



UNIVERSITY OF LEEDS

This is a repository copy of *A lexicon-based method for detecting eye diseases on microblogs*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/179650/>

Version: Accepted Version

Article:

Sarsam, SM and Al-Samarraie, H orcid.org/0000-0002-9861-8989 (2022) A lexicon-based method for detecting eye diseases on microblogs. *Applied Artificial Intelligence*, 36 (1). 1993003. pp. 1-12. ISSN 0883-9514

<https://doi.org/10.1080/08839514.2021.1993003>

© 2021 Taylor & Francis. This is an author produced version of an article published in *Applied Artificial Intelligence*. Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

A lexicon-based method for detecting eye diseases on microblogs

Samer Muthana Sarsam¹ (Corresponding author), Hosam Al-Samarraie^{2&3}

¹ Department of Business Analytics, Sunway University Business School, Sunway University,

5 Selangor, Malaysia, samers@sunway.edu.my

² School of Design, University of Leeds, Leeds, UK

³ Centre for Instructional Technology and Multimedia, Universiti Sains Malaysia, Penang, Malaysia

Abstract

10 This paper explored the feasibility of detecting eye diseases on microblogs. A
lexicon-based approach was developed to provide an early recognition of common
eye disease from social media platforms. The data were obtained using Twitter free
streaming Application Programming Interface (API). A cluster analysis was
applied to extract instances that share similar characteristics. We extracted three
15 types of emotions (positive, negative, and neutral) from users' messages (tweets)
using SentiStrength. A time-series method was used to determine the applicability
of predicting emotional changes over a period of seven months. The relevant
disease symptoms were extracted using Apriori algorithm with prediction accuracy
of 98.89 %. This study offers a timely and effective method that can be

20 implemented to help healthcare decision makers and researchers reduce the spread
of eye diseases in a population specific manner.

Keywords: social media mining; eye disease; lexicon-based approach; time series

1. Introduction

25 Social media mining has been widely applied in many sectors, such as
commerce, healthcare, and education, to help decision makers understand how to
transform data into knowledge for decision support (Sarsam, Al-Samarraie,
Alzahrani, et al., 2021; Sarsam, Al-Samarraie, Alzahrani, et al., 2020). The use of
social media sites (e.g., Twitter and Facebook) has been found to offer an effective
30 way for studying users' emotional experiences. Such experiences have been found
to be important for detecting various events. Health-related topics published on
social media sites have been extensively regarded as a rich resource for researchers
seeking to understand health changes in a population over time (Ru & Yao, 2019).
An example would be glaucoma which is one of the leading causes of blindness
35 globally (Ban et al., 2018). Common eye disease such as cataract and glaucoma are
two conditions that affect the neuroretinal rim of the nerve (Maetschke et al., 2019).
However, the association between eye disease and users' emotions is not fully
studied. One reason to this can be linked to the difficulty in mapping the structural
and functional abnormalities when a person experience a range of eye conditions

40 (Dervisevic et al., 2016). This has motivated the need for developing mechanisms
capable of detecting various eye disease. For example, Mehta et al. (2021)
developed a multimodal model to automate glaucoma detection. The authors
trained a multimodal model with multiple deep neural nets (trained on macular
optical coherence tomography volumes and color fundus photographs) based on
45 demographic and clinical data. The detection results showed an accuracy of 97%.
Kim et al. (2017) introduced a machine learning approach capable at diagnosing
glaucoma by using several feature selection techniques. The authors found that
random forest achieved the highest performance (98%) for the detection of
glaucoma. Another study by Ahn et al. (2018) used a deep learning technique to
50 diagnose glaucoma using 1542 images. These images were used to build simple
logistic classification and convolutional neural network that achieved a
performance of 99%. According to Stefan et al. (2020), different machine learning,
deep learning and transfer learning techniques are likely to offer different solutions
for retinal image analysis. Despite these studies, there is a limited use of textual
55 information shared in social media sites for the diagnosis of glaucoma disease. This
study examined the potential of using time series and association rules mining in
the detection of common eye disease from social media texts. We used Twitter to
extract and analyze users' sentimental features in an attempt to categorize disease-
related emotions.

