Classification of Obese and Healthy Children Based on Machine Learning Algorithms using Choroidal Thickness Features

Erkan Bulut (erkanbulut@outlook.com)
Beylikduzu Public Hospital https://orcid.org/0000-0003-3488-1515
Sumeyra Koprubasi
Sancaktepe Şehit Prof Dr Ilhan Varank Training and Research Hospital: Sancaktepe Sehit Prof Dr Ilhan Varank Egitim ve Arastirma Hastanesi
Ozlem Dayi
Beylikduzu Public Hospital
Hatice Bulut
Istanbul Gelisim Univercity

Research Article

Keywords: Choroidal thickness, feature selection, machine learning, obese children, optical coherence tomography

Posted Date: October 19th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-790701/v1

Abstract

Purpose: To analyse the effect of macular choroidal thickness (MCT) and peripapillary choroidal thickness (PPCT) on the classification of obese and healthy children by comparing the performance of Random Forest (RF) or Support Vector Machine (SVM), and Multilayer Perceptrons (MLP) algorithms.

Methods: 59 obese children and 35 healthy children aged 6 to 15 years were studied in this prospective comperative study using optical coherence tomography. MCT and PPCT were measured at 500 μm, 1000 μm, and 1500 μm distances from fovea and optic disc. Three different feature selection algorithms were used to determine the most prominent features of all extracted features. The classification efficiency of the extracted features was analyzed using RF, SVM, and MLP algorithms, demonstrating their efficacy for distinguishing obese from healthy children. The precision and reliability of measurements were assessed using Kappa analysis.

Result: Correlation Feature Selection algorithm produced the most successful classification results among the different feature selection methods. The most prominent features for distinguishing the obese and healthy groups from each other were PPCT temporal 500 μm, PPCT temporal 1500 μm, PPCT nasal 1500 μm, PPCT inferior 1500 μm, and subfoveal MCT. The classification rates for RF, SVM, and MLP algorithms were 98.6%, 96.8% and 89%, respectively.

Conclusion: Obesity-related metabolic alterations have an effect on the choriocapillaries of children, particularly in the subfoveal region and the outer semi-circle at 1500 μm from optic disc head. Both the RF and SVM algorithms are effective and accurate at classifying obese and healthy children.

Introduction

Childhood obesity is an exceedingly prevalent health issue in the world. The World Health Organization (WHO) has declared obesity to be an "escalating global epidemic" [1]. Worldwide, 22 million children under the age of five and 150 million school-age children have been reported to be severely overweight, with the prevalence of childhood obesity estimated to be 10% [2]. While there are several parameters that indicate the child's nutrition and growth status, the parameter recommended by WHO in 2006 is the Z-score. The Z-score system displays a set of standard deviations (SD) from the reference median or mean. It allows for more accurate assessments by standardizing measurements based on age and gender [3]. The Z-score system, which is used to determine children's nutritional and obesity status, can be calculated using a variety of anthropometric values such as weight-for-age-Z-scores (WAZ), height-for-age-Z-scores (HAZ), weight-for-height-Z-scores (WHZ), body mass index for age-Z-scores (BMIZ). BMIZ has been stated to be the most useful measure in evaluating obesity and adipose tissues [4]. However, in ophthalmological research, the body mass index (BMI) is the most widely used to classify children's nutrition and development.

Obesity has been associated with multiple ocular diseases, including cataract, glaucoma, dry eye, diabetic retinopathy, and age-related macular degeneration [5–7]. Although the cause of the connection

between obesity and eye disorders is unknown, it is thought that it could be due to obesity-related chronic oxidative stress, endothelial dysfunction, and vascular damage [6]. The choroid is the vascular layer of the eye and is responsible for approximately 70% of the blood supply of the ocular structures. The choroidal layer has many important roles, such as supplying oxygen and nutritional support for the outer retinal layers, thermoregulation and control of intraocular pressure [8]. Changes in choroidal thickness have been observed in various systemic diseases, including diabetes, hypertension, and endocrine diseases [9, 10]. There is not enough research in the literature studying the impact of obesity on the eyes. Just subfoveal choroidal thickness was assesed in a small number of ophthalmological studies on ocular changes in childhood obesity, but no comprehensive macular choroidal thickness (MCT) and peripapillary choroidal thickness (PPCT) assessment was performed.

As a result of the improvement of computing technology, artificial intelligence has begun to take place instead of conventional parametric tests in data analysis. Machine learning, the most important subset of artificial intelligence, makes it possible to interpret information, classify data and make predictions for the future by analyzing the structures and texture patterns of a large number of computer data [11, 12]. Various machine learning algorithms may be used to classify data, such as artificial neural networks, decision trees, vector machines and classifier ensembles. Machine learning algorithms have been found to be more efficient, effective and accurate than conventional statistical methods in the analysis of a large number of complex data [11, 13, 14]. The significance of algorithms of ensemble classification such as neural network ensembles, random forest, bagging and boosting is rising. Ensemble learning algorithms have the ability to make multiple classifications of data using the same or different classifiers that can be single or multiple [15, 16].

