Activity Classification for Daily Lifelogs

Dongyun Nie*
ADAPT Centre
School of Computing
Dublin City University
Dublin, Ireland
dongyun.nie@dcu.ie

Cathal Gurrin
ADAPT Centre
School of Computing
Dublin City University
Dublin, Ireland
cathal.gurrin@dcu.ie

Michael Scriney
Insight Centre for Data Analytics
School of Computing
Dublin City University
Dublin, Ireland
michael.scriney@dcu.ie

ABSTRACT

In recent years, researchers have emphasized interactive question-answering (QA) systems integrated with lifelog retrieval for their prompt query resolution and ability to accommodate various types of data. Lifelog datasets, collected via wearables, serve as valuable resources for multimedia retrieval and human behaviour exploration across diverse fields like healthcare and sports. Accurate lifelong activity prediction is pivotal for understanding daily behaviours, necessitating precise Activities of Daily Living (ADL) prediction. This paper reframes lifelogging's retrieval task as a question-answering challenge, which can be applied to auto-labelling extensive unseen lifelog activity data using classification algorithms. Leveraging machine learning methodologies enables lifelog retrieval systems to analyze and interpret lifelog data, improving ADL predictions and system performance.

CCS CONCEPTS

• **Information systems → Clustering and classification**; **Question answering**.

KEYWORDS

Classification, Lifelog Retrieval, Activities of Daily Living

ACM Reference Format:
Dongyun Nie, Cathal Gurrin, and Michael Scriney. 2024. Activity Classification for Daily Lifelogs. In *The 1st ACM Workshop on AI-Powered Q&A Systems for Multimedia (AIQAM '24), June 10, 2024, Phuket, Thailand*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3643479.3662058

1 INTRODUCTION

In recent years, researchers have increasingly focused on interactive question-answering systems integrated with lifelog retrieval [21, 23]. These systems have garnered attention not only for their capacity to promptly address queries but also for their versatility in accommodating diverse data types. Lifelog datasets, gathered through wearable devices by individuals known as "lifeloggers", have emerged as valuable resources for multimedia retrieval [2, 14, 17] and the exploration of human behaviour [3]. They intersect with various disciplines such as healthcare [12], sports [15], and numerous other applications [11].

In essence, accurate lifelong activity prediction is crucial for understanding an individual's daily behaviours and experiences. An essential role of predicting Activities of Daily Living (ADL) is to ensure the delivery of precise question-answering systems. This task significantly influences the system's overall performance. The Covid-19 management [18, 25] serves as a prime sample of this phenomenon. The integrated ADL and QA systems can assist with common manifestations of the disease; rehabilitation recommendations in the acute hospital setting, recommendations for inpatient rehabilitation and special considerations. Consequently, employing machine learning techniques, particularly classification algorithms, becomes imperative to enhance the accuracy of lifelogger activity prediction.

By leveraging machine learning methodologies, lifelog retrieval systems can analyze and interpret lifelog data to infer underlying patterns and behaviours [13]. Classification algorithms, in particular, enable the categorization of activities based on contextual cues, temporal dependencies, and user-specific preferences. Through the integration of diverse datasets and advanced analytics techniques, these algorithms facilitate more accurate ADL predictions, thereby improving the performance and relevance of lifelog retrieval systems.

The subsequent sections of this paper are structured as follows: First, we delve into a discussion of the literature review. Secondly, introduce the framework and its associated data model for classifying activities. Then, the implementation of the framework will be evaluated. Finally, we conclude this paper.

2 RELATED RESEARCH

In their comprehensive study [17], researchers have outlined the diverse applications of lifelogging across five key domains: Daily activities, Event segmentation, Healthcare, Summarization, and Retrieval. These applications are supported by more diverse datasets encompassing visual, audio, location, physical activity, and physiological signals such as heart rate, ECG (electrocardiogram), EEG (electroencephalogram), EMG (electromyogram), blood pressure, body temperature, blood glucose, blood oxygen saturation, and breathing rate.

