HİTİT MEDICAL JOURNAL
HİTİT ÜNİVERSİTESİ TIP FAKÜLTESİ DERGİSİ

e-ISSN: 2687-4717 Cilt|Volume: 7 • Sayı|Issue: 1 - Şubat|February 2025

Evaluation of Ultrasound Imaging in Developmental Hip Dysplasia with Artificial Intelligence

Gelişimsel Kalça Displazisinde Ultrason Görüntülemenin Yapay Zeka ile Değerlendirilmesi

Hacı Ali Olçar[1,*] | Ahmet Sertol Köksal[2] | Onur Altıntaş[3] | Bülent Turan[3] | Göker Yurdakul[1] Satuk Buğrahan Yinanç[4] | Gürol Göksungur[5] | Burak Çakar[6] | Murat Korkmaz[1]

[1]Yozgat Bozok University, Department of Orthopaedics and Traumatology, Yozgat, Türkiye
[2]Yozgat Bozok University, Department of Computer Engineering, Yozgat, Türkiye
[3]Gaziosmanpaşa University, Department of Computer Engineering, Tokat, Türkiye
[4]Yozgat Bozok University, Sarıkaya School of Physiotherapy and Rehabilitation, Yozgat, Türkiye
[5]Yozgat City Hospital, Department of Radiology, Yozgat, Türkiye
[6]Erciyes University, Department of Orthopaedics and Traumatology, Kayseri, Türkiye

Sorumlu Yazar | Correspondence Author
Hacı Ali Olçar
h.ali.olcar@bozok.edu.tr
Address for Correspondence: Çapanoğlu Mahallesi Celal Köksal Sokak No 10 Yozgat /Merkez

Makale Bilgisi | Article Information
Makale Türü | Article Type: Araştırma Makalesi | Research Article
Doi: https://doi.org/10.52827/hititmedj.1472551
Geliş Tarihi | Received: 23.04.2024
Kabul Tarihi | Accepted: 16.10.2024
Yayım Tarihi | Published: 25.02.2025

Atıf | Cite As
Olçar HA, Köksal AS, Altıntaş O, et. al. Evaluation of Ultrasound Imaging in Developmental Hip Dysplasia with Artificial Intelligence. Hitit Medical Journal 2025;7(1):78-87. https://doi.org/10.52827/hititmedj.1472551

Hakem Değerlendirmesi: Alan editörü tarafından atanan en az iki farklı kurumda çalışan bağımsız hakemler tarafından değerlendirilmiştir.
Etik Beyanı: Çalışma, Yozgat Bozok Üniversitesi Tıp Fakültesi Etik Kurulu tarafından 10.02.2022 tarihinde onaylandı (EK No: 2017-KAEK-189_2022.02.10_03).
İntihal Kontrolleri: Evet (iThenticate)
Çıkar Çatışması: Yazarlar çalışma ile ilgili çıkar çatışması beyan etmemiştir.
Şikayetler: hmj@hitit.edu.tr
Katkı Beyanı: Fikir/Hipotez: HAO Tasarım: MK, GY Veri Toplama/Veri İşleme: GY, GG, ASK Veri Analizi: ASK, OA, BT, SBY, BÇ Makalenin Hazırlanması: HAO, ASK
Hasta Onamı: Katılımcıların ebeveynlerinden yazılı onam alındı.
Finansal Destek: Bu çalışma ile ilgili herhangi bir finansal kaynaktan yararlanılmamıştır.
Bilgi: Yozgat Bozok Üniversitesinin 16-17-18 Mayıs 2024 tarihlerinde gerçekleştireceği “Endüstri 5.0, Yapay Zekâ ve Dijitalleşme” temalı sempozyuma bildiri olarak gönderilmiştir.

Peer Review: Evaluated by independent reviewers working in the at least two different institutions appointed by the field editor.
Ethical Statement: The study was approved by the Yozgat Bozok University Faculty of Medicine Ethics Committee on 10.02.2022 (IRB No. 2017-KAEK-189_2022.02.10_03)
Plagiarism Check: Yes (iThenticate)
Conflict of Interest: The authors declared that, there are no conflicts of interest.
Complaints: hmj@hitit.edu.tr
Authorship Contribution: Idea/Hypothesis: HAO Design: MK, GY Data Collection/Data Processing: GY, GG, ASK Data Analysis: ASK, OA, BT, SBY, BÇ Manuscript Preparation: HAO, ASK
Informed Consent: Our University Clinical Research Ethics Committee has confirmed that no consent for publication from patients is required.
Financial Disclosure: There are no financial funds for this article.
Information: It was sent as a paper to the symposium themed “Industry 5.0, Artificial Intelligence and Digitalization” to be held by Yozgat Bozok University on 16-17-18 May 2024.

