

Original Paper

# Subspace Clustering of Physiological Data From Acute Traumatic Brain Injury Patients: Retrospective Analysis Based on the PROTECT III Trial

Sina Ehsani<sup>1</sup>, MSc; Chandan K Reddy<sup>2</sup>, PhD; Brandon Foreman<sup>3</sup>, MD; Jonathan Ratcliff<sup>4</sup>, MD, MPH; Vignesh Subbian<sup>1,5</sup>, PhD

<sup>1</sup>Department of Systems and Industrial Engineering, College of Engineering, The University of Arizona, Tucson, AZ, United States

<sup>2</sup>Department of Computer Science, Virginia Polytechnic Institute and State University, Arlington, VA, United States

<sup>3</sup>Department of Neurology & Rehabilitation Medicine, University of Cincinnati, Cincinnati, OH, United States

<sup>4</sup>Department of Emergency Medicine, Emory University School of Medicine, Emory University, Atlanta, GA, United States

<sup>5</sup>Department of Biomedical Engineering, College of Engineering, The University of Arizona, Tucson, AZ, United States

**Corresponding Author:**

Vignesh Subbian, PhD

Department of Biomedical Engineering

College of Engineering, The University of Arizona

1127 E James E Rogers Way

Tucson, AZ, 85721

United States

Phone: 1 5206216559

Email: [vsubbian@email.arizona.edu](mailto:vsubbian@email.arizona.edu)

## Abstract

**Background:** With advances in digital health technologies and proliferation of biomedical data in recent years, applications of machine learning in health care and medicine have gained considerable attention. While inpatient settings are equipped to generate rich clinical data from patients, there is a dearth of actionable information that can be used for pursuing secondary research for specific clinical conditions.

**Objective:** This study focused on applying unsupervised machine learning techniques for traumatic brain injury (TBI), which is the leading cause of death and disability among children and adults aged less than 44 years. Specifically, we present a case study to demonstrate the feasibility and applicability of subspace clustering techniques for extracting patterns from data collected from TBI patients.

**Methods:** Data for this study were obtained from the Progesterone for Traumatic Brain Injury, Experimental Clinical Treatment–Phase III (PROTECT III) trial, which included a cohort of 882 TBI patients. We applied subspace-clustering methods (density-based, cell-based, and clustering-oriented methods) to this data set and compared the performance of the different clustering methods.

**Results:** The analyses showed the following three clusters of laboratory physiological data: (1) international normalized ratio (INR), (2) INR, chloride, and creatinine, and (3) hemoglobin and hematocrit. While all subclustering algorithms had a reasonable accuracy in classifying patients by mortality status, the density-based algorithm had a higher F1 score and coverage.

**Conclusions:** Clustering approaches serve as an important step for phenotype definition and validation in clinical domains such as TBI, where patient and injury heterogeneity are among the major reasons for failure of clinical trials. The results from this study provide a foundation to develop scalable clustering algorithms for further research and validation.

(*JMIR Biomed Eng* 2021;6(1):e24698) doi: [10.2196/24698](https://doi.org/10.2196/24698)

**KEYWORDS**

cluster analysis; unsupervised machine learning; traumatic brain injury

## Introduction

Traumatic brain injury (TBI) is broadly defined as disruption in normal brain function or other evidence of brain pathology as a result of mechanical force directed at the head or a rapid acceleration/deceleration event. TBI is the most common cause of death and disability in children and adults aged less than 44 years [1]. However, there has been little change in TBI-related deaths despite advancements in care delivery [2]. Additionally, a major challenge to both TBI-related clinical research and acute care is reliably identifying candidates for targeted interventions [3]. While there have been substantial advances in technological and computational approaches to TBI phenotyping [4-6], there is still a dearth of actionable information that can be used for pursuing secondary clinical research in this domain.

Existing approaches to stratification of patients based on clinical presentation does not adequately address the heterogenous nature of TBI, whereas data mining and machine learning techniques have shown promise in identifying subgroups [5], predicting outcomes [7], and prognosticating among TBI patients [8]. In particular, clustering-based techniques serve as an important step for phenotype definition and have the potential to uncover previously unrecognized relationships between various physiologic variables [9]. For example, in other clinical domains, traditional cluster analyses have been helpful in identifying unique subgroups of patients. These studies include application of k-means cluster analysis for identifying distinct phenotypes of asthma patients [10], as well as using hierarchical clustering to identify both new and known relationships between physiologic variables collected from critically ill patients [9]. In this study, we applied *subspace clustering* (or subclustering) methods on physiologic data collected from TBI patients and compared the performance of different subspace clustering methods (density-based, cell-based, and clustering-oriented methods). The rationale for applying subspace clustering over traditional clustering methods (eg, k-means) is the ability to account for the multiple low-dimensional subspace structure of higher dimensional data [11]. In terms of critical illnesses, such as acute TBI, we hypothesize that the complex latent relationships between various physiologic variables are better represented in subspaces and thus better captured by subclustering methods than traditional methods that are often limited to spatial proximity of data points in individual clusters.

