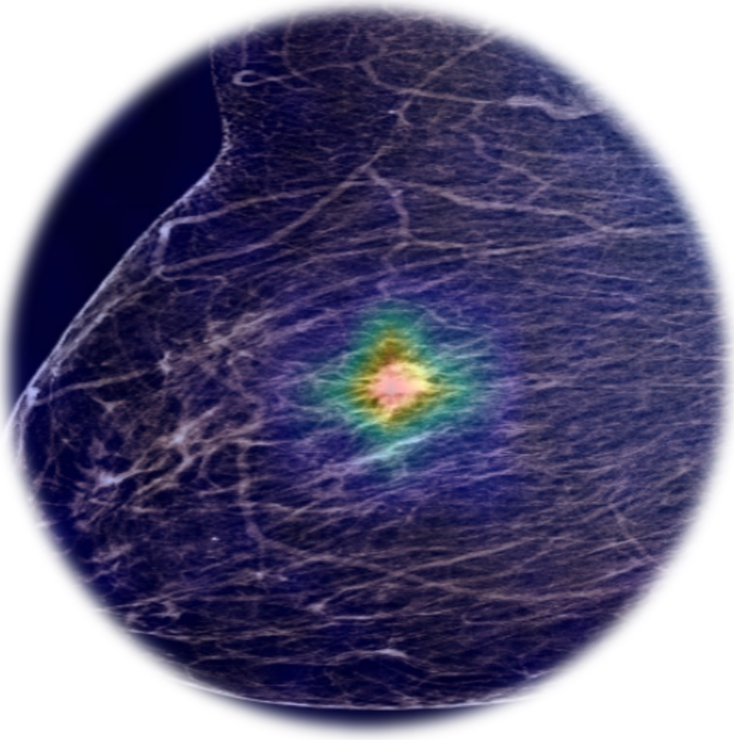

ARTIFIN

“Mammography Breast Cancer Detection”
Project Report - Team 2

For the module “*ARTIFIN*” in the *MSc ITDS* at the
Lucerne University of Applied Sciences and Arts
School of Computer Science and Information Technology

Oluwadarasimi Emmanuel Falade	oluwadarasimi.falade@stud.hslu.ch
Amr Ashraf Kandil	amr.kandil@stud.hslu.ch
Manuel Fernando Lopez	manuel.lopez@stud.hslu.ch
Idan Andrei Banag Paguio	idanandrei.paguio@stud.hslu.ch

Disclaimer
All sources of information (including images) are cited using the IEEE standard, either in-, above-, or below-text. References are automatically ordered by the order of citation (i.e. numerical order based on when they appear in the document), not alphabetically (A–Z). This is by design, as IEEE style is a numeric citation style, and the reference list must reflect the order in which citations appear in the text.

Table of Contents

1 Introduction

Breast cancer detection through mammography is a complex and demanding task that requires a high level of expertise and attention to detail. Mammographic images often contain subtle and overlapping features that can make it difficult to distinguish between benign and malignant findings. Factors such as breast density, image quality, and patient history add further layers of complexity to the diagnostic process. This challenge is compounded by the fact that early-stage tumors can appear as faint or ambiguous patterns, requiring sharp visual acuity and pattern recognition skills to detect.

Radiologists face significant pressure in interpreting these images accurately, as missing an early sign of cancer can have serious consequences. The task is made even more difficult by the repetitive nature of image reading and the sheer volume of cases, which can lead to fatigue and reduced concentration. Human error, even among experienced professionals, remains a risk in this environment. As a result, there is growing interest in leveraging computer-aided detection tools and artificial intelligence to support radiologists in identifying potential abnormalities, enhancing diagnostic accuracy, and ultimately improving patient outcomes.

1.1 Purpose and scope of this technology report

This report is for the module "ARTIFIN" in the Master of Science program "IT, Digitalization & Sustainability" at Hochschule Luzern (HSLU) [1]. It serves as a summary and reasoning document for our project work.

