Ultrasound in Med. & Biol., Vol. 42, No. 6, pp. 1337–1356, 2016

0301-5629/$ - see front matter

http://dx.doi.org/10.1016/j.ultrasmedbio.2016.01.024

● *Original Contribution*

AN ADVANCED QUANTITATIVE ECHOSOUND METHODOLOGY FOR FEMORAL NECK DENSITOMETRY

Sergio Casciaro,* Marco Peccarisi,† Paola Pisani,* Roberto Franchini,* Antonio Greco,† Tommaso De Marco,† Antonella Grimaldi,‡ Laura Quarta,‡ Eugenio Quarta,‡ Maruizio Muratore,‡ and Francesco Conversano*

*National Research Council, Institute of Clinical Physiology, Lecce, Italy; †Echolight srl, Lecce, Italy; and ‡Operative Unit of Rheumatology, Galateo Hospital, San Cesario di Lecce, Lecce, Italy

(*Received* 14 *August* 2015; *revised* 26 *January* 2016; *in final form* 27 *January* 2016)

Abstract—The aim of this paper was to investigate the clinical feasibility and the accuracy in femoral neck densitometry of the Osteoporosis Score (O.S.), an ultrasound (US) parameter for osteoporosis diagnosis that has been recently introduced for lumbar spine applications. A total of 377 female patients (aged 61–70 y) underwent both a femoral dual X-ray absorptiometry (DXA) and an echographic scan of the proximal femur. Recruited patients were sub-divided into a reference database used for ultrasound spectral model construction and a study population for repeatability assessments and accuracy evaluations. Echographic images and radiofrequency signals were analyzed through a fully automatic algorithm that performed a series of combined spectral and statistical analyses, providing as a final output the O.S. value of the femoral neck. Assuming DXA as a gold standard reference, the accuracy of O.S.-based diagnoses resulted 94.7%, with $k = 0.898$ ($p < 0.0001$). Significant correlations were also found between O.S.-estimated bone mineral density and corresponding DXA values, with r^2 up to 0.79 and root mean square error = 5.9–7.4%. The reported accuracy levels, combined with the proven ease of use and very good measurement repeatability, provide the adopted method with a potential for clinical routine application in osteoporosis diagnosis. (E-mail: sergio.casciaro@cnr.it) **© 2016 World Federation for Ultrasound in Medicine & Biology.**

Key Words: **Osteoporosis diagnosis, Quantitative Ultrasound, Bone Mineral Density Measurement, Femoral Neck, Bone Densitometry, Radiofrequency Signal Analysis.**

INTRODUCTION

Osteoporosis is a highly prevalent bone disease characterized by a decrease in bone mass accompanied by microarchitectural alterations, resulting in bone fragility and increased fracture risk (Liu et al. 2011). Vertebral and hip fractures are the most frequent, expensive and disabling osteoporotic fractures, leading to reductions in patient quality of life and also causing significant mortality rates (Albanese et al. 2011; Cooper 1997; Ensrud et al. 2000).

According to the operational definition provided by the World Health Organization, osteoporosis is diagnosed when bone mineral density (BMD) measured at lumbar spine or proximal femur is at least 2.5 standard deviations (SDs) lower than the young adult mean (*i.e.,* T-score ≤ -2.5) (Genant et al. 1999; Kanis 1994). Analogously, a patient will be classified as "osteopenic" if $-2.5 <$ T-score < -1.0, or "healthy" if T-score ≥ -1.0. For instance, according to the National Health and Nutrition Examination Survey database, young adult white women (aged in 20–25 y) had an average femoral neck BMD (mean ± SD) of 0.847 ± 0.110 g/cm^2, therefore a generic white woman will be classified as "osteoporotic" if her femoral neck BMD is ≤0.572 g/cm^2, "osteopenic" if her BMD is internal to the range 0.572–0.737 g/cm^2, or "healthy" if BMD ≥0.737 g/cm^2.

Dual X-ray absorptiometry (DXA) currently represents the most common method for osteoporosis diagnosis and it is considered the gold standard reference for BMD assessments (Baim and Leslie 2012; Link 2012; Schnitzer et al. 2012). However, DXA also has some important intrinsic limitations that prevent its use

Address correspondence to: Sergio Casciaro, Consiglio Nazionale delle Ricerche, Istituto di Fisiologia Clinica, c/o Campus Ecotekne (Ed. A7), via per Monteroni, 73100 Lecce, Italy. E-mail: sergio.casciaro@cnr.it

for population mass screenings (*i.e.,* exposure to ionizing radiation with associated risks, high costs, and the need for dedicated structures with certified operators) (Pisani et al. 2013), and this resulted in the development of quantitative ultrasound (QUS) approaches to bone health assessment and osteoporosis diagnosis (Breban et al. 2010; Nayak et al. 2006; Paggiosi et al. 2012; Pais et al. 2010; Schnitzer et al. 2012; Trimpou et al. 2010). In fact, QUS methods have several potential advantages over DXA, including absence of ionizing radiation, lower costs, portable machines and availability in primary care settings without requiring dedicated structures or certified operators.

Commercially available QUS devices are presently usable only on peripheral bone districts, such as the calcaneus, and numerous studies focused on evaluating their diagnostic effectiveness compared to DXA measurements on the axial reference sites (lumbar spine and proximal femur) obtained contradictory results (Breban et al. 2010; Dane et al. 2008; El Maghraoui et al. 2009; Iida et al. 2010; Kwok et al. 2012; Liu et al. 2012; Moayyeri et al. 2012; Schnitzer et al. 2012; Stewart et al. 2006; Trimpou et al. 2010). However, work in this field is ongoing and interesting experimental results are being continuously published, including, for instance, encouraging multi-site measurements of apparent integrated backscatter (AIB) and the integrated reflection coefficient (Karjalainen et al. 2012); extended clinical validations of calcaneal measurements of AIB and spectral centroid shift (Jiang et al. 2014); and improvements of precision in the measurement of speed of sound (SOS) at the heel (Daugschies et al. 2015).

Nevertheless, despite the number of available papers, the International Society for Clinical Densitometry (ISCD) has restricted the actual diagnostic usefulness of QUS methods to validated calcaneal devices used on people aged 65 and older, only in combination with clinical risk factor assessment and only in order to identify those patients at very low risk of fractures who therefore do not require further investigations (ISCD 2013).

As a consequence, clinical routine management of osteoporosis is based on the combined evaluation of axial DXA outcomes and clinical risk factors (*i.e.,* previous fragility fracture, parent with a fractured hip, smoking, use of glucocorticoids, rheumatoid arthritis, causes of secondary osteoporosis, alcohol abuse) (Ferrari et al. 2012), resulting in high rates of under-diagnosis and under-treatment—it was estimated that approximately 75% of osteoporotic patients are not diagnosed and thus never receive appropriate treatment (Curtis and Safford 2012; Nguyen et al. 2004; van den Bergh et al. 2012). In fact, the best osteoporosis treatment would be the prevention of the disease, and of subsequent fractures, through a very early diagnosis, ideally when the patient presents a BMD reduction still in the osteopenic range, before becoming osteoporotic. In most cases, this condition can be effectively managed by targeted corrections to lifestyle habits (*e.g.,* increasing physical exercise, reducing alcohol consumption, *etc.*). Otherwise, a good therapy for osteoporosis should inhibit bone resorption and/or stimulate bone formation—an effective osteoporosis treatment significantly reduces fracture rate by 40–70% in a few months. The most commonly used therapeutic agents include estrogens, selective estrogen receptor modulators, bisphosphonates, agents derived from the parathyroid hormone, strontium ranelate and denosumab (Muratore et al. 2014).

To improve this situation, attention has recently been turned to the investigation of ultrasound (US) approaches for osteoporosis diagnosis directly applicable on lumbar spine and/or proximal femur (Barkmann et al. 2010; Conversano et al. 2015; Karjalainen et al. 2012). In particular, the femur has become the target of many experimental works focused on the translation of peripheral QUS measurement to the proximal femur (Barkmann et al. 2007, 2008a, 2008b, 2010; Dencks et al. 2008; Grimal et al. 2013; Haiat et al. 2005; Karjalainen et al. 2012; Padilla et al. 2008). The most significant clinical results were obtained by Barkmann et al. (2010), employing a "through transmission" US approach for measuring SOS and broadband ultrasound attenuation (BUA), and by Karjalainen et al. (2012), who used a backscatter approach to measure AIB and integrated reflection coefficient.

In this context, a different US methodology for osteoporosis diagnosis on lumbar spine has been recently introduced by our research group, together with a novel diagnostic parameter, the Osteoporosis Score (O.S.) (Conversano et al. 2015). The basic idea underlying this approach is that unfiltered radiofrequency (RF) signals, acquired during an echographic scan of the investigated bone district, can be employed to assess the bone health status through advanced comparisons with previously derived reference spectral models of healthy and osteoporotic bones. Moreover, the implemented RF signal analysis method is natively integrated with echographic imaging, which is needed for both the identification of the region of interest (ROI) for diagnostic calculations and the simultaneous acquisition of several RF signals for each frame, in order to provide a solid and reliable statistical basis for subsequent spectral analyses. This close integration in the processing of echographic images and "raw" unfiltered RF signals has been recently defined as the "echosound approach" (Casciaro et al. 2015).

The aim of the present work was to extend the application of this new methodology to the femoral neck, investigating its clinical feasibility and diagnostic

accuracy in both patient classification and BMD estimation. It is important to note that the effectiveness of O.S.-based diagnoses on the femoral neck was not obvious *a priori*, since, given the numerous differences between lumbar vertebrae and proximal femur, both the acquisition protocol and the image segmentation procedure were entirely redesigned. Important modifications and specific parameter optimizations were also required for the signal processing steps, which in turn involve the calculation of reference spectral models. Therefore, the resulting approach presents several novelties whose effectiveness had to be verified through dedicated studies. Furthermore, the specific clinical benefits that could be brought in by its possible routine adoption are discussed in the present work.

METHODS

Patients

The study was conducted at the Operative Unit of Rheumatology of Galateo Hospital in San Cesario di Lecce in Lecce, Italy, and included all consecutive female patients who fulfilled the following enrollment criteria: Caucasian ethnicity, aged 61–70 y, body mass index (BMI) < 25 kg/m^2, medical prescription for a hip DXA, absence of previous femoral fractures and absence of significant deambulation impairment.

Since the main objective of the present paper was to assess the feasibility and the effectiveness of O.S. calculation on femoral neck, the enrollment criteria were established following the same approach of the paper that introduced the O.S. calculation on lumbar vertebrae (Conversano et al. 2015). Apart from setting the target age range at 61–70 y (because this interval includes the majority of women referred for a hip DXA), the recruitment strategy of the present study was the same as that of our previous work. In particular, the enrollment of only female patients was motivated by practical considerations—typically, over 90% of the patients referred for a DXA examination at the aforementioned hospital are women; therefore, we decided to target this investigation to only female patients in order to ensure the recruitment of a statistically significant number of patients in a reasonable time interval.

A total of 377 patients were recruited in 9 mo. Table 1 reports the average anthropometric characteristics of the enrolled patients and also summarizes the patient distribution among different sub-groups. In fact, the enrolled patients were sub-divided between a reference database and a study population, which was in turn composed of two different groups: one destined to repeatability assessments and the other to accuracy evaluations (Table 1). The patients were grouped into two 5-y intervals according to their age—61–65 y (Group A) and 66–70 y (Group B). For each group, the first 100 patients were included in the reference database, while the remaining individuals represented the study population (Table 1). The size of the reference database was established according to widely adopted rules (Engelke and Gluer 2006; Hou et al. 2008).

All enrolled patients underwent both a femoral DXA and an echographic scan of proximal femur, as detailed in the proceeding sections.

The study protocol was approved by the hospital ethics review board and all patients gave informed consent.

