PEOPLE'S DEMOCTRATIC REPUBLIC OF ALGERIA MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

UNIVERSITY KASSDI MERBAH - OUARGLA
FACULTY OF NEW TECHNOLOGIES OF INFORMATION AND COMMUNICATION
DEPARTEMENT OF ELECTRONIC AND COMMUNICATION

THESIS

OF END OF STUDIES FOR OBTAINING THE MASTER'S DEGREE IN AUTOMATION AND SYSTEMS

TITLE

Integrated Machine Learning Techniques for Enhanced Lymphoma Classification

Presented by:

- **Mr. Ahmed Fouad Abid**
- **Mr. Ammar Guerrida**

In front of the jury :

-	**Pr.**	**Djamel SAMAI**	**Chairman**
-	**Dr.**	**Meriem MEBARKIA**	**Supervisor**
-	**Dr.**	**Sabra BENKRINAH**	**Examiner**

Academic Year 2024 / 2025

Acknowledgments

First and foremost, we thank Allah, the Almighty, for granting us patience, strength, and guidance throughout these years of study. Without His blessings, this accomplishment would not have been possible.

We express our deepest gratitude to **Dr. Meriem Mebarkia**, our supervisor, for her unwavering dedication, insightful guidance, and invaluable advice. Her patience, expertise, and encouragement were instrumental in shaping this research and achieving meaningful results. We are profoundly grateful for her contributions to the success of this academic endeavor.

We extend our sincere thanks to the members of the jury, **Pr. Djamal Samai** and **Dr. Samira Souri**, for graciously accepting to review and evaluate this work. Their time and feedback are deeply appreciated.

Our heartfelt appreciation goes to all our professors and teachers for their wisdom, support, and mentorship throughout our academic journey. Their lessons and encouragement have left an indelible mark on our growth.

Lastly, we thank our colleagues and peers for their camaraderie and shared perseverance, which made this journey both enriching and memorable.

Dedication

To My Family

With pleasure, with an open heart, and with great joy, I dedicate my work to my dear father and mother: "He gave me life, tenderness, and the courage to succeed. All I can give you cannot express my love and gratitude. In this testimony, I offer you humble thanks for your work, sacrifices, and passion with which you have always surrounded me."

I dedicate this work to: my dear brothers Marouane, Farouk, and Khaled, and to the star of our home, my dear sister.

To My Dear Friends

To those who helped me build ideas and correct my many mistakes, motivating me by sharing words of advice and constant encouragement with the perfect blend of insight and humor throughout my years of study and during the process of researching and writing this treatise.

To all my friends who have accompanied me in life, and to all those who have passed through my life. Special thanks go to my colleague Ammar.

Ahmed FOUAD

Dedication

I dedicate this work with all my heart to my family, my greatest source of love and support.

To my late father:

Your memory has been my guiding light. In moments when all seemed impossible, thoughts of you gave me the strength to persevere. This achievement is yours as much as mine.

To my beloved mother:

Your endless prayers and unconditional love have been my shelter through every storm. No words can fully express my gratitude for your sacrifices.

To my dear friends:

Thank you for standing by me with unwavering encouragement, wise advice, and steadfast support during the most challenging times. Your presence transformed this journey, making it both lighter and more meaningful.

Ammar GUERRIDA

Abstract

Recent advancements in digital pathology have underscored the critical role of automated medical image analysis for disease diagnosis. While texture-based approaches are widely employed, conventional handcrafted feature extraction methods often yield suboptimal results due to high inter-class correlations in medical imagery. This thesis proposes a novel hybrid framework that synergizes Local Phase Quantization (LPQ) with chaotic-weighted Gaussian filtering and Bat Algorithm optimization to enhance lymphoma subtype classification. Our methodology begins with a deep texture analysis of histopathological images using an adaptive filter bank, where chaotic systems dynamically weight filter parameters. The extracted descriptors are processed via LPQ for robust feature representation, while the Bat Algorithm optimizes filter configurations to maximize discriminative power. Validated on a multicenter lymphoma dataset with inherent staining variability, the system achieves **96.17%** Accuracy, surpassing state of the art handcrafted techniques.

Index Terms: Computer Aided Diagnosis (CAD), Feature Extraction, Local Phase Quantization (LPQ), Chaotic Systems, Optimization, Medical Image Analysis.

الملخص

أبرزت التطورات الحديثة في علم ألمراض الرقمي الدور الحاسم للتحليل آللي للصور الطبية في تشخيص ألمراض. على الرغم من االنتشار الواسع لألساليب القائمة على تحليل الملمس، فإن طرق استخراج الميزات التقليدية غالباً ما تعطي نتائج دون المستوى ألمثل بسبب االرتباطات العالية بين الفئات المختلفة في الصور الطبية. تقترح هذه ألطروحة إطاراً هجيناً يجمع بين تكميم الطور المحلي والترشيح الغاوسي الموزون بالفوضى وتحسين خوارزمية الخفاش لتحسين تصنيف جديداً أنواع سرطان الغدد اللمفاوية تبدأ منهجيتنا بتحليل عميق لملمس الصور النسيجية المرضية باستخدام بنك مرشحات متكيف، حيث تقوم ألنظمة الفوضوية لتمثيل ميزات قوي، بينما تقوم خوارزمية LPQ بوزن معاملات المرشحات ديناميكياً. تتم معالجة الواصفات المستخرجة عبر الخفاش بتحسين تكوينات المرشحات لتعزيز القدرة التمييزية. عند التحقق على مجموعة بيانات متعددة المراكز لسرطان الغدد .اللمفاوية مع تباين طبيعي في التلوين، حقق النظام دقة بلغت ،96.17% متفوقاً على أحدث التقنيات التقليدية الكلمات المفتاحية :التشخيص بمساعدة الحاسوب، استخراج الميزات، تكميم الطور المحلي ، ألنظمة الفوضوية، التحسين، تحليل الصور الطبية.

الكلمات المفتاحية :التشخيص بمساعدة الحاسوب، استخراج الميزات، تكميم الطور المحلي ، ألنظمة الفوضوية، التحسين، تحليل الصور الطبية.

Contents

List of Figures

List of Tables

General Introduction

Medical imaging has become an indispensable diagnostic tool for most diseases, including cancers. The information derived from medical images plays a pivotal role in clinical decision making at various stages of patient care, such as detection, characterization, treatment response evaluation, and follow-up. Despite significant advancements in medical imaging technologies, the analysis and interpretation of these images remain a major challenge for pathologists. Any diagnostic error can lead to incorrect treatments, potentially endangering patients' lives. Due to the increasing workload and variability in expertise among pathologists, there is a pressing need for reliable decision-support tools to enhance diagnostic accuracy.

Computer Aided Diagnosis (CAD) systems, powered by Artificial Intelligence (AI), have emerged as transformative solutions in oncology. These systems assist in analyzing complex medical data, reducing diagnostic variability, and improving patient outcomes. Lymphoma, a heterogeneous group of cancers originating in the lymphatic system, presents unique diagnostic challenges due to its diverse subtypes such as Chronic Lymphocytic Leukemia (CLL), Follicular Lymphoma (FL), and Mantle Cell Lymphoma (MCL) each is requiring precise identification for optimal treatment planning. Traditional diagnostic methods, reliant on histopathological analysis, are often subjective and time consuming.

This dissertation addresses these challenges by proposing an Adaptive AI driven framework for lymphoma subtype classification. Our approach integrates Local Phase Quantization

(LPQ) with chaotic systems and bio inspired optimization algorithms to enhance feature extraction and classification Accuracy. Specifically, we:

1. **Develop a hybrid feature extraction pipeline:** Combining Gaussian filtering (weighted by Chaotic parameters) with LPQ to capture discriminative texture patterns in histopathological images.
2. **Optimize system parameters:** Using the Bat Algorithm, achieving subtype specific Accuracy up to **96.97%** for FL through red-channel analysis.
3. **Validate the framework:** On a multicenter dataset with real-world variability, ensuring clinical applicability.

This work is organized into three chapters: Chapter 1 provides an overview of CAD systems, AI techniques in Lymphoma Diagnosis, and current challenges. Chapter 2 details our proposed methodology, including Chaotic weighted filtering, LPQ based feature extraction, and Bat Algorithm optimization. Chapter 3 presents experimental results, comparative analysis, and performance validation.

By bridging traditional feature extraction with adaptive AI, this research contributes to computational pathology, offering a scalable tool to reduce diagnostic latency and improve lymphoma management.

Chapter I. Lymphoma recognition and Diagnosis

Introduction

Lymphoma is a group of cancers that originate in the lymphatic system, an essential component of the immune system. Due to its complex nature and subtle symptoms in early stages, timely recognition and diagnosis are vital. In this chapter, we explore the nature of lymphoma, its various types, and the revolutionary changes in medical diagnosis. With advancements in Artificial Intelligence (AI) and Computer-Aided Diagnosis (CAD), early detection is becoming more accurate, reducing mortality and improving patient outcomes [01] [02].

I.1. Understanding Lymphoma

Lymphoma is a complex group of blood cancers that originate in the lymphatic system, a critical part of the body's immune defense. To fully grasp its impact, it is essential to explore its fundamental nature, classifications, and progression. The following sections delve into lymphoma's definition as a cancer, its distinct types, and the evolving landscape of its treatment and research. This foundational knowledge will provide clarity on the disease's mechanisms and current medical advancements.

I.1.1. Lymphoma Cancer

The lymphatic system plays a crucial role in the body's immune response. It helps protect the body against infections by destroying pathogens such as bacteria and viruses. Additionally, it prevents fluid buildup in body tissues by transporting excess fluid and waste through lymphatic vessels for elimination. Lymphocytes can be affected by many diseases, among which cancer (lymphoma) is the most dangerous. When cancer infects these cells, it can pose a serious threat to the entire body by attacking the human body's main line of defense. Lymphoma is a type of

cancer that forms in the cells of the lymphatic system, which is part of the immune system. The lymphatic system is responsible for producing and transporting lymphocytes, immune cells that help fight infections and diseases. Lymphoma occurs when lymphatic cells grow abnormally and multiplies uncontrollably, forming tumors in the lymph nodes [03].

I.1.2. Lymphoma types

There are primarily three subtypes of lymphoma, classified based on the specific characteristics of lymphocytic cells and their clinical behavior [04]:

- ✓ **CLL (Chronic Lymphocytic Leukemia):**
  - A type of B-cell lymphoma characterized by the accumulation of abnormal mature B lymphocytes in the bone marrow, blood, and lymph nodes.
  - Typically slow-growing, often unnoticed for years.
  - Common in older adults but can affect younger individuals.
  - Symptoms: Persistent fatigue, swollen lymph nodes, enlarged spleen/liver, frequent infections, abnormal bleeding, and unexplained weight loss.
  - Diagnosis: Blood tests and bone marrow biopsy. Prognostic factors (disease stage, chromosomal abnormalities) guide treatment.
- ✓ **MCL (Mantle Cell Lymphoma):**
  - An aggressive non-Hodgkin lymphoma originating from B cells in the mantle zone of lymph nodes.
  - Usually diagnosed in older adults (median age ~60).
  - Symptoms: Swollen lymph nodes, night sweats, fatigue, weight loss, fever, and spleen/liver enlargement.
  - Highly aggressive, spreading quickly to bone marrow, distant lymph nodes, digestive tract, and CNS.
  - Diagnosis: Lymph node biopsy, supported by CT/PET scans to assess disease spread.
- ✓ **FL (Follicular Lymphoma):**
  - A non-Hodgkin subtype marked by abnormal B-cell growth in lymphoid follicles.
  - Symptoms: Painless lymph node swelling, fatigue, weight loss, night sweats, unexplained fever, and sometimes itchy skin.
  - Diagnosis: Medical exams, blood tests, biopsies, and imaging (CT/PET scans) to determine disease stage.

