

## Research Article

# The Classification of Valid and Invalid Beats of Three-Dimensional Nystagmus Eye Movement Signals Using Machine Learning Methods

Martti Juhola,<sup>1</sup> Heikki Aalto,<sup>2</sup> Henry Joutsijoki,<sup>1</sup> and Timo P. Hirvonen<sup>2</sup>

<sup>1</sup> Computer Science, School of Information Sciences, University of Tampere, 33014 Tampere, Finland

<sup>2</sup> Department of Otorhinolaryngology & Head and Neck Surgery, University of Helsinki and Helsinki University Central Hospital, HUS, 00029 Helsinki, Finland

Correspondence should be addressed to Martti Juhola; [martti.juhola@sis.uta.fi](mailto:martti.juhola@sis.uta.fi)

Received 30 June 2013; Revised 27 September 2013; Accepted 18 October 2013

Academic Editor: Christian Mayr

Copyright © 2013 Martti Juhola et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nystagmus recordings frequently include eye blinks, noise, or other corrupted segments that, with the exception of noise, cannot be dampened by filtering. We measured the spontaneous nystagmus of 107 otoneurological patients to form a training set for machine learning-based classifiers to assess and separate valid nystagmus beats from artefacts. Video-oculography was used to record three-dimensional nystagmus signals. Firstly, a procedure was implemented to accept or reject nystagmus beats according to the limits for nystagmus variables. Secondly, an expert perused all nystagmus beats manually. Thirdly, both the machine and the manual results were united to form the third variation of the training set for the machine learning-based classification. This improved accuracy results in classification; high accuracy values of up to 89% were obtained.

## 1. Introduction

Nystagmus is a repetitive, reflexive eye movement that may be congenital, induced physiologically by vestibular or optokinetic stimuli, or occur spontaneously in vestibular patients. It is formed by to-and-fro, saw tooth-like beats that can be recorded in the horizontal, vertical, and torsional directions. Older recording techniques, such as electrooculography (EOG), recorded only horizontal and vertical movements. Video-oculography (VOG) using two small video cameras, one for each eye, also enables the recording of torsional movements. A nystagmus beat contains a slow phase followed by a shorter fast phase that returns the eyes in the opposite direction (Figure 1). Nystagmus beats are repetitive, but their configuration changes often in the course of even a short measurement.

For the analysis of nystagmus signals, it is important to distinguish slow and fast phases and separate noisy or corrupted signal locations so that they do not impair values of the nystagmus variables to be computed. The slow phase characteristics are significant for the diagnostics of vestibular

neuritis, positional vertigo, acoustic neuroma, and Ménière's disease, and the fast phases returning the eyes to the centre are of central origin (the brain). Our present data consists of 107 otoneurological patients with spontaneous nystagmus, for whom the velocity and direction of the slow phase is the most important part.

Several detection methods have been published for nystagmus signal analysis since the 1970s [1–12]. One of the research details designed has been to detect nystagmus beats as accurately as possible. Most of the designs have required, more or less, the control of a user. Typically, they have been on the basis of the use of nystagmus direction and especially velocity, derived as an approximation of the first derivative of a positional nystagmus signal. The direction or sign of velocity changes at the beginning and end of a slow or fast phase. The direction of nystagmus beats, defined as the velocity of their fast phases, is normally constant but may infrequently change. The directions are left and right for horizontal, up and down for vertical, and clockwise and anticlockwise for torsional eye movements.

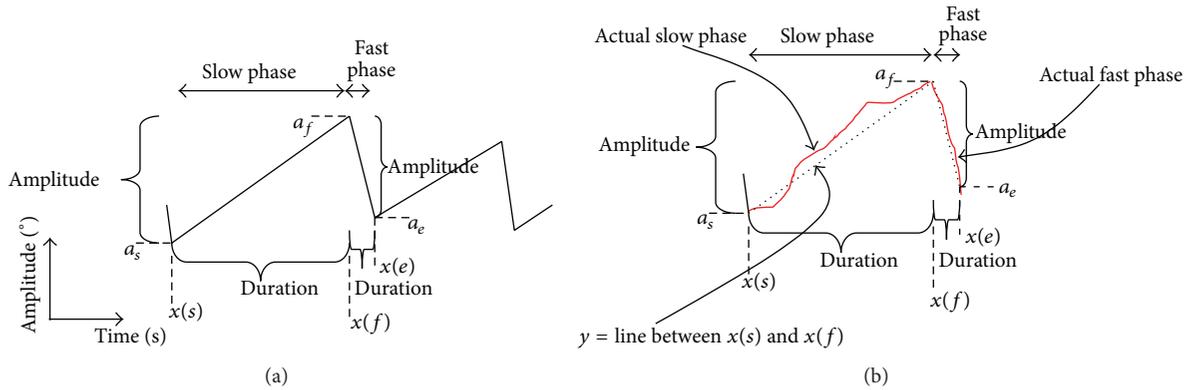


FIGURE 1: (a) A nystagmus beat from an ideal eye movement signal, including no noise or corruptions, contains a slow phase immediately followed by its shorter fast phase. Their amplitudes and durations are basic nystagmus variables. Durations are computed from the beginnings  $x(s)$  and  $x(f)$  of slow and fast phases and the end  $x(e)$ . Amplitudes are the differences from  $a_s$ ,  $a_f$ , and  $a_e$ . (b) The example shows a more realistic nystagmus beat including slightly less smooth nystagmus phases.

Filtering, thresholds, and the estimated functional relation between slow and fast phases of nystagmus beats (Figure 1) were applied to recognize the difference between slow and fast phases [1]. The authors in [1] also applied their functional broken line relation to detect rapid eye movements, saccades (as extraneous, voluntary movements, not fast phases of nystagmus), and eye blinks present in signals as strong peak artefacts, provided that these were in the same direction as the slow phases of a nystagmus signal. The detection of minimum and maximum locations within a window of preset width was applied to find the beginnings and ends of slow and fast phases [2]. The authors constructed a spike detection algorithm also usable for extraneous saccades, but since this computed the mean square error between eye movement and stimulation signals, such a procedure would not be possible for our spontaneous nystagmus without any stimulation signal. Furthermore, a procedure based on velocity thresholds was presented in which noise was dampened with low-pass filtering, but no specific artefact or saccade detection was introduced [3]. In some earlier studies, we applied filtering and syntactic recognition methods to both detecting nystagmus beats and eliminating extraneous saccades and artefacts at the same time [4, 5]. A method was introduced on the basis of filtering and wavelet processing to separate slow phases from fast phases [6], but problems caused by noise, artefacts, and saccades were not considered.

Optokinetic nystagmus was studied by using chaos theory [7] to detect nystagmus beats. The method was described as being tolerant of noise, but the noise tests were limited to sheer simulated Gaussian noise to the nystagmus signals. A study of congenital nystagmus in infants was performed in which waveforms of different nystagmus beats were analysed [8], but the procedure for the computer analysis was not described. Congenital nystagmus containing waveform analysis was also studied in [9] for both children and adults by applying time-frequency analysis with the fast Fourier transform. The influence of blinking artefacts was identified along with their narrow frequency band in the spectrum.