DXA measurements

DXA scans were performed according to standard procedures used for hip investigations. During the examination, the patient's femur was straight on the table, in a position such that the shaft was parallel to the vertical edge of the obtained image, and with 15°–25° of internal

Table 1. Distribution of enrolled patients among reference database and study population for each age range considered (average anthropometric details for each group are also included). Values are expressed as mean ± standard deviation

		Study population		
Age range (y)	Reference database (n)	Repeatability measurements (n)	Accuracy measurements (n)	Total
61–65 (Group A)	100	15 (intra-operator)	70	185
Age (y)	62.9 ± 1.4	62.7 ± 1.2	62.8 ± 1.4	62.8 ± 1.4
BMI (kg/m^2)	22.7 ± 1.6	23.0 ± 1.4	23.1 ± 1.4	22.9 ± 1.5
Height (cm)	159.8 ± 6.0	160.6 ± 5.6	159.6 ± 5.5	159.8 ± 5.8
Weight (kg)	58.0 ± 5.5	59.7 ± 5.1	59.0 ± 5.8	58.5 ± 5.6
66–70 (Group B)	100	30 (inter-operator)	62	192
Age (y)	68.1 ± 1.5	68.2 ± 1.1	68.1 ± 1.3	68.1 ± 1.4
BMI (kg/m^2)	22.7 ± 1.9	22.3 ± 1.4	23.0 ± 1.8	22.7 ± 1.8
Height (cm)	159.9 ± 6.5	159.6 ± 4.9	159.2 ± 5.5	159.6 ± 6.0
Weight (kg)	58.1 ± 6.7	59.5 ± 5.7	58.3 ± 6.0	58.4 ± 6.3
Total	200	45	132	377

BMI = body mass index.

rotation, which was achieved by using a dedicated positioning device.

The adopted DXA scanner was a Discovery W (QDR Series, Hologic, Waltham, MA, USA), which performed a BMD measurement over the femoral neck, expressing the resulting value in g/cm^2. Moreover, for each patient, Hologic software also provided the conversion of the BMD in the corresponding T-score value, defined as the number of SDs from the peak BMD of young women found in the standard Hologic reference database for Caucasian women. According to the commonly used World Health Organization definitions, patients were classified as "osteoporotic" if their T-score was ≤ -2.5, "osteopenic" if $-2.5 <$ T-score < -1.0 or "healthy" if T-score ≥ -1.0.

DXA equipment underwent daily quality control and regular maintenance for the entire study period. Specific attention was paid to ensure adherence to correctness criteria for DXA scanning and data analysis as reported in ISCD 2013 guidelines (ISCD 2013) and also in the user manual of the adopted DXA system (Hologic 2000). In particular, all DXA acquisitions were double-checked by two different experienced operators in order to avoid possible errors associated with patient positioning, data analysis, presence of artifacts, and demographic characteristics, as identified by recent literature (Messina et al. 2015). In fact, such errors could provide wrong BMD measurements, potentially resulting in inappropriate diagnostic classifications.

US acquisitions

US scans of femoral neck were performed employing an echographic device (Echo Blaster 128, Telemed Medical Systems, Milan, Italy), equipped with a convex transducer (C3.5/60/128 Z, Telemed Medical Systems) operating at the nominal frequency of 3.5 MHz. The transducer had a -20 dB bandwidth from 1.4 MHz to 5.8 MHz and its center frequency was chosen as the best compromise between the highest possible spatial resolution and a suitable US penetration into femoral neck tissue. In theory, a higher frequency (corresponding to a better spatial resolution) could provide more accurate results, but the US frequencies higher than 3.5 MHz available on convex echographic transducers did not provide a satisfactory signal-to-noise ratio (SNR) for signals backscattered from the internal part of the femoral neck. For the entire study duration, the correct calibration of the US transducer spectrum was verified monthly to ensure no appreciable differences with respect to the initial calibration provided by the manufacturer. These experimental verifications were carried out in a water tank by employing a polyvinylidene fluoride membrane hydrophone (Precision Acoustics Ltd, Dorchester, UK).

The echographic device was provided in a modified research configuration allowing the acquisition of both the conventional B-mode echographic images and the corresponding unprocessed "raw" RF signals. These signals underwent the following pre-processing steps, performed through a dedicated board (Echolight srl, Lecce, Italy): 1-kHz high-pass filtering, 18-dB amplification and analogue-to-digital conversion (40 MS/sec, 16 bits). Finally, the obtained digital signals were stored on a personal computer hard disk for offline data analysis.

Each patient underwent a femoral neck scan performed through the following protocol. The operator initially placed the convex probe on the patient proximal femur, orienting the transducer to visualize the transversal femoral neck section. The probe was then rotated about 90°, until the typical proximal femur profile, including femoral head, neck and trochanter interfaces, was visible—this defined the start of the US data acquisition, which lasted for about 40 s and generated 50 frames of RF data (frame rate $\sim$1.25 fps). During data acquisition, the operator slightly tilted the US probe in a very slow way, always keeping the femoral neck interface in the horizontal position, aligned with the focus reference line and in the central part of the echographic image. The tilting operation was performed to acquire different sections of the longitudinal proximal femur interface.

For all the performed acquisitions, transducer focus and scan depth were set to 4.5 cm and 9 cm, respectively, in order to have femoral neck interface in the beam focal zone and in the central part of the image. Other echographic settings were power = 45%; mechanical index = 0.4; gain = 0 dB; and linear time gain compensation.

US acquisitions on patients to be included in the reference database (n = 200) were performed by a sonographer experienced in musculoskeletal US investigations, in order to ensure that the database included only high-quality acquisitions. US acquisitions on patients to be included in the study population were, instead, performed with the double aim of assessing the repeatability of the obtained results and evaluating the accuracy of the adopted approach on data sets acquired by inexperienced operators. Therefore, patients to be included in the study population were managed in the following way: the first 15 patients enrolled in Group A underwent three consecutive US investigations performed by the same experienced operator, with patient repositioning between the scans, and the corresponding data were used to assess intra-operator repeatability; the first 30 patients enrolled in Group B underwent two consecutive US investigations performed by two different experienced operators, with patient repositioning between the scans, and the corresponding data were used to assess inter-operator repeatability; all the subsequent patients enrolled in both

Group A and Group B underwent a single US examination, alternately performed by one of two inexperienced operators who had previously received only a 3-h specific training session, and the corresponding data were used for accuracy assessments.

US data analysis

Offline analysis of US data was conducted through a customized implementation of a fully automatic algorithm that we have recently employed for the estimation of lumbar spine BMD (Conversano et al. 2015).

In the present work, we implemented a different version of the previous algorithm, in which a novel segmentation method was employed to achieve a highly selective automatic identification of femoral neck interfaces within the sequence of acquired echographic images, and dedicated models of osteoporotic and healthy femoral necks were used in the RF signal analysis for O.S. calculation.

Calculation of reference models. Figure 1 shows the block diagram of the aforementioned procedure used for the calculation of reference model spectra from the database patients for both 5-y age ranges (61–65 y and 66–70 y).

For each age range, the US data sets of all database patients classified as "osteoporotic" by DXA were processed to obtain the corresponding "osteoporotic model." This procedure was presented in our previous paper referring to the lumbar spine application (Conversano et al. 2015), but some parameters had to be adapted to the peculiarities of the femoral neck case. Therefore, we summarize herein the main processing steps, detailing just the settings that characterize femoral neck analysis:

1. For the *k*th osteoporotic patient considered, five different femoral neck images were manually selected among those acquired during the tilting movement of the probe, with each image including E_i ($i = 1, \ldots, 5$) echographic lines crossing the femoral neck interface.
2. For the *j*th echographic line crossing the femoral neck surface in the *i*th selected image ($j = 1, \ldots, E_i$; $i = 1, \ldots, 5$), a segment of the corresponding RF signal was selected using a 150-point Hamming window located immediately after the echo from the femoral neck surface. The *i*th ROI was therefore identified by the E_i signal segments of the *i*th frame. Actually, the placement of the starting point of the Hamming window had been carefully optimized through dedicated preliminary studies, aimed at, in particular, avoiding as much as possible the contamination of the scattering from the bone interior by the ring-down effect from bone surface echo. Our final choice was to define the window starting point as the first position in which the signal envelope was less than 15% of its peak value corresponding to the bone interface. In fact, we initially tried to place the selection window at a fixed delay after the peak amplitude to ensure enough distance from bone surface echo, but in most cases the final part of the target RF signal segment was characterized by an SNR that was too low. We then decided to reduce the delay after the peak as much as it was necessary to always have a suitable SNR in the target region, but, because of the variable properties of the cortical layer (*e.g.,* thickness), in several cases it was evident that we were too close to the bone surface, whose signal was contaminating the one coming from the underlying region. Therefore, we decided to optimize the window placement with respect to a signal amplitude threshold, and the 15% of the peak value resulted as the best compromise between the two requirements of minimum distance from bone interface and suitable SNR in the target region.
3. The fast Fourier transform power spectrum was calculated for each obtained *j*th RF signal segment belonging to the *i*th selected ROI and compensated to take into account the probe transfer function. The obtained spectrum was then normalized with respect to its maximum value and labeled as $P_{Norm_{ij}}(f)$.
4. The spectrum representing the *i*th ROI (called "ROI mean spectrum" $\overline{P_i}(f)$) was obtained through the same iterative procedure employed in our previous work to obtain the "vertebra mean spectrum." Typically, the ROI was composed of about 60 echographic lines, 35% of which contributed to the final $\overline{P_i}(f)$ at the end of the iterative selection.
5. The five resulting $\overline{P_i}(f)$ spectra were averaged to obtain a spectrum that was normalized with respect to its maximum value and assumed as representative of the *k*th osteoporotic patient considered ("osteoporotic patient mean spectrum" $\overline{PS}_{Ost_k}(f)$).

In an analogous manner, for each *z*th "healthy" patient belonging to the reference database, the related US data were analyzed to obtain a representative, patient-dependent "healthy patient mean spectrum" $\overline{PS}_{Heal_z}(f)$. In this way, each "osteoporotic" or "healthy" patient was represented by a single spectrum. For each 5-y age interval, we then considered, one by one, the "osteoporotic patient mean spectra" $\overline{PS}_{Ost_k}(f)$ and calculated the Pearson correlation coefficient r_{kz} between the considered $\overline{PS}_{Ost_k}(f)$ and each $\overline{PS}_{Heal_z}(f)$ belonging to the same age interval. Among all the calculated r_{kz} values, we selected, for each 5-y age interval, the lowest three r_{kz} values, which identified the three pairs of spectra showing the minimum mutual correlation (selected r_{kz} values were in the range 0.87–0.89 for Group A and 0.85–0.88 for Group B). These three pairs of spectra were used to identify a final pair of "osteoporotic" and "healthy" model spectra, each