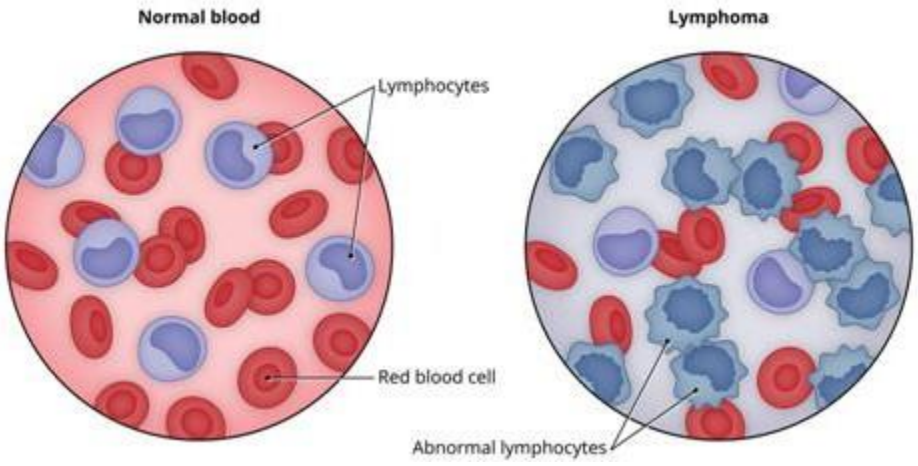

Fig. **1**. Normal vs abnormal blood cell [05].

I.1.3. Lymphoma cancer revolution

Lymphoma begins when genetic mutations disrupt the normal life cycle of lymphocytes, immune cells residing in the lymph nodes, spleen, and bone marrow. These mutations, triggered by factors like viral infections (e.g., EBV), environmental toxins, or inherited predispositions, cause lymphocytes to proliferate uncontrollably, evading apoptosis. Early-stage lymphoma often manifests subtly, with painless lymph node enlargement, fatigue, or night sweats, as malignant cells accumulate locally. Over time, clonal evolution fosters aggressive subpopulations that spread hematogenously or via lymphatic vessels, infiltrating distant organs (stage III/IV). In advanced disease, immunosuppression and organ dysfunction arise from bone marrow suppression (e.g., anemia), hepatosplenomegaly, or CNS involvement. Histopathological grading (e.g., indolent follicular lymphoma vs. aggressive DLBCL) further dictates tempo, underscoring the interplay between molecular aberrations (e.g., BCL2 translocations) and clinical lethality. Despite its systemic reach, modern therapies, from R-CHOP to CAR-T, leverage this biology to transform once-fatal diagnoses into manageable chronic conditions or even cures [06].

I.2. Clinical Diagnosis

I.2.1. Physical Examination: A Medical Examination, also known as a Physical Examination, is a routine medical examination once a year that aims to check a person's physical health and note any changes or signs indicating possible health problems. A person does not have to be sick to have a thorough medical and physical examination; this is also an opportunity to speak with a doctor about any pain or symptoms they are having, which is especially important for people

over the age of 50. The physical exam is typically performed by a medical professional in a doctor's office or private hospital [07], and it may begin with areas of the patient's body that have lumps that could indicate cancer. During the physical exam, the expert can look for any abnormalities, such as a change in skin color or enlargement of an organ that could indicate cancer. Benefits of a Physical Examination [08] [09]:

- ✓ Early detection of any disease symptoms and confirmation of possible diseases, aid in the initiation of treatment.
- ✓ Increase the chances of complete recovery and avoid any serious complications.
- ✓ Determine whether the patient has any problems or risk factors that could lead to disease in the future.
- ✓ Determine if the patient needs additional tests and investigations.
- ✓ Updating the patient's vaccinations as needed.
- ✓ Ensure that the patient maintains a healthy lifestyle.
- ✓ Develop a relationship between the patient and the caregivers.

I.2.2. Laboratory test: Laboratory tests, such as urine and blood tests, can help in the detection of abnormal changes that may be caused by cancer, it is used for:

- ✓ Confirmation of the diagnosis of the disease: some of these tests are confirmatory, and others are indicative of the disease.
- ✓ Aid in Differential Diagnosis: Laboratory tests help diagnoses diseases.
- ✓ To know the prognosis: Laboratory tests help to know the progress of the disease.
- ✓ To determine the prognosis: Laboratory tests.
- ✓ Screening tests: Laboratory tests are carried out in the form of panels to diagnose the disease in the community.

I.2.3. Imaging Examination: Imaging plays a crucial role in diagnosing, staging, and monitoring lymphoma. Techniques such as computed tomography (CT)**,** positron emission tomography (PET)**,** and magnetic resonance imaging (MRI) help evaluate tumor size, location, and metastasis. PET-CT, in particular, is highly sensitive for detecting metabolic activity in lymphoma cells, distinguishing between active tumors and scar tissue. Imaging also guides biopsies and assesses treatment response, making it indispensable in clinical management [10].

I.2.4. Biopsy: A biopsy is a procedure in which a sample of tissue from the body is removed for examination. Biopsy specimens can be obtained in a variety of ways. Needle biopsy is used when a lesion or tumor can be felt or identified in imaging studies [11]. During the procedure, a thin, hollow needle is inserted into the lesion and guided through an X-ray or ultrasound ma- chine. A small amount of tissue is taken for examination. Doctors use a microscope to examine cell samples:

- ✓ Normal Cells: appear uniform and similar in size and organization.
- ✓ Cancer Cells: appear less organized and appear in different sizes and without a clear ar- rangement.
- ✓ The appropriate biopsy procedure is determined by the type and location of cancer, in most cases, a biopsy is the only way to confirm a diagnosis of cancer

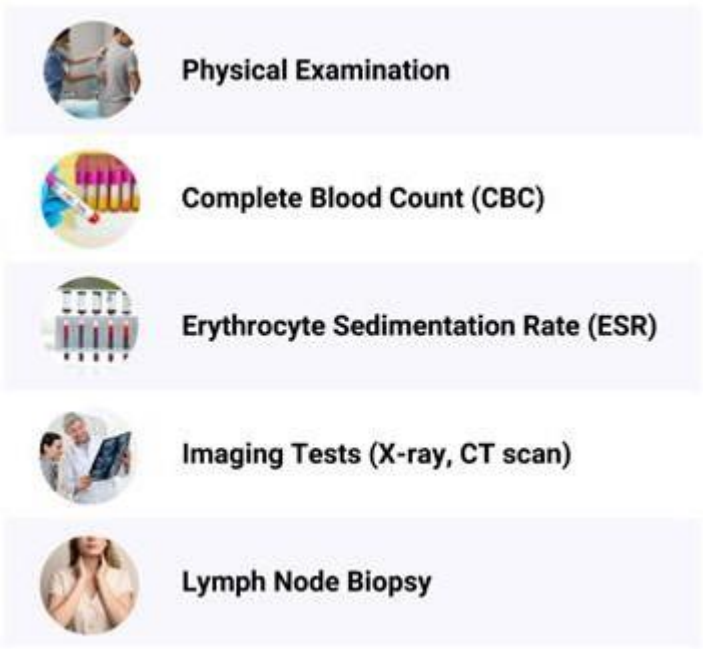

Fig. 2. Typical types of Lymphoma blood cell clinical Diagnosis [12].

I.3. Diagnosis based AI

Recent advances in artificial intelligence have revolutionized lymphoma diagnosis, enabling more accurate and efficient detection methods. AI systems can analyze complex medical data at unprecedented scales, assisting clinicians in making faster and more precise diagnostic decisions [13]. These technologies are particularly valuable in lymphoma diagnosis due to the disease's complex heterogeneity and subtle diagnostic features.

I.3.1. Types of Artificial Intelligence

Modern AI applications in lymphoma diagnosis primarily utilize three key approaches:

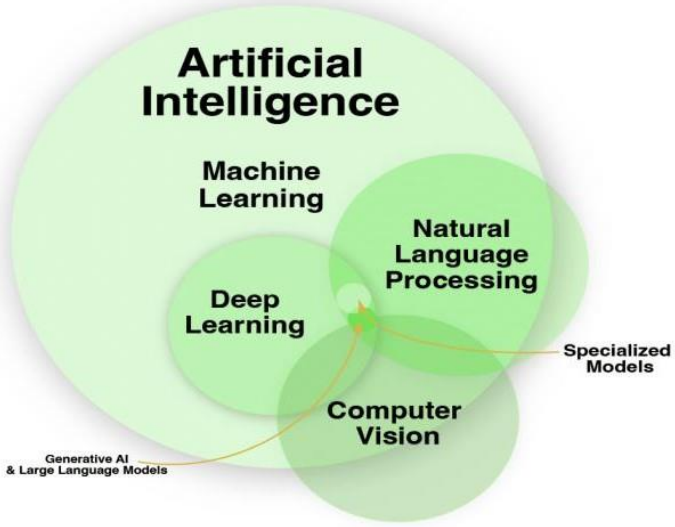

Fig. 3. The evolution of AI types [14].

✓ **Machine Learning (ML):**

Supervised learning algorithms are trained on large datasets of lymphoma cases to recognize diagnostic patterns in pathology images and clinical data. For instance, random forest models have achieved 92% accuracy in distinguishing between Hodgkin and non-Hodgkin lymphoma subtypes [15].

✓ **Deep Learning (DL):**

Convolutional neural networks (CNNs) excel at analyzing medical imaging. Recent studies demonstrate DL models can detect malignant lymph nodes in CT scans with sensitivity exceeding 94%, significantly reducing false negatives [16].

✓ **Natural Language Processing (NLP):**

Advanced NLP systems extract clinically relevant information from unstructured physician notes and medical literature. One study showed NLP could identify 87% of key lymphoma diagnostic criteria from electronic health records [17].

I.3.2. Branches of Artificial Intelligent

AI applications in lymphoma diagnosis span several specialized domains:

Computer Vision: Automated image analysis systems can quantify malignant cell percentages in bone marrow biopsies with precision matching expert hematopathologists (mean concordance 89%) [18]. these systems significantly reduce inter-observer variability in lymphoma grading.

Predictive Analytics: Machine learning models incorporating clinical, genomic and treatment response data can predict disease progression with 85% accuracy for follicular lymphoma patients [19]. Such models enable personalized surveillance strategies.

I.3.3. Computer Aided Diagnosis

Computer-Aided Diagnosis (CAD) systems represent a transformative partnership between artificial intelligence and clinical expertise in lymphoma diagnosis. These sophisticated tools analyze medical images and pathology slides with remarkable precision, identifying subtle patterns that might escape human detection during lengthy diagnostic workflows. By processing vast datasets of historical cases, CAD provides quantitative assessments of tumor characteristics, measuring everything from lymph node enlargement to metabolic activity, while maintaining consistency across evaluations.

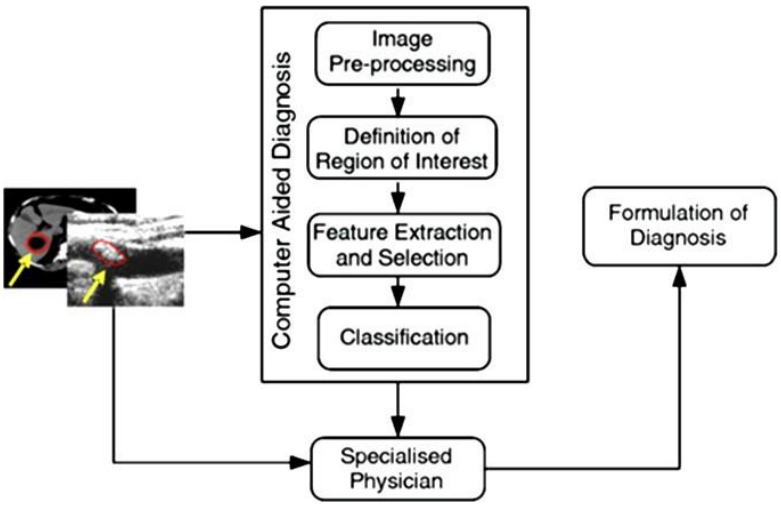

Fig. 4. Computer-aided Diagnosis for disease detection diagram [20].

