



AI-Based Age Estimation from Mammograms

Fabia Afzal
Charitha Dissanayake Lekamlage

This thesis is submitted to the Faculty of Computing at Blekinge Institute of Technology in partial fulfilment of the requirements for the degree of Master of Science in Computer Science. The thesis is equivalent to 20 weeks of full time studies.

The authors declare that they are the sole authors of this thesis and that they have not used any sources other than those listed in the bibliography and identified as references. They further declare that they have not submitted this thesis at any other institution to obtain a degree.

Contact Information:

Author(s):

Fabia Afzal

E-mail: faaf18@student.bth.se

Charitha Dissanayake Lekamlage

E-mail: didi18@student.bth.se

University advisor:

Abbas Cheddad (Senior lecturer/Associate professor)

Department of Computer Science

Faculty of Computing
Blekinge Institute of Technology
SE-371 79 Karlskrona, Sweden

Internet : www.bth.se
Phone : +46 455 38 50 00
Fax : +46 455 38 50 57

Abstract

Background. Age estimation has attracted attention because of its various clinical and medical applications. There are many studies on human age estimation from biomedical images such as X-ray images, MRI, facial images, dental images etc. However, there is no research done on mammograms for age estimation. Therefore, in our research, we focus on age estimation from mammogram images.

Objectives. The purpose of this study is to make an AI-based model for estimating age from mammogram images based on the pectoral muscle segment and check its accuracy. At first, we segment the pectoral muscle from mammograms. Then we extract deep learning features and handcrafted features from the pectoral muscle segment as well as other regions for comparison. From these features, we built models to estimate the age.

Methods. We have selected an experiment method to answer our research question. We have used the U-net model for pectoral muscle segmentation. After that, we have extracted handcrafted features and deep learning features from pectoral muscle using ResNet-50 and Xception. Then we trained Support Vector Regression and Random Forest models to estimate the age based on the pectoral muscle of mammograms. Finally, we observed how accurately these models are in estimating the age by comparing the MSE and MAE values. We have considered breast region (BR) and the whole MLO to answer our research question.

Results. The MAE values for both SVR and RF models from handcrafted features is around 10 in years in all cases. On the other hand, with deep learning features MAE is less as compared to handcrafted features. In our experiment, the least observed error value for MAE was around 8.4656 years for the model that extracted the features from the whole MLO using ResNet50 and SVR as the regression model.

Conclusions. We have concluded that the breast region (BR) is more accurate in estimating the age compared to PM by having least MAE and MSE values in its models. Moreover, we were able to observe that handcrafted feature models are not as accurate as deep feature models in estimating the age from mammograms.

Keywords: Machine Learning, Mammograms, Image Segmentation, Deep Learning, Feature Extraction, Regression Analysis, Age Estimation

Acknowledgments

We would like to express our special thanks of gratitude to our supervisor Abbas Cheddad, for his continuous support, expert advice and wonderful supervision. His guidance and motivation helped us throughout the research in writing our thesis. We were grateful to have him as our supervisor.

Secondly, we would like to pay special thanks to our parents, family and friends for their love and support throughout this journey.

Contents

Abstract	i
Acknowledgments	ii
1 Introduction	1
1.1 Problem Statement	2
1.2 Aim and Objectives	2
1.3 Research Question	3
1.4 Thesis Structure and Length	3
2 Background	4
2.1 Image Segmentation	4
2.2 Deep Learning for Segmentation	4
2.3 U-net Architecture	5
2.4 U-net Evaluation Metrics	5
2.5 Handcrafted Features	6
2.5.1 Mean Intensity	6
2.5.2 Standard Deviation	6
2.5.3 Entropy	6
2.5.4 Haralick Features	7
2.6 Feature Extraction using Deep Learning	7
2.6.1 Xception	7
2.6.2 ResNet	8
2.7 Regression Analysis	8
2.7.1 Support Vector Regression	8
2.7.2 Random Forest	9
2.8 Performance Measures	9
2.8.1 Mean Square Error	9
2.8.2 Mean Absolute Error	10
3 Related Work	11
4 Method	15
4.1 Experiment	15
4.2 Database	16
4.2.1 Ethical Aspects	17
4.3 Software Environment Setup	17
4.4 Dataset Cleaning	17

4.5	Implementation	18
4.5.1	Data Labeling for U-net Segmentation	18
4.5.2	U-net Implementation	19
4.5.3	Handcrafted Feature Extraction	20
4.5.4	Deep Learning Feature Extraction	21
4.5.4.1	ResNet50	21
4.5.4.2	Xception	21
4.5.5	Fitting models for Age Estimation	22
4.5.5.1	Random Forest	22
4.5.5.2	SVR	22
5	Results and Analysis	25
5.1	U-net Segmentation	25
5.2	Handcrafted Features	26
5.2.1	PM Segmented Image Features	26
5.2.2	BR Segmented Image Features	27
5.2.3	PM+BR Image Features	28
5.2.4	Handcrafted with whole MLO(BR+PM)- 7 Features	29
5.3	Deep Learning Features	30
5.3.1	Resnet50 PM Segmented Image Features	30
5.3.2	Resnet50 BR Segmented Image Features	31
5.3.3	Resnet50 PM+BR Image Features	32
5.3.4	Xception PM Segmented Image Features	33
5.3.5	Xception BR Segmented Image Features	34
5.3.6	Xception PM+BR Image Features	35
5.4	Summary	36
5.4.1	Significance Test	40
6	Discussion	42
6.1	Discussion	42
6.2	Validity Threats	44
6.2.1	Internal Validity	44
6.2.2	External Validity	44
6.2.3	Conclusion Validity	45
7	Conclusions and Future Work	46
7.1	Conclusion	46
7.2	Future Work	46
	References	48

List of Figures

4.1	Workflow part 1 of Experiment	16
4.2	Workflow part 2 of Experiment	16
4.3	Age distribution of each catagory and of the over all samples (bottom-right)	18
4.4	Labelled Image. Blue:PM; Yellow:BR; Red:Background	18
4.5	Trained U-net Network	20
4.6	Function to train RF	22
4.7	Function to train SVR	24
5.1	An Image of Sementic Segmentation	25
5.2	An Image of Segmented Regions	26
5.3	SVR with handcrafted+PM Features	27
5.4	RF with handcrafted+PM Features	27
5.5	SVR with handcrafted+BR Features	28
5.6	RF with handcrafted+BR Features	28
5.7	SVR with handcrafted+PM+BR Features	29
5.8	RF with handcrafted+PM+BR Features	29
5.9	SVR with handcrafted+PM+BR(whole MLO) Features	30
5.10	RF with handcrafted+PM+BR(whole MLO) Features	30
5.11	SVR with Resnet50+PM Features	31
5.12	RF with Resnet50+PM Features	31
5.13	SVR with Resnet50+BR Features	32
5.14	RF with Resnet50+BR Features	32
5.15	SVR with ResNet50+PM+BR(whole MLO) Features	33
5.16	RF with ResNet50+PM+BR(whole MLO) Features	33
5.17	SVR with Xception+PM Features	34
5.18	RF with Xception+PM Features	34
5.19	SVR with Xception+BR Features	35
5.20	RF with Xception+BR Features	35
5.21	SVR with Xception+PM+BR(whole MLO) Features	36
5.22	RF with Xception+PM+BR(whole MLO) Features	36
5.23	MAE Results Graph	37
5.24	MSE Results Graph	39

List of Tables

4.1	Environment Setup Tools	17
4.2	Training options for U-net Segmentation	19
4.3	Feature Extraction Built-in Functions	21
5.1	Segmentation Evaluation Metrics Results	25
5.2	Handcrafted+PM Features Results	26
5.3	Handcrafted+BR Features Results	27
5.4	Handcrafted+PM+BR(14 Features) Results	28
5.5	Handcrafted+PM+BR(whole MLO) Results	29
5.6	ResNet50+PM Features Results	30
5.7	Resnet50+BR Features Results	31
5.8	Resnet50+PM+BR(whole MLO) Features Results	32
5.9	Xception+PM Features Results	33
5.10	Xception+BR Features Results	34
5.11	Xception+PM+BR(whole MLO) Results	35
5.12	MAE Results Summary	37
5.13	MSE Results Summary	38
5.14	Results of Test1	40
5.15	Results of Test2	41

LIST OF ABBREVIATIONS

MRI: Magnetic Resonance Imaging
BMI: Body Mass Index
FFDM: Full-Field Digital Mammography
APD: Area Per Density
MIP: Mean Intensity of Pectoral Muscle
AI: Artificial Intelligence
MAE: Mean Absolute Error
MSE: Mean Square Error
CNN: Convolutional Neural Network
CRN: Convolutional Residual Network
RNN: Recurrent Neural Network
ReLU: Rectified Linear Unit
GLCM: Grey Level co-occurrence Matrix
AAE: Automatic Age Estimation
SVR: Support Vector Regression
SVM: Support Vector Machine
RF: Random Forest
DCNN: Deep Convolutional Neural Network.
TW3: Tanner-Whitehouse 3
ANN: Artificial Neural Network
KNN: K-Nearest Neighbour
RMSE: Root Mean Square Error
3D-CNN: 3-Dimensional Convolutional Neural Networks
BR: Breast Region
PM: Pectoral Muscle
DDSM: Digital Database for Screening Mammography
MLO: Mediolateral Oblique
CC: Cranio-Caudal
GDPR: General Data Protection Regulation
ResNet: Residual Neural Network

Biomedical imaging is a technique of creating a visual representation of the body that can be used for medical diagnoses and clinical analyses. Biomedical imaging involves the use of various technologies such as X-Rays, CT- scans, magnetism (MRI), ultrasound, mammography, light (endoscopy, OCT) or radioactive pharmaceuticals (nuclear medicine: SPECT, PET) for detecting and treating medical conditions of patients. Along with that biomedical images can be used for estimating the age of a person.

We were able to find many studies on human age estimation from face images, dental images, MRI, X-rays etc for different purposes. For example, in one of the studies [28], researchers work on predicting the age of a patient from a chest X-ray by using deep learning methods. Similarly, some researchers presented a software-based solution for estimating age automatically based on 3D MRI images of the hand [53]. There are many studies related to age estimation from biomedical images. However, to our knowledge there is no research done related to mammograms in age estimation. Therefore, in our research we focus on mammogram images.

Mammography is specially designed medical imaging that uses low dose X-ray to view inside breast which helps in early diagnosis of breast diseases in women. On mammograms we can observe fat, fibroglandular tissue and pectoral muscle region. Research conducted by [11] investigated the association of pectoral muscle as a risk predictor in breast cancer using full-field digital mammography(FFDM). This study includes 3 image-based measures which are area percent density (APD), volumetric percent density (VPD), and the mean intensity of the pectoral muscle (MIP). They also studied the relationship of mean intensity of pectoral muscle with age and BMI. In another study, the authors contemplated possible explanations of why MIP is associated with breast cancer risk [12]. This resulted in a strong significant association between age and pectoral muscle mean intensity. Therefore, we plan on focusing on the pectoral muscle area in mammograms in estimating the age by considering other statistical features as well as mean intensity and to deploy deep/machine learning into the process.

In our work, our goal is to make an AI-based model for estimating age from mammograms images based on the pectoral muscle segment. This research will be helpful in filling out the missing values in the age index (null values) in many mammography databases. These null values can be caused by many reasons. For

example, it is quite common that in data collection some patients may not want to reveal their age, therefore when we put patient data in an excel sheet or database we find null values of age, or there can be mistakes in data entry. Therefore, for such scenarios normally researchers either discard the data which consist of these null values or take the average age in filling the data which is not an optimal way to fill null values. This will result in a waste of resources as well as affect the prediction power in many studies which consider age as a factor. In many breast cancer studies, it is proven that the age of the patient is a risk factor in breast cancer. Therefore, it is crucial for studies like breast cancer prediction models to have accurate age values than average ones [27][42]. However, if our chosen method provides positive results, we could use this to estimate the age in a specific manner.

1.1 Problem Statement

From the existing literature, we found that there are many studies related to age estimation based on biomedical imaging such as x-rays images of hand, teeth, knees etc. or studies related to MRI images of the brain, facial images. But based on our literature review, we were unable to find any research based on mammogram images related to age estimation and we noticed the need of this research as mentioned previously. Therefore, our research is based on age estimation from mammograms by considering the pectoral muscle. This research will be helpful for the researchers to estimate the age of patients and fill out values of age to conduct their research and in many other such scenarios.

1.2 Aim and Objectives

Our aim of this research is to estimate the age based on the pectoral muscle area in mammograms using machine learning techniques. In order to achieve our aim following objectives are considered.

1. To segment the pectoral muscle from mammograms(MLO view).
2. To extract deep learning features and handcrafted features from pectoral muscle segment.
3. To build models based on pectoral muscle features in mammograms for estimating age.
4. To compare the accuracy of each model based on mammogram regions for estimating the age.

