■ RESEARCH ARTICLE

Prediction of Scoliosis Risk in Adolescents with Machine Learning Models

Murat Ali Çınar [a†], **İbrahim Küçükcan** [b]

[a] Department of Physiotherapy and Rehabilitation, Hasan Kalyoncu University, Gaziantep, Türkiye
[b] Ütopya Physiotherapy Consultancy Center, Gaziantep, Türkiye
[†] muratali.cinar@hku.edu.tr, corresponding author

RECEIVED MARCH 25, 2024
ACCEPTED APRIL 16, 2024

CITATION Çınar, M.A. & Küçükcan, İ. (2024). Prediction of scoliosis risk in adolescents with machine learning models. *Artificial Intelligence Theory and Applications, 4(*1), 33-42.

Abstract

When considering classifications of scoliosis, 'idiopathic scoliosis' emerges as the most common type. Alongside spinal alterations, individuals with scoliosis undergo changes in stability and gait while standing. Despite existing literature exploring the progression of scoliosis and its impact on foot pressure among those diagnosed with adolescent idiopathic scoliosis, no studies have been found regarding the prediction of scoliosis risk in healthy adolescents. This study aims to develop an machine learning based decision support system capable of forecasting scoliosis risk in adolescents using foot pressure analysis values and machine learning models.

The study encompassed 20 patients diagnosed with adolescent idiopathic scoliosis and 43 healthy adolescents exhibiting similar demographic characteristics, totaling 63 patients. Plantar pressure distributions of all participants were measured both statically and dynamically.

The data collected for all patients comprised: age, sex, percentage of right hindfoot static plantar pressure, percentage of left hindfoot static plantar pressure, percentage of right forefoot static plantar pressure, percentage of left forefoot static plantar pressure, percentage of right foot dynamic plantar pressure, and percentage of left foot dynamic plantar pressure. A dataset including pressure percentages and the presence of scoliosis diagnosis was constructed, consisting of 8 input variables and 1 outcome variable for each patient.

The most effective predictors of adolescent idiopathic scoliosis risk were identified as follows: Subspace KNN (100%), RUS Boosted Trees (100%), Weighted KNN (100%), Bagged Trees (100%), and Fine KNN (100%)

Keywords: adolescent idiopathic scoliosis; plantar pressure distribution; machine learning

1. Introduction

The term "scoliosis" comes from the Greek and means "crooked" or "curved". It was first defined and introduced into the literature by Hippocrates [1-3]. In the Scientific Society on Scoliosis Orthopedic and Rehabilitation Treatment (SOSORT) guide published in 2016, scoliosis was described as a group of conditions that result in various deformities in the shape of the spine, thorax and trunk [1]. Although there are different definitions of scoliosis, they all converge on the fact that it involves a lateral curvature of the spine of more than 10° (Figure 1) [1-4]. There are many classifications of scoliosis; in 1973, the Scoliosis Research Society (SRS) divided scoliosis into two groups: structural and non-

Artificial Intelligence Theory and Applications, ISSN: 2757-9778. ISBN: 978-605-69730-2-4 © 2024 İzmir Bakırçay University

structural [4].In non-structural functional scoliosis, spinal curvature develops due to causes outside the spine. There is often shortness of the lower extremities or asymmetry in the tone of the paraspinal muscles. A person with non-structural scoliosis can correct posture. In structural scoliosis, the person has a loss of flexibility and needs treatment to correct the curvature [1-4]. This classification by the SRS is shown in Table-1 [1-4].

Figure 1. Radiological image of an individual with scoliosis [1-4].

Table 1. Scoliosis classification of SRS.

Structural scoliosis	**Non-structural scoliosis (functional scoliosis)**
Idiopathic scoliosis Infantile Juvenile Adolescent	Postural scoliosis
Neuromuscular scoliosis Neuropathic Myopathic	Hysterical
Congenital scoliosis	Caused by nerve root irritation
Neurofibromatosis	Caused by hip contractures
Scoliosis due to connective tissue disorder	Caused by leg length inequality
Osteochondrodystrophy	Inflammatory (appendicitis etc.) related
Due to metabolic disorders	
Traumatic	
Scoliosis caused by tumors or infection	
Scoliosis due to rheumatic diseases	
Scoliosis due to pathologies in the lumbosacral region	

When looking at classifications of scoliosis, "idiopathic scoliosis" appears to be the most common type of scoliosis [1-4]. This term, which was introduced into the literature in 1922, is defined as situations in which no specific disease-causing deformity of the spine can be found [5].