1.2 Methodology

This report is based on our project and research conducted through internet sources. Information was gathered from reputable websites, academic papers, material from the lectures of Dr. Sc. ETH Javier Montoya, industry reports, and news platforms.

2 Overview

2.1 Use-Case

Mammography breast cancer detection is a critical process for early diagnosis and treatment. The radiologist or an AI system analyzes mammogram images to identify abnormalities that may indicate cancer. After the patient undergoes a mammogram, the images are reviewed for suspicious areas. AI can assist by highlighting potential issues, which the radiologist then verifies. If abnormalities are found, further tests or a biopsy may be recommended. Early detection significantly increases survival rates and improves treatment outcomes, making mammography an essential tool in reducing breast cancer mortality. Research in this area could maybe help to bring technology and knowledge like this also sooner to third world countries. This ultimately benefits the SDGs *3 (Good health and well-being)*, *5 (Gender equality)*, and *10 (Reduced inequalities)*.

2.2 Problem

Detecting breast cancer requires significant medical expertise and considerable time, as mammogram images contain a vast amount of complex information. Artificial Intelligence (AI) has the potential to greatly assist in this process by pre-analyzing mammography results, identifying areas of concern, and providing a second layer of review to support doctors in making accurate diagnoses.

According to the *World Health Organization (WHO)*, breast cancer is the most frequently diagnosed cancer globally. In 2020 alone, 2.3 million new cases were reported, resulting in 685'000 deaths. However, high-income countries have seen a 40% decline in breast cancer mortality since the 1980s, largely due to the introduction of regular mammography screening for at-risk age groups. Early detection and timely treatment remain essential to lowering cancer-related deaths. Machine learning expertise could play a key role in enhancing how radiologists interpret screening mammograms [2].

At present, the early detection of breast cancer depends on the skills of highly trained radiologists, making screening programs costly and resource intensive. A growing shortage of radiologists in several regions threatens to exacerbate these challenges. Additionally, mammography can yield a high rate of false positives, causing unnecessary stress, follow-up procedures, further imaging, and often invasive tissue sampling such as needle biopsies.

2.3 Goals

Help detect breast cancer from mammography images. We want to try to experiment and combine CNN and Auxiliary data.

2.4 Models & Dataset

This project uses the "*RSNA Screening Mammography Breast Cancer Detection*" dataset from a Kaggle competition [2], and EfficientNet-B4 [3] as the model.

This competition is hosted by the Radiological Society of North America (RSNA), a non-profit organization representing 31 radiologic subspecialties across 145 countries. RSNA is committed to advancing patient care and health care delivery through education, research, and innovation.

3 Dataset

This dataset [4] comprises **54’713 files**, totaling approximately **314.72 GB**, primarily in **DICOM (.dcm)** and **CSV** formats. It includes mammogram images and associated metadata for both training and test sets. It includes both imaging and clinical features to support multi-modal analysis.

Files

- [train/test]_images/[patient_id]/[image_id].dcm are the mammograms in the **DICOM format** (many of the images use the jpeg 2000 format). Each patient typically has 4 images.
- [train/test].csv and contain detailed metadata for each image and patient.
- sample_submission.csv is a partial example of a valid submission format.

Key Metadata Fields

`site_id`	ID code for the source hospital.
`patient_id`	ID code for the patient.
`image_id`	ID code for the image.
`laterality`	Whether the image is of the left or right breast.
`view`	The orientation of the image. The default for a screening exam is to capture two views per breast.
`age`	The patient's age in years.
`implant`	Whether or not the patient had breast implants. Site 1 only provides breast implant information at the patient level, not at the breast level.
`density`	A rating for how dense the breast tissue is, with A being the least dense and D being the most dense. Extremely dense tissue can make diagnosis more difficult. Only provided for train.
`machine_id`	An ID code for the imaging device.
`cancer`	Whether or not the breast was positive for malignant cancer. The target value. Only provided for train.
`biopsy`	Whether or not a follow-up biopsy was performed on the breast. Only provided for train.
`invasive`	If the breast is positive for cancer, whether or not the cancer proved to be invasive. Only provided for train.
`BIRADS`	0 if the breast required follow-up, 1 if the breast was rated as negative for cancer, and 2 if the breast was rated as normal. Only provided for train.
`prediction_id`	The ID for the matching submission row. Multiple images will share the same prediction ID. Test only.
`difficult_negative_case`	True if the case was unusually difficult. Only provided for train.