The task of activity and event segmentation within lifelog data has been extensively explored in prior research [3, 5, 8, 20]. This process involves partitioning continuous lifelog streams into meaningful segments corresponding to distinct activities or events. Automated segmentation techniques have played a crucial role in facilitating the analysis and interpretation of lifelog data, enabling

*Corresponding author

AIQAM '24, June 10, 2024, Phuket, Thailand

ACM ISBN 979-8-4007-0547-2/24/06
https://doi.org/10.1145/3643479.3662058

researchers to extract valuable insights into individuals' daily routines, behaviour patterns, and contextual interactions.

Furthermore, the applications of lifelogging extend far beyond mere documentation, encompassing domains such as healthcare, where lifelog data serves as a valuable resource for monitoring patients' health status, detecting anomalies, and informing personalized interventions [12]. In the realm of event summarization, lifelogging techniques enable the condensation of extensive lifelog streams into concise representations, aiding in retrospective recall and knowledge extraction [17].

Moreover, lifelog retrieval systems leverage advanced indexing and querying mechanisms to facilitate efficient access to relevant lifelog segments based on user-defined criteria [23]. This enables users to retrieve specific information or relive past experiences effortlessly, enhancing the utility and accessibility of lifelog data for personal, professional, and research purposes.

For lifelog retrieval, some people work on activities related to GPS data. Determining indoor/outdoor. Sit/stand/working/, etc [22]. Some had wide categories [5], and some labelled a small number of topics [22]. Activity segmentation using a multimodal for lifelog data has been conducted in [9]. Instead of using the traditional static segmentation approach, the author deployed a static segmentation and proved the improvement in different types of activities. The broad topics like walking, cooking, and cleaning. Narrow topics like walking at work, brainstorming and ponytails. On average, the dynamic model works slightly better in improving the broad topic.

Many techniques have been applied to the lifelog dataset. In research [11], they used Medical Lifelog Ontology (MELLO) to identify lifelog concepts and relationships between concepts, and it provides clear definitions by following ontology development methods; support the classification and semantic mapping of lifelog data from diverse health self-tracking. The MELLO concepts are divided into two levels. The primary terms are behaviour, body measurement, environment etc. The secondary terms are body region, lifelog apps & devices.

Many researchers have been working on using machine learning algorithms to improve the accuracy of segmenting lifelog data. Hidden Markov Model (HMM) played an important role in identifying the lifelog activities in research [24]. It did achieve prediction for 16 popular everyday activities with an average F-score of 0.9. Artificial Neural Network (ANN) has been used in [13]. It managed to predict the classifier of Mild Cognitive Impairment (MCI) with a good capacity AUC (Area under the Receiver Operating Characteristic Curve) of around 80%. In lifelog image classification, Convolutional Neural Networks(CNN), pre-trained the VGG19 model with XGBoost to best perform the accuracy over 80% [16].

In our research, we would like to apply the classification methods to improve the result of predicting the lifelog activities. This gives the QA systems the ability to ingest unseen data and provide QA functionality without the need for manual annotation.

3 METHODOLOGY

3.1 Data preprocessing & feature selection

The dataset used was available LSC2018 lifelog test collection [10]. The dataset included 31,400 records taken during May 2018. It was annotated with attributes, categories, concepts and ADL. Table 1 details the feature types within the original source dataset.

Table 1: Properties of features in the NCTIR-14 dataset

Type	Description	Example
ADL	A human-annotated description of the events within an image	"eating/drinking" "dish washing" "using desktop computer"
Category	The top five categories indicating the location of the current image and associated confidence scores	"beauty_salon" "television_studio" "drugstore"
Concept	Objects found within each image and associated confidence scores	"person" "bottle" "laptop"
Attribute	The top five attributes found within a specific image with a corresponding score indicating	"man-made" "glossy", "indoor lighting"
Datetime	The date & time of the image in `yyyy-mm-dd hh:mm:ss` format	"2018-05-03 06:00:00" "2018-05-31 22:10:01"

The dataset was flattened to list all concepts, attributes and categories with their associated scores and each ADL was converted into a binary variable which was utilised to construct classifiers. The dates and times were encoded using sine and cosine transformations. This resulted in a final dataset of 31,400 rows and 530 columns as shown in Table 2. This dataset was subsequently utilised for model training, testing and validation.