Wavelet analysis was used for nystagmus signals [10]. Several filters were used with a threshold to detect fast phases of nystagmus in electrooculographic signals [11]. The analysis of three-dimensional nystagmus signals has been described recently, for example, in [12].

Most previous studies using the EOG technique have applied one-dimensional (horizontal) or two-dimensional (horizontal and vertical) directions only, because the torsional direction can be recorded only with the magnetic scleral coil technique and by some VOG techniques. The earlier technique used was frequently EOG. The majority of nystagmus signal algorithms have considered the detection of nystagmus beats as a phenomenon, but not specifically the classification of valid beats from noisy, corrupted beats or “disinformation” without any actual beats. Recently, we introduced an algorithm for three-dimensional nystagmus signals [13] and also dealt with the preceding problem. In the present research, we concentrated on this classification topic in order to design a method based on machine learning. The idea was to utilize a dataset of accepted and rejected nystagmus beat candidates detected from nystagmus signals either manually or with a selection algorithm, and to enable reliable classification of nystagmus beats by combining this selection information collected in both manners.

We present the average values of accepted and rejected nystagmus beat candidates of several nystagmus variables to demonstrate how they differ in the dataset collected. In addition, we studied which variables best separate the accepted and the rejected nystagmus beats in the dataset measured. This is useful since some poor variables in this respect might be abandoned in future research if they do not provide useful information for medicine. Variable or feature evaluation and selection is an important phase in data analysis, as in, for example, [14, 15], because it is necessary to find which variables most affect classification and also those which are less influential. If there are a particularly large number of variables, variable selection is important. Sometimes leaving out poor variables may also improve classification results somewhat.

TABLE 1: Nystagmus variables defined with symbols given in Figure 1. In the variable names,  $sp$  and  $fp$  are slow and fast phases of nystagmus,  $a$  refers to amplitude,  $d$  duration,  $v$  mean velocity,  $c$  correlation,  $mv$  maximum velocity,  $q$  quality signal, and subscripts  $h$ ,  $v$ , and  $t$  indicate horizontal, vertical, and torsional components.

Variables	Name	Definition
$spa_h, spa_v, spa_t$	Amplitudes of slow phase	$ a_f - a_s $
$fpa_h, fpa_v, fpa_t$	Amplitudes of fast phase	$ a_e - a_f $
$spd_h$	Duration of slow phase	$(x(f) - x(s)) / fr$ ( $fr$ sampling frequency)
$fpd_h$	Duration of fast phase	$(x(e) - x(f)) / fr$ ( $fr$ sampling frequency)
$spv_h, spv_v, spv_t$	Mean velocity of slow phase	Slope computed with linear regression from samples of $\{x(s), x(s+1), \dots, x(f)\}$
$fpv_h, fpv_v, fpv_t$	Mean velocity of fast phase	Slope computed with linear regression from samples of $\{x(f), x(f+1), \dots, x(e)\}$ and $\{y(s), y(s+1), \dots, y(f)\}$ (see Figure 1(b))
$spc_h, spc_v, spc_t$	Correlations of slow phase	correlation coefficient between samples of $\{x(s), x(s+1), \dots, x(f)\}$ and $\{y(s), y(s+1), \dots, y(f)\}$ (Figure 1(b))
$fpmv_h$	Maximum velocity of fast phase	Slopes $z(i)$ with linear regression from samples of $\{x(i-2), x(i-1), x(i), x(i+1), x(i+2)\}$ and $\{y(i-2), y(i-1), y(i), y(i+1), y(i+2)\}$ , $i = f+2, \dots, e-2$ , and then by taking $\max_{i \in \{f+2, \dots, e-2\}} \{z(i)\}$ (Figure 1(b))
$q_t$	Mean of torsional quality values	Mean of quality values during slow phase of interval $[s, f]$ (Figure 1(b))

## 2. Eye Movement Data

We measured nystagmus signals from 107 otoneurological patients (mean age approximately 50 years) suffering mainly from acute, unilateral, peripheral loss of vestibular function (vestibular neuritis or having had surgery for acoustic neuroma).

An alert subject was seated in a fixed chair and instructed not to move his or her head during the measurement. At first, a calibration measurement was performed by asking a subject to look alternately at nine dots situated symmetrically on a wall in the visual field. Actual nystagmus measurements were run with the eyes covered in the dimmed, stationary laboratory by applying an eye movement tracking system of two video cameras, one for each eye (SensoMotoric Instruments, Berlin, Germany). The video camera systems applied to eye movement studies include a built-in image processing program to recognize the pupil in each image in order to measure horizontal and vertical eye movements. Using the angles of the iris between successive images, torsional eye movements can be computed. The appropriate measuring circumstances enabled the lack of gaze fixation that was important to obtain the best nystagmus eye movements. Each measurement took 30 seconds. This short duration was normally sufficient to include 20–80 nystagmus beats and was preferred to avoid subject fatigue. Nystagmus was spontaneous for some patients, while for other cases, rapid horizontal head shaking was used to generate head-shaking nystagmus. The sampling frequency was 50 Hz. The horizontal and vertical amplitude resolutions of the camera system were  $0.05^\circ$  and that of the torsional direction was  $0.1^\circ$ .

The eye movement tracking system gave several signals as its outputs. For each eye, the outputs utilised were three-dimensional signals of eye movements: horizontal, vertical,

and torsional. In addition, we employed the torsional quality signals of both eyes that included values from an interval of  $[0, 1]$ : the higher the value, the better the quality of torsional signal. The system estimated noise here, and noisy torsional signal segments could be detected with these quality signals. In the subsequent description, three-dimensional signals are given as three one-dimensional signals in order to express and visualize the nystagmus beats clearly.

## 3. Computation Methods

The method used for the detection of nystagmus beats followed our previous publication [13] with a minor extension and modification. The principle of the method was to apply angular velocity for the recognition of beginnings and ends of nystagmus beats and to reject nystagmus beat candidates that were clearly outliers or corrupted by noise. In the following section, we describe how a method based on classification was formed to improve the separation of valid and poor nystagmus beat candidates.

For the classification of nystagmus beat candidates, we used the nystagmus variables explained below (see also the precise definitions of the variables in Table 1). Please note that some of the variables are probably not interesting for medical diagnostic purposes. Nevertheless, they could be useful simply for classification. First, after the detection of a signal's nystagmus beat candidates, variable values were computed for every nystagmus beat candidate. The first three variables were the slow phase amplitudes of the horizontal, vertical, and torsional components.

Second, the duration of a horizontal slow phase was computed. These basic variables are illustrated in Figure 1(a). Third, the mean angular velocities of the slow phases of the three components were estimated by means of linear

regression between the locations of the beginning and end of a nystagmus beat candidate (Figure 1(b)). The slope given by linear regression is directly equal to the mean velocity between two locations. Fourth, the amplitudes of horizontal, vertical, and torsional components of the fast phases were calculated. Fifth, the duration of a horizontal fast phase was computed.

Sixth, the mean velocities of the three components of the fast phases were estimated using linear regression. Such a slope value given by linear regression directly estimates the first derivative, that is, angular velocity. It is good to remember that not all nystagmus signal types are as close to linear as the slow and fast phases of vestibular nystagmus. Thus, for other nystagmus types, the variables used should be modified.

