Development of an Automated Bone Mineral Density Software Application: Facilitation Radiologic Reporting and Improvement of Accuracy

I-Ta Tsai[1] · Meng-Yuan Tsai[1] · Ming-Ting Wu[1] · Clement Kuen-Huang Chen[2]

Abstract The conventional method of bone mineral density (BMD) report production by dictation and transcription is time consuming and prone to error. We developed an automated BMD reporting system based on the raw data from a dual energy X-ray absorptiometry (DXA) scanner for facilitating the report generation. The automated BMD reporting system, a web application, digests the DXA's raw data and automatically generates preliminary reports. In Jan. 2014, 500 examinations were randomized into an automatic group (AG) and a manual group (MG), and the speed of report generation was compared. For evaluation of the accuracy and analysis of errors, 5120 examinations during Jan. 2013 and Dec. 2013 were enrolled retrospectively, and the context of automatically generated reports (AR) was compared with the formal manual reports (MR). The average time spent for report generation in AG and in MG was 264 and 1452 s, respectively ($p < 0.001$). The accuracy of calculation of T and Z scores in AR is 100 %. The overall accuracy of AR and MR is 98.8 and 93.7 %, respectively ($p < 0.001$). The mis-categorization rate in AR and MR is 0.039 and 0.273 %, respectively ($p = 0.0013$). Errors occurred in AR and can be grouped into key-in errors by technicians and need for additional judgements. We constructed an efficient and reliable automated BMD reporting system. It facilitates current clinical service and potentially prevents human errors from technicians, transcriptionists, and radiologists.

✉ I-Ta Tsai
md@tsai.it

1   Department of Radiology, Kaohsiung Veterans General Hospital, Kaohsiung, Taiwan

2   Department of Radiology, Chi Mei Foundation Hospital, Tainan, Taiwan

Keywords Bone mineral density · DXA · Radiology reporting · Workflow · Open source

Background

Osteoporosis is a metabolic disease characterized by low bone mass, leading to bone fragility and an increase in risk of fracture, which affects millions of people worldwide [1]. Dual energy X-ray absorptiometry (DXA) is considered the gold standard for screening and diagnosis of osteoporosis by measuring bone mineral density (BMD) at the spine, hip, and/or one third (33 %) of the radius site [2]. In current clinical practice in our institute, DXA is used for the diagnosis of osteoporosis, predicting the risk of bone fracture, assessing therapeutic efficacy, and as a screening item in adult periodic health examination.

According to the 2013 official International Society for Clinical Densitometry (ISCD) Adult Positions [2], T score is defined as the difference between measured bone density and average bone density of young women divided by the standard deviation, whereas Z score is population specific (each ethnicity and gender has different reference standards) and the calculation is based on the patient's self-reported ethnicity. The manufacturers use different databases as the reference standard for T scores: data from the third National Health and Nutrition Examination Survey (NHANES III) for femoral neck and total hip T scores and their own databases for those of the lumbar spine. Osteoporosis may be diagnosed in postmenopausal women and in men age 50 and older if the T score is −2.5 or less. In females prior to menopause and in males younger than age 50, a Z score of −2.0 or lower is defined as "below the expected range for age," and a Z score above −2.0 is "within the expected range for age."

The reports of BMD usually contain six to nine numeric data from the DXA scanner and the conclusion is made according to the categorization algorithm aforementioned. The conventional method of report production in our institute is dictation by one attending radiologist working with one transcriptionist. Manual keying in of these data is time consuming and there may be mistakes. We aimed to construct a web application using Ruby on Rails, an open-source web application framework, to improve the reporting process, including the report generation speed and the accuracy of report. By directly importing the data generated by the DXA scanner, our web application has been able to automatically generate structured reports and integrate with the electronic medical record system.

Methods

Application Design

For achieving the purpose of automated generation of structured bone mineral density reports, we built an open-source project—LiberaBMD (Appendix) [3]. In our institute, the bone mineral densities of spine, hip, and/or forearm were measured by the technicians using a Hologic Delphi series DXA scanner (software version 12.7.3; Hologic Inc., Bedford, MA, USA), which supports exporting the scan data and the reference curve data to a file that is compatible with the Microsoft Access database (.mdb) (Redmond, WA, USA) without encryption. In order to import the database from the bone densitometry into LiberaBMD, an open-source software called MDB Tools [4] is used for converting the MDB database file to a structured query language (SQL) file, which is more common and popular in the open-source community.

