Disease discovery-based emotion lexicon: A heuristic approach to characterise sicknesses in microblogs

Abstract

The analysis of microblogging data has been widely used to discover valuable resources for timely identification of critical illness-related incidents and serious epidemics. Despite the numerous efforts made in this field, making an accurate and timely prediction of incidents and outbreaks based on certain clinical symptoms remains a great challenge. Hence, providing an investigative method can be crucial in characterising a disease state. This study proposes a heuristic mechanism by using an unsupervised learning technique to efficiently detect disease incidents and outbreaks from the tweet content. We categorised the types of emotions that are highly linked to a specific disease and its related terminologies. Emotions (anger, fear, sadness, and joy) and diabetes-related terminologies were extracted using the NRC Affect Intensity Lexicon and a part-of-speech tagging tool. A two-cluster solution was established and validated. The classification result showed that K-means clustering with 2 centroids had the highest classification accuracy (96.53%). The relationship between diabetes-related terms (in the form of tweets) and emotions were established and assessed using the association rules mining technique. The results showed that diabetes-related terms were exclusively associated with fear emotions. This study offers a novel mechanism for disease recognition and outbreak detection in microblogs which is useful in making informed decisions about a disease state.

Keywords: diabetes; emotion lexicon; disease detection; part-of-speech tagging; association rules mining; Twitter

1. Introduction

Social media websites offer a promising opportunity for tracking people’s opinions about various health-related issues around the world. These websites provide users with the means to exchange thoughts and feelings on health-related topics. Using social media platforms can increase the chance of sharing personal experience among users at any time and place. Currently, social media is being used by public health organisations as a platform to help them reach a wider audience (Sharma & Kaur, 2017). Using social media sites to map people's health behaviour, disease transmission procedure and experience has boosted information diffusion (Benton, Coppersmith, & Dredze, 2017). Health-related topics that are shared between online users can be help health experts improve clinicians' skills, training simulations, and infectious disease monitoring.

Twitter is one popular example of social media sites that offer a reliable medium for users to freely exchange various health-related information (Motamedi, Jamshidi, Rejaie, & Willinger, 2020; Nejad, Delghandi, Bali, & Hosseinzadeh, 2020; Xu et al., 2016). Data extracted from Twitter data can help health organisations, such as state health departments and large healthcare systems, to use timely health statistics for decision-making (Mejova, Weber, & Macy, 2015). This has motivated many scholars to use Twitter in an attempt to extract knowledge related to public health and chronic disease (e.g., diabetes) (Lu, Wu, Liu, Li, & Zhang, 2017).

Diabetes is one of the most complicated diseases that annually imposes heavy costs on economies worldwide. There are two types of diabetes: type 1 (often abbreviated as T1 or T1D on Twitter) which occurs when the pancreas does not produce sufficient insulin. Type 2 (T2 or T2D) develops when the body does not effectively respond to insulin. With T2, most patients tend to experience certain behavioural changes, such as adopting a healthy diet and increasing physical activity, in an attempt to reduce diabetes-related complications (Tamarai, Bhatti, & Reddy, 2019). Type 2 diabetes is the leading cause of premature deaths, it can lead to a number of health issues, including heart diseases, stroke, kidney disease, blindness, nerve damage, leg and foot amputations (Asif, 2014). In general, patients with diabetes are advised to control blood sugar levels which includes eating healthy food, regular exercise, and medicines (Wilt, Kansagara, Horwitch, & Barry, 2018). However, all these aspects and many others are shared by diabetic patients via online social networking websites. Diabetes-related information, such as medication and outcomes, may offer an effective means for characterising individuals' health, thus providing key insights into their disease state. This leads us to argue that extracting latent information related to diabetes symptoms from microblogs can help improve the efficiency of early disease detection methods.

Many scholars have examined the potential of different machine learning algorithms in tracking users' sharing of information (Yavary, Sajedi, & Abadeh, 2020). Machine learning is one example that gives the system the capability to automatically learn and improve over time (Habib, Asghar, Khan, Habib, & Khan, 2019). Classification algorithms have been extensively used by many scholars to process social media content and predict a specific outcome (Choubey, Kumar, Tripathi, & Kumar, 2020). In a health context, there has been an increasing attention on the use of social media and the Internet for general disease surveillance (Priyadarshi & Saha, 2020). For instance, a study by Sarsam, Al-Samarraie, Ismail, Zaqout, and Wright (2020b) proposed a real-time bio surveillance mechanism in order to predict the early-stage of migraine disease. The authors researchers used machine learning algorithms to extract migraine-related symptoms based on the interconnection between emotional and climatic factors embedded in Twitter messages. Molaei, Khansari, Veisi, and Salehi (2019) introduced an entropy-based method for minimising errors as well as time complexity in the event of predicting the influenza-like illness (ILI) based on data derived from Twitter. The authors found that deep neural network and entropy-based methods help in minimising the mean average error by up to 25% compared to other nonlinear methods. Another study by Ji, Chun, Wei, and Geller (2015) investigated the main issues of spreading public concern about epidemics. They distinguished personal tweets into negative and positive tweets using Naïve Bayes classifier. From a recognition perspective, Rani (2018) proposed the “Bigram Text Classification” to track the diabetes disease using special features generated from “High Ranked Feature Extraction Algorithm”, followed by classifying the tweets using support vector machine classifier (SVM).