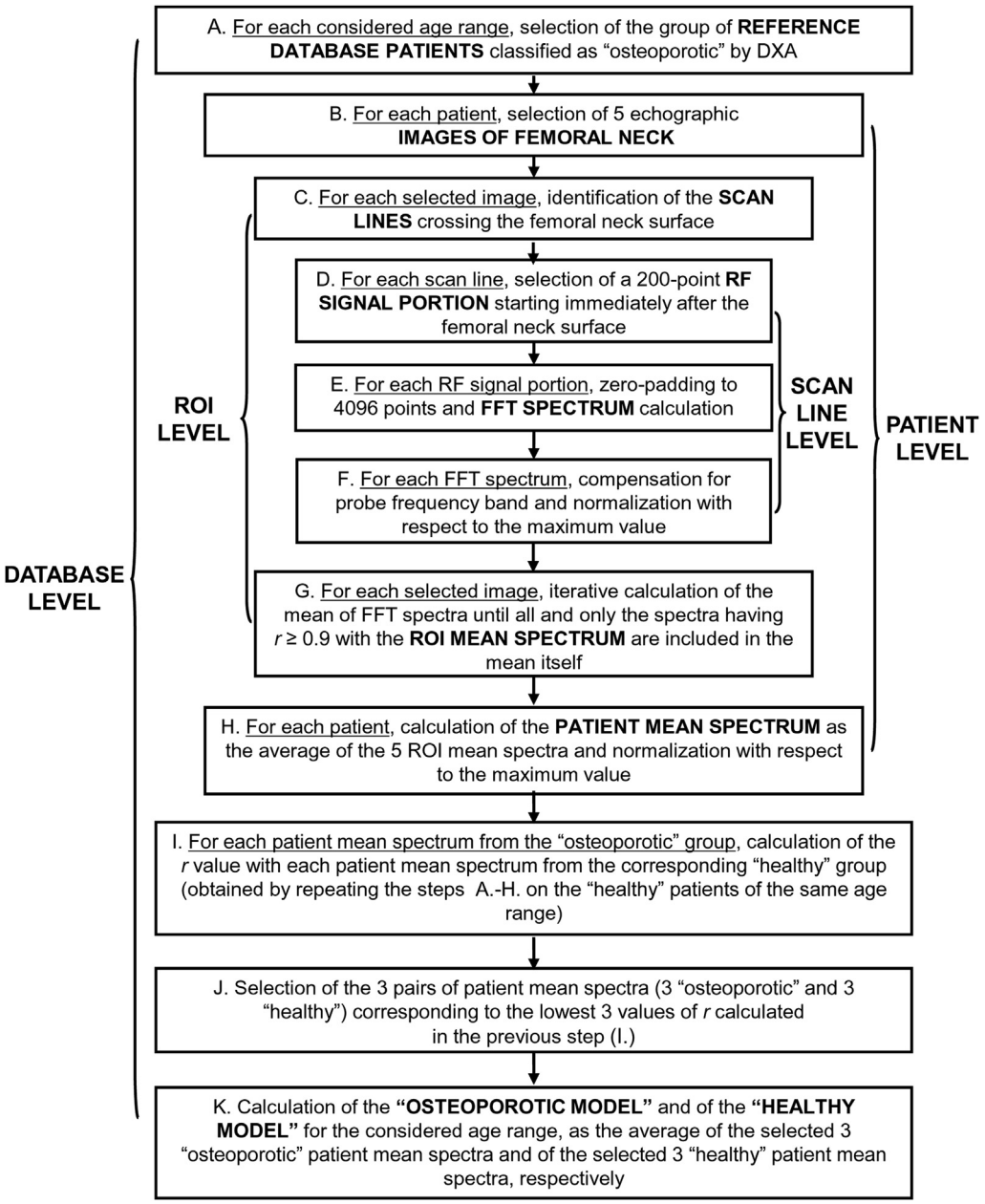

Fig. 1. Schematic illustration of the data processing steps implemented to calculate the pairs of reference model spectra for each 5-y age range. DXA = dual X-ray absorptiometry; FFT = fast Fourier transform; RF = radiofrequency; ROI = region of interest.

obtained as the average of the three corresponding "osteoporotic/healthy patient mean spectra" normalized with respect to its maximum value.

The decision to consider just three pairs of spectra for each age interval was based on the results obtained in our previous paper referring to the lumbar application (Conversano et al. 2015), in which we varied the number of spectra included in each model from one to five and verified that three was the number providing the best diagnostic accuracy. In the present work, we directly started with the employment of three patient spectra in each model and then verified *a posteriori* that both higher and lower numbers of patient spectra worsened the final diagnostic performance. Analogously to the lumbar case, this finding can be explained by the fact that the inclusion of a lower number of patients in the model gives an excessive weight to the peculiar features of those specific patients, whereas, on the contrary, the employment of a higher number of patients augments the similarity between the "healthy" model and the corresponding "osteoporotic"

one, producing a drop in the discrimination power. The inclusion of the spectra from three patients in each model was the best trade-off between the two phenomena.

The next sub-section describes the typical operations performed by the algorithm on the US data acquired on a generic patient in order to automatically identify the femoral neck interface and the corresponding ROI for the calculation of the O.S. value through the appropriate reference models.

Automatic identification of femoral neck interface and O.S. calculation. For each considered *k*th patient, the algorithm performed the automatic identification of femoral neck interfaces within the sequence of acquired echographic images. This is achieved by performing the following steps (1–10) on each acquired frame according to the related indicated criteria (the corresponding block diagram is shown in Fig. 2). Figure 3 illustrates how the following processing steps are applied to the image frame reported in Figure 3a in order to automatically identify the proximal femur interface, detecting the profile tracts corresponding to head, neck and trochanter regions and selecting the proper femoral neck ROI. The single image processing steps are simply listed here together with the references to different parts of Figure 3 to provide a qualitative visualization of the results of each step (quantitative details on the corresponding operations are reported in the Appendix):

1. Image pre-processing: noise removal around the femoral interface and convex image construction (Fig. 3b).
2. Gray-level thresholding and conversion to a binary image (Fig. 3c).
3. Median filter application (Fig. 3d).
4. Horizontal axis image dilation (Fig. 3e).
5. Asymmetric morphologic closing, consisting of an *m*-pixel dilation followed by *n*-pixel erosion (with $m < n$) (Fig. 3f).
6. Morphologic evaluations, including (i) selective thresholding based on geometric distribution of white pixel clusters and (ii) proximal femur profile fitting function (Fig. 3g).
7. ROI identification and classification of the proximal femur interface in head, neck and trochanter profiles (Fig. 3h).
8. Anti-convex image construction for the identification of the echographic lines crossing the femoral neck interface (Fig. 3i).
9. Spectral validation (Fig. 3j): non-femoral interfaces are filtered out if no correlation is found with either of the two appropriate reference models (see Appendix for further details).
10. New analysis is started on the next frame.

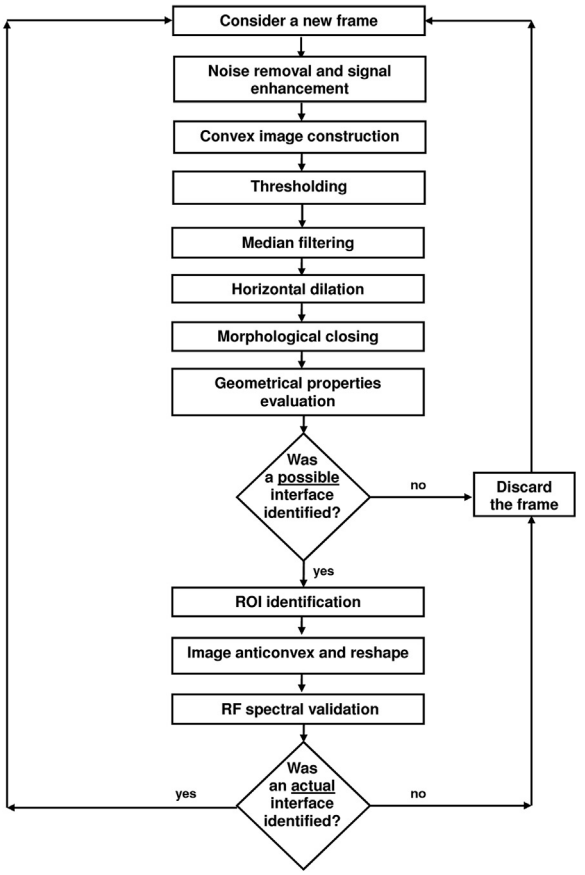

Fig. 2. Schematic illustration of the automatic identification of femoral neck interfaces within the sequence of US data frames acquired on each patient. RF = radiofrequency; ROI = region of interest.

Once the described process has been iterated until all the frames belonging to the US data set of the *k*th patient have been analyzed, a number n_k of femoral neck interfaces has been identified; if $n_k < 5$, the data set was discarded and O.S. value was not calculated. Otherwise, the following diagnostic calculations were performed on each RF spectrum of the identified ROIs, selected as in the case of model construction (150-point Hamming-windowed signal portions starting immediately after the echo from the femoral neck surface, when the amplitude of the RF signal envelope reached 15% of its peak value).

The *j*th RF spectrum of the *i*th ROI ($P_{Norm_{ij}}(f)$) was classified as "osteoporotic" if the value of its Pearson correlation coefficient $r_{Ost_{ij}}$ with the age-matched osteoporotic model spectrum was higher than the corresponding $r_{Heal_{ij}}$ value with the related healthy model spectrum; otherwise it was classified as "healthy". The number of spectra classified as "osteoporotic" for the *i*th ROI was labeled as $E_{i_{ost}}$ (with $E_{i_{ost}} \leq E_i$). Then, the

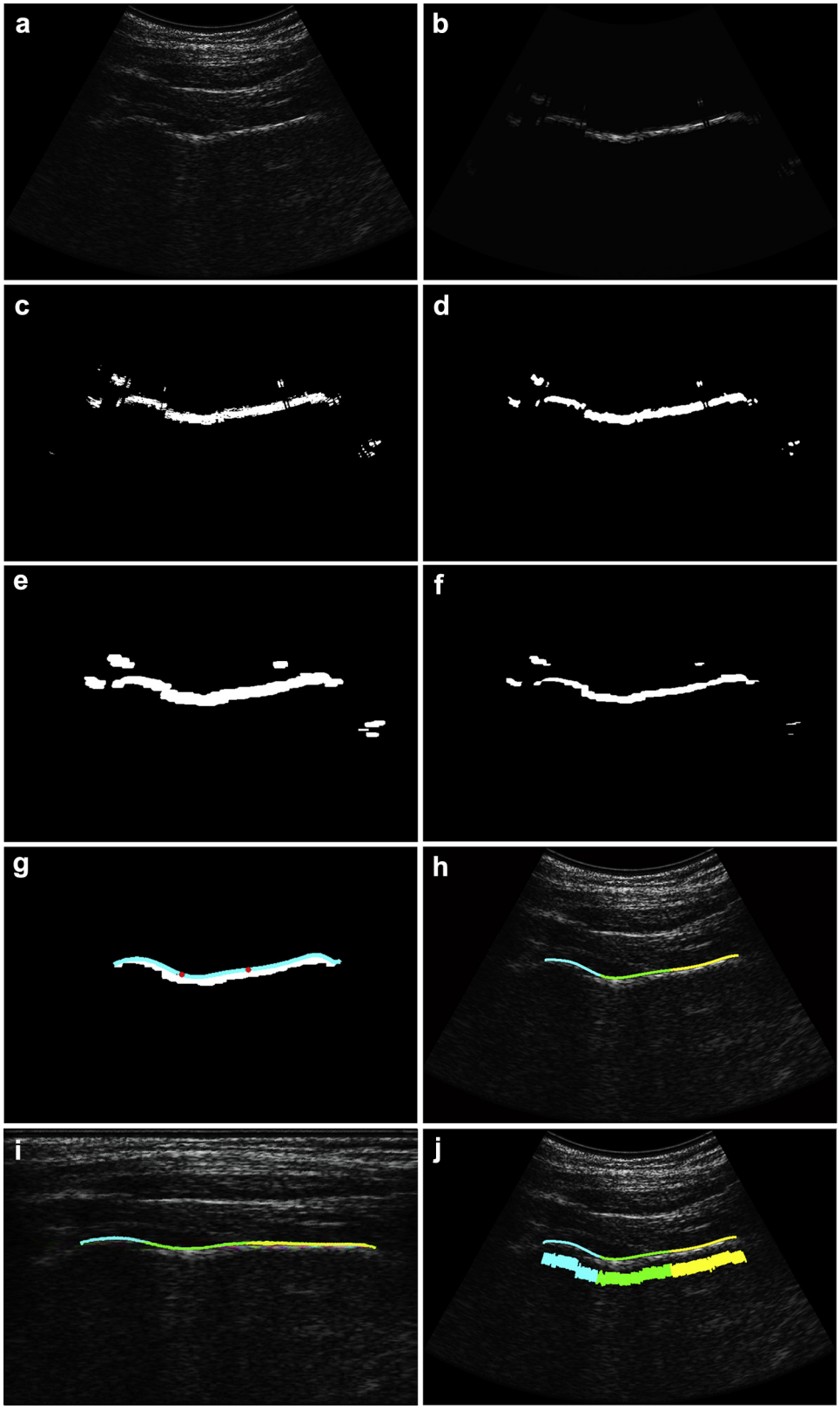

Fig. 3. Application of the processing steps for automatic femoral neck interface identification to a typical echographic image frame: (a) original frame; (b) signal noise removal around the femoral interface and convex image construction; (c) gray-level thresholding and conversion to binary image; (d) median filter application; (e) horizontal axis image dilation; (f) asymmetric morphologic closing; (g) morphologic evaluations; (h) ROI identification and classification of the femoral interface, discriminating between femoral head (light blue), femoral neck (green) and trochanter (yellow); (i) anti-convex image construction; (j) final identification of femoral neck ROI (green), below the previously identified bone interface. (For the sake of having a simple visual feedback, image data are always shown in a rearranged convex image; see the Appendix for details).

percentage of analyzed spectra classified as "osteoporotic" represented the O.S. of the considered ROI:

$$O.S._i = \frac{E_{i_{ost}}}{E_i} \cdot 100 \quad (1)$$

The same evaluations were repeated for each identified ROI and the final O.S. related to the kth analyzed patient was obtained as the average of the O.S. values calculated for single ROIs:

$$O.S._k = \frac{\sum_{i=1}^{n_k} O.S._i}{n_k} \quad (2)$$

where n_k represents the number of ROIs identified in the US data set of the kth patient.

