

ARTICLE OPEN



An automatic speech analytics program for digital assessment of stress burden and psychosocial health

Amanda M. Y. Chu¹, Benson S. Y. Lam², Jenny T. Y. Tsang³, Agnes Tiwari^{4,5}, Helina Yuk⁶, Jacky N. L. Chan⁷ and Mike K. P. So⁷✉

The stress burden generated from family caregiving makes caregivers particularly prone to developing psychosocial health issues; however, with early diagnosis and intervention, disease progression and long-term disability can be prevented. We developed an automatic speech analytics program (ASAP) for the detection of psychosocial health issues based on clients' speech. One hundred Cantonese-speaking family caregivers were recruited with the results suggesting that the ASAP can identify family caregivers with low or high stress burden levels with an accuracy rate of 72%. The findings indicate that digital health technology can be used to assist in the psychosocial health assessment. While the conventional method requires rigorous assessments by specialists with multiple rounds of questioning, the ASAP can provide a cost-effective and immediate initial assessment to identify high levels of stress among family caregivers so they can be referred to social workers and healthcare professionals for further assessments and treatments.

npj *Mental Health Research* (2023)2:15; <https://doi.org/10.1038/s44184-023-00036-9>

INTRODUCTION

Psychosocial wellness, which encompasses mental, emotional, social, and spiritual well-being, is an essential component of health. However, psychosocial health issues are increasingly recognized as a public health crisis and a global economic burden. According to the World Health Organization, there has been a 13% rise in mental and substance use disorders in the last decade, that is, between 2007 and 2017¹. Mental disorders cause 1 in 5 years lived with disability. Between 1990 and 2019, the global number of disability-adjusted life years lost due to mental disorders increased from 80.8 million to 125.3 million². The direct and indirect costs of psychosocial health issues have amounted to over 4% of the global GDP. These costs totaled more than the combined cost of cancer, diabetes, and chronic respiratory diseases³. The Global Burden of Disease 2019 showed that psychosocial health issues were among the top ten leading causes of disease burden around the globe².

These issues were markedly intensified by the COVID-19 pandemic. Being one of the biggest global crises in recent generations, the COVID-19 pandemic has had long-lasting and far-reaching repercussions for health systems, economies, and societies. Countless people have been infected and many of those are still suffering from long COVID. Businesses have gone bankrupt and millions of people have fallen below the poverty line. Children and young people have missed out on precious learning and socializing opportunities⁴. It has, therefore, created enormous stress among the global population. The number of people reporting symptoms of stress-induced anxiety and depression has profoundly increased by 30% amid the pandemic⁵. In 2021, the World Health Organization announced that depression has become a leading cause of disability worldwide⁶. Psychosocial health issues resulting from acute panic, anxiety,

obsessive behaviors, hoarding, paranoia, and depression served as major contributors to the overall global burden of disease⁷.

Family caregivers are particularly prone to developing psychosocial health issues. According to the Family Caregiver Alliance⁸, all unpaid individuals who assist in the activities of daily living for other family members are family caregivers. In the context of an aging population, lengthening life expectancy, increasing chronic illnesses, and diversifying family structures, the number of family caregivers is on the rise^{9,10}. Caregiving activities create additional physical, psychosocial, and financial burdens for caregivers, thereby leading to stress¹¹. Previous reports have indicated that high levels of stress burden are related to the psychosocial health issues of family caregivers, particularly those with low family resilience¹². Family resilience refers to effective coping strategies and adaptation in the face of losses, hardships, or adversities. It involves the ability to get through crises successfully and become stronger as a result. The capability of a family to manage and move forward from a disruptive experience will influence both the immediate and long-term adaptation of the entire family unit and thus is closely correlated with the caregiver's stress burden¹³. The challenges of caregiving have been exacerbated by the COVID-19 pandemic; family caregivers reported increased duties and stress burden together with a decrease in family resilience, thereby resulting in psychosocial problems^{14,15}.

The early detection of, and intervention in, psychosocial health issues can prevent disease progression and lessen long-term disability^{16,17}. The underlying diseases of psychosocial health issues can be effectively treated at an early stage and at a relatively low cost. Primary healthcare, therefore, plays an important role in the management of these issues. Moreover, identifying individuals with psychosocial health issues—specifically those belonging to vulnerable groups, such as family caregivers—for targeted support is a public health priority.

¹Department of Social Sciences and Policy Studies, The Education University of Hong Kong, Tai Po, Hong Kong, China. ²Department of Mathematics, Statistics and Insurance, The Hang Seng University of Hong Kong, Shatin, Hong Kong, China. ³School of Nursing, Tung Wah College, Ho Man Tin, Hong Kong, China. ⁴School of Nursing, The University of Hong Kong, Pokfulam, Hong Kong, China. ⁵School of Nursing, Hong Kong Sanatorium & Hospital, Happy Valley, Hong Kong, China. ⁶Department of Social Work, The Chinese University of Hong Kong, Shatin, Hong Kong, China. ⁷Department of Information Systems, Business Statistics and Operations Management, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China. ✉email: immkps@ust.hk

Although it is emphasized that psychosocial health is a fundamental human right and is essential to the development of all countries², only five out of ten people that require psychosocial health care services can access the care they need in high-income countries. In low- and middle-income countries, this number is even lower, constituting one out of ten people^{18,19}. The cumbersome and subjective assessment process used to identify psychosocial health issues is a major obstacle to eradicating these issues. Detailed individual interviews are typically considered the gold standard of psychosocial health assessments. This conventional method involves rigorous assessments by specialists based on validated instruments such as questionnaires. It contains multiple rounds of questioning for a wide range of symptoms of mental health problems and social adjustment disorders and therefore, it may take up to a couple of hours to complete²⁰. If interviews are recorded and need to be manually transcribed into text before data processing, it will take even longer and require more human resources. These time- and labor-consuming assessments impede the early detection of, and intervention in, psychosocial problems. Moreover, the assessments derived from clients' responses in the interviews can be subjective; for example, owing to the stigma of psychosocial problems, clients may find it embarrassing to openly express their true feelings. The ability to identify subtle evidence of psychosocial health issues therefore largely depends on the experience of the assessors. These limitations affect both the efficiency and accuracy of the psychosocial health assessments.

