Artificial intelligence and breast screening: French Radiology Community position paper

Isabelle Thomassin-Naggara, Corinne Balleyguier, Luc Ceugnart, Patrice Heid, Greg Lenczner, Aurélien Maire, Brigitte Séradour, Laurent Verzaux, Patrice Taourel

Artificial intelligence and breast screening:

A position paper for the French radiological community

Isabelle Thomassin-Naggara[1,2], Corinne Balleyguier[3], Luc Ceugnart[4], Patrice Heid[5],

Greg Lenczner[6], Aurélien Maire[7], Brigitte Séradour[8], Laurent Verzaux[9], Patrice Taourel[10]

Sous l'égide du Conseil National Professionnel Collège des Enseignants de Radiologie de France (CERF), de la Société Française de Radiologie (SFR) / Fédération Nationale des Médecins Radiologues / Syndicat des Radiologues Hospitaliers
En association avec la Société d'Imagerie de la FEMme (SIFEM)

1- Sorbonne Université – APHP - Paris
2- Collège des enseignants de Radiologie de France (CERF)
3- Institut Gustave Roussy, Service de radiologie Villejuif
4- Centre Oscar Lambret, Lille
5- Centre Arcades, Marseille
6- Fédération Nationale des Médecins Radiologues (FNMR)
7- APHP, Equipe Wind, Paris
8- Centre Arcades, Marseille
9- Société Française de Radiologie (SFR)
10- Université de Montpellier, Hôpital Lapeyronie

Short title: Artificial intelligence in breast imaging

Type of manuscript: Review / Position paper

Key words: Breast cancer, artificial intelligence, French screening program

Correspondence:

Isabelle Thomassin-Naggara,

Service de Radiologie,

Hôpital Tenon, 4 rue de la Chine,

75020, Paris, France

Tel.: +33 1 56 01 78 30

Fax: +33 1 56 01 64 02

Abstract

The objective of this paper was to evaluate the evidence currently available on the clinical value for artificial intelligence in breast imaging. Nine experts from the disciplines involved in breast disease management – including physicists and radiologists–met in a meeting on 2019, June, 3th and discussed the evidence for the use of this technology in plenary and focused sessions. Prior to the meeting, the group performed a literature review on predefined topics and the paper presents the consensus reached by this working group on the recommendations for the future use and research issue of AI.

Introduction

Over 10 years, the number of publications on artificial intelligence (AI) in radiology have exponentially increased up to 700–800 per year and breast screening is one of the major applications in radiological imaging: First, breast cancer is a public health issue in France in 2017 with 58 968 incident cancer and 11 883 annual number of deaths. Second, breast screening is based on a clinical and radiographic examination named mammography which includes four 2D incidences (2MLO and 2CC) with an exam every two years in Europe and annually in Unites States. Third, the limitations of mammography are a wide variability of interpretive performance, (1) a masking effect with dense tissue, the needed of double-reading of each mammogram (2). Furthermore, in mammography, we have a long history with Computer Aided Diagnosis (CAD) which were as quickly implemented in United states at twenty years ago and nowadays left due to its inefficiency because of a high level of false positives (3,4). Thus, a certain degree of skepticism exists in breast radiologist's mind about this concept.

The French society of Radiology and the College des Enseignants de Radiologie (SFR and CERF) organised a consensus meeting in Nimes on 3–4nd june 2019 to evaluate the evidence currently available on the clinical value for artificial intelligence in breast imaging. Nine experts from the disciplines involved in breast disease management – including physicists and radiologists– discussed the evidence for the use of this technology in plenary and focused sessions. Prior to the meeting, the group performed a literature review on predefined topics; defined questions to be answered at the meeting were identified. This paper presents the consensus reached by this working group on the recommendations for the future use and research issue of AI. Following the meeting, the literature review has been updated to June 2019. The working group intends to further update these recommendations as and when new relevant evidence becomes available.

Background

In the field of breast imaging, as well in all radiology fields, five main groups of application exist in artificial intelligence

1) **lesion classification** that consists in predicting the nature of a group of pixels, such as determining tumor versus normal, and malignant versus benign.
2) **image processing, mainly including tissue and lesion segmentation** that consists in Identifying which pixels are part of a structure of interest, identifying which pixels are abnormal within an identified structure, and labeling each pixel in an image with its type (semantic segmentation). Some authors like Erickson from Mayo Clinics suggests that "Image-omics" may be better than tissue genomics. This is the domain of radiological classifiers and precision medicine for prognostic imaging, useful for NAT response for example.
3) **Generative tasks** are on going that consists in creating new images based on current images
4) **Regression** that consists in predicting a continuous variable from inputs, such as predicting age from a hand radiograph
5) **Workflow and efficiency** that may allow to reduce dose and acquisition time

Under the term “artificial intelligence”, there are many different training techniques including artificial neural networks (ANNs), machine learning (ML), and deep learning (DL) (5) The machine learning is a data driven approach learning with a mathematical model based on the observed “training” data in which there are two main types of models: If the learning is based on the features from the labelled images this is supervised AI, if training data has no diagnosis or normal/abnormal labels, this is unsupervised AI:

a) **Supervised AI** (decrescent frequency) includes artificial neural networks (ANNs) including Convolutional neural network (CNN), Support vector machine (SVM),

random forest (RF), linear discriminant analysis (LDA), and decision tree. The highest accuracy in the SVM method is observed in the results of a research, which used an appropriate segmentation method for obtaining the desired area in the image. The shape and intensity of the extracted features had the most effect in the classification. The combination of gray-level co-occurrence matrix (GLCM) and ratio features along with morphological features resulted in the highest accuracy. These types of algorithms are classical machine learning (type ANN) that need relatively low computational requirements in comparison with deep learning architectures algorithms (type CNN) that need millions of parameters and thus high-performance computing hardware.

b) Unsupervised AI (clustering) represented mainly (77%) by k-nearest neighbor (k-NN). One of the most famous application in breast pathology is the description of the intrinsic molecular subtypes for breast cancer in the journal Nature in 2000 (6)

While supervised learning algorithms were primarily used, with the AUC value from ROC analysis ranging from 0.74 to 0.98 (median, 0.87) and with that from prognostic imaging ranging from 0.62 to 0.88 (median, 0.80), unsupervised learning are mainly used for image processing purposes. (7). Actually, artificial neural networks (ANN) , support vector machines (SVM), and clustering are the most frequently used algorithms, accounting for 66% of AI imaging publications (7).

This paper will focus the application of artificial intelligence in 5 main fields of breast imaging : Image acquisition, automatic assessment of breast density, the actual AI CAD working on 2DMG and breast tomosynthesis, synthetic mammography and personalized screening. Moreover, two chapters detail the specificity of AI implemantation in France related to specific french screening organization and present the different french databases.

Chapter 1 : 2D mammography : Present and next future : What do we expect for artificial intelligence ?

1-1 Radiation dose optimisation and quality control

Deep-learning algorithms have been developed for improving images, including speeding up acquisition time and outperforming traditional noise reduction techniques in image reconstruction. In particular, generative adversarial networks (GANs) are going to have a huge impact in radiology, according to Erickson.

In order to reduce dose manufacturers have developed different tools. A first approach was the development by Siemens of a new optional software PRIME (Progressive Reconstruction Intelligently Minimizing Exposure), which may be used for breast thicknesses up to 70mm and is able to do a scatter correction without the use of anti-scatter-grid. Instead of the system using a grid, the software identifies structures in the breast that cause scatter, and subtracts the calculated scatter. The mAs is then lower as the X-rays are not absorbed by a grid. The dose saving depends on breast thickness and structure.