The described procedure was used to assess the effectiveness of the previously calculated reference models through dedicated tests on study population patient data sets, aiming at the evaluation of the method repeatability (*i.e.,* intra- and inter-operator variability) and at the assessment of its diagnostic power with respect to DXA, assumed as a gold standard reference.

The algorithm programming environment was MATLAB R2012a (The MathWorks, Natick, MA, USA). The US data analysis of a single patient, including automatic identification of target bone interfaces and calculation of O.S. values, was completed within 90 s on a personal computer equipped with an Intel i7 Core™ i7-3610QM processor at 2.3 GHz, 8 GB of RAM, 64 bits.

Statistical analysis

Preliminary reference model test and establishment of O.S. diagnostic thresholds. For each age interval considered, the finally obtained pair of reference spectral models were preliminarily tested on the database patients belonging to the same age range but not included in the final reference models (n = 94 for each age range). The O.S. value was calculated for each patient and two specific O.S. diagnostic thresholds were automatically determined for each age interval, as the values providing the best agreement between DXA-based patient classification as osteoporotic, osteopenic or healthy, and a corresponding classification based on O.S. values. The achieved level of agreement was quantified through the calculation of accuracy (*i.e.,* correctly classified patients/analyzed patients).

The automatic determination of the O.S. diagnostic thresholds for a given age interval was achieved through the following procedure:

- Initially, to determine the healthy/osteopenic threshold, only the patients classified by DXA as "healthy" or "osteopenic" were considered.
- All corresponding O.S. values were put in ascending order and the initial tentative threshold t was fixed equal to the lowest O.S. value.
- All patients that had O.S. $\leq t$ were temporarily classified as "healthy" and all patients that had O.S. $> t$ were temporarily classified as "osteopenic."
- The percentage of diagnostic agreement between the O.S.-based patient classification performed in the previous step and the DXA classification was calculated.
- The value t was incremented by 0.1 and the previous two steps were repeated.
- The procedure was iterated until t reached the highest O.S. value.
- The final diagnostic threshold to discriminate between "healthy" and "osteopenic" patients was fixed equal to the t value providing the maximum diagnostic agreement (in cases of more than one value providing the same diagnostic accuracy, the lowest one was chosen).

The osteopenic/osteoporotic threshold was determined in a similar way, starting from the highest O.S. value available among the patients classified as either "osteopenic" or "osteoporotic" by DXA and decreasing the tentative threshold value by 0.1 at each iteration.

Intra-operator variability. Intra-operator variability was assessed in terms of "short-term precision" as defined by Engelke and Gluer (2006) using the data acquired on the first 15 patients of Group A included in the study population, who underwent three consecutive US examinations performed by the same operator.

For each considered patient, the method precision was calculated, as recommended by the ISCD, through specific statistical parameters, such as the root mean square (RMS) average of the standard deviation (RMS-SD) of repeated measurements, the RMS coefficient of variation (RMS-CV) and the least significant change (LSC) for a 95% confidence level. To facilitate the interpretation of numerical data and the comparison with previously published articles, all the results related to precision and repeatability of the proposed methods were expressed in terms of BMD (g/cm^2) by employing a linear regression approach to estimate femoral neck BMD from O.S. values, as described later in the text.

Inter-operator variability. Inter-operator variability was assessed based on the data acquired from the first 30 patients of Group B included in the study population, who underwent two consecutive US investigations performed by two different operators. Quantifications were carried out in a similar manner to those related to intra-operator variability. For each considered patient, the SD of the repeated US measurements was calculated and inter-operator variability was then expressed as RMS-

SD and RMS-CV. LSC for a 95% confidence level and the inter-observer Cohen's kappa (Cohen 1960) were also calculated.

Accuracy of the novel US method versus DXA. The remaining study population patients, not used for repeatability analyses, were employed for the assessment of the diagnostic accuracy of the proposed US method compared to DXA measurements of BMD, assumed as a gold standard reference. For each age interval considered, the acquired US data were analyzed employing the age-matched pair of reference models and the corresponding O.S. diagnostic thresholds established during the preliminary test on database patients.

On the basis of DXA diagnosis, in each age range considered, the study population patients underwent a further diagnostic category sub-division, distinguishing among osteoporotic, osteopenic and healthy patients. This was necessary to assess the effectiveness of O.S. values in terms of patient discrimination power. Mean and SD of O.S. values were calculated for each patient group and the statistical significance of the differences between the calculated mean values was evaluated using an unpaired two-tailed Student's *t*-test.

The degree of agreement in patient classification between DXA and the proposed US approach was assessed through the calculation of accuracy and Cohen's kappa, whereas the effectiveness of a linear regression approach employed to estimate femoral neck BMD from O.S. values was quantified through the Pearson correlation coefficient (r), the coefficient of determination (r^2) and the root mean square error (RMSE). Furthermore, the agreement between DXA-measured BMDs and US-derived values was also evaluated as recommended by Altman and Bland (1983).

RESULTS

Preliminary test of reference models

The obtained pairs of reference models are shown in Figure 4 for both the considered age intervals. As expected, these two pairs of spectral models, built up for femoral neck analyses, qualitatively confirmed the characteristics of those obtained in our previous work for lumbar spine (Conversano et al. 2015). Because of the higher compactness and lower porosity compared to osteoporotic ones, healthy bones acted as a kind of low-pass filter on the backscattered US as the frequency value increased.

To obtain a preliminary measure of the diagnostic effectiveness of the calculated spectral models for both the age intervals, they were tested on the database patients not included in the models themselves for computation of O.S. values and diagnostic classification. The percentage of reference database patients that received concordant

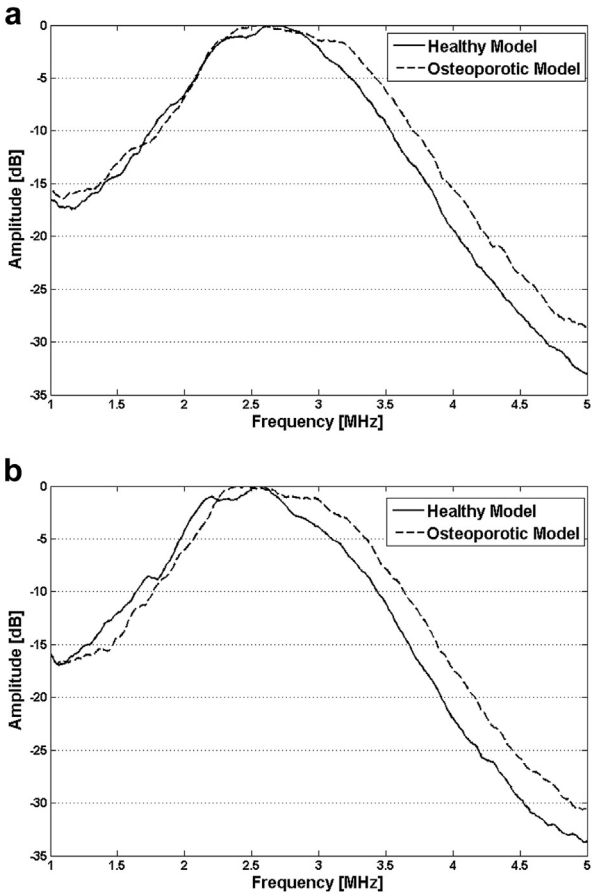

Fig. 4. Finally obtained pairs of reference models for both the considered age intervals: (a) healthy model and osteoporotic model for patients aged 61–65 y (Group A); (b) healthy model and osteoporotic model for patients aged 66–70 y (Group B). Each model included data from three patients, as described in the text.

diagnoses from O.S. and DXA was 96.8% in Group A and 95.7% in Group B.

The automatically determined O.S. thresholds that provided the reported accuracy levels were the following: patients aged 61–65 y (Group A) were classified as "osteoporotic" if O.S. $\geq$ 55.0; "osteopenic" if 46.3 $<$ O.S. $<$ 55.0; or "healthy" if O.S. $\leq$ 46.3. Patients aged 66–70 y (Group B) were classified as "osteoporotic" if O.S. $\geq$ 53.6; "osteopenic" if 45.7 $<$ O.S. $<$ 53.6; or "healthy" if O.S. $\leq$ 45.7.

Precision and repeatability

Intra-operator variability and inter-operator repeatability were specifically quantified to evaluate the precision and the reproducibility of the proposed approach.

Precision (*i.e.,* intra-operator variability), expressed as RMS-SD, was 0.003 g/cm^2 (RMS-CV = 0.27%) and the corresponding LSC for a 95% confidence level was 0.007 g/cm^2 (RMS-CV = 0.75%). Analogous

Table 2. Inter-operator variability of the proposed ultrasound-based method. Cross-tabulation of the diagnostic classifications of 30 patients independently investigated by two different operators

	Operator A			
	Healthy	Osteopenic	Osteoporotic	Total
Operator B				
Healthy	6	1	0	7
Osteopenic	0	16	1	17
Osteoporotic	0	1	5	6
Total	6	18	6	30

calculations were performed to assess inter-operator variability, producing the following results: RMS-SD = 0.004 g/cm^2 (RMS-CV = 0.36%) and LSC = 0.010 g/cm^2 (RMS-CV = 0.99%).

The degree of inter-observer correlation was also assessed through Cohen's kappa. The obtained results gave $k = 0.826$ ($p < 0.0001$), confirming the optimal agreement between the operators. The corresponding data are also reported in Table 2.

Accuracy of the US method compared with DXA

To evaluate the actual accuracy of the proposed method, diagnostic measurements were also conducted on a separate group of patients included in the study population (see Table 1), whose US scans were performed by inexperienced operators and whose corresponding data sets were not used for model derivation nor for repeatability assessments. All the US acquisitions resulted of suitable quality for actual diagnostic calculations (*i.e.*, $n_k \geq 5$ for every considered kth patient), documenting a significant user-friendliness of the adopted US approach even for inexperienced operators.

Table 3 reports the distribution of the study population patients enrolled for accuracy measurements according to age range and DXA diagnostic classification, also showing, for each obtained patient group, the average O.S. value together with the statistical significance of the difference with respect to the corresponding average O.S. values calculated for the patients belonging to different DXA-based groups in the same age range.

From Table 3, it is also possible to observe how, in each considered age interval, the average O.S. values were correctly ordered with respect to the DXA diagnostic categories, since the highest average O.S. value was always associated to the osteoporotic patients and the lowest average O.S. value was always found in correspondence of the healthy patients ($p < 0.0001$ for all the considered O.S. value differences).

Table 4 summarizes the level agreement between O.S.-based and DXA diagnoses, as assessed through the computation of accuracy and Cohen's kappa values. Accuracy was above 93% for both the age groups and the corresponding Cohen's k values were well above 0.85, therefore documenting an excellent diagnostic agreement between the two techniques.