The database schema from the bone densitometry includes tables to store the data of patient information, examination information, scanning results by regions of interest (ROIs), and reference curve data. The models in the application were set corresponding to these tables, such as *patients*, *scan_analyses*, *spines*, *hips*, and *forearms*. Then, the relationship among the models was also established. Figure 1 briefly illustrates the database schema and the relationships among the tables.

Preliminary Report Generation

While doing the analysis, the *T* and *Z* scores are calculated dependent on the reference curve, and these analytic values, including the peak reference and age-matched percentage, are not maintained in the database nor for exporting either. Therefore, we also have to set models of *reference_curves* and *points* to calculate these important values in the reports. A reference curve for calculation is selected by matching the

gender, ethnicity, and the scanning range. The *T* score is calculated as comparing with the peak bone density age reference and the *Z* score with the age-matched reference by linear interpolating two reference points with an age to the precision of days.

A pre-defined reporting template contained the important values of bone density of spines, hips, and/or forearms from the available scan data and the calculated *T* score with peak reference or *Z* score with age-matched percentage according to the instruction of categorization from the 2013 official ISCD Adult Positions [2]. A conclusion of diagnostic category, such as within or below normal limit in groups using *Z* score and normal, low bone mass, or osteoporosis in groups using *T* score, was made by these scores. These composed a preliminary report, which was integrated into the electronic medical record (EMR) system with a web application programming interface (API) in JavaScript Object Notation (JSON) format. Figure 2 demonstrates result images generated by the DXA scanner and the corresponding preliminary report generated by LiberaBMD.

Integration into the EMR

We used an AutoHotKey [5] (an open-source macro-creation and automation software utility in Microsoft Windows) script for the integration between LiberaBMD and the EMR system. Additionally, another AutoHotKey script was used to automate the comparison between the automatically generated and the formal manually generated reports, which helped to efficiently find discrepancies.

In daily practice, the technicians finished a period of examinations and then exported the database to the application server, which automatically digested the data, and the radiologist could use a hotkey to load the preliminary report into the EMR system. After confirming the accuracy and/or doing some modifications, the report was made formal. Figure 3 shows a flow diagram of the proposed automated system.

Evaluation and Statistical Analysis

For comparison of reporting speed, in Jan. 2014, 500 examinations were randomized into an automated group (AG) and a manual group (MG). With 25 examinations per test unit, the time spent on report generation in AG and in MG (including dictation and check), with 95 % confidence interval, was determined.

