Automatic Detection of Genetic Diseases in Pediatric Age Using Pupillometry

ERNESTO IADANZA, (Senior Member, IEEE), FRANCESCO GORETTI, MICHELE SORELLI, PAOLO MELILLO, (Member, IEEE), LEANDRO PECCHIA, (Member, IEEE), FRANCESCA SIMONELLI, AND MONICA GHERARDELLI

1Department of Information Engineering, University of Florence, 50139 Florence, Italy
2Eye Clinic, Multidisciplinary Department of Medical, Surgical and Dental Sciences, University of Campania Luigi Vanvitelli, 80100 Naples, Italy
3School of Engineering, University of Warwick, Coventry CV4 7AL, U.K.

Corresponding author: Ernesto Iadanza (ernesto.iadanza@unifi.it)

This work was supported by the Italian Ministry of Education, Research and University under Grant PRIN 2015 (Ref. 20158Y77NT - LS5).

ABSTRACT Inherited retinal diseases cause severe visual deficits in children. They are classified into outer and inner retina diseases, and often cause blindness in childhood. The diagnosis for this type of illness is challenging, given the wide range of clinical and genetic causes (with over 200 causative genes). It is routinely based on a complex pattern of clinical tests, including invasive ones, not always appropriate for infants or young children. A different approach is thus needed, that exploits Chromatic Pupillometry, a technique increasingly used to assess outer and inner retina functions. This paper presents a novel Clinical Decision Support System (CDSS), based on Machine Learning using Chromatic Pupillometry in order to support diagnosis of Inherited retinal diseases in pediatric subjects. An approach that combines hardware and software is proposed: a dedicated medical equipment (pupillometer) is used with a purposely designed custom machine learning decision support system. Two distinct Support Vector Machines (SVMs), one for each eye, classify the features extracted from the pupillometric data. The designed CDSS has been used for diagnosis of Retinitis Pigmentosa in pediatric subjects. The results, obtained by combining the two SVMs in an ensemble model, show satisfactory performance of the system, that achieved 0.846 accuracy, 0.937 sensitivity and 0.786 specificity. This is the first study that applies machine learning to pupillometric data in order to diagnose a genetic disease in pediatric age.

INDEX TERMS Artificial intelligence, clinical decision support systems, machine learning, pupillometry, python, rare diseases, retinitis pigmentosa, retinopathy, support vector machine.

I. INTRODUCTION
Inherited Retinal Diseases (IRDs) represent a significant cause of severe visual deficits in children [1]. They frequently cause blindness in childhood in Established Market Economies (1/3000 individuals). IRDs can be divided into diseases of the outer retina, namely photoreceptor degenerations (e.g., Leber Congenital Amaurosis, Retinitis Pigmentosa, Stargardt disease, Cone Dystrophy, Acromatopsia, Choroideremia, etc.), and diseases of the inner retina, mainly retinal ganglion cell degeneration (e.g. congenital glaucoma, dominant optic atrophy, Leber hereditary optic neuropathy). Both conditions are characterized by extremely high genetic heterogeneity with over 200 causative genes identified to date, which represent a remarkable obstacle to a rapid and effective diagnosis (https://sph.uth.edu/retnet/disease.htm), also considering that the same gene could cause different and heterogeneous clinical phenotypes.

A. CURRENT CLINICAL EVALUATION METHODS
The clinical evaluation of IRDs is routinely based on a complex pattern of clinical tests, including invasive ones, that are not always appropriate for infants or young children. For example, electrophysiological testing, that represents the most informative clinical investigation for the diagnosis of inner and outer retinal diseases, often requires sedation of the children. Sedation affects the retinal response and requires a complex healthcare environment (e.g., operating room, pediatric, anesthesiologist, dedicated instrumentation, etc.)
with high costs for the health system. Therefore, the clinical diagnosis is not easy and requires specialized centers. Consequently, it takes a long time for the young patients and their relatives to receive a correct and complete screening.

In many cases the electrophysiological responses are below the noise level (for example, extinguished scotopic electroretinogram response is the condition confirming the diagnosis). These responses are therefore not suitable for monitoring changes in visual functionality, that is relevant for evaluating disease progression and therapy efficacy.

**B. PUPILLOMETRY**

A novel approach to support the diagnosis of IRDs would be useful. To this regard, chromatic pupillometry has been proposed as a highly sensitive and objective test to quantify the function of different light-sensitive retinal cells and, therefore, it has been shown helpful to detect the retinal dysfunction caused by IRDs as summarized in the following [2]–[6].

Photoreceptor cells (rods and cones) exhibit fast temporal kinetics and cause a brisk pupillary constriction in response to light, whereas the inner retinal melanopsin containing intrinsic photosensitive Retinal Ganglion Cells (ipRGCs) exhibits slower temporal kinetics and elicits a sustained pupillary constriction to light stimuli, persisting after light cessation [2]. The relative contributions of the three receptor types (rod, cone, and melanopsin photopigments) to the Pupillary Light Reflex (PLR) have been examined by manipulating the characteristics of large-field (90) flash stimuli and the adaptation conditions (light vs. dark adapted) [3]. For example, high-luminance, long-wavelength (red) flashes presented against a rod-suppressing adapting field elicit a PLR that is predominately cone-mediated whereas low-luminance, short-wavelength (blue) flashes presented to the dark-adapted eye elicits a PLR that is primarily rod-mediated. For high-luminance, short-wavelength flashes presented to the dark-adapted eye, there is an initial transient pupillary constriction (rod- and cone-mediated) that is followed by a melanopsin-mediated sustained constriction that can last for more than 30s after stimulus offset. The prolonged melanopsin-mediated constriction has been used in clinical protocols to assess inner-retina function [4]–[6]. Thus, the use of chromatic pupil responses may be a novel way to diagnose and monitor diseases affecting either the outer or inner retina [2].

