Faculty of Medicine Ramathibodi Hospital,

Mahidol University

Predicting BI-RADS by applying deep learning to breast ultrasound images and clinical data

Research Proposal

Chaiyawat Suppasilp

6238023

Doctor of Philosophy Program in Data Science for Health Care

(International Program)

Contents

PREDICTING BI-RADS BY APPLYING DEEP LEARNING TO BREAST ULTRASOUND IMAGES AND CLINICAL DATA

List of Investigators and Affiliation (Including E-mail Address)

Chaiyawat Suppasilp (t_chaiyawat@hotmail.com)

Supervisors

Assoc.Prof. Thunyarat Anothaisintawee, M.D., Ph.D., thunyarat.ano@mahidol.ac.th

Dr. Anuchate Pattanateepapon Ph.D., anuchate.gab@mahidol.edu (Co-advisor)

Dr. Cholatip Wiratkapun, M.D. (Co-advisor)

Prof. Dr. Ammarin Thakkinstian, Ph.D., ammarin.tha@mahidol.edu(Co-advisor)

Department of Family medicine Faculty of Medicine Ramathibodi Hospital

Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine, Ramathibodi Hospital, Mahidol University

CHAPTER I
BACKGROUND & RATIONALE

1.1 Background and Rationale

Breast cancer is a major public health problem for women worldwide with age-standardized incidence rate of 46.3 per 100,000 person-years contributing as the first most common cancer following by prostate and lung cancer of 29.3 and 22.5 per 100,000 person-years, respectively[1]. And breast cancer is the second leading cause of cancer-related death in women[2].

The sufficient evidence supports using mammogram as a breast cancer screening tool for women aged from 50 to 69 years old that can reduce the risk of mortality from breast cancer of 24% and 40% among women who are invited and who are attended to mammography screening program, respectively. Even though there is limited evidence to establish the benefit of applying the mammography screening program for women aged between 40 to 49 years of age, it shows the risk reduction of breast cancer mortality about 20%[3]. Ultrasonography is usually used as adjunctive tool to mammography especially for dense breast and for negative mammogram[3]. Using ultrasonography adjunct to mammogram increases the breast cancer detection rate about 40% in patients who have dense breast. And the additional detected cases is 3.8 (95%CI, 3.4 to 4.2) per 1,000 mammography-negative women[4]. There is insufficient evidence to support using ultrasound alone for breast cancer screening[3], however, from the systematic reviews and meta-analysis of Rupali S. et. al. (2019) supports the potential use of ultrasound as an effective primary detection tool in case of infeasibility of mammography, but the meta-analyses showed high heterogeneity. The pooled sensitivity and specificity are 0.75 (95% CI, 0.64 to 0.83) and 0.87 (95% CI, 0.74 to 0.94), respectively, that are comparable to that of mammography of 0.56 (95% CI, 0.45 to 0.66) and 0.94 (95% CI, 0.86 to 0.98), respectively[5].

For mammogram, ultrasound, and also magnetic resonance imaging (MRI), Breast Imaging Reporting and Data System or BI-RADS is initially developed in 1993 to standardize breast imaging reports, improve communication, and provide improved quality assurance tool[6]. Nowadays, the 5th edition for mammography including the 2nd

edition for ultrasound and MRI are published simultaneously in 2014[7]. There are 7 classifications numbered from 0 to 6. Category 0 represents an incomplete study, additional imaging evaluation and/or comparison to prior mammograms is needed. Category 6 is known biopsy-proven malignancy. The rest of the categories (1, 2, 3, 4a, 4b, 4c, 5) represent the risk of malignancy, which the higher numbers represent higher likelihood of cancer[7].

Breast cancer risk prediction model from demographic data, reproductive history, and external hormone usage is another tool for screening and patient stratification. Unfortunately, the systematic reviews and meta-analyses showed limited performance of the risk prediction model[8, 9]. Javier L. et. al (2019) reported that maximum area under receiver operating characteristic curve value of individualized breast cancer risk prediction model is about 0.71 with moderate quality of the studies. So, it is a challenge to recommend using any individualized risk prediction models in clinical setting due to their limited quality and discrimination performance[9].

Obvious benefit of screening mammography with or without adjunctive ultrasonography, it has been recommended as a screening procedure for detecting breast cancer worldwide. For example, the national comprehensive cancer network (NCCN) and American College of Radiology (ACR) recommend annual screening in women over 40 years old without any risk, woman with risk factors or associated symptoms should be underwent early investigations. Thus, workload of radiologist is large especially for Thailand that has limited number of radiologists. Computer-aided decision (CAD) support systems to help radiologist's interpretation or triaging model to identify the mammograms that do not need radiologists to interpret the results would reduce workload and benefit to the radiologist worldwide.

Artificial intelligence (AI) as computer-aided detection systems for mammography has been extensively implemented in USA following initial promising results and reimbursement introduced since 2001[10]. In the last decade, a lot of AI were mainly applied for predicting the probability of malignancy of the breast tumor and were predominantly small retrospective studies. The median of area under ROC curve is 88.2% (range from 69.2% to 97.8%) with heterogeneous techniques of developed AI and highly selected image datasets that may not represent the proportion of cancers in the real clinical setting[11]. In the field of breast ultrasonography, Several semi-automated,

required hand-engineered features, and complete automated breast tumor classification models were proposed for predicting the probability of malignancy[12-15]. However, multi-category classification corresponding to BIRADS may better assist physicians in relieving the diagnose burden. At present, there are few studies developed BIRADS classification models using breast ultrasound images with accuracy around 0.734 to 0.998 for the five classes[16-18].

For Thailand, breast cancer was the second most common cancer with the age-standardized incidence rate of 21.8 per 100,000 person-years in 2010 and it was rising to be the most common cancer in 2012 with age-standardized incidence rate of 27.9 per 100,000 person-years. The predicted age-standardized incidence rate in 2025 is 30.5 per100,000 person-years, thus this problem has been being the important public health problem of Thailand[19]. As a developing country, screening with mammogram was not feasible and inequitable because more than 50% of total mammographic machines are only in the capital. Furthermore, there was only one radiologist in 63 provinces out of 77 provinces that was also insufficient[20]. Similar to majority of the individualized risk prediction models, a primary study in Thailand of Anothaisintawee et al. (2014) showed fair discriminative performance of external validation with C-statistic of 0.609 (95% CI, 0.511 to 0.706) even though it showed good model calibration[21]. Mammogram and ultrasound are still the main tools for breast cancer screening as the recommendation of National cancer Thailand (2017) for women after 40 years of age[22]. Because of the limitation of mammographic machines in Thailand, ultrasound is maybe an effective option used for breast cancer screening in Thailand.

The previous published works used deep learning as artificial intelligence to classify breast ultrasound images into BIRADS categories without concerning about clinical data used in risk prediction model. To the best of our knowledge, this is the first study that aims to develop artificial intelligence to predict BIRADS categories using breast ultrasound images combined with clinical data used in risk prediction models. Thus, it may facilitate physicians in clinical practice for breast cancer screening more efficient.

1.2 Research Question

1.2.1 Can we improve model performance of artificial intelligence by combined breast ultrasound images with clinical data that was used in risk prediction model?

1.2.2 Among different input data, image refinement techniques, transfer learning algorithms, model architectures, and clinical data from risk prediction models, which model provides the highest model performance?

1.3 Research Objectives

1.3.1 Primary Objective

To develop artificial intelligence to predict BIRADS categories from breast ultrasound images combined with clinical data that was used in risk prediction models

1.3.2 Secondary Objectives

To compare model performances among different input data, image refinement techniques, transfer learning algorithms, model architectures, and clinical data from risk prediction models

CHAPTER II
LITERATURE REVIEW

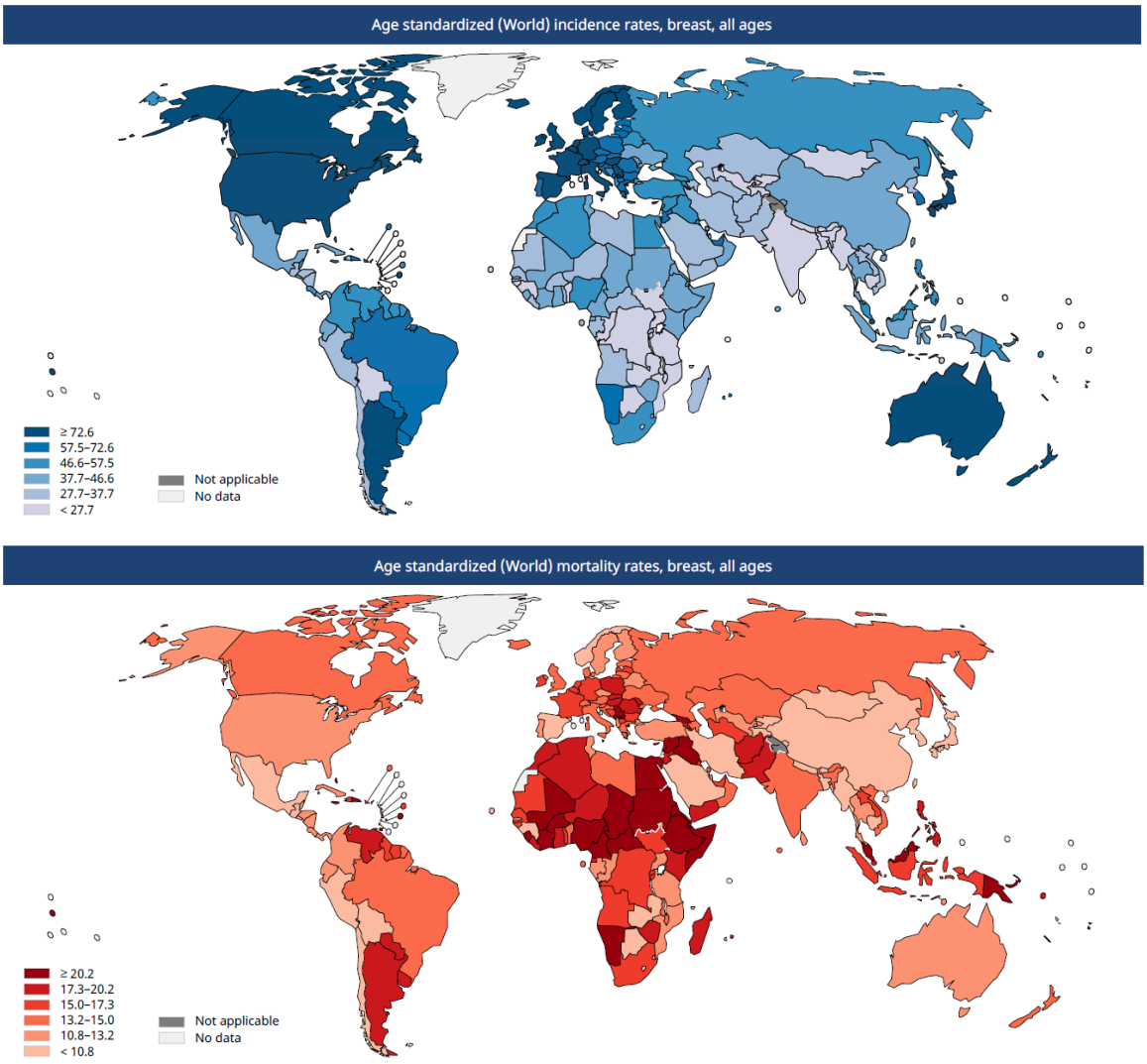
This chapter consists of 7 parts. The first part will give the reader the details about breast cancer and its epidemiologic profiles in both global and Thailand perspective. The second part, effective screening methods will be explained including reporting the results of mammograms and breast ultrasound using Breast Imaging-Reporting and Data System (BIRADS). The individual risk prediction model for breast cancer is also reviewed in this part. The third part will describe the terms of 'artificial intelligence' and its subgroups: machine learning and deep learning. Application of deep learning for mammograms and breast ultrasound images will be reviewed in fourth and fifth parts, respectively, while the sixth showed its application whether cooperating the clinical data in the image analytical models. The last part will detail about conceptual framework of this study.

2.1 Epidemiology of breast cancer

Breast cancer is a major public health problem for women worldwide with age-standardized incidence rate of 46.3 per 100,000 person-years contributing as the first most common cancer following by prostate and lung cancer of 29.3 and 22.5 per 100,000 person-years, respectively[1]. And breast cancer is the second leading cause of cancer-related death in women[2]. The heatmap of age-standardized incidence rates and age-standardized mortality rates of breast cancer in 2020 shown in Figure 2-1, replicated from International Agency for Research on Cancer[22] of world health organization (WHO).

For Thailand, breast cancer was the second most common cancer with the age-standardized incidence rate of 21.8 per 100,000 person-years in 2010 and it was rising to be the most common cancer in 2012 with age-standardized incidence rate of 27.9 per 100,000 person-years. The predicted age-standardized incidence rate in 2025 is 30.5 per100,000 person-years, thus this problem has been being the important public health problem of Thailand[19].