In recent years, natural language processing (NLP) techniques^{21–23} are used to detect mental illness including depression, suicide, stress, anorexia, etc. Both traditional machine learning and deep learning methods are employed although deep learning methods receive more attention and perform better. However, these research works mainly study social media posts such as Twitter, Facebook, Reddit, etc. These may miss non-netizens such as elderly people who seldom share their thoughts through social media platforms. Other than that, the learning methods may not provide possible reasons for the causes or explanatory factors of mental illness. The methods, especially most of the deep learning techniques are treated as black-box methods that are hard to know why the methods work well. The causes are important for clinicians that can guide them to mental health diagnosis.

In the face of the tremendous surge in demand for psychosocial services, we developed an automatic speech analytics program (ASAP) for the detection of psychosocial health issues based on clients' speech. ASAP is a digital health technology that involves a novel use of topic modeling, which is a technique used in machine learning. Topic modeling can discover the abstract "topic" in human speech by analyzing the words said. As the theory on linguistic behavior demonstrates, our verbal communication contains both conscious and unconscious components²⁴. "What we say" in a conversation is the conscious component, containing messages that we intend to convey. Meanwhile, "how we say" something in a conversation is the unconscious component, reflecting our psychosocial status that we do not intend to reveal²⁵. Owing to this behavior, when people are talking about certain topics, they will unconsciously include certain keywords in their speech²⁶; for example, when people are upset, their speech will contain more negative words. Accordingly, identifying keywords in clients' speech gives us a better chance of assessing their psychosocial status more accurately. Previous studies indicated that analyzing the frequency of word count in spoken text can be an indicator of various psychosocial disorders, including depression²⁷, generalized anxiety disorder²⁸, and schizophrenia²⁹. Moreover, the ASAP can efficiently obtain adequate data for analysis by asking clients only a few questions. It therefore reduces the time required for an interview and increases the efficiency of the assessment. Furthermore, the automated program allows for cost-effective assessments as well

as making psychosocial health care services more accessible. In this study, we successfully distinguish family caregivers with either high or low stress levels through analyzing their speech using ASAP, thus demonstrating the feasibility of providing easily accessible, efficient, accurate, and cost-effective psychosocial health care services using digital health technologies.

METHODS

ASAP development

ASAP was developed to perform a stress burden assessment using speech analysis. The workflow of ASAP is shown in Fig. 1. The system contains five parts, namely, automated speech recognition (ASR), Text Pre-processing (TP), term frequency-inverse document frequency (TF-IDF) calculation, text analytics (TA), and visual analysis (VA).

ASAP part 1: Automated speech recognition (ASR)

We converted the interview audio tracks into the same file format (for example, wav). Then, we used the Google Cloud Speech API service to perform the speech to text processes to obtain the audio transcript for text analysis³⁰. ASR allows for the automatic conversion of speech into text. The system can also separate the speech signal into chunks to be grouped into different word tokens^{31,32}. To ensure the entire accuracy of the conversion, we performed common Chinese or Cantonese data cleansing before passing the transcripts into further analysis, such as removing punctuations and whitespaces³³. Following that, the text was further analyzed using TP.

Speech Recordings

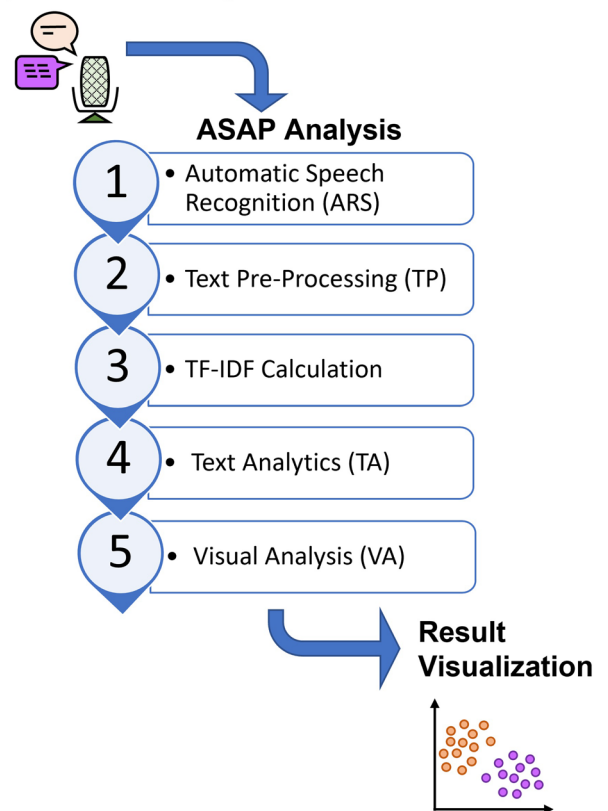


Fig. 1 ASAP workflow diagram. ASAP contains five parts, which are (i) automatic speech recognition (ASR), (ii) text pre-processing (TP), (iii) TF-IDF calculation, (iv) text analytics (TA), and (v) visual analysis (VA). Speech recordings of caregivers were analyzed by ASAP and the clustering results were generated automatically.

ASAP part 2: Text Pre-processing (TP)

Cantonese is the spoken language of the family caregivers and is very different from written Chinese such as Traditional Chinese. It consists of foreign terminology and has a large set of Hong Kong specific terms. This makes the language analysis challenging. Common linguistic tools such as Linguistic Inquiry and Word Count are not applicable to analyze the textual form of Cantonese. Therefore, we adopted a recently developed Python package, PyCantonese, to analyze the textual form of Cantonese. PyCantonese is a Python library for Cantonese linguistics and NLP. However, it has limited accuracy to part-of-speech tagging and sentence parsing. This leads us to employ only two of its functions, including word segmentation and stop words processing. Cantonese is different from English in that Cantonese does not use a space between two words—for example, “解決問題” means “solving problems”. Cantonese also contains numerous word segments, which carry different meanings from the words constituting them—for example, “家人” (family members) is different from “家” (family) and “人” (people). We applied word segmentation to divide the text into meaningful words and word segments using the Python package, PyCantonese. Thereafter, we obtained a set of unique word tokens, though it may contain different languages (for example, English, Chinese, or Cantonese), symbols (including punctuation), and some spoken numbers. Given that our aim was to analyze the content based on the frequency of meaningful word tokens in each document, frequently shown symbols, numbers, and stopwords were not necessary. Thus, we removed them by matching them to the stopwords library (PyCantonese) to clean up the text. After the word segmentation, a list of unique word tokens was identified for subsequent TF-IDF analysis.

