UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA

Image Reconstruction aided by Machine Learning for Breast Microwave Imaging

Tiago Alexandre Marinho Machado da Silva

Mestrado em Engenharia Biomédica e Biofísica

Dissertação orientada por:
Prof. Dra. Daniela Godinho
Prof. Dra. Raquel Conceição

2025

Agradecimentos

Gostaria de começar por expressar a minha gratidão às minhas orientadoras, Prof.ª Dr.ª Daniela Godinho e Prof.ª Dr.ª Raquel Conceição, por todo o apoio, orientação e paciência ao longo de todo este processo. Estendo também os meus agradecimentos a todos os professores que, de alguma forma, contribuíram para o meu percurso académico durante a minha estadia na Faculdade de Ciências da Universidade de Lisboa, e que intensificaram o meu gosto pelo conhecimento, em especial pela área da Engenharia Biomédica e Biofísica.

Gostaria igualmente de agradecer à Ana Catarina Pelicano, por ter disponibilizado os dados utilizados nesta dissertação.

Agradeço igualmente a todos os colegas e amigos que fiz ao longo destes últimos 5+1 anos, acompanhando-me durante toda a jornada ou apenas em partes dela, com um agradecimento especial à Ana Beatriz, Catarina, Diogo, Gonçalo, Hugo, Inês, Joana, Marco, Rita, Tiago.

À minha família, os mais sinceros agradecimentos por serem os meus pilares e por me oferecerem sempre apoio incondicional ao longo da minha vida, especialmente aos meus pais, à minha irmã e à minha avó.

Por fim, agradeço a mim mesmo pela resiliência, esforço e dedicação ao longo de todo o meu percurso académico.

Abstract

Breast cancer remains the most prevalent malignancy among women worldwide, and early detection is essential to improve treatment outcomes and survival rates. However, current imaging modalities, such as mammography, ultrassound, and MRI, present limitations in terms of cost, comfort, availability, and diagnostic sensitivity, particularly in women with dense breast tissue. In this context, Microwave Imaging (MWI) has emerged as a promising alternative, offering non-ionizing, cost-effective, and portable solutions for breast cancer detection. Nonetheless, traditional MWI reconstruction techniques such as Delay-And-Sum (DAS) are constrained by low resolution, noise sensitivity, and limited image quality. To address these challenges, this dissertation explores the application of Machine Learning (ML) algorithms, namely K-Nearest Neighbors (KNN), Decision Trees (DT), and Support Vector Machines (SVM), to undergo image reconstruction in breast MWI. A scoping literature review was conducted, highlighting current trends, challenges, and research gaps in the application of ML to MWI. A methodology was developed, involving signal acquisition using numerical breast phantoms, Feature Engineering (FE) pipelines (with optional normalization and window trimming), and model optimization via cross-validation. The classification performance was assessed using metrics such as sensitivity, specificity, accuracy, precision, and F1-score, while reconstruction quality was evaluated using Intersection over Union (IoU), Dice Similarity Coefficient (DSC), Relative Volume Error (RVE), and Matthews Correlation Coefficient (MCC). Results show that ML-based reconstructions outperform DAS in visual assessments, demonstrating improved tumor localization and reduced noise. Among the models tested, SVM generally achieved superior classification and reconstruction performance across multiple datasets. This work confirms the potential of integrating ML techniques in MWI for breast cancer imaging and provides a reproducible pipeline that can be further expanded with more advanced models and experimental data. Overall, this dissertation contributes to advancing reliable, accessible, and non-invasive imaging methods for breast cancer detection.

Keywords: Microwave Imaging, Breast Cancer, Machine Learning, Feature Engineering.

Resumo

O cancro da mama continua a ser a neoplasia mais prevalente entre mulheres em todo o mundo e uma das principais causas de mortalidade feminina por cancro. De acordo com a Organização Mundial da Saúde, em 2022 foram diagnosticadas 2,3 milhões de mulheres com cancro da mama, tendo ocorrido aproximadamente 670 mil mortes atribuídas a esta doença. A deteção precoce desempenha um papel fundamental na eficácia dos tratamentos, permitindo intervenções menos invasivas e aumentando significativamente as taxas de sobrevivência. Contudo, os métodos convencionais de imagem médica, nomeadamente a mamografia, a ressonância magnética (*Magnetic Resonance Imaging* – MRI) e a tomografia por emissão de positrões (*Positron Emission Tomography* – PET), apresentam várias limitações, como o custo elevado, o desconforto físico, o uso de radiação ionizante e a diminuição da sensibilidade diagnóstica em mulheres com mamas densas.

Neste contexto, a imagem por micro-ondas (*Microwave Imaging* – MWI) tem vindo a destacar-se como uma alternativa promissora. Esta técnica utiliza campos eletromagnéticos na gama de frequência das micro-ondas (300 MHz a 30 GHz) para reconstruir imagens internas do corpo humano com base nas propriedades dielétricas dos tecidos. O seu carácter não ionizante, o baixo custo, a portabilidade e a ausência de desconforto ou compressão mamária tornam-na particularmente apelativa para rastreios frequentes, sobretudo em regiões com recursos limitados. No entanto, as abordagens tradicionais de reconstrução de imagem em MWI, como o algoritmo *Delay-and-Sum* (DAS), sofrem de baixa resolução espacial, elevada sensibilidade ao ruído e presença de artefactos, comprometendo a fiabilidade do diagnóstico.

Foi realizada uma revisão sistemática da literatura sobre o uso de aprendizagem automática (*Machine Learning* – ML) em MWI, tanto para tarefas de classificação como de reconstrução. Esta revisão, submetida como artigo científico, identificou 124 estudos relevantes a partir de 828 documentos analisados, dos quais 37 dizem respeito a tarefas de reconstrução e 87 a tarefas de classificação. Os principais desafios identificados incluem a escassez de conjuntos de dados padronizados, a baixa diversidade de casos clínicos e a necessidade de validação com dados experimentais e clínicos reais. A revisão também evidencia a tendência crescente da utilização de redes neuronais profundas, embora modelos mais simples ainda apresentem desempenho competitivo com menor complexidade computacional.

Para ultrapassar as limitações das abordagens tradicionais, esta dissertação propõe a aplicação de algoritmos de ML como ferramenta de auxílio na reconstrução de imagens MWI da mama. Os algoritmos explorados são k-vizinhos mais próximos (*K-Nearest Neighbors* – KNN), árvores de decisão (*Decision Tree* – DT) e máquinas de vetores de suporte (*Support Vector Machines* – SVM), escolhidos pela sua simplicidade, versatilidade e potencial de generalização em tarefas de classificação. O principal objetivo consiste em validar a eficácia destes modelos na deteção de tumores mamários, aumentando a qualidade das reconstruções.

A metodologia desenvolvida nesta dissertação é composta por várias etapas fundamentais. A aquisição dos sinais utilizados foi realizada fora do âmbito deste trabalho; assim, o primeiro passo con-

sistiu na realização de um pré-processamento, incluindo a subtração do sinal saudável, a aplicação de janelas temporais e a extração de características relevantes. Foram concebidas diferentes sequências de processamento de engenharia de características (*Feature Engineering* – FE): F24 (baseada em características derivadas do sinal), DWTD, DWTA e DWTA2 (baseadas em coeficientes de transformadas de *wavelet*), com a opção de adicionar a normalização das características e utilizar uma janela temporal mais curta (*Window Trimming* – WT). Cada observação correspondia a um píxel na imagem, acompanhado de um rótulo binário indicando a presença ou ausência de tumor. Inicialmente, foi testada a subtração do sinal do espaço livre, ou seja, o sinal correspondente a simulações apenas com as antenas; no entanto, devido à fraca qualidade dos resultados, optou-se por subtrair o sinal sem a resposta do tumor, deixando apenas a resposta tumoral — uma abordagem ideal, mas não realista.

Foi realizada uma análise dos conjuntos de dados após a aplicação das técnicas de FE com diferentes estratégias de extração de características. Para essa análise, foram utilizados diagramas de caixa e histogramas das quatro características com maior desvio padrão. Os histogramas foram construídos de forma distinta para os dados com rótulo de presença de tumor e para os de ausência de tumor, resultando em oito histogramas por combinação de conjunto de dados e estratégia de FE. Foram também obtidas matrizes de correlação para cada combinação, separadamente para dados com e sem tumor.

Seguidamente, os dados foram divididos em conjuntos de treino e teste, sendo aplicada validação cruzada para otimização de hiperparâmetros. A performance dos modelos foi avaliada com base em métricas de classificação, sensibilidade, especificidade, exatidão, precisão e *F1-score*, e de reconstrução de imagem, interseção sobre união (*Intersection over Union* – IoU), coeficiente de similaridade de Dice (*Dice Similarity Coefficient* – DSC), erro relativo do volume (*Relative Volume Error* – RVE) e coeficiente de correlação de Matthews (*Matthews Correlation Coefficient* – MCC). A reconstrução final das imagens foi realizada píxel a píxel, com o objetivo de gerar mapas binários de probabilidade de tumor, posteriormente limiares (>0,9) e refinados por remoção de pequenos aglomerados.

Durante a fase de treino e otimização dos hiperparâmetros, a estratégia de extração de características que produziu os melhores modelos foi a F24, para os algoritmos SVM e KNN, e a DWTA para o algoritmo DT. Entre os algoritmos testados, o SVM destacou-se por apresentar, de forma consistente, os melhores resultados em termos de F1-score e métricas de reconstrução em múltiplos conjuntos de dados (D1, D2 e D3), abrangendo diferentes tipos de tumores (benignos e malignos). O KNN também demonstrou bom desempenho, particularmente com FE normalizadas, enquanto o DT apresentou desempenho ligeiramente inferior.

Adicionalmente, foram analisadas visualmente as imagens reconstruídas, permitindo comparar diretamente a performance dos modelos. Em todos os conjuntos de dados, as reconstruções baseadas em ML apresentaram localização mais precisa dos tumores e menor número de artefactos em comparação com o DAS. Observou-se uma elevada sobreposição com a área tumoral verdadeira, conforme indicado pelos valores de IoU e DSC. As melhores métricas foram obtidas com o conjunto de dados D2, utilizando o modelo SVM treinado com a FE F24→Norm; o modelo KNN, também treinado com a FE F24→Norm e com o D2, apresentou resultados bastante próximos. Os resultados obtidos demonstram que todos os modelos de ML superaram o algoritmo DAS em termos de qualidade de imagem.

Os principais contributos desta dissertação são os seguintes: (i) desenvolvimento de técnicas de FE reprodutíveis para reconstrução de imagens em MWI utilizando ML; (ii) comparação sistemática entre diferentes algoritmos de classificação e estratégias de FE; (iii) validação quantitativa com múltiplas métricas e conjuntos de dados diversos; e (iv) proposta de um fluxo de trabalho completo que pode ser aplicado em futuros trabalhos, inclusive com dados reais. Esta abordagem representa um passo na transição da MWI do domínio experimental para a prática clínica, sobretudo em rastreios acessíveis,

seguros e de baixo custo.

Em suma, os resultados confirmam que a integração de algoritmos de aprendizagem automática na reconstrução de imagens por micro-ondas pode aumentar significativamente a qualidade das imagens, melhorar a localização de tumores e reduzir as limitações associadas a métodos tradicionais como o DAS. A utilização de modelos supervisionados, combinados com estratégias de extração e seleção de atributos, permite explorar eficazmente a riqueza dos sinais de micro-ondas, tornando a deteção precoce do cancro da mama mais eficaz e viável em diferentes contextos clínicos.

Palavras chave: Imagem por Micro-ondas, Cancro da Mama, Aprendizagem Automática, "Feature Engineering".

Contents

List of Figures

List of Tables

List of Acronyms

AD Alzheimer’s Disease

AdaBoost Adaptive Boosting

AI Artificial Intelligence

ANN Artificial Neural Networks

AUC Area Under the Curve

BIM Born Iterative Method

BMT one Benign and one Malignant tumor

BT Benign tumor

CART Classification And Regression Trees

CNN Convolutional Neural Network

CORR correlation coefficient

CSI Contrast-Source Inversion

CV-MWINet Complex-Valued combined model

CVNN Complex-Valued Neural Network

DAS Delay-And-Sum

DBIM Distorted Born Iterative Method

DBT Double Benign tumor

DL Deep Learning

DMT Double Malignant tumor

DSC Dice Similarity Coefficient

DT Decision Trees

FE Feature Engineering

FEM Finite Element Method

FN False Negatives

FP False Positive

gdi Gini’s Diversity Index

GDMI Graph Degree Mutual Information

GRS Gaussian Random Spheres

H Healthy

HDI Human Development Index

IARC International Agency for Research on Cancer

ICH Intracranial Hemorrhage

IoU Intersection over Union

IQR Interquartile Range

IS Ischemic Stroke

KNN K-Nearest Neighbors

LDA Linear Discriminant Analysis

LR Logistic Regression

MBINet Microwave Brain Image Network

MCC Matthews Correlation Coefficient

ML Machine Learning

MRI Magnetic Resonance Imaging

MSE Mean Squared Error

MT Malignant tumor

MWI Microwave Imaging

MWT Microwave Tomography

NB Naive Bayes

NCC Normalized Cross-Correlation

NH Non-Healthy

NNRGA Neural Network Real Genetic Algorithm

NRMSE Normalized Root Mean Squared Error

NT Non-tumor

PET Positron Emission Tomography

PG-SACC-CNN Physics-Guided Structurally-Aware Complex Cascaded Convolutional Neural Network

QDA Quadratic Discriminant Analysis

QIF Quantitative Imaging Function

RBF Radial Basis Function

RF Random Forest

RMSE Root-Mean-Square Error

rMWI Radar Microwave Imaging

RV-CNN Real-Valued Convolutional Neural Network

RV-DNN Real-Valued Deep Neural Network

RV-MWINet Real-Valued combined architecture

RVE Relative Volume Error

RVNN Real-Valued Neural Network

Self-ONN Self-Organized Operational Neural Network

SFP Synthetic Focal Points

SSIM Structural Similarity Index Measure

Std Standard Deviation

SVM Support Vector Machine

TN True Negative

TP True Positive

UWB Ultra-WideBand

UWCEM University of Wisconsin Computational Electromagnetics

WHO World Health Organization

WT Window Trimming

Chapter 1

Introduction

1.1 Motivation

According to the World Health Organization (WHO), in 2022, 2.3 million women worldwide were diagnosed with breast cancer, and 670,000 died from the disease. In 157 out of 185 countries, breast cancer was the most prevalent cancer in women. WHO also highlights disparities in both the incidence and mortality rates of breast cancer based on the Human Development Index (HDI). In countries with a very high HDI, one in 12 women will be diagnosed with the disease, and one in 71 will die from it. Conversely, in low-HDI countries, one in 27 women is diagnosed with breast cancer, and one in 48 will die from the disease [1].

In Portugal, according to the International Agency for Research on Cancer (IARC), 8,954 women were diagnosed with breast cancer in 2022, representing 28.2% of all cancers diagnosed in women. Breast cancer was responsible for 15.5% of cancer-related deaths among women, with 2,211 fatalities recorded that year [2].

The early detection of breast cancer increases the effectiveness and tolerance of treatments and allows for less invasive procedures, such as more conservative surgeries [1]. An example is lumpectomy, which is the treatment of choice in early stages (I and II) and has outcomes comparable to mastectomy when combined with adjuvant radiotherapy [3].

Regarding early detection, several diagnostic imaging techniques have been developed over the years in order to improve breast cancer detection and improve the sensitivity and specificity of screening protocols. However, these techniques have limitations. For example, the most commonly used imaging method for breast cancer diagnosis, X-ray mammography, has notable disadvantages. These include the discomfort caused by breast compression during the procedure and the use of ionizing radiation (X-rays). Additionally, the high cost of other common imaging modalities, particularly Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET), contributes to the disparity in breast cancer diagnoses, where countries with lower HDI detect fewer cases than those with higher HDI. On the other hand, lower-HDI countries often experience higher mortality rates due to limited financial resources and restricted access to advanced healthcare technologies. Although ultrasound represents a cost-effective alternative and is frequently used for breast cancer detection, its effectiveness is often hindered by the need for specialized expertise, further complicating its implementation in resource-limited settings. Despite these challenges, the most significant issue with traditional screening technique is the high false-positive rate. This is largely due to mammography's sensitivity being heavily dependent on breast density. The similarity in attenuation coefficients between fibroglandular tissue and malignant breast tumor tissue makes it difficult to distinguish them in denser breasts [4]. As a result, mammograms performed on women with dense breasts show a higher false-positive rate compared to those with less dense breasts. For extremely dense breasts, sensitivity drops significantly, ranging from 30% to 48% [5].

A promising alternative that has been under investigation for over 20 years is Microwave Imaging (MWI). This method is particularly appealing due to its cost-effectiveness, speed, portability, non-ionizing nature, and noninvasiveness. MWI is defined as the imaging of an object's internal composition using electromagnetic fields within the microwave frequency range of 300 MHz to 30 GHz [6].

In breast imaging, MWI further distinguishes itself as a painless, non-invasive and less uncomfortable modality compared to conventional imaging techniques, making it a compelling candidate for improving early breast cancer detection. However, this technique also has drawbacks, such as high sensitivity to noise, computational complexity, and reliance on the accuracy of the image algorithms used, such as Delay-And-Sum (DAS). These disadvantages can be addressed or minimized through the use of Machine Learning (ML) algorithms to reduce noise, enhance the quality of image reconstruction, and improve the reliability of results.

In the last decade, there has been a notable increase in the application of ML in the field of medical imaging. Despite this progress, there is still room for improvement, as ML algorithms require not only large amounts of data but also well-designed data processing and Feature Engineering (FE) pipelines for effective training and validation. Nevertheless, ML demonstrates significant potential to reduce the time, cost, and error associated with medical imaging diagnosis [7, 8, 9].

This dissertation is motivated by the ongoing research conducted by the IBEB's research group at the Faculdade de Ciências da Universidade de Lisboa. The work began with Conceição et al. (2010) [10] and was more recently advanced in the Master's dissertation of Moura [11].

1.2 Objectives

The primary objective of this Master's dissertation is to validate the use of ML algorithms for the reconstruction of microwave images, focusing on enhancing breast tumor detection accuracy and reliability. To achieve this, the following secondary objectives are defined:

1. **Data Processing:** Preprocess datasets derived from microwave signals acquired through simulation, ensuring they are suitable for ML training and testing.

2. **Algorithm Optimization:** Select, parameter-tune, and train ML algorithms tailored for MWI reconstruction tasks.

3. **Model Evaluation:** Rigorously test the performance of trained models on a separate test set, ensuring generalization and robustness to unseen data.

4. **Performance Metrics and Validation:** Use well-defined classification and image reconstruction metrics to evaluate the effectiveness of the proposed ML models.

5. **Comparative Analysis:** Conduct a comparative study between the ML-based image reconstructions and those produced by the DAS algorithm to assess relative performance in terms of image quality and tumor detection accuracy.

1.3 Dissertation Overview

This dissertation is structured as follows.

In Chapter 2, a theoretical background is provided, covering the key concepts required to understand the work conducted in this study. The chapter begins with the definition and explanation of MWI and the typical components of MWI equipment. It then introduces the theory behind beamforming algorithms, focusing on the DAS algorithm. The chapter concludes with a subsection dedicated to ML, including an introduction to ML concepts, a description of the algorithms used in this dissertation, and the evaluation metrics employed during both training and testing phases.

Chapter 3 provides a comprehensive review of ML algorithms applied to image reconstruction in MWI. The chapter begins with an analysis of relevant studies in the literature and concludes by summarizing existing research in the field. Additionally, it offers insights into future directions for ML-based MWI approaches. A paper based on this review is currently under consideration for publication in a journal, following minor changes.

In Chapter 4, the methodology employed in this dissertation is detailed. The chapter starts with a description of microwave signal acquisition, including details about the phantoms used, as well as the setup of the equipment for simulations. Subsequently, the construction of the datasets with the acquired signals is discussed, along with a description of the processing framework and the statistical analysis applied to the datasets. The training process of the models is then explained, including hyperparameter tuning (optimized parameters and their ranges), cross-validation implementation, and the used metric for both parameter tuning and model testing. The chapter concludes by describing the implementation of the reconstruction algorithms of microwave images and the metrics used to evaluate the resulting images.

Chapter 5 presents the results, including the statistical analysis of the datasets, the optimal hyperparameters for each model and dataset, the evaluation metrics assessing model performance on the test set, and the reconstructed images generated using the ML algorithm. The chapter concludes with the presentation of the metrics used to evaluate the reconstructed images produced by the ML algorithms and a comparative analysis between ML and the DAS method.

Finally, Chapter 6 summarizes the results obtained, confirms that the proposed objectives were achieved, provides a concise analysis of the impact of this dissertation, and outlines future perspectives for research in this area.

1.4 Contributions

As part of this thesis, a scoping literature review was conducted and published as a peer-reviewed article [12]. This review comprehensively analyzes the current state of microwave breast imaging using machine learning for both classification and image reconstruction tasks. A total of 124 studies were included out of 828 papers screened. The review critically examines the datasets commonly used in the field, highlighting limitations related to dataset size, diversity, and lack of standardization. It also identifies key challenges and proposes future research directions, offering a strong foundation for further investigation. Section 3 provides a condensed version of the scoping review.

Chapter 2

Background Theory

2.1 Microwave Imaging

Microwave Imaging (MWI) is defined as the imaging technique of the internal composition of an object using electromagnetic fields within the microwave frequency range from 300 MHz to 30 GHz [6]. This imaging modality is possible due to the distinct dielectric properties of biological tissues within the microwave spectrum. The differentiation and imaging of tissues rely on varying dielectric properties, particularly the significant differences observed in tissues with high water content (such as tumor) and those with low water content (such as fat or bone) [13, 14, 15].

MWI in medical diagnostics offers notable advantages. Its appeal lies in cost-effectiveness, speed, non-ionizing characteristics, portability, and noninvasiveness. In breast imaging, MWI further stands out for its painless and less discomfort-inducing nature compared to conventional imaging techniques. Despite these benefits, the progress of MWI techniques has been delayed by distinct hardware requirements and data processing speed constraints. Encouragingly, recent advancements in wireless communication and computing technologies have paved the way for continued research and development in this field [15, 16].

MWI can be divided into two major types, Microwave Tomography (MWT) and Radar Microwave Imaging (rMWI). In both approaches, a set of antennas transmits low-power microwave signals into the body, and in turn, these antennas measure the scattered microwave signals [17]. MWT is a quantitative imaging technique that estimates an object's dielectric properties from scattered signals, typically using narrowband data. It relies on iterative algorithms to solve a non-linear inverse scattering problem, which requires regularization to ensure convergence to a stable and physically meaningful solution due to the problem's ill-posed nature [17].

rMWI, also referred to as Ultra-WideBand (UWB) radar or Confocal MWI, is a qualitative technique that relies on the contrast of dielectric properties of various tissues, for instance the contrast between healthy and malignant tissue for the detection of cancerous tissue. The goal is to build an image that shows regions with distinct dielectric properties without attempting to reconstruct the complete profile of dielectric properties [17]. The first algorithm for radar-based breast cancer detection was proposed by Hagness et al. in 1998 [18, 19] and since then different algorithms have been proposed.

MWI has various medical applications, from imaging bones inside the leg [20], brain imaging [21] or breast imaging [22], to classification or detection across different body parts, including the breast tumor [23], brain tumor[24], colon tumor [25], and the stage of Alzheimer's Disease [26].

2.1.1 Delay-And-Sum Beamforming

The traditional algorithms used for rMWI reconstruction are the beamforming algorithms, namely, variations of the Delay-And-Sum (DAS) beamformer. The basis of the DAS beamformer is applying a time delay to the signal from each antenna and summing the outputs together. This process can be mathematically described as:

$$y(\mathbf{r}) = \sum_{i=1}^{N} w_i x_i \big(t - \tau_i(\mathbf{r})\big). \tag{2.1}$$

where:

- $y(\mathbf{r})$ is the reconstructed signal at the focal point $\mathbf{r}$.
- $x_i(t)$ is the signal received by the i^{th} antenna.
- $\tau_i(\mathbf{r})$ is the time delay applied to the signal from the i^{th} antenna based on the focal point $\mathbf{r}$.
- w_i represents the weight assigned to the i^{th} antenna's contribution (typically set to 1 for simplicity in DAS).
- N is the total number of antennas.

The DAS beamformer essentially aligns the signals from different antennas by compensating for the propagation delay and sums them to form the image. This simple approach is widely used due to its ease of implementation and computational efficiency.

Advantages of DAS Beamforming

- **Simplicity**: The DAS algorithm is computationally efficient and straightforward to implement.
- **Low Cost**: It requires minimal hardware and processing resources, making it suitable for low-cost imaging systems.

Disadvantages of DAS Beamforming

- **Low Resolution**: The DAS algorithm struggles with fine detail reconstruction, especially in heterogeneous tissues where variations in permittivity hinder imaging.
- **Sensitivity to Noise**: DAS performs poorly when the data quality is low or contains significant noise, leading to degraded reconstructions.
- **Artifacts**: Reconstructions generated by DAS may exhibit artifacts due to multiple scattering and reflections, which are not compensated for in the simple time-delay model.

2.2 Machine Learning

Machine Learning (ML) is a subdivision of Artificial Intelligence (AI) that aims to develop algorithms and statistical models enabling computers to learn patterns to make predictions and/or decisions without explicit programming for specific tasks. Instead, the model improves its performance over time by interacting with new data. ML has various applications in healthcare, with the major ones being disease identification and diagnosis, medical imaging, drug discovery and manufacturing, disease prediction, and smart health records [27, 28].

Within ML, there are several subtypes, including supervised learning, where models learn from labeled data where input-output pairs are known a priori. The algorithms used in this dissertation fall under this subtype. Supervised learning models can be categorized into several types based on their mathematical and algorithmic foundations. The main categories include:

- **Probabilistic Models**: These models use probabilities to model uncertainties and predict the likelihood of future outcomes. (e.g., Naïve Bayes, Logistic Regression).
- **Tree-based Models**: These utilize a hierarchical tree structure to divide the feature space to perform tasks, whether regression or classification (e.g., Decision Trees (DT), Random Forest).
- **Distance-Based Models**: These rely on the distance between data points in the feature space to make predictions (e.g., Support Vector Machine (SVM), K-Nearest Neighbors (KNN)).

Given that this dissertation aims to highlight the usefulness of ML algorithms in performing MWI reconstruction, it is important to demonstrate their advantages over traditional MWI reconstruction techniques. Therefore, traditional and relatively simple ML algorithms were selected for this task, namely DT, SVM, and KNN.

2.2.1 Decision Tree

As mentioned above, DT is a model that falls into the category of tree-based models within the subtype of ML algorithms, supervised learning.

A decision tree is a structure similar to a flowchart (as shown in an example in Figure 2.1), where each node represents a split of the feature space through a split decision applied to one of the features of the data provided to the model. The resulting branches of that node indicate the test outcome. In summary, the structure of a decision tree is as follows [29].

- **Root Node**: A node that represents the entire dataset, where the first split decision on the data is performed, leading to the first split.
- **Internal Nodes**: Represent subsequent split decision conducted after the data split at the root node.
- **Branches**: Represent the outcomes of the split decision that create a link between a node and another internal node or, ultimately, the leaf nodes.
- **Leaves**: Represent the final decision outcome, i.e., the prediction result of the model. No further splits occur beyond these nodes.

To make a split decision, it is necessary to select the optimal feature for the split based on its distribution. To achieve this, various metrics are employed, among which is the Gini Impurity. The Gini Impurity quantifies the likelihood of a misclassification occurring if a new instance is randomly classified according to the class distribution within the dataset. The mathematical formula to calculate the Gini Impurity is presented in equation 2.2, where C represents the total number of classes, p_i represents the probability of an instance belonging to class i. Another commonly used metric for split decision-making is entropy (or deviance), which measures the amount of impurity or disorder within the dataset, equation

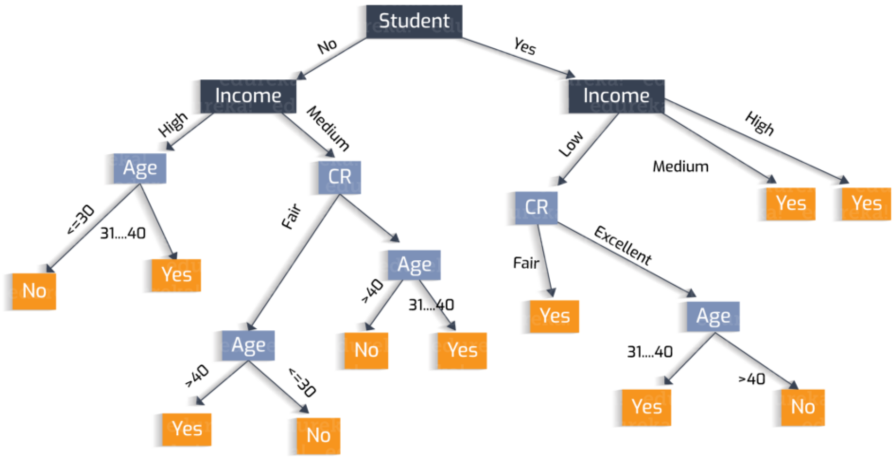


Figure 2.1: Example of a Decision Tree [30].

2.3, where C represents the total number of classes, p_i represents the probability of an instance belonging to class i.

$$\text{Gini Impurity} = 1 - \sum_{i=1}^{C} p_i^2 \tag{2.2}$$

$$\text{Entropy} = -\sum_{i=1}^{C} p_i \log_2(p_i) \tag{2.3}$$

Once a split decision is made at a node, this process is recursively applied to all subsequent nodes until a pre-defined stopping criterion is met. Such criteria may include conditions like all data within a node belonging to the same class, the maximum pre-defined tree depth being reached, or the number of instances in a node falling below the minimum threshold.

DT have both advantages and disadvantages, as is the case with all algorithms. These can be summarized as follows:

Advantages:

- **Ease of Interpretability:** Decision trees are easy to interpret. Their visual representations closely resemble human decision-making processes.
- **Versatility:** Decision trees can be used for both classification and regression tasks.
- **Robustness:** Decision trees require minimal effort for data pre-processing. They are not significantly affected by outliers or missing values.
- **Non-Parametric Nature:** Decision trees are non-parametric, meaning they make no assumptions about the data distribution, the structure of the classifier, or initial seed values.

Disadvantages:

- **Overfitting:** Decision trees are prone to overfitting, especially if they are deep with many nodes.
- **Instability:** Small variations in the data can lead to significant changes in the generated tree.

- **Bias Towards Features with Many Levels:** Features with numerous levels may dominate the tree structure, potentially skewing the results.

2.2.2 Support Vector Machines

SVM is a supervised learning model that belongs to the category of distance-based models. While it can be employed for both classification and regression tasks, SVM is particularly effective in solving classification problems. The primary objective of SVM is to identify a hyperplane that optimally and effectively separates data points into distinct classes within the feature space. To achieve this, the model maximizes the margin, defined as the distance between the hyperplane and the closest data points from each class, thereby enhancing its ability to generalize to unseen data [29].

A notable feature of SVM is the *kernel trick*, which enables the model to handle non-linear decision boundaries by mapping the input data to higher-dimensional spaces. Commonly used kernels in SVM include the following:

- **Linear Kernel:**

  $$K(x,y) = x \cdot y \tag{2.4}$$

  This kernel is used when the data is linearly separable in the original feature space. It calculates the dot product of two input vectors x and y.

- **Polynomial Kernel:**

  $$K(x,y) = (x \cdot y + c)^d \tag{2.5}$$

  Here, d is the degree of the polynomial, and c is a free parameter that trades off the influence of higher-order versus lower-order terms. This kernel is useful for capturing non-linear relationships in the data.

- **Radial Basis Function (RBF) Kernel:**

  $$K(x,y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \tag{2.6}$$

  This kernel maps data into an infinite-dimensional space. The parameter σ controls the spread of the kernel, making it effective for datasets with complex, non-linear decision boundaries.

- **Sigmoid Kernel:**

  $$K(x,y) = \tanh(\alpha x \cdot y + c) \tag{2.7}$$

  In this equation, α and c are parameters that control the shape of the kernel. The sigmoid kernel is often used in neural networks and can mimic the behavior of a multilayer perceptron.

SVM have both advantages and disadvantages, summarized as follows:
Advantages:

- **Effective in High-Dimensional Spaces:** SVM perform well with datasets that have a large number of features, making them suitable for high-dimensional data.

- **Memory Efficient:** SVM use a subset of training points, known as support vectors, in the decision function, which reduces memory requirements.

- **Versatility:** The model allows the use of different kernel functions for the decision function, enabling flexibility in handling both linear and non-linear relationships.
- **Robust to Overfitting (with Proper Parameter Tuning):** By focusing on maximizing the margin, SVM reduce the risk of overfitting, especially in high-dimensional spaces, provided hyperparameters are properly tuned.

Disadvantages:

- **Computational Complexity:** Training an SVM can be computationally expensive, as it involves solving a quadratic optimization problem, particularly for large datasets.
- **Poor Performance with Large Datasets:** SVM scale poorly with dataset size, both in terms of computational cost and memory requirements.
- **Sensitivity to Hyperparameters:** The performance of SVM depends heavily on the choice of hyperparameters, such as the regularization parameter and kernel-specific parameters, which balance the trade-off between maximizing the margin and minimizing classification errors.

2.2.3 K-Nearest Neighbors

The KNN algorithm is a supervised learning model that belongs to the category of distance-based models. It makes predictions by utilizing the K closest points in the feature space. To identify these points, the model calculates the distance between data points using various distance metrics, such as Euclidean, Manhattan, and others. Once the K nearest points are identified, the model assigns the majority class of these K points to the new sample [29].

Common Distance Metrics in KNN includes:

- **Euclidean Distance:**

  $$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{2.8}$$

  This is the most commonly used distance metric, representing the straight-line distance between two points in n-dimensional space.

- **Manhattan Distance:**

  $$d(x,y) = \sum_{i=1}^{n}|x_i - y_i| \tag{2.9}$$

  Also known as the "city block" distance, this metric calculates the sum of absolute differences between the coordinates of two points.

- **Minkowski Distance:**

  $$d(x,y) = \left(\sum_{i=1}^{n}|x_i - y_i|^p\right)^{1/p} \tag{2.10}$$

  This is a generalization of both Euclidean ($p = 2$) and Manhattan ($p = 1$) distances. The parameter p allows flexibility in defining the distance.

- **Chebyshev Distance:**

  $$d(x,y) = \max_{i=1}^{n}|x_i - y_i| \tag{2.11}$$

This metric calculates the maximum absolute difference between the coordinates of two points. It is useful when the largest difference in any dimension dominates the distance.

KNN also has its advantages and disadvantages:
Advantages:

- **Ease of Implementation:** KNN is simple to implement and requires low computational complexity in its algorithmic structure.
- **Adaptability:** KNN easily adapts to new data, as the algorithm stores all training samples in memory and dynamically incorporates new data points for predictions.
- **Few Hyperparameters:** The KNN algorithm has only a few hyperparameters to tune, primarily the number of neighbors (K) and the distance metric.

Disadvantages:

- **Sensitivity to Noise:** Because KNN relies on the nearest neighbors for predictions, it is sensitive to noise and outliers in the data, which can adversely affect performance.
- **Curse of Dimensionality:** KNN struggles with high-dimensional data, as the distance between points becomes less meaningful, reducing the algorithm's accuracy.
- **Prone to Overfitting:** Due to the curse of dimensionality [31], KNN is also susceptible to overfitting, especially when the number of features is large relative to the number of samples.

2.2.4 Data Partitioning

In supervised learning, available data are typically divided into distinct subsets to enable model training, tuning, and evaluation. The training set is used to adjust the model's parameters, allowing it to learn the underlying patterns from the data. The validation set is employed to assess the model's performance during development and to guide hyperparameter tuning, helping to prevent overfitting. Finally, the test set is kept separate and used only for the final evaluation to estimate the model's generalization to unseen data. When cross-validation is adopted, the train set is repeatedly partitioned into training and validation folds, eliminating the need for a dedicated validation set. In this case, the model is trained and validated across multiple folds, and the averaged performance provides a more reliable measure of its expected behavior on new data.

2.2.5 Training and Evaluation of ML Models

To identify the best ML model for the task at hand, whether by finding the optimal parameters for a specific model or by comparing models against one another, it is essential to have an objective method to evaluate their performance. This is equally important to assess whether the model successfully addresses the problem or not. The purpose of this subchapter is to present the metrics selected to evaluate the models used in this dissertation. These metrics are applied both during the training phase, where parameter tuning is performed using cross-validation, and during the evaluation phase.

Since this dissertation employs classification algorithms aimed at performing MWI reconstruction, it is crucial to consider metrics for both classification and image reconstruction tasks.

2.2.5.1 Classification Metrics

For classification tasks, it is essential to select a set of metrics that evaluate different aspects of model performance. Sensitivity, for instance, measures the model's ability to correctly identify true positives within a category, while specificity evaluates its ability to correctly classify true negatives. Although these two metrics can counterbalance each other to some extent, they do not individually provide a comprehensive interpretation of the results. For example, if sensitivity equals 1 (i.e., the model correctly detects all true positives), it does not necessarily imply strong classification ability, as the model could simply classify all samples as positive.

Even when using both sensitivity and specificity, an additional metric is required for optimizing hyperparameters during training. The F1-Score is commonly used for this purpose, as it provides a balanced evaluation of performance, especially in cases of imbalanced datasets or tasks where the true positive and true negative rates are of similar importance, so in this case, the F1-score was used both during the optimization process and afterwards to evaluate the performance of the model obtained with the test set.

Since the primary objective in classifying healthy and tumorous tissue is to correctly identify tumorous samples, the precision metric is particularly relevant. Precision measures the proportion of correctly classified tumorous samples among all samples predicted as tumorous, making it crucial in medical applications where false positives and false negatives have significant implications.

Considering these aspects, five metrics were selected to evaluate the models: sensitivity, specificity, accuracy, precision, and F1-Score. Accuracy was included due to its frequent use in the literature for classification tasks, while the F1-Score was chosen as the optimization metric for model tuning.

The expressions for these four metrics are provided below, where True Positive (TP) refers to correctly predicted positive cases, True Negative (TN) to correctly predicted negative cases, False Positive (FP) to incorrectly predicted positive cases, and False Negatives (FN) to incorrectly predicted negative cases.

- **Accuracy:**
$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{2.12}$$
Accuracy measures the proportion of correctly classified samples among the total samples.

- **Sensitivity (Recall):**
$$\text{Sensitivity} = \frac{TP}{TP+FN} \tag{2.13}$$
Sensitivity measures the ability of the model to correctly identify positive cases.

- **Specificity:**
$$\text{Specificity} = \frac{TN}{TN+FP} \tag{2.14}$$
Specificity measures the ability of the model to correctly identify negative cases.

- **Precision:**
$$\text{Precision} = \frac{TP}{TP+FP} \tag{2.15}$$
Precision assesses the proportion of true positive predictions among all positive predictions.

- **F1-Score:**
$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{2.16}$$

The F1-score is the harmonic mean of precision and sensitivity, providing a balance between the two.

2.2.5.2 Image Reconstruction Metrics

In order to quantitatively assess the quality of tumor image reconstructions, several evaluation metrics are employed to compare the predicted regions against the ground truth. These metrics capture different aspects of reconstruction performance, such as spatial overlap, volumetric accuracy, and overall classification balance. The following measures — Intersection over Union (IoU), Dice Similarity Coefficient (DSC), Relative Volume Error (RVE), and Matthews Correlation Coefficient (MCC) — provide a comprehensive framework for evaluating how well the reconstructed images align with the actual tumor shape, size, and location. The justification for the use of these metrics will be provided later in Section 4.4.

- **IoU / Jaccard Index:**

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{2.17}$$

  IoU measures the overlap between the predicted region and the ground truth region, indicating how well the reconstruction aligns with the actual tumor shape.

- **DSC:**

$$\text{DSC} = \frac{2 \cdot |\text{Predicted} \cap \text{Ground Truth}|}{|\text{Predicted}| + |\text{Ground Truth}|} \tag{2.18}$$

  DSC evaluates the similarity between the predicted and actual regions, balancing precision and recall.

- **RVE:**

$$\text{RVE} = \frac{|\text{Volume}_{\text{Predicted}} - \text{Volume}_{\text{Ground Truth}}|}{\text{Volume}_{\text{Ground Truth}}} \tag{2.19}$$

  RVE quantifies the difference in volume between the reconstructed and actual tumor, indicating whether the model overestimates or underestimates the tumor size.

- **MCC:**

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{2.20}$$

  MCC provides a balanced evaluation of the reconstruction's performance by considering TP, TN, FP, and FN. It is particularly useful for assessing model performance in unbalanced datasets.

Chapter 3

State of the art

This chapter aims to present the state of the art in the application of Machine Learning (ML) algorithms to Microwave Imaging (MWI). It is divided into two parts: the first part reviews the literature on the use of ML for breast image reconstruction, while the second part gives an overview of other applications, particularly classification. The image reconstruction section encompasses 22 studies (Section 3.1), whereas the classification section covers 9 studies (Section 3.2). The review presented is a condensed version of a scoping literature review that was conducted and published as a peer-reviewed article [12].

3.1 Reconstruction using Machine Learning Models

Table 3.1 summarizes the literature on the application of ML algorithms to reconstruct breast images using MWI data.

Kerhet et al. (2006) [32] proposed a 3D approach using a Support Vector Machine (SVM) classifier with a Gaussian kernel, to obtain the probability maps of tumor presence in a phantom of the breast. The approach was evaluated on synthetic data generated with the finite element method. The dataset used in this study contains 41,250 samples, each comprising 99 features: 96 time-steps of the signal and 3 geometrical coordinates of each pixel. Each sample is labeled with a binary class, where the positive class indicates a tumor-affected area, and the negative class represents a tumor-free area. There were 28,750 samples for the training set, 8,750 samples for the validation set, and 3,750 samples for the test set. All sets were scaled to a range from -1 to 1. The authors tested the approach in two different scenarios: one where the dielectric properties of the breast tissue were known a priori, and another where these properties were unknown. Additionally, the authors attempted training on a reduced dataset for the noiseless case of the first scenario (dielectric properties known a priori). They reported that the probability maps obtained for the first scenario showed that the region near the tumor location tended to clearly stand out against the background. For the second scenario (dielectric properties unknown), they observed that although the quality of results was lower, the highest probability values corresponded to the tumor voxels. The specific method used to evaluate the model's performance was not reported, nor were objective results presented, making it unfeasible to ensure reproducibility or to draw conclusions supported by the model's performance.

Ashtari et al. (2010) [33] proposed a Neural Network Real Genetic Algorithm (NNRGA) for breast cancer detection using MWI. The dataset comprised 200 numerical breast profiles and 200 random profiles, which featured randomly distributed tissue types. The data were divided into training, validation, and test sets using a 60/20/20% split. Four breast tissue types were used in the study: mostly fatty, scattered fibroglandular, heterogeneously dense, and very dense. The NNRGA achieved its best permittivity

Table 3.1: Overview of Breast Image Reconstruction Papers in the State of the Art.

Reference	**Algorithm**	**Microwave Image Type**	**Type of Data**
Kerhet et al. [32]	SVM	Radar	Numerical
Ashtari et al. [33]	NNRGA	Radar	Numerical
Conceição et al. [34]	SVM	Radar	Numerical
Shah et al. [35]	CNN	Tomography	Numerical
Khoshdel et al. [36]	CNN	Tomography	Numerical
Khoshdel et al. [37]	CNN	Tomography	Numerical
Mojabi et al. [38]	CNN	Tomography	Numerical
Ambrosiano et al. [39]	ANN	Tomography	Numerical
Mojabi et al. [40]	CNN	Tomography	Numerical
Ambrosiano et al . [41]	ANN	Tomography	Numerical
Ambrosiano et al. [42]	ANN	Tomography	Numerical
Costanzo et al. [43]	U-Net	Tomography	Numerical
Costanzo et al. [44]	U-Net	Tomography	Numerical
Noël et al. [45]	PG-SACC-CNN	Tomography	Numerical
Qin et al. [46]	CNN	Tomography	Numerical
Fontaine et al. [47]	CNN	Radar	Experimental
Bicer [48]	CV-MWINet, RV-MWINet	Radar	Experimental
Costanzo et al. [49]	CVNN	Tomography	Numerical
Khoshdel et al. [50]	CNN	Tomography	Numerical
Borghouts et al. [51]	CNN	Tomography	Numerical
Franceschini et al. [52]	ANN	Tomography	Numerical
Ambrosanio et al. [53]	ANN	Tomography	Numerical

Artificial Neural Network (ANN), Complex-Valued combined model (CV-MWINet), Complex-Valued Neural Network (CVNN), Convolutional Neural Network (CNN), Neural Network Real Genetic Algorithm (NNRGA), Physics-Guided Structurally-Aware Complex Cascaded Convolutional Neural Network (PG-SACC-CNN), Real-Valued combined model (RV-MWINet), Support Vector Machines (SVM)

reconstruction performance with the scattered fibroglandular breast type, reaching a relative error of just 2.7%.

Conceição et al. (2017) [34] used an SVM to help in the diagnosis of breast cancer, while using a Radar Microwave Imaging (rMWI) prototype developed at the University of Bristol. This study extends the authors' previous work [54]. In this study, 36 numerical phantoms were used, developed with layers of tissue mimicking materials. The model considered in the study consisted of 1558 Synthetic Focal Points (SFP), which corresponded to voxels in these numerical phantoms, comprising a total of 1,925,455 measurements, classified as either "hits" (tumor tissue) or "misses" (healthy tissue) in the 36 numerical phantoms. Of the SFP, 779 were "hits" and 779 were "misses." An SVM classifier was trained using 24 features extracted from the signal of each SFP, aiming to distinguish between "hits" and "misses." The dataset was divided into training and testing sets, and a k-fold cross-validation with $k = 10$ was implemented. The authors concluded that the SVM classification approach outperformed the Linear Discriminant Analysis (LDA) classifier presented in their previous work [54].

Shah et al. (2018) [35] employed a Convolutional Neural Network (CNN) to estimate the total electric field from the Born-approximated electric field, treating the problem as an image-to-image transformation. To train the network, the authors used a dataset composed of nine Magnetic Resonance Imaging (MRI)-derived breast numerical phantoms from a public database [55], generating corresponding permittivity and conductivity maps. From these, 75% of the 2D coronal slices from seven randomly selected numerical phantoms were used for training, 25% for validation, and the remaining two numerical phantoms were reserved for testing. The images were normalized and had a resolution of 0.5 mm, while the simulated MWI images had a lower resolution of 1.5 mm. Training was performed using 33×33 image patches. This learning-based approach significantly reduced the l_2 distance between the predicted and true contrast distributions from 57.75 to 28, indicating its effectiveness in capturing the nonlinearities of microwave imaging. The authors concluded that estimating the total electric field from the Born-approximated field using a CNN is a feasible and promising strategy for improving reconstruction accuracy in MWI.

Khoshdel et al. (2019) [36] proposed a deep learning approach based on CNN to enhance MWI reconstructions in a dual-modality microwave–ultrasound system. The study employed 1,200 numerically simulated breast phantoms, with half containing one tumor and the other half containing two. A hybrid reconstruction strategy was introduced, where the output of the Contrast-Source Inversion (CSI) method—augmented with ultrasound-derived tissue profiles—was used as input to the CNN. The CNN significantly outperformed CSI in both Root-Mean-Square Error (RMSE) and Area Under the Curve (AUC), achieving an RMSE of 0.122 and AUC of 0.987, compared to CSI's RMSE of 2.199 and AUC of 0.897. In a subsequent study, Khoshdel et al. (2020) [37] extended this work to 3D MWI using a 10-channel 3D U-Net architecture trained on synthetic data. The dataset comprised 600 numerical breast phantoms with tumor diameters ranging from 1.1 to 1.5 cm, equally divided between single- and double-tumor cases. The CNN was trained on CSI reconstructions obtained using a Finite Element Method (FEM)-based inversion and evaluated on both synthetic and experimental test sets. On synthetic data, the CNN achieved an AUC of 0.957 and RMSE of 1.161, while on experimental data, it obtained an AUC of 0.938 and RMSE of 1.172. In both cases, the CNN outperformed CSI, which achieved AUC of 0.935 (synthetic) and 0.794 (experimental), and RMSE of 1.436 and 1.250, respectively.

Mojabi et al. (2020) [38] introduced a CNN with U-Net architecture to reconstruct MWI images from quantitative dielectric and/or ultrasonic property maps. The dataset included 400 numerical breast phantoms: 150 derived from 3D MRI intensity models capturing diverse fibroglandular patterns, and 250 from 2D MRI intensity models. Among these, 50 phantoms contained tumors across eight distinct configurations. The CNN was trained to map quantitative reconstructions of dielectric and/or ultrasonic properties to corresponding tissue-type and uncertainty maps. Although dataset partitioning was not specified, results showed that the CNN outperformed a Bayesian model in terms of the number of correctly classified pixels, highlighting improved accuracy in tissue classification. In follow-up work, Mojabi et al. (2021) [40] employed two U-Net-based CNNs—one with two depth levels and the other with four—to predict the complex nonlinear relationship between ultrasound compressibility and dielectric properties. The two-level U-Net used 400 numerical phantoms derived from eight tumor-free, MRI-based breast models representing fatty, dense, and heterogeneous tissue types (350 for training, 50 for testing). The four-level U-Net dataset was augmented with 11 rotations at 30° intervals, expanding it to 4,800 phantoms (4,200 training, 600 testing). Inputs were ultrasound compressibility images reconstructed under the Born approximation, and outputs were the real and imaginary parts of the true relative complex permittivity. The CNNs yielded predictions closely matching the true dielectric properties in both shape and magnitude, demonstrating their potential for generating accurate prior information for MWI algorithms. However, the study noted that training under the Born approximation may lead to the omission of tumor regions in some reconstructions. More recently, Khoshdel et al. (2023) [50] proposed a multi-branch CNN architecture to reconstruct two-dimensional permittivity maps for breast cancer detection using MWI. The model comprises two input branches—one for background tissue information and another for scattered field data—and was trained on 4,800 synthetic 2D images generated from MRI-based numerical breast phantoms developed in earlier studies [38, 40]. While the data split was not disclosed, the test set included tumor-free, single-tumor, and two-tumor cases. Reconstruction quality was assessed through qualitative visual analysis, with the authors reporting that the CNN was effective in addressing the electromagnetic inverse scattering problem in breast MWI.

Ambrosanio et al. (2020) [39] investigated the application of Artificial Neural Networks (ANN) for MWI of numerical breast phantoms containing tumors. A dataset of 50,000 numerically simulated breast profiles was employed, divided into training (80%), testing (15%), and validation (5%) sets. Although no quantitative metrics were reported, qualitative visual analysis revealed that the ANN-based reconstructions yielded superior tumor localization and contrast recovery compared to the traditional Distorted Born Iterative Method (DBIM). Building upon this work, Ambrosanio et al. (2022) [41] explored ANN-based quantitative breast MWI using a dataset of 120,000 simulated profiles categorized into four breast tissue classes (A to D), ranging from predominantly adipose (Class A) to highly fibroglandular tissue (Class D). Data were split into 85% for training, 10% for validation, and 5% for testing. The ANN was evaluated against two conventional techniques, AMTISTA and CC-CSI, using Structural Similarity Index Measure (SSIM), Normalized Root Mean Squared Error (NRMSE), and correlation coefficient (CORR) for both permittivity and conductivity reconstructions. The ANN outperformed both methods across all metrics and tissue classes for permittivity, and in most cases for conductivity, except for SSIM in Classes B and D, where CC-CSI slightly surpassed the ANN. The best ANN results for conductivity reconstruction were in Class D (SSIM: 0.417, NRMSE: 0.087, CORR: 0.871), while the highest SSIM for permittivity was in Class A (0.458), the lowest NRMSE in Classes A and B (0.102), and the highest CORR in Classes B and C (0.818). These outcomes highlight the ANN's capability to produce high-fidelity reconstructions across varying tissue compositions. A complementary study by Ambrosanio et al. (2022) [42] employed a fully connected ANN trained on the same dataset (split 85/10/5%) to reconstruct both permittivity and conductivity, achieving NRMSE values of 0.10 and 0.26, respectively.

Borghouts et al. (2023) [51] applied a CNN to image reconstruction and classification of breast MWI data into malignant or healthy categories. The dataset consisted of 160,000 simulated breast profiles, divided into 128,000 for training, and 16,000 for validation and 16,000 for testing, balanced across both classes. Reconstruction performance was assessed using Soft-Dice, Normalized Cross-Correlation (NCC), and NRMSE, while classification performance was evaluated with accuracy, sensitivity, specificity, precision, and F1-score. The CNN achieved a Soft-Dice score of 0.144, NCC of 0.337, and NRMSE of 0.809. For classification, it demonstrated excellent performance, achieving 99.95% accuracy, 99.96% sensitivity, 99.94% specificity, 99.94% precision, and a 99.95% F1-score, confirming the potential of deep learning for end-to-end breast cancer detection in MWI. Using the same dataset, Franceschini et al. (2023) [52] implemented an ANN for breast cancer detection, evaluating reconstruction performance on a pixel-wise basis resulting in accuracy (99.5%), sensitivity (98.9%), specificity (99.9%), and AUC (100%). Despite employing the same dataset as Borghouts et al., direct comparisons are limited due to differences in methodologies.

Extending their work, Ambrosanio et al. (2024) [53] developed a fully connected ANN to generate tumor probability maps from breast MWI data. Using the same dataset of 160,000 simulated breast profiles (80% training, 10% validation, 10% testing), the model achieved a Soft-Dice score of 0.102, NCC of 0.287, and NRMSE of 0.856, indicating reasonable performance and reinforcing the utility of ANNs for probabilistic tumor localization in MWI.

Costanzo et al. have published several studies utilizing data from the University of Wisconsin Computational Electromagnetics (UWCEM) repository [55], in which the output of the quadratic Born Iterative Method (BIM) was used as input to ML models. In Costanzo et al. (2022) [43], a CNN was employed; however, the total number of samples and the dataset partitioning were not fully disclosed. The CNN achieved a mean relative error of 8.09% and an average reconstruction accuracy of 91.91%

across four different breast phantoms. In a complementary study from the same year, Costanzo et al. (2022) [44] used a more controlled dataset consisting of 177 images from two numerical phantoms, with 95% of the images allocated for training and 5% for testing. The CNN achieved relative errors of 3.8% and 7.18% for low- and high-density phantoms, respectively. However, these results may reflect overfitting due to the small dataset size and limited phantom diversity. In subsequent work, Costanzo et al. (2023) [49] introduced a Complex-Valued Neural Network (CVNN) and compared its performance with two Real-Valued Neural Network (RVNN), one trained to reconstruct permittivity and the other conductivity. The dataset comprised 1,500 images, divided into 80% for training and 20% for validation. The CVNN achieved a validation loss of 8.4205, closely matching the RVNN for permittivity (8.0074) and outperforming the RVNN for conductivity (16.2775), suggesting that the CVNN provides comparable or improved performance, particularly when handling complex-valued data.

