Genetic algorithm based on support vector machines for computer vision syndrome classification in health personnel

Eva María Artime Ríos[1], Ana Suarez Sánchez[2*], Sergio Luis Suárez Gómez[3], Fernando Sánchez Lasheras[4], María del Mar Seguí Crespo[5]

[1] Central University Hospital of Asturias, Oviedo, Spain
[2] Department of Business Administration, University of Oviedo, Spain
[3] Prospecting and Exploitation of Mines Department, University of Oviedo. Spain
[4] Department of Mathematics, University of Oviedo, Spain
[5] Department of Optics, Pharmacology and Anatomy, University of Alicante, Spain

[*]Correspondence: suarezana@uniovi.es

Abstract The inclusion in workplaces of video display terminals has brought multiple benefits for the organization of work. Nevertheless, it also implies a series of risks for the health of the workers, since it can cause ocular and visual disorders, among others.
In this research, a group of eye and vision-related problems associated to prolonged computer use (known as computer vision syndrome) are studied. The aim is to select the characteristics of the subject that are most relevant for the occurrence of this syndrome, and then, to develop a classification model for its prediction.
The estimate of this problem is made by means of support vector machines for classification. This machine learning technique will be trained with the support of a genetic algorithm. This provides the training of the support vector machine with different patterns of parameters, improving its performance.
The model performance is verified in terms of the area under the ROC curve, which leads to a model with high accuracy in the classification of the syndrome.

Keywords support vector machines; genetic algorithms; computer vision syndrome; health personnel.

1 Introduction

Medical records mean all the documents containing data, assessments and information of any kind on a patient's status and clinical development throughout the care process. They include identifying of all the professionals who have intervened and aim to obtain the fullest possible integration in the clinical documentation of every patient, at least in the area of each center [1]. The current legislation allows the clinical information to be processed in electronic format, an informed consent being the only exception [1]. Since the introduction of Electronic Health Records (EHR) in the National Health Service, health personnel have become users of video display terminals (VDT).

VDT are alphanumeric or graphical screens, regardless of the method of visual representation used, and represent one of the most typical elements of computerized work equipment. Most workplaces with VDT are designed for sedentary work. Nevertheless, with health personnel the workplace design frequently includes VDT and working standing up; this happens, for example, with workstations for the administration of medication or with a portable terminal, which makes it possible to work beside the bed.

In Spain, the regulation on health protection of the workers exposed to the risks derived from working with VDT is fundamentally set out in the Royal Decree 488/1997 [2]. The National Institute of Safety and Hygiene at Work (Instituto Nacional de Seguridad e Higiene en el Trabajo, INSHT) provides criteria and recommendations for the interpretation and application of the said Royal Decree [3]. According to this regulation "workers using VDT" are, among others, those who effectively work more than 4 hours a day or 20 hours a week with this equipment. The "portable" systems are excluded when they are not continuously used at work.

Working with VDT has brought multiple benefits, but it also implies a series of risks for the health of the workers, since it demands certain requirements of physical and mental load. It can cause ocular and visual disorders, musculoskeletal disorders or other alterations, such as mental fatigue. The use of VDT during long periods of time has been associated with intense eye strain [4-5] as well as with changes in the ocular surface and in the condition of the tear film [6, 7]. In these working places, Kroemer [8] considers the screen, the source document and the keyboard as visual targets. If these targets are separate, in direction or distance, the eyes must refocus constantly while they sweep from one target to another, with continuous changes in posture and convergence. This also requires good coordination of eye movements in order to merge the images of both eyes and to obtain a suitable binocular vision [4]. In addition, during computer work blinking frequency decreases, increasing the evaporation rate of the tear film, which compromises the good condition of the ocular surface. The American Optometric Association [9] defines computer vision syndrome (CVS) as a group of eye and vision-related problems that result from prolonged computer use. The most common symptoms associated with CVS are eyestrain, blurred vision, dry eyes and headaches, among others. In Spain, there are two instruments that have recently been developed and validated to measure this syndrome [10-11].

The prevalence of CVS is high, presenting variations in different studies, according to the characteristics of the sample, the method and the instruments used for data collection. It is related to individual and work-related factors. The higher the exposure to VDT, the greater the prevalence of CVS. Ranashinghe et al. [12], estimate the prevalence of CVS's at 67.4% among a group of 2,210 computer office workers of Sri Lanka. A lower prevalence (53%) has been observed in a study carried out with a sample of 426 office workers in Spain [13] and in a group of 476 call center operators in Brazil (54.6%) [14]. According to the bibliographical review, to date no studies have been identified that evaluate the effects on visual health caused by exposure to VDT in health personnel. The most common occupations in the samples studied in the revised bibliography include office workers, bank employees, high-tech workers, graphical editors and call center operators. Only two studies carried out in Turkey included a sample of workers who used computers from two hospitals. The first one [15] included secretaries, computer operators and hospital data management system users. The second one [7] did not specify whether they were health personnel or not.

