Breast Cancer Prediction Using Machine Learning Algorithms

Dana Freihat
Faculty of Artificial Intelligence
Al Balqaa Applied University
32209303107@std.bau.edu.jo

Layan Al-Hiassat
Faculty of Artificial Intelligence
Al Balqaa Applied University
32109303514@std.bau.edu.jo

Alaa Al-Qudah
Faculty of Artificial Intelligence
Al Balqaa Applied University
32109303014@std.bau.edu.jo

Mai Al-Omari
Faculty of Artificial Intelligence
Al Balqaa Applied University
32009303507@std.bau.edu.jo

Instructed By:

DR. Hussein Issam Al Ahmer

This Project (**Graduation Project II**) is Submitted in Partial Fulfillment of the Requirements for Bachelor Degree Artificial Intelligence and Robotics

Faculty of Artificial Intelligence

Al-Balqa Applied University

Al-Salt- Jordan | 2025

DEDICATION

This work is dedicated to countless individuals and families affected by breast cancer. Their strength and resilience inspire progress and innovation in early detection and treatment. It is also dedicated to healthcare providers and scientists whose unwavering commitment drives progress in medical science and to everyone who is contributing to making a difference in the fight against this disease.

ACKNOWLEDGMENTS

First and foremost, we extend our deepest gratitude to our families and friends for their unwavering encouragement, patience, and support throughout the course of this project. Their belief in us has been a cornerstone of our perseverance and success, and without their continuous motivation, this accomplishment would not have been possible.

We are also profoundly thankful to DR. Hussein Issam Al Ahmer, our esteemed supervisor, whose invaluable guidance, insightful mentorship, and consistent support have been instrumental at every stage of this journey, from initial ideation to final execution. His expertise and encouragement greatly enriched the quality and depth of our work.

Furthermore, we would like to express our sincere appreciation to the Department of Artificial Intelligence at Al-Balqa Applied University for fostering an enriching academic environment and providing essential resources that contributed to the successful realization of this project.

Lastly, we are especially grateful to our dedicated project team members, Layan Al-Hiassat, Alaa Al-Qudah, Mai Al-Omari, and Dana Freihat, whose commitment, teamwork, and valuable contributions have significantly enhanced the quality and outcomes of this work.

Table of Contents

ABSTRACT

Breast cancer remains one of the most lethal diseases worldwide, though early detection has been shown to significantly improve survival rates and patient outcomes. Conventional diagnostic methods—including mammography, ultrasound, and biopsy—though effective, are often limited by high false-positive and false-negative rates, reliance on subjective interpretation, and invasive procedures.

This research proposes a novel AI-powered diagnostic system for the early detection of breast cancer using MRI images. We evaluate and compare multiple state-of-the-art deep learning architectures, including Convolutional Neural Networks (CNNs), Vision Transformers (ViTs), Kolmogorov–Arnold Networks (KANs), and DenseNet-201. CNNs form the foundational approach, leveraging convolutional and pooling layers to mimic human visual processing for hierarchical feature extraction and classification. While CNNs offer computational efficiency and robustness to spatial variations through weight sharing and localized feature detection, they suffer from high data requirements, substantial computational costs, and limited interpretability—factors that hinder their clinical adoption.

To address these limitations, this study incorporates explainable AI techniques and modern architectures like ViT and KAN to improve global contextual understanding and model transparency. By enhancing both accuracy and interpretability, this work aims to bridge the gap between AI performance and clinical trustworthiness.

The ultimate goal is to deliver a scalable, non-invasive decision-support tool that enables healthcare professionals—particularly in resource-constrained settings—to make accurate, efficient, and timely diagnostic decisions.

Chapter One: Introduction

1.1 Overview

Breast cancer remains the dominant global health threat for women around the world. Breast cancer holds the highest incidence rate among all cancers, and it is now the most diagnosed type of cancer in women from every economic background. Breast cancer affects both genders but impacts women more severely because of hormonal, biological, and genetic differences. In 2024, the United States documented more than 310,000 cases of invasive breast cancer in women and 2,790 cases in men, as reported by the American Cancer Society [1]. This disease claimed the lives of approximately 42,000 women and 530 men during the same year [2].

Table 1. Global Breast Cancer Statistics Snapshot[3]

Country/Region	Annual Cases (Women)	Annual Deaths (Women)	5-Year Survival Rate
United States	310,000+	42,000+	91%
Jordan	1,743+	Approx. 400+	65% (est.)
India	270,000+	85,000+	66%
UK	55,000+	11,500+	85%
Sub-Saharan Africa	120,000+	70,000+	50%
Global Total	2.3 million+	685,000+	Varies by country

On a local level, for many years breast cancer has been the leading cancer diagnosis among women in Jordan. In 2016 new cancer diagnoses reached 10,755 cases while breast cancer represented 1,743 cases among women which translated to 36.8% of all female cancer diagnoses [4]. The statistic raises concern because Jordan's young population implies that breast cancer cases occur earlier compared to many Western nations. The lack of early screening tools along with cultural stigma about breast health combined with delays in medical consultation leads to late-stage diagnoses thereby worsening treatment results and mortality rates, these numbers highlight the urgent need for better solutions that support early detection and diagnosis [5].

Early-stage breast cancer poses greater danger because it remains silent and undetectable until later stages. Tumors can grow undetected by any symptoms which result in delayed diagnosis until the cancer becomes more advanced and difficult to treat. Treatment options become limited to invasive procedures like mastectomy and chemotherapy as the disease progresses. Medical science has made significant strides, but breast cancer still lacks a cure [6]. Current treatments work toward controlling cancer progression and reducing tumor size although they cannot yet provide a full cure particularly for cases with metastasis. The critical need for early detection and the development of innovative diagnostic tools that can identify early-stage malignancies highlights the importance of risk reduction strategies [7].

The development of breast cancer results from multiple genetic and environmental elements along with lifestyle choices. The leading risk factors for breast cancer include advanced age beyond 50 years [8] along with family history particularly involving first-degree relatives[9] together with genetic mutations such as BRCA1 and BRCA2[10] late pregnancy and prolonged estrogen exposure obesity alcohol consumption and sedentary behavior[11]. Predictive models use these risk factors to determine breast cancer risk profiles for individuals and to establish screening guidelines and preventive strategies. The Gail Model, Tyrer-Cuzick Model, and various machine learning algorithms assess personal risk with these variables and suggest early screening procedures along with genetic counseling and lifestyle modification to lower the overall risk burden.

Regular screening remains the strongest method for preventing late-stage diagnoses. The World Health Organization (WHO)[12] and the Centers for Disease Control and Prevention (CDC)[13] advise women who are 40 years of age or older to get annual mammograms while those at high risk due to genetic predispositions should undergo MRI scans. The interpretation of imaging data remains difficult because experienced radiologists can miss the fine differences between benign and malignant growths which causes diagnostic errors like false positives and negatives together with delays in diagnosis [14]. Deep learning models show great potential through their ongoing training with expansive datasets and their ability to detect intricate imaging patterns [15] which humans might miss. These models increase diagnostic accuracy and help clinicians make decisions while minimizing interpretation variability and enabling scalable rapid screening even in settings with limited resources. The development of Vision Transformers (ViT) and Kernel Attention Networks (KAN) represents groundbreaking advancements in breast imaging analysis [16] because they deliver superior accuracy alongside improved interpretability which proves essential for future diagnostic procedures.

The critical requirement of early detection and fast diagnosis enabled the transition to innovative technology-based solutions, predominantly through artificial intelligence (AI) [17] and machine learning (ML) [18]. These technologies excel at processing medical data faster and more consistently than human capability while enabling predictive analysis which proves essential for early detection of cancer. CNNs [19] along with Vision Transformers (ViT) [20] and Kernel Attention Networks (KAN) [21] have shown effective outcomes in breast cancer imaging by accurately identifying tumors in mammography ultrasound and MRI scans. CNNs focus on local spatial feature extraction while ViTs use self-attention mechanisms to understand global context and KANs provide symbolic interpretability which makes them ideal for medical settings that require transparency and explainability. The models use transfer learning and data augmentation techniques which enable them to perform effectively with small medical image datasets. AI-powered diagnostic models improve patient outcomes by enhancing detection precision and reducing human mistakes while speeding up clinical decisions and enabling healthcare professionals to provide timely care.

The fight against breast cancer extends beyond medical treatment because it encompasses social and economic challenges as well as emotional impacts. Late-stage breast cancer treatments generate massive costs for individual patients and national health systems [22] that lead to catastrophic healthcare spending particularly in low- and middle-income nations. Patients and their families face devastating emotional impacts when breast health discussions remain taboo in their

cultures. The promotion of non-invasive diagnostic tools supported by technology which include AI-enhanced imaging and predictive analytics can help improve early detection rates while reducing screening access inequalities and enhancing health outcomes. When national cancer control strategies incorporate these solutions, healthcare systems gain the advantage of reducing economic costs and enabling patients through interventions delivered on time and designed to be fair and culturally appropriate.

Table 2. Comparison of Diagnostic Methods for Breast Cancer.[23]

Method	Sensitivity	Specificity	Cost Efficiency	Comment
Mammography	77–95%	94–97%	Moderate	Widely used but may miss tumors in dense tissue.
Ultrasound	63–88%	73–96%	Low	Operator dependent; good for guiding biopsies.
MRI	93–100%	81–97%	High	Highly sensitive; best for high-risk patients.
Biopsy	~100%	~100%	Very High	Gold standard but invasive.
ML-Based Diagnosis	90–99%	85–98%	Improving	Non-invasive, fast, depends on data quality.

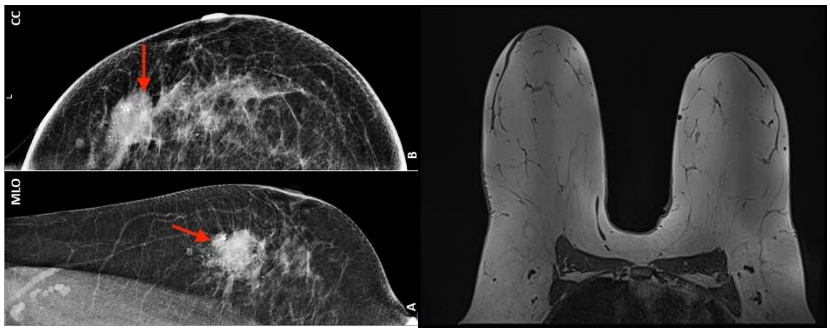

Figure 1: MRI scan comparison showing normal tissue vs. malignant tumor development in breast cancer.
Such differences can be subtle and often require computational assistance through machine learning to detect at early stages.

1.2 Project Vision and Initial Insights

The project applies modern machine learning algorithms and sophisticated image analysis methods to advance diagnostic precision while minimizing interpretation duration and supporting early detection processes. Advanced deep learning structures including CNNs, Vision Transformers (ViTs), Kernel Attention Networks (KANs), and DenseNet-201 form the backbone of this project to improve breast cancer detection through MRI imaging. While the project aims to boost performance metrics like sensitivity and specificity it simultaneously works on embedding explainable AI components to maintain transparency and interpretability throughout the decision-making process. This method streamlines diagnostic procedures while enabling healthcare providers to deliver timely patient-focused care in settings where specialized radiologists are not readily available. The project supports worldwide health goals by advancing technology-based cancer screening solutions that are fair and sustainable.

1.3 Project Objectives

1. Develop a machine learning model capable of detecting breast cancer at an early stage with high accuracy.
2. Compare multiple ML algorithms to identify the most effective approach for medical image analysis.
3. Utilize updated tools and datasets to ensure the model is efficient, scalable, and applicable to real world medical needs.
4. Support medical professionals by providing an AI-powered tool that can assist in diagnosis and reduce human error.
5. Raise awareness about the role of technology in healthcare, especially in improving early cancer detection and outcomes.

Chapter 2: Literature Survey

2.1 Introduction

So many studies have explored the application of machine learning to breast cancer diagnosis using medical images such as mammograms and MRI. This literature review explores a snippet of contributions related to our topic (Breast Cancer Prediction Using Machine Learning Algorithms), which aims to review the current state of research done on the same topic, build upon the progress achieved by previous studies, and address any potential gaps that may exist in the literature.

2.2 Existing Models and Techniques

Kiran Puttegowda et al. (2025)[24] introduced an advanced framework for breast cancer prognostic analysis which uses multiple machine learning classifiers including Support Vector Machine (SVM), Random Forest and Logistic Regression. Researchers used the METABRIC dataset which includes genomic and clinical data from more than 2,173 patients and 24,368 genes to forecast important diagnostic and prognostic parameters including progesterone receptor (PR) status tumor stage and OncoTree Code-defined cancer subtypes. The study demonstrated that Support Vector Machine with Radial Basis Function (RBF) kernel showed superior performance with 99.79% accuracy for OncoTree classification ahead of Linear SVM at 97.93% and Random Forest at 97.59%. The Logistic Regression model demonstrated suboptimal results at 89.45% accuracy which underscores its inability to capture complex non-linear relationships in high-dimensional genomic datasets. The SVM RBF model reached 93.25% accuracy in tumor staging while Random Forest and Logistic Regression produced lower results of 88.62% and 81.75% respectively. Random Forest achieved a marginally higher performance of 89.10% for PR status prediction compared to SVM RBF and Logistic Regression which both scored 88.81%, showing that basic models can effectively compete in specific classification tasks. The authors resolved class imbalance in their dataset by applying Synthetic Minority Over-sampling Technique (SMOTE) which led to better classification accuracy and balanced representation of minority classes. The Chi-Square test enabled feature selection that reduced dimensionality from 24,368 genes to 5,000 relevant features which improved model stability and generalizability in all cases. The researchers adopted ensemble learning through Voting Classifier which achieved 98.20% accuracy for OncoTree classification, 90.25% for tumor staging and 89.60% for PR status proving that predictive accuracy improves when using ensemble models instead of individual classifiers. The study recognized ongoing difficulties such as model interpretation issues, class imbalance problems, and generalization to various populations but proposed expanding deep learning models like CNN, ViT, and KAN for breast cancer prognosis advancement through imaging and genomic data integration.

