




## Article

# Classification of Distal Growth Plate Ossification States of the Radius Bone Using a Dedicated Ultrasound Device and Machine Learning Techniques for Bone Age Assessments

Lukas Brausch <sup>1</sup>, Ruth Dirksen <sup>2</sup>, Christoph Risser <sup>1</sup>, Martin Schwab <sup>3</sup>, Carole Stolz <sup>4</sup>, Steffen Tretbar <sup>1</sup>,  
Tilman Rohrer <sup>2</sup> and Holger Hewener <sup>1,\*</sup>

<sup>1</sup> Ultrasound Department, Fraunhofer Institute for Biomedical Engineering, Joseph-von-Fraunhofer-Weg 1, 66280 Sulzbach, Germany; lukas.brausch@ibmt.fraunhofer.de (L.B.); christoph.risser@ibmt.fraunhofer.de (C.R.); steffen.tretbar@ibmt.fraunhofer.de (S.T.)

<sup>2</sup> Clinic for General Paediatrics and Neonatology, Saarland University Medical Center, Kirrberger Straße 100, 66424 Homburg, Germany; ruth.dirksen@uks.eu (R.D.); tilman.rohrer@uks.eu (T.R.)

<sup>3</sup> CEMEC Intelligente Mechanik GmbH, Gewerbepark Hügelmühle 30, 91174 Spalt, Germany; martin.schwab@cemec.de

<sup>4</sup> Hope for Freedom e.V., Friedenstraße 60, 63743 Aschaffenburg, Germany; c.stolz@hopeforfreedom.de

\* Correspondence: holger.hewener@ibmt.fraunhofer.de

**Featured Application:** This work features a mobile ultrasound measurement system that uses machine learning to classify finished bone growth of girls and women. It relies on 1D ultrasound radio frequency signals and can be a radiation-free alternative to certain bone age assessments using X-ray imaging.



**Citation:** Brausch, L.; Dirksen, R.; Risser, C.; Schwab, M.; Stolz, C.; Tretbar, S.; Rohrer, T.; Hewener, H. Classification of Distal Growth Plate Ossification States of the Radius Bone Using a Dedicated Ultrasound Device and Machine Learning Techniques for Bone Age Assessments. *Appl. Sci.* **2022**, *12*, 3361. <https://doi.org/10.3390/app12073361>

Academic Editor: Kelvin K.L. Wong

Received: 25 February 2022

Accepted: 23 March 2022

Published: 25 March 2022

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

**Abstract:** X-ray imaging, based on ionizing radiation, can be used to determine bone age by examining distal growth plate fusion in the ulna and radius bones. Legal age determination approaches based on ultrasound signals exist but are unsuitable to reliably determine bone age. We present a low-cost, mobile system that uses one-dimensional ultrasound radio frequency signals to obtain a robust binary classifier enabling the determination of bone age from data of girls and women aged 9 to 24 years. These data were acquired as part of a clinical study conducted with 148 subjects. Our system detects the presence or absence of the epiphyseal plate by moving ultrasound array transducers along the forearm, measuring reflection and transmission signals. Even though classical digital signal processing methods did not achieve a robust classifier, we achieved an  $F_1$  score of approximately 87% for binary classification of completed bone growth with machine learning approaches, such as the gradient boosting machine method CatBoost. We demonstrate that our ultrasound system can classify the fusion of the distal growth plate of the radius bone and the completion of bone growth with high accuracy. We propose a non-ionizing alternative to established X-ray imaging methods for this purpose.

**Keywords:** bone age; growth plate fusion; mobile ultrasound; machine learning

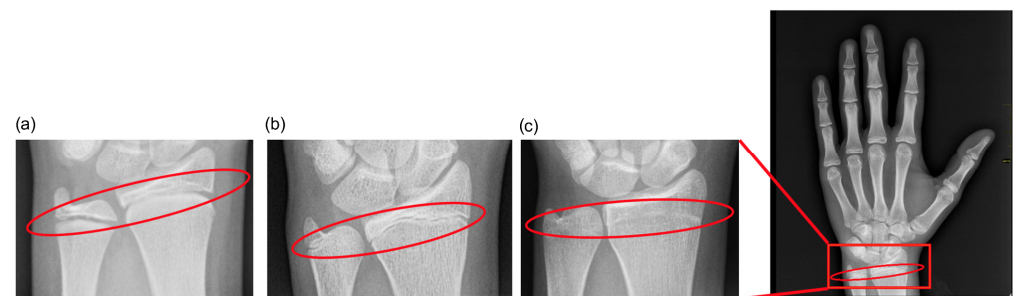


**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Bone age correlates with the skeletal maturity of an individual and is often used by pediatricians and endocrinologists to diagnose diseases or predict the final adult height [1,2]. Bone age is also a helpful metric in forensic medicine to estimate the chronological age and has applications in the anthropological analysis of human skeletal remains [3]. Other potential but controversial uses for bone age assessments include age determination of asylum seekers or individuals within the context of human trafficking [4]. Different approaches to determine bone age based on X-ray imaging, magnetic resonance imaging (MRI), or ultrasonography exist, but there is no standard method as each method has its own shortcomings [1]. Bone

age assessments often rely on visualizations of hand and wrist, teeth, clavicle, iliac bone, or femoral head. In children, radiographs of hand and wrist have been described as the most commonly used methods [1]. Each bone begins as a cartilaginous structure, developing and ossifying throughout the period of physical growth, ultimately forming the adult bone structure. At the end of growth and development, the epiphyseal centers ossify, and the growth plates fuse. Over time, this fusion is consolidated by a further increase in bone density at the fused plate. In the forearm, there are two main bones: ulna and radius. In a study based on data from 80 healthy males and females, the secondary ossification centers of the radius and ulna closed completely for females between 17 and 18 years and for males between 18 and 19 years [5]. Ossification of the various bone centers finishes at different ages. They become radio-opaque on radiographs and change significantly at different skeletal development stages. This staging (see Figure 1) is the basis for various bone age assessment atlases.



**Figure 1.** Bone age radiographs showing different degrees of growth plate ossification of the distal radius and ulna in girls aged 12 years (a), 15 years (b), and 18 years (c).

Influential bone age assessment publications include the atlases by Greulich and Pyle (GP) [6], Tanner et al. (TW3) [7], and Gilsanz and Ratib (GR) [8]. The GP atlas was published in 1959 (6) and is considered the most commonly used atlas for bone age measurement in the Western world [1]. The TW3 atlas [7] has been described as being more accurate and reproducible than the GP atlas in white populations [1,2], while the GR atlas has been found to be much more precise and of better quality than the GP atlas [1]. A 2009 study found that, due to the risk of age overestimation in females, the GR atlas “only seems to be suitable for forensic age diagnostics to a limited extent” [9]. A review and meta-analysis published in 2019 concludes that the ethnicity or origin of a child can influence the applicability of the GP atlas. It was found to be imprecise and should be used with caution in Asian, certain Hispanic, and African populations, particularly when assessing age for forensic or legal purposes [10]. Endocrine or nutritional factors, certain medications, non-endocrine chronic diseases, or certain medical syndromes can influence bone development [2]. Additionally, differences in bone age development exist for different ethnicities and genders [9,10].

Aside from methods based on radiographs, bone age assessment methods based on MRI of the left hand have also been tested [11], and the predictable ossification pattern of the patella, tibia, fibula, and femur on knee MRI has also been found to enable accurate bone age assessments [12]. However, the acquisition of radiographs requires ionizing radiation, and MRI is an expensive and less accessible technology. To address the shortcomings of MRI or radiographs, the usage of non-invasive quantitative ultrasound (QUS) has been proposed. For example, in the related field of osteoporosis diagnosis, ultrasound (US) measurements of the bones are used. These systems work within a typical US frequency range from 0.2 to 1.0 MHz to measure bone mineral content of the calcaneus and allow the determination of bone loss. The onset and progress of osteoporosis can be classified using conventional signal processing [13] or, more recently, using multi-channel convolutional neural networks based on 1D US radio frequent signals [14].