## Methods

### Data Source

Data for this study were obtained from the Progesterone for Traumatic Brain Injury, Experimental Clinical Treatment-Phase III (PROTECT III) study. The PROTECT III trial included a cohort of 882 TBI patients [12], who were originally recruited for a randomized clinical trial to study the effect of progesterone on patients with acute TBI. Patients were randomly assigned to a treatment group that received progesterone within 4 hours of injury or placebo. While the PROTECT III clinical trial showed that there was no difference in patients between the two study groups, the longitudinal data from the trial were made available for secondary analyses and continued research.

This data set included patient demographics, baseline assessment data, 6-month outcome data, including the Glasgow Outcome Scale Extended scale, and mortality status. The temporal data in this study included laboratory test results for the first 7 days of stay. Other clinical and radiologic data were not included in this analysis. Deidentified data were obtained in collaboration with the PROTECT III investigators and are now available through the Federal Interagency Traumatic Brain Injury Research informatics system. The inclusion criteria for this analysis were as follows: (1) subjects were alive for at least 3 days, (2) subjects were not excluded from the parent study, and (3) their baseline laboratory results were available. The 3-day criterion is used because subjects who do not survive for at least 3 days or 72 hours have likely experienced devastating brain injury or other forms of severe trauma, which often require aggressive interventions [13]. Additionally, the first 72 hours of observation is the time interval used for determining the preliminary effect of the injury and is thus recommended for valid prognostication [14,15].

### Subspace Clustering

Subspace or projective clustering is a clustering method that emphasizes on clustering in subspaces of high-dimensional spaces, that is, it tries to find clusters in smaller subspaces and builds up to form larger clusters by using overlapping subspaces [16]. Subspace clustering can be classified into the following three main categories: density-based approaches, cell-based approaches, and clustering-oriented approaches. Density-based approaches define subspaces in dense areas [17]. In cell-based approaches, subspaces are formed by predefining the width of grid cells and the number of objects within each cell [18]. Clustering-oriented approaches define properties of the entire set of clusters, as opposed to definition of the cluster itself, and then assign objects to the cluster with the most relevant properties [19].

### Density-Based Approach

One of the commonly used clustering algorithms is density-based spatial clustering of applications with noise (DBSCAN) [20]. The key idea of DBSCAN is that after detecting a cluster using density-based grids, it looks at the neighborhood of each cluster point in a defined radius; any point that exists in this radius is contributed to the cluster.

Every cluster  $C$  in a subspace projection is defined by a set of objects  $O$ , that is a subset of database  $DB$  and a set of relevant dimensions  $S$  out of the set of all dimensions  $D$ .

$$C = (O, S), O \subseteq DB, S \subseteq D \quad (1)$$

A clustering result  $R$  is a set of clusters  $k$  found in the respective subspace projections as follows:

$$R = \{C_1, \dots, C_k\}, C_i = (O_i, S_i) \text{ for } i = \{1, \dots, k\} \quad (2)$$

A density-based subspace cluster  $(O, S)$  in a two-dimensional space is defined with respect to parameters *minPoints* and  $\epsilon$ -neighborhood  $N_\epsilon(p) = \{q \in DB \mid dist^S(p, q) \leq \epsilon\}$ , where  $dist^S$  represents a distance function constrained to the dimensions  $S$ , as follows [20]:

(1)  $\epsilon$ -neighborhood of a point: Let  $p$  and  $q$  be two points of the sample, and the distance equation between these two points is defined by  $dist(p, q)$ . The distance could be defined with Manhattan distance, Euclidean distance, or other different distance methods. The  $\epsilon$ -neighborhood of a point is defined as follows:

$$N_{\epsilon}(P) = \{q \in DB | dist(p, q) \leq \epsilon\} \quad (3)$$

(2) *Directly density reachable*: A point  $p$  is directly density reachable from a point  $q$  with respect to  $\epsilon$  and  $MinPts$  if

$$p \in N_{\epsilon}(q) \text{ and } |N_{\epsilon}(P)| \geq MinPts \quad (4)$$

(3) *Density reachable*: A point  $p$  is density reachable from a point  $q$  with respect to  $\epsilon$  and  $MinPts$  if all the points in a chain of points (including  $q$  and  $p$ ) are directly density reachable from each another.

(4) *Density connected*: A point  $p$  is density connected to a point  $q$  if only there is point  $o$ , which both  $p$  and  $q$  are density reachable from.

