NEURON CELL CLASSIFICATION USING MACHINE LEARNING ALGORITHMS: METHODOLOGICAL CONSIDERATIONS

Slade. Matthews, Ian. Spence
Department of Pharmacology
University of Sydney
Camperdown, Australia
slade.matthews, ian.spence@sydney.edu.au

CraigS. McLachlan
Rural Clinical School
University of New South Wales
Sydney, Australia
reperfusion@hotmail.com

Herbert F. Jelinek
Australian School of Advanced Medicine
Macquarie University
Sydney, Australia
Centre for Research in Complex Systems
Charles Sturt University
Albury, Australia
hjelinek@csu.edu.au

ABSTRACT
Retinal ganglion cells of different species have been categorised using different paradigms and resulting in different number of suggested classes/types based on traditional morphological parameters such as cell body size or dendritic diameter. The inherent nature of the neuron’s branching pattern has also been shown to play a role in signal processing and therefore additional features such as fractal dimension and lacunarity were added. Machine learning algorithms (MLA) provide a basis for classification tasks based on large numbers of features. However there are numerous ways of presenting data, different algorithms, validation methods and determination of performance. No studies have been undertaken that investigate the influence of model validation on small datasets with diverse feature parameters. This paper outlines the differences when balanced and imbalanced data is used in combination with six supervised MLAs and two different validation algorithms (LOO and 10-fold) as well as interpreting the results using two performance measures (AUC or accuracy). Our results indicate that the largest effect on MLA outcomes is whether data is balanced or imbalanced. AUC is a more robust decision rule compared to accuracy. The best classifiers for our data were neural networks and logistic regression with an AUC of greater than 0.9.

KEY WORDS
Machine learning algorithms, balanced datasets, verification, performance, neurons.

1. Introduction
Previous work has shown that a machine learning algorithm (MLA) using supervised learning back propagation neural networks and clustering algorithms correctly classified cat retinal ganglion cells [1, 2]. In these cases the membership was well defined. Further work explored additional complex measures including fractal analysis, lacunarity, entropy of orientation, multiscale bending energy, circularity and the correlation dimension [3-8]. This study examines the application of MLAs to classification of rat retinal ganglion cells using these variables.

2.1 Rat retinal ganglion cells
Of the neurons in the retina, the retinal ganglion cells are the most interesting due to their role in information processing. Ganglion cells receive data from photoreceptor cells via connections with bipolar, horizontal and amacrine cells [9]. Many animal retinas including cat, rat, mouse, rabbit, and some primates have been investigated and cell classes, based on their morphology have been suggested [10-16]. The use of logistic regression analysis and artificial neural network analysis (ANN) provides a way of ascertaining the importance of each variable within a classification task [17-19]. Neither logistic regression analysis nor ANN analysis have been used extensively in neuroscience although these methods have been shown to be useful in clinical decision making and other classification tasks [20-23]. A study of rat retinal ganglion cells using ANNs and leave-one-out cross validation as well as logistic regression (LR) analysis correctly identified up to 98% of the cells [24].

Rat RGC types have been shown to be structurally different depending on the location of their dendritic arborisation in different parts of the inner plexiform layer (IPL) of the retina [12-14]. This makes the classification task more difficult because a particular class of retinal ganglion cell may possess different morphology depending on the location. Outer alpha cells have denser dendritic branching patterns and are smaller in size at all retinal locations (eccentricities) compared to the inner alpha cells. Delta cells differ from alpha cells by having
thinner primary dendrites and their dendritic segments are more curving and wavy (Figure 1).

![Image](image.jpg)

Figure 1. Schematic of alpha cell characteristics.

No morphological differences have been reported between the inner and outer delta cells. However the dendritic tree of peripheral delta cells was larger and less densely branched compared to central delta cells. Both alpha and delta cell soma size increases with eccentricity. Such a problem may favour the use of neural network models because they readily use all data simultaneously and hence take into account the eccentricity (location) value and stratification.

