Sensors and Materials, Vol. 34, No. # (2022) 1–##
MYU Tokyo

S & M ####

Applying Machine Learning to Determine the Behavioral Characteristics of Rodents with Traumatic Brain Injury in an Eight-arm Maze

Shu-Cing Wu,[1] Chi-Yuan Lin,[1,2*] Liang-Jyun Hong,[3] and Chi-Chun Chen[3*]

[1]Ph.D. Program, Prospective Technology of Electrical Engineering and Computer Science, National Chin-Yi University of Technology, No. 57, Sec. 2, Chung shan Rd., Taiping District, Taichung City 41170, Taiwan
[2]Department of Computer Science and Information Engineering, National Chin-Yi University of Technology, No. 57, Sec. 2, Chung shan Rd., Taiping District, Taichung City 41170, Taiwan
[3]Department of Electronic Engineering, National Chin-Yi University of Technology, No. 57, Sec. 2, Chung shan Rd., Taiping District, Taichung City 41170, Taiwan

(Received November 4, 2021; accepted January 20, 2022; online published, March 9 2022)

Keywords: cognitive parameters, eight-arm maze, machine learning, traumatic brain injury, support vector machine, decision tree, random forest, k-nearest neighbor

In this study, we identified the cognitive parameters of rats with traumatic brain injury (TBI) in an eight-arm radial maze and used them for TBI classification through machine learning models. A total of 16 cognitive parameters were derived using a sensing trajectory bitmap in the eight-arm maze. Of these 16 parameters, five (i.e., short-term memory error, latency, total distance, frequency of movement from an arm without food to an arm with food, and frequency of entry into the arm on the right after exiting an arm) were selected as representative parameters and were input into four machine learning models, namely, support vector machine (SVM), decision tree, random forest, and k-nearest neighbor (KNN) models, to classify and compare sham rats and rats with TBI. The performance evaluation results for the machine learning models revealed that the SVM model had the best performance among the models. Its overall accuracy, sensitivity, and area under the receiver operating characteristic curve (AUC) were >85, 98, and >94%, respectively. At some postsurgical time points, the sensitivity and AUC of the SVM model even approached 100%. The random forest and KNN models had satisfactory performance on Day 28 postsurgery. Overall, the SVM model had satisfactory performance in classifying both mild and severe TBI. Our findings can serve as a reference for future research on TBI feature classification.

1. Introduction

A traumatic brain injury (TBI), also known as a head injury, refers to an injury to the brain caused by an external physical impact, such as that from a road traffic accident, fall, or physical altercation. Its clinical manifestations are a concussion, a cerebral contusion, and an intracranial hemorrhage.(1) Among patients with TBI, 80% have mild TBI.(2) In the early stage of TBI, patients

*Corresponding author: e-mail: chichun@ncut.edu.tw

ISSN 0914-4935 © MYU K.K.

do not experience obvious symptoms; therefore, they often overlook the subsequent effects of the injury. In recent years, studies have indicated that TBI is a risk factor for Alzheimer disease[3] because TBI shares numerous pathological features with Alzheimer disease, including brain amyloidosis, tangled nerve fibers, neuronal synaptic function loss, and cognitive impairment.[4] In several studies, basic experiments to assess TBI-related effects on learning and memory in rats have been carried out.[5,6] Hence, cognitive assessment instruments for evaluating learning and memory and for verifying the effectiveness of related treatments are essential. Our research team has recently developed an automated eight-arm radial maze for assessing cognition; this maze system considerably simplifies manual observation and recording procedures. In this study, we applied machine learning to examine rats' behavioral characteristics in this eight-arm maze and to determine the differences in characteristics between sham rats and rats with TBI.

The Barnes maze,[7] Y-maze,[8] T-maze,[9] Morris water maze,[10] and radial arm maze[11] are commonly used for assessing learning and memory abilities. Among these, the Morris water and radial arm mazes are considered the most reliable memory assessment apparatuses, primarily because they can be used to detect the hippocampus-dependent learning and memory abilities of rats.[12] In a water maze, a standing platform is placed at a point below the water surface, and rats are compelled to remember the location of the platform to avoid drowning. In a radial eight-arm maze, food is placed at the end of certain arms to determine the spatial memory ability of rats. Accordingly, the training time required for a water maze is considerably shorter than that required for a radial eight-arm maze. However, in a water maze, experimental results usually vary substantially because of differences in rat survival instincts.[13] By contrast, a radial eight-arm maze can be effectively used to evaluate consistency through repeated testing;[13] the derived experimental results for the same pathological characteristic do not vary substantially. Accordingly, in this present study, we applied the radial eight-arm maze as the basic testing tool to assess the spatial learning and memory abilities and cognitive behavioral characteristics of rats with TBI.

Machine learning is a family of algorithms often used to solve classification problems.[14] Such algorithms reveal patterns through an automatic analysis of data and through learning. Predictions about unknown data are then made according to these revealed patterns. Machine learning classification models can be roughly divided into linear and nonlinear models. Linear models can handle only simple classification problems; therefore, most studies have instead employed nonlinear models to solve relatively complex problems. For example, deep learning has been applied for classifying and tracking multiple animal species.[15] Crisler *et al.* used a support vector machine (SVM) model to integrate electrocorticographic signals to identify different stages of sleep in rats.[16] By using hemodynamic, cardiac function, and blood gas parameters, Lucas *et al.* established a logistic regression model for predicting recovery from hemorrhagic shock.[17] Lysenko-Martin *et al.* processed age and concussion data using three-dimensional multiple-object tracking and constructed a decision tree to assess patients' concussion status.[18] Haveman *et al.* applied a random forest model to analyze electroencephalographic signals in order to predict the outcome of patients with moderate to severe TBI.[19] Othman *et al.* employed a k-nearest neighbor (KNN) algorithm along with near-infrared spectroscopy and electroencephalography to identify acute brain injury in patients.[20]

Various machine learning models have been reported to have high accuracy in medical decision-making applications.[21,22] Studies have revealed that machine learning can be divided into two stages. In the first stage, relevant machine learning parameters are determined; these parameters are typically related to the final assessment objective. In the second stage, training and verification are performed. In this present study, we executed a basic experiment to observe and analyze cognitive parameters of rats in the eight-arm maze, and we determined the differences between sham rats and rats with TBI using machine learning classification models. On the basis of the classification results, we consider that our study approach and findings can serve as a reference for determining TBI symptoms in the future.