For evaluation of the accuracy, 5120 examinations during Jan. 2013 and Dec. 2013 were enrolled retrospectively. The reinterpreting process included three steps: (1) interpreting the DXA scanner-generated report images by a musculoskeletal radiologist as a new report, (2) comparing the new report with the formal manual report, and (3) comparing the context of automatically generated reports (AR) with the formal manual
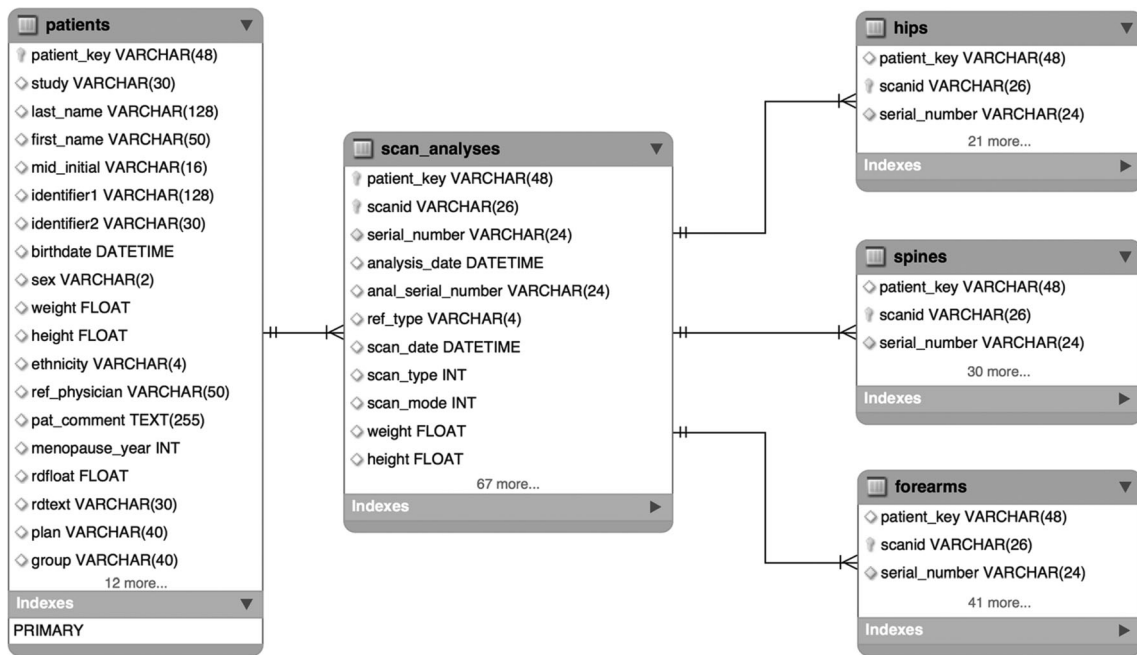
Fig. 1 A brief illustration of the database schema consisting of tables and columns for patient and study information and scanning results of different regions of interest (ROIs). There are also tables for reference curve data, which are not shown

reports (MR) using an AutoHotKey script. Discrepancy among the reports, defined as different ROI areas, numeric data, and the conclusion, was recorded, and the possible causes of discrepancy were analyzed. The evaluation also included the accuracy of calculation of *T* and *Z* scores in AR, which is the most important part for the conclusion of the report.

The statistical analysis was performed using R version 3.1.0. The speed of report generation was compared with two-tailed *T* test, and the accuracy, the subgroup of error rate, and the miscategorization rate of AR and MR were compared with one-sample *Z* test for proportions. Mis-categorization refers to wrong conclusion in the report caused by incorrect interpretation of the

bone density data. For example, an incorrect interpretation might cause a patient to be categorized from osteoporotic to normal, which is thought to be a significant error in reporting as compared with simply using wrong numeric data.

Results

Performance in Report Generation

For comparison of reporting speed, in Jan. 2014, 500 examinations were randomized into an AG and a MG. With 25

PRELIMINARY REPORT

Average bone mineral density (BMD) of L1 to L4 is ***0.939*** gm/cm2, about ***93%*** of the mean of young reference value (T-score = ***-0.6***).

The BMD of right proximal femur is ***0.711*** gm/cm2, and is about ***84%*** of the mean of young reference value (T-score = ***-1.4***).

The BMD of right 1/3 forearm is ***0.666*** gm/cm2, about ***97%*** of the mean of young reference value (T-score = ***-0.3***).

Conclusion:

The BMD meets the criteria of ***low bone mass***, according to the WHO (World Health Organization) classification.