We have conducted an experiment in order to achieve our objectives. U-net has been used to segment the pectoral muscle from mammograms. We have used ResNet50, and Xception models to extract the deep learning features from the pectoral muscle. Random forest and Support Vector Machines are used to estimate the age based on extracted features. The performance of each model is evaluated based on Mean

Absolute Error(MAE) and the Mean Square Error(MSE). In order to see the effect of age estimation from mammograms, we have also considered other regions of the mammogram. We have considered the Breast region (BR) of the mammogram in estimating the age as a comparison of this study.

1.3 Research Question

We would like to investigate the relationship between the pectoral muscle in mammograms and the age in order to estimate the age based on the pectoral muscle features. Therefore, we have formulated our research question as follows.

R1: How accurate it is to consider the pectoral muscle segment in mammograms to estimate age?

This research question will help us to find the level of association between pectoral muscle region and age, if any. From the results we can conclude how accurately we could estimate patient age from pectoral muscle region in mammograms.

1.4 Thesis Structure and Length

The structure of the thesis is divided into seven chapters. The first chapter is about the introduction of the thesis along with aims and objectives, research questions and motivation. The second chapter consists of background explaining important concepts related to our research. The third chapter consists of related work. Fourth chapter describes how the research has been conducted and the motivation behind the method. The fifth chapter discusses the results and analysis. The sixth chapter is about the discussion of results and answering the research question. The seventh chapter is about the conclusion and the future work.

In this section, we have discussed the concepts we have addressed in our thesis. The content in these sections will help the readers to understand the theme and the approach followed in the thesis. Here we have discussed the concepts related to image segmentation, feature extraction, regression models, deep learning models and performance measures.

2.1 Image Segmentation

Image segmentation is a process of partitioning an image into multiple regions. It is one of the approaches used in image processing to separate objects and textures within an image. In the medical field, image segmentation plays an important role in clinical analysis and diagnosis by segmenting the medical images. There are several different ways of image segmentation. In the past, most common methods used for image segmentation were naive thresholding, K-means clustering, histogram-based image segmentation, edge-detection etc.[19]. Many other methods have introduced different ways to segment the pectoral muscle area from mammograms [51][18][37]. These methods required significant human effort and expertise which is considered less efficient with the development of modern image segmentation techniques. These modern techniques involve deep learning technology.

2.2 Deep Learning for Segmentation

In one of the studies [23], the authors have presented an overview of popular methods that use deep learning techniques for segmentation of medical images. This paper mainly focuses on machine learning approaches applied by many researchers for segmentation of medical images and discusses their problems and strength. The authors put light on CNN, 2D CNN, 2.5D CNN, 3D CNN, FCN, cascaded FCN, Focal FCN, multi-stream FCN, 2D U-net, 3D U-net, V-Net, Convolutional Residual Networks (CRNs) and Recurrent Neural Networks (RNNs) architectures. Authors argue that deep learning techniques play a significant role in segmenting medical images. In this paper it is also highlighted that U-net is capable of fast and precise image segmentation. U-net was first introduced by Ronneberger et al stating that it is effective even with a limited amount of dataset images [43]. In one of the papers [49], U-net architecture was used to segment the breast region and pectoral muscle from digital mammogram images. It was concluded that U-net achieved a median dice similarity

coefficient of 0.8879 on pectoral muscle segmentation and performance was improved by training with normalized images.

2.3 U-net Architecture

The U-net architecture consists of a contracting path and an expansive path. The contracting path consists of two 3x3 convolutions each of which followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling. The feature channels are doubled in each downsampling unit. The expansive path consists of an upsampling of the feature map followed by a 2x2 convolution (“up-convolution”) that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU. The final layer consists of a 1x1 convolution which is used to map each 64 component feature vector to the desired number of classes [43].

2.4 U-net Evaluation Metrics

Segmentation evaluation metrics are used to evaluate the results of semantic segmentation against ground truth [3]. Some of which are discussed below:

Mean Accuracy: Accuracy metric is used to determine how well each class identifies pixels correctly. It is the ratio of correctly identified pixels to the total amount of pixels. For every image, mean accuracy gives the average accuracy of all classes present in an image [3].

Mean BFscore: BFscore identifies how well the true boundary of each class coordinates with the predicted boundary. This score is calculated between true segmentation in the ground truth and predicted segmentation in prediction. For every image, mean BFscore gives the average BFscore of all classes present in an image [3].

Global Accuracy: Global accuracy metric is used to make a quick estimate of correctly identified pixels in an image, regardless of class. It is the ratio of correctly classified pixels to the total number of pixels in an image [3].

Mean IoU: IoU (Intersection over Union) is the most commonly used metric for segmentation evaluation. It is the ratio of correctly identified pixels to the total amount of predicted pixels and ground truth in the class. For every image, Mean IoU gives the average IoU score of all classes present in an image [3].

Weighted IoU: This metric is used to minimize the impact of the error on classes when classes have disproportionality in their size. It is the average IoU weighted by the number of pixels in each class [3].

2.5 Handcrafted Features

Hand craft features can be described as properties derived from image information using various approaches. In our thesis we have focused only on seven statistical features: Mean intensity, Standard Deviation, Correlation, Contrast, Energy, Homogeneity and Entropy. There are many statistical features used for analysing data in different fields. Such as in digital image processing, researchers have used many statistical features such as Mean intensity, Standard Deviation, Correlation, Contrast, Energy, Homogeneity, variance etc. in their work for different purposes [47][46][45]. Paper [25] provides a comprehensive survey of the texture feature extraction methods. They have mentioned seven categories in text feature extraction. They are statistical approaches, structural approaches, transform-based approaches, model-based approaches, graph-based approaches, learning-based approaches, and entropy-based approaches. In statistical approaches, they discuss different approaches such as grey level co-occurrence matrix (GLCM), grey level run-length matrix, auto correlation based approaches, histogram of gradient magnitudes etc. In order to classify the texture in GLCM, each metric should be condensed to a few numbers. Therefore, they have highlighted a few studies that propose measures to be selected. They have discussed one paper which proposes 14 measures and another mentioning only 5 of these measures are sufficient from 14. These 5 measures are energy, entropy, correlation, homogeneity, and contrast. Authors of this paper also mentioned some studies which highlighted the GLCM method which gives better results where textures are visually easily separable, and they are easy to implement. Also, his paper has mentioned that the GLCM method has shown good results in large fields of applications. The 4 measures contrast, Homogeneity, Correlation and energy are defined as Haralick features [20] which we have demonstrated under Haralick Features.

2.5.1 Mean Intensity

Intensity is defined as a numerical value of the pixel and mean intensity refers to the average of these pixel values. In one of the studies mean intensity has been considered as a statistical feature in feature extraction from mammograms [46].

2.5.2 Standard Deviation

In terms of image processing standard deviation shows how much pixel variation exists from the expected value or average mean value. If the standard deviation is low it indicates that pixel intensity values are very close to mean and if it is high it shows that pixel intensity values are spread on a large range [30].

2.5.3 Entropy

One of the studies states that entropy is the measure of uncertainty associated with a random variable. In information sense, it can be the statistical measure of randomness for a given image to characterize the texture [45].

2.5.4 Haralick Features

In one of the studies, Haralick, Robert M. , extracted textual features of an image block by computing grey tone spatial-dependence probability-distribution matrices [20]. We have briefly discussed these features as follows:

Contrast: In a video image or photograph contrast is the measure as a difference between dark and light areas in an image. It also affects the ability to capture image details [45] . It is basically the difference between minimum and maximum pixel intensity values within an image.

Homogeneity: Homogeneity in image is defined as a change in intensity that occurs in a region of an image. It shows the uniformity of various elements and indicates how similar they are. For example, If the region in an image is darkened or lightened uniformly then homogeneity does not change.

Correlation: Correlation is used to extract information from an image [45]. This statistical feature measures how a pixel is correlated with its neighbour in an image.

Energy: Energy is used for describing the measure of information when some operation is formulated under a probability framework such as (maximum a priori) MAP estimation in conjunction with Markov random fields [45].

2.6 Feature Extraction using Deep Learning

For feature extraction using deep learning, we have used pre-trained models as feature extractor in which input images propagate forward, stop at specified layer and take output of that layer as features. From the literature review we found that researchers have used several deep learning architectures for extracting features from images such as Xception, ResNet, GoogleNet etc. Some of these networks are discussed below:

2.6.1 Xception

Xception (Extreme Inception) is a convolution neural network architecture introduced by google in 2017 [13]. It is built based on separable depth wise convolution layers. The architecture of Xception consists of 36 convolutional layers which form the base of a network for feature extraction. These 36 layers are structured into 14 modules. Except the first and last module, these all modules have a linear residual connection around them. The separable convolutional layers with residual connection makes the architecture easy to define and modify [13]. In the paper [13], Xception was compared with the Inception V3 architecture. Results show that Xception performs slightly better than Inception V3 on the ImageNet dataset and also on large image classification dataset. In one of the studies [40], the author uses the Xception architecture along with the other models such as VGG19, VGG16, ResNET-50, Inception V3 for learning the aging pattern from facial images in order to estimate the human age. It is found that, with unfrozen layers (No Transfer Learning) Xception model is the best model followed by Inception V3, ResNet-50, VGG19, and

VGG16 for automatic age estimation (AAE). It is concluded that the method based on Xception architecture performs best among all state-of-art methods for automatic age estimation. When pre-trained on ImageNet and CASIA-Web face dataset it obtained a MAE of 2.35 and 2.01 years respectively. Thus, we have selected Xception as one of the deep learning model for feature extraction.

2.6.2 ResNet

ResNet was introduced in 2015 and won first place in the ILSVRC competition on tasks of ImageNet localization, ImageNet detection, COCO segmentation and COCO detection [21]. ResNet are deep architectures achieved good coverage and high accuracy in many classification and recognition problems [38]. This deep architecture is designed on the principle of residual learning. There are several versions that belong to the family of ResNet such as ResNet-101, ResNet-50, ResNet-34, ResNet-18 and ResNet-152. The number of layers in each architecture is the same as the number given in the name of the network [38]. In one of the studies [40], authors presented the comparative analysis of many deep learning architectures for automatic age estimation (AAE) from facial images. For extracting deep learning features they use different models such as Resnet50, VGG19, VGG16, InceptionV3 and Xception. Authors concluded that ResNet50, Xception and InceptionV3 outperforms the other models and learn more information. In another study[44], authors have used ResNet50 for age estimation based on chest radiography. They were able to achieve MAE around 4.94 years, which comparative to other models scored in lowest MAE. Therefore, we were motivated to use ResNet50 as one of the deep learning models for feature extraction.

2.7 Regression Analysis

Regression analysis is a dynamic technique that is used to determine a relationship between two or more variables. The process of regression analysis allow us to examine which factor is important, which factor can be avoided and how they affect each other. Regression analysis can be used for multiple problems such as time series problems, forecasting and examining relationship between independent variables (X) and dependent variable (Y). It is one of the important techniques for analysing and modeling data. In the process of regression analysis we try to fit a line between data points in such a way that distance can be minimized between data points and line/curve. There are many regression techniques that can be used for making predictions.

2.7.1 Support Vector Regression

Support Vector Machine is a machine learning algorithm which comes under the category of supervised learning algorithms. SVM can be used both for classification and regression problems. In classification problems, the idea of a support vector machine is to create a line or a hyperplane which divides data into separate classes. So that after separation any new case can guess the target class. While in regression

problems, support vector regression (SVR) use the same principle as SVM and return continuous-valued output. But the main idea of this algorithm is to reduce the error and determine a hyperplane that maximizes the margin [9]. In recent years, SVM has attracted many researchers in machine learning areas. There are many studies that use a support vector machine model for classification and regression problems. In one of the studies [36], authors have used SVM and SVR for training the classifier and building regression function for age estimation from face images. The results show the best MAE with this model. In another study [14], authors presented a method by using support vector machines for large scale regression problems which yield outstanding performance on many time series prediction problems and regression problems. There is another study [35], which presents regression-based model named as Hybrid Constraint Support Vector Regression (HC-SVR) based on a support vector regression (SVR) for human age estimation from facial images. The proposed method is also compared with standard SVR. Results suggests that HC-SVR model performed well on accurate age estimation and standard SVR also gives the good accuracy.

2.7.2 Random Forest

Random forest is a machine learning algorithm which is flexible and easy to use, and also produces good results most of the time. It is a supervised learning algorithm and can be used for both regression and classification problems. Random forest builds multiple decision trees and combines them to get a more stable and accurate prediction. This algorithm can be used in different fields like the stock market, e-commerce, banking and medicine. In the paper[16], the author highlights the importance of different approaches of machine learning in medicine. It states that models such as random forest: decision tree algorithms and k-nearest neighbour perform remarkably well when working with complex interactions between features. The author gives special attention to the random forest stating it as an innovative and highly effective algorithm while referring to many other studies. It also states that the final majority vote across many trees is remarkably accurate. In one of the studies [38], the authors have used random forest as a regression model for estimating the age based on hand X-ray images with ResNet as a feature extractor. In this study they were able to achieve accuracy up to 100% the other models such as k-nearest neighbour and linear discriminant. Therefore, we decided to select random forest as one of the regression models together with the other selected deep learning models.

