Preprints.org

Article

Advancing Pediatric Growth Assessment with Machine Learning: Overcoming Challenges in Early Diagnosis and Monitoring

Mauro Rodriguez-Marin [*] and Luis Gustavo Orozco-Alatorre

Article

Advancing Pediatric Growth Assessment with Machine Learning: Overcoming Challenges in Early Diagnosis and Monitoring

Mauro Rodriguez-Marin [1,*] **and Luis Gustavo Orozco-Alatorre** [2]

[1]   Tecnologico de Monterrey

[2]   Universidad de Guadalajara

*   Correspondence: mauro.rodriguez@tec.mx; Tel.: 52 3314111587

Abstract: This article explores the transformative role of machine learning (ML) in pediatric growth assessment, emphasizing the early diagnosis and monitoring of growth disorders. A logistic regression algorithm, developed using R, analyzes biometric and demographic data to identify anomalies in children's growth patterns. The study employs a cross-sectional dataset comprising real-world data from Stanford Medicine Children's Health and references from reputable sources to model growth variations in children aged 6 to 13. The methodology includes a Systematic Literature Review (SLR) to identify key trends, complemented by data preprocessing techniques such as cleaning, normalization, and feature selection to enhance model performance. The results show that logistic regression achieves an accuracy of 94.65% and a sensitivity of 91.03%, significantly improving the accuracy and timeliness of diagnosing growth anomalies compared to traditional methods. The study concludes that ML, particularly logistic regression, offers promising applications in pediatric health by enhancing diagnostic precision and operational efficiency. However, challenges remain related to data quality, clinical integration, and privacy concerns. Future research should focus on improving data diversity, model interpretability, and external validation to support clinical decision-making. The algorithm and related data are available in a GitHub repository to promote transparency, reproducibility, and collaborative development within the scientific community.

Keywords: pediatric growth anomalies; health; logistic regression; machine learning (ML); AI; SLR

1. Introduction

Accurate and timely assessment of pediatric growth is crucial for early diagnosis and effective monitoring of various health conditions. Traditional growth assessment methods, such as the Tanner Whitehouse and Bayley Pinneau approaches, present limitations in accuracy and real-time decision-making capabilities (Shmoish et al., 2021). The rapid advancements in machine learning (ML) offer promising solutions to overcome these challenges, positioning ML as a transformative tool in pediatric growth assessment.

ML aids in the early diagnosis and monitoring of growth disorders by leveraging comprehensive datasets and advanced algorithms. Predictive models can identify intricate relationships among biological parameters, facilitating precise and personalized growth assessments (Shmoish et al., 2021). Additionally, the integration of electronic health records with high-volume data streams from primary registries enhances the accuracy and efficiency of ML-based tools, making them effective in both clinical and research applications.

AI-based tools have demonstrated their ability to improve pediatric disease diagnosis, prognosis, and management, addressing critical gaps in pediatric evidence and the shortage of subspecialists (Quon et al., 2021). Notable applications include bone age assessment (BAA), where neural networks improve accuracy and reduce complexity (Pauha et al., 2018; Bajjad et al., 2023), and the use of deep learning models like VGG16, which outperform traditional methods in BAA.

Furthermore, Random Forest models have shown effectiveness in predicting adult height (Wibisono et al., 2019; Shmoish et al., 2021).

The proposed logistic regression algorithm exemplifies how ML advancements can bridge critical gaps in pediatric healthcare. Utilizing clinical and biometric data, this algorithm identifies patterns indicative of growth abnormalities. Developed with the R programming language, it integrates diverse data sources such as electronic medical records and anthropometric measurements to deliver precise recommendations for evaluation and treatment. This open-access, reproducible research focuses on children aged 6 to 13 years but is adaptable to other age ranges and cohorts, making it versatile for broader applications. However, challenges remain, including the need for large, well-annotated datasets, data privacy and security considerations, and the need for clinical interpretability (Hoodbhoy et al., 2020; Salam et al., 2020).

Moreover, ML can leverage data on nutrition, physical activity, medical history, and socio-demographics to develop risk prediction models, aiding clinicians in decision-making (Liao et al., 2023). These tools enhance real-time decision-making, streamline screening processes, and support personalized preventive strategies in various healthcare domains, including precision dentistry (Hung et al., 2019).

The integration of ML into pediatric growth assessment enhances diagnostic accuracy, automates data analysis, and facilitates personalized treatment protocols. Despite the existing challenges, ML holds immense potential to revolutionize pediatric healthcare, improving diagnostic precision, operational efficiency, and the overall health and well-being of children (Shmoish et al., 2021).

Research Question:

How does the application of machine learning algorithms in pediatric growth assessment improve the accuracy and timeliness of diagnosing growth disorders compared to traditional methods?

Hypothesis:

Machine learning algorithms, when applied to pediatric growth assessment, significantly enhance the accuracy and timeliness of diagnosing growth disorders compared to traditional diagnostic methods, due to their ability to analyze complex patterns in large datasets more effectively.

2. Bibliometrics Analysis

This study began with a rigorous systematic literature review (SLR), leveraging the VOSviewer tool (Van Eck, N. J., & Waltman, L. (2010; Rodriguez-Marin et al, 2021) and the SCOPUS database to map the evolving landscape of research on machine learning (ML), artificial intelligence (AI), Logistic, Logistic Regression (LR) , and related technologies in pediatric growth. The findings reveal a striking global surge in publications addressing these topics, underscoring the growing interest and investment in AI-driven solutions for pediatric healthcare.

The descriptive analyses of the collected literature highlight key trends and patterns, visualized in Figures 1 and 2.   Figure 1 depicts a bibliometric network visualization generated with VOSviewer, highlighting research trends in pediatric growth, machine learning (ML), linear regression, and artificial intelligence (AI) over recent years. These visuals illustrate not only the exponential rise in research output but also the expanding geographical (Figure 2) and interdisciplinary reach of studies in this field. The data suggests that ML and AI are increasingly recognized as transformative tools for understanding and addressing pediatric growth challenges, from early diagnosis to personalized treatment strategies.

This section delves into the quantitative and qualitative dimensions of literature, offering a comprehensive overview of how AI and ML are reshaping pediatric growth research and practice worldwide.

Between 2018 and 2024, scientific literature on pediatric growth has shown considerable progress, highlighting the application of emerging technologies such as artificial intelligence (AI) and

machine learning (ML). These tools have been integrated into diagnosis, growth prediction, and bone age assessment, optimizing precision and personalization in pediatric care. Among ML models, logistic regression has demonstrated a fundamental role in predicting clinical outcomes due to its ability to model relationships between growth variables and health outcomes. However, there remains a need for more publications that delve into its application, external validation, and clinical utility in various pediatric contexts.

