Computers in Biology and Medicine 114 (2019) 103466

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Phasic dopamine release identification using convolutional neural network

Gustavo H.G. Matsushita [a,*], Adam H. Sugi [b,c], Yandre M.G. Costa [d], Alexander Gomez-A [e], Claudio Da Cunha [b,c], Luiz S. Oliveira [a]

[a] Department of Informatics, Federal University of Parana, Curitiba, PR, Brazil
[b] Department of Biochemistry, Federal University of Parana, Curitiba, PR, Brazil
[c] Department of Pharmacology, Federal University of Parana, Curitiba, PR, Brazil
[d] Department of Informatics, State University of Maringa, Maringa, PR, Brazil
[e] Bowles Center for Alcohol Studies, University of North Carolina, Chapel Hill, NC, USA

ARTICLE INFO

Keywords:
Phasic dopamine release
Fast-scan cyclic voltammetry
Convolutional neural network
YOLO
Pattern recognition
Machine learning

ABSTRACT

Dopamine has a major behavioral impact related to drug dependence, learning and memory functions, as well as pathologies such as schizophrenia and Parkinson's disease. Phasic release of dopamine can be measured *in vivo* with fast-scan cyclic voltammetry. However, even for a specialist, manual analysis of experiment results is a repetitive and time consuming task. This work aims to improve the automatic dopamine identification from fast-scan cyclic voltammetry data using convolutional neural networks (CNN). The best performance obtained in the experiments achieved an accuracy of 98.31% using a combined CNN approach. The end-to-end object detection system using YOLOv3 achieved an accuracy of 97.66%. Also, a new public dopamine release dataset was presented, and it is available at https://web.inf.ufpr.br/vri/databases/phasicdopaminerelease/.

1. Introduction

The research to understand the role of the neurotransmitter dopamine (DA) in brain functions under normal and pathological states is one of the most active areas of neuroscience. Dopamine is the main modulator of the glutamate synapses in a brain system dedicated to action-selection [1]. When something better than expected happens, a phasic (fast) increase in dopamine release causes a long-lasting strengthening in the glutamatergic synapses, between neurons that encode the context in which the action was taken, and neurons that initiated the action [2,3]. It increases the likelihood that the same action will be selected in the same context. In addition, in the presence of the same stimuli which were present when that action happened, it causes a new phasic release of dopamine that acts to increase the activation of those neurons that start the learned action [4]. Therefore, failures in dopaminergic neurotransmission is implicated in movement disorders (e.g. Parkinson's disease [5]), psychiatric disorders (e.g. obsessive–compulsive disorder and schizophrenia [6]), and reward/motivation related disorders (e.g. drug addiction [4]).

Hence, to better understand the role of DA in these normal and abnormal functions, it is critical to have a method to quantify the synaptic DA release, which is reliable in terms of specificity and presents a good temporal resolution to capture an event that last less than a second (synaptic release and reuptake of dopamine). To data, the technique used to measure phasic DA release/reuptake *in vivo* is the fast-scan cyclic voltammetry (FSCV), a type of cyclic voltammetry that uses a high scan rate to improve selectivity [7]. The FSCV is an electrochemical method which consists of applying a fast triangular ramp of electrical potentials to a carbon-fiber electrode implanted in the brain. In the ascending phase of the ramp, dopamine oxidizes to a quinone that is reduced back to DA in the descending phase of the ramp. When dopamine oxidizes, it donates electrons to the electrode and the electrode donates these electrons to reduce the DA quinone back into DA. Each ramp lasts less than 10 ms and it is usually repeated 10 times per second. In addition to dopamine, several other molecules that are present in the brain tissue around the electrode are oxidized and reduced. However, their concentration does not vary as fast as the concentration of dopamine varies when it is released and reuptaken. Therefore, the currents generated by oxi-reduction of other molecules can be cleaned when the currents recorded at each cycle are subtracted from those recorded in 2 s cycle before. The shape of the oxi-reduction voltammogram (the potential vs. current plot) is characteristic of DA.

FSCV data are stored in a numerical matrix, which can be processed into images represented by the applied potential on the y-axis and the cycle (time) on the x-axis, and the current is represented by the pixel intensity [8]. Fig. 1 shows an image generated from FSCV data, in which there is DA release highlighted between vertical red lines.

* Corresponding author.
E-mail address: gustavomatsushita@ufpr.br (G.H.G. Matsushita).

https://doi.org/10.1016/j.compbiomed.2019.103466
Received 27 July 2019; Received in revised form 23 September 2019; Accepted 23 September 2019
Available online 25 September 2019

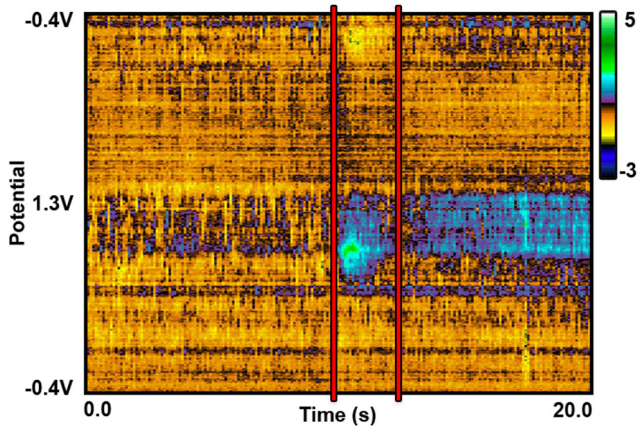

Fig. 1. Example of image generated from FSCV data using a standard color scheme [10], in which a dopamine release is highlighted between vertical red lines.

This color plot image is a visual representation using a standard false color palette used by FSCV analysis softwares [9]. This color scheme is nonlinear and designed to enhance transitions. It also avoids the use of colors that are not apparent to a red-green colorblind person [10].

Fast-scan cyclic voltammetry is more adequate than other techniques to measure phasic dopamine release. However, the high temporal resolution of this method generates a lot of data to be analyzed. Moreover, noise signals present in the recordings make the identification of DA releases a difficult task [9]. Even for a specialist, manual analysis of experiment results is a repetitive activity and requires a lot of time.

FSCV is not only used to analyze dopamine. In addition to dopamine, fast-scan cyclic voltammetry is also used to detect *in vivo* and *ex vivo* changes in extracellular concentration of other neuronstrasmitters, such as serotonin [11–13], noradrenaline [11,14,15] and acetylcholine [16]. Besides the neuroscience field, cyclic voltammetry is used in industrial research to study the behavior of doped materials with specific semiconductor and electrochemical properties [17], as well as in the synthesis and characterization of inorganic metal–polymer compounds [18]. Furthermore, Park et al. [19] examined zirconium electrochemical redox behaviors based on cyclic voltammetry results. Masek et al. [20] investigated, using cyclic and differential pulse voltammetry, the process and the kinetics of the electrochemical oxidation of morin in an anhydrous electrolyte. And Borman et al. [21] presented an algorithm for transient adenosine detection, but the method does not use visual image information. An analysis of current data is performed at specific adenosine oxidation voltages, and thus the peak moments at these points are verified.

