frontiers | Frontiers in Neurology

TYPE Original Research
PUBLISHED 29 March 2023
DOI 10.3389/fneur.2023.1027160

Check for updates

OPEN ACCESS

EDITED BY
Carol Di Perri,
University of Edinburgh,
United Kingdom

REVIEWED BY
Beril Mat,
University of Wisconsin–Madison,
United States
Andrea I. Luppi,
University of Cambridge,
United Kingdom

*CORRESPONDENCE
Thanh G. Phan
✉ Thanh.Phan@monash.edu

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Neurotrauma, a section of the journal Frontiers in Neurology

RECEIVED 24 August 2022
ACCEPTED 14 March 2023
PUBLISHED 29 March 2023

CITATION
Liyana Arachige M, Seneviratne U, John N, Ma H and Phan TG (2023) Mapping topography and network of brain injury in patients with disorders of consciousness. *Front. Neurol.* 14:1027160. doi: 10.3389/fneur.2023.1027160

COPYRIGHT
© 2023 Liyana Arachige, Seneviratne, John, Ma and Phan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Mapping topography and network of brain injury in patients with disorders of consciousness

Manoj Liyana Arachige[1,2], Udaya Seneviratne[1,2†], Nevin John[1,2†], Henry Ma[1,2†] and Thanh G. Phan[1,2*]

[1]Department of Medicine, School of Clinical Sciences at Monash Health, Monash University, Melbourne, VIC, Australia, [2]Department of Neurology, Monash Health, Clayton, VIC, Australia

Background: There is a growing interest in the topography of brain regions associated with disorders of consciousness. This has caused increased research output, yielding many publications investigating the topic with varying methodologies. The objective of this study was to ascertain the topographical regions of the brain most frequently associated with disorders of consciousness.

Methods: We performed a cross-sectional text-mining analysis of disorders of consciousness studies. A text mining algorithm built in the Python programming language searched documents for anatomical brain terminology. We reviewed primary PubMed studies between January 1st 2000 to 8th February 2023 for the search query "Disorders of Consciousness." The frequency of brain regions mentioned in these articles was recorded, ranked, then built into a graphical network. Subgroup analysis was performed by evaluating the impact on our results if analyses were based on abstracts, full-texts, or topic-modeled groups (non-negative matrix factorization was used to create subgroups of each collection based on their key topics). Brain terms were ranked by their frequency and concordance was measured between subgroups. Graphical analysis was performed to explore relationships between the anatomical regions mentioned. The PageRank algorithm (used by Google to list search results in order of relevance) was used to determine global importance of the regions.

Results: The PubMed search yielded 24,944 abstracts and 3,780 full texts. The topic-modeled subgroups contained 2015 abstracts and 283 full texts. Text Mining across all document groups concordantly ranked the thalamus the highest (Savage score=11.716), followed by the precuneus (Savage score=4.983), hippocampus (Savage score=4.483). Graphical analysis had 5 clusters with the thalamus once again having the highest PageRank score (PageRank=0.0344).

Conclusion: The thalamus, precuneus and cingulate cortex are strongly associated with disorders of consciousness, likely due to the roles they play in maintaining awareness and involvement in the default mode network, respectively. The findings also suggest that other areas of the brain like the cerebellum, cuneus, amygdala and hippocampus also share connections to consciousness should be further investigated.

KEYWORDS

disorders of consciousness, neuroanatomy, text mining, non-negative matrix factorisation, graph theory

Introduction

Disorders of Consciousness (DoC) are sequelae of severe acute brain injury (SABI) with an incidence of 0.7 per 100,000 SABI cases per year (1, 2). Patients with DoC have impaired awareness and/or arousal and are often dependent on full-time nursing care in specialized wards and a multidisciplinary team of clinicians (3). DoC have a profound impact on patients and their families (4). The cost to the healthcare system, such as the total inpatient cost of care, often exceeds $1,000,000 USD per patient (5, 6).

Diagnosis and prognosis of DoC remain challenging, in part from our incomplete understanding of this condition. There is a pervasive idea that DoC has a universally poor prognosis in the first 28 days; this has led to patients with DoC being labeled as a "neglected group of brain-injured patients" (3, 7). A recent review on the subject drew attention to the different categories within the disorder and which subgroups may have late improvement (7). Causative injuries can involve multiple areas of the brain and vary depending on the underlying etiology (8–11). Individual studies, reviews or commentary articles on the topic have often discussed dysfunction in terms of large-scale functional networks such as the default mode network (DMN), and may refer to selected regions within the DMN (12–14). Depending on the nature of the publications some review articles do not discuss the specific brain regions involved (1–3, 15). While papers exploring neuroanatomical substrates of DoC have often utilized novel imaging techniques to localize brain injuries causing DoC, clinicians not directly involved in research do not yet have standardized guidelines to direct the interpretation of these results, and assist decision making and communication of results to families (15–17). We postulate that the aggregated data from these studies and its visual representation will be invaluable to clinicians to understand DoC in their patients (18). This information may also assist future research such as targeted imaging studies in patients with DoC. In addition, the caregivers of patients may want to know the nature of the imaging abnormalities and how this may fit into the concept of DoC. Such data may be used to better communicate this information to caregivers.

Recognizing that only selected brain regions are mentioned in papers on DoC, we have opted not to perform a traditional meta-analysis but to take a data analytics approach. This takes the form of text-mining literature on the topic to determine the brain structures involved in DoC. By text mining a large body of publications, the brain regions most commonly implicated in DoC according to the literature may be aggregated in an unbiased fashion.

Methods

We performed a cross-sectional bibliometric (text mining) analysis of DoC publications.

Because this article did not directly involve human subjects, while using only data from published articles, an institutional review board approval was not required. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline (19).

Data sources

We performed a search query for "Disorders of Consciousness" using the PubMed Entrez Programming Utilities (20). Primary articles for the search query published from 1st January 2000 to 8th February 2023 were retrieved. All studies from the following search query were included:esearch -db pubmed -query "(((Disorders of consciousness) NOT (\"review\"[Publication Type])) NOT (\"systematic review\"[Publication Type])) NOT (\"meta analysis\"[Publication Type])"

Two initial sets of texts (referred to henceforth as *corpora*) were created: the first corpus contained all abstracts from the PubMed e-search; the second contained all freely available full-texts from the PubMed e-search. Scripts for acquiring the abstracts and full texts can be found here: https://github.com/Mango117/BMedScDOC_2023.