Based on these, it can be observed that most of the utilised methods in previous studies were supervised in nature. In addition, the problem of filtering which content is relevant to a given topic of interest is still a matter of debate (Carvalho, Rosa, Brogueira, & Batista, 2017). According to Carvalho et al. (2017), it is very unlikely that more than a few thousand tweets are relevant to the search query. Therefore, this study aims at proposing a new approach to characterising sicknesses in microblogs. For this purpose, we have chosen diabetes as a case for this study. The proposed approach provides a heuristic way for detecting diabetes using emotions (embedded in the body of tweets) and association rules mining (unsupervised learning technique). In this study, we have three main objectives: extracting the diabetes-related groups from Twitter; identifying the types of emotions prominent in each group; and finding the types of emotions that are associated with diabetes-related terms. To accomplish our goals, an unsupervised learning technique was used to cluster diabetes groups from the collected tweets and to identify potential relationships between diabetes-related terms and diabetes-related emotions.

2. Literature review

The rapid increase in user-generated online content has become a means for clinicians and public health practitioners to gather health-related data about. All publicly available are is invaluable resources to obtain some set of goals (Lee, Agrawal, & Choudhary, 2013). Because of that, social media platforms can be effectively used as new surveillance systems, especially when the traditional methods of disease surveillance take a long time to detect a potential outbreak (St Louis & Zorlu, 2012). Social media websites for clinicians, patients, and general public can offer a rich context for users to share healthy lifestyles, which helps in making informed medical decisions and improves personal health management (Zhou, Zhang, Yang, & Wang, 2018). Twitter, an example of such websites, provides an effective and efficient way for users to communicate and share their personal experience (Wakamiya, Kawai, & Aramaki, 2018). This also includes a rich reflection of individuals' emotions which influence their decision-making and attitude (Wang, Chen, Thirunarayan, & Sheth, 2012). Predicting user's emotion from shared messages on social media is referred to as "sentiment analysis" (Arora & Kansal, 2019). Sentiment analysis can be used to identify and map online opinions shared between users in the social media space (Chen, Hossain, & Zhang, 2020). Decisions that are linked to health domain and individuals' behaviours usually take place in emotionally-laden contexts (Ferrer & Mendes, 2018). Sarsam, Al-Samarraie, Ismail, Zaqout, and Wright (2020a) explored the interconnection between certain emotional and climatic factors related to the migraine disease. The authors found that sad emotions can play a significant role in disease recognition. In the same context, a work by Mannix et al. (2016) highlighted that frustration, depression, and anxiety might be linked to specific disease conditions. For example, the emotions that reflect anger can potentially influence individuals' health by causing both chemical and hormonal imbalance (Yadav, Yadav, &

Sapkota (2017). In addition, fear has the potential to trigger anger response which can influence individuals' blood pressure (Yadav et al., 2017). Also, in diabetes-related studies, a work by Ghaffari, Salsali, Rahnavard, and Parvizy (2014) revealed that several diabetic patients are likely to experience fear and anxiety responses. Therefore, analysing the emotions in tweets can offer some important applications to public opinion (Bravo-Marquez, Frank, Mohammad, & Pfahringer, 2016), healthcare (Dadich & Olson, 2017), and disease recognition (Schwartz et al., 2018). Many previous studies have been conducted to examine ways for characterising disease or interspecific associations from Twitter messages. For instance, Hayate, Wakamiya, and Aramaki (2016) studied the role of analysing tweets to predict Influenza outbreak in a specific time and place. Another study by Lee et al. (2013) used data obtained from Twitter to develop a real-time disease surveillance system in an attempt to track both Influenza and cancer activities. The authors analysed the textual data and investigated the popularity of terms based on the type of disease, symptoms, and treatments. They also used Natural Language Processing (NLP) in order to analyse emotions within text (Giatsoglou et al., 2017). In addition, important document features can be further used to provide a successful sentiment detection process using frequent terms, parts-of-speech, emotional terms and phrases (Chatzakou & Vakali, 2015).

Based on these, it can be concluded that emotions of online users can be effectively used in conjunction with supervised learning algorithms in order to effectively detect diseases. In this study, we proposed a heuristic approach to characterise diabetes by examining patient-related emotions in tweets. Using affect intensity lexicon method, we were able to extract four types of emotions: anger, fear, sadness, and joy. Disease-related terminologies were also extracted by means of the part-of-speech tagging technique. The relation between diabetes-related terminologies and diabetes-related emotions were shaped and assessed using the association rules mining technique.

3. Procedure

Figure 1 summarises the main stages in this study, including data collection, data pre-processing, text clustering, emotion extraction, part-of-speech tagging, association rules mining, and validation.

