

RESEARCH ARTICLE

In silico identification of critical proteins associated with learning process and immune system for Down syndrome

Handan Kulan *, Tamer Dag

Computer Engineering Department, Kadir Has University, Istanbul, Turkey

* 20161102001@stu.khas.edu.tr



Abstract

Understanding expression levels of proteins and their interactions is a key factor to diagnose and explain the Down syndrome which can be considered as the most prevalent reason of intellectual disability in human beings. In the previous studies, the expression levels of 77 proteins obtained from normal genotype control mice and from trisomic Ts65Dn mice have been analyzed after training in contextual fear conditioning with and without injection of the memantine drug using statistical methods and machine learning techniques. Recent studies have also pointed out that there may be a linkage between the Down syndrome and the immune system. Thus, the research presented in this paper aim at in silico identification of proteins which are significant to the learning process and the immune system and to derive the most accurate model for classification of mice. In this paper, the features are selected by implementing forward feature selection method after preprocessing step of the dataset. Later, deep neural network, gradient boosting tree, support vector machine and random forest classification methods are implemented to identify the accuracy. It is observed that the selected feature subsets not only yield higher accuracy classification results but also are composed of protein responses which are important for the learning and memory process and the immune system.

OPEN ACCESS

Citation: Kulan H, Dag T (2019) In silico identification of critical proteins associated with learning process and immune system for Down syndrome. PLoS ONE 14(1): e0210954. <https://doi.org/10.1371/journal.pone.0210954>

Editor: Wajid Mumtaz, Zapadoceska univerzita, CZECH REPUBLIC

Received: April 11, 2018

Accepted: January 6, 2019

Published: January 28, 2019

Copyright: © 2019 Kulan, Dag. 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 author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: HK holds a PhD Scholarship from Tubitak 2211 A National Scholarship Programme for PhD Students. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Down syndrome (DS) is a very common identifiable genetic cause of intellectual disability (ID) and affects approximately one in 700 live births [1]. In addition to ID, people with DS are at risk for certain types of blood diseases, like leukemia, autoimmune disorders and Alzheimer's disease (AD) [2, 3].

The characteristics of DS can be diagnosed by the observation of the extra copy of whole or a portion of the long arm of human chromosome 21 (Hsa21). Hsa21 is responsible for nearly 160 protein-coding genes and five microRNAs [4]. Over expression of these proteins which include transcription factors, cell surface receptors, protein modifiers, adhesion molecules, RNA splicing factors and components of many biochemical pathways can cause the learning and memory (L/M) deficits. In addition for a person diagnosed with DS, the number of

neurons and cellular morphology are not normal in brain regions, such as the cortex, cerebellum and hippocampus [5–7].

Scientists have been using mice to find a treatment for the DS. However, it is compelling to model DS in mice since orthologs of the Hsa21 genes map to many mouse chromosomes, chromosomes 10, 16 and 17. However, Ts65Dn trisomic mice consisting 88 orthologs of Hsa21 protein coding genes and 5 microRNA genes can be used as a DS mouse model [8, 9]. For the treatment of the DS, many efforts are in progress in order to develop drugs. More than 20 drugs which have diverse properties, such as N-methyl-D-aspartate receptor (NMDAR) antagonist, γ -aminobutyric acid A (GABAA) receptor antagonists, acetylcholinesterase inhibitors and the green tea component have been shown to be effective for rescuing performance in L/M tasks [10–18].

One of these drugs called memantine, is an NMDAR antagonist and it modulates excitatory neurotransmission through antagonizing the activity by binding the N-methyl D-aspartate 2A (NR2A) and N-methyl D-aspartate 2B (NR2B) subunits with high on and off rates [13, 19, 20]. The NMDARs gated by glutamate plays an essential role in excitatory transmission and L/M process. When excessive amounts of glutamate binds to NMDARs, they generate free radicals and cause the synaptic dysfunction [21]. However, when memantine binds to NMDARs, it prevents glutamate binding and thus prevents cognitive and memory deficits [22]. By inspecting the protein profiles of normal and trisomic mice with and without memantine treatment, the impact of memantine on learning capability can be evaluated. In order to understand which protein expressions in control mice are important for successful learning, which abnormalities for Ts65Dn trisomic mice cause failed learning, and which changes by memantine give rise to rescued learning for Ts65Dn trisomic mice, protein expression data has been evaluated by computational learning methods [23, 24].

In this study, we applied supervised learning methods to protein expression data for 77 proteins (thus a 77dimensional space) taken from the cortex of control and Ts65Dn trisomic mice, with and without memantine treatment and with and without contextual fear conditioning (CFC). We compared our results with previous studies where Self Organizing Map (SOM) was used to pinpoint functional or regulatory similarities among proteins with similar expression profiles.

In previous works, it was shown that discriminating proteins were enriched in processes, such as mTOR signaling pathway, AD, MAPK signaling pathway and apoptosis [25]. In addition, it was also stated that DS could be related to the immune system and considered the DS as an immune disorder [26]. It was shown that interferon response which happened in response to the presence of several pathogens, such as parasites, viruses, bacteria and also tumor cells were consistently activated in cells obtained from individuals with the DS and could cause autoimmune disorders and leukemia, and perhaps AD. Because of this reason, we have inspected the feature subsets to understand their role in the immune system.

Per our findings, the selected protein subsets that we found can result in more accurate classification models of mice than those selected protein subsets chosen in previous studies. We have achieved better results by using different preprocessing steps and feature selection methodology when compared to previous studies. To select the best parameters for different classification methods to differentiate control and Ts65Dn trisomic mice, we applied the grid search method. To build a robust and reliable classification model, cross validation is used. Thus, our results are not only more accurate, but also composed of protein expressions that are important in the L/M process and the immune response. We made these conclusions after inspecting the literature to understand the importance of proteins in selected feature subsets. As a consequence, we believe that the protein subsets selected by applying the method described in this

paper can be utilized to understand the effects of proteins on L/M task and can be used to develop effective drugs.

Related work

Protein abnormalities in DS were studied by using different techniques to select proteins in the literature as stated in Table 1. Firstly, Ahmed et al [27, 28] studied a three level mixed effects (3LME) statistical analysis model of the Ts65Dn trisomic and normal mice protein profiles with and without exposure to CFC. Later, Higuera et al [25] examined the profiles using unsupervised learning, SOM to find important proteins for three cases; successful learning, rescued learning with memantine and failed learning. However, Eicher et al [29] believed that the problem was more suitable as a classification problem instead of a clustering problem and applied the linear SVM to find out proteins are discriminatory between two classes or groups of classes. A. Block et al [12] applied 3LME and used another drug RO4938581 for rescuing protein anomalies. B. Feng et al [30] used adaptive boosting (AdaBoost) method for feature selection and applied random forest, SVM and decision tree classification techniques for differentiating normal and trisomic mice.

Ahmed et al [27, 28] measured 84 protein expression levels in the hippocampus and the cortex of normal mice to evaluate learning capability. Context shock (CS) and shock context (SC) classes were partitioned into memantine or saline injected subclasses yielding four different classes and these four classes were analyzed. Memantine usage improved L/M capability in patients with AD [22]. Thus, the effects of memantine were assessed for comparison with DS. They showed that more than half of the protein levels changed significantly in the hippocampus. The number of proteins showing important changes in the cortex was smaller [27]. Furthermore, they applied this study to Ts65Dn trisomic mice to understand their protein dynamics for learning capabilities. They showed that there were indicative differences between the normal and Ts65Dn trisomic mice profiles [28].