This evidence suggested that a clinical decision support system (CDSS) based on chromatic pupillometry could be developed in order to support diagnosis of IRDs.

**C. THE RESEARCH PROJECT**

Our activity was performed within a research project, which main goal is defining effective protocols and systems for an early diagnosis and monitoring through chromatic pupillometry. The team that worked to this project is structured in three operative units: Department of Information Engineering of the University of Florence, Eye Clinics at the University of Campania Luigi Vanvitelli and at the University of Milan. This team designed a novel CDSS for diagnosis of Retinitis Pigmentosa (RP) in pediatric subjects. The activity included the following steps:

- Development of the pupillometric based protocol [7]. In the first phase of the project, the team focused on subjects with RP (one of the IRDs with the highest prevalence) aged eight to 16 years old.
- Design and creation of an Information Technology (IT) cloud web-platform by the operative unit of Florence. The components of other operative units have taken advantage of this web-platform to share results and data obtained in the project’s partner institutions [8], [9].
- Analysis of Machine Learning (ML) techniques which could be instrumental in the development of the CDSS (see the next Section)
- Deployment of the overall system in an existing web application realized within the same project [8], [9].

**D. STRUCTURE OF THE ARTICLE**

Section II of this paper briefly illustrates the analysis of the state of the art carried out before approaching the presented research. Section III provides some information about patients that contributed to this activity and describes the system structure. Specifically, this section describes the acquisition of raw data by the pupillometric equipment, the analysis of the filtering chain for removing artefacts, the extraction of the features of interest and the implementation of the ML classifier itself. Section IV and Section V, respectively, illustrate the obtained results and the system performance.

**II. STATE OF THE ART**

**A. MACHINE LEARNING**

ML tools have been proven as very effective in supporting the decision process, as widely confirmed by literature [10]–[13]. One of the most important applications can be found in the clinical domain [14], [15]. Indeed in previous works the authors successfully used ML for creating CDSSs dedicated to chronic diseases such as congestive heart failure [16], [17] [18] or chronic obstructive pulmonary disease [19]. A considerable body of literature supports the extreme effectiveness of these techniques also in the field of ophthalmology, as recently reviewed by Hogarty et al. [20] and by Kapoor [21]. The problem arises when dealing with ML technique and chromatic pupillometry applied for the diagnosis of IRDs.

**B. LITERATURE OVERVIEW**

A study of the state of the art was developed at the beginning of the activity. The search for previous articles in the literature was done on Scopus, IEEE Xplore and PubMed, using the following keywords: “clinical decision support system”, “eye diseases”, “rare eye diseases”, “CDSS”, “DSS”, “pupillometry”, “retinitis pigmentosa” and “machine learning”. No articles including all the above keywords were found. None of the found articles use both pupillometry and ML techniques. Most of the found articles refer to “clinical
decision support system”, “machine learning” and “eye diseases”. The number of studies decreases when it deals with systems for “rare diseases”, “retinitis pigmentosa” and “pupillometry”. Among all the found articles, the seven resumed below were chosen based on regency and variety, so as to have different views of general approaches when ML interfaces with eye diseases. Brancati et al. [22] apply ML supervised techniques for detecting pigment signs on fundus images acquired with a digital retinal camera to study patients affected by RP. Gao et al. [23] apply the ML random forest algorithm on optical coherence tomography (OCT) images to support the diagnosis of choroideremia by detecting intact choriocapillaris. Four more articles apply similar supervised ML algorithms to common eye diseases such as age-related macular degenerations [24], [25] diabetic retinopathy [26] and glaucoma [27]. Gargeya et al. [28] bring a different approach to support the diagnosis of diabetic retinopathy using deep learning. The results from the studies just cited are summarized in Table 1.

III. MATERIALS AND METHODS

A. PARTICIPANTS AND EXPERIMENTAL SETUP

This study was approved by the local Ethics Committees of the involved clinical centre (University of Campania and University of Milan) and was conducted in accordance with the guidelines of the Declaration of Helsinki [29]. All the involved participants received detailed information on the research protocol and signed an informed consent form prior to the measurement sessions. 20 patients affected by RP and 18 control subjects, characterized by the absence of any ocular diseases and an absolute refraction error lower than 5 dioptres, were enrolled in the present multi-centric research study. The subjects were recruited and evaluated at the Eye Clinic of the Multidisciplinary Department of Medical, Surgical and Dental Sciences (University of Campania Luigi Vanvitelli, Naples), and at the Department of Clinical Sciences and Community Health of the University of Milan. Subjects underwent a standardized evaluation of pupillary response to chromatic stimulation, carried out with a customized DP-2000 binocular pupillometer (NeurOptics, US) showed in Fig.1.

This pupillometry system enables the simultaneous imaging and assessment of both eyes and, furthermore, supports the automatic detection of the pupil contour and the video tracking of its dynamic response to red (λ = 622 nm), green (λ = 528 nm), blue (λ = 462 nm) and white light stimuli. In detail, videos were captured at a 30-Hz frame rate, with an 8-bit grey-level resolution and a spatial resolution of 0.05 mm. In order to build a homogeneous set of pupillometric data, a standard protocol was agreed between the different participating institutions, in accordance with the current literature [2], [6], [30]–[32]. The identified stimulation sequence required a proper customization of the original firmware, carried out by the manufacturer. Expressly, each measurement session included 10 min of preliminary adaptation to dark, followed by six different stimulation patterns, each applied consecutively to both eyes for three times. Thus, each subject was associated with 18 traces for each pupil, i.e. 36 signals. More specifically, low-intensity impulses against a dark background were used for evaluating rods’ function; high intensity impulses for evaluation of intrinsically photosensitive retinal ganglion cells’ (ipRGCs) function. Moreover, a high-intensity stimulation on a blue background was performed in order to evaluate cones. Each light stimulus lasted 1 s. The phases of the protocol are summarized below, in Fig. 2. Eight out of the 38 test subjects were associated with significantly corrupted signals and were accordingly discarded from the study. Indeed, pupillometric responses may be affected by blinking or eye movements, which could affect the measurement of pupil diameters. The system is actually capable of blinking detection, although this capability was not included in this first study: in the future the protocol could be optimized in order to repeat the stimulation in case of unreliable measurements. The quality of the remaining 30 test subjects was judged by the clinical partners in Naples and Milan. As evident by comparing the signals in Fig. 3 and Fig. 5, this evaluation could be performed by simply plotting the signals and assessing their shapes.