For tomosynthesis implementation in breast cancer screening, manufacturers should optimize image quality in synth2D mammography. Today, only « real » 2D mammography with an image acquisition can be used in Europe. Meaning that the dose can be double for a complete examination with 2D and 3D images. The Synth2D mammography is actually implemented on some systems and is a work in progress for others.

In US, Hologic received FDA approval for C-View in 2013. Thus, synthetic MG may replace the conventional 2D image. Approval was based on a Hologic study showing that 3D + C-View is non-inferior to conventional 2D digital mammography (8).

FDA has also approved « high definition » breast tomosynthesis technology from Siemens in 2017. The technology incorporates a software call EMPIRE (enhanced multiple parameter iterative reconstruction), a combination of iterative and machine learning

algorithms. This technology from SIEMENS has been approved as a 3D-only exam. But it also includes "Insight 2D and 3D", a synthetic software that generated tomosynthesis volumes in 3D and that allows to obtain a 2D image from the EMPIRE 3D slides without added dose for the 2D exposition.

The last point is Quality Control. Well defined and applied for Full Field Digital Mammography (FFDM), quality control guidelines have not yet been finalized for tomosynthesis. In European countries, tomosynthesis is not yet approved in national screening programmes. Only some experimental breast cancer screening experiments are done (as TOMMY Trial in UK). The EUREF group have done a huge work for publishing tomosynthesis QC guidelines (available on their internet web site since 2015). In UK, the National Co-ordinating Centre for the Physics of Mammography (NCCPM) is run QC tests based on the EUREF guidelines on behalf of the NHS Breast Screening Programme and has developed tools to analysed 3D QC images. But the time required to analyze QC data is important today.

French situation is more complicated and less clear. More than 400 units have been used daily without any QC (daily/weekly/monthly radiographer's tests or semestrial physicists tests) since years. Not in the breast cancer screening programme but for all survey mammograms and additional diagnostic exams. The first evaluation tests done in France, based on the EUREF guidelines, shows a huge disparity between manufacturers in terms of image quality and dose, even on manufacturers adjustments for the same brand of systems. Without a normalized evaluation and a strict QC, it is not possible to know the physical performances of the installed systems and therefore the potential clinical results. In the meantime, the QC is carried out in France by private companies using technicians not always well trained. The lack of medical physics experts in France and the number of private

radiology centers doing mammography (more than 2,500) poses a problem for the implementation of effective quality control in tomosynthesis and in synthetic view evaluation.

1-2 Image quality acquisition

1-2.1 Positioning, parameters

Another important issue on breast cancer screening is radiation dose which is correlated with breast thickness and therefore the correct positioning of the patient (9). One way to use artificial intelligence would be to help the technician achieving optimal positioning, by defining the right compression force and showing the right exposure parameters. Volpara® has developed software that can give feedback to the radiographers and radiologists on the quality of all these parameters. After each exposure, anonymized data are sent to an external cloud database that is able to analyze the final image quality of each diagnostic image and give an advice on all acquisition parameters. General Electric presented similar tools in development at last radiology congresses, allowing an analysis of each examination on qualitative criteria. These different softwares checking the quality of the exams can therefore be used as a continuous training.

1-2.2 Automatic breast density assessment

Dense breasts are associated with higher risk of breast cancer (10–12). In addition to breast cancer risk prediction, breast density assessment is also crucial because associated with masking of cancers leading to interval cancers in mammographic screening (13).

A wide degree of variability is well known among the radiologists and also for the same radiologist at two different time: In a large prospective multicentric observational study (N=216 783 including 34 271 patients seen several times), low inter-observer agreement was found and low intra observer agreement also with 17% of women differently categorized on

the different successive mammograms (14) requiring the need of more reproducible softwares.

Ten years ago, first software were created to allow automatic assessment of breast density hoping a better performance and reproducibility. Most of them were based on segmentation techniques and did not reach or just reach the accuracy of BI-RADS subjective assessment to predict breast cancers (QUANTRA, Hologic – Volpara – Densitas..). More recently, these software were redesigned to integrate deep learning model (QUANTRA 2.2, ..) and further evaluations are needed

In a recent publication (15), a deep conventional neural network was compared with human analysis of breast density according to BI-RADS lexicon (2013) on a set of 20 578 mammographic images issued after data augmentation from 12 932 MLO and CC images. This study demonstrated a very good agreement to split mammography between fatty breast (BD rated A and B) from dense breast (BD rated C and D) with an agreement reached 99% for MLO views and 96% for CC views.

This study is in line with another publication in Radiology for the team of C.Lehman et al. who developed also a deep convolutional neural network, ResNet-18 (20), with PyTorch (2018, version 0.31; pytorch.org) using 58 894 randomly selected digital mammograms from 39 272 women screened between January 2009 and May 2011. This paper tested this AI algorithm in 10 763 consecutive screening digital mammograms from January to May of 2018 and shew that the DL model matched the radiologist interpretation in 78% of mammograms for four-way BI-RADS categorization, and in 94 for binary categorization of dense or non dense breasts (16)

Others authors have correlated an automatic assessment of breast density with commercial software based on artificial intelligence technique with screening population characteristics in Norway (n= 107 949) on 307 015 MG (17). They concluded that screening

examinations of women with dense breasts assessed by using automated software resulted in higher recall rate, lower sensitivity, larger tumor diameter, and more lymph node–positive disease compared with women with nondense breasts (17)

Finally, in addition to breast density, recently another descriptor of breast parenchyma was built to describe the complexity of the gland (18). Indeed, radiomic phenotypes were defined and studied in Kontos et al ‘s article published in Radiology in 2019. On a cohort of 2241 women with MG + DBT (both views), they performed an unsupervised hierarchical clustering in a population training (n=1339) and a validation set (n=690) and classified the parenchyma in 4 levels of complexity. In this article, they demonstrated that low or low intermediate complex parenchyma have an OR higher than 2 to develop a breast cancer, independently than breast density (18).

Research issues

- To develop cloud software analyzing automatically daily images and able to detect defaults or instabilities to help medical physicists to decrease the analysis time during QC
- To create automatic analysis of quality control criterial to simplify daily QC and decreasing time for technologists
- To propose automatic assessment patient control quality for technologists (positioning, compression, flou, artifact) (report for technologist’s self-assessment)
- To optimize automatic assessment of parenchymal evaluation including density, complexity, heterogeneity to reach equal accuracy than BI-RADS classification for predicting breast cancer risk for radiologists

1-3 Which are AI algorithm trained and validated on 2DMG?

1-3-1 CAD performance

First CAD in breast imaging was approved in 1998 and use from 2002 after reimbursement act in USA. CAD is use in 92 % of all screening mammograms in 2006 (19). There are 2 components for CAD; CAD for detection and another for diagnosis who help to classify and interpret. In our topics, CAD for detection is probably the most important but the new CAD generation are able to provide both for detection and characterization.

First CAD generation were able to mark 86 % of missed calcifications and 72 % Masses and in other studies 42 % of very subtle mammographic cancer finding deemed occult for radiologist. But, in clinical practice, the majority of these marks were considered as False-positive and negatived by radiologist. Nevertheless, if this traditional CAD did not identify all the cancer detected by a radiologist, they improve sensitivity and even reached a second reader accuracy for detection but decrease specificity. In a 2007 study by Fenton et al (3), specificity significantly decrease and result in a 20% increase in the biopsy rate, lowering overall accuracy (AUC à,807 versus 0,919). In summary, traditional CAD have very high sensitivity for calcifications (99%), lower sensitivity for masses (75-99%) and poor sensitivity for architectural distortion (38%). The use of CAD in general practice was evaluated in a large retrospective study from Breast Cancer Surveillance Consortium in 600.000 mammograms read with and without CAD by 271 radiologists across 66 facilities (4). With a cancer detection rate at 4, 1 %, sensitivity and specificity were identical with or without use of CAD. The 107 radiologists even decrease their performances with CAD especially for sensitivity (83% with CAD versus 89 % without CAD).