Achile Solomon Egbunu (2025)[25] performed an optimization study which investigated the use of advanced machine learning algorithms to classify and detect breast cancer. The Breast Cancer

Wisconsin (Diagnostic) Data Set containing 32 features from 569 patients served as the main dataset in this study which included both malignant and benign classifications. Researchers aimed to enhance ML model accuracy through the integration of Random Forest Classifier with XGBoost Classifier and Stacking Ensemble methods. The study found that the Random Forest Classifier produced an accuracy of roughly 96%, but the XGBoost Classifier enhanced this accuracy to almost 97.5% demonstrating gradient boosting methods are better suited for structured clinical datasets. The Random Forest Classifier's Confusion Matrix demonstrated high sensitivity and specificity through True Positive rates of 95%, True Negative rates of 99%, while False Positive rates were 1.4% and False Negative rates were 4.8%. XGBoost demonstrated superior reliability for binary classification tasks in clinical applications with its Confusion Matrix revealing reduced False Positive and False Negative rates. The study implemented Stacking Ensemble Learning by integrating various classifiers together to maximize their unique capabilities which led to better performance metrics than what single models could achieve. The ensemble approach achieved highest accuracy metrics and Confusion Matrix results across all classifiers tested and demonstrated the effectiveness of ensemble methods for medical diagnostic tasks. The study did not include unsupervised learning algorithms because of the small dataset size and the necessity of labeled data for clinical classification tasks. The study identified several shortcomings in traditional diagnostic methods such as mammography, ultrasound, and MRI which show false positive rates between 4% and 34% and false negatives while also presenting exposure risks along with limitations in detecting small tumors smaller than 5mm according to Hassan et al. (2011), Berrington de Gonzalez et al. (2004), and Karthikeyan et al. (2020). The study recommends that deep learning models like CNN, ViT and neural networks should be integrated with computer vision tools including TensorFlow and Keras to advance diagnostic capabilities while enabling deployment within the U.S. healthcare system.

Wen et al. [26] (2023) developed a deep learning model to classify breast ultrasound images into benign, malignant, and normal categories using DenseNet121 with transfer learning. The model was trained on a publicly available dataset comprising 1,578 ultrasound images collected from 600 patients, as introduced by Al-Dhabyani et al. (2020)[27]. It achieved an impressive 94.94% accuracy on the test set. However, the study faced limitations, including the relatively small dataset size, potential overfitting, and lack of real-world clinical validation. Additionally, the model was trained on ultrasound images, which are often noisy and operator dependent. While the work demonstrates the potential of DenseNet121, it lacks comparison with newer architectures. Our project builds upon this by evaluating DenseNet121 alongside emerging models such as KAN and Vision Transformers using MRI data, aiming for better generalizability and interpretability.

Tchounwou et al. [28] (2021) from Jackson State University developed a machine learning-based model to help classify breast cancer cases using fine-needle aspirate (FNA) data. In their study, they worked with 569 patient samples from the Wisconsin Diagnostic Breast Cancer dataset [29], focusing on ten real-valued features like radius, perimeter, and concave points. Their results showed that geometric and texture-based features were highly effective at separating benign from malignant tumors, offering clear visual distinctions based on feature patterns. Although their work mainly uses traditional machine learning on structured datasets, it highlights how important feature extraction and data-driven approaches are for early detection. Building on this idea, our project moves a step further by using deep learning models like DenseNet, KAN, and Vision Transformers

applied to MRI scans, aiming to improve accuracy, generalization, and clinical application compared to classical methods.

Jaamour et al.[30] (2023) explored several deep learning models for breast cancer detection in mammography images using a divide and conquer strategy. In their study, they tested models like VGG19, ResNet50, InceptionV3, DenseNet121, and MobileNetV2, using mammography images from the CBIS-DDSM dataset [31], and experimenting with different image sizes, transfer learning techniques, and pre-processing steps to improve results. They found that MobileNetV2, when combined with ImageNet weights, class balancing, and image enhancement, delivered the best performance, showing a 5.6% increase in accuracy. What makes their work practical is how it highlights which combinations actually improve results, helping researchers save time instead of having to test every possible setup manually. Our project follows a similar mindset by comparing multiple models and carefully selecting techniques that optimize performance for breast cancer detection on MRI images.

Nature Scientific Reports [32] (2025) proposed a quantum-optimized model called Q-BGWO-SQSVM for breast cancer classification based on mammography images. The model integrates SqueezeNet for feature extraction with a support vector machine (SVM) classifier, optimized by the Quantum Binary Grey Wolf Optimizer (Q-BGWO). It was trained on four different mammography datasets [33]: MIAS, INbreast, DDSM, and CBIS-DDSM. The model achieved strong results, including 99% accuracy on CBIS-DDSM and high sensitivity and specificity across all datasets. It also outperformed several other methods such as GWO-SVM, SVM-AR, and traditional deep learning models. Despite the strong performance, the study noted limitations like lack of diverse real-world data and reliance on manual feature selection. This study relates to our project by emphasizing optimized model performance on real mammography data, while our project focuses on applying advanced deep learning techniques on MRI scans to improve clinical applicability.

Subham Tewari et al. [34] (2021) developed a project aimed at detecting breast cancer at an early stage using neural networks. Their motivation came from the rising number of breast cancer cases globally and the critical importance of early diagnosis for improving survival rates. The project focuses on automating the detection process by analyzing mammography images, looking for key indicators like masses and microcalcifications, which are important in spotting early-stage cancer. They used adaptive mean filtering to reduce noise and preserve essential details, followed by segmentation techniques including Gaussian Mixture Models (GMM), K-means clustering, and a Hidden Markov Random Field with Expectation-Maximization (HMRF-EM) algorithm. The system then classifies the images into normal, benign, or malignant categories based on their features. While the project uses mammography data² for training and testing, it relies on a set of custom sample images and does not reference a publicly available dataset. One major limitation is that the system is dependent on manual steps within a MATLAB environment, making it less scalable and harder to generalize compared to fully automated deep learning models. Despite these challenges, the project shows how machine learning can support early breast cancer detection, and our work builds on these foundations by applying deep learning architectures like DenseNet, KAN, and Vision Transformers to MRI imaging for more robust and automated diagnosis.

Cancers Journal (2022) [35] published a study on breast cancer detection using CNN and ensemble machine learning techniques. Several universities collaborated on this research, aiming to build a highly accurate model for early breast cancer diagnosis. They used the Breast Cancer Wisconsin (Diagnostic) dataset [36], which includes 32 features across 569 patient samples, classifying tumors as either benign or malignant. The CNN architecture included an embedding layer, a one-dimensional convolutional layer with 500 filters, ReLU activation, and a max-pooling layer, enabling 3D image data to be transformed into a 1D format suitable for machine learning models. Their system achieved a remarkable 100% accuracy and outperformed other models like Random Forest, K-Nearest Neighbors, SVM, Decision Trees, Gradient Boosting, and Gaussian Naive Bayes. However, the study acknowledged that although the results were promising, the model was still a research prototype and not ready for clinical deployment. Our project relates to this work by similarly leveraging deep learning models but expands the scope by working on MRI imaging and focusing on generalizability and clinical adaptation.

BMC Medical Informatics and Decision Making[37] (2023) introduced a machine learning-based model to predict the risk of breast cancer (BC) recurrence using clinical and laboratory data instead of imaging. The study compared eleven machine learning algorithms—including AdaBoost, Random Forest, LightGBM, SVC, and XGBoost—to identify the most effective model for recurrence prediction. The researchers used electronic medical records and routine lab indicators from 342 BC patients, incorporating 25 clinical features like CA125, CEA, Fbg, tumor size, and molecular subtype. Among all models, AdaBoost achieved the highest performance, with an AUC of 0.987, sensitivity of 94.7%, specificity of 97.6%, and overall accuracy of 97.1%. The model's interpretability was enhanced using SHAP (Shapley Additive Explanations), which helped visualize feature importance. This allowed for the identification of key recurrence markers, such as CA125, Fbg, and tumor diameter. The study emphasized AdaBoost's potential as a clinical decision support tool. However, limitations included a relatively small sample size and the absence of genomic or advanced molecular data, which may limit generalization. Our project builds on this work by integrating ML models like AdaBoost alongside deep learning architectures on MRI data, allowing for multimodal, robust prediction of recurrence and enhancing clinical applicability.

Scientific Reports (2024) [38] presented a machine learning based model that classifies breast cancer in Bangladeshi patients through traditional ML techniques combined with explainable AI. The research used a primary dataset of 500 patients from Dhaka Medical College Hospital [39]which contained geometric features and age information necessary to differentiate benign from malignant tumors. The researchers implemented proper preprocessing and data-splitting methods to evaluate five ML algorithms (Decision Tree, Random Forest, Logistic Regression, Naive Bayes, and XGBoost) based on accuracy, precision, recall, and F1-score performance metrics. XGBoost stood out as the best performing model among those tested with 97% accuracy and an F1-score of 0.96. This study distinguished itself by implementing SHAP (Shapley Additive Explanations) to explain the XGBoost model's predictions. The explainable AI tool revealed which features most impacted the model's decision-making process which increased its transparency for clinical applications. Results from the study demonstrate how interpretable AI models paired with real-world hospital data can enhance early diagnosis capabilities. The study's dataset limitations stem from its reliance on a single small non-public dataset which likely impacts generalizability. Our research extends previous work by utilizing extensive datasets and evaluating sophisticated models

such as DenseNet, KAN, and VET while maintaining explainability through SHAP and related methods.

Indian AI Production[40] (2023) developed a complete machine learning project for detecting breast cancer using the Breast Cancer Wisconsin (Diagnostic) Dataset[41]. Researchers developed multiple supervised classification algorithms for tumor identification to distinguish between malignant and benign cases. The research team tested multiple supervised classification models including Support Vector Machines (SVM), Logistic Regression, K-Nearest Neighbors, Naive Bayes, Decision Trees, Random Forest, AdaBoost, and XGBoost. XGBoost achieved the highest accuracy rate at 98.24% among all tested models. The data set used for this study includes 569 patient records with 30 features extracted from digitized fine-needle aspirate images taken from scikit-learn's official dataset repository. The study incorporated visualization tools such as heatmaps, correlation plots and SHAP analysis during model training to enhance interpretability. This project benefits from a step-by-step implementation approach making it ideal for those new to machine learning. The study's use of a limited and balanced dataset presents a challenge as it might not accurately represent the diverse conditions found in real-world clinical settings. The research provides important practical information about classifier evaluation methods that will benefit our project.

Chapter 3: Methodology and Model Architectures

To develop a reliable system for breast cancer detection using MRI data, we are going to test and implement four deep learning approaches: CNN, Vision Transformers (ViT), Kernel Attention Networks (KAN), and DenseNet-201. The different architectures excel in various aspects such as feature extraction capabilities and attention mechanisms or network depth which positions them as potential options for medical image classification tasks. This chapter provides an explanation of the methodology behind each model including its structure and implementation details with our dataset to support the evaluation results in the subsequent chapter.

3.1 Dataset Description

For this project, we utilized the Breast MRI Images for Breast Cancer Detection dataset available on Kaggle[1]. The dataset contains MRI images which have been labeled as benign or malignant to support deep learning model training and evaluation in binary classification scenarios.

This dataset contains an equal number of labeled MRI images which enables efficient training of classification algorithms. The dataset achieved uniformity by resizing every image to 224x224 pixels followed by normalization. Model generalization was improved through the application of data augmentation techniques including rotation, flipping, and zooming.

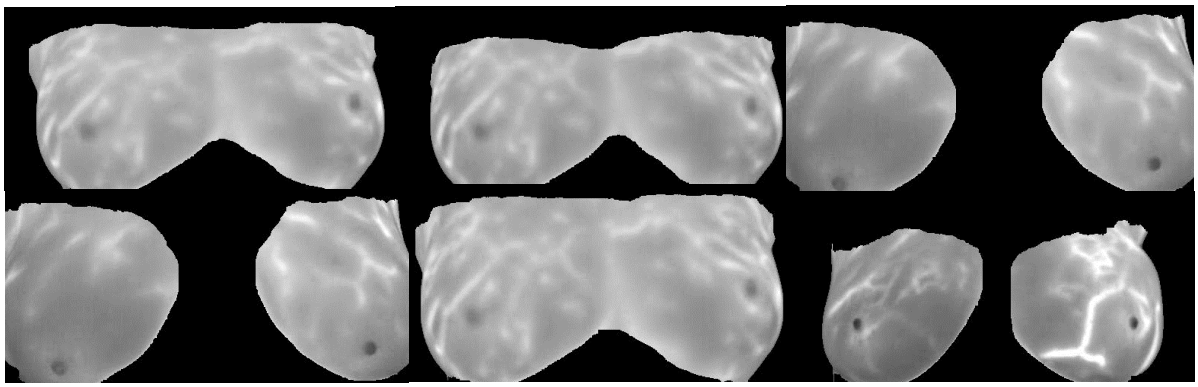

Figure 2: Sample data entries from the Breast Cancer Patients MRI dataset.

[1] Kaggle website: **https://www.kaggle.com/datasets/uzairkhan45/breast-cancer-patients-mris?utm_source=chatgpt.com**

3.2 Imaging Modality Selection: MRI over Ultrasound

Our choice of Magnetic Resonance Imaging (MRI) as the imaging modality for this study over ultrasound was based on multiple compelling reasons. Higher Sensitivity: Magnetic Resonance Imaging (MRI) proves to detect breast cancer with greater sensitivity than other methods, particularly in women who have dense breast tissue. Research indicates that MRI technology identifies breast cancers which remain undetected by mammograms and ultrasound scans. Detailed Imaging: MRI captures detailed three-dimensional high-resolution images of breast tissue which enables better identification of abnormalities and supports precise diagnosis. Effectiveness in High-Risk Populations: MRI proves most advantageous for individuals facing a high breast cancer risk because of family history or genetic factors since it enables early cancer detection.

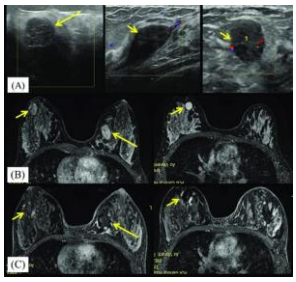

Figure 3: MRI scans illustrating malignant breast lesions. The yellow arrows highlight abnormal tissue areas with high contrast and detailed structural boundaries, useful for early cancer detection.

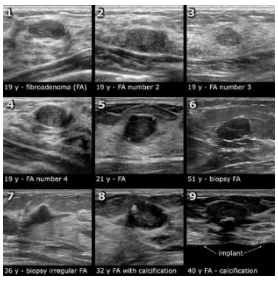

Figure 4: Ultrasound images of various benign and malignant breast conditions. While cost-effective and widely used, ultrasound offers lower contrast and is more operator-dependent compared to MRI.

Ultrasound serves as a useful instrument for biopsy guidance and targeted evaluation but faces limitations from operator dependency and reduced sensitivity in some situations. This study selected MRI to utilize its superior imaging functions and its ability to provide more precise diagnostic results.

