
ANALYZING INSTAGRAM POST FOR DEPRESSION FORECASTING USING ML AND NLP APPROACH

Pratibha

Research Scholar, Dr. K. N. Modi University Newai, Rajasthan

Dr. Neha Gupta

Associate Professor, Dr. K. N. Modi University Newai, Rajasthan

Abstract

Despite the fact that depression is a significant problem in terms of mental health, millions of individuals all over the world do not get treatment for it. This is due to prejudice and misinformation related to the condition. As the number of people using social media continues to rise, websites such as Instagram provide a wealth of information that might shed light on how individuals are now feeling. Through the examination of the linguistic information (such as captions, comments, and hashtags) as well as the visual features (such as images) of the postings that people make, our objective is to develop a system that is capable of automatically identifying instances in which individuals are exhibiting signs of depression. Our strategy includes the collection of data from Instagram, the use of text and photo preparation techniques, and the use of machine learning (ML) models such as support vector machines (SVMs), random forests, and deep learning-based classifiers. Important natural language processing techniques, such as topic modelling, sentiment analysis, and linguistic feature extraction, are used in order to identify patterns that are associated with depressive behaviour.

Keywords: Depression Detection, Social Media Analysis, Instagram Posts, Machine Learning (ML), Natural Language Processing (NLP).

INTRODUCTION

The advent of social media presents a promising new opportunity for early detection and intervention in psychiatric disorders. Predictive screening methods have successfully analyzed online media to detect a number of harmful health conditions. All of these studies relied on text analysis, however, and none have yet harnessed the wealth of psychological data encoded in visual social media, such as photographs posted to Instagram. In this report, we introduce a methodology for analyzing photographic data from Instagram to predictively screen for depression.

There is good reason to prioritize research into Instagram analysis for health screening. Instagram members currently contribute almost 100 million new posts per day, and Instagram's rate of new users joining has recently outpaced Twitter, YouTube, LinkedIn, and even Facebook. A nascent literature on depression and Instagram use has so far either yielded results that are too general or too labor-intensive to be of practical significance for predictive analytics. In particular, only attempted to correlate Instagram usership with depressive symptoms, and employed a time-consuming qualitative coding method which

the authors acknowledged made it ‘impossible to qualitatively analyze’ Instagram data at scale (p.4). In our research, we incorporated an ensemble of computational methods from machine learning, image processing, and other data-scientific disciplines to extract useful psychological indicators from photographic data. Our goal was to successfully identify and predict markers of depression in Instagram users’ posted photographs.

Positive emotions are love, laughter, and happiness. Negative emotions can be anger and sadness or depression. Negative emotions can be very serious and enormous and can lead to many deaths. Depression means a common and serious illness that can harm physical & mental wellbeing. It affects how someone feels, how he or she thinks or how he or she acts. Depression is a mental and activity disorder. Scientifically, various events in personal life, major illness, chemical imbalance in the brain, etc. are blamed for depression. Suicide is a major and probable inevitable public health problem. Depression is the main cause of suicide. People with depression suffer not only mental difficulties but also physical problems and the suicide rate become risen. 80% of people who attempt suicide suffer from depression. Depression can be caused for various types of reasons such as Abuse of physical, sexual, emotional, drug-taking personal conflict with anyone, losing someone special, and long-time major illness. Also, childbirth, menopause, unemployment, stress, low income, hassle, jealousy, separation, and social rejection lead a person to depression. According to the World Health Organization (WHO) 322 million people around the world suffer from depression, which is 4.4% of the world's population. People who encounter the COVID-19 area are forever afraid and those who suffered from COVID become depressed more. In the last year, 14 thousand 436 men and women have committed suicide all over the country. Lower or middle-income countries account for 75% or three-quarters of all suicides worldwide. Bangladesh ranks tenth on this list. According to a survey by the national institute of mental health, an average of 26 people commit suicide every day in the country. Most of them are women. Women are about twice as likely to become depressed as men are. Depression is so common in the young generation because today’s youth are indoors and addicted to video games and mobile phones. Most of the total suicides among men and women are between 21 and 30 years of age. Depression has led to a rise in alcohol consumption, and a study found that college students are much higher on the list. The more people stay indoors, the more they will be physically and mentally diseased.

