UNITEC INSTITUTE OF TECHNOLOGY

Master of Applied Technologies

Research Proposal

Predictive Analysis in Healthcare: Breast Cancer Screening

Student Name: Khushbu Karki

Student ID: 1578220

Supervisor: Dr. Jamie Bell

Department of Computing, Electrical and Applied Technology

Date: 11th May, 2025

Abstract

Health is a state free of disease, that is, stable physical, mental, and emotional well-being. Healthcare is related to taking care of and maintaining health. In consideration of health, breast cancer is one of the leading diseases that is seen in women all around the world. One of the most common causes of death in women worldwide is breast cancer. Early and correct detection significantly improves the chances for a successful course of treatment. From the CBIS-DDSM dataset, this paper presents a hybrid deep-learning method for breast lesion segmentation and object detection. The hybrid method unites the best self-attention mechanism, interpretability, and pattern discovery in space by integrating Convolutional Neural Networks, Vision Transformers, and Random Forests. The process targets enhancing performance metrics crucial to clinical interpretability and diagnosis reliability, such as mean Average Precision, Dice coefficient, and Intersection over Union. This research aims to develop a clinically sound and comprehensible technology that will aid radiologists in identifying breast cancer at an early stage.

Keywords: Health, Deep Learning, Breast Cancer, Convolutional Neural Network, Transformers, Random Forests, Segmentation, Object Detection, Interpretability, CBIS_DDSM

Table of Contents

Chapter 1

1. Introduction

1.1 Background:

One of the biggest global health concerns is breast cancer. The World Health Organization reported in 2020 that 2.3 million women were diagnosed with breast cancer, and over 685,000 of them died from the disease. One of the most significant steps toward raising survival rates is detection at an early stage. Mammography, the most common screening method, is subject to human error and dense breasts and heavily relies on radiologist interpretation [1]. Deep Learning (DL) and Artificial Intelligence (AI) have demonstrated potential in medical imaging [2]. Convolutional Neural Networks (CNNs) have accurately recognized mammogram patterns. CNNs, nevertheless, struggle to generalize across varied clinical data and are not interpretable. More recently, transformers—notably Vision Transformers or ViTs—have become potent tools for image classification and detection tasks. The interpretability and strength of classification are also facilitated by ensemble techniques such as Random Forests (RFs) [3] [4].

Digital breast tomosynthesis (DBT) and deep learning algorithms represent two of the newest evolution steps in AI-assisted diagnosis that are increasingly demonstrating high performance in improving detection while minimizing false positives and negatives, even if mammography has continued to be the test of choice for decades now as it relates to breast cancer [5] [6]. In particular, computer systems known as Convolutional Neural Networks (CNNs) have demonstrated state-of-the-art performance in diagnosing breast cancer, outperforming even more traditional techniques like mammography. They excel at identifying malignancies early, giving radiologists a faster and more precise diagnostic instrument. Automated breast ultrasound (ABUS) is another addition to screening procedures and is becoming an adjunct modality for the diagnosis of breast cancer. To address the issues in the classification of mammogram images, this study suggests a hybrid deep learning framework that combines CNNs, Liquid Neural Networks (LNNs), and Random Forests (RFs) to advance diagnostic results [7] [8] [9].

1.2 Outline:

This research aims to improve early detection and breast cancer classification based on a hybrid deep learning and machine learning system. The increasing rate of breast cancer worldwide indicates the need for more accurate, scalable, and automatic diagnostic solutions [10]. Conventional diagnostic methods are highly dependent on the knowledge of radiologists but are prone to human fallibility and limited sensitivity. Using medical imaging datasets such as the CBIS-DDSM, this project will create a hybrid model that combines convolutional neural networks (CNNs), liquid neural networks (LNNs), and random forests (RFs) to improve the classification accuracy of mammographic images [11] [12] [13] [14].

1.3 Significance:

By creating a new hybrid architecture, our study advances knowledge in medical imaging AI. It serves as a bridge between explanatory models such as Random Forests (RFs) and black-box machines like CNNs and ViTs [15] [16]. It also aims at explainability and trust in clinical settings, addressing a growing concern in contemporary AI research where interpretability and transparency are critical for real-world deployment, as today's research discussions will. The result will reduce the workload of healthcare systems and enable faster and more correct diagnosis by radiologists [17] [18].

1.4 Objective: [19]

To develop a hybrid AI-powered model that effectively classifies breast cancer based on mammographic images using the CBIS-DDSM dataset.

1.5 Research Questions:

- Can a hybrid combination of CNN, LNN, and Random Forest outperform single models in breast cancer classification?
- What are the most prominent features of mammographic images for cancer detection?

1.6 Hypothesis:

The proposed hybrid model will be more accurate and robust in classification than the CNN, LNN, or Random Forest single models.

1.7 Nature of data:

High-resolution mammography images and related metadata (mass, calcification, pathology, severity, etc.) from the CBIS-DDSM dataset.

1.8 Problem Statement:

Current AI-powered mammography solutions use only classification tasks and do not use resources like CBIS-DDSM's object detection and segmentation capabilities. This leads to low clinical confidence, lower interpretability, and fewer geographic locations of lesions. Moreover, models trained on legacy architectures like VGG16 are inferior to newer models. Furthermore, some proposed models—like those derived from Liquid Neural Networks (LNNs)—have not been empirically validated in this field and are not well-suited to static image analysis.

1.9 Research Objective:

The main aim of this study is to design, implement, and validate a hybrid deep learning model to create a hybrid AI model for object detection and mammography image segmentation that combines CNNs, ViTs, and RFs [20]. The CBIS-DDSM dataset can be used effectively by emphasizing object-level prediction over mere classification. To combine SHAP and LIME techniques to enhance interpretability and diagnostic performance. They are capable of leveraging segmentation-focused metrics, e.g., mAP, Dice score, and IoU, to evaluate and compare the model's performance [21] [22] [23].