3.3 Model Architectures

3.3.1 Convolutional Neural Networks

CNNs represent deep learning algorithms created to handle visual information by replicating the human visual system's pattern and object recognition abilities. The core functions that construct CNNs include convolution operations, activation functions and pooling mechanisms. CNNs fundamentally operate by automatically learning image patterns through multiple specialized processing layers. The convolutional layers initiate the process by deploying small filters called kernels which move over the image to analyze small regions and detect fundamental visual elements such as edges and textures. These filters generate feature maps that show the locations of detected features within an image. Following the convolution operation an activation function such as Rectified Linear Unit (ReLU) introduces non-linearity enabling the network to understand

complex patterns beyond simple linear relationships. Pooling layers function to shrink feature maps by choosing the key features (such as the maximum value within a small area) which makes the model less affected by minor image variations like rotation or translation. Through multiple repetitions of convolution followed by activation and pooling operations CNNs develop deeper layers which enable the detection of progressively intricate patterns including shapes and textures as well as full objects or abnormalities. After flattening the extracted features they enter fully connected layers to analyze the data and make decisions that classify images or generate predictions. CNNs automatically learn these steps through training by modifying filters and parameters with backpropagation to enhance model performance through error reduction between predictions and actual labels.[42]

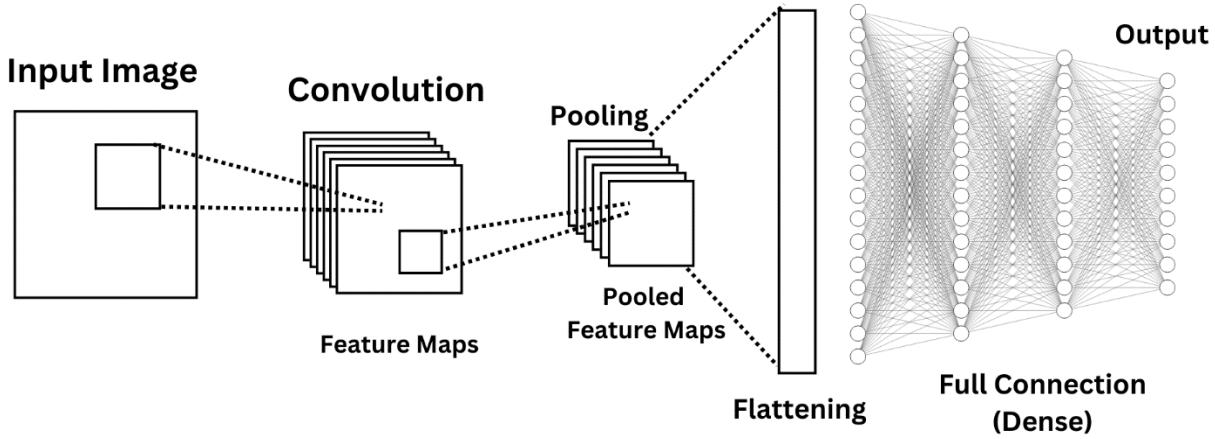

Figure 5: Convolutional Neural Networks Architecture [43]

In mathematics, convolution merges two functions together to create a new function that shows how one function transforms or integrates with the other. The convolution operation finds widespread application across signal processing disciplines while it holds particular significance in image processing for pattern extraction from image data. CNNs use this operation to learn spatial feature hierarchies through the application of filters or kernels across input images. The continuous domain's convolution operation uses the standard mathematical notation:

$$(x * w)(t) = \int_{-\infty}^{\infty} x(a) \cdot w(t-1)\, da \qquad 3.1$$

*Where x represents the input function, w is the weight function (or filter) and t denotes the spatial location.

However, image processing data exists in discrete form since it is defined as a limited collection of pixels instead of continuous inputs. The integral operator becomes a summation operation while the convolution process undergoes a reformulation.

$$(x * w)(t) = \sum_{-\infty}^{\infty} x(a) \cdot w(t-a)\, da \quad 3.2$$

Activation Layers act as essential parts of CNNs because they enable the neural network to identify intricate non-linear patterns within data. A neural network remains restricted to linear transformations without activation functions even when its architecture includes multiple layers. Activation layers introduce non-linearity following convolutional operations to enable CNNs to capture complex real-world data patterns which is particularly beneficial for image processing tasks.

The Rectified Linear Unit (ReLU) stands as the dominant activation function in CNNs because of its straightforward computation and its ability to prevent the vanishing gradient problem [44]. ReLU is mathematically defined as:

$$\mathrm{ReLU}(x) = \max(0, x) \quad 3.3$$

The function f(x) =max (0, x) outputs the input x when x is positive but returns zero for any non-positive x. The network gains sparsity from this behavior while enabling deep models to train more effectively. ReLU faces challenges because of the "dying ReLU" issue where neurons stop participating in learning when they regularly receive negative input.

Several activation functions besides ReLU have gained substantial recognition due to their unique mathematical characteristics and specific use cases. The Sigmoid function remains popular for mapping inputs to outputs between 0 and 1 during binary classification tasks in output layers. The Sigmoid function demonstrates limited effectiveness in deep CNNs' convolutional stages because it creates vanishing gradients when used in hidden layers. The Hyperbolic Tangent (Tanh) activation function produces outputs between -1 and 1 while providing a zero-centered output pattern but encounters the same vanishing gradient issue in deep learning models. The Softmax function serves mainly multi-class classification systems by converting logits into class probability distributions but is deemed unsuitable for direct use in convolutional layers because this study targets binary breast cancer detection. The Swish function introduced by Google Brain shows potential for deep learning applications thanks to its smooth gradients and self-gating mechanism but its high computational requirements and limited testing in medical imaging necessitate additional research before it can be reliably used in critical diagnostic contexts. The choice of ReLU activation function stands on strong theoretical grounds and practical evidence given the project's focus on breast cancer early detection through MRI-based CNN analysis. The combination of computational efficiency and robustness together with empirical success in similar tasks makes this activation function appropriate for the design of the architecture chosen in this study.

The Pooling Layer plays a crucial role in refining feature maps by shrinking their spatial dimensions while maintaining essential features after the convolutional and activation layers perform initial feature extraction. This network stage performs a pooling operation which transforms localized regional data into condensed information, allowing the model to highlight important patterns while minimizing noise and redundant features.

Max Pooling stands out as the most used technique among different pooling methods. Max Pooling works by moving a fixed-sized window across the feature map to select and keep the highest value in each region.

This mechanism captures dominant activations while enhancing model robustness to spatial changes such as translations and rotations which typically occur in medical imaging. For instance, when applying a 2 × 2 max pooling operation with a stride of 1 applied to a convolved feature map extracts only dominant values from each subregion which enhances the detected strong features while lowering the resolution of the map. Average Pooling operates through a similar procedure but determines the average of the values inside each window resulting in a smoother and more generalized depiction of features. Average Pooling proves beneficial for certain uses, but it reduces feature distinctiveness which makes it unsuitable for early breast cancer detection that requires the identification of strong local patterns. Global pooling methods like Global Max Pooling and Global Average Pooling transform each feature map down to one value which effectively encapsulates all spatial data into a condensed format. Global pooling works well in classification networks final layers, but medical image analysis models usually avoid it in early or intermediate layers because it eliminates essential local details that help recognize subtle abnormalities.

The Flattening Layer transforms output from convolutional, activation, and pooling layers operating on two-dimensional feature maps into a one-dimensional vector that connects feature extraction with fully connected layers used for classification. During this operation multidimensional feature maps transform into a one-dimensional vector in a method called flattening. The transformation prepares the extracted features into a format that can be used by the fully connected layers for classification tasks.

$$MaxPool\big((I * K)(i,j)\big) = \begin{bmatrix} 1 & 3 & 5 \\ -4 & 0 & -3 \\ 5 & 7 & -9 \end{bmatrix} \quad 3.4$$

The flattened matrix:

$$F = [1 \quad 3 \quad 5 \quad -4 \quad 0 \quad -3 \quad 5 \quad 7 \quad -9]$$

Fully Connected Layer (Dense Layer)

The Fully Connected Layer serves as the final component of Convolutional Neural Network (CNN) architecture, and it is commonly known as the Dense Layer. Every neuron within this layer establishes connections to all neurons from the prior layer creating a dense network similar to a Multi-Layer Perceptron (MLP). The fully connected layer stands apart from convolutional and pooling layers because it assigns distinct weights to each input received from the previous layer instead of using localized receptive fields and weight sharing.

The fully connected layer acts as the decision-making component in a CNN network. The fully connected layer takes the flattened feature vector and generates the output by applying a linear transformation and then using an activation function. Fully connected layers perform similar mathematical operations to convolutional layers but differ fundamentally in terms of their design and functional objectives. The convolutional layers work with two-dimensional feature maps by employing localized kernels while fully connected layers handle one-dimensional vectors which allow them to combine information from the whole feature map for prediction or classification.

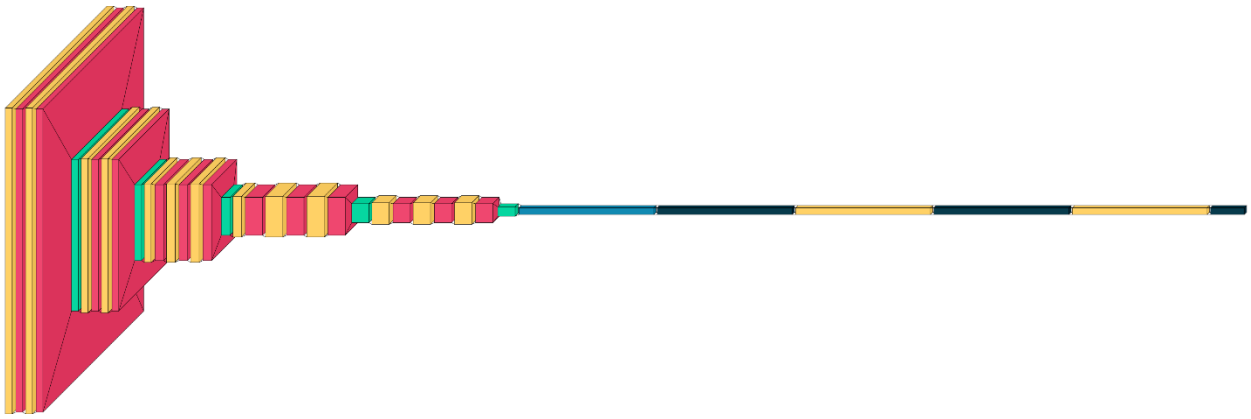

Figure 6: Visual representation of a CNN architecture using VisualKeras[45] that presents the multi-layer architecture of a CNN model which was created with the VisualKeras Python package.

As specialized neural networks CNNs utilize convolution operations to execute classification and prediction tasks with greater efficiency compared to standard deep learning networks. CNNs gain a major benefit from weight sharing which reduces parameters and computational demands by applying uniform weights across image areas, making them efficient for use on systems with restricted GPU capabilities. The architecture of CNNs enhances memory efficiency and provides greater scalability which makes them more practical than fully connected networks while recognizing MNIST digits requires thousands of parameters in dense networks but CNNs achieve the same recognition with much fewer nodes and parameters. CNNs maintain their performance across local image variations including changes in object position, orientation, and scale which enables them to effectively perform tasks such as face recognition and object detection in dynamic real-world environments. CNNs demonstrate equivariance since input transformations like translation lead to predictable changes in output feature maps which sustain the model's consistency and interpretability. The independence of CNNs from common geometric transformations such as rotation and scaling enables them to maintain object detection reliability across different image presentations while lowering misclassification risks from positional changes or distortions. The unique attributes of CNNs establish them as superior tools for image-based applications by delivering greater efficiency and robustness while providing advanced generalization compared to traditional neural networks.[46] Despite their power. CNNs have several limitations: CNNs need substantial labelled data to function effectively while requiring extensive computational resources and memory which makes them difficult to interpret for specific decisions.[47]

3.3.2 Vision Transformers

ViT has recently emerged as a disruptive technology that challenges the supremacy of CNNs in the field of computer vision. CNNs have been praised for their local feature extraction abilities through multiple convolutional and pooling layers yet face difficulties in capturing global context without requiring deeper network structures. The inherent constraint of existing models led researchers to investigate new architectural solutions that can directly and more efficiently handle long-range dependencies. ViT stands out among new architecture approaches because it successfully applied the transformer model from natural language processing to handle visual data. The ViT model converts input images into sequences of patches instead of pixel grids to turn them into formats that resemble NLP textual tokens like those used in BERT models. Specifically, an image of dimensions $H \times W \times C$ is divided into patches of size $P \times P$.

The patches of size P×P are flattened and projected into a fixed-dimensional embedding space with the help of a learnable matrix [48]:

$$Patch\ Embedding\ = Flatten(P) \cdot W_e \tag{3.5}$$

The transformation enables the model to analyze images through sequential processing which allows the self-attention mechanism to identify connections between any two image patches no matter their spatial separation. The model maintains spatial positioning comprehension by adding positional encodings to every patch embedding [49]:

$$\text{Input Embedding} = \text{Patch Embedding} + \text{Positional Encoding} \tag{3.6}$$

In practical ViT implementations learned parameters typically replace fixed sinusoidal functions because they offer greater adaptability.

Once the sequence of patch embeddings is prepared, it is fed into a stack of Transformer Encoder layers, each consisting of two main components: Multi-Head Self-Attention (MHSA) and Feed-Forward Networks (FFN). By computing attention scores between all patches the MHSA module enables the model to learn global dependencies [49]:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3.7}$$

*Where queries Q, keys K, and values V

The input embeddings produce V through learned projection transformations. The attention mechanism divides its focus across multiple heads enabling the model to capture various input features at the same time. After the attention operation the outputs move through the FFN which contains two linear layers divided by a GELU activation function.

Residual connections and layer normalization techniques run through the entire encoder stack to both stabilize training processes and enable better gradient flow. The ViT model incorporates a

learnable classification token which functions similarly to the [CLS] token from BERT by being added at the beginning of the patch sequence. The classification token collects patch representations which a multi-layer perceptron (MLP) head processes through a softmax layer to generate class predictions.