While the selection of a single search term and the inclusion of all search results may be unorthodox for meta-analyses, this approach was deemed optimal for text-mining data collection. We chose to focus solely on the keyword search term "Disorders of Consciousness" due to its broad coverage encompassing a variety of papers related to DoC, while avoiding the many classifications of DoC that can vary between labs and literature (21, 22). Three main features of our data collection maximized the integrity of our approach:

1. Firstly, by including thousands of search results, the impact of outliers or papers with outsized impact was minimized given the large volume of the corpora. Unlike a meta-analysis or systematic review comprised of 5–20 papers in which a single paper may skew results, our dataset for text-mining would be 2–3 orders of magnitude larger. This is an important distinction for our choice of methodology and selection of data sources (23).
2. To further minimize the impact of any papers that may skew results, the PubMed query was filtered to only include primary research. Systematic reviews, meta-analyses and any other review articles were excluded as these article types would not provide any novel evidence about disorders of consciousness but may possibly have an outsized impact on text-mining.
3. Finally, a second, smaller pair of datasets was also created for subgroup analysis. These subgroups were selected by using non-negative matrix factorization (NMF). NMF is a multivariate topic modeling algorithm that can perform unsupervised clustering of documents into "hidden topics." NMF algorithms factorize a *term-document matrix* (TDM) into a *term-feature matrix* and a *feature-document matrix*. The *features* of the two new matrices constitute the clustered "hidden topics" within the corpus, and are described by a series of keywords (24). For example, a topic may be described by the keywords: [*mcs, doc, crsr, uws, vegetative, vsuws, state, disorder, consciousness*].

The total number of topics for each dataset was determined by calculating a coherence score – a statistical test measuring the relative distance between words within a topic to derive the optimal number of topics for a specified corpus (25, 26). Using this method, relevant topics were filtered from the abstract and the full-texts corpora, then their documents separated for the subgroups (Figure 1) (26). Thus, 4 main datasets were created for analysis:

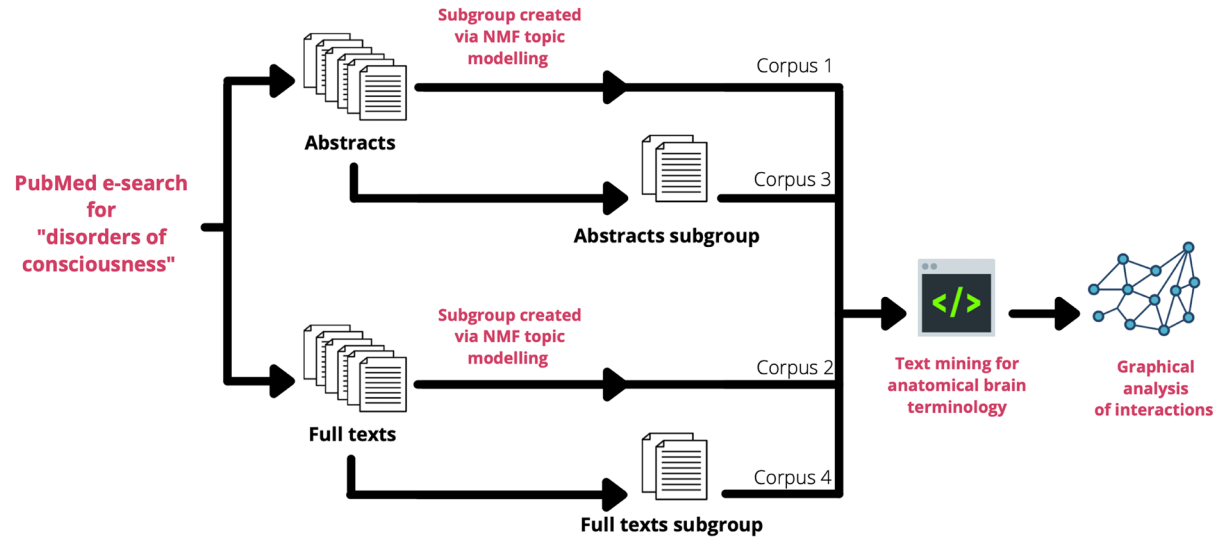

FIGURE 1
Flowchart of how each corpus was prepared from the original PubMed search query.

Corpus 1: an abstracts corpus
Corpus 2: a full-texts corpus
Corpus 3: a subgroup corpus of selected abstracts based on their NMF topics
Corpus 4: a subgroup corpus of selected full-texts based on NMF topics. The code used to perform this NMF can be found here: https://github.com/Mango117/BMedScDOC_2023/blob/main/2-Subgrouping.ipynb.

Anatomical dictionary

To perform the bibliometric analysis, an anatomical dictionary of brain terms fit for text mining was chosen from a well-established neuroanatomical library. These terms were used to search for keywords within the corpora of PubMed documents.

The automated anatomical labeling atlas 3 (AAL3) dictionary was used, which lists 120 discrete regions of the brain and is a commonly used in neuroimaging research (27). The AAL3 anatomical dictionary was modified slightly by making some linguistic changes to optimize the accuracy of the text mining process, and the total dictionary volume was reduced to 50 anatomical regions by combining unilateral terms (i.e., combining "*Left Lateral orbital gyrus*" and "*Right Lateral orbital gyrus*" into "*Lateral orbital gyrus*") (Table 1).

Text mining

After establishing the corpora of documents and the dictionary of anatomical terms, text-mining analysis was performed on all four corpora. Each document was pre-processed to clean data for text mining, ensuring optimal machine readability (23). This included various steps such as:

1. The replacement of complex compound anatomical regions with terms that had whitespace substituted for underscores (i.e., "inferior frontal gyrus, opercular part" to frontal_inf_oper").
2. Filtering of "stop words" to remove common English words such as articles and prepositions (i.e., "I, he, was, and, on, do not") that provided little value to the analysis.
3. Lemmatisation to simplify word variations such as plural nouns, verb tenses and participles.
4. Tokenisation to separate words into a "bag of words" model for unigram analysis, where no importance was placed on the order of words.

After pre-processing, a TDM was created by calculating the frequency of each term from the anatomical dictionary in each corpus.

Statistical analysis

Statistical analysis was performed with Python (Python Software Foundation; version 3.9.5)[1] (28). The term frequency (TF) and mean term frequency-inverse document frequency (TF-IDF) were calculated. Both TF and mean TF-IDF are important measures of sentiment and importance in text-mining analysis. TF describes the frequency of terms within a corpus of documents, while the TF-IDF is a weighted measure of the importance of a term within a corpus of documents. TF-IDF is calculated as follows:

1 https://www.python.org/

TABLE 1 Brain region dictionary of 50 anatomical brain regions created from the AAL3 Anatomical Atlas.