Noël et al. (2022) [45] proposed a Physics-Guided Structurally-Aware Complex Cascaded Convolutional Neural Network (PG-SACC-CNN) for breast MWI, integrating both microwave and ultrasound modalities. The performance of the PG-SACC-CNN was compared to that of various CNN architectures, including a ResNet-based model. All networks were trained and tested using breast phantom data from the UWCEM repository [55]. The dataset included 2,920 samples, divided into 2,336 for training and 584 for testing. Network performance was assessed using Intersection over Union (IoU), reconstruction error, and SSIM. The PG-SACC-CNN achieved superior results across all metrics, with an IoU of 0.755, a reconstruction error of 0.156, and an SSIM of 0.810, indicating enhanced reconstruction accuracy and preservation of structural features.

Similarly, Qin et al. (2022) [46] employed several CNN variants to jointly invert microwave and ultrasonic data for breast imaging, using the same UWCEM breast phantom dataset. In this case, the dataset comprised 2,180 samples, with 1,920 allocated for training and 240 for testing (there are 20 samples unaccounted for). Among the tested models, the CNN architecture that employed multistream input and multitask learning demonstrated the highest performance, achieving an IoU of 0.7504. These results further support the effectiveness of hybrid CNN-based strategies for multimodal breast imaging tasks.

Fontaine et al. (2023) [47] used a CNN in a portable microwave system, aimed to reduce the device's cost, size, and complexity. The CNN was trained to directly reconstruct rod phantoms from their S_{11} sinograms. The training and testing data consisted of pairs of numerical rod phantoms and their corresponding S_{11} sinograms. In total, 10,000 phantoms were generated, but by randomly flipping and rotating the pairs, the dataset was expanded to 250,000 image-sinogram pairs. The entire dataset was then split into a training set (80%) and a test set (20%). The authors concluded that the CNN's predictions were promising compared to Delay-And-Sum reconstructions and that similar devices, using inexpensive materials, could serve as a viable alternative to conventional image reconstruction methods.

Bicer (2023) [48] evaluated four deep learning models for microwave radar-based breast imaging: a Real-Valued Deep Neural Network (RV-DNN), a Real-Valued Convolutional Neural Network (RV-CNN), a Real-Valued combined architecture (RV-MWINet), and a Complex-Valued combined model (CV-MWINet). The dataset consisted of 1,000 scattered signals acquired from experimental breast phantoms. The RV-DNN and RV-CNN models were evaluated using Mean Squared Error (MSE) and SSIM, while RV-MWINet and CV-MWINet were assessed using SSIM and classification accuracy. All models were evaluated with 10-fold cross-validation. The best performance was achieved by RV-MWINet and CV-MWINet, both attaining an SSIM of 0.999 and a classification accuracy of 99.3%. In comparison, the RV-DNN and RV-CNN models achieved SSIM values of 0.914 and 0.911, and MSE values of 197.401 and 162.089, respectively, demonstrating the higher performance of the combined architectures for both reconstruction and classification of breast tumor profiles.

Table 3.2 summarizes the characteristics of the acquisition systems employed in the studies discussed in this section. These include the number of antenna positions, the configuration of the acquisition system, the mobility of the antenna array, and the operating frequencies. The “No. of Antenna Positions" column shows that most studies use setups with more than 15 antenna elements, commonly ranging from 20 to 30, with only a few exceptions reporting lower counts—such as the study by Kerhet et al. (1 T and 16 R), Ashtari et al. (4 T 16 R), and those by Costanzo et al., which used 10 or 18 antennas for both transmission and reflexion. The “Multistatic vs. Monostatic" column reveals a clear dominance of multistatic configurations, where each transmitted signal is received by all antennas in the array. Monostatic or quasi-multistatic (defined as the transmitted signal is received by all antennas except the transmitting one) systems are less common, and bistatic configurations appear rarely, as in Kerhet et al.’s work. The “Fixed vs. Moving" column highlights a strong incidence of fixed antenna systems, used in nearly all studies, with Bicer et al. being the only example employing a moving array. The “Frequencies" column indicates that 1 GHz is the most frequently used operating frequency across studies, although several works adopt broader or multiple frequency bands—such as Conceição et al. (3–8 GHz), Fontaine et al. (0.7–3 GHz), and Bicer et al. (1–10 GHz)—to enhance imaging quality. Overall, the table reflects a consistent trend in the adoption of fixed, multistatic systems with moderate-to-high numbers of antenna elements and operating frequencies centered around 1 GHz. These choices suggest a growing consensus on hardware configurations that balance complexity, resolution, and practicality in breast microwave imaging research.

Table 3.3 summarizes key characteristics of the datasets used in the reviewed studies. The “Dataset” column describes both the dataset size and what the reported size refers to, which varies across works—for instance, it may represent the number of raw signal samples, image pairs, or patient profiles. This variability reflects the lack of standardization in how dataset size is reported. The “Input Sample” column specifies the actual data type provided to the model, such as scattered fields, dielectric profiles, or reconstructed images. The distinction between the “Dataset” and “Input Sample” columns is necessary because most studies do not clearly define how many input samples are derived from the dataset. To address this, the table descriminates general dataset size from specific model inputs. The “Metric” column reports the performance metric used and its corresponding value, with notable variation across studies. While some consistency exists, MSE and SSIM are frequently reported, other metrics such as IoU, AUC, accuracy, and F1-score also appear. The diversity in data types, input formats, and evaluation criteria highlights the heterogeneous nature of research in this domain. However, more recent studies tend to use larger datasets, more comprehensive performance metrics, and achieve higher accuracy or reconstruction quality, indicating a trend toward increasingly robust and data-driven approaches.

This section reveals a clear predominance of Deep Learning (DL) algorithms in breast image reconstruction tasks. Although synthetic (numerical) data remains more commonly used, several studies employing experimental data—such as phantom-based measurements—have reported promising results. Additionally, when comparisons are possible using the same evaluation metrics, there is a visible trend of improved performance as dataset sizes increase. These encouraging outcomes highlight the importance of continued development in this field. Advancing these applications will require the creation of larger and more diverse datasets, especially incorporating clinical data, to ensure that models are trained with information reflective of real-world conditions.

Table 3.2: Summary table on the equipment used for microwave signal acquisition (Target Zone: Breast). Label “T" represents transmitters, while “R" represents receivers. When these labels are absent it means the respective study did not specify transmitters and receivers. The (*) indicates that the nomenclature used in this paper differs from that employed by the authors of the paper.

Reference	No. of Antenna Positions	Monostatic vs Multistatic	Moving vs Fixed	Frequencies
Kerhet et al. [32]	1 T 16 R	Bistatic	Fixed	6 GHz
Ashtari et al. [33]	4 T 16 R	Multistatic	Fixed	1 GHz
Conceição et al. [34]	60	Multistatic	Fixed	3 - 8 GHz
Shah et al. [35]	24 T 24 R	Multistatic	Fixed	2 GHz
Khoshdel et al. [36]	NA	Multistatic	NA	NA
Khoshdel et al. [37]	24	Multistatic	Fixed	1.1 - 1.5 GHz
Mojabi et al. [38]	30	Multistatic	Fixed	1, 1.5 and 2 GHz
Ambrosiano et al. [39]	30	Quasi-Multistatic*	Fixed	1 GHz
Mojabi et al. [40]	30	Multistatic	Fixed	1 GHz
Ambrosiano et al . [41]	30	Quasi-Multistatic*	Fixed	0.2 - 1 GHz
Ambrosiano et al. [42]	30	Quasi-Multistatic*	Fixed	1 GHz
Costanzo et al. [43]	18 T 18 R	Multistatic	Fixed	1 GHz
Costanzo et al. [44]	18 T 18 R	Multistatic	Fixed	1 GHz
Noël et al. [45]	20	Multistatic	Fixed	1 GHz
Qin et al. [46]	20	Multistatic	Fixed	1 GHz
Fontaine et al. [47]	26	Multistatic	Fixed	0.7 - 3 GHz
Bicer [48]	90 T 90 R	Monostatic	Moving	1 - 10 GHz
Costanzo et al. [49]	NA	Multistatic	NA	NA
Khoshdel et al. [50]	30	Multistatic	Fixed	1, 1.5 and 2 GHz
Borghouts et al. [51]	30	Quasi-Multistatic*	Fixed	NA
Franceschini et al. [52]	30	Quasi-Multistatic*	Fixed	1 GHz
Ambrosanio et al. [53]	30	Quasi-Multistatic*	Fixed	1 GHz

3.2 Other Applications of Machine Learning in Microwave Imaging

ML has been used in the field of MWI for classification purposes for nearly fifteen years, beginning with the work of Conceição et al. in 2010 [10]. The classification tasks range from simpler distinctions, such as distinguishing between benign and malignant tumors in the breast [56], to other classification tasks, such as discerning between Intracranial Hemorrhage (ICH) and Ischemic Stroke (IS) [57]. The studies presented in this section use a variety of algorithms, as well as different types of data. The data used in these studies come from different sources: 1) computational simulation; 2) measurements made using experimental MWI prototypes and phantoms mimicking dielectric properties of tissues; 3) combination of experimental data with phantoms and simulation; 4) clinical studies, which represent the most advanced type of data. Table 3.4 summarizes studies that apply machine learning algorithms to classification tasks using MWI data for both brain and breast imaging applications.

Starting with applications in stroke classification, Zhu et al. (2020) [57] introduced a novel Graph Degree Mutual Information (GDMI) method to directly differentiate between ICH and IS using electromagnetic signals. This method involved data from 100 realistic brain models, derived from complex imaging modalities and segmented MRI images, using an antenna array operating in the 0.7–2 GHz band [62, 63, 64]. Post-stroke patient signals were collected via a MWI device, transformed using an inverse Fast Fourier Transform into 256 (16×16) time series signals from each model, and then converted

Table 3.3: Summary of Dataset Characteristics and Evaluation Metrics from studies listed in Table 3.1.

Reference	Dataset	Input Sample	Metric
Kerhet et al. [32]	41,250 (Signals and Coordinates of the cell)	Signals and coordinates of the cell	NA
Ashtari et al. [33]	200+200 (Breasts + Random Profiles)	FV	Relative Error: 2.7%
Conceição et al. [34]	1,558 (SFPs)	Signal	NA
Shah et al. [35]	NA	NA	L2 Distance: 28.54
Khoshdel et al. [36]	1200 (Breast phantoms)	CSI Reconstruction	RMSE, AUC: 0.122, 0.987
Khoshdel et al. [37]	600 (Training Set, Breast Phantoms)	CSI Reconstructions	RMSE, AUC (Syntetic, Experimental): 1.161, 0.957; 1.172, 0.938
Mojabi et al. [38]	400 (Breast Phantoms)	Dieletric and/or Ultrasound Reconstructions	Overall Error: 0.024
Ambrosiano et al. [39]	50,000 (Breasts Profiles)	Breast Profiles	NA
Mojabi et al. [40]	400, 4,800 (Breast Phantoms)	Reconstructed Ultrasonic (Breast Compressibility)	MSE: 0.097, 0.038
Ambrosiano et al . [41]	120,000 (Breast Profiles)	Breast Profiles	NRMSE, SSIM, CORR: (0.102, 0.458, 0.795
Ambrosiano et al. [42]	120,000 (Breast Profiles)	Breast Profiles	NRMSE (Permittivity, Conductivity): 0.10, 0.26
Costanzo et al. [43]	NA	Quadratic BIM output	Acc, RE: 91.91%, 8.09%
Costanzo et al. [44]	177 (BIM Images)	Quadratic BIM output	Relative error: 3.87%
Noël et al. [45]	2920 (Samples)	Scattered Field	IoU, Error, SSIM: 0.755, 0.156, 0.810
Qin et al. [46]	2180 (Samples)	Scattered Field	IoU: 0.7504
Fontaine et al. [47]	250,000 (Sinogram-Image pairs)	Sinogram-Image pairs	MSE, Accuracy: 0.137, 63%
Bicer [48]	1000 (scattered fields)	Scattered Field	Acc, SSIM: 99.3%, 99.9%
Costanzo et al. [49]	1500 (Contrast Maps)	Quadratic BIM output	Validation Loss: 8.4205
Khoshdel et al. [50]	4800 (Breast Phantoms)	Scattered Field	NA
Borghouts et al. [51]	160,000 (Breast Profiles)	Scattered Field	Acc, Sen, Spe, Prec, F1: 99.95%, 99.96%, 99.94%, 99.94%, 99.95%
Franceschini et al. [52]	160,000 (Breast Profiles)	Scattered Field	Acc, Sen, Spe, AUC: 99.5%, 98.9%, 99.9%, 100%
Ambrosanio et al. [53]	160,000 (Breast Profiles)	Breast Profiles	NRMSE, NCC, Soft-DICE: 0.856, 0.287, 0.102

Accuracy (Acc), Area Under the Curve (AUC), Breast Imaging Modality (BIM), Contrast-Source Inversion (CSI), Correlation Coefficient (CORR), F1-Score (F1), Intersection over Union (IoU), Mean Squared Error (MSE), Not Available (NA), Normalized Cross-Correlation (NCC), Normalized Root Mean Square Error (NRMSE), Precision (Prec), Relative Error (RE), Sensitivity (Sen), Specificity (Spe), Structural Similarity Index Measure (SSIM)

into graphs through the Fast Weighted Horizontal Visibility Algorithm. Features were extracted using the GDMI algorithm and fed into an SVM classifier. The method was evaluated with 50 ICH and 50 IS signals, achieving a sensitivity of 91%, specificity of 98%, and an accuracy of 94% without any noise. When noise was introduced (40 dB, 25 dB, and 10 dB Signal-to-Noise Ratio), the accuracies were 93%, 88%, and 77% respectively, with an overall accuracy of 89%. This demonstrated the GDMI method's robustness and significant potential for stroke classification using MWI.

Adding to the investigation on stroke classification, Pokorny et al. (2023) [59] utilized two datasets, each containing three classes: IS, ICH, and noStroke. The authors developed a multiclass classifier by combining multiple binary classifiers. To improve algorithm performance, they optimized hyperparameters of the SVM algorithm using Bayesian optimization. Acknowledging that increasing data variability might reduce SVM efficiency, they also employed five alternative algorithms: Logistic Regression (LR), Discriminant Analysis, K-Nearest Neighbors (KNN), Naive Bayes (NB), and Decision Trees (DT). The authors concluded that the simulation study demonstrated improved classification accuracy for brain strokes by separating adjacent antennas in MWI systems. Training and testing on the new dataset led to an increased classification accuracy from 66.2% to 68.7%. Additionally, the Cohen's kappa value rose

Table 3.4: Classification using Machine Learning Models (Target Zones: Brain and Breast)

Reference	**Algorithm**	**Type of Data**
Zhu et al. [57]	SVM	Numerical
Conceição et al. [56]	NB, DT, KNN	Numerical
Saied et al. [26]	LR, LDA, KNN, DT, NB, SVM	Numerical
Fasoula et al. [58]	NB, QDA	Clinical
Pokorny et al. [59]	SVM	Numerical
Ullah et al. [60]	KNNs; Random Forests CARTs.	Numerical
Hossain et al. [24]	MBINet	Experimental
Janjic et al. [61]	AdaBoost	Clinical
Papini et al. [23]	Random Forests and SVM	Clinical

Support Vector Machines (SVM), Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Decision Tree (DT), Naïve Bayes (NB), Quadratic Discriminant Analysis (QDA), Random Forests, Adaptive Boosting (AdaBoost), Classification and Regression Trees (CART), Microwave Brain Image Network (MBINet).

from 0.24 to 0.29, indicating a better classification success rate than random chance for both datasets. Although the classification accuracy did not increase compared to previous studies, this article proved valuable by demonstrating that separating adjacent antennas can enhance the accuracy of stroke classification.

Regarding in breast cancer classification, Conceição et al. (2020) [56] employed three different classifiers—NB, DT, and KNN—to study the application of ML algorithms in distinguishing between malignant and benign tumors, using a dataset comprising the signals from 13 benign and 13 malignant tumors with a total of 3,744 signals, with tumor shapes modeled to approximately match the Gaussian Random Spheres (GRS) models used in previous studies [65, 66], where spiculated and microlobulated shapes represented malignant tumors, while round and oval shapes represented benign tumors. Among the classifiers, KNN outperformed DT and NB, achieving an accuracy of 96.2%.

Continuing with breast cancer-related applications, Fasoula et al. (2021) [58] conducted a clinical investigation involving 25 symptomatic patients to explore the capability of the Wavelia semi-automated Quantitative Imaging Function (QIF) in detecting and classifying breast lesions as malignant or benign. This study presents the methodology used in the Wavelia QIF to support breast lesion detection based on lesion persistence, sizing, and characterization within a feature space that encompasses both shape-based and texture-based features. The low-dimensional feature space covered shape-based features such as solidity, and texture-based features including correlation and busyness. The features were extracted from the images. In the Wavelia QIF, a NB and Quadratic Discriminant Analysis (QDA) classifier were trained for the purpose of distinguishing between malignant and benign breast lesions. The results showed a notable level of separability between malignant and benign breast lesions, with a classification loss of 11.5% estimated using 10-fold cross-validation for the trained QDA classifier.

Adding further to the breast cancer classification investigation, Janjic et al. (2023) [61] used a dataset that included 113 breast samples, with 70 categorized as benign and 43 as malignant. These samples were obtained from biopsies conducted on patients with BI-RADS category 4 or 5. This study utilized an Adaptive Boosting (AdaBoost) model to assess whether variations in backscattered signals could effectively classify tumors as benign or malignant. The results revealed a sensitivity of 79%, a specificity of 77%, and an overall accuracy of 78%, suggesting that the AdaBoost model shows promise in differentiating between benign and malignant tumors based on the differences in backscattered signals.

Another notable study using data from a MammoWave device is the one by Papini et al. (2023) [23], who utilized various classification machine learning algorithms, such as SVM, Random Forest (RF), DT, KNN, among others, to classify Non-Healthy (NH) and Healthy (H) breasts in a clinical trial. Data from 697 breasts (354 patients) were analyzed, with 123 categorized as NH and 574 as H. Principal component analysis was applied to retain 95% of data variance and oversampling and undersampling methods ensured at least 40% representation of NH breasts. Each classifier underwent cross-validation with the objective of tuning the parameters. The selection of parameters was based on statistical metrics such as F1-Score, Accuracy, Precision, among others. The two classifiers with the best F1-Scores were tested in four scenarios: raw data for all breasts, raw data from dense breasts only, features from MWI images for all breasts, and features from MWI images for dense breasts only. The results indicated that for raw data of all breasts, SVM achieved the best performance with an accuracy of 88%, sensitivity of 86%, and specificity of 89%. For dense breasts using raw data, RF performed best with an accuracy of 85%, sensitivity of 83%, and specificity of 90%. For all breasts using MWI image features, SVM achieved 83% accuracy, 77% sensitivity, and 85% specificity. For dense breasts using MWI features, SVM attained 87% accuracy, 79% sensitivity, and 88% specificity. These results underscore the potential of these algorithms to differentiate between NH and H breasts.

Shifting focus to applications in neurological diseases, the first study of Alzheimer's Disease (AD) using MWI data was presented by Saied et al. (2021) [26]. They used five machine learning algorithms, namely LR, KNN, DT, NB, and SVM, with the aim of classifying various stages of AD using data obtained from numerical model simulations. Additionally, LDA was utilized in their study. These models were developed in the CST Microwave Studio Suite, with measurements acquired from previous studies [67, 68]. Each stage of AD (normal, mild AD, moderate AD, and severe AD) was represented by changes in the dielectric properties of specific regions and tissues. Simulations were conducted using six antennas, and the signals from each antenna were aggregated into a single dataset for each simulation case. This dataset was then divided into a training set, comprising approximately 78% of the complete simulation cases, and a validation set, which accounted for the remaining 22%. The results of the study indicated that LR achieved the highest estimated average accuracy score, reaching 98.97%, while LDA exhibited an estimated average accuracy of 95.56%. However, the other algorithms displayed accuracy levels below 80%.

Complementing this work, another study in the field of neurological diseases was conducted by Ullah et al. in 2023 [60], where the authors proposed a machine learning-based classification method using KNN, RF, and a DT: Classification And Regression Trees (CART) for the early diagnosis of acute neurological conditions, focusing on AD. The data used in this study were obtained through simulations performed on realistic numerical brain phantoms using the CST Studio Suite to capture the scattered signals. These head models comprised various tissue layers, including the skull, skin, blood, white matter, and gray matter. The models represented three states of AD: normal, mild AD, and severe AD. After data preprocessing, the authors obtained a dataset of 3600 samples. The results of this study yielded accuracies of 79.0%, 83.2%, and 81.0% for the KNN, RF, and CART algorithms, respectively. The authors concluded that the proposed machine learning-based classification method, coupled with MWI algorithms, could potentially be used to monitor AD at its early stages. They intend to evaluate the proposed method on more realistic data obtained from AD patients in future studies. Comparing these results with those presented in [26], it is evident that the latter achieved superior performance. Specifically, the best-performing model in that study was LR with an accuracy of 98.97%.

Finally, turning to applications in brain cancer, Hossain et al. (2023) [24] introduced an eight-layered lightweight classifier named Microwave Brain Image Network (MBINet), employing a Self-Organized Operational Neural Network (Self-ONN) to classify microwave images into six classes: Non-tumor (NT), Benign tumor (BT), Malignant tumor (MT), Double Benign tumor (DBT), Double Malignant tumor (DMT), and one Benign and one Malignant tumor (BMT). The dataset utilized comprised a total of 1,320 microwave images, including 300 NT, 215 BT, 215 MT, 200 DBT, 200 DMT, and 190 BMT samples. These images were obtained via a sensor-based microwave brain imaging system and a fabricated tissue-mimicking brain phantom model, reflecting actual brain properties, as documented in prior research by Hossain et al. (2022) [69]. Subsequently, the authors applied image resizing and normalization techniques for image preprocessing, resulting in 13,200 microwave images. A five-fold cross-validation technique was employed for training, validation, and testing, with 80% of the total images allocated for training and 20% for testing, repeated across five folds. Additionally, 20% of the training dataset was utilized for validation to mitigate overfitting. The proposed approach achieved an accuracy of 96.97%. The MBINet model was compared with four Self-ONN, two vanilla CNNs, ResNet50, ResNet101, and DenseNet201 pre-trained models, demonstrating superior classification performance.

Table 3.5: Summary table on the equipment used for signal acquisition (Target Zones: Brain and Breast). Label “T" represents transmitters, while “R" represents receivers. When these labels are absent, the respective study did not distinguish between transmitters and receivers. The (*) indicates that the nomenclature used in this paper differs from that employed by the authors of the paper.

Reference	**No. of Antenna Positions**	**Monostatic vs Multistatic**	**Moving vs Fixed**	**Frequencies**
Zhu et al. [57]	16	Multistatic	Fixed	0.7 – 2 GHz
Pokorny et al. [59]	10	Multistatic	Fixed	1 GHz
Conceição et al. [56]	144	Monostatic	Moving	1 – 6 GHz
Fasoula et al. [58]	18	Multistatic	Fixed	0.5 – 4 GHz
Janjic et al. [61]	36 T 36 R	Bistatic	Moving	1 – 8 GHz
Papini et al. [23]	15 T 80 R	Bistatic*	Moving	1 – 9 GHz
Saied et al. [26]	6	Monostatic	Fixed	0.2 – 3 GHz
Ullah et al. [60]	6	Multistatic	Fixed	0.2 – 3 GHz
Hossain et al. [24]	50 T 50 R	Multistatic	Moving	1 – 4 GHz

Table 3.5 summarizes key characteristics of the simulations from the studies discussed above. These characteristics include the number of antenna positions and information related to the device configuration used for signal acquisition in each study, presented in the column “Multistatic vs. Monostatic". In addition to the configurations previously introduced, this column includes additional configurations: “Monostatic," which refers to a setup where each antenna, if more than one is present, operates as both transmitter and receiver simultaneously. The (*) in this column indicates that the nomenclature used in this article differs from the terminology adopted by the authors in their study. The table also provides information on whether the antenna array was fixed during signal acquisition and the frequencies used in the acquisition process. There is not a clear pattern for the number of antenna positions depending on the target zone of the application, but the number of antennas is ranging from 6 antennas to 144 antennas positions. The prevalence of the "Multistatic" setup in the “Monostatic vs Multistatic” column can be observed. Unlike the previous section, here it is evident that some studies feature antenna arrays that are not fixed during signal acquisition.

Table 3.6 shares many similarities with the structure of Table 3.3, with the addition of the column “Class Balance", which indicates the distribution of the classes within the dataset for which the respective studies performed classification tasks. It is notable that most studies present a balanced dataset across

Table 3.6: Summary of dataset characteristics and evaluation metrics for Breast and Brain applications.

Ref.	Dataset	Class Balance	Input Sample	Metric
Zhu et al. [57]	25,600 (Time Series Signals)	50/50 (ICH/IS)	Time Series Signals	Acc, Sen, Spe: 94%, 91%, 96%
Pokorny et al. [59]	NA	NA	Feature Vector	Acc: 68.7%
Conceição et al. [56]	26 (TM)	13/13 (Ben/Mal)	Feature Vector	Acc: 96.2%, 92.3%
Fasoula et al. [58]	24 (Patients)	11/8/5 (BPBC/UC/BPBBL)	Solidity, Correlation, Busyness	CL: 11.50%
Janjic et al. [61]	113 (Breast Exams)	70/43 (Ben/Mal)	S_{ii} Matrix	Acc, Sen, Spe: 78%, 79%, 77%
Papini et al. [23]	697 (Breast Exams)	574/123 (H/NH)	Signals/Feature Vector	Acc, Sen, Spe: 88%, 86%, 89% 83%, 77%, 85%
Saied et al. [26]	36 (Simulation Cases)	9/9/9/9 (H/MiAD/MoAD/SAD)	Signal	Acc: 98.97%
Ullah et al. [60]	3,600 (Signals)	1,200/1,200/1,200 (H/MiAD/SAD)	Signal	Acc: 83.2%
Hossain et al. [24]	13,677 (Images)	3,108/2,228/2,228/ 2,072/2,072/1,969 (H/BT/MT/ BBT/MMT/BMT)	Images	Acc, Rec, Spe, Prec, F1: 96.97%, 96.85%, 97.95%, 96.93%, 96.83%

Accuracy (Acc), Sensitivity (Sen), Specificity (Spe), Precision (Prec), Recall (Rec), F1-Score (F1), Classification Loss (CL), Not Available (NA), Healthy (H), Not Healthy (NH), tumor Models (TM), Benign (Ben), Malignant (Mal), Biopsy-Proven Breast Cancer (BPBC), Unaspirated Cysts (UC), Biopsy-Proven Benign Breast Lesions (BPBBL), Intracerebral Hemorrhagic Stroke (ICH), Ischemic Stroke (IS), Mild/Moderate/Severe Alzheimer's Disease (MiAD/MoAD/SAD), Brain tumor (BT), Malignant tumor (MT), Benign Brain tumor (BBT), Malignant Brain tumor (MMT), Benign Malignant tumor (BMT), S_{ii} Matrix

the classes they aim to classify. Regarding the evaluation metrics used, accuracy is the most common choice. However, despite the datasets being balanced, accuracy may not always be the best metric to evaluate the models, particularly in studies distinguishing H from NH cases. If the main objective of the study is to detect cancer, the focus should be on reducing the false negative rate, which accuracy alone does not adequately reflect. Considering this, it is always relevant to use more than one metric to assess the classifiers comprehensively.

3.3 Discussion

The review conducted in this chapter underscores the potential of combining MWI and ML, highlighting the variety of applications this integration offers. These range from breast disease detection to strokes, subdural hematomas, leg imaging, and brain diseases such as AD. MWI provides less invasive, faster, and more comfortable options for patients.

Over time, there has been an evident increase on the sophistication of the algorithms used. Early studies relied on classical ML models such as SVM and KNN. However, more recent studies have employed DL algorithms, such as CNNs, as demonstrated by Mojabi et al. (2021) [40]. In addition, innovative approaches, such as the one presented by Hossain et al. (2023) [24], show promising results.

Despite these advances, limitations persist. For example, although there has been an increase in the variability and complexity of the datasets used, there remains a need for continued efforts to expand these characteristics, as well as the size of the datasets. One possible improvement would be the public availability of datasets used in various studies. This practice has proven beneficial in other fields, as replicability is a critical feature in ML research. Without dataset availability, it becomes impossible to replicate results, which poses a significant barrier to progress in this field.

Another important aspect observed in this review is the frequent use of simulated data, particularly in reconstruction studies, where most works rely on numerical simulations rather than experimental or clinical data. While simulations are essential for initial methodological development, it is critical to validate such approaches using experimental phantoms and/or clinical datasets to ensure their applicability in real-world scenarios. Moreover, simulations, especially in two-dimensional settings, may not fully capture the complexities of real data, such as those present in three-dimensional environments with more pronounced multipath effects. These discrepancies highlight the need for caution when interpreting results from simulated data and emphasize the importance of progressive validation stages in future research.

Even with these limitations, some studies have integrated MWI systems like Wavelia, as shown in Fasoula et al. (2021) [58] presenting encouraging results. Such cases emphasize the importance of confirming results obtained with phantoms and simulations through *in vivo* studies, highlighting the need for a collective effort in preclinical and clinical research.

Another key takeaway from this review is the importance of using appropriate metrics for the classification tasks being performed. Although the most reported metric in the reviewed studies, accuracy may not always be the best measure to evaluate models. For instance, in classification tasks distinguishing between H and NH cases, minimizing false negatives is often more critical than minimizing false positives. In such cases, sensitivity becomes more significant than accuracy. It is also important to note that, in some cases, a single metric may not suffice to evaluate how effectively the model performs the task. In such instances, employing a set of complementary metrics is recommended to provide a more comprehensive assessment of the model's overall performance. In addition, evaluating model generalization would benefit from the regular use of cross-validation, along with reporting the average and standard deviation of the evaluation metrics, as this would illustrate the model's behavior when retrained on different subsets of data.

In tasks that classify malignant versus benign conditions (e.g., in brain or breast cancer), when the dataset is balanced, accuracy, despite providing useful information about model performance, should be complemented with additional metrics such as sensitivity and specificity to offer a more complete evaluation. Furthermore, accuracy can be misleading in cases where the algorithm is trained on imbalanced datasets. In such scenarios, the resulting model might be “lazy," achieving high accuracy by biasing toward the dominant class. However, such a model could perform poorly in real-world applications. Metrics like Precision and Recall can be better suited for these cases. Therefore, the F1-score is often a more appropriate metric in these situations, as it accounts for both precision and recall, offering a more balanced assessment of model performance.

Finally, it is important to note that the analysis of the reviewed articles is subject to interpretation errors. Additionally, challenges were encountered in obtaining complete information from the articles, as seen in the tables summarizing dataset details, particularly in the “Dataset Size" and “Type of Sample".

3.4 Conclusion

In summary, this chapter highlighted the advancements and challenges associated with the use of MWI and ML in various medical applications. The results presented in the literature suggest great potential for these techniques but also emphasize the need for additional efforts to overcome limitations related to model complexity and realism, the diversity of tested data, validation with *in vivo* data, and the selection of appropriate metrics depending on the task at hand. It is crucial to acknowledge that while this field has significant potential in medical applications, there is still room for improvement. Collaboration within the scientific community remains vital to achieve the shared goal of improving diagnostic conditions, aiding diagnoses as early as possible.

Chapter 4

Materials and Methods

This chapter describes the methodology employed in this dissertation, outlining the procedure followed step-by-step . It begins with data acquisition, detailing the sources and methods used to obtain the datasets. Next, it covers data preprocessing, including feature extraction, Window Trimming (WT), and normalization, essential steps to prepare the data for machine learning applications. The chapter then addresses the optimization of machine learning models, discussing the strategies used to enhance their performance. Finally, it presents the application of these models in the image reconstruction task, demonstrating their role in improving microwave imaging for breast tumor detection.

4.1 Data Acquisition

The data used in this dissertation were acquired by Ana Catarina Pelicano, a PhD student from my research group, through the simulation of multiple scenarios. The data for these simulations were based on breast tissue models segmented from MRI scans collected at Hospital da Luz-Lisboa between 25/09/2019 and 29/06/2022, under clinical study references CES/44/2019/ME and CES/34/2020/ME [70]. The MRI scans were obtained using a 3.0T MAGNETOM Vida clinical magnetic resonance scanner (Siemens Healthineers, Erlangen, Germany) with an 18-channel dedicated breast coil.

A total of 83 tumor models, segmented from MRI scans of 55 patients, were used in the simulations. Of these, 46 were benign, and 37 were malignant. The breast phantom was composed of healthy tissues extracted from the MRI of a single patient. Tumors were artificially placed in the breast model to create a combined model of breast and tumor tissue. Each tumor was simulated in three different breast phantoms: one composed solely of homogeneous adipose tissue, another with homogeneous fibroglandular and adipose tissues, and a third with heterogeneous fibroglandular and adipose tissues. To enhance data variability, tumors were positioned in off-center locations within each phantom.

The simulations were conducted using CASE, a 3D Finite Difference Time-Domain simulation tool implemented in MATLAB 2020b. The setup included 10 antennas arranged in a circular pattern around the breast. A modulated Gaussian pulse at 3 GHz was used, and a Perfectly Matched Layer was applied to minimize boundary reflections [71].

In total, 249 simulations were performed, forming the dataset used in this dissertation. These simulations were divided into three datasets of 83 simulations each, based on the phantom composition:

- **D1**: Homogeneous adipose tissue phantom
- **D2**: Homogeneous fibroglandular and adipose tissue phantom

- **D3**: Heterogeneous fibroglandular and adipose tissue phantom

4.2 Data Processing

After data acquisition, it is necessary to process the signals in a way that enables training ML models to distinguish voxels containing tumor tissue (class 1) from those containing healthy tissue (class 0). To achieve this, different Feature Engineering (FE) pipelines were created as part of the data processing stage.Each pipeline specifies the type of feature extraction used, whether normalization was applied (further explained in Section 4.3), or whether WT was performed. In this dissertation, four types of feature extraction methods were considered: detail coefficients from the wavelet transform (**DWTD**), approximation coefficients from the wavelet transform (**DWTA**), approximation coefficients from the second level of wavelet decomposition (**DWTA2**), and twenty-four signal-derived tailored features (**F24**). Wavelet coefficients were chosen for their ability to capture both high- and low-frequency components of the signal. High-frequency components (**DWTD**) are often associated with abrupt changes caused by abnormalities, while low-frequency components (**DWTA** and **DWTA2**) preserve the general shape and trends of the signal, allowing the detection of broader structural differences. The second level of decomposition (**DWTA2**) was included to investigate whether an additional level of decomposition could improve classification performance. The **F24** set captures various aspects of the signal's shape and behavior. This diverse set increases the likelihood of capturing subtle but meaningful differences in signal morphology between healthy and tumor tissues. The **F24** features were previously used in work conducted by my research group [72]. A detailed list of the features included in **F24** is provided below.

1. **Maximum Amplitude** (`maxAmp`):
2. **Location of Maximum Amplitude** (`maxAmpInd`)
3. **Minimum Amplitude** (`minAmp`)
4. **Location of Minimum Amplitude** (`minAmpInd`)
5. **Variance** (`variance`)
6. **Root-Mean-Squared Error** (`error`)
7. **Number of Positive Peaks** (`numMaxs`)
8. **Number of Negative Peaks** (`numMins`)
9. **Average Amplitude of Positive Peaks** (`avg_maxs`)
10. **Average Amplitude of Negative Peaks** (`avg_mins`)
11. **Average Full-Width Half-Maximum (FWHM) of Positive Peaks** (`fwhm_max`)
12. **Average FWHM of Negative Peaks** (`fwhm_min`)
13. **Average Distance Between Positive Peaks** (`avg_distance_max`)
14. **Average Distance Between Negative Peaks** (`avg_distance_min`)
15. **Number of Zero Crossings** (`phaseCounter`)

16. **Area Under the Curve** (`auc`)
17. **Absolute Area Under the Curve** (`aucAbs`)
18. **Positive Percentage Area** (`aucPos`)
19. **Negative Percentage Area** (`aucNeg`)
20. **Average Autocorrelation** (`avg_autocorr`)
21. **Number of Autocorrelation Peaks** (`num_peaks_autocorr`)
22. **Average Amplitude of Autocorrelation Peaks** (`avg_pks_autocorr`)
23. **Average FWHM of Autocorrelation Peaks** (`fwhm_autocorr`)
24. **Average Distance Between Autocorrelation Peaks** (`avg_distance_autocorr`)

For the application of FE (Figure 4.1), only signals in which the same antenna acted as both transmitter and receiver (i.e. monostatic) were used. Initially, the signal obtained from the free space (i.e. only antennas) configuration was subtracted. However, due to the poor quality of the resulting reconstructions, an ideal scenario was considered where the signals obtained from simulations of the same phantom without the presence of a tumor were subtracted. This substitution places the reconstructions presented in this dissertation in an ideal scenario that does not reflect real-world conditions—primarily because, in practice, it is not feasible to acquire tumor-free signals from actual patients. Nonetheless, this approach was adopted to facilitate the task for the ML models implemented in this work by removing background contributions common to both healthy and pathological cases—such as reflections from the skin and structural components of the phantom—enhancing the contrast associated with the presence of a tumor, and evaluate the feasibility of this approach. After subtraction, feature extraction required iterating through the signal. For each voxel, the distances to the antennas were calculated and used to identify the corresponding signal portions for each antenna.

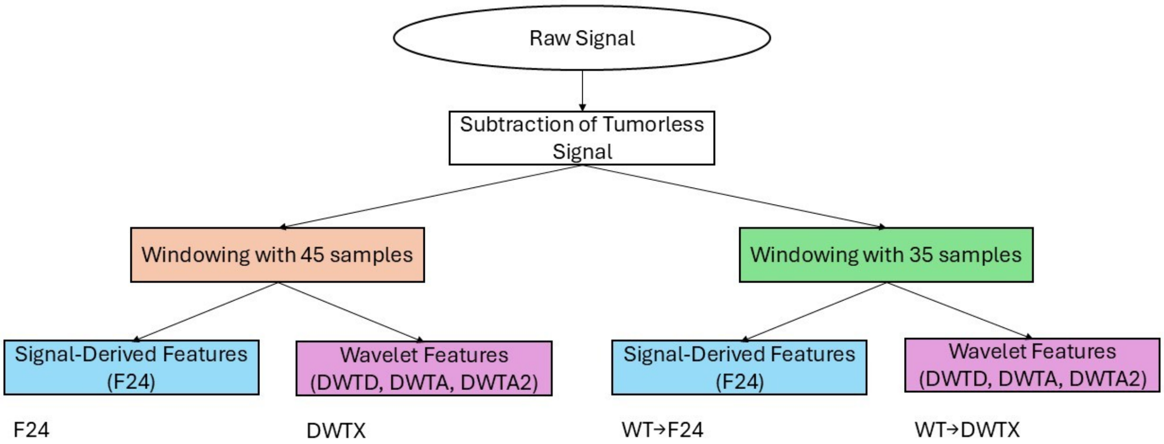


Figure 4.1: Overview of feature extraction pipelines using two windowing strategies (45 and 35 samples) applied to the baseline-subtracted signal. Extracted features include signal-derived (F24) and wavelet-based descriptors (DWTD, DWTA, DWTA2), resulting in four distinct feature sets for model input.

For each signal portion, features were extracted at each time sample using a window of 45 time samples—22 before and 22 after the central time sample, which corresponded to the voxel localization. When wavelet transform (WT) was applied as part of the feature extraction (FE) process, a reduced window of 35 time samples (17 before and 17 after) was used (Figure 4.2). This adjustment aimed to assess whether a shorter window length would be less prone to capturing noisy segments. Feature sets using trimmed windows are denoted with “WT”. Since the phantom model provided ground truth tissue labels for each voxel, it was possible to determine whether a given voxel corresponded to healthy tissue (class 0) or tumorous tissue (class 1). To maintain a balanced dataset with equal representation of both classes, and to ensure consistency in data extracted across simulations, a specific selection process was applied. First, the number of class 1 samples extracted from each simulation was recorded, as the number of class 1 data points was substantially lower than that of class 0 in the simulations. After processing all simulations, the smallest number of available class 1 samples (n) from a single simulation was identified. Then, n samples of each class were randomly selected from each simulation. Finally, the dataset was split into training and test sets, with the test set comprising approximately 20% of the total dataset size.

The naming convention for each FE includes the dataset it was applied to, the type of features extracted from the signal, and, where applicable, indicators for normalization and/or WT. For example, an FE extracting F24 features, with WT and normalization applied to D1, would be named “D1 WT→F24→Norm.” Normalization is performed during model training.

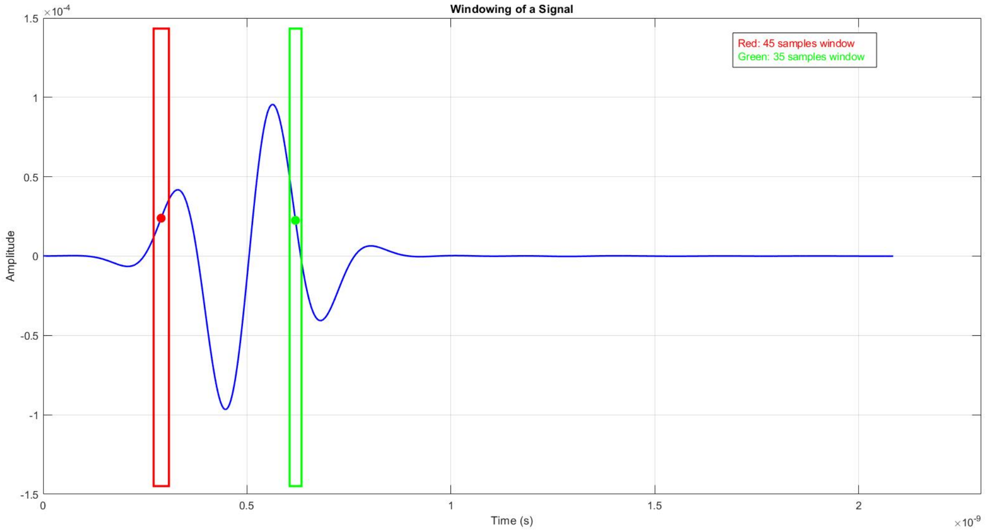


Figure 4.2: Example of a time-domain signal used in this study, illustrating the application of windowing. Two temporal windows are highlighted: a red window comprising 45 samples and a green window comprising 35 samples. The blue curve represents the received signal, and the colored dots indicate the center points of each window. These windows are used to extract time-localized features for subsequent analysis.

4.3 Model Optimization

After data processing, the next step in the framework was model optimization, which involved hyperparameter tuning and model training. As shown in Figure 4.3, this stage encompasses the hyperparameter tuning and training process for the models. As mentioned in Chapter 1, the algorithms used in this study were SVM, KNN, and DT.

During model optimization, cross-validation was employed. Since the training set consisted of samples obtained from acquisitions of 66 different tumors, 11 folds were used to ensure that samples from the same tumor were included in the same fold. Each fold therefore contained samples from six tumors. In FE pipelines that included normalization, Min-Max scaling was applied to the training set folds at

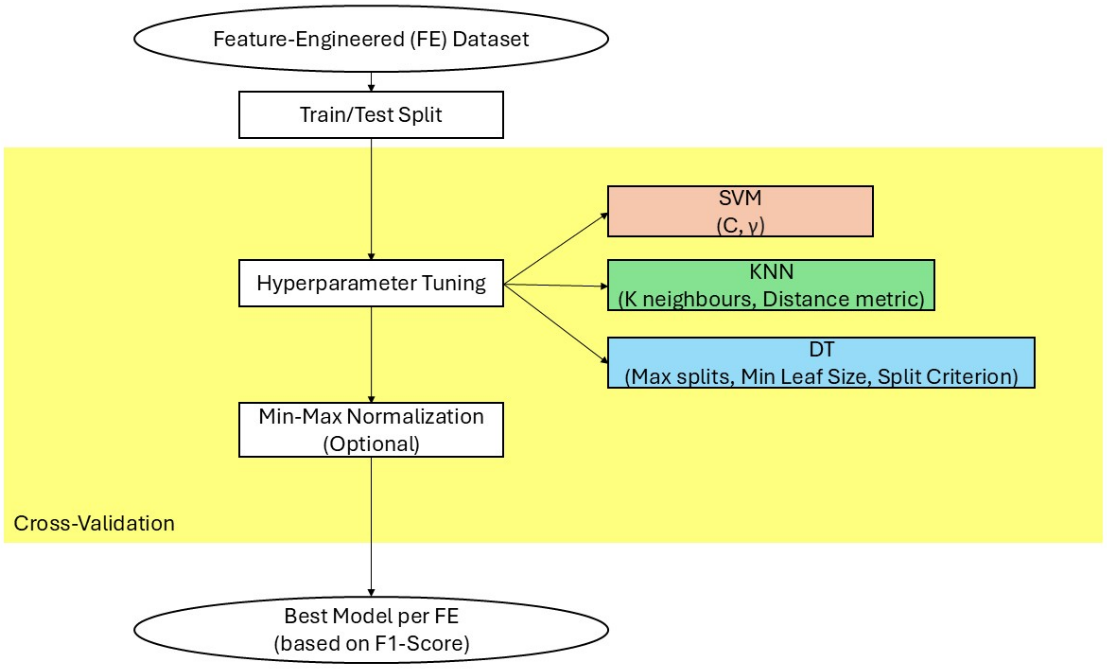


Figure 4.3: Machine learning pipeline for classifying a feature-engineered dataset. After splitting the data into training and testing sets, hyperparameter tuning is conducted for three classifiers—SVM, KNN, and DT. Optionally, Min-Max Normalization is applied. Cross-validation is employed to select the best model based on the F1-score.

each iteration of the cross-validation process. The same scaling parameters were then used for the corresponding validation fold. Min-Max normalization follows Equation 4.1, where X represents the original feature values, X_{min} and X_{max} are the minimum and maximum values of the feature in the training set, and X' is the normalized value within the range $[0,1]$. This ensures consistency in data distribution between training and validation sets.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{4.1}$$

The hyperparameters optimized for each model were as follows:

- **SVM:** The regularization parameter (C) and gamma (γ) were varied over the range $[0.01, 0.1, 1, 10, 100]$, with all SVM models using the Radial Basis Function (RBF) kernel.
- **KNN:** The number of neighbors (k) was varied from 1 to 19 in steps of 2, and different distance metrics were tested, including Euclidean, Minkowski, Chebyshev, and Cityblock (Manhattan) distances.
- **DT:** Three parameters were optimized:
  - Maximum number of splits, ranging from 2 to 10 in steps of 2.
  - Minimum leaf size, ranging from 5 to 50 in steps of 5.
  - Split decision criterion, tested using both Gini's Diversity Index (gdi) and entropy.

All optimizations were performed using the F1-score as the evaluation metric, ensuring a balanced trade-off between false positives and false negatives. Once trained with the optimal parameters, models were evaluated on the test set using Sensitivity, Specificity, Accuracy, Precision, and F1-score.

The best-performing model for each algorithm within each dataset optimized in terms of F1-score, yielding a total of nine trained models for microwave image reconstruction.

4.4 Image Reconstruction

Similar to the approach described in “Data Processing” (Section 4.2), image reconstruction from microwave signals was performed voxel by voxel for the ML-based approach, and using Equation 2.1 for the Delay-And-Sum (DAS) case, both after subtracting the signal acquired without the tumor from the signal with the tumor. By calculating the distance between each antenna and each voxel, it is possible to determine the signal delay and, consequently, the portion of the signal corresponding to that voxel. In ML-based reconstruction the features are directly fed into the model, which predicts the likelihood of each pixel corresponding to tumor tissue. Signals from 10 antennas around the breast are used. The final value for each pixel is obtained by averaging the predicted values from all antennas. After reconstructing the entire image, two post-processing steps are applied. First, a threshold is set, i.e., only pixels with a value greater than 0.9 are classified as class 1 (tumor tissue), while the rest are classified as class 0 (healthy tissue). Second, clusters smaller than 30% of the size of the largest cluster obtained in the first step are removed, to eliminate any artifacts obtained during reconstruction. This pipeline is depicted in Figure 4.4. Both thresholds used in these steps were empirically defined, after testing different values.

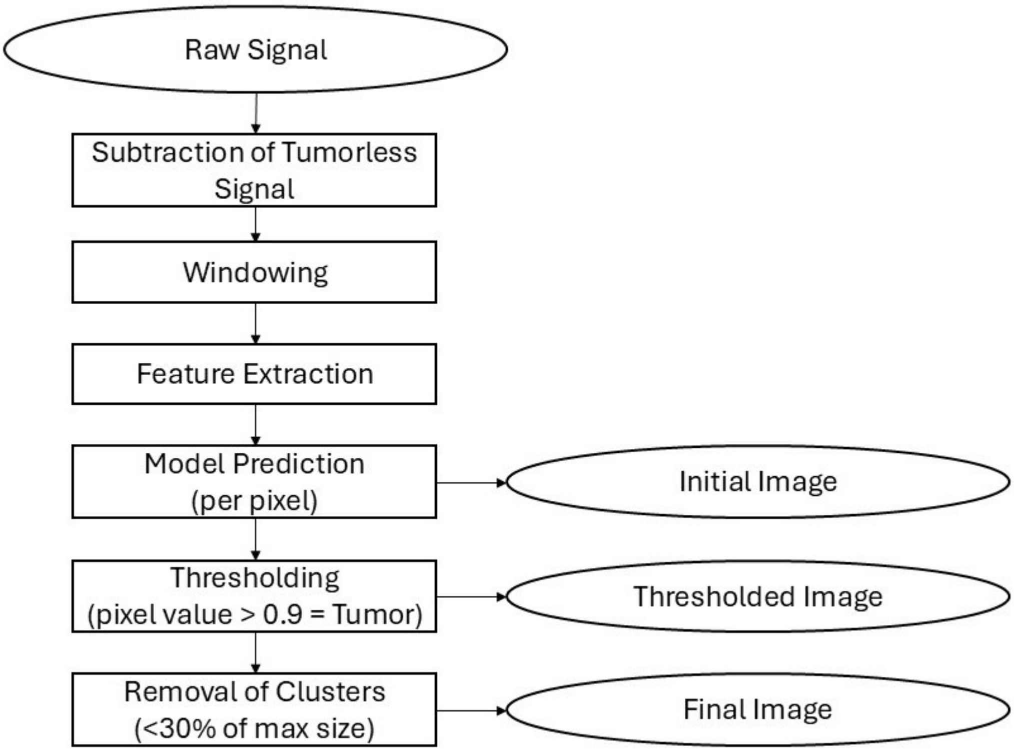


Figure 4.4: Sequential pipeline for pixel-wise tumor detection from raw signal data, incorporating tumorless signal subtraction, windowing, feature extraction, model prediction, thresholding (tumor defined as pixel > 0.9), and removal of small clusters to generate the final segmented image.

Since a single metric cannot objectively evaluate the performance of models trained for image reconstruction, the reconstructions were assessed using a group of metrics that express different characteristics, ensuring a comprehensive assessment of the image accuracy and reliability. The selected metrics were IoU, Dice Similarity Coefficient (DSC), Relative Volume Error (RVE), and Matthews Correlation Coefficient (MCC), considering the tumor the ground truth mask. Together, these metrics provide a comprehensive understanding of the performance of the model, capturing spatial accuracy, volumetric precision, and overall reconstruction quality. Particularly, the IoU and DSC metrics are used to evaluate spatial accuracy, offering insights into how well the reconstructed regions align with the ground truth in terms of shape and location. In contrast, RVE measures volumetric precision by indicating the extent to which the volumes of the reconstructed image and the ground truth differ. Lastly, the MCC accounts for true and false positives and negatives, providing a balanced view of the reconstruction quality, particularly in cases of class imbalance, such as the one considered in this dissertation.

Chapter 5

Results and Discussion

This chapter comprises the dataset analysis, specifically focusing on the data after the Feature Engineering (FE) pipelines, including boxplots, histograms, and correlation matrices. It also presents the results of model training and discussion, detailing the performance of each model combined with the different FE approaches. The best combination of hyperparameters for each FE is identified, along with the corresponding F1-score and other evaluation metrics. Finally, the chapter presents the reconstructions obtained using the best-performing models for each algorithm and dataset. These reconstructions are analyzed using four of the 17 simulations from the test group and are evaluated using the average values of multiple evaluation metrics computed across the 17 simulations. A visual comparison is also made between the Delay-And-Sum (DAS) reconstruction and the best reconstruction model for each dataset, using one representative simulation from the test set.

5.1 Dataset Analysis

In total, 249 simulations were divided into three datasets based on the composition of the phantoms, resulting in 83 simulations per dataset. D1 corresponds to a homogeneous adipose tissue phantom, D2 to a homogeneous fibroglandular and adipose tissue phantom, and D3 to a heterogeneous fibroglandular and adipose tissue phantom. From each simulation, 180 observations were extracted. Therefore, each dataset contained 14,940 observations after FE. These were split into training and testing sets, with approximately 80% (11,880 observations) used for training and 20% (3,060 observations) for testing. Observations from the same simulation were kept within the same subset to avoid data leakage.

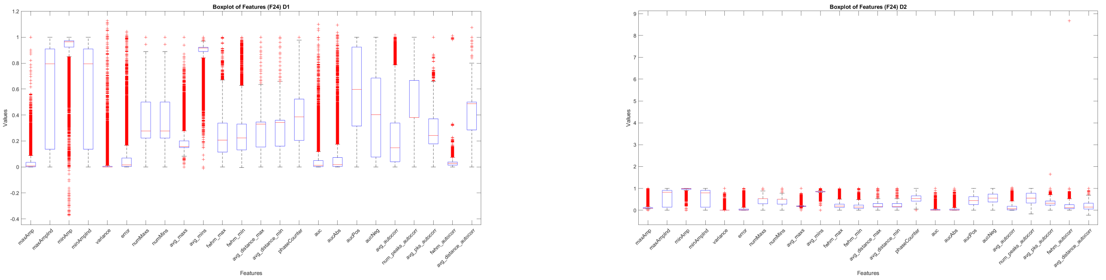
In this section, boxplots are presented for the entire dataset, normalized between 0 and 1 using Min-Max normalization (Equation 4.1) based on the training set parameters, as the magnitude of the features ranged from $10e^{-19}$ to 10. Additionally, histograms of the four features/coefficients with the highest Standard Deviation (Std) values and the correlation matrices will be shown. The histograms and correlation matrices will be presented separately for $y = 0$ and $y = 1$ for each feature extraction method, with y being the label for each sample. Only the FE pipelines without the Window Trimming (WT) step were considered for this analysis.

Due to the similarity of the different elements among the various types of wavelet coefficients, the FE pipelines with wavelet coefficients were grouped into a single subsection. One example will be presented for FE pipelines, except in cases where further illustration is deemed appropriate. Therefore, the analysis is divided into two subsections, one for each type of FE: F24 and wavelet coefficients.