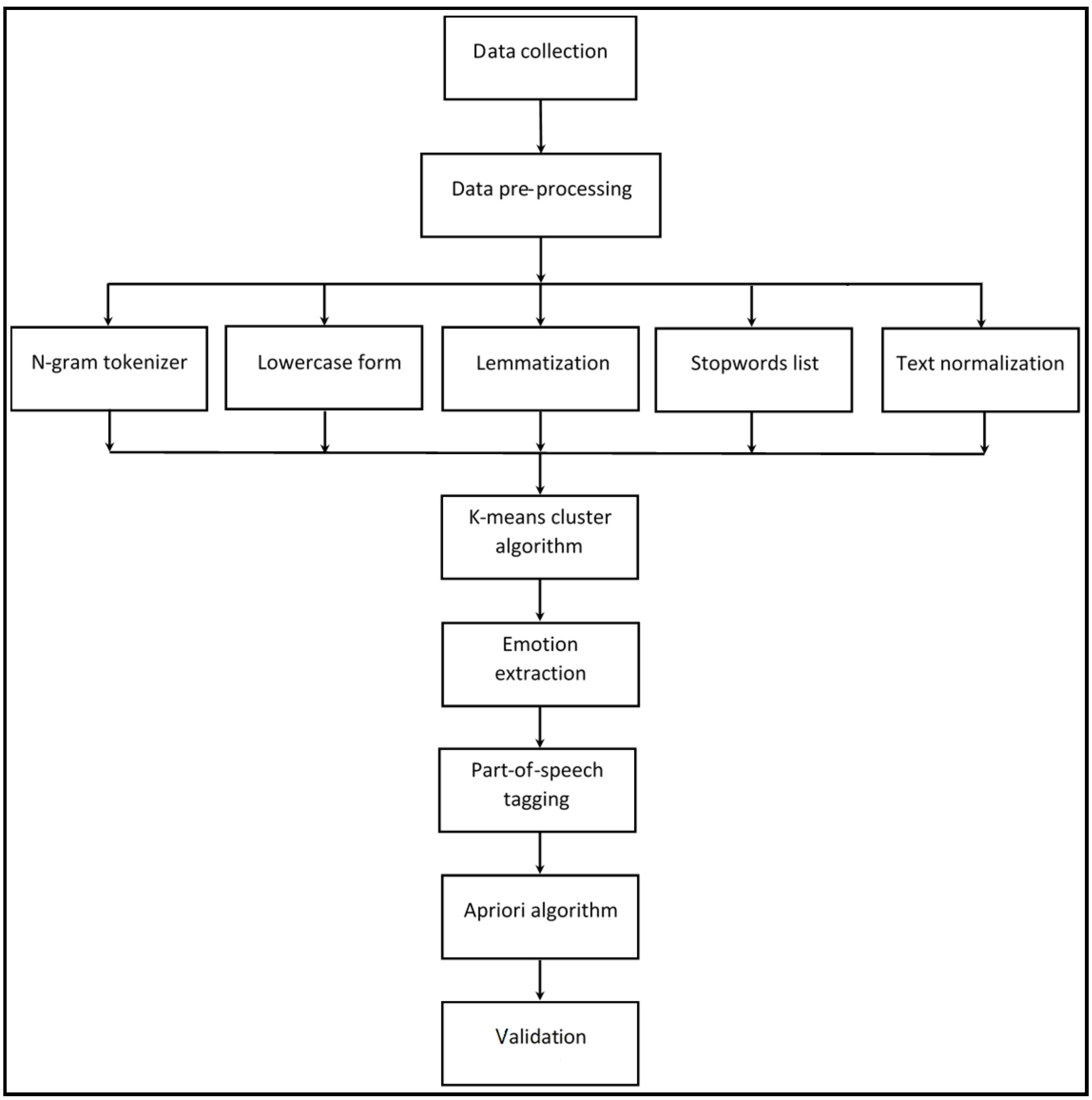

Figure 1: Data collection and pre-processing procedure

3.1 Data collection

A total of 573,718 English tweets were collected within a time span of six months (January 15th 2019 till mid-June 2019). The data collection process was accomplished by using the Twitter free streaming Application Programming Interface (API) based on the recommendation of Sarsam, Al-Samarraie, and Omar (2019). The main keywords used in tweet searching were 'diabetes' and 'diabetes disease'.

3.2 Data pre-processing and text clustering

Several pre-processing techniques were applied to process the collected data for future analysis. Data transformation was applied to the collected tweets in order to obtain a manageable representation. This is due to that most of the popular machine learning algorithms may experience difficulties in processing textual data smoothly. This is why we considered the option of using the bag-of-words (or set-of-words) via n-gram tokenizer method to extract the features of the retrieved data. All the tokens were transformed to a lowercase form before applying the lemmatisation technique. Lemmatisation, in general, uses vocabulary and morphological analysis of word and removes inflectional endings to convert words to a dictionary form (Balakrishnan, Humaidi, & Lloyd-Yemoh, 2016). The stop-words method was applied on the lemmatised words. The length of each tweet was normalised using the L2 norm. The L2 norm was calculated based on the minimisation of the sum of the squares of the residuals, that permits the estimation of the values of unspecified parameters (Inal, Yetkin, Bulbul, & Bilgen, 2018).

After the data pre-processing stage, the K-means clustering algorithm was utilised based on the recommendations of previous works (e.g., Patowary, Sarmah, & Bhattacharyya, 2020; Sarsam & Al-Samarraie, 2018a) to extract information about diabetes disease. This was achieved by categorising the data into *k* groups. K-means calculates the *k* centroid and assigns each point to the relevant cluster centroid (the centroid for each cluster is the point to which the sum of distances from all the instances in that cluster is minimised). In order to determine the number of clusters using K-means algorithm, we used the “Log-likelihood” technique (Witten, Frank, Hall, & Pal, 2016). According to Witten et al. (2016), log-likelihood is a probabilistic clustering approach that can be measured over the training set used by the clustering algorithm. The larger the log-likelihood value, the better the model fit. Here, we implemented a density-based method to each cluster produced by the K-means algorithm. The performance of K-means was examined with different number of centroids (2, 3, 4, and 5). The log-likelihood values were: 35.04 for 2 centroids, 21.94 for 3 centroids, 19.86 for 4 centroids, and 12.73 for 5 centroids. Figure 2 shows the feature values for each cluster (the higher the value is, the darker the colour). After finding the number of groups that share similar features, we extracted the embedded emotions from the two clusters to deduce the features relevant to diabetes disease. In this study, we argued that there is a particular type of emotion (in each group) that is mainly related to diabetes disease. Next section explains the emotion extraction process in details.

