Algorithm for Testing Behavioural Phenotypes in a Zebrafish Model of Parkinson's Disease

Angela Pimentel[1], Hugo Gamboa[1,2], Sérgio Reis Cunha[3] and Ana Dulce Correia[4]
[1]*CEFITEC, Physics Department, FCT-UNL, Lisbon, Portugal*
[2]*PLUX - Wireless Biosignals, Lisbon, Portugal*
[3]*Faculty of Engineering, Porto University, Porto, Portugal*
[4]*Instituto de Medicina Molecular, Faculty of Medicine, University of Lisbon, Lisbon, Portugal*

Keywords: Parkinson's Disease (PD), Zebrafish, Behaviour, Biosensor MOBS, Machine Learning.

Abstract: Parkinson's disease (PD) is one of the neurodegenerative diseases with an increased prevalence widely studied by the scientific community. Understanding the behaviour related to the disease is an added value for diagnosis and treatment. Thus the use of an animal model for PD that develops similar symptoms to the human being allows to the clinic a larger vision over the health of a patient. Zebrafish can be used to study some human diseases including PD. This work describes the development of an algorithm for the characterization of behaviour in this specie. The biosensor called Marine On-line Biomonitor System (MOBS) is connected electrically to chambers where the specimen of zebrafish moves freely providing a signal that is related with the fish activity. Using the developed algorithm based on signal processing, statistic analysis and machine learning techniques we present classification of a fish as normal or ill and characterize its behaviour.

1 INTRODUCTION

Biosensors are an essential control and safety tool for our environmental and health quality and commonly used in medicine. Many of today's biosensor applications use living organisms which respond to toxic substances or other stressors at a much lower level than us to warn us of their presence. Under this scope, the MOBS was developed, an automated system for recording behavioural responses of marine and fresh water species. This device has been applied successfully in the environmental field, and the next challenging step is to bring this technology into other research areas. In particular, by sensing behavioural changes in organisms as an indication of stress or disease. A suitable model candidate is the zebrafish, a freshwater specie which has been used in medical research during the past years, e.g in development studies (Lepage and Bruce, 2008), drug toxicity assessments (Usenko et al., 2008) and neurodegenerative diseases (Bretaud et al., 2004).

1.1 PD and Zebrafish

The PD is characterized by tremor, muscle rigidity, a slowing of physical movement, and can also cause cognitive and mood disturbances. It results of the loss of nerve cells in part of the brain known as the substancia nigra. These cells are called dopaminergic (DA) neurons as they produce the neurotransmitter, dopamine, which is used to send messages to the parts of the brain that co-ordinate movement (Fish for Science, 2012). Most insights into human disease are a result of experiments that would be unethical or unfeasible to perform on humans. Instead biomedical research uses models to look at the functions of the genes involved in maintaining healthy organisms in order to obtain vital clues about the causes and progression of human diseases. Zebrafish are an ideal model organism to bridge the gap between too simple (yeast) and too complex (mice or rats). They are vertebrates and have similar body plans (and similar tissues and organs) to humans, and they're much easier and with reduced cost to breed than mice and rats. Zebrafish mutations phenocopy many human disorders and the genome sequence of zebrafish is near completion. The DA nervous system in zebrafish is well characterized in both embryos and adult zebrafish. Some toxins known to induce DA cell loss in other animal models have now also been tested in adult zebrafish, as for example, the *6-hydroxydopamine* (6-

Pimentel A., Gamboa H., Reis Cunha S. and Dulce Correia A..
Algorithm for Testing Behavioural Phenotypes in a Zebrafish Model of Parkinson's Disease.
DOI: 10.5220/0004238101960202
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing* (BIOSIGNALS-2013), pages 196-202
ISBN: 978-989-8565-36-5

OHDA) which is a neurotoxin that induces death of the DA cells. After injecting the neurotoxin via intramuscular, locomotor activity and dopamine levels of the brain decreases (Kalueff and Cachat, 2010; McGrath, 2012; Breese et al., 2005; Flinn et al., 2008). Thus the evaluation of swimming behaviour can be related with the loss of DA cells, and consequently with the PD. In the work performed by (Correia, Ana Dulce and Soares, Rui S. and Sousa, Sara and Outeiro, Tiago F. and Afonso, Nuno and Willemsen, Rob and Herma van der Linde, 2012) a new transgenic line of zebrafish was developed to study the DA neurons, which were validated with the use of the neurotoxin 6-OHDA and with the behaviour analysis using the biosensor MOBS. They verified behavioural changes that were related with the death of the DA neurons. The algorithm to be developed can be a contribution for this work: an algorithm that is sensible in the behaviour characterizations to allow the responses to be comparable with the loss of the DA neurons.

