الجمهوريــة الجزائريــة الديمقراطيــة الشعبيــة		
Democratic and Popular Republic of Algeria		
وزارة التعلـيـم الـعـالي والبـحـث الـعلـمـي		
Ministry of Higher Education and Scientific Research		
University Mustapha Stambouli of Mascara		جامعة مصطفى اسطمبولي معسكر

Faculty of Exact Sciences — كلية العلوم الدقيقة

Department of Computer Science — قسم الإعلام الآلي

Thesis Presented to Obtain the Degree of

Doctorate in Computer Science

Intitled:

Design of Bio-Inspired Metaheuristics for Medical Image Classification

Presented by:

Brahim KHALDI

Defense date: Thursday, April 9, 2026 at 9:00 AM

In front of the Jury Committee composed of:

President	Boufera Fatma	Professor	University of Mascara
Examiner	Zagane Mohammed	MCA	University of Mascara
Examiner	Mekkaoui Kheireddine	Professor	University of Saida
Examiner	Bouziane Abdelghani	Professor	University of Naama
Supervisor	Debakla Mohammed	Professor	University of Mascara
Co-Supervisor	Djemal Khalifa	Professor	University of Evry-Paris Saclay, France
Guest	Bouougada Benamar	MCA	University of Naama

ACKNOWLEDGEMENT

All praise is due to God Almighty, who granted me the strength and perseverance to complete this work.

I extend my sincere thanks and appreciation to my esteemed supervisor ***Dr. Mohammed Debakla*** *and my co-supervisor* ***Khalifa Djemal****, for the continuous support of my related research, for guiding me throughout this academic journey, offering valuable advice, and providing corrections, motivation and immense knowledge that greatly contributed to the success of this work.*

I extend my thanks to all my esteemed professors for their knowledge and support, and to my friends who have been a great source of help and companionship.

Lastly, I would also like to express my deep gratitude to my parents, my wife and beloved family for their continuous support, patience, and encouragement, which gave me the strength and determination to reach this stage.

Dedication

To my beloved parents.

To my wife and children.

To my brother, sisters, and all my dear family, thank you for your boundless love and support.

To my friends and colleagues who have shared both the challenging and joyful moments with me, my deepest gratitude.

With love,

Brahim

Abstract

Breast cancer diagnosis relies fundamentally on histopathological examination of tissue samples; however, manual microscopic evaluation by pathologists is subjective, time-consuming, and prone to inter-observer variability. While deep learning has demonstrated remarkable success in medical image analysis, conventional single-architecture models exhibit inherent limitations. Convolutional neural networks excel at local feature extraction but fail to capture global contextual relationships, whereas Vision Transformers effectively model long-range dependencies but may be inefficient in representing fine-grained local details that are critical for histopathological interpretation. Moreover, optimizing deep learning models through hyperparameter tuning and feature selection remains computationally expensive and often suboptimal when relying solely on gradient-based optimization techniques.

This thesis proposes HNet, a novel hybrid deep learning architecture that strategically integrates three complementary neural paradigms for histopathological image classification. HNet combines EfficientNet for efficient local feature extraction from tissue morphology, an Advanced Vision Transformer (AVT) for capturing global contextual relationships and long-range tissue patterns, and Capsule Networks (CapsNet) for explicitly modeling spatial hierarchies and part-whole relationships inherent in tissue organization. To address the challenges of high-dimensional feature spaces and suboptimal parameter tuning, a Genetic Algorithm (GA)-based feature selection mechanism is incorporated as a critical preprocessing step between the concatenated EfficientNet-AVT feature representations and the CapsNet input. This metaheuristic-driven optimization enables automatic identification of the most discriminative features while reducing dimensionality and computational complexity.

Comprehensive experimental validation is conducted on the BreakHis dataset, encompassing binary (benign vs. malignant) breast cancer classification tasks. Detailed ablation studies quantify the individual contributions of each architectural component as well as the impact of GA-based feature selection on overall performance. Comparative evaluation against recent state-of-the-art methods demonstrates that the proposed HNet architecture achieves superior classification accuracy, sensitivity, specificity, and F1-score, establishing a new benchmark for breast cancer histopathology classification. The integration of metaheuristic-driven optimization with hybrid deep learning significantly enhances model robustness, generalization to unseen data, and computational efficiency compared to non-optimized approaches.

Beyond empirical performance gains, this work demonstrates how hybrid metaheuristic, deep learning frameworks can be effectively integrated into clinical workflows by addressing real-world deployment constraints, including inference latency, computational resource utilization, and model interpretability. The proposed methodology provides a generalizable template for applying metaheuristic-optimized

hybrid deep learning to a wide range of medical image classification tasks, offering substantial potential for advancing automated computer-aided diagnosis systems and digital pathology.

Keywords: Breast cancer classification, histopathological image analysis, hybrid deep learning, EfficientNet, Vision Transformers, Capsule Networks, genetic algorithm, feature selection, metaheuristic optimization, computer-aided diagnosis, BreakHis dataset.

Résumé

Le diagnostic du cancer du sein repose fondamentalement sur l'examen histopathologique des échantillons tissulaires ; toutefois, l'évaluation microscopique manuelle réalisée par les pathologistes est subjective, chronophage et sujette à une variabilité inter-observateurs. Bien que l'apprentissage profond ait démontré des performances remarquables en analyse d'images médicales, les modèles conventionnels à architecture unique présentent des limitations intrinsèques. Les réseaux de neurones convolutionnels excellent dans l'extraction de caractéristiques locales mais peinent à capturer les relations contextuelles globales, tandis que les Vision Transformers modélisent efficacement les dépendances à longue portée mais peuvent être moins performants pour représenter les détails locaux fins, essentiels à l'interprétation histopathologique. De plus, l'optimisation des modèles d'apprentissage profond par l'ajustement des hyperparamètres et la sélection de caractéristiques demeure coûteuse en calcul et souvent sous-optimale lorsqu'elle repose uniquement sur des techniques d'optimisation basées sur le gradient.

Cette thèse propose HNet, une architecture d'apprentissage profond hybride novatrice qui intègre stratégiquement trois paradigmes neuronaux complémentaires pour la classification d'images histopathologiques. HNet combine EfficientNet pour une extraction efficace des caractéristiques locales de la morphologie tissulaire, un Advanced Vision Transformer (AVT) pour la capture des relations contextuelles globales et des motifs tissulaires à longue portée, ainsi que les réseaux de capsules (CapsNet) pour la modélisation explicite des hiérarchies spatiales et des relations partie-tout inhérentes à l'organisation des tissus. Afin de relever les défis liés aux espaces de caractéristiques de grande dimension et à l'optimisation sous-optimale des paramètres, un mécanisme de sélection de caractéristiques basé sur un algorithme génétique (AG) est intégré comme étape de prétraitement clé entre les représentations concaténées issues d'EfficientNet et de l'AVT et l'entrée du CapsNet. Cette optimisation métaheuristique permet l'identification automatique des caractéristiques les plus discriminantes tout en réduisant la dimensionnalité et la complexité computationnelle.

Une validation expérimentale approfondie est menée sur le jeu de données BreakHis, couvrant les tâches de classification binaire du cancer du sein (bénin vs malin). Des études d'ablation détaillées quantifient la contribution individuelle de chaque composant architectural ainsi que l'impact de la sélection de caractéristiques basée sur les AG sur les performances globales. Une évaluation comparative avec des méthodes récentes de l'état de l'art démontre que l'architecture HNet proposée atteint des performances supérieures en termes de précision, de sensibilité, de spécificité et F1-score, établissant ainsi un nouveau référentiel pour la classification histopathologique du cancer du sein. L'intégration de l'optimisation métaheuristique avec l'apprentissage profond hybride améliore significativement la robustesse du modèle, sa capacité de généralisation à des données non vues et son efficacité computationnelle par rapport aux approches non optimisées.

Au-delà des améliorations empiriques des performances, ce travail montre comment les cadres hybrides combinant métaheuristiques et apprentissage profond peuvent être efficacement intégrés dans les flux de travail cliniques en prenant en compte les contraintes réelles de déploiement, notamment la latence d'inférence, l'utilisation des ressources computationnelles et l'interprétabilité des modèles. La méthodologie proposée fournit un cadre généralisable pour l'application de l'apprentissage profond hybride optimisé par métaheuristiques à un large éventail de tâches de classification d'images médicales, offrant un fort potentiel pour l'avancement des systèmes de diagnostic assisté par ordinateur et de la pathologie numérique.

Mots-clés : Classification du cancer du sein, analyse d'images histopathologiques, apprentissage profond hybride, EfficientNet, Vision Transformers, réseaux de capsules, algorithme génétique, sélection de caractéristiques, optimisation métaheuristique, diagnostic assisté par ordinateur, jeu de données BreakHis.

الملخص

يعتمد تشخيص سرطان الثدي بشكل أساسي على الفحص النسيجي المرضي لعينات الأنسجة؛ إلا أن التقييم المجهري اليدوي الذي يجريه أطباء علم الأمراض يُعد عملية ذاتية، تستغرق وقتًا طويلاً، ومعرضة لاختلافات بين الملاحظين. وعلى الرغم من أن التعلم العميق قد حقق نجاحًا ملحوظًا في مجال تحليل الصور الطبية، فإن النماذج التقليدية ذات البنية الواحدة تعاني من قيود جوهرية. إذ تتميز الشبكات العصبية الالتفافية بقدرتها العالية على استخراج الخصائص المحلية، لكنها تفشل في التقاط العلاقات السياقية العالمية، في حين أن نماذج Vision Transformersقادرة على تمثيل الاعتماديات بعيدة المدى بفعالية، لكنها قد تكون أقل كفاءة في تمثيل التفاصيل المحلية الدقيقة الضرورية لتفسير الصور النسيجية المرضية. علاوة على ذلك، فإن تحسين نماذج التعلم العميق من خلال ضبط المعاملات الفائقة واختيار الخصائص يظل مكلفًا حسابيًا وغالبًا ما يكون غير مثالي عند الاعتماد فقط على طرق التحسين القائمة على التدرج.

تقدم هذه الأطروحة HNet، وهي بنية هجينة جديدة للتعلم العميق تدمج بشكل استراتيجي ثلاثة نماذج عصبية متكاملة من أجل تصنيف الصور النسيجية المرضية. تجمع HNet بين EfficientNetلاستخراج الخصائص المحلية بكفاءة من مورفولوجيا الأنسجة، و Advanced Vision Transformer (AVT)لالتقاط العلاقات السياقية العالمية والأنماط النسيجية بعيدة المدى، وشبكات الكبسولات (CapsNet) من أجل نمذجة الهياكل المكانية والعلاقات الجزء-الكل المتأصلة في تنظيم الأنسجة. ولمواجهة تحديات فضاءات الخصائص عالية الأبعاد والتحسين غير المثالي للمعاملات، تم دمج آلية لاختيار الخصائص تعتمد على الخوارزمية الجينية (GA) كمرحلة معالجة مسبقة أساسية بين الخصائص المدمجة المستخرجة من EfficientNet و AVTومدخل .CapsNet وتُمكّن هذه العملية التحسينية الميتاهيوريستية (الخوارزميات الموحاة بالطبيعة) من تحديد الخصائص الأكثر تمييزًا تلقائيًا مع تقليل الأبعاد والتعقيد الحسابي.

تم إجراء تقييم تجريبي شامل باستخدام مجموعة بيانات BreakHis، شمل مهام التصنيف الثنائي لسرطان الثدي (حميد مقابل خبيث). وقد مكّنت دراسات الإزالة التفصيلية من قياس المساهمة الفردية لكل مكوّن معماري وكذلك تأثير اختيار الخصائص القائم على الخوارزمية الجينية على الأداء الكلي. كما أظهرت المقارنة مع أحدث الطرق في الأدبيات أن بنية HNet المقترحة تحقق أداءً متفوقًا من حيث الدقة، والحساسية، والنوعية، f1-score، مما يرسخ معيارًا جديدًا لتصنيف الصور النسيجية المرضية لسرطان الثدي. ويؤدي دمج التحسين الميتاهيوريستي مع التعلم العميق الهجين إلى تحسين ملحوظ في متانة النموذج، وقدرته على التعميم على بيانات غير مرئية، وكفاءته الحسابية مقارنة بالأساليب غير المحسّنة.

إضافةً إلى تحسينات الأداء التجريبية، يوضح هذا العمل كيفية دمج الأطر الهجينة التي تجمع بين الميتاهيوريستيات والتعلم العميق بفعالية ضمن سير العمل السريري، مع مراعاة متطلبات التطبيق العملي مثل زمن الاستدلال، واستهلاك الموارد الحسابية، وقابلية تفسير النماذج. وتوفر المنهجية المقترحة إطارًا عامًا لتطبيق التعلم العميق الهجين المحسّن بالميتاهيوريستيات على مجموعة واسعة من مهام تصنيف الصور الطبية، مما يفتح آفاقًا واعدة لتطوير أنظمة التشخيص بمساعدة الحاسوب وعلم الأمراض الرقمي.

الكلمات المفتاحية :تصنيف سرطان الثدي، تحليل الصور النسيجية المرضية، التعلم العميق الهجين، EfficientNet، Vision Transformers، شبكات الكبسولات، الخوارزمية الجينية، اختيار الخصائص، التحسين الميتاهيوريستي، التشخيص بمساعدة الحاسوب، مجموعة بيانات.BreakHis

Content

General Introduction

1 Problem Statement

In recent years, the field of medical image analysis has witnessed significant advancements, driven largely by the integration of deep learning techniques and optimization methods. Medical image analysis, which involves extracting clinically meaningful information from digital imaging data, plays a crucial role in early disease detection, diagnosis, treatment planning, and patient monitoring across numerous medical specialties. Despite substantial progress, challenges still exist in terms of diagnostic accuracy, computational efficiency, and generalization to diverse patient populations and imaging protocols, particularly when dealing with complex histopathological images and high-dimensional feature spaces.

Breast cancer remains the most prevalent malignancy in women globally, representing 11.7% of all new cancer cases annually. Histopathological examination of tissue samples obtained through biopsy is the gold standard for confirming breast cancer diagnosis and determining tumor grade and subtype, which are critical for guiding treatment decisions and predicting patient outcomes. However, manual microscopic evaluation of histopathological images is subjective, time-consuming, and heavily dependent on the expertise and experience of individual pathologists, leading to significant inter-observer variability and potential diagnostic delays. The advent of digital pathology and whole-slide imaging (WSI) technology has enabled high-resolution acquisition of histopathological images at multiple magnification levels, generating vast quantities of image data that far exceed human visual processing capacity and create bottlenecks in clinical workflows.

Deep learning, a subset of machine learning based on artificial neural networks, has revolutionized medical image analysis and interpretation. Convolutional neural networks (CNNs), in particular, have demonstrated state-of-the-art performance in medical imaging tasks including object detection, image segmentation, and classification. The ability of deep learning models to automatically learn hierarchical features from raw images without manual feature engineering has made them indispensable in medical image analysis. However, despite their success, conventional single-architecture deep learning models often have significant limitations. Convolutional neural networks excel at local feature extraction and spatial invariance but struggle to capture long-range global contextual relationships; conversely, Vision Transformers capture global context and relationships through self-attention mechanisms but may be less efficient at modeling fine-grained local details critical for histopathological analysis. Furthermore, deep learning models often require large amounts of labeled training data and are prone to overfitting, especially with insufficient data diversity or when applied to new patient populations or imaging protocols. This has led researchers to explore various strategies, including data augmentation, transfer learning, and advanced optimization techniques, to improve model robustness and generalization.

Optimization plays a fundamental role in enhancing both the efficiency and accuracy of medical image analysis tasks. Optimization algorithms are employed at multiple stages: from fine-tuning the hyperparameters of deep learning models (learning rates, dropout rates, regularization coefficients) to selecting the most relevant features for improving classification or segmentation results, to optimizing fusion weights in multimodal imaging systems. Traditional optimization methods, such as gradient descent, have long been used in training deep neural networks, but gradient-based approaches alone are insufficient for many medical imaging optimization challenges. The rise of metaheuristic optimization techniques, including genetic algorithms, particle swarm optimization, grey wolf optimization, and other nature-inspired methods, has opened new avenues for solving complex, non-convex medical image analysis problems. These algorithms leverage randomness and iterative search processes to explore broader solution spaces, thereby improving the diversity and quality of optimization results without requiring explicit gradient information.

A critical gap exists at the intersection of multiple challenges: (1) single-architecture deep learning models fail to capture complementary aspects of histopathological images (local vs. global, spatial hierarchy vs. contextual relationships); (2) hyperparameter optimization and feature selection in high-dimensional medical imaging spaces remain computationally expensive and often suboptimal; and (3) while deep learning and metaheuristics have been successfully applied independently, their strategic integration remains underdeveloped for medical imaging applications. Current approaches typically employ either conventional single-network architectures or manual ensemble design without systematic optimization, failing to leverage the synergistic potential of combining complementary neural network paradigms with metaheuristic-driven optimization.

This thesis aims to address these challenges by exploring the strategic integration of hybrid deep learning architectures with metaheuristic optimization in the context of histopathological image analysis and breast cancer classification. By combining the complementary strengths of multiple deep learning paradigms (convolutional networks for local features, vision transformers for global context, and capsule networks for spatial hierarchies) with genetic algorithm-based feature selection and optimization, we seek to improve the accuracy, efficiency, robustness, and generalization capabilities of medical image analysis systems. The proposed HNet architecture and its metaheuristic-optimized variant will be comprehensively validated through extensive experiments on the BreakHis breast cancer histopathology dataset, demonstrating the potential to overcome existing limitations and advance the state-of-the-art in digital pathology and computer-aided diagnosis.

2 Research Contributions

The primary contributions of this research include:

- **Comprehensive Literature Review**: A thorough synthesis of deep learning architectures (CNNs, Vision Transformers, Capsule Networks), metaheuristic optimization algorithms (genetic algorithms, swarm intelligence), and their applications in medical image analysis, establishing the theoretical foundation for hybrid approaches.
- **Hybrid Multi-Modal Deep Learning Architecture (HNet)**: Development of a novel three-component hybrid architecture that strategically integrates EfficientNet (for efficient local feature extraction), Advanced Vision Transformer/AVT (for global contextual relationships), and Capsule Networks (for spatial hierarchy preservation) specifically optimized for histopathological image classification. This architecture overcomes the limitations of single-network approaches by leveraging complementary strengths of each paradigm.
- **Metaheuristic-Optimized Feature Selection**: Integration of Genetic Algorithm (GA)-based feature selection as a critical preprocessing step between concatenated features from EfficientNet and AVT extraction and Capsule Network input. This optimization automatically identifies discriminative features from the combined high-dimensional feature space, reducing dimensionality, eliminating feature redundancy, and improving downstream classification while reducing computational complexity.
- **Comprehensive Experimental Validation and Ablation Studies**: Extensive evaluation of the proposed HNet architecture on the BreakHis dataset for both binary (benign vs. malignant) and multi-class breast cancer classification across multiple magnification levels (40X, 100X, 200X, 400X). Detailed ablation studies quantifying the individual contribution of each architectural component and the impact of GA-based feature selection on overall performance.
- **Clinical Benchmarking and State-of-the-Art Comparison**: Comparative evaluation against recent state-of-the-art methods for breast cancer histopathology classification, demonstrating superior performance and establishing a new benchmark for the BreakHis dataset using hybrid metaheuristic-deep learning approaches.
- **Practical Integration Framework**: Demonstration of how hybrid metaheuristic-deep learning approaches can be practically integrated into clinical workflows, addressing real-world deployment constraints including computational efficiency, inference latency, interpretability, and generalization to new patient cohorts.

3 Thesis Organization

The structure of the thesis is organized as follows:

- **General Introduction**: Outlines the research motivation in medical image analysis and histopathological breast cancer diagnosis, identifies the research problem and existing gaps, presents the thesis contributions, and provides an overview of the thesis structure.
- **Chapter 1: Medical Image Diagnosis**: Offers a comprehensive overview of medical imaging modalities (X-ray, CT, MRI, ultrasound, PET), the clinical diagnostic workflow in medical imaging, the concept and evolution of Computer-Aided Diagnosis (CAD) systems, and clinical applications in brain, breast, and pulmonary imaging. This chapter establishes the clinical context and motivation for automated diagnostic systems.
- **Chapter 2: Medical Image Classification Techniques**: Provides a thorough review of classification approaches in medical imaging, including classical machine learning methods (SVM, Random Forests, k-NN), deep learning architectures (CNNs, Transformers), feature extraction methodologies (handcrafted vs. automatic), data augmentation techniques, and evaluation metrics. This chapter establishes the technical landscape and identifies limitations of single-architecture approaches and non-optimized methods.
- **Chapter 3: Bio-inspired Metaheuristics for Medical Image Classification**: Presents a comprehensive taxonomy of metaheuristic algorithms including evolutionary methods (Genetic Algorithms, Differential Evolution), swarm intelligence (Particle Swarm Optimization, Grey Wolf Optimizer, Whale Optimization), collective behavior algorithms (Ant Colony Optimization, Firefly Algorithm), and physical phenomena-inspired methods (Simulated Annealing, Harmony Search). This chapter reviews their applications in feature selection, hyperparameter optimization, and classifier design for medical imaging. It establishes the theoretical and empirical foundation for the proposed integration of metaheuristics with deep learning.
- **Chapter 4: Deep Learning for Medical Image Analysis**: Provides detailed technical exposition of deep learning architectures fundamental to the proposed approach, including Convolutional Neural Networks (CNNs), Vision Transformers with attention mechanisms, Capsule Networks and their spatial hierarchy modeling, advanced architectures (EfficientNet, U-Net), and their individual capabilities and limitations. This chapter explains the architectural components comprising HNet and motivates their integration into a hybrid framework.
- **Chapter 5: HNet-Based Breast Cancer Classification**: Presents the complete proposed HNet architecture, including detailed descriptions of the EfficientNet module for local feature extraction, the Advanced Vision Transformer (AVT) module for global context, the Capsule Network component for spatial relationships, and the novel integration of Genetic Algorithm-based feature selection between AVT-EfficientNet feature concatenation and CapsNet input. This chapter includes comprehensive experimental methodology, results on the BreakHis

dataset (binary and multi-class classification), detailed ablation studies, comparison with state-of-the-art methods, and analysis of computational efficiency. It demonstrates how metaheuristic-optimized hybrid deep learning achieves superior classification performance for breast cancer histopathology.

- **Conclusion and Future Works**: Synthesizes key findings, discusses implications for digital pathology and CAD systems, outlines limitations of the current work, and identifies promising directions for future research including extension to other cancer types, multimodal imaging integration, and clinical deployment considerations.

1. Medical Image Diagnosis

1.1. Introduction

Medical image diagnosis is a key part of modern healthcare since it helps doctors find, describe, and keep an eye on a wide spectrum of disorders across time. Imaging modalities, including radiography, ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), and nuclear medicine, provide non-invasive access to anatomical and functional data, facilitating earlier interventions and more accurate treatment planning compared to clinical examination alone [1].

In everyday practice, nevertheless, most diagnostic judgments still depend on how human specialists read pictures. This procedure is susceptible to inter and intra-observer variability, cognitive stress, and weariness, and it faces growing challenges due to the increasing volume and complexity of picture data in contemporary hospitals. It is possible to miss subtle or diffuse patterns, and it is not always possible to reproduce results across readers, institutions, and time. This can affect the accuracy of diagnoses and the health of patients [2].

These constraints have led to the creation of Computer-Aided Diagnosis (CAD) systems. These systems use computer methods to find, measure, and analyze picture elements to give a standardized "second opinion" that adds to, but doesn't replace, human knowledge. CAD has already shown its therapeutic relevance in a number of areas, such as screening for breast cancer and lung cancer [4]. It continues to grow as machine learning and deep learning improve. The aim of this chapter is to create a broad conceptual framework for image-based diagnosis, outlining the process from picture capture to clinical decision-making. This will set the stage for the computational methods that will be discussed in later chapters.

1.2. Overview of Medical Imaging Modalities

1.2.1. X-ray Imaging

Based on the differential attenuation of ionizing electromagnetic radiation as it travels through the body, X-ray imaging creates a two-dimensional projection image in which low-density structures like air-filled lungs appear radiolucent (dark) and dense tissues like bone appear radiopaque (white). When electrons are accelerated into a metal target (often tungsten) in an X-ray tube, bremsstrahlung and characteristic interactions result in a polyenergetic photon beam that is shaped and filtered before it reaches the patient. A detector (film-screen, computed radiography, or flat-panel digital detector) records the remaining transmitted photons to create the radiographic image. As the beam passes through various tissues, photoelectric absorption and Compton scattering remove photons from the primary beam in a tissue-dependent manner [5].

Figure 1.1 Hand Bone Imaging Using X-ray

The figure 1.1 shows a bilateral hand X-ray, in which ionizing radiation is used to visualize the bony structures of both hands. The image highlights the phalanges, metacarpals, and carpal bones with high contrast, allowing assessment of fractures, deformities, and degenerative changes.

Plain radiography is still one of the most used imaging modalities in clinical settings because it is quick, affordable, and easily accessible. It also has good spatial resolution and a high sensitivity for fractures, lung disease, and some foreign items. Its primary drawbacks are the use of ionizing radiation, the projection nature of the pictures (superposition of structures), and comparatively low soft-tissue contrast, which can make it difficult to characterize soft-tissue disease or mask minor lesions. To avoid artifacts like motion blur and scatter-induced contrast loss and to balance picture quality against radiation dosage, exposure parameters (tube voltage, current-time product, filtration, and source-to-image distance) must be optimized [6].

X-ray imaging remains one of the most widely used diagnostic modalities due to its speed, low cost, and utility in evaluating skeletal structures, the thorax, and dental anatomy. Clinically, it is a standard tool for detecting fractures, assessing lung pathology, and screening for common thoracic conditions such as infections and heart enlargement. Its strengths include rapid acquisition, broad accessibility, and good spatial resolution for high-density tissues. However, X-ray radiography has several weaknesses: it provides poor soft-tissue contrast, involves exposure to ionizing radiation, and produces two-dimensional projections that can obscure pathology due to overlapping anatomical structures. Subtle lesions may be missed, and image quality can be affected by noise, motion, and patient positioning.

1.2.2. Computed Tomography (CT)

Computed Tomography (CT) is an X-ray-based tomographic imaging modality in which a rotating X-ray tube and a ring of detectors acquire multiple projections around the patient that are reconstructed into cross-sectional images. CT estimates the spatial distribution of X-ray attenuation coefficients, expressed in Hounsfield Units, and provides high-resolution slices that can be reformatted in arbitrary planes or as three-dimensional volumes, offering detailed visualization of bone, lungs, and many

soft-tissue structures, as shown in figure 1.2. Modern multidetector CT scanners enable very fast acquisitions, making CT a cornerstone in emergency imaging, oncologic staging, trauma assessment, and pre-operative planning, although the use of ionizing radiation and iodinated contrast agents requires careful dose optimization and patient selection [7].

Figure 1.2 Contrast enhanced CT scan, demonstrating an abdominal aortic aneurysm

CT is a first-line imaging modality in emergency medicine due to its rapid acquisition, high spatial resolution, and excellent visualization of bone, calcifications, and acute hemorrhage. Clinically, it is indispensable for trauma assessment, cerebrovascular emergencies, lung imaging, abdominal pathologies, and oncological staging. Its major strengths include fast imaging times, wide availability, and highly detailed anatomical representation. However, CT's weaknesses stem mainly from the use of ionizing radiation, which limits repeated examinations, especially in children and pregnant patients. Its soft-tissue contrast is inferior to MRI, and images can suffer from artifacts such as beam hardening, streaking near metal implants, and issues caused by patient motion. Additionally, iodinated contrast agents may cause allergic reactions or nephrotoxicity in vulnerable patients.

1.2.3. Magnetic Resonance Imaging (MRI)

A powerful static magnetic field, gradient fields, and radiofrequency (RF) pulses are used in magnetic resonance imaging (MRI), a tomographic imaging technique that primarily takes advantage of the nuclear magnetic properties of hydrogen protons in fat and water to produce detailed cross-sectional images of internal anatomy. Tissue protons align with the main field when a patient is placed in the magnet. This net magnetization is then tipped away from equilibrium by an RF pulse, and the signal released during relaxation ,which is indicated by T1 and T2 times, is spatially encoded by gradients and reconstructed into images with superior soft-tissue contrast [8].

Due to its high intrinsic soft-tissue contrast, multiparametric capabilities (T1/T2 weighting, diffusion, perfusion, spectroscopy), and lack of ionizing radiation, magnetic resonance imaging (MRI) is especially useful for brain, as shown in figure 1.3, spinal cord, musculoskeletal, cardiac, and pelvic imaging. When choosing between MRI and other modalities in a diagnostic pathway, one must take into account its limitations, which include relatively long acquisition times, higher cost, sensitivity to motion and metal-induced artifacts, and contraindications in patients with specific implants or severe claustrophobia [6].

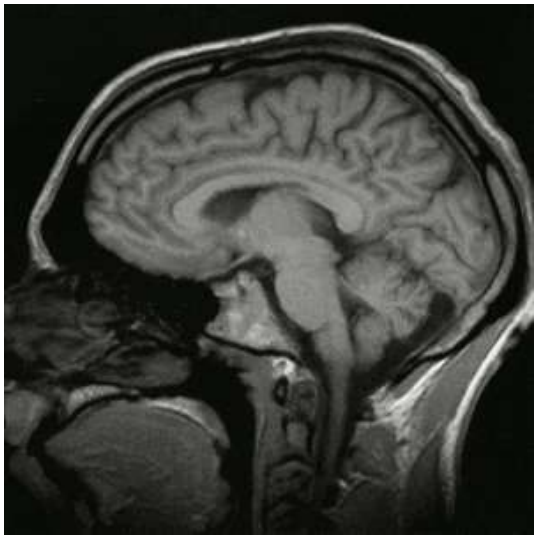


Figure 1.3 Brain MRI

MRI is widely used across clinical disciplines due to its exceptional ability to visualize soft tissues with high contrast, making it particularly valuable in neuroimaging, musculoskeletal assessment, cardiovascular evaluation, and oncological applications. Its primary strengths include excellent soft-tissue differentiation, multiple contrast mechanisms (T1, T2, diffusion, perfusion), and the ability to acquire functional and quantitative information without ionizing radiation. MRI is also highly versatile, enabling advanced techniques such as fMRI, DWI, MRS, and cardiac cine imaging. However, MRI has several weaknesses: long acquisition times, sensitivity to motion, high operational costs, and contraindications for patients with certain implants or metallic objects. Additionally, artifacts such as susceptibility distortions, motion blur, and inhomogeneities can reduce diagnostic reliability.

1.2.4. Ultrasound (US)

Ultrasound (US) is an imaging technique that creates real-time pictures of interior structures by using high-frequency mechanical sound waves that are produced and received by piezoelectric crystals in a

transducer. Short ultrasonic pulses are sent into the body by the probe, which then listens for echoes that are reflected back at tissue interfaces. The timing and amplitude of these echoes are processed to create a grayscale picture whose brightness represents the echogenicity of the tissue [9].

Clinically, ultrasound is widely used in obstetrics, cardiology, abdominal, vascular, and musculoskeletal imaging because it is portable, relatively inexpensive, and does not involve ionizing radiation, making it suitable for bedside and repeated examinations. Its main limitations include strong operator dependence, reduced image quality in obese patients or in regions with poor acoustic windows (e.g., air-filled bowel, bone), and susceptibility to artifacts such as acoustic shadowing, reverberation, and speckle, which can obscure or mimic pathology. The ability to capture real-time motion also makes it valuable for guiding minimally invasive procedures. However, ultrasound has notable weaknesses: it is highly operator-dependent, has limited penetration in obese patients, and cannot effectively image structures obscured by bone or gas. Image quality is often affected by acoustic shadowing, speckle noise, and variability in probe placement, which can reduce diagnostic consistency [9].

1.2.5. Nuclear Medicine Imaging: PET and SPECT

By providing tomographic pictures of the in vivo distribution of radiotracers, nuclear medicine imaging methods like positron emission tomography (PET) and single photon emission computed tomography (SPECT) enable the visualization of physiological and molecular processes rather than merely anatomy. In PET, a biologically active molecule is linked to a positron-emitting radionuclide (such as fluorine 18 labeled fluorodeoxyglucose, or 18F FDG). Following injection, the positrons are annihilated with electrons to produce pairs of 511 keV photons that are detected in temporal coincidence around the patient, enabling the reconstruction of three-dimensional activity maps. SPECT uses spinning gamma cameras with collimators to scan gamma-emitting tracers like technetium 99m or iodine 123. These cameras capture single photons from various angles, which are subsequently reconstructed into volumetric pictures of tracer uptake [10].

PET is often utilized in cancer, neurology, and cardiology for tumor staging, therapeutic response evaluation, and measurement of brain metabolism or receptor expression because it provides more sensitivity and spatial resolution than SPECT. Though its resolution and quantitative accuracy are usually lower, SPECT is generally more affordable and more widely used in cardiac perfusion imaging, bone scans, and targeted radionuclide therapy. In order to enable accurate anatomical localization and attenuation correction of functional signals, both modalities are now frequently combined with CT (PET/CT, SPECT/CT) and, more recently, with MRI (PET/MRI). This strengthens their role in integrated functional–anatomical characterization within precision medicine workflows [11].

PET and SPECT provide functional and molecular imaging by detecting radioactive tracers, allowing clinicians to visualize metabolic activity, perfusion, receptor binding, and other physiological processes. These modalities are highly valuable in oncology for tumor detection, staging, and treatment monitoring, as well as in neurology for assessing neurodegenerative diseases and in cardiology for evaluating

myocardial perfusion. Their greatest strengths lie in their ability to detect early biochemical changes before structural abnormalities appear, offering high sensitivity for disease processes. However, their weaknesses include low spatial resolution compared to CT and MRI, dependence on radiopharmaceutical availability, high cost, and significant radiation exposure. PET and SPECT images often require fusion with CT or MRI for anatomical localization, and image quality can be degraded by noise, scatter, and motion artifacts.

1.3. Clinical Diagnostic Workflow in Medical Imaging

In medical imaging, the clinical diagnostic workflow is an organized series of procedures intended to guarantee effective, precise, and secure patient condition evaluation. It offers the framework for the smooth integration of CAD and AI-assisted technologies. It incorporates clinical decision making, picture capture, technological optimization, expert interpretation, and follow-up activities. Therefore, when developing intelligent technologies that really assist radiologists rather than interfere with ordinary practice, a thorough grasp of this process is crucial [12], [13].

The workflow begins with the clinical indication and referral, where the physician evaluates symptoms, medical history, and risk factors to decide whether imaging is justified and which modality is most appropriate [14]. Evidence-based referral criteria and decision-support systems help select between MRI, CT, ultrasound, X-ray, or PET/SPECT according to the suspected pathology, patient characteristics, and radiation-protection principles of justification and optimization, thereby reducing unnecessary or redundant examinations. Proper modality selection at this stage improves diagnostic yield, minimizes radiation exposure, and streamlines downstream workflow [15].

After an examination is warranted, image acquisition is carried out, usually by technicians or radiographers who set up the scanner, prepare the patient, and carry out the scan. In order to balance image quality, examination time, and patient safety, acquisition protocols, such as slice thickness, kV/mAs, contrast-agent timing for CT, pulse sequences for MRI, or transducer selection for ultrasound, must be adjusted. Inadequate positioning, motion, or suboptimal parameters can introduce artifacts that compromise diagnostic accuracy. Numerous studies highlight how important radiographers and procedure standardization are to preserving uniform picture quality throughout patients and institutions [16].

To produce pictures that are clinically interpretable, raw measures are reconstructed and preprocessed after capture. This comprises noise reduction, contrast enhancement, spatial normalization, k-space processing and artifact correction in MRI, and tomographic reconstruction methods in CT, PET, and SPECT. When using CAD and quantitative imaging biomarkers, advanced iterative and model-based reconstruction techniques, which are frequently enhanced by machine learning, may enhance picture quality and dosage efficiency, but they can also change noise texture and potentially impact quantitative data [17].

In order to arrive at a diagnosis or differential diagnosis, radiologists methodically evaluate pictures, compare them with previous studies, and integrate imaging results with clinical information. This process forms the basis of the workflow. At this point, CAD and AI systems are being incorporated more and more to help with tasks like lesion detection, segmentation, and disease classification. Research indicates that careful integration into the reading environment can decrease inter-observer variability and interpretation time while preserving or increasing diagnostic accuracy. However, the degree to which they are in line with current reporting norms and reading habits will determine how beneficial they are [18].

After interpretation, radiologists provide an organized report that conveys important discoveries, diagnostic conclusions, and evidence-based suggestions for additional imaging or therapy. A number of studies emphasize the necessity of systems that track follow-up suggestions to prevent loss to follow-up, and timely and clear reporting is crucial, especially in time-sensitive situations like stroke or trauma. In the last stage, imaging results are incorporated into more comprehensive clinical management, directing therapeutic choices, surgical planning, treatment response monitoring, and long-term surveillance, frequently in multidisciplinary settings like tumor boards. For many diseases, this results in a cyclical imaging pathway whereby repeated studies improve diagnosis and customize care over time. To guarantee that CAD and AI tools improve patient outcomes, efficiency, and diagnostic reliability rather than acting as standalone technological add-ons, it is crucial to design them with this end-to-end process in mind [1].

1.4. Computer-Aided Diagnosis (CAD) Systems

By automatically identifying, classifying, or quantifying problematic findings and presenting them as a "second reader" rather than a substitute for human competence, computer-aided diagnosis (CAD) systems are computational tools created to help doctors analyze medical pictures. In order to support tasks like lung nodule identification, polyp detection in CT colonography, and mammographic microcalcification detection, early CAD research concentrated on handcrafted features and classical machine learning. The main objectives were to decrease perceptual oversight and increase sensitivity while keeping a reasonable false-positive rate [19].

Preprocessing, segmentation, feature extraction, and classification are all conceptually integrated into an end-to-end pipeline by CAD systems, which produce lesion-level or patient-level judgments that radiologists can accept, reject, or look into further [20]. In order to achieve significant improvements in detection and classification accuracy across modalities like X-ray, CT, MRI, ultrasound, and digital pathology, modern CAD increasingly depends on data-driven feature learning from large annotated datasets thanks to the development of deep learning, especially convolutional neural networks. These developments have made it possible for clinically tested systems to perform as well as or better than skilled radiologists in tasks like lung nodule identification, breast cancer screening, and TB triage [21].

Recent research highlights the need for CAD to be closely integrated into radiology processes, going beyond simple performance and addressing concerns of usability, interpretability, and prospective validation prior to widespread deployment [22]. Large, varied training cohorts, resilience to domain transition, handling false positives, and a precise description of CAD's function, whether as a contemporaneous reader, triage, or retrospective quality-assurance tool, are among the difficulties. However, CAD and AI-enhanced systems are generally recognized as important facilitators of precision imaging, providing more quantitative, objective, and repeatable image evaluation that can support treatment planning, risk assessment, and long-term response monitoring across a wide range of diseases [23].

1.4.1. Conceptual Definition of CAD

CAD systems can be conceptually defined as computer-based decision-support tools that evaluate medical images and produce quantitative results, such as candidate lesions, probability scores, or measurements, with the goal of supporting the interpreting radiologist rather than taking their place. By highlighting subtle abnormalities and providing standardized, algorithmically derived information, CAD serves as a "second reader" that enhances human visual assessment in this paradigm [24]. The radiologist then combines these outputs with clinical data, previous studies, and personal expertise. As a result, the radiologist continues to be the ultimate decision-maker and is fully accountable for the diagnosis and report. The decrease of human variability in picture interpretation is a major driving force behind CAD. Numerous studies have demonstrated that fatigue, experience level, and reading circumstances have an impact on observer performance in tasks including lung nodule diagnosis, polyp identification, and mammography screening, resulting in significant inter- and intra-observer variability. When used properly as an assistive tool, CAD can help standardize detection thresholds and quantitative assessments by applying consistent algorithms across all cases. This improves reproducibility and, in many settings, increases sensitivity without a clinically unacceptable rise in false positives [25].

1.4.2. Clinical and Historical Context

Growing demands on radiology services, rising picture volumes, and worries about perceptual mistakes in image interpretation gave rise to Computer-Aided Diagnosis (CAD). Early research in the 1980s and 1990s concentrated on chest radiography and mammography, where missed lesions in screening programs brought attention to the need for instruments that might serve as a reliable "second reader" to increase sensitivity without unduly adding to the burden. Significant research in mammographic CAD showed that computer-generated prompts might assist radiologists in identifying more malignancies.

This led to regulatory approvals and extensive clinical implementation, especially in breast cancer screening programs in the US and Japan [20].

Thanks to developments in image processing, machine learning, and computing power, CAD ideas evolved throughout time from basic computer-aided detection to more complex systems capable of characterization and diagnosis across modalities including CT, MRI, and colonography. However, significant drawbacks of traditional CAD systems were also revealed, such as high false-positive rates, inconsistent effects on reader performance, and difficulties integrating into regular workflow. These problems have spurred critical reevaluation and the shift to contemporary AI-driven systems. CAD is now seen as part of a larger artificial intelligence ecosystem in medical imaging, where data-driven deep learning models seek to provide quantitative biomarkers and decision support in addition to lesion detection, while clinical and regulatory guidelines place an increasing emphasis on human oversight, transparency, and prospective validation [26].

1.4.3. Types of CAD

1.4.3.1. CADe: Computer-Aided Detection

The main purpose of computer-aided detection (CADe) systems is to automatically identify and locate questionable areas in medical pictures, serving as a "second reader" that alerts radiologists to possibly anomalous discoveries. In order to reduce observational errors and false-negative interpretations, CADe uses pattern-recognition algorithms to analyze the entire image volume and produce marks or candidate lists for structures like pulmonary nodules, mammographic masses or microcalcification clusters, and colonic polyps [27]. In order to maintain human accountability while taking use of the system's remarkable sensitivity to minute anomalies, radiologists in clinical practice usually initially analyze the pictures without assistance and then re-examine CADe-marked areas before delivering the final report [28].

1.4.3.2. CADx: Computer-Aided Diagnosis

Instead of concentrating on the initial identification of discoveries, Computer-Aided Diagnosis (CADx) systems concentrate on the diagnostic characterisation of previously discovered findings. Following the radiologist's or a previous CADe step's outline of a lesion or region of interest, CADx extracts quantitative features (such as intensity, shape, texture, and kinetics) and uses classification models to estimate the likelihood of malignancy, assign a pathology class, or stratify risk. While the radiologist remains the ultimate arbiter of the diagnosis, the resulting probabilities or decision scores are presented as decision-support information that can help standardize interpretation, reduce inter-observer variability, and support management decisions like biopsy recommendation or imaging follow-up [29]. CADe and CADx systems are often integrated into a sequential diagnostic pipeline to complement each other’s strengths, as summarized in table 1.1. In this workflow, CADe first scans medical images to

detect and highlight potential abnormal regions, serving as a preliminary step that ensures suspicious areas are not overlooked. These candidate regions are then analyzed by CADx, which provides diagnostic classification, such as assessing the likelihood of malignancy or characterizing the type of lesion. By combining detection and classification, this integrated approach enhances overall diagnostic performance: sensitivity is improved through comprehensive identification of potential abnormalities, specificity is increased by accurately characterizing true lesions and reducing false positives, and radiologists benefit from greater efficiency and confidence in their decision-making.

Aspect	**CADe (Computer-Aided Detection)**	**CADx (Computer-Aided Diagnosis)**
Primary Goal	Detect suspicious regions or abnormalities	Classify or characterize detected abnormalities
Main Output	Highlighted/marked candidate regions	Diagnostic likelihood (e.g., benign vs malignant)
Focus	Sensitivity – ensuring no abnormality is missed	Specificity – reducing false positives and refining diagnosis
Role in Workflow	Acts as a "second reader" to the radiologist	Assists in decision-making and risk assessment
Typical Applications	Detection of tumors, nodules, microcalcifications, polyps	Lesion classification, malignancy prediction, disease staging
Required Model Complexity	Lower: often region proposal or anomaly detection	Higher: feature extraction, classification, decision prediction
Dependence on Prior Detection	Independent: detects candidates from scratch	Dependent: analyzes regions typically provided by CADe or radiologist
Strengths	Reduces oversight, increases sensitivity	Improves diagnostic accuracy and reduces false positives
Weaknesses	May generate many false positives	Requires high-quality input regions; risk of misclassification
Examples in Practice	Breast cancer screening (microcalcification detection), lung nodule detection	Malignancy scoring of breast lesions, classification of lung nodules

Table 1-1: Key Features and Differences Between CADe and CADx Systems

1.4.4. Clinical Benefits of CAD

Computer-Aided Diagnosis (CAD) systems offer several clinical benefits when they are properly validated and integrated into routine workflows.

- **Reduction of diagnostic errors**

By serving as an automated "second reader" that draws attention to small or easily missed lesions, CAD may reduce perceptual mistakes, particularly in screening activities like lung nodule identification and mammography. Research indicates that although careful tuning is required to prevent excessive false positives, CAD boosts sensitivity for many readers and aids in the detection of new tumors or lesions that would otherwise go undetected [30].

- **Automated double reading**

CAD offers a scalable substitute or supplement in situations where human double reading is logistically or financially challenging, providing consistent case-by-case assessment without additional human manpower. In high-volume screening systems, where CAD may methodically reexamine each picture and encourage radiologists to reevaluate highlighted regions prior to final reporting, this automated double reading is very helpful [31].

- **Faster analysis in high-workload settings**

CAD and contemporary AI techniques may save reading times and assist radiologists in managing growing workloads by pre-localizing worrisome spots, prioritizing studies, or pre-computing measures. It has been shown that workflow-integrated solutions that prioritize normal or low-risk patients or that provide quick quantitative analysis increase productivity while preserving diagnostic performance [21].

- **Better standardization of diagnosis**

CAD may decrease inter- and intra-observer variability by applying the same algorithms to all instances, resulting in more consistent measurements and decision thresholds across readers and institutions. Quantitative imaging biomarkers and risk-stratification schemes in contemporary precision medicine are supported by this standardization, which also makes multi-center investigations and longitudinal follow-up easier [17].

1.4.5. Scientific and Clinical Limitations of CAD

Conventional CAD systems are quite sensitive to acquisition circumstances, particularly those that rely on handmade features. Robustness between institutions and over time might be limited by minor changes to scanner technology, imaging procedures, reconstruction kernels, or noise characteristics, which can drastically affect picture appearance and thus deteriorate CAD performance. Reusing or redeploying a system without carefully fine-tuning or retraining on data obtained with the new settings is challenging due to this reliance [22].

Anatomical variability and picture imperfections are further challenges for CAD algorithms. While motion blur, beam hardening, partial volume effects, or ultrasonic speckle might produce false positives or conceal tiny lesions, normal variations, surgical alterations, or unique presentations may be mistaken

for disease. Because of this, radiologists must carefully interpret CAD results, and when faced with uncommon instances, systems built for a single organ, modality, or disease are often less accurate [20]. It is still very difficult to generalize to different populations. When applied to various demographics, disease prevalences, or imaging technologies, many CAD systems may perform worse since they were trained and tested on comparatively homogenous datasets from single sites. These systems are less adaptable than contemporary deep learning-based methods because their architectures and feature sets are closely linked to the original design, making it challenging to adapt them to new tasks or anatomies. Strong clinical validation is crucial and resource-intensive from a translational perspective. To prove safety, effectiveness, and influence on clinical outcomes, prospective multi-center trials, reader-study assessments, and post-market monitoring are necessary; nevertheless, these studies involve significant effort, data, and funds. Reproducibility, explainability, and clearly specified indications for use are further stressed in regulatory approval procedures, which many early CAD systems failed to completely meet. This underscores the need of thorough, continuous review prior to broad deployment [25].

1.6. Clinical Applications of Medical Image Diagnosis

Early CAD applications focused on tasks such as breast cancer detection in mammography, lung nodule identification in chest CT scans, and brain tumor segmentation in MRI images [32], [33]. More recent studies have expanded CAD usage to cardiovascular diseases, retinal disorders, and musculoskeletal abnormalities, demonstrating its potential to improve diagnostic accuracy, reduce inter-observer variability, and support clinical decision-making. As a result, CAD has become an important complementary tool in medical image diagnosis rather than a replacement for clinicians [34].

1.6.1. Brain Imaging

Medical imaging is central to the diagnosis and management of neurological diseases, including stroke, brain tumors, demyelinating disorders, epilepsy, and neurodegenerative diseases. CT is widely used for rapid assessment of acute stroke and traumatic brain injury, while MRI provides superior soft-tissue contrast for detecting white- and gray-matter abnormalities, characterizing tumors, and evaluating inflammatory or demyelinating lesions such as multiple sclerosis. Functional techniques such as fMRI and PET enable evaluation of cerebral perfusion, metabolism, and receptor binding, supporting presurgical planning in epilepsy and tumor resection, as well as research and diagnosis in movement and cognitive disorders [35].

1.6.2. Breast Imaging

In breast cancer screening and diagnosis, mammography remains the first-line modality for detecting microcalcifications and masses, often complemented by ultrasound in dense breasts and MRI in high-risk patients or for preoperative staging. Digital breast tomosynthesis, contrast-enhanced

mammography, and breast MRI improve lesion conspicuity and assessment of multifocal or multicentric disease, while ultrasound elastography and MRI-based functional techniques contribute information on lesion stiffness, vascularity, and kinetics that aid in differentiating benign from malignant findings. These imaging tools play a pivotal role across the continuum of care, from population screening and diagnostic work-up to treatment planning, response assessment, and surveillance for recurrence [36].

1.6.3. Pulmonary Imaging

Thoracic imaging is essential for evaluating a broad spectrum of pulmonary and mediastinal diseases, including infections, interstitial lung disease, pulmonary embolism, and primary or metastatic lung cancer. Chest radiography is typically used as an initial, low-dose survey, while CT, particularly high-resolution and low-dose protocols, provides detailed assessment of lung parenchyma, airways, and pulmonary vasculature, and underpins lung-cancer screening programs. PET/CT further contributes metabolic information for staging lung cancer, assessing treatment response, and differentiating active disease from post-treatment changes, reinforcing the combined anatomical and functional role of imaging in pulmonary diagnosis and management [36].

1.7. Conclusion

A broad conceptual framework for medical image diagnosis has been presented in this chapter, connecting the clinical functions and physical principles of the main imaging modalities with the actual diagnostic process from referral to follow-up. Computer Aided Diagnosis (CAD) systems were defined within this framework as decision support tools that supplement human visual assessment, taking into account their functional architectures, the differences between CADe and CADx, and the ways in which they can both introduce and reduce variability in image interpretation. The chapter highlights the need for reliable, transparent, and clinically verified algorithmic solutions by looking at actual clinical applications, describing the clinical and scientific limits of existing systems, and briefly discussing regulatory and ethical issues. The shift to next chapters, which will concentrate on the development, training, and assessment of machine learning and deep learning techniques for image-based diagnosis, is made possible by this preparation.