Higuera et al [25] claimed that the statistical analysis performed by Ahmed et al [27, 28] was not satisfactory to determine all changes in protein profiles. They proposed that machine learning methods might fulfill these needs. They applied SOM to cluster protein profiles by using 77 proteins rather than 84. They described a set of class-specific clusters which were constituted from a set of adjacent nodes containing only samples from a single class or a node with at least 80% of its samples obtained from one mouse. Then, they applied the Wilcoxon rank-sum test and detected that protein levels were significantly different between each pair of clusters and specified those proteins as discriminatory between two classes.

A. Block et al [12] used $GABA_A\alpha5$ - selective modulator, RO4938581, for rescuing protein anomalies of trisomic Ts65Dn mice. In their work, 91 protein levels relevant to brain functions were measured by applying the 3LME. 44 of the 52 anomalies in trisomic Ts65Dn mice were corrected by RO4938581.

Eicher et al [29] believed that the problem was naturally related to classification problem rather than clustering problem since the determination of proteins that can separate two

Table 1. Studied techniques in the literature.

	Techniques
Ahmed et al [27]	3LME statistical model
Higuera et al [25]	Unsupervised Learning(SOM)+Wilcoxon rank-sum test
Eicher et al [29]	Supervised Learning(Linear SVM)+Wilcoxon rank-sum test
A. Block et al [12]	3LME statistical model
B. Feng et al [30]	Supervised Learning(AdaBoost)

<https://doi.org/10.1371/journal.pone.0210954.t001>

classes or groups of classes was required. In addition, they stated that classification methods could give higher accuracy than clustering methods as re-labeling clusters might lower the result accuracy and also accuracy could be measured more efficiently by using quantitative methods like cross validation, training and testing prediction rather than a visual basis. Therefore, they applied linear SVM for differentiating proteins. The classification performance of the linear SVM algorithm was better than the methods used in previous studies. However, for determining important proteins for more than two classes as an input to Higuera et al [25], Eicher et al [29] aggregated classes to constitute new positive and negative classes. These aggregated class results are not compared with the Higuera's work efficiently. Since linear SVM was not efficient for multi-class classification of proteins, multiclass classification methods were needed.

B. Feng et al [30] reduced feature subset from 77 to 30 features by applying AdaBoost method and applied Random Forest, Decision Tree and SVM classification methods to distinguish normal and trisomic Ts65Dn mice. They showed that selected protein datasets gave higher classification results. However, they did not consider control and Ts65Dn mice, with and without memantine treatment and with and without CFC stimulation subgroups. They were able to only differentiate control group from the trisomic group. Thus, their work did not show systematic analysis which was carried out with Higuera's work by inspecting the subgroups. Also, AdaBoost has been a very efficient method for solution of the two-class classification problem. However, in going from two-class to multiclass classification, naive AdaBoost algorithm has restricted to the reduction of the multiclass classification problem to multiple two-class problems. Thus, multiclass classification algorithms are needed to determine which proteins are discriminatory when there are more than two classes. For this reason, using naive Bayes learner which is one of the machine learning algorithms for multiclass classification [31, 32], we applied forward feature selection technique in our previous work [33] for the determination of important proteins for the DS. After selecting features, DNN, random forest and SVM classification methods are used to differentiate control and trisomic Ts65Dn mice. The accuracy result of our work turned out to be higher than B. Feng's work for all classification methods.

In this study, naive Bayes learner in forward feature selection method is used for learning process and features are selected based on their effects of the improvement in our model. After selecting features, DNN, gradient boosted tree, random forest and SVM classification methods are used to differentiate control and trisomic Ts65Dn mice. The control and Ts65Dn mice with and without memantine treatment and with and without CFC stimulation subgroups are analyzed and the accuracy results of different classification methods are compared with the accuracy results of feature subsets selected in Higuera's work in which systematic analysis was carried out by implementing SOM for three cases, successful learning, rescued learning and failed learning.

Materials and methods

Datasets

The dataset that we used in this paper are obtained from University of California Irvine Machine Learning Repository [34]. The same data was also used in Higuera's work [25] with which we will compare our results. The data contains of the expression levels of 77 proteins obtained from the nuclear fraction of cortex. In the dataset, there are 38 control mice and 34 trisomic Ts65Dn mice. 15 samples (three replicates of a five-point dilution series) are extracted from each mouse, resulting in 1080 samples.

The dataset is divided into eight classes of mice based on the protein profiles of 77 proteins after training in CFC with and without injection of memantine. These 77 proteins have roles for brain function, structure or development. Table 2 describes format of dataset in which

Table 2. Description of protein expression data.

Mice	P1	P2	P77	Class
mouse 1	0.504	0.747				1.676	c-cs-m
mouse 2	0.515	0.689				1.744	c-cs-m
mouse 3	0.509	0.730				1.926	c-cs-m
..							
mouse n							

<https://doi.org/10.1371/journal.pone.0210954.t002>

rows show the individual mice and columns show the expression levels of the 77 proteins and the class of each mice.

Table 3 shows eight classes of mice in the dataset based on their types, their exposure to CS or SC, applied drugs, number of mice in each class and their learning outcomes.

In CFC protocols [35], CS group are placed in a cage, waiting several minutes to explore the context. Later, an electric shock is applied. It is expected from wild type mice to link the context with an electric shock and would freeze after re-exposure to the same cage. The SC group is placed in a cage for controlling the effect of the shock alone. After placement in the cage, the electric shock is given immediately. It is expected that wild type mice do not learn to link the cage with shock and do not freeze after re-exposure the same cage. However, the trisomic Ts65Dn CS group of mice can not to learn and they do not freeze. However, if the Ts65Dn CS group of mice is injected with memantine, learning can be rescued [13].

After determining the groups, protein expression levels of each mice are measured with reverse phase protein arrays (RPPA) [36] which provides a quantitative analysis of the differential expression of proteins.

Data preprocessing

For some of the mice in the dataset, one or more protein level measurements have missing values. The missing values are replaced by the average expression levels of the corresponding sample of the mice in the same class. For example, if a mouse is missing the first sample expression level information, the missing value is replaced by the average value of the first sample protein expression of other mice in the same class.

The replacement method that we use is different from previous studies. In the previous studies, missing values were replaced with the average value of all protein expression levels in same class mice. 15 tissue samples that are three replicates of a five-point dilution series were obtained per mouse. We considered the effect of dilution ratio and applied different calculations to handle missing values. In addition to replacing the missing parts, all measurements are normalized with Z-score normalization to prevent proteins with higher values influence on the classification result erroneously. In order Z-score normalization to preserve range (maximum and minimum), we applied Z-score normalization rather than max-min normalization which was applied in Higuera’s work [25]. With Z-score normalization as shown in Eq (1), mean of the scores is subtracted from each score and then divided into the standard deviation [37].

$$Z = \frac{x - \mu}{\sigma}, \tag{1}$$

Feature selection

Before building a classification model, dimensionality reduction is very crucial for the understanding the information about the class. Dimensionality reduction is the process of

Table 3. Classes in the dataset.

Class	Type of Mice	Type of Experiment	Treatment	Number of Mice	Learning Outcome
c-SC-s	Control	Shock Context	Saline	9	No Learning
c-SC-m	Control	Shock Context	Memantine	10	No Learning
c-CS-s	Control	Context Shock	Saline	9	Normal Learning
c-CS-m	Control	Context Shock	Memantine	10	Normal Learning
t-SC-s	Trisomic	Shock Context	Saline	9	No Learning
t-SC-m	Trisomic	Shock Context	Memantine	9	No Learning
t-CS-s	Trisomic	Context Shock	Saline	7	Failed Learning
t-CS-m	Trisomic	Context Shock	Memantine	9	Rescued Learning

<https://doi.org/10.1371/journal.pone.0210954.t003>

decreasing the number of features for identification of the most relevant and important variables. It has the effect of decreasing the computational cost. For dimensionality reduction, feature selection and feature extraction methods can be used. Feature selection chooses a subset of features, while feature extraction generates a new feature set of original features.