Clinical studies demonstrate CAD's particular value in complex diagnostic scenarios, such as distinguishing between indolent and aggressive lymphoma subtypes or detecting early transformation. The technology doesn't replace pathological judgment, but rather enhances it, serving as a decision-support tool that reduces diagnostic variability. Recent implementations have shown CAD can decrease diagnosis time by 30% to 40% while maintaining diagnostic accuracy above 90% [21]. As these systems evolve, they're increasingly incorporating multimodal data, combin-

ing imaging findings with genomic and clinical information, to support truly personalized diagnostic approaches.

I.4. CAD System Design

I.4.1. Training Phase

During the training phase, the classifier is constructed by learning patterns and features from a labeled dataset. This process involves feeding the model a curated and representative sample of data, allowing it to iteratively adjust its internal parameters to minimize prediction errors. The selection of training data is critical, as the model's performance heavily depends on the quality, diversity, and relevance of the dataset.

In the context of lymphoma classification, the training set typically consists of histopathological images, genomic data, or clinical records, each annotated with corresponding diagnostic labels. The dataset must be sufficiently large to capture biological variability and mitigate overfitting, while also being carefully balanced to avoid biases that could skew the model's predictions.

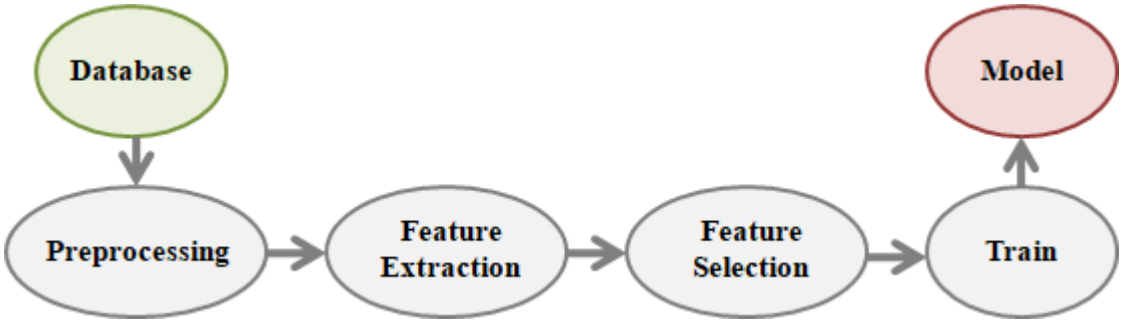

Fig. 5. Train Phase Model.

Beyond medical imaging, training data can encompass diverse modalities, such as textual re- ports, sensor-derived measurements, or multimodal clinical data. Preprocessing steps including normalization, augmentation, and feature extraction are often applied to enhance the model's ability to generalize to unseen cases. The choice of algorithm further influences the training dynamics and eventual diagnostic accuracy.

❖ **Data Preprocessing**

Medical images, whether acquired for diagnostic or therapeutic purposes, are often affected by sensor artifacts, illumination noise, and other acquisition related distortions. To ensure reliable analysis, a preprocessing stage is essential, involving a series of computational operations designed to enhance image quality before further processing.

In computer aided diagnosis (CAD) systems, preprocessing typically involves three key operations:

- **Noise reduction**

To eliminate random pixel variations and sensor artifacts, is achieved through adaptive filtering algorithms (e.g., non-local means or wavelet transforms) that suppress stochastic pixel variations while preserving critical pathological features.

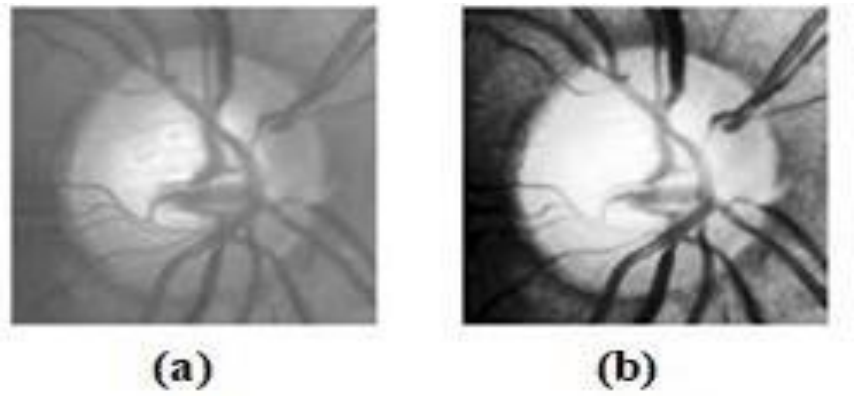

Fig. 6. (a) Gray scale image (b) Noise reduction and Contrast Enhancement image [22].

- **Contrast enhancement**

To improve image clarity and feature visibility, utilizes histogram equalization and gamma correction to amplify subtle intensity differences between tissues, particularly improving visibility of low-contrast lesions.

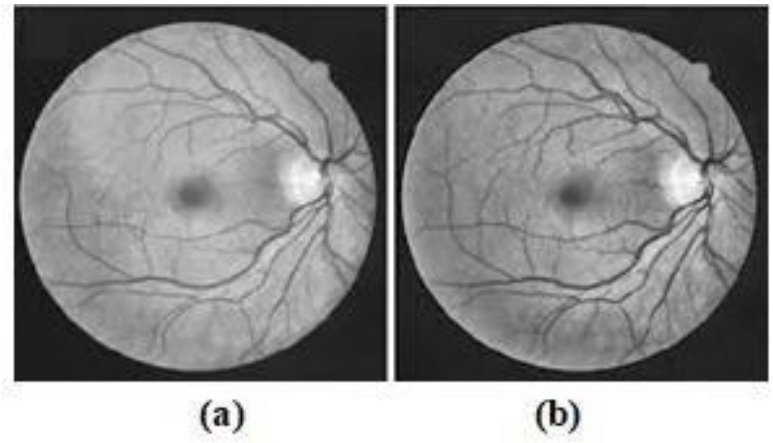

Fig. 7. (a) Gray scale image (b) Contrast Enhancement image [23].

- **Edge sharpening**

To better define anatomical boundaries, applies anisotropic diffusion and gradient-based operators to enhance anatomical boundaries without amplifying background noise

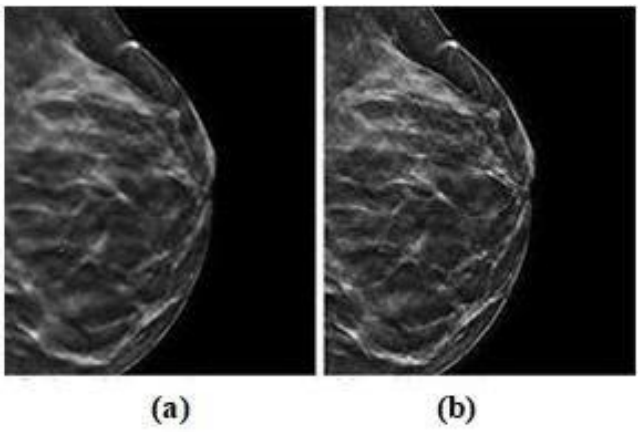

Fig. 8. (a) Before sharpening (b) After sharpening [24].

The primary objectives of medical image preprocessing are:

- To correct acquisition-related defects and normalize image quality
- To reduce noise while preserving diagnostically relevant features
- To standardize images by compensating for variations in exposure and illumination
- To enhance the differentiation between anatomical structures (e.g., distinguishing between organs, bones, and pathological tissues)
- To ensure consistency with established anatomical references

For lymphoma imaging specifically, these preprocessing steps are crucial for:

- Accurate identification of lymph node boundaries
- Clear visualization of malignant masses
- Reliable differentiation between pathological features and normal anatomy
- Elimination of confounding elements such as rib shadows in thoracic imaging
- Alignment with anatomical reference data for consistent interpretation

❖ **Feature Extraction**

The primary objective of feature extraction in CAD (Computer Aided Diagnosis) is to transform raw medical data into more meaningful and informative representations. This process reduces data complexity, identifies relevant aspects, and provides a robust foundation for subsequent analysis and interpretation. The importance of this step lies in its ability to enhance diagnostic system accuracy and support medical decision making. By identifying discriminative and in-

formative features within medical data, it improves pathology detection and classification, predicts complications, and tracks disease progression [25].

- **Roles of Feature Extraction**

Feature extraction plays a critical role in facilitating medical analysis and decision making [26]:

- **Dimensionality Reduction:** Medical datasets often contain high-dimensional variables that complicate analysis and visualization. Feature extraction reduces dimensionality by identifying the most discriminative and informative features, simplifying analysis and improving data interpretability.
- **Identification of Relevant Information:** Medical data may include noisy, redundant, or irrelevant information. Feature extraction isolates clinically relevant aspects by highlighting key features associated with pathologies, symptoms, or test results, effectively filtering essential information while eliminating noise.
- **Diagnostic Accuracy Improvement:** By extracting distinctive pathological features, CAD systems achieve higher precision in disease detection, classification, and prediction. This directly enhances diagnostic reliability and reduces medical errors.
- **Decision Making Optimization:** Feature extraction provides actionable insights to support evidence-based clinical decisions. Extracted features enable risk assessment, treatment personalization, and outcome prediction, leading to more accurate and efficient medical decision making.

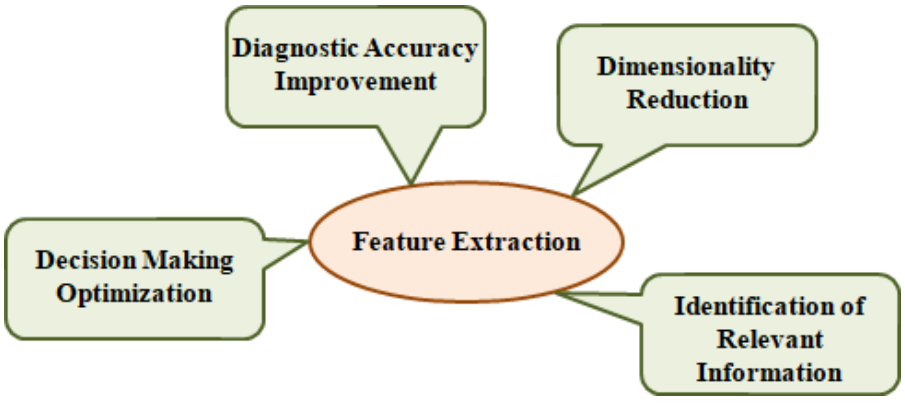

Fig. 9. The role of Feature extraction in medical analysis and decision making.

- **Types of Feature Extraction**

In medical imaging-based CAD (Computer Aided Diagnosis), different types of image features can be used to represent medical data and facilitate analysis and decision making. It is important

to note that the selection of appropriate features depends on the specific CAD task and the type of data being processed [27].

- **Line Based Features:** Line based feature extraction involves detecting and representing lines within an image. These lines may include edges, contours, or more complex linear structures, providing critical information about an image's geometry and spatial organization. Various feature extraction techniques can be applied to capture line-based features, including edge detection algorithms (e.g., Canny, Sobel), Gabor filters, and thresholding methods.
- **Texture Based Features:** Texture-based feature extraction focuses on recurring patterns and local intensity variations in an image. Textures may exhibit regular patterns (e.g., tiled structures) or stochastic properties (e.g., granularity). This approach captures information about an image's local composition and structural properties.
- **Appearance Based Features:** Appearance-based feature extraction emphasizes global visualattributes such as color, shape, size, and spatial distribution of objects. This holistic representation is particularly useful for tasks like object recognition or image classification.

❖ **Feature Extraction Methods**

Feature extraction is the process of transforming raw pixel values from an image into more meaningful and useful information that can be utilized in a computer vision application. For this reason, these processes have been carefully designed by specialists (manual or handcrafted creation) using their knowledge and expertise. Typically, these techniques are applied to an image without prior knowledge of its content.

To improve their performance, these techniques have been enhanced with prior contextual knowledge through machine learning (learned handcrafted). In recent years, with the advent of deep learning, neural networks themselves automatically perform the work of experts in convolutional layers.