Fig. 2 The images and data tables of different regions of interest (cropped and rearranged for illustration) shown in the *left column* were generated by the dual energy X-ray absorptiometry scanner. The *right* column shows one of the templates of preliminary report with the numeric data and the conclusion text (in *bold italic type*) which were automatically calculated by LiberaBMD
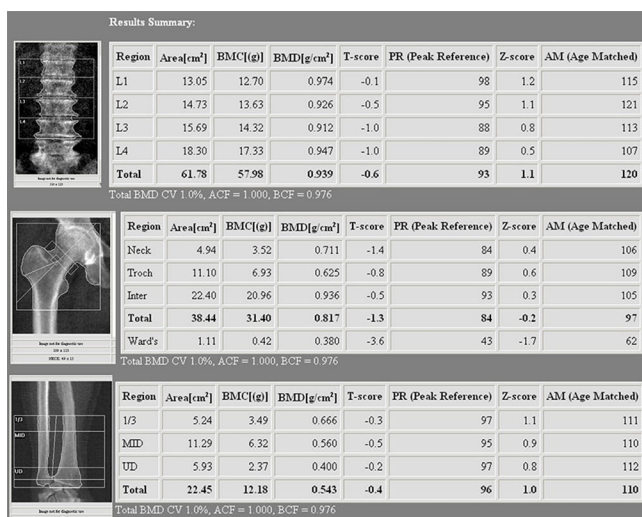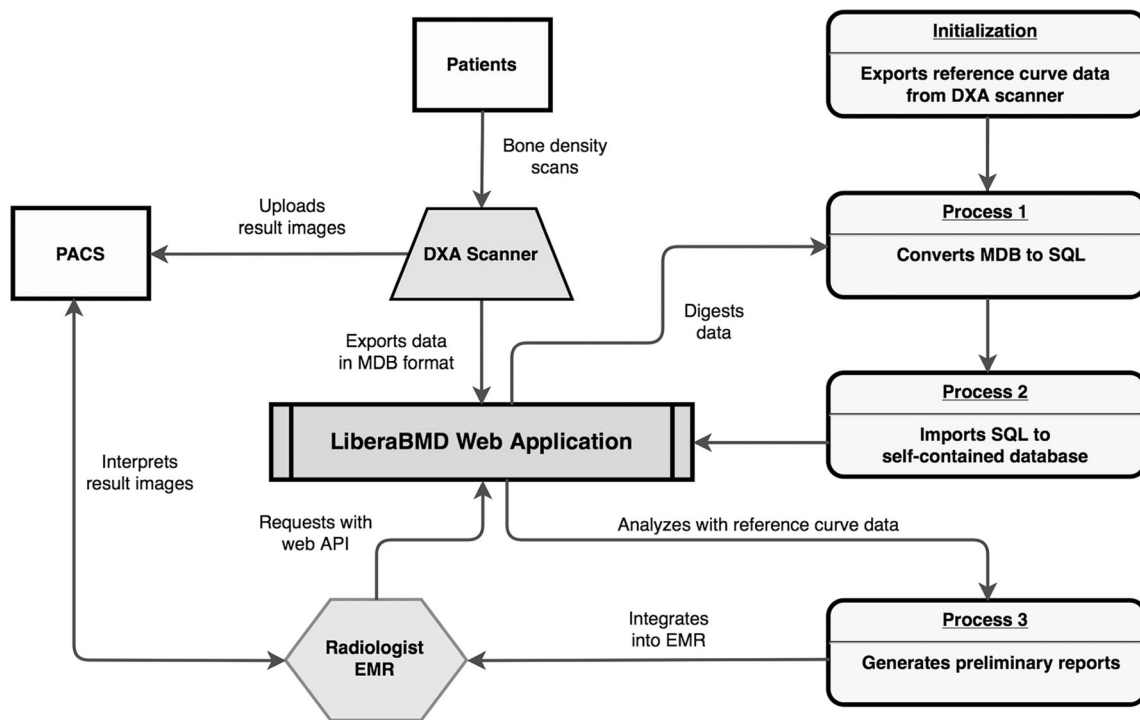
Fig. 3 The flow diagram of LiberaBMD illustrates the system's initialization, processing of the dual energy X-ray absorptiometry (DXA) scanner's data, and integration with the electronic medical record (EMR) system

examinations per test unit, the average time spent for report generation in the AG and in the MG (dictation and check) was 264 s (95 % confidence interval [CI]=235–293) and 1452 s (95 % CI=1366–1538), respectively. The speed of report generation in the AG is significantly faster than that in the MG (p <0.001). The performance of automated report generation is presented in Fig. 4.

Evaluation of the Accuracy

In the evaluation of the accuracy, the AutoHotKey automation script found 383 discrepant reports in 5120 examinations during Jan. 2013 and Dec. 2013. The context of AR was compared with the MR. The accuracy of calculation of T and Z scores in AR is 100 %.