Emotion Recognition and Depression Detection by Machine Learning

Research made a dataset from 887 Reddit users containing 10-2,000 documents. Their goal was to classify which post-depression, in chronological order and obtain results from five models of a classifier based on bag of words (BOW), paragraph-vector, latent semantic analysis (LSA), recurrent neural networks (RNN), and long short-term memory (LSTM). concluded teen suicide is one of the leading causes of death in Australia. Most suicides are linked to depressive disorders & symptoms. To understand and prevent depression and suicide in teens, they have carried out several studies. anxiety-depressive disorder (AD) is predominantly associated with erratic thought processes, insomnia, and restlessness. AD prediction model for anxious depression prediction in real-time tweets is proposed. The proposed model achieves a classification accuracy of 85.09% for tweets of sampled 100 users. tested over 2,500 sentences for emotion prediction from Twitter dataset. The validation of the dataset gives more accuracy than other existing individual classifiers for early detection.

Emotion recognition and depression detection are difficult tasks for single classifiers. Showed that all these methods of separation are based on linguistic style, emotional process, and temporary process and to effectively release the effect of depressing emotions, all aspects must be familiar to. They used 21 types of linguistic inquiry and word count (LIWC) software for stress detection. Explored different sets of features for the task of depression detection in Instagram post. ML techniques for depression detection evaluated using a set of various psycholinguistic and textual features. Then the result shows that in different experiments, the DT gives the highest accuracy. This paper performs an analytical study for depression analysis in a Instagram post context. Supervised classifier algorithms are very popularly used in ML. It is a highly scalable algorithm and a very popular method for text categorization.

Data mining is the significant extraction of underlying, previously unknown, and potentially useful knowledge from data in large data repositories. Nowadays data mining and ML plays a vital role in detecting depression. Our study is to identify whether someone is depressed or not using DMML. We collect information from different sources and implement specific knowledge through data mining to get a better solution & detect depressive user on Instagram post. Generally, ML is a method of analyzing data in a system, it analyses a set of data from our survey & predicts someone's mental health. This paper is sorted as follows: section 2 represents the research methodology including a description of the dataset, implementation process, and classifiers description. Section 3 discusses the experimental result, other findings, and comparative analysis. Finally, section 4 provides a conclusion and future plan.

Necessity of Sentiment Analysis

When individuals take action based on their feelings, it is because they are attempting to convey their feelings to others and to convince others to respond in the same manner. A result of this is that academics hunt for methods to apply computational Sentiment Analysis (SA) to make use of these sentiments; eventually, the conclusion is decided by SA. In the context of this inquiry into human-machine communication, opinion mining is one of the most significant applications of natural language processing (NLP). The process of establishing a system to gather and arrange data from various social network sites, such as online forums, web-based surveys, social media, and blogs, is referred to as opinion mining. An example of this kind of system is online surveys.

Due to the fact that people make decisions based on the resources they have available to them, business houses have the opportunity to learn about the buyer's attitude, appraisal, mood, and opinion. Social media may help with this to a certain degree. Politicians and policymakers are also responsible for making adjustments in response to the concerns of the general population. For the purpose of categorising, finding, or analysing information, there is now the possibility of developing a new piece of software that makes use of data from the real world, with a particular emphasis on textual knowledge discovery techniques, and Twitter in particular. The fact that Twitter is used to send messages via social media and blog sites is the reason why new challenges arise as a result of the many distinctive characteristics that are present in tweets. These characteristics also have an effect on the various domains and methodologies that are covered in sentiment analysis. Below, for your convenience, are some of the most essential aspects of tweets, which will help you better comprehend them:

- **Length of Messages:** As a result of the character limit of 280, reviews of products and films, which are comparable to long text classifiers, are dependent on particular correlations to prior feelings.
- **Writing System:** The meaning of the message is not altered in any way by irregular spellings, cyber slang, acronyms, or characters that are not very common.
- **Accessibility:** When compared to Facebook, far more tweets are available to the general public, and an application programming interface (API) makes it simple to collect all of these tweets at once.
- **Topics:** Through the use of certain domains, Twitter is able to differentiate itself from its predecessors. On the one hand, there are several websites that focus on a specific topic, and on the other hand, the range of topics that are covered is rather vast.
- **Real-Time:** As a result of the limited amount of time available, blogs are often longer and typically having longer intervals between posts.