Technical limitations of traditional CADs were small datasets, poor quality image (digitized image, no quality standard), insufficiency of computer processing (with an impossibility to include multiples views and prior studies), the absence of dynamic

improvement (only with periodic software upgrades) and a selection of cases and images reference depending of human expertise.

1-3-2 New CAD with emerging Deep Learning algorithm

The most important factor of technical CADs evolution is the huge capability from computer processing and developing technics of deep learning and Circumvolution Neural network (CNN). Now, softwares named "AI -CAD "are able to work with very large database and improves their levels by learning from new cases. Thus, developing new Cad is not so easy even with huge computer potentiality. It needs to work with large database which can be supervised (each image is labelled by human, costly and not exempt from mistakes and approximations) or unsupervised (computers alone discern from non-labelled database the image characteristic). This requires high quality raw data in Full resolution mammogram in each view and if possible prior images which generate very high data volume. Second major problem is the complexity of the algorithm who can comprise between 30-150 layers and users need to understand how it works and assess learned parameters to avoid overfitting (learning about idiosyncratic variation but not understand the clinical impact of these variations). Four ways to implement CAD in screening process is its performance: 1) For a given high sensitivity, the major drawback is the false-positive flags who alters performance of screening test and this is the major way of research. In a recent study(20), IA CAD (cmAssit from CureMetrix) is used to reduce false positive results in screening comparing with traditional CAD (ImageChecker- Hologic). There was 69 % reduction in False positive mark lesions with AI CAD with same performance for both masses and calcifications. 2) Interpretation time with CAD increases by approximately 20 %. Reading time is one of important feature for centralized screening program but is not a major concern from French program in first reading setting, but can be a key issue for second reading session. In this

same study (20), the reading time was decrease by 64% with AI CAD comparing traditional CAD. This implies potential time saving may be use to reed 10 % more screening. The main limitation is the retrospective design of this work and the very small sample size and few cancer cases (250 FFDM) 3) Regarding the work flow, using CAD in current practice implies complete integration in the post-acquisition process especially working on the current workstation and without delay in presenting image on screen. Thus, AI CAD need to be compatible with all manufacturers and in a same way manufacturers need to open theirs system to CAD providers. 4) Finally, cost is a major concern especially in France. None CAD is actually reimbursed in France in opposite to some other countries. For implementation CAD system need to be very cost effective. One way seems interesting in term for saving money but need evaluation is the capacity for CAD to dismiss a mammogram volume from L2 (actually 4 euros for radiologist reading and 20 millions / year for L2 process).

1-3-3 Presentation of the level of development of the different algorithms existing on the market

Many AI algorithm mainly based on deep conventional network techniques has been trained during the last five years. However, their level of validation for clinical implementation differ and should be well understood before any clinical use (Table 1).

Transpara® was developed by Screen point and is actually distributed by INCEPTO® in France. This algorithm was developed on the basis on 9000 true positive and 180 000 true negative mammograms provided by different constructors. This algorithm was trained on 2D Mammo and breast tomosynthesis images. The radiologists provided a risk of malignancy based on the Breast Imaging and Reporting Data System (BI-RADS) assessment scale that ranges from 0 to 6 while this AI system offers three different decisions tools : an interactive decision tool that provides local cancer likelihood score (1-100) activated by

clicking on a specific breast region, a traditional lesion markers for computer-detected abnormalities and a proprietary examination-based cancer likelihood score with a score ranged from 1 to 10, with a score calibrated such that the number of mammograms in each category is approximately equal

In 2019, a multicenter and multireader study demonstrated the non-inferiority of this AI system to the average of 101 radiologists (AUC= 0.840 versus 0.814) (21). In this study, each dataset consisted in 2DMG acquired with different systems from the four different vendors (GE, Siemens, Hologic and Philips) and the reference standard was either histopathological analysis or follow up in a total of 2652 exams (prevalence of malignancy 653/2652: 24.6%). The AI system was more performant than 61.4% of radiologists. Sensitivity and specificity of the system was also found to be better than majority of the radiologists, but always worse than the best radiologist, which is not surprising. The authors suggest that AI system could be used as an independent stand-alone first or second reader in countries with lack of experienced breast radiologist (22). or as an interactive decision support tool (23). In this setting, the same authors compare breast cancer detection of radiologists reading 2DMG unaided versus supported by this AI system on an enriched cohort of 240 women (100 showing cancers, 40 false positive and 100 normal 2DMG) (24). In this study, the AUC of AI system was similar to the average of 14 certified radiologists but lower than AUC of a radiologist supported by AI system. Reading time per case was similar (146 second/149 seconds). Even if these results are clearly very interesting, the high prevalence of breast cancers in these validation sets must have probably overestimated the good accuracy of the AI system and further studies are needed especially for a use as a stand-alone technique. Moreover, a stand-alone approach makes asks the question of who would take ultimate responsibility for breast cancers missed (which remain the most litigious situation for medical

malpractice lawsuits) by an imperfectly performing AI algorithm? (25). This algorithm is CE and FDA approved

Other artificial intelligence algorithms were trained but not yet externally validated in clinical conditions: **Therapixel®** was created in 2013 by two researchers from the French National Institute for computer science and applied mathematics (INRIA) and took the joint 1st place of the Digital DREAM Mammography Challenge which was the biggest international competition ever organized in deep learning applied to mammography. This competition organized jointly by the National Cancer Institute, the Group Health Cooperative, the Icahn School of Medicine at Mount Sinai, the FDA, Apple and IBM, gathered about 1,200 participants to compare the best breast cancer prediction algorithms based on screening mammograms ((Accessed 20/07/2017, at (23)) This challenge was based on a set of 320 000 2DMG with 1200 breast cancers (Prevalence : 3.7/1000 patients). After four consecutive rounds, Therapixel reached the co first place of the challenge with an accuracy of 75%. However, no comparison with radiologist interpretation was available.

Mammography Intelligent Assessment (MIA)® is an AI algorithm developed by Kheiron as a part of an NHS grant. This algorithm was trained on more than one million of Mammography images from United Kingdom breast screening program and validated in a retrospective multicenter study with 3854 cases from 4 UK screening sites (Prevalence 6.9%). MIA displayed a sensitivity of 85%-97% and a specificity of 50% to 94% in the different 4 sites. No comparison was available with radiological interpretation in this study only presented in an industrial workshop at RSNA 2018. This algorithm is CE approved and waiting for an FDA approval

Arterys® is a big American firm located in San Francisco which is also developing an AI algorithm with French collaborations to train and validate their new algorithm. They are on a preliminary step of training

AI Research issues

1- To validate AI model in French population

2- To test the ability of AI system integrating the comparison with previous mammograms

3- To test the performance of AI algorithm on combined two views MG analysis

4- To compare 2nd reading versus 1st reading + AI software

5- To test AI algorithm on 2DMG according the different vendors

Chapter 2: Breast tomosynthesis: The future of mammography

2-1 Are there any AI algorithm trained and validated on breast tomosynthesis?

According to the 6 published meta-analysis on the value of breast tomosynthesis for screening (26–31), breast tomosynthesis added to 2DMG has been demonstrated to be more sensitive and more specific than 2D mammography with higher detection rate of invasive cancers. Moreover, recall rate was demonstrated in retrospective studies. Thus, in the next future, breast tomosynthesis will become the standard technique for breast screening. Thus, the would be logical that AI algorithm be trained and validated as soon as possible on breast tomosynthesis data.

