CONVOLUTIONAL NEURAL NETWORK FOR EARLY DIAGNOSING OF BREAST CANCER

A DISSERTATION IN
Mathematics
And
Biomedical and Health Informatics

Presented to the Faculty of the University
Of Missouri-Kansas City in partial fulfillment of
the requirements for the degree

DOCTOR OF PHILOSOPHY

by ZHIHENG ZHANG

B. S. Shanghai JiaoTong University, 2014
M.S. Stony Brook University, 2016
M.S. University of Missouri-Kansas City, 2020

Kansas City, Missouri
2022

CONVOLUTIONAL NEURAL NETWORK FOR EARLY DIAGNOSING OF BREAST CANCER

Zhiheng Zhang, Candidate for the Doctor of Philosophy Degree

University of Missouri-Kansas City

ABSTRACT

Due to the high demand for mammograms, radiologists are swamped with many patients' mammograms. Radiologists' workload has increased dramatically over the last 15 years during on-call hours. This rise is due to an increase in the number of computed tomography (CT). Previous research published in the early 2000s found a 22% increase in radiology examinations during on-call hours in the United States over four years. The shortage of radiologists has recently worsened due to the COVID-19 pandemic. Based on their training and experience, radiologists classify mammography images as benign or malignant. Automating this diagnostic process with a machine learning algorithm may improve diagnosis speed and accuracy. This research aims to propose an effective automated algorithm system that lessens the workload of radiologists while improving speed and accuracy.

This study uses Convolution Neural Networks (CNN) to assist radiologists in classifying lesions by incorporating repeated imaging. In addition, longitudinal electronic health record (EHR) data prior to diagnosis will augment the ML algorithm to improve its accuracy. This study takes a novel approach to the fusion model because only a few models have been developed to integrate both clinical and

imaging data for diagnostic purposes, and few of them have included both EHR and longitudinal imaging data. Therefore, this study compared various multi-modal fusion and single models that can use pixel-based data (image) and EHR data from a large urban medical center.

APPROVAL PAGE

The faculty listed below, appointed by the Dean of the School of Graduate Studies have examined a dissertation titled “Convolutional Neural Networks for Early Diagnosing of Breast Cancer” presented by Zhiheng Zhang, candidate for the Doctor of Philosophy, and certify that in their opinion it is worthy of acceptance.

Supervisory Committee

An-Lin Cheng, PhD., Committee Chair and Advisor

Department of Mathematics and Statistics

Jenifer Allsworth, PhD., Co-discipline Advisor

Department of Biomedical and Health Informatics

Majid Bani-Yaghoub, Ph.D., Member

Department of Mathematics and Statistics

Yong Zeng, Ph.D., Member

Department of Mathematics and Statistics

Noah Rhee, Ph.D., Member

Department of Mathematics and Statistics

CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

ACKNOWLEDGEMENTS

This research would never have reached completion without the support of my family, advisors, and University Health (formerly Truman Medical Center). My parents' encouragement was instrumental in making progress. Dr. AnLin Cheng was a great source of guidance and direction in my research. I want to express my thanks to the database staff in the radiology department of University Health. I would also like to thank everyone that agreed to be a part of my doctoral committee: Dr. Jenifer Allworth, Dr. Majid Bani-Yaghoub, Dr. Yong Zeng, and Dr. Noah Rhee; Thank you all very much for your time and support.

CHAPTER 1

INTRODUCTION

1.1 Background and Literature Review

Breast cancer is the most common cancer globally in the United States. Breast cancer is the second most common type of cancer in women. According to the National Cancer Institute, nearly one in eight females in the United States will develop ductal carcinoma a common type of early-stage breast cancer during their lifetime. In 2019, there were 268,600 cases of inflammatory breast cancer, and approximately 2,670 cases were diagnosed in men (Street, 2019). If a female's sister, mother, or daughter has been diagnosed with breast cancer, the chance of developing breast cancer will nearly triple (Ginsburg,2020). This trend may also reflect an increase in the detection of asymptomatic metastases due to the increased use of advanced imaging (DeSantis, 2019).

The overall breast cancer death rate increased by 0.4% per year from 1975 to 1989 but decreased rapidly, with a total decline of 40% through 2017 because of advanced imaging methods for early detection. As a result, 375,900 breast cancer deaths were averted in US women from 1989 to 2017 (Street, 2019).

According to the report (Stewart, 2014), In 2012, breast cancer accounted for 12% of newly diagnosed cancers and 25% of all female cancers. Breast cancer develops when the mammary gland cell begins to grow abnormally. These abnormal cells can combine to form tumors, which can be benign or malignant. A lump in the breast, a change in breast shape, dimpling of the skin, fluid coming from the nipple, a newly inverted nipple, or a red or scaly patch of skin are signs of breast cancer (Stewart, 2014). According to the table below, females in the United States are at an

increased risk of developing breast cancer. The last five annual Cancer Statistics Review reports show the following estimates of lifetime risk of breast cancer.

12.32%, based on statistics for 2010 through 2012
12.43%, based on statistics for 2011 through 2013
12.41%, based on statistics for 2012 through 2014
12.44%, based on statistics for 2013 through 2015
12.83%, based on statistics for 2014 through 2016

Table 1: Trend of Risk of breast Cancer

According to the American Cancer Society, in terms of malignant breast cancer tumors, Invasive ductal carcinoma and invasive lobular carcinoma are the most common types. Invasive ductal carcinoma accounts for roughly 70% to 80% of all breast cancers(American Cancer Society, n.d.). Breast cancer is classified into five stages: stage 0 (zero), which is non-invasive ductal carcinoma in situ (DCIS), and stages I through IV (1–4), which are used for invasive breast cancer. The stage provides a common way for doctors to describe cancer to collaborate to plan the best treatments.

Stage describes the extent of the breast cancer, including the size of the tumor, whether it has spread to lymph nodes, whether it has spread to distant parts of the body, and what biomarkers are present. The TNM system is doctors' most commonly used tool to describe the stage. Doctors use diagnostic tests and scan results to answer the following questions:

I: Tumor (T): How large is the primary tumor in the breast? What are its biomarkers?

II: Node (N): Has the tumor spread to the lymph nodes? If so, where, what size, and how many?

III: Metastasis (M): Has cancer spread to other parts of the body?

A mammogram is a type of medical imaging that uses a low-dose x-ray system to examine the inside of the breasts. A mammography exam, also known as a mammogram, helps women detect and diagnose breast diseases early (Centers for Disease Control [CDC], 2022). Usually, Mammograms are classified into two types: digital and conventional. Both use X-ray radiation to create an image of the breast, but traditional mammograms are read and stored on film,in contrast,digital mammograms are read and stored in a computer, allowing the data to be enhanced, magnified, or manipulated for further evaluation(Joe, 2014). In 2010, 37 million mammograms performed annually in the United States were for screening purpose(Chubak, 2010). In 2021, according to the FDA (Food and Drug Administration in the US). Nearly 39 million mammograms are performed yearly(U. S. Food and Drug Administration, 2021).

A mammogram is still considered the best method for screening women for breast cancer. The reason is that mammogram is very helpful in identifying breast cancer to avoid false-negative results (Miglioretti, 2011). Screenings accounted for one-fourth of breast cancers detected. Women who do not receive regular screening were six times more likely to be grade III (the cancer is more extensive and may have spread to the surrounding tissues and the lymph nodes) at the time of diagnosis and had 3.5 times the risk of death compared to women who take regular screening (Niraula, 2020).

	Women with Cancer		Women without Cancer		Cancer Detection Rate	
Time Since Prior Mammography	No. of Cancers	No. of False-Negative Findings	No. of Noncancers	No. of False-Positive Findings	No. of Examinations	No. of Detected Cancers
No prior mammography	201	38	64057	10419	64258	152
Mammography performed within prior 2 years	772	210	239487	25493	240259	499
Mammography performed 3 or more years prior	215	32	56232	7553	56447	166

Figure 1: Prior mammogram With Possibilities of Cancer

In patients' diagnosing process upon their visit to the medical center. The first step in detecting breast tissue is using a mammogram. It is a fast-screening process. Both the patient and radiologist can see the result in 24-48 hours when possible. The results of the procedure can indicate and determine the condition of the breasts, as well as develop treatment plans for patients. Breast density, defined as the proportion of parenchymal to fatty tissue in mammograms, is a significant and prevalent risk factor. With increasing breast density, the risk of having breast cancer that is masked or hidden on mammography increases, as does the risk of developing breast cancer in the future. Many mammograms show microcalcification, and there are well-defined patterns that help distinguish benign from potentially malignant changes. The lexicon of the Breast Imaging-Reporting and Data System promotes consistency in nomenclature and provides descriptions to distinguish between benign and malignant changes(Sickles, 2013). The picture below illustrates characteristic appearances of benign and malignant calcification(Wilkinson,2017), which illustrates the slight difference between benign and malignant features.

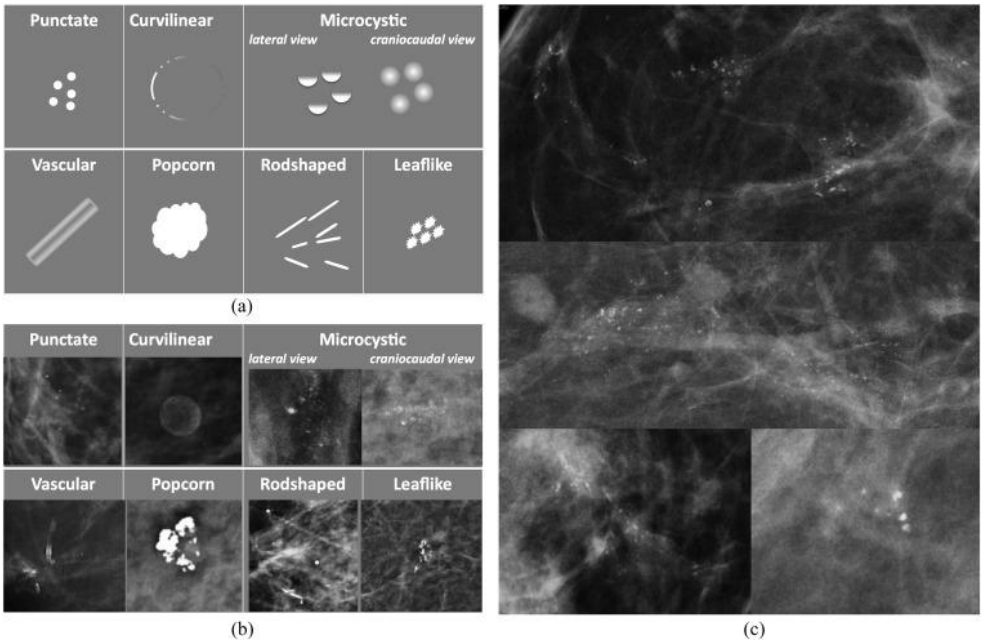


Figure 2 (a) Patterns of calcification associated with a benign change; (b) examples of calcification associated with a benign change; and (c) examples of malignant calcification.

However, mammography has limitations as a main screening tool. Despite the presence of breast cancer, a false-negative mammograms may appear normal. Women with thick breasts are more susceptible to false-negative test findings (Kerlikowske,2019). Even if there is no cancer, a false-positive mammography will seem abnormal. For a definitive diagnosis, abnormal mammograms typically demand further testing (diagnostic mammography, ultrasound, and sometimes MRI or breast biopsy) (Blackwood,1998).

Radiologists are medical doctors (MDs) or doctors of osteopathic medicine (DOs) who have completed a four-year radiology residency program. In treating a disease, a radiologist may act as a consultant to another doctor caring for the patient or as the patient's primary doctor(John Hopkins Medicine n.p). Breast imaging radiologist is one of the specific areas which needs additional training. The training process is relatively longer than other medical imaging diagnosis training in society which causes the shortage of this profession.

According to the journal of Nuclear Medicine, the US has approximately 34,000 expert radiologists, or 100 per million (Acuff, 2017). Many parts of the world are experiencing a shortage of radiologists. Patients' access to imaging is impacted when there is an uneven workforce distribution. Over the last few decades, the growing demand for imaging has also put a strain on resources (ESR&ACR, 2016). Especially during the COVID-19 pandemic, the shortage situation is severe (Tang,2020). If a mammogram reveals any abnormalities or characteristics consistent with a tumor, the patient will undergo further examination. Mammograms are easy to obtain because they are included in the annual physical examination for females over 40 in the United States. Following a suspicious mammogram, people are asked to undergo ultrasound and MRI (Magnetic resonance imaging).

In current medical fields, diagnosis is from a medical image mainly depending on the experience of the radiologist. The radiologist's subjective thinking will influence the diagnosis, resulting in misdiagnosed problems (Brady, 2017). The artificial intelligence (AI) techniques have potential to help the radiologist improve the accuracy as a second reader. In radiology, there are higher expectations for AI's application, which can "see" more than human radiologists in terms of tumor size, shape, morphology, texture, and kinetics — enabling better care through earlier detection or more precise reports (Pakdemirli,2019).

1.2 Basic Method of Classifying System

Artificial intelligence (AI) has significant potential to play an essential role in biology and medicine. According to recent literature, the number of AI publications in diagnostic imaging alone has increased from about 100–150 per year in 2007–2008 to 1000–1100 in 2017–2018. Researchers have used AI to recognize complex patterns in imaging data and provide quantitative assessments of radiographic characteristics.

A proper intelligent learning algorithm can learn from data, extract knowledge, and make decisions in the same way humans do. As a result, AI has always been inspired by human behaviors such as how they learn, extract knowledge, and make decisions. It will impact almost everything in medicine and improve patient outcomes in various ways. For example, it is expected that radiologists' reports will improve and become more precise (Sorantin,2021).

The application of AI in the identification process has become popular due to the high demand for reporting such as automated reporting, systems and computer-based data centers. An accurate AI algorithm can dramatically reduce the burden on medical staff (Bohr,2020). AI can help to liberate healthcare professionals from the strain of tedious and repetitive tasks so they can focus on caring for patients.

Furthermore, it is widely known that AI tools will facilitate and enhance human work rather than completely replace the work of physicians and other healthcare personnel (Alexander,2017). AI is ready to assist healthcare personnel with a wide range of tasks, including administrative workflow, clinical documentation, and patient outreach, as well as specialized assistance in image analysis, medical device automation, and patient monitoring. (Bohr,2020).

AI refers to a broader concept in which robots can execute activities smartly. Machine learning is based on the premise that machines should be able to learn and adapt via experience. Machine learning concept is incorporated into the whole AI systems. Machine learning (ML) is the study of computer algorithms that can improve automatically through experience and by the use of data. It is based on a model of brain cell interaction in part. Donald Hebb developed the model and term in his 1949 book The Organization of Behavior. Hebb's theories on neuron excitement and neuron communication are presented in the book. After nearly 20 years' effort, the first introduced chatbot, which was a conversational tool that recreated the conversation between a psychotherapist and a patient. That also was the early days of applying artificial intelligence and rules-based systems to the interaction between patients and their care providers (Joseph,1966).

Ideally, a practical, automated, and accurate system for early-stage breast cancer detection should be developed to reduce the burden on radiologists while also extending patient survival years. This system will be trained to produce the best results. After each training, the algorithm can learn like a human and gain experience. Modelers can manipulate and adjust the parameters based on the algorithm in the first part to fit testing needs through medical images such as mammograms and other types of medical images.

In traditional machine learning, the learning process is supervised. The programmer must be extremely specific when telling the computer what types of things it should be looking for: For example, if the goal of the ML is to look for a dog profile in an image, dog is a parameter that need to be specified in advanced. The disadvantage of traditional machine learning is that training the classifier is separate from the process of feature extraction, which lacks continuous connections. In addition, the extracted features may not be suitable for the upcoming classification model. The process plot below (Figure.1) is the architecture of traditional machine learning and deep learning approaches, consisting of two independent stages - extraction and classification. Extraction stages are extracted manually by programmers, and this layer can be called handcrafted features.

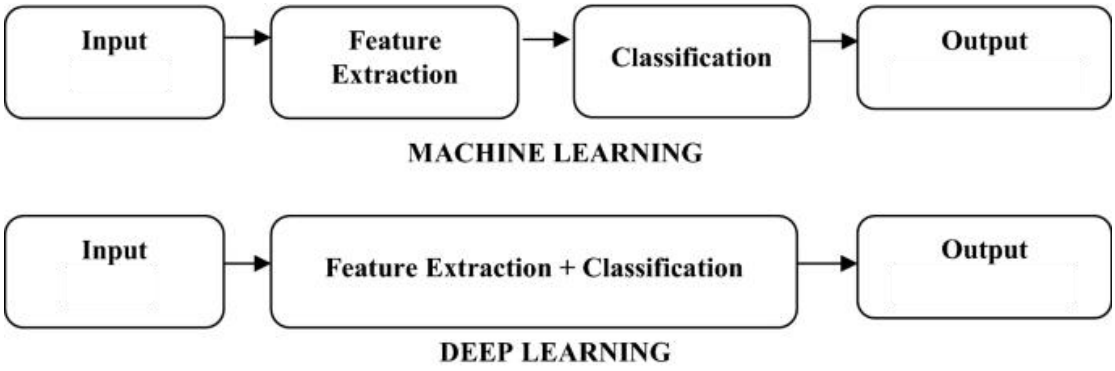


Figure. 3 Classification process of typical machine learning and deep learning

Deep learning is a descriptive learning algorithm with several nonlinear transfer structures. While many medical imaging machine learning models take a traditional machine learning approach, deep learning models have several advantages. The advantage of deep learning is that the program recognizes the feature set by itself without supervision, and it is an end-to-end process, as shown figure.1 showed above. It consists of several layers of the fully connected layer that can function well in an overall network to make predictions (Li, 2014). One example of the model is deep belief networks, where every layer in this network is a typical restricted Boltzmann

machine, which can do unsupervised training layer by layer (Li, 2014). Unlike traditional neural network structures, convolution neural networks reduce the number of parameters that must be optimized. Furthermore, 2D and 3D images can be directly fed into the CNN, making them suitable for extracting partial vision-related information. (Natalie Reznikov, 2020). Deep learning applies the learned knowledge to solve problems in practical situations, reducing the subjective disturbance conducted by radiologists. The deep learning model has widespread application in speech processing, image processing, translation, and natural language processing (NLP). Therefore, it is a natural extension to apply the deep learning method as an effective tool to advance medical image analysis (Mahony,2019).