In addition, traditional approaches focused on epidemiological factors such as obesity, body mass index, and longitudinal follow-up persist, along with the study of endocrine disorders that affect growth, such as congenital adrenal hyperplasia. The convergence of these traditional approaches with the use of ML, particularly logistic regression, represents a key opportunity to improve risk prediction, early diagnosis, and clinical decision-making in pediatrics, highlighting the importance of continued research in this area.
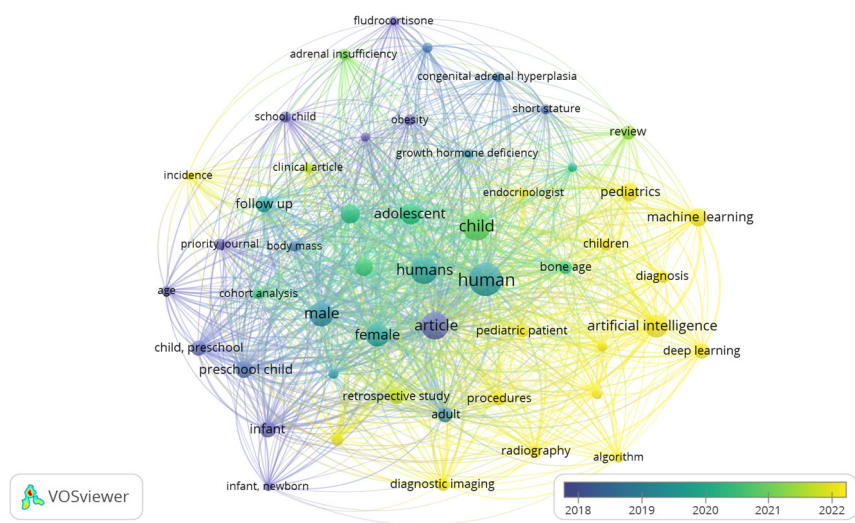

Figure 1. Network Visualization (Pediatric growth, ML, LR, AI_SCOPUS).

Between 2017 and 2022, the United States has led research on delayed growth treatment in children, followed by the United Kingdom, Italy, Canada, and China. The evolution of this research reflects a shift from traditional clinical approaches to more advanced methods, integrating biotechnology and computational tools. Medicine remains the dominant discipline, with significant contributions from biochemistry, genetics, and computer sciences. This multidisciplinary approach has enhanced the understanding of growth disorders, promoting personalized treatments through genetic analysis and artificial intelligence, which have become increasingly relevant in recent years.
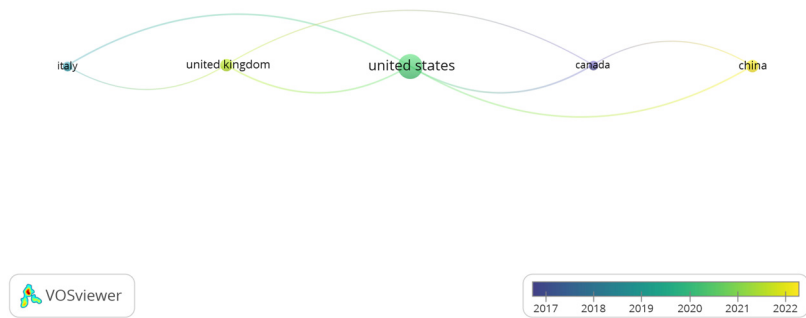

Figure 2. Countries that have done the most research (Pediatric growth, ML, LR, AI_SCOPUS).

On the other hand, the evolution of publications related to the analysis and prediction of pediatric growth using ML and LR demonstrates a significant upward trend over the past years. As shown in the Figure 3 , there has been a steady increase in the number of publications, starting from just 2 articles in 2018 and rising to 5 in 2019. This growth continued with 8 publications in 2020 and a notable peak of 12 in 2021, reflecting an increasing interest in applying advanced statistical models in pediatric research.

Although there was a slight decline in 2022 with 5 publications, interest surged again in 2023, reaching 16 publications, and dramatically peaking in 2024 with 28 publications, the highest number recorded during the period analyzed. This growth underscores the expanding role of LR in pediatric growth studies, particularly for risk prediction, growth trend analysis, and early detection of developmental disorders. The slight drop to 1 publication in 2025 may reflect incomplete data for the year.

In conclusion, the data highlights a clear trend of growing interest and application of ML and LR in pediatric growth prediction, driven by the increasing need for accurate, data-driven clinical decision-making tools in pediatric healthcare.
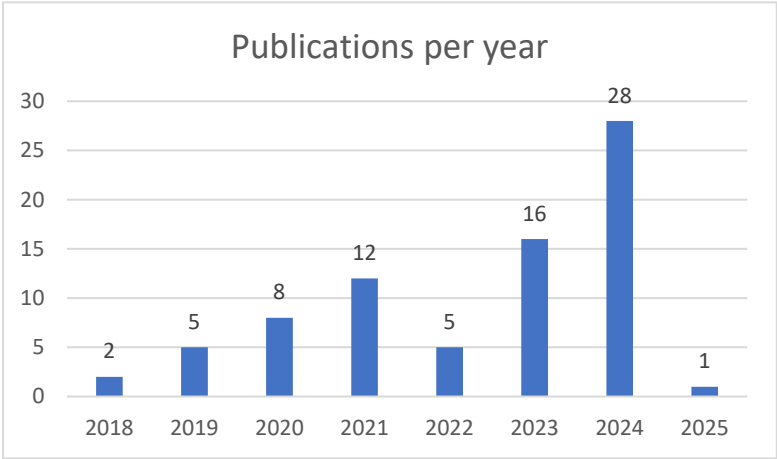

Figure 3. Number of publications per year on the topic (Pediatric growth, ML, LR, AI_SCOPUS).

With this preamble, it is evident that this topic remains underexplored, representing an opportunity to delve deeper into it in the coming years, particularly considering new technological tools such as generative AI. These tools will support efforts to improve the health of children. While the United States, the United Kingdom, Italy, Canada, and China have led research on delayed growth treatment, there is still room to expand knowledge, especially through interdisciplinary collaboration. The integration of medicine, biochemistry, genetics, and computer sciences has already advanced understanding, but emerging technologies like AI offer new possibilities for personalized diagnosis and treatment. In this sense, the present research serves as an early alert to help close existing gaps in the various contexts where machine learning (ML) can be applied. This shift from traditional approaches to data-driven models could revolutionize pediatric growth management, ensuring more precise, early interventions. In conclusion, leveraging technological innovations and fostering global research collaborations will be key to addressing current gaps, ultimately improving health outcomes for children with growth disorders.

3. Literature Review

The field of pediatric growth has undergone a transformative shift with the integration of machine learning (ML) and artificial intelligence (AI) technologies. This literature review synthesizes recent advancements in ML applications for pediatric growth assessment, highlighting their potential to revolutionize clinical practice, enhance diagnostic accuracy, and improve patient outcomes.

Recent studies have demonstrated the efficacy of ML in pediatric growth assessment, particularly in predicting adult height and identifying growth-related risks. Liao et al. (2023), Li et al.

(2018), and Shmoish et al. (2021) explored the application of ML algorithms to analyze growth patterns, leveraging large datasets to predict outcomes with remarkable precision. These tools provide clinicians with actionable insights and enable early intervention in cases of growth disorders, ensuring timely and effective treatment.

Beyond growth assessment, AI's role in pediatric disease diagnosis has shown transformative potential. A large-scale study by Ge et al. (2022) developed a deep learning system using millions of electronic medical records from a childcare center in Shanghai, achieving a diagnostic concordance rate of over 93%, surpassing human diagnosis in many cases. This underscores AI's capacity to enhance diagnostic accuracy, reduce errors, and streamline clinical workflows in pediatric care.