To estimate femoral neck BMD from O.S. values, a linear regression approach was also used. In each age range, an appreciable and statistically significant correlation was found between DXA-measured BMD values and the corresponding US estimates: $r = 0.89$ in Group A ($p < 0.001$) and $r = 0.85$ in Group B ($p < 0.001$). A further confirmation of the good diagnostic performance provided by the US method was given by the corresponding high values of the coefficient of determination ($r^2 = 0.79$ in Group A and $r^2 = 0.73$ in Group B) and the low residual errors (RMSE = 0.039 g/cm^2 [5.9%] in Group A and RMSE = 0.047 g/cm^2 [7.4%] in Group B). Figure 5 shows the scatterplot of BMD values provided by DXA and US measurements for all the patients belonging to the study population and employed for accuracy evaluations, together with the global value of Pearson correlation coefficient (r = 0.87, $p < 0.001$). US-estimated BMD values were calculated from the corresponding O.S. values through the following equation:

$$BMD_{US} = m \cdot O.S. + q \quad (3)$$

Where m and q were respectively −0.0192 and 1.6217 for Group A and −0.0211 and 1.6991 for Group B (these

Table 3. Results of DXA measurements and US-based O.S. calculations for study population patients used in accuracy measurements

		DXA			US
Age range (y)	Analyzed Patients (n)	DXA diagnosis	n	BMD (mean ± SD) [g/cm^2]	O.S. (mean ± SD)
61–65 (Group A)	70	Osteoporotic	10	0.523 ± 0.022	56.0 ± 1.2*
		Osteopenic	46	0.654 ± 0.042	50.1 ± 2.0*
		Healthy	14	0.773 ± 0.034	44.3 ± 2.2*
66–70 (Group B)	62	Osteoporotic	14	0.515 ± 0.032	55.8 ± 1.5*
		Osteopenic	40	0.649 ± 0.045	50.8 ± 2.4*
		Healthy	8	0.782 ± 0.059	43.8 ± 2.0*

BMD = bone mineral density; DXA = dual X-ray absorptiometry; O.S. = Osteoporosis Score; SD = standard deviation; US = ultrasound.
* $p < 0.0001$.

Table 4. Diagnostic effectiveness of US-based O.S. Agreement with DXA results for the study population patients used in accuracy measurements

Age range (y)	Analyzed Patients (n)	Agreement between US and DXA	
		Accuracy	Cohen's kappa
61–65 (Group A)	70	95.7%	0.914*
66–70 (Group B)	62	93.5%	0.879*
Total	132	94.7%	0.898*

DXA = dual X-ray absorptiometry; O.S. = Osteoporosis Score; US = ultrasound.
* $p < 0.0001$.

values were obtained from a linear regression applied on database patients' data). The corresponding Bland-Altman plot is reported in Figure 6. The overall average difference in BMD measurement (expressed as bias ± 2 SDs) was −0.002 ± 0.086 g/cm^2 and the single age intervals provided the following values: −0.001 ± 0.078 g/cm^2 for Group A and −0.003 ± 0.095 g/cm^2 for Group B.

DISCUSSION

This work demonstrated the feasibility of osteoporosis diagnosis on the femoral neck by employing the O.S., a novel US-based parameter that we have recently introduced for lumbar spine applications (Conversano et al. 2015). O.S. was designed to quantify the degree of similarity to US spectral models derived from patients with an osteoporotic BMD level (T-score ≤ −2.5) with respect to those derived from patients having a normal BMD (T-score ≥ −1.0). In fact, the original algorithm had been thought to perform diagnostic calculations on specific RF signal portions identified within the automatically segmented vertebrae, with the aim of quantifying the percentage of analyzed RF signals whose spectral features correlated better with those of an osteoporotic vertebral structure than with those of a healthy one. The obtained similarity degree was synthesized by the O.S. value, which was then used for patient diagnostic classification.

In the present work, as a preliminary step, a reference database consisting of 100 patients for each of the two considered 5-y age intervals was built, allowing the derivation of two pairs of model spectra (each composed by a "healthy" model and an "osteoporotic" one).

A first assessment of the diagnostic effectiveness of these models was conducted on the database patients not included in the models themselves (n = 188, acquisitions performed by an experienced operator). The direct comparison with DXA resulted in a diagnostic accuracy of 96.3% in patient classification. This impressive degree of diagnostic accuracy was also confirmed by the overall algorithm performance obtained on study population patients (total accuracy = 94.7%), whose acquisitions were performed by inexperienced operators (n = 132). Importantly, for each age range, the O.S. diagnostic thresholds to discriminate among osteoporotic, osteopenic and healthy patients were established during the preliminary test on database patients and were then kept constant to those values during the subsequent independent analysis of study population patients.

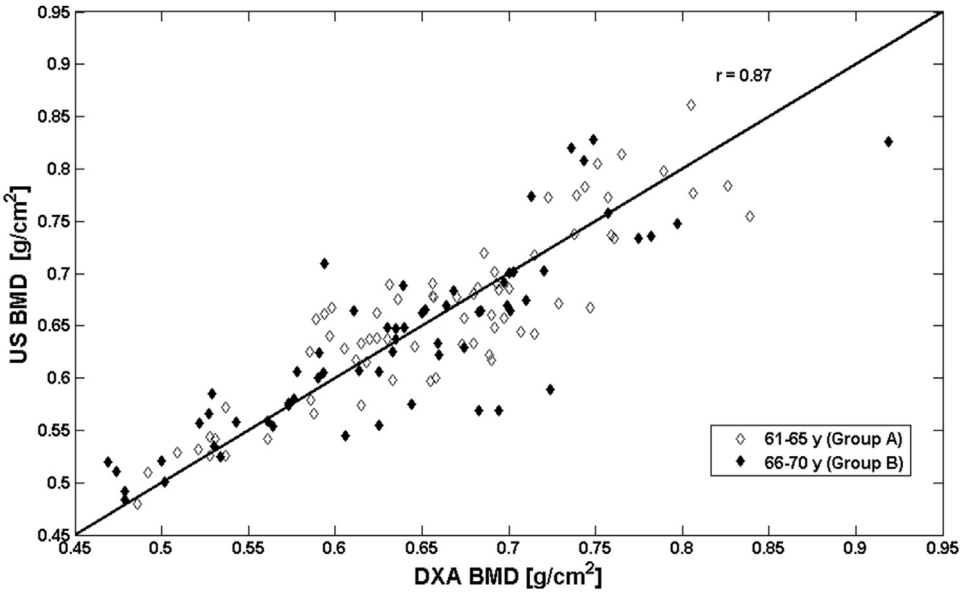

Fig. 5. Scatterplot of BMD values provided by DXA and US measurements for the study population patients used in accuracy measurements. The line of equality and the global Pearson correlation coefficient are also shown. BMD = bone mineral density; DXA = dual X-ray absorptiometry; US = ultrasound.

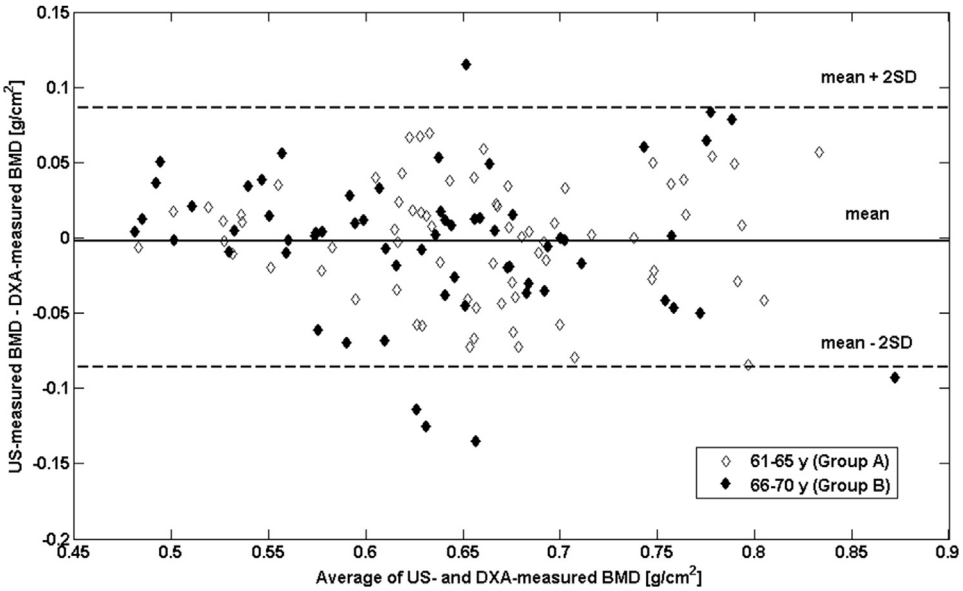

Fig. 6. Bland-Altman plot for comparison of US- and DXA-measured BMD values for the study population patients used in accuracy measurements. BMD = bone mineral density; DXA = dual X-ray absorptiometry; US = ultrasound.

Study population data sets were also employed to obtain O.S.-based estimates of BMD values, which resulted in being highly correlated with the corresponding DXA outcomes. The coefficient of determination (r^2) was 0.79 in Group A (n = 70) and 0.73 in Group B (n = 62), with RMSE being 5.9% and 7.4%, respectively, confirming the robustness of the reported results.

Referring to previous literature, we can observe that the highest and most significant correlations between US parameters measured at the proximal femur and site-matched BMD levels were reported for "through transmission" measurements on *ex vivo* specimens dissected from fresh cadavers (Barkmann et al. 2007; Dencks et al. 2007; Grimal et al. 2013; Haiat et al. 2005; Padilla et al. 2008). Among these studies, the one reporting the best results is probably the work by Haiat et al. (2005), documenting a strong correlation between SOS and BMD (r^2 = 0.86, n = 38), which was also improved by the use of a linear multiple regression using both BUA and SOS (r^2 = 0.95, n = 38). However, these impressive correlation levels have never been reported for analogous measurements performed *in vivo*. In fact, the most powerful results obtained from through transmission US measurements on the proximal femur of living patients are those obtained by Barkmann et al. (2010), who found r^2 = 0.72 (RMSE ~10%, n = 62) between DXA results

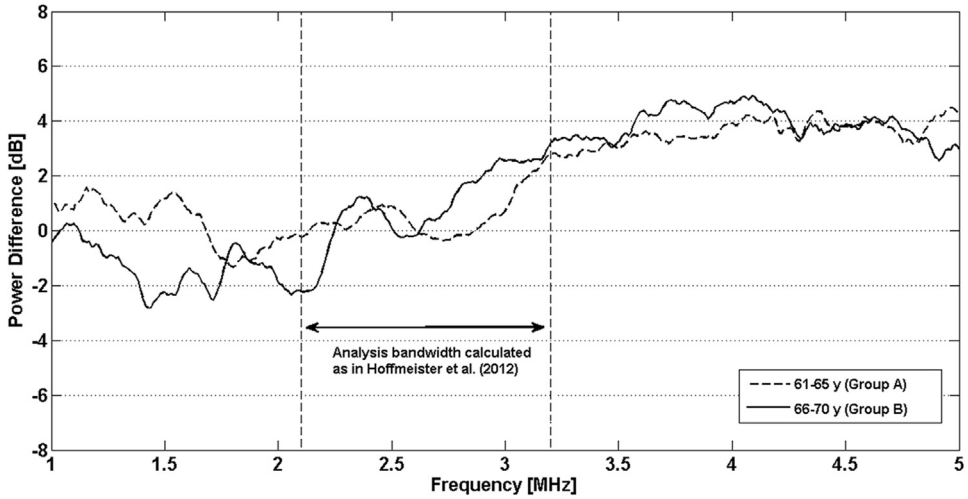

Fig. 7. Difference spectra obtained for each pair of the employed reference models.

and QUS-based estimates, and are therefore definitely comparable with the results obtained in the present study, although our backscatter approach was advantageous in terms of bulkiness and complexity of the employed device.

Alternative backscatter parameters measured on proximal femur resulted to be highly correlated with the corresponding BMD values when measured on conveniently shaped pure trabecular samples, but their effectiveness decreased when measured on intact proximal femurs or in living patients. Regarding *in vitro* tests, two of the most interesting works were reported by Hoffmeister et al. (2012), who found r^2 up to 0.90 between mean of backscatter difference spectrum and density of cube-shaped specimens extracted from human femoral heads (n = 22), and by Malo et al. (2014), who measured the same US parameter on intact human proximal femurs and found $r^2 = 0.45$ (n = 16) with site-matched assessments of femoral neck BMD. Regarding the *in vivo* measurements of US backscatter on proximal femur, the most promising approach until now published was based on the assessment of AIB on the femoral neck. Karjalainen et al. (2012) found $r^2 = 0.27$ on 26 elderly women with and without previous hip fractures.