Current state-of-the-art deep learning models for radiology applications take only pixel-value information (image) into account with no clinical context data. However, relevant and accurate non-imaging data based on the clinical history and laboratory data allow physicians to interpret imaging findings in the appropriate clinical context, resulting in higher diagnostic accuracy, informative clinical decision making, and improved patient outcomes(Huang,2020).

Neural Network, a subset of machine learning, are the heart of deep learning algorithms and a set of algorithms that attempts to recognize underlying relationships in a set of data using a process similar to how the human brain works, the applications of neural network have been well documented in recent articles (Shorten,2019). Usually, a neural network, like the superficial nerve in the human brain, which has one or more hidden layers. These properties can be used to model complex nonlinear problems. The remaining layers of the superficial nerve can perform higher-level abstraction to improve the model's power and accuracy. A Convolution Neural Network (CNN) is one of the neural network structures that can take in an image,

assign importance (learnable weights and biases) to various aspects in the image, and distinguish one from the other. Its architecture is regarded as an End-to-End straight learning box, which means that the feature extraction process is entirely automatic and can operate concurrently with the classification process. It demonstrates that features are learned and picked up from training data without human intervention, a process known as unsupervised learning. This level of independence comes at the expense of having a large amount of training data to train the classifier, which is time-consuming. Mathias discovers that increasing the amount of training data has little effect on performance in the case of classical machine learning approaches (Figure.2).(Mathias Kraus,2020). I will limit the investigation to the deep learning method since it has the best stretching power and performance.

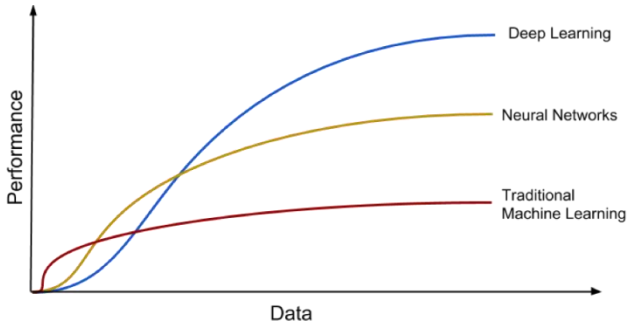


Figure.4 The trend of performance with increasing sample size

Compared to other classification algorithms, the amount of pre-processing required by a CNN is significantly less.CNN can learn these filters/characteristics with enough training. It is a highly adaptable neural network that can be used for various medical image analyses, including covid-19 CT-based diagnosis. (Hartono, 2021). Yadav uses CNN to classify pneumonia in the lung, leading the way for different kinds of convolution neural network structures and choosing methods among them. (Yadav,2019). Dominik Müller creates the open-source MIScnn (medical image segmentation convolution neural network), a simple API (application programming interface) that allows fast setup of the medical image segmentation pipeline. He

proposed a flexible API framework for CT images (Müller, 2021). CNN can be used for AI-assisted diagnosis in various tasks such as image classifying, result predicting and text mining. The initial research is centered on applying a single image that does not take time into account.

Medical image analysis is vital in disease detection, classification, and treatment. With the advancement of the internet and computer technology, computer science's role in modern technology is rapidly expanding. Deep learning applications gradually mature and are applied to various new areas such as data analysis, informatics, bioinformatics, and computer vision (Li, 2014). Deep learning's breakthrough raises the prospect of a simple application in medical image analysis (Kim, 2019). The trained neural network model is used to simulate the real brain's procedures for data analysis and understanding and making predictions for the input image. Finally, the result will be set to a binary variable or the tumor's grade rank (Yadav,2019).

A convolutional neural network (CNN) is a type of neural network commonly used in image analysis. For example, Venugopalan uses multi-modal merging information from MRI (magnetic resonance imaging) and PET (positron emission tomography) to predict Alzheimer's Disease and measure the progression of MCI and early AD (Venugopalan,2021). Likewise, Cruz-Roa employs mixed CNN to automatically extract mammogram features for qualitative classification and early cancer detection (Cruz-Roa,2013).

According to recent literature, the best single model achieved a per-image AUC*(accuracy) of 0.88 on an independent test set of digitized film mammograms from the Digital Database for Screening Mammography (Shen,2019). Molla Hosseini proposed an early detection of breast cancer based on CNN structure using discrete

Shearlet transformation, which increased the accuracy of early detection by an average of 4.6% than the usual CNN structure (Rezaeilouyeh, 2016). Han Silu has explored the unsupervised stacked auto-encoder, three-dimensional convolution neural network, transfer learning, and other deep learning methods for mammography images. (Han,2021). Deep learning methods are also used in the direction of disease evaluation. Raghu has created a high natural image subset for alternative method based on medical images. He developed a new convolution neural network using computer assistance and transfer learning. This model can provide image-assisted diagnosis, particularly for Alzheimer's disease. (Raghu,2019). Cuixia Liang used a multi-weights fusion model to build a breast cancer classification model by combining traditional image properties with deep learning (Liang,2018). They all focus on the single image classification and usually have a mature processing system.

In recent years, deep learning has led to stunning breakthroughs in a variety of research fields such as text extracting (Liang, 2017), image analysis (Yadav,2019), and object detection and auto labeling(Ivars,2019). It can automatically extract deep layers and abstract unique properties, which can then be used to perform quantity and quality analysis and change the way clinical diagnoses are performed. It significantly improves diagnostic efficiency and accuracy. According to abundant literature and the Kaggle competition (competition of optimized algorithms for machine learning), deep learning has a significant advantage in dealing with various types of images.

1.3 Proposed Method And Goal of Research

This study proposes and evaluates a machine learning-based algorithm that can detect breast cancer early and classify tumors as benign or malignant. Medical image analysis is a valuable tool for accurately and quickly recognizing and

diagnosing various lesion-based diseases such as cancer; earlier breast cancer diagnosis at an early stage is linked to a better prognosis (Ginsburg, 2020).

We used in this research an artificial neural network, a series of algorithms that mimic the operations of a human brain to recognize relationships between vast amounts of data (Schmidhuber,2015). This algorithm can mimic the process of human thoughts. In addition, it helps to identify specific essential features of images. The components above can be concluded to the field of application of artificial intelligence.

This research aims to create an algorithm system that will aid doctors and pathologists in early detection by utilizing various machine learning and deep learning methods. The fusion model will be applied to several scenarios and situations that have advantages over the EHR-only model or image-only model (Huang,2020). This algorithm can be set up using both medical images and Electronic Health Records (EHR) and expect to improve the accuracy rate in this research study. Researchers can make the best decision when selecting models based on method comparison.

Electronic health records can be included in the type of longitudinal data as they track patients over time. This type of information is commonly used in the medical field for numerical analysis (Juan Zhao, 2019), particularly in chronic diseases. Electronic health records are real-time, patient-based records that make information available instantly and securely to authorized users. The EHR contains some of the following information:

I.Patient’s medical history

II.Diagnoses and medications

III.Treatment plans

IV.Immunization dates and medical records

V.Laboratory and Test results

Longitudinal data analysis models outcome while accounting for the process of change, which is superior to a single statistics test at each time point. As a result, longitudinal methods are instrumental when studying developmental and lifespan issues. Researchers can observe how certain things change throughout a person's life and investigate the reasons for these changes and developments (Juan Zhao, 2019).

Mammograms are most commonly used for early detection and diagnosis. We know from the previous section that deep learning is an effective tool for image analysis and disease diagnosis. However, because of medical images' unique shape and sequence, extracting dynamic features from longitudinal images is challenging to master. Moreover, it makes applying the deep learning method to time-varying images difficult. This research will propose a new deep learning-based method for extracting features in proper time order and a new algorithm for classifying images into benign or malignant groups. Longitudinal electronic health records will also be incorporated into the algorithm to improve the accuracy of previous image classification using the fusion model (Huang,2020).

In the next chapter, I will address the math concept and basic structure of convolution neural network from simple elements or neurons to the whole network, This is also the fundamental composing part for the whole classifying algorithm system.

CHAPTER 2

FOUNDATION

2.1 Structure

The previous chapter provided background on current breast cancer diagnosis methods and contextualized the need for new, more accurate methods. This research aims to create a deep learning algorithm that will be applied to longitudinal mammography images while accounting for EHR data. Therefore, this chapter will focus on the primary method that should be used and fundamental mathematical concepts.

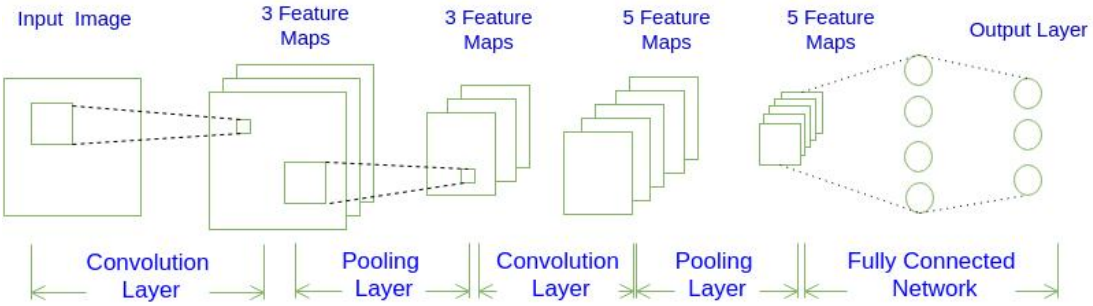


Figure.5 The typical structure of convolution neural network

The above graph depicts the general structure of a deep learning neural network. In the research, the pooling layer and convolution layer extract image features, and the output is a 0 or 1 situation (benign or malignant for breast cancer tumor info). As shown in the diagram above, In the first phase, each image pixel can be treated as a small matrix. A finished CNN comprises several convolution layers, pooling layers, and a fully connected layer. The general structure model of the CNN is

INPUT -> [[CONV]*N -> POOL] *M -> [FC]*K----------(1)

CONV: Convolution layer; POOL: Pooling layer; FC: Fully Connected layer

For the CNN in the figure the structure:

INPUT -> CONV -> POOL -> CONV -> POOL -> FC -> FC-------(2)

Following the form of (1), the figure can be written as

INPUT -> [[CONV]*1 -> POOL] *2 -> [FC]*2----------(3)

A convolution neural network is made up of input layers, hidden layers, and output layers. (Li, 2014) Any middle layers in a feed-forward neural network are hidden because the activation function and final convolution mask their inputs and outputs. The hidden layers typically include convolution layers (directly accumulation plus) in a neural network. This does a dot product of the convolution kernel and the input matrix of the layer. This product is commonly known as the Frobenius inner product, and its activation function is ReLU, which stands for rectified linear activation function, and is a piece-wise linear function that outputs the input directly if it is positive, otherwise, it outputs zero. When the convolution kernel slides along the layer's input matrix, the convolution operation generates a feature map, which contributes to the next layer's input. Other layers such as pooling, fully connected, and normalization layers are added to normalize and deal with the image.

The Convolution neural network has two significant advantages compared to traditional machine learning methods. Firstly, the partial connections have a relatively small number of parameters (existence of pooling layers). Secondly, weight sharing in each connection can be used to reduce parameters and speed up calculation (M. Ahmadi,2019). That is why CNN is the best option for image-related classification. To get a good result on the classification task, CNN can keep the necessary parameters and eliminate many unnecessary parameters when conducting analysis.

2.2 Math Concept and Internal Training Procedures.

Here is a simple example of how to calculate convolution. We can develop the project's key concepts and calculation methods. Assume we have a 5*5 image, use a 3*3 filter to convolve it, and our task is to create a 3*3 feature map. Below is the

example transformation. It illustrates the process of simple convolution process in each cells in the neural network.

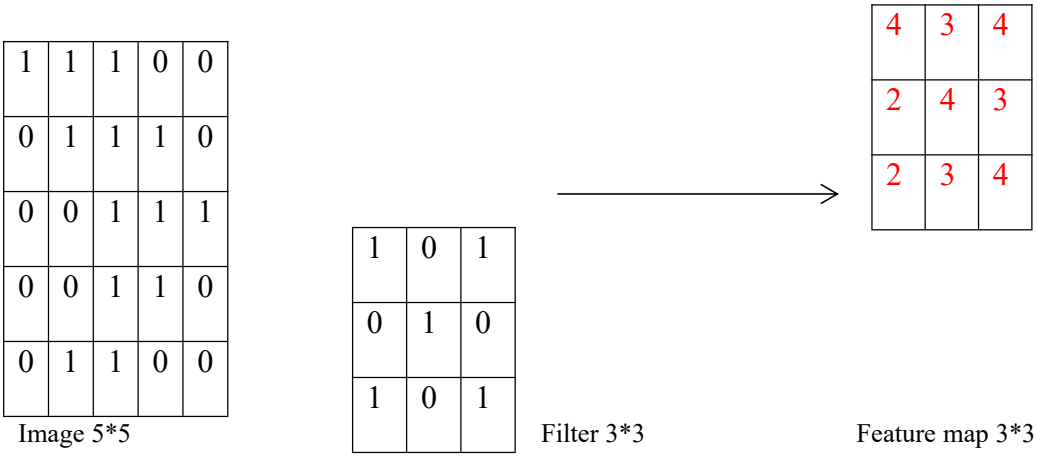


Figure 6: Simple Convolution Calculation Process-1

In order to clearly describe the convolution calculation process, we first number all pixels in the image. $X_{i,j}$ to illustrate i^{th} column and j^{th} element. Then numbering all weigh in the filter using $W_{m,n}$. W_0 is the bias term. In addition, the activation function is numbering all elements in the feature maps called $a_{i,j}$.f (take ReLU function as an example). We have:

$$a_{i,j} = f(\sum_{m=0}^{2}\sum_{n=0}^{2} w_{m,n}x_{i+m,j+n} + w_b) \text{----------(1)}$$

For example, for the left corner element $a_{0,0}$, the calculation method will be

$$a_{0,0} = f(\sum_{m=0}^{2}\sum_{n=0}^{2} w_{m,n}x_{m+0,n+0} + w_b)$$
$$= relu(\sum_{m=0}^{2}\sum_{n=0}^{2} w_{m,n}x_{m+0,n+0} + w_b)$$
$$= relu(1+0+1+0+1+0+0+0+1+0)$$
$$= relu(4) = 4$$

Using the same convolution method (ReLU), a similar result can be filled in the above grid. Then make sure all blanks in the grid are filled. In the above calculation, we set

the stride (the number of pixels shifts over the input matrix) as 1. For the next step, we will try when the stride is equal to 2.

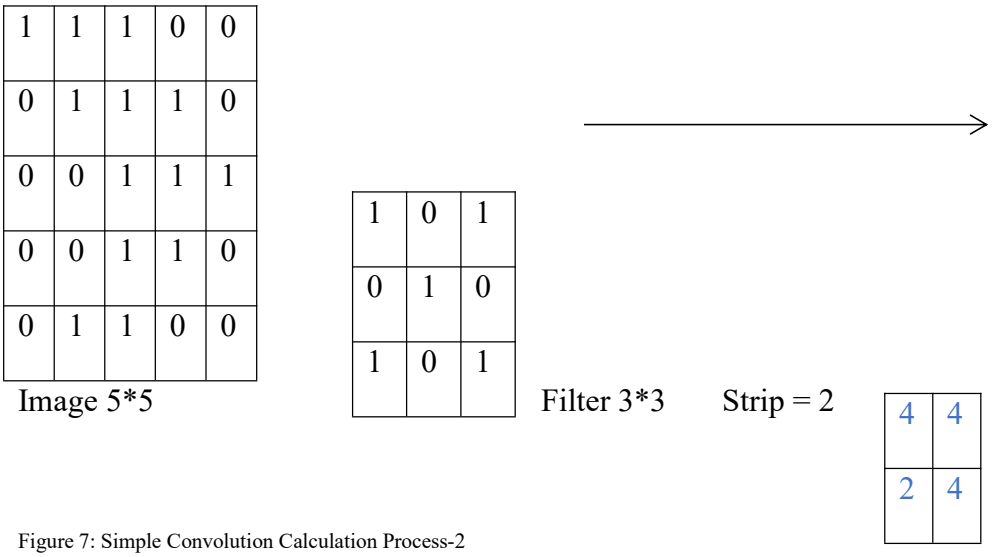


Figure 7: Simple Convolution Calculation Process-2

That indicates that the image size stride has a close relationship with feature map size.

In the following equations, W_2 represents the width of the convolution feature map. W_1 represents the width of image before convolution. F is the width of filter. P is the number of zero padding. If p=1, the image will be added a surrounding 0. which is good for image extracting near the side. S is the stride, H_2 is the height of the feature map after convolution. H_1 is the width of the image before convolution. The (2) and (3) are the same. Using the above two equations, we can easily calculate the dimension of the feature map (see above result).

$$W_2 = (W_1 - F + 2P)/S + 1 \text{----------} (2)$$
$$H_2 = (H_1 - F + 2P)/S + 1 \text{----------} (3)$$

Above is the condition when image depth is equal to 1. We are now considering the image depth equal to 2 or more. If the image depth is equal to 2 before convolution. Filter depths are also equal to 2. Furthermore, if the image depth is equal

to D before convolution, Filter depths are also equal to D. Expanding the equation (1) we get(D>1):

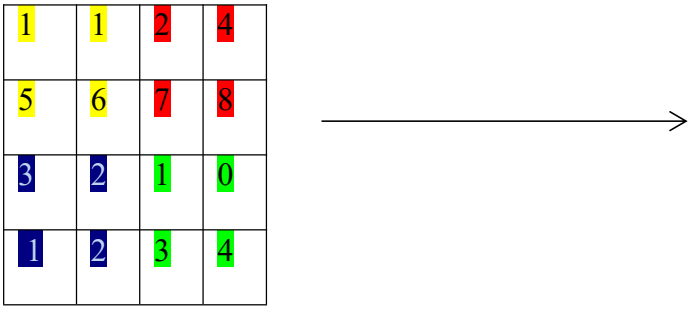
$$a_{i,j} = f\left(\sum_{d=0}^{D-1}\sum_{m=0}^{F-1}\sum_{n=0}^{F-1} W_{d,m,n} X_{d,i+m,j+n} + W_b\right) \text{----------(4)}$$

In (4), D is the depth, and F is the size of the filter (same width and height). $W_{d,m,n}$ represents the weight of the layer m^{th} row and n^{th} columns.$a_{d,i,j}$ represents the pixel of d^{th} layer i^{th} row and j^{th} columns. Other elements are the same with (1).