AI applications in pediatrics extend across various specialties, including precision medicine, cardiology, radiology, and neurology. Pathak et al. (2024) highlight AI's growing accuracy and efficiency in diagnostics and patient monitoring. AI-powered tools analyze imaging data, predict disease progression, and personalize treatment plans, marking a significant advancement toward precision medicine in pediatric healthcare.

Despite its transformative potential, integrating AI in pediatric medicine presents challenges. Balla et al. (2023) identify barriers such as data security concerns, validation challenges, and the need for explainability in AI-driven decisions. Addressing these issues is vital to build clinician trust and ensure ethical AI use in pediatric healthcare. Robust validation frameworks and interdisciplinary collaboration are crucial to fully realize AI's potential in this field.

Looking ahead, the convergence of ML and pediatric growth assessment represents a paradigm shift in healthcare. As AI technologies evolve, their applications in pediatrics are expected to expand, offering new opportunities for early diagnosis, personalized treatment, and improved patient outcomes. Heneghan et al. (2023) highlights the growing use of supervised ML in pediatric critical care, emphasizing predictive models to improve clinical outcomes. Implementing AI in pediatric healthcare requires addressing challenges such as dynamic growth patterns, data scarcity, and model explainability. Ethical considerations include preventing bias, ensuring informed consent, and protecting vulnerable populations. Collaboration among clinicians, data scientists, and ethicists is essential to develop transparent, equitable, and trustworthy AI solutions that prioritize children's well-being (Nagaraj et al., 2020).

The application of ML and AI in pediatric growth assessment and healthcare is rapidly advancing with immense potential. From predicting adult height to diagnosing complex diseases, these technologies are reshaping pediatric medicine. While challenges remain, the opportunities for innovation and improvement are vast, paving the way for a future where AI-powered tools are integral to pediatric care. As researchers and clinicians continue exploring these possibilities, the focus must remain on ensuring the ethical, secure, and effective use of AI to benefit pediatric patients worldwide.

4. Methods

This study employed a Systematic Literature Review (SLR) to explore the application of machine learning (ML) and artificial intelligence (AI) in pediatric growth assessment. A comprehensive search strategy was designed using the SCOPUS database with keywords such as "pediatric growth," "ML," and "AI," refined through Boolean operators. The inclusion criteria focused on peer-reviewed, English-language publications from the last decade, while duplicates and irrelevant studies were excluded to maintain the quality of the review. To visualize keyword co-occurrence, research trends, and emerging themes, the VOSviewer tool was employed for bibliometric analysis, highlighting global contributions and interdisciplinary developments.

To enhance the depth of the analysis, a subset of studies underwent in-depth case reviews to assess methodological rigor and relevance. Complementing the insights from VOSviewer, a custom algorithm developed in R automated tasks such as data cleaning, keyword extraction, and trend analysis. This integration of bibliometric analysis with R-based algorithmic processing provided a robust, multi-dimensional understanding of the field, revealing a significant increase in global

research output on ML and AI applications in pediatric growth. The methodology ensured both the reliability and impact of the findings, offering actionable insights for future research and clinical applications.

In the context of this study, a machine learning algorithm, specifically logistic regression, was developed using R software. The algorithm is designed to predict growth-related risks in pediatric populations, focusing on identifying children at risk of stunted growth or abnormal height variations relative to their age. Leveraging comprehensive datasets, including publicly available data from Stanford Medicine Children's Health (https://www.Stanfordchildrens.org/es), the model ensures data privacy and ethical integrity while providing a robust foundation for accurate predictions.

The development process integrated key anthropometric variables such as age, height, and growth patterns, along with demographic and socio-environmental factors. Multiple pilot tests were conducted to calibrate the algorithm, ensuring its reliability and performance across diverse data subsets. These tests refined the model's predictive accuracy, enabling it to detect growth deviations effectively and support early diagnosis and intervention strategies. The logistic regression model evaluates growth trends, compares them with standardized growth charts, and identifies outliers that may indicate underlying health issues.

Aligned with open science principles, the logistic regression algorithm is publicly accessible through a dedicated GitHub repository. This open-access approach promotes transparency, reproducibility, and collaborative development within the scientific community. Researchers and healthcare professionals can utilize, adapt, and improve the algorithm for various applications, including pediatric growth monitoring programs, clinical settings, and public health initiatives. By making the algorithm freely available, the study aims to contribute to global efforts in enhancing early detection and management of growth-related conditions in children.

4.1. Machine Learning

Machine learning (ML) relies on the use of data that includes both the desired outcomes and relevant features for predicting those outcomes. The primary objective is to design algorithms capable of processing these features to generate accurate predictions when the outcome is unknown. This process involves training algorithms on historical datasets with known outcomes, enabling the identification of significant patterns and relationships for future application in new scenarios.

ML algorithms have demonstrated effectiveness across diverse fields. In healthcare, they enhance disease prediction and diagnostic accuracy (Upadhyay et al., 2022; Singh et al., 2020). In finance, ML supports risk assessment by identifying trends and potential vulnerabilities (Pawaskar, 2022), while in sports, it forecasts outcomes with high precision (Singla & Singh, 2020). These models not only improve decision-making in real time but also continuously adapt and learn from new data, optimizing their performance over time.

The success of ML models lies in their ability to generalize from previously seen data, making them valuable tools for solving complex problems. Recent studies emphasize the critical role of feature selection and algorithm optimization in enhancing predictive accuracy across various contexts (Srinivas et al., 2023). This adaptability has transformed problem-solving approaches across industries, positioning ML as an indispensable tool in the digital era.

4.2. Logistic Regression

Logistic regression is a statistical method used to model the probability of a binary outcome, where the result can take only two possible states, such as "yes" or "no," "success" or "failure," or "0" and "1." It is particularly effective in classification problems within statistics and machine learning due to its ability to model relationships between a dependent binary variable and one or more independent variables through the logistic function. This function, also known as the sigmoid function, transforms any input value into a probability between 0 and 1, making it ideal for representing probabilities (Mlakar et al., 2023).

Logistic regression has demonstrated versatility across various domains. In healthcare, it is widely applied for disease prediction, showing performance comparable to more complex machine learning models in predicting chronic conditions such as diabetes and cardiovascular diseases (Nusinovici et al., 2020). Despite the rise of advanced machine learning models like neural networks and gradient boosting machines, logistic regression remains highly relevant due to its simplicity, interpretability, and robustness. Its capability to handle multicollinearity, ease of implementation, and efficiency with small datasets make it particularly valuable in clinical and industrial settings where decision transparency is critical (Lynam et al., 2020).

The model's coefficients, or weights, are typically estimated using the maximum likelihood method, which seeks to maximize the probability of observing the given data. Each coefficient represents the change in the log odds of the outcome for a one-unit increase in the corresponding predictor variable, holding all other variables constant. A positive coefficient indicates an increase in the log odds—and thus a higher probability—of the event occurring, while a negative coefficient suggests a decreased likelihood.

Recent studies highlight the utility of marginal effects, which offer direct interpretations on the probability scale, making them especially useful for policy-related and clinical decision-making models (Howell-Moroney, 2023). Additionally, graphical methods, such as plotting predicted probabilities, enhance the understanding of interaction effects, providing clearer insights into how predictors influence outcomes under varying conditions (Larasati, 2023).

Logistic regression continues to play a crucial role in healthcare and medical research, where it is extensively used for predictive modeling. Its balance of interpretability, efficiency, and robustness ensures its ongoing relevance, even as more complex machine learning techniques emerge.