Deep learning algorithms are an alternative approach to artificial intelligence. But the performance of this approach depends on the quality and, in particular, the number of datasets [17]. Deep learning algorithms may only be used if there is a large amount of training data which is not easily obtainable in medical studies. Transfer learning is the reuse of a pre-trained model on a new dataset and performs well with a small training dataset. However, transfer learning is another way to reduce overfitting in models. Optimization of the hyper-parameters and determining the optimal learning rate of the neural net for different layers are very challenging tasks due to the limited knowledge about the relevance of the large number of architectural and training hyper-parameters. Moreover, the success rate of the transfer learning-based systems is highly dependent on the dataset used for training. To our knowledge, there are no pre-trained weights of a neural net similar to our dataset.

The determination and selection of features and the classification algorithm are decisive for machine discrimination abilities. It is common knowledge that all image structures have inherent features which influence segmentation and classification. In this study, we examined and compared the performance of Random Forest (RF), Support Vector Machine (SVM), and Multilayer Perceptrons (MLP) algorithms to classify MCT and PPCT in obese and healthy children. By this way, we aimed to examine the impact of childhood obesity on choroidal thickness and to recognise early clinical changes that could pose a risk for multiple ocular diseases by using machine learning algorithms, a modern method of analysis.

2.material And Method

This research was reviewed by an independent ethical review board and confirmed with the principles and applicable guidelines for the protection of human subjects in biomedical research.

In this prospective comperative study, healthy and obese children between 6 to 15 years of age who admitted to the Department of Pediatrics and Ophthalmology at Beylikduzu State Hospital for routine control were recruited from 1 June 2020 to 1 December 2020. The exclusion criteria are as follows: prevalence of chronic diseases such as diabetes, hypertension, heart disease, obstructive sleep apnea syndrome; history of any medication use; ocular diseases such as strabismus, cataracts, glaucoma, amblyopia, uveitis, optic disc anomaly, retinal disease; history of prior eye surgery; more than 2D of spherical or cylindrical refractive error; cornea, lens or vitreous opacity which does not allow qualified OCT imaging, children who do not cooperate well enough for OCT imaging.

2.1. Physical Examination

Height and weight measurements of patients were taken using a digital scale and a wall-mounted Harpender stadiometer. For children under the age of 5 years, Z-scores were measured by WHO Anthro software (www.who.int/childgrowth/software/en/index.html) to assess the participant's position on the age and sex matched development map [18]. For children over the age of 5 years, Z-scores were measured by WHO Anthro plus software (www.who.int/tools/growth-reference-data-for-5to19-years/application-tools). Obesity was defined as greater than + 2 SD, while normal was defined as between 1 and + 1 SD for both BMIZ and HAZ. Pubertal staging was performed using the Tanner procedure [19]. After the rest time, blood pressure was calculated by using the automatic sphygmomanometer (Omron M2 HEM7121E, Omron Healthcare Co, Japan) at least three times a 10-minute period. Blood pressure was measured as the average of a total of three consecutive measurements taken after the required resting time. Children with systolic and/or diastolic blood pressure levels greater than the 95th percentile were defined as hypertensive [20].

2.2. Ophthalmological Examination

A detailed ophtalmological examination, including measures of best-corrected visual acuity, spherical equivalent, slit-lamb-biomicroscophy, intraocular pressure (IOP), central corneal thickness (CCT), axial length (AXL), and anterior chamber depth (ACD) and OCT imaging was performed for each participant by an experienced opthalmologist. The research examined only the participants' right eyes. Autokeratorefractometry (Topcon KR-800, Topcon Medical Systems, Inc. Fukuoka, Japan) was used for refractive measurements. IOP was measured using Goldmann applanation tonometry and CCT was measured using a non-contact tonopachymeter (NT-530P, Nidek CO. Gamagori, Japan). AXL and ACD were measured using optic biometry (Nidek Axial Length-Scan, Nidek CO. Gamagori, Japan). Retinal and choroidal thicknesses were assessed using the Spectralis OCT method (Cirrus HD OCT, Carl Zeiss Meditec, Dublin, CA, USA) following further ophthalmological examinations.