Anatomical brain terms					
1	Amygdala	21	Lateral orbital gyrus	41	Superior frontal gyrus, medial orbital
2	Angular gyrus	22	Lingual	42	Superior occipital gyrus
3	Anterior cingulate	23	Medial orbital gyrus	43	Superior parietal lobule
4	Anterior orbital gyrus	24	Middle cingulate	44	Superior temporal gyrus
5	Calcarine fissure	25	Middle frontal gyrus	45	Supplementary motor area
6	Caudate nucleus	26	Middle occipital gyrus	46	Supramarginal gyrus
7	Cerebellum	27	Middle temporal gyrus	47	Temporal pole: middle temporal gyrus
8	Cerebellar crus	28	Olfactory cortex	48	Temporal pole: superior temporal gyrus
9	Cuneus	29	Pallidum	49	Thalamus
10	Fusiform gyrus	30	Paracentral lobule	50	Vermis
11	Gyrus rectus	31	Parahippocampal		
12	Heschl's gyrus	32	Postcentral gyrus		
13	Hippocampus	33	Posterior cingulate		
14	Inferior frontal gyrus, opercular part	34	Posterior orbital gyrus		
15	Inferior frontal gyrus, pars orbitalis	35	Precentral gyrus		
16	Inferior frontal gyrus, triangular part	36	Precuneus		
17	Inferior occipital gyrus	37	Putamen		
18	Inferior parietal lobule	38	Rolandic operculum		
19	Inferior temporal gyrus	39	Superior frontal gyrus		
20	Insular cortex	40	Superior frontal gyrus, medial		

Arranged in alphabetical order.

$$TFIDF_{x,y} = TF_{x,y} \times \log\left(\frac{N}{DF_x}\right)$$

Where x and y represent term x within document y, $TF_{x,y}$ is the frequency of x in y, DF_x is the number of documents containing x and N is the total number of documents (29). The TF-IDF values were averaged for each term across all documents.

The TF and mean TF-IDF of dictionary terms in each corpus were first ranked, with the aim of measuring concordance between the 4 corpora and providing a combined ranking of brain regions (29). The statistical test commonly used to assess concordance between rankings is Kendall's rank correlation coefficient (Also referred to as Kendall's Tau). However, in this study, it was more appropriate to place a greater statistical emphasis on agreement between highly ranked terms, rather than uniformly assessing concordance across all rankings. In other words, the brain areas ranked 1st, 2nd, and 3rd in a corpus should be weighted more than those placed 20th, 21st, and 22nd.

Based on this protocol, a Savage score was calculated for each set of rankings to measure top-down correlation (30, 31).

Savage scores assign an exponential weighting scheme to rankings assigned to terms in each corpus, ensuring that the most important terms receive a higher weighting when concordance is calculated. This allowed for concordance between each of the 4 corpora to be tested using the Kendall rank correlation coefficient computed on Savage scores:

$$C_T = \frac{1}{b^2\left(n - S_1\right)}\sum_{i=1}^{n} S_i^{\,2} - b^2 n,$$

Where C_T is the Kendall coefficient, and S_i is the sum of the Savage scores assigned to the *i*th object taken over all b sets of rankings. The final Kendall's value computed on Savage scores is a score between 0 and 1, with 0 indicating no concordance and 1 indicating full concordance (30, 31). The final Savage scores allow for a ranking to be created combining multiple datasets – in this case Corpus 1 to 4.

In addition, the full-texts corpus was selected for further graphical analysis using R (The R Foundation for Statistical Computing; version 4.1.0) to explore the associations between brain regions. The network manipulation software, Gephi (Gephi 0.9.2), was used for visualization of the data, including cluster analysis and calculation of a PageRank value. PageRank is an eigenvector type clustering method that intuitively shows the interactions between each node in a graphical network. Each node in this study represented a brain area from our anatomical dictionary, while each edge represented the strength of each vector relationship between a pair of nodes based on data from the TDM. Thus, anatomical terms that were in "close proximity" to each other in a text and regularly associated in articles would be linked *via* edges of varying weights to represent the strength of these associations in the corpus. From this, the PageRank value could be calculated to quantify the cumulative strength of these connections to each node.

Through this method, our graphical analysis was able to visualize the relationships between each of the brain regions used in the compound dictionary, and the relative importance of each one in the full text subgroup corpus studied (32–34).

Results

The PubMed search yielded a total of 24,239 results for the query "Disorders of Consciousness." Data extraction garnered all 24,239 abstracts and 3,780 full-text studies that were open access and thus

freely accessible for download. The abstracts and full-texts acquired were allocated into separate corpora. Analysis of the corpora revealed that the mean wordcount was 132.42 [95% CI, 131 to 134] for the abstract corpus and 2010.16 [95% CI, 1960 to 2060] for the full-texts corpus. NMF of the abstracts uncovered 25 topics by coherence score, and 2 of these topics were selected as containing keywords in close proximity to DoC, yielding a subgroup of 2015 abstracts. NMF of the full-texts corpus uncovered a total of 40 topics by coherence score, and 4 of these topics were selected as containing keywords in close proximity to DoC, yielding a subgroup of 283 documents.

The final 2 abstracts topics selected by NMF were described by the keywords:

"state mcs patient conscious vegetative consciousness doc minimally"

"injury tbi brain traumatic head severe"

The final 4 full-texts topics selected by NMF were described by the keywords:

"mcs doc crsr uws consciousness vsuws disorder vegetative"

"tbi injury traumatic brain head mtbi"

"connectivity network functional brain region consciousness cortex"

"score gcs scale coma outcome Glasgow"

As seen in the listed topics, NMF, while imperfect, was able to cluster corpus documents into topics based on the content contained in each text.

Notably, the thalamus was the most frequent dictionary term across all corpora. It had 2,789 mentions across the 4 datasets and was the highest ranked in each one (Table 2). The 4 corpora were considered significantly coherent with a Kendall's Tau computed on Savage scores of 0.7350 ($p < 0.001$). Globally, the thalamus was ranked the highest (Savage score = 11.716), followed by the precuneus (Savage score = 4.983) and the hippocampus (Savage score = 4.483) (Table 3).

Graphical analysis was performed on Corpus 4 and yielded 5 clusters: cluster 1 had 9 members with the parahippocampus [PageRank = 0.0330] as the most important member; cluster 2 had 11 members with the caudate nucleus [PageRank = 0.0337] being most important; cluster 3 had 6 members with the inferior parietal gyrus [PageRank = 0. 0324] being most important; cluster 4 had 4 members with the postcentral gyrus being the most important member [PageRank = 0.0321]; and cluster 5 had 6 members with the thalamus being the most important member [PageRank = 0.0344] (Figure 2). An interactive online visualization of this graph can be seen here: https://mango117.github.io/BMedScDOC/network.html along with details of each group included in the supplementary material (Table 4). A 3D visualization of these brain regions in is depicted in Figure 3.

