









Article

Machine Learning Evaluation of Biliary Atresia Patients to Predict Long-Term Outcome after the Kasai Procedure

Martina Caruso ^{1,*}, Carlo Ricciardi ^{2,3}, Gregorio Delli Paoli ¹, Fabiola Di Dato ⁴, Leandro Donisi ^{1,3}, Valeria Romeo ¹, Mario Petretta ^{4,†}, Raffaele Iorio ⁴, Giuseppe Cesarelli ^{3,5}, Arturo Brunetti ¹ and Simone Maurea ¹

- ¹ Department of Advanced Biomedical Sciences, University of Naples “Federico II”, 80131 Naples, Italy; greg.dellipaoli@gmail.com (G.D.P.); lean.donisi@gmail.com (L.D.); valeria.romeo@unina.it (V.R.); brunetti@unina.it (A.B.); maurea@unina.it (S.M.)
- ² Department of Electrical Engineering and Information Technology, University of Naples “Federico II”, 80125 Naples, Italy; carloricciardi93@gmail.com
- ³ Bioengineering Unit, Institute of Care and Scientific Research Maugeri, 82037 Telese Terme, Italy; giuseppe.cesarelli@unina.it
- ⁴ Department of Translational Medical Sciences, University of Naples “Federico II”, 80131 Naples, Italy; fabydd88@gmail.com (F.D.D.); petretta@unina.it (M.P.); raffaele.iorio@unina.it (R.I.)
- ⁵ Department of Chemical, Materials and Production Engineering, University of Naples “Federico II”, 80125 Naples, Italy
- * Correspondence: caruso.martina90@gmail.com; Tel.: +39-0817463560
- † Current affiliation: IRCCS SDN, 80143 Naples, Italy.



Citation: Caruso, M.; Ricciardi, C.; Delli Paoli, G.; Di Dato, F.; Donisi, L.; Romeo, V.; Petretta, M.; Iorio, R.; Cesarelli, G.; Brunetti, A.; et al. Machine Learning Evaluation of Biliary Atresia Patients to Predict Long-Term Outcome after the Kasai Procedure. *Bioengineering* **2021**, *8*, 152. <https://doi.org/10.3390/bioengineering8110152>

Academic Editor: Luca Mesin

Received: 23 September 2021
Accepted: 19 October 2021
Published: 22 October 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Kasai portoenterostomy (KP) represents the first-line treatment for biliary atresia (BA). The purpose was to compare the accuracy of quantitative parameters extracted from laboratory tests, US imaging, and MR imaging studies using machine learning (ML) algorithms to predict the long-term medical outcome in native liver survivor BA patients after KP. Twenty-four patients were evaluated according to clinical and laboratory data at initial evaluation (median follow-up = 9.7 years) after KP as having ideal ($n = 15$) or non-ideal ($n = 9$) medical outcomes. Patients were re-evaluated after an additional 4 years and classified in group 1 ($n = 12$) as stable and group 2 ($n = 12$) as non-stable in the disease course. Laboratory and quantitative imaging parameters were merged to test ML algorithms. Total and direct bilirubin (TB and DB), as laboratory parameters, and US stiffness, as an imaging parameter, were the only statistically significant parameters between the groups. The best algorithm in terms of accuracy, sensitivity, specificity, and AUCROC was naive Bayes algorithm, selecting only laboratory parameters (TB and DB). This preliminary ML analysis confirms the fundamental role of TB and DB values in predicting the long-term medical outcome for BA patients after KP, even though their values may be within the normal range. Physicians should be alert when TB and DB values change slightly.

Keywords: artificial intelligence; bilirubin; ultrasound; magnetic resonance; shear-wave elastography

1. Introduction

In the past decades, due to the growth of medical information digitalization and thanks to the availability of increasingly sophisticated technological quantitative tools, large volumes of patient data have become widely available. In this scenario, new approaches from computational sciences can be used to analyze medical data to extract critical health information that can help clinicians in the decision-making process and prognostic evaluation [1]. In particular, machine learning (ML) has gained great interest thanks to cheaper computing power and inexpensive memory and also because it is agnostic to the domain of application. It is a methodology of data analysis, a branch of artificial intelligence, that enables systems to learn and improve from data [2]. The ML methodology is

spreading in clinical research with applications in several medical fields, such as neurology, cardiology, ophthalmology, pediatrics, and fetal monitoring [3–11].

Biliary atresia (BA) is a rare cholangiopathy of unknown etiology, which is characterized by inflammatory obliteration of both intrahepatic and extrahepatic bile ducts [12–15]; an early diagnosis is needed, and Kasai portoenterostomy (KP) represents the treatment of choice [16,17]. During the post-surgical follow-up, diagnostic evaluation consists of monitoring clinical and laboratory data as well as performing abdominal ultrasound (US) and magnetic resonance imaging (MR) [18–22]. In the literature, a good correlation of qualitative imaging findings using US and/or MR with the medical outcome of BA patients with native liver after KP during follow-up was described, as well as the potential role of US and MR findings in predicting the long-term medical outcome in such patients [21,22].

The aim of this study was to compare the accuracy of several quantitative parameters extracted from different methodologies, such as laboratory tests, US, and MR imaging, using ML algorithms in predicting the long-term medical outcome for native liver survivor patients with BA who have undergone KP.

2. Materials and Methods

2.1. Patient Population

Native liver survivor patients with BA after KP were retrospectively enrolled from the pediatric liver unit (January 2012 to December 2019). Exclusion criteria were (1) patients with liver transplantation and (2) patients with a time interval between acquisition of imaging studies (US and MR) greater than 30 days. Patients were initially evaluated by clinical, laboratory, and imaging (US and MR) studies to assess the medical outcome after KP. Patients were classified as having an ideal or a non-ideal medical outcome after KP following the criteria suggested by Ng et al. [23] and modified by Lee et al. [18]. An ideal medical outcome was defined as normal laboratory parameters with no evidence of medical complications of chronic liver disease (CLD), while a non-ideal medical outcome was based on at least one abnormal laboratory parameter and/or one CLD medical complication [18], including cholangitis, portal hypertension, variceal bleeding, fractures, hepatopulmonary syndrome, and portopulmonary hypertension. Successively, patients were similarly re-evaluated during long-term follow-up from initial evaluation to assess the disease course as stable or non-stable. The disease course was considered stable when the patient medical outcome remained unchanged at re-evaluation, whereas the disease course was considered non-stable when the patient medical outcome changed at re-evaluation to ideal from non-ideal or to non-ideal in progression; in particular, the status of non-ideal in progression consisted of the occurrence of at least one additional laboratory or clinical abnormality.