Important studies have been carried out in the area of pattern recognition, which addresses the classification problems according to certain classes, or categories, predicted in a domain of a problem [22]. In pattern recognition the classification task can be understood as the assignment of a class to a feature vector, extracted from a sample to be classified. In supervised learning problems there is a predefined number of classes, in which the input and the desired output data are provided. An optimal scenario will allow the correct classification for unseen instances. However, the available attributes to characterize the samples do not always differentiate each class.

The classical approach to the development of pattern recognition systems foresees three well-defined steps: preprocessing, feature extraction and classification [22]. And these types of recognition systems have been successfully employed in the tasks of automatic class recognition in the most diverse application domains [23].

Despite the importance of dopamine, there is a gap in the literature on automatic identification of DA release. To the best of our knowledge, Matsushita et al. [9] introduced the first and only public dataset of FSCV images. However, it was limited to few images generated from 9 different experimental recordings, with a total of 295 phasic dopamine release. From these images, a classification system was created using extracted patches from regions containing DA release or not. The approach uses texture descriptors to perform feature extraction, and support vector machines (SVM) as a classifier. SVM is a widely used classifier which presents competitive results in various application domains [24]. The best f-measure obtained in Matsushita et al. [9] was 77.23%. The presented approach has a larger number of non-release regions, and consequently results in several false positives.

Therefore, in this paper the authors propose to fill a gap in the literature regarding the development of new classification systems using convolutional neural networks (CNN) and object detection techniques. The proposed methodology uses combined approaches of entire images and extracted patches, avoiding the previous problem of unbalanced sample numbers between different classes. The best obtained result was an accuracy of 98.31%. We also introduce a new and larger public dataset with 1005 evoked dopamine release images and 1005 images without dopamine release, which is available for research purposes upon request.

2. Dataset

Fast-scan cyclic voltammograms color plot images were obtained from the Laboratory of Central Nervous System of the Federal University of Parana (UFPR) at Curitiba, Brazil and from D. Robinson's Laboratory of the University of North Carolina (UNC) at Chapel Hill, United States of America. The two laboratories used different animals and FSCV setups. The UFPR laboratory used 29 male Swiss mice and the UNC laboratory used 6 male Sprague Dawley rats.

Each animal was anesthetized with 1.5–1.8 mg/kg urethane (i.p.) and mounted in a stereotaxic frame. A scalpel was used to make a midline incision exposing the skull bone surface, and a stainless steel burr was used to drill two circular opening above the nucleus accumbens (NAc) and ventral tegmental area (VTA), respectively. The openings were centered in the following stereotaxic coordinates: (i) Mice: NAc, AP +1.2 mm, ML +1.2 mm; VTA, AP -3.8 mm; ML +0.2 mm. (ii) Rats: NAc, AP +1.2, ML −1.2; VTA, AP −3.8, ML −0.2. Another hole was opened above the contralateral frontal cortex to insert an Ag/AgCl-reference electrode just below the dura mater. A recording carbon fiber electrode was inserted in the NAc and an stainless steel electrode was inserted into the VTA.

Dopamine release was evoked by electrical stimulation of the ventral tegmental area (20 pulses, 0.5 ms per pulse). FSCV measurements at UFPR were taken with a Wireless Instantaneous Neurotransmitter Concentration Sensor system (WINCS, Mayo Clinic, Rochester, MN, USA) and processed using WINCSware with MINCS software (version 2.10.4, Mayo Clinic, Rochester, MN, USA). Every 100 ms, a triangular wave form potential of −0.4 V to +1.0 V to −0.4 V was applied at a rate of 300 V/s to the carbon-fiber recording electrode versus the Ag/AgCl-reference electrode. Oxidative and reductive currents were continuously sampled at 100,000 samples/s and 944 samples/scan. FSCV measurements at UNC were taken with a customized setup controlled by a computer using Lab VIEW instrumentation software (National Instruments, Austin, TX, USA) and the potential applied to the carbon-fiber was of −0.4 V to +1.3 V at a rate of 400 V/s.

The images were generated from 30 different experimental recordings with a total of 1005 electrically evoked dopamine release. Each recording has dopamine release evoked with different magnitudes of electrical stimulation, resulting in different patterns of form and intensity. All experiments were performed in accordance with the NIH Guide for the Care and Use of Laboratory Animals with procedures approved by the Institutional Animal Care and Use Committee of the University of North Carolina, and the Institutional Ethics Committee for Animal Experimentation of the Federal University of Parana (Protocol 638).

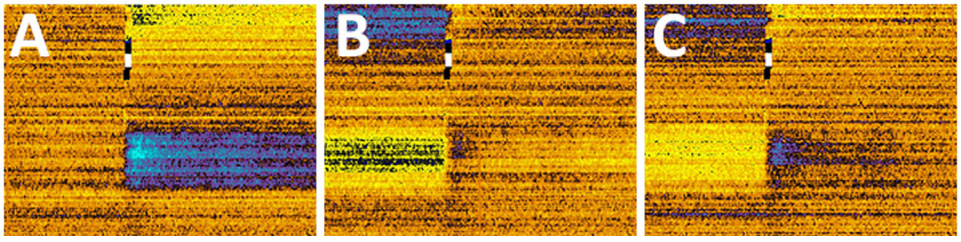

Fig. 2. Images generated using different background positions. In A, the background position was selected from the beginning of the image; In B, the background position was selected from the middle of the image; In C, the background position was selected from the end of the image.

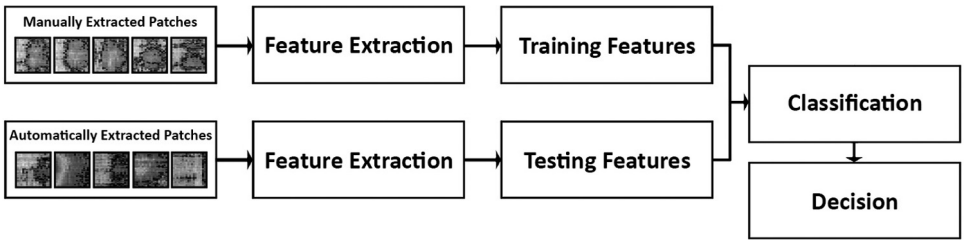

Fig. 3. An overview of the image patches approach [9], in which manually extracted patches features are used as training set for the classification and the automatically extracted ones are used as testing set.

This new dataset version is composed of images generated from all these experimental recordings, which include those in Matsushita et al. [9] Dataset. Unlike the first version, the new one has not only color plots of phasic dopamine release episodes, but also longer recordings that include periods with no DA release. In total there are 2010 images, 1005 of each of these classes, with resolution of 875 × 656 pixels representing a 20 s recording. During the generation of these FSCV images, a background subtraction is commonly used before applying a fake color palette. Normally for each image, one column or more columns are selected to subtract the values from the others. In the case of the first version of the dataset [9], this process was done manually during its generation. In this new one, since we must consider that we do not know where the dopamine release is present, the process is done automatically choosing 3 different background positions: the Background A was selected from a column at the beginning of each image (0.5 s), the Background B from the middle of each image (10 s), and the Background C from the end of each image (19.5 s).