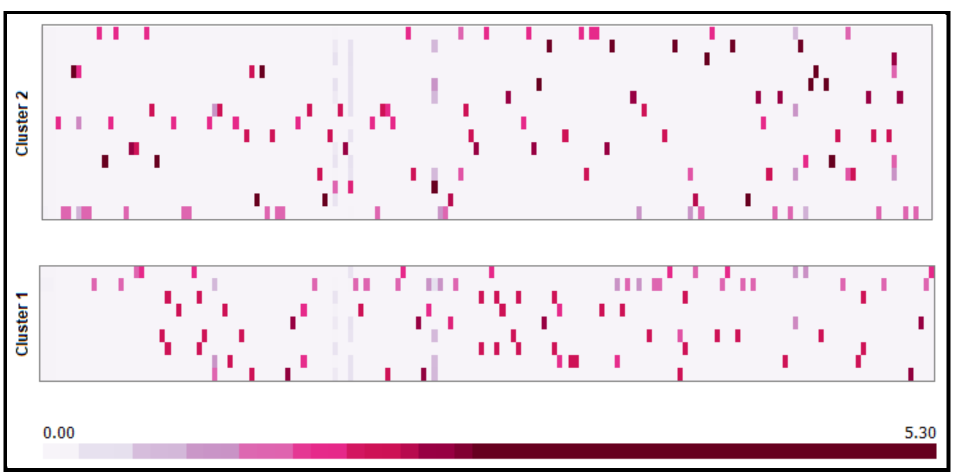

Figure 2: K-means clustering result

3.3 Emotion extraction

Emotions were extracted from the obtained clusters using the AffectiveTweets package. The NRC Affect Intensity Lexicon technique was applied, as recommended by Mohammad (2017), to extract the embedded emotions from the collected tweets. This method consists of a list of English words and their associations that we used to represent four basic emotions (anger, fear, sadness, and joy). For a given word and emotion X, the scores ranged between 0 and 1. A score of 1 means that the word conveys the highest amount of emotion X. A score of 0 means that the word conveys the lowest amount of emotion X. Then, the emotional features for each tweet were calculated by adding the relevant associations of words for a given lexicon.

3.4 Part-of-speech tagging

After finding the dominant emotions in each cluster, the part-of-speech tagging technique was applied to each cluster in order to extract the relevant features. Part-of-speech tagging is commonly used in social media analysis due to its role in identifying words that can be used in different parts of speech (Ritter, Etzioni, & Clark, 2012). In this study, we extracted sentences from each tweet using the Penn State Treebank tokenizer in conjunction with the Document Pre-processor approach. Then, the Penn State Treebank tokenizer was applied again to obtain relevant words before applying the Probabilistic context-free grammar parser. This process enabled us to extract the ‘noun’ words from the sentence that were analysed using the association rules mining technique. The noun words were used to form the terminologies of diabetes. Finally, the relationship between these terminologies and the type of emotions in the two clusters was established using association rules.

3.5 Apriori algorithm

The Apriori algorithm was used in this study to define patterns within a set of items and to make a meaningful relationship between the data features. We configured the Apriori algorithm by setting the delta value at 0.05 in order to reduce the support until a minimum support is reached. The minimum metric score was set at 0.9, while the upper bound and lower bound support were set at 1.0. Then, we invoked Apriori to find the emotional types associated with certain disease-related terminologies.

3.6 Validation stage

The validation process of the unsupervised learning phase was carried out in several steps using the Waikato Environment for Knowledge Analysis (WEKA) tool. First, we set the ground truth by labelling the obtained data with two target labels: 'Yes' (instances related to diabetes) and 'No' (instances not related to diabetes). Second, the prediction validity of the K-means algorithm was also established using the "ClassificationViaClustering" classifier (Witten et al., 2016). This was achieved by assessing the ground truth of the two classes. ClassificationViaClustering is a meta-classifier that allows converting a clustering task into a classification task. In this way, the majority of classes in each cluster can be used in the prediction process. In addition, ClassificationViaClustering was also used to find the minimum error assignment of class labels in which a particular class label can only be assigned to one cluster. The K-means algorithm was applied within the ClassificationViaClustering classifier for four times. The classification results are explained in section 4.2.

4. Results

In each cluster, four types of emotions were extracted at the very early stage: anger, fear, sadness, and joy. Our results showed that among these emotions, only ‘sadness’ and ‘joy’ had a zero value (e.g., tweets with no sadness and joy emotions). The percentage of the ‘fear’ type (92%) in Cluster 1 was higher than Cluster 2 (8%) (see Figure 3a). In contrast, the percentage of the anger emotion category was higher in Cluster 2 (90%) than in Cluster 1 (10%) (see Figure 3b). As a result, only one type of emotions was found to be dominant in each cluster. This led us to label Cluster 1 as ‘Fear group’ and Cluster 2 as ‘Anger group’.

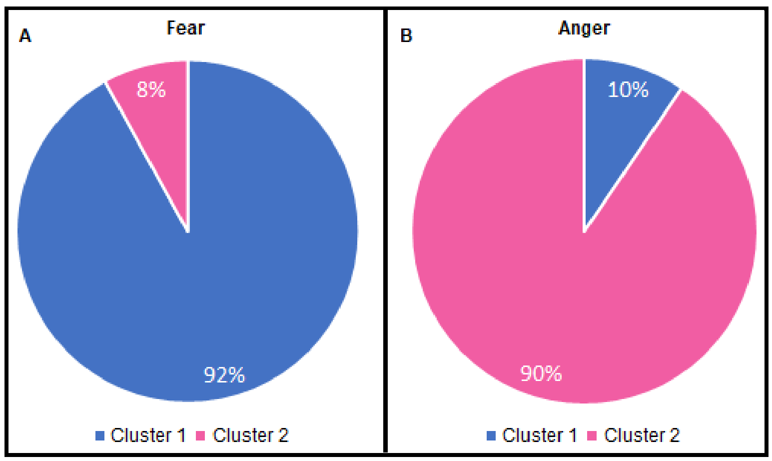

Figure 3: Fear vs. anger emotions in the two clusters

4.1 Results of association rules mining and part-of-speech tagging methods

After recognising the type of emotion in each cluster, the Apriori algorithm was invoked on the data/nouns. The constructed rules were evaluated using both support and confidence scores as shown in Figure 4. Figure 4a shows the highly associated rules (top-ten rules) with the fear emotion in Cluster 1. Figure 4b shows the top-ten words associated with the anger emotion type in Cluster 2. From Figure 4 it can be observed that the terms associated with the fear emotion in Cluster 1 are more relevant to the recognition of diabetes disease than the terms found in Cluster

2 (anger emotion). As such, it can be concluded that the fear emotion can be effectively used to characterise and predict diabetes in Twitter.