Figure 2-1 Age-standardized incidence rates and age-standardized mortality rates of breast cancer in 2020 (International Agency for Research on Cancer[22], WHO 2020)



The diagnosis of breast cancer is by biopsy-proven malignancy or pathology of the carcinoma after tumor removal. There are several histological subtypes that are infiltrating ductal carcinoma, infiltrating lobular carcinoma, tubular carcinoma, mucinous carcinoma, medullary carcinoma, invasive micropapillary carcinoma, metaplastic carcinoma, and adenoid cystic carcinoma. Each subtype has its own characteristics including the aggressiveness of the carcinoma. The most common subtype, in American women, is

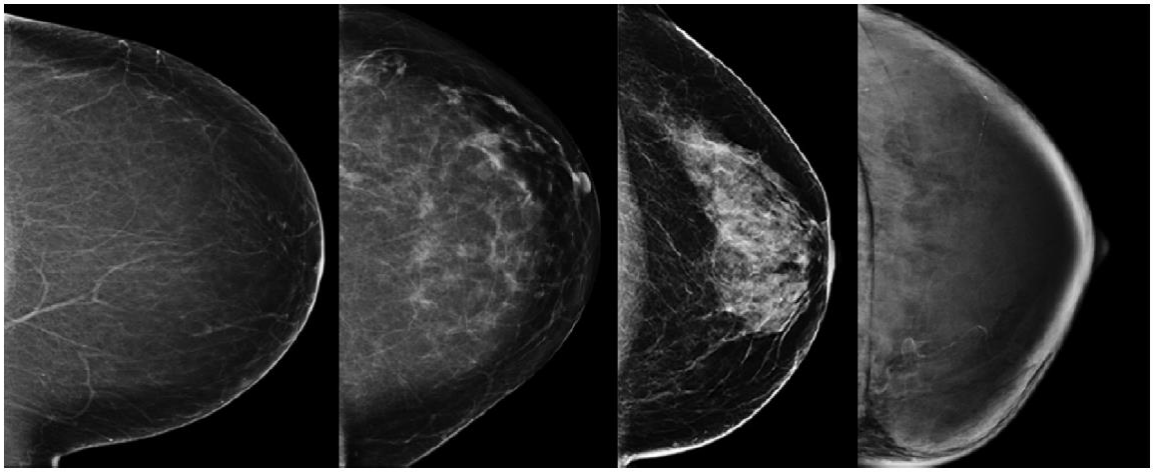
infiltrating ductal carcinoma, which is found about 76% of all breast cancer[23]. The patients with breast cancer will be classified by tumor node metastasis (TNM) system that can guide physicians about disease prognosis and using to decide the treatment. The latest 8th edition was published in 2018 and updated intervally[24]. Breast cancer treatment depends on TNM staging. For stage I, II, and IIIA, early stages without clinical evidence of lymph node involvement, the patients will undergo local treatment such as surgery without without radiotherapy. In contrast with early stages, patients with locally advanced stages should receive neoadjuvant systematic treatment prior to undergo surgery.

2.2 Screening tools for breast cancer

2.2.1 Mammogram

Mammography using X-ray imaging systems introduced by Salomon in Germany in 1913. The mammography screening is and effective breast cancer screening program, the performance of mammography showed sensitivity above 80% and specificity between 88% and 96%[25]. But two main factors reducing the performance are dense breast and size of the tumor[25]. Dense breast decreases contrast between the lesion and surrounding soft tissue, thus the lesion cannot be distinguished from random fluctuation, or noise, of the image. Sensitivity also depends on the size of the lesion, generally it is much easier to detect large cancers because they provide greater contrast, and whether microcalcifications, which increases the possibilities of being cancer, are present.

Figure 2-2 Example of mammograms with subjective assessment of breast density. Craniocaudal mammograms show findings characterized as almost entirely fatty (far left), scattered areas of fibroglandular density (second from left), heterogeneously dense (second from right), and extremely dense (far right)[26].



Double reading is practiced in some screening programs to increase screening performance. Double reading can be implemented in several possible ways:

1) two readers individually interpret the mammography examination, and the of them reports a suspicious finding

2) the readers interpret the examination independently and then create a consensus opinion, upon which assessment is based

3) after independent interpretation, a third radiologist arbitrates only if the two findings are different.

In a population screening programs using screen-film mammography, the detection rate of the tumor increased 10% to 15%[27, 28], but the specificity was decreased about 1.8%. The double reading programs was better than single reading for the detection of small invasive cancers, defined as less than 1.5 cm, but it did not increased the detection rate for larger cancer[29]. However, double reading is labor-intensive and many locations lack of radiologists.

The sufficient evidence supports using mammogram as a breast cancer screening tool for women aged from 50 to 69 years old that can reduce the risk of mortality

from breast cancer of 24% and 40% among women who are invited and who are attended to mammography screening program, respectively. Even though there is limited evidence to establish the benefit of applying the mammography screening program for women aged between 40 to 49 years of age, it shows the risk reduction of breast cancer mortality about 20%[3].

Obvious benefit of screening mammography with or without adjunctive ultrasonography, it has been recommended as a screening procedure for detecting breast cancer worldwide. For example, the national comprehensive cancer network (NCCN) and American College of Radiology (ACR) recommend annual screening in women over 40 years old without any risk, woman with risk factors or associated symptoms should be underwent early investigations

The breast cancer screening programs in Thailand provided by National cancer Thailand in 2007[30] have two purpose, which are breast mass screening and voluntary screening, without breast mass, with different recommended follow-up time. However, different center may have their own screening program. In practice, Thailand usually preforms mammography for breast cancer annually as same as the recommendation of American College of Radiology and American Cancer Society. However, as a developing country, screening with mammogram was not feasible and inequitable because more than 50% of total mammographic machines are only in the capital. Furthermore, there was only one radiologist in 63 provinces out of 77 provinces that was also insufficient[20].

Table 2-1 Thailand breast cancer screening program

Age	Mass Screening			Voluntary Screening		
(years)	BSE	CBE	Mammogram	BSE	CBE	Mammogram
20-39	1 month	-	-	1 month	3 years	-
40-69	1 month	1 year	-	1 month	1-2 years	1-2 years
70+	individual	individual	individual	-	-	-

BSE: breast self-examination, CBE: clinician-breast examination

2.2.2 Breast ultrasound

Breast ultrasound is not first line imaging for breast cancer screening due to its reproducibility, strongly depend on diagnostic skill of the physician or high inter-reader variability. However, ultrasound is low cost, high accessibility, and no radiation as compared to mammogram. Thus, ultrasonography is usually used as adjunctive tool to mammography especially for dense breast and for negative mammogram[3]. Using ultrasonography adjunct to mammogram increases the breast cancer detection rate about 40% in patients who have dense breast. And the additional detected cases is 3.8 (95%CI, 3.4 to 4.2) per 1,000 mammography-negative women[4]. There is insufficient evidence to support using ultrasound alone for breast cancer screening[3], however, from the systematic reviews and meta-analysis of Rupali S. et. al. (2019) supports the potential use of ultrasound as an effective primary detection tool in case of infeasibility of mammography, but the meta-analyses showed high heterogeneity. The pooled sensitivity and specificity are 0.75 (95% CI, 0.64 to 0.83) and 0.87 (95% CI, 0.74 to 0.94), respectively, that are comparable to that of mammography of 0.56 (95% CI, 0.45 to 0.66) and 0.94 (95% CI, 0.86 to 0.98), respectively[5].

Due to accessibility of ultrasound, this study made the decision of using breast ultrasound images rather than mammogram images as input of the models. The model may help physicians for better screening program regardless the available of mammography machine.

2.2.3 Breast Imaging-Reporting and Data System (BIRADS)

For mammogram, ultrasound, and also magnetic resonance imaging (MRI), Breast Imaging Reporting and Data System or BI-RADS published by the American College of Radiology (ACR), was initially developed in 1993 to standardize breast imaging reports, improve communication, and provide improved quality assurance tool[6]. Nowadays, the 5th edition for mammography including the 2nd edition for ultrasound and MRI are published simultaneously in 2014[7]. There are 7 classifications numbered from 0 to 6. Category 0 represents an incomplete study, additional imaging evaluation and/or comparison to prior mammograms is needed. Category 6 is known biopsy-proven

malignancy. The rest of the categories (1, 2, 3, 4a, 4b, 4c, 5) represent the risk of malignancy, which the higher numbers represent higher likelihood of cancer[7], shown in Table 2-2 with clinical recommendation. The Interobserver agreement with the new BI-RADS terminology is good[31] with cohen's kappa around 0.5-0.6, but the final assessment showed fair agreement with kappa of 0.28, which is comparable to the another study[32] of 0.37.

Table 2-2 BIRADS categories with clinical recommendation

Categories	Meaning	Likelihood of cancer	Clinical recommendation
0	Incomplete study	N/A	
1	Negative	0%	Routine screening
2	Benign	0%	Routine screening
3	Probably Benign	<2%	Short interval-follow-up
4	Suspicious for malignancy		Tissue diagnosis
	Category 4A: Low	2% - 10%	
	Category 4B: Moderate	10% - 50%	
	Category 4C: High	50% - 95%	
5	Highly Suggestive of Malignancy	>95%	Tissue diagnosis
6	Known Biopsy-Proven Malignancy	N/A	

The ultrasound lexicons were suggested to use in clinical practice divided into 5 main topics: tissue composition, masses, calcifications, associated features, special cases.

Breast Composition:

a. Homogeneous background echotexture – fat

b. Homogeneous background echotexture – fibroglandular

c. Heterogeneous background echotexture

Masses:

a. Shape

- Oval
- Round
- Irregular

b. Orientation

- Parallel
- Not parallel

c. Margin

- Circumscribed
- Not circumscribed
- Indistinct
- Angular
- Microlobulated
- Spiculated

d. Echo pattern

- Anechoic
- Hyperechoic
- Complex cystic and solid
- Hypoechoic
- Isoechoic
- Heterogeneous

e. Posterior features

- No posterior features
- Enhancement
- Shadowing
- Combined pattern

Calcifications:

a. Calcifications in a mass

b. Calcifications outside of a mass

c. Intraductal calcifications

Associated features:

a. Architectural distortion

b. Duct changes

c. Skin changes

- Skin thickening
- Skin retraction

d. Edema

e. Vascularity Absent

- Internal vascularity
- Vessels in rim

f. Elasticity assessment

- Soft
- Intermediate
- Hard

Special cases:

a. Simple cyst

b. Clustered microcysts

c. Complicated cyst

d. Mass in or on skin

e. Foreign body including implants

f. Lymph nodes – intramammary

g. Lymph nodes – axillary

h. Vascular abnormalities

- AVMs (arteriovenous malformations/ pseudoaneurysms)
- Mondor disease

i. Postsurgical fluid collection

j. fat necrosis

Radiologists use these lexicons to describe lesions and conclude to a final assessment category using BIRADS categories 0 to 6. If Mammography and US are

performed, the overall assessment should be based on the most abnormal of the two breasts, based on the highest likelihood of malignancy. At Ramathibodi hospital, radiologists always perform both mammogram and breast ultrasound and conclude the final BIRADS category with clinical recommendation.

BIRADS 0: Need additional imaging evaluation and/or prior Mammograms for Comparison

Category 0 or BI-RADS 0 is utilized when further imaging evaluation (e.g., additional views or ultrasound) or retrieval of prior examinations is required. When additional imaging studies are completed, a final assessment is made. For example, a patient presented with a inconclusive mass based on the mammogram at screening, which was assigned as BI-RADS 0 (needs additional imaging evaluation). After additional breast ultrasound was performed, it demonstrated that the mass was caused by an intramammary lymph node. Thus, the final assessment is BI-RADS 2, benign findings.

BIRADS 1: Negative

The breasts are symmetry and there is no mass or other suspicious findings.

BIRADS 2: Benign findings

BIRADS 2 category is different from BIRADS 1 because there is/are lesions or findings in either mammogram or ultrasonography, but the lesion(s) is considerably benign. The example of benign findings are listed as following:

- Involuting, calcified fibroadenomas
- Multiple large, rod-like calcifications
- Intramammary lymph nodes
- Vascular calcifications
- Implants
- Architectural distortion clearly related to prior surgery.
- Fat-containing lesions such as oil cysts, lipomas, galactoceles and mixed-density hamartomas

BIRADS 3: Probably benign findings

Findings in this BIRADS 3 category have less than a 2% risk of malignancy. Initial short-interval follow-up is suggested, for example, 6-month interval follow-up is suggested instead of routinely 1-year follow-up. Lesions appropriately placed in this category include:

- Non-calcified circumscribed mass on a baseline mammogram (unless it can be shown to be a cyst, an intramammary lymph node, or another benign finding)
- Focal asymmetry which becomes less dense on spot compression view
- Solitary group of punctate calcifications

BIRADS 4: Suspicious abnormalities

This category is reserved for findings that do not have the classic appearance of malignancy but are sufficiently suspicious to justify a recommendation for biopsy. Because BIRADS 4 has a wide range of probability of malignancy, from 2% to 95%, so false positive results as compared to tissue diagnosis as reference standard are usually occurred. If our proposed model can distinguish BIRADS 4 into its subdividing categories of 4A, 4B, 4C, it would be benefit to the patients and it can reduce the overall cost of the screening program.

Category 4A findings:

- Partially circumscribed mass, suggestive of (atypical) fibroadenoma
- Palpable, solitary, complex cystic and solid cyst
- Probable abscess

Category 4B findings:

- Group amorphous or fine pleomorphic calcifications
- Nondescript solid mass with indistinct margins

Category 4C findings:

- New group of fine linear calcifications
- New indistinct, irregular solitary mass

BIRADS 5: Highly suggestive of malignancy

BIRADS 5 category is reserved for findings that are classic breast cancers, with more than 95% likelihood of malignancy. Thus, the tissue diagnosis is required. The common findings of lesions in BIRADS 5 are listed below:

- Spiculated, irregular high-density mass.
- Segmental or linear arrangement of fine linear calcifications.
- Irregular spiculated mass with associated pleomorphic calcifications.

BIRADS 6: Known biopsy-proven malignancy

Usually, the malignancy should be gotten rid of after the diagnosis of carcinoma. But incomplete excision may be occurred, the follow-up mammogram and ultrasound may report as BIRADS 6 category. Beside incomplete excision or surgery, the lesions after monitoring response to neoadjuvant chemotherapy can be classified into this category.

Each finding has its own risk of malignancy. For example, cluster microcysts are usually benign caused by fibrocystic change or apocrine metaplasia, the positive predictive value of the lesion is low as 0.5% or 0.005 of being carcinoma, shown in Figure 2-3. Not only findings in the lesions represent the risk of being malignancy, but also findings anterior or superficial to the lesion and posterior to the lesion can represent the risk of malignancy. One of good examples is posterior acoustic features, which are no posterior features, enhancement, shadowing, and combined pattern. The risk of malignancy for no posterior features, enhancement, shadowing, and combined pattern, are 32%, 28%, 35%, 5%, respectively. So, posterior acoustic shadowing has the highest risk of malignancy. From these points of views, we decided to experiment on outside-lesion features in margin experiment rather than using only segmented lesions that may lose the important features of the breast ultrasound images.