4 Implementation

As per conversations with our professor in class and via email, the implementation and documentation of our project / solution can be found in the Jupyter Notebook. In this section we will elaborate the model used.

4.1 EfficientNet

EfficientNet [3] has proven to be a powerful convolutional neural network (CNN) architecture widely adopted in medical image analysis due to its balance of accuracy and efficiency [5]. It has been effectively applied in critical tasks such as COVID-19 diagnosis [6], lung cancer classification [7], and melanoma detection [8], among many others.

In this project, **EfficientNet-B4** was selected as the optimal model, offering a suitable tradeoff between performance and computational cost. While larger variants like B5 to B7 offer improved accuracy, B4 delivers sufficient learning capacity for region-of-interest (ROI) enhanced inputs and remains efficient for fine-tuning on available hardware. The larger variants have been tried, but did not work due to hardware constraints. Future work may explore the use of more advanced versions to further enhance model performance.

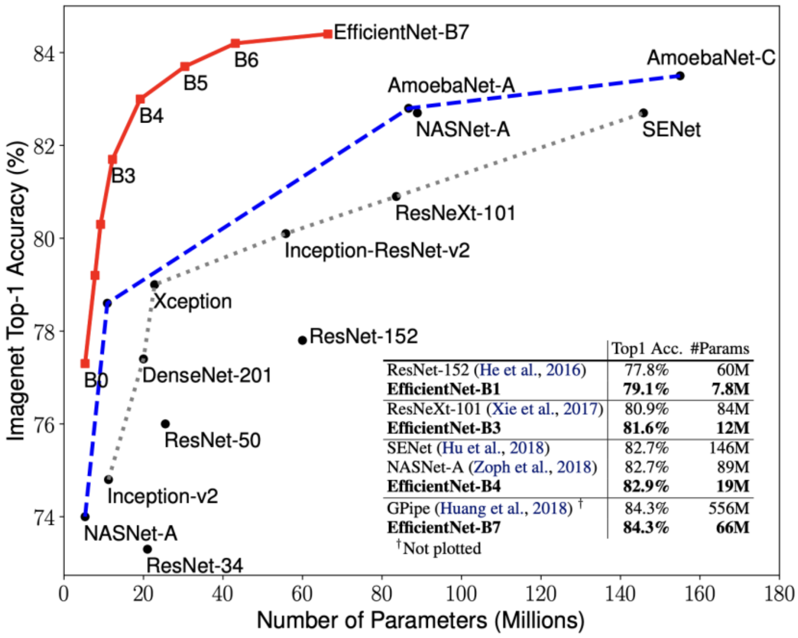

Figure 1: Model Size vs. ImageNet Accuracy [3, Fig. 1]

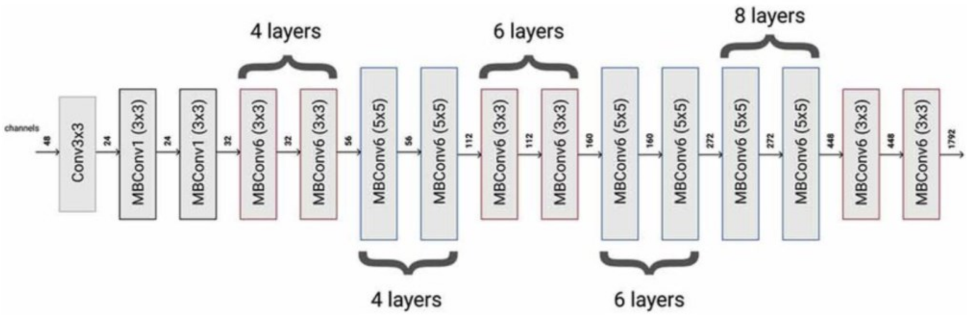

Figure 2: Architecture of efficientNet-B4 model [9]

5 Experiments

We tested various combinations of hyperparameters, though our exploration was constrained by computational limitations.