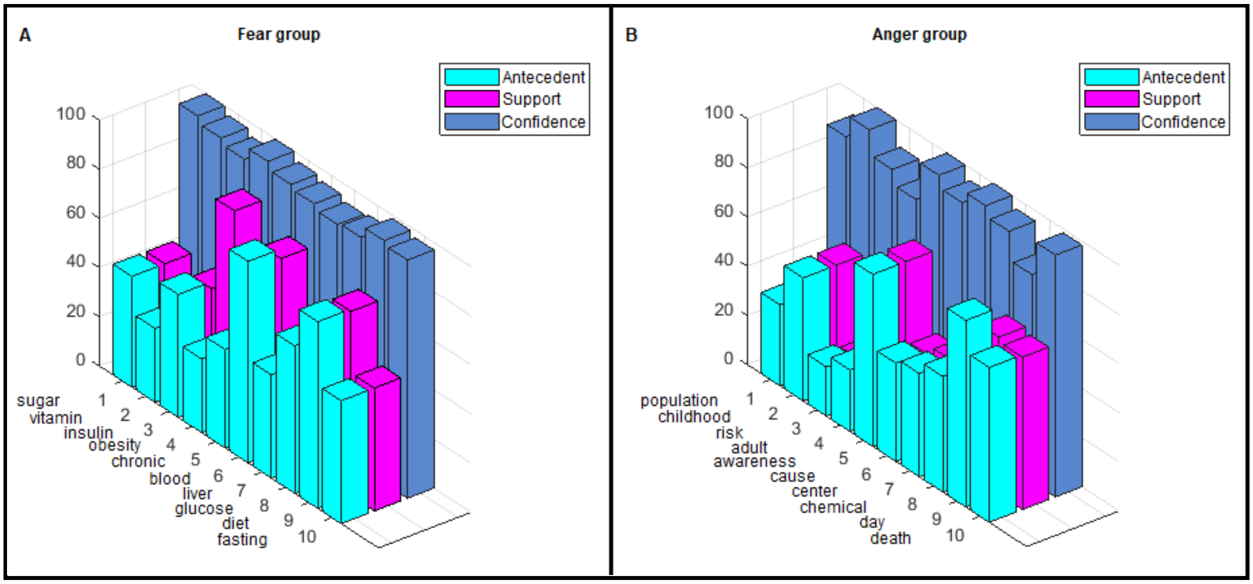

Figure 4: Results from the association rules method

4.2 Validation results

In this study, a specific number of centroids (2, 3, 4, and 5) was used for K-means at each runtime in each classification procedure. A stratified tenfold cross-validation was also used to evaluate the overall learning process of the K-means algorithm. Several evaluation metrics were also used such as Accuracy, Kappa statistic, Root Mean Squared Error (RMSE), and Receiver Operating Characteristic (ROC) to evaluate the performance of the classification result. The classification result (see Table 1) showed that K-means clustering with 2 centroids had the highest classification accuracy (96.53%), followed by 3 centroids (72.54%), 4 centroids (64.68 %), and 5 centroids (54.51 %). In addition, K-means clustering with 2 centroids was found to result in a higher kappa statistic result (95.43%) than 3 centroids (67.27 %), 4 centroids (51.27 %), and 5 centroids (43.57 %), respectively. The results also showed that K-means clustering with 2 centroids had a lower RMSE result (10.03 %). Furthermore, the prediction model (with 2 centroids) had the highest ROC value than prediction models with 2 centroids, 3 centroids, 4 centroids, and 5 centroids, respectively (see Figure 5). In sum, it can be said that K-means clustering with 2 centroids is more efficient in detecting instances related to diabetes.

Table 1: Summary of the evaluation result of K-means clustering with different centroids

Learning algorithm	**Accuracy (%)**	**Kappa statistic (%)**	**RMSE (%)**
2 centroids	96.53	95.43	10.03
3 centroids	72.54	67.27	28.00
4 centroids	64.68	51.27	49.28
5 centroids	54.51	43.57	64.66