Table 2: Truncated sample of data after pre-prossessing

beauty_salon	person	man-made	sin_hour	cos_hour	ADL_ eating/ drinking
0.2	0.9	0.6	0.596	0.803	1
0.117	0.9	0.3	0.596	0.803	1
0	0	0.2	0.596	0.803	0
0	0	0.2	0.596	0.803	0

Subsequently, The dataset was filtered to remove ADL classifications with low frequencies. Fig 1 details the frequency of each ADL classification. The class `adl-other` with 5,040 entries was removed from the analysis as it was a placeholder for ADLs which did not fit into other prescribed ADLs. Table 3 details the frequency of the

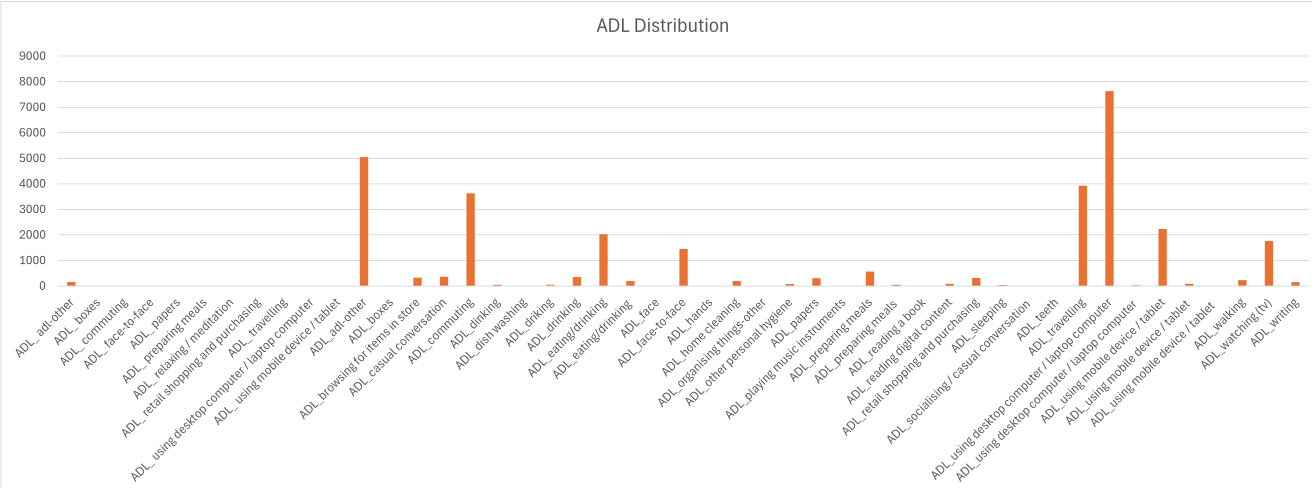

Figure 1: Frequency of all ADL classifications

Table 3: Top 5 most frequent ADL classifications (excl adl-other)

ADL Category	Count
using desktop computer/laptop computer	7633
travelling	3932
commuting	3631
using mobile device/tablet	2230
eating/drinking	2016

top 5 most frequent ADL classifications which were used for our experiments.

3.2 Experimental setup

Each individual ADL class was treated as a binary classification task. For each class four sampled datasets were created `oversampled`, `undersampled`, `SMOTE` and `no sampling`.

For each dataset, the following classification methods were utilised:

- Decision Trees - a method of binary classification by splitting the target dataset along defined criteria (e.g information gain) [19].
- Gradient Boosted Decision trees (GDBT) - an ensemble method where multiple decision trees are combined [7].
- K-Nearest Neighbor (KNN) - a clustering method where K centroids are calculated grouping the dataset using a pre-defined distance function [4]
- Logistic Regression - a binary classification method using a logistic function to fit the data [6].
- XGBoost - an ensemble based tree boosting method for classification [1].

Each sampling & classification method underwent hyperparameter optimisation to determine the optimal configuration. The experimental metrics used were; precision, recall, accuracy and f1-score. For the 5 most frequent ADLs a total of 100 classifiers were constructed. All experiments were run on a Windows 11 machine with an Intel i7-4770K processor, 32GBs of RAM using Python 3.8 with packages scikit-learn v1.1.2, scipy v1.7.1 and XGBoost v1.7.5. For all models, a train, test and validation split of 80-10-10 was used.