ASAP part 3: Term frequency-inverse document frequency (TF-IDF) calculation

TF-IDF calculation can eliminate both common and rare words, which contain minimal information for differentiating caregivers' stress burden levels³⁴. TF-IDF measures the importance of a word in a document and consists of two parts: term frequency (TF) and inverse document frequency (IDF). TF measures the frequency of a word in a document, whereas IDF measures the frequency of a word across all the documents. Intuitively, a word that appears frequently in a document demonstrates its importance. However, this situation is often not the case if the word also appears frequently in all the other documents. For example, the word “only” appears frequently in all the documents in our data. Evidently, this word does not contain any discriminative information for identifying a caregiver's stress burden level. To eliminate these types of common words, IDF counts the word frequency

across all documents and weighs the contributions of these words. Given that TF counts the occurrence of words, the combination of TF and IDF (that is, TF-IDF) can discard words that appear frequently in all documents and retain words that appear frequently in a document. Not only common words, but rare words also contain minimal information. For example, the word “WhatsApp” only appeared in a few documents and did not appear in other documents in our data; that is, this word does not contain discriminative information and should be discarded as well. We introduced rules to eliminate common and rare words, including (i) the total weights of the TF-IDF of a word had to be at least 0.5 and (ii) the ratio between the maximum weight and the total weights of the TF-IDF of a word had to be at least 0.1. Similar rules are adopted to filter out less discriminative information³⁵. The total weights of the TF-IDF of a word measures the total contributions of a word across all the documents. A small value implies that the occurrence of this word is low in all documents and therefore, it is potentially a rare word and should be discarded. The second rule is designed to eliminate common words; if a word commonly appears in most documents, then the TF-IDF weights of this word are similar across different documents. These similarities can be measured using the ratio between the maximum weight and the total weights of the TF-IDF of that word. A small value implies that this word is a common word. After performing the TF-IDF analysis, a list of meaningful words that do not include common or rare words are ready for TA analysis.

Other than TF-IDF, word embeddings³⁶ and n-gram³⁷ are popular features in text analysis. Word embeddings represent words using numerical vectors that put similar words close to each other. However, this approach requires many samples to obtain an accurate estimation. N-gram constructs a sequence of n nearby words. However, as different combinations of nearby Cantonese characters can have very different meanings, n-gram may create words that may be difficult to understand.

ASAP part 4: Text analytics (TA)

TA can analyze the text and discover the abstract “topic” in human speech using the topic ensemble method. The ensemble method^{38–42} combines multiple learning algorithms to analyze complicated datasets and achieve a better performance than a single learning algorithm. It has been successfully applied in a wide range of classification tasks and clustering problems, including facial recognition⁴³, object recognition⁴³, and speech recognition⁴⁴. The proposed ensemble method consists of two phases and they are (1) Ensemble Member Generation; and (2) Consensus Function. More explanations can be found in Fig. 2.

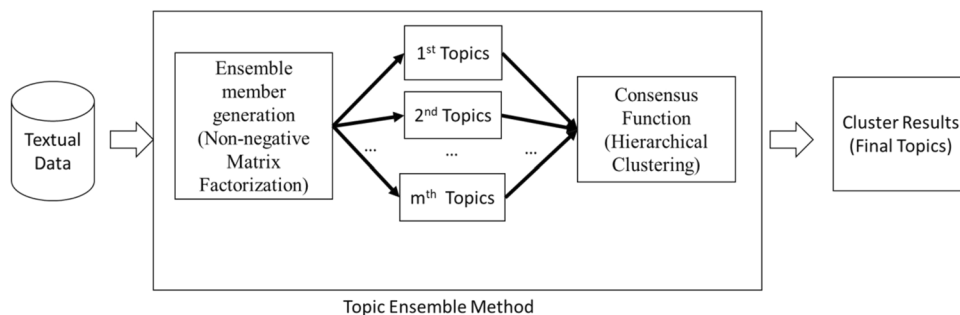


Fig. 2 Illustration of the proposed ensemble modeling. In phase 1, Non-negative Matrix Factorization (NMF)⁵⁸, a popular topic modeling tool, is applied to discover 10 different topics of the documents. Similar to other topic modeling tools, NMF generates different topics with different random number generators. To remove the random effect, we apply NMF with 70 different random number seeds. This produces 70 different sets of topics. Each set has 10 topics. These 70 × 10 topics are the ensemble members. In phase 2, the hierarchical clustering (HC)⁵⁹ with complete linkage is adopted to integrate the ensemble members obtained in phase 1 and form two clusters. The HC is a clustering method that groups similar samples together. In other words, the HC identifies two groups by combining 700 topics.

Table 1. Interview questions related to Walsh's family resilience theory.

Walsh's Theory (Three broad processes)	Interview Questions
Family Belief System	1. What make you stressful? 2. When you feel bad, what you will do to make yourself feel better? 3. What does it mean by adversity according to your experience? 4. What keeps you persevering in the face of adversity?
Organizational Pattern	5. Who is on your mind when you feel weak and why? 6. How many new friends did you make at the Center ^a ? (The HKSKH Lady MacLehose Center) 7. Do you think making new friends are helpful to you? 8. Is there anything make you feel deeply at the Center?
Communicational Pattern	9. What troubles you the most when taking care of your family members? 10. Who do you talk to when you are unhappy and why? 11. Who in your family motivates you the most and why? 12. What gives you the most satisfaction in caring for your family members?

^athe participating non-profit organization providing integrated family and community services in this study.

ASAP part 5: Visual analysis (VA)

We used a symmetric diagram to display the two clusters, which were obtained by the proposed ensemble method, based on the analysis of the keywords that appeared in the respondents' speech and then used multi-dimensional scaling to visualize the clustering results. We also listed the top ten keywords of each topic from caregivers with low and high stress burden levels in the current study.

ASAP application on stress burden analysis: study design

An ASAP has been applied to analyze the speech of family caregivers. Eligible family caregivers were asked 12 open-ended questions regarding their families (Table 1). An ASAP analysis identified family caregivers with a different stress burden.