Nowadays, Machine Learning techniques are used extensively in medicine. Classification is a popular method that plays an important role in different medical diagnoses. Classification methods can help to integrate computer techniques in health care environments to improve medical services quality and efficiency. In recent years, some studies have employed classification machine learning techniques in different medical applications. For example, support vector machines (SVM) [16] have been employed in order to classify patients who will survive in the long term and those that not in oral squamous cell carcinoma cases, using as input variables clinicopathological parameters and molecular markers. Also, another study employed artificial neural networks for breast cancer diagnosis [17]; the model developed is able to determine which women are more likely to suffer from a particular kind of tumour before they undergo a mammography, which is helpful in order to increase the time from one screening control to the next. Among the classification techniques, genetic algorithms are adaptive search methods for finding optimal or near optimal solutions, premised on the evolutionary ideas of natural selection. This methodology, has already been employed in medical research [18], a remarkable application is the one that employed a genetic algorithm-based model to detect longitudinal changes in white matter fiber-bundles of patient with multiple sclerosis.

The aim of the present work was to select the characteristics of the subject most relevant for the occurrence of CVS, and then to develop a classification model for its prediction in health personnel. For this purpose, the prosed algorithm combines SVM and genetic algorithms.

2 Materials and methods

2.1 Case of Study

This was an observational cross-sectional epidemiological study, based on the completion of two self-administered questionnaires among health personnel of the Monte Naranco Hospital (Oviedo, Spain). This hospital is specialized in geriatrics and palliative care and it started using EHR in 2007.

The study included health personnel that were using VDT at work in the following occupational categories: physicians and surgeons, residents, nurses, advanced practice nurses (APNs) in training and auxiliary nurses. Exclusion criteria applicable to all participants were: not being a user of VDT at work, being a student, working in another occupational category, having seniority of less than a year, and currently suffering from, and/or being under treatment for, a diagnosed ocular disease.

Of a total of 172 workers, 151 (87.79 %) took part in the study. In the end the sample that fulfilled the inclusion and exclusion criteria was finally of 139 workers from 20 different services or hospitalization units. The reasons of exclusion were that they were suffering the following diseases when the questionnaires were collected: dry eye (3), amblyopia (1), non-surgically controlled cataracts (2), vitreous diseases (3) and retinal disease (1). Also, one worker was excluded because he had been in his job for less than one year and another one for the lack of information about their seniority.

All the participants signed an informed consent form accepting their participation, in which data confidentiality was guaranteed during the entire process. The study was authorized by the local health authority and approved both by the Research Ethics Committee of the Principality of Asturias and by the Ethics Committee of the University of Alicante, Spain (the coordinating institution of the study), in accordance with the tenets of the Declaration of Helsinki.

2.2 Data collection

The health workers who took part in the study answered the following self-administered questionnaires:

1. Anamnesis and History of Exposure Questionnaire, which was specifically developed for this study, to gather information about sex, age, use of ophthalmics and/or contact lenses use, history of eye disease and treatment, thereof previous eye surgeries, occupational categories (physicians and surgeons including residents, nurses including APNs in training and auxiliary nurses) and seniority, work schedule (morning shifts, evening shifts, rotating shifts without nights, rotating shifts including nights and morning shifts with on-call shifts), current departments and nursing units and its seniority, information about the ease of use of the software application and daily VDT usage at and outside work.
2. The Computer Vision Syndrome Questionnaire (CVS-Q), designed and validated by Seguí et al. in 2015 [10], was used to measure perceived ocular and visual symptoms during or immediately following computer work. This questionnaire evaluates the frequency (never, occasionally or often/always) and the intensity (moderate or intense) of 16 ocular and visual symptoms: burning, itching, foreign body sensation, tearing, excessive blinking, eye redness, eye pain, heavy eyelids, dryness, blurred vision, double vision, difficulty in focusing for near vision, increased sensitivity to light, colored halos around objects, feeling that eyesight is worsening, and headaches. Subjects with a score of 6 or more on the questionnaire are classified as symptomatic (suffering CVS).

2.3 Support Vector Machines

The SVM are machine learning techniques. Among other mathematical models for similar problems [19–21], SVM are used to model physical systems through the adaptation of their parameters [22–25]. These methods are broadly known for their usage in classification and regression problems [26-27]. In the case of the present research, the SVM is used as a classificator. This technique has been selected due to its well-known performance. The performance of SVM relies on the adjustment of the model to data previously taken from the system to be modelled, as its training data set. To train a SVM to model a classification problem, the vectors from the training data are used to map hyperplanes that define the separation of classes. The output estimation provided by a trained SVM can be formulated as:

$$\hat{y}_i = a^T \Phi(x_i) + b \quad (1)$$

Where x_i corresponds to the input vectors from the training set. The function $\Phi(x_i)$ linearizes the influences between inputs and outputs. In this scenario, the parameters are a and b, which are a vector of the same dimension as the image of Φ, and a coefficient, respectively. These parameters are determined by finding an optimized solution to the following problem and with restrictions:

