SDF: psychological Stress Detection Framework from Microblogs using Pre-defined rules and Ontologies

Mohammed Mahmood Ali[1*], Mohd Tajuddin[2], M. Kabeer[2]

Abstract: Spreading of Unwanted microblogs from Social Networking Sites (SNS) is pervasive in social media that leads to unaccountable disturbances such as Mental disorders, Wastage of precious time, Break-up of relationships, Stressness giving birth to psychological health problems and many more. To overcome these problems, the immense necessity is to ignore those unwanted microblogs in SNS, which is uncontrollable by humans due to addiction towards social media. Even the literate people fall prey to psychological stress from SNS. This seriousness of stress related issues is very rarely attended by researchers, to tackle such vicious microblogs. The prediction strategy is proposed named as Stress Detection Framework (SDF) to analyze the stress in microblog. SDF is developed using *Ontology based Information Extraction technique using Probabilistic Model (GSHL & TreeAlignment Algorithm)*, *set of pre-defined knowledge based logical rules* that constitutes of low-level attributes (simple textual, linguistic words) and visual features (emoticons & Images) and social Interaction (Likes and Dislikes) to detect and predict stress in microblog messages.SDF is compared with TeniStrength that has shown an increase of 94.2% of stress detection rate. The experimental results obtained will aid to take precise decision for blocking/eradicating/ segregating stress related microblogs from Social media (especially SNS).

Keywords: Psychological stress, Social Networking Sites (SNS), microblogs, Ontology, Stress Detection Framework (SDF), low-level attributes, Social interaction attributes.

1. Introduction

Predicting of Stress related messages in social media especially from Social Networking Sites (SNS) has made adverse affect on human health condition and their behavior. Pew Research reports briefed the upsurge of social media and its impact on civic life due to sharing of information leading towards psychological stress on human minds [1]. The dark face of Facebook SNS is it could collect and then uncover innumerable hidden facts of the individual's privacy unintentionally with other individuals unknowingly. The affordances of Facebook is programmed in such a way that explore the individual profiles, likes, dislikes, relationships, negative psychological emotions, thoughts, events, visibility and persistence resulting in unusual stressfulness [2]. The Surveillance of Facebook Offline and Online contacts had revealed the ex-partners relationships, break-ups and their association of sexual desire ended in degradation of personal growth and stressful life [3]. No doubt, the feasibility of Facebook's social interaction provides the users with uncountable advantages of Fast and economic communication for sharing of information, providing updates, personal satisfaction through entertainment and establishing reliable connectivity with other groups along the wide area networks at cheaper costs.

The accumulation of psychological stress words in microblogs from various SNS (Facebook, WhatsApp, Linkedin, Twitter) of same user is bit difficult task due to different architectures of SNS. As it is well-known fact that, Facebook has its own methodology for collecting of individual profile information at different point of time, where as WhatsApp does not support this

[1] *Associate Professor, Osmania University, Telangana – 500034, INDIA*
[2] *B.S Abdur Rahman Crescent Inst.of Sci. & Techn., Chennai, INDIA*
* Corresponding Author:Email: mahmoodedu@gmail.com

facility instead, it uses mobile number through which various activities of the users are attached at one location. Similarly, Twitter has its own affordance of collecting, sharing and responses of user's tweets from different users.

In SNS, Identifying the behavior of user varies from person to person as it depends on various artifacts such as linguistic, usage of patterns, words and their mentality of responses for a given context. To detect those words, WordNet Ontology can be effectively used for this purpose, WordNet Ontology is an intelligent logical dictionary that assists to detect the various synonyms for a single or multiple words [4]. Further, the Probabilistic Greater Similarity Hierarchy Learning (GSHL) and TreeAlignment Algorithms assists in categorizing the predicted synonym of a particular word precisely under a specified Ontology through learning process [16].

Predicting the context from set of existing words is an intelligent mining process, for this machine learning techniques are used which relies on historical knowledge. Most of these techniques are efficient, but are deployed on well-defined data to get the results from which logical rules are formulated.

The hindrance of machine learning technique is they make use of statistical measures which exploit topic-related terms, but the results obtained are not satisfactory in practical applications [5]. TensiStrength program uses the lexical approach with set of logical rules to predict the stress from tweets has given better results when compared to machine learning and sentimental analysis techniques [5]. Thus, studies suggest embedding of well-defined rules in real-time applications, assist for fast and quick retrieval of useful information.