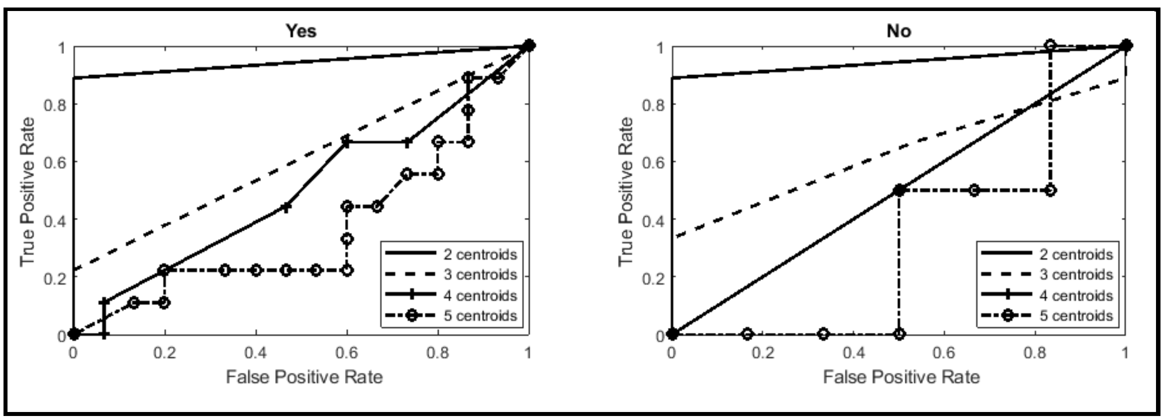

Figure 5: ROC curve result

5. Discussion

This study proposed a novel heuristic approach for detecting diabetes disease based on the characteristics of certain types of emotions embedded within tweets. Our result showed that fear emotions were the most effective in characterising and predicting diabetes disease than all other types of emotions. The results confirmed that predicting diabetes from microblogs can be achieved by building the disease-based emotion lexicon using the terminologies of the disease. Our results are in line with some previous studies that have found strong associations between fear emotions and hypoglycemia in diabetic patients. Hypoglycemia is popular among diabetes patients in which an increase in hypoglycemic level in a person can increase the probability of fear emotion (Willis, Diago-Cabezudo, Madec-Hily, & Aslam, 2013). In addition, severe hypoglycemic event (SHE) is an unpleasant experience, which may involve convulsions, coma and hypothermia. Based on this, any SHE may increase patients' fears of future hypoglycemic episodes (Currie et al., 2006). Such fear places an increased psychologic burden on patients, which is often not shared or discussed with healthcare professionals.

In contrast, previous work found different types of emotions that can be associated with diabetes when investigating the associations between several forms of emotional stress and the development of type 2 diabetes. For example, Pouwer, Kupper, and Adriaanse (2010) stated that not only depression can be correlated with an increased risk of type 2 diabetes, but also general emotional stress and anxiety, sleeping problems, anger, and hostility.

In line with our approach, some previous studies considered emotions embedded in social media texts as key features for machine learning algorithms to perform disease recognition. For example, our results support the work of Ofoghi, Mann, and Verspoor (2016) who suggested that emotions found in Twitter messages may be utilised effectively in the detection and monitoring of a disease outbreak. This includes supporting the use of Ekman's six basic emotions (anger, disgust, happiness, sadness, surprise, and fear) to extract embedded emotions in tweets. Our findings also support the work of Chen, Sykora, Jackson, and Elayan (2018), which reported the potential of emotion in predicting mental health conditions using stories shared/posted on social media websites. The results from this study are in line with the finding of Zuccon et al. (2015) who developed a prediction system to detect the outbreak of Influenza-Like Illness (ILI) through classifying textual features including n-gram, stemmed words, hashtags, URLs, and emoticons.

Nevertheless, the use of emotions for disease outbreak detection were applied not only on Twitter, but also in other microblogs like Sina Microblog. For example, Sun, Ye, and Ren (2014) utilised an unsupervised Bayesian technique based on Markov Network and microblog emotion to recognise early flu-like symptoms. Our work exhibited that the fear emotions can be linked to diabetes. This finding support the conclusion of Gabarron, Dorronzoro, Rivera-Romero, and Wynn (2019) who found that type 2 diabetes tweets are usually accompanied with negative emotions.

6. Implications

To the best of our knowledge, this is the first study to propose a heuristic mechanism (using an unsupervised learning technique) by establishing a relationship between diabetes-related terms and the emotions embedded in tweets. Our approach highly contributes to the health informatics and bioinformatics domains. The proposed mechanism can be treated as an investigative method for characterising a disease state to benefit public health. The detection scenario in this work can benefit health care systems by providing a cost-effective way to characterise and detect early stage of a disease from microblogs. The proposed approach can help assist the decision-making process by providing timely understanding of a disease outbreak. Moreover, our method paves the way for developing advanced surveillance systems to extract critical disease-related information from technologies like electronic health (eHealth) and mobile health (mHealth). For example, the proposed method can effectively contribute to the tracking process of disease progression trajectory, thus promoting disease surveillance and control programs.

The use of clustering techniques in disease discovery can promote the adaptability of healthcare delivery plans by grouping patients according to similarities in their needs. This can also help primary care practices develop multiple health needs-based delivery systems to predict disease spread in real world situations. In addition, the use of emotion lexicon in the prediction of disease in microblogs can help promote sentimental orientation toward disease relations. The proposed method has the ability to identify real-world latent infectious diseases that have not yet been identified by national public health authorities.

7. Limitations and future works

This research has some limitations. First, we examined online diabetes disease, mainly because it is one of the largest global health emergencies. Second, our study focused primarily on English tweets because it is the most common language used in Twitter. Therefore, future studies could incorporate other languages using our proposed approach. Third, we used a limited number of keywords to search for users' tweets. Perhaps other missing tweets could have additional information to characterise the disease within social media posts. Finally, we considered only four types of emotions; in the future, scholars may consider other types of emotions and examine their relations with a particular disease.

8. Conclusion

In this study, we proposed a novel heuristic mechanism to efficiently detect disease incidents and outbreaks from the tweet content. We categorised the types of emotions that are highly linked to diabetes disease and its related terminologies. Emotions (anger, fear, sadness, and joy) and diabetes-related terminologies were extracted using the NRC Affect Intensity Lexicon and a part-of-speech tagging tool. The relationship between diabetes-related terms (in the form of tweets) and emotions were established and assessed using an association rules mining technique. The result showed that diabetes-related terms were exclusively associated with the fear emotions, which is in line with the extant literature on diabetes, and that supports the robustness and effectiveness of the proposed mechanism. The current study offers a novel investigative mechanism for disease recognition and outbreak detection in microblogs which is useful in making informed decisions about a disease state.