5.1.1 F24

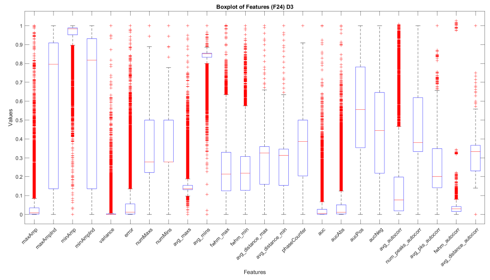
The boxplots obtained from the data processed by FE with F24 processing were similar across the three different datasets, with D2 showing slight differences due to an outlier in the feature `fwhm_autocorr` (Figure 5.1 illustrates the boxplots for D1, D2, and D3 using F24 as their feature set).

By analyzing the boxplots from D1 and D3, shown in Figures 5.1a and 5.1c respectively, it is evident that many features yielded several outliers, represented by red markers. Additionally, certain features, such as `maxAmp` and `minAmp`, exhibit longer whiskers—upper whiskers for `maxAmp` and lower whiskers for `minAmp`—indicating right- and left-skewed distributions, respectively. The boxplots for D2 (Figure 5.1b) show similar tendencies; however, there is an outlier with a value nearly nine times higher in the test set than the maximum value found in the training set.



(a) F24 D1

(b) F24 D2



(c) F24 D3

Figure 5.1: Boxplots of F24 features from the datasets. Each box represents the distribution of values for a specific feature. The red lines indicate the median, the blue boxes represent the interquartile range (IQR), and the red crosses denote outliers. The x-axis labels correspond to features names.

The distribution of features such as `maxAmpInd`, `minAmpInd`, `aucPos`, and `aucNeg`, displays a wider Interquartile Range (IQR), suggesting higher variance. This observation aligns with the nature of the features: the tumor location varies between simulations, consequently affecting amplitude of the reflection signals, and for the `auc`-type features, it may result from the variance in tumor size. In contrast, some features, such as `fwhm_autocorr`, exhibit very small IQRs, indicating lower dispersion in the data. Furthermore, a few features, such as `fwhm_min` and `numMaxs`, appear more symmetric, with their medians centered within the boxes.

Figure 5.2 shows the distributions of the features `maxAmpInd` and `minAmpInd`, which are among the four features with the highest Std across all three datasets. The distributions are notably similar across datasets, each exhibiting two distinct density clusters: one at the lower end and another at the higher end of the feature range. When $y = 0$, the densities of both clusters are relatively balanced. However, for $y = 1$, the higher-end cluster tends to dominate, suggesting that these features may have greater discriminative power in distinguishing between classes. This pattern indicates their potential relevance for classification models tasked with differentiating between $y = 0$ and $y = 1$.

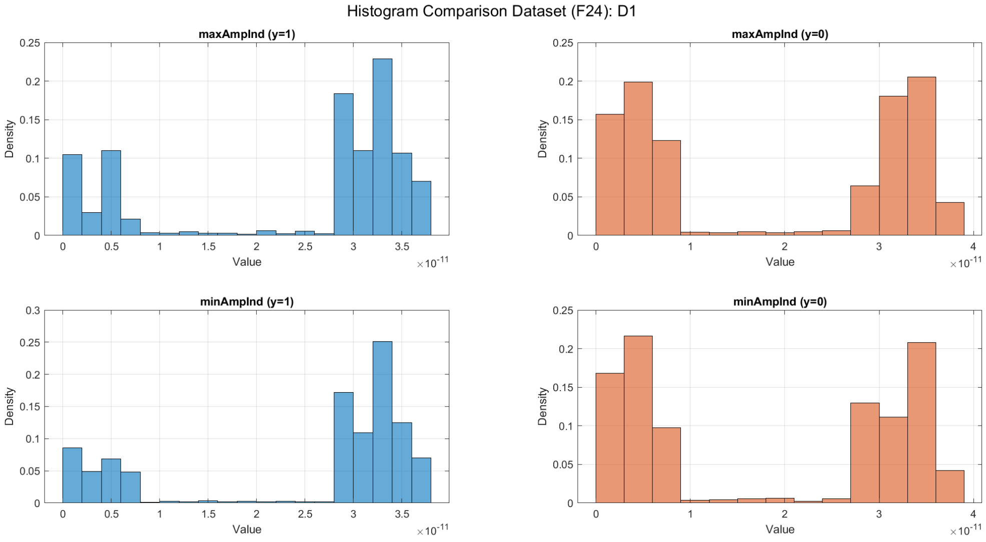
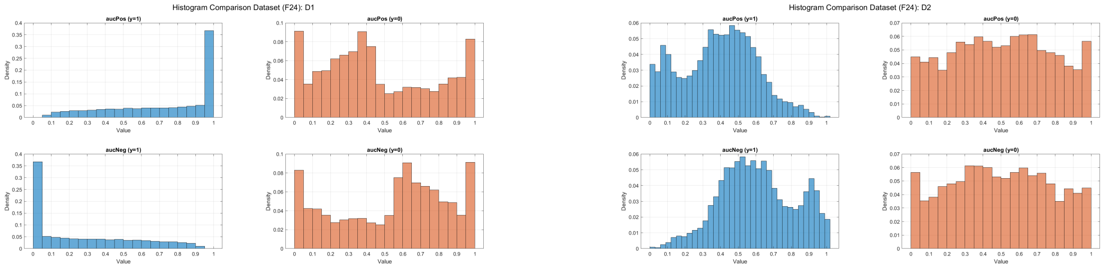
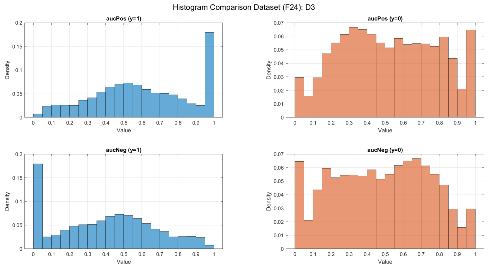


Figure 5.2: Histograms of the features: Location of Maximum Amplitude (top row) and Location of Minimum Amplitude (bottom row). The left column represents data for $y = 1$, while the right column represents data for $y = 0$, from dataset D1.



(a) F24 D1

(b) F24 D2



(c) F24 D3

Figure 5.3: Histograms of the features: Positive Percentage Area (top row) and Negative Percentage Area (bottom row). The left column represents data for $y = 1$, while the right column represents data for $y = 0$.

Figure 5.3 shows the distribution of the features `aucPos` and `aucNeg` across the three datasets. In both D1 and D3, samples with $y = 1$ exhibit pronounced density at the extremes, i.e, `aucPos` values are concentrated near 1, while `aucNeg` values are concentrated near 0. In contrast, the $y = 0$ samples tend to display a more uniform distribution across the value range. This separation between classes suggests that these features may possess strong discriminative power in D1 and D3. For dataset D2, however, the distributions are less distinct. Both features show a higher density of $y = 0$ samples concentrated around the middle of the value range, while the $y = 1$ samples are more dispersed. This overlap indicates a reduced capacity of `aucPos` and `aucNeg` to distinguish between the classes in D2, suggesting that these features may be less effective in this context, although the reason for this difference remains unclear.

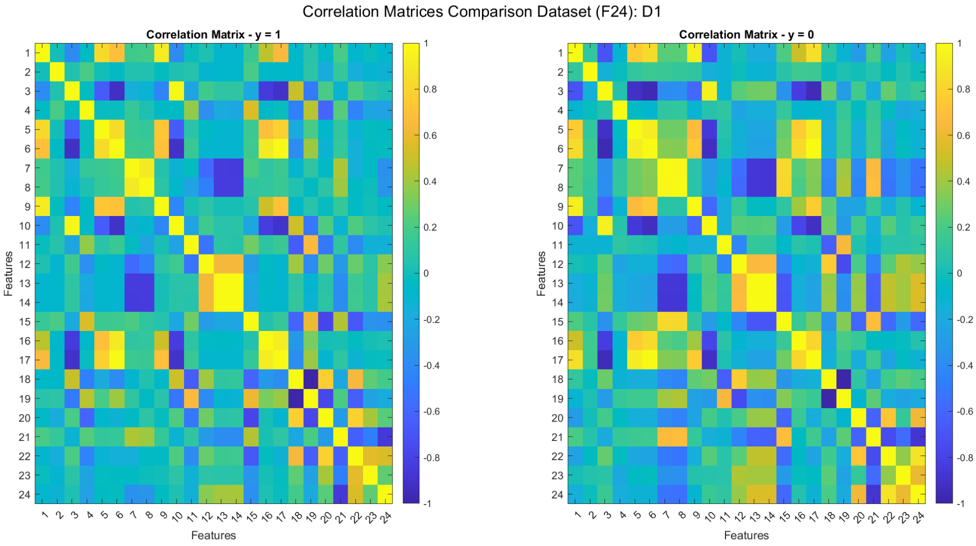


Figure 5.4: Comparison of correlation matrices for F24 applied to D1. The left matrix corresponds to samples with $y = 1$, and the right matrix corresponds to samples with $y = 0$. Color intensity represents the correlation strength between features, ranging from -1 (blue, strong negative correlation) to 1 (yellow, strong positive correlation).

The three datasets exhibit nearly identical correlation matrices, as exemplified in Figure 5.4 for dataset D1. The correlation patterns remain consistent across datasets, with only minor variations in correlation intensity, suggesting that the complexity of the used phantom does not significantly influence the inter-feature correlations. Among the 24 features, both strong positive and negative correlations are observed, which indicates that, even though feature selection was not performed in this dissertation, such an approach could be a promising direction for future work. Nevertheless, most feature relationships are characterized by weak correlations. Furthermore, the correlation matrices show minimal differences between the classes $y = 0$ and $y = 1$, indicating that correlation structure is largely class-independent.

5.1.2 Wavelet Coefficients

The boxplots of the wavelet coefficients are highly similar across the three levels of phantom complexity, displaying narrow IQR values and a high number of outliers both above and below the interquartile range. This pattern is illustrated in Figure 5.5, which presents the boxplots for dataset D1 using DWTD. As all boxplots show a balanced distribution of upper and lower whiskers, there is no clear indication of skewness in the distributions.

The wavelet coefficients with the highest Std vary across different FE methods and even between datasets. However, their distributions are generally similar—centered around zero with a sharp peak. The main differences lie in the skewness and the width of the tails, as illustrated in the examples in Figure 5.6. The differences between the distributions for $y = 0$ and $y = 1$ are not pronounced enough to suggest that feature selection would be beneficial for these FE methods.

The correlation matrices exhibit consistent patterns across all datasets and between classes ($y = 0$ vs. $y = 1$), as illustrated in Figure 5.7a, which compares the correlation matrices for $y = 0$ and $y = 1$ in DWTD features from D1. Strong positive correlations between coefficients are evident in both classes. The differences between the two matrices are minimal, with the $y = 0$ class showing slightly higher correlation values overall. Most relationships reflect positive correlations, while a few in the $y = 1$ matrix

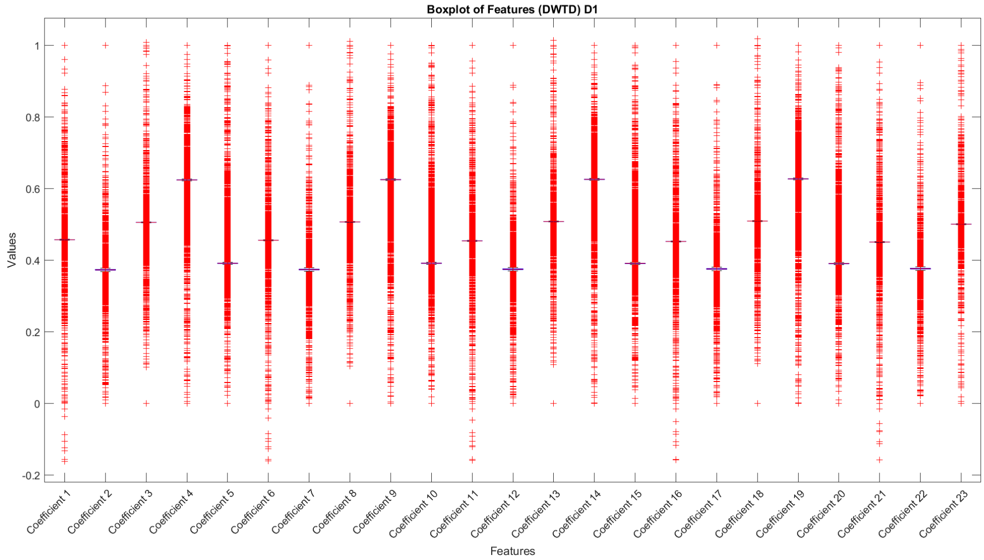
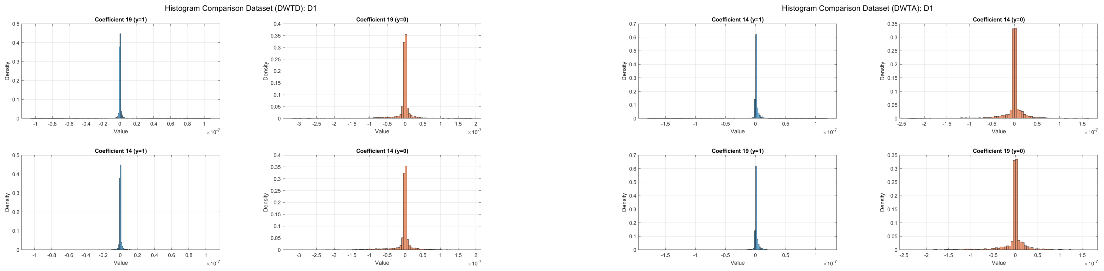
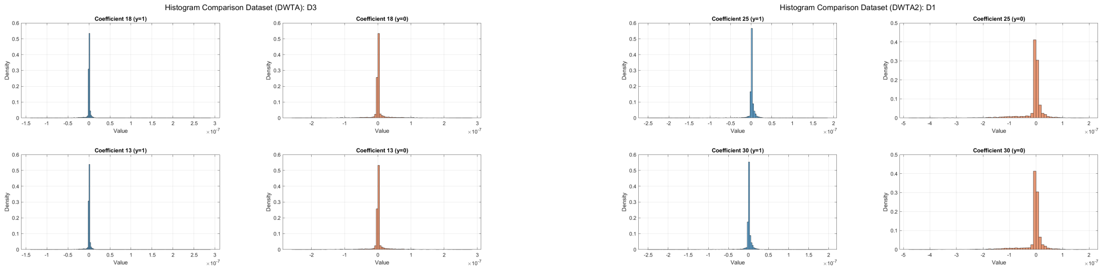


Figure 5.5: Boxplots of DWTD coefficients from D1. Each box represents the distribution of values for a specific feature. The red lines indicate the median, the blue boxes represent the IQR, and the red crosses denote outliers. The x-axis labels correspond to their coefficient number.



(a) Dataset D1 (DWTD): Coefficients 5 and 10

(b) Dataset D1 (DWTA): Coefficients 4 and 9

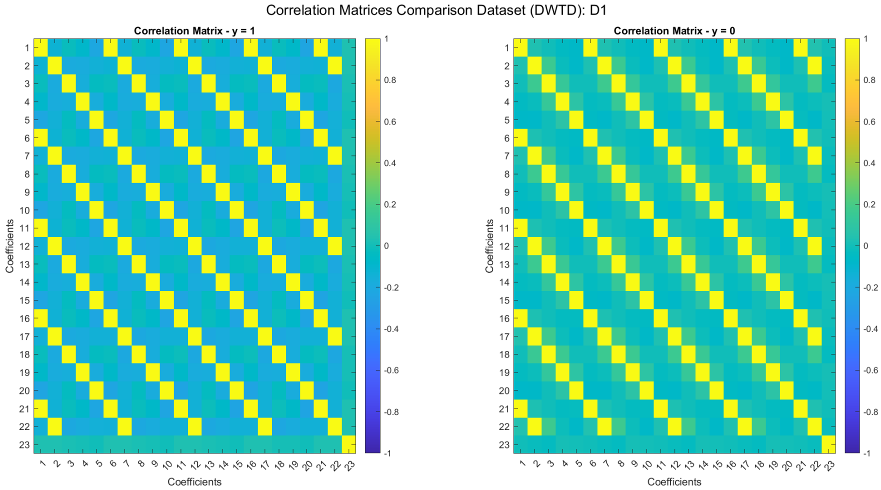


(c) Dataset D3 (DWTA): Coefficients 18 and 13

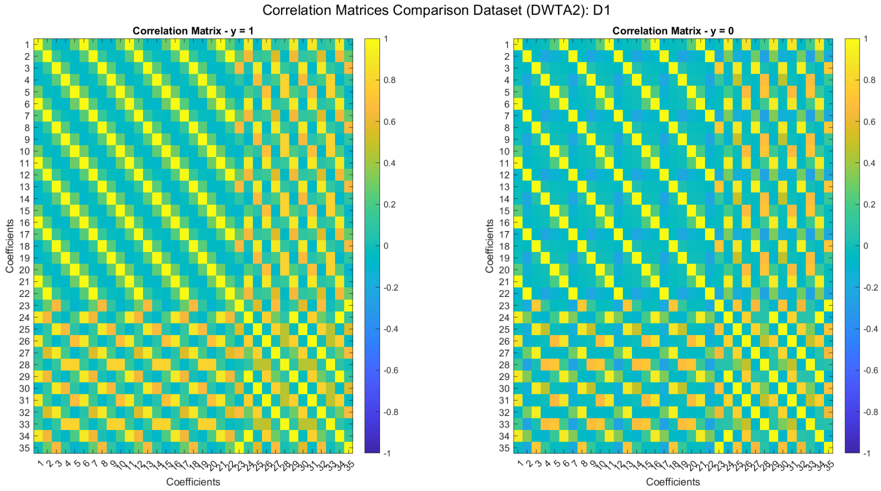
(d) Dataset D1 (DWTA2): Coefficients 25 and 30

Figure 5.6: Histograms of the highest Std values from different datasets and wavelet coefficient as its features. The left column represents data for $y = 1$, while the right column represents data for $y = 0$.

are negative, although these are relatively weak. Since DWTA2 extends DWTA by including a second level of decomposition, its correlation matrix follows the same pattern but contains more coefficients, as shown in Figure 5.7b.



(a) Dataset D1 DWTD



(b) Dataset D1 DWTA2

Figure 5.7: Comparison of correlation matrices for DWTD and applied to D1. The left matrix corresponds to samples with $y = 1$, and the right matrix corresponds to samples with $y = 0$. Color intensity represents the correlation strength between features, ranging from -1 (blue, strong negative correlation) to 1 (yellow, strong positive correlation).

5.2 Model Training

The results of model training are presented in Tables 5.1-5.9 and a Summary Table 5.10. Each table displays the results of models from a single algorithm (K-Nearest Neighbors (KNN), Decision Trees (DT), Support Vector Machine (SVM)) trained with FE applied to a specific dataset. These results include the evaluation metrics: Sensitivity, Specificity, Accuracy, Precision, and F1-Score. The tables also present the parameters obtained during cross-validation, as well as F1-Score, which was used to optimize the models.

The best values for each metric are highlighted in bold, along with the name of the FE whose model achieved the highest F1-Score in the testing phase. F1-Score was the metric used to optimize a model from each algorithm and dataset group.

5.2.1 K-Nearest Neighbors

Table 5.1 presents the classification metrics, the optimal hyperparameters, and the F1-score obtained during training for the KNN models trained with the FE applied to dataset D1. The best results were achieved by two models trained with F24 features, both of which were normalized. The only difference between them was the size of the window applied: the model with a smaller window (**WT→F24→Norm**) achieved the best sensitivity (0.9824), accuracy (0.9827), and F1-score (0.9827), while the model without WT (**F24→Norm**) obtained the best specificity (0.9843) and precision (0.9842). The best F1-score during training was also achieved by the **WT→F24→Norm** model. The optimal hyperparameters for **WT→F24→Norm** were one neighbour and the cityblock (Manhattan) distance metric.

For F24 and DWTD, models that included normalization as part of their FE exhibited slightly better performance than those that did not. Additionally, FE incorporating WT consistently produced models with superior performance compared to their unprocessed counterparts. For DWTA and DWTA2, normalization had either no impact or a slightly negative effect, although window-trimmed models still outperformed the non-trimmed ones. During training, the best-performing distance metric for all FE was the cityblock metric. The number of neighbors varied: five for **DWTA2→Norm**; three for **F24→Norm**, **WT→DWTD**, **DWTD**, and **WT→DWTA2**; and one for the remaining FE configurations. It is noteworthy that models using F24 features achieved the best overall performance, whereas those using DWTD features yielded the lowest results.

Table 5.1: Performance metrics of K-Nearest Neighbors model across different FE of the dataset group D1: Sensitivity, Specificity, Accuracy, Precision, F1-Score, Number of Neighbors, Distance Metric, and F1-Score During Training. WT: Window Trimmed, Norm: Normalized.

K-Nearest Neighbors for D1

	Train			Test				
Feature Engineering	**F1-Score During Training**	**NumNeighbors**	**Distance**	**Sensitivity**	**Specificity**	**Accuracy**	**Precision**	**F1-Score**
F24	0.9527	1	cityblock	0.9510	0.9614	0.9562	0.9610	0.9560
F24→Norm	0.9843	3	cityblock	0.9745	**0.9843**	0.9794	**0.9842**	0.9793
WT→F24	0.9571	1	cityblock	0.9523	0.9739	0.9631	0.9733	0.9627
WT→F24→Norm	**0.9849**	1	cityblock	**0.9824**	0.9830	**0.9827**	0.9830	**0.9827**
DWTD	0.9273	3	cityblock	0.9405	0.9458	0.9431	0.9455	0.9430
DWTD→Norm	0.9280	1	cityblock	0.9425	0.9536	0.9480	0.9531	0.9477
WT→DWTD	0.9307	3	cityblock	0.9471	0.9523	0.9497	0.9520	0.9495
WT→DWTD→Norm_WT	0.9329	1	cityblock	0.9458	0.9595	0.9526	0.9589	0.9523
DWTA	0.9400	1	cityblock	0.9673	0.9536	0.9605	0.9542	0.9607
DWTA→Norm	0.9394	1	cityblock	0.9601	0.9536	0.9569	0.9539	0.9570
WT→DWTA	0.9474	1	cityblock	0.9810	0.9614	0.9712	0.9622	0.9715
WT→DWTA→Norm	0.9447	1	cityblock	0.9745	0.9516	0.9631	0.9527	0.9635
DWTA2	0.9451	1	cityblock	0.9745	0.9510	0.9627	0.9521	0.9632
DWTA2→Norm	0.9420	5	cityblock	0.9621	0.9438	0.9529	0.9448	0.9534
WT→DWTA2	0.9436	3	cityblock	0.9569	0.9647	0.9608	0.9644	0.9606
WT→DWTA2→Norm	0.9404	1	cityblock	0.9608	0.9595	0.9601	0.9595	0.9602

Table 5.2 presents the classification metrics, best hyperparameters, and F1-score values obtained during training for the KNN models trained with the FE applied to dataset D2. Similarly to dataset D1, D2 has more than one FE achieving the best performance. The **DWTA** model achieved the highest specificity (0.9451) and precision (0.9402), while the **WT→F24→Norm** model achieved the highest sensitivity (0.9327). The **F24→Norm** model achieved the highest accuracy (0.9261) and F1-score (0.9261). Consequently, the model trained with **F24→Norm** was selected for reconstruction using the D2-trained KNN models. This model's hyperparameters were three neighbors and the cityblock distance metric. The best F1-score during training was obtained by the **WT→F24→Norm** model, with a value of 0.9149.

For F24 and DWTD, incorporating normalization yielded better results, whereas for DWTA and DWTA2, normalization had little to no impact or a slightly negative effect on performance. Window-trimmed FE produced better results than their non-window-trimmed counterparts for DWTD and

DWTA2, while the opposite was observed for F24 and DWTA. All models used the cityblock distance parameter. Contrary to the D1 models, most D2 models achieved the best performance with three neighbors, with two models performing best with one neighbor and three models with five neighbors. F24 remained the top-performing FE pipeline, while DWTD continued to yield the lowest performance across the classification metrics.

Table 5.2: Performance metrics of K-Nearest Neighbors model across different FE of the dataset group D2: Sensitivity, Specificity, Accuracy, Precision, F1-Score, Number of Neighbors, Distance Metric, and F1-Score During Training. WT: Window Trimmed, Norm: Normalized.

K-Nearest Neighbors for D2

	Train			**Test**				
Feature Engineering	**F1-Score During Training**	**NumNeighbors**	**Distance**	**Sensitivity**	**Specificity**	**Accuracy**	**Precision**	**F1-Score**
F24	0.8721	3	cityblock	0.9105	0.8660	0.8882	0.8717	0.8907
F24→Norm	0.9140	3	cityblock	0.9255	0.9268	**0.9261**	0.9267	**0.9261**
WT→F24	0.8747	3	cityblock	0.9275	0.8758	0.9016	0.8819	0.9041
WT→F24→Norm	**0.9149**	1	cityblock	**0.9327**	0.9137	0.9232	0.9153	0.9239
DWTD	0.8685	5	cityblock	0.8431	0.9098	0.8765	0.9034	0.8722
DWTD→Norm	0.8625	3	cityblock	0.8405	0.9150	0.8761	0.9079	0.8711
WT→DWTD	0.8720	3	cityblock	0.8425	0.9105	0.8765	0.9039	0.8721
WT→DWTD→Norm	0.8667	3	cityblock	0.8405	0.9190	0.8797	0.9121	0.8748
DWTA	0.8957	3	cityblock	0.8634	**0.9451**	0.9042	**0.9402**	0.9002
DWTA→Norm	0.8917	3	cityblock	0.8614	0.9418	0.9016	0.9367	0.8975
WT→DWTA	0.8979	5	cityblock	0.8490	0.9386	0.8938	0.9325	0.8888
WT→DWTA→Norm	0.8942	5	cityblock	0.8451	0.9340	0.8895	0.9275	0.8844
DWTA2	0.8922	1	cityblock	0.8641	0.9307	0.8974	0.9258	0.8938
DWTA2→Norm	0.8855	3	cityblock	0.8562	0.9340	0.8951	0.9284	0.8909
WT→DWTA2	0.8897	1	cityblock	0.8745	0.9275	0.9010	0.9234	0.8983
WT→DWTA2→Norm	0.8871	3	cityblock	0.8627	0.9399	0.9013	0.9348	0.8973

Table 5.3 presents the classification metrics, best hyperparameters, and F1-score values for the KNN models trained with the FE applied to dataset D3. For D3, two different models achieved the highest F1-score (0.9577): **F24→Norm** and **WT→F24→Norm**, both of which were selected for image reconstruction. The latter model (**WT→F24→Norm**) achieved the highest sensitivity (0.9621) and accuracy (0.9575). The **DWTD→Norm** model scored the highest specificity (0.9647) and precision (0.9626). The highest F1-score during training, 0.9503, was obtained by the **WT→F24→Norm** model. The optimal hyperparameters for the **F24→Norm** model were one neighbor and the cityblock distance metric.

For the F24 FE, normalization had a significant impact on the evaluation metrics, improving scores by approximately 0.03–0.05, while WT had negligible influence. For the other FE, both normalization and WT had little effect on performance. All models used the cityblock distance metric. Eight models achieved their best results with one neighbor, six models with three neighbors, and two models with five neighbors. F24 produced both the best- and worst-performing models within this dataset, while DWTD consistently remained the weakest feature set overall.

The KNN-based models achieved metrics above 0.9 in nearly all cases, except for a few instances in dataset D2, indicating strong overall performance. Notably, the best results were obtained from D1, the dataset composed of homogeneous adipose tissue. Specifically, **WT→F24→Norm** achieved a sensitivity of 0.9824, an accuracy of 0.9827, and an F1-score of 0.9827. Meanwhile, **F24→Norm** achieved the best specificity (0.9843) and precision (0.9842).

The models with the highest F1-scores were all trained using the F24 FE, whereas the worst results came from models trained with the DWTD FE. This supports the idea that datasets with greater variation and clearer class separations ($y = 0$ vs. $y = 1$) tend to yield better results, as discussed in Section 5.1. Apart from the F24 FE, normalization and WT had little to no impact, or even a negative impact, on performance. Regarding model hyperparameters, the only distance metric selected after optimization was the cityblock distance. As for the number of neighbors, the majority of models used 1 or 3 neighbors,

Table 5.3: Performance metrics of K-Nearest Neighbors model across different FE of the dataset group D3: Sensitivity, Specificity, Accuracy, Precision, F1-Score, Number of Neighbors, Distance Metric, and F1-Score During Training. WT: Window Trimmed, Norm: Normalized.

K-Nearest Neighbors for D3

	Train			Test				
Feature Engineering	**F1-Score During Training**	**NumNeighbors**	**Distance**	**Sensitivity**	**Specificity**	**Accuracy**	**Precision**	**F1-Score**
F24	0.9169	1	cityblock	0.9275	0.9216	0.9245	0.9220	0.9247
F24→Norm	0.9462	1	cityblock	**0.9621**	0.9529	**0.9575**	0.9534	**0.9577**
WT→F24	0.9166	3	cityblock	0.9105	0.9294	0.9199	0.9280	0.9192
WT→F24→Norm	**0.9503**	1	cityblock	0.9614	0.9536	0.9575	0.9540	**0.9577**
DWTD	0.9065	3	cityblock	0.9209	0.9601	0.9405	0.9585	0.9393
DWTD→Norm	0.9056	5	cityblock	0.9092	**0.9647**	0.9369	**0.9626**	0.9351
WT→DWTD	0.9045	3	cityblock	0.9013	0.9562	0.9288	0.9537	0.9267
WT→DWTD→Norm	0.9007	5	cityblock	0.8902	0.9588	0.9245	0.9558	0.9218
DWTA	0.9189	3	cityblock	0.9235	0.9582	0.9408	0.9567	0.9398
DWTA→Norm	0.9138	1	cityblock	0.9333	0.9634	0.9484	0.9623	0.9476
WT→DWTA	0.9145	3	cityblock	0.9203	0.9627	0.9415	0.9611	0.9402
WT→DWTA→Norm	0.9082	3	cityblock	0.9190	0.9582	0.9386	0.9565	0.9373
DWTA2	0.9169	1	cityblock	0.9327	0.9556	0.9441	0.9545	0.9435
DWTA2→Norm	0.9137	1	cityblock	0.9320	0.9549	0.9435	0.9538	0.9428
WT→DWTA2	0.9161	1	cityblock	0.9320	0.9575	0.9448	0.9564	0.9441
WT→DWTA2→Norm	0.9155	1	cityblock	0.9314	0.9608	0.9461	0.9596	0.9453

although in some cases 5 neighbors were selected.

5.2.2 Decision Tree

It is important to note that, in the case of the DT algorithm, due to its nature, each FE and normalized FE yielded identical results. Therefore, the non-normalized version was chosen for the reconstruction phase, as it requires fewer preprocessing steps.

Table 5.4 presents the classification metrics, best hyperparameters, and corresponding optimizer metric values obtained during training for the DT models trained with D1. The FE that produced the models achieving the best metrics were **WT→DWTA**, with a specificity of 0.9725, accuracy of 0.9588, precision of 0.9718, and F1-score of 0.9583, and **WT→F24**, with a sensitivity of 0.9758. The model selected for the reconstruction phase had the following hyperparameters: a maximum number of splits of 10, a minimum leaf size of 20, and deviance as the split criterion.

WT had a positive impact on specificity and precision in most cases. For example, for F24, WT increased specificity from 0.9059 to 0.9490 and precision from 0.9120 to 0.9490. However, for the other three metrics, WT had little to no impact, or even a negative effect. The maximum number of splits took two values, 6 and 10, while the minimum leaf size varied between 5, 15, and 20. Regarding the split criterion, the majority of models selected deviance as optimal, with only **WT→F24** and its normalized version selecting Gini's Diversity Index (gdi). It is evident that DWTD features produced the weakest models, while DWTA, DWTA2, and F24 yielded very similar performances. DWTD models exhibited lower sensitivity, indicating difficulty in identifying positive cases, whereas DWTA models exhibited lower specificity, suggesting challenges in distinguishing negative cases.

Table 5.5 presents the results and best parameters for the DT models trained with different FE methods applied to D2. By analyzing the table, the model with the best overall results was trained with **WT→DWTA** as its FE, achieving a sensitivity of 0.9523, an accuracy of 0.9464, and an F1-score of 0.9467. The highest values for the remaining metrics were obtained by the model trained with **DWTA** as its FE, reaching 0.9556 specificity and 0.9547 precision; this model also achieved an accuracy of 0.9464.

The impact of WT was minimal for F24, DWTD, and DWTA FE, but notably improved the results for DWTA2, increasing the F1-score from 0.8156 to 0.8362. The D2 group exhibited two different values for the maximum number of splits (8 and 10), and four values for the minimum leaf size (5, 20, 25, and

Table 5.4: Performance metrics of Decision Tree model across different FE of the dataset group D1: Sensitivity, Specificity, Accuracy, Precision, F1-Score, Maximum Number of Splits, Minimum Leaf Size, Split Criterion, and F1-Score During Training. WT: Window Trimmed, Norm: Normalized, gdi: Gini's Diversity Index.

Decision Tree for D1

	Train				Test				
Feature Engineering	**F1-Score During Training**	**MaxNumSplits**	**MinLeafSize**	**SplitCriterion**	**Sensitivity**	**Specificity**	**Accuracy**	**Precision**	**F1-Score**
F24	0.9450	10	5	deviance	**0.9758**	0.9059	0.9408	0.9120	0.9428
F24→Norm	0.9450	10	5	deviance	**0.9758**	0.9059	0.9408	0.9120	0.9428
WT→F24	0.9433	10	20	gdi	0.9490	0.9490	0.9490	0.9490	0.9490
WT→F24→Norm	0.9433	10	20	gdi	0.9490	0.9490	0.9490	0.9490	0.9490
DWTD	0.8565	10	5	deviance	0.9118	0.9020	0.9069	0.9029	0.9073
DWTD→Norm	0.8565	10	5	deviance	0.9118	0.9020	0.9069	0.9029	0.9073
WT→DWTD	0.8843	10	5	deviance	0.9275	0.8190	0.8732	0.8367	0.8797
WT→DWTD→Norm	0.8843	10	5	deviance	0.9275	0.8190	0.8732	0.8367	0.8797
DWTA	0.9355	6	5	deviance	0.9261	0.9621	0.9441	0.9607	0.9431
DWTA→Norm	0.9355	6	5	deviance	0.9261	0.9621	0.9441	0.9607	0.9431
WT→DWTA	**0.9455**	10	20	deviance	0.9451	**0.9725**	**0.9588**	**0.9718**	**0.9583**
WT→DWTA→Norm	**0.9455**	10	20	deviance	0.9451	**0.9725**	**0.9588**	**0.9718**	**0.9583**
DWTA2	0.9372	6	5	deviance	0.9425	0.9523	0.9474	0.9518	0.9471
DWTA2→Norm	0.9372	6	5	deviance	0.9425	0.9523	0.9474	0.9518	0.9471
WT→DWTA2	0.9401	6	15	deviance	0.9366	0.9542	0.9454	0.9534	0.9449
WT→DWTA2→Norm	0.9401	6	15	deviance	0.9366	0.9542	0.9454	0.9534	0.9449

50). Regarding the optimal split criterion, the majority of models selected gdi, except for **WT→DWTA2** and its normalized counterpart, which selected deviance. In terms of F1-score performance, DWTA clearly produced the best models, followed by DWTD with a margin, while DWTA2 yielded the worst-performing models.

Table 5.5: Performance metrics of Decision Tree model across different FE of the dataset group D2: Sensitivity, Specificity, Accuracy, Precision, F1-Score, Maximum Number of Splits, Minimum Leaf Size, Split Criterion, and F1-Score During Training. WT: Window Trimmed, Norm: Normalized, gdi: Gini's Diversity Index.

Decision Tree for D2

	Train				Test				
Feature Engineering	**F1-Score During Training**	**MaxNumSplits**	**MinLeafSize**	**SplitCriterion**	**Sensitivity**	**Specificity**	**Accuracy**	**Precision**	**F1-Score**
F24	0.8477	10	20	gdi	0.8353	0.8569	0.8461	0.8537	0.8444
F24→Norm	0.8477	10	20	gdi	0.8353	0.8569	0.8461	0.8537	0.8444
WT→F24	0.8483	10	5	gdi	0.8242	0.8725	0.8484	0.8661	0.8446
WT→F24→Norm	0.8483	10	5	gdi	0.8242	0.8725	0.8484	0.8661	0.8446
DWTD	0.8488	10	5	gdi	0.8719	0.8529	0.8624	0.8557	0.8637
DWTD→Norm	0.8488	10	5	gdi	0.8719	0.8529	0.8624	0.8557	0.8637
WT→DWTD	0.8478	10	5	gdi	0.9000	0.8255	0.8627	0.8376	0.8677
WT→DWTD→Norm	0.8478	10	5	gdi	0.9000	0.8255	0.8627	0.8376	0.8677
DWTA	0.9023	10	50	gdi	0.9372	**0.9556**	**0.9464**	**0.9547**	0.9459
DWTA→Norm	0.9023	10	50	gdi	0.9372	**0.9556**	**0.9464**	**0.9547**	0.9459
WT→DWTA	**0.9071**	8	25	gdi	**0.9523**	0.9405	**0.9464**	0.9412	**0.9467**
WT→DWTA→Norm	**0.9071**	8	25	gdi	**0.9523**	0.9405	**0.9464**	0.9412	**0.9467**
DWTA2	0.8467	10	5	gdi	0.8301	0.7948	0.8124	0.8018	0.8157
DWTA2→Norm	0.8467	10	5	gdi	0.8301	0.7948	0.8124	0.8018	0.8157
WT→DWTA2	0.8648	10	5	deviance	0.8745	0.7830	0.8288	0.8012	0.8362
WT→DWTA2→Norm	0.8648	10	5	deviance	0.8745	0.7830	0.8288	0.8012	0.8362

Table 5.6 presents the results and parameters for the DT models trained with several combinations of FE applied to D3. The model trained with **DWTA** as its FE achieved the highest values in accuracy (0.9363) and F1-score (0.9362). **WT→DWTA2** obtained the best sensitivity (0.9359), while **F24** achieved the highest specificity (0.9765) and precision (0.9743). The best F1-score during training was obtained by **WT→F24**, with a value of 0.9020. The **DWTA**-based model was trained with a maximum number of splits set to 10, a minimum leaf size of 5, and deviance as the split criterion.

The negative impact of WT was significant for F24 and DWTD, particularly with regards to F1-score, while it had a relatively positive effect for DWTA2. For example, in the case of F24, the F1-score decreased from 0.9311 to 0.9176 after applying WT; in contrast, DWTA2 saw an increase from 0.8874 to 0.9290. For the D3 dataset group, all models used a maximum number of splits of 10. The minimum leaf size varied between 5 and 20, with only the **F24** model and its normalized counterpart using a minimum leaf size of 20. Regarding the split criterion, three model pairs used deviance, while the remainder used gdi. Although **DWTA** achieved the highest F1-score, the **F24**-based models demonstrated more consistent performance, with both model pairs achieving values above 0.9 for most metrics. In contrast, models trained with **DWTD** features yielded the weakest results, with most metrics falling below 0.9 and at least one instance of sensitivity dropping below 0.8.

Table 5.6: Performance metrics of Decision Tree model across different FE of the dataset group D3: Sensitivity, Specificity, Accuracy, Precision, F1-Score, Maximum Number of Splits, Minimum Leaf Size, Split Criterion, and F1-Score During Training. WT: Window Trimmed, Norm: Normalized, gdi; Gini's Diversity Index.

Decision Tree for D3

	Train				Test				
Feature Engineering	**F1-Score During Training**	**MaxNumSplits**	**MinLeafSize**	**SplitCriterion**	**Sensitivity**	**Specificity**	**Accuracy**	**Precision**	**F1-Score**
F24	0.9017	10	20	gdi	0.8915	**0.9765**	0.9340	**0.9743**	0.9311
F24→Norm	0.9017	10	20	gdi	0.8915	**0.9765**	0.9340	**0.9743**	0.9311
WT→F24	**0.9020**	10	5	deviance	0.8804	0.9614	0.9209	0.9580	0.9176
WT→F24→Norm	**0.9020**	10	5	deviance	0.8804	0.9614	0.9209	0.9580	0.9176
DWTD	0.8253	10	5	gdi	0.8111	0.7954	0.8680	0.9152	0.8600
DWTD→Norm	0.8253	10	5	gdi	0.8111	0.7954	0.8680	0.9152	0.8600
WT→DWTD	0.8271	10	5	gdi	0.7980	0.8895	0.8438	0.8784	0.8363
WT→DWTD→Norm	0.8271	10	5	gdi	0.7980	0.8895	0.8438	0.8784	0.8363
DWTA	0.8674	10	5	deviance	0.9353	0.9373	**0.9363**	0.9371	**0.9362**
DWTA→Norm	0.8674	10	5	deviance	0.9353	0.9373	**0.9363**	0.9371	**0.9362**
WT→DWTA	0.8386	10	5	gdi	0.9242	0.7954	0.8598	0.8188	0.8683
WT→DWTA→Norm	0.8386	10	5	gdi	0.9242	0.7954	0.8598	0.8188	0.8683
DWTA2	0.8939	10	5	gdi	0.8732	0.9052	0.8892	0.9021	0.8874
DWTA2→Norm	0.8939	10	5	gdi	0.8732	0.9052	0.8892	0.9021	0.8874
WT→DWTA2	0.8864	10	5	deviance	**0.9359**	0.9209	0.9284	0.9221	0.9290
WT→DWTA2→Norm	0.8864	10	5	deviance	**0.9359**	0.9209	0.9284	0.9221	0.9290

Considering only the F1-score values, the D1-based models of DT demonstrated the best overall performance, with only the DWTD-based models scoring below 0.94. In contrast, more than half of the models from the D2 and D3 datasets achieved F1-score values below 0.9. The impact of WT was most pronounced in the D3-based models, with the largest observed difference reaching 0.0679 in the DWTA models (from 0.8683 to 0.9362). All models with the best performance included **DWTA** as one of their FE steps, whereas the worst-performing models were produced using either **DWTD** or **DWTA2** as their FE approach.

5.2.3 Support Vector Machines

The results from the SVM models trained with each FE method applied to D1, along with the corresponding parameters obtained during cross-validation, are summarized in Table 5.7. The model with the highest F1-score was **F24→Norm**, achieving a value of 0.9892. This model also reached the highest sensitivity (0.9863), accuracy (0.9892), and precision (0.9921). The **F24** model (non-normalized) also achieved a sensitivity of 0.9863. Regarding specificity, the **DWTA** model obtained the highest score; however, it performed poorly in all other metrics, indicating that it was only effective at detecting one class. This reinforces the use of the F1-score as the primary optimization metric among the five used in this dissertation. The selected model (**F24→Norm**) was trained with a penalty parameter $C = 10$ and a kernel coefficient $gamma = 1$.

Apart from the F24-based models—where the non-normalized model achieved a similar F1-score to its normalized counterpart—normalization generally had a clear positive impact, improving F1-score values by up to 0.2 in some cases. The effect of WT was less pronounced, showing a slight improvement for DWTD-based models and a negative impact for the others. The C hyperparameter varied among the values 0.1, 1, 10, and 100, with 100 being the most frequently selected. The $gamma$ hyperparameter took values of 0.01, 0.1, and 1, with 0.1 being the most commonly used.

Table 5.8 presents the results of the SVM models trained on D2 using different FE methods. The best overall performance was achieved by the model trained with **F24→Norm** as its FE, which obtained the highest scores in accuracy (0.9542), precision (0.9609), and F1-score (0.9549). The model

Table 5.7: Performance metrics of Support Vector Machines model across different FE of the dataset droup D1: Sensitivity, Specificity, Accuracy, Precision, F1-Score, Regularization Parameter (C), Kernel Coefficient (Gamma), and F1-Score During Training. WT: Window Trimmed, Norm: Normalized, NaN: Not a Number.

Support Vector Machine for D1

	Train			Test				
Feature Engineering	**F1-Score During Training**	**C**	**Gamma**	**Sensitivity**	**Specificity**	**Accuracy**	**Precision**	**F1-Score**
F24	0.9436	100	1	**0.9863**	0.9680	0.9448	0.9664	0.9445
F24→Norm	**0.9893**	10	1	**0.9863**	0.9922	**0.9892**	**0.9921**	**0.9892**
WT→F24	0.9566	100	1	0.9307	0.9732	0.9520	0.9720	0.9509
WT→F24→Norm	0.9892	100	1	0.9824	0.9895	0.9859	0.9895	0.9859
DWTD	0.6937	10	0.1	0.9725	0.2771	0.6248	0.5736	0.7216
DWTD→Norm	0.9107	100	0.01	0.9078	0.9529	0.9304	0.9507	0.9288
WT→DWTD	0.6902	0.01	0.1	0.9784	0.2425	0.6105	0.5636	0.7152
WT→DWTD→Norm	0.9139	100	0.01	0.9190	0.9588	0.9389	0.9571	0.9376
DWTA	0.6860	0.1	1	0.0000	**1.0000**	0.5000	NaN	NaN
DWTA→Norm	0.9103	100	0.1	0.9516	0.9072	0.9294	0.9111	0.9309
WT→DWTA	0.6863	10	0.1	0.9621	0.3614	0.6618	0.6011	0.7399
WT→DWTA→Norm	0.9005	100	0.1	0.9373	0.8987	0.9180	0.9025	0.9195
DWTA2	0.6829	1	1	0.9712	0.2752	0.6232	0.5762	0.7205
DWTA2→Norm	0.9237	100	0.1	0.9379	0.9327	0.9353	0.9330	0.9355
WT→DWTA2	0.6820	0.1	0.1	0.9314	0.4588	0.6951	0.6325	0.7534
WT→DWTA2→Norm	0.9163	100	0.1	0.9366	0.9216	0.9291	0.9227	0.9296

WT→F24→Norm achieved the highest sensitivity (0.9562). The highest specificity (1.0000) was attained by six models: **WT→DWTD**, **DWTD**, **WT→DWTA**, **DWTA**, **WT→DWTA2**, and **DWTA2**; however, all of these models had a sensitivity of 0.0000, indicating a "lazy" model that classified all samples as negative. The best F1-score during training was also obtained by the **F24→Norm** model, which was trained using a penalty parameter $C = 100$ and a kernel coefficient $gamma = 1$.

The impact of normalization was significant only for the F24 feature set; for other FE strategies, the non-normalized models classified all samples as class 0. Nevertheless, normalization led to a substantial mean improvement of approximately 0.6 in F1-score. Except for F24 (where WT had no effect), WT contributed to performance improvements across other methods, with the most notable increase observed in the DWTD-based models, where the F1-score improved from 0.8326 to 0.9431. The F24-based models consistently produced the best performance, whereas the DWTA2-based models yielded the worst results. The most frequently selected penalty parameter was $C = 100$, and the most common kernel coefficient was $gamma = 0.01$.

As shown in Table 5.9, which summarizes the results and parameters of the SVM models applied to D3, the best-performing model was trained with the **WT→F24→Norm** FE . This model achieved a sensitivity of 0.9797, an accuracy of 0.9745, and an F1-score of 0.9746. The highest precision (0.9721) was obtained by the **WT→DWTA2→Norm**-based model, while the highest specificity (1.0000) was achieved by the model using **DWTD** as its FE. However, similarly to other models scoring 1.0000 in specificity, the DWTD-based model recorded a sensitivity of 0.0000, also indicating a "lazy" model that classified all samples as belonging to class 0. The chosen model also achieved the highest F1-score during training (0.9668) and was optimized with a penalty parameter $C = 100$ and kernel coefficient $gamma = 1$.

Normalization had a substantial positive impact on wavelet coefficient-based FE. For DWTA, the F1-score increased from 0.6953 to 0.9265, and for DWTA2, from 0.6898 to 0.9152. For F24-based FE, normalization improved the F1-score from 0.9195 to 0.9654. The application of WT also showed various effects: it slightly improved F24 from 0.9195 to 0.9216 and DWTA2 from 0.6898 to 0.6988, but negatively affected DWTA, reducing its F1-score from 0.6953 to 0.6933. Normalization also had a positive effect when applied to FE pipelines that included WT; for instance, in the F24 case, the F1-score

Table 5.8: Performance metrics of Support Vector Machines model across different FE of the dataset droup D2: Sensitivity, Specificity, Accuracy, Precision, F1-Score, Regularization Parameter (C), Kernel Coefficient (Gamma), and F1-Score During Training. WT: Window Trimmed, Norm: Normalized, NaN: Not a Number.

Support Vector Machines for D2

	Train			Test				
Feature Engineering	**F1-Score During Training**	**C**	**Gamma**	**Sensitivity**	**Specificity**	**Accuracy**	**Precision**	**F1-Score**
F24	0.8656	100	10	0.9209	0.8484	0.8846	0.8586	0.8887
F24→Norm	**0.9371**	100	1	0.9471	0.9614	**0.9542**	**0.9609**	**0.9539**
WT→F24	0.8753	100	1	0.8974	0.9078	0.9026	0.9069	0.9021
WT→F24→Norm	0.9344	100	1	**0.9562**	0.9458	0.9510	0.9463	0.9512
DWTD	0.6667	0.01	10	0.0000	**1.0000**	0.5000	NaN	NaN
DWTD→Norm	0.7937	100	0.01	0.7641	0.9288	0.8464	0.9147	0.8326
WT→DWTD	0.6667	0.1	10	0.0000	**1.0000**	0.5000	NaN	NaN
WT→DWTD→Norm	0.7930	100	0.01	0.7745	0.9373	0.8559	0.9251	0.9431
DWTA	0.6667	0.01	10	0.0000	**1.0000**	0.5000	NaN	NaN
DWTA→Norm	0.8144	100	0.01	0.7516	0.9575	0.8546	0.9465	0.8379
WT→DWTA	0.6667	0.01	10	0.0000	**1.0000**	0.5000	NaN	NaN
WT→DWTA→Norm	0.8126	100	0.01	0.7693	0.9444	0.8569	0.9326	0.8431
DWTA2	0.6667	0.01	100	0.0000	**1.0000**	0.5000	NaN	NaN
DWTA2→Norm	0.7969	100	0.01	0.7020	0.9523	0.8271	0.9364	0.8024
WT→DWTA2	0.6667	0.01	100	0.0000	**1.0000**	0.5000	NaN	NaN
WT→DWTA2→Norm	0.8035	100	0.01	0.7346	0.9569	0.8458	0.9445	0.8265

improved from 0.9216 to 0.9746. Most models used a penalty parameter $C = 100$ (13 out of 16), while the others used 1 (1 model) or 0.1 (2 models). Regarding the kernel coefficient $\gamma = 0.01$, values were distributed among 0.01 (used in 6 models), 1 (5 models), and 0.1 (5 models).

Table 5.9: Performance metrics of Support Vector Machines model across different FE of the dataset droup D3: Sensitivity, Specificity, Accuracy, Precision, F1-Score, Regularization Parameter (C), Kernel Coefficient (Gamma), and F1-Score During Training. WT: Window Trimmed, Norm: Normalized, NaN: Not a Number.

Support Vector Machines for D3

	Train			Test				
Feature Engineerin	**F1-Score During Training**	**C**	**Gamma**	**Sensitivity**	**Specificity**	**Accuracy**	**Precision**	**F1-Score**
F24	0.9169	100	1	0.9294	0.9078	0.9186	0.9098	0.9195
F24→Norm	0.9656	100	1	0.9669	0.9647	0.9654	0.9648	0.9654
WT→F24	0.9182	100	1	0.9294	0.9124	0.9209	0.9139	0.9216
WT→F24→Norm	**0.9668**	100	1	**0.9797**	0.9693	**0.9745**	0.9696	**0.9746**
DWTD	0.6873	0.1	1	0.0000	**1.0000**	0.5000	NaN	NaN
DWTD→Norm	0.8641	100	0.01	0.8902	0.9582	0.9242	0.9551	0.9215
WT→DWTD	0.6779	100	0.1	0.9654	0.2484	0.6069	0.5622	0.7106
WT→DWTD→Norm	0.8693	100	0.01	0.8752	0.9582	0.9167	0.9544	0.9131
DWTA	0.6747	0.1	0.1	0.9471	0.2229	0.5850	0.5493	0.6953
DWTA→Norm	0.8533	100	0.01	0.8856	0.9739	0.9297	0.9713	0.9265
WT→DWTA	0.6753	100	0.1	0.9131	0.2791	0.5961	0.5588	0.6933
WT→DWTA→Norm	0.8520	100	0.01	0.8974	0.9582	0.9278	0.9555	0.9255
DWTA2	0.6781	100	0.1	0.9418	0.2444	0.5843	0.5502	0.6898
DWTA2→Norm	0.8499	100	0.01	0.8680	0.9712	0.9196	0.9679	0.9152
WT→DWTA2	0.6755	1	0.1	0.9418	0.2464	0.5941	0.5555	0.6988
WT→DWTA2→Norm	0.8462	100	0.01	0.8895	0.9745	0.9320	**0.9721**	0.9290

The best-performing SVM models were those trained using the F24 feature extraction, particularly for the D1 dataset, where both models that included normalization in their FE process achieved F1-scores above 0.98. Excluding the models that classified all samples as class 0, the performances of the DWTD, DWTA, and DWTA2-based models did not differ significantly from one another.

5.2.4 Model Performance Summary

Table 5.10 summarizes the results of the FE that produced the best models for each algorithm trained on each dataset. Among these, the algorithm that yielded the best models was the SVM, while the worst-performing algorithm was DT. The best overall model (SVM) was trained using D1 with **F24→Norm** as its FE, achieving an F1-Score of 0.9892. This model also yielded the highest values for the rest of metrics. A close second, in terms of F1-score, was the model KNN trained with D1 **WT→F24→Norm** with a 0.9827. The D2 produced the weakest models, while the models trained on D1 and D3 showed more comparable results.

Table 5.10: Performance metrics of the best models across different datasets: Sensitivity, Specificity, Accuracy, Precision, F1-Score. WT: Window Trimmed, Norm: Normalized.

