Medical Imaging and Machine Learning

Trent R. Merkins
Division of Computer Science
University of Minnesota, Morris
Morris, Minnesota, USA 56267
merki002@morris.umn.edu

ABSTRACT

This paper will talk about how machine learning can be used to assist with the medical imaging process. It will talk specifically in regards to breast cancer medical images and go through some of the diagnosis process for breast cancer. Also there will be information on difficulties that occur when using machine learning. This paper will end with a possible solution to these difficulties as well as results from two different experiments.

1. INTRODUCTION

Breast cancer affects many men and women around the world. It is a deadly disease the affects the breast tissue. Breast cancer is an illness in which abnormal cell growth is found in the breast region and is dangerous to men and women.

Machine learning is the study of computer algorithms that improve automatically through experience. This is an important concept that will be referenced and explained throughout the paper. The sections throughout all focus on various aspects and types of machine learning.

In sections three and four there are descriptions of Artificial Neural Networks and Convolutional Neural Networks; these are neural networks used in machine learning techniques. These sections describe the structure and how they are used to assist with medical imaging. Section five is about computer aided detection and diagnosis. This describes how the computer aided detection works as well as the steps typically involved.

In section six it describes computer aided detection and diagnosis in regards to mammography. Section seven is about an additional method that is used to assist with medical imaging. This section is focused more on protocol setting and image quality.

In section ten a possible solution to the difficulties of machine learning and medical imaging is addressed. This section describes using an experiment how to efficiently increase sample database size.

2. BACKGROUND

Radiologists look at ultrasounds for various elements in the image and oncologists determines treatment. This paper will focus on radiologists and detecting and diagnosing cancerous cells. Radiologists look for masses and calcifications in the ultrasound image. These could be cancerous or benign, but are indicators for change in breast tissue. Calcifications are calcium deposits within the breast tissue that show up as white spots in an ultrasound.

These calcium deposits can be categorized into two different classifications: macrocalcifications and microcalcifiations. Macrocalcifications are larger calcium deposits that can be associated with aging of the breast arteries, old injuries and inflammation. Microcalcifications are the smaller calcium deposits and are more concerning. The shape and layout of these microcalcifications help to determine if they are due to cancer. If they are suspicious looking then a biopsy is used to confirm whether the calcifications are due to cancer or not.

The other things that they look at are masses that are present. A mass is a dense area of tissue that has a shape and edges that are different then the surrounding areas. Masses are an important change in a mammogram. They can be cysts (fluid-filled sacs) or solid masses. Solid masses can be concerning, but most are not cancerous.

Machine learning can be used to analyze these images similar to the way that radiologists do. They can be programmed using active learning methods and various other techniques to analyze the data more thoroughly than the radiologists could because of the amount of data that a machine can handle. Machine learning can be used as a second opinion to help improve the detection and diagnosis process.

This paper will focus directly on the application of Machine learning in the detection and diagnosis of breast cancer. It will look at the various elements that are used in order to improve the effectiveness of these diagnoses.

3. ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks are also known as ANNs for short. ANNs and neural networks in general are computational models. The information taken from biological systems that exist in the world help to create a foundation for the system. The systems take the information given to them and learn from the data similar to the living systems they are based on. They can take and process the information and then use this data to provide information to be utilized by an outside source. The part of the network that takes in the information as an input and transforms it into a valid output is called a *neuron*.

The neurons in an ANN are combined together in a group of layers. All of these layers are connected between each

UMM CSci Senior Seminar Conference, April 2020 Morris, MN.

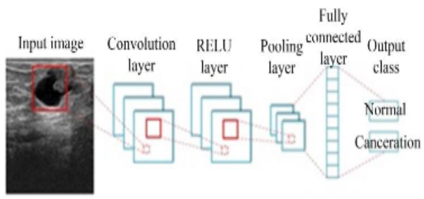

Figure 1: This shows a diagram of what the convolutional layers look like.

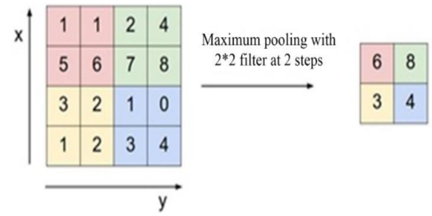

Figure 2: Diagram of Pooling functions

other to help with the transfer of information. The inputs that are received have a mathematical algorithm that is applied to them. This application allows the network to learn about the data and print out the correct output. The learning algorithms are organized into separate classes.