Figure 2-3 Cluster microcysts, findings in BIRADS 3 category

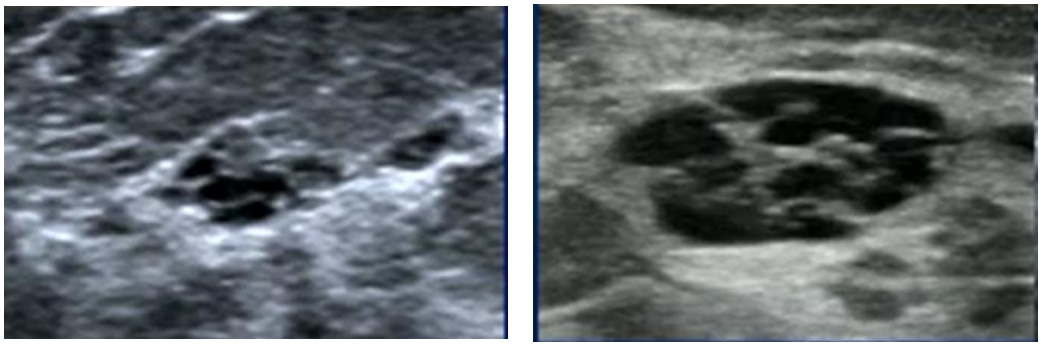
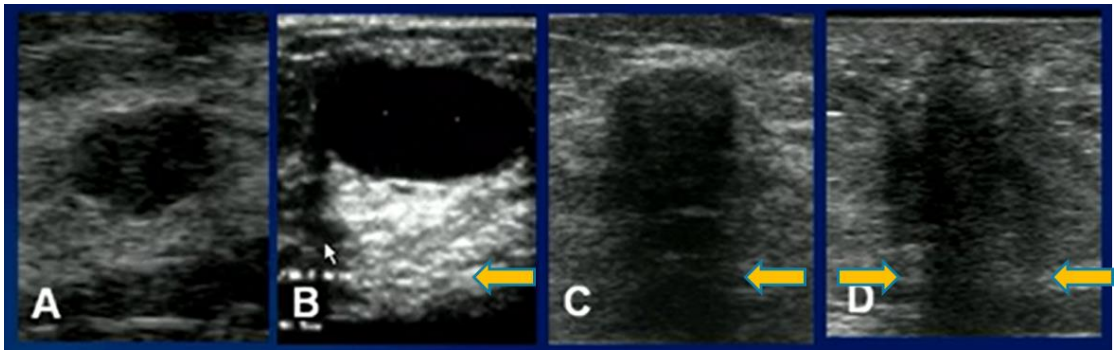


Figure 2-4 Posterior acoustic features from ultrasonography

(A: no posterior features, B: enhancement, C: shadowing, D: combined pattern)



2.2.4 Individual risk prediction model

Breast cancer risk prediction model from demographic data, reproductive history, and external hormone usage is another tool for screening and patient stratification. Unfortunately, the systematic reviews and meta-analyses showed limited performance of the risk prediction model[8, 9]. Javier L. et. al (2019) reported that maximum area under receiver operating characteristic curve value of individualized breast cancer risk prediction model is about 0.71 with moderate quality of the studies. So, it is a challenge to recommend using any individualized risk prediction models in clinical setting due to their limited quality and discrimination performance[9]. . Similar to majority of the individualized risk prediction models, a primary study in Thailand of Anothaisintawee et. al. in 2014 showed

fair discriminative performance of external validation with C-statistic of 0.609 (95% CI, 0.511 to 0.706) even though it showed good model calibration[21].

Among individual risk prediction models[8, 9], there are two landmark models from Gail et. al.[33] in 1989 to estimate the probability of develop breast cancer in American women given age and risk factors over the specified time interval and Rosner et. al.[34] in 1996 considered both reproductive and non-reproductive medical risk factors to develop the prediction model. Hence force, these risk factors were used with some modifications to create the risk prediction model.

IARC Handbooks of Cancer Prevention for breast cancer screening[3] broadly grouped associated risk factors into five groups, which are

- hormonal and reproductive factors
- lifestyle factors and environmental exposures
- risk factors that are not modifiable
- exposure to ionizing radiation
- genetic factors

The details of interesting risk factors and relative risk (RR) with 95% confident interval (95%CI) shown in Table 2-3, but the exposure to ionizing radiation and genetic factors did not be included in the table due to high heterogeneities. Among the established risk factors for breast cancer, genetic factors are of particular importance. For example, BRCA1 and BRCA2 mutation carriers, the cumulative risk to age 80 years was shown to reach 90% and 41%, respectively, for breast cancer[35].

In this study, we will consider clinical data from risk prediction model of Anothaisintawee et. al.[21] that divided into 3 categories including demographic data, reproductive history, and external hormone usage. There is no feature about exposure to ionizing radiation and genetic factors in this dataset. This information will be described in detail in chapter 3 research methodology.

Table 2-3 Established risk factors for breast cancer with magnitude of relative risk

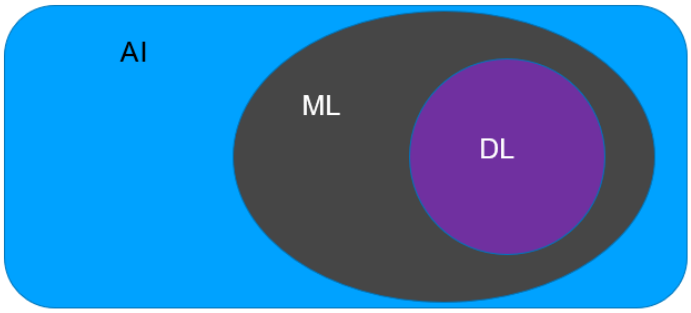
Risk factor	Categories	Relative Risk (95%CI)
Hormonal and reproductive factors		
Age at menarche (years)	11	1.0 (reference)
	15	0.69 (0.65-0.74)
Parity	Nulliparous	1.0 (reference)
	Parous	1.26 (1.10–1.44)
Age at first full-term pregnancy (years)	20	0.73 (0.63–0.86)
	30	1.16 (0.96–1.41)
Breastfeeding	Per 12 months of total breastfeeding	0.96 (0.94–0.97)
Age at menopause (years)	45	1.0 (reference)
	55	1.44 (1.26–1.64)
Type of menopause	Natural	1.0 (reference)
	Bilateral oophorectomy	0.89 (0.80–0.98)
Postmenopausal hormone use	None	1.0 (reference)
	Estrogen only	1.18 (1.08–1.30)
	Combined estrogen–progestogen for > 5 years	1.63 (1.22–2.18)
Lifestyle factors		
Alcohol consumption	Per 12 g/day	1.12 (1.09–1.14)
	Premenopausal	1.09 (1.01–1.17)
	Postmenopausal	1.08 (1.05–1.10)
Tobacco smoking (pack–years)	≥ 20	1.28 (1.17–1.39)
Weight increase (per 5 kg/m2 increase in BMI)	Postmenopausal	1.12 (1.08–1.16)
	Premenopausal	0.92 (0.88–0.97)

Table 2-3 (continue)		
Risk factor	**Categories**	**Relative Risk (95%CI)**
Lifestyle factors (continue)		
Physical activity, high vs low (METs)	Premenopausal	0.87 (0.84–0.92)
	Postmenopausal	0.77 (0.72–0.84)
	Moderate physical activity (3–5.9 METs)	0.81 (0.72–0.92)
Non-modifiable factors		
Height (per 5 cm increase)	Premenopausal	1.09 (1.05–1.14)
	Postmenopausal	1.11 (1.09–1.13)
	Any age	1.03 (1.01–1.04)
Age (years)	< 50	1.0 (reference)
	50-59	6.6 (6.5–6.7)
	60-69	9.2 (9.1–9.3)
	70-79	11.1 (10.9–11.2)
	≥ 80	10.1 (10.0–10.3)
Benign breast disease	No	1.0 (reference)
	Non-epithelial proliferative hyperplasia	1.57 (1.43–1.73)
	Common epithelial hyperplasia	1.5–2.0
	Atypical epithelial hyperplasia	2.5–4.0
Breast density	Dense area, mean: 59.92–201.49 sq.cm.	1.57 (1.18–1.67)
Exposure to ionizing radiation		
Genetic factors		

2.3 Artificial intelligence

Nowadays, artificial intelligence (AI) has been developed rapidly including application of artificial intelligence in healthcare system. The definition of AI is still debated, but we can roughly explain AI for medical aspects as a system's ability, that may learn from data, can achieve specific goals and tasks[36]. Machine learning (ML) is computational program and also a specific type of AI that can learn from data and can achieve specific goals and tasks through flexible adaptation upon the data that the machine learning had learnt. Further, deep learning (DL) is a machine learning that has complex programming; thus, it is a specific type of AI. Nowadays, deep learnings usually use artificial neural network as its algorithms, called deep neural networks, and there are several types of deep neural networks with thousands of model architectures.

Figure 2-5 Conceptual diagram of artificial intelligence (AI), machine learning (ML), and deep learning (DL)



For machine learning, including deep learning, there are 4 main learning algorithm that are supervised learning, semi-supervised learning, unsupervised learning, reinforcement learning.

Supervised learning is a learning algorithm that can learn from the data under supervision. The word "under supervision" in this specific meaning is that we give the answers, we want them to predict, to them. The training set with the answers is called labeled data. After the machine have learnt, the test set without the answer will be used as an input for the machine learning, then the machine learning will predict the answer by

using the data from the test set. The performance of the machine learning can be evaluated by comparing the predicted answer with the real answer or ground truth.

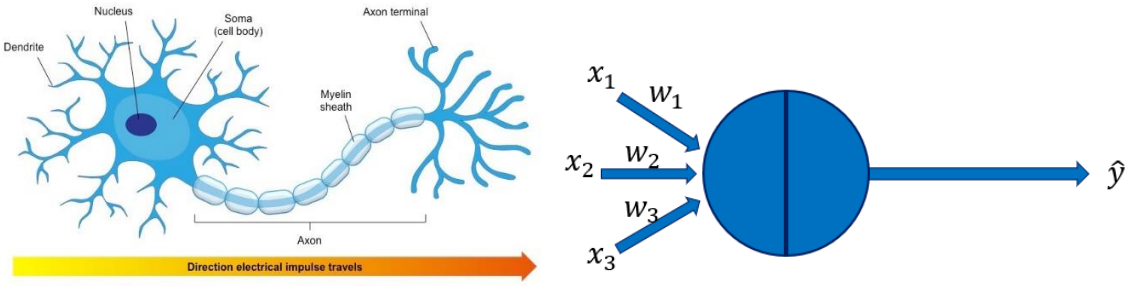
In contradiction with supervised learning, unsupervised learning will make a prediction from the whole data set without the needs of known answers. Unsupervised learning uses clustering, grouping, probability, etc. to classify the data into subgroups with the same characteristics, called features. Please note that, even though unsupervised learning does not use labeled data, but the performance may be better than supervised learning in specific nature of the dataset, furthermore, some problem we cannot directly label the data. The examples of unsupervised learnings are principal component analysis, k-nearest neighbor algorithm, etc.

Semi-supervised learning and reinforcement learning will not be used in our study. Briefly, semi-supervised learning makes predictions by using partially labeled training dataset, and reinforcement learning learns through interactions between machine learning and response from human to adjust the policies in order to make the best policies achieving the specific tasks.

Artificial neuron is one of the most popular supervised learning algorithms that imitate the way biological neuron functions. While the biological neuron receives signal via dendrite to cell body and delivers the signal through out the axon, the artificial neuron uses the input data as input signal, processing in the artificial neuron, and sent the results as the output. The output of the artificial neuron will be inputs of other artificial neuron in the neural network model, thus one neuron has only one output, but it can be inputs of many other neurons as same as the way biological neuron does: one axon with many synapses. Figure 2-6 shows the comparison between biological and artificial neurons.

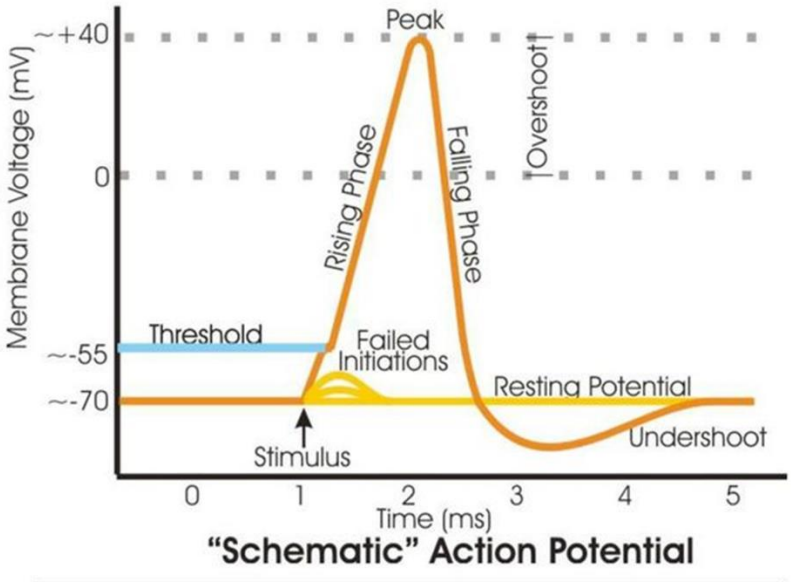
Another concept adopted from the activity of biological neuron is activation function. Activation function uses all-or-none algorithm. If the inducted cell membrane is not higher than the action potential, the neuron will not be activated and go resting as the same as before. But, if the inducted neural membrane is higher than the action potential, the neuron will be activated and sent the signal further to other neurons, as shown in Figure 2-7, the pale orange line failed to initiate the activation, while the dark orange line doses.