2.8 Performance Measures

We have used Mean Square Error and Mean Absolute Error to evaluate the performance of our models.

2.8.1 Mean Square Error

Mean square error is a statistical method that calculate the average of the set of errors. Mean square error gives an idea of how close are the points to the regression

line. This is done by taking the distance of the regression line from points and then squaring them. The distance between points and the regression line is known as error. The formula to calculate the MSE is given where n is the number of samples, y_k is the actual age for the sample k and \hat{y}_k is the predicted age.

$$MSE = \frac{1}{n} \sum_{k=1}^n (y_k - \hat{y}_k)^2 \quad (2.1)$$

2.8.2 Mean Absolute Error

Mean absolute error is a statistical method used to measure accuracy of a model. Mean absolute error measures the average of all absolute errors in the set of predictions. It is the average of absolute differences between actual and predicted observations where the weight of each individual difference is equal. Mean absolute error can be calculated using following formula where n is the number of samples, y_k is the actual age for the sample k , \hat{y}_k is the predicted age and $|\cdot|$ denotes the absolute value.

$$MAE = \frac{1}{n} \sum_{k=1}^n |y_k - \hat{y}_k| \quad (2.2)$$

In this section, we have reviewed a few of the related studies in order to support our research. There is absolutely no prior studies on age estimation from mammographic images, hence, the research gap. This section is dedicated to related work on age estimation from medical scans (knees, hands, etc) and others like age estimation on human faces using deep learning.

In one of the papers, authors presented a software-based solution for estimating age automatically based on 3D MRI images of the hand[53]. The method proposed in this article is handling the problems occurring in radiological methods such as the subjective influence of the examiner, exposure to ionizing radiation and the necessity of defining a new MRI staging system. Other than this, comprehensive analyses about machine learning methods such as deep convolutional neural networks and random forest have been presented in this paper regarding image simplifications. It is concluded that a DCNN based approach reaches new state-of-art accuracy in comparison with other deep learning based approaches. Moreover, the results obtained from the age regression method have shown that it works accurately for age up to 18 years.

Similarly, in another article[28], authors work on predicting the age from chest X-ray images using a deep learning approach. They have trained convolutional neural networks in a regression fashion on the dataset. They try to predict which areas on chest X-ray is important for CNN for predicting age by using activation maps. Authors found that areas near the shoulder, spine, clavicles and mediastinum are activated and thus believed to be important. They received an accuracy of 94% for PA CXRs.

In one of the research studies[8], the authors focus on evaluating several convolution neural networks (CNN) architectures for age estimation from dental X-ray images. For this purpose, they train several convolutional neural network (CNN) architectures using the dataset consisting of 2000 images belonging to seven different classes. They also used the concept of transfer learning for training CNN architectures, such as VGGNet, ResNet and AlexNet for estimating age. Authors apply Capsule Network on pre-trained models to compare with CNNs and transfer learning approach. Results suggest that CNN gives less accuracy (40%) during the experiment on rotated objects, overlapped teeth and missing teeth. And Capsule network overcomes the limitations of CNN by giving an accuracy of 76 %.

There is another study related to age estimation [15], which proposed an automated approach for estimating the age of adults and youth in the age range (14 to 21 years of age) based on deep learning technique using knee MRI. The dataset included MRI scans of 221 Males and 181 females. The proposed method comprises two CNN models, in which the first model selects the MRI which is most informative in terms of age assessment and the second model is responsible for age estimation. For testing, various CNN architectures were used and trained from scratch along with the transfer learning. Authors concluded that the GoogLeNet CNN architecture provides the best result. And the accuracy of 98.1% for males and 95.0 % for females were achieved.

Article [17], focus on an automatic age estimation method based on a deep learning algorithm. Authors proposed a framework based on deep learning for the task of age classification in which face images are assigned to labels having age range. The proposed method makes use of deep convolutional neural networks CNN architecture for extracting complex visual features relating to age in order to predict the age range of input images. Because of unlabeled data, authors used transfer learning for training CNN model. Authors argued that the trained model was able to extract facial features and the results of the experiment suggested that the proposed model outperforms other state-of-the-art methods.

In one of the studies, the authors proposed a method for estimating the age of adolescents based on a machine learning approach using MRI images of hand and wrist[48]. This method is based on an artificial neural network approach. The authors use MRI scans of 12-17 years old Chinese adolescents. For analysis of artificial neural networks (ANN) they include some maturity indicators such as body weight, body height, composition intensity of bone marrow from MRI and skeleton age using Tanner-Whitehouse 3 (TW3) method. From the results, the authors concluded that a machine learning approach based on ANN gives a more accurate solution for estimating age than the TW3 method for MRI.

There is another paper which aims at applying deep learning techniques in medical images for age estimation using hand X-ray images[31]. The proposed method presented in the paper is defined as a regression problem for estimating age with hand X-ray as input and age as output. The authors trained a regression model with hand X-ray images. The use of the deep learning framework known as Caffe is also demonstrated in the experiment. The results presented in the study suggest that the deep learning neural network approach can be used for estimating age from hand X-ray images.

Authors in article [50], design a framework based on deep learning models for estimating bone age. The framework is built in two steps. Firstly, they extract hand regions from x-ray images using Mask R-CNN to avoid other objects. Secondly, they applied a residual attention subnet on the network to focus on regions that are important in x-ray images and make predictions. This proposed method has been evaluated on a large bone age dataset and results suggest the importance of the proposed method over previous methods.

In one of the studies [26], the authors presented a method for age estimation designed on deep convolutional neural network framework. This method is evaluated on children between the age ranges 1 to 17 years old. It uses labelled and comprehensive orthopantomography data of 456 patients. In the proposed approach, firstly global fuzzy segmentation was done for labelling and then features were extracted using an intensity projection technique. Finally, the deep CNN model was designed to extract a large set of features. Results suggest that the developed methods can be used to classify images and can estimate age with high accuracy.

In article [38], the authors presented an automated system for bone age assessment from hand X-ray images. The proposed method uses a transfer learning concept for feature extraction. In this method firstly they preprocessed the hand X-ray images and then features were extracted by using pre-trained deep neural networks (ResNet-101 and AlexNet). Then, age range based on X-ray images was determined using classification models such as K-nearest neighbour (KNN), decision tree, support vector machine and linear discriminant. The results of this study suggested that features extracted from ResNet-101 are more effective and decision tree classifier gives better accuracy up to 100% among other classifiers.

Jaeyoung, Woong and Kyu-Hwan in [29], evaluate the applicability of deep learning algorithms for developing an automatic system for estimating chronological age by using convolutional neural networks. The database used for applying and evaluating model consist of panoramic x-ray images. Authors train the CNN model based on DenseNet-121 [24] to predict chronological age from x-ray images. The model was trained using two training strategies named curriculum learning strategy in which subjects were picked up to 11 year old and then include 12 -19 years or above, and random sampling strategy in which authors picked subjects from all age groups. The results of research suggests that a deep learning model based on curriculum learning strategy gives promising results for estimating age using dental x-ray images.

There is another study which focuses on estimating age from radiological data such as MRI volumes. This paper [52], presents a multi-factorial method for age estimation using MRI data for age range up to 25 years old. This proposed method is built based on deep convolutional neural networks. This multi-factorial age estimation is based on three anatomical sites such as teeth, calves and hands. From the result, it is concluded that the proposed method can be used to estimate the age up to the age range of 25 years.

In one of the research [33], authors have proposed a deep learning method for automatically estimating age by using fine tuned convolutional neural networks (CNN). For the development of the deep learning model authors picked up 1408 pelvic X-ray radiographs of individuals between age ranges 18 to 25 years. The performance of the model was accessed by comparing the obtained results with the age estimated using existing cubic regression models [33]. Authors calculated root mean square error (RMSE) and mean absolute error (MSE) for both models. The results of research shows that convolutional neural networks achieve better results compared to existing cubic regression models for estimating bone age based on pelvic radiography images.

In article [1], authors presented an age estimation method which is built using 3-Dimensional convolutional neural networks (3D-CNN) from T1-weighted MRI brain images. The proposed architecture consists of four convolution blocks having a 3D convolution layer, 3D batch-normalization layer, rectified linear unit (ReLU) activation layer and max pooling layer of which the last three layers are connected to form feature vectors. The output value is the predicted age. The experiment was conducted using Japanese MR Images. The results demonstrate that the proposed method performed age estimation efficiently as compared to other conventional methods.

In one of the studies [34], authors presented a review on the use of several deep learning contributions for medical image analysis, object detection, image segmentation, age estimation and other tasks from recent years. The authors mainly focus on studies with application areas such as breast, digital pathology, neuro, retinal, cardiac, abdominal, and musculoskeletal. This paper puts light on various challenges and also presents a critical discussion of various state-of-the-art approaches. The authors reviewed more than 300 articles and found that deep learning has a great impact on medical imaging analysis. Various research authors also put light on papers that use deep learning methods for age estimation from MRI images and X-ray images. Such as the paper presented by Stern et al. (2016) which proposed an age estimation method based on MRI using 2D regression CNN architecture. Another paper presented by Spampinato et al. (2017) reviews many deep learning approaches for the skeleton bone age estimation[34]. This survey provides knowledge about researchers related to deep learning in medical image analysis which could help us to find what work has been done in this area and which methods have been proposed by researchers for image analysis tasks. Based on this literature review, we were able to observe that many studies have been conducted on age estimation from X-ray images of bone, MRI, dental images and face images but we were unable to find any work related to age estimation from mammograms.

We have selected an experiment method to answer our research question. We have implemented several models to estimate the age based on the pectoral muscle of mammograms and observe how accurately these models are in estimating the age. We cannot conduct a literature review since there is no research done previously based on pectoral muscle to estimate the age. Furthermore, a case study is not suitable since this is not about an observation.

4.1 Experiment

In our experiment, we have used the U-net architecture to segment the pectoral muscle from the mammograms. Then we have extracted the features from the segmented image. First, we have extracted seven statistical features: mean intensity, standard deviation, entropy and four Haralick Features. Then we have trained the SVR and Random Forest to estimate the age and obtained the MSE and MAE values. Secondly, we have used two pre-trained deep learning models (Resnet50, Xception) to extract deep features from the segmented images. Then we have trained SVR and RF separately for each set of extracted features while observing the errors (MAE and MSE). We have conducted the same procedure for the whole MLO image and the breast region(BR) segmented image (Figure 4.1). In order to observe the effect of the increased number of handcrafted features in SVR and RF, we have concatenated the features from the PM segment and the BR (Figure 4.2). Thereby SVR and RF will be trained on 14 features.

In training the RF and SVR, our independent variables are handcrafted features and deep learning features taken from each region while the dependent variable is age.

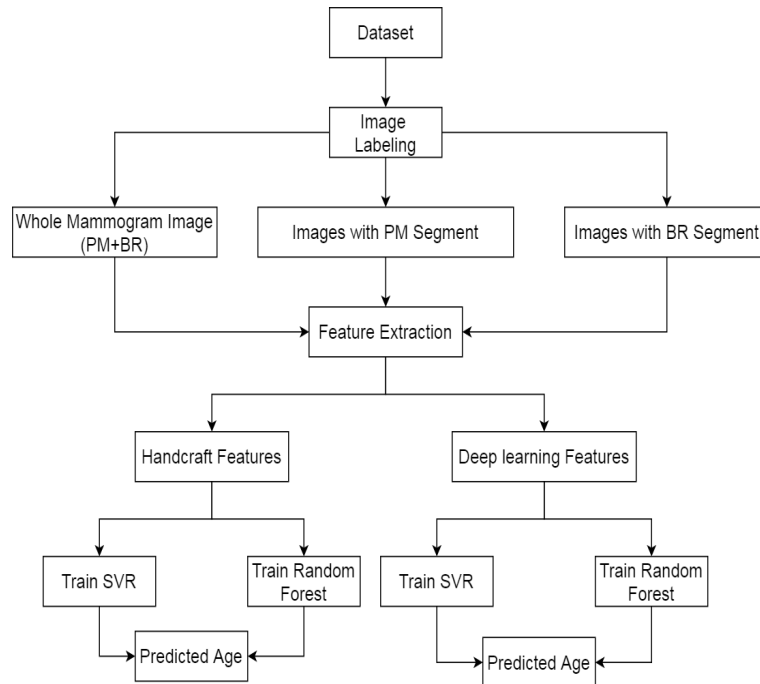


Figure 4.1: Workflow part 1 of Experiment

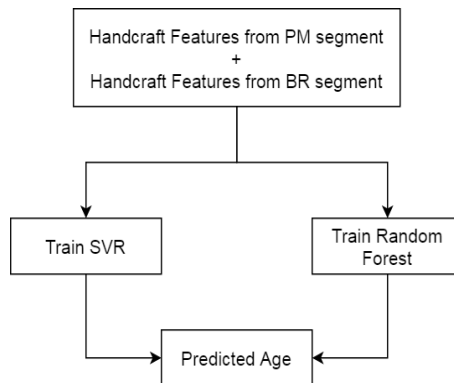


Figure 4.2: Workflow part 2 of Experiment

4.2 Database

We have used the publicly available 'Digital Database for Screening Mammography' dataset [22][32]. This is the only publicly available dataset of mammograms with age variables. This dataset consists of volumes such as normal, cancer and benign. Within these volumes, there are different cases which will represent various patients. The DDSM dataset consists of 2620 cases in 43 volumes. A "case" is a collection of images and information corresponding to one mammography exam of one patient. A "volume" is simply a collection of cases collected together for purposes of ease of distribution [22][32].