1.2 Current Approach

The current algorithm used to characterize the behaviour of zebrafish consists in the evaluation of a specific locomotion behaviour, with a series of bursts in the domain of MOBS corresponding to the tail-flip activity of zebrafish. Thus the outcome reflects the number of tail-flips per minute per individual fish (Correia et al., 2011). The behaviour detection is based on the derivative peaks resulted from the strong bursts in the signal. However, these peaks require a threshold for the behaviour detection, and this is accomplished using the standard deviation multiplied by a factor so that these two parameter, standard deviation and derivative, may be comparable. It's essential to confirm if the current algorithm is in fact detecting the right behaviour, the tail-flips. The first intention of this research would be to understand and improve the current algorithm, however it will be proved the need to create a new one using supervised learning.

1.3 Supervised Learning

By Arthur Samuel (1959), machine learning is the field of study that gives computers the ability to learn without being explicitly programmed. There are different types of machine learning algorithms, the main two types are: unsupervised and supervised learning.

With supervised learning, the scheme operates under supervision by being provided with the actual outcome for each of the training examples. In this type of machine learning is included regression problems that predicts continuous valued outputs and classification problems which intends to predict discrete valued outputs (Machine Learning, 2012). For classification problems, a known method is the Support Vector Machine (SVM) which looks for the optimal hiperplane between two classes by maximizing the margin. A non-linear separator is possible by projection the data points to higher-dimension space to become linearly separable (projection with kernel techniques) (Machine Learning, 2012). Also the method Naïve Bayes which applies Bayes theorem to estimate the probability with the "naïve" assumption of independence between each feature. For validation, a possible statistic test is *leave one out*, which given a dataset of m instances, only one instance is left out as the validation set (instance) and training uses the $m-1$ instances (Witten et al., 2011).

2 METHODS

2.1 MOBS

The main device is controlled via an USB port by external processing software which produces signals in the digital domain (at 48000 samples/s or 48 kHz). These are converted by the main device into analogical electrical signals, power amplified and transmitted to the independent testing units at which they are conducted into the water by a pair of non-invasive stainless steel electrodes. In response to the behavioural signatures of the organisms as a change in impedance of the water, the amplitudes of the electrical signals are modulated and then received by a second pair of electrodes. In the main device they are amplified and converted back to the digital domain at 48000 samples/s, before filtered, demodulated and downsampled at 100 Hz by the external computer software. Upon processing, the system provides a signal in the frequency band of 0.2 Hz to 40 Hz that is correlated with the fish activity (Cunha et al., 2008). With MOBS, locomotion can be presented with a series of bursts in the time domain, and can cover a broad frequency spectrum, at which ventilation is occasionally present. Typically ventilation generates waves of triangular shape with a higher frequency and smaller amplitude than the most of the energy located for locomotion. However ventilation will not be studied with zebrafish given its high level of activity.

2.2 Experimental Design

2.2.1 Test Animals and 6-OHDA

The zebrafish (D. rerio Hamilton 1822) strain used for this work was the AB line (Zebrafish Facility, IMM, Portugal). Animals were maintained under standard conditions and experiments were approved by the Institutional Animal Care and Use Committee. A master stock solution of 6-hydroxydopamine hydrochloride (6-OHDA, Sigma-Aldrich, USA) was prepared in 0.2% ascorbic acid solution (analytical grade, Sigma) and stored at -20°C.