- **Traditional Methods:** Also referred to as handcrafted or manually designed methods, these involve feature extraction based on specific knowledge and rules defined by ex- perts. In this approach, descriptors are manually designed to capture relevant information in images.

For example, in edge detection, operators such as the Laplacian or gradient operators can be used to extract edge features.

Handcrafted methods allow for interpretability and control over the feature extraction process but may be limited by the experts' ability to define relevant descriptors [28].

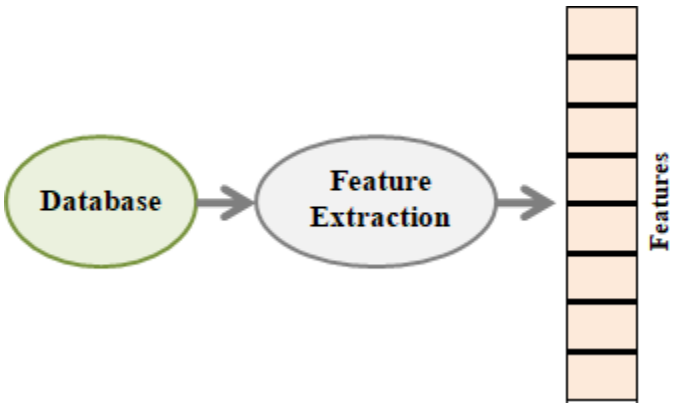

Fig. 10. Traditional feature extraction method.

- **Learned Traditional Methods:** Learned handcrafted methods combine aspects of handcrafted techniques and machine learning. They involve training manually designed descriptors from a learning dataset. In this approach, experts initially design a set of de- scriptors and then use machine learning techniques to select and weight the most relevant ones.

This automates part of the descriptor design process and adapts them to specific data [29].

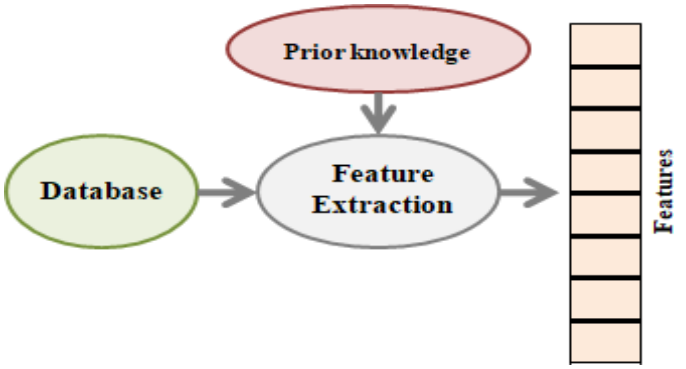

Fig.11. Learned traditional feature extraction method.

- **Deep Methods:** Deep learning based methods have revolutionized feature extraction. These methods use deep neural network architectures to automatically learn descriptors from raw data. Deep neural networks can learn hierarchical and abstract representations of data, enabling them to capture complex and nonlinear features. Popular models such as CNNs (Con-

volutional Neural Networks) and RNNs (Recurrent Neural Networks) are used in computer vision tasks to extract features from images, videos, or temporal sequences. Deep methods often outperform traditional feature extraction techniques but typically require large da- tasets and significant computational resources for training. Additionally, interpreting the learned features can be more challenging compared to traditional methods, whether they in- volve learning or not [30].

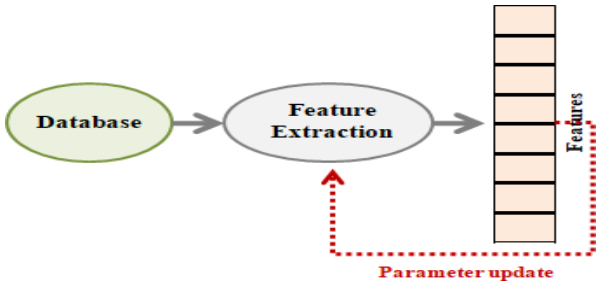

Fig. 12. Deep feature extraction method.

I.4.2. Classification phase

The system processes extracted descriptors through specialized classification algorithms that: Assign segmented objects to diagnostic categories based on morphological characteristics and determine mass type through feature-based pattern recognition and perform spatial probability mapping to:

- Evaluate structural features region by region
- Calculate detection confidence scores
- Differentiate true positives from artifacts

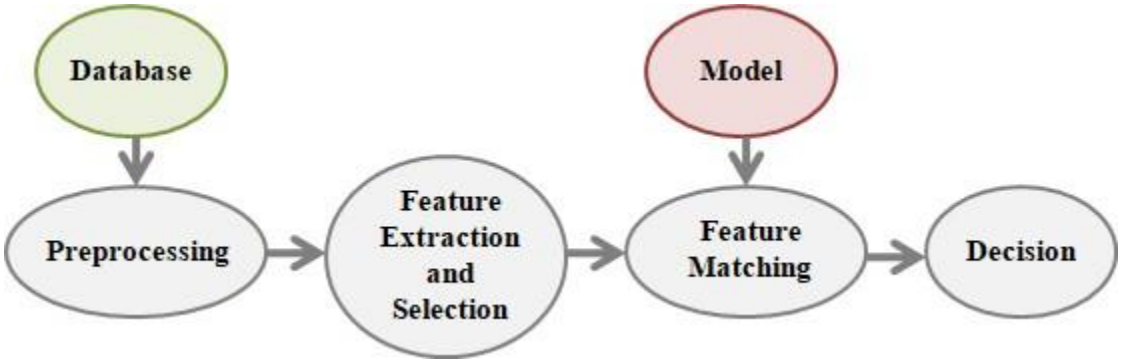

Fig. 13. Classification Phase Model.

- **Most tasks of Classification**

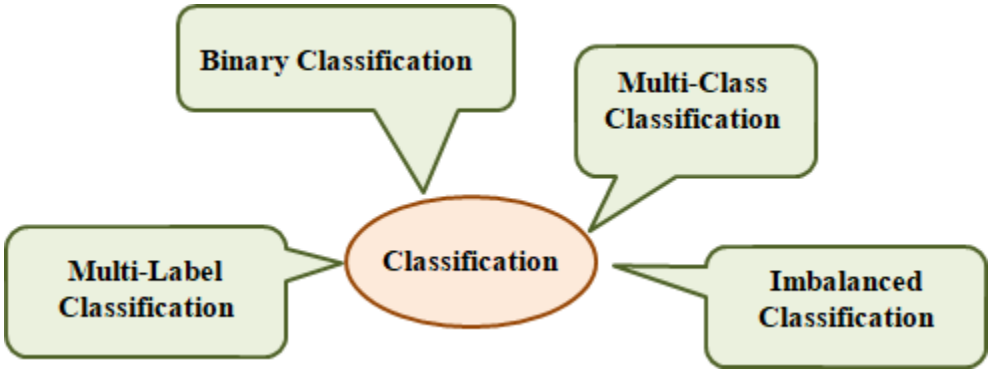

Fig. 14. The most Classification task.

✓ **Binary Classification:**

Yields one of two possible results: positive or negative.

✓ **Multi-Class Classification:**

In multi-class classification systems, instances are exclusively categorized into a single class label from three or more possible categories, following a mutually exclusive assignment paradigm.

✓ **Multi-Label Classification :**

Each sample gets a single label from multiple possible classes.

✓ **Imbalanced Classification**

This dataset exhibits class imbalance, with some categories (majority classes) appearing more frequently than others (minority classes).

I.5. Conclusion

The diagnosis of lymphoma has advanced from traditional methods to include AI technologies that improve accuracy and speed. While tools like machine learning and CAD assist in pattern detection and tumor analysis, they support rather than replace clinical expertise. Challenges re- main, such as the need for better datasets and continued human oversight. Future developments aim to combine AI with genomic and clinical data for personalized diagnostics, blending technology with physician judgment to enhance patient care globally.

Chapter II. Lymphoma features Learning

Introduction

This chapter presents the theoretical foundation for our proposed texture analysis system in lymphoma diagnosis. We introduce Local Phase Quantization (LPQ), and detail its quantization and decorrelation processes. The framework is enhanced by Gaussian filtering (for noise reduction) and chaotic systems (for parameter optimization), with the Bat Algorithm dynamically tuning filter configurations. These components form a hybrid pipeline designed to improve discriminative feature extraction for lymphoma subtype classification, bridging traditional techniques with adaptive AI Driven Optimization.

II.1. Theoretical Foundation

This section outlines the preliminary requirements for the proposed texture analysis process.

II.1.1. Local Phase Quantization

Local Phase Quantization (LPQ) [31] captures phase information from local image regions by leveraging the blur-invariant property of Fourier phase responses in slightly blurred images.

- **Local Frequency Analysis:**

LPQ operator on an image pixel is done by employing Short-term Fourier Transform (STFT) over an $M \times M$-sizedwindow(W_n). Thus, for all pixel locations $x = \{x_1, x_2, \ldots, x_{HW}\}$ in an $H \times W$-sized image $(f(x))$, the local image patches in W_n are defined as follows:

$$f_x(y) = f(x - y), \quad \forall y \in W_n \tag{01}$$

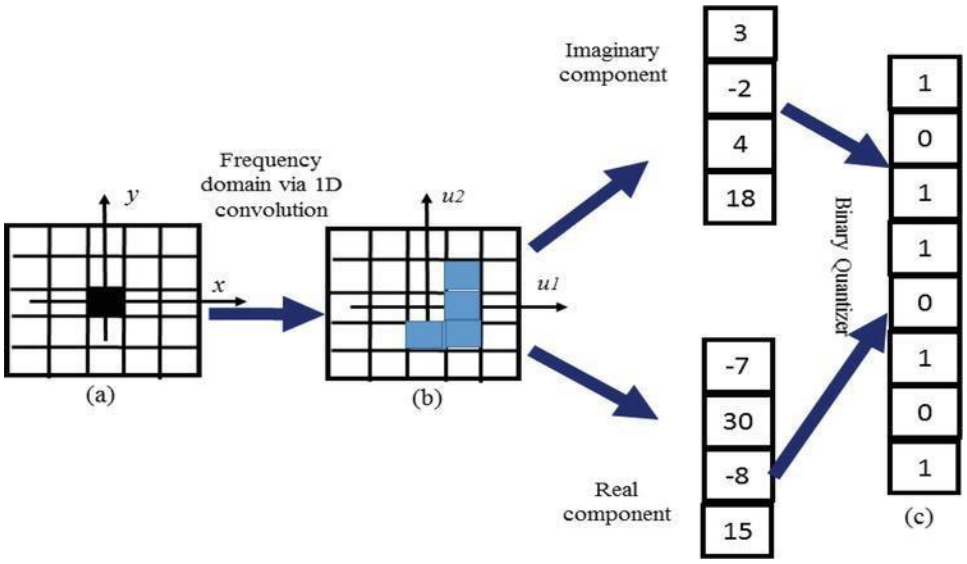

Fig. 15. Example of LPQ encoding scheme. (a) Input 5x5 patch. (b) Frequency domain representation. (c) LPQ code [32].

Applying STFT to W_n gives us the local frequency domain representation:

$$F_x(u) = \sum_{y_i \in W_n} f_x(y) e^{-j2\pi u^T y_i} \qquad (02)$$

where $i = \mathbf{1}, \ldots, M^2$. The variable u comprises l frequency variables $(u_1, u_2, \ldots u_l)$ that serve as blur-insensitive local descriptors. Eq. (02) can be rewritten as follows:

$$F_x(u) = \psi_u^T f_x, \quad \psi_u^T(y) = e^{-j2\pi u^T y} \qquad (03)$$

Where

$$\psi_u^T = [\psi_u(y_1), \psi_u(y_2), \ldots, \psi_u(y_{M^2})] \qquad (04)$$

And

$$f_x = [f_x(y_1), f_x(y_2), \ldots, f_x(y_{M^2})] \qquad (05)$$

To find the LPQ codeword, a phase quantization process is performed. Indeed, if the coefficients to be quantized are statistically independent, scalar quantization is performed directly; otherwise, the samples must be decorrelated before quantization.