The most peculiar characteristic of our proposed approach is the natively integrated processing of echographic images and unfiltered RF signals, both acquired during an echographic scan of the target bone district. The combined statistical analysis is also facilitated by the use of a convex array transducer in place of the single-element US probes commonly used in reported studies, since this allows the fast acquisition of a significant number of B-mode images and corresponding RF signals without requiring neither time-consuming scan procedures nor complex image reconstruction algorithms.

A further characteristic feature of the adopted approach is the extreme ease of use, resulting from a simple and easy to learn acquisition procedure (inexperienced operators who had received only a 3-h dedicated training were able to perform US scans of suitable quality in 100% of cases) followed by fast and fully automated data processing, which eliminated possible residual sources of error related to operator experience and allowed the achievement of an objective diagnosis in about 2 min (40 s for echographic scan and about 90 s for data analysis).

Currently, US-based commercially available devices for osteoporosis diagnosis present some important limitations, because they are applicable only for peripheral bone districts and their degree of correlation with bone properties at axial sites was extremely variable (Breban et al. 2010; Dane et al. 2008; El Maghraoui et al. 2009; Iida et al. 2010; Kwok et al. 2012; Liu et al. 2012; Moayyeri et al. 2012; Schnitzer et al. 2012; Stewart et al. 2006; Trimpou et al. 2010). In particular, through a review of available papers reporting the correlation between a QUS parameter measured on a peripheral bone district and DXA-measured femoral neck BMD, we found that the highest coefficient of determination was reported by Schnitzer et al. (2012), who reported $r^2 = 0.52$ between calcaneal QUS T-score and femoral neck BMD T-score in 66 patients with spinal cord injury.

Referring to larger study populations, the best results are those obtained by Dane et al. (2008), who analyzed the correlation between BUA and femoral neck BMD in 351 pre- and post-menopausal women, reporting $r^2 = 0.36$ for the latter and $r^2 = 0.19$ for the former. In fact, a very recent paper (Daugschies et al. 2015) introduced a novel QUS device that measures the speed of sound at the heel with better precision, thanks to the use of dedicated mechanics to finely adjust the US beam direction and a foot temperature sensor to avoid errors related to wrong estimations of soft tissue temperature. A first test on 60 volunteers documented a significant improvement of short- and mid-term precision in speed of sound measurement with respect to a commercially available device, but the actual diagnostic effectiveness of this innovative approach has not yet been tested in comparison with DXA measurements on axial sites.

A different approach was used by Karjalainen et al. (2012), who combined US-measured cortical thickness at distal and proximal tibia with age and weight of the patient, providing a significant estimate of femoral neck BMD ($r^2 = 0.74$, $p < 0.001$); however, this study was conducted on a relatively small population of elderly female volunteers (n = 30) characterized by a narrow age distribution (74.1 ± 3.0 y), and data on larger patient cohorts have not yet been reported. Therefore, because of variable correlations with DXA outcomes and extreme heterogeneity of diagnoses, the recognized value of commercially available QUS devices is limited to fragility fracture prediction for patients over the age of 65 y through calcaneal measurements, whose outcome has to be used in conjunction with clinical risk factors (ISCD 2013).

On the other hand, as expected and previously mentioned, markedly improved correlations between DXA outcomes and US measurements were found when the latter were performed on the axial reference sites. The use of O.S. for spinal investigations resulted in a diagnostic agreement with DXA of 91.1% (n = 79; $k = 0.859$, $p < 0.0001$) and O.S.-based BMD estimates also showed a significant correlation with DXA-measured values ($r^2 = 0.71$, $p < 0.001$) (Conversano et al. 2015); similar results were reported for a linear

combination of SOS values measured in different tissues in the proximal femur region (cortical bone, trabecular bone and soft tissue), which provided a strong correlation with total hip BMD ($r^2 = 0.72$, $p < 0.0001$) (Barkmann et al. 2010). This level of agreement was confirmed by the present study.

We also documented the measurement precision of the proposed approach. The obtained value (RMS-CV = 0.27%) is comparable with values reported for SOS measurements performed at the heel (Njeh et al. 2000) or at the proximal femur (Barkmann et al., 2008), whereas our reported precision was better than the typical corresponding values for both DXA BMD (El Maghraoui et al. 2006; Kiebzak and Morgan 2011) and other QUS parameters evaluated at either the proximal femur (Karjalainen et al. 2012) or the heel (Njeh et al. 2000).

In principle, QUS approaches can be also used to obtain 2-D parametric images (as shown, for instance, by Barkmann et al. 2010), which could be potentially useful for better treatment guidance and monitoring. Nevertheless, with traditional QUS approaches, the finally obtained image resolution is limited by the diameter of the US beam, typically in the order of a few millimeters (Barkmann et al. 2010). In our case, we can superimpose on the original grey-scale echographic image a tailored color map illustrating the spatial location of the bone segments whose RF spectra were classified as "osteoporotic" or "healthy" during O.S. calculation. However, since the response of a given bone segment is influenced by the connections of its trabeculae with those of the neighboring segments, the best way to obtain a reliable parametric image should be the employment of a kind of "moving average" computing the status of each considered bone segment through a (weighted) average of the data obtained from a group of neighboring echographic lines. Nevertheless, this approach will need dedicated investigations to prove its effectiveness.

Regarding the diagnostic accuracy of our method for femoral neck densitometry (overall agreement with DXA output = 94.7%, $r^2 = 0.76$, n = 132), the registered improvement with respect to our previous lumbar spine application (overall agreement with DXA = 91.1%, $r^2 = 0.71$, n = 79) (Conversano et al., 2015) was probably due to the specific attention we paid to ensure the achievement of reliable DXA reports and to the better reproducibility of femoral neck US scans with respect to lumbar ones.

In general, our adopted method, which had been introduced for lumbar spine densitometry and whose application has now been extended to femoral neck, differs from any previous available work because it is not focused on the extraction of a single parameter from the US signal or the corresponding spectrum, but is based on overall correlations between different spectra, each of them considered as a whole. Moreover, the described RF signal analysis method is natively integrated with echographic imaging, since B-mode echographic images are needed for two reasons: 1) the ROI for diagnostic calculations within the investigated bone is identified frame-by-frame by a fully automatic segmentation algorithm, and 2) the simultaneous acquisition of several RF signals, corresponding to the echographic scan lines of the considered frame, is necessary to provide a solid and reliable statistical basis for subsequent spectral analyses.

In regard to the theoretical interpretation of our reference models, even in this case, analogously to what we did for the preliminary interpretation of our lumbar spine spectral models (Conversano et al. 2015), we can adapt the approach proposed by Hoffmeister et al. (2012) for calculating the backscatter difference spectrum between two signal portions gated at different depths on the same signal backscattered from a trabecular bone sample. In our case, for each pair of spectral models, the osteoporotic model is considered as the spectrum of the first gated signal (inferior depth) and the healthy model as the spectrum of the second gated signal (superior depth). The obtained difference spectra are reported in Figure 7, which also shows the analysis bandwidth calculated as in Hoffmeister et al. (2012). The fact that in the evidenced analysis bandwidth (2.1–3.2 MHz) both the considered difference spectra were similar to monotonically increasing quasi-linear functions of frequency represents a first-approximation confirmation of the suitability of the underlying theoretical assumptions, which can be summarized in the "single scattering assumption" (Hoffmeister et al. 2012) combined with the described statistical approach adopted in RF signal analysis to account for local variations in the soft tissues between the transducer and the target bone interface (Conversano et al. 2015).

Referring to the two difference spectra reported in Figure 7, which were obtained by subtracting from each other the two spectra belonging to each pair of reference models reported in Figure 4, we can note that Group A models show almost no difference in the range 1–3 MHz (*i.e.,* the corresponding difference spectrum is close to zero), then the difference slowly increases up to a kind of "plateau" in the range 4–5 MHz. Group B models showed a slightly different behavior. The corresponding curve emphasized a visible negative difference in the range 1.4–2.2 MHz, followed by an almost linear increment up to a broad peak in the range 3.6–4.2 MHz and a slight decrease in the final part of the curve. Therefore, for patients in Group A, the calculated O.S. values were essentially related to spectral differences in the range 3–5 MHz, whereas for patients in Group B measurable spectral differences were also found in the range 1.4–2.2 MHz.

These behaviors resulted qualitatively similar to those observed in our previous work referring to lumbar spine spectral models (Conversano et al. 2015). In fact, we can think that in the younger considered patients (Group A) the alterations due to osteoporosis are visible only in the right side of the spectrum (*i.e.,* the part after the peak in Fig. 4a), whereas in the older patients (Group B) such alterations are more evident and also gradually start affecting the left side of the spectrum (*i.e.,* the part before the peak in Fig. 4b). Interestingly, the "threshold age" at which the spectral alterations caused by osteoporosis start becoming visible also in the left side of the spectrum can be a function of the investigated anatomic site; for instance, in the present study on femoral neck we started seeing these alterations in patients aged in 66–70 y, but in our previous work on lumbar spine similar alterations were already visible in spectra of patients aged in 55–60 y (Conversano et al. 2015). These differences can be partly due to peculiar features of the considered anatomic site (*e.g.,* spectral alterations due to osteoporosis could manifest earlier on lumbar spine than on femoral neck), but could be also attributed to specific patient characteristics (*e.g.,* BMI), affecting the shape of calculated spectra and in turn influencing the "threshold age." However, this will be the objective of future studies, in which the proposed diagnostic approach based on the calculation of O.S. values starting from the described RF spectra will be applied to larger study populations, including wider age and BMI intervals.

The possible routine adoption of the proposed method for femoral neck densitometry would bring important benefits from a clinical point of view, both as a stand-alone site-specific application and as the integration of the corresponding lumbar spine application. In fact, the availability of an effective US method for osteoporosis diagnosis will allow the efficient management of this pathology in primary care settings, with immediate advantages in terms of accessibility to diagnostic tests, prevention and earlier diagnosis. Moreover, a common clinical routine procedure for osteoporosis diagnosis is to perform DXA scans on both lumbar spine and femoral neck and to carry out an integrated evaluation of the two reports. In this context, the possibility of analyzing both reference sites through a widely available echographic approach will further contribute to reduce waiting lists and to increase diagnosis accessibility. This will in turn lead to improved clinical decisions thanks to the possible very early detection not only of osteoporosis but also of osteopenia. The former typically implies a specific drug therapy prescription, whereas the latter can be a warning for a timely correction of lifestyle habits.

To appreciate the specific benefits brought in by the femoral application of the echosound methodology, we should take into account that hip fractures are the most frequent and expensive osteoporotic fractures (620,000 new hip fractures in the EU in 2010, with total costs of about €19 billion; Svedbom et al. 2013), and that BMD measurements have a reasonable fracture risk prediction capability only when referred to the directly measured anatomic site. This provides the proposed method for femoral neck densitometry with a specific added value. However, our approach to femoral densitometry will have an important role also in the prediction of general osteoporotic fracture risk.

In fact, it has been demonstrated that BMD is only one of the factors determining the actual fracture risk, whose accurate estimation requires the combined evaluation of the relevant clinical risk factors (CRFs) (McCloskey 2013). Presently, an effective approach to the pooled evaluation of BMD and CRFs is represented by FRAX® (the Fracture Risk Assessment tool developed by the WHO Collaborating Centre for Metabolic Bone Diseases at Sheffield, UK) (Kanis et al. 2007), a software algorithm that takes into account a variety of patient data (age, sex, ethnicity, *etc.*), including a series of CRFs (*e.g.,* smoking habits, history of previous fragility fractures, *etc.*), and combines them with the DXA-measured BMD value of the femoral neck in order to calculate the 10-y probability of an osteoporotic fracture at the hip and the 10-y probability of a generic major osteoporotic fracture (occurring at hip, spine, humerus or forearm) (Kanis et al. 2008). The possibility of introducing into FRAX a US-measured BMD value of the femoral neck will significantly improve the accessibility of reliable fracture risk predictions. Finally, an additional clinical advantage will be the possibility of more effective therapeutic monitoring, thanks to the intrinsic suitability of US imaging methods for repeated measurements in short time intervals combined with the optimal reproducibility of our US assessments of proximal femur (intra-operator RMS-CV = 0.27%), which is one order of magnitude better than the values typically reported for DXA measurements (El Maghraoui et al. 2006; Kiebzak and Morgan 2011).