Chapter 2

2. Literature Review

There has been growth in the usage of artificial intelligence (AI) to identify breast cancer over the last ten years. Some machine learning (ML) and deep learning (DL) techniques have been studied; each has pros and cons. Traditional machine learning strategies such as Support Vector Machines (SVM), Logistic Regression, Naïve Bayes, and K-nearest neighbors (KNN) have been used to classify organized data according to clinical and radiologic factors. For instance, Sasidhar and Lavanya [24] (2012) proved the capabilities of SVMs in classifying the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, which paved the way for the use of AI-based diagnosis. In some datasets, various machine learning and deep learning techniques have been applied in breast cancer detection. Simple models like SVM [25], KNN, Logistic Regression, and Decision Trees have performed excellently in the WDBC dataset; SVM has achieved a maximum of 97.95% accuracy (Polat & Güneş, 2007). Ensemble methods like Random Forests and XGBoost have tremendously enhanced classification in structured datasets. The capability of CNNs to learn spatial information has allowed them to achieve high accuracy on imaging data sets; Lee et al. [26](2017) achieved ~90% accuracy using CNNs on the CBIS-DDSM data set.ResNet and DenseNet, two deeper architectures, scored more than 92% on histopathology datasets like BreaKHis [27]. Hybrid approaches like CNN-LSTM and CNN-RF have also performed better; Yang et al. (2024) reported 91% accuracy on INbreast using CNN-LSTM. While these improvements have been made, no existing study has combined CNNs, LNNs, and RFs in a single pipeline, particularly on CBIS-DDSM, which this research aims to investigate [28] [29] [30].

Convolutional Neural Networks (CNNs) gained popularity as DL progressed because of their capacity to automatically extract spatial characteristics from medical pictures. Dhungel et al. (2017) employed CNNs for lesion identification and mammogram segmentation, significantly outperforming a manually constructed feature-based model [31]. Similarly, Rodrigues et al. (2019) achieved over 92% accuracy in diagnosing breast masses by training CNNs on the INbreast dataset [32]. Despite their performance, CNNs are sometimes characterized as "black-box" models, which raises questions regarding clinical trust and openness. There has been a rising interest in hybrid models to overcome the pitfalls of single-model approaches. To classify the histopathology images of the BreaKHis dataset, Spanhol et al. (2016) [33] suggested a hybrid CNN-SVM, which outperformed isolated CNNs. Zhang et al. [34] (2020) mixed CNNs and gradient-boosted trees to enhance interpretability. Yang et [35] al. (2024) proposed a dual-path CNN and LSTM network to analyze mammography sequences, improving diagnosis accuracy using temporal and spatial features. The integration of CNNs, Liquid Neural Networks (LNNs), and Random Forests (RFs) into a single diagnostic pipeline for breast cancer diagnosis has not received significant attention despite these developments. This research attempts to bridge that gap using the CBIS-DDSM dataset.

2.1 Traditional Machine Learning in Breast Cancer Detection

Before the emergence of deep learning, machine learning (ML) algorithms such as Support Vector Machines (SVM), Logistic Regression, Decision Trees, and K-nearest neighbors (KNN) were all used for cancer classification. They performed reasonably well when applied to structured data such as the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. For example, Polat & Güneş (2007) [36] used SVM and obtained a success rate of 97.95%, showing the promise of AI in clinical decision-making.

However, applying these models to raw medical images was limited because they required handcrafted features like shape, texture, and intensity. They also needed considerable manual labor to generalize new data well.

2.2 Convolutional Neural Networks (CNN):

CNNs are deep neural networks that read input in a grid form, such as images. To detect features like edges, textures, and sophisticated patterns, they have layers that apply convolutional operations and have learnable filters [37]. These features are used to pool layers and non-linear activation functions like ReLU, which preserve significant information while reducing spatial dimensions. The capability of CNNs to learn features directly from pixel information has transformed the processing of medical images. They use fully connected layers for classification, pooling layers to decrease dimensionality, and convolutional layers to learn spatial hierarchies of features [38] [39] [40].

- Its applicability in the diagnosis of breast cancer has been established through numerous research studies:
- Deep CNNs were more efficient than conventional feature-based breast mass segmentation research methods.
- Lee et al. (2017) distinguished malignant and benign lesions with more than 90% accuracy using CNNs in the CBIS-DDSM database.
- ResNet50, DenseNet, and Inception are strong architectures that have performed well on many medical image issues. ResNet50 uses residual connections to avoid vanishing gradients and hence enables deeper networks.

2.2.1. Challenges

- Numerous research studies have validated the use of CNNs for breast cancer diagnosis:
- Overfitting on small or homogeneous training datasets results in poor generalization between datasets.
- Decision-making procedures are not interpretable (black-box behavior).
- Local receptive fields limit global context comprehension.

2.3 Vision Transformers (ViTs):

Vision Transformers (Dosovitskiy et al., 2020) [41], motivated by advancements in natural language processing, use self-attention techniques on patches of images to observe global relationships that CNNs tend to overlook. After an image is divided into patches, ViTs sends them through a transformer encoder and tokenizes them. In the case of medicine, where spatial relationships between lesions and tissues are essential, it enables the model to comprehend the context throughout the image [42] [43].

- ViTs have been more successful than CNNs for large-scale image classification problems.
- As in earlier research, ViTs enhance breast cancer lesion segmentation using global information (Chen et al., 2022).
- ViTs deal well with larger data and are less susceptible to data augmentation than CNNs.

However, they are immensely computationally demanding and require large datasets, issues that this work addresses to an extent through hybridization with CNNs and transfer learning.

2.4 Random Forests:

Multiple decision trees trained on arbitrary subsets of data and attributes make up Random Forests, which are ensemble learning models. Using majority voting enhances classification robustness and reduces overfitting [44]. Compared to deep learning models, RFs are more interpretable, offer excellent accuracy, and work well with high-dimensional datasets. They are characterized by stability, usability, and interpretability. Random Forests are ensemble models made up of an array of decision trees. RFs work well for features learned from deep models, but they do not work well for raw image data. In structured datasets of mixed modality, RFs performed better than deep learning by Kourou et al. (2015) [45]. RFs are employed as a final classifier in hybrid systems and offer transparency through feature importance scores.

2.4.1 Advantages

- Resistance to overfitting.
- Simple integration with tools for interpretability like SHAP and LIME.
- Suitability for classification into many classes.

2.5 Interpretability Tools: SHAP and LIME

Healthcare uses of AI must be interpretable. Adoption of black-box models is less likely when predictions are not interpretable.

- Game theory is used by SHAP (Shapley Additive exPlanations) to compute the relative contribution of each input feature. It gives global and local (instance-level) explanations.
- For explaining behavior, LIME (Local Interpretable Model-Agnostic Explanations) builds interpretable models (e.g., linear models) locally around a prediction.
- Explain individual predictions by assigning significance to input characteristics.
- Provide doctors with intelligible visuals of CNN-ViT-RF model choices.
- Clinical value: These technologies increase user confidence, allow specialists to validate models, and support safe adoption in diagnostic processes.