2.2. Laboratory Tests

The following laboratory parameters were used: white blood cell (WBC) count (n.v. $> 4000/\text{mm}^3$), platelet (PLT) count (n.v. $> 150,000/\text{mm}^3$), total bilirubin (TB; n.v. $< 1.2 \text{ mg/dL}$), direct bilirubin (DB; n.v. $< 0.5 \text{ mg/dL}$), albumin (n.v. $> 3.5 \text{ g/dL}$), international normalized ratio (INR; n.v. < 1.3), alanine aminotransferase (ALT; n.v. $< 40 \text{ IU/L}$), aspartate aminotransferase (AST; n.v. $< 40 \text{ IU/L}$), and γ -glutamyl transpeptidase (GGT; n.v. $< 55 \text{ IU/L}$).

2.3. US and MR Imaging Acquisition and Processing

US and MR studies were acquired using imaging protocols, as previously reported [21].

For US quantitative analysis, the right hepatic lobe diameter and portal vein diameter were measured, as well as liver stiffness being analyzed using shear-wave elastography (SWE). In particular, the right hepatic lobe diameter (mm) was obtained on the midclavicular plane using the upper margin of the liver as the uppermost edge under the dome of the diaphragm, while the lower margin was taken as the lowermost edge of the lobe [24]; the portal vein was visualized in its longitudinal axis, and the greatest anteroposterior diameter at the liver hilum was measured in millimeters. SWE evaluates tissue stiffness, expressed as

Young's modulus (kPa) [25–27]. The position of the regions of interest (ROIs) is selected by the operator in real-time grayscale mode imaging, allowing them to choose a homogeneous vessel-free area placed at least 1 cm below the liver capsule [28]. For spleen diameter measurement (mm), the longitudinal dimension in the coronal plane was obtained; of note, the longitudinal measurement was performed between the most superomedial and the most inferolateral points [24].

For MR quantitative analysis, liver and spleen volumes were measured using a semi-automatic method with OsiriX[®] version 3.3 software. An expert abdominal radiologist manually traced liver and spleen contours at different levels on T2-weighted images with the closed polygon selection tool under the ROI tool button; the Grow Region (2D/3D Segmentation) tool in the ROI dropdown menu made it possible to automatically outline the remaining boundaries. The automatic generated outlines were hand-adjusted with the closed polygon selection tool and the repulsor tool to optimize the ROIs. After selecting all of the ROIs within the series, OsiriX[®] automatically calculated the volume by multiplying the surface and slice thickness and then adding up individual slice volumes. OsiriX[®] also provided 3D images using the ROI volume tool (Figure 1) [29]. Furthermore, the portal vein diameter was measured in millimeters on the axial T2-weighted sequence at the liver hilum.

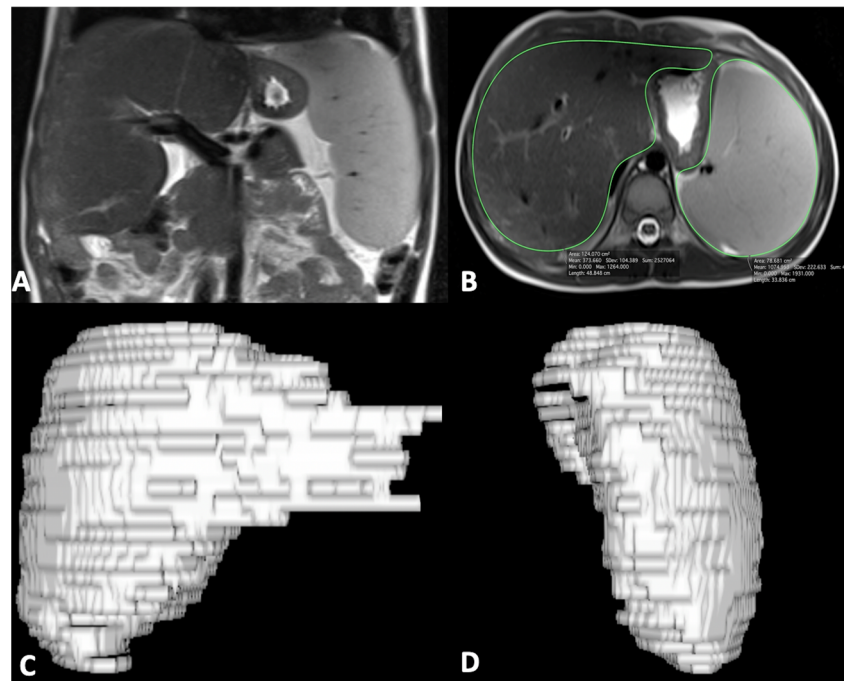


Figure 1. Coronal MR image shows liver and spleen enlargement (A); axial MR image shows ROI analysis of liver and spleen (B) to obtain 3D liver (C) and spleen (D) volume reconstruction images; of note, ROI analysis was performed on multiple sequential slices for completely including the liver and spleen.

2.4. Statistical Analysis

A preliminary statistical analysis was performed analyzing the data of each methodology, both by the laboratory tests and imaging, for giving an input to ML algorithms. In the light of the small sample size, a non-parametric Mann–Whitney test was performed to distinguish stable (group 1) from non-stable (group 2) patients, considering each quantitative variable associated with the three diagnostic methodologies under examination, namely laboratory, US, and MR parameters. A Wilcoxon signed-rank test was performed to compare paired data. Moreover, a chi-square test was performed to compare the evaluation metrics (accuracy, sensitivity, specificity) of the different methodologies, since the augmentation of the data made the dataset not paired; the first two best evaluation metrics among laboratory tests, US, and MR parameters were compared. For all statistical tests, a

two-tailed p -value of <0.05 was considered statistically significant. All the statistic tests were implemented using IBM SPSS Statistics (version 26).

2.5. Machine Learning: Tools and Algorithms

The data of all parameters were merged, and a selection method was used to understand the most important parameter among all diagnostic methodologies. Considering the small sample size and to make a fair comparison among the methodologies through machine learning analysis, an oversampling technique was performed, resulting in the generation of artificial data, namely the Synthetic Minority Oversampling Technique (SMOTE) proposed by Chawla et al., in order to double the amount of data [30,31]. This technique creates synthetic examples in the feature space from randomly selected pairs of real word–feature examples.