B. CLINICAL DECISION SUPPORT SYSTEM

The main stages for the implementation of the RP classifier are shown in Fig. 4, namely: import and pre-processing of the pupillary diameter signals, pupillary feature extraction and reduction, hyperparameters optimization and, finally, training of the supervised classifier. These stages are discussed in the following paragraphs.

1) SIGNAL PRE-PROCESSING

A first preliminary stage of the CDSS is devoted to the analysis of the raw files, produced by the binocular pupillometer after each measurement session, for the export of the following relevant data:

- Patient ID;
- Bilateral pupillary diameter signals related to each phase of the protocol;
- Diagnosis, i.e. “Pathologic” or “Healthy”, as performed by a clinical specialist.
As detailed in the following, the field diagnosis is used to label the subjects and their related data during the training process of the ML system, whereas the above diameter signals are used to extract clinically motivated features of the pupillary reactivity and for building the input dataset of the supervised classifier. However, before the extraction of the feature set, the raw pupillometric signals must be properly processed in order to attenuate noisy components and, particularly, to cope with potential eye-blink artefacts. Involuntary eye blinking during video capture is indeed associated with abrupt spurious spikes, which might significantly corrupt the resultant traces of the pupil diameter, thus reducing the reliability of the features of interest. In detail, the pre-processing module first involves the application of a Savitzky-Golay

| Article Reference | Algorithm                        | Dataset                                                                 | Performance Evaluation | Input                                  | Results                                                                 |
|-------------------|----------------------------------|-------------------------------------------------------------------------|------------------------|----------------------------------------|------------------------------------------------------------------------|
| Gao et al. (2017) [23] | Random forest                    | 30 images from 19 participants                                          | 10 images for training 20 images for test | OCT images                            | Accuracy: 92.16%          |
|                   |                                   |                                                                        |                        |                                        | Precision: 93.20%          |
|                   |                                   |                                                                        |                        |                                        | Recall: 92.20%            |
|                   |                                   |                                                                        |                        |                                        | F-measure: 92.10%          |
| Khalid et al. (2017) [25] | SVM                              | 1447 healthy 40 central serious chorioretinopathy 650 retinal edema 752 age-related macular degeneration | 40 training 2819 validation | OCT images                            | Accuracy: 99.92%          |
|                   |                                   |                                                                        |                        |                                        | Sensitivity: 100.00%       |
|                   |                                   |                                                                        |                        |                                        | Specificity: 99.86%       |
| Piri et al. (2017) [26] | Custom ensemble algorithm         | >1.4 million diabetic patients from the Cerner Corporation’s Health Facts data warehouse | the exact partition for training and validation set are not specified | Blood test results and demographic data | Accuracy: 92.76%          |
|                   |                                   |                                                                        |                        |                                        | Sensitivity: 90.22%       |
|                   |                                   |                                                                        |                        |                                        | Specificity: 95.30%       |
|                   |                                   |                                                                        |                        |                                        | AUC: 97.90%               |
| Kim et al. (2017) [27]  | Random forest                     | 297 with glaucoma 202 without glaucoma                                  | 319 training 80 validation 100 test | Retinal nerve fiber layer thickness and visual field data | Accuracy: 98.00%          |
|                   |                                   |                                                                        |                        |                                        | Sensitivity: 98.30%       |
|                   |                                   |                                                                        |                        |                                        | Specificity: 97.50%       |
|                   |                                   |                                                                        |                        |                                        | AUC: 97.90%               |
| Gargeya et al. (2017) [28] | Custom deep convolutional neural networks | 75137 images from eye-PACS public dataset | 5-fold stratified cross-validation | Fundus images | Sensitivity: 94.00%          |
|                   |                                   |                                                                        |                        |                                        | Specificity: 98.00%       |
|                   |                                   |                                                                        |                        |                                        | AUC: 97.00%               |
FIGURE 2. Phases of pupillometric protocol, central color is the light one and the side color represent background; numbers in the central part are the intensity of the light stimuli and on top of the scheme there is time express in seconds.

FIGURE 3. Example of irrecoverable signal, the quantity and size of the spikes and artefacts makes it impossible to extract features.

... (SG) of a third-order smoothing filter [33] and window span of 55 samples. This FIR filter generally improves the original SNR levels without greatly distorting the underlying pupillometric signal. Afterwards, the residual between the original data and the SG-smoothed signal is computed: blink-related artefacts are then identified with values exceeding a properly tailored maximum threshold (0.2 mm) and removed accordingly. Finally, possible gaps produced by the above operation are filled by means of a cubic interpolation, and the resulting trace is once again filtered by a low-pass filter so as to give final smoothness to the pupillary trace. A sample filtered signal is shown in Fig. 5.

2) FEATURE EXTRACTION
We selected the most predictive features based on the following literature [3], [4], [5], [6], [34]. After the pre-processing stage, the following 8-elements vector of features is extracted from each pupillometric signal:
- MAX: maximum pupil diameter at baseline;
- MIN: minimum diameter in correspondence with the peak constriction;
- DELTA: absolute difference between the above values;
- CH: percentage maximum constriction (with respect to the pupillary diameter at rest);
- LATENCY: delay between the light stimulus and the onset of the pupillary constriction;
- MCV: mean constriction velocity;
- MDV: mean dilation velocity;
- CVmax: maximum constriction velocity.