$$\min_{a,\varepsilon,\eta_i,\eta_i'} \frac{1}{2} a^T a + c\left(\frac{1}{N}\sum_{i=1}^{N}(\eta_i + \eta_i') + v\varepsilon\right) \quad (2)$$

$$a^T \Phi(x_i) + b - y_i \leq \varepsilon + \eta_i (3)$$

$$y_i - a^T \Phi(x_i) - b \leq \varepsilon + \eta_i' (4)$$

Where c is a regularization parameter, ε is the tolerance error for each input x_i. Both η and η' are the slack variables, which take positive values. Finally, v is a parameter for the adjustment of the tolerance. Therefore, the output of SMV [28] can be expressed as:

$$\hat{y}_i = F(x) = \sum_{i=1}^{N}(\beta_i' - \beta)\Phi(x_i)^T \Phi(x) + b \quad (3)$$

In this expression, β and β' are the Lagrange multiplayers corresponding to the restrictions above. In this context, the kernel function K, can be defined as $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$. Consequently, the estimate of SVM turns into the following expression

$$\hat{y}_i = \sum_{i=1}^{N}(\beta_i' - \beta)K(x_i, x) + b \ (4)$$

In general, the SVM for classification can be determined with the parameters c and v, since, as said before, a and b can be obtained as the optimal solution to the quadratic problem [26]. Depending on the sort of function chosen as kernel, other parameters should be determined as well.

In the case of the present research, four different kernel functions are employed. One of them is the linear kernel, which is the sum of cross products. The limitations of linear learning machines are well-known and in general, to solve real-world applications, more complex hypothesis spaces are required. Kernel representations are the solution as they allow data to be projected into a high dimensional features space [29].

According to the Mercer's Theorem [30], Let X be a finite input space with $K(X, z)$ a symmetric function on X. Then $K(X, z)$ is a kernel function if and only if the matrix K is a positive semidefinite or, in other words, has non-negative eigenvalues.

Please note that the use of kernels can overcome the curse of dimensionality in both computation and generalisation.

For the reasons set out in the paragraphs above, three other kinds of non-linear kernels, are employed in the present research:

They are as follow:

- Polynomial: $(\phi(x^1\mu) + 1)^p$, where p represents the degree.
- Radial basis function: $exp\ (-\sigma\|x - \mu\|^2)$

- Hyperbolic tangent: $\tanh(\phi(x^1\mu) + 1)$

In the equations listed above, ϕ and σ represent the scaling parameters. As is well-known in existing literature, the performance of the kernel depends on the problem to be solved. That is the reason why four different ones are tested in our algorithm.

Finally, we would like to remind the reader of the kernel trick which says: "Given an algorithm which is formulated in terms of a positive definite kernel K, one can construct an alternative algorithm by replacing K by another positive definite kernel K.

2.4 Genetic algorithms

In general, in artificial intelligence, an evolutionary algorithm is a kind of algorithm that is based on the evolution of certain possible solutions set by means of a metaheuristic optimization algorithm. The Genetic Algorithms are evolutionary procedures developed to simulate the evolution of a population in terms of optimizing the survival of the next generation. These algorithms were first developed for chromosomic studies [31], but now genetic algorithms work with the premise of improving fitness over the iterations until a solution for an optimization problem is reached. For each generation, the adjustments to the elements are based on four basic genetic operators used as criteria: crossover, mutation, reproduction and elitism [32]. The optimization problem and the iterations of the genetic algorithm can be formulated as follows [33]:

For a function $f: D \to R$, and a set of restrictions, the minimization problem consists of looking for the best value x in the domain D such that $f(x) \leq f(y)$ for all y in the domain D. The value x' in the domain D is a local minimum of f, when a neighborhood $N(x')$ of x' exists where for all z in $N(x')$, $f(x') \leq f(z)$.

In this scenario, the genetic algorithm begins with an initial population $P_0 = \{I_0^1, \dots, I_0^n\}$. Each step of the algorithm, the objective function is calculated along with its correspondent performance measures, then, generates a new population of which the elements are selected with a determined rule from the four genetic operators defined above. After m steps, the population is denoted as P_m. The algorithm stops when the performance measures are not significantly improved in a chosen number of iterations.

The crossover operation is defined as follows. Let two individuals I_1 and I_2 of the population P_m. A new individual I'_{12} (string) is creating by randomly choosing a crossover point in the string of I_1 and I_2 and splitting them and recombining the two parts of the individuals in such a way that the length continues being the same.

The mutation operation consists on given certain individual I_1, the change in certain values of its individual bit components (either from '0' to '1' or from '1' to '0'). In order to avoid convergence problems, the mutation rate must be low.

Finally, the elitism mechanism in a GA simply involves passing to the next generation of a certain amount of those individuals of a certain generation that have performed the best performance when evaluated according to the fitness function.

2.5 The proposed algorithm

The aim of the proposed algorithm is to find the adequate variables and parameters with which the SVM will model the proposed classification problem properly. The iterations of the genetic algorithm are focused on maximizing the value of the AUC (Area Under the ROC Curve) of each of the SVM models calculated. The steps of the algorithm are as Figure 1 shows.