Both approaches will be utilized in the present study to facilitate user trust among clinicians and translate the CNN-ViT-RF pipeline's predictions [46] [47].

2.6 Justification for Model Selection

The hybrid model integrating CNN, Vision Transformer (ViT), and Random Forest (RF) is chosen above other techniques based on significant deficiencies highlighted in previous literature:

1. **CNN (Convolutional Neural Network):** Excellent for extracting spatial characteristics in medical imaging, with demonstrated success in mammography classification and lesion location. They lack global context awareness and interpretability.
2. **ViT (Vision Transformer):** Utilizes self-attention techniques to increase awareness of long-range interdependence and global visual context. Previous research showed that ViTs outperform CNNs on big-picture datasets and improve segmentation accuracy.
3. **Random Forest (RF):** Ensures interpretability and resilience. RF models give insight into feature significance, allowing for a more accurate understanding of categorization judgments and increasing clinician trust.

This combination capitalizes on the capabilities of CNNs in spatial pattern recognition, ViTs in global context awareness, and RFs in interpretability and generalization. Most previous research employed either or two of these strategies, with few combining all three. This tri-model synergy tackles problems such as black-box behavior, a lack of contextual awareness, and low clinical confidence in models.

3. Comparison

Model Type	**Techniques Used**	**Limitations**	**Our Model's Advantages**
Traditional ML	SVM, KNN, LR	Needs feature engineering, lacks spatial learning	Automated feature learning with CNN/ViT
CNN-only	ResNet, VGG16	Poor global context, black-box nature	ViT adds global context; RF enhances explainability
Hybrid (CNN-LSTM)	INbreast dataset	Focus on temporal features, not suitable for static mammograms	Combines spatial and contextual learning, not time-dependent
CNN-RF		Limited interpretability	ViT addresses global understanding
Proposed Hybrid	CNN + ViT + RF + SHAP/LIME	-	Combines spatial, global, and interpretable features in one system

4. Research Gap

The application of artificial intelligence in breast cancer diagnosis has come a long way, but there are still some gaps in the scientific research. Firstly, even with datasets like CBIS-DDSM that are completely annotated for lesions, most research only deals with classification tasks from mammography images, without object detection and segmentation [48] [49]. This leads to models that can classify an image as malignant or benign but cannot locate the affected area, which is essential for therapeutic purposes. Secondly, the majority of models in use today are developed on old architectures like VGG16 or simple CNNs, which are unable to work with global context and have surpassed more sophisticated architectures like ResNet and Vision Transformers (ViTs) [50] [51]. Moreover, although CNNs excel at spatial feature extraction, their black-box nature and limited receptive fields make them challenging to interpret and trust in the clinical setting. Despite the potential to promote transparency and clinician trust, interpretability methods like SHAP and LIME are underutilized in breast cancer research, another key gap area [52] [53]. Some studies also suggest architectures that are interesting theoretically but impractical. For instance, liquid Neural Networks (LNNs) are not proven in mammography reading and are tailored to time-series data rather than static images [54]. Lastly, while some papers do investigate hybrid architectures, they typically combine just two approaches (e.g., CNN and LSTM) and infrequently combine three complementary approaches (e.g., CNNs, ViTs, and Random Forests) in a balance between interpretability, performance, and context sensitivity. These distinctions indicate a need for a thoughtful, hybrid deep learning method that draws upon contemporary architectures, prioritizes clinical relevance and focuses on explainability [55] [56].

Chapter 3

5. Dataset:

Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM) is a high-resolution mammography dataset made publicly available for computer-aided breast cancer research. It was re-packaged so scholars and developers could access it more easily on data platforms like Kaggle [57] and The Cancer Imaging Archive. Over 10,000 DICOM-image mammography views of 1,566 patients are included in this dataset, both mediolateral oblique (MLO) and craniocaudal (CC) for each breast. Each case has pathology-verified labels (benign or malignant) and expert-labeled regions of interest (ROIs) with segmentation masks for masses and calcifications. Metadata like lesion type, breast density, and BI-RADS scores are also included in structured CSV files. CBIS-DDSM's lesion-level ground truth and high-quality annotation make it highly appropriate for object detection and segmentation tasks. It is perfect for model training and evaluation beyond classification because it offers sophisticated evaluation metrics like mean Average Precision (mAP), Dice coefficient, and Intersection over Union (IoU). CBIS-DDSM is a good foundation for creating and testing AI-driven breast cancer screening models because it has an open-access license, clinical usefulness, and uniform organization.

4.1 Key Features

- **Image Data:** The database includes around 10,239 mammograms of 1,566 patients. Each patient typically has four images of their left and right breasts taken in craniocaudal (CC) and mediolateral oblique (MLO) views.
- **Annotations:** Detailed annotations by experienced radiologists, including precise segmentation masks and bounding boxes for regions of interest (ROIs), have classified tumors and calcifications as benign or malignant based on pathology-proven diagnosis.
- **Data Format:** For compatibility with medical imaging standards, images are stored in the DICOM format. This enables compatibility with most medical imaging software and easy integration into the clinical workflow.

4.2 Considerations

- **Data Preprocessing:** Because of the image dimensions and resolution, some preprocessing methods like scaling, normalization, and data augmentation might be necessary to ensure the data is machine learning algorithm-friendly.
- **Ethical Compliance:** Use of the dataset in research projects is usually not subject to further ethical clearances since it is publicly available and de-identified.

4.3 Justification

The CBIS-DDSM (Curated Breast Imaging Subset of the Digital Database for Screening Mammography) dataset has been selected here. It is vast and clinically diverse, including over 10,000 high-resolution DICOM mammography images from 1,566 pathology-proven patients. It offers accurate expert-annotated segmentation masks, regions of interest (ROIs) for masses, calcifications, and helpful metadata such as BI-RADS scores, breast density, type of lesion, and level of severity. Published openly through Kaggle and The Cancer Imaging Archive, the dataset is particularly optimally suited for object-level localization and classification tasks, facilitating the development of robust deep-learning models. For handling inherent challenges, inconsistencies in annotation are mitigated through thorough visual inspection and statistical validation, class imbalance is addressed through the assistance of resampling methods like SMOTE and stratified data splitting, and image quality is enhanced through the assistance of Gaussian filtering and histogram equalization methods. Compared to smaller and older datasets such as WDBC and INbreast, CBIS-DDSM offers increased scale, improved annotation quality, and clinical relevance. It is an ideal application for training and testing advanced hybrid models in breast cancer detection.