Figure 2-6 Comparison between biological and artificial neurons (Biological neuron, https://ib.bioninja.com.au/standard-level/topic-6-human-physiology/65-neurons-and-synapses/neurons.html)



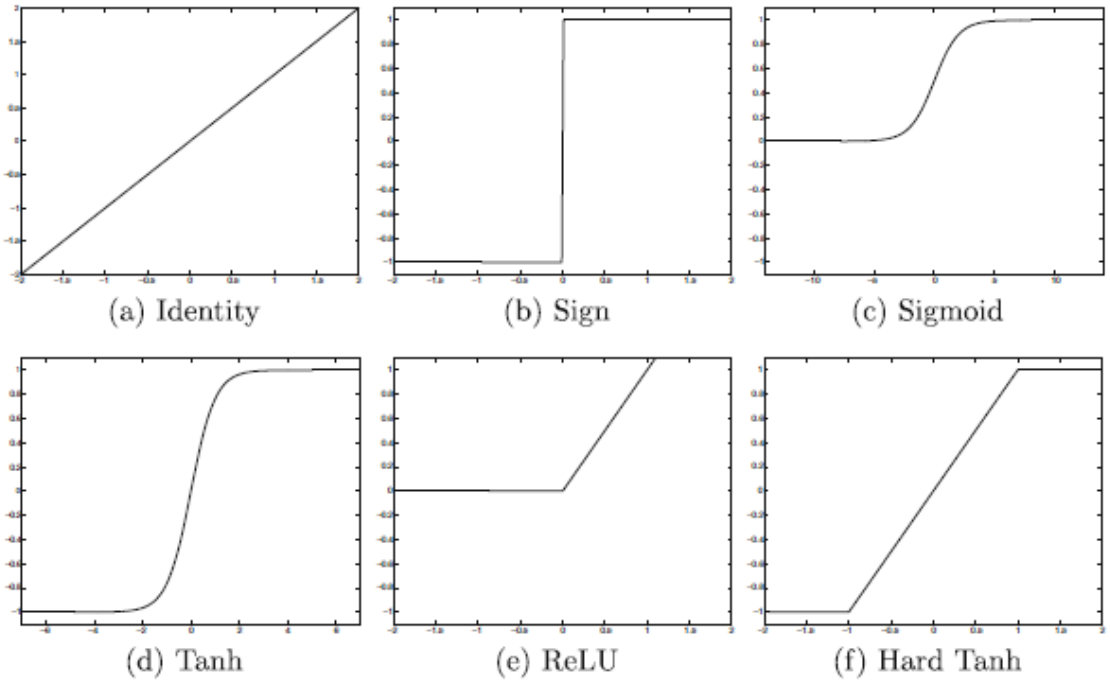
Biological neuron	Artificial neuron
Soma (cell body)	Neuron
Dendrite	Input
Axon	Output
Synapse	Weight

Figure 2-6 Action potential of biological neuron (Mohammed El Majdoubi)



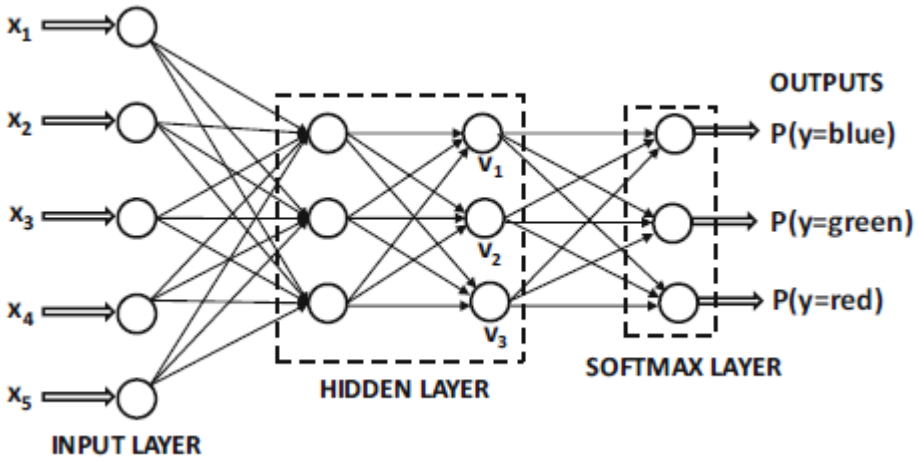
Similar but not the same, activation functions in artificial neurons may perform with all-or-none fashion as shown in Figure 2-7, Sign activation function. But, for example, sigmoid and hyperbolic tangent (Tanh) activation functions are remarkably similar but different from biological activation function that a small number around the value zero will give a value. Identity activation function do not use all-or-none concept. In our study, rectified linear activation function (ReLU) and softmax activation function will be used for internal activation of neural networks and for the last layer as multiclass classification activation function, respectively. ReLU activation function will convert negative value into zero and keep positive values as it is. The softmax activation function, in the last layer, will use probabilistic cutoff to make a prediction of the input as the chance of being grouped in each multiclass.

Figure 2-7 An example of activation functions in artificial neurons[37]



Once, we combine multiple artificial neurons together, it will be called artificial neural network. There are 3 main layers of an artificial neural network, which are input layer, hidden layers, and output layer (using softmax activation, thus can be called softmax layer), shown in Figure 2-8. Please note that there are one input layer and one output layer, but an artificial neural network may contain multiple hidden layers for complex architecture. If there are several hidden layers, no exact cutoff number, we can call it a deep neural network according to complex architectures as same as other deep learning algorithm.

Figure 2-8 An example of artificial neural network with softmax activation function for categorical multiclass classification[37]



There are several types of deep neural networks, for example, convolutional neural networks (CNNs) for image analysis, recurrent neural networks (RNNs) for time-series and natural language processing (NLP), deep reinforcement learning for complex tasks that have interactions with environments such as self-driving car, generative adversarial networks (GANs) for generation of data such as fake faces to develop dataset. For our study, we will only use convolutional neural network to classify medical ultrasound images into BIRADS categories.

Convolutional neural networks (CNNs) typically contain 3 parts: convolution layers, pooling layers, and one fully connected layer, shown in Figure 2-9. Convolution layers rely on convolutional operation or application of filter or kernel to the image matrix as an input, for example, smoothening or sharpening of the images. After an appropriate convolutional operation, the features or characteristics of interest may be enhanced and can be used for better classification or prediction. For better demonstration, Figure 2-10 the rightest kernel function is the best among these three to convolute the image for kidney segmentation purpose. The pooling layer groups feature map, ultrasound image matrix, into lower resolution feature map to increase the effective scope or receptive field. The lower resolution of the image but contextually rich information. The commonly used pooling algorithm in biomedical image is maximum pooling or max pooling, which choose the highest value in matrix of interest because the original image may contain a lot of zero-intensity pixel, black pixel, thus it is not appropriate to use others such as mean pooling or minimum pooling. The example of max pooling is shown in Figure 2-11, the down sampling or pooling processes decreased 512x512 pixels of CT image to 32x32 pixels, thus the resolution is lower but better represent the kidney fields in the image. CNNs will use convolutional layers and max pooling layers to create the last layer called flattened feature maps, or fully connected layer, which contain several or many artificial neurons working together, and the final classification will be made according to this last layer. In biomedical field, sigmoid activation function is commonly used for binary classification as same as an example demonstrated in Figure 2-9 and softmax activation function is used for multiclass classification that will be used for our study, which plans to develop BIRADS prediction model.

Figure 2-9 Convolutional neural networks (CNNs)[38]

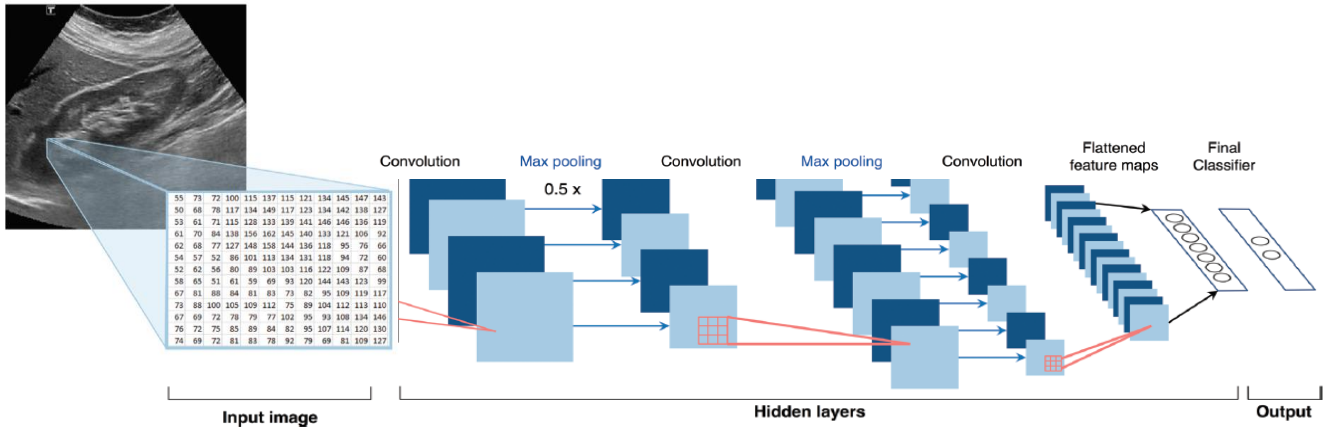


Figure 2-10 Example of different kernel functions for convolutional layer[38]

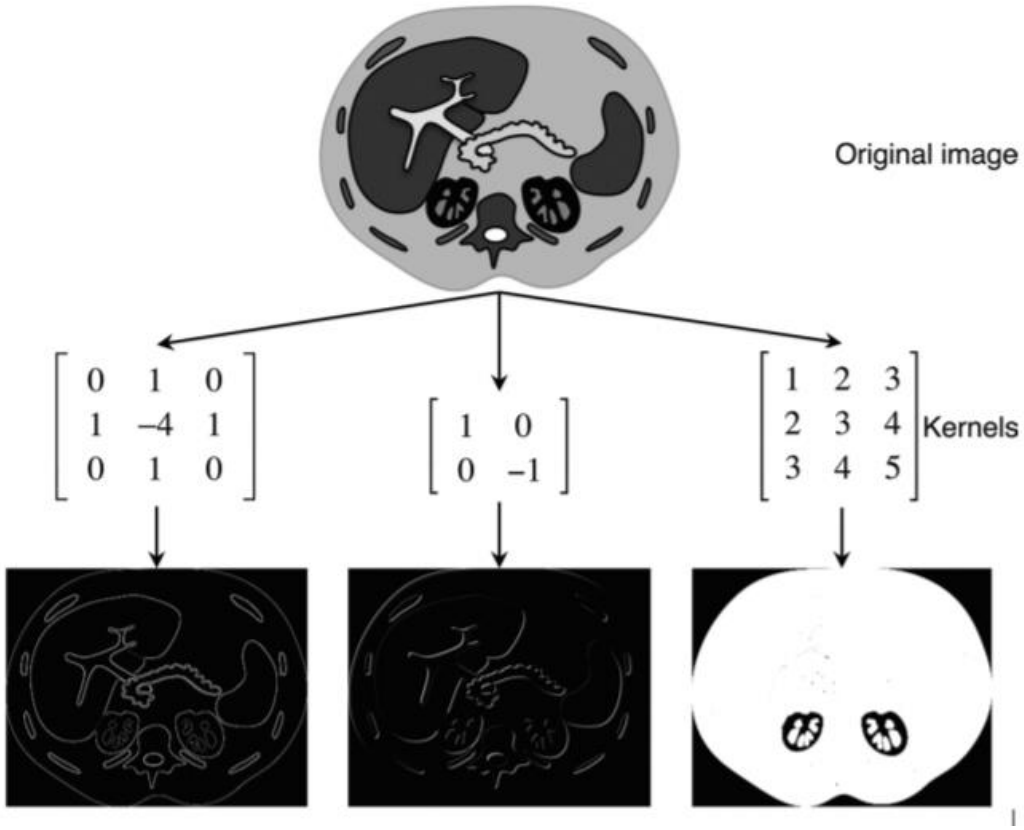
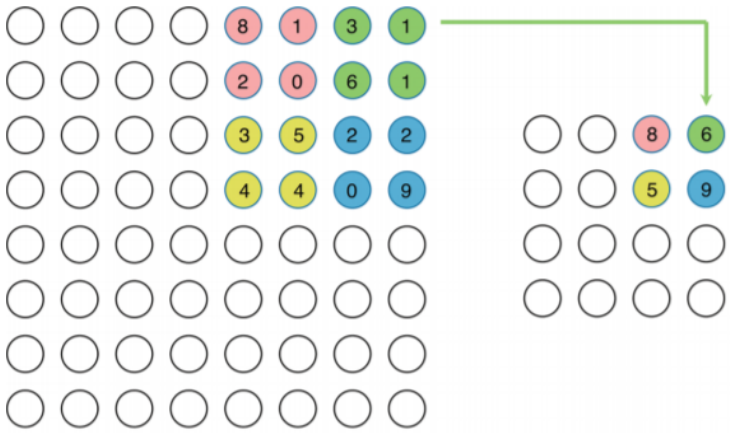
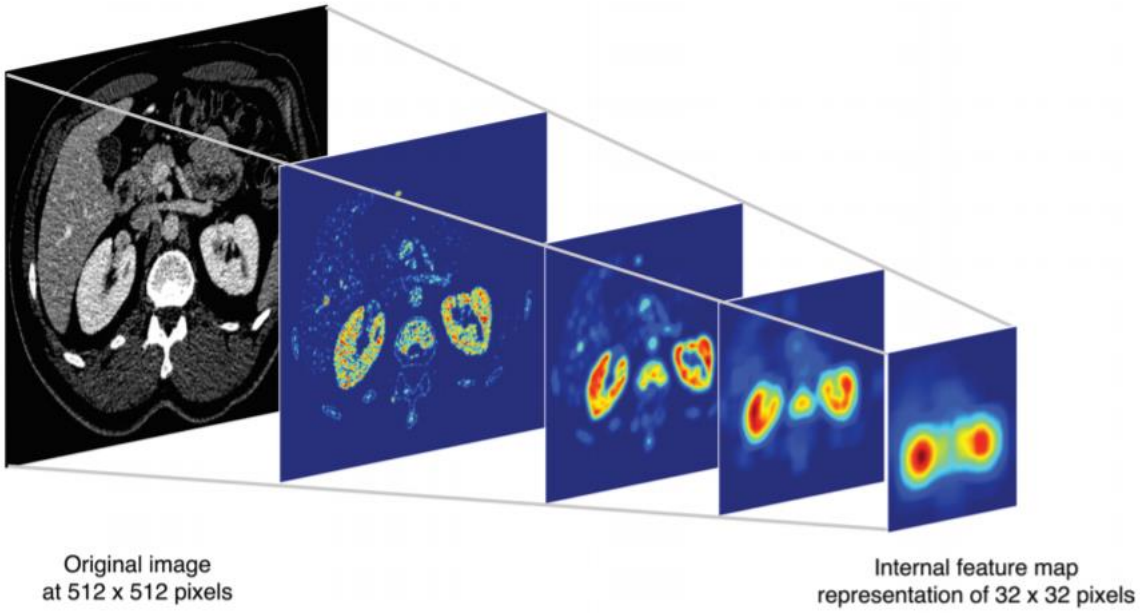