Because of the negative effect of irrelevant attributes on most ML schemes, it is common to precede learning with a feature selection stage that strives to eliminate all the redundant and irrelevant attributes for the classification and to identify the most informative features for the specific classification task. Dimensionality reduction yields a more compact and easily interpretable representation of the target concept, focusing the user's attention on the most relevant variables. A wrapper method was used for feature selection before the final classification procedure when the features of all the methodologies were merged [32]. To classify the prognosis of the patients (stable or non-stable), different ML classification techniques were tested to ensure the best performance. As a result, classification methods, including supervised learning with a random forest (RF), naive Bayes (NB) algorithm, k -nearest neighbor (kNN) algorithm, and support vector machine (SVM), were evaluated [33,34]. In particular, an RF is composed of a large number of decision trees, which are mainly used to correct the overfitting problem of decision trees, which is surely an added value in this study with a small sample size. In this technique, multiple decision trees, trained from different subsets of the same training set, are averaged, and overfitting is avoided by reducing the variance of the system. The training algorithm works by applying bagging and randomization to tree learners. In this paper, the RF was made up of 100 models, used the information gain ratio as a split criterion, and had a tree depth of 10. Differently, the NB algorithm is a probabilistic ML algorithm based on Bayes' theorem that calculates the probability of each class for a specified instance and then returns the class with the highest probability. This algorithm, requiring little data for training and little storage space, is suitable for the small size of the data sets at disposal. The kNN algorithm is an instance-based statistical method that works on the idea that the instances of a dataset are in proximity with other instances that have similar characteristics. In this classification approach, a test example is classified by observing the class label of its adjacent neighbors. The kNN algorithm finds out the k -nearest instances to the one to be classified and identifies its class on the basis of the most common class label. In this study, a k value was set equal to 3 and the Euclidean distance was used as the distance metric to identify the closest neighbors. Another instance-based algorithm is the SVM, which creates, in a binary classification, a hyperplane that separates data from two different classes. The largest possible distance is established between the separating hyperplane by maximizing the margin, thus creating the separation. The choice of kernel determines the separation boundary of the classes. The radial basis function (RBF) or Gaussian kernels are the most popular kernels used as default for any nonlinear model; polynomial kernels are also popular. An SVM with an RBF kernel was considered in this study.

The feature importance of the best subset of the features was computed according to the information gain for one of the best algorithms.

Leave-one-out cross-validation (LOOCV) was performed to evaluate the performance of the predictive models [35]. In LOOCV, every instance is in turn used to test the model induced from the other instances, ensuring the instance independence assumption, namely every prediction in LOOCV is independent of the other. This technique uses for each

train/test round the biggest-possible train set, thus reducing the errors and being the most reliable validation method.

Standard evaluation metrics such as accuracy, sensitivity, and specificity, as well as the area under the curve of the receiver operating characteristic (AUCROC) were used to evaluate the models' performance [36]. The AUCROC was computed by using as input a column with the real class and a second one with the probabilities that a record is classified as being from the selected class. The ML analysis was performed by means of the KNIME analytics platform (version 4.1.3) [10,37–39].

3. Results

3.1. Patient Population

The study population consisted of 24 patients (15 male; median age = 9.25 years, range = 5–25 years) according to inclusion and exclusion criteria. The median timing between the birth and KP surgical intervention was 67.5 days (range = 38–119 days). At initial evaluation, 15 patients had an ideal medical outcome, while the remaining 9 had a non-ideal medical outcome after KP. The median follow-up timing at initial evaluation after KP treatment was 9.7 years (range = 5–25 years) for all patients. At re-evaluation, after additional 4 years of long-term follow-up, 12 (50%) patients were stable (group 1) in their disease course, of which 9 had an ideal medical outcome and 3 a non-ideal medical outcome (Table 1); the other 12 (50%) patients had a non-stable (group 2) disease course, of which 6 patients changed from an ideal to a non-ideal medical outcome and 6 patients showed clinical disease progression (Table 2).

Table 1. Clinical results of stable patients (group 1).

#	Sex	Age (years)	Medical Status *	Laboratory Abnormalities °	CLD Complications
1	M	6	Ideal	-	-
2	M	13	Ideal	-	-
3	F	10	Ideal	-	-
4	M	9	Ideal	-	-
5	F	13	Ideal	-	-
6	M	6	Ideal	-	-
7	M	5	Ideal	-	-
8	F	9	Ideal	-	-
9	M	14	Ideal	-	-
10	M	11	Non-ideal	AST, ALT, WBC, PLT	Portal hypertension, cholangitis
11	M	9	Non-ideal	AST, ALT, GGT, WBC, PLT	Portal hypertension
12	M	25	Non-ideal	TB, PLT	Portal hypertension

* The medical status was established according to the criteria of Ng et al. [23] and Lee et al. [18]. ° Abnormal values out of the normal range. - = not present.

3.2. Descriptive Analysis

The results of each diagnostic parameters, either by laboratory tests or imaging (US and MR), are reported in Table 3; in particular, TB and DB, as laboratory parameters, and US stiffness, as the imaging parameter, were the only statistically significant parameters between groups 1 and 2. In detail, TB and DB were significantly higher in patients of group 2 compared to those of group 1, even though the corresponding values in group 2 were still in the normal ranges. However, in patients of group 2, the mean values of TB (1.23 ± 0.43 vs. 0.74 ± 0.25 ; $p = 0.005$) and DB (0.53 ± 0.18 vs. 0.29 ± 0.12 ; $p = 0.006$) were significantly increased at re-evaluation during the long-term follow-up; in particular, in the majority (75%) of patients of group 2, a significant increase in TB and DB values beyond

the high normal limit was observed. Finally, US liver stiffness by SWE was significantly higher in patients of group 2 compared with those of group 1 (Figure 2).

Table 2. Clinical results of non-stable patients (group 2).

#	Sex	Age (years)	Medical Status at Initial Evaluation *	Laboratory Abnormalities at Re-Evaluation	CLD Complications at Re-Evaluation	Long-Term Medical Outcome
1	M	13	Ideal	TB	-	Non-ideal
2	M	10	Ideal	TB	-	Non-ideal
3	M	12	Ideal	TB	Cholangitis	Non-ideal
4	M	5	Ideal	ALT, PLT	-	Non-ideal
5	F	14	Ideal	TB	Cholangitis	Non-ideal
6	F	6	Ideal	WBC	-	Non-ideal
7	M	6	Non-ideal ^a	TB, PLT	Portal hypertension	Clinical progression
8	M	5	Non-ideal ^b	WBC	-	Clinical progression
9	F	7	Non-ideal ^c	AST, ALT, WBC,	-	Clinical progression
10	F	10	Non-ideal ^d	TB	-	Clinical progression
11	M	7	Non-ideal ^e	WBC	-	Clinical progression
12	F	7	Non-ideal ^f	TB	-	Clinical progression

* The medical status was established according to the criteria of Ng et al. [23] and Lee et al. [18]. ^o Abnormal values out of the normal range. - = not present. ^a Increased values of AST and ALT associated with the presence of cholangitis; ^b decreased values of PLT associated with the presence of portal hypertension; ^c decreased values of PLT associated with the presence of cholangitis and portal hypertension; ^d abnormal values of PLT and INR associated with the presence of cholangitis, portal hypertension, and variceal bleeding; ^e abnormal values of AST, ALT, GGT, and PLT associated with the presence of portal hypertension; ^f abnormal values of AST, ALT, INR, albumin, WBC, and PLT associated with the presence of cholangitis and portal hypertension.

Table 3. Laboratory and imaging results in group 1 and group 2.