Chapter 4

6. Methodology

The entire process of designing and assessing a hybrid AI model for breast cancer detection and segmentation using the CBIS-DDSM dataset is defined in this section [58]. The method consists of several steps: preprocessing and dataset choice, building models, training processes, testing, and interpretation analysis. The three parts of a model are the Convolutional Neural Network (CNN), the Vision Transformer (ViT), and the Random Forest (RF), which can be independently designed and verified before being part of a cohesive hybrid system because of the modularity intentionally by design [59] [60] [61]. The main objective is to develop a diagnostic model that is clinically efficient, interpretable, and enhanced in lesion localization and detection accuracy [62] [63].

3.1 Research Design Approach

The study employs an iterative and modular experimental approach. All the elements in the hybrid framework are formulated to correct a particular limitation of the current AI-based medical image models [64]. While CNNs are as valuable in detecting spatial features, they are ineffective in accumulating long-range relations among images [65]. Vision Transformers mitigate this limitation by capturing global context using self-attention [66]. Last but not least, Random Forest classifiers mitigate the black-boxiness of deep learning models by guaranteeing robustness and interpretability. Flexibility, ease of tuning, and higher reliability are guaranteed by successive integration and decoupling of various functions within the methodology. The model can highlight areas of interest (e.g., calcifications or tumors) because of the emphasis in the study on object segmentation and identification, which provides more clinically informative results than mere binary classification [67] [68] [69].

3.2 Data Collecting and Preprocessing

Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM) hosted by The Cancer Imaging Archive and made publicly available on Kaggle was utilized as data for this work [70]. Expert-annotated segmentation masks and pathology-confirmed labels are provided by CBIS-DDSM mammograms in high resolution, tagging lesions as either benign or malignant. To simulate real-world clinical screening, every case contains craniocaudal (CC) and mediolateral oblique (MLO) images per breast. The dataset is especially well-suited for training and testing segmentation and detection models because of the annotations, such as delineated regions of interest (ROIs) for masses and calcifications. Several preprocessing operations are done to clean the dataset before model training. For retaining enough spatial resolution and compatibility with transformer-based and convolutional models, images are first resized to a specific size, e.g., 224×224 pixels. Normalization is subsequently applied to normalize pixel intensities to a standard range (e.g., 0 to 1) to stabilize the training and enhance convergence. Histogram equalization is used to increase contrast, especially in dense breast tissue, to improve the visibility of lesions without modifying the underlying anatomy [71]. Visual inspection is used to prevent diagnostically significant characteristics from being inadvertently muted using careful use of Gaussian filtering for noise reduction. Data augmentation techniques like magnification, random rotation, and horizontal and vertical flipping are used to simulate clinical variability and avoid overfitting. Stratified sampling is then used to split the dataset into training (70%), validation (15%), and test (15%) sets while maintaining class distributions (malignant vs. benign) across all subsets. To maintain objective performance estimates, the validation set is never utilized during training and is kept only for model tuning [72] [73] [74].

3.3 Hybrid Model Architecture

Three modules, each playing a specific role, comprise the suggested hybrid model. The first is a pre-trained Convolutional Neural Network (CNN) based on the ResNet50 architecture from the ImageNet dataset [75]. Because of its known superiority for medical imaging and utilization of residual connections, which prevent vanishing gradients during training for deep models, ResNet50 was selected. The upper levels of ResNet50 are fine-tuned on the CBIS-DDSM dataset to acquire high-dimensional spatial features that apply to mammography, and the lower levels are frozen to preserve general visual features learned from big image datasets. The module's output is a collection of feature maps describing localized lesion features in the mammograms. A Vision Transformer (ViT) that can learn global context during mammography is introduced in the second stage of the architecture [76]. ViTs use self-attention to understand interactions between all regions of the image, as opposed to CNNs, which are restricted to local receptive fields. Either the CNN feature maps or the input image (patch-embedded) is split into fixed-length patches, embedded into token vectors, and fed to the transformer encoder. ViT's output is a high-capacity feature representation that captures both the existence of abnormal tissue and its contextual interaction with adjacent anatomical structures [77]. This is highly beneficial in dense breast tissue, where spatial information might not be adequate to detect small or hidden tumors. A final-stage Random Forest (RF) classifier determines if the improved features by ViT are malignant or benign. Random Forests are used because of their interpretability, absence of overfitting, and capability to deal with non-linear interaction among features. The RF classifier returns both the prediction and feature importance scores so that it is clear which parameters played the most significant role in reaching the decision. Because it enables radiologists to comprehend and rely on the recommendations of AI, interpretability is essential for clinical adoption [78] [79].