The two classes are supervised and unsupervised learning. The supervised learning has a system in it that helps with error correcting and works to balance the input and correct the values. The unsupervised algorithms focuses on self reorganization and corrections. Both of them make corrections, but the supervised algorithms use deviation between the result and the input to make an informed correction. [3]

3.1 Architecture

The architecture of a neural network is built by using the structure of the interconnections and the smaller computing units. The structure is usually based on a layer structure. It splits each of the neurons into layers. Then each of these layers are connected above and below by more layers. The layer receives information from the layer above or below them and then outputs information to the layers directly above or below. This connection focuses on making sure that interactions are only between the layers above and below and not in the layer itself. This also means helps to simplify the complexity of the structure by focusing on the layers independently.

The number of characteristics that are expected to be learned by the network depend on the structure of that network. It can vary from in the thousands into even many millions depending on the size. The size of a typical medical imaging model is rather small compared to the size of the larger classification systems. Further in depth information can be found [3].

4. CONVOLUTIONAL NEURAL NETWORKS

Convolutional Neural Networks have four distinct layers. The convolution layer, RELU layer, pooling layer and the fully-connected layer (See Figure 1). Each layer works together to identify the damaged tissue and mark it for the radiologist or system for further classification. It is important to note that it can distinguish these features apart from their location. The first layer is the convolutional layer. The information from this layer is very linear as it slides across the image filtering out lesions.

The next layer is the RELU layer. This layer takes the information from the convolutional layer and makes it more realistic. It uses an equation to help transform the linear data.

The pooling layer lies between these layers and helps to reduce the amount of invariance caused by rotation. It uses a matrix and takes the largest value as shown in Figure 2.

Filter procedures are used across the entire image and allowing the same feature to be found at different locations. Simple features are transferred to the next layer. Abstracted layer by layer, combined with basic features to show various objects.

The full connection layer takes the output of the RELU layer or pooling layer as input, and finds the probability of a classification. The recurring features are also shown in this layer. Further information can be found [4].

5. COMPUTER AIDED DETECTION AND DIAGNOSIS

Computer Aided Detection and Diagnosis also known as CAD is used with medical imaging. In order to assist with the detection and diagnosis of various diseases. CADe stands for Computer Aided Detection and CADx stands for Computer Aided Diagnosis. CADe will be discussed first since this is the first part of the process. [3]

5.1 Computer Aided Detection

CADe specializes in the detection process of Computer Aided Detection and Diagnosis. CADe typically follows a very specific outline when determining pathological from the surrounding tissues. Pathological features are damaged tissue caused by diseases.

1. The segmentation of the organ studied.
2. Pre-processing stage: this is usual a lesion signal enhancement with respect to the surrounding structures.
3. The detection of the lesion candidates.
4. Segmentation of the lesion candidate area.
5. Feature extraction: calculation of mathematical characteristics that describe the findings.
6. Classification of Lesion Findings according to the calculated features.
7. Application of false-positive reduction methods.
8. Final Result: computer output. Detected lesions with associated characteristics.

Segmentation is the dividing of something into segments or different parts. This is done to help the program better analyze whether their are lesions. Lesions are the region of damage caused by disease. The next step helps to enhance and show the damaged tissue area apart from the healthy tissue.

The third step focuses on detecting the damaged tissue area that may have been caused by cancer or a disease. Now that the lesion candidates have been found it splits these areas up into smaller parts. The splitting of these parts into smaller segments allows the computer detection system to analyze the smaller components further. The fourth step uses a mathematical algorithm that takes the characteristics found and computes whether the lesion findings are cancerous.

This process takes each of the calculations on the lesions and uses it to separate them into separate categories. This is not the same as classifying whether it is malignant or benign. This separates and classifies the lesions into categories grouping them with their associated characteristics. These methods are used to help with providing accurate lesion detection. The false-positive reduction methods help to reduce the amount of false lesion detections.

This final step returns a computer output that can be used by the radiologist or CADx system. The CADe system only provides an output with associated characteristics. This system does not diagnose the lesions as cancerous. This is an important distinction for going forward.

It is important that no lesion is missed in the detection process. One of the main disadvantages noted is that CADe systems tend to have a high false positive detection rate. This discourages radiologists from using them because of the additional work required to sort through the false positives. [3]

5.2 Computer Aided Diagnosis

CADx specializes in the diagnosis process of Computer Aided Detection and Diagnosis Process. CADx actually does the diagnosing not the CADe system. Even though CADx can be used to diagnose lesions it is still treated as a second opinion to help the Radiologist. CADx also has a group of processes that it follows.