All the images were compressed as LOSSLESS JPEG which can only be extracted

using the software they have provided. Since the software is outdated we were unable to uncompress the images. Therefore, we had to collect the thumbnail images which were in png format for each case through web crawling and the age variable was stored in spreadsheets. The age variables were recorded in years.

The dataset included 1364 MLO images from the normal category, 1798 from the Cancer category and 1680 From the Benign category.

4.2.1 Ethical Aspects

We have selected a publicly available dataset which is used for the research purposes. It is public data and it is not related to any individual. According to the definition of GDPR (General Data Protection Regulation), our dataset is not considered as personal data belonging to any individual. Therefore, there will not be any ethical issues. Also, published dataset does not reveal any patient identity as well as during our experiment. Furthermore, we will use public MatLab libraries for implementation.

4.3 Software Environment Setup

We have used the Matlab® R2020a version in our experiment. We have installed the following toolboxes and features to carry out our experiment.

Table 4.1: Environment Setup Tools

Toolbox	Version
Computer Vision Toolbox	Version 9.2
Deep Learning Toolbox	Version 14.0
Image Processing Toolbox	Version 11.1
Statistics and Machine Learning Toolbox	Version 11.7
Deep learning Toolbox model for Xception Network	Version 20.1.0
Deep learning Toolbox model for ResNet-50 Network	Version 20.1.0

4.4 Dataset Cleaning

Once the dataset is downloaded we have checked the distribution of the Age variables in each folder (Figure 4.3). In which we were able to identify some outliers. These were incorrect values such as age=1. Therefore, we removed these records from the dataset.

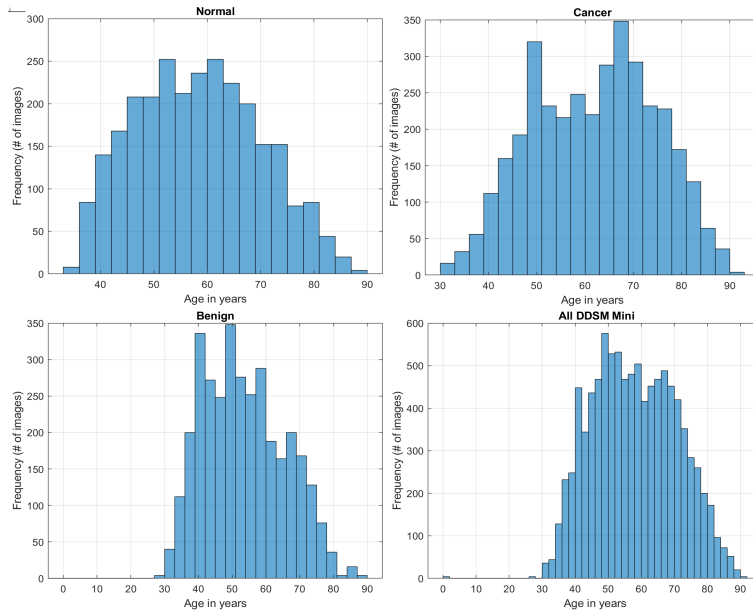


Figure 4.3: Age distribution of each category and of the over all samples (bottom-right)

4.5 Implementation

4.5.1 Data Labeling for U-net Segmentation

We used the Image Labeler application in Matlab Computer Vision Toolbox to label 120 MLO images. We randomly picked 40 images from each Normal, Cancer and Benign folder in which 20 images Right MLO and 20 Left MLO. As shown in ((Figure 4.4) we have labelled the mammogram into 3 sections: PM, BR and Background. Then we stored these labelled pixel data in a ground truth object to be used in the next step.

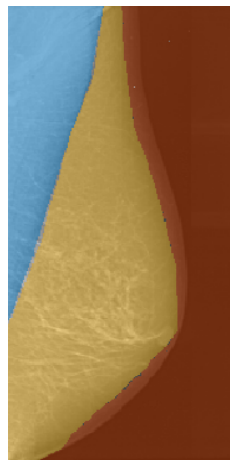


Figure 4.4: Labelled Image. Blue:PM; Yellow:BR; Red:Background

4.5.2 U-net Implementation

Since each labelled image has different sizes, we have resized the labelled images into 32 by 32 pixels. The reason for selecting this image size is to speed up the training process of segmentation. Then we have divided the labelled data into training and testing giving each set 115, 5 respectively. Matlab builtin function 'unetlayers' was used to access the U-net layers. This function required image size and the number of classes as input arguments. As the output, it returned the U-net network. We have used the default encoder and decoder depths which is 4((Figure 4.5). We have set some training options to improve the U-net segmentation (table 4.2) .

Solver algorithm was set to stochastic gradient descent with momentum (SGDM) and momentum to 0.9. In order to reduce the overfitting, we have set a regularization term called 'L2Regularization' to 0.0001 [10][39]. Gradient threshold values were set to prevent gradient explosion and to stabilize the training at higher learning rates [41]. The U-net model was then trained using the builtin function 'trainNetwork' by parsing the training data, U-net layers and the training options. Finally, the network was tested and evaluated using the 'semanticseg' and 'evaluateSemanticSegmentation' built-in functions. Results are reported in the next chapter. Trained U-net network((Figure 4.5) was stored to be used in the sections.

Table 4.2: Training options for U-net Segmentation

Input Argument	Value
solverName	'Sgdm' stochastic gradient descent with momentum
'Momentum'	0.9
'InitialLearnRate'	0.05
'LearnRateSchedule'	'piecewise'
'L2Regularization'	0.0001
'Shuffle'	'every-epoch'
'GradientThresholdMethod'	'l2norm'
'GradientThreshold'	0.05
'MaxEpochs'	150
'MiniBatchSize'	16

Before feeding the images to the trained U-net network for segmentation we resized them into 32 by 32. The function 'semanticseg' was used to segment each

region. This function took the image to be segmented and the network as input arguments and returned a 32 by 32 categorical object. Then we converted it back to unit8 and resized back to original size to obtain the probable regions. Before, extracting the features we had to revert back to the original image and mask it by the segmented image. Finally, all the images were resized to 224 by 224 to extract the features since images have different sizes.

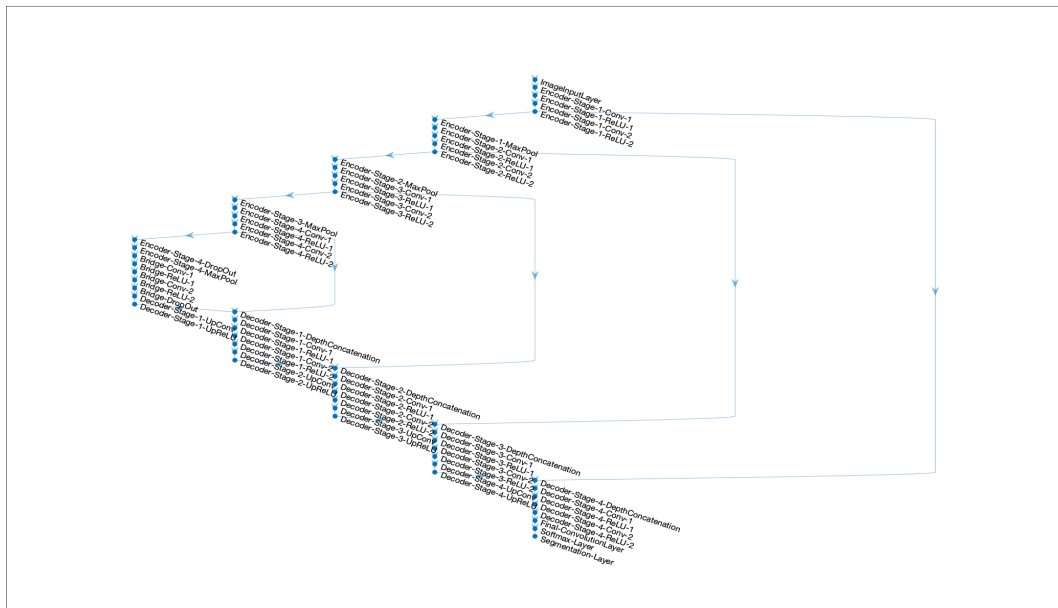


Figure 4.5: Trained U-net Network

4.5.3 Handcrafted Feature Extraction

Here we have focused on extracting 7 statistical features from the mammograms. First, we extracted the features from PM segmented images then BR and whole MLO mammograms respectively. In order to extract the features from all the MLO images, we stored them in image datastores.

The table 4.3 shows the built-in functions that were used to extract the features.

Table 4.3: Feature Extraction Built-in Functions

Feature	Function
Mean intensity	mean2(image)
Standard Deviation	std2(image)
Entropy	entropy(image)
Haralick Features	<ol style="list-style-type: none"> 1. glcms=graycomatrix(image) 2. graycoprops(glcms)

4.5.4 Deep Learning Feature Extraction

We have used three pre-trained deep learning models to extract the deep features from the images. In order to extract the deepest features, we activated the layer before the classification layer in each model.

4.5.4.1 ResNet50

Resnet50 is a 50 layer deep network. We have made use of the Matlab pre-trained renet50 network for our experiment. This network was trained on more than a million images from ImageNet database [6].

First, we loaded the pre-trained network into our workspace to analyze the network. Then we observed the input layer size, the image size to be fed into the network, and the layer names. In ResNet50, input image size is 224-by-224. Therefore, we augmented our images by resizing them according to the input layer size. Then we specified the layer to be activated which is a fully connected layer('fc1000') with 1000 features. The builtin 'activation' function was used to activate the layer. The function input arguments were network, augmented image data store, and the activation layer. It returned features as rows and then we stored them in an object for the next step.

4.5.4.2 Xception

Xception is a 71 layer deep CNN. We have used the pre-trained Xception network for extracting the deep features from mammograms. This network was trained on more than a million images from ImageNet database [7].

We have followed the same steps as in Resnet to extract the features. First, we loaded the pre-trained network into our workspace to analyze the network. Then we observed the input layer size, the image size to be fed into the network, and the layer names. In Xception, input image size is 299-by-299. Therefore, we augmented our images by resizing them according to the input layer size. Then we specified the layer to be activated which is a fully connected layer('predictions') with 1000 features.

The builtin 'activation' function was used to activate the layer. The function input arguments were network, augmented image data store, and the activation layer. It returned features as rows and then we stored them in an object for the next step.

4.5.5 Fitting models for Age Estimation

We have trained the SVR and RF for estimating the age from the features. In order to evaluate the models, we have used the Builtin functions 'loss' to achieve MSE and 'mae' to achieve MAE.

4.5.5.1 Random Forest

The handcrafted and deep features were loaded as tables and treated the missing values as 'NA'. Then removed the rows which contained the missing values by using the 'rmmissing' function. First, we randomly selected an equal number of images from each category(Normal, Cancer, Benign). The sample size was 1364 from each folder which in total 4092(1364x3) for training the model. Then the dataset was randomly divided into 70% training and 30% testing.

Matlab has a builtin function called 'fitrensemble' which we used for fitting the RF model [5]. This function uses the same technique as in RF in fitting the model based on predictors/features by creating decision trees [2]. We have passed the training features, training age labels, learner template and hyperparameter optimization values (Figure 4.6).

```

% Hyperparameter optimization process (requested from an individual expected improvement plus)
t = templateTree('Reproducible',true);
mdl = fitrensemble(Xtrain,Ytrain,'Method','Bag','OptimizeHyperparameters',...
    {'NumLearningCycles','MinLeafSize','MaxNumSplits'},'Learners',t,...
    'HyperparameterOptimizationOptions',struct('AcquisitionFunctionName','expected-improvement-plus'));

```

Figure 4.6: Function to train RF

We have optimized the model by setting the 'OptimizeHyperparameters' argument by passing the parameters to be optimized. By setting this argument we were able to train the model to find hyperparameters that minimize five-fold cross-validation loss by using automatic hyperparameter optimization [5]. We have used the random seed function 'rng' and 'HyperparameterOptimizationOptions' value to 'expected-improvement-plus' for reproducibility. Also, we have set the value of 'Reproducible' to true in tree learners.