These images with different background end up generating different results, as it is possible to be observed in Fig. 2. Thus, it is possible to explore different approaches of training and testing, since for each DA release 3 images were generated. Each image containing DA was manually labeled with the approximated information of each release interval and peak. As in Matsushita et al. [9], all images were randomly divided into 3 folds with same amount of samples from each of the two classes: (1) phasic DA release images and (2) non-release images. A new enhanced division was also performed, in which the images were grouped into 10 folds with class balance, and images with DA release from the same experiment were placed in the same fold. This ensures that all dopamine releases from the same animal are not present in both training set and testing set.

It was possible to notice the common dopamine release region [9] between the pixel 320 and 520 of the y-axis. The common area was used to delimit the suitable region to extract patches. Patches with size of 200 × 200 pixels were manually extracted using the labeled information. For each DA release patch created, one from a non-release region was also extracted creating a balanced training dataset. Following the testing dataset approach presented in Matsushita et al. [9], patches were extracted automatically by applying a sliding window over the original images and moving horizontally every 135 pixel inside the common area, in which the patches were separated into two classes using the approximated release interval information.

3. Proposed approach

Due to the limitations of the first public dataset, Matsushita et al. [9] presented an approach wherein patches were extracted from regions where there was dopamine release (Class 1) and from regions with no release (Class 2). Fig. 3 illustrates this previous approach, which used only image patches. Briefly, hand-crafted features were extracted from labeled patches applying texture descriptors, then manually extracted patches features were used as training set for the classification, while the automatically extracted ones were used as testing set. The new dataset presented is not only larger than the existing one, but also provided the exploration of new approaches. Our proposed approach follows the methodology of representation and classification presented in Matsushita et al. [9], however it also uses original samples.

Fig. 4 shows an overview of this approach, which contains two main steps: the original images approach and the image patches approach. In the first main step, features are extracted from the original images and then classified. In addition to the original samples, two zoning variations were tested (Fig. 5): the first one uses only the common region of phasic dopamine release, and the second one also adds a region of the top of the original sample (between the pixel 0 and 90 of the y-axis), which in some releases has visual information that could be important for extracting features.

In the second main step, once it was decided if an input contains a DA release or not, patches were extracted from those images classified as DA Release (Class 1). As it is assumed that each image contains only one release of dopamine, the classification of these patches allows a more precise decision of its location. If all patches from a single image are classified as Class 2 (non-release), the image from which they were extracted and reclassified to the other class. Otherwise, it means that at least one patch has been identified as containing the DA release and, if correct, the image itself was considered as a hit. With this metric it was possible to calculate hits and errors per image, keeping the dataset analysis more balanced and providing results of possible dopamine release images as well as more precise regions within them.

3.1. Feature extraction

The representation or feature extraction stage is quite important in the development of pattern recognition systems. Although the main objective of this work is the use of convolutional neural networks, the

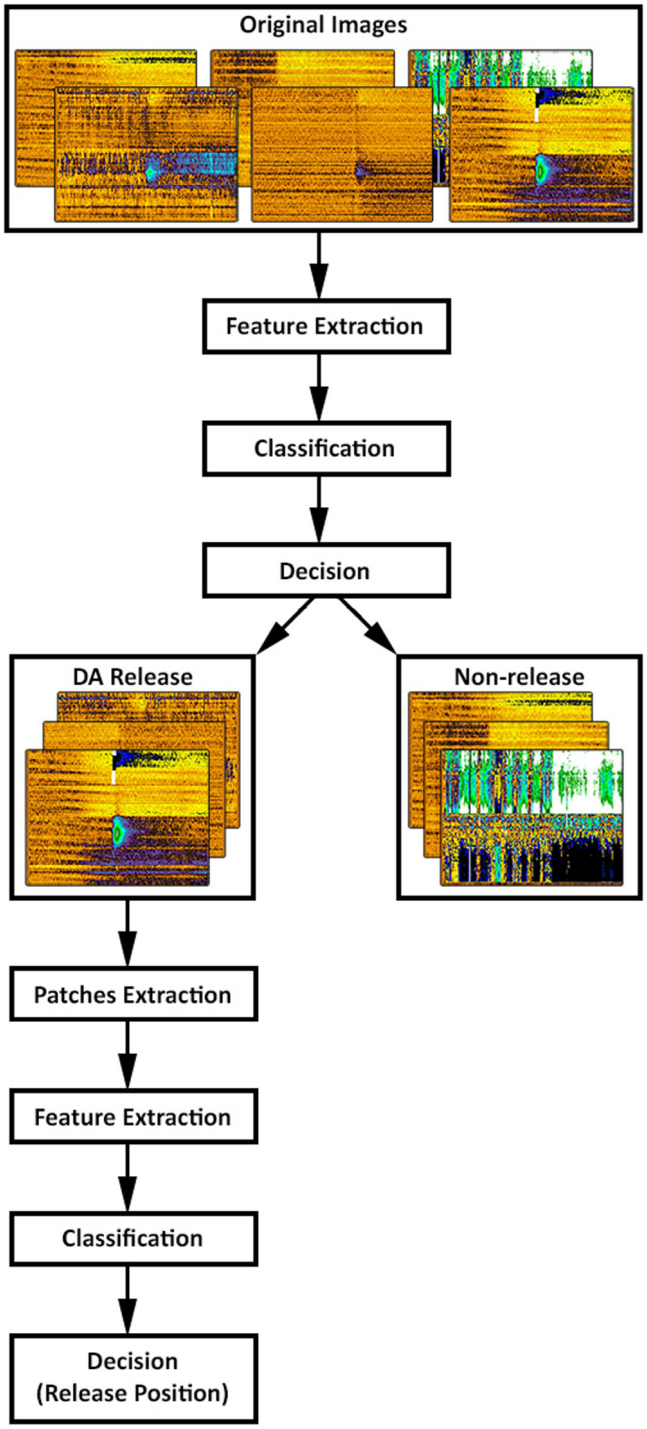

Fig. 4. An overview of the combined approach, which the first main step performs the classification of the original images, and from this decision the extraction and classification of the patches are carried out.

initial tests were performed using hand-crafted texture features, which has been successfully used in several application domains: medical diagnosis [25], face recognition [26], musical genre recognition [27], handwriting identification [28] and classification of bird species [24]. Texture corresponds to a visual pattern which is usually related to the pixel distribution in a region and properties of the image object such as color, brightness, and size. Thus, this attribute contains significant information about the content of the image.