## 2. Methods

### 2.1 Rat retinal ganglion cells

The cells of the present study were provided by Leo Peichl [12]. Adult rats were deeply anaesthetized with chloral hydrate and eyes enucleated. Retinal ganglion cells were intracellularly injected with Lucifer Yellow to visualize their morphology. Lucifer Yellow labelled cells were digitized and the binary representations of the α and δ cells at different eccentricities analysed using the public domain Image J software (http://rsb.info.nih.gov/nih-image/). The areas of both the soma and dendritic field of each cell were measured in a way as described in previous publications [12]. After the measurements, each cell was subjected to complexity analysis.

### 2.2 Complexity Measures

Prior to analysis the cell body and axon were removed digitally. The box counting method and lacunarity analysis were used for this study [25, Error! Hyperlink reference not valid.]. The box-counting analysis covers the image with sets of squares. Each set is characterised by the box size $r$ of the square edge. Box sizes were taken from 1 to 1024 pixels. The number of squares $N(r)$ necessary to cover the object is presented as a function of $r$. The fractal dimension ($D$) is determined from the slope $S \text{ of } \log N(r)/\log r$, as $D = 1-S$.

Lacunarity is determined from the probability distribution for pixels. The distribution is determined from the number of pixels per box as a function of box size or scale ($\epsilon$), which is inversely proportional to the box size [25].

Roundness is a measure of the irregularity of the image boundary and defined as the ratio of the curvature of the corners and edges of an object to the curvature of the object as a whole [27].

Circularity is the mean radial distance to the tumour boundary divided by the standard deviation of the tumour boundary [28]. It increases as the object becomes more circular and is therefore a measure of the regularity of the shape of the object.

### 2.3 Machine Learning Algorithms

The dataset contained 67 alpha cells and 14 delta cells. It contained 5 complexity features, (fractal dimension, mean box count, lacunarity, roundness and circularity) as well as eccentricity, cell area, and maximum dendritic length. Baysian Network, Naïve Baysian, Neural Network, Decision Tree and 10-Nearest neighbour (10-NN) and the more commonly used logistic regression were employed for separation of the retinal ganglion cells. All MLA were calculated as part of the Weka analysis program [29, 30].

### 2.4 Power Analysis

Power analysis was used to determine the number of rounds of 10-fold cross validation required to provide a reasonable chance of detecting a difference in the performance of the algorithms. The $n$ value for the comparison of predictive performance is equal to the number of results per test set multiplied by the number of rounds of cross validation [31]. Power calculations for comparison of proportions were then used to calculate $n$ values that would provide an 80% chance of detecting a difference of at least 10% between the proportion of correctly identified cells for each algorithm.

### 2.4 Validation

Two methods of model validation were used in the present work to evaluate performance of the RGC classification. The first was 10-fold cross validation. This technique splits the data into 10 “folds” and trains the model on the first 9 folds and tests the ability of the model to predict the target variable (cell classification) on the tenth fold. The process is repeated until all ten folds have been used for validation based on a combination of the other nine. The second method for model evaluation was leave-one-out cross validation (LOO). This method works in almost exactly the same way as 10-fold cross validation except that each fold consists of only a single example from the data. This method is far more computationally expensive than 10-fold but may provide a better estimate of model performance because the composition of the training dataset can not be randomly assigned to have more of one target class than another.
3. Results

The stratification of alpha and delta cells for the imbalanced dataset was: delta (10 outer: 4 inner) and alpha (24 outer: 43 inner). For the balanced set the stratification was delta (10 outer: 4 inner) and alpha (8 outer: 6 inner).

3.1 10-fold cross validation

The first validation task was the 10-fold task (Figure 2). Some differences in ROC between the 6 MLAs were noted with the Neural Network model performing slightly better than the others except for the decision tree application.

![Figure 2. ROC curves for imbalanced data using 10-fold cross validation.](image)

The ROC curves for each algorithm averaged from ten folds for the balanced dataset are shown in Figure 3. The average curve for the nearest neighbour appears to be much lower than those of the other algorithms. There is also a cluster of curves in the middle area of the graph and then a gap between those and that of logistic regression.

![Figure 3. ROC curves for balanced data using 10-fold cross validation.](image)

Table 1 compares area under the curve (AUC) and accuracy as an indicator of performance and compared using Duncan’s multiple range test (DMRT). The DMRT placed the top five algorithms in the same range and consistently put the decision tree in a lower group. For the balanced data the order of increasing performance is slightly different but not significant (Table 2).