4.5.5.2 SVR

The features were loaded as tables and treated the missing values as 'NA'. Then removed the rows which contained the missing values by using the 'rmmissing' function. First, we randomly selected an equal number of images from each category(Normal, Cancer, Benign). The sample size was 1364 from each folder which in total 4092(1364x3) for training the model. Then the dataset was randomly divided into 70% training and 30% testing.

Matlab has a builtin function called 'fitrsvm' which we used for fitting the SVR model [4]. We have optimized the model by setting the 'OptimizeHyperparameters' argument by passing the parameters to be optimized. By setting this argument we were able to train the model to find hyperparameters that minimize five-fold cross-validation loss by using automatic hyperparameter optimization[2]. We have used the random seed function 'rng' and 'HyperparameterOptimizationOptions' value to 'expected-improvement-plus' for reproducibility. Since we do not know much about the data, we have set the kernel function to 'gaussian'. Furthermore, we have set the standardize argument to true to standardize the predictors. Train model is shown as follows(Figure 4.7).

```
mdl = fitrsvm(Xtrain,Ytrain,'KernelFunction','gaussian',...  
  'OptimizeHyperparameters',{'BoxConstraint','KernelScale','Epsilon','Standardize'},...  
  'HyperparameterOptimizationOptions',struct('AcquisitionFunctionName',...  
  'expected-improvement-plus'));
```

Figure 4.7: Function to train SVR

5.1 U-net Segmentation

Once the U-net model is trained we have evaluated the model using the test set which included 5 labelled images. The function 'evaluateSemanticSegmentation' returned the data matrices as shown in table 5.1.

Table 5.1: Segmentation Evaluation Metrics Results

Global Accuracy	Mean Accuracy	Mean Iou	Weighted IoU	Mean BFScore
0.97079	0.96549	0.9	0.92911	0.7231

Figure 5.1 shows the semantic segmentation for the given image. Since the U-net model is trained on 32 by 32 image size, it returns in the same size. As mentioned in the implementation we have reverted back to the original image and segmented the regions as shown in figure 5.2. We will use segmented regions shown on figure 5.2 for the feature extraction.

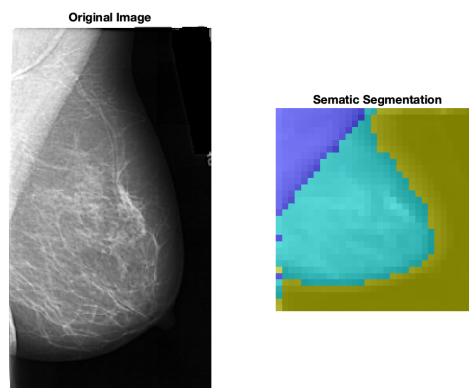


Figure 5.1: An Image of Sementic Segmentation

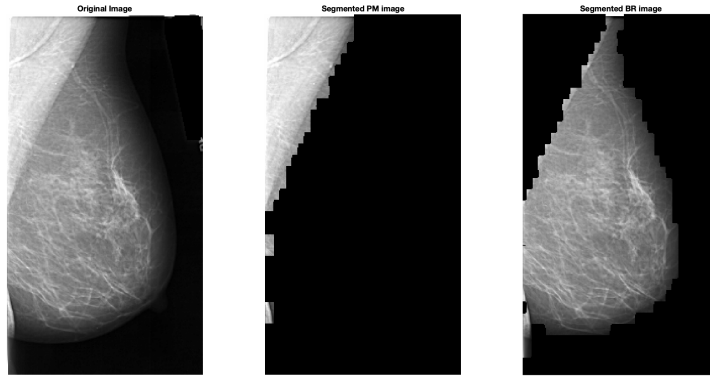


Figure 5.2: An Image of Segmented Regions

As seen in table 5.1, we can observe that both accuracy levels are over 90% as well as the IoUs. BF score is over 72% for the 5 images. As seen in the figure 5.1 and 5.2, we can see that PM and BR are segmented accurately enough to extract the features from each region. Therefore, we can consider U-net as a good model in segmenting the regions of Mammograms. Furthermore, we can assure that the U-net will segment the images accurately as possible to be used in the further experiment.

5.2 Handcrafted Features

5.2.1 PM Segmented Image Features

After the PM segmentation, we extracted the handcrafted features from the PM segmented images. Then we passed these features to train SVR and RF. The accuracies of these models are evaluated by calculating MAE and MSE. The results are shown in the table 5.2. MAE values are calculated in years and MSE is calculated in squared years. We can see that the MAE values for both models are quite similar. But in the RF model MSE value is a bit higher than the SVR.

Table 5.2: Handcrafted+PM Features Results

Model	MSE(Years) ²	MAE(Years)
SVR	161.7684	10.7066
RF	162.1853	10.7104

By looking at the fitted and actual plots of each trained model, it can be seen that both SVR and RF models estimated age values are closer to the fitted line. Furthermore, estimated age values are scattered between age 50-60 for both models.

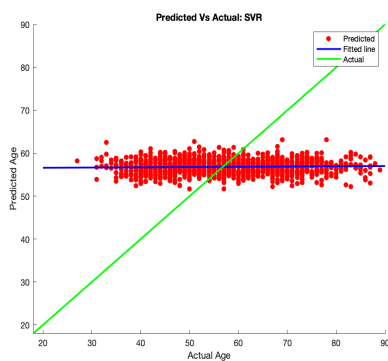


Figure 5.3: SVR with hand-crafted+PM Features

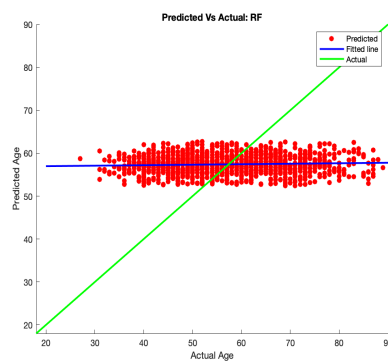


Figure 5.4: RF with hand-crafted+PM Features

5.2.2 BR Segmented Image Features

We have extracted the handcrafted features from the BR segmented images to train the SVR and RF models. The results are shown in table 5.3. It can be observed that both models get MAE values around 10(Years). However, the MSE values are different for both models. SVR obtains less value as compared to RF.

Table 5.3: Handcrafted+BR Features Results

Model	MSE(Years) ²	MAE(Years)
SVR	153.4198	10.1725
RF	155.3585	10.3212

If we look at the graph of trained models, we can see that the SVR model performs better compared to the RF model. The predicted ages by the RF model are much closer to the fitted line. On the other hand, it is observed that in SVR some values are closer to the line as compared to the RF model. Furthermore, estimated age values are scattered between age 40-80 in SVR model while on RF it is between 45-70.

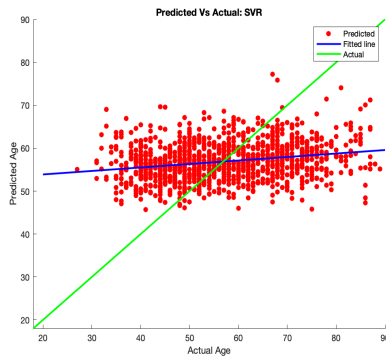


Figure 5.5: SVR with hand-crafted+BR Features

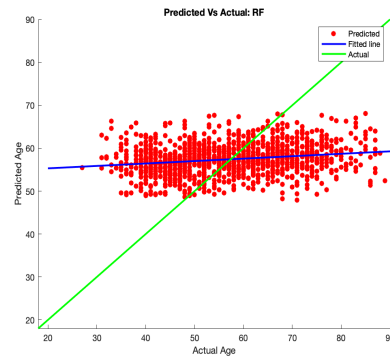


Figure 5.6: RF with hand-crafted+BR Features

5.2.3 PM+BR Image Features

We have extracted features from the BR region and PM region images as discussed in chapter 4. We have concatenated the features(14 features) obtained from both types of images and then passed it to train our models. The obtained results are shown in the table 5.4. It can be observed that MAE value for RF has lesser error value than SVR. However, they do not differ that much comparatively. SVR got MAE around 10.2684(years) while RF achieved MAE around 10.2687(years). In comparison to the MSE values of both models, RF has achieved a lesser error value than RF.

Table 5.4: Handcrafted+PM+BR(14 Features) Results

Model	MSE(Years) ²	MAE(Years)
SVR	153.9969	10.2684
RF	152.9581	10.2687

From the plotted graphs for trained models, it can be observed that in both models predicted values are closer to the fitted line for most of the cases. Furthermore, estimated age values are scattered between age 40-80 for both models.

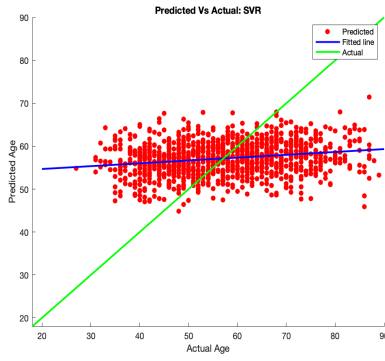


Figure 5.7: SVR with handcrafted+PM+BR Features

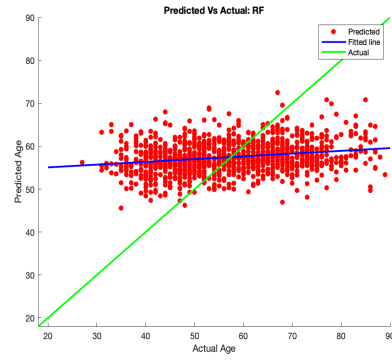


Figure 5.8: RF with handcrafted+PM+BR Features

5.2.4 Handcrafted with whole MLO(BR+PM)- 7 Features

We have extracted the handcrafted features from BR and PM regions by feeding the entire MLO to train SVR and RF models. The results of the evaluation are shown in the table 5.5. It can be seen that the MAE value for both models is around 10(Years). However, the MSE value for SVR is less as compared to the RF model.

Table 5.5: Handcrafted+PM+BR(whole MLO) Results

Model	MSE(Years) ²	MAE(Years)
SVR	152.3889	10.2286
RF	154.6714	10.3697

If we look at the graph plotted for both models, it can be observed that for the RF model predicted values of age is closer to the fitted line in many cases. But in the SVR model, there are fewer values close to the line as compared to RF. Furthermore, estimated age values are scattered between age 45-75 for SVR model while on the other hand for RF between 45-70.

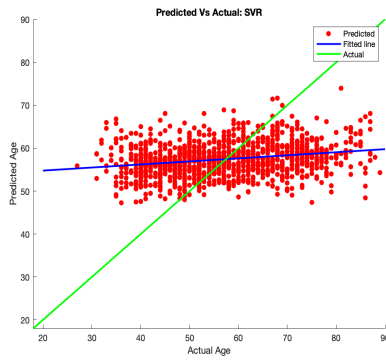


Figure 5.9: SVR with hand-crafted+PM+BR(whole MLO) Features

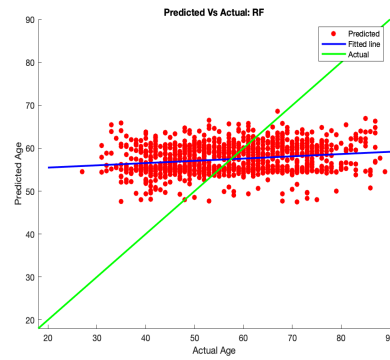


Figure 5.10: RF with hand-crafted+PM+BR(whole MLO) Features

5.3 Deep Learning Features

5.3.1 Resnet50 PM Segmented Image Features

We extracted the deep features from PM segmented images using Resnet50 and trained the SVR and RF models. The MAE and MSE values for both models are shown in table 5.6. It is seen that MAE values for both models are around 10(years). However, the MSE value for RF is less than SVR.

Table 5.6: ResNet50+PM Features Results

Model	MSE(Years) ²	MAE(Years)
SVR	150.6311	10.1495
RF	149.6521	10.1689

If we look at the graph of trained models, we can see that the predicted ages by the RF model are much closer to the fitted line. On the other hand, it is observed that in SVR some values are further to the line as compared to the RF model. Furthermore, estimated age values are scattered between age 40-70 in SVR model while on RF it is between 45-70.

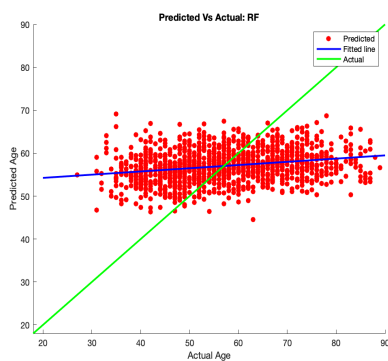


Figure 5.11: SVR with Resnet50+PM Features

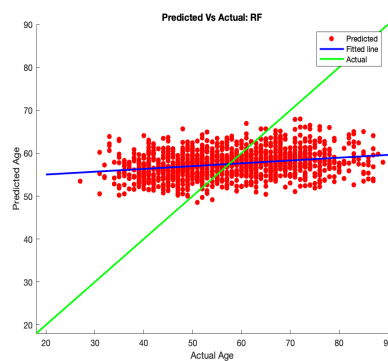


Figure 5.12: RF with Resnet50+PM Features

5.3.2 Resnet50 BR Segmented Image Features

We have extracted features from the BR segmented images and used them for training the SVR and RF models. The evaluation results of both models are shown in the table 5.7. It can be seen that MSE and MAE value for the SVR model is less than the RF model. SVR performs better than the RF model.