The algorithm begins with the setting of the parameters for the genetic algorithm, such as crossover, mutation, elitism and population size. After this, an initial population should be created. As was stated by Galán et al. [33], although the setting depends heavily on in the data considered, a range of optimal parameters can be determined; for the present research, the probability values applied for crossover were those from 0.5 to 1 in steps of 0.1, while the mutation probability employed values of 0.1, 0.2 and 0.3. The elitism probabilities were 0.01, 0.05, 0.1 and 0.2. Different sizes for population of 10, 25 and 50 individuals were considered.

The set of the initial population must be a vector with size of the possible variation of parameters for the SVM. We will consider different types of initial populations, depending on the type of kernel. To avoid this problem, we consider branching the algorithm. Specifically, we will consider four ways in parallel; for each branch, the genetic algorithm is performed with each possible type of kernel. Now the sets of initial populations can be created. Each item of the initial population has the following form, depending on the kernel used: Linear: $(c, v, x_1, \ldots, x_k)$. Polynomial: $(\gamma, \alpha_0, \alpha, c, v, x_1, \ldots, x_k)$. Radial basis: $(\gamma, c, v, x_1, \ldots, x_k)$. Sigmoid: $(\gamma, \alpha_0, x_1, \ldots, x_k)$. The parameters common to all the populations are: the cost c of constraints violation which corresponds to the constant of the regularization term in the Lagrange formulation of the SVM, ranging from 10^{-2} to 10^{2}; the tolerance error v for the determination of the SVM which takes values from $5 \cdot 10^{-4}$ to $5 \cdot 10^{-3}$; variables$x_1, \ldots, x_k$, with values 0 or 1, depending if the variable is taken into account for the SVM model or not. For the other parameters, γ is needed in all kernels except linear and it takes values that range from $1/(2 \cdot datadimension)$ to $1/(datadimension)$; α_0 works as coefficient for the sigmoid and polynomial kernel (from -1 to 1), and finally, α for the degree determination of the polynomial kernel (from 3 up to 5). The SVM model is then estimated with the previous values for each of the branches, and its AUC with a validation subset of the data is calculated. The stop criteria will be satisfied if the AUC does not change more than 0.01% in the last 100 iterations of the algorithm.

When the stop criteria is not satisfied, a new population has to be created. This is performed through implementing the crossover, mutation and elitism. With the elitism process, the populations were sorted based on their value of the AUC, and only the ones with higher AUC were considered. Crossover will combine sections of the chosen populations, and finally, with the mutation process an aleatory modification is introduced to the population. Then the whole process is repeated until the stop criteria are satisfied.

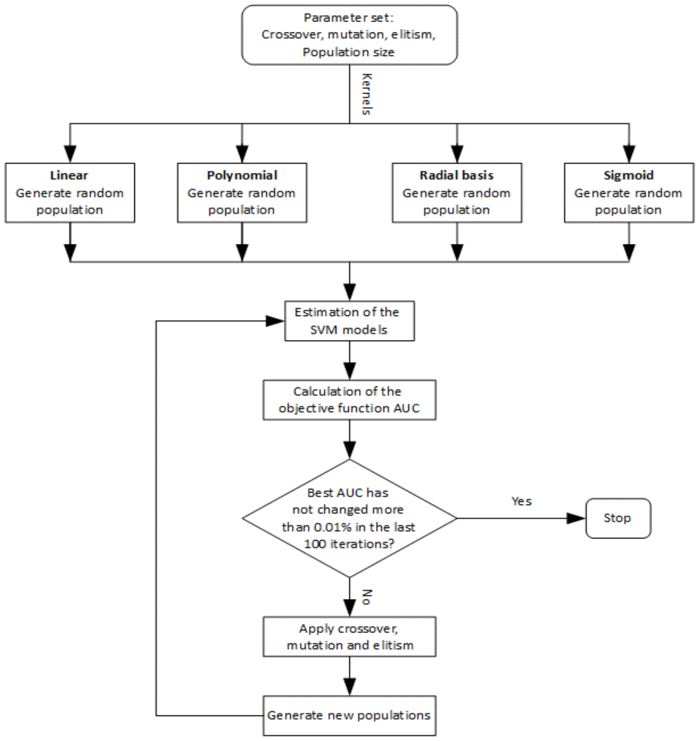

Fig. 1 Algorithm diagram

3 Results

3.1 Study population and prevalence of CVS

The characteristics of the participants are summarized in Table 1. The mean age of the 139 workers included in the study was 46.75 ± 10.40 years (range 22-64 years), and 82% were women. A majority (46.8%) of the study population worked in a geriatric, palliative or internal medicine service. Most of the workers were employed as a nurse (63.3%). In the occupational category and in the current department or nursing unit of work, the mean seniority was 17.85 ± 10.72 years and 12.19 ± 10.61 years, respectively. They worked with VDT an average of 4.93 ±1.92 hours a day and used a computer outside work an average of 1.43 ±1.15 hours a day.