Idiopathic scoliosis is divided into four groups based on the age of onset: infantile (0-3 years), juvenile (3-10 years), adolescent (10-18 years) and adult (18 years and older) [4-6].

Table 2. Classifications of idiopathic scoliosis [4-6]

Chronological	**Angular**		**Topographic**		
			Apex		
Age at diagnosis	**Cobb degrees**			**from**	**to**
Infantile	Low	Up to 20	Cervical	–	Disc C6–7
Juvenile	Moderate	21–35	Cervico-thoracic	C7	T1
Adolescent	Moderate to severe	36–40	Thoracic	Disc T1–2	Disc T11–12
Adult	Severe Severe to very severe Very severe	41–50 51–55 56 or more	Thoraco-lumbar Lumbar	T12	L1 Disc L1–2

Relationship Between Scoliosis and Plantar Pressure

In patients with scoliosis, changes in the spine are accompanied by changes in stability during standing and walking [7]. Because scoliosis affects the biomechanics of the spine in three dimensions, changes in spinal mobility and posture occur, causing movement patterns to change with each step [6-8]. The deformed spine shifts the body's center of mass to help maintain trunk balance, resulting in asymmetry and various gait abnormalities [6-8].

When examining the biomechanics of gait, the literature indicates that the pelvis and spine are intimately involved in the gait process [9]. A study conducted in 2023 highlighted that both static and dynamic plantar pressures are abnormally altered in individuals with adolescent idiopathic scoliosis and require treatment [10]. It has been suggested in the literature that these changes in plantar pressure may aid in the diagnosis of scoliosis, highlighting the need for further research in this area [11-13].

While there have been studies on the progression of scoliosis and its effects on base pressure in individuals diagnosed with adolescent idiopathic scoliosis in the literature, no studies have been encountered predicting the risk of scoliosis in healthy adolescents. The aim of this study is to develop an artificial neural network (ANN)-based decision support system that can predict the risk of scoliosis in adolescents using foot pressure analysis values and machine learning models. In addition, the dataset obtained from this study can serve as a preliminary study for researchers working in the field of scoliosis who wish to conduct research in the field of artificial intelligence.

2. Materials and Methods

The study conducted at Hasan Kalyoncu University Physiotherapy and Rehabilitation Department included 20 patients who were diagnosed with adolescent idiopathic scoliosis (AIS) by a specialist physician and who applied to Gaziantep Utopya Physiotherapy Consultancy Center to receive physiotherapy. 43 healthy adolescent individuals with similar demographic characteristics to the 20 included patients were also included in the study, and a data set was created with the data of a total of 63 participants.

The plantar pressure distributions of all participants were measured using two methods: static and dynamic. Both static and dynamic measurements were performed using the Ottobock Esco Scan device (Germany) and the Presto-Scan, Class I Rule 1, per MDD 93/42/EEC Annex IX, USA software (see Figure 1). The device is approximately 5 mm thick and has a sensor area of 44 x 37 cm with a total of 2288 sensors. It uses resistive sensor technology and can collect pressure and force data up to forty Hertz. Static measurements were taken while the subjects stood in a relaxed position, concentrating on a fixed point in front of them. Percentage values of the total contact area of both feet, including forefoot and hindfoot, were obtained through static evaluation. For dynamic measurements, a two-step protocol was used, utilizing the device's ability to colour code foot pressure points based on pressure percentages.

