A Comparison of Doubly Hierarchical Discriminant Analyses for Multiple Class Longitudinal Data from EEG Experiments

WOUTERS, Kristien; CORTINAS ABRAHANTES, Jose; MOLENBERGHS, Geert; Ahnaou, Abdellah; Drinkenburg, W.H.I.M. & BIJNENS, Luc (2008) A Comparison of Doubly Hierarchical Discriminant Analyses for Multiple Class Longitudinal Data from EEG Experiments. In: JOURNAL OF BIOPHARMACEUTICAL STATISTICS, 18(6). p. 1120-1135.

DOI: 10.1080/10543400802369111
Handle: http://hdl.handle.net/1942/9001

A Comparison of Doubly Hierarchical Discriminant Analyses for Multiple Class Longitudinal Data from EEG Experiments

Kristien Wouters, José Cortiñas Abrahantes, Geert Molenberghs,
Universiteit Hasselt, Center for Statistics, Agoralaan, B3590 Diepenbeek, Belgium
Abdellah Ahnaou, Wilhelmus H.I.M. Drinkenburg, and Luc Bijnens
Johnson & Johnson Pharmaceutical Research and Development, Turnhoutseweg 30,
B2340 Beerse, Belgium

Abstract

This paper proposes a general and simple procedure that can be applied to establish classification rules for application with multiple-class longitudinal data. The procedure is applied to pre-clinical pharmaco-electroencephalogram (EEG) studies aiming at characterizing psychotropic drug effects on the basis of spectral EEG analysis. It is a flexible hierarchical supervised learning tool, allowing to take the specific nature of the multiple drug classes into account, as well as the longitudinal aspect of the data. Several variations to this procedure are applied to the EEG data, generally producing comparable results, in particular similar association between the sleeping stages and the psychotropic drug classes.

Keywords

Linear mixed model; linear discriminant analysis; flexible discriminant analysis; mixture discriminant analysis; fractional polynomials.

1 Introduction

Classification techniques are used in a wide range of human activity. One such use, for example, is the preliminary diagnosis of a patient's disease in view of instantaneously selecting treatment while awaiting conclusive test results. In fact, the term could be used in any context in which some decision is made on the basis of available information, and a classification procedure is then some formal method for making such judgments in a particular situation. In this research, we will construct a method for application with a continuously accruing sequence of cases, in which each newly collected case must be assigned to one of

a number of pre-defined classes on the basis of observed features. To build such a classification procedure from a set of data for which we know the true classes, we will employ discrimination.

For longitudinal data classical discriminant analysis is not necessarily one's best option, since it ignores the correlation between measurements on the same subject. Of course, developing classification rules for complex data structures, such as multiple-class problems within a longitudinal design, is a non-trivial task and requires appropriately tailored methods. The combination of precisely these features is encountered in so-called pharmaco-electroencephalogram (EEG) experiments, conducted to establish classification rules for differing psychotropic drug classes. The potential of using EEG-derived parameters (pharmaco-EEG) and characteristic fingerprints on rodent sleep-wake architecture for the classification of drugs has been recognized for several decades and is used as a valuable tool in both preclinical drug discovery and clinical drug development (Fink (1959), Krijzer *et al* (1993), Ruigt *et al* (1993), Depoortere *et al* (1995), Edgar (2002), Drinkenburg and Ahnaou (2004), Uchida *et al* (2007)). In addition EEG technology is used to identify biomarkers that have predictive validity for clinical, pharmacological activity and even for possible efficacy (i.e. as surrogate endpoint) for some diseases such as Major Depressive Disorder (MDD) (Mucci *et al* (2006), Murck *et al* (2003), Staner *et al* (2004)). These experiments motivate our research. Classical discriminant analysis is not suited to handle the combination of a multiple-class problem and a longitudinal design. A flexible two-step procedure, termed doubly hierarchical discriminant analysis (Wouters *et al*, 2007), has been proposed to cope with such problems.

In this paper, we aim to classify psychotropic drugs based on the sleep wake behaviour of rats. The so-called doubly hierarchical discriminant analysis or DHDA, introduced by Wouters *et al* (2007), has been shown to perform adequate in the experimental data described in the next section, but there was room for improvement, as acknowledged by aforementioned authors. Here, we introduce a more general form of the DHDA procedure, and apply several particular instances of the procedure to the data.

In the next section, the data are described and some background on the experiments is provided. In Section 3, the methodology is explained, starting from the general form of the doubly hierarchical discriminant analysis, followed by a number of different versions. In Section 4, these variations to the theme are put to the test on the data and afterwards the results will be compared.

2 Data Description

Many different recording technologies exist today for measuring brain activity. A graphical record of electrical activity of the brain, produced by an electro-encephalograph (EEG), is one of them. EEG experiments have been used for many purposes. We are interested in particular in preclinical EEG studies aiming at characterizing psychotropic drug effects on the basis of spectral EEG analysis. Psychotropic drugs can be divided into 5 major classes: antidepressants, antipsychotics, anxiolytics, hypnotics and stimulants (Deniker, 1982; Oughourlian, 1984; Cohen and Cailloux-Cohen, 1995), each one typically named according to its main indication in psychiatry. Classifying drugs solely based on chemical structure would create numerous categories, which would not necessarily be indicative of their therapeutic use. New chemical entities ideally should be classified based on their potential therapeutic activity as early as possible in the drug discovery process. Availability of an advanced classification model or tool that uses a standardized physiological read-out, such as the preclinical sleep-wake EEG, would greatly aid efficient determination of psychoactive properties of novel chemical entities and allow for better selection or to more reliably prioritize new molecules to be developed further in highly costly clinical trials. From a translational medicine point of view, the use of rodent sleep-wake EEG data allows for direct extrapolation to human volunteers (phase 1) or patients (phase 2/3). Sleep-wake studies in human and patients to date have provided some important points towards the importance of higher sigma frequency range in the sleep EEG of non-REM sleep and REM density as a marker of treatment response (Murck *et al* (2003)).