Table 1 Characteristics of the study population and prevalence (P) of Computer Vision Syndrome

Variables	Nr. of subjects (%)	P (%)
Total	139 (100.0)	84 (60.4)
Gender		
Men	25 (18.0)	12 (48.0)
Women	114 (82.0)	72 (63.2)
Age (years)		
≤ 30	14 (10.1)	8 (57.1)
31-40	24 (17.3)	14 (58.3)
41-50	47 (33.8)	34 (72.3)
51-60	42 (30.2)	23 (54.8)
> 60	12 (8.6)	5 (41.7)
Ophthalmic lens wearers		
No	40 (28.8)	22 (55.0)
Yes	99 (71.2)	62 (62.6)
Contact lens wearers		
No	124 (89.2)	72 (58.1)
Yes	15 (10.8)	12 (80.0)
Ocular Surgery		
No	121 (87.1)	72 (59.5)
Yes	18 (12.9)	12 (66.7)
Occupational categories		
physicians and surgeons including residents	43 (30.9)	23 (53.5)
nurses including APNs in training	88 (63.3)	57 (64.8)
auxiliary nurses	8 (5.8)	4 (50.0)
Work schedule		
morning shifts	63 (45.3)	36 (57.1)
evening shifts	1 (0.7)	0 (0.0)
rotating shifts without nights	16 (11.5)	9 (56.3)
rotating shifts including nights	39 (28.1)	28 (71.8)
morning shifts with on-call shifts	20 (14.4)	11 (55.0)
Easy software application		
No	12 (8.6)	10 (83.3)
Yes	127 (91.4)	74 (58.3)
Use of VDT at work (hours/day)		
<2	3 (2.2)	1 (33.3)
2-4	61 (43.9)	34 (55.7)
>4	75 (54.0)	49 (65.3)
Use of VDT outside work (hours/day)		
No	23 (16.5)	11 (47.8)
Yes	116 (83.5)	73 (62.9)

The prevalence of CVS was 60.4%. A higher prevalence was detected among women (63.2%), those who were between 41 and 50 years of age (72.3%) and individuals who worked as a nurse (64.8%), more than 4 hours a day with PVD (65.3%) or with rotating shifts including nights (71.8%). The same happened with wearers of ophthalmic or contact lenses (62.6% and 80.0% respectively) and those who had undergone ocular surgery (66.7%), as well as with individuals who used VDT outside work (62.9%) or those who did not consider the software application as "easy" (83.3%). The most frequently occurring symptoms included "feeling that eyesight is worsening" (64.7%), "headaches" (61.9%), "difficulty in focusing for near vision" (61.9%), "increased sensitivity to light" (58.3%) and "itching" (56.1%).

3.2 Classification model

The chosen model was selected due to its AUC. The value of this performance metric was 0.9433029, giving a high performance over the validation data. The average AUC of the models for each iteration is shown in Figure 2. For each iteration the AUC is computed and the average value presented in this figure. As can be observed, the performance of the models increases quickly in the first iterations, and afterwards oscillates between the iterations 400 and the 900. Then it raises again until the performance level is high, and finally the average value remains stable for, at least 100 iterations and consequently the algorithm stops, determining the model chosen. This model is set with a sigmoid kernel, and parameters $\gamma = 2{,}1 \cdot 10^{-4}$, $\alpha_0 = 0$, $c = 1$ and $\upsilon = 5 \cdot 10^{-4}$. Figure 3 shows a color map in which the AUC variations are compared for different values of γ and υ when the sigmoid kernel is employed for $\alpha_0 = 0$ and $c = 1$.