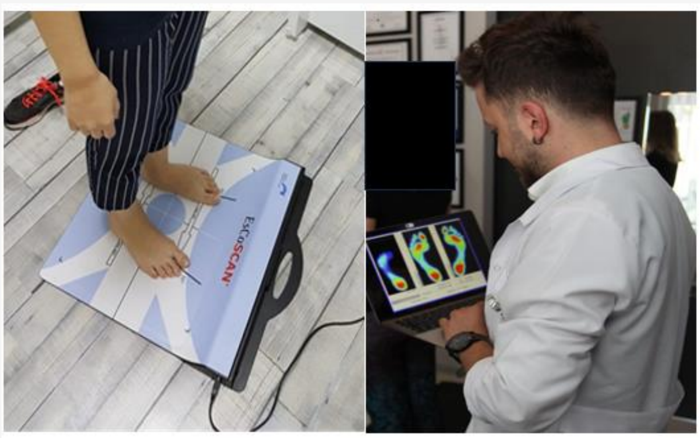

Figure 2 Plantar pressure analysis (representative image)

The data set for all patients was compiled with the following information: age, sex, percentage of right hindfoot static plantar pressure, percentage of left hindfoot static plantar pressure, percentage of right forefoot static plantar pressure, percentage of left forefoot static plantar pressure, percentage of right foot dynamic plantar pressure, percentage of left foot dynamic plantar pressure, and presence of scoliosis diagnosis (see Table 3). The dataset was randomly divided into two separate sets: 70% of the data was allocated for training the artificial neural network (ANN), while the remaining 30% was reserved for testing the model's performance. Table 3 below presents the dataset, comprising 8 input variables and 1 output variable collected for each patient.

Table 3. Dataset of features collected from patients

Features
Gender
Age
Right hindfoot static plantar pressure percentage
Left hindfoot static plantar pressure percentage
Right forefoot static plantar pressure percentage
Left forefoot static plantar pressure percentage
Right foot dynamic plantar pressure percentage
Left foot dynamic plantar pressure percentage
Outcome- Adolescent idiopathic scoliosis

The study was approved by the Hasan Kalyoncu University Health Sciences Ethics Committee. Informed consent forms were signed and permission to use the data was obtained from all patients included in the study.

2.1 Accuracy

Accuracy is a common metric for assessing the performance of a model, but there are situations where it should be considered. In particular, if there are unbalanced classes (i.e. large differences in sample counts between classes), accuracy may not be a sufficient metric and other metrics (e.g. precision, sensitivity) should also be considered.

In this study, the accuracy metric is used to evaluate the performance of machine learning methods.

2.2 K-fold cross validation

K-fold cross validation is a widely used method for evaluating the performance of a machine learning model. In this method, the data set is randomly divided into k parts (usually 5 or 10). Then one of these k parts is used as the test set, while the other k-1 parts are used as the training set. The model is trained once, and each time a different part is selected as the test set. The results are combined and the overall performance is measured. This method is used to assess how generalizable the model is, as it is tested on different pieces of data. In this way, the overall performance of the model can be more reliably assessed without relying on a single test set. In this study, 3,5 and 10 k were tested.

3. Experimental Results

This study used 25 different machine learning techniques. Each algorithm used different activation functions, optimization algorithms and loss functions, as detailed in Table 4, with or without PCA. All these algorithms were implemented using the machine learning toolbox available in the MATLAB programming language. The numerical results were derived using MATLAB R2021b on an Intel processor running on the Windows 10 platform.

The best performers in predicting the risk of adolescent idiopathic scoliosis were determined to be: Subspace KNN (100%), RUS Boosted Trees (100%), Weighted KNN (100), Bagged Trees (100%), Fine KNN (100%). In addition, the PCA method was used to try different parameter variations and the best results are shown in Table 4. The values of the most and least successful algorithms (confusion matrix) of the dataset are shown in Figures 3 and 4. In this study, the Principal Component Analysis (PCA) method has been used to demonstrate whether there is an improvement in the results from a feature engineering perspective. By applying the PCA method with a ratio of 7/8, it is possible to achieve the same success in the results obtained in the experimental study with the best 7 features out of 8. This was tested to reduce the computational complexity. It was observed that, due to the small number of features, it did not have a positive impact on the performance, as can be seen in Table 4.