<table>
<thead>
<tr>
<th></th>
<th>DT</th>
<th>NB</th>
<th>BN</th>
<th>10-NN</th>
<th>LR</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>0.76</td>
<td>0.88</td>
<td>0.89</td>
<td>0.92</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.89</td>
<td>0.78</td>
<td>0.85</td>
<td>0.83</td>
<td>0.93</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Table 1. Algorithm performance on imbalanced dataset (10-fold CV)

For the balanced set the nearest neighbour and neural network retained the same placement while all others change their relative order of performance. Both AUC and accuracy estimates are smaller for all algorithms. Range comparisons for the average AUC measure from each set of ten folds were conducted using Duncan’s multiple range test (Table 2). Duncan’s test failed to find any difference between the AUC for the algorithms, which suggests that the change in order of the algorithms is simply due to chance variation in the resulting averaged AUC values.

<table>
<thead>
<tr>
<th></th>
<th>10-NN</th>
<th>BN</th>
<th>NB</th>
<th>NN</th>
<th>DT</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>0.70</td>
<td>0.78</td>
<td>0.83</td>
<td>0.83</td>
<td>0.85</td>
<td>0.90</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.57</td>
<td>0.89</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
</tr>
</tbody>
</table>

Table 2. Algorithm Performance on Balanced RGC Dataset (10-fold CV)

3.2 Leave-one-out validation

In order to determine the possible gains in resolution of true error estimation that may be achieved using LOO cross validation the six algorithms were trained on the imbalanced and balanced datasets.

The ROC curves for each algorithm trained on the imbalanced dataset as estimated by LOO cross validation are displayed in Figure 4. All the curves are very close together which indicates that their classification performance is very similar.

![Figure 4. ROC Curves for Imbalanced Dataset using LOO Cross Validation.](image)
Table 3 shows the AUC and accuracy scores for the different algorithms arranged in order of increasing AUC score. The order has changed slightly compared to the results of the 10-fold cross validation error estimates. This time the nearest neighbour (10-NN) has slipped back to 5th place and the order of logistic regression and neural network has switched. Decision tree is still at the bottom of the performance rankings although its score is not as far below the others when estimated with LOO cross validation.

<table>
<thead>
<tr>
<th></th>
<th>DT</th>
<th>10-NN</th>
<th>NB</th>
<th>BN</th>
<th>LR</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>0.88</td>
<td>0.92</td>
<td>0.92</td>
<td>0.93</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.85</td>
<td>0.83</td>
<td>0.77</td>
<td>0.84</td>
<td>0.94</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Table 3. Algorithm Performance on Imbalanced Dataset (LOO CV)

The balanced data was once again presented to each of the six algorithms, this time building models on all but one instance and performing one test for each instance (leave-one-out cross validation). The ROC curves for each algorithm shown in Figure 5.

Figure 5. ROC Curves for balanced Dataset using LOO Validation

These are not averaged curves as with the 10-fold cross validation results because as stated earlier it is not necessary to compensate for the variation due to the constitution of each of the sets. The curve for the nearest neighbour algorithm is lower and has fewer points than the other curves. The curve for decision tree also has few points. The reduced number of points on the curve indicates that the predictions made by the algorithm could be separated at fewer places using thresholds. This is because the algorithms gave fewer unique answers in terms of the value of the output probability distribution of the algorithm which is then compared to the threshold and used to calculate the proportion of True and False positive results.

The order of error estimations are tabulated in order of increasing AUC value (Table 4). This order has changed slightly from that seen with the 10-fold cross validation estimation. The positions for nearest neighbour and logistic regression did not change but the rankings for the other four algorithms were different with LOO estimation.

<table>
<thead>
<tr>
<th></th>
<th>10-NN</th>
<th>DT</th>
<th>NN</th>
<th>BN</th>
<th>NB</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>0.56</td>
<td>0.86</td>
<td>0.88</td>
<td>0.88</td>
<td>0.91</td>
<td>0.93</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.54</td>
<td>0.86</td>
<td>0.86</td>
<td>0.93</td>
<td>0.89</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Table 4. Algorithm performance on balanced dataset (LOO CV)

4. Conclusion

Rat ganglion cell morphology has been analysed previously indicating a far greater variety and complexity than initially thought. Brown first proposed two classes, followed by the work of Perry, who suggested 3 classes and cumulating in Huxlin and Goodchild suggesting 3 main classes consisting of at least 10 groups [14, 15, 32]. These classification systems are based primarily on cell body size and visual interpretation of the dendritic field complexity.