All OCT imaging and evaluations were performed by the same experienced opthalmologist without pupil dilatation. Retinal thickness, mean ganglion cell layer (GCL) and inner plexiform layer (IPL) thickness were measured using automated segmentation values of the Spectralis OCT system with a macular cube position of 512x128. The OCT HD 1-line-edi protocol's high resolution scan through the fovea was used for MCT measurements. The choroidal thickness was assessed manually from the outer surface of the hyperreflective line referring to the retinal pigment epithelium to the sclera's inner layer. MCT measurements were performed at the fovea's center and at a distance of 500 μm, 1000 μm, and 1500 μm away from the center of the fovea in the nasal and temporal regions (Fig. 1). The scans were carried out in vertical and horizontal planes through the middle of the optic disc using the OCT HD 5-Line Raster-Edi protocol for PPCT assessment [21]. In this scan, the optic disc is divided into two equal sections in both the horizontal and vertical planes. Then, in each of the nasal, temporal, superior, and inferior regions, PPCT measurements were taken at distances of 500 μm, 1000 μm, and 1500 μm from the optical disc's boundary (Fig. 1). Both MCT and PPCT measurements were performed by two masked ophthalmologists (EU, OD) for inter-observer reproducibility at 100 per cent magnification during different sessions. The OCT Disc Cube 200x200 protocol was used for retinal nerve fiber layer thickness (RNFLT) and cup to disc ratio analysis. The superior, inferior, nasal, temporal and average RNFLT were calculated automatically. Both measurements were taken between 9:00 a.m. and 11:00 a.m. to eliminate diurnal differences.

2.3. Feature Extraction and Selection

We manually measured the image features that we believe affect our hypothesis and tested whether these parameters validated our hypothesis or not. All of the manually extracted features were given in Table 1.

Table 1
All Extracted Features

Physical examination based features	Ocular examination based features	OCT images based PPCT features	OCT images based MCT features	OCT images based other features
Age	Spherical eqivalent	PPCT Temporal 500	MCT Fovea	GCL + IPL complex
Gender	AXL	PPCT Temporal 1000	MCT Temporal 500	MT
Height	ACD	PPCT Temporal 1500	MCT Temporal 1000	Average c/d
Weight	IOP	PPCT Nasal 500	MCT Temporal 1500	Vertical c/d
BMI	Pachymeter	PPCT Nasal 1000	MCT Nasal 500	RNFLT Temporal
BMIZ		PPCT Nasal 1500	MCT Nasal 1000	RNFLT Nasal
HAZ		PPCT Superior 500	MCT Nasal 1500	RNFLT Superior
Systolic blood pressure		PPCT Superior 1000		RNFLT Inferior
Diastolicblood pressure		PPCT Superior 1500		RNFLT Average
		PPCT Inferior 500		
		PPCT Inferior 1000		
		PPCT Inferior 1500		

ACD: anterior chamber depth, **AXL**: axial length, **BMI**: body mass index, **BMIZ**: Z score for body mass index, **GCL + IPL complex**: Ganglion cell layer + Internal plexiform layer complex, **HAZ**: Z score for height, **IOP**: intraocular pressure, **MCT**: macular choroidal thickness, **MT**: macular thickness, **PCT**: peripapillary choridal thickness, **RNFLT**: retinal nerve fiber layer thickness.

Feature selection approaches are designed to minimize the number of parameters to those deemed most useful for the model in order to predict the target variable. Feature selection primarily focuses on removing non-informative or irrelevant predictors from the model. To produce an easier and faster classification system, we used three feature selection algorithms: Variable Ranking (VR), Correlation Feature Selection (CFS), and Principle Component Analysis (PCA). All the extracted features were entered into the VR, CFS, and PCA algorithms, and the most prominent features were selected to form the feature vector. This feature vector is used as an input for the classification algorithms (Fig. 2).

2.4. Classifiers for Machine Learning

We analyzed and compared the efficiency of RF, SVM and MLP in order to distinguish selected features, whether or not they are discriminatory for obese and healthy children. RF is a grouping, correlation, and other task-specific ensemble learning process. It works through building a large number of decision trees during training and then extracting the test [22].

SVM is a regulated classification algorithm with corresponding learning methods that require appropriate data for classification and correlation analysis. Through providing a training dataset, the SVM algorithm attributes features to just one or another subclass, and makes it a binary and linear classifier which is not predictable. The SVM algorithm enables learning of multidimensional functions and non-linear classifications successfully [23].

MLP is a well-known correlation algorithm for determining the relationship between a continuous dependent variable and two or more independent variables [24]. It will be used to determine which variable has the largest influence on the expected output and which variable now relates to each other.

2.*5. Statistical Analysis*

The distribution of parameters in two different classes was tested using the Shaphiro Wilk test. The Student t-test was used to compare parameters with normal distribution in two independent groups, and the Mann Whitney-u test was used to compare non-normally distributed parameters in two independent groups.

3.results

This study includes 59 (35 female, 24 male) patients with obesity as the study group and 35 (21 female, 14 male) healthy subjects as the control group.

The classification efficiency of different systems is influenced by their capabilities in data classification. In addition, to achieve the highest efficiency, we employed various feature selection algorithms such as VR, CFS, and PCA. First, we analyzed a variety of features to choose which ones were most effective for the classification of healthy and obese groups. Second, depending on the chosen features, the device compared various machine learning algorithms, such as RF, SVM, and MLP. The efficiency of the different versions can vary, since they are structured differently.