The above eight features, calculated on the filtered signal, were chosen in accordance to the literature about pupilometry in several pathologies [35]–[37] and in biometric authentication [38]. The same features are regularly used by the clinicians involved in this project and are also provided by the equipment itself in its output files. The time interval used to derive the above features was properly restricted so as to minimize the risk of inaccurate values: namely, MAX and LATENCY are computed in the first second whereas the others are obtained using a 5-s window. The adopted features are represented below, in Fig. 6. The rationale behind...
TABLE 2. Pupillary reactivity: extracted features.

| Feature | Description                                         | Expression |
|---------|-----------------------------------------------------|------------|
| MAX     | maximum diameter at baseline                       | \( \text{MAX}(r(t)) \) |
| MIN     | minimum diameter corresponding to the peak constriction | \( \text{MIN}(r(t)) \) |
| DELTA   | difference between Max and Min                      | \( \text{MAX} - \text{MIN} \) |
| CH      | percentage maximum constriction                     | \( \frac{\text{DELTA}}{\text{MAX}} \) |
| LATENCY | delay between stimulus and onset of the pupillary constriction | Computed using custom script |
| MCV     | mean constriction velocity                          | \( \frac{\text{DELTA}}{t_{\text{min}} - \text{LATENCY}} \) |
| MDV     | mean dilation velocity                              | \( \frac{t_{90\%} - t_{\text{min}}}{t_{\text{max}} - t_{\text{min}}} \) |
| \( CV_{\text{max}} \) | maximum constriction velocity                      | \( \text{MIN}\left(\frac{dr(t)}{dt}\right) \) |

FIGURE 6. Graphic representation of features.

the definition of the pupillary LATENCY (see Table 2) is that the contraction starts a few milliseconds after the light stimulus is applied. In detail, this parameter is estimated as follows: the first derivative \( d'(t) \) of the pupillometric signal is computed; then, starting from its absolute minimum, the array of values is checked backwards and the time instant corresponding to \( d'(t) = 0 \) is identified. The detection of an inflection point is avoided since, despite the preliminary SG-smoothing, \( d'(t) \) signals are characterized by significant noisy components and zero-crossings are less sensitive to flickering signals. Although this might seem the easiest strategy, it was chosen not to identify the inflection point because it is not possible to have a perfectly smoothed signal which determines a noisy derivative graph. Conversely, the zero-crossing detection is less influenced by flickering signals. In Fig. 7 the latency is highlighted in red while the zero axis in green.

3) SUPPORT VECTOR MACHINES

Support vector machines (SVMs) are supervised linear binary classifiers, first introduced by Vapnik [39]. From a conceptual standpoint, SVMs are formally based on the definition of an optimal linear hyperplane of equation [40]:

\[
w^T x + b = 0
\]

which separates the feature space into two regions, corresponding to the binary classes of the training data. Specifically, the identification of the above decision boundary is performed via the maximization of the geometric margin between the classes:

\[
M_{\text{SVM}} \propto \frac{1}{||w||}.
\]

Maximizing \( M_{\text{SVM}} \) is theoretically equivalent to minimizing the term \( \frac{1}{2}||w||^2 \); accordingly, the training process of an SVM
classifier corresponds to the following optimization problem:

$$\frac{1}{2}||w||^2.$$  \hspace{1cm} (3)

subject to:

$$y_i(w^T x + b) \geq 1 \quad i = 1, \ldots, N.$$  \hspace{1cm} (4)

where $y_i \in \{-1, +1\}$ are the labels identifying the binary classes. However, in practical classification tasks, real input datasets often cannot be directly separated by a linear boundary. Thus, during training, some instances may be allowed to lie either inside the margin $M_{SVM}$ or on the wrong side of the decision hyperplane, leading to the so-called soft margin SVM:

$$\frac{1}{2}||w||^2 + C \sum_{i=1}^{N} \epsilon_i.$$  \hspace{1cm} (5)

subject to:

$$y_i(w^T x + b) \geq 1 - \epsilon_i \quad i = 1, \ldots, N.$$  \hspace{1cm} (6)

where $\epsilon_i$ are the so-called slack variables, corresponding to misclassified input instances, whereas the cost $C \in \mathbb{R}^+_0$ is an internal parameter of the classifier. This boundary constant determines the relative weight given to training accuracy and margin $M_{SVM}$ maximization. More specifically, high values of $C$ will penalize the presence of misclassified cases, thus leading to narrow margins between the classes; on the other hand, small coefficients would tolerate the incorporation of misdetections during the training process, and are thus related to wider geometric margins. Numerically, the above optimization is performed via the method of Lagrange multipliers [41], which identifies boundary feature vectors of expression:

$$\hat{w} = \sum_{i=1}^{N} \hat{a}_i y_i x_i.$$  \hspace{1cm} (7)

In the above equation, the input observations $x_i$ related to non-zero $\hat{a}_i$ coefficients are the support vectors, which lie on the decision hyperplane or inside the corresponding margin, thus defining the boundary between the classified clusters. The SVM, however, might achieve improved training accuracy through a nonlinear transformation of the original training dataset. Indeed, this strategy can lead to the generation of more flexible boundaries with respect to an elementary linear hyperplane. Therefore, in addition to the original linear SVM, the present study also explored the performance attainable by transforming the pupillometric features according to a Gaussian radial basis function (RBF): $\Phi(x_i - x_j) = e^{-\gamma ||x_i - x_j||^2}$, a popular kernel function which is widely adopted in SVM-based classification, for handling non-linearly separable data clusters. Its scale $\gamma$ is the second tenable internal parameter of the SVM. Namely, small values of $\gamma$ favour the identification of smooth classification boundaries which, however, may be comparable to a linear hyperplane for extremely small $\gamma$ and thus lead to underfitting. Conversely, high values tend to produce more flexible, sinuous margins; still, $\gamma$ needs proper configuration since very high values may, on the other hand, compromise the generalization capability of the SVM due to overfitting of the training dataset.