For spying of suspicious messages from Instant Messaging Applications, the Suspicious Message Detection (SMD) framework was developed using Ontology and domain expert

knowledge guided with pre-defined rules for classification of various e-crimes [6]. Similarly, Anti Phishing Detection (APD) system aids to predict the various deceptive phishing attacks from microblogs in SNS, this APD system was developed using data mining technique in which well-defined phishing rules were effectively used [7]. Formulating of knowledge based logical rules and extracting the best features for decision making process requires extensive domain expert knowledge in a specific field.

Adverse effects on human minds with continuous usage of SNS as well as strategies exist to overcome stress from SNS are discussed in this section. The Section 2 briefs the detailed history of psychological stress detection strategies developed earlier in literature, beginning from Manual testing stress detection methods till automated tools. Section 3, elaborates the proposed Stress Detection Framework (SDF) strategy for detection of stress using WordNet Ontology and Set of pre-defined logical rules. In Section 4, SDF strategy is undergone through testing phase and the experimental results obtained for a given scenario that constitutes of chatting sessions (i.e., microblogs). Further, this SDF strategy is compared with TensiStrength stress detection technique. Finally, Section 5, we conclude that the results from SDF strategy obtained are far better than TensiStrength technique which could be able to detect the patterns of words from various users by understanding language linguistics in SNS. No doubt, Sentimental analysis program and machine learning techniques are good enough to retrieve the results using various algorithms. But, in real time applications with the use of pre-defined rules the accuracy of stress detection rate is increased up to 94%. Future scope of this research is to identify the stress from multilingual languages, short-form words among different age groups that vary from childhood to young-aged, middle-aged and old-aged people. It is well-known fact that with the development of innovative high-end technologies, electronic gadgets has become cheaper, which could be affordable by everyone and usage frequency has drastically increased. Its usage is found more between the age group of 17 to 40 as per PEW trends report [1].

2. Problem Statement And Related Work

These days, Psychological Stress is turning into a risk to individual's wellbeing. It is of critical significance to recognize and oversee the stress before it transforms into serious issues. Psychological Stress Detection (PSD) constitutes of four different methods namely i) psychological evaluation, ii) Evaluation of physiological signals, iii) Behavioral responses (Twitter) iv)and Visual & Social media interactions (Facebook, WhatsApp). In this paper we considered Textual behavioral responses, Visuals and social media interactions for predicting the PSD from Social media, especially through SNS.

In 2014, H. Lin and et. al., detected the psychological stress from Text, Visual, and Social interaction attributes from short text, images properties, and count responses received through tweets using Neural networks that involves excessive training [12]. The same authors H.Lin and et. al., extended their work in the year 2017, using Convolutional Neural Network (CNN) and Factor Graph Model (FGM) approaches to increase the stress detection rate i.e., 87.2% to 91.4% in which tweets constitutes of Text, Visual, Behavior and Social interaction attributes, among these attributes behavior is an extra attribute is considered for analysis of stress [13].

Identifying Psychological stress in SNS is rarely attended by researchers. The "blue whale challenge game" shocked the world due to multiple suicides of young children's with the use of social networks. The victims of stress-buster game are found to use keywords of suspicious activities inside their personal SNS accounts [8]. Many of these keywords are pointed towards psychological stress-levels. Decision for committing of suicides is due to increase of acute stress-levels. The cybercrime department needs to surveillance the mciroblogs of those offenders for inciting suspicious actions in SNS and then initiate rigorous steps to entrap these e-criminals [6].

Fig. 1. Twitter snippet post in *Text*, *Image*, *Likes and comments*, and a few *"stress"* words

The real time snippet of user posts is taken from one of the users twitter account as shown in Fig. 1, explores a short post in which stress related words through which the intention of emotions is conveyed by using keywords such as "unfinished work", "feeling sad" marked in red color. Similarly, an Image and Emojis (":(" or "☹") are also posted that shows emotions of the user. Identifying the stress features from user's textual words, images and Emoticons is not an easy task. The precise decision related to stress from SNS could be taken after keenly examining all these (3) three features of user's post as discussed in Section 1. Efficient models were developed earlier that spy the textual posts from SNS exists for surveillance of phishing and suspicious words [6][7]. These models embedded the Learning Ontologies using Probabilistic GSHL and TreeAlignment Algorithms which had enhanced the accuracy for categorizing the predicted words precisely under the various suspicious cybercrime domains in SNS [7] [16].