Seventh, mean torsional quality during a slow phase was computed. Eighth, correlation between an “ideal slow phase” and that of an actual signal part was estimated. An ideal slow phase corresponds to that illustrated in Figure 1(b) where there is a line between the beginning and end of the slow phase of a nystagmus beat. In an actual slow phase, there is some noise or other nonlinearity between those locations in the slow phase. This deviation was estimated by computing a Pearson correlation coefficient between the ideal and actual slow phases. The nearer the (absolute) correlation coefficient to 1, the better the slow phase, and the nearer to 0, the poorer the slow phase.

Ninth, the maximum velocity of a horizontal fast phase was computed using successive segments of 5 samples through a fast phase. The horizontal component was emphasized compared with the others, since frequently (at least in the present data) it is dominant compared with the other components: its amplitude and velocity values are greater. In total, there were 19 variables for the classification of nystagmus candidates.

To form training and test data for classification, we used all nystagmus beat candidates found by the recognition program. We sorted nystagmus candidates to be either valid or invalid at first manually and secondly with the procedure given in the following section. Manual selection was executed independently of the automatic selection. At first, the better eye signal from those of the left and right eye was chosen by computing the mean of torsional quality values through the whole signal and selecting the eye with a higher mean. In Figures 2–4 there are 10-second long examples in which the eye movement signal used (left or right) is also given.

**3.1. Selection Procedure for Nystagmus Candidates.** The following selection criteria were computed for every nystagmus beat candidate to label them either accepted or rejected.

- (1) The mean of torsional quality values were computed for successive signal segments of 2 seconds including at least one [13] beat candidate. If the mean torsional quality of such a segment was low, below 0.2–0.4 depending on the mean of the whole signal, that segment was rejected. Usually such a segment included abundant spikes, such as a few probable dropouts from the video images in Figure 4. Torsional quality

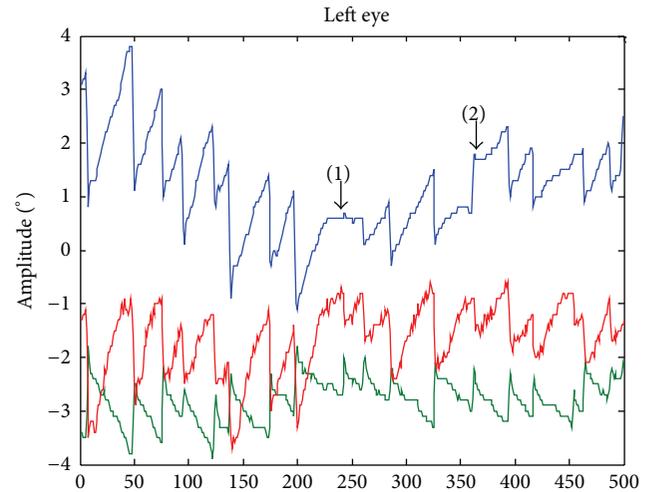


FIGURE 2: Abnormally slow (1) and fast (2) segments of horizontal (blue or uppermost) slow phases compared with the surrounding valid nystagmus beats.

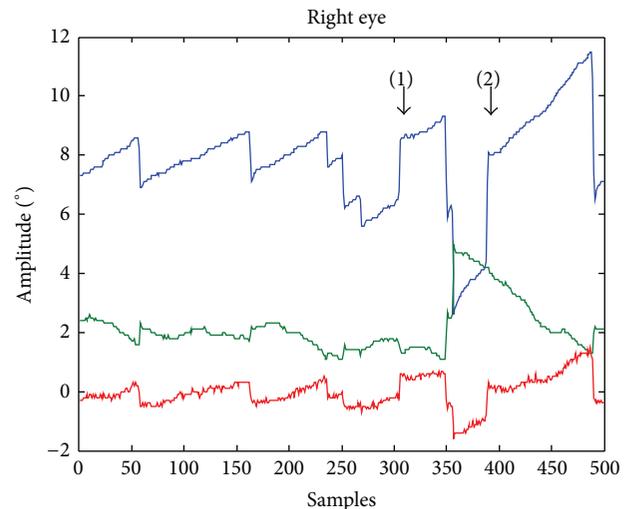


FIGURE 3: Abrupt upward “steps” (1) and (2) in the horizontal (blue or uppermost) and torsional (red or lowest) signals are obviously rapid voluntary eye movements (saccades) which would distort the slow phase velocities of two nystagmus beats.

values given by the recording system are in the range  $[0, 1]$ .

- (2) A lower lb and upper bound ub were computed for mean slow phase velocities of all nystagmus candidates according to their means  $\mu_i$  and standard deviations  $\sigma_i$  for the three components,  $i = h, v, t$  (horizontal, vertical, and torsional):

$$\begin{aligned} \text{lb} &= \mu_i - 1.96\sigma_i, \\ \text{ub} &= \max\{\mu_i + 1.96\sigma_i, \text{spv}_{\max}\}. \end{aligned} \quad (1)$$

In principle, this would prune 2.5% of nystagmus beat candidates from each end of the distribution if the

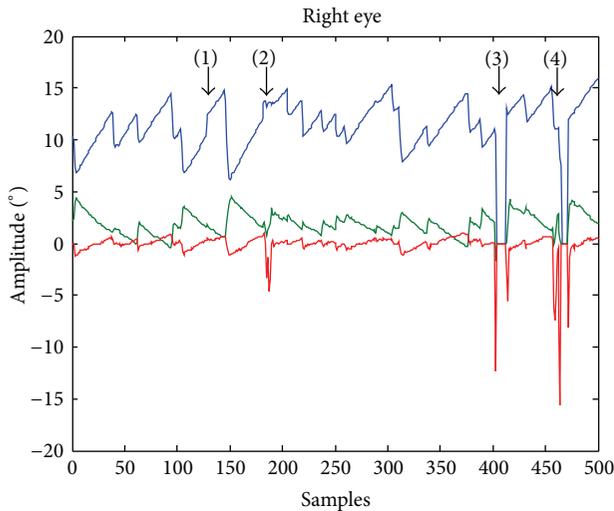


FIGURE 4: Deflection (1) could be a small saccade, (2) is noise, and (3) and (4) could be caused by short periods when the image recognition procedure of the eye movement camera lost the image of the pupil. Distortion (1) only occurred in the horizontal (blue or uppermost) signal, but (2), (3), and (4) were also present in the vertical (green or middle) and torsional (red or lowest) signals.

distribution were normal. In addition, an absolute maximum of  $spv_{\max} = 75^\circ/\text{s}$  was used to leave out extraordinarily high mean velocities that were probably high peaks, for example, eye blinks. This maximum was chosen on the basis of the data used. Note that a nystagmus beat is rejected if even one of the three component values is unacceptable. This condition could affect the rejections of locations (2), (3), and (4) in Figure 4.