CFS algorithm produced the most successful classification results among the three different feature selection methods. The CFS algorithm determined that subfoveal choroidal thickness is the most distinguishing feature, along with PPCT measurement locations including temporal 500 μm, temporal 1500 μm, nasal 1500 μm, and inferior 1500 μm. In addition to these features, PCA algorithm selected the spherical equivalent value feature. However, when the spherical equivalent feature was absent, the classification results showed a higher success rate.

A 10-fold cross-validation process was used to test the stability and reliability of the RF, SVM, and MLP algorithms. The dataset was divided into two sections, with 70% of the data used for training and 30% for

testing. To reduce selection bias, random sampling was conducted ten times to generate separate training and testing sets from the dataset.

The confusion matrix and classification rates of the RF algorithm to classify choridal thickness as normal or obese were shown in Table 2. The overall accuracy rate of our system based on RF was 98.9%.

Table 2
Classification results of obese and healthy children based on choroidal thickness by Random Forest Algorithm

Class	TP Rate	FP Rate	Precision	Recall	F Measure	Confusion Matrix	
Obese	1	0.029	0.983	1	0.992	59	**0**
Normal	0.971	0.000	1	0.971	0.986	**1**	34
Weigted Average	0.989	0.018	0.990	0.989	0.989		
TP: True positive, FP: False positive							

The classification accuracy of SVM and the confusion matrix to classify choroidal thickness as normal or obese were demonstrated in Table 3. The overall accuracy of our system based on SVM was 96.8%. Although the RF and SVM algorithms were equally successful at classifying the healthy group, RF was more successful in tagging the obese group. While the RF algorithm identified all obese data sets correctly as healthy, the SVM algorithm incorrectly classified two obese datasets as healthy. The thickness of the choroidal layer differed between obese and healthy children, and this difference was crucial in classifying groups using both RF and SVM algorithms.

Table 3
Classification results of obese and healthy children based on choroidal thickness by Support Vector Machine Algorithm

Class	TP Rate	FP Rate	Precision	Recall	F Measure	Confusion Matrix	
Obese	0.966	0.029	0.983	0.966	0.974	57	**2**
Normal	0.971	0.034	0.944	0.971	0.958	**1**	34
Weigted Average	0.968	0.031	0.968	0.968	0.968		
TP: True positive, FP: False positive							

The confusion matrix of the MLP algorithm is given in Table 4. It shows to classiffication rates of as healthy or obese. The overall accuracy of our system based on MLP was 89.4%. Despite using different learning rates and architecture, its success just increased from 85.83–89.36%. The reason of small change is most probably because of the limited dataset, falls into the local extremum and lacks spatial information.

Table 4
Classification result of obese and healthy children based on choroidal thickness by Multilayer Perceptrons Algorithm

Class	TP Rate	FP Rate	Precision	Recall	F Measure	Confusion Matrix	
Obese	0.983	0.257	0.866	0.983	0.921	58	**1**
Healthy	0.743	0.017	0.963	0.743	0.839	**9**	26
Weigted Average	0.894	0.168	0.902	0.894	0.890		
TP: True positive, FP: False positive							
Tables Legends:							

Precision (positive predictive), recall (sensitivity), and F-measure can be used to measure the relevance of a classification system based on artificial intelligence. The proportion of relevant instances among the retrieved instances is known as precision. The proportion of relevant instances that were retrieved is known recall. Unlike precision, which only considers accurate positive predictions out of all positive predictions, recall considers failed positive predictions.

The F-Measure gives us the harmonic mean of the values of precision and recall. The main purpose of using the F-Measure value is not to choose an inaccurate model of datasets that are not uniformly distributed. The F-measure is a means to integrate precision and recall into one measure capturing all properties.

The overall RF and SVM precision rates were 98.9% and 96.8%, respectively. However, the overall precision rate of the MLP system was 89.4%, relatively unsatisfactory. Similarly, the overall F-Measurement results of RF and SVM were both high (98.9% and 96.8%), whereas the result of MLP was low (89%). The overall recall rates for the RF and SVM systems were 98.9% and 96.8%, respectively. However, recall values for the obese group for the RF and SVM systems were 100% and 96.6%, respectively, which confirms the power of the proposed system's capability to recognize choroidal thickness measurements. (Table 2–3). The average recall rate for the MLP system was 89.4%. However, the recall values of the obese and healthy groups were 98.3% and 74.3%, respectively (Table 4).

We conducted a Kappa analysis to assess the reliability and accuracy of our measurements. The Kappa value ranges from 0 to + 1. System reliability improves when the kappa value converges to 1 [25]. The Kappa coefficients for RF, SVM and MLP were 97.71%, 93.05%, and 76% respectively.

4. Discussion

According to the results of the current study, which compared obese children to a healthy control group, obesity had an effect on choroidal thickness at distinct measurement points, but not at all measurement points. Obesity-related metabolic alterations have an effect on the choriocapillaries, particularly in the subfoveal region and the outer semi-circle at 1500 µm from optic disc head. This research is noteworthy

because it is not only evaluating MCT and PPCT in obese children, but it also utilizes machine learning algorithms in their analysis.