Evaluation of Ultrasound Imaging in Developmental Hip Dysplasia with Artificial Intelligence

ABSTRACT

Objective: Developmental hip dysplasia is a common condition that starts in infancy. With the introduction of machine learning (artificial intelligence, AI) into medicine, the early diagnosis of disease and the success of treatment have increased significantly. This study aims to determine the accuracy of ultrasound images from ultrasound videos used in the developmental hip dysplasia screening program using machine learning techniques.

Material and Method: The study involved the extraction of ultrasound image features using the Local Binary Pattern (LBP) method. The ultrasound image dataset was then prepared to evaluate the effectiveness of various machine learning approaches, including Decision Tree (DT), Random Forest (RF), K-Nearest Neighbour (KNN), Gradient Boosting (GB), Support Vector Machines (SVM), Naïve Bayes (NB), Logistic Regression (LR), and Multilayer Perceptron (MLP).

Results: RF algorithm performed very well, recording the highest correct image rate. The study was generally considered successful and it is believed that the resulting model will be useful in the early diagnosis of developmental hip dysplasia.

Conclusion: RF algorithm recorded the highest correct image rate, performing very well at 87.62% compared to other tested algorithms. The study was generally considered successful and the resulting model is believed to be useful in the early diagnosis of developmental hip dysplasia.

Keywords: Algorithms, classification, deep learning, developmental dysplasia of the hip, machine learning, artificial intelligence, ultrasonography.

ÖZET

Amaç: Gelişimsel kalça displazisi, bebeklik döneminde ortaya çıkan yaygın bir hastalıktır. Makine öğreniminin (yapay zeka, AI) tıp alanına girmesi ile hastalıkların erken tanısını ve tedavi başarısını önemli oranda artırmaktadır. Bu çalışma, makine öğrenme tekniklerini (yapay zeka) kullanarak gelişimsel kalça displazisi tarama programında kullanılan ultrason videolarından ultrason görüntüsünün doğruluğunun belirlenmesi amaçlamaktadır.

Gereç ve Yöntem: Çalışma, Lokal İkili Model (LBP) metodolojisi aracılığıyla ultrason görüntü özelliklerinin çıkarılmasını içeriyordu. Daha sonra, 'Decision Tree (DT), Random Forest (RF), K-Nearest Neighbour (KNN), Gradient Boosting (GB), Support Vector Machines (SVM), Naïve Bayes (NB), Linear Regression (LR), and Multilayer Perceptron (MLP)' dahil olmak üzere farklı makine öğrenimi yaklaşımlarının etkinliğini değerlendirmek için ultrason görüntü veri kümesini hazırlandı.

Bulgular: RF algoritması, test edilen diğer algoritmalarla karşılaştırıldığında %87,62 ile çok iyi performans göstererek en yüksek doğru görüntü oranını kaydetti. Çalışma genel olarak başarılı kabul edildi ve ortaya çıkan modelin gelişimsel kalça displazisinin erken teşhisinde faydalı olacağına inanılıyor.

Sonuç: RF algoritması en yüksek doğru görüntü oranını kaydederek çok iyi bir performans sergilemektedir. Çalışma genel olarak başarılı sayıldı ve elde edilen modelin gelişimsel kalça displazisi' nin erken teşhisinde yardımcı olacağı düşünülmektedir.

Anahtar Sözcükler: Algoritma, derin öğrenme, gelişimsel kalça displazisi, makine öğrenmesi, yapay zeka, sınıflandırma, ultrasonografi.

Introduction

Developmental dysplasia of the hip (DDH) is a frequently-occurring ailment among infants, caused by genetic, intrauterine and cultural factors. It has an incidence rate ranging between 2% and 5% (1). The condition can be grouped into various types, including acetabular dysplasia, instability, subluxation and complete dislocation (2). Failure to promptly receive a diagnosis and treatment may result in more intricate procedures, including hip replacement, in the future (3).

Current clinical practice for DDH entails clinical examination, assessment of risk factors, and radiological imaging (4,5). Radiography is widely regarded as an effective diagnostic tool for DDH, notwithstanding its potential drawbacks in terms of radiation exposure and limited information for neonatal hips due to incomplete ossification of cartilage. Ultrasonography is an efficient method for screening DDH in the first six months of an infant's life because it facilitates comprehensive static and dynamic imaging of the hip joint. However, ultrasound imaging of the hip joint poses technical challenges. The ultrasound technique for assessing the condition of infants' hips was introduced by Graf in the 1980s (6). Currently, Graf's approach is the most widely used technique for diagnosing DDH using coronal plane ultrasound images. However, there are numerous ongoing debates about the effectiveness of ultrasound in enabling early and precise diagnosis and guidance for treatment (4,5). In keeping with Graf's methodology, measurement and classification are executed based on the standard plane, encompassing the iliac crest, the base of the acetabulum, and the acetabular labrum. However, establishing the precise plane that demarcates the image slice of a neonatal hip is often difficult, and drawing three lines heavily depends on the physician's precise application (7,8).

