Classification of 2D Ultrasound Breast Cancer Images with Deep Learning

Jack Ellis

Kofi Appiah
Department of Computer Science
University of York
York, YO10 5GH, UK
(je976, kofi.appiah)@york.ac.uk

Emmanuel Amankwaa-Frempong
Sweden-Ghana Medical Center
Accra, Ghana
amankwaafrempong@gmail.com

Sze Chai Kwok
Division of Natural & Applied Sciences, Duke Kunshan University, Jiangsu, China
Affiliated Mental Health Center, School of Psychology & Cognitive Sci., ECNU, China
Shanghai Changning Mental Health Center, Shanghai, China
szechai.kwok@duke.edu

Abstract

Breast cancer is the second most prevalent form of cancer and is the "leading cause of most cancer-related deaths in women". Most women living in low- and middle-income countries (LMIC) have limited access to the existing poor health systems, restricted access to treatment facilities, and in general lack of breast cancer screening programmes. The likelihood of women living in LMIC attending a health facility with advanced-stage breast cancer is very high and the chances of them being able to afford treatment at that stage, even if the treatment is available, is very low. In this work, we evaluate the capabilities of deep learning as a classification tool with the aim of detecting cancerous ultrasound breast images. We aim to deploy a simple classifier on a mobile device with an inexpensive handheld ultrasound imaging system to pick up breast cancer cases that will need medical attention. We demonstrate in this work that with minimal ultrasound images, a de novo system trained from scratch can achieve accuracy of close to 64% *and about* 78% *when the same model is pre-trained.*

1. Introduction

Many people in their lives will face cancer at some point [29]. In 2020, it was the leading cause of death around the world, accounting for more than 10 million lives lost [16] and countless more, changed. As the population ages at large, cancer will become more of a problem, overwhelming hospitals and increasing waiting times for those with more immediate issues. It is a disease that gets increasingly more difficult to treat as it progresses, it could start in one part of the body (cells dividing rapidly in the skin), then find its way very quickly to other organs through the lymph nodes, to the heart, kidneys, lungs, breasts, liver and others [26]. If we can find it quickly, we could cut it out at the source, halting any serious progression and stopping any further potential harm. This is why early surgery is often regarded as the best form of cancer treatment.

Breast cancer is the second most prevalent form of cancer and is the "leading cause of most cancer-related deaths in women in the United States" [21]. In 2021, it was predicted that around 281,550 women were to be diagnosed with breast cancer and around 43,600 women were predicted to die due to breast cancer. It has been shown that "*almost all* women with breast cancer survive their disease for 5 years or more, if diagnosed at the earliest stage. This falls to around 3 in 10 women when breast cancer is diagnosed at the most advanced stage" [28]. Ultrasound together with digital mammography are the mainstay of breast cancer screening and diagnosis.

Computed Tomography (CT) scan and Magnetic Resonance Imaging (MRI) are the mainstay for determining the extent of spread of breast cancer in patients who have the disease, because of their higher soft tissue resolution and extent of body coverage. These facilities (Mammogram/MRI) are however not widely available in low- and middle-income countries (LMIC) and where available, are very expensive. Considering the comparatively higher availability and access to Ultrasound (USG) and the ease of use of mobile (portable) USG in even hard to reach terrains and now well supported observation of USG as an effective substitute tool for detection of breast cancer [23], in this work, we aim to enhance the effectiveness of USG in the early diagnosis of breast cancer in LMIC. Recent advances in artificial intelligence (AI) and machine learning (ML) have provided consecutive successes for helping medical doctors in making diagnoses [18]. This work will build on the successes of AI by using machine learning models to further develop the effectiveness of USG in speeding up breast cancer treatment.

2. Related Work

Image classification problems have become more prevalent over the last few decades as we realised that we can harness this new power of machine learning to do a lot more than we initially imagined. Vailaya et al [30] describe a method in which to classify images from landscapes to cities using machine learning and asked different people to first classify the images into specific groups and then used this training data to train their machine learning model; the input is the image to classify and the output is a one-hot representation of each class with its corresponding confidence level. They [30] managed to achieve an accuracy of around 93% throughout their study using a *k-NN classifier* [9] to classify the images.