The descriptors chosen to be used were the Local Binary Pattern (LBP) [29] and the Local Phase Quantization (LPQ) [30], which were the same used in Matsushita et al. [9], allowing also the comparison of the results obtained in the two datasets. The texture feature extraction was performed using grayscale input images, that there is no lost information and it does not compromise the performance of the classifier. LBP operates on an image pixel and its adjacent ones to find a histogram of local binary patterns. To be able to operate textures of different scales, it can create patterns considering different pre-defined quantities of neighbors for its operation. Such variations are identified by $LBP_{P,R}$ in which P is the number of neighboring pixels existing in a region of radius R around the central pixel. A $LBP_{8,2}$ was applied using the so-called uniform patterns, which results in a feature vector with 59 values. A LBP variant was also tested, and its known as Robust Local Binary Pattern (RLBP) [31]. LPQ is based on the blur invariance property and uses the local phase information extracted using the 2D Discrete Fourier Transform (DFT) calculated in a rectangular neighborhood, which is a local window for each pixel of the image. The best LPQ results were obtained using a window size of 3, which generates a vector with 256 values. For some experiments, different texture operators were combined using Early Fusion technique, that is the concatenation of the LPQ vector with a LBP/RLBP one, generating a larger single vector.

3.2. Classification

The classification with hand-crafted features was performed using Support Vector Machine (SVM). This classifier introduced by Vapnik [32] has been widely used with success. As well as the chosen texture descriptors, SVM has presented competitive results compared to other classifier methods in the most diverse applications [24,27,28,33,34], besides having been previously used for dopamine release identification [9].

Support vector machine is a set of supervised learning techniques able to analyze data, recognize patterns and classify them. The standard SVM is defined as a non-probabilistic binary linear classifier, that inputs a set of data, and is able to predict for each input which of the possible classes it is part of. Initially, with a training algorithm and set of examples already defined to which category each belongs, the SVM constructs a model that assigns the new examples to one category or another. The SVM experiments were carried out using a RBF kernel, and gamma and cost parameters were optimized by using grid-search to attain better results. A cross-validation was used, in which when one fold was used as testing set, the other two were used as training.

For each sample, SVM outputs present an estimate of probability for each class in the classification system, thus it is possible to perform a combination of classifiers. In Kittler et al. [35], some merging rules to combine the predictions of different classifiers were proposed. The best results were obtained when the Sum Rule was used: it sums all the predictions values of each class in all classifiers, and chooses the class with the highest final value.

3.3. Convolutional neural networks

The performance of a classifier system is heavily dependent on the choice of data representation or features used. The inability to extract and organize discriminative information from the data impacts the results that can be obtained [36]. Automatic learning representation can make it easier to extract important information to build a classification system. Among the ways of learning representations, there are the deep learning methods, like convolutional neural networks (CNN). Convolutional neural networks have been applied successfully in different problems such as breast cancer classification [37], diagnosis of seizures [38], forest species identification [39], and it was used by replacing the two previous steps: feature extraction and classification.

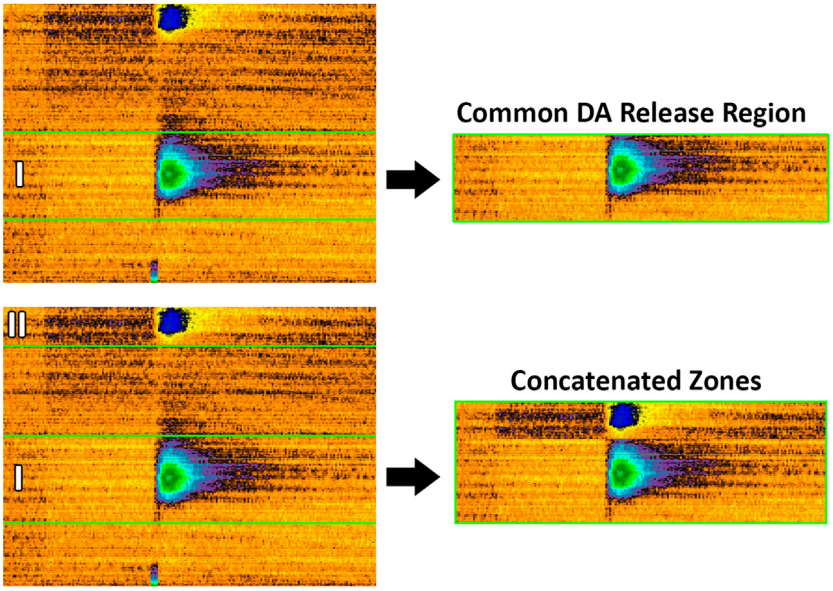

Fig. 5. Examples of the original image and two zoning variations: A common DA release region (I); and a concatenated zones (I and II).

CNN is a variation of multi-layer perceptrons network and consists of layers with different functions. Initially, it is common to apply the data to input layers known as convolutional layers [40]. These layers are composed of neurons, and each neuron is responsible for applying a trainable filter to a specific image area. Basically a neuron is being connected to a set of pixels of the previous layer and for each connection a weight is applied. The respective weights of its connections produce an output passed to the next layer. The weights assigned to the connections of a neuron can be interpreted as a matrix representing the filter of a convolution of images in the spatial domain (kernel). The weights are shared across neurons from a same layer, leading the filters to learn patterns which occur in any part of the image.

In the CNN convolutional layers, it is not necessary to specify which filters or features to be used. It defines only the architecture of the filters: sizes, stride and quantity per layer. The learning process of the network changes the weights throughout the training, searching automatically for the best values for the input dataset. A very important layer commonly used after the convolutions is the pooling layer. The function of this layer is to reduce the dimensionality of the data in the network. This reduction is important for training faster, but also to create spatial invariance. The pooling layer only reduces the height and width of a map. When it is desired to perform a classification, it is appended after the set of the convolutional and pooling layers at least one fully-connected layer. This fully-connected layer is responsible for tracing a decision path to each class based on results of the filters from previous layers. After the fully-connected layers the last step is the classification function.

The use of neural networks requires lot of training samples for good performance. Although this second version dataset has already a significantly larger number of images than the first one, a data augmentation technique has been applied for some tests. The technique consisted of adding up to three random variation in training images including: rotation, translation, brightness, blur, saturation, and sharpening. Fig. 6 shows examples after the data augmentation, in which an original training sample generates 14 new samples increasing the training set.

The main convolutional neural network model used in this work was proposed by Roecker et al. [41]. It was designed with principles to simplify the model and use low-resolution images, useful to development of systems with limited resources. Table 1 describes the model architecture. It receives as input an RGB image, in which the input passed through a stack of convolutional layers with variable number of filters 3×3. The convolution output was set up with a leaky rectifier activation function (LReLU). Then, a spacial pooling performed a maximum-value subsampling with a 2×2 window and a stride of 2. Fully-connected layers have structure similar to multilayer perceptron (MLP) receiving the previously stages results as input [41]. The only existing difference of the model used in this work for the one presented by [41], is the last layer which does the classification and has 2 units, since in this case we have a binary problem, which did not happen in that work. In this output, a softmax (normalized exponential function) was applied to result into probabilities. The best results were using a learning rate of 0.0001 and the batch size was 50. These tests were run for 100 epochs. In addition to the default dropout parameters in this model, we took care to make sure no overfitting occurs during training. For all CNN models tested, not only the accuracy/error rate of the testing set was analyzed, but also in the training set throughout the epochs. Thus, the values for the epoch/iteration parameters were chosen before any overfitting situation occurred.

Table 1
Roecker et al. [41] CNN architecture model.