Algorithm	Feature Engineering	Sensitivity	Specificity	Accuracy	Precision	F1-Score
KNN	D1 WT→F24→Norm	**0.9824**	**0.9830**	**0.9827**	**0.9830**	**0.9827**
	D2 F24→Norm	0.9255	0.9268	0.9261	0.9267	0.9261
	D3 F24→Norm	0.9621	0.9529	0.9575	0.9534	0.9577
	D3 WT→F24→Norm	0.9614	0.9536	0.9575	0.9540	0.9577
DT	D1 WT→DWTA	0.9451	**0.9725**	**0.9588**	**0.9718**	**0.9583**
	D2 WT→DWTA	**0.9523**	0.9405	0.9464	0.9412	0.9467
	D3 DWTA	0.9353	0.9373	0.9363	0.9371	0.9362
SVM	D1 F24→Norm	**0.9863**	**0.9922**	**0.9892**	**0.9921**	**0.9892**
	D2 F24→Norm	0.9471	0.9614	0.9542	0.9609	0.9539
	D3 WT→F24→Norm	0.9797	0.9693	0.9745	0.9696	0.9746

5.3 Image Reconstruction

This section is organized by dataset. Each subsection presents a table containing the evaluation metrics for the models selected in Section 5.2. Additionally, for each model, four reconstructed images are shown: two corresponding to malignant tumors (one relatively small and one relatively large) and two benign tumors (one relatively small and one significantly large), the rest of reconstructions from the test group are in the Appendix A.2 organized by tumor to facilitate the comparison between models. The title of each image consists of two parts. The first indicates the Magnetic Resonance Imaging (MRI) exam number from which the tumor originated, the tumor type (benign or malignant), and the simulation number within the test set. The second part includes the dataset name, the model used, the FE approach, and the reconstruction step—for example, "188_benign_4 + D1_KNN_F24_Norm: Declustered."

The metrics used for evaluation are Intersection over Union (IoU), Dice Similarity Coefficient (DSC), Relative Volume Error (RVE), and Matthews Correlation Coefficient (MCC). It is important to note that the only metric for which lower values indicate better reconstruction performance is RVE, as this metric evaluates the difference in volume between the ground truth and the predicted output. So, higher values correspond to a greater discrepancy between the size of the ground truth and the one predicted by the model.

After the presentation of each dataset, a separate subsection will compare the reconstructions produced by the best-performing model for that dataset with those obtained using the DAS method. A direct visual comparison was chosen over quantitative evaluation, as it provides a more practical assessment. This approach avoids the degradation of the image produced by the DAS method, which would otherwise be necessary if it were to be binarized for evaluation using the same metrics applied to the Machine Learning (ML) reconstructions.

Figure 5.8 presents an example of a reconstruction obtained using only the Free Space signal subtraction, illustrating the poor quality of the resulting image and reinforcing the decision to use the signal

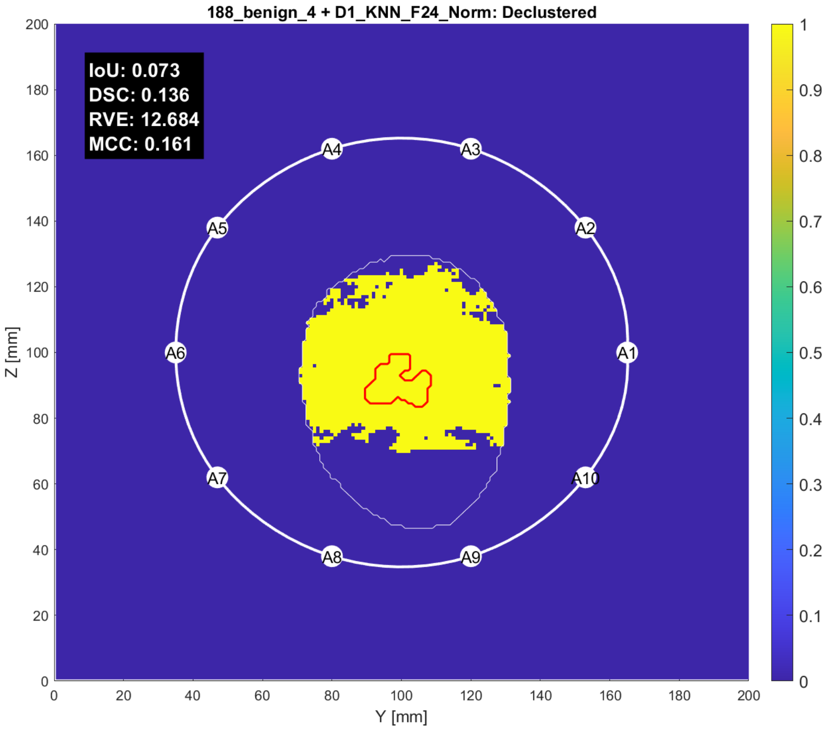


Figure 5.8: Reconstruction of a breast coronal plane obtained using the KNN-based algorithm trained with the FE pipeline F24→Norm, based on signal subtraction from the Free Space simulation. The thicker white contour represents the antenna array, with antennas labeled from A1 to A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

from the phantom without the tumor in order to improve reconstruction quality, even though this approach moves further from real-world conditions and instead represents an idealized scenario.

The reconstruction process, illustrated in Figure 4.4, can also be visualized in Figure 5.9. The top-left image represents the first step of the reconstruction, obtained by averaging the predicted pixel values from each antenna. The top-right image shows the result after applying a threshold of 0.9, a value empirically defined by testing different thresholds. The final result, shown in the bottom image, corresponds to the application of a declustering technique, where all clusters smaller than 30% of the size of the largest cluster are removed from the image.

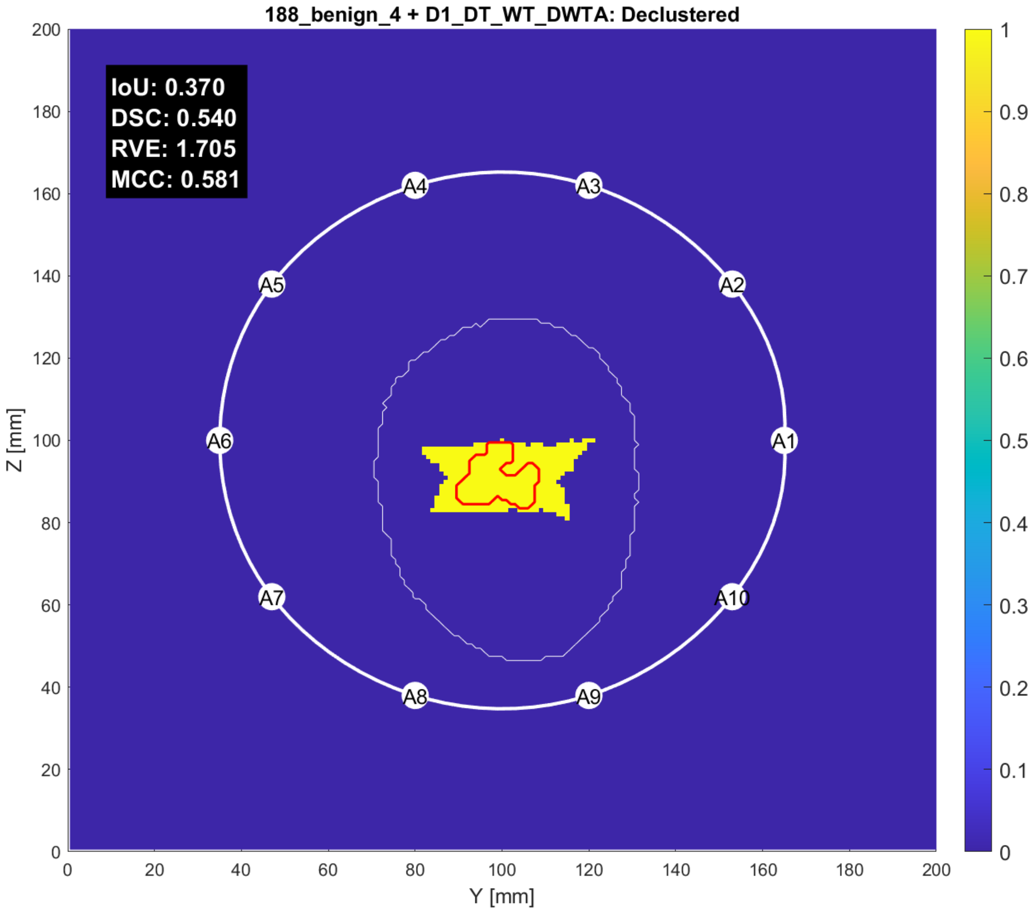


Figure 5.9: Sequential pipeline for pixel-wise tumor detection from raw signal data, incorporating tumorless signal subtraction, windowing, feature extraction, and model prediction (image on the left of the top row titled “Initial,” indicating no post-processing), followed by thresholding (tumor defined as pixels > 0.9; image on the right of the top row titled “Threshold 0.9”) and removal of small clusters to generate the final segmented image (bottom-row image titled “Declustered”). The thicker white contour represents the antenna array, with antennas labeled A1 to A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

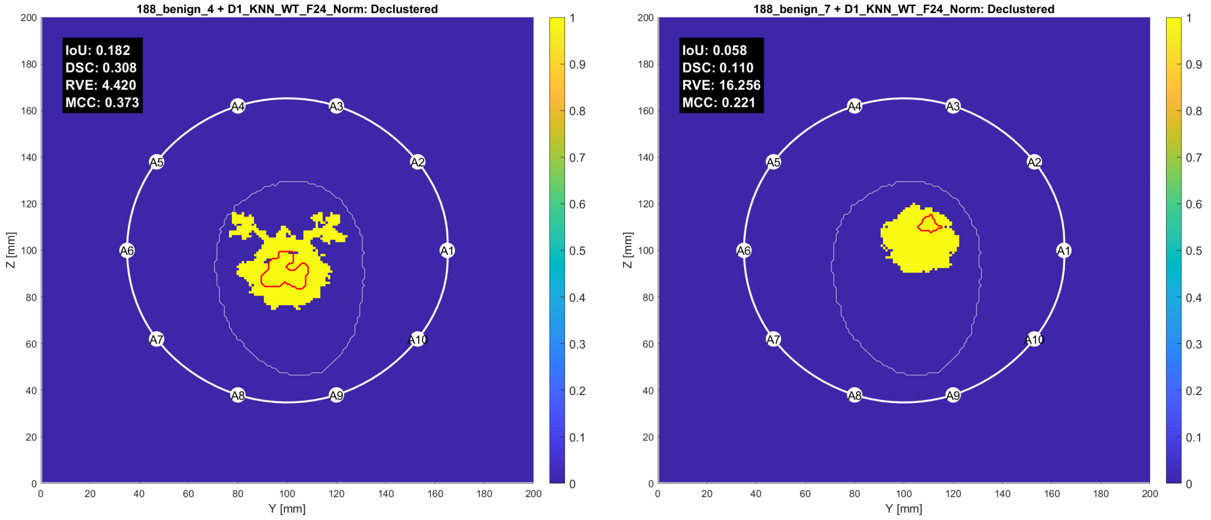
5.3.1 Dataset D1

Table 5.11 presents the results obtained by the models selected for the D1 dataset. Among the evaluated configurations, the DT model achieved the highest scores across all metrics. Although the performance of the SVM model was close, the KNN model recorded the lowest scores in every metric. All configurations produced negative RVE values greater than 1, indicating a systematic tendency to substantially over-predict class 1, with the predicted volume surpassing twice the size of the ground truth.

Table 5.11: Average performance metrics—IoU, DSC, RVE, and MCC—for the best combinations of FE methods and algorithms applied to dataset D1.

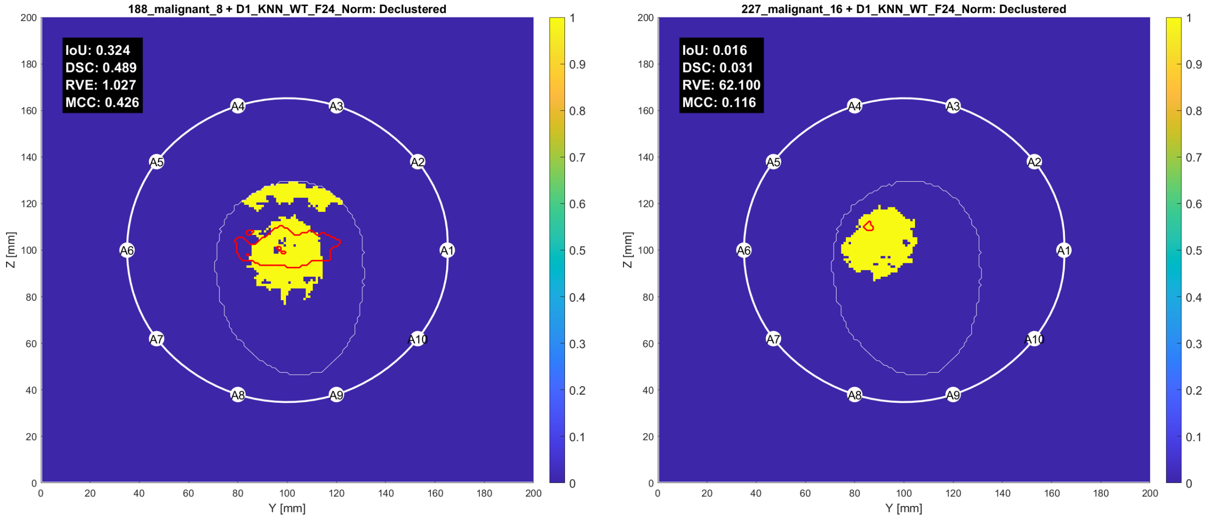
Feature Engineering	Algorithm	IoU	DSC	RVE	MCC
WT→F24→Norm	KNN	0.0997	0.1728	16.14	0.2613
WT→DWTA	DT	0.1382	0.2170	15.06	0.2810
F24→Norm	SVM	0.1056	0.1794	15.91	0.2681

Figure 5.10 shows images reconstructed using the KNN algorithm and the FE pipeline **WT→F24→Norm** from four different tumors in the D1 phantom dataset: two benign (top row) and two malignant (bottom row). The reconstructions demonstrate that the KNN-based model tends to significantly overestimate tumor volumes, as indicated by the RVE, particularly for smaller tumors, such as those in (b) and (d). This overestimation is visually evident in the extent of the predicted regions (yellow) compared to the actual tumor locations and sizes. The discrepancy is also reflected in the low IoU, DSC, and MCC scores in these cases. Specifically, the differences in performance between small and large tumors can also be observed through the obtained metrics. While the lowest scores among the four reconstructions were obtained for the largest tumor, the highest scores were associated with the smallest tumor. This is most evident in the RVE, which changed from 1.027 in (c) to 62.100 in (d). Based on the metrics from each reconstruction, there appears not to be a clear influence of tumor malignancy (benign vs. malignant) on the reconstruction quality.



(a) 188_benign_4

(b) 188_benign_7



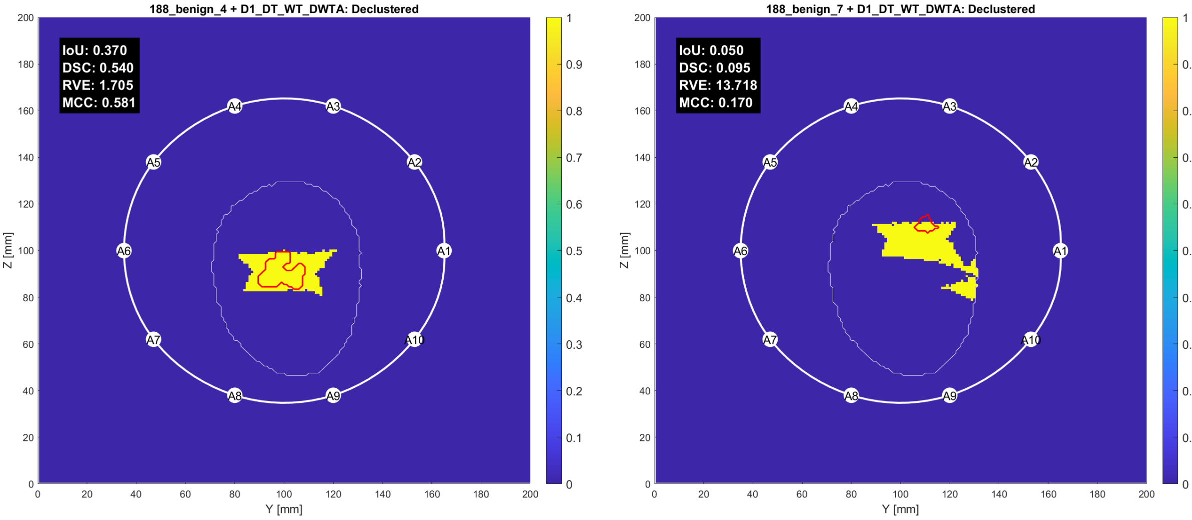
(c) 188_malignant_8

(d) 227_malignant_16

Figure 5.10: Images reconstructed of breast coronal plane using the KNN algorithm and the FE pipeline WT→F24→Norm from four different tumors in the phantom corresponding to the D1 dataset: two benign (top row) and two malignant (bottom row). The thicker white contour represents the antenna array, with antennas labeled from A1 to A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

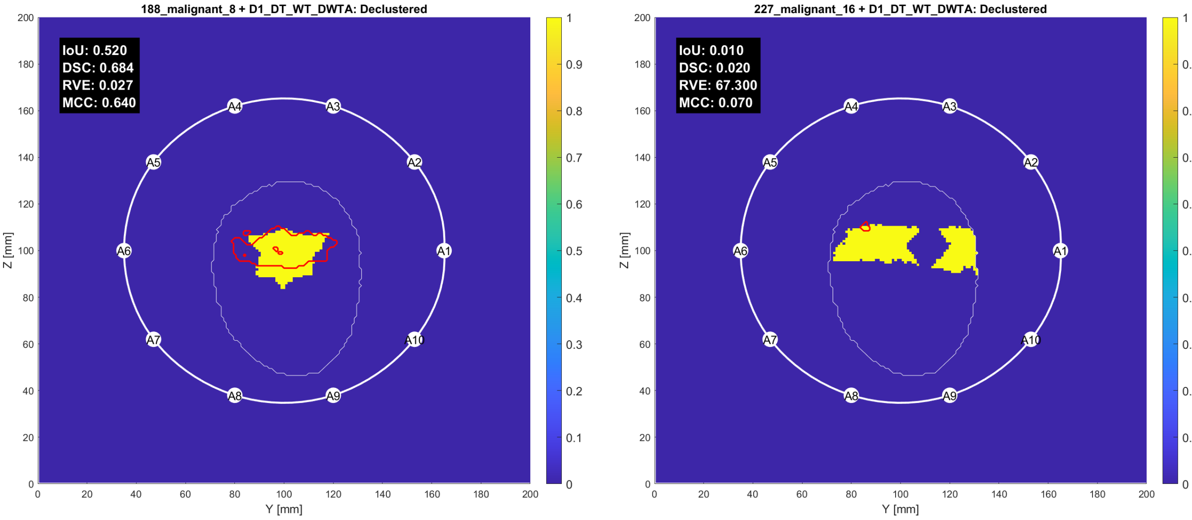
Although the DT-based model using the **WT→DWTA** FE achieved the highest overall scores for the D1 dataset, the reconstructed images are not necessarily the best for each individual tumor. This can be observed by comparing image (d) from Figure 5.10, reconstructed by the KNN-based model, with image (d) from Figure 5.11, where the former outperformed the latter across all evaluated metrics. Similar to the KNN-based model, the DT-based model also demonstrated better performance on larger tumors, with image (c) yielding the best scores and image (d) the worst. Additionally, the DT-based reconstructions exhibit a clear tendency to overestimate tumor regions.

Another notable aspect is the difficulty this model faces in accurately delineating tumor boundaries, resulting in more abrupt transitions between tumor and healthy tissue. This effect might be considered appropriate for malignant tumors due to their more irregular and invasive characteristics. Nevertheless, as observed previously, tumor malignancy does not appear to influence the model's reconstruction performance.



(a) 188_benign_4

(b) 188_benign_7

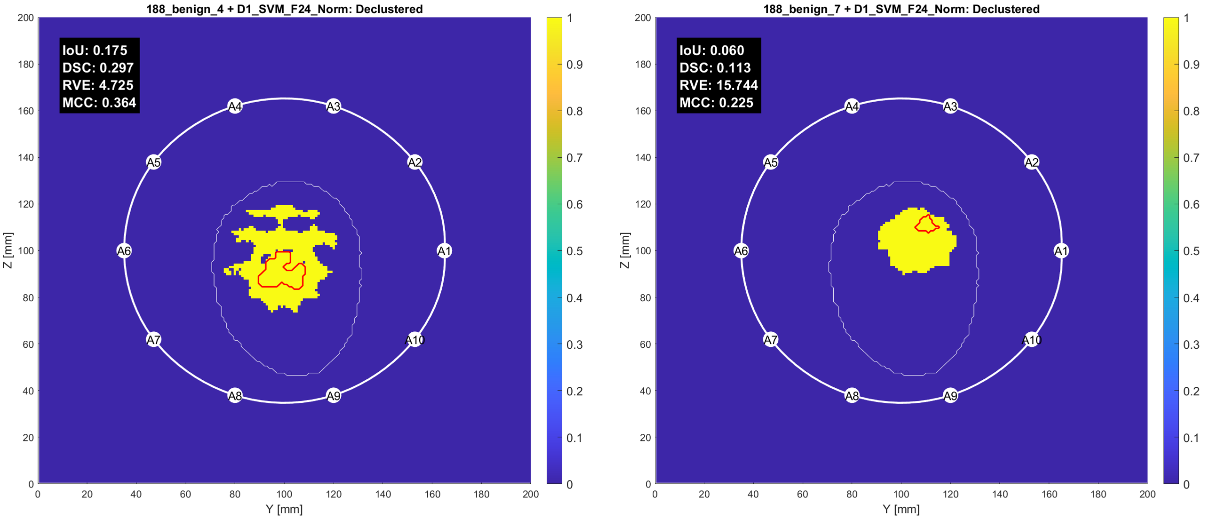


(c) 188_malignant_8

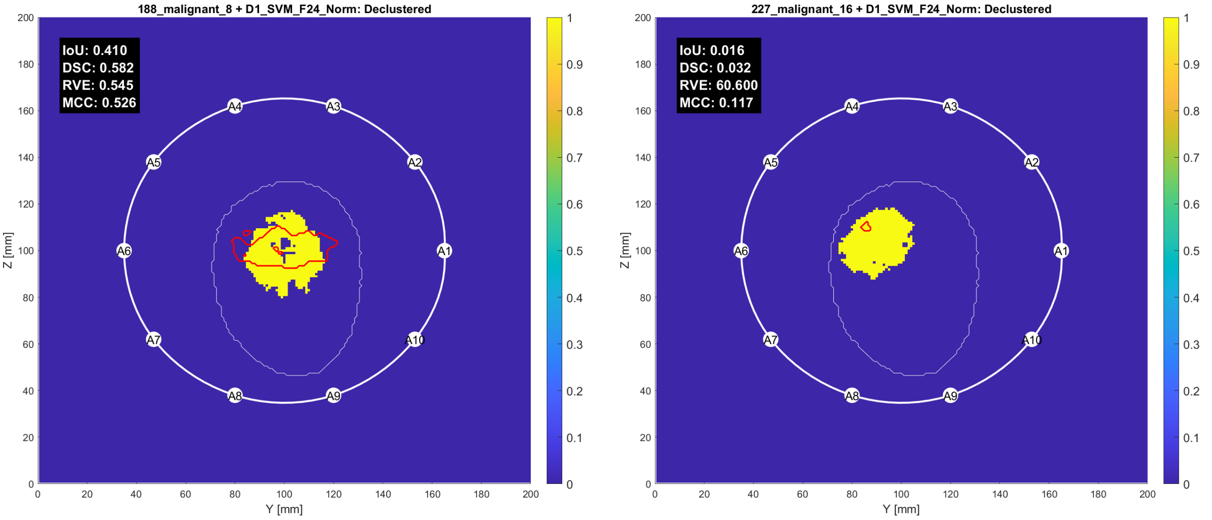
(d) 227_malignant_16

Figure 5.11: Images reconstructed using the DT algorithm and the FE pipeline WT→DWTA from four different tumors in the phantom corresponding to the D1 dataset: two benign (top row) and two malignant (bottom row). The thicker white contour represents the antenna array, with antennas labeled from A1 to A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

As when comparing to the evaluation metrics, the images reconstructed by the SVM-based model using the **F24→Norm** FE (Figure 5.12) closely resemble those obtained with the KNN-based model, both in structure and in metric values. Consequently, the same patterns are observed: an overprediction of tumor pixels relative to the actual tumor size, and a consistent trend where larger tumors yield better reconstruction results—image (c) achieving the highest scores and image (d) the lowest. Finally, it is also evident that tumor class (benign vs. malignant) does not influence the model's performance, with tumor size emerging as the most decisive factor.



(a) 188_benign_4 (b) 188_benign_7



(c) 188_malignant_8 (d) 227_malignant_16

Figure 5.12: Images reconstructed using the SVM algorithm and the FE pipeline F24→Norm from four different tumors in the phantom corresponding to the D1 dataset: two benign (top row) and two malignant (bottom row). The thicker white contour represents the antenna array, with antennas labeled from A1 to A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

5.3.2 Dataset D2

Contrary to the results observed with the D1 dataset, the DT-based model showed the weakest performance on D2 (Table 5.12), while the KNN and SVM models performed similarly, with a slight edge for SVM. Overall, the results improved compared to D1, possibly due to the greater complexity of the D2 phantom, suggesting that model performance tends to benefit from increased structural detail.

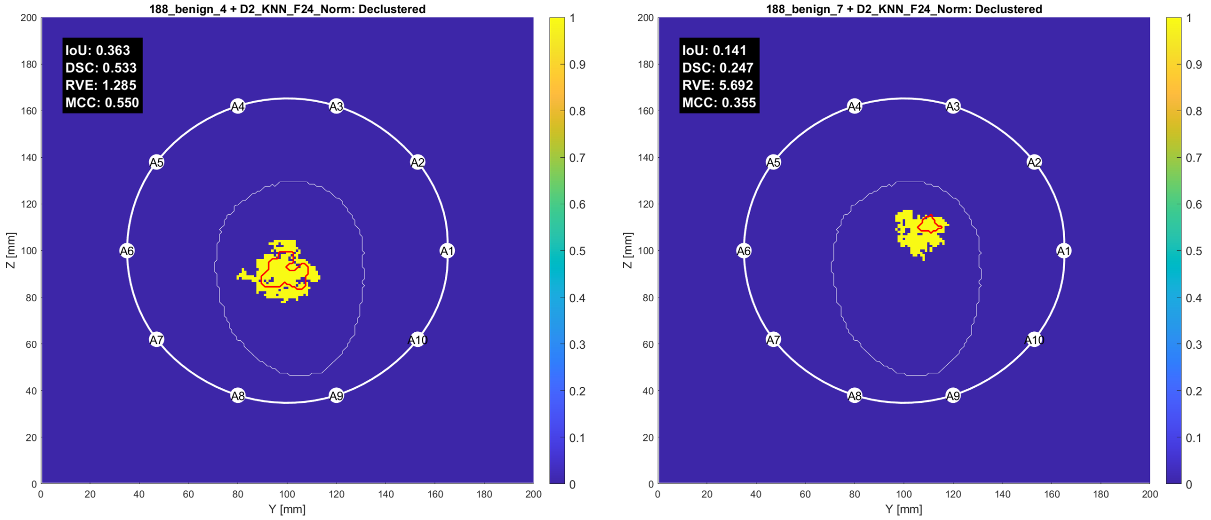
Table 5.12: Average performance metrics—IoU, DSC, RVE, and MCC—for the best combinations of FE methods and algorithms applied to dataset D2.

Feature Engineering	Algorithm	IoU	DSC	RVE	MCC
F24→Norm	KNN	0.1840	0.2999	6.023	0.3815
WT→DWTA	DT	0.0570	0.1044	21.09	0.1579
F24→Norm	SVM	0.1924	0.3107	5.760	0.3848

Figure 5.13 displays reconstructions produced by the KNN algorithm with the **F24→Norm** FE pipeline for four tumors in the D2 dataset—two benign (top row) and two malignant (bottom row). These results show marked improvement over D1, both quantitatively—except for case (c)—and visually. Overestimation is notably reduced in D2, especially in case (a), as reflected by lower RVE values.

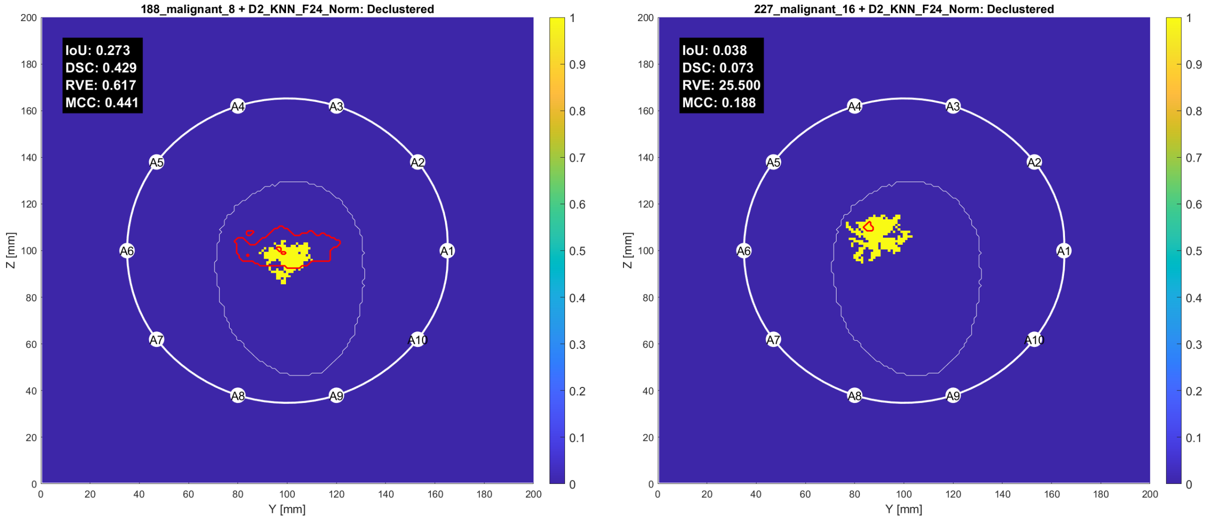
In contrast to the stable performance of the KNN-based models, DT models using the **WT→DWTA** FE pipeline produced poorer reconstructions in D2 (Figure 5.14), particularly for tumor (d), where the model failed to predict any tumor pixels. While metrics tend to improve with larger tumors, this trend is not clearly supported by visual assessment. Most reconstructions—aside from tumor (d)—exhibit a horizontal artifact extending across the image that is mistakenly identified as tumor tissue.

The reconstructions from the SVM-based model using the FE pipeline **F24→Norm** for the D2 dataset are presented in Figure 5.15. These reconstructions are visibly superior to those obtained by the same model on the D1 dataset, which—along with the KNN-based model's performance—suggests that these algorithms may perform better as the complexity of the phantom increases. The best reconstruction was obtained for tumor (a), which challenges the previous observation that reconstruction quality correlates with tumor size. Overall, the results from all models trained with the D2 dataset provide no evidence that reconstruction quality is influenced by tumor classification (benign or malignant).



(a) 188_benign_4

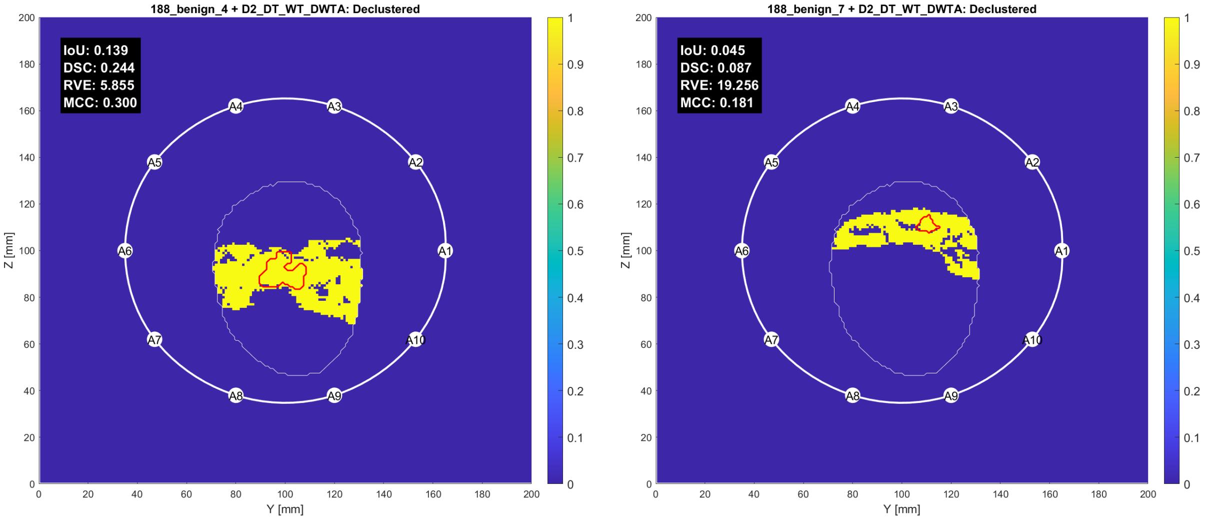
(b) 188_benign_7



(c) 188_malignant_8

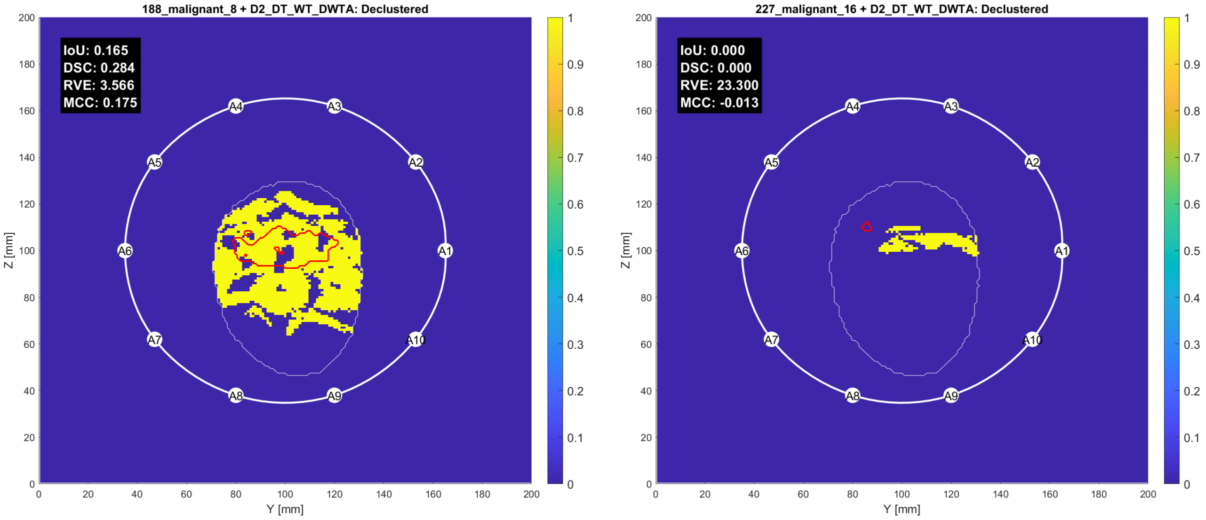
(d) 227_malignant_16

Figure 5.13: Images reconstructed using the KNN algorithm and the FE pipeline F24→Norm from four different tumors in the phantom corresponding to the D2 dataset: two benign (top row) and two malignant (bottom row). The thicker white contour represents the antenna array, with antennas labeled from A1 to A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



(a) 188_benign_4

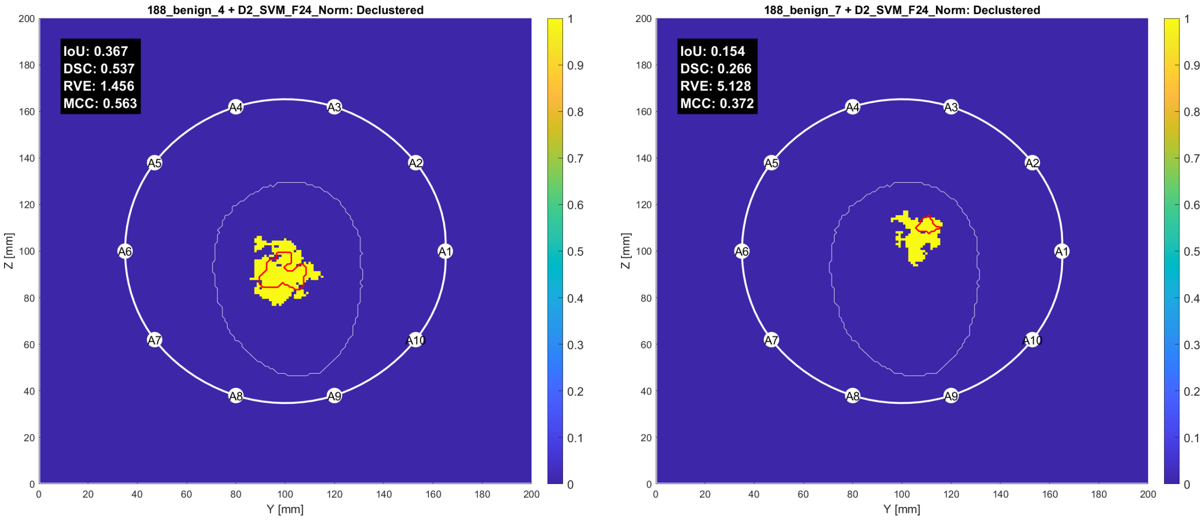
(b) 188_benign_7



(c) 188_malignant_8

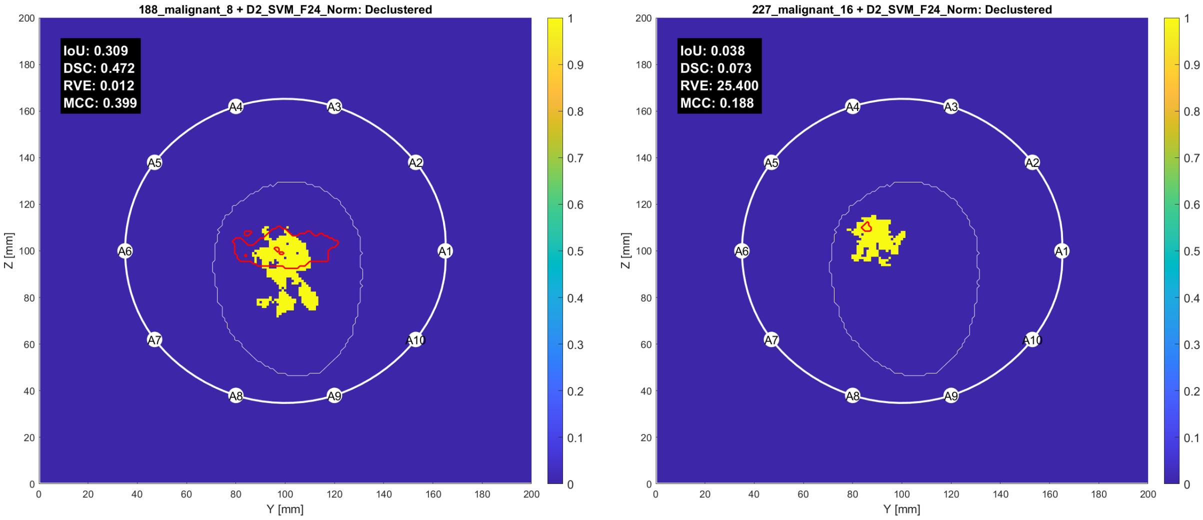
(d) 227_malignant_16

Figure 5.14: Images reconstructed using the DT algorithm and the FE pipeline WT→DWTA from four different tumors in the phantom corresponding to the D2 dataset: two benign (top row) and two malignant (bottom row). The thicker white contour represents the antenna array, with antennas labeled from A1 to A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



(a) 188_benign_4

(b) 188_benign_7



(c) 188_malignant_8

(d) 227_malignant_16

Figure 5.15: Images reconstructed using the SVM algorithm and the FE pipeline F24→Norm from four different tumors in the phantom corresponding to the D2 dataset: two benign (top row) and two malignant (bottom row). The thicker white contour represents the antenna array, with antennas labeled from A1 to A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

5.3.3 Dataset D3

For the D3 dataset, two KNN-based models were selected due to both achieving the highest F1-Score values during model optimization (Section 5.2). As shown in Table 5.13, which presents the evaluation metrics for the four different combinations of algorithms and FE, the same performance ranking observed in D2 is verified—SVM-based models performed best, followed by KNN-based models, while DT-based models performed worst. However, the margin between the KNN- and SVM-based models is larger than in D2. Overall, the evaluation metrics for D3 are slightly worse than those observed for D2, although the DT-based model exhibited improved performance compared to its results in D2. Despite this, the DT model still produced the lowest scores across all metrics.

Table 5.13: Average performance metrics—IoU, DSC, RVE, and MCC—for the best combinations of FE methods and algorithms applied to dataset D3.

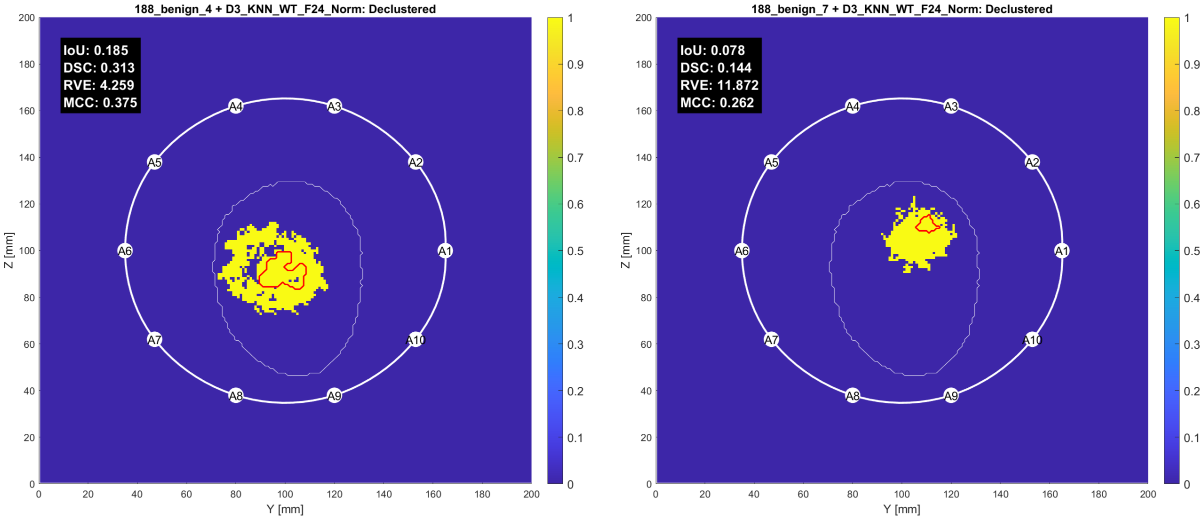
Feature Engineering	**Algorithm**	**IoU**	**DSC**	**RVE**	**MCC**
WT→F24→Norm	KNN	0.1139	0.1973	11.94	0.2880
F24→Norm	KNN	0.1282	0.2169	10.88	0.3082
DWTA	DT	0.0935	0.1627	-0.8493	13.08
WT→F24→Norm	SVM	0.1491	0.2484	8.895	0.3378

Figure 5.16 presents the images reconstructed using the KNN-based model with the FE pipeline **WT→F24→Norm**. These reconstructions clearly show lower quality compared to those produced by the KNN-based model for D2, but they remain superior to those from D1. This trend suggests that the complexity level of the phantom influences the reconstruction quality: a moderate increase in complexity may enhance model performance, but beyond a certain threshold, it may have a detrimental effect. In the D3 reconstructions, the model's performance once again appears to improve with increasing tumor size.

Figure 5.17 presents the image reconstructions from the D3 dataset generated by the KNN-based model using the FE pipeline **F24→Norm**. Compared to the KNN-based model, this FE achieved slightly better performance across all metrics and tumor cases. The same reconstruction patterns are observed, though with a slightly reduced overestimation of tumor pixels. The most notable improvement appears in reconstruction (c), where the differences are particularly visible: approximately 0.1 higher in IoU, DSC, and MCC, and nearly 0.4 lower in RVE.

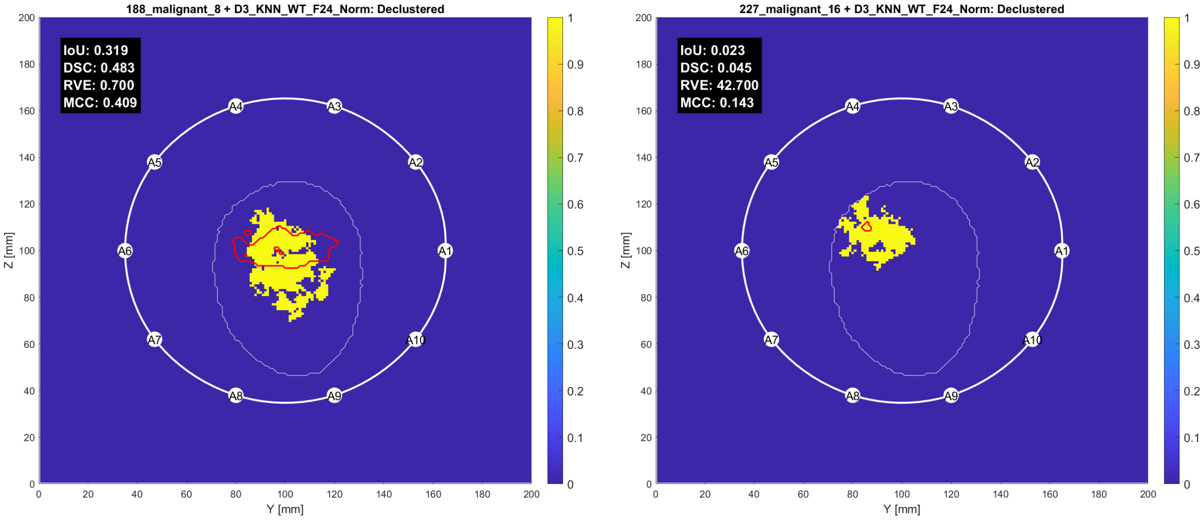
The reconstructions generated using the DT-based model with the DWTA FE pipeline on the D3 dataset (Figure 5.18) show clear improvements compared to those from D2, particularly in cases (a) and (d). The improvement in (d) is mainly due to the model's ability to correctly identify the true tumor pixels in this instance. In contrast to the D2 reconstructions, the D3 results do not exhibit the artifact bands observed previously—one of the most noticeable visual improvements. Among the four reconstructions, case (a) achieved the highest scores across all evaluation metrics, unlike in D2 where it was only the best in MCC, while case (c) yielded the best values in the remaining metrics.

The images reconstructed using the SVM-based model with the **WT→F24→Norm** FE pipeline on the D3 dataset are presented in Figure 5.19. These reconstructions are of lower quality compared to those produced by the same algorithm on the D2 dataset. They follow the same general pattern observed throughout this dissertation: the larger the tumor, the better the reconstruction metrics tend to be. In this specific case, this trend holds for all metrics except the MCC, where the highest score was obtained in reconstruction (a), while the remaining metrics reached their peak in reconstruction (c). Overall, the results from all models trained with the D3 dataset provide no evidence that reconstruction quality is influenced by tumor classification (benign or malignant).