Node sizes are weighted by betweenness centrality, and communities are color coded by modularity cluster. (Green) Cluster Group 1. (Pink) Cluster Group 2. (Orange) Cluster Group 3. (Yellow) Cluster Group 4. (Blue) Cluster Group 5. A description of each group is in Table 4.

An interactive online visualization of this graph can be seen here: https://mango117.github.io/BMedScDOC/network.html.

Brain regions are colored in a red to light blue gradient, with the most frequent region being a dark red. An animated 3D GIF of this visualization can be seen here: https://mango117.github.io/BMedScDOC/braingif.html.

Discussion

Our results indicated that the thalamus was the most frequently mentioned anatomical brain regions in DoC studies. This finding is not surprising given it has a well-established link with awareness and arousal in current DoC literature (35–37). The novelty of our approach is in taking written text in scientific literature and producing visualization tools for both clinicians and the lay public. In this paper, we were also able to apply graph theoretical analysis to discern the underlying structure of the connections among these brain structures. These findings were performed by collating available techniques and using open-source software and open data; these methods are documented in this paper and the code has been made available on Github. This reproducible approach provides a way for clinicians and other citizen scientists to undertake similar research on this topic or related topics in neuroscience and other fields.

TABLE 2 Bibliometric search results highlighting the top 3 highest ranked brain regions in each corpus by TF-IDF.

Corpus	Rank	Dictionary term	Mentions	Mean TF-IDF
Corpus 1: Abstracts	1	Thalamus	386	0.000664
	2	Cerebellum	180	0.000366
	3	Hippocampus	123	0.000309
Corpus 2: Full Texts	1	Thalamus	625	0.002343
	2	Hippocampus	110	0.001328
	3	Cerebellum	206	0.001280
Corpus 3: Abstracts NMF	1	Thalamus	174	0.003456
	2	Precuneus	54	0.001295
	3	Anterior cingulate	29	0.000960
Corpus 4: Full Texts NMF	1	Thalamus	1,604	0.008453
	2	Precuneus	176	0.002875
	3	Amygdala	509	0.002809

Anatomical findings

The high mean TF-IDF of the thalamus in the full-texts subgroup (Corpus 4) (TF-IDF = 0.008453) and high Savage score (Savage = 11.716) is highly congruent with current literature. The thalamus inhabits the dorsal component of the diencephalon, between the cerebral cortex and midbrain. This location allows it to act as a hub to relay information between subcortical areas and the cerebral cortex (38). Most sensory inputs are sent to the thalamus before being redirected *via* many thalamocortical radiations to their final destinations in the outer cortex (39). This key anatomical feature aligns with our graph theory analysis. The high PageRank score of 0.03439 indicates many highly weighted connections between the thalamus and other nodes in the network, with 91.89% of nodes sharing an edge with it. Its high centrality on the network visualization (Figure 2) indicates that it shares many connections with various other cortical regions rather than being highly connected to a single region of the brain. In the literature, the thalamus is strongly connected with both awareness and arousal in DoC. Awareness relies on high connectivity between the thalamus and cortex, and arousal requires brainstem signals to be forwarded to the thalamus (35, 36). Its further importance as a treatment target for deep brain stimulation indicates that downregulation of the thalamus likely has considerable implications on conscious state (40).

TABLE 3 The top 10 highest ranked dictionary terms by Savage score.

Dictionary terms	Ranking	Savage score
Thalamus	1	11.716
Precuneus	2	4.983
Hippocampus	3	4.483
Cerebellum	4	4.415
Amygdala	5	3.849
Caudate nucleus	6	3.373
Anterior cingulate cortex	7	2.511
Posterior cingulate cortex	8	2.488
Pallidum, lenticular nucleus	9	1.547
Superior frontal gyrus, dorsolateral	10	0.636

Components of the cingulate cortex also received a high TF-IDF across the corpora. The cingulate cortex is in the medial aspect of the cerebral hemispheres. It is often categorized into three parts – the anterior cingulate cortex (ACC), the midcingulate, the posterior cingulate cortex (PCC) (41). The ACC (PageRank = 0.0309, Savage score = 2.511) is often linked to emotion due to dense interconnectedness with limbic system structures such as the amygdala, sharing a strong edge weighting on the graphical visualization (Figure 2). The ACC's role in self-perception has been previously correlated to the level of consciousness. A fMRI study by Qin et al. illustrated ACC signal changes when DoC patients were exposed to self-related auditory stimuli which further correlated with the patient's level of consciousness (42). The PCC (PageRank = 0.0343, Savage score = 2.488) has previously well-established links with consciousness. The PCC forms a central node in the DMN, active during "wakeful rest." (16).

Similarly, the precuneus was highlighted as a notable brain region across the DoC corpus (Savage Score = 4.983, PageRank = 0.0332). The

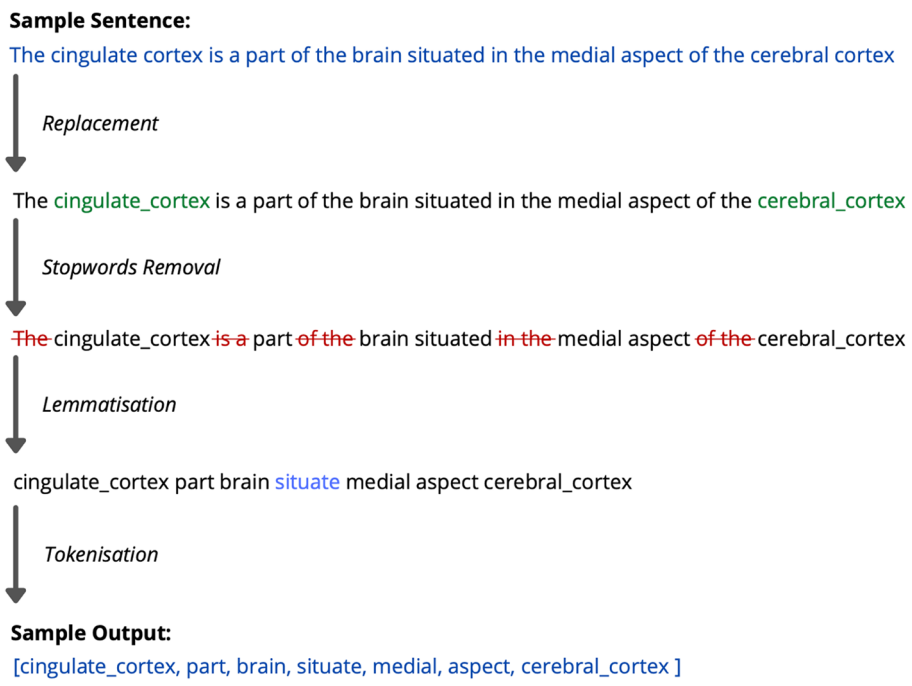

FIGURE 2
Flowchart illustrating some of the pre-processing steps undertaken during the text mining process. An example sentence is used.