In a more prospective view, this new method could provide also important information on the microstructural bone architecture directly reporting a quantitative evaluation of the specific patient fracture risk, as has been demonstrated in several works available in literature illustrating the huge potentialities associated with US investigations of bone "quality" (Gluer 2008; Raum et al. 2014).

Obviously, the present study also suffered from some limitations, which are herein summarized. First of all, the experimental nature of the proposed approach, although supported by several statistical considerations and some previously reported theoretical assumptions, implies that the physiologic variability of acoustic properties of soft tissues from patient to patient was taken

into account only indirectly through the aforementioned sequences of averaging operations. However, the adopted approach has documented its feasibility and diagnostic effectiveness on both lumbar spine and femoral neck, therefore encouraging further investigations dedicated to achieve a more exhaustive theoretical interpretation of the obtained results.

Secondly, our method analyzed only the spectral features of US signals backscattered from the trabecular region of the investigated bone district, even if recent literature has demonstrated the important contribution of the cortical layer to bone mechanical strength, especially in the case of femoral neck (Grimal et al. 2013). However, the assessment of actual bone strength independently of BMD was beyond the scope of the present paper and would also require the use of a technique different from DXA as a reference gold standard (e.g., quantitative computed tomography).

Finally, as in the case of our previous work on the application of O.S. to lumbar spine investigations (Conversano et al. 2015), an obvious influence on the study population was exerted by the involvement of a single clinical center and patient enrolment criteria based on a medical prescription for a proximal femur DXA scan independently from the motivation, although this represented the real field of use of medical devices for osteoporosis diagnosis.

Future studies will be specifically aimed at addressing the mentioned limitations of the present work, providing, in particular, a more detailed explanation of the involved physical phenomena through dedicated *in vitro* tests on excised human bone samples, an integrated evaluation of cortical layer properties through the use of quantitative computed tomography as an additional or alternative gold standard reference, and a multicenter clinical validation of the proposed approach on larger and less homogeneous study populations.

CONCLUSION

Clinical feasibility and effectiveness in femoral neck densitometry were demonstrated for the O.S., a US parameter for osteoporosis diagnosis that had been previously introduced for lumbar spine applications.

The diagnostic accuracy was evaluated in a study population of 132 women aged in 61–70 y, providing an overall agreement with DXA patient classification of 94.7%, with k = 0.898 ($p < 0.0001$). Significant correlations were also found between O.S.-estimated BMD values and corresponding DXA-measured values ($r = 0.87$, $p < 0.001$).

The obtained accuracy levels, combined with a very good result reproducibility (intra-operator RMS-CV = 0.27%; inter-operator RMS-CV = 0.36%) suggest that the proposed approach has an interesting potential for clinical routine applications.

Acknowledgement—This work was partially funded by FESR P.O. Apulia Region 2007–2013, Action 1.2.4 (Grant No. 3Q5AX31: ECHO-LIGHT Project).

REFERENCES

Albanese CV, De Terlizzi F, Passariello R. Quantitative ultrasound of the phalanges and DXA of the lumbar spine and proximal femur in evaluating the risk of osteoporotic vertebral fracture in postmenopausal women. Radiol Med 2011;116:92–101.

Altman DG, Bland JM. Measurements in medicine: The analysis of method comparison studies. Statistician 1983;32:307–317.

Baim S, Leslie WD. Assessment of fracture risk. Curr Osteoporos Rep 2012;10:28–41.

Barkmann R, Laugier P, Moser U, Dencks S, Klausner M, Padilla F, Haiat G, Gluer C-C. A device for *in vivo* measurements of quantitative ultrasound variables at the human proximal femur. IEEE Trans Ultrason Ferroelectr Freq Control 2008a;55:1197–1204.

Barkmann R, Laugier P, Moser U, Dencks S, Padilla F, Haiat G, Heller M, Gluer C-C. A method for the estimation of femoral bone mineral density from variables of ultrasound transmission through the human femur. Bone 2007;40:37–44.

Barkmann R, Laugier P, Moser U, Dencks S, Klausner M, Padilla F, Haiat G, Heller M, Gluer C-C. *In vivo* measurements of ultrasound transmission through the human proximal femur. Ultrasound Med Biol 2008b;34:1186–1190.

Barkmann R, Dencks S, Laugier P, Padilla F, Brixen K, Ryg J, Seekamp A, Mahlke L, Bremer A, Heller M, Gluer CC. Femur ultrasound (FemUS)–first clinical results on hip fracture discrimination and estimation of femoral BMD. Osteoporos Int 2010;21:969–976.

Breban S, Padilla F, Fujisawa Y, Mano I, Matsukawa M, Benhamou CL, Otani T, Laugier P, Chappard C. Trabecular and cortical bone separately assessed at radius with a new ultrasound device, in a young adult population with various physical activities. Bone 2010;46:1620–1625.

Casciaro S, Conversano F, Pisani P, Muratore M. New perspectives in echographic diagnosis of osteoporosis on hip and spine. Clin Cases Miner Bone Metab 2015;12:143–151.

Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37–46.

Conversano F, Franchini R, Greco A, Soloperto G, Chiriacò F, Casciaro E, Aventaggiato M, Renna MD, Pisani P, Di Paola M, Grimaldi A, Quarta L, Quarta E, Muratore M, Laugier P, Casciaro S. A novel ultrasound methodology for estimating spine mineral density. Ultrasound Med Biol 2015;41:281–300.

Cooper C. The crippling consequences of fractures and their impact on quality of life. Am J Med 1997;103:12S–17S.

Curtis JR, Safford MM. Management of osteoporosis among the elderly with other chronic medical conditions. Drugs Aging 2012;29:549–564.

Dane C, Dane B, Cetin A, Erginbas M. The role of quantitative ultrasound in predicting osteoporosis defined by dual-energy X-ray absorptiometry in pre- and postmenopausal women. Climacteric 2008;11:296–303.

Daugschies M, Brixen K, Hermann P, Rohde K, Gluer CC, Barkmann R. Quantitative ultrasound measurements at the heel: Improvement of short- and mid-term speed of sound precision. Ultrasound Med Biol 2015;41:858–870.

Dencks S, Barkmann R, Padilla F, Haiat G, Laugier P, Gluer CC. Wavelet-based signal processing of in vitro ultrasonic measurements at the proximal femur. Ultrasound Med Biol 2007;33:970–980.

Dencks S, Barkmann R, Padilla F, Laugier P, Schmitz G, Gluer CC. Model-based estimation of quantitative ultrasound variables at the proximal femur. IEEE Trans Ultrason Ferroelectr Freq Control 2008;55:1304–1315.

El Maghraoui A, Achemlal L, Bezza A. Monitoring of dual-energy X-ray absorptiometry measurement in clinical practice. J Clin Densit 2006;9:281–286.

El Maghraoui A, Morjane F, Mounach A, Ghazi M, Nouijai A, Achemlal L, Bezza A, Ghozlani I. Performance of calcaneus quantitative ultrasound and dual-energy X-ray absorptiometry in the discrimination of prevalent asymptomatic osteoporotic fractures in postmenopausal women. Rheumatol Int 2009;29: 551–556.

Engelke K, Gluer CC. Quality and performance measures in bone densitometry. Part 1: Errors and diagnosis. Osteoporos Int 2006;17: 1283–1292.

Ensrud KE, Thompson DE, Cauley JA, Nevitt MC, Kado DM, Hochberg MC, Santora AC 2nd, Black DM. Prevalent vertebral deformities predict mortality and hospitalization in older women with low bone mass. Fracture Intervention Trial Research Group. J Am Geriatr Soc 2000;48:241–249.

Ferrari S, Bianchi ML, Eisman JA, Foldes AJ, Adami S, Wahl DA, Stepan JJ, de Vernejoul MC, Kaufman JM. IOF Committee of Scientific Advisors Working Group on Osteoporosis Pathophysiology. Osteoporosis in young adults: Pathophysiology, diagnosis, and management. Osteoporos Int 2012;23:2735–2748.

Genant HK, Cooper C, Poor G, Reid I, Ehrlich G, Kanis J, et al. Interim report and recommendations of the World Health Organization task force for osteoporosis. Osteoporos Int 1999;10:259–264.

Gluer CC. A new quality of bone ultrasound research. IEEE Trans Ultrason Ferroelectr Freq Control 2008;55:1524–1528.

Grimal Q, Grondin J, Guerard S, Barkmann R, Engelke K, Gluer CC, Laugier P. Quantitative ultrasound of cortical bone in the femoral neck predicts femur strength: Results of a pilot study. J Bone Miner Res 2013;28:302–312.

Haiat G, Padilla F, Barkmann R, Dencks S, Moser U, Gluer CC, Laugier P. Optimal prediction of bone mineral density with ultrasonic measurements in excised human femur. Calcif Tissue Int 2005;77:186–192.

Hoffmeister BK, Wilson AR, Gilbert MJ, Sellers ME. A backscatter difference technique for ultrasonic bone assessment. J Acoust Soc Am 2012;132:4069–4076.

Hologic. QDR series user's guide. Bedford: Hologic; 2000.

Hou YL, Liao EY, Wu XP, Peng YQ, Zhang H, Dai RC, Luo XH, Cao XZ. Effects of the sample size of reference population on determining BMD reference curve and peak BMD and diagnosing osteoporosis. Osteoporos Int 2008;19:71–78.

Iida T, Chikamura C, Aoi S, Ikeda H, Matsuda Y, Oguri Y, Ono Y, Katada K, Ishizaki F. A study on the validity of quantitative ultrasonic measurement used the bone mineral density values on dual-energy X-ray absorptiometry in young and in middle-aged or older women. Radiol Phys Technol 2010;3:113–119.

Jiang YQ, Liu CC, Li RY, Wang WP, Ding H, Qi Q, Ta D, Dong J, Wang WQ. Analysis of apparent integrated backscatter coefficient and backscatter spectral centroid shift in calcaneus in vivo for the ultrasonic evaluation of osteoporosis. Ultrasound Med Biol 2014;40: 1307–1317.

Kanis JA. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: Synopsis of a WHO report. WHO Study Group. Osteoporos Int 1994;4:368–381.

Kanis JA, Burlet N, Coope C, Delmas PD, Reginster JY, Borgstrom F, Rizzoli R, European Society for Clinical and Economic Aspects of Osteoporosis and Osteoarthritis (ESCEO). European guidance for the diagnosis and management of osteoporosis in postmenopausal women. Osteoporos Int 2008;19:399–428.

Kanis JA, Oden A, Johnell O, Johansson H, De Laet C, Brown J, Burckhardt P, Cooper C, Christiansen C, Cummings S, Eisman JA, Fujiwara S, Glüer C, Goltzman D, Hans D, Krieg MA, La Croix A, McCloskey E, Mellstrom D, Melton LJ 3rd, Pols H, Reeve J, Sanders K, Schott AM, Silman A, Torgerson D, van Staa T, Watts NB, Yoshimura N. The use of clinical risk factors enhances the performance of BMD in the prediction of hip and osteoporotic fractures in men and women. Osteoporos Int 2007;18: 1033–1046.

Karjalainen JP, Riekkinen O, Toyras J, Hakulinen M, Kroger H, Rikkonen T, Salovaara K, Jurvelin JS. Multi-site bone ultrasound measurements in elderly women with and without previous hip fractures. Osteoporos Int 2012;23:1287–1295.

Kiebzak GM, Morgan SL. Long-term versus short-term precision of dual-energy X-ray absorptiometry scans and the impact on interpreting change in bone mineral density at follow-up. J Clin Densit 2011; 14:108–115.

Kwok T, Khoo CC, Leung J, Kwok A, Qin L, Woo J, Leung PC. Predictive values of calcaneal quantitative ultrasound and dual energy X-ray absorptiometry for non-vertebral fracture in older men: Results from the MrOS study (Hong Kong). Osteoporos Int 2012;23: 1001–1006.

Link TM. Osteoporosis imaging: State of the art and advanced imaging. Radiology 2012;263:3–17.