2 Medical images classification techniques

2.1 Introduction

Image processing represents the foundational field enabling all image-based analysis, including a wide set of computational and mathematical techniques designed to enhance, transform, and extract meaningful information from digital images. Image processing covers basic operations including noise reduction, contrast enhancement, edge detection, and feature extraction, enabling the preparation and conditioning of raw pixel data for subsequent analysis. These preprocessing techniques are essential for improving image quality, standardizing data representations, and making visual information more accessible for human and machine interpretation. By systematically manipulating image data through filtering, segmentation, and enhancement operations, image processing transforms raw visual inputs into refined representations suitable for higher-level analysis and decision-making across diverse applications ranging from satellite imagery to medical diagnostics[1].

Building upon the foundation of image processing, image classification emerges as a fundamental machine learning task that assigns categorical labels to digital images based on their visual content and extracted features. Image classification represents the process of automatically categorizing images into predefined classes through machine learning algorithms, leveraging handcrafted features (edges, textures, shapes) or automatically-learned feature representations (deep learning) to make predictions. This classification process typically involves four sequential stages: image preprocessing and enhancement, feature extraction or learning, classifier training, and performance evaluation. Image classification finds widespread applications across numerous domains, from autonomous vehicle object detection to industrial quality control, facial recognition systems, agricultural crop monitoring, and content-based image retrieval, demonstrating the fundamental importance of accurate image categorization across diverse sectors [2].

Medical image classification, a specialized application of image classification adapted for healthcare contexts, represents a critical clinical tool that leverages advanced machine learning and deep learning methodologies to automatically analyze medical images for disease detection, diagnosis, and treatment planning. Medical image classification extends beyond general-purpose image analysis by incorporating domain-specific knowledge about human anatomy, pathological patterns, and clinical requirements, combined with sophisticated algorithms to categorize medical images according to diagnostic criteria established by radiologists and pathologists. Unlike generic image classification, medical image classification must address unique challenges including extreme class imbalance (rare diseases represented by small data samples), stringent accuracy requirements (diagnostic errors carry serious clinical consequences), complex multi-scale features (requiring detection of both large-scale anatomical structures and subtle pathological indicators), and regulatory compliance requirements (ensuring model transparency and clinical validity). The transformative impact of medical image classification on healthcare delivery stems from its capacity to automate analysis of high-volume imaging data, reduce

human error, improve diagnostic consistency, enable early disease detection, and ultimately enhance patient outcomes through more accurate and timely clinical decision-making[39].

The application of classification algorithms greatly enhances decision-making processes, and their use goes far beyond the theoretical domain. AI-based categorization systems, for instance, may help with traffic management in the transportation industry by forecasting patterns of congestion and enabling more effective routes. By lowering emissions from idle cars, these developments not only maximize operational effectiveness but also support sustainable urban planning initiatives [40].

In this chapter, we explore the comprehensive landscape of medical image classification techniques, examining traditional machine learning approaches, state-of-the-art deep learning architectures, and complementary methodologies essential for developing robust diagnostic systems. We discuss the fundamental importance of classification in medical imaging contexts, identify specific technical and clinical challenges inherent to medical image analysis, present contemporary classification models spanning both classical and neural network-based approaches, and establish evaluation frameworks critical for clinical validation. Through systematic examination of data augmentation strategies, feature extraction paradigms, and performance metrics, this chapter provides a structured foundation for understanding how advanced classification methodologies revolutionize medical imaging practice and contribute to precision medicine initiatives globally.

2.2 Classification in general

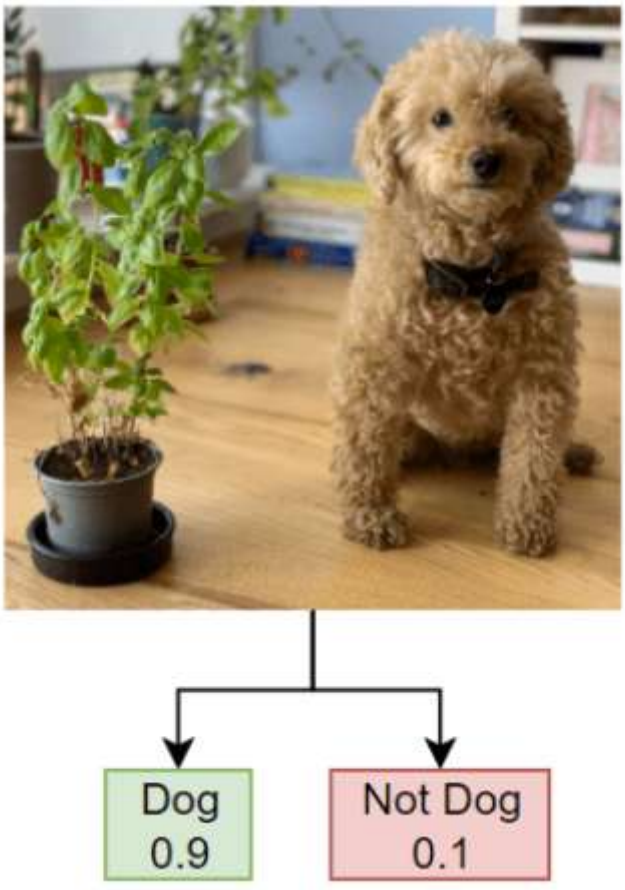
In artificial intelligence, classification is the process of giving a data point a label or category based on its attributes and predetermined criteria. Classification is essential to data processing because it makes it possible to organize, analyze, and understand information sets, which makes it easier to make decisions and automate difficult processes. Classification is very important in medical imaging since it serves as the foundation for clinical decision assistance, early pathology identification, and guided diagnosis. This chapter's goals are to outline the fundamental concepts of categorization, examine the several methods that are used, and demonstrate how they are applied in the particular subject of medical imaging [40], [41].

Formally, classification involves establishing a correspondence between input data and predefined labels or classes. Binary classification (two classes), multiclass classification (more than two classes), and multilabel classification (several alternative labels for a single data point) are some of the many forms of classification. A classification pipeline's basic procedure consists of the following steps: preprocessing the data, extracting pertinent features, classifying the data, and then evaluating the model's performance. This methodical approach guarantees the durability of models applied to complicated medical data and optimizes the quality of results [42].

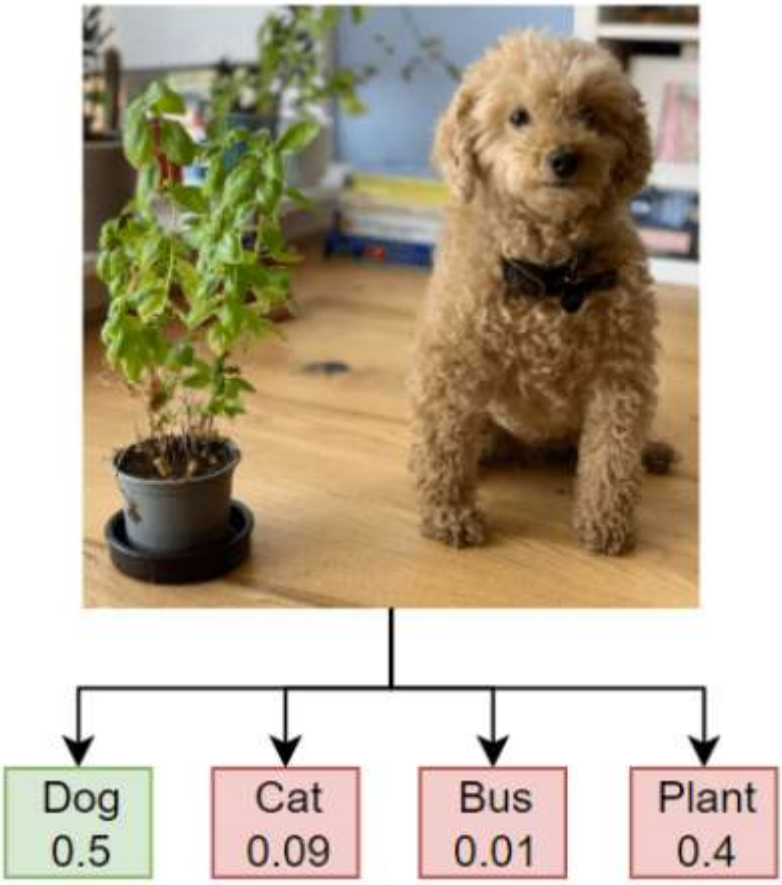
Classification types

Classification tasks in machine learning are fundamentally categorized into three distinct types based on the nature of the target variable and the number of permissible class assignments.

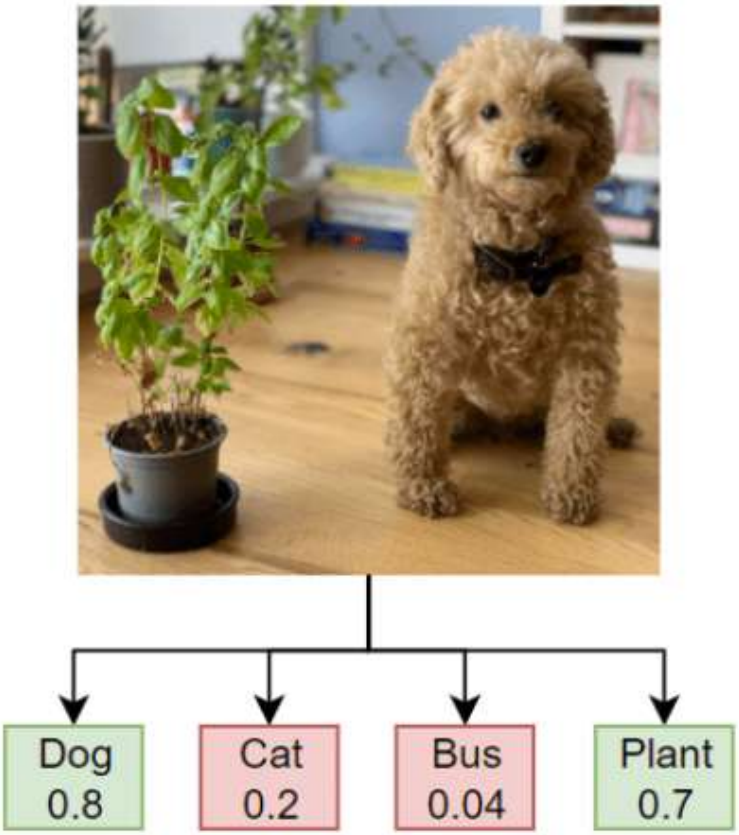
Binary classification represents the foundational classification paradigm, where data instances are categorized into exactly two mutually exclusive classes, such as spam versus non-spam emails or benign versus malignant tumors in medical imaging. Binary classifiers traditionally employ techniques such as logistic regression, SVM, and binary decision trees, utilizing sigmoid activation functions and binary cross-entropy loss functions during training[43].



Multiclass classification extends this framework to scenarios involving three or more classes, requiring each instance to be assigned to precisely one class from a predefined set of categories. Multiclass problems are typically addressed using softmax activation functions, categorical cross-entropy loss, and algorithms such as multinomial logistic regression, k-nearest neighbors, random forests, and gradient boosting methods [44].



Multilabel classification represents a more complex scenario where a single instance can be simultaneously assigned multiple labels, reflecting real-world problems where objects possess overlapping or shared characteristics. Examples include assigning multiple topic tags to documents, identifying multiple objects within an image, or categorizing movies with multiple genres. Multilabel classification employs specialized approaches such as Binary Relevance (BR), which decomposes the multilabel problem into multiple independent binary classification tasks; Classifier Chains (CC), which account for label dependencies; and Label Powerset (LP), which transforms multilabel problems into multiclass single-label problems. Additionally, adapted neural network variants and ensemble methods are employed to capture label interdependencies and improve prediction accuracy [44], [45].



2.3 Importance of classification in medical imaging

Classification in medical imaging is crucial for improving diagnostic accuracy, enabling early disease detection, and supporting clinical decision-making. By automating the analysis of medical images, classification algorithms help reduce human error, increase diagnostic efficiency, and facilitate the management of large volumes of data in clinical settings. Advanced classification techniques, particularly deep learning and convolutional neural networks, have demonstrated remarkable performance in critical tasks including tumor detection, organ segmentation, and disease classification, fundamentally transforming healthcare workflows and significantly impacting patient outcomes. Furthermore, the integration of machine learning-based classification systems into clinical practice has been shown to enhance radiologist productivity, reduce diagnostic variability, and enable more timely interventions, making it an indispensable component of modern medical imaging infrastructure [46], [47].

The importance of classification in medical imaging is underlined by its fundamental role in improving diagnostic processes through the systematic categorization of medical data. The evolution of medical imaging techniques from traditional modalities such as X-rays and CT scans to advanced imaging technologies such as MRI and PET scans has facilitated unprecedented insights into human anatomy and pathology. However, the increasing complexity and volume of imaging data requires sophisticated approaches for effective interpretation and clinical decision making. In this context, classification

algorithms have emerged as essential tools that leverage large amounts of imaging data to assist healthcare providers in diagnosing diseases more accurately and efficiently.

Recent advances in machine learning and deep learning methodologies have further revolutionized classification in medical imaging. These approaches use complex neural networks that can learn hierarchical features from imaging data, thus outperforming traditional image processing techniques in terms of accuracy and speed [48]. The translation of these algorithms into clinical practice is indicative of a paradigm shift, as they enable automated analysis, significantly reducing the reliance on manual interpretations that are vulnerable to human error. For example, CNN have been effectively employed to classify images for various diseases, such as identifying malignant tumors in mammograms or distinguishing between types of pneumonia in chest x-rays.

Furthermore, the integration of classification algorithms not only improved diagnostic accuracy but also contributed to personalized treatment planning. By using data from different patient demographics and leveraging large datasets from various institutions, machine learning models can provide more personalized recommendations, thereby improving patient outcomes. This highlights a critical intersection between technology and personalized medicine, where classification algorithms function as powerful decision support systems. Their ability to process and classify imaging data at scale represents an essential evolution in the healthcare landscape, facilitating a more precise and dynamic understanding of patient health.

In summary, the continued integration of classification methodologies into medical imaging reflects an increasing reliance on technological innovations to address traditional limitations in diagnostic practices. As healthcare continues to embrace these advances, the potential to improve patient care through greater diagnostic accuracy and informed treatment strategies becomes increasingly tangible. These developments promise not only to transform clinical workflows, but also pave the way for future innovations that will further streamline healthcare delivery. The interaction between classification, imaging modalities and machine learning advances therefore represents a significant growth point for modern medicine, deserving continued exploration and investment in this vital research area. The application of advanced classification techniques in medical images has revolutionized diagnostic accuracy, particularly with the incorporation of deep learning models that significantly outperform traditional image analysis methods. Recent studies illustrate the effectiveness of these approaches in identifying various diseases from various imaging modalities, thereby providing a robust platform for clinical decision making [13].

The continued integration of classification methodologies into medical imaging reflects an increasing and justified reliance on technological innovations to systematically address longstanding limitations inherent in traditional diagnostic practices. By automating image analysis and feature extraction, classification algorithms help reduce human error, increase diagnostic efficiency, and facilitate the

management of large volumes of data in complex clinical settings ultimately enabling early disease detection, accurate treatment planning, and improved patient outcomes. As healthcare systems continue to embrace and implement these technological advances, the potential to improve patient care through greater diagnostic accuracy, reduced diagnostic variability, and more informed treatment strategies becomes increasingly tangible and measurable. These developments promise not only to fundamentally transform clinical workflows and radiologist productivity, but also to pave the way for future innovations including real-time diagnostic tools, personalized medicine approaches, and advanced clinical decision support systems that will further streamline healthcare delivery and expand access to high-quality diagnostics globally. The dynamic interaction between classification methodologies, diverse imaging modalities, and accelerating advances in machine learning therefore represents a significant and increasingly vital growth point for modern medicine, deserving continued exploration, research investment, and clinical validation in this transformative area[48].

2.4 Specific challenges in medical imaging classification

Medical image classification faces several interconnected and compounding challenges that significantly impact the development and deployment of robust diagnostic algorithms.

Image noise represents a fundamental and pervasive challenge, arising from inherent limitations of imaging acquisition equipment, sensor artifacts, and environmental factors. Medical images are often corrupted by Gaussian noise, Poisson noise, and shot noise. To maintain diagnostic information while reducing noise abnormalities, sophisticated denoising approaches including Gaussian filtering, bilateral filtering, and deep learning-based denoising networks are required [50].

Class imbalance is one of the biggest and most common problems in medical imaging classification. Disease prevalence often exhibits an uneven distribution in clinical practice [51]. For example, malignant samples greatly outweigh benign samples in the BreakHis breast cancer histopathology dataset, while normal/healthy cases greatly outnumber pathological cases in general medical imaging. Despite obtaining excellent overall accuracy, conventional deep learning models trained on unbalanced data often show significant bias toward majority classes and low sensitivity and specificity for minority classes [52]. Imbalance-aware loss functions derived from classification metrics like the Matthews correlation coefficient (MCC) and F1 score can significantly improve performance, as shown by a thorough study by Scholz et al. [53]. When combined with cross-entropy loss, these loss functions can achieve improvements of up to 12% in balanced accuracy and up to 51% in class-wise F1 score for minority classes. To successfully address this issue, recent sophisticated approaches include strategies including oversampling of minority classes, undersampling of majority classes, synthetic data generation (SMOTE), cost-sensitive learning with class-weighted losses, focused loss functions, and ensemble methods [54].

Intra-class variability describes the notable heterogeneity and diversity seen within each diagnostic category, which may be caused by a variety of reasons, including changes in patient anatomy, variations in illness presentation, variations in imaging protocols, and variations in scanner manufacturers. Medical pictures show both intra-operator and inter-operator variability [55]. Depending on the particular imaging job and anatomical location, intraclass correlation values might range from low to outstanding. Significant intra-class variations and subtle inter-class differences in medical images are explicitly addressed by the SequencesNet model proposed for MRI sequence classification, which shows that accounting for these variations through fine-grained prototype learning significantly improves classification accuracy and generalization [21]. Sophisticated methods like data augmentation strategies, domain adaptation techniques, transfer learning from large datasets, and architecture designs that explicitly capture and preserve discriminative features while remaining robust to class-specific variations are needed to address intra-class variability [57].

Dataset size greatly impacts model performance, generalization, and clinical deployment, which is a basic and enduring problem in medical image categorization. Developing robust deep learning models requires the availability of sufficiently large, well-curated, and representative training datasets [58]. However, healthcare environments often face severe data scarcity because of annotation costs, privacy regulations (HIPAA, GDPR), and the fragmented nature of medical imaging data across institutions. Using CNN on computed tomography (CT) images across six anatomical classes, a groundbreaking study by Cho et al. methodically examined the relationship between training dataset size and classification accuracy in medical image analysis[59].

Additionally, these issues often work in concert: class imbalance and noisy labels produce particularly difficult situations where minority class samples are more likely to be incorrectly classified as noisy, which results in their removal from training data and exacerbates the imbalance issue. Collaborative learning frameworks with curriculum-based sample selection and noise balance losses that leverage rather than discard potentially noisy minority class samples are examples of integrated solutions needed to address these compounded challenges, as noted by Er et al. in their framework for imbalanced medical image segmentation with pixel-wise noisy labels [60].

2.5 Medical images classification process

The medical image classification process is usually described as a pipeline of sequential stages that transform raw data into clinically meaningful predictions. These stages range from image acquisition and preprocessing to feature extraction, classification, and rigorous performance evaluation. Each step has a direct impact on the robustness, generalization ability, and clinical usefulness of the final system.

2.5.1 Image acquisition

Images are obtained from several modalities such as X-ray, MRI, CT, mammography, ultrasound, PET, or histopathology scanners, often following standardized clinical protocols. Acquisition parameters (contrast agent, slice thickness, resolution) strongly influence image characteristics and later classification performance [61].

2.5.2 Preprocessing and segmentation

The raw images produced by clinical scanners frequently contain noise, intensity inhomogeneities, motion artifacts, and irrelevant background structures. Preprocessing aims to correct or reduce these factors in order to standardize the data before analysis. Typical operations include noise reduction (e.g. Gaussian, median, or non-local means filtering), intensity normalization and resampling to a common resolution, bias-field correction (in MRI), and image registration to a reference anatomical space. In many applications, segmentation is also performed at this stage to isolate the region of interest (ROI), such as a specific organ, lesion, or tissue compartment. ROI extraction can be done manually, semi-automatically, or by using automatic segmentation algorithms (thresholding, region growing, deformable models, or deep learning–based models such as U-Nets). Accurate preprocessing and segmentation are crucial, as errors introduced here propagate through the subsequent stages and may degrade classification performance.

2.5.3 Data Augmentation Techniques

In order to overcome the underlying problem of data scarcity inherent in medical imaging, data augmentation plays a crucial and extensively used approach for artificially enlarging training datasets via systematic changes and synthetic data synthesis. Data augmentation greatly improves model generalization, reduces overfitting, and strengthens classification robustness by applying various transformations to existing images while maintaining their diagnostic information and clinically-relevant features [62]. This is especially important when the availability of labeled medical imaging data is severely constrained [63].
Several geometric and intensity-based augmentation methods remain widely employed in medical imaging applications:

- **Rotation:** Rotation: Models may identify pathogenic patterns irrespective of lesion orientation by applying angular transformations (usually 10–90 degrees depending on anatomical context). This is crucial for dermatological imaging, because skin lesions display arbitrary orientations. Rotating brain MRI may produce artificial configurations rarely found in clinical settings, however rotation validity is dependent on anatomical context [64].
- **Flipping (Horizontal/Vertical):** Both horizontal and vertical flips efficiently double training samples for symmetric organs while maintaining clinical plausibility by taking use of

anatomical symmetries in certain organ systems (e.g., kidneys, lungs) where bilateral symmetry is predicted [65].

- **Translation and Shifting:** Variations in lesion location within anatomical structures are simulated by spatial shifts by pixels in both horizontal and vertical directions, with translation amounts that vary per organ (often 5–20% of picture dimensions) [66].
- **Scaling and Zoom:** Models can identify pathological patterns at multiple scales by simulating different distances between imaging sensors and anatomical structures through resizing images and zooming into specific regions. This is especially crucial for histopathology, where magnification varies significantly between laboratory microscopes [67].
- **Intensity Variations:** In order to handle the crucial issue of domain shift when deploying models across various hospital imaging equipment, brightness/contrast modifications and histogram equalization mimic scanner calibration discrepancies and variable imaging parameters across institutions [66].

2.5.4 Feature extraction

Once the images are preprocessed, the next step is feature learning, or feature extraction. This phase involves identifying and isolating key attributes present within the medical images that can contribute to accurate classification. Traditional methods historically employed techniques such as scale-invariant feature transform (SIFT) and histogram of oriented gradients (HOG). However, recent advances have seen a significant shift toward deep learning methodologies, specifically CNNs, which autonomously learn hierarchical feature representations from raw data, significantly reducing the need for manual feature engineering [68]. CNNs leverage multiple convolutional layers to progressively abstract features, thereby achieving a more robust representation essential for accurate classification tasks.

this step involves transforming raw pixel data into meaningful, discriminative qualities that facilitate efficient disease diagnosis and diagnostic decision-making. The process of identifying and depicting distinguishing features, such as edges, textures, spatial patterns, anatomical structures, or pathological markers, that separate healthy tissue from diseased areas and discriminate across disease subtypes is known as feature extraction. Model performance is directly influenced by the quality, relevance, and comprehensiveness of extracted features; discriminative feature sets allow high-performance classification even with simpler classifiers, while insufficient or irrelevant features result in poor classification regardless of algorithmic sophistication [69].

2.5.5 Traditional Feature Extraction Approaches

Conventional feature extraction relies on domain expert knowledge to design and calculate features from medical images, employing mathematically-defined descriptors capturing specific image characteristics [70]:

- **Texture Features:** Gray Level Co-Occurrence Matrix (GLCM) captures local texture patterns through analyzing pixel co-occurrence relationships, providing statistical measures including contrast, correlation, energy, and homogeneity that characterize tissue roughness and structure. Gabor filters and Gabor wavelets analyze multi-scale, multi-oriented texture patterns, with research demonstrating that Gabor wavelets represent the most effective and accurate method among texture feature extraction techniques for medical imaging applications [71].
- **Morphological Features:** Shape descriptors (area, perimeter, circularity, eccentricity) and size characteristics characterize anatomical structures and lesion morphology, particularly valuable for detecting abnormal shapes indicative of pathology [70].
- **Gray Level Run Length Matrix (GLRLM):** Quantifies the distribution of consecutive pixels with identical gray values along specific directions, capturing structural patterns and tissue organization patterns [71].
- **Color and Intensity Features:** Histogram-based descriptors characterize pixel intensity distributions, enabling detection of intensity variations associated with pathological changes [70].

2.5.6 Automatic Feature Extraction Through Deep Learning

Automatic feature learning eliminates the need for manual engineering by allowing models to discover optimal data representations directly from the raw input.

- **Automatic feature learning:** Medical image classification was completely transformed by automatic feature learning using deep neural networks, especially convolutional neural networks, which eliminated the need for human feature engineering and allowed for the hierarchical identification of appropriate feature representations. Over the course of several layers, deep convolutional networks gradually learn more abstract features: early layers identify basic visual primitives (edges at different orientations, simple textures); intermediate layers integrate primitive features into higher-level patterns (corners, curves, simple shapes); deep layers acquire intricate, task-specific representations (anatomical structures, disease morphologies) [71], [72].
- Beaglehole et al. (2023) proposed the CNN Ansatz, identifying the fundamental mechanism of deep feature learning in CNNs by establishing that filter covariances in convolutional layers are

proportional to the average gradient outer product (AGOP) of input patches, with extensive empirical evidence demonstrating high correlation between filter covariances and patch-based AGOPs across standard architectures (AlexNet, VGG, ResNets) pre-trained on ImageNet [73]. This theoretical framework illuminates how convolutional networks discover discriminative feature representations, revealing that Deep ConvRFM (Convolutional Random Feature Maps) based on the ansatz recovers similar features to deep convolutional networks including emergence of edge detectors, and overcomes limitations of fixed convolutional kernels by achieving superior performance through adaptation to local image signals.

2.5.7 Feature selection and dimensionality reduction

The feature extraction step often yields high-dimensional descriptors, many of which may be redundant or weakly informative. Feature selection and dimensionality reduction aim to retain the most discriminative information while discarding noise and redundancy. This step helps to mitigate the curse of dimensionality, reduce computational cost, and improve model generalization, especially when the number of available training samples is limited. Techniques such as filter-based selection (e.g. mutual information, Fisher score), wrapper or embedded methods (e.g. recursive feature elimination, LASSO), and projection-based approaches (e.g. principal component analysis, linear discriminant analysis, t-SNE for visualization) are widely used. In deep learning frameworks, dimensionality reduction is often implicitly performed through learned bottleneck layers or global pooling operations.

2.5.8 Classification

In the classification stage, the selected features are mapped to diagnostic labels, such as benign versus malignant, normal versus pathological, or multi-class disease subtypes. Classical machine learning approaches include SVM, k-NN, decision trees, random forests, and ensemble methods, which operate on hand-crafted or radiomic features. In end-to-end deep learning systems, the final layers of the network (fully connected layers with softmax or sigmoid activations) act as the classifier and are optimized jointly with the feature extraction layers. The choice of classifier and loss function (e.g. cross-entropy, focal loss, class-balanced loss) must account for common challenges in medical imaging, such as class imbalance, label noise, and the need for calibrated probabilistic outputs to support clinical decision-making.

2.5.9 Evaluation and validation

The final stage of the pipeline concerns the evaluation and validation of the classification system. Robust assessment requires the use of appropriate experimental protocols (train/validation/test split, k-fold cross-validation, or nested cross-validation) and the strict separation of training and test data to avoid information leakage. Performance is typically quantified using metrics such as accuracy, sensitivity, specificity, precision, F1-score, area under the ROC curve (AUC), and, where relevant, free-response

ROC (FROC) analysis. For clinical translation, external validation on multi-center datasets, robustness analyses with respect to acquisition variability, and statistical significance testing are essential. In addition, explainability tools (saliency maps, class-activation maps, attention mechanisms) are increasingly used to verify that the model bases its decisions on anatomically plausible image regions, thereby fostering trust and acceptance among clinicians.

2.6 Classification models

Classification models form the foundational architecture of machine learning systems, enabling algorithms to learn patterns from labeled training data and make accurate predictions on new, unseen instances. From traditional statistical techniques to complex neural network structures, there is a wide range of categorization systems. Each has unique benefits, computing needs, and applicability for certain data properties and problem areas. Developing successful solutions in medical image analysis, where model selection significantly affects diagnostic accuracy, computational efficiency, and clinical feasibility, requires an understanding of the advantages, disadvantages, and suitable contexts for deployment of these various model types [74].

2.6.1 Machine learning-based models

Traditional machine learning classification models are tried-and-true, mathematically understandable methods that have worked effectively in a variety of applications. They are especially useful in situations when computing resources are few or datasets are tiny. Support vector machines (SVMs) maximize the margin between class boundaries through hyperplane separation and use kernel tricks to handle non-linear relationships, offering solid theoretical underpinnings and outstanding performance on moderately sized datasets. Logistic regression, on the other hand, learns linear decision boundaries by modeling class probabilities through a sigmoid function and is especially effective for linearly separable data with high interpretability and minimal computational overhead. Random forests combine several decision trees to significantly increase robustness, minimize overfitting, and successfully manage intricate non-linear patterns while retaining a moderate level of interpretability. Decision trees recursively divide feature space using hierarchical if-then rules and offer clear, understandable decision paths appropriate for both classification and feature importance analysis. K-nearest neighbors (k-NN) are computationally efficient for inference but difficult for prediction on big datasets since they categorize instances based on the majority vote of adjacent training samples and do not need an explicit training step. Despite its simplifying assumptions, Bayes works very well on text and categorical data when using probabilistic inference based on feature independence assumptions.

*2.6.1.1 Support Vector machines (**SVMs**)*

SVMs were first introduced as large margin classifiers [75] [76]. The margin Δ of a linear classifier is defined as the minimal distance between the points of the two classes, measured perpendicularly to the separating hyperplane, for a linearly separable training set, as shown in Figure 2.1. This hyperplane is taken into consideration in its canonical form for the SVMs, which means that its parameters w and b are normalized such that the training points nearest to the hyperplane fulfill | ⟨w, xi⟩ + b| = 1. The margin in this instance is equal to 2/kwk. A learning algorithm may manage the model's complexity and choose the best separating hyperplane from all the hyperplanes that divide the two classes of the training set by maximizing this margin. Controlling the complexity (or capacity) is a crucial component that enables the model to generalize on unknown data.

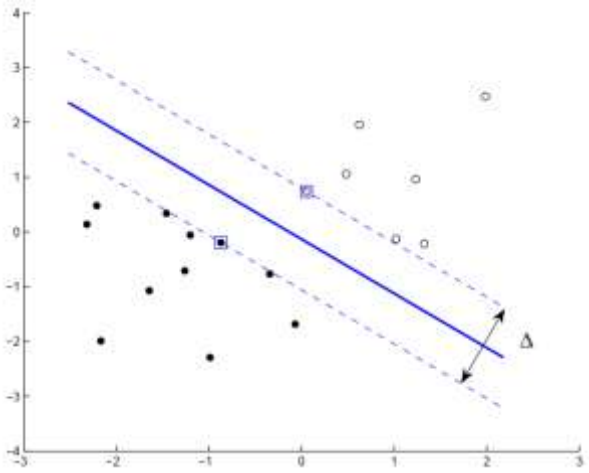


Figure 2-1: SVM Classification with Maximum Margin Hyperplane

A SVM classifier is built from a training set of N labeled samples (xi, yi), where xi ∈ Rp is the input vector corresponding to the ith sample labeled by yi ∈ {-1, +1} depending on its class (only binary problems are considered here). For binary classification, the machine implements the decision function:

$$f(x) = sign(\langle w, x \rangle + b) \qquad 2.1$$

where $w \in \mathbb{R}^p$ and $b \in \mathbb{R}$. This function determines on which side of the separating hyperplane (⟨w, x⟩ + b = 0) the sample x lies.

In other manner:

$$\varphi(x) = \begin{cases} c1 \; if \; b + \sum_{j=1}^{p} x_j \, w_j \\ c2 \qquad otherwise \end{cases} \qquad 2.2$$

Mathematically, SVM are maximum-margin linear models of the form of Equation 2.2 [76] [77]. assuming without loss of generality that Y = {−1, 1} and that b = 0, SVM are trained by resolving the following primal optimization problem:

$$\min_{w,\xi}\left\{\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i\right\} \quad (2.28)$$

subject to

$$y_i(\mathbf{w}\cdot\mathbf{x}_i) \geq 1-\xi_i, \quad \xi_i \geq 0. \quad (2.29)$$

In its dual the form, the optimization problem is

$$\max_{\alpha}\left\{\sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j \mathbf{x}_i\cdot\mathbf{x}_j\right\} \quad (2.30)$$

subject to

$$0 \leq \alpha_i \leq C, \quad (2.31)$$

where C is an hyper-parameter that controls the degree of misclassification of the model, in case classes are not linearly separable. From the solution of dual problem, we have

$$\min_{w,\xi}\left\{\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i\right\} \quad 2.3$$

subject to

$$y_i(\boldsymbol{w}\cdot\boldsymbol{x_i}) \geq 1-\xi_i, \ \xi_i \geq 0. \quad 2.4$$

In its dual the form, the optimization problem is

$$\max_{\alpha}\left\{\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j x_i x_j\right\} \quad 2.5$$

Subject to

$$0 \leq \alpha_i \leq C$$

where C is a hyper-parameter that regulates the model's level of misclassification when classes are not linearly separable. We can ultimately express the final linear model from the dual problem solution.

$$w = \sum_{i=1}^{n}\alpha_i y_i x_i \quad 2.6$$

By projecting the original input space into a high-dimensional space (the so-called kernel technique), SVM may potentially find a separating hyperplane, hence extending to non-linear classification. It's interesting to note that the dual optimization issue is precisely the same, with the exception that a kernel

$K(x_i, x_j)$, which corresponds to the dot product of x. and x. in the new space, replaces the dot product $x_i.x_j$.

2.6.1.2 Decision trees

Among the most popular algorithms for machine learning are decision trees. Decision trees are widely used because they may provide prediction models that are both trustworthy and intelligible.
When compared to other algorithms, decision trees have four main advantages that contribute to their success in the machine learning field. First, if there is enough training data, decision trees can simulate arbitrarily complicated learning problems since they are non-parametric [78]. Secondly, heterogeneous input vectors including numerical, ordinal, and category values are supported. Thirdly, they are resilient to missing, irrelevant, and noisy attribute values. Fourth, several cutting-edge prediction models, such random forest [79], boosting, and shapelet trees in the context of data series categorization, are built on different types of decision trees [80].

A decision tree is a tree-structured model in which the input space is partitioned by each node. The material below focuses on binary decision trees, where each internal node has precisely two offspring, even if decision trees in the general case may be created for various partitions at each node.

Every internal node (t) in a decision tree has a splitting criteria (st) that separates the input space into distinct subspaces. The input instances for which the condition is met make up one of the subspaces for binary decision trees, while the examples for which it is not make up the other subspace. As a result, X is partitioned by the root node t0, and Xt is the partitioning (local training set) at node t for each node. The most likely output Ψ yt 2 Yt of the output value, among the instances in Zt reaching t, is labeled on the terminal nodes of a decision tree. Ψ yt(x) represents the most likely class label. In this manner, the best estimate among the instances that reach the leaf is used to identify the terminal nodes. Formally, the label of the leaf that is reached by navigating the tree in accordance with the splits st is the prediction of an instance x.

2.6.1.3 Random Forest

A random forest is a classifier consisting of a collection of tree-structured classifiers $\{h(\boldsymbol{x}, \Theta k), k = 1, \ldots\}$ where the $\{ \Theta k \}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input **x** [79].

With the training set selected at random from the distribution of the random vector Y, X, and an ensemble of classifiers $h_1(x)$, $h_2(x)$,... $h_K(x)$, define the margin function as

$$mg(X,Y)) = av_k I(h_k(X) = Y) - \max_{j \neq Y} av_k I(h_k(X) = j) \qquad 2.7$$

Where I(.) is the indicator function, The margin quantifies the degree to which the average vote for the correct class at X, Y is higher than the average vote for any other class. The degree of confidence in the categorization increases with the margin. The mistake in generalization is provided by

$$PE* = P_{\boldsymbol{X},Y}(mg(\boldsymbol{X},Y) < 0) \quad 2.8$$

An upper limit on the generalization error for random forests may be obtained in terms of two parameters that represent the accuracy of the individual classifiers and their interdependence. The basis for comprehending how random forests function is provided by the interaction between these two[81].

2.6.1.4 K-Nearest Neighbors (k-NN)

One of the first and most researched algorithms in the history of learning algorithms is the Nearest Neighbors Algorithm. It makes use of an approximation function that utilizes the target values of the training dataset vectors that are closest to the unknown vector in the Hermitian Space to average the target values for the unknown vector. By using the goal values of the K-nearest vectors from the training dataset, the K-Nearest Neighbors method [82] advances the Nearest Neighbors method. Both classification and regression use K-Nearest Neighbors. K-Nearest Neighbors is widely used in medical imaging including brain tumor detection, cardiac disease diagnosis [83], gastric cancer lymph node metastasis detection, dermatological lesion classification, medical image retrieval systems, infectious disease identification, echocardiographic analysis, and multi-organ disease detection [84].

2.6.2 Deep learning-based models

Deep learning (DL) has fundamentally revolutionized medical image analysis through its capability to automatically learn hierarchical feature representations directly from raw data, it has completely transformed medical image analysis by doing away with the requirement for human feature engineering that standard machine learning algorithms need [85]. Deep learning algorithms gradually extract increasingly abstract and clinically relevant features across multiple layers, in contrast to traditional methods that rely on handcrafted features created by domain experts. While deeper layers learn complex anatomical structures and pathological patterns indicative of disease, earlier layers capture basic visual primitives like edges and textures [86]. In medical imaging, where images are naturally high-dimensional, frequently contain noise, show subtle abnormalities that are challenging for humans to detect, and exhibit complex inter and intra-class similarities across various imaging modalities (MRI, CT, X-ray, ultrasound, dermoscopy), this hierarchical and automatic feature learning paradigm has proven especially helpful [87].

Beyond only increasing accuracy, deep learning has revolutionized medical imaging by streamlining clinical procedures, cutting down on radiologist interpretation time, enabling earlier illness identification, facilitating prompt action, and ultimately improving patient outcomes.

Among the numerous deep learning architectures available, three models have emerged as particularly prominent and influential in medical image classification: **VGG16 (Visual Geometry Group)** for their uniform architecture design and strong transfer learning capabilities from ImageNet pre-training [88], **Capsule Networks (CapsNet)** for their ability to preserve spatial hierarchies and demonstrate robustness to transformations through dynamic routing mechanisms [89], and **EfficientNet** for their parameter efficiency and scalability across computational constraints [90]. Each architecture offers distinct advantages and is suited to different clinical applications and institutional contexts.

2.6.2.1 VGG16 (Visual Geometry Group)

VGG16, introduced by Simonyan & Zisserman [91], represents a landmark deep convolutional neural network architecture characterized by its uniform design of 16 weight layers comprising 13 convolutional layers with 3×3 filters followed by ReLU activations, and 3 fully connected (FC) layers with 4096 neurons each. The architecture employs five max-pooling layers (2×2) to reduce spatial dimensions while preserving key features, culminating in a final softmax output layer originally trained on 1,000 ImageNet classes. VGG16 achieved approximately 92.7% accuracy on the ImageNet dataset and remains widely adopted as a backbone for transfer learning applications in medical imaging[92].

Figure 2-2: VGG16 Architecture

2.6.2.2 Capsule Networks (CapsNet)

Capsule Networks (CapsNets), introduced by Hinton et al. [93], represent a paradigm shift in neural network design, replacing traditional pooling layers with dynamic routing algorithms and capsule-based modeling that preserve spatial hierarchies and part-whole relationships between features capabilities that conventional CNNs fundamentally lack [94]. Unlike CNNs that produce scalar outputs through pooling (which discards spatial information), CapsNets output vectors encoding both feature presence and orientation, making them inherently adept at recognizing spatial relationships, pose variations, and complex patterns critical in medical imaging [95].

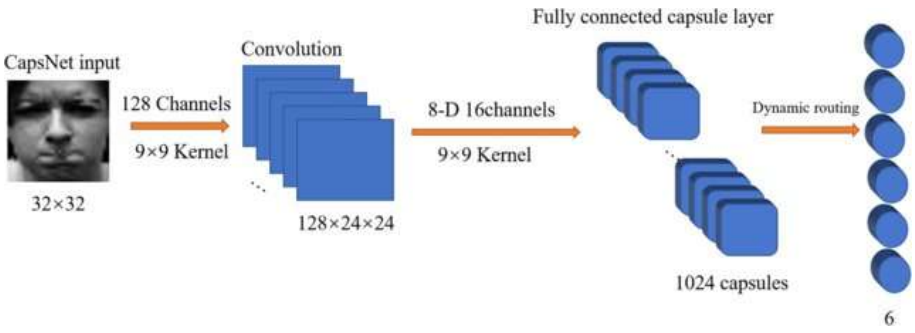


Figure 2-3: Capsule Network structure

2.6.2.3 EfficientNet

EfficientNet, developed by Tan & Le [96], which enables deployment on both high-performance and resource-constrained devices by optimizing accuracy-efficiency tradeoffs by systematic scaling of network depth, breadth, and resolution. By maintaining ideal proportions between depth, breadth, and resolution dimensions, the compound scaling technique produces models that outperform conventional CNNs in terms of accuracy while using fewer parameters. With 8.4x fewer parameters and strong diagnostic performance on par with or better than ResNet, EfficientNet-B0 through B7 variants allow deployment across a variety of clinical settings, from smartphones and edge devices at resource-constrained facilities to high-performance hospital imaging systems [97]. In international healthcare settings, where computing resources differ greatly across institutions, EfficientNet is very helpful [98].

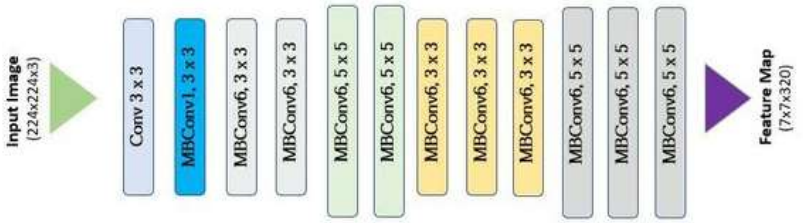


Figure 2-4: EfficientNet architecture

2.7 Evaluation metrics

Evaluation metrics are quantifiable measurements that evaluate the performance of classification models, allowing for the objective comparison of algorithms, the identification of their advantages and disadvantages, and the assessment of clinical readiness for diagnostic deployment. Choosing the right evaluation metrics is crucial in medical imaging applications because different metrics show different aspects of model behavior, and choosing the wrong metrics can mislead researchers about true clinical performance. This is especially problematic when class imbalance, unequal error costs, and diagnostic consequences differ between false positives and false negatives. Medical diagnostics requires a thorough metric assessment that takes into consideration the clinical environment and the repercussions of

diagnostic failures, in contrast to ordinary computer vision applications where accuracy is sufficient [99].

- **Accuracy**, represents the proportion of correct predictions among all predictions. While intuitive and commonly reported, accuracy proves misleading in medical imaging with class imbalance, a model predicting "normal" for all cases achieves high accuracy in datasets where abnormalities are rare, while failing to detect pathology. Consequently, accuracy should never be the sole evaluation metric in medical imaging. calculated as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad 2.9$$

- **Precision**, measures the proportion of positive predictions that are correct. High precision indicates few false alarms, particularly important in screening scenarios where unnecessary further investigation of false positives creates patient anxiety and healthcare costs. calculated as

$$Precision = \frac{TP}{TP + FP} \quad 2.10$$

- **Recall** (sensitivity), measures the proportion of actual positive cases detected by the model. High recall proves critical in disease detection where missing a true positive (false negative) has serious clinical consequences, a cancer screening system with high recall minimizes missed diagnoses. calculated as

$$Recall = \frac{TP}{TP + FN} \quad 2.11$$

Precision and recall often demonstrate inverse relationships: maximizing precision (requiring high confidence predictions) reduces recall (missing some true cases), while maximizing recall (accepting lower confidence predictions) reduces precision (increasing false alarms).

2.8 Conclusion

Medical image analysis and clinical diagnostics have undergone a fundamental transformation thanks to the integration of advanced classification methodologies, which include cutting-edge deep learning architectures (CNNs, Vision Transformers, CapsNets, EfficientNet, VGG16), conventional machine learning algorithms (SVMs, k-NN), and complementary techniques like data augmentation and thorough feature extraction. Deep learning techniques have proven to perform exceptionally well in a variety of imaging modalities, such as MRI, CT scans, X-rays, ultrasound, and digital pathology. They have achieved diagnostic accuracy on specific, well-defined classification tasks that is on par with or better than that of expert radiologists, while also automating laborious feature extraction procedures and enabling large-scale image analysis that is not possible through manual review. With applications

ranging from cancer detection in mammograms and CT scans to cardiovascular disease classification in echocardiograms, neurological disorder detection in brain MRIs, and infectious disease identification across multiple modalities, Shobayo et al. showed how deep learning has transformed medical image analysis through automated, efficient, and highly accurate diagnostic solutions[100].

However, a number of significant obstacles still prevent medical image classification algorithms from being widely used and successfully translated into clinical settings. Despite decades of progress in medical imaging, data availability and annotation scarcity still exist; labeled datasets are still much smaller than natural image datasets, which restricts the training and generalization of deep learning models. Physicians need clear diagnostic reasoning to incorporate AI systems into clinical workflows, yet many deep learning models operate as "black boxes" that make predictions without clinical support. Interpretability and explainability are persistent issues. Models trained on data from individual institutions or particular imaging equipment often fail when deployed on other scanners, imaging procedures, or varied patient groups, making generalization across institutions and populations challenging and requiring ongoing validation and modification. Deployment in resource-constrained healthcare settings is limited by computational requirements and real-time performance, especially in developing nations with limited computational infrastructure. Ethical and regulatory frameworks are still developing; issues including data protection, fair access to AI-powered diagnostic tools, and culpability attribution in the event of AI diagnostic mistakes are yet not sufficiently addressed [101].

3 Bio-inspired Metaheuristics for Medical Image Classification

3.1 Introduction

Medical imaging generates vast amounts of data daily, and the computational challenges in extracting diagnostic information have become increasingly acute. Feature selection, hyperparameter tuning, and architecture design in machine learning pipelines require navigating high-dimensional, non-convex solution spaces where gradient-based methods often fail to find optimal or near-optimal solutions. Metaheuristic algorithms, population-based stochastic search methods inspired by natural phenomena, offer a principled yet flexible approach to these complex optimization problems, balancing computational feasibility with solution quality.

In the context of Computer-Aided Diagnosis (CAD), the integration of metaheuristics has emerged as a promising strategy to enhance the robustness, generalizability, and efficiency of diagnostic systems. Traditional manual tuning of machine learning models is labor-intensive, often yields suboptimal results, and is sensitive to dataset characteristics and imaging protocols, limiting reproducibility across institutions. Metaheuristic algorithms, such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Grey Wolf Optimizer (GWO), provide automated, systematic methods for optimizing radiomic feature sets, selecting discriminative features, and tuning hyperparameters of classifiers ranging from classical machine learning (SVM, Random Forest) to deep neural networks.

The scope of metaheuristic applications in medical imaging spans multiple levels of the diagnostic pipeline: feature extraction and dimensionality reduction, classification algorithm parameter tuning, ensemble weight optimization, and increasingly, architecture design for deep learning models. Recent advances in hybrid approaches, such as PSO-tuned SVMs, GA-optimized CNNs, and GWO-enhanced ensemble classifiers, have demonstrated marked improvements in diagnostic accuracy, robustness to domain shift, and computational efficiency compared to non-optimized baselines. These successes underscore the critical importance of integrating intelligent optimization strategies into the design of next-generation CAD systems that must operate reliably across diverse patient populations, imaging devices, and clinical settings.

The present chapter reviews the conceptual foundations of metaheuristic optimization, surveys the principal algorithms applicable to medical imaging, and discusses their roles in improving feature selection, hyperparameter tuning, and model fusion. By establishing this theoretical and practical grounding, the chapter prepares the foundation for subsequent sections that will detail the application of metaheuristics, specifically Genetic Algorithms, to optimize hybrid deep learning architectures for histopathological image classification in breast cancer diagnosis.

3.2 Metaheuristics: Concepts and Foundations

3.2.1 Definition of Metaheuristics

High-level optimization techniques such as metaheuristics are created to effectively resolve challenging, nonlinear, or nonconvex optimization problems in which traditional deterministic algorithms either fail or become computationally expensive. The term "metaheuristic" was initially formalized by Glover [102], who defined it as a framework that directs underlying heuristics to search beyond local optimality and more efficiently explore the solution space. These methods were referred to as "modern heuristics" prior to the term's widespread use [103]. The stochastic aspect of metaheuristics, which is purposefully introduced using randomization techniques like probabilistic selection, random perturbation, or randomized initialization, is one of its distinguishing features. For high-dimensional or multimodal search spaces in particular, this stochasticity improves resilience and prevents premature convergence to poor solutions [104].

Furthermore, a lot of metaheuristics use memory structures that affect search behavior over time. For instance, evolutionary algorithms preserve and spread knowledge about the best people over generations [105], while Tabu Search employs a tabu list to avoid cycling back to previously visited solutions [102]. Both short-term guiding and long-term strategic adaptability are made possible by these memory processes.

The need of striking a balance between exploration and exploitation is a basic tenet of metaheuristic design [106]:

The algorithm is encouraged by exploration to look into other or uncharted areas of the search field. To improve the quality of the solution, exploitation concentrates the search on regions that show promise. It's important to strike the correct balance since too much exploitation might imprison the algorithm in local minima, while too much exploration could stall convergence. This trade-off is often emphasized in the literature as being crucial to developing a successful search strategy [106] [104].

The trade-off between convergence speed and solution quality is another crucial factor to take into account. While more exploratory approaches often provide better results at the expense of more computing time, algorithms built for quick convergence may compromise accuracy or global search capabilities. Effective metaheuristics use learning-based processes, hybridization, or parameter control to adaptively manage these conflicting goals [107].

3.2.2 Categories of Bio-inspired Metaheuristics

Bio-inspired metaheuristics are typically grouped into several major categories; each associated with a distinct natural phenomenon. Figure 3.1 summarizes these categories, highlighting evolutionary algorithms, swarm intelligence models, collective-behavior–based systems, and physics-inspired algorithms. In the following sections, we describe some of these representative algorithms, outlining their core principles and highlighting their relevance in medical imaging and CAD applications.