- **Phase Quantization:**

The blur-insensitive representation uses l frequencies, producing an l -length feature vector for each pixel. In practice, the phase is quantized in four quadrants using the following quantizer:

$$Q(F_x(u)) = (Re\{F_x(u)\} > 0) + 2(Im\{F_x(u)\} > 0) \qquad (06)$$

This quantization can be expressed with 2 bits per frequency for each pixel, for a total of $2l$ bits; the concatenation of the codes for the l frequency components results in a single codeword. To achieve a blur-insensitive representation, only low-frequency components are employed, as they contain the most image energy. In practice, the local coefficients F_x are computed for each pixel at four low frequencies($l = 4$):

$$[u_1, u_2 u_3, u_4] = \begin{bmatrix} a & 0 & a & a \\ 0 & a & a & -a \end{bmatrix} \quad (07)$$

The scalar a is the highest frequency whose point spread function (PSF) is positive and it is usually calculated from $M\left(e.i.a = \frac{1}{M}\right)$. Finally, an 8-bit codeword ($[q_7, q_6, q_5, \ldots, q_0]$) can be obtained to describe the local texture surrounding each pixel (W_n) this codeword can easily be transformed into a decimal number (in the range $0-255$) by simple binary decoding:

$$LPQ_i = \sum_{j=1}^{8} q_i 2^{j-1} \quad (08)$$
(08)

- **Decorrelation:**

Vector quantization is typically more efficient incases when the coefficients to be quantized are correlated. To decorrelate the frequency coefficients, we first separate the real and imaginary components of $F_x(u)$ and then concatenate them as follows:

$$F_x(u) = [F_x(u_1), F_x(u_1), \ldots, F_x(u_l)] \quad (09)$$

$$F_x = [Re\{F_x(u)\} \quad Im\{F_x(u)\}] = [F_x^R F_x^I] \quad (10)$$

The STFT transform (Eq. (10)) shows that F_x and f_x are linearly dependent, so we can write:

$$F_x = \psi . f_x \quad (11)$$

Where

$\psi = [\psi_R \psi_I]^T$

And

$\psi_R = Re\{[\psi_{u_1}, \psi_{u_2}, \ldots, \psi_{u_l}\}]$

And

$\psi_I = Im\{[\psi_{u_1}, \psi_{u_2}, \ldots, \psi_{u_l}\}]$

To decorrelate F_x, Ojansivu et al. [33] applied the following transform:

$G_x = V^T . F_x$

Where V is an orthogonal matrix obtained from the matrix (D) mentioned below:

$D = U\sum V^T$

D is the covariance matrix of F_x and can be obtained by:

$D = \psi C \psi^T$

where C is the covariance matrix of $M \times M$ samples in W_n. After the decorrelation process, a scalar quantizes the j^{th} coefficient (g_j) of G_x:

$$q_j = \begin{cases} 1, if\ g_j \geq 0 \\ 0, \ otherwise \end{cases}$$

The resulting codeword is then transformed into a decimal number. After computing the LPQ code words for each pixel in the image, and since the LPQ range is 0 to 255, an image containing the LPQ information of the input image is produced. To calculate a feature vector (H) for the input image in practice, we divide the LPQ image into non-overlapping $r \times c$-sized sub-regions and compute the histograms for each sub-region. Then, these histograms are concatenated to provide an LPQ feature vector for the whole input image.

$H = [h_1, h_2, \dots, h_{rc}]$

This division is used to keep some information about the spatial arrangement of the patterns since the histograms of the whole image discard all of this information.

II.1.2. Gaussian Filter

The Gaussian filter is an image processing technique used to smooth data and reduce noise. Based on the Gaussian (normal) distribution, a low-pass Gaussian filter is defined by the following impulse response:

$$h(x, y) = e^{\left(\frac{x^2+y^2}{2\sigma^2}\right)}$$

- **Determining σ**

To select the value of σ, we choose a small ε for the element H at coordinates $(x, y) = (P, 0)$ (located at the center of a border pixel). The center value $(x, y) = (0,0)$ is 1, while border values must be negligible relative to the center.

$$\sigma = \frac{P}{\sqrt{-2\ln\ (\varepsilon)}}$$

- **Normalization**

The second step involves normalizing the matrix H by dividing all coefficients so their sum equals 1. This preserves the image's average grayscale level.

❖ **Operation Principle**

The Gaussian filter computes a weighted average of neighboring pixels around each image pixel. Weights are assigned based on the Gaussian distribution [34]:

Higher weights: are given to pixels closer to the center.

Lower weights: are assigned to distant pixels.

II.1.3. Chaotic Systems

A chaotic system (Logistic Maps or chaotic dynamics) is a nonlinear dynamical system highly sensitive to initial conditions. These systems exhibit complex, unpredictable behavior where minor variations in starting parameters lead to dramatically divergent outcomes.

Logistic maps are canonical examples of one-dimensional chaotic systems, extensively studied in chaos theory and nonlinear dynamics. They model population growth using a recursive nonlinear function defined as:

$$x_{n+1} = \mu x_n (1 - x_n)$$

- **Parameters:**

x_n: Population at time step n (normalized between 0 and 1)

μ: Growth rate parameter (controls system behavior)

- **Dynamical Regimes**

Extinction ($\mu < 1$): Population converges to zero.

Stable Equilibrium ($1 < \mu < 3$): Population reaches a steady-state value.

Chaos ($\mu > 3.75$):

Future values oscillate unpredictably.

Exhibits sensitive dependence on initial conditions infinitesimal changes in x_0 yield entirely divergent trajectories [35].

II.1.4. Bats algorithm

The Bat Algorithm is a nature-inspired metaheuristic optimization technique developed by Xin-She Yang in 2010. This method computationally mimics the echolocation behavior exhibited by

microbats during prey hunting. The algorithm operates through the following biological principles [36]:

❖ **Echolocation Mechanism**:

Microbats emit high-frequency sound pulses (typically 20-200 kHz)

These acoustic waves propagate through the search space

Reflected echoes provide spatial information about prey location

❖ **Computational Analog**:

Each virtual bat represents a potential solution

Pulse emission corresponds to solution evaluation

Echo reception guides the search direction

Frequency modulation controls exploration/exploitation balance

The micro-bats could locate a target thanks to an echo received this phenomenon is called echolocation. Each microbat is characterized by five variables, the frequency f_i , the velocityv_i, the position x_i, the loudnessA_i , the pulse rater_i.

These variables are calculated using the following equations

- **Frequency**

$$f_i = f_{min} + (f_{max} - f_{min})\beta$$

With, f_{min} and f_{max} Depending on the nature of the problem, β is a random number between 0 and 1.

- **Velocity**

$$v_i(t+1) = v_i(t) + (X_i(t) - X_{best})\, f_i$$

$v_i(t+1)$ is the new velocity of bat i, and $X_i(t)$ is its current position, $X_{best}$is the current best position.

- **Position**

$$X_i(t+1) = X(t) + v_i(t+1)$$

- **Loudness and Pulse rate**

Initially, it is calculated using the following equation:

$$A_i = rand \times (A_{max} - A_{min}) + A_{min}$$

$$r_i = rand \times (r_{max} - r_{min}) + r_{min}$$

If the random number is less than A_i and $f(X_i) < f(X_{best})$, Thereafter decreases A_i and r_i in each iteration according to the following equation:

$A_i = \alpha\, A_i^t$ with $0 < \alpha < 1$

$r_i = r_i^t\left[1 - e^{-\gamma^t}\right]$ $with\ 0 < \gamma$

With, f denote the objective function

The organization chart of the BAT algorithm is shown as follows

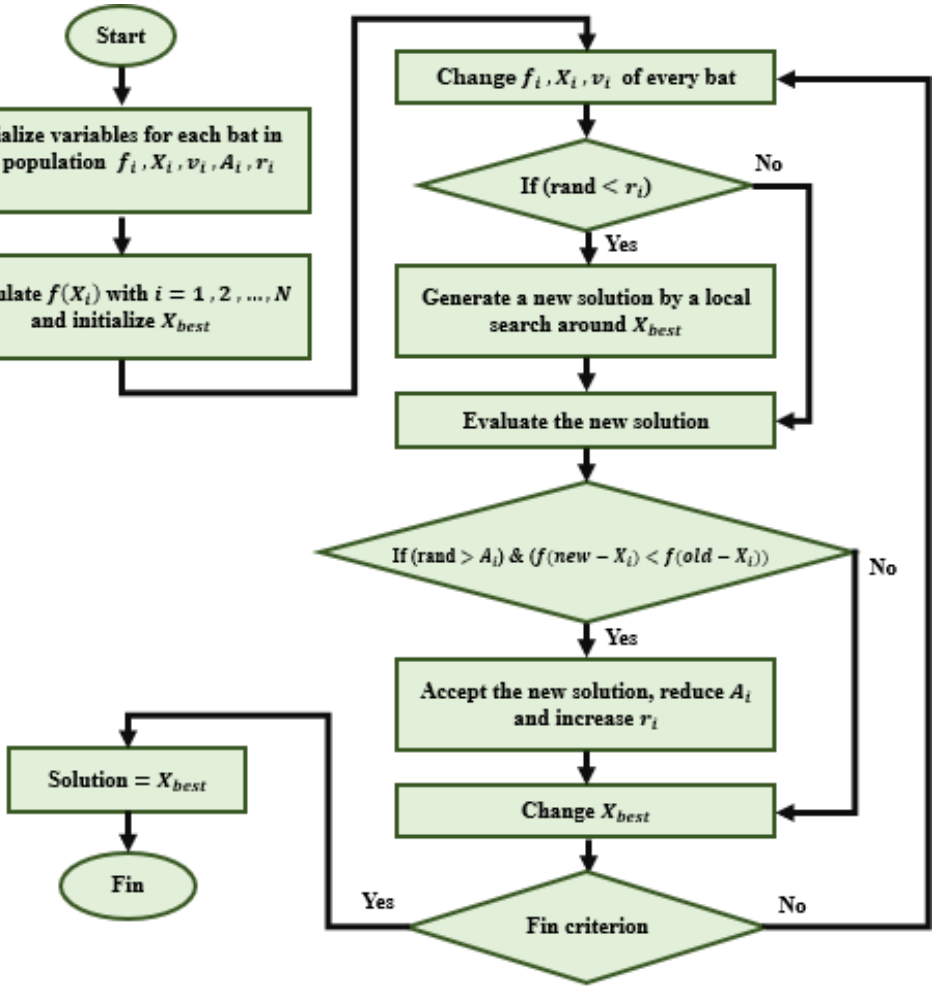

Fig. 16. The organization chart of the BAT algorithm.

II.2. Proposed System

To describe image texture in a Computer Aided Design system, we propose a Deep Convolution based Analysis Method. All images were initially preprocessed to enhance contrast.

In the following sections, we detail the techniques used in the proposed approach.

II.2.1. Deep Feature Extraction for Medical Image Classification

In classification problems, due to the high correlation between images of different classes, the goal is to extract feature vectors with high intra-class correlation and low inter-class correlation. However, in pathological tissue images, experts often struggle to detect diseases with the naked eye. Traditional (handcrafted) feature extraction methods fail to provide precise and discriminative vectors for automated diagnosis.

Inspired by the advantages of deep learning, we propose a novel feature extraction method for medical image classification. The approach consists of three main steps:

- **Gaussian Filter Bank with Chaotic Weighting**

The input image is filtered using a bank of Gaussian filters. Filter coefficients are weighted with chaotic values (logistic map-generated). This operation enhances edges, ridges, singular points, and texture patterns.

- **Data Reduction & Descriptor Fusion**

Filtered images are combined into a single descriptor to reduce dimensionality. Local Phase Quantization (LPQ) is applied to extract the final feature vector.