-	Parameter	Group 1 (Mean ± SD)	Group 2 (Mean ± SD)	p-Value
Laboratory	AST (IU/L)	31 ± 11	40 ± 25	0.443
	ALT (IU/L)	29 ± 21	33 ± 20	0.291
	GGT (IU/L)	23 ± 19	25 ± 22	0.887
	TB (mg/dL)	0.38 ± 0.34	0.74 ± 0.25	0.001
	DB (mg/dL)	0.13 ± 0.09	0.29 ± 0.12	0.001
	INR	1.06 ± 0.07	1.12 ± 0.11	0.198
	Albumin (g/dL)	4.74 ± 0.24	4.44 ± 0.50	0.114
	WBC (cells/mm ³)	6567 ± 2293	6122 ± 1873	0.551
	PLT (cells/mm ³)	242083 ± 115800	188667 ± 93292	0.378
US	Portal vein (mm)	9.75 ± 1.60	9.08 ± 2.11	0.932
	Liver diameter (mm)	129.17 ± 23.53	114.00 ± 21.56	0.078
	Spleen diameter (mm)	118.00 ± 23.83	124.92 ± 25.65	0.443
	Liver stiffness (kPa)	5.95 ± 1.28	10.47 ± 7.32	0.020
MR	Portal vein (mm)	9.92 ± 1.38	8.75 ± 2.05	0.198
	Liver volume (cm ³)	923.46 ± 250.47	823.97 ± 282.75	0.242
	Spleen volume (cm ³)	300.64 ± 199.82	356.17 ± 142.86	0.198

Note: the parameters statistically significant are marked in bold.

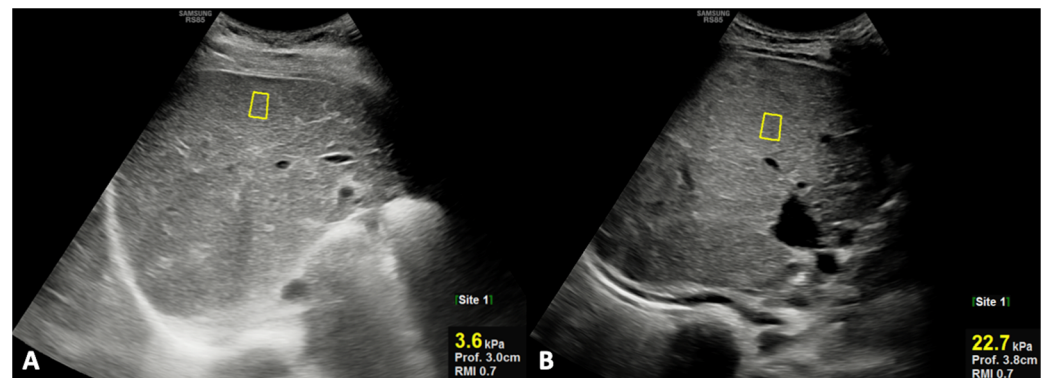


Figure 2. Ultrasound oblique scans under the right rib, showing shear-wave elastography measurements using ROI analysis in a patient of group 1 ((A) #5; Table 1—liver stiffness = 3.6 kPa) and in a patient of group 2 ((B) #11; Table 2—liver stiffness = 32.5 kPa).

3.3. Machine Learning

The result of SMOTE assessment increased the dataset from 24 to 48 subjects. Then, ML algorithms were implemented to classify the outcomes for all subjects using laboratory, US, and MR parameters by performing LOOCV (Tables 4–6). Table 3 contains the list of laboratory and imaging parameters that were given as input to the algorithms. For laboratory algorithms (Table 4), the RF was the best according to accuracy, sensitivity, and specificity values, even though the kNN algorithm achieved the highest AUCROC value. For US algorithms (Table 5), the RF was the best according to accuracy, sensitivity, and AUCROC, while NB and kNN algorithms obtained the highest specificity. For MR algorithms (Table 6), the kNN and SVM were the best according to accuracy, sensitivity, and specificity values, even though the kNN algorithm showed the highest AUCROC. The comparison of the mean performance between laboratory and imaging algorithms showed that the laboratory algorithms achieved the best results in terms of accuracy, sensitivity, and specificity values, as well as the AUCROC. The comparison between the first two best evaluation metrics (each best one is marked in bold for each methodology in Tables 4–6) among all the methodologies (laboratory, US, and MR) showed that the accuracy and sensitivity obtained through the RF applied on the laboratory data were greater than the others in a statistically significant way (p -value = 0.046 for both). When laboratory or imaging parameters were merged and analyzed as input to ML algorithms, using the wrapper technique as the feature selection method, the best algorithm was the NB algorithm using only laboratory parameters, such as TB and DB; however, the same result was obtained with the RF and kNN algorithms but using either laboratory or imaging parameters (Table 7). For the NB algorithm, the feature importance was also computed, thus determining that the TB contributed to the prediction with 56%, while DB contributed with 44%.

Table 4. Results using laboratory features after the SMOTE technique in predicting long-term medical outcomes.

Algorithms	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUCROC
RF	95.8	95.8	95.8	0.991
NB	72.9	62.5	83.3	0.866
kNN	93.8	91.7	95.8	0.997
SVM	89.6	87.5	91.7	0.896
Mean performance	88.0	84.4	91.7	0.937

Note: the best value for each evaluation metric is marked in bold.

Table 5. Results using US imaging features after the SMOTE technique in predicting long-term medical outcomes.

Algorithms	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUCROC
RF	79.2	79.2	79.2	0.868
NB	64.6	41.7	87.5	0.642
kNN	79.2	70.8	87.5	0.818
SVM	75.0	70.8	79.2	0.750
Mean performance	74.5	65.6	83.4	0.769

Note: the best value for each evaluation metric is marked in bold.

Table 6. Results using MR imaging features after the SMOTE technique in predicting long-term medical outcomes.

Algorithms	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUCROC
RF	79.2	79.2	79.2	0.878
NB	60.4	41.7	79.2	0.677
kNN	83.3	83.3	83.3	0.908
SVM	83.3	83.3	83.3	0.833
Mean performance	76.6	71.9	81.3	0.824

Note: the best value for each evaluation metric is marked in bold.

Table 7. Results using merged laboratory and imaging features after the SMOTE technique in predicting long-term medical outcomes.