There are only a few studies in the literature that assess the impact of childhood obesity on ocular structures. Baran et al. [26] found that obese children had higher IOP and lower RNFLT than healthy children. They reported that childhood obesity may contribute to the development of glaucoma. They assessed choroidal thickness in the central subfoveal region alone and discovered no statistically significant differences. However, they did not conduct a comprehensive evaluation of MCT and PPCT. Bulus et al. [27] discovered that obese children have thicker MCT than healthy children, but did not evaluate PPCT. Additionally, they also used the BMI standard deviation score, which is equal to the BMIZ for childhood nutrition and growth classification reported by the World Health Organization in 2006. Buluş et al. reported that there was a strong positive correlation between BMI standard deviation score and subfoveal MCT. Consistent with this study, we found that subfoveal MCT is affected by obesity and is a distinguishing feature between the obese and control groups.

While there are several literature studies assessing the MCT in various diseases, there are few studies evaluating the PPCT. Read et al. [28] identified normal PPCT values and variations in healthy children and confirmed that myopic refractive errors cause thinning in PPCT. Ozcimen et al. [29] documented thinning in both PPCT and MCT in chronic obstructive pulmonary diseases. He claimed that choroidal thinning is caused by vascular resistance due to hypoxia. Koma et al. [30] evaluated PPCT and subfoveal choroidal thickness in healthy and glaucoma subjects using spectral domain OCT and swept source OCT. She discovered that choroidal thickness was significantly thicker in glaucoma subjects than in controls in the peripapillary region, but not in the macular region, using swept source OCT.

This is the first research that we are aware of that evaluates PPCT in childhood obesity. Furthermore, conventional statistical methods have been employed in previous studies, including the choroidal evaluation of various disorders. There is no prior study in the current literature that evaluates both MCT and PPCT by using machine learning algorithms.

In machine learning, feature selection helps boost classification efficiency by avoiding over-fitting, creating a time-saving model, and making the designed model more human-friendly. There are several feature selection approaches in the literature to minimize the number of features for classification purposes. Different subsets can be created with each feature selection method. We ran all of the data through a feature selection process using three different algorithms: VR, CFS, and PCA. None of the parameters associated with MCT and PPCT were excluded in any of the three analyses, and they were found to be distinctive in all of them. According to the results, obese and healthy children have significantly different choroidal thicknesses at specific measurement points. These measurement points were PPCT temporal 500 μm, PPCT temporal 1500 μm, PPCT nasal 1500 μm, PPCT inferior 1500 μm, and subfovealregions. The spherical equivalent value was chosen in the PCA algorithm in addition to the distinguishing features chosen in the CFS algorithm. There was no statistically significant difference between the two groups' spherical equivalent values ($p > 0.05$). The CFS algorithm outperforms PCA in

classification because of the spherical equivalent value is not a distinguishing feature for these groups. While machine learning algorithms identify distinct features in classification for the two groups, they do not show the relative value of these features in each group. Due to the fact that machine learning algorithms reveal the importance of features, classification is performed on all of the selected features.

In this study, the results of three different classification algorithms, which included RF, SVM, and MLP were compared. Because it is difficult to predict which machine learning algorithm will perform better in classification. We selected RF because it is a good comparison and classification technique and can very well detect outliers. SVM is a very robust technique for solving high-dimensional problems and creating accurate classifications. MLP is an attainable technique with the ability to create a simple architecture, easily build it and quickly calculate the model. The risk of being introduced into the local extremum, weak overfitting skills, a lack of theoretically-based rigid design programs, and difficulty managing the training program are disadvantages of the MLP algorithm. SVM may be more determinant in some cases, even though the RF algorithm is generally more successful in classification. Due to the limited and unbalanced datasets used in this study, we encountered some difficulties when implementing SVM and MLP algorithms. To overcome this challenging situation, we focused on kernel selection, which had an effect on the kernel's success in implementing the SVM algorithm. We used polynomial and radial base kernels to improve classification efficiency by reducing our margin of error. Additionally, the MLP algorithm's success was influenced by the network structure. The more complicated the network's structure, the more successful it will be. However, we did not increase the number of layers in order to reduce the margin of error.

While RF produces better results against outliers and noise than SVM, it is not as successful in handling the dataset imbalance problem. Although our dataset was slightly unbalanced, the results with RF were quite successful. MLP was found to be less successful than SVM and RF in the classification according to the choroidal thickness.

The MLP algorithm had the highest rate of misclassification of all of the other classification techniques. The MLP algorithm misclassified the ten samples. At the same time, three of them were also misclassified by the SVM algorithm. We found no similarities in terms of features such as height or weight in cases misclassified by the MLP algorithm. In terms of group classification, we discovered that the SVM algorithm outperforms the MLP algorithm. The main reason for misclassification based on the SVM algorithm is that they were at the limit of obesity according to the BMIZ value (respectively 2,01 and 2,02). As a result, classification success of SVM algorithm is higher in obese cases with a high BMIZ value.