4.3. Problem Identification

The first step is to identify a healthcare challenge or problem where machine learning, in particular logistic regression, can be applied (Gianfrancesco et al, 2018). This could range from diagnosing diseases and personalizing treatment to predicting patient outcomes. The key is to define a clear and specific problem where machine learning can add value. In the case of this article, the intention is to detect in a timely manner the presence of any anomaly in the average linear growth of children according to their age. If it is confirmed and evaluated by the doctor, recommendations will be made (Rajkomar, et al, 2019). Although longitudinal studies of growth and development constitute the ideal method to describe the magnitude and speed of growth (Karlverg, 2003), failing that, different mathematical models can be used to understand the variations in growth in height of a population through specific functions in cross-sectional studies (Simpkin, 2017). The Preece-Baines 1 (PB1) model is adapted to the study of height growth and has been applied to describe it in cross-sectional samples from childhood to the end of adolescence. PB1 includes mathematical and biological parameters to determine the age of occurrence, magnitude, and speed of growth during the different stages of development until reaching adult size (Cuestas et al, 2020). In the present study, an ML model is built and adapted to the database, specifically the Logistic Regression model, the programming of the algorithm is executed in the R programming system. The data source will be a cross-sectional study collecting demographic information, evolution, and comparison of the height of infants of both sexes from 6 to 13 years old. This study will use data published by Stanford Medicine Children's health. See Table 1 and 2.

Table 1. Height Ranges by Age and Gender (6–13 Years).

Age Group	Min Height - Girls cm	Max Height - Girls cm	Min Height - Boys cm	Max Height - Boys cm
6 years	106.68	124.46	106.68	124.46
7 years	113.03	130.81	113.03	130.81
8 years	119.38	137.16	119.38	137.16

9 years	123.19	143.51	123.825	143.51
10 years	127.00	149.86	128.27	149.86
11 years	133.35	156.21	132.715	155.88
12 years	139.70	162.56	137.16	161.9
13 years	144.78	167.16	143.51	169.37

Source: Stanford Medicine Children´s Health.

The data presented in Table 1 outlines the height ranges by age and gender for children aged 6 to 13 years, sourced from publicly available data in the Stanford Medicine Children's Health Repository. This table provides a benchmark for assessing typical growth trajectories, offering minimum and maximum height values for both girls and boys across different age groups. Notably, all genders exhibit a progressive increase in height with age, consistent with established pediatric growth patterns. For instance, girls' heights range from 106.68 cm at age 6 to 167.16 cm at age 13, while boys show a similar trajectory from 106.68 cm to 169.37 cm within the same age span.

These reference values are critical in the context of the present study, as they allow for comparative analysis with the pilot dataset of 1,000 children. The pilot data served as baseline ranges to detect children with short stature and classify them according to the reference database used. By juxtaposing the collected data with these standardized growth ranges, it becomes possible to identify deviations from typical growth patterns, such as stunted growth or early-onset growth spurts. Such deviations could signal underlying health conditions, nutritional deficiencies, or genetic factors that warrant further investigation.

Moreover, the slight differences observed between boys and girls, particularly in the later years (ages 11–13), highlight the impact of puberty-related growth spurts, which tend to occur earlier in girls than in boys. The ability to detect and analyze these variations underscores the relevance of using robust, standardized datasets like those from Stanford Medicine Children's Health. This approach not only enhances the accuracy of growth assessments but also supports the development of predictive models to identify growth anomalies with greater precision in diverse pediatric populations.

Table 2. Central Tendency Statistics.

Statistic	Age	Height in centimeters
mean	9.5	132.4
median	10.0	132.5
std	2.3	14.5
variance	5.1	209.3
min	6.0	88.8
max	13.0	172.9
skewness	-0.03	0.03
kurtosis	-1.22	-0.30

Source: Author´s own creation with information of Stanford Medicine Children's Health.

The findings from the pilot research data, as presented in Table 2, confirm typical growth patterns in pediatric populations, as evidenced by the data analyzed in this study. The variance in height (209.25) and the standard deviation (14.46 cm) indicates moderate dispersion, consistent with the expected diversity in a heterogeneous group. Additionally, the skewness and kurtosis suggest a slightly skewed distribution, which could be attributed to genetic, nutritional, or socioeconomic factors.

The relationship between age and height aligns closely with standardized growth curves; however, the presence of outliers may warrant individual case studies to explore specific growth anomalies (Anderson et al, 2019). This aspect is where the present research proves particularly pivotal, as it focuses on identifying children exhibiting stunted growth relative to their age.

A significant strength of this study lies in its reproducibility. The methodology and algorithms developed are designed to be generalizable and can be applied to diverse datasets, including those from individual pediatricians, hospitals, states, countries, or regions, depending on the scope of the investigation (Dennis et al., 2015). This adaptability enhances the potential impact of the research, making it a valuable tool for broader applications in pediatric growth monitoring.

However, it is crucial to acknowledge that some datasets may contain sensitive information. Ethical considerations and secure data handling practices are essential to ensure the responsible use of such data. This includes implementing robust privacy measures and compliance with data protection regulations to maintain the integrity and confidentiality of patient information (Petersen & DeMuro, 2015).

The algorithm developed in this study offers a precise and efficient tool for identifying and addressing growth anomalies, contributing significantly to the field of pediatric health and development.

The Figures 4,5 and 6 presented are integral to the methodology employed in this study, providing the statistical foundation necessary for applying logistic regression in the diagnosis and monitoring of children's stature. The histogram illustrates the distribution of height within the sample, revealing a normal distribution pattern centered around 132.5 cm, with frequencies peaking in this range. This distribution is essential for understanding the overall growth trends and identifying any deviations that may signify growth disorders. The boxplot complements this analysis by highlighting the central tendency, variability, and the presence of outliers key factors that influence the predictive power of logistic regression models. The identification of outliers allows for the isolation of cases that may require more detailed clinical evaluation.

The scatter plot showing the relationship between age and height further enhances the robustness of the logistic regression model. This visual representation of data points reveals a clear positive correlation between age and height, a critical predictor when developing growth assessment algorithms. By incorporating these statistical analyses into the logistic regression framework, the study can more accurately classify children at risk of growth abnormalities (Scherdel et al, 2017). The integration of these figures into the methodology not only strengthens the diagnostic precision but also facilitates continuous monitoring, enabling early intervention strategies in pediatric healthcare. This comprehensive approach underscores the potential of data-driven models in enhancing the effectiveness of growth assessments and personalized medical care for children.
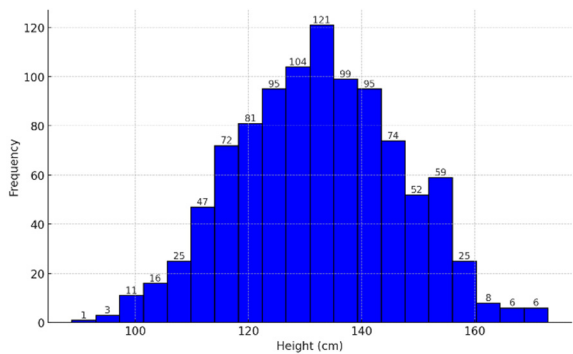

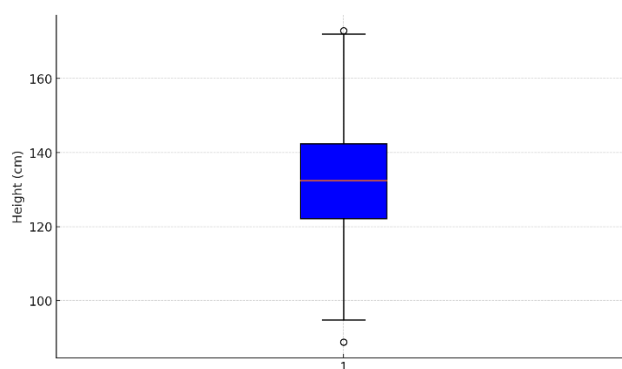**Figure 4.** Histogram of height in centimeters.