The extensive utilization of machine learning and deep learning techniques has been observed for the diagnosis of various diseases (9). Among deep learning architectures that have been described in existing literature, Convolutional Neural Networks (CNNs) are frequently applied (10). Automatic classification of hip dysplasia is a recent and developing field, involving data collection, data preprocessing, feature extraction, training, and testing on images (11).

In recent years, researchers have demonstrated that a close connection between engineering labs and real-world practice is necessary to achieve meaningful results in ultrasound research (12). Hip ultrasonography is the established diagnostic tool for DDH necessitating precise assessment with standard plane images to guarantee proper diagnosis. Nonetheless, it has been reported that imaging may be undependable, and, thus, accurate imaging tests are indispensable for precise diagnosis. Incorrect body positioning or poor image quality may compromise the quality of images, significantly affecting diagnosis impacts. In recent years, scholars have conducted research on the efficacy of deep learning methods to offer automated assurance of precise diagnosis by operators (11,13,14). This can prove to be a useful aid for inexperienced operators.

The evaluation of the ultrasound image used in the developmental hip dysplasia screening programme consists of three steps. The first step is to correctly locate the reference points on the ultrasound, the second step is to draw lines from the correctly located reference points and make angle measurements according to these lines, and the third step is to stage the hip according to the degree of measurement as a normal hip or a DHH hip. There are studies on the second and third steps (11, 14). There are not enough studies in the literature on the first step, where physician experience is important.

The study's significance lies in the scarcity of research into the precision of images used in diagnosing DDH and the approaches taken. It offers evidence that machine learning techniques can automatically identify appropriate images. A novel dataset and unexplored methodology are employed in the study, which fills a gap in the research and results in acceptable precision and swift functionality. This highlights the unique value of our study. The aim of our study is to use artificial intelligence to distinguish between the presence (correct image) or absence (incorrect image) of reference points used in the diagnosis of DDH in ultrasound images taken during DDH scans. Thus, artificial intelligence that can perform ultrasound evaluation can pave the way for young physicians to meet their need for experience.

Material and Method

Ethical Approval

The study was approved by the Yozgat Bozok University Faculty of Medicine Ethics Committee on 10.02.2022 (IRB No. 2017-KAEK-189_2022.02.10_03) in accordance with the Declaration of Helsinki, and written informed consent was obtained from the parents of the participants.

Study Design and Data Collection

In this study, ultrasound records "ACUSON S2000 Ultrasound System, HELX Evolution with Touch Control (Siemens Healthineers, Germany") of both the right and left hips of 100 infants collected from Yozgat City Hospital were examined. The Local Binary Patterns (LBP) texture analysis method was utilised for gathering features from these images, and subsequently building a dataset. In the study, 80% of the dataset was used as training dataset to train machine learning algorithms to classify images according to their correctness. The remaining 20% of the dataset were reserved for testing purposes. A range of machine learning algorithms, such as Support Vector Machines (SVM), Naive Bayes (NB), Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbours (KNN), Decision Tree (DT), Gradient Boosting (GB), and Multilayer Perceptron (MLP), were employed to perform this classification. Inclusion criteria: Infants aged 40-45 days, standard cross-sectional coronal plane ultrasound scans performed in the lateral decubitus position, using a 7.5 MHz linear ultrasound probe, acquisition of images representing normal or dysplastic hips, with measurement reference points specified in the hip scan programme.

Exclusion criteria: Infants with a history of previous hip surgery or unrelated congenital anomalies, presence of neuromuscular diseases or syndromes that may contribute to congenital hip dislocation, images taken in positions other than lateral decubitus, hip ultrasound measurements outside the specified age range of 40-45 days, situations where the infant's distress prevents imaging, ultrasound scans performed for indications other than DDH.

The study utilized Python® programming language, along with the Scikit-Learn® and OpenCV® libraries. The experimental setup included a computer with a 12th generation Intel® Core™ i7 processor, 32 GB RAM, and a 1 TB M.2 SSD.

Dataset and Data Augmentation

In the study, 200 video recordings were created from both the right and left hips of one hundred newborns. Then, eighteen images were captured from each video recording at equal intervals (three images per second). Thus, a dataset consisting of three thousand six hundred images was obtained. Neonatal hip ultrasound was performed with the infant in the lateral decubitus position, with the hip and knee in semi-flexion and 15-20 degrees of internal rotation (6). In this position, the greater trochanter of the femur was imaged exactly laterally, under the probe, in the anteroposterior plane. The anatomical structures of the tissues of a healthy developed hip are shown schematically (Figure I). A standard section was used in the coronal plane so that the ultrasound scans could be compared and everyone could take measurements in the same plane (6). There are three important reference points that should be present in the standard section: The ilium should be parallel to the skin, the labrum should be visible, and the ossified end of the ilium should be visible in the acetabulum (6).