(5) *Noise*: The sets of points in database  $DB$  that are not assigned to any cluster are called noise.

To find a cluster, the DBSCAN algorithm starts with a random point  $p$  and finds all density reachable points with respect to  $\epsilon$  and  $MinPts$ . DBSCAN also merges two clusters together if the distance between two sets of points is defined as follows:

$$(S_1, S_2) = \min\{dist(p, q) | p \in S_1, q \in S_2\} \quad (5)$$

Density-connected subspace clustering (SUBCLU) is a greedy algorithm built on an adaption of the DBSCAN algorithm for high-dimensional data. It computes all density-connected sets hidden in subspaces of high-dimensional data. Studies have shown that SUBCLU can outperform other subspace clustering methods based on different measures [18,20,21]. SUBCLU is capable of detecting arbitrarily shaped clusters using the DBSCAN algorithm in subspaces. To use DBSCAN in each subspace, let  $DB$  be a  $d$ -dimensional feature vectors data set with  $n$  objects  $DB \subseteq R^d$ . Let  $A = \{a_1, a_2, \dots, a_d\}$  be the set of all attributes  $a$  of  $DB$ . Any subset  $S \subseteq A$  is called a subspace. The projection of an object  $o$  into a subspace  $S$  is denoted by  $\pi_S(o)$ , and the distance function is denoted by  $dist$ . For instance, the  $\epsilon$ -neighborhood of  $o$  in  $S$  is the same as DBSCAN, but projected in  $S$  subspace as follows:

$$N_{\epsilon}^S(o) = \{x \in DB | dist(\pi_S(o), \pi_S(x)) \leq \epsilon\} \quad (6)$$

The core object is defined as follows:

$$CORE_{\epsilon, m}^S(o) \Leftrightarrow |N_{\epsilon}^S(o)| \geq m \quad (7)$$

The algorithm begins by generating all one-dimensional clusters using the DBSCAN algorithm. For each detected cluster, it checks whether the cluster also exists in higher dimensions or not. For each  $k$ -dimensional subspace  $S \in S_k$ , the algorithm searches all other  $k$ -dimensional subspaces  $T \in S_k$  having  $(k-1)$  attributes in common and combines them to generate  $(k+1)$ -dimensional candidate subspaces. Based on prior studies [21], we chose the  $Midpts$  to be in the range from 8 to 128 (with

five steps) and the  $\epsilon$ -neighborhood to be from 0.01 to 0.25 (with nine steps). For this study, the initial  $Midpts$  value was set to 8 and increased by 30 after each run until it reached 128. The  $\epsilon$ -neighborhood value was initially set to 0.01 and was increased by 0.03 until a maximum of 0.25.

### Cell-Based Approach

Cell-based clustering is centered on cell estimate of the data space. The width of the cells is parametrized by  $w$ . A cluster  $R$  contains a set of cells, and each cell contains at least  $\tau$  number of data points. One of the popular cell-based methods is the *MineClus* algorithm, which describes each of these cells as the objects of the cluster by a hypercube with width  $w$ . These hypercubes are arbitrarily positioned to define a region with frequent data patterns.

A cell-based subspace cluster  $(O, S)$  is defined with respect to the minimum number of objectives  $\tau$  in cells  $CS$  of  $w$  width specified by intervals  $I_i$  per dimension  $\forall i \in S$ . Each interval is part of the common domain  $I_i = [l_i \dots u_i] \subseteq [0 \dots v]$  with lower and upper bounds  $l_i$  and  $u_i$ . For all irrelevant dimensions  $\forall j \in D \setminus S$ , the interval is the full domain  $I_j = [0 \dots v]$ , and the cluster objects  $O = \{o | o \in DB \cap CS\}$  fulfill  $|O| \geq \tau$  [21].

### Clustering-Oriented Approach

Clustering-oriented approaches focus on the clustering result  $R$  by specifying objective functions. PROCLUS [22], one of the first top-down subspace clustering algorithms, forms the clusters first and iteratively improves the clustering model. In the PROCLUS algorithm, the number of clusters and the average dimensionality are used as parameters, and data are partitioned into  $k$  clusters with the average dimension being  $l$ . A clustering-oriented approach is defined with respect to objective function  $f(R)$ , which is based on the entire clustering result  $R$ , and an optimal value parameter  $optF$  is a result set  $R$  with  $f(R) = optF$ .

In this case study, we adapted the aforementioned subspace clustering techniques to analyze the PROTECT III data set. Analyses were performed using OpenSubspace [21,23], an open-source framework that extends the WEKA platform [24,25]. All laboratory values were normalized to a scale between 0 and 10 before applying the algorithms.