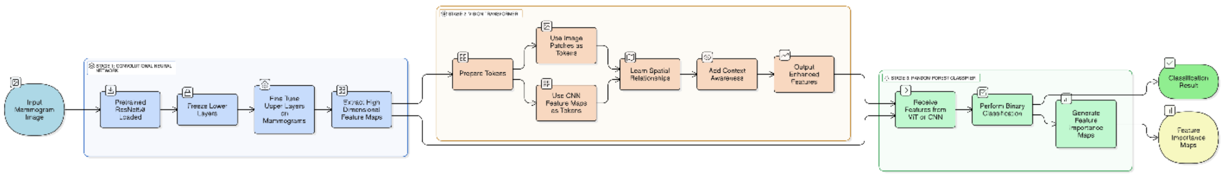

Figure 1: Model Architecture

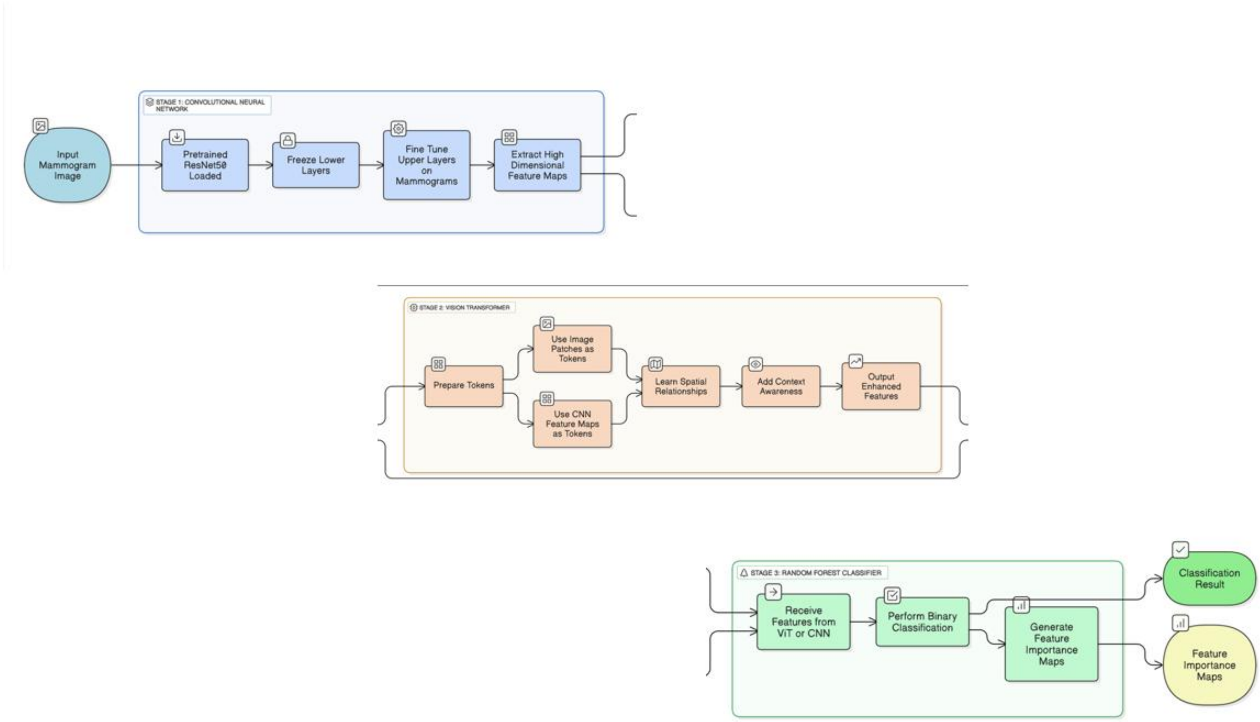

Figure 2: Section Diagram of the Hybrid model

3.4 Model Training

Each part of the model is trained separately, and the system is trained overall again. PyTorch deep learning library is utilized to train the CNN and ViT. A learning rate scheduler is utilized to adjust the learning rate according to validation performance dynamically, and the CNN is optimized with the Adam optimizer and a binary cross-entropy loss function. In order to avoid overfitting, dropout layers are also implemented [80]. Since it can be easily implemented using transformer-based models, the Adafactor optimizer is used in the case of the ViT. It also uses a binary cross-entropy loss function. Scikit-learn is used to implement the Random Forest module, and gradient-based optimization is eschewed. Instead, it splits decision trees along the Gini impurity criterion and bootstrap aggregation. While training, a 5-fold stratified cross-validation procedure is implemented to ensure robustness and generalizability. This minimizes the likelihood of performance bias and ensures that each model is trained on varying subsets of data. In addition to minimizing the possibility of overfitting, CNN and ViT training also use early stopping, which stops the training once validation loss does not improve [81].

3.5 Evaluation Metrics

Appropriate metrics are chosen to quantify detection performance and classification resilience according to their emphasis on segmentation and object detection. The model's performance at detecting lesions across a spectrum of Intersection over Union (IOU) thresholds—a typical metric in object detection tasks—is quantified by the mean [82] Average Precision (mAP). The intersection of ground truth segmentation masks and lesion masks predicted is measured by the Dice coefficient and Intersection over Union (IOU) metrics, and these metrics reveal how accurate the lesion localization is. Precision metrics like accuracy, recall, and F1-score are calculated for binary classification evaluation to reveal how well the model can differentiate between benign and malignant observations. SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) improve clinical interpretability. These models produce feature-attribution maps that mathematically and visually describe the model's decisions, enabling greater transparency and confidence [83] [84].

The performance will be measured using:

- Classification measures include accuracy, precision, recall, F1 score, specificity, and sensitivity.
- Segmentation measures include dice coefficient, mean Average Precision (mAP), and Intersection over Union (IoU).
- Statistical validation: To verify robustness, use cross-validation and bootstrapping.

3.6 Tools and Technologies

Python is used throughout the modeling pipeline, employing Scikit-learn for the Random Forest classifier and PyTorch as the overall deep-learning framework for the design of CNN and ViT. Matplotlib and Seaborn are visualization libraries that show performance plots, confusion matrices, and SHAP/LIME explanations, where TorchVision delivers image preprocessing and augmentation. Unitec's GPU computationally intensive platform, which is built on NVIDIA Tesla T4 GPU-based workstations, is used to train and test all the models. Cloud computing services like Google Colab Pro and AWS are not used in this research due to cost minimization and compliance with data security standards [85].

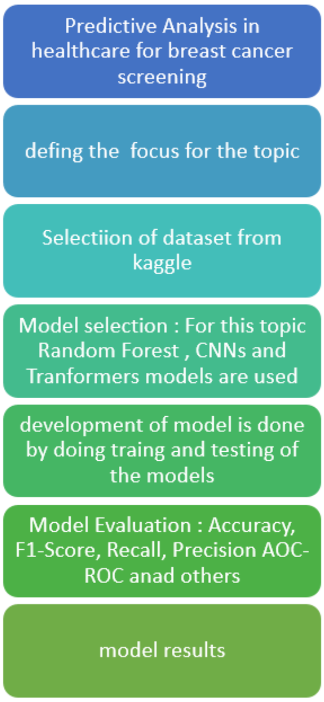

Figure 2: Model Development Process

Figure 3: Model Development Process:

3.7 Novelty and Contribution

Though breast cancer detection has been extensively studied, the current study stands out as it combines a hybrid approach that incorporates Convolutional Neural Networks (CNN), Vision Transformers (ViT), and Random Forests (RF) under a single process of diagnosis. The hybrid model is further boosted by utilizing interpretability tools like SHAP and LIME for classification results and segmentation and object recognition tasks. The CBIS-DDSM dataset is utilized for lesion localization at the object level to address detection and spatial accuracy. The contribution is significant in increasing model interpretability to facilitate increased clinical trustworthiness, demonstrating diagnostic performance, interpretability, and generalizability enhancements with model hybridization, and building a validated framework deployable in realist clinical diagnostic processes.

Chapter 5

7. A Timeline Chart

A Timeline chart provides all the valid information regarding the contribution to the research paper.