The performance of machine learning algorithms, as well as the complexity of the models used, are influenced by the quality and quantity of data. To the best of our knowledge, there is no open dataset in the literature that is comparable to our dataset. The drawback of our analysis is the limited size of the dataset. However, the majority of medical research faces difficulty in achieving a sufficient number of cases. Obtaining large quantities of high-quality data for medical research is a time-consuming and

difficult task. There are medical research in the literature that use machine learning algorithms with small number of datasets. Hidalgo et al. [31] used machine learning algorithms to classify keratoconus using 5 Pentacam-derived parameters of 131 eyes. An et al. [32] developed classification criteria that could aid in the clinical management of glaucoma by using machine learning algorithms to classify 163 glaucomatous optic discs. Cartes et al. [33] evaluated the variability of tear osmolarity in 20 patients with dry eye using machine learning techniques. It has been demonstrated that machine learning algorithms can conduct self-diagnosis and classification analyses of OCT images with high accuracy, speed, and consistency [34]. However, in the classification tests, we measured Kappa values to ensure that the small dataset did not affect the reliability of our results and to maximize success. The Kappa value is a measure that contrasts the observed precision with the predicted precision (random chance). This is a far more reflective indicator of model efficiency. Kappa values were measured as 97.71%, 93.05%, and 76% for the RF, SVM, and MLP analyses, respectively. According to the Kappa statistics, RF is the most accurate test, but the reliability of SVM is also very similar to RF. Despite the limited number of datasets, Kappa analyses showed that both RF and SVM were very successful and reliable in the classification of obese and healthy children.

5.conclusion

The current study indicates that MCT and PPCT differ from each other in obese and healthy children and are effective at categorizing two groups using machine learning algorithms, especially when RF or SVM algorithms were used. Additionally, obesity has been shown to impact choroidal thickness at some distinct measurement points, when compared to healthy children. The current study emphasizes the importance of subfoveal MCT, as well as PPCT measurement regions, including temporal 500 μm, temporal 1500 μm, nasal 1500 μm, and inferior 1500 μm, in classifying children as obese or healthy. To improve classification performance, further deep learning studies with larger datasets are needed.

Declarations

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study was approved by the Ethics Committee of Biruni University (Document number: 2020/40-06)

Consent to participate Informed consent and oral consent was obtained from all individual participants and/or their legal guardians

Consent to publish Additional informed consent was obtained from all individual participants and/or their legal guardians for whom identifying information is included in this article

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest The authors declare that there are no any competing financial interests in relation to the work described

Acknowledgements

Special Thanks for Aysun Sezer who is postdoctoral researcher at École nationale supérieure de techniques avancées (ENSTA) in Unité d'Informatique et Ingénierie des Systèmes. She worked as a data scientist and supported all statistical tests and machine learning algorithms for our research.

References

1. Kosti RI, Panagiotakos DB (2006) The Epidemic of Obesity in Children and Adolescents in the World. Cent Eur J Public Health 14:151–159
2. Lobstein T, Baur L, Uauy R (2004) Obesity in Children and Young People: A Crisis in Public Health. Obes Rev 1:4–104
3. Mei Z, Grummer-Strawn LM (2007) Standard Deviation of Anthropometric Z-Scores as a Data Quality Assessment Tool Using the 2006 Who Growth Standards: A Cross Country Analysis. Bull World Health Organ 85:441–448
4. Cole TJ, Faith MS, Pietrobelli A, Heo M (2005) What Is the Best Measure of Adiposity Change in Growing Children: Bmi, Bmi %, Bmi Z-Score or Bmi Centile? Eur J Clin Nutr 59:419–425
5. Areds Study Group (2001) Risk Factors Associated with Age-Related Nuclear and Cortical Cataract: A Case-Control Study in the Age-Related Eye Disease Study, Areds Report No. 5. J Ophthalmol 108:1400–1408
6. Abramson N, Abramson S (2001) Hypercoagulability: Clinical Assessment and Treatment. South Med J 94:1013–1020
7. Karti O, Nalbantoglu O, Abali S et al (2017) The assessment of peripapillary retinal nerve fiber layer and macular ganglion cell layer changes in obese children: a cross-sectional study using optical coherence tomography. Int Ophthalmol 37:1031–1038
8. Nickla DL, Wallman J (2010) The Multifunctional Choroid. Prog Retin Eye Res 29:144–168
9. Aiello LP, Avery RL, Arrigg PG et al (1994) Vascular Endothelial Growth Factor in Ocular Fluid of Patients with Diabetic Retinopathy and Other Retinal Disorders. N Eng J Med 331:1480–1487
10. Alam AA, Mitwalli AH, Al-Wakeel JS et al (2004) Plasma Fibrinogen and Its Correlates in Adult Saudi Population. Saudi Med J 25:1593–1602
11. Hansen M, Dubayah R, Defries R (1996) Classification Trees: An Alternative to Traditional Land Cover Classifiers. Int J Remot Sens 17:1075–1081
12. Huang C, Davis LS, Townshend JRG (2002) An Assessment of Support Vector Machines for Land Cover Classification. Int J Remot Sens 23:725–749
13. Foody GM (2009) Sample Size Determination for Image Classification Accuracy Assessment and Comparison. Int J Remot Sens 30:5273–5291