Table 4. Machine learning techniques for comparison (accuracy %)

Machine Learning Models	TV		3-Fold CV		5-Fold CV		10-Fold CV	
	PCA Disable	PCA Enable	PCA Disable	PCA Enable	PCA Disable	PCA Enable	PCA Disable	PCA Enable
Fine Tree	90.3%	87.1%	66.1%	58.1%	72.6%	67.7%	74.2%	67.7%
Medium Tree	90.3%	87.1%	66.1%	58.1%	72.6%	67.7%	74.2%	67.7%
Coarse Tree	87.1%	82.3%	66.1%	58.1%	72.6%	74.2%	72.6%	64.5%
Linear Discriminant	77.4 %	72.6%	66.1%	69.4%	72.6%	71.0%	72.6%	69.4%
Logistic Regression	77.4%	75.8%	67.7%	71.0%	75.8%	71.0%	72.6%	69.4%
Gaussian Naive Bayes	79.0%	79.0%	69.4%	75.8%	75.8%	74.2%	75.8%	77.4%
Kernel Naive Bayes	77.4%	79.0%	66.1%	75.8%	67.7%	67.7%	71.0%	72.6%
Linear SVM	75.8%	74.2%	69.4%	69.4%	71.0%	69.4%	71.0%	69.4%
Quadratic SVM	88.7%	82.3%	**59.7%****	61.3%	64.5%	72.6%	66.1%	66.1%
Cubic SVM	96.8%	93.5%	62.9%	66.1%	62.9%	66.1%	64.5%	69.4%
Fine Gaussian SVM	98.4%	91.9%	69.4%	71.0%	67.7%	71.0%	71.0%	71.0%
Medium Gaussian SVM	83.9%	82.3%	66.1%	64.5%	66.1%	67.7%	72.6%	67.7%
Coarse Gaussian SVM	67.7%	67.7%	67.7%	67.7%	67.7%	67.7%	67.7%	67.7%
Fine KNN	**100%***	**100%***	66.1%	67.7%	64.5%	72.6%	69.4%	75.8%
Medium KNN	74.2%	77.4%	69.4%	64.5%	71.0%	64.5%	71.0%	62.9%
Coarse KNN	67.7%	67.7%	67.7%	67.7%	67.7%	67.7%	67.7%	67.7%
Cosine KNN	77.4%	74.2%	62.9%	67.7%	67.7%	69.4%	69.4%	67.7%
Cubic KNN	79.0%	77.4%	71.0%	69.4%	66.1%	69.4%	72.6%	64.5%
Weighted KNN	**100%***	**100%***	62.9%	67.7%	66.1%	71.0%	74.2%	71.0%
Boosted Trees	67.7%	67.7%	67.7%	67.7%	67.7 %	67.7%	71.0%	67.7%
Bagged Trees	**100%***	**100%***	62.9%	66.1%	74.2 %	69.4%	74.2%	67.7%
Subspace Discriminant	77.4%	75.8%	71.0%	71.0%	**75.8%**	72.6%	71.0%	71.0%
Subspace KNN	**100%**	**100%**	69.4%	67.7%	**69.4%**	**69.4%**	**74.2%**	72.6%
RUS Boosted Trees	**100%**	**100%**	64.5%	58.1%	72.6%	72.6%	77.4%	71.0%

* Best accuracy, ** worst accuracy

In this study, since the total input vector consists of 63 cases, the k-fold cross validation method did not improve the performance. Therefore, this method is not recommended for studies with a small input vector. The experimental results of the study are presented in Table 4 and analyzed in terms of feature engineering techniques. As a result, the traditional value data splitting method is recommended for studies with low input vector.

In order to prevent the model from overlearning, both cross-validation and feature selection methods have been applied and performance degradation has been observed in the results. To overcome this problem, performance can be improved by increasing the amount of data.