It is the convolution layer calculation method, which contains sparse connectivity and sharing weights and biases. Each layer of cells is only connected to previous cell parts (convolution calculation rules).

2.3 Pooling Layer

The primary goal of creating a pooling layer is to reduce the dimension of the matrix. The goal is to remove the unimportant parameters. The most common method is max pooling. The following is an example of a max-pooling method.



Max pool with 2*2 filter and stride 2

*Above is the ReLU function.

Figure 8: Max-pooling process

As previously stated, Pooling methods are used independently for each layer for the depth D of the feature map. As a result, the depth of the feature map after pooling will also be D depth.

2.4 Training of CNN (Gradient Descent Procedure)

The goal of CNN training is to find accurate weights. We will use the chain rule to compute the partial derivative, followed by gradient descent to recalculate the weight. Backpropagation is the training algorithm. There are three steps to this:

1. Forward propagation toward the output of every neural cell a_j.

2. Backpropagation towards loss of every neural cell σ_j. which is called sensitivity in some articles. It is the loss function-related derivation of Net, which is

$$\sigma_j = \frac{\partial E_d}{\partial net_j}$$

3. Calculate the connection weight of every neural cell Wij (the weight from neural cell i to the neural cell j). $\frac{\partial E_d}{\partial W_{ji}} = a_i \sigma_j$ a_i stands for output of cell i

We were finally recalculating the weight using the gradient descent procedure. The next section will discuss the internal mathematical concept and reasoning of the convolution and pooling layers.

For the convolution layer, repeat step 2 of above. The question is how to transfer the loss term to the previous layer. Then referring to the third step for calculating the slope of weight w of the filter.

Layer 1

$\sigma_{1,1}$	$\sigma_{1,2}$	$\sigma_{1,3}$
$\sigma_{2,1}$	$\sigma_{2,2}$	$\sigma_{2,3}$
$\sigma_{3,1}$	$\sigma_{3,2}$	$\sigma_{3,3}$

Input 3*3

$w_{1,1}$	$w_{1,2}$
$w_{2,1}$	$w_{2,2}$

Filter 2*2

$\delta_{1,1}$	$\delta_{1,2}$
$\delta_{2,1}$	$\delta_{2,2}$

Feature map 2*2

Figure 9: Typical settings of input, filter and feature map

In the feature above, we assign a number to every element. $\delta_{i,j}^{l-1}$ They are placed to represent the L-1-layer i[th] row and j[th] column loss term $w_{m,n}$ and the weight of the m[th] row and nth column. W_b As the bias. $a_{i,j}^{l-1}$ Represent the output of L-1 layers i[th] row and j[th] column. $ne_{i,j}^{l-1}$ It is the weighted sum of every neural cell f^{l-1} and the activation function for layer l-1. Their relation equation is written as:

$$net^l = conv(W^l, a^{l-1}) + w_b \text{ ----------(5)}$$
$$a_{i,j}^{l-1} = f^{l-1}(net_{i,j}^{l-1}) \text{ ----------(6)}$$

*Conv stands for the convolution step.

*Y=ax + b is the linear relationship

In this example, assuming every δ^l of L has been calculated. What we should do is to calculate the loss term δ^{l-1} of the L-1 layer. According to the chain rule of derivation steps:

$$\delta_{i,j}^{l-1} = \frac{\partial E_d}{\partial net_{i,j}^{l-1}} = \frac{\partial E_d}{\partial a_{i,j}^{l-1}} \frac{\partial a_{i,j}^{l-1}}{\partial net_{i,j}^{l-1}} \text{ ----------(7)}$$

Firstly, we need to do a calculation for the first term and see if we can come up with some general idea

Approach I: Calculate $\frac{\partial E_d}{\partial a_{1,1}^{l-1}}$, $a_{1,1}^{l-1}$ are only connected to the calculation of $net_{1,1}^l$

$$net_{1,1}^{j} = w_{1,1}a_{1,1}^{l-1} + w_{1,2}a_{1,2}^{l-1} + w_{2,1}a_{2,1}^{l-1} + w_{2,2}a_{2,2}^{l-1} + w_b$$

$$\frac{\partial E_d}{\partial a_{1,1}^{l-1}} = \frac{\partial E_d}{\partial net_{1,1}^{l}} \frac{\partial net_{1,1}^{l}}{\partial a_{1,1}^{l-1}} = \delta_{1,1}^{l} w_{1,1} \text{----------(8)}$$

So that

Approach II: Prove the relationship between $\frac{\partial E_d}{\partial \alpha_{1,2}^{l-1}}$, $\alpha_{1,2}^{l-1}$ and $net_{1,1}^{l}$, $net_{1,2}^{l}$

We have the (8) and based on (8). we find that

$$net_{1,2}^{j} = w_{1,1}a_{1,2}^{l-1} + w_{1,2}a_{1,3}^{l-1} + w_{2,1}a_{2,2}^{l-1} + w_{2,2}a_{2,3}^{l-1} + w_b$$

So that we can find the derivative of two factors like what we have done in (8)

$$\frac{\partial E_d}{\partial a_{1,2}^{l-1}} = \frac{\partial E_d}{\partial net_{1,1}^{l}} \frac{\partial net_{1,1}^{l}}{\partial a_{1,2}^{l-1}} + \frac{\partial E_d}{\partial net_{1,2}^{l}} \frac{\partial net_{1,2}^{l}}{\partial a_{1,2}^{l-1}} = \delta_{1,1}^{l} w_{1,2} + \delta_{1,2}^{l} w_{1,1} \text{----------(9)}$$

Concluded from the above two approaches, we can quickly expand the equation to multiple layers so that based on (8) and (9) equations, we find that in order to

Calculate $\frac{\partial E_d}{\partial a^{l-1}}$ which is equivalent to doing zero padding in L^{th} layers. Then rotating 180 degrees of the filter matrix to do cross-correlation manipulation. Because the convolution is equivalent to the 180-degree rotation of the filter matrix. Then we can form the general equation of training equation like below, and second equation is the expanding form:

$$\frac{\partial E_d}{\partial a_l} = \delta^l * W^l$$

$$\frac{\partial E_d}{\partial a_l} = \sum_m \sum_n w^l_{m.n} \delta^l_{i+m,i+n}$$

For equation (6)

$$a^{l-1}_{i,j} = f(net^{l-1}_{i,j}) \text{----------} (6)$$

And we do the derivative of the left-hand value

$$\frac{\partial a^{l-1}_{i,j}}{\partial net^{l-1}_{i,j}} = f^{'}(net^{l-1}_{i,j}) \text{----------} (10)$$

$$\delta^{l-1}_{i,j} = \frac{\partial E_d}{\partial net^{l-1}_{i,j}} = \frac{\partial E_d}{\partial a^{l-1}_{i,j}} \frac{\partial a^{l-1}_{i,j}}{\partial net^{l-1}_{i,j}} = \sum_m \sum_n w^l_{m,n} \delta^l_{i+m,j+n} f^{'}(net^{l-1}_{i,j}) \text{----------} (11)$$

Then change the equation (11) into the convolution form as

$$\delta^{l-1} = \delta^l * w^l \otimes f^{'}(net^{l-1}) \text{----------} (12)$$

Which stands for the element-wise product, and all parameters in equation (12) are matrix. Equation 1-12 illustrates the simplest situations (stride =1, deep size=1, number of filters=1)

An essential step of training CNN is the pooling layer. They do not have a learning parameter, consequently, in typical CNN training, whether it is max pooling or mean pooling. The function of the pooling layer is just transferring to the loss of the previous layer.

1): Max pooling loss transferring

Like the picture below, assuming that L-1 is a 4*4 matrix, and the pooling layer is a 2*2 layer and strip equal to 2. After the max-pooling, the L^{th} layer will be a 2*2 matrix. Assuming that σ value has been calculated. The task should be calculating the $(L-1)^{th}$ σ value.

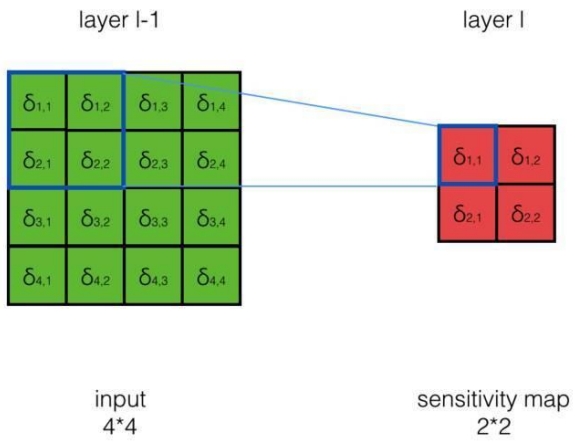


Figure 10: Max Pooling Layer

Assuming $net_{1,1}^{t}$ as the weighted input for (L-1)th layer, according to the definition of the max-pooling method.

$$net_{1,1}^{l} = \max(net_{1,1}^{l-1}, net_{1,2}^{l-1}, net_{2,1}^{l-1}, net_{2,2}^{l-1})$$

Only the max number in the matrix can affect $net_{1,1}^{t}$. So, we assume that $net_{1,1}^{l-1}$ as the max value in this matrix. So, the above formula can be written as:

$$net_{1,1}^{l} = net_{1,1}^{l-1}$$

we can quickly obtain some partial derivatives below:

$$\frac{\partial net_{1,1}^{l}}{\partial net_{1,1}^{l-1}} = 1 \text{----------} (13)$$

$$\frac{\partial net_{1,1}^{l}}{\partial net_{1,2}^{l-1}} = 0 \text{----------} (14)$$

$$\frac{\partial net_{1,1}^{l}}{\partial net_{2,1}^{l-1}} = 0 \text{----------} (15)$$

$$\frac{\partial net_{1,1}^{l}}{\partial net_{2,2}^{l-1}} = 0 \text{----------} (16)$$

So that as the convolution layer in last part:

$$\delta_{1,1}^{l-1} = \frac{\partial E_d}{\partial net_{1,1}^{l-1}} = \frac{\partial E_d}{\partial net_{1,1}^{l}} \frac{\partial net_{1,1}^{l}}{\partial net_{1,1}^{l-1}} = \delta_{1,1}^{l} \text{----------(17)}$$

And

$$\delta_{1,2}^{l-1} = \frac{\partial E_d}{\partial net_{1,1}^{l-1}} = \frac{\partial E_d}{\partial net_{1,1}^{l}} \frac{\partial net_{1,1}^{l}}{\partial net_{1,2}^{l-1}} = 0 \text{----------(18)}$$

$$\delta_{2,1}^{l-1} = \frac{\partial E_d}{\partial net_{1,1}^{l-1}} = \frac{\partial E_d}{\partial net_{1,1}^{l}} \frac{\partial net_{1,1}^{l}}{\partial net_{2,1}^{l-1}} = 0 \text{----------(19)}$$

$$\delta_{2,2}^{l-1} = \frac{\partial E_d}{\partial net_{1,1}^{l-1}} = \frac{\partial E_d}{\partial net_{1,1}^{l}} \frac{\partial net_{1,1}^{l}}{\partial net_{2,2}^{l-1}} = 0 \text{----------(20)}$$

Now we have a general idea about the loss transferring. For the max-pooling layer, the loss of the next layer will transfer to the corresponding cell of the previous layer, but for other cells, the loss will be 0.

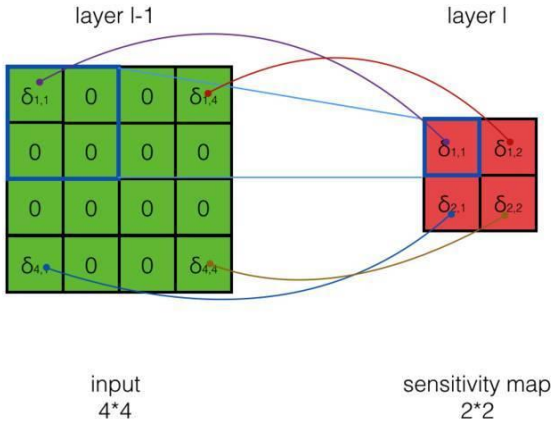


Figure 11: loss function-max pooling

2): Mean pooling loss transferring

Mean pooling is an alternative way when improving the performance of the whole convolution neural network. Here below is the calculation step.

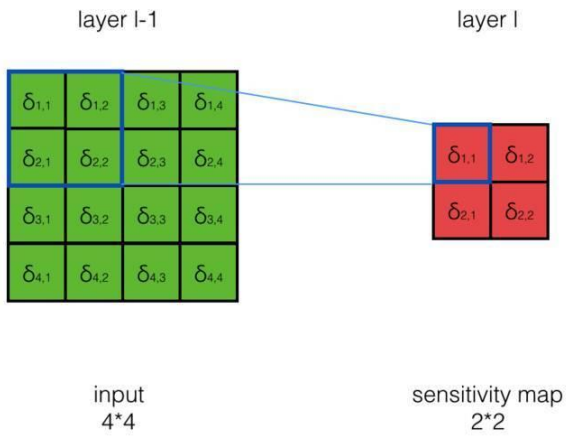


Figure 12: Mean pooling layer

As the picture showed above. According to the definition of the mean pooling method. The equation is as below:

$$net_{1,1}^{l} = \frac{1}{4}(net_{1,1}^{l-1} + net_{1,2}^{l-1} + net_{2,1}^{l-1} + net_{2,2}^{l-1})$$

Then we can quickly obtain some partial derivatives below:

$$\frac{\partial net_{1,1}^{l}}{\partial net_{1,1}^{l-1}} = \frac{1}{4} \text{----------} (21)$$

$$\frac{\partial net_{1,1}^{l}}{\partial net_{1,2}^{l-1}} = \frac{1}{4} \text{----------} (22)$$

$$\frac{\partial net_{1,1}^{l}}{\partial net_{2,1}^{l-1}} = \frac{1}{4} \text{----------} (23)$$

$$\frac{\partial net_{1,1}^{l}}{\partial net_{2,2}^{l-1}} = \frac{1}{4} \text{----------} (24)$$

So that as the max-pooling layer calculation in the last part:

$$\delta_{1,1}^{l-1} = \frac{\partial E_d}{\partial net_{1,1}^{l-1}} = \frac{\partial E_d}{\partial net_{1,1}^{l}} \frac{\partial net_{1,1}^{l}}{\partial net_{1,1}^{l-1}} = \frac{1}{4}\delta_{1,1}^{l} \text{----------} (25)$$

And

$$\delta_{1,2}^{l-1} = \frac{\partial E_d}{\partial net_{1,1}^{l-1}} = \frac{\partial E_d}{\partial net_{1,1}^{l}} \frac{\partial net_{1,1}^{l}}{\partial net_{1,2}^{l-1}} = \frac{1}{4}\delta_{1,1}^{l} \text{----------} (21)$$

$$\delta_{2,1}^{l-1} = \frac{\partial E_d}{\partial net_{1,1}^{l-1}} = \frac{\partial E_d}{\partial net_{1,1}^{l}} \frac{\partial net_{1,1}^{l}}{\partial net_{2,1}^{l-1}} = \frac{1}{4}\delta_{1,1}^{l} \text{ ----------(22)}$$

$$\delta_{2,2}^{l-1} = \frac{\partial E_d}{\partial net_{1,1}^{l-1}} = \frac{\partial E_d}{\partial net_{1,1}^{l}} \frac{\partial net_{1,1}^{l}}{\partial net_{2,2}^{l-1}} = \frac{1}{4}\delta_{1,1}^{l} \text{ ----------(23)}$$

After the mean pooling method, the general idea is that loss from the next layer will be averagely distributed evenly to every cell in previous layers, as the picture showed below:

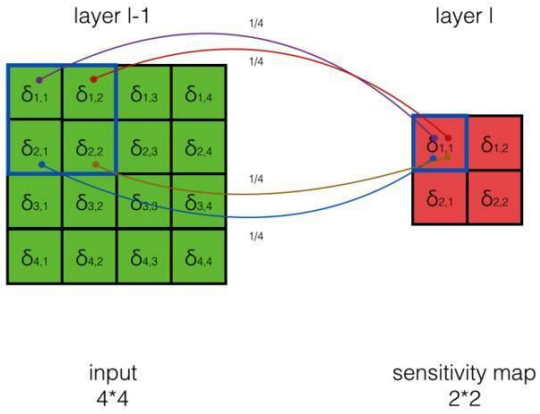


Figure 13: loss function-mean pooling

This is the fundamental part of constructing a convolution neural network, the next chapter will move the flow on the actual application on method settings.

CHAPTER 3

RESEARCH METHOD (CNN)

3.1 Data Preparation

I obtained both Electronic Health Record (EHR) and breast mammograms from the radiology department at Truman Medical Center (TMC) to conduct this study. The UMKC IRB approved this project (ORA#21-010) and the TMC research committee. Citrix, a cloud server, is where the data is stored. The sample contains 122 benign tumor patients and 112 malignant tumor patients over three years, excluding null and empty data. Each patient has 1-4 images from the most recent time points after 2017. There are 371 benign images and 331 malignant images in total.

The data set includes 692 images from 234 patients, including 122 benign and 112 malignant female patients over 40 from 2017 to 2020. There are 192 benign longitudinal records and 313 longitudinal malignant records in the EHR section. It contains some basic information about the patient and some risk factors for breast cancer, such as family history, breast characteristics, hormone factors and breast disorder conditions. Here below is the summary table:

	Breast Disorder	Hormore factor	Birth Control	Hysterectomy	MenoPausal status	Child	Family history	Self history
Yes(1)	93	84	112	240	26	675	330	133
No(0)	562	583	547	424	638	4	359	535
Total	655	667	659	664	664	679	689	668

Table 2: EHR Data Summary

3.2 Overview of CNN Architecture

The first chapter discussed the differences between traditional machine learning approaches and deep learning approaches before deciding on the deep learning method. Looking at CNN's architecture, it is an end-to-end structure, which means that the feature extraction process is completely automated. Without human intervention, features are extracted from training data. After each training, the algorithm can learn like a human and gain experience (Glasmachers,2017). Modelers can manipulate and adjust the parameters based on the algorithm in the first part to fit testing needs through medical images such as mammograms and other types of medical images (Kim, 2019).