### Evaluation

Evaluation of unsupervised learning methods, such as cluster analysis, is typically informed by domain expertise. For this work, two clinicians (coauthors of this work [BF and JR]) independently evaluated the results and validated the clusters based on their experiences in the clinical management of TBI as well as clinical research in neurotrauma. The informatician on the team (VS) coordinated the clinician validation process. Mechanistic interpretations for potential markers or associations indicated by clusters were offered based on clinical expertise. To demonstrate alignment of subclustering solutions to a clinical outcome, mortality at 6 months after injury was examined.

Additional evaluation metrics used in this study included F1 score, entropy, coverage, average dimension, and accuracy of classification. The F1 value, a common metric for evaluating

clustering algorithms, is defined as the harmonic mean of precision and recall. Entropy is a metric that accounts for clarity of clustering [26]. Coverage characterizes how clusters cover the input data space. Average dimension is the average of number of dimensions that the clusters cover in each run. Accuracy of classification compares the patterns detected in the model in relation to labeled data, such as outcome. Here, the mortality status of TBI patients was used as the outcome. Finally, the performance of subspace clustering algorithms was compared to traditional k-means clustering, which partitions  $n$  data points into  $k$  clusters, placing each observation in one of the clusters with nearest mean representation. While k-means rely on distance metrics and proximity of observations within individual clusters, subspace methods group data points based on their lower-dimensional subspaces. Given these distinct algorithmic differences between subspace and k-means

clustering in formulation of the clustering problem, a direct comparison of the clusters formed and interpretation of clusters may not be appropriate. Instead, we report performance metrics for comparison purposes.

## Results

### Subject Characteristics

Of the 882 study subjects in the parent PROTECT III trial, 643 subjects met the inclusion criteria for this study. Table 1 shows the characteristics of these study subjects at baseline. Ten different laboratory results were used in this study, including blood serum chemistry and hematology results at baseline (Table 2). Coagulation tests, such as the international normalized ratio (INR) and activated partial thromboplastin time, were also included.

**Table 1.** Patient characteristics.

Characteristic	Value (N=643)
Age (years), mean (range)	34 (17-93)
Male sex, n (%)	475 (73.9)
Black people, n (%)	105 (16.3)
Hispanic people, n (%)	97 (15.1)
<b>Cause of injury, n (%)</b>	
Motor vehicle accident	242 (37.7)
Motorcycle or scooter accident	121 (18.8)
Pedestrian struck by a moving vehicle	78 (12.1)
Other	202 (31.4)

**Table 2.** Laboratory results.

Laboratory parameter	Value, mean (range)
Glucose, mg/dL	151.6 (68-554)
Creatinine, mg/dL	1.015 (0.3-4.2)
Potassium, mmol/L	3.667 (1.5-5.8)
Sodium, mmol/L	139.8 (125-157)
Chloride, mmol/L	105.4 (88-130)
Bicarbonate, mmol/L	22.77 (8.0-34.0)
Hemoglobin, g/dL	13.66 (4.9-18.6)
Hematocrit, %	40.31 (14.6-54.2)
Total white blood cell count, $\times 10^9/L$	14.85 (3.2-41.40)
Platelet count, $\times 10^3/mm^3$	249.7 (51-700)

### Application of Subspace Clustering Algorithms to PROTECT III Data

All three types of subspace clustering algorithms (density-based [SUBCLU], cell-based [MineClus], and clustering-oriented [PROCLUS] algorithms) were applied to the PROTECT III data set. The INR, which characterizes the clotting tendency of blood, was identified as one of the distinct clusters. This could represent coagulopathy, a marker of secondary insult in TBI

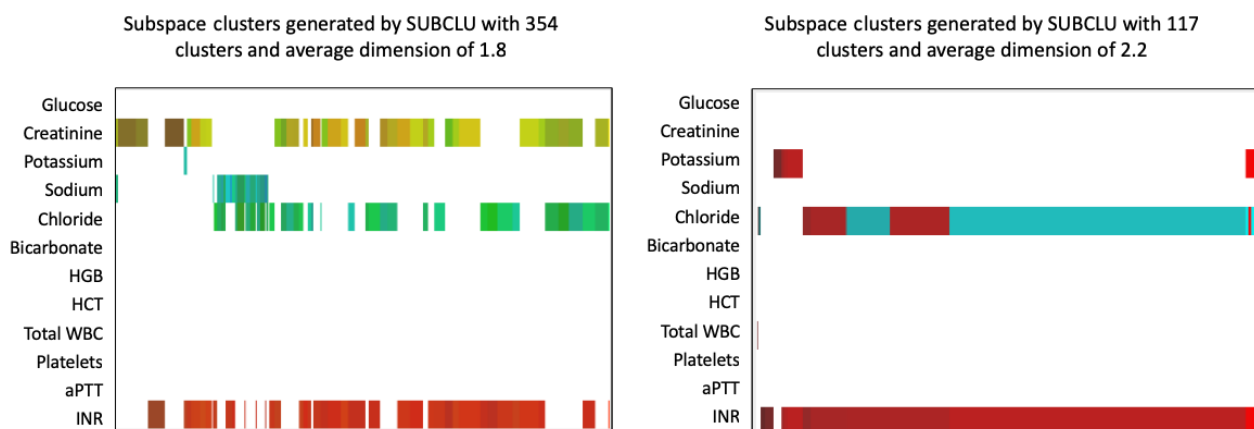
patients [27]. For example, coagulopathy is associated with increased risk of ongoing bleeding and expansion of any intracranial traumatic hemorrhage. One of the clinicians also noted that progressive coagulopathy, which is resistant to correction, is further associated with worse outcomes in TBI patients.