Error Analysis

Table 1 contains definition and examples of error for both the AR and MR groups. Of the 5120 BMD examinations with automated generated reports, key-in errors by technicians accounted for 33 (0.64 %) errors, and the need for additional judgements accounted for 29 (0.57 %) errors. The key-in errors include the following: (1) missing images on picture archiving and communication system (PACS), (2) accession number changed, (3) missing menopause age, and (4) different analysis parameters. The need-for-additional-judgements

errors include the following: (1) choosing one of the repeated scans and (2) exclusion of abnormal values. Of the 5120 BMD reports generated conventionally, misuse of T or Z score accounted for 128 (2.50 %) errors, containing subgroups of (a) T score in premenopausal women, (b) men younger than 50 years of age, (c) Z score in postmenopausal women, and (d) men older than 50 years of age. Misuse of ROI accounted for 93 (1.82 %) errors, containing subgroups of (a) choosing incorrect ROI of hip and (b) choosing incorrect ROI of spine. Dictation error accounted for 69 (1.35 %) errors, containing subgroups of (a) incorrect peak reference percentage, (b) incorrect T score, (c) incorrect Z score, (d) incorrect age-matched percentage, and (e) using an inappropriate report template. Data omission accounted for 30 (0.59 %) errors, containing subgroups of missing report of hip, forearm, and spine. Table 2 demonstrates the detailed proportions of errors in groups and subgroups for the AR and MR.

Overall Accuracy and Mis-categorization

The overall accuracy of the AR and MR is 98.8 and 93.8 %, respectively. The overall accuracy of AR is statistically significantly higher than that of MR (p<0.001).

The mis-categorization rate of BMD in the AR and MR is 0.039 % (2/5120) and 0.273 % (14/5120), respectively. The AR group has significantly less mis-categorizations than the MR group (p=0.0013) (Table 2).

Fig. 4 The time spent for report generation in the automated group (AG) and manual group (MG). The average time spent on generation of 25 reports in the AG and in the MG was 264 s (95 % CI=235–293) and 1452 s (95 % CI=1366–1538), respectively (p <0.001)
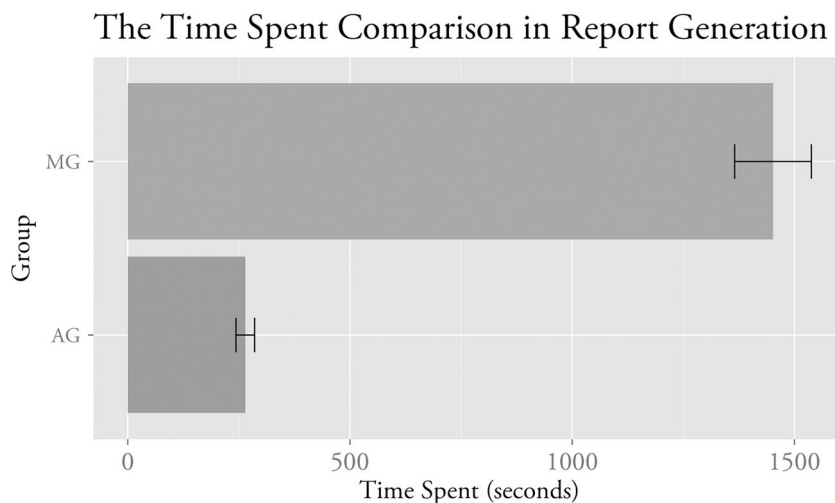
Discussions

Our study constructed an open-source tool to provide an accurate and efficient automated mechanism for assisting BMD report production. The application can be set up on a simple personal computer, and there is no need to modify or to communicate with the current workstation of the DXA scanner. By exporting the database after each session of examination or periodically, the application digests the data for preliminary report generation and backup of the raw data of the DXA scanner simultaneously.

Iv et al. [6] reported a method to facilitate preliminary report generation by extracting information from the Digital Imaging and Communications in Medicine (DICOM) files of the DXA result images. However, the software we use does not support generating a structured reporting (SR) object. Without purchasing or upgrading to a newer DXA scanner or the next generation of software, there are two possible ways to efficiently retrieve the numeric data in bone density examinations. The first is direct access to the database of the DXA scanner, and the other is performing optical character recognition (OCR) on the result images. The analytic values, such as T and Z scores and the peak reference and age-matched percentage, are dependent on the reference curve and therefore not maintained in the database. They should be calculated in our application by following the manufacturer's algorithm to fit a reference curve, which is also exported from the DXA scanner.