Hip ultrasound images were obtained from healthy or dysplastic newborns aged 40-45 days in the newborn hip screening programme. Ultrasound images from the screening programme were graded by an experienced radiologist and a paediatric orthopaedic specialist on the basis of whether the 3 reference points, for which physician experience is important, were correctly included in the image area. Failure to detect the three reference points used to diagnose and classify DDH (acetabular labrum, acetabular edge and inferior iliac edge) was used as a faulty image, and the ability to detect this was used as a correct image in the artificial intelligence training.

The images were evaluated by a radiologist and an experienced paediatric orthopaedic physician. The radiologist categorized the images as either correct or incorrect (Figure I), identifying 1522 as correct and 2078 as incorrect. These experts also evaluated these images and determined that 8 hips were Type 2 and 192 hips were Type 1. This shows that the dataset contains data from both healthy and unhealthy individuals.

Feature extraction is a widely used procedure in machine learning that involves picking a subset of data-specific features to apply a learning algorithm. The ultimate goal is to determine the most minimal number of features for a given problem domain that precisely represent it (15). One well-known method for accomplishing this is LBP.

The LBP works with greater efficiency and reliability on grayscale images due to its capacity to process only one operation per pixel. Besides being user-friendly, it offers improved speed advantages for handling larger datasets and real-time applications (16). This quality of LBP was the primary reason for utilizing it in this study.

The LBP method of extracting features was employed on each classified image, resulting in 10 column features being extracted. An extra column containing the class feature (correct or incorrect) was also included, resulting in a 1x11 array for each image. The features of every image were then added as lines in a CSV file, resulting in the creation of a dataset measuring 3600x11. Data augmentation is a technique that can enhance the effectiveness of machine learning and deep learning algorithms and regularize the datasets used by (17).

Initially, it was believed that any lack of balance between the classes within the dataset could negatively impact the success rate. To address this concern, different data augmentation methods were employed, including vertical and horizontal shift, as well as cropping and scaling techniques. As a consequence, the quantity of data assigned to the correct class rose to 2078, thus rectifying the imbalance between the classes. This resulted in a dataset consisting of 4156 data points, which were utilised to develop and test the model as per the machine learning methodologies detailed in section 2. The success rates achieved are presented in the Results section.

After rectifying the unevenness between the classes in the dataset, data expansion techniques were implemented once more to amplify the dataset. Nevertheless, as no enhancement on the achievement rate was witnessed, the study continued with a dataset comprising 4156 data points.

Machine Learning Algorithms

The research employed the widely accepted K-Nearest Neighbours machine learning algorithm that is commonly utilised in regression and classification tasks. This algorithm allows a point to be classified based on the closest previously classified K number of points. Its primary aim is to use majority voting for the label prediction of a test data point (18).

Decision trees represent a classification algorithm that iteratively segments the dataset. This tree features a root node, internal nodes as well as terminal nodes (leaves), while the nodes of the tree consist of characteristics from the dataset, and the leaves are composed of predetermined values located in the outcome column. Each new example in the dataset is allocated to an appropriate class within the tree based on the pre-existing attribute values. To perform this placement, the relevant instance is moved within the tree and its classification is determined (19).

Gradient Boosting is a machine learning algorithm used for classification. The process involves training models in succession, with each subsequent model aiming to reduce errors from the previous one. The following steps are taken: Firstly, a model is created for predicting from the training dataset. Residuals are then calculated based on the difference between observed and predicted values. A new model is subsequently generated from the residuals to produce fresh predictions. Following this, the sequence is repeated until the residuals are reduced to a minimum or a specific threshold value is attained (19).

The Random Forest algorithm is a machine learning technique that was developed in 2001 by Leo Breiman (20). It can be used for both regression and classification problems, and its structure is composed of numerous decision trees. By processing data using multiple decision trees, the algorithm is able to provide precise predictions by averaging the acquired predictions. This approach also helps to prevent overfitting, which is a common issue in the decision tree method.

Naive Bayes is a statistical algorithm for classification based on probabilities. It is trained for supervised learning and applicable in practical real-world scenarios. The approach is simple as it uses Bayes' theorem to calculate probabilities, making it easy to understand and useful in limiting complexity. The algorithm determines the probability of an instance belonging to a target category, making it a valuable tool in various applications (21).

Table I. Evaluation metrics for machine learning methods employed

Model	RF	GB	KNN	DT	MLP	LR	NB	SVM
Accuracy	0.8762	0.8582	0.8474	0.8125	0.7536	0.6478	0.6214	0.5853
Sensitivity	0.8945	0.8784	0.8601	0.8326	0.7638	0.7041	0.7706	0.7569
Specificity	0.8561	0.8359	0.8333	0.7904	0.7424	0.5859	0.4571	0.3965
PPV	0.8725	0.8549	0.8503	0.8139	0.7655	0.6518	0.6098	0.5800
F1 Score	0.8834	0.8665	0.8552	0.8231	0.7646	0.6770	0.6809	0.6567
Youden J Index	0.7506	0.7143	0.6934	0.6230	0.5062	0.2900	0.2277	0.1533

RF: Random Forest, GB: Gradient Boosting, KNN: K-Nearest Neighbours, DT: Decision Tree, MLP: Multilayer Perceptron, LR: Logistic Regression, NB: Naive Bayes, SVM: Support Vector Machine, PPV: Positive Predictive Value.