Figure 2-11 Example of max pooling algorithm for pooling layer by using 2x2 matrix (a) and the CT image of 256x256 pixels down sampled to 32x32 pixels[38]

(a)



(b)



2.4 Deep learning for breast cancer prediction

Deep learning is successfully applied to mammogram image as a computer-aid decision support system. Computer-aid decision support system (CAD) of mammogram interpretation was approved by The United States Food and Drug Administration (FDA). Gromet et al. conducted prospective clinical trials to compare single reading, double reading, and single reading but with CAD[39]. The results showed that single reading with CAD provides the highest sensitivity of 90.4% as compared to single reading of 81.4% and double reading of 88.0%. Furthermore, there are available commercial artificial intelligence algorithms for independent assessment of screening mammograms[40]. The area under the receiving operating curve for cancer detection was 0.956 with 95% CI of 0.948 to 0.965. Because of high diagnostic performance of commercial AI on mammograms, there is extremely limited possibility to develop and further study in this fields. Nevertheless, there is limit number of mammogram machine in Thailand[20]. Thus, this study will focus on using breast ultrasound images for prediction of BIRADS categories.

From the review, the performance of deep learning algorithm on breast ultrasound images the performance, either from validation set or independent test set, showed very promising performance with area under the receiving operating curve more than 0.872 to 0.93 for different models, as shown in Table 2-4. The accuracies of the models range from 82% to 95.8%. Moreover, Byra et al. and Fujioka et al. showed that CNNs surpass the performance of the radiologists with AUC of 0.936 and 0.913 as compared to the radiologists of 0.806 to 0.882 and 0.728 to 0.845, respectively. Huang et al. used imagenet-pretrained modified VGG-16 deep convolutional neural networks to classify 2,238 ultrasound images into BIRAD categories of 3, 4A, 4B, 4C, and 5. The results showed that the accuracies range from 0734 to 0.998 for the five classes, which is very promising.

Table 2-4 Studies using deep learning approach on breast ultrasound images

Publication	Year	Training set	Validation set	Test set	Performance
Anropova et al.	2017	2393 ROIs	5-fold CV	-	AUC = 0.872-0.902
Han et al.	2017	6579 masses	10-fold CV	829 masses	AUC = 0.958
Xiao et al.	2018	1,647 images	411 images	411 images	AUC = 0.91-0.93
Zhou et al.	2018	400 images	45 images	45 images	Accuracy of 95.8%
Lee et al.	2018	Study1: 143 Study2: 210	Study1: 27 Study2: 40	-	Accuracy: Study1: 82% Study2: 83%
Byra et al.	2019	582 masses	150 masses	150 masses	AUC = 0.936
Fujioka et al.	2019	240 masess	120 masses	-	AUC = 0.913

CHAPTER III
METHODOLOGY

3.1 Study design and setting

This study is retrospective cross-sectional study, which all subjects are in the subset of derived phase in the research of "Development and Validation of a Breast Cancer Risk Prediction Model for Thai Women: A Cross-Sectional Study" published by Thunyarat et. al. in 2014[21], which you may look in details. Briefly, the data were consecutively collected from September 2011 to Septimber 2012 at Ramathibodi Hospital, which is a tertiary care hospital and also a school of medicine hospital in center of Bangkok, Thailand. There are about 5,000 out-patients per day with 1,000 in-patient beds. The breast screening cases are about 16,000 cases per year that every patient was performed both mammogram and breast ultrasound. The patient's data including pathological reports were recorded in electronic medical record (EMR) system and all radiographic images and reports are in Picture Archiving and Communication System (PACS). All participants provided written informed consents.

This study will use patient's characteristics, risk factors in individual risk prediction model for breast cancer, and breast ultrasonographic images as input data for deep convolutional neural networks (DNNs) to predict Breast Imaging Reporting and Data System (BIRADS) score via supervised machine learning algorithm. The details about supervised machine learning and deep learning are in Chapter 2 and also in proposed method in this Chapter3.

3.2 Study subjects

As stated, all subjects are in the subset of derived phase in the research of "Development and Validation of a Breast Cancer Risk Prediction Model for Thai Women: A Cross-Sectional Study" published by Thunyarat et. al. in 2014[21] that included 15,718 eligible subjects after excluded 1,842 women according to exclusion criteria; 1234 previous

breast cancer or other cancers, 529 not willing to participate, 76 breast augmentations, 2 primary amenorrhea, 1 age less than 18 years old. The sample size was calculated using breast cancer prevalence at Ramathibodi Hospital in 2010 of 0.6% with 95% confidence interval from 0.5% to 0.7% and one proportion testing with 5% type-1 error and confidence interval width of 0.0015[21]. However, the sample size may not perfectly dedicate for this study because of the different target outcome and methodology. Further, the following inclusion and exclusion criteria will be considered as following to serve the aims of the study. The result of the study included 107 cases of breast cancer and 15,611 cases of non-breast cancer.

3.2.1 Inclusion criteria

- All female subjects with age of 18 years old and over in derived phase of the mentioned study[21]

3.2.2 Exclusion criteria

- Subjects with BIRADS score of 0 (incomplete study) or 1 (no breast ultrasound lesion) or 6 (evidence proof of malignancy)

- No final BIRADS score in mammogram and ultrasound reports corresponding to the time of collecting the data or more than 1 month from the patient's data collected.

- No ultrasound images corresponding to the final BIRADS score

- Poor quality of ultrasound image or it do not represent the BIRADS score, for example, the lesion in an image is not at the location that the report stated.

3.3 Features

3.3.1 Features

The input data in this study consists of two parts: ultrasound images and patient's clinical data. Because deep convolutional neural network technique does not require feature engineering, thus ultrasound image can be used as an input directly. And our

proposed two-stage model will deal with the ultrasound image that the method will be described later in data analysis part.

Clinical data are mainly categorized into three parts: Demographic data, Reproductive history, and External hormone usage. All the data was from interviewing the participants by well-trained staffs and recorded in structed data record forms. Regardless, from the systematic review and meta-analysis[8], the individual risk prediction model for breast cancer shows low discrimination performance with concordance statistics of 0.53 to 0.66 for internal validation and 0.56 to 0.63 for external validation, however, these data have no proof of its utility to improve the DNNs model to predict BIRADS. Thus, one of the aims of this study is to assess the incremental model performance by combining both ultrasound image and these clinical data compared to using of ultrasound image alone.

3.3.1.1 Demographic data

Demographic data is one of the most used data in individual risk prediction model. The landmark study from Gail et. al. (1989)[33] included Age, age at menarche, age at first live birth, family history of breast cancer, numbers of previous breast biopsy, history of atypical hyperplasia as risk factors in the multiple logistic regression model to predict invasive breast cancer with CIS. Henceforth, these risk factors were usually included in the individual risk prediction models.

In this study, the demographic data includes age, body mass index (BMI), risk behavior such as smoking and alcohol consumption, family history of breast and ovarian cancers in the first-degree relatives, history of breast biopsy, and underlying diseases, which are chronic kidney disease (CKD), dyslipidemia (DLP), and diabetes mellitus (DM). These data were obtained from interviews and subsequently verified with International Classification of Diseases 10 (ICD-10) databases.

Demographic variables

- Date of birth
- Living place

- Educational level that was categorized into 3 categories
    - No education or less than primary school
    - High school
    - College degree or higher
- Weight
- Height
- Body mass index (BMI) calculated by dividing weight in kilogram (kg) with height in meter square.

Risk behaviors

- History of smoking was defined as follows:

    Active smoker: currently smoke or have ever smoked at least one cigarette per day for more than 6 months prior to study date. Active smoker was categorized into current smoker and past smoking.

    Passive smoker: women who have ever lived in the same houses or same workplaces with smoking persons.

The duration and number of cigarettes per day were also recorded and it was converted into pack per year.

- History of alcohol drinking was defined as participants who have ever drunk more than three times a week for 6 months prior to study date.

- Types of alcoholic beverages (for example; rice whisky, beer, wine, whisky) were recorded and alcohol consumption in grams/day of pure alcohol was calculated according to standardized questionnaire from Health Technology and Intervention Assessment Program briefly as follow. Ethanol content (grams/day) equals to numbers of alcohol (drinks/day) × 0.79 × % alcohol in each alcoholic beverage.

- Average daily alcohol intake was categorized into four groups according to: no alcohol, low intake (from 0.1 to 4.9 g/day), medium intake (from 5.0 to 9.9 g/day), and high intake (more than 10 g/day)

Family history of cancer

- History of breast or ovarian cancers in the first and second degree relatives

- Numbers and relationship with study participants (i.e., mother, daughter, or sister) of the first- and second-degree relatives who have been diagnosed with invasive breast cancer

Underlying diseases (in this cross-sectional study)

- Hypertension (HT) was defined as systolic blood pressure equals or higher than 140 mmHg or diastolic blood pressure equals or higher than 90 mmHg.

- Dyslipidemia (DLP) was defined as serum total cholesterol higher than 200 mg/dl

- Chronic kidney disease (CKD) was defined as estimated glomerular filtration rate (GFR) lower than 60 mL/min/1.73 m2

- Diabetes mellitus (DM) was defined according to following standard criteria:

1) Fasting plasma glucose ≥ 126 mg/dL at least two consecutive times or

2) Random plasma glucose ≥ 200 mg/dL or

3) OGTT ≥ 200 mg/dL or

4) HbA1C ≥ 6.5% or

5) Taking oral diabetic drugs or injection of insulin

Diagnosis of DM was confirmed with the ICD-10 code E10-14. Moreover, Types of diabetic drugs: Oral diabetic drug and/or Insulin injection and age at diagnosis were also recorded.

3.3.1.2 Reproductive history

The reproductive history includes Age at menarche, Age at first live birth, Age at menopause, and history of breast feeding. While the first three reproductive history were included in the model[8], the recent model[21] in Thai population included new variable of history of breast feeding that is one of the important factors. Thus, it will be included in our study as well.

Age at menarche defined as age at first menstrual period

Marital status

- Single
- Married
- Widowed
- Separated

Menopausal status

- Premenopausal
- Menopausal which was defined as women whose menstrual periods have been ceased for more than 6 consecutive months before the study date or women whose both ovaries were surgically removed or damaged by radiation or drugs
- Unknown menopausal status includes women who have undergone hysterectomy without bilateral oophorectomy and women who did not report their menopausal status. Women in this category were classified as menopausal women if their age were more than 49 years old which was the average age at menopause in Thai women.

Age at menopause was referred to age at menstrual period ceased in natural menopause or age at having bilateral oophorectomy in surgical menopause. For women who have undergone hysterectomy without bilateral oophorectomy, age at menopause was defined as average age at menopause in Thai women.

Age at birth of first child was defined as age at delivering the first live birth.

Parity was defined as the number of giving birth or a fetus with a gestational age of 24 weeks or more, regardless of whether the child was born alive or was stillborn.

History of breastfeeding

- Have ever breastfed was defined as having ever breastfed for more than 1 month.
- Have never breastfed was defined as having never breastfed or never pregnant.

For women who have ever breastfed, data about breastfeeding was gathered as follows:

- Duration of breast feeding (in month) for each child
- Cumulative duration of breast feeding
- Average duration of breastfeeding per one child

3.3.1.3 External hormone usage

The external hormone usage such as hormonal replacement therapy (HRT), oral contraceptives (OC), and medroxyprogesterone injection, will be included in the models. External hormone users were defined as follows: current users if they were currently used or used within the last 12 months, past users if they had stopped using longer than 12 months, and never users if they had never or used for less than 1 month.

History of Hormonal replacement therapy (HRT)

- "Have ever used of HRT" was defined as women who have ever used HRT continuously more than 1 months. Thus, "Have never used of HRT" was referred to women who have never used or HRT used less than 1 months before the study date.

- In women who have ever used HRT, associated data would be recorded as follow:

  - Age of first use of HRT
  - Age of last use of HRT
  - Duration of HRT usage

History of oral contraceptives use

- Have ever used OCs was defined as women who have ever used OCs continuously for more than 1 months

- Have never used of OCs was referred to women who have never used OCs or used less than 1 months before the study date.

- In women who have ever used OCs, data about OCs would be gathered as follows:

  - Age of first use of OCs
  - Age of last use of OCs
  - Duration of OCs usage

3.3.2 Target outcome

The target outcome of the study is to predict BIRADS categories. BIRADS classifies mammographic or ultrasound findings into 7 categories[7] as shown in the Table 3-1. However, in this study, we excluded categories 0 and 6 because category 0 is incomplete study, and category 6 is definitely positive for malignancy proven by biopsy. Therefore, we will have 5 possible target outcomes that are category 1, 2, 3, 4, and 5.