In general, the aim of pharmaco-electroencephalographical studies is to characterize psychotropic drug effects. When EEG measurements are combined with movement monitoring and the recording of muscle activity, clearly defined states of vigilance can be separated out and used to classify psychotropic agents. Typically, six sleep-wake stages are distinguished: (1) *active wake (AW)*; (2) *passive wake (PW)*; (3) *light sleep (SWS1)*; (4) *deep sleep (SWS2)*; (5) *intermediate stage sleep (IS)*; (6) *Rapid Eye Movement or REM Sleep (RS)*. Further details on this type of data can be found in Wouters *et al* (2007)

The study considered here includes 26 psychoactive agents at 4 different doses, including a zero dose constituting a large database of well-controlled rat experiments that were carried out in a highly

standardized way using clinically active reference drugs within the framework of pharmaceutical drug discovery (Drinkenburg and Ahnaou (2004)). For each of these compounds, 8 rats were randomly assigned to each of the 4 doses. The brain signals of the rats are monitored during a period of 16 hours, divided into a light period (10 hours) and a period of darkness (6 hours). The treatment is administered at the beginning of the light period and after each experiment 3 weeks of washout period are considered before using the same rat in another experiment. The effects of the compounds on sleep-waking behavior are assessed using several hypnogram parameters. For every interval of 30 minutes the time spent in each of the six sleeping stages is measured (in minutes).

From the 104 compound dose combinations in the original dataset, only those for which the psychotropic drug class is known are used in the training set. These compound-dose combinations are also extensively used in clinical practice. The training data now consists of 64 compound-dose combinations: 26 placebos, 14 antidepressants, 7 antipsychotics, 2 anxiolytics, 5 hypnotics, and 10 stimulants.

3 Doubly Hierarchical Discriminant Analysis

Doubly hierarchical discriminant analysis (DHDA) has been proposed by Wouters *et al* (2007). In this paper, a more general form of this procedure is proposed.

Figure 1, About Here.

The procedure is schematically represented in Figure 1. In a first stage, the longitudinal profiles are modeled and appropriate summaries extracted from the model fit. These summary measures are then used, in the second stage, as input for the discriminant analysis, in view of classifying the data. A similar approach has been used by MacLachlan (1992). This second stage proceeds in a hierarchical fashion. Let us zoom in on each of the two stages in turn.

3.1 Stage I: Modelling the Longitudinal Data

In the first stage, we model the longitudinal data so as to obtain relevant summaries from the profiles. A modelling approach that allows for capturing complexities and intricacies in the data, while lending itself

easily to the obtention of simple summaries is to be preferred. While several approaches are possible, we will use so-called fractional polynomial mixed models (FPMM). Linear mixed effects models (LMM) are a widely used tool for modelling longitudinal data (Verbeke and Molenberghs, 2000). To capture the irregular trends in our profiles, we combine the LMM with the use of fractional polynomial predictor functions (Royston and Altman, 1994). In our case, for each compound-dose combination and each sleeping stage, separate models are fitted to the light, the dark and the first three hours periods. Not only the coefficients, but also the fractional polynomial powers, denoted by subscripted p's, are allowed to differ across compound-dose combinations. The details of this approach can be found in Wouters *et al* (2007).

Let us now turn to the second stage.

3.2 Stage II: Hierarchical Discriminant Analysis

The continuation of the classification procedure necessitates informative summaries of the highly variable longitudinal profile available for each rat. To this end, the parameters of the models in the first stage, i.e., the collection made up of $\beta_{0i} + b_{0ij}$, $\beta_{1i} + b_{1ij}$, $\beta_{2i} + b_{2ij}$, p_{1il}, p_{2il}, $\gamma_{0i} + c_{0ij}$, $\gamma_{1i} + c_{1ij}$, $\gamma_{2i} + c_{2ij}$, p_{1id}, p_{2id}, $\delta_{0i} + d_{0ij}$, $\delta_{1i} + d_{1ij}$, $\delta_{2i} + d_{2ij}$, p_{1if}, and p_{2id}, will be used as input to the discriminant procedure.

Note that this leads to important economy. Indeed, the rat-specific model parameters are used as predictors. We have 5 parameters per model (p_1, p_2, intercept, and the coefficients of both covariates), with 6 models (for the 6 sleep-wake stages) in the light period, 6 in the dark period and 6 for the first 3 hours. Since the light period on the one hand and the first three hours on the other hand present overlapping information, each sleeping stage is at most used for one of these two periods. This results in at most $12 \times 5 = 60$ parameters, much less than the number of observations in the dataset, brought down from 496 in the first step to 258 in the last step.

To establish and optimize a flexible classification rule, we proceed in a stepwise, hierarchical way. In a first step, we discriminate, for example, stimulants from the other psychotropic classes, using the parameters describing the longitudinal profile pertaining to some of the sleep-waking stages for the three different periods considered (first 3 hours, light period, and dark period). Then, focus shifts to the remaining five classes. This process continues until a complete decision tree has been built. The order in which the

classes are discriminated is determined based on the performance in the training dataset. Different orders are checked and the one that leads to the best classification results using cross-validation, is retained.

Various supervised learning techniques can be used at this stage. Numerous comparisons of different classification methods have been made in different areas; examples include Bauer and Kohavi (1999), Caruana and Niculescu-Mizil (2005), Lee *et al* (2005). In this paper, we will focus only on discriminant analysis which uses s procedure similar to the linear discriminant analysis (LDA), but more flexible when constructing the classification rule. In particular, we will focus on flexible (FDA) and mixture (MDA) discriminant analysis, since they both extend LDA methodology.