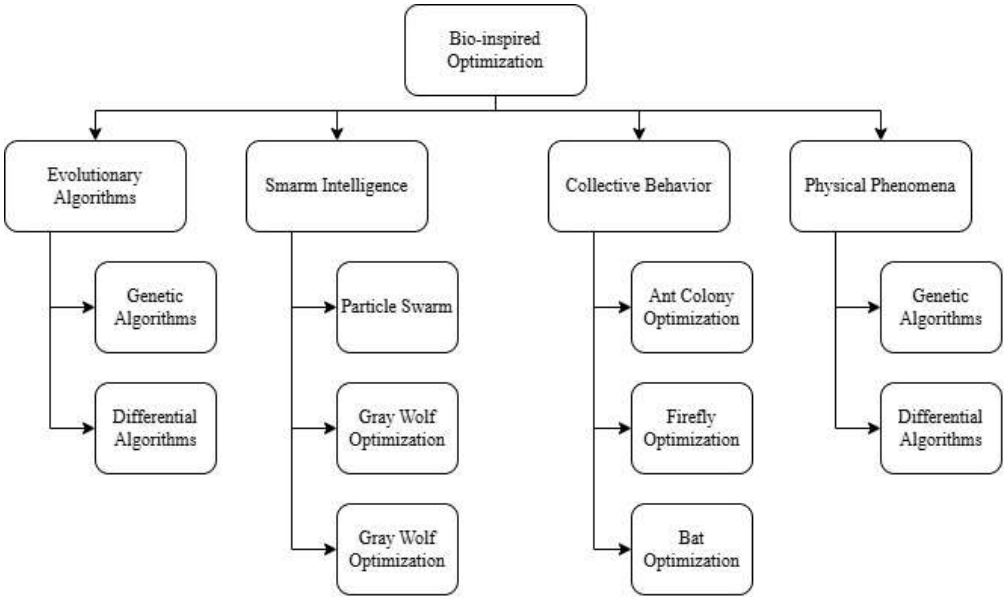


Figure 3.1: Overview of Bio-Inspired Optimization Algorithms [108]

3.2.2.1 Evolutionary Algorithms: GA

In order to explain the evolution of living things, C. Darwin proposed a hypothesis in 1859 that was founded on the idea of natural selection. The author defines natural selection as follows: "I have called this principle, by which any variation, however insignificant it may be, is preserved and perpetuated if it is useful, natural selection." Three concepts form the foundation of this theory. The first is the principle of variation, which is defined as the difference between individuals within a population; natural selection requires this difference, no matter how minor. The second principle is adaptation, which leads to differences in people's characteristics. The people that are most adaptable to their surroundings benefit from these changes, or mutations.

The last principle is heredity, which stipulates that characteristics must be passed on to progeny. Although evolutionary algorithms were first used to solve engineering problems in the 1950s, it wasn't until four separate approaches, genetic algorithms [109], genetic programming [110], evolutionary programming [111], and evolution strategy [112], were published that they became widely accepted. The same procedural ideas underpin all of these methods. The general process of an evolutionary algorithm is shown in Figure 2.1. Since genetic algorithms are unquestionably the most popular

approach among evolutionary algorithms and metaheuristics in general, We will briefly discuss these methods in the following sections, including both Genetic Algorithms (GA) and Differential Evolution (DE).

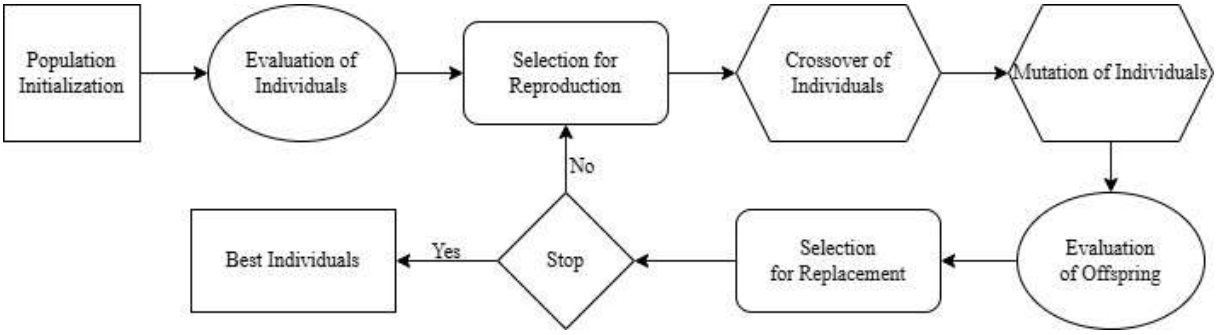


Figure 3.2 Evolutionary algorithm process [113]

Although J. Holland first suggested genetic algorithms [109], they didn't become well-known until 14 years later, when D.E. Goldberg's groundbreaking work [114] was published. The people (population) that go through evolution in genetic algorithms are a collection of answers to the issue that has to be addressed. Using selection and variation operators, the present population (parents) creates a new generation (offspring) at each iteration. Selection operations enable the selection of a person for replacement or reproduction (crossover). In general, an individual's chances of getting chosen increase with their level of efficiency. Crossover operators and mutation operators are the two types of variation operators. Crossover operators produce one or more offspring by merging several parents, often two. By making changes to one person, mutation operators create a new individual.

3.2.2.2 Swarm Intelligence: PSO, BCO, GWO.

Swarm Intelligence (SI) represents a major subcategory of bio-inspired metaheuristics, grounded in the collective behavior of decentralized, self-organizing biological systems such as bird flocks, fish schools, insect colonies, and animal social hierarchies. Unlike evolutionary algorithms, which rely on reproduction and selection, SI methods operate through the coordinated interaction of multiple agents that adapt their movement based on simple behavioral rules and shared information. This emergent cooperation enables the swarm to efficiently explore high-dimensional search spaces, avoid premature convergence, and adapt dynamically to complex optimization landscapes, capabilities that are particularly valuable in medical imaging applications where global optima are often hidden within irregular, multimodal spaces [115].

Among the most widely used SI algorithms are Particle Swarm Optimization (PSO), inspired by flocking dynamics; Bee Colony Optimization (BCO), modeled after foraging behavior in honeybee colonies; Grey Wolf Optimizer (GWO), based on the leadership hierarchy and hunting strategies of grey wolves;

and Whale Optimization Algorithm (WOA), which simulates the bubble-net hunting mechanism of humpback whales. In the following sections, representative algorithms from each SI family are presented to illustrate their core principles, search dynamics, and practical relevance to CAD optimization tasks in medical imaging [116].

3.2.2.2.1 Bee Colony Optimization

Bees in the wild display very intricate mating, reproductive, and food-foraging habits. These natural activities that take place inside or originate from bee colonies serve as an inspiration for a number of optimization techniques. The MBO (Marriage in Honey Bees Optimization) method is one of the algorithms that draws inspiration from bee mating and reproductive activity [117].

The MBO algorithm starts with a single queen without a family and works its way up to a colony with one or more queen families. The HBMO (Honey-Bees Mating Optimization) method [118] , the FMHBO (Fast Marriage in Honey Bees Optimization) algorithm [119], and the HBO (Honey-Bees Optimization) algorithm [120] are some of the variations of the MBO technique that have been suggested in the literature.

The way bees naturally forage is the basis for another class of bee swarm algorithms. To find interesting locations, these algorithms use either a random exploratory search or a regular evolutionary search. In order to locate the global optimum, they also use exploitative search around the most promising locations. The way bees forage for food served as the model for the following algorithms. ABC (Artificial Bee Colony) [121], BS (Bee System) [122], BCO (Bee Colony Optimization) [123], and BA (Bees Algorithm) [124].

The Genetic Algorithm (GA) has been refined into the BS algorithm. In order to improve local search while maintaining GA's global search capabilities, it takes inspiration from GA.
To address combinatorial optimization issues, the BCO method was put out. It is based on the forward pass and backward pass phases. Through individual investigation and group experience, a partial solution is created during the forward pass, which is then used in the backward pass. In the backward pass, probabilistic information is utilized to choose whether to start investigating the neighborhood of newly chosen solutions or to continue investigating the present solution in the subsequent forward pass. Roulette Wheel Selection and other probabilistic methods are used to find the new answer.

Karaboga introduced the Artificial Bee Colony (ABC) optimization technique in 2005 [125]. Employed bees, onlooker bees, and scout bees are the groups of bees that form the basis of the ABC algorithm.

After using resources, employed bees provide knowledge about the environment back to the hive. Onlooker bees are then given access to this search data, and they use a probabilistic method (such as Roulette Wheel Selection) to assess the data and start a neighborhood search.

To guarantee exploration, scout bees conduct haphazard searches in the meanwhile. Additionally, both local and global search procedures are used by the Bees Algorithm (BA). In contrast to the ABC algorithm, which employs a probabilistic method for neighborhood search, the BA algorithm bases its search strategy on the development of fitness [126].

3.2.2.2.2 Particle Swarm Optimization

Inspired by the statistical models created by Reynolds et al. 1987 [127] and Heppner et al. [128], which enable the simulation of the movement of flocks of birds and schools of fish, J. Kennedy and R. Eberhart first proposed Particle Swarm Optimization (PSO) as an optimization metaheuristic in 1995 [129].

A population of people known as particles conducts the search in the particle swarm method. Every particle is seen as a possible solution to the issue as it flies throughout the search space in quest of the global optimum. A particle uses two sorts of information to decide its flight direction: information from its own experience and information from the swarm's experience.

The following equations (Eqs. I.1 and I.2) control how the particles travel.

$$V^{(k+1)} = \omega . V^{(k+1)} + c_1 . rand_1 . \left(Pbest^{(k)} - X^{(k)}\right) + c_2 . rand_2 . \left(Gbest^{(k)} - X^{(k)}\right)$$

$$X^{(k+1)} = X^{(k)} + V^{(k+1)}$$

Where :

X*: the position of the particles*
V*: the velocity of the particles*
w*: the inertia parameter*
Pbest*: the personal best position*
Gbest*: the swarm's best solution*
rand1, rand2*: random variables between 0 and 1*
c1, c2*: positive constants*
k*: the iteration index*
The three terms of the velocity equation can be interpreted as follows:

$\omega . V^{(k)}$*: represents a physical inertia component, which encourages each particle to follow its current direction of movement.*

$c_1.rand_1.\left(Pbest^{(k)} - X^{(k)}\right)$: *represents a cognitive component, which encourages the particle to return to the best position it has previously visited.*

$c_2.rand_2.\left(Gbest^{(k)} - X^{(k)}\right)$: *represents a social component, which encourages the particle to move toward the best position found by its neighbors.*

Figure I.2 illustrates an example of a particle's flight direction in a search space.
The general procedure of particle swarm optimization algorithms is presented in PSO Algorithm.

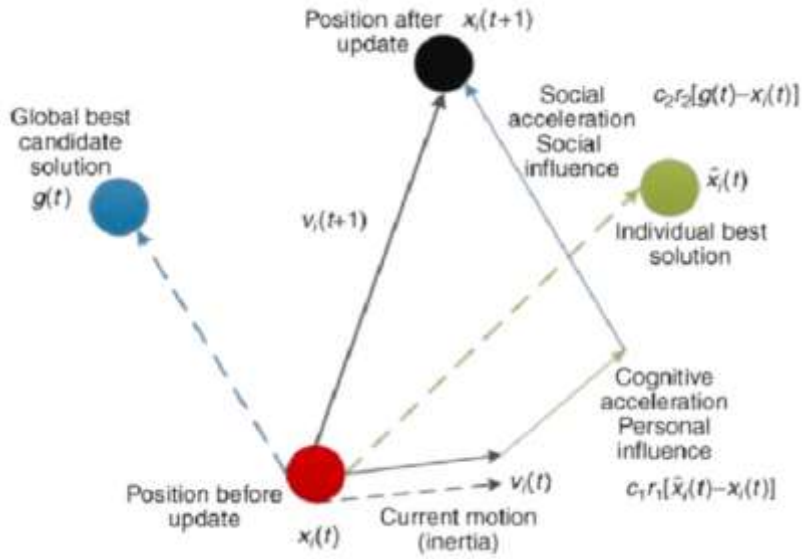


Figure 3.3:Illustration of the PSO Velocity and Position Update Process

PSO Algorithm
Requires: *The objective function f, the population size, and the parameters of equation (I.1).*
Generate for each particle an initial position and velocity.
Evaluate for each particle the ***fitness*** *of the objective function.*
Assign to each particle's personal best solution its initial position.
Determine the global best solution.
While *the stopping condition is not satisfied* ***Do***
Move the particles according to equations (I.1) and (I.2).
Evaluate the objective function for the new particle positions.
Update the personal best solutions of the particles and the global best solution.
End while
Return: *The global best solution.*

3.2.2.2.3 Grey Wolf Optimization

The grey wolf optimization method, created in 2014 by MIRJALILI et al. [130] , mimics the group hunting hierarchy of wolves. The CANIDAE family includes the gray wolf. This family is distinguished

by a rigid social order and often likes to live in a group. In order to replicate the grey wolf leadership structure, Figure 3 defines four groups: α, β, δ, and ω.

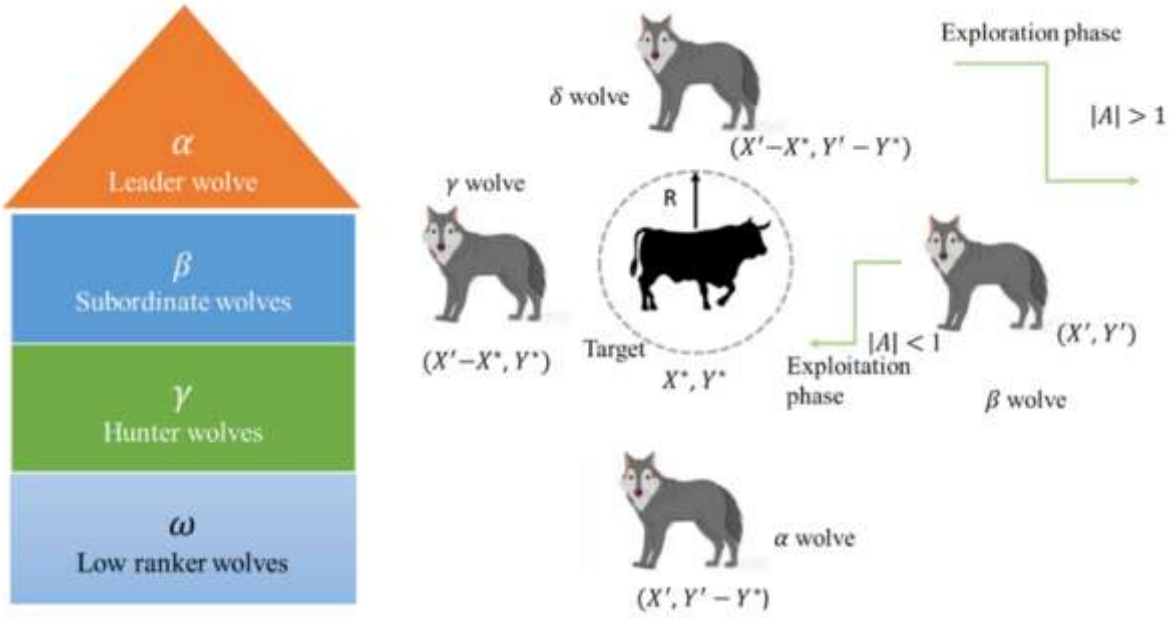


Figure 3.4: General representation of Grey Wolf Optimization [131], [132]

The leader, known as Alpha, may be either male or female and is often in charge of making choices. The group of wolves Beta, which are subordinate wolves that aid in decision-making, must obey the instructions of the dominant wolves. Furthermore, the gray wolf ω has the lowest rank in the hierarchy of leadership, whereas wolf β is one of the group's α advisors. The wolf that controls ω and reports to α and β is referred to as δ if it is not either of α, β, or ω. To create and execute the GWO algorithm, hunting strategies and the wolf social hierarchy were mathematically modeled. The algorithm outperformed artificial intelligence methods, according on the results of testing it using established test functions. Additionally, the approach was effectively used to resolve a number of engineering optimization issues. The majority of swarm strategies utilized to address optimization issues are unable to maintain a leader throughout the processing time.

In GWO, where grey wolves have a certain hierarchical social sequence, this issue is often resolved. Furthermore, the GWO method is simpler to implement than its predecessors since it only requires a few parameters [133], [134].

The GWO algorithm steps can be summarized as follows:

- **Initialization**
  - Define the population of grey wolves (candidate solutions).

- Initialize their positions randomly within the search space.
- Set algorithm parameters, including maximum iterations.

➢ **Evaluate Fitness and Identify the Leading Wolves**

- Compute the fitness of each wolf.
- Rank the population and identify the three best solutions, referred to as:
  - **α (alpha):** the best solution
  - **β (beta):** the second-best solution
  - **δ (delta):** the third-best solution
- These wolves guide the rest of the pack (ω wolves).

➢ **Encircling the Prey**

- Grey wolves update their positions based on the positions of α, β, and δ.
- The encircling behavior is mathematically modeled using adaptive coefficient vectors **A** and **C**, which gradually reduce exploration and increase exploitation as iterations progress.

➢ **Hunting (Search Guidance)**

- Each wolf updates its position by estimating the prey's possible location as a weighted combination of α, β, and δ positions.
- This collaborative mechanism ensures that the pack converges toward promising regions in the search space.

➢ **Attacking the Prey (Exploitation)**

- As iterations advance, the vector **A** decreases linearly, encouraging wolves to transition from large, exploratory movements to more refined, exploitation-focused updates.
- The search becomes increasingly focused near α, eventually leading to convergence.

➢ **Termination**

- Repeat the evaluation and update steps until the maximum number of iterations is reached or a stopping criterion is satisfied.
- Return the final α wolf as the best-found solution

3.2.2.3 Collective Behavior: ACO, Firefly

3.2.2.3.1 Ant Colony Optimization

A class of optimization metaheuristics known as "ant colony algorithms" draws inspiration from the way actual ants follow and deposit pheromone trails. Inspired by the work of Deneubourg et al. [135], who approximated the random behavior of ants, Dorigo et al. [136] [137] proposed these algorithms.

The first optimization algorithm based on ant colonies, called the **Ant System (AS)**, was proposed to solve the Traveling Salesman Problem (TSP) [137]. Since then, numerous improvements and variants have emerged and have been applied in several domains with varying degrees of success.

It was observed in [138], [139] that real ants are capable of selecting the shortest path between their nest and a food source through collaborative and collective behavior, even though no single individual has a global view of the environment. Ant behavior is based on the following principles:

- **Self-organization:** ants are capable of solving complex problems by individually performing simple tasks. The simple and local behavior of each individual gives rise to a global-level pattern through interactions.
- **Stigmergy:** this is the communication mechanism between individuals (ants), achieved by dynamically modifying the environment in which they operate. As ants move, they deposit pheromones[1] to mark the path they have traveled. In the absence of this substance, ants move randomly through the environment. In contrast, when pheromone is present, an ant can detect it and follow its trail with a probability proportional to its intensity. Thus, the more a trail is used, the more attractive it becomes.
- **Decentralized control:** this means that no decision is made at a central level or by a single individual. Each ant performs relatively simple actions based solely on local environmental information, without any global view of the problem.
- **Dense heterarchy:** in contrast to a hierarchical structure where the population is directed by a leader, a dense heterarchy is a horizontal structure in which individuals are strongly interconnected, influencing the global properties of the system.

Figure IV.1 illustrates an example of the path-optimization process between an ant nest and a food source. At the beginning of the experiment (Figure IV.1(a)), ants arrive at point (A); since there are no pheromone trails yet, they move randomly, choosing between the two possible paths with equal probability. Ants that take the path A-C-D logically reach the food source more quickly than the others and return sooner. As a result, the amount of pheromone deposited along path A-C-D will be greater than that deposited along path A-B-D. Since a trail with more pheromone is more likely to be followed, more ants will choose the path A–C–D (Figure IV.1(b)). Consequently, the shortest path becomes increasingly reinforced and will ultimately be used by the majority of individuals. Moreover, considering pheromone evaporation, after some time all individuals will end up taking the shortest path.

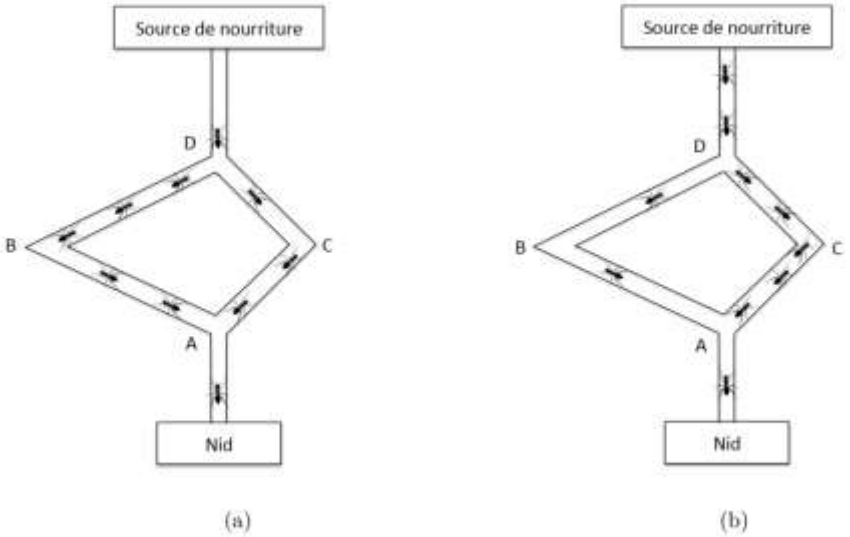


Figure 3.5:Path optimization by an ant colony. (a) At the beginning of the search. (b) At the end of the search.

An analogy is formed between the food supply and the objective function, between pheromone trails and an artificial kind of memory, and between the environment in which ants travel and the search area of the optimization problem in order to take use of this behavior in an optimization method [140]. Ants in artificial systems need to have features not seen in natural insects, such as explicit memory, partial environmental visibility, and discrete time steps [141]. In order to convert natural stigmergic behavior into an effective computing process, these changes are necessary [115].

3.2.2.3.2 Firefly Algorithm (FA)

FA is a kind of stochastic, nature-inspired, meta-heuristic algorithm that may be used to solve the most challenging optimization problems (also known as NP-hard problems). It is one of the more recent swarm intelligence techniques created by Yang [142] in 2008. This method falls under the category of stochastic algorithms. This implies that by looking for a collection of answers, it employs a kind of randomness. The flashing lights of fireflies in the natural world served as inspiration. Heuristic is defined as "to find" or "to find solutions by trial and error" [142]. In actuality, there is no assurance that the best answer will be discovered in a timely manner. Lastly, meta-heuristic refers to "higher level," where a trade-off between randomization and local search influences the search method used in algorithms [142]. The "lower level" (heuristic) of the firefly algorithm chooses the optimal solution for survival by focusing on the creation of new solutions inside a search space. Randomization, on the other hand, makes it possible for the search process to prevent the solution from being stuck in local optima. A candidate solution is improved by the local search until improvements are found, which puts the solution in the local optimum.

Keep in mind that FA is population-based. When compared to single-point search algorithms, population-based algorithms provide the following benefits [143]:

- Crossover is the process of assembling building blocks from several solutions.
- Concentrating a search once again depends on the crossover, which implies that if a variable has the same value in both parents, the offspring will also have the same value.
- Distractions in the scenery are ignored using low-pass filtering.
- The algorithm's chance to learn optimal parameter values in order to strike a balance between exploration and exploitation;
- Hedging against poor luck in its beginning positions or judgments.

Mathematically, the attractiveness β of a firefly is defined as:

$$\beta(r) = \beta_0 e^{-\gamma r^2}$$

where β represents the maximum attractiveness at zero distance, γ is the light absorption coefficient, and r denotes the Euclidean distance between two fireflies. The movement of a firefly *i* toward a more attractive firefly *j* is expressed as:

$$x_i^{t+1} = x_i^t + \beta(r_{ij})(x_j^t - x_i^t) + \alpha\epsilon$$

where α controls the randomization strength and ϵ is a random vector drawn from a uniform or Gaussian distribution

The features of fireflies that inspired the development of the firefly algorithm will be briefly covered in the remainder of this section.

Fireflies are mostly distinguished by their flashing light. Attracting mating mates and alerting possible predators are the two main purposes of these lights. But there are other physical laws that the flashing lights follow. On the one hand, the expression I / 1=r2 indicates that the light intensity I drops as the distance r rises. Yang [142] was motivated to create the firefly algorithm by this phenomena. Conversely, the firefly functions as an oscillator, charging and discharging (firing) light at regular intervals, that is, at θ = 2π. A mutual coupling takes happen when two fireflies are put close to each other. The answer to graph coloring difficulties was particularly influenced by this behavior of fireflies.

The Firefly Algorithm's natural equilibrium between exploration and exploitation is one of its main benefits. While the randomization term facilitates exploration and helps prevent premature convergence to local optima, the attractiveness-based movement encourages exploitation of interesting locations in the search space. FA may successfully solve non-convex, multimodal, and high-dimensional optimization problems without the need for derivative knowledge, in contrast to gradient-based optimization techniques.

3.2.2.4 ***Physical Phenomena****: Simulated Annealing, Harmony Search .*

3.2.2.4.1 Harmony Search

A search strategy called Harmony Search (HS) is based on how jazz musicians improvise [144]. Jazz performers attempt to maximize overall harmonies by varying their pitches in order to achieve aesthetic goals. They start with a few harmonies and use improvisation to try to get better harmonies. Instead of using harmonics to maximize a particular objective function, search heuristics may be derived using this analogy. In this case, the decision variables are associated with the musicians, and the solutions are represented by the harmonies. The HS algorithm repeatedly generates new answers based on previous solutions and random alterations, much how jazz players develop new harmonies via improvisation. Although there is a lot of room for interpretation within this framework, the fundamental HS method is consistently stated in the literature as follows.

The Harmony Memory (HM) is initialized by the HS algorithm using solutions that are created at random. The Harmony Memory Size (HMS) determines how many solutions are kept in the HM. Next, a new solution is developed iteratively as follows. Every choice variable is created using either random selection or memory consideration and a potential extra modification. Harmony Memory Considering Rate (HMCR) and Pitch Adjusting Rate (PAR) are the parameters employed in the process of creating a new solution. With a probability of HMCR, each decision variable is set to the value of the relevant variable of one of the solutions in the HM. A further modification of this value is carried out with a probability of PAR. If not, the decision variable is assigned to a random value with a probability of 1-HMCR. Once a new solution has been developed, it is assessed and contrasted with the HM's worst option. It substitutes the worst answer in the HM if its objective value is higher than that of the worst option. Until a termination requirement is met, this procedure is repeated. The following Algorithm uses pseudo code to provide an overview of the HS algorithm.

Harmony Search Algorithm

1. Initialize the HM with HMS randomly generated solutions
2. **repeat**
3. Create a new solution in the following way
4. **for all** decision variables **do**
5. With probability HMCR use a value of one of the solutions in the harmony memory and additionally change this value slightly with probability PAR
6. Otherwise (with probability 1-HMCR) use a random value for this decision variable
7. **end for**
8. **if** the new solution is better than the worst solution in the harmony memory **then**
9. Replace the worst solution by the new one
10. **end if**
11. **until** Termination criterion is fulfilled

return The best solution in the harmony memory

3.2.3 Evaluation Criteria

Several important assessment criteria are taken into consideration in order to compare and objectively evaluate the efficacy of metaheuristic optimization algorithms in computer-aided diagnostic and medical imaging applications. Both optimized performance and practical viability in actual clinical situations are captured by these criteria.

Exploration–Exploitation Balance: An efficient metaheuristic has to keep exploration and exploitation in a proper balance. In order to lower the chance of premature convergence to local optima, the algorithm must be able to explore a variety of uncharted territory in the search space. In contrast, exploitation focuses on thoroughly investigating potential areas in order to develop superior solutions. In high-dimensional and multimodal optimization issues that are often encountered in medical imaging, a well-balanced trade-off is crucial.

Convergence Rate: Over the course of many iterations, an algorithm's convergence rate indicates how rapidly it gets closer to an ideal or nearly optimal solution. In medical applications, algorithms with quicker convergence are often favored since they save computation and training time. Convergence behavior is a crucial comparison parameter since quick convergence shouldn't compromise the quality of the solution.

Stability and Robustness to Noise: Variability in acquisition techniques, inter-patient variations, and noise may all have an impact on medical imaging data. Consequently, when the objective function is noisy or partly unclear, a trustworthy metaheuristic algorithm should show consistent performance across many runs and retain resilience. Reproducible and reliable optimization requires consistent convergence behavior and low susceptibility to random initialization.

Computational Complexity: The algorithm's time and memory needs are reflected in computational complexity. Large datasets and deep learning models are used in many medical imaging applications, therefore optimization strategies must continue to be computationally viable. Efficient metaheuristics are appropriate for both high-performance computer settings and healthcare systems with limited resources because they strike a good compromise between optimization quality and computational overhead.

Algorithm	***Exploration–Exploitation Balance***	***Convergence Rate***	***Stability & Robustness to Noise***	***Computational Complexity***
Genetic Algorithm (GA)	*Strong exploration due to mutation and population diversity; exploitation via selection and crossover*	*Moderate (depends on population size and operators)*	*High robustness; stable across runs but sensitive to parameter settings*	*High (population-based with crossover and mutation overhead)*
Particle Swarm Optimization (PSO)	*Faster exploitation; exploration controlled by inertia and stochastic terms*	*Fast convergence in early iterations*	*Moderate robustness; may suffer from premature convergence in noisy landscapes*	*Low to moderate (simple velocity–position updates)*
Firefly Algorithm (FA)	*Adaptive balance; attraction favors exploitation, randomization ensures exploration*	*Moderate to fast; improves with proper parameter tuning*	*High robustness to multimodal and noisy objective functions*	*Moderate (pairwise attraction computations increase cost)*
Grey Wolf Optimizer (GWO)	*Well-balanced via adaptive control of*	*Fast and stable convergence*	*High stability and low sensitivity to*	*Low (few parameters and*

	exploration and exploitation		*noise and initialization*	*simple update equations)*

3.3 Popular Metaheuristics in Medical Imaging

Complex, nonlinear, and computationally expensive optimization issues are often encountered in medical imaging activities such image segmentation, registration, reconstruction, feature selection, and parameter optimization. High-dimensional search spaces, multimodal goal functions, and sensitivity to noise and acquisition variability are characteristics of these tasks. Under these circumstances, traditional deterministic or gradient-based optimization techniques often fail, either converging to local optima or becoming computationally impractical. Inspired by biological, physical, or natural phenomena, metaheuristic algorithms provide adaptable and reliable alternatives that may successfully explore such difficult solution spaces.

This section examines a number of well-known metaheuristic algorithms that have been extensively used in computer-aided diagnostic (CAD) and medical imaging systems, emphasizing their fundamental workings and typical uses.

Complex, nonlinear, and computationally demanding issues are often involved in medical imaging operations, including image segmentation, registration, reconstruction, feature selection, and parameter optimization. High-dimensional search spaces, noise, and multimodality may be challenges for conventional optimization methods. Inspired by physical or natural processes, metaheuristic algorithms provide reliable and adaptable substitutes. Several popular metaheuristics in medical imaging are presented in this section.

3.3.1 Evolutionary Algorithms

Evolutionary algorithms are inspired by the principles of natural evolution, including selection, reproduction, and survival of the fittest. These algorithms maintain a population of candidate solutions that evolve over successive generations, making them well suited for global optimization problems.

- **Genetic Algorithm (GA)**:

One of the most popular metaheuristics in medical imaging applications is GA. Each potential solution in GA is represented by a chromosome, and a population of chromosomes changes as a result of genetic operators like selection, crossover, and mutation. In order to preserve variety and prevent premature convergence, mutation adds random variants, crossover mixes genetic material from parent solutions to create offspring, and selection favors fitter individuals [145].

Sharma and Kumar [146] claim that the use of GAs in medical imaging has revolutionized the field by enabling better quality, accuracy, and customization in a variety of imaging modalities, such as computed tomography (CT) and magnetic resonance imaging (MRI).

The effectiveness of genetic algorithms (GAs) in medical imaging has been shown by several applications supported by empirical research. One important area where GAs have a big impact is image segmentation, a critical process for accurate diagnosis and analysis. The use of GAs to enhance image segmentation techniques, namely in MRI and CT scans, was examined by Torse et al. [147]. When compared to traditional methods, they discovered that using GAs significantly improved segmentation accuracy and decreased segmentation mistakes by 15%.

A key use of GAs in medical imaging is feature selection, which lowers the dimensionality of data while preserving diagnostic information. Feature selection for GA-based breast cancer diagnosis was examined by Al-Najdawi et al. [148]. Their findings showed that using GAs not only identified the most relevant traits from a pool of potential predictors but also enhanced the classification performance of several machine learning models, achieving classification accuracies above 90%.

- **Differential Evolution (DE)**:

Differential evolution (DE) has emerged as a potent optimization technique in the field of medical imaging. Research has shown that DE is excellent at preserving convergence speed while reducing the possibility of stalling at local optima, which are frequent problems in traditional optimization techniques [149]. This quality is especially important in medical imaging, since fast and precise findings may have a big impact on patient outcomes.

The growing convergence of machine learning and enhanced illness detection methods further increases the significance of DE in this field. Optimized parameters are becoming more important as machine learning methods, including convolutional neural networks (CNNs), gain popularity in the analysis of medical pictures. DE has been effectively used to tune these models' hyperparameters, improving the detection rates for a number of illnesses, including as cancer and heart disease [150]. DE is a crucial part of contemporary imaging methods as it may greatly increase the effectiveness of machine learning models by repeatedly improving population members depending on performance criteria.

Adaptive mutation techniques, which dynamically modify their mutation parameters in response to input from the optimization process, have emerged as a result of recent developments. In the context of medical image registration, for example, Farda et al. introduced a self-adaptive mutation strategy that uses feedback from prior iterations to inform the magnitude of perturbations, improving convergence rates [151]. Because medical surroundings are dynamic, this method is especially useful in real-time imaging applications where quick adaption is required. In a similar vein, Chen et al. investigated hybrid mutation techniques that enhance picture segmentation procedures by fusing the concepts of genetic

algorithms with conventional DE [152]. According to their results, in scenarios with high dimensionality and noise, these adaptive approaches perform noticeably better than their conventional equivalents.

Moreover, the development of recombination techniques has advanced significantly with the introduction of multimodal differential evolution (DE). The multimodal framework is especially well-suited for complicated medical imaging problems, where heterogeneity in forms and structures often complicates the optimization process, since it stresses the capacity to investigate many peaks in the solution terrain [153]. This method efficiently leverages the advantages of both exponential crossover and classification-based recombination to improve the whole image analysis process by allowing the simultaneous optimization of several segments inside a single picture.

One such use is in the field of remote sensing, where DE has been used to optimize image classification systems' parameter settings. The usefulness of DE in refining spectral unmixing techniques was shown by Ramadas and Abraham [154], who also highlighted its capacity to reduce pixel categorization mistakes and improve the overall accuracy of remote-sensed pictures

In a similar vein, Thakare et al. examined the use of DE in the analysis of electroencephalogram (EEG) signals, emphasizing the improvement of feature extraction and selection procedures [155]. DE was used in their work to determine the most relevant characteristics from the EEG data for mental state categorization tasks. The scientists demonstrated DE's ability to traverse the intricate, high-dimensional parameter spaces characteristic of EEG data by using crossover and mutation procedures to improve signal quality and classification accuracy

3.3.2 Swarm Algorithms

- **Particle Swarm Optimization (PSO)**:

The use of PSO has shown to be important in the context of medical imaging, especially in fields as crucial as computed tomography (CT) and magnetic resonance imaging (MRI). Advanced optimization strategies that may improve imaging efficiency without sacrificing the authenticity of the results are required due to the rising complexity and demand for high-quality medical images.

For example, enhancing signal-to-noise ratio (SNR), contrast resolution, and overall diagnostic capabilities in magnetic resonance imaging (MRI) requires fine-tuning parameters such as repetition time (TR), echo time (TE), and flip angles [156]. Similar to this, in CT imaging, lowering radiation doses without sacrificing picture quality requires adjusting factors including tube current, rotation time, and slice thickness [157].

Particles' trajectory toward ideal solutions gets increasingly accurate as they update their velocities via the inertia weight and cognitive and social components. These velocity modifications have significant effects on MRI and CT imaging, especially with regard to picture resolution and diagnostic precision.

For example, Ma and Hu show how the optimization of MRI pulse sequence parameters may be significantly improved by dynamically adjusting particle velocities [158]. The research shows that adjusting the velocity not only improves MRI scan contrast resolution but also drastically cuts down on image collection time without sacrificing diagnostic quality.

Similar to this, El Amoury et al. emphasize how velocity updates affect CT parameter optimization, particularly with regard to dose reduction strategies [159]. Their study shows that the capacity to get lower radiation doses while preserving picture integrity is directly correlated with precisely calibrated velocity updates. The scientists contend that PSO makes it easier to find ideal CT settings that strike a compromise between patient safety and diagnostic performance by more effectively directing the particles across the search area. In a therapeutic context, when reducing radiation exposure is of utmost importance, this is especially important.

Recent research has shed light on how adaptive inertia weight processes affect PSO convergence rates in medical imaging applications. For example, Ekanem et al. showed that an adaptive approach to inertia weight may improve the algorithm's capacity to focus on optimum solutions more successfully than static inertia weights [160].

Furthermore, utilizing adaptive inertial techniques in PSO to optimize imaging parameters, Adhikari et al. observed significant increases in picture quality and acquisition times [161]. According to their research, the adjustable inertia weight accelerated the convergence to high-quality imaging results by facilitating a more effective exploration of the parameter space.

PSO was used to improve the parameters required for MRI tumor classification tasks in a groundbreaking research by Thangamani et al. [162]. The study focuses on tumor segmentation in MRI images, where precise tumor delineation and subsequent diagnosis depend on the proper setup of imaging parameters. The authors were able to determine the ideal set of parameters, including echo duration, repetition time, and flip angle, by using PSO.

- **Artificial Bee Colony (ABC)**:

According to Shaban and Yasin, the ABC algorithm's complex structure has many phases, including the scout, employed, and observer phases, each of which has a unique function in efficiently exploring the search universe of possible features [163]. These stages improve the algorithm's capacity to converge toward ideal feature subsets that improve the interpretability of mammography pictures while still being pertinent.

The ABC algorithm is becoming more and more useful in medical diagnostics, especially when it comes to mammography, according to recent research. Tawil and Dakkak draw attention to a growing corpus of research that illustrates the use of ABC in a variety of medical domains, showcasing its potential to

expedite feature selection and enhance diagnostic results[164]. A wide range of feature candidates are taken into consideration during the scout phase, which is in charge of investigating new areas of the feature space. In mammography, where the correlations between characteristics may be complex and multivariate, this is vital. The scout phase greatly improves the robustness of diagnostic models by enabling the algorithm to investigate feature combinations that were previously ignored.

As research in this field continues to advance, the potential advantages of the ABC algorithm in medical imaging, particularly through its structured phases, warrant further investigation and development [165]. This stage is critical to the identification of pertinent aspects in mammography that may not have been previously taken into account.

Finding important characteristics in mammography is crucial because they have a direct impact on the precision and dependability of diagnostics carried out on pictures used for breast cancer screening. According to Ali et al., scout bee-led investigation often reveals characteristics that might greatly improve the distinction between benign and malignant tissues in mammography, hence strengthening the diagnostic process's robustness[166]. In mammography, characteristics including texture, shape, and intensity changes are crucial markers. Given the intricacy of the required differentiations in medical pictures, where even little alterations might signify severe disease changes, scout bees' ability to identify such traits is very crucial.

Kolli and Parvathala, who showed that incorporating the scout phase into the ABC algorithm produced a more thorough search for pertinent features in mammographic datasets, leading to optimal performance outcomes in classification tasks, offer supporting evidence for the scout phase's efficacy in enhancing feature sets for medical imaging tasks [167]. Their results demonstrate how adding a robust scout phase allows the algorithm to adapt dynamically to the data, making it easier to locate novel features that improve model performance without being limited by traditional or already defined characteristics.

- **Grey Wolf Optimizer (GWO)**:

Since GWO is crucial for improving the effectiveness and efficiency of picture classification and segmentation algorithms, its application in the context of medical imaging has proved beneficial. For diagnosis and treatment planning in medical settings, precise picture interpretation from sources including MRIs, CT scans, and X-rays is essential. However, the proper tuning of the hyperparameters of machine learning classifiers, including Random Forests (RF) and Support Vector Machines (SVM), is crucial to their performance. Optimizing the parameters that control the learning process is known as hyperparameter tuning, and it may have a big impact on the predictive power and performance of the model [168].

Using improved optimization techniques like GWO guarantees that classifiers like SVM and RF are well-suited to the difficulties of actual clinical settings, given the growing amount and complexity of medical imaging data. Grey wolves (Canis lupus) have a complex social structure within their packs, characterized by hierarchical leadership and cooperative hunting strategies that maximize their foraging efficiency. The ability to precisely tune hyperparameters greatly contributes to achieving high performance in medical imaging tasks, highlighting the relevance of the GWO in advancing both the accuracy and efficiency of machine learning-driven diagnostic tools in the medical field.

Similarly, by utilizing GWO's cooperative behavior, the hyperparameters pertaining to the number of trees, maximum depth, and minimum samples per split for RF classifiers can be effectively adjusted, guaranteeing a varied exploration of hyperparameter combinations while focusing on those that produce better classification outcomes. Thus, an understandable framework for designing optimization techniques is provided by the wolf pack behavior analogy. In the end, GWO shows a considerable capacity to improve medical imaging efficiencies via hyperparameter tuning for SVM and RF classifiers, resulting in improved diagnostic accuracy and performance [169]. Hyperparameter tuning is a crucial step in developing effective classifiers for medical imaging applications, enhancing their predictive capabilities.

Studies have shown that GWO can perform faster and more accurately than conventional tuning techniques. Specifically, a more comprehensive search of the hyperparameter space is made possible by its capacity to escape local optima via the variety of its pack members. Muryadi et al. demonstrated significant gains in classification accuracy over conventional techniques, highlighting the importance of adopting GWO for hyperparameter tweaking in SVMs in medical imaging[170]. The effectiveness of GWO in fine-tuning RF classifiers was also shown by Abuya et al. (2024), who showed that the leader-following strategy built into GWO significantly improves predictive ability [171].

3.4 Role of Metaheuristics in Medical Image Classification

In order to handle the complex optimization problems present in medical image classification pipelines, metaheuristic algorithms have become effective tools. Metaheuristics use principles inspired by nature to effectively explore high-dimensional, multimodal solution spaces, in contrast to standard deterministic approaches that often need explicit issue formulation and gradient information. From feature engineering to model tuning, its application covers many crucial steps in the classification process, allowing practitioners to get around computational constraints and increase diagnostic accuracy without requiring a lot of human effort.

3.4.1 Feature Selection and Optimization

In medical imaging, where the "curse of dimensionality" may worsen model generalization and raise computing costs, choosing pertinent radiomic or deep features from high-dimensional spaces is a crucial

difficulty. By automatically identifying compact and discriminative feature subsets while maintaining discriminative power, metaheuristic-based feature selection techniques lower model complexity and enhance interpretability.Selecting relevant radiomic or deep features from high-dimensional spaces [172].

In this situation, GA have proven very effective. They repeatedly evolve a population of potential feature subsets based on fitness criteria that are generated from the performance of downstream classifiers. For instance, GA-based feature selection has proven successful in identifying relevant radiomics characteristics from hundreds of candidates in the diagnosis of breast cancer. By selecting characteristics that are most predictive of the illness state, harmony search and genetic algorithms have also been used to diagnose sarcopenia in medical picture analysis, exceeding baseline classifiers and outperforming conventional statistical techniques [173].

3.4.2 Hyperparameter Optimization

Model performance is greatly impacted by the computationally demanding process of fine-tuning hyperparameters, which include learning rates, dropout rates, regularization coefficients, SVM kernel parameters, and KNN neighborhood sizes. Given the exponential increase of the hyperparameter search field, manual grid search or random search are often unfeasible, especially when designing lightweight CNN systems for mobile or edge deployment [174].

In order to overcome this difficulty, genetic algorithms automatically evolve ideal hyperparameter configurations over the course of subsequent generations, enabling the fine-tuning of network depth, filter sizes, activation functions, and training dynamics without the need for a thorough enumeration. In comparison to traditional random or Bayesian optimization techniques, recent research show that ensemble genetic algorithm approaches combined with CNN models produce much greater classification accuracy (e.g., 98.46% in acute lymphoblastic leukemia identification). Additionally, particle swarm optimization has been effectively used to dynamically adjust CNN learning parameters and optimize SVM hyperparameters, allowing for quick convergence and better generalization on difficult medical imaging datasets like Alzheimer's disease classification from MRI [175].

3.4.3 Feature Fusion and Weight Optimization

Optimizing the fusion weights that balance contributions from each modality is crucial for integrating complementary diagnostic information while reducing noise and artifacts in multimodal medical imaging systems that combine complementary data from various acquisition modalities (e.g., CT+PET, MRI+ultrasound). This is a nonlinear optimization issue in and of itself that greatly benefits from the use of metaheuristic techniques.

Fusion weight matrices in multimodal CT-MRI fusion systems have been effectively optimized using particle swarm optimization in conjunction with sophisticated image decomposition algorithms, yielding higher fused picture quality while preserving computing efficiency. In a similar vein, the gradient compass-based fusion of multimodal pictures has been optimized using a variable-order fractional Darwinian particle swarm optimization algorithm, which has shown better results than state-of-the-art methods in terms of image quality measures. Additionally, discrete wavelet transforms and arithmetic optimization methods have been used to automatically improve fusion rule parameters, allowing for the smooth integration of complementary functional and anatomical imaging data from various modalities [176].

3.4.4 Hybrid Classifiers

Compared to single classifiers, hybrid classification frameworks that include many base learners (such as deep neural networks, support vector machines, and closest neighbor techniques) with metaheuristic-driven optimization provide improved accuracy and resilience. In 3D medical image registration and multimodal alignment tasks, hybrid particle swarm optimization, which integrates genetic algorithm concepts like subpopulation and crossover into traditional PSO, has proven to perform better than gradient descent, standard GA, and traditional PSO techniques [177].

By combining CNN-based deep feature extraction, large margin nearest neighbor metric learning, and swarm intelligence-based feature selection (PSO and GWO), a new hybrid framework for brain tumor classification achieves four-class classification of gliomas, meningiomas, pituitary tumors, and healthy tissue with high accuracy and computational efficiency. This cohesive strategy establishes a model for sophisticated CAD systems that use many optimization paradigms at once by showcasing the synergistic advantages of integrating deep learning, metric learning, and swarm optimization [178].

3.5 Comparison of Metaheuristics in CAD Systems

The selection of an appropriate metaheuristic algorithm for medical imaging and computer-aided diagnosis tasks requires careful evaluation of algorithmic strengths, limitations, and the specific characteristics of the optimization problem at hand. This section provides a systematic comparison of major metaheuristic families, highlighting their convergence properties, computational efficiency, and suitability for different CAD scenarios.

3.5.1 Strengths and Limitations

When investigating high-dimensional search spaces that include discrete and continuous choice factors, genetic algorithms (GA) and differential evolution (DE) show remarkable adaptability. While GA's crossover and mutation operators efficiently recombine promising partial solutions, its chromosome-based representation naturally encodes both continuous parameters (learning rates, regularization coefficients) and categorical choices (network architecture decisions, classifier type selection). In

comparison to GA, DE's vector difference-based mutation process offers better convergence on unimodal and smooth continuous problems, often needing much fewer function evaluations and population members. DE is especially appealing for parameter optimization problems when fitness assessment (model training and validation) is computationally costly because of its efficiency [179].

However, compared to swarm-based systems, evolutionary methods show much slower convergence rates, usually needing higher population sizes and more iterations to obtain acceptable solutions. The computational expense of evolutionary approaches becomes prohibitive in medical imaging applications, such as real-time CAD systems, surgical guiding, or emergency department screening, when quick hyperparameter optimization or interactive model refining is required. Additionally, the number of generations and population size have a significant impact on the quality of the solution; too many generations squander computing resources investigating declining advances, while too few generations run the danger of premature convergence to inferior local optima. Because of this sensitivity, rigorous problem-specific calibration is required, which lowers algorithm resilience across a variety of medical imaging applications [179].

Compared to evolutionary methods, swarm intelligence algorithms (PSO, WOA, and GWO) achieve noticeably faster convergence, allowing deployment in clinical settings with limited processing power, such as edge computing systems, mobile health applications, and real-time point-of-care diagnostic devices. They significantly lessen the burden of algorithm-specific hyperparameter tuning because to their intrinsic simplicity, which usually just requires population size and two to four control coefficients. The algorithm can automatically move from global discovery (early iterations) to local refinement (later iterations) thanks to the dynamic adjustment of exploration-exploitation balance provided by PSO's inertia weight decay, GWO's hierarchical leadership structure with adaptive encircling-to-attacking behavior, and WOA's spiral contraction mechanism.

When navigating highly multimodal objective landscapes with many isolated local optima, a typical situation in medical imaging feature selection where the search space may contain thousands of locally optimal feature subsets, swarm algorithms' vulnerability to premature convergence and stagnation is a critical limitation. Particles or wolves lack the variety and exploratory vigor to flee and find far superior solutions if the population prematurely converges on a suboptimal area. Large feature dimensions (hundreds to thousands of radiomic features), high-dimensional parameter spaces (deep network architectures with numerous configurable layers), noisy or non-smooth objective functions (validation performance evaluated on limited clinical datasets), and multimodal landscapes (many different configurations achieving similar intermediate accuracy but differing significantly in downstream clinical utility) are some of the inherent characteristics of medical imaging that exacerbate this vulnerability.

This trade-off is empirically demonstrated by comparative studies on the diagnosis of cardiovascular disease: GA-based approaches eventually found better feature subsets and classifier configurations by preserving population diversity and continuing exploration even after initial convergence, while PSO achieved faster initial performance improvement and required fewer iterations to reach the first acceptable solution. Additionally, the robustness of swarm algorithms is highly sensitive to population size; larger populations improve exploration coverage at the expense of increased computational time, creating a problematic trade-off in time-sensitive clinical deployment scenarios, while smaller populations speed up convergence rate but increase stagnation probability [179].

3.6 Metaheuristics and AI: Toward Hybrid Models

3.6.1 Deep Network Optimization

Deep learning has significantly advanced medical image analysis due to its ability to automatically learn hierarchical feature representations from complex imaging data. Convolutional Neural Networks (CNNs), in particular, have become the backbone of medical image classification, segmentation, and detection tasks, enabling improved performance in applications such as tumor detection, anatomical structure segmentation, and disease diagnosis. Lower layers of CNNs typically capture low-level features such as edges and textures, while deeper layers learn high-level abstractions relevant to clinical interpretation.

Despite their effectiveness, the performance of deep learning models is highly dependent on the choice of network architecture and hyperparameters, including learning rate, batch size, number of layers, filter sizes, and optimization strategies. Poor hyperparameter selection can significantly degrade model performance, leading to overfitting, underfitting, or inefficient generalization to unseen data [180]. Given the high stakes associated with medical decision-making, achieving optimal model performance is essential to ensure reliable and accurate diagnostic outcomes.