Later, we will conduct the experiment to find the performance of the model if we separate category 4 into 3 sub-categories due to expected clinical utility. Because the range of the likelihood of cancer of category 4 is large from 2% to 95%, thus it would be better to categorized into 4A, 4B, 4C, as the likelihood of cancer shown in the Table 3-1. So, there are 7 possible target outcomes: 1, 2, 3, 4A, 4B, 4C, and 5. Further, the likelihood of cancer of categories 1 and 2 is 0%. Category 1 is definitely negative that means there is no mammographic or ultrasonographic lesion to be evaluated, and the findings in ultrasonography of category 2 may be subtle or small, so the performance of identifying the lesions may be poor. Thus, we may combine these two categories together and defined as benign group in order to improve the overall model performance and it correlates with clinical context that categories 1 and 2 lesions are required routine screening, while category 3 lesions are required screening more frequently and category 4 and above are required biopsy[7].

Table 3-1 BI-RADS assessment categories with likelihood of cancer

Categories	Meaning	Likelihood of cancer
0	Incomplete study	N/A
1	Negative	0%
2	Benign	0%
3	Probably Benign	More than 0% to 2%
4	Suspicious	
	Category 4A: Low suspicion for malignancy	2% not more than10%
	Category 4B: Moderate suspicion for malignancy	10% not more than 50%
	Category 4C: High suspicion for malignancy	50% but less than 95%
5	Highly Suggestive of Malignancy	95% and above
6	Known Biopsy-Proven Malignancy	N/A

3.4 Source of data and data retrieval

There are 2 set of data used in this study; ultrasound images and clinical data to assess the benefit of using ultrasound images combined with clinical data rather than using ultrasound images alone, in order to predict BIRAD classification. As mentioned above, all subjects are in the subset of derived phase in the research of "Development and Validation of a Breast Cancer Risk Prediction Model for Thai Women: A Cross-Sectional Study" published by Thunyarat et. al. in 2014[21]. The study collected the data from interviewing patients who came for breast cancer screening program at Ramathibodi hospital, all of them underwent both mammography and breast ultrasonography at radiology department in the same visit of interviewing. However, there is no ultrasound image stored in the research database, but in the server of radiology department.

3.4.1 Ultrasound images

Routinely, breast cancer screening program in Ramathibodi hospital contains both mammogram and breast ultrasound. All images are stored in Digital Imaging and

Communications in Medicine (DICOM) format in picture archiving and communication system (PACS) in the server of radiology department. Each subject may have multiple visits of breast cancer screening, only ultrasound images in the study corresponding to the date of interviewing will be retrieved. In that ultrasound study, there are several images in a study, all images will be retrieved and stored under the same ID code corresponding to clinical data. Theoretically, we can classify into BIRADS category by considering only mammogram or only breast ultrasound, but in clinical practice of Ramathibodi hospital, radiologists use both mammogram and ultrasound findings together for reporting only one BIRADS category. We will define this category as "final BIRADS" that is the reference standard of this study. The final BIRADS will retrieved from the reports in PAC system. In summary, we will get all ultrasound images and the final BIRADS reference standard from PAC system.

3.4.2 Clinical data

All clinical data were collected and stored in structured tabulated data. Clinical data are mainly categorized into three parts: Demographic data, Reproductive history, and External hormone usage. The details have been already described in the prior session.

3.5 Data preparation

The data preparation is mainly for ultrasound images because the clinical data is structured tabulated data and cleaned. The missing clinical data will not be imputed and left blank. The missing data were shown in detailed in Table 3-2 below.

Table 3-2 Missing clinical data

Characteristics	Numbers of missing data (%)
Demographic data	
BMI, kg/m^2, mean (SD)	12 (0.07)
Family history of breast cancer	21 (0.13)
No	
Yes	
Family history of ovarian cancer	40 (0.25)
No	
Yes	
History of breast biopsy	29 (0.18)
No	
Yes	
History of smoking	16 (0.10)
Never	
Ever	
History of alcohol drinking	17 (0.11)
Never	
Ever	
Reproductive history	
Age at menarche, years, mean (SD)	253 (1.60)
Age at first live birth, years, mean (SD)	4 (0.04)
Age at menopause, years, mean (SD)	771 (7.04)
History of breastfeeding	13 (0.13)
Never	
Ever	
External hormone use	
History of HRT use	71 (0.65)
Never	
Ever	
History of OC use	28 (0.18)
Never	
Ever	
History of hormone injection use	44 (0.28)
Never	
Ever	

The ultrasound studied are stored in Digital Imaging and Communications in Medicine (DICOM) files that consist 10 parts but simply divided into image dataset (e.g., patient's characteristic, some ultrasound's parameter) and DICOM image (image pixel

intensity data)[41]. After data retrieval process, all data that can refer to the patients will be removed for privacy and bias reducing, and DICOM image will be identified by using research ID. Based on literature review, DICOM image can be directly used for further analytic process. However, if there is any analytic problem, I will convert DICOM image to Tagged Image File Format (TIFF image), which is high quality and acceptable for publication. Other format such as Joint Photographic Experts Group (JPEG), Graphics Interchange Format (GIF), or Portable Networks Graphics (PNG) can be used but less acceptable for publication or lower resolution even though the file is compressed into smaller size as compared to TIFF image[41]. Unfortunately, if TIFF image cannot be used JPEG image, which I have more experience dealing with this file type, will be the next format to be used.

In each ultrasound study, there are several images captured during the study procedure, which may contain only one lesion or more than one lesions. Each lesion can be categorized into different BIRADs category, but in the report only the highest category will be a representative of the study indicating the highest risk of malignancy. In this study, the only one image will be selected to represent the lesion in the same BIRADS category corresponding to the report, called final BIRADS. Because of the report format of Ramathibodi hospital, a school of medicine and also a tertiary care center, the radiologists always stated the lesion that has the highest BIRADS category in the report and they captured the associated images to the lesion. Thus, it is not difficult to identify and choose an image among several stored breast ultrasound images in PACS. The physicians, who are responsible for image selection, are Suppasilp C. (nuclear medicine radiologist) and Sapankaew T. (general physician). The weighted Cohen's kappa coefficient will be assessed for inter-observer agreement between Suppasilp C., Sapankaew T. and Wiratkapun C. (an experienced radiologist working in the field of mammogram and breast ultrasound) for about 10% of total cases for each BIRADS categories. Please note that there are both non-randomized and randomized processes for image selection the inter-observer agreement will be assessed only for non-randomized processes.

The details of image selection are described as follow:

BIRADS 1, negative study that has no abnormal lesion, the representative image will be simply randomized from several stored images.

BIRADS 2, benign(s) lesion that has about 0% likelihood of malignancy, the representative image will be selected as the most suspicious lesion according to the report. If there is no the most suspicious lesion reported, the representative will be simply randomized.

BIRADS 3, probably benign(s) lesion has less than 2% likelihood of malignancy, the representative image will be selected as the most suspicious lesion according to the report. If there are more than one BIRADS-3 lesion, the representative will be simply randomized from all BIRADS-3 lesions, abandon all the lesions below BIRADS 3.

BIRADS 4 and 5, suspicious for malignancy from low to highly suggestion, the representative image will be the lesion underwent biopsy, BIRADS recommends to biopsy BIRADS-4,5 lesion. If there are more than one biopsy lesions or there is no biopsy lesion, the representative will be selected in the same fashion as the BIRADS-3 selection.

After this selection, each record will consist of one ultrasound image, a matrix of corresponding clinical data, and BIRADS category as reference standard and it is ready for analysis.

3.6 Data analysis/ data mining technique

3.6.1 Overview

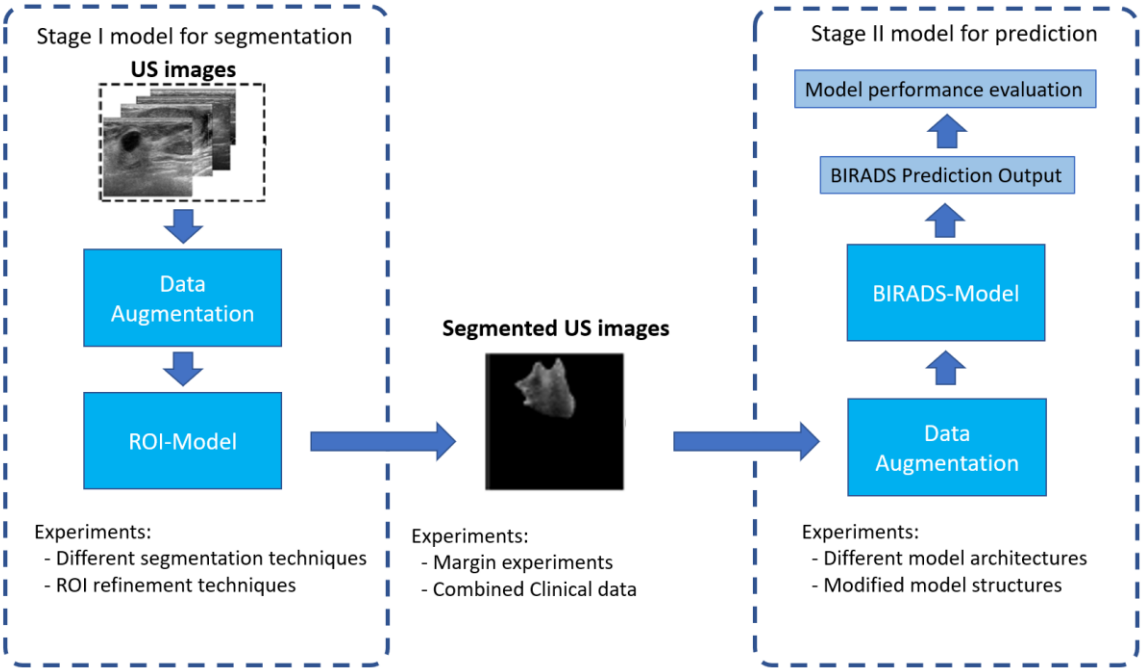
This study proposes two-stage model for predicting BIRADS categories using ultrasound images and clinical data. These two stages are the model for region-of-interest (ROI) segmentation and using those ROI-image as inputs for the second model that is for predicting BIRADS categories, which is one of the aims of the study as shown in Figure 3-1 below.

The records will be divided into train and test set in the ratio of 70:30 and for the test set will be separated equally into validation and actual test set. Thus, the ratio of

training: validation: test is 70: 15: 15. About missing clinical data, we plan to do not perform data imputation due to small percentage of missing valves shown in Table 3-2.

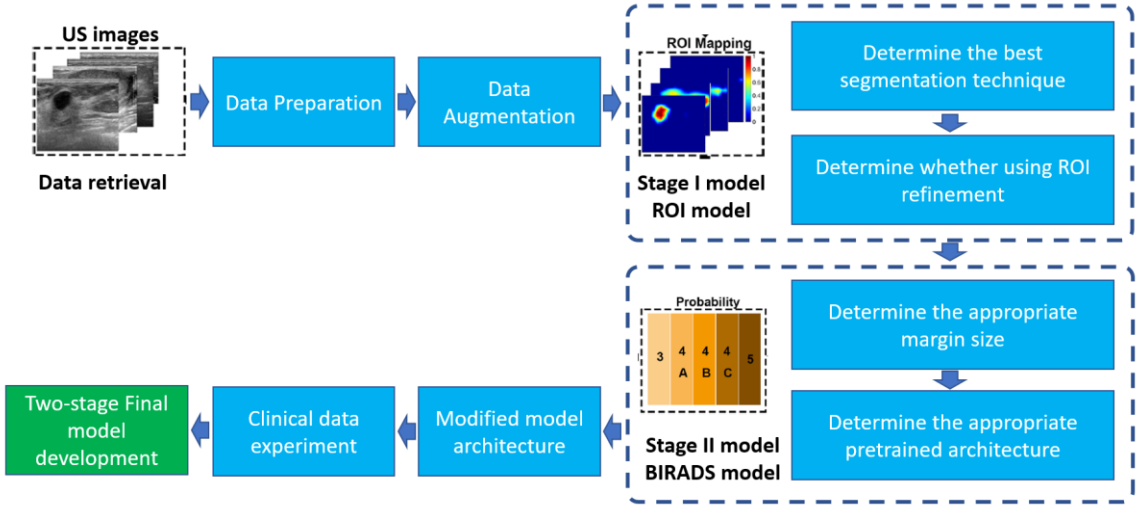
All models will be done by using Python programming, and the software versions including Anaconda and their libraries will be fixed accordingly for reproducibility issues.

Figure 3-1 Overall concept of proposed two-stage model



Because of complex model and many experiments, the study flow is shown below in Figure 3-2. The details of each steps will be stated accordingly in the following sessions.

Figure 3-2 Study flow



3.6.2 Data augmentation

Machine learning often required sufficient amount of training dataset, which is usually large, in order to achieve the acceptable performance. Unfortunately, medical images are limited, therefore data augmentation becomes a commonly used method to increased number of training data. Different augmentation strategies were utilized such as rotation, horizontal flip, vertical flip, random crop, scaling, translation, noising, random brightness, random zoom, etc[42]. Among several techniques, vertical flip is not appropriate because skin and connective tissue below the skin may have important characteristic such as skin retraction, which increases risk of being malignancy. Also, posteriorly to the lesion, for example, posterior acoustic shadowing from the lesion has different risk of being malignancy. Thus, vertical flip, swapping anterior and posterior area of the lesion, is not appropriate. Importantly, the augmentation will implement for the training dataset only. The validation and test dataset will be preserved originally.

3.6.3 Stage 1: image segmentation model (ROI model)

The first stage of the model is breast ultrasound segmentation, it will be considered as “ROI model” from now on. The aim of ROI model is to segment the lesion

from normal background of ultrasound image. Several techniques, both semi-automate and automate segmentation, were adopted in past two decades. Semi-automate methods, user usually begins with specifying a region of interest (ROI) containing the lesion, or a seed in the lesion, or an initial boundary of the lesion, while fully automatic segmentation needs no intervention during the segmentation process at all. However, breast ultrasound segmentation is still changing and in state of art due to low image quality caused by speckle noise, intensity inhomogeneity, different machine parameter, low contrast, weak boundary, and artifacts[43, 44]. Aim of ROI model is only for lesion segmentation and be used in the next stage and we do not aim to contribute new proposed model or to reach better model performance or to create new improved model. Our hypotheses are the segmentation will improve the final model as compared to using the whole image and will focus on unsupervised fully automate segmentation due to large number of sample size and real-world clinical application.