The emotions are categorized into (2) Two different types which are either Positive or Negative. Every emotion is not a stress it depends on various parameters such as context, time, day, event and occasion in which the keywords are used [11]. The stress from image is identified by extracting the features such as dullness, brightness and clearness which is calculated using the mean between the specified threshold values [9]. Similarly, the saddish emojis are used to express the emotions in Facebook which ultimately leads to stress through social media [10]. Likes and Comments, expresses the strength for the given post from 'n' number of users and the comments raised for the shared post which are collected in the form of Text, emojis or images that has to be analyzed to find stress [11]. Recently, a new approach named as TensiStrength program uses the lexical approach with set of logical rules to predict the stress from tweets has given better results when compared to machine learning and sentimental analysis techniques, but its use is restricted to Textual keywords [5].

In this paper knowledge based well-defined rules are used that constitutes Two (2) types of attributes namely, Low-level and Social Interaction attributes, in which the terms Linguistic and emoticons are mentioned. The term "linguistic" means hidden meaning for words in minds are expressed via short posts in natural language which are frequently used by children's and adults to convey the intent of message through set of words, the

term "emoticons" represents emojis from which the thoughts are expressed using small digital icons or images where as the term social interaction represents "Likes & Comments". Thus, short posts either in Linguistic words, emoticons format or through social interaction mode that shows emotions which leads to stress is predicted from SNS/ Instant Messaging System (IMS).

3. Proposed Stress Detection Framework (SDF)

In this Section the operational phases of Stress Detection Framework (SDF) for predicting the stress from SNS is elaborately illustrated in Fig. 3. The Stress Detection (SD) algorithm initiate the steps to capture the short posts/microblogs that are sent between the clients/users and stores them into database for identifying Stress using Set of pre-defined logical rules (SPSWDB) shown in Table I, and Ontology based Information Extraction (OBIE) technique. The Schematic-cum-flowchart of SD algorithm is shown in Fig. 2. DFDF
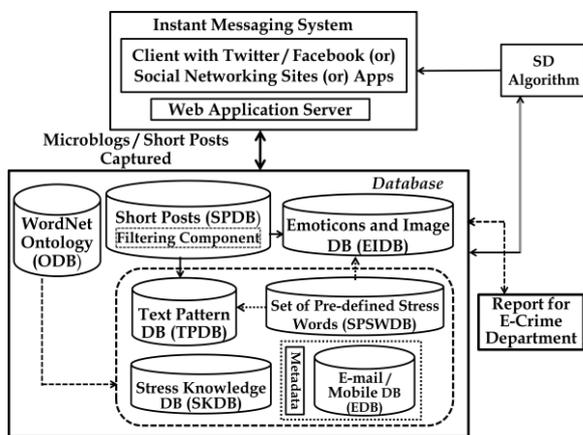

Fig. 2. Stress Detection Framework for detection of Psychological Stress

For, the proposed SDF architecture (Fig. 2), the SD Algorithm (Fig. 6) is applied that initiates the microblogs to be captured into SPDB. The words stored in SPDB are checked for unnecessary words and are filtered out using Information retrieval techniques [14]. After filtering of irrelevant words, the Textual words, Emoticons, and Images are found to exist. The Textual words are pushed into TPDB, where as Emoticons and Images are pushed into EIDB [13]. The relevant words stored in TPDB are mapped with set of pre-defined logical rules stored in SPSWDB words to find the Stress words, on detection these stress words are pushed into SKDB. Subsequently, Emoticons and Images that are stored in EIDB are mapped with SPSWD, to find the stress related emojis, on detection, these emoticons are also pushed into SKDB where as for images the saturation (brightness) is calculated using mean value and then the stress value obtained is pushed into SKDB [9].

In the next stage, the stress words, emoticons and image mean value which are stored altogether in SKDB are checked for threshold values of stress level. When, the pre-defined constraint of threshold is satisfied the stress is identified and reported to e-crime department. The Email-id or Mobile number, used during the chat session is stored in EDB. Further, metadata stores the timestamp related information such as Time, Date, No.of Connections with peers, Total No. of messages, Place & Location, IP & ISP Address, No. of Comments, Likes and

Dislikes information are stored. WordNet Ontology assists in creating of Synonyms, each of pre-defined stress related words stored in SPSWDB words uses WordNet ontology (OBIE) support internally [15]. If the words of TPDB are not mapped with SPSWDB then the WordNet Ontology logically tries to create the ontology for these TPDB words such that it could appropriately map with SPSWDB words to improve the efficiency in predicting of stress words.