#	Layer	Parameters	Stride (x, y)
1	Convolutional	$3 \times 3 \times 32$	(1, 1)
2	Convolutional	$3 \times 3 \times 32$	(1, 1)
3	Pooling	2×2	(2, 2)
4	Convolutional	$3 \times 3 \times 64$	(1, 1)
5	Convolutional	$3 \times 3 \times 64$	(1, 1)
6	Pooling	2×2	(2, 2)
7	Fully-connected	512	–
8	Fully-connected	512	–
9	Fully-connected	2	–
10	Softmax	2	–

Some experiments were also carried out using the Inception v3 network, which is a deep convolutional neural network architecture with a design that allows increasing the depth and also the width of the network, while keep the computational cost constant [42]. Images from the same class may have huge variations in the location and size of the information. Choosing the best kernel size for the convolution is not so easy. So the Inception uses multiple size filters operating on the same level.

This network model consisted of inception modules stacked upon each other, with occasional max-pooling layers to reduce the resolution

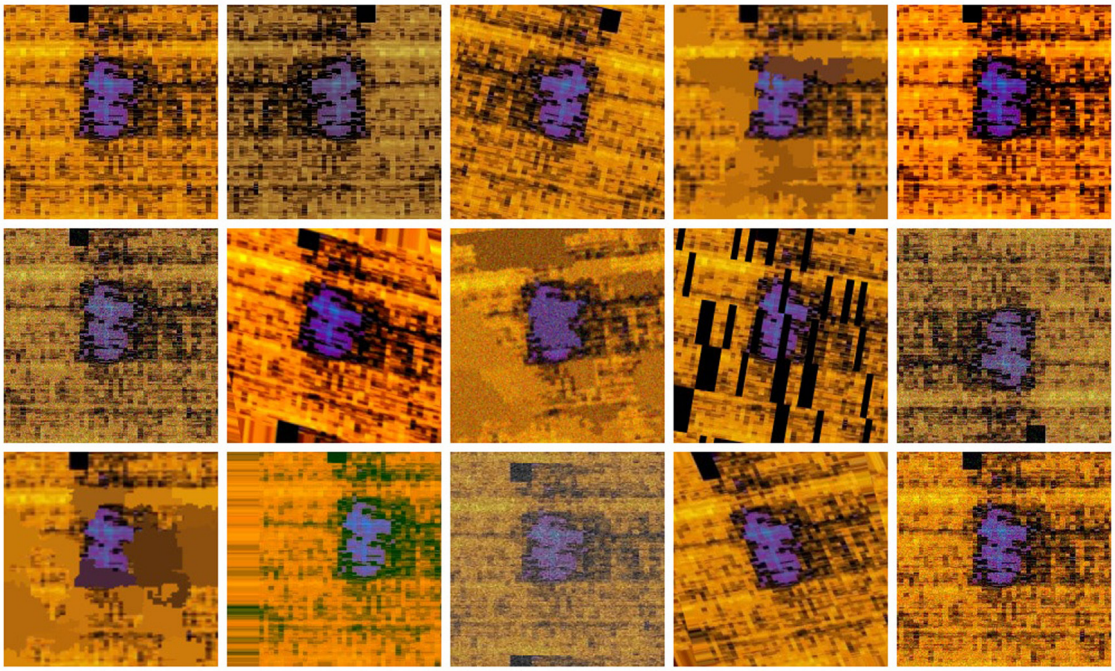

Fig. 6. Examples of Data Augmentation, in which a patch image generates 14 new samples, adding up to three random variations such as rotation, translation, brightness, blur, saturation, sharpening.

of the grid. The use of this dimension reduced inception module, a neural network known as GoogLeNet (Inception v1) was built. Through Inception v2 and v3, several improvements were applied in the network model [43], such as smart factorization methods. A filter $n \times n$ can be factored into a combination of $1 \times n$ and $n \times 1$ convolutions to improve computational speed.

For our experiments, RGB input images were resized to $299 \times 299 \times 3$ which is the default value of the Inception v3 model used. For these tests, the transfer learning technique was applied, using the pre-trained parameters on ImageNet dataset [44]. The only altered layer of the original network was the end of the fully-connected layers, so that it had the output for our two classes. In transfer learning, we reused these transferred weights for the feature extraction layers, which is the most complex part of the model. The parameters learned are transferred to the target task, except for the last layer. The classification part (fully-connected layers) was re-trained using our dataset as input. The best results were obtained using a learning rate of 0.01 and the batch size was 100. These tests were run for 15,000 iterations.

3.3.1. Object detection

Finally, some experiments were performed using a YOLOv3 end-to-end system. This approach is based on object detection, which has also been used successfully in different applications [33,45,46]. YOLO is an accurate and fast approach for object detection [47], in which a single neural network evaluation predicts bounding boxes and class probabilities directly from full images. You Only Look Once (YOLO) at a sample to predict the classes and their locations [48].

The YOLO model initially divides the input sample into an $S \times S$ grid. Each grid cell predicts a fixed number bounding boxes. If the center of an object falls into the cell, it is responsible for predicting that object. Each cell predicts B bounding boxes and each box has a confidence score. This score shows how confident the YOLO is that there is an object in the box and how accurate it is [48]. The confidence score should be zero if there is no object in that grid cell. Each bounding box also consists of: (x, y, w, h). The box width w and height h are normalized, x and y are offsets to the corresponding cell. x, y, w, h are between 0 and 1. Each cell also predicts C conditional class probabilities. And it is the probability that the detected object belongs to a specific class. All of these predictions are encoded as an (S, S, B×5+C) tensor.

YOLOv1 predicted the bounding boxes using fully-connected layers, which were removed since its second version and know uses anchor boxes. So instead of directly predicting a bounding box, YOLOv2 and v3 predict offsets from a predetermined set of predetermined boxes. The YOLOv3 network for performing feature extraction uses successive 3×3 and 1 convolutional layers [49]. The model is significantly larger than other versions (v1, v2, fast and tiny) with 53 convolutional layers.

In this way, it was possible to use original images as input and provide an accurate result of the location of the dopamine release in the sample, without the necessity of the two different CNN steps with patches extraction. YOLO identifies the Class 1 (phasic dopamine release) in the input samples. And, unlike all other approaches, there will be only one class to be identified. When it is not detected, then it is assumed that there is no release of dopamine (Class 2).

The DA release labels were converted to the YOLO pattern of bounding boxes. Some training sessions were carried out with only files of the DA images, and others with all the images. The evaluation metrics are similar to the previous one. If a dopamine release is identified in the correct location, compared with the original labels, it is already considered a hit. If there is any identification in an image of Class 2, it will be considered a miss classification.

Two models were used: Yolo and Tiny Yolo with weights pre-trained on ImageNet, both in version 3. From its default settings, only the input resolution size and the output classes has been changed. Different tests were performed, some using the default input size of 416×416 and others using its variable of random resize. Since the version 2 model, the convolutional and pooling layers can be resized on the fly, it means that instead of using only one input size, for each 10 iterations, the network size will be randomly resized (input and output) to size between 320×320 and 608×608. The best results were obtained using a learning rate of 0.001 and the batch size was 64 and 8 subdivision. These tests were run for 25,000 iterations.