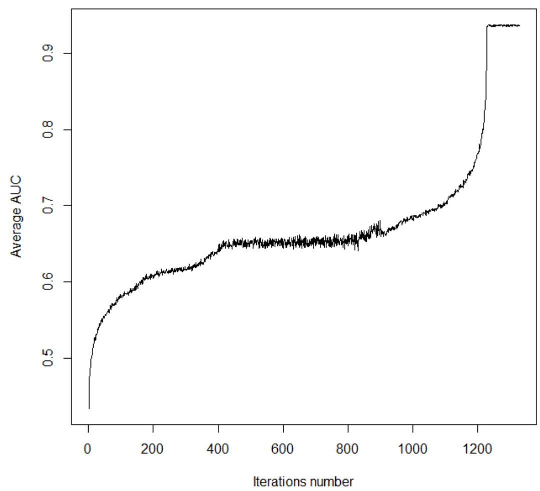

Fig. 2 Average AUC of the estimated models over the genetic algorithm iterations

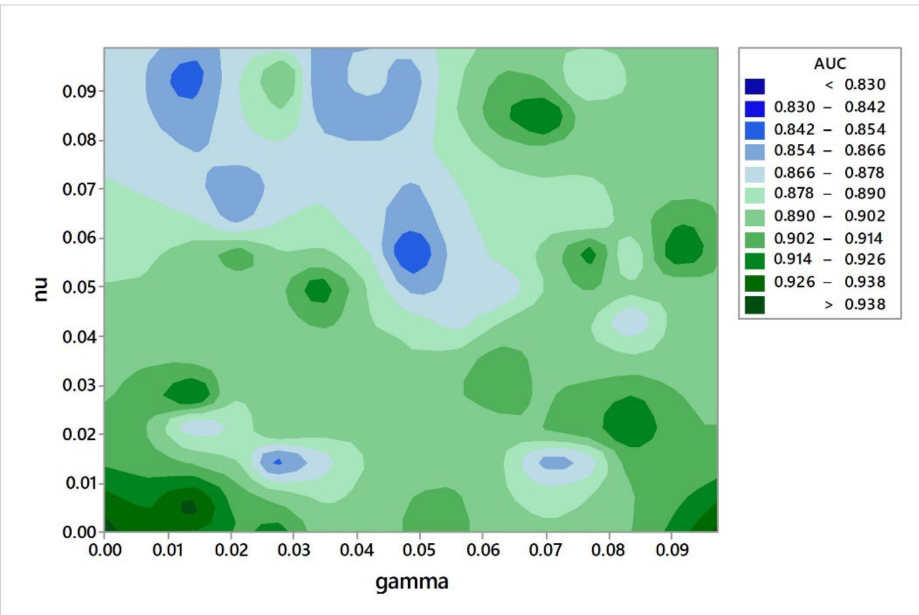

Fig. 3 Color map of AUC for different values of γ and ν when the sigmoid kernel is employed for $\alpha = 0$ and $c = 1$.

The variables identified by the model as the most relevant in the development of CVS, together with their description, are summarized in Table 2.

The AUC was compared with another performance metric, the Youden index, to corroborate the robustness of the model. The comparison is performed via the correlation between both metrics, giving a result of 0.8594404, which implies a high degree of correspondence between the calculated models.

4 Discussion and conclusions

According to the model developed, workers who used a computer outside work and those exposed to VDT more hours a day at the workplace were more likely to suffer CVS. In other studies the total number of hours a day using VDT (both at work and outside the workplace) was found to be related to visual discomfort [34-35], and prolonged VDT use at work was also connected to an increased risk of developing CVS [12-13]. Our results are similar to those of Tauste et al. [13] who observed that workers who use VDT more than six hours a day at work had a prevalence of CVS of 63% and therefore were more likely to develop the syndrome (OR=2.28; 95%CI 0.96-5.37). Likewise Uchino et al. [36] identified that more than 8 hour-a-day VDT users had a significantly increased risk of dry eye disease (OR=1.94; 95%CI 1.22-3.09); this coincides with two of the categories identified in this study: people with dryness and those who use lubricant eye drops.

Table 2 Variables and categories selected by the model as the most relevant for the development of CVS

Variable	**Description**	**Category**
Gender	Sex of the worker.	Women
Age	Age of the worker.	41 – 50 years
Contact lens wearers	Daily or occasional use of contact lenses by the worker.	Yes
Occupational category	Health personnel's groups of similar jobs identified with a common occupational title. Jobs and occupations can be described not only in terms of tasks, but also in terms of associated characteristics such as skill, responsibility, earnings, entry qualifications and prestige.	Nurses (including APNs in training)
Work schedule	Time periods during which different groups of workers perform their tasks (shift work). Shift work is an employment practice designed to make use of, or provide service across, all 24 hours of the clock each day of the week. This is common among health personnel.	Rotating shifts including nights
Easy software application	User-friendliness of the software.	No
Use of VDT at work	Number of hours per day of usage of VDT at the workplace.	>4 hours/day
Use of VDT outside work	Number of hours per day of usage of VDT outside the workplace.	Yes
Current eye complaint	The main current complaint that the worker reports with respect to their eyes.	Dryness
Current eye care	Eye care the worker is currently undergoing.	Lubricant eye drops (artificial tears)

The model also associates the female gender with a higher risk of developing CVS as well. Many studies have reported an association between the female gender and the prevalence of CVS [12-14]. However, when considering individual symptoms, Ranasinghe et al. [12] reported higher prevalence of red eyes, changes in visualizing colours and excessive tearing (p<0,05) in men. According to our results, working as a nurse is also associated with CVS; this may be explained because 95.5% in this occupational category are women.

In our study, age is also identified as a factor related to CVS. The prevalence of CVS was higher among those aged from 41 to 50 years. The average age of those with CVS (45.92±10.06 years) was lower than the average age of those without CVS (47.12±11.31 years). Rosenfield [37], in a review of the scientific literature on CVS, stated that it is unclear whether asthenopia during computer use is associated with age. As a matter of fact, there are studies where no significant change with age was observed in computer-related symptoms [12, 35].

Another relevant result of the model is that contact lens wearers are more likely to suffer CVS. This is consistent with previous studies. Tauste et al. [13] found, in the group of workers exposed to the computer for more than 6 hours a day, that contact lens wearers were almost five times more likely to suffer CVS (OR=4.85; 95%CI 1.25-18.80). Kojima et al. [38] also observed lower tear reservoir volumes in contact lens wearers, as well as higher visual symptom scores associated with blurred vision and visual difficulties.