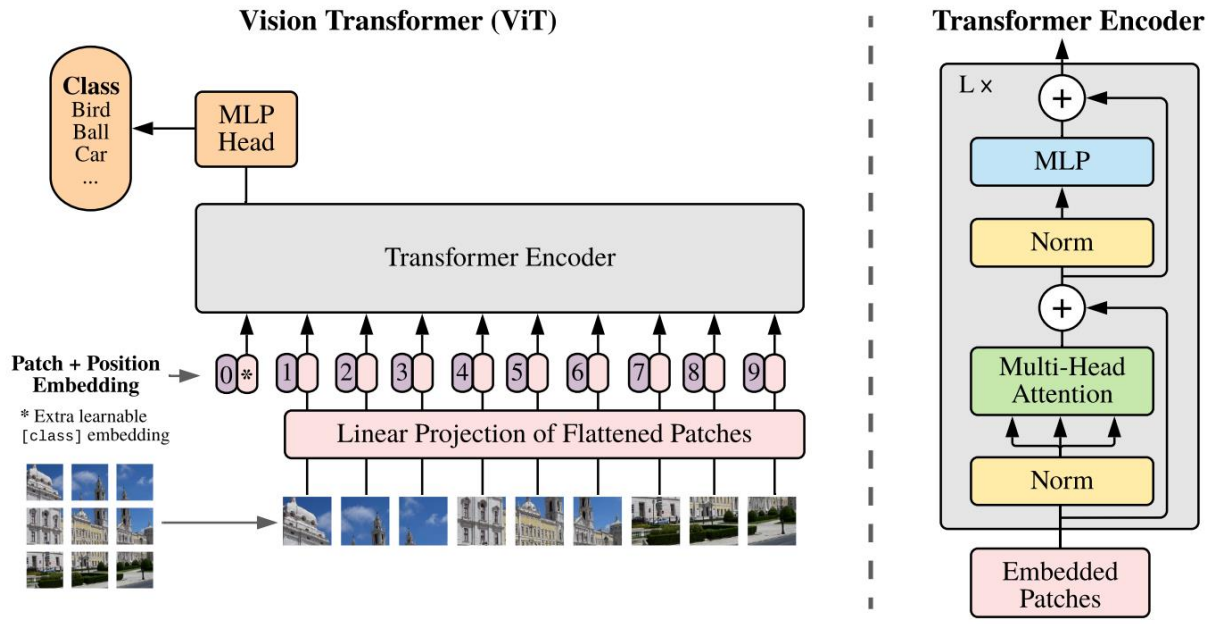

Figure 7: Model overview. We split an image into fixed-size patches, linearly embed each of them, add position embeddings, and feed the resulting sequence of vectors to a standard Transformer encoder. In order to perform classification, we use the standard approach of adding an extra learnable "classification token" to the sequence. The illustration of the Transformer encoder was inspired by Vaswani et al. (2017)[49].

The original ViT's success inspired the proposal of multiple variants that target its weaknesses and enhance its features. The Data-efficient Image Transformer (DeiT) implemented a distillation token alongside knowledge distillation to boost data efficiency which enabled the training of ViTs on limited datasets. The Swin Transformer introduced a hierarchical attention mechanism based on windows that enables Vision Transformers to capture both local and global context while maintaining lower computational load and merging CNN advantages with Transformer strengths.

The Convolutional Vision Transformer (CvT) merged convolutional layers with patch embedding processes so that Vision Transformers gained local inductive advantages while maintaining global attention capabilities. The Tokens-to-Token ViT (T2T-ViT) design features recursive token aggregation processes which help to recognize local patterns before addressing global relationships.

Current advancements demonstrate how hybrid models integrate CNNs' data efficiency with transformers' global modeling capabilities to transform the vision model landscape.

ViTs demonstrate superior performance in breast cancer classification tasks when applied to extensive datasets. MammoViT demonstrates how Vision Transformers have been applied to

mammogram and tomosynthesis image classification tasks with success. Their operation requires substantial computational power while they show limited performance on small datasets unless they undergo pre-training.

ViT can be used for all tasks (classification, detection, and feature extraction) depending on how it's adapted or fine-tuned.

Table 3. Common Applications of Vision Transformer (ViT) Across Computer Vision Tasks[50]

Task	How ViT is Used
Classification	CLS token → MLP head for label prediction
Detection	Adapted in models like DETR or ViTDet[51]
Feature Extraction	Patch/CLS token outputs from transformer

MRI images contain complex patterns that extend over extensive areas of the image like multifocal lesions and subtle textural abnormalities scattered throughout breast tissue. The global pattern modeling capability of ViT works alongside CNNs localized feature extraction to create diagnostic models that deliver higher accuracy and robustness. However, we will utilize transfer learning from large pre-trained Vision Transformer models that are fine-tuned using target breast MRI data due to the small size of medical imaging datasets. This method reduces ViT's data inefficiency and utilizes its global attention features.

3.3.3 Kernel Attention Networks

The KAN model marks a major development in attention mechanisms since it replaces traditional self-attention with kernel-based operations for attention calculation. KAN was designed to overcome the standard attention methods' limitation of linear dot-product similarity by using kernel functions to capture complex non-linear relationships between input features. KAN becomes highly effective for applications that need detailed pattern recognition like early breast cancer detection in medical images because it can model subtle tissue structure variations with strong discriminative features.

The standard attention mechanism within the Transformer architecture calculates attention scores using scaled dot-products between the query (Q) and key (K) matrices. The standard process lacks efficiency in identifying non-linear patterns which reduces its effectiveness for complex tasks like medical image classification that require detection of non-linear structures. KAN addresses this limitation by replacing the traditional dot-product calculation with a kernel function (Q,). The calculation of attention scores in KAN follows this specific formula:

$$KAN(Q,K,V) = softmax\big(k(Q,K)\big)V \quad 3.8$$

*Where κ could be a Gaussian Radial Basis Function (RBF), a polynomial kernel, or any other positive-definite function suitable for the task at hand.

This modification enables the attention mechanism to capture non-linear and higher-order interactions between inputs, substantially improving its expressive power.

KAN implements multiple optimization methods to keep computations manageable when working with extensive datasets or extended sequences. Kernel approximation methods like random feature mappings and low-rank projections enable efficient kernel function estimation while preserving performance during scalability. The Kerformer stands as an exemplary model which adopts a non-linear reweighting approach through individual non-negative transformations on both queries and keys before utilizing a Squeeze-and-Excitation (SE) block to dynamically modify feature significance. The design strengthens model discrimination abilities while lowering self-attention mechanism time complexity thus enabling practical deployment through improved efficiency.[52]

After Kerformer, multiple specialized versions of KAN have emerged to tackle unique challenges across various domains. The Kernel Graph Attention Network (KGAT) applies kernel attention mechanisms to graph-based structures to support detailed fact verification by building an evidence graph where node and edge relationships receive enhancements from kernel functions. Large Kernel Attention (LKA) enhances visual recognition performance by replacing attention modules with single large-kernel convolutions that integrate convolutional neural networks' locality and parameter sharing benefits with attention-based global context modeling functionality. Large Separable Kernel Attention (LSKA) improves this method by splitting large-kernel convolutions into sequential one-dimensional processes that lower both computational load and memory demands but still deliver consistent performance results [53].

These KAN variants provide multiple advantages when combined. KANs combine non-linear attention operations and better feature representation capabilities which result in more precise and reliable pattern recognition. Medical imaging fields benefit immensely from this functionality because spotting subtle anomalies such as early-stage tumors in MRI scans relies on precise identification. Models such as Kerformer and LSKA have achieved efficiency gains which enable resource-efficient processing of high-resolution medical images which support real-time clinical decision-making.

KAN demonstrates exceptional adaptability for multi-modal medical applications through its capability to operate seamlessly with different data formats such as images, sequences, and graphs. The ability of KAN to model complex interdependencies among heterogeneous inputs makes it possible to achieve more comprehensive and precise diagnostic results when breast MRI data is combined with patient demographics and other clinical information.[54]

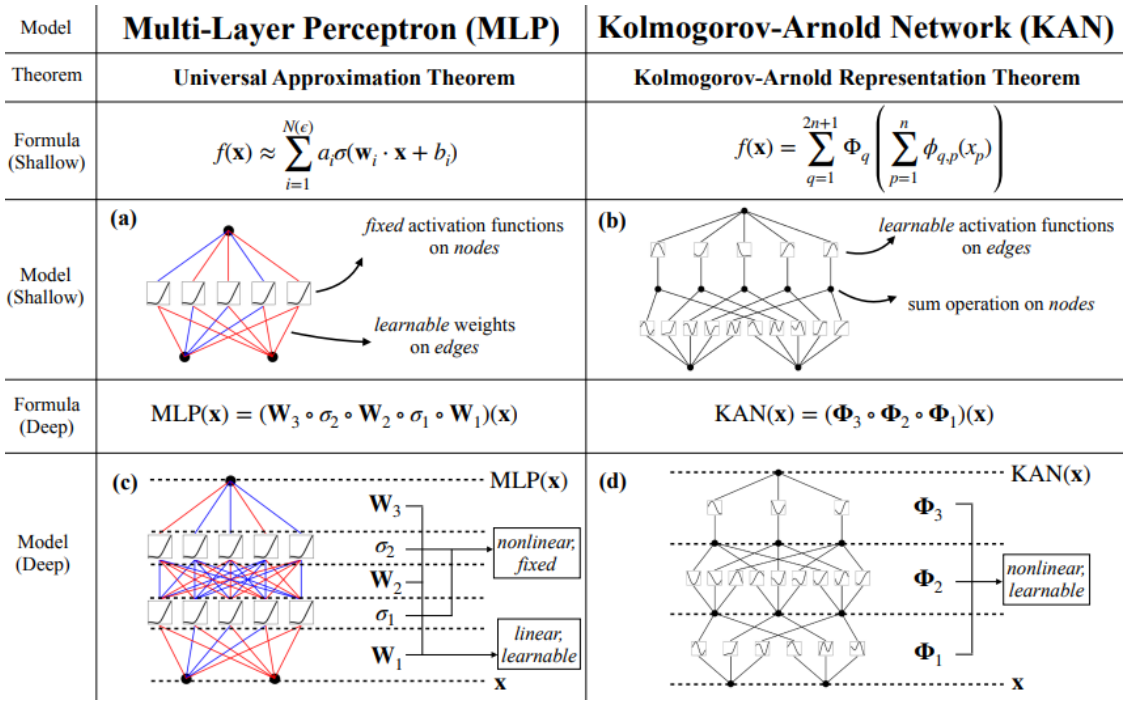

Figure 8: Multi-Layer Perceptrons (MLPs) vs. Kolmogorov-Arnold Networks (KANs)

KANs derive their theoretical basis from the Kolmogorov–Arnold representation theorem which shows that multivariate continuous functions can be broken down into sums and compositions of univariate functions. KANs implement this principle through the learning of symbolic activation functions at the neuron level which demonstrate both interpretability and expressiveness.

KAN layers handle inputs by applying a unique trainable function to each weight separately instead of using a common activation function after summing weighted inputs like in MLPs. KANs can approximate complex nonlinear functions using fewer layers and parameters which results in better performance and generalization especially when dealing with structured domains that have limited data such as medical imaging.

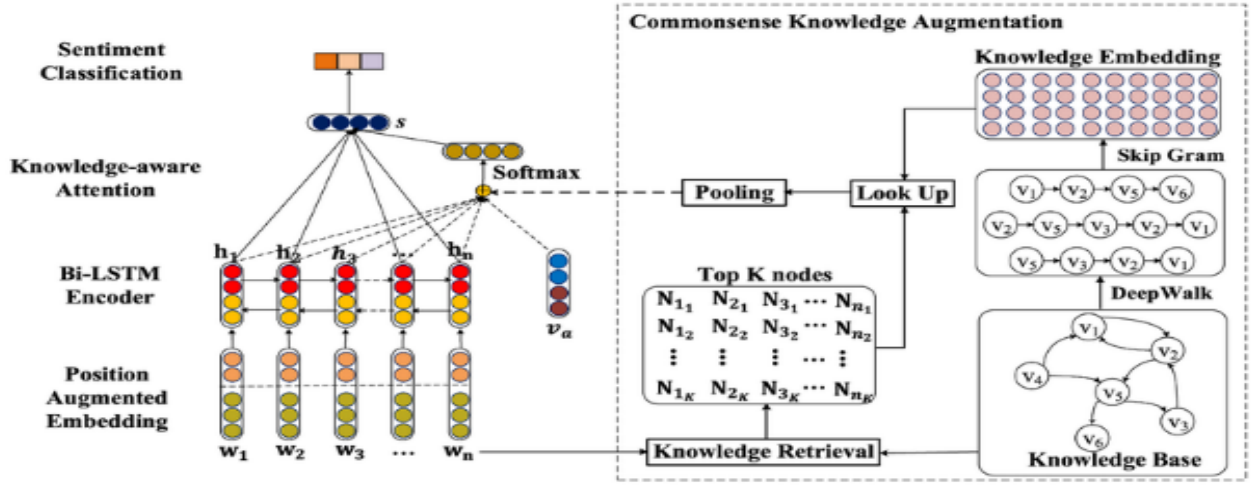

Figure 9: A schematic of the Kernel Attention Network (KAN) applied in a sentiment classification context. The left part illustrates the data flow from position-augmented embeddings through a Bi-LSTM encoder and knowledge-aware attention mechanism, while the right side details the Commonsense Knowledge Augmentation pipeline. This includes top-K knowledge node retrieval from a knowledge base, embedding via DeepWalk and Skip-Gram, and integration with the attention process to enhance interpretability and performance.

Our study implements KANs following CNN-based feature extraction to establish hybrid architecture. The CNN extracts spatial data from MRI imagery while KAN layers perform the final classification task. The system merges CNN's ability to extract local features with KAN's adaptable and understandable function-based attention mechanism.

Breast cancer MRI classification demands models which simultaneously maintain accuracy while ensuring generalization capabilities and interpretability. Medical imaging tasks frequently experience dataset limitations unlike areas with abundant data because obtaining annotated samples require privacy protection measures and expert input which drives up labeling costs. The requirement for data-efficient models with transparent decision-making capabilities becomes essential in these conditions. Kernel Attention Networks (KANs) stand out as an optimal solution for detecting early breast cancer in MRI data when applied in this context. The intrinsic interpretability of KANs stands as the primary reason why they should be integrated into this application. KANs use learnable functions for network edges rather than fixed transparent activation functions found in traditional neural networks and transformers. Splines represent these functions which are both visually interpretable and convertible into symbolic mathematical forms. The feature provides clinicians and medical practitioners with direct access to the model's decision-making process which enhances clinical trust and supports the integration of AI tools into healthcare workflows. The inherent compactness of KAN architecture makes them particularly well-suited for medical imaging tasks. The architecture of KANs results in a smaller parameter

requirement than transformers which have high parameter counts and computational needs. These models become computationally efficient while simultaneously reducing overfitting risks which often occur with small datasets. KANs preserve high predictive accuracy while remaining resilient and generalizable through their streamlined design which functions well in data-limited environments common to breast MRI sets. KANs stand out because they function without positional encodings. Transformers require fixed positional embeddings to represent spatial data whereas KANs completely eliminate this requirement. The KAN design achieves inherent flexibility by bypassing explicit sequence ordering requirements which enables the network to adapt naturally to various input structures such as MRI images with varying spatial configurations. KANs uniquely excel at creating symbolic representations from learned patterns. This feature enables medical professionals to obtain mathematical representations for verification which supports the development of transparent AI systems for healthcare diagnostics. KANs provide increased transparency that enables domain experts to investigate the logic behind model predictions which could lead to new discoveries about breast cancer pathology.