- **Bat Algorithm-Based Optimization**

Filter bank parameters (logistic map control parameter μ and initial value x_0) are optimized using the bat algorithm during training to maximize classification accuracy.

$$[\mu^{obest}, \mu_0^{obest}] = \mathcal{F}_{OPT}(S_{taux})$$

S_{taux}: optimizing classification rate (the objective function)

II.2.2. Functional Architecture

The complete pipeline (Fig. 17) comprises three layers:

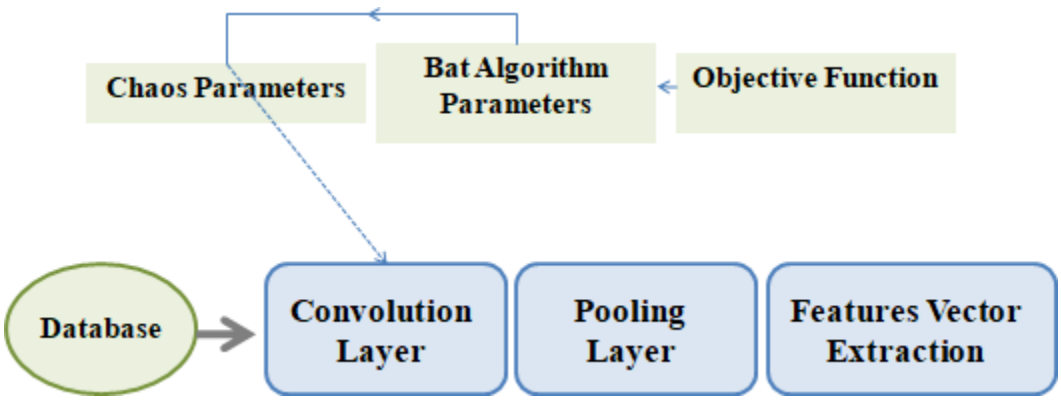

Fig. 17. The Proposed system layers.

- **Convolution Layer**

Enhances dominant features via 2D convolution between input images and chaotic-weighted filters

$$W_{i,\lambda_i} = M_{\lambda_i}\, Gauss(k_1, k_2) \in \mathbb{R}^{k_1 \times k_2}, \qquad \lambda_i = (\mu^i, x_0^i)$$

M_{λ_i} : Weight matrix generated via logistic map

$\lambda_i = (\mu^i, x_0{}^i)$: Logistic map parameters

μ^i: control parameter

$x_0{}^i$: initial value

$$\hat{I}_{ji} = I_j * W_{i,\lambda_i}, \quad i \in [1 \ldots N], \qquad et\ j \in [1 \ldots N_{train}]$$

N: Number of filters in bank

N_{train}: the number of images in the training set

$*$: the 2D convolution operation

$\hat{I}_{ji}$: the filtered output images

- **Pooling Layer**

Reduces data volume by binarizing filtered images (Eq. 23) and merging them into an integer valued image (Eq. 24).

$$I_{ji}^b(x,y) = \begin{cases} 0 & si \quad \hat{I}_{ji}(x,y) < \tau_b \\ 1 & si \quad \hat{I}_{ji}(x,y) \geq \tau_b \end{cases}, \quad i \in [1 \ldots . N]$$

τ_b: the binarization threshold

In our implementation, the binarization threshold was set to 0 since the filtered image coefficients exhibited equal probability of being positive or negative. The N binarized images I^b_{ji} are then converted into a single integer valued image through the following binary to integer transformation: $I_{jd}(x,y) = \sum_{i=1}^{N} I^b_{ji}(x,y) \cdot 2^{i-1}$

As a result, we obtain a single image to which we apply a traditional feature extraction method to derive its features.

- **Feature Vector Extraction**

LPQ is applied to generate the final feature vector V_j (Eq. 25), refined via Fisher based selection (Eq. 26–27) for optimal discriminative power.

$$V_j = \mathcal{F}_{LPQ}(I_{jd})$$

$$\tilde{V}_j = \mathcal{F}_{QNTZ}\left(v_j(x)\right) = \begin{cases} v_j(x) & if\ v_j(x) \geq \lambda_0 . \rho_v \\ 0 & otherwise \end{cases},\ \tilde{V}_j \in \mathbb{R}^{1\times L}, x \in [1 \dots N]$$

$$V_j = \mathcal{F}_{SLCT}(V_j)$$

V_j: represents the feature vector reorganized according to the coordinates obtained by Fisher's selection method resulting in a vector with a predefined number of coordinates (L) in descending order of importance, ρ_v is the mean value of v_j, and λ_0 is a predefined threshold value [37].

- **Classification**

In classification problems, discriminative machine learning enables the identification of a function that can correctly predict instance labels. As one of the most important techniques in discriminative learning, Support Vector Machine (SVM) classification [38] identifies the optimal hyperplane that correctly separates two classes with maximal margin.

A Support Vector Machine (SVM) classifier works by:

Concept:

- Mapping training data points into a high-dimensional space
- Establishing clear separation boundaries between classes

- Maximizing the margin (distance) between these boundaries and the nearest data points (support vectors)

Prediction Mechanism:

- New data points are classified based on which side of the decision boundary they fall
- The classifier determines the most probable category membership for each new point

Advantages:

- Exceptional performance in high-dimensional feature spaces
- Memory efficiency (relies only on support vectors for decision making)
- Versatility in handling:

Nonlinear data distributions (via kernel tricks)

Regression tasks

Unsupervised learning problems

Optimization Objective

- Identifies the optimal hyperplane that maximizes the margin between classes
- The margin is defined as the perpendicular distance to the nearest training points

II.3. Conclusion

In summary, this chapter established a robust theoretical framework for lymphoma texture analysis, combining LPQ's phase-based feature extraction with chaotic-weighted Gaussian filtering and Bat Algorithm optimization. The methodology addresses key challenges in medical image analysis including noise sensitivity and feature discriminability while maintaining computational efficiency. By integrating these components, we lay the groundwork for an accurate and adaptive classification system, validated experimentally in subsequent chapters. This approach demonstrates the potential of hybrid traditional AI methods in computational pathology, offering both interpretability and high performance.

Chapter III. Experimental Results

Introduction

Feature extraction plays a crucial role in image processing and analysis by identifying and isolating key information from visual data. In this study, we introduce an adaptive feature extraction technique where the classifier dynamically optimizes its parameters.

In this study, we evaluate our proposed methodology for distinguishing these subtypes through experimental validation. This chapter presents our findings, including a rigorous performance analysis of the technique within the framework of our research objectives.

III.1. Clinical Background and Challenges

The framework developed in this study is designed for generic image analysis and classification tasks, with adaptable parameters for diverse applications. In this investigation, we specifically evaluate its performance in the domain of medical image analysis, focusing on lymphoma subtype classification.

The proposed system offers potential clinical utility by:

- Assisting physicians in lymphoma diagnosis and monitoring through automated image analysis.
- Enabling lesion detection and characterization to support treatment decisions.

It is important to emphasize that our solution is specialized for lymphoma classification, leveraging morphological features and tissue patterns from medical imaging [38] to distinguish between disease subtypes with enhanced precision.

III.1.1. Lymphatic Cancer (Lymphoma):

The lymphatic system serves as a critical component of the body's immune defense network. Its primary functions include:

- **Pathogen elimination:** Neutralizing infectious agents (e.g., bacteria, viruses) through lymphocyte activity.
- **Fluid homeostasis:** Preventing tissue edema by draining excess interstitial fluid and metabolic waste via lymphatic vessels.

Among the pathologies affecting lymphocytes, lymphoma emerges as the most life-threatening malignancy. This cancer compromises the body's primary defense mechanisms by infiltrating and dysregulating immune cells. Specifically, lymphoma originates in the lymphatic system's cells a key subsystem of the immune network responsible for lymphocyte production and circulation.

The disease manifests when genetic mutations trigger uncontrolled proliferation of lymphoid cells, leading to tumor formation in lymph nodes [39]. These malignant transformations severely impair immune competence while enabling systemic spread.

III.2. Dataset and Imaging Characteristics

Robust assessment of such diagnostic systems requires validation against established benchmark datasets. This approach enables:

- Accurate measurement of classification performance metrics
- Optimization of feature extraction parameters
- Comparative analysis against state-of-the-art methods

The evaluation process must employ medically representative data to ensure clinical relevance and statistical significance of the results. Standardized datasets allow for reproducible testing of the system's diagnostic capabilities while facilitating parameter tuning for optimal feature selection.

III.2.1 Dataset Rationale

This dataset was created to help identify three common lymphoma subtypes (CLL, FL, and MCL) using standard H&E-stained tissue samples. Currently, doctors need expensive special tests to tell these subtypes apart. However, experienced pathologists have shown that regular microscope slides can work for diagnosis, suggesting that computer based methods could help too.

The dataset includes real-world cases with two important features:

- Samples prepared by different pathologists at multiple hospitals
- Deliberately varied staining quality (more than normally seen)

This makes the dataset more challenging but better represents actual hospital conditions, helping develop practical diagnostic tools.

III.2.2 Dataset Description

The dataset contains 374 microscopic images with 1388×1040 pixels and Class distribution which are (CLL with 113 images, FL with 139 images and MCL with 122 images).

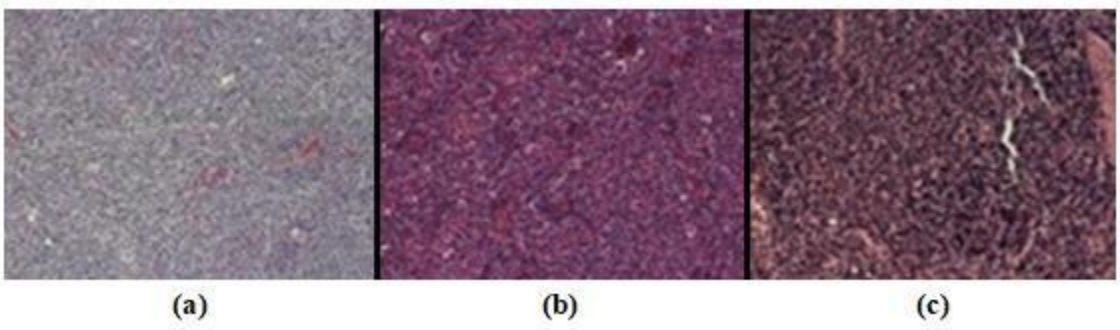

Fig.18. (a), (b), (c) Examples of the three lymphoma subtypes (CLL, FL, and MCL), respectively

The dataset was compiled from multiple clinical sources to create a representative cohort that includes natural variations in staining and scanning conditions commonly encountered in diagnostic practice. This multi-center approach ensures the dataset reflects real-world clinical diversity, accounting for differences in sample preparation and imaging protocols. **Fig.17.** provides visual examples of the three lymphoma subtypes (CLL, FL, and MCL) contained in this collection [40]. The inclusion of such variability strengthens the dataset's utility for developing robust computational analysis methods that can perform effectively across different laboratory settings.

III.3. Methodology Implementation

Before conducting system evaluation experiments, it is essential to define both the testing protocol and the equipment used to perform these experiments. This clarification ensures methodological rigor and reproducibility of results.

III.2.1. Testing Protocol

Our system is designed for one vs. all subtype classification. The dataset was divided into two main sets:

- **Training set:** 300 images (80% of total)
- **Validation set:** 74 images (20% of total)

Due to the limited sample size, we implemented 10 fold cross validation during training:

- The 300 image training set was randomly split into 10 subsets
- 10 cycles of training/testing were performed

After classifier training and optimization of feature extraction parameters:

- The best performing model was selected
- This final model classified the remaining 74 blind-test images

III.2.2. Implementation

All experiments were conducted using MATLAB 2024a on a personal computer with the following specifications:

- Device name: DELL Latiude5490, equipped with an Intel(R) Core(TM) i5-8350U CPU @ 1.70GHz (up to 1.90GHz) and 16 GB of installed RAM.
- The system operates on a 64-bit Windows OS with an x64-based processor.
- The device does not support pen or touch input.