The Support Vector Machine constitutes another algorithm often employed in the realm of machine learning. SVM aims to distinguish between classes by measuring distances among instances belonging to respective categories. Significantly, to achieve high classification success, the hyperplane must remain distant from data points of other classes, a feat accomplished through the formation of hyperplanes. Furthermore, support vectors are determined by calculating the points closest to the classifier's margin (22).

Table II. Resource usage of machine learning techniques

Model	RF	NB	MLP	LR	KNN	DT	GB	SVM
Training time (sec)	0.8604	0.003	3.7491	0.0312	0	0.0227	3.6078	0.1884
Memory usage in training (MB)	3.4922	0.0117	1.6094	0.25	0.0977	0	1.4844	0.5977
CPU usage in training (%)	41.8	0	0	0	0	0	61.5	83.6
Test time (sec)	0	0	0	0	0.0156	0	0	0.0313
Memory usage in test (MB)	0.0078	0	0.2734	0	0.0547	0	0.0117	0.1445
CPU usage in test (%)	0	0	0	0	0	0	0	48.8

(RF: Random Forest, GB: Gradient Boosting, KNN: K-Nearest Neighbours, DT: Decision Tree, MLP: Multilayer Perceptron, LR: Logistic Regression, NB: Naive Bayes, SVM: Support Vector Machine).

Logistic regression is a widely utilized model in machine learning, particularly effective in binary classification problems. This model analyses relationships within a dataset, modelling the connection between input variables and the probability of an event, making it a machine learning model that generates predictions (23).

Table III. Resource usage of machine learning techniques

Model	RF	NB	MLP	LR	KNN	DT	GB	SVM
Training time (sec)	0.8604	0.003	3.7491	0.0312	0	0.0227	3.6078	0.1884
Memory usage in training (MB)	3.4922	0.0117	1.6094	0.25	0.0977	0	1.4844	0.5977
CPU usage in training (%)	41.8	0	0	0	0	0	61.5	83.6
Test time (sec)	0	0	0	0	0.0156	0	0	0.0313
Memory usage in test (MB)	0.0078	0	0.2734	0	0.0547	0	0.0117	0.1445
CPU usage in test (%)	0	0	0	0	0	0	0	48.8

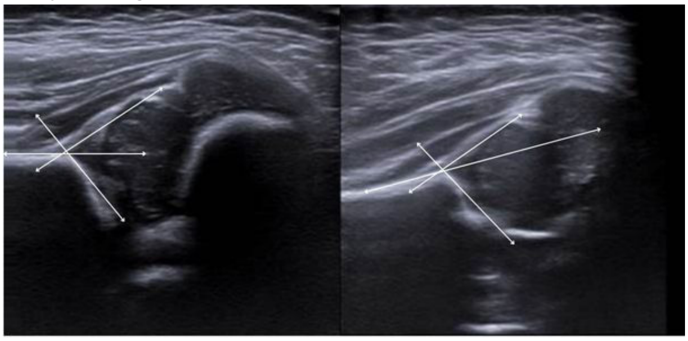
RF: Random Forest, GB: Gradient Boosting, KNN: K-Nearest Neighbours, DT: Decision Tree, MLP: Multilayer Perceptron, LR: Logistic Regression, NB: Naive Bayes, SVM: Support Vector Machine

Multilayer perceptions are often used for tasks such as recognizing patterns, interpolating, and classifying data. They are an improved type of the Perception neural network, which was developed in the early 1960s and has several limitations (24).

Statistical Analysis

The learning performance of different algorithms was evaluated using the traditional Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC). Accuracy, sensitivity, specificity, positive predictive value, F1 score, and Youden J Index were calculated for each algorithm (Figure II). The F1 Score is used in statistical analysis to assess the balance between precision and recall in classification tasks, providing a single value that combines both measures and is particularly useful for evaluating the overall performance of machine learning models (25). The Youden J Index, employed in statistical analysis, quantifies the overall accuracy of a diagnostic or classification test by optimizing the trade-off between sensitivity and specificity, offering a single metric that encapsulates the model's ability to correctly identify both positive and negative instances (26). "IBM SPSS Statistics for Windows, Version 27.0" program was used for the ROC (receiver operating characteristic) curve and AUC (Area under the ROC Curve). The main criterion for evaluating the effectiveness of a classification algorithm is the accuracy rate (27).