- (3) Any occasional beat in the opposite direction from beats before it or after it was rejected. Infrequently, the direction of nystagmus may change during the recording period of a nystagmus signal. The “correct” direction was then decided by the majority of candidates, and those in the opposite direction were rejected.
- (4) Nystagmus candidates with higher torsional amplitudes than  $a_t = 15^\circ/\text{s}$  were rejected since this was seen as the physiological limit for torsional rotations of the eye. This could yield rejections because of the torsional spikes (4) in Figure 4. These spikes were not actual eye movements, but disturbances of the eye movement tracking system. Obviously, the video camera system has failed to analyse the images for a short period, resulting in video dropouts.
- (5) Nystagmus candidates with shorter durations of slow phases than  $spd_{\min} = 0.080 \text{ s}$  (4 sampling intervals for 50 Hz) were rejected as improbable nystagmus beats. This lower bound was found experimentally earlier [16].
- (6) The fast phase of a nystagmus beat immediately follows its slow phase to turn the eyes in the opposite

direction. For values of the horizontal fast phase, the velocity maximum  $fpv_{\max} = 400^\circ/\text{s}$  estimated earlier [16] was used to reject the corresponding nystagmus candidate. Usually noise spikes are steep, generating high velocity values higher than  $fpv_{\max}$ , and are not related to actual eye movements.

- (7) Noisiness during a slow phase segment was assessed by applying the correlation coefficient as described above. If a horizontal, vertical, or torsional component was less than the bound  $spvc_{\min} = 0.5$  derived experimentally, the nystagmus candidate was omitted. This cleaning often discarded candidates such as a “plain” (1) or “steps” (2) in Figure 2, or (1) and (2) in Figure 3 from among valid nystagmus beats. In addition, horizontal deflection (1) in Figure 4 could be identified on the basis of the present condition.

In addition to the manual and automatic selections of nystagmus beat candidates, we combined the above mentioned to study whether this would improve later classification results. Thus, in the third mode the accepted nystagmus beats were formed taking those that were accepted by both manual and automatic selections. Those that were rejected by either or both were considered rejected.

The entire data collected seemed to be rather overlapping, subject to variable values of the accepted and rejected nystagmus candidates. In other words, some values of rejected cases were close to those accepted by the same variable (see means and standard deviations given in Table 2). This is dealt with in more detail in the following section.

Properties (variable values) of valid nystagmus beats can vary considerably between subjects. Even within an individual signal, nystagmus beats may change in the course of a short recording time. The frequency of nystagmus beats (number of beats per second) may vary considerably between subjects. Consequently, relatively rapid mean velocity values of slow phases of one subject might be rejected for another subject that had much smaller mean velocity values. Thus, we designed a straightforward cleaning procedure to delete some rejected nystagmus candidates from the dataset to be used because there were more rejected than accepted candidates. Moreover, this action also equalized their numbers. This can sometimes be useful for some classification algorithms, as the two classes of a classification task are of roughly equal size. First, we computed the centres of both classes (the accepted and rejected) in the 19-dimensional Euclidean variable space. Then, we discarded, one by one, those rejected nystagmus candidates that were the closest to the centre of the accepted candidates. This procedure slowly cleaned out rejected nystagmus candidates between the class centres by reducing the sphere of influence around the class centre of the rejected candidates and those not yet rejected. In principle, this separated the two classes slightly from each other. Deletions were made until the classes were of equal size. The cleaning procedure was carried out for all three selection modes: manual, automatic, and both. Since the three selection modes were also performed without a cleaning procedure, there were six different test setups altogether.

TABLE 2: Manually, automatically, or both manually and automatically accepted (/rejected) nystagmus beats: means and standard deviations. In the variable names, *sp* and *fp* are slow and fast phases of nystagmus, *a* refers to amplitude, *d* duration, *v* mean velocity, *c* correlation, *mv* maximum velocity, *q* quality signal, and subscripts *h*, *v*, and *t* indicate horizontal, vertical, and torsional components.

	Manual		Automatic		Both	One, both
	Acceptance	Rejection	Acceptance	Rejection	Acceptance	Rejection
Numbers	2171	3818	2517	3472	1645	4344
Variables						
$spa_h$ (°)	2.8 ± 3	2.1 ± 4.6	3 ± 4.8	1.9 ± 4.8	3.1 ± 2.4	2.1 ± 4.6
$spa_v$ (°)	1.5 ± 2	1.6 ± 4	1.5 ± 1.9	1.6 ± 4.2	1.4 ± 1.3	1.6 ± 3.9
$spa_t$ (°)	1.9 ± 2.2	2.8 ± 5.6	1.9 ± 2.1	2.9 ± 5.8	2 ± 2	2.7 ± 5.3
$spd_h$ (s)	0.55 ± 0.45	0.34 ± 0.49	0.54 ± 0.47	0.33 ± 0.48	0.55 ± 0.43	0.37 ± 0.5
$spv_h$ (°/s)	6.5 ± 6	7.6 ± 27.8	7.1 ± 6.7	7.4 ± 28.9	7.4 ± 6	7.2 ± 26.1
$spv_v$ (°/s)	4 ± 6.6	6.2 ± 21.8	3.8 ± 5.8	6.5 ± 22.9	3.8 ± 5.2	6 ± 20.7
$spv_t$ (°/s)	6.2 ± 11.9	16.5 ± 50.2	5.7 ± 9.2	17.9 ± 52.8	5.9 ± 9.5	15.3 ± 47.6
$fpa_h$ (°)	3 ± 3.1	2.2 ± 5.1	3 ± 3.1	2 ± 5.2	3.2 ± 2.7	2.2 ± 5
$fpa_v$ (°)	1.4 ± 2.3	1.6 ± 4.3	1.5 ± 2.4	1.6 ± 4.3	1.4 ± 1.8	1.6 ± 4.2
$fpa_t$ (°)	1.8 ± 2.2	2.9 ± 5.6	1.9 ± 2.6	2.9 ± 5.7	1.9 ± 2.1	2.7 ± 5.4
$fpd_h$ (s)	0.07 ± 0.04	0.08 ± 0.08	0.07 ± 0.05	0.08 ± 0.08	0.07 ± 0.04	0.08 ± 0.07
$fpv_h$ (°/s)	52.2 ± 60.5	35 ± 107.3	51.9 ± 57.2	33.4 ± 112.1	56.3 ± 44.8	35.5 ± 105.7
$fpv_v$ (°/s)	23.8 ± 50.1	26.4 ± 96.3	23.3 ± 40.1	27 ± 103	22.8 ± 25.5	26.5 ± 95.7
$fpv_t$ (°/s)	31.6 ± 42.1	53.5 ± 148.9	33.3 ± 54.9	54.5 ± 152.6	33.1 ± 38.1	50.3 ± 141
$spc_h$	0.94 ± 0.16	0.65 ± 0.38	0.94 ± 0.1	0.61 ± 0.4	0.97 ± 0.07	0.67 ± 0.38
$spc_v$	0.88 ± 0.21	0.62 ± 0.38	0.9 ± 0.14	0.58 ± 0.39	0.93 ± 0.11	0.63 ± 0.37
$spc_t$	0.81 ± 0.26	0.53 ± 0.35	0.85 ± 0.16	0.47 ± 0.36	0.89 ± 0.14	0.53 ± 0.35
$fpmv_h$ (°/s)	41.3 ± 41	33.6 ± 71	43.8 ± 45	31 ± 71.3	44.4 ± 37.8	33.3 ± 68.7
$q_t$	0.78 ± 0.1	0.64 ± 0.2	0.76 ± 0.11	0.64 ± 0.2	0.78 ± 0.1	0.66 ± 0.19

When the 19 variables applied were from quite different scales, for instance, fast phase durations as small as 0.04 s and fast phase velocities of hundreds of °/s-scaling were useful for some classification algorithms. We standardized (normalized) the data (subtracting the mean from a variable value and then dividing with the standard deviation), variable by variable, in order to set all variables to the same scale. The classification methods used were run with and without standardization.