The feature selection method, named as forward feature selection, is used in our work. It is the heuristic method which tries to find the optimal feature subset by iteratively selecting features based on the classifier performance. It begins with an empty feature subset and adds one feature at a time for each round. This one feature is taken from the pool of all features that are not in the feature subset and when added it results in best classifier performance. The above process is repeated until the required number of features are added. It does not examine all possible subsets and does not give a guarantee to find the optimal subset. However, it reduces the search time when compared to exhaustive feature selection [38].

In this study, forward feature selection is applied [39] using the Knostanz Information Miner (KNIME). The logic of program is a search loop. Inside the loop, the dataset is divided into a learning set (70%) and a validation set (30%). Learning set is used for the construction of the model in the current selection of the variables and validation set computes an unbiased error rate estimation. For the learning process, naive Bayes learner which was applied to multi classification problem is used. In spite of the underlying simplifying assumption of conditional independence, naive Bayes performs well with more than two classes problem. [33, 40]. In previous studies, the applied algorithms suffered from an efficient multiclass classification technique. In our studies, we eliminated this deficiency with naive Bayes algorithm in forward feature selection method.

Classification methods

After selecting features, classification methods are applied for differentiating mice in different subgroups. We carried out four classification methods, DNN, gradient boosted tree, random forest and SVM. These classification methods are implemented by using Python and Scikit Learn package [41]. In order to select the most appropriate parameters of classification methods, grid search method [42] is applied. Also, for building robust and reliable classification model, 5 fold cross validation is applied. Thanks to cross validation, a learner can generalize to an unknown data set. In K Fold cross validation [43], the data is partitioned into k subsets. Only one of these subsets is used as the test set and the others are constituted to a training set at each time. This procedure is repeated k times. The error estimation is averaged over all k trials to get total effectiveness. This way significantly decreases bias since we are using most of the data for fitting. It also significantly reduces variance as most of the data is also being used

in validation set. In the rest of this section, a brief discussion on the four types of classification methods that we have used in our study is described.

Deep Neural Network

DNN is type of a neural network with multiple layers between the input and output layers. Neural networks are inspired from human brain as they acquire knowledge through learning and composed of connected units called artificial neurons which is analogous to biological neurons in a brain. Each connection between neurons can transmit a signal to another neuron and may also have a weight that can increase or decrease the strength of the signal. Neurons are generally organized in layers and signals travel between layers. In order turn the input into the output, DNN tries to find the relationship whether linear or not. The network moves through the layers calculating the probability of each output [44].

Gradient boosted tree

Boosting is a sequential ensemble method that converts weak learners to a strong learner by promoting previously mislabeled data with higher weight. Thus, the subsamples of data have an different probability of appearing in subsequent models and the ones with the highest probability of error appear most [45]. Gradient boosting builds the model in a sequential way. At each step the decision tree $h_m(x)$ that is base learner is selected to minimize a loss function L given the current model $F_{m-1}(x)$ as shown in Eq (2).

$$F_0(x) = \operatorname{argmin} \sum_{i=1}^n (L(y_i), \gamma), \quad (2)$$

$$F_m(x) = F_{m-1}(x) + \gamma h_m(x)$$

In above equation, m is the number of iterations, $F_m(x)$ is the model and γ is the learning rate.

Support vector machines

SVM is a supervised machine learning classification method which uses a data set d -dimensional Euclidean space. The number of d represents the number of features in the data set. Later, SVM finds an optimal $(d-1)$ dimensional hyperplane as given in Eq (3) to separate the data by class. In this equation, w represents a weight vector of length d and b represents a bias term. The distance between the hyperplane and the nearest data point from either part of the hyperplane is known as the margin. In order to classify new data correctly, the distance between between the hyperplane and any point within the training set must be higher [46].

$$w \cdot x + b = 0 \quad (3)$$

Random forest

Random forest is composed of many decision trees which are selected from a random subset of training set. It constructs random forest by combining a large number of decision trees and outputs the class that is the mode of the classes or mean prediction of the individual trees [47]. Random forest classification methodology is described in Fig 1.

Model is tuned with two parameters n_{tree} and n_{try} to get optimized forest architecture. The parameter n_{tree} specifies how many trees are to be built to populate the random forest

Step 1: Draw n_{tree} bootstrap samples from the original data

Step 2: For each of the samples, produce an unpruned classification tree, with the following modification:

At each node, rather than choosing the best split among all predictors, randomly sample n_{try} of the predictors and choose the best split from among those variables

Step 3: Predict new data by aggregating the predictions of the n_{tree} trees

Fig 1. Random forest classification algorithm.

<https://doi.org/10.1371/journal.pone.0210954.g001>

where n_{try} specifies the number of variables that will be considered at any time in deciding how to partition the dataset.

Results

Using the KNIME tool [39], forward feature selection technique is used to obtain the feature subsets for identifying the critical proteins in successful learning, rescued learning and failed learning cases. Afterwards, in order to validate importance of selected proteins, principal component analysis (PCA) is carried out. After determination and validation crucial proteins, DNN, gradient boosted tree, random forest and SVM classification methods are executed. PCA and application of classification methods are carried out with Python and Scikit learn package [41]. Also, grid search which is the parameter optimization technique [42] and 5 fold cross validation are done for obtaining robust and reliable classification results. The below subsections successively show the results of feature selection method, PCA and classification methods for successful learning, rescued learning and failed learning.

Feature selection results

Forward feature selection method is applied with KNIME tool [39] and then results of selected feature subsets are compared with Higuera's work [25]. In that work, three feature subsets were highlighted for normal learning, rescued learning and failed learning. Higuera et al [25] analyzed control mice and trisomic mice separately and together in order to understand the changes in protein levels. To understand which of the protein expression level changes are required for successful learning, all groups of normal mice were inspected in the first case. To determine important proteins in rescued learning, trisomic mice exposed to CFC with and without memantine were analyzed in the second case. The third case found out important protein abnormalities in failed learning by comparing normal and trisomic mice protein expression levels.

In our work, similar to Higuera’s work, we also selected three feature subsets to understand critical proteins in normal learning, rescued learning and failed learning. The number of features in feature subsets are selected based on the number stated in Higuera’s work [25] for comparison purposes. The difference in preprocessing and feature selection methods affect results in a positive manner and important proteins that were not highlighted in the previous work are found in normal learning, rescued learning and failed learning cases.

1. Feature subset of data from control mice. Table 4 shows the selected features and their accuracy with normal learning when selected feature is added to the subset. Under the first case, feature subset is selected from control group mice data. By comparing control group mice with and without memantine treatment and with and without CFC stimulation, critical proteins in successful learning can be understood. When compared with Higuera’s work [25], there are 4 common proteins (SOD1, pGSK3B, S6, CaNA) out of 11 proteins which are shown in bold. After literature review, it can be deduced that the selected proteins in successful learning are related to the L/M pathway and the immune responses [48–62].

2. Feature subset of data from trisomic Ts65Dn mice. In the second case, to understand the important proteins in rescued learning, features are selected from data consisting of trisomic mice which are exposed to CFC with and without memantine. When exposed to CFC, the trisomic mice fail to learn if they are not treated with memantine which rescues the learning performance. Table 5 shows the selected features and their accuracy results obtained when the corresponding feature is added to feature subset in rescued learning case. There are 2 common proteins (BRAF, CDK5) when compared with previous work. Literature search shows us that selected proteins in rescued learning are related to L/M process and immune response [63–69].