Some 100 Cantonese-speaking family caregivers were recruited for the study at a non-profit organization (HKSKH Lady MacLehose Center) that provides integrated family and community services in Hong Kong. Potential participants were approached and screened for caregiver stress burden through social workers' professional judgment and were then verified by the Caregiver Burden Inventory (CBI)^{45,46}. Before interviewing the potential participants, the social workers explained the study and obtained signed informed consent from the family caregivers. All the family caregivers took part on a voluntary basis. This project was approved by the Human and Artefacts Research Ethics Committee of The Hong Kong University of Science and Technology (Protocol no.: HREP-2021-0213).

The CBI is a 24-item self-report scale that evaluates caregiver burden through a multi-dimensional approach⁴⁵. It consists of five subscales, including time-dependence, developmental, physical, social, and emotional burden. A validated Chinese version of the CBI was adopted in this study. The Cronbach's alpha of the Chinese CBI was 0.95⁴⁷. Participants in this study were asked to assess each item on the Chinese CBI using a five-point Likert scale, where 0 is not at all descriptive and 4 is very descriptive. A higher score indicates a greater caregiver burden. A total score of 36 or lower indicates that low levels of stress are present while a score that totals over 36 indicates a high degree of stress⁴⁶.

Finally, 47 caregivers with high stress levels and 53 caregivers with low stress levels were recruited on a voluntary basis. A total of 100 family caregivers were asked 12 interview questions about their family conditions and their responses were recorded for the ASAP analysis.

ASAP application on stress burden analysis: Interview questions design

Twelve general questions regarding their family and family resilience were designed for the interview and are listed in Table 1. The 12 questions cover the three broad processes of Walsh's family resilience theory: family belief systems, organizational pattern, and communication pattern. Family belief systems indicate the ability to overcome a crisis through making meaning in the face of adversity, maintaining a positive outlook, and maintaining spiritual belief. Organizational pattern indicates the presence of supportive family relationships that are flexible, connected, and have easy access to social networks and economic resources. Communication pattern indicates the capacity of family members to communicate effectively using communication skills that ensure clarity, allow for open emotional expression, and facilitate problem-solving in the face of adversity⁴⁸.

According to our previous studies, family resilience is closely related to caregiver stress burden where high family resilience is correlated with low stress levels. Asking open-ended questions related to family resilience is a less sensitive topic than asking questions about caregiver stress. These questions allow participants to casually talk about their daily life and express their true feelings. Since family resilience is strongly correlated with stress levels, these non-sensitive questions can reflect the stress burden of the participants. The 100 eligible caregivers were asked to answer all 12 questions and their answers were recorded for subsequent analysis.

RESULTS

Clustering results

The results of this study indicated that the ASAP could distinguish between family caregivers with high or low stress burden levels. According to the ASAP analysis, caregivers could be grouped into two clusters—that is, Clusters A or B—based on the types and frequency of keywords in their speech. Cluster A contains 38 of the 53 caregivers with low stress levels while cluster B contains 34 of the 47 caregivers with high stress levels (Fig. 3 and Table 2). The results indicated that the accuracy rate of distinguishing between caregivers with high or low stress levels is 72% (that is, $(38 + 34) / 100 \times 100\% = 72\%$). Figure 4 visualizes the clustering results of the proposed method using multi-dimensional scaling. It further indicated a strong correlation between caregivers with low levels of stress and cluster A as well as between caregivers with high levels of stress and cluster B.

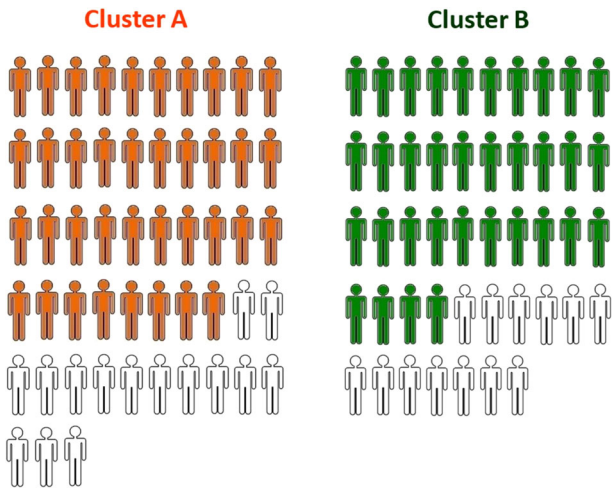


Fig. 3 Symmetric diagram of ASAP analysis result. According to ASAP analysis, the 100 caregivers can be grouped into cluster A (orange) and cluster B (green) based on the types and frequency of keywords appeared in their speech. Cluster A contains 38 of the 53 low-stress caregivers while cluster B contains 34 of the 47 high-stress caregivers.

Table 2. Confusion Matrix of the Proposed Method.

	Low-stress burden	High-stress burden
Cluster A	38	13
Cluster B	15	34

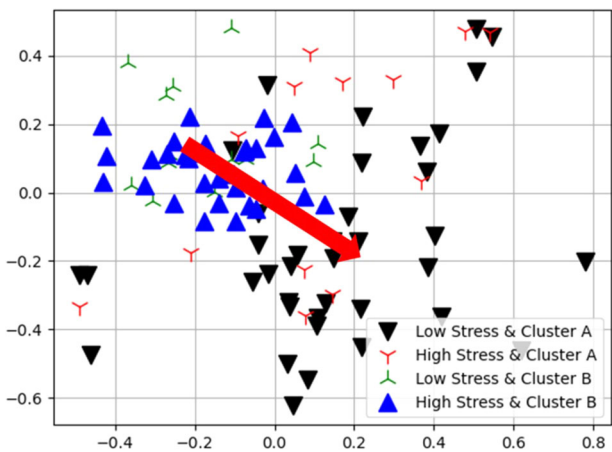


Fig. 4 Visualization of the clustering results using multi-dimensional scaling (MDS). The blue up and dark down triangles represent the caregivers with low stress that belongs to cluster A and high stress that belongs to cluster B, respectively. We can see that these two groups are separated apart. The red arrow indicates the discriminative direction that differentiates the two groups of caregivers.