60 **2. Literature review**

Online sentiments in the form of tweets or messages can provide valuable information about users' state. The changes in users' sentiments and emotions over time have been used by many studies to investigate various emergent phenomena (Su et al., 2017). This motivated many researchers to study changes in user
65 sentiments in an attempt to describe various health conditions and needs (Denecke & Deng, 2015). Public's opinion/reaction to personal circumstances can be mapped with the help of sentiment analysis. The Twitter platform is once source of data that contain emotions represented in a text form. The texts/tweets are used to explain users' experiences and their opinions towards certain events and topics (Kessler &
70 Schmidt-Weitmann, 2021). However, in order to make a good use of online opinions shared between users on social media sites, a number of analytical procedures should be carried out (Chen et al., 2020). In a healthcare context, analyzing users' sentiments is a complex process which requires a lot of information to be retrieved and annotated (Ahmet & Abdullah, 2020). This has
75 motivated many scholars to examine health-related content available on Twitter in an attempt to support healthcare decision-makers. For example, a study by Sarsam, Al-Samarraie and Al-Sadi (2020) used sentiment analysis technique to analyze social media content in the detection of diabetes using a set of machine learning models. The authors developed a heuristic mechanism to efficiently detect diabetes-

80 related incidents and outbreaks on Twitter. The result showed that certain terminologies were highly linked to users' emotions. Along the same line, another study by Sarsam, Al-Samarraie, Ismail, et al. (2020) examined the possibility of mapping certain emotional types and climatic factors in the detection of migraine disease. The authors demonstrated the potential of using the Twitter platform as a reliable real-time data source to aid in early-stage disease recognition. The literature (e.g., Sarsam, Al-Samarraie, Alzahrani, et al., 2021) also showed the role of sentiments, in the form of tweets, in detecting suicide-related contents. The authors used a complex analytical procedure to extract and characterize six types of emotions: anger, fear, sadness, joy, positive, and negative. They used a semi-90 supervised learning approach via the YATSI (Yet Another Two-Stage Idea) classifier to predict suicide-related tweets through the identification of certain emotions such as fear, sadness, and negative.

Based on these observations, it can be concluded that the use of social media platform can be efficient in characterizing users' emotions related to their health conditions. This can help us develop a robust mechanism for diagnosing95 diseases at the very early stage. We develop a lexicon-based approach to provide an effective means for extracting emotions related to common eye disease. We used time series to determine the applicability of predicting emotional changes over a specific period of time. Both NRC Affect Intensity Lexicon and SentiStrength were

100 used to cluster users' emotions. It is believed that this method can help provide
robust and more efficient estimates of common eye diseases on Twitter.

3. Proposed method

Our method and procedure are summarized in Figure 1. This procedure
105 consists of data collection, data pre-processing, cluster analysis, emotion
extraction, association rule mining and time series, and disease detection. These
stages are explained in the following sections.

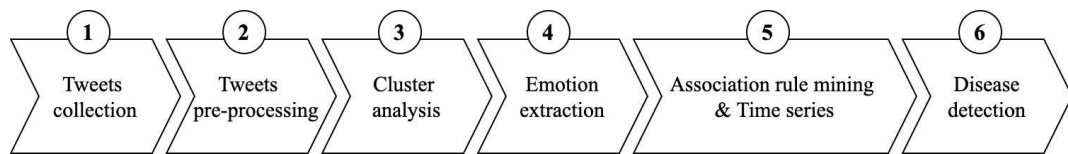


Figure 1: The study procedure

110

3.1 Data collection

A total of 2,448,250 English tweets were collected within a time span of
seven months (October 2020 till April 2021). We collected messages in relation to
the common eye diseases using the Twitter free streaming Application
115 Programming Interface (API). The search keywords consisted of: Refractive
Errors; Age-Related Macular Degeneration; Cataract; Diabetic Retinopathy;

Glaucoma; Amblyopia; and Strabismus. Then, the data pre-processing stage was conducted to produce reliable data for the analytical stage.

120 ***3.2 Data pre-processing***

In this study, several pre-processing methods were used to provide the necessary resources for the recognition of various eye diseases. We applied the bag-of-words model to extract words associated to a sentiment level from the collected tweets. After that, all the extracted words were converted to a lowercase form
125 before deleting the unnecessary words via the Stopwrods list technique. Lastly, the length of the tweets was normalized using the L2 norm.

3.3 Cluster analysis

We performed cluster analysis to obtain semantic labels for the collected
130 queries. We clustered the data using a density-based method via the K-means algorithm by fitting normal distributions and discrete distributions within each produced cluster. Finally, two ophthalmologists (with 14 years of experience) were involved to assess these terminologies together with their relation to various eye diseases. The embedded emotions were extracted from each cluster using the
135 lexicon-based method.

3.4 Emotion extraction

In order to extract users' emotion from their tweets, lexicon-based method was implanted via SentiStrength (Khaira et al., 2020; Sarsam, Al-Samarraie, Bahar, et al., 2021). In this sense, scores were assigned to each tweet in which the score ranges from '+1' for 'not positive' to '+5' for 'extremely positive' and '-1' for 'not negative' to '-5' for 'extremely negative'. Based on these scores, we labeled the tweets with +5 as 'Positive' tweets, -5 as 'Negative' tweets, and -1/+1 as 'Neutral' tweets (Culpeper et al., 2018; Thelwall, 2017). After that, time series prediction was implemented in each cluster to examine periodical changes in users' emotions.