1. Automatic or manual detection of lesions. This software can start from a previous CADe output or a set of lesion candidates from a radiologist.
2. Feature extraction and analysis of the input detected lesions.
3. Classification of the lesions based on the extracted features.
4. Diagnosis of the lesions extracted.

This first step can come from the output that was mentioned above from the CADe system. If the radiologist would rather not use that system, however they can still input their own candidate results. This next step is the part of the analysis that works looks at the lesions given to the system. It works to analyze each of the lesions characteristics. It extracts each of the features for the next step.

During the third step is where the features get sorted into various classifications. It splits each of these extracted features into separate categories. This final step is the part of the process where the CADx system actually provides a diagnosis of the damaged tissue and whether it is cancerous or not. This classification step not only assigns the lesion (damaged tissue) into a normal or abnormal or benign or malignant category. Benign means that the tumor or tissue is non-cancerous and malignant means that the tissue is cancerous. It also assigns a probability of the tissue belonging to a certain class. [3]

5.3 Advantages and Limitations

There are some advantages and limitations to types of CAD systems. Some of the advantages are as follows: Potential for assisting radiologists in image interpretation, reduction of the variability of image interpretation, providing access to hidden features and complex relationships and processing large amounts of information. Some of the limitations are as follows: Overfitting and the detection of a high number of False-Positives. Overfitting happens when the system starts to model the noise present and does not improve its capabilities even after the training process continues. [3]

6. CAD CLINICAL APPLICATIONS MAMMOGRAPHY

The small differences in density between lesions and healthy breast tissue make interpretation of mammography very difficult. CADe can be used in the detection of masses and detection of microcalcification clusters. CADx system based on ANNs are not only for differentiating between benign and malignant lesions, but also for predicting cancer probability on an individual basis.[3]

7. IMAGING PROTOCOL SELECTION

Another area that machine learning can applied to in order to assist with medical imaging is image protocol selection. Image protocol selection is normally performed by the radiologist. They take the information given by the patient in order to set the correct protocol for the image being taken. The protocol specifies various settings for the image being taken. If the given protocol is incorrect then the image can be distorted or unusable and must be retaken.

> "75 percent of the patients underwent CT, and of these the majority (58 percent) had repeated imaging at the trauma center; outside of head CT, the majority of the repeat scans were caused by inadequate technique at the local hospital." [1]

This shows the need for methods to be changed in order to increase the accuracy and efficiency of a method. Retaking exams of a patient can be dangerous for the patient and reflect poorly on the clinic. [1]

8. IMAGING PROTOCOL SELECTION USING MACHINE LEARNING

In order to understand which protocol is needed for medical imaging it becomes similar to a categorizing or classification problem. The information is obtained from a survey and then it is used to find the correct protocol. Each protocol acts as a category and the information gets sorted into each protocol to choose the correct one. The vast amount of information to sort through to obtain the optimal protocol setting makes this extremely difficult. There are many different types of information that all needs to be sorted. Knowledge is needed about the body structure, the way the disease interacts with the body, the way the body functions and current practices for medical imaging.

There are some components necessary for machine learning to be able to understand how to sort medical images. It uses The Unified Medical Language system created by the National Library of Medicine. It uses this in order to how to take survey info and apply it to medical ideas that can be used for protocols. Then it takes this info and creates a Meta-map with graph nodes (see Figure 3).

Machine learning must be trained through a set of training data. This training data contained exams with protocols

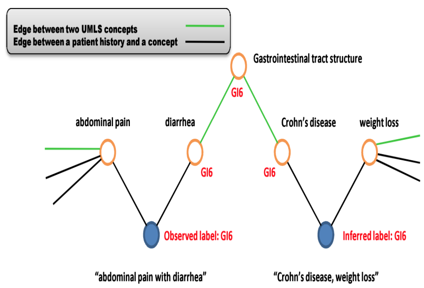

Figure 3: This is an image that shows an example of the algorithm. The patient history "abdominal pain with diarrhea" is labeled with protocol GI6 during the training phase. As a result the patient history "Crohn's disease, weight loss" is given the same protocol during the testing phase [1]

that the radioligists had used. This provides the graph with some of the labels that will be used. A technique called adsorption creates more protocol labels on this graph. MetaMap can then use the input to assign nodes and choose the protocol based on this information.