Figure I. Correct and Incorrect Images from the dataset, the iliac bone should be parallel to the skin, the Labrum should be visible, and the ossified end of the ilium should be seen in the acetabulum (6). Change in the drawing due to the iliac bone not being parallel to the skin, left is correct, right is incorrect. sample image.



Results

The average age of the babies in the study was 40-45 days. 47% of the participants were girls and 53% were boys. The distribution of hips studied was 50% left hip and 50% right hip. Of the infants, 60 were the first babies in the family and 40 were the next babies in the family. There were 45 caesarean sections and 55 normal births.

Figure II. The calculation formulas (TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative, PPV: Positive Predictive Value).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{F1 Score} = 2 * \frac{\text{Sensitivity} * \text{PPV}}{\text{Sensitivity} + \text{PPV}}$$

$$\text{J Statistic Youden Index} = \text{Sensitivity} + \text{Specificity} - 1$$

Table I of the study displays the accuracy, sensitivity, specificity, positive predictive value, F1 score, and Youden J Index of the applied machine learning techniques. Examination of the results suggests that the most effective approach was the RF method, as evidenced by sensitivity, specificity, positive predictive value, F1 score, and Youden J Index of 0.8762, 0.8945, 0.8561, 0.8725, 0.8834, and 0.7506, respectively. The ROC curves of different machine learning techniques are shown in Figure III.

Figure III. ROC Curves of Machine Learning Techniques (RF: Random Forest, GB: Gradient Boosting, KNN: K-Nearest Neighbours, DT: Decision Tree, NB: Naive Bayes, MLP: Multilayer Perceptron, SVM: Support Vector Machine, LR: Logistic Regression)

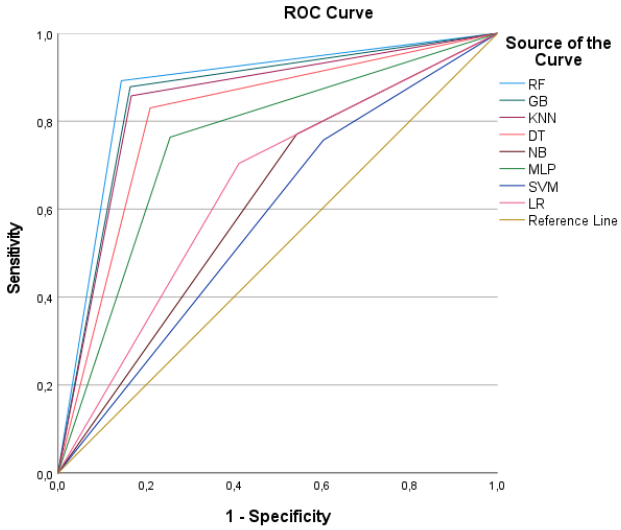


Table II presents a comparison of the training and test times, training and test memory usage, and training and test processor usage across the algorithms utilized in the study. Some of the algorithms recorded a resource usage of zero due to their minimal resource consumption. Notably, the RF algorithm demonstrated impressive performance while utilizing the least resources in test time (0), test memory usage (0.0078), and test processor usage (0). The test time and processor usage are displayed as zero because they are very close to zero. The system under study displays them as zero.

Discussion

Acquiring the appropriate image is crucial for automating an accurate diagnosis of DDH. This research employed machine learning algorithms to obtain accurate images from ultrasound recordings. A specially prepared dataset measuring 4156x11 was utilised in this study. Eight varied machine learning algorithms were tested, and the Random Forest algorithm presented the highest rate of achievement and productivity.

AI's image detection capabilities have been reported in various fields, including the detection of lung CT nodules, automatic detection of COVID-19 using chest X-ray images, and the identification of brain tumors and Alzheimer's lesions through brain MRI scans (13). CNNs are purportedly benefit from diagnosing DDH by examining anteroposterior pelvic radiographs (28). Zhang and colleagues (28) have reported that a deep learning system was trained and optimized using 9081 radiographs. Subsequently, 1138 test radiographs were employed to compare diagnoses made by both the deep learning system and clinicians. The deep learning system's diagnostic accuracy for identifying hip dislocations was determined by the area under the receiver operating characteristic curve (AUC), sensitivity and specificity, scoring 0.975, 276/289 (95.5%) and 1978/1987 (99.5%), respectively.

Park and colleagues (29) assessed a deep learning algorithm's diagnostic accuracy in automatically detecting DDH using anteroposterior radiographs. The study included 5076 hip images from patients up to 12 months old with suspected DDH. The deep learning algorithm exhibited a sensitivity of 98.0%, specificity of 98.1%, positive predictive value of 84.5%, and negative predictive value of 99.8%. There was no significant disparity in DDH diagnosis between the algorithm and experienced paediatric radiologists. However, radiologists lacking in paediatric radiology experience had lower sensitivity, specificity, and positive predictive value when compared with the proposed model. Xu et al. carried out another x-ray imaging study to detect DDH, achieving a classification precision of 95% through the usage of 1265 patient images. Mask RCNN was employed for detecting local features (30). Pham et al. created an automated measurement for migration percentage on pelvis radiographs with respect to hip dysplasia using CNN. Several deep learning algorithms were tested, and maximum accuracy of 94.5% was achieved (31).