3.5 Association rule mining and time series

To extract the emotional features in each tweet (i.e., words), we used the Apriori algorithm to define patterns within a set of items and establish a meaningful relationship between the data features. We set the delta value of the Apriori algorithm at 0.05 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. After that, the emotional changes during the data collection time span were captured and analyzed using the time series technique.

3.6 Disease detection

To detect the glaucoma using the extracted emotional features, three classification algorithms were compared to find the best detection algorithm. These were: Deep learning algorithm (Witten et al., 2016), sequential minimal optimisation (SMO) or support vector machine (SVM) (Al-Ajeli et al., 2020), and Bagging classifier (Kabiraj et al., 2020; Sarsam, Al-Samarraie, & Alzahrani, 2021). These algorithms were implemented using Weka software (Waikato Environment for Knowledge Analysis) using settings similar to previous experiments (Al-Samarraie et al., 2018; Al-Samarraie et al., 2016). The stratified tenfold cross-validation method was applied to assess the learning process of the three algorithms. Finally, the best algorithm was selected using several evaluation metrics, including Accuracy, Kappa statistic, and Confusion matrix, as per the recommendations of previous studies (Al-Samarraie et al., 2016; Sarsam, Al-Samarraie, Alzahrani, et al., 2021).

4. Results

4.1 Cluster analysis and emotion extraction results

The results for the k-means showed three different clusters (groups) with instances in each group sharing similar characteristics. To label each cluster, two ophthalmologists named the first, second, and third cluster as ‘Symptoms’,

'Lifestyle', and 'Advice', respectively. Figure 2 exhibits the result of SentiStrength (positive, negative, and neutral) in each of the three clusters where negative emotions were found to be dominant over (with highest score of 91 %) the positive (5 %) and neutral (4 %) emotions. In second cluster (lifestyle), positive emotions were found to be the dominant type (score of 94 %) as compared to the negative (3 %) and neutral (3 %) emotions. In the third cluster (advise), neutral emotions were found to be dominant (96 %) over the positive (2 %) and negative (2 %) emotions. From these, it can be said that identifying the polarity of emotions in the tweets can be useful in labeling training samples.