The clustering models also showed a strong relation among INR, chloride, and creatinine. Both clinicians noted and agreed

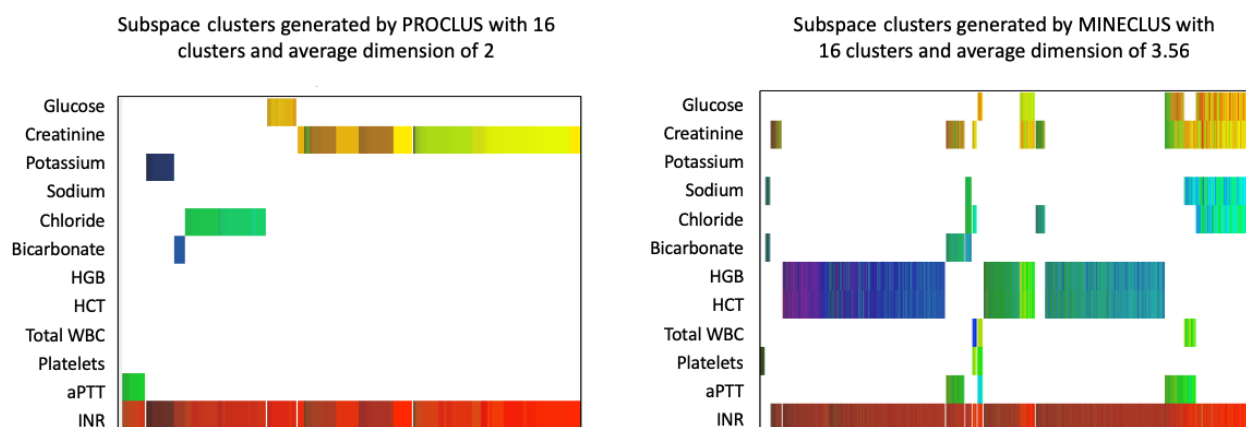
that elevations in chloride levels are often related to fluid administration for treatment of intracranial hypertension or a shock (hypoperfusion) state. Therefore, elevations in these parameters may also be indicators that the clinical team needed to treat a sicker patient more aggressively. Creatinine may be elevated at baseline in patients with chronic illness or may indicate that secondary kidney failure may impact outcome as a complication of TBI. One of the clinicians noted that there is

a further relationship between elevated chloride and subsequent elevation in creatinine, though a wide variety of insults may lead to elevations in creatinine. Finally, in models with higher dimensions, a relationship between the hemoglobin level and hematocrit percentage was noted. This relationship is quite intuitive, given that both measure similar properties. These observations are demonstrated in Figure 1 and Figure 2.

**Figure 1.**



**Figure 2.**



The performances of different subspace clustering methods as well as the traditional k-means algorithm on the PROTECT III data set were compared using various evaluation metrics and the mortality status as the outcome (Table 3). The density-based algorithm (SUBCLU) had higher F1 and coverage. The cell-based algorithm (MineClus) had good performance on the

F1 measure while having lower number of clusters. The clustering-oriented algorithm (PROCLUS) performed reasonably in terms of accuracy and entropy, while it had the lowest F1 compared to other models. K-means, given its simplicity, was the fastest algorithm, but performed worst in all other metrics.



**Table 3.** Comparison of subspace clustering algorithms.

Evaluation metric	Density-based algorithm (SUBCLU <sup>a</sup> ), min-max	Cell-based algorithm (MineClus), min-max	Clustering-oriented algorithm (PROCLUS), min-max	K-means, min-max
F1	0.45-0.69	0.42-0.64	0.36-0.44	0.21-0.30
Entropy	0.45-0.59	0.44-0.55	0.48-0.63	0.51-0.53
Coverage	0.9-1	0.78-0.97	0.43-0.82	0.11-1
Number of clusters	6-1024	6-64	8-32	2-32
Average dimensions	2.3-9	3.2-6.1	2-9	12
Accuracy (%)	81-88	88-88	88-88	51-63
Runtime (s)	367-745,785	58-194	155-402	0.07

<sup>a</sup>SUBCLU: density-connected subspace clustering.