Table 5.7: Resnet50+BR Features Results

Model	MSE(Years) ²	MAE(Years)
SVR	112.3545	8.5248
RF	129.8398	9.3602

From the plotted graphs for trained models, it can be seen that predicted values for both models are further to the fitted line in most cases. However, the SVR model shows more good predictions as compared to the RF model. Furthermore, estimated age values are scattered between age 30-90 in SVR model while on RF it is between 40-75.

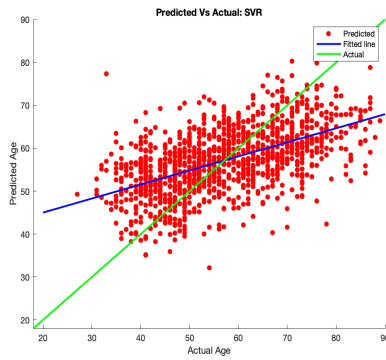


Figure 5.13: SVR with Resnet50+BR Features

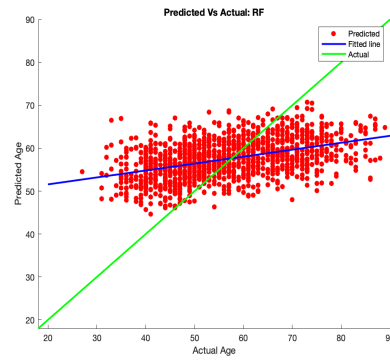


Figure 5.14: RF with Resnet50+BR Features

5.3.3 Resnet50 PM+BR Image Features

We have extracted deep features from the BR region and PM region images using the Resnet50 model as discussed in chapter 4. We have concatenated the features obtained from both types of images and then passed it to train our models. The obtained results are shown in the table 5.8. It can be seen that the MAE value for the SVR model is around 8 (Years) and for RF it is around 9(Years) which is higher than SVR. Similarly, the MSE value of SVR is less than the RF model.

Table 5.8: Resnet50+PM+BR(whole MLO) Features Results

Model	MSE(Years) ²	MAE(Years)
SVR	111.8621	8.4656
RF	128.2817	9.2427

If we look at the graph for trained models, it can be seen that predicted values for SVR model are further to the fitted line in most cases as compared to RF. The SVR model shows more good predictions as compared to the RF model. Furthermore, estimated age values are scattered between age 35-85 in SVR model while on RF it is between 40-75.

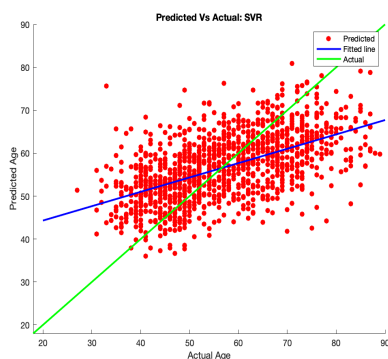


Figure 5.15: SVR with ResNet50+PM+BR(whole MLO) Features

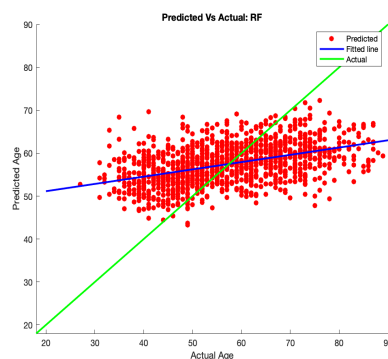


Figure 5.16: RF with ResNet50+PM+BR(whole MLO) Features

5.3.4 Xception PM Segmented Image Features

We have extracted deep learning features from Xception architecture and used these features to train SVR and RF models. The results are shown in table 5.9. The obtained results show that both models have quite similar MAE values around 10 (years). However, the MSE value for RF is a bit less than SVR.

Table 5.9: Xception+PM Features Results

Model	MSE(Years) ²	MAE(Years)
SVR	150.0633	10.1490
RF	149.8927	10.2465

If we look at the graph of trained models, we can see that the predicted ages by the RF model are much closer to the fitted line. On the other hand, it is observed that in SVR some values are further to the line as compared to the RF model. Furthermore, estimated age values are scattered between age 40-70 in SVR model while on RF it is between 50-65.

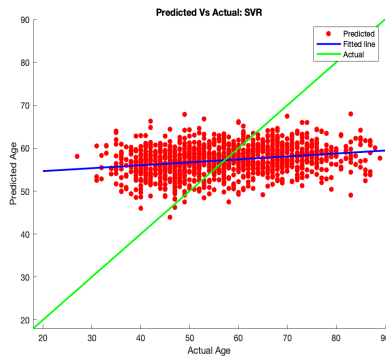


Figure 5.17: SVR with Xception+PM Features

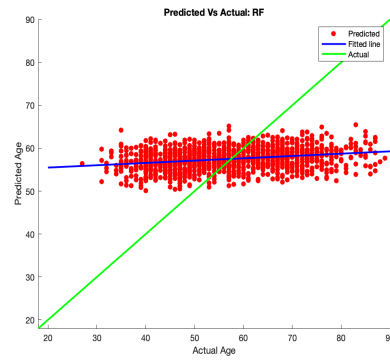


Figure 5.18: RF with Xception+PM Features

5.3.5 Xception BR Segmented Image Features

We have extracted the features from BR segmented images by using the Xception model and use these features to train SVR and RF models. The results obtained for both models are shown below in table 5.10. It can be seen that the MSE value for the SVR model is less than the RF model. On the other hand, MAE values for both models are around 9.

Table 5.10: Xception+BR Features Results

Model	MSE(Years) ²	MAE(Years)
SVR	132.8169	9.2287
RF	134.6028	9.5892

From the graphs of trained models, it can be observed that the SVR model shows better performance than the RF model. The points are much closer to the fitted line in SVR. However, the RF model also gives good predictions for some cases but not as compared to the SVR while having estimated age values closer to the fitted line. Furthermore, estimated age values are scattered between age 30-80 in SVR model while on RF it is between 45-70.

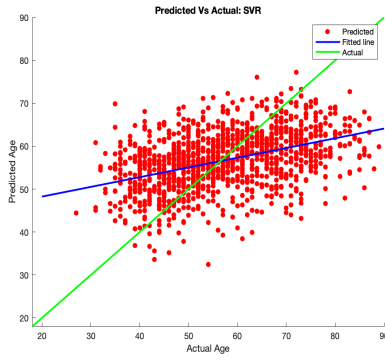


Figure 5.19: SVR with Xception+BR Features

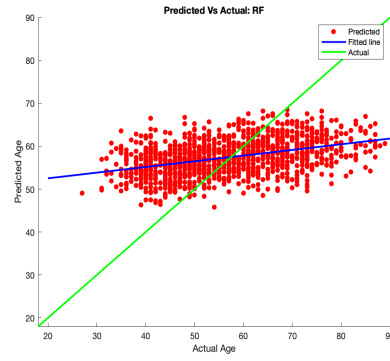


Figure 5.20: RF with Xception+BR Features

5.3.6 Xception PM+BR Image Features

We have extracted deep features from the BR region and PM region images using the Xception model as discussed in chapter 4. We have concatenated the features obtained from both types of images and then passed it to train our models. The obtained results are shown in the table 5.11. It can be seen that the MAE value for the SVR model is around 8 (Years) and for RF it is around 9 (Years) which is higher than SVR. Similarly, the MSE value of SVR is less than the RF model.

Table 5.11: Xception+PM+BR(whole MLO) Results

Model	MSE(Years) ²	MAE(Years)
SVR	126.7388	8.9866
RF	134.3715	9.5016

From the graphs plotted for SVR and RF trained models, it can be seen that predicted values for SVR model are further to the fitted line in most cases as compared to RF. The SVR model shows more good predictions as compared to the RF model. Furthermore, estimated age values are scattered between age 30-850 in SVR model while on RF it is between 40-70.

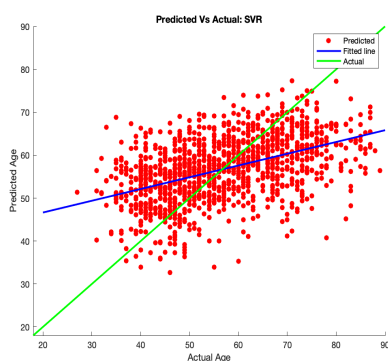


Figure 5.21: SVR with Xception+PM+BR(whole MLO) Features

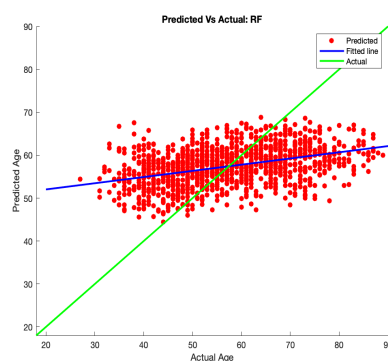


Figure 5.22: RF with Xception+PM+BR(whole MLO) Features

5.4 Summary

In table 5.12 we have compared the MAE values between each model we have trained. Comparatively, SVR achieves lesser MAE values in estimating age based on PM. However, RF also achieves MAE much closer to SVR when estimating age based on PM and we cannot see much difference between these two. Among the handcrafted features and the deep learned features, deep learning feature models achieve lower MAE values in estimating age based on PM region. Between ResNet50 and Xception features, Xception with SVR achieved the lowest MAE for estimation from the PM region.

As we compared the MAE values achieved from BR, we can see that SVR achieved lesser error values than RF. Between handcrafted and deep features in BR, deep feature models achieve significantly lesser MAE values. Furthermore, Resnet50 with SVR achieves the lowest MAE value in estimating age based on BR.

When we compare the whole MLO view(PM+BR), deep learning feature models achieve significantly less MAE values as compared to the handcrafted features. In comparison to the SVR and RF in the whole MLO view, SVR achieves much lesser MAE value than RF. ResNet50 with SVR achieves the lowest MAE value in estimating age based on the whole MLO.

When we compare handcrafted feature models by feeding the whole MLO (7 features) and PM+BR images separately (14 features), we observed SVR achieves lesser MAE than RF for only 7 features. When SVR and RF were having 14 handcrafted features, SVR achieved lesser MAE value than RF. However, these models differ less in between compared to the other models. The lowest MAE we have achieved for age estimation based on Mammograms was ResNet50 with SVR for the features we extracted in the whole MLO by having only 8.4656.

Table 5.12: MAE Results Summary

Features	SVR	RF
handcrafted+PM	10.7066	10.7104
handcrafted+BR	10.1725	10.3212
handcrafted+PM+BR(14 features)	10.2684	10.2687
handcrafted+PM+BR(whole MLO)	10.2286	10.3697
ResNet50+PM	10.1495	10.1689
ResNet50+BR	8.5248	9.3602
ResNet50+PM+BR(whole MLO)	8.4656	9.2427
Xception+PM	10.1490	10.2465
Xception+BR	9.2287	9.5892
Xception+PM+BR(whole MLO)	8.9866	9.5016

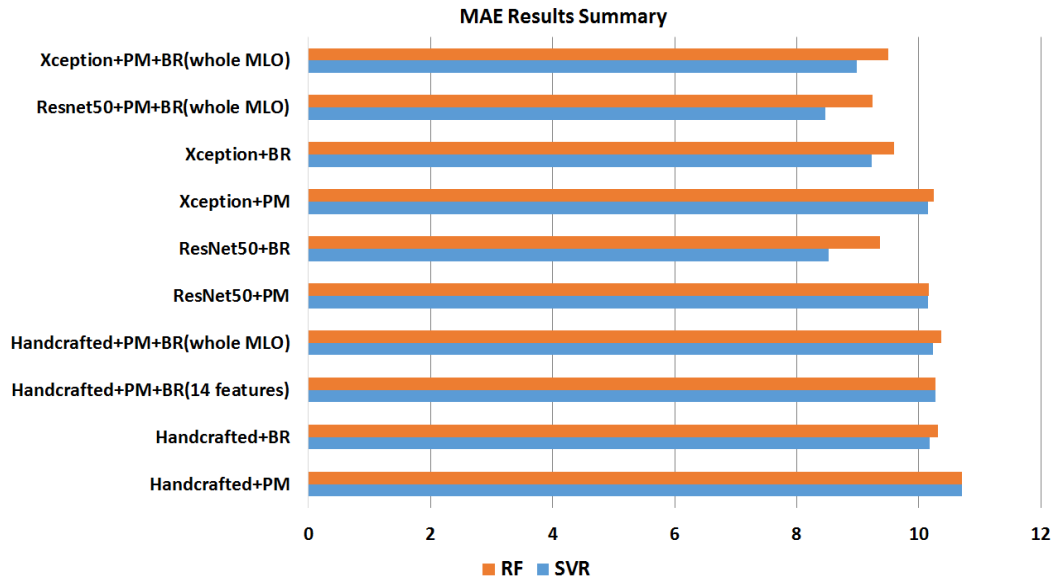


Figure 5.23: MAE Results Graph

The table 5.13 shows the MSE values we have obtained from each model. These MSE values are reported in square years.