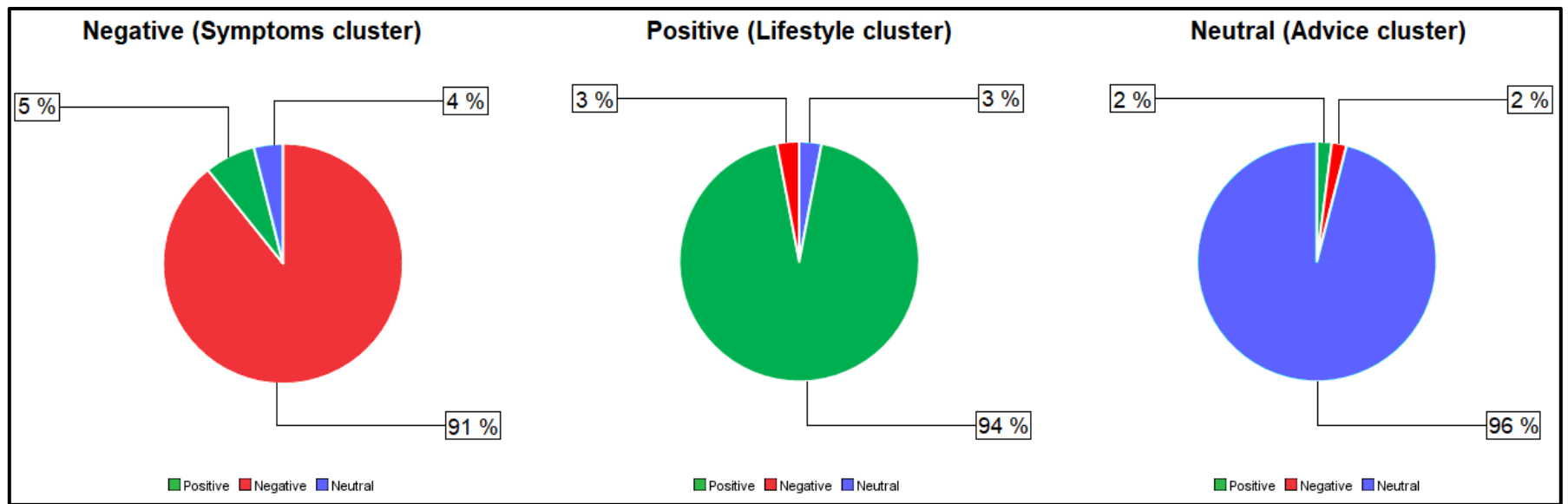


Figure 2: SentiStrength results

4.2 Association rule mining and time series results

190 Figure 3 shows the highly associated rules in each cluster solution. In this sense, Figure 3a outlines the rules that are highly associated with the main symptoms of eye diseases in the form of eye pain, nausea, blurred vision, headache, and vomiting. We can also note that some of the rules were highly associated with the patients' lifestyle. An example is the type of food consumed
195 by the affected users such as eggs, carrots, milk, and broccoli. The figure also reveals another set of rules that were highly associated with users seeking information (advise) about their eye disease such as information about the types of eye screening and eye drops for glaucoma.

To further understand users' emotional reactions in each cluster,
200 emotional changes were extracted together with their temporal features, which were analyzed using time series in the period of seven months (see Figure 3b). From the figure, it can be observed that negative emotions were the only type that faced a substantial increase in number of tweets comparing to other types of emotions. The increasing number of negative emotions was located in the
205 cluster that contained symptoms of eye disease.

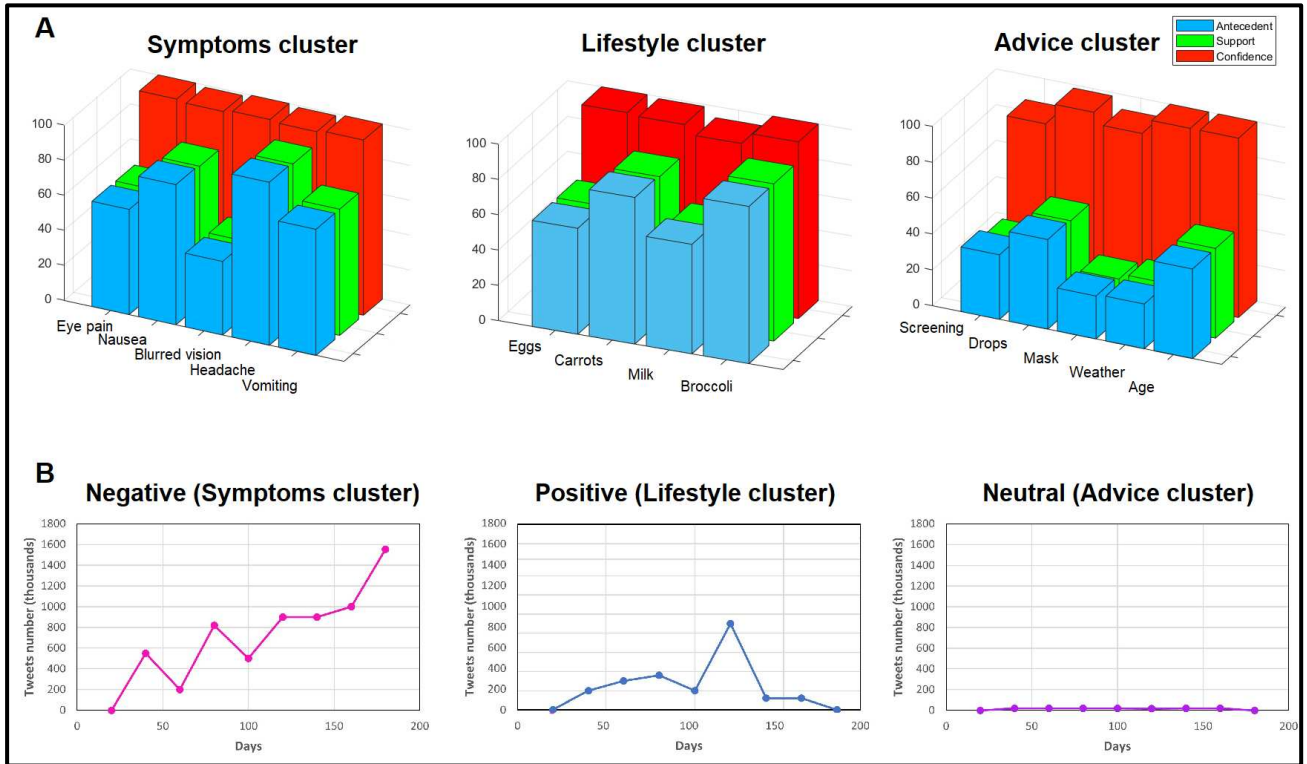


Figure 3: Apriori algorithm and time series results

210 **4.3 Disease detection results**

Our classification results are summarized in Table 1 which shows that deep learning algorithm achieved the highest classification accuracy (98.89 %), followed by SMO (65.13 %), and Bagging (57.43 %). The classification results also revealed that deep learning had the highest kappa statistic value (97 %) compared to SMO (73 %), and Bagging (60 %), respectively. In contrast, the Bagging classifier

215

produced the highest RMSE value (89 %) followed by SMO (45 %), and deep learning (3%).