First attempts in applying QUS to assess bone age by performing B-mode imaging of the femoral head articular cartilage (FHC) were proposed in 1995 [15]. However, clinical

use of methods based on signals stemming from the FHC was strongly discouraged in a publication stating that the sensitivity of these methods was too low [16]. A study evaluated the QUS system BonAge™ on 37 children aged 5 to 15 years and stated it was accurate at assessing the bone age [17]. This system computes the bone age by combining US propagation speeds and the distance between emitter and receiver and comparing those parameters to a database containing reference data based on sex and ethnicity [18]. In two studies, the BonAge™ system was evaluated on 152 and 65 subjects, respectively, and demonstrated sufficient accuracy of bone ages [18,19]. However, a follow-up paper found the potential of US for assessing bone properties to be largely unexploited at the time [20], while another study concluded that US assessment should not yet be considered a valid replacement for radiographic bone age determination [21]. In the latter study, the authors report multiple reasons not to consider US as a valid replacement for radiographs. They found a higher variability of bone age estimates, compromised readings due to diseases that increase the thickness of cartilage, the need for an operator to perform a time-consuming examination, and greater variability in correlation studies [21].

A study examined the QUS device SonicBone™ and found it to be safe, convenient, and non-painful, with highly reproducible results [22]. The same study stated that the BonAge™ system was not commonly used due to its lack of consistency and confirmatory data in large groups. Furthermore, this study claimed that the SonicBone™ device takes most of the drawbacks of both the GP atlas and previously suggested QUS-based devices into consideration [22]. A successor to the SonicBone™ device is the BAUS™ system, which also performs assessments of bone age using US. Its results are highly reproducible and comparable to bone age assessments based on radiographs [23]. The differences between the determined bone ages of BAUS™ and assessments of GP and TW3 atlases were found to be non-significant. However, two different parameters from three different locations had to be taken into account to achieve this result. All these systems determine absolute bone age, are mainly used to analyze and predict growth in children, and are unsuitable for identifying an ossified growth plate.

In contrast to QUS devices, systems relying on B-mode US images have also been proposed. A study published in 2016 evaluated the postnatal skeletal development of the proximal femoral epiphysis, distal femoral epiphysis, and proximal tibial epiphysis in 178 premature and mature newborns, between 25 and 47 weeks of biological age, and found that B-mode musculoskeletal ultrasound “is very useful in monitoring the skeletal development of premature and mature newborns” [24]. Another study published in 2018 examined the wrists of 120 children from 10 to 17 years and found that “hand-wrist radiographs, which are traditional bone age determination methods, can be replaced with [ultrasonography] in the determination of bone age and especially in determining the extent of pubertal growth in the field of dentistry” [25]. More recently, a study published in 2021 found that approaches based on B-mode ultrasound “are of good accuracy in comparison to hand X-ray [but these approaches are] rarely used in daily practice. This may be because the examination needs to be performed by a trained specialist or there is a need for a specific device” [26]. Another study published in the same year examined the use of ultrasonography for age estimation by investigating the degree of ossification of the distal radial epiphysis on a Turkish population. To this end, “the left wrist of 688 (322 males, 366 females) patients between the ages of 9 and 25 years was prospectively evaluated by ultrasonography”, and it was found that “the data obtained may help [to] determine legally critical age limits of 14 and 15” [27].

In this work, we propose a low-cost, mobile, and custom-designed device, which determines the bone age solely by analyzing 1D US radiofrequency (RF) signals (i.e., A-scans) from the distal epiphyseal plates of the radius bone. We did not design the system to assess legal ages, as there is still a lack of a distinct, universally applicable gold standard due to biases toward white populations in existing approaches [1,2,9]. Our device assesses signals from the ulna and radius bones of 120 female subjects from a clinical study to determine the end of bone growth.

We evaluated the system using digital signal processing (DSP) and machine learning (ML) methods. Those were traditional algorithms artificial neural networks (ANNs) and gradient boosting machines (GBMs). The traditional algorithms consist of the 1-nearest neighbors algorithm exploiting the dynamic time warping distance (1-NN DTW) [28], support vector machines [29], and logistic regression [30]. The ANN methods include fully convolutional networks (FCNs) [31,32], multilayer perceptrons (MLPs) [31,32], residual networks (ResNets) [31,32], and radial basis function (RBF) neural networks [33]. The GBM approaches are XGBoost [34], LightGBM [35], and CatBoost [36].

To substitute our hypothesis that the determination of the end of bone growth by analyzing 1D US RF signals is possible, we extend previously published preliminary results [37]. We include a variety of different dimensionality reduction techniques (DRTs) and ML models to visualize the acquired data and classify the signals.

## 2. Materials and Methods

### 2.1. Measurement Principle

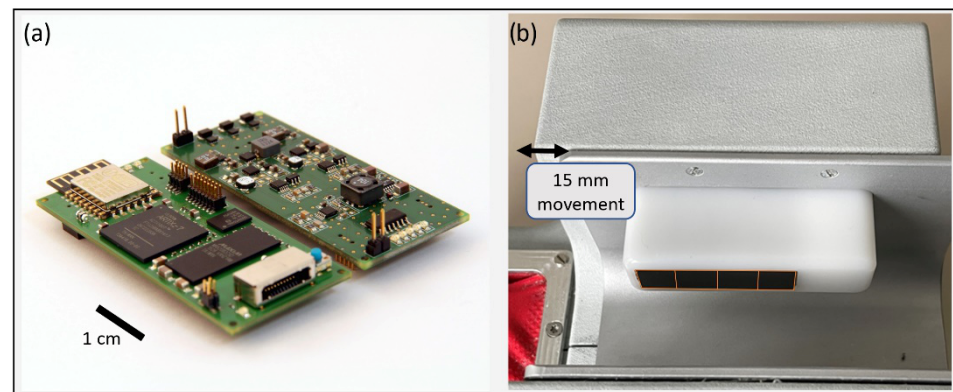
Figure 2 shows the device we designed and built, featuring a grab handle and a fixture containing US transducers.



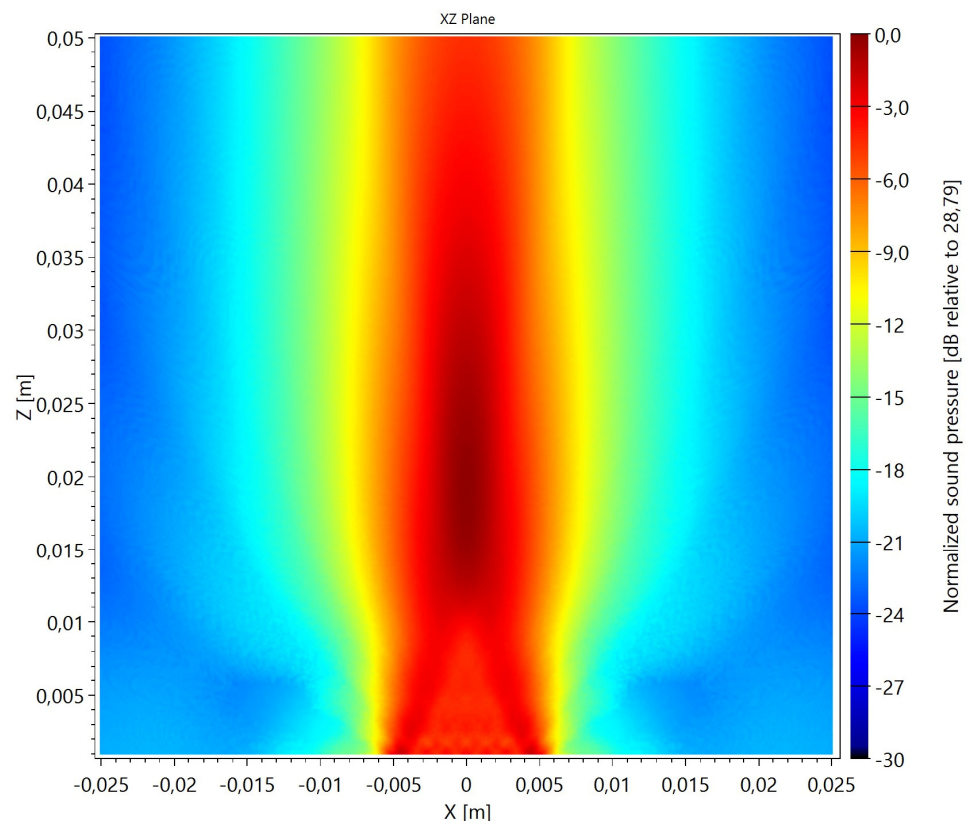
**Figure 2.** The portable integrated ultrasound-based measurement device tested and evaluated.