Algorithms	Accuracy	Sensitivity	Specificity	AUCROC	Features Selected
RF	100	100	100	1	TB, US liver diameter, MR portal vein diameter
NB	100	100	100	1	TB, DB
kNN	100	100	100	1	TB, DB, WBC, US Stiffness, MR portal vein diameter
SVM	93.3	100	87.5	0.938	TB, INR

4. Discussion

In BA patients surviving with native liver after KP, the evaluation of the disease course and biliary cirrhosis occurrence is clinically relevant during follow-up [18,23]. For this purpose, clinical evaluation as well as laboratory tests and imaging studies are conventionally used. Imaging exams such as US and/or MR are able to depict liver and spleen anatomic conditions, providing a series of specific imaging parameters to assess the disease course [21]. Therefore, a wide spectrum of diagnostic parameters (clinical, laboratory, and imaging) is available in this setting, even though it is not well established how to use them and whether a complementary role may be hypothesized. In this study, the accuracy of several diagnostic quantitative parameters extracted from different methodologies, such as laboratory tests and imaging exams (US and MR), using ML algorithms was compared to predict the long-term medical outcome for native liver survivor patients with BA who have undergone KP. In detail, the patient population consisted of 24 patients, of which 50% were stable (group 1) in their disease course as an ideal ($n = 9$) or anon-ideal ($n = 3$) long-term medical outcome; conversely, the other 50% of the patients showed a long-term non-stable (group 2) disease course, since 6 patients changed from the ideal to the non-ideal medical status, while 6 patients had clinical disease progression. In this investigation, to predict the long-term medical outcome, laboratory parameters such as WBC and PLT counts, TB, DB, albumin, INR, ALT, AST, and GGT values were considered, as well as quantitative imaging parameters of liver and spleen conditions by US (right hepatic lobe diameter,

portal vein diameter, and liver stiffness) and MR (liver and spleen volumes and portal vein diameter) imaging modalities. In this setting, laboratory data reflect mainly liver function, while imaging parameters are an expression of liver and spleen morphological changes, the liver parenchyma structure using the assessment of liver stiffness by US, and portal hypertension by measuring the portal vein diameter by both imaging techniques. ML algorithms with different operating principles were used to obtain a wider range of investigation. The overall results of the ML analysis showed that TB and DB as laboratory tests and US liver stiffness as the imaging parameter were the only significant parameters that were able to distinguish stable from non-stable patients in predicting the long-term medical outcome. These findings are reasonable since they reflect liver conditions, either directly in terms of the liver structure by US stiffness or indirectly by TB and DB reflecting liver function. These observations are concordant and confirm previous experiences in which a predictive role of serum bilirubin levels and US liver stiffness has been suggested in patients with BA treated with KP during early and long-term follow-up [19,40–44]. In particular, among the used ML algorithms, the RF algorithm was the best either for laboratory or for US parameters, while SVM and kNN algorithms were the best according to MR parameters. However, the evaluation of the mean performance of laboratory and imaging algorithms showed that laboratory algorithms achieved the best results in terms of accuracy, sensitivity, and specificity values, as well as the AUCROC. Furthermore, when all the diagnostic parameters, either by laboratory tests or by imaging, were merged and analyzed as input to ML algorithms, the best algorithm was the NB algorithm using only TB and DB, even though the same result was obtained also with RF and kNN algorithms using either laboratory or imaging parameters. Of course, the high evaluation metrics achieved could make a reader think of overfitting, since a high number of computations on a small sample of data through simple cross-validation provide an optimistic estimation of the model, as reported by Tsamardinos et al. [45]. However, it is worth underlining that the best results were obtained by using the combination of LOOCV and the RF, both of which are used to reduce the chance of overfitting. Moreover, it should be emphasized that the purpose of the article was not to obtain a perfect model, since the dataset had obtained an injection of artificial data, but to understand the weight and importance of the parameters extracted from US, MRI, and laboratory tests in predicting the long-term outcome for native liver survivor patients with BA after KP. Indeed, ML has already been used to compare different clinical methodologies to predict an outcome (both diagnostic and prognostic) in cardiology or choose the best resolution for ultrasound [9,46,47].

Thus, this preliminary ML evaluation confirms that laboratory tests, specifically TB and DB, represent powerful parameters to predict the long-term medical outcome in native liver survivor patients with BA after KP, supporting previous observations that already suggested a main role of serum bilirubin levels for this purpose [19]. These preliminary results and those of previous investigations may have significant advantages in terms of clinical patient management and cost-effectiveness, since TB and DB plasma measurements as laboratory tests are easily performed, widely available, and not expensive [19,42]. However, even though the values of TB and DB were able to predict the long-term medical outcome, they were still in the normal range but tended toward the upper limit; of note, this trend was confirmed by increased values of TB and DB beyond the high normal limit at re-evaluation in the majority of patients with non-stable disease.

To date, ML methods have been applied in clinical research with applications in several medical fields, of which many are in pediatric diagnostic imaging [48–50]. In particular, the ML methodology has been applied to assess skeletal maturity on hand X-rays [51], to diagnose and classify acute appendicitis using laboratory tests and US [52], to identify MR biomarkers of the autistic spectrum [53], and to evaluate CLD using clinical data and MR [54]. Furthermore, recent studies have suggested a role of ML methodologies also in patients with BA, focusing on disease diagnosis. In detail, Hoshino et al., using an ML algorithm, realized an iPhone application (Baby-Poop) able to capture subtle differences in stool color that may be undetectable by a layperson to get early diagnosis of BA [55]. A

similar ML application with the same purpose was made by Angelico et al., who created PopòApp [56]. Moreover, Zhou et al. developed an ensembled deep learning model to facilitate the diagnosis of BA for non-expert radiologists using DB values and US images as well as videos of the gallbladder [57]. In this setting, this pilot experience is the first that reports an ML evaluation using laboratory and imaging parameters with long-term predictive purposes in patients with BA after KP, supporting the main role of laboratory tests in the follow-up of such patients. A future development could be the use of deep learning algorithms on the images to further test their feasibility to predict the outcome.

Some limitations of this study should be addressed. Mainly, the small sample size and the retrospective type of the investigation might be not optimal, but the low incidence of BA, a rare pediatric disease, should be considered; therefore, additional experiences in a larger patient population are required. The data used in ML analysis to establish the long-term medical outcome consisted of laboratory and quantitative imaging parameters as continuous variables requested by ML algorithms; therefore, the presence or absence of CLD medical complications was not included in the analysis for the lack of continuous quantification; similarly, patients with asplenia or poly-splenia, possible findings in children with BA, may be not included. Moreover, technical ML limitations were also present, particularly due to the implementation of SMOTE for augmenting data; nevertheless, predicting the outcome was not the main purpose of the research, since the aim was to compare the accuracy of several diagnostic parameters extracted from different methodologies. Therefore, the use of SMOTE, which was used to augment the dataset with artificial data, as already done in a previous study, rather than to balance a minority class, as is usually employed [58], might have a limited impact on the analysis; in comparison with traditional logistic regression, ML has the advantage of not requiring the assessment of assumptions to be performed, such as the detection of outliers or a strict limit between subjects and variables. Moreover, ML algorithms have demonstrated empirically their powerfulness in several fields. The main disadvantage of ML algorithms is the black-box style, since the input and output of the algorithms are known but a numerical model is not provided; nevertheless, ML algorithms may be used as clinical support decision-making systems since they provide users with a probability for each subject of being part of a fixed class.