Seven breast ultrasound image segmentation techniques were summarized by Huang Q. et. al.[44] including:

1. Thresholding-based
2. Clustering-based
3. Watershed-based
4. Graph-based
5. Active contour model (ACM)
6. Markov random field (MRF)
7. Neural network model (NN)

Thresholding-based, clustering-based, watershed-based, MRF, NN can be fully automatic procedures, while the others are semi-automate methods. In recent years, convolutional neural networks (CNNs), a branch of deep learning, have become more popular and they achieve highest overall model accuracy, but deep learning in medical images requires labeled ground truth, which is very time consuming. Thresholding-based clustering-based, watershed-based methods are not complicated and widely used to segmenting breast ultrasound images. The performances for the images are about 86% to 92% true positive[44] that were good but not excellent, however, it can be improved by using

multi-stage[45] or cooperate with some adjustment such as expectation maximization (EM) algorithm[46]. Thus, this study will focus only on unsupervised fully automate methods, which are thresholding-based, clustering-based, watershed-based, MRF, and the performances of the models will be compared by using test dataset that manually contoured by a physician and by the results from the BIRADS models. Again, the ROI model is just for data preprocessing before entering the images into the final BIRADS model and the ROI-models performance evaluation is for selecting the best method that is the most appropriate to our dataset. ROI-models performance will be discussed in a following session. The best ROI model will be selected and used for processing all the image for the next phase.

3.6.4 ROI refinement experiment

As mentioned, ROI model alone might be not good enough for image segmentation, using multi-stage[45] or cooperate with some adjustment such as expectation maximization (EM) algorithm[46] may improve the overall model performance. Rahman M. published an unsupervised segmentation algorithm for breast ultrasound images using local histogram features[45], combining Dirichlet Process Mixture Model (DPMM) Based non-parametric Bayesian clustering with postprocessed through merging different segments to obtain the final image segmentation. The results show better Dice Similarity
Coefficient and Boundary Displacement Error (BDE) as compared to the prior model[47]. In application of predicting malignancy breast ultrasound lesion, Huang Y. et. al.[16] showed that refined ROI-CNN can perform the better DSC average value and lowest average error distance, thus establishing the most similarity to the ground truth lesion. More importantly, the BIRADS prediction performances of refined ROI are better regardless the BIRADS categories 3, 4A, 4B, 4C, and 5.

Active contour model is semi-automate segmentation required initial boundary prior to complete the segmentation. Thus, after submitting the images into our ROI-model, the non-refined ROI will be initial boundary for active contour model. Combining ROI model with active contouring, the overall model is still automated image segmentation model but with refined ROI as the result. There are several active contour models, such as

snake model, geodesic model, but most of them rely on the edge-function depending on the image gradient to stop the curve evolution. So, strong gradient is necessary to reach satisfy performance. Breast ultrasound image is usually noisy and blurred edge, classical active contour model may not provide enough accuracy defining lesion edges. In 2001, Chan-Vese proposed the active contour model without edges[48] called C-V model, that is modified active contour model without using stopping edge-function, thus it is not based on the gradient of the image for the stopping process. The method is based on Mumford-Shah segmentation techniques, so the model can detect contours using both with or without gradient between the lesion and surrounding soft tissue. Mimicking the process of Huang et. al.[16] with some modifications, following steps will be applied:

1. Removing the connected small regions and less than 40% of the max area
2. If there are more than one region, the largest area or the connected region closest to the image center
3. Refining the boundary with a typical C–V level-sets methodology by following mathematical formula

$$E(C) = \mu_1 \int_{inside(C)} |I(x,y) - c_1|^2 dxdy + \mu_2 \int_{outside(C)} |I(x,y) - c_2|^2 dxdy + \alpha\kappa$$

Where I is the image, C refers to the boundary of the segmented region, c1 and c2 are the respective averages of I inside and outside C, and κ is the curvature of C.

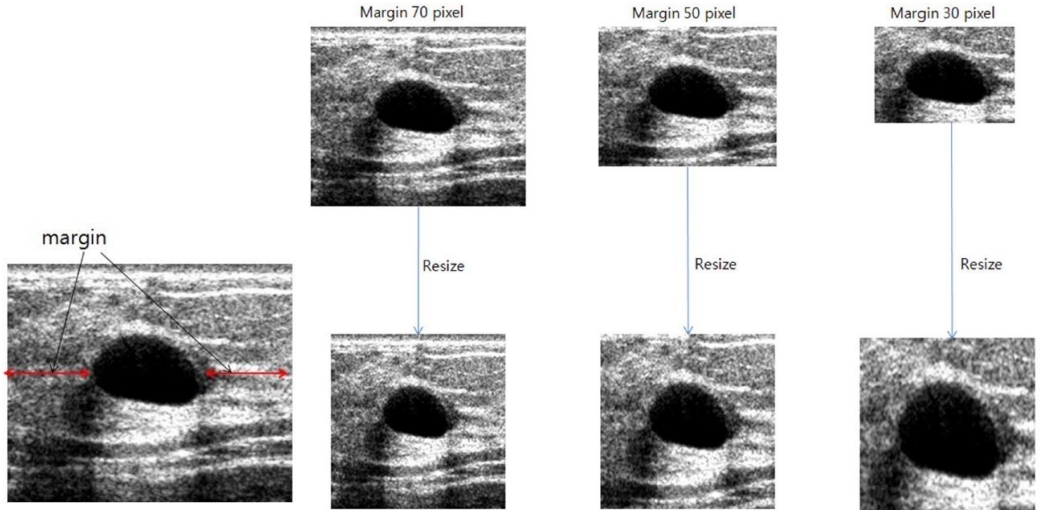
These processes are performed to ensure that only the lesion will be used as an input to subsequent BIRADS model and according to our hypothesis these methods should improve the accuracy of the final BIRADS categorization. This study will experiment about ROI refinement on vary image segmentation methods and will choose the best image segmentation and ROI refinement method with the highest incremental performance for using as an image pre-processing prior to use those images in our BIRADS model.

3.6.5 Margin experiment

After the ROI model, the region of interest (ROI) should contain only the interested area representing the breast lesion in an ultrasound image. As mentioned, we hypothesized that refined ROI should be better performance than non-refined ROI. Rather than using only lesion's characteristics, BIRADS categories are determined by using outside-the-lesion characteristics. For example, posterior acoustic features of a lesion; no posterior acoustic feature, enhancement, shadowing, or mixed pattern, which have different likelihood of cancer; 32%, 28%, 35%, 5%, respectively[49]. So, using characteristics outside of the lesion may improve the BIRADS models. Margin in this specific context is defined as the distance between the lesion boundary and the boundary of the cropped image itself in squared shape, as shown by the example in the Figure 3-3 below. This concept is adopted from Seokmin Han et. al.[50] exploring different size of margins from 0 to 270 pixels, and the results showed that 180-pixel margin improves the accuracy to 90.21% validation accuracy (SD of 0.3076) as compared to the 0-pixel margin of 87.85% accuracy (SD of 0.2354) and the highest performance were reached when combined 0-pixel with 180-pixel margin images, accuracy of 91.23% (SD of 0.1087).

This study will experiment on not only the size of margins that give the best discriminating performance but also comparing non-refined ROI with or without margin, refined ROI with or without margin. We hypothesized that non-refined ROI with margin will give similar performance compared to the refined ROI, and the ROI with margin combining with 0-pixel margin image will show the highest model performance as the recent paper[50]. Furthermore, our study will firstly contribute the model performance comparison among non-refined ROI, refined ROI, and margin experiment. Again, we will select the best segmentation method with refinement technique and the best margin as the pre-processing prior to put the images into the stage 2 or BIRADS model.

Figure 3-3 The concept of margin and examples in the experiment



(Images are from the publication of Seokmin Han et. al.[50])

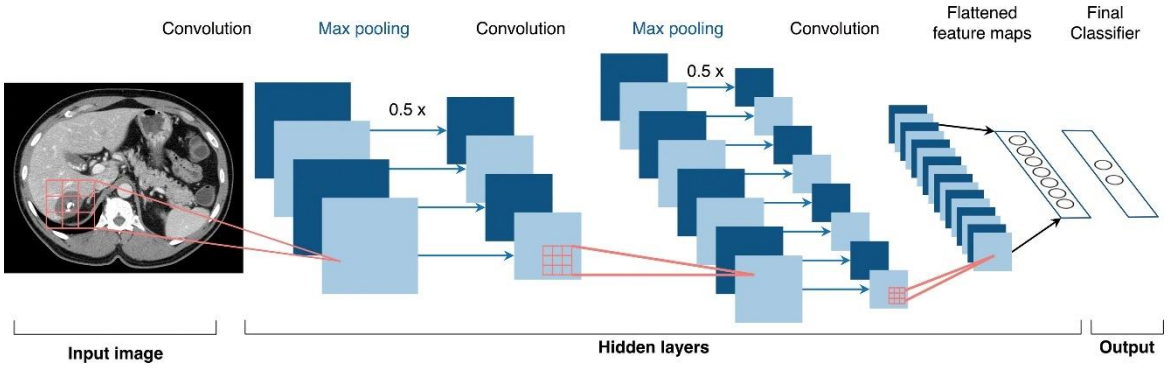
3.6.6 Stage 2: BIRADS grading model (BIRADS model)

Convolutional neural network (CNN) is a special machine learning algorithm called deep learning that are the most popular architectures for image analysis nowadays. Briefly, CNNs have characteristic layers that perform convolutional operation aiming to filter the pixel image matrix to extract spatially correlated features of the input image and offer some shift invariance, then the feature map is created as a linear output. The feature map created is initially passed through an activation function, commonly used a rectified linear unit (ReLU), which convert negative values to 0 and keep positive values as the same, enables it to be an output. The next layer is pooling layer that is down-sampling of the feature map into the concise and smaller feature matrix, which can still include the important values of the feature map. The convolutional layers, including kernel or filler layers and pooling layers, were design specifically, which may contain hundreds of layers, to gain the small output feature mapping. The latest feature mapping will be passed into the fully connected layer to classify the final outcome, BIRADS category in our model. Figure 3-4[38] below shows concept of CNNs of by using a trans axial upper abdominal image as an input image. The input image is passed through a series of convolution and pooling layers, using maximum pooling technique, producing a stack of features maps. The flattened

feature maps or fully connected layers will be used to make the final classification. In the figure, there are two final classification, but for our study there are more than two as multiclass classification as mentioned above.

We will adopt four established neural network architectures as the pretrained models for comparing different model architectures to find the best one that suits for our dataset, that are GoogleNet, imagenet, VGG, and restnet. Only the best one will be used for the next step. The architecture will be modified in before creating the final two-stage model in the last step shown in the Figure 3-2.

Figure 3-4 General convolutional neural network architectures



3.6.7 Clinical data integration

Few studies make a prediction of breast cancer or BIRADS categories by cooperating clinical data in the model. Thomas el. al.[51] showed that combine mammograms with previous examination and clinical and demographic risk factor data resulted in a higher area under the receiving operating curve of 0.943 and achieved a significantly improved specificity to 92.0% at the same sensitivity that are better than using images alone. That information elucidates us to experiment about using clinical data that are from individual risk prediction for breast cancer and accesses the potential of using clinical data in ultrasound images model. Based on our reviews, this is the first study assessing the benefit of clinical data to predict BIRADS categories by using ultrasound images with CNNs techniques.

3.6.8 Model performance evaluation

There are two model in this study, segmentation model and classification model that use different indexes for evaluation. Overall, the precision of each parameter will be represented by using 95% confidence interval and the difference between parameters will be tested statistically using level of significance of 5%.

For segmentation

To quantitatively evaluate the segmentation results, the difference between computed contour and manually sketched contour will be measured using several indexes described as following[44, 52], and the formula and figure are shown in the Table 3-3 and Figure 3-5, respectively:

True positive (TP) is a commonly used that indicates the total fraction of tissue in the "true" tumor region with which segmented region overlaps.

False positive (FP) denotes the amount of tissue falsely identified by the segmentation method as a fraction of the total amount of tissue in the "true" tumor region.

False negative (FN) denotes the fraction of tissue defined in the "true" tumor region that is missed by the segmentation method

Similarity (SI or Jaccard index) measures the overlap ratio, which Jaccard and dice similarity coefficient is interchangeably as using the formular of J=DICE/(2-DICE). For publication, only one that the most represent the similarity will be chosen.

Dice coefficient (DICE) is also a commonly used index for medical image segmentation, which is also regarded as the overlap index, computed by directly comparing the ground truth and the automatic segmentation results, through the measure of spatial overlap rate between two binary images

Precision ratio (PR) denotes the precision ratio between the manually determined contours and the automatically detected contours, that equals to the ratio of the number of pixels that differ between the manually determined contour and the automatically determined contour over the number of pixels in the manual contour.

Normalized residual value (NRV) denotes the areas outside the overlaps between two areas of ground truth and the segmented one as shown by the formula in Table 3-3.

Figure 3-5 Two segmented lesions, A_m is the region covered by manually sketched contours used as the reference standard and A_n is the region covered by contours generated by machine learning algorithm.