This study is based on ultrasound images and machine learning for automatic detection of DDH. Automatic detection studies for DDH, based on ultrasound images, are prevalent in the literature with numerous publications supporting the findings of this study. Hu et al. developed a multi-task framework that automated the evaluation of DDH through the use of Mask R-CNN. Their recommended approach yields 93% accuracy in identifying alpha angles below 5 degrees (32). Liu et al. (33) proposed a feature attention network to improve the accuracy of angle measurement for neonatal femur segmentation using ultrasound. The network utilises 400 images obtained from a publicly available dataset, resulting in a reduction of doctors' error rates from 6-10% to 2%. Another related study by Chen et al. presented a deep learning-based computer-aided framework for diagnosing DDH. The framework automatically detects standard planes and measures angles for Graft type I and type II hips. They reported a classification accuracy of 94.71%. Additionally, they explored a standard plane scoring module to calculate the scoring formula and identify the most suitable ultrasound image, which is similar to the proposed study (14).

Huang et al. (34) constructed a network in accordance with the guidelines of the American College of Radiology (ACR) and the American Institute of Ultrasound in Medicine (AIUM). Their goal was to devise an innovative technique for the automatic measurement of ultrasound-based DDH using deep neural networks (DNN).

Based on the findings presented in Table III, our study demonstrates that integrating the model into a real-time decision support system is easily achievable. The model demonstrated outstanding performance with the highest accuracy value of 0.8762. The precision, recall, and F1-score metrics are consistent with this outcome, indicating the reliability of the model. Additionally, the confusion matrix in Figure I shows that the model has an accuracy rate of 89.45% in identifying false images. These results indicate that the likelihood of selecting an incorrect image is only 10.55%.

Well-trained and accurate ultrasound images are crucial in all of these studies. If individuals lack the experience to precisely measure the angles of the images, the diagnosis of DDH may be compromised. The acquisition of reliable and precise images is essential for correctly diagnosing DDH (32, 34). Our goal is to mitigate the issues arising from inaccurate image acquisition and evaluation. We assume that it is possible to help less experienced people by defining what is true and false of the image. This problem is our motivation to carry out this study using 3600 photographs taken from 100 subjects. In this study, 8 different machine learning algorithms were used to develop the correct image decision-making support system.

The study has limitations, notably a small sample size. Further research with a larger sample size is required to enable generalization of the results. Additionally, our study did not employ the Graf classification system, precluding knowledge of the sample types according to Graf. This is an important consideration for treatment decisions. Although the determination of measurement points is a matter of physician experience, in our study there were no reference points that could not be determined due to image quality and baby age. The effect of the ultrasound machine used, the image quality and the baby's age on the clinician's ultrasound scan can be evaluated in future studies.

Conclusion

Machine learning algorithms are capable of evaluating ultrasound images for the diagnosis of developmental dysplasia of the hip (DDH). Early-stage diagnostic accuracy can be enhanced through the identification of appropriate images by physicians. Furthermore, the integration of machine learning into ultrasound devices may reduce the occurrence of false evaluations in DDH diagnosis.

Our study also offers valuable insights into diagnostic approaches by demonstrating the potential of artificial intelligence to differentiate between normal hip cases and DDH cases without the need for angle measurements in ultrasound images.