Two array transducers with four elements each gently press against the arm laterally and medially, respectively. Both transducer arrays on each side of the arm are used for acquiring pulse-echo reflection signals and receiving transmission signals. To increase the chances of measuring the potentially existing growth plate, the system moves all transducers in distal and proximal directions during data acquisition while the subject is holding a grab handle tightly with her hand. The use of US array transducers eliminates the need to reposition the transducer as the skin's elasticity permits epidermal movements of up to 15 mm. The grab handle restrains the subject's arm movements, preventing any involuntary actions, while the transducers are moving along the wrist and arm. It also serves as a reference point for the arm, reducing the variation in measurement location. The linear orientation of the transducer elements enables measurement locations of subjects with different arm lengths to be taken into account. We designed the system for forearm sizes ranging from 31 to 55 mm at the epiphyseal plate of the radius bone, which was suitable in most cases. Our system (see Figure 3a) does not perform 2D US imaging and acquires only raw A-scans. We use rectangular tristate burst excitation signals with a length of 3 or 5 periods at center frequency for each measurement path. In each measurement modality, 25 A-scans containing 2048 values each are sampled as a motorized sled moves the transducers along the forearm (see Figure 3b). The lateral resolution of subsequent ultrasound measurements is at least 0.75 mm in the middle of the motion area and slightly

smaller at the sides due to acceleration and deceleration of the motor. This lateral resolution is smaller than the lateral extent of the ultrasound beam resolution of a single array element. Figure 4 shows the simulated acoustic pressure distribution of such a single array element. For this simulation, we assume an average speed of sound of 2500 m/s for the measurement through the wrist consisting of tissue (average speed of sound of 1540 m/s) and bone (average speed of sound of 3900 m/s for girls around maturity [38]) in non-equal parts. The processing and analysis presented later do not rely on any assumptions regarding the average speed of sound through the tissue and bone.



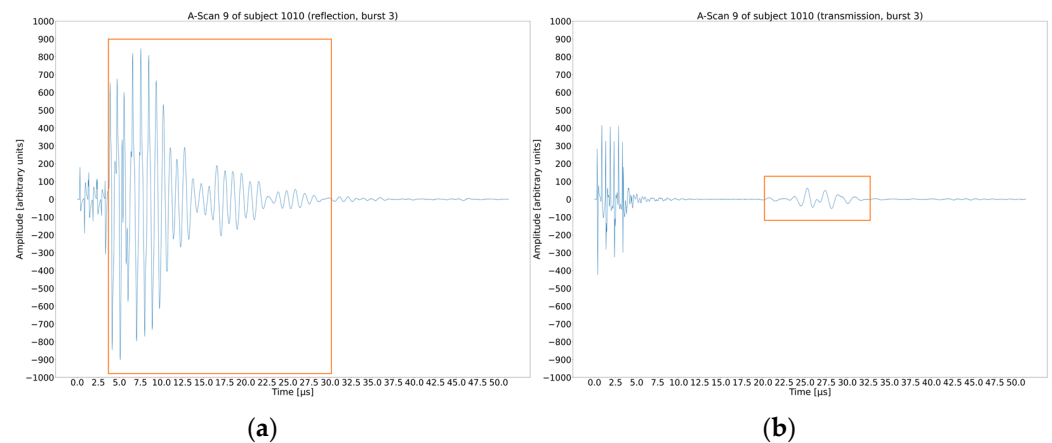
**Figure 3.** The compact, stackable ultrasound electronic module used in the device (a); the inbuilt ultrasound array (b).



**Figure 4.** The simulated acoustic pressure distribution of a single array element for a speed of sound of approximately 2500 m/s in bone.

Figure 5 depicts a sample reflection (left) and transmission (right) signal. The signals include clipped digitization of the transmitted burst pattern in the first approximately 5  $\mu$ s. Depending on the measurement mode the reflected ultrasound wave is directly

measured again at the same transducer (left), or the transmitted ultrasound wave is detected after traveling through the arm or wrist (right). While the reflected ultrasound signal is much stronger being measured after traveling only through skin and tissue layers, the transmitted signal amplitude is much lower, caused by the attenuation of the bone. The reflected signal is also detected over a longer period of time caused by multiple reflections between the transducer and bone. The transmission signal is only a superposition of the direct propagation through bone, the growth plate (if present), and the ultrasound wave traveling around the bone. In all cases, we analyze all significant parts of the signal not only containing the first arriving signal (FAS) but also multi-path propagations and multi-reflections (highlighted with orange boxes in Figure 5).



**Figure 5.** A-scan samples illustrating the reflection (a) and transmission (b) signals of the most distal element(s) acquired with a 3-cycle burst. The orange boxes indicate signal parts containing the most significant information.

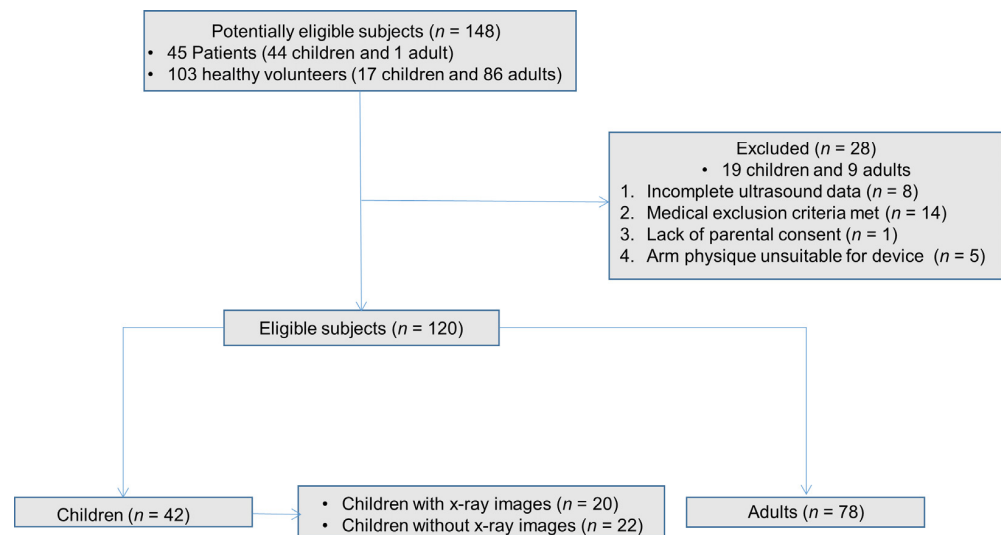
## 2.2. System Design

Our mobile, low-cost, and low-power battery-powered US electronics (see Figure 3a) communicate with tablet or laptop devices via a WIFI module. The transmission pulser generates rectangular, bipolar burst signals, and the received US echoes are digitized at a sampling rate of 40 MHz using eight parallel electrical channels. A pair of US array transducers with four elements in total are used for transmission and reflection measurements. Each element has an aperture size of  $11 \times 11$  mm, a center frequency of 1 MHz, and can be used individually for measurements. The transducers are built out of custom 1-3-composite materials and have a natural focus at a depth of 1.5 cm inside the radius or ulna bone. The acoustic pressure distribution of our system's natural focus is designed to provide an optimal energy yield and detection possibilities for a human wrist. All components were integrated into an electromagnetically shielded housing (see Figure 2), and the acoustic output was evaluated according to medical standards in collaboration with certified laboratories. A detailed description of the system is available [37].

## 2.3. Study Design

We evaluated the system in a clinical study with 148 girls and women at Saarland University Medical Center, including patients from pediatric endocrinology and healthy volunteers. The medical association of Saarland approved the study design (ID: 255/15). For patients with a clinical indication, an experienced pediatric endocrinology consultant, blinded to the age, height, and weight of each patient, determined the bone age from X-ray images using the GP atlas [6]. We excluded five subjects from the study whose arm physique was unsuitable for the device. This occurred in patients whose wrists were too slim, causing acoustic coupling issues, and in obese subjects, where the motor powering the US probe was unable to move along the oversized forearm. None of the subjects found the scans to be uncomfortable with the top layer of skin being pulled slightly (within a 15 mm

range) and the US transducer array on the medial side being pushed toward the forearm. We excluded subjects with known medical conditions affecting bone age when radiographs were not available. Bone age as derived from available radiographs or chronological age was used as annotation ground truth to distinguish between open and ossified growth plates. Figure 6 shows the STARD flowchart of the study cohort and details the reasons for the exclusion of 28 potentially eligible subjects, leaving 120 included subjects.