In comparison to the MSE values we have achieved for models trained in the PM region, deep learning feature models have less error than handcrafted features. handcrafted features achieved MSE over 161 for the PM region while deep feature

models achieved MSE around 150 or less. ResNet50 with RF achieved the lowest MSE for age estimation based on PM region.

The models trained on extracted features from BR achieved noticeably lesser values compared to the MSE values from the PM region. In comparison to the deep feature models and handcrafted feature models in BR, deep feature models achieved much lesser MSE values. ResNet50 with SVR achieved the lowest MSE for age estimation based on BR having only 112.35 in squared years.

Once we compared the MSE values from the whole MLO, we can observe that deep feature models achieved much lesser values as compared to handcrafted feature models. SVR achieved lower MSE values as compared to the RF for the features extracted from the whole MLO. ResNet50 with SVR achieved the lowest MSE for the whole MLO by having only 111.86 and this was also the lowest MSE we achieved in age estimation based on the Mammograms.

Table 5.13: MSE Results Summary

Features	SVR	RF
handcrafted+PM	161.7684	162.1853
handcrafted+BR	153.4198	155.3585
handcrafted+PM+BR(14 features)	153.9969	152.9581
handcrafted+PM+BR(whole MLO)	152.3889	154.6714
ResNet50+PM	150.6311	149.6521
ResNet50+BR	112.3545	129.8398
ResNet50+PM+BR(whole MLO)	111.8621	128.2817
Xception+PM	150.0633	149.8927
Xception+BR	132.8169	134.6028
Xception+PM+BR(whole MLO)	126.7388	134.3715

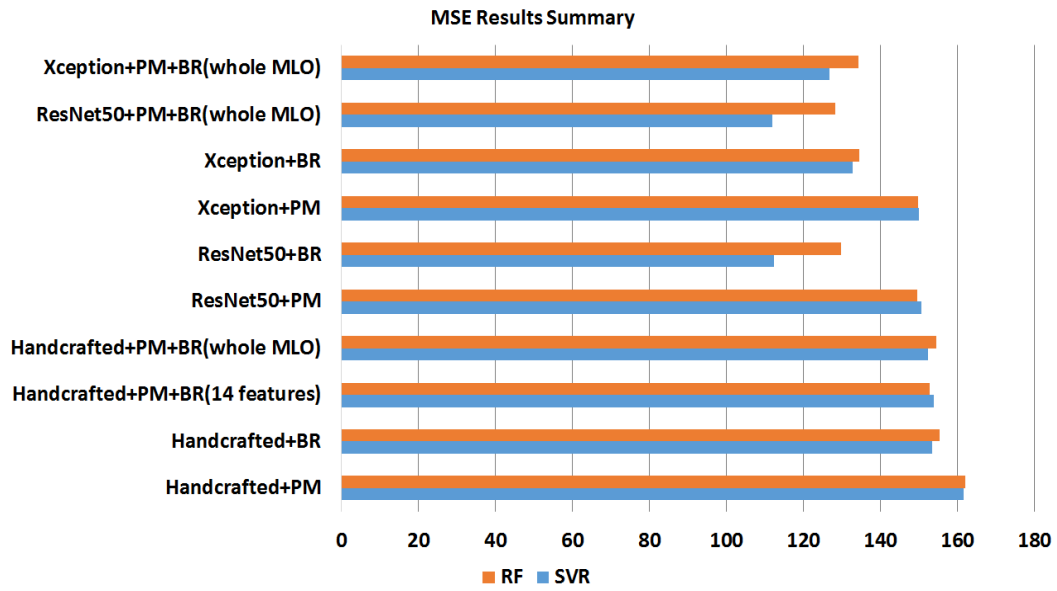


Figure 5.24: MSE Results Graph

5.4.1 Significance Test

Since we were able to notice the similarity between errors in our best 4 models, we decided to perform significance tests to see if our models differ significantly between BR and PM.

Models compared:

Model1: ResNet50+BR (RF)(MAE=9.3602)

Model2: ResNet50+PM+BR (RF)(MAE=9.2427)

Model3: ResNet50+BR (SVR) (MAE=8.5248)

Model4: ResNet50+PM+BR (SVR) (MAE=8.4656)

We perform two significant tests, Model1 against Model2(best RF models) and Model3 against Model4(best SVR models). By this, we could conclude whether there is a positive shift in the median of observed absolute errors of BR+PM models compared to the median MAEs of BR models. We have used Wilcoxon's rank-sum test which is a non-parametric test as our significance test.

Test1:

H₀(null hypothesis): There is no positive shift in the median (improvement) of observed absolute errors from Model1 to Model2 at the 1% or 5% significance levels.

H₁(alternative hypothesis): There is a positive shift in the median (improvement) of observed absolute errors from Model1 to Model2 at the 1% or 5% significance levels. In other words, the median of Model1 is less than the median of Model2.

Table 5.14: Results of Test1

Alpha	P-Value	H
0.01	0.7099	0
0.05	0.7099	0

H = 1 indicates rejection of the null hypothesis at the 100 * Alpha% significance level.

H = 0 indicates a failure to reject the null hypothesis at the 100 * Alpha% significance level.

Test2:

H₀(null hypothesis): There is no positive shift in the median (improvement) of observed absolute errors from Model3 to Model4 at the 1% or 5% significance levels.

H₁(alternative hypothesis): There is a positive shift in the median (improvement) of observed absolute errors from Model3 to Model4 at the 1% or 5% significance levels. In other words, the median of Model3 is less than the median of Model4.

Table 5.15: Results of Test2

Alpha	P-Value	H
0.01	0.6401	0
0.05	0.6401	0

H=1 indicates rejection of the null hypothesis at the $100 * \text{Alpha}\%$ significance level.

H=0 indicates a failure to reject the null hypothesis at the $100 * \text{Alpha}\%$ significance level.

The interpretation (statistical inference):

The Wilcoxon's test at both significance levels (i.e., alpha 1% and alpha 5%), strongly suggests that there is no statistical evidence of the results of both models (RF, SVR) to conclude that there is a positive shift (improvement) in the median of observed absolute errors (AEs) of the breast region (BR) + PM compared to the median AEs of the BR alone.

6.1 Discussion

Our main research question was, how accurate it is to consider the pectoral muscle segment in mammograms to estimate age. This research question was formed based on the research paper [12] which highlighted the association between the mean intensity of pectoral muscle and the age. Therefore, we wanted to investigate if the pectoral muscle (PM) features can be considered in estimating the age from mammograms. For the purpose of comparison, other regions of the mammogram were considered (breast region (BR)).

In order to achieve our aim of the study, we have implemented an experiment to estimate age from PM and compared the error measures with BR and the whole mammogram. In which the whole mammogram image contains both PM and BR. The compared errors, MAE and MSE values are reported in chapter 5. The lowest MAE we were able to achieve was 8.4656 years which was for the ResNet50 features with SVR model and the features were extracted from the whole mammogram. This model also achieved the lowest MSE in our experiment.

As compared to the models trained on the PM region, much fewer errors were observed for the BR. This indicates that considering BR is much accurate in estimating the age than the PM region. However, the models that consisted of both regions (whole mammogram) achieved the least errors. This indicates that the PM also contributes some features/feature that help to estimate the age. One of these contributing features from PM can be the mean intensity of PM since in previous studies we were able to see the association between these two [12]. Although, the handcrafted feature models that included the mean intensity of PM were not as accurate as BR featured models. Another explanation for this may be that BR contributes more features that could help in estimating the age from Mammograms. Even though mean errors indicate this argument, according to the significant test we performed on section 5.4.1, we were unable to obtain statistical evidence to support that BR models and whole mammogram models perform differently.

Furthermore, when we compared the feature extraction method, models with deep learning features achieved lesser MAE and MSE values than handcrafted feature models. This indicates that the deep features are more accurate in estimating the age than handcrafted features. However, we were able to observe less MSE

and MAE values for ResNet50 feature models than Xception feature models. The pre-trained Xception model was a 71 layer deep model whereas ResNet is 50 layers deep. In most of the age estimation studies, the pre-trained Xception model achieved better results as compared to ResNet50 [40]. However, in our experiment it was different and we were unable to see any patterns that might cause this difference.

One can argue that the number of handcrafted features is not enough to estimate the age. In order to defend this argument, we increased the number of features by concatenating the BR and PM features to have 14 features. Once we compared the error values obtained from this model with values from the whole mammogram which had 7 features from both regions, the values were differing from small error rates. On the one hand, MSE with SVR differed from 1.608 while giving many benefits to the 7 feature model, and on the other MSE with RF differed from 1.7133 by benefiting the 14 feature model. Another explanation for this difference can be the characteristics of SVR and RF. Therefore, we can not completely ignore the fact that handcrafted features might not be enough or the fact the number of selected handcrafted features are enough in estimating the age from mammograms.

To the best of our knowledge, there are no previous studies conducted in relation to age estimation from mammograms and we cannot guarantee that the models we fitted achieved the best possible results in estimating the age from mammograms. However, when we studied the related work in age estimation from images, we were able to observe MAE values around 5.35 -14.56 in years for different trained models in different areas of study such as age estimation from the face images [35]. Also, there were much lesser MAE values with proposed methods, fine-tuning the features, and using the transfer learning concept [48][50]. In one of the studies, they have evaluated automatic age estimation from face images by comparing many aspects such as feature extraction method and dataset size [40]. In this paper, they have reported MAE values ranging from 2 years to 14 years by discussing how to improve these models. Therefore, we can assume that if we could also use any of those concepts such as fine-tuning the layers for transfer learning, we could be able to improve our results. But this was not within our scope of the study.

We have taken equal random samples from each subset (3 types of mammograms) which we believe would remove any bias that may be caused by these types. If we divide the data set into the three types and develop the model separately then the samples will be too few to deploy deep machine learning (it is known that deep learning requires large data samples). We need to provide a more general model that could be applied regardless of the cancer status. Restricting the model to certain disease status would require extra work/information (disease status) at clinics which may or may not be attainable. The age estimation is meant to improve 'assist other risk factors' (e.g., BMI, mammography percent density, parity, HRT) prediction of the disease status, therefore, it is not feasible to classify our age prediction models based on disease status that the final aim is to help predict.

As seen in our plots in the results section, many estimated values are scattered around the middle of the plot. This might be due to the fact in our dataset, many

age values are between the ages 40-75 [Figure 4.3]. In mammogram datasets, mainly patients are from age 40-75. Therefore, the models will have more data points around that age range to be trained and tested. This might cause new data points that differ from this range to have higher error rates than the new data points which fall within this age range. We could estimate the ages by dividing the data into age groups before training the models as in many other studies [15][36]. As we mentioned in the introduction, one could consider these methods in estimating age to fill the null values in the mammogram databases. However, it is recommended that to improve the results before filling the null values since age is a crucial factor in breast cancer predictions. If one could improve the estimated age by considering the results from our study, that would be more sustainable than deleting the empty records.

Our research would not only be beneficial in filling the null values of databases. As for an example, there are women in underdeveloped countries who do not have a birth certificate (i.e, women living in the Sahara or in rural areas). In which one could use an improved version of our models to estimate the age to check whether they belong to the risk group. Furthermore, our model will be useful for those researchers who have access to patients' mammograms but not to their identity (due to GDPR) to retrieve their ages, therefore we provide a way to automatically estimate the age.

6.2 Validity Threats

6.2.1 Internal Validity

Internal validity is concerned about how well the research has been conducted to answer the research question. In our research, this validity is achieved by selecting the model that could segment the pectoral muscle region more accurately. Also, the choice of models for estimating the age was concerned in order to answer the research question. Moreover, we also filter data and remove outliers from the dataset. We have also considered the reproducibility of the results by using appropriate Matlab functions such as 'rng' which helps us to produce the same random sample data in training and testing the models.

A potential internal validity threat can be that image resizing parameter. In handcrafted images and ResNet50 resizing parameter was 224 by 224 while in Xception it was 229 by 229. For the pre-trained deep learning models Xception and ResNet50 these parameters were fixed by default for the input layer size. This was one of the limitations we faced during the implementation. By taking the same resizing parameter for handcrafted images as in ResNet50 we tried to eliminate this threat. However, we could not extend the experiment to implement this for Xception since the time taken to train these models and evaluate them were a limitation.