2.2.2 Behaviour Assay

Before the experiments, small groups of female fish (24 animals, body weight 0.5 $\pm$ 1 g) were acclimatized to the experimental testing conditions (temperature 22 °C $\pm$ 1 °C, 10 h:12 h light-dark cycle) in 17 litre glass aquaria under static conditions and for a minimum of one week. Food was not provided 24 h before or during the experiments. The behaviour analysis was divided in two groups: non-treated (12 fish) and for that considered as normal fish in which no injection was administered, and treated (12 fish) also considered as ill or less active where 5μL of 6-OHDA (33 mg/kg) was injected via intramuscular. During the injection they were in a medium-to deep-plane level of anaesthesia (tricaine 50 mg/L) and had lost their reflex responses and muscular control. Afterwards they returned to their original test chambers and allowed 30 min to recover from the anaesthesia. On the day of experiments, either the treated or non-treated groups of fish were placed individually in the test chambers (22 °C $\pm$ 1 °C) and acclimated for 30 min. Then individual baseline responses were monitored using MOBS and recorded using video (property of 25 frames per second) for five minutes between 10 and 12 a.m. After behavioural recording, treated fish were sacrificed with tricaine.

2.3 Synchronism

The signal in the time domain is delayed in relation to the instant of acquisition start. This delay is caused by the main device, which makes it difficult to compare a video where the fish movements are present, with its respective signal from MOBS. The Open Signals is a platform designed and programmed by *PLUX - Wireless Biosignals*. Using this platform, synchronism is possible with a visible stimulus in the signal and video. This stimulus must be sufficient to not be confused with the fish activity in the signal. A touch in the chamber is a possible stimulus and to not corrupt the signal from the fish activity for further analysis the stimulus should be produced at the end of the recording.

2.3.1 Visual Analysis

To verify what the algorithm is detecting a detailed analysis using Open Signals was necessary after synchronism. This analysis using the video frame by frame consisted in the detection of the behaviour tail-flip. The tail-flip is characterized as an abrupt and fast change of direction implying a strong burst in the tail. The visual analysis will consist in counting the number of tail-flips detected and divide it by the total time in minutes. Since the visual analysis is a long process, 24 study cases were made, 12 of them were non-treated and the rest were submitted to the drug 6-OHDA. Each visual analysis consisted in 3 minutes of the video. Since the visual analysis depends of the user that is interpreting the data, it's important to test other user and compare the results. A visual test using a different user was made. The test consisted in a precise analysis frame by frame using a signal with 30 seconds, and for this time both users detected 46 abrupt tail-flips. After the User 1 detect the abrupt tail-flip it was considered an interval of 0.25 seconds in which the User 2 had also to detect the same abrupt tail-flip to be a valid success.

2.4 Current Algorithm Evaluation

In this subsection is intended to compare the visual analysis with the algorithm result using linear regression for each group (treated and non-treated) and estimate the relative error with the *leave one out* method. This was chosen because the number of points analysed is small. Also in consideration is the correlation coefficient which is a numerical value that indicates the degree and direction of relationship between two variables (O'Toole, 2006). The relative error obtained will show the need to improve the algorithm.

The multiplicative factor in the current algorithm is used so that the derivative can be comparable to the standard deviation thus allowing to detect the behaviour abrupt tail-flips. Given that, to improve the algorithm the multiplicative factor should be analysed. The value used so far has been 0.1. To understand which is the best factor value, it was decided to vary the factor according to the outcome of the algorithm, and with the visual analysis choose the factor that was closer to reality. A unique study case isn't sufficient to choose the ideal factor, thus with all data analysed for each group it's estimated the average relative error

from the current algorithm result with the visual analysis. In the end we chose the factor that minimizes the relative error.

2.5 New Algorithm

2.5.1 Behaviour Characterization

To characterize the behaviour in number of tail-flips per minutes it was necessary to use the parameter zero crossing rate. The zero crossing rate it is defined as the number of time-domain zero crossings within a defined region of signal, divided by the number of samples of that region (Gouyon et al., 2000). Each data was divided by its standard deviation, so that all data is at the same scale to be comparable and because the signal is centred at zero, it wasn't necessary to subtract the average. Also the signal was smoothed using a Hanning window of 0.05 seconds. To validate this parameter it was used the statistic analysis *leave one out*. This was chosen because the number of points analysed is small. This study also considered the correlation coefficient.