## Discussion

Currently, clinical data used to predict outcomes after TBI come from modeling and validation performed across two older clinical studies in TBI encompassing more than 15,000 patients [28,29]. The covariates that were significant in these prior regression models included glucose and hemoglobin, in addition to clinical predictors such as age and clinical examination. However, the area under the curve of these models is suboptimal. Clusters of data may also incorporate clinical knowledge such as the observation that the combination of lactic acidosis, hypothermia, and coagulopathy at presentation after major trauma imparts poor prognosis. Furthermore, many of these patients do not survive the 72 hours required for inclusion in the current analysis.

Lack of access to multiple data sources has limited further external validation of the proposed methods. Nonetheless, clinician validation is important to inform analyses of data from ongoing observational studies and provide valuable insights

into the development of clinically relevant tools for TBI management. This case study serves as a demonstration for such applications. As a next step, focus on temporal data and methods for time-series analyses are warranted.

## Conclusion

This study explored the application and feasibility of subspace clustering techniques for a specific clinical condition, TBI, using clinical data from a randomized clinical trial. The analyses showed the following three clusters of laboratory physiological data: (1) INR, (2) INR, chloride, and creatinine, and (3) hemoglobin and hematocrit. While all subclustering algorithms had a reasonable accuracy in classifying patients by mortality status, the density-based algorithm had a higher F1 score and coverage. Clustering approaches serve as an important step for phenotype definition and validation in clinical domains, such as TBI, where patient and injury heterogeneity are among the major reasons for failure of clinical trials. Results from this study also provide a foundation to develop scalable clustering algorithms for further research and validation.

## Acknowledgments

The authors acknowledge the Progesterone for Traumatic Brain Injury, Experimental Clinical Treatment–Phase III (PROTECT III) principal investigator, Dr David Wright (Emory University), for providing deidentified data for this study. This material is based upon work supported by the National Science Foundation under grants #1838730 and #1838745. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## Conflicts of Interest

None declared.

## References

1. Taylor CA, Bell JM, Breiding MJ, Xu L. Traumatic Brain Injury-Related Emergency Department Visits, Hospitalizations, and Deaths - United States, 2007 and 2013. *MMWR Surveill Summ* 2017 Mar 17;66(9):1-16 [FREE Full text] [doi: [10.15585/mmwr.ss6609a1](https://doi.org/10.15585/mmwr.ss6609a1)] [Medline: [28301451](https://pubmed.ncbi.nlm.nih.gov/28301451/)]
2. Ghajar J. Traumatic brain injury. *The Lancet* 2000 Sep;356(9233):923-929. [doi: [10.1016/s0140-6736\(00\)02689-1](https://doi.org/10.1016/s0140-6736(00)02689-1)]
3. Subbian V, Ratcliff JJ, Korfhagen JJ, Hart KW, Meunier JM, Shaw GJ, et al. A Novel Tool for Evaluation of Mild Traumatic Brain Injury Patients in the Emergency Department: Does Robotic Assessment of Neuromotor Performance Following Injury Predict the Presence of Postconcussion Symptoms at Follow-up? *Acad Emerg Med* 2016 Apr 21;23(4):382-392 [FREE Full text] [doi: [10.1111/acem.12906](https://doi.org/10.1111/acem.12906)] [Medline: [26806406](https://pubmed.ncbi.nlm.nih.gov/26806406/)]