(a) 188_benign_4

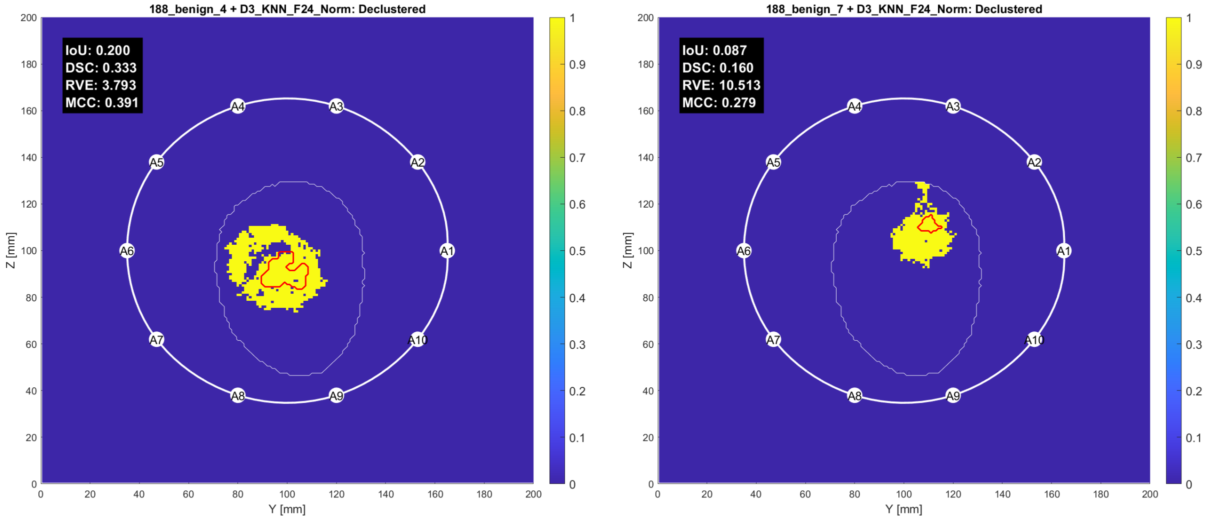
(b) 188_benign_7



(c) 188_malignant_8

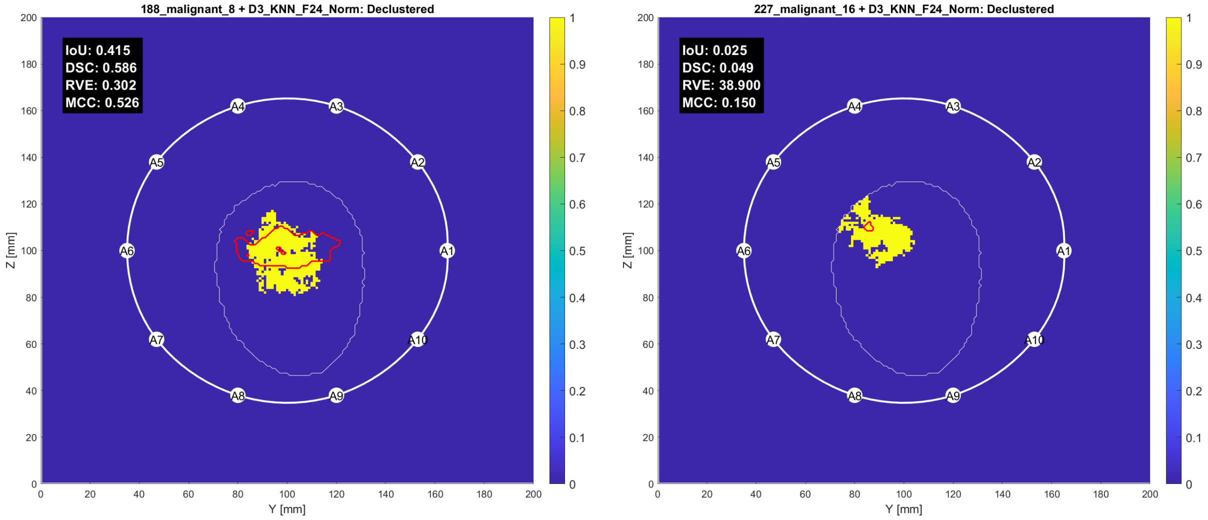
(d) 227_malignant_16

Figure 5.16: Images reconstructed using the KNN algorithm and the FE pipeline WT→F24→Norm from four different tumors in the phantom corresponding to the D3 dataset: two benign (top row) and two malignant (bottom row). The thicker white contour represents the antenna array, with antennas labeled from A1 to A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



(a) 188_benign_4

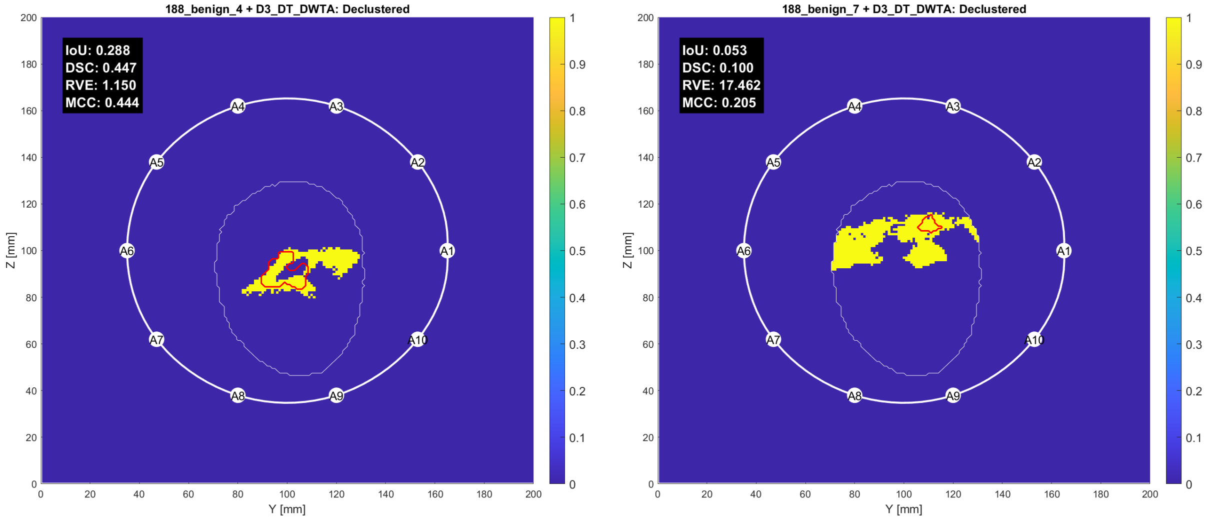
(b) 188_benign_7



(c) 188_malignant_8

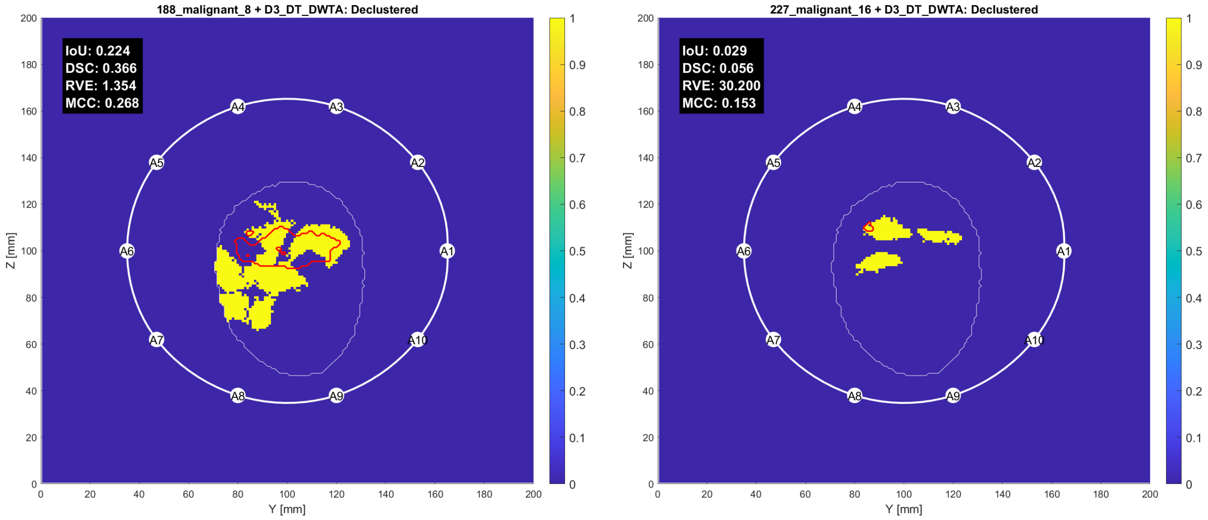
(d) 227_malignant_16

Figure 5.17: Images reconstructed using the KNN algorithm and the FE pipeline F24→Norm from four different tumors in the phantom corresponding to the D3 dataset: two benign (top row) and two malignant (bottom row). The thicker white contour represents the antenna array, with antennas labeled from A1 to A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



(a) 188_benign_4

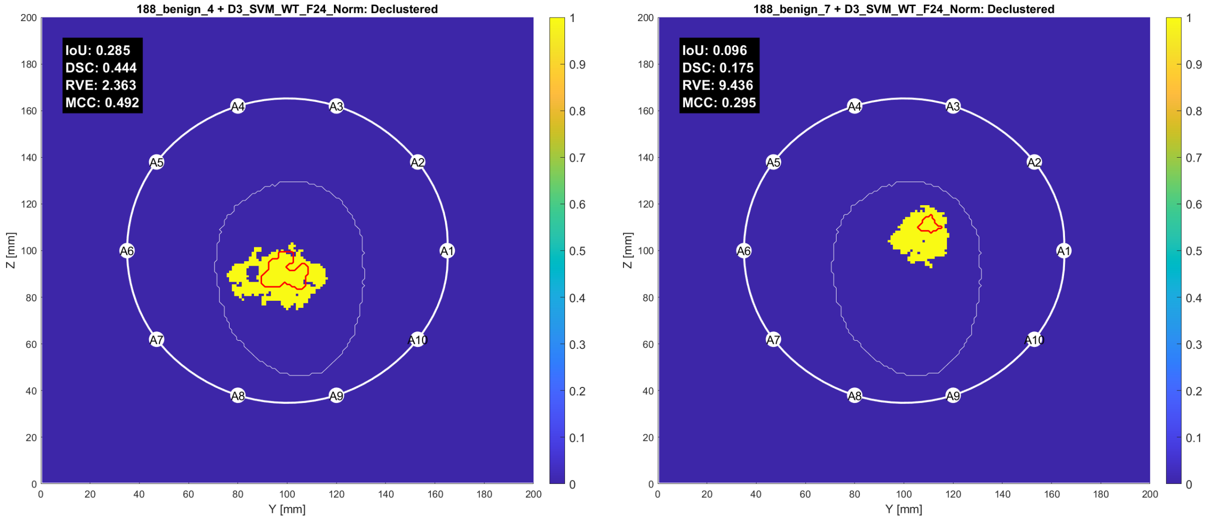
(b) 188_benign_7



(c) 188_malignant_8

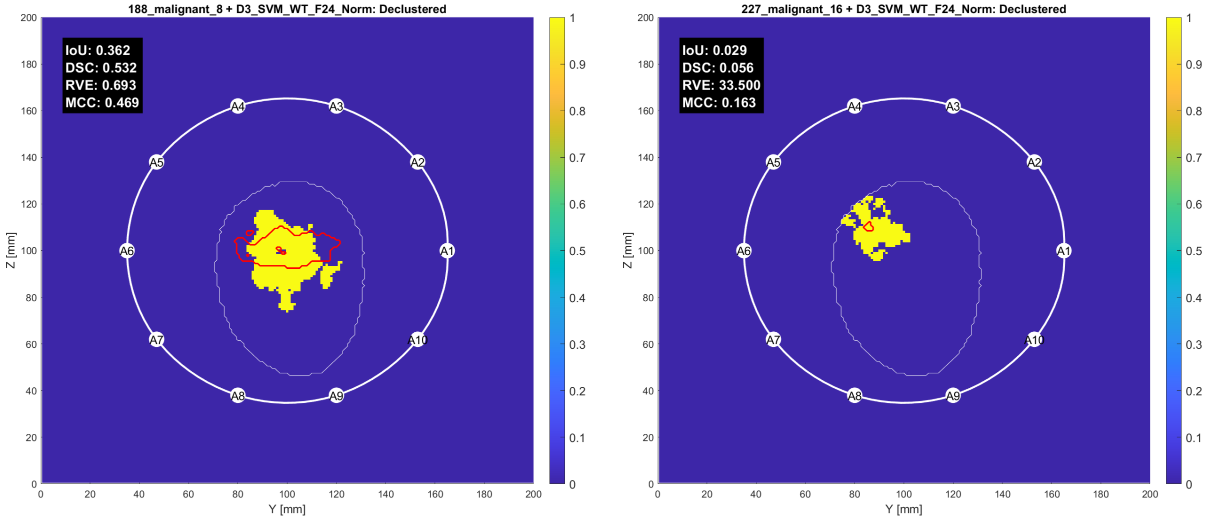
(d) 227_malignant_16

Figure 5.18: Images reconstructed using the DT algorithm and the FE pipeline DWTA from four different tumors in the phantom corresponding to the D3 dataset: two benign (top row) and two malignant (bottom row). The thicker white contour represents the antenna array, with antennas labeled from A1 to A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



(a) 188_benign_4

(b) 188_benign_7



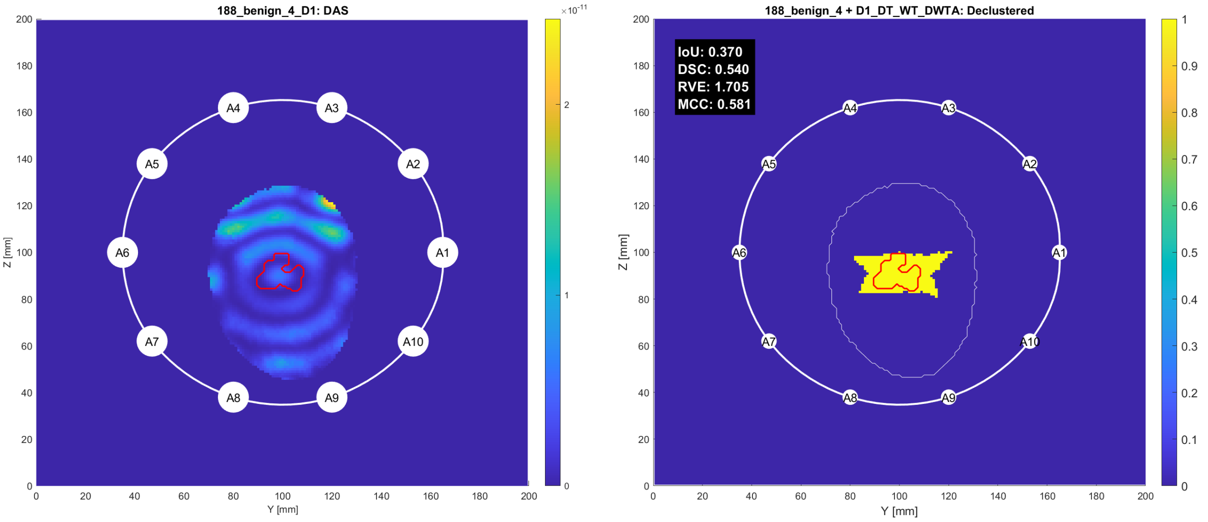
(c) 188_malignant_8

(d) 227_malignant_16

Figure 5.19: Images reconstructed using the SVM algorithm and the FE pipeline WT→F24→Norm from four different tumors in the phantom corresponding to the D3 dataset: two benign (top row) and two malignant (bottom row). The thicker white contour represents the antenna array, with antennas labeled from A1 to A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

5.3.4 Comparision between ML and DAS

The comparison between the reconstructions obtained using the ML-based model—specifically the DT with the FE pipeline **WT→DWTA**—and the DAS approach for the D1 dataset is shown in Figure 5.20. It is evident that in the DAS reconstruction, there is a slight concentration of intensity in the tumor region; however, the highest intensity areas are located in the upper part of the breast, particularly between antennas 2 and 3 hampering successful tumor detection. Considering this, the reconstruction produced by the ML-based approach suggests superior quality and effectiveness in tumor detection, despite the presence of some overestimation and a noticeable angular tendency in the boundaries between pixels predicted as tumor and those predicted as healthy tissue.



(a) DAS approach

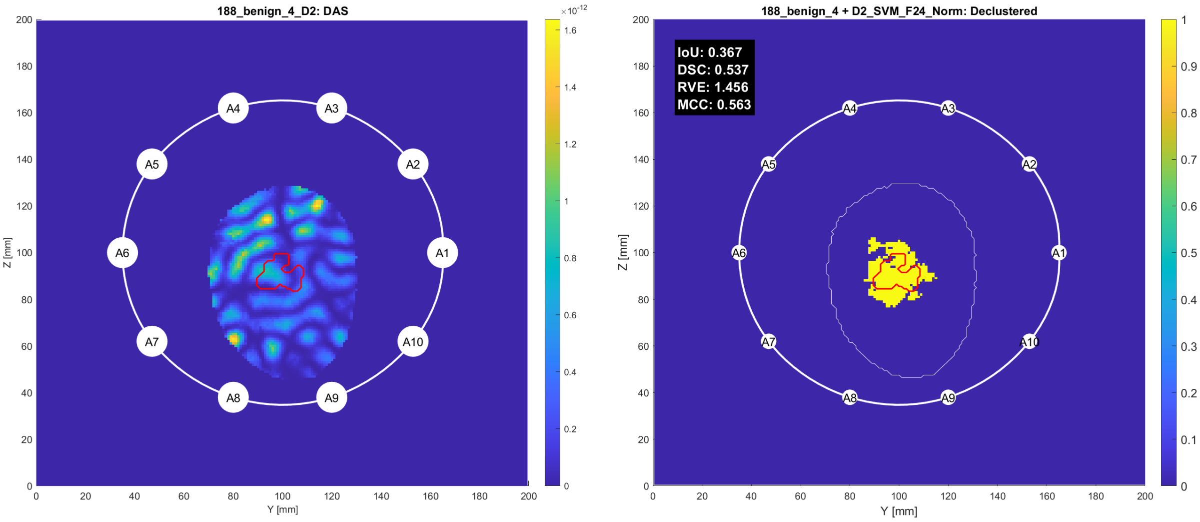
(b) Decision Tree: WT→DWTA

Figure 5.20: Comparison of images reconstructed with the ML approach and the traditional DAS method for a tumor in the phantom corresponding to the characteristics of dataset D1. The thicker white contour represents the antenna array, with antennas labeled from A1 to A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

The reconstructions obtained for the D2 dataset, both from the DAS-based approach and the ML-based approach—specifically the SVM model with the FE pipeline **F24→Norm**—are shown in Figure 5.21. Although there is a noticeable improvement in the quality of the reconstruction using the DAS-based approach compared to D1, the ML-based model still demonstrates superior reconstruction quality. This model also benefited from the increased phantom complexity. In the DAS-based reconstruction, the regions of highest intensity remain distant from the actual tumor location, whereas in the ML-based reconstruction, a reduction in false positives and an improvement in the definition of the boundaries between healthy and tumor tissue are clearly visible.

Finally, the comparison of the reconstructions for the D3 dataset is presented in Figure 5.22, with the DAS-based result shown on the left and the ML-based result—using the SVM model with the **WT→F24→Norm** FE pipeline—on the right. Compared to D2, there is a clear decline in the reconstruction quality of both methods, with both approaching the level of performance observed in D1. Nevertheless, the ML-based approach still shows a distinct advantage in reconstruction quality, as the main intensity regions in the DAS-based reconstruction are not aligned with the actual tumor location, but rather appear near the upper boundary of the breast phantom, close to antenna A5.

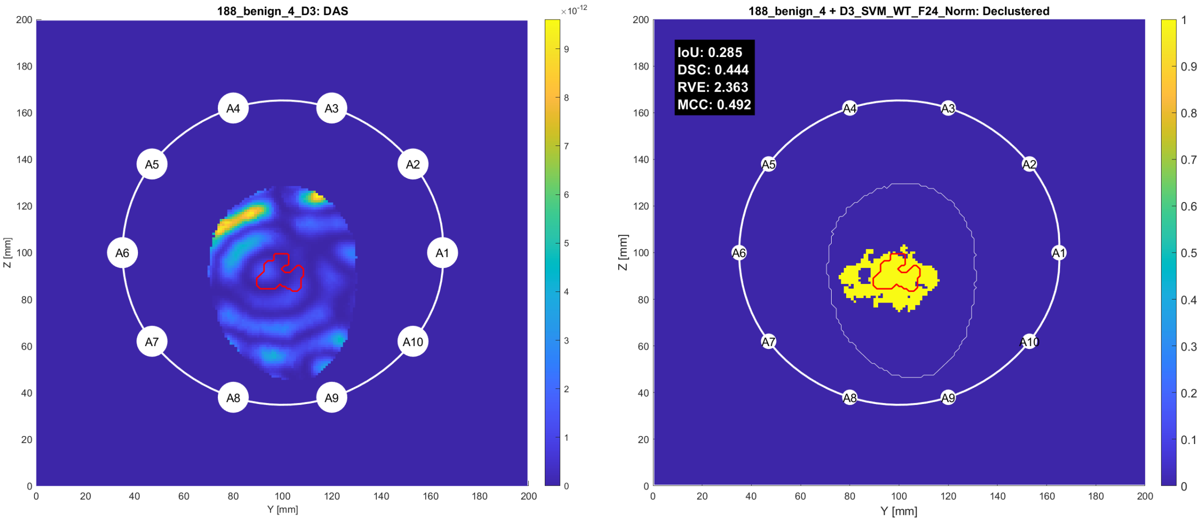
Although the ML-based approach consistently produced higher-quality reconstructions, it remains far from perfect. A substantial proportion of false positives is still present, as evidenced by the significant



(a) DAS approach

(b) Support Vector Machines: F24→Norm

Figure 5.21: Comparison of images reconstructed with the ML approach and the traditional DAS method for a tumor in the phantom corresponding to the characteristics of dataset D2. The thicker white contour represents the antenna array, with antennas labeled from A1 to A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



(a) DAS approach

(b) Support Vector Machines: WT→F24→Norm

Figure 5.22: Comparison of images reconstructed with the ML approach and the traditional DAS method for a tumor in the phantom corresponding to the characteristics of dataset D3. The thicker white contour represents the antenna array, with antennas labeled from A1 to A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

overestimation of tumor regions, which is reflected in the RVE metric values. It is important to emphasize that this comparison was made using the best-performing model out of a total of 48 models (16 per algorithm) for each dataset. This highlights that a pipeline involving ML algorithms does have the potential to outperform traditional approaches, given that the appropriate combination of algorithm and FE is identified.

It is also worth noting that, in some cases, the reconstructions appeared visually better than what the metrics alone would suggest. While the selected metrics were effective in comparing reconstructions relative to one another, they were less reliable when evaluating a reconstruction in isolation. Therefore, future work should focus on developing or identifying objective and quantitative evaluation methods suitable for both ML-based and DAS-based reconstructions. This would enable more accurate and robust analyzes, ultimately aiding in the selection of the most effective approach for detecting and classifying both malignant and benign tumors.

Chapter 6

Conclusions and Future Work

The main objective of this dissertation was to demonstrate the potential of Machine Learning (ML)-based approaches for image reconstruction from breast microwave signals, by comparing the results against a traditional method, the Delay-And-Sum (DAS) algorithm. To this end, a processing pipeline was developed, which led to the establishment of secondary objectives. Specifically, Feature Engineering (FE) pipelines were created to process data obtained from simulations carried out outside the scope of this dissertation. This resulted in 16 combinations of FE configurations, based on the possible variations in the preprocessing steps (as shown in Figure 4.1), including four types of feature extraction and two optional processing steps: reduction of the time window size (from 45 to 35 samples-Window Trimming (WT)- see Figure 4.2) and feature normalization.

Subsequently, each ML algorithm was optimized for each FE pipeline by tuning its hyperparameters, and the resulting models were evaluated using five performance metrics. For each algorithm and dataset combination, the best-performing model was selected based on its F1-Score. This selection resulted in three models for datasets D1 and D2, and four models for D3, due to a tie between two models in F1-Score. The best models and their respective metrics are summarized in Table 5.10. The analysis did not reveal a consistent advantage in applying WT or normalization; their effectiveness appeared to depend on the specific context.

Following this, the image reconstructions generated by each model across the different datasets were presented and discussed. Based on the evaluation metrics, one model per dataset was selected to be compared with the traditional DAS-based method. This comparison led to the conclusion that the ML-based approach achieved superior performance in microwave image reconstruction, thus fulfilling the main goal of this work and confirming its potential for such applications.

Despite the demonstrated superiority of the ML-based approach and the successful achievement of the primary objective, there remains room for improvement, as the work conducted considered only an idealized scenario, focusing solely on the tumor response. Future work could include the application of feature selection techniques or Principal Component Analysis (PCA), as suggested by the correlation matrix analysis in subsection 5.1. Additionally, the development or identification of a single, reliable metric capable of objectively evaluating reconstruction quality is essential. While the current set of metrics effectively captures individual aspects of reconstruction—such as volume, accuracy, and overlap—they may not fully reflect overall image quality, as some reconstructions were visually better than their metric values suggested.

Lastly, it would be worthwhile to apply the pipeline proposed in this dissertation to datasets generated from more complex phantoms to validate the feasibility of this approach, and potentially to extend the approach to deep learning algorithms for further performance gains.

References

[1] W. H. Organization, “Breast cancer.” https://www.who.int/news-room/fact-sheets/detail/breast-cancer, 2024. Accessed: 2025-01-13.

[2] I. A. for Research on Cancer, “Cancer today.” https://gco.iarc.who.int/today/en/dataviz/pie?mode=cancer&types=1&sexes=2&populations=620, 2024. Accessed: 2025-01-13.

[3] S. N. de Saúde, “Registo oncológio nacional.” https://ron.min-saude.pt/pt/tumor/top5/mama/terapeutica-e-prevencao/, 2020. Accessed: 2025-01-13.

[4] E. White, P. Velentgas, M. T. Mandelson, C. D. Lehman, J. G. Elmore, P. Porter, Y. Yasui, and S. H. Taplin, “Variation in mammographic breast density by time in menstrual cycle among women aged 40–49 years,” *Journal of National Cancer Institute*, vol. 90, no. 12, pp. 906–910, 1998.

[5] J. H. Youk and E.-K. Kim, “Supplementary screening sonography in mammographically dense breast: pros and cons,” *Korean journal of radiology*, vol. 11, no. 6, pp. 589–593, 2010.

[6] E. C. Fear, P. M. Meaney, and M. A. Stuchly, “Microwaves for breast cancer detection?,” *IEEE potentials*, vol. 22, no. 1, pp. 12–18, 2003.

[7] H.-H. Tseng, L. Wei, S. Cui, Y. Luo, R. K. Ten Haken, and I. El Naqa, “Machine learning and imaging informatics in oncology,” *Oncology*, vol. 98, no. 6, pp. 344–362, 2020.

[8] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, “Machine learning for medical imaging,” *radiographics*, vol. 37, no. 2, pp. 505–515, 2017.

[9] M. L. Giger, “Machine learning in medical imaging,” *Journal of the American College of Radiology*, vol. 15, no. 3, pp. 512–520, 2018.

[10] R. C. Conceicao, M. O’Halloran, M. Glavin, and E. Jones, “Support vector machines for the classification of early-stage breast cancer based on radar target signatures,” *Progress In Electromagnetics Research B*, vol. 23, pp. 311–327, 2010.

[11] C. B. Moura, “Reconstruction of microwave imaging using machine learning,” master’s thesis, Universidade de Lisboa, Faculdade de Ciências, 2023. Available at http://hdl.handle.net/10451/56884.

[12] T. M. M. Silva, R. C. Conceicao, and D. M. Godinho, “Machine and deep learning applied to medical microwave imaging: a scoping review from reconstruction to classification,” *Progress in Biomedical Engineering*, vol. 7, no. 4, 2025.

[13] S. Gabriel, R. Lau, and C. Gabriel, "The dielectric properties of biological tissues: Ii. measurements in the frequency range 10 hz to 20 ghz," *Physics in medicine & biology*, vol. 41, no. 11, p. 2251, 1996.

[14] C. Polk and E. Postow, *Handbook of Biological Effects of Electromagnetic Fields, -2 Volume Set*. CRC press, 1995.

[15] S. Semenov, "Microwave tomography: review of the progress towards clinical applications," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1900, pp. 3021–3042, 2009.

[16] A. Mobashsher and A. Abbosh, "Developments of tomography and radar-based head imaging systems," in *2015 International Symposium on Antennas and Propagation (ISAP)*, pp. 1–3, IEEE, 2015.

[17] X. Li, E. J. Bond, B. D. Van Veen, and S. C. Hagness, "An overview of ultra-wideband microwave imaging via space-time beamforming for early-stage breast-cancer detection," *IEEE Antennas and Propagation Magazine*, vol. 47, no. 1, pp. 19–34, 2005.

[18] S. C. Hagness, A. Taflove, and J. E. Bridges, "Two-dimensional fdtd analysis of a pulsed microwave confocal system for breast cancer detection: Fixed-focus and antenna-array sensors," *IEEE transactions on biomedical engineering*, vol. 45, no. 12, pp. 1470–1479, 1998.

[19] A. Naghibi and A. R. Attari, "Near-field radar-based microwave imaging for breast cancer detection: A study on resolution and image quality," *IEEE Transactions on antennas and propagation*, vol. 69, no. 3, pp. 1670–1680, 2020.

[20] I. T. Rekanos, "Neural-network-based inverse-scattering technique for online microwave medical imaging," *IEEE transactions on magnetics*, vol. 38, no. 2, pp. 1061–1064, 2002.

[21] L.-Y. Xiao, R. Hong, L.-Y. Zhao, H.-J. Hu, and Q. H. Liu, "A hybrid neural network electromagnetic inversion scheme (hnnemis) for super-resolution 3-d microwave human brain imaging," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 8, pp. 6277–6286, 2022.

[22] R. C. Conceição, D. M. Godinho, D. Byrne, and I. Craddock, "Support vector machines to aid breast cancer diagnosis using a microwave radar prototype," in *2017 XXXIInd General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS)*, pp. 1–3, IEEE, 2017.

[23] L. Papini, M. Badia, L. Sani, S. P. Rana, D. Álvarez Sánchez-Bayuela, A. Vispa, A. Bigotti, G. Raspa, N. Ghavami, C. R. Castellano, D. Bernardi, A. Tagliafico, M. Calabrese, M. Ghavami, and G. Tiberi, "Breast cancer detection using machine learning approaches on microwave-based data," in *2023 17th European Conference on Antennas and Propagation (EuCAP)*, pp. 1–5, 2023.

[24] A. Hossain, M. T. Islam, S. K. Abdul Rahim, M. A. Rahman, T. Rahman, H. Arshad, A. Khandakar, M. A. Ayari, and M. E. Chowdhury, "A lightweight deep learning based microwave brain image network model for brain tumor classification using reconstructed microwave brain (rmb) images," *Biosensors*, vol. 13, no. 2, p. 238, 2023.

[25] A. Garrido, J. Romeu, M. Guardiola, and L. Jofre, “Rapid progress in early detection of colorectal cancer using microwaves,” in *The Advancing World of Applied Electromagnetics: In Honor and Appreciation of Magdy Fahmy Iskander*, pp. 471–500, Springer, 2023.

[26] I. M. Saied, T. Arslan, and S. Chandran, “Classification of alzheimer’s disease using rf signals and machine learning,” *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 6, no. 1, pp. 77–85, 2021.

[27] M. Shehab, L. Abualigah, Q. Shambour, M. A. Abu-Hashem, M. K. Y. Shambour, A. I. Alsalibi, and A. H. Gandomi, “Machine learning in medical applications: A review of state-of-the-art methods,” *Computers in Biology and Medicine*, vol. 145, p. 105458, 2022.

[28] L. B. Holder, M. M. Haque, and M. K. Skinner, “Machine learning for epigenetics and future medical applications,” *Epigenetics*, vol. 12, no. 7, pp. 505–514, 2017.

[29] P. Flach, *Machine learning: the art and science of algorithms that make sense of data*. Cambridge university press, 2012.

[30] Fritz, “Understanding the mathematics behind decision trees.” `https://fritz.ai/wp-content/uploads/2023/09/1jojTznh4HOX_8cGw_04ODA-1536x864.png`, 2023. Accessed: 2025-01-08.

[31] M. Köppen, “The curse of dimensionality,” in *5th online world conference on soft computing in industrial applications (WSC5)*, vol. 1, pp. 4–8, 2000.

[32] A. Kerhet, M. Raffetto, A. Boni, and A. Massa, “A svm-based approach to microwave breast cancer detection,” *Engineering Applications of Artificial Intelligence*, vol. 19, no. 7, pp. 807–818, 2006.

[33] A. Ashtari, S. Noghanian, A. Sabouni, J. Aronsson, G. Thomas, and S. Pistorius, “Using a priori information for regularization in breast microwave image reconstruction,” *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 9, pp. 2197–2208, 2010.

[34] R. C. Conceição, D. M. Godinho, D. Byrne, and I. Craddock, “Support Vector Machines to aid breast cancer diagnosis using a microwave radar prototype,” in *2017 32nd General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS)*, vol. 2017-January, pp. 1–3, 2017.

[35] P. Shah, G. Chen, and M. Moghaddam, “Learning nonlinearity of microwave imaging through deep learning,” in *2018 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting*, pp. 699–700, IEEE, 2018.

[36] V. Khoshdel, A. Ashraf, and J. LoVetri, “Enhancement of multimodal microwave-ultrasound breast imaging using a deep-learning technique,” *Sensors*, vol. 19, no. 18, p. 4050, 2019.

[37] V. Khoshdel, M. Asefi, A. Ashraf, and J. LoVetri, “Full 3d microwave breast imaging using a deep-learning technique,” *Journal of Imaging*, vol. 6, no. 8, p. 80, 2020.

[38] P. Mojabi, V. Khoshdel, and J. Lovetri, “Tissue-type classification with uncertainty quantification of microwave and ultrasound breast imaging: A deep learning approach,” *IEEE Access*, vol. 8, pp. 182092–182104, 2020.

[39] M. Ambrosanio, S. Franceschini, F. Baselice, V. Pascazio, *et al.*, “Artificial neural networks for quantitative microwave breast imaging.,” in *Bioimaging*, pp. 204–208, 2020.

[40] P. Mojabi, M. Hughson, V. Khoshdel, I. Jeffrey, and J. LoVetri, "Cnn for compressibility to permittivity mapping for combined ultrasound-microwave breast imaging," *IEEE Journal on Multiscale and Multiphysics Computational Techniques*, vol. 6, pp. 62–72, 2021.

[41] M. Ambrosanio, S. Franceschini, V. Pascazio, and F. Baselice, "An end-to-end deep learning approach for quantitative microwave breast imaging in real-time applications," *Bioengineering*, vol. 9, no. 11, p. 651, 2022.

[42] M. Ambrosanio, M. M. Autorino, S. Franceschini, F. Baselice, and V. Pascazio, "Microwave breast imaging via deep learning," in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pp. 1–4, IEEE, 2022.

[43] S. Costanzo, A. Flores, and G. Buonanno, "Machine learning approach to quadratic programming-based microwave imaging for breast cancer detection," *Sensors*, vol. 22, no. 11, p. 4122, 2022.

[44] S. Costanzo, A. Flores, and G. Buonanno, "Machine learning methods for microwave imaging in cancer detection," in *2022 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, pp. 1–5, IEEE, 2022.

[45] V. Noël, Y. Qin, T. Rodet, and D. Lesselier, "Breast imaging by cascaded cnn from joint microwave and ultrasonic data," in *2022 30th European Signal Processing Conference (EUSIPCO)*, pp. 917–921, IEEE, 2022.

[46] Y. Qin, P. Ran, T. Rodet, and D. Lesselier, "Breast imaging by convolutional neural networks from joint microwave and ultrasonic data," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 8, pp. 6265–6276, 2022.

[47] G. Fontaine and S. Pistorius, "Machine learning based reconstruction of point-like scatterers in a portable microwave detection device," in *2023 17th European Conference on Antennas and Propagation (EuCAP)*, pp. 1–5, IEEE, 2023.

[48] M. B. Bicer, "Radar-based microwave breast imaging using neurocomputational models," *Diagnostics*, vol. 13, no. 5, p. 930, 2023.

[49] S. Costanzo and A. Flores, "Cvnn-based microwave imaging approach," in *2023 IEEE Conference on Antenna Measurements and Applications (CAMA)*, pp. 728–731, IEEE, 2023.

[50] V. Khoshdel, P. Mojabi, and J. LoVetri, "A multi-branch deep learning architecture for microwave-ultrasound breast imaging," in *2023 XXXVth General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS)*, pp. 1–4, IEEE, 2023.

[51] M. Borghouts, M. Ambrosanio, S. Franceschini, M. M. Autorino, V. Pascazio, and F. Baselice, "Microwave breast sensing via deep learning for tumor spatial localization by probability maps," *Bioengineering*, vol. 10, no. 10, p. 1153, 2023.

[52] S. Franceschini, M. M. Autorino, M. Ambrosanio, V. Pascazio, and F. Baselice, "A deep learning approach for diagnosis support in breast cancer microwave tomography," *Diagnostics*, vol. 13, no. 10, p. 1693, 2023.

[53] M. Ambrosanio, M. Borghouts, S. Franceschini, M. M. Autorino, V. Pascazio, and F. Baselice, “Enhanced deep-learning-based microwave sensing technology for breast cancer localization,” in *2024 IEEE International Conference on E-health Networking, Application & Services (HealthCom)*, pp. 1–5, IEEE, 2024.

[54] R. C. Conceição, D. Byrne, J. A. Noble, and I. Craddock, “Initial study for the investigation of breast tumour response with classification algorithms using a microwave radar prototype,” in *2016 10th European Conference on Antennas and Propagation (EuCAP)*, pp. 1–2, 2016.

[55] M. J. Burfeindt, T. J. Colgan, R. O. Mays, J. D. Shea, N. Behdad, B. D. Van Veen, and S. C. Hagness, “Mri-derived 3-d-printed breast phantom for microwave breast imaging validation,” *IEEE antennas and wireless propagation letters*, vol. 11, pp. 1610–1613, 2012.

[56] R. C. Conceição, H. Medeiros, D. M. Godinho, M. O’Halloran, D. Rodriguez-Herrera, D. Flores-Tapia, and S. Pistorius, “Classification of breast tumor models with a prototype microwave imaging system,” *Medical physics*, vol. 47, no. 4, pp. 1860–1870, 2020.

[57] G. Zhu, A. Bialkowski, L. Guo, B. Mohammed, and A. Abbosh, “Stroke classification in simulated electromagnetic imaging using graph approaches,” *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 5, no. 1, pp. 46–53, 2020.

[58] A. Fasoula, L. Duchesne, J. D. Gil Cano, B. M. Moloney, S. M. Abd Elwahab, and M. J. Kerin, “Automated breast lesion detection and characterization with the wavelia microwave breast imaging system: Methodological proof-of-concept on first-in-human patient data,” *Applied Sciences*, vol. 11, no. 21, 2021.

[59] T. Pokorny, O. Fiser, T. Drizdal, and J. Vrba, “2d numerical dataset for microwave svm-based brain stroke classification,” in *2023 Photonics & Electromagnetics Research Symposium (PIERS)*, pp. 1705–1711, IEEE, 2023.

[60] R. Ullah, Y. Dong, T. Arslan, and S. Chandran, “A machine learning-based classification method for monitoring alzheimer’s disease using electromagnetic radar data,” *IEEE Transactions on Microwave Theory and Techniques*, 2023.

[61] A. Janjic, I. Akduman, M. Cayoren, O. Bugdayci, and M. E. Aribal, “Microwave breast lesion classification–results from clinical investigation of the safe microwave breast cancer system,” *Academic Radiology*, vol. 30, pp. S1–S8, 2023.

[62] I. G. Zubal, C. R. Harrell, E. O. Smith, Z. Rattner, G. Gindi, and P. B. Hoffer, “Computerized three-dimensional segmented human anatomy,” *Medical physics*, vol. 21, no. 2, pp. 299–302, 1994.

[63] A. Christ, W. Kainz, E. G. Hahn, K. Honegger, M. Zefferer, E. Neufeld, W. Rascher, R. Janka, W. Bautz, J. Chen, *et al.*, “The virtual family—development of surface-based anatomical models of two adults and two children for dosimetric simulations,” *Physics in Medicine & Biology*, vol. 55, no. 2, p. N23, 2009.

[64] E. Lee, W. Duffy, R. Hadimani, M. Waris, W. Siddiqui, F. Islam, M. Rajamani, R. Nathan, and D. Jiles, “Investigational effect of brain-scalp distance on the efficacy of transcranial magnetic stimulation treatment in depression,” *IEEE Transactions on Magnetics*, vol. 52, no. 7, pp. 1–4, 2016.

[65] R. C. Conceicao, M. O'Halloran, M. Glavin, and E. Jones, "Support vector machines for the classification of early-stage breast cancer based on radar target signatures," *Progress In Electromagnetics Research B*, vol. 23, pp. 311–327, 2010.

[66] R. Conceicao, M. O'halloran, M. Glavin, and E. Jones, "Evaluation of features and classifiers for classification of early-stage breast cancer," *Journal of Electromagnetic Waves and Applications*, vol. 25, no. 1, pp. 1–14, 2011.

[67] I. Saied, T. Arslan, S. Chandran, C. Smith, T. Spires-Jones, and S. Pal, "Non-invasive rf technique for detecting different stages of alzheimer's disease and imaging beta-amyloid plaques and tau tangles in the brain," *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 4060–4070, 2020.

[68] I. Saied, M. Bashri, T. Arslan, C. Smith, and S. Chandran, "Dielectric measurements of brain tissues with alzheimer's disease pathology in the microwave region," in *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1–6, IEEE, 2019.

[69] A. Hossain, M. T. Islam, and A. F. Almutairi, "A deep learning model to classify and detect brain abnormalities in portable microwave based imaging system," *Scientific Reports*, vol. 12, no. 1, p. 6319, 2022.

[70] A. C. Pelicano, M. C. T. Gonçalves, D. M. Godinho, T. Castela, M. L. Orvalho, N. A. M. Araújo, E. Porter, and R. C. Conceição, "Development of 3d mri-based anatomically realistic models of breast tissues and tumours for microwave imaging diagnosis," *Sensors*, vol. 21, no. 24, 2021.

[71] A. C. Pelicano, N. A. Araújo, D. M. Godinho, and R. C. Conceição, "A preliminary study on the impact of model complexity in classification in breast microwave imaging," in *2025 19th European Conference on Antennas and Propagation (EuCAP)*, pp. 1–5, IEEE, 2025.

[72] R. C. Conceição and D. M. Godinho, "Extracting features from multistatic signals in a radar microwave imaging system for breast cancer detection," *2nd URSI AT-RASC, Gran Canaria, Spain*, 2018.

Appendix A

Appendices

To provide additional supporting material, two appendices are included. Appendix A.1 summarizes the scoping review article with two illustrative figures, offering an overview of the included studies and the selection process. Appendix A.2 presents the remaining reconstructed images from the test group that were not shown in the main text, organized by tumor to facilitate comparison between models and provide a more comprehensive view of the experimental results.

A.1 Supplementary Information for Chapter 1 – Introduction

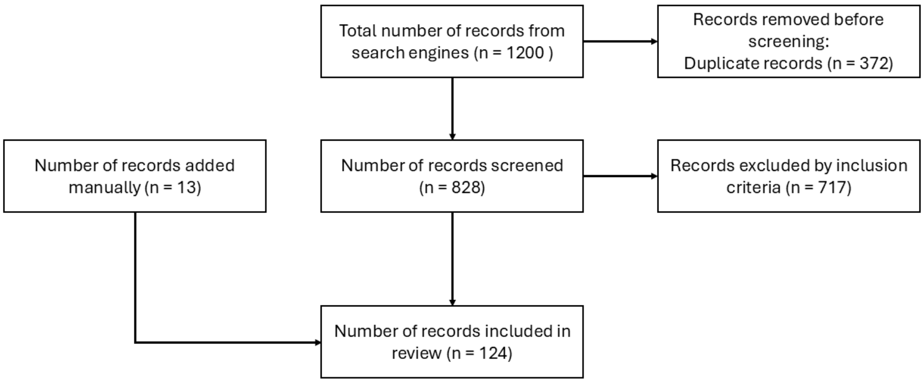


Figure A.1: Flowchart of the study selection process. A total of 1,200 records were initially retrieved from search engines, with 372 duplicates removed prior to screening. An additional 13 records were manually included. After screening 828 records, 717 were excluded based on inclusion criteria, resulting in 124 studies included in the final review [12].

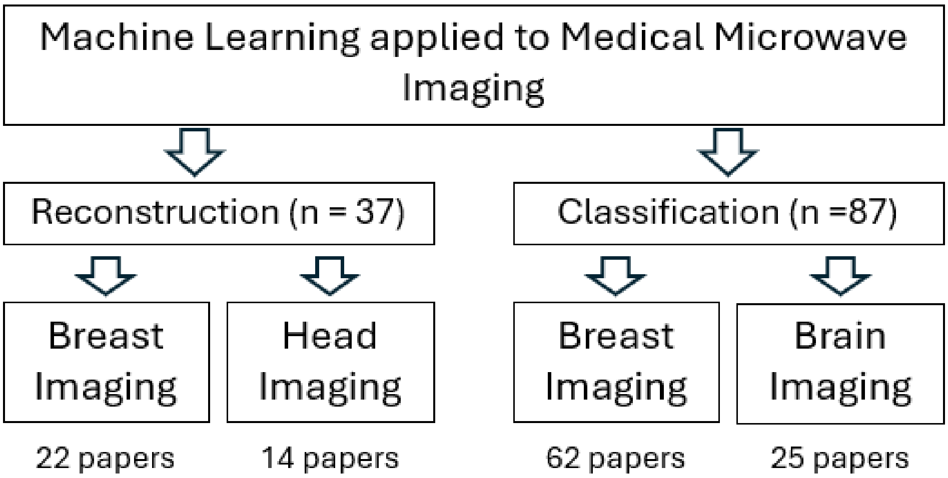
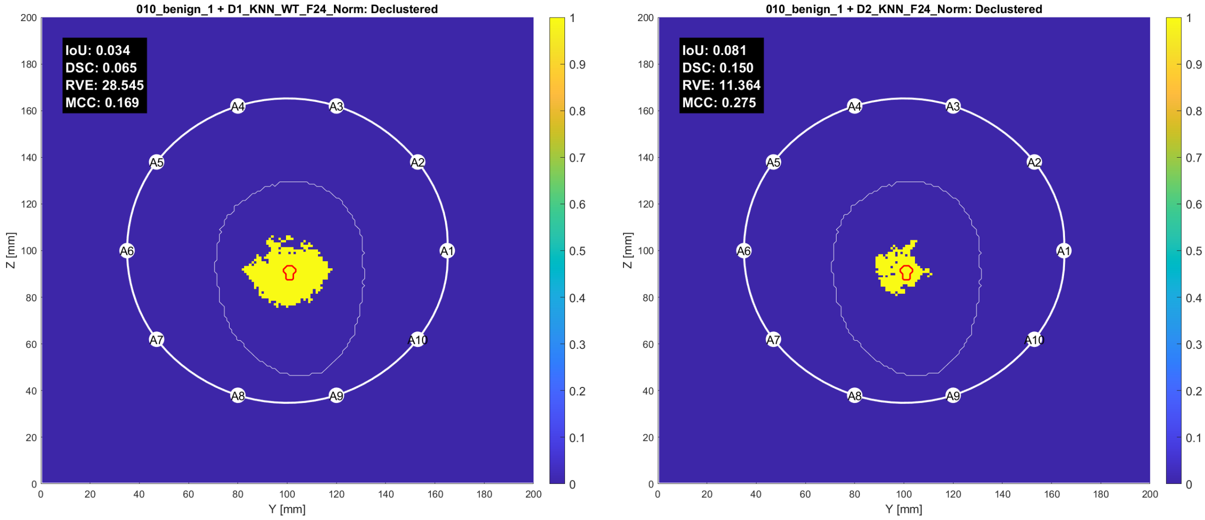


Figure A.2: Overview of the reviewed papers categorized by their focus within machine learning applications in medical microwave imaging. The studies are divided into image reconstruction (37 papers)—including breast imaging (22) and brain imaging (14)—and classification (87 papers)-including breast imaging (62) and brain imaging (25) [12].

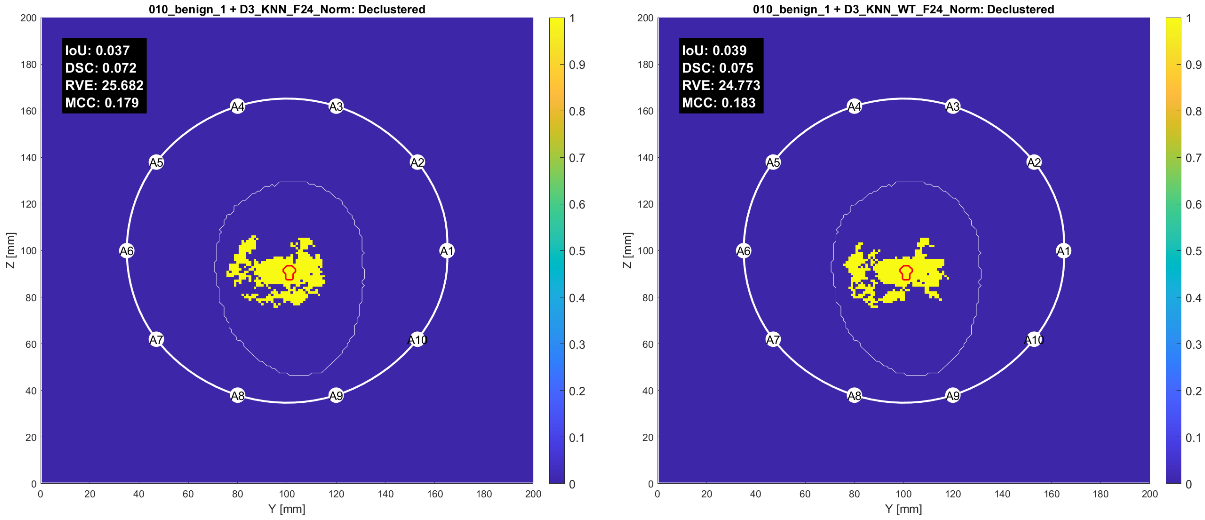
A.2 Extended Results for Chapter 5 – Results and Discussion

A.2.1 010_benign_1



(a) D1, WT→F24→Norm

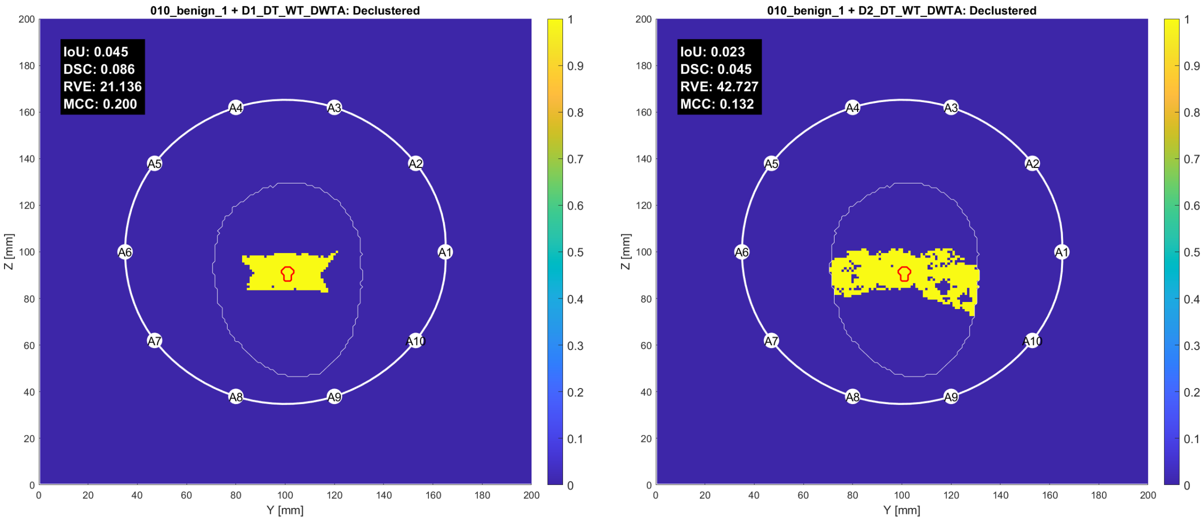
(b) D2, F24→Norm



(c) D3, F24→Norm

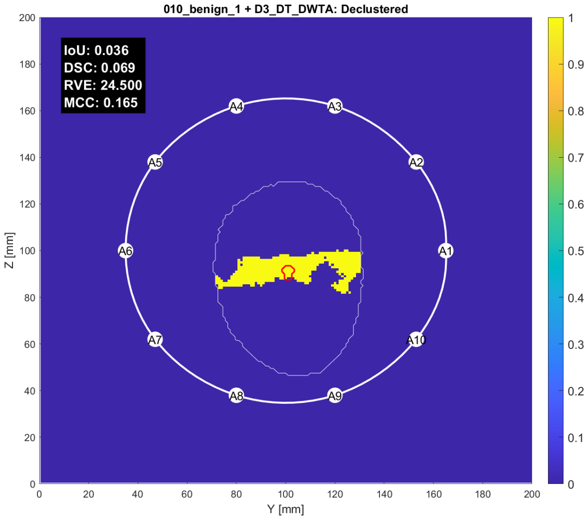
(d) D3, WT→F24→Norm

Figure A.3: Images reconstructed from simulation 010_benign_1 using the KNN algorithm: (a) D1, (b) D2, (c) D3, (d) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



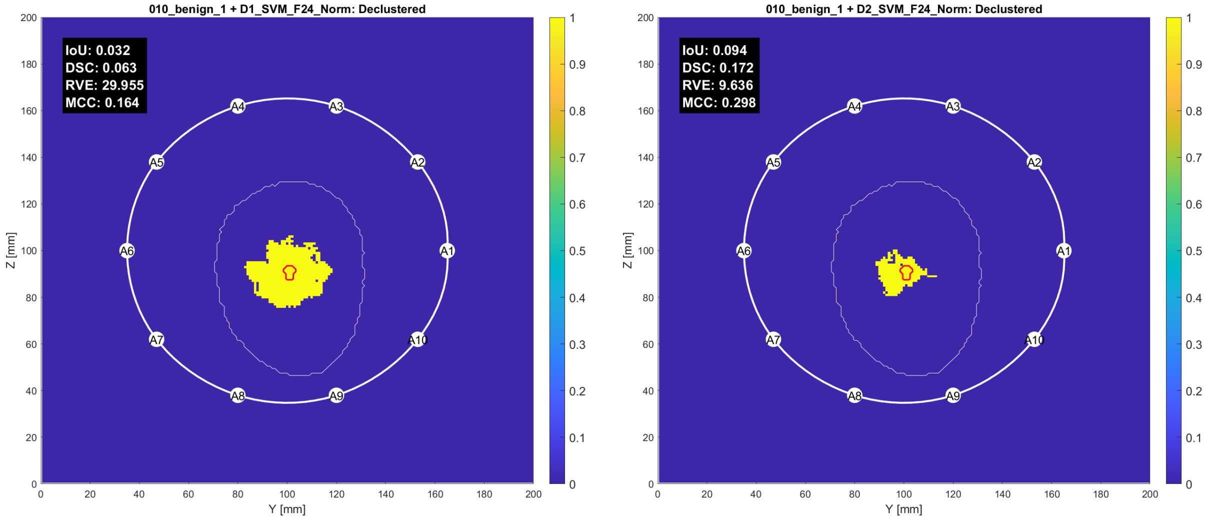
(a) D1, WT→DWTA

(b) D2, WT→DWTA



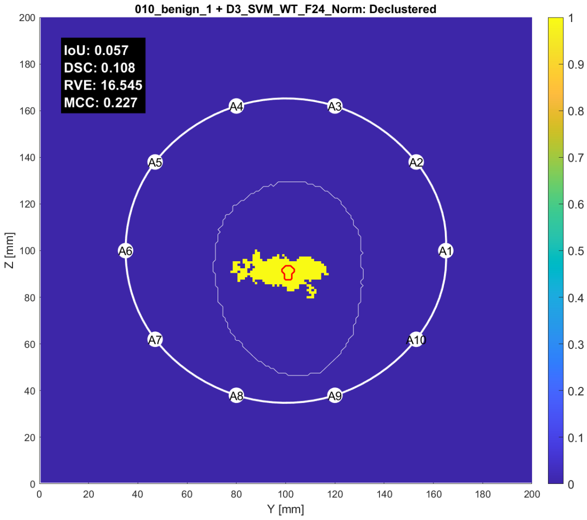
(c) D3, DWTA

Figure A.4: Images reconstructed from simulation 010_benign_1 using the DT algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



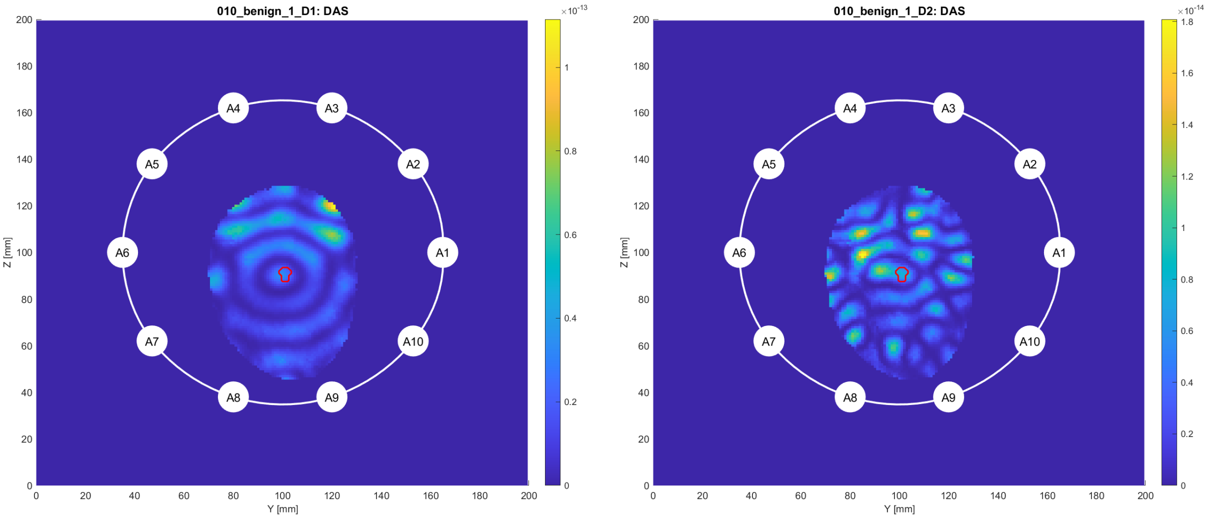
(a) D1, F24→Norm

(b) D2, F24→Norm



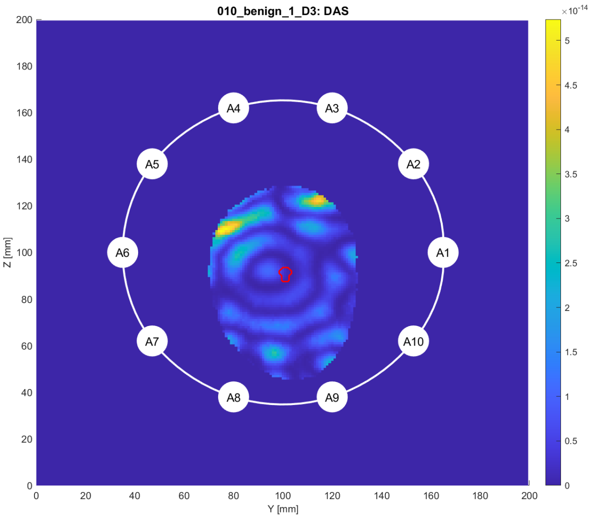
(c) D3, WT→F24→Norm

Figure A.5: Images reconstructed from simulation 010_benign_1 using the SVM algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



(a) D1

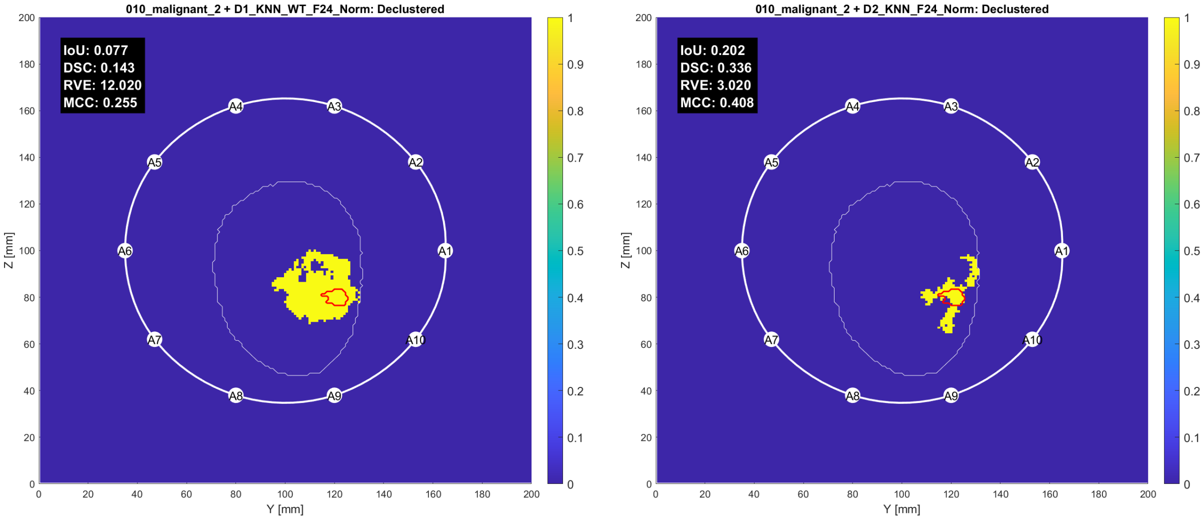
(b) D2



(c) D3

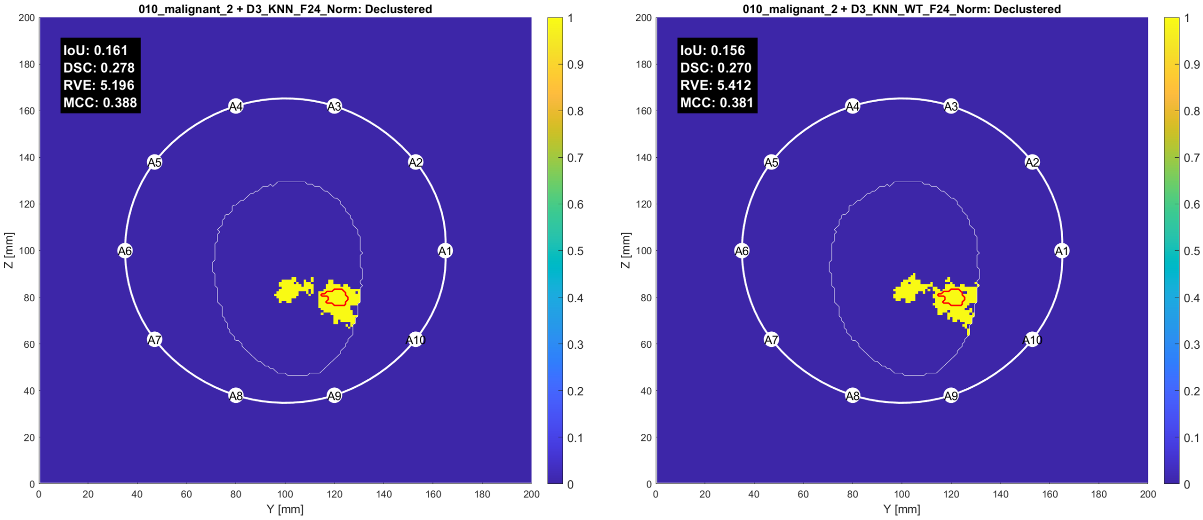
Figure A.6: Images reconstructed from simulation 010_benign_1 using the DAS algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