2.5.2 Classification

The Orange is a software suitable for machine learning. It is a free software and open source. It allows to use data mining through visual programming and Python scripting (Curk et al., 2005). The classifier was studied with the methods SVM and Naïve Bayes. The validation used the statistic analysis *leave one out* to provide the accuracy for each method used (SVM or Naive Bayes) which is the proportion of correctly classified examples (Curk et al., 2005). Thus varying the number of parameters obtained from the data we choose the ones that give higher accuracy for the respective method. The parameters extracted from each data were the zero crossing rate, the standard deviation, the maximum power using the periodogram, the maximum number of occurrences using the histogram and the current algorithm output. Also the optimal values for the SVM, namely the Cost parameter and the gamma value for the kernel function were chosen by the Orange software which uses the LIBVSM library. Since the classifier doesn't require the visual analysis as output, which is a long process, instead of using the data obtained so far (24 study cases), it was used data from a previous work to provide more points to the classifier (108 study cases with equal number for each class). This work developed at the Instituto de Medicina Molecular provides data with non-treated and treated fish (submitted to the drug 6-OHDA).

3 RESULTS AND DISCUSSION

3.1 Synchronism

3.1.1 Visual Analysis

In 46 detections between both users, 44 were accepted, leading to an error of 4.35%. The agreement between the users characterizing the behaviour, leads that the visual result can be a valid information to be compared with the current algorithm or with future works.

3.2 Current Algorithm Evaluation

We can now compare the algorithm output with the visual analysis. The results are shown in figure 1. It

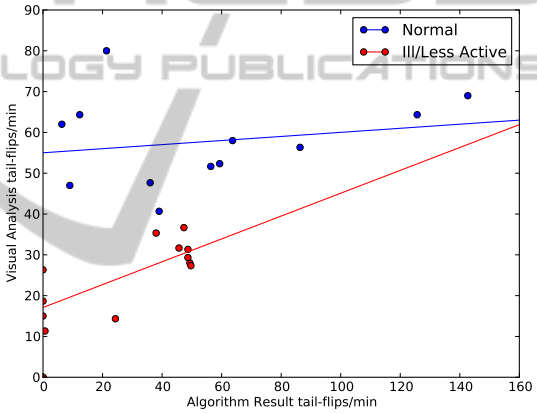

Figure 1: Comparison between the visual analysis and the algorithm result.

is visible that there is no direct relation between the visual analysis and the algorithm output as it would be expected both for treated and non-treated fish. After applying linear regression in each group it was estimated the relative error with the method *leave one out* which resulted in an error of 17.29% for the non-treated and 25.31% for treated. Also the correlation coefficient obtained was 0.20 and 0.76 for the non-treated and treated respectively which can be considered as a poor relation between the visual analysis with the algorithm output. These errors imply an improvement in the algorithm, more specifically in the multiplicative factor. To choose the best factor it was decided to study the error associated with the visual analyse. Figure 2 indicates the minimum error accepted as well as the error used with the actual factor for the treated and non-treated fish. The error using the actual factor is 55.26% and 68.79%

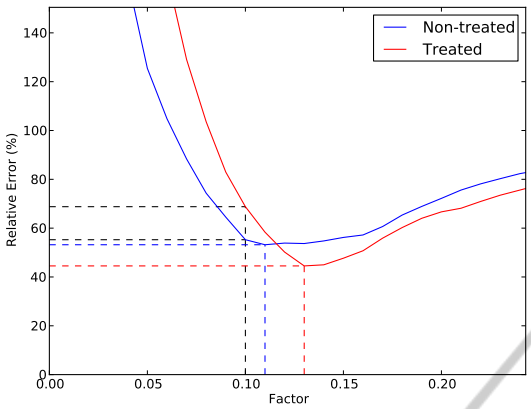

Figure 2: Relative error in percentage. Black dotted lines: Actual multiplicative factor (0.1); Red dotted lines: Best multiplicative factor for treated; Blue dotted lines: Best multiplicative factor for non-treated.