Importance of Sentiment Analysis

Sentiment analysis offers essential insights that may be used for the purpose of reorganising organisational plans and policies, whether it be in the context of public discourse or as the chief marketing officer of some multinational organisation. The following are some of the advantages:

- **Marketing Approach:** Social media platforms provide companies with an excellent platform from which to advertise their goods and services, and doing continual sentiment research may assist these firms in refining their marketing strategy.
- **Enlarge Product Quality:** As a result of the fact that customer opinion is taken into consideration and unquestionably influences the demand of choices made by other people, marketing research makes it easier to determine the quality of the product.
- **Improve Customer Service:** The consumers are the micro-influencers; hence, the most important thing for a business to do is to give the greatest customer service possible, to deliver items on time, to compensate customers appropriately for faulty products, to respond actively on social media, and so on. Under these conditions, sentiment analysis is used to address the adverse effects that are being experienced by the consumers.
- **Crisis Management:** Keeping a close check on social media at all times helps limit the likelihood of unanticipated crises and safeguards the image of the business or brand from further damage. This is because social media has an influence on current events.

Challenges of Sentiment Analysis

Even though natural language processing (NLP) and text analysis methods (TAL) are helpful in discovering and deriving the precise facts, there are still challenges that need to be resolved. These issues

include determining the appropriate polarity of emotions and providing evidence that supports the correct perception of emotions. Listed here are some of the difficulties:

- **Sentiment of the speaker:** Despite the fact that customers are not usually persuaded by user evaluations, a customer's perspective on a product may nevertheless impact their decision to acquire that product.
- **Multiple Thoughts:** Due to the fact that we do not all have the same feelings about the same thing, it may be challenging to reach an agreement on the services, attributes, and presentation of the things.
- **Fake Opinion:** Due to the fact that customers are often misled by fake reviews, sentiment assessments are rendered worthless in a variety of scenarios where spam is prominent.
- **Challenge in Application:** In spite of the fact that sentiment analysis is widely used, organisations in the entertainment industry, political arena, and commercial sector often fail to develop monitoring systems that are successful.
- **Language Problem:** It is difficult to conduct Opinion Mining (OP) in languages other than English because of the use of lexicons, corpora, and dictionaries. As a consequence, researchers are tasked with the responsibility of producing resources for these languages.
- **Sarcasm Detection:** Individuals often choose positive adjectives to describe bad concepts when they are writing in a sarcastic vocabulary. It is possible that techniques of sentiment analysis are not expressly designed to take into consideration this possibility; thus, this element makes it easier to use sarcasm in order to deceive them. One of the most typical areas to encounter sarcasm is in the comment sections of social networking websites, tweets, and comments. When it comes to sentiment analysis, sarcasm detection is not exactly a stroll in the park, particularly when one does not have sufficient previous information on the topic at issue.
- **Detecting negativity:** When it comes to the study of semantics, the process of negation is a technique that is used to reverse the meaning of words, phrases, and sentences. Researchers make use of a variety of semantic concepts in order to get an understanding of whether or not negation exists. Nevertheless, it is necessary to specify the scope of the words that are impacted by negation terms.
- **Word Ambiguity:** Word ambiguity is another problem that arises when attempting to categorise emotions. The problem of word ambiguity makes it difficult to establish a predefined definition of polarity since the meaning of polarity is highly dependent on the context of the sentence for certain terms.
- **Multi-polarity:** In the process of being studied, it is not very usual for a phrase, document, or other text unit to display multi-polarity. It is possible that relying just on the overall conclusions of the analysis might be deceptive in some circumstances, similar to how an average may occasionally obscure significant details about all of the input data.

OBJECTIVES

1. To apply PRISMA standards to a scoping evaluation of social media and machine learning research on depression and anxiety.
2. To find and examine machine learning models that are used to diagnose sadness and anxiety on social media.

RESEARCH METHODOLOGY

Depression is a prevalent mental illness that is associated with an increased risk of suicide that ranges from 15 to 25 percent. Twelve hundred and fifty million people all around the globe are affected by depression. According to the Disability Adjusted Life Years (DALY) framework, it is expected that it will come in second place overall for all age combinations. It holds the fourth spot on the list of countries with the highest global illness burden.