References

1. Harsanyi S, Zamborsky R, Krajciova L, Kokavec M, Danisovic L. Developmental Dysplasia of the Hip: A Review of Etiopathogenesis, Risk Factors, and Genetic Aspects. Medicina (Kaunas, Lithuania) 2020; 56.
2. Tao Z, Wang J, Li Y, Zhou Y, et al. Prevalence of developmental dysplasia of the hip (DDH) in infants: a systematic review and meta-analysis. BMJ Paediatr Open 2023;7(1):e002080.
3. He J, Chen T, Lyu X. Analysis of the results of hip ultrasonography in 48 666 infants and efficacy studies of conservative treatment. J Clin Ultrasound 2023;51:656-662.
4. Shipman SA, Helfand M, Moyer VA, Yawn BP. Screening for developmental dysplasia of the hip: a systematic literature review for the US Preventive Services Task Force. Pediatrics 2006;117:e557-576.
5. Omeroğlu H. Use of ultrasonography in developmental dysplasia of the hip. J Child Orthop 2014;8:105-113.
6. Graf R. Classification of hip joint dysplasia by means of sonography. Arch Orthop Trauma Surg (1978) 1984;102:248-255.
7. Jaremko JL, Mabee M, Swami VG, Jamieson L, Chow K, Thompson RB. Potential for change in US diagnosis of hip dysplasia solely caused by changes in probe orientation: patterns of alpha-angle variation revealed by using three-dimensional US. Radiology 2014; 273:870-878.
8. Pedrotti L, Crivellari I, Degrate A, De Rosa F, Ruggiero F, Mosconi M. Interpreting neonatal hip sonography: intraobserver and interobserver variability. J Pediatr Orthop B 2020;29:214-218.
9. Arslan RS, Ulutas H, Köksal AS, Bakir M, Çiftçi B. Sensitive deep learning application on sleep stage scoring by using all PSG data. Neural Computing and Applications 2023; 35:7495-7508.
10. Sharma N, Jain V, Mishra A. An analysis of convolutional neural networks for image classification. Procedia computer science 2018;132:377-384.
11. Sezer A, Sezer HB. Deep Convolutional Neural Network-Based Automatic Classification of Neonatal Hip Ultrasound Images: A Novel Data Augmentation Approach with Speckle Noise Reduction. Ultrasound Med Biol 2020;46:735-749.
12. Pandey M, Fernandez M, Gentile F, et al. The transformational role of GPU computing and deep learning in drug discovery. Nature Machine Intelligence 2022;4:211-221.
13. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med 2020;121:103792.
14. Chen T, Zhang Y, Wang B, et al. Development of a Fully Automated Graf Standard Plane and Angle Evaluation Method for Infant Hip Ultrasound Scans. Diagnostics (Basel) 2022; 12.

15. Ladha L, Deepa T. Feature selection methods and algorithms. International journal on Computer Science and Engineering 2011; 3:1787-1797.
16. Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. Pattern Recognition 1996; 29:51-59.
17. Nalepa J, Marcinkiewicz M, Kawulok M. Data Augmentation for Brain-Tumor Segmentation: A Review. Front Comput Neurosci 2019; 13:83.
18. Mucherino, A., Papajorgji, P.J., Pardalos, P.M. (2009). k-Nearest Neighbor Classification. In: Data Mining in Agriculture. Springer Optimization and Its Applications, vol 34. Springer, New York, NY. 2009:83-106.
19. Natekin A, Knoll A. Gradient boosting machines, a tutorial. Frontiers in Neurorobotics 2013;7:21.
20. Breiman L. Random forests. Machine Learning 2001; 45:5-32.
21. Soria D, Garibaldi JM, Ambrogi F, Biganzoli EM, Ellis IO. A 'non-parametric'version of the naive Bayes classifier. Knowledge-Based Systems 2011; 24:775-784.
22. Mammone A, Turchi M, Cristianini N. Support vector machines. Wiley Interdisciplinary Reviews: Computational Statistics 2009; 1:283-289.
23. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. Journal of Biomedical Informatics 2002; 35:352-359.
24. Murtagh F. Multilayer perceptrons for classification and regression. Neurocomputing 1991; 2:183-197.
25. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 2020; 21:6.
26. Ruopp MD, Perkins NJ, Whitcomb BW, Schisterman EF. Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection. Biom J 2008; 50:419-430.
27. Lu D, Weng Q. A survey of image classification methods and techniques for improving classification performance. International journal of Remote sensing 2007; 28:823-870.
28. Zhang SC, Sun J, Liu CB, Fang JH, Xie HT, Ning B. Clinical application of artificial intelligence-assisted diagnosis using anteroposterior pelvic radiographs in children with developmental dysplasia of the hip. Bone Joint J 2020; 102-b: 1574-1581.
29. Park HS, Jeon K, Cho YJ, et al. Diagnostic Performance of a New Convolutional Neural Network Algorithm for Detecting Developmental Dysplasia of the Hip on Anteroposterior Radiographs. Korean J Radiol 2021; 22:612-623.
30. Xu W, Shu L, Gong P, et al. A Deep-Learning Aided Diagnostic System in Assessing Developmental Dysplasia of the Hip on Pediatric Pelvic Radiographs. Front Pediatr 2021; 9:785480.
31. Pham TT, Le MB, Le LH, Andersen J, Lou E. Assessment of hip displacement in children with cerebral palsy using machine learning approach. Med Biol Eng Comput 2021; 59:1877-1887.
32. Hu X, Wang L, Yang X, et al. Joint Landmark and Structure Learning for Automatic Evaluation of Developmental Dysplasia of the Hip. IEEE J Biomed Health Inform 2022; 26:345-358.
33. Liu R, Liu M, Sheng B, et al. NHBS-Net: A Feature Fusion Attention Network for Ultrasound Neonatal Hip Bone Segmentation. IEEE Trans Med Imaging 2021; 40:3446-3458.
34. Huang B, Xia B, Qian J, et al. Artificial Intelligence-Assisted Ultrasound Diagnosis on Infant Developmental Dysplasia of the Hip Under Constrained Computational Resources. J Ultrasound Med 2023; 42:1235-1248.