With image classification being one of the more useful cases of machine learning, it has allowed us to create much better facial recognition algorithms, better self-driving car functionality, detecting obstacles such as people, other cars or cyclists, better scientific imaging such as cancer and injury detection, amongst many others [25] [19] [14]. Machine Learning refers to computational models that, to the best of our knowledge [5], mimic our own brains by way of many inter-connected neurons all computing weighted values in which we perceive as the act of *learning*. Convolutional Neural Networks (CNNs), as described by Sun et al [24], have ”exceptional superiority in visual recognition tasks” and showed that detecting traffic light signals, biological imaging and generic image classification all have positive results when using the convolutional network method. There have been many studies that prove that it is possible to detect breast cancer using different machine learning techniques with some being more accurate than others. Some people use a combination of feature extraction and simple deep learning models to classify the output and others use image based techniques such as CNNs. For the feature extraction method, the dataset used, is a list of attributes provided by radiologists, for example: cell size, clump size and cell uniformity, the most popular being the Wisconsin data set [31]. Sometimes, the features are extracted automatically, using a CNN, by detecting edges and computing pixel difference.

In [12], they describe how they first pre-processed their data to remove noise and redundant, ambiguous data from the dataset. The dataset contained 16 rows with missing values and so to stabilise the results, they removed all data with those missing values. In order to select which features were best fit for their model, they used a method known as Recursive Feature Elimination, an approach which iteratively removes features that are deemed unsuitable or unnecessary for the training process. Once they had selected all of the essential features, they split their data into subsets for training and testing; this was done randomly, without following any sequences. The Deep Neural Network used contains an input layer with four input nodes, three hidden layers with 10, 20, and 10 hidden nodes, and an output layer with a single node. The results presented in [12] are very promising, showing that the 80-20 split had an average accuracy of 98.62% where the other splits had lower accuracy’s, 97.66% and 96.88%.

Another example of where deep neural networks (DNNs) are used for classification is [7]. They use a very similar technique to [12], using the Wisconsin data set and using a feature elimination technique to determine which features to use in their model. For their experiments, they tested a range of numbers of features with differing classifier techniques and recorded the accuracy of each test. They noted that with 19 features (out of a possible 32), they had the best accuracy across the board with an average accuracy of 97%. After they had found the number of features that provided the best accuracy, they determined how many hidden layers and nodes to use in their model. Eventually, they found that 3 hidden layers with 60 hidden nodes provided the best accuracy of 99.42%; it is worth noting that in their data, it seems to have a linear correlation, every time a node or a layer is added, the accuracy seemed to improve.

In [15], Nasser and Yusof explained the most popular methods that are currently being used to tackle the issue of breast cancer detection. They explain that there is an absolute need for automated learning to supplement the current human effort in order to correctly diagnose breast cancer, with at least the accuracy of a human. They list the techniques used, and go on to describe the **convolutional neural network** to be the most popular method utilised in breast cancer detection as of the publish date. CNNs used for breast cancer detection can be grouped into two: the transfer learning-based model and the de novo trained model; a model which is trained from scratch. The transfer learning-based model uses previously trained networks such as AlexNet [4], ResNet [10] or VGG [22] to significantly speed up the results process as the models have already been partially trained. Using previously trained networks can also improve the accuracy of a model as already well-known accurate weights can be partially retrained to fit a new problem.

Ting et al [27] describe how they use a CNN to aid with the diagnosis of breast cancer using mammographic images. The images used were provided by the Mammographic Image Analysis Society (MIAS), provided in the Portable Gray Map image format and verified by a doctor with 21 benign, 17 malignant and 183 normal cases. A CNN of 28 hidden layers and 1 output layer was used as part of their network model, a dropout rate of 0.5, a learning rate of 0.002, fully connected hidden layer of 1024 neurons, and Rectified Linear Units. Data augmentation is also applied to the image patches (128x128 pixels) to *"overcome the overfitting issue faced by other researchers"*, a rotation and a flip to each patch is also applied as part of the augmentation stage. They also use an average pool layer to reduce the dimensions of the hidden layer so the subsequent layer has the correct dimensions needed for the output. They explain that for pooling, the exact feature location is less important than the *"approximate location relative to other features"* and so can control the *"over-learning"* tendencies. They [27] achieved a sensitivity, specificity, and accuracy scores of 89.47%, 90.71%, and 90.50% respectively, using their Convolutional Neural Network Improvement for Breast Cancer Classification (CNNI-BCC) technique.