4) FEATURE REDUCTION

According to the adopted measurement protocol, previously detailed in the “Participants and experimental setup” section, a total of 288 features was extracted from the 36 pupil reactivity signals, available for each subject to be classified. Due to the relatively high number of features, feature reduction represented a key preliminary operation which was applied to avoid overfitting of the training dataset. In ML applications, a general rule of thumb is to keep the dimension of the input
feature space below one fifth of the total number of observations, i.e. the best subjects. In the present study, the set of selected features comes from the results of a recent study [34], which has identified a subset of pupillary features with superior discriminant capability regarding the clinical diagnosis of RP: in detail, the values of the maximum pupil diameter (i.e. MAX) before stimuli 1 and 2 appear to be significantly higher in RP patients which, furthermore, are associated with a reduced pupillary constriction (i.e. higher MIN following stimuli 1, 2, 3 and 4; lower DELTA for all stimuli except 6). Consistently, the following six features were chosen for characterizing each subject: DELTA1, DELTA2, MIN3, MIN4, DELTA5, DELTA6. Although there is an evidence for considering also MAX1, MAX2, MIN1, and MIN2, these four properties are implicitly related to DELTA1 and DELTA2, and were thus discarded in order to keep the final dimension of the feature space as low as possible (see the above consideration). Among the three repeated measurements available for each light stimulus, the features related to just one acquisition were selected, according to criteria reported in Table 3.

5) SVM TRAINING AND SUBJECT CLASSIFICATION

The overall scheme of the developed CDSS is shown in Fig. 8. In general, the system is designed (so as) to separately label the left and right eyes and, then, to classify the related subject by means of an OR logical operator, i.e. the subject is diagnosed with RP if at least one of the eyes is assigned with the “Pathologic” label (thus improving the global sensitivity of the CDSS). This choice is related to the fact that the artifacts might be not equally distributed between the two eyes. For example, a patient with a frequent blinking in his/her left eye would generate a cleaner signal for his/her contralateral eye. An SVM was selected as supervised (eye) classification algorithm because of its proven solidity and versatility for classification problems [42]. Each SVM classifier was fed with the pupillometric feature vectors acquired from the left and right eyes of 30 of the enrolled subjects (see Participants and experimental setup). As previously mentioned, linear and RBF kernels were alternatively used for both the left- and right-eye classifiers, so as to explore and compare their performances. The optimization of the hyper-parameters of the SVM, i.e. the boundary constant C and the scale λ of the non-linear RBF kernel, represents a fundamental step for the achievement of improved classification performances (see the above paragraph “Support Vector Machines”).

The tuning process was carried out separately for each eye by means of a grid search [43], [44](over the values listed in Table 4) and a leave-one-out cross-validation strategy, aimed at identifying the best combination of internal parameters associated to the highest average classification accuracy over the 30 subjects [45]. This approach allowed us to exclude possible over-fitting. It also ensured that every single element in the available dataset was exploited both as training set and as validation set, with no fortuity in the partitioning of the feature vectors, as opposed to a k-fold cross-validation (see Fig. 9). Detailed description of this method could be found in Chapter 7 of [45] and in [46].

### IV. RESULTS

The optimal combination of the SVM hyperparameters, returned by the data-driven tuning process, are reported in Tables 5 and 6, alongside the related classification accuracy achieved on the 30 available subjects. Tables 7 and 8 summarize sensitivity, specificity and accuracy for the final ensemble model (schematized in Fig. 8). In details, these performance scores were derived by comparing the actual class of the subject - as assigned by the physician - with the class obtained by applying an OR logical operation to the two labels separately returned by the tuned SVMs for each subject.
eye. As expected, this strategy determines an increase in the overall sensitivity of the CDSS. It is worth to specifying that only one table is reported because both the linear and RBF kernel functions gave the same results in the ensemble logic.

V. DISCUSSION

IRDs are degenerative diseases that affect the eye starting from the first years of life. Accurate diagnosis of inherited retinal disease is a relevant clinical issue. In clinical practice, the diagnosis relies on invasive test, particularly electroretinogram, which requires sedation in children/non-collaborative patients. This aspect brings the necessity of a non-invasive and accurate system to make fast diagnoses in pediatric age. We propose the adoption of chromatic pupillometry to support the screening and we achieved an excellent sensitivity 93.7 % (due to one false negative) with a satisfactory specificity (78.6 %). We privileged the sensitivity over the specificity because this novel technique will be, at least in these first stages, mostly applied for screening purposes. It is planned to test the accuracy of the method in a successive study on a larger sample of pediatric patients, which should undergo electroretinogram to confirm the diagnosis. The non-invasiveness is granted by adopting the proposed pupillometric method, which requires no specific patient preparations with drugs or collyriums. If compared with other standard diagnostic techniques, particularly, electrorheological test, in this case no electrodes need to be placed on the patient skin: this is particularly convenient when dealing with pediatric patients. Particularly, in younger children the electrophysiological testing are usually performed in sedation, thus requiring a more complex clinical setting (i.e. availability of operating theater together with anesthesiologist). Chromatic pupillometry has been proven to be effective in diagnosis of RP [34]. The protocol adopted in the current study included dark-adaption and light adaptation lasting less than 20 minutes, as opposed to standard electroretinogram protocol requiring about 45 minutes without sedation or more than one hour if performed in sedation. However, the procedure for pupillometry is not standardized and the results of chromatic pupillometry are not easily interpretable by ophthalmologists. For this reason, a CDSS could be useful to lighten the time needed for the diagnosis phase. To this aim, given the relatively small amount of available
signals, we decided not to include a feature selection stage, in order to avoid over-fitting but we selected the most predictive features based on the existing literature. This is a very conservative choice: in the future, when more data will be available, it will be interesting comparing the set of features we chose, with the output of a feature extraction step (e.g. Correlation-based Feature Selection for machine learning) [47], [48]. Regarding the classification algorithm, in order to have a fast evaluation, we chose to design a CDSS relying on SVMs optimized in terms of kernel, gamma and C. The CDSS showed satisfactory performance in terms of specificity and sensitivity in this pilot study. It is relevant to remind that, to our knowledge, there are no studies about ML applied to pupillometric data for genetic diseases in pediatric age, so the comparison with the good results of other similar works must be appropriately weighted. Actually, all the previous studies on pupillometric examinations for IRDs relied on statistical analysis without adopting any automatic classification algorithms, even if there is an increasing interest on the using AI for ophthalmological applications (as already shown above in Table 1): almost all the selected studies focused on age-related eye diseases and on retinal imaging and only one ML-system has been proposed for supporting diagnosis of RP. The system, proposed by Brancati et al. [22], was based on fundus imaging and achieved a sensitivity of 75.66% and a specificity of 99.57%. The CDSS proposed in this study outperformed the system proposed by Brancati in terms of sensitivity and consequently seems more suitable for screening of the disease. However, our system, being based on a relatively small data set, needs to be further tested on a larger data set in future works.