Ultimately, we focused on comparing a pure CNN model with our hybrid CNN + Auxiliary model. We adjust the weight ratio between different losses to control how much influence each task has on the model's learning. If one loss is too big or too small, it can dominate or get ignored. By tuning the ratio, we make sure the model learns from all tasks in a balanced and effective way, depending on what matters most. Our goal was to identify the optimal balance that would enhance the model's overall learning performance.

We evaluated from 0.0 to 1.0 (at 0.1 increments), which affected our total loss calculation:

```
total_loss = (1 - CFG.aux_loss_weight) * cancer_loss + CFG.aux_loss_weight * aux_loss
```

Weighted Average Loss

CNN + AUX Model	F1_thr	Accuracy	Threshold	Fold
Weight 1	0.4067	0.7330	0.4082	2
Weight 2	0.3973	0.7337	0.4286	2
Weight 3	0.3985	0.6256	0.3265	2
Weight 4	0.3956	0.6649	0.3673	2
Weight 5	0.3824	0.7141	0.4286	2
Weight 6	0.3781	0.7802	0.4082	0
Weight 7	0.3830	0.7083	0.3878	0
Weight 8	0.3973	0.8060	0.4082	0
Weight 9	0.3793	0.7931	0.4082	0
Weight 10	0.3189	0.2632	0.4694	2
CNN ONLY	0.3936	0.7873	0.4082	1

Figure 3: Weighted Average Loss

6 Challenges & Limitations

Breast cancer detection using AI and machine learning faces numerous challenges, starting with the quality and quantity of available data. High-quality annotated medical images are crucial for training effective models, yet such data is often scarce due to privacy concerns and the high cost of expert labeling. Even when datasets are available (like ours), they may lack quality, diversity, or consistency, leading to biased or underperforming models. Domain knowledge is another major hurdle; understanding the nuances of medical imaging requires collaboration with domain experts like radiologists, which can be time-consuming and resource intensive.

Technical and operational limitations further complicate the development process. Medical AI applications typically require significant computational power for training and inference, but developers may be restricted by limited server uptime, RAM, or overall system performance. In our case we could not really use a better version of EfficientNet-B4 due to the hardware constraints of HSLU Lab Services GPUHub.

Accessing data and development environment (GPUHub) through the HSLU VPN and browser-based platforms like Jupyter Notebook, adds latency and hinders workflow efficiency. We also noticed that if the browser or computer stopped, the notebook / execution also stopped on GPUHub. Time is also a critical constraint, both in terms of model training duration and the pressure to deliver clinically useful solutions rapidly.

Together with the challenges and limitations mentioned above, these factors made it rather challenging for us to work on this project.

7 Results

Our findings indicate that integrating auxiliary inputs with convolutional neural networks (CNN + AUX) can potentially improve standalone CNN models (when specific lower weight thresholds are applied). The combined architecture demonstrated improved predictive accuracy and robustness in identifying early-stage breast cancer from mammographic images. The f1 metric, which accounts for precision, recall, and confidence thresholds, proved valuable in evaluating real-world applicability and helped isolate the most clinically relevant configurations. The training metrics indicate stable learning, with key indicators such as f1, where the best threshold was computed, steadily increasing and the loss consistently decreasing, which reflects ideal behavior.