4 EVALUATION

In this section we detail the results of our models created for each of the five ADLs [9] shown in Table 3. For each ADL we report on the top five model configurations detailing the Machine Learning method, the sampling method and the precision, recall and f1-score for the model.

ADL - using desktop computer / laptop computer

Table 4 details the results of the top five experimental configurations for the ADL `using desktop computer / laptop computer`. XGBoost on an unsampled dataset provided the highest f1-score of 0.893, in addition, it held the highest precision and recall of 0.893 and 0.896 respectively. The lowest experimental configuration was a Logistic Regression on an oversampled dataset with an f1-score of 0.79. These high results indicate it is relatively easy to classify this ADL which may be attributed to the nature of the ADL in the context where a desktop/laptop computer would be visibly present within the image and appear as a concept.

Table 4: Top 5 performing models for ADL `desktop computer / laptop computer`

Method	Sampling	Precision	Recall	F1-score
XGBoost	unsampled	0.893	0.896	0.893
GDBT	oversampled	0.892	0.884	0.886
XGBoost	smote	0.885	0.885	0.885
GDBT	smote	0.882	0.882	0.882
GDBT	unsampled	0.874	0.877	0.875

ADL - travelling

Table 5 details the top five experimental configurations for the ADL `travelling`. The top-performing model was XGBoost on an unsampled dataset with an f1-score of 0.938 followed by GDBT on a synthetic sampled dataset with an f1-score of 0.929. The worst-performing configuration for this ADL was an oversampled DecisionTree with an f1-score of 0.690. In this instance, it appears the categories which indicate location aid the performance of these classifiers with the top three categories identified for the ADL travelling being “car interior“, “airplane cabin“ and “parking lot“.

Table 5: Top 5 performing models for ADL `travelling`

Method	Sampling	Precision	Recall	F1-score
XGBoost	unsampled	0.938	0.940	0.938
GDBT	smote	0.928	0.930	0.929
GDBT	unsampled	0.925	0.929	0.926
KNN	unsampled	0.923	0.928	0.922
XGBoost	smote	0.920	0.915	0.917

ADL - commuting

Similarly to travelling, the ADL `commuting` yields high results outlined in Table 6. XGBoost unsampled yielded the highest results with an f1-score of 0.940. The lowest performing model of this experimental configuration was a DecisionTree with SMOTE with a score of 0.784. Commuting holds a marginally higher f1-score which may be attributed to the presence of additional categories such as “bus interior“ and “train interior“.

Table 6: Top 5 performing models for ADL `commuting`

Method	Sampling	Precision	Recall	F1-score
XGBoost	unsampled	0.940	0.943	0.940
XGBoost	smote	0.933	0.936	0.934
GDBT	unsampled	0.922	0.928	0.923
GDBT	smote	0.919	0.922	0.920
XGBoost	oversampled	0.926	0.906	0.913

ADL - mobile device/tablet

The results for the ADL `mobile device/tablet` can be seen in Table 7. Interestingly XGBoost with SMOTE performed the best with an f1-score of 0.958 with XGBoost unsampled coming in second with an f1-score of 0.965, however, the unsampled configuration had a marginally higher precision score of 0.959. The lowest-performing configuration for this set was a KNN on unsampled data with an f1-score of 0.829. Due to the nature of this ADL, the presence of a cell phone appearing as a *concept* within the dataset undoubtedly contributed to these results.

Table 7: Top 5 performing models for ADL `mobile device/tablet`

Method	Sampling	Precision	Recall	F1-score
XGBoost	smote	0.958	0.961	0.958
XGBoost	unsampled	0.959	0.961	0.956
KNN	unsampled	0.952	0.956	0.951
GDBT	oversampled	0.951	0.947	0.949
GDBT	unsampled	0.946	0.952	0.946

ADL - eating/drinking

Table 8 outlines the results for the ADL `eating/drinking` outlined the results of the top five experimental configurations for the ADL `eating/drinking`. Similar to other ADLs XGBoost performed the highest on an unsampled dataset with an f1-score of 0.956. Similar to the ADL `using mobile device/tablet` the worst performing method was KNN on an unsampled dataset with an f1-score of 0.859. The high results in this instance may also be attributed to concepts and categories found within this ADL with categories “restaurant“ and “coffee shop“ and the presence of concepts such as “bottle“, “cup“ and “wine glass“.