We deployed several classification algorithms: linear, quadratic, and logistic discriminant analysis, *k*-nearest neighbour searching with different numbers of *k*, the naïve Bayes rule, multilayer perceptron networks, and support vector machines. We also experimented with *k*-means clustering, but as this did not converge, we dropped it from further tests.

#### 4. Results

To begin with, we computed the means and standard deviations of the 19 nystagmus variables listed above in order to study how they differed between the accepted and rejected classes of nystagmus beat candidates. These average results are shown in Table 2. Please note that durations were computed for horizontal components only, because virtually they do not vary between the three components, and the quality signal given by the video camera system exists only for torsional signals.

In order to accept or reject, we ran three different alternatives, which were (1) manually accepted or rejected, (2) automatically accepted or rejected by the preceding selection procedure, or (3) accepted both manually and automatically or rejected by one or both of the two procedures. We were interested to see which of the first two procedures was better and whether using both could improve results in classification tests. Manual selections were first made by an experienced person accustomed to the assessment of nystagmus signals, but who did not know the decisions of automatic selections.

The signals of 107 patients included 5,989 nystagmus beat candidates. On the basis of Table 2, the means of most variables undoubtedly differed between the accepted and rejected alternatives. On the other hand, standard deviations were so large that distributions of individual variables overlapped notably. We performed an unpaired two-sampled *t*-test for every variable and for both alternatives. For  $\alpha = 0.001$ , excluding the vertical amplitude of slow phase  $spa_v$ , the horizontal mean velocity of slow phase  $spv_h$ , the vertical amplitude of fast phase  $fpa_v$ , and the vertical mean velocity of fast phase  $fpv_v$ , the other 15 variables differed statistically significantly between two alternatives. For  $\alpha = 0.05$ , even  $fpa_v$  differed fairly significantly. These outcomes indicated that the 15 variables were certainly useful for later classification tests, and the other four were possibly useful. Although there were only small differences between the means of these four variables, the standard deviations of the rejected nystagmus candidates were high, but relatively small for the accepted

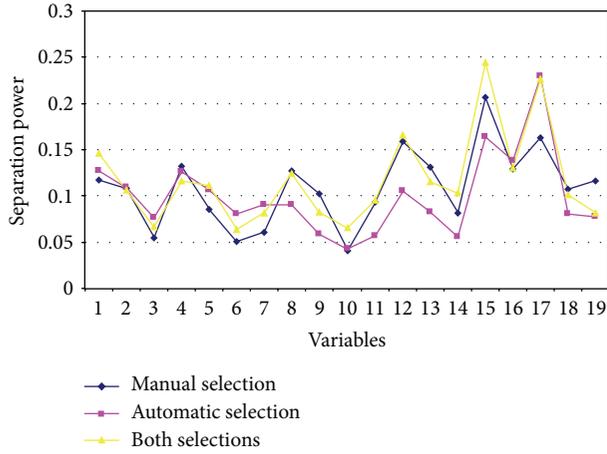


FIGURE 5: Separation power values given by the Scatter method. The higher the value of a variable, the better the classification. Variables (1)–(19) are in the same order as in Tables 1–3:  $spa_h$ ,  $spa_v$ ,  $spa_t$ ,  $spd_h$ ,  $spv_h$ ,  $spv_v$ ,  $spv_t$ ,  $fpa_h$ ,  $fpa_v$ ,  $fpa_t$ ,  $fpd_h$ ,  $fpv_h$ ,  $fpv_v$ ,  $fpv_t$ ,  $spc_h$ ,  $spc_v$ ,  $spc_t$ ,  $fpmv_h$ , and  $q_t$ .

ones. The underlying property seemed to be that the subset of the accepted nystagmus candidates was faintly compact, but the subset of the rejected dispersed strongly.

In order to study the importance of the 19 variables for classification, we applied our Scatter method [17, 18] which runs a nearest neighbour search in a dataset from a sample to its nearest neighbour sample according to Euclidean distance throughout the whole set and counts how many changes between classes (the accepted and rejected) have occurred. The fewer class changes that are encountered, the more compact the classes; that is, they are probably more separable. The more class changes that are detected, the worse the separable the classes. The Scatter method gave us values called separation powers: the larger the value, the more separable the classes. Figure 5 includes the results for the accepted nystagmus beats only because those of the rejected were very similar. Calculating the average of the separation powers of the three different selections used for each variable shows that the most important variables with the highest separation powers for classification were, in descending order, (17)  $spc_t$ , (15)  $spc_h$ , (12)  $fpv_h$ , (16)  $spc_v$ , (1)  $spa_h$ , (4)  $spd_h$ , and (8)  $fpa_h$ .

Correspondingly, the least important were, in ascending order, (10)  $fpa_t$ , (6)  $spv_v$ , (3)  $spa_t$ , (7)  $spv_t$ , (14)  $fpv_t$ , (9)  $fpa_v$ , (11)  $fpd_h$ , (19)  $q_t$ , (18)  $fpmv_h$ , (5)  $spv_h$ , (2)  $spa_v$ , and (13)  $fpv_v$ . Among the poorest there were (3), (9), (2), and (13); that is, those four given by the  $t$  test to have statistically insignificant differences ((9) with  $\alpha = 0.01$  only). Since separation powers between the best variable (17) and the worst variable (10) did not differ greatly, and that of (10) as the worst of all was not relatively close to 0 in comparison to the other variables, we kept all variables for further analysis.

As explained in the preceding section, we executed all tests with three selection modes: manual, automatic, and both. Each of these was performed either with or without the cleaning of rejected nystagmus candidates as possible outliers. Furthermore, six combinations were run either with

or without standardization of the whole dataset, variable by variable, as the preprocessing stage before classification. In addition to linear, quadratic, and logistic discriminant analysis, we exploited  $k$ -nearest neighbour searching with  $k$  equal to 1, 3, 5, 11, 15, 21, 25, 45, and 65, and the naïve Bayes rule classification. We ran the same tests with multilayer perceptron networks with 19 inputs and 6 hidden and 2 output nodes and used the Levenberg-Marquardt training algorithm.

Finally, we tested with support vector machines by applying three different kernel functions as reported in Table 3 in order to search for the best regulation parameter values. For the linear and quadratic kernels, 20 box constraint values from the set of  $\{0.5, 1.0, 1.5, \dots, 10\}$  were employed. For the radial basis kernel,  $20 \times 20 = 400$  pairs of parameter values were explored as Cartesian products of box constraint and standard deviation  $\sigma$  each from  $\{0.5, 1.0, 1.5, \dots, 10\}$ . These results were computed without standardization of the variable values.