The interest of Artificial Intelligence in Digital Breast Tomosynthesis is three-folds: 1) To detect more lesions by keeping an acceptable rate of False positive 2) To improve the characterization of breast lesion, whatever they were detected by mammography or by DBT 3) To decrease the time of reading, which may be relevant for a screening method which has been shown to double the reading time. There are still several issues in the use of Artificial Intelligence in Digital Breast Tomosynthesis: must the model use projection views images, DBT reconstruction slices, combination of both, 3D reconstructed volume or the derivatives of the reconstructed images such as synthetic mammograms ? must AI systems use large data

collected from mammography (32) or should the model be built exclusively with DBT data ?, how to use the region-based conventional neural networks (RCNN) which is computationally expensive and is very time-consuming in a process with a great set of data such as DBT without overly slowing down the reading and what is the value of faster RCNN (33).

To the best of our knowledge, two societies have industrially developed IA in DBT: Icad Profound AI and ScreenPoint which has developed Transpara. **Transpara™® for DBT** performs analysis and interpretation using the full 3D information from the DBT volumes. Similar to Transpara™ for mammography it delivers interactive decision support including the detection of soft-tissue lesions and calcifications, interpretation of suspicious regions and automated linking of MLO and CC views. Furthermore, Transpara™ for DBT uses synthetic images for intelligent navigation in both CC and MLO views. However, no results have been shown in the use of Transpara for DBT in clinical conditions

Recently, E.Conant presented in ECR congress 2019 the first results of a new version of a commercial software named **Icad Profound AI**® which was trained and validated on breast DBT. In this retrospective, fully-crossed, multi-reader with 24 radiologists study based on 260 cases with 127 dense breast and 133 non-dense breast, they demonstrated a better performance of radiologists with AI algorithm than without AI in both dense and non-dense breasts. Sensitivity and specificity significantly increased by 7% and 9.9% in dense breasts and 9% and 4% in non-dense breasts. Moreover, reading times decreases for the 24 radiologists with AI by 57.4% in dense breasts and 47.6% in non-dense breast. The main limit of this study is the fact that only one vendor was represented, and these results must be consolidated on a cohort with all vendors represented. This algorithm has been trained on a database of 12.000 DBT cases including 4000 proved cancer cases. It is able to detect and diagnose calcified and non-calcified suspicious lesions. According the workstation version, detection sensitivity may be changed by the radiologist, allowing 3 different levels: low level

(88% overall Se, and 72% Sp; medium level: 91 % overall Se and 59 % Sp; and high level 95 % overall Se and 31% Sp). This program provides two types of information: case score (probability that the entire case should be malignant, and lesion score (probability that a lesion marked should be malignant), Probability score increases from 0 to 100 %. A higher score indicates a higher level of confidence in the malignancy of the detection or case. The scores were calibrated on an enriched cohort with a 25% prevalence of cancer. Thus, these scores should be interpreted as a probability to detect a cancer correctly interpretated in a population of 25% cancers and 75% non-cancers. This is a main limitation for an application for breast screening as positive predictive value is clearly probably highly overestimated.

2-1 Are AI algorithm useful to optimize synthetic mammography ?

Two-view breast tomosynthesis is better than one-view tomosynthesis (34). Thus, SM is really necessary to limit radiation dose. Since Synthetic mammography will become the standard in DBT, in order to avoid 2D acquisition and subsequent radiation exposition, it is mandatory to know the tool used by the different constructors to perform synthetic mammography. A recent study (Oslo Trial) (8) in which the sensitivity of 3D +Synthetic Mammography (SM) was not superior to the sensitivity of 2D, that is the first one to conclude this "negative" result has underlined the potential difference of reliability of SM among the different constructors. SM uses more and more IA, as shown by the evolution of SM in Hologic, brand which has the most published experience in DBT in general and in SM in particular.

The first generation of SM was an algorithm using machine learning techniques to generate a synthesized 2D image from the 100-micron tomosynthesis reconstructions. This is much more than a basic MIP function of a volume. In order to avoid superimposed tissue mimicking suspicious area and to improve visibility of true structures, the C-View algorithm will analyze each tomo slice as well as adjacent slices (above and below) to differentiate

normal structures from suspicious ones and to find microcalcifications barely visible in conventional 2D due to breast thickness attenuation.

The second version of SM was built on advanced machine learning technique, it operates on Hologic 70µm tomosynthesis reconstructions to generate a synthesized 2D image. In order to avoid superimposed tissue mimicking suspicious area and to improve visibility of true structures, the I2D algorithm will analyse each tomosynthetic slice as well as adjacent slices (above and below) to differentiate normal structures from suspicious ones and to find microcalcifications barely visible in conventional 2D due to breast thickness attenuation. These findings will be better depicted in I2D thanks to the advanced AI identification, which increases the conspicuity of the identified lesions. The higher resolution of the tomosynthesis reconstructions, along with the advanced AI algorithms, theoretically enables better identification of suspicious lesions while reducing the enhancement of false positives, compared to synthetic 2D MG. In addition, IA was used to take in account details of the images such as breast density and parenchymal arrangement when creating the synthesized image, in order to permit a look of the synthesized image being very close to a conventional 2D image, while at the same time maintaining the increased conspicuity of suspicious lesions.

AI Research issues

1-To compare the accuracy of AI algorithm of one view/two views of breast DBT by assessing the accuracy of reading with and without AI of one view and 2 views breast DBT with synthetic reconstruction

2-To evaluate the added value of AI in terms of characteristics of cancer detected (size, grade)

3- To evaluate the modification due to IA in terms of false positive and of false positive requiring a biopsy (since it is a potential issue of DBT

4-To evaluate the time of reading with and without DBT

5- To optimize synthetic MG quality thanks to AI algorithm

6- To test AI algorithm on synthetic MG and according the different vendors

Chapter 3 : Organization of breast screening in France in 2019 : What are the expected impact of artificial intelligence ?

In France, the number of breast screening mammography performed each year represents **2,5 millions of women**. Digital mammography (DR) represents 70% of MG units as well as Computed Radiography represents 30% of MG units. The number per regional center and per year is between **5000 to 70 000 mammograms**. The number of 1st reader radiologists (>500 mammograms read per year) per department is between 5 to 200 while the number of 2nd readers (>2500 mammograms read per year) is between 5 to 50. 2nd reader interpretation is performed with the previous mammograms, the knowledge of clinical examination

In France, **175 000 mammograms** (7%) of breast screening are considered as **abnormal at the 1st reading** (rated BI-RADS 0, 3 ,4, 5) while **2 325 000 mammograms** are considered as **normal** at 1st reading and are referred for 2nd reading. In addition, 7000 mammograms (4%) considered as abnormal before complementary incidences and or ultrasonography (performed in the same time than 1st reading) and are reclassified at normal at the end and thus also referred for 2nd reading. Thus, **2 332 000 "normal mammograms »** are referred to regional screening center for 2nd reading in which 1% are reclassified as "suspicious" (n=23 320). After complementary incidences and or ultrasonography performed in these mammograms considered as positive but the 2nd reader, 22% are confirmed to be abnormal and will correspond to **0 ,3 -0 ,4 /1000 cancers**. Finally, French breast screening program detects **7 cancers per thousand screened women** including a L2 contribution /of 5 % of breast cancer detected in 2019. This percentage has decreased during the last years form about 10% to 5%.