Looking at Figure 3 there are 5 nodes that are open and 2 nodes that are filled in. The nodes that are filled in from the survey information that the patient took. The node on the left is training information and the node on the right is from the test using the machine learning. The links between the nodes are considered what the node is relevant to. You can see that both are filled in nodes connect to a GI6 labeled node. The GI6 is a protocol from the list. These nodes both point to the GI6 node making the GI6 the correct protocol to be chosen.

There are two main things to focus on for the protocol. One is knowledge about the body and the other is about which part of the the disease affects. If you have both on the map this is what forms the connections to the relevant protocol. Extra information that is not relevant must be removed.[1]

9. CSAL MODEL

One of the main shortcomings of machine learning is the fact that the current Medical Imaging database is small. CSAL works to assist in correcting that and making sure that machine learning is as viable as possible.

The CSAL model is a machine learning model that combines CNN-based Seq2Seq-Attention Model and Alipy Active Learning model. This model performs well in image sequence conversion and sample selection

The reason that the CNN-based Seq2Seq attention model is used is because the original medical image data is too large. The two dimensional array is 3328*4084. In order to use the Active learning algorithm the image must be converted to a smaller size. The CNN-based Seq2Seq-Attention model can reduce this to a 40*40 two dimensional array. It also helps to reduce the amount of information lost during this process.

Firstly, deal with the grayed-out image data by CNN-based Seq2Seq-Attention model to obtain sequence data with small loss rate. After that apply these data to Study of Mammography Medical Imaging Sample Selection Based on CSAL the active learning model, through multiple iterations until the termination condition is reached (See Figure 4).[2]

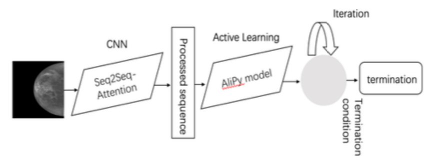

Figure 4: CSAL Model Architecture

9.1 CNN-based Seq2Seq-Attention Model

Seq2Seq is used across a variety of fields. It is used in image voice and even more. The Seq2Seq model consists of two parts: encoder and decoder (See Figure 5). The sample of the model takes the X as the input and maps it to the output Y. It does this using the hidden vector C.

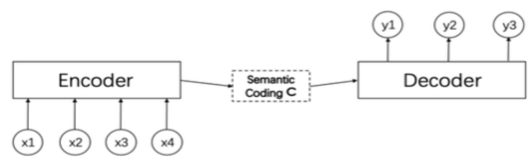

Figure 5: Encoder-Decoder

The biggest problem with the classical image is that it strictly encodes and decodes for a fixed C value. This doesn't allow the information to be entirely expressed and too much information will overwrite the current information. This will cause a lot of information to be lost. To correct this issue the Attention model was used. The Attention model architecture is shown in Figure 6.

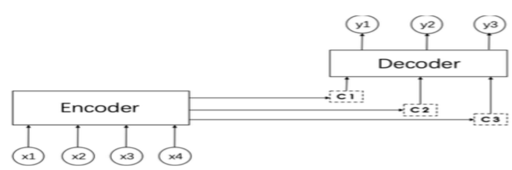

Figure 6: Attention Model

This model allows for more vector values of c and increases the non-linear capabilities. This allows more information to be stored and decreases the amount lost. The amount of information lost is based on the loss equation below.

$$loss = -1/N \Sigma \log P(Y_n|X_n, \Theta) \tag{1}$$

N is the amount of samples and X and Y are the input and output of the data. Each $P(Y_n|X_n, \Theta)$ is generated by the Encoder- Decoder framework. The Seq2Seq model used typically is based on RNN. CNN can replace the encoder and decoder to improve results. [2]

9.2 Active Learning

The layout of active learning uses A=(C,L,S,Q,U). C stands for the classifiers, L stands for the labeled training data, Q is the amount of times it queries in unlabeled samples, U is the entire unlabeled set and S is where the radiologist can mark the unlabeled samples (See Figure 7).