Linear, quadratic, and logistic discriminant analysis and the naïve Bayes rule are insensitive to data standardization giving identical values independent of whether standardization is made or not. Meanwhile,  $k$ -nearest neighbour searching was sensitive to data standardization. Thus, the results of the better alternative of each classification method are only given in Figure 6. Since the runs with  $k$  equal to 5, 11 and 15 generated results a few per cent better than those of smaller  $k$  and around 1% better than those of greater  $k$  values, results of  $k$  equal to 15 are only depicted in Figure 6. As usual, accuracies in percentages were computed as the ratio between the sum of true positive and true negative classification outcomes and the number of all  $n$  tested cases.

We applied the leave-one-out method for training and testing, in which  $n - 1$  cases were used for a training set and 1 as the only test case, and then we repeated this for each case. Therefore, there were 5,889 tests for dataset variations (1), (3), and (5), and after cleaning 4,342 for (2), 5,034 for (4), and 3,290 for (6) in Figure 6. For (2), (4), and (6), the originally larger class of the rejected nystagmus candidates was cleaned (balanced) to be of equal size according to the smaller class of the accepted nystagmus candidates.

The results in Figure 6 show how cleaning improved the accuracies of the linear and quadratic discriminant analysis methods by 2%–7% but strongly decreased those of the other methods, even from 80%–90% down to 40% of correct classifications. Results of the support vector machines are shown in Figure 7. Cleaning improved the results that were generally among the best. The quadratic kernel was the best of the three kernels.

It was a surprising result that data cleaning impaired the results of other methods so much other than linear and quadratic discriminant analysis and support vector machines. We therefore studied the data in more detail. We calculated the means and the standard deviations of the reduced sets of the rejected nystagmus candidates after cleaning the data according to the three methods described above. These results are presented in Table 4. Comparing the values of its three columns to those corresponding to Table 2, it can be seen that there are differences in the means among the rejected

TABLE 3: The parameter values for the kernel functions that produced the best classification results with support vector machines.

	Linear kernel: box constraint	Quadratic kernel: box constraint	Radial basis function: box constraint and $\sigma$	
Manual selection	0.5	9.5	2.5	0.5
Manual selection with cleaning	3.5	0.5	3.0	10.0
Automatic selection	7.5	3.5	8.5	0.5
Automatic selection with cleaning	2.5	8.0	6.5	10.0
Both manual and automatic selections	1.5	1.0	0.5	0.5
Both selections with cleaning	1.0	0.5	7.5	10.0

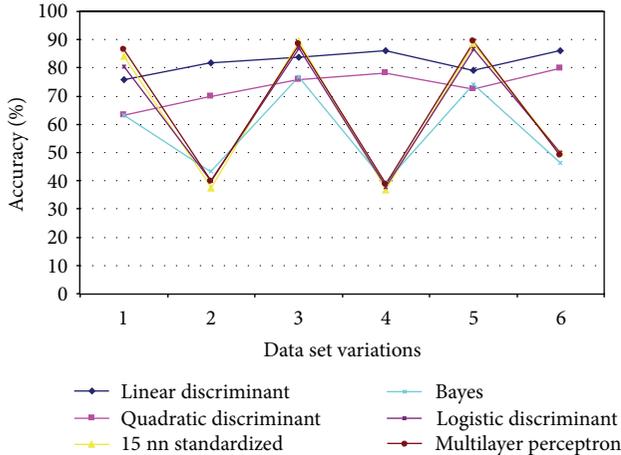


FIGURE 6: Classification results of the different methods used (nn is the nearest neighbour) as accuracies (as percentages) with dataset variations: (1) manual selection and no cleaning of rejected nystagmus candidates, (2) manual selection and with cleaning of rejected candidates, (3) automatic selection and no cleaning, (4) automatic selection with cleaning, (5) both manual and automatic selection and no cleaning, and (6) both manual and automatic selection with cleaning.

nystagmus beat candidates, so that the differences between the rejected nystagmus beat candidates in Table 4 and the accepted nystagmus beat candidates in Table 2 increased. In spite of cleaning, the standard deviations of the rejected still remained large in Table 4. Therefore, the three ways to apply the cleaned datasets did not alleviate the problem of considerably overlapping classes of the accepted and rejected nystagmus beat candidates caused by their large standard deviations. At the same time, differences between the means of both classes (their centres) increased by 44%, 22%, and 56% for the three data cleaning methods. Despite separating the set of the rejected from the accepted, the large standard deviation showed the rather similar overlap that was present before cleaning.

We nevertheless computed a principal component analysis and present its visualization results in Figure 8 for the situations of manual selection with no cleaning (1) and with cleaning (4) (from Figure 6). Figure 8 shows that the overlap between the classes of the accepted and rejected is still rather substantial after cleaning; in other words, cleaning

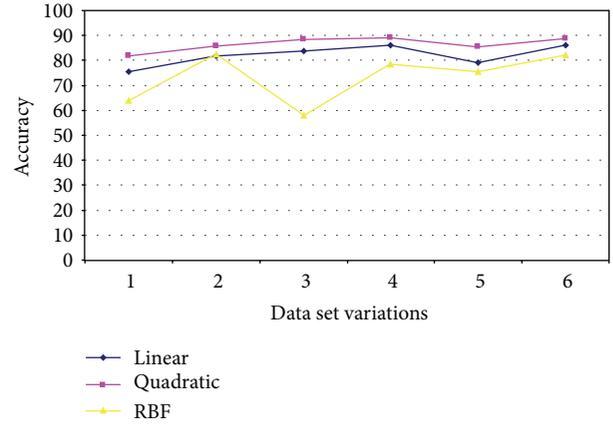


FIGURE 7: Classification results with support vector machines after using the best parameter selections from Table 3 for the linear, quadratic, and radial basis function (RBF) kernels with dataset variations: (1) manual selection and no cleaning of rejected nystagmus candidates, (2) manual selection and with cleaning of rejected candidates, (3) automatic selection and no cleaning, (4) automatic selection with cleaning, (5) both manual and automatic selection and no cleaning, and (6) both manual and automatic selection with cleaning.

did not change the data distribution to more favourable in classification between the accepted and rejected nystagmus beat candidates.

## 5. Discussion and Conclusions

We found useful information about nystagmus variables analysed here for the purpose of nystagmus recognition in respect to the separation of valid and invalid nystagmus beats. On the basis of Figure 5, supported by the results in Table 2, we found that variables (17)  $spc_t$ , (15)  $spc_h$ , (12)  $fpv_h$ , (16)  $spc_v$ , (1)  $spa_h$ , (4)  $spd_h$ , and (8)  $fpa_h$  were the best to separate valid and poor nystagmus beats in the present dataset. In future research, it is more preferable to concentrate on these variables and perhaps abandon the other analysed variables if not important for medical diagnostics.

Based on the results in Figure 6, we found out how the classifications with nearest  $k$ -neighbour searching, the naïve Bayes rule, logistic discriminant, and multilayer perceptron networks were very sensitive to cleaning in the current data.