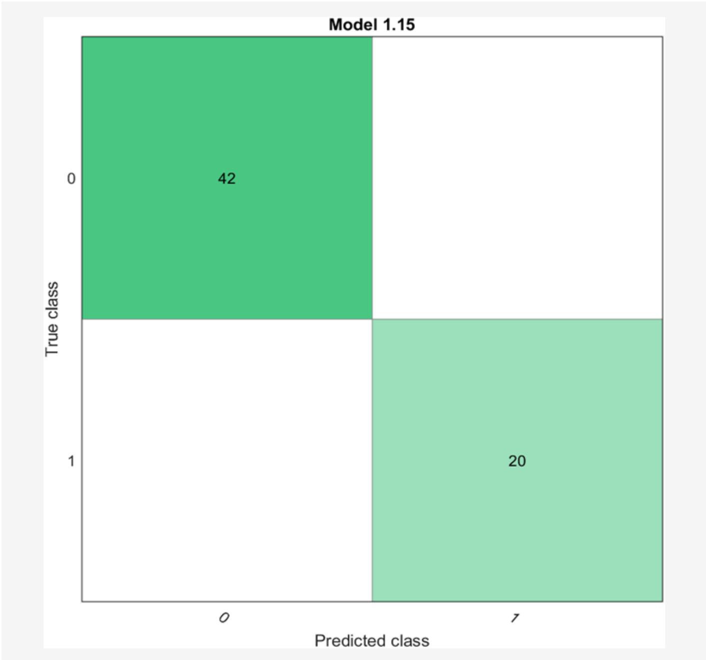

Figure 3. Confusion Matrix for the most successful Fine KNN algorithm.

Figure 4. Confusion Matrix for the worst successful Quadratic SVM algorithm.

The ROC curves of the best-performing Fine KNN algorithm and the worst-performing Quadratic SVM algorithms are shown in Figure 5 and Figure 6, respectively.

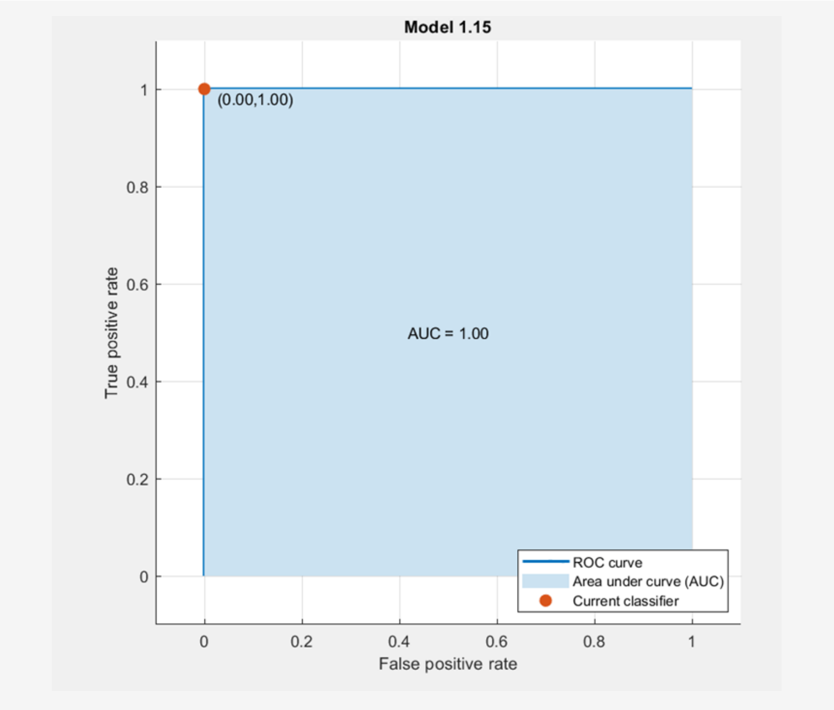

Figure 5. ROC curve for the best performing Fine KNN algorithm.