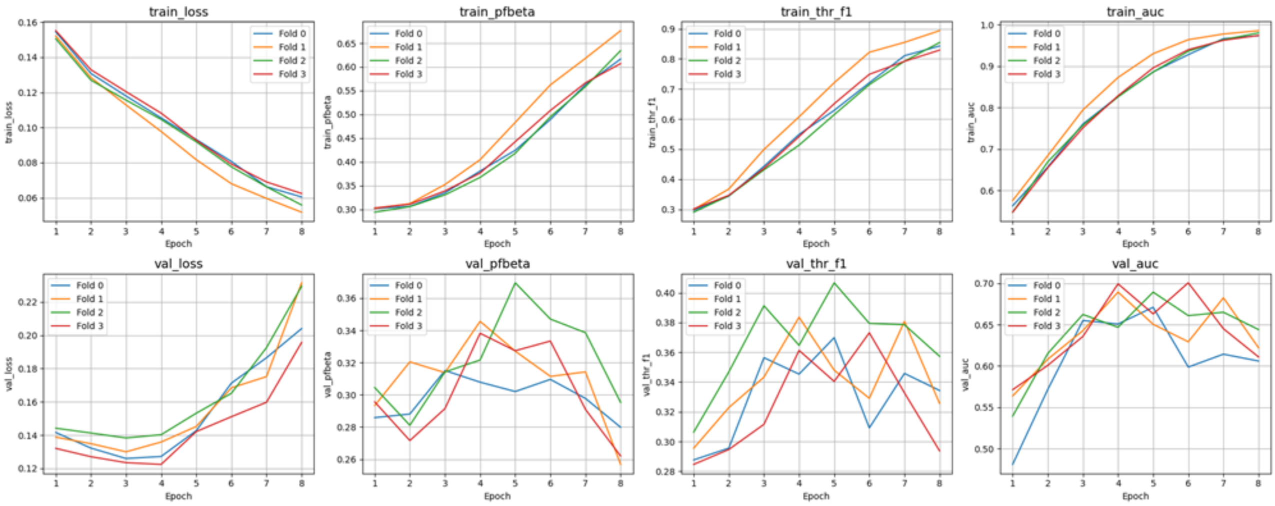

Figure 4: Training result of best version of CNN + AUX model

Several insights emerged from our experiments that point to avenues for future improvements. First, architectural modifications, especially those that incorporate domain-specific auxiliary features, offer promising gains. Second, enhancements in image processing, such as refining the region of interest (ROI), can help focus learning on clinically significant areas. Grad-CAM visualizations supported this by showing that more accurate models tend to concentrate on relevant breast tissue regions. Fourth, we would want the Grad-CAM visualizations to focus more on the breast tissue regions, and not on the regions around it (which did happen on occasion). Fifth, after epoch 4, the validation loss is rising which suggests overfitting (this could be tackled by e.g., dropouts, learning rate scheduler, etc.). Finally, our results reaffirm the importance of both data volume and quality; diverse and well-labeled datasets remain fundamental to training models that generalize effectively across patient populations.

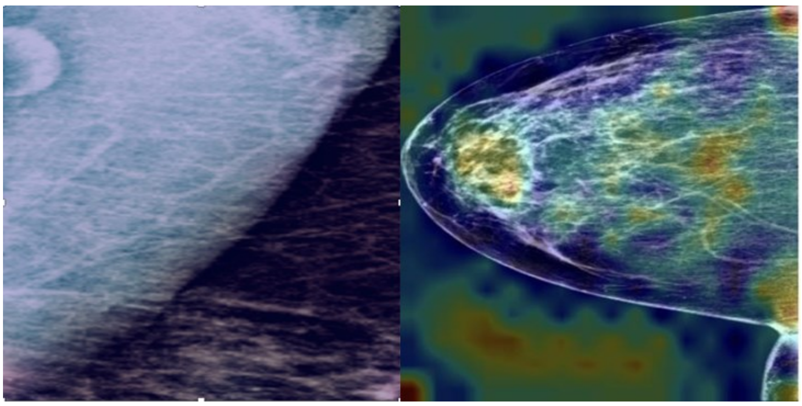

Figure 5: Examples of bad ROI and unwanted Grad-CAM visualizations

8 Conclusion & Reflection

This project has been a valuable learning opportunity, offering deep insights into the challenges and potential of using AI for the medical sector (i.e., in our case: breast cancer detection). Working at the intersection of medicine and machine learning has highlighted not only the technical complexities but also the profound impact this work can have on early diagnosis and patient care. A certain level of domain knowledge of both areas is crucial to perform better than average.