In [20], Shen et al described their methods in classifying breast cancer mammogram data using an image patching process. They proposed what they called an "end-to-end" approach where they pre-train a model to detect the image patches with fully-annotated Regions of Interest (ROI). They used a combination of a VGG [22] and ResNet [10] architecture to classify the ROIs, more specifically the VGG-16 and ResNet-50 flavours. Their dataset consisted of 2478 mammography images from 1249 women from the CBIS-DDSM with training, validation and testing sets split as 1903, 199 and 376, respectively. The images were downsized to 224x224 pixels because of limitations to their hardware and the fact that this was big enough to cover most of the image patches. Their network training was split into two parts, the "patch classifier" and the "whole image classifier". The patch classifier was trained using a pre-trained classifer and transfer learning, the model used was the ImageNet model; they only needed to train the final few layers to adjust the weights to their specific problem set. Then they trained the whole image classifier using their VGG structure and compared it against the ResNet structure to see which one showed better results. To improve their final model, data augmentation was performed using some random transformations of the images: horizontal and vertical flips, rotation, zoom and intensity shift. They achieved a per-image accuracy of around 88% with their best model using a ResNet50 [10] and VGG-16 [22] combination.

Using CNNs to classify breast cancer with MRIs is described in [2]. Their dataset was split into training and validation sets in the ratio 3:1 or 214:72 cases and trained on a RetinaNet [13] classifier. The RetinaNet structure uses a Feature Pyramid Network backbone, as well as a ResNet [10] architecture to *"generate a rich multi-scale convolutional feature pyramid"*. Their results show that the machine learning approach is significantly more accurate than the human counterparts: 92.6% and 82.8% (sensitivity and specificity respectively) for the machine learning system with 84.7% and 84.1% for human readers (four radiologist). They go on to explain that the study is the *"first report to focus on building an AI system based on RetinaNet that detects and diagnoses lesions of MIPs of DCE breast MRI and compares the diagnostic performance of the AI to that of human readers"* and as previously stated, their machine learning outcomes were far greater than that of the human readers. However, a limitation of their study was that of the image size; they converted all of their images to 512x512 pixels which could reduce the accuracy of the network as information will be lost due to the compression algorithm used.

Jafari et al [11] describe their methods to predict breast cancer within mammogram data. They used the Radiological Society of North America (RSNA) dataset' from roughly 11,000 patients and 54,713 images in the DICOM format. Before feeding any data into their models, they first pre-processed their images using a range of techniques, from normalisation (converting all pixel values to a range of 0-255), Region of Interest Selection which enabled them to crop the image to the section they were most interested in analysing, to feature extraction, using pre-trained models for extracting features and choosing the ones that will make the most impact for classification. Once they had those features, they classified them using different models such as k-NN, random forest (RF), SVM, and NN. They found that the NN approach showed the best results, obtaining an accuracy of 92%, using two fully connected layers with a hidden layer of 96 neurons and a single-neuron classification layer.

In their study for the New York University dataset [6], Yiqiu Shen et al described their methods in creating a classifier for pre-processing their ultrasound images. With each image, they use a combination of cropping and rotation in order to remove any of the unnecessary background information that may impact the training. They used the methods of Dilation and Erosion in order to remove the surrounding areas: first they needed to make the images grey-scale, then they needed to obtain a non-zero mask of the image and for the devices where the background wasn't zero using an algorithm to fetch the modal value of the background pixels. After they had detected the section that needed to be removed, they used an erosion technique on the non-zero binary mask to remove that section. In the dataset, there are images where text or other annotations are used for the radiographers benefit but to the hindrance of the machine learning model as it obstructs the image underneath. These annotations are called "burnt in annotations" and they need to be detected and removed before the model can be trained to stop the network from 'relying' on the annotations. They first trained a ResNet-18 [10] network in order to detect the annotations, they then used a U-Net [17] classifier combined with image in-painting [1] to remove the detected annotations.