Table 8: Top 5 performing models for ADL `eating/drinking`

Method	Sampling	Precision	Recall	F1-score
XGBoost	unsampled	0.957	0.961	0.956
XGBoost	smote	0.950	0.955	0.951
GDBT	smote	0.946	0.951	0.947
KNN	unsampled	0.946	0.952	0.945
GDBT	unsampled	0.943	0.950	0.945

Overall results

XGBoost on unsampled data consistently held high results. These results may indicate model bias due to the presence of key concepts or categories which influence results (e.g. the presence of a laptop in an image). Recall the objective of this work is to investigate whether classification methods may be employed to overcome the manual task of applying an ADL to an image. In these instances, the presence of a concept or category may be enough to accurately attribute an ADL to an image.

5 CONCLUSION

The classification of ADLS is of importance for providing a high-quality question-and-answer system. In practitioner domains, this is a largely manual task which proves difficult to manage considering the longitudinal nature of lifelog data. In this work, we investigate the use of ML classification approaches to automatically assign ADLs to images based on the concepts, categories and attributes created by researchers [10]. We employed several experimental configurations for the top five most frequent ADLs within the **LSC2018 Dataset**. For all configurations, XGBoost on unsampled data appears to be a consistently high-performing configuration. Our analysis indicates that for the most frequent ADLs representing the most common actions within the dataset, their classification is indicative of the presence of key categories (locations) or concepts (objects) detected within the data.

Our future work will focus on collaborating with practitioners to apply classifications of less frequent ADLs and employing eXplainable Artificial Intelligence (XAI) approaches to provide further analysis of our approach.

ACKNOWLEDGMENTS

This publication has emanated from research supported by Science Foundation Ireland (SFI) under Grant Numbers SFI/12/RC/2289_P2 and 13/RC/2106_P2, co-funded by the European Regional Development Fund.

REFERENCES

[1] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 785–794.

[2] Tai-Te Chu, Chia-Chun Chang, An-Zi Yen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2020. Multimodal retrieval through relations between subjects and objects in lifelog images. In *Proceedings of the third annual workshop on lifelog search challenge*. 51–55.

[3] Seungeun Chung, Chi Yoon Jeong, Jeong Mook Lim, Jiyoun Lim, Kyoung Ju Noh, Gague Kim, and Hyuntae Jeong. 2022. Real-world multimodal lifelog dataset for human behavior study. *ETRI Journal* 44, 3 (2022), 426–437.

[4] Padraig Cunningham and Sarah Jane Delany. 2020. k-Nearest Neighbour Classifiers: 2nd Edition (with Python examples). *CoRR* abs/2004.04523 (2020). arXiv:2004.04523 https://arxiv.org/abs/2004.04523

[5] Aiden R Doherty and Alan F Smeaton. 2008. Automatically segmenting lifelog data into events. In *2008 ninth international workshop on image analysis for multimedia interactive services*. IEEE, 20–23.

[6] Stephan Dreiseitl and Lucila Ohno-Machado. 2002. Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics* 35, 5-6 (2002), 352–359.

[7] Jerome H Friedman. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* (2001), 1189–1232.

[8] Rashmi Gupta and Cathal Gurrin. 2018. Approaches for event segmentation of visual lifelog data. In *MultiMedia Modeling: 24th International Conference, MMM 2018, Bangkok, Thailand, February 5-7, 2018, Proceedings, Part I 24*. Springer, 581–593.

[9] Rashmi Gupta and Cathal Gurrin. 2019. Activity recognition and segmentation approaches to multimodal Lifelog data. In *2019 international conference on content-based multimedia indexing (CBMI)*. IEEE, 1–7.

[10] Cathal Gurrin, Klaus Schoeffmann, Hideo Joho, Andreas Leibetseder, Liting Zhou, Aaron Duane, Duc-Tien Dang-Nguyen, Michael Riegler, Luca Piras, Minh-Triet Tran, et al. 2019. [Invited papers] comparing approaches to interactive lifelog search at the lifelog search challenge (LSC2018). *ITE Transactions on Media Technology and Applications* 7, 2 (2019), 46–59.