Typically, a completed structure of CNN consists of two major steps: feature learning part and classification part. Chapter 2 discussed the convolution layer, pooling layer, and activation layer (ReLu). They are all included in the feature learning part. They can repeat several times to adjust for the objective of the research. There are usually three parts in the classification part: flatten layer, fully connected layer, and soft-max layer. Soft-max layer is considered the final classification layer. The following sections will explain how these layers work and the internal functions of each layer in the whole CNN structure.

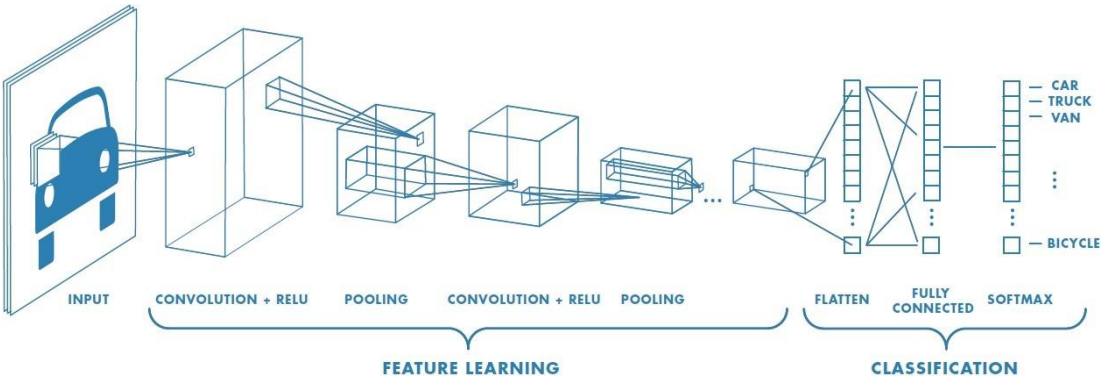


Figure 14: Feature Learning and classification part of CNN

3.3 Layer Details

As mentioned in Chapter 2, the first operation of CNN is the convolution layer. Therefore, the image data will first pass through a convolution layer, followed by a series of filters in the following layers. In Chapter 2, we discussed the convolution process and zero paddling. The function of zero paddling is to unify image sizes so that they all have the same dimension matrix. The equation is as follows:

$$y[i,j]=\sum_{m}\sum_{n} filter(m,n)image(i-m,j-n) \text{----------}(24)$$

In this study, the pre-trained CNN was made up of 32 filters. After the convolution operation is completed, the output of this layer is passed to the ReLU (Rectified Linear Unit) function. As mentioned in Chapter 2, the ReLU function will set all negative values to 0.

Early papers found that training a deep network with ReLU tended to converge much faster and more reliably than training a deep network with sigmoid activation (Brownlee, 2019). Recall that a ReLU is defined as $h=max\ (0,\ a)$ where $a=Wx+b$. One significant advantage is that the gradient is less likely to vanish. It occurs when a >0. The gradient has a constant value in this regime. In contrast, as the absolute value of x increases, the gradient of the sigmoid decreases. Reclus' constant

gradient results in faster learning, which accelerates learning. The other advantage of ReLUs is their sparsity. Sparsity arises when a≤0—the greater the number of such units in a layer, the sparser the resulting representation. The function of Max-pooling layers can reduce data volume to reduce the workload of CNN in further calculation and training.

After the data passes through the repeated primary layers in the last part (convolution layers, ReLU activation layers, and pooling layers), the data is analyzed, it must be reshaped from high dimension to lower dimension format. The flattened layer is used to obtain data in the form of a single dimension. The fully connected layer follows the flattened layer. It is made up of several neurons that are all fully connected to the previous flattened layer. Assuming a numerical example, *X* comes from the flattened layer and can be referred to as output *X*. It is possible to give it to the neuron. For example, we have one neuron that is connected to the previous layer, and we have weights for these links that include bias. After passing through the fully connected layer, find output y. The equation will be solved using the ReLU activation (*w* stands for the weight)

$$y = Max(0, wx + bias) \text{----------} (25)$$

The classifier is the model's final layer. There are two types of cancer in this study: benign and malignant. Assume there are two neurons in the final output layer. One is benign, while the other is cancerous. Then, using the above equation, obtain the output y. Then, using the Soft-Max classifier equation as shown below:

$$SoftMax(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_i}} \text{----------} (26)$$

It can be used to normalize the result range starting at [0.1]. Then, the researcher will determine the threshold for justifying the binary conditions (usually be 0.5).

3.4 CNN Model Training

The training set will comprise 60% of the total data, with the remaining 40% serving as the test set. Random extractions from the entire data set will be performed for each training process to select different training and test data, but their proportions remain constant. The proportion is derived from trial-and-error experience (Dobbin,2011).

Normalization, loss function, back-propagation, and other techniques are used in CNN model training, as discussed in Chapter 2. In this section, I will review the training procedure in this research.

1): Data normalization, Cross entropy, loss function and back-propagation

Normalization is critical because it makes training more stable and efficient. The normalization procedure involves changing the pixel values because the input file is an image of a different size. In this research, the training images are in greyscale format, which means that the pixel values range from 0 to 255. If the pixels have a similar normal distribution, the scaling process implies that the pixel values should be changed to be in the range of 0 to 1, which could be accomplished by multiplying each pixel value by the ratio of 1/255. The re-scaled distribution will be in the 0 to 1 range. The mean is then moved to the origin, and the variance is reduced to one. The equations are as follow:

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \sigma^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu)^2 \qquad x_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \xi}}$$

After completing equations (26), two probabilities for each class were calculated, and the labels for the classes were assigned (benign=0 and malignant=1). The loss will result from this manipulation. To calculate the error and determine how this instance produced much error. The cross-entropy formula is as follow:

$$CrossEntrophy(CE) = -\sum_{c} L\log(softMax(x_i)) - (1-L)\log(1-softMax(x_i))$$

2): Back-propagation & forward pass process.

The weights and bias of CNN will be set to a random value before we begin the training process. Then, it will be processed as described and the loss calculated. It is referred to as a forward pass. The main idea behind training is to minimize loss by updating the weights and biases of the convolution neural network. Take a look at the diagram below.

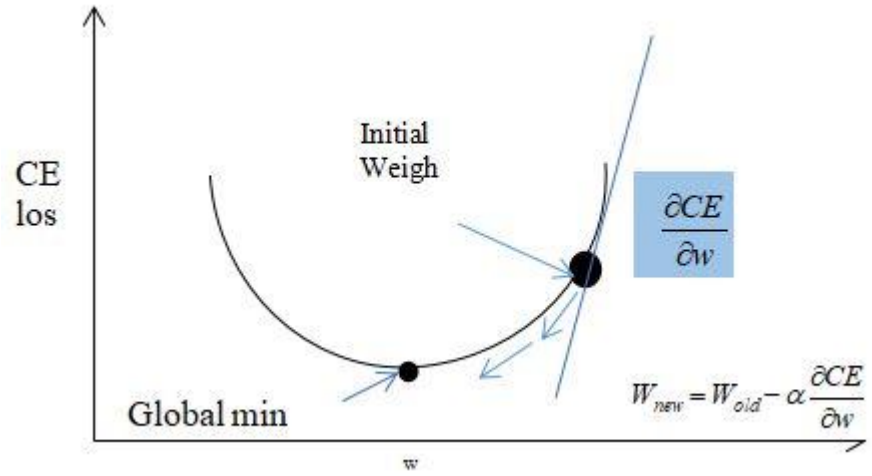


Figure 15: Derivative of training process in loss function optimization

We noted the starting weight. Now, we can see from the feature that in order to minimize the loss. As the above arrow of feature indicates in feature 15. The derivative of cross-entropy loss with respect to w will be used. Essentially, the derivative is the slope of the line that passes through this point. The derivative is positive, and the new value of w equals the old value minus the learning rate

multiplied by the derivative. The learning rate is how much we are adjusting the weights of our network with respect the loss gradient. The lower the value, the slower we travel along the downward slope shown in figure.15. The alpha (learning rate) used to be 0.001 in the literature. This technique is known as gradient descent, and it is used to update the parameters of CNN while ensuring global minimum loss (Ruder, 2016).

After that, the whole process was connected. Suppose we are interested in measuring the relationship between variables with respect to the cross-entropy loss. According to $\boldsymbol{y_1=Wx+b}$. $\boldsymbol{w}$ here will affect $\boldsymbol{y_1}$. In return, the $\boldsymbol{y_1}$ will affect the loss to calculate the derivative of loss with respect to $\boldsymbol{w}$. We can use the chain rule, equal to the derivative of $\boldsymbol{y_1}$ with respect to $\boldsymbol{w}$ times the derivative of output of the softmax classifier equation.

$$CrossEntrophy(CE) = -\sum_{c} L\log(out_1) - (1-L)\log(1-out_1)$$

$$SoftMax(x_i) = out_1 = e^{x_i} \Big/ \sum_{j=1}^{n} e^{x_i}$$

3.5 Model Evaluation

After several steps in prior training and calculation, the training loss will be stable and converge to certain value which means the training is complete. We already have a trained neural network that can perform prediction. The model's functionality should be tested through evaluation. Assume there are five benign and five malignant cases being processed by the neural network to determine the prediction result. The confusion matrix is used to calculate the statistical error. True positive is the first metric in the confusion matrix. The number of correctly predicted cases is referred to as a true positive. The second metric to consider is false positives. It will calculate the

number of incorrectly classified cases as abnormal. The third metric is the false negative measure, which counts the number of cases incorrectly predicted as normal but abnormal.

Assuming we have the confusion matrix below:

	Predicted Abnormal	Predicted Normal
Actual Abnormal	3(True Positive **TP**)	2(False Negative **FN**)
Actual Normal	1(False Positive **FP**)	4(True Negative **TN**)

Table 3: Typical Confusion Matrix

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} = 0.7 = 70\%$$

$$precision = \frac{TP}{TP + FP} = 0.75 = 75\% \qquad recall = \frac{TP}{TP + FN} = 0.6 = 60\%$$

$$f1score = 2 * \frac{precision * recall}{precisio + recall} = 0.67 = 67\%$$

The percentage of correctly predicted cases among the total number of cases used in the test can be used to calculate accuracy. For example, several cases were correctly predicted seven out of ten in this example, so the accuracy is 70%.

Another metric is precision, which quantifies the impact of false-positive cases on actual positive cases. It is simply the type II error rate multiplied by itself. The F-score is a measure of how accurate a test is. It is derived from the test's precision and recall. The F-score is a method of combining the model's precision and recall, and it is defined as the harmonic mean of the model's precision and recall. It is a measure of a model's accuracy on a data set. It is used to assess binary classification systems that categorize examples as 'benign' or 'malignant.'

ROC Curve relates to accuracy rate, called Receiver Operating Characteristic. The relationship between the ROC and AUC is as below:

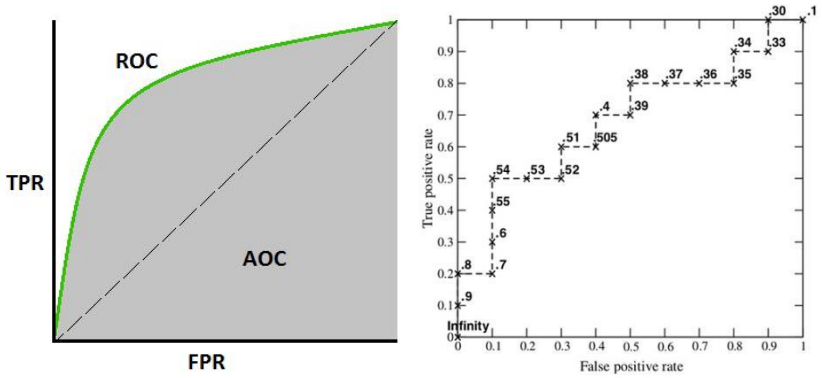


Figure 16: ROC Curve Definition

The smooth curve is drawn from various points along the curve. The x-axis represents the false positive rate, while the y-axis represents the valid positive rate. It describes all situations of each image/data, which means that the larger the valid positive rate, the larger the area should be. The ROC curve depicts the actual trend of the AUC rate and the model's stretching power.

3.6 Proposed Method to Account for Longitudinal Images Data Files

3.6.1 Literature Review of Stacking

The most critical step in this study is to account for temporal fluctuations in the input image. While the preceding discussion concentrated on single image classification, the purpose of this study is to evaluate the images collected through two to three years of time period. Because this is a patient-level study, the image data must be pre-processed. The Stacking RGB method is a method which has the potential to account for the time component in the collected image data. When stacking the single images in time order. It is possible to concentrate on the characteristics from the image that networks are learning (Zdolsek,2021). Thus, the discriminative image areas employed by neural networks to recognize features may be

seen by using stacking method (Zdolsek,2021). The function of the stacking method is to identify the hidden feature in multiple mammograms. With this technique, researcher can identify the hidden characteristic of the feature to make sure it contains key classification features.

Time series data are collected in radiology tests such as dynamic contrast-enhanced CT/MRI or dynamic radio isotope (RI) and positron emission tomography (PET). Previous research classified liver tumors using 2D-CNN on CT image sets acquired during three phases (non-enhanced CT, enhanced CT in arterial and delayed phases) (Yasaka,2018). The research employed triphasic CT scans as 2D image with three color channels, which correspond to the RGB color channels in computer vision, for 2D-CNN. The research demonstrated that 2D convolution neural networks trained on triphasic CT image outperformed those trained on biphasic or monophasic CT images. This demonstrates that RGB image has the potential to be applied to account for time component while doing deep learning-based classification.

The process of obtaining a sequence of colors of light is the foundation of our study. The normal pattern that allows human eyes to distinguish between colors on a screen is called the “color sequence”. The ability of the human eye to distinguish colors is based on the varying sensitivity of different cells in the retina to light of different wavelengths. Humans are more sensitive on the blue light over other two types of Primary color lights (Judd,1975). The other colors sensor is shown below:

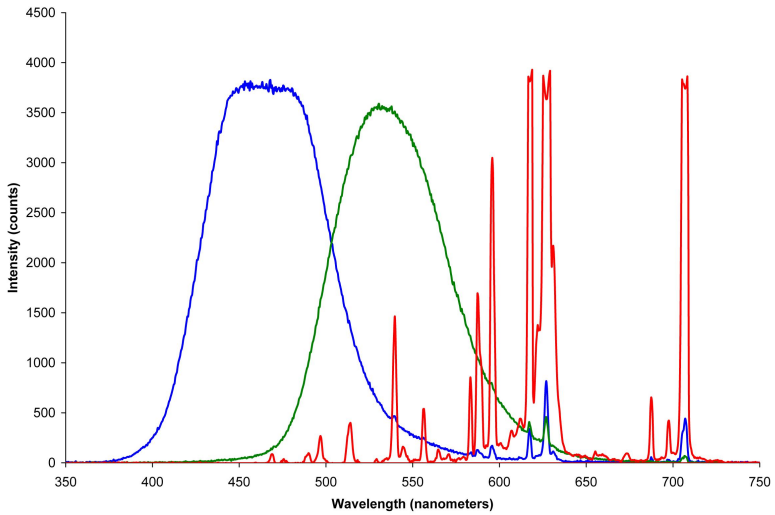


Figure 17: Wavelength and receiving order of three different light

In this research, we are stacking three images in chronological sequence by using a number of different image processing programs, such as Photoshop (to crop) and ImageJ (to stack). The most recent mammography will be placed on top, which, according to the illustration, will be the blue layer. Other medical images of each individual sufferer will be arranged chronologically at a later time. The patient's three most recent mammograms and the electronic health record that corresponds to them make up the archived document for the patient. The mammograms will be stacked from the top down, and the dye that is applied to them will go from blue to green to red. This is done because blue light is the most receptive of the visible light spectrum.

RGB stacking is a relatively better method for time related neural network-based classification. A method that utilizes Recurrent Neural Networks (RNN) for the analysis of skeletal data and 3D convolutions for the analysis of RGB films was suggested by Zhao (Zhao,2012). Because the authors included the time dimension in the convolution stack, it became possible to extract spatial-temporal patterns from the video.

The previous paragraphs demonstrate two benefits of the RGB stacking technique. Firstly, since RGB images have three channels, they hold more information

than greyscale images. Convolution may retain more details than a greyscale image. Second, the stacking method can account for time-varying medical images, allowing for the construction of patient-level information in addition to image-level information. This is why I'm using the stacking approach to maintain both functions and construct an idiosyncratic database for each participant in this study. Following the data pre-processing, the sample database of most participant comprises the EHR and the post-processed RGB image. When we are selecting participants for this study. We are selecting the patient who has the most comprehensive database of mammograms as well as electronic health records (EHR).

Human color space is defined by the set of all possible tristimulus values (Hirakawa, 2005). Therefore, in this study, I manually stack a single mammography in the R-G-B eye receiving sequence from bottom to top to imitate eye activity. Briefly describe how do you make this data pre-processing here.

3.6.2 Proposed RGB-Stacking Method

The RGB-Stacking method is a technique that can combine all color channels. This method is the application of multi-channel color combinations. Usually, researchers divide the channels to identify the shape features and expand the sample size. However, in this research, the objective of this manipulation is to amplify the time-varying features. In other words, it makes the significant features more detectable. This method is also compatible with the CNN structure. The limitation of this method is that some greyscale images are not suitable for merging color channels. The merging result may have image noise.

It is necessary to deal with time-varying images, especially for time-sensitive medical image objects like brain MRI and lung transformation after COVID-19. There is some existing way to deal with time-varying images in the literature. Huang

firstly treated the continuous image as video and trained this framework which finds applications ranging from model-based video coding and facial expression analysis to face recognition address for the facial changing in years(Huang,1997). However, this method depends on the snapshot of the face in continuous order, which is not suitable for regular examination images in breast mammograms. Ito's research work proposed a new way to put every feature to manipulate the time-varying in bacteria(Garcia-Perez, 2021). Garcia-Perez's methodology is a good reference for detecting mammograms and is a typical method dealing with the longitudinal image. I will fix and try this method in this research. However, the drawback is that this method should manually extract features and order, which causes bias on classification.

After considering the existing classification method, the kernel consideration for selecting method should be low bias, great AUC rate after training, and relatively quick training speed in the application.