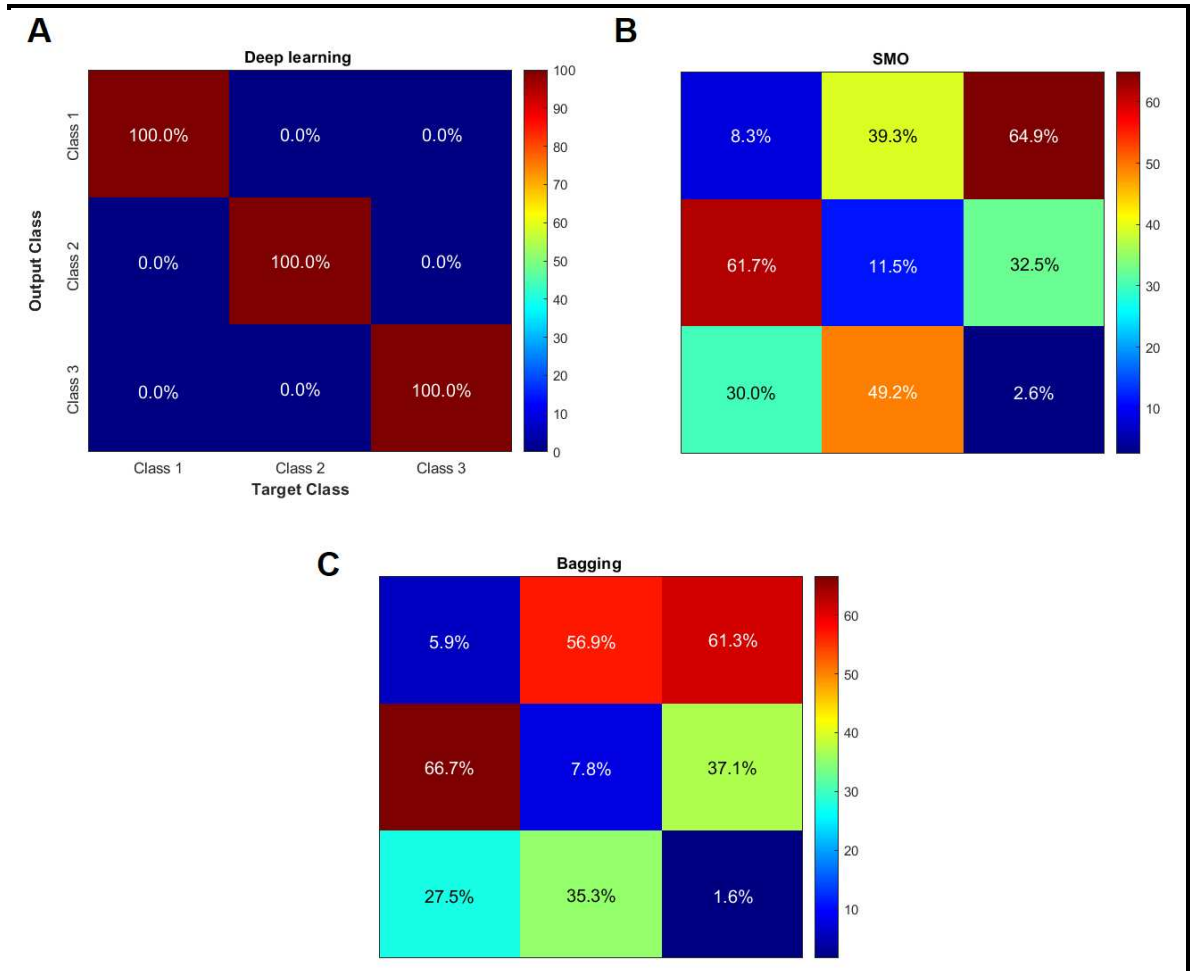
Table 1: Classification results

Learning algorithm	Accuracy (%)	Kappa statistic (%)	RMSE (%)
<i>Deep learning</i>	98.89	97	3
<i>SMO</i>	65.13	73	45
<i>Bagging</i>	57.43	60	89

220

In addition to the previous evaluation metrics, we also used the “Confusion matrix” approach to evaluate our algorithms by measuring the relationship between the predicted and the actual instances when demonstrating instances along the diagonal of the confusion matrix. Our results (see Figure 4) revealed that Deep learning classifier had the highest predictive capability between actual and predicted classes, i.e., 100 %, for the three classes.