4. Results

In Matsushita et al. [9], the experiments were accomplished using their first version of DA release dataset, and they were performed

Table 2
Best results obtained in Matsushita et al. [9] using the automatically extracted patches approach.

Patch size	Descriptor	F-Measure(%)
120 × 120	**LPQ_9**	**77.23 ± 1.20**
120 × 120	$LBP_{8,2}$	62.32 ± 0.43
120 × 120	$RLBP_{8,2}$	69.85 ± 0.40

only following the image patches approach with texture descriptors. The training set used was composed of 120 × 120 manually extracted patches from the original samples and 120 × 120 automatically extracted patches as testing set. An automatic extraction generates different numbers of samples between the two classes, in which there are more patches without the DA release, and consequently, f-measures were used to evaluate these specific tests.

Table 2 summarizes the best [9] results obtained using features extracted with LPQ (window size 9), $LBP_{8,2}$ and $RLBP_{8,2}$ descriptors, and SVM as classifier. As each type of test was performed three times, one for each fold, the mean results are presented with the respective standard deviation. The LPQ performance was superior to the other descriptors, and the best obtained f-measure was 77.23% ± 1.20. The authors also tried to use combinations of feature vectors (early fusion) or predictions (late fusion), but they did not improve the final result.

The second and more complete version of the dataset, introduced here, allows us to better explore new approaches. As in the previous experiments [9], the initial tests were made exploring texture descriptors and their best parameters. However, the samples used to compose the training and the testing set were only the original images, without patches extraction.

4.1. Texture descriptors

There are three different training sets, each generated with different backgrounds as presented. Table 3 shows the results obtained by training and testing full images with a specific background and different texture descriptors: $LBP_{8,2}$, $RLBP_{8,2}$ and LPQ (window size 3), and SVM as classifier. All of these initial tests were performed using 3 folds division. However, for each dataset, tests were also performed with their two zoning variations: the first containing only the common DA release region and the second containing two regions of the image concatenated. It was possible to notice that of all the varied images, the best results were always using the background positions A and C. And the best result in these preliminary tests was using the common DA release region, the background position C dataset and LPQ as descriptor, with an accuracy of 75.72% ± 0.60.

In addition to the experiments performed separately, tests were also done combining the training sets of different background positions. Thus, the respective folds of positions A, B and C were grouped, keeping the same original images together. But the best result was 73.88% ± 1.17 (testing set using background C, descriptor LPQ and SVM classifier) and it was not better than the previously one (75.72%) besides having a higher cost for the training.

In addition to the support vector machines some other classifiers were tested. However, they did not obtain better results. Random Forests and Gradient Boosting classifiers obtained accuracy of 72.28% ± 1.24 and 72.19% ± 0.25 respectively, both using the common DA release region, background position C dataset and LPQ as descriptor.

Using the best results with background positions A and C, late fusion tests were also performed using the sum rule. The combination of SVM score obtained through LPQ features resulted in 79.15% ± 0.67 accuracy, and it was the best result obtained with texture descriptors and these samples. To finalize the investigations with texture descriptors, the same tests using automatically extracted patches approach [9] were performed. The best automatically extracted patches f-measure was of 74.78% ± 0.93, slightly inferior to the 77.23% in the literature. And as in Matsushita et al. [9], early and late fusions did not improve results.

Table 3
Results obtained with texture descriptors.

Background position	Descriptor	Accuracy(%)
Background A	**$LBP_{8,2}$**	**73.77 ± 0.31**
Background A	$RLBP_{8,2}$	73.28 ± 0.64
Background A	LPQ_3	73.44 ± 0.78
Background B	$LBP_{8,2}$	69.20 ± 0.58
Background B	$RLBP_{8,2}$	69.10 ± 1.89
Background B	**LPQ_3**	**72.04 ± 2.72**
Background C	$LBP_{8,2}$	70.15 ± 1.52
Background C	$RLBP_{8,2}$	70.70 ± 1.85
Background C	**LPQ_3**	**73.93 ± 0.31**
Background C	LPQ_5	73.13 ± 1.29

4.2. Convolutional neural networks

Our new dataset brought greater complexity, and many images without dopamine release. In these images, there are the most diverse variations and noises. Despite the greater challenge, it was possible to explore new approaches such as the use of Convolutional Neural Networks. The first 3 folds experiments were made using the Inception v3 and the model presented in Roecker et al. [41]. As shown in the previous section, the Inception tests were performed using RGB input images (resized to 299 × 299 × 3) and pre-trained weights on ImageNet. These experiments were run for 15,000 iterations, using learning rate of 0.01 and the batch size of 100. The best accuracy was 95.72% ± 0.51 using data augmentation in the original images training set, which was already much better than all results obtained with texture descriptors.

Unlike the Inception tests, no pre-trained parameters were used for the Roecker et al. [41] model. Each of the training sets were performed for 100 epochs and initially only images generated using the background position A were used. Initial tests without data augmentation served to test different input sizes. From them, only the best parameters were used for the tests with the other training sets and using data augmentation (Table 4). An accuracy of 97.31% ± 0.64 was the best result obtained, using a training set of zoned images in the common DA release region and generated by background B. In spite of the excellent result obtained, when we analyzed the predictions, it was possible to observe that both false positives and false negatives had very high values. Perhaps for this reason, no predictions fusion tests showed improvements, no longer being an interesting approach to use.

Tests using automatically extracted patches approach and this model with data augmentation also got competitive results. The best result had an f-measure of 95.64% ± 1.04 using the background C and patch size 200 × 200. However, for 1005 DA release, there are more than 9000 patches without release, generating hundreds of false positives.

The alternative to this problem was to extract the patches directly from the images classified as positive of the best experiment from Table 4, and then classify these patches. Following the methodology presented in the previous chapter, this approach allowed even a revision of the false positives: Considering the 41 FP occurrences there were only 17 now, and the final accuracy of the complete process was 98.36%±2.39. This classifier was very versatile in identifying phasic dopamine release in different situations, even in small amounts or in the middle of noises.

All experiments have been performed using a random 3-fold division with class balance proposed by Matsushita et al. [9]. However, some DA release images from the same animal may be present in both the testing set and the training set. Thus, experiments were also performed with an enhanced 10-fold division, in which all dopamine release samples from the same experimental recording were placed in the same fold. This not only increases the complexity of the problem, but also the computational cost of performing 10 training models with more images each. Thus, the best parameters previously obtained were replicated for this new dataset organization. The 10-fold complete approach using the Roecker et al. [41] model, training set of zoned images in the common DA release region and generated by background B, resulted in a final accuracy of 97.35%±5.84.

Table 4
Results obtained with Data Augmentation using the Roecker et al. [41] CNN model.