Liu JM, Ma LY, Bi YF, Xu Y, Huang Y, Xu M, Zhao HY, Sun LH, Tao B, Li XY, Wang WQ, Ning G. A population-based study examining calcaneus quantitative ultrasound and its optimal cut-points to discriminate osteoporotic fractures among 9352 Chinese women and men. J Clin Endorinol Metab 2012;97:800–809.

Liu XS, Shane E, McMahon DJ, Guo XE. Individual trabecula segmentation (ITS)-based morphological analysis of microscale images of human tibial trabecular bone at limited spatial resolution. J Bone Miner Res 2011;26:2184–2193.

Malo MKH, Toyras J, Karjalainen JP, Isaksson H, Riekkinen O, Jurvelin JS. Ultrasound backscatter measurements of intact human proximal femurs–Relationships of ultrasound parameters with tissue structure and mineral density. Bone 2014;64:240–245.

McCloskey E. Assessing fracture risk in patient with osteoporosis. Practitioner 2013;257:19–21.

Messina C, Bandirali M, Sconfienza LM, D'Alonzo NK, Di Leo G, Papini GDE, Ulivieri FM, Sardanelli F. Prevalence and type of errors in dual-energy x-ray absorptiometry. Eur Radiol 2015;25: 1504–1511.

Moayyeri A, Adams JE, Adler RA, Krieg MA, Hans D, Compston J, Lewiecki EM. Quantitative ultrasound of the heel and fracture risk assessment: An updated meta-analysis. Osteoporos Int 2012;23: 143–153.

Muratore M, Conversano F, Renna MD, Pisani P, Villani V, Casciaro S. Social impact of osteoporotic fractures: Early diagnosis and possible therapies. Int J Meas Technol Instrum Eng 2014;4:39–53.

Nayak S, Olkin I, Liu H, Grabe M, Gould MK, Allen IE, Owens DK, Bravata DM. Meta-analysis: Accuracy of quantitative ultrasound for identifying patients with osteoporosis. Ann Intern Med 2006; 144:832–841.

Nguyen TV, Center JR, Eisman JA. Osteoporosis: Under-rated, underdiagnosed and undertreated. Med J Aust 2004;180:S18–S22.

Njeh CF, Hans D, Li J, Fan B, Fuerst T, He YQ, Tsuda-Futami E, Lu Y, Wu CY, Genant HK. Comparison of six calcaneal quantitative ultrasound devices: Precision and hip fracture discrimination. Osteoporos Int 2000;11:1051–1062.

Official Positions of the ISCD(International Society for Clinical Densitometry) as updated in 2013. Available at: http://www.iscd.org/official-positions/2013-iscd-official-positions-adult/. Accessed February 29, 2016.

Otsu N. A threshold selection method from gray-level histograms. Automatica 1975;11:23–27.

Padilla F, Jenson F, Bousson V, Peyrin F, Laugier P. Relationships of trabecular bone structure with quantitative ultrasound parameters: *In vitro* study on human proximal femur using transmission and backscatter measurements. Bone 2008;42:1193–1202.

Paggiosi MA, Barkmann R, Gluer CC, Roux C, Reid DM, Felsenberg D, Bradburn M, Eastell RA. European multicenter comparison of quantitative ultrasound measurement variables: The OPUS study. Osteoporos Int 2012;23:2815–2828.

Pais R, Campean R, Simon S-P, Bolosiu CR, Muntean L, Bolosiu HD. Accuracy of quantitative ultrasound parameters in the diagnosis of osteoporosis. Centr Eur J Med 2010;5:478–485.

Pisani P, Renna MD, Conversano F, Casciaro E, Muratore M, Quarta E, Di Paola M, Casciaro S. Screening and early diagnosis of osteoporosis through X-ray and ultrasound based techniques. World J Radiol 2013;5:398–410.

Raum K, Grimal Q, Varga P, Barkmann R, Gluer CC, Laugier P. Ultrasound to assess bone quality. Curr Osteoporos Rep 2014;12: 154–162.

Schnitzer TJ, Wysocki N, Barkema D, Griffith J, Lent V, Romba M, Welbel R, Bhuva S, Manyam B, Linn S. Calcaneal quantitative ultrasound compared with hip and femoral neck dual-energy X-ray absorptiometry in people with a spinal cord injury. PM R 2012;4: 748–755.

Stewart A, Felsenberg D, Eastell R, Roux C, Glüer CC, Reid DM. Relationship between risk factors and QUS in a European population the OPUS study. Bone 2006;39:609–615.

Svedbom A, Hernlund E, Ivergard M, Compston J, Cooper C, Stenmark J, McCloskey EV, Jönsson B, Kanis JA, EU Review Panel of IOF. Osteoporosis in the European Union: A compendium of country-specific reports. Arch Osteoporos 2013;8:137.

Trimpou P, Bosaeus I, Bengtsson BA, Landin-Wilhelmsen K. High correlation between quantitative ultrasound and DXA during 7 years of follow-up. Eur J Radiol 2010;73:360–364.

van den Bergh JP, van Geel TA, Geusens PP. Osteoporosis, frailty and fracture: Implications for case finding and therapy. Nat Rev Rheumatol 2012;8:163–172.

APPENDIX

Details of the procedure for the automatic identification of femoral neck interfaces

The steps carried out on each data frame acquired on the *k*th patient for the automatic identification of the femoral neck are herein detailed (the same steps were applied to all the acquired patient data sets, belonging to either the reference database or the study population):

1. Image preprocessing: noise removal around the femoral interface and convex image construction. To reduce the computational burden and to speed up the segmentation process, for each echographic line of the acquired data frame, a fixed range of data points, assumed to contain the entire femoral interface, is selected (the specific considered range, defined as the interval [0.30; 0.75] x N_{pix}, where N_{pix} is the total number of data points in each scan line, was calculated experimentally over all the acquired frames of each patient included in the reference database, giving a 100% probability of finding the proximal femur profile in this interval). Subsequently, the signal portion outside the fixed range is set to zero. For each echographic line, the position in which the signal envelope has a global maximum is calculated and a 300-point window, centered on the global maximum position, is recorded. Signal portion outside the 300-point window is again set to zero. The resulting image data points are then organized to reproduce the shape of a typical image provided by a conventional clinically available convex probe (Fig. 3b).
2. Gray-level thresholding and conversion to a binary image. Image intensity values are opportunely adjusted: the gray-scale values of the image *I* are mapped to new values in *J* in such a way that 1% of the data is saturated at lowest and highest intensities of *I*. This increases the contrast of the output image *J,* which subsequently undergoes a global image thresholding using the Otsu's method (Otsu 1975). This technique chooses the threshold to minimize the intra-class variance of the black and white pixels. Since each image is different from the others, the threshold provided by Otsu's method is not fixed but depends on the considered image intensity value distribution. The image is then converted to a binary image (Fig. 3c).
3. Median filter application. Each output pixel contains the median value in the *m-by-n* neighborhood around the corresponding pixel in the input image (Fig. 3d). This non-linear operation is therefore used to reduce the "salt and pepper" noise. A median filter is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges. The values [*m n*] are set to [9 9].
4. Horizontal axis image dilation. Binary image is dilated using a rectangular structural element (*SE*) of non-negative integers, of dimension $M \times N$. Applying *SE* to the binary image consists in a sequential examination of each pixel of the image, performed in such a way that the origin of the *SE* is superimposed to the pixel under study. The value of each point in the output image is based on, then, a comparison between the corresponding pixel of the input image and its neighbors. Applying therefore a rectangular *SE* of dimension [5 15] to the binary input image leads to a horizontal axis image dilation with the only purpose of making uniform the femoral interface and eliminating its discontinuities due to the application of previous steps (Fig. 3e).
5. Asymmetric morphologic closing. The morphologic closing operation is defined by the combination of morphologic dilation followed by morphologic erosion with the same structural element. Such a composed operator has the effect of filling the holes in the images and strengthening the connection between objects. It is also a good tool for the removal of isolated pixels of the background binary images (Fig. 3f). However the *SE* used for dilation and for erosion is in square shape but with different dimension (respectively [8 8] for dilation and [12 12] for erosion) to compensate the dilation of the precedent step.
6. Morphologic evaluations. All clusters of white pixels representing the foreground (connected components) in the binary image are labeled. For each of the *l* connected components (connection = 8), area (number of pixels of the cluster), bounding box (the smallest rectangle containing the region), eccentricity (eccentricity of the ellipse that has the same second-moments as the region) and orientation (the angle between the horizontal *x*-axis and the major axis of the ellipse that has the same second-moments as the region) are calculated. Quantitative cut-offs are then performed over these parameters starting with the selection of the only cluster of pixel which has the maximum area and is in agreement with the following criteria: eccentricity >0.90, absolute value of orientation (in degrees) $< 12°$, and length of the femoral interface in the range 50–100 mm. If no clusters satisfy the mentioned criteria, the frame is discarded and a new frame is considered; otherwise, the selected connected component of pixels undergoes also qualitative cut-offs. The coefficients of a polynomial *p* of degree *n* ($n = 8$) that fits the data in a least squares sense is found and, from the polynomial evaluation, maximum, minimum and inflection points are found in order to get knowledge of the anatomic landmarks (Fig. 3g). Pearson's linear correlation coefficient (*r*) between the function that approximates the femoral profile and a reference fitting function of femoral profile is evaluated and frame is selected if $r \geq 0.90$. Otherwise, little adjustments on the tails of the function that approximate the femoral profile (with an overall function change of about 2% with respect to the original one), are performed, with a smoothing effect. Pearson's linear correlation coefficient is recalculated and the frame is selected if $r \geq 0.90$, otherwise it is discarded.

The reference fitting function expressed in terms of coefficients of a polynomial $p(x)$ of grade 8, with polynomial coefficients in descending powers, is $p(x) = 1.125 \cdot 10^{-17}\, x^8 - 3.326 \cdot 10^{-14}\, x^7 + 4.037 \cdot 10^{-11}\, x^6 - 2.564 \cdot 10^{-8}\, x^5 + 8.879 \cdot 10^{-6}\, x^4 - 1.507 \cdot 10^{-3}\, x^3 + 4.899 \cdot 10^{-2}\, x^2 + 18.45\, x - 1.553 \cdot 10^3$.

7. ROI identification and classification of the proximal femur interface. Using the qualitative morphologic evaluations of the previous step it is possible to discriminate among different profiles, that is, femoral head, femoral neck and trochanter (Fig. 3h). Finally, an evaluation of the femoral neck length (FNL) and of the distance between the highest point of the head profile and the lowest point of the neck profile (HND) is performed. Frame acceptance thresholds are used to select the femoral neck ROI or discard the frame (15 mm $<$ FNL $<$ 55 mm, 3 mm $<$ HDN $<$ 16 mm). If the frame meets these additional selection criteria, it is labeled as a "possible femoral neck interface."
8. Anti-convex image construction. To proceed to the spectral validation of the "possible femoral neck interface," image data points corresponding to the single echographic scan lines are organized in a rectangular matrix, having size $N_{pix} \times N_{lines}$, where N_{lines} is the number of echographic lines (Fig 3i). After image re-arrangement, by calculating the position of the points of the proximal femur profile that separate femoral neck from femoral head and trochanter, this step provides the selection of only the signal lines referred to the femoral neck (vertical echographic lines crossing only the femoral neck profile).
9. Spectral validation. The algorithm analyzes the corresponding RF segments corresponding to the ROI selected as in the case of model construction (150-point Hamming-windowed signal portions, located immediately after the echo from the femoral neck surface when the amplitude of RF signal envelope reached 15% of its peak value, and zero-padded to 4096 points). A comparative analysis is performed on the corresponding fast Fourier transform power spectra, which are compensated and normalized as previously described in order to put them in the form $P_{Norm_{ij}}(f)$ (where *i* identifies the considered ROI and *j* indicates the specific echographic line). The "possible femoral neck interface" is finally labeled as an "actual femoral neck interface" if at least 70% of the spectra of the identified ROI have Pearson's linear correlation coefficient $r \geq 0.85$ with at least one of the appropriate reference model spectra (*i.e.,* the age-matched "healthy" and "osteoporotic" models), otherwise the frame is labeled as "noisy" and discarded.
10. A new analysis starts on the subsequent frame.