VI. CONCLUSION

This paper describes a new approach for supporting clinical decision for diagnosis of retinitis pigmentosa starting from analysis of pupil response to chromatic light stimuli in pediatric patients. The system was developed to clean artefacts, extract features and help the diagnosis of RP using a ML approach based on an ensemble model of two fine-tuned SVMs. Performances were evaluated with a leave-one-out cross-validation, also used to identify the best combination of internal parameters of the SVM, separately for both the left and right eyes. The class assigned to each eye were combined in the end with an OR-like approach so as to maximize the overall sensitivity of the CDSS; the ensemble system achieved 84.6% accuracy, 93.7% sensitivity and 78.6% specificity. The small amount of data available for this work, calls for further tests with a larger data pool for validating the performance of the system. Future scope includes testing the same approach with different devices. A problem that came out with great evidence, at the signal acquisition stage, is the frequent presence of movement artifacts. This is due to the particular shape of the device, together with the young age of the enrolled patients. Devices with different frame, including also systems based on smartphones, are going to be investigated. Moreover, considering the duration of the whole acquisition protocol, the procedure would benefit of some systems to capture the attention of the young patient (and his/her sight).

ACKNOWLEDGMENT

The activities presented in this paper have been performed within the project entitled “Toward new methods for early diagnosis and screening of genetic ocular diseases in childhood”. This project is one of National Interest Research Projects (PRIN). PRIN projects are funded by the Italian Ministry of Education, University and Research, for three years.

REFERENCES

[1] X.-F. Huang, F. Huang, K.-C. Wu, J. Wu, J. Chen, C.-P. Pang, F. Lu, J. Qu, and Z.-B. Jin, “Genotype–phenotype correlation and mutation spectrum in a large cohort of patients with inherited retinal dystrophy revealed by next-generation sequencing,” Genet. Med., vol. 17, no. 4, pp. 271–278, Apr. 2015.
[2] R. Kardon, S. C. Anderson, T. G. Damarjian, E. M. Grace, E. Stone, and A. Kawasaki, “Chromatic pupil responses. Preferential activation of the melanopsin-mediated versus outer photoreceptor-mediated pupil light reflex.” Ophthalmology, vol. 116, no. 8, pp. 1564–1573, 2009.
[3] J. C. Park, A. L. Moura, A. S. Raza, D. W. Rhee, R. H. Kardon, and D. C. Hood, “Toward a clinical protocol for assessing rod, cone, and melanopsin contributions to the human pupil response,” Invest. Ophthalmo- mol. Vis. Sci., vol. 52, no. 9, pp. 6624–6635, Aug. 2011.
[4] A. Kawasaki, S. V. Crippa, R. Kardon, L. Leon, and C. Hamel, “Char-acterization of pupil responses to blue and red light stimuli in autosomal dominant retinitis pigmentosa due to NR2E3 mutation,” Investigative Ophthalmo- mol. Vis. Sci., vol. 53, no. 9, pp. 5562–5569, 2012.
[5] A. Kawasaki, F. L. Munier, L. Leon, and R. H. Kardon, “Pupillometric quantification of residual rod and cone activity in Leber congenital amauro- sis,” Arch. Ophthalmol., vol. 130, no. 6, pp. 798–800, Jan. 2012.
[6] A. Kawasaki, S. Collomb, L. Léon, and M. Münch, “Pupil responses derived from outer and inner retinal photoreception are normal in patients with hereditary optic neuropathy,” Exp. Eye Res., vol. 120, pp. 161–166, Mar. 2014.
[7] P. Melillo, A. de Benedictis, E. Villani, M. C. Ferraro, E. Iadanza, M. Gherardelli, F. Testa, S. Bantli, P. Nucci, and F. Simonelli, “Toward a novel medical device based on chromatic pupillometry for screening and monitoring of inherited ocular disease: A pilot study,” in Proc. IFMBE, vol. 68, 2019, pp. 387–390.
[8] E. Iadanza, R. Fabbrì, A. Laschi, F. Gavazzi, P. Melillo, F. Simonelli, and M. Gherardelli, “ORAO: RESTful cloud-based ophthalmologic medical record for chromatic pupillometry,” in Proc. IFMBE, vol. 73, 2020, pp. 713–720.
[9] E. Iadanza, R. Fabbrì, A. Laschi, P. Melillo, and F. Simonelli, “A col- laborative RESTful cloud-based tool for management of chromatic pupillometry in a clinical trial,” Health Technol., pp. 1–14, Aug. 2019, doi: 10.1007/s12553-019-00362-z.
[10] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, “Supervised machine learning: A review of classification techniques,” Emerg. Artif. Intell. Appl. Comput. Eng., vol. 160, pp. 3–24, Jun. 2007.
[11] J. A. Alzubi, “Optimal classifier ensemble design based on coopera- tive game theory,” Res. J. Appl. Sci., Eng. Technol., vol. 11, no. 12, pp. 1336–1343, Jan. 2016.
[12] J. Alzubi, A. Nayyar, and A. Kumar, “Machine learning from theory to algorithms: An overview,” J. Phys., Conf. Ser., vol. 1142, Nov. 2018, Art. no. 012012.
[13] O. A. Alzubi, J. A. Alzubi, S. Tedmori, H. Rashaidah, and O. Almomani, “Consensus-based combining method for classifier ensembles,” Int. Arab J. Inf. Technol., vol. 15, no. 1, pp. 76–86, Jan. 2018.
[14] P. Sajda, “Machine learning for detection and diagnosis of disease,” Annu. Rev. Biomed. Eng., vol. 8, no. 1, pp. 537–565, Aug. 2006.
[15] J. A. Alzubi, B. Bharathikannan, S. Tanwar, R. Manikandan, A. Khanna, and C. Thaventhiran, “Boosted neural network ensemble classification for lung cancer disease diagnosis,” Appl. Soft Comput., vol. 80, pp. 579–591, Jul. 2019.
[16] G. Guidi, L. Pollonini, C. C. Dasco, and E. Iadanza, “A multi-layer monitoring system for clinical management of congestive heart failure,” BMC Med. Inform. Decis. Making, vol. 15, p. S3, Dec. 2015.