Guan and Loew [8] explain that training CNNs from scratch may be inefficient as they require a large amount of labeled data in order for them to be accurate. They [8] explain how they *applied the generative adversarial network (GAN) to generate synthetic mammographic images from the digital database for screening mammography (DDSM)*. The problem with classic image augmentation is that they won't be used as new training images but instead only versions of images that already exist so the accuracy of the network cannot be significantly improved. They go on to explain that instead, one can generate images using GANs, keeping the original features of the original images, but making it a completely different image; this would improve the ease of accessing datasets because you could generate new mammographic images whenever you needed to. The GAN is a generative neural network that takes existing data and creates 'fake' data following the same patterns seen in the real data.

3. Our Approach

Machine learning is a technique known to be accurate (when trained correctly) and time efficient when dealing with classification problems. This will reduce the number of errors made, when compared to human counterparts and can exploit the fact that computers can effectively run for 24 hours a day, 7 days a week, detecting cancer even when a human may not be around. Machine learning algorithms may be expensive to train in terms of time, storage and data available, but in the long term, their accuracy and usability are worth the use of resources.

The online resources used are informational websites such as the UK National Health Insurance (NHS) website or Cancer Research website [29], providing knowledge on breast cancer statistics and research papers describing the techniques for creating and running machine learning algorithms to detect breast cancer. We conducted background research into how neural networks work and the best types of neural network structure to use for this particular problem using credible sources from various universities and institutions around the world. We also researched which programming languages and libraries to use in order to find an effective solution; Python with TensorFlow being the final platform with many machine learning problems and TensorFlow is an easy framework to work with within the Python ecosystem.

3.1. Dataset selection

First, we will need to have a dataset consisting of ultrasound images with which to train our network on. We chose the publicly available dataset [3] [6] for the input images to the network as the folder structure is easy to work with and the

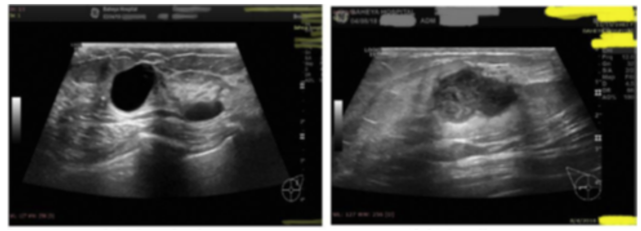

Figure 1. Sample images from the dataset in [3]. The image to the left is a sample benign ultrasound image and to the right is a sample malignant ultrasound image.

Figure 2. Sample corresponding ground-truth images showing the marked out locations of any tumour in the ultrasound images taken from [3] dataset.

images are of a high quality. The data is set out into 3 folders: benign, malignant and normal each consisting of high quality ultrasound images in the PNG format, 210 benign, 210 malignant and 130 normal. Sample ultrasound images taken from the [3] dataset are as shown in Figure 1. Each non-normal image also had an accompanying positional image which shows exactly where the tumour is, if we wanted to train our network to not only detect but to show where the tumour is, as shown in Figure 2. For this preliminary work, we removed the positional images as we deemed them not necessary. We also use ultrasound images provided to us by radiologists specifically for this work (York Ultrasound Images) for testing. The aim is to test the robustness of the system in recognising ultrasound images in a format and a dataset that it hasn't seen before. After we have collected our ultrasound images, we pre-process them so that our network has the **best** chance of correctly identifying the hidden patterns between them. To do this, we will have to employ a range of techniques such as data augmentation, removal of unnecessary background information and grouping, amongst others.