In conclusion, the results of this preliminary ML investigational study of native liver survivor patients with BA who have undergone KP, integrating laboratory and imaging quantitative diagnostic data, showed that TB and DB represent the fundamental parameters to predict the long-term medical outcome after treatment, confirming the results of previous studies that demonstrated a main predictive role of serum bilirubin levels in such patients during early follow-up. In particular, the values of TB and DB may be within the normal range but with a slight increase; therefore, clinicians should be alert when the values of these laboratory parameters show subtle changes. Furthermore, US liver stiffness, reflecting liver parenchyma changes, is the best imaging parameter for this purpose.

Author Contributions: Conceptualization, M.C. and S.M.; methodology, M.P.; software, C.R.; formal analysis, L.D. and G.C.; investigation, F.D.D.; data curation, G.D.P.; writing—original draft preparation, M.C. and C.R.; writing—review and editing, V.R. and R.I.; supervision, A.B. and S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: The study was conducted according to the guidelines of the Declaration of Helsinki. Authors declare that in view of the retrospective nature of the study, since all the procedures being performed were part of the routine care, all the collected data were anonymized, and no information is linked or linkable to a specific person, no ethical approval and no consent declarations were required.

Data Availability Statement: Data are not available due to privacy policy.

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

References

- Maity, N.G.; Das, S. Machine learning for improved diagnosis and prognosis in healthcare. In Proceedings of the 2017 IEEE Aerospace Conference, Big Sky, MT, USA, 4–11 March 2017; pp. 1–9.
- Cerri, R.; Da Silva, R.R.O.; De Carvalho, A.C.P.L.F. Comparing methods for multilabel classification of proteins using machine learning techniques. *Lect. Notes Comput. Sci.* **2009**, *5676 LNBI*, 109–120.
- Daldrup-Link, H. Artificial intelligence applications for pediatric oncology imaging. *Pediatr. Radiol.* **2019**, *49*, 1384–1390. [[CrossRef](#)] [[PubMed](#)]
- Booz, C.; Yel, I.; Wichmann, J.L.; Boettger, S.; Al Kamali, A.; Albrecht, M.H.; Martin, S.S.; Lenga, L.; Huizinga, N.A.; D’Angelo, T.; et al. Artificial intelligence in bone age assessment: Accuracy and efficiency of a novel fully automated algorithm compared to the Greulich-Pyle method. *Eur. Radiol. Exp.* **2020**, *4*, 6. [[CrossRef](#)]
- Scruggs, B.A.; Paulchan, R.V.; Kalpathy-Cramer, J.; Chiang, M.F.; Peter Campbell, J. Artificial intelligence in retinopathy of prematurity diagnosis. *Transl. Vis. Sci. Technol.* **2020**, *9*, 1–10. [[CrossRef](#)] [[PubMed](#)]
- Ricciardi, C.; Cantoni, V.; Green, R.; Improta, G.; Cesarelli, M. Is It Possible to Predict Cardiac Death? In Proceedings of the Mediterranean Conference on Medical and Biological Engineering and Computing, São Francisco, Portugal, 26–28 September 2019; Springer: Cham, Switzerland, 2019; pp. 847–854.
- Ricciardi, C.; Valente, A.S.; Edmund, K.; Cantoni, V.; Green, R.; Fiorillo, A.; Picone, I.; Santini, S.; Cesarelli, M. Linear discriminant analysis and principal component analysis to predict coronary artery disease. *Health Inform. J.* **2020**, *26*, 2181–2192. [[CrossRef](#)] [[PubMed](#)]
- Ricciardi, C.; Amboni, M.; De Santis, C.; Ricciardelli, G.; Improta, G.; Iuppariello, L.; D’Addio, G.; Barone, P.; Cesarelli, M. Classifying Different Stages of Parkinson’s Disease Through Random Forests. In Proceedings of the Mediterranean Conference on Medical and Biological Engineering and Computing, São Francisco, Portugal, 26–28 September 2019; Springer: Cham, Switzerland, 2019; pp. 1155–1162.
- Cantoni, V.; Green, R.; Ricciardi, C.; Assante, R.; Zampella, E.; Nappi, C.; Gaudieri, V.; Mannarino, T.; Genova, A.; De Simini, G.; et al. A machine learning-based approach to directly compare the diagnostic accuracy of myocardial perfusion imaging by conventional and cadmium-zinc telluride SPECT. *J. Nucl. Cardiol.* **2020**. [[CrossRef](#)] [[PubMed](#)]
- Scrutinio, D.; Ricciardi, C.; Donisi, L.; Losavio, E.; Battista, P.; Guida, P.; Cesarelli, M.; Pagano, G.; D’Addio, G. Machine learning to predict mortality after rehabilitation among patients with severe stroke. *Sci. Rep.* **2020**, *10*, 20127. [[CrossRef](#)] [[PubMed](#)]
- Improta, G.; Ricciardi, C.; Amato, F.; D’Addio, G.; Cesarelli, M.; Romano, M. Efficacy of Machine Learning in Predicting the Kind of Delivery by Cardiotocography. In Proceedings of the Mediterranean Conference on Medical and Biological Engineering and Computing, São Francisco, Portugal, 26–28 September 2019; Springer: Cham, Switzerland, 2019; pp. 793–799.
- Hartley, J.L.; Davenport, M.; Kelly, D.A. Biliary atresia. *Lancet* **2009**, *374*, 1704–1713. [[CrossRef](#)]
- Neto, B.; Borges-Dias, M.; Trindade, E.; Estevão-Costa, J.; Campos, J.M. Biliary Atresia-Clinical Series. *GE Port. J. Gastroenterol.* **2018**, *25*, 68–73. [[CrossRef](#)]
- Govindarajan, K.K. Biliary atresia: Where do we stand now? *World J. Hepatol.* **2016**, *8*, 1593. [[CrossRef](#)]
- Feldman, A.G.; Mack, C.L. Biliary Atresia: Clinical Lessons Learned. *J. Pediatr. Gastroenterol. Nutr.* **2015**, *61*, 167–175. [[CrossRef](#)] [[PubMed](#)]
- Baumann, U.; Ure, B. Biliary atresia. *Clin. Res. Hepatol. Gastroenterol.* **2012**, *36*, 257–259. [[CrossRef](#)] [[PubMed](#)]
- Nio, M.; Wada, M.; Sasaki, H.; Tanaka, H.; Okamura, A. Risk factors affecting late-presenting liver failure in adult patients with biliary atresia. *J. Pediatr. Surg.* **2012**, *47*, 2179–2183. [[CrossRef](#)] [[PubMed](#)]
- Lee, W.S.; Ong, S.Y.; Foo, H.W.; Wong, S.Y.; Kong, C.X.; Seah, R.B.; Ng, R.T. Chronic liver disease is universal in children with biliary atresia living with native liver. *World J. Gastroenterol.* **2017**, *23*, 7776–7784. [[CrossRef](#)]
- Jeon, T.Y.; Yoo, S.-Y.; Kim, J.H.; Eo, H.; Lee, S.-K. Serial ultrasound findings associated with early liver transplantation after Kasai portoenterostomy in biliary atresia. *Clin. Radiol.* **2013**, *68*, 588–594. [[CrossRef](#)]
- Takahashi, A.; Hatakeyama, S.I.; Kuroiwa, M.; Suzuki, N.; Toki, F.; Suzuki, M.; Suehiro, T.; Shimura, T.; Kuwano, H. Time-course changes in the liver of biliary atresia patients on magnetic resonance imaging. *Pediatr. Int.* **2009**, *51*, 66–70. [[CrossRef](#)]
- Caruso, M.; Cuocolo, R.; Di Dato, F.; Mollica, C.; Vallone, G.; Romeo, V.; Petretta, M.; Liuzzi, R.; Mainenti, P.P.; Iorio, R.; et al. Ultrasound, shear-wave elastography, and magnetic resonance imaging in native liver survivor patients with biliary atresia after Kasai portoenterostomy: Correlation with medical outcome after treatment. *Acta Radiol.* **2020**, *61*, 1300–1308. [[CrossRef](#)]
- Caruso, M.; Di Dato, F.; Mollica, C.; Vallone, G.; Romeo, V.; Liuzzi, R.; Mainenti, P.P.; Petretta, M.; Iorio, R.; Brunetti, A.; et al. Imaging prediction with ultrasound and MRI of long-term medical outcome in native liver survivor patients with biliary atresia after kasai portoenterostomy: A pilot study. *Abdom. Radiol.* **2021**, *46*, 2595–2603. [[CrossRef](#)]
- Ng, V.L.; Haber, B.H.; Magee, J.C.; Miethke, A.; Murray, K.F.; Michail, S.; Karpen, S.J.; Kerkar, N.; Molleston, J.P.; Romero, R.; et al. Medical Status of 219 Children with Biliary Atresia Surviving Long-Term with Their Native Livers: Results from a North American Multicenter Consortium. *J. Pediatr.* **2014**, *165*, 539–546.e2. [[CrossRef](#)]
- Konuş, O.L.; Ozdemir, A.; Akkaya, A.; Erbaş, G.; Celik, H.; Işik, S. Normal liver, spleen, and kidney dimensions in neonates, infants, and children: Evaluation with sonography. *Am. J. Roentgenol.* **1998**, *171*, 1693–1698. [[CrossRef](#)]