Our screening system is very different from other countries and few experimentations can be implemented in France. First, screening attendance rate is very heterogeneous varying from less than 30% in Paris to more than 65 % in some department from west or center of the country. This can be partially explained by the level of individual screening in some regions especially in South of France and in Ile de France, but there are also huge differences between towns and even quarters. The attendance rate varies also between age of screening with higher participation for women between 55-65 years old and lesser participation for older women after 70 years old. There are also multiple reasons for this heterogenous participation especially in socio-economic life conditions. A first application of Using AI from national French medical-database cross with other like taxes or family data-base or unemployment benefits registry may be helpful to identify and communicate specifically with these populations.

A second application of artificial intelligence is related to our second reading. In France, the Program is organized by regions and done by experimented radiologists but with a great difference in comparison with other countries: In our program, radiologists perform immediately complementary incidence or sonographic examination named immediat Diagnosis work up, if they detect clinically or MG abnormality. Thus, with this French specificity, the R2 recall rate is only 1,1% of all attenders, sending back to the first radiologist for a Differed Diagnosis work up. 22% of these recalled women will have a positive diagnosis resulting in detection of 6% of all cancers or 0,4 cancer for 1000 women. The R2 reading performance is variable between regions with unclarified reasons. The second reading is always done on screen film (actually there are some small experimentations for dematerialization without publishing data) with a heavy cost (film transfert to the centralized R2 office, film display on light box by technician, radiologist fees, data monitoring, screening

results and film sending back to women and medical practitioner). The total cost of this process is about 20 Millions per year

The value of second reading is still in debate especially in this period after the recent governmental decision to prohibit screen film and the ongoing transfer from CR to DR as CR system detects less cancer than DR system. Moreover, the potential implementation in screening program of Digital Breast Tomosynthesis which improve R1 cancer detection rate and decrease the recall rate is also an argue against R2. Results of OSLO trial just published shows that DM plus DBT performs better than FFDM and double reading (8)

The challenges for using CAD as a second reader are the necessity of a great improvement in terms of False positive mark, the interpretation time with old CAD that increased by approximately 20 %. (Reading time is one of important feature for centralized screening program but is not a major concern from French program in first reading setting because a low volume of examination done per radiologist and per day), the development of an easy work flow (Using CAD in current practice implies complete integration in the post-acquisition process especially working on the current workstation and without delay in presenting image on screen. Thus, AI CAD need to be compatible with all manufacturers and in a same way manufacturers need to open theirs system to CAD providers) and finally the cost. Cost is a major concern because none CAD is actually reimbursed in France in opposite to some other countries. For implementation CAD system need to be very cost effective. One way seems interesting for saving money but need evaluation is the capacity for CAD to dismiss a significant mammogram volume from R2 (actually 4 euros for radiologist reading and 20 millions / year for R2 process).

Thus, in France, breast screening organization may benefit from the development of artificial intelligence in these different fields:

- **improvement of image quality** helping technician to optimize positioning or blurring and validating this step before sending mammography to radiologist for interpretation,
- **better selection of women** who may benefit of **breast ultrasonographic examination**. Actually Breast US is performed in 20% of all attenders after normal mammography for dense breast and is the single examination done in 76 % women during BDI, but detect only 2,4 % cancers. We need software who can determine not only global density but also, like radiologist do, focal high breast density which is sometime good indication for breast ultrasound. Other way of research will be the development of IA CAD for breast ultrasound and there are 2 devices actually potentially in use. One for ABUS system called Qview Medical (https://www.qviewmedical.com/) and another cloud based CAD system developed by Koios Medical able to works on PACS image (Koios medical https://koiosmedical.com/solutions/) but there is no study published at term.
- **better first round triage** and possibly dismissing some mammography from Second reading, lowering the cost and improving the results' delay to patient and Medical referee. One interesting study, published in April 2019 (35) tested the ability of an AI algorithm (Transpara, Screenpoint) to exclude exams with the lowest likelihood ratio of malignancy. This study was conducted on 2562 examinations and demonstrated that excluding mammograms with a score 1 or 2 allowed to decrease of 17% the number of exams and 5 % of false positive case (minoring the recall rate) with only missing 1% of cancers. Radiologist's performance was unchanged in the new cohort with higher prevalence of breast cancer.

- **improved level of general radiologist** which are the majority in our decentralized program and so benefit for all the women and reduce the contribution of second reading.

AI Research issues

1) Organization issue

- To identify in French national medical database, reasons to non-participating at breast screening program
- To analyze the Cancer detection rate difference at local level (town, quarters) to better select higher risk women
- To establish predicting parameters for Improving cost effectiveness of organized breast screening program

2) Reading Issues

- To predicts comorbidity (Vascular calcification correlated with heart attack)
- To improve screening performance of different experienced level radiologist?
- To better select patient who need US evaluation

Chapter 4: What do we need in France to develop artificial intelligence for breast screening?

In the journal Nature last year, Dexter Hadley estimated that screening algorithms should be trained on millions of mammograms suggesting that AI researchers should embrace bitcoin technology. (15 mars 2008 Nature). Actually, many validation cohorts are enriched with a great proportion of cancer-positive mammograms than detected in clinical screening routine. This type of cohort may induce bias with a greater radiologist scrutinity that may increase

recall rate and decrease accuracy. We can split in 3 main phases the development of an AI algorithm for detect breast can using deep learning CNN: First, the model is built on an enriched data cohort with a high volume of breast cancers, ideally including all types of cancers on mammography (spiculated mass, round mass, cluster of microcalcifications, architectural distortion …). Most of time, there is another data set in the same cohort not used for the training which will be used for internally validation. Second, the model needs to be externally validated in another cohort with a lower prevalence of cancer and possibly with more subtle cancers to improve its accuracy. Then, the last step is to test the model on an independent data set with a prevalence representative of screening population before clinically validate the model in a randomized trial comparing the accuracy of the model with those of the radiologists. We need all these steps because incremental improvement in the AUC is not directly translatable to improved patient outcomes in the clinical setting. It is uncertain what proportion of exams that a commercial AI system would flag as having more than 2% malignancy, requiring additional diagnostic workup under our current clinical practice thresholds. Moreover, we learned from the experience of computer-aided detection in mammography that adopting promising new technologies too quickly could be a costly mistake; later found to lead to more false positives without improved cancer detection (36)

Thus, an algorithm needs more than one million of 2D MG/DBT to truly demonstrate its efficacy and the development of dedicated platform must be elaborated to build larger validation data sets more representative of a screening population. In France, the government created in 2018 the Health Data Hub to pull in same environment data from clinical data public/private practice. In France, breast screening clinical data and follow up are recorded in dedicated regional screening cancers dedicated to screening program and images are available Health Data Warehouse either in public or private centers. Thus, a connection needs to be created

4.1 What can we expect from regional breast screening centers?

In breast screening centers, are collected all data related to 2nd reading including epidemiological informations such as the age, menopausal status, personal history of breast biopsy, familial history of breast cancer, the presence/absence of an abnormal MG image, the side, the type of MG abnormality (cluster of microcalcifications, architectural distortion, mass or asymmetric density) and breast density (according BI-RADS classification) as well as follow up data for all abnormal MG (type of biopsy, type of surgery, histopathological findings)

Moreover, the actual role of breast screening center is to record all cancers appeared after normal 1st or 2nd reading (named interval cancers) which is a major data to evaluate the presence/absence of positive impact of a breast screening program.

The regional breast screening center collects also 1st reading data regarding personal history context, clinical examination, the type of vendors, type of MG abnormalities, the number of ultrasonography performed after normal MG, the type of supplementary incidences performed (including breast tomosynthesis), the type of pathological findings detected.