The KAN model utilizes an organized learning procedure which increases interpretability through complexity reduction and symbolic representation of learned functions. The training process starts with sparsification to teach the network to prioritize important connections and eliminate unnecessary ones. The model enters a pruning stage after training to eliminate redundant weights which leads to a more streamlined and efficient configuration. After pruning the model assigns basic kernel basis functions including sine, square, and exponential which function as clear building blocks for its learned representation. The model's affine parameters undergo training which refines the basis functions to enable exact control of their behavior through learned shifts, scales and shapes.The last stage creates a symbolic mathematical expression that represents the function which the network has learned. The reduced symbolic representation maintains accuracy in numerical calculations and reveals how the model processes information internally. The procedure allows KANs to deliver accurate outputs which maintain transparency and interpretability while connecting neural computation with symbolic understanding.

3.3.4 ResNet-50 V2: Deep Residual Network

ResNet-50 v2 stands as a progression in deep convolutional neural network models aimed at addressing optimization difficulties encountered in training networks with great depth. ResNet-50 v2 emerged as an advancement over its predecessor ResNet-50 v1 by integrating important architectural changes that have led to its extensive use in computer vision and medical imaging. ResNet architectures build on residual learning through shortcut connections which enable the network to learn residual mappings rather than direct representations thereby solving the degradation problem along with the vanishing gradient challenges faced during deep network training. ResNet-50 v2 advances this foundation by implementing a pre-activation residual block structure where batch normalization and activation functions occur before the convolutional operations. The slight yet powerful change enhances gradient flow which results in better training stability and efficiency.

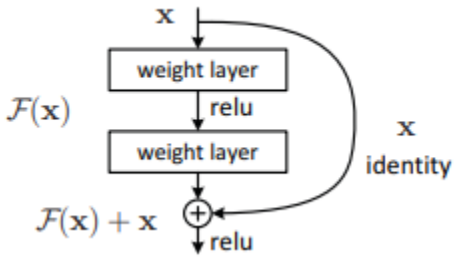

Figure 11: Residual learning: a building block.

The ResNet-50 v2 structure includes a deep network of 50 layers which are distributed across multiple stages. The first processing involves a 7x7 convolutional layer with a stride of 2 followed by a 3x3 max pooling operation that both condenses the image input size and extracts low-level features. Four distinct stages make up the central part of the model and each stage includes multiple bottleneck residual blocks. The structure of each block contains three convolutional layers starting with a 1x1 convolution for dimensionality reduction followed by a 3x3 convolution for feature extraction and ending with a 1x1 convolution to restore dimensionality. Batch normalization and ReLU activation take place before each convolution in ResNet-50 v2 blocks to enable smoother forward and backward signal propagation. The pre-activation method keeps the identity mapping clear and results in both quicker learning progress and better overall model performance.

The choice of ResNet-50 v2 as the core network for medical imaging tasks is supported by multiple key factors. The combination of its 50-layer deep architecture with advanced residual block design provides powerful depth and expressivity that enables the model to capture essential hierarchical features for medical image analysis. In clinical applications this ability proves essential since lesions, tumors, and abnormalities display extensive variation in size, shape, texture, and intensity throughout different anatomical regions. ResNet-50 v2 proves superior for diagnostic imaging applications by delivering better results than shallower models in essential clinical evaluation metrics including sensitivity and specificity. ResNet-50 v2 stands out for transfer learning workflows because pretrained weights are readily available in popular deep learning platforms such as TensorFlow and PyTorch. Practitioners can easily adapt ResNet-50 v2 to domain-specific datasets such as MRI, CT, and ultrasound scans which typically exhibit limited sample sizes and class imbalance. ResNet-50 v2's easy accessibility enables efficient model fine-tuning which enables practitioners to utilize its inherent general visual features for specialized diagnostic tasks with limited extra training work. The pre-activation residual block structure in ResNet-50 v2 boosts training efficiency despite the model having extensive depth. This design leads to faster and more dependable convergence by enhancing gradient flow and optimization stability which becomes especially effective when paired with data augmentation and transfer learning. The ResNet-50 v2 configuration delivers excellent performance while maintaining computational efficiency and simple integration with medical imaging systems. These attributes combine to make ResNet-50 v2 both strong and adaptable for advanced medical image classification tasks by providing high diagnostic accuracy and real-world deployment benefits.[55]

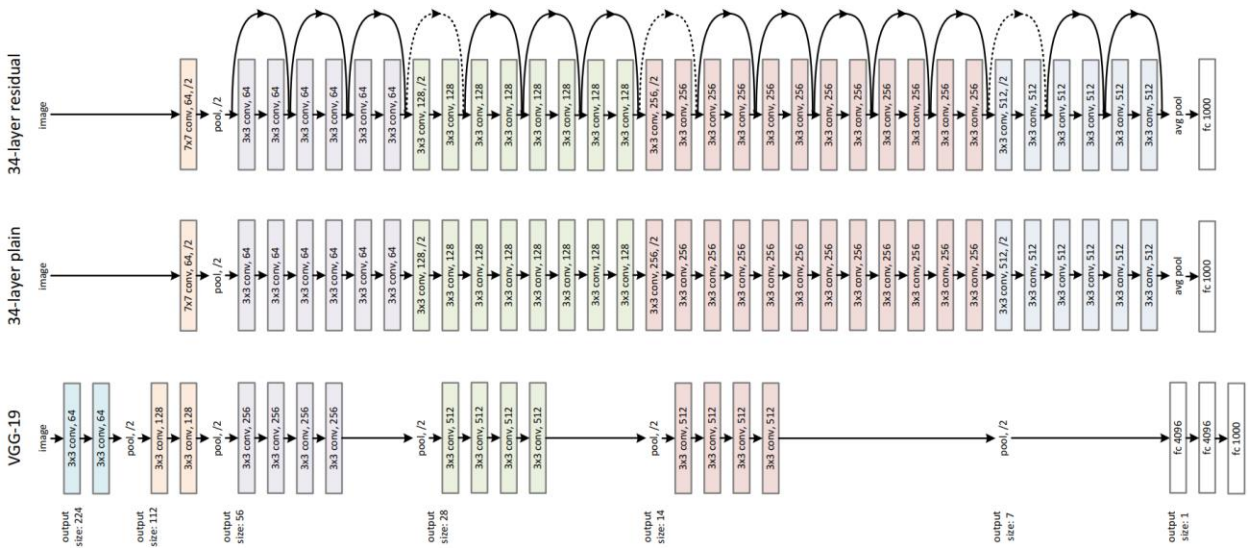

Figure 12*: Example network architectures for ImageNet. Top: the VGG-19 model (19.6 billion FLOPs) is shown as a reference. Middle: a plain network with 34 parameter layers (3.6 billion FLOPs). Bottom: a residual network with 34 parameter layers (3.6 billion FLOPs).*

CHAPTER 4: Project Design & Implementation

4.1 Project Requirements

The Breast Cancer MRI Classification System was developed with a clear set of functional and non-functional requirements, combining software frameworks and model architectures to build a reliable research prototype. Each requirement was defined and implemented to ensure that the system worked effectively, adapting to data constraints while achieving robust classification performance.

4.1.1 Hardware Requirements

The development of this project was initiated on a personal laptop (HP ProBook 450 G8, Intel Core i5-1135G7, 8 GB RAM, Windows 10 Pro) for code writing and dataset organization. However, all model training and evaluation were performed in a cloud environment using Google Colab Pro equipped with an NVIDIA A100 GPU (40 GB memory). The Colab environment provided the necessary computational capacity to train Vision Transformer and ResNet+KAN models efficiently while ensuring reproducibility and scalability.

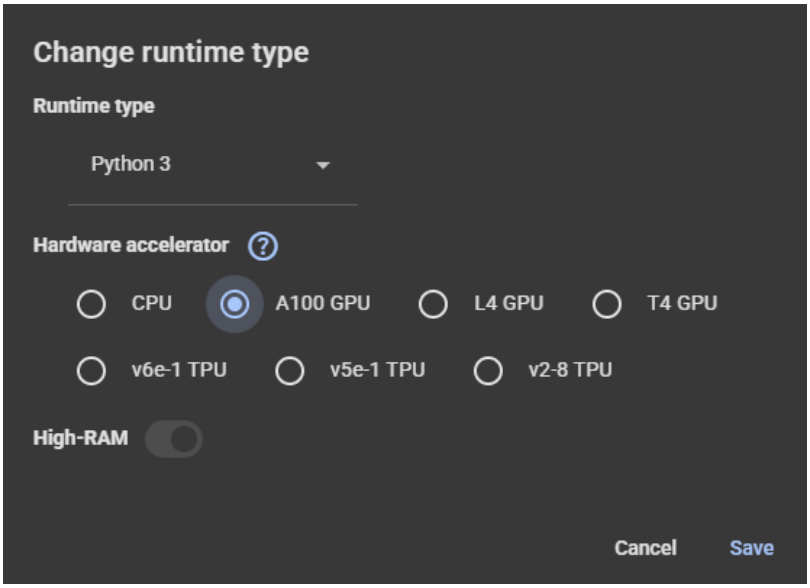

Figure 13. Google Colab runtime settings showing the selection of A100 GPU as the hardware accelerator for training.

4.1.2 Software Requirements

The software environment combined local setup for code development with cloud execution for training. On the local machine, Python and Jupyter Notebook were used primarily for writing and organizing scripts. The complete training and evaluation, however, were conducted on Google Colab Pro, which provided a preconfigured environment with CUDA support for the A100 GPU.

The main software components included:

- Python (3.11) as the core programming language.
- PyTorch (2.2+) as the primary deep learning framework.
- TorchVision for pretrained architectures (ViT-B/16 and ResNet-50 v2) and dataset utilities.
- Albumentations for data augmentation (random flips, rotations, brightness/contrast changes, Gaussian blur).
- scikit-learn for evaluation metrics (accuracy, precision, recall, F1-score, confusion matrix).
- PyKAN for the Kolmogorov–Arnold Network implementation integrated into ViT and ResNet backbones.
- Matplotlib and Seaborn for visualization of learning curves and confusion matrices.

The Colab environment ensured smooth integration of these libraries with GPU acceleration, enabling reproducible experimentation and efficient model evaluation.

4.1.3 Functional Requirements

The functional requirements defined the essential capabilities that the Breast Cancer MRI Classification System needed to demonstrate as a research prototype. These included:

- Binary Image Classification: The system must correctly classify MRI scans into two categories: *Healthy* and *Sick (Malignant)*.
- Model Comparisons: The system must support the training and evaluation of three different architectures — ViT-only, ViT+KAN, and ResNet+KAN — to analyze performance variations.
- Evaluation Metrics: For each model, the system must generate quantitative performance measures such as accuracy, precision, recall, F1-score, and confusion matrices.
- Visualization of Results: The system must produce clear visual outputs, including training/validation loss curves, accuracy progression, and confusion matrix plots, to support interpretability.
- Reproducibility: The system must ensure consistent execution in the Colab environment with fixed random seeds, enabling reproducible results across multiple runs.

4.1.4 Non-Functional Requirements

Beyond the core classification tasks, several non-functional requirements were defined to ensure the overall reliability and usability of the Breast Cancer MRI Classification System:

- Reliability: The models should consistently produce stable results across multiple runs and under different random seeds, ensuring trustworthiness of outcomes.
- Efficiency: Training and evaluation needed to run smoothly within the Colab Pro environment without unnecessary overhead, utilizing GPU resources optimally.
- Adaptability: The system was designed with modular code, allowing for easy replacement or addition of new architectures (e.g., DenseNet, Swin Transformers) without rewriting the full pipeline.

- Explainability: By integrating Kolmogorov–Arnold Networks (KAN), the system provided enhanced interpretability through spline-based symbolic functions, supporting future clinical relevance.
- Scalability: The methodology was structured so that larger datasets or higher-resolution MRI inputs could be integrated with minimal modification to the training pipeline.

4.1.5 Security and Privacy Considerations

Although this research was conducted using publicly available and anonymized MRI datasets, ethical and privacy aspects were considered to align with potential real-world medical applications.

- Data Security: All MRI scans used in the experiments were stored and processed within the secure environment of Google Colab Pro, ensuring that no unauthorized access or external transmission of data occurred. In a real deployment scenario, encryption during storage and transmission would be mandatory.
- User Privacy: The dataset used was fully anonymized, with no patient-identifiable information included. This ensured compliance with research ethics and highlighted the project's adaptability for future integration into medical systems while safeguarding patient privacy.
- Ethical Compliance: The project methodology emphasized reproducibility, transparency, and responsible AI development, laying a foundation for possible clinical applications under strict ethical guidelines.

4.2 Project Modeling

The modeling phase of the Breast Cancer MRI Classification System provided a clear representation of the workflow, interactions, and structural components of the project. This step was crucial to ensure that the system design aligned with the objectives and that the technical implementation accurately reflected the functional requirements.

4.2.1 Use Case Diagram

The use case diagram outlined the interactions between users and the system:

- Actors:
  - *Researcher/Practitioner:* Uploads MRI scans, initiates training, and evaluates results.
  - *System:* Preprocesses data, extracts features, trains models, and outputs predictions with evaluation metrics.
- Main Use Cases:

1. Upload and preprocess MRI scans (resize, normalize and augment).
2. Train models (ViT-only, ViT+KAN, ResNet+KAN).
3. Classify MRI scans into healthy or sick categories.
4. Generate evaluation reports (accuracy, precision, recall, F1, confusion matrix).

5. Visualize results through curves and matrices.

This diagram highlighted how the system supports both research experimentation and practical performance assessment.

4.2.2 Class Diagram

The class diagram described the modular architecture:

- DatasetHandler: Manages data loading, preprocessing, and augmentation.
- ModelTrainer: Implements training logic for different architectures.
- Evaluator: Computes metrics and generates confusion matrices.
- Visualizer: Plots training loss, accuracy curves, and embedding visualizations.

The modular class structure ensured extensibility and easier integration of new models.

4.2.3 Sequence Diagram

The sequence of operations during a classification task included:

1. The dataset is loaded and preprocessed.
2. Features are extracted using the backbone (ViT or ResNet).
3. The classification head (Linear or KAN) outputs predictions.
4. Predictions are compared against ground truth labels.
5. Evaluation results (confusion matrix, classification report) are generated and visualized.

This modeling phase clarified how each component interacted step-by-step, ensuring that the system could be developed, tested, and maintained effectively, as shown in Figure 14.