TABLE 4 Distribution of brain regions between modularity cluster groups in graphical network visualization.

Group 1	Group 2	Group 3
Calcarine	Amygdala	Cingulate_mid
Frontal_inf_oper	Caudate	Parietal_inf
Frontal_inf_tri	Cingulate_ant	Rolandic_oper
Fusiform	Cingulate_post	Temporal_inf
Heschl	Frontal_mid	Temporal_mid
Occipital_inf	Frontal_sup	Temporal_pole_sup
Occipital_mid	Hippocampus	
Occipital_sup	Parietal_sup	
Parahippocampal	Putamen	
	Supramarginal	
	Vermis	
Group 4	**Group 5**	
Ofclat	Cuneus	
Ofcpost	Pallidum	
Postcentral	Paracentral_lobule	
Temporal_sup	Precuneus	
	Supp_motor_area	
	Thalamus	

Listed in alphabetical order.

precuneus is situated in the medial aspect of the posterior parietal lobe. It notably has strong corticocortical connections with the prefrontal cortex as well as the ACC (43, 44). The precuneus also has well established links to both the NCC and the DMN, with fMRI and functional connectivity studies illustrating its importance in recovery from UWS, supporting the recovery of consciousness (45–47).

Of greater novelty, however, are the findings linking brain regions that are not as commonly associated with DoC. The amygdala (PageRank = 0.01987, Savage score = 6.291) and hippocampus (PageRank = 0.01881, Savage score = 6.5231) are midline areas strongly linked to the limbic system functions of emotional responses and memory (48). Both share input from the thalamus, as mirrored by the high centrality of all three locations on the graphical visualization (Figure 2) (49). The high frequency of these two regions alongside the ACC may suggest limbic involvement in DoC. This has been highlighted by a few previous studies into pain perception during DoC, but this conclusion cannot be made based on text mining analysis alone (50–52). The cuneus (PageRank = 0.01956, Savage score = 5.8743) is another midline area. Previous literature utilizing white matter connectometry analysis has demonstrated an increase fiber density in the cuneus associated with increasing conscious state during DoC recovery, but once again, literature on this relationship is sparse (53–55).

Perhaps most interesting is the prominence of the cerebellum in this analysis. The cerebellum is a neuron-dense hindbrain structure that has typically been associated purely with ipsilateral motor control (56). However, there is newer research highlighting the role of the cerebellum in nonmotor functions such cognition and affect, with connections to cortical and subcortical areas like the limbic system and prefrontal cortex (57–59). In our text-mining analysis, the cerebellum was found to have a high significance across all 4 corpora (Savage Score = 4.415), which is seemingly consistent with this smaller body of research associating the cerebellum with cognition and emotion (60–62). Conversely however, it must be noted that case reports of cerebellar agenesis suggest that the structure might not be directly involved in maintaining any NCCs at all (63). Given the contradictory evidence, it seems that further research into the cerebellum's role in nonmotor function and consciousness is required (Figure 4).

Comments on experimental design

Text mining, sometimes referred to as natural language processing (NLP), is a way of converting unstructured text into structured data (23). This allows for automated analysis of large corpora that would otherwise be infeasible for a human researcher to analyze. A human reader would have trouble processing, memorizing and understanding 14,945 abstracts and 2,178 full-texts on this topic. A domain knowledge expert would have this knowledge having spent many years working in the field. By contrast, a newcomer to the topic would not be able to do this.

In this study, we have illustrated a different approach to text mining scientific literature. In the case of DoC literature, there are different methods between studies due to varying imaging techniques or diagnostic criteria used (64–66). Text mining also allows some overlooked details to be picked up on, such as the highlighting of limbic system structures in this study. This is often because of the "objective" analysis a computer is able to perform without any pre-existing ideas or biases regarding a research topic.

Limitations

There are limitations when using text mining to analyze scientific literature (67). Many anatomical terms we searched for were long tail keywords (i.e., "Left supplementary motor area") comprised of multiple nouns and adjectives. The linguistic structure of these terms presents a limitation for text mining (68). In writing, these terms are rarely fully written out by researchers and often have a high degree of variation (i.e., using acronyms or shorthand). It was noted that different publications interchangeably used words such as "region," "gyrus," "area" and "cortex" when describing brain regions. We utilized pre-processing of data in each corpus to account for such linguistic idiosyncrasies (23).

Unlike human researchers, a text mining algorithm is unable to grasp the context of each individual word or document. This is unavoidable in the case of such a large dataset with millions of words, unless a human were to perform this study entirely by hand. Countermeasures such as subgroup analysis using NMF topic modeling provide an added layer of accuracy to text mining (24). By separating a large dataset into smaller corpora based on topics, we can compare different subgroups to better understand the corpora holistically. With a large enough dataset, effective pre-processing, and topic modeling, the limitations of text mining can be minimized.