[17] G. Guidi, R. Miniati, M. Mazzola, and E. Iadanza, “Case study: IBM Watson analytics cloud platform as analytics-as-a-service system for heart failure—early detection,” Future Internet, vol. 8, no. 3, p. 32, Jul. 2016.

[18] M. Porumb, E. Iadanza, S. Massaro, and L. Pecchia, “A convolutional neural network approach to detect congestive heart failure,” Biomed. Signal Process. Control, vol. 55, Jan. 2020, Art. no. 101597.

[19] E. Iadanza, V. Mudura, P. Melillo, and M. Gherardelli, “An automatic supporting system clinical decision for chronic obstructive pulmonary disease,” Health Technol., pp. 1–12, Mar. 2019, doi: 10.1007/978-3-030-03137-9.

[20] D. T. Hogarty, D. A. Mackey, and A. W. Hewitt, “Current state and future prospects of artificial intelligence in ophthalmology: A review,” Clin. Exp. Optom., vol. 47, no. 1, pp. 128–139, Jan. 2019.

[21] R. Kapoor, S. P. Walters, and L. A. Al-Aswad, “The current state of artificial intelligence in ophthalmology,” Surv. Optom., vol. 64, no. 2, pp. 233–240, Mar. 2019.

[22] N. Brancati, M. Frucci, D. Gragnaniello, D. Riccio, V. Di Iorio, L. Di Perna, and F. Simonelli, “Learning-based approach to segment pigment signs in fundus images for Retinitis Pigmentosa analysis,” Neurocomputing, vol. 308, pp. 159–171, Sep. 2016.

[23] S. S. Gao, R. C. Patel, N. Jain, M. Zhang, R. G. Weiher, D. Huang, M. E. Pennesi, and Y. Jia, “Choriocapillaris evaluation in choroideremia using optical coherence tomography angiography,” Biomed. Opt. Express, vol. 8, no. 1, p. 48, Jan. 2017.

[24] A. García-Floriani, Á. Ferreira-Santiago, O. Camacho-Nieto, and C. Yáñez-Márquez, “A machine learning approach to medical image classification: Detecting age-related macular degeneration in fundus images,” Comput. Elect. Eng., vol. 75, pp. 218–229, May 2019.

[25] S. Khalid, M. U. Akram, T. Hassan, A. Nasim, and A. Jameel, “Fully automated robotic system to detect retinal edema, central serous choroidal neovascularization, and age related macular degeneration from optical coherence tomography images,” BioMed Res. Int., vol. 2017, pp. 1–15, 2017.

[26] S. Piri, D. Delen, T. Liu, and H. M. Zolbanin, “A data analytics approach to building a clinical decision support system for diabetic retinopathy: Developing and deploying a model ensemble,” Decis. Support Syst., vol. 101, pp. 12–27, Sep. 2017.

[27] S. J. Kim, K. J. Cho, and S. Oh, “Development of machine learning models for diagnosis of glaucoma,” PLoS ONE, vol. 12, no. 5, May 2017, Art. no. e0177726.

[28] R. Gargeya and T. Leng, “Automated identification of diabetic retinopathy using deep learning,” Ophthalmology, vol. 124, no. 7, pp. 962–969, Jul. 2017.

[29] B. Gandevia and A. Tovell, “Declaration of helsinki,” BMJ, vol. 355, no. 7237, Aug. 2015.

[30] A. J. Roman, S. B. Schwartz, T. S. Aleman, A. V. Cideciyan, B. Gandevia and A. Tovell, “Declaration of helsinki,” J. Ophthalmol., vol. 2017, pp. 1–8, 2017.

[31] F. Zennifa and K. Iramina, “Quantitative formula of blink rates—The Elements of Statistical Learning,” New York, NY, USA: Springer-Verlag, 2009.

[32] K. T. Choi, K. F. Tsang, H. R. Chi, B. W. K. Ling, and C. K. Wu, “An accurate ECG-based transportation safety drowsiness detection scheme,” IEEE Trans. Ind. Informat., vol. 12, no. 4, pp. 1438–1452, Aug. 2016.