Background position	Zoning type	Accuracy(%)
Background A	Original Samples (175 × 131)	96.47 ± 0.61
Background A	DA Release Region (175 × 40)	96.47 ± 0.67
Background A	Concatenated Zones (175 × 58)	96.22 ± 1.42
Background B	Original Samples (175 × 131)	96.72 ± 0.12
Background B	**DA Release Region** (175 × 40)	**97.31 ± 0.64**
Background B	Concatenated Zones (175 × 58)	96.62 ± 0.55
Background C	Original Samples (175 × 131)	94.78 ± 1.06
Background C	DA Release Region (175 × 40)	96.22 ± 0.14
Background C	Concatenated Zones (175 × 58)	95.77 ± 0.78

4.2.1. YOLO

Despite excellent results already obtained, the previous approach requires two different steps of training and classification with neural networks. Using YOLO, it was possible to directly identify the original image without the need to extract patches. The best 3-fold accuracy obtained using the Tiny model and RGB inputs with size of 416 × 416 was 96.82% ± 1.13. The training set was composed only of phasic dopamine release samples, and background position A. These same parameters were replicated to the complete YOLOv3 model, which is larger and has a higher computational cost. However, the accuracy did not improve, being 96.57% ± 0.68. The last two tests performed were accomplished again using the best parameters and using the random resize. New anchors were calculated using kmeans in the new dataset to result in the same default number of the model: 6. The result was 97.56% ± 0.49, but the best accuracy was 97.66%±0.67 using the default anchors. This best model was also used for a 10-fold test and resulted in an accuracy of 96.62%±5.94.

YOLO only identifies the phasic dopamine release class, but it is possible to analyze true and false positives predictions. As in the results of the other model, false positives also have a high value. These high values present in CNN also made it impossible to apply a good threshold for rejection, which may improve the final decision. Like the other classifier, YOLO also makes correct DA identification in the most varied situations.

5. Conclusion

In this paper, we aimed to fill a gap in the literature by mitigating new approaches and parameters to identify phasic dopamine release in fast-scan cyclic voltammetry images. The methodology proposed using neural networks presented results with higher accuracy than those previously reported. But for that, it was necessary to create a larger dataset and it is now publicly available to the scientific community.

The new presented dataset not only allowed to explore/investigate new approaches but also a way to circumvent the problem of the unbalanced classes. The tests in this dataset, which contains greater diversity in phasic dopamine releases and also complete images without any release, resulted in similar values using texture descriptors and automatically extracted patches. The f-measure with automatically extracted patches was 74.78% using LBP. And the accuracy with entire images approach using Late Fusion of different background positions predictions was 79.15%. But by far the best results were achieved using Convolutional Neural Networks.

The model presented in Roecker et al. [41] was used to perform 3-fold classification of entire images with an accuracy of up to 97.31%. By itself, this result is already very good by classifying DA release within 20 s of an experimental recording. But using a combined approach, extracting patches from images classified as Class 1, it was also possible to provide a more accurate result within each of the images with an accuracy of 98.31%. The best accuracy obtained for the enhanced 10-fold dataset was 97.35%.

The combined approach provides an excellent result, but YOLO tests resulted in an accuracy of 97.66% using 3-fold random division, and 96.62% using 10-fold division by animals. Even though this value is somewhat lower, the YOLO allows the identification of phasic DA release directly from the original samples, without the need for two different training and classification steps. This method also results in identifications in varied positions and sizes, being more versatile and accurate than patches, which are always restricted to a defined size. Thus, in a problem of identifying a substance that varies many patterns, YOLO also becomes an excellent choice for an automatic classification. Although there were visual variations in the images, there was no definitive background position that was always better than the others. It was possible to achieve great results in all of them.

Dataset availability

The dataset generated during the current study are available for research purposes at: https://web.inf.ufpr.br/vri/databases/phasicdopaminerelease/.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Gustavo H.G. Matsushita: Investigation, Methodology, Writing - original draft. **Adam H. Sugi:** Data curation, Validation, Writing - original draft. **Yandre M.G. Costa:** Supervision, Validation, Writing - review & editing. **Alexander Gomez-A:** Data curation, Writing - review & editing. **Claudio Da Cunha:** Conceptualization, Validation, Writing - review & editing. **Luiz S. Oliveira:** Conceptualization, Project administration, Writing - original draft.

Acknowledgments

The authors would like to thank the Pró-Reitoria de Pesquisa e Pós-Graduação (PRPPG), Brazil at Federal University of Parana (UFPR), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 - and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil, which financed in part this study; The D. Robinson's Laboratory of the University of North Carolina (UNC) for providing part of the FSCV data.

References

[1] C.R. Gerfen, D.J. Surmeier, Modulation of striatal projection systems by dopamine, Annu. Rev. Neurosci. 34 (2011) 441–466.
[2] C. Da Cunha, A. Gómez-A, C.D. Blaha, The role of the basal ganglia in motivated behaviour, Rev. Neurosci. 28 (5–6) (2012) 747–767.
[3] W. Schultz, Behavioral dopamine signals, Trends Neurosci. 30 (5) (2007) 203–210.
[4] N.D. Volkow, R.A. Wise, R. Baler, The dopamine motive system: implications for drug and food addiction, Nat. Rev. Neurosci. 18 (12) (2017) 741.
[5] C. Da Cunha, S.L. Boschen, A. Gómez-A, E.K. Ross, W.S. Gibson, H.K. Min, K.H. Lee, C.D. Blaha, Toward sophisticated basal ganglia neuromodulation: Review on basal ganglia deep brain stimulation, Neurosci. Biobehav. Rev. 58 (2015) 186–210.
[6] R.A. McCutcheon, A. Abi-Dargham, O.D. Howes, Schizophrenia, dopamine and the striatum: From biology to symptoms, Trends Neurosci. (2019).
[7] J.G. Roberts, L.A. Sombers, Fast-scan cyclic voltammetry: chemical sensing in the brain and beyond, Anal. Chem. 90 (1) (2017) 490–504.
[8] D.L. Robinson, B.J. Venton, M.L. Heien, R.M. Wightman, Detecting subsecond dopamine release with fast-scan cyclic voltammetry in vivo, Clin. Chem. 49 (10) (2003) 1763–1773.
[9] G.H.G. Matsushita, L.E.S. Oliveira, A.H. Sugi, C. da Cunha, Y.M.G. Costa, Automatic identification of phasic dopamine release, in: 2018 25th International Conference on Systems, Signals and Image Processing, IWSSIP, IEEE, 2018, pp. 1–5.

[10] D. Michael, E.R. Travis, R.M. Wightman, Peer reviewed: color images for fast-scan CV measurements in biological systems, Anal. Chem. 70 (17) (1998) 586A–592A.

[11] U. Kumar, J.-S. Medel-Matus, H.M. Redwine, D. Shin, J.G. Hensler, R. Sankar, A. Mazarati, Effects of selective serotonin and norepinephrine reuptake inhibitors on depressive- and impulsive-like behaviors and on monoamine transmission in experimental temporal lobe epilepsy, Epilepsia 57 (3) (2016) 506–515.

[12] R.A. Saylor, M. Hersey, A. West, A.M. Buchanan, H.F. Nijhout, M.C. Reed, J. Best, P. Hashemi, In vivo serotonin dynamics in male and female mice: Determining effects of acute escitalopram using fast scan cyclic voltammetry, Front. Neurosci. 13 (2019) 362.