During the year 2017, it was projected that around 18.3 million individuals in the United States had suffered major depression on average. Almost 7.2% of all adults in the United States are included in this category. It was observed that 8.6% of women have serious depressive episodes, which is much more than the 5.4% of males who do so. It was shown that the age group of 18 to 25 years old had the greatest incidence of people (13.1%) who had experienced a serious episode of depression. It was discovered that those who reported being of two or more races had the greatest prevalence of major depression, which was 11. three percent. In the year 2017, the frequency of major depressive episodes among older persons in the United States is shown in Figure 1.

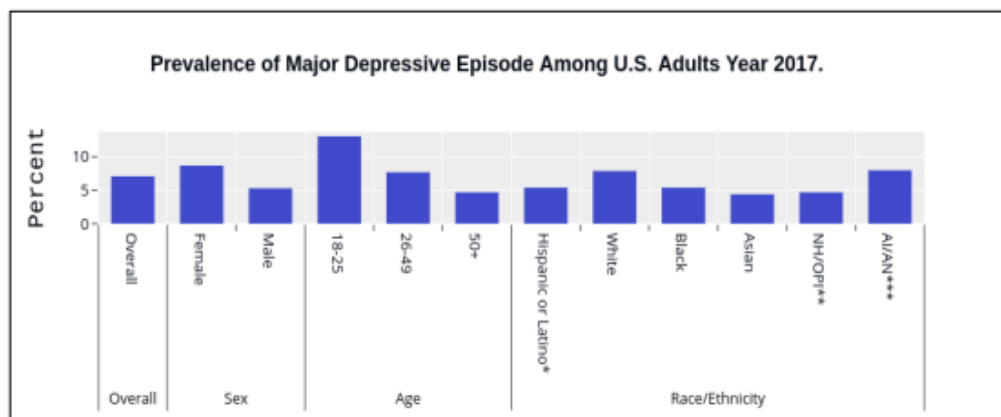


Figure 1: Frequency of major incidence of depression in U.S. citizens aged ≥ 18 in 2017. (All other groups are non-Hispanic or Latino | **NH/OPI = Native Hawaiian / Other Pacific Islander | *AI/AN = American Indian / Alaskan Native)**

Exploration and Pre-Processing

Over the course of the last couple of decades, data from social media platforms have been gathered and examined, and the results of these analyses are shown in figure 4.5. A percentage breakdown of the usage of social media sites among adults in the United States is shown in figure 4.5(a), which is organised according to age. More than seventy percent of individuals in the United States are now using social

networking sites, as shown in figure 4.5(a), according to observations made. Figure 4.5(b) and 4.5(c) respectively illustrate how individuals in the United States use social media sites in accordance with their gender and annual income based on the information provided. According to the degree of education, it is noted from figure-4.5(d) that around eighty percent of college graduates in the United States utilise social networking sites. As demonstrated in figure 4.5(e), the use of social media sites is not only limited in urban areas, but it is also practically equally utilised in suburban areas and rural areas. This is shown by the fact that urban areas are the most restricted. Figure 4.5(f) illustrates the proportion of adult persons in the United States who make use of each social networking platform.

In this study, we carried out our assignment by analysing the comments made by users of Twitter in order to conduct a thorough investigation and identification of the individuals involved. When this occurs, information is gathered from the many social networks together. As part of our efforts to acquire linguistic and geo-positioning specifics, we have collected data from tweets that were published between the dates of 22 April 2020 and 19 May 2020, which corresponds to the third quarter of the COVID-19 threat. As shown in figure 2, the operating concept of Twitter data analysis for the detection of depression is illustrated. Due to the fact that Twitter provides its users with the ability to engage in microblogging, users often utilise acronyms, misspell words, and apply emoticons, in addition to include other characters that express certain concepts. Here is an example of a succinct phrase that is associated with tweeting:

RESULT

In tables 1 to 4, we provide an estimation of the accuracy, remind, fl-score, prop up, and accuracy of the k-nearest neighbours categorisation technique, the prop up vector machine, the decision tree, and the random forest machine learning approach. In the following, we have a look at the possibilities of experiencing either positive or negative depression in the tables 2, 3, 4, and 5. For the purpose of this examination, we take into consideration a number of factors, including age, gender, depression, and word count. Various characteristics, such as gender, the number of words that describe depression, and the age of those who suffer from depression, are used. For the purpose of this investigation, we make use of a collection of machine learning algorithms that have been administered in order to speculate on the projecting capability of our proposed setback. As shown in table 4, the accuracy achieved by using the SVM technique is 71%, whereas the accuracy achieved by employing the KNN method is claimed to be 62% in table 2. On the other hand, the accuracy of the Random Forest approach is 54%, as indicated in table-6, while the accuracy of the Decision Tree method is claimed to be 52% in table-5. The conclusion that can be drawn from this is that, among other things, SVM gives superior results in this respect.