[33] M. A. Hall, “Correlation-based feature selection for machine learning,” Ph.D. dissertation, Dept. Comput. Sci., Waikato Univ., Hamilton, New Zealand, 1999.

[34] F. Zennifa and K. Iramina, “Quantitative formula of blink rates—pupillometry for attention level detection in supervised machine learning,” IEEE Access, vol. 7, pp. 96263–96271, 2019.

[35] E. Iadanza (Senior Member, IEEE) received the B.M.E. degree in CE, and the M.Sc. and Ph.D. degrees from the University of Florence. He is currently an Adjunct Professor in clinical engineering with the Department of Information Engineering, University of Florence. He has authored more than 145 publications on international books, scientific journals, volumes, and conference proceedings. He is also a member of the IFMBE Administrative Council, the Chairman of the International Federation for Medical and Biological Engineering/Health Technology Assessment Division Board (IFMBE/HTAD), and an Immediate Past Chairman of the Clinical Engineering Division Board (IFMBE/CED) and of the International Union for Physical and Engineering Sciences in Medicine/Education and Training Committee (IUPESM). He is the Senior Member of the IEEE Engineering in Medicine and Biology Society (EMBS). He is a member of the Scientific Committee, a Track Chair, and a Session Chairman of national and international scientific conferences in biomedical engineering. He received the IBM Faculty Award, in 2013, and the IFMBE/CED Teamwork Award, in 2019. He is an Editor-in-Chief of the Clinical Engineering Handbook (Academic Press, Second Ed., 2020). He is an Associate Editor of Health & Technology and Future Internet, a Section Editor of the International Journal of Clinical Engineering and Healthcare Technology Assessment (CEHTA), a member of the Editorial Board of China Medical Devices Journal and of the Journal of Healthcare Technology. He is a Guest editor for the journal Health and Technology Special Issue of Global issues in Clinical Engineering. He has been an Organizer of postgraduate and postmaster courses in clinical engineering and healthcare engineering, and HTA with the University of Florence, since 2007. He is a Supervisor of more than 190 graduation theses.
FRANCESCO GORETTI received the bachelor’s degree in electronic engineering from the University of Florence, in 2018, where he is currently pursuing the master’s degree in biomedical engineering.

Michele Sorelli was born in Florence, Italy, in 1989. He received the master’s degree (cum laude) in biomedical engineering and the Ph.D. degree in information engineering from the University of Florence, in 2015 and 2019, respectively. He is currently a Research Fellow in bioengineering with the Department of Information Engineering, University of Florence. His research interests mainly include bio signal processing, physiological modeling, and machine learning techniques, with a special focus on the noninvasive assessment of microvascular perfusion by means of optical modalities.

PAOLO MELILLO (Member, IEEE) was born in Naples, in June 1985. He received the M.Sc. degree (Hons.) in biomedical engineering from the University of Naples “Federico II”, Italy, in 2008, and the Ph.D. degree from the University of Naples “Federico II,” in 2012. He is currently a Senior Assistant Professor with the Multidisciplinary Department of Medical and Dental Sciences, University of Campania Luigi Vanvitelli. He has authored or coauthored about 50 journal and conference papers in the fields of health management, telemedicine, biomedical signal, and image processing. He is a member of the IEEE Engineering in Medicine and Biology Society, of the Italian Association of Medical and Biological Society (AIIMB), of the International Federation of Medical and Biological Society (IFMBE), and of the Italian Mathematical Union. He was an Editorial Board Member of the IEEE Journal of Translational Engineering in Health and Medicine.

LEANDRO PECCHIA received the degree in biomedical engineering and the Ph.D. degree in biomedical engineering and management of healthcare services from the University of Naples “Federico II”, in 2005 and 2009, respectively. Since 2013, he has been with the University of Warwick, U.K., where he is currently an Associate Professor of biomedical engineering and directs the Applied Biomedical Signal Processing and Intelligent eHealth Lab (ABSPIE). He authored more than 100 peer-reviewed articles on journals, books, and conferences in the fields of health technology assessment (HTA), machine learning and biomedical signal processing applied to healthy ageing, chronic diseases, and fall prediction in the later life. He is a Secretary General of the IUPESM, a Treasurer of the IFMBE Clinical Engineering Division, and a Chair of the Public Affairs Working Group of the EAMBES. He served the IFMBE Healthcare Technology Assessment Division as a Chair, from 2015 to 2018, and a Treasurer, from 2012 to 2015.

FRANCESCA SIMONELLI was born in Nola, Italy, in 1959. She received the degree (Hons.) in medicine and surgery from the University of Naples Federico II, Naples, Italy, in 1983, and the Residency in ophthalmology from the School of Medicine, University of Naples Federico II, in 1987. She is currently a Full Professor of ophthalmology with the Multidisciplinary Department of Medical, Surgical and Dental Sciences, University of Campania Luigi Vanvitelli, Naples. She has authored or coauthored about 100 journal and conference papers in ophthalmology. Prof. Simonelli is a member of the Association for Research in Vision and Ophthalmology, the President of the Italian Society of Ophthalmologic Genetics, a member of the Eye Working Group of the Telethon Institute of Genetics and Medicine, of the National Fighting Blindness Committee of the Italian Ministry of Health, and of the Italian Society of Ophthalmology, and the President of the Scientific Committee of Retina Italia Onlus.

MONICA GHERARDELLI received the M.S. degree in electronic engineering from the University of Florence, Italy, in 1981, and the Ph.D. degree in information engineering from Padua University, Italy, in 1987. She is currently a Professor with the University of Florence and scientific responsible of agreements between the Information Engineering Department, University of Florence and University Hospitals, Tuscany, Italy. She has authored articles in biomedical engineering.

* * *