Thus, a lot of data are available in the regional breast screening centers recently grouped from departmental breast screening cancers connected with administrative informations of each patient (name, birth date, "numero de securité sociale"..)? These huge amounts of very informative data should be correlated with "entrepot de données de santé" where are located mammograms such as DRIM IA or APHP

4.2 What can we expect from an hospital Health Data Warehouse? The APHP model

L'Assistance Publique des Hôpitaux de Paris (APHP) is a unique international structure that groups 39 public hospitals federated in a unique legal entity. Each day, this information System collects many and various medical data in a wide variety of software and databases used to monitore the patients' care courses. Their exploitation for artificial

intelligence research especially in the domain of deep learning needs new new big-data approaches and the creation of a specific database usually named Health Data Warehouse (EDS in french).

With more than 7 millions of patients and more than 20 millions of radiological examinations in DICOM format (Figure 1), the APHP via l'EDS APHP is a unique database in Europe for biomedical research in imaging. Actually, around 150 000 mammography are available in EDS (from 2010 to 2018) in more than 75 000 patients with more than 13 000 patients with a 2year follow up by mammography. Moreover, more than 27 000 digital breast tomosynthesis are available (period 2015-2018) in more than 26 000 patients. These mammography have been performed on more than 3 manufacturers with basic french image quality. Thus, several industrial partnerships have been contracted to train and validate different types of AI algorithms, and a large study named EZ mammo is ongoing.

In this setting, the creation of the EDS database and specifically for medical picture data (DICOM data) was associated with the definition of legal and ethical rules that take account of data specificity, especially on rule access of data for intern and extern scientist or industrial. The hospital created a process to inform patient of possible re-use of her data and allow him to opposite, define the limit of use of data and the rule of access and define data valorization rules. In APHP, a specific committee named CSE (Comité Scientifique et Ethique) composed by medical, patient and technical members is charged to validate the rule respect and evaluate each access request of data collected in EDS APHP. As all data are unified in unique structure, the scientist can't be used the database directly because a risk exists if multiple and big request are performed with a possibility of stress on this database that can negatively impact the operation of care services. Thus, specific research databases are created by collecting periodically and automatically the data. During the collect of research data in EDS, you can apply a specific pipeline for data like de-identification process. In EZ

mammo study, deep learning technique are used to automatically search words in the imaging and pathological report in order to constitute different cohorts with different prevalence of malignancy.

Regarding DICOM data at the APHP scale, this process would require to duplicate all DICOM file in PACS. As the duplication is not recommended, the solution tested by EDS APHP consist to collect only minimal metadata information of DICOM like id, description or modality information. When the scientist builds her research cohort it can used this metadata to select DICOM of interest for his project. Once done, the DICOM data was collected by a PACS to PACS (C-MOVE) process. During this process the DICOM file metadata or pixel will be also de-identified.

At the end of process the scientist must be to have access at this research database that is composed by structured data and a DICOM repository. This data will be exported out of hospital server but for security and valorization aspect APHP has chosen to not export data. For this, it created a data management and research platform that permit to data-scientist to analyze data with a megadata cluster without exporting them off the secure servers of the APHP (Figure 2).

During research projects, the database will undergo a lot of transformation and will need to be structured to answer to a specific question. For example, the adding of image annotations needs to be reintegrated in the database that will be enriched with the different research projects.

Generally, the radiologist can write annotation information in private tag DICOM. This means all DICOM file must be re-sended to PACS and if no annotation management is designed, the annotation in private tag may be erased if another annotation is created. Many solutions in development like an extension of MedInria or a solution of Inception start-up want to create a specific annotation data format that not need to extend DICOM format. In EDS AP-HP a

specific collect annotation web service compatible with PACS solution named SPHERE will be added. At the end of data labellisation and research processing, the data could be used to test artificial intelligence algorithm.

In brief, the challenges for a large data PACS adapted solution is to permit a high speed transfer of DICOM file and a less cost storage, to define specific data model and structure for medical imaging data researcher, to find and build solution to assist and secure annotation process, and to report and facilitate the creation of re-usable dataset.

4.3 What can we expect from individual breast screening: An example of radiologist's initiative: DRIM France IA

In France a database exists regarding individualized breast screening performed in private structures outside form the organized program named SENOLOG. This database contains all informations from mammographic report (clinical history, clinical examination, radiological features and BI-RDS classification) and especially the follow up of mammography classified Bi-RADS 1 or 2. The first challenge for private radiologists will be to standardize our work methods as a huge heterogeneity exists regarding to MG units, archiving methods, report standardization. Recently, an initiative for radiological community named DRIM France IA consisted in platform dedicated to radiological examination and designed to train and test artificial intelligence algorithm in a centralized system. Several questions are until now unsolved: Will the anonymization be done by the radiologists, will it be done by DRIM, or done by the software providers? How are the radiologists going to get the flow of data? Will it be DRIM that will recover and store the data via a common channel or are the builders master installer? If it is DRIM that retrieves the data, besides the major problem of the RGPD and the data security, it will require storage servers with huge capacities.

As Dexter Hadley proposed in nature in March 2018, a system should be built that allows people to share their medical data with researchers easily and securely — and retain

control over it. Their method, which is based on the blockchain technology that underlies the cryptocurrency Bitcoin, will soon be put to the test ». Perhaps, this could be a solution to build massive database that are needed to train deep learning algorithms. In any case, there should be a complementarity and a partnership between the radiologist and the AI.

AI Research issues

1- To create universal annotation Tools

2- To develop IA tool for increasing virtually enriched cohort

3- To structure platform dedicated to radiological volume

Chapter 5: Better assessing the risk of breast cancer: Artificial intelligence and personalized program

In France, national recommendations for the management of high-risk women were published in 2014 by the Haute Autorité de Santé (HAS) and are presented in Table 2.
For intermediate risk women no strict recommendation exists, and a defiance exists for the women with normal risk of breast cancer on the actual breast screening program. The future will probably include a personalized approach depending on the risk of developing a breast cancer. For the women with the lowest risk of breast cancer, the interval between two rounds of mammographic screening might increase whereas for women with a higher risk of breast cancer, breast screening might be annually performed. Two international trials have began in Europe and United States. WISDOM TRIAL (USA) and MyPEBS (Europe). Based on an European H2020 grant (http://mypebs.eu/fr/) (37). The main endpoint in this study is the ability to decrease the number of cancer stage higher that 2 of a personalized breast screening in comparison with the conventional approach: this trial will include all women between 40 and 70 year-old without any risk factor of breast cancer. 85000 women will be invited in 5

countries (France, Belgium, Italy, Israël, United-Kingdom). In France, 20 000 women will be concerned. Women will be randomized in two arms: standard arm (screening according each country recommendation), and risk-stratified arm. The risk will be evaluated with BCSC score (Breast Cancer Screening Consortium) or Tyrer-Cuzick if there is more than one familial breast cancer at the first degree. Women in this arm will be proposed to perform a score of 300 polymorphisms and a risk prediction score. The final risk score will determine the strategy depending on the risk (Table 2). This prediction risk model based on artificial intelligence named MammoRisk® has been developed by a French start up named Predlife ® and includes: 1st degree breast cancer, personal history of breast biopsy, breast density. In this setting, the software algorithm was developed using a concept of Manhattan distance to compare a patient's mammographic image to reference mammograms with an assigned BMD category. Reference databases were built from a total of 2289 pairs (cranio-caudal and medio-lateral oblique views) of 2D full-field digital mammography (FFDM). A validation set of additional 800 image pairs was evaluated for BMD both by the software and seven blinded radiologists specialized in breast imaging. The software showed a substantial agreement with the radiologists' consensus (unweighted kappa = 0.68, 95% CI 0.64-0.72) when considering the four breast density categories, and an almost perfect agreement (unweighted kappa = 0.84, 95% CI 0.80-0.88) when considering clinically significant non-dense (A-B) and dense (C-D) categories. (38)

A third risk-based screening trial has been led in UK between 2011 and 2013 (PROCAS study). Women who were invited for their 3-yearly mammogram were eligible for this trial. 53.596 women had been recruited, (37% of the 68 % of women attending for NHSBSP screening). A total of 10.000 women in PROCAS were recruited, by invitation, to a DNA collection study using saliva. The risk information data comprised age at menarche/menopause, HRT, family history of breast cancer, weight/height, breast biopsies.