**Figure 6.** STARD 2015 flowchart of the study cohort. Exclusion criteria were previous fracture of the left wrist, hand, or underarm; precocious puberty; pubertas tarda; growth hormone deficiency; congenital adrenal hyperplasia; previous radio- or chemotherapy; previous high-dose steroid therapy; Turner syndrome; Down syndrome; transgender individuals undergoing/previous hormone therapy; tall stature individuals undergoing hormone therapy; non-treated hyper- or hypothyroidism or severe obesity (BMI > 40).

The data acquisition workflow consisted of eight subsequent measurement blocks (one for each transducer), acquiring reflection and transmission data for each subject. The resulting four back-and-forth motions of the motorized sled with the transducers were performed in less than one minute in total per subject.

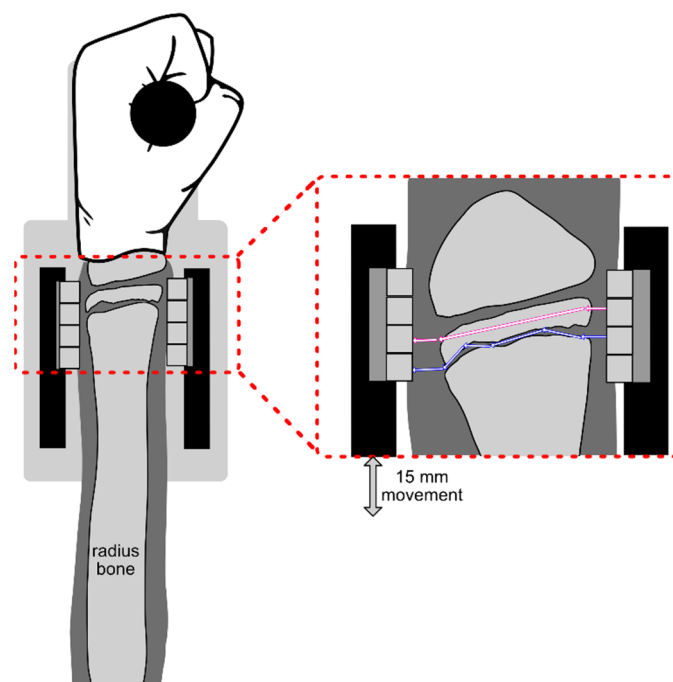
#### 2.4. Analysis

To focus on relevant echoes only, we extracted a window in the digitized data ranging from 25 to 50  $\mu$ s. In addition to processing the time signal, we also analyzed its spectral representation by performing a fast Fourier transform (FFT). Thus, we obtain a combination of 32 data sets in total, resulting from 2 possible signal modes (reflection or transmission), 2 possible excitation burst lengths (3 or 5 periods at transducer center frequency), 4 transducer elements, and 2 DSP options (raw signals or signals processed with a FFT). We applied classical DSP and ML methods to classify these series of measurements.

#### 2.5. Classical Digital Signal Processing

The presence of an open growth plate before its full ossification affects the ultrasonic wave propagation through the forearm and the wrist, which needs to be detected by an algorithm. Instead of previous approaches used by other researchers and vendors of bone age ultrasound devices, we do not measure the average speed of sound of an ultrasound wave through the bone to correlate this value with empirically determined speed of sound values for certain ossification states and bone ages. Our approach is based on the fact that the ultrasound signal differs more while the ultrasound transducer is moved over an open growth plate than it changes while being moved over fully ossified bone. Hence, the analysis must look for signal differences caused by changes in relative orientation of

the bone and the US transducers during transducer movements. In subjects with an open growth plate, the measurements may differ more significantly from the signals measured next to the epiphyseal gap, through the bone, or around the bone. This can be caused by less attenuation and change in US propagation speed due to different stages of ossification or changes in bone surface structure that reflect and diffract the US wave at the boundaries of the growth plate. The presence of a growth plate potentially leads to larger changes in reflectivity or ratio of transmitted US waves as the reflectivity and intensity of transmitted US waves differ more than the signal changes while moving over an ossified growth plate. A growth plate without full ossification results in a slower US wave as it travels through more non-bony tissue and less-ossified bone tissue. Due to the lateral extension of the acoustic beam, parts of the wave still travel through the bone tissue located outside the growth plate. This results in an overlay of both traveling paths, with the signal through the open growth plate submerging into the other signal (see Figure 7).



**Figure 7.** The left forearm is placed in the device with the transducers of the motorized sled positioned around the assumed location of the growth plate (left). The acoustic wave propagation through the growth plate (blue) and the radius bone (pink) is depicted in the enlarged detail inset on the right.

The exact location of a potentially open growth plate varies between individuals. However, in order to detect full growth plate ossification, identifying the exact location of the growth plate is not necessary, only its potential influence on signal change when the transducer moves across it.

Our classical signal processing approach computes the integral changes in signal amplitude of the received signals over all measurements during transducer motion over the possible open growth plate area. It also computes the sum of all differences over each consequential A-scan difference and the sum of all changes over all A-scans. All permutations of transmit/receive ultrasound transducer elements are used for both transmission scans through the wrist/arm and reflection measurements on a single side using the same array transducer. According to our hypothesis, these integrated changes of subsequent measurements during the movement should be higher in subjects with an open growth plate compared to those with a fully ossified growth plate.



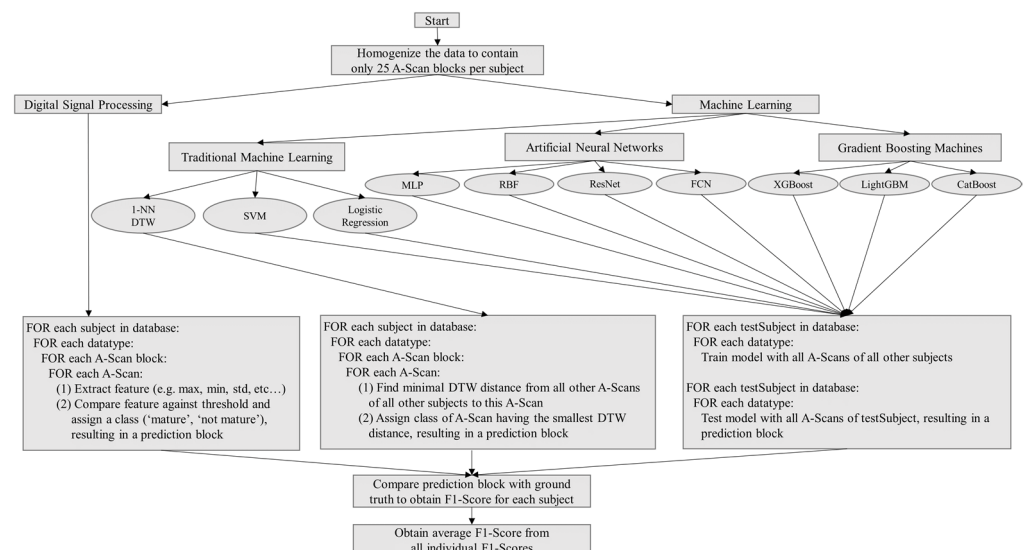
## 2.6. Machine Learning

### 2.6.1. Dimensionality Reduction Techniques

As a first step to analyzing the data, we used DRTs to gain a better understanding of the underlying structure of the contained information. Even though many approaches exist, we focused on the t-distributed stochastic neighbor embedding (t-SNE) technique, which is well suited for 2D visualizations and able to handle nonlinear data [39], and principal component analysis (PCA), which is a deterministic DRT aiming to reduce information losses [40].

### 2.6.2. Classification Methods

We treated the A-scans as time series with up to 2048 elements and thus faced a non-trivial time series classification (TSC) task. We embedded all methods in a general data processing pipeline (see Figure 8) and deployed a strict leave-one-out cross-validation (LOOCV) regime for all ML-based classifications. Thus, to evaluate the performance of a model, all 25 A-scans of a given subject were used as input data for a model that has been trained on all A-scans of all other subjects. Those A-scans were either raw unprocessed or preprocessed 1D ultrasound radio frequency signals.