A recent publication has described an algorithm using OCR technology to fetch the analyzed results of bone mineral density from the images on the PACS with a high accuracy of OCR in the numeric data. They reported accuracy of T scores of the OCR of 95.4–100 %, with a necessity of correcting the OCR errors, such as segmentation errors and misrecognition [7]. In our application, the accuracy of calculation of T and Z scores was 100 %. Direct access to the raw data from the database is thought to be a more reliable way for these

numeric data. Besides, our application had implanted the following steps after retrieving the T or Z scores, leading to a preliminary conclusion and categorization of the examinations, which also facilitates the report production.

The retrospective evaluation of the automatically generated reports shows 98.8 % overall accuracy for LiberaBMD, and the causes of error were key-in errors by technicians (0.64 %) and need for additional judgements (0.57 %) (Table 2). Most of the errors in the automated reports presented with a discordant total number between the preliminary report generated by LiberaBMD and the actual result images on the PACS, except for "accession number changed" and "missing menopause age" subgroups. To detect such situations to prevent possible errors, the radiologist may be warned by the application during the assisted report production by comparing the total number of examinations in the preliminary report with that in result images on the PACS. When "accession number changed" error occurs, no preliminary report could be generated. Switching to the conventional manual method is mandatory. For "missing menopause age" error, both the raw data from the DXA scanner and the result images cannot show the information of menopause age. In the current clinical scenario, correction of this error needs to be double-checked by the referred clinician or the staff of the Physical Examination Center.

In a previous conventional workflow, ancillary radiological comments such as degenerative change of spine with relative high T scores to other ROIs, which implies the possibility of overestimation of bone density status, were made in the report, and they are not generated in the current version of the application. Even though the ISCD does not set any criteria on this condition, a revised algorithm could be adopted according to the clinical circumstances or the experts' consensus and would give a warning message in future automated reporting. Besides, this overestimation might have little influence on the conclusion if other skeletal sites are measured according to the instruction of the ISCD official positions [2].

Table 1 The definitions and examples of the category of cause of error for both the AR and MR groups

Cause of error	Definition and example
Automated report (AR)	
Key-in errors by technicians	
Missing images on PACS	A part of the examination was performed and the scan data was stored in the database, but the analysis result image was not uploaded to the picture archiving and communication system (PACS)
Accession number changed	The accession number had been changed which makes the application unable to find the data
Missing menopause age	Missing input of menopause age in female patients in perimenopausal age leading to incorrect using of the T or Z score by the application
Different analysis parameters	Reanalyzing the data with different parameters causing 2 different result images on the PACS
Need for additional judgements	
Choose 1 of the repeated scans	Repeated scans of 1 body part and 1 of them was chosen, which results in more examination items than those on the PACS
Exclusion of abnormal values	One part of the scans was excluded due to abnormally high bone density
Manual report (MR)	
Misuse of T or Z score	
T score in premenopausal women	Z scores are preferred in premenopausal women [2]
T score in men <50	Z scores are preferred in men younger than age 50 [2]
Z score in postmenopausal women	T scores are preferred in premenopausal women [2]
Z score in men ≥50	T scores are preferred in men younger than age 50 [2]
Misuse of region of interest (ROI)	
Incorrect hip ROI	Use femoral neck or total proximal femur, whichever is lowest [2]
Incorrect spine ROI	Use PA L1–L4 and only exclude vertebrae that are affected by local structural change or artifact. Use 3 vertebrae if 4 cannot be used and 2 if 3 cannot be used [2]
Dictation error	
Incorrect peak reference percentage	For a specified ROI, all numeric data was correct except the peak reference percentage
Incorrect T score	For a specified ROI, all numeric data was correct except for the T score
Incorrect Z score	For a specified ROI, all numeric data was correct except for the Z score
Incorrect age-matched percentage	For a specified ROI, all numeric data was correct except for the age-matched percentage
Inappropriate report template	The report has a ROI which was not included in the examination. In this case, the report states ROI of right femur but filled with the results of the right forearm
Data omission	
Missing report of hip	The analysis of hip ROI existed in both the automatically generated report and the result image on PACS, but missed in the manual report
Missing report of forearm	The analysis of forearm ROI existed in both the automatically generated report and the result image on PACS, but missed in the manual report
Missing report of spine	The analysis of spine ROI existed in both the automatically generated report and the result image on PACS, but missed in the manual report