Ten-year risks were identified. Breast density assessment was done visually for analog MG, and with 2 software (Volpara and Quantra). 18 SNPS were identified. 632 prospective breast cancers had occurred in 53.184 women. For density evaluation, Visual assessment (VAS) gave the best prediction with an OR of 3.59. Using a Mammogram Density-adjusted TC model in PROCAS improved risk stratification (AUC = 0.6) and identified significantly higher rates (4.7 per 10,000 vs. 1.3 per 10,000; $p < 0.001$) of high-stage cancers in women with above-average breast cancer risks. The model performed particularly well in predicting higher stage 2+ invasive cancer. This combined approach using Tyrer-Cuzick, mammographic density assessment and polygenic risk score provides accurate risk stratification, particularly for poor prognosis cancers. (39)

Recently, Yala et al. published a breast cancer risk model based on deep learning mammography (40) on 88994 consecutive screening MG in 39571 women. In this study, the authors combined traditional risk factors and mammograms in a hybrid DL model and demonstrated this model was better with AUCs of 0.70 than classical Tyrer-Cuzick model (0.62; $P < .001$) or a risk-factor-based logistic regression model (0.67; $P = .01$).

AI Research issues

- To develop AI model that helps to better define breast risk assessment select patient for different way of screening (HR Patient, breast density B3 lesion, etc) – Better define intermediate risk patients
- To build AI model that integrates of all epidemiological and clinical data for better (physical activity..

CONCLUSION

As C.Kuhl said during ISMRM In Montreal 2019 « A golden age is currently underway for deep learning and imaging-based risk prediction of breast cancer, enabling generations of imaging biomarkers to help make better recommendations for patients"

Finally, the main question is how the radiologist will cohabit with artificial intelligence in the future? Maybe we have to imagine this as an airplane, with human pilot and atomic one. The radiologist will always be the last bastion to AI errors. The radiologist will have to know the strengths and weaknesses of the AI so that the union of better will be stronger than either taken individually. This also raises the question of training for residents: like simulators that allow pilots to train to handle critical situations, will the radiologist of tomorrow have to be even better trained on the most complicated cases? This huge data base will give the opportunity for residents to use all the mammogram to train themselves. Senior radiologists should be able to create annotations on mammograms even if the AI said everything is normal. This will help the AI to be stronger and better, faster than the machine would do alone.

References

1. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. N Engl J Med. 1994;331(22):1493–1499.
2. Trister AD, Buist DSM, Lee CI. Will Machine Learning Tip the Balance in Breast Cancer Screening? JAMA Oncol. 2017;3(11):1463–1464.
3. Fenton JJ, Taplin SH, Carney PA, et al. Influence of computer-aided detection on performance of screening mammography. N Engl J Med. 2007;356(14):1399–1409.
4. Lehman CD, Wellman RD, Buist DSM, et al. Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection. JAMA Intern Med. 2015;175(11):1828–1837.
5. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 2017;2(4):230–243.
6. Perou CM, Sørlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747–752.
7. Codari M, Schiaffino S, Sardanelli F, Trimboli RM. Artificial Intelligence for Breast MRI in 2008-2018: A Systematic Mapping Review. AJR Am J Roentgenol. 2019;212(2):280–292.
8. Skaane P, Bandos AI, Niklason LT, et al. Digital Mammography versus Digital Mammography Plus Tomosynthesis in Breast Cancer Screening: The Oslo Tomosynthesis Screening Trial. Radiology. 2019;291(1):23–30.

9. Brnić Z, Hebrang A. Breast compression and radiation dose in two different mammographic oblique projections: 45 and 60 degrees. Eur J Radiol. 2001;40(1):10–15.
10. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev. 2006;15(6):1159–1169.
11. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. N Engl J Med. 2007;356(3):227–236.
12. Brandt KR, Scott CG, Ma L, et al. Comparison of Clinical and Automated Breast Density Measurements: Implications for Risk Prediction and Supplemental Screening. Radiology. 2016;279(3):710–719.
13. Kerlikowske K, Zhu W, Tosteson ANA, et al. Identifying women with dense breasts at high risk for interval cancer: a cohort study. Ann Intern Med. 2015;162(10):673–681.
14. Sprague BL, Conant EF, Onega T, et al. Variation in Mammographic Breast Density Assessments Among Radiologists in Clinical Practice: A Multicenter Observational Study. Ann Intern Med. 2016;165(7):457–464.
15. Ciritsis A, Rossi C, Vittoria De Martini I, et al. Determination of mammographic breast density using a deep convolutional neural network. Br J Radiol. 2018;20180691.
16. Lehman CD, Yala A, Schuster T, et al. Mammographic Breast Density Assessment Using Deep Learning: Clinical Implementation. Radiology. 2019;290(1):52–58.
17. Moshina N, Sebuødegård S, Lee CI, et al. Automated Volumetric Analysis of Mammographic Density in a Screening Setting: Worse Outcomes for Women with Dense Breasts. Radiology. 2018;288(2):343–352.
18. Kontos D, Winham SJ, Oustimov A, et al. Radiomic Phenotypes of Mammographic Parenchymal Complexity: Toward Augmenting Breast Density in Breast Cancer Risk Assessment. Radiology. 2019;290(1):41–49.
19. Keen JD, Keen JM, Keen JE. Utilization of Computer-Aided Detection for Digital Screening Mammography in the United States, 2008 to 2016. J Am Coll Radiol. 2018;15(1 Pt A):44–48.
20. Mayo RC, Kent D, Sen LC, Kapoor M, Leung JWT, Watanabe AT. Reduction of False-Positive Markings on Mammograms: a Retrospective Comparison Study Using an Artificial Intelligence-Based CAD. J Digit Imaging. 2019;
21. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, et al. Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists. J Natl Cancer Inst. 2019;
22. Gilbert FJ, Astley SM, Gillan MGC, et al. Single reading with computer-aided detection for screening mammography. N Engl J Med. 2008;359(16):1675–1684.
23. Hupse R, Samulski M, Lobbes MB, et al. Computer-aided detection of masses at mammography: interactive decision support versus prompts. Radiology. 2013;266(1):123–129.
24. Rodríguez-Ruiz A, Krupinski E, Mordang J-J, et al. Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System. Radiology. 2019;290(2):305–314.
25. Arleo EK, Saleh M, Rosenblatt R. Lessons learned from reviewing breast imaging malpractice cases. J Am Coll Radiol. 2014;11(12 Pt A):1186–1188.
26. Lei J, Yang P, Zhang L, Wang Y, Yang K. Diagnostic accuracy of digital breast tomosynthesis versus digital mammography for benign and malignant lesions in breasts: a meta-analysis. Eur Radiol. 2014;24(3):595–602.
27. Houssami N, Lång K, Bernardi D, Tagliafico A, Zackrisson S, Skaane P. Digital breast tomosynthesis (3D-mammography) screening: A pictorial review of screen-detected cancers and false recalls attributed to tomosynthesis in prospective screening trials. Breast.