[13] K.M. Wood, P. Hashemi, Fast-scan cyclic voltammetry analysis of dynamic serotonin reponses to acute escitalopram, ACS Chem. Neurosci. 4 (5) (2013) 715–720.

[14] J. Park, P. Takmakov, R. Wightman, In vivo comparison of norepinephrine and dopamine release in rat brain by simultaneous measurements with fast-scan cyclic voltammetry, J. Neurochem. 119 (5) (2011) 932–944.

[15] E.N. Nicolai, J.K. Trevathan, E.K. Ross, J.L. Lujan, C.D. Blaha, K.E. Bennet, K.H. Lee, K.A. Ludwig, Detection of norepinephrine in whole blood via fast scan cyclic voltammetry, in: 2017 IEEE International Symposium on Medical Measurements and Applications, MeMeA, IEEE, 2017, pp. 111–116.

[16] R. Asri, B. O'Neill, J. Patel, K. Siletti, M. Rice, Detection of evoked acetylcholine release in mouse brain slices, Analyst 141 (23) (2016) 6416–6421.

[17] W. Brütting, Introduction to the physics of organic semiconductors, Phys. Org. Semiconduct. (2005) 1–14.

[18] Y.N. Zeng, N. Zheng, P.G. Osborne, Y.Z. Li, W.B. Chang, M.J. Wen, Cyclic voltammetry characterization of metal complex imprinted polymer, J. Mol. Recognit. 15 (4) (2002) 204—208, http://dx.doi.org/10.1002/jmr.578.

[19] J. Park, S. Choi, S. Sohn, K.-R. Kim, I.S. Hwang, Cyclic voltammetry on zirconium redox reactions in LiCl-KCl-ZrCl4 at 500 °C for electrorefining contaminated zircaloy-4 cladding, J. Electrochem. Soc. 161 (3) (2014) H97–H104.

[20] A. Masek, E. Chrzescijanska, M. Zaborski, Electrooxidation of morin hydrate at a Pt electrode studied by cyclic voltammetry, Food Chem. 148 (2014) 18–23.

[21] R.P. Borman, Y. Wang, M.D. Nguyen, M. Ganesana, S.T. Lee, B.J. Venton, Automated algorithm for detection of transient adenosine release, ACS Chem. Neurosci. 8 (2) (2017) 386–393.

[22] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, John Wiley & Sons, 2012.

[23] J.G. Martins, Y.M.G. Costa, D. Bertolini, L.S. Oliveira, Uso de descritores de textura extraídos de GLCM para o reconhecimento de padrões em diferentes domínios de aplicação, in: XXXVII Conferencia Latinoamericana de Informática, 2011, pp. 637–652.

[24] R.H.D. Zottesso, G.H.G. Matsushita, D.R. Lucio, Y.M.G. Costa, Automatic segmentation of audio signal in bird species identification, in: Computer Science Society (SCCC), 2016 35th International Conference of the Chilean, IEEE, 2016, pp. 1–11.

[25] A. Khademi, S. Krishnan, Medical image texture analysis: A case study with small bowel, retinal and mammogram images, in: Electrical and Computer Engineering, 2008. CCECE 2008. Canadian Conference on, IEEE, 2008, pp. 1949–1954.

[26] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: Application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 2037–2041.

[27] Y.M.G. Costa, L.S. Oliveira, A.L. Koerich, F. Gouyon, J.G. Martins, Music genre classification using LBP textural features, Signal Process. 92 (11) (2012) 2723–2737.

[28] D. Bertolini, L.S. Oliveira, E. Justino, R. Sabourin, Texture-based descriptors for writer identification and verification, Expert Syst. Appl. 40 (6) (2013) 2069–2080.

[29] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.

[30] V. Ojansivu, J. Heikkilä, Blur insensitive texture classification using local phase quantization, in: Image and Signal Processing, Springer, 2008, pp. 236–243.

[31] J. Chen, V. Kellokumpu, G. Zhao, M. Pietikäinen, RLBP: Robust Local Binary Pattern, in: Proceedings of the British Machine Vision Conference, 2013, pp. 1–12.

[32] V. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, 1995.

[33] E. Severo, R. Laroca, C.S. Bezerra, L.A. Zanlorensi, D. Weingaertner, G. Moreira, D. Menotti, A benchmark for iris location and a deep learning detector evaluation, in: 2018 International Joint Conference on Neural Networks, IJCNN, 2018, pp. 1–7.

[34] G.Z. Felipe, R.L. Aguiar, Y.M.G. Costa, C.N. Silla, S. Brahnam, L. Nanni, S. McMurtrey, Identification of infants' cry motivation using spectrograms, in: 2019 International Conference on Systems, Signals and Image Processing, IWSSIP, 2019, pp. 181–186.

[35] J. Kittler, M. Hatef, R.P.W. Duin, J. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) (1998) 226–239.

[36] Y. Bengio, A. Courville, P. Vincent, Representation learning: A review and new perspectives, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1798–1828.

[37] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, Breast cancer histopathological image classification using convolutional neural networks, in: Neural Networks (IJCNN), 2016 International Joint Conference on, IEEE, 2016, pp. 2560–2567.

[38] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, H. Adeli, Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals, Comput. Biol. Med. 100 (2018) 270–278.

[39] L.G. Hafemann, L.S. Oliveira, P. Cavalin, Forest species recognition using deep convolutional neural networks, in: 2014 22nd International Conference on Pattern Recognition, IEEE, 2014, pp. 1103–1107.

[40] A.C.G. Vargas, A. Paes, C.N. Vasconcelos, Um Estudo sobre Redes Neurais Convolucionais e sua Aplicação em Detecção de Pedestres, in: Conference on Graphics, Patterns and Images, vol. 29, IEEE, 2016, pp. 1–4.

[41] M.N. Roecker, Y.M.G. Costa, J.L.R. Almeida, G.H.G. Matsushita, Automatic vehicle type classification with convolutional neural networks, in: 2018 25th International Conference on Systems, Signals and Image Processing, IWSSIP, IEEE, 2018, pp. 1–5.

[42] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[43] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.

[44] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252.

[45] Y. Zhou, H. Nejati, T.-T. Do, N.-M. Cheung, L. Cheah, Image-based vehicle analysis using deep neural network: A systematic study, in: Digital Signal Processing (DSP), 2016 IEEE International Conference on, IEEE, 2016, pp. 276–280.

[46] R. Laroca, E. Severo, L.A. Zanlorensi, L.S. Oliveira, G.R. Gonçalves, W.R. Schwartz, D. Menotti, A robust real-time automatic license plate recognition based on the YOLO detector, in: 2018 International Joint Conference on Neural Networks, IJCNN, 2018, pp. 1–10.

[47] J. Redmon, A. Farhadi, YOLO9000: Better, faster, stronger, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017, pp. 6517–6525.

[48] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779–788.

[49] J. Redmon, A. Farhadi, YOLOv3: An incremental improvement, 2018, CoRR arXiv:abs/1804.02767.