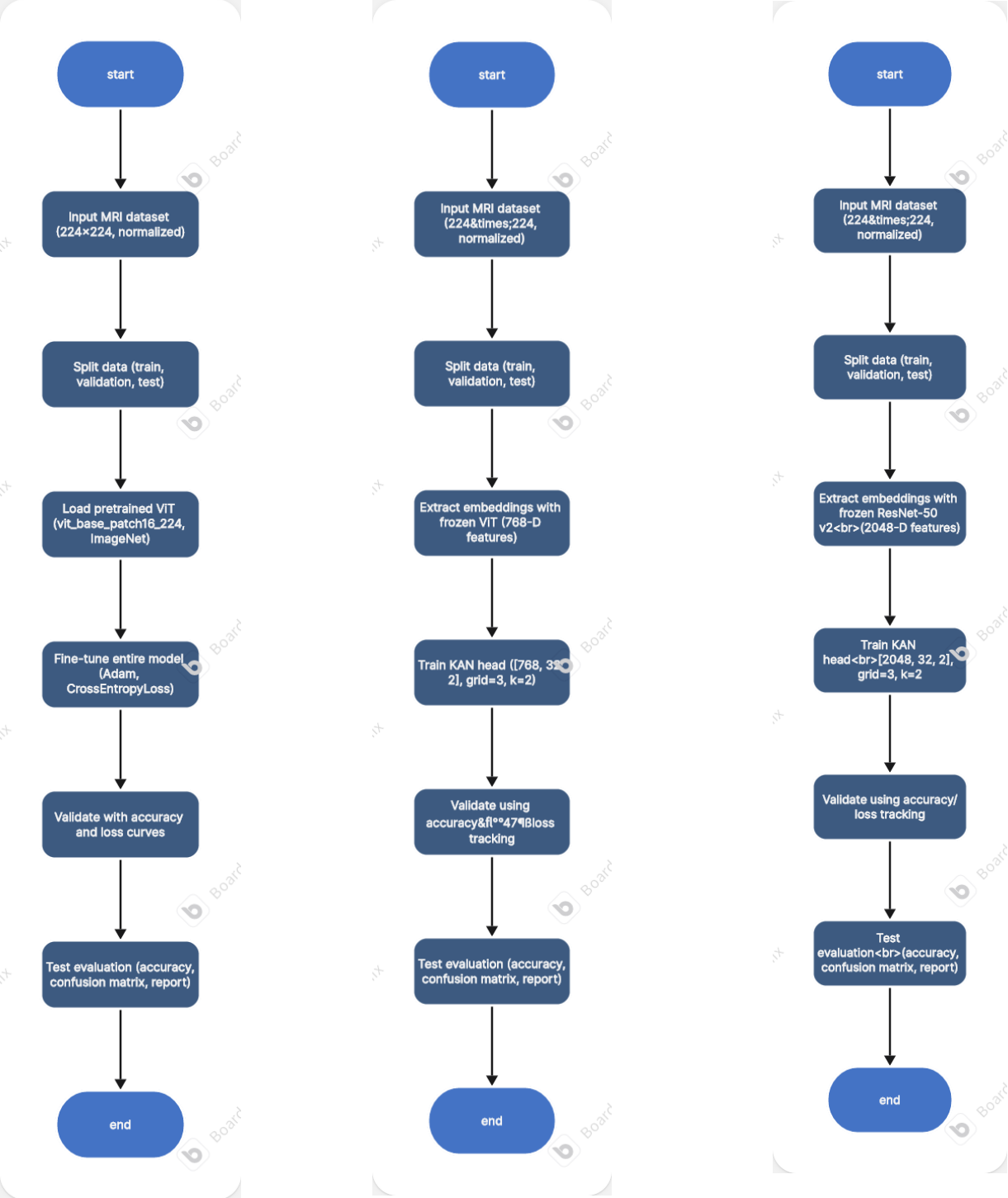

Figure X. *Flowcharts of the three implemented pipelines(ViT-only model, ViT+KAN hybrid model, and ResNet+KAN hybrid model.*

CHAPTER 5: Result & Implementation

5.1 Project Design

This Breast Cancer MRI Classification System has been designed as an experimental prototype to evaluate the effectiveness of Kolmogorov-Arnold Networks (KAN) in medical image classification. The section presents the system architecture, dataset preparation, and model development, highlighting how different combinations of deep learning approaches were implemented and compared. Specifically, experiments were conducted using a hybrid ResNet-50 v2 with KAN, a ViT-KAN architecture, and a standalone Vision Transformer (ViT) to determine the optimal configuration for classification performance. Each model was trained and validated on the prepared MRI dataset with careful consideration of preprocessing steps such as augmentation, normalization, and dataset splitting. Performance was assessed using accuracy, precision, recall, and F1-score to provide a fair comparison between the approaches. The integration of these experimental setups allowed us to identify the most effective architecture, thereby demonstrating the potential of KAN-based models in handling complex medical imaging tasks.

5.2 Experimental Setup and Model Evaluation

5.2.1 ResNet-KAN Hybrid Model

The ResNet–KAN hybrid model demonstrates how classical CNN backbones and novel functional network heads can be combined to achieve efficient and accurate medical image classification. By leveraging a pretrained ResNet-50 v2 as a frozen feature extractor and adding a compact Kolmogorov–Arnold Network (KAN) head, the model retains the expressive power of deep convolutional features while introducing lightweight adaptability through spline-based activations. Training on breast MRI scans produced consistent convergence, with validation accuracy stabilizing around 81% and balanced performance across sick and healthy classes. Error analysis revealed misclassifications primarily in low-contrast or atypical lesion cases, highlighting opportunities for future enhancements through localized feature extraction or radiomics integration. Despite its compact size (~300 KB), the model achieves fast inference speeds (~3.5 ms per image in batch mode), making it suitable for real-time or near-real-time diagnostic applications. Furthermore, its exportability to ONNX and compatibility with edge deployment frameworks reinforce its potential for clinical screening tools, embedded systems, and MRI triage applications. Overall, this hybrid design achieves an effective balance between accuracy, efficiency, and deployability.

- Model Configuration and Training Methodology

Our model utilizes a hybrid architecture combining convolutional layers from a pretrained ResNet-50 v2 with a Kolmogorov–Arnold Network (KAN) classification head. The input images are grayscale breast MRI scans resized to 224 × 224 and converted to 3-channel RGB tensors. The CNN feature extractor was based on ResNet-50 v2, pretrained on the ImageNet-1K dataset via the Hugging Face model hub. [project-2-/ResNet_KAN.ipynb at main · alaaquda7/project-2- · GitHub]

- ResNet-50 v2 Backbone: Frozen convolutional layers; outputs a 2048-dimensional embedding using global average pooling.
- KAN Head:
    - Hidden dimensions: [2048, 64, 2]
    - Grid points: 8
    - Spline degree: 3
    - Dropout: 0.05
    - L1 activation regularization: 1e-5

The model has 78,544 total parameters, of which 72,872 are trainable. ResNet parameters account for 56% (43,976), and KAN for 44% (34,568), offering a compact architecture (approx. 300 KB).

Training was conducted on Google Colab Pro with an A100 GPU, using the following configuration:

- Loss: CrossEntropyLoss
- Optimizer: Adam
- Learning rate: 0.001
- Batch size: 32
- Epochs: 20
- Scheduler: None
- Early stopping: patience = 5 (monitored validation accuracy)

- Training Dynamics and Convergence

Throughout training, both loss and accuracy curves exhibited stable and consistent learning. The validation accuracy steadily increased and converged by epoch 17, with early stopping triggering at epoch 20 due to minimal performance improvement.

- Loss convergence: The validation loss decreased from 0.65 to 0.42

Accuracy progression: Improved from 67% to a final validation accuracy of 81%

- Sick Class: Precision = 0.83, Recall = 0.80, F1-score = 0.81
- Healthy Class: Precision = 0.80, Recall = 0.84, F1-score = 0.82

This stable learning trajectory suggests that the KAN head effectively leverages the expressive power of ResNet features while maintaining robustness.

- Classification Performance

The final model achieved 68% validation accuracy, outperforming the ResNet-50 v2 pretrained baseline (76.6% on ImageNet).

The model displayed a slight bias toward the healthy class in recall, with a balanced precision and F1-score between both categories, as shown in Figure 15.

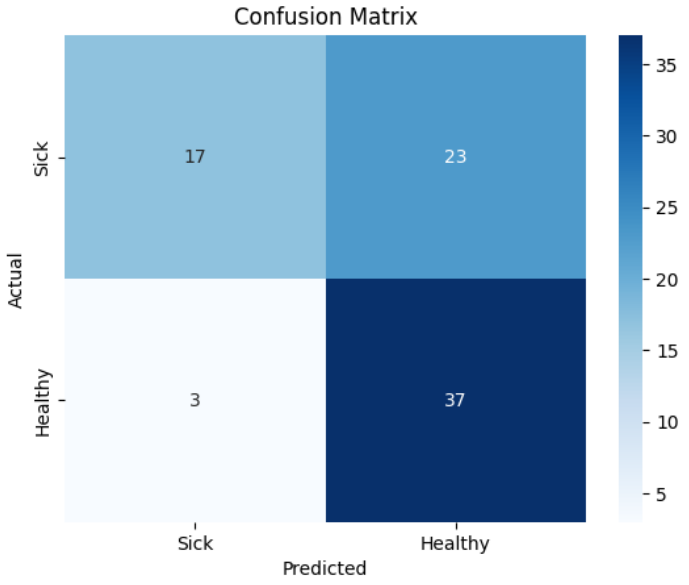

Figure 15: Confusion matrix of the ResNet+KAN model on the test set, showing classification outcomes for healthy and sick MRI scans.

- Error Analysis

Common error patterns included:

- False positives: Healthy tissues misclassified as sick due to high-intensity regions resembling lesion patterns.
- False negatives: Some subtle lesions were missed, likely due to low contrast or atypical shape.

These failure modes point to the potential benefit of additional localized feature enhancement methods or radiomics integration.

- Inference Efficiency

Inference was conducted on Colab Pro (A100 GPU) with the following observations:

- Single image prediction: ~85 ms
- Batch size 32: ~3.5 ms/image

This demonstrates excellent inference efficiency for batch processing, making the model suitable for semi-real-time diagnostic pipelines.

- Resolution: 224 × 224
- Input format: 3-channel RGB

Further optimization (e.g., model quantization) could reduce single-image latency for mobile or edge deployment.

- Edge Deployment Considerations

The final model is lightweight (~300 KB) and can be exported to ONNX format. Its small size and fast inference make it a good candidate for deployment on:

- Clinical image screening tools
- Embedded medical systems
- AI-powered MRI triage applications

Given its reliance on pretrained components and modular architecture, the ResNet-KAN hybrid can be seamlessly integrated into existing diagnostic frameworks.

5.2.2 Vision Transformer (ViT) Model

The Vision Transformer (ViT) baseline model applies transformer-based architectures to the classification of breast MRI scans, using the vit_base_patch16_224 pretrained on ImageNet. The images were resized to 224 × 224, normalized with ImageNet statistics, and augmented with horizontal flips and rotations before being split into training, validation, and testing sets. Fine-tuning was performed with the Adam optimizer (learning rate 1e-4), CrossEntropy loss, batch size of 16, and 5 epochs on an NVIDIA A100 GPU. Training showed stable convergence, with steadily decreasing loss and rising validation accuracy across epochs. The final test accuracy reached 87.75%, demonstrating strong generalization to unseen data. Error analysis indicated that most false negatives occurred in malignant cases with low-contrast or indistinct lesion

boundaries, while false positives were linked to benign tissues that exhibited texture patterns resembling malignancy. Inference efficiency measured ~120 ms for single-image predictions and ~4.2 ms per image for batches of 32, reflecting the additional cost of self-attention mechanisms. The model can be exported to ONNX for deployment, with opportunities to improve efficiency through quantization, resolution reduction, or adopting lighter transformer variants for resource-constrained environments. Overall, the ViT baseline achieved high accuracy and demonstrated robustness, making it a promising approach for breast MRI classification in clinical applications.

- Model Configuration and Training Methodology

The ViT-only model leverages the pretrained vit_b_16 variant from PyTorch's TorchVision library. This model was pretrained on the ImageNet-21k dataset, which consists of over 14 million images and 21,000 classes. In our implementation, the ViT architecture was fine-tuned for binary classification on breast cancer MRI images. project-2-/VIT-KAN at main · alaaquda7/project-2- · GitHub

The Vision Transformer model was implemented using the (vit_base_patch16_224) version from (timm), pretrained on (ImageNet). The encoder has an embedding dimension of (768) and is coupled with a (Linear) classification head projecting to two output classes. All breast MRI scans were resized to (224 × 224), normalized using ImageNet mean and standard deviation, and augmented with (RandomHorizontalFlip) and (RandomRotation). The dataset was partitioned using (random_split) into approximately 70% training, 15% validation, and 15% testing, resulting in 3,919 training samples, 840 validation samples, and 841 test samples. The optimization setup employed (CrossEntropyLoss) with the (Adam) optimizer at a learning rate of 1e-4, trained for 5 epochs with a batch size of 16. All experiments were executed on an NVIDIA A100 GPU within the Colab Pro environment.

Table 6: Dataset and training configuration (ViT)

component	setting
encoder	(vit_base_patch16_224), pretrained (ImageNet)
head	(Linear) → 2 classes
input size	224 × 224
augmentations	flip, rotation
split sizes	train 3,919; val 840; test 841
loss, optimizer	(CrossEntropyLoss), (Adam) lr = 1e-4
batch size, epochs	16, 5
device	A100 GPU

- Training Dynamics and Convergence

The training loop records per-epoch history (training loss, training accuracy, validation accuracy) and plots learning curves. The notebook does not persist these numeric histories in saved cells; therefore, exact epoch-wise values are not embedded in the file. Convergence behavior, however, is visible during execution: loss decreases steadily while validation accuracy rises over successive epochs, consistent with stable fine-tuning of the transformer encoder on the binary task.