**Figure 8.** General workflow of the complete data processing pipeline.

We obtained a binary prediction for each A-scan of each participant (i.e., 25 predictions for each of the 120 eligible participants). Those predictions were then compared to the ground truth to obtain the  $F_1$  score for each participant. These  $F_1$  scores of each participant were then averaged to obtain the final  $F_1$  score for the whole model.

The ML methods used for TSC were 1-NN DTW, SVM, logistic regression, MLP, FCN, ResNet, RBF neural networks, XGBoost, LightGBM, and CatBoost.

We include deep learning algorithms based on 1D US RF signals as those have recently been shown to be accurate in similar classification scenarios, such as the “diagnosis of nonalcoholic fatty liver disease and hepatic fat fraction quantification when other causes of steatosis are excluded” [41]. Other methods have been included due to their general TSC performance.

MLPs work well as TSC baseline but do not exhibit any spatial invariance, as each time stamp has its own weight and the temporal information is lost [32]. It has been shown that FCNs can work well in TSC settings [31,32] if convolutions are applied as sliding windows on the input signal. By applying several convolutions, it becomes possible for FCNs to learn multiple discriminative features. If the same convolutions are used for all time series, an FCN learns filters that are invariant across the time dimension [32]. ResNet was first adapted for TSC in 2017 [31] and has been refined in later work [32], in which it performed comparatively well. We included this refined architecture in our evaluation as

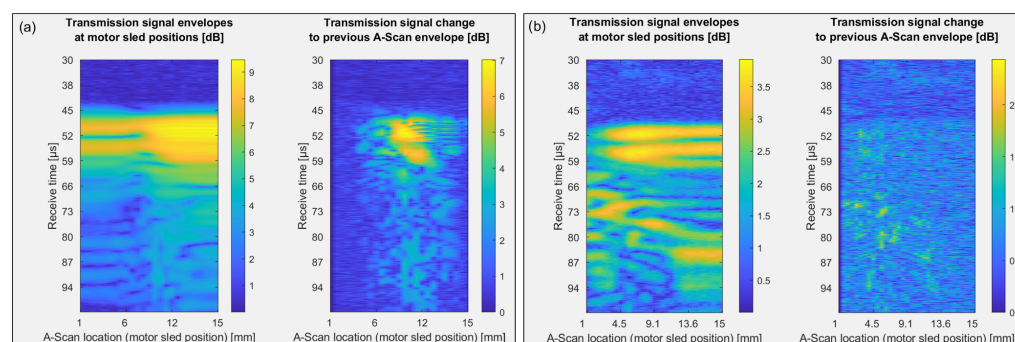
we expected it to be superior to other relatively shallow architectures. RBF neural networks use radial basis functions as activation functions and typically only consist of an input layer, a hidden layer with a nonlinear RBF activation function, and a linear output layer [33]. We included this approach in our evaluation as it historically performed well on 1D US RF signals [42]. k-NN DTW algorithms classify signals based on their respective DTW distance to each other. The DTW metric is able to match patterns of different signals, regardless of their positions. Even though this algorithm has been known for decades, a comprehensive evaluation of different TSC algorithms concluded that many of the tested algorithms were no better than 1-NN DTW in a 2016 publication [24]. GBM models, consisting of ensembles of weak prediction models to achieve superior accuracies, are usually based on decision trees and have shown promising results in the past. We have already observed suitable results with GBMs in other work classifying similar 1D US RF signals to identify advanced hepatic steatosis and fibrosis using ML algorithms in patients with non-alcoholic fatty liver disease [43].

Our aim was to obtain binary predictions to distinguish between ongoing or finished bone growth in the epiphyseal plate of the radius bone. To this end, we categorized all subjects with bone ages  $< 18$  years and bone ages  $\geq 18$  years into different groups. For DSP and 1-NN DTW, the pipeline processes each A-scan separately. For the ML methods, iterations over each subject create a model trained on data from all other subjects of the database. We evaluated each model by presenting it with the data from the test subject, performing a leave-one-out cross-validation.

### 3. Results

#### 3.1. Classical Digital Signal Processing

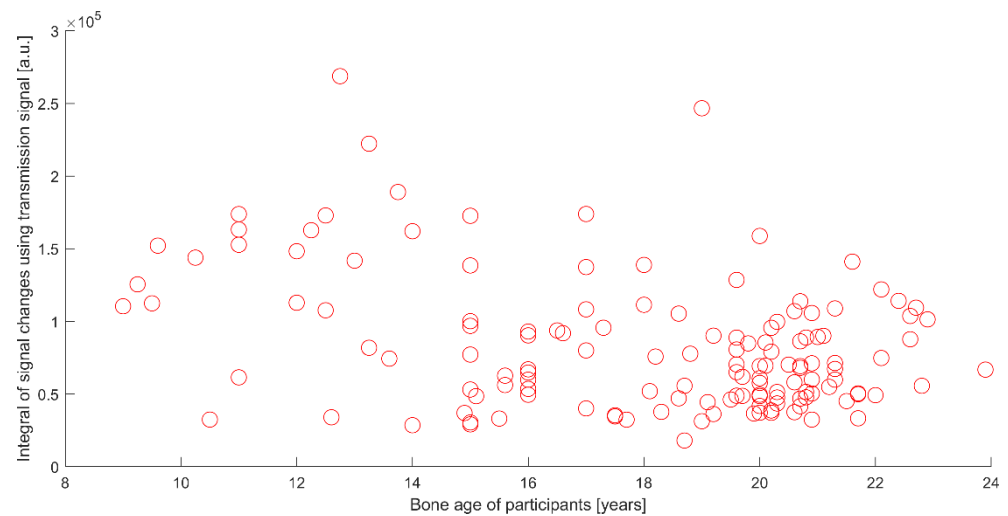
For all subjects, all reflection and transmission measurements were analyzed for all permutations of two transducer elements of the arrays. It is not possible to classify a single US wave traveling through an open growth plate in single A-scans with only small changes in amplitude, frequency content, or phase shifts caused by wave cancellation. On the other hand, the signal changes over time during sled movement. Both the intensity value and the signal change in terms of the difference of consequential A-scans were used for processing (see Figure 9).



**Figure 9.** Analysis of measurement data plotted as A-scans (color coded) over motor movement (x-axis) and measurement time (y-axis) of a person with an open growth plate (a) and an ossified growth plate (b); transmission signal envelope amplitude (left-hand plot in panels a and b) and transmission signal change (difference signal change to previous A-Scan, right-hand plot in panels a and b).

Intensity fluctuations were seen best in the difference image, which removed the static background signal. The integral sum over all amplitude intensities moving over a possible growth plate provided us with a relevant feature. While subjects with a fully ossified growth plate showed only slight signal changes during transducer movement along the forearm, subjects with an open growth plate showed a stronger signal change, caused by diffraction, scattering, and reflection of the US wave through the open growth plate. Although a difference in integrated signal change is seen for younger ( $< 15$  years of bone

age) and older ( $\geq 15$  years of bone age) women, a linear clustering based on such a feature extraction does not look very promising (see Figure 10).



**Figure 10.** Plot of integral values of transmission signal changes during motor-based transducer motion (y-axis) in relation to subject bone age (x-axis) using the most distal transducer element. Each circle represents a single individual.

Using a combination of multiple additional features extracted by signal processing could lead to a more robust classifier but optimizing such a higher dimensional nonlinear classification is also challenging.

With this result, we proceeded with ML approaches, as those proved most promising in creating a robust classifier for our purposes.