3. Feature subset of data from control and trisomic Ts65Dn mice. Under the third case, for identifying proteins that are critical in failed learning with trisomic mice, features are selected from protein expression levels of trisomic mice exposed to CFC without memantine and protein expression levels of control mice which are exposed to CFC with and without memantine. Table 6 shows the selected features and accuracy results of feature subset in failed learning. There are 2 common proteins (P38, GluR3) out of 10 proteins when compared with former work. Selected proteins in failed learning play important roles in signaling pathway [70, 71].

Validation of selected proteins subsets

In this section, we conducted PCA for both selected protein subsets and original protein sets for three cases; successful learning, rescued learning and failed learning. We projected these

Table 4. Feature subset of normal learning.

Feature No	Accuracy of Feature	Feature Subset	Feature Subset of Previous Work [25]
1	0.656	SOD1	DYRK1A
2	0.751	Ubiquitin	ITSN1
3	0.852	pGSK3B	pERK
4	0.873	S6	BRAF
5	0.905	CaNA	SOD1
6	0.921	IL1B	pNUMB
7	0.937	BAX	pGSK3B
8	0.942	pNR2A	CDK5
9	0.942	BDNF	S6
10	0.942	pJNK	GFAP
11	0.942	pCFOS	CaNA

<https://doi.org/10.1371/journal.pone.0210954.t004>

Table 5. Feature subset of rescued learning.

Feature No	Accuracy of Feature	Feature Subset	Feature Subset of Previous Work [25]
1	0.762	BRAF	DYRK1A
2	0.838	S6	pERK
3	0.85	CDK5	BRAF
4	0.887	BDNF	CDK5
5	0.887	pCREB	RRP1
6	0.9	PKCA	GFAP
7	0.912	SOD1	GluR3
8	0.925	PSD95	P3525
9	0.925	pNR2A	Ubiquitin

<https://doi.org/10.1371/journal.pone.0210954.t005>

feature sets into 3D spaces in order to validate the selected protein subsets specified in Feature Selection Results. As shown in Figs 2, 3 and 4, the PCA of selected protein subsets can better discriminate the class of mice instances when compared with the PCA of the original protein sets for three cases.

Fig 2 shows the PCA of successful learning. In this case, there are four classes which are normal genotype of control mice with and without memantine and with and without CFC stimulation. As seen in Fig 2, classes are better discriminated with selected proteins.

Fig 3 shows the result of PCA for rescued learning case. In this case, there are two classes which are trisomic mice exposed to CFC with and without memantine. It can be seen that better discrimination of classes can be done with selected proteins.

Fig 4 shows the result of PCA for failed learning. In this case, there are three classes which are trisomic mice exposed to CFC without memantine and normal mice with CFC simulation with and without memantine. Fig 4 also shows better discrimination of classes with selected proteins.

Classification results

After determination of the different feature subsets for the three cases, classification is performed for differentiating mice in different classes. DNN, gradient boosted tree, random forest and SVM classification methods are executed by using Python and Scikit learn package [41]. Parameters of classifiers are determined based on the grid search hyper-parameter optimization technique which is useful in computational biology problems. With the grid search

Table 6. Feature subset of failed learning.

Feature No	Accuracy of Feature	Feature Subset	Feature Subset of Previous Work [25]
1	0.636	P38	pNR1
2	0.713	pPKCAB	APP
3	0.775	CAMKII	MTOR
4	0.814	pCAMKII	P38
5	0.868	GluR3	NR2B
6	0.891	DSCR1	RAPTOR
7	0.907	nNOS	S6
8	0.915	BAX	Tau
9	0.93	pCFOS	GluR3
10	0.93	ERK	EGR1

<https://doi.org/10.1371/journal.pone.0210954.t006>

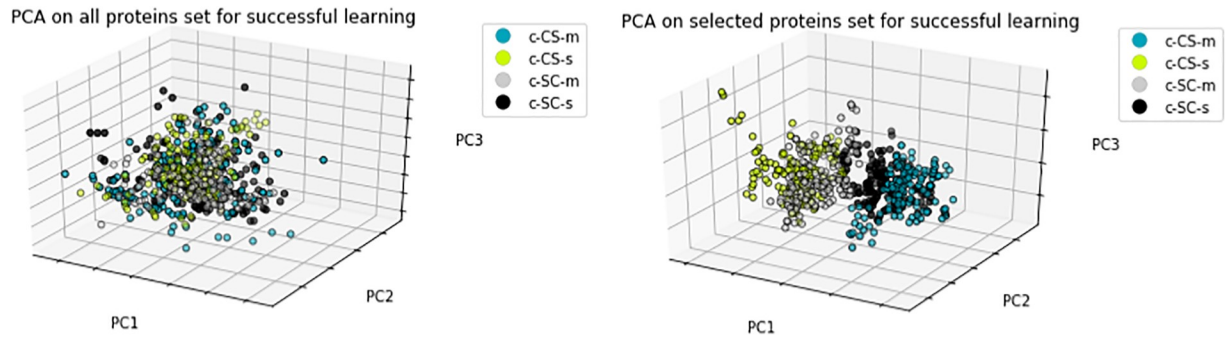


Fig 2. PCA of all proteins set and selected proteins subset for successful learning.

<https://doi.org/10.1371/journal.pone.0210954.g002>

method, the most suitable parameters for different classification methods are found. In addition, five fold cross validation is applied for preventing overfitting. Together with grid search method, cross validation affects classification accuracy in a positive manner.

The accuracy results of feature subsets shown in the previous Feature Selection Results part are compared with the Higuera's presented accuracy results.

Table 7 shows the classification accuracies of feature subsets selected in our work and Higuera's work [25] for successful learning. It can be seen that our feature subset gives higher accuracy results for all classification techniques. For example, while the accuracy of Random Forest was 0.902, it is increased to 0.963. Also, it is observed that the highest accuracy is obtained with SVM with a value of 0.981.

Table 8 shows the comparison of rescued learning classification results. The accuracy results of our feature subset are higher than previous work for all classification methods. For example, the accuracy of Random Forest is increased by % 6.3 from 0.883 to 0.946. For rescued learning, the highest accuracy is achieved with DNN and SVM with a value of 0.971.

Table 9 shows the comparison of classifications for failed learning. Similar to previous case, DNN and SVM give highest accuracy results with a value of 0.926. In addition, classification results of our feature subsets are again higher than previous work for all classification methods implemented.

Discussion

Pharmacotherapies of ID are largely unknown as the abnormalities at the complex molecular level which causes ID are difficult to understand. DS which is the prevalent reason of ID and

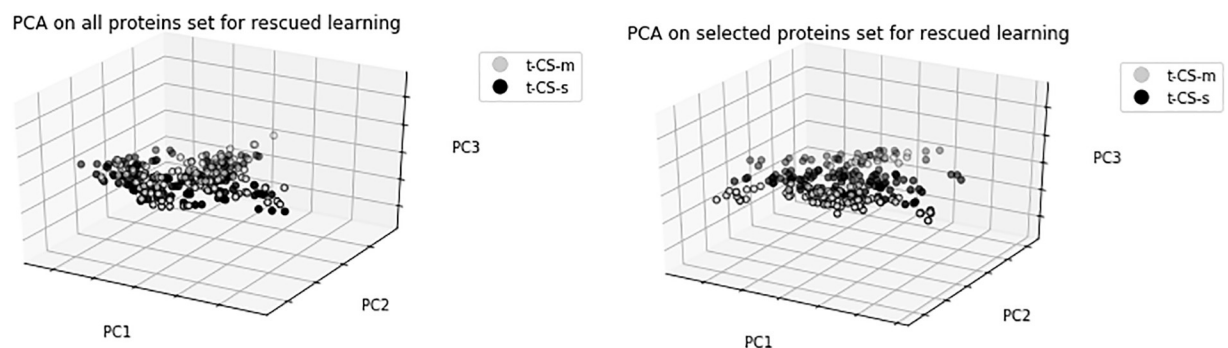


Fig 3. PCA of all proteins set and selected proteins subset for rescued learning.