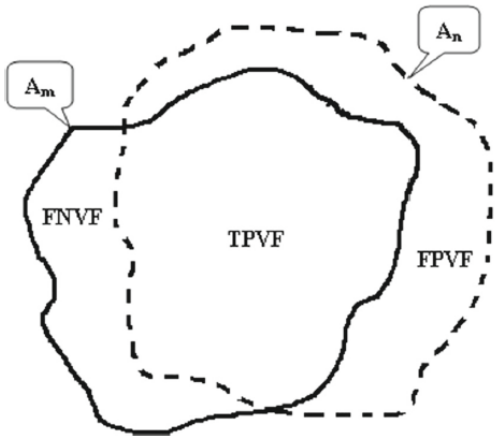


Table 3-3 Evaluation indexes for segmentation

Measurement	Meaning	Formula
TP	True positive	$\frac{A_m \cap A_n}{A_m}$
FP	False positive	$\frac{A_m - A_m \cap A_n}{A_m}$
FN	False negative	1-TP
PR	Precision ratio	$\frac{A_m \cap A_n}{A_n}$
SI	Similarity or Jaccard index	$\frac{A_m \cap A_n}{A_m \cup A_n}$
DICE	Dice coefficient	$\frac{2 * \lvert A_m \cap A_n \rvert}{\lvert A_m \rvert + \lvert A_n \rvert}$
NRV	Normalized residual value	$\frac{(A_m - A_m \cap A_n) \cup (A_n - A_m \cap A_n)}{A_m}$

For classification

To evaluate performance of the classification or BIRADS models, the terms positive (P) and negative (N) refer to the prediction from the model and true (T) or false (F) refer to whether the prediction corresponds to the reference standard. Our model will make the prediction as multiclass classification, so the results will be filled in 2x2 contingency table to be evaluated by these following measurements shown in Table 3-4. The most appropriate measurement will be F1-score, rather than accuracy, to represent overall classification performance of the model due to imbalance data, other measurements will also be considered as adjunct measurements for model selection.

Table 3-4 Classification model performance evaluation

Measurement	Meaning	Formula
TP	True positive	$\frac{P}{T}$
TN	True negative	$\frac{N}{F}$
FP	False positive/ Type I error	1-TN
FN	False negative/ Type II error	1-TP
TPR	True positive rate/Sensitivity	$\frac{TP}{TP+FN}$
TNR	True negative rate/Specificity	$\frac{TN}{TN+FP}$
ACC	Accuracy	$\frac{TP+TN}{TP+FN+TN+FP}$
F1-score	F1-score	$\frac{2TP}{2TP+TN+FP}$

3.6.9 Model performance evaluation per study analysis

Once the best final 2-stage model is selected, the overall model performance per patient's study will be evaluated. One patient's study contains about 10 to 20 ultrasound images, the model will predict BIRADS category for every image and the highest category will be the representative of the predictive BIRADS. This process is similar to the clinical practice that radiologists use for reporting the result. The evaluation will be made in the same fashion as mentioned above, for classification.

3.7 Ethics considerations

This study followed the principles of the Helsinki Declaration and regarding the Medical Research Involving Human Subjects Act. The protocol will be proposed for agreement by the ethics committee of the Faculty of Medicine Ramathibodi Hospital Mahidol University prior to start conducting the research.

3.8 Budget

The estimated budget required for the study with a breakdown to the costs of different activities shown in Table 3-5 below.

Table 3-5 Budget estimation

Category	Budget
Research assistant 15,000 baht/FTE/month x 0.5 FTE x 2 people	180,000
Analysis cost	300,000
Publication cost	100,000
Materials and Supplies 3,000 baht per month	36,000
Total	**616,000**

3.9 Time Frame

Topic	2020					2021									2022		
	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
Proposal defend																	
EC submission																	
Data collection																	
Data preparation																	
ROI model																	
BIRADS model																	
Final Model																	
Model Evaluation																	
Thesis Development																	
Thesis Defend																	
Manuscript																	

ACKNOWLEDGEMENTS

This study proposal is an important part of Chaiyawat Suppasilp training in Ph.D. (Data Science for Health Care) the Faculty of Medicine Ramathibodi Hospital and Faculty of Graduate Studies, Mahidol University.

REFERENCES

1. International Agency for Research on Cancer (IARC) WHOW. Cancer today. 2020.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA: A Cancer Journal for Clinicians. 2020;70(1):7-30.
3. International Agency for Research on Cancer (IARC) WHOW. Breast Cancer Screening: IARC Handbooks of Cancer Prevention Volume 15. 2016;Volume 15.
4. Rebolj M, Assi V, Brentnall A, Parmar D, Duffy SW. Addition of ultrasound to mammography in the case of dense breast tissue: systematic review and meta-analysis. British Journal of Cancer. 2018;118(12):1559-70.
5. Sood R, Rositch AF, Shakoor D, Ambinder E, Pool K-L, Pollack E, et al. Ultrasound for Breast Cancer Detection Globally: A Systematic Review and Meta-Analysis. Journal of Global Oncology. 2019(5):1-17.
6. Spak DA, Plaxco JS, Santiago L, Dryden MJ, Dogan BE. BI-RADS® fifth edition: A summary of changes. Diagnostic and Interventional Imaging. 2017;98(3):179-90.
7. Radiology ACo, D'Orsi CJ. ACR BI-RADS Atlas: Breast Imaging Reporting and Data System; Mammography, Ultrasound, Magnetic Resonance Imaging, Follow-up and Outcome Monitoring, Data Dictionary: ACR, American College of Radiology; 2013.
8. Anothaisintawee T, Wiratkapun C, Lerdsitthichai P, Kasamesup V, Wongwaisayawan S, Srinakarin J, et al. Risk factors of breast cancer: a systematic review and meta-analysis. Asia Pac J Public Health. 2013;25(5):368-87.
9. Louro J, Posso M, Hilton Boon M, Román M, Domingo L, Castells X, et al. A systematic review and quality assessment of individualised breast cancer risk prediction models. British Journal of Cancer. 2019;121(1):76-85.
10. Le EPV, Wang Y, Huang Y, Hickman S, Gilbert FJ. Artificial intelligence in breast imaging. Clin Radiol. 2019;74(5):357-66.
11. Houssami N, Kirkpatrick-Jones G, Noguchi N, Lee CI. Artificial Intelligence (AI) for the early detection of breast cancer: a scoping review to assess AI's potential in breast screening practice. Expert Rev Med Devices. 2019;16(5):351-62.
12. Zhou Z, Wu W, Wu S, Tsui P-H, Lin C-C, Zhang L, et al. Semi-automatic Breast Ultrasound Image Segmentation Based on Mean Shift and Graph Cuts. Ultrasonic Imaging. 2014;36(4):256-76.
13. Levman J, Warner E, Causer P, Martel A. Semi-automatic region-of-interest segmentation based computer-aided diagnosis of mass lesions from dynamic contrast-enhanced magnetic resonance imaging based breast cancer screening. J Digit Imaging. 2014;27(5):670-8.
14. Shan J, Cheng HD, Wang Y. Completely Automated Segmentation Approach for Breast Ultrasound Images Using Multiple-Domain Features. Ultrasound in Medicine & Biology. 2012;38(2):262-75.
15. Fujioka T, Kubota K, Mori M, Kikuchi Y, Katsuta L, Kasahara M, et al. Distinction between benign and malignant breast masses at breast ultrasound using deep learning method with convolutional neural network. Jpn J Radiol. 2019;37(6):466-72.
16. Huang Y, Han L, Dou H, Luo H, Yuan Z, Liu Q, et al. Two-stage CNNs for computerized BI-RADS categorization in breast ultrasound images. Biomed Eng Online. 2019;18(1):8.

17. He T, Puppala M, Ogunti R, Mancuso JJ, Yu X, Chen S, et al., editors. Deep learning analytics for diagnostic support of breast cancer disease management. 2017 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2017; 2017.
18. Luo W-q, Huang Q-x, Huang X-w, Hu H-t, Zeng F-q, Wang W. Predicting Breast Cancer in Breast Imaging Reporting and Data System (BI-RADS) Ultrasound Category 4 or 5 Lesions: A Nomogram Combining Radiomics and BI-RADS. Scientific Reports. 2019;9(1):11921.
19. Virani S, Bilheem S, Chansaard W, Chitapanarux I, Daoprasert K, Khuanchana S, et al. National and Subnational Population-Based Incidence of Cancer in Thailand: Assessing Cancers with the Highest Burdens. Cancers (Basel). 2017;9(8):108.
20. Putthasri W, Tangcharoensathien V, Mukem S, Jindawatana W. Geographical distribution and utilization of mammography in Thailand. Reg Health Forum. 2004;8:84-91.
21. Anothaisintawee T, Teerawattananon Y, Wiratkapun C, Srinakarin J, Woodtichartpreecha P, Hirunpat S, et al. Development and validation of a breast cancer risk prediction model for Thai women: a cross-sectional study. Asian Pac J Cancer Prev. 2014;15(16):6811-7.
22. Laoitthi P PN. Breast cancer: Epidemiology, prevention and screening recommendations. Chula Med J 2016;60(5):11.
23. Li CI, Uribe DJ, Daling JR. Clinical characteristics of different histologic types of breast cancer. British journal of cancer. 2005;93(9):1046-52.
24. Giuliano AE, Edge SB, Hortobagyi GN. Eighth Edition of the AJCC Cancer Staging Manual: Breast Cancer. Annals of Surgical Oncology. 2018;25(7):1783-5.
25. Stout NK, Lee SJ, Schechter CB, Kerlikowske K, Alagoz O, Berry D, et al. Benefits, harms, and costs for breast cancer screening after US implementation of digital mammography. J Natl Cancer Inst. 2014;106(6):dju092.
26. Rao AA, Feneis J, Lalonde C, Ojeda-Fournier H. A Pictorial Review of Changes in the BI-RADS Fifth Edition. Radiographics. 2016;36(3):623-39.
27. Thurfjell EL, Lernevall KA, Taube AA. Benefit of independent double reading in a population-based mammography screening program. Radiology. 1994;191(1):241-4.
28. Anderson EDC, Muir BB, Walsh JS, Kirkpatrick AE. The efficacy of double reading mammograms in breast screening. Clinical Radiology. 1994;49(4):248-51.
29. Bennett RL, Sellars SJ, Blanks RG, Moss SM. An observational study to evaluate the performance of units using two radiographers to read screening mammograms. Clin Radiol. 2012;67(2):114-21.
30. Information Technology Division NCI. Thailand breast cancer screening: practice and guideline. 2017.
31. Lazarus E, Mainiero MB, Schepps B, Koelliker SL, Livingston LS. BI-RADS Lexicon for US and Mammography: Interobserver Variability and Positive Predictive Value. Radiology. 2006;239(2):385-91.
32. Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment. AJR Am J Roentgenol. 2000;174(6):1769-77.
33. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst. 1989;81(24):1879-86.
34. Rosner B, Colditz GA. Nurses' health study: log-incidence mathematical model of breast cancer incidence. J Natl Cancer Inst. 1996;88(6):359-64.

35. Offit K. BRCA mutation frequency and penetrance: new data, old debate. J Natl Cancer Inst. 2006;98(23):1675-7.
36. Kaplan A, Haenlein M. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons. 2019;62(1):15-25.
37. Aggarwal C. Neural Networks and Deep Learning: A Textbook2018.
38. Chartrand G, Cheng PM, Vorontsov E, Drozdzal M, Turcotte S, Pal CJ, et al. Deep Learning: A Primer for Radiologists. RadioGraphics. 2017;37(7):2113-31.
39. Gromet M. Comparison of Computer-Aided Detection to Double Reading of Screening Mammograms: Review of 231,221 Mammograms. American Journal of Roentgenology. 2008;190(4):854-9.
40. Salim M, Wåhlin E, Dembrower K, Azavedo E, Foukakis T, Liu Y, et al. External Evaluation of 3 Commercial Artificial Intelligence Algorithms for Independent Assessment of Screening Mammograms. JAMA Oncol. 2020;6(10):1581-8.
41. Varma DR. Managing DICOM images: Tips and tricks for the radiologist. Indian J Radiol Imaging. 2012;22(1):4-13.
42. Hussain Z, Gimenez F, Yi D, Rubin D. Differential Data Augmentation Techniques for Medical Imaging Classification Tasks. AMIA Annu Symp Proc. 2018;2017:979-84.
43. Xian M, Zhang Y, Cheng HD, Xu F, Huang K, Zhang B, et al. A Benchmark for Breast Ultrasound Image Segmentation (BUSIS). arXiv e-prints. 2018:arXiv:1801.03182.
44. Huang Q, Luo Y, Zhang Q. Breast ultrasound image segmentation: a survey. International Journal of Computer Assisted Radiology and Surgery. 2017;12(3):493-507.
45. Rahman MM, editor An unsupervised segmentation algorithm for breast ultrasound images using local histogram features. 2016 International Conference on Medical Engineering, Health Informatics and Technology (MediTec); 2016 17-18 Dec. 2016.
46. Baby D, Devaraj SJ, Mathew S, Anishin Raj MM, Karthikeyan B. A Performance Comparison of Supervised and Unsupervised Image Segmentation Methods. SN Computer Science. 2020;1(3):122.
47. Prabusankarlal KM, Thirumoorthy P, Manavalan R. Segmentation of Breast Lesions in Ultrasound Images through Multiresolution Analysis Using Undecimated Discrete Wavelet Transform. Ultrasonic Imaging. 2015;38(6):384-402.
48. Chan TF, Vese LA. Active contours without edges. IEEE Transactions on Image Processing. 2001;10(2):266-77.
49. Stavros AT, Thickman D, Rapp CL, Dennis MA, Parker SH, Sisney GA. Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. Radiology. 1995;196(1):123-34.
50. Han S, Kang HK, Jeong JY, Park MH, Kim W, Bang WC, et al. A deep learning framework for supporting the classification of breast lesions in ultrasound images. Phys Med Biol. 2017;62(19):7714-28.
51. Schaffter T, Buist DSM, Lee CI, Nikulin Y, Ribli D, Guan Y, et al. Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms. JAMA Netw Open. 2020;3(3):e200265.
52. Almajalid R, Shan J, Du Y, Zhang M, editors. Development of a Deep-Learning-Based Method for Breast Ultrasound Image Segmentation. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA); 2018 17-20 Dec. 2018.

TABLES

FIGURES

Figure 2-1 Age-standardized incidence rates and age-standardized mortality rates of breast cancer in 2020