A graphical display of the doubly hierarchical discriminant analysis, when either LDA, FDA, or MDA are used, is presented in Figure 2.

FIGURE 2, ABOUT HERE.

We will now briefly outline each of the three choices in turn.

3.2.1 Linear Discriminant Analysis

In linear discriminant analysis (Hastie, Tibshirani and Friedman, 2001), each class is assumed to follow a multivariate normal distribution with common variance-covariance matrix, leading to a linear decision rule. This linearity makes it easy to implement and interpret the decision boundaries. Unfortunately, in a number of situations, linear decision boundaries are not adequate to separate the classes. To account for this, Hastie, Tibshirani, and Friedman (2001) propose generalizations of LDA, such as flexible discriminant analysis (FDA), mixture discriminant analysis (MDA), and penalized discriminant analysis (PDA). In what follows, we will confine attention to FDA, because of its ability to model irregular decision boundaries, and MDA, which allows us to use more than one prototype per class.

Another way to think about linear discriminant analysis is by assuming one disposes of observations with a qualitative response, G say, falling into one of C classes, $\Omega = \{1, \ldots, C\}$, for which some features X are measured. Suppose now that we have got a function assigning scores to the classes $\theta : \Omega \to \mathbb{R}$, such that a linear regression on X optimally predicts the class labels. For a sample of the form (g_i, x_i),

$i = 1, \ldots, n$, one then needs to solve

$$\min_{\beta,\theta} \sum_{i=1}^{n} (\theta(g_i) - x_i^T \beta)^2,$$

with restrictions imposed on θ to avoid a trivial solution.

More generally, we can find up to $L \leq C - 1$ sets of independent scorings for each of the class labels, $\theta_1, \ldots, \theta_L$. Scores θ_l and β_l are then chosen to minimize the average squared residuals

$$ASR = \frac{1}{n} \sum_{\ell=1}^{L} \left[\sum_{i=1}^{n} (\theta_\ell(g_i) - x_i^T \beta_\ell)^2 \right].$$

The scores are assumed to be mutually orthogonal and normalized, to prevent trivial zero solutions.

3.2.2 Flexible Discriminant Analysis

The linear discriminant analysis can be regarded as a sequence of linear regression followed by classification to the closest class centroid in the space of fits. The linear regression will now be generalized to a more flexible one (Hastie, Tibshirani and Friedman, 2001). In this more general form, the regression problems are defined via

$$ASR = \frac{1}{n} \sum_{\ell=1}^{L} \left[\sum_{i=1}^{n} (\theta_\ell(g_i) - f(x_i))^2 + \lambda J(f) \right],$$

where J is a regularizing function, specific choices of which correspond to specific non-parametric regression techniques.

In our particular case, we use Multivariate Adaptive Regression Splines (MARS) models (Friedman, 1991). The input space is partitioned into regions, each with its own linear regression equation. The MARS equation is given by

$$f(x) = \gamma_0 + \sum_{m=1}^{M} \gamma_m h_m(x),$$

where M is the number of non-constant terms in the model and h_m is a basis function in the collection

$$\mathcal{C} = \{(X_j - t)_+, (t - X_j)_+ | t \in \{x_{1j}, x_{2j}, \ldots, x_{nj}\}, j = 1, 2, \ldots, p\},$$

with n is the number of observations.

3.2.3 Mixture Discriminant Analysis

Mixture discriminant analysis (Hastie, Tibshirani and Friedman, 2001) is an extension of LDA, to be viewed as a prototype classifier with each class represented by its centroid. We assign an observation to the closest centroid using an appropriate distance measure. In many situations, a single prototype per class is not sufficient, in which case mixture models can be used. Assume classes have several prototypes, thence a Gaussian mixture model for the k^{th} class could be considered. The corresponding density is

$$P(X|C=k) = \sum_{r=1}^{R_k} \pi_{kr}\phi(X;\mu_{kr},\Sigma),$$

where the mixing proportions satisfy $\sum_{r=1}^{R_k}\pi_{kr} = 1$, R_k is the prototype for the k^{th} class and Σ the covariance matrix used as a metric throughout. For class k with *a priori* probabilities Π_k, we estimate the parameters by maximizing the joint log-likelihood:

$$\sum_{k=1}^{K}\sum_{g_i=k}\log\left[\sum_{r=1}^{R_k}\pi_{kr}\phi(X;\mu_{kr},\Sigma)\Pi_k\right].$$

The expectation-maximization (EM) algorithm is a convenient mode to obtain maximum likelihood estimates. In order to obtain the maximum likelihood estimates, we use the EM algorithm (Dempster, Laird, and Rubin, 1977). The algorithm consists of iterating between the expectation (E) and maximization (M) steps, until convergence. In our situation, they take the following forms.

E-step: Given the current values for the parameters, compute the weights associated with the subclasses c_{kr}:

$$W(c_{kr}|x_i,g_i) = \frac{\pi_{kr}\phi(x_i;\mu_{kr},\Sigma)}{\sum_{l=1}^{R_k}\pi_{kl}\phi(x_i;\mu_{kl},\Sigma)}. \quad (1)$$

M-Step: Compute weighted MLEs for the parameters of each of the component Gaussian densities, within each of the classes, using the weights obtained from (1).