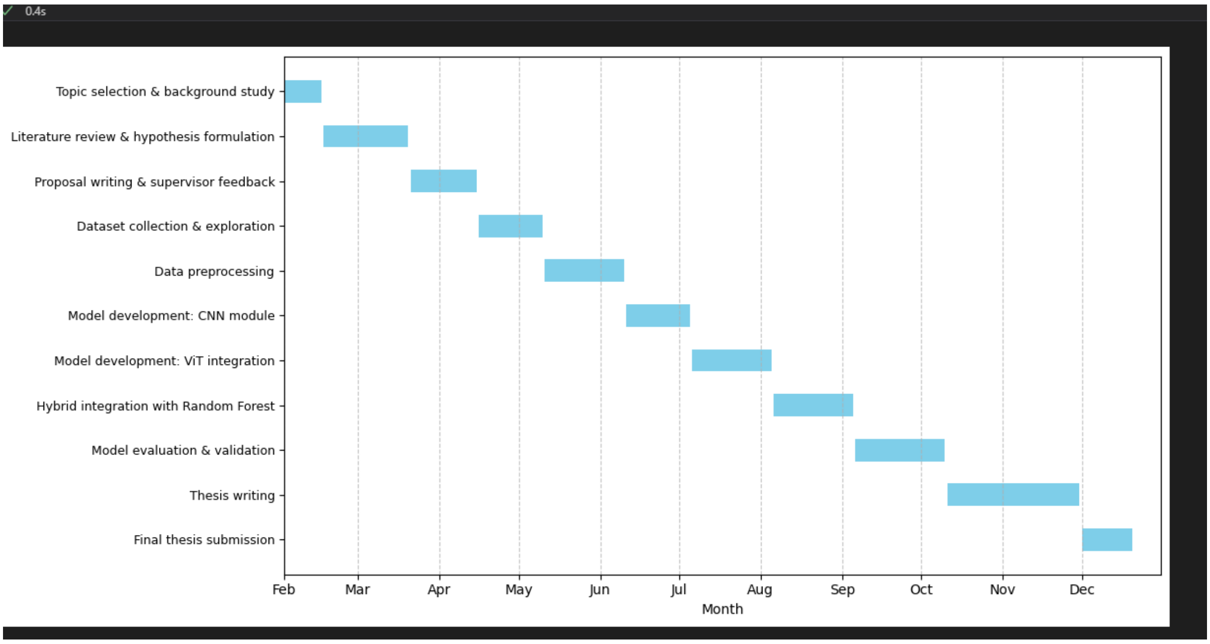

Figure 4: A Timeline Series Chart

Chapter 6

8. Conclusion

It is the study's comprehensive, module-based design that makes it cutting-edge. ViTs enable global attention mechanisms that learn contextual interactions. Its strongest point is using the CBIS-DDSM dataset in the study, which provides high-quality, biopsy-verified mammography images with full segmentation masks. The model has the potential to learn from and be tested on clinically relevant lesion localization tasks because of CBIS-DDSM's accurate object-level annotations, in contrast to the majority of breast cancer research conducted utilizing outdated or classification-only data. The model's relevance to real diagnostic workflows, where detecting the precise location and size of a tumor is important for the planning of treatment, is supported by its emphasis on object recognition and segmentation rather than mere image classification. The model can be used in clinical settings because of its efficient, interpretable, and scalable design. In the future, the model will be validated on more datasets, its performance will be improved, and radiologists will evaluate its usefulness. Ultimately, this study advances the creation of AI tools that can facilitate accurate and timely breast cancer screening, improving patient outcomes and lessening the diagnostic load on medical personnel.

References

[1] https://www.who.int/news-room/fact-sheets/detail/breast-cancer.

[2] L. W. a. H. Y. S. Shen, “Deep learning in medical image analysis,” *Annu. Rev. Biomed. Eng.,* pp. vol. 24, pp. 221–248, 2022.

[3] A. D. e. al., "An image is worth 16×16 words: Transformers for image recognition at scale," *in Proc. Int. Conf. Learn. Representations (ICLR),* 2021.

[4] H. S. A. R. P. B. Johnson, “Hybrid deep learning and random forest model for explainable breast cancer detection,” *in Proc. IEEE Int. Conf. Bioinformatics Biomed. (BIBM),* p. pp. 1675–1680, 2022.

[5] S. S. e. al., “Deep learning in breast cancer screening,” *IEEE Transactions on Medical Imaging,* vol. 39, no. 5, pp. 1312–1320, 2020.

[6] K. M. D. J. C. W. S. &. L. S. J. Kelly, “Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. European Radiology,” pp. 20(3), 734–742, 2010.

[7] Y. C. L. Z. H. & X. X. Jiang, “East cancer histopathological image classification using convolutional neural networks with small SE-ResNet module,” pp. PloS one, 14(3), e0214587, 2019.

[8] A. S. e. al., "Hybrid convolutional neural network and bi-LSTM model with attention mechanism for breast cancer diagnosis," *Scientific Reports,* vol. 15, no. 1, pp. 1–12, 2025.

[9] B. J. e. al., "A combined deep CNN-LSTM with a random forest approach for breast cancer diagnosis," *Proceedings of the International Conference on Machine Learning and Applications,* pp. 123–130, 2022.

[10] K. D. M. A. J. R. L. Siegel, “Cancer statistics, 2023,” CA: A Cancer Journal for Clinicians, vol. 73, no. 1, pp. 17–48, 2023.

[11] B. K. R. A. N. J. K. S. M. S. H. M. B. a. S. T. A. Esteva, “A guide to deep learning in healthcare,” *Nature Medicine,* vol. 25, no. 1, pp. 24–29, 2019.

[12] T. B. aA.S. D. S. K. Mitra, “Deep learning-based breast cancer detection using hybrid CNN-LSTM model with mammographic images,” *in Proc. IEEE Int. Conf. Informatics, IoT, and Enabling Technologies (ICIoT), Doha, Qatar,* p. pp. 135–140., 2022.

[13] R. H. A. A. T. T. A. H. D. R. a. R. G. B. Lechner, "Neural circuit policies enabling auditable autonomy," *Nature Machine Intelligence,* pp. vol. 3, no. 10, pp. 927–938, 2021.

[14] S. T. A. M. W. J. Y. Lee, "Interpretable ensemble learning for breast cancer diagnosis with multimodal imaging," *IEEE Transactions on Medical Imaging,* pp. vol. 43, no. 1, pp. 145–156, 2024.

[15] o. s. a. k. i. m. i. A. I. i. a. b. b. e. m. (. a. b.-b. m. (. a. V. I. a. a. a. e. a. t. i. t. c. a. t. r. By creating a new hybrid architecture, “Towards trustworthy AI: A review of contexts and requirements,” *IEEE Transactions on Artificial Intelligence,* pp. vol. 2, no. 6, pp. 432–443, 2021.