2. Materials and Methods

The procedures for examining the cognitive parameters of the rats in the eight-arm maze are illustrated in Fig. 1. We used our eight-arm maze system to locate and identify the experimental rats. During the experiment, the coordinates of all the rats were noted in an Excel file in chronological order and could be presented as a trajectory bitmap. The cognitive parameters could be obtained using the trajectory bitmap with time information. In addition to conventional parameters, such as long-term memory errors, short-term memory errors, and latency, we calculated the number of arm-to-arm selections (i.e., the number of times a rat moved between arms) and further classified these selections in accordance with whether the arms selected did or did not have food. The number of direction choices that a rat made after exiting an arm was also calculated. Cognitive parameters with greater differences were selected and input into machine learning models to classify sham rats and rats with TBI.

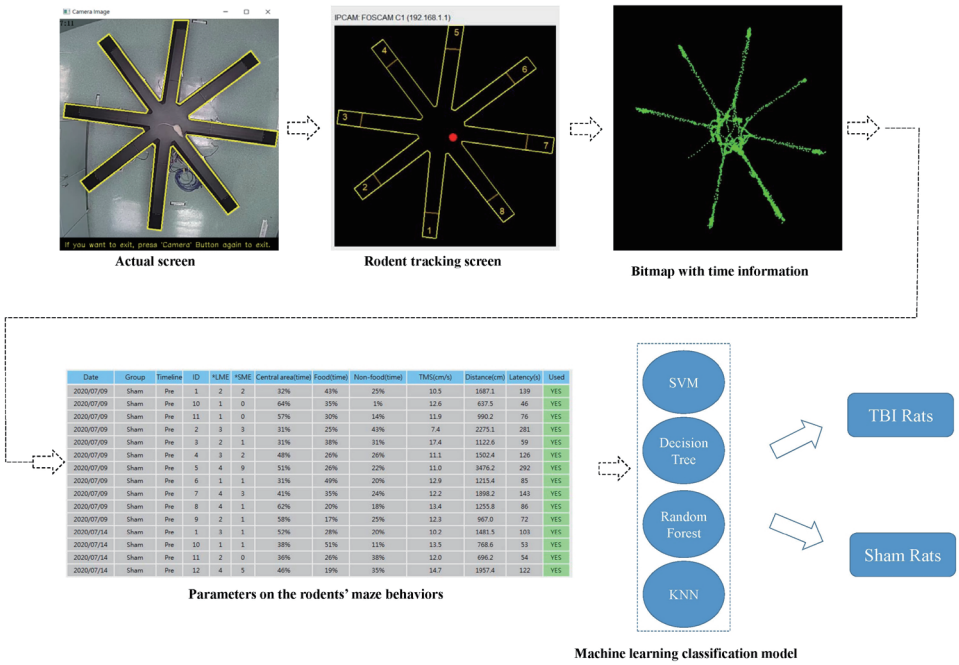

Fig. 1. (Color online) Processing procedures for examining rodents' cognitive parameters in the eight-arm maze.

2.1 Cognitive parameters of rats in eight-arm maze

On the basis of the sensing trajectory data derived in the eight-arm maze, we obtained 16 cognitive parameters (Table 1) that could be divided into two types: those related to direction and those unrelated to direction. The parameters unrelated to direction were total distance, average speed, and percentage of time spent in each area. The parameters related to direction mostly pertained to search strategies, such as the arm-to-arm selections and the direction choice after exiting an arm.

2.2 Machine learning classification model

To classify sham rats and rats with TBI, we applied four commonly used supervised learning models, namely, SVM, decision tree, random forest, and KNN models. Each model is described as follows.

Table 1
Explanations of the cognitive parameters.

Cognitive parameters		Explanation
Direction-irrelevant parameters:		
1.	Long-term memory error (number of times)	Number of times rats enter an arm without food
2.	Short-term memory error (number of times)	Number of times rats reenter an arm
3.	Latency (s)	Total time spent in entering all four arms with food
4.	Total distance (cm)	Total distance walked throughout the experiment
5.	Average speed (cm/s)	Total distance/latency
6.	Percentage of time spent in the central area (%)	Total time spent in the central area/latency
7.	Percentage of time spent in arms with food (%)	Total time spent in the arms with food/latency
8.	Percentage of time spent in arms without food (%)	Total time spent in the arms without food/latency
Direction-relevant parameters:		
9.	Moving from an arm with food to another arm with food (number of times)	Number of times a rat moved from an arm with food to another arm with food following the arm entrance order
10.	Moving from an arm with food to an arm without food (number of times)	Number of times a rat moved from an arm with food to an arm without food following the arm entrance order
11.	Moving from an arm without food to an arm with food (number of times)	Number of times a rat moved from an arm without food to an arm with food following the arm entrance order
12.	Moving from an arm without food to another arm without food (number of times)	Number of times a rat moved from an arm without food to another arm without food following the arm entrance order
13.	Reentering the exit arm (number of times)	Number of times a rat reentered an arm after exiting the same arm following the arm entrance order
14.	Entering the arm on the left (number of times)	Number of times a rat entered the arm on the left after exiting an arm following the arm entrance order
15.	Entering the arm on the right (number of times)	Number of times a rat entered the arm on the right after exiting an arm following the arm entrance order
16.	Entering the arm opposite the front arm (number of times)	Number of times a rat entered the arm opposite the arm it exited following the arm entrance order