A.2.2 010_malignant_2



(a) D1, WT→F24→Norm

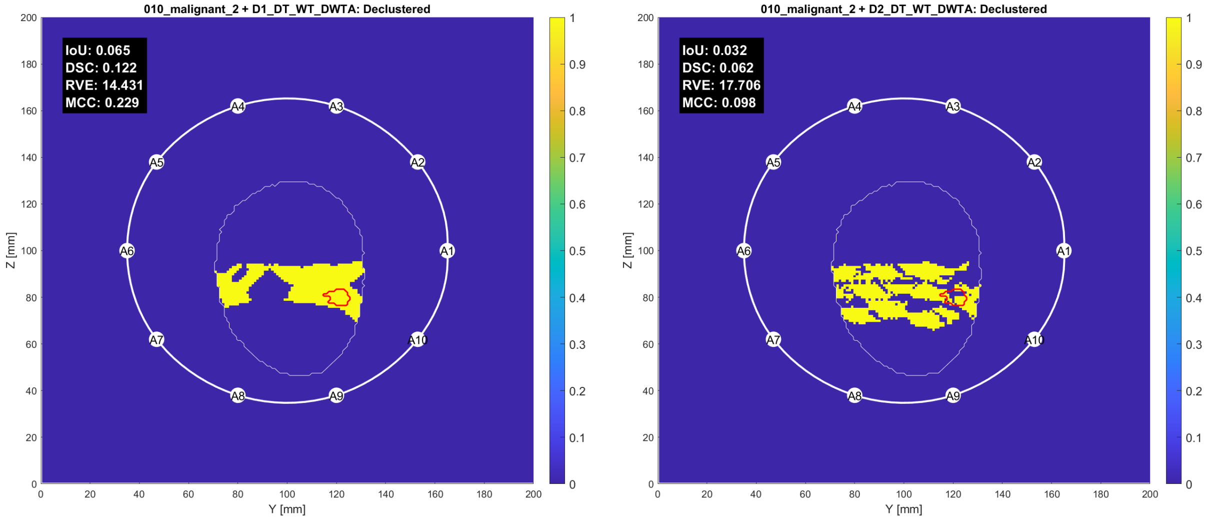
(b) D2, F24→Norm



(c) D3, F24→Norm

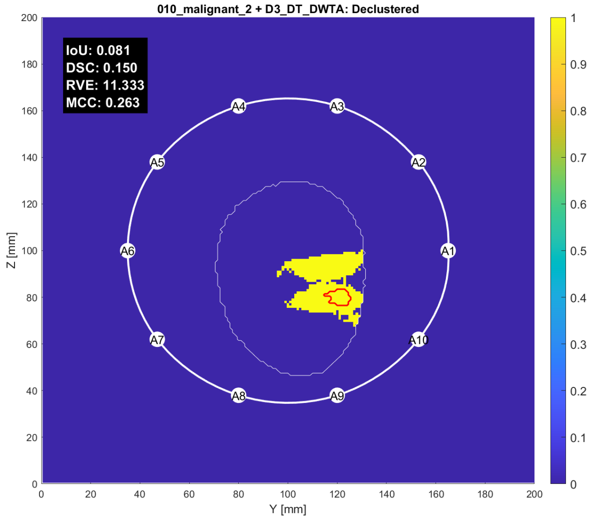
(d) D3, WT→F24→Norm

Figure A.7: Images reconstructed from simulation 010_malignant_2 using the KNN algorithm: (a) D1, (b) D2, (c) D3, (d) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



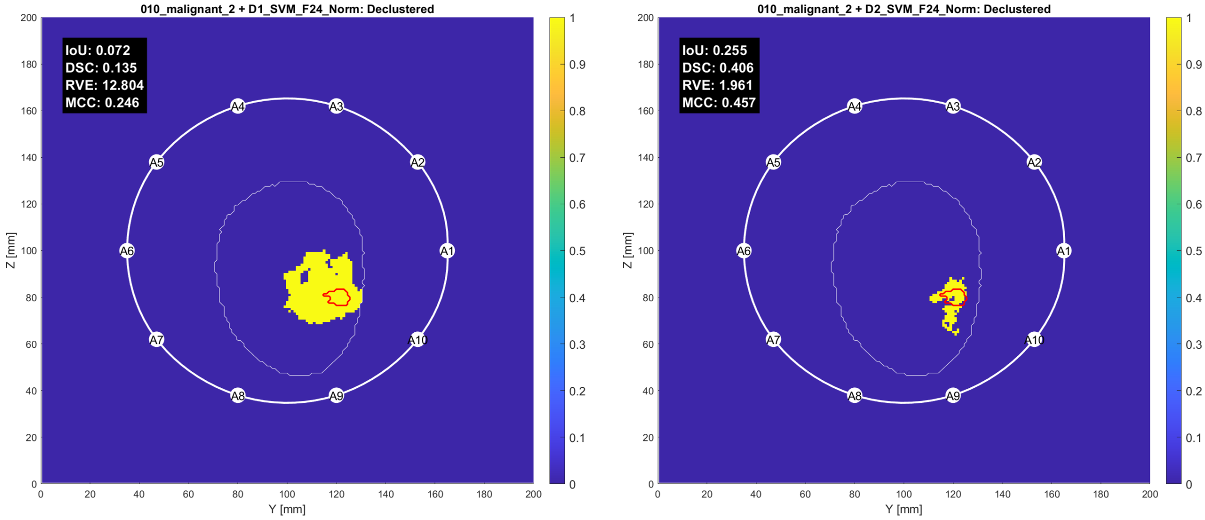
(a) D1, WT→DWTA

(b) D2, WT→DWTA



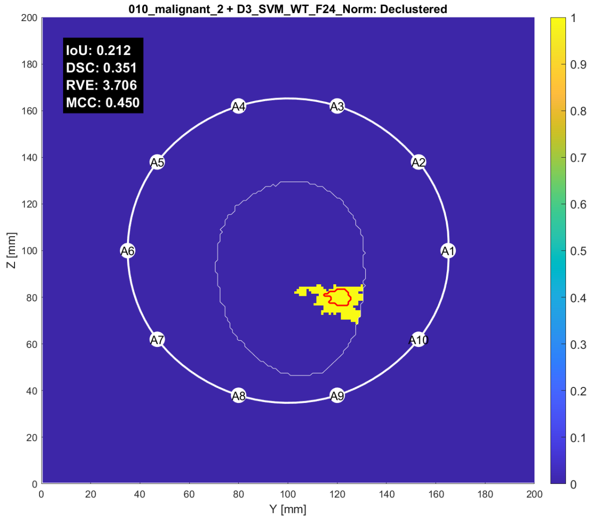
(c) D3, DWTA

Figure A.8: Images reconstructed from simulation 010_malignant_2 using the DT algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



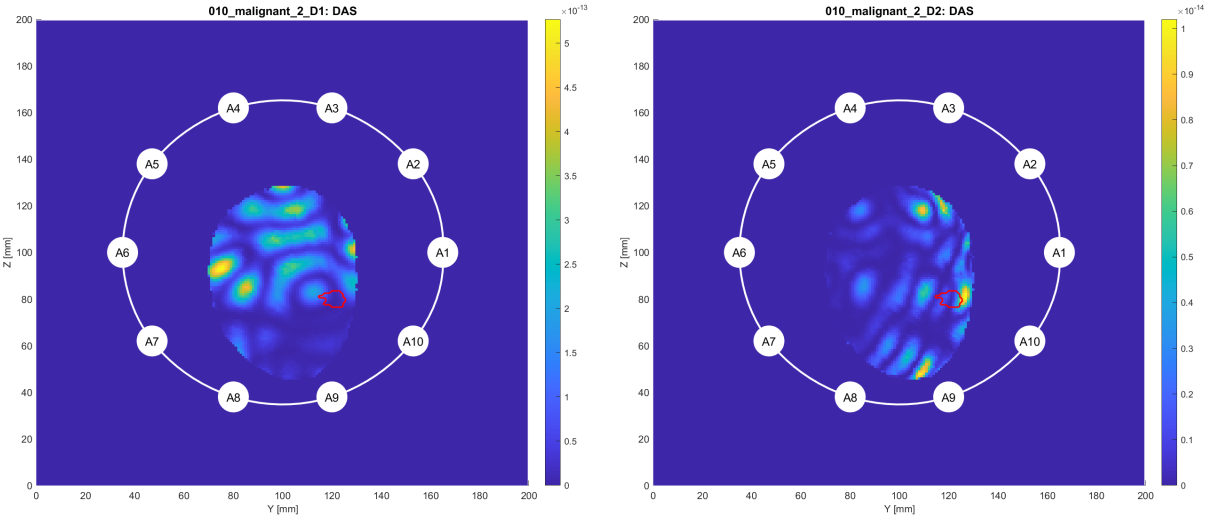
(a) D1, F24→Norm

(b) D2, F24→Norm

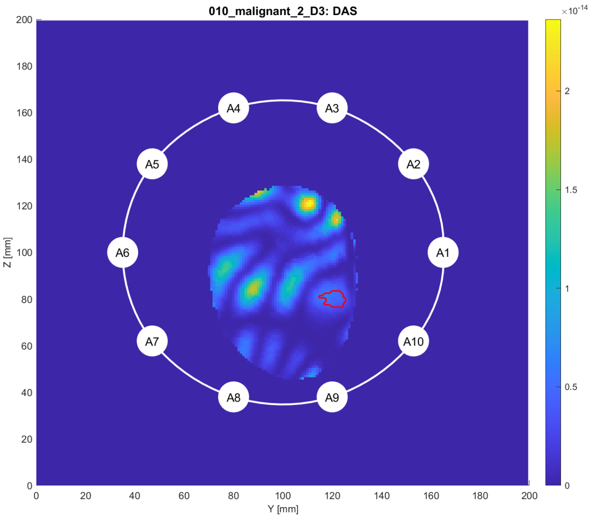


(c) D3, WT→F24→Norm

Figure A.9: Images reconstructed from simulation 010_malignant_2 using the SVM algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



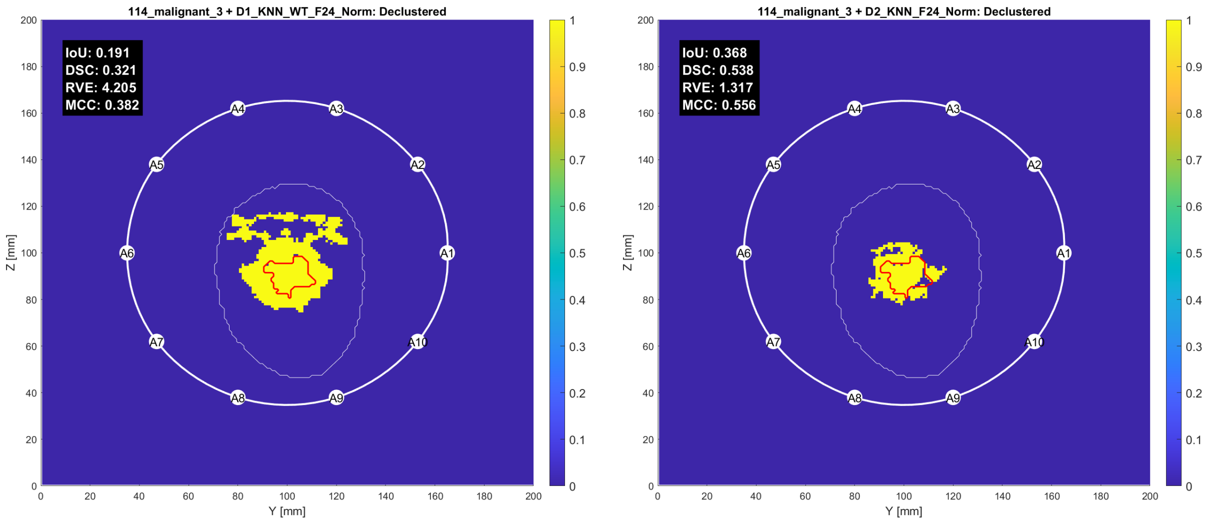
(a) D1 (b) D2



(c) D3

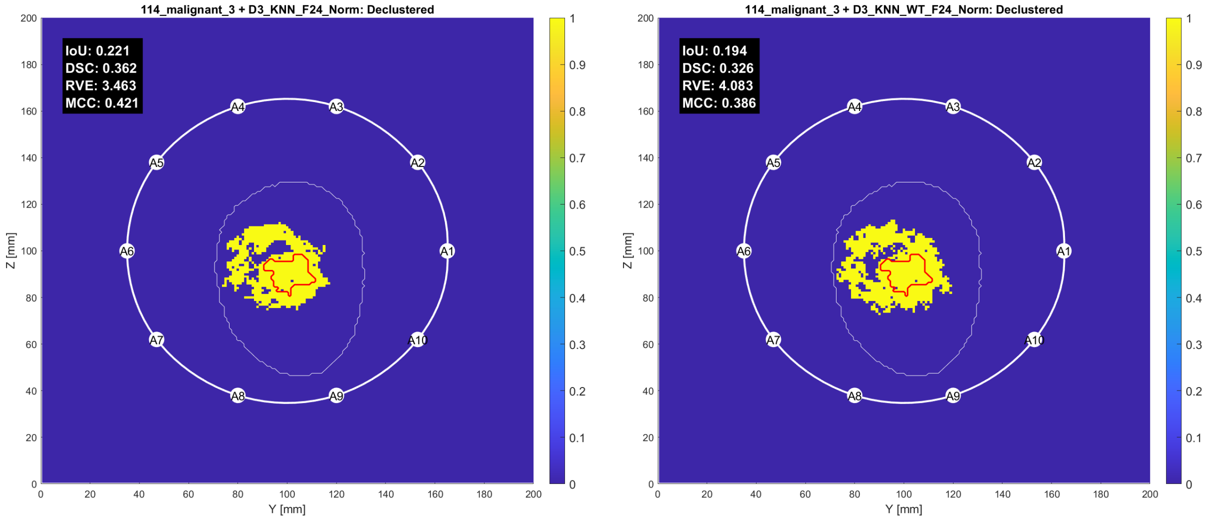
Figure A.10: Images reconstructed from simulation 010_malignant_2 using the DAS algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

A.2.3 114_malignant_3



(a) D1, WT→F24→Norm

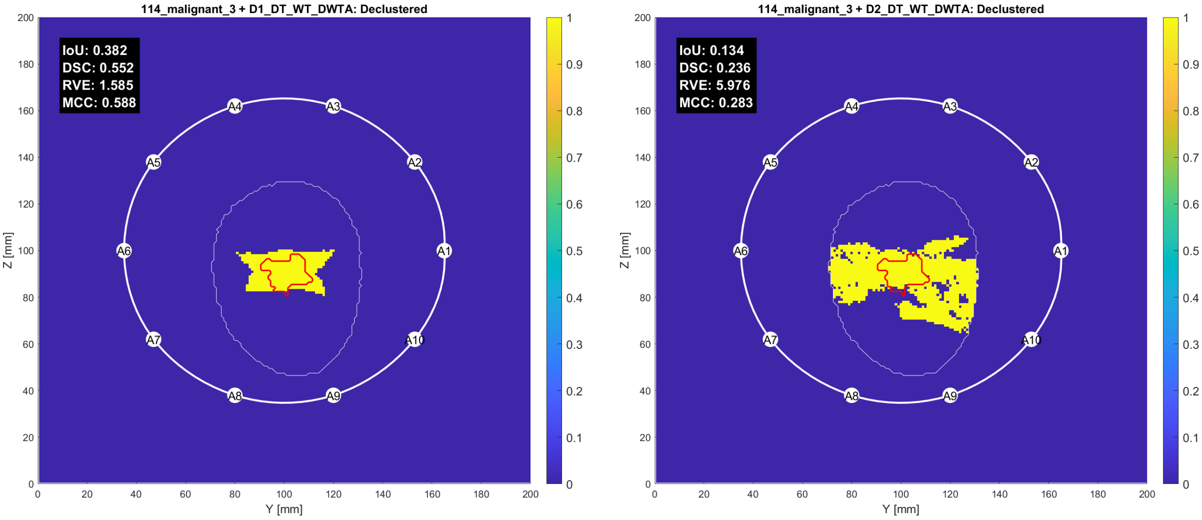
(b) D2, F24→Norm



(c) D3, F24→Norm

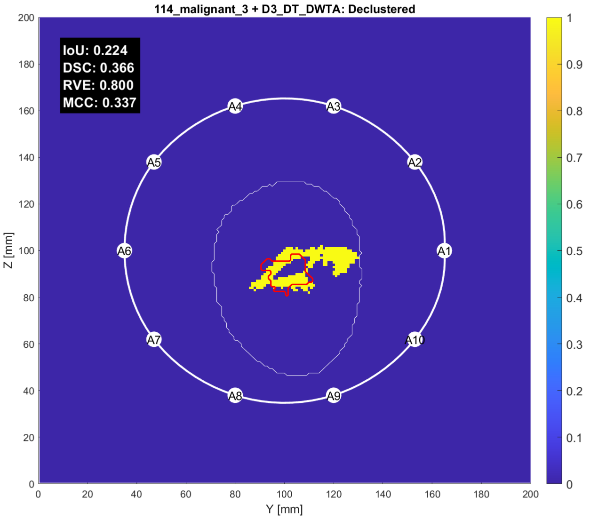
(d) D3, WT→F24→Norm

Figure A.11: Images reconstructed from simulation 114_malignant_3 using the KNN algorithm: (a) D1, (b) D2, (c) D3, (d) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



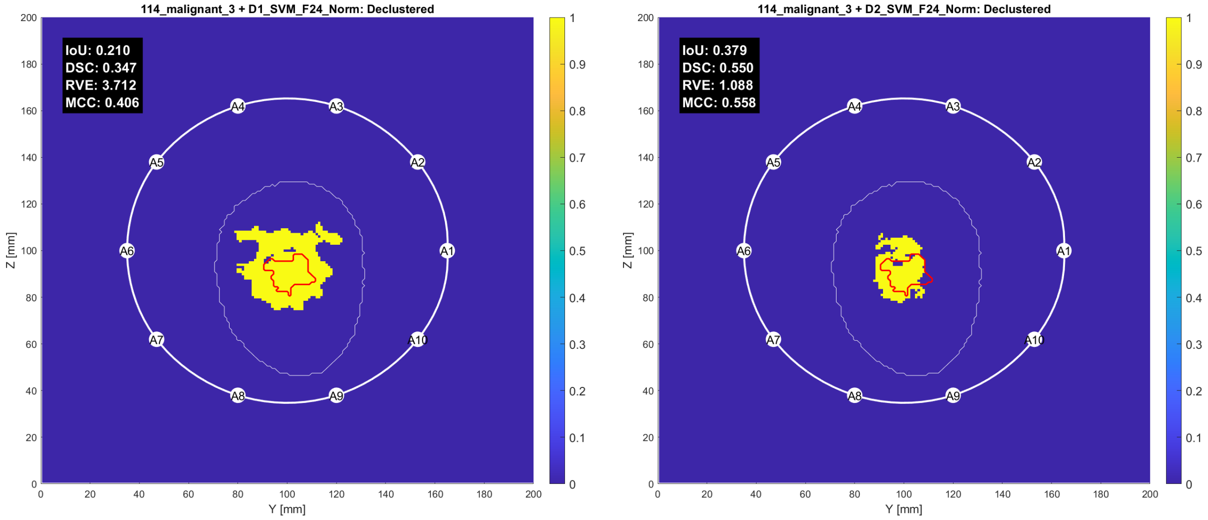
(a) D1, WT→DWTA

(b) D2, WT→DWTA



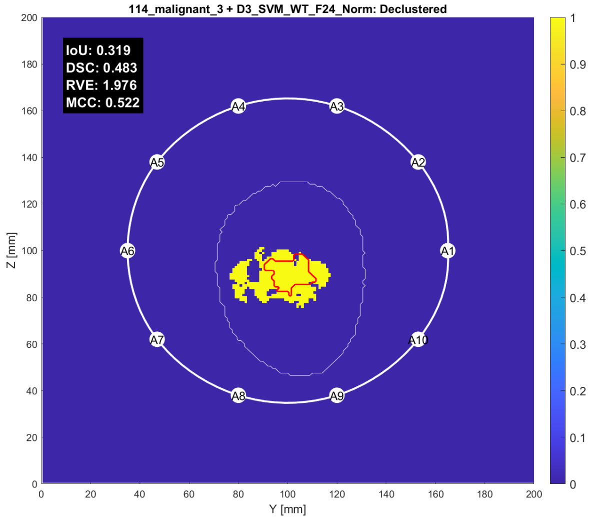
(c) D3, DWTA

Figure A.12: Images reconstructed from simulation 114_malignant_3 using the DT algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



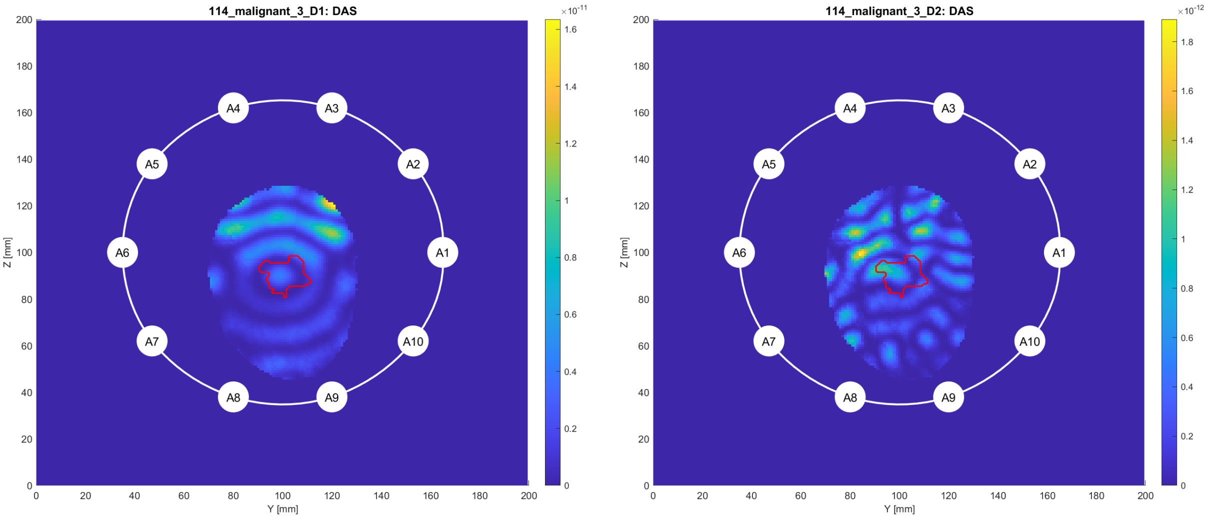
(a) D1, F24→Norm

(b) D2, F24→Norm

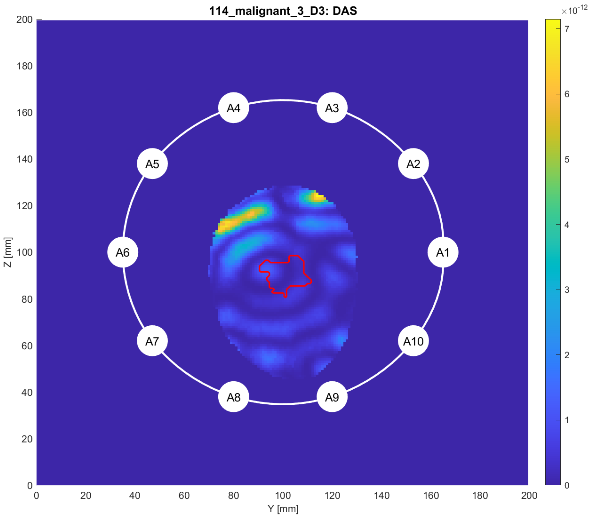


(c) D3, WT→F24→Norm

Figure A.13: Images reconstructed from simulation 114_malignant_3 using the SVM algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



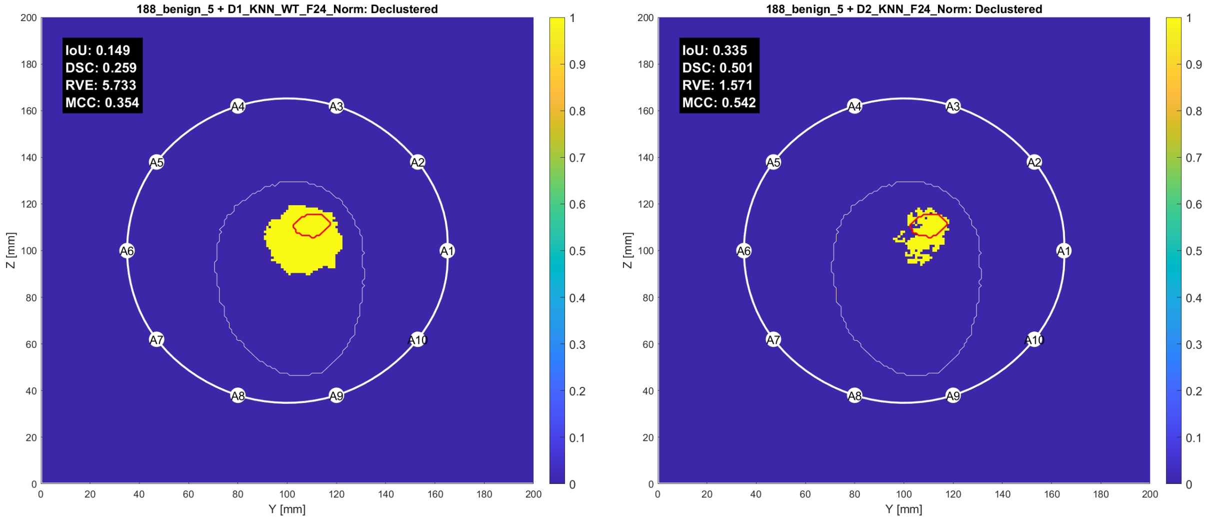
(a) D1 (b) D2



(c) D3

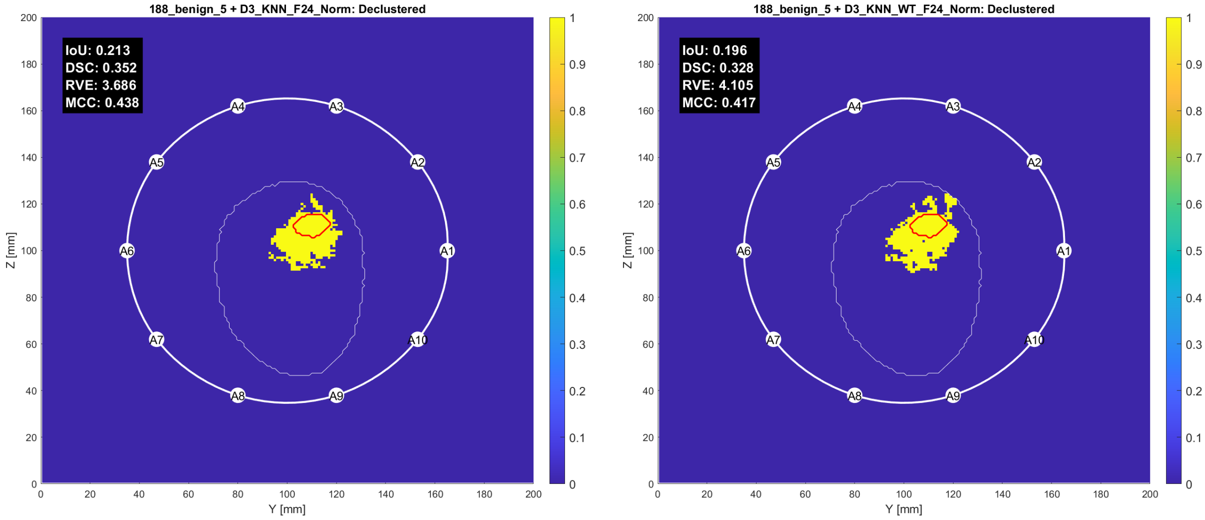
Figure A.14: Images reconstructed from simulation 114_malignant_3 using the DAS algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

A.2.4 188_benign_5



(a) D1, WT→F24→Norm

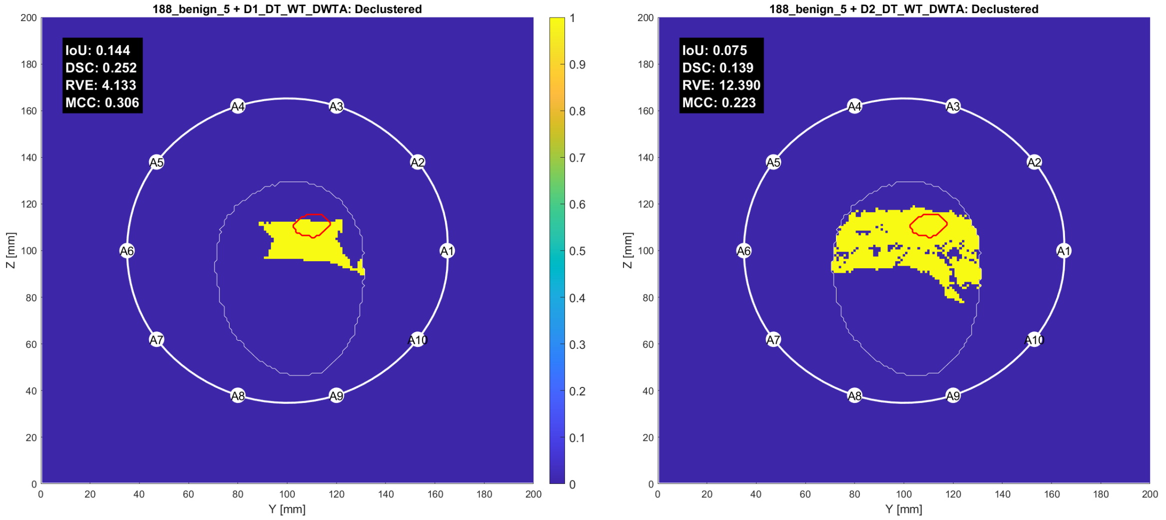
(b) D2, F24→Norm



(c) D3, F24→Norm

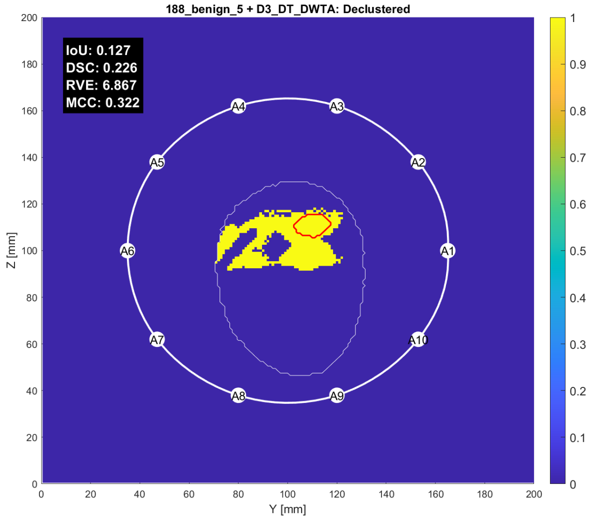
(d) D3, WT→F24→Norm

Figure A.15: Images reconstructed from simulation 188_benign_5 using the KNN algorithm: (a) D1, (b) D2, (c) D3, (d) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



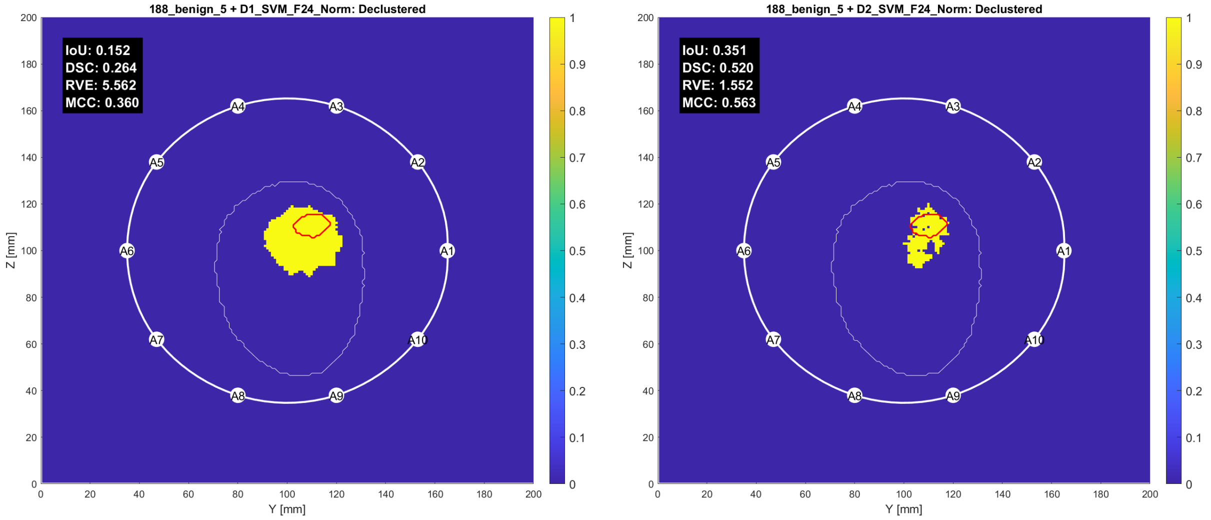
(a) D1, WT→DWTA

(b) D2, WT→DWTA



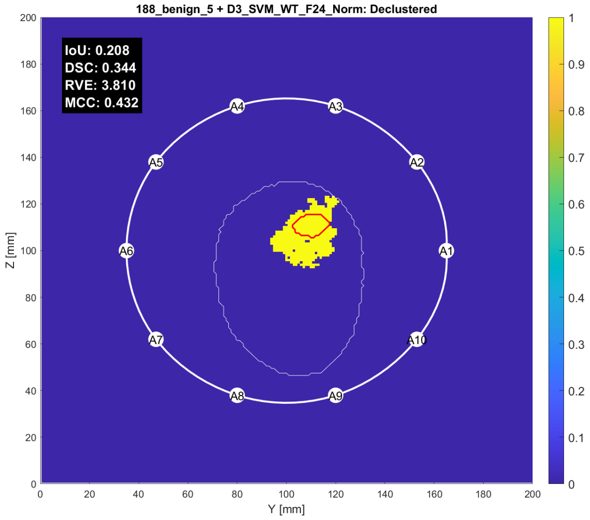
(c) D3, DWTA

Figure A.16: Images reconstructed from simulation 188_benign_5 using the DT algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



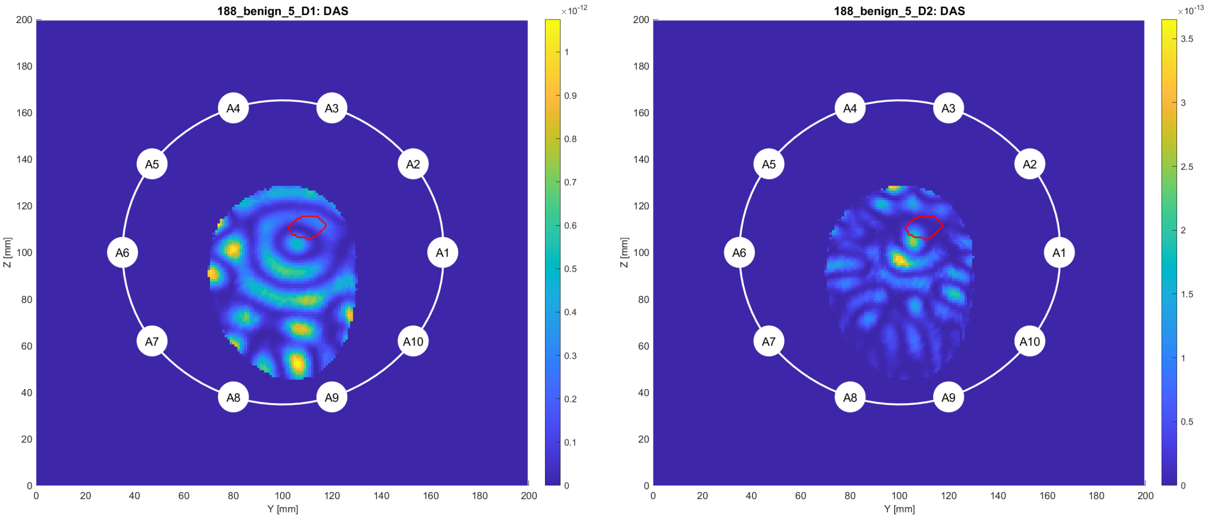
(a) D1, F24→Norm

(b) D2, F24→Norm



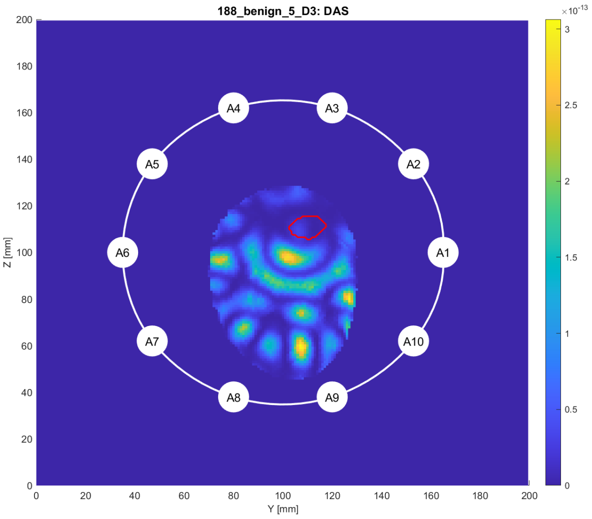
(c) D3, WT→F24→Norm

Figure A.17: Images reconstructed from simulation 188_benign_5 using the SVM algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



(a) D1

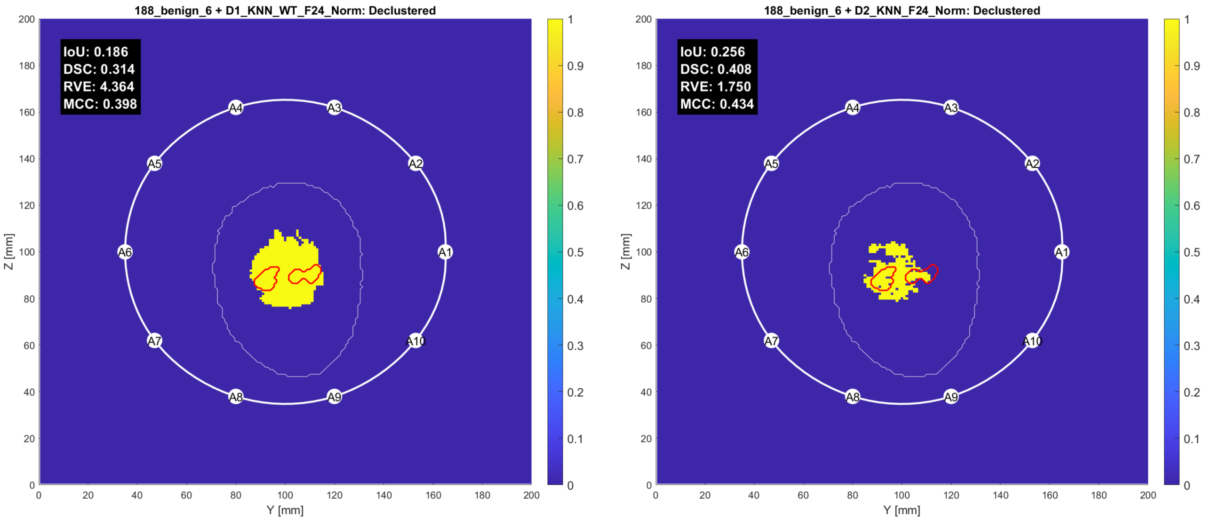
(b) D2



(c) D3

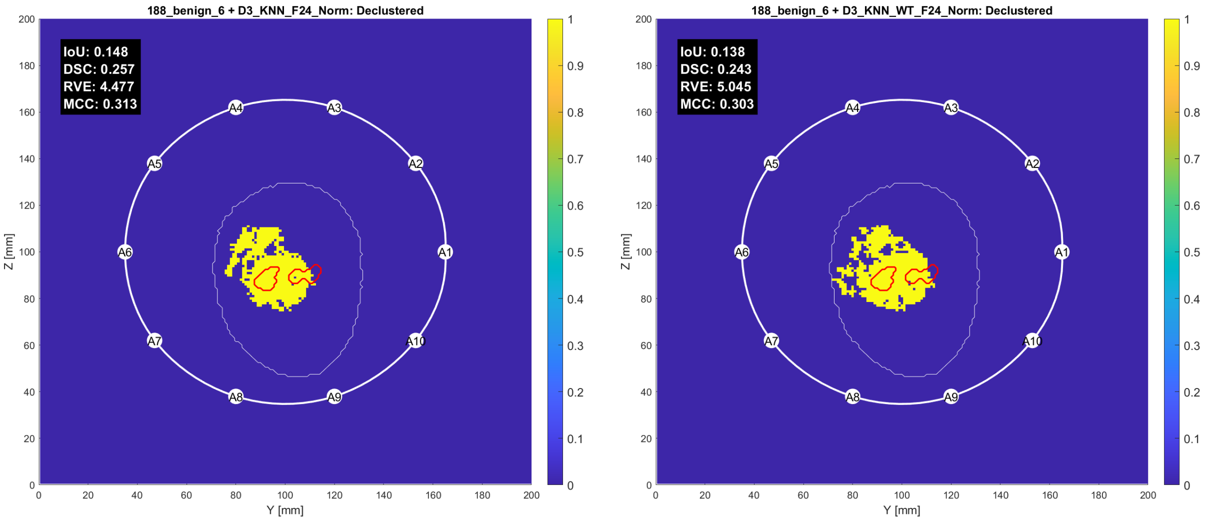
Figure A.18: Images reconstructed from simulation 188_benign_5 using the DAS algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

A.2.5 188_benign_6



(a) D1, WT→F24→Norm

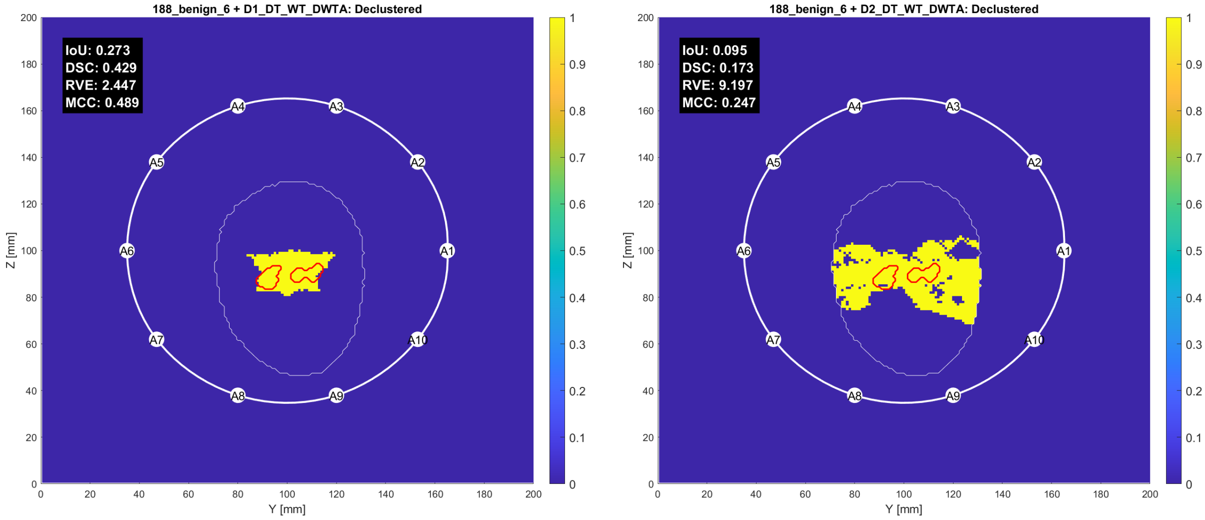
(b) D2, F24→Norm



(c) D3, F24→Norm

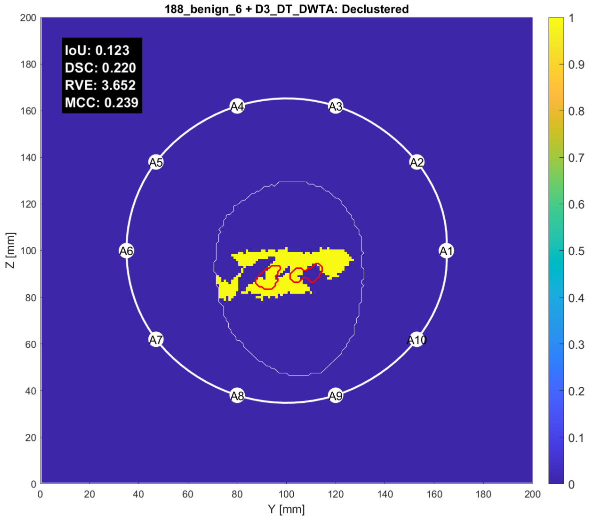
(d) D3, WT→F24→Norm

Figure A.19: Images reconstructed from simulation 188_benign_6 using the KNN algorithm: (a) D1, (b) D2, (c) D3, (d) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



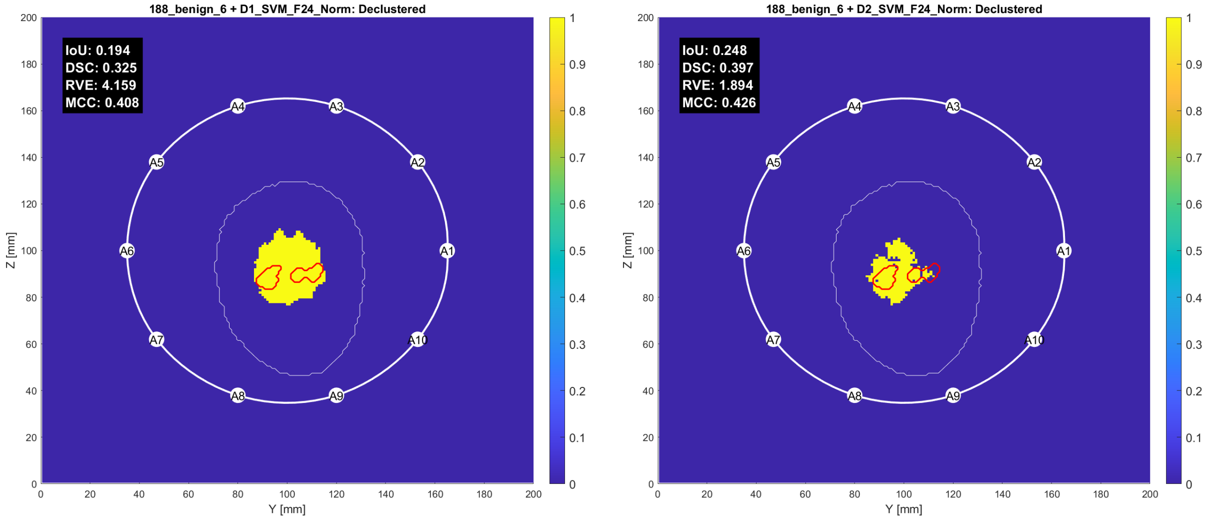
(a) D1, WT→DWTA

(b) D2, WT→DWTA



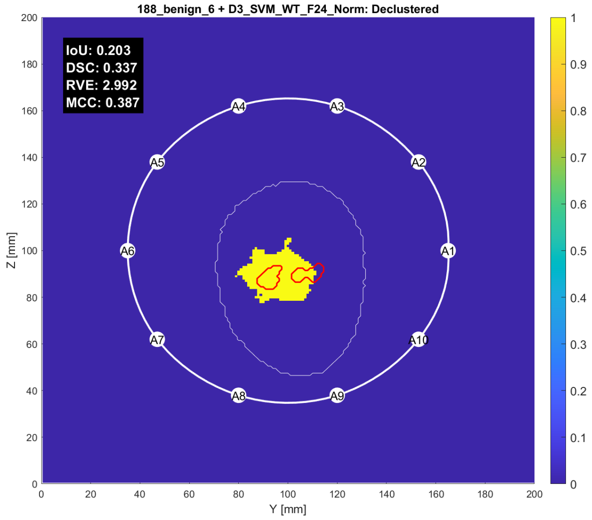
(c) D3, DWTA

Figure A.20: Images reconstructed from simulation 188_benign_6 using the DT algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



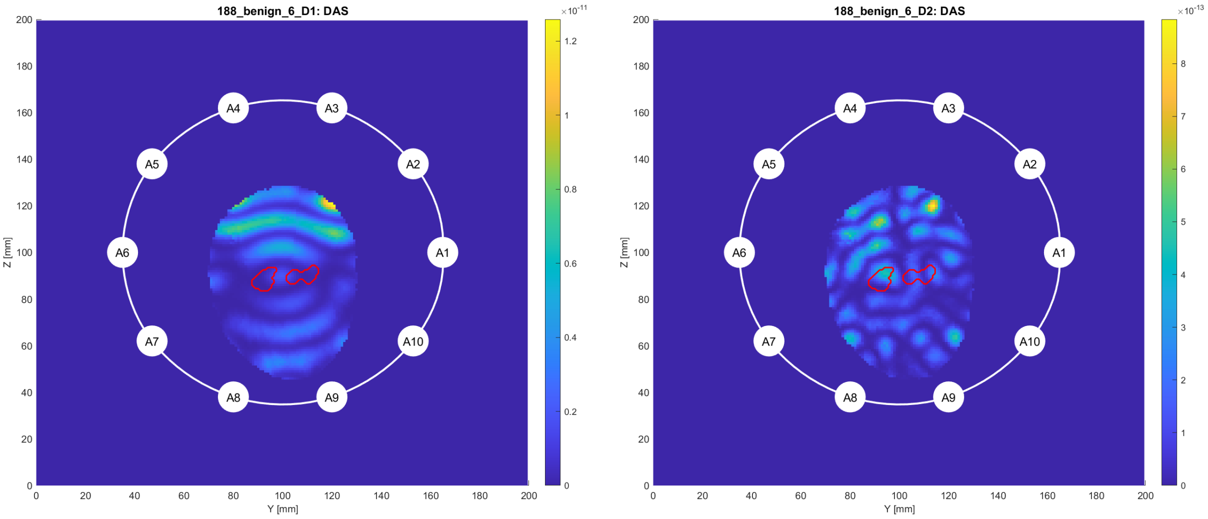
(a) D1, F24→Norm

(b) D2, F24→Norm



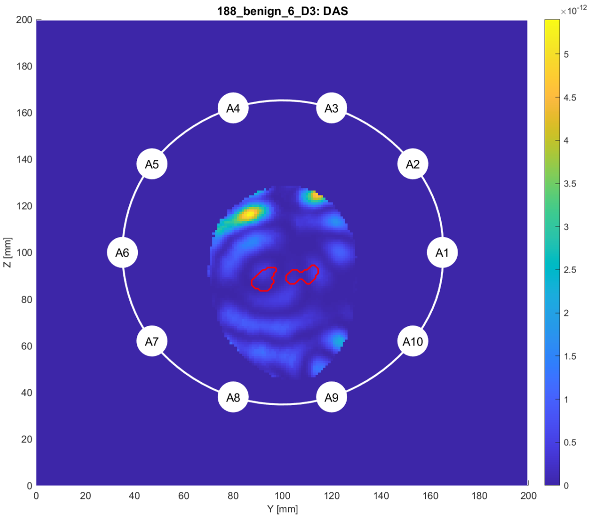
(c) D3, WT→F24→Norm

Figure A.21: Images reconstructed from simulation 188_benign_6 using the SVM algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