Metaheuristic optimization techniques offer a powerful solution to deep network optimization challenges. Algorithms such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO) are capable of efficiently exploring complex, high-dimensional, and multimodal hyperparameter spaces. Unlike conventional grid search or gradient-based approaches, metaheuristics can avoid local optima and identify globally optimal or near-optimal configurations. As demonstrated by Kadhim et al., these techniques enable systematic and adaptive optimization of deep learning models, significantly enhancing performance in medical imaging tasks [181].

3.6.2 Limitations of Non-optimized Deep Learning Models

Although deep learning models have demonstrated remarkable success in medical imaging, non-optimized models suffer from several critical limitations. One major challenge is *overfitting*, particularly due to the limited size, high dimensionality, and noise characteristics of medical imaging datasets. When

models overfit, they memorize training data instead of learning generalizable patterns, resulting in reduced diagnostic reliability when applied to new clinical cases [182].

Another limitation lies in the computational burden associated with training deep neural networks. Large architectures with poorly tuned hyperparameters can lead to excessive training times, inefficient resource utilization, and impractical deployment in real-world healthcare environments. This is especially problematic in time-sensitive scenarios such as emergency diagnostics or intraoperative imaging, where rapid and accurate decision-making is crucial.

Traditional optimization techniques, including manual tuning and gradient-based optimization, are often insufficient to address these challenges. Such approaches are prone to becoming trapped in local minima and typically require extensive computational effort to explore large hyperparameter spaces. Consequently, non-optimized deep learning models may exhibit suboptimal accuracy, poor generalization, and limited robustness across heterogeneous datasets from different institutions.

3.6.3 Toward Hybrid Solutions

To overcome the limitations of non-optimized deep learning models, hybrid frameworks combining deep learning with metaheuristic optimization algorithms have emerged as a promising research direction. These hybrid approaches leverage the powerful feature-learning capabilities of deep neural networks while using metaheuristic algorithms to optimize hyperparameters, network architectures, and feature selection processes.

Numerous studies demonstrate the effectiveness of this integration. For instance, Bohmrah and Kaur reported significant improvements in MRI abnormality detection by optimizing CNN hyperparameters using a metaheuristic algorithm, resulting in enhanced accuracy, specificity, and generalization [183]. Similarly, Shetty et al. employed PSO for feature selection in deep learning-based medical image retrieval, achieving superior diagnostic performance and reduced computational complexity [184].

Genetic Algorithms have also proven effective in optimizing deep learning architectures. Singh et al. showed that GA-based hyperparameter tuning improved cancer diagnosis accuracy by enhancing model robustness and mitigating overfitting [185]. Likewise, Awotwe et al. demonstrated that PSO-optimized deep neural networks significantly reduced training time while increasing detection sensitivity in lung imaging applications [186].

Beyond imaging, hybrid approaches have been successfully applied to time-series medical data. Mohamed et al. integrated Ant Colony Optimization with RNNs to improve predictive modeling in chronic disease management, highlighting the versatility of metaheuristic-enhanced deep learning frameworks [187].

Overall, the synergy between deep learning and metaheuristic optimization offers a robust pathway toward building accurate, efficient, and generalizable diagnostic systems. These hybrid models address key challenges in medical image analysis, including high-dimensional optimization, computational efficiency, and cross-institutional generalization. As medical data continues to grow in volume and complexity, such integrated approaches are expected to play a pivotal role in the future of intelligent healthcare systems.

3.7 Conclusion

In medical imaging, metaheuristic algorithms have become essential tools for developing, refining, and implementing reliable computer-aided diagnostic systems. Metaheuristics offer adaptable, problem-agnostic frameworks that can traverse the intricate, multimodal, and high-dimensional solution spaces present in medical imaging tasks, in contrast to conventional deterministic optimization techniques that demand explicit gradient information, smoothness assumptions, or convex problem structure. Metaheuristics are a fundamental part of modern CAD pipelines because of their proven efficacy in a variety of optimization domains, including feature selection and dimensionality reduction, hyperparameter tuning of machine learning and deep learning models, multimodal fusion weight optimization, and hybrid classifier design [188].

A pragmatic, problem-adapted selection technique rather than algorithm-agnostic application is justified by the complimentary strengths and drawbacks shown by a comparative examination of the main metaheuristic families. Neural architecture search and classifier selection tasks where solutions span both categorical (layer types, activation functions) and continuous (regularization weights, learning rates) domains are especially well-suited for evolutionary methods (GA, DE) because they perform well in mixed discrete-continuous spaces and maintain robust diversity. The real-time and computational limitations of clinical deployment situations are addressed by swarm-based algorithms (PSO, WOA, GWO), which provide excellent convergence speed and little parameter adjustment. For combinatorial issues like feature subset selection and medical picture thresholding, collective techniques (ACO) provide specialized efficacy at the expense of higher computing overhead. Principled algorithm selection that is in line with particular medical imaging applications and clinical deployment constraints is made possible by an understanding of these context-dependent trade-offs between exploration and exploitation, convergence speed and solution stability, and computational efficiency and solution quality.

Metaheuristics have been shown to be useful in conventional machine learning, and their incorporation into hybrid deep learning systems is a logical and experimentally supported extension. With millions to billions of trainable parameters, intricate non-convex loss landscapes with many saddle points and local minima, sensitivity to initialization and learning rate schedules, and susceptibility to overfitting on small clinical datasets, deep neural networks present significant new optimization challenges. Metaheuristics

handle complementary but crucial optimization tasks, such as automated network architecture discovery (number and type of layers, filter dimensions, skip connections), systematic tuning of training hyperparameters (learning rate schedules, dropout rates, batch normalization parameters, regularization strengths), and intelligent ensemble design combining multiple specialized networks for multimodal or multitask medical imaging objectives. Gradient-based training (backpropagation with stochastic gradient descent variants) is still computationally required for forward-backward pass execution. Recent developments in hybrid ensemble frameworks, GA-based neural architecture search, and PSO-optimized deep feature selection show that metaheuristics can significantly enhance final model performance (finding non-obvious configurations better than hand-crafted designs) as well as the design process (lessening the burden of manual architecture engineering).

This synergistic integration is demonstrated and a model for wider clinical translation is established by the domain-specific application of metaheuristic-enhanced CAD systems to specific medical imaging challenges, such as the detection and grading of breast cancer from histopathological imagery, the classification and segmentation of brain tumors from multimodal MRI, and the detection of pulmonary disease from chest radiography and CT. Large publicly accessible annotated datasets that enable robust algorithm evaluation and cross-validation, a variety of imaging modalities that enable multimodal fusion studies, distinct clinical decision-making endpoints that enable objective performance assessment, and significant disease burden and mortality that justify investment in improved diagnostic systems are all shared by these domains, making them perfect testbeds.

4 Deep Learning for Medical Image Analysis

4.1 Introduction

Medical imaging has revolutionized the field of healthcare, allowing for non-invasive visualization of the human body and facilitating the detection and diagnosis of diseases. With the advent of deep learning, medical image analysis has witnessed significant advancements, enabling computers to learn from large datasets and make accurate predictions. Deep learning, a subset of machine learning, has shown remarkable potential in medical imaging, from automating image analysis tasks to improving diagnosis accuracy.

Traditional machine learning algorithms can handle many important tasks effectively, but they have not been able to address core AI challenges such as complex medical image processing. The emergence of deep learning was driven by the inability of these conventional methods to generalize well to complex AI problems.

For these reasons, working with high-dimensional data makes it exponentially harder to generalize to new examples, and the mechanisms used in traditional machine learning to achieve generalization are not enough to learn complex functions in high-dimensional spaces. High computational expenses are also frequently associated with such environments. To get around these and other challenges, deep learning was created.

Deep learning represents one of the most rapidly evolving disciplines, spanning both academic research and industrial applications. This field encompasses multiple neural network architectures, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and Deep Belief Networks (DBNs), among others. The ability of these networks to extract high-level abstractions from large datasets while outperforming more conventional machine learning techniques has led to their widespread recognition. Deep learning demonstrates versatility across various data modalities, including images, audio, and textual information. Given our focus on image-based datasets, this section emphasizes deep learning applications in image analysis broadly and medical imaging specifically [189].

The integration of deep learning technologies into medical imaging has revolutionized diagnostic capabilities, offering unprecedented accuracy, efficiency, stability, and scalability in clinical applications. Recent advances in deep learning based medical image analysis have enabled breakthrough achievements across multiple human body systems, including

neurological, cardiovascular, digestive, and skeletal disorders, establishing deep learning as an indispensable tool in modern healthcare diagnostics [189].

4.2 Deep Learning Concepts and Motivation

Deep learning is a specialized subfield within machine learning that involves the use of multi-layered artificial neural networks (ANN) to analyze and interpret data. These multi-layered models are designed to learn and represent data through increasingly complex, hierarchical abstractions.

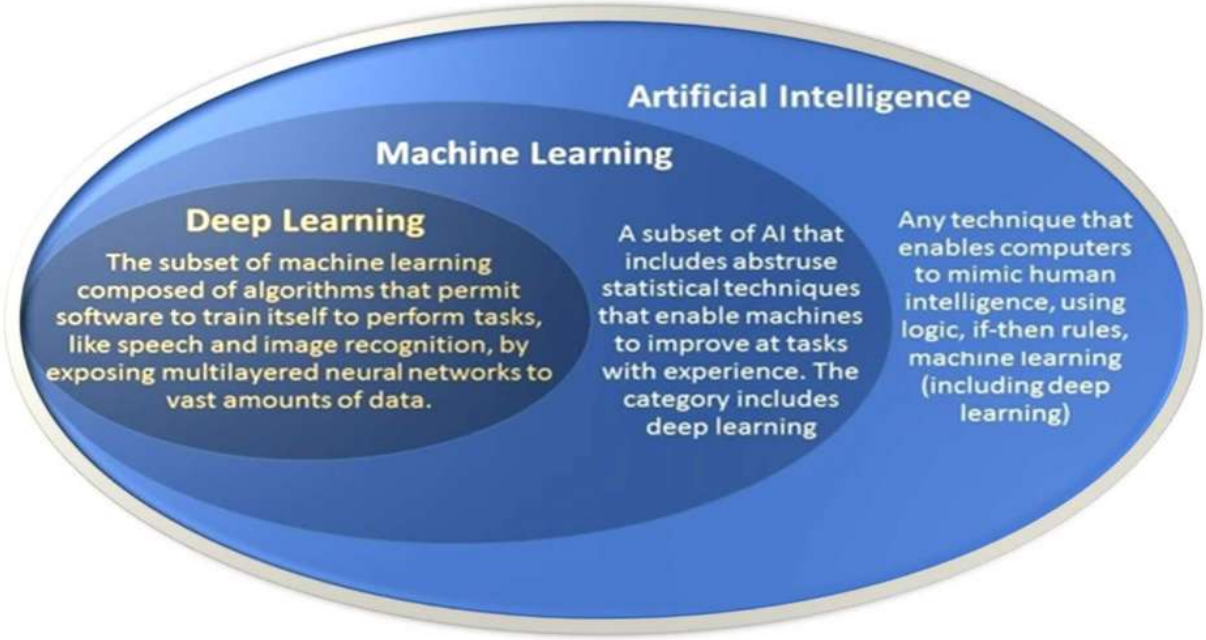


Figure 4.1: Overview of Fields: AI, ML, DL, and Data Science [190]

McCulloch and Pitts developed the first mathematical framework for neural computation, proposing that biological neurons could be described as computational devices capable of performing logical operations. Their pioneering work, titled "A Logical Calculus of the Ideas Immanent in Nervous Activity," established the theoretical groundwork for artificial neural networks by demonstrating how simple binary elements, when interconnected in various configurations, could execute all logical functions [191]. Perceptron was introduced by Rosenblatt in 1958[192], At that time, it was regarded as a highly innovative ANN. The limits of the perceptron were primarily exposed in a 1969 book titled "Perceptrons" by Minsky et al. [193]. Minsky demonstrated that the perceptron cannot learn a non-linearly separable function such as XOR. The backpropagation algorithm [194] was introduced in the late 1980s, which revitalised research on neural networks and restored their strength.

A Multi-Layer-Perceptron is an ANN that consists of an input and output layer, as well as one or more hidden layers, each of which contains multiple hidden units.

Gradient Descent (GD), a first-order technique used to minimize the error function used to update the parameters, is used to create a backpropagation process in order to train an ANN. Despite the long training time, this method has been successfully used for training ANNs for a long time.

A significant limitation of GD lies in its computational inefficiency, as it processes the entire training dataset to execute a single parameter adjustment, resulting in considerable time overhead. To address this computational bottleneck, Stochastic GD (SGD) was developed, which employs a subset of training samples rather than the complete dataset for parameter updates [195], [196].

Contemporary optimization techniques have emerged that demonstrate superior performance compared to SGD, including advanced algorithms such as ADAM and RMSprop. These modern optimizers have established themselves as more effective alternatives for training artificial neural networks [196]. These days, several ANN optimization methods, such ADAM [197] and RMSprop [198], have shown themselves to be even more successful than SGD.

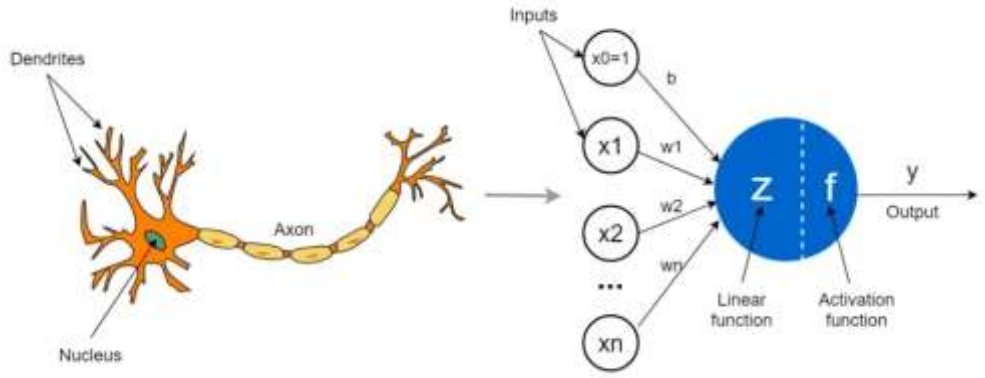


Figure 4.2: An artificial neuron model based on the morphology of a real neuron.

Since their inception, artificial neural networks (ANNs) have undergone significant advancements, with a major breakthrough occurring in 2006 when Hinton and his collaborators introduced the greedy layer-wise training algorithm for deep belief networks [199]. This technique enabled effective training of deeper neural architectures, marking the rise of deep learning as a distinct field within machine learning. Several factors have propelled the success

of deep learning over traditional machine learning approaches, most notably the development of high-performance computational resources such as GPUs and TPUs, which have mitigated the challenges associated with training very deep models. Additionally, the abundance of large datasets has played a crucial role in enabling these algorithms to learn complex patterns with greater accuracy. Overall, machine learning, and particularly deep learning, can be categorized into three primary approaches: supervised, semi-supervised, and unsupervised learning [200]

4.2.1 Supervised learning

Supervised learning represents a methodology that leverages annotated datasets to establish associations between specific inputs and their target outputs. The approach involves iterative refinement of model parameters to reduce loss functions until the system achieves near-optimal prediction accuracy. Following successful training, the model demonstrates capability to generate accurate predictions for novel, previously unobserved data. Essentially, the primary objective of supervised learning frameworks is to reliably forecast appropriate output categories for new input instances [201].

Supervised methodologies primarily address two fundamental computational challenges: classification and regression tasks. Classification algorithms receive training examples accompanied by predetermined labels during the learning phase. The central purpose of these classification systems is to establish mappings between input features and their appropriate categories through knowledge acquired during prior training processes.

Classification frameworks operate in either binary configurations (involving two distinct classes) or multi-class scenarios. For instance, categorizing pulmonary nodules as benign versus malignant exemplifies binary classification, whereas distinguishing among various canine breeds represents a multi-class classification problem.

Numerous algorithmic approaches exist for addressing classification challenges, with selection criteria determined by dataset characteristics and problem requirements. Notable classification methodologies include ANNs, CNN, Random Forest algorithms and Support Vector Machines (SVM).

Conversely, regression constitutes a predictive analytical framework that seeks to establish relationships between dependent and independent variables. Unlike classification approaches,

regression techniques focus on forecasting continuous numerical values including temporal measurements, revenue figures, performance metrics, and similar quantitative outcomes.

The regression domain encompasses various algorithmic implementations, including linear regression, logistic regression, and polynomial regression approaches, each designed for specific types of continuous prediction scenarios.

4.2.2 Unsupervised learning

A subfield of machine learning called unsupervised learning uses datasets without labeled outputs to train algorithms. The system looks for hidden groups, patterns, or structures in the data. Unsupervised techniques investigate the underlying distribution of data to extract meaningful representations, in contrast to supervised learning, which depends on input–output pairings [202].

In the literature, unsupervised learning techniques are discussed in relation to association or grouping issues. Clustering is the process of assembling or grouping data pieces into groups that are believed to be connected in some way. Finding the ideal criterion for grouping data is the difficult part of clustering. Cosine, Jacard, and Euclidean distances are just a few of the proximity metrics that may be used to discover a specific clustering solution. Numerous clustering techniques, including K-means, fuzzy K-means, mixture of Gaussian, and hierarchical clustering, use such proximity metrics.

Clustering algorithms are often used in the area of medical imaging to find underlying patterns or structures in complicated biological data. To distinguish between healthy and sick tissues, for example, unsupervised clustering may be used to segment regions of interest in MRI or CT images without previous annotations. Because they allow the identification of intrinsic data categories based just on picture attributes like intensity, texture, or form, such methods are especially useful when labeled datasets are few or unavailable. Additionally, clustering may promote more individualized and accurate diagnostic procedures by helping with patient stratification, illness subtype identification, and anomaly detection.

4.2.3 Semi-supervised learning

This kind of learning, as its name suggests, lies in the middle between supervised and unsupervised learning approaches. Unlabeled data typically outnumbers labelled data. More unlabeled data may often improve the learning process. For example, Jeff Bezos claims that

Amazon Alexa's additional unlabeled data helped it achieve more accuracy than it had previously(reference). Reinforcement learning is one of the most widely used learning approaches that uses a semi-supervised learning strategy. In recent years, this learning method has become increasingly well-liked and has shown itself to be highly reliable and successful. It is extensively utilized in self-driving automobiles and game-based algorithms [203].

The foundation of reinforcement learning is rewards and punishments depending on the actions taken. Its goal of reinforcement learning is to increase the total reward. The Google Deep Mind Group introduced reinforcement learning in 2013 [204] when they presented the deep Q-network algorithm, which not only performs very well in the video game Breakout but also surpasses all other machine learning techniques.

Learning Type	**Data Used**	**Goal**	**Example in Medical Imaging**
Supervised	Labeled data	Learn input → output mapping	Classifying tumor as benign/malignant
Unsupervised	Unlabeled data	Find hidden structure/patterns	Grouping images with similar features
Semi-Supervised	Few labeled + many unlabeled data	Improve performance with limited labels	Medical Image Classification Combined with Unsupervised Deep Clustering

4.3 Convolutional Neural Networks

Machine learning algorithms have the capacity to discern hidden relationships within data features, enabling them to make independent decisions without relying on explicit supervision. Most machine learning techniques have been introduced since the early 1980s, aiming to simulate human cognitive behaviors in processing diverse data types such as vision [205].

Nevertheless, Machine learning algorithms generally struggled to reach high levels of abstraction because of their limitations in handling data. This challenge led to the emergence

of a specialized type of neural network called CNNs in the late 1990s, designed specifically to process and interpret image data effectively. CNNs possess several key features that have enabled them to outperform traditional computer vision methods. These include hierarchical learning, which allows the model to learn features at multiple levels of complexity; automatic feature extraction, eliminating the need for manual intervention in identifying important image attributes; multi-tasking capabilities that allow the same network to perform various related tasks; and weight sharing, which reduces the number of parameters and computational overhead required for training and inference [205].

4.3.1 General concepts of CNNs

CNNs are among the best methods for processing image data, and they have shown excellent performance in a range of image recognition applications, such as segmentation, classification, and image detection [206], [207], [208]. Beyond academic research, the field of medical imaging has seen substantial interest from both industry and healthcare technology organizations. Numerous hospitals, research institutes, and medical AI companies have invested in developing and refining CNNs to improve diagnostic accuracy and efficiency. These efforts have led to innovative CNN architectures tailored for various imaging modalities, such as X-ray, CT, MRI, and histopathology slides enabling automated detection, segmentation, and classification of diseases. As a result, CNN-based systems are increasingly integrated into clinical workflows to assist radiologists, pathologists, and other medical professionals in delivering faster and more reliable diagnoses

The power of CNNs lies in their ability to make use of spatial relationships within data. The CNN architecture is made up of several processing stages, including convolutional layers, nonlinear activation functions, subsampling (or pooling) layers, and classification layers. These networks are multi-layered and hierarchical, where each layer applies a series of transformations. For instance, convolutional layers extract important features like shapes and patterns by scanning across the input with filters that capture various kinds of correlations. Following this, nonlinear activation functions introduce complexity into the feature maps by enabling the network to learn abstract representations. These activations highlight different patterns, helping the model distinguish meaningful differences within the images. Often, a normalization layer follows the activation to stabilize and speed up training [209]. To stabilize the outputs from the previous layer, normalization techniques are applied to keep the mean activation close to zero and the standard deviation near one. Following normalization,

subsampling (or pooling) layers are used to condense the feature maps, reducing their spatial size and helping to protect against distortions in the input data. Thanks to this integrated automatic feature extraction process, CNNs do not require separate manual feature engineering. As a result, CNNs can learn high-level and robust representations directly from new image inputs without the need for extensive preprocessing.

The very first concept resembling a CNN architecture was introduced under the name "Noncognition" or more precisely, the Neocognitron [210], developed by Hubel and Wiesel's work in neuroscience. Inspired by the biological visual cortex, the Neocognitron featured alternating layers that performed convolution-like operations and downsampling, enabling it to recognize visual patterns invariant to shifts, this architectural design was inspired by. Consequently, pattern recognition often adheres to the fundamental architecture of the human visual brain. For example, The dorsal visual stream or *how/where stream* (from the occipital cortex into the parietal cortex) processes the spatial relationships between objects and motion [211], [212]. Figure 4.3 shows a simplified view of how the visual system works compared to a CNN-based model. (a) When the eye captures visual information, it first passes through the lateral geniculate nucleus (LGN) before reaching the visual cortex located in the occipital lobe of the brain. This pathway processes the visual input, which then splits into two streams. The ventral stream is responsible for identifying and recognizing different parts of the image, while the dorsal stream focuses on understanding the spatial relationships and positioning between those parts. (b) In our CNN model, dataset images serve as inputs. The model is structured into multiple layers, including a baseline network that extracts important features from each image. These features are then fed into two separate branches: the ventral branch, which classifies the objects within the image, and the dorsal branch, which determines the objects' locations and the distances between them.

However, Yann LeCun's work "Gradient-based Learning Applied to Document Recognition" [213] in 1990 established an architecture influenced on Neocognitron for processing matrix-like topological data, and it was this work that brought CNN to prominence.

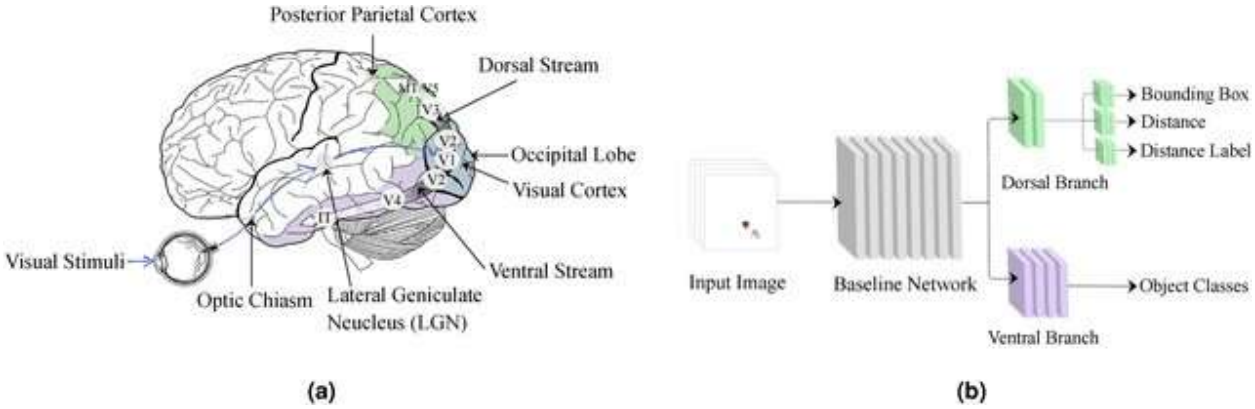


Figure 4.3: Visual Cortex and CNN: A Structural Comparison [213].

CNNs became popular in image recognition because they can learn features in a step-by-step way, from simple to complex. Deep CNNs outperform shallow ones when analyzing complicated data because they stack many layers that combine linear and non-linear processing. Additionally, the rise of big data and improvements in hardware like GPUs have played a major role in the recent success of deep CNNs, enabling them to reach and even surpass human performance levels [214], [215].

4.3.2 Overview of CNN Building Blocks

CNNs are widely recognized as one of the most popular machine learning techniques across various fields, especially those involving visual data. They excel at processing grid-like data structures through two key stages: feature extraction and feature classification. In the feature extraction phase, CNNs automatically learn to identify important characteristics in the data. Then, during feature classification, these learned features are used to categorize the input. A typical CNN architecture consists of several layers arranged in a pattern that alternates convolutional layers with activation functions and pooling layers, followed by one or more fully connected layers, as illustrated in Figure 3.4

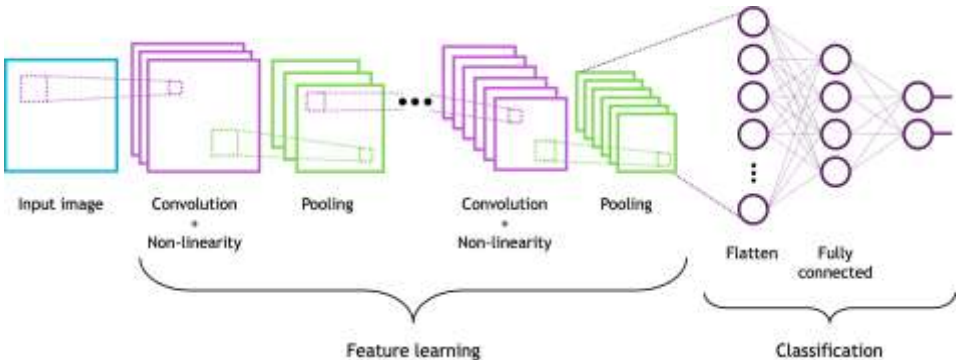


Figure 4.4: Basic CNN architecture.

In addition to using various mapping functions, CNN architectures also incorporate normalization techniques and dropout layers to boost performance [209], [216] . The next section will provide a detailed explanation of each component of the CNN architecture and how they contribute to achieving optimal results

4.3.2.1 Convolutional layer

A convolutional layer produces a map of abstract features by performing a convolution operation using relatively small, parameterized filters of size N × N. The nature of these filters determines the type of features extracted, such as edges, horizontal lines, shapes, or patterns. Formally, the convolution operation can be defined mathematically as follows:

$$y_k = f(W_k * x) \qquad 4.1$$

Where x represents the input image, W_k represents the convolution filter associated to the K^{th} feature map and the multiplication sign denotes the convolution operator. The function $f()$ applied to the feature map is the activation function.

Generally, when a convolution operation is performed with same padding, the shape of the input tensor remains unchanged. This means the height and width of the output feature map are the same as those of the input. However, if valid padding is applied instead, no extra pixels are added around the input edges, causing the output size to shrink compared to the input. In this way, the output dimensions depend on the chosen padding scheme.

This convolution operation, combined with non-linear transformations provided by activation functions (which will be discussed later), helps simulate the task of the visual cortex. The goal of applying these non-linear activations is to enhance the contrast of meaningful features, allowing the convolution layers to extract features at higher, more abstract levels.

There are different types of convolutional layers considering their purpose in dealing with the input volume, from which we can site the following:

- **Simple convolution** It is the commonly used convolution type on a standard CNN, it is the dot product of the same filter with some width/height shape as shown in the following figure3.5:

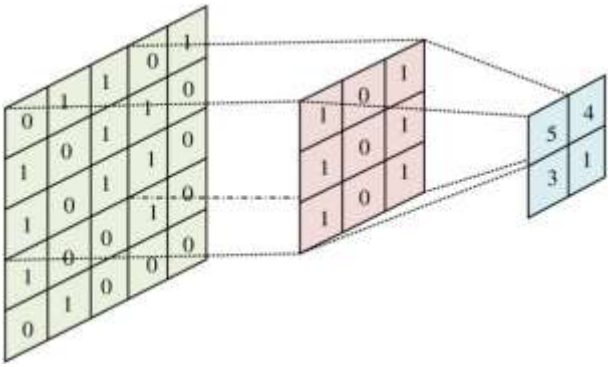


Figure 4.5: Simple convolution operation.

- 1 × 1 **convolution** This particular type of convolution was initially introduced in the "network-in-network" architecture [217], they were later incorporated into the Inception architecture [215]. The 1×1 convolution is used to reduce the dimensionality within the filter space, which helps lower the computational cost of the network. The figure 3.6 demonstrates how a 1×1 convolution effectively reduces the number of channels in the feature maps while preserving the spatial dimensions.

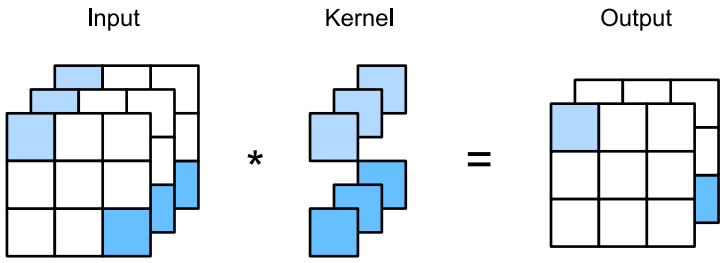


Figure 4.6: 1 ×1 convolution.

- **Flattened convolution** It operates with the same purpose as the 1x1 convolution — to reduce dimensionality within the network. However, instead of reducing just the feature dimension to 1, it reduces one of the spatial dimensions (either width or height) to 1. This is achieved through a sequence of one-dimensional filters applied along different directions in the 3D space of the feature map. Flattened convolution has been shown to deliver comparable results to traditional CNNs while significantly reducing the number of learnable parameters, leading to faster performance during both training and inference [218]. As shown in Figure 3.7, flattened convolution effectively decreases the number of learnable parameters while maintaining comparable performance to traditional 3D convolutions, leading to enhanced speed during training and inference

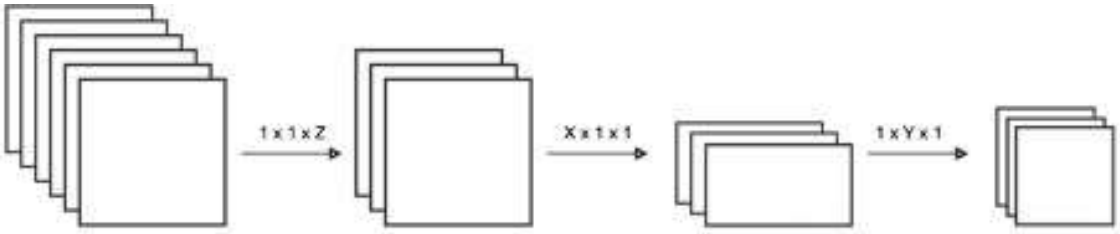


Figure 4.7: Flattened convolution

- **Spatial and cross-channel convolution** It is widely used in architectures like Inception. The main idea is to separate the operations that analyze cross-channel correlations from those that analyze spatial correlations, making the process more efficient. As shown in figure 3.8, instead of applying a single 3×3 filter that jointly considers spatial and channel information, the operation might first use a 3×1 filter to analyze one dimension, followed by a 1×3 filter for the other dimension. This decoupling allows the network to learn more effectively by independently focusing on channel-wise and spatial features, which can lead to improved performance and reduced computational costs.

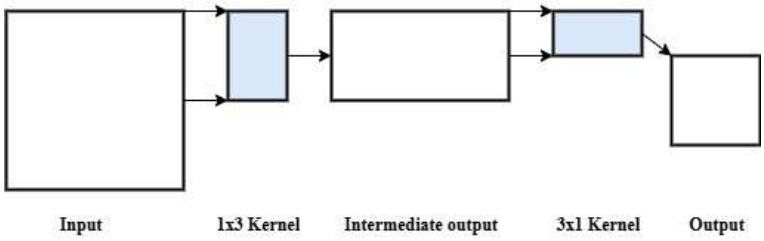


Figure 4.8: Cross-channel convolution.

- **Depth-wise convolution** Depth-wise convolution differs from spatial separable convolutions in that it applies convolutional filters independently over each channel (depth) of the input feature map without mixing information across channels. Specifically, instead of using filters that operate across all channels simultaneously, depth-wise convolution uses separate filters for each channel. This approach reduces computational complexity and the number of parameters while preserving spatial feature extraction within each channel. Figure 3.9 illustrates this operation:

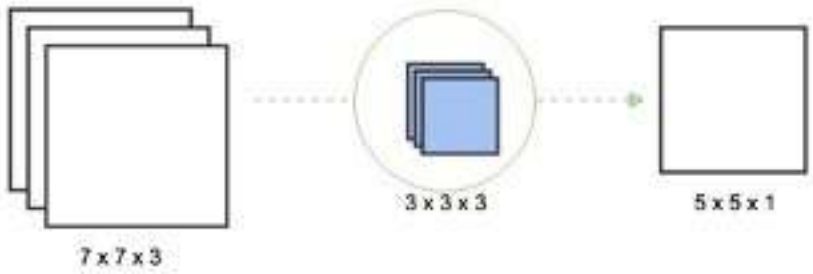


Figure 4.9: Depth-wise convolution.

- **Grouped convolution** It was first introduced in the AlexNet architecture [216]. The motivation behind using grouped convolutions was to optimize network performance by reducing computational complexity. This was achieved by dividing features into groups and distributing the computation across multiple GPUs, which helped overcome hardware memory limitations. The next figure illustrates the working scheme of grouped convolution.

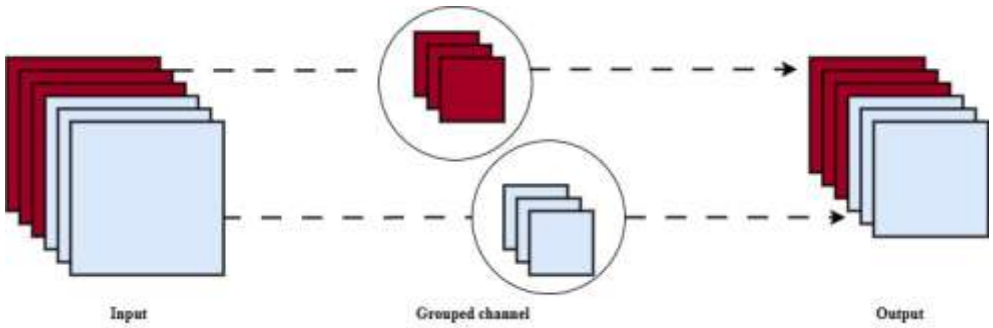


Figure 4.10 : Grouped convolution.

- **Shuffled grouped convolution** The concept of shuffled grouped convolution originates from Shuffle Net. The primary goal of this technique is to counteract the limitation where features derived from a certain channel are only influenced by a small subset of input channels, which can restrict information flow. As shown in Figure 3.11, shuffled grouped convolution rearranges the channels between groups, enabling better interaction across different channel groups and improving the network's representation power.

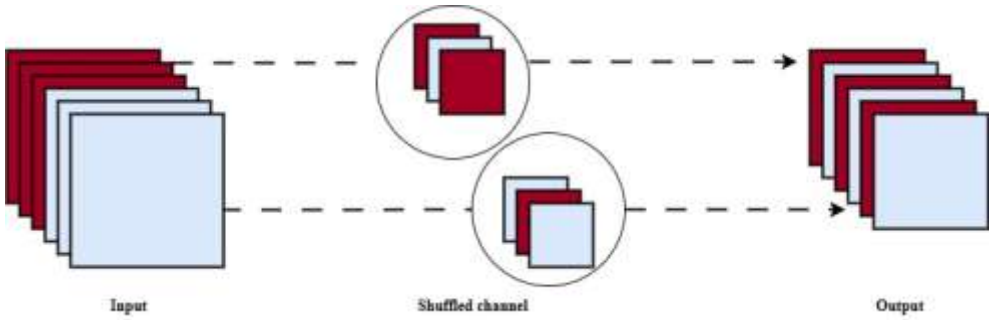


Figure 4.11: Shuffled-grouped convolution.

4.3.2.2 Activation functions

Activation functions are a fundamental component of deep convolutional neural networks, as they determine the network's output, influence predictive accuracy, and affect the computational efficiency of training. Their impact is substantial, often determining whether a network successfully converges or fails to do so. Consequently, selecting an appropriate activation function is a critical task, since an unsuitable choice may hinder convergence altogether.

Mathematically, activation functions define the transformation applied to a neuron's output, thereby controlling which units become activated based on their relevance to the learning process. This mechanism facilitates faster convergence while also constraining outputs within specific ranges, typically between 0 and 1 or −1 and 1. As CNN architectures have grown deeper, the computational burden on activation functions has increased, motivating the design of more efficient alternatives such as Swish [219], ReLU [220], and their numerous variants.

Among the different categories—binary step, linear, and non-linear—non-linear activation functions have emerged as the most widely adopted in CNNs. Their ability to capture complex data representations enables the network to learn intricate features and deliver more accurate predictions. For this reason, the discussion in this work will focus primarily on non-linear activation functions, given their sophistication and dominance in modern deep learning applications.

- **Sgmoid** Sigmoid function [221] is widely used, especially in binary classification tasks. It maps any input to a value between 0 and 1, making the output interpretable as a probability., as shown in the figure 3.12.

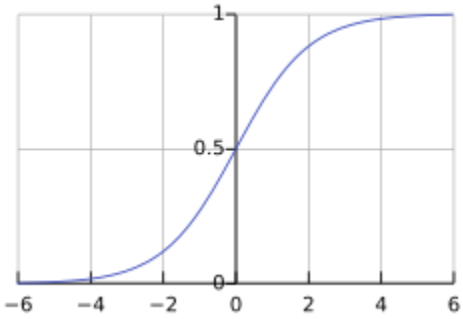


Figure 4.12: Sigmoid function.

The sigmoid function is distinguished by its smooth gradient, which ensures a continuous transition in output values without abrupt changes. Furthermore, for relatively large positive or negative input values, the function drives the output toward the asymptotic boundaries of the curve, thereby producing confident predictions. Its mathematical formulation is given as:

$$sig(x) = \frac{1}{1 + e^{-x}} \tag{4.2}$$

Despite its popularity, sigmoid functions have drawbacks such as the vanishing gradient problem, where gradients approach zero for large input values, slowing down learning. Moreover, its output is not zero-centered, which can affect the stability of gradient updates. Nonetheless, sigmoid remains important in output layers for tasks requiring probabilistic interpretation of outputs [222].

- **Softmax** The Softmax function, shown in figure 3.13, can be regarded as a generalization of the sigmoid function. While the sigmoid function is typically employed in binary classification tasks, Softmax is predominantly used for multiclass classification problems. Mathematically for an input vector $x=[x_1,x_2,\ldots,x_k]$, the Softmax function for each class i is expressed as follows:

$$sofmax(x_i) = \frac{e^{x_i}}{\sum_{k=1}^{K} e^{x_k}} \tag{4.3}$$

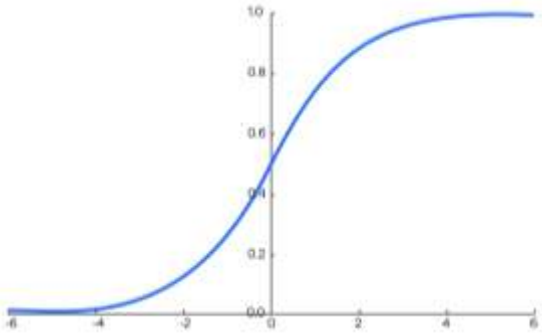


Figure 4.13: Softmax function.

- **Hyperbolic tangent** Usually commonly referred to as Tanh, as illustrated in figure 3.14, is another widely used non-linear activation function. Compared to the sigmoid function, Tanh has demonstrated greater effectiveness, as it maps input values to the range [-1,1], resulting in outputs that are zero-centered and often leading to faster convergence during training.

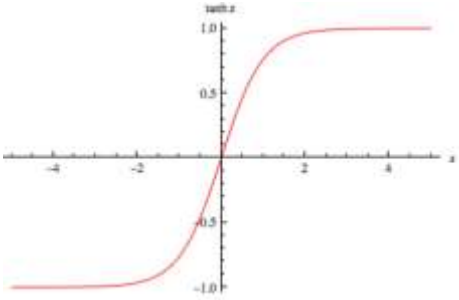


Figure 4.14: Hyperbolic tangent function.

The *tanh* formula is represented as follows:

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \qquad 4.4$$

- **ReLU (Rectified Linear Unit):** It is a non-linear activation function that has demonstrated remarkable effectiveness with the rise of deep neural networks over the past decade.

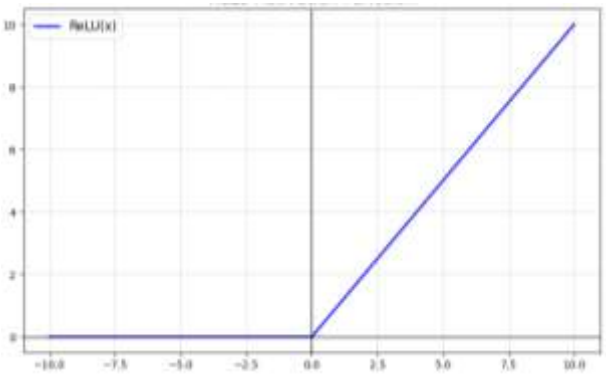


Figure 4.15: ReLU function.

ReLU [220] ReLU became popular because of its computational speed and its capability to activate only a subset of neurons at a time. This property enhances training efficiency and reduces computational overhead compared to the Sigmoid and Tanh functions. Mathematically, ReLU is defined as follows:

$$ReLU(x) = \max(0, x) \tag{4.5}$$

However, ReLU suffers from a limitation known as the *dying ReLU* problem. When input values are negative or approach zero, the function's gradient becomes zero, thereby halting weight updates during backpropagation and preventing the affected neurons from learning from data.

- **Leaky ReLU** The leaky ReLU [223] aims to improve the ReLU function by presenting a solution to the dying ReLU problem.

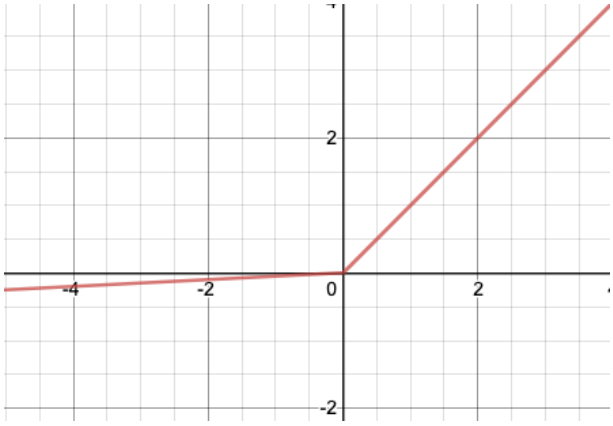


Figure 4.16: Leaky ReLU activation function representation.

Instead of resulting a 0 for negative input values of x, the function is defined as an extremely small linear component of x as expressed by the following formula:

$$LReLU(x) = \begin{cases} 0.01x, x < 0 \\ x, x \geq 0 \end{cases} \quad 4.6$$

- **Parametric ReLU** It is another variant of ReLU function [223] that aims to address the problem of gradient becoming zero when the input value is negative. It is the same as leaky ReLU, but instead of using a constant 0.01 we use a parameter a as shown in the following formula:

$$LReLU(x) = \begin{cases} ax, x < 0 \\ x, x \geq 0 \end{cases} \quad 4.7$$

- **Swish** is a newly discovered activation function by google research group. According to [223], it performs better than ReLU on deeper models with the same efficient computation.

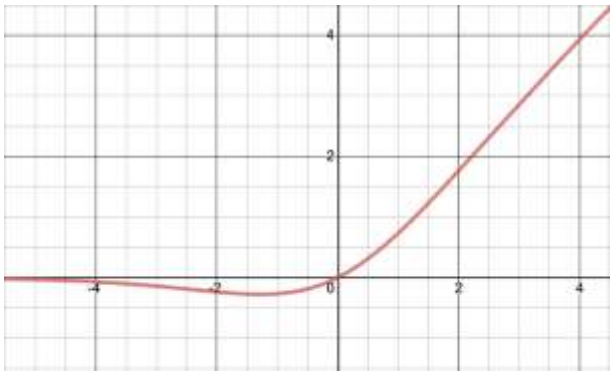


Figure 4.17: Swish activation function representation.

The output range of this function extends from negative infinity to positive infinity. It is mathematically expressed as follows:

$$\text{swish}(x) = \frac{x}{1 + e^{x}} \quad 4.8$$

4.3.2.3 Pooling layer

A pooling layer is a common component in CNNs designed to reduce the spatial dimensions (width and height) of feature maps while preserving important information. Pooling works by sliding a small filter over each channel of a feature map and summarizing the region covered

by the filter, often using operations like taking the maximum value or the average value. This process decreases the amount of computation and memory needed, helps control overfitting, and makes the network more robust to variations and distortions in input data.

Pooling also enables the network to build hierarchical feature representations, focusing on increasingly abstract features as the data moves through deeper network layers. Popular types of pooling include max pooling, average pooling, and global pooling, each with its own advantages for different computer vision tasks [224].

Pooling layers are typically applied after the activation non-linearity to the convolutional feature maps. Several variants of pooling exist, including max pooling, average pooling, and hybrid approaches such as mixed pooling, mixed max–average pooling, and gated pooling functions. Given the diversity of pooling strategies, the selection of a particular method for a given task has traditionally relied on empirical evaluation, as discussed in [225]. However, Boureau *et al.* [226] [40] provided theoretical insights offering guidance on selecting the most suitable pooling operation under specific conditions

Bellow, we present some of the popular pooling functions.

- **Max-pooling** A max-pooling operator [227] can be applied to down-sample the convolutional feature maps, thereby reducing their variability. This operation propagates the maximum value from each group of R activations to the next layer. Consequently, the mth max-pooled feature map is formed from a set of J related filters

$$p_m = [p_{1,m}, p_{2,m}, \dots, p_{j,m,} \dots \dots p_{J,m,}] \in R^J: \quad p_{j,m} = \max\left(h_{j,(m-1)N+r}\right) \tag{4.9}$$

where $N \in \{1, \dots, R\}$ is a pooling shift allowing for overlap between pooling regions when $N < R$. The pooling layer decreases the output dimensionality from K convolutional bands to $M = (K - R)/N + 1$ pooled bands and the resulting layer is $p = [p_1, \dots, p_M] \in R^{M.J}$

An example of the Max-Pooling operation is shown in Figure 3.18

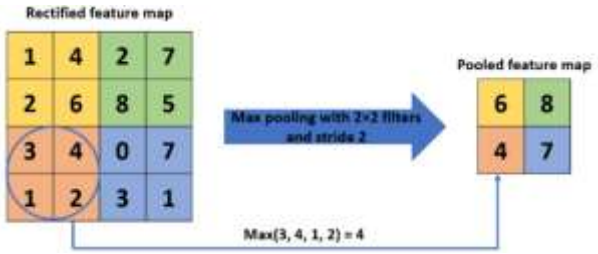


Figure 4.18: Max Pooling operation

- **Average-pooling:** The concept of using average, or mean, pooling for feature extraction was first introduced in [213], which represents the first convolution-based deep neural network. As illustrated in Figure 3.19, an average pooling layer performs down-sampling by partitioning the input feature map into rectangular regions and computing the mean value within each region.

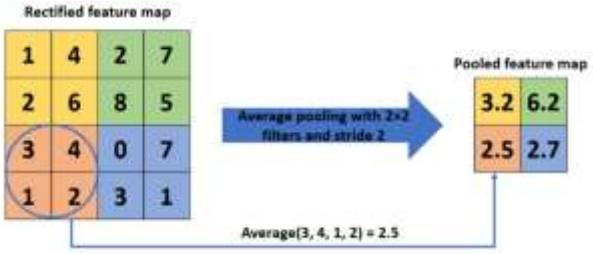


Figure 4.19: Average pooling operation

It can be mathematically represented by the following formula:

$$y_{ij}(x) = \frac{1}{|R_{ij}|} \sum_{(p,q)\in R_{ij}} x_{k_{ij}} \qquad 4.10$$

Where $x_{k_{ij}}$ represents the element at location (p,q) covered by the pooling region R_{ij}.

- **Mixed-pooling** Max pooling retains only the strongest activation within a region, while average pooling combines all activations by averaging them, which may reduce the impact of prominent features. To address this limitation, Yu et al. [228] introduced a hybrid pooling strategy that combines both average and max pooling operations. This method draws inspiration from regularization techniques such as Dropout [229] and DropConnect [230]. The mixed pooling operation can be mathematically expressed as shown in Equation 3.11.

$$s_j = \lambda \max_{i \epsilon R_j} a_i + (1 - \lambda) \frac{1}{|R_j|} \sum_{i \epsilon R_j} a_i \quad 4.11$$

The parameter λ determines whether the operation performs max pooling or average pooling. Its value is randomly assigned as either 0 or 1—where λ=0 corresponds to average pooling, and λ=1 corresponds to max pooling. The value of λ is stored during the forward propagation phase and subsequently utilized in the backpropagation process. Yu *et al.* demonstrated the superiority of this mixed pooling approach over both max and average pooling by conducting image classification experiments on three different datasets.

Mixed max-average-pooling This approach was proposed by Chen-Yu Lee et al. in [231]. It consists of proportionately blending the two pooling procedures, max and average, instead than only picking one operation to conduct. The mixed max-average pooling procedure may be described by the following formula:

$$y_{ij}(x) = \alpha \max_{(p,q) \epsilon R_{ij}} x_{k_{ij}} + (1 - \alpha) \frac{1}{|R_{ij}|} \sum_{(p,q) \epsilon R_{ij}} x_{k_{ij}} \quad 4.12$$

Where $x_{k_{ij}}$ represents the element at location (p, q) covered by the pooling region R_{ij} and α is a scalar representing the mixing proportion which specifies the exact amount of combination of max and average pooling, $\alpha \in [0 - 1]$. We can see that mixed-pooling is a generalization of the mixed max-average pooling method when $\alpha = 0$ or $\alpha = 1$.

