
[27] S. Tokala, V. Gambhir, and A. Mukherjee, "Deep Learning for social media health text classification," in *Proc. 3rd Social Media Mining Health Appl. Workshop*, Brussels, Belgium, Oct. 2018, pp. 61–64.
[28] Medical Thesaurus. (Jan. 6, 2019). *Sogou Input Method Library*. [Online]. Available: https://pinyin.sogou.com/dict/cate-/index/132
[29] D. Hu, "An introductory survey on attention mechanisms in NLP problems," Nov. 2018, *arXiv:1811.05544*, [Online]. Available: https://arxiv.org/abs/1811.05544
[30] Z. Li, Y. Zhang, Y. Wei, Y. Wu, and Q. Yang, "End-to-end adversarial memory network for cross-domain sentiment classification," in *Proc. IJCAI*, Aug. 2017, pp. 2237–2243.
[31] Z. Li, W. Ying, Z. Yu, and Q. Yang, "Hierarchical attention transfer network for cross-domain sentiment classification," in *Proc. 32nd Assoc. Adv. Artif. Intell. Conf. Artif. Intell.*, New Orleans, LA, USA, Apr. 2018, pp. 1–8.
[32] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by back-propagation," in *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn.*, Lille, France, Junl. 2015, pp. 1180–1189.
[33] S. Kim, I. Kang, and N. Kwak, "Semantic sentence matching with densely-connected recurrent and Co-attentive information," Nov. 2018, *arXiv:1805.11360*. [Online]. Available: https://arxiv.org/abs/1805.11360 5.11360
[34] J. Mitchell and M. Lapata, "Composition in distributional models of semantics," *Cogn. Sci.*, vol. 34, no. 8, pp. 1388–1429, 2010.
[35] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

**KAN LIU** received the Ph.D. degree in computer application from Wuhan University, China, in 2002. He is currently a Professor with the School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan. His research area covers machine learning, business intelligence, social network analysis and data visualization. He has published over 50 papers in related journals and conferences.

**LU CHEN** received the B.S. degree in information management and information system from the Anhui University of Finance and Economics, Bengbu, China, in 2017. She is currently pursuing the M.S. degree with the Natural Language Processing Group, Zhongnan University of Economics and Law, Wuhan. Her research interests include machine learning, business intelligence, and text mining.