3.2. Machine learning techniques

After the pre-processing, we build the machine learning model. The best approach to building this model, is to use a convolutional neural network (as we are working with images) in either the ResNet-18 [10], or VGG-16 [22] structure, as they have proven to be accurate when dealing with image-based problems. We also use data augmentation and dropout to make sure that the model doesn't over-fit the data and also make sure that for each epoch, the dataset is shuffled so that the model doesn't just learn which image will come next. Once the model is trained, we compare it against common baselines or to an accuracy threshold set, in this work, 75%. To see how we can improve; we employ trial and error to our methods so that we choose the correct implementation for this use case, for instance, by changing the activation or optimisation functions, or changing the number of nodes and layers.

3.3. Cleaning the dataset

We have chosen to use the NYU dataset [6] supplemented by the York Ultrasound Images. In order to use the images from the radiologists (York Ultrasound Images), we remove any personal information from the images using a very simple Keras application which learns where the text is in the image and removes it. We also remove any images with manually marked-out or segmented tumour areas by the radiologist, to avoid the introduction of any biases. Once we have the dataset, we will split it into two groups: training and testing. We take around 10 images from each class for the York Ultrasound Data and add them to their corresponding testing class from the NYU dataset, for example, we move a benign image to the benign testing folder. We first tested the small datasets on three different VGG structure (VGG-12, VGG-16 and VGG-19) and test how accurately each model predicts the outcome correctly.

We train each model 3 times and take accuracy measurements each time to get an average accuracy for that model. If the accuracy of the model is deemed unsuitable, we switch to InceptionNet to compare the results. By way of testing we have automated the process with a script that takes the model generated and will run it against all of the test input. If the output of the model for a specific input image matches the class folder, then *1* will be added to the 'score' and after we have completed every test image, we divide the 'score' by the total number of images seen, to get the final accuracy. This is a good approach because this is the model being tested on unseen data and will simulate a real world application. We use trial and error to create at least 5 models with varying node and layer counts to find the best fit for this problem.

3.4. Implementation

We split the data into 3 sets: training, validation and testing (the validation set was determined programatically in TensorFlow). For training, we have **202** in the benign set, with **7** from our York Ultrasound Images (YUI) dataset, **202** malignant, with **7** from our YUI dataset and **125** normal with none from the YUI dataset. The validation set was created as a 0.1 split of the training data with the seed *123* so that for each test, the data would be split the same. For testing, we have **17** in the benign set, with **2** from our YUI dataset, **17** malignant, with **2** from our YUI dataset and **8** normal with **0** from our YUI dataset. The idea for the test set is that, the data is from the same set (all from the NYU dataset and our YUI dataset) but the machine will not see them in the learning process and so we label them as *unseen* data.

3.5. Preliminary Results

To test our hypothesis that with transfer learning we can train a deep learning model with minimal medical images to reach an acceptable level of classifying an ultrasound image as malignant or benign, we have tested four different models with the same dataset by first training the models from scratch and then using transfer learning. The four models are three different VGG architectures (VGG-12, VGG-16 and VGG-19) and ResNet50. We found that on the average the accuracy increased when the model is pre-trained and even with the de novo models, most of the misclassified images are benign ultrasound images misclassified as malignant. Any false positive in cancer diagnosis and for that matter breast cancer is unacceptable and should be minimised.

False positives in breast cancer screening using ultrasound can be confusing to the patient involved as well as the healthcare system. Patients can end up with serious anxiety, depression constant stress and just feel crappy in general, even after finding out it was a mistake. And their families go through a lot of worry too, which can take a toll on relationships. Some people stay scared about their health for a long time after, making them not want to do tests in the future or not loose truth. There is also the financial side; false positives usually mean doing more tests like biopsies, MRIs, and follow-up appointments to confirm. All those extra procedures cost money that didn't need to be spent. Basically, falsely diagnosing cancer causes bad emotional and financial consequences. It shows why its so important to make sure scan results are accurate.

In all cases from our preliminary experiments, none of the normal images are being misclassified as either benign or malignant, which is inline with our expectation that the model should be capable of identifying individuals with some sort of tumour for further medical examination by way of preventing cancerous tumours developing any further and unknowingly to the patient. As shown in Table 1, the pre-trained ResNet50 obtained the highest accuracy of 77.77%.