TABLE 4: After cleaning, manually, automatically, or using both manually and automatically rejected nystagmus beats kept in the data set: means and standard deviations. In the variable names  $sp$  and  $fp$  are slow and fast phases,  $a$  refers to amplitude,  $d$  duration,  $v$  mean velocity,  $c$  correlation,  $mv$  maximum velocity,  $q$  quality signal, and subscripts  $h$ ,  $v$ , and  $t$  indicate horizontal, vertical, and torsional components.

	Manual	Automatic	Manual or automatic
Numbers	2171	2517	1645
Variables			
$spa_h$ ( $^\circ$ )	$1.8 \pm 4.9$	$1.9 \pm 5.2$	$2.1 \pm 5.7$
$spa_v$ ( $^\circ$ )	$1.7 \pm 4.6$	$1.7 \pm 4.7$	$2.1 \pm 5.4$
$spa_t$ ( $^\circ$ )	$3.2 \pm 6.3$	$3.1 \pm 6.1$	$3.4 \pm 6.5$
$spd_h$ (s)	$0.29 \pm 0.50$	$0.30 \pm 0.47$	$0.33 \pm 0.52$
$spv_h$ ( $^\circ/s$ )	$7.9 \pm 33.0$	$7.5 \pm 31.3$	$8.6 \pm 35.4$
$spv_v$ ( $^\circ/s$ )	$7.2 \pm 25.7$	$7.1 \pm 24.5$	$8.4 \pm 28.9$
$spv_t$ ( $^\circ/s$ )	$21.0 \pm 61.3$	$20.2 \pm 58.3$	$22.1 \pm 64.3$
$fpa_h$ ( $^\circ$ )	$2.0 \pm 5.7$	$2.0 \pm 5.8$	$2.3 \pm 6.6$
$fpa_v$ ( $^\circ$ )	$1.9 \pm 4.9$	$1.7 \pm 4.8$	$2.1 \pm 5.6$
$fpa_t$ ( $^\circ$ )	$3.2 \pm 6.2$	$3.1 \pm 6.0$	$3.4 \pm 6.4$
$fpd_h$ (s)	$0.08 \pm 0.08$	$0.08 \pm 0.08$	$0.08 \pm 0.08$
$fpv_h$ ( $^\circ/s$ )	$27.4 \pm 78.8$	$30.8 \pm 114.9$	$32.7 \pm 97.7$
$fpv_v$ ( $^\circ/s$ )	$27.2 \pm 109.0$	$30.5 \pm 117.5$	$33.5 \pm 124.2$
$fpv_t$ ( $^\circ/s$ )	$63.5 \pm 173.7$	$61.8 \pm 168.5$	$69.2 \pm 182.0$
$spc_h$	$0.63 \pm 0.39$	$0.60 \pm 0.40$	$0.65 \pm 0.38$
$spc_v$	$0.60 \pm 0.39$	$0.58 \pm 0.40$	$0.61 \pm 0.39$
$spc_t$	$0.53 \pm 0.35$	$0.48 \pm 0.36$	$0.54 \pm 0.35$
$fpmv_h$ ( $^\circ/s$ )	$31.5 \pm 80.3$	$31.4 \pm 79.6$	$36.3 \pm 91.2$
$q_t$	$0.61 \pm 0.18$	$0.62 \pm 0.19$	$0.62 \pm 0.19$

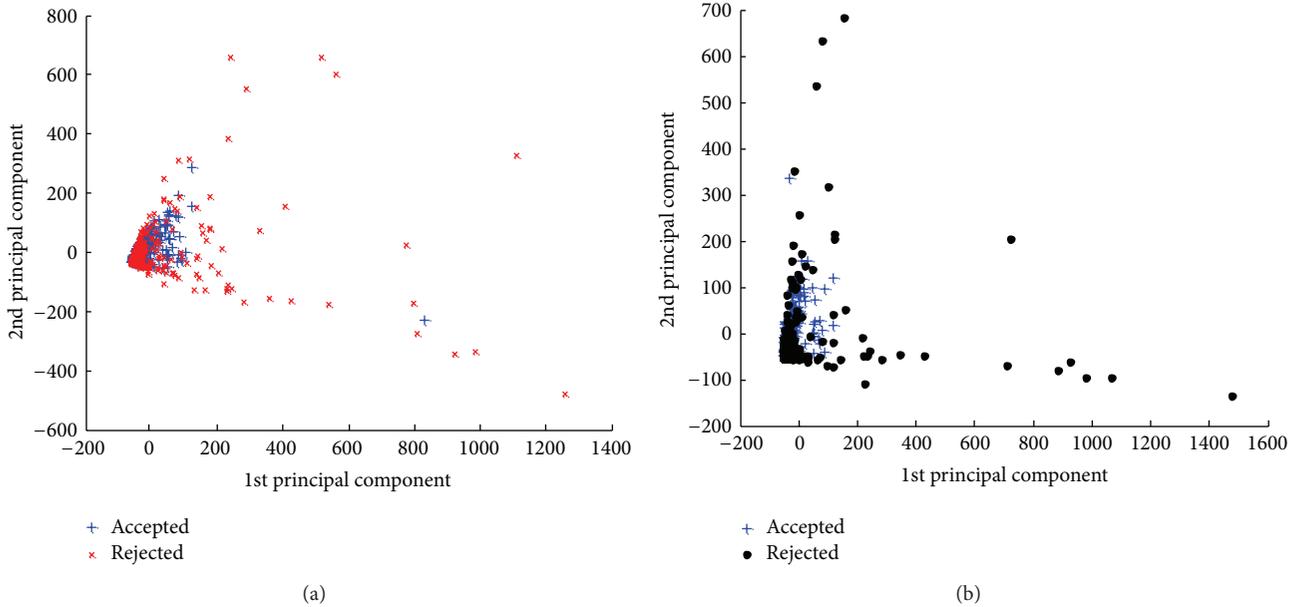


FIGURE 8: The two greatest principal components were used to map how accepted and rejected nystagmus candidates by manual selection were distributed before (a) and after (b) cleaning (situations (1) and (2) from Figure 6). No more than one tenth of nystagmus candidates are mapped to make the figures clear, since the rejected and accepted candidates overlap and the rejected candidates are strongly concentrated in the “left corner” of the distribution. The two principal components corresponded to (a) 80.0% and (b) 82.1% from variance in the data.

This type of cleaning was definitely disadvantageous, and although it balanced the two classes, it did not affect their distribution so that the classification task was more successful. Instead, for linear and quadratic discriminant analysis and support vector machines, it gave accuracies a few percent better than no cleaning. Linear discriminant analysis was 6%–12% better for all situations than quadratic discriminant analysis. Nevertheless, the support vector machines using the quadratic kernel were among the best. When not cleaned, the standardized nearest neighbour searching was 6%–16% better than the nonstandardized one. The nearest neighbour searching with  $k$  equal to 5, 11, and 15 gave virtually the same accuracies as those tested above 15. The naïve Bayes rule was mostly the poorest choice except for those cleaned.