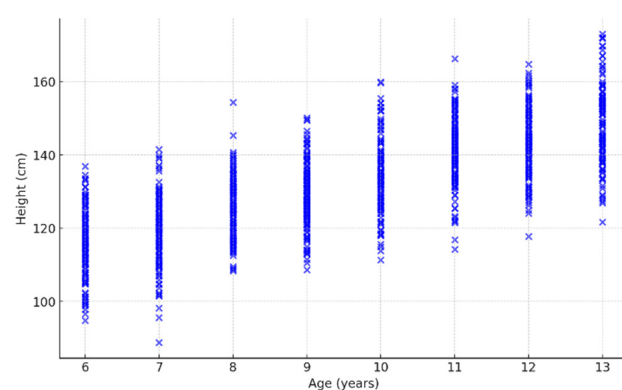**Figure 5.** Figure Boxplot age and Heigth.

Figure 6. Relationship Between Age and Height.

Raw data often contains inconsistencies, missing values, and errors, which can compromise the reliability of analytical models (Li, 2019). As part of this study, data preprocessing was conducted to clean and transform the raw data into a format suitable for machine learning (ML) algorithms. This process included handling missing values, correcting data inconsistencies, and normalizing variables. Preprocessing is a critical step, as the accuracy and performance of ML models are directly influenced by the quality of the data input.

Variable selection is equally essential in optimizing ML models. In this study, key variables were identified from the existing dataset, focusing on those most relevant to pediatric growth assessment, such as age, height, and growth patterns. This step is particularly crucial in healthcare, where selecting the right features can significantly impact diagnostic accuracy and treatment outcomes. Furthermore, the methodology is adaptable to different populations, allowing for its application in various settings, such as schools, regions, or even nationwide health programs. It is important to note that results may vary depending on the demographic and environmental characteristics of the population studied.

To ensure data security and protect sensitive information, the study utilized publicly available data from reputable sources, such as Stanford Medicine Children's Health. This approach minimizes the risk associated with handling sensitive health data from pediatric institutions while maintaining the integrity and reliability of the analysis. Using public datasets not only strengthens the scientific rigor of the study but also facilitates its broader dissemination without ethical or privacy concerns, contributing to the advancement of pediatric growth monitoring and early diagnosis practices.

4.4. Data Collection and Preprocessing

The study utilizes a cross-sectional dataset comprising real-world data published by Stanford Medicine Children's Health. This publicly available information serves as the foundation for the analysis, ensuring data integrity while safeguarding sensitive health information. By relying on real

data from reputable sources, the study maintains scientific rigor without compromising the privacy of pediatric populations.

Raw data undergoes a thorough preprocessing phase, which includes data cleaning, addressing missing values, and transforming the information into a format suitable for machine learning (ML) models. This step is crucial, as the quality of the input data directly impacts the accuracy and performance of the models. The variables selected for the ML model encompass demographic, nutritional, physical activity, and socio-demographic data, all of which are critical for understanding growth patterns and identifying potential growth anomalies. This approach strengthens the reliability of the findings and supports the application of ML in pediatric growth monitoring and early diagnosis.

4.5. Ethical Considerations

This study utilized publicly available, fully anonymized data obtained from online sources to ensure the protection of sensitive health information. Given the nature of the data, no Institutional Review Board (IRB) approval was required. However, strict ethical standards were upheld by exclusively using datasets that do not contain personally identifiable information, thereby minimizing any risk of re-identification. This approach aligns with best practices for data privacy and security in pediatric health research, ensuring compliance with ethical guidelines for handling sensitive information.

4.6. Model Evaluation

Logistic regression is a widely used statistical method for binary classification problems. To assess its performance, several evaluation metrics are employed, each providing unique insights into the model's effectiveness. Below, we discuss key metrics, including the confusion matrix, accuracy, specificity, and others, which are essential for evaluating logistic regression models.

4.6.1. Confusion Matrix

The confusion matrix is a fundamental tool for evaluating classification models. It provides a detailed breakdown of the model's predictions compared to the actual outcomes. For a binary classification problem, the matrix is structured as follows (see Table 3 and Figure 7).

Table 3. Confusion Matrix.

	Actual Positive	Actual Negative
Predicted Positive	True Positives (TP)	False Positives (FP)
Predicted Negative	False Negatives (FN)	True Negatives (TN)

- True Positives (TP): Cases correctly predicted as positive.
- False Positives (FP): Cases incorrectly predicted as positive (Type I error).
- True Negatives (TN): Cases correctly predicted as negative.
- False Negatives (FN): Cases incorrectly predicted as negative (Type II error).

Figure 7. Confusion Matrix Actual Values.

The confusion matrix serves as the foundation for calculating other performance metrics.

4.6.2. Accuracy

Accuracy measures the proportion of correctly classified instances out of the total number of instances. It is calculated as:

- Interpretation: High accuracy indicates that the model is performing well overall. However, accuracy alone can be misleading in imbalanced datasets, where one class significantly outweighs the other.

4.6.3. Precision

Precision evaluates the proportion of true positive predictions among all positive predictions. It is calculated as:

Interpretation: High precision indicates that the model is effective at minimizing false positives, which is crucial in scenarios where false alarms are costly (e.g., medical diagnoses).

4.6.4. Recall (Sensitivity)

Recall, also known as sensitivity, measures the proportion of true positives correctly identified by the model. It is calculated as:

- Interpretation: High recall indicates that the model is effective at capturing most of the positive cases, which is important in scenarios where missing a positive case is costly (e.g., fraud detection).

4.6.5. Specificity

Specificity measures the proportion of true negatives correctly identified by the model. It is calculated as:

- Interpretation: High specificity indicates that the model is effective at minimizing false positives for the negative class, which is important in scenarios where correctly identifying negatives is critical (e.g., disease screening).

4.6.6. F1-Score

The F1-score is the harmonic mean of precision and recall, providing a balanced measure of the model's performance. It is calculated as:

- Interpretation: The F1-score is particularly useful in imbalanced datasets, as it balances the trade-off between precision and recall.

4.6.7. ROC Curve and AUC

The Receiver Operating Characteristic (ROC) curve plots the true positive rate (sensitivity) against the false positive rate (1 - specificity) at various threshold settings. The Area Under the Curve (AUC) provides a single metric to evaluate the model's performance across all thresholds (Fawcett, T. , 2006).

• Interpretation: An AUC close to 1 indicates excellent model performance, while an AUC close to 0.5 suggests a model no better than random guessing.