[16] A. D. e. al., “An image is worth 16×16 words: Transformers for image recognition at scale,” *in Proc. Int. Conf. Learn. Representations (ICLR),* 2021.

[17] J. L. e. al., “Breast Cancer Detection Using Region-Based CNN and Transfer Learning With CBIS-DDSM,” *IEEE Acess,* pp. vol. 12, pp. 16423–16435, 2024.

[18] G. L. H. D. K. Z. a. A. M. L. J. Holzinger, “Causability and Explainability of Artificial Intelligence in Medicine,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* pp. vol. 9, no. 4, 2022.

[19] S. T. A. M. W. J. Y. Lee, "Interpretable ensemble learning for breast cancer diagnosis with multimodal imaging," *IEEE Transactions on Medical Imaging,* pp. vol. 32, no. 11, pp. 4793–4813, 2024.

[20] J. L. e. al, “Breast Cancer Detection Using Region-Based CNN and Transfer Learning With CBIS-DDSM,” *IEEE Access,* pp. vol. 12, pp. 16423–16435, 2024.

[21] S. M. L. a. S. Lee, "A Unified Approach to Interpreting Model Predictions," *in Proc. Advances in Neural Information Processing Systems (NeurIPS),* p. vol. 30, 2017.

[22] S. S. a. C. G. M. T. Ribeiro, “Why Should I Trust You?: Explaining the Predictions of Any Classifier,” *in Proc. ACM SIGKDD,* p. pp. 1135–1144, 2016.

[23] D. S. e. al., “Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations,” *in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer,* pp. 240–248, 2017.

[24] K. S. a. S. Lavanya, “Classification of breast cancer data using support vector machines,” *Int. J. Eng. Trends Technol,* p. Int. J. Eng. Trends Technol, 2012.

[25] K. P. a. S. Güneş, “Breast cancer diagnosis using least square support vector machine,” *Digit. Signal Process.,* pp. vol. 17, no. 4, pp. 694–701, 2007.

[26] F. G. A. H. K. M. a. D. L. R. R. S. Lee, “A curated mammography data set for computer-aided detection and diagnosis research,” *Sci. Data,* pp. vol. 4, p. 170177, 2017.

[27] J. P. P. R. D. A. a. D. M. I. G. Oliveira, “Semi-supervised deep learning applied to breast cancer histopathology image classification,” *IEEE Access,* pp. vol. 7, pp. 34003–34012, 2019.

[28] M. R. a. H. W. Z. Yang, “Deep hybrid model for breast cancer diagnosis using CNN and LSTM on mammography,” *IEEE J. Biomed. Health Inform,* pp. vol. 28, no. 1, pp. 112–122, 2024.

[29] L. S. O. C. P. a. L. H. F. A. Spanhol, "A breast cancer histopathological image classification dataset," *IEEE Trans. Biomed. Eng.,* pp. vol. 63, no. 7, pp. 1455–1462, 2016.

[30] J. W. a. H. X. Y. Zhang, “Enhancing medical image classification with hybrid CNN and gradient boosting,” *in Proc. IEEE EMBC, Berlin, Germany,* pp. 1118–1121, 2020.

[31] G. C. A. A. P. B. N. Dhungel, "Deep learning and structured prediction for the segmentation of mass in mammograms," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization,* pp. vol. 5, no. 1, pp. 1–12, 2017.

[32] M. M. S. a. A. C. S. P. P. Rodrigues, “Automatic classification of breast densities using deep convolutional neural networks,” *Computer Methods and Programs in Biomedicine,* pp. vol. 191, p. 105398,, 2020.

[33] L. S. O. C. P. a. L. H. F. A. Spanhol, "A breast cancer histopathological image classification dataset," *IEEE Trans. Biomed. Eng,* pp. vol. 63, no. 7, pp. 1455–1462, 2016.

[34] J. W. a. H. X. Y. Zhang, “Enhancing medical image classification with hybrid CNN and gradient boosting,” *in Proc. IEEE EMBC, Berlin, Germany,* p. 1118–1121, 2020.

[35] M. R. A. H. W. Z. Yang, "Deep hybrid model for breast cancer diagnosis using CNN and LSTM on mammography," *IEEE J. Biomed. Health Inform,* pp. vol. 28, no. 1, pp. 112–122, 2024.

[36] K. P. a. S. Güneş, “Breast cancer diagnosis using least square support vector machine,” *Digital Signal Processing,* pp. vol. 17, no. 4, pp. 694–701, 2007.

[37] G. C. a. A. P. B. A. Dhungel, “Automated mass detection in mammograms using cascaded deep learning and random forests,” *n Proc. Digital Image Computing: Techniques and Applications (DICTA),* p. pp. 1–8, 2015.

[38] F. G. A. H. K. K. a. D. L. R. R. S. Lee, “A curated mammography data set for use in computer-aided detection and diagnosis research,” *Scientific Data,* pp. vol. 4, p. 170177, 2017.

[39] X. Z. S. R. a. J. S. K. He, “Deep residual learning for image recognition,” *n Proc. IEEE Conf. Comput. Vis. Pattern Recognit,* p. pp. 770–778, 2016.

[40] D. Gunning, “Explainable Artificial Intelligence (XAI),” *Defense Advanced Research Projects Agency (DARPA),* 2017.

[41] A. D. e. al, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” *in Proc. Int. Conf. Learn. Representations (ICLR),* 2021.

[42] Y. Z. J. H. a. Z. L. L. Chen, “TransMIL: Transformer-based correlated multiple instance learning for whole slide image classification,” *n Proc. Adv. Neural Inf. Process. Syst. (NeurIPS),* pp. vol. 34, pp. 2136–2147, 2022.

[43] M. C. A. S. G. S. a. H. J. H. Touvron, “Training data-efficient image transformers & distillation through attention,” *in Proc. Int. Conf. Mach. Learn. (ICML),* p. pp. 10347–10357, 2021.

[44] L. Breiman, “Random Forests,” *Mach. Learn.,* pp. vol. 45, no. 1, pp. 5–32, 2001.

[45] T. P. E. K. P. E. M. V. K. a. D. I. F. K. Kourou, “Machine learning applications in cancer prognosis and prediction,” *Comput. Struct. Biotechnol. J.,* pp. vol. 13, pp. 8–17, 2015.