225



230 Figure 4: Confusion matrix results

235 **5. Discussions**

The proposed lexicon-based approach in this study showed promising results in detecting common eye diseases. We found that certain types of emotions can be linked to certain eye conditions. The results obtained from NRC Affect Intensity Lexicon and SentiStrength revealed that negative, fear, and sadness
240 sentiments were strongly associated with symptoms of eye diseases. It is believed that users' sharing of personal stories about their eye conditions on social media sites is driven by the need for information and support (Zhang & Zhou, 2020). In addition, people in general are interested in similar stories submitted by other users, which can help them make informed decisions about their symptoms (Moorhead et
245 al., 2013). The findings support the work of Jampel et al. (2007), who reported that glaucoma patients are likely to express fear-related emotions in relation to vision loss. Ciuraru (2016) stated that concerns about losing vision from glaucoma or other similar eye disease can increase people's anxiety and depression, which explains the negative sentiment that were found in our results. Based on this, it can
250 be said that patients who exhibit fear of becoming blinds are likely to share negative emotions with others (Dorison et al., 2020). Our findings are also in line with the work of Stamatiou et al. (2021) who confirmed the associations between glaucoma patients and the negative sentiments which were characterized in the form of depression.

255 The main topics shared between users on Twitter were mostly related to the symptoms of eye diseases. Several eye conditions were posted by social media users in relation to their eye disease such as dry eye, temporary loss of vision, eye pain, nausea, blurred vision, and halos. Other topics shared between social media users were related to the type of food consumed by people with certain eye diseases.

260 The results also showed that another group of people were mostly interested in sharing topics related to eye disease prevention and treatment advice. We believe that these topics are essential in characterizing the types of emotions people express on social media sites. Our review of the literature showed that previous studies have frequently discussed certain eye symptoms (e.g., dry eye, head pain, and nausea) in

265 relation to people experiencing early eye disease signs (Bartlett et al., 2015). The results also showed that people with certain eye conditions were very interested in learning about the types of food to take or to avoid when taking certain medications. This finding is in line with previous studies, which have shown that certain types of food and fruits can be effective in reducing intra ocular pressure in patients, as

270 well as possessing an analgesic effect that is relative to the powerful opioid morphine, aspirin, and indomethacin (Rasmussen & Johnson, 2013). Also, the consumption of fruits and vegetables rich in vitamin A, vitamin C, and carotenoids has been reported to help reduce the risk of glaucoma (Jabbehdari et al., 2020). In addition, the consumption of blackcurrants was found to be helpful in reducing the

275 eye fatigue among patients (Smeriglio et al., 2016). Patients were also found to seek
advice about certain symptoms of eye diseases and treatment or preventive
solutions. This information seeking pattern is probably the most common of all the
patterns identified in this study. According to Smailhodzic et al. (2016), six groups
of patients use social media: emotional, information, esteem, network support,
280 social comparison and emotional expression. These groups use social media sites
for improved self-management and control, enhanced psychological well-being,
and enhanced subjective well-being, diminished subjective well-being, addiction to
social media, loss of privacy, and being targeted for promotion. As a result, it can
be said that users' queries about certain eye treatments or conditions can be
285 effective in the prediction of eye disease. From the findings of this study, one can
observe that the discussed topics on Twitter were highly associated with the
literature on various eye diseases.

6. Implications

290 This study provides a first step for the use of sentiment analysis in
extracting and characterizing eye diseases from Twitter. The use of emotions in the
form of tweets was found to be useful in predicting the common symptoms of eye
diseases. The proposed method contributes to the development of clinical decision
support systems that aim at examining changes in users' vision based on certain

295 sentimental features embedded in their social media posts. The proposed
mechanism is cost-effective and can be effectively used in the detection of various
eye diseases. Meanwhile, the proposed mechanism can be used not only for
detecting specific eye diseases, but also for other diseases/conditions. The use of
time series and Apriori algorithm offers an additional way of building associations
300 between tweets/emotions and should be used to complement traditional detection
methods.

7. Limitations and future works

This study has several limitations. For example, we only used English
305 tweets in the analysis and detection of eye diseases. The collection period of the
data was limited to seven months in which three cluster solution were identified
and evaluated. This study was also limited to certain types of emotions found in the
collected tweets. The prediction process was limited to the use of Apriori algorithm
to establish associations between the common symptoms of eye diseases and
310 emotions. In addition, this study used and compared three classifiers for the
prediction of eye diseases. Although the utilized classifiers were recommended by
previous studies, other more complex classifiers can be considered in the future.
Future work can also consider the use of different social media platforms to collect
and retrieve different types of emotions capable of representing a larger population.