<https://doi.org/10.1371/journal.pone.0210954.g003>

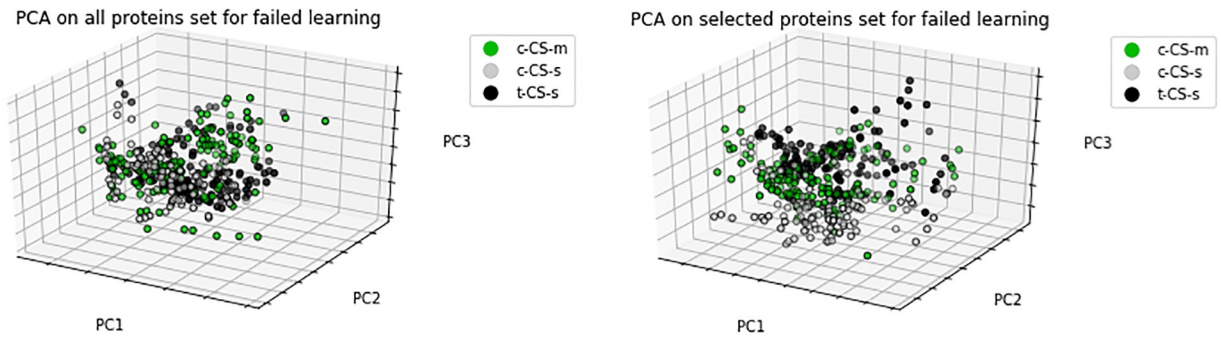


Fig 4. PCA of all proteins set and selected proteins subset for failed learning.

<https://doi.org/10.1371/journal.pone.0210954.g004>

caused by an extra copy of the Hsa21 has been investigated on protein levels. Due to the increase in trisomic genes, protein expression levels of corresponding genes are elevated. Furthermore, in addition to expression of genes on 21 chromosome, protein coding genes on other chromosomes play important roles in DS. Thus, understanding the abnormalities in the protein expressions are very important for developing drugs to rescue learning. For this reason, critical roles of proteins have been analyzed by comparing protein expression levels of normal mice and trisomic mice which are exposed to CFC with or without memantine treatment. In order to find critical proteins in DS, statistical analysis and machine learning methods are used.

In our work, we implemented forward feature selection technique for selecting protein subsets and applied DNN, gradient boosted tree, SVM and random forest classifiers to classify mice more accurately. The classification accuracy results of selected proteins are compared with Higuera et al work [25] in which SOM was applied for clustering of protein based on the similarities in their expression levels and Wilcoxon rank test was done for identifying significantly different protein levels between clusters.

Higuera et al [25] implemented SOM for three cases, successful learning, rescued learning and failed learning, respectively. In the first case, four classes of control mice protein profiles were analyzed for understanding critical proteins in successful learning. In the second case, trisomic mice which are exposed to CFC with or without applying drug memantine were investigated to understand rescue performance of memantine on trisomic mice learning capability. In the last case, using control and trisomic mice, protein profile patterns were analyzed for understanding important factors in learning impairment. They reduced feature subsets from 77 proteins to 11 proteins, 9 proteins and 10 proteins for the three cases, respectively. In this work, we applied naive Bayes classification technique in forward feature selection method rather than SOM which is the clustering technique to group protein levels. We constituted our feature subsets with the same number of proteins selected in Higuera’s work [25] in order to compare the results effectively.

Table 7. Accuracy result comparison of normal learning.

	Accuracy Result of Our Work	Previous Work Accuracy Result [25]
Deep Neural Network	0.972	0.967
Gradient Boosted Tree	0.935	0.902
Random Forest	0.963	0.902
SVM	0.981	0.961

<https://doi.org/10.1371/journal.pone.0210954.t007>

Table 8. Accuracy result comparison of rescued learning.

	Accuracy Result of Our Work	Previous Work Accuracy Result [25]
Deep Neural Network	0.971	0.954
Gradient Boosted Tree	0.933	0.892
Random Forest	0.946	0.883
SVM	0.971	0.921

<https://doi.org/10.1371/journal.pone.0210954.t008>

Before selecting feature subsets, the preprocessing step consisted of filling missing part and then normalization is carried out. Missing values are replaced by the average protein expression level of corresponding sample in the same class. This replacement is different from previous works where missing values were handled by the mean value of protein expression levels in same class mice. 15 tissue samples that are three replicates of a five-point dilution series were obtained per mouse. This dilution ratio affects expression level of proteins. Thus, we considered this effect and applied a different technique described in data processing part to handle missing values. For normalization, Z score normalization rather than max-min normalization which is used in Higuera’s work [25] is applied to prevent higher influence of proteins with higher values on the classification outcome.

After preprocessing steps, the forward feature selection algorithm is used to select feature subsets for three cases and these feature subsets are compared with Higuera et al work [25]. For learning process, naive Bayes learner which has been applied to multiclass classification problem is used. In spite of the underlying simplifying assumption of conditional independence, naive Bayes performs well with more than two classes problem. These distinct preprocessing and feature selection methods affected results in a good way and important proteins that were not highlighted in the previous works are found.

Critical proteins in DS have been found to be related with different pathways and processes, such as MAPK and MTOR pathways, immediate early genes (IEGs), AD, neurotrophin signaling pathway and apoptosis. In our work, in addition to these pathways and processes, we evaluated proteins according to their relations with immune system. It was hypothesized that DS causes an increase in interferon signaling which triggers the protective defenses of the immune system [26]. Thus, we searched proteins for finding a connection between L/M and immune response.

Eleven proteins were found to be significantly different in successful learning case: SOD1, ubiquitin, pGSK3B, S6, CaNA, IL1B, BAX, pNR2A, BDNF, pJNK and pCFOS. These proteins play important roles in L/M, immune response, MAPK pathway, mTOR pathway and AD. When compared with Higuera’s work [25], 4 proteins (SOD1, pGSK3B, S6 and CaNA) are found to be common. Three of the proteins (SOD1, pGSK3B and CaNA) are related to immune system.

SOD1 found on chromosome 21 causes immune abnormalities in Amyotrophic lateral sclerosis (ALS) disease [48] and increases reactive oxygen in DS [49]. Ribosomal Protein S6 and

Table 9. Accuracy result comparison of failed learning.

	Accuracy Result of Our Work	Previous Work Accuracy Result [25]
Deep Neural Network	0.926	0.921
Gradient Boosted Tree	0.879	0.844
Random Forest	0.892	0.859
SVM	0.926	0.910

<https://doi.org/10.1371/journal.pone.0210954.t009>

pGSK3B are components of mTOR pathway which takes action in learning [50]. Also, in the literature it is noted that GSK3 inhibitors provide to prevent excessive inflammation and ameliorate the autoimmune disease [51]. CaNA and IL1B are known to be pathogenesis of AD [52, 53]. In addition, it is known that IL1B is natural suppressor of innate inflammatory [54]. BAX and ubiquitin play critical roles in apoptosis and immune response [55, 56]. BDNF takes action in L/M [57] and bridges inflammation and neuroplasticity [58, 59]. pNR2A has well established roles in learning [60]. pJNK is component of MAPK pathway which is associated with L/M [61]. pCFOS is an IEG and is important in long term memory and neurological function seen in AD [62]. In the first case, by analyzing protein expression levels of control group mice, it can be deduced that proteins which are related to the L/M pathway and the immune responses are critical in successful learning.