The Processing steps of SDF architecture is shown in Schematic-cum-algorithmic Steps as shown below:

1. The First step is for capturing the Short Post and Microblog messages sent between the chat mates of Social Networking Site (SNS). These messages are saved in the database (SPDB).
2. In this Second step, Text posts are taken, these text message are Segregated to plain text by removing of stop-words (such as articles, preposition), Emoticons and Images.
3. Third Step constitutes of Three (3) important tasks for detection of stress are given below (*refer Table 1*):
   a. Each of the Textual keywords of the plain text are picked and starts comparing with pre-defined logical words defined in SPSWDB (Rule 1) to find out linguistic lexicon stress words (Stress Lexicon, Negative emotion Lexicon and Negating words Lexicon) present in the Text. During this process WordNet Ontology is effectively utilized if mismatch occurs.
   b. Calculate mean value of brightness of image using formulae given in Equation (1) :

Brightness / Saturation Formulae:
$$S_{HSV} = \frac{\max(R,G,B) - \min(R,G,B)}{\max(R,G,B)} \quad \text{…................... (1)}$$

$$V_{HSV} = \max(R,G,B)$$

   c. Similarly, Calculate the mean value for the posted likes and comments using the forumulae given below:

$$M_{L,C} = \frac{\text{Total No. of Likes + Comments}}{2}$$

4. The Values calculated from the above steps are stored in SKDB as follows.
   a. The stress words that are compared with SPSWDB, on detection are shortlisted and then stored in SKDB. Subsequently, Emoticons for which a match is found after comparing with SPSWDB are also stored in SKDB.
   b. Whereas the calculated Mean Image value retrieved after mapping with pre-defined threshold value (SPSWDB (Rule 2)) given in SD algorithm is stored in SKDB.
   c. Similarly, the calculated Mean value of social interactions i.e., likes and comments are checked with the pre-defined threshold value if it is met then that value is stored in SKDB.
5. All the features retrieved from Textual attribute, Emoticons attribute, Image and Social interaction attribute which is stored in SKDB are fed into the probabilistic SD algorithm to predict the stress levels of the user.

Finally, in last step results obtained regarding stress-levels and its seriousness are predicted, which is stored in the form of report. If the chatting session activities are encounter into stressful activities then a recommendation is submitted automatically to e-crime department for appropriate action for stress management in SNS.
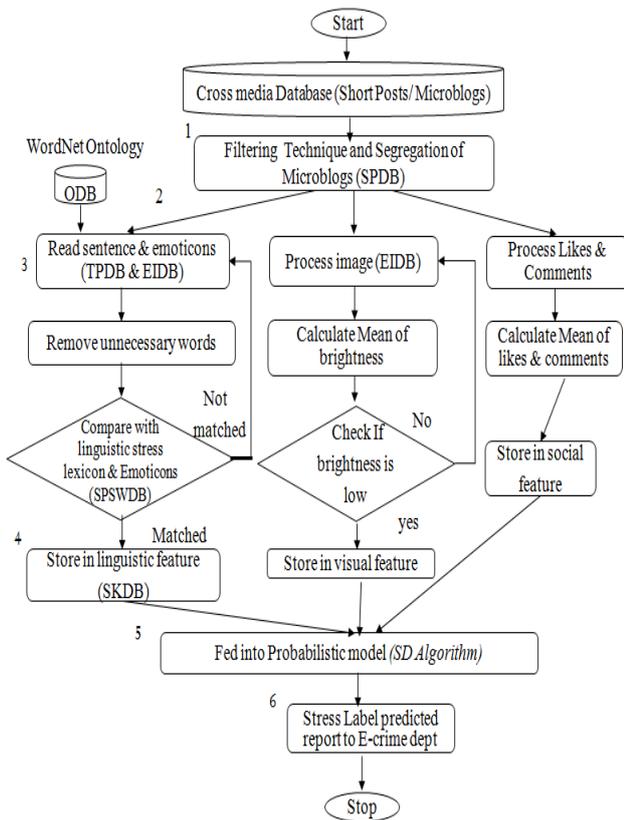
Fig. 3. Schematic-cum-algorithmic Steps of SDF is Illustrated

The OBIE, is a probabilistic learning method that predicts the domain to which these stress words belongs [16]. We used different database tables namely SPDB, TPDB, EIDB, ODB, SPSWDB, SKDB, EDB and Metadata. In this Framework, SPDB (Short Posts database) is used to store the online messages/posts that are communicated in between the users (chatmates). ODB (Ontology Database) is a lexical database that identifies terms, Synonyms, Concepts, Taxonomy (concept hierarchy), relations, Axioms and Rules [4] (Table I).