Retrospective analysis of error in conventional manual reports shows a significant lower overall accuracy of 93.8 %. The causes of error can be reviewed in two points of view, from the misreading of the reporting radiologist and during dictation, which include misuse of T or Z score (2.5 %), misuse of region of interest (1.82 %), dictation error (1.35 %), and data omission (0.59 %). Error in radiology is not seldom and the level of error varies depending on the radiological

Table 2 The reasons of error in the automatically generated reports (AR) and the formal manually made reports (MR). The overall accuracy in the AR is significantly higher than that in the MR ($p<0.001$), whereas the mis-categorization rate in the AR is significantly lower than that in the MR ($p=0.0013$)

Cause of error	Error rate ($n=5120$)	Mis-categorization	Overall accuracy (%)
Automated report (AR)	1.21 % (62)	0.04 % (2)	98.8
Key-in errors by technicians	0.64 % (33)	0.04 % (2)	
Missing images on PACS	0.41 % (21)	0.02 % (1)	
Accession number changed	0.12 % (6)	0	
Missing menopause age	0.08 % (4)	0.02 % (1)	
Different analysis parameters	0.04 % (2)	0	
Need for additional judgements	0.57 % (29)	0	
Choose 1 of the repeated scans	0.47 % (24)	0	
Exclusion of abnormal values	0.10 % (5)	0	
Manual report (MR)	6.25 % (320)	2.73 % (14)	93.8
Misuse of T or Z score	2.50 % (128)	0.16 % (8)	
T score in premenopausal women	2.27 % (116)	0.12 % (6)	
T score in men <50	0.18 % (9)	0.04 % (2)	
Z score in postmenopausal women	0.04 % (2)	0	
Z score in men ≥50	0.02 % (1)	0	
Misuse of region of interest (ROI)	1.82 % (93)	0.02 % (1)	
Incorrect hip ROI	1.70 % (87)	0	
Incorrect spine ROI	0.12 % (6)	0.02 % (1)	
Dictation error	1.35 % (69)	0.06 % (3)	
Incorrect peak reference percentage	0.43 % (22)	0	
Incorrect T score	0.39 % (20)	0.04 % (2)	
Incorrect Z score	0.29 % (15)	0.02 % (1)	
Incorrect age-matched percentage	0.21 % (11)	0	
Inappropriate report template	0.02 % (1)	0	
Data omission	0.59 % (30)	0.04 % (2)	
Missing report of hip	0.39 % (20)	0.04 % (2)	
Missing report of forearm	0.16 % (8)	0	
Missing report of spine	0.04 % (2)	0	

Fig. 5 Screenshot of one of the LiberaBMD's statistical page to summarize the mean BMD and standard deviation in each ROI in a specific patient group. Serving as a duplication of the bone densitometry's database, it has potentialities for further epidemiological statistics

investigation, and the range is 2–20 % for clinically significant or major error [8]. In one study comparing the accuracy of voice-recognized and dictation-transcription radiology reports, McGurk at el. reported an error rate of 2.1 % in dictation-transcription reports [9], whereas the present study has a similar result in the conventional manual group of an overall accuracy and a dictation error rate of 93.8 and 1.35 %, respectively.

Mis-categorization refers to wrong conclusion in the report caused by incorrect interpretation of the bone density data and is thought to be a more significant error than simply using wrong numeric data. There were mis-categorizations in both groups, two (0.039 %) in the automated group and 14 (0.273 %) in the manual group. In the AR group, the causes of mis-categorization are "missing images on PACS" and "missing menopause age." The first error can be possibly prevented by the alert mechanism aforementioned, and even without discarding this potentially correctable error, the AR group has significantly less mis-categorizations than the MR group ($p = 0.0013$).