2.2.1 SVM model

An SVM model maps data points onto a high-dimensional plane and then identifies a hyperplane that both separates two different data samples and maximizes their distance from the hyperplane (Fig. 2). In a high-dimensional cube, the hyperplane can be considered as a linearly separable, two-dimensional plane. In Fig. 2, the red line that completely separates the two data samples is the segmentation hyperplane. An SVM model can identify the optimal segmentation hyperplane, maximize the distance of data samples from the margin, and obtain a maximum margin.

To identify the optimal segmentation hyperplane, we can define the equation of the red line in Fig. 3 *as* $ax + by + c = 0$ and convert it into matrix form as follows: $w^T x + b = 0$. This equation serves as the equation of the hyperplane. Because the objective of the SVM model is to solve a classification problem, we can define the equation of one of the dotted lines as $w^T x + b - k = 0$ and that of the other dotted line as $w^T x + b + k = 0$, where k is an arbitrary constant. These two equations can be combined to obtain $y(w^T x + b) \geq k$.

Subsequently, the vectors $\mathbf{X_1}$ and $\mathbf{X_2}$ can be projected onto the plane W (Fig. 4), and the boundary length can be derived, as indicated in Eq. (1).

$$\text{Margin} = \frac{w \cdot \overline{x_1 - x_2}}{|\vec{w}|} = \frac{k}{|\vec{w}|} \tag{1}$$

A mapping function must be used to convert low-dimensional coordinates into high-dimensional ones. Accordingly, we used a sigmoid function for conversion.

2.2.2 Decision tree model

A decision tree is a tree-structured algorithm that typically contains three components: a root node, internal nodes, and leaf nodes. In accordance with the concept of a tree structure, beginning from the root node, each internal node represents a feature. The data cluster is divided

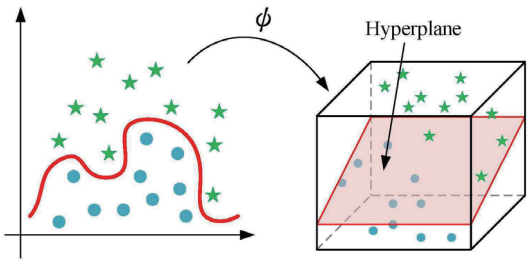

Fig. 2. (Color online) Mapping the coordinate plane onto the hyperplane.

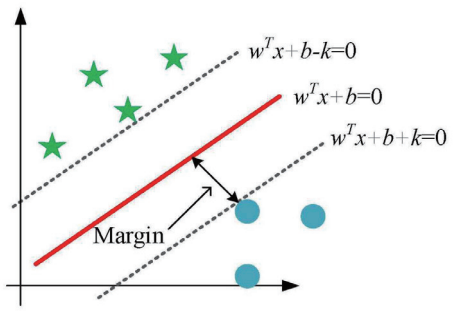

Fig. 3. (Color online) Equation of the segmentation hyperplane.

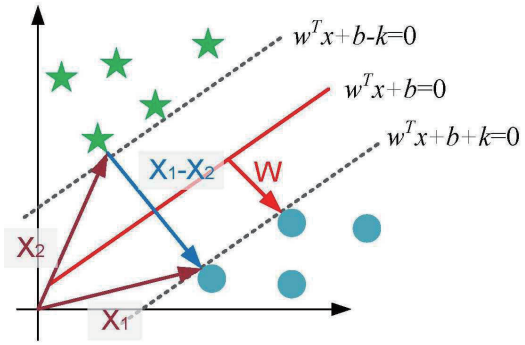

Fig. 4. (Color online) Calculation of the maximum margin.

into two components by each internal node. These components are further divided into other branches until each data point is classified.

In a decision tree, feature selection is based on information gain (IG). IG is the quantity of data that a feature contributes to the classification system, and a higher IG can generate a more satisfactory decision-making result. A feature with a larger quantity of information is more important than other features.

Equation (2) presents the IG equation, where $Info(D_p)$ is the original amount of information, $Info_A(D_p)$ is the amount of information after segmentation, and m is the number of segmentations. Equation (3) presents the binary classification model ($m = 2$).

$$\mathrm{IG}\left(D_p\right) = Info\left(D_p\right) - Info_A\left(D_p\right) = Info\left(D_p\right) - \sum_{j=1}^{m} \frac{N_j}{N_p} Info\left(D_j\right) \tag{2}$$

$$\mathrm{IG}\left(D_p\right) = Info\left(D_p\right) - \frac{N_{left}}{N_p} Info\left(D_{left}\right) - \frac{N_{right}}{N_p} Info\left(D_{right}\right) \tag{3}$$

Entropy and Gini impurity are commonly applied to calculate the amount of information. In this study, we employed entropy for calculation, as presented in Eq. (4), where K is the number of categories a data sample should be divided into and P_k is the score of a category.

$$Info_H = -\sum_{k=1}^{K} P_k \log_2 P_k \tag{4}$$

2.2.3 Random forest model

A random forest is formed by multiple randomly created classification and regression trees. Each decision tree has an independent decision-making process, and the final output result is selected through majority voting or averaging. Decision trees that differ considerably from other

trees are required for calculation; therefore, different data sets should be produced. The most common methods of generating trees with large differences are bagging (bootstrap aggregation) and boosting.