3.1 Topic (Domain) Hierarchy Construction GSHL Algorithm

Wang Wei and et al., had developed Global Similarity Hierarchy Learning (GSHL), this algorithm recursively finds for the most alike topics of the current "root" topic and removes of those words under the domain which does not fulfill the condition of the difference of K_L divergence. This algorithm starts with an initial topic as the root node and look for the top n most similar topics according to (dis)similarity measures. The parameters used in the algorithm are given below:

- N — The total number of topics (domains i.e. root words).
- M_c — The maximum number of sub-nodes (words) for a particular node (root node, i.e. domain/topic).
- TH_s and TH_d — The thresholds for similarity and divergence measures.
- TH_n — The noise factor, defined by the difference between two K_L divergence measures $D_{KL}(P//Q)$ and $D_{KL}(Q//P)$.
- I — Maximum number of iterations.

The parameters TH_s, TH_d, TH_n are user-specified values, which can be changed to obtain desirable precision and accuracy. Particularly, in our experiment, we noticed that setting TH_s, TH_d,

and TH_n within some narrow range results in negligible variation of precision values. The pairwise measures of Cosine similarity, JS divergence, and KL divergence are collectively denoted as the M_s matrix. But we have used KL divergence for *dis* (similarity).

The algorithm will terminate according to the conditions specified in the while loop. The pseudo code for *GSHL* algorithm is shown in Fig. 4.

Table 1. Depicts set of knowledge based pre-defined logical rules internally supported with WordNet Ontology for Rule 1 Linguistic stress lexicon

RULE 1 (SPSWDB)	
Categories of Stresses (Domain)	**Stress words/Emoticons or Images to be detected**
Stress Lexicon →	Anxious,sad,depress,anger,pain,grief,restless,sleepless,failure,sorrow,lonely,quit,worry,nervous,mental,break,anguish,argu,arrogan,atrocious, bereave, brokenhearted
Negative emotion Lexicon →	Hurt, ugly, anger, annoy, boredom, sadness, grief, disapproval, low self-esteem, callous, despair, devastate, hate
Negating words Lexicon →	Cannot, cant, couldn't, could not, didn't, did not, doesn't, does not, don't, do not, hasn't, has not, isn't, is not, never, no, not, shouldn't, should not, won't, wont, wouldn't
Emoticons and Images → *(Mean Value)*	"%-(", ")-:", "):", ")o:" , ":(", ":*(", ":,(", ":-", ":-&", ":-(",
RULE 2	
Images (Brightness/Saturation) → *(Mean Value)*	Mean of brightness and contrast of images is calculated. When Mean is less than 0%, image has low brightness. When Mean is greater than 0%, image has high brightness. (low < 0% >high)
Social Interaction *(Mean Value)* →	No. of Likes, Dislikes and Comments
RULE 3 (threshold value)	
Check the user-defined threshold value for the stem words that may belongs to multiple domains, using precision and Mean values *(Rule 1 and Rule 2)* **SD algorithm** *(Fig. 6)*	
RULE 4 (undetected words to be ignored)	
Sometimes, Special characters, unknown image formats are also sent via SNS which are ignored	

3.2. OBIE Model for Root word (Domain) Extraction using Treealignment algorithm

Ontology refers to intelligent information constitutes a set of words within a domain and predicts the association among those words within the specific domain and among the other successive domain(s). The hidden stress words are explored from the short posts or microblogs using pre-defined logical rules (SPSWDB) given in Table I, and stress domain words are identified i.e., Rule 1 *(Stress Lexicon, Negative emotion Lexicon and Negating words Lexicon, Emoticons)* using ontology, Rule 2 *(Images and Social interaction attributes)* are found in our Stress Detection Framework. The Ontology actively assists in forming of partial tree (constrained to 2-levels, i.e. *"parent-child"*) using algorithm in [15] [17]. During this Tree building phase, the stem words *(TPDB)*, participates in mapping with the Domain words *(SPSWDB)*, at the same time threshold value for each Domain is checked using SD algorithm shown in Fig. 6.

Algorithm 1. *GSHL (root)*
Require: Initialize *V, Ms, I, THs, THd, THn,* and *Mc*.
Ensure: A terminological ontology with "broader" and "related" relations.

```
1   Initialize V, Ms, I, THs, THd, THn, and Mc;
2   while (i < I and V is not empty) do
3     Add current root into V ;
4     Select most similar Mc nodes (words) of
5       root word (topic) from Ms;
6     Add similar nodes into Vtemp;
7     Remove nodes in Vtemp against//similarity and divergence
8   //difference condition
9     for (all nodes n ,in Vtemp) do
10      if (Sim(ni, root) > Sim(ni, Sibling(root))) then
11        Assert broader relations between root and
12      topic ni;  //relationship between topics is broader
13      else default
14        Assert related relation between root and
15        topic ni;  //relationship between topics is nearer
16      end if
17      Move topic ni from Vtemp to V ;
18      Increment i by 1;
19    end for
20    Remove current root from V ;
21  end while
```