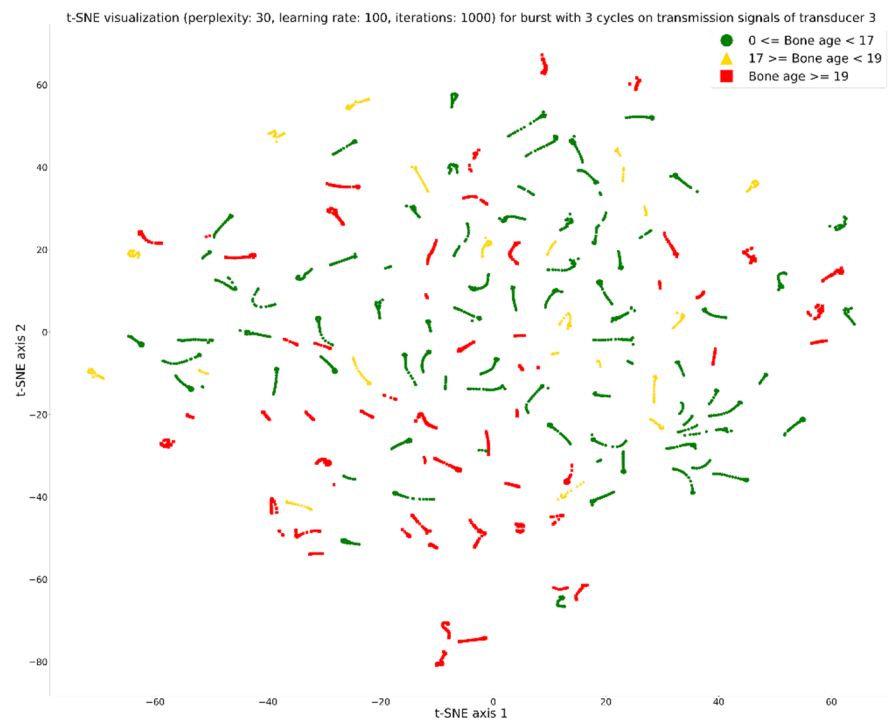
### 3.2. Machine Learning

We deployed t-SNE to visualize the distribution of the signals and found that they tend to cluster according to certain age segments but cluster for each subject much more densely (see Figure 11).

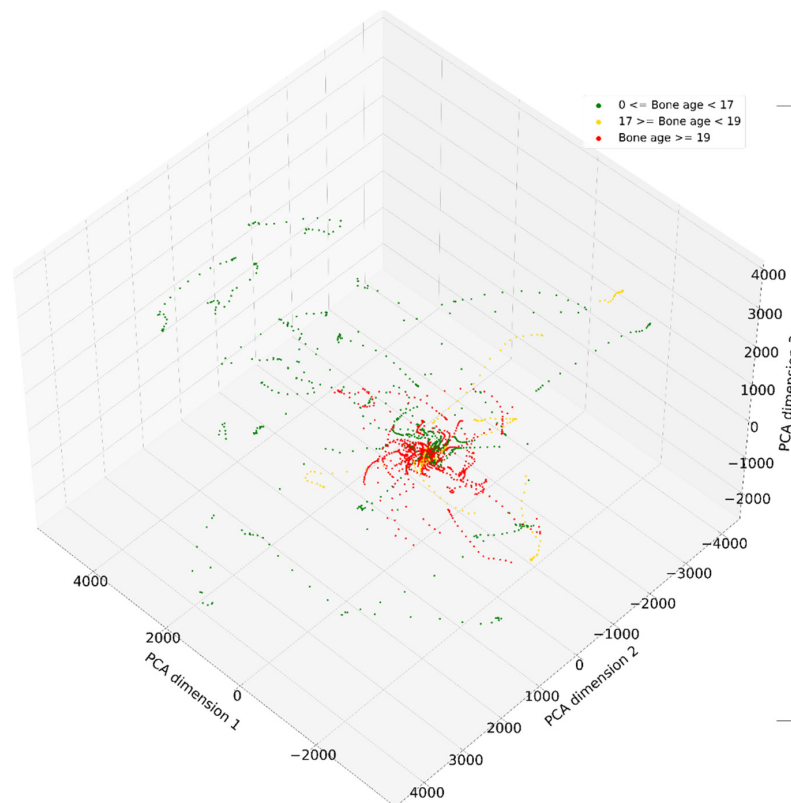
We also performed a PCA on the same signals to quantify the clustering. Figure 12 visualizes the first three principal components (PCs) of the PCA, which shows that the signals cluster strongly according to age segments.

Even though there are some overlapping data points, the overall distribution indicates the data points belonging to the age segment  $> 19$  years cluster more densely than the data points belonging to other age segments. Figure 13 provides an overview of the ratio of explained variance of the first 10 PCs.

These analyses indicate that A-scans from similarly aged subjects tend to cluster together in high-dimensional spaces. The evaluation metric of our ML models was the  $F_1$  score, which is defined as the harmonic mean of precision and recall. Table 1 shows a comparison of the performance of different ML methods by comparing the times needed to train and evaluate each model and the average  $F_1$  score for certain subsets of signals that were acquired from transducer #3 with a burst excitation with a length of 3 periods at US center frequency.



**Figure 11.** Two-dimensional t-SNE visualization showing the interrelation of all A-scans. Color codes for bone age groups: green  $\leq 17$  years, yellow = 17 to <19 years, and red  $\geq 19$  years.



**Figure 12.** Three-dimensional PCA showing the relation of all A-scans to each other. Green dots represent A-scans belonging to a subject younger than 17 years, orange dots represent A-scans belonging to a subject between the age of 17 and 19 years, while red dots represent A-scans belonging to subjects being 19 years or older.

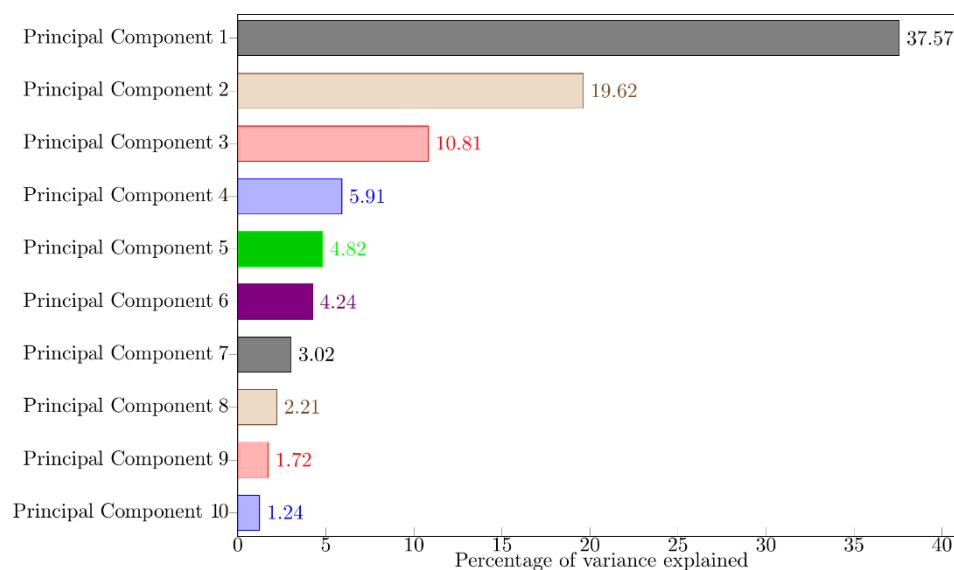


Figure 13. Ratio of explained variance of the first 10 principal components of the data.

Table 1. Average F<sub>1</sub> score and corresponding time for training and evaluation of data stemming from transducer 3, excited with a 3 cycles burst for all subjects. The best performance for each column is depicted in green, and the second-best performance is depicted in yellow.

Method	Average F <sub>1</sub> Score for All Subjects (%)	Average F <sub>1</sub> Score for All Subjects under 17 years (%)	Average F <sub>1</sub> Score for All Subjects under 18 years (%)	Average F <sub>1</sub> Score for All Subjects between 17 and 19 years (%)	Average F <sub>1</sub> Score for All Subjects over 18 years (%)	Average F <sub>1</sub> Score for All Subjects over 19 years (%)	Time for Training and Evaluation (in h)
CatBoost (10,000 iterations)	86.93	83.68	78.48	82.57	90.72	91.35	26.08
CatBoost (1000 iterations)	86.40	76.38	91.30	81.19	83.86	91.20	2.65
XGBoost	85.23	80.22	76.67	78.86	90.61	91.85	0.08
CatBoost (100 iterations)	84.77	77.24	88.41	81.84	79.43	89.45	0.23
LightGBM	82.57	74.70	72.95	77.71	90.38	90.33	0.06
FCN	82.30	77.62	72.29	70.00	89.80	91.71	186.18
1-NN DTW	81.50	65.62	64.86	75.00	91.54	91.42	4.31
ResNet	78.53	61.84	59.43	74.29	91.19	91.93	271.75
MLP	77.73	66.49	63.81	66.00	86.14	85.75	47.99
SVM	77.00	70.59	66.95	67.43	82.97	85.86	0.18
Logistic Regression	75.20	61.51	59.71	71.00	85.74	86.55	0.82
Radial Basis Functions	73.10	50.03	47.33	78.71	88.00	85.60	0.23
Neural Network							

Previously, we achieved an average F<sub>1</sub> score of approximately 84% with a ResNet implementation for all subjects [37]. To increase the robustness and improve the times for training and evaluation, we conducted further analyses, including different input data types (raw unprocessed and processed 1D US RF signals) and a variety of different ML models.