(a) D1

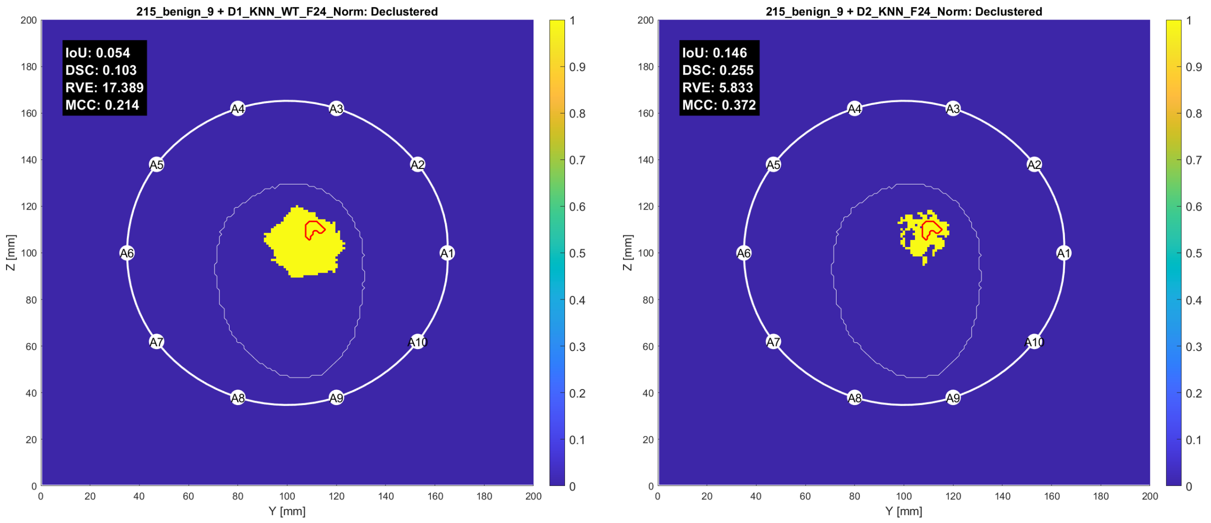
(b) D2



(c) D3

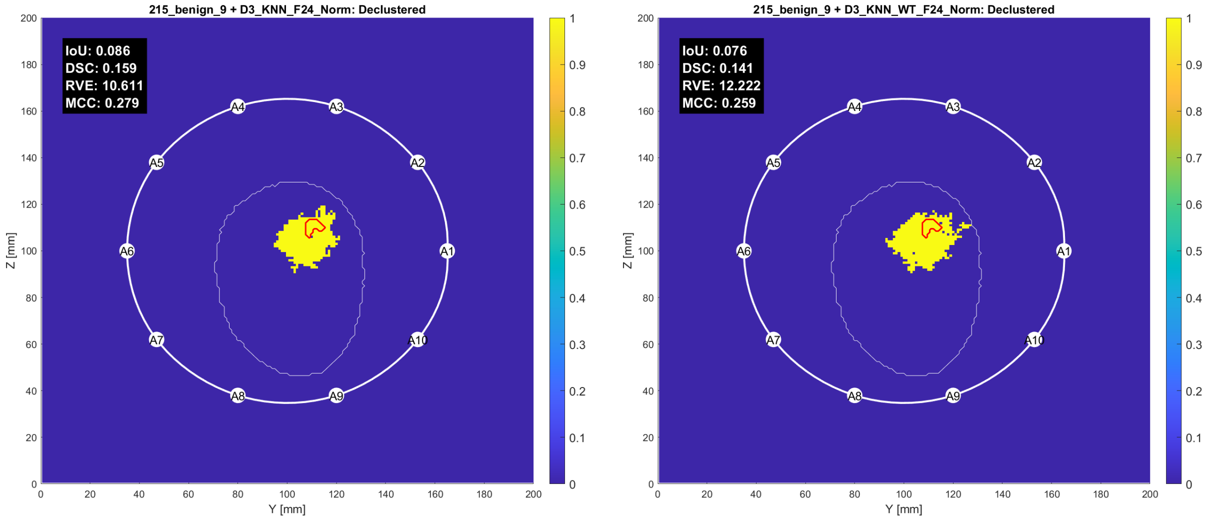
Figure A.22: Images reconstructed from simulation 188_benign_6 using the DAS algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

A.2.6 215_benign_9



(a) D1, WT→F24→Norm

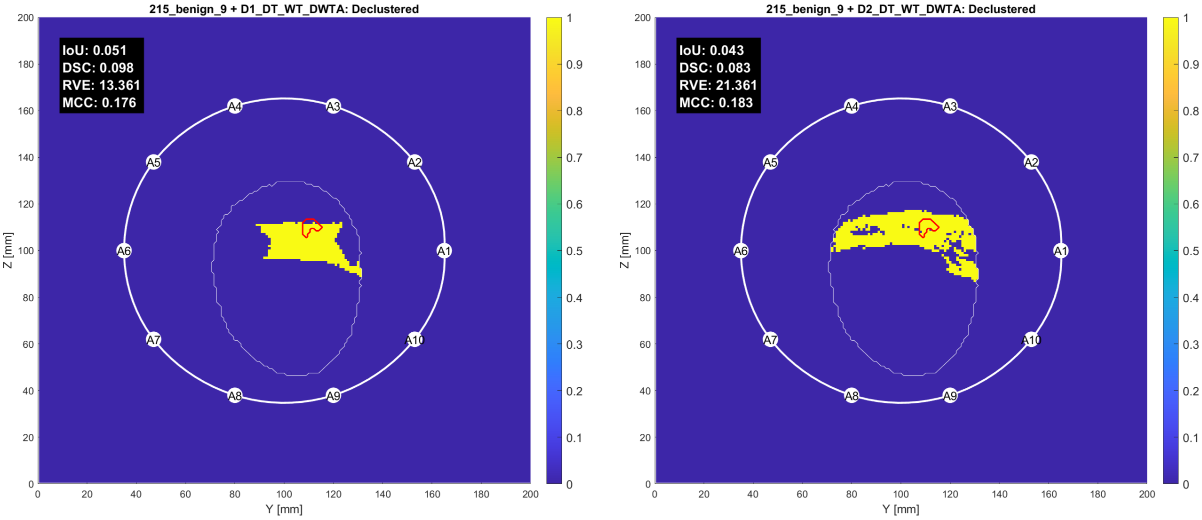
(b) D2, F24→Norm



(c) D3, F24→Norm

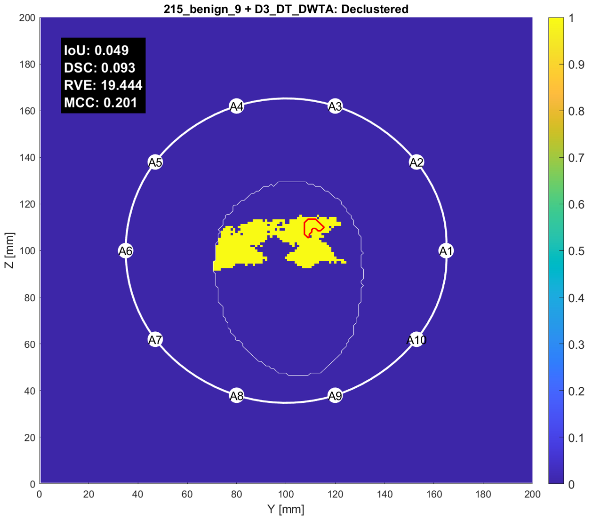
(d) D3, WT→F24→Norm

Figure A.23: Images reconstructed from simulation 215_benign_9 using the KNN algorithm: (a) D1, (b) D2, (c) D3, (d) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



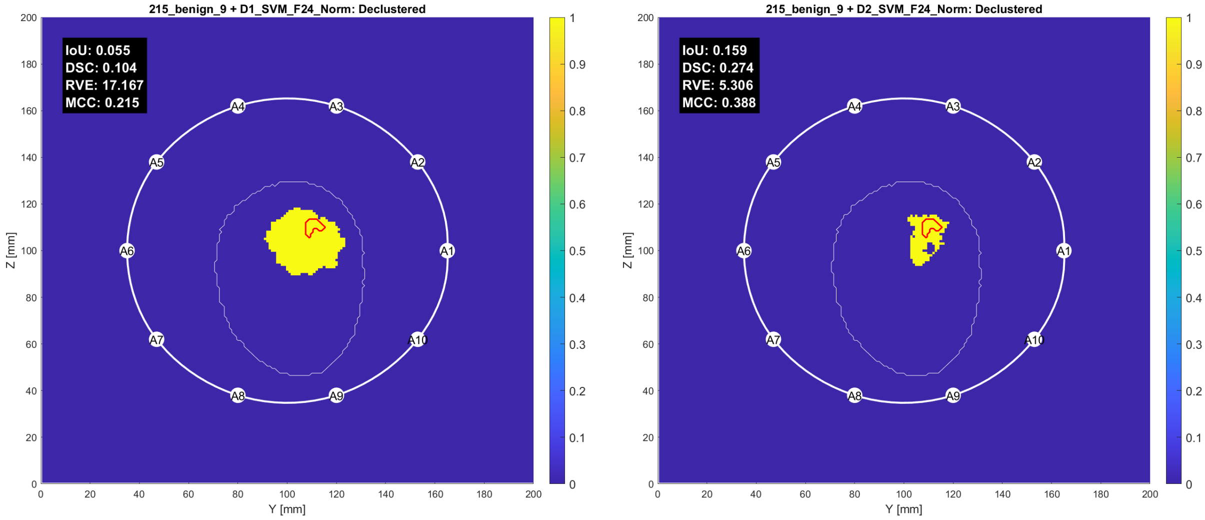
(a) D1, WT→DWTA

(b) D2, WT→DWTA



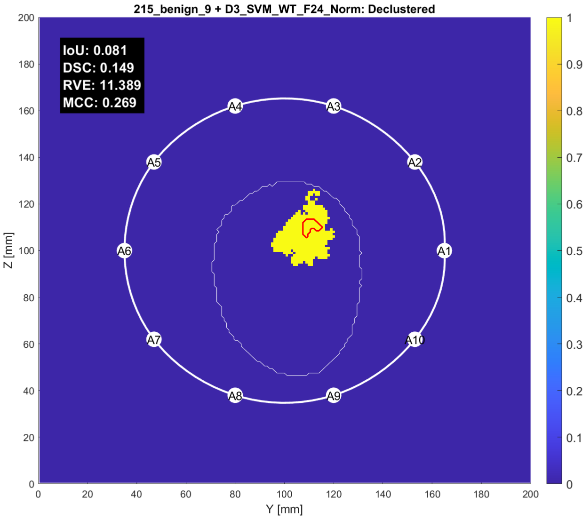
(c) D3, DWTA

Figure A.24: Images reconstructed from simulation 215_benign_9 using the DT algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



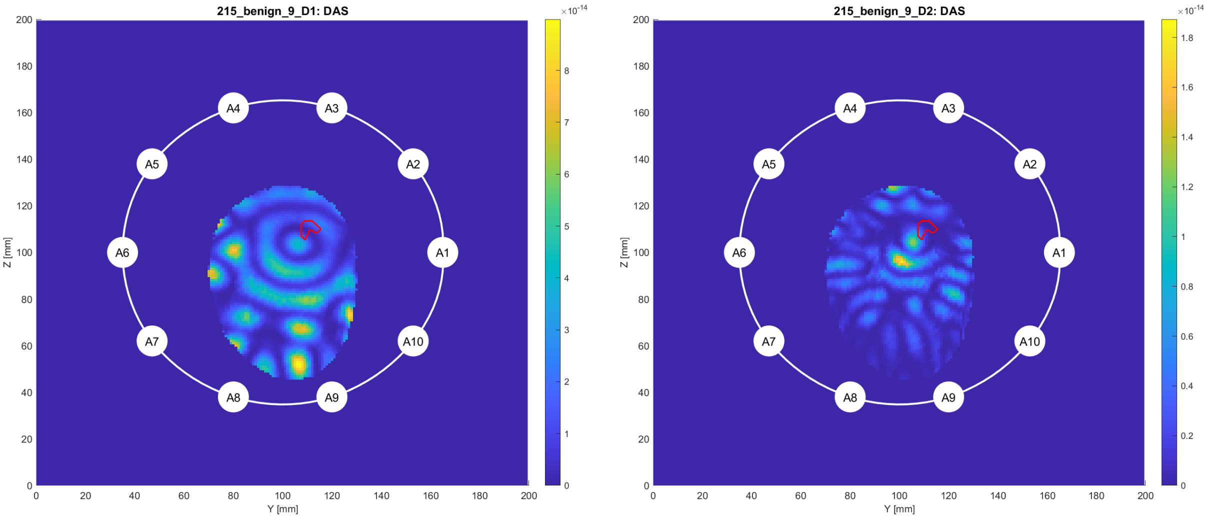
(a) D1, F24→Norm

(b) D2, F24→Norm



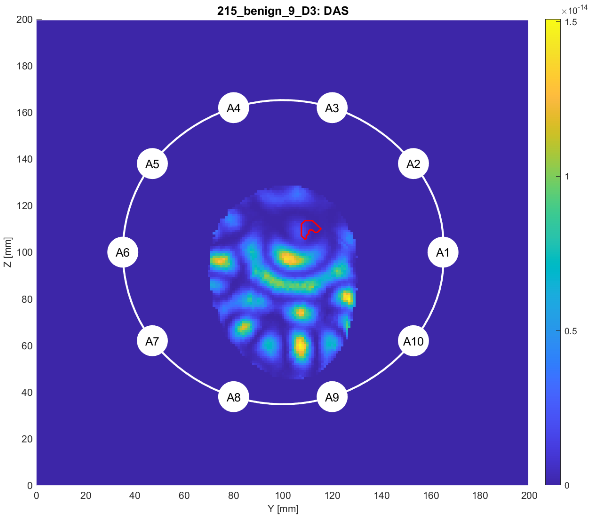
(c) D3, WT→F24→Norm

Figure A.25: Images reconstructed from simulation 215_benign_9 using the SVM algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



(a) D1

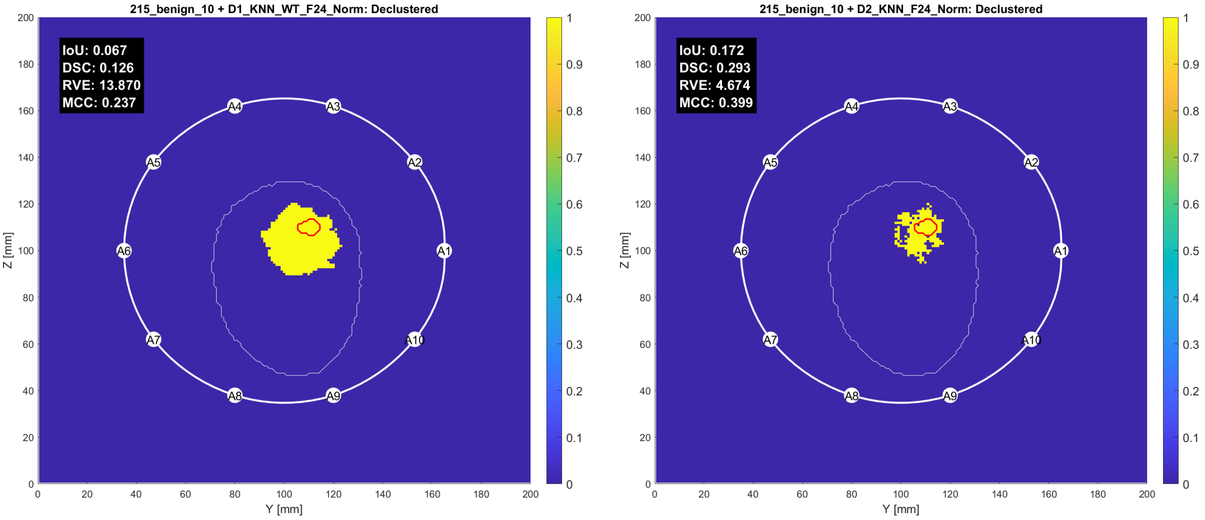
(b) D2



(c) D3

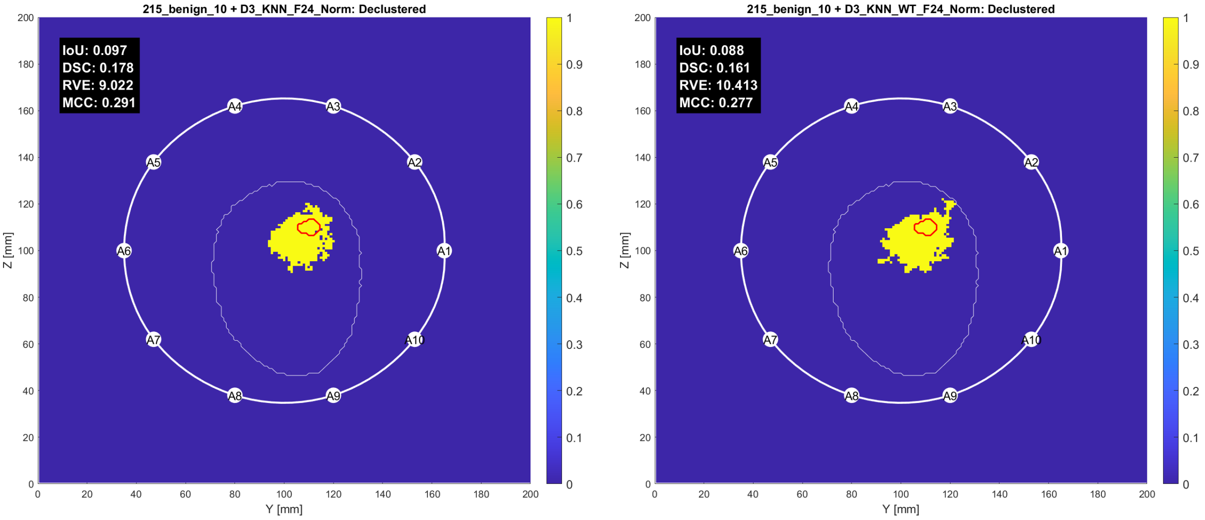
Figure A.26: Images reconstructed from simulation 215_benign_9 using the DAS algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

A.2.7 215_benign_10



(a) D1, WT→F24→Norm

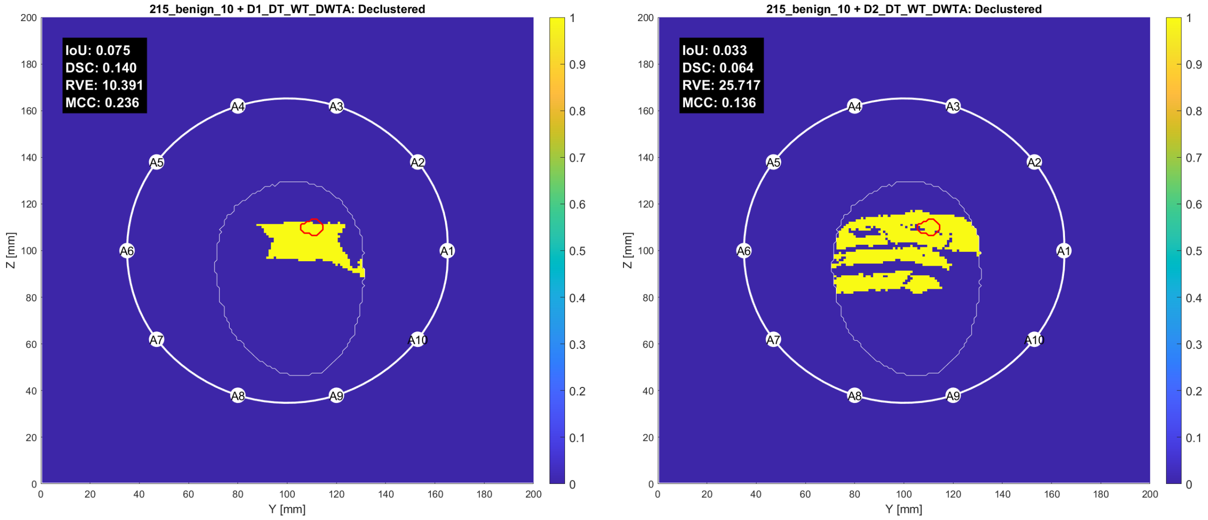
(b) D2, F24→Norm



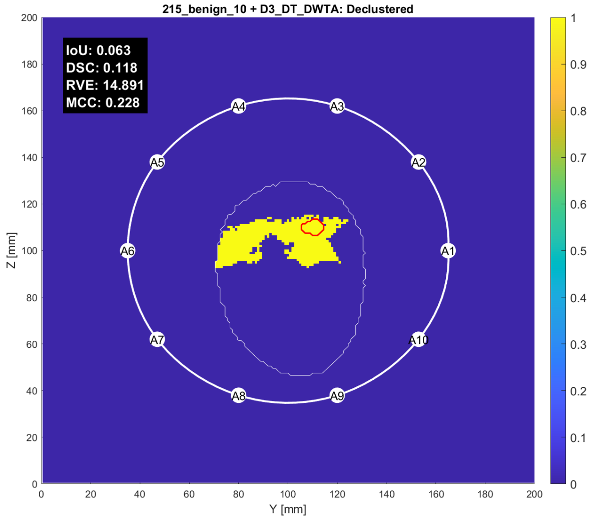
(c) D3, F24→Norm

(d) D3, WT→F24→Norm

Figure A.27: Images reconstructed from simulation 215_benign_10 using the KNN algorithm: (a) D1, (b) D2, (c) D3, (d) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

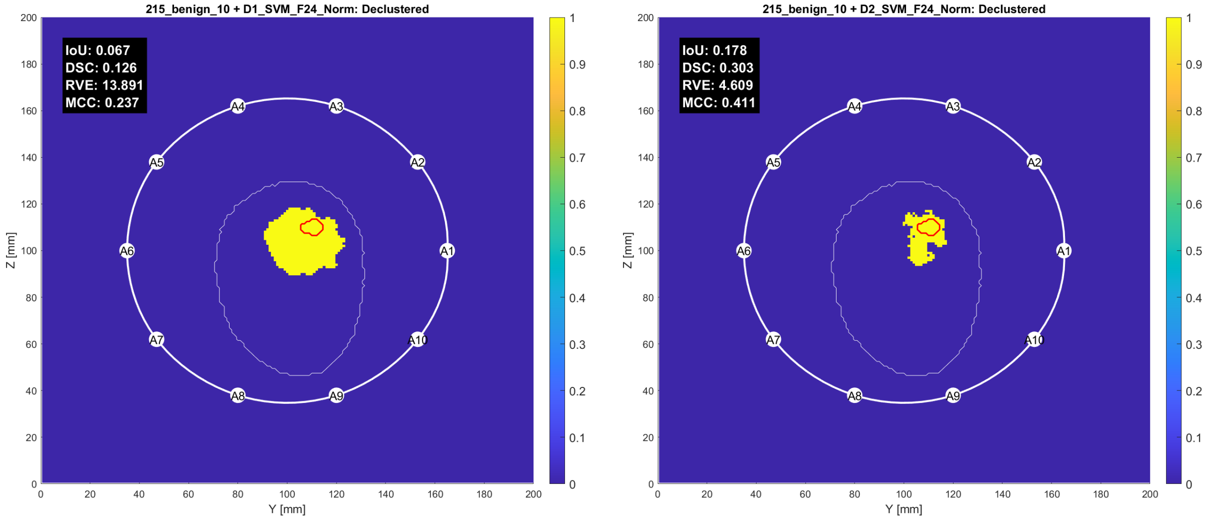


(a) D1, WT→DWTA (b) D2, WT→DWTA



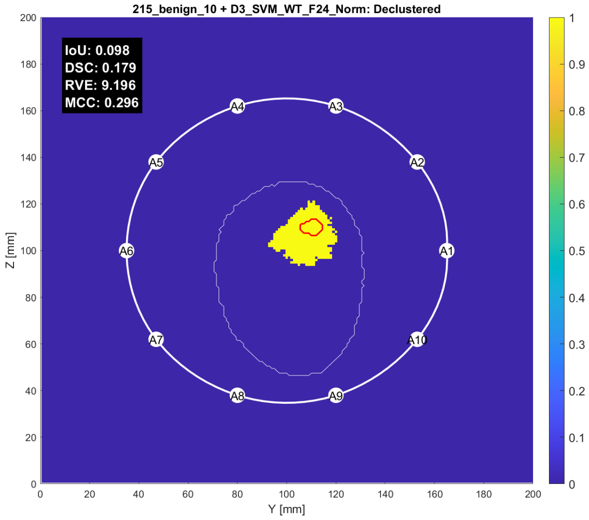
(c) D3, DWTA

Figure A.28: Images reconstructed from simulation 215_benign_10 using the DT algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



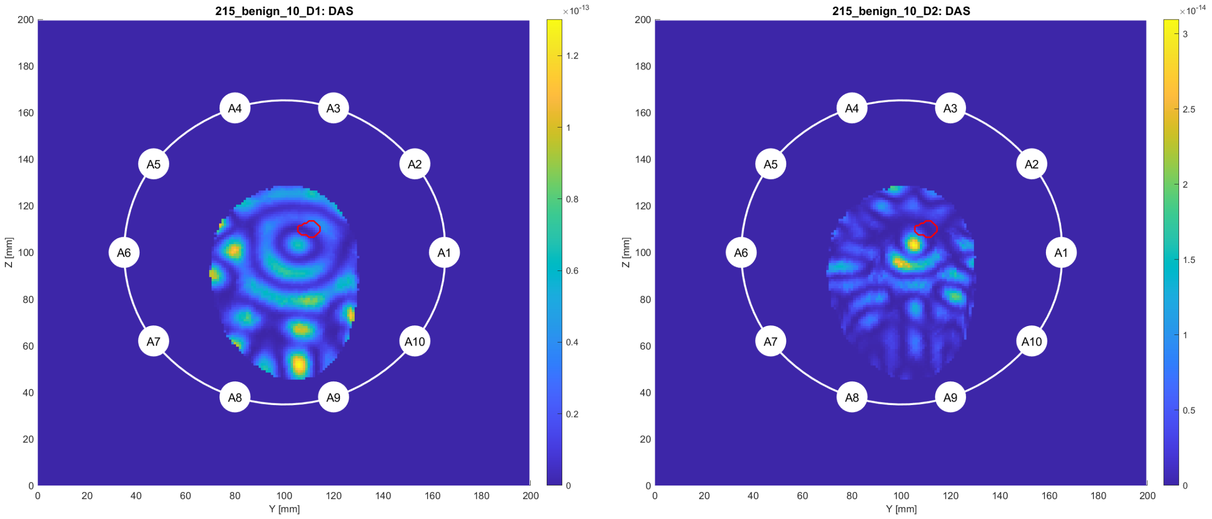
(a) D1, F24→Norm

(b) D2, F24→Norm

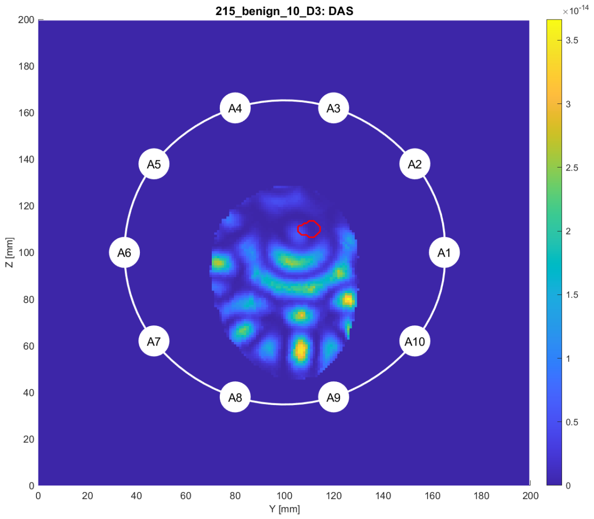


(c) D3, WT→F24→Norm

Figure A.29: Images reconstructed from simulation 215_benign_10 using the SVM algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



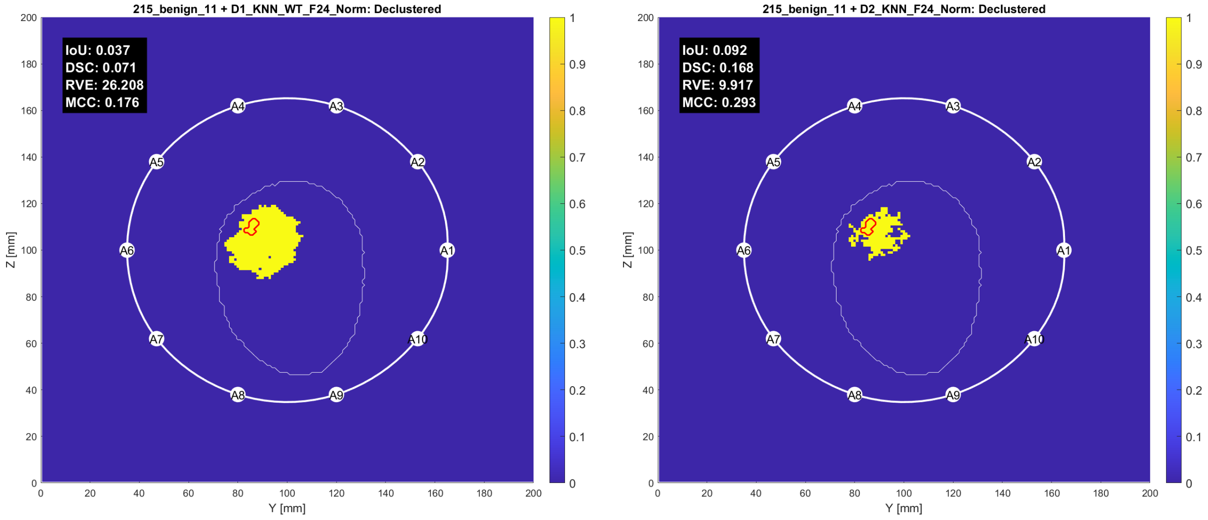
(a) D1 (b) D2



(c) D3

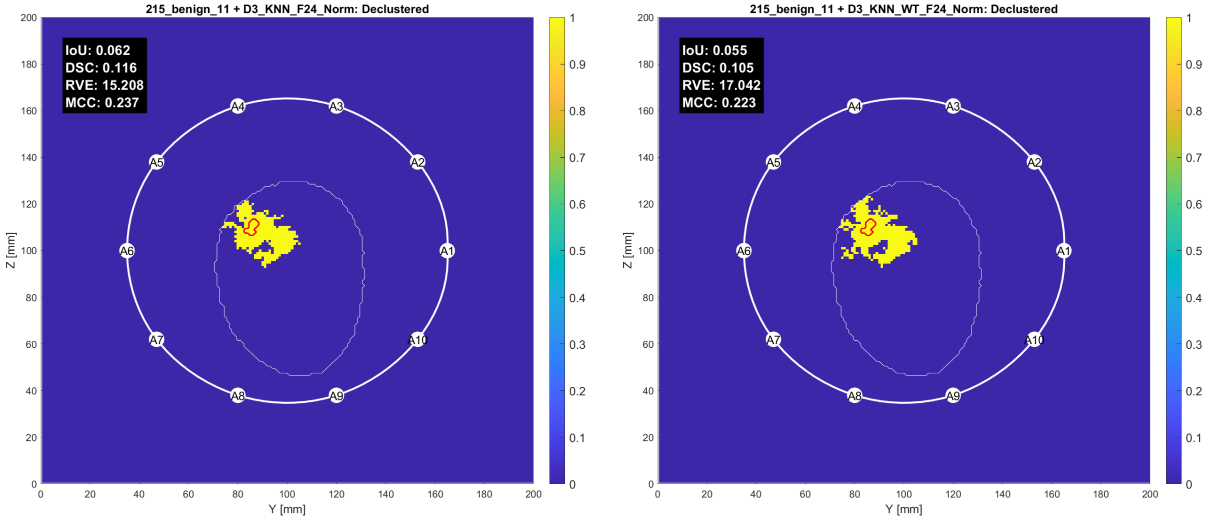
Figure A.30: Images reconstructed from simulation 215_benign_10 using the DAS algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

A.2.8 215_benign_11



(a) D1, WT→F24→Norm

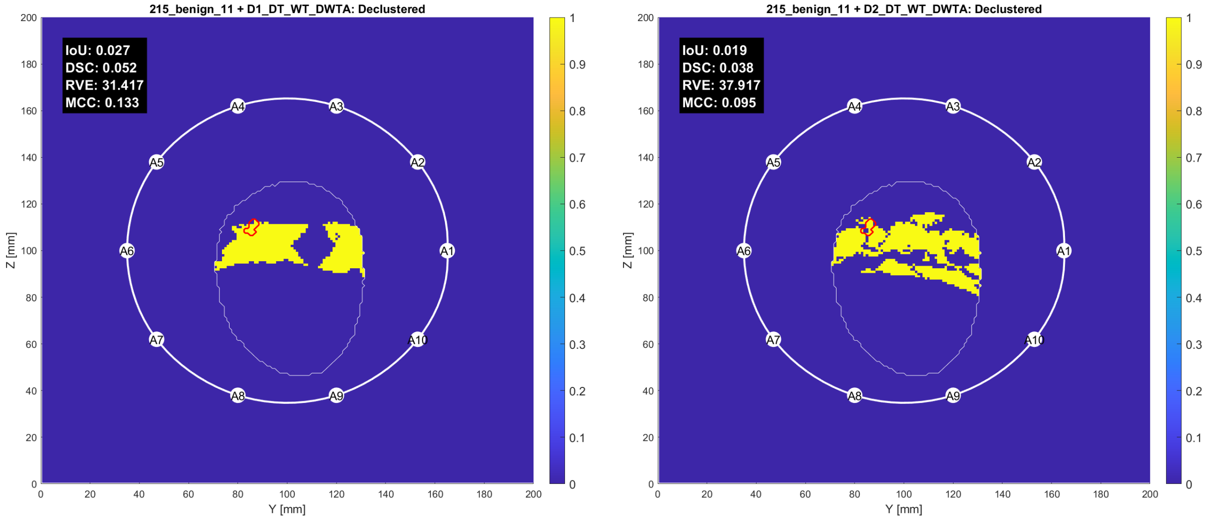
(b) D2, F24→Norm



(c) D3, F24→Norm

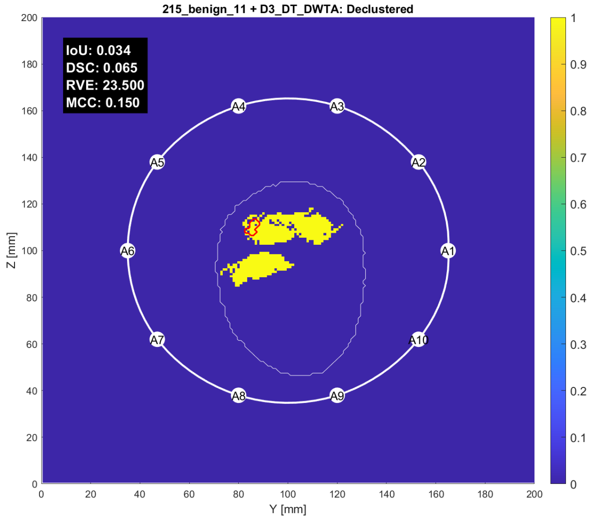
(d) D3, WT→F24→Norm

Figure A.31: Images reconstructed from simulation 215_benign_11 using the KNN algorithm: (a) D1, (b) D2, (c) D3, (d) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



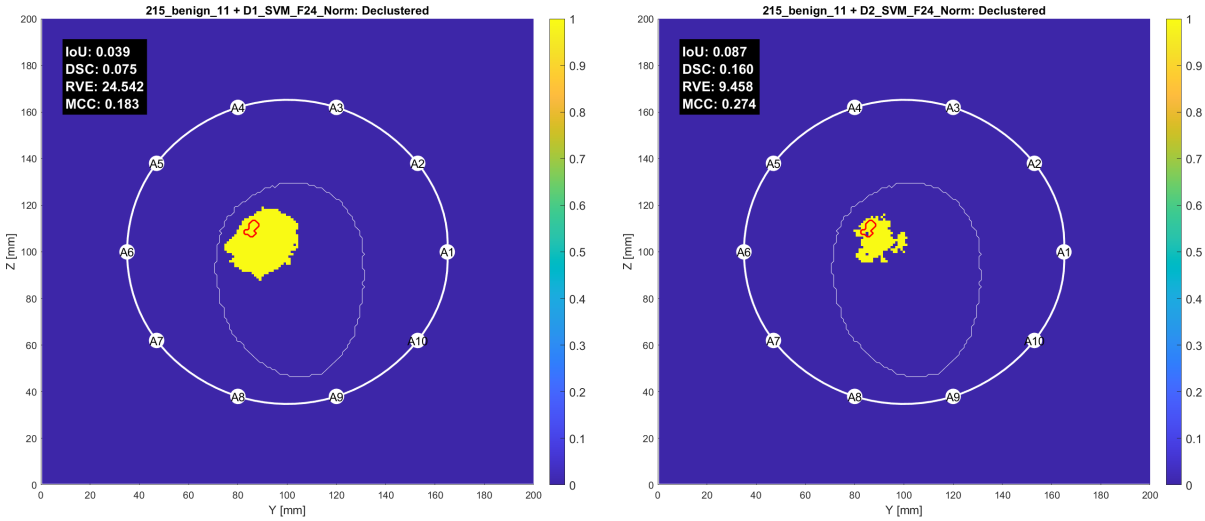
(a) D1, WT→DWTA

(b) D2, WT→DWTA



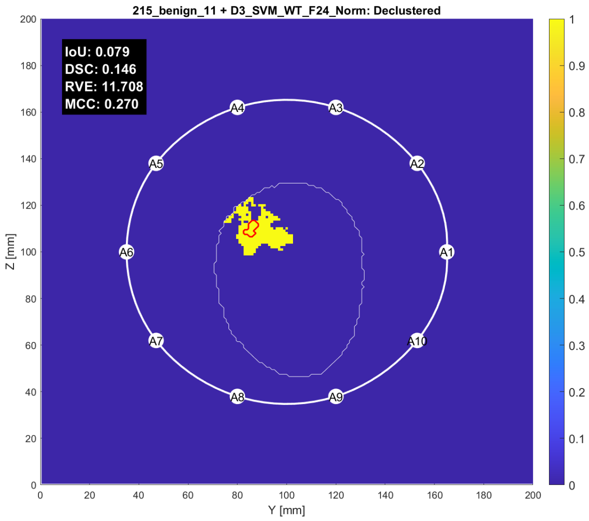
(c) D3, DWTA

Figure A.32: Images reconstructed from simulation 215_benign_11 using the DT algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



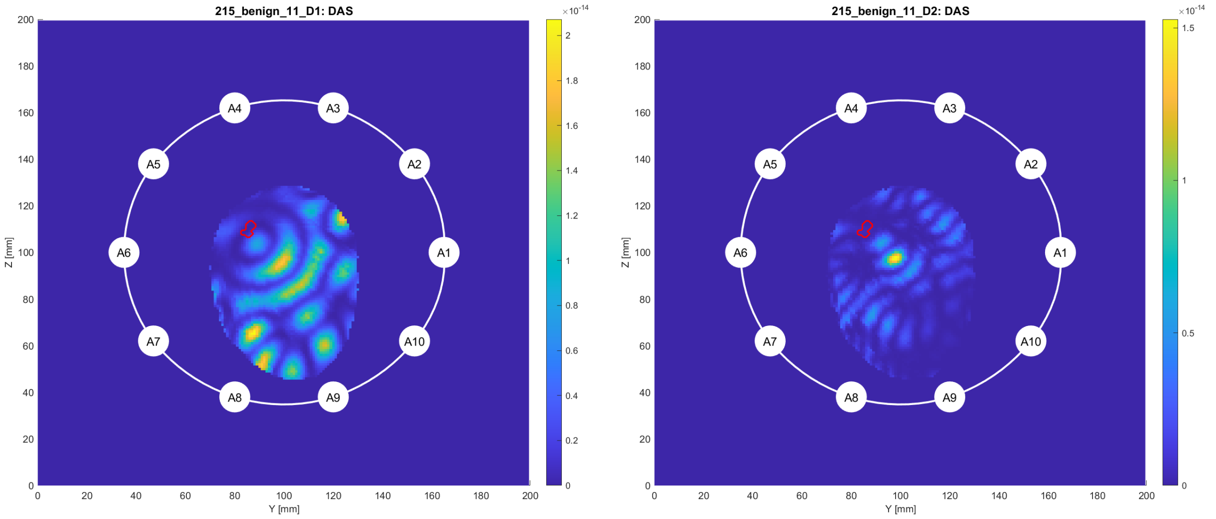
(a) D1, F24→Norm

(b) D2, F24→Norm



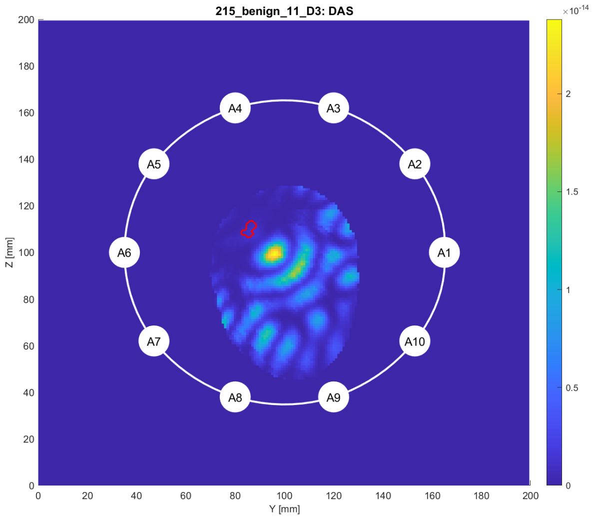
(c) D3, WT→F24→Norm

Figure A.33: Images reconstructed from simulation 215_benign_11 using the SVM algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



(a) D1

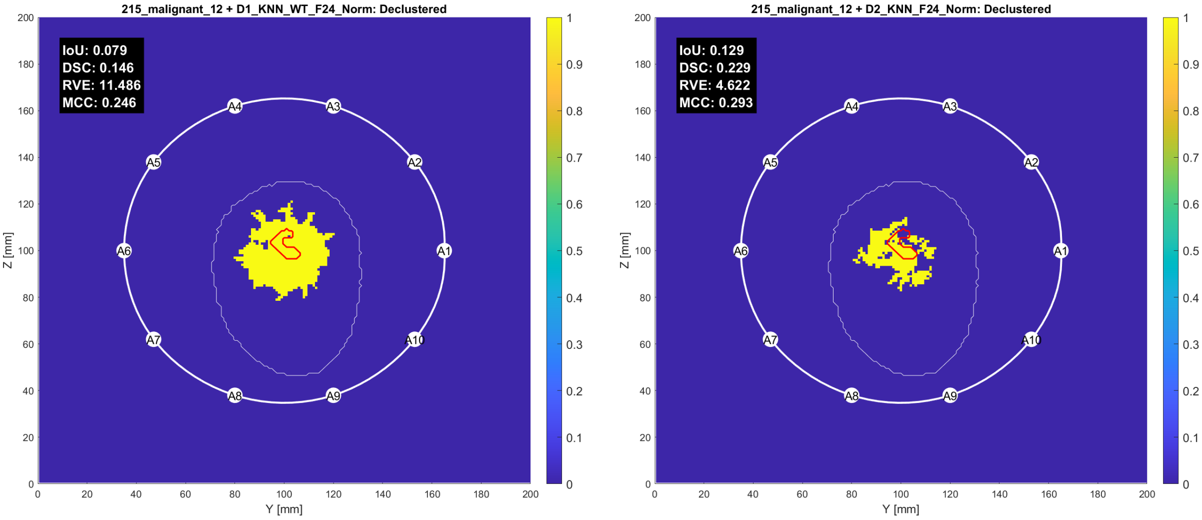
(b) D2



(c) D3

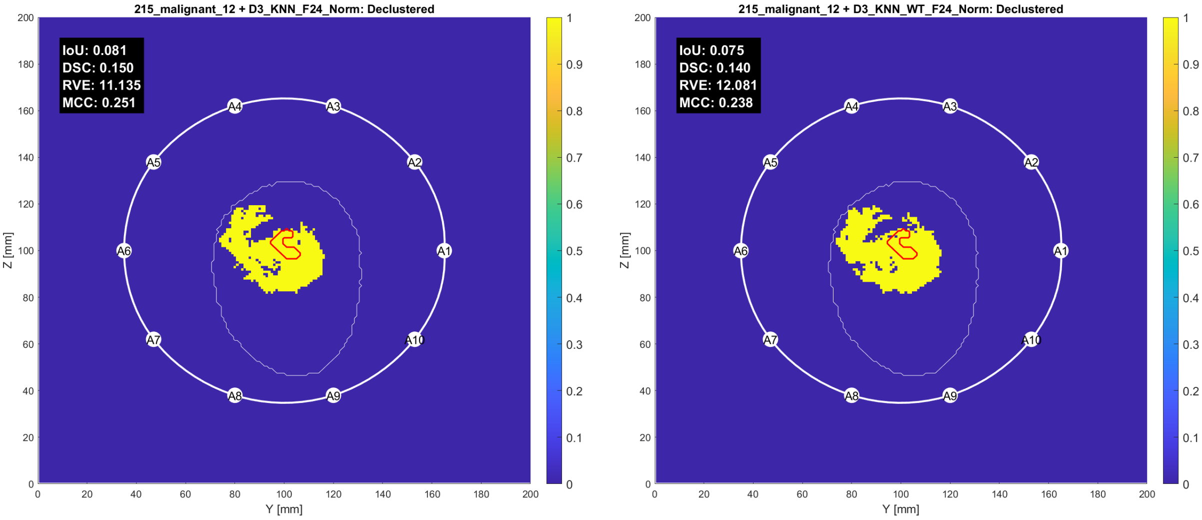
Figure A.34: Images reconstructed from simulation 215_benign_11 using the DAS algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

A.2.9 215_malignant_12



(a) D1, WT→F24→Norm

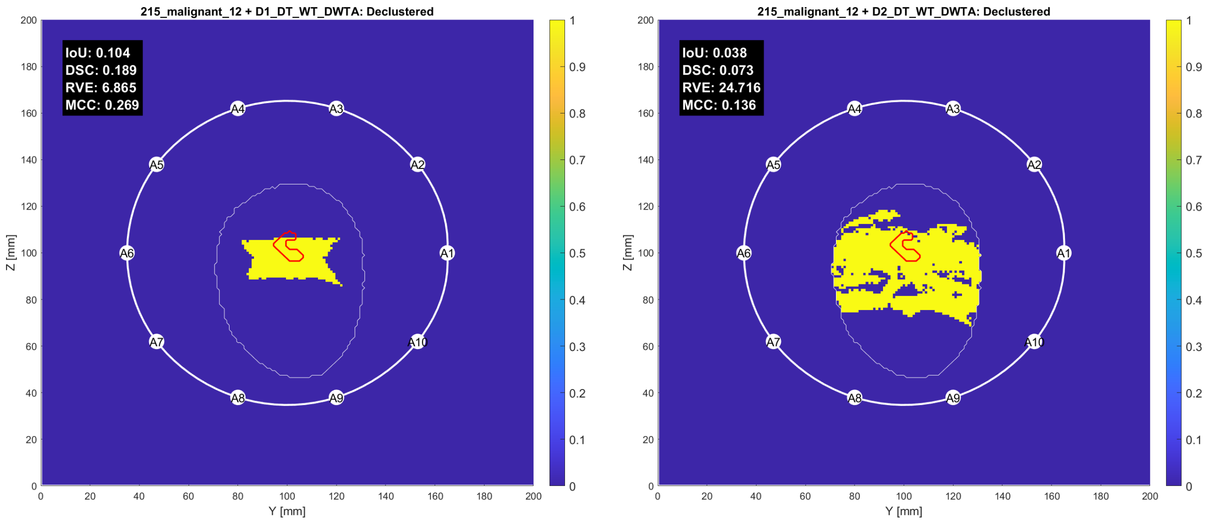
(b) D2, F24→Norm



(c) D3, F24→Norm

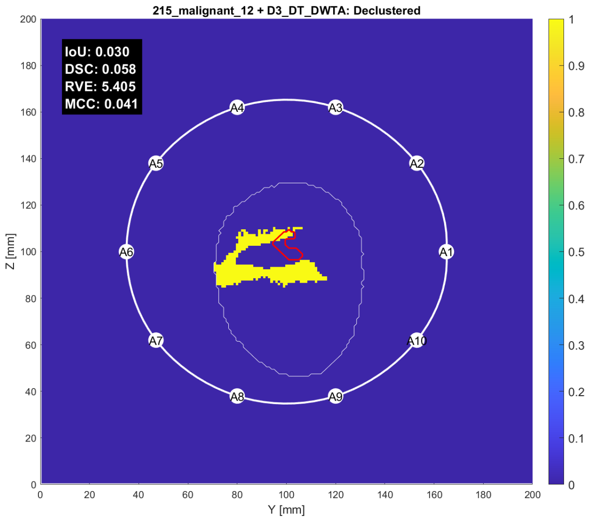
(d) D3, WT→F24→Norm

Figure A.35: Images reconstructed from simulation 215_malignant_12 using the KNN algorithm: (a) D1, (b) D2, (c) D3, (d) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



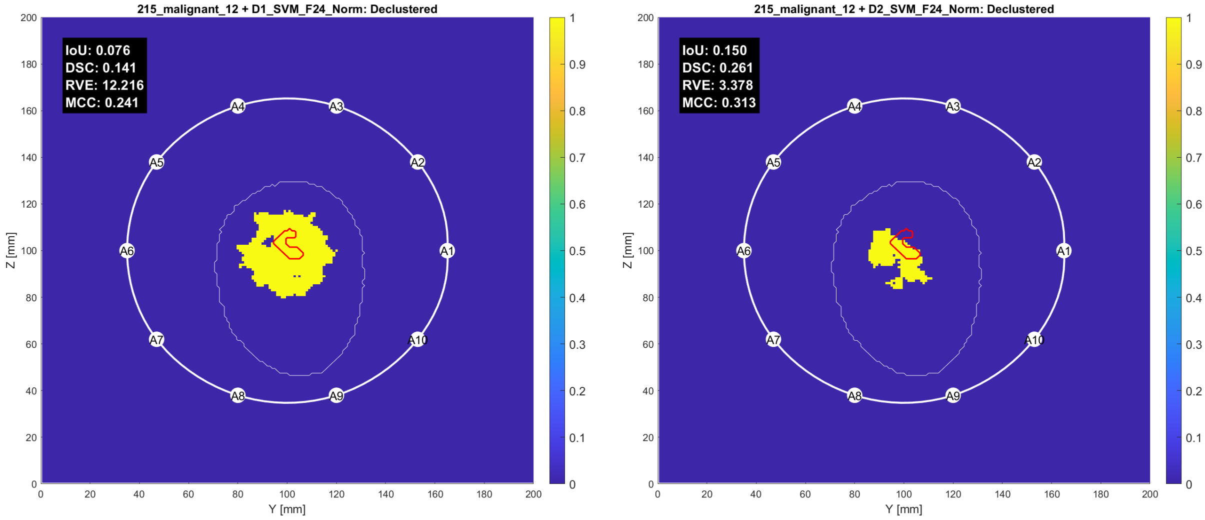
(a) D1, WT→DWTA

(b) D2, WT→DWTA



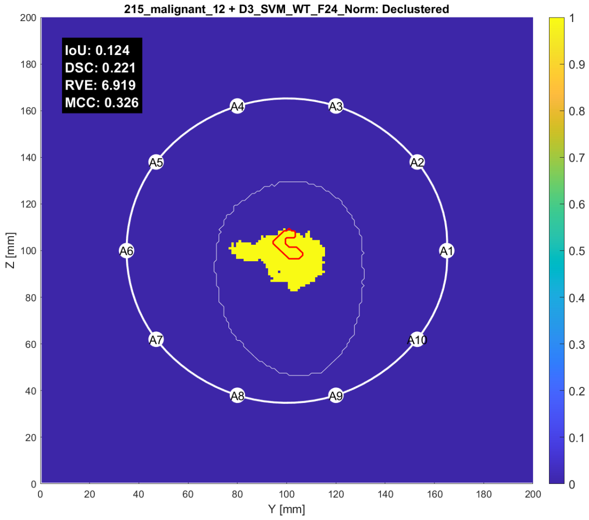
(c) D3, DWTA

Figure A.36: Images reconstructed from simulation 215_malignant_12 using the DT algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



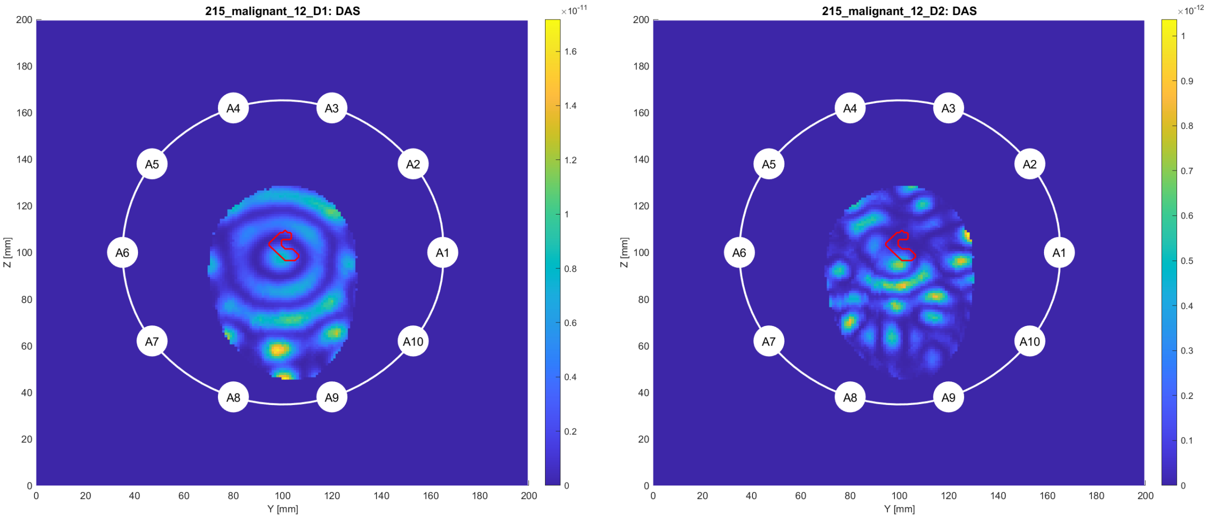
(a) D1, F24→Norm

(b) D2, F24→Norm



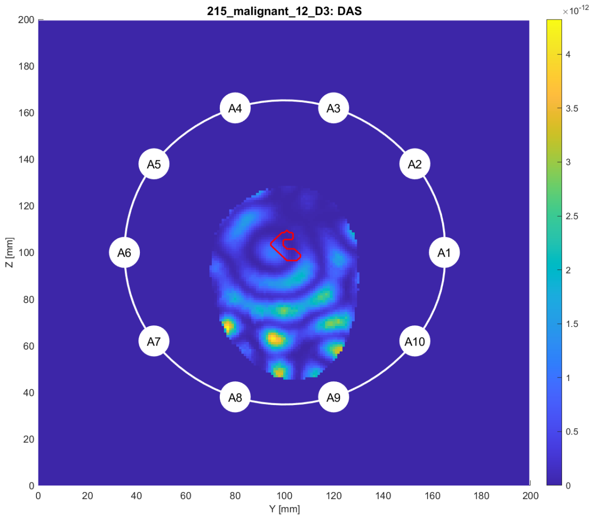
(c) D3, WT→F24→Norm

Figure A.37: Images reconstructed from simulation 215_malignant_12 using the SVM algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