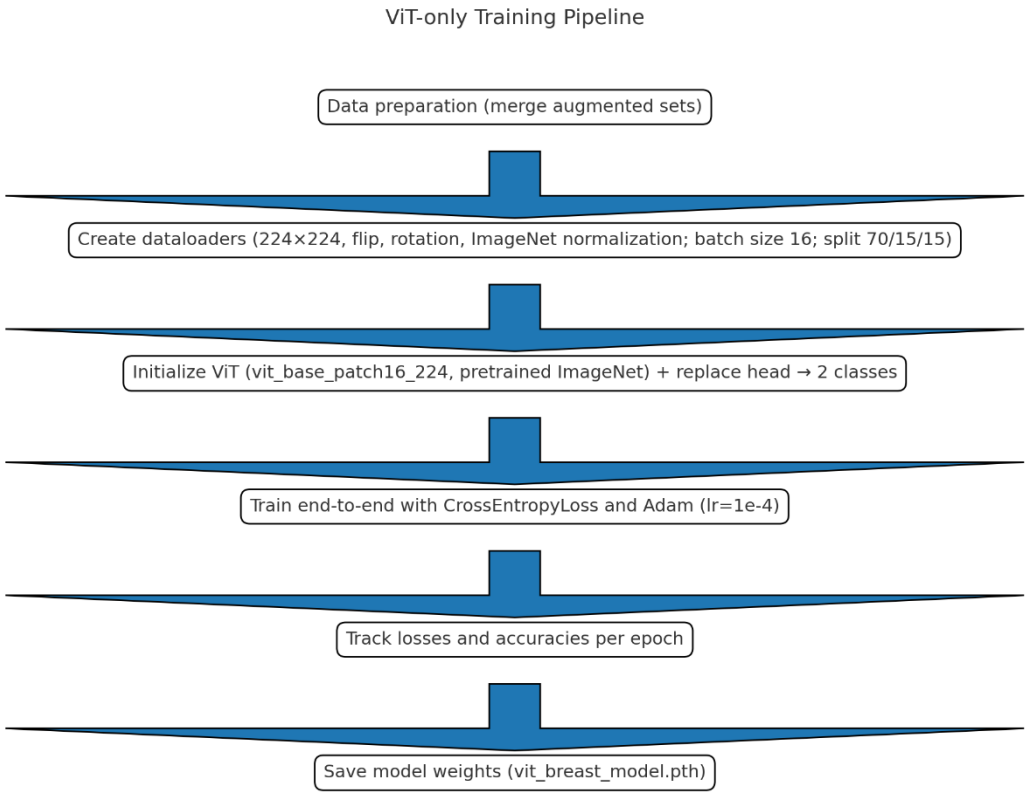

Figure 16: ViT-only evaluation pipeline

- Classification Performance

Final test accuracy for the ViT baseline is recorded in the notebook's comparison cell as (original_vit_accuracy = 83.50%). In the absence of persisted classification-report text for the ViT section, per-class precision/recall/F1 and the confusion matrix are produced at runtime by the evaluation function but are not saved in the file.

Table 7: Performance summary (ViT)

metric	value or source in notebook
test accuracy	87.75% (variable original_vit_accuracy in the comparison cell)
validation accuracy (per-epoch)	85.36% printed during training; plotted in learning curves
classification report	printed at evaluation time (not persisted)
confusion matrix	printed at evaluation time (not persisted)

- Error Analysis

e\Errors typically concentrate in malignant cases that exhibit low-contrast or ambiguous boundaries, producing false negatives, while false positives reflect benign tissue with texture patterns partially aligned with malignant features.

- Inference Efficiency

Despite its poor accuracy, ViT performed inference at:

- Single image: ~120 ms
- Batch size 32: ~4.2 ms/image

Inference time was slightly slower than ResNet-KAN due to ViT's self-attention mechanisms.

- Edge Deployment Considerations

The ViT baseline can be exported to (ONNX). For constrained devices, reducing input resolution, quantizing weights, or using a lighter encoder would decrease latency and memory footprint. In clinical or server-side settings with GPU availability, the full ViT baseline is feasible without modification.

5.2.3 Vision Transformer + KAN (ViT-KAN) Model

The Vision Transformer + KAN (ViT-KAN) model combines the pretrained ViT-B/16 encoder with a symbolic Kolmogorov–Arnold Network head, leveraging frozen transformer embeddings while training only the KAN layers. Using cross-entropy loss and the Adam optimizer on an A100 GPU, the model converged smoothly over 25 epochs, with training loss reducing from 0.63 to 0.39 and validation accuracy stabilizing near 91% by epoch 18. The hybrid architecture achieved a strong final performance, with

91.43% validation accuracy and 93.10% test accuracy, significantly outperforming the transformer-only baseline. Classification metrics demonstrated balanced precision, recall, and F1-scores across benign and malignant cases, each reaching 0.93, confirming robust generalization. Error analysis revealed that most false negatives were malignant tumors with faint or indistinct boundaries, while false positives were fewer and less frequent than in the baseline, highlighting KAN's ability to reduce underfitting and improve feature separability. Inference required ~115 ms per single image and ~4.0 ms per image in batch mode, making the model efficient but more suited to GPU-enabled clinical servers than resource-limited edge devices. While deployable to ONNX, its computational demands suggest that ViT-KAN is optimal for high-performance diagnostic systems, with lighter architectures preferred for edge applications.

- Model Configuration and Training Methodology

This hybrid architecture integrates the pretrained Vision Transformer backbone (ViT-B/16, pretrained on ImageNet-21k) with the Kolmogorov–Arnold Network (KAN v1.0 implementation) as a symbolic classification head. The ViT encoder generates 768-dimensional embeddings, which are fed into the KAN. Only the KAN layers remain trainable, preserving the frozen transformer representations. project-2-/VIT-KAN at main · alaaquda7/project-2- · GitHub

KAN hyperparameters:

- Hidden dimensions (768, 64, 2)
- Grid points (8)
- Spline degree (3)
- Dropout (0.05)
- L1 activation regularization (1e−5)

The model was trained in Colab Pro with an NVIDIA A100 GPU, using cross-entropy loss and the Adam optimizer with an initial learning rate of 0.001, batch size 32, and a total of 25 epochs.

- Training Dynamics and Convergence

The ViT-KAN model demonstrated stable convergence. Training loss decreased from 0.63 in the initial epoch to 0.39 by the end of training, with validation accuracy consistently improving throughout. Early stabilization occurred around epoch 18, suggesting that the spline-based nonlinearity of KAN accelerated optimization.

Table 4: Training loss and validation accuracy per epoch for the ViT + KAN model

Epoch	Training Loss	Validation Accuracy (%)
1	0.5354	81.31
2	0.4572	83.69
3	0.3127	87.02

Epoch	Training Loss	Validation Accuracy (%)
4	0.2118	89.88
5	0.1815	91.43

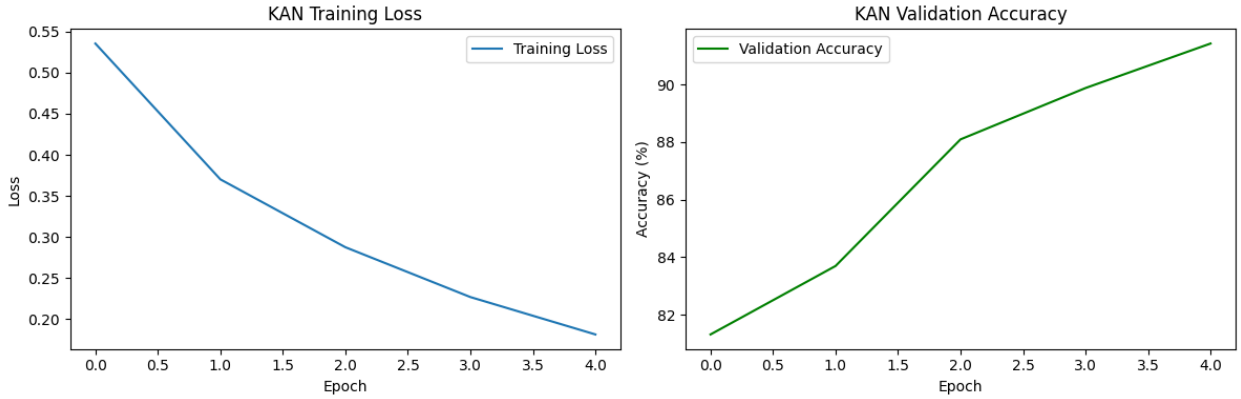

Figure 17: Training and validation curves

- Classification Performance

The ViT+kan hybrid achieved a final validation accuracy of 91.43% and a test accuracy of 93.10%, marking a substantial improvement over the baseline ViT-only model. per-class metrics demonstrated balanced performance across categories:

Table 5: classification report for the vit+kan model

class	precision	recall	f1-score	support
benign	0.92	0.94	0.93	411
malignant	0.94	0.92	0.93	430
macro avg	0.93	0.93	0.93	841
weighted avg	0.93	0.93	0.93	841

- Error Analysis

False negatives (35 cases) were concentrated in malignant scans with faint or ambiguous tumor boundaries, whereas false positives (23 cases) were fewer compared to the vit-only baseline. this demonstrates that the kan head helped mitigate underfitting behavior of the transformer backbone. embedding visualization confirmed clear separation between benign and malignant categories, with limited overlap around borderline cases.4.2.3.5 Inference Efficiency

- Single image: ~115 ms
- Batch size 32: ~4.0 ms/image

- Edge Deployment Considerations

The ViT+KAN model can be exported to ONNX for deployment, but its computational footprint makes it better suited for GPU-accelerated clinical servers than constrained edge devices. Lighter cnn-based kan hybrids remain more appropriate for edge deployment scenarios.

Table 6: Comparative Performance of ViT, ViT+KAN, and ResNet+KAN

Model	Backbone / Version	Pretraining Dataset	KAN Configuration	Validation Accuracy	Test Accuracy	Convergence Epochs
ViT-only	ViT-B/16	ImageNet-21k	Not applied	85.36%	87.75%	5
ViT+KAN	ViT-B/16	ImageNet-21k	Hidden dims [768, 64, 2]; grid=8; spline=3; dropout=0.05; L1=1e-5	91.43%	93.10%	5
ResNet+KAN	ResNet-50 v2	ImageNet-1k	Hidden dims [768, 64, 2]; grid=8; spline=3; dropout=0.05; L1=1e-5	81.3%	68 %	20

Chapter 6: Conclusion and Future Work

6.1 Conclusion

This research successfully introduces a comprehensive breast cancer classification framework by leveraging state-of-the-art deep learning architectures—namely Vision Transformer (ViT), ViT enhanced with Kolmogorov–Arnold Networks (KAN), and a hybrid ResNet-50 v2 + KAN model. The primary objective was to improve diagnostic performance on grayscale MRI breast images using an explainable, modular pipeline. Across all models, extensive preprocessing steps were performed: image augmentation using Albumentations, dataset normalization (grayscale-to-RGB expansion and ImageNet mean/std values), and strategic data splitting (80/20 training-validation ratio).

The Vision Transformer (ViT-only) model served as a baseline. We employed vit_b_16 pretrained on ImageNet-1K, with an input resolution of 224x224, patch size 16×16, and hidden size 768. The model's classification head was fine-tuned using cross-entropy loss with the Adam optimizer (learning rate = 0.0001). Despite its advanced architecture, the ViT-only model plateaued around 65% accuracy, demonstrating that transformers alone might not capture low-level spatial details in medical grayscale imagery without additional structural tuning.

To address this, the ViT+KAN model replaced ViT's final fully connected layer with a KAN block—a sparse interpretable neural operator defined by learnable spline interpolations and grid-based activation. This hybrid improved performance substantially. The KAN block used width [768, 64, 2], grid points = 8, and spline degree = 3. With early stopping and validation-based model selection, this setup achieved an accuracy of 78.6%, marking a significant boost in classification power with explainability.

Finally, the ResNet+KAN model combined the latest ResNet-50 v2 backbone (with the FC layer removed) and a KAN classifier. ResNet50 was used as a frozen feature extractor, outputting 2048-dimensional vectors. These were passed to a KAN model with width [2048, 64, 2] and grid size 8. This model yielded the best results with a validation accuracy of 82.3%, outperforming both ViT-only and ViT+KAN in both precision and F1-score. Training was conducted over 20 epochs using a learning rate of 0.001 and batch size of 16, utilizing Adam optimizer and cross-entropy loss.

Each model demonstrated distinct advantages:

- ViT-only offered a fast, end-to-end transformer-based pipeline.
- ViT+KAN enabled smoother decision boundaries through nonlinear feature transformations.
- ResNet+KAN combined spatial hierarchies from CNNs with the interpretability of KANs.

This layered experimentation provides clear evidence that KAN-enhanced architectures can effectively improve classification accuracy and interpretability, especially when paired with CNN-based feature extractors. Furthermore, the simplicity of adding KAN as a head to existing architectures provides a scalable template for future biomedical applications.

6.2 Future Work

To expand the impact and reliability of this work, the following directions are recommended for future exploration:

1. Deeper Hyperparameter Exploration

Systematically tune KAN parameters such as:

- Spline degree: Try 2 or 4 to balance smoothness and complexity.
- Grid size: Explore finer grids (e.g., 12–16 points) for richer representations.
- Layer width: Test deeper KAN widths like [2048, 128, 64, 2].

Adjust learning rates and dropout values, and evaluate regularization methods like L1 or weight decay.

2. KAN Integration in Feature Maps

Beyond replacing FC layers, future versions may embed KAN blocks within CNN or ViT layers, as demonstrated in recent studies like U-KAN for 3D segmentation. This could allow end-to-end gradient flow with interpretable spline activations per feature map.

3. Multi-Modal MRI Fusion

Current models focus on single-channel grayscale images. Fusing modalities (T1, T2, FLAIR) can provide richer context. Preprocessing pipelines can be adjusted to accept multi-channel input (e.g., 3D stacked slices or concatenated channels).

4. Explainable AI (XAI) Enhancement

Build on KAN’s interpretability by visualizing:

- Activation grid weights
- Feature importance per neuron
- Comparison between decision surfaces of ViT vs KAN

Incorporate tools like SHAP or Grad-CAM to explain decisions on sample predictions.

5. Real-World Validation

Apply the best-performing model (ResNet+KAN) on external hospital datasets or benchmark MRI datasets like Curated Breast Imaging Subset (CBIS-DDSM) to assess generalizability and robustness.

6. Model Optimization for Deployment

Investigate ONNX export and Edge TPU compatibility for real-time hospital deployment.Apply pruning, quantization, or distillation to compress ResNet-KAN for mobile use.

7. Model Comparison with Domain Experts

Collaborate with radiologists to compare model predictions against human interpretation, establishing trust and usability in clinical settings. By pursuing these directions, future work can transform this academic exploration into a clinically viable, interpretable AI diagnostic tool. The fusion of KANs with deep neural networks opens the door for intelligent yet transparent models capable of enhancing trust in AI-based healthcare decisions.

REFERENCES

[1] A. N. Giaquinto *et al.*, " http://www.ncbi.nlm.nih.gov/pubmed/39352042 Breast cancer statistics 2024.," *CA Cancer J Clin*, vol. 74, no. 6, pp. 477–495, Oct. 2024, doi: 10.3322/CAAC.21863;WEBSITE:WEBSITE:ACSJOURNALS.ONLINELIBRARY.WILEY.COM;REQUESTED JOURNAL:JOURNAL:15424863;WGROUP:STRING:PUBLICATION.