References

Arora, M., & Kansal, V. (2019). Character level embedding with deep convolutional neural network for text normalization of unstructured data for twitter sentiment analysis. *Social Network Analysis and Mining*, *9*(1), 12-23.

Asif, M. (2014). The prevention and control the type-2 diabetes by changing lifestyle and dietary pattern. *Journal of education and health promotion, 3*.

Balakrishnan, V., Humaidi, N., & Lloyd-Yemoh, E. (2016). Improving document relevancy using integrated language modeling techniques. *Malaysian Journal of Computer Science, 29*(1), 45-55.

Benton, A., Coppersmith, G., & Dredze, M. (2017). *Ethical research protocols for social media health research.* Paper presented at the Proceedings of the First ACL Workshop on Ethics in Natural Language Processing.

Bravo-Marquez, F., Frank, E., Mohammad, S.M., & Pfahringer, B. (2016). *Determining word-emotion associations from tweets by multi-label classification.* Paper presented at the 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI).

Carvalho, J.P., Rosa, H., Brogueira, G., & Batista, F. (2017). Misnis: An intelligent platform for twitter topic mining. *Expert Systems with Applications*, *89*, 374-388.

Chatzakou, D., & Vakali, A. (2015). Harvesting opinions and emotions from social media textual resources. *IEEE Internet Computing*, *19*(4), 46-50.

Chen, J., Hossain, M.S., & Zhang, H. (2020). Analyzing the sentiment correlation between regular tweets and retweets. *Social Network Analysis and Mining, 10*(1), 13.

Chen, X., Sykora, M.D., Jackson, T.W., & Elayan, S. (2018). *What about mood swings: Identifying depression on twitter with temporal measures of emotions.* Paper presented at the Companion Proceedings of the The Web Conference 2018.

Choubey, D.K., Kumar, P., Tripathi, S., & Kumar, S. (2020). Performance evaluation of classification methods with pca and pso for diabetes. *Network Modeling Analysis in Health Informatics and Bioinformatics, 9*(1), 5.

Currie, C.J., Morgan, C.L., Poole, C.D., Sharplin, P., Lammert, M., & McEwan, P. (2006). Multivariate models of health-related utility and the fear of hypoglycaemia in people with diabetes. *Current medical research and opinion, 22*(8), 1523-1534.

Dadich, A., & Olson, R.E. (2017). How and why emotions matter in interprofessional healthcare. *International Journal of Work Organisation and Emotion, 8*(1), 59-79.

Ferrer, R.A., & Mendes, W.B. (2018). Emotion, health decision making, and health behaviour: Taylor & Francis.

Gabarron, E., Dorronzoro, E., Rivera-Romero, O., & Wynn, R. (2019). Diabetes on twitter: A sentiment analysis. *Journal of diabetes science and technology, 13*(3), 439-444.

Ghaffari, F., Salsali, M., Rahnavard, Z., & Parvizy, S. (2014). Compliance with treatment regimen in women with gestational diabetes: Living with fear. *Iranian journal of nursing and midwifery research, 19*(7 Suppl1), S103.

Giatsoglou, M., Vozalis, M.G., Diamantaras, K., Vakali, A., Sarigiannidis, G., & Chatzisavvas, K.C. (2017). Sentiment analysis leveraging emotions and word embeddings. *Expert Systems with Applications, 69*, 214-224.

Habib, A., Asghar, M.Z., Khan, A., Habib, A., & Khan, A. (2019). False information detection in online content and its role in decision making: A systematic literature review. *Social Network Analysis and Mining, 9*(1), 50.

Hayate, I., Wakamiya, S., & Aramaki, E. (2016). *Forecasting word model: Twitter-based influenza surveillance and prediction.* Paper presented at the Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers.

Inal, C., Yetkin, M., Bulbul, S., & Bilgen, B. (2018). Comparison of l1 norm and l2 norm minimisation methods in trigonometric levelling networks. *Tehnički vjesnik, 25*(Supplement 1), 216-221.

Ji, X., Chun, S.A., Wei, Z., & Geller, J. (2015). Twitter sentiment classification for measuring public health concerns. *Social Network Analysis and Mining, 5*(1), 13.

Lee, K., Agrawal, A., & Choudhary, A. (2013). *Real-time disease surveillance using twitter data: Demonstration on flu and cancer.* Paper presented at the Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining.

Lu, Y., Wu, Y., Liu, J., Li, J., & Zhang, P. (2017). Understanding health care social media use from different stakeholder perspectives: A content analysis of an online health community. *Journal of medical Internet research, 19*(4), e109.

Mannix, S., Skalicky, A., Buse, D.C., Desai, P., Sapra, S., Ortmeier, B., . . . Hareendran, A. (2016). Measuring the impact of migraine for evaluating outcomes of preventive treatments for migraine headaches. *Health and Quality of Life Outcomes, 14*(1), 143.

Mejova, Y., Weber, I., & Macy, M.W. (2015). *Twitter: A digital socioscope*: Cambridge University Press.

Mohammad, S.M. (2017). Word affect intensities. *arXiv preprint arXiv:1704.08798*.

Molaei, S., Khansari, M., Veisi, H., & Salehi, M. (2019). Predicting the spread of influenza epidemics by analyzing twitter messages. *Health and Technology*, 1-16.

Motamedi, R., Jamshidi, S., Rejaie, R., & Willinger, W. (2020). Examining the evolution of the twitter elite network. *Social Network Analysis and Mining, 10*(1), 1.

Nejad, M.Y., Delghandi, M.S., Bali, A.O., & Hosseinzadeh, M. (2020). Using twitter to raise the profile of childhood cancer awareness month. *Network Modeling Analysis in Health Informatics and Bioinformatics, 9*(1), 3.