This field is both interesting and important, with the potential to transform healthcare delivery, especially in underserved regions where access to expert radiologists is limited. By integrating AI and machine learning into medical diagnostics, we can take meaningful steps toward achieving Sustainable Development Goals (SDGs) related to good health and well-being, reduced inequalities, and innovation.

Another learning for us is that the addition of auxiliary targets shows potential to improve the CNN model's performance, particularly when their influence is kept small through lower auxiliary loss weights (just enough to supplement the main task of cancer classification without overwhelming it). Our results indicate that smaller weights (e.g., 10-30%) can slightly boost F1 score, implying that auxiliary targets help guide the shared backbone to learn more structured representations. However, at higher weights, we observed a drop in F1 score even when accuracy increased. This rise in accuracy can be misleading, as it may reflect the model's bias toward the majority class (in our case “no cancer”) rather than true improvements in cancer detection. This reinforces the importance of using F1, and not accuracy, as the main evaluation metric for our task. Since F1 balances precision and recall, it better captures how well the model identifies true positives, which is essential in medical applications where missing a cancer case has serious consequences.

Looking ahead, several promising directions can further enhance this work. Refining the model architecture, experimenting with data augmentation techniques, and improving preprocessing steps could lead to more accurate and robust predictions. Incorporating our auxiliary targeting approach through multi-modal learning into the “best”-performing convolutional neural network (CNN) could significantly boost model performance by leveraging additional contextual data. Continued exploration and iteration will not only advance the technical quality of the models but also bring us closer to creating accessible, reliable tools that can support medical professionals and improve outcomes for patients worldwide.

9 List of Abbreviations

AUX	Auxiliary
CNN	Convolutional neural network
ROI	Region of interest
WHO	World Health Organization

10 List of Figures

11 List of References / Bibliography

[1] "Master of Science in IT, Digitalization & Sustainability," Hochschule Luzern, bildet – forscht – begeistert. Accessed: Feb. 19, 2025. [Online]. Available: https://www.hslu.ch/de-ch/informatik/studium/master/information-technologies-digitalization-und-sustainability/

[2] "RSNA Screening Mammography Breast Cancer Detection." Accessed: May 17, 2025. [Online]. Available: https://kaggle.com/rsna-breast-cancer-detection

[3] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," Sep. 11, 2020, *arXiv*: arXiv:1905.11946. doi: 10.48550/arXiv.1905.11946.

[4] "RSNA Screening Mammography Breast Cancer Detection | Kaggle." Accessed: May 17, 2025. [Online]. Available: https://www.kaggle.com/competitions/rsna-breast-cancer-detection/data

[5] "Improved EfficientNet Architecture for Multi-Grade Brain Tumor Detection." Accessed: May 17, 2025. [Online]. Available: https://www.mdpi.com/2079-9292/14/4/710#B13-electronics-14-00710

[6] G. Marques, D. Agarwal, and I. de la Torre Díez, "Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network," *Appl. Soft Comput.*, vol. 96, p. 106691, Nov. 2020, doi: 10.1016/j.asoc.2020.106691.

[7] R. Raza *et al.*, "Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images," *Eng. Appl. Artif. Intell.*, vol. 126, p. 106902, Nov. 2023, doi: 10.1016/j.engappai.2023.106902.

[8] "Improved EfficientNet-B4 for Melanoma Detection | IEEE Conference Publication | IEEE Xplore." Accessed: May 17, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/9389915

[9] Pak et al., *EfficientNet-B4 - Distribution of images across classes*. 2020. Accessed: May 16, 2025. [Online]. Available: https://www.researchgate.net/figure/Architecture-of-efficientNet-B4-model_fig2_343654628