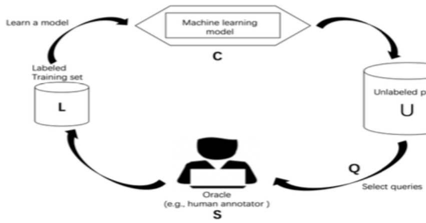

Figure 7: Active Learning

The basic steps are as follows:

1. Select appropriate machine learning model, actively select strategy Q, and divide the data (Training Set, Testing Set, Unlabeled Data Set).
2. Initialization: model C is initialized by training set.
3. Use the trained model C to predict U, obtain the prediction result of each sample, and query the sample n1 with large amount of information through Q.
4. Label n1 by the oracle, update the label set n=n+n1.
5. Train and update model C based on training set n.
6. Use model C to verify on the test set, stop the iteration in accordance with the convergence condition, otherwise loop execution 3-5.[2]

9.3 AliPy Active Learning Model

The AliPy active learning model provides an implementation of a module-based active learning framework. AliPy takes and combines multiple learning algorithms in order to help the machine learning system sort through data. It is easier to compare existing data and use this to understand how well it is performing. Uncertainty selection was one of the methods implemented to test CSAL. There are typically two choices for the uncertainty method: 1) Use probability to indicate the degree of uncertainty. 2) Use distance to indicate the degree of uncertainty. Of the two classification problems with probability indicating uncertainty, the method of selecting the lowest confidence is the most commonly used uncertainty measure (see [2] for further explanation).

10. EXPERIMENTS

Below are experiments and there various setups that were used in order to test the application of machine learning and medical imaging.

10.1 Experiment for Imaging Protocol Selection

There were about 1,000 abdominal CT studies at the institution where the CSAL experiment was conducted. Each had a short history attached with them that were 10-30 words. The majority of these included the protocol that the radiologist had selected. Overall there were 26 protocols in the dataset, but of those the 7 most common are included in the table 8.

The ONC1 protocol appears the most often and relates to cancer diagnoses. The other protocols in the table have specific diagnosis uses. [1]

10.2 CSAL Experiment

This experiment is based on Python, which contains a lot of proven machine learning framework, such as Random-forest-classifier that was used.

Dataset 1 comes from the breast-cancer standard dataset of sklearn.datasets. The data set is divided into two categories, the total sample size was 569, of which 212 were benign samples, 357 were malignant samples and has 6 types of sample labels, as shown in.

Dataset 2 comes from INbreast includes 115 cases, 410 images and corresponding 117 case reports. Each mammogram has different benign and malignant markers, corresponding to different targets. The same label is used for the INbreast data and the breast-cancer data and have a uniform resolution and pre-processing, while excluding the impact of Incomplete image, severe image tilt, blurred image, etc. [2]

10.3 CSAL Pre-process

First initialize the dataset and set the same settings for Dataset 1 and Dataset 2. Trained the model with 20 percent of all samples as the initial labeled sample set, 20 percent of the sample is used as the test set to test the performance of the base model, the remaining 60 percent of the samples are unlabeled sample sets.

Second set the value of the termination condition “number of query” to 6. The query number represents the true number of queries to oracle for unlabeled samples. Then add the oracle-tagged sample to the labeled sample set for training the base model. As the number of marked samples increases the more samples are obtained. More samples and comparisons increases the accuracy of the system.

Thirdly, set the value of “split-count” to 5 and the data set is divided into 5 according to the ratio set as above. Because each time the data set is randomly divided the labeled, unlabeled, and test sets contain different samples between each partition. This improves the credibility of the experiment. Finally, set the value of “round” to 3. “round=3” means that three parallel experiments are performed, and the experimental results are the average of the three experiments. The purpose of this is to further enhance the credibility of the experiment. [2]

11. RESULTS

Included in this section are the various results that were reached for the experiments. This also includes some of the graphs and other representations provided by the sources.

11.1 Imaging Protocol Results

The imaging protocol result data has multiple components involved to help prove that it is valid. Overall there were five runs performed. The data from these were split into different sections. 70 percent of this was training and the remaining 30 percent was testing data. The baseline or basic method predicts what protocol is being used based on how popular the method is. 60 percent of the time it is correct and 75 percent of the time the protocol chosen by the radiologist is in the top two choices. This does not prove that the radiologist has chosen the correct answer. This method just focuses on predicting the protocol chosen (see Figure 9). [1]

11.2 CSAL Experiment Results

Two existing active learning methods (QBC Query Committee and Uncertainty Selection Method) were used on the public dataset. The validity of the model proven by comparing the breast-cancer dataset on sklearn.datasets and the

Protocol	Count	Indications
ONC1	486	lymphoma, colon, gastric, prostate, testicular, ovarian, cervical, lung cancers
GI1	69	appendicitis, diverticulitis, IBD, abscess,fever of unknown origin, small-bowel obstruction
GI9	65	pancreatic lesion, pancreatitis
GI6	50	enterography, IBD
GI3	46	liver, cirrhosis, hepatoma, indeterminate liver lesion
ONC2	43	breast, islet cell, carcinoid, pheochromocytoma, thyroid, melanoma, sarcomas, choriocarcinoma
GU2	43	history of TCC, hematuria

Figure 8: This includes the information for the table below. It includes all abbreviations so that the protocols in the table can be understood.