4. Subbian V, Ratcliff J, Meunier J, Korfhagen J, Beyette F, Shaw G. Integration of New Technology for Research in the Emergency Department: Feasibility of Deploying a Robotic Assessment Tool for Mild Traumatic Brain Injury Evaluation. *IEEE J. Transl. Eng. Health Med* 2015 Apr 23;1-1. [doi: [10.1109/jtehm.2015.2424224](https://doi.org/10.1109/jtehm.2015.2424224)]
5. Nielson JL, Cooper SR, Yue JK, Sorani MD, Inoue T, Yuh EL, TRACK-TBI Investigators. Uncovering precision phenotype-biomarker associations in traumatic brain injury using topological data analysis. *PLoS One* 2017 Mar 3;12(3):e0169490 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0169490](https://doi.org/10.1371/journal.pone.0169490)] [Medline: [28257413](https://pubmed.ncbi.nlm.nih.gov/28257413/)]
6. Subbian V, Meunier J, Korfhagen J, Ratcliff J, Shaw G, Beyette F. Quantitative assessment of post-concussion syndrome following mild traumatic brain injury using robotic technology. 2014 Presented at: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 26-30, 2014; Chicago, IL, USA p. 5353-5356. [doi: [10.1109/EMBC.2014.6944835](https://doi.org/10.1109/EMBC.2014.6944835)]
7. de Toledo P, Rios PM, Ledezma A, Sanchis A, Alen JF, Lagares A. Predicting the outcome of patients with subarachnoid hemorrhage using machine learning techniques. *IEEE Trans Inf Technol Biomed* 2009 Sep;13(5):794-801. [doi: [10.1109/TITB.2009.2020434](https://doi.org/10.1109/TITB.2009.2020434)] [Medline: [19369161](https://pubmed.ncbi.nlm.nih.gov/19369161/)]
8. Hemphill JC, Andrews P, De Georgia M. Multimodal monitoring and neurocritical care bioinformatics. *Nat Rev Neurol* 2011 Jul 12;7(8):451-460. [doi: [10.1038/nrneurol.2011.101](https://doi.org/10.1038/nrneurol.2011.101)] [Medline: [21750522](https://pubmed.ncbi.nlm.nih.gov/21750522/)]
9. Cohen MJ, Grossman AD, Morabito D, Knudson MM, Butte AJ, Manley GT. Identification of complex metabolic states in critically injured patients using bioinformatic cluster analysis. *Crit Care* 2010;14(1):R10 [[FREE Full text](#)] [doi: [10.1186/cc8864](https://doi.org/10.1186/cc8864)] [Medline: [20122274](https://pubmed.ncbi.nlm.nih.gov/20122274/)]
10. Halder P, Pavord ID, Shaw DE, Berry MA, Thomas M, Brightling CE, et al. Cluster Analysis and Clinical Asthma Phenotypes. *Am J Respir Crit Care Med* 2008 Aug;178(3):218-224. [doi: [10.1164/rccm.200711-1754oc](https://doi.org/10.1164/rccm.200711-1754oc)]
11. Elhamifar E, Vidal R. Sparse Subspace Clustering: Algorithm, Theory, and Applications. *IEEE Trans. Pattern Anal. Mach. Intell* 2013 Nov;35(11):2765-2781. [doi: [10.1109/tpami.2013.57](https://doi.org/10.1109/tpami.2013.57)]
12. Wright DW, Yeatts SD, Silbergleit R, Palesch YY, Hertzberg VS, Frankel M, et al. Very Early Administration of Progesterone for Acute Traumatic Brain Injury. *N Engl J Med* 2014 Dec 25;371(26):2457-2466. [doi: [10.1056/nejmoa1404304](https://doi.org/10.1056/nejmoa1404304)]
13. Jochems D, van Wessel KJP, Houwert RM, Brouwers HB, Dankbaar JW, van Es MA, et al. Outcome in Patients with Isolated Moderate to Severe Traumatic Brain Injury. *Crit Care Res Pract* 2018 Sep 23;2018:3769418-3769417 [[FREE Full text](#)] [doi: [10.1155/2018/3769418](https://doi.org/10.1155/2018/3769418)] [Medline: [30345113](https://pubmed.ncbi.nlm.nih.gov/30345113/)]
14. Eriksson EA, Barletta JF, Figueroa BE, Bonnell BW, Sloffer CA, Vanderkolk WE, et al. The first 72 hours of brain tissue oxygenation predicts patient survival with traumatic brain injury. *The Journal of Trauma and Acute Care Surgery* 2012;72(5):1345-1349. [doi: [10.1097/ta.0b013e318249a0f4](https://doi.org/10.1097/ta.0b013e318249a0f4)]
15. Souter MJ, Blissitt PA, Blosser S, Bonomo J, Greer D, Jichici D, et al. Recommendations for the Critical Care Management of Devastating Brain Injury: Prognostication, Psychosocial, and Ethical Management : A Position Statement for Healthcare Professionals from the Neurocritical Care Society. *Neurocrit Care* 2015 Aug 18;23(1):4-13. [doi: [10.1007/s12028-015-0137-6](https://doi.org/10.1007/s12028-015-0137-6)] [Medline: [25894452](https://pubmed.ncbi.nlm.nih.gov/25894452/)]
16. Theeramunkong T, Kijssirikul B, Cercone N, Ho TB. *Advances in Knowledge Discovery and Data Mining*. Berlin, Germany: Springer; 2009.
17. Chu Y, Huang J, Chuang K, Yang D, Chen M. Density Conscious Subspace Clustering for High-Dimensional Data. *IEEE Trans. Knowl. Data Eng* 2010 Jan;22(1):16-30. [doi: [10.1109/TKDE.2008.224](https://doi.org/10.1109/TKDE.2008.224)]
18. Sembiring R, Zain J. Cluster Evaluation of Density Based Subspace Clustering. *J Comput* 2010;2(11):2151-9617 [[FREE Full text](#)]
19. Aggarwal CC, Wolf JL, Yu PS, Procopiuc C, Park JS. Fast algorithms for projected clustering. *SIGMOD Rec* 1999 Jun 01;28(2):61-72. [doi: [10.1145/304181.304188](https://doi.org/10.1145/304181.304188)]
20. Ester M, Krieger H, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. 1996 Aug Presented at: Second International Conference on Knowledge Discovery and Data Mining; August 2-4, 1996; Portland, Oregon p. 226-231. [doi: [10.5555/3001460.3001507](https://doi.org/10.5555/3001460.3001507)]
21. Müller E, Günnemann S, Assent I, Seidl T. Evaluating clustering in subspace projections of high dimensional data. *Proc. VLDB Endow* 2009 Aug 01;2(1):1270-1281. [doi: [10.14778/1687627.1687770](https://doi.org/10.14778/1687627.1687770)]
22. Agrawal R, Gehrke J, Gunopulos D, Raghavan P. Automatic subspace clustering of high dimensional data for data mining applications. 1998 Jun Presented at: 1998 ACM SIGMOD International Conference on Management of Data; June 1998; Seattle, WA, USA p. 94-105. [doi: [10.1145/276305.276314](https://doi.org/10.1145/276305.276314)]
23. Assent I, Müller E, Krieger R, Jansen T, Seidl T. Pleiades: Subspace Clustering and Evaluation. In: Daelemans W, Goethals B, Morik K, editors. *Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2008. Lecture Notes in Computer Science*, vol 5212. Berlin, Heidelberg: Springer; 2008.
24. Witten I, Frank E, Mark H. *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington, MA: Morgan Kaufmann; 2011.
25. Frank E, Hall M, Witten I. The WEKA Workbench. In: *Data Mining (Fourth Edition)*. Burlington, MA: Morgan Kaufmann; 2016:553-571.
26. Sequeira K, Zaki M. SCHISM: a new approach to interesting subspace mining. *IJBIDM* 2005;1(2):137. [doi: [10.1504/ijbidm.2005.008360](https://doi.org/10.1504/ijbidm.2005.008360)]