25. Serai, S.D.; Trout, A.T.; Sirlin, C.B. Elastography to assess the stage of liver fibrosis in children: Concepts, opportunities, and challenges. *Clin. Liver Dis.* **2017**, *9*, 5–10. [[CrossRef](#)]
26. Dillman, J.R.; Heider, A.; Bilhartz, J.L.; Smith, E.A.; Keshavarzi, N.; Rubin, J.M.; Lopez, M.J. Ultrasound shear wave speed measurements correlate with liver fibrosis in children. *Pediatr. Radiol.* **2015**, *45*, 1480–1488. [[CrossRef](#)]
27. Lurie, Y.; Webb, M.; Cytter-Kuint, R.; Shteingart, S.; Lederkremer, G.Z. Non-invasive diagnosis of liver fibrosis and cirrhosis. *World J. Gastroenterol.* **2015**, *21*, 11567–11583. [[CrossRef](#)]
28. Tang, A.; Cloutier, G.; Szeverenyi, N.M.; Sirlin, C.B. Ultrasound Elastography and MR Elastography for Assessing Liver Fibrosis: Part 2, Diagnostic Performance, Confounders, and Future Directions. *Am. J. Roentgenol.* **2015**, *205*, 33–40. [[CrossRef](#)] [[PubMed](#)]
29. Van Der Vorst, J.R.; Van Dam, R.M.; Van Stiphout, R.S.A.; Van Den Broek, M.A.; Hollander, I.H.; Kessels, A.G.H.; Dejong, C.H.C. Virtual Liver Resection and Volumetric Analysis of the Future Liver Remnant using Open Source Image Processing Software. *World J. Surg.* **2010**, *34*, 2426–2433. [[CrossRef](#)] [[PubMed](#)]
30. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
31. Wong, S.C.; Gatt, A.; Stamatescu, V.; McDonnell, M.D. Understanding Data Augmentation for Classification: When to Warp? In Proceedings of the 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, QLD, Australia, 30 November–2 December 2016.
32. Witten, I.H.; Hall, M.A. *Practical Machine Learning*, 3rd ed.; Packt Publishing Ltd.: Birmingham, UK, 2016.
33. Al-Aidaros, K.M.; Abu Bakar, A.; Othman, Z. Naïve Bayes variants in classification learning. In Proceedings of the 2010 International Conference on Information Retrieval & Knowledge Management (CAMP), Shah Alam, Malaysia, 17–18 March 2010; pp. 276–281.
34. Vapnik, V.N. An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **1999**, *10*, 988–999. [[CrossRef](#)] [[PubMed](#)]
35. Wong, T.T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognit.* **2015**, *48*, 2839–2846. [[CrossRef](#)]
36. Hossin, M.; Sulaiman, M.N. A Review on Evaluation Metrics for Data Classification Evaluations. *Int. J. Data Min. Knowl. Manag. Process.* **2015**, *5*, 01–11. [[CrossRef](#)]
37. Tougui, I.; Jilbab, A.; El Mhamdi, J. Heart disease classification using data mining tools and machine learning techniques. *Health Technol.* **2020**, *10*, 1137–1144. [[CrossRef](#)]
38. Ricciardi, C.; Donisi, L.; Cesarelli, G.; Pagano, G.; Coccia, A.; D’addio, G. Feasibility of Machine Learning applied to Poincaré Plot Analysis on Patients with CHF. In Proceedings of the 2020 11th Conference of the European Study Group on Cardiovascular Oscillations (ESGCO), Pisa, Italy, 27–29 April 2020; pp. 16–17.
39. Donisi, L.; Ricciardi, C.; Cesarelli, G.; Pagano, G.; Amitrano, F.; D’addio, G. Machine Learning applied on Poincaré Analysis to discriminate different cardiac issues. In Proceedings of the 2020 11th Conference of the European Study Group on Cardiovascular Oscillations (ESGCO), Pisa, Italy, 27–29 April 2020; pp. 17–18.
40. Hahn, S.M.; Kim, S.; Park, K.I.; Han, S.J.; Koh, H. Clinical benefit of liver stiffness measurement at 3 months after Kasai hepatoportoenterostomy to predict the liver related events in biliary atresia. *PLoS ONE* **2013**, *8*, e80652. [[CrossRef](#)]
41. Jain, V.; Burford, C.; Alexander, E.C.; Sutton, H.; Dhawan, A.; Joshi, D.; Davenport, M.; Heaton, N.; Hadzic, N.; Samyn, M. Prognostic markers at adolescence in patients requiring liver transplantation for biliary atresia in adulthood. *J. Hepatol.* **2019**, *71*, 71–77. [[CrossRef](#)] [[PubMed](#)]
42. Hanquinet, S.; Courvoisier, D.S.; Rougemont, A.L.; Wildhaber, B.E.; Merlini, L.; McLin, V.A.; Anooshiravani, M. Acoustic radiation force impulse sonography in assessing children with biliary atresia for liver transplantation. *Pediatr. Radiol.* **2016**, *46*, 1011–1016. [[CrossRef](#)] [[PubMed](#)]
43. Yan, H.; Du, L.; Zhou, J.; Li, Y.; Lei, J.; Liu, J.; Luo, Y. Diagnostic performance and prognostic value of elastography in patients with biliary atresia and after hepatic portoenterostomy: Protocol for a systematic review and meta-analysis. *BMJ Open* **2021**, *11*, e042129. [[CrossRef](#)] [[PubMed](#)]
44. Liu, Y.; Peng, C.; Wang, K.; Wu, D.; Yan, J.; Tu, W.; Chen, Y. The utility of shear wave elastography and serum biomarkers for diagnosing biliary atresia and predicting clinical outcomes. *Eur. J. Pediatr.* **2021**. [[CrossRef](#)]
45. Tsamardinos, I.; Rakhshani, A.; Lagani, V. Performance-Estimation Properties of Cross-Validation-Based Protocols with Simultaneous Hyper-Parameter Optimization. *Int. J. Artif. Intell. Tools* **2015**, *24*, 1540023. [[CrossRef](#)]
46. Mannarino, T.; Assante, R.; Ricciardi, C.; Zampella, E.; Nappi, C.; Gaudieri, V.; Mainolfi, C.G.; Di Vaia, E.; Petretta, M.; Cesarelli, M.; et al. Head-to-head comparison of diagnostic accuracy of stress-only myocardial perfusion imaging with conventional and cadmium-zinc telluride single-photon emission computed tomography in women with suspected coronary artery disease. *J. Nucl. Cardiol.* **2021**, *28*, 888–897. [[CrossRef](#)]
47. Ricciardi, C.; Cuocolo, R.; Verde, F.; Improta, G.; Stanzione, A.; Romeo, V.; Maurea, S.; D’Armiento, M.; Sarno, L.; Guida, M.; et al. Resolution Resampling of Ultrasound Images in Placenta Previa Patients: Influence on Radiomics Data Reliability and Usefulness for Machine Learning. In Proceedings of the European Medical and Biological Engineering Conference, Portorož, Slovenia, 29 November–3 December 2020; Springer: Cham, Switzerland; 2020; pp. 1011–1018.
48. Davendralingam, N.; Sebire, N.J.; Arthurs, O.J.; Shelmerdine, S.C. Artificial intelligence in paediatric radiology: Future opportunities. *Br. J. Radiol.* **2021**, *94*, 20200975. [[CrossRef](#)]