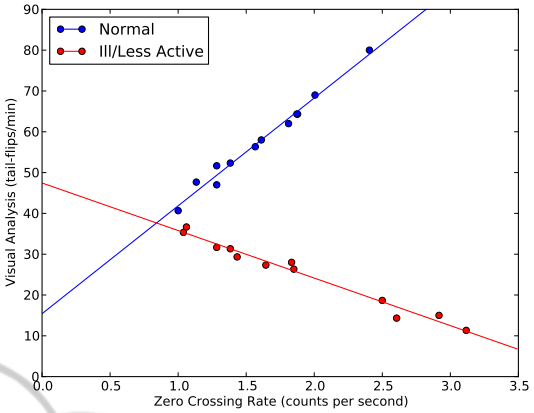

Figure 3: Comparison between the zero crossing rate with the visual analysis for normal and ill fish with a window of 180 seconds.

for non-treated and treated respectively, and even improving the factor, the minimal error accepted would be 53.20% for non-treated which leads to a best factor of 0.11 and 44.53% for treated with a best factor of 0.13. To be able to choose the best factor these errors obtained should be as close to zero as possible which indicates that even with these improvements the best multiplicative factor cannot be certain. Therefore, and considering that the visual analysis is a valid measure, it is suggested the development of a new algorithm.

3.3 New Algorithm

With the visual analysis it will be possible to study new parameters using supervised learning, more precisely, regression models.

3.3.1 Behaviour Characterization

Figure 3 shows visually that there is a linear tendency between the zero crossing rate results with the visual analysis both for treated and non-treated fish. Considering first the normal fish for validation, it was used the statistic analysis *leave one out*. The result leaded to a error of 2.55%. The relative error of 2.55% compared with the 17.29% from the previous algorithm can be considered as an excellent improvement. The user test from the previous section showed an error of 4.35%. Given that, the reason why this parameter shows a smaller error (2.55%) it's because it suits the user that performed this analysis. If User 2 had also performed this analysis, it should be expected a bigger error. The correlation coefficient obtained in this case was 0.99, indicating that there is a very good positive relation between the zero crossing rate and the visual analysis. Finally using all points for a window of 180 seconds, linear regression can be applied to define our hypothesis:

$$h_\theta(x) = 15.42 + 26.43x \quad (1)$$

To characterize the behaviour for ill fish, Figure 3 shows that this group presents an inverse linear tendency between the zero crossing rate and visual analysis, which means that the higher the number of counts per second the less active the fish is. Again it was used the *leave one out* method to validate this parameter. The relative error obtained was 5.75% which can be a good estimative even thought it's higher than the error obtained to characterize normal fish (2.55%). This error in comparative to the 25.31% from the previous algorithm can also be considered as an excellent improvement. The correlation coefficient was -0.99, meaning there is a very good inverse relation between the visual analysis and zero crossing rate.

Using all points for a window of 180 seconds linear regression can be applied to define our hypothesis:

$$h_\theta(x) = 47.45 - 11.65x \quad (2)$$

The value of 47.45 tail-flips per minute limits the fish activity, which means that ill fish won't show a higher value of activity than 47.45 tail-flips per minute. Also for a fish that doesn't present any activity (0 tail-flips per minute) it should be expected a value of 4.07 counts per second. Since both groups use different equations to characterize the behaviour, to know which equation to use for the development of this algorithm a classifier is needed to distinguish between normal or ill fish.

3.3.2 Classification

Now our output is defined by two classes: normal and ill fish. The parameters used that leaded to a higher accuracy for the SVM were the zero crossing rate, the standard deviation, the maximum power using the periodogram, the maximum number of occurrences using the histogram, and the previous algorithm output. The learning options used were for the kernel function the Sigmoid function ($tanh(8 * x.y)$), a Cost of 2.0 (Model Complexity - penalty parameter) and a numeric precision of 0.001. The accuracy obtained using *leave one out* for the SVM method was 100%, meaning that all cases analysed were correctly classified. On the other hand, the Naive Bayes method based on the relative frequency presented a maximum accuracy of 67.59% using the parameters standard deviation, the maximum power using the periodogram and the previous algorithm output.

As we want to choose the classifier that predicts the classes with a higher accuracy value we choose the method SVM to build our final classifier. Because the Orange program is open source, with the access to the functions that build the classifier SVM we can use them to construct the final algorithm in python.