3.3 Lack-of-classification Measure

To define the goodness of a model we first introduce some notation. Let g be the initial number of classes, $\mathrm{ERR}_{kl}^{(i)}$ the misclassification percentage from class k into class l for model $m^{(i)}$, $\mathrm{PP}_{kl}^{(i)}$ the posterior

probability for rats belonging to class k to be classified in class l for model $m^{(i)}$, C_s the class discriminated in step s, and C_{-s} the classes retained in the dataset in step s. Further, write

$$\begin{aligned} \text{Error1}_s^{(i)} &= \text{ERR}^{(i)}_{C_s C_{-s}} + \left(1 - \text{PP}^{(i)}_{C_s C_s}\right) \\ \text{Error2}_s^{(i)} &= \text{ERR}^{(i)}_{C_{-s} C_s} + \sum_{k \neq C_s} \text{PP}^{(i)}_{k C_s} \end{aligned}$$

Now, we can define the following lack-of-classification measure (LC) for model m_i in step s as a weighted sum of $\text{Error1}_s^{(i)}$ and $\text{Error2}_s^{(i)}$

$$\text{LC}_s^{(i)} = w_{s1} \cdot \text{Error1}_s^{(i)} + w_{s2} \cdot \text{Error2}_s^{(i)}.$$

The first term in this formula is monitoring the false-negative cases, while the second one focuses on the false positives. Different weights w_{s1} and w_{s2} can be chosen, depending on the type of application. In our particular case, we chose the following weights $w_{s1} = s + 1$ and $w_{s2} = 2 \cdot (g - s)$. Along the process, more weight is given to the false-negatives whereas the weight given to the false-positives is decreased. The choice of these weights is based on the fact that the algorithm discriminates in the first steps the classes that are well differentiated from the rest whereas in the final steps the classes are less clearly separated.

The lack-of-classification measure is now standardized and corrected for the number of parameters in the model by multiplying with a decreasing function of the number of sleep-wake stages used, given by $F(\text{ss})$:

$$\text{LC}'^{(i)}_s = 1 - \left(1 - \frac{\text{LC}_s^{(i)}}{2 \cdot w_{s1} + (g - s + 1) \cdot w_{s2}}\right) \cdot F(\text{ss}).$$

Again, different choices can be entertained for $F(ss)$. We decided to work here with $F(\text{ss}) = 0.999^{\text{ss}}$. With this choice of F, a difference of at least 0.001 in LC' is enough to motivate an extra sleep-wake stage in the model. This difference of 0.001 in LC' corresponds to a difference in LC varying from 0.022 in step 1 to 0.01 in step 4.

3.4 Selection Procedure

We consider two different selection procedures, both based on 10-fold cross-validation, a technique to be described next, inspired by the fact that the dataset can be divided randomly at each of two different hierarchical levels.

In the first approach (Selection Procedure I), we use rats as the unit of analysis. The 512 rats comprising the dataset are then randomly divided into ten groups. For every parameter combination obtained from the fractional polynomial models and for each sleep-waking stage, one of the 10 samples is used as a test dataset, while the remaining 9 samples are assigned the role of training sets. For the test dataset, both the misclassification error and the posterior probabilities are calculated. The combination of sleep-waking stages resulting in the lowest lack-of-classification measure is retained. This is repeated for every step in the DHSLA.

Selection Procedure II uses 10-fold cross-validation at the compound-dose combination level. We randomly divide the 64 such combinations into ten groups and then proceed in the same way it was described above.

The posterior probabilities of belonging to each of the six drug classes are determined in an iterative way. At the first split of the agents into two subclasses, posterior probabilities are calculated for each of them. Generally, given that k splits have been made, the values of the posterior probabilities at split $k+1$ are multiplied with the posterior probabilities of not being classified at the previous steps in the class we were interested to discriminate from the rest (Wouters *et al*, 2007).

For each selection procedure, the error count is calculated at both levels, i.e., rat and compound-dose combination. The first is computed as the average of the percentage of misclassified rats in each class ($\text{error}_{\text{rat}}$), while the second uses the percentages of compound-dose combinations that are misclassified in a particular class ($\text{error}_{\text{c-d}}$).

4 Results

Upon building a fractional polynomial mixed model for the light and dark periods, as well as for the first three hour period, separately for each of the compound-dose combinations observed, the parameters of all these models are used in a stepwise discriminant analysis. As an illustration the fitted fractional polynomial for the number of minutes spent in Active Wake for one particular compound-dose combination is shown in Figure 3 together with the mean profile for that compound-dose combination.

In all three discriminant procedures, the order in which the classes are separated is the same, but the sleeping stages used in every step are allowed to differ. We sequentially discriminate stimulants, then anxiolytics, antipsychotics, antidepressants; finally hypnotics are separated from placebo. The sleeping stages used in each step of Stage II are obtained by means of both selection procedures described in Section 3.4. Because the first three hour period is part of the light period, the first three hours are excluded from the latter to avoid double use.

In what follows, the results obtained with linear discriminant analysis, flexible discriminant analysis built on MARS, and mixture discriminant analysis with 2 subclasses per group are compared with respect to the sleeping stages used in each step and the performance with 10-fold cross-validation at the rat level as well as at the compound-dose level.

Due to lack of information about the anxiolytic class (only 2 compound-dose combinations), it was decided to exclude this class from the classification procedure.

4.1 Linear Discriminant Analysis

In Table 1, the sleeping stages retained per step for the linear discriminant analyses, obtained with 10-fold cross-validation on the rat level and the compound-dose combination level are shown.

Table 1, About Here.

The number of sleeping stages needed in each step is similar for the two selection procedures. Some sleeping stages are selected by both selection procedures for some steps. For example, to discriminate stimulants, Passive Wake, Light Sleep and Deep Sleep during the light period and Active Wake and Light Sleep in the dark period are selected by both analyzes. This lines up with expectation because a stimulant generally increases Active Wake and reduces Light Sleep and Deep Sleep. For the other classes, we observe some further similarities between the two selection procedures. Passive Wake, Deep Sleep, and Intermediate Stage Sleep in the light period,together with Active Wake in the dark period are always selected for the classification of antipsychotics. For antidepressants, Active Wake and Light Sleep in the first three hour period, Intermediate Stage Sleep, and REM Sleep in the light period are selected by both procedures.

Finally for hypnotics, Light, Deep, and REM Sleep in the light period or in the first three hours are retained by both selection procedures.