The experimental data collected in Table 1 were collected using a single *NVIDIA A40* GPU using 50 epochs for each test and no early stopping. These are the steps taken to arrive at the values in table 1:

- 1. Load the model into the script
- 2. For each image in the test set, load the image and the label into numpy arrays

Model	VGG-12	VGG-16	VGG-19	ResNet50
De-Novo	56.80%	57.93%	62.69%	64.37%
PreTrained	60.25%	62.69%	73.80%	77.77%

Table 1. The accuracy in percentages of the three models tested when trained from scratch and when trained with pre-trained parameters.

- 3. Run the image through the model
- 4. If the output is correct, add one to the score
- 5. Finally, divide the score by the total to receive the accuracy.

For each network in Table 1, we ran the training algorithm three times and tested the accuracy. The accuracy reported is hence the average of three test runs for each model.

4. Conclusion

There is an estimated 44% reduction in mortality rate amongst women exposed to breast cancer screening; an estimate based on women who can easily access health facilities, and in most cases the cause of death is easily identified. Unfortunately, a large percentage of women do not have access to medical facilities and may die from an undetected or undiagnosed form of cancer. This work is the first step in using deep learning with minimal or limited medical data (images) to supplement the unavailable medical facilities in various part of the world to minimise some of the avoidable deaths related to cancer. We have demonstrated in this work that with hundreds of images, coupled with transfer learning, we are able to train a deep earning model to have an accuracy of approximately 78%. Our overall aim is to increase the training data with the aim of improving on the accuracy of the models and yet fit the entire model on a mobile device to aid quick and early tumour detection.

Acknowledgement

Special thanks goes to George Asafu Adjaye Frimpong (radiologist), Spectra Health, Ghana for his continuous support in providing this work with the much needed ultrasound images with ground truth.

References

[1] Telea A. An image inpainting technique based on the fast marching method. *Journal of graphics tools 9(1):23–34.*, 2004. 4

[2] Mio Adachi, Tomoyuki Fujioka, Mio Mori, Kazunori Kubota, Yuka Kikuchi, Wu Xiaotong, Jun Oyama, Koichiro Kimura, Goshi Oda, Tsuyoshi Nakagawa, Hiroyuki Uetake, and Ukihide Tateishi. Detection and diagnosis of breast cancer using artificial intelligence based assessment of maximum intensity projection dynamic contrast-enhanced magnetic resonance images. *Diagnostics*, 10(5), 2020. 3

[3] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. Dataset of breast ultrasound images. *Elsevier Journal of Data in Brief*, 28, 2020. 4

[4] Ilya Sutskever Alex Krizhevsky and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, 2012. 2

[5] MIT Anne Trafton. The brain may learn about the world the same way some computational models do, 2023. 2

[6] Jan Witowskid Jamie Oliverd Kawshik Kannanf Nan Wuc Jungkyu Parkd Beatriu Reigd Linda Moyd Laura Heacockd Farah Shamouta, Yiqiu Shenc and Krzysztof J. Gerasd. The nyu breast ultrasound dataset v1.0. 2021. 3, 4, 5

[7] Md. Omaer Faruq Goni, Fahim Md. Sifnatul Hasnain, Md. Abu Ismail Siddique, Oishi Jyoti, and Md. Habibur Rahaman. Breast cancer detection using deep neural network. In *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, pages 1–5, 2020. 2

[8] M. Guan, S. Loew. Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks. *Journal of medical imaging (Bellingham, Wash.), 6(3), 031411. https://doi.org/10.1117/1.JMI.6.3.031411*, 2019. 4

[9] Gongde Guo, Hui Wang, David Bell, Yaxin Bi, and Kieran Greer. Knn model-based approach in classification. In *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, pages 986–996, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg. 2

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 2, 3, 4

[11] Zahra Jafari and Ebrahim Karami. Breast cancer detection in mammography images: A cnn-based approach with feature selection. *Information*, 14(7), 2023. 3

[12] S. et al Karthik. Breast cancer classification using deep neural networks. *Margret Anouncia, S., Wiil, U. (eds) Knowledge Computing and Its Applications. Springer, Singapore. https://doi.org/10.1007/978-981-10-6680-1*$_1$2*b*, 2018. 2