4.3.2.4 Batch normalization

Deep Neural Networks (DNNs) are considered among the most advanced and powerful learning algorithms in modern machine learning [232]. Despite their remarkable success, the training process of such models remains highly complex and challenging, primarily due to their sensitivity to initial parameter settings [233]. To address these challenges and improve convergence toward optimal and robust solutions, several optimization algorithms have been proposed and widely adopted in practice [234], [235].

Stochastic Gradient Descent (SGD) [236] is one of these methods that has shown to be an excellent means of training DNNs. SGD has demonstrated significantly more successful when applied in mini-batches. However, as resilient as SGD is, it still requires careful setup of

parameter values for the network. This seems to be an issue for the training process due to the link between the inputs to each layer and the parameters of all previous levels. Therefore, when the network expands deep-wise, modest modifications to the network parameters cause the layers' inputs rapidly expand.

To address the aforementioned issue, one intuitive approach is to modify the distribution of inputs across network layers. However, while seemingly effective, this strategy introduces another challenge known as *covariate shift* [237], which typically arises when the training and testing data distributions differ and can be mitigated through domain adaptation techniques. Nonetheless, a similar phenomenon can also occur within a deep neural network itself, affecting subsets of the model such as specific sub-networks or layers. Applying the same corrective strategy across the entire model can lead to a substantial increase in the dimensionality of layer inputs, thereby slowing down convergence.

Sergey et al. [209] introduced as BN an effective technique to reduce such internal shifts, thereby accelerating the training process. BN achieves the same accuracy with up to 14 times fewer training steps and often surpasses previous benchmarks, demonstrating its effectiveness in both speeding up training and improving model performance. For computational efficiency, normalization is applied over mini-batches rather than across the entire dataset, as normalizing each layer's inputs individually is not practical. Consequently, two key simplifications were proposed in the BN method to optimize this normalization process. Firstly, instead of jointly normalizing the features in layer's inputs and outputs, each scalar feature is independently normalized as shown in Equation 3.13 For a layer of dimension d and an input $x = (x^1, \ldots, x^k)$ each dimension is normalized as follows:

$$Z^i = \frac{x^i Mean(x^i)}{\sqrt{Var(x^i)}} \qquad 4.13$$

Here, the mean and variance are computed over the training set. In practice, constraining each layer's activations to a fixed mean of 0 and variance of 1 may reduce the network's representational capacity. To overcome this limitation, BN introduces two learnable parameters, γ and β, which enable the network to adjust the normalized outputs. These parameters allow the model to rescale and shift the normalized activations, thereby restoring the network's ability to learn more expressive and flexible representations:

$$y^i = \gamma^i z^i + \beta^i \quad 4.14$$

The parameters used in Equation 3.14. are learned jointly with the original model parameters during training. By setting $\gamma^i = \sqrt{Var(x^i)}$ = and $\beta^i = Mean(xi)$ the original activations can be effectively restored if necessary.

The second simplification involves estimating the mean and variance of each activation using the samples within each mini-batch, as mini-batches are employed during stochastic gradient training. Consequently, the parameters used for normalization are fully integrated into the gradient backpropagation process, allowing them to be updated jointly with the other model parameters.

BN can be applied to any set of activations within a neural network. In the case of CNNs, it is typically applied to transformations consisting of a convolution operation followed by an element-wise non-linear activation function:

$$Z = f\,(Wu + b) \quad 4.15$$

Here, W and b denote the learnable parameters of the model, and $f(\cdot)$ represents the non-linear activation function. This transformation applies to both fully connected and convolutional layers. Batch Normalization (BN) is inserted just before the non-linearity, i.e., it normalizes the term x=Wu+bx=Wu+b. Although it is possible to normalize the layer inputs directly, u is typically the output of a preceding non-linearity whose distribution may vary significantly during training. Thus, constraining only the first and second moments of u would not effectively eliminate the covariate shift.

In contrast, the pre-activation term Wu+b tends to exhibit a more symmetric and less dispersed (approximately Gaussian) distribution, making it a better candidate for normalization. Consequently, normalization at this stage helps maintain stable activation distributions. The bias term b can be omitted, as its contribution is nullified by the subsequent mean subtraction during normalization. Therefore, the expression Z=f(Wu+b) can be reformulated as Z=f(BN(Wu)), where Batch Normalization is done individually to each dimension of x=Wu, with different pairings of learnable parameters γ_k and β_k for every dimension. In the case of convolution layers, the normalization needs to follow the convolution property, such that different parts of the same feature map are normalized in the same way at various places. To do

this, all activations are normalized in a mini-batch, at all locations. In this situation, to execute BN, Values of x in a mini-batch $B = x_{1...m}$, B is going to be a collection of all the values of a feature map on both the elements of a mini-batch and the spatial locations. Thus, for a mini-batch of size s with feature maps of size i × j, the size of the mini-batch $s' = s.i.j$ is utilized. This leads to learn a set of parameters γk and βk per feature map, instead of per activation

4.3.2.5 Dropout

Since their inception, Neural Networks have encountered several challenges that complicate the training process, one of the most prominent being *overfitting*. Numerous approaches have been proposed to mitigate this issue. On one hand, techniques such as the use of pooling layers and careful hyperparameter tuning have shown promising results in reducing overfitting. On the other hand, regularization methods like L1 and L2 regularization [191] help constrain model complexity by keeping the network's weights as small as possible. Moreover, Srivastava *et al.* [236] introduced a simple yet effective technique known as *Dropout*, shown in figure 3.20, which prevents overfitting by randomly deactivating a subset of neurons during training effectively creating an ensemble of different subnetworks that improve generalization.

For many machine learning methods, combining multiple models can enhance overall performance. However, this approach significantly increases computational costs due to the need to fine-tune numerous hyperparameters for each model. Furthermore, it demands large volumes of training data, which might not always be available to effectively train multiple distinct networks. Even if training several large networks were feasible, the inference process would be inefficient and impractical because of the need for rapid prediction times during testing. To address challenges in term of computation complexity, limited data availability, and inference speed, dropout was introduced as an efficient alternative that simulates model combination without the associated drawbacks [238].

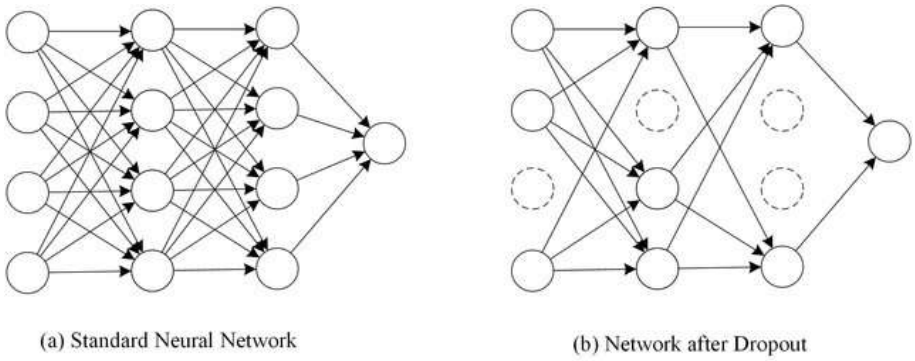


Figure 4.20: Neural network transition from basic to dropout.

The fundamental principle of dropout is to provide an efficient approximation of combining multiple neural network architectures, thereby addressing the overfitting problem effectively. Dropout operates by randomly deactivating selected units during the neural network training process. When a unit is dropped out, it is temporarily removed from the network along with all its incoming and outgoing connections.

The dropout mechanism introduces a key hyperparameter known as the "retaining probability," typically denoted as p*p*, which represents the probability that each unit will remain active during training. According to the seminal work by Srivastava et al., the value of p*p* can be determined using a validation set or simply set to 0.5, which appears to be near-optimal for many networks and tasks. However, selecting the appropriate value for p*p* is crucial for model performance [53].

Research indicates that for input layers, *p* should be closer to 1.0 rather than 0.5, while for hidden layers, optimal values typically range between 0.5 and 0.8, with 0.5 being the most commonly used default. The nature of the data also influences the optimal retaining probability. For instance, when working with image data, the optimal value of *p* for input layers is typically around 0.8, as demonstrated in empirical studies [52].

4.3.2.6 Fully connected layers

Fully connected layers (also known as dense layers) are neural network layers where every neuron in the current layer is connected to every neuron in the previous layer. This complete connectivity distinguishes them from other layer types like convolutional layers, which only connect to local regions of the input.

In a fully connected layer, each neuron computes a weighted sum of all inputs from the previous layer, adds a bias term, and applies an activation function. The mathematical operation can be expressed as:

$$z_j = \sum_i w_{ij} x_i + b_j \quad 4.16$$

Where w_{ij} represents the weight connecting neuron ii from the previous layer to neuron j, x_i is the input from neuron i, and b_j is the bias term.

The basic components of FC layers are as follows:

- ✓ **Input layer**: it contains the flattened feature vector extracted from previous layers.
- ✓ **Weights**: they represent the percentage of importance of a node in a layer, hence the final output prediction.
- ✓ **Hidden layers**: it's a type of layers that their inputs and outputs are supposed to be uncontrollable, they contain a number of nodes called neurons stacked on top of each other.
- ✓ **Output layer**: after that data is fed to the input and passed through hidden layers, the output layer decides which real value to output in case of regression or a set of probabilities in case of classification.

Fully connected layers are commonly used as the final layers in neural networks for classification or regression tasks, where they transform high-level features extracted by earlier layers into class probabilities or output predictions. However, they can be computationally expensive due to the large number of parameters required, making them susceptible to overfitting, especially with limited training data.

4.4 Transformers

In natural language processing (NLP), self-attention-based architectures, specifically, Transformers [241], have emerged as the preferred paradigm. The standard method is to fine-tune on a smaller, task-specific dataset after pretraining on a big corpus of text. Transformers' scalability and computational efficiency have made it feasible to train models of previously unheard-of sizes. Attempting to integrate CNNs with self-attention have been made recently by Chen et al. [242], and some models even completely do away with convolutions. Due to the usage of unique attention patterns, these models have not yet been successfully scaled on contemporary hardware accelerators, despite their promise for efficiency.

Comparing to sequential models, the Transformer design has quickly shown good performance and increased in popularity. The necessity to convey intra-relevancy inside a phrase while simultaneously having the ability to scale up to a bigger collection of tokens drove the development of the Self-Attention mechanism. Larger models can now be trained at scale thanks to Transformers' performance and scalability, and performance is still not showing any signs of saturation. However, applying self-attention to pictures is difficult since it requires that every pixel pay attention to every other pixel, which results in a cost that is quadratic to the number of pixels and is impractical for actual input sizes. Dosovitskiy et al. [243] addressed this by treating each picture as a series of patches and making only minor changes to the original Transformers for NLP, shown In figure 3.21. After that, each patch is regarded as a token, transformed into an embedding, and supplied into the Transformer. Certain methods may be trained end-to-end for supervised classification and scale up to 16x16 pixel pictures.

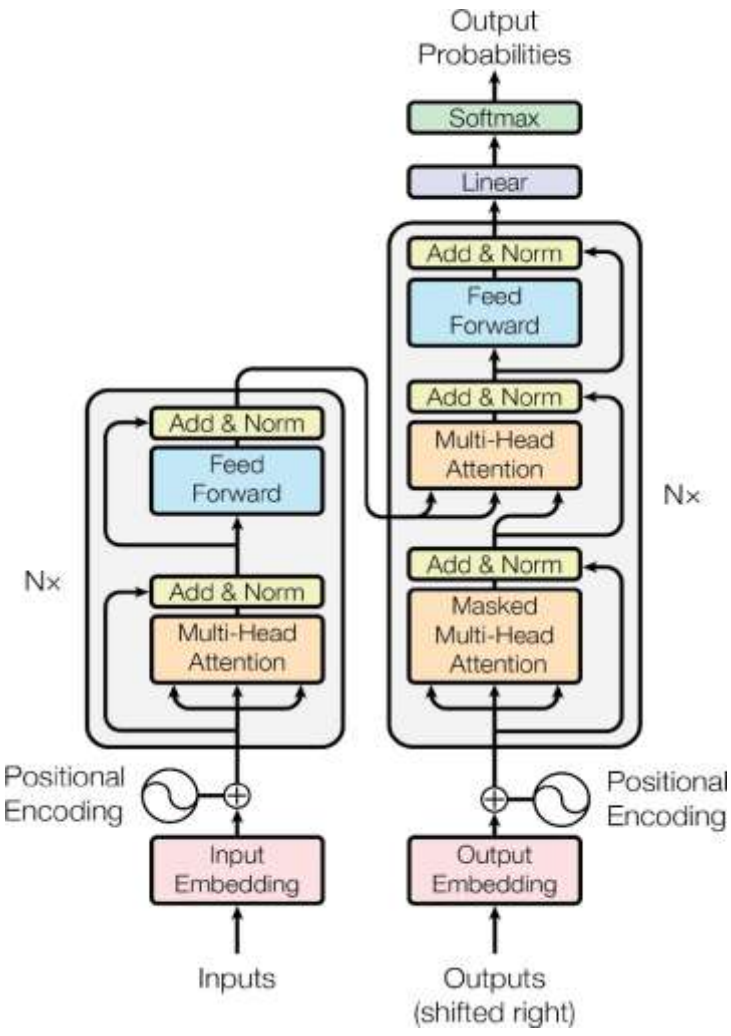


Figure 4.21: Transformer architecture [241]

4.4.1 Towards Transformer SeqtoSeq

In the simulator proposed by Shah et al. [244], Transformers aim to learn a representation of the input sequence that is independent of the sequence's elemental order. Self-attention, a method that enables a model to concentrate on various input sequence points in order to calculate a representation of that position, is used to do this.

Before exploring the Transformer components, we must first address the Sequence to sequence (SeqtoSeq) [245] design. Which consists of two multilayered LSTMs, one for encoding and one for decoding as shown in Figure 2.8. The input sequence is initially mapped to a hidden representation with a predetermined number of dimensions. The hidden representation is decoded into an output sequence in the second phase. The input and output are constructed in the same way as the transformer. Since machine translation is the primary development effort for both models, we provide it first. In machine translation, the input vector is basically a phrase or collection of words that are fed into the encoder to create the previously discussed hidden representation. Instead, the hidden representation and a start sentence token are the two inputs to the decoder. The decoder functions dynamically in this scenario; the start token is fed into the decoder, which produces a word as the output. The procedure is repeated until an end sentence token is created by sending the first output word back into the decoder.

The main challenge here is that all of the information is compressed into the hidden representation, which is obviously a SeqtoSeq method bottleneck.

To overcome the bottleneck, one suggestion to update this model was to include an extra attention mechanism [246], which we shall go into more depth about in the next section subsection 3.4.3. The fundamental concept is to provide the decoder direct access to all of the encoder's input words so that it may concentrate on a specific input sequence segment at each time step.

Attention enhanced greatly the performance and it helped with the gradient loss difficulties and the bottleneck. It also brought additional interpretability to the model, but it lacked still the scalability as RNN design remain sequential. Ultimately this challenge resulted to the invention of the Transformer [241] that employs a completely feed forward design.

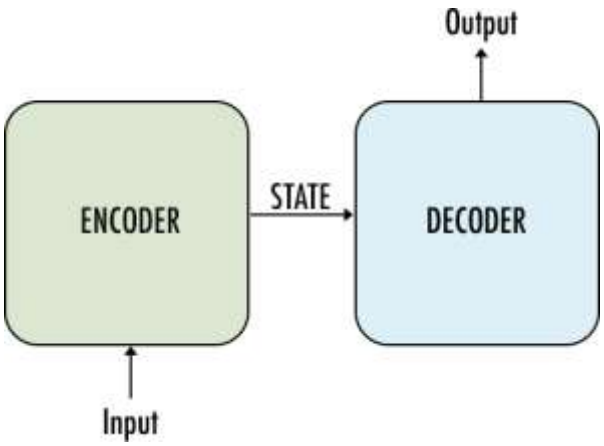


Figure 4.22: Encoder decoder architecture.

4.4.2 Embedding

In machine learning embedding is to encode a data to a vector representation of different dimension. It is used for numerous applications, for example while neural network cannot be fed with raw sequences of character, we may use an embedding to encode the information into a numeric vector and proceed with the deep learning analysis.

In the Transformer architecture several embeddings are employed. Word2Vec [238] is used to turn words into vectors, then a second embedding, termed positional embedding is utilized to provide the model a spacial information. As we shall see in the next section the Transformer's architecture is insensible to any operation of permutation on input. So to keep the sequential information of a sentence positional embedding has been devised.

The original positional embedding established by the creators of Transformer was fixed, meaning it doesn't change throughout training and is defined as following:

$$p_t^{(i)} = f(t)^{(i)} := \begin{cases} \sin(w_k . t), if\ i = 2k \\ \cos(w_k . t), if\ i = 2k+1 \end{cases} \quad 4.17$$

Where:

$$w_k = \frac{1}{10000^{\frac{2k}{d}}} \quad 4.18$$

Index *t* denotes the location in the sequence of token word, whereas i runs from 1 to *d*, the dimension of the embedding. If *i* is even we will use the sine function, whereas if it is odd we will use the cosine. ω is termed the frequency. The authors picked this specific combination because relative positioning may be described simply by a linear function.

$$\bar{p}_{t+\phi} = M_{\phi \bar{p}_t} \tag{4.19}$$

We can make a simple example with dimension 2, it can be proven that:

$$M_{\phi,q} = \begin{bmatrix} \cos(\omega_k . \phi) & \sin(\omega_k . \phi) \\ -\sin(\omega_k . \phi) & \cos(\omega_k . \phi) \end{bmatrix} \tag{4.20}$$

So equation 3.19 becomes:

$$\begin{bmatrix} \sin(\omega_k . (t+\phi)) \\ \cos(\omega_k . (t+\phi)) \end{bmatrix} = \begin{bmatrix} \cos(\omega_k . \phi) & \sin(\omega_k . \phi) \\ -\sin(\omega_k . \phi) & \cos(\omega_k . \phi) \end{bmatrix} \begin{bmatrix} \sin(\omega_k . t) \\ \cos(\omega_k . t) \end{bmatrix} \tag{4.21}$$

Learned positional embedding, on the other hand, are merely a fully connected layer. The model learns them during training using the current data-driven paradigm. As shown in Figure 3.21, both fixed and learned positional embedding are only added to the output of the input embedding.

4.4.3 Attention

A key component of the Transformer design is the attention mechanism. This system aims to strengthen one input characteristic over another. Q, K, and V are our three input matrices. These are referred to as values, keys, and queries, respectively. The concept is straightforward: we look for similar items in the key matrix for every query vector. For every key vector, there is only one value vector. Based on how closely our query resembles the item's key, we give each value a weight. When using formulae:

$$Att = f(Q,K)V, f(Q,k) = softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \quad 4.22$$

Where $\boldsymbol{V, K, and\ Q} \in \mathbb{R}^{\boldsymbol{L\times d}}$, d is the embedding dimension, and L is the length of the input sequence. The use softmax function to weight the values has been proposed by the authors, however it has led to some serious concerns of computing complexity. The $\sqrt{d}$ term is there simply for regularization purposes and doesn't impact the conclusion of the weighted sum.

We can rewrite the attention in another way as

$$Att_{\leftrightarrow}(Q,K,V) = D^{-1}AV\ (2.27) \quad 4.23$$

Where we define

$$A = \exp\left(\frac{QK^T}{\sqrt{d}}\right), \quad D = diag(A1_L) \quad 4.24$$

Where 1_L is basically a unitary vector of dimension L. Since information moves in both directions with respect to the input sequence, the aforementioned attention is sometimes referred to as bidirectional. In contrast to bidirectional attention, we might consider single-directional attention, often known as casual attention, in which information is derived only from prior tokens. Assume that the input vector at time t can only access keys from the past. This is used in the decoder during inference to prevent the transformer architecture from accessing future values [247].

To reach this goal we just multiply a triangular matrix tril of 0,1 to block future entry scores. Thus, we can write the attention matrix as

$$Att_{\leftrightarrow}(Q,K,V) = \widetilde{D}^{-1}\tilde{A}V \quad 4.25$$

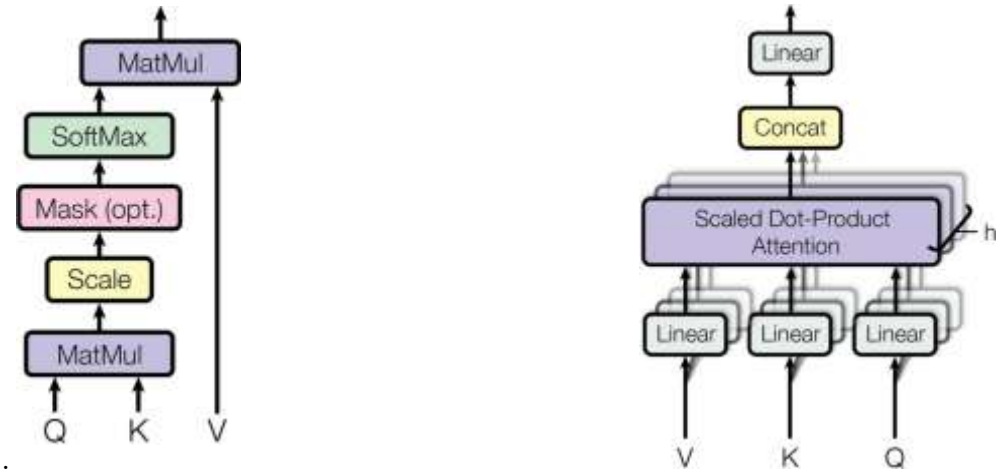
Where we define

$$\tilde{A} = tril(A)$$
$$\tilde{D} = diag(\tilde{A}1_L) \quad 4.26$$

4.4.4 Encoder

As shown in Figure 3.21 this is the whole design of the transformer. It offers an encoder-decoder structure exactly as SeqtoSeq [245].

The input matrix is received by the encoder as the total of the several embedding layers. The multi-head attention layer, which is shown in detail in Figure 2.9, receives the input after that. The authors choose to divide the computation of the attention matrix, decreasing the hidden dimensions h of each query, key, and value. This layer is just a multiple attention computed in parallel. Each brain may concentrate on different aspects of the input data because to this divide.



(a) (b)

Figure 4.23: Attention block of Transformer, scaled dot product attention (a) and multi head attention (b) [241]

Let $h = e/d$, with e and d indicating respectively the embedding dimension and the h hidden dimension of the attention, and $X \in \mathrm{R}^{L \times e}$. For each head i, with

$\{i \in \mathrm{N}, i < h\}$, the attention Att^i is computed as:

$$Att^i{}_{\leftrightarrow} = Att^i{}_{\leftrightarrow}(W_i^Q X, W_i^k X, W_i^V X) \quad 4.27$$

Where $W_i^Q X, W_i^k X, W_i^V X \in \mathbb{R}^{e \times d}$ are values, keys and queries weight matrices, representing dense layer. The output of all heads is concatenated as:

$$Out = W^C Concat(Att^1{}_{\leftrightarrow}, \ldots; Att^h{}_{\leftrightarrow}) \quad 4.28$$

where W^C is another thick layer's weight matrix. The previous equation and the Figure 3.21 both demonstrate how values, keys, and queries are generated as a linear combination of the input matrix. Because the score attention matrix is derived from the input matrix and only weights the input itself, this attention, which is used in Transformer [241], is also known as self-attention.

A Feed Forward layer comes after the multi-head attention layer. This simple block consists of a fully linked layer, a ReLU activation function, and another fully connected layer. Before the ReLU, the hidden dimension h in this block is expanded four times before being condensed once again to maintain its initial dimensions. The authors made this decision, while other researchers have chosen a different approach.

The output matrix moves on to a batch normalization layer and a residual layer after each of these two blocks. This completes the encoder design, which may be sequentially repeated several times. The encoder's output is then sent to the decoder for further use in the cross-attention.

4.4.5 Decoder

The decoder and the encoder share some of the same block architecture. We can see that there are two distinct attention blocks, each with minor variations from the encoder one. The first is known as causal attention since, as was indicated in the previous part of Eq. (2.30), it enforces causality on the decoder input using a triangular mask. The matrices Q, K, and V are functions of the decoder input in this case as well, and it is referred to as self-attention. Rather than being a self-attention block, the second attention block uses the encoder's hidden representation—that is, its output—to calculate keys and values. Since the idea is similar to SeqtoSeq with attention [245], [246], which provides the decoder with direct access to the encoder input, queries are still a function of the decoder input. The later focus is frequently referred to as cross attention because of this.

Similar to the encoder, a batch normalization layer and a residual layer follow each attention block or feed forward neural network. Additionally, the decoder structure can be repeated several times, and the output will immediately resolve the job that was chosen, without the need for a task-specific activation function or dimensional space shift.

4.4.6 Training and Inference

In order to repeat the process until the end, dynamic inference, which was introduced in section 1.3, means to output a single data point at each iteration and then enter the new point as the decoder input. If we consider it, we will use dynamic inference to propagate the error for successive timesteps if there is one at timestep t. This is particularly noticeable in a simple method that directly lowers the loss at each location; more sophisticated methods, like beam search, can be used to improve it to some extent [248].

While inference can only be done dynamically in the actual world when the model is deployed, we can use a different strategy known as teacher forcing during training.

The decoder input is forced to be the target sequence's ground truth by the teacher. There are significant advantages and disadvantages to this. Since any error will be contained in the output, we first solve the issue of the fault propagating into subsequent timesteps. However, because deployment conditions differ from training conditions, this adds bias into our network. In neural machine translation practice, instructor forcing is often used for transformers because it has the advantage of computing the output in a single sweep, eliminating the need for a dynamic process and significantly reducing training times

4.4.7 Time series forecasting with Transformers

The transformer was initially only employed for neural machine translation [241], but several variations have been suggested to expand its applicability to time series forecasting [249], [250], [251]. We will discuss this in more detail in the next chapter 3, but for now, keep in mind that the transformer doesn't scale well with multivariate time series, so we need to modify the design.

It goes without saying that we also need to change the encoding; NLP tasks use Word2Vec [252] encoding, which is inappropriate for time series forecasting. We frequently replace it with a Time2Vec that is more suitable [253]. Time2Vec is a size k + 1 embedding that accepts a scalar time value t as input. The authors designed it to be time rescaling invariant, meaning it

can function with timeseries with time step units of seconds and ones with hours. The definition of Time2Vec

$$v_t^{(i)} := \begin{cases} w_i t + \theta_i \; if \; i = 0 \\ \mathcal{F}(w_i t + \theta_i) if \; 1 \leq i \leq k \end{cases} \quad 4.29$$

While function F is a periodic activation function and ωi and θi are learnable patterns, it bears some resemblance to positional encoding as given in Eq. (2.21). Time2Vec's creators [253] demonstrated that this embedding can learn any periodic pattern using a variety of periodic functions, including sin, mod, and triangle, with sin doing better than the other two.

It should be noted that the first component, $v_t^{(0)}$, encodes the liner scaling of time, while the hyperparameter, k, specifies the number of periodic patterns to be learned.

4.5 Advances in Deep Learning for Medical Imaging Tasks

Medical images differ substantially from natural images in several aspects [254]. They often contain highly sensitive and clinically relevant information, which plays a critical role in diagnosis and treatment planning. However, this information can be easily degraded or lost during image acquisition or processing stages [255], posing a significant challenge for medical image analysis.

Depending on the anatomical structure being examined and the desired image quality, various imaging modalities can be employed for medical data acquisition. The most commonly used modalities in clinical practice include Computed Tomography (CT) [254], Positron Emission Tomography (PET) [256], and Magnetic Resonance Imaging (MRI) [58]. Each of these techniques captures organ information differently — for example, MRI uses magnetic fields and radio waves, whereas CT relies on X-rays to produce detailed cross-sectional images.

In the nascent stages of medical image analysis, practitioners predominantly employed sequential low-level pixel processing techniques. These methods included shape detection filters, region growing algorithms, and mathematical modeling approaches such as fitting geometric shapes (lines, circles, ellipses) to anatomical structures. These techniques formed the foundation of rule-based systems designed to perform specific interpretation tasks. Such expert systems, often categorized as "good old-fashioned artificial intelligence," operated similarly to

traditional rule-based image processing frameworks, relying on predefined logic and human expertise to interpret medical images [257].

The introduction of machine learning marked a paradigmatic shift in medical image analysis, with supervised learning techniques becoming the dominant approach. This era was characterized by the widespread adoption of methods including active shape models, atlas-based segmentation, handcrafted feature extraction, and statistical classifiers for tasks such as segmentation, detection, and diagnosis. While these approaches represented a significant advancement from human-designed rule-based systems toward computer-trained models requiring minimal manual intervention, they remained heavily dependent on carefully engineered features designed by domain experts [65], [66].

The emergence of deep learning has brought about a revolutionary transformation in medical image analysis by automating the feature extraction process entirely. Deep learning architectures integrate layers responsible for both feature extraction—capturing high-level abstractions from raw image data—and feature learning, leveraging hierarchical representations from previous layers to make inferences. This integration eliminates the need for manual feature engineering and represents a key factor behind the exceptional performance of modern medical imaging systems[260], [261], [262]

These technological advances collectively form the foundation of Computer-Aided Diagnosis (CAD) systems—sophisticated tools designed to assist clinicians and radiologists in making accurate diagnostic decisions. The effectiveness of CAD systems is typically evaluated based on three critical criteria: diagnostic accuracy, computational efficiency, and level of automation [263]

For instance, a typical CAD system designed for breast cancer diagnosis follows four major stages: (1) segmentation of the breast region to isolate the area of interest, (2) detection of potential lesions or masses within the breast tissue, (3) segmentation of the detected abnormalities, and (4) classification of the segmented regions into benign or malignant categories.

4.5.1 Detection

In image processing, object detection is one of the fundamental tasks. It involves identifying instances of semantic objects belonging to a specific class within an image. In general terms,

object detection may be considered an extension of image classification, as it determines whether or not an image contains a particular object. However, it is important to distinguish between object classification, object localization, and object detection.

- In **object classification**, a class label is assigned to the entire image.
- In **object localization**, the model identifies and encloses the object within a bounding box.
- **Object detection** combines both processes—it draws bounding boxes around objects and assigns them corresponding class labels.

The performance of an object detection model is typically assessed using precision and recall, which evaluate how accurately the model identifies known objects across the best-matching bounding boxes.

Having clarified these concepts, the following paragraphs present recent state-of-the-art deep learning architectures used for object detection and their applications in the medical imaging domain, where detection generally refers to identifying lesions or organs in medical images.

Several deep learning-based object detection frameworks have gained popularity, including R-CNN, Fast R-CNN, Faster R-CNN, and the YOLO family of models.

There are numerous object detection architectures in deep learning that are very popular, such as R-CNN, fast R-CNN, faster R-CNN, YOLO in all its versions etc.

In [264], R-CNN was applied to prostate cancer detection based on Gleason grading using histopathological images. Wenyuan Li et al. demonstrated that by adopting a multitask R-CNN model, they were able to capture complementary contextual information, leading to better performance than single-task approaches. Their model achieved state-of-the-art results in epithelial cell detection, with an accuracy of 99.07%.

Girshick et al. [265] introduced Fast R-CNN for object detection. This variant unified the three separate components of R-CNN into a single, more efficient model, significantly improving detection speed. Building upon this, Shaoqing Ren et al. [266] introduced Faster R-CNN in 2016, a framework that significantly improved both training efficiency and detection accuracy. Unlike previous approaches that depended on the computationally intensive selective search algorithm to generate candidate regions, Faster R-CNN integrates a Region Proposal

Network (RPN) that learns to produce region proposals directly from shared convolutional features. This end-to-end design eliminates the bottleneck of external proposal generation, enabling the model to jointly optimize feature extraction, region proposal, and object classification within a unified architecture.

Subsequent YOLO variants have progressively improved both accuracy and efficiency. YOLOv2 [267] introduced batch normalization, higher-resolution inputs, and k-means clustering for bounding box initialization. Further architectural refinements were presented in YOLOv3 [268] , which increased network depth and enhanced performance without sacrificing speed. Continuous improvements have led to even more advanced versions such as YOLOv5, offering superior accuracy and real-time detection capabilities.

YOLO architectures have also been successfully applied in the medical imaging domain. For example, YOLOv3 achieved promising results in kidney detection from CT scans, with Dice scores of 0.851 in 2D and 0.742 in 3D [269] . Other applications include breast mass detection and classification [270], lung nodule detection in CT scans [271], and cholelithiasis and gallstone detection in abdominal CT images [272].

4.5.2 Segmentation

As it entails dividing an image into meaningful and clinically relevant sections, image segmentation is a crucial step in the medical image analysis pipeline. This is an intrinsically difficult endeavor that calls for models that retain computational economy while achieving great accuracy.

Medical image segmentation [273] plays a crucial role in isolating and analyzing specific regions of interest (ROIs), such as lung tissues, the spleen, or the brain. Numerous medical applications rely on segmentation, including brain tumor boundary extraction in MRI slices, cancer detection in lung CT scans, and identification of affected regions in chest X-rays. Given the shortage of domain experts, a growing number of well-designed algorithms have been proposed in the literature to enhance the accuracy and speed of diagnosis [274].

Several studies have successfully implemented deep learning-based segmentation architectures in the medical domain. For example, in [275], the authors proposed a novel two-pathway CNN architecture for brain tumor semantic segmentation. This model was designed to capture both local features that describe fine anatomical details and global contextual features

that provide broader structural understanding. Moreover, they introduced a two-phase training strategy that effectively addresses label imbalance issues during training.

In another study, Baumgartner et al. [276] developed a 3D CNN architecture for cardiac MRI segmentation, focusing on the delineation of the left and right ventricular cavities as well as the myocardium. Similarly, Wang et al. [277] performed pneumothorax segmentation in X-ray images using a CNN framework enhanced with spatial and channel Squeeze-and-Excitation mechanisms to improve feature representation.

Additionally, the authors of [278] designed a CNN-based architecture for brain tissue segmentation in MRI images, demonstrating robust performance across various tissue types. Building upon the Fully Convolutional Network (FCN) framework, Zhang et al. [273] proposed an FCN-based model for liver segmentation in CT scans, while Christ et al. [279] introduced a two-stage cascaded FCN architecture for liver segmentation. In their approach, the initial FCN performed a coarse segmentation, and the final result was refined using a dense 3D Conditional Random Field (CRF) to enhance boundary precision

4.5.3 Classification

Classification represents a crucial task for achieving accurate and reliable diagnosis. The performance of classification models is highly dependent on the quality of preceding stages in the image analysis pipeline such as preprocessing, segmentation, and feature extraction as well as on the robustness of the employed classification algorithm.

Image classification is a supervised learning process in which a model is trained to recognize specific target labels from a given set of annotated examples and subsequently predict these labels for unseen input images.

In the field of deep learning, transfer learning has emerged as a widely used strategy to improve classification accuracy, especially when working with limited medical datasets. There are two common transfer learning approaches: (1) using a pre-trained network as a fixed feature extractor, and (2) fine-tuning a pre-trained model on new medical data. Antony et al. [83] employed transfer learning to fine-tune a CNN on medical datasets and achieved 57.6% accuracy in knee osteoarthritis grade assessment, outperforming the feature extraction approach, which achieved 53.4%. Conversely, Kim et al. [73] reported the opposite trend in cytopathology image classification, where the feature extraction strategy yielded 70.5% accuracy, slightly surpassing fine-tuning (69.1%). Although results across studies are

sometimes inconsistent, several works [84], [85] have shown that fine-tuning pre-trained models generally provides superior performance compared to using them solely as feature extractors. For instance, fine-tuning a pre-trained Inception v3 model on medical datasets has achieved performance levels approaching that of human experts.

Deep learning architectures have also been successfully tailored for domain-specific medical classification tasks. Shen et al. [86] proposed a multi-scale CNN for lung nodule classification, where three independent CNNs processed nodules at different scales, and their outputs were combined to form a comprehensive feature representation. A similar multi-scale approach was adopted by Kawahara and Hamarneh [276] for skin lesion classification, demonstrating improved robustness to variations in lesion size and texture.

While most computer vision models are designed for 2D natural images, medical imaging often involves 3D volumetric data. Several studies have thus explored the integration of 3D information into deep networks. Nie et al. [88] developed a 3D CNN for classifying high-grade gliomas using MRI volumes, while Setio et al. [89] implemented a 3D multi-stream CNN to distinguish between nodule and non-nodule regions in chest CT scans, showing the potential of three-dimensional feature learning in medical diagnosis.

4.6 U-net based brain tumor segmentation

U-net is developed by Olaf Ronneberger et al. in 2015 [280], it's an improved fully convolutional neural network (FCN) architecture [281] which has an encoder path in which spatial information is reduced whereas feature information increases and decoder path in which image size in increased back to its original size with feature channel decreasing and also pooling operations and pooling operations are changed to transposed convolutions, which transform the shrunken feature maps to larger ones.

U-Net is a form of CNN that has been more popular in the last few years due its impressive in image segmentation step. Several series of convolutional and max-pooling layers compose the design, followed by upsampling and concatenation layers. This architecture allows the model to capture contextual information at different levels, which makes it especially good for tasks like brain tumor segmentation.

Skip connections are an important part of the U-Net design. They let the network save spatial information from previous levels and send it to later layers. This is especially helpful for segmentation tasks, where keeping the spatial connections between pixels is quite important.

Another essential part of the U-Net architecture is the use of upsampling layers, which let the network make the feature maps more detailed and provide a high-resolution output. Traditional CNNs, on the other hand, generally generate low-resolution outputs because they employ max-pooling layers.

4.6.1 U-Net Architecture

4.6.2 Convolutional Block

As shown in figure 3.24 The U-Net architecture consists of up of a succession of convolutional blocks, each of which has two convolutional layers and a max-pooling layer. The first layer has 64 filters while the second layer has 128 filters. Each convolutional layer has a kernel size of 3x3 and a stride of 1. The max-pooling layer has a kernel size of 2x2 and a stride of 2, which cuts the feature maps' spatial dimensions in half.

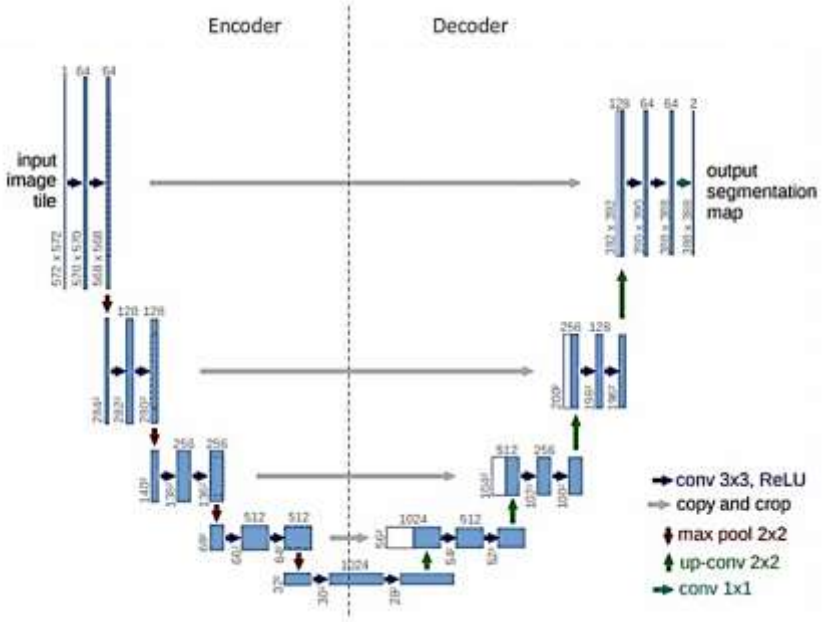


Figure 4.24: The architecture of U-net network

4.6.3 Encoder-Decoder Structure

The U-Net architecture has an encoder-decoder structure, with the encoder consisting of four convolutional blocks and the decoder consisting of four upsampling layers. The encoder captures the contextual information at multiple scales, while the decoder produces a high-resolution output by upsampling the feature maps.

- **Encoder path**

The encoder path, often referred to as the contracting path, is where the network must apply consecutive convolutions to create the feature map [282]. A ReLU activation function comes after each 3x3 convolution [283]. The ReLU's output, known as the skip connection, is in charge of giving the relevant decoder block more information or location data in order to enhance the final segmentation's quality and provide better results [284]. The feature map dimensions are cut in half by applying downsampling using 2x2 max-pooling in addition to convolution [283]. This method often results in fewer trainable parameters [284].

- **Decoder path**

Also known as the spreading path The segmentation mask will be obtained in this section by passing the encoder's output [283]. The resulting skip connection feature map will be concatenated with each layer of this block after a 2x2 transposition [285]. The output of the final layer is then subjected to a 1x1 convolution in addition to the sigmoid activation function, which yields the segmentation mask of the pixel-wise classification [284].

- **Bridge**

This component is in responsible of connecting the encoder and decoder components [284]. It is composed of two 3x3 convolutions followed by a ReLU activation function [283].

- **Skip Connections**

The U-Net design relies heavily on the skip connection to transfer local information from the encoder portion to the layers in the decoder section without losing it. It implies that it helps the decoder component to provide better outcomes [285]. Additionally, skip connections enhance network efficiency to get better result representation and facilitate rapid convergence[283].

The model shown in figure 3.24 was proposed for brain tumor segmentation [286] , the suggested method for segmenting and classifying brain tumors using figshare dataset, which contains 3064 patient images. As indicated in Figure 3.25, images with an axial view were used in this research with three types of brain tumors: meningioma, glioma, and pituitary tumor the well-known U-Net architecture, which has been quite successful in analyzing medical images. U-Net has a symmetric encoder-decoder structure. The encoder (contracting path) is made up of a series of convolutional and max pooling layers that gradually lower the spatial dimensions of the feature maps while increasing the depth of the representation. This step lets the network understand the semantic meaning of the picture, or the "what," but it loses the ability to precisely locate things in space. The decoder (expanding route) does the opposite by employing up-

sampling (transposed convolutions) to slowly increase the picture resolution and decrease the depth to get the "where" information back. Importantly, skip connections are added between the encoder and decoder's respective layers. This makes sure that fine-grained characteristics that were lost during down-sampling are restored so that exact segmentation borders may be drawn. After each concatenation step, new convolutional layers let the network combine local and contextual signals, which makes the segmentation outputs more accurate and stronger. This end-to-end design effectively captures both context and detail, rendering it exceptionally ideal for identifying intricate brain tumor structures and differentiating between tumor and healthy tissue in medical imaging applications.

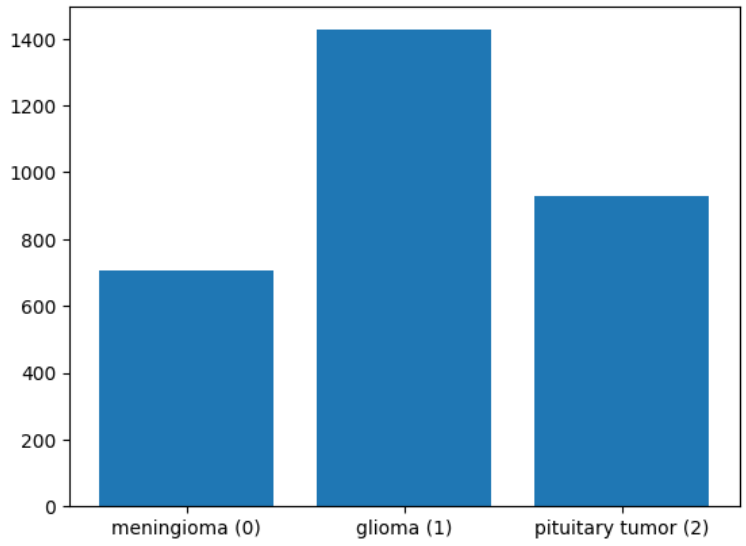


Figure 4.25: Summary of the Brain Tumor Figshare Dataset

The segmentation results presented demonstrate the practical application of U-Net architecture for automated brain tumor detection and delineation in MRI scans. As shown in figure 3.27, the three representative samples demonstrate the practical effectiveness of U-Net architecture for precise brain tumor segmentation across diverse tumor presentations and anatomical contexts. This dataset, widely utilized in medical imaging research, provides ground truth tumor masks enabling rigorous evaluation of segmentation model performance.

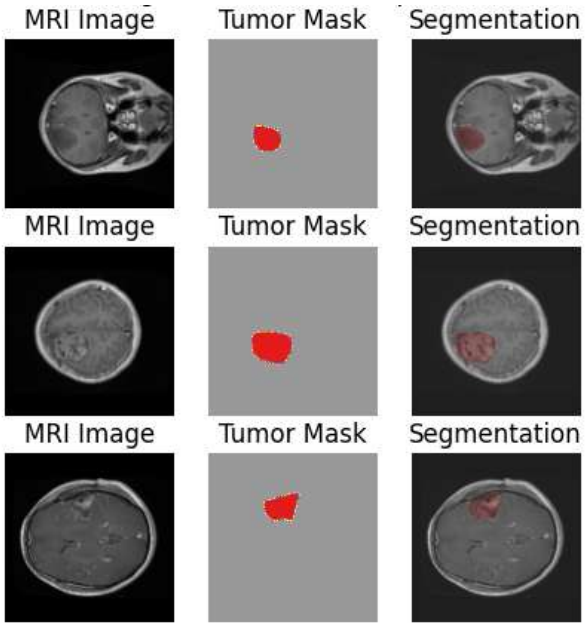


Figure 4.26: MRI Images with Original and Predicted Tumor Masks

The first case presents a small, well-circumscribed tumor located in the right hemisphere. The segmentation result closely replicates the tumor mask's spatial location and morphology. This case tests the model's ability to detect small lesions, particularly challenging for deep learning models due to limited pixel representation relative to background tissue. The successful segmentation of small tumors reflects U-Net's skip connections preserving fine-grained spatial details through direct concatenation of high-resolution encoder features.

The second sample shows a larger, slightly irregular tumor in the brain parenchyma. The predicted segmentation exhibits strong spatial correspondence with the ground truth mask, including accurate boundary delineation around the tumor periphery. This morphology tests the model's capacity to learn irregular shapes, U-Net successfully captures non-circular tumor boundaries through its multi-scale feature extraction, where early convolutional layers detect edge irregularities while deeper layers integrate contextual information.

The third case demonstrates a tumor with complex boundaries adjacent to the brain ventricles. The segmentation result shows high fidelity to the ground truth mask, accurately delineating tumor-CSF (cerebrospinal fluid) interfaces. This challenging case highlights U-Net's effectiveness in distinguishing tumors from anatomically adjacent structures with different tissue properties, facilitated by attention mechanisms in modern U-Net variants that learn to focus on diagnostically relevant features.

The training curves presented in Figure 3.27 demonstrate the U-Net model's convergence behavior and segmentation performance during the 80-epoch training procedure on the brain tumor dataset. Loss curves (left panel) reveal rapid convergence with the training loss (blue line) declining sharply from approximately 0.95 at epoch 0 to <0.1 by epoch 20, then continuing to decrease gradually toward near-zero values by epoch 80, indicating effective model optimization and feature learning. The validation loss (orange line) exhibits a similar rapid initial descent from ~0.95 to approximately 0.15 by epoch 20, then plateaus around 0.10-0.15 for the remaining epochs, suggesting stable model performance on held-out validation data. The modest divergence between training and validation loss curves indicates well-controlled overfitting, a critical requirement in medical image segmentation where models must generalize across diverse patient anatomies and imaging protocols. This moderate training-validation gap reflects appropriate regularization through techniques such as dropout, batch normalization, and combined loss functions (likely binary cross-entropy and Dice loss) that prevent excessive memorization of training patterns while maintaining discriminative feature learning.

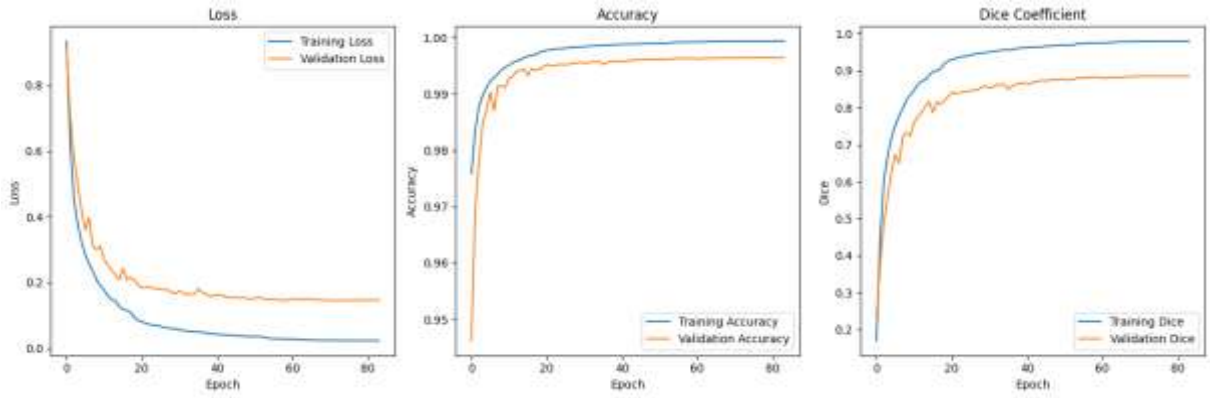


Figure 4.27: Training Curves for proposed U-Net Brain Tumor Segmentation

4.7 Conclusion

In conclusion, image classification plays a crucial role in the medical image analysis pipeline by converting the characteristics that have been retrieved by previous processing stages into diagnostic conclusions that are clinically significant. The accuracy and resilience of classification systems across a variety of medical imaging modalities have been greatly improved by the development of deep learning, especially via convolutional neural networks and transfer learning. Current research continues to show impressive progress in resolving these

shortcomings, despite obstacles including high inter-patient variability, data paucity, and the need for model interpretability. The potential of deep neural networks to attain near-human performance is further supported by the incorporation of domain-specific fine-tuning techniques and three-dimensional feature learning. All things considered, categorization is a vital component of computer-aided diagnosis, connecting automated picture analysis with practical clinical use.

5 HNet Optimization for Breast Cancer Classification

5.1 Introduction

Many existing deep learning models applied to medical image analysis are purely data-driven and often suffer from limited interpretability, which can hinder clinical adoption. Conversely, traditional diagnostic approaches rely heavily on domain knowledge and expert-crafted features, offering interpretability but often lacking the flexibility and scalability needed to handle the complex variability of histopathological images. While conventional CNNs excel at capturing local texture patterns, they struggle to model global contextual information and spatial hierarchies inherent in breast tissue architecture. This limitation motivates the development of learned feature representations that adapt dynamically to the diverse morphological characteristics present in the data, enabling more accurate and robust classification.

In this chapter, we propose HNet, a novel hybrid deep learning architecture that integrates EfficientNet for local feature extraction, Advanced Vision Transformers (AVT) for modeling global dependencies, and Capsule Networks for preserving spatial hierarchies. By leveraging the complementary strengths of convolutional, transformer-based, and capsule-based learning paradigms, HNet aims to enhance both classification accuracy and interpretability in breast cancer histopathological image analysis. The proposed framework is evaluated on multi-magnification images from the BreakHis dataset, addressing key challenges such as tissue heterogeneity and class imbalance.