2016;26:119–134.
28. Yun SJ, Ryu C-W, Rhee SJ, Ryu JK, Oh JY. Benefit of adding digital breast tomosynthesis to digital mammography for breast cancer screening focused on cancer characteristics: a meta-analysis. Breast Cancer Res Treat. 2017;164(3):557–569.
29. Phi X-A, Tagliafico A, Houssami N, Greuter MJW, de Bock GH. Digital breast tomosynthesis for breast cancer screening and diagnosis in women with dense breasts - a systematic review and meta-analysis. BMC Cancer. 2018;18(1):380.
30. Marinovich ML, Hunter KE, Macaskill P, Houssami N. Breast Cancer Screening Using Tomosynthesis or Mammography: A Meta-analysis of Cancer Detection and Recall. J Natl Cancer Inst. 2018;110(9):942–949.
31. Marinovich ML, Macaskill P, Bernardi D, Houssami N. Systematic review of agreement between tomosynthesis and pathologic tumor size for newly diagnosed breast cancer and comparison with other imaging tests. Expert Rev Med Devices. 2018;15(7):489–496.
32. Samala RK, Chan H-P, Hadjiiski L, Helvie MA, Wei J, Cha K. Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography. Med Phys. 2016;43(12):6654.
33. Fan M, Li Y, Zheng S, Peng W, Tang W, Li L. Computer-aided detection of mass in digital breast tomosynthesis using a faster region-based convolutional neural network. Methods. 2019;
34. Zackrisson S, Lång K, Rosso A, et al. One-view breast tomosynthesis versus two-view mammography in the Malmö Breast Tomosynthesis Screening Trial (MBTST): a prospective, population-based, diagnostic accuracy study. Lancet Oncol. 2018;19(11):1493–1503.
35. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, et al. Can we reduce the workload of mammographic screening by automatic identification of normal exams with artificial intelligence? A feasibility study. Eur Radiol. 2019;
36. Fenton JJ, Lee CI, Xing G, Baldwin L-M, Elmore JG. Computer-aided detection in mammography: downstream effect on diagnostic testing, ductal carcinoma in situ treatment, and costs. JAMA Intern Med. 2014;174(12):2032–2034.
37. Delaloge S, Bachelot T, Bidard F-C, et al. [Breast cancer screening: On our way to the future]. Bull Cancer. 2016;103(9):753–763.
38. Balleyguier C, Arfi-Rouche J, Boyer B, et al. A new automated method to evaluate 2D mammographic breast density according to BI-RADS® Atlas Fifth Edition recommendations. Eur Radiol. 2019;29(7):3830–3838.
39. Evans DGR, Harkness EF, Brentnall AR, et al. Breast cancer pathology and stage are better predicted by risk stratification models that include mammographic density and common genetic variants. Breast Cancer Res Treat. 2019;176(1):141–148.
40. Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction. Radiology. 2019;292(1):60–66.

TABLES

Table 1: Degree of validation of the main available intelligence artificial algorithms for 2D breast screening based on deep learning convolutional neural networks

	Internal validation	External validation Publications	Positive points	Negative points
Transpara (Screenpoint/ incepto)	Over 9000 MG with cancer (one-third of which are presented as lesions with calcifications) and 180000 MG without abnormalities. The MG originate from devices from four different vendors (Hologic; Siemens; General Electric; Philips; Fujifilm) and institutions across Europe, the United States, and Asia. Validation is performed on an independent dataset representative of screening population with enriched prevalence of cancer	• Rodriguez-Ruiz et al, Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System, Radiology 2019; 290:2, 305-314 • Rodriguez-Ruiz et al, Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists, Journal of the National Cancer Institute 2019;djy222 • Rodriguez-Ruiz et al. Can we reduce the workload of mammographic screening by automatic identification of normal exams with artificial intelligence? A feasibility study, Eur Radiol (2019). (37) • Lång K et al. Can artificial intelligence identify normal mammograms in screening? SB-0696, Presented at ECR 2019, Vienna	Provides 3 different outputs to aid radiologists in detection, classification and triaging/workflow optimization. This system can be applied to processed (ie, "for presentation") DM images and DBT volumes (no raw data) Multi-vendor Support for combined exams (exams with both 2D mammography - no synthetic - and 3D digital breast DBT)	Currently, the AI system does not use information from prior mammograms (when available). Two different modules to detect calcifications and soft tissue lesion
Therapixel	Multi-constructeurs : HOLOGIC, GE, SIEMENS, PHILIPS, FUJI Origin USA and Europe (France, UK) Quantity : plusieurs centaines de milliers Enriched cohort with annotations of all cancers Validation is performed on an independent dataset with a prevalence of cancer : 1cas /10	Data challenge : 320 000 cases on a dataset representative of screening population with enriched prevalence of cancer A new study is on going – Multicentric design	Enriched cohort Multicentric cases	
Icad	Origin North America and Europe 2000+ cases to train and internally test the algorithm, including 4000+ cancer cases. Tomosynthesis vendors included GE, Hologic and Siemens. Roughly 50% of the cases were used for training, and 50% for testing. Prevalence of cancer is 50% cancer 50% non-cancer	A separate independent regulatory set was used for the reader study and reporting the standalone performance, as in the user manual. Reader study was performed by independent company, Intrinsic Imaging		

Table 2. Recommandations de dépistage spécifique chez les femmes à haut risque de cancer du sein

Situation de haut risque	Recommandation	Niveau de preuve
Variant germinal délétère des gènes BRCA1, BRCA2, PALB2	IRM annuelle dès 30 ans et jusqu'à 65 ans Mammographie annuelle dès 30 ans Examen clinique bi-annuel Cf. recommandations nationales INCa et réseaux de prise en charge, HAS 2014 [3]	Grade A, niveau 1
Variant germinal délétère d'autres gènes de prédisposition établis associés à un haut risque	Identique à BRCA Cf. gènes identifiés par le groupe génétique et cancer (http://www.unicancer.fr/recherche/les-groupes-recherche/groupe-genetique-et-cancer-ggc, accès 21/03/2019)	Grade C, niveau 4
Histoire familiale avec probabilité de mutation > 20% mais pas de mutation identifiée	Identique à BRCA Cf. recommandations HAS 2014 [3]	Grade C, niveau 4
Antécédent de radiothérapie thoracique avant 25 ans	IRM annuelle dès 30 ans Mammographie 1 incidence dès 30 ans Cf. recommandations HAS 2014 [3]	Grade B, niveau 2 (études de cohortes)
Antécédent personnel de carcinome du sein	Mammographie annuelle	Grade B, niveau 2
Antécédent de lésion histologique atypique du sein [4]	Mammographie annuelle pendant 10 ans puis dépistage standard Cf. recommandations HAS 2014 [3]	Grade C, niveau 3
Autre situation de haut risque identifiée par un score composite [5]	Pas de recommandation existante, essais cliniques en cours	

Table 4 : Type de prise en charge dans étude MyPebs en fonction du risque de cancer du sein à 5ans

- En cas de risque faible (< 1 % à 5 ans) : une mammographie sera réalisée au début de l'étude puis à 4 ans (fin de l'étude)
- En cas de risque moyen (1-1.67 %) : le dépistage standard sera proposé (mammographie tous les deux ans de 50-74 ans en France)
- En cas de risque élevé (1.67-6%) : comparable à une histoire personnelle de cancer du sein d'hyperplasie atypique, une mammographie tous les ans, plus une échographie en cas de forte densité mammaire sera proposée.
- En cas de risque très élevé (> 6%) : comparable à être porteur d'une mutation génétique héréditaire BRCA1 ou BRCA2, une mammographie tous les ans plus une IRM seront proposées.

Figure 1: EDS- APHP