Performance assessment

We compare the performance of the proposed unsupervised learning method in ASAP with five supervised learning methods. The five methods are three versions of support vector machine (SVM)⁴⁹ and two deep learning methods. The kernels adopted in SVM are linear kernel (SVM (Lin)), radial basis function kernel (SVM (Rbf)), and sigmoid function kernel (SVM (Sig)). The two deep learning methods are word embedding (WE)⁵⁰ approach and

recurrent neural network (RNN)⁵¹ approach. For the supervised learning methods, we use 10-fold cross validation technique to validate their performance. As there are 100 documents in total, this implies there are 90 documents for training and 10 documents for testing. The accuracy rate, true positive (low-stress) rate and true negative (high-stress) rate are reported in Table 3. The accuracy of the proposed method is the highest. It is better than the second-best method (SVM (Rbf)) over 10%. Besides, the ratio of true positive (low-stress) rate and true negative (high-stress) rate of the proposed method is close to one (71.70/72.34 ≈ 1). This balanced ratio implies that the proposed method is not biased to any one of the two classes. The performance to classify high-stress and low-stress are above 70%. However, SVM and Deep Learning methods are generally biased to the true positive (low-stress) rate. The true negative (high-stress) of SVM (Lin) and Deep Learning (WE) are even close to zero. This means they are not able to classify the high-stress cases. The following provides some possible reasons. The proposed method adopts the topic modeling tool, Non-negative Matrix Factorization, to identify words to form topics. Different topics may consist of different words of the documents. In other words, not all the words are used in the classification. However, for the three versions of SVM, they use all the words for classification. As some of the words may be noise, they can confuse SVM and limit their applicability to this data. This may be the reason why the proposed method performs better than SVM. The proposed method is also better than the two deep learning methods. The deep learning methods require a large sample size to generate good results. However, in this study, there are only 90 documents for training. This may not be enough for deep learning to train up a high-performance classifier.

We study the impacts of the number of words toward the performance of the proposed method. We rank all the words using mutual information score⁵². The score measures the similarity between the TF-IDF features of words and the class information (i.e., low-stress or high-stress). A larger score implies a higher association with the class information and thus implies the feature is more important. We select the top 10%, 20%, ..., 90%, 100% of words as an input to the proposed method. The accuracy rates are shown in Table 4. The highest accuracy rate is reached at 60%. The accuracy rate is then dropped and go back to 72. The results imply that some words are not relevant and confuse the proposed method. Those words may be from interview questions that are less related to the stress level. One possible question is the interview question 8 in Table 1. Both low-stress caregivers and high-stress caregivers can feel deeply at the Center.

The keywords identified in the two clusters also demonstrated substantial differences (Table 5). More salient topics were identified for caregivers with a lower stress level (eight topics) than in caregivers with a higher stress level (three topics).

DISCUSSION

The digital health program, ASAP, was successfully developed in this research study. The ASAP can distinguish between family caregivers with high or low stress levels by analyzing their speech. These research findings indicate the potential of using digital health technology to assist in psychosocial health assessments. ASAP can provide an accurate, efficient, and cost-effective initial assessment to identify higher stress burden among family caregivers. The information can be referred to social workers and healthcare professionals for follow-up actions.

The ASAP can identify family caregivers with a low or high stress burden with high accuracy. According to the results of this study, the accuracy for the ASAP to distinguish between a low or high stress burden is 72%. The accuracy is the highest when we compared the performance of the proposed unsupervised learning method in ASAP with five supervised learning methods,

Table 3. Performance Comparison (in Percentage) Using Different Machine Learning Methods.

	True Positive (low-stress) Rate (%)	True Negative (high-stress) Rate (%)	Accuracy Rate (%)
SVM (Lin)	100.00	2.00	54.00
SVM (Rbf)	79.71	41.08	61.00
SVM (Sig)	60.86	54.76	59.00
Deep Learning (WE)	100.00	0.00	53.00
Deep Learning (RNN)	71.24	43.19	56.00
Proposed Method	71.70	72.34	72.00

Table 4. Percentage of words selected as an input for the proposed method.

Percentage of Words	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Accuracy Rate	60	67	69	70	71	75	71	65	69	72

including SVM(Lin), SVM (Rbf), SVM (Sig), WE, and RNN. Identifying stress burden by analyzing people's verbal responses to non-sensitive questions can greatly reduce the biases in psychosocial health assessments. Answering non-sensitive questions allows people to casually talk about themselves and express their true feelings. It can help to reduce bias since due to the stigma of psychosocial health problems, people will be more willing to openly express themselves if the questions are not directly concerned with these problems. Moreover, the digital health program can analyze data based on a technique from machine learning, which may provide more objective and consistent assessments.

Our results in Table 5 found that more salient topics were identified for family caregivers with a low stress burden than those with high stress burden. One possible reason is that caregivers with a lower stress level are more willing to share their experiences and exchange information with others, whereas caregivers with a higher stress level are less willing. This is also reflected in Fig. 4. We can observe in Fig. 4 that the distribution of the low-stress burden caregivers was more spread out than that of the high-stress burden caregivers. Table 5 shows that the low-stress burden caregivers shared more relaxing issues such as "Travel", "Interest", "Appointment", "Wandering", while the high-stress caregivers mainly shared issues about their family members and the words they always used included "Mother", "Father", and "Husband". By reviewing the interview scripts, we found that one of the common stressors of the high-stress burden caregivers was from their family members, with whom they may not be satisfied with their relationship. On the contrary, the low-stress burden caregivers expressed more gratitude than those with high stress burden. Among the eight topics of the low-stress burden caregivers, "happy" appeared three times and "thank you" appeared four times. However, these words did not appear in the high-stress burden caregivers. This echoes the psychology that people who are more stressful are generally less likely to be happy. These results indicated that the proposed ASAP is feasible to assess the caregivers' stress burden.

The efficiency of the psychosocial health assessment can be improved by using the ASAP. The analysis was based on participants' answers to 12 questions regarding their family. The average interview time was around 30 minutes, which is shorter than the average time of a conventional assessment that requires 1 to 2 hours to complete several validated instruments. The ASAP therefore demonstrates the potential to provide highly efficient psychosocial health assessments.