Finally, a work schedule with shifts including nights was also associated by the model with the development of CVS. Previous studies [39-41] suggest an association between rotating night shift work and several diseases, including cardiovascular disease, cancer risk, diabetes, hypertension, chronic fatigue, developing excess weight and obesity, sleeping problems and early spontaneous pregnancy loss.

Taking into account that the AUC means the ability of the model to detect computer vision syndrome and also the results obtained by means of the Youden index (a single statistic that captures the performance of dichotomous diagnostic test), the results of the proposed model are promising.

Nevertheless, these results should be interpreted with caution given the limitations of our study. The first of these is that it was a cross-sectional design, and we cannot be sure that the cause precedes the effect. A second one is that we did not include ophthalmic examinations that inform us of the workers' refractive state.

Despite these limitations, the use of a validated questionnaire to measure CVS is a particular strength of this research, and this is the first study that describes the relationship between CVS and its associated factors among health personnel by means of genetic algorithms and support vector machines. Both machine-learning techniques are combined. SVM is used in order to perform classifications due to its well-known performance, while genetic algorithms are employed in order to optimize the SVM parameters.

Finally, we would like to remark that the main contribution of the present research is a novel hybrid methodology which is able to determine the most important variables in a classification problem.

Conflict of interest statement: The authors report no conflicts of interest in this work.

References

1. Spain (2002) Ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica. Boletín Oficial del Estado, Nº 274 (15-11-2002), pp 40126-40132
2. Spain (1997) Real decreto 488/1997 sobre disposiciones mínimas de seguridad y salud relativas al trabajo con equipos que incluyen pantallas de visualización.

Boletín Oficial del Estado, Nº 97 (23-04-1997), pp 12928–12931.
3. Instituto Nacional de Seguridad e Higiene en el trabajo (2006) Guía Técnica para la evaluación y prevención de los riesgos relativos a la utilización de equipos con pantallas de visualización [Internet]. http://www.insht.es/InshtWeb/Contenidos/Normativa/GuiasTecnicas/Ficheros/pantallas.pdf. Accessed 15 January 2018.
4. Scheiman M (1996) Accommodative and binocular vision disorders associated with video display terminals: diagnosis and management issues. J Am Optom Assoc 67(9): 531–539
5. Bergqvist UO, Knave BG (1994) Eye discomfort and work with visual display terminals. Scand J Work Environ Health 20(1): 27–33
6. Fenga C, Aragona P, Di Nola C, Spinella R (2014) Comparison of ocular surface disease index and tear osmolarity as markers of ocular surface dysfunction in video terminal display workers. Am J Ophthalmol 158(1): 41–48
7. Ünlü C, Güney E, Akçay BÍS, Akçali G, Erdoğan G, Bayramlar H (2012) Comparison of ocular-surface disease index questionnaire, tearfilm break-up time, and Schirmer tests for the evaluation of the tearfilm in computer users with and without dry-eye symptomatology. Clin Ophthalmol 6: 1303–1306
8. Kroemer KHE (1997) Design of the Computer Workstation. In: Helander M, Landauer TK, Prasad VE (ed) Handbook of Human-Computer Interaction, 2nd edn. Elsevier, Amsterdam, pp 1395-1414
9. American Optometric Association. Computer Vision Syndrome [Internet]. https://www.aoa.org/patients-and-public/caring-for-your-vision/protecting-your-vision/computer-vision-syndrome. Accessed 16 January 2018
10. Seguí M del M, Cabrero-García J, Crespo A, Verdú J, Ronda E (2015) A reliable and valid questionnaire was developed to measure computer vision syndrome at the workplace. J Clin Epidemiol 68(6): 662–673
11. González-Pérez M, Susi R, Antona B, Barrio A, González E (2014) The Computer-Vision Symptom Scale (CVSS17): development and initial validation. Invest Ophthalmol Vis Sci 55(7): 4504–4511
12. Ranasinghe P, Wathurapatha WS, Perera YS et al (2016) Computer vision syndrome among computer office workers in a developing country: an evaluation of prevalence and risk factors. BMC Res Notes 9: 150
13. Tauste A, Ronda E, Molina MJ, Seguí M (2016) Effect of contact lens use on Computer Vision Syndrome. Ophthalmic Physiol Opt 36(2): 112-119
14. Sa EC, Ferreira Junior M, Rocha LE (2012) Risk factors for computer visual syndrome (CVS) among operators of two call centers in São Paulo, Brazil. Work 41 (Suppl 1): 3568–3574
15. Yazici A, Sari ES, Sahin G et al (2015) Change in tear film characteristics in visual display terminal users. Eur J Ophthalmol 25(2): 85–89
16. Rosado P, Lequerica-Fernández P, Villallaín L, Peña I, Sanchez-Lasheras F, de Vicente JC (2013) Survival model in oral squamous cell carcinoma based on clinicopathological parameters, molecular markers and support vector machines. P Rosado. Expert Syst Appl 40 (12): 4770-4776
17. Alvarez-Menéndez L, de Cos Juez FJ, Lasheras FS, Riesgo JAA (2010) Artificial neural networks applied to cancer detection in a breast screening programme. Math Comput Model 52 (7-8): 983-991