Table 2 shows the classification results obtained with DHSLA when fractional polynomials and linear discriminant analysis are used in each of the two stages and for both selection procedures. For every psychotropic class, the adjusted posterior probabilities for the six classes are given. The observed adjusted posterior probabilities obtained without cross-validation are presented parenthetically.

Table 2, About Here.

As expected, the adjusted posterior probabilities, obtained with Selection Procedure I, are higher than those coming from Selection Procedure II. Leaving out one tenth of the rats will rarely result in leaving out a whole compound-dose combination, hence there is still information on all compound-dose combinations in the training dataset, leading to better classification. This is different when applying Selection Procedure II, when good results are obtained only when a representative sample of the compound-dose combination population is taken for the training of the procedure.

The adjusted posterior probabilities for correct classification obtained with Selection Procedure I are all very high and above 90%. The error count for this selection procedure is 0.00 at the rat level as well as at the level of the compound-dose combinations. For Selection Procedure II, the adjusted posterior probabilities for antidepressants, and stimulants remain high (above 85%). The ones for placebo, antipsychotics, and hypnotics are somewhat lower, but still above 70%. The error counts for this selection procedure are 0.12 and 0.11 at the rat and compound-dose levels, respectively.

4.2 Flexible Discriminant Analysis

The sleeping stages selected per step when flexible discriminant analysis is used in the DHSLA for the selection procedures described in Section 3.4 are shown in Table 3. Similar to the case when the linear discriminant analysis is used in the DHSLA, we see that the number of sleeping stages retained does not differ much for both selection procedures. Also, we note that for some classes both selection procedures arrive at selecting the same sleeping stages, indicating association between the class and its effect on a

particular sleeping stage. For example, for stimulants, we retain Active Wake in the dark period, with either selection procedure. For antipsychotics, Deep Sleep in the light period and Active Wake, Passive Wake and Deep Sleep in the dark period are common to both selection procedures. Active Wake, Passive Wake and Deep Sleep either in the light period or in the first three hours are retained with both selection procedures for antidepressants. Finally, for hypnotics, Deep Sleep in the light period is selected with both selection procedures.

Table 3, About Here.

Table 4, About Here.

Regarding the adjusted posterior probability, Table 4 displays very high posterior probabilities for the correct classification with flexible discriminant analysis, obtained with Selection Procedure I. For Selection Procedure II, high posterior probabilities for placebo, antidepressants, and hypnotics are obtained too, while those for antipsychotics and stimulants are somewhat lower but still above 70%.

4.3 Mixture Discriminant Analysis

The corresponding results for mixture discriminant analysis are presented in Tables 5 and 6, respectively.

Table 5, About Here.

Similar conclusions can be drawn with respect to the sleeping stages retained at each step. Furthermore, the number of stages retained is similar for both selection procedures, and some are selected by both. For example, for stimulants, Deep Sleep in the light period and Light Sleep in the first three hours are chosen irrespective of the selection procedure used. Similar conclusions can be drawn for the other classes in the hierarchical procedure.

Table 6, About Here.

For the adjusted posterior probabilities, we observe, once more, very promising results for Selection Procedure I. All posterior probabilities for the correct classes are above 90%. For Selection Procedure II, all adjusted posterior probabilities are above 70%. The error count for this procedure amounts to 7%.

5 Discussion

In this paper, we presented a method for classifying potentially active compounds into a predefined set of psychotropic classes. Sleep and wake EEG data, longitudinally collected in rats, are used to this effect. The method proceeds by first analyzing the longitudinal profiles, using the flexible fractional polynomial regression linear mixed model, and then continuing in a second stage with one of three forms of discriminant analysis: linear, flexible, and mixture discriminant analysis. The second stage is an instance of supervised learning. Selection of compounds is done using either the individual rat or the compound by dose combination as unit of analysis.

While the selection of the sleeping stages was done ad-hoc in Wouters *et al* (2007), the lack-of-classification measure now provides us with a formal procedure to define the sleeping stages that are needed in each step of the discriminant analysis to classify a particular psychotropic class. The sleeping stages retained by both methods are showing similarities, but for future analysis, the last method will be coherent, while the ad-hoc method would not necessarily be.

The number of sleeping stages used by the method at each step is very stable across the three discriminant techniques for both selection procedures. Some sleeping stages are retained in all analyzes irrespective of the discriminant analysis or the selection procedure.

It appears that the level on which the cross-validation is performed plays an important role in the selection of the sleeping stages. In the second step, Deep Sleep is selected in the dark period for all three discriminant techniques for Selection Procedure II, but does not show up in any of the analyses when Selection Procedure I is applied. The same is observed for Active Wake in the light period and Passive Wake in the dark period, for the third step. On the other hand, we have some sleeping stages that are needed in all three discriminant analyses when using Selection Procedure I, but not at all when Selection

Procedure II is used, such as Deep Sleep in the dark period in the fourth step. For the last step, the same combination of sleeping stages is retained when Selection Procedure II is combined with linear, flexible, and mixture discriminant analysis.

Variables that appear in a certain step for all three discriminant analyses, regardless of the selection procedure, can be seen as important variables for the discrimination of that particular class from the rest. The first three hours are part of the light period; therefore we will consider a variable as common when it is used either in the light period or in the first three hours. In general, Light Sleep and Deep Sleep in either the light period or the first three hours, and Active Wake in the dark period are showing up in the first step, designed to discriminate stimulants from the rest. This agrees with expectation, because stimulants generally decrease the time spent in light and deep sleep. Deep Sleep and Intermediate Stage Sleep in the light or the first three hour period and Active Wake in the dark, appear in all six analyses in the third step to classify antipsychotics. Active Wake, Light Sleep, Intermediate Stage Sleep, and REM Sleep in the light or the first three hour period seem to be crucial to classify antidepressants. Finally, for hypnotics we see that Active Wake, Deep Sleep, and Intermediate Stage Sleep in either the light period or the first three hours are retained in all the analyses. The error counts in Selection Procedure I are lower than those obtained with Selection Procedure II for all discriminant analyses. The error counts on the rat level are higher than those on the compound-dose level for both selection procedures and in all the discriminant analyses. As all three discriminant procedures produce comparable results in terms of posterior probabilities and error counts, it would be fair to recommend the use of linear discriminant analysis in similar settings, also in view of its simplicity.