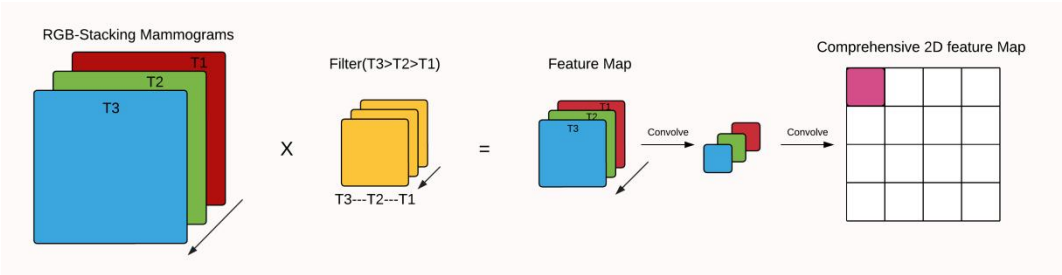


Figure 18: Proposed Method Convolution steps

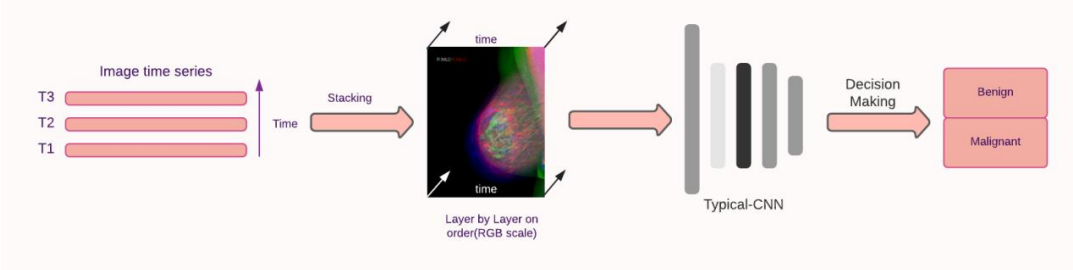


Figure 19: Proposed Method Convolution structures

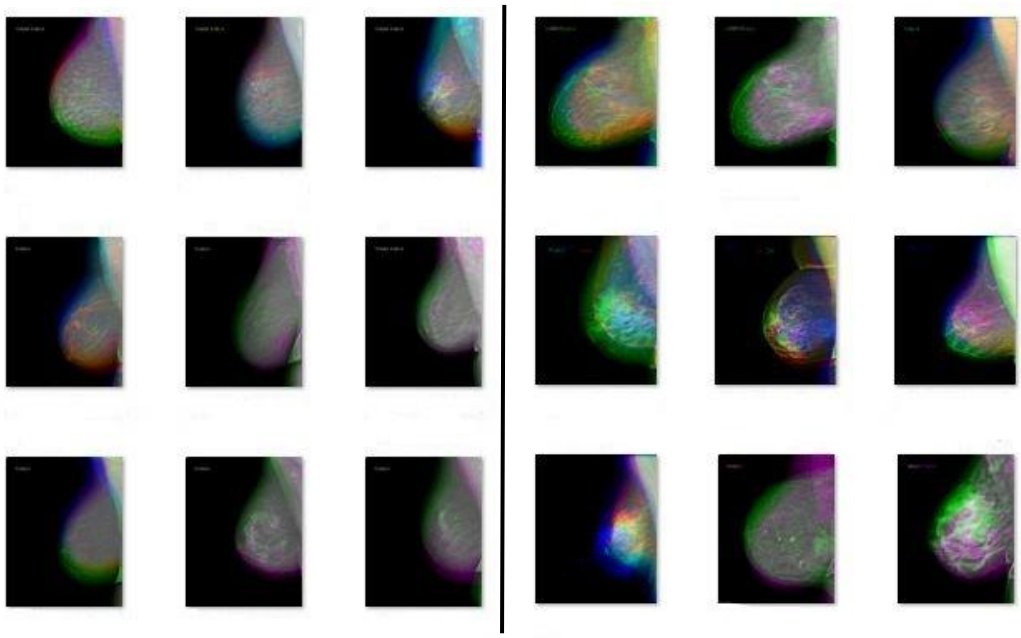
The primary process of the first steps of convolution is depicted above. We can see that the image at the top will be given the most significant weight when the

feature map results are calculated. Furthermore, this method's result matrix has the same structure as the greyscale image, making them comparable in all steps. As a result, it incorporates all features from the patients' mammogram history.

This general idea came from the literature (Park, 2010). In this article, the CNN is intended to be capable of prioritizing RGB images for nanosatellite applications. Furthermore, several pieces are spliced using the layer combination, which increases the CNN classifying efficiency when the weather condition is not good. Consequently, converting the greyscale to an RGB image will improve the model, which is verified in this paper.

Since the cancer progression is closely related to time, it is crucial to consider the time effect in the analysis. The stacking method is preferred for different color channel image analyses (Cai, 2017). I marked and dyed the color in their image sequence using RGB tunnel using image editing software (Photoshop, GIMP, and ImageJ). These tools can quickly stack 1-3 images and place the most recent image on top.

The images below show that benign tumors are frequently lighter in color than malignant tumors. This characteristic is caused by naturally metastasized abnormal tissue that has taken over normal tissue in the breast. However, when we combine the images in chronological order, the color of malignant features is more noticeable than that of benign tumors.



Benign **Malignant**

Figure 20: Comparison between Benign and Malignant tumors

In order to make the RGB stacking image go through the existing CNN smoothly. Under this scenario, we changed the input data shape from 1 channel to 3 channels and changed the training and testing data shape from 1 channel to 3 channels. As a result, every layer will adjust for three channels stuff. Such that image data shape will shift from 1 channel to 3 channels.

The typical-adjusted CNN methods mentioned earlier represent the CNN-SoftMax in previous paragraphs because such structure has been shown to be the best underlying model when dealing with single-image problems. The brighter color represents the most recent layer of the image, the most recent mammogram of patients in this study. Because CNN stimulates the existing neural network in the human brain, the top feature will be prioritized for extraction by the neural network since humans are always drawn to the brighter colors.

CHAPTER 4

RESULT

This research project used the convolution neural network (CNN) as the foundational method to classify patients' cancer status. The criteria for evaluating model performance is the accuracy rate, F-score, precision, and recall rate mentioned in chapter 3. It is known that model with higher accuracy index can perform well in the classification procedure. The objective is to improve the AUC rate and other indexes above to select the best deep learning model. The goal is to propose an underlying deep learning model that will account for medical images and electronic health records.

During this whole study, I used the personal computer (GTX 1650) for training and testing. Based on this strategy, I trained the neural network for 25 epochs to ensure convergence and stability. Images were chosen at random throughout each training session, ensuring a random approach and minimizing selection bias. In the simulation procedure, the neural network is trained 30 times, yielding 30 sets of evaluation indices. Each evaluation index showed in the chart is the thirty-times average index.

4.1 Basic CNN & Microsoft ResNet Network-The Naive Model

The basic CNN approach is a classic method for classifying the actual properties of images. The network structure contains several convolution layers, activation layer, and max-pooling layer, as stated in chapter 3. It can be performed as an essential method and preliminary approach. The research approaches can be modified to adjust to input matrix and functions based on this.

This method is the primary method without considering the longitudinal data type and assigned all images into two groups in proper proportion: test groups and training groups. As a result, they occupied a 60/40 proportion in this research.

The Microsoft research team developed the ResNet model in 2016(K. He,2016). The developers of ResNet increased the depth of the network up to 152 layers. It is an advanced research approach compared to the primary CNN-softmax method, which contains more deep layers when dealing with high dimension images and matrices. This model also avoids regular overfitting problems.

Both basic CNN and the Microsoft ResNet model can be performed as a bottom layer for future research, including longitudinal and combined research. All results from breast cancer data are summarized in the chart below using five indexes:

	Accuracy	Precision	Recall	F1 Score	ROC AUC
CNN-Softmax	0.679	0.788	0.501	0.612	0.775
CNN-ResNet	0.472	0.22	0.47	0.301	0.481

Table 4: Underlying Structure Comparison

Based on the result above, CNN-Softmax will act as the basic structure for the single image classification based on research data. This testing sample contains all collected images. However, in Chapter 3, I mentioned that single image classification has its limitation. The classification label is the imaging unit. The patient is the unit practically in this research. Longitudinal images are considered instead of a single patient unit. So, based on this, I conducted another research on patients' levels using the underlying CNN-softmax method.

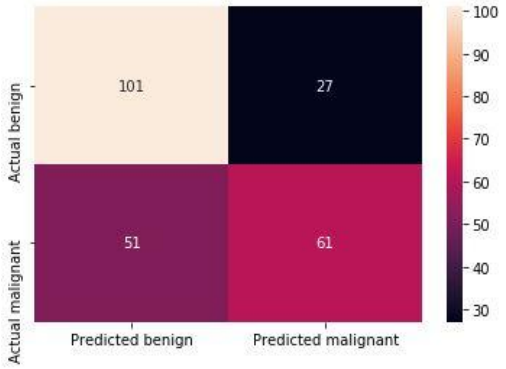
CNN's kernel function is to convolve and extract the significant feature, resulting in a combined one-dimensional vector. Suppose the result is derived from well-assigned weights of various time points images. The accuracy may be superior to a basic CNN single-image approach. We can combine the images based on their time

sequences. The most recent one will be at the top, implying that it will be given the most weight when the convolution process is carried out (Cai, 2017).

4.2 Method Comparison

In order to choose the most suitable one, evaluation using five indexes is used for comparison. Four methods have been chosen for image-only classification models. To select the best model, I have created criteria for comparison. These criteria contain five indexes. The accuracy is the overall general accuracy of this classification model. Precision is the ratio between the true positives and all the positives which is the criteria for evaluation of positive rate. Another assessment metric is the recall rate, which is a measure of how well our model detects true positives. As a result, recall rate informs us how many patients we accurately recognized as having breast cancer out of all those who have it. ROC accuracy is special index different from the previous four indexes. It represents the stretching power of the classification when changing the classification threshold.

1) CNN-Softmax



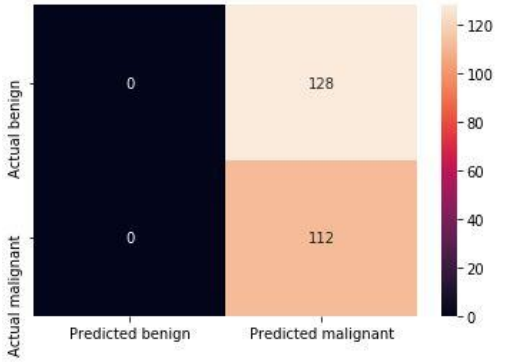
	precision	recall	f1-score	support
benign	0.66	0.79	0.72	128
malignant	0.69	0.54	0.61	112
accuracy			0.68	240
macro avg	0.68	0.67	0.67	240
weighted avg	0.68	0.68	0.67	240

Figure 21: CNN-SoftMax evaluation result

The above graphs and evaluating result shows that for samples like a single image, the accuracy rate is 66% when diagnosing benign tumor and 69% when recognizing malignant tumors. The overall accuracy is 68%. Then we can come up with other indexes. The overall precision rate is 0.788, the recall rate is 0.501 and the F-1 score is 0.612. These data shows that CNN-Softmax is a decent model among all models used for classification. The ROC accuracy for this model is 0.775, which is also an important criterion for model selection.

2): CNN-ResNet

According to the data obtained. This method is not a proper method for a limited number of the sample after the trial, and the sensitivity is very low, especially for the benign prediction, as the graph is shown below:



	precision	recall	f1-score	support
benign	0.00	0.00	0.00	128
malignant	0.47	1.00	0.64	112
accuracy			0.47	240
macro avg	0.23	0.50	0.32	240
weighted avg	0.22	0.47	0.30	240

Figure 22: ResNet evaluation result

According to the figure above, this model is not good selection for collected data. It lacks sensitivity on my breast cancer data. Although ResNet is the classic method serving for single image, it is not suitable for my study based on the index.

Since the two methods have their limitations, we are trying to mix the layers of CNN and put the patients as a unit to predict the benign or malignant features. First, however, we analyze the patient-level images using the longitudinal method mentioned in the previous paragraph.

Unlike many different activation layers in the previous neural network, the CNN is designed to avoid the over fitting problem. Since for patient-level. There are only 231 patients in the sample. Therefore, the test set will be small and have high sensitivity on the augmentation of original data.

3) : CNN-Fixed-Patient Level-longitudinal

	precision	recall	f1-score	support
benign	0.73	0.68	0.71	98
malignant	0.58	0.63	0.60	67
accuracy			0.66	165
macro avg	0.65	0.66	0.65	165
weighted avg	0.67	0.66	0.66	165

Figure 23: CNN-longitudinal evaluation result

After 30 trials, we get this result in patient level, according to the result above. However, the overall accuracy is 67%, which is little bit smaller than the simple CNN model. According to the partial accuracy, this model is more sensitive to predicting benign tumors than image-related techniques. Benign patients often have externally fewer mammograms than malignant ones. The precision and recall rate is not good as single image model. Also, this model's stretching power is also decreasing.

Consequently, the longitudinal method is not the best model when accounting for time variable.

4): New proposed method-Time related stacking method.

Below is the preliminary result of this method which shows some improvement over the previous method, especially for detecting malignant tumors in typical-adjusted CNN structures.

	precision	recall	f1-score	support
benign	0.68	0.93	0.79	57
malignant	0.86	0.49	0.62	49
accuracy			0.73	106
macro avg	0.77	0.71	0.70	106
weighted avg	0.76	0.73	0.71	106

Figure 24: CNN-Stacking evaluation result

This method's performance is better for the malignant patients due to the high precision rate, which means that when stacking the features on different time points, the mammograms' essential features will be amplified and be efficiently recognized by the neural network and make the right decision. The predicted malignant cases have the most significant chance to be the actual malignant cases. Although recall rate of malignant is not as good as we think. 49% of malignant cases in this sample are predicted right. The combined accuracy, stretching power, and other index looks better than other approaches. Consequently, It is necessary to add the Electronic Health Record as a reference to improve the whole model quality in this research.

Below is the complete review for each method.

	Accuracy	Precision	Recall	F1 Score	ROC AUC
CNN-Softmax	0.679	0.788	0.501	0.612	0.775
CNN-ResNet	0.47	0.22	0.47	0.3	0.481
CNN-longitudinal	0.671	0.67	0.66	0.678	0.614
CNN-Stacking	**0.726**	**0.776**	**0.715**	**0.705**	**0.842**

Table 5: All Four Method Evaluation indexes Comparison

The Stacking-RGB method has excellent potential for detecting the cancer malignant tumor associated with time. Therefore, this method will be chosen as the basic structure for combining longitudinal EHR to form the improved fusion model in the next chapter. In next part, this research will address on longitudinal EHR and the combination process and structure.

CHAPTER 5

COMBIMNATION OF IMAGING DATA AND EHR IN CNN

Previous chapters covered the machine learning technique for image categorization. The result demonstrates that the Stacking-RGB method is the best suitable technique for classifying the longitudinally variable time image compared to the image-only model. Therefore, we must use the the electronic health records (EHR) to develop a new algorithm capable of enhancing the performance of the image-only model. Electronic Health Records are an essential diagnostic resource. They should be treated as a supplement to the model employed in chapter 4 in order to enhance the performance of the image-only model. This chapter will explore the precise property, function, characteristic, and practicability of the EHR and its combination.

5.1 EHR Data Summary

Electronic health data plays a vital role in diagnosing patients. Therefore, EHR has been increasingly deployed within health care organizations to improve the safety and quality of care since the 2000s (Poissant,2005). The EHR comes from two resources in this research: patients' surveys and physician notes. Family history comes from regular patient visiting surveys. However, self report data has a selection bias because most patients are unsure about their health condition and relatives' cancer history (Verheij,2018).

Many components of the emerging EHR connect patients, practices, clinics, imaging centers, hospitals, health insurance, laboratories, and pharmacies in a confidential, secure, and standardized manner over the Internet(Ambinder, 2005). As a result, the change will improve the quality of modern medical care.

EHR is rapidly increasing its influence on medical centers and the health-related industry. It allows access to evidence-based tools providers can use to make

decisions about a patient's care on the cloud. A vital feature of an EHR solution is easy to access to digital information. During a patient visit, independent physicians can enter and review information about the patient. The physician can then seamlessly collaborate with other providers to provide coordinated care. In addition, EHR eliminates the need to request records from specialty providers, laboratories, or healthcare facilities.

Hospitals have begun to use electronic health record (EHR) systems in recent years. This digitalization of enormous volumes of medical data presents a once-in-a-lifetime potential for deep learning to enhance healthcare by predicting diagnoses, lowering healthcare expenditures, and modeling the temporal link between medical occurrences, among other things. The intrinsic longitudinal and multi-modal character of EHR data, on the other hand, provides a level of complexity that is missing from standard academic datasets like ImageNet and WMT, which are frequently used to construct deep learning models (Xu,2021). the pre-processing work is important to normalize both the type and format of collected EHR data.