49. Rajpurkar, P.; Irvin, J.; Zhu, K.; Yang, B.; Mehta, H.; Duan, T.; Ding, D.; Bagul, A.; Ball, R.L.; Langlotz, C.; et al. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv* **2017**, arXiv:1711.05225.
50. Cherukuri, V.; Ssenyonga, P.; Warf, B.C.; Kulkarni, A.V.; Monga, V.; Schiff, S.J. Learning Based Segmentation of CT Brain Images: Application to Postoperative Hydrocephalic Scans. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 1871–1884.
51. Larson, D.B.; Chen, M.C.; Lungren, M.P.; Halabi, S.S.; Stence, N.V.; Langlotz, C.P. Performance of a Deep-learning neural network Model in assessing skeletal Maturity on Pediatric hand radiographs 1 PEDIATRIC IMAGING: Neural Network to Assess Skeletal Maturity on Pediatric Hand Radiographs Larson et al. Materials and Methods Data Acquisit. *Radiology* **2018**, *287*, 313–322. [[CrossRef](#)] [[PubMed](#)]
52. Reismann, J.; Romualdi, A.; Kiss, N.; Minderjahn, M.I.; Kallarackal, J.; Schad, M.; Reismann, M. Diagnosis and classification of pediatric acute appendicitis by artificial intelligence methods: An investigator-independent approach. *PLoS ONE* **2019**, *14*, e0222030. [[CrossRef](#)] [[PubMed](#)]
53. Chen, T.; Chen, Y.; Yuan, M.; Gerstein, M.; Li, T.; Liang, H.; Froehlich, T.; Lu, L. The development of a practical artificial intelligence tool for diagnosing and evaluating autism spectrum disorder: Multicenter study. *JMIR Med. Inform.* **2020**, *8*, e15767. [[CrossRef](#)] [[PubMed](#)]
54. He, L.; Li, H.; Dudley, J.A.; Maloney, T.C.; Brady, S.L.; Somasundaram, E.; Trout, A.T.; Dillman, J.R. Machine learning prediction of liver stiffness using clinical and T2-Weighted MRI radiomic data. *Am. J. Roentgenol.* **2019**, *213*, 592–601. [[CrossRef](#)] [[PubMed](#)]
55. Hoshino, E.; Hayashi, K.; Suzuki, M.; Obatake, M.; Urayama, K.Y.; Nakano, S.; Taura, Y.; Nio, M.; Takahashi, O. An iPhone application using a novel stool color detection algorithm for biliary atresia screening. *Pediatr. Surg. Int.* **2017**, *33*, 1115–1121. [[CrossRef](#)] [[PubMed](#)]
56. Angelico, R.; Liccardo, D.; Paoletti, M.; Pietrobattista, A.; Basso, M.S.; Mosca, A.; Safarikia, S.; Grimaldi, C.; Saffioti, M.C.; Candusso, M.; et al. A novel mobile phone application for infant stool color recognition: An easy and effective tool to identify acholic stools in newborns. *J. Med. Screen.* **2020**, *28*, 230–237. [[CrossRef](#)] [[PubMed](#)]
57. Zhou, W.; Yang, Y.; Yu, C.; Liu, J.; Duan, X.; Weng, Z.; Chen, D.; Liang, Q.; Qing, F.; Zhou, J.; et al. An ensemble deep learning model outperforms human experts in diagnosing biliary atresia from sonographic gallbladder images. *medRxiv* **2020**, *12*, 1259.
58. Stanzione, A.; Ricciardi, C.; Cuocolo, R.; Romeo, V.; Petrone, J.; Sarnataro, M.; Mainenti, P.P.; Improta, G.; De Rosa, F.; Insabato, L.; et al. MRI Radiomics for the Prediction of Fuhrman Grade in Clear Cell Renal Cell Carcinoma: A Machine Learning Exploratory Study. *J. Digit. Imaging* **2020**, *33*, 879–887. [[CrossRef](#)] [[PubMed](#)]