For all three discriminant techniques, the adjusted posterior probabilities, obtained with Selection Procedure I, are higher than the ones obtained with Selection Procedure II. This was expected, because leaving out one tenth of the rats at random rarely results in leaving out a whole compound-dose combination. Therefore, all compound-dose combinations in a certain test dataset are still present in the corresponding training dataset. This alleviates the classification of a rat in the test dataset.

In this paper, we have only compared three discriminant techniques, driven by the similarities between them in methodological terms. Nevertheless, they allow for different degrees of flexibilities to

construct the classification rule to be used. Of course, other supervised learning techniques can be used as well, such as non-parametric discriminant analysis, support vector machines, neural networks, random forest, boosting methods, etc.

As a final remark, the methods developed here are tightly linked to the motivating problem, coming from the wish to classify potentially active psychotropic compounds or, rather, compound-by-dose combinations. It is evident that the methodology can be used in a variety of similar preclinical and clinical settings, across the widest range of therapeutic areas.

The method has been tuned for this particular dataset, but the general idea as presented in Figure 1 can be applied to any other dataset. The techniques used in the first and the second phase of the procedure can be tuned to the situation.

The interested reader can obtain the software programs from the authors upon request.

References

Bauer, E., Kohavi, R. (1999). An empirical comparison of voting classification algorithms: bagging, boosting and variants. *Machine Learning* **36 (1-2)**, 105–139.

Caruana, R., Niculescu-Mizil, A. (2005). An empirical comparison of supervised learning algorithms using different performance metrics. *Technical Report* TR2005-1973, Cornell University, Ithaca,USA (2005)

Caruana, R., Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. *Proceedings of the 23rd international conference on machine learning, June 25–29, Pittsburgh, Pennsylvania* 161–168.

Cohen, D. and Cailloux-Cohen, S. (1995). *Guide critique des médicaments de l'âme.* Québec, Les Editions de l'Homme.

Deniker, P. (1982). Vers une classification automatique des psychotropes à travers un fichier informatisé de leurs propriétés. *Annales Médico-psychologiques*, **1**, 25–27.

Dempster, A.P., Laird, N.M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.

Depoortere, H., Francon, D., Van Luijtelaar, E.L.J.M., Drinkenburg, W.H.I.M. and Coenen, A.M.L. (1995). Differential effects of midazolam and zolpidem on sleep-wake states and epileptic activity in WAG/Rij rats. *Pharamacology, Biochemistry and Behavior*, **51**, 571–576.

Drinkenburg, W.H.I.M. and Ahnaou, A. (2004). The Use of pEEG in Preclinical Models in Drug Discovery *Essentials and Applications of EEG Research in Preclinical and Clinical Pharmacology.*, Berlin, International Pharmaco-EEG Group, –.

Edgar, D.M. (2002). Signature profiles in sleep-wake drug discovery. *Methods and Findings in Experimental and Clinical Pharmacology*, **24 Suppl. D**, 71–72.

Fink, M. (1959). EEG and behavioural effects of psychopharmacology agents. *Neuropsychopharmacology*, **1**, 441–446.

Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine . *Annals of Statistics*, **29**, 1189–1232.

Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The elements of statistical learning: data mining, inference, and prediction.* Springer Series in Statistics, Springer-Verlag, New-York.

Krijzer, F., Koopman, P. and Olivier, B. (1993). Classification of psychotropic drugs based on pharmaco-electrocorticographic studies in vigilance-controlled rats. *Neuropsychobiology* **28**, 122–137.

Lee, J.W., Lee, J.B., Park, M., Song, S.H. (2005). An extensive comparison of recent classification tools applied to microarray data. *Computational Statistics and Data Analysis* bfseries 48 (4), 869–885.

MacLachlan, G.J. (1992). *Discriminant analysis and statistical pattern recognition.* New York, NY, Wiley.

Mucci, A., Volpe, U., Merlotti, E., Bucci, P., Galderise, S. (2006). Pharmaco-EEG in psychiatry. *Clinical EEG and Neuroscience*, **37**, 81–98.

Murck, H., Nickel, T., Künzel, H., Antonijevic, I.A., Schill, J., Zobel, A., Steiger, A., Sonntag, A. and Holsboer, F. (2003) State markers of depression in sleep EEG: Dependency on drug and gender in patients treated with tianeptine or paroxetine. *Neuropsychopharmacology*, **28**, 348–358.

Oughourlian, J.M. (1984). *La personne du toxicomane. Psychosociologie des toxicomanies actuelles dans la jeunesse occidentale.* Toulouse, Privat.

Royston, P, Altman, D.G. (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling (with Discussion). *Applied Statistics* **43**, 429–467.

Ruigt, G.S., Engelen, S., Gerrits, A., and Verbon, F. (1993). Computer-based prediction of psychotropic drug classes based on a discriminant analysis of drug effects on rat sleep. *Neuropsychobiology* **28** 138–153.

Staner, L.P.J., Luthringer, R., and Macher, J-P. (2004) Antidepressant induced alteration of sleep EEG as a surrogate marker of drug activity: A window to the neurobiology of depression. *The World Journal of Biological Psychiatry*, **5**, Supp. 1, 8034.