Nine proteins are highlighted in the rescued learning case: BRAF, S6, CDK5, BDNF, pCREB, PKCA, SOD1, PSD95 and pNR2A. Two out of nine proteins are common with Higuera's work [25]. Four of these proteins (S6, BDNF, SOD1 and pNR2A) are also found in successful learning case and their importance is explained above. BRAF and PKCA are associated with MAPK pathway and important in learning [63, 64]. CDK5 is synaptic protein and plays a critical role in long-term memory [65]. Also, it regulates the evasion of tumors from the immune system [66]. PSD95 is a neuropathological marker of AD observed in later stage of DS [67]. In addition, PSD95 colocalizes with major histocompatibility complex class I (MHCI) which is the signature of its expressed proteins and is important for the immune system to differentiate self from nonself [68]. CREB regulates crucial cell stages and participates in neuronal plasticity [69]. Thus, it can be concluded from the second case that proteins which are important in rescued learning are relevant to the L/M and the immune response.

In the case of failed learning, ten proteins are found to be significant: p38, pPKCAB, CAMKII, pCAMKII, GluR3, DSCR1, nNOS, BAX, pCFOS and ERK. Two of these proteins (BAX and pCFOS) were also highlighted in successful learning and described above. The remaining selected proteins are largely connected to MAPK signaling pathway, such as P38, pPKCAB, CAMKII, pCAMKII and ERK. GluR3 is related to glutamate receptors which cause memory deficit if excess amount of glutamate binds to receptor [70]. DSCR1 is known to be over expressed in DS and affects signaling pathway [71]. Failed learning case also shows us the importance of signaling pathway in the learning process.

PCA is also done for both selected protein subsets and original protein sets for the three cases; successful learning, rescued learning and failed learning. It is shown that selected protein subsets can better discriminate the class of mice instances when compared with the PCA of the original protein sets for all the indicated cases.

After finding critical proteins for three cases, DNN, gradient boosted tree, random forest and SVM classification methods are applied. The parameters of classifiers are optimized with grid search. Also, 5 fold cross validation is done to prevent overfitting. The accuracy results of our feature subsets are found to be higher than previous work for all classifiers. DNN and SVM achieved the highest overall classification accuracy followed by random forest and then gradient boosted tree.

Multiple layers in a deep learning model can learn features from a wide perspective with higher flexibility. Thus, it is logical to obtain good results with DNN. SVM maps data to a feature space and then classify the data. It explicitly determines the decision boundary directly from the training data. Parameter optimization step is required to build an efficient SVM model. Using grid search method, parameters are selected and SVM with the selected parameters gives higher accuracy results. Accuracy result of random forest is lower than SVM and Deep Neural Network. The reason of lower accuracy can be the size of data as random forest generally needs larger number of instances for performing its randomization concept in a

good way. Also, decision trees used as base learners in the random forest cannot exactly learn many of soft linear boundaries at the decision surface which can cause lower success than the SVM non linear boundaries. Gradient boosted tree is prone to overfitting as it tries to find optimal linear combination of trees in relation to given train data. This tuning stage may be the reason of the lowest accuracy obtained by gradient boosted tree.

In conclusion, our work described in this paper provides better learning model and shows that proteins which are found to be related to the L/M and the immune system are critical in successful learning. Therefore, by extracting information from these protein subsets, the effective drugs can be developed for the treatment of DS.

Supporting information

S1 Dataset. Protein expression profiles of 77 proteins obtained from control and trisomic Ts65Dn mice. These data are a subset of those used in Ahmed's work [27](ZIP). (ZIP)

Author Contributions

Conceptualization: Handan Kulan, Tamer Dag.

Data curation: Handan Kulan, Tamer Dag.

Formal analysis: Handan Kulan, Tamer Dag.

Investigation: Handan Kulan, Tamer Dag.

Methodology: Handan Kulan, Tamer Dag.

Supervision: Tamer Dag.

Writing – original draft: Handan Kulan, Tamer Dag.

Writing – review & editing: Handan Kulan, Tamer Dag.

References

1. Parker SE, Mai CT, Canfield M, Rickard R, Wang Y, Meyer RE, et al. Updated National Birth Prevalence estimates for selected birth defects in the United States, 2004-2006. *Birth Defects Res A Clin Mol Teratol*. 2010; 88(12):1008–16. <https://doi.org/10.1002/bdra.20735> PMID: 20878909
2. Head E, Lott IT, Wilcock DM, Lemere CA. Aging in Down syndrome and the development of Alzheimer's disease neuropathology. *Curr Alzheimer Res*. 2016; 13(1):18–29. <https://doi.org/10.2174/1567205012666151020114607> PMID: 26651341
3. Lott IT. Neurological phenotypes for Down syndrome across the life span. *Progress in brain research*. 2012; 197:101–121. <https://doi.org/10.1016/B978-0-444-54299-1.00006-6> PMID: 22541290
4. Sturgeon X, Gardiner KJ. Transcript catalogs of human chromosome 21 and orthologous chimpanzee and mouse regions. *Mamm Genome*. 2011; 22:261–271. <https://doi.org/10.1007/s00335-011-9321-y> PMID: 21400203
5. Chapman RS, Hesketh LJ. Behavioral phenotype of individuals with Down syndrome. *Ment Retard Dev Disabil Res Rev*. 2000; 6:84–95. [https://doi.org/10.1002/1098-2779\(2000\)6:2<84::AID-MRDD2>3.0.CO;2-P](https://doi.org/10.1002/1098-2779(2000)6:2<84::AID-MRDD2>3.0.CO;2-P) PMID: 10899801
6. Silverman W. Down syndrome: cognitive phenotype. *Ment Retard Dev Disabil Res Rev*. 2007; 13:228–36. <https://doi.org/10.1002/mrdd.20156> PMID: 17910084
7. Nadel L. Down's syndrome: a genetic disorder in biobehavioral perspective. *Genes Brain Behav*. 2003; 2:156–66. <https://doi.org/10.1034/j.1601-183X.2003.00026.x> PMID: 12931789
8. Davison MT, Schmidt C, Reeves RH, Irving NG, Akeson EC, Harris BS, et al. Segmental trisomy as a mouse model for Down syndrome. *Prog Clin Biol Res*. 1993; 384:117–133. PMID: 8115398