This resulted in a total average F<sub>1</sub> score of approximately 87% using a CatBoost model with 10,000 iterations, which took ca. 26 h to complete. A CatBoost model training for 1000 iterations performed only slightly worse and achieved a total average F<sub>1</sub> score of ca. 86% in roughly 2 h and 40 min. XGBoost achieved the third-best total average F<sub>1</sub> score of ca. 85% in less than 5 min. The total average F<sub>1</sub> score of all other models was significantly less. We examined several age segments separately to illustrate our model’s feasibilities for different scenarios. Notably, the age segments > 18 years and >19 years of bone age

performed better for almost all models than the age segments < 17 years and <18 years of bone age (see Table 1).

#### 4. Discussion

Employing DSP, we demonstrated that the presence of an open growth plate influences the signal changes during transducer movements over the distal end of the forearm. However, we found that classification could not be achieved using naive threshold DSP algorithms or linear classifiers. We, therefore, developed robust ML classifiers using data from 120 girls and women considered eligible for inclusion out of a cohort of 148 potential subjects. Notably, GBMs outperformed ANNs in terms of accuracy and speed in almost all age segments. These results are in line with our expectations because MLPs do not exhibit any spatial invariance, and temporal information is lost. FCNs can learn filters that are invariant across the time dimension but are comparatively shallow. A deeper convolutional architecture is ResNet, which has been shown to be superior to shallow architectures [32]. ResNet performs comparatively well for the bone age segment > 18 years, while RBF neural networks are very shallow [33] and achieved the lowest performance. The fastest models, SVM, logistic regression, LightGBM, and XGBoost, yielded results within minutes, while the slowest model, ResNet, needed more than 11 days to complete. The best total average  $F_1$  score of approximately 87%, obtained using a CatBoost model, showed that robust classifications are possible with our system. The XGBoost classifier yielded an  $F_1$  score of approximately 85% within less than 5 min for training and evaluation. In the bone age segment > 18 years, all models performed better compared to younger bone age segments. We suspect that premature closing of the growth plate or artifacts from stiffer soft tissue regions might be responsible for misclassification in some younger subjects. We consider  $F_1$  scores of around 80% for bone age segments with younger subjects and  $F_1$  scores of more than 90% for bone age segments with older subjects provide a reasonable indicator, although further refinements are needed if the device is to be used for medical purposes. Our study presents an alternative approach to existing solutions based on X-ray imaging or B-mode ultrasound [24–27]. A potential limitation of our study is its dependence on data from girls and women of a single, comparatively homogeneous population and the partial determination of bone age from X-ray images using the GP atlas. Therefore, the results may not be representative of other populations, and the models we present may not be applicable to all ethnicities and to all sexes [9,10]. This limitation warrants further research, for example, multi-center studies that also include male and ethnically diverse subjects.

#### 5. Conclusions

We demonstrate that our ultrasound system can classify the fusion of the distal growth plate of the radius bone and with it the end of bone growth with a high accuracy offering the possibility of being a radiation-free alternative to established X-ray imaging methods. The results show that it is possible to construct robust ML models based on 1D US RF signals that can be acquired with low-cost equipment. Future clinical applications will require a larger and more diverse database, including male subjects and subjects of other ethnicities, as it is known that these factors have an impact on bone growth. Furthermore, future research also needs to take inter- and intra-observer reliabilities into account to quantify whether measurements differ significantly for different participants.

**Author Contributions:** Conceptualization, C.S., S.T., T.R. and H.H.; methodology, C.R., M.S., S.T., T.R. and H.H.; software, L.B., C.R. and H.H.; validation, R.D., T.R. and H.H.; formal analysis, L.B., R.D., T.R. and H.H.; investigation, R.D.; resources, S.T. and T.R.; data curation, L.B. and H.H.; writing—original draft preparation, L.B., R.D. and H.H.; writing—review and editing, S.T. and T.R.; visualization, L.B. and H.H.; supervision, S.T. and T.R.; project administration, T.R. and H.H.; funding acquisition, S.T., T.R. and H.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Federal Ministry of Education and Research of Germany (BMBF) in the project PRIMSA (PRIMSA: Prevention of and intervention in sex trafficking) in the framework “Research for Civil Security 2012–2017”, Funding code: 13N13519 and 13N13520.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of “Ärztchamber des Saarlandes” (protocol code Bu 255/15, 18 April 2016).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

## References

- Mughal, A.M.; Hassan, N.; Ahmed, A. Bone age assessment methods: A Critical Review. *Pak. J. Med. Sci.* **2014**, *30*, 211. [[CrossRef](#)] [[PubMed](#)]
- Creo, A.L.; Schwenk, W.F. Bone age: A Handy Tool for Pediatric Providers. *Pediatrics* **2017**, *140*, e20171486. [[CrossRef](#)] [[PubMed](#)]
- Ubelaker, D.H.; Khosrowshahi, H. Estimation of age in forensic anthropology: Historical Perspective and Recent Methodological Advances. *Forensic Sci. Res.* **2019**, *4*, 1–9. [[CrossRef](#)] [[PubMed](#)]
- Sauer, P.J.; Nicholson, A.; Neubauer, D. Age determination in asylum seekers: Physicians Should Not Be Implicated. *Eur. J. Pediatr.* **2016**, *175*, 299–303. [[CrossRef](#)]
- Hassan, N.; Noor, F.; Ahmad, S.; Fazili, K.M. Age of fusion of the distal radial and ulnar epiphyses from hand radiographs—A study in Kashmiri population. *Sci. Justice* **2016**, *56*, 431–436. [[CrossRef](#)]
- Greulich, W.; Pyle, S. Radiographic atlas of skeletal development of the hand and wrist. *Am. J. Phys. Anthr.* **1950**, *8*, 518–520. [[CrossRef](#)]
- Tanner, J.M.; Cameron, N. *Assessment of Skeletal Maturity and Prediction of Adult Height (TW3 Method)*, 3rd ed.; Saunders: London, UK, 2001.
- Gilsanz, V.; Ratib, O. *Hand Bone Age: A Digital Atlas of Skeletal Maturity*; Springer: Berlin, Germany, 2005; pp. 11–19. [[CrossRef](#)]
- Schmidt, S.; Nitz, I.; Schulz, R.; Tsokos, M.; Schmeling, A. The digital atlas of skeletal maturity by Gilsanz and Ratib: A Suitable Alternative for Age Estimation of Living Individuals in Criminal Proceedings? *Int. J. Leg. Med.* **2009**, *123*, 489–494. [[CrossRef](#)]
- Alshamrani, K.; Messina, F.; Offiah, A.C. Is the Greulich and Pyle atlas applicable to all ethnicities? A systematic review and meta-analysis. *Eur. Radiol.* **2019**, *29*, 2910–2923. [[CrossRef](#)]
- Hojreh, A.; Gamper, J.; Schmook, M.T.; Weber, M.; Prayer, D.; Herold, C.J.; Noebauer-Huhmann, I.M. Hand MRI and the Greulich-Pyle atlas in skeletal age estimation in adolescents. *Skelet. Radiol.* **2018**, *47*, 963–971. [[CrossRef](#)]
- Pennock, A.T.; Bomar, J.D.; Manning, J.D. The creation and validation of a knee bone age atlas utilizing MRI. *JBJS* **2018**, *100*, e20. [[CrossRef](#)]
- Langton, C.M.; Palmer, S.B.; Porter, R.W. The measurement of broadband ultrasonic attenuation in cancellous bone. *Eng. Med.* **1984**, *13*, 89–91. [[CrossRef](#)] [[PubMed](#)]
- Chen, Z.; Luo, W.; Zhang, Q.; Lei, B.; Wang, T.; Chen, Z.; Liu, J. Osteoporosis Diagnosis Based on Ultrasound Radio Frequency Signal via Multi-channel Convolutional Neural Network. In Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Guadalajara, Mexico, 1–5 November 2021; IEEE: New York, NY, USA, 2021; pp. 832–835. [[CrossRef](#)]
- Castriota-Scanderbeg, A.D.M.V.; De Micheli, V. Ultrasound of femoral head cartilage: A New Method of Assessing Bone age. *Skelet. Radiol.* **1995**, *24*, 197–200. [[CrossRef](#)] [[PubMed](#)]
- Castriota-Scanderbeg, A.; Sacco, M.C.; Emberti-Gialloreti, L.; Fraracci, L. Skeletal age assessment in children and young adults: Comparison between a Newly Developed Sonographic Method and Conventional Methods. *Skelet. Radiol.* **1998**, *27*, 271–277. [[CrossRef](#)] [[PubMed](#)]
- Shimura, N.; Koyama, S.; Arisaka, O.; Imataka, M.; Sato, K.; Matsuura, M. Assessment of measurement of children’s bone age ultrasonically with Sunlight BonAge. *Clin. Pediatr. Endocrinol.* **2005**, *14*, S24\_17–S24\_20. [[CrossRef](#)]
- Mentzel, H.J.; Vilser, C.; Eulenstein, M.; Schwartz, T.; Vogt, S.; Böttcher, J.; Kaiser, W.A. Assessment of skeletal age at the wrist in children with a new ultrasound device. *Pediatr. Radiol.* **2005**, *35*, 429–433. [[CrossRef](#)]
- Mentzel, H.J.; Vogt, S.; Vilser, C.; Schwartz, T.; Eulenstein, M.; Böttcher, J.; Kaiser, W.A. Abschätzung des Knochenalters mit einer neuen Ultraschallmethode. In *RöFo-Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*; Georg Thieme Verlag: New York, NY, USA, 2005; Volume 177, pp. 1699–1705. [[CrossRef](#)]
- Laugier, P.; Talmant, M.; Thien-Ly, P.H.A.M. Quo vadis, ultrasonics of bone? Present state and future trends. *Arch. Acoust.* **2008**, *33*, 553–564.
- Khan, K.M.; Miller, B.S.; Hoggard, E.; Somani, A.; Sarafoglou, K. Application of ultrasound for bone age estimation in clinical practice. *J. Pediatr.* **2009**, *154*, 243–247. [[CrossRef](#)]