Uchida, M., Suzuki, M., Shimizu, K. (2007) Effects of Urocortin, Corticotropin-Releasing Factor (CRF) Receptor Agonist, and Astressin, CRF Receptor Antagonist, on the Sleep-Wake Pattern: Analysis by Radiotelemetry in Conscious Rats. *Biological and Pharmaceutical Bulletin*, **30** , No. 10 1895.

Verbeke, G. and Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* Springer Series in Statistics, Springer-Verlag, New-York.

Wouters, K., Ahnaou, A., Cortiñas, J., Molenberghs, G., Geys, H., Bijnens, L., and Drinkenbrug, W.H.I.M. (2007). Psychotropic drug classification based on sleep-wake behaviour of rats. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* **56 (2)**, 223–234.

Lee, J.W., Lee, J.B., Park, M., Song, S.H. (2005). An extensive comparison of recent classification tools applied to microarray data. *Computational Statistics and Data Analysis* bfseries 48 (4), 869–885.

Bauer, E., Kohavi, R. (1999). An empirical comparison of voting classification algorithms: bagging, boosting and variants. *Machine Learning* **36 (1-2)**, 105–139.

Caruana, R., Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. *Proceedings of the 23rd international conference on machine learning, June 25–29, Pittsburgh, Pennsylvania* 161–168.

Caruana, R., Niculescu-Mizil, A. (2005). An empirical comparison of supervised learning algorithms using different performance metrics. *Technical Report* TR2005-1973, Cornell University, Ithaca,USA (2005)

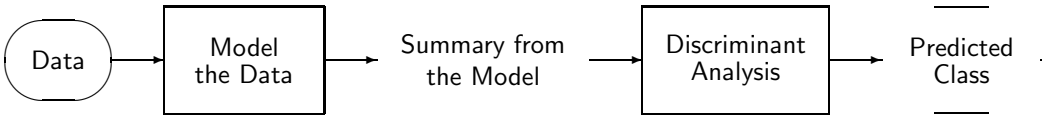

Figure 1: Diagram representing doubly hierarchical discriminant analysis.

Figure 2: Diagram representing doubly hierarchical discriminant analysis, when a fractional polynomial mixed model (FPMM) is used in Stage I and linear (LDA), flexible (FDA) or mixture discriminant analysis (MDA) are used in Stage II.

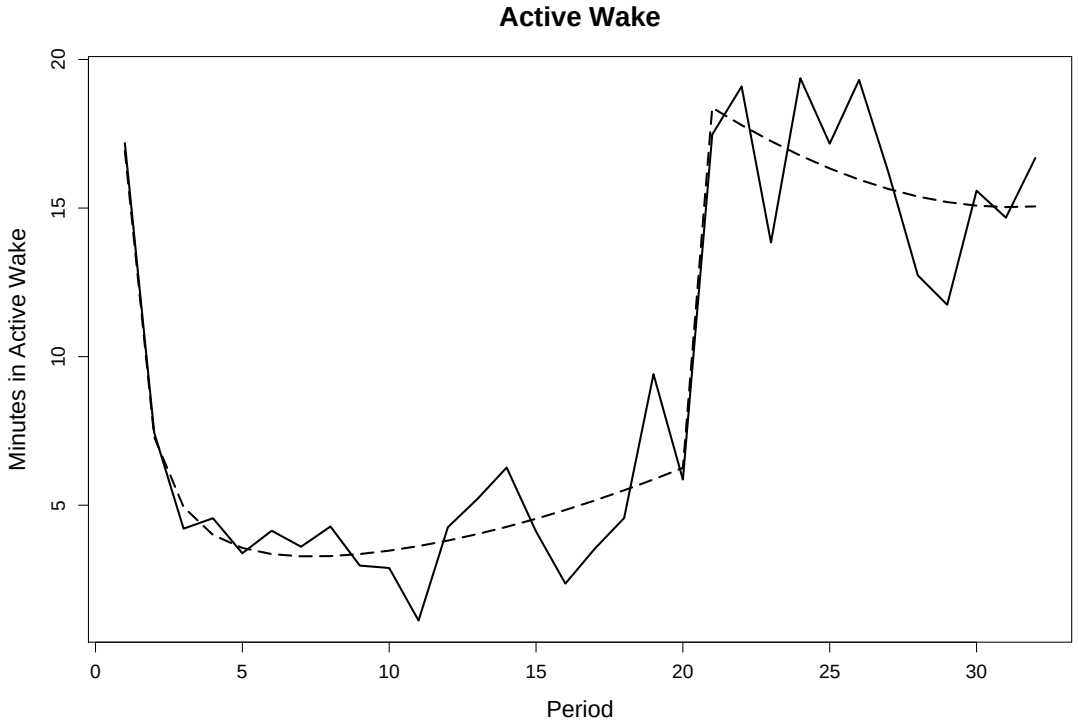

Figure 3: Mean number of minutes spent in Active Wake for one treatment together with the corresponding fitted fractional polynomial profile (dashed line).

Table 1: Linear Discriminant Analysis. Sleeping stages used in each step of the doubly hierarchical discriminant analysis with linear discriminant analysis for both selection procedures.

LDA - Selection Procedure I			
Step	Light period	Dark period	First 3 hours
(1) Stimul	PW SWS1 SWS2	AW PW SWS1 SWS2	
(3) Antipsy	PW SWS2 IS	AW PW SWS2	AW
(4) Antidep	PW IS RS	SWS1 IS	AW SWS1
(5) Hypno	AW SWS2 IS RS		
LDA - Selection Procedure II			
Step	Light period	Dark period	First 3 hours
(1) Stimul	PW SWS1 SWS2	AW SWS1 RS	AW
(3) Antipsy	PW SWS2 IS RS	AW IS	
(4) Antidep	IS RS	PW SWS2 IS	AW SWS1
(5) Hypno	PW SWS2	AW PW SWS2	AW IS

Table 2: Linear Discriminant Analysis. Adjusted posterior probabilities obtained when FPMM and LDA with Selection Procedure I (upper panel) and Selection Procedure II (lower panel) are applied.