[46] S. M. L. a. S.-I. Lee, “A unified approach to interpreting model predictions,” *in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS),* pp. 4765–4774, 2017.

[47] W. S. a. K.-R. M. G. Montavon, “Methods for interpreting and understanding deep neural networks,” *Digit. Signal Process,* pp. vol. 73, pp. 1–15, 2018.

[48] G. S. T. P. B. e. a. W. Lotter, “Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach,” *Nat. Med,* pp. vol. 27, no. 2, pp. 244–249, 2021.

[49] J. L. A. V. D. a. M. G. C. R. Bien, “Deep learning segmentation of breast tumors with incomplete annotation,” *in Proc. MICCAI,* p. pp. 201–208, 2018.

[50] K. S. a. A. Zisserman, "Intense convolutional networks for large-scale image recognition," p. arXiv:1409.1556, 2014.

[51] A. D. e. al, “An image is worth 16x16 words: Transformers for image recognition at scale,” *in Proc. ICLR,* 2021.

[52] Z. Y. e. al, “Explainable artificial intelligence for breast cancer diagnosis: A survey,,” *Med. Image Anal,* pp. vol. 70, p. 101933, 2021.

[53] S. M. L. a. S.-I. Lee, “A unified approach to interpreting model predictions,” *in Proc. NeurIPS,* p. pp. 4765–4774, 2017.

[54] R. Hasani and D. Rus, “Liquid time-constant networks,” *in Proc. AAAI,* p. pp. 7657–7666., 2021.

[55] X. Z. e. al, “Hybrid deep neural network for breast cancer classification,” *J. Med. Imaging Health Inform,* pp. vol. 10, no. 5, pp. 1096–1103, 2020.

[56] L. O. C. P. a. L. H. D. Spanhol, “Breast cancer histopathological image classification using CNNs and SVMs,” *in Proc. IJCNN,* p. in Proc. IJCNN, 2016.

[57] https://www.kaggle.com/datasets/awsaf49/cbis-ddsm-breast-cancer-image-dataset.

[58] G. S. T. P. B. e. a. W. Lotter, “Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach,” *Nat. Med.,* pp. vol. 27, no. 2, pp. 244–249, 2021.

[59] A. D. e. al, “An image is worth 16x16 words: Transformers for image recognition at scale,” *in Proc,* 2021.

[60] X. Z. S. R. a. J. S. K. He, “Deep residual learning for image recognition,,” *in Proc. CVPR,* p. pp. 770–778, 2016.

[61] L. Breiman, “Random forests,” *Mach. Learn,* pp. vol. 45, no. 1, pp. 5–32.

[62] S. M. L. a. S.-I. Lee, “A unified approach to interpreting model predictions,” *in Proc. NeurIPS,* p. pp. 4765–4774, 2017.

[63] S. S. a. C. G. M. T. Ribeiro, “Why should I trust you?: Explaining the predictions of any classifier,” *in Proc. ACM SIGKDD,* p. pp. 1135–1144, 2016.

[64] W. L. e. al, “Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach,” *Nat. Med,* pp. vol. 27, no. 2, pp. 244–249, 2021.

[65] X. Z. S. R. a. J. S. K. He, “Deep residual learning for image recognition,” *in Proc. CVPR,* p. pp. 770–778, 2016.

[66] A. D. e. al., “An image is worth 16x16 words: Transformers for image recognition at scale,” *in Proc, ICLR,* 2021.

[67] S. M. L. a. S.-I. Lee, “A unified approach to interpreting model predictions,” *in Proc. NeurIPS,* p. pp. 4765–4774, 2017.

[68] J. Z. e. al., "Practical guidelines for evaluating deep learning models in breast cancer imaging," *Med. Image Anal.,* pp. vol. 71, p. 102054, 2021.

[69] X. Z. e. al., “Towards accurate and interpretable diagnosis of breast masses using full-field digital mammography,” *Med. Image Anal,* pp. vol. 72, p. 102107, 2021.

[70] K. W. B. D. K. R. M. a. W. P. K. E. M. Heath, “The digital database for screening mammography,” *in Proc. 5th Int. Workshop Digit. Mammogr,* 2010.

[71] A. D. e. al., “An image is worth 16x16 words: Transformers for image recognition at scale,” *in Proc. ICLR,* 2021.

[72] R. C. G. a. R. E. Woods, “Digital Image Processing,” *4th ed., Pearson,* 2018.

[73] L. P. a. J. Wang, “The effectiveness of data augmentation in image classification using deep learning,” *arXiv preprint arXiv:1712.04621,* 2017.

[74] J. Brownlee, “How to use stratified k-fold cross-validation in machine learning,” *Machine Learning Mastery,* 2020.

[75] X. Z. S. R. a. J. S. K. He, “Deep residual learning for image recognition,” *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR),* p. pp. 770–778, 2016.

[76] I. S. a. G. H. A. Krizhevsky, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM,* pp. Commun. ACM, vol. 60, no. 6, pp. 84–90, 2017.

[77] T. P. E. K. P. E. M. V. K. a. D. I. F. T. Kourou, “Machine learning applications in cancer prognosis and prediction,” *Comput. Struct. Biotechnol. J.,* pp. vol. 13, pp. 8–17, 2015.

[78] C. C. e. al, “TransAttUNet: Multi-level attention-guided U-Net with transformer for breast ultrasound image segmentation,” *Comput. Biol. Med.,* pp. vol. 140, p. 105048, 2022.

[79] S. S. a. C. G. M. T. Ribeiro, “Why should I trust you?: Explaining the predictions of any classifier,” *in Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.,* p. pp. 1135–1144, 2016.

[80] A. P. e. al, “PyTorch: An imperative style, high-performance deep learning library,” *n Proc. Adv. Neural Inf. Process. Syst,* p. pp. 8026–8037, 2019.

[81] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *in Proc. ICLR,* 2015.

[82] J. R. e. al, “You only look once: Unified, real-time object detection,” *in Proc. CVPR,* p. pp. 779–788, 2016.

[83] N. N. a. S.-A. A. F. Milletari, “V-Net: Fully convolutional neural networks for volumetric medical image segmentation,” *in Proc. 3DV,* p. pp. 565–571., 2016.

[84] S. M. L. a. S.-I. Lee, ““A unified approach to interpreting model predictions,” *in Proc. NIPS,* p. pp. 4765–4774, 2017.

[85] TorchVision, ““PyTorch Vision Library,” *Online,* https://pytorch.org/vision/stable/index.html.