22. Rachmiel, M.; Naugolani, L.; Mazor-Aronovitch, K.; Levin, A.; Koren-Morag, N.; Bistrizter, T. Bone age assessment by a novel quantitative ultrasound based device (SonicBone), is comparable to the conventional Greulich and Pyle method. *Horm. Res. Pediatr.* **2013**, *80*, 35.
23. Aref Elnasasra, M.D.; Hilmi Alnsasra, M.D.; Rozalia Smolyakov, M.D.; Klaris Riesenber, M.D.; Lior Nesher, M.D. Bone age assessments by quantitative ultrasound (SonicBone) and hand X-ray based methods are comparable. *Isr. Med. Assoc. J. IMAJ* **2017**, *9*, 533–538.
24. Windschall, D.; Pommerenke, M.; Haase, R. Ultrasound assessment of the skeletal development of the proximal tibial, proximal femoral, and distal femoral epiphyses in premature and mature newborns. *Ultrasound Med. Biol.* **2016**, *42*, 451–458. [[CrossRef](#)]
25. Torenek Ađırman, K.; Bilge, O.M.; Milođlu, Ö. Ultrasonography in determining pubertal growth and bone age. *Dentomaxillofacial Radiol.* **2018**, *47*, 20170398. [[CrossRef](#)] [[PubMed](#)]
26. Prokop-Piotrkowska, M.; Marszałek-Dziuba, K.; Moszczyńska, E.; Szalecki, M.; Jurkiewicz, E. Traditional and new methods of bone age assessment-an overview. *J. Clin. Res. Pediatr. Endocrinol.* **2021**, *13*, 251. [[CrossRef](#)] [[PubMed](#)]
27. Ekizoglu, O.; Er, A.; Buyuktoka, A.D.; Bozdog, M.; Karaman, G.; Moghaddam, N.; Grabherr, S. Ultrasonographic assessment of ossification of the distal radial epiphysis for estimating forensic age. *Int. J. Leg. Med.* **2021**, *135*, 1573–1580. [[CrossRef](#)] [[PubMed](#)]
28. Bagnall, A.; Lines, J.; Bostrom, A.; Large, J.; Keogh, E. The great time series classification bake off: A Review and Experimental Evaluation of Recent Algorithmic Advances. *Data Min. Knowl. Discov.* **2017**, *31*, 606–660. [[CrossRef](#)] [[PubMed](#)]
29. Mitsa, T. *Temporal Data Mining*; Chapman and Hall/CRC: London, UK, 2010. [[CrossRef](#)]
30. Susto, G.A.; Cenedese, A.; Terzi, M. Time-series classification methods: Review and Applications to Power Systems Data. *Big Data Appl. Power Syst.* **2018**, 179–220. [[CrossRef](#)]
31. Wang, Z.; Yan, W.; Oates, T. Time series classification from scratch with deep neural networks: A Strong Baseline. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; IEEE: New York, NY, USA, 2017; pp. 1578–1585. [[CrossRef](#)]
32. Ismail Fawaz, H.; Forestier, G.; Weber, J.; Idoumghar, L.; Muller, P.A. Deep learning for time series classification: A Review. *Data Min. Knowl. Discov.* **2019**, *33*, 917–963. [[CrossRef](#)]
33. Vidnerová, P. RBF-Keras: An RBF Layer for Keras Library. 2019. Available online: [https://github.com/PetraVidnerova/rbf\\_keras](https://github.com/PetraVidnerova/rbf_keras) (accessed on 15 February 2022).
34. Chen, T.; Guestrin, C. Xgboost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794. [[CrossRef](#)]
35. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Liu, T.Y. Lightgbm: A Highly Efficient Gradient Boosting Decision Tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3149–3157.
36. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased Boosting with Categorical Features. *arXiv* **2018**, arXiv:1706.09516.
37. Hewener, H.; Risser, C.; Brausch, L.; Rohrer, T.; Tretbar, S. A mobile ultrasound system for majority detection. In Proceedings of the 2019 IEEE International Ultrasonics Symposium (IUS), Glasgow, UK, 6–9 October 2019; IEEE: New York, NY, USA, 2019; pp. 502–505. [[CrossRef](#)]
38. Zadik, Z.; Price, D.; Diamond, G. Pediatric reference curves for multi-site quantitative ultrasound and its modulators. *Osteoporos. Int.* **2003**, *14*, 857–862. [[CrossRef](#)]
39. van der Maaten, L.; Hinton, G. Visualizing data using t-sne. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
40. Jolliffe, I.T.; Cadima, J. Principal component analysis: A Review and Recent Developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20150202. [[CrossRef](#)] [[PubMed](#)]
41. Han, A.; Byra, M.; Heba, E.; Andre, M.P.; Erdman, J.W., Jr.; Loomba, R.; O'Brien, W.D., Jr. Noninvasive diagnosis of nonalcoholic fatty liver disease and quantification of liver fat with radiofrequency ultrasound data using one-dimensional convolutional neural networks. *Radiology* **2020**, *295*, 342–350. [[CrossRef](#)] [[PubMed](#)]
42. Guo, J.Y.; Zheng, Y.P.; Xie, H.B.; Koo, T.K. Towards the application of one-dimensional sonomyography for powered upper-limb prosthetic control using machine learning models. *Prosthet. Orthot. Int.* **2013**, *37*, 43–49. [[CrossRef](#)] [[PubMed](#)]
43. Brausch, L.; Tretbar, S.; Hewener, H. Identification of advanced hepatic steatosis and fibrosis using ML algorithms on high-frequency ultrasound data in patients with non-alcoholic fatty liver disease. In Proceedings of the 2021 IEEE UFFC Latin America Ultrasonics Symposium (LAUS), Gainesville, FL, USA, 4–5 October 2021; IEEE: New York, NY, USA, 2021; pp. 1–4. [[CrossRef](#)]