Table 1: Confusion Matrix for SVM

Confusion Matrix					
	Precision	recall	f1-score	Support	Accuracy
Positive	0.35	0.63	0.45	261	71%
Negative	0.83	0.61	0.70	764	

Avg. / Total	0.71	0.61	0.64	1025	
--------------	------	------	------	------	--

Table 2: Confusion Matrix for KNN

Confusion Matrix					
	precision	recall	f1-score	support	Accuracy
Positive	0.46	0.53	0.49	370	62%
Negative	0.71	0.64	0.67	655	
Avg. / Total	0.62	0.60	0.61	1025	

Table 3: Confusion Matrix for RF

Confusion Matrix					
	precision	recall	f1-score	Support	Accuracy
Recurrence	0.41	0.42	0.41	395	54%
No Recurrence	0.63	0.62	0.62	630	
Avg. / Total	0.54	0.54	0.54	1025	

Table 4: Confusion Matrix for DT

Confusion Matrix					
	precision	Recall	f1-score	support	Accuracy
Recurrence	0.46	0.43	0.44	456	52%
No Recurrence	0.56	0.59	0.58	569	
Avg. / Total	0.52	0.52	0.52	1025	

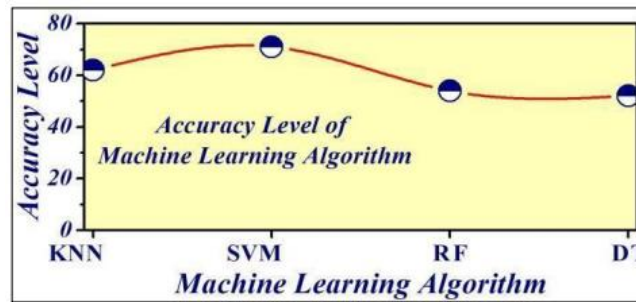


Figure 2: Accuracy level of Machine Learning algorithm

CONCLUSION

We made an effort to focus on the benefits that come with using an online communication system that is based on an internet connection. As a result, a strategy that is connected to scalability would lead to machine learning in conjunction with longitudinal data, which would result in an increase in the screening risk for nonfatal suicide attempts among teenagers. This chapter offers a comprehensive and analytical review of the material that has been published. In the case of a problem such as suicide, its detection and prevention are very important factors, particularly in cases where depression is the driving force behind the death. In recent years, there has been a paradigm change in the clinical care of suicide across diagnostics and assessment. This shift has been made possible by the benefits of technologies such as machine learning and data analytics. The current works are broken down into a few different categories, and both the positive and negative aspects of each are discussed in this chapter. However, the researchers are still facing a hurdle in terms of boosting the accuracy efficiency while maintaining the confidentiality of the data.

REFERENCES

- [1] Biradar, Abhilash & Totad, Shashikumar. (2024). Detecting Depression in Social Media Posts Using Machine Learning. 10.1007/978-981-13-9187-3_64.
- [2] Patil, Prof. (2023). DEPRESSION DETECTION BY ANALYZING SOCIAL MEDIA POST OF USER: A REVIEW. INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT. 08. 1-5. 10.55041/IJSREM32471.
- [3] Farhan, A & Redoy, Farhan & Hossain, Md. (2022). Depression Detection From Social Media Textual Data Using Natural Language Processing and Machine Learning Techniques. 10.13140/RG.2.2.20338.89283.
- [4] Chiu, Chun & Lane, Hsien-Yuan & Koh, Jia-Ling & Chen, Arbee. (2021). Multimodal depression detection on instagram considering time interval of posts. Journal of Intelligent Information Systems. 56. 10.1007/s10844-020-00599-5.
- [5] Gan, Lin & Guo, Yingqi & Yang, Tao. (2024). Machine Learning for Depression Detection on Web and Social Media: International Journal on Semantic Web and Information Systems. 20. 10.4018/IJSWIS.342126.