Fig. 4. Pseudo code of Terminological ontology to find the root word from domain (topics) using pre-defined threshold value(s) tunable

Algorithm 2. *TreeAlignment(RootNode,node)* for Domain (Topic) extraction using Ontology concepts

Require: Initialize *RootNode, Node(s), MinThreshold* value
Ensure: Domain (topic) extraction from stem words.

```
1   for (int i:1 to numDomainTopic {
2   //traverse tree for DomainTopic i
3    NumLevels=2;  //size of tree is 2(parent & child)
4    for (int j:1 to NumLevels){
5    RootNode[ ]Nodes=Empty // top most level 1
6    Node[ ]nodes=getNodesAtlevel(nodes);
7    //All the stem words from the TCDB at level 2
8    RootNode[]Nodes=Node[]nodes
9    //stem words assigned to RootNode
10     checkForRepetitiveData(Nodes);
11     { call GSHL(Root) }//check the stem words(TPDB)
12     //matched in other domain i.e SPSWDB using Ontology
13     checkFordisjunctiveData (nodes);
14     { call GSHL(Root, MinThreshold) } //check for stem
15      //words that belongs to more than one domain (topic)
16    }//end for
17 }//end for
```

Fig. 5. Pseudo code of Tree Alignment for classification of Domain(s) Rule 1 *(SPSWDB)* for stress stem words *(TPDB)*

3.3. Role of Stress Detection(SD) Algorithm for the proposed Stress Detection Framework.

Detection Framework, as already discussed in previous section that has initiated the overall progress starting from storing text messages in *SPDB* till finding the stress keywords by providing a detailed report from S*KDB* and *EDB* databases to E-crime department on detection of stress words. SDF schematic cum algorithmic steps of Fig. 3, are revisited again are shown in Fig. 6.

Algorithm 3. *Stress Detection Algorithm (SD)*
Input: Short posts/microblogs stored in Text Database (SPDB) (day to day) from IMS/SD Framework.
Output: Report to E-crime when Stress messages are detected

```
1 Do{ //Apply Ontology based IE technique for filtering unnecessary
     //words and pick stress words from SPDB (if found) and push to
     //(TPDB) for Text and (EIDB) emojis and images which
     //include stem words mapped with pre-defined
    //knowledge-based  rules stored in (SPSWDB) rule 1  & rule
    //discussed in section 3.
     Push Messages to SPDB //instant messages stored in SPDB
2:    Do{//Scan SPDB for relevant stress words patterns if found
         //store it in TPDB and perform mapping with SPSWDB
        //(rule 1) using
        // Ontology (OBIE)  building taxonomy of stem words
3     Call TreeAlignment algorithm{ // algorithm discussed in Fig. 5
        // initially all stem words (TPDB) mapped to empty root
        //node using OBIE model forming tree
4       Call GSHL algorithm {//Check threshold values of stem
    //words (TPDB) with root nodes stored in SPSWDB(rule 1) using
         //algorithm 1 shown in Fig. 4 and finds stress type(s)
         //activity i.e. domain topic(s)
5       Scan TPDB
6       Push patterns to SKDB //stem words
        } //end of call GSHL
7     Compare TPDB with SPSWDB{
8      If TPDB==SPSWDB Then Push patterns to SKDB
         //stem words with Stress Domain(s) stored
       else
        Do Nothing
9      End If
      } //end compare
      } //end  call TreeAlignment
10    }until TPDB!=NULL //end of do
11   If TPDB==SKDB Then //Stress words found  with Domain
12   Check SKDB {//check the knowledge database using E-crime
         // monitoring system program for type of cyber threat
         // activity (i.e stress linguistic stem words along with stress
     // category ( i.e domain) using Treealignment algorithm (Fig. 4)
13       if SKDB=='TRUE' then {// if Stress words match then
14         Check EDB {//trace the profile details (emailid, phone
                //number, ISP IP address and location details
15         Report to E-crime Department //detailed report
          //including threat activity details traced from KDB/EDB
16       }//end of check EDB
17     }//end if
18   }//end of check SKDB
19 } //end of do
```

Fig 6. Pseudo code showing the Overall working of our proposed SD Framework for identifying stress messages and reporting to E-crime dept