We adopted bagging in this study. We randomly retrieved z samples from the original data samples. After each sampling, z classifiers (trees) were produced. Subsequently, the retrieved samples were returned to the original data population. Although trees shared some samples, the trained classifiers (trees) differed from each other overall. The final classifier (tree) was obtained through voting.

2.2.4 KNN model

In a KNN model, each new data point is compared with its "neighbors" within an area. The data point is then classified into the most common category among the categories of its neighbors. KNN classification involves the following steps.

1. Calculate the distance between the new data point and the training set data points.
2. Determine the k data points with the shortest distance to the new data point.
3. Categorize the new data points by a majority vote among the k nearest neighbors.

Numerous methods are available for distance estimation; among these, the Euclidean distance and Manhattan distance are the most commonly applied methods. In this study, the Euclidean distance was used and can be expressed as

$$D_{\text{Euclidean}} = \sqrt{\left(x_2 - x_1\right)^2 + \left(y_2 - y_1\right)^2}. \tag{5}$$

2.3 Evaluation indices for machine learning

The most commonly used indices for evaluating machine learning classification models are a confusion matrix, the receiver operating characteristic (ROC) curve, and Cohen's kappa.

2.3.1 Confusion matrix

A confusion matrix is used to evaluate the predictive performance of a classifier on the basis of the number of positive (correct) and negative (incorrect) prediction results. Therefore, four conditions exist, as shown in Table 2.

On the basis of the confusion matrix, several evaluation indices for classification models have been developed, as follows.

Table 2
Confusion matrix.

		Actual condition	
		Positive	Negative
Predict condition	Positive	True positive (TP)	False positive (FP)
	Negative	False negative (FN)	True negative (TN)

- Accuracy: Percentage of correct predictions in the overall data set, as shown by Eq. (6).

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) \tag{6}$$

- True-positive rate: Actual rate of positives (i.e., diseased) detected by the model, also referred to as sensitivity or recall, as indicated by Eq. (7).

$$Sensitivity = TP/(TP + FN) \tag{7}$$

- True-negative rate: Actual rate of negatives (i.e., not diseased) detected by the model, also referred to as specificity, as indicated by Eq. (8).

$$Specificity = TN/(TN + FP) \tag{8}$$

- Positive predictive value: Ratio of true disease predictions to all disease predictions, also referred to as precision, as indicated by Eq. (9).

$$Precision = TP/(TP + FP) \tag{9}$$

- Negative predictive value (NPV): Ratio of false nondiseased predictions to all nondiseased predictions, as indicated by Eq. (10).

$$NPV = TN/(FN + TN) \tag{10}$$

2.3.2 ROC curve analysis

An ROC curve is a visual analytic tool, and its horizontal axis represents specificity, whereas its vertical axis represents sensitivity, as illustrated in Fig. 5.

In binary classification, the output of each classification is likely to indicate the probability of a classification rather than directly outputting values of 0 or 1. An ROC curve is a plot of the

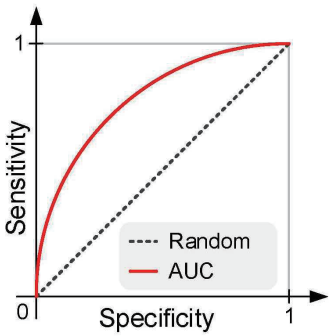

Fig. 5. (Color online) ROC curve.

sensitivity of results against the specificity of results. Through the modification of a threshold value for classifying an output value between 0 and 1 as true, the true-positive rate can be modified. Threshold values associated with higher true-positive rates result in ROC curves further from the diagonal line (i.e., the line representing random guessing); thus, the area under the ROC curve (AUC) is greater for models with higher classification performance. An AUC value of ≥0.8 suggests higher accuracy and greater authenticity.

2.3.3 Cohen's kappa

Cohen's kappa is a measure of the consistency and reliability of ratings provided by two raters on the same topic. If the raters have high consistency, the rating results are sufficiently reliable.

Cohen's kappa can be used to evaluate multivariate classification models. An example in matrix form is presented in Table 3. Each item represents the number of classifications of each type as a percentage; that is, C_{21} represents the number of predictions that an item in Class 2 is predicted as being in Class 1 divided by the total number of predictions. Summing the elements along the diagonal results in the observational consistency P_0, that is, the probability of a correct prediction. P_c is the expected consistency—representing the probability of random guessing—and is calculated by multiplying the total number of items in a class by the total number of predictions for that class (i.e., $S_{A1} \times S_{P1}$ for Class 1 in Table 3) and summing these terms across all classes. Thus, the ratio of $P_0 - P_c$ to $1 - P_c$ represents the performance of the model compared with that of a model with a perfect observational consistency of 1; this ratio is Cohen's kappa. Cohen's kappa ranges between –1 and 1, and higher values imply better consistency. Typically, a value of >0.75 suggests excellent consistency.

2.4 Experimental process

In our memory assessment experiment conducted in the eight-arm maze, the rats were subjected to two weeks of early-stage training five days per week before surgery. The first week was the adaptation phase, which was divided into two stages. In the first two days, no food was placed in the eight-arm maze [Fig. 6(a)]. The rats were allowed to walk freely in the maze for

Table 3
Application matrix for Cohen's kappa.