14. Friedl MA, Brodley CE, Strahler AH (1999) Maximizing Land Cover Classification Accuracies Produced by Decision Trees at Continental to Global Scales. IEEE Trans Geosci Remote Sens 37:969–977
15. Zhang Z, Krawczyk B, Garcìa S et al (2016) Empowering One-Vs-One Decomposition with Ensemble Learning for Multi-Class Imbalanced Data. Knowl-Based Syst 106:251–263
16. Vong CM, Du J (2020) Accurate and Efficient Sequential Ensemble Learning for Highly Imbalanced Multi-Class Data. Neural Netw 128:268–278
17. Medeiros FA, Jammal AA, Thompson AC (2019) From Machine to Machine: An Oct-Trained Deep Learning Algorithm for Objective Quantification of Glaucomatous Damage in Fundus Photographs. J Ophthalmol 126:513–521
18. Edalat A, Abbaszadeh M, Eesvandi M, Heidari A (2014) The Relationship of Severe Early Childhood Caries and Body Mass Index in a Group of 3- to 6-Year-Old Children in Shiraz. J Dent 15:68–73
19. Tanner JM, Whitehouse RH (1976) Clinical Longitudinal Standards for Height, Weight, Height Velocity, Weight Velocity, and Stages of Puberty. Arch Dis Child 51:170–179
20. Lurbe E, Agabiti-Rosei E, Cruickshank JK et al (2016) European Society of Hypertension Guidelines for the Management of High Blood Pressure in Children and Adolescents. J Hypertens 34:1887–1920
21. Ho J, Branchini L, Regatieri C et al (2011) Analysis of Normal Peripapillary Choroidal Thickness Via Spectral Domain Optical Coherence Tomography. J Ophthalmol 118:2001–2007
22. Svetnik V, Liaw A, Tong C et al (2003) Random Forest: A Classification and Regression Tool for Compound Classification and Qsar Modeling. J Chem Inf Comput Sci 43:1947–1958
23. Jayadeva, Khemchandani R, Chandra S (2007) Twin Support Vector Machines for Pattern Classification. IEEE Trans Pattern Anal Mach Intell 29:905–910
24. Liu M, Wang M, Wang J (2013) Comparison of Random Forest, Support Vector Machine and Back Propagation Neural Network for Electronic Tongue Data Classification: Application to the Recognition of Orange Beverage and Chinese Vinegar. **Sens Actuators B Chem** 177:970–980
25. Landis JR, Koch GG (1977) The Measurement of Observer Agreement for Categorical Data. Biometrics 33:159–174
26. Baran RT, Baran SO, Toraman NF et al (2019) Evaluation of Intraocular Pressure and Retinal Nerve Fiber Layer, Retinal Ganglion Cell, Central Macular Thickness, and Choroidal Thickness Using Optical Coherence Tomography in Obese Children and Healthy Controls. Niger J Clin Pract 22:539–545
27. Bulus AD, Can ME, Baytaroglu A et al (2017) Choroidal Thickness in Childhood Obesity. Ophthalmic Surg Lasers Imaging Retina 48:10–17
28. Read SA, Alonso-Caneiro D, Vincent SJ, Collins MJ (2015) Peripapillary Choroidal Thickness in Childhood. Exp Eye Res 135:164–173
29. Ozcimen M, Sakarya Y, Kurtipek E et al (2016) Peripapillary Choroidal Thickness in Patients with Chronic Obstructive Pulmonary Disease. Cutan Ocul Toxicol 35:26–30

30. Komma S, Chhablani J, Ali MH et al (2019) Comparison of Peripapillary and Subfoveal Choroidal Thickness in Normal Versus Primary Open-Angle Glaucoma (Poag) Subjects Using Spectral Domain Optical Coherence Tomography (Sd-Oct) and Swept Source Optical Coherence Tomography (Ss-Oct). BMJ Open Ophthalmol 4:e000258
31. Ruiz Hidalgo I, Rozema JJ, Saad A et al (2017) Validation of an Objective Keratoconus Detection System Implemented in a Scheimpflug Tomographer and Comparison with Other Methods. Cornea 36:689–695
32. An G, Omodaka K, Tsuda S et al (2018) Comparison of Machine-Learning Classification Models for Glaucoma Management. J Healthc Eng 6874765
33. Cartes C, López D, Salinas D et al (2019) Dry Eye Is Matched by Increased Intrasubject Variability in Tear Osmolarity as Confirmed by Machine Learning Approach. Arch Soc Esp Oftalmol 94:337–342
34. Tan Z, Scheetz J, He M (2019) Artificial Intelligence in Ophthalmology: Accuracy, Challenges, and Clinical Application. Asia-Pac J Ophthalmol (Phila) 8:197–199