4. Implementation And Experimental Results

4. 1. Evaluation method for data sets

Precision metric is used as given in Equation (2) [14] for evaluation of tweets in SD Framework. The stress words extracted efficacy are based on two factors, the number of actual words available in the pre-defined database i.e. *SPSWDB* with respect to stress domain, to that of the number of extracted stress words from tweet chat session:

$$Precision\ (P) = \frac{Correctly\ Extracted}{Total\ Extracted\ Correctly} \times 100 \qquad (2)$$

4.2. Preparation of data sets

Stress related datasets in social media are not freely accessible due to restriction of privacy policies of FISA Act. Many of the stress detection strategies tested their developed frameworks for detection of stress from data collected through their local databases from social networking sites which includes Twitter, Facebook, Instagram, WhatsApp and other social media applications [2, 3, 5, 9-13]. To test our SDF architecture, we have used pre-defined rules that constitute of stress related words totaling of 57 words that exclude emoticons & Images which

counts 27 in numbers. Further, each of these words are supported with WordNet ontology that had created synonyms for the existing pre-defined words summing up to 432 textual stress words.

4.3. Tested using SD Framework and TensiStrength

The real chatting session is intentionally conducted and the experimental results are demonstrated for the conversation happened between the two users, as shown in Fig. 7.
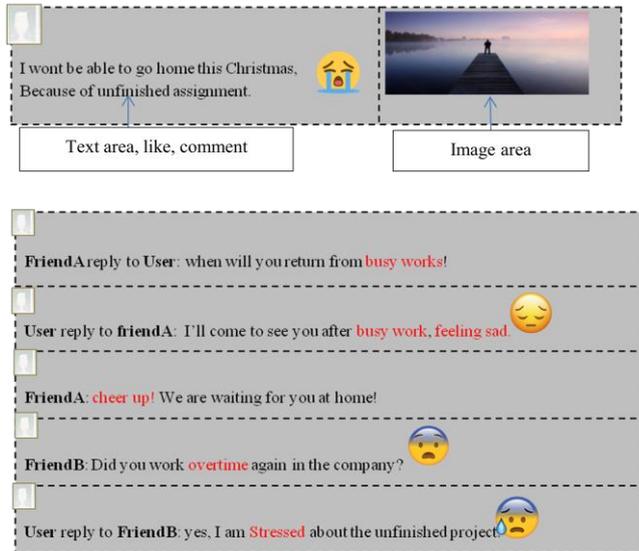
Fig. 7. Real Tweet, constitutes of Linguistic stress lexicon words, Emoticons and Image

For the twitter chatting session (Fig. 7), is tested using SD algorithm (Fig. 6), where the threshold value is set to 7%. The stress are in red-color are detected namely "busy works", "feeling sad", "unfinished" (not detected), "overtime", "stress". Apart from that, Emoticon which are exactly mapped with our pre-defined database (SPSWDB) rule 2, are Four (4), where as stressful image mean value appeared to be highly saturated is One (1). The precision rate obtained by Tensistrength is 58.32%, whereas 94.2% with SDF system as shown in Table II.

Table 2. Precision Rate Obtained For Real Sample Twitter Chatting Session (Fig. 7)

Parameters used	TensiStrength	SDF system
Textual data	98.11%	94.43%
Textual+Emoticons	68.53%	96.21%
Textual+Emoticons+Social Interactions	58.32%	94.20%
Textual+ Social Interactions	97.49%	97.82%

5. Challenges And Future Work

As it is of significant importance to detect and manage stress before it turns into severe problems. Traditional methods are mainly based on interviews conducted by psychologists or using sensors which is time consuming process. The proposed SDF strategy detects psychological stress by employing the probabilistic model [17]. In SDF model two unique features are added which were not done earlier, except TensiStrength which has used only one feature of pre-defined rules that to only textual words are considered. In our research work, the First feature is we have not only embedded set of predefined logical rules

comprising of Linguistic stress lexicon words, but also Emoticons & Images and Social interaction attributes. The Second feature is integration of OBIE technique (Probabilistic GSHL Algorithm) which was not used earlier by any of the state-of-art systems. The comparative features of SD Framework and TensiStrength are shown in Table III.

Table 3. Features of Sdf & Tensistrensth

Features	TensiStrength	SDF System
Support for Social Interaction	Supported	Supported
Pre-Defined rules	Very less	More (because of WordNet)
Support for Emoticons and Images	Not Supported	Supported
Stress Activity Detection	Static Detection (time consumed is less)	Dynamic Detection (time consumed is more)
Report Generation for e-crime department	No Report	Report with details (email, Mobile No, etc)
Ontology support	No	Yes

The SDF is developed using Ontology based Information Extraction technique using Probabilistic Model (GSHL & TreeAlignment Algorithm), set of pre-defined knowledge based logical rules that constitutes of low-level attributes and visual features and social Interaction to detect and predict stress in microblog messages. In future we plan to detect stress from multilingual words and Multimedia contents. Apart from English language we would like to enhance this work in other languages based on various features like location-based, cultures, different people and their nature of use.