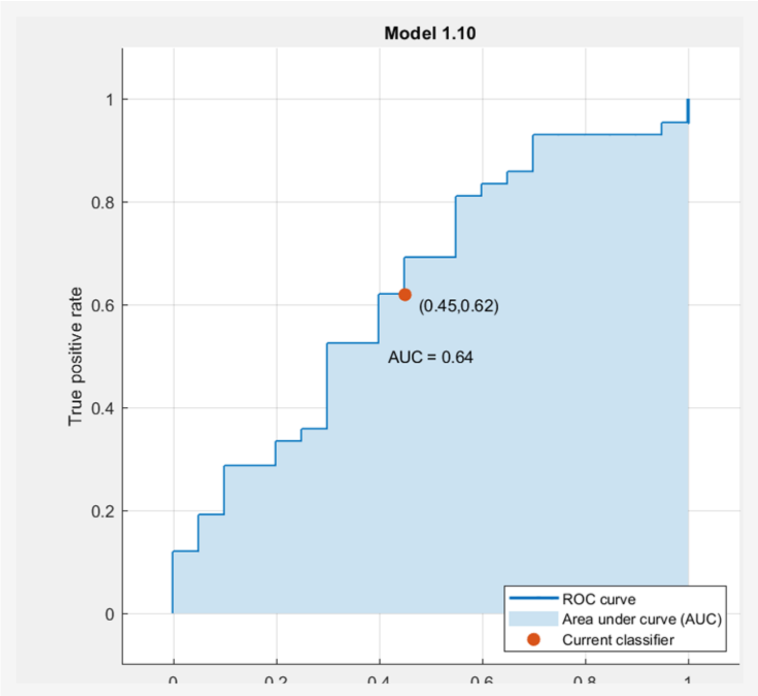

Figure 6. ROC curve for the worst-accuracy performance of Quadratic SVM algorithm.

4. Discussion

In this study, machine learning models were employed for predicting the risk of idiopathic scoliosis in adolescents. The utilized machine learning models incorporated plantar pressure distribution data, revealing models that exhibited 100% performance (see Table 4). Leveraging artificial intelligence for scoliosis risk prediction in this study holds promising prospects. Additionally, we believe that the data obtained from this study serves as a preliminary exploration towards establishing a decision support system based on artificial neural networks (ANN) capable of predicting scoliosis risk in adolescents.
In a 2019 study by XU et al. [14], genetic factors potentially linked to the prognosis of AIS in diagnosed individuals were investigated. They highlighted the potential influence of 10 genetic variants on AIS susceptibility [14]. These variants identified in genetic factor analysis were deemed influential [14], although it's noted that individual testing for each patient may be necessary, posing potential cost challenges [14]. Considering the financial constraints associated with genetic testing for every patient, we propose that plantar pressure analysis coupled with machine learning models offers a cost-effective and rapid alternative for scoliosis risk prediction. However, we acknowledge that plantar pressure analysis alone may not suffice for predicting prognosis in diagnosed individuals. A systematic review published in 2021 emphasized the necessity of developing a patient-specific prediction system for the progression of scoliosis [15]. The review emphasized the insufficiency of relying solely on radiological findings and classification systems [15]. We believe that the dataset obtained from our study holds promise for developing such prediction systems.

In a study by Lv et al. [16], the efficacy of machine learning models for predicting scoliosis risk was assessed. Data including sitting height, biomechanical properties of the lumbar region, pelvis, and shoulder were utilized across five different machine learning models [16]. Radiological imaging was employed in these methodologies, culminating in the creation of a dataset derived from calculations performed on radiological images obtained from patients. Within this dataset, five distinct machine learning models were implemented alongside their respective sets: the Random Forest Model (RFM), Support Vector Machine Model, Artificial Neural Network Model (ANNM), Decision Tree Model (DTM), and Generalized Linear Model (GLM). In our investigation, a total of 25 diverse machine learning models were utilized. We posit that our study holds potential to significantly enrich the existing literature in this domain. Notably, the plantar pressure analysis conducted in our study incurred no costs for either patients or healthy individuals, and the utilized pressure analysis method is devoid of any harmful radiation. A notable strength of our study lies in the absence of adverse effects on patients stemming from the obtained dataset.

5. Conclusion

Radiological evaluations and related algorithms are available for predicting the prognosis of scoliosis in diagnosed individuals. However, predicting the risk of scoliosis in healthy adolescents remains challenging. The machine learning models derived from this study can offer a solution for predicting scoliosis risk in healthy adolescents. Moreover, we contend that incorporating data from plantar pressure analysis into machine learning models in this study will yield significant contributions to the literature. Consequently, predicting scoliosis risk in adolescents using plantar pressure analysis values and machine learning models can provide valuable insights for clinicians in this field. Furthermore, we believe that the dataset obtained from this study can serve as a valuable resource for researchers in the field of scoliosis and those interested in conducting research in artificial intelligence.

References

[1] Negrini, S., Donzelli, S., Aulisa, A. G., Czaprowski, D., Schreiber, S., de Mauroy, J. C., ... & Zaina, F. (2018). 2016 SOSORT guidelines: orthopaedic and rehabilitation treatment of idiopathic scoliosis during growth. Scoliosis and Spinal Disorders, 13, 1-48.