Ofoghi, B., Mann, M., & Verspoor, K. (2016). *Towards early discovery of salient health threats: A social media emotion classification technique.* Paper presented at the Biocomputing 2016: Proceedings of the Pacific Symposium.

Patowary, P., Sarmah, R., & Bhattacharyya, D.K. (2020). Developing an effective biclustering technique using an enhanced proximity measure. *Network Modeling Analysis in Health Informatics and Bioinformatics, 9*(1), 1-17.

Potegal, M., Stemmler, G., & Spielberger, C. (2010). *International handbook of anger: Constituent and concomitant biological, psychological, and social processes*: Springer.

Pouwer, F., Kupper, N., & Adriaanse, M.C. (2010). Does emotional stress cause type 2 diabetes mellitus? A review from the european depression in diabetes (edid) research consortium. *Discovery medicine, 9*(45), 112-118.

Priyadarshi, A., & Saha, S.K. (2020). Web information extraction for finding remedy based on a patient-authored text: A study on homeopathy. *Network Modeling Analysis in Health Informatics and Bioinformatics, 9*(1), 1-12.

Rani, V.V. (2018). *Efficient tool for diabetes tracking through layered bigram approach.* Paper presented at the 2018 IADS International Conference on Computing, Communications & Data Engineering (CCODE).

Ritter, A., Etzioni, O., & Clark, S. (2012). *Open domain event extraction from twitter.* Paper presented at the Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining.

Sarsam, S.M., & Al-Samarraie, H. (2018a). A first look at the effectiveness of personality dimensions in promoting users' satisfaction with the system. *SAGE Open*, *8*(2), 2158244018769125.

Sarsam, S.M., & Al-Samarraie, H. (2018b). Towards incorporating personality into the design of an interface: A method for facilitating users' interaction with the display. *User Modeling and User-Adapted Interaction*, *28*(1), 75-96.

Sarsam, S.M., Al-Samarraie, H., Ismail, N., Zaqout, F., & Wright, B. (2020a). A real-time biosurveillance mechanism for early-stage disease detection from microblogs: A case study of interconnection between emotional and climatic factors related to migraine disease. *NetMAHIB*, *9*(1), 32.

Sarsam, S.M., Al-Samarraie, H., Ismail, N., Zaqout, F., & Wright, B. (2020b). A real-time biosurveillance mechanism for early-stage disease detection from microblogs: A case study of interconnection between emotional and climatic factors related to migraine disease. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *9*(1), 32. doi: 10.1007/s13721-020-00239-6

Sarsam, S.M., Al-Samarraie, H., & Omar, B. (2019). *Geo-spatial-based emotions: A mechanism for event detection in microblogs.* Paper presented at the Proceedings of the 2019 8th International Conference on Software and Computer Applications.

Schwartz, H.A., Giorgi, S., Kern, M.L., Park, G., Sap, M., Labarthe, D.R., . . . Ungar, L.H. (2018). More evidence that twitter language predicts heart disease: A response and replication.

Sharma, P., & Kaur, P.D. (2017). Effectiveness of web-based social sensing in health information dissemination—a review. *Telematics and Informatics, 34*(1), 194-219.

St Louis, C., & Zorlu, G. (2012). Can twitter predict disease outbreaks? *Bmj, 344*, e2353.

Sun, X., Ye, J., & Ren, F. (2014). *Real time early-stage influenza detection with emotion factors from sina microblog.* Paper presented at the Proceedings of the Fifth Workshop on South and Southeast Asian Natural Language Processing.

Tamarai, K., Bhatti, J. S., & Reddy, P. H. (2019). Molecular and cellular bases of diabetes: Focus on type 2 diabetes mouse model-TallyHo. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1865*(9), 2276-2284.

Wakamiya, S., Kawai, Y., & Aramaki, E. (2018). Twitter-based influenza detection after flu peak via tweets with indirect information: Text mining study. *JMIR public health and surveillance*, *4*(3), e65.

Wang, W., Chen, L., Thirunarayan, K., & Sheth, A.P. (2012). *Harnessing twitter" big data" for automatic emotion identification.* Paper presented at the 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing.

Willis, W.D., Diago-Cabezudo, J.I., Madec-Hily, A., & Aslam, A. (2013). Medical resource use, disturbance of daily life and burden of hypoglycemia in insulin-treated patients with diabetes: Results from a european online survey. *Expert review of pharmacoeconomics & outcomes research, 13*(1), 123-130.

Wilt, T., Kansagara, D., Horwitch, C., & Barry, M. (2018). Hemoglobin a1c targets for nonpregnant adults with type 2 diabetes mellitus.

Witten, I.H., Frank, E., Hall, M.A., & Pal, C.J. (2016). *Data mining: Practical machine learning tools and techniques*: Morgan Kaufmann.

Xu, S., Markson, C., Costello, K.L., Xing, C.Y., Demissie, K., & Llanos, A.A. (2016). Leveraging social media to promote public health knowledge: Example of cancer awareness via twitter. *JMIR public health and surveillance*, *2*(1), e17.

Yadav, Yadav, & Sapkota. (2017). Anger; its impact on human body.

Yavary, A., Sajedi, H., & Abadeh, M.S. (2020). Information verification in social networks based on user feedback and news agencies. *Social Network Analysis and Mining, 10*(1), 2.

Zhou, L., Zhang, D., Yang, C.C., & Wang, Y. (2018). Harnessing social media for health information management. *Electronic commerce research and applications*, *27*, 139-151.

Zuccon, G., Khanna, S., Nguyen, A., Boyle, J., Hamlet, M., & Cameron, M. (2015). Automatic detection of tweets reporting cases of influenza like illnesses in australia. *Health information science and systems*, *3*(S1), S4.