27. Stein SC, Smith DH. Coagulopathy in Traumatic Brain Injury. NCC 2004;1(4):479-488. [doi: [10.1385/ncc.1.4.479](https://doi.org/10.1385/ncc.1.4.479)]
28. Murray GD, Butcher I, McHugh GS, Lu J, Mushkudiani NA, Maas AI, et al. Multivariable prognostic analysis in traumatic brain injury: results from the IMPACT study. J Neurotrauma 2007 Feb;24(2):329-337. [doi: [10.1089/neu.2006.0035](https://doi.org/10.1089/neu.2006.0035)] [Medline: [17375997](https://pubmed.ncbi.nlm.nih.gov/17375997/)]
29. Steyerberg EW, Mushkudiani N, Perel P, Butcher I, Lu J, McHugh GS, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. PLoS Med 2008 Aug 05;5(8):e165; discussion e165 [FREE Full text] [doi: [10.1371/journal.pmed.0050165](https://doi.org/10.1371/journal.pmed.0050165)] [Medline: [18684008](https://pubmed.ncbi.nlm.nih.gov/18684008/)]

## Abbreviations

**DBSCAN:** density-based spatial clustering of applications with noise

**INR:** international normalized ratio

**PROTECT III:** Progesterone for Traumatic Brain Injury, Experimental Clinical Treatment–Phase III

**SUBCLU:** density-connected subspace clustering

**TBI:** traumatic brain injury

*Edited by G Eysenbach; submitted 01.10.20; peer-reviewed by H Zhang, G Lim; comments to author 23.10.20; revised version received 31.12.20; accepted 16.01.21; published 02.02.21*

*Please cite as:*

*Ehsani S, Reddy CK, Foreman B, Ratcliff J, Subbian V*

*Subspace Clustering of Physiological Data From Acute Traumatic Brain Injury Patients: Retrospective Analysis Based on the PROTECT III Trial*

*JMIR Biomed Eng 2021;6(1):e24698*

*URL: <http://biomedeng.jmir.org/2021/1/e24698/>*

*doi: [10.2196/24698](https://doi.org/10.2196/24698)*

*PMID:*

©Sina Ehsani, Chandan K Reddy, Brandon Foreman, Jonathan Ratcliff, Vignesh Subbian. Originally published in JMIR Biomedical Engineering (<http://biomedeng.jmir.org>), 02.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on <http://biomedeng.jmir.org/>, as well as this copyright and license information must be included.