18. Stamile C, Kocevar G, Cotton F, Sappey-Marinier D (2017) A genetic algorithm-based model for longitudinal changes detection in white matter fiber-bundles of patient with multiple sclerosis. Comput Biol Med 84 (1): 182-188
19. Antón JCÁ, Nieto PJG, de Cos Juez FJ, Lasheras FS, Viejo CB, Gutiérrez NR (2013) Battery state-of-charge estimator using the MARS technique. IEEE Trans Power Electron 28(8): 3798–3805
20. De Cos Juez FJ, Lasheras FS, García Nieto PJ, Suárez MAS (2009) A new data mining methodology applied to the modelling of the influence of diet and lifestyle on the value of bone mineral density in post-menopausal women. Int J Comput Math 86(10–11): 1878–1887
21. Sánchez-Lasheras F, de Andrés J, Lorca P, de Cos Juez FJ (2012) A hybrid device for the solution of sampling bias problems in the forecasting of firms' bankruptcy. Expert Syst Appl 39(8): 7512–7523
22. Osborn J, De Cos Juez FJ, Guzman D et al (2012) Using artificial neural networks for open-loop tomography. Opt Express 20(3): 2420–2434
23. Guzmán D, de Cos Juez FJ, Myers R, Guesalaga A, Lasheras FS (2010) Modeling a MEMS deformable mirror using non-parametric estimation techniques. Opt Express 18(20): 21356–21369
24. Guzmán D, de Cos Juez FJ, Lasheras FS, Myers R, Young L (2010) Deformable mirror model for open-loop adaptive optics using multivariate adaptive regression splines. Opt Express 18(7): 6492–6505
25. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3): 273–297
26. de Cos Juez FJ, Nieto PJG, Torres JM, Castro JT (2010) Analysis of lead times of metallic components in the aerospace industry through a supported vector machine model. Math Comput Model 52(7): 1177–1184
27. Lasheras FS, Vilán JAV, Nieto PJG, del Coz Díaz JJ (2010) The use of design of experiments to improve a neural network model in order to predict the thickness of the chromium layer in a hard chromium plating process. Math Comput Model 52(7): 1169–1176
28. Vapnik VN, Vapnik V (1998) Statistical learning theory, vol. 1. Wiley, New York
29. Shawe-Taylor J, N Cristianini (2004) Kernel Methods for Pattern Analysis, 1st Edition. Cambridge University Press.
30. Mercer J (1909) Functions of positive and negative type and their connection with the theory of integral equations. Philosophical Transactions of the Royal Society A, 209 (441–458): 415–446.
31. Holland JH (1992) Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. MIT press
32. De Falco I, Della Cioppa A, Tarantino E (2002). Mutation-based genetic algorithm: performance evaluation. Appl Soft Comput 1(4): 285–299
33. Galán CO, Lasheras FS, de Cos Juez FJ, Sánchez AB (2017) Missing data imputation of questionnaires by means of genetic algorithms with different fitness functions. J Comput Appl Math 311: 704–717
34. Robertson MM, Huang YH, Larson N (2016) The relationship among computer

work, environmental design, and musculoskeletal and visual discomfort: examining the moderating role of supervisory relations and co-worker support. Int Arch Occup Environ Health 89(1): 7–22
35. Portello JK, Rosenfield M, Bababekova Y, Estrada JM, Leon A (2012) Computer-related visual symptoms in office workers. Ophthalmic Physiol Opt 32(5): 375–382
36. Uchino M, Yokoi N, Uchino Y et al (2013) Prevalence of dry eye disease and its risk factors in visual display terminal users: the Osaka study. Am J Ophthalmol 156 (4): 759–766
37. Rosenfield M (2011) Computer vision syndrome: a review of ocular causes and potential treatments. Ophthalmic Physiol Opt 31(5): 502-515
38. Kojima T, Ibrahim OM, Wakamatsu T et al (2011) The impact of contact lens wear and visual display terminal work on ocular surface and tear functions in office workers. Am J Ophthalmol 152(6): 933–940
39. Ramin C, Devore EE, Wang W, Pierre-Paul J, Wegrzyn LR, Schernhammer ES (2015) Night shift work at specific age ranges and chronic disease risk factors. Occup Environ Med 72(2): 100-107
40. Gu F, Han J, Laden F et al (2015) Total and cause-specific mortality of U.S. nurses working rotating night shifts. Am J Prev Med 48(3): 241-252
41. Stocker LJ, Macklon NS, Cheong YC, Bewley SJ (2014) Influence of shift work on early reproductive outcomes: a systematic review and meta-analysis. Obstet Gynecol 124(1): 99-110