9. Rueda N, Flórez J, Martínez-Cué C. Mouse models of Down syndrome as a tool to unravel the causes of mental disabilities. *Neural Plast.* 2012; 2012:584071. <https://doi.org/10.1155/2012/584071> PMID: 22685678
10. Gardiner KJ. Pharmacological approaches to improving cognitive function in Down syndrome: current status and considerations. *Drug Design, Development and Therapy.* 2015; 9:103–125. <https://doi.org/10.2147/DDDT.S51476> PMID: 25552901
11. Braudeau J, Delatour B, Duchon A, Pereira PL, Dauphinot L, de Chaumont F, et al. Specific targeting of the GABAA receptor 5 subtype by a selective inverse agonist restores cognitive deficits in Down syndrome mice. *Journal of Psychopharmacology (Oxford, England).* 2011; 25(8):1030–1042. <https://doi.org/10.1177/0269881111405366>
12. Block A, Ahmed MM, Rueda N, Hernandez MC, Martinez-Cué C, Gardiner KJ. The GABAA α 5-selective Modulator, RO4938581, Rescues Protein Anomalies in the Ts65Dn Mouse Model of Down Syndrome. *Neuroscience.* 2017; 372. <https://doi.org/10.1016/j.neuroscience.2017.12.038> PMID: 29292072
13. Costa AC, Scott-McKean JJ, Stasko MR. Acute injections of the NMDA receptor antagonist memantine rescue performance deficits of the Ts65Dn mouse model of Down syndrome on a fear conditioning test. *Neuropsychopharmacology.* 2008; 33:1624–1632. <https://doi.org/10.1038/sj.npp.1301535> PMID: 17700645
14. Chang Q, Gold PE. Age-related changes in memory and in acetylcholine functions in the hippocampus in the Ts65Dn mouse, a model of Down syndrome. *Neurobiol Learn Mem.* 2008; 89:167–177. <https://doi.org/10.1016/j.nlm.2007.05.007> PMID: 17644430
15. Corrales A, Martínez P, García S, Vidal V, García E, Flórez J, et al. Long-term oral administration of melatonin improves spatial learning and memory and protects against cholinergic degeneration in middle-aged Ts65Dn mice, a model of Down syndrome. *J Pineal Res.* 2013; 54:346–358. <https://doi.org/10.1111/jpi.12037> PMID: 23350971
16. Das I, Park JM, Shin JH, Jeon SK, Lorenzi H, Linden DJ, et al. Hedgehog agonist therapy corrects structural and cognitive deficits in a Down syndrome mouse model. *Sci Transl Med.* 2013; 5:201ra120. <https://doi.org/10.1126/scitranslmed.3005983> PMID: 24005160
17. Busciglio J, Capone G, O'Bryan J, Gardiner KJ. Down syndrome: genes, model systems, and progress towards pharmacotherapies and clinical trials for cognitive deficits. *Cytogenet Genome Res.* 2013; 141:260–271. <https://doi.org/10.1159/000354306> PMID: 24008277
18. Gardiner KJ. Molecular basis of pharmacotherapies for cognition in Down syndrome. *Trends in pharmacological sciences.* 2010; 31(2):66. <https://doi.org/10.1016/j.tips.2009.10.010> PMID: 19963286
19. Chen HS, Lipton SA. Pharmacological implications of two distinct mechanisms of interaction of memantine with N-methyl-D-aspartate-gated channels. *J Pharmacol Ex. Ther.* 2005; 314:961–97. <https://doi.org/10.1124/jpet.105.085142>
20. Lipton SA. Pathologically-activated therapeutics for neuroprotection: mechanism of NMDA receptor block by memantine and S-nitrosylation. *Curr Drug Targets.* 2007; 8:621–632. <https://doi.org/10.2174/138945007780618472> PMID: 17504105
21. Kamat PK, Rai S, Swarnkar S, Shukla R, Ali S, Najmi AK, et al. Okadaic acid-induced Tau phosphorylation in rat brain: role of NMDA receptor. *Neuroscience.* 2013; 238:97–113. <https://doi.org/10.1016/j.neuroscience.2013.01.075> PMID: 23415789
22. Olivares D, Deshpande VK, Shi Y, Lahiri DK, Greig NH, Rogers JT, et al. N-Methyl D-Aspartate (NMDA) Receptor Antagonists and Memantine Treatment for Alzheimer's Disease, Vascular Dementia and Parkinson's Disease. *Curr Alzheimer Res.* 2012 Jul; 9(6):746–758. <https://doi.org/10.2174/156720512801322564> PMID: 21875407
23. Larranaga P, Calvo B, Santana R, Bielza C, Galdiano J, Inza I, et al. Machine learning in bioinformatics. *Brief Bioinform.* 2006; 7(1):86–112. <https://doi.org/10.1093/bib/bbk007> PMID: 16761367
24. Nguyen CD, Costa AC, Cios KJ, Gardiner KJ. Machine learning methods predict locomotor response to MK-801 in mouse models of down syndrome. *J Neurogenet.* 2011; 25(1–2):40–51. <https://doi.org/10.3109/01677063.2011.558606> PMID: 21391779
25. Higuera C, Gardiner KJ, Cios KJ. Self-Organizing Feature Maps Identify Proteins Critical to Learning in a Mouse Model of Down Syndrome. *PLoS ONE.* 2015; 10(6):e0129126. <https://doi.org/10.1371/journal.pone.0129126> PMID: 26111164
26. Sullivan KD, Evans D, Pandey A, Hraha TH, Smith KP, Markham N, et al. Trisomy 21 causes changes in the circulating proteome indicative of chronic autoinflammation. *Scientific Reports.* 2017; 7(1):14818. <https://doi.org/10.1038/s41598-017-13858-3> PMID: 29093484
27. Ahmed MM, Dhanasekaran AR, Block A, Tong S, Costa ACS, Stasko M, et al. Protein dynamics associated with failed and rescued learning in the Ts65Dn mouse model of Down syndrome. *Di Cunto F, ed. PLoS ONE.* 2015; 10(3):e0119491. <https://doi.org/10.1371/journal.pone.0119491> PMID: 25793384

28. Ahmed MM, Dhanasekaran AR, Block A, Tong S, Costa ACS, Gardiner KJ. Protein Profiles Associated With Context Fear Conditioning and Their Modulation by Memantine. *Molecular Cellular Proteomics: MCP*. 2014; 13(4):919–937. <https://doi.org/10.1074/mcp.M113.035568> PMID: 24469516
29. Eicher T. A support vector machine approach to identification of proteins relevant to learning in a mouse model of Down Syndrome [dissertation]. Wichita State University;2016.
30. Feng B, Hoskins W, Zhou J, Xu X, Tang J. Using Supervised Machine Learning Algorithms to Screen Down Syndrome and Identify the Critical Protein Factors. *International Conference on Intelligent and Interactive Systems and Applications*. 2018;302–308.
31. Tsoumakas G, Katakis I. Multi-Label Classification: An Overview. *International Journal of Data Warehousing and Mining*. 3; 1–13. <https://doi.org/10.4018/jdwm.2007070101>
32. Aly M. Survey on Multiclass Classification Methods. Technical report. California Institute of Technology;2005.
33. Kulan H, Dag T. Using Machine Learning Classifiers to Identify the Critical Proteins in Down Syndrome. *2nd International Conference on Bioinformatics and Computational Intelligence*. 2018.
34. Dua D, Karra Taniskidou E. UCI Machine Learning Repository [Internet] Irvine, CA: University of California, School of Information and Computer Science. 2017. Available from: <https://archive.ics.uci.edu/ml/datasets/Mice+Protein+Expression>.
35. Fanselow MS. Factors governing one-trial contextual conditioning. *Anim Learn Behav*. 1990; 18(3):264–70. <https://doi.org/10.3758/BF03205285>
36. Tibes R, Qiu Y, Lu Y, Hennessy B, Andreeff M, Mills GB, et al. Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. *Mol. Cancer Ther*. 2006; 5:2512–21. <https://doi.org/10.1158/1535-7163.MCT-06-0334> PMID: 17041095
37. Abdi H, Williams LJ. Normalizing data. In: Salkind Neil, editor. *Encyclopedia of research design*. Thousand Oaks,CA: Sage;2010. p.935–8.
38. Kumar V, Minz S. Feature selection: A literature review. *Smart Computing Review*. 2014; 4:211–229.
39. Berthold MR, Cebron N, Dill F, Di Fatta G, Gabriel TR, Georg F, et al. *KNIME: the Konstanz Information Miner. Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*. 2007; Heidelberg-Berlin: Springer-Verlag.
40. Chen J, Huang H, Tian S, Qu Y. Feature selection for text classification with Naïve Bayes. *Expert Syst. Appl*. 2009; 36(3): 5432–5435. <http://dx.doi.org/10.1016/j.eswa.2008.06.054>.
41. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011; 12,2825–2830.
42. Bergstra J, Bardenet R, Bengio Y, Kégl B. Algorithms for hyper-parameter optimization. *Advances in Neural Information Processing Systems*. 2011; 2546–2554.
43. Wong TT. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*. 2015 48. <https://doi.org/10.1016/j.patcog.2015.03.009>
44. Hinton G, LeCun Y, Bengio Y. Deep learning. *Nature*. 2015; 521:436–444. <https://doi.org/10.1038/nature14539> PMID: 26017442
45. Natekin A, Knoll A. Gradient boosting machines, a tutorial. *Frontiers in Neuroinformatics*. 2013; 7:21. <https://doi.org/10.3389/fnbot.2013.00021> PMID: 24409142
46. Ortes C, Vapnik V. Support vector network. *Mach. Learn*. 1995; 20,1–25.
47. Breiman L. Random forests. *Machine Learning*. 2001; 45(1):5–32. <https://doi.org/10.1023/A:1010933404324>
48. McCombe PA, Henderson RD. The Role of Immune and Inflammatory Mechanisms in ALS. *Current Molecular Medicine*. 2011; 11(3):246–254. <https://doi.org/10.2174/156652411795243450> PMID: 21375489
49. Iannello RC, Crack PJ, deHaan JB, Kola I. Oxidative stress and neural dysfunction in Down syndrome. *J Neural Transm Suppl*. 1999; 57:257–67. PMID: 10666681
50. Hoeffler CA, Klann E. mTOR signaling: at the crossroads of plasticity, memory and disease. *Trends Neurosci*. 2010; 33(2):67–75. <https://doi.org/10.1016/j.tins.2009.11.003> PMID: 19963289
51. Beurel E, Grieco SF, Jope RS. Glycogen synthase kinase-3 (GSK3): regulation, actions, and diseases. *Pharmacology & therapeutics*. 2015; 0:114–131. <https://doi.org/10.1016/j.pharmthera.2014.11.016>
52. Reese LC, Tagliatalata G. A Role for Calcineurin in Alzheimer's Disease. *Current Neuropharmacology*. 2011; 9(4):685–692. <https://doi.org/10.2174/157015911798376316> PMID: 22654726