315 Researchers are also encouraged to focus on a specific type of eye diseases in order
to confirm the efficiency of our proposed approach. Finally, future studies are also
invited to adopt our technique to diagnose other types of diseases on social media
platforms.

320 **8. Conclusion**

This study proposed a lexicon-based approach the detection of eye diseases
using tweets. We applied a cluster analysis to extract instances that share similar
characteristics. Three types of emotions were extracted (positive, negative, and
neutral) from users' messages (tweets) using SentiStrength. Changes in users'
325 emotions were analyzed using time series. In addition, Apriori algorithm was used
to extract the common symptoms of eye diseases from the tweets. The results
showed that deep learning can be effective in detecting eye disease with an accuracy
of 98.89 %. The proposed approach can contribute to the development of intelligent
health monitoring systems and clinical decision-support systems.

330

Conflict of interest statement: The author declare that no conflict of interest
exists.

335 **References**

- Ahmet, A., & Abdullah, T. (2020). Recent trends and advances in deep learning-based sentiment analysis. In *Deep Learning-Based Approaches for Sentiment Analysis* (pp. 33-56). Springer.
- Ahn, J. M., Kim, S., Ahn, K.-S., Cho, S.-H., Lee, K. B., & Kim, U. S. (2018). A
340 deep learning model for the detection of both advanced and early glaucoma using fundus photography. *PloS one*, *13*(11), e0207982.
- Al-Ajeli, A., Alubady, R., & Al-Shamery, E. S. (2020). Improving spam email detection using hybrid feature selection and sequential minimal optimization. *Indonesian Journal of Electrical Engineering and Computer
345 Science*, *19*(1), 535-542.
- Al-Samarraie, H., Sarsam, S. M., Alzahrani, A. I., & Alalwan, N. (2018). Personality and individual differences: the potential of using preferences for visual stimuli to predict the Big Five traits. *Cognition, Technology & Work*, *20*(3), 337-349.
- 350 Al-Samarraie, H., Sarsam, S. M., & Guesgen, H. (2016). Predicting user preferences of environment design: A perceptual mechanism of user interface customisation. *Behaviour & Information Technology*, *35*(8), 644-653.

- Ban, N., Siegfried, C. J., & Apte, R. S. (2018). Monitoring neurodegeneration in
355 glaucoma: therapeutic implications. *Trends in molecular medicine*, 24(1),
7-17.
- Bartlett, J. D., Keith, M. S., Sudharshan, L., & Snedecor, S. J. (2015). Associations
between signs and symptoms of dry eye disease: a systematic review.
Clinical ophthalmology (Auckland, NZ), 9, 1719.
- 360 Chen, J., Hossain, M. S., & Zhang, H. (2020). Analyzing the sentiment correlation
between regular tweets and retweets. *Social network analysis and mining*,
10(1), 1-9.
- Ciuraru. (2016). anxiety and defense mechanisms of people diagnosed with
glaucoma.
- 365 Culpeper, J., Findlay, A., Cortese, B., & Thelwall, M. (2018). Measuring emotional
temperatures in Shakespeare's drama. *English Text Construction*, 11(1),
10-37.
- Denecke, K., & Deng, Y. (2015). Sentiment analysis in medical settings: New
opportunities and challenges. *Artificial intelligence in medicine*, 64(1), 17-
370 27.
- Dervisevic, E., Pavljasevic, S., Dervisevic, A., & Kasumovic, S. S. (2016).
Challenges in early glaucoma detection. *Medical Archives*, 70(3), 203.

- Dorison, C. A., Wang, K., Rees, V. W., Kawachi, I., Ericson, K. M., & Lerner, J. S. (2020). Sadness, but not all negative emotions, heightens addictive substance use. *Proceedings of the National Academy of Sciences*, *117*(2), 943-949.
- 375
- Jabbehdari, S., Chen, J. L., & Vajaranant, T. S. (2020). Effect of dietary modification and antioxidant supplementation on intraocular pressure and open-angle glaucoma. *European Journal of Ophthalmology*, 1120672120960337.
- 380
- Jampel, H. D., Frick, K. D., Janz, N. K., Wren, P. A., Musch, D. C., Rimal, R., Lichter, P. R., & Group, C. S. (2007). Depression and mood indicators in newly diagnosed glaucoma patients. *American journal of ophthalmology*, *144*(2), 238-244. e231.
- 385
- Kabiraj, S., Akter, L., Raihan, M., Diba, N. J., Podder, E., & Hassan, M. M. (2020). Prediction of Recurrence and Non-recurrence Events of Breast Cancer using Bagging Algorithm. 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT),
- Kessler, S. H., & Schmidt-Weitmann, S. (2021). Diseases and emotions: An automated content analysis of health narratives in inquiries to an online health consultation service. *Health communication*, *36*(2), 226-235.
- 390