6.2.2 External Validity

External validity refers to the generalization of the obtained results from the experiment. In our research, we took the real-world data of mammogram images. Also, we are taking all volumes into consideration from the dataset in sample selection as

well as selecting them randomly, so that implementation and results could be more generalized.

6.2.3 Conclusion Validity

Conclusion validity threat is concerned about the accuracy of the results we conclude from the research. In our research, we treated this validity by following the proper steps in implementation and selecting the right method to conduct the research. Also, we selected performance metrics in order to interpret the results correctly and reached the conclusions.

7.1 Conclusion

In our thesis, we implemented an experiment to estimate age based on the pectoral muscle(PM) region in mammograms and to address its accuracy. Despite the association we have seen between PM and the age from previous studies, we were able to observe that the breast region(BR) is more accurate in estimating the age compared to PM by having least MAE and MSE values in its models. Moreover, we were able to observe that handcrafted feature models are not as accurate as deep feature models in estimating the age. However, in our experiment, the least observed error value for MAE was around 8.4656 years. This generally means that if a patient is around 50 years, our best model will estimate that age between the boundaries 50 ± 8.4656 . Therefore, we will conclude our thesis by stating that age estimation solely based on PM is not as accurate as we expected. However, one could consider BR or the whole mammogram in estimating the age based on the mammograms.

7.2 Future Work

This thesis acts as a proof of concept and to protect the idea for further investigation. Therefore, there are many possible future works we could discuss. In this section, we have mentioned a few of the future studies out of many that could be conducted using this concept.

In future studies, one could try to improve the results by getting access to a large full resolution dataset (as we used thumbnails in this thesis). By using higher resolution images they would be able to access more image features in the mammograms as well as remove defects in thumbnail images. Also, it may improve the results of U-net segmentation.

We could consider our research as another imputation method. This research could potentially help in predicting disease status in case there are some missing age entries. Testing this would require adding another dataset with disease status such as MINI-MIAS database (does not have age variable), which is worth investigating.

As we mentioned in the discussion that our age variable is mostly distributed around 40-75 years, it might be causing the mean errors to differ. Therefore, we think we could be able to improve the results in future work if we divide the data

between age groups. Furthermore, in future work others could consider increasing the number of different handcrafted features rather than the features we have discussed here: mean intensity, standard deviation, entropy, and Haralik features. Another study that could be conducted is that instead of MLO images consider the 'Cranial-Caudal' (CC) images from the mammogram dataset where BR is also visible to estimate the age.

References

- [1] An Age Estimation Method Using 3D-CNN From Brain MRI Images - IEEE Conference Publication.
- [2] Ensemble algorithms supported by statistics and machine learning toolbox™. <https://se.mathworks.com/help/stats/ensemble-algorithms.html#bsw8at7> (visited on 2020-04-12).
- [3] Evaluate semantic segmentation data set against ground truth. <https://se.mathworks.com/help/vision/ref/evaluatesemanticsegmentation.html#d120e149007> (visited on 2020-03-20).
- [4] Fit a support vector machine regression model. <https://se.mathworks.com/help/stats/fitrsvm.html#bvdtavm-1> (visited on 2020-04-12).
- [5] Fit ensemble of learners for regression. <https://se.mathworks.com/help/stats/fitrensemble.html#bvdwj92> (visited on 2020-04-14).
- [6] Resnet-50 convolutional neural network. <https://www.mathworks.com/help/deeplearning/ref/resnet50.html#description> (visited on 2020-05-1).
- [7] Xception convolutional neural network. <https://se.mathworks.com/help/deeplearning/ref/xception.html#description> (visited on 2020-05-2).
- [8] Sultan Alkaabi, Salman Yussof, and Sameera Al-Mulla. Evaluation of Convolutional Neural Network based on Dental Images for Age Estimation. In *2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, pages 1–5, November 2019.
- [9] Mariette Awad and Rahul Khanna. Support vector regression. In *Efficient Learning Machines*, pages 67–80. Springer, 2015.
- [10] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.
- [11] Abbas Cheddad, Kamila Czene, Per Hall, and Keith Humphreys. Pectoral Muscle Attenuation as a Marker for Breast Cancer Risk in Full-Field Digital Mammography. *Cancer Epidemiology and Prevention Biomarkers*, 24(6):985–991, June 2015.
- [12] Abbas Cheddad, Kamila Czene, John A Shepherd, Jingmei Li, Per Hall, and Keith Humphreys. Enhancement of mammographic density measures in breast cancer risk prediction. *Cancer Epidemiology and Prevention Biomarkers*, 23(7):1314–1323, 2014.

- [13] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.
- [14] Ronan Collobert and Samy Bengio. Svmtorch: Support vector machines for large-scale regression problems. *Journal of machine learning research*, 1(Feb):143–160, 2001.
- [15] Ana Luiza Dallora, Johan Sanmartin Berglund, Martin Brogren, Ola Kvist, Sandra Diaz Ruiz, André Dübbel, and Peter Anderberg. Age Assessment of Youth and Young Adults Using Magnetic Resonance Imaging of the Knee: A Deep Learning Approach. *JMIR Medical Informatics*, 7(4):e16291, 2019.
- [16] Deo Rahul C. Machine Learning in Medicine. *Circulation*, 132(20):1920–1930, November 2015.
- [17] Yuan Dong, Yinan Liu, and Shiguo Lian. Automatic age estimation based on deep learning algorithm. *Neurocomputing*, 187:4–10, April 2016.
- [18] Cyrille Feudjio, James Klein, Alain Tiedeu, and Olivier Colot. Automatic extraction of pectoral muscle in the MLO view of mammograms. *Physics in medicine and biology*, 58(23):8493–8515, 2013.
- [19] Karthikeyan Ganesan, U. Rajendra Acharya, Kuang Chua Chua, Lim Choo Min, and K. Thomas Abraham. Pectoral muscle segmentation: A review. *Computer Methods and Programs in Biomedicine*, 110(1):48–57, April 2013.
- [20] Robert M Haralick, Karthikeyan Shanmugam, and Its’ Hak Dinstein. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6):610–621, 1973.
- [21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [22] Michael Heath, Kevin Bowyer, Daniel Kopans, P Kegelmeyer, Richard Moore, Kyong Chang, and S Munishkumaran. Current status of the digital database for screening mammography. In *Digital mammography*, pages 457–460. Springer, 1998.
- [23] Mohammad Hesam Hesamian, Wenjing Jia, Xiangjian He, and Paul Kennedy. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *Journal of Digital Imaging*, 32(4):582–596, August 2019.
- [24] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely Connected Convolutional Networks. pages 4700–4708, 2017.
- [25] Anne Humeau-Heurtier. Texture Feature Extraction Methods: A Survey. *IEEE Access*, 7:8975–9000, 2019.

- [26] Seyed M. M. Kahaki, Md. Jan Nordin, Nazatul S. Ahmad, Mahir Arzoky, and Waidah Ismail. Deep convolutional neural network designed for age assessment based on orthopantomography data. *Neural Computing and Applications*, August 2019.
- [27] Marzena Kamińska, Tomasz Ciszewski, Karolina Łopacka-Szatan, Paweł Miotła, and Elżbieta Starosławska. Breast cancer risk factors. *Przeгляд menopauzalny= Menopause review*, 14(3):196, 2015.
- [28] Alexandros Karargyris, Satyananda Kashyap, Joy T. Wu, Arjun Sharma, Mehdi Moradi, and Tanveer Syeda-Mahmood. Age prediction using a large chest X-ray dataset. *arXiv:1903.06542 [cs]*, March 2019.
- [29] Jaeyoung Kim, Woong Bae, Kyu-Hwan Jung, and In-Seok Song. Development and Validation of Deep Learning-based Algorithms for the Estimation of Chronological Age using Panoramic Dental X-ray Images. April 2019.
- [30] Vijay Kumar and Priyanka Gupta. Importance of statistical measures in digital image processing. *International Journal of Emerging Technology and Advanced Engineering*, 2(8):56–62, 2012.
- [31] Jang Hyung Lee and Kwang Gi Kim. Applying Deep Learning in Medical Images: The Case of Bone Age Estimation. *Healthcare Informatics Research*, 24(1):86–92, January 2018.
- [32] Rebecca Sawyer Lee, Francisco Gimenez, Assaf Hoogi, Kanae Kawai Miyake, Mia Gorovoy, and Daniel L Rubin. A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific data*, 4:170177, 2017.
- [33] Yuan Li, Zhizhong Huang, Xiaoai Dong, Weibo Liang, Hui Xue, Lin Zhang, Yi Zhang, and Zhenhua Deng. Forensic age estimation for pelvic X-ray images using deep learning. *European Radiology*, 29(5):2322–2329, May 2019.
- [34] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A. W. M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, December 2017.
- [35] Jianyi Liu, Yao Ma, Lixin Duan, Fangfang Wang, and Yuehu Liu. Hybrid constraint svr for facial age estimation. *Signal Processing*, 94:576–582, 2014.
- [36] Khoa Luu, Karl Ricanek, Tien D Bui, and Ching Y Suen. Age estimation using active appearance models and support vector machine regression. In *2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–5. IEEE, 2009.
- [37] Xiangyuan Ma, Jun Wei, Chuan Zhou, Mark A. Helvie, Heang-Ping Chan, Lubomir M. Hadjiiski, and Yao Lu. Automated pectoral muscle identification on MLO-view mammograms: Comparison of deep neural network to conventional computer vision. *Medical Physics*, 46(5):2103–2114, 2019.

- [38] Noor Mualla, Essam H Houssein, and M R Hassan. AUTOMATIC BONE AGE ASSESSMENT USING HAND X-RAY IMAGES. . *Vol.*, (02):12, 2005.
- [39] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.
- [40] Alice Othmani, Abdul Rahman Taleb, Hazem Abdelkawy, and Abdenour Haddid. Age estimation from faces using deep learning: A comparative analysis. *Computer Vision and Image Understanding*, page 102961, 2020.
- [41] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training recurrent neural networks. In *International conference on machine learning*, pages 1310–1318, 2013.
- [42] Megan S Rice, Kimberly A Bertrand, Tyler J VanderWeele, Bernard A Rosner, Xiaomei Liao, Hans-Olov Adami, and Rulla M Tamimi. Mammographic density and breast cancer risk: a mediation analysis. *Breast Cancer Research*, 18(1):94, 2016.
- [43] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597 [cs]*, May 2015.
- [44] Carl F Sabottke, Marc A Breaux, and Bradley M Spieler. Estimation of age in unidentified patients via chest radiography using convolutional neural network regression. *Emergency radiology*, 2020.
- [45] Monika Sharma, R B Dubey, and S K Gupta. Feature Extraction of Mammograms. *International Journal of Advanced Computer Research*, 2(3):10.
- [46] H. S. Sheshadri and A. Kandaswamy. Experimental investigation on breast tissue classification based on statistical feature extraction of mammograms. *Computerized Medical Imaging and Graphics*, 31(1):46–48, January 2007.
- [47] Holalu Seenappa Sheshadri and Arumugam Kandaswamy. Breast Tissue Classification Using Statistical Feature Extraction Of Mammograms. page 3.
- [48] Fuk Hay Tang, Jasmine L.C. Chan, and Bill K.L. Chan. Accurate Age Determination for Adolescents Using Magnetic Resonance Imaging of the Hand and Wrist with an Artificial Neural Network-Based Approach. *Journal of Digital Imaging*, 32(2):283–289, April 2019.
- [49] Kaier Wang, Nabeel Khan, Ariane Chan, Jonathan Dunne, and Ralph Highnam. Deep Learning for Breast Region and Pectoral Muscle Segmentation in Digital Mammography. In Chilwoo Lee, Zhixun Su, and Akihiro Sugimoto, editors, *Image and Video Technology*, Lecture Notes in Computer Science, pages 78–91, Cham, 2019. Springer International Publishing.
- [50] Eric Wu, Bin Kong, Xin Wang, Junjie Bai, Yi Lu, Feng Gao, Shaoting Zhang, Kunlin Cao, Qi Song, Siwei Lyu, and Youbing Yin. Residual Attention Based Network for Hand Bone Age Assessment. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1158–1161, April 2019. ISSN: 1945-8452.

- [51] Kaiming Yin, Shiju Yan, Chengli Song, and Bin Zheng. A robust method for segmenting pectoral muscle in mediolateral oblique (MLO) mammograms. *International Journal of Computer Assisted Radiology and Surgery*, 14(2):237–248, February 2019.
- [52] Darko Štern, Philipp Kainz, Christian Payer, and Martin Urschler. Multi-factorial Age Estimation from Skeletal and Dental MRI Volumes. In Qian Wang, Yinghuan Shi, Heung-Il Suk, and Kenji Suzuki, editors, *Machine Learning in Medical Imaging*, Lecture Notes in Computer Science, pages 61–69, Cham, 2017. Springer International Publishing.
- [53] Darko Štern, Christian Payer, and Martin Urschler. Automated age estimation from MRI volumes of the hand. *Medical Image Analysis*, 58:101538, December 2019.