		Actual condition			Sum
		Class 1	Class 2	Class 3	
Predict condition	Class 1	C_{11}	C_{21}	C_{31}	S_{P1}
	Class 2	C_{12}	C_{22}	C_{32}	S_{P2}
	Class 3	C_{13}	C_{23}	C_{33}	S_{P3}
Sum		S_{A1}	S_{A2}	S_{A3}	Total

$$K = \frac{(P_0 - P_c)}{(1 - P_c)} = \frac{(C_{11} + C_{22} + C_{33}) - (S_{A1}S_{P1} + S_{A2}S_{P2} + S_{A3}S_{P3})}{1 - (S_{A1}S_{P1} + S_{A2}S_{P2} + S_{A3}S_{P3})} \quad (11)$$

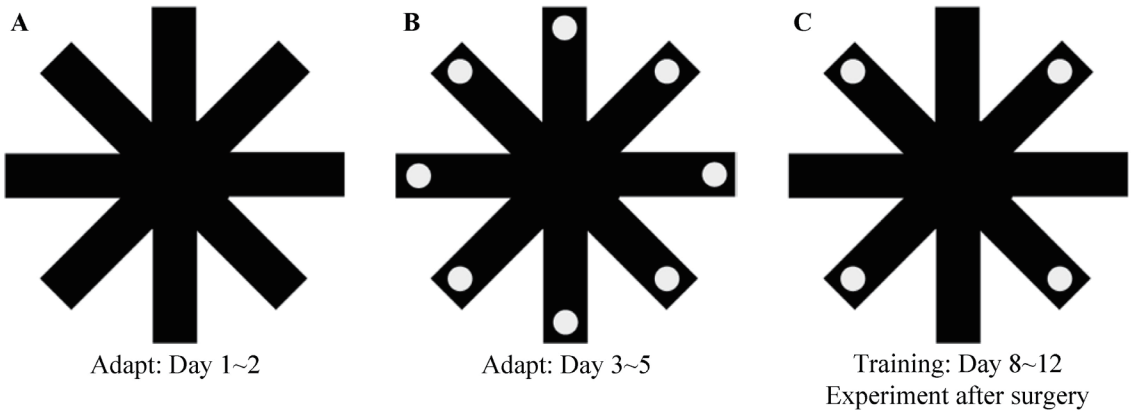

Fig. 6. Food placements in the eight-arm maze before surgery.

Fig. 7. (Color online) Experimental process in eight-arm maze.

5 min to adapt to the maze environment. In the subsequent three days, food was placed at the end of all arms [Fig. 6(b)] to teach the rats to expect and search for food in the maze. The white dots in Fig. 6 indicate the food placements. The second week was the training phase. During this phase, food was placed only at the ends of the four designated arms [Fig. 6(c)]. The training ended after the rats walked through the four arms with food.

On the last day in the training phase, a formal maze test was conducted. Surgery was performed on the next day. Maze tests were conducted on Days 7, 14, and 28 after surgery to investigate the rats' memory ability. The testing method was the same as that in the training phase. After the rats walked through the four arms with food, the timer was stopped immediately, as illustrated in Fig. 7. The animal trials and TBI surgery were both conducted at Chi Mei Medical Center and approved by the relevant Animal Ethics Committee.

3. Results

3.1 Quantified cognitive parameters

The statistical analysis results for all experimental parameters are presented in Table 4. We observed that for the parameters unrelated to direction, the two groups (i.e., sham rats and rats with TBI) differed significantly in terms of the number of short-term memory errors. The number of errors in the TBI group exhibited an increasing trend. The latency of the two groups did not differ considerably on Day 7; however, statistically significant differences were evident on Day 14. The latency of the TBI group increased with time. After surgery, the walking

Table 4
Statistical analysis results for cognitive parameters.

Cognitive parameters	Time points	Sham (n = 8)	TBI (n = 8)
Direction-irrelevant parameters:			
Long-term memory error (times)	Day 7 post-surgery	2.7 ± 0.8	3.08 ± 0.3
	Day 14 post-surgery	2.7 ± 0.9	3.62 ± 0.4
	Day 28 post-surgery	2.9 ± 0.4	4.00 ± 0.0*
Short-term memory error (times)	Day 7 post-surgery	1.1 ± 0.7	4.0 ± 2.0*
	Day 14 post-surgery	1.2 ± 0.4	4.2 ± 1.7*
	Day 28 post-surgery	2.3 ± 1.4	7.6 ± 3.3*
Latency (s)	Day 7 post-surgery	153.1 ± 29.3	170.8 ± 57.1
	Day 14 post-surgery	117 ± 26.3	247 ± 83.3*
	Day 28 post-surgery	171 ± 57.9	490 ± 374.4*
Total distance (cm)	Day 7 post-surgery	1433.6 ± 526.3	2225.4 ± 742.7*
	Day 14 post-surgery	1531.8 ± 550.4	2642.0 ± 802.9*
	Day 28 post-surgery	1807.9 ± 479.3	3337.6 ± 1396.5*
Average speed (cm/s)	Day 7 post-surgery	9.4 ± 2.8	16.2 ± 4.7*
	Day 14 post-surgery	12.9 ± 3.0	10.2 ± 2.5
	Day 28 post-surgery	11.0 ± 2.7	8.4 ± 2.6
Percentage of time spent in the central area (%)	Day 7 post-surgery	54 ± 17	36 ± 10*
	Day 14 post-surgery	56 ± 15	46 ± 17
	Day 28 post-surgery	56 ± 8	40 ± 17*
Percentage of time spent in arms with food (%)	Day 7 post-surgery	17 ± 6	32 ± 8*
	Day 14 post-surgery	23 ± 13	29 ± 13
	Day 28 post-surgery	21 ± 7	29 ± 10
Percentage of time spent in arms without food (%)	Day 7 post-surgery	29 ± 14	32 ± 9
	Day 14 post-surgery	21 ± 9	25 ± 10
	Day 28 post-surgery	23 ± 8	31 ± 11
Direction-relevant parameters:			
Moving from an arm with food to another arm with food (number of times)	Day 7 post-surgery	1.25 ± 0.8	1.8 ± 0.6
	Day 14 post-surgery	1.62 ± 0.9	1.8 ± 1.1
	Day 28 post-surgery	2.12 ± 1.0	1.5 ± 1.3
Moving from an arm with food to an arm without food (number of times)	Day 7 post-surgery	2.25 ± 0.9	3.0 ± 0.8
	Day 14 post-surgery	2.00 ± 0.8	3.6 ± 0.8*
	Day 28 post-surgery	2.12 ± 0.3	4.7 ± 1.5*
Moving from an arm without food to an arm with food (number of times)	Day 7 post-surgery	2.6 ± 0.9	3.6 ± 0.8
	Day 14 post-surgery	2.5 ± 1.0	4.0 ± 0.8*
	Day 28 post-surgery	2.7 ± 0.6	5.3 ± 2.1*
Moving from an arm without food to another arm without food (number of times)	Day 7 post-surgery	1.2 ± 0.8	1.8 ± 0.6
	Day 14 post-surgery	1.6 ± 0.9	1.8 ± 1.1
	Day 28 post-surgery	2.1 ± 1.0	1.5 ± 1.3
Reentering the front arm (number of times)	Day 7 post-surgery	0.0 ± 0.0	0.0 ± 0.0
	Day 14 post-surgery	0.0 ± 0.0	0.0 ± 0.0
	Day 28 post-surgery	0.0 ± 0.0	0.8 ± 0.6*
Entering the arm on the left (number of times)	Day 7 post-surgery	4.0 ± 1.9	3.3 ± 2.1
	Day 14 post-surgery	3.6 ± 2.4	4.0 ± 2.1
	Day 28 post-surgery	4.5 ± 2.7	3.5 ± 1.9
Entering the arm on the right (number of times)	Day 7 post-surgery	2.3 ± 2.2	6.7 ± 3.3*
	Day 14 post-surgery	2.5 ± 2.4	8.1 ± 1.9*
	Day 28 post-surgery	3.3 ± 1.6	9.3 ± 3.0*
Entering the arm opposite to the front arm (number of times)	Day 7 post-surgery	0.5 ± 0.5	0.7 ± 1.3
	Day 14 post-surgery	0.8 ± 0.6	0.7 ± 1.0
	Day 28 post-surgery	1.1 ± 1.0	0.8 ± 1.7

Values are means ± S.D. Student's t-test was used to compare variables for two groups.
*$p < 0.05$ indicated statistical significance compared with the sham group.

distance of the TBI group at each time point was significantly farther than that of the sham group.

For the parameters related to direction, we observed that the rats most frequently moved from "an arm with food to an arm without food" and "from an arm without food to an arm with food," and these movements thus constituted the most frequent movement types. The frequencies of these movements differed significantly between the two groups on Days 14 and 28. For all tests after surgery, the rats in the TBI group more frequently entered the arm on the right after exiting from an arm.

On the basis of the preceding statistical analysis results, we selected five parameters (i.e., short-term memory error, latency, total distance, frequency of movement to the arms with food, and frequency of movement to the arm on the right) as the feature parameters for machine learning model training.

3.2 Analytical results for machine learning

The experimental data for the sham group and the TBI group were collected. Data were collected at three time points after surgery (Days 7, 14, and 28) for each group, and eight data values were collected at each time point (one for each rat); thus, a total of 48 data values were obtained. The data were partitioned into a training set and a test set at a 6:4 ratio. The training set was input into the four classification models for training. After the training was completed, we employed six evaluation indices to analyze the quality of the models in the experiment.

The equipment used for machine learning model training was an Acer Veriton M4640G host with an Intel Core i5-6500 CPU at 3.20 GHz and Micron Crucial 8 GB DDR4-2400 RAM. The programming language used was Python 3.7.4. The Scikit-learn machine learning library was used for machine learning model training and testing. This library was developed using Python and includes many classification and regression models.

3.2.1 Evaluation of overall classification effectiveness

Table 5 presents the evaluation results regarding the overall performance of the four machine learning classification models. The data were partitioned randomly at a 6:4 ratio. Partitioning was conducted six times, and the results of the six partitioning procedures were averaged. Among the models, the SVM model had the best accuracy, sensitivity, AUC, and Cohen's kappa, whereas the random forest model had the best performance in terms of specificity and accuracy.

3.2.2 Classification evaluation at different time points

The classification performance of each of the models was investigated at the three postsurgical time points, as shown in Table 6. We collected eight pieces of data for both the sham and TBI groups (comprising a data set with 16 data values) for each time point. The data set was partitioned into a training set and a test set at a 6:4 ratio and input into the four machine learning models. After six independent training iterations, the mean values were recorded.

Table 5
Evaluation results regarding the performance of the four machine learning classification models.

Model	Evaluation indices	Mean ± standard deviation
SVM	Accuracy	**0.850 ± 0.03**
	Sensitivity	**0.983 ± 0.04**
	Specificity	0.762 ± 0.07
	Precision	0.737 ± 0.04
	AUC	**0.940 ± 0.03**
	Cohen's kappa	**0.702 ± 0.05**
Decision tree	Accuracy	0.750 ± 0.03
	Sensitivity	0.812 ± 0.09
	Specificity	0.705 ± 0.07
	Precision	0.650 ± 0.04
	AUC	0.758 ± 0.03
	Cohen's kappa	0.495 ± 0.06
Random forest	Accuracy	0.842 ± 0.04
	Sensitivity	0.875 ± 0.06
	Specificity	**0.822 ± 0.08**
	Precision	**0.768 ± 0.10**
	AUC	0.913 ± 0.03
	Cohen's kappa	0.675 ± 0.10
KNN	Accuracy	0.758 ± 0.05
	Sensitivity	0.897 ± 0.05
	Specificity	0.670 ± 0.09
	Precision	0.645 ± 0.10
	AUC	0.888 ± 0.03
	Cohen's kappa	0.527 ± 0.10

The sensitivity and AUC values for the SVM model at the three postsurgical time points were substantially superior to those obtained for the other models; some values approached 100%. The SVM model also exhibited satisfactory performance in accuracy, specificity, precision, and Cohen's kappa. Thus, the SVM model had excellent performance in classifying TBI at different severity levels. On Day 28, the AUC for the random forest model reached 100%, and its sensitivity was >90%. For the data collected on Days 14 and 28, the sensitivity of the KNN model also reached 100%, and its AUC values at all three time points were between 90 and 95%.

4. Discussion

We quantified 16 cognitive parameters on the basis of rats' trajectory points in the eight-arm maze and selected five representative TBI feature parameters for machine learning model training. Among the five parameters, three were unrelated to direction and two were related to direction. The analysis results for the three direction-unrelated parameters imply that the rats with TBI were likely to have forgotten their previous movement trajectories and destinations, increasing their searching time and total walking distance. The analysis results for the direction-related parameters reveal that the frequency of rats with TBI entering the arm on the right after exiting an arm increased with time postsurgery; this frequency differed significantly from that

Table 6
Means of the five evaluation indices for the models at the three time points.

Model	Evaluation indices	On the 7th day after surgery (including standard deviation)	On the 14th day after surgery (including standard deviation)	On the 28th day after surgery (including standard deviation)
SVM	Accuracy	**0.81 ± 0.07**	**0.86 ± 0.08**	**0.91 ± 0.16**
	Sensitivity	**0.96 ± 0.09**	**1.00 ± 0.00**	**1.00 ± 0.00**
	Specificity	**0.71 ± 0.11**	**0.76 ± 0.12**	**0.89 ± 0.23**
	Precision	0.71 ± 0.15	**0.74 ± 0.15**	**0.85 ± 0.23**
	AUC	**0.99 ± 0.03**	**0.99 ± 0.03**	**1.00 ± 0.00**
	Cohen's kappa	**0.75 ± 0.14**	**0.81 ± 0.16**	**0.85 ± 0.28**
Decision tree	Accuracy	0.71 ± 0.18	0.71 ± 0.12	0.71 ± 0.08
	Sensitivity	0.62 ± 0.30	0.75 ± 0.38	0.74 ± 0.21
	Specificity	0.66 ± 0.30	0.71± 0.19	0.87 ± 0.07
	Precision	**0.54 ± 0.25**	0.64 ± 0.27	0.84 ± 0.10
	AUC	0.71 ± 0.14	0.72 ± 0.13	0.75 ± 0.10
	Cohen's kappa	0.39 ± 0.30	0.41 ± 0.27	0.48 ± 0.17
Random forest	Accuracy	0.76 ± 0.23	0.78 ± 0.11	0.81 ± 0.16
	Sensitivity	0.88 ± 0.28	0.96 ± 0.09	0.92 ± 0.19
	Specificity	0.64 ± 0.29	0.75 ± 0.12	0.83 ± 0.26
	Precision	0.66 ± 0.26	0.75 ± 0.15	0.83 ± 0.26
	AUC	**0.91 ± 0.03**	0.99 ± 0.07	**1.00 ± 0.00**
	Cohen's kappa	0.55 ± 0.37	0.61 ± 0.17	0.67 ± 0.28
KNN	Accuracy	0.79 ± 0.10	0.81 ± 0.11	0.83 ± 0.21
	Sensitivity	0.88 ± 0.19	**1.00 ± 0.00**	**1.00 ± 0.00**
	Specificity	0.64 ± 0.15	0.72 ± 0.12	0.86 ± 0.31
	Precision	0.67 ± 0.19	0.74 ± 0.16	0.82 ± 0.27
	AUC	**0.90 ± 0.06**	0.93 ± 0.04	0.95 ± 0.08
	Cohen's kappa	0.60 ± 0.18	0.67± 0.17	0.68 ± 0.34

observed in the sham group. This might reveal a behavioral characteristic in the movements of patients with TBI.

We applied four machine learning models to evaluate and compare the five cognitive parameters of the rats with TBI. The SVM model had a fairly favorable performance in the entire experiment and at individual time points. Accordingly, the SVM model was determined to be the most suitable choice for our experiment. The sensitivity and AUC values derived for the random forest and KNN models on Day 28 after surgery were both favorable, indicating that the two models exhibited satisfactory performance in classifying severe TBI. The decision tree model had the poorest performance among the four models. A possible reason is that it only had one decision-maker. Its classification results were unstable. A completely different tree could be created for each new variable. Regarding the six model evaluation indices, the specificity and precision of the four models were relatively low. Therefore, the models had poor performance in terms of making positive predictions and identifying negative features. Therefore, when the data features were insufficiently clear, the models were highly likely to classify the data as positive.

Symptoms of mild TBI are not obvious. However, the performance of the TBI group with respect to the various cognitive parameters differed from that of the sham group over time. This result is consistent with those of most studies in the relevant literature. For the four models, the

classification results obtained on Day 28 typically were superior to those obtained at the previous two time points. This result verifies that the severity of dementia increases over time. Accordingly, the early detection of the signs of dementia can enable the early treatment of the disease and thus improve the treatment effectiveness. However, the obtained results reveal that only the SVM model had satisfactory classification performance in the early stage of the disease.

Feature parameters for machine learning training are key factors affecting the final accuracy of models. Therefore, parameter selection markedly affects classification success. Parameter selection was particularly crucial in this study because the sample size was small. Conventionally, only three parameters are used for an eight-arm maze. To increase accuracy and compensate for the data insufficiency in this study, we included multiple additional parameters. Data variety can affect classification results, and different models have individual strengths. Hence, we increased data variety, applied different machine learning models for analysis and comparison, and selected the most suitable classification model. The SVM model attained a satisfactory level of reliability and objectivity.

5. Conclusions

We identified five representative cognitive parameters for TBI by using an animal maze and revealed differences between sham rats and rats with TBI. Among the four models used in this study, the SVM model had the highest performance. Its sensitivity and AUC reached 98 and >94%, respectively. These values even approached 100% at some time points. These results indicate that this model has excellent performance for TBI classification. Our findings can serve as a reference for identifying patients with TBI in the future.

Acknowledgments

We thank Prof. Ching-Ping Chang of the Department of Medical Research, Chi Mei Medical Center, Tainan City 701, Taiwan, for her valuable assistance in the experiments with the rats. We also acknowledge the support provided to this study by the Ministry of Science and Technology, Taiwan, under grant number MOST 109-2221-E-167-002-MY3.

References

1 K. Hackenberg and A. Unterberg: Nervenarzt **87** (2016) 203. https://doi.org/10.1007/s00115-015-0051-3
2 W. Heegaard and M. Biros: Emerg. Med. Clin. North Am. **25** (2007) 655. https://doi.org/10.1016/j.emc.2007.07.001
3 A. J. Hicks, A. C. James, G. Spitz, and J. L. Ponsford: J. Neurotrauma **36** (2019) 3191. https://doi.org/10.1089/neu.2018.6346
4 R. C. Mannix and M. J. Whalen: Int. J. Alzheimer's Dis. **2012** (2012) 608732. https://doi.org/10.1155/2012/608732
5 E. M. Sikoglu, M. E. Heffernan, K. Tam, K. M. Sicard, B. T. Bratane, M. Quan, M. Fisher, and J. A. King: Behav. Brain Res. **259** (2014) 354. https://doi.org/10.1016/j.bbr.2013.11.008
6 S. J. Song, X. Y. Kong, S. Acosta, V. Sava, C. V. Borlongan, and J. Sanchez-Ramos: Int. J. Mol. Sci. **18** (2017) 1418. https://doi.org/10.3390/ijms18071418
7 M. Yamada and Y. Sakurai: Cogn. Neurodyn. **12** (2018) 519. https://doi.org/10.1007/s11571-018-9493-1
8 Y. Fu, Y. M. Chen, L. E. Li, Y. M. Wang, X. Y. Kong, and J. H. Wang: Int. J. Dev. Neurosci. **60** (2017) 8. https://doi.org/10.1016/j.ijdevneu.2017.03.010

9 L. Y. Xiang, A. Harel, H. Y. Gao, A. E. Pickering, S. J. Sara, and S. I. Wiener: Sci. Rep. **9** (2019) 1361. https://doi.org/10.1038/s41598-018-37227-w
10 R. Morris: J. Neurosci. Methods **11** (1984) 47. https://doi.org/10.1016/0165-0270(84)90007-4
11 D. S. Olton: Physiol. Behav. **40** (1987) 793. https://doi.org/10.1016/0031-9384(87)90286-1
12 D. Puzzo, L. Lee, A. Palmeri, G. Calabrese, and O. Arancio: Biochem. Pharmacol. **88** (2014) 450. https://doi.org/10.1016/j.bcp.2014.01.011
13 H. Hodges: Cogn. Brain Res. **3** (1996) 167. https://doi.org/10.1016/0926-6410(96)00004-3
14 B. Noori: Appl. Artif. Intell. **35** (2021) 567. https://doi.org/10.1080/08839514.2021.1922843
15 P. C. Ravoor and T. S. B. Sudarshan: Comput. Sci. Rev. **38** (2020) 100289. https://doi.org/10.1016/j.cosrev.2020.100289
16 S. Crisler, M. J. Morrissey, A. M. Anch, and D. W. Barnett: J. Neurosci. Methods **168** (2008) 524. https://doi.org/10.1016/j.jneumeth.2007.10.027
17 A. Lucas, A. T. Williams, and P. Cabrales: IEEE J. Transl. Eng. Health Med. **7** (2019) 1900509. https://doi.org/10.1109/jtehm.2019.2924011
18 M. R. Lysenko-Martin, C. P. Hutton, T. Sparks, T. Snowden, and B. R. Christie: J. Neurotrauma **37** (2020) 1777. https://doi.org/10.1089/neu.2019.6842
19 M. E. Haveman, M. Van Putten, H. W. Hom, C. J. Eertman-Meyer, A. Beishuizen, and M. C. Tjepkema-Cloostermans: Crit. Care **23** (2019) 401. https://doi.org/10.1186/s13054-019-2656-6
20 M. H. Othman, M. Bhattacharya, K. Moller, S. Kjeldsen, J. Grand, J. Kjaergaard, A. Dutta, and D. Kondziella: Neurocrit. Care **34** (2021) 31. https://doi.org/10.1007/s12028-020-00971-x
21 A. T. Jamal, A. Ben Ishak, and S. Abdel-Khalek: Neural Comput. Appl. **33** (2021) 7773. https://doi.org/10.1007/s00521-020-05518-x
22 H. C. Lee, H. K. Yoon, K. Nam, Y. J. Cho, T. K. Kim, W. H. Kim, and J. H. Bahk: J. Clin. Med. **7** (2018) 322. https://doi.org/10.3390/jcm7100322