53. Nicoll JAR, Mrak RE, Graham DI, Stewart J, Wilcock G, MacGowan S, et al. Association of Interleukin-1 Gene Polymorphisms with Alzheimer's Disease. *Annals of neurology*. 2000; 47(3):365–368. [https://doi.org/10.1002/1531-8249\(200003\)47:3%3C365::AID-ANA13%3E3.0.CO;2-G](https://doi.org/10.1002/1531-8249(200003)47:3%3C365::AID-ANA13%3E3.0.CO;2-G) PMID: 10716257
54. Dinarello CA. Interleukin-1 in the pathogenesis and treatment of inflammatory diseases. *Blood*. 2011; 117(14):3720–3732. <https://doi.org/10.1182/blood-2010-07-273417> PMID: 21304099
55. Tano T, Okamoto M, Kan S, Nakashiro K, Shimodaira S, Koido S, et al. Prognostic Impact of Expression of Bcl-2 and Bax Genes in Circulating Immune Cells Derived from Patients with Head and Neck Carcinoma. *Neoplasia* (New York, NY). 2013; 15(3):305–314. <https://doi.org/10.1593/neo.121528>
56. Sujashvili R. Advantages of Extracellular Ubiquitin in Modulation of Immune Responses. *Mediators of Inflammation*. 2016; 2016:4190390. <https://doi.org/10.1155/2016/4190390> PMID: 27642236
57. Cunha C, Brambilla R, Thomas KL. A Simple Role for BDNF in Learning and Memory? *Frontiers in Molecular Neuroscience*. 2010; 3:1. <https://doi.org/10.3389/neuro.02.001.2010> PMID: 20162032
58. Calabrese F, Rossetti AC, Racagni G, Gass P, Riva MA, Molteni R. Brain-derived neurotrophic factor: a bridge between inflammation and neuroplasticity. *Frontiers in Cellular Neuroscience*. 2014; 8:430. <https://doi.org/10.3389/fncel.2014.00430> PMID: 25565964
59. Yu X, Lu L, Liu Z, Yang T, Gong X, Ning Y, et al. Brain-derived neurotrophic factor modulates immune reaction in mice with peripheral nerve xenotransplantation. *Neuropsychiatric Disease and Treatment*. 2016; 12:685–694. <https://doi.org/10.2147/NDT.S98387> PMID: 27099498
60. Li F, Tsien JZ. Memory and the NMDA Receptors. *The New England journal of medicine*. 2009; 361(3):302–303. <https://doi.org/10.1056/NEJMcibr0902052> PMID: 19605837
61. Shen CP, Tsimberg Y, Salvatore C, Meller E. Activation of Erk and JNK MAPK pathways by acute swim stress in rat brain regions. *BMC Neuroscience*. 2004; 5:36. <https://doi.org/10.1186/1471-2202-5-36> PMID: 15380027
62. Kidambi S, Yarmush J, Berdichevsky Y, Kamath S, Fong W, SchianodiCola J. Propofol induces MAPK/ERK cascade dependant expression of cFos and Egr-1 in rat hippocampal slices. *BMC Research Notes*. 2010; 3:201. <https://doi.org/10.1186/1756-0500-3-201> PMID: 20637119
63. Lee YS, Ehninger D, Zhou M, Oh JY, Kang M, Kwak C, et al. Mechanism and treatment for the learning and memory deficits associated with mouse models of Noonan syndrome. *Nature neuroscience*. 2014; 17(12):1736–1743. <https://doi.org/10.1038/nn.3863> PMID: 25383899
64. Zhang G, Liu M, Cao H, Kong L, Wang X, O'Brien JA, et al. Improved spatial learning in aged rats by genetic activation of protein kinase C in small groups of hippocampal neurons. *Hippocampus*. 2009; 19(5):413–423. <https://doi.org/10.1002/hipo.20506> PMID: 18942114
65. Pollonini G, Gao V, Rabe A, Palmiello S, Albertini G, Alberini CM. Abnormal Expression of Synaptic Proteins and Neurotrophin-3 in the Down Syndrome Mouse Model Ts65Dn. *Neuroscience*. 2008; 156(1):99–106. <https://doi.org/10.1016/j.neuroscience.2008.07.025> PMID: 18703118
66. Shupp A, Casimiro MC, Pestell RG. Biological functions of CDK5 and potential CDK5 targeted clinical treatments. *Oncotarget*. 2017; 8(10):17373–17382. <https://doi.org/10.18632/oncotarget.14538> PMID: 28077789
67. Shao CY, Mirra SS, Sait HBR, Sacktor TC, Sigurdsson EM. Postsynaptic degeneration as revealed by PSD-95 reduction occurs after advanced A β and tau pathology in transgenic mouse models of Alzheimer's disease. *Acta neuropathologica*. 2011; 122(3):285–292. <https://doi.org/10.1007/s00401-011-0843-x> PMID: 21630115
68. Marin I, Kipnis J. Learning and memory . . . and the immune system. *Learning & Memory*. 2013; 20(10):601–606. <https://doi.org/10.1101/lm.028357.112>
69. Ortega-Martínez S. A new perspective on the role of the CREB family of transcription factors in memory consolidation via adult hippocampal neurogenesis. *Frontiers in Molecular Neuroscience*. 2015; 8:46. <https://doi.org/10.3389/fnmol.2015.00046> PMID: 26379491
70. Ahmed AH, Wang Q, Sondermann H, Oswald RE. Structure of the S1S2 Glutamate Binding Domain of GluR3. *Proteins*. 2009; 75(3):628–637. <https://doi.org/10.1002/prot.22274> PMID: 19003990
71. Rahmani Z, Blouin JL, Créau-Goldberg N, Watkins PC, Mattei JF, Poissonnier M, et al. Down syndrome critical region around D21S55 on proximal 21q22.3. *Am J Med Genet*. 1990; Suppl7:98–103.