The best accuracies of 88%–89% were gained with standardized nearest neighbour searching with  $k$  equal to 5, 11 and 15 or above 15 and without cleaning using automatic or manual and automatic selections. Multilayer perceptron networks and support vector machines gave accuracies virtually as good as those of the nearest neighbour searching. However, the radial basis function kernel of the support vector machines did not succeed in the class of the accepted for the three selections without cleaning but almost completely failed with them, although the class of the rejected was classified very well. Thus, this kernel could not be recommended for the current data without cleaning. It is natural that support vector machines benefitted from cleaning, since their principle is based on searching for the maximally separating boundary between two classes. Linear discriminant analysis also gave high accuracies of 86% provided that automatic selection or manual and automatic selections with cleaning were included. Logistic discriminant analysis favoured no cleaning, yielding accuracies of 87% for automatic or manual and automatic selections.

Calculating pairwise between manual selection and automatic selection through all methods and with or without cleaning, automatic selection was approximately 3% superior to manual selection on average. Correspondingly, between automatic selection and jointly manual and automatic selections, the latter choice was approximately 4% superior to the former on average.

To build a training set, automatic or preferably both manual and automatic selection modes are recommended, since for most methods these gave better results than the manual selection. Among the classification methods, nearest neighbour searching with data standardization and no cleaning gave the best results. The support vector machines applied with the quadratic kernel function were also among the best. Logistic discriminant analysis without cleaning and linear discriminant analysis with cleaning were almost equally effective.

To build a good training set, joint manual and automatic selection seem to be useful, because an experienced expert can recognize nystagmus candidates corrupted in some exceptional way that the automatic selection might accept. For example, in Table 2 there were more manually than automatically rejected beats. Overall, the latter obviously gave slightly better results because the fixed lower and upper bounds of the variable values created stricter boundaries

in a variable space than the subjective (somewhat varying) manual selection.

A training set created as the dataset in the present research can be used, together with machine learning methods, to separate valid nystagmus beats from those corrupted in various ways. This could aid the current difficulties considered in various research articles since the 1970s. However, such a dataset is probably not generally suitable for nystagmus signals, since values of nystagmus variables may vary slightly between measurement systems and sampling frequencies. For instance, the mean velocities of slow or fast phases might vary slightly if they were generally steeper for one system than another. On the other hand, the advantage of many machine learning methods, for example, multilayer perceptron networks and support vector machines, is that they can adaptively function with varying data. Therefore, this problem may, after all, be minor.

In the future we could test some other classification methods and more useful cleaning procedures. However, it may be difficult to develop cleaning because of the complicated distribution of the dataset. An advantage of the tested classification methods was their fast running times, with the exception of logistic discriminant analysis and multilayer perceptron networks. The duration of building and testing 30333 models (3.0 GHz dual CPU) in the tests executed was 21 h 40 min for logistic discriminant analysis and even longer for multilayer perceptron networks and support vector machines with MATLAB. In comparison, the other methods required less than 10 minutes. However, such execution times would be meaningless in routine use, because one model would only be needed and this could be built in advance if a rather stable training set was applied.

## References

- [1] J. H. J. Allum, J. R. Tole, and A. D. Weiss, "MITNYS II. A digital program for on line analysis of nystagmus," *IEEE Transactions on Biomedical Engineering*, vol. 22, no. 3, pp. 196–202, 1975.
- [2] C. Wall III and F. O. Black, "Algorithms for the clinical analysis of nystagmus eye movements," *IEEE Transactions on Biomedical Engineering*, vol. 28, no. 9, pp. 638–646, 1981.
- [3] G. R. Barnes, "A procedure for the analysis of nystagmus and other eye movements," *Aviation Space and Environmental Medicine*, vol. 53, no. 7, pp. 676–682, 1982.
- [4] M. Juhola, "Detection of nystagmus eye movements using a recursive digital filter," *IEEE Transactions on Biomedical Engineering*, vol. 35, no. 5, pp. 389–395, 1988.
- [5] M. Juhola, "A syntactic method for analysis of nystagmus and smooth pursuit eye movements," *Computer Methods and Programs in Biomedicine*, vol. 26, no. 1, pp. 53–62, 1988.
- [6] P. Augustyniak, "Advanced method of Nystagmus-phase separation using adaptive modification of time-frequency signal representation," in *Proceedings of the 3rd International Conference on Signal Processing (ICSP '96)*, pp. 351–354, October 1996.
- [7] T. Aasen, D. Kugiumtzis, and S. H. G. Nordahl, "Procedure for estimating the correlation dimension of optokinetic nystagmus signals," *Computers and Biomedical Research*, vol. 30, no. 2, pp. 95–116, 1997.

- [8] R. W. Hertle and L. F. Dell'Osso, "Clinical and ocular motor analysis of congenital nystagmus in infancy," *Journal of AAPOS*, vol. 3, no. 2, pp. 70–79, 1999.
- [9] M. Hosokawa, S. Hasebe, H. Ohtsuki, and Y. Tsuchida, "Time-frequency analysis of electronystagmogram signals in patients with congenital nystagmus," *Japanese Journal of Ophthalmology*, vol. 48, no. 3, pp. 262–267, 2004.
- [10] L. A. Abel, Z. I. Wang, and L. F. Dell'Osso, "Wavelet analysis in infantile nystagmus syndrome: limitations and abilities," *Investigative Ophthalmology and Visual Science*, vol. 49, no. 8, pp. 3413–3423, 2008.
- [11] T. Pander, R. Czabański, T. Przybyła et al., "A new method of saccadic eye movement detection for optokinetic nystagmus analysis," in *Proceedings of the 34th Annual International Conference of the IEEE EMBS*, pp. 3464–3467, San Diego, Calif, USA, 2012.
- [12] S. Tominaga and T. Tanaka, "3-Dimensional analysis of nystagmus using video and image processing," in *The Society of Instrument and Control Engineers Annual Conference (SICE '10)*, pp. 89–91, Taipei, Taiwan, August 2010.
- [13] M. Juhola, H. Aalto, T. Jutila, and T. P. Hirvonen, "Signal analysis of three-dimensional nystagmus for otoneurological investigations," *Annals of Biomedical Engineering*, vol. 39, no. 3, pp. 973–982, 2011.
- [14] T. C. Lu and C. Y. Tseng, "Hemodialysis key features mining and patients clustering technologies," *Advances in Artificial Neural Systems*, vol. 2012, Article ID 835903, 11 pages, 2012.
- [15] F. Steinbruecker, A. Meyer-Bease, T. Schlossbauer, and D. Cremers, "Evaluation of a nonrigid motion compensation technique based on spatiotemporal features for small lesion detection in breast MRI," *Advances in Artificial Neural Systems*, vol. 2012, Article ID 808602, 10 pages, 2012.
- [16] M. Juhola, H. Aalto, and T. Hirvonen, "Using results of eye movement signal analysis in the neural network recognition of otoneurological patients," *Computer Methods and Programs in Biomedicine*, vol. 86, no. 3, pp. 216–226, 2007.
- [17] M. Juhola and V. Siermala, "A scatter method for data and variable importance evaluation," *Integrated Computer-Aided Engineering*, vol. 19, no. 2, pp. 137–139, 2012.
- [18] Scatter method, [http://www.uta.fi/sis/cis/research\\_groups/darg/publications.html](http://www.uta.fi/sis/cis/research_groups/darg/publications.html).



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