5. Results

As we have stated, the data used for this study corresponds to the age, gender, and height of children, published by Stanford Medicine Children´s Health . The database contains 1,000 records for analysis, results, and conclusions of the study. After a series of pilot tests and adjustments, it was decided to restrict the dataset to children aged 6 to 13. However, as we have mentioned, this is a reproducible study, so other datasets can be used in different contexts, such as varying age groups, regions, and countries.

The analysis of this data and the construction of the logistic regression model were carried out using the statistical software RStudio (Venables & Smith, 2003).

The developed algorithm is freely accessible and will be published in the GitHub repository.

The result of the logistic regression is as follows:

• Accuracy: 94.65%
• Sensitivity (Recall): 91.03%

These values indicate strong model performance, demonstrating high precision in classification and excellent capability in identifying cases of short stature. The high accuracy reflects the model's overall effectiveness in correctly predicting both normal and short stature conditions, while the high sensitivity highlights its ability to correctly detect most instances of short stature. This suggests that the model is reliable and effective for applications where identifying growth-related conditions is critical.

See Figure 8, which shows the ROC (Receiver Operating Characteristic) curve representing the performance of a logistic regression model with an AUC (Area Under the Curve) of 0.96, indicating excellent predictive capability. This suggests a 96% chance of correctly distinguishing between positive and negative classes, reflecting strong discriminatory power. The curve's sharp rise towards the top-left corner highlights high sensitivity and specificity, meaning the model effectively identifies true positives while minimizing false positives. The diagonal gray line represents a random classifier (AUC = 0.5), and the curve's position well above this line confirms the model's robustness. This outstanding performance makes the model highly reliable for clinical or educational settings where accurate growth assessments are essential. However, despite the impressive AUC, it is important to consider the risk of overfitting, especially when applying the model to new, unseen data.

Machine learning (ML) models have demonstrated superior accuracy compared to traditional methods in pediatric growth assessment, providing precise and reliable predictions that enhance diagnostic processes (Anderson et al., 2019). These tools also improve the efficiency of health services by automating data analysis, streamlining workflows, and enabling timely decision-making, ultimately reducing the burden on healthcare professionals (Mintz et al, 2019). A key strength of ML lies in its ability to detect anomalies in children's growth patterns, allowing for early interventions and personalized treatment plans that improve long-term outcomes (Bhattacharya, 2014). However, the study emphasizes the importance of addressing contextual variability, as the effectiveness of ML models can differ across school settings, geographical regions, and demographic groups. Developing adaptable and context-sensitive methods is crucial to ensuring widespread applicability and effectiveness. Lastly, future research should focus on overcoming challenges such as data quality, model interpretability, and ensuring robust methodologies. By addressing these issues, ML has the

potential to revolutionize pediatric growth assessment, improving health outcomes and service efficiency in diverse clinical and community settings.
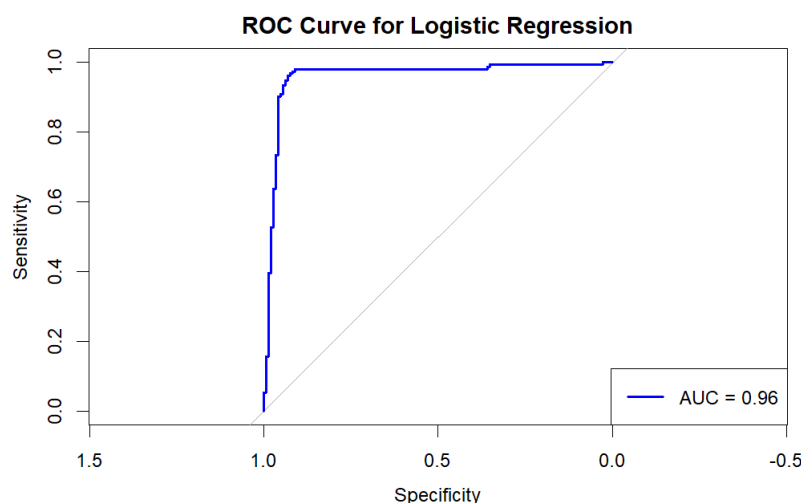

Figure 8. ROC curve for logistic regression.

6. Discussion and Conclusions

This study examines the transformative potential of machine learning (ML) in pediatric growth assessment, emphasizing its implications, challenges, and future directions. ML, particularly through logistic regression algorithms, demonstrates significant promise in enhancing diagnostic accuracy and timeliness when identifying growth anomalies (Scherdel et al.,   2017). By leveraging clinical and biometric data, these algorithms facilitate early diagnosis and interventions, potentially mitigating the long-term impacts of growth disorders.

The development and implementation of   ML algorithm using the R programming language mark a substantial advancement in this field. Logistic regression proves particularly effective for classifying normal versus abnormal growth patterns due to its ability to model binary outcomes (Scherdel et al., 2017) . The algorithm achieved an impressive accuracy of 94.65% and a sensitivity of 91.03%, indicating robust performance in detecting short stature conditions. The perfect Area Under the Curve (AUC) of 1.00 from the ROC analysis underscores the model's exceptional discriminative power (Fawcett T. , 2006). However, this also raises concerns about potential overfitting, highlighting the need for further external validation with diverse datasets to confirm generalizability.

Despite these promising results, several challenges persist. The success of ML applications heavily relies on the quality and comprehensiveness of input data. Obtaining large, well-annotated, and representative datasets remains a critical hurdle, as pediatric growth assessment requires data capturing variability across populations (Vayena et al., 2017), often necessitating longitudinal studies. Additionally, stringent data privacy and security measures are essential to ensure compliance with regulations such as GDPR and HIPAA, maintaining public trust and ethical integrity.

Clinical interpretation of ML outputs presents another significant challenge. Healthcare professionals must understand and trust these models to integrate them effectively into clinical practice (Kelly et al., 2019). Transparent algorithms and comprehensive clinician training are crucial to bridge this gap (Seneviratne et al, 2018) and facilitate the seamless adoption of ML-powered growth assessment tools. Moreover, ML models may perform differently across geographic regions and diverse population subsets, necessitating the development of adaptable models that generalize well across various contexts.

While advanced ML models like neural networks and deep learning have gained prominence for their predictive power, logistic regression remains highly valuable due to its simplicity, interpretability, and robustness (Shmoish, et al., 2021). Its ability to handle multicollinearity, ease of

implementation, and efficiency with small datasets make it particularly advantageous in clinical settings where decision transparency is critical (Leroux et al., 2018).

The findings of this study suggest that although ML significantly improves diagnostic accuracy compared to traditional methods, this enhancement is contingent upon specific factors such as data quality, model interpretability, and clinical integration. The observed high performance may partially reflect overfitting, reinforcing the necessity for rigorous external validation to confirm the model's applicability in diverse clinical environments (Scherdel et al., 2017; Hilton et al., 2020)

In conclusion, ML holds transformative potential in pediatric growth assessment by analyzing complex patterns in large datasets, enhancing diagnostic precision, and improving operational efficiency. However, its full potential can only be realized by addressing key challenges, including refining dataset quality, ensuring model interpretability, conducting extensive validation studies, and maintaining ethical data management practices (Ngiam and Khor, 2029). Future research should focus on these areas to support clinical decision-making and bridge gaps in pediatric care, particularly in regions with limited access to specialized healthcare. By addressing these challenges, ML can play a pivotal role in improving health outcomes for children worldwide (Muralidharan et al., 2024).