The chart below is the actual variable name corresponding to the detailed description of certain variables. The collected electronic health data consists of 8 variables associated with the time variable. Time variables are ranked as 1,2,3 to show each patient's time points of visiting times. The variable name and binary definition are listed below:

Variable Name	Definition for 0 (No)	Definition for 1 (Yes)
Breastdisorder	No previous breast disorder in the survey	Have previous breast disorder in survey
hormone factor	Have not received previous hormone treatment	Have received previous hormone treatment
Birthcontrol	No previous contraception	Have used birth control pills
Hysterectomy	No previous hysterectomy surgery	Have a history of hysterectomy surgery
Menopausalstatus	Menopause	Period
Child	No giving birth to a Child	Have one or more children
Familyhistory	No family history of breast cancer	Have a family history of breast cancer
Selfhistory	No other gynecological type cancer	history of gynecological types of cancer

Table 6: Definition of Variables in Electronic Health Record

Each variable has at least three active visits. If the patient has more than three visits, take the latest three visits as the reference to clean the data to co-responding to the previous three stacking images. The cleaning task is to match every record to every patient's image. Here below is the cleaning result for the EHR data:

Variable name	Time point 1	Time point 2	Time point 3	Tracking patients numbers (sequence by time points)
Breastdisorder	Y:37 (15.8%) N:197 (84.2%)	Y:32 (14.5%) N:188 (85.5%)	Y:24 (11.9%) N:177 (88.1%)	Time point 1:234 Time point 2:220 Time point 3:201
hormorefactor	Y:30 (14.7%) N:204 (85.3%)	Y:30 (13.5%) N:191 (86.5%)	Y:24 (11.3%) N:188 (88.7%)	Time point 1:234 Time point 2:221 Time point 3:212
Birthcontrol	Y:46 (19.6%) N:188 (80.3%)	Y:37 (16.9%) N:181 (83.1%)	Y:29 (14.0%) N:178 (86.0%)	Time point 1:234 Time point 2:218 Time point 3:207
Hysterectomy	Y:85 (36.3%) N:149 (63.6%)	Y:81 (36.0%) N:144 (64.0%)	Y:74 (36.1%) N:131 (63.9%)	Time point 1:234 Time point 2:225 Time point 3:205
Menopausalstatus	Y(period): 10 (4.2%) N(pause): 224 (95.8%)	Y(period): 8 (3.7%) N(pause): 210 (96.3%)	Y(period): 8 (3.8%) N(pause): 204 (96.2%)	Time point 1:234 Time point 2:218 Time point 3:212
Child	Y:232 (99.1%) N:2 (0.9%)	Y:222 (99.1%) N:2 (0.9%)	Y:221 (100%) N:0 (0%)	Time point 1:234 Time point 2:224 Time point 3:221
Familyhistory	Y:112 (47.86%) N:122 (52.14%)	Y:110 (47.82%) N:120 (52.17%)	Y:108 (48.0%) N:117 (52.0%)	Time point 1:234 Time point 2:230 Time point 3:225
Selfhistory	Y:48 (20.5%) N:186 (79.5%)	Y:44 (19.8%) N:178 (80.2%)	Y:41 (19.1%) N:174 (79.5%)	Time point 1:234 Time point 2:222 Time point 3:215

Table 7: Descriptive Statistics of EHR Variables

After sorting and dealing with the null value in the collected data, we have 165 patients who meet the criteria of at least three mammogram screening visits. It consists of 86 benign cases and 79 malignant cases. Furthermore, 40% of data will serve as a test set and 60% as a training set.

5.2 EHR Combination Methods

One hundred sixty five patients have three time points electronic records co-responding to the RGB(3 layers) images after data cleaning. The feature map extraction can represent the aggregate features, as the picture shows below. We aggregated features across the three visiting observation windows and then constructed a tensor representation on temporal values from all certified patients.(Zhao, J, 2019).

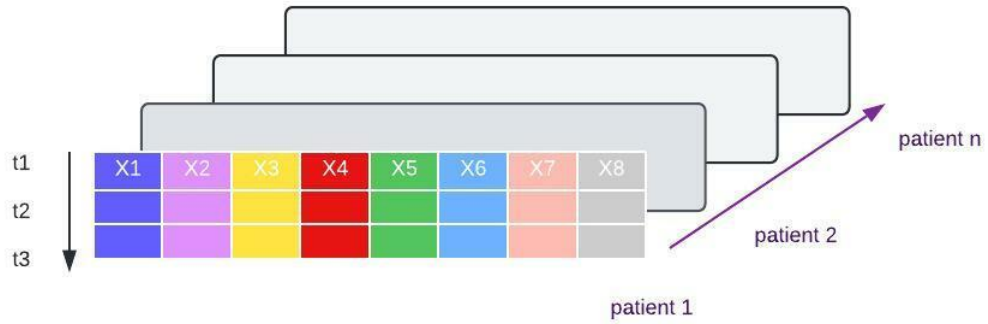


Figure 25: Temporal value fit in feature map matrix

The above figure shows the basic structure of the aggregating method. The x-axis is the feature type of each patient. The y-xis stands for the time sequence. This single feature map represents one patient's electronic health record in three visits, ready to combine with the image matrix. Z-axis will serve as the different patients. This matrix-like feature map is suitable and convenient for combining with an image matrix in a convolution neural network.

This study utilizes a matrix to summarize patients' personal longitudinal information. The X_1 to X_8 represent the 8 following factors mentioned. The dimension for the matrix is 3 multiply 8 (3x8). The reason to make the EHR as a matrix is to make them go through CNN quickly and keep the critical feature of each patient. In addition, this type of matrix is convenient for combination with image data. In the previous chapter, we knew that CNN layers(fully-connected layers) could adjust the matrix and transform them into 1 dimension vector, which is easy to classify into different group.

5.3 Combination with Longitudinal Image and EHR

Chapter 4 lays out the primary method for the image-only model in mammograms. Section 5.2 illustrates the feature of the electronic health record and how to combine this type of data into a convolution neural network. Image and electronic health records can be fused to improve the model's accuracy. This section will address the whole structure and evaluation of all three fusion models and select the best model for classifying.

Using multi-modality aims to extract and combine important information from various resources and then use this integrated feature to solve a given problem. As a result, the expected output will be richer in representation and performance than the individual modalities. Multi-modal data analysis is a practical solution for many fields of study especially for medical field.

5.3.1 Literature Review

Radiologists utilize data from different sources to determine to what degree a patient is at risk for breast cancer in breast imaging(Heidi,2012; Valerie,2006).They may have access to imaging results, demographic and clinical data (age, gender, and clinical indication), and information on risk factors for cancer (comorbidity and

family history)(Valerie,2006).A model that integrates and learns from these disparate sources of information may outperform a uni-modal (single-source) approach in terms of prediction performance.

Modern medicine's diagnostic tool is significantly reliant on the synthesis of information and data from a variety of sources; these sources include imaging pixel data, organized laboratory data, unstructured narrative data, and, in certain situations, audio or observational data. This is especially true in the field of medical image interpretation, where extensive clinical training and experience is often required to make diagnostic choices. The majority (87%) of radiologists who participated in the survey expressed the opinion that the clinical information of the patient had a significant impact on interpretation of data (Boonn, 2009). Consequently patient's clinical context and information are critical for the successful interpretation of imaging data in many disciplines.

In Chapter 1, I mentioned about the shortage and burden of radiologist. The amount of radiology imaging examinations is increasing in the modern digital era. To fulfill this increasing task requirement, radiologist may be required to interpret an image every 3–4 seconds over the course of an 8-hour workday, resulting in tiredness and an increased mistake rate(Leslie,2020). Therefore, deep learning is gaining traction in healthcare owing to the possibility of effective automated systems augmenting or offloading cognitive work from busy radiologist (Dean, 2015).

Convolutional neural networks (CNN) have been shown to be very good in image recognition and classification tasks and are often used to analyze medical imaging. The example includes diabetic retinopathy and chest X-rays (Gulshan,2016; Kallianos, 2019). However, the current application of CNN on medical imaging use pixel data with a single input modality without contextualizing other clinical

information in the model. This disadvantage has limitations on performance and accuracy of the CNN model. Accurate and meaningful medical imaging interpretation can be better ascertained with the use of an electronic health record (EHR). As is the case with human physicians, automated detection and classification systems that effectively integrate medical imaging data with clinical data from the EHR, such as patient demographics, past diagnoses, and laboratory findings, may result in more accurate and clinically useful models. For example, the authors of the paper (Akselrod, 2019) merged mammography data with clinical data in a two-stage process and saw comparable gains in AUC based on CNN baseline (from 0.88 to 0.91).

A fusion model is defined as multi-modal (characterized by several different modes and source of activity or occurrence) deep learning models which are capable of ingesting pixel data in addition to other data types. It has shown effectiveness in areas other than medicine, including autonomous driving and video categorization. For example, a multi-modal social media video classification pipeline that leverages both visual and textual features can increase classification accuracy to 88.0%, significantly higher than single modality neural networks such as Google's InceptionV3 that achieved 76.4% accuracy on the same task (Trzcinski,2018). Fusion model incorporates contextual information in addition to images, overcoming limitations of image-only models.

I covered the concept of fundamental convolution neural network structure which were employed in this study in prior chapters. The proposed CNN is consisted of three types of layers: convolution layers, pooling layers, and fully connected layers. The input image is passed through kernels or filters in the convolution layer to create various feature maps. The pooling layer compresses the size of each of the feature maps in order to minimize the number of weights. This is sometimes referred to as

downsampling or sub-sampling. There are several pooling strategies, including maximum pooling and average pooling. Following these layers, the fully connected layer is utilized to convert two-dimensional feature maps to a one-dimensional vector suitable for classification. The basic architecture of the CNN is shown in the below figure X. Following each convolution layer is a pooling layer. The last pooling layer's output is routed to a fully connected layer and a final output layer. (J. Cong,2014)

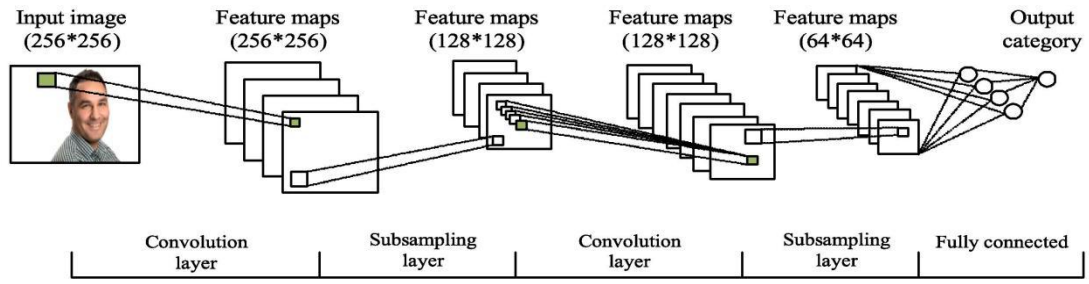


Figure 26: Advanced S Structure of CNN

Nguyen's study (Nguyen,2017) presented a unique end-to-end deep learning technique dubbed Deepr for extracting critical information from medical record and predicting abnormalities. To predict unexpected readmission after discharge, a convolution neural network is applied to a series of discrete components.

Consequently, fusion model shows potential and power to increase the important indexes over single information-based deep learning model. It contains more information and reference for classification. Based on the type and stage of fusion. The fusion method mainly consists of three types: early fusion, joint fusion and late fusion.

(1): Early Fusion

The assumption behind early data fusion is that numerous data sources are conditionally independent. Different modalities might include information that is

connected at a higher level, according to the article. So, each modality's output can be assumed to be separately processed in practical application.

There are two disadvantages in using early-stage data fusion method. Firstly, before fusion, a considerable quantity of data will be removed from the modalities in order to create a common ground. This is one of the method's primary drawbacks. The second drawback of early fusion is the need to synchronize the timestamps of the various modalities. Collecting data at a consistent sample rate is a typical method for overcoming this drawback. It began from the data collecting step. In this research, we are using corresponding values in both EHR and image data to avoid the risk.

(2): Joint Fusion

Joint fusion transforms input data into a higher level of features using many layers. Each layer executes linear and nonlinear functions that change the scale, skew, and swing of the input data, resulting in a new characteristic of the original input data. In a deep learning multi-modal setting, joint fusion is the fusing of distinct modalities' features into a single hidden layer so that the model may learn a joint feature of each modality. Features can be learned from different kinds of layers including 2D convolution and fully connected layer in this research.

Different modalities can be merged into a single common representation layer either simultaneously or progressively utilizing multiple modalities. Although it is possible to combine the features or weights of many modalities in a single layer, doing so may lead to model over fitting or prevent the network from learning the relationship between each modality. Our method for enhancing the performance of deep multi-modal fusion is to reduce the dimension of the data and normalize it before input. In contrast to early level fusion and late level fusion, joint fusion allows for the fusing of features at different depths and stag.

(3): Late Fusion

The prevalence of ensemble classifiers has affected the development of late data fusion. This methodology is substantially easier than the early fusion method, especially when the sampling rate, data dimension, and unit of measurement of the data sources are significantly different. Late fusion often results in improved performance since mistakes from many models may be addressed individually, meaning that errors are no longer associated with one another.

There are a variety of criteria that may be used to identify the best strategy to decide how to ultimately integrate each of the models that were individually trained. The Bayes rules, max-fusion, and average-fusion are examples of rules. When input data streams differ greatly in terms of size and sampling rate, late fusion provides a more straightforward and adaptable solution.

In order to solve a variety of research questions, these three fusion procedures are often used. When compared to the technique based on a single modality, the approach based on multiple modalities could include more features, which is beneficial for the further classification process. I will discuss the actual way of combining toward our research in next part.

5.3.2 Fusion Method Concept

In contrast to employing a single data modality, the fusion model refers to integrating data from various modalities to collect complementary and more comprehensive information for better performing deep learning models.

Advancements in deep learning techniques have the potential to contribute to healthcare, especially in sectors where medical imaging is used for diagnosis, prognosis, and treatment decisions. Current state-of-the-art deep learning models for radiology applications take just pixel-value information (image matrix) into account,

with no clinical context data. However, relevant and accurate non-imaging data based on the clinical history and laboratory data allow clinicians to interpret imaging results in the proper clinical context, resulting in increased diagnostic accuracy, insightful clinical decision making, and improved patient outcomes (Huang.S,2020).

Chapter 1 discussed how the deep learning algorithm simulates the human brain to make decisions. Based on that, the deep learning algorithm should incorporate clinical text data and health records as part of radiologists' decisions. In order to attain a similar aim using deep learning, medical imaging pixel-based models must be able to handle contextual data from electronic health records (EHR) and pixel data.

Through the use of deep learning, this study works to combine the fusion CNN method with EHR to improve the accuracy of early detection of breast cancer. In order to achieve this goal. Several fusion methods are experimented with to fit both the image and EHR data, accounting for time. In recent literature (Huang, 2020), the author has listed several fusion strategies using deep learning, including early fusion, joint fusion, and late fusion. Our research mainly depends on that setting. There are three techniques used for multi-modal data fusion (Lahat,2015) (Khalegh,2013).

Early fusion concatenates original or extracted features at the input level. Joint fusion is joining features after processing. Late fusion aggregates predictions at the decision level. Three basic methods have their advantages and disadvantages, which should be considered in the method evaluation process.

I have experimented with several methods of fusion type and tried to find a proper detailed way to fit the model with collected data. However, model fusion is different in specific scenarios, so I need to try several times to make sure the fusion

method can fit the collected data in this research and adjust the model parameters to achieve a better result.

5.4 Type and Structure of Fusion Methods

I: Early Fusion-Feature Fusion

Early fusion is a traditional way of fusing multiple data before conducting the analysis. It is to merge the data at the beginning of the learning process and let them go through the neural network, which refers to combining numerous input modalities into a single feature vector before feeding it into a single deep learning model for training. The basic structure is as shown in the figure below:

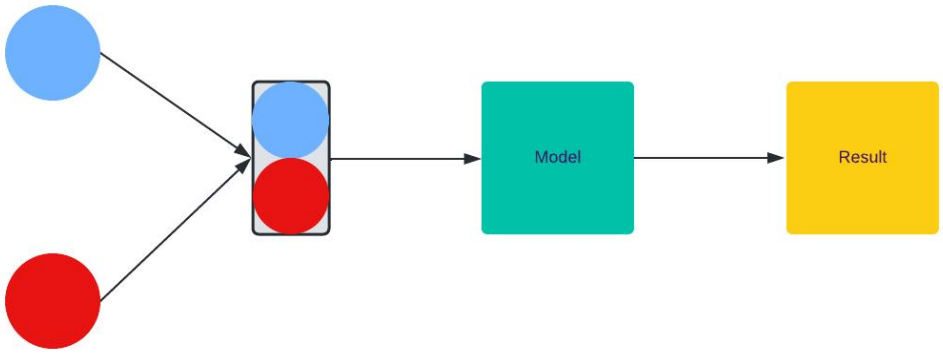


Figure 27: Early-Fusion Basic Structure

Early fusion is applicable on raw data or pre-processed data obtained from resources. In this approach, we first simplify the image data to a vital feature matrix, combine the EHR data in an early step, and then let the combined matrix go through the neural network. The whole structure graph is as below. In this feature, the pink box represents the fully connected layer. The n and m are the numbers of filters used in this process which can be adjusted to fit for the training requirement. Green box is regular ReLU layers to make extracted features non-linear. The blue box classifies the layer-softmax function layer to make the final classification result.

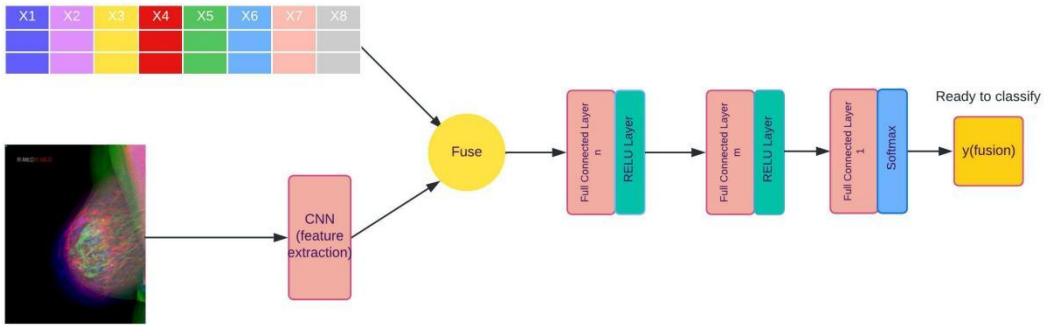


Figure 28: Early Fusion Detailed Structure

II: Joint Fusion- Learned feature Fusion

The joint fusion is based on the deep learning neural network. This strategy is the most adaptable, allowing for data fusion at difference phases of model development. It contains lots of possibilities through fusing in different stages. It is the technique of combining learned features from intermediate layers of neural networks as input to a final model with features from other modalities. In contrast to early fusion, the training loss is fed back to the feature extraction neural networks during training, resulting in improved feature representations for each training iteration. The basic structure of joint fusion is as shown in the feature below.

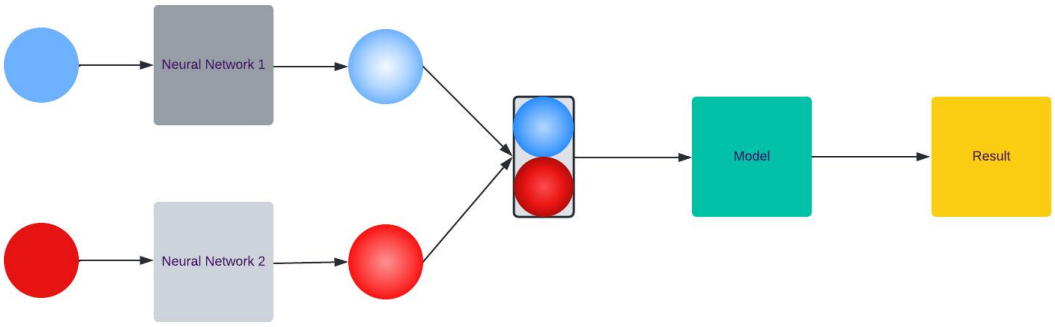


Figure 29: Joint-Fusion Basic Structure

Below is the structure of the Joint Fusion structure. Learned feature fusion learns features from the image and the electronic health record simultaneously, then concatenates learned feature vectors from each modality before learning from this combined vector to produce a final prediction. This approach is comprehensive and

can be the best theoretically since it contains more feature extraction content than the early fusion method and fewer calculation steps than the late fusion method.