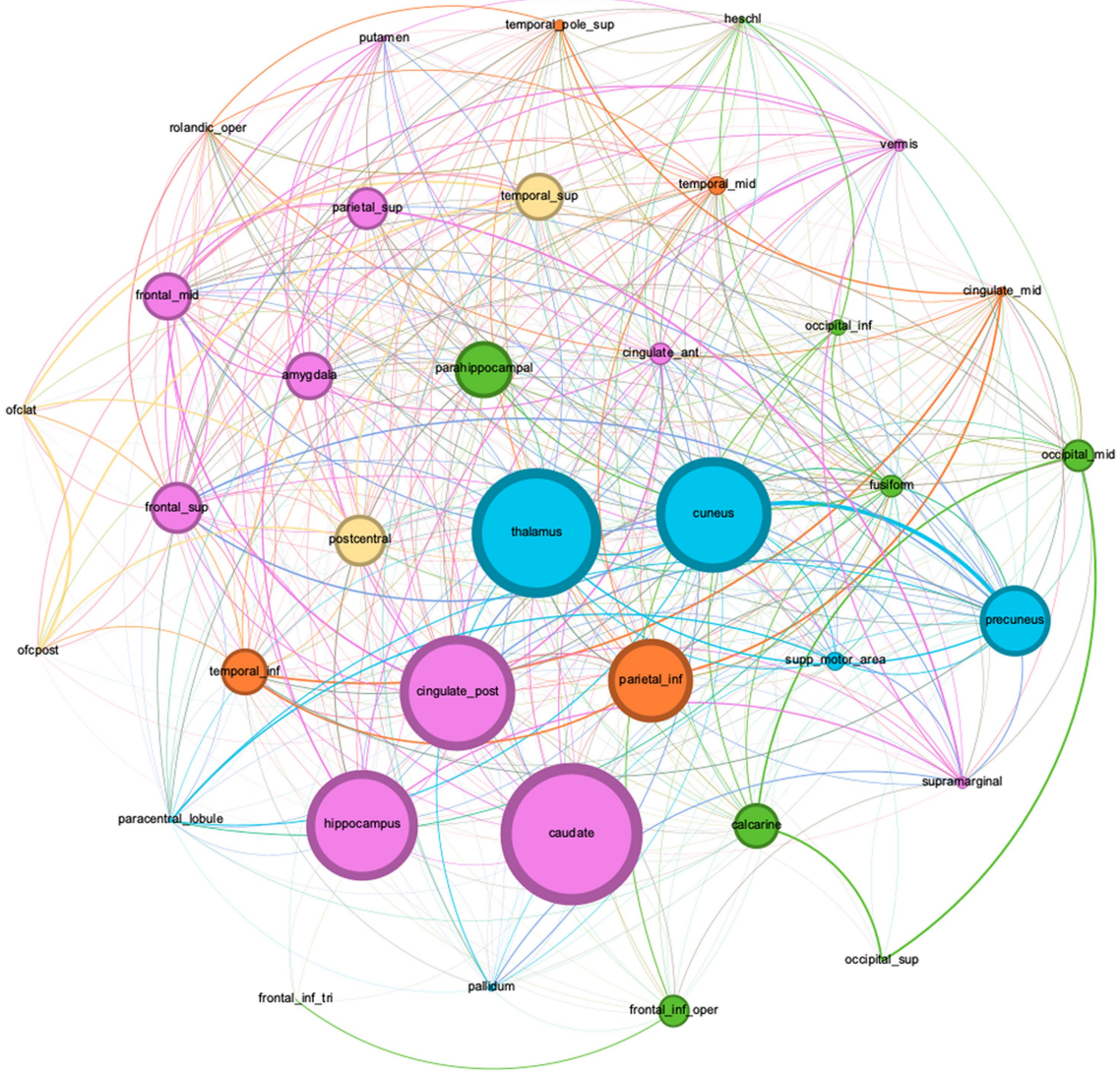

FIGURE 3
Graph network visualization of dictionary terms found in the full-texts subgroup corpus (Corpus 4).

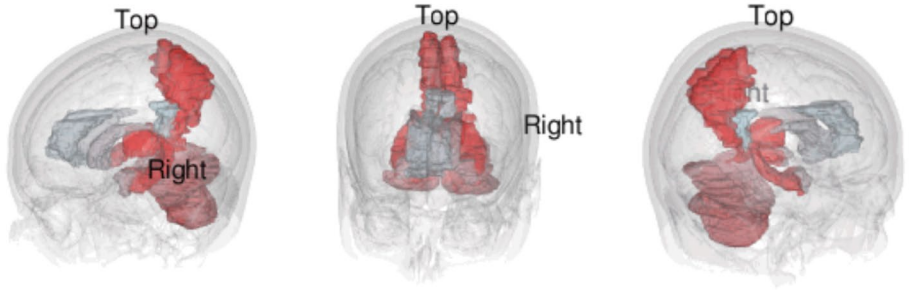

FIGURE 4
Still images of a 3D visualization of the top 10 most frequently mentioned terms from text-mining of the full-texts corpus.

Conclusion

In this study, text mining of DoC literature revealed the thalamus, cingulate cortex and precuneus are strongly associated with DoC, likely due to the roles they play in maintaining awareness and comprising the Default Mode Network (DMN). Areas of the brain such as the cerebellum, amygdala and hippocampus, were also prominently mentioned in literature.

These structures and their relationships are documented in our figures and on the web. Their relationship to DoC should be further investigated.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Author contributions

ML: data curation, formal analysis, investigation, methodology, software, and writing (original). US: supervision and writing (review). NJ and HM: writing (review). TP: conceptualization, project administration, supervision, and writing (review). All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Gray M, Lai S, Wells R, Chung J, Teraoka J, Howe L, et al. A systematic review of an emerging consciousness population: focus on program evolution. *J Trauma*. (2011) 71:1465–74. doi: 10.1097/TA.0b013e31821f82f5

2. Eapen BC, Georgekutty J, Subbarao B, Bavishi S, Cifu DX. Disorders of consciousness. *Phys Med Rehabil Clin N Am*. (2017) 28:245–58. doi: 10.1016/j.pmr.2016.12.003

3. Giacino JT, Katz DI, Schiff ND, Whyte J, Ashman EJ, Ashwal S, et al. Practice guideline update recommendations summary: disorders of consciousness: report of the guideline development, dissemination, and implementation subcommittee of the American Academy of neurology; the American congress of rehabilitation medicine; and the National Institute on Disability, Independent Living, and Rehabilitation Research. *Neurology*. (2018) 91:450–60. doi: 10.1212/WNL.0000000000005926

4. Chinner A, Pauli R, Cruse D. The impact of prolonged disorders of consciousness on family caregivers' quality of life – A scoping review. *Neuropsychol Rehabil*. (2021) 32:1643–66. doi: 10.1080/09602011.2021.1922463

5. Wade DT. How many patients in a prolonged disorder of consciousness might need a best interests meeting about starting or continuing gastrostomy feeding? *Clin Rehabil*. (2018) 32:1551–64. doi: 10.1177/0269215518777285

6. National Insts of Health, Consensus Development Panel, Rehabilitation of Persons with Traumatic Brain Injury. Rehabilitation of persons with traumatic brain injury. *JAMA*. (1999) 282:974–83. doi: 10.1001/jama.282.10.974

7. Fins JJ, Bernat JL. Ethical, palliative, and policy considerations in disorders of consciousness. *Neurology*. (2018) 91:471–5. doi: 10.1212/WNL.0000000000005927

8. Giacino JT, Fins JJ, Laureys S, Schiff ND. Disorders of consciousness after acquired brain injury: the state of the science. *Nat Rev Neurol*. (2014) 10:99–114. doi: 10.1038/nrneurol.2013.279

9. Sigl JC, Chamoun NG. An introduction to bispectral analysis for the electroencephalogram. *J Clin Monit*. (1994) 10:392–404. doi: 10.1007/BF01618421