ASAP is an automated digital health program. Together with a high accuracy rate and efficiency, ASAP has the potential to

greatly improve the accessibility of psychosocial health care services. Limited health care resources reduce access and is one of the major problems of psychosocial health care services. It was reported that only 10% of people who require psychosocial health care services can access the care they need in low- and middle-income countries¹⁸. The ASAP can speed up the initial psychosocial health assessments and provide cost-effective psychosocial health screenings to many people. For those who were identified to have psychosocial health issues, they can be referred to health care professionals for further assessments and treatments. The ASAP is therefore a fully scalable system for assessing a wide variety of psychosocial problems in the future. It has the potential to assess different psychosocial problems, including acute panic, anxiety, obsessive behaviors, hoarding, paranoia, and depression. It can also greatly increase the accessibility of psychosocial support services. The ASAP, therefore, has the potential to be an effective tool for society to deal with the ever-growing psychosocial health crisis.

This work also provides a showcase for analyzing text in non-dominant languages and is applicable to other languages. The language we analyzed was Cantonese, which is the language of the family caregivers. However, few NLP tools were available. This made our analysis difficult because the text contained a lot of irrelevant information. In English, many Python language tools can use part-of-speech tagging and word association to extract relevant information. Without the help of these tools, we adopted TF-IDF calculation and introduced a set of rules to filter out less discriminant information. These techniques are applicable to other languages and particularly useful to languages with few analysis tools.

Further research is needed to improve the use of the ASAP in psychosocial health assessments. Firstly, the program can be modified to further improve the accuracy in distinguishing family caregivers with high or low stress burden. Secondly, the number of questions and the content of each question can be modified to further improve the efficiency of the digital psychosocial health assessment. Thirdly, the ASAP can be further developed for the auto-detection of a wide range of psychosocial health issues, including acute panic, anxiety, obsessive behaviors, depression, social phobia, and communication disorders. Fourthly, similar to other AI chatbots^{53–55} in the field of mental health, the ASAP can also empower large language models (LLM) such as ChatGPT or GPT-4 for mental illness detection. The interview questions and topic modeling results discussed in this paper can serve as a foundation for the detection of the mental illness and the interpretation of the detection results.

Table 5. Top 10 keywords of each topic of low-stress and high-stress burden caregivers.

Low-Stress Caregivers					High-Stress Caregivers					
Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 1	Topic 2	Topic 3
多謝 (Thank you)	即時 (Immediate)	活動 (Activity)	交換 (Exchange)	阿女 (Daughter)	右病冇痛 (No illness and pain)	旅行 (Travel)	阿仔 (Son)	媽媽 (Mother)	解決 (Resolve)	先生 (Husband)
旅行 (Travel)	成長 (Growth up)	阿仔 (Son)	電話 (Phone)	家姐 (Elder sister)	開開心心 (Happy)	多謝 (Thank you)	隨便 (casual or informal)	見下 (Meet)	方法 (Method)	阿女 (Daughter)
右病冇痛 (No illness and pain)	溝通 (Communicate)	定係 (Or else)	邊一個 (Which one)	煮飯 (Cook)	解決 (Resolve)	辛苦 (Hard and exhausting)	旅程 (Journey)	行下 (Walk)	需要 (Need)	學校 (School)
開開心心 (Happy)	帶畀 (Bring to)	鍾意 (Like)	約出 (Appointment)	行街 (Wandering)	多謝 (Thank you)	裏面 (Inside)	籃球 (Basketball)	放假 (Vacation)	本身 (Self)	姑娘 (Nurse)
辛苦 (Hard and exhausting)	多謝 (Thank you)	希望 (Hope)	阿女 (Daughter)	關係 (Relationship)	動力 (propulsion)	學習 (Learn)	有得 (Have)	多數 (Majority)	處理 (Deal with)	繼續 (Continue)
裏面 (Inside)	諗起 (Think of)	平時 (Most of the time)?	健康 (Healthy)	出去 (Go out)	堅強 (Strong)	細路仔 (Kids)	狀態 (Condition)	出街 (Go out)	心情 (Mood)	出去 (Go out)
解決 (Resolve)	鍾意 (Like)	工作 (Work)	有少少 (A little bit)	改變 (Change)	話畀 (Tell)	朋友 (Friend)	點解會 (Why)	爸爸 (Father)	冇乜 (Not much)	家姐 (Elder sister)
動力 (propulsion)	快樂 (Happy)	咩嘢 (What up)	開朗 (Cheerful)	同人 (With others)	旅程 (Journey)	多數 (Majority)	定係 (Or else)	最緊要 (The most important)	最主要 (Mainly)	頭先 (A while ago)
學習 (Learn)	見面 (Meet up)	瞓覺 (Sleep)	當中 (In the middle)	經常 (Frequently)	嚟自 (Come from)	感覺 (Feeling)	平時 (Most of the time)	其他人 (The others)	媽媽 (Mother)	少少 (A little bit)
遇嚟 (Come here)	另外 (In addition)	興趣 (Interest)	希望 (Hope)	脾氣 (Temper)	需要 (Required or needed)	大概 (Probably)	即刻 (Immediate)	今日 (Today)	地方 (Place)	以前 (In the past)

In the current study, we employed a clear cut-off for the CBI of either side of 36 to indicate low level of stress (a total score of 36 or lower) and high level of stress (a total score over 36). This may affect the classification accuracy. Further evidence is needed to support the cut-off and the classification accuracy. There are also some limitations that may impede the applications of ASAP. First, there is little evidence that digital health program can be successfully incorporated into healthcare system⁵⁶. Although the usefulness of digital programs has been demonstrated in research studies, few reports have confirmed their usage in clinical settings. One problem is that research has assumed the technology itself can be a stand-alone service, while paying less attention to the ecosystem in clinical settings, such as human support and organizational factors⁵⁶. Our research overcomes this obstacle by designing and developing ASAP as a user-friendly tool for healthcare professionals. It can assist healthcare professionals by streamlining the health assessment process and therefore can be easily integrated into the healthcare system to improve the efficiency of healthcare services. Second, technological barriers can limit the accessibility of digital health programs⁵⁷. Most digital services required the clients to have devices with internet connectivity to access the services. In our research design, it is not necessary for the clients to possess special devices. ASAP can perform health assessment whenever the clients can make an ordinary phone call or come to the healthcare institutes in-person.