[13] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017. 3

[14] Mpho Mokoatle, Vukosi Marivate, Darlington Mapiye, Riana Bornman, and Vanessa. M. Hayes. A review and comparative study of cancer detection using machine learning: Sbert and simcse application. *BMC Bioinformatics*, 24, 2023. 2

[15] Maged Nasser and Umi Kalsom Yusof. Deep learning based methods for breast cancer diagnosis: A systematic review and future direction. *Diagnostics (Basel, Switzerland), 13(1), 161.*, 2023. 2

[16] World Health Organisation. Cancer, 2022. 1

[17] Brox T Ronneberger O, Fischer P. U-net: Convolutional networks for biomedical image segmentation in international conference on medical image computing and computer-assisted intervention. *(Springer), pp. 234–241.*, 2015. 3

[18] Ebrahim Mohammed Senan, Mukti E. Jadhav, Taha H. Rassem, Abdulaziz Salamah Aljaloud, Badiea Abdulkarem

Mohammed, and Zeyad Ghaleb Al-Mekhlafi. Early diagnosis of brain tumour mri images using hybrid techniques between deep and machine learning. *Computational and Mathematical Methods in Medicine*, 2022(5):1–17, 2022. 2

[19] Chirag Sharma. Self driving car using deep learning technique. *International Journal of Engineering Research and*, V9, 2020. 2

[20] Li Shen, Laurie R. Margolies, Joseph H. Rothstein, Eugene Fluder, Russell McBride, and Weiva Sieh. Deep learning to improve breast cancer detection on screening mammography. *Scientific Reports, 2045-2322, https://doi.org/10.1038/s41598-019-48995-4, 10.1038/s41598-019-48995-4*, 2019. 3

[21] Rebecca L et al Siegel. Cancer statistics. *CA: a cancer journal for clinicians vol. 71,1 (2021): 7-33. doi:10.3322/caac.21654*, 2021. 1

[22] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015. 2, 3, 4

[23] Rupali Sood, Anne F Rositch, Delaram Shakoor, Emily Ambinder, Kara-Lee Pool, Erica Pollack, Daniel J Mollura, Lisa A Mullen, and Susan C Harvey. Ultrasound for breast cancer detection globally: A systematic review and meta-analysis. *Journal of Global Oncology*, 14(5):1–17, 2019. 1

[24] Yanan Sun, Bing Xue, Mengjie Zhang, and Gary G. Yen. Evolving deep convolutional neural networks for image classification. *IEEE Transactions on Evolutionary Computation*, 24(2):394–407, 2020. 2

[25] KH Teoh, RC Ismail, SZM Naziri, R Hussin, MNM Isa, and MSSM Basir. Face recognition and identification using deep learning approach. *Journal of Physics: Conference Series*, 1755(1):012006, 2021. 2

[26] Leanne C. Huysentruyt Thomas N. Seyfried. On the origin of cancer metastasis. *Crit Rev Oncog. 2013;18(1-2):43-73. doi: 10.1615/critrevoncog.v18.i1-2.40. PMID: 23237552; PMCID: PMC3597235.*, 2014. 1

[27] Fung Fung Ting, Yen Jun Tan, and Kok Swee Sim. Convolutional neural network improvement for breast cancer classification. *Expert Systems with Applications Volume 120, 15 April 2019, Pages 103-115*, 2018. 2, 3

[28] Cancer Research UK. Why is early cancer diagnosis important?, 2023. 1

[29] Cancer Research UK. Cancer incidence statistics, 2024. 1, 4

[30] ADITYA VAILAYA, ANIL JAIN, and HONG JIANG ZHANG. On image classification: City images vs. landscapes. *Pattern Recognition*, 31(12):1921–1935, 1998. 2

[31] William Wolberg, Olvi Mangasarian, Nick Street, and W. Street. Breast cancer wisconsin (diagnostic). UCI Machine Learning Repository, 1995. DOI: https://doi.org/10.24432/C5DW2B. 2