The morphological approach to classification has been shown to increase our understanding of cell function as evidenced by the finding that the inner and outer stratifying alpha cells in cat retina correspond to ON centre and OFF centre physiological types [33]. It is not possible to classify all rat retinal ganglion cells on the basis of their physiology because some are very small. Rat retinal ganglion cells range in diameter from as small as 5 to 24 microns [34].

The main question this study set out to answer was whether the MLAs could classify rat retinal ganglion cells using histological measurements. When the complete dataset (all the cells and all the measurements) was used all of the supervised algorithms were capable of separating the cells into their different classes using the data presented to them. This was true regardless of the way performance was assessed. The algorithms all performed similarly on the classification task in contrast to previous work, which suggested different classifiers do not usually give the same results on neural morphological classification tasks [35].

3.2 The problem of imbalanced data

All supervised algorithms were able to classify the retinal ganglion cells into alpha and delta with high degrees of accuracy on the imbalanced dataset. Training these algorithms on the balanced data improved classification performance. This improvement was unexpected due to the reduction in sample size as the predictive accuracy of neural networks and MLAs has been shown to depend on training dataset size [36]. On the other hand there have been some but infrequent references in the literature to the effect of imbalanced data on MLA performance. Provost reminds us that the use of imbalanced data can be a trap for researchers wishing to build predictive models from
datasets are small (less than 100) because of the small validation. It is frequently suggested that it be used when any additional benefit from the use of leave one out cross validation takes a fraction of the time it would take one with a multipl

Accuracy is the first casualty of the imbalanced dataset. If an algorithm over predicts the dominant class then overall accuracy will remain high. Even in a balanced dataset ROC area provides a truer representation of the ability to detect in the dataset a signal indicative of output class. If the MLA modelled the data well then both estimates, AUC and accuracy provide a reasonable measure of performance. If the MLA is unable to model the data, accuracy tends to overestimate the performance, as mentioned especially with imbalanced datasets. Thus our results confirm that AUC is the preferred method for evaluation of MLA performance [37].

3.3 Cross validation

Range comparisons for the 10-fold cross validation estimates of AUC for the complete imbalanced dataset showed that decision tree’s performance was significantly lower than that of the other algorithms. No differences could be detected using LOO cross validation between the classifiers. Given that decision tree gave the lowest AUC scores for both 10-fold and LOO cross validation the former may have detected a small real difference in the performance that the latter could not. This result would be enough to at least guide algorithm selection by a researcher toward more linear discriminating algorithms for the present data. Duncan's multiple range test for the 10-fold cross validation AUC estimate for the balanced dataset showed the presence of two possible ranges. A lower range including nearest neighbour and decision tree and an upper range including the remaining four algorithms.

These results suggest that decision tree and nearest neighbour should not be the first choice with this data and that any of the other four algorithms should perform very well. It appears that 10-fold cross validation as compared with a multiple range test is more powerful than leave-one-out cross validation. Furthermore given that 10-fold cross validation takes a fraction of the time it would take to run LOO it is a much more economical method.

None of the results on all of these datasets indicate any additional benefit from the use of leave-one-out cross validation. It is frequently suggested that it be used when datasets are small (less than 100) because of the small amount of training data however the results of the present work suggest that it is unnecessary [38].

Machine learning algorithms are of great benefit to the biological sciences due to their ability to delve into complex and sometimes noisy data and bring into light complex and implicit relationships. This paper demonstrates excellent performance in rat retinal ganglion cell classification and outlines methodology issues in the use of machine learning for the modelling of classification decisions made by experts in the field of neurobiology thus supporting the foundation of neurobiological classification.

References

351 vicinity of the optic disk in digital retinal images. W, Cree MJ. Towards vessel characterisation in the
Jelinek HF, Depardieu C, Lucas C, Cornforth D, Huang W, Cree MJ. Towards vessel characterisation in the
vicinity of the optic disk in digital retinal images. Image and Vision Computing New Zealand 2005, Otago; 2005,
351-7.