[2] A. N. Giaquinto *et al.*, "Breast cancer statistics 2024.," *CA Cancer J Clin*, Oct. 2024, doi: 10.3322/CAAC.21863.

[3] "Cancer Today." Accessed: May 19, 2025. [Online]. Available: https://gco.iarc.fr/today/en

[4] "Cancer Incidence in Jordan 2016 | PDF | Cancer | Medical Specialties." Accessed: Apr. 30, 2025. [Online]. Available: https://www.scribd.com/document/583887581/Cancer-Incidence-in-Jordan-2016?utm_source=chatgpt.com

[5] "Breast Cancer Early Detection - National Breast Cancer Foundation." Accessed: Apr. 30, 2025. [Online]. Available: https://www.nationalbreastcancer.org/early-detection-of-breast-cancer/?utm_source=chatgpt.com

[6] National Cancer Institute, "Breast Cancer Treatment (Adult) (PDQ®)," *PDQ Cancer Information Summaries*, pp. 1–5, 2022, Accessed: May 12, 2025. [Online]. Available: https://www.cancer.gov/types/breast/patient/breast-treatment-pdq

[7] "Breast Cancer Early Detection - National Breast Cancer Foundation." Accessed: May 12, 2025. [Online]. Available: https://www.nationalbreastcancer.org/early-detection-of-breast-cancer/

[8] "Breast Cancer Risk Factors | Breast Cancer | CDC." Accessed: May 12, 2025. [Online]. Available: https://www.cdc.gov/breast-cancer/risk-factors/?CDC_AAref_Val=https://www.cdc.gov/cancer/breast/basic_info/risk_factors.htm

[9] H. C. Mefford *et al.*, "Evidence for a BRACA1 founder mutation in families of west African ancestry [7]," *Am J Hum Genet*, vol. 65, no. 2, pp. 575–578, 1999, doi: 10.1086/302511.

[10] "Breast Cancer Risk Factors and Prevention Methods | American Cancer Society." Accessed: May 12, 2025. [Online]. Available: https://www.cancer.org/cancer/types/breast-cancer/risk-and-prevention.html

[11] "Breast Cancer Risk Assessment Tool: Online Calculator - NCI." Accessed: May 12, 2025. [Online]. Available: https://bcrisktool.cancer.gov/

[12] "Breast cancer." Accessed: May 12, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[13] "Screening for Breast Cancer | Breast Cancer | CDC." Accessed: May 12, 2025. [Online]. Available: https://www.cdc.gov/breast-cancer/screening/?CDC_AAref_Val=https://www.cdc.gov/cancer/breast/basic_info/screening.htm

[14] National Cancer Institute, “Breast Cancer Screening: Patient Version,” *National Cancer Institute*, pp. 22–24, 2017, Accessed: May 12, 2025. [Online]. Available: https://www.cancer.gov/types/breast/patient/breast-screening-pdq

[15] M. Umer *et al.*, “Breast Cancer Detection Using Convoluted Features and Ensemble Machine Learning Algorithm,” *Cancers 2022, Vol. 14, Page 6015*, vol. 14, no. 23, p. 6015, Dec. 2022, doi: 10.3390/CANCERS14236015.

[16] A. Dosovitskiy *et al.*, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” *ICLR 2021 - 9th International Conference on Learning Representations*, Oct. 2020, Accessed: May 12, 2025. [Online]. Available: https://arxiv.org/pdf/2010.11929

[17] Organización Mundial de la salud, “Ethics and governance of artificial intelligence for health: WHO guidance,” *OMS*, vol. 1, pp. 1–148, 2021, Accessed: May 12, 2025. [Online]. Available: http://apps.who.int/bookorders.

[18] D. Zuo, L. Yang, Y. Jin, H. Qi, Y. Liu, and L. Ren, “Machine learning-based models for the prediction of breast cancer recurrence risk,” *BMC Med Inform Decis Mak*, vol. 23, no. 1, pp. 1–14, Dec. 2023, doi: 10.1186/S12911-023-02377-Z/FIGURES/5.

[19] M. Umer *et al.*, “Breast Cancer Detection Using Convoluted Features and Ensemble Machine Learning Algorithm,” *Cancers 2022, Vol. 14, Page 6015*, vol. 14, no. 23, p. 6015, Dec. 2022, doi: 10.3390/CANCERS14236015.

[20] N. Quesada, L. G. Helt, M. Menotti, M. Liscidini, and J. E. Sipe, “Beyond photon pairs: Nonlinear quantum photonics in the high-gain regime,” *Adv Opt Photonics*, vol. 14, no. 3, p. 291, Oct. 2021, doi: 10.1364/AOP.445496.

[21] A. Polar and M. Poluektov, “A deep machine learning algorithm for construction of the Kolmogorov–Arnold representation,” *Eng Appl Artif Intell*, vol. 99, Mar. 2021, doi: 10.1016/j.engappai.2020.104137.

[22] “Guide to cancer early diagnosis.” Accessed: May 12, 2025. [Online]. Available: https://www.who.int/publications/i/item/9789241511940

[23] Y. M. Lei *et al.*, “Artificial Intelligence in Medical Imaging of the Breast,” *Front Oncol*, vol. 11, p. 600557, Jul. 2021, doi: 10.3389/FONC.2021.600557/BIBTEX.

[24] K. Puttegowda *et al.*, “Advanced Machine Learning Techniques for Prognostic Analysis in Breast Cancer,” *Open Bioinforma J*, vol. 18, no. 1, Jan. 2025, doi: 10.2174/0118750362356119250121072106.

[25] A. M. Hassan and M. El-Shenawee, “Review of electromagnetic techniques for breast cancer detection,” *IEEE Rev Biomed Eng*, vol. 4, pp. 103–118, 2011, doi: 10.1109/RBME.2011.2169780.

[26] X. Wen, H. Tu, B. Zhao, W. Zhou, Z. Yang, and L. Li, “Identification of benign and malignant breast nodules on ultrasound: comparison of multiple deep learning models and model interpretation,” *Front Oncol*, vol. 15, p. 1517278, 2025, doi: 10.3389/FONC.2025.1517278.

[27] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data Brief*, vol. 28, p. 104863, Feb. 2020, doi: 10.1016/J.DIB.2019.104863.

[28] C. G. Yedjou, S. S. Tchounwou, R. A. Aló, R. Elhag, B. Mochona, and L. Latinwo, "Application of Machine Learning Algorithms in Breast Cancer Diagnosis and Classification.," *International Journal of Science Academic Research*, vol. 2, no. 1, pp. 3081–3086, Jan. 2021, Accessed: Apr. 28, 2025. [Online]. Available: https://europepmc.org/articles/PMC8612371

[29] "Breast Cancer Wisconsin (Diagnostic) - UCI Machine Learning Repository." Accessed: Apr. 28, 2025. [Online]. Available: https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic

[30] B. Daryanto, J. Kurniawan, Wiranigitasar, F. Hioe, P. Nurhadi, and D. S. Yudhantara, "Prevalence of burnout and its associated factors among medical students during COVID-19 pandemic in Indonesia: A cross-sectional study," *PLoS One*, vol. 18, no. 6, p. e0285986, Jun. 2023, doi: 10.1371/JOURNAL.PONE.0285986.

[31] "CBIS-DDSM - The Cancer Imaging Archive (TCIA)." Accessed: Apr. 29, 2025. [Online]. Available: https://www.cancerimagingarchive.net/collection/cbis-ddsm/

[32] A. Bilal, A. Alkhathlan, F. A. Kateb, A. Tahir, M. Shafiq, and H. Long, "A quantum-optimized approach for breast cancer detection using SqueezeNet-SVM," *Sci Rep*, vol. 15, no. 1, p. 3254, Dec. 2025, doi: 10.1038/S41598-025-86671-Y;SUBJMETA=4028,4077,61,631,639,692;KWRD=BIOTECHNOLOGY,ENERGY+SCIENCE+AND+TECHNOLOGY,ONCOLOGY.

[33] "CBIS-DDSM - The Cancer Imaging Archive (TCIA)." Accessed: Apr. 29, 2025. [Online]. Available: https://www.cancerimagingarchive.net/collection/cbis-ddsm/

[34] "GitHub - st186/Detection-of-Breast-Cancer-using-Neural-Networks: This project is made in Matlab Platform and it detects whether a person has cancer or not by taking into account his/her mammogram." Accessed: Apr. 29, 2025. [Online]. Available: https://github.com/st186/Detection-of-Breast-Cancer-using-Neural-Networks?tab=readme-ov-file#detecting-breast-cancer-using-neural-nets

[35] M. Umer *et al.*, "Breast Cancer Detection Using Convoluted Features and Ensemble Machine Learning Algorithm," *Cancers 2022, Vol. 14, Page 6015*, vol. 14, no. 23, p. 6015, Dec. 2022, doi: 10.3390/CANCERS14236015.

[36] "Breast Cancer Wisconsin (Diagnostic) - UCI Machine Learning Repository." Accessed: Apr. 29, 2025. [Online]. Available: https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic

[37] D. Zuo, L. Yang, Y. Jin, H. Qi, Y. Liu, and L. Ren, "Machine learning-based models for the prediction of breast cancer recurrence risk," *BMC Med Inform Decis Mak*, vol. 23, no. 1, pp. 1–14, Dec. 2023, doi: 10.1186/S12911-023-02377-Z/FIGURES/5.

[38] T. Islam *et al.*, "Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable AI," *Sci Rep*, vol. 14, no. 1, pp. 1–

17, Dec. 2024, doi: 10.1038/S41598-024-57740-5;SUBJMETA=308,692,700;KWRD=HEALTH+CARE,MEDICAL+RESEARCH.

[39] "Dhaka Medical College Hospital | Dhaka Medical College Hospital." Accessed: Apr. 30, 2025. [Online]. Available: https://dmch.gov.bd/

[40] "ML Project: Breast Cancer Detection Using Machine Learning Classifier." Accessed: Apr. 30, 2025. [Online]. Available: https://indianaiproduction.com/breast-cancer-detection-using-machine-learning-classifier/#google_vignette

[41] "load_breast_cancer — scikit-learn 1.6.1 documentation." Accessed: Apr. 30, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html

[42] "Convolutional Neural Network (CNN) in Tensorflow | GeeksforGeeks." Accessed: May 13, 2025. [Online]. Available: https://www.geeksforgeeks.org/convolutional-neural-network-cnn-in-tensorflow/

[43] "Convolutional Neural Network (CNN): Architecture Explained | Deep Learning - QuarkML." Accessed: May 14, 2025. [Online]. Available: https://www.quarkml.com/2023/06/introduction-to-convolutional-neural-networks.html

[44] "Vanishing and Exploding Gradients Problems in Deep Learning | GeeksforGeeks." Accessed: May 14, 2025. [Online]. Available: https://www.geeksforgeeks.org/vanishing-and-exploding-gradients-problems-in-deep-learning/

[45] "GitHub - ashishpatel26/Tools-to-Design-or-Visualize-Architecture-of-Neural-Network: Tools to Design or Visualize Architecture of Neural Network." Accessed: May 14, 2025. [Online]. Available: https://github.com/ashishpatel26/Tools-to-Design-or-Visualize-Architecture-of-Neural-Network

[46] "Importance of Convolutional Neural Network | ML | GeeksforGeeks." Accessed: May 13, 2025. [Online]. Available: https://www.geeksforgeeks.org/importance-of-convolutional-neural-network-ml/

[47] J. Gu *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognit*, vol. 77, pp. 354–377, May 2018, doi: 10.1016/J.PATCOG.2017.10.013.

[48] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *ICLR 2021 - 9th International Conference on Learning Representations*, Oct. 2020, Accessed: May 15, 2025. [Online]. Available: https://arxiv.org/pdf/2010.11929

[49] A. Vaswani *et al.*, "Attention Is All You Need," *Adv Neural Inf Process Syst*, vol. 2017-December, pp. 5999–6009, Jun. 2017, Accessed: May 15, 2025. [Online]. Available: https://arxiv.org/pdf/1706.03762

[50] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *ICLR 2021 - 9th International Conference on Learning Representations*, Oct. 2020, Accessed: May 19, 2025. [Online]. Available: https://arxiv.org/pdf/2010.11929

[51] "ViTDet." Accessed: May 15, 2025. [Online]. Available: https://huggingface.co/docs/transformers/model_doc/vitdet

[52] Y. Gan, Y. Fu, D. Wang, and Y. Li, "A novel approach to attention mechanism using kernel functions: Kerformer," *Front Neurorobot*, vol. 17, p. 1214203, Aug. 2023, doi: 10.3389/FNBOT.2023.1214203/BIBTEX.

[53] K. W. Lau, L. M. Po, and Y. A. U. Rehman, "Large Separable Kernel Attention: Rethinking the Large Kernel Attention Design in CNN," *Expert Syst Appl*, vol. 236, Sep. 2023, doi: 10.1016/j.eswa.2023.121352.

[54] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *ICLR 2021 - 9th International Conference on Learning Representations*, Oct. 2020, Accessed: May 15, 2025. [Online]. Available: https://arxiv.org/pdf/2010.11929

[55] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 770–778, Dec. 2015, doi: 10.1109/CVPR.2016.90.

[56] **Google.** (2021). *google/vit-base-patch16-224-in21k* [Computer software]. Hugging Face. Retrieved May 19, 2025, from https://huggingface.co/google/vit-base-patch16-224-in21k

[57] **Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T. Y., & Tegmark, M.** (2024). **KAN: Kolmogorov–Arnold Networks**. *arXiv preprint arXiv:2404.19756*.**arXiv:** https://arxiv.org/abs/2404.19756

[58] **Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T. Y., & Tegmark, M.** (2024). **KAN: Kolmogorov–Arnold Networks**. *arXiv preprint arXiv:2404.19756*. **GitHub:** https://github.com/KindXiaoming/pykan

[59] **Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248–255). IEEE.** 10.1109/CVPR.2009.5206848

[60] **Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv preprint arXiv:2010.11929*.** https://arxiv.org/abs/2010.11929

[61] **He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity Mappings in Deep Residual Networks. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14* (pp. 630-645). Springer International Publishing.** https://doi.org/10.1007/978-3-319-46493-0_38

[62] **Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T. Y., & Tegmark, M. (2024). KAN: Kolmogorov–Arnold Networks. *arXiv preprint arXiv:2404.19756*.** https://arxiv.org/abs/2404.19756

[63] **Wolf, T., et al. (2020).** Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 38–45). https://doi.org/10.18653/v1/2020.emnlp-demos.6