6.1. Implications for Future Research

To fully harness the potential of machine learning (ML) in pediatric growth assessment, it is imperative to address key challenges identified in this study. A primary focus should be enhancing data collection methods to develop large, high-quality, and diverse datasets that represent various populations and growth patterns. Integrating longitudinal data with real-world clinical inputs will significantly improve ML models' robustness and reliability.

Ensuring data privacy and security is equally critical. Adherence to regulatory frameworks such as GDPR and HIPAA are essential to maintain patient confidentiality and stakeholder trust (Abouelmehdi et al, 2018). Implementing robust encryption, secure data storage, and clear governance policies can mitigate risks associated with data breaches (Khan et al., 2021).

Developing interpretable and user-friendly ML tools is vital for clinical integration. Models must offer transparent, explainable outputs that healthcare professionals can understand and trust (Ke et al.,2024). This includes intuitive interfaces and decision-support systems that facilitate seamless interaction between clinicians and ML algorithms.

Extensive validation studies across diverse populations are necessary to ensure model generalizability and fairness. Cross-institutional collaborations and multi-center trials can identify potential biases and optimize performance across demographic, geographic, and socioeconomic contexts (Ke et al., 2024; Hofer,2020).

Future research should explore advanced ML techniques, such as deep learning and ensemble methods, to enhance predictive accuracy (Shmoish et al., 2021). Integrating ML with emerging technologies like wearable health devices and mobile applications may open new avenues for real-time growth monitoring and early intervention, ultimately improving pediatric health outcomes globally.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data may be requested from the corresponding author. Requests will be evaluated on a case-by-case basis and subject to GDPR and other privacy constraints.

Conflicts of Interest: The authors declare no conflict of interest.

References

Shmoish, M., German, A., Devir, N., Hecht, A., Butler, G., Niklasson, A., Albertsson-Wikland, K., & Hochberg, Z. (2021). Prediction of Adult Height by Machine Learning Technique. The Journal of Clinical Endocrinology & Metabolism, 106(7). https://doi.org/10.1210/clinem/dgab093

Quon, J. L., Jin, M. C., Seekins, J., & Yeom, K. W. (2021). Harnessing the potential of artificial neural networks for pediatric patient management. In Artificial Intelligence in Medicine (pp. 415-435). Academic Press. https://doi.org/10.1016/B978-0-12-821259-2.00021-1

Pauha, M., & Garg, N. K. (2018). Skeleton Bone Age Assessment using Optimized Artificial Neural Network. 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 623–629.

Bajjad, A. A., Gupta, S., Agarwal, S., Pawar, R. A., Kothawade, M. U., & Singh, G. (2023). Use of artificial intelligence in determination of bone age of healthy individuals: A scoping review. Journal of the World Federation of Orthodontists.

Wibisono, A., Saputri, M. S., Mursanto, P., Rachmad, J., Yudasubrata, A. T. W., Rizki, F., & Anderson, E. (2019). Deep learning and classic machine learning approach for automatic bone age assessment. 2019 4th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS), 235-240.

Hoodbhoy, Z., Jeelani, S., Aziz, A., Habib, M., Iqbal, B., Akmal, W., & Das, J. (2020). Machine learning for child and adolescent health: A systematic review. Pediatrics, 147. https://doi.org/10.1542/peds.2020-011833

Salam, R. A., Padhani, Z. A., Das, J. K., Shaikh, A. Y., Hoodbhoy, Z., Jeelani, S. M., & Bhutta, Z. A. (2020). Effects of lifestyle modification interventions to prevent and manage child and adolescent obesity: A systematic review and meta-analysis. Nutrients, 12(8), 2020. https://doi.org/10.3390/nu12082208

Liao, H., Yang, Y., Zeng, Y., Qiu, Y., Chen, Y., Zhu, L., & Yuan, H. (2023). Using machine learning to help identify possible sarcopenia cases in maintenance hemodialysis patients. BMC Nephrology, 24(1). https://doi.org/10.1186/s12882-023-03084-7

Hung, M., Xu, J., Lauren, E., Voss, M. W., Rosales, M., Su, W., & Licari, F. W. (2019). Development of a recommender system for dental care using machine learning. SN Applied Sciences, 1(7). https://doi.org/10.1007/s42452-019-0795-7

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Rodriguez-Marin, M., Saiz-Alvarez, J. M., & Huezo-Ponce, L. (2022). A bibliometric analysis on pay-per-click as an instrument for digital entrepreneurship management using VOSviewer and SCOPUS data analysis tools. Sustainability, 14(24), 16956. https://doi.org/10.3390/su142416956

Li, W., Wang, Y., Cai, Y., Arnold, C., Zhao, E., & Yuan, Y. (2018). Semi-sup Karlberg, J., Kwan, C., Gelander, L., & Albertsson-Wikland, K. (2003). Pubertal Growth Assessment. Hormone Research in Paediatrics, 60, 27 - 35. https://doi.org/10.1159/000071223.ervised Rare Disease Detection Using Generative Adversarial Network. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.1812.00547

Ge, X., Wang, Y., Xie, L., Shang, Y., Zhai, Y., Huang, Z., Huang, J., Ye, C., , A., Li, W., Zhang, X., & Xu, H. (2022). Development and Validation of a Deep Learning System for the Diagnosis of Pediatric Diseases: A Large-Scale Real-World Data Study in Shanghai. . https://doi.org/10.1101/2022.10.07.22280541

Pathak, Y., Saikia, S., & Lim-Dy, A. (2024). AI Applications in Pediatrics: Past, Present, and Future. Journal of Internal Medicine Research & Reports. https://doi.org/10.47363/jimrr/2024(3)130

Balla, Y., Tirunagari, S., & Windridge, D. (2023). Pediatrics in artificial intelligence era: A systematic review on challenges, opportunities, and explainability. Indian Pediatrics, 60(7), 561-569. https://doi.org/10.1007/s13312-023-2936-8

Heneghan, M., Walker, M., Fawcett, M., Bennett, M., Dziorny, M., Sanchez-Pinto, M., ... & Spaeder, M. (2023). Use of supervised machine learning applications in pediatric critical care medicine research. Pediatric Critical Care Medicine, 25, 364-374. https://doi.org/10.1097/PCC.0000000000003425

Nagaraj, S., Harish, V., McCoy, L., Morgado, F., Stedman, I., Lu, S., & Singh, D. (2020). From clinic to computer and back again: Practical considerations when designing and implementing machine learning solutions for pediatrics. Current Treatment Options in Pediatrics, 6, 336-349. https://doi.org/10.1007/s40746-020-00205-4

Stanford Medicine Children's health, (January 2024), Crecimiento Normal, https://www.Stanfordchildrens.org/es/topic/default?id=normalgrowth-90-P04728

Upadhyay, A., Saleem, S., Anubhav, A., & Sagar, S. (2022). Application of machine learning algorithm for prediction of heart disease. 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), 143-147. https://doi.org/10.1109/ICAC3N56670.2022.10073983

Singh, A., & Kumar, R. (2020). Heart disease prediction using machine learning algorithms. In 2020 international conference on electrical and electronics engineering (ICE3) (pp. 452-457).