3.3.3 Final Algorithm

Now it's possible to built the final algorithm. First we prepare the data with the removal of the initial peak from the main device, the application of a filter to exclude possible noise, the normalization of the data and the smooth of the signal using a Hanning window of 0.05 seconds. Then we use the classifier to predict if the fish is normal or ill. Consequently, according to the classification it's possible to characterize the behaviour in number of tail-flips per minute using the corresponding hypothesis that consists in the use of the parameter zero crossing rate. The final result will present the classification, the probability for that classification, and the number of tail-flips per minute.

4 CONCLUSIONS

A new algorithm was developed to classify and characterize the behaviour of zebrafish. To facilitate its use, the algorithm should be integrated in the platform Open Signals. The fact that this algorithm uses classification can be an advantage as it may bring an efficient separation between a healthy fish from one that has been genetically modified to have PD. Also, the algorithm should be applied in a case study as executed by (Correia, Ana Dulce and Soares, Rui S. and Sousa, Sara and Outeiro, Tiago F. and Afonso, Nuno and Willemsen, Rob and Herma van der Linde, 2012), to verify that the responses are in agreement with the fish behaviour and literature. This algorithm may be useful for further studies not only related with PD, but any other that uses zebrafish behaviour as an end point to study human diseases.

REFERENCES

Breese, G. R., Knapp, D. J., Criswell, H. E., Moy, S. S., Papadeas, S. T., and Blake, B. L. (2005). The neonate-6-hydroxydopamine-lesioned rat: a model for clinical neuroscience and neurobiological principles. *Brain research reviews*, 48(1):57–73.

Bretaud, S., Lee, S., and Guo, S. (2004). Sensitivity of zebrafish to environmental toxins implicated in parkinson's disease. *Neurotoxicology and teratology*, 26(6):857–864.

Correia, A. D., Cunha, S. R., Scholze, M., and Stevens, E. D. (2011). A novel behavioral fish model of nociception for testing analgesics. *Pharmaceuticals*, 4(4):665–680.

Correia, Ana Dulce and Soares, Rui S. and Sousa, Sara and Outeiro, Tiago F. and Afonso, Nuno and Willemsen, Rob and Herma van der Linde (2012). Green fluorescent protein labeling of dopaminergic neurons in zebrafish for the study of the molecular basis of parkinson's disease (submitted).

Cunha, S. R., Gonçalves, R., Silva, S. R., and Correia, A. D. (2008). An automated marine biomonitoring system for assessing water quality in real-time. *Ecotoxicology*, 17(6):558–564.

Curk, T., Demsar, J., Xu, Q., Leban, G., Petrovic, U., Bratko, I., Shaulsky, G., and Zupan, B. (2005). Microarray data mining with visual programming. *Bioinformatics*, 21(3):396–398.

Fish for Science (2012). http://www.fishforscience.com/.

Flinn, L., Bretaud, S., Lo, C., Ingham, P. W., and Bandmann, O. (2008). Zebrafish as a new animal model for movement disorders. *Journal of Neurochemistry*, 106(5):1991–1997. PMID: 18466340.

Gouyon, F., Pachet, F., and Delerue, O. (2000). On the use of zero-crossing rate for an application of classification of percussive sounds. In *Proceedings of the COST G-6 conference on Digital Audio Effects (DAFX-00), Verona, Italy*.

Kalueff, A. V. and Cachat, J. M., editors (2010). *Zebrafish Models in Neurobehavioral Research: 52*. Humana Press, 1st edition. edition.

Lepage, S. E. and Bruce, A. E. E. (2008). Characterization and comparative expression of zebrafish calpain system genes during early development. *Developmental Dynamics*, 237(3):819–829.

Machine Learning (2012). https://class.coursera.org/ml/lecture/preview.

McGrath, P. (2012). *Zebrafish: Methods for Assessing Drug Safety and Toxicity*. John Wiley & Sons.

O'Toole, M. T. (2006). Miller-keane encyclopedia & dictionary of medicine, nursing & allied health-second revised reprint. *Recherche*, 67:02.

Usenko, C. Y., Harper, S. L., and Tanguay, R. L. (2008). Fullerene c_{60} exposure elicits an oxidative stress response in embryonic zebrafish. *Toxicology and applied pharmacology*, 229(1):44–55.

Witten, I. H., Frank, E., and Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier.