農学生命科学研究科獣医学専攻
平成２９年度博士課程入学
氏名　張　育薇
指導教員名　松脇　貴志

論文題目　Establishment of a diagnostic method for premenstrual dysphoric disorder using machine learning algorithms and hippocampal factors in rats
（機械学習を用いたラット海馬での発現遺伝子を指標とする月経前不快気分障害の診断法の確立）

Premenstrual dysphoric disorder (PMDD) is a disease involving emotional and physical symptoms during the luteal phase in women of childbearing age. The main symptoms include anxiety, depression, irritability, and mood lability. The prevalence of PMDD is about 1.8%-5.8%, and up to 80% of women are affected by one or more of those syndromes. The pathological mechanism has been considered that the chronic exposure and rapid withdrawal of the metabolites of progesterone, allopregnanolone, in the luteal phase makes gamma-aminobutyric acid-A receptors (GABAR) insensitive, and thus induces anxiety- and depression-like behaviors. However, due to the inconsistent experimental methods and results in the previous studies, the important pathological factors of PMDD remain unknown, and the clinical diagnostic procedure is still subjective and ineffective. To break through the obstacle, the key is to find the important factors of PMDD symptoms and use them to build an objective diagnostic method. Moreover, because there are multiple symptoms and involved factors in PMDD pathology, new tools which can deal with more than one feature at the same time were used to achieve the goal.

The chosen tool is machine learning in artificial intelligence. Machine learning is the study of computer algorithms used for data analysis. It has the ability to improve automatically through learning by experience. There are two major approaches in machine learning, which are supervised learning and unsupervised learning. Supervised learning is a method using input vectors of data to find the target vector. Unsupervised learning is a method to use when the data only have input vectors but no target vector. Besides, there is still another type of machine learning approach which is popular in recent years. Deep learning is an analytic method with a construction similar to the human brain. Due to its multilayer structure and the ability of autotuning, it is a method with high accuracy and is suitable for solving complex problems, such as image recognition.

In the present study, pseudopregnant rats were used to mimic the hormone environment during the luteal phase in humans, and clustered by the severity of PMDD-like symptoms using unsupervised machine learning implementing the k-means algorithm. Next, the factors most related to symptom severity were extracted through an analysis of the RNA sequencing (RNA-seq) results using an original two-step feature selection. The selected factors were considered as a diagnostic set and used to build predictive machine learning models. Another group of rats that underwent the same experimental procedure was then used to evaluate the feasibility of the diagnostic method.

Chapter 1 includes the introduction of the five methods and tools used for data analysis. First, DeepLabCut (DLC) was used to analyze the videos of behavioral tests. DLC is a deep learning package. It can capture the coordinates of the target rat in every frame of the video after training. The biggest advantage of DLC is that it recognizes animals in video directly but not by the difference in gray scale of the animal and the background. This property substantially reduces the wrong tracking results generated by mistaking the animal and the background or reflected light. Second, an unsupervised learning algorithm, k-means, was used to cluster animals by their behavioral test results. K-means is a method that can put the data points with similar characteristics together to form clusters. The advantage of k-means is that it is sensitive to noise and thus can help to get rid of outliers. In addition, the number of centroids in k-means can be customized which makes it easy to adjust the results after observing the distribution of data. Third, the RNA-seq results were processed by Usegalaxy, an online biological platform for analysis of genomics. Fourth, an original two-step feature selection method was used to identified the important factors of PMDD symptoms from the RNA-seq results. The traditional ways of differential expression analysis, such as DESeq and edgeR, can only tell the group of genes with significant difference but not their ranking of importance. Thus, the feature selection method was used. It was built by a ranking algorithm based on univariate statistical tests, SelectKBest, and a supervised learning algorithm, Linear Support Vector Machine (Linear SVM) in a cross-validation pipeline, Grid Search CV. SelectKBest can obtain all the combinations of all the features. Linear SVM can separate data by a hyperplane with the maximal margin. Grid Search CV can give out model accuracies of all the assigned conditions. This feature selection method is able to find the combination of the features with the best mean model accuracy. The genes in this combination will be the set of the most important genes. However, the dataset in the present study was a small dataset with high dimensions which made the best mean accuracy of classification low and the selected gene set suspicious. Thus, another manual extraction step before the feature selection method was added. The preprocessing step helped to remove the disturbing genes and reduce the dimensions of the dataset, making the feature selection possess have higher accuracy and explanatory. Finally, six supervised machine learning models were built to classify the data of another group of rats undergoing the same experiments by their severity of PMDD symptoms using the selected gene set from the last step. The six classifiers included Linear SVM, Random Forest (RF), Bag SVM, Neural Network (NN), Light Gradient Boosting Machine (Light GBM), and Extreme Gradient Boosting (XGBoost). All of them are the most popular classifiers in machine learning with different principles. They were evaluated and compared to see which is the best classifier for the present task.

Chapter 2 is the experiment that contains the attempt to mimic the hormone environment of human luteal phase in experimental animals by an exogenous method, and the behavioral tests and their analyses by EthoVision (version: XT). First, a three-week progesterone daily injection plan was chosen to construct the luteal phase-like environment in the rats. Then, the rats underwent a series of behavioral tests including open field test (OFT), elevated plus-maze test (EPM), and forced swim test (FST) to evaluate the degree of anxiety- and depression-like behaviors. After the experiment, there were two concerns. First, the progesterone concentration was higher in the daily injection group (168 ± 32 ng/ml) than that in a natural pregnant period. This suggested that the daily injection plan might be a method inducing excessive progesterone. Second, there were no statistical significances in all the behavioral test results after the analysis by EthoVision. Thus, the methods of progesterone evocation and behavioral test analysis were alternated in the main experiment.

Chapter 3 is the experiment that contains the use of pseudopregnant rats in the process of the same behavioral tests as above followed by an analysis of DLC and custom programming scripts, hippocampal sampling, analysis of RNA-seq results, and evaluation and prediction of the supervised learning models. After the raw behavioral data of pseudopregnant rats were analyzed by DLC, the accurate behavioral test results were obtained. The main results of the tests: time ratio in center zone in OFT, time ratio in closed arms in EPM, and immobile time ratio in FST served as input vectors, and were used to cluster the animals by k-means. The clustering result showed that there were three clusters with mild to severe anxiety- and depression-like behaviors, being also referred to as cluster

1, 2, and 3. Then, the analysis of the RNA-seq results of hippocampal factors of cluster 1, 2, and 3 were performed by the original two-step feature selection method mentioned before. In the classification of cluster 1 versus 2 versus 3, the best accuracy was only 69.40% and the number of features remaining was only 1. Due to the difficulty of separating cluster 1 and 2, the two-class classification of cluster 1+2 versus 3 was also performed. This time, the number of features remaining dropped gradually and the accuracy kept at 100.00% when the number of features remaining was 16. The only genes remaining in the three-class classification and the 16 genes remaining in the two-class classification were then selected to form the set of important genes. To confirm the RNA-seq results, quantitative polymerase chain reaction (qPCR) was performed to verify the gene expression of the 17 identified genes, and the result of qPCR was used to build the supervised learning models. The six models based on Linear SVM, RF, Bag SVM, NN, Light GBM, and XGBoost were used for both three- and two-class classifications. Both the model accuracies and the predictive accuracies, the measurement of a model correctly classifying the training data and the test data respectively, were calculated to evaluate the performance of the models. The training data and the test data in the present study is the qPCR results of the set of important genes from the pseudopregnancy group and another group of rats undergoing the same experimental procedure. In the results, NN had the best model accuracies in both classifications (80.00% and 100.00%); while Linear SVM and XGBoost had considerably lower accuracies (both 54.55%) in the three-class and XGBoost had the lowest accuracy (72.73%) in the two-class classification. As for the predictive accuracies, RF had the best accuracies (96.00% and 100.00%); while NN in the three-class (78.00%) and Light GBM in the two-class (91.00%) had the lowest. By considering all the performances, RF was the best classifier in our case. The opposite, NN, the one with higher model accuracies but lower predictive accuracies, might have a problem with overfitting.

In the present study, a diagnostic method for PMDD syndromes of anxiety and depression using machine learning by hippocampal factors in rats was successfully established. In the clustering result, the three clusters had different symptom severity corresponding to healthy, premenstrual syndrome (PMS), and PMDD. PMS is the term used in the patients whose symptoms do not meet the criteria of PMDD. The differentiation of PMS and PMDD is also difficult in the clinical setting. The present result showed that it is possible to separate PMS and PMDD simultaneously when diagnosing PMDD. Besides, an original two-step feature selection method for RNA-seq small dataset analysis were also employed. By using it, the feature dimensions were reduced of 84.49% and a set of diagnostic genes was provided instead of only one key gene. Regarding the 17 identified genes, although they are not closely related to each other, almost all of them have been reported as involved in the development of PMDD syndromes. Besides, some of them have already been known for their relationship with progesterone or GABAR. For example, G protein-coupled receptor 101 (Gpr101) was found expressed by a subgroup of GABAR in the limbic system. The results suggested that our method can identify factors that have actual relationships with and close to PMDD symptoms and the hypothesized pathological pathway. Furthermore, the methods and tools in the present study can be used not only for PMDD, but also other complex and unsolved diseases. They can also be used to perform differential diagnosis between two or more similar diseases and contribute to tailored treatment. However, there are also limitations in the present study. First, there were only a small number of individuals induced symptoms of severe level. There may be some level of difference among clusters not being reflected. Second, the hippocampal factors were used but they are difficult to be practiced directly in the clinical setting. Further studies using serum or cerebrospinal fluid are required. Finally, the experimental subjects are rats. The same experiments and methods are also needed to be tested in other experimental animals and humans for further understanding.

In conclusion, a methodology using machine learning to construct efficient and objective diagnostic procedures for pluricausal diseases, including PMDD, was established. This is the first study applying machine learning to syndrome classification of experimental animals. The procedures and tools developed in the present study can be expected to be applied to other mood disorders and complex diseases in the future.