- [6] Merayo, N. & Lanchares, Alba & González Sanguino, Clara. (2024). Machine learning and natural language processing to assess the emotional impact of influencers' mental health content on Instagram. *PeerJ Computer Science*. 10:e2251. 10.7717/peerj-cs.2251.
- [7] Asad, Nafiz & Pranto, Md. Appel Mahmud & Afreen, Sadia & Islam, Md. Maynul. (2019). Depression Detection by Analyzing Social Media Posts of User. 13-17. 10.1109/SPICSCON48833.2019.9065101.
- [8] Purude, Vaishali & Kumari, P Lalitha. (2020). Analysis of Machine Learning Algorithms for Predicting Depression. 1-4. 10.1109/ICCSEA49143.2020.9132963.
- [9] Dabhane, Suyash & Chawan, Pramila. (2021). Depression Detection on social media using Machine Learning Techniques. 9. 2321-0613.
- [10] Asahara, G & Swarnakantha, N.H.P. Ravi Supunya & Chathurika, Bhagyanie & B.N.W, Subasinghe & K.A, Aththanayake & W.D.M.U.P, Waidyarathna. (2023). Sentiment Analysis in Social Media Data for Depression Detection System. *International Research Journal of Innovations in Engineering and Technology*. 7. 639-647. 10.47001/IRJIET/2023.710083.
- [11] Vasha, Zannatun & Sharma, Bidyut & Esha, Israt & Nahian, Jabir & Polin, Johora. (2023). Depression detection in social media comments data using machine learning algorithms. *Bulletin of Electrical Engineering and Informatics*. 12. 987-996. 10.11591/eei.v12i2.4182.
- [12] Bin Tahir, Waleed & Khalid, Shah & Almutairi, Sulaiman & Abohashrh, Mohammed & Memon, Sufyan & Khan, Jawad. (2025). Depression Detection in Social Media: A Comprehensive Review of Machine Learning and Deep Learning Techniques. *IEEE Access*. PP. 1-1. 10.1109/ACCESS.2025.3530862.
- [13] Tejaswini, Vankayala & Babu, Korra & Sahoo, Bibhudatta. (2022). Depression Detection from Social Media Text Analysis using Natural Language Processing Techniques and Hybrid Deep Learning Model. *ACM Transactions on Asian and Low-Resource Language Information Processing*. 23. 10.1145/3569580.
- [14] Safa, Ramin & Edalatpanah, S A. & Sorourkhah, Ali. (2023). Predicting mental health using social media: A roadmap for future development. 10.48550/arXiv.2301.10453.
- [15] Ayyalasomayajula, Madan Mohan Tito & Agarwal, Akshay & Khan, Shahnawaz. (2024). Reddit social media text analysis for depression prediction: using logistic regression with enhanced term frequency-inverse document frequency features. *International Journal of Electrical and Computer Engineering (IJECE)*. 14. 5998-6005. 10.11591/ijece. v14i5.pp5998-6005.
- [16] Obagbuwa, Ibidun & Danster, Samantha & Chibaya, Colin. (2023). Supervised machine learning models for depression sentiment analysis. *Frontiers in Artificial Intelligence*. 6. 10.3389/frai.2023.1230649.

- [17] Chiong, Raymond & Budhi, Gregorius & Dhakal, Sandeep & Chiong, Fabian. (2021). A textual-based featuring approach for depression detection using machine learning classifiers and social media texts. Computers in Biology and Medicine. 135. 104499. 10.1016/j.combiomed.2021.104499.
- [18] Salas Zárate, Rafael & Alor-Hernández, Giner & Salas Zarate, María & Paredes-Valverde, Mario & Bustos-López, Maritza & Sánchez-Cervantes, José. (2022). Detecting Depression Signs on Social Media: A Systematic Literature Review. Healthcare. 10. 291. 10.3390/healthcare10020291.
- [19] Alsagri, Hatoon & Ykhlef, Mourad. (2020). Machine Learning-Based Approach for Depression Detection in Twitter Using Content and Activity Features. IEICE Transactions on Information and Systems. E103.D. 1825-1832. 10.1587/transinf.2020EDP7023.