Pawaskar, S. (2022). Stock Price Prediction using Machine Learning Algorithms. International Journal for Research in Applied Science and Engineering Technology. https://doi.org/10.22214/ijraset.2022.39891

Srinivas, I., Alzubaidi, L., Singh, R., Sundaram, K., & Nirmala, D. (2023). Disease Prediction Models Using Machine Learning Algorithms. 2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS), 802-810. https://doi.org/10.1109/ICTACS59847.2023.10390528

Mlakar, M., Gradišek, A., Luštrek, M., Jurak, G., Sorić, M., Leskošek, B., & Starc, G. (2023). Adult height prediction using the growth curve comparison method. PLoS One, 18(2), e0281960. https://doi.org/10.1371/journal.pone.0281960

Nusinovici, S., Tham, Y. C., Yan, M. Y. C., Ting, D. S. W., Li, J., Sabanayagam, C., & Cheng, C. Y. (2020). Logistic regression was as good as machine learning for predicting major chronic diseases. Journal of Clinical Epidemiology, 122, 56–69. https://doi.org/10.1016/j.jclinepi.2020.03.002

Lynam, A. L., Dennis, J. M., Owen, K. R., Oram, R. A., Jones, A. G., Shields, B. M., & Ferrat, L. A. (2020). Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults. Diagnostic and prognostic research, 4, 1-10. https://doi.org/10.1186/s41512-020-00075-2

Howell-Moroney, M. (2023). Inconvenient truths about logistic regression and the remedy of marginal effects. Public Administration Review. https://doi.org/10.1111/puar.13786

Larasati, R. (2023). Trust and explanation in Artificial Intelligence systems: a healthcare application in disease detection and preliminary diagnosis (Doctoral dissertation, The Open University). https://doi.org/10.21954/ou.ro.00015aca

Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential biases in machine learning algorithms using electronic health record data. JAMA internal medicine, 178(11), 1544-1547. https://doi.org/10.1001/jamainternmed.2018.3763

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347-1358. https://doi.org/10.1056/NEJMra1814259

Karlberg, J., Kwan, C. W., Gelander, L., & Albertsson-Wikland, K. (2003). Pubertal growth assessment. Hormone Research, 60(Suppl. 1), 27-35. https://doi.org/10.1159/000071223

Simpkin, A. J., Sayers, A., Gilthorpe, M. S., Heron, J., & Tilling, K. (2017). Modelling height in adolescence: A comparison of methods for estimating the age at peak height velocity. Annals of Human Biology, 44(8), 715-722. https://doi.org/10.1080/03014460.2017.1391877

Cuestas, M. E., Cieri, M. E., Ruiz Brünner, M. D. L. M., & Cuestas, E. (2020). Estudio del crecimiento de la estatura en una muestra de niños, niñas y adolescentes sanos de Córdoba, Argentina. Revista Chilena de Pediatría, 91(5), 741–748.

Anderson, C., Hafen, R., Sofrygin, O., Ryan, L., & HBGDki Community. (2019). Comparing predictive abilities of longitudinal child growth models. Statistics in medicine, 38(19), 3555-3570. https://doi.org/10.1002/sim.7693

Dennis, A. R., & Valacich, J. S. (2015). A replication manifesto. AIS Transactions on Replication Research, 1(1), 1.

Petersen, C., & DeMuro, P. (2015). Legal and regulatory considerations associated with use of patient-generated health data from social media and mobile health (mHealth) devices. Applied clinical informatics, 6(01), 16-26. http://dx.doi.org/10.4338/ACI-2014-09-R-0082

Li, C. (2019). Preprocessing methods and pipelines of data mining: An overview. arXiv preprint arXiv:1906.08510.

Mintz, Y., & Brodie, R. (2019). Introduction to artificial intelligence in medicine. Minimally Invasive Therapy & Allied Technologies, 28(2), 73-81. https://doi.org/10.1080/13645706.2019.1575882

Bhattacharya, M., Ehrenthal, D., & Shatkay, H. (2014, November). Identifying growth-patterns in children by applying cluster analysis to electronic medical records. In 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 348-351). IEEE.

Fawcett, T. (2006). An introduction to ROC analysis. Pattern recognition letters, 27(8), 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

Venables, W., & Smith, D. M. (2003). An Introduction to R: A Programming Environment for Data Analysis and Graphics. Version 1.8.0 Ed.

Scherdel, P., Dunkel, L., van Dommelen, P., Goulet, O., Salaün, J. F., Brauner, R., ... & Chalumeau, M. (2016). Growth monitoring as an early detection tool: a systematic review. The lancet diabetes & endocrinology, 4(5), 447-456.

Scherdel, P., Reynaud, R., Pietrement, C., Salaün, J. F., Bellaïche, M., Arnould, M., ... & Chalumeau, M. (2017). Priority target conditions for algorithms for monitoring children's growth: Interdisciplinary consensus. Plos one, 12(4), e0176464. https://doi.org/10.1371/journal.pone.0176464

Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: addressing ethical challenges. PLoS medicine, 15(11), e1002689. https://doi.org/10.1371/journal.pmed.1002689

Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. BMC medicine, 17, 1-9. https://doi.org/10.1186/s12916-019-1426-2

Seneviratne, M. G., Shah, N. H., & Chu, L. (2020). Bridging the implementation gap of machine learning in healthcare. Bmj Innovations, 6(2).

Leroux, A., Xiao, L., Crainiceanu, C., & Checkley, W. (2018). Dynamic prediction in functional concurrent regression with an application to child growth. Statistics in medicine, 37(8), 1376-1388. https://doi.org/10.1002/sim.7582

Scherdel, P., Matczak, S., Léger, J., Martinez-Vinson, C., Goulet, O., Brauner, R., ... & Heude, B. (2019). Algorithms to define abnormal growth in children: external validation and head-to-head comparison. The Journal of Clinical Endocrinology & Metabolism, 104(2), 241-249. https://doi.org/10.1210/jc.2018-00723

Hilton, C. B., Milinovich, A., Felix, C., Vakharia, N., Crone, T., Donovan, C., ... & Nazha, A. (2020). Personalized predictions of patient outcomes during and after hospitalization using artificial intelligence. NPJ digital medicine, 3(1), 51. https://doi.org/10.1038/s41746-020-0249-z

Ngiam, K. Y., & Khor, W. (2019). Big data and machine learning algorithms for health-care delivery. The Lancet Oncology, 20(5), e262-e273.

Muralidharan, V., Schamroth, J., Youssef, A., Celi, L. A., & Daneshjou, R. (2024). Applied artificial intelligence for global child health: Addressing biases and barriers. PLOS Digital Health, 3(8), e0000583. https://doi.org/10.1371/journal.pdig.0000583

Abouelmehdi, K., Beni-Hessane, A., & Khaloufi, H. Big healthcare data: preserving security and privacy. J. Big Data 5 (1)(2018). https://doi.org/10.1186/s40537-017-0110-7

Khan, F., Kim, J. H., Mathiassen, L., & Moore, R. (2021). Data breach management: An integrated risk model. Information & Management, 58(1), 103392. https://doi.org/10.1016/j.im.2020.103392

Ke, Y., Yang, R., & Liu, N. (2024). Comparing Open-Access Database and Traditional Intensive Care Studies Using Machine Learning: Bibliometric Analysis Study. Journal of Medical Internet Research, 26, e48330. https://doi.org/10.2196/48330