[2] Choudhry, M. N., Ahmad, Z., & Verma, R. (2016). Adolescent Idiopathic Scoliosis. The Open Orthopaedics Journal, 10(1), 143. https://doi.org/10.2174/1874325001610010143

[3] Zagalaz-Anula, N., León-Morillas, F., Andradre-Ortega, J. A., Ibáñez-Vera, A. J., de Oliveira Sousa, S. L., & Lomas-Vega, R. (2020). Case Report: Conservative Treatment of Adolescent Idiopathic Scoliosis Can Alter the Perception of Verticality. A Preliminary Study. Frontiers in Pediatrics, 8, 609555. https://doi.org/10.3389/FPED.2020.609555

[4] Gámiz-Bermúdez, F., Obrero-Gaitán, E., Zagalaz-Anula, N., & Lomas-Vega, R. (2022). Corrective exercise-based therapy for adolescent idiopathic scoliosis: Systematic review and meta-analysis. Clinical Rehabilitation, 36(5), 597-608.

[5] Szopa, A., & Domagalska-Szopa, M. (2017). Correlation between respiratory function and spine and thorax deformity in children with mild scoliosis. Medicine, 96(22).

[6] Negrini, S., Aulisa, A. G., Aulisa, L., Circo, A. B., De Mauroy, J. C., Durmala, J., ... & Maruyama, T. (2012). 2011 SOSORT guidelines: orthopaedic and rehabilitation treatment of idiopathic scoliosis during growth. Scoliosis, 7(1), 1-35.

[7] Daryabor, A., Arazpour, M., Sharifi, G., Bani, M. A., Aboutorabi, A., & Golchin, N. (2017). Gait and energy consumption in adolescent idiopathic scoliosis: A literature review. Ann Phys Rehabil Med., 60(2), 107–116. doi: 10.1016/j.rehab.2016.10.008

[8] Park, H. J., Sim, T., Suh S. W., Yang, J. H., Koo, H., & Mun, J. H. (2016). Analysis of coordination between thoracic and pelvic kinematic movements during gait in adolescents with idiopathic scoliosis. Eur Spine J., 25, 385 –394. doi: 10.1007/s00586-015-3931-0

[9] Hmida, J., Tomschi, F., Strauss, A. C., & Hilberg, T. (2023). Relationship between foot pressure and spinal parameters in healthy adults–A systematic review. Gait & Posture.

[10] Arslan, M., & Görgü, S. Ö. (2023). Effect of short-term spinal orthosis and insoles application on cobb angle, plantar pressure and balance in individuals with adolescent idiopathic scoliosis. Clinical Biomechanics, 110, 106121.

[11] Yağcı, G., & Yakut, Y. (2017). Effects of A Spinal Brace on The Functional Profile of The Feet in Adolescent Idiopathic Scoliosis. Acıbadem Üniversitesi Sağlık Bilimleri Dergisi, (4), 207-213.

[12] Altun, Y., Dilek, B., & Yakut, Y. (2021). Effects of brace on pedobarographic parameters in individuals with adolescent idiopathic scoliosis. Journal of Exercise Therapy and Rehabilitation, 8(1), 44-53.

[13] Gruić, I., Cebović, K., & Medved, V. (2016, November). Comparison of pedobarographic profile in young males with left and right scoliotic posture. In International Congress on Sport Sciences Research and Technology Support (Vol. 2, pp. 89-95). SCITEPRESS.

[14] Xu, L., Wu, Z., Xia, C., Tang, N., Cheng, J. C., Qiu, Y., & Zhu, Z. (2019). A genetic predictive model estimating the risk of developing adolescent idiopathic scoliosis. Current Genomics, 20(4), 246-251.

[15] Lenz, M., Oikonomidis, S., Harland, A., Fürnstahl, P., Farshad, M., Bredow, J., ... & Scheyerer, M. J. (2021). Scoliosis and Prognosis—a systematic review regarding patient-specific and radiological predictive factors for curve progression. European Spine Journal, 30, 1813-1822.

[16] Lv, Z., Lv, W., Wang, L., & Ou, J. (2023). Development and validation of machine learning-based models for prediction of adolescent idiopathic scoliosis: A retrospective study. Medicine, 102(14), e33441.