10. Fridman EA, Beattie BJ, Broft A, Laureys S, Schiff ND. Regional cerebral metabolic patterns demonstrate the role of anterior forebrain mesocircuit dysfunction in the severely injured brain. *Proc Natl Acad Sci U S A*. (2014) 111:6473–8. doi: 10.1073/pnas.1320969111

11. Albrechtsen SS, Riis RGC, Amiri M, Tanum G, Bergdal O, Blaabjerg M, et al. Impact of MRI on decision-making in ICU patients with disorders of consciousness. *Behav Brain Res*. (2022) 421:113729. doi: 10.1016/j.bbr.2021.113729

12. Monti MM, Vanhaudenhuyse A, Coleman MR, Boly M, Pickard JD, Tshibanda L, et al. Willful modulation of brain activity in disorders of consciousness. *N Engl J Med*. (2010) 362:579–89. doi: 10.1056/NEJMoa0905370

13. Sporns O. Structure and function of complex brain networks. *Dialogues Clin Neurosci*. (2013) 15:247–62. doi: 10.31887/DCNS.2013.15.3/osporns

14. Friston KJ. Functional and effective connectivity: a review. *Brain Connect*. (2011) 1:13–36. doi: 10.1089/brain.2011.0008

15. Phan TG. Disorders of consciousness: are we ready for a paradigm shift? *Lancet Neurol*. (2013) 12:131–2. doi: 10.1016/S1474-4422(12)70289-2

16. Fernández-Espejo D, Soddu A, Cruse D, Palacios EM, Junque C, Vanhaudenhuyse A, et al. A role for the default mode network in the bases of disorders of consciousness. *Ann Neurol*. (2012) 72:335–43. doi: 10.1002/ana.23635

17. Roquet D, Foucher JR, Froehlig P, Renard F, Pottecher J, Besancenot H, et al. Resting-state networks distinguish locked-in from vegetative state patients. *Neuro Image Clin*. (2016) 12:16–22. doi: 10.1016/j.nicl.2016.06.003

18. Fernandez-Espejo D, Bekinschtein T, Monti MM, Pickard JD, Junque C, Coleman MR, et al. Diffusion weighted imaging distinguishes the vegetative state from the minimally conscious state. *NeuroImage*. (2011) 54:103–12. doi: 10.1016/j.neuroimage.2010.08.035

19. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ*. (2007) 335:806–8. doi: 10.1136/bmj.39335.541782.AD

20. Sayers E. *A general introduction to the E-utilities*. Bethesda, MD: National Center for Biotechnology Information (US) (2010).

21. Gosseries O, Vanhaudenhuyse A, Bruno M-A, Demertzi A, Schnakers C, Boly MM, et al. Disorders of consciousness: coma, vegetative and minimally conscious states In: D Cvetkovic and I Cosic, editors. *States of consciousness: Experimental insights into meditation, waking, sleep and dreams*. Berlin, Heidelberg: Springer Berlin Heidelberg (2011). 29–55.

22. Song H, Li Q, Li S, Kang W, Yang J, Yang Y, et al. Classification of patients with disorder of consciousness based on DTI sequence analysis. 2017 IEEE International Conference on Smart Cloud (Smart Cloud), New York, NY, USA. (2017). 3–5.

23. Žižka J, Dařena F, Svoboda A. *Text mining with machine learning: Principles and techniques*. Boca Raton, FL: CRC Press (2019).

24. Pauca P, Shanaz F, Berry M.W., Plemmons R.J. Text mining using non-negative matrix factorization society for industrial and applied mathematics. *Proceedings of the 2004 SIAM International Conference on Data Mining (SDM)*. (2004). 87–90.

25. Röder M, Both A, Hinneburg A. Exploring the space of topic coherence measures. Proceedings of the eighth ACM international conference on web search and data mining. Shanghai, China: Association for Computing Machinery. (2015). 399–408.

26. Mifrah S, Benlahmar ELH. Topic modeling coherence: A comparative study between LDA and NMF models using COVID'19 corpus. *Int J Adv Trends Comput Sci Eng*. (2020) 9:5756–61. doi: 10.30534/ijatcse/2020/231942020

27. Rolls ET, Huang C-C, Lin C-P, Feng J, Joliot M. Automated anatomical labelling atlas 3. *NeuroImage*. (2020) 206:116189. doi: 10.1016/j.neuroimage.2019.116189

28. Kluyver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, Frederic J, et al. Jupyter notebooks – A publishing format for reproducible computational workflows In: F Loizides and B Scmidt, editors. *Positioning and Power in Academic Publishing: Players, Agents and Agendas*: IOS Press (2016)

29. Sharma U, Jain A, Anshita . Term frequency inverse document frequency. *J Multimedia Technol Recent Adv*. (2020) 7:1–6.

30. Iman R, Conover W. A measure of top–down correlation. *Technometrics*. (1987) 29:351–7. doi: 10.1080/00401706.1987.10488244

31. Kendall MG. *Rank correlation methods*. New York, NY: Hafner Publishing Company (1962).

32. Bastian M, Heymann S, Jacomy M. Gephi: An open source software for exploring and manipulating networks. Proceedings of the third international AAAI conference on weblogs and social media (ICWSM'09). San Jose: American Journal of Sociology. (2009).

33. Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project. (1999).

34. Sporns O. Graph theory methods: applications in brain networks. *Dialogues Clin Neurosci*. (2018) 20:111–21. doi: 10.31887/DCNS.2018.20.2/osporns

35. Bernat JL. Chronic disorders of consciousness. *Lancet*. (2006) 367:1181–92. doi: 10.1016/S0140-6736(06)68508-5

36. Garcia-Rill E. *Reticular activating system*: Academic Press. (2009).137–143.

37. BL. *Walter AGS*. Midbrain: Academic Press. (2014). 28–33

38. Gazzaniga MS, Ivry RB, Mangun GR. *Cognitive neuroscience: The biology of the mind*. New York: W.W. Norton (2019).

39. Yeh FC, Panesar S, Fernandes D, Meola A, Yoshino M, Fernandez-Miranda JC, et al. Population-averaged atlas of the macroscale human structural connectome and its network topology. *NeuroImage*. (2018) 178:57–68. doi: 10.1016/j.neuroimage.2018.05.027

40. Giacino J, Fins JJ, Machado A, Schiff ND. Central thalamic deep brain stimulation to promote recovery from chronic posttraumatic minimally conscious state: challenges and opportunities. *Neuromodulation*. (2012) 15:339–49. doi: 10.1111/j.1525-403.2012.00458.x

41. Leech R, Sharp DJ. The role of the posterior cingulate cortex in cognition and disease. *Brain*. (2014) 137:12–32. doi: 10.1093/brain/awt162