By exporting the database from the bone densitometry, the database in LiberaBMD could be considered as a duplication of the DXA scanner. With an accumulative and complete database, further epidemiological statistics such as normal range of our population could be relatively easily performed (Fig. 5).

This application has some limitations. First, its database schema design is currently only compatible with that of Hologic Delphi DXA system. Conversion of the database schema and relinking the models in the application are necessary for application to other DXA systems. Second, the compatibility for other software versions is uncertain due to unavailable source for testing. The restrictions may be less significant because LiberaBMD is an open-source project that anyone can easily access the source code to modify for his/her own circumstances [10]. Third, our study was conducted at a single institution. With different reporting configurations, the experience of the radiologists and transcriptionists, and the cooperation between them, this project may have varied impact on the report generation efficiency and accuracy. Fourth, many of the product manufacturers encrypt the data before storing the data into the database, and directly manipulating the database is not feasible. To our knowledge, the software of Lunar iDXA (GE Healthcare) has an encrypted database and does not support exporting unencrypted data in neither SQL nor MDB format. The application is not suitable for these bone densitometries.

Conclusion

We constructed an efficient and reliable web application for automated BMD report generation. To our knowledge, we are the first to describe such a fashion to extract the data from the DXA's database. It facilitates current clinical service, reduces the radiologists' working load, and potentially prevents human errors from technicians, transcriptionists, and radiologists. Some potential academic applications could be developed based on the complete database. As an open-source project, there is a lower threshold in modification for other specific circumstances.

Appendix

The software and scripts used in the open-source project, LiberaBMD, are described briefly as below.

LiberaBMD was developed with Ruby on Rails (http://rubyonrails.org/), which is a popular web framework that supplies features of model-view-controller architectural pattern, ActiveRecord (an object-relational mapping system for database access). These features help the application designer to focus more on software logistics than low-level database interaction. It handles data storage and calculation of T and Z scores and creates web APIs to communicate with the EMR system.

AutoHotKey (AHK) (http://ahkscript.org/) is an open-source macro-creation and automation software utility in Microsoft Windows. We created two AHK scripts to retrieve the preliminary report generated by LiberaBMD and facilitated the accuracy evaluation by comparing the fields in the formal manual reports and the automated reports.

References

1. Consensus development conference: diagnosis, prophylaxis, and treatment of osteoporosis. Am J Med 94:646–650, 1993
2. Schousboe JT, Shepherd JA, Bilezikian JP, Baim S: Executive summary of the 2013 international society for clinical densitometry position development conference on bone densitometry. J Clin Densitom Off J Int Soc Clin Densitom 16:455–466, 2013
3. LiberaBMD. Available at https://github.com/tsaiid/libera-bmd. Accessed Nov 4, 2014
4. MDB Tools—Unlocking Your Data. Available at http://mdbtools.sourceforge.net/. Accessed Nov 4, 2014
5. AutoHotkey. Available at http://ahkscript.org/. Accessed 5/24/2015
6. Iv M, Patel MR, Santos A, Kang YS: Informatics in radiology: use of a macro scripting editor to facilitate transfer of dual-energy X-ray absorptiometry reports into an existing departmental voice recognition dictation system. Radiographics Rev Publ Radiol Soc N Am Inc 31:1181–1189, 2011
7. Lee YH, Song H-T, Suh J-S: Quantitative computed tomography (QCT) as a radiology reporting tool by using optical character recognition (OCR) and macro program. J Digit Imaging 25:815–818, 2012
8. Goddard P, Leslie A, Jones A, Wakeley C, Kabala J: Error in radiology. Br J Radiol 74:949–951, 2001
9. McGurk S, Brauer K, Macfarlane TV, Duncan KA: The effect of voice recognition software on comparative error rates in radiology reports. Br J Radiol 81:767–770, 2008
10. The Open Source Definition | Open Source Initiative. Available at http://opensource.org/docs/osd. Accessed Nov 4, 2014