(a) D1

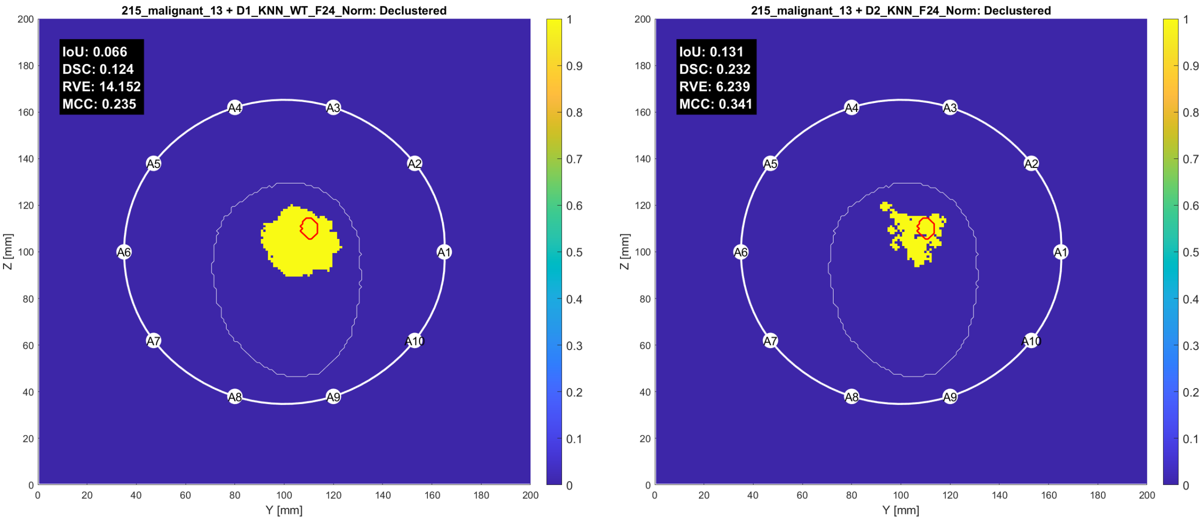
(b) D2



(c) D3

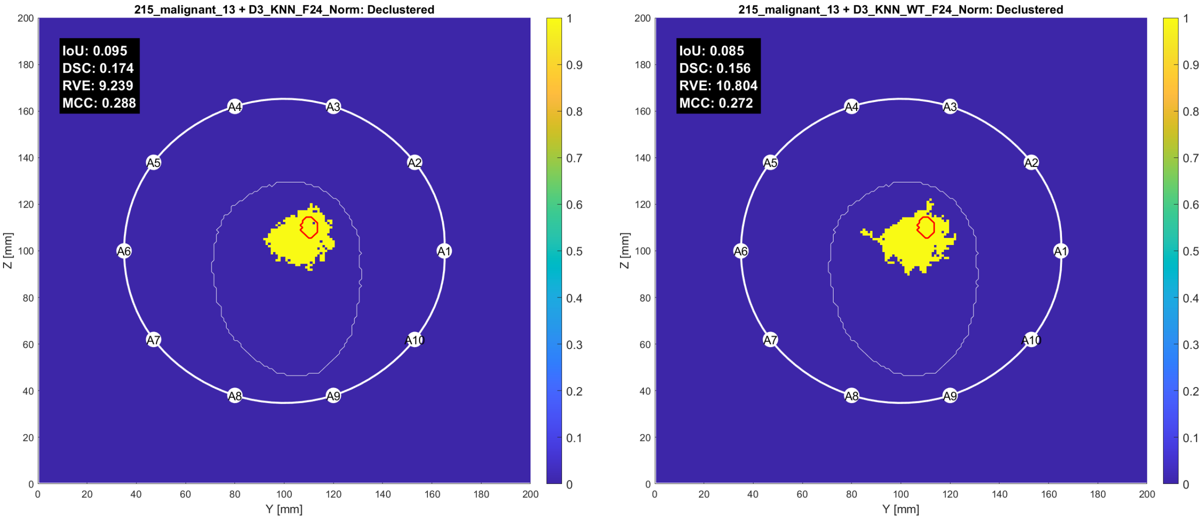
Figure A.38: Images reconstructed from simulation 215_malignant_12 using the DAS algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

A.2.10 215_malignant_13



(a) D1, WT→F24→Norm

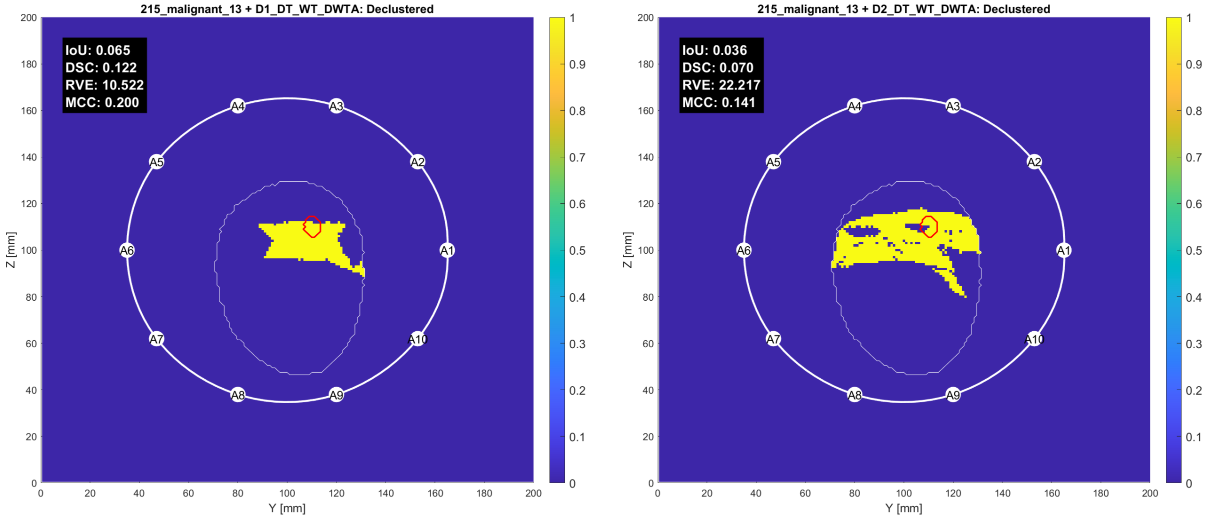
(b) D2, F24→Norm



(c) D3, F24→Norm

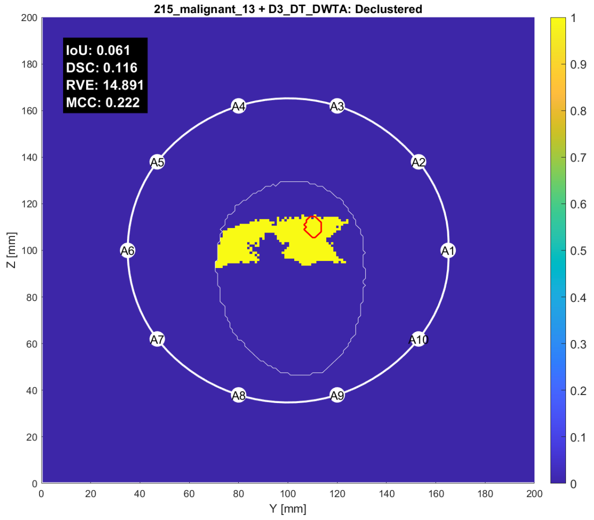
(d) D3, WT→F24→Norm

Figure A.39: Images reconstructed from simulation 215_malignant_13 using the KNN algorithm: (a) D1, (b) D2, (c) D3, (d) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



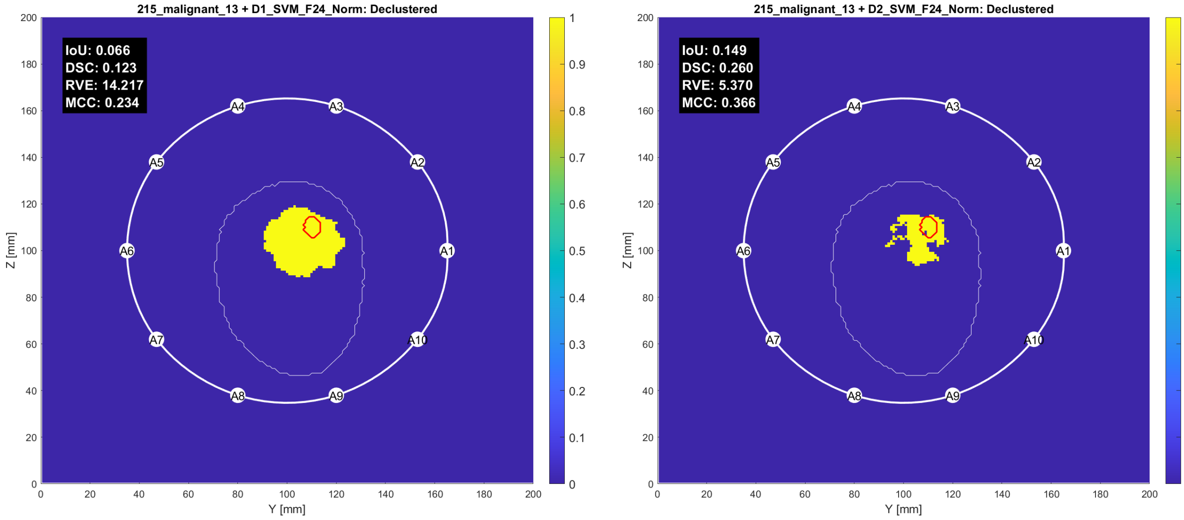
(a) D1, WT→DWTA

(b) D2, WT→DWTA



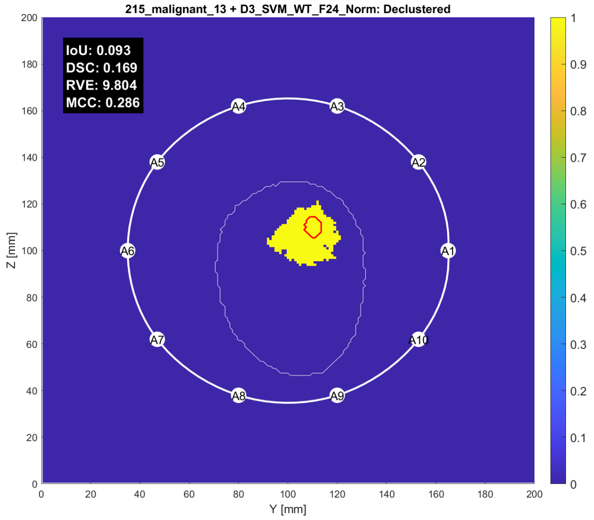
(c) D3, DWTA

Figure A.40: Images reconstructed from simulation 215_malignant_13 using the DT algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



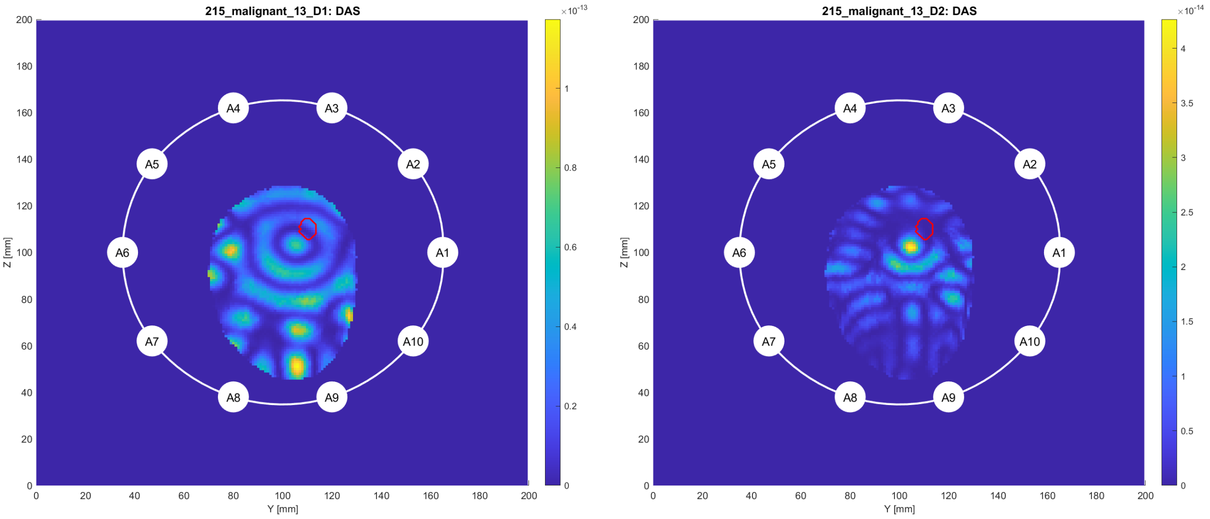
(a) D1, F24→Norm

(b) D2, F24→Norm



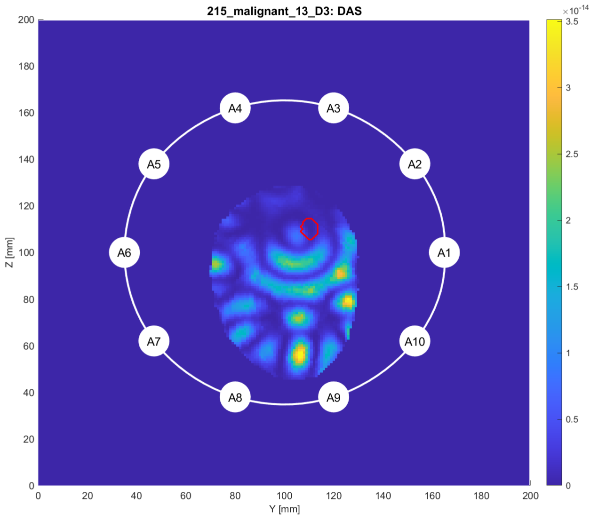
(c) D3, WT→F24→Norm

Figure A.41: Images reconstructed from simulation 215_malignant_13 using the SVM algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



(a) D1

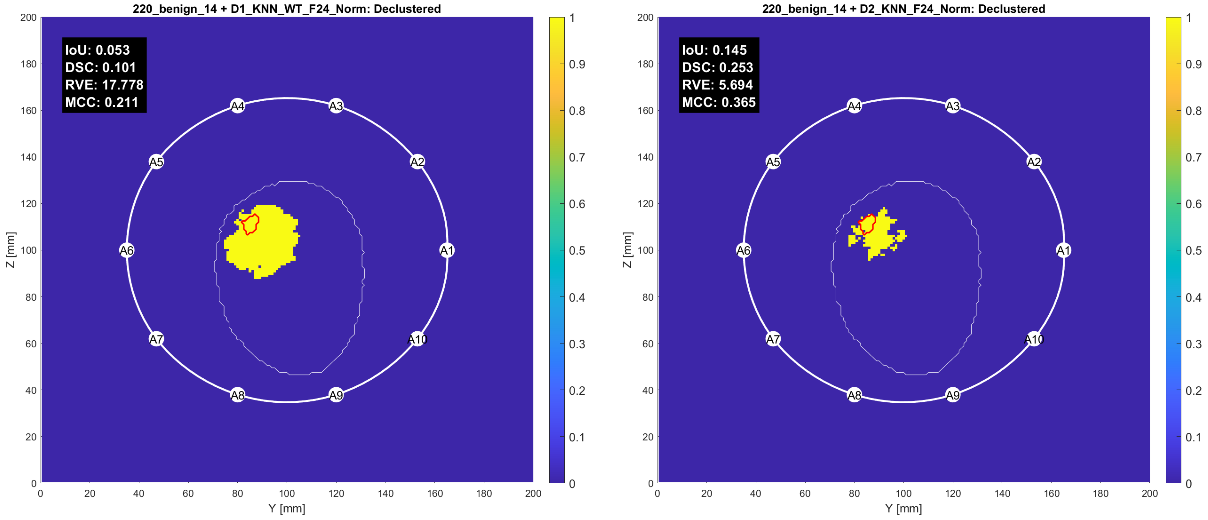
(b) D2



(c) D3

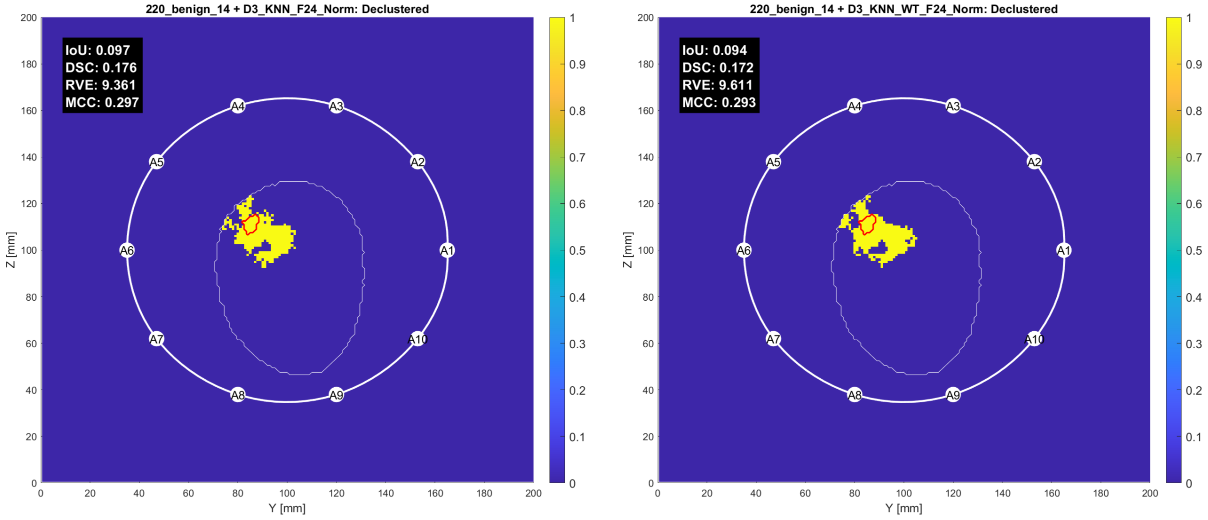
Figure A.42: Images reconstructed from simulation 215_malignant_13 using the DAS algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

A.2.11 220_benign_14



(a) D1, WT→F24→Norm

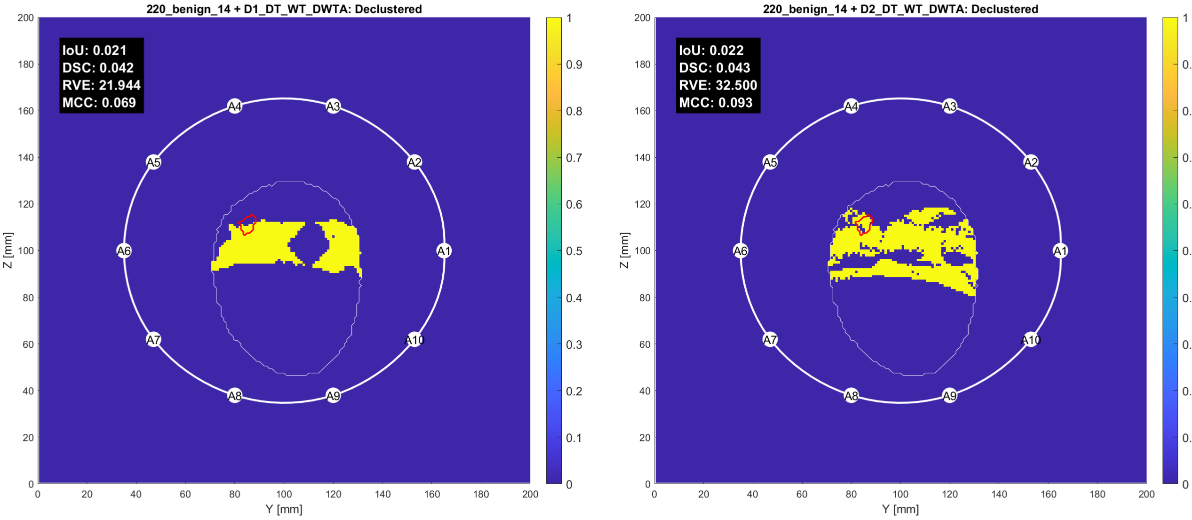
(b) D2, F24→Norm



(c) D3, F24→Norm

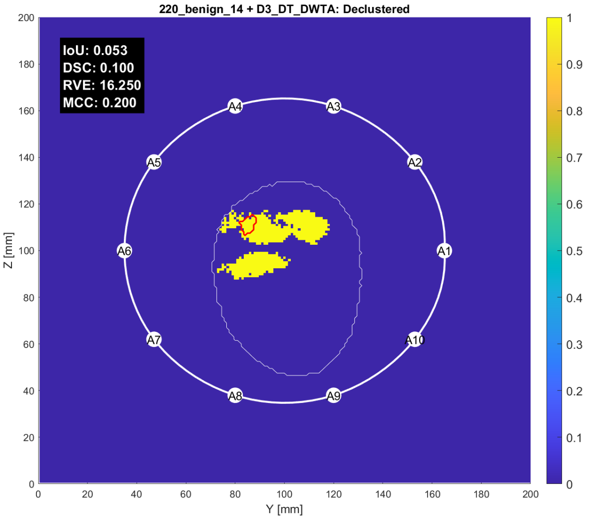
(d) D3, WT→F24→Norm

Figure A.43: Images reconstructed from simulation 220_benign_14 using the KNN algorithm: (a) D1, (b) D2, (c) D3, (d) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



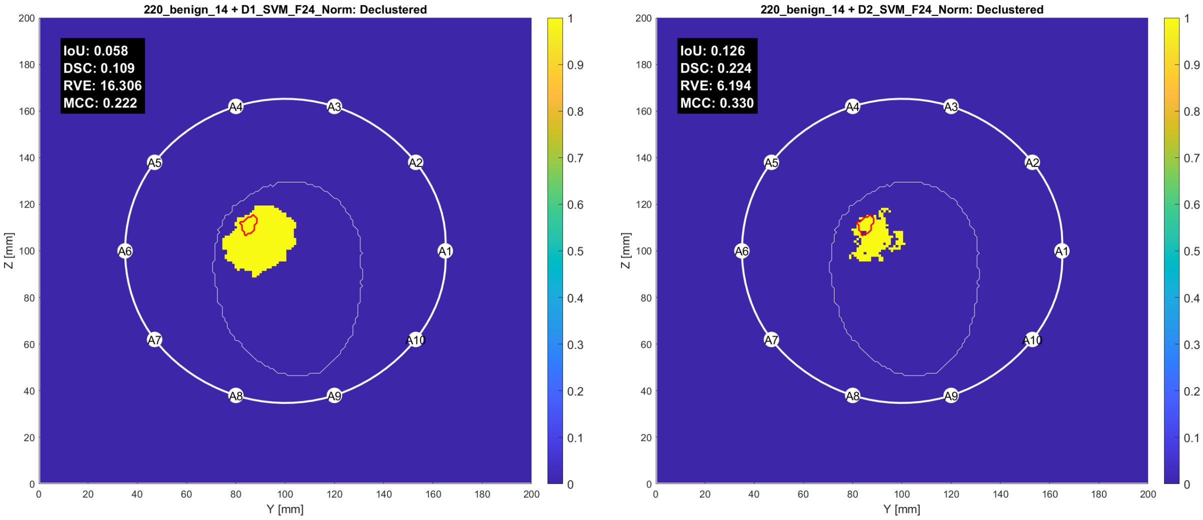
(a) D1, WT→DWTA

(b) D2, WT→DWTA



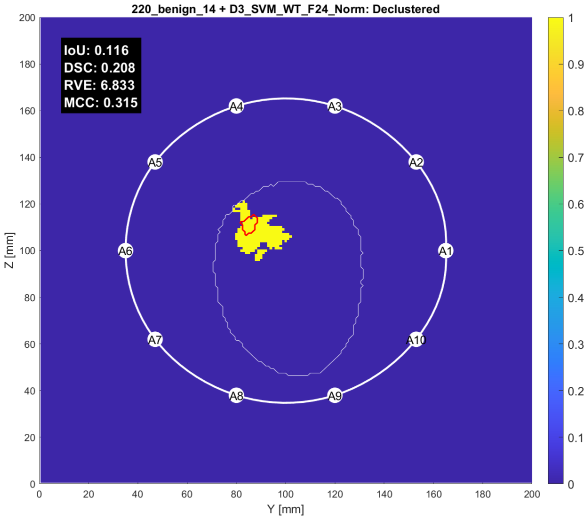
(c) D3, DWTA

Figure A.44: Images reconstructed from simulation 220_benign_14 using the DT algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



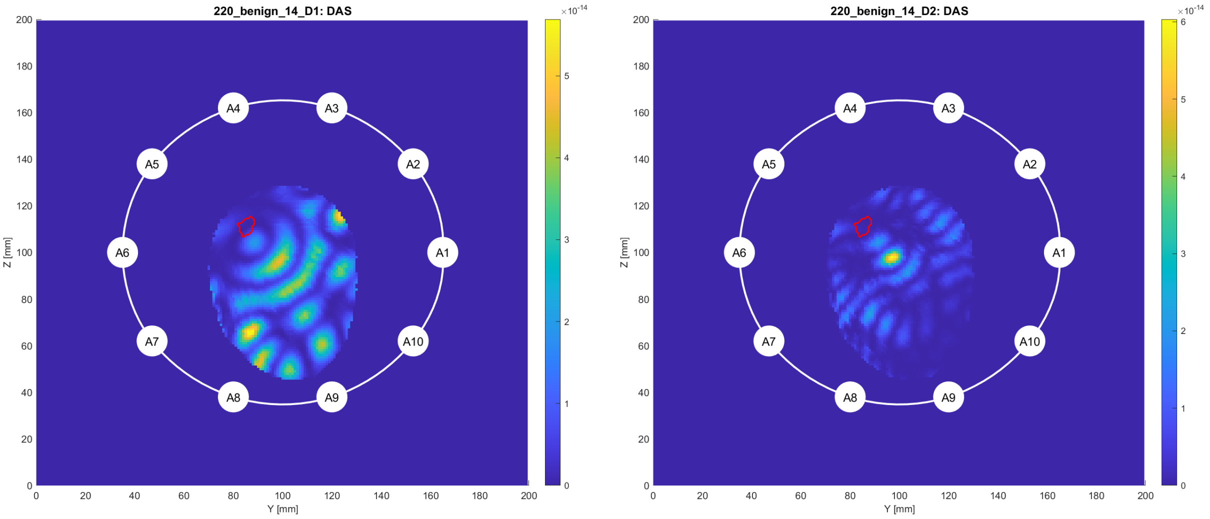
(a) D1, F24→Norm

(b) D2, F24→Norm



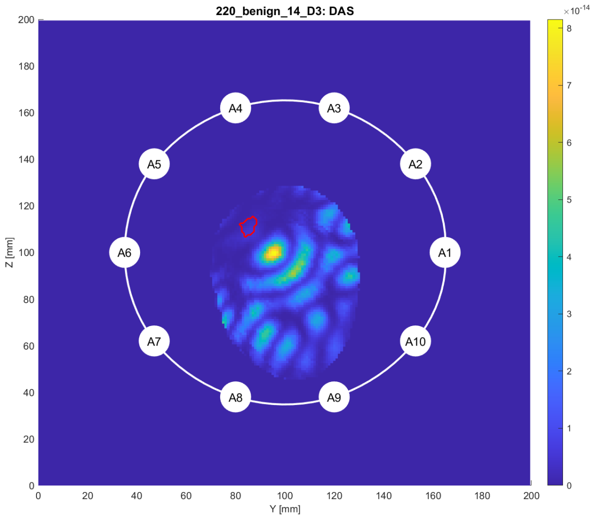
(c) D3, WT→F24→Norm

Figure A.45: Images reconstructed from simulation 220_benign_14 using the SVM algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



(a) D1

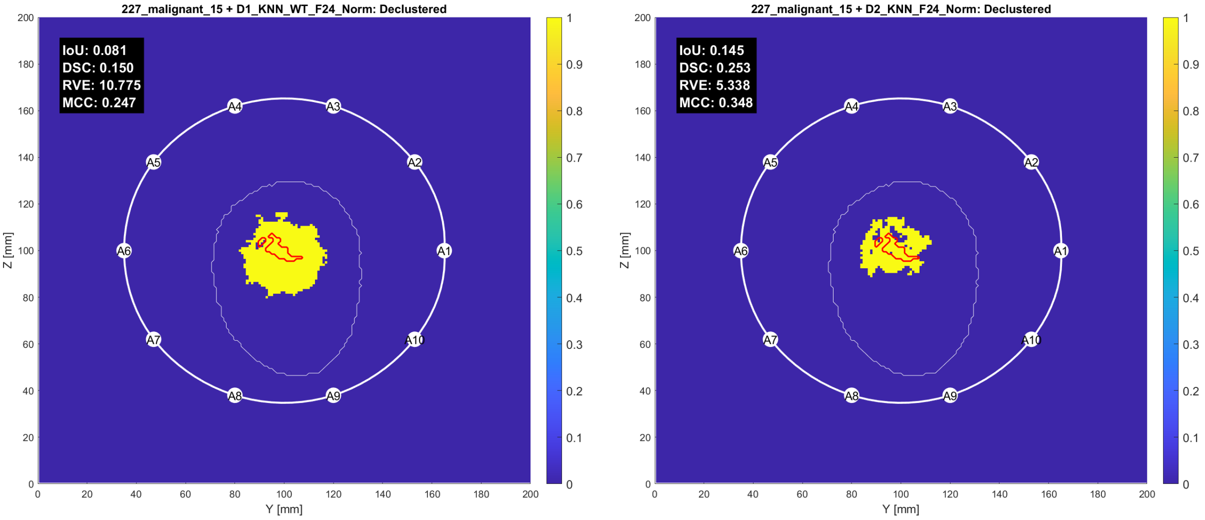
(b) D2



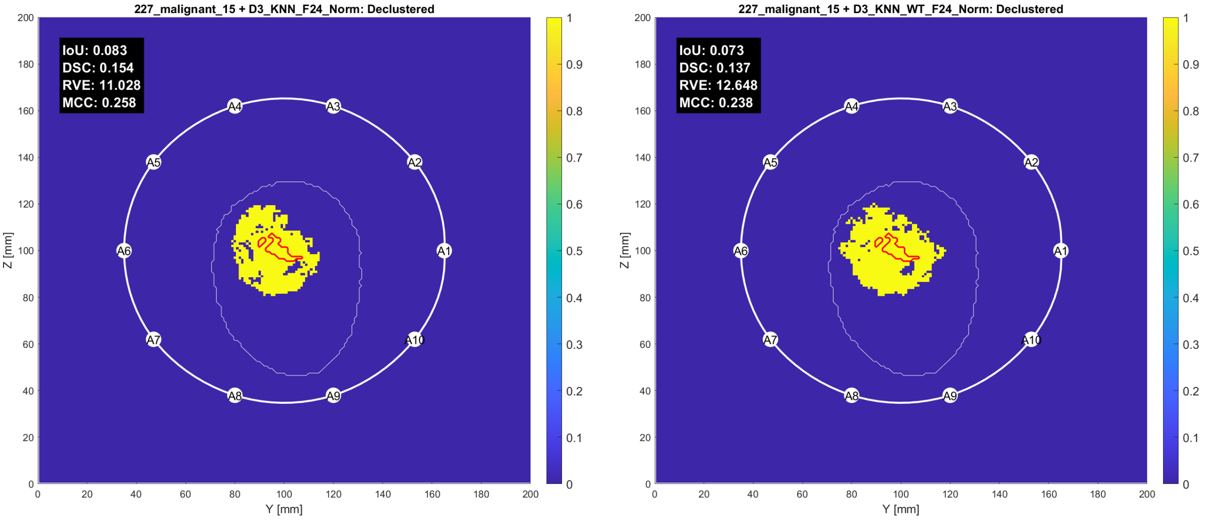
(c) D3

Figure A.46: Images reconstructed from simulation 220_benign_14 using the DAS algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

A.2.12 227_malignant_15

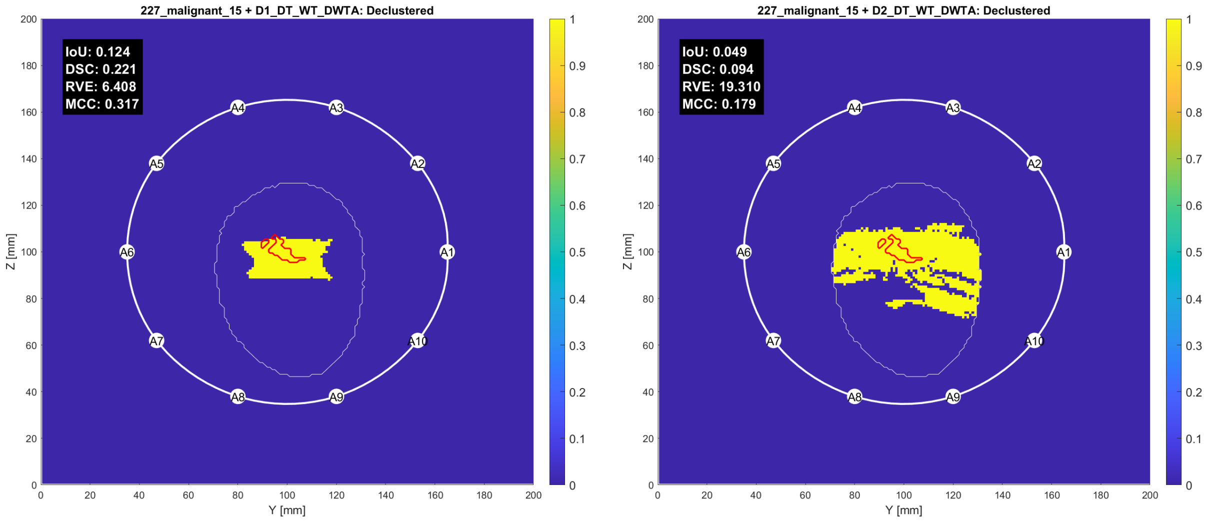


(a) D1, WT→F24→Norm (b) D2, F24→Norm



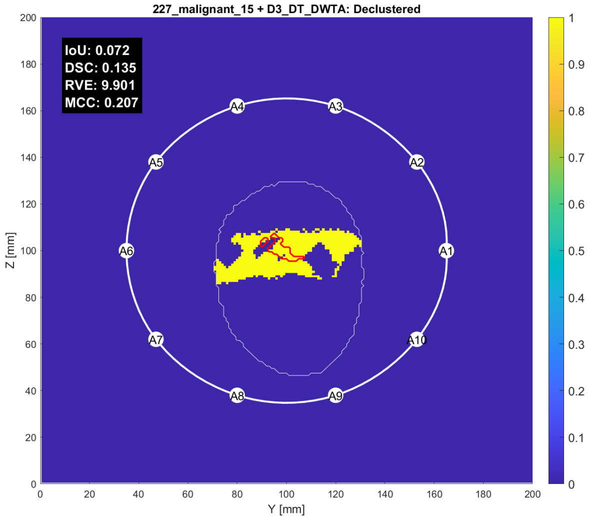
(c) D3, F24→Norm (d) D3, WT→F24→Norm

Figure A.47: Images reconstructed from simulation 227_malignant_15 using the KNN algorithm: (a) D1, (b) D2, (c) D3, (d) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



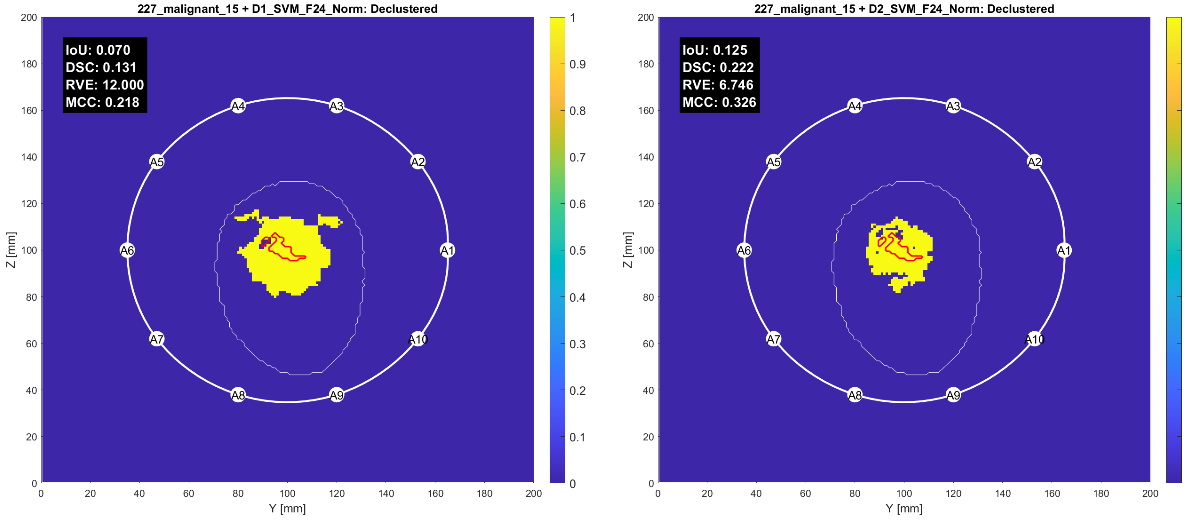
(a) D1, WT→DWTA

(b) D2, WT→DWTA



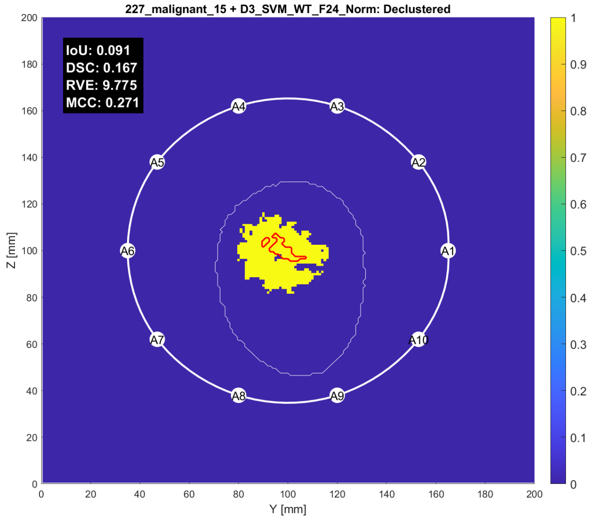
(c) D3, DWTA

Figure A.48: Images reconstructed from simulation 227_malignant_15 using the DT algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



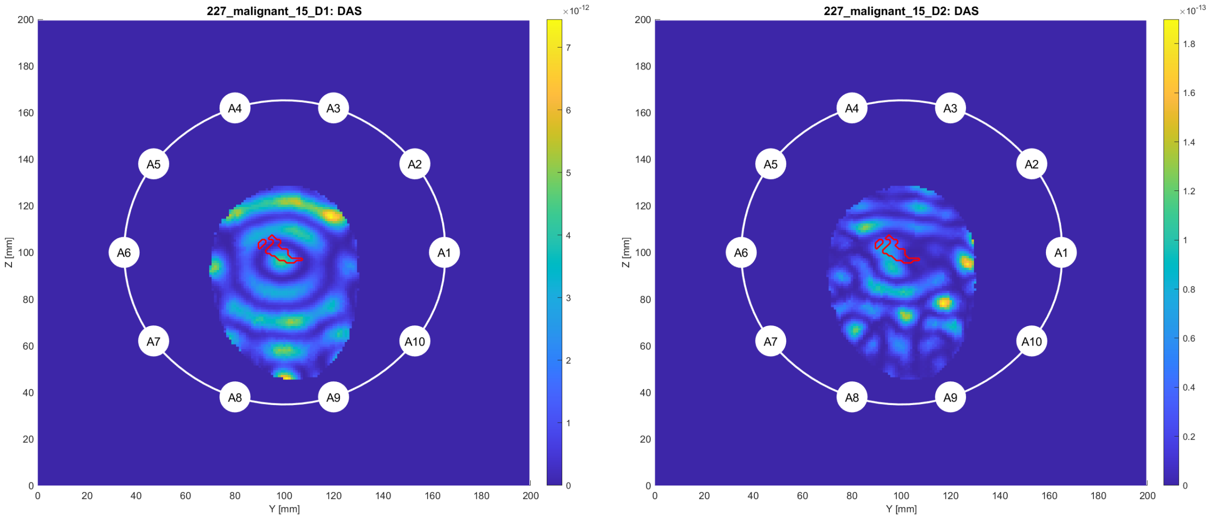
(a) D1, F24→Norm

(b) D2, F24→Norm



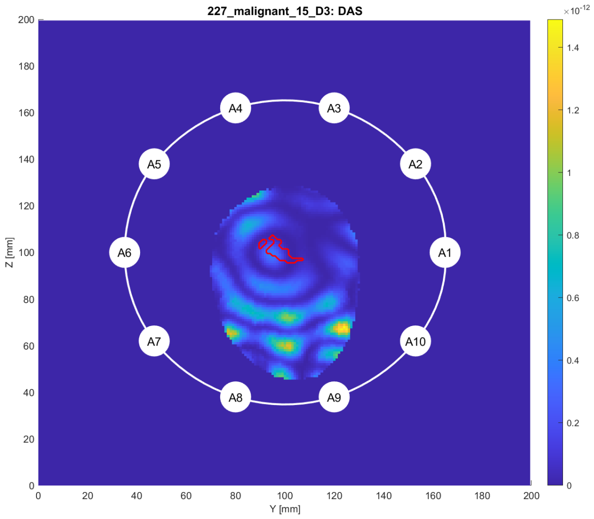
(c) D3, WT→F24→Norm

Figure A.49: Images reconstructed from simulation 227_malignant_15 using the SVM algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



(a) D1

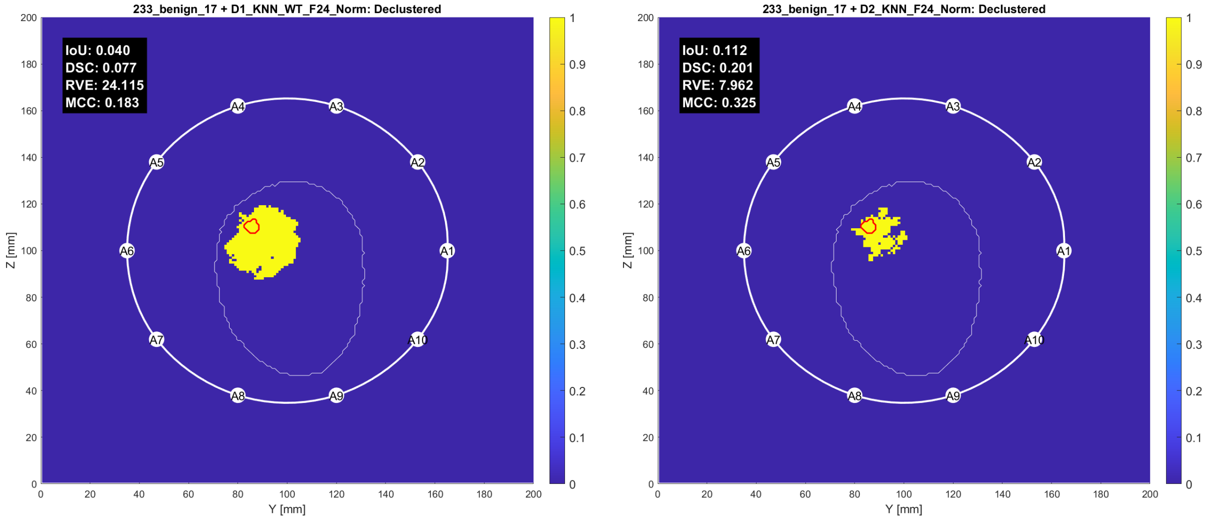
(b) D2



(c) D3

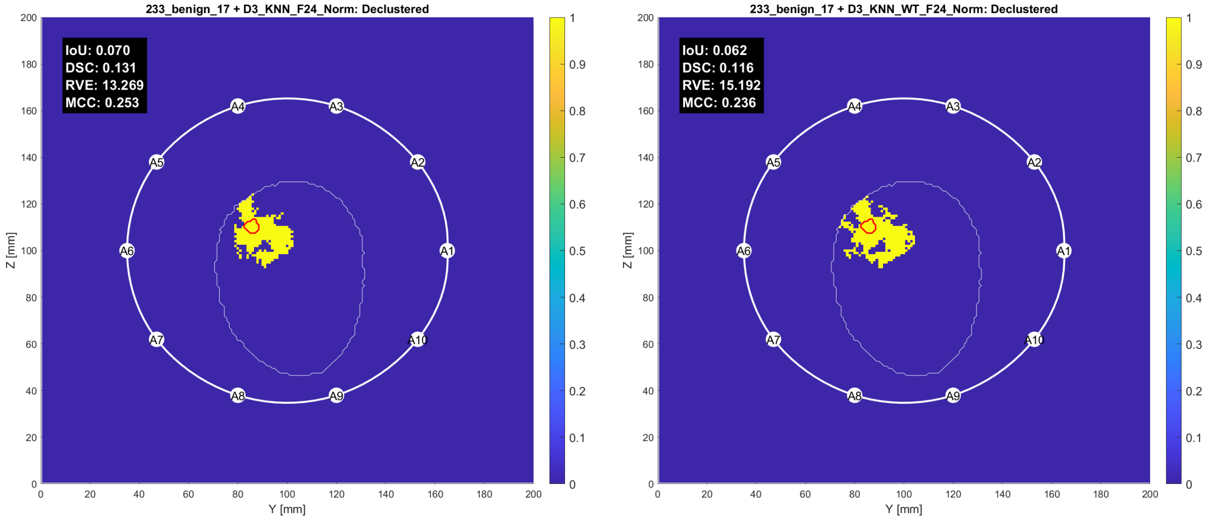
Figure A.50: Images reconstructed from simulation 227_malignant_15 using the DAS algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.

A.2.13 233_benign_17



(a) D1, WT→F24→Norm

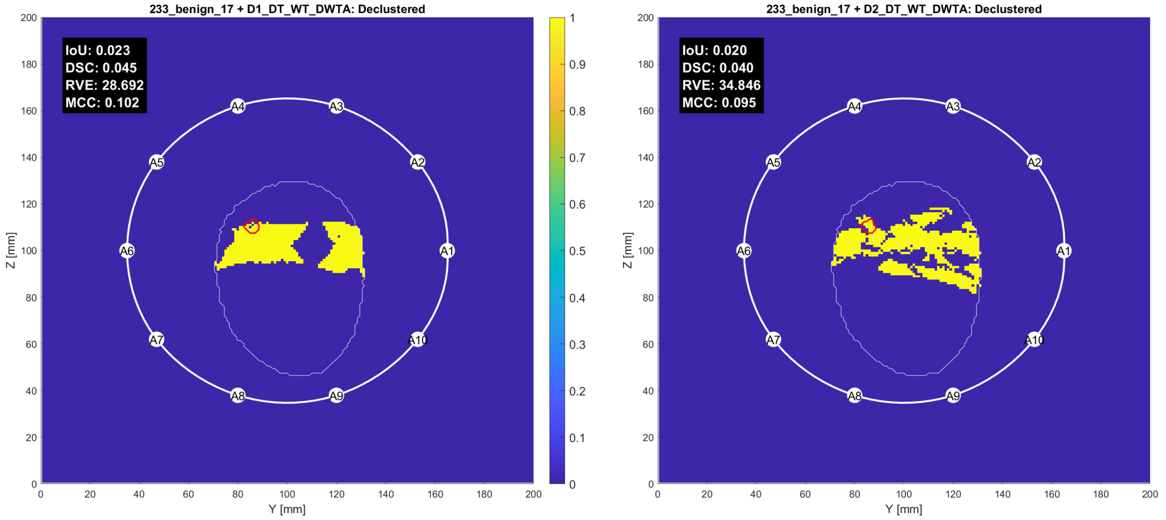
(b) D2, F24→Norm



(c) D3, F24→Norm

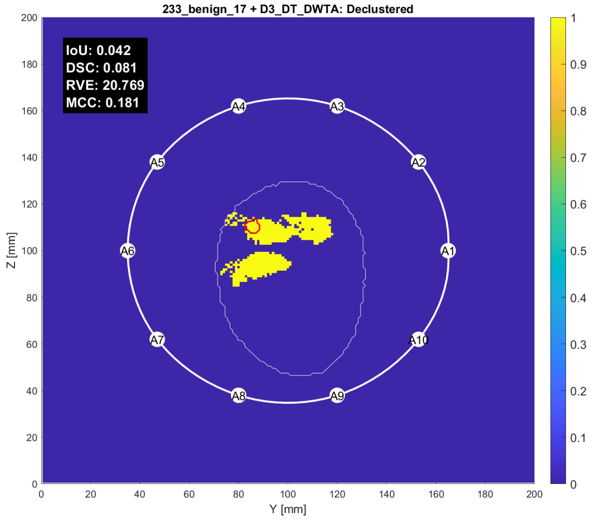
(d) D3, WT→F24→Norm

Figure A.51: Images reconstructed from simulation 233_benign_17 using the KNN algorithm: (a) D1, (b) D2, (c) D3, (d) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



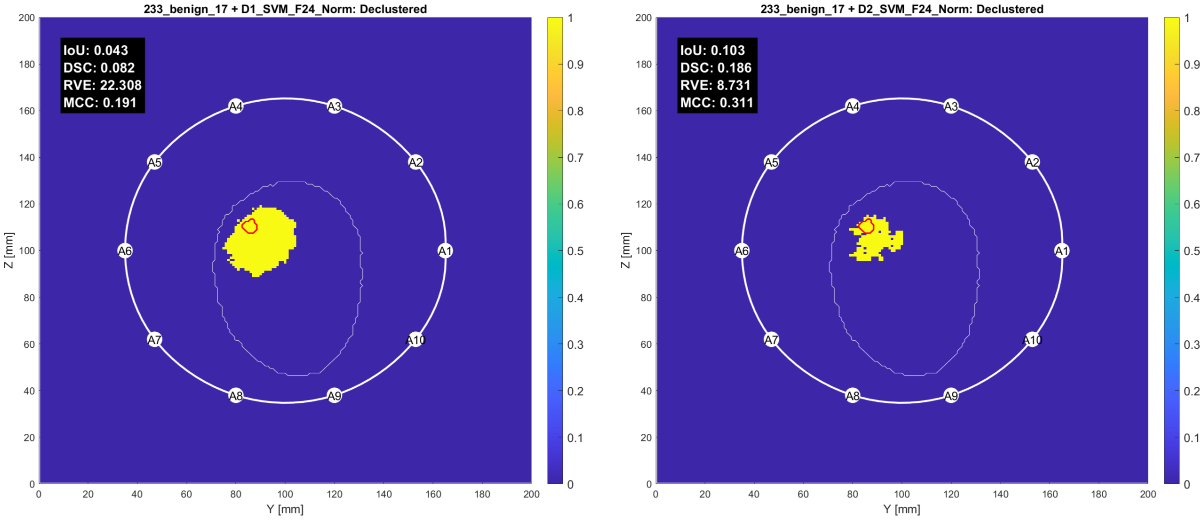
(a) D1, WT→DWTA

(b) D2, WT→DWTA



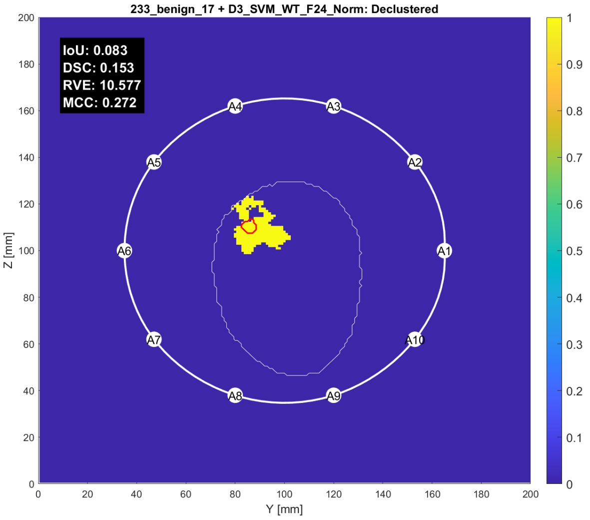
(c) D3, DWTA

Figure A.52: Images reconstructed from simulation 233_benign_17 using the DT algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



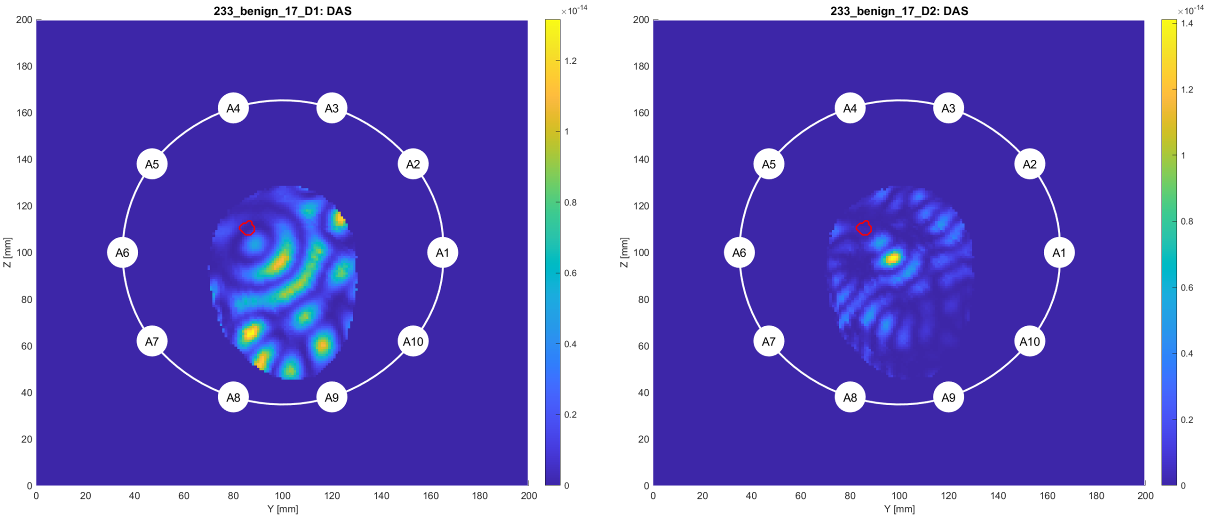
(a) D1, F24→Norm

(b) D2, F24→Norm



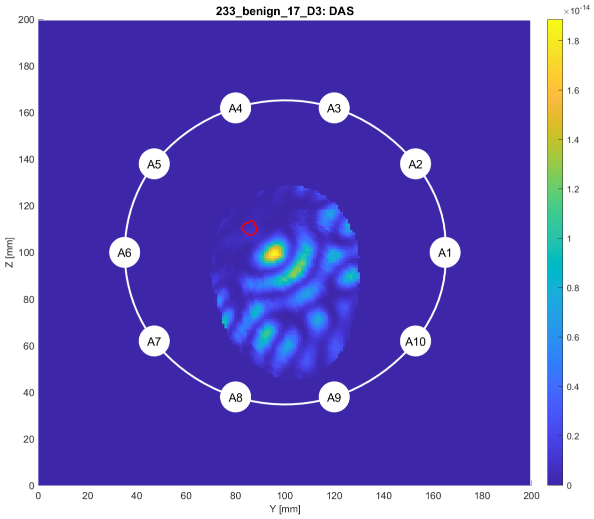
(c) D3, WT→F24→Norm

Figure A.53: Images reconstructed from simulation 233_benign_17 using the SVM algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.



(a) D1

(b) D2



(c) D3

Figure A.54: Images reconstructed from simulation 233_benign_17 using the DAS algorithm: (a) D1, (b) D2, (c) D3. The thicker white contour represents the antenna array, with antennas labeled A1 through A10; the thinner white contour outlines the phantom, and the red contour indicates the tumor location.