To further strengthen the discriminative capability of the fused deep representations, we subsequently extend HNet by incorporating a GA–based feature optimization stage. The GA operates on the concatenated feature vectors extracted by EfficientNet and AVT, reducing redundancy and selecting the most informative features prior to capsule-based reasoning. This evolutionary optimization enhances robustness and generalization while maintaining computational efficiency. This chapter details the design, implementation, and training methodology of both the baseline HNet architecture and its GA-enhanced variant, demonstrating their effectiveness in real-world diagnostic scenarios.

5.2 Overview of the HNet Architecture

The proposed HNet framework is designed to enhance the classification of histopathological breast cancer images by integrating the strengths of multiple deep learning paradigms. As illustrated in Figure 1, the architecture is organized into three primary stages. First, input images undergo preprocessing and data augmentation to improve robustness and generalization. Next, the images are processed in parallel by two branches: EfficientNet, which efficiently extracts local and fine-grained visual features, and the AVT, which captures global contextual information via self-attention mechanisms. The feature maps from both branches are then fused and passed to a Capsule Network module, responsible for preserving spatial hierarchies and modeling part-whole relationships within the tissue. Finally, the capsule outputs are flattened and fed to fully connected layers for the final classification. This staged and hybrid design

allows HNet to effectively leverage complementary information from local texture, global context, and spatial structure, providing a powerful and interpretable framework for breast cancer diagnosis.

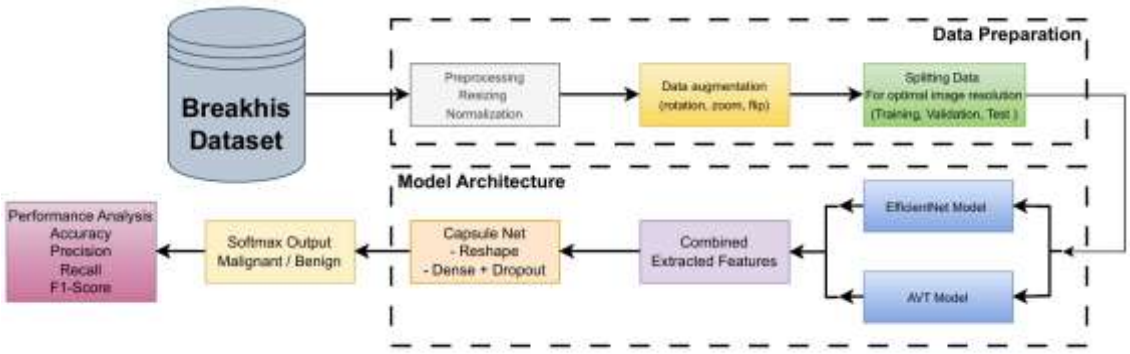


Figure 5.1 : Proposed HNet flowchart

5.2.1 EfficientNet Module

The EfficientNet-B0 architecture represents the baseline model in the EfficientNet family, implementing a systematic compound scaling approach that balances network depth, width, and resolution for optimal computational efficiency. The flowchart illustrates the key architectural components that transform input histopathological images through a series of specialized layers designed for feature extraction and classification [287].

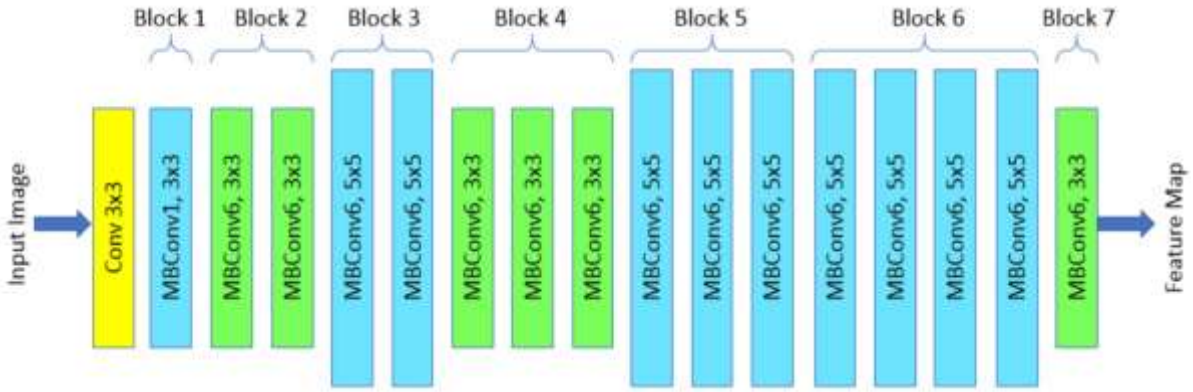


Figure 5.2 : Architecture of EfficientNet network

Input Layer Configuration

The architecture begins with an Input Layer that accepts histopathological images with dimensions (None, 128, 128, 3), where the batch size remains flexible (None), and each image maintains 128×128 pixel resolution with three RGB channels. This input specification is optimized for medical imaging applications where consistent image dimensions facilitate reproducible feature extraction while maintaining sufficient resolution for cellular-level pattern recognition [288]

Feature Extraction through EfficientNet-B0

The core feature extraction is performed by the EfficientNetV2-B0 layer, a refined version of the original EfficientNet-B0 that incorporates Mobile Inverted Bottleneck (MBConv) blocks as fundamental building components. This layer transforms the input from (None, 128, 128, 3) to (None, 4, 4, 1280), demonstrating significant spatial dimension reduction while expanding the channel depth for rich feature representation.

The EfficientNetV2-B0 implementation utilizes compound scaling methodology with optimized coefficients: depth coefficient $\alpha = 1.2$, width coefficient $\beta = 1.1$, and resolution coefficient $\gamma = 1.15$. This scaling approach ensures balanced resource utilization across network dimensions, achieving superior accuracy with computational efficiency compared to traditional CNN architectures [263]

MBConv blocks within the architecture employ depthwise separable convolutions combined with inverted residual connections, reducing computational complexity while preserving representational capacity. Each MBConv block incorporates Squeeze-and-Excitation (SE) optimization that enables the model to focus on essential features through channel-wise attention mechanisms.

Global Average Pooling Layer

The Global Average Pooling (GAP) layer serves as a critical dimensionality reduction component, transforming the feature maps from (None, 4, 4, 1280) to (None, 1280). This operation computes the average value across spatial dimensions for each feature channel, effectively reducing the 4×4 spatial feature maps to single representative values per channel.

GAP provides several advantages over traditional fully connected layers: parameter reduction (eliminating spatial-to-vector conversion weights), overfitting mitigation (through inherent regularization), and translation invariance (maintaining robustness to input variations). For medical image analysis, GAP ensures that diagnostic features extracted from different spatial locations contribute equally to the final classification decision.

Batch Normalization

The Batch Normalization layer maintains the dimensionality at (None, 1280) while normalizing the feature distribution across the batch dimension. This layer addresses internal covariate shift by normalizing inputs to have zero mean and unit variance, accelerating training convergence and improving model stability.

In medical imaging applications, batch normalization is particularly valuable for handling variations in histopathological image characteristics across different specimens, staining protocols, and acquisition

conditions. The normalization ensures consistent feature scaling that enhances the discriminative power of subsequent classification layers.

Final Dense Layer

The final Dense layer implements the classification function, transforming the 1280-dimensional feature vector to the target class space with (None, 256) output dimensions. This fully connected layer employs ReLU activation to introduce non-linearity while maintaining computational efficiency.

The 256-dimensional output suggests either an intermediate classification layer in a more complex architecture or a multi-class classification scenario with extensive class granularity. For breast cancer histopathology, this could represent fine-grained subtype classification or feature encoding for subsequent processing stages.

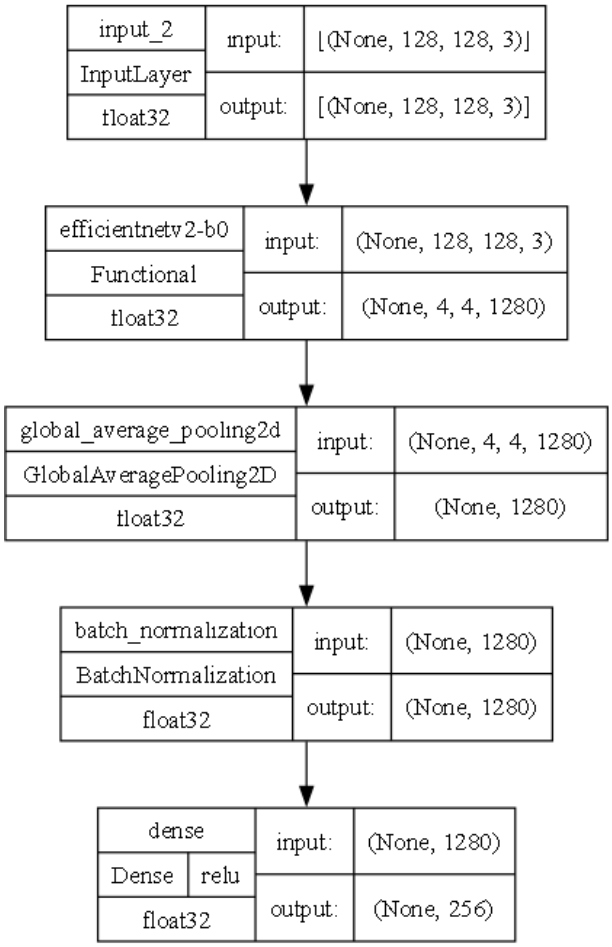


Figure 5.3 : Proposed EfficienNet diagram

5.2.2 Advanced Vision Transformer (AVT) for Global Context

To address the critical challenge of capturing both fine-grained cellular details and broader architectural contexts in histopathological images from the BreakHis dataset, we propose a novel hybrid deep learning architecture termed the Advanced Vision Transformer (AVT). This model synergistically integrates the representational power of Inception modules, the training stability of residual connections, and the global contextual awareness of a self-attention mechanism, specifically tailored for the high-resolution and complex nature of medical imagery.

The input to our network is a pre-processed image patch from the BreakHis dataset, typically of dimensions (height, width, 3), representing the RGB channels. Our architecture is designed to hierarchically learn features from this input, progressing from local texture patterns to global tissue structures.

5.2.2.1 Core Building Blocks

The foundation of the AVT is constructed using two custom-designed blocks:

- **The Enhanced Residual Block (**residual_block**)**: This block mitigates the vanishing gradient problem, which is crucial for training very deep networks on complex data. It consists of three convolutional layers, each followed by Batch Normalization and a Swish activation function. A skip connection adds the original input (or a linearly transformed version of it via a 1x1 convolution if dimensionalities do not match) to the output of the final batch normalization layer. This addition is subsequently passed through the Swish activation. The use of L2 kernel regularization on all convolutional layers further prevents overfitting, a common concern with the limited dataset sizes often encountered in medical domains.

- **The Inception-inspired Block (**inception_block**)**: To efficiently capture multi-scale information within a single image patch, a vital capability for identifying nuclei of varying sizes and glandular structures, we employ a modified Inception module. This block processes the input through four parallel pathways:

1. A 1x1 residual branch followed by max-pooling.
2. A cascade of two 1x1 residual blocks.
3. A cascade of two 5x5 residual blocks, designed to capture a larger receptive field.
4. A max-pooling operation followed by a 1x1 residual block and a dropout layer for regularization.

The outputs of these parallel pathways are concatenated, allowing the network to learn and combine features at different scales simultaneously.

5.2.2.2 Multi-Series Attention on Patches (multi_series_attention_on_patches)

A key innovation of our model is the incorporation of a self-attention mechanism applied directly to image patches. This component allows the model to dynamically weigh the importance of different regions within the feature maps, effectively learning long-range dependencies and contextual relationships that are spatially separated, a task convolutional layers struggle with alone.

The process begins by using the TensorFlow extract_patches operation to divide the feature maps into non-overlapping patches of a specified size (e.g., 3x3). These patches are then projected into a lower-dimensional space via a 1x1 convolution. The core attention mechanism is implemented as a multi-head attention layer. For each head, separate convolutional layers generate Query (Q), Key (K), and Value (V) matrices from the patched features. The self-attention scores are computed as the softmax of the scaled dot-product between Q and K, which are then used to create a weighted sum of the Values. The outputs from all attention heads are concatenated and projected back to the original channel dimension. This "Multi-Series Attention" mechanism enables the model to focus on multiple, distinct informative regions concurrently.

5.2.2.3 Overall Architectural Integration

The complete AVT model (advanced_vision_transform) integrates these components into a coherent pipeline:

1. The input tensor is first processed by an Inception block to extract rich, multi-scale features.
2. The features are normalized via Batch Normalization to stabilize training.
3. The normalized features are passed through the Multi-Series Attention on Patches module, which recalibrates them based on global context.
4. The attentive features are then downsampled by a max-pooling layer to reduce spatial dimensionality and increase translational invariance.
5. A dropout layer is applied to further enhance generalization.
6. Finally, a Global Average Pooling layer condenses the feature maps into a compact 1D feature vector, which serves as a powerful descriptor for the input BreakHis image patch. This vector is the output of the model and can be fed into a final classification layer (not shown here) for benign/malignant discrimination.

This hybrid design is philosophically grounded in the understanding that while convolutional operations are excellent local feature extractors, the global reasoning capability provided by self-attention is indispensable for accurate histopathological image analysis. The AVT model is thus posited to be particularly well-suited for the nuanced classification task presented by the BreakHis dataset.

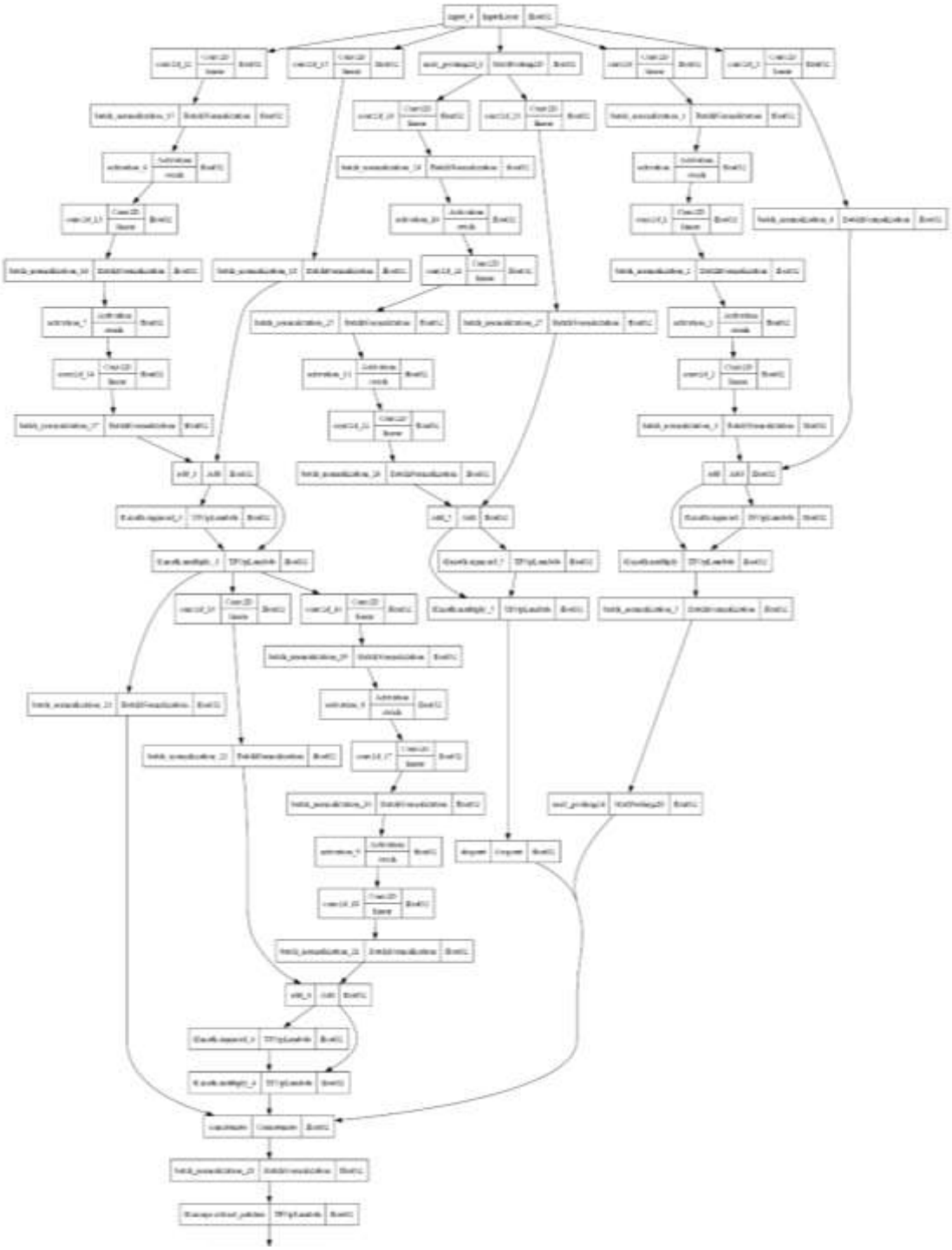


Figure 5.4: Proposed AVT Diagram

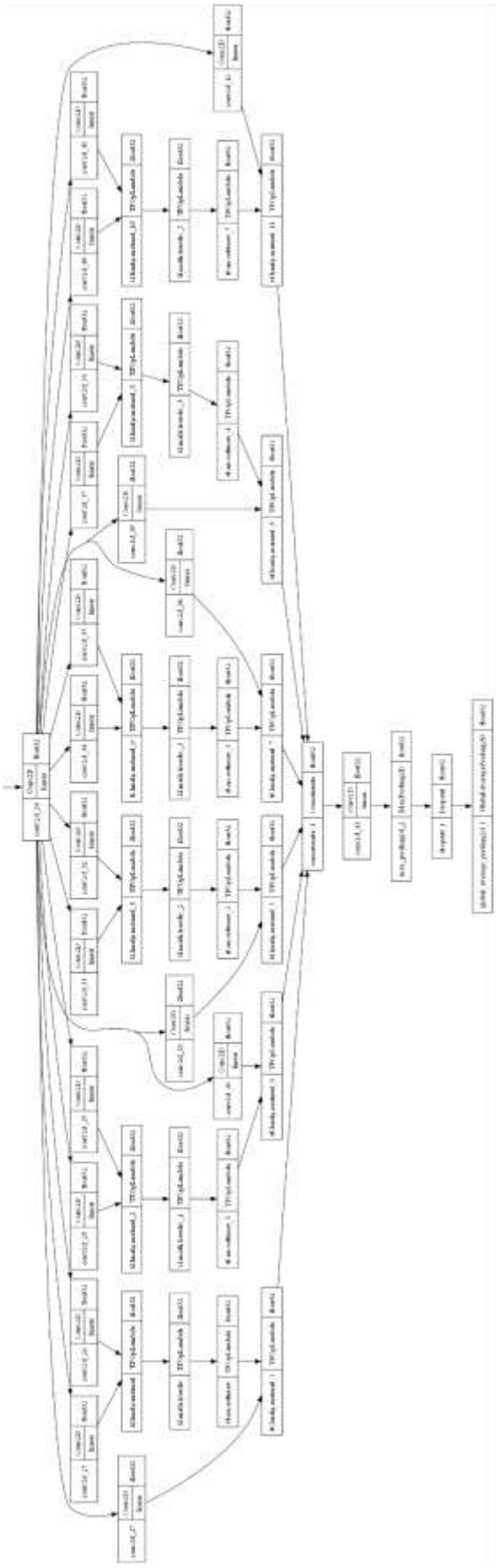


Figure 5.5 : Overview of the Vision Transformer proposed model

5.2.3 Capsule Networks (CapsNets)

CApsule networks are an advanced type of neural network architecture designed to overcome some limitations of traditional Convolutional Neural Networks (CNNs), particularly regarding the preservation of spatial hierarchies and relationships between features.

Unlike CNNs, which output scalar activations representing the presence of features, CapsNets organize neurons into capsules that output vectors or matrices. These vectors encode not only the probability that a feature is present but also its instantiation parameters, such as pose (position, orientation, size), deformation, texture, and other attributes. This vector representation allows CapsNets to capture more detailed information about how parts relate to a whole object.

A key innovation in CapsNets is the dynamic routing-by-agreement mechanism, which iteratively determines the strength of connections between capsules in consecutive layers based on how well their predictions agree. This process enables the network to model part-whole relationships explicitly, facilitating spatial hierarchy preservation and improving robustness to viewpoint changes and affine transformations.

In the context of medical imaging, particularly histopathological breast cancer detection, CapsNets help maintain the structural organization of tissue features such as nuclei, glands, and tumor shapes. This capability allows the network to better distinguish between benign and malignant lesions, where subtle spatial relationships are diagnostically important.

In the HNet framework, CapsNet is used as the final stage to process fused feature vectors from EfficientNet and Advanced Vision Transformer branches. The capsule network comprises two 4D capsules and employs three dynamic routing iterations. The capsules produce vector outputs encoding both the presence and spatial configuration of features, which are then passed to a softmax classification layer to predict the class (benign or malignant).

Overall, CapsNet enhances the model's interpretability and diagnostic accuracy by preserving spatial hierarchies and enabling relational reasoning beyond what standard CNNs can achieve.

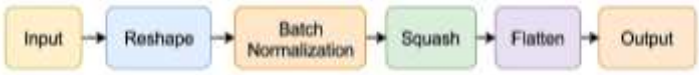


Figure 5.6 : Capsule Network flowchart

5.3 BreakHis Dataset Description

In order to completely examine malignancy in breast cancer tissues, biopsy procedures are routinely performed. The biopsy method (see Fig. 5.7.) entails obtaining tissue samples, mounting them to microscope slides, then staining these slides to clearly examine the cytoplasm and nucleus. Next, the conduct microscopic study of these slides was taken by pathologists to reach a final diagnosis of breast cancer [289]. The breast tissue picture is characterized by carrying a lot of useful information which is undetectable in the image such as morphological information, structural information of tissues, and other features of deep tissues. Specifically Kowal et al recovered 42 morphological, topological and structural characteristics from segmented nuclei, similarly, Filipczuk et al collected 25 shape and structure-based features from nuclei [289].

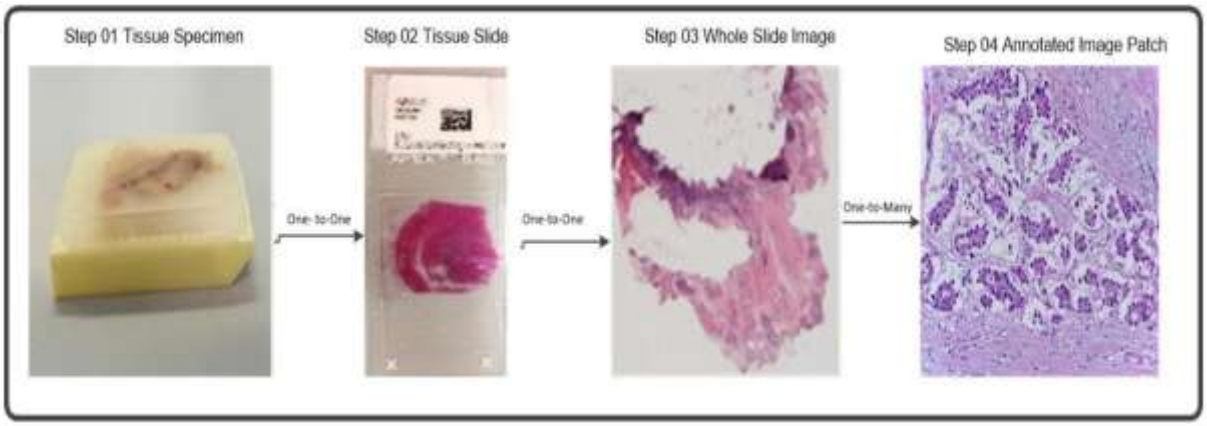


Figure 5.7 : The biopsy procedure for collecting tissue samples[289].

5.3.1 Dataset Origin and Purpose

The BreakHis dataset was developed by Spanhol [290] to provide a comprehensive collection of annotated histopathological images for breast tumor classification research. It aims to facilitate the development and benchmarking of computational models that can accurately differentiate between benign and malignant breast lesions.

5.3.2 Composition and Class Distribution

The dataset comprises a total of 7,909 breast histopathological images collected from 82 patients. These images are divided into two primary classes:

- Benign tumors: 2,480 images
- Malignant tumors: 5,429 images

- This imbalance reflects real clinical prevalence and poses a challenge for machine learning models.

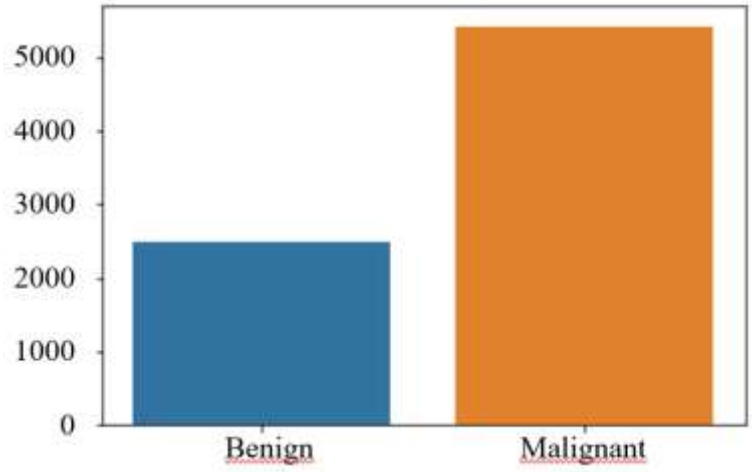


Figure 5.8:Distribution of Benign and Malignant Samples in the BreaKHis Dataset

5.3.3 Image Resolution and Magnifications

The images are RGB color micrographs captured at four different magnification levels: 40×, 100×, 200×, and 400×. Each image has a spatial resolution of 700 × 460 pixels, with magnification affecting the scale and detail of tissue features presented.

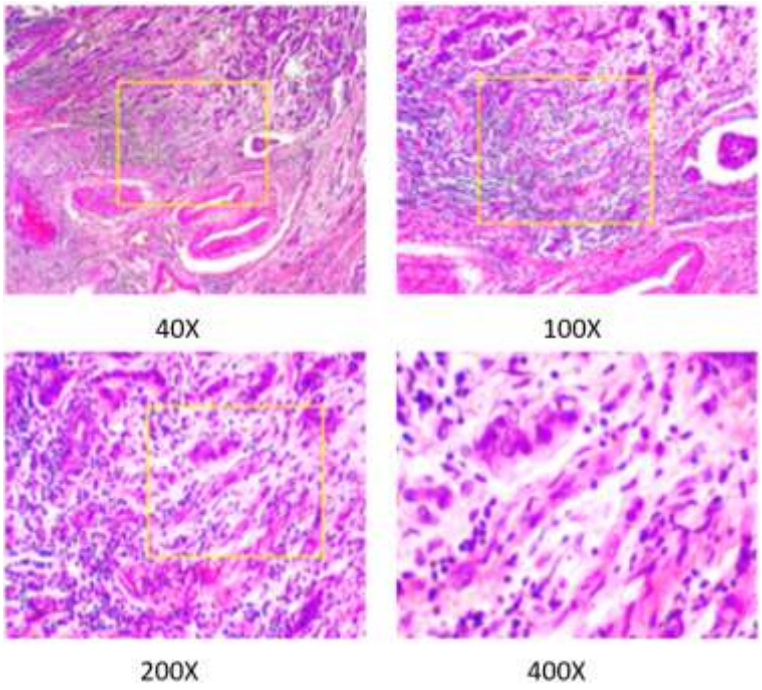


Figure 5.9 : A slide of breast malignant tumor seen in different magnification factors

The table below illustrates the distribution of images in the BreakHis dataset across four magnification levels (40×, 100×, 200×, and 400×) for both benign and malignant breast tumor classes. This organized structure enables multi-scale analysis and supports the development of robust deep learning models for breast cancer classification tasks.

Magnification	Benign	Malignant	Total
40x	652	1,370	1,995
100x	644	1,437	2,081
200x	623	1,390	2,013
400x	588	1,232	1820
Total	**2,480**	**5,429**	**7,909**

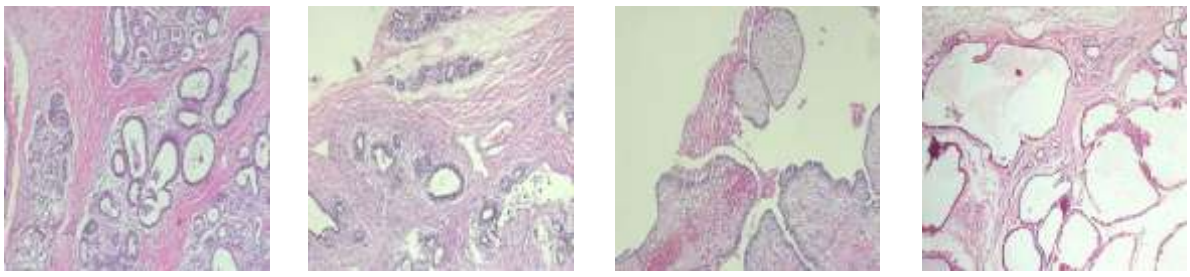
Table 5-1 BreakHis Dataset Composition Across Magnification Levels

5.3.4 Subcategories of Benign and Malignant Classes

Within the two main classes, images are further categorized into histological subtypes to capture tumor heterogeneity:

Benign Tumor Subtypes

The benign category encompasses four histologically distinct tumor types: Adenosis (A) with 444 images (17.9%), Fibroadenoma (F) with 1,014 images (40.9%), Phyllodes Tumor (PT) with 569 images (22.9%), and Tubular Adenoma (TA) with 453 images (18.3%). These lesions are characterized by the absence of malignancy criteria such as marked cellular atypia, mitosis, and basement membrane disruption [291].

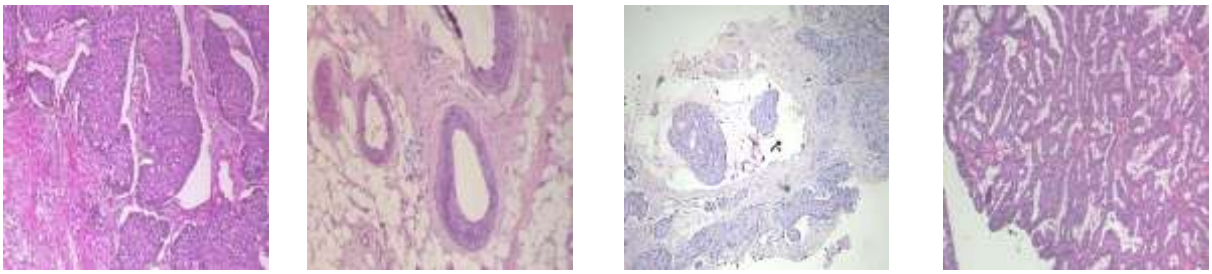


Adenosis **Fibroadenoma** **Phyllodes** **Tubular**

Figure 5.10 : Examples of Benign Subtypes from the BreakHis Dataset in x40 zoom

Malignant Tumor Subtypes

The malignant category includes four primary breast cancer types: Ductal Carcinoma (DC) with 3,451 images (63.6%), Lobular Carcinoma (LC) with 626 images (11.5%), Mucinous Carcinoma (MC) with 792 images (14.6%), and Papillary Carcinoma (PC) with 560 images (10.3%). These represent invasive breast cancers capable of local tissue destruction and distant metastasis [292].



Ductal carcinoma **Lobular carcinoma** **Mucinous carcinoma** **Papillary carcinoma**

Figure 5.11 : Examples of Malignant Subtypes from the BreakHis Dataset in x40 zoom

The figure bellow illustrates the distribution of the benign and malignant subcategories within the BreakHis dataset across four magnification levels: 40×, 100×, 200×, and 400×. Each histopathological subtype, including adenosis, fibroadenoma, phyllodes tumor, and tubular adenoma for benign cases, and carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma for malignant cases, is represented at each magnification level. This visualization highlights the multi-scale nature of the dataset, with images spread relatively evenly across magnifications to capture tissue details at varying resolutions. Such a distribution enables the model to learn scale-invariant features and enhances generalization to diverse clinical imaging conditions.

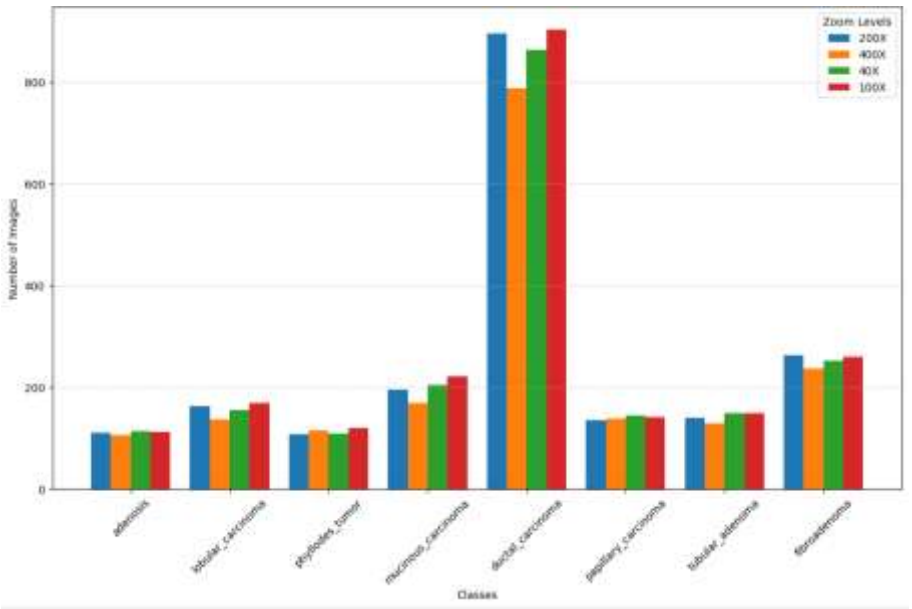


Figure 5.12 : Class distribution across zoom levels

Technical Specifications and Data Acquisition

Imaging Equipment and Parameters

The dataset was acquired using an Olympus BX-50 system microscope equipped with a 3.3× magnification relay lens and connected to a Samsung SCC-131AN digital color camera. The camera utilizes a 1/3" Sony Super-HAD (Hole-Accumulation Diode) interline transfer CCD with a pixel size of 6.5 μm × 6.25 μm and supports four objective lens magnifications: 40×, 100×, 200×, and 400×.

Image Characteristics

All images are captured as 700 × 460-pixel RGB images with 24-bit color depth (8-bit per channel) and saved in PNG format. The corresponding pixel sizes in the object plane are 0.49 μm at 40×, 0.20 μm at 100×, 0.10 μm at 200×, and 0.05 μm at 400× magnification [293].

Sample Collection and Processing Methodology

Biopsy Procedure

All tissue samples were collected using the Surgical Open Biopsy (SOB) method, also known as partial mastectomy or excisional biopsy. This procedure removes larger tissue samples compared to needle biopsy methods and is performed under general anesthesia in a hospital setting [294].

Histological Preparation

Tissue sections underwent standard Hematoxylin and Eosin (H&E) staining protocols. Hematoxylin stains cell nuclei purplish-blue through binding to nucleic acids, while eosin stains cytoplasm and extracellular matrix pink, providing essential morphological contrast for pathological assessment [295].

5.4 Data preparation

This section details the three core steps applied to histopathology images prior to model training: preprocessing, resizing, and normalization. The goals are to reduce acquisition artifacts and stain variability, standardize spatial scale for batch processing, and place inputs in a numerically stable range aligned with the backbone's training distribution.

5.4.1 Preprocessing

Purpose: Improve signal quality, reduce nuisance variability, and prepare images for consistent downstream processing while preserving diagnostically relevant morphology.

5.4.2 Resizing.

This step standardizes all images to a fixed spatial resolution required by the backbone and batching pipeline while preserving diagnostically relevant morphology. Given that histopathology contains fine-grained structures such as nuclei, gland boundaries, and stromal textures, the target size is selected to

balance detail retention and computational cost (e.g., 96×96, 128×128, or 224×224 depending on the experiment). When source images share a uniform native size (such as 700×460 in BreakHis), the aspect ratio is handled consistently across all splits to avoid distribution shift: either a center crop or padding (letterboxing) is applied to reach the desired aspect ratio before resizing, or direct resizing is used if validated to have negligible distortion on classification performance. For downscaling, area or bicubic interpolation is employed to minimize aliasing and preserve structural cues; for upscaling, bicubic interpolation is preferred to avoid block artifacts and maintain edge continuity. In multi-magnification settings (40×/100×/200×/400×), the same resizing policy is applied across magnifications, and complementary scale-aware augmentations can be used to mitigate magnification bias. The entire resizing pipeline is deterministic and identical for training, validation, and testing to ensure reproducibility and prevent inadvertent covariate shift.

5.4.3 Normalization.

After resizing, pixel intensities are mapped to a numerically stable range aligned with the initialization of the backbone to improve optimization and generalization. First, 8-bit RGB pixels are converted to floating point and scaled to the range by dividing by 255. This is followed by channel-wise standardization using either the reference mean and standard deviation of the pretrained backbone (e.g., ImageNet statistics for EfficientNet) or dataset-specific statistics computed strictly on the training set to avoid information leakage. Formally, each pixel is normalized as x_norm = (x/255 − mean) / std, applied per RGB channel. If stain normalization or color harmonization is used earlier in preprocessing, normalization is applied afterwards so that the standardized values reflect the stabilized color space. The same normalization configuration is used for training, validation, and testing, and tensors are stored in float32. Pixel-space augmentations that assume raw intensity distributions are performed prior to normalization; augmentations that assume standardized inputs are applied after normalization. This disciplined ordering, resizing first, then normalization, ensures geometric operations act on native intensities while the model ultimately receives standardized inputs conducive to stable and reproducible training.

The BreakHis dataset contains an imbalanced distribution of breast cancer histopathological images, with 2,480 benign samples and 5,429 malignant samples. To address this imbalance, upsampling techniques were applied to the benign class to equalize the number of samples between benign and malignant groups. This balancing step is critical to prevent the model from becoming biased towards the majority malignant class and to improve its ability to correctly classify benign tumors. By generating synthetic or augmented benign samples, the training dataset becomes more representative and equitable, enabling the deep learning model to learn discriminative features across both classes effectively. This approach leads to better generalization performance and reduces the risk of overfitting to the dominant malignant class in the BreakHis dataset.

5.4.4 Data augmentation

The dataset was augmented on-the-fly using a stochastic policy implemented with Keras' ImageDataGenerator. The policy included random zooming, in-plane rotations, large translations, and mirror flips along both axes. Specifically, zoom_range=1.2 sampled isotropic zoom factors within a bounded range to vary the effective field of view while preserving tissue morphology. rotation_range=90 introduced rotation-invariance to the classifier by randomly rotating patches within ±90°, which is appropriate for histopathology where diagnostic structures are orientation-agnostic. width_shift_range=0.5 and height_shift_range=0.5 applied random translations up to 50% of the image size in each axis, encouraging robustness to tissue placement and local context variation. Finally, horizontal_flip=True and vertical_flip=True mirrored images across x- and y-axes, further improving invariance to tissue orientation and mitigating overfitting. Augmentations were applied only to the training split and sampled independently per batch, ensuring that each epoch exposed the network to new, yet histologically plausible, appearances without altering class semantics.

Expert recommendations for histopathology

- Shift ranges: 0.5 (50%) is very large and can push most of the tissue out of view on small inputs. For 96–224 px inputs, consider 0.05–0.2 unless you use fill_mode='reflect' and verify that context remains informative.
- Zoom: zoom_range=1.2 is typically safe; if you want symmetric zoom-in/out, use a tuple (0.8, 1.2) so patches sometimes zoom out to include more context.
- Rotations: ±90° is sensible for H&E; some groups allow 0–180° or 0–360°. If you want uniform coverage, set rotation_range=180 or use RandomRotation with fill_mode='reflect'.
- Flips: Horizontal and vertical flips are both acceptable for H&E patches; keep both enabled.
- Fill mode: When rotating/translating/zooming, set fill_mode='reflect' (or 'nearest') to avoid black borders that can become spurious cues.
- Order in pipeline: Apply augmentations before normalization if they operate in pixel space; keep validation/test unaugmented.
- Reproducibility: Fix a seed for deterministic experiments, and log the policy in the paper's Data Preparation section.

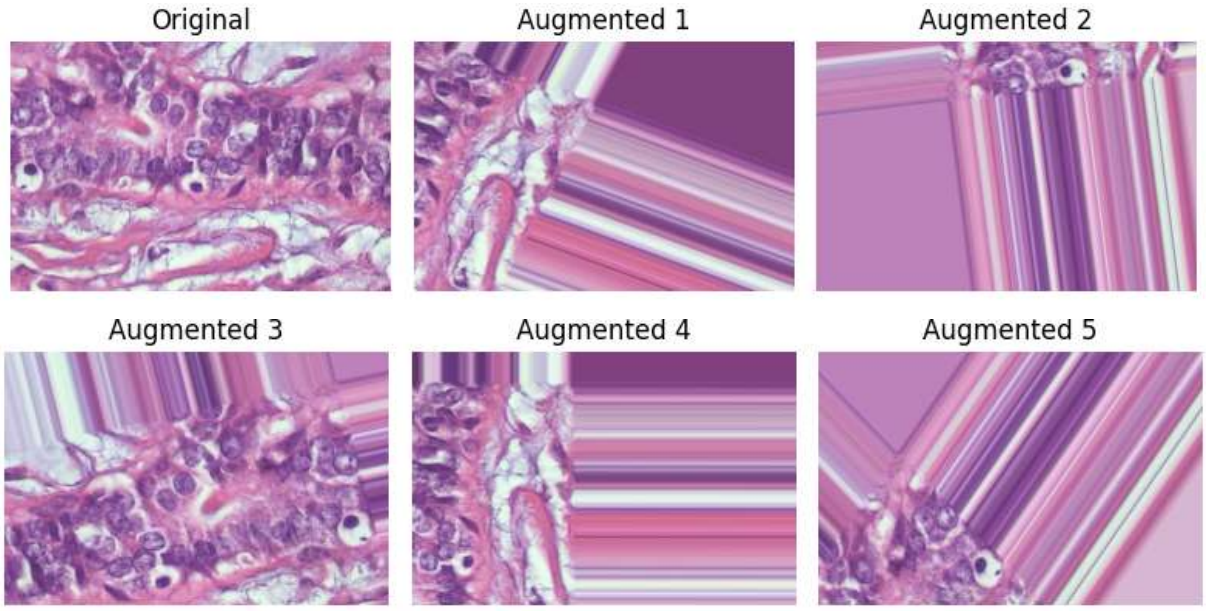


Figure 5.13 : Examples of Augmented Breast Cancer Images Using Data Augmentation Techniques

5.5 Results and discussion

5.5.1 Overview of Experiments

We present a comprehensive experimental evaluation of the proposed HNet architecture on the BreakHis dataset under various conditions and configurations. The experiments were systematically designed to assess multiple dimensions of model performance, including classification accuracy, robustness across different data availability scenarios, sensitivity to input resolution, and computational efficiency. All experiments were conducted on a high-performance computing system equipped with an NVIDIA RTX 3000 GPU (6GB VRAM), Intel i7-10850H CPU, and 32GB RAM, utilizing TensorFlow 2.8 and Python 3.9.

The experimental framework evaluates HNet using standard classification metrics including accuracy, precision, recall, and F1-score, providing a multi-faceted assessment of diagnostic performance. To ensure reproducibility and statistical validity, multiple experimental runs were conducted with different random seeds, and results were averaged across these runs. The evaluation methodology encompasses six distinct experimental scenarios examining data splitting strategies, three resolution configurations, ablation studies to assess component contributions, confusion matrix analysis for class-wise performance evaluation, training convergence behavior analysis, comparative assessment against state-of-the-art methods, and computational cost profiling.

5.5.2 Experimental Scenarios

The experimental design incorporates multiple scenarios to comprehensively evaluate HNet's performance characteristics such as different split configurations and image.

Data Split Configuration Experiments: Six different train-validation-test split ratios were evaluated, for example for setting of 70, 20 and 10% of training, validation and test set respectively was initially adopted, after that, different settings were adopted to get the best choise in term of different metrics such as 70,15 and 15%, 75-15 and 10%, 80,10 and 10%, 85,10 and 5%, and 90, and 5% to assess model robustness under varying data availability conditions. This analysis helps determine the minimum training data requirements while maintaining acceptable diagnostic performance, which is critical for scenarios with limited annotated datasets.

Multi-Resolution Analysis: in our experiments, we have proposed three input image resolutions (96×96, 128×128, and 224×224 pixels) those were systematically evaluated to investigate the trade-off between computational efficiency and classification accuracy. This analysis provides practical guidance for deployment scenarios with different computational constraints, from resource-limited mobile devices to high-performance clinical workstations.

Component-Level Ablation: Individual components of the HNet architecture (EfficientNet, Advanced Vision Transformer, and Capsule Network) were evaluated both in isolation and in combination to quantify each module's contribution to overall performance. This component-wise analysis validates the hybrid design philosophy and demonstrates the synergistic benefits of combining complementary architectures.

Class-Wise Performance Analysis: Detailed confusion matrix analysis was performed to examine the model's discriminative capabilities for benign versus malignant classification, identifying potential biases and class-specific strengths or weaknesses.

5.5.3 Impact of Data Splitting Strategies

To assess HNet's robustness and generalization capability under different data availability conditions, we systematically evaluated six train-validation-test split configurations using a fixed input resolution of 128×128 pixels. This resolution was selected as it represents an optimal balance between computational efficiency and classification performance, as detailed in Section 5.4.

The results presented in Table 5.2 demonstrate that HNet maintains consistently high performance across all split ratios, with accuracy ranging from 93.65% to 97.15%. The optimal performance was achieved with the 70-20-10 split configuration, yielding an accuracy of 97.15%, precision of 99.04%, recall of 95.38%, and F1-score of 97.18%. This configuration provides sufficient training data for effective learning while allocating adequate samples for robust validation and testing.

Data Split (Train-Val-Test)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
70-20-10	97.15	99.04	95.38	97.18
70-15-15	94.98	97.22	92.92	95.02

75-15-10	95.40	96.07	94.76	95.41
80-10-10	93.65	96.41	91.28	93.78
85-10-5	95.76	94.75	96.82	95.77
90-5-5	96.13	94.75	96.15	95.45

Table 5-2 : Evaluation Metrics Across Various Data Splits for the HNet Model

Interestingly, the 90-5-5 split, despite allocating the largest proportion to training (90%), achieved an accuracy of 96.13%, slightly lower than the 70-20-10 configuration. This counterintuitive result suggests that the extremely small validation set (5%) may provide insufficient samples for effective model selection and early stopping, potentially leading to suboptimal hyperparameter choices. The 80-10-10 split showed the lowest performance (93.65% accuracy), possibly due to reduced validation data limiting the model's ability to monitor generalization during training.

The relatively small performance variation across configurations (range of 3.5 percentage points) demonstrates HNet's strong generalization capability and robustness to data availability. This characteristic is particularly valuable for medical imaging applications where annotated datasets are often limited due to the high cost and expertise required for manual labeling. The consistent high precision across all splits (ranging from 94.75% to 99.04%) indicates that HNet maintains low false positive rates, which is crucial for clinical applications where unnecessary biopsies or treatments resulting from false alarms impose significant physical and psychological burden on patients.

5.5.4 Influence of Input Image Resolution

Image resolution is a critical parameter in histopathological image analysis, as it directly impacts the preservation of fine-grained morphological structures essential for accurate diagnosis. To investigate this relationship systematically, we evaluated HNet at three different input resolutions: 96×96, 128×128, and 224×224 pixels, using the optimal 70-20-10 data split identified in Section 5.3.

The results, summarized in Table 5.3, reveal a clear positive correlation between input resolution and classification performance. The highest performance was achieved at 224×224 resolution, with an accuracy of 97.52%, precision of 98.52%, recall of 95.43%, and F1-score of 96.95%. This superior performance can be attributed to the preservation of high-frequency spatial details and fine-grained tissue structures that are critical for distinguishing between benign and malignant lesions. At higher resolutions, cellular nuclei boundaries, glandular architecture, and stromal texture patterns remain more distinct, providing richer discriminative features for the model.

Resolution	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
96×96	94.70	91.40	95.50	93.40

128×128	97.15	99.04	95.38	97.18
224×224	97.52	98.52	95.43	96.95

Table 5-3 : Performance of HNet at Different Input Image Resolutions

In contrast, the 96×96 resolution yielded the lowest performance metrics, with accuracy dropping to 94.70%. This degradation suggests that excessive downsampling leads to loss of diagnostically relevant morphological information. When histopathological images are compressed to very small dimensions, critical features such as nuclear pleomorphism, mitotic figures, and architectural distortion may be obscured or eliminated entirely, limiting the model's discriminative capacity.

The 128×128 resolution provides an excellent compromise, achieving 97.15% accuracy—only 0.37 percentage points lower than the highest resolution—while requiring significantly less computational resources. This resolution was therefore selected as the default configuration for subsequent experiments. The marginal performance gain from 128×128 to 224×224 (0.37% improvement in accuracy) must be weighed against the substantially increased computational cost, which will be discussed in detail in Section 5.9. For real-time clinical deployment scenarios or resource-constrained environments, the 128×128 configuration offers near-optimal diagnostic performance with practical computational efficiency.

5.5.5 Ablation Study and Component-Level Analysis

To rigorously quantify the individual contributions of each architectural component and validate the synergistic benefits of the hybrid design, we conducted comprehensive ablation experiments. All ablation studies were performed using the 128×128 resolution and the 70-20-10 data split to ensure fair comparison.

Three model configurations were evaluated: (1) EfficientNet-only configuration, utilizing solely the convolutional backbone for local feature extraction; (2) Advanced Vision Transformer (AVT)-only configuration, employing only the transformer branch for global context modeling; and (3) Full HNet configuration, integrating EfficientNet, AVT, and Capsule Networks in the complete hybrid architecture. The comparative results are presented in Table 5.4.

Configuration	**Components**	**F1-Score (%)**	**Key Observations**
EfficientNet	CNN only	96.40	Strong local feature extraction, limited global context
AVT	Transformer only	92.67	Effective global modeling but weak local detail capture

HNet	CNN + AVT + Capsule	96.95	Optimal balance of local, global, and spatial features

Table 5-4: Performance Analysis of Different Architectural Components in HNet

The EfficientNet-only configuration achieved an F1-score of 96.40%, demonstrating the strong capability of convolutional architectures for extracting rich local textures and morphological features from histopathological images. EfficientNet's compound scaling approach and efficient use of depthwise separable convolutions enable effective capture of fine-grained cellular patterns, nuclear morphology, and tissue texture characteristics. However, the standalone CNN architecture struggles with modeling long-range dependencies and global contextual relationships across spatially distant tissue regions, which limits its ability to capture architectural patterns that extend beyond local receptive fields.