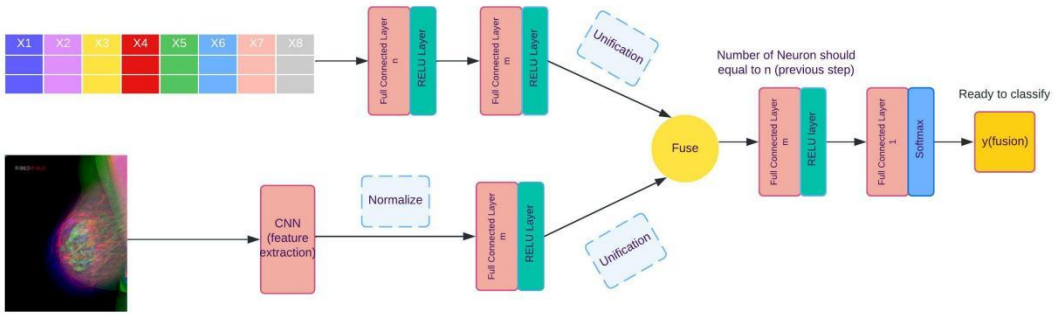


Figure 30: Joint Fusion Detailed Structure

III: Late fusion-probability fusion

Late fusion uses data sources independently followed by fusion at a decision-making stage. This method is easier than the early fusion method, especially when the data sources differ greatly in terms of sampling rate, data complexity, and unit of measurement. It frequently provides higher performance since mistakes from many models are dealt with individually, resulting in uncorrelated errors.

Late fusion is the process of combining predictions from several models to produce a final choice. Different modalities are utilized for independent training models. Furthermore, the ultimate choice is determined by combining the predictions of numerous models using an aggregation function. In this research, we only consider two modalities. The basic structure of late fusion is as the feature below:

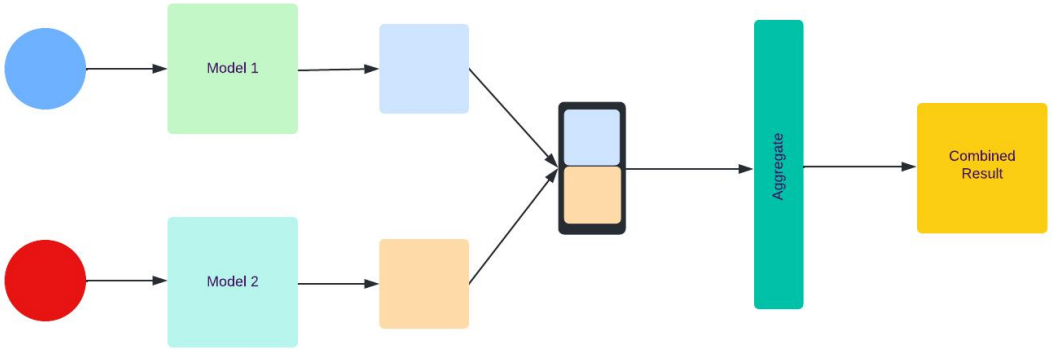


Figure 31: Late-Fusion Basic Structure

Below is the structure and setup of the late fusion (probability fusion) model. The neurons of the fully connected layer (green) are similar in co-responding steps. The blue box is the softmax function layer. As mentioned before, this function can return the number between 0 and 1. when setting the threshold. The model can easily classify the categories. The critical step is to merge the output of each result y(EHR) and y(image) because, after the softmax layer, the output of the two sides will be normalized to the same

dimension.

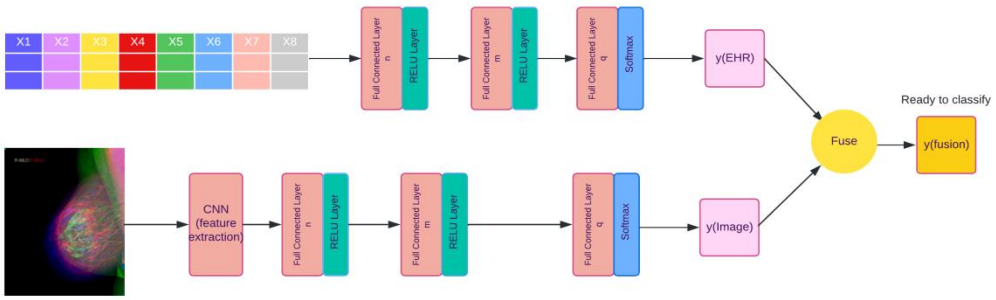


Figure 32: Late Fusion Detailed Structure

In the next section, I will concentrate on the evaluation method mentioned in the end section of chapter 4 to explore the best (accuracy) and the most efficient model (time-consuming in the training process).

5.5 Evaluation and Execution and Comparison of Model

I used the personal computer (GTX 1650) for training testing and simulation in this study. We separately analyze three techniques and 25 epochs of training to make them converge like Figure 15.In previous sections, I explained that the training set and testing set samples are selected at random to eliminate selection bias in sampling. We trained 30 times and produced 30 sets of indices. Therefore, the evaluation index is the average index of all 30 sets. It is the same setup as Chapter 4 illustrates.

The evaluation criteria contain five indices, including four model performances and one model stretching index (ROC). After a long training, test, and adjustment in all 4 models (including the image-only model and three multi-modal fusing models). I selected the best model based on accuracy rate and other indices (training times bigger than 20). Since the sample size is small, the experiment must be run many times to avoid bias. More runs will avoid the contingency.

The figures below illustrate the ROC curve associate with all four methods.Model performance on breast cancer prediction for all four model architectures evaluated is compared. The Image-Only model was trained only on mammography pictures, whilst the other fusion models were trained on both images and clinical characteristics. Fusion models outperformed single element model greatly. AUC is an abbreviation for the area under the receiver operating characteristic (ROC) curve.

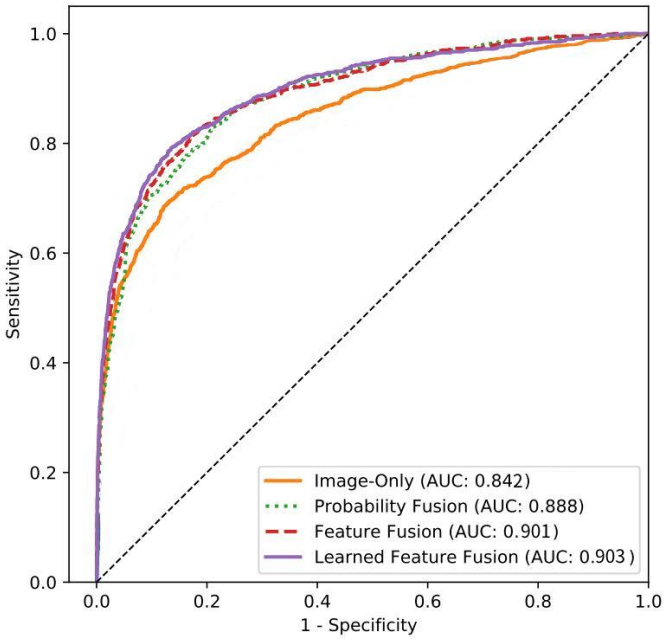


Figure 33: ROC Curve for four methods

In Chapter 4, the five indexes measure the performance and power of model. So I use these five indexes to evaluate both the image-only model and fusion model.

Model	AUC	Precision	Recall	F-1 Score	ROC AUC	Running time per epoch
Image-only (RGB Stacking)	0.726	0.857	0.715	0.705	0.842	78s
Latefusion-probability fusion(EHR+Image)	0.865	0.821	0.752	0.785	0.888	104s
Earlyfusion-Feature Fusion(EHR+Image)	0.872	0.806	0.771	0.788	0.901	84s
JointFusion-Learnedfeature Fusion(EHR+Image)	0.876	0.871	0.744	0.802	0.903	91s

Table 8 : Model Comparison Based on Indexes

In the above result, we find that the accuracy rate improves dramatically with the electronic health record combined compared to the RGB stacking method in chapter 4. The joint fusion method is the best and most efficient method among all fusion methods.

Joint fusion (learned feature fusion) served as the best model among all choices. According to the neural network structure, this training process is expected to be the best. However, based on the result above, every fusion method has the potential to be the best one because their accuracy rate is quite close.

5.6 Limitations of The Fusion Method

Even though the fusion approach improves the accuracy rate of the prediction of early diagnosis, there is still a limitation to this line of research. The three significant problems are longitudinal effects, small sample size, and over-fitting problems.

The amount of experiment data is relatively small, which causes a random effect on the method selection, meaning that the accuracy rate shows disturbance in the training process. We can expand the sample size to see if the model works well in future study. In this research, we keep only 70% of the collected data. Therefore, the utilization of collected data is relatively low. The data collection procedure should be optimized to expand the valid sample size.

Longitudinal studies take a long time and are often quite expensive. As a result, these studies frequently have fewer subjects, making it difficult to extrapolate the findings to a larger population. One hospital has a limited number of patients visiting. If we want to expand the scale of such research, cooperation with several local medical centers is necessary, which costs much money.

Within the context of this combination study approach, the possibility of over-fitting is more prevalent than the issue with a single image. The fusion model allows for a multitude of non-linear layers to simultaneously learn and output complicated correlations. Because of the intricate relations, an overfitting problem will arise, which has the potential to lessen the effectiveness of the training set. This question can be appropriately resolved by the dropout function built into each training stage (N.Srivastava,2014). It has the potential to both enhance the sparsity and decrease the amount of time needed for calculations. In addition, the dropout function has the ability to govern and maintain the essential neurons that are employed in decision-making. Including a dropout layer after each ReLU layer is an important step in optimizing the model.

CHAPTER 6

CONCLUSION AND FUTURE WORK

This study presents a strategy that may be used to predict early-stage breast cancer. Mammograms are the initial step in detecting cancer; thus, recognizing the exact kind of cancer in mammograms will save time and money. In the present classification procedure, SVM (support vector machine) and other machine learning approaches are commonly used. However, in this study, we use the deep learning technique since it is the most effective at classifying picture collections and is also the simplest to manipulate. In addition, the deep learning neural network provides a number of benefits, including a reasonably quick processing time and a great rate of accuracy. We construct the classifier using a convolution neural network (CNN) since its structure is very adaptable and simple to combine with other elements.

The second chapter is separated into two sections. The first section describes the usual structure of a convolution neural network. The second section primarily illustrates the precise mathematical equations deducing in the convolution stage, the activation layer (ReLU), and the pooling layer. Mathematical concepts are crucial to this study since the whole training framework and assessment procedure are based on several mathematical equations.

The third chapter introduces the configuration of convolution neural networks. It depicts the framework of the training and testing procedure, as well as the manner of training evaluation. It describes the parameters, loss functions, and evaluation criteria that will be utilized in subsequent chapters. In contrast to Chapter 2, Chapter 3 focuses more on the configurations of specific CNN structures and computations for this study.

The objective of Chapter 4 is to determine the optimal strategy for accounting for longitudinal medical images. Several techniques are compared using Chapter 3's assessment criteria. Finally, we choose RGB stacking to ensure that all layers at various time points are processed concurrently by the neural network. The stacking strategy enables CNN to scan over layers in chronological sequence. In addition, we may manually adjust the settings to address additional time-dependent problems. Compared to other procedures, the stacking method is the most effective of all potential strategies.

The fifth chapter covers electronic health records (EHR) and the fusion method. When people visit a medical institution, EHR is the initial source of information. It includes a variety of time-sensitive data, such as patient surveys and physician notes. It may be turned into a feature map while preserving the essential variables, making it acceptable and ready for fusion with the image matrix. When combined with EHR, the fusion technique is more precise than image-only models. According to current research (Huang, 2020), we use three kinds of fusion methods: early fusion, joint fusion, and late fusion, after a thorough evaluation of accuracy and training duration. The optimum method of fusion is joint fusion.

There are still some limitations in this research. The main limitation can be included in three aspects. Firstly, this research is a small sample analysis that focuses on the method proposing. This model needs more training data to enhance its function and performance. Secondly,the time and cooperation issues among medical centers are hard to unify the data type in different medical center, However, the training process need much data from same categories in different centers. Thirdly, this research model contains over fitting problems. The third problem can be partially

solved by adding dropout layers but it is also depend on the objective of certain convolution layer.

Future work will focus on developing a web page platform to simultaneously process and categorize server-based data and do analysis. This model's performance can be enhanced by storing training data on the server for future reference. All data is kept on a cloud server, which improves sample size and reduces the cost of data collection. The challenge for this is the privacy of data when storing the patients' data on the cloud.

The primary objective of this study is to offer a fusion strategy that takes longitudinal image data and EHR information into consideration. Depending on the kind of cancer tumor, the procedure may be modified. To increase the accuracy of early diagnosis, however, more simulations and the development of neural networks must be conducted on a greater number and variety of tumor images.

REFERENCES

Acuff, S., Whittle, B., Taylor, R., Green, C., Findeiss, L., Leschak, S., Stephens, C., & Osborne, D. (2017). Comparison of interventional radiologist and the nuclear medicine Technologist radiation exposure when using TheraSph ere and SIR-Spheres Y90 microspheres for liver cancer therapy. *Journal of Nuclear Medicine, 58(supplement 1)*, 802.

Ahmadi, M., Vakili, S., Langlois, J. M., & Gross, W. (2018). Power reduction in CNN pooling layers with a preliminary partial computation strategy. *2018 16th IEEE International New Circuits and Systems Conference (NEWCAS)*. https://doi.org/10.1109/newcas.2018.8585433

Akselrod-Ballin, A., Chorev, M., Shoshan, Y., Spiro, A., Hazan, A., Melamed, R., Barkan, E., Herzel, E., Naor, S., Karavani, E., Koren, G., Goldschmidt, Y., Shalev, V., Rosen-Zvi, M., & Guindy, M. (2019). Predicting breast cancer by applying deep learning to linked health records and mammograms. *Radiology, 292*(2), 331–342. https://doi.org/10.1148/radiol.2019182622

Ambinder E. P.(2005). Electronic health records. *Journal of Oncology Practice, 1*(2), 57–63. https://doi.org/10.1200/JOP.2005.1.2.57

Brady A. P. (2017). Error and discrepancy in radiology: inevitable or avoidable?. *Insights Into Imaging, 8(1)*, 171–182. https://doi.org/10.1007/s13244-016-0534-1

Breast cancer - stages. *American Society of Clinical Oncology*. (2022, May 24). Retrieved July 14, 2022, from https://www.cancer.net/cancer-types/breast-cancer/stages

Breast cancer facts & figures 2019-2020 - American Cancer Society. (n.d.). Retrieved February 14, 2022, from https://www.cancer.org/content/dam/cancer-

org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2019-2020.pdf

Blackwood, M. A., & Weber, B. L. (1998). BRCA1 and BRCA2: From molecular genetics to clinical medicine. *Journal of Clinical Oncology : Official Journal of The American Society of Clinical Oncology*, *16*(5), 1969–1977. https://doi.org/10.1200/JCO.1998.16.5.1969

Bohr, A., & Memarzadeh, K. (2020). The rise of artificial intelligence in healthcare applications. *Artificial Intelligence in Healthcare*, (pp. 25–60) . https://doi.org/10.1016/B978-0-12-818438-7.00002-2

Boonn, W. W., & Langlotz, C. P. (2009). Radiologist use of and perceived need for patient data access. *Journal of Digital Imaging, 22*(4), 357–362. https://doi.org/10.1007/s10278-008-9115-2

BW, S., & CP, W. (n.d.). *World cancer report 2014*. IARC Publications Website. Retrieved October 14, 2022, from https://publications.iarc.fr/Non-Series-Publications/World-Cancer-Reports/World-Cancer-Report-2014

Cai, Z.(2017). Performance evaluation of deep feature learning for RGB-D image/video classification. *Information Sciences, 385*. pp. 266-283.