- Khaira, U., Johanda, R., Utomo, P. E. P., & Suratno, T. (2020). Sentiment Analysis Of Cyberbullying On Twitter Using SentiStrength. *Indonesian Journal of Artificial Intelligence and Data Mining*, 3(1), 21-27.
- 395 Kim, S. J., Cho, K. J., & Oh, S. (2017). Development of machine learning models for diagnosis of glaucoma. *PloS one*, 12(5), e0177726.
- Maetschke, S., Antony, B., Ishikawa, H., Wollstein, G., Schuman, J., & Garnavi, R. (2019). A feature agnostic approach for glaucoma detection in OCT volumes. *PloS one*, 14(7), e0219126.
- 400 Mehta, P., Petersen, C. A., Wen, J. C., Banitt, M. R., Chen, P. P., Bojikian, K. D., Egan, C., Lee, S.-I., Balazinska, M., & Lee, A. Y. (2021). Automated detection of glaucoma with interpretable machine learning using clinical data and multi-modal retinal images. *American Journal of Ophthalmology*.
- Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A., & Hoving, C. (2013). A new dimension of health care: systematic review of the uses, 405 benefits, and limitations of social media for health communication. *Journal of medical Internet research*, 15(4), e1933.
- Rasmussen, H. M., & Johnson, E. J. (2013). Nutrients for the aging eye. *Clinical interventions in aging*, 8, 741.
- 410 Ru, B., & Yao, L. (2019). A literature review of social media-based data mining for health outcomes research. *Social Web and Health Research*, 1-14.

- 415 Sarsam, S. M., Al-Samarraie, H., & Al-Sadi, A. (2020). Disease discovery-based emotion lexicon: a heuristic approach to characterise sicknesses in microblogs. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 9(1), 1-10.
- Sarsam, S. M., Al-Samarraie, H., & Alzahrani, A. I. (2021). Influence of personality traits on users' viewing behaviour. *Journal of Information Science*, 0165551521998051.
- 420 Sarsam, S. M., Al-Samarraie, H., Alzahrani, A. I., Alnumay, W., & Smith, A. P. (2021). A lexicon-based approach to detecting suicide-related messages on Twitter. *Biomedical Signal Processing and Control*, 65, 102355.
- Sarsam, S. M., Al-Samarraie, H., Alzahrani, A. I., & Wright, B. (2020). Sarcasm detection using machine learning algorithms in Twitter: A systematic review. *International Journal of Market Research*, 62(5), 578-598.
- 425 Sarsam, S. M., Al-Samarraie, H., Bahar, N., Shibghatullah, A. S., Eldenfria, A., & Al-Sa'Di, A. (2021). Detecting Real-Time Correlated Simultaneous Events in Microblogs: The Case of Men's Olympic Football. *International Conference on Human-Computer Interaction*,
- 430 Sarsam, S. M., Al-Samarraie, H., Ismail, N., Zaqout, F., & Wright, B. (2020). 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, 1-14.

435 Smailhodzic, E., Hooijsma, W., Boonstra, A., & Langley, D. J. (2016). Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals. *BMC health services research*, 16(1), 1-14.

440 Smeriglio, A., Barreca, D., Bellocco, E., & Trombetta, D. (2016). Chemistry, pharmacology and health benefits of anthocyanins. *Phytotherapy Research*, 30(8), 1265-1286.

Stamatiou, M.-E., Kazantzis, D., Theodossiadis, P., & Chatziralli, I. (2021). Depression in glaucoma patients: A review of the literature. *Seminars in Ophthalmology*,

445 Stefan, A.-M., Paraschiv, E.-A., Ovreiu, S., & Ovreiu, E. (2020). A Review of Glaucoma Detection from Digital Fundus Images using Machine Learning Techniques. 2020 International Conference on e-Health and Bioengineering (EHB),

450 Su, L. Y.-F., Cacciatore, M. A., Liang, X., Brossard, D., Scheufele, D. A., & Xenos, M. A. (2017). Analyzing public sentiments online: Combining human-and computer-based content analysis. *Information, Communication & Society*, 20(3), 406-427.

- Thelwall, M. (2017). The Heart and soul of the web? Sentiment strength detection in the social web with SentiStrength. In *Cyberemotions* (pp. 119-134). Springer.
- 455 Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Zhang, X., & Zhou, S. (2020). Sharing health risk messages on social media: Effects of fear appeal message and image promotion. *Cyberpsychology: Journal of psychosocial research on cyberspace*, 14(2).