LDA - Selection Procedure I ($error_{rat}$ = 0.003 / $error_{c\text{-}d}$ =0.000)					
Drugclass	Placebo	Antipsy	Antidep	Hypnotic	Stimulant
Placebo	**0.96**	0.02	0.02	0.00	0.00
Antipsychotic	0.01	**0.89**	0.04	0.00	0.06
Antidepressant	0.00	0.03	**0.96**	0.00	0.01
Hypnotic	0.00	0.00	0.00	**1.00**	0.00
Stimulant	0.00	0.00	0.01	0.00	**0.99**
LDA - Selection Procedure II ($error_{rat}$ = 0.121 / $error_{c\text{-}d}$ =0.105)					
Drugclass	Placebo	Antipsy	Antidep	Hypnotic	Stimulant
Placebo	**0.77**	0.04	0.09	0.10	0.00
Antipsychotic	0.00	**0.71**	0.12	0.09	0.08
Antidepressant	0.05	0.07	**0.85**	0.03	0.00
Hypnotic	0.18	0.06	0.00	**0.76**	0.00
Stimulant	0.03	0.03	0.06	0.02	**0.86**

Table 3: Flexible Discrimant Analysis. Sleeping stages used in each step of the doubly hierarchical discriminant analysis with flexible discriminant analysis for both selection procedures.

FDA - Selection Procedure I			
Step	Light period	Dark period	First 3 hours
(1) Stimul	PW SWS2	AW PW SWS1	RS
(3) Antipsy	AW SWS2 RS	AW PW SWS2 IS	
(4) Antidep	AW PW SWS1	PW RS	IS RS
(5) Hypno	AW SWS2 IS RS		
FDA - Selection Procedure II			
Step	Light period	Dark period	First 3 hours
(1) Stimul	SWS1	AW SWS2	AW
(3) Antipsy	PW SWS1 SWS2 IS	AW PW SWS2	
(4) Antidep	IS RS	PW IS	AW PW SWS1
(5) Hypno	PW SWS2	AW PW SWS2	AW IS

Table 4: Flexible Discriminant Analysis. Adjusted posterior probabilities obtained when FPMM and FDA with Selection Procedure I (upper panel) and Selection Procedure II (lower panel) are applied.

FDA - Selection Procedure I ($error_{rat} = 0.016$ / $error_{c\text{-}d} = 0.000$)					
Drugclass	Placebo	Antipsy	Antidep	Hypnotic	Stimulant
Placebo	**0.96**	0.00	0.04	0.00	0.00
Antipsychotic	0.00	**0.98**	0.02	0.00	0.00
Antidepressant	0.05	0.03	**0.88**	0.04	0.00
Hypnotic	0.00	0.00	0.13	**0.87**	0.00
Stimulant	0.00	0.00	0.01	0.00	**0.99**
FDA - Selection Procedure II ($error_{rat} = 0.093$ / $error_{c\text{-}d} = 0.059$)					
Drugclass	Placebo	Antipsy	Antidep	Hypnotic	Stimulant
Placebo	**0.82**	0.04	0.09	0.05	0.00
Antipsychotic	0.03	**0.73**	0.18	0.00	0.06
Antidepressant	0.06	0.11	**0.78**	0.04	0.01
Hypnotic	0.00	0.03	0.04	**0.93**	0.00
Stimulant	0.00	0.04	0.12	0.03	**0.81**

Table 5: Mixture Discriminant Analysis. Sleeping stages used in each step of the doubly hierarchical discriminant analysis with mixture discriminant analysis for both selection procedures.

MDA - Selection Procedure I			
Step	Light period	Dark period	First 3 hours
(1) Stimul	SWS2 IS	AW SWS1	AW SWS1
(3) Antipsy	PW SWS2 IS	AW PW SWS2	AW
(4) Antidep	AW SWS1 SWS2 IS RS	SWS1 SWS2	
(5) Hypno	AW SWS1 SWS2 IS RS		
MDA - Selection Procedure II			
Step	Light period	Dark period	First 3 hours
(1) Stimul	AW PW SWS2	PW	SWS1
(3) Antipsy	SWS2	AW PW SWS2	AW IS
(4) Antidep	IS RS	PW SWS2 IS	AW SWS1
(5) Hypno	PW SWS1 SWS2	RS	AW IS

Table 6: Mixture Discriminant Analysis. Adjusted posterior probabilities obtained when FPMM and MDA with Selection Procedure I (upper panel) and Selection Procedure II (lower panel) are applied.

MDA - Selection Procedure I ($error_{rat} = 0.020$ / $error_{c\text{-}d} = 0.000$)					
Drugclass	Placebo	Antipsy	Antidep	Hypnotic	Stimulant
Placebo	**0.99**	0.00	0.01	0.00	0.00
Antipsychotic	0.02	**0.95**	0.01	0.00	0.02
Antidepressant	0.00	0.04	**0.95**	0.01	0.00
Hypnotic	0.00	0.03	0.02	**0.95**	0.00
Stimulant	0.00	0.06	0.03	0.00	**0.91**
MDA - Selection Procedure II ($error_{rat} = 0.108$ / $error_{c\text{-}d} = 0.065$)					
Drugclass	Placebo	Antipsy	Antidep	Hypnotic	Stimulant
Placebo	**0.82**	0.07	0.05	0.06	0.00
Antipsychotic	0.17	**0.73**	0.04	0.03	0.04
Antidepressant	0.05	0.03	**0.89**	0.02	0.00
Hypnotic	0.04	0.01	0.01	**0.95**	0.00
Stimulant	0.08	0.11	0.06	0.00	**0.76**