With practical measures to minimize the limitations of digital health services, the development of ASAP has a great potential to promote psychosocial health. With the long-lasting effects of COVID-19 and increasing number of global conflicts, the healthcare system of societies and the psychosocial health of people are in jeopardy around the world. Improving the use of digital health technologies in healthcare services is a potential way to handle the surge in cases of psychosocial health problems. It is also critical for ensuring the productivity of societies for their future development.

DATA AVAILABILITY

Aggregated data that support the findings of this study may be available upon request by contacting the corresponding author, Prof. Mike K. P. So. Any request for data will be evaluated and responded to in a manner consistent with policies intended to protect participant confidentiality and language in the study protocol and informed consent form.

CODE AVAILABILITY

Code for all analyses was written in Python 3.9. This code may be available upon request by contacting the corresponding author, M.K.P.S. Any request for Python scripts will be evaluated and responded to in a manner consistent with policies intended to protect participant confidentiality and language in the study protocol and informed consent form.

Received: 17 January 2023; Accepted: 24 August 2023;
Published online: 13 September 2023

REFERENCES

- World Health Organization. *Mental health* https://www.who.int/health-topics/mental-health#tab=tab_2 (2022).
- GBD 2019. Mental disorders collaborators. global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Psychiatr.* **9**, 137–150 (2022).
- Bloom, D. E. et al. The global economic burden of noncommunicable diseases. *Geneva: WEF* (2011).
- World Health Organization. *The impact of COVID-19 on mental health cannot be made light of* <https://www.who.int/news-room/feature-stories/detail/the-impact-of-covid-19-on-mental-health-cannot-be-made-light-of> (2022).