References

[1] A. Perrin, "Social Media Usage: 2005-2015," Pew Internet & American Life Project, Washington, October 8th 2015, pp. 4-10,

[2] [Online]. Available: http://www.pewinternet.org/2015/10/08/social-networkin g-usage-2005-2015.

[3] J. Fox and J. J. Moreland, "The dark side of social networking sites: An exploration of the relational and psychological stressors associated with Facebook use and affordances," *Journal of Computers in Human Behavior, Elsevier,* vol. 45, pp. 168-176, 2015.

[4] Tara C. Marshall, "Facebook Surveillance of Former Romantic Partners: Associations with Post Breakup Recovery and Personal Growth," *Cyberpsychology, Behavior, and Social Networking, Liebert Open access journal,* vol. 15, No. 10, pp. 521-526, 2012.

[5] WordNet Ontology, 2018. [Online]. Available: http:www.ontologyportal.org. Accessed on Feb 21 2018.

[6] Mike Thelwall, "TensiStrength: Stress and relaxation magnitude detection for social media texts," *Journal of Information Processing and Management*, vol. 53, pp. 106-121, Elsevier, 2017.

[7] Mohammed Mahmood Ali, Khaja Moizuddin Mohammed, and Lakshmi Rajamani. "Framework for surveillance of instant messages in instant messengers and social networking sites using data mining and ontology," in *Proc.* of *Students' Technology Symposium (TechSym)*, pp. 297-302, IEEE, 2014.

[8] Mohd Mahmood, and Lakshmi Rajamani, "APD: ARM Deceptive Phishing Detector System Phishing Detection in Instant Messengers Using Data Mining Approach," Krishna P.V., Babu M.R., Ariwa E. (eds), Global Trends in Computing and

Communication Systems, Communications in Computer and Information Science, Springer, vol. 269, pp. 490-502, 2012.

[9] A. Kumar, S. N. Pandey, V. Pareek, M. Faiq, N. Khan, and V. Sharma "Psychobiological determinants of 'Blue Whale Suicide Challenge' victimization: A proposition for the agency mediated mental health risk in new media age," *Etiologically Elusive Disorders Research Network (EEDRN)*, 2017.

[10] Huijie Lin *et. al.,* "Psychological Stress Detection From Cross-Media Microblog Data Using Deep Sparse Neural Network," IEEE, 2014.

[11] Ye tian *et al.,* "Facebook Sentiment: Reactions and Emojis", in *Proc. of the Fifth International Workshop on Natural Language Processing for Social Media*, Association for Computational Linguistics, Valencia, Spain, 2017, pp. 11-16.

[12] AL Zell and L Moeller, "Are you happy for me… on Facebook? The potential importance of "likes" and comments," *Computers in Human Behavior*, Elsevier, vol. 78, pp. 26-33, 2018.

[13] H. Lin, J. Jia, Q. Guo, Y. Xue, J. Huang, L. Cai, and L. Feng, "Psychological stress detection from cross-media microblog data using deep sparse neural network," in *Proc. of Int. Conf. Multimedia Expo,* IEEE, pp. 1–6, 2014.

[14] H. Lin, J. Jia, J. Qiu, Y. Zhang, G. Shen, L. Xie, J. Tang, L. Feng, T.S. Chua, "Detecting stress based on social interactions in social networks*," IEEE Trans. Knowl. Data Engg*., vol. 13, No. 9, pp 1-14, 2017.

[15] C.D. Manning, P. Raghavan, and H. Schutze, "Introduction to Information Retrieval," 2008, Cambridge University Press.

[16] Jer Lang Hong, "Data Extraction for Deep Web Using WordNet", *IEEE Transactions on systems, man and cybernetics*, vol. 41, Issue 6, pp. 854-868, 2011.

[17] Wang Wei, Payam Barnaghi, and Andrzej Bargiela, "Probabilistic Topic Models for Learning terminological ontologies*," IEEE Tran., on Knowledge and data engineering*, vol. 22, No. 7, pp. 1028-1040, July, 2010.

[18] Y. Zhai and B. Liu, "Web data extraction based on partial tree alignment," in *Proc. of International World Wide Web Conference Committee (IW3C2)*, ACM, 2005.