Figures

Figure 1

Example of macular and peripapillary choroidal thickness measurements (right eye). 1 Macular choroidal thickness was measured at the central fovea. Right: Lines denote placements for the nasal (left) and temporal (right) quadrants. Left: Line denotes where scan was taken relative to the fundus. 1 Peripapillary choroidal thickness measurements at horizontal planes through the middle of the optic disc. Right: Lines denote placements for the nasal (left) and temporal (right) quadrants. Left: Line denotes where scan was taken relative to the fundus. 4) Peripapillary choroidal thickness measurements at vertical planes through the middle of the optic disc. Right: Lines denote placements for the superior (right) and inferior(left) quadrants. Left: Line denotes where scan was taken relative to the fundus

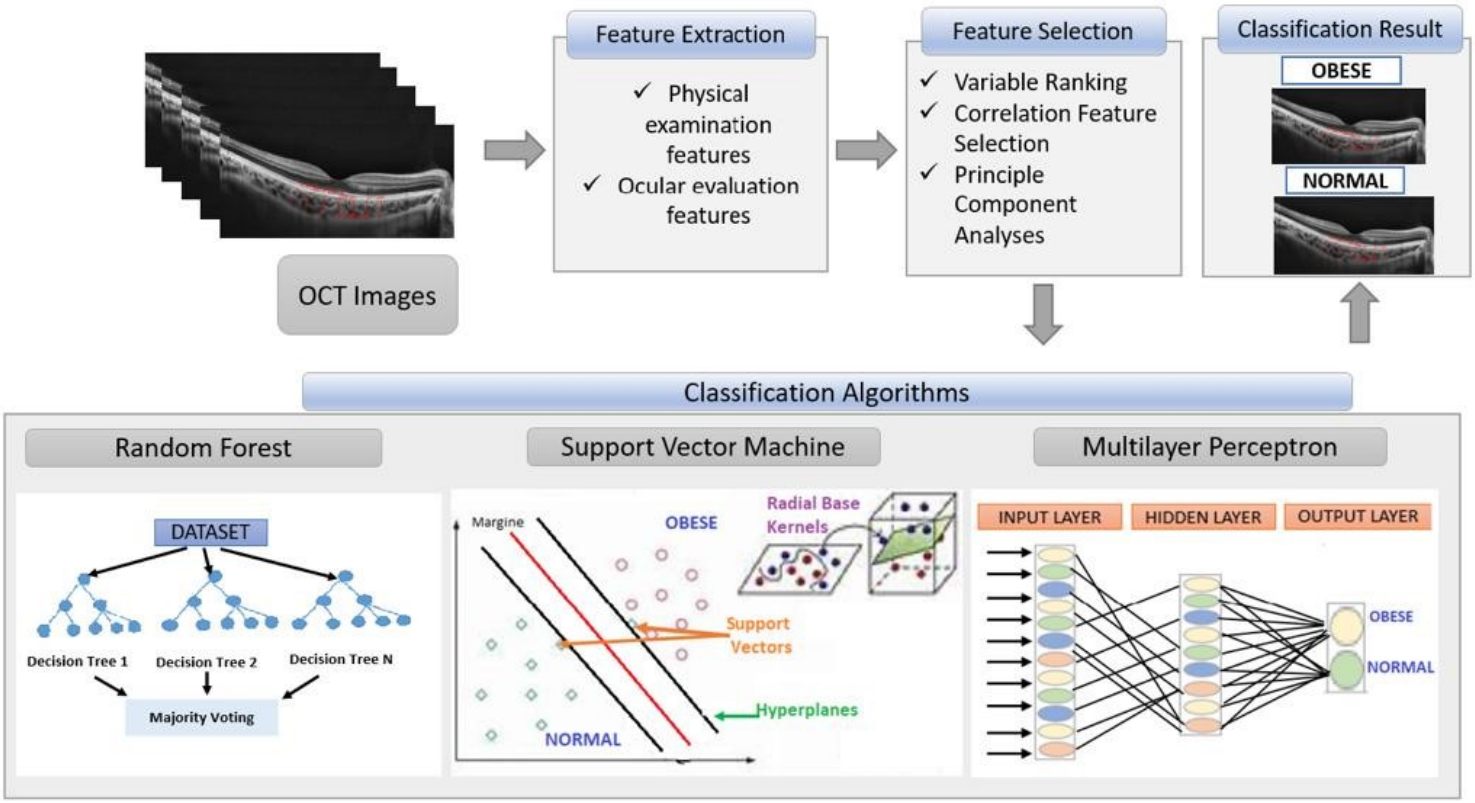

Figure 2

Flow chart of proposed recognition system