42. Qin P, Di H, Liu Y, Yu S, Gong Q, Duncan N, et al. Anterior cingulate activity and the self in disorders of consciousness. *Hum Brain Mapp*. (2010) 31:1993–2002. doi: 10.1002/hbm.20989

43. Leichnetz GR. Connections of the medial posterior parietal cortex (area 7m) in the monkey. *Anat Rec*. (2001) 263:215–36. doi: 10.1002/ar.1082

44. Thier P, Andersen RA. Electrical microstimulation distinguishes distinct saccade-related areas in the posterior parietal cortex. *J Neurophysiol*. (1998) 80:1713–35. doi: 10.1152/jn.1998.80.4.1713

45. Liu X, Li S-J, Hudetz AG. Increased precuneus connectivity during propofol sedation. *Neurosci Lett*. (2014) 561:18–23. doi: 10.1016/j.neulet.2013.12.047

46. Horovitz SG, Braun AR, Carr WS, Picchioni D, Balkin TJ, Fukunaga M, et al. Decoupling of the brain's default mode network during deep sleep. *Proc Natl Acad Sci U S A*. (2009) 106:11376–81. doi: 10.1073/pnas.0901435106

47. Tagliazucchi E, von Wegner F, Morzelewski A, Brodbeck V, Borisov S, Jahnke K, et al. Large-scale brain functional modularity is reflected in slow electroencephalographic rhythms across the human non-rapid eye movement sleep cycle. *NeuroImage*. (2013) 70:327–39. doi: 10.1016/j.neuroimage.2012.12.073

48. Amunts K, Kedo O, Kindler M, Pieperhoff P, Mohlberg H, Shah NJ, et al. Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: Intersubject variability and probability maps. *Anat Embryol*. (2005) 210:343–52. doi: 10.1007/s00429-005-0025-5

49. Carlesimo GA, Lombardi MG, Caltagirone C. Vascular thalamic amnesia: A reappraisal. *Neuropsychologia*. (2011) 49:777–89. doi: 10.1016/j.neuropsychologia.2011.01.026

50. Boly M, Faymonville ME, Schnakers C, Peigneux P, Lambermont B, Phillips C, et al. Perception of pain in the minimally conscious state with PET activation: an observational study. *Lancet Neurol*. (2008) 7:1013–20. doi: 10.1016/S1474-4422(08)70219-9

51. Calabro RS, Naro A, Manuli A, Leo A, De Luca R, Lo Buono V, et al. Pain perception in patients with chronic disorders of consciousness: what can limbic system tell us? *Clin Neurophysiol*. (2017) 128:454–62. doi: 10.1016/j.clinph.2016.12.011

52. Aubinet C, Murphy L, Bahri MA, Larroque SK, Cassol H, Annen J, et al. Brain, behavior, and cognitive interplay in disorders of consciousness: A multiple case study. *Front Neurol*. (2018) 9:665. doi: 10.3389/fneur.2018.00665

53. Tan X, Gao J, Zhou Z, Wei R, Gong T, Wu Y, et al. Spontaneous recovery from unresponsive wakefulness syndrome to a minimally conscious state: early structural changes revealed by 7-T magnetic resonance imaging. *Front Neurol*. (2018) 8:741. doi: 10.3389/fneur.2017.00741

54. Di Perri C, Bahri MA, Amico E, Thibaut A, Heine L, Antonopoulos G, et al. Neural correlates of consciousness in patients who have emerged from a minimally conscious state: a cross-sectional multimodal imaging study. *Lancet Neurol*. (2016) 15:830–42. doi: 10.1016/S1474-4422(16)00111-3

55. Bodien YG, Chatelle C, Edlow BL. Functional networks in disorders of consciousness. *Semin Neurol*. (2017) 37:485–502. doi: 10.1055/s-0037-1607310

56. Hodos W. Evolution of cerebellum In: MD Binder, N Hirokawa and U Windhorst, editors. *Encyclopedia of neuroscience*. Berlin, Heidelberg: Springer Berlin Heidelberg (2009). 1240–3.

57. Wolf U, Rapoport MJ, Schweizer TA. Evaluating the affective component of the cerebellar cognitive affective syndrome. *J Neuropsychiatry Clin Neurosci*. (2009) 21:245–53. doi: 10.1176/jnp.2009.21.3.245

58. Schmahmann JD, Pandya DN. The cerebrocerebellar system. *Int Rev Neurobiol*. (1997) 41:31–60. doi: 10.1016/S0074-7742(08)60346-3

59. Ramnani N. The primate cortico-cerebellar system: Anatomy and function. *Nat Rev Neurosci*. (2006) 7:511–22. doi: 10.1038/nrn1953

60. Turner BM, Paradiso S, Marvel CL, Pierson R, Boles Ponto LL, Hichwa RD, et al. The cerebellum and emotional experience. *Neuropsychologia*. (2007) 45:1331–41. doi: 10.1016/j.neuropsychologia.2006.09.023

61. Al Abdulgader AA. Human consciousness: the role of cerebral and cerebellar cortex, vagal afferents, and beyond In: JB Stavros, editor. *Cerebral and cerebellar cortex*. Rijeka: Intech Open (2020)

62. Strata P, Scelfo B, Sacchetti B. Involvement of cerebellum in emotional behavior. *Physiol Res*. (2011) 60:S39–48. doi: 10.33549/physiolres.932169

63. Yu F, Jiang Q-J, Sun X-Y, Zhang R-W. A new case of complete primary cerebellar agenesis: clinical and imaging findings in a living patient. *Brain*. (2014) 138:e353. doi: 10.1093/brain/awu239

64. Giacino JT, Katz DI, Schiff ND, Whyte J, Ashman EJ, Ashwal S, et al. Comprehensive systematic review update summary: disorders of consciousness. Report of the guideline development, dissemination, and implementation Subcommittee of the American Academy of neurology; the American congress of rehabilitation medicine; and the National Institute on disability, Independent Living, and Rehabilitation Research. *Arch Phys Med Rehabil*. (2018) 99:1710–9. doi: 10.1016/j.apmr.2018.07.002

65. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. *Lancet*. (1974) 2:81–4. doi: 10.1016/S0140-6736(74)91639-0

66. Galanaud D, Perlbarg V, Gupta R, Stevens RD, Sanchez P, Tollard E, et al. Assessment of white matter injury and outcome in severe brain trauma: A prospective multicenter cohort. *Anesthesiology*. (2012) 117:1300–10. doi: 10.1097/ALN.0b013e3182755558

67. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Syst Rev*. (2015) 4:5. doi: 10.1186/2046-4053-4-5

68. Feldman R, Sanger J, Press CU. *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge: Cambridge University Press (2007).