Carroll T. J. (2003). Trends in on-call workload in an academic medical center radiology department 1998-20021. *Academic Radiology, 10(11)*, 1312–1320. https://doi.org/10.1016/s1076-6332(03)00381-7

Center for Devices and Radiological Health. (n.d.). *MQSA National Statistics*. U.S. Food and Drug Administration. Retrieved April 11, 2022, from https://www.fda.gov/radiation-emitting-products/mqsa-insights/mqsa-national-statistics

Centers for Disease Control and Prevention. (2022, September 26). *What is a mammogram?* Centers for Disease Control and Prevention. Retrieved October 7, 2022, from https://www.cdc.gov/cancer/breast/basic_info/mammograms.htm

Chubak, J., Boudreau, D. M., Fishman, P. A., & Elmore, J. G. (2010). Cost of breast-related care in the year following false positive screening mammograms. *Medical Care, 48(9)*, 815–820. https://doi.org/10.1097/MLR.0b013e3181e57918

Cong, J., & Xiao, B. (2014). Minimizing computation in Convolutional Neural Networks. *Artificial Neural Networks and Machine Learning – ICANN 2014*, 281–290. https://doi.org/10.1007/978-3-319-11179-7_36

Cruz-Roa, A. A., Arevalo Ovalle, J. E., Madabhushi, A., & González Osorio, F. A. (2013). A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. Medical image computing and computer-assisted intervention : *MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention, 16*(Pt 2), 403–410. https://doi.org/10.1007/978-3-642-40763-5_50

Dean, N. C., Jones, B. E., Jones, J. P., Ferraro, J. P., Post, H. B., Aronsky, D., Vines, C. G., Allen, T. L., & Haug, P. J. (2015). Impact of an electronic clinical decision support tool for emergency department patients with pneumonia. *Annals of Emergency Medicine, 66*(5), 511–520. https://doi.org/10.1016/j.annemergmed.2015.02.003

Deng, L. (2018, October 17). *Achievements and challenges of Deep Learning*. Microsoft Research. Retrieved May 22, 2022, from

https://www.microsoft.com/en-us/research/publication/achievements-and-challenges-of-deep-learning/

DeSantis, C. E., Ma, J., & Jemal, A. (2019). Trends in stage at diagnosis for young breast cancer patients in the United States. *Breast Cancer Research and Treatment*, *173*(3), 743–747. https://doi.org/10.1007/s10549-018-5042-1

Dobbin, K.K., Simon, R.M.(2011). Optimally splitting cases for training and testing high dimensional classifiers. *BMC Medical Genomics 4*,31. https://doi.org/10.1186/1755-8794-4-31

European Society of Radiology (ESR), & American College of Radiology (ACR) (2016). European Society of Radiology (ESR) and American College of Radiology (ACR) report of the 2015 global summit on radiological quality and safety. *Insights into imaging, 7*(4), 481–484. https://doi.org/10.1007/s13244-016-0493-6

Fowler, E. E., Sellers, T. A., Lu, B., & Heine, J. J. (2013). Breast Imaging Reporting and Data System (BI-RADS) breast composition descriptors: Automated measurement development for full field digital mammography. *Medical Physics, 40*(11), 113502. https://doi.org/10.1118/1.4824319

Garcia-Perez, C., Ito, K., Geijo, J., Feldbauer, R., Schreiber, N., & Zu Castell, W. (2021). Efficient detection of longitudinal bacteria fission using transfer learning in deep neural networks. *Frontiers in Microbiology, 12*, 645972. https://doi.org/10.3389/fmicb.2021.645972

Ginsburg, O., Yip, C. H., Brooks, A., Cabanes, A., Caleffi, M., Dunstan Yataco, J. A., Gyawali, B., McCormack, V., McLaughlin de Anderson, M., Mehrotra, R., Mohar, A., Murillo, R., Pace, L. E., Paskett, E. D., Romanoff, A., Rositch, A. F., Scheel, J. R., Schneidman, M., Unger-Saldaña, K., Vanderpuye, V., …

Anderson, B. O. (2020). Breast cancer early detection: A phased approach to implementation. *Cancer, 126 Suppl 10* (Suppl 10), 2379–2393. https://doi.org/10.1002/cncr.32887

Glasmachers, T. (2017). Limits of end-to-end learning. *Proceedings of the Ninth Asian Conference on Machine Learning, in Proceedings of Machine Learning Research, 77*: 17-32. Available from https://proceedings.mlr.press/v77/glasmachers17a.html.

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA, 316*(22), 2402–2410. https://doi.org/10.1001/jama.2016.17216

Han, S., Liang, X., Li, T., Yin, F. F., & Cai, J.(2021). Slice-stacking T2-weighted MRI for fast determination of internal target volume for liver tumor. *Quantitative Imaging in Medicine and Surgery, 11*(1), 32–42. https://doi.org/10.21037/qims-20-41

Hartono, A. P., Luhur, C. R., Indriyani, C. A., Wijaya, C. R., Qomariyah, N. N., & Purwita, A. A. (2021). Evaluating deep learning for CT scan COVID-19 automatic detection. *2021 International Conference on ICT for Smart Society (ICISS)*. https://doi.org/10.1109/iciss53185.2021.9533224

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2016.90

Hirakawa, K., and T.W. Parks.(2005). Chromatic adaptation and white-balance problem. *IEEE International Conference on Image Processing 2005, Genova, Italy: IEEE, III–984*. http://ieeexplore.ieee.org/document/1530559/ (May 10, 2022).

Huang, S.-C., Pareek, A., Seyyedi, S., Banerjee, I., & Lungren, M. P. (2020). Fusion of medical imaging and electronic health records using deep learning: A systematic review and implementation guidelines. *Npj Digital Medicine*, *3*(1). https://doi.org/10.1038/s41746-020-00341-z

Huang, T.S., R. Lopez, and A. Colmenarez.(1997). Time-varying image processing for 3D model-based video coding in time-varying Image processing and moving object recognition, 4, *Elsevier.*(pp.79–86). https://linkinghub.elsevier.com/retrieve/pii/B9780444823076500114.

Joe, B. N., & Sickles, E. A. (2014). The evolution of breast imaging: past to present. *Radiology*, *273*(2 Suppl), S23–S44. https://doi.org/10.1148/radiol.14141233

Judd, Deane B.; Wyszecki, Günter.(1975). Color in business, science and industry. *Wiley Series in Pure and Applied Optics (third ed.). New York: Wiley-Interscience.* p. 388. ISBN 978-0-471-45212-6.

Kallianos, K., Mongan, J., Antani, S., Henry, T., Taylor, A., Abuya, J., & Kohli, M. (2019). How far have we come? Artificial intelligence for chest radiograph interpretation. *Clinical Radiology, 74*(5), 338–345. https://doi.org/10.1016/j.crad.2018.12.015

Kerlikowske, K., Miglioretti, D. L., & Vachon, C. M. (2019). Discussions of dense Breasts, breast cancer risk, and screening choices in 2019. *JAMA, 322(1)*, 69–70. https://doi.org/10.1001/jama.2019.6247

Khaleghi, B., Khamis, A., Karray, F. O., & Razavi, S. N. (2013). Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*, *14*(1), 28–44. https://doi.org/10.1016/j.inffus.2011.08.001

Kim, M., Yun, J., Cho, Y., Shin, K., Jang, R., Bae, H. J., & Kim, N.(2019). Deep learning in medical imaging. *Neurospine, 16*(4), 657–668. https://doi.org/10.14245/ns.1938396.198

Kraus, M., Feuerriegel, S., & Oztekin, A. (2020). Deep learning in business analytics and operations research: models, applications and managerial implications. *European Journal of Operational Research*, *281*(3), 628–641. https://doi.org/10.1016/j.ejor.2019.09.018

Lahat, D., Adali, T., & Jutten, C. (2015). Multimodal data fusion: an overview of methods, challenges, and prospects. *Proceedings of the IEEE*, *103*(9), 1449–1477. https://doi.org/10.1109/jproc.2015.2460697

Leslie, A., Jones, A. J., & Goddard, P. R. (2000). The influence of clinical information on the reporting of CT by radiologists. *The British Journal of Radiology, 73*(874), 1052–1055. https://doi.org/10.1259/bjr.73.874.11271897

Liang, C., Bian, Z., Lyu, W., Zeng, D., & Ma, J. (2018). A deep features-based Radiomics model for breast lesion classification on FFDM. *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*. https://doi.org/10.1109/nssmic.2018.8824722

Liang, H., Sun, X., Sun, Y. et al.(2017). Text feature extraction based on deep learning: a review. *J Wireless Com Network 2017*, 211. https://doi.org/10.1186/s13638-017-0993-1

Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. *Zeitschriftfur Medizinische Physik, 29(2)*, 102–127. https://doi.org/10.1016/j.zemedi.2018.11.002

McCormack, V. A., & dos Santos Silva, I. (2006). Breast density and parenchymal patterns as markers of breast cancer risk: A meta-analysis. Cancer epidemiology, biomarkers & prevention : A publication of the American Association for Cancer Research, cosponsored by the *American Society of Preventive Oncology, 15*(6), 1159–1169. https://doi.org/10.1158/1055-9965.EPI-06-0034

Miglioretti, D. L., Walker, R., Weaver, D. L., Buist, D. S., Taplin, S. H., Carney, P. A., Rosenberg, R. D., Dignan, M. B., Zhang, Z. T., & White, E. (2011). Accuracy of screening mammography varies by week of menstrual cycle. *Radiology, 258(2)*, 372–379. https://doi.org/10.1148/radiol.10100974

Müller, D., & Kramer, F. (2021). MISCNN: A Framework for medical image segmentation with convolutional neural networks and Deep Learning. *BMC Medical Imaging*, *21*(1). https://doi.org/10.1186/s12880-020-00543-7

Namatevs, I., Sudars, K., & Polaka, I. (2019). Automatic data labeling by neural networks for the counting of objects in videos. *Procedia Computer Science*, *149*, 151–158. https://doi.org/10.1016/j.procs.2019.01.118

Nelson, H. D., Zakher, B., Cantor, A., Fu, R., Griffin, J., O'Meara, E. S., Buist, D. S., Kerlikowske, K., van Ravesteyn, N. T., Trentham-Dietz, A., Mandelblatt, J. S., & Miglioretti, D. L. (2012). Risk factors for breast cancer for women aged 40 to 49 years: a systematic review and meta-analysis. *Annals of Internal Medicine, 156*(9), 635–648. https://doi.org/10.7326/0003-4819-156-9-201205010-00006

Nguyen, P., Tran, T., Wickramasinghe, N., & Venkatesh, S. (2017). A convolutional net for medical records. *IEEE Journal of Biomedical and Health Informatics*, *21*(1), 22–30. https://doi.org/10.1109/jbhi.2016.2633963

Niraula, S., Biswanger, N., Hu, P., Lambert, P., & Decker, K. (2020). Incidence, characteristics, and outcomes of interval breast Cancers compared with screening-detected breast cancers. *JAMA Network Open*, *3*(9), e2018179. https://doi.org/10.1001/jamanetworkopen.2020.18179

O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., Riordan, D., & Walsh, J. (2019). Deep learning vs. Traditional computer vision. *Advances in Intelligent Systems and Computing*, 128–144. https://doi.org/10.1007/978-3-030-17795-9_10

Pakdemirli, E. (2019). Artificial intelligence in radiology: friend or foe? where are we now and where are we heading? *Acta Radiologica Open*, *8*(2), 205846011983022. https://doi.org/10.1177/2058460119830222

Park, J.H., Inamori, T., Hamaguchi, R., Otsuki, K., Kim, J.E., & Yamaoka, K.(2020). RGB image prioritization using convolutional neural network on a microprocessor for nanosatellites. *Remote Sensing., 12*, 3941.

Poissant, L., Pereira, J., Tamblyn, R., & Kawasumi, Y.(2005). The impact of electronic health records on time efficiency of physicians and nurses: a systematic review. Journal of the American Medical Informatics Association : *JAMIA, 12*(5), 505–516. https://doi.org/10.1197/jamia.M1700

Raghu, M., Zhang, C., Kleinberg, J.M., & Bengio, S.(2019). Transfusion: Understanding transfer learning with applications to medical imaging. *ArXiv*, abs/1902.07208.

Rezaeilouyeh, H., Mollahosseini, A., & Mahoor, M. H.(2016). Microscopic medical image classification framework via deep learning and shearlet transform. *Journal of Medical Imaging (Bellingham, Wash.), 3*(4), 044501. https://doi.org/10.1117/1.JMI.3.4.044501

Reznikov, N., Buss, D. J., Provencher, B., McKee, M. D., & Piché, N. (2020). Deep learning for 3D imaging and image analysis in biomineralization research. *Journal of Structural Biology, 212(1)*, 107598. https://doi.org/10.1016/j.jsb.2020.107598

Ruder.S. (2016). An overview of gradient descent optimization algorithms. *ArXiv*, abs/1609.04747.

Schmidhuber, J.(2015). Deep learning in neural networks: an overview. *Neural Networks. 61: 85–117.* arXiv:1404.7828. doi:10.1016/j.neunet.2014.09.003. PMID 25462637. S2CID 11715509.

Shen, L., Margolies, L.R., Rothstein, J.H.et al.(2019). Deep learning to improve breast cancer detection on screening mammography. *Sci Rep 9*, 12495. https://doi.org/10.1038/s41598-019-48995-4

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0197-0

Sorantin, E., Grasser, M. G., Hemmelmayr, A., Tschauner, S., Hrzic, F., Weiss, V., Lacekova, J., & Holzinger, A. (2022). The augmented radiologist: Artificial intelligence in the practice of radiology. *Pediatric Radiology, 52*(11), 2074–2086. https://doi.org/10.1007/s00247-021-05177-7

Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Jornal of Machine Learning Research., 15*, 1929-1958.

Tang, L., Wang, Y., Zhang, Y., Zhang, X. Y., Zeng, X. C., & Song, B. (2020). COVID-19: a review of what radiologists need to know. *World Journal of Clinical Cases, 8(22)*, 5501–5512. https://doi.org/10.12998/wjcc.v8.i22.5501

Tang X. (2019). The role of artificial intelligence in medical imaging research. *BJR Open, 2*(1), 20190031. https://doi.org/10.1259/bjro.20190031

The Radiologist. Johns Hopkins Medicine. (2019, November 19). Retrieved March 6, 2022, from https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/the-radiologist

Trzcinski, T. (2018). Multimodal social media video classification with deep neural networks. *Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2018*. https://doi.org/10.1117/12.2501679

Types of breast cancer: about breast cancer. American Cancer Society. (n.d.). Retrieved February 11, 2022, from https://www.cancer.org/cancer/breast-cancer/about/types-of-breast-cancer.html

Venugopalan, J., Tong, L., Hassanzadeh, H. R., & Wang, M. D. (2021). Multimodal deep learning models for early detection of alzheimer's disease stage. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-020-74399-w

Verheij, R. A., Curcin, V., Delaney, B. C., & McGilchrist, M. M.(2018). Possible sources of bias in primary care electronic health record data use and reuse. *Journal of Medical Internet Research, 20*(5), e185. https://doi.org/10.2196/jmir.9134

Weizenbaum, J. (1966). Eliza—a computer program for the study of natural language communication between man and Machine. *Communications of the ACM*, *9*(1), 36–45. https://doi.org/10.1145/365153.365168

Wilkinson, L., Thomas, V., & Sharma, N. (2017). Microcalcification on mammography: Approaches to interpretation and biopsy. *The British Journal of Radiology*, *90*(1069), 20160594. https://doi.org/10.1259/bjr.20160594

Xu, Z., So, D. R., & Dai, A. (2021). Mufasa: multimodal fusion architecture search for electronic health records. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(12), 10532–10540. https://doi.org/10.1609/aaai.v35i12.17260

Yadav, S. S., & Jadhav, S. M. (2019). Deep convolutional neural network based medical image classification for disease diagnosis. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0276-2

Yasaka, K., Akai, H., Abe, O., & Kiryu, S. (2018). Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. *Radiology, 286*(3), 887–896. https://doi.org/10.1148/radiol.2017170706

Zdolsek, G., Chen, Y., Bögl, H. P., Wang, C., Woisetschläger, M., & Schilcher, J. (2021). Deep neural networks with promising diagnostic accuracy for the classification of atypical femoral fractures. *Acta Orthopaedica, 92*(4), 394–400. https://doi.org/10.1080/17453674.2021.1891512

Zhao, J., Feng, Q.,Wu, P. et al.(2019). Learning from longitudinal data in electronic health record and genetic data to improve cardiovascular event prediction. *Scientific Reports 9*, 717. https://doi.org/10.1038/s41598-018-36745-x

Zhao, Y., Liu, Z, Yang, L, Cheng, H.(2012). Combing rgb and depth map features for human activity recognition. *Proceedings of The Asia Pacific Signal and Information Processing Association Annual Summit and Conference, pp. 1–4. IEEE.*

APPENDIX

MEMORANDUM

DATE: November 23,

2020 TO: Kelly

Overman, MD
Nasim Ahmadiyeh, MD

FROM: Mark Hecker, PharmD, MBA - Research Administration

RE: Research project – **Development of Deep Learning Algorithm for Early Stage Breast Cancer Detection using Longitudinal Images and Electronic Health Record.** ORA #21-010

Attached is the TMC Research Application Form signed by a Department Manager and Chairperson (or residency program director). In accordance with TMC policies, patient health information and other confidential information should never be emailed, stored on a flash or thumb drive, or maintained on a TMC laptop that is not encrypted.

You may conduct your research project after receipt of **final** approval from the Institutional Review Board (IRB).

If you have any questions, please contact Research Administration: *HSDResearch@umkc.edu* or (816) 235-6015.

UMKC Institutional Review Board
University of Missouri-Kansas City

5319 Rockhill Road
Kansas City, MO 64110
816-235-5927
umkcirb@umkc.edu

December 15, 2020

Principal Investigator: An-Lin Cheng
Department: Biomedical/Health Informatics

Your IRB Application to project entitled "Development of Deep Learning Algorithm for Early Stage Breast Cancer Detection using Longitudinal Images and Electronic Health Record." was reviewed and determined to qualify for IRB exemption according to the terms and conditions described below:

IRB Project Number	2025582
IRB Review Number	267845
Initial Application Approval Date	December 15, 2020
IRB Expiration Date	N/A
Level of Review	Exempt
Project Status	Active - Exempt
Exempt Categories	45 CFR 46.101b(4)
Risk Level	Minimal Risk
HIPAA Category	HIPAA Waiver

Approved Documents
REVISED HIPAA Waiver of Authorization, Edits to section 7B
Revised Data Collection Tool (remove identifiers)
memo-TMC Office of Research Administration approval
TMC Office of Research Administration approval

The principal investigator (PI) is responsible for all aspects and conduct of this study. The PI must comply with the following conditions of the determination:

1. No subjects may be involved in any study procedure prior to the determination date.
2. Changes that may affect the exempt determination must be submitted for confirmation prior to implementation utilizing the Exempt Amendment Form.
3. The Annual Exempt Form must be submitted 30 days prior to the determination anniversary date to keep the study active or to close it.
4. Maintain all research records for a period of seven years from the project completion date.

If you are offering subject payments and would like more information about research participant payments, please click here to view the UM system Policy on Research Subject Payments: https://www.umsystem.edu/oei/sharedservices/apss/nonpo_vouchers/research_subject_payments

If you have any questions, please contact the IRB at 816-235-5927 or umkcirb@umkc.edu.

VITA

On January 15, 1990, Zhiheng Zhang was born in a small town in Sichuan Province, China. From 2010 to 2014, he attended Shanghai JiaoTong University and earned a bachelor of science in industrial engineering. Then he moved to the United States to pursue a master's degree in industrial engineering, where he discovered a keen interest in statistics. So, following a brief stint as an SAS programmer for a contract research organization, he came to UMKC to pursue a master's and doctorate in statistics. He received a master's degree in statistics after five years of hard work and is now a candidate for a doctorate. He is now trying to complete his Ph.D. His co-discipline was biomedical informatics, and he acquired a strong interest in bio-statistics and decided to pursue a career as a bio-statistician. He will attend Northwestern University's Feinberg School of Medicine as a statistician to further develop his talents in medical-related statistics analysis.