		Predicted						
		ONC2	GU2	GI6	GI3	GI1	GI9	ONC1
Actual	ONC2	5	3	0	0	1	2	2
	GU2	3	7	0	0	1	1	1
	GI6	0	2	5	0	6	1	1
	GI3	0	2	0	7	1	3	1
	GI1	2	2	0	1	13	2	1
	GI9	0	0	0	0	1	19	0
	ONC1	2	1	0	1	4	7	10

Figure 9: The provides a comparison between the predicted method and the actual protocols picked by the attending radiologist.

INbreast dataset. Comparing results in between Unc, QBC and Rnd. Both Unc and QBC outperformed the Rnd (Short for random)(See Figure 10).

Methods	Number_of _quiries	Number_of_ different_split	Run time	Performance
Unc	6	3	1min	0.962±0.00
QBC	6	3	1.2mins	0.955±0.00
Rnd	6	3	3mins	0.943±0.01

Figure 10: Data set 1 Table

The model performed much better on the Breast-Cancer dataset than the INbreast dataset (See Figure 11). The Breast-Cancer data set has had the noise and defects removed from it. However, the INbreast dataset still has some noise in the data. The INbreast-based experiment results based on active learning methods are significantly better than Rnd. This shows that the method provided will assist with increasing the sample pool size effectively.

Figure 12 refers to the data from the first data set. This is the data set with the errors cleaned up. This data performed much better than the data that still contained errors (See Figure 13). This is important to consider because the data that is being looked at may not always be completely error free. This program is meant to search through data and provide samples for the radiologist to annotate. This system being used on a data set that has not been cleaned up can save time on that process so that the radiologists focus on the important images. [2]

12. CONCLUSION

The conclusion is that machine learning helps when used with medical imaging. There are some shortcomings and challenges when applying machine learning, however as it continues to develop machine learning continues to become

Methods	Number_of _quiries	Number_of_ different_split	Run time	Performance
Unc	6	3	2.2mins	0.802±0.02
QBC	6	3	3.1mins	0.809±0.01
Rnd	6	3	5.5mins	0.754±0.02

Figure 11: Data set 2 Table

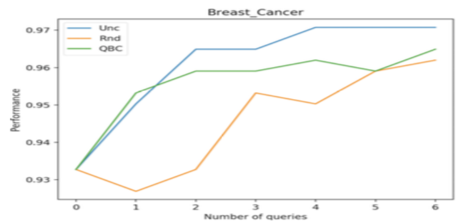

Figure 12: Breast-Cancer Dataset 1 Graph

a valuable asset to oncologists, radiologists and doctors. Medicine is continuously evolving in order to provide better assistance to patients and a key part of that future is machine learning.

13. REFERENCES

[1] A. Bhat, G. Shih, and R. Zabih. Automatic selection of radiological protocols using machine learning. In *Proceedings of the 2011 Workshop on Data Mining for Medicine and Healthcare*, DMMH '11, page 52–55, New York, NY, USA, 2011. Association for Computing Machinery.

[2] J. Li and W. Mi. Study of mammography medical imaging sample selection based on csal. In *Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence*, ACAI 2019, page 485–490, New York, NY, USA, 2019. Association for Computing Machinery.

[3] J. H. Rodríguez, F. J. C. Fraile, M. J. R. Conde, and P. L. G. Llorente. Computer aided detection and diagnosis in medical imaging: A review of clinical and educational applications. In *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality*, TEEM '16, page 517–524, New York, NY, USA, 2016. Association for Computing Machinery.

[4] G. Yunchao and Y. Jiayao. Application of computer vision and deep learning in breast cancer assisted diagnosis. In *Proceedings of the 3rd International Conference on Machine Learning and Soft Computing*,

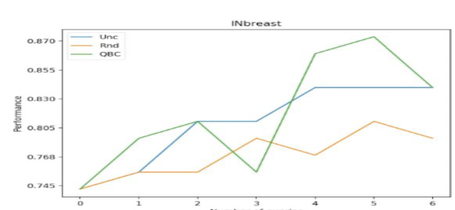

Figure 13: INbreast Dataset 2 Graph

ICMLSC 2019, page 186–191, New York, NY, USA, 2019. Association for Computing Machinery.