In contrast, the AVT-only configuration yielded the lowest performance with an F1-score of 92.67%. While the transformer's self-attention mechanism excels at capturing global dependencies and contextual information across the entire image, it demonstrates limitations in extracting fine-grained local features such as individual cell boundaries, nuclear characteristics, and subtle textural variations. This finding aligns with recent research indicating that pure transformer architectures, when applied to medical imaging tasks without extensive pre-training or large datasets, may underperform compared to convolutional approaches for tasks requiring detailed local feature discrimination.

The full HNet model achieved the highest F1-score of 96.95%, outperforming both individual configurations. This superior performance validates the hybrid design philosophy, demonstrating that the complementary strengths of convolutional networks, transformers, and capsule networks create synergistic benefits that enhance both discriminative power and generalization capacity. The integration of EfficientNet's local feature extraction, AVT's global context modeling, and Capsule Networks' spatial hierarchy preservation enables HNet to capture multi-scale representations ranging from fine cellular details to broad architectural patterns, while explicitly modeling part-whole relationships critical for histopathological diagnosis.

5.5.6 Confusion Matrix and Class-Wise Evaluation

Confusion matrix analysis provides detailed insights into class-specific performance characteristics, revealing the balance between sensitivity (recall) and specificity, and identifying potential systematic biases in predictions. We analyzed confusion matrices for three model configurations, AVT-only, EfficientNet-only, and full HNet, to understand how architectural choices affect class-wise discrimination capabilities.

Model	True Negatives	False Positives	False Negatives	True Positives
AVT	1033	46	118	975
EfficientNet	1060	26	51	1035
HNet	1063	16	51	1042

Table 5-5: Confusion Matrix Comparison Across AVT, EfficientNet, and HNet Models

The full HNet model demonstrated optimal class-wise discrimination, achieving 1042 true positives and 1063 true negatives while maintaining relatively low false positives (16) and false negatives (51). This configuration provides an excellent balance between sensitivity (recall = 95.38%) and specificity, making it highly suitable for clinical decision support where both false alarms and missed diagnoses carry significant consequences.

EfficientNet exhibited slightly fewer false positives (26) compared to HNet but achieved fewer true positives (1035), suggesting a more conservative prediction strategy that prioritizes specificity at the expense of sensitivity. The AVT-only configuration showed the highest misclassification rates with 46 false positives and 118 false negatives, indicating comparatively lower precision and recall. The high false negative rate is particularly concerning in medical applications, as it corresponds to malignant cases being incorrectly classified as benign, potentially delaying necessary treatment.

The HNet model's superior performance in minimizing false positives (16) is particularly noteworthy, as it indicates high precision (99.04%) for malignant classification. In clinical practice, reducing false positive rates decreases unnecessary biopsies, invasive procedures, and patient anxiety, while the maintained high sensitivity ensures that malignant cases are not overlooked. This balance makes HNet a reliable tool for clinical decision support in breast cancer histopathology.

5.5.7 Training Convergence and Learning Behavior

Understanding training dynamics through learning curves is essential for diagnosing potential issues such as overfitting, underfitting, or convergence problems, and for validating that the model has learned meaningful patterns rather than memorizing training data. We monitored accuracy and loss metrics over 100 training epochs for the HNet model configured at 224×224 resolution with the 70-20-10 data split. The training curves demonstrated smooth and stable convergence characteristics. The validation loss stabilized after approximately 60 epochs, indicating that the model reached an optimal state where further training provided diminishing returns. This convergence pattern is indicative of effective

learning, where the model progressively discovers underlying data patterns without overfitting to training-specific idiosyncrasies.

The implementation of early stopping and learning rate scheduling mechanisms effectively mitigated overfitting risks. Early stopping monitors validation loss and terminates training when no improvement is observed over a predefined patience period, preventing the model from continuing to fit noise in the training data. The reduce-on-plateau learning rate scheduler halves the learning rate after five epochs of stagnation, allowing the optimizer to make finer adjustments as it approaches local minima, thereby improving convergence quality.

The learning curves exhibited several desirable characteristics that indicate healthy training dynamics. First, the curves showed smoothness with gradual and consistent performance improvements, suggesting stable optimization without erratic fluctuations. Second, both training and validation losses decreased in tandem without a significant gap emerging, indicating good generalization to unseen data. Third, the curves converged to stable low values, demonstrating that the model reached its learning capacity for the given dataset and architecture.

These convergence patterns reinforce HNet's reliability for breast cancer histopathological image classification, demonstrating that the model learns robust, generalizable features rather than overfitting to spurious correlations in the training data.

5.5.8 Comparison with State-of-the-Art

To contextualize HNet's performance within the broader landscape of breast cancer classification research, we conducted a comprehensive comparative analysis with several state-of-the-art methods previously applied to the BreakHis dataset. The comparison includes traditional CNN-based approaches, attention-enhanced architectures, ensemble methods, and recent transformer-based models.

Model	**Study**	**Accuracy (%)**	**Key Features**
CBAM-EfficientNetV2	Sengodan 2025	99.01	EfficientNetV2-XL with Convolutional Block Attention Module
DeepBreastCancerNet	Ben Atitallah 2025	99.35	Ensemble of ResNet18, ShuffleNet, and Inception-V3Net
CBAM-VGGNet		98.96	VGG16 and VGG19 fusion with attention mechanisms

Inception-ResNet-v2 + Gradient Boosting		96.82	Feature extraction with ensemble classifiers
EfficientNet + Hybrid Attention		91.30	EfficientNet with hybrid attention mechanisms
HNet (Proposed Model)	**Current Study**	**97.52**	**EfficientNet + AVT + Capsule Networks**

Table 5-6: Comparison of State-of-the-Art Breast Cancer Classification Models

The proposed HNet framework achieves an impressive accuracy of 97.52%, positioning it competitively within the state-of-the-art landscape. While DeepBreastCancerNet (99.35%) and CBAM-EfficientNetV2 (99.01%) report marginally higher accuracies, these models employ substantially more complex ensemble strategies or deeper architectures that significantly increase computational requirements and model interpretability challenges.

HNet offers several advantages over these top-performing models. First, it avoids the excessive parameter load and inference complexity of deep ensemble methods, which require training and maintaining multiple independent models. Second, the integration of Capsule Networks provides enhanced interpretability through explicit modeling of part-whole spatial relationships, a feature absent in pure CNN or attention-based approaches. Third, HNet achieves its strong performance with a single unified model rather than relying on ensemble voting or complex fusion strategies, making it more practical for real-time clinical deployment.

Compared to Inception-ResNet-v2 with Gradient Boosting (96.82%) and EfficientNet with Hybrid Attention (91.30%), HNet demonstrates superior performance while maintaining architectural elegance and computational efficiency. The hybrid design philosophy, combining complementary strengths of CNNs, transformers, and capsule networks, enables HNet to capture rich multi-scale representations without resorting to extremely deep networks or complex ensemble architectures.

These results validate that HNet provides a compelling balance between diagnostic accuracy, computational efficiency, and model interpretability, making it particularly suitable for clinical decision-support systems where all three attributes are equally critical

5.5.9 Computational Cost and Efficiency

Understanding computational requirements is crucial for assessing model feasibility in real-world clinical deployment, particularly in resource-constrained environments or scenarios requiring real-time inference. We systematically profiled training time across different resolution configurations and model architectures to quantify the computational trade-offs associated with various design choices.

Model	Resolution	Data Split	Execution Time
EfficientNet	128×128	70-20-10	45 minutes
AVT	128×128	70-20-10	1 hour
HNet	128×128	70-20-10	1 hour 18 minutes
HNet	224×224	70-20-10	27 hours

Table 5-7: Execution Time Comparison of Different Models at Various Resolutions

Training time varied considerably depending on model complexity and input resolution. At 128×128 resolution with the 70-20-10 split, EfficientNet completed training in approximately 45 minutes, demonstrating the computational efficiency of the convolutional backbone. The AVT model required approximately 1 hour, reflecting the additional computational burden of self-attention mechanisms, which have quadratic complexity with respect to sequence length. The full HNet model took approximately 1 hour and 18 minutes under the same configuration, representing a reasonable computational overhead for the substantial performance improvements achieved through hybrid architecture integration.

However, increasing resolution to 224×224 resulted in a dramatic escalation in training time, with HNet requiring nearly 27 hours to complete training and evaluation. This substantial increase, more than 20× longer than the 128×128 configuration, underscores the significant computational demands of high-resolution image processing in deep learning workflows. The computational cost scales superlinearly with resolution due to multiple factors: increased number of pixels requiring processing, larger feature map dimensions propagating through network layers, higher memory requirements potentially forcing smaller batch sizes, and increased number of parameters when using adaptive architectures.

These findings have important practical implications for deployment strategies. The 128×128 configuration offers an optimal balance for scenarios requiring rapid model iteration, limited computational resources, or near-real-time inference requirements, while the 224×224 configuration is preferable when maximum diagnostic accuracy is paramount and computational resources are available. For production deployment, inference time (forward pass only) would be substantially lower than training time, making even the higher-resolution model potentially viable for clinical use on modern GPU-equipped workstations.

5.5.10 Summary of Findings

The comprehensive experimental evaluation of HNet across multiple dimensions provides strong evidence for its effectiveness as a breast cancer histopathology classification system. The key findings can be summarized across several critical aspects of model performance and practical viability.

Robust Generalization Across Data Availability Scenarios: HNet demonstrated consistent high performance across six different data split configurations, with accuracy ranging from 93.65% to 97.15%. The optimal 70-20-10 split achieved F1-score of 97.18%, while even the most constrained configuration-maintained accuracy above 93%, indicating robust generalization capabilities with limited training data, a crucial characteristic for medical imaging applications where annotated datasets are often scarce.

Resolution-Accuracy Trade-offs: The multi-resolution analysis revealed that 224×224 input achieved the highest accuracy (97.52%), but the 128×128 configuration provided an excellent compromise with 97.15% accuracy while requiring substantially less computational resources. This 0.37 percentage point difference represents a practical trade-off that enables deployment in resource-constrained environments without significant diagnostic performance degradation.

Validated Hybrid Architecture Design: Ablation studies conclusively demonstrated the synergistic benefits of the hybrid approach, with the full HNet model (F1-score: 96.95%) outperforming both EfficientNet-only (96.40%) and AVT-only (92.67%) configurations. This validates that combining local feature extraction, global context modeling, and spatial hierarchy preservation creates complementary representations that enhance discriminative power beyond what any single component can achieve.

Excellent Class-Wise Discrimination: Confusion matrix analysis revealed that HNet achieves optimal balance between sensitivity and specificity, with only 16 false positives and 51 false negatives, significantly outperforming standalone architectures. The high precision (99.04%) for malignant classification is particularly valuable for clinical applications, minimizing unnecessary interventions while maintaining high sensitivity for cancer detection.

Stable Training Dynamics: Learning curve analysis confirmed smooth convergence without overfitting, with validation loss stabilizing after approximately 60 epochs. The implementation of early stopping and learning rate scheduling effectively regulated training, ensuring the model learns generalizable patterns rather than memorizing training data.

Competitive State-of-the-Art Performance: With 97.52% accuracy, HNet positions competitively among leading methods on the BreakHis dataset while offering advantages in computational efficiency and interpretability compared to complex ensemble approaches. The model achieves this performance

through a single unified architecture rather than requiring multiple independent models or complex fusion strategies.

Practical Computational Requirements: While the 224×224 configuration requires approximately 27 hours for complete training, the 128×128 configuration trains in just 1 hour and 18 minutes with minimal performance compromise. This flexibility enables adaptation to different deployment scenarios, from rapid prototyping to high-accuracy production systems.

These findings collectively demonstrate that HNet represents a significant advancement in breast cancer histopathology classification, offering a well-balanced solution that addresses the key challenges of diagnostic accuracy, computational efficiency, and model interpretability required for real-world clinical deployment.

5.6 GA-based Optimization for HNet

The resultant fused feature space may still include redundant, correlated, or non-discriminative components even if the concatenation of feature vectors generated by EfficientNet and the Advanced Vision Transformer (AVT) offers a rich and complementary representation. When training data are limited, as is often the case in medical imaging applications, such redundancy may have a detrimental effect on classification performance, raise computing overhead, and restrict generalization. Our proposed HNet architecture overcomes this limitation by incorporating GA as an evolutionary feature optimization stage in order to overcome this restriction.

The GA acts immediately on the combined feature vector after feature concatenation and before to the Capsule Network. Finding the ideal collection of discriminative features that optimizes classification performance while reducing feature redundancy is its goal. A binary chromosome is used to encode each potential solution in the GA population, with each gene representing a feature dimension in the concatenated vector. A gene value of "0" implies feature exclusion, whereas a value of "1" suggests feature selection.

An initial population of chromosomes is generated randomly to ensure diversity and broad exploration of the search space. The population is then evolved iteratively using standard genetic operators, namely selection, crossover, mutation, and elitism.

- ✓ **Selection** is performed using a tournament-based strategy that favors high-quality feature subsets while preserving population diversity.
- ✓ **Crossover** enables information exchange between parent solutions, facilitating the combination of complementary feature subsets and accelerating convergence toward optimal solutions.
- ✓ **mutation** introduces controlled stochastic variations by randomly altering a small fraction of genes to avoid premature convergence and enhance exploration of the search space.
- ✓ In addition, **elitism** is employed to preserve the best-performing solutions across successive generations, ensuring that high-quality feature subsets are not lost during evolution.

The evolutionary process continues until a predefined termination criterion is satisfied. Specifically, the algorithm stops either when a maximum number of generations is reached or when the fitness value converges, indicating no significant improvement over successive generations. At termination, the chromosome with the highest fitness is selected as the optimal feature subset.

A composite objective function that strikes a compromise between feature compactness and classification performance is used to assess each chromosome's fitness. In particular, a lightweight classifier's validation-based performance measures (accuracy and F1-score), which are penalized by the percentage of chosen features, are included into the fitness function to encourage high diagnostic performance while penalizing excessive feature dimensionality, promoting compact and discriminative representations. (Equation 5.1). By encouraging the GA to find discriminative and compact feature ubsets, this architecture enhances computing efficiency and generalization.

$$Fitness = \alpha.F1 - score + \beta.Accuracy - \gamma.\frac{N_s}{N} \quad 5.1$$

Where:

- Ns = number of selected features
- N = total concatenated features
- α, β, γ = weighting coefficients

This formulation encourages the GA to identify feature subsets that achieve high diagnostic performance while minimizing redundancy, thereby enhancing computational efficiency and generalization capability.

Once the GA converges, the optimized feature vector is passed to the Capsule Network, where part-whole relationships and spatial hierarchies are modeled. By producing a compact and noise-reduced representation, the GA enhances classification robustness and improves the efficiency of the capsule network's dynamic routing.

This deep-evolutionary integration not only boosts classification accuracy but also improves interpretability and mitigates overfitting, while maintaining high diagnostic performance

The Capsule Network module, which carries out part-whole relationship modeling and spatial reasoning, receives the optimized feature vector generated by the GA. The GA improves classification robustness and boosts the efficiency of dynamic routing by giving the capsule network a more accurate and noise-free representation.

All things considered, the use of GA-based feature optimization enhances the suggested HNet architecture by supplementing evolutionary selection processes with deep learnt representations. By highlighting the most relevant characteristics for breast cancer histopathology image classification, our hybrid deep–evolutionary approach improves interpretability, lowers the danger of overfitting, and increases diagnostic accuracy.

The diagram illustrated in Figure 5.18 shows the GA section placement. After data preparation, features are extracted using EfficientNet and AVT models and then combined. A GA module optimizes these

features through selection and reduction before feeding them into a Capsule Network for final classification into benign or malignant, with performance evaluated using standard metrics.

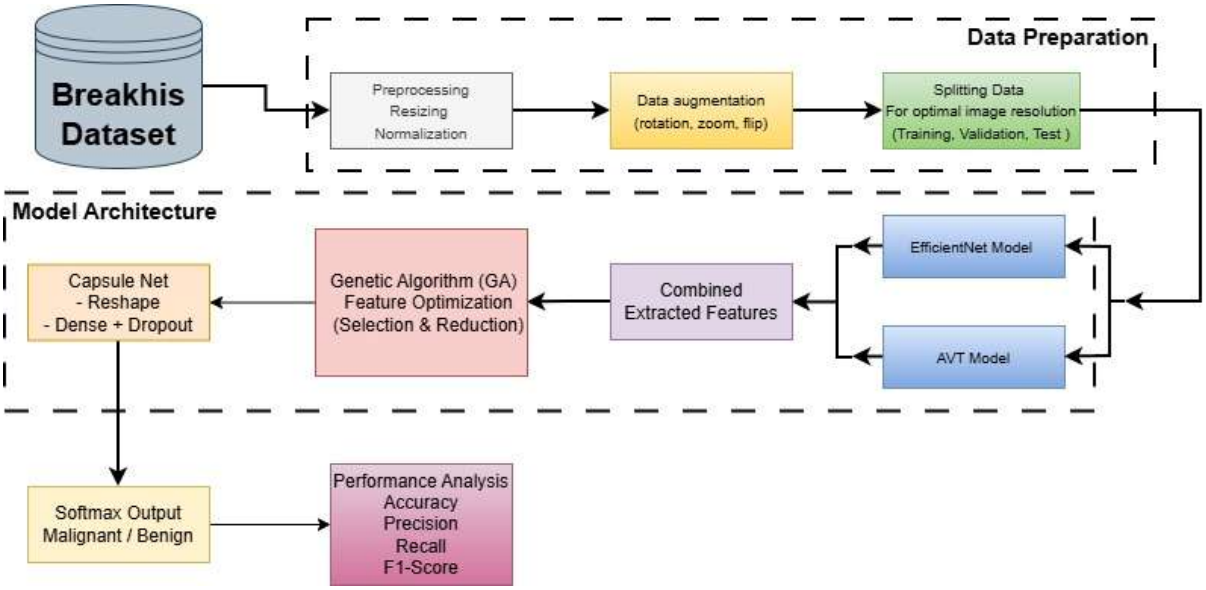


Figure 5.14: Updated HNet architecture incorporating GA-based feature optimization

5.6.1 GA-Based Feature Selection in HNet

To clearly illustrate the operational steps of the proposed GA-based feature selection mechanism within the HNet framework, Algorithm 5.1 summarizes the evolutionary process used to identify the optimal subset of deep fused features. The algorithm details the initialization of the population, fitness evaluation, application of genetic operators, and selection of the final optimized feature subset that is subsequently forwarded to the Capsule Network for classification.

Algorithm 5.1: GA-Based Feature Selection

Input: Fused feature vector F of dimension N
Output: Optimized feature subset F^*

1. Initialize population P with binary chromosomes of length N
2. Evaluate fitness of each chromosome using Eq. (5.1)
3. while termination condition not met do
4. Select parents using tournament selection
5. Apply crossover to generate offspring
6. Apply mutation to offspring
7. Evaluate new population fitness
8. Apply elitism to retain best solutions
9. Update population P
10. end while
11. Select best chromosome C^*
12. Generate optimized feature subset F^*
13. Return F^*

5.6.2 GA Parameters and Settings

These parameters were selected empirically based on preliminary experiments to ensure stable convergence and optimal performance.

Parameter	Baseline settings	Optimized settings
Population size	30	50
Generations	50	100
Selection	Tournament (k = 3)	
Crossover rate	0.7	0.8
Mutation rate	0.01	0.05
Elitism	5%	10%
Termination	Max generations or convergence	

Table 5-8: Genetic Algorithm Parameters and Configuration Settings

These parameters were selected empirically based on preliminary experiments to ensure stable convergence and optimal performance.

The GA parameters were determined empirically through preliminary experiments to balance exploration capability, convergence stability, and computational cost. Table 5.8 summarizes the baseline and optimized parameter settings used in this study.

The population size was increased from 30 to 50 to improve diversity in candidate feature subsets and enhance the search capability of the algorithm. A larger population allows better exploration of the feature space, reducing the risk of premature convergence.

The number of generations was extended from 50 to 100 to provide sufficient evolutionary iterations for convergence toward an optimal feature subset.

Tournament selection (k = 3) was adopted because it provides a good balance between selection pressure and population diversity while maintaining low computational complexity.

The crossover rate was increased from 0.7 to 0.8 to encourage greater information exchange between parent chromosomes, facilitating faster convergence toward high-quality feature combinations.

The mutation rate was increased from 0.01 to 0.05 to enhance exploration and avoid local optima, which is particularly important in high-dimensional fused feature spaces.

The elitism rate was increased from 5% to 10% to ensure that high-fitness solutions are preserved across generations, improving convergence stability.

The algorithm terminates when either the maximum number of generations is reached or the fitness value converges, ensuring both computational efficiency and solution stability.

5.6.3 GA-Based Feature Optimization's Effect (Results and discussion)

To evaluate the effectiveness of the proposed Genetic Algorithm feature-optimization module, a comparative ablation study was conducted between the baseline HNet architecture and the enhanced HNet+GA model. Both models were trained using identical experimental settings, including the same data split (70/20/10), training strategy, and input resolution (128×128), ensuring a fair comparison.

Table 5.9 shows that incorporating GA-based feature selection leads to consistent improvements across all evaluation metrics. The baseline HNet model achieved an F1-score of 97.18%, while the HNet+GA model using baseline GA settings improved the F1-score to 97.48% with a 20% reduction in feature dimensionality.

Model	Feature Optimization	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Feature Reduction
HNet	✗ None	97.15	99.04	95.38	97.18	-
HNet+GA	✓ Baseline settings	97.42	99.18	95.72	97.48	20%
HNet+GA	✓ Optimized settings	**97.85**	**99.32**	**96.01**	**97.95**	40%

Table 5-9: HNet vs HNet+GA Results Table

Further performance gains were observed with the optimized GA configuration. The HNet+GA (optimized) model achieved the best results, reaching an accuracy of 97.85% and an F1-score of 97.95%, while reducing the fused feature space by 40%. This demonstrates that GA-based feature selection effectively removes redundant and less informative components from the fused EfficientNet-AVT feature representation.

The reduction in feature dimensionality contributes to improved generalization and more efficient capsule-network routing by providing a compact and discriminative input representation. This is particularly beneficial in medical-image classification tasks where training data are limited and overfitting is a common concern.

Although the GA introduces additional optimization steps during training, the resulting model benefits from reduced feature size and improved inference efficiency. These results confirm that evolutionary feature optimization enhances the discriminative quality of deep fused representations and improves overall classification performance within the HNet framework.

5.7 Conclusion

The proposed HNet framework provides a robust and interpretable solution for breast cancer histopathological image classification by integrating EfficientNet for local feature extraction, the

Advanced Vision Transformer (AVT) for global context modeling, Capsule Networks for spatial hierarchy preservation, and a Genetic Algorithm for feature optimization. Experimental results on the BreakHis dataset demonstrate strong generalization across different data-split configurations and input resolutions, achieving an accuracy of up to 97.52% and an F1-score of 96.95%.

To evaluate the impact of evolutionary feature optimization, two GA configurations were investigated: baseline GA settings and optimized GA settings. The baseline configuration improved classification performance while reducing the fused feature dimensionality by approximately 20%. Further improvements were obtained with the optimized GA configuration, which achieved the best performance (97.85% accuracy and 97.95% F1-score) while reducing the feature space by 40%. These results confirm that GA-based feature selection effectively removes redundant information from the fused EfficientNet-AVT representation.

By producing a compact and discriminative feature representation, GA optimization enhances capsule-network routing efficiency and improves generalization in limited-data scenarios. The hybrid architecture demonstrates the complementary strengths of convolutional feature extraction, transformer-based global modeling, capsule-based spatial reasoning, and evolutionary feature selection.

HNet+GA achieves improved robustness, computational efficiency, and diagnostic performance compared with standalone CNN and transformer-based approaches, making it a practical framework for breast cancer histopathology image classification.

6 General conclusion

Breast cancer is still one of the most important global health issues, and improving patient outcomes and survival rates depends critically on early and precise detection. Computer-aided diagnostic systems have benefited greatly from the growing availability of medical imaging data, but creating models that are both very accurate and clinically interpretable remains a considerable issue. In order to overcome these obstacles, this thesis investigated sophisticated deep learning and hybrid optimization methods for the processing of histological breast cancer images.

The research described in this thesis examined the constraints of traditional convolutional neural networks in terms of maintaining the spatial hierarchies found in histological pictures and collecting global contextual information. A brand-new hybrid framework known as HNet was put out to get around these restrictions. HNet combines Capsule Networks to maintain spatial hierarchies and part-whole interactions, an Advanced Vision Transformer to represent long-range dependencies and global context, and EfficientNet for scalable and effective local feature extraction. By using the complementing capabilities of several learning paradigms, this architectural design makes it possible to create feature representations that are more reliable and understandable.

Additionally, this thesis presented an evolutionary feature optimization technique based on a genetic algorithm to improve the discriminative strength of the fused deep representations. The GA efficiently eliminated duplication and chose the most informative features before classification by working on the concatenated feature space. This hybrid deep-evolutionary technique demonstrated the advantages of combining evolutionary optimization with deep feature learning in medical imaging applications by improving generalization, decreasing overfitting, and increasing computing efficiency.

The efficacy of the suggested techniques across various input resolutions and data split configurations was confirmed by extensive experimental assessments carried out on the BreakHis dataset. The results showed that the suggested framework offers significant resilience to data fluctuation, balanced sensitivity and specificity, and excellent classification accuracy. Ablation experiments demonstrated the synergistic benefit of integrating convolutional, transformer-based, capsule-based, and evolutionary optimization strategies and further validated the contribution of each architectural component.

All things considered, this thesis offers a thorough and adaptable framework for the categorization of breast cancer histopathology images that strikes a compromise between computational efficiency, interpretability, and diagnostic accuracy. The suggested method has a great chance of being included into clinical decision-support systems, where scalability, transparency, and dependability are crucial. The techniques and insights discussed in this paper may be applied to various medical imaging applications requiring intricate spatial structures and sparse annotated data, in addition to breast cancer detection.

Extending the framework to multi-class breast cancer subtype classification, assessing cross-dataset generalization on broader and more varied benchmarks, and integrating sophisticated explainability methods like attention visualization and capsule activation analysis are some future research directions. Furthermore, combining multi-scale and multi-modal imaging data offers a viable way to enhance clinical relevance and diagnostic performance.

Future Work

Although the findings in this thesis show how well the suggested HNet framework and its GA-enhanced extension work for classifying breast cancer histopathology images, there are still a number of interesting avenues for further study.

Future research will first concentrate on expanding the suggested framework from binary classification to multi-class categorization of breast cancer subtypes. Higher inter-class similarity and intra-class variability make it more difficult to distinguish between various benign and malignant subtypes, which is clinically significant for prognosis and treatment planning.

Second, cross-dataset studies on other public benchmarks like BACH, Camelyon16, and other whole-slide picture datasets may be used to further assess the framework's capacity for generalization. Such investigations would aid in evaluating robustness under various staining procedures, population characteristics, and acquisition settings.

Third, characteristics collected at various magnification levels (40×, 100×, 200×, and 400×) will be collaboratively modeled in future research on multi-scale and multi-magnification learning procedures. This strategy may increase scale invariance and the model's capacity to represent patterns at the cellular and tissue levels.

Fourth, the methodical integration of metaheuristic optimization techniques at various deep learning pipeline stages is a viable avenue. Metaheuristics like Particle Swarm Optimization, Differential Evolution, Ant Colony Optimization, and Grey Wolf Optimizer might be investigated for hyperparameter tweaking, network architecture optimization, attention weight calibration, and training strategy refinement in addition to feature selection. Classification performance, convergence stability, and generalization may be further improved by using metaheuristics to adjust learning rates, layer configurations, routing parameters in capsule networks, or fusion weights between many feature branches. The creation of adaptable and self-optimizing deep learning frameworks for medical picture categorization would be made possible by this field of inquiry.

Fifth, in order to promote clinical adoption, improved model explainability will be attempted. Medical professionals may get clear and understandable diagnostic insights by combining methods like Grad-

CAM, attention visualization, capsule activation analysis, and metaheuristic-based feature priority ranking.

Lastly, deployment-oriented issues such model compression, inference-time acceleration, and integration with clinical decision-support systems will be the focus of future research. Translating the suggested framework into clinical practice will need working with pathologists to assess performance on whole-slide pictures and actual diagnostic procedures.

Bibliography

[1] D. H. Ballard *et al.*, 'The role of imaging in health screening: overview, rationale of screening, and screening economics', *Acad. Radiol.*, vol. 28, no. 4, pp. 540–547, Apr. 2021, doi: 10.1016/j.acra.2020.03.038.

[2] D. A. Castro, A. A. Naqvi, D. Manson, M. P. Flavin, E. VanDenKerkhof, and D. Soboleski, 'Novel Method to Improve Radiologist Agreement in Interpretation of Serial Chest Radiographs in the ICU', *J. Clin. Imaging Sci.*, vol. 5, Jul. 2015, doi: 10.4103/2156-7514.161848.

[3] 'Computer Assisted Diagnosis - an overview | ScienceDirect Topics'. Accessed: Dec. 09, 2025. [Online]. Available: https://www.sciencedirect.com/topics/medicine-and-dentistry/computer-assisted-diagnosis

[4] L. H. Eadie, P. Taylor, and A. P. Gibson, 'Recommendations for research design and reporting in computer-assisted diagnosis to facilitate meta-analysis', *J. Biomed. Inform.*, vol. 45, no. 2, pp. 390–397, Apr. 2012, doi: 10.1016/j.jbi.2011.07.009.

[5] M. Berger, Q. Yang, and A. Maier, 'X-ray Imaging', in *Medical Imaging Systems: An Introductory Guide*, A. Maier, S. Steidl, V. Christlein, and J. Hornegger, Eds, Cham (CH): Springer, 2018. Accessed: Dec. 09, 2025. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK546155/

[6] L. M. Hamberg, 'Basic Physics of Diagnostic X-ray Imaging'.

[7] D. Magid, J. S. Thompson, and E. K. Fishman, 'Computed Tomography of the Hand and Wrist', *Hand Clin.*, vol. 7, no. 1, pp. 219–233, Feb. 1991, doi: 10.1016/S0749-0712(21)01321-4.

[8] A. Pai, R. Shetty, B. Hodis, and Y. S. Chowdhury, 'Magnetic Resonance Imaging Physics', in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2025. Accessed: Dec. 09, 2025. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK564320/

[9] S. P. Grogan and C. A. Mount, 'Ultrasound Physics and Instrumentation', in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2025. Accessed: Dec. 09, 2025. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK570593/

[10] F. F. Alqahtani, 'SPECT/CT and PET/CT, related radiopharmaceuticals, and areas of application and comparison', *Saudi Pharm. J. SPJ*, vol. 31, no. 2, pp. 312–328, Feb. 2023, doi: 10.1016/j.jsps.2022.12.013.

[11] A. Rahmim and H. Zaidi, 'PET versus SPECT: strengths, limitations and challenges', *Nucl. Med. Commun.*, vol. 29, no. 3, pp. 193–207, Mar. 2008, doi: 10.1097/MNM.0b013e3282f3a515.

[12] R. Noumeir, 'Radiology interpretation process modeling', *J. Biomed. Inform.*, vol. 39, no. 2, pp. 103–114, Apr. 2006, doi: 10.1016/j.jbi.2005.07.001.

[13] E. Dikici, M. Bigelow, L. M. Prevedello, R. D. White, and B. S. Erdal, 'Integrating AI into radiology workflow: levels of research, production, and feedback maturity', *J. Med. Imaging*, vol. 7, no. 1, p. 016502, Jan. 2020, doi: 10.1117/1.JMI.7.1.016502.

[14] Internationale Atomenergie-Organisation, *Justification of medical exposure in diagnostic imaging: proceedings of an international workshop on justification of medical exposure in diagnostic imaging, held in Brussels, 2 - 4 September 2009*. in Publication / Division of Scientific and Technical Information, International Atomic Energy Agency, no. 1532. Vienna: Internat. Atomic Energy Agency, 2011.

[15] M. del Rosario Pérez, 'Referral criteria and clinical decision support: radiological protection aspects for justification', *Ann. ICRP*, vol. 44, no. 1_suppl, pp. 276–287, Jun. 2015, doi: 10.1177/0146645314551673.

[16] T. P. Szczykutowicz *et al.*, 'A General Framework for Monitoring Image Acquisition Workflow in the Radiology Environment: Timeliness for Acute Stroke CT Imaging', *J. Digit. Imaging*, vol. 31, no. 2, pp. 201–209, Apr. 2018, doi: 10.1007/s10278-018-0055-1.

[17] M. Mansourian, S. Khademi, and H. R. Marateb, 'A Comprehensive Review of Computer-Aided Diagnosis of Major Mental and Neurological Disorders and Suicide: A Biostatistical Perspective

on Data Mining', *Diagnostics*, vol. 11, no. 3, p. 393, Feb. 2021, doi: 10.3390/diagnostics11030393.
[18] R. Noumeir, 'Radiology interpretation process modeling', *J. Biomed. Inform.*, vol. 39, no. 2, pp. 103–114, Apr. 2006, doi: 10.1016/j.jbi.2005.07.001.
[19] R. A. Castellino, 'Computer aided detection (CAD): an overview', *Cancer Imaging*, vol. 5, no. 1, pp. 17–19, Aug. 2005, doi: 10.1102/1470-7330.2005.0018.
[20] K. Doi, 'Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential', *Comput. Med. Imaging Graph. Off. J. Comput. Med. Imaging Soc.*, vol. 31, no. 4–5, pp. 198–211, 2007, doi: 10.1016/j.compmedimag.2007.02.002.
[21] M. Khalifa and M. Albadawy, 'AI in diagnostic imaging: Revolutionising accuracy and efficiency', *Comput. Methods Programs Biomed. Update*, vol. 5, p. 100146, Jan. 2024, doi: 10.1016/j.cmpbup.2024.100146.
[22] K. Doi, H. MacMahon, S. Katsuragawa, R. M. Nishikawa, and Y. Jiang, 'Computer-aided diagnosis in radiology: potential and pitfalls', *Eur. J. Radiol.*, vol. 31, no. 2, pp. 97–109, Aug. 1999, doi: 10.1016/S0720-048X(99)00016-9.
[23] Z. Guo *et al.*, 'A review of the current state of the computer-aided diagnosis (CAD) systems for breast cancer diagnosis', *Open Life Sci.*, vol. 17, no. 1, pp. 1600–1611, Dec. 2022, doi: 10.1515/biol-2022-0517.
[24] N. A. Alhamdan, 'The Role of Deep Learning in Advancing Computer-Aided Diagnosis in Medical Imaging: A Comprehensive Review', *Rev. Contemp. Philos.*, vol. 22, pp. 4531–4540, 2023.
[25] M. K. Santos, J. R. Ferreira Júnior, D. T. Wada, A. P. M. Tenório, M. H. N. Barbosa, and P. M. de A. Marques, 'Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: advances in imaging towards to precision medicine', *Radiol. Bras.*, vol. 52, no. 6, pp. 387–396, 2019, doi: 10.1590/0100-3984.2019.0049.
[26] Z. Z. Qin *et al.*, 'Comparing the accuracy of computer-aided detection (CAD) software and radiologists from multiple countries for tuberculosis detection in chest X-Rays', *Sci. Rep.*, vol. 15, no. 1, p. 22540, Jul. 2025, doi: 10.1038/s41598-025-06164-w.
[27] W. Jorritsma, F. Cnossen, and P. M. A. van Ooijen, 'Improving the radiologist–CAD interaction: designing for appropriate trust', *Clin. Radiol.*, vol. 70, no. 2, pp. 115–122, Feb. 2015, doi: 10.1016/j.crad.2014.09.017.
[28] G. J. Bansal and K. G. Thomas, 'Imaging techniques in breast cancer', *Surg. Oxf.*, vol. 28, no. 3, pp. 117–124, Mar. 2010, doi: 10.1016/j.mpsur.2009.12.004.
[29] 'Computer-Aided Detection and Diagnosis in Mammography', in *Handbook of Image and Video Processing*, Academic Press, 2005, pp. 1195–1217. doi: 10.1016/B978-012119792-6/50130-3.
[30] A. A. Peters *et al.*, 'Performance of an AI based CAD system in solid lung nodule detection on chest phantom radiographs compared to radiology residents and fellow radiologists', *J. Thorac. Dis.*, vol. 13, no. 5, May 2021, doi: 10.21037/jtd-20-3522.
[31] A. Davidson, 'Computer-Aided Diagnosis: Advancements in Medical Imaging and Healthcare', *J. Med. Diagn. Methods*, vol. 12, no. 3, pp. 1–2, Jun. 2023, doi: 10.35248/2168-9784.23.12.421.
[32] K. Doi, 'Computer-aided diagnosis in medical imaging: historical review, current status and future potential', *Comput. Med. Imaging Graph. Off. J. Comput. Med. Imaging Soc.*, vol. 31, no. 4–5, pp. 198–211, 2007, doi: 10.1016/j.compmedimag.2007.02.002.
[33] M. L. Giger, N. Karssemeijer, and S. G. Armato, 'Computer-aided diagnosis in medical imaging', *IEEE Trans. Med. Imaging*, vol. 20, no. 12, pp. 1205–1208, Dec. 2001, doi: 10.1109/tmi.2001.974915.
[34] K. Suzuki, 'Overview of deep learning in medical imaging', *Radiol. Phys. Technol.*, vol. 10, no. 3, pp. 257–273, Sep. 2017, doi: 10.1007/s12194-017-0406-5.

[35] 'Modern Diagnostic Imaging Technique Applications and Risk Factors in the Medical Field: A Review - PMC'. Accessed: Oct. 11, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9192206/

[36] B. Remeseiro and V. Bolon-Canedo, 'A review of feature selection methods in medical applications', *Comput. Biol. Med.*, vol. 112, p. 103375, Sep. 2019, doi: 10.1016/j.compbiomed.2019.103375.

[37] M. Marques, A. Almeida, and H. Pereira, 'The Medicine Revolution Through Artificial Intelligence: Ethical Challenges of Machine Learning Algorithms in Decision-Making', *Cureus*, vol. 16, no. 9, p. e69405, Sep. 2024, doi: 10.7759/cureus.69405.

[38] I. H. Sarker, 'Machine Learning: Algorithms, Real-World Applications and Research Directions', *Sn Comput. Sci.*, vol. 2, no. 3, p. 160, 2021, doi: 10.1007/s42979-021-00592-x.

[39] E. I. Fernandez *et al.*, 'Artificial intelligence in the IVF laboratory: overview through the application of different types of algorithms for the classification of reproductive data', *J. Assist. Reprod. Genet.*, vol. 37, no. 10, pp. 2359–2376, Oct. 2020, doi: 10.1007/s10815-020-01881-9.

[40] J. Aftab, M. A. Khan, S. Arshad, S. ur Rehman, D. A. AlHammadi, and Y. Nam, 'Artificial intelligence based classification and prediction of medical imaging using a novel framework of inverted and self-attention deep neural network architecture', *Sci. Rep.*, vol. 15, no. 1, p. 8724, Mar. 2025, doi: 10.1038/s41598-025-93718-7.

[41] G. Litjens *et al.*, 'A Survey on Deep Learning in Medical Image Analysis', *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[42] E. Miranda, M. Aryuni, and E. Irwansyah, 'A survey of medical image classification techniques', in *2016 International Conference on Information Management and Technology (ICIMTech)*, Nov. 2016, pp. 56–61. doi: 10.1109/ICIMTech.2016.7930302.

[43] Meng Joo Er, R. Venkatesan, and Ning Wang, 'An online universal classifier for binary, multi-class and multi-label classification', in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Budapest, Hungary: IEEE, Oct. 2016, pp. 003701–003706. doi: 10.1109/SMC.2016.7844809.

[44] M. V. C. Aragão *et al.*, 'A practical evaluation of AutoML tools for binary, multiclass, and multilabel classification', *Sci. Rep.*, vol. 15, no. 1, p. 17682, May 2025, doi: 10.1038/s41598-025-02149-x.

[45] E. A. Cherman, M. C. Monard, and J. Metz, 'Multi-label Problem Transformation Methods: a Case Study', *CLEI Electron. J.*, vol. 14, no. 1, pp. 4–4, Apr. 2011.

[46] L. Dao and N. Q. Ly, 'Recent Advances in Medical Image Classification', *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 7, 2024.

[47] Manmeet Kaur, 'A Comprehensive Overview of Artificial Intelligence-Based Classification Techniques', *Int. J. Sci. Res. Arch.*, vol. 11, no. 2, pp. 125–129, Mar. 2024, doi: 10.30574/ijsra.2024.11.2.0387.

[48] R. Aggarwal *et al.*, 'Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis', *Npj Digit. Med.*, vol. 4, no. 1, p. 65, Apr. 2021, doi: 10.1038/s41746-021-00438-z.

[49] J. Aftab, M. A. Khan, S. Arshad, S. ur Rehman, D. A. AlHammadi, and Y. Nam, 'Artificial intelligence based classification and prediction of medical imaging using a novel framework of inverted and self-attention deep neural network architecture', *Sci. Rep.*, vol. 15, no. 1, p. 8724, Mar. 2025, doi: 10.1038/s41598-025-93718-7.

[50] C. Yuan, X. Jia, L. Wang, and C. Yang, 'Fine-grained Prototype Network for MRI Sequence Classification', *Curr. Med. Imaging*, vol. 21, p. e15734056361649, 2025, doi: 10.2174/0115734056361649250717162910.

[51] L. Gao, C. Liu, D. Arefan, A. Panigrahy, M. L. Zuley, and S. Wu, 'Medical Knowledge-Guided Deep Learning for Imbalanced Medical Image Classification', Apr. 14, 2022, *arXiv*: arXiv:2111.10620. doi: 10.48550/arXiv.2111.10620.

[52] L. Cui *et al.*, 'Towards Reliable Healthcare Imaging: A Multifaceted Approach in Class Imbalance Handling for Medical Image Segmentation', *Interdiscip. Sci. Comput. Life Sci.*, vol. 17, no. 3, pp. 614–633, 2025, doi: 10.1007/s12539-025-00726-2.
[53] D. Scholz, 'Imbalance-aware loss functions improve medical image classification'.
[54] L. Gao, L. Zhang, C. Liu, and S. Wu, 'Handling Imbalanced Medical Image Data: A Deep-Learning-Based One-Class Classification Approach', *Artif. Intell. Med.*, vol. 108, p. 101935, Aug. 2020, doi: 10.1016/j.artmed.2020.101935.
[55] F. Renard, S. Guedria, N. D. Palma, and N. Vuillerme, 'Variability and reproducibility in deep learning for medical image segmentation', *Sci. Rep.*, vol. 10, no. 1, p. 13724, Aug. 2020, doi: 10.1038/s41598-020-69920-0.
[56] J. Liu, X. Cai, and M. Niranjan, 'Medical image classification by incorporating clinical variables and learned features', *R. Soc. Open Sci.*, vol. 12, no. 3, p. 241222, Mar. 2025, doi: 10.1098/rsos.241222.
[57] C. Yuan, X. Jia, L. Wang, and C. Yang, 'Fine-grained Prototype Network for MRI Sequence Classification', *Curr. Med. Imaging*, vol. 21, p. e15734056361649, 2025, doi: 10.2174/0115734056361649250717162910.
[58] L. Alzubaidi *et al.*, 'Novel Transfer Learning Approach for Medical Imaging with Limited Labeled Data', *Cancers*, vol. 13, no. 7, p. 1590, Mar. 2021, doi: 10.3390/cancers13071590.
[59] J. Cho, K. Lee, E. Shin, G. Choy, and S. Do, 'How much data is needed to train a medical image deep learning system to achieve necessary high accuracy?', Jan. 07, 2016, *arXiv*: arXiv:1511.06348. doi: 10.48550/arXiv.1511.06348.
[60] 'Imbalanced Medical Image Segmentation with Pixel-dependent Noisy Labels'. Accessed: Nov. 17, 2025. [Online]. Available: https://arxiv.org/html/2501.06678v1
[61] M. Li, Y. Jiang, Y. Zhang, and H. Zhu, 'Medical image analysis using deep learning algorithms', *Front. Public Health*, vol. 11, p. 1273253, Nov. 2023, doi: 10.3389/fpubh.2023.1273253.
[62] Z. Hussain, F. Gimenez, D. Yi, and D. Rubin, 'Differential Data Augmentation Techniques for Medical Imaging Classification Tasks', *AMIA. Annu. Symp. Proc.*, vol. 2017, pp. 979–984, Apr. 2018.
[63] F. Garcea, A. Serra, F. Lamberti, and L. Morra, 'Data augmentation for medical imaging: A systematic literature review', *Comput. Biol. Med.*, vol. 152, p. 106391, Jan. 2023, doi: 10.1016/j.compbiomed.2022.106391.
[64] E. Goceri, 'Medical image data augmentation: techniques, comparisons and interpretations', *Artif. Intell. Rev.*, pp. 1–45, Mar. 2023, doi: 10.1007/s10462-023-10453-z.
[65] S. Dutta, P. Prakash, and C. G. Matthews, 'Impact of data augmentation techniques on a deep learning based medical imaging task', in *Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications*, SPIE, Mar. 2020, pp. 168–177. doi: 10.1117/12.2549806.
[66] M. Kim and H.-J. Bae, 'Data Augmentation Techniques for Deep Learning-Based Medical Image Analyses', 대한영상의학회지, vol. 81, no. 6, pp. 1290–1304, Nov. 2020, doi: 10.3348/jksr.2020.0158.
[67] M. Cossio, 'Augmenting Medical Imaging: A Comprehensive Catalogue of 65 Techniques for Enhanced Data Analysis', Mar. 02, 2023, *arXiv*: arXiv:2303.01178. doi: 10.48550/arXiv.2303.01178.
[68] J. Wang, H. Zhu, S.-H. Wang, and Y.-D. Zhang, 'A Review of Deep Learning on Medical Image Analysis', *Mob. Netw. Appl.*, vol. 26, no. 1, pp. 351–380, Feb. 2021, doi: 10.1007/s11036-020-01672-7.
[69] 'Feature extraction and feature selection in medical images', in *Intelligent Computing Techniques in Biomedical Imaging*, Academic Press, 2025, pp. 83–97. doi: 10.1016/B978-0-443-15999-2.00008-6.

[70] R. Josphineleela, S. Preethi, A. M, M. Srikanth, E. Ramesh, and V. A. Kolluru, 'Feature Extraction Techniques in Medical Imaging: A Systematic Review', *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 11, no. 5, p. 23, doi: 10.17762/IJRITCC.V11I5.6521.

[71] Z. Lai and H. Deng, 'Medical Image Classification Based on Deep Features Extracted by Deep Model and Statistic Feature Fusion with Multilayer Perceptron', *Comput. Intell. Neurosci.*, vol. 2018, p. 2061516, Sep. 2018, doi: 10.1155/2018/2061516.

[72] M. Z. Ahmed and C. Mahesh, 'An Efficient Image Based Feature Extraction and Feature Selection Model for Medical Data Clustering Using Deep Neural Networks', *Trait. Signal*, vol. 38, no. 4, pp. 1141–1148, Aug. 2021, doi: 10.18280/ts.380425.

[73] D. Beaglehole, A. Radhakrishnan, P. Pandit, and M. Belkin, 'Mechanism of feature learning in convolutional neural networks', Sep. 01, 2023, *arXiv*: arXiv:2309.00570. doi: 10.48550/arXiv.2309.00570.

[74] T. Maruyama *et al.*, 'Comparison of medical image classification accuracy among three machine learning methods', *J. X-Ray Sci. Technol.*, vol. 26, no. 6, pp. 885–893, 2018, doi: 10.3233/XST-18386.

[75] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY: Springer, 2000. doi: 10.1007/978-1-4757-3264-1.

[76] B. E. Boser, I. M. Guyon, and V. N. Vapnik, 'A training algorithm for optimal margin classifiers', in *Proceedings of the fifth annual workshop on Computational learning theory*, in COLT '92. New York, NY, USA: Association for Computing Machinery, juillet 1992, pp. 144–152. doi: 10.1145/130385.130401.

[77] C. Cortes and V. Vapnik, 'Support-vector networks', *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[78] C. C. Aggarwal, *Data Mining: The Textbook*. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-14142-8.

[79] L. Breiman, 'Random Forests', *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[80] L. Ye and E. Keogh, 'Time series shapelets: a novel technique that allows accurate, interpretable and fast classification', *Data Min. Knowl. Discov.*, vol. 22, no. 1, pp. 149–182, Jan. 2011, doi: 10.1007/s10618-010-0179-5.

[81] Y. Amit and D. Geman, 'Shape Quantization and Recognition with Randomized Trees', *Neural Comput.*, vol. 9, no. 7, pp. 1545–1588, Oct. 1997, doi: 10.1162/neco.1997.9.7.1545.

[82] T. Cover and P. Hart, 'Nearest neighbor pattern classification', *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967, doi: 10.1109/TIT.1967.1053964.

[83] A. Heena, N. Biradar, N. M. Maroof, S. Bhatia, R. Agarwal, and K. Prasad, 'Machine learning based biomedical image processing for echocardiographic images', arXiv.org. Accessed: Nov. 18, 2025. [Online]. Available: https://arxiv.org/abs/2303.09103v1

[84] D. Gupta, R. Loane, S. Gayen, and D. Demner-Fushman, 'Medical Image Retrieval via Nearest Neighbor Search on Pre-trained Image Features', Oct. 05, 2022, *arXiv*: arXiv:2210.02401. doi: 10.48550/arXiv.2210.02401.

[85] V. K. Prasad *et al.*, 'Revolutionizing healthcare: a comparative insight into deep learning's role in medical imaging', *Sci. Rep.*, vol. 14, no. 1, p. 30273, Dec. 2024, doi: 10.1038/s41598-024-71358-7.

[86] J. Aftab, M. A. Khan, S. Arshad, S. ur Rehman, D. A. AlHammadi, and Y. Nam, 'Artificial intelligence based classification and prediction of medical imaging using a novel framework of inverted and self-attention deep neural network architecture', *Sci. Rep.*, vol. 15, no. 1, p. 8724, Mar. 2025, doi: 10.1038/s41598-025-93718-7.

[87] G. K. Thakur, A. Thakur, S. Kulkarni, N. Khan, and S. Khan, 'Deep Learning Approaches for Medical Image Analysis and Diagnosis', *Cureus*, vol. 16, no. 5, p. e59507, doi: 10.7759/cureus.59507.

[88] H. Pant, G. Joshi, and B. Rawat, 'Precision Medicine for the Lungs: Deep Learning Applications in Thoracic Imaging', *Biomed. Pharmacol. J.*, vol. 18, no. 4, Nov. 2025, Accessed: Nov. 18,

2025. [Online]. Available: https://biomedpharmajournal.org/vol18no4/precision-medicine-for-the-lungs-deep-learning-applications-in-thoracic-imaging/
[89] M. N. Srinivasan, M. Y. Sikkandar, M. Alhashim, and M. Chinnadurai, 'Capsule network approach for monkeypox (CAPSMON) detection and subclassification in medical imaging system', *Sci. Rep.*, vol. 15, no. 1, p. 3296, Jan. 2025, doi: 10.1038/s41598-025-87993-7.
[90] P. K. Mall *et al.*, 'A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities', *Healthc. Anal.*, vol. 4, p. 100216, Dec. 2023, doi: 10.1016/j.health.2023.100216.
[91] K. Simonyan and A. Zisserman, 'Very Deep Convolutional Networks for Large-Scale Image Recognition', Apr. 10, 2015, *arXiv*: arXiv:1409.1556. Accessed: Apr. 06, 2023. [Online]. Available: http://arxiv.org/abs/1409.1556
[92] A. R. Ismail, S. Q. Nisa, S. A. Shaharuddin, S. I. Masni, and S. A. S. Amin, 'Utilising VGG-16 of Convolutional Neural Network for Medical Image Classification', *Int. J. Perceptive Cogn. Comput.*, vol. 10, no. 1, pp. 113–118, Jan. 2024, doi: 10.31436/ijpcc.v10i1.460.
[93] S. Sabour, N. Frosst, and G. E. Hinton, 'Dynamic Routing Between Capsules', Nov. 07, 2017, *arXiv*: arXiv:1710.09829. doi: 10.48550/arXiv.1710.09829.
[94] K. Sabapathi *et al.*, 'Advancing Medical Imaging with Capsule Networks for Diagnostic Accuracy', *Int. J. Comput. Exp. Sci. Eng.*, vol. 11, no. 2, Apr. 2025, doi: 10.22399/ijcesen.1082.
[95] M. Tran, V.-K. Vo-Ho, K. Quinn, H. Nguyen, K. Luu, and N. Le, 'CapsNet for Medical Image Segmentation', Mar. 16, 2022, *arXiv*: arXiv:2203.08948. doi: 10.48550/arXiv.2203.08948.
[96] Y. Huang *et al.*, 'Comparative Analysis of ImageNet Pre-Trained Deep Learning Models and DINOv2 in Medical Imaging Classification', Feb. 13, 2024, *arXiv*: arXiv:2402.07595. doi: 10.48550/arXiv.2402.07595.
[97] G. K. Thakur, A. Thakur, S. Kulkarni, N. Khan, and S. Khan, 'Deep Learning Approaches for Medical Image Analysis and Diagnosis', *Cureus*, vol. 16, no. 5, p. e59507, doi: 10.7759/cureus.59507.
[98] P. K. Mall *et al.*, 'A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities', *Healthc. Anal.*, vol. 4, p. 100216, Dec. 2023, doi: 10.1016/j.health.2023.100216.
[99] D. Müller, I. Soto-Rey, and F. Kramer, 'Towards a guideline for evaluation metrics in medical image segmentation', *BMC Res. Notes*, vol. 15, p. 210, Jun. 2022, doi: 10.1186/s13104-022-06096-y.
[100] O. Shobayo and R. Saatchi, 'Developments in Deep Learning Artificial Neural Network Techniques for Medical Image Analysis and Interpretation', *Diagnostics*, vol. 15, no. 9, p. 1072, Apr. 2025, doi: 10.3390/diagnostics15091072.
[101] C. Patrício, J. C. Neves, and L. F. Teixeira, 'Explainable Deep Learning Methods in Medical Image Classification: A Survey', Sep. 19, 2023, *arXiv*: arXiv:2205.04766. doi: 10.48550/arXiv.2205.04766.
[102] F. Glover, 'Future paths for integer programming and links to artificial intelligence', *Comput. Oper. Res.*, vol. 13, no. 5, pp. 533–549, Jan. 1986, doi: 10.1016/0305-0548(86)90048-1.
[103] C. Reeves, 'Modern heuristic techniques for combinatorial problems', 1993. Accessed: Dec. 10, 2025. [Online]. Available: https://www.semanticscholar.org/paper/Modern-heuristic-techniques-for-combinatorial-Reeves/b649d30268c12b88f6267489bfce5f6c036c44af
[104] A. N. Benaichouche, 'Conception de métaheuristiques d'optimisation pour la segmentation d'images : application aux images IRM du cerveau et aux images de tomographie par émission de positons', Theses, Université Paris-Est, 2014. Accessed: Dec. 10, 2025. [Online]. Available: https://theses.hal.science/tel-01143778
[105] S. Bandyopadhyay and U. Maulik, 'An evolutionary technique based on K-Means algorithm for optimal clustering in RN', *Inf. Sci.*, vol. 146, no. 1, pp. 221–237, Oct. 2002, doi: 10.1016/S0020-0255(02)00208-6.

[106] Siarry Patrick, *Métaheuristiques: recuit simulé, recherche avec tabous, recherche à voisinages variables, méthode GRASP, algorithmes évolutionnaires, fourmis artificielles, essaims particulaires et autres méthodes d'optimisation*. in Algorithmes. Paris: Eyrolles, 2014.
[107] A. Ghoumari, 'Métaheuristiques adaptatives d'optimisation continue basées sur des méthodes d'apprentissage', Theses, Université Paris-Est, 2018. Accessed: Dec. 10, 2025. [Online]. Available: https://theses.hal.science/tel-02085935
[108] G. Dhiman, 'ESA: a hybrid bio-inspired metaheuristic optimization approach for engineering problems', *Eng. Comput.*, vol. 37, no. 1, pp. 323–353, Jan. 2021, doi: 10.1007/s00366-019-00826-w.
[109] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Accessed: Dec. 10, 2025. [Online]. Available: https://direct.mit.edu/books/monograph/2574/Adaptation-in-Natural-and-Artificial-SystemsAn
[110] JohnR. Koza, 'Genetic programming as a means for programming computers by natural selection', *Stat. Comput.*, vol. 4, no. 2, Jun. 1994, doi: 10.1007/BF00175355.
[111] L. J. Fogel, A. J. Owens, and M. J. Walsh, 'Intelligent decision making through a simulation of evolution', *Behav. Sci.*, vol. 11, no. 4, pp. 253–272, 1966, doi: 10.1002/bs.3830110403.
[112] I. Rechenberg and M. Eigen, *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. in Problemata, no. 15. Stuttgart-Bad Cannstadt: Frommann-Holzboog, 1973.
[113] J. Dreo, A. Petrowski, P. Siarry, and E. Taillard, *Métaheuristiques pour l'optimisation difficile*. in Algorithmes. EYROLLES, 2003. Accessed: Dec. 10, 2025. [Online]. Available: https://hal.science/hal-00843020
[114] D. E. Goldberg and J. H. Holland, 'Genetic Algorithms and Machine Learning', *Mach. Learn.*, vol. 3, no. 2, pp. 95–99, Oct. 1988, doi: 10.1023/A:1022602019183.
[115] E. Bonabeau, M. Dorigo, and G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press, 1999. doi: 10.1093/oso/9780195131581.001.0001.
[116] 'Kennedy, J., Eberhart, R.C. and Shi, Y. (2001) Swarm Intelligence. Morgan Kaufmann Publishers, Burlington. - References - Scientific Research Publishing'. Accessed: Dec. 11, 2025. [Online]. Available: https://www.scirp.org/reference/referencespapers?referenceid=1586695
[117] H. A. Abbass, 'MBO: marriage in honey bees optimization-a Haplometrosis polygynous swarming approach', in *Proceedings of the 2001 Congress on Evolutionary Computation*, May 2001, pp. 207–214 vol. 1. doi: 10.1109/CEC.2001.934391.
[118] O. B. Haddad, A. Afshar, and M. A. Mariño, 'Honey-Bees Mating Optimization (HBMO) Algorithm: A New Heuristic Approach for Water Resources Optimization', *Water Resour. Manag.*, vol. 20, no. 5, pp. 661–680, Oct. 2006, doi: 10.1007/s11269-005-9001-3.
[119] C. Yang, J. Chen, and X. Tu, 'Algorithm of Marriage in Honey Bees Optimization Based on the Nelder-Mead Method', presented at the International Conference on Intelligent Systems and Knowledge Engineering 2007, Atlantis Press, Oct. 2007, pp. 886–892. doi: 10.2991/iske.2007.151.
[120] P. Curkovic and B. Jerbic, 'Honey-bees optimization algorithm applied to path planning problem', *Int. J. Simul. Model.*, vol. 6, pp. 154–164, Sep. 2007, doi: 10.2507/IJSIMM06(3)2.087.
[121] T. Sato and M. Hagiwara, 'Bee System: Finding solution by a concentrated search', in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Institute of Electrical and Electronics Engineers Inc., Dec. 1997, pp. 3954–3959. Accessed: Dec. 11, 2025. [Online]. Available: https://keio.elsevierpure.com/en/publications/bee-system-finding-solution-by-a-concentrated-search/
[122] T. Sato and M. Hagiwara, 'Bee System: Finding solution by a concentrated search', in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Institute of Electrical and Electronics Engineers Inc., Dec. 1997, pp. 3954–3959. Accessed: Dec. 11, 2025. [Online]. Available: https://keio.elsevierpure.com/en/publications/bee-system-finding-solution-by-a-concentrated-search/

[123] D. Teodorović and M. Dell'Orco, 'Bee colony optimization - A cooperative learning approach to complex transportation problems', *Adv. AI Methods Transp.*, pp. 51–60, Jan. 2005.
[124] D. Pham, A. Ghanbarzadeh, E. Koç, S. Otri, S. Rahim, and M. Zaidi, 'The Bees Algorithm Technical Note', *Manuf. Eng. Cent. Cardiff Univ. UK*, pp. 1–57, Sep. 2005.
[125] D. Karaboga and B. Basturk, 'On the performance of artificial bee colony (ABC) algorithm', *Appl. Soft Comput.*, vol. 8, no. 1, pp. 687–697, Jan. 2008, doi: 10.1016/j.asoc.2007.05.007.
[126] B. Yuce, M. Packianather, E. Mastrocinque, D. Pham, and A. Lambiase, 'Honey Bees Inspired Optimization Method: The Bees Algorithm', *Insects*, vol. 4, pp. 646–662, Nov. 2013, doi: 10.3390/insects4040646.
[127] C. W. Reynolds, 'Flocks, herds and schools: A distributed behavioral model', *SIGGRAPH Comput Graph*, vol. 21, no. 4, pp. 25–34, août 1987, doi: 10.1145/37402.37406.
[128] A. Cavagna *et al.*, 'Marginal speed confinement resolves the conflict between correlation and control in natural flocks of birds', *Nat. Commun.*, vol. 13, no. 1, p. 2315, May 2022, doi: 10.1038/s41467-022-29883-4.
[129] J. Kennedy and R. Eberhart, 'Particle swarm optimization', in *Proceedings of ICNN'95 - International Conference on Neural Networks*, Nov. 1995, pp. 1942–1948 vol.4. doi: 10.1109/ICNN.1995.488968.
[130] R. Maity, 'Optimization Techniques'.
[131] M. Varan, A. Erduman, and F. Menevşeoğlu, 'A Grey Wolf Optimization Algorithm-Based Optimal Reactive Power Dispatch with Wind-Integrated Power Systems', *Energies*, vol. 16, no. 13, p. 5021, Jan. 2023, doi: 10.3390/en16135021.
[132] S. Mirjalili, S. M. Mirjalili, and A. Lewis, 'Grey Wolf Optimizer', *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014, doi: 10.1016/j.advengsoft.2013.12.007.
[133] A. Rezaee Jordehi and J. Jasni, 'Parameter selection in particle swarm optimisation: a survey', *J. Exp. Theor. Artif. Intell.*, vol. 25, no. 4, pp. 527–542, Dec. 2013, doi: 10.1080/0952813X.2013.782348.
[134] K. R. Das, D. Das, and J. Das, 'Optimal tuning of PID controller using GWO algorithm for speed control in DC motor', in *2015 International Conference on Soft Computing Techniques and Implementations (ICSCTI)*, Oct. 2015, pp. 108–112. doi: 10.1109/ICSCTI.2015.7489575.
[135] J. L. Deneubourg, J. M. Pasteels, and J. C. Verhaeghe, 'Probabilistic behaviour in ants: A strategy of errors?', *J. Theor. Biol.*, vol. 105, no. 2, pp. 259–271, Jan. 1983, doi: 10.1016/S0022-5193(83)80007-1.
[136] M. Dorigo, V. Maniezzo, and A. Colorni, 'DIPARTIMENTO DI ELETTRONICA - POLITECNICO DI MILANO'.
[137] M. Dorigo, V. Maniezzo, and A. Colorni, 'Ant system: optimization by a colony of cooperating agents', *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 26, no. 1, pp. 29–41, Feb. 1996, doi: 10.1109/3477.484436.
[138] S. Goss, S. Aron, J. L. Deneubourg, and J. M. Pasteels, 'Self-organized shortcuts in the Argentine ant', *Naturwissenschaften*, vol. 76, no. 12, pp. 579–581, Dec. 1989, doi: 10.1007/BF00462870.
[139] R. Beckers, J. L. Deneubourg, and S. Goss, 'Trails and U-turns in the selection of a path by the ant Lasius niger', *J. Theor. Biol.*, vol. 159, no. 4, pp. 397–415, 1992.
[140] M. Dorigo, V. Maniezzo, and A. Colorni, 'The Ant System: Optimization by a colony of cooperating agents'.
[141] M. Dorigo and L. M. Gambardella, 'Ant colony system: a cooperative learning approach to the traveling salesman problem', *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 53–66, Apr. 1997, doi: 10.1109/4235.585892.
[142] X.-S. Yang, *Nature-Inspired Metaheuristic Algorithms*. Luniver Press, 2008.
[143] A. Prügel-Bennett, 'Benefits of a Population: Five Mechanisms That Advantage Population-Based Algorithms', *IEEE Trans. Evol. Comput.*, vol. 14, no. 4, pp. 500–517, Aug. 2010, doi: 10.1109/TEVC.2009.2039139.

[144] Z. W. Geem, J. H. Kim, and G. V. Loganathan, 'A New Heuristic Optimization Algorithm: Harmony Search', *SIMULATION*, vol. 76, no. 2, pp. 60–68, Feb. 2001, doi: 10.1177/003754970107600201.
[145] M. F. Dar and A. Ganivada, 'Deep learning and genetic algorithm-based ensemble model for feature selection and classification of breast ultrasound images', *Image Vis. Comput.*, vol. 146, p. 105018, Jun. 2024, doi: 10.1016/j.imavis.2024.105018.
[146] S. Sharma and V. Kumar, 'Application of Genetic Algorithms in Healthcare: A Review', in *Next Generation Healthcare Informatics*, B. K. Tripathy, P. Lingras, A. K. Kar, and C. L. Chowdhary, Eds, Singapore: Springer Nature, 2022, pp. 75–86. doi: 10.1007/978-981-19-2416-3_5.
[147] D. A. Torse *et al.*, 'Optimal feature selection for COVID-19 detection with CT images enabled by metaheuristic optimization and artificial intelligence', *Multimed. Tools Appl.*, vol. 82, no. 26, pp. 41073–41103, Nov. 2023, doi: 10.1007/s11042-023-15031-7.
[148] N. A. Al-Najdawi, A. F. Al-Shawabkeh, S. Tedmori, I. I. Ikhries, and O. Dorgham, 'Comprehensive evaluation of optimization algorithms for medical image segmentation', *Sci. Rep.*, vol. 15, no. 1, p. 37190, Oct. 2025, doi: 10.1038/s41598-025-14261-z.
[149] S. Chakraborty, A. K. Saha, A. E. Ezugwu, J. O. Agushaka, R. A. Zitar, and L. Abualigah, 'Differential Evolution and Its Applications in Image Processing Problems: A Comprehensive Review', *Arch. Comput. Methods Eng.*, vol. 30, no. 2, pp. 985–1040, Mar. 2023, doi: 10.1007/s11831-022-09825-5.
[150] T. Egling, 'Differential evolution optimization algorithms and its application in machine learning based disease detection'.
[151] I. Farda, A. Thammano, and J. Morris, 'Differential Evolution With Self-Adaptive Mutation and Population Improvement Strategy for Optimization Problems', *IEEE Access*, vol. 12, pp. 131809–131829, 2024, doi: 10.1109/ACCESS.2024.3460385.
[152] Y.-G. Chen, Y. Cao, K. Lu, Q. Yang, Y. Chen, and Y. Ping, 'Differential evolution with classified mutation for parameter extraction of photovoltaic models', *PLOS ONE*, vol. 20, no. 10, p. e0332083, Oct. 2025, doi: 10.1371/journal.pone.0332083.
[153] D. Chauhan, Shivani, D. Jung, and A. Yadav, 'Advancements in Multimodal Differential Evolution: A Comprehensive Review and Future Perspectives', *Artif. Intell. Rev.*, vol. 58, no. 11, p. 335, Aug. 2025, doi: 10.1007/s10462-025-11314-7.
[154] M. Ramadas and A. Abraham, 'Segmentation on remote sensing imagery for atmospheric air pollution using divergent differential evolution algorithm', *Neural Comput. Appl.*, vol. 35, no. 5, pp. 3977–3990, Feb. 2023, doi: 10.1007/s00521-022-07922-x.
[155] A. Thakare, A. M. Anter, and A. Abraham, 'Seizure disorders recognition model from EEG signals using new probabilistic particle swarm optimizer and sequential differential evolution', *Multidimens. Syst. Signal Process.*, vol. 34, no. 2, pp. 397–421, Jun. 2023, doi: 10.1007/s11045-023-00870-2.
[156] S. Saifullah and R. Dreżewski, 'Advanced Medical Image Segmentation Enhancement: A Particle-Swarm-Optimization-Based Histogram Equalization Approach', *Appl. Sci.*, vol. 14, no. 2, p. 923, Jan. 2024, doi: 10.3390/app14020923.
[157] K. Krämer, S. Müller, and M. Kosterhon, 'Deep Learning-Tuned Adaptive Inertia Weight in Particle Swarm Optimization for Medical Image Registration':, in *Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, Porto, Portugal: SCITEPRESS - Science and Technology Publications, 2025, pp. 307–318. doi: 10.5220/0013122000003912.
[158] J. Ma and J. Hu, 'An improved particle swarm optimization for multilevel thresholding medical image segmentation', *PLOS ONE*, vol. 19, no. 12, p. e0306283, déc 2024, doi: 10.1371/journal.pone.0306283.
[159] S. El Amoury, Y. Smili, and Y. Fakhri, 'Design of an Optimal Convolutional Neural Network Architecture for MRI Brain Tumor Classification by Exploiting Particle Swarm Optimization', *J. Imaging*, vol. 11, no. 2, p. 31, Feb. 2025, doi: 10.3390/jimaging11020031.

[160] I. I. Ekanem, A. E. Ekanem, and E. S. Abia, 'A Systemic Review on the Adoption of Particle Swarm Optimization Algorithms in Biomedical Engineering Diagnostics and Simulations', *Ann. Healthc. Syst. Eng.*, vol. 2, no. 1, pp. 1–15, Jan. 2025, doi: 10.22105/ahse.v2i1.25.
[161] 'Medical image analysis using swarm intelligence: A survey', in *Recent Trends in Swarm Intelligence Enabled Research for Engineering Applications*, Academic Press, 2024, pp. 89–130. doi: 10.1016/B978-0-443-15533-8.00012-6.
[162] M. Thangamani, S. Satheesh, R. Lingisetty, S. Rajendran, and B. D. Shivahare, 'Mathematical Model for Swarm Optimization in Multimodal Biomedical Images', in *Swarm Optimization for Biomedical Applications*, CRC Press, 2025.
[163] A. A. Shaban and H. M. Yasin, 'Applications of the artificial bee colony algorithm in medical imaging and diagnostics: a review', *Int. J. Sci. World*, vol. 11, no. 1, pp. 21–29, Feb. 2025, doi: 10.14419/yszxm607.
[164] M. A. Tawil and O. Dakkak, 'edmABC: an improved artificial bee colony algorithm on detecting breast cancer for mammography images', *Evol. Syst.*, vol. 16, no. 2, p. 42, Feb. 2025, doi: 10.1007/s12530-025-09666-0.
[165] D. T. Pham, M. Castellani, and L. Baronti, 'Nature-Inspired Optimisation with the Bees Algorithm', in *Intelligent Optimisation with the Bees Algorithm: Concepts and Applications*, D. T. Pham, M. Castellani, and L. Baronti, Eds, Cham: Springer Nature Switzerland, 2025, pp. 23–59. doi: 10.1007/978-3-031-87286-0_2.
[166] M. Ali, M. Danyal, T. Riaz, L. Ullah, and S. Ullah, *ABCNN: A Hybrid Artificial Bee Colony Neural Network for Robust Classification*. 2025.
[167] S. Kolli and B. R. Parvathala, 'A Novel Assessment of Lung Cancer Classification System Using Binary Grasshopper with Artificial Bee Optimisation Algorithm with Double Deep Neural Network Classifier', *J. Inst. Eng. India Ser. B*, vol. 105, no. 5, pp. 1129–1143, Oct. 2024, doi: 10.1007/s40031-024-01027-w.
[168] H. N. Fakhouri, F. Hamad, and A. Al Hwaitat, 'SVM Hyperparameter Tuning Using Grey Wolf Optimizer for Heart Disease Prediction', in *2025 1st International Conference on Computational Intelligence Approaches and Applications (ICCIAA)*, Apr. 2025, pp. 1–8. doi: 10.1109/ICCIAA65327.2025.11012972.
[169] E. Hassan, A. Saber, S. El-Sappagh, and N. El-Rashidy, 'Optimized ensemble deep learning approach for accurate breast cancer diagnosis using transfer learning and grey wolf optimization', *Evol. Syst.*, vol. 16, no. 2, p. 59, Apr. 2025, doi: 10.1007/s12530-025-09686-w.
[170] E. I. Muryadi, I. Futri, and D. C. E. Saputra, 'iGWO-RF: an Improved Grey Wolfed Optimization for Random Forest Hyperparameter Optimization to Identification Breast Cancer', *Breast Cancer*, vol. 10, no. 4, 2024.
[171] T. K. Abuya, W. C. Waithera, and C. W. Kipruto, 'Augmented Lung Cancer Prediction: Leveraging Convolutional Neural Networks and Grey Wolf Optimization Algorithm', *Open Access Libr. J.*, vol. 11, no. 4, pp. 1–25, Apr. 2024, doi: 10.4236/oalib.1111172.
[172] J. Lee, Y. Yoon, J. Kim, and Y.-H. Kim, 'Metaheuristic-Based Feature Selection Methods for Diagnosing Sarcopenia with Machine Learning Algorithms', *Biomimetics*, vol. 9, no. 3, p. 179, Mar. 2024, doi: 10.3390/biomimetics9030179.
[173] D. Khafaga *et al.*, 'Meta-heuristics for Feature Selection and Classification in Diagnostic Breast Cancer', *Comput. Mater. Contin.*, vol. 73, no. 1, pp. 749–765, 2022, doi: 10.32604/cmc.2022.029605.
[174] L. F. Rodrigues, A. R. Backes, B. A. N. Travençolo, and G. M. B. de Oliveira, 'Optimizing a Deep Residual Neural Network with Genetic Algorithm for Acute Lymphoblastic Leukemia Classification', *J. Digit. Imaging*, vol. 35, no. 3, pp. 623–637, Jun. 2022, doi: 10.1007/s10278-022-00600-3.
[175] L. F. Rodrigues, A. R. Backes, B. A. N. Travençolo, and G. M. B. de Oliveira, 'Optimizing a Deep Residual Neural Network with Genetic Algorithm for Acute Lymphoblastic Leukemia Classification', *J. Digit. Imaging*, vol. 35, no. 3, pp. 623–637, Jun. 2022, doi: 10.1007/s10278-022-00600-3.

[176] C. E. Ogbuanya, A. Obayi, S. Larabi-Marie-Sainte, A. O. Saad, and L. Berriche, 'A hybrid optimization approach for accelerated multimodal medical image fusion', *PLOS One*, vol. 20, no. 7, p. e0324973, Jul. 2025, doi: 10.1371/journal.pone.0324973.
[177] C.-L. Lin, A. Mimori, and Y.-W. Chen, 'Hybrid Particle Swarm Optimization and Its Application to Multimodal 3D Medical Image Registration', *Comput. Intell. Neurosci.*, vol. 2012, p. 561406, 2012, doi: 10.1155/2012/561406.
[178] A. Yonar, 'A swarm intelligence-driven hybrid framework for brain tumor classification with enhanced deep features', *Sci. Rep.*, vol. 15, p. 37543, Oct. 2025, doi: 10.1038/s41598-025-23820-3.
[179] T. Bai *et al.*, 'Exploration and comparison of the effectiveness of swarm intelligence algorithm in early identification of cardiovascular disease', *Sci. Rep.*, vol. 15, no. 1, p. 4647, Feb. 2025, doi: 10.1038/s41598-025-87598-0.
[180] Q. S. Hamad, H. Samma, and S. A. Suandi, 'Optimization of Convolutional Neural Network Hyperparameter for Medical Image Diagnosis using Metaheuristic Algorithms: A short Recent Review (2019-2022)', Dec. 23, 2024, *arXiv*: arXiv:2412.17956. doi: 10.48550/arXiv.2412.17956.
[181] Y. A. Kadhim, M. S. Guzel, and A. Mishra, 'A Novel Hybrid Machine Learning-Based System Using Deep Learning Techniques and Meta-Heuristic Algorithms for Various Medical Datatypes Classification', *Diagnostics*, vol. 14, no. 14, p. 1469, Jan. 2024, doi: 10.3390/diagnostics14141469.
[182] R. N. Ravikumar, S. Aarthi, S. Kurbanova, S. Polvanov, B. Matchanova, and K. Sathya, 'Hybrid Metaheuristic Optimization for Neural Networks in Biomedical Imaging', in *Metaheuristic Algorithms and Optimizing Neural Networks for Biomedical Image Processing*, IGI Global Scientific Publishing, 2026, pp. 197–234. doi: 10.4018/979-8-3373-0523-3.ch008.
[183] M. K. Bohmrah and H. Kaur, 'Advanced Hybridization and Optimization of DNNs for Medical Imaging: A Survey on Disease Detection Techniques', *Artif. Intell. Rev.*, vol. 58, no. 4, p. 122, Feb. 2025, doi: 10.1007/s10462-024-11049-x.
[184] R. Shetty, V. S. Bhat, and J. Pujari, 'Content-based medical image retrieval using deep learning-based features and hybrid meta-heuristic optimization', *Biomed. Signal Process. Control*, vol. 92, p. 106069, Jun. 2024, doi: 10.1016/j.bspc.2024.106069.
[185] K. V. Singh, A. Singh, H. Kaur, and B. Moharana, 'Applications of nature-inspired metaheuristic algorithms for medical image analysis', in *Nature-inspired Metaheuristic Algorithms*, CRC Press, 2025.
[186] S. Awotwe, A. T. Dufera, and W. Yi, 'Recent advancement of metaheuristic optimization algorithms-based learning for breast cancer diagnosis: a review', *Memetic Comput.*, vol. 17, no. 3, p. 31, Jun. 2025, doi: 10.1007/s12293-025-00467-1.
[187] N. Mohamed, R. L. Almutairi, S. Abdelrahim, R. Alharbi, F. M. Alhomayani, and A. A. Elhag, 'Towards precise chronic disease management: A combined approach with binary metaheuristics and ensemble deep learning', *J. Radiat. Res. Appl. Sci.*, vol. 17, no. 4, p. 101092, Dec. 2024, doi: 10.1016/j.jrras.2024.101092.
[188] J. Lee, Y. Yoon, J. Kim, and Y.-H. Kim, 'Metaheuristic-Based Feature Selection Methods for Diagnosing Sarcopenia with Machine Learning Algorithms', *Biomimetics*, vol. 9, no. 3, p. 179, Mar. 2024, doi: 10.3390/biomimetics9030179.
[189] M. M. Ahsan, S. A. Luna, and Z. Siddique, 'Machine-Learning-Based Disease Diagnosis: A Comprehensive Review', *Healthcare*, vol. 10, no. 3, p. 541, Mar. 2022, doi: 10.3390/healthcare10030541.
[190] J. H. Moore, N. Raghavachari, and Workshop Speakers, 'Artificial Intelligence Based Approaches to Identify Molecular Determinants of Exceptional Health and Life Span-An Interdisciplinary Workshop at the National Institute on Aging', *Front. Artif. Intell.*, vol. 2, p. 12, Aug. 2019, doi: 10.3389/frai.2019.00012.
[191] W. S. Mcculloch and W. Pitts, 'A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY'.

[192] F. Rosenblatt, 'The perceptron: A probabilistic model for information storage and organization in the brain.', *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, 1958, doi: 10.1037/h0042519.
[193] H. D. Block, 'A review of "perceptrons: An introduction to computational geometry≒', *Inf. Control*, vol. 17, no. 5, pp. 501–522, Dec. 1970, doi: 10.1016/S0019-9958(70)90409-2.
[194] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, 'Learning representations by backpropagating errors', 1986.
[195] C. Leo, 'The Math Behind Stochastic Gradient Descent', Towards Data Science. Accessed: Oct. 09, 2025. [Online]. Available: https://towardsdatascience.com/stochastic-gradient-descent-math-and-python-code-35b5e66d6f79/
[196] 'Gradient Descent vs. Mini-Batch Gradient Descent vs. Stochastic Gradient Descent: An Expert Comparison - LUNARTECH'. Accessed: Oct. 09, 2025. [Online]. Available: https://www.lunartech.ai/blog/gradient-descent-vs-mini-batch-gradient-descent-vs-stochastic-gradient-descent-an-expert-comparison
[197] S. Ruder, 'An overview of gradient descent optimization algorithms', Jun. 15, 2017, *arXiv*: arXiv:1609.04747. doi: 10.48550/arXiv.1609.04747.
[198] G. Hinton, 'Neural Networks for Machine Learning'.
[199] G. E. Hinton, S. Osindero, and Y.-W. Teh, 'A Fast Learning Algorithm for Deep Belief Nets', *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006, doi: 10.1162/neco.2006.18.7.1527.
[200] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, 'Greedy layer-wise training of deep networks', in *Proceedings of the 20th International Conference on Neural Information Processing Systems*, in NIPS'06. Cambridge, MA, USA: MIT Press, décembre 2006, pp. 153–160.
[201] S.-C. Huang, A. Pareek, M. Jensen, M. P. Lungren, S. Yeung, and A. S. Chaudhari, 'Self-supervised learning for medical image classification: a systematic review and implementation guidelines', *Npj Digit. Med.*, vol. 6, no. 1, p. 74, Apr. 2023, doi: 10.1038/s41746-023-00811-0.
[202] K. P. Murphy, *Machine Learning - A Probabilistic Perspective*. in Adaptive Computation and Machine Learning. Cambridge: MIT Press, 2014.
[203] J. E. van Engelen and H. H. Hoos, 'A survey on semi-supervised learning', *Mach. Learn.*, vol. 109, no. 2, pp. 373–440, Feb. 2020, doi: 10.1007/s10994-019-05855-6.
[204] N. K. Logothetis, 'The ins and outs of fMRI signals', *Nat. Neurosci.*, vol. 10, no. 10, pp. 1230–1232, Oct. 2007, doi: 10.1038/nn1007-1230.
[205] P. Ravindran, A. Costa, R. Soares, and A. C. Wiedenhoeft, 'Classification of CITES-listed and other neotropical Meliaceae wood images using convolutional neural networks', *Plant Methods*, vol. 14, p. 25, 2018, doi: 10.1186/s13007-018-0292-9.
[206] O. Ronneberger, P. Fischer, and T. Brox, 'U-Net: Convolutional Networks for Biomedical Image Segmentation', in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds, in Lecture Notes in Computer Science, vol. 9351. , Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.
[207] Z. Xian, R. Huang, D. Towey, and C. Yue, 'Convolutional Neural Network Image Classification Based on Different Color Spaces', *Tsinghua Sci. Technol.*, vol. 30, no. 1, pp. 402–417, Feb. 2025, doi: 10.26599/TST.2024.9010001.
[208] B. A. Skourt, N. S. Nikolov, and A. Majda, 'Feature-Extraction Methods for Lung-Nodule Detection: 3rd International Conference on Intelligent Systems and Advanced Computing Sciences, ISACS 2019', *Proc. - 2019 Int. Conf. Intell. Syst. Adv. Comput. Sci. ISACS 2019*, Dec. 2019, doi: 10.1109/ISACS48493.2019.9068871.
[209] S. Ioffe and C. Szegedy, 'Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift', Mar. 02, 2015, *arXiv*: arXiv:1502.03167. doi: 10.48550/arXiv.1502.03167.
[210] 'Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position - PubMed'. Accessed: Oct. 06, 2025. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/7370364/

[211] K. Fukushima, 'Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position', *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, 1980, doi: 10.1007/BF00344251.
[212] Y. El-Shamayleh, R. D. Kumbhani, N. T. Dhruv, and J. A. Movshon, 'Visual response properties of V1 neurons projecting to V2 in macaque', *J. Neurosci. Off. J. Soc. Neurosci.*, vol. 33, no. 42, pp. 16594–16605, Oct. 2013, doi: 10.1523/JNEUROSCI.2753-13.2013.
[213] 'MNIST handwritten digit database, Yann LeCun, Corinna Cortes and Chris Burges'. Accessed: Oct. 06, 2025. [Online]. Available: https://yann.lecun.org/exdb/mnist/
[214] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, 'Squeeze-and-Excitation Networks', May 16, 2019, *arXiv*: arXiv:1709.01507. doi: 10.48550/arXiv.1709.01507.
[215] C. Szegedy *et al.*, 'Going Deeper with Convolutions', Sep. 17, 2014, *arXiv*: arXiv:1409.4842. doi: 10.48550/arXiv.1409.4842.
[216] A. Krizhevsky, I. Sutskever, and G. E. Hinton, 'ImageNet classification with deep convolutional neural networks', *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
[217] M. Lin, Q. Chen, and S. Yan, 'Network In Network', Mar. 04, 2014, *arXiv*: arXiv:1312.4400. doi: 10.48550/arXiv.1312.4400.
[218] J. Jin, A. Dundar, and E. Culurciello, 'Flattened Convolutional Neural Networks for Feedforward Acceleration', Nov. 20, 2015, *arXiv*: arXiv:1412.5474. doi: 10.48550/arXiv.1412.5474.
[219] P. Ramachandran, B. Zoph, and Q. V. Le, 'Searching for Activation Functions', Oct. 27, 2017, *arXiv*: arXiv:1710.05941. doi: 10.48550/arXiv.1710.05941.
[220] V. Nair and G. E. Hinton, 'Rectified Linear Units Improve Restricted Boltzmann Machines'.
[221] S. Narayan, 'The generalized sigmoid activation function: Competitive supervised learning', *Inf. Sci.*, vol. 99, no. 1, pp. 69–82, Jun. 1997, doi: 10.1016/S0020-0255(96)00200-9.
[222] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, 'Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark', Jun. 28, 2022, *arXiv*: arXiv:2109.14545. doi: 10.48550/arXiv.2109.14545.
[223] K. Xu *et al.*, 'Show, Attend and Tell: Neural Image Caption Generation with Visual Attention', Apr. 19, 2016, *arXiv*: arXiv:1502.03044. doi: 10.48550/arXiv.1502.03044.
[224] L. Zhao and Z. Zhang, 'A improved pooling method for convolutional neural networks', *Sci. Rep.*, vol. 14, no. 1, p. 1589, Jan. 2024, doi: 10.1038/s41598-024-51258-6.
[225] Y. Yuan, M. Chao, and Y.-C. Lo, 'Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks With Jaccard Distance', *IEEE Trans. Med. Imaging*, vol. 36, no. 9, pp. 1876–1886, Sep. 2017, doi: 10.1109/TMI.2017.2695227.
[226] Y.-L. Boureau, J. Ponce, and Y. LeCun, 'A Theoretical Analysis of Feature Pooling in Visual Recognition'.
[227] M. Ranzato, Y.-L. Boureau, and Y. LeCun, 'Sparse Feature Learning for Deep Belief Networks'.
[228] 'Mixed Pooling for Convolutional Neural Networks | SpringerLink'. Accessed: Oct. 10, 2025. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-11740-9_34
[229] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, 'Improving neural networks by preventing co-adaptation of feature detectors', Jul. 03, 2012, *arXiv*: arXiv:1207.0580. doi: 10.48550/arXiv.1207.0580.
[230] 'Yu, D., et al. Mixed pooling for convolutional neural networks. in International conference on rough sets and knowledge technology. 2014. Springer.'.
[231] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, 'Automatic Brain Tumor Segmentation using Cascaded Anisotropic Convolutional Neural Networks', vol. 10670, 2018, pp. 178–190. doi: 10.1007/978-3-319-75238-9_16.
[232] L. Wan, M. Zeiler, S. Zhang, Y. LeCun, and R. Fergus, 'Regularization of neural networks using dropconnect', in *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, in ICML'13. Atlanta, GA, USA: JMLR.org, juin 2013, p. III-1058-III–1066.
[233] X. Glorot and Y. Bengio, 'Understanding the difficulty of training deep feedforward neural networks'.

[234] S. Ruder, 'An overview of gradient descent optimization algorithms', Jun. 15, 2017, *arXiv*: arXiv:1609.04747. doi: 10.48550/arXiv.1609.04747.
[235] D. P. Kingma and J. Ba, 'Adam: A Method for Stochastic Optimization', Jan. 30, 2017, *arXiv*: arXiv:1412.6980. doi: 10.48550/arXiv.1412.6980.
[236] R. Ward, 'Stochastic gradient descent: where optimization meets machine learning | EMS Press'. Accessed: Oct. 10, 2025. [Online]. Available: https://ems.press/books/standalone/279/5594
[237] D. Silver *et al.*, 'Mastering the game of Go with deep neural networks and tree search', *Nature*, vol. 529, no. 7587, pp. 484–489, Jan. 2016, doi: 10.1038/nature16961.
[238] M. Mohaimenuzzaman, Z. S. Abdallah, J. Kamruzzaman, and B. Srinivasan, 'Effect of Hyper-Parameter Optimization on the Deep Learning Model Proposed for Distributed Attack Detection in Internet of Things Environment', Jun. 19, 2018, *arXiv*: arXiv:1806.07057. doi: 10.48550/arXiv.1806.07057.
[239] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, 'Dropout: A Simple Way to Prevent Neural Networks from Overfitting', *J. Mach. Learn. Res.*, vol. 15, no. 56, pp. 1929–1958, 2014.
[240] '[1806.07057] Effect of Hyper-Parameter Optimization on the Deep Learning Model Proposed for Distributed Attack Detection in Internet of Things Environment'. Accessed: Oct. 10, 2025. [Online]. Available: https://arxiv.org/abs/1806.07057
[241] A. Vaswani *et al.*, 'Attention Is All You Need', Aug. 02, 2023, *arXiv*: arXiv:1706.03762. doi: 10.48550/arXiv.1706.03762.
[242] J. Chen, Z. Liang, and X. Lu, 'A dual attention and cross layer fusion network with a hybrid CNN and transformer architecture for medical image segmentation', *Sci. Rep.*, vol. 15, no. 1, p. 35707, Oct. 2025, doi: 10.1038/s41598-025-19563-w.
[243] A. Dosovitskiy *et al.*, 'An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale', Jun. 03, 2021, *arXiv*: arXiv:2010.11929. doi: 10.48550/arXiv.2010.11929.
[244] S. Shah, D. Dey, C. Lovett, and A. Kapoor, 'AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles', Jul. 18, 2017, *arXiv*: arXiv:1705.05065. doi: 10.48550/arXiv.1705.05065.
[245] I. Sutskever, O. Vinyals, and Q. V. Le, 'Sequence to Sequence Learning with Neural Networks', Dec. 14, 2014, *arXiv*: arXiv:1409.3215. doi: 10.48550/arXiv.1409.3215.
[246] M.-T. Luong, H. Pham, and C. D. Manning, 'Effective Approaches to Attention-based Neural Machine Translation', Sep. 20, 2015, *arXiv*: arXiv:1508.04025. doi: 10.48550/arXiv.1508.04025.
[247] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. in Adaptive computation and machine learning. Cambridge, Massachusetts: The MIT Press, 2016.
[248] M. Freitag and Y. Al-Onaizan, 'Beam Search Strategies for Neural Machine Translation', in *Proceedings of the First Workshop on Neural Machine Translation*, 2017, pp. 56–60. doi: 10.18653/v1/W17-3207.
[249] H. Zhou *et al.*, 'Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting', Mar. 28, 2021, *arXiv*: arXiv:2012.07436. doi: 10.48550/arXiv.2012.07436.
[250] C.-Z. A. Huang *et al.*, 'Music Transformer', Dec. 12, 2018, *arXiv*: arXiv:1809.04281. doi: 10.48550/arXiv.1809.04281.
[251] B. Lim, S. O. Arik, N. Loeff, and T. Pfister, 'Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting', Sep. 27, 2020, *arXiv*: arXiv:1912.09363. doi: 10.48550/arXiv.1912.09363.
[252] T. Mikolov, K. Chen, G. Corrado, and J. Dean, 'Efficient Estimation of Word Representations in Vector Space', Sep. 07, 2013, *arXiv*: arXiv:1301.3781. doi: 10.48550/arXiv.1301.3781.
[253] S. M. Kazemi *et al.*, 'Time2Vec: Learning a Vector Representation of Time', Jul. 11, 2019, *arXiv*: arXiv:1907.05321. doi: 10.48550/arXiv.1907.05321.
[254] 'Details for: Basic Science of PET Imaging / › Malawi College of Health Sciences - LL Campus OPAC catalog'. Accessed: Oct. 06, 2025. [Online]. Available:

https://mchs.bestbookbuddies.com/cgi-bin/koha/opac-detail.pl?biblionumber=11836&query_desc=Provider%3ASpringer%2C
[255] L. Lévêque, M. Outtas, H. Liu, and L. Zhang, 'Comparative study of the methodologies used for subjective medical image quality assessment', *Phys. Med. Biol.*, vol. 66, no. 15, Jul. 2021, doi: 10.1088/1361-6560/ac1157.
[256] M. Winkels and T. S. Cohen, '3D G-CNNs for Pulmonary Nodule Detection', Apr. 12, 2018, *arXiv*: arXiv:1804.04656. doi: 10.48550/arXiv.1804.04656.
[257] D. Zeng, C. Noteboom, K. Sutrave, and R. Godasu, 'A Meta-Analysis of Evolution of Deep Learning Research in Medical Image Analysis'.
[258] M. Avanzo, J. Stancanello, G. Pirrone, A. Drigo, and A. Retico, 'The Evolution of Artificial Intelligence in Medical Imaging: From Computer Science to Machine and Deep Learning', *Cancers*, vol. 16, no. 21, p. 3702, Nov. 2024, doi: 10.3390/cancers16213702.
[259] S. Kumari and P. Singh, 'Data efficient deep learning for medical image analysis: A survey', Oct. 10, 2023, *arXiv*: arXiv:2310.06557. doi: 10.48550/arXiv.2310.06557.
[260] '[2310.06557] Data efficient deep learning for medical image analysis: A survey'. Accessed: Oct. 10, 2025. [Online]. Available: https://arxiv.org/abs/2310.06557
[261] M. Avanzo, J. Stancanello, G. Pirrone, A. Drigo, and A. Retico, 'The Evolution of Artificial Intelligence in Medical Imaging: From Computer Science to Machine and Deep Learning', *Cancers*, vol. 16, no. 21, p. 3702, Nov. 2024, doi: 10.3390/cancers16213702.
[262] A. D A, M. S. Shekar, A. Bharadwaj, N. Vineeth, and M. L. Neelima, 'Deep Learning in Medical Image Analysis: A Survey', in *2024 International Conference on Innovation and Novelty in Engineering and Technology (INNOVA)*, Dec. 2024, pp. 1–5. doi: 10.1109/INNOVA63080.2024.10847040.
[263] N. Sengodan, 'Breast Cancer Histopathology Classification using CBAM-EfficientNetV2 with Transfer Learning', May 13, 2025, *arXiv*: arXiv:2410.22392. doi: 10.48550/arXiv.2410.22392.
[264] W. Li *et al.*, 'Path R-CNN for Prostate Cancer Diagnosis and Gleason Grading of Histological Images', *IEEE Trans. Med. Imaging*, vol. 38, no. 4, pp. 945–954, Apr. 2019, doi: 10.1109/TMI.2018.2875868.
[265] R. Girshick, 'Fast R-CNN', Sep. 27, 2015, *arXiv*: arXiv:1504.08083. doi: 10.48550/arXiv.1504.08083.
[266] S. Ren, K. He, R. Girshick, and J. Sun, 'Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks', Jan. 06, 2016, *arXiv*: arXiv:1506.01497. doi: 10.48550/arXiv.1506.01497.
[267] J. Redmon and A. Farhadi, 'YOLO9000: Better, Faster, Stronger', Dec. 25, 2016, *arXiv*: arXiv:1612.08242. doi: 10.48550/arXiv.1612.08242.
[268] J. Redmon and A. Farhadi, 'YOLOv3: An Incremental Improvement', Apr. 08, 2018, *arXiv*: arXiv:1804.02767. doi: 10.48550/arXiv.1804.02767.
[269] A. Lemay, 'Kidney Recognition in CT Using YOLOv3', Oct. 03, 2019, *arXiv*: arXiv:1910.01268. doi: 10.48550/arXiv.1910.01268.
[270] J. Antony, K. McGuinness, N. E. O. Connor, and K. Moran, 'Quantifying Radiographic Knee Osteoarthritis Severity using Deep Convolutional Neural Networks', Sep. 08, 2016, *arXiv*: arXiv:1609.02469. doi: 10.48550/arXiv.1609.02469.
[271] B. A. Skourt, 'Medical Image Analysis Using Deep Learning'.
[272] L. Shen and L. Bai, 'A review on Gabor wavelets for face recognition', *Pattern Anal Appl*, vol. 9, no. 2–3, pp. 273–292, Oct. 2006.
[273] K.-L. Ng, J. Yazer, M. Abdolell, and P. Brown, 'National survey to identify subspecialties at risk for physician shortages in Canadian academic radiology departments', *Can. Assoc. Radiol. J. J. Assoc. Can. Radiol.*, vol. 61, no. 5, pp. 252–257, Dec. 2010, doi: 10.1016/j.carj.2010.02.007.
[274] 'National survey to identify subspecialties at risk for physician shortages in Canadian academic radiology departments - PubMed'. Accessed: Oct. 06, 2025. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/20382499/

[275] C. F. Baumgartner, L. M. Koch, M. Pollefeys, and E. Konukoglu, 'An Exploration of 2D and 3D Deep Learning Techniques for Cardiac MR Image Segmentation', Oct. 10, 2017, *arXiv*: arXiv:1709.04496. doi: 10.48550/arXiv.1709.04496.

[276] S.-H. Wang, Y.-D. Lv, Y. Sui, S. Liu, S.-J. Wang, and Y.-D. Zhang, 'Alcoholism Detection by Data Augmentation and Convolutional Neural Network with Stochastic Pooling', *J. Med. Syst.*, vol. 42, no. 1, p. 2, Nov. 2017, doi: 10.1007/s10916-017-0845-x.

[277] 'Alcoholism Detection by Data Augmentation and Convolutional Neural Network with Stochastic Pooling - PubMed'. Accessed: Oct. 06, 2025. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29159706/

[278] M. Havaei *et al.*, 'Brain tumor segmentation with Deep Neural Networks', *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017, doi: 10.1016/j.media.2016.05.004.

[279] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, and P.-A. Heng, '3D Deeply Supervised Network for Automatic Liver Segmentation from CT Volumes', Jul. 03, 2016, *arXiv*: arXiv:1607.00582. doi: 10.48550/arXiv.1607.00582.

[280] O. Ronneberger, P. Fischer, and T. Brox, 'U-Net: Convolutional Networks for Biomedical Image Segmentation', in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds, in Lecture Notes in Computer Science, vol. 9351. , Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[281] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, 'Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks', *ArXiv170503820 Cs*, Jun. 2017, Accessed: May 16, 2021. [Online]. Available: http://arxiv.org/abs/1705.03820

[282] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, 'KNN Model-Based Approach in Classification', in *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, vol. 2888, R. Meersman, Z. Tari, and D. C. Schmidt, Eds, in Lecture Notes in Computer Science, vol. 2888. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 986–996. doi: 10.1007/978-3-540-39964-3_62.

[283] O. Ronneberger, P. Fischer, and T. Brox, 'U-Net: Convolutional Networks for Biomedical Image Segmentation', May 18, 2015, *arXiv*: arXiv:1505.04597. doi: 10.48550/arXiv.1505.04597.

[284] J. S. Suri *et al.*, 'UNet Deep Learning Architecture for Segmentation of Vascular and Non-Vascular Images: A Microscopic Look at UNet Components Buffered With Pruning, Explainable Artificial Intelligence, and Bias', *IEEE Access*, vol. 11, pp. 595–645, 2023, doi: 10.1109/ACCESS.2022.3232561.

[285] D. Wu, Y. Wang, S.-T. Xia, J. Bailey, and X. Ma, 'Skip Connections Matter: On the Transferability of Adversarial Examples Generated with ResNets', Feb. 14, 2020, *arXiv*: arXiv:2002.05990. doi: 10.48550/arXiv.2002.05990.

[286] D. M. Bayram and D. A. Seçer, '9th (Online) International Conference on Applied Analysis and Mathematical Modeling (ICAAMM21) June 11-13, 2021, Istanbul-Turkey Abstracts Book', 2021.

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

[288] H. Ali, N. Shifa, R. Benlamri, A. A. Farooque, and R. Yaqub, 'A fine tuned EfficientNet-B0 convolutional neural network for accurate and efficient classification of apple leaf diseases', *Sci. Rep.*, vol. 15, no. 1, p. 25732, Jul. 2025, doi: 10.1038/s41598-025-04479-2.

[289] Z. Hameed, S. Zahia, B. Garcia-Zapirain, J. Javier Aguirre, and A. María Vanegas, 'Breast Cancer Histopathology Image Classification Using an Ensemble of Deep Learning Models', *Sensors*, vol. 20, no. 16, p. 4373, Aug. 2020, doi: 10.3390/s20164373.

[290] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, 'A Dataset for Breast Cancer Histopathological Image Classification', *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, Jul. 2016, doi: 10.1109/TBME.2015.2496264.

[291] P. Ramamoorthy, B. R. Ramakantha Reddy, S. S. Askar, and M. Abouhawwash, 'Histopathology-based breast cancer prediction using deep learning methods for healthcare applications', *Front. Oncol.*, vol. 14, p. 1300997, Jun. 2024, doi: 10.3389/fonc.2024.1300997.
[292] 'Breast Cancer Histopathological Database (BreakHis) – Laboratório Visão Robótica e Imagem'. Accessed: Aug. 22, 2025. [Online]. Available: https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/
[293] Y. Benhammou, B. Achchab, F. Herrera, and S. Tabik, 'BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights', *Neurocomputing*, vol. 375, pp. 9–24, Jan. 2020, doi: 10.1016/j.neucom.2019.09.044.
[294] M. Tafavvoghi, L. A. Bongo, N. Shvetsov, L.-T. R. Busund, and K. Møllersen, 'Publicly available datasets of breast histopathology H&E whole-slide images: A scoping review', *J. Pathol. Inform.*, vol. 15, p. 100363, Feb. 2024, doi: 10.1016/j.jpi.2024.100363.
[295] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, 'A Dataset for Breast Cancer Histopathological Image Classification', *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, Jul. 2016, doi: 10.1109/TBME.2015.2496264.