5. Panchal, N., Saunders, H., Rudowitz, R., & Cox, C. *The implications of COVID-19 for mental health and substance use* <https://www.kff.org/coronavirus-covid-19/issue-brief/the-implications-of-covid-19-for-mental-health-and-substance-use/> (2021).
6. World Health Organization. *Depression* <https://www.who.int/news-room/fact-sheets/detail/depression> (2021).
7. Dubey, S. et al. Psychosocial impact of COVID-19. *Diab. Metab. Syndr.: Clin. Res. Rev.* **14**, 779–788 (2020).
8. Family Caregiver Alliance. *Caregiver statistics: Demographics* <https://www.caregiver.org/caregiver-statistics-demographics> (2019).
9. Chu, A. M. Y., Tsang, J. T. Y., Tiwari, A., Yuk, H. & So, M. K. P. Measuring family resilience of Chinese family caregivers: psychometric evaluation of the family resilience assessment scale. *Fam. Relat.* **71**, 130–146 (2022).
10. Turcotte, M. Family caregiving: what are the consequences? *Stat. Can. (Ed.), Insight. Can. Soc.*, 2013001 (2013).
11. Adelman, R. D., Tmanova, L. L., Delgado, D., Dion, S. & Lachs, M. S. Caregiver burden: a clinical review. *JAMA* **311**, 1052–1060 (2014).
12. Li, Y., Wang, K., Yin, Y., Li, Y. & Li, S. Relationships between family resilience, breast cancer survivors' individual resilience, and caregiver burden: a cross-sectional study. *Int. J. Nurs. Stud.* **88**, 79–84 (2018).
13. Walsh, F. The concept of family resilience: crisis and challenge. *Fam. Process* **35**, 261–281 (1996).
14. Beach, S. R., Schulz, R., Donovan, H. & Rosland, A. M. Family caregiving during the COVID-19 pandemic. *Gerontologist* **61**, 650–660 (2021).
15. Chan, A. C. Y., Piehler, T. F. & Ho, G. W. K. Resilience and mental health during the COVID-19 pandemic: findings from Minnesota and Hong Kong. *J. Affect. Disord.* **295**, 771–780 (2021).
16. Kraus, C., Kadriu, B., Lanzenberger, R., Zarate, C. A. Jr. & Kasper, S. Prognosis and improved outcomes in major depression: a review. *Transl. Psychiatr.* **9**, 127 (2019).
17. Rowe, E. Early detection of eating disorders in general practice. *Aust. Fam. Physician* **46**, 833–838 (2017).
18. PAHO. *Mental health problems are the leading cause of disability worldwide, say experts at PAHO Directing Council side event* https://www3.paho.org/hq/index.php?option=com_content&view=article&id=15481:mental-health-problems-are-the-leading-cause-of-disability-worldwide-say-experts-at-paho-directing-council-side-event&Itemid=72565&lang=en (2019).
19. World Health Organization. *WHO highlights urgent need to transform mental health and mental health care* <https://www.who.int/news/item/17-06-2022-who-highlights-urgent-need-to-transform-mental-health-and-mental-health-care> (2022).
20. Davison, T. E., McCabe, M. P. & Mellor, D. An examination of the “gold standard” diagnosis of major depression in aged-care settings. *Am. J. Geriatr. Psychiatry* **17**, 359–367 (2009).
21. Rezaei, N., Wolff, P. & Price, B. H. Natural language processing in psychiatry: the promises and perils of a transformative approach. *BJPsych* **220**, 251–253 (2022).
22. Zhang, T., Schoene, A. M., Ji, S. & Ananiadou, S. Natural language processing applied to mental illness detection: a narrative review. *NPJ Digit. Med.* **5**, 46 (2022).
23. Jain, P., Srinivas, K. R. & Vichare, A. Depression and suicide analysis using machine learning and NLP. *J. Phys.: Conf. Ser.* **2161**, 012034 (2022).
24. Sedivy, J. *Language in mind: an introduction to psycholinguistics* (2nd ed.). Oxford University Press (2020).
25. Seuren, P. A. M. Unconscious elements in linguistic communication: language and social reality. *Eur. J. Philos. Commun.* **6**, 185–194 (2015).
26. Horvath, I. Interpreter behaviour: a psychological approach. *Budapest: Prime Rate Kft* (2012).
27. Smirnova, D. et al. Language patterns discriminate mild depression from normal sadness and euthymic state. *Front. Psychiatry* **9**, 105 (2018).
28. Teferra, B. G. & Rose, J. Predicting generalized anxiety disorder from impromptu speech transcripts using context-aware transformer-based neural networks: model evaluation study. *JMIR Ment. Health* **10**, e44325 (2023).
29. Voppel, A. E., de Boer, J. N., Bredero, S. G., Schnack, H. G. & Sommer, I. E. C. Quantified language connectedness in schizophrenia-spectrum disorders. *Psychiatry Res.* **304**, 114130 (2021).
30. Halpern, Y. et al. Contextual prediction models for speech recognition. In *INTERSPEECH 2016* (pp. 2338–2342) (2016).
31. Arora, S. J. & Singh, R. P. Automatic speech recognition: a review. *Int. J. Comput. Appl.* **60**, 34–44 (2012).
32. Ghai, W. & Singh, N. Literature review on automatic speech recognition. *Int. J. Comput. Appl.* **41**, 42–50 (2012).
33. Pentland, S. J., Fuller, C. M., Spitzley, L. A., & Twitchell, D. P. Does accuracy matter? Methodological considerations when using automated speech-to-text for social science research. *Int. J. Soc. Res. Methodol.* 1–17; <https://doi.org/10.1080/13645579.2022.2087849> (2022).
34. Rajaraman, A. & Ullman, J. D. *Data mining. Mini. Massive Datasets* (pp. 1–17. Cambridge University Press, Cambridge, 2011).
35. Akash, A. & Bagui, S. Sentiment analysis on Twitter data using term frequency-inverse document frequency. *J. Comput. Commun.* **10**, 117–128 (2022).
36. Alina, A. K. Theoretical foundations and limits of word embeddings: what types of meaning can they capture? *Sociol. Methods Res.* <https://doi.org/10.1177/00491241221140142> (2022).
37. Zhu, L., Huang, M., Chen, M. & Wang, W. A N-gram based approach to auto-extracting topics from research articles. *J. Intell. Fuzzy Syst.* **43**, 6137–6146 (2021).
38. Alqurashi, T. & Wang, W. Clustering ensemble method. *Int. J. Mach. Learn. Cybern.* **10**, 1227–1246 (2019).
39. Opitz, D. & Maclin, R. Popular ensemble methods: an empirical study. *J. Artif. Intell. Res.* **11**, 169–198 (1999).
40. Polikar, R. Ensemble based systems in decision making. *IEEE Circuits Syst. Mag.* **6**, 21–45 (2006).
41. Rokach, L. Ensemble-based classifiers. *Artif. Intell. Rev.* **33**, 1–39 (2010).
42. Wang, W. Some fundamental issues in ensemble methods. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)* (pp. 2243–2250) (2008).
43. Beluch, W. H., Genewein, T., Nürnberger, A., & Köhler, J. M. The power of ensembles for active learning in image classification. In *IEEE CVPR* (pp. 9368–9377) (2018).
44. Le Roux, J., Watanabe, S., & Hershey, J. R. Ensemble learning for speech enhancement. In *WASPAA* (pp. 1–4) (2013).
45. Novak, M. & Guest, C. Application of a multidimensional caregiver burden inventory. *Gerontologist* **29**, 798–803 (1989).
46. So, M. K. P., Yuk, H., Tiwari, A., Cheung, S. T. Y. & Chu, A. M. Y. Predicting the burden of family caregivers from their individual characteristics. *Inform. Health Soc. Care* **47**, 211–222 (2022).
47. Chou, K. R., Jiann-Chyun, L. & Chu, H. The reliability and validity of the Chinese version of the caregiver burden inventory. *Nurs. Res.* **51**, 324–331 (2002).
48. Walsh, F. Family resilience: a framework for clinical practice. *Fam. Process* **42**, 1–18 (2003).
49. Wei, L., Wei, B. & Wang, B. Text classification using support vector machine with mixture of kernel. *J. Softw. Eng. Appl.* **5**, 55–58 (2012).
50. Saihan, L. & Bing, G. Word embedding and text classification based on deep learning methods. *MATEC Web Conf.* **336**, 06022 (2021).
51. Hu, H., Liao, M., Zhang, C. & Jing, Y. Text classification based recurrent neural network. In *2020 IEEE ITOEC* (pp. 652–655) (2020).
52. Ross, B. C. Mutual information between discrete and continuous data sets. *PLoS One* **9**, e87357 (2014).
53. Ranjan, B., Kaustav, C. & Rjarshi, N. ChatGPT and its application in the field of mental health. *J. SAARC Psychiat. Fed.* **1**, 6–10 (2023).
54. Lamichhane, B. Evaluation of ChatGPT for NLP-based mental health Applications. <https://doi.org/10.48550/arXiv.2303.15727> (2023).
55. Chen, S., et al. LLM-empowered chatbots for psychiatrist and patient simulation: application and evaluation. <https://doi.org/10.48550/arXiv.2305.13614> (2023).
56. Mohr, D. C., Weingardt, K. R., Reddy, M. & Schueller, S. M. Three problems with current digital mental health research... and three things we can do about them. *Psychiatr. Serv.* **68**, 427–429 (2017).
57. World Health Organization. *Digital health not accessible by everyone equally, new study finds* <https://www.who.int/europe/news/item/21-12-2022-digital-health-not-accessible-by-everyone-equally-new-study-finds> (2022).
58. Hassani, A., Iranmanesh, A. & Mansouri, N. Text mining using nonnegative matrix factorization and latent semantic analysis. *Neural Comput. Appl.* **33**, 13745–13766 (2021).
59. Maimon, O., & Rokach, L. (Eds.). *Data mining and knowledge discovery handbook* (Vol. 2, No. 2005). New York: Springer. (2005).

ACKNOWLEDGEMENTS

This work is partially supported by the Research Matching Grant from the Research Grants Council of the Hong Kong Special Administrative Region (project: 700006 Applications of SAS Viya in Big Data Analytics) and the Big Data Intelligence Centre of The Hang Seng University of Hong Kong.

AUTHOR CONTRIBUTIONS

A.M.Y.C. and M.K.P.S. conceptualized the study. M.K.P.S., B.S.Y.L. and J.N.L.C. collected and analyzed the data. A.M.Y.C., B.S.Y.L., M.K.P.S., J.T.Y.T., A.T., and H.Y. interpreted the results. A.M.Y.C., J.T.Y.T., B.S.Y.L. and J.N.L.C. drafted the manuscript. M.K.P.S., A.T. and H.Y. finalized the manuscript. All authors read and approved the final manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Mike K. P. So.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023