[11] Hye Hyeon Kim, Soo Youn Lee, Su Youn Baik, and Ju Han Kim. 2015. MELLO: Medical lifelog ontology for data terms from self-tracking and lifelog devices. *International journal of medical informatics* 84, 12 (2015), 1099–1110.

[12] Yongjin Kwon, Kyuchang Kang, Changseok Bae, Hee-Joon Chung, and Ju Han Kim. 2014. Lifelog agent for human activity pattern analysis on Health Avatar platform. *Healthcare informatics research* 20, 1 (2014), 69.

[13] Sang-ho Lee, Won-seok Kang, and Cheil Moon. 2018. Lifelog-based classification of mild cognitive impairment using artificial neural networks. In *2018 International Conference on Electronics, Information, and Communication (ICEIC)*. IEEE, 1–2.

[14] Tai Tan Mai, Duc-Tien Dang-Nguyen, Quang-Linh Tran, Ly-Duyen Tran, Tu Ninh, and Cathal Gurrin. 2024. AIQAM'24: The First ACM Workshop on AI-Powered Question Answering Systems for Multimedia. *Proceedings of the 2024 International Conference on Multimedia Retrieval (ICMR'24)*. https://doi.org/10.1145/3652583.3658891

[15] Van-Tu Ninh, Tu-Khiem Le, Liting Zhou, Luca Piras, Michael Alexander Riegler, Pål Halvorsen, Mathias Lux, Minh-Triet Tran, Cathal Gurrin, and Duc Tien Dang Nguyen. 2020. Overview of imageclef lifelog 2020: lifelog moment retrieval and sport performance lifelog. (2020).

[16] Akanksha Rajpute, Tejal Nijai, and Graham Healy. 2020. An Investigation of Transfer Learning for a Lifelog Dataset.. In *AICS*. 265–276.

[17] Ricardo Ribeiro, Alina Trifan, and António JR Neves. 2022. Lifelog retrieval from daily digital data: narrative review. *JMIR mHealth and uHealth* 10, 5 (2022), e30517.

[18] Michael Scriney, Mohan Timilsina, Edward Curry, Lukasz Porwol, Dongyun Nie, Darren Dahley, Jaime B Fernandez, Mathieu D'Aquin, and Mark Roantree. 2023. Engineering Data Assets for Public Health Applications: A Covid-19 Case Study. In *2023 IEEE International Conference on Big Data (BigData)*. IEEE, 1853–1862.

[19] Yan-Yan Song and LU Ying. 2015. Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry* 27, 2 (2015), 130.

[20] Katsuhiro Takata, Jianhua Ma, Bernady O Apduhan, Runhe Huang, and Qun Jin. 2008. Modeling and analyzing individual's daily activities using lifelog. In *2008 International Conference on Embedded Software and Systems*. IEEE, 503–510.

[21] Ly-Duyen Tran, Thanh Cong Ho, Lan Anh Pham, Binh Nguyen, Cathal Gurrin, and Liting Zhou. 2022. LLQA-Lifelog Question Answering Dataset. In *International Conference on Multimedia Modeling*. Springer, 217–228.

[22] Ly-Duyen Tran, Dongyun Nie, Liting Zhou, Binh Nguyen, and Cathal Gurrin. 2023. VAISL: Visual-aware identification of semantic locations in lifelog. In *International Conference on Multimedia Modeling*. Springer, 659–670.

[23] Ly-Duyen Tran, Liting Zhou, Binh Nguyen, and Cathal Gurrin. 2024. Interactive Question Answering for Multimodal Lifelog Retrieval. In *International Conference on Multimedia Modeling*. Springer, 68–81.

[24] Peng Wang and Alan F Smeaton. 2013. Using visual lifelogs to automatically characterize everyday activities. *Information Sciences* 230 (2013), 147–161.

[25] Simge J Yonter, Katherine Alter, Matthew N Bartels, Jonathan F Bean, Martin B Brodsky, Marlís González-Fernández, David K Henderson, Helen Hoenig, Holly Russell, Dale M Needham, et al. 2020. What now for rehabilitation specialists? Coronavirus disease 2019 questions and answers. *Archives of physical medicine and rehabilitation* 101, 12 (2020), 2233–2242.