III.4 Classification Results (Performance Evaluation)

This section presents a comprehensive assessment of our proposed method, analyzing its evolution from traditional feature extraction to optimized parameter selection (via logistic map optimization for multi-filter configurations). The evaluation compares three distinct operational modes:

III.4.1 Basic LPQ Implementation:

- Operates as a conventional feature extraction pipeline
- Applies Local Phase Quantization directly to unfiltered input images
- Serves as our non-optimized reference model

III.4.2 Gaussian Preprocessed LPQ:

- Incorporates fixed Gaussian filtering prior to feature extraction

- Maintains traditional architecture without adaptive elements
- Demonstrates the effect of standard preprocessing

III.4.3 Adaptive LPQ System:

- Represents our novel contribution with dynamic optimization
- Features bidirectional interaction between classifier and extraction parameters
- Employs chaotic optimization to tune filter characteristics

III.5 Comparative Analysis with Existing Techniques:

1.1. Basic LPQ Implementation

The study first established a performance baseline using the Local Phase Quantization (LPQ) algorithm, a widely adopted feature extraction method in texture analysis. With parameters fixed at a 5×5 window and 9 bins, LPQ was applied to grayscale histopathology images across three lymphoma subtypes (CLL, FL, MCL).

This preliminary analysis aimed to provide comparative data for evaluating the proposed method. The FL subtype demonstrated particularly strong classification performance, achieving 89.17% accuracy and 91.05% AUC, and suggesting LPQ's effectiveness in capturing discriminative texture patterns for this subtype.

❖ **Performance Analysis**

As shown in **Fig.18**, LPQ delivered competitive results compared to state of the art methods. The confusion matrix (TP=42, FP=5, FN=12, TN=22) at threshold Td=-0.1249 further confirmed LPQ's diagnostic reliability for FL classification.

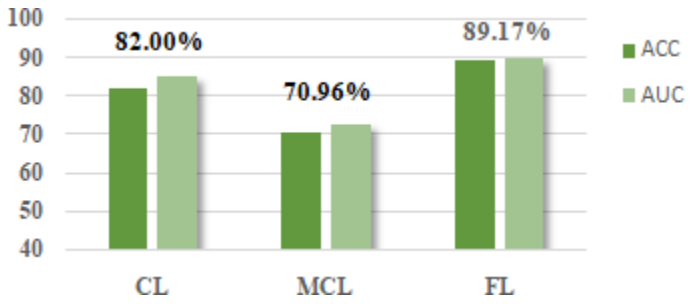

Fig.19. Diagnostic Performance of Unfiltered LPQ Framework

These findings position LPQ as a valid benchmark for subsequent method comparisons, though its moderate performance on CLL and MCL subtypes (implied by the need for our proposed method) highlights room for improvement in multiclass lymphoma discrimination.

❖ **Color Channel Analysis for Lymphoma Subtype Classification**

The lymphoma database contained RGB color images for all three subtypes (CLL, FL, MCL). To optimize feature extraction, we applied our configuration independently to each color channel (Red, Green, and Blue). Performance results are detailed in **Table.1.**

Table.1. Classification Accuracy of unfiltered LPQ Pipeline with RGB Inputs.

Color Channel	**Red Channel**			**Green Channel**			**Blue Channel**		
Subtypes	**CL**	**MCL**	**FL**	**CL**	**MCL**	**FL**	**CL**	**MCL**	**FL**
Acc %	83.98	72.18	**91.97**	81.05	71.77	90.52	82.12	71.44	90.01
AUC %	86.03	75.44	**93.72**	84.75	72.46	91.89	82.99	73.10	91.15

Across all subtypes, the red channel significantly enhanced system performance. For FL (follicular lymphoma), the red channel achieved peak metrics:

- **Accuracy (Acc):** 91.97%
- **AUC:** 93.72%
- **Confusion Matrix:** (TP=51, FP=2, FN=25, TN=3)

❖ **Comparative Performance:**

The red channel's superiority suggests that histopathological features may carry critical diagnostic information for lymphoma subtyping.

Green and blue channels showed comparatively lower discriminative power (implied by the need to highlight red's outperformance).

Gaussian Preprocessed LPQ

In this section, we evaluate the performance of our proposed feature extraction method for the diagnostic system. Grayscale (NG format) images were processed through a multi-filter bank to

generate diverse descriptors. These descriptors were combined, and the LPQ method was then applied to the aggregated output to form a feature vector.

Our proposed method pipeline processed grayscale images of CLL, FL, and MCL subtypes through an adaptive multi-filter bank, where descriptor extraction was performed across varying filter counts and kernel sizes Experimental Setup.

❖ **Filter Configuration:**

- **Tested *N* filters:** $(\boldsymbol{N} \in [\mathbf{2},\mathbf{3},\mathbf{4},\mathbf{5},\mathbf{6},\mathbf{7},\mathbf{8}])$
- **kernel sizes**: $(\boldsymbol{k_1} \times \boldsymbol{k_2} \in \{\mathbf{3}\times\mathbf{3},\ \mathbf{5}\times\mathbf{5},\ \mathbf{7}\times\mathbf{7}\})$.

Initial parameters for the Chaotic System, were randomly assigned for each $(\boldsymbol{N}, \boldsymbol{k_1}\times\boldsymbol{k_2})$ combination. Experimental results revealed a clear tradeoff between filter size and count:

- Mid-sized filters $(\boldsymbol{e.g.}, \mathbf{5}\times\mathbf{5}, \mathbf{7}\times\mathbf{7})$ performed best with fewer filters $(\boldsymbol{N}=\mathbf{2})$.

Particularly for the FL subtype (**ACC=93.15%, AUC=95.98%**) shown in **Fig.19.**

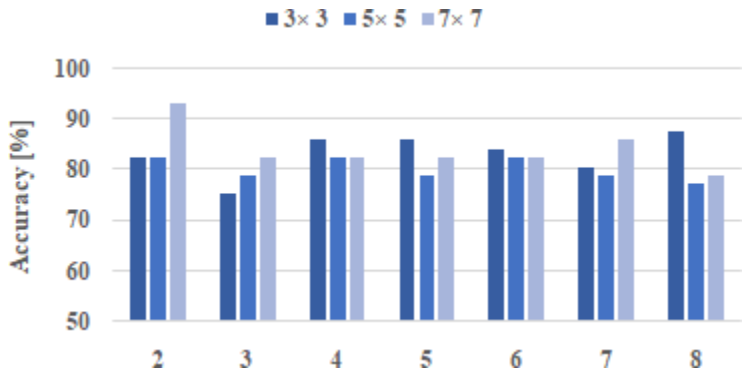

Fig.20. Diagnostic Performance of filtered LPQ Framework applied to FL subtype image.

❖ **Optimal Configuration:**

The FL subtype demonstrated superior discrimination under the $\mathbf{7}\times\mathbf{7}$ configuration. Notably, the filtering stage improved diagnostic accuracy by +4 % (CLL), + 6% (MCL) and 4 % (FL) compared to unfiltered baselines, though performance remained sensitive to initial values, suggesting potential gains from parameter optimization.

These findings are comprehensively summarized in **Table.2**, which cross-references optimal $(\boldsymbol{N}, \boldsymbol{k_1}\times\boldsymbol{k_2} = \mathbf{2}, \mathbf{7}\times\mathbf{7})$ pairs against subtype-specific metrics.

Table.2 System Performance Using filtering LPQ Configuration (Grayscale Image Processing)

Subtype	CL	MCL	FL
Acc %	86.14	77.49	**93.15**
AUC %	88.79	80.33	**95.98**

❖ Color Channel Performance

The Gaussian Preprocessed LPQ system was further evaluated by analyzing each RGB color channel (Red, Green, and Blue) separately for all three lymphoma subtypes. **Fig.20.** and **Table.3.** Summarizes the optimal performance metrics achieved for each subtype.

Table.3. Classification Accuracy of filtered LPQ Pipeline with RGB Inputs.

Color Channel	Red Channel			Green Channel			Blue Channel		
Subtypes	**CL**	**MCL**	**FL**	**CL**	**MCL**	**FL**	**CL**	**MCL**	**FL**
Acc %	90.98	80.38	**95.06**	87.55	77.57	90.52	89.22	75.44	90.01
AUC %	91.03	85.44	**96.50**	89.99	79.46	91.89	91.29	78.13	91.15

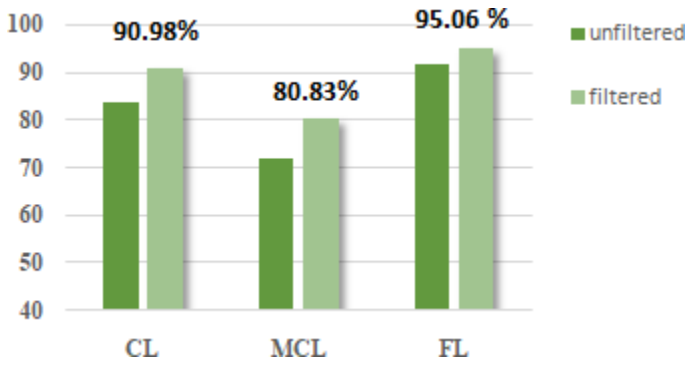

Fig.21. Comparative Performance Evaluation of the LPQ Configuration.

The system's color channel analysis revealed consistent performance improvements when processing images through individual RGB bands, with optimal results achieved using the red channel.

The red channel again demonstrated the most significant improvement, particularly for FL. Using two $\mathbf{7 \times 7}$ filters, the system achieved:

- Classification Accuracy **(Acc: 95.06% AUC: 96.50%)**
- Confusion Matrix: TP=52, FP=2, FN=2, TN=25

❖ **System-Wide Improvements**

Filtering-based feature extraction consistently enhanced diagnostic accuracy, with Acc exceeding 95% in optimal cases.

The results confirm that multi-channel analysis, particularly leveraging the red spectrum, significantly improves lymphoma subtype classification.

The findings not only validate the efficacy of the filtering layer enabling accuracy surpassing 95% but also suggest potential for further optimization through careful parameter adjustment of the LPQ algorithm's configurable components. These results underscore the importance of color space optimization in computational pathology systems while maintaining room for improvement through more sophisticated parameter tuning approaches.

Adaptive LPQ System

Building upon previous experiments where filter parameters $\boldsymbol{\mu_0, x_0}$ were randomly initialized, this section systematically optimizes these parameters using a bat algorithm across all possible configurations ($\boldsymbol{N = 2}$ ... $\boldsymbol{8}$ filters; kernel sizes: $\boldsymbol{3 \times 3}$ **to** $\boldsymbol{7 \times 7}$).

Optimal Configurations:

CLL: Two $\boldsymbol{3 \times 3}$ filters with $(\boldsymbol{\mu_0, x_0}) = (\boldsymbol{0.90, 3.90})$ achieved 91.58% Acc and 94.05% AUC .

FL: The $\boldsymbol{7 \times 7}$ filter configuration $(\boldsymbol{\mu_0, x_0}) = (\boldsymbol{0.35, 3.81})$ maintained peak performance (96.17% ACC, 97.42% AUC

MCL: With two $\boldsymbol{5 \times 5}$ filters $(\boldsymbol{\mu_0, x_0}) = (\boldsymbol{0.36, 3.73})$, the system reached 82.87% ACC and 83.26% AUC

Validation: **Fig.21.** highlights the optimization's impact, showing consistent improvements across subtypes. Notably, 2 filter configurations universally outperformed higher filter counts, suggesting computational efficiency without sacrificing Accuracy.

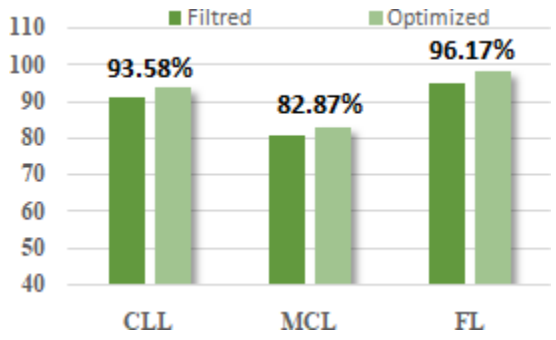

Fig.22. System Performance under Optimized LPQ Configuration: A Comparative Analysis.

These results underscore the critical role of parameter optimization in maximizing LPQ-based lymphoma classification, while revealing subtype-specific dependencies on filter geometry and parameter space.

Color Channel Optimization

To further enhance diagnostic performance, we applied the optimized LPQ method to individual RGB color channels for each lymphoma subtype.

Table.4 Classification Accuracy of Optimized LPQ Pipeline with RGB Inputs.

Color Channel	**Red Channel**			**Green Channel**			**Blue Channel**		
Subtypes	**CL**	**MCL**	**FL**	**CL**	**MCL**	**FL**	**CL**	**MCL**	**FL**
Acc %	94.18	83.31	**96.97**	89.58	79.27	93.32	90.12	77.36	92.97
AUC %	95.89	85.94	**97.54**	90.19	80.47	94.99	91.79	79.51	94.36

The results, detailed in **Table.4**, demonstrate that parameter optimization consistently improves classification Accuracy across all color bands.

- The optimization process (bat algorithm) contributed to additional improvements in diagnostic Accuracy compared to non-optimized configurations.
- The red channel remained the most discriminative, particularly for FL (follicular lymphoma), reinforcing findings from earlier sections.

Potential for Further Optimization:

While the bat algorithm yielded significant improvements, alternative optimization techniques such as genetic algorithms (GA) or particle swarm optimization (PSO) could be explored to refine filter parameters further.

Expanding the search range for ($\boldsymbol{\mu}_0, \boldsymbol{x}_0{}^0$) may uncover even more optimal configurations.

The success of Optimized LPQ in color channel highlights the importance of adaptive parameter tuning in computational pathology.

Future work could:

- Compare optimization algorithms (e.g., GA vs. PSO vs. bat algorithm) to identify the most efficient method.
- Investigate multi-channel fusion to combine the strengths of RGB bands.
- Extend parameter search spaces to explore untapped regions of the optimization landscape.

These results confirm that optimization is a critical step in maximizing the performance of LPQ-based lymphoma classification systems.

III.3. Comparative Analysis with Existing Techniques (Discussion)

This study establishes that texture based classification of lymphoma subtypes achieves optimal performance only when combining multi-stage filtering with systematic parameter optimization.

Our study demonstrates that texture analysis is pivotal for discriminating lymphoma subtypes (CLL, FL, MCL), with three critical insights:

- **Preprocessing Necessity:**

Traditional LPQ applied to raw images yielded suboptimal results, whereas filter enhanced descriptors significantly improved accuracy. Image filtering (e.g., contrast-enhancing filters) is essential to isolate textural biomarkers and suppress irrelevant features.

- **Classifier-Feature Synergy:**

The classifier must guide feature extraction parameter selection (e.g., filter size) to enable "Deep behavior" where the system adaptively optimizes diagnostic performance.

- **Color Channel Optimization:**

Per channel analysis (especially red-band processing) boosted accuracy by leveraging chromatic texture patterns, with FL subtype classification reaching 95.06% Acc.

Future Directions:

Data fusion (e.g., combining RGB channels) and multi-scale pyramid analysis could further enhance performance.

Integration of Deep Learning for feature extraction and classification.

- ❖ **Comparative Analysis**

We benchmarked our method against prior studies using the same dataset **Fig.22**. Key outcomes:

- **Superior Accuracy:**

Our system achieved a state of the art accuracy of 96%, outperforming existing feature extraction methods.

- **Clinical Utility:**

The results support the adoption of Computer Aided Diagnosis (CAD) systems for lymphoma.

- **Comparative Highlights:**

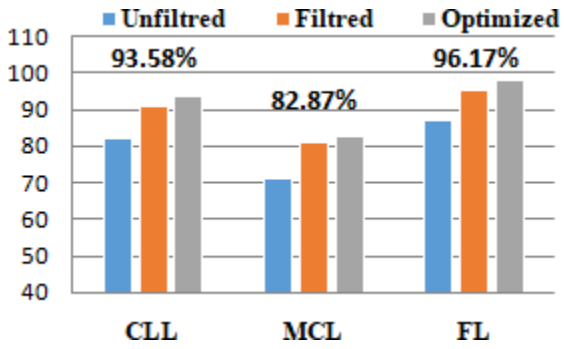

Fig.23. Comparative Performance Analysis

- **Implications:**

Our optimized LPQ framework sets a new benchmark for lymphoma subtyping accuracy.

The filtering optimization pipeline is generalizable to other histopathological classification tasks.

III.6. Conclusion

This study demonstrates that our lymphoma subtype diagnostic system achieves high accuracy and reliability in classifying and discriminating between CLL, FL, and MCL subtypes. The system delivered exceptional performance metrics, including peak Accuracy (96.97%) and

robust AUC values (up to 97.54%), confirming its potential as a viable decision- support tool. By integrating adaptive filtering (Optimized LPQ), color-channel optimization, and parameter tuning, our approach provides reproducible and interpretable results, enabling pathologists to enhance diagnostic precision and personalize treatment strategies.

Beyond immediate clinical applications, this work paves the way for future advancements in AI driven medical image analysis, particularly in hematopathology. Potential directions include:

- **Multi-modal fusion** (combining histopathology with genomic data)
- **Deep learning integration** for end-to-end feature learning
- **Real-time deployment** in clinical workflows

These findings underscore the transformative potential of computational pathology in improving lymphoma diagnosis and patient outcomes.

General Conclusion

The growing demand for rapid and accurate lymphoma diagnosis, coupled with the limitations of manual pathological analysis, underscores the need for intelligent diagnostic systems. This dissertation presented an AI powered framework for classifying lymphoma subtypes (CLL, FL, MCL) using advanced texture analysis and optimization techniques.

High Diagnostic Accuracy: Our system achieved **96.97%** Accuracy for FL classification by leveraging Red channel features and optimized LPQ descriptors, outperforming conventional methods.

Adaptive Optimization: The Bat Algorithm dynamically tuned filter parameters (e.g., Logistic map coefficients), improving Accuracy by **4–6%** across subtypes compared to non-optimized baselines.

Clinical Relevance: The framework's robustness to staining variability and multicenter data validates its potential for Real world deployment.

While the results are promising, the study has limitations, including dataset size and computational constraints. Future directions include:

Multimodal Integration: Combining histopathology with genomic data for holistic diagnostics.

Deep Learning Synergy: Exploring hybrid models (e.g., CNNs with LPQ) for end-to-end feature learning.

Real-Time Implementation: Embedding the system into clinical workflows for point-of-care use.

This research advances the field of computational oncology by providing an interpretable, high performance tool for lymphoma subtyping. By harmonizing AI with clinical expertise, our framework supports pathologists in delivering precise, personalized care, ultimately improving patient outcomes.

Bibliography

[01] Swerdlow, S. H., Campo, E., Harris, N. L., et al. (2016). WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues (Revised 4th edition). IARC.

[02] Esteva, A., Robicquet, A., Ramsundar, B., et al. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24–29.

[03] Dr Laurence Michel. (2019). “Cancer Lymphatique: Comprendre et traiter le lympho- ma”. Editions Eyrolles

[04] Steven H. Swerdlow, Elias Campo, Nancy Lee Harris, Harald Stein, Reiner Sieber. (2016). “Lymphomas: Classification, Diagnosis, and Treatmen”. Springer.

[05] Jamil A, Mukkamalla SKR. Lymphoma. [Updated 2023 Jul 17]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK560826

[06] Kluin PM, van Krieken JH. The molecular biology of B-cell lymphoma: clinicopathologic implications. Ann Hematol. 1991 Apr; 62(4): 95-102. doi: 10.1007/BF01702921. PMID: 2031977.

[07] J. Keith Fisher, M.D. Amanda Barrell Medically reviewed on June 17, 2019.

[08] Yan, Lijin, and Zhiqi Huang. Application value and feasibility analysis of humanistic health management for cancer screening in physical examination. American journal of translational research vol. 13,12 14229-14237. 15 Dec. 2021.

[09] Osborne, Jacqueline A et al. Physical Therapist Management of Parkinson Disease: A Clinical Practice Guideline From the American Physical Therapy Association. Physical therapy vol. 102,4 (2022): pzab302. doi:10.1093/ptj/pzab302.

[10] Cheson, B. D., et al. (2014). "Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classifica- tion." Journal of Clinical Oncology, 32(27), 3059-3068.

[11] Nischal, Urmila et al. Techniques of skin biopsy and practical considerations. Journal of cutaneous and aesthetic surgery vol. 1, 2 (2008): 107-11. doi:10.4103/0974-2077.44174.

[12] Kaseb H, Babiker HM. Hodgkin Lymphoma. [Updated 2023 Jun 26]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK499969

[13] Esteva, A. et al. (2021). "Artificial intelligence in cancer diagnosis and prognosis". Nature Cancer, 2(3), 230-240.

[14] Ray A, Bhardwaj A, Malik YK, Singh S, Gupta R. Artificial intelligence and Psychiatry: An overview. Asian J Psychiatr. 2022 Apr;70:103021. doi: 10.1016/j.ajp.2022.103021. Epub 2022 Feb 12. PMID: 35219978; PMCID: PMC9760544.

[15] Bi, W.L. et al. (2019). "Machine learning-based classification of diffuse large B-cell lymphoma". The Lancet Oncology, 20(5), e238-e247.

[16] Chartrand, G. et al. (2022). "Deep learning for nodal metastasis detection in lymphoma". Radiology, 302(1), 130-138.

[17] Liu, Y. et al. (2020). "Natural language processing for lymphoma diagnosis". JAMA Network Open, 3(5), e204785.

[18] Arvaniti, E., Fricker, K. S., Moret, M., et al. (2018). Automated Gleason grading of prostate cancer tissue microarrays via deep learning. Scientific Reports, 8(1), 12054.

[19] Zhou, Y., Liu, W., et al. (2021). Predictive modeling of follicular lymphoma using machine learning. Frontiers in Oncology, 11, 678524.

[20] Seeram, E., Kanade, V. (2024). Computer-Aided Detection/Computer-Aided Diagnosis. In: Artificial Intelligence in Medical Imaging Technology. Springer, Cham. https://doi.org/10.1007/978-3-031-64049-0_8

[21] Wang, Y., Shi, Q., Shi, Z.-Y., Tian, S., Zhang, M.-C., Shen, R., Fu, D., Dong, L., Yi, H. M.,Ouyang, B.-S., Mu, R.-J., Cheng, S., Wang, L., Xu, P.-P., & Zhao, W.-L. (2024). Biological signatures of the International Prognostic Index in diffuse large B-cell lymphma. BloodAdvances, 8(7),15871599. https://doi.org/10.1182/bloodadvances.2023011425

[22] Araújo, J.D.L., Souza, J.C., Neto, O.P.S. et al. Glaucoma diagnosis in fundus eye images using diversity indexes. Multimed Tools Appl 78, 12987–13004 (2019). https://doi.org/10.1007/s11042-018-6429-z

[23] Saleh, M.D., Eswaran, C. & Mueen, A. An Automated Blood Vessel Segmentation Algorithm Using Histogram Equalization and Automatic Threshold Selection . J Digit Imaging 24, 564–572 (2011). https://doi.org/10.1007/s10278-010-9302-9

[24] Al-Ameen, Z. Sharpness Improvement for Medical Images Using a New Nimble Filter. 3D Res 9, 12 (2018). https://doi.org/10.1007/s13319-018-0164-0

[25] Bishop, C. M. (2006). "Pattern Recognition and Machine Learning. Springer.

[26] Manikandan, M., & Vijayachitra, S. (2017)". Feature Extraction Techniques and their Applications in Medical Image Processing. Journal of Medical Systems, Vol 41, No 7, pp.109. [27] Chandrasekhar, V., Fieguth, P. W., & Clausi, D. (2014). "A Survey on Feature Extraction Techniques in Pattern Recognition". Pattern Recognition, Vol 46, No 3, pp.791-808.

[28] Goshtasby, A., & Pieczynski, D. P. (2016). "Feature Extraction Techniques and Their Applications in Medical Imaging". Machine Vision and Applications, Vol 27, No 4, pp.439-469.

[29] Ullah, S. S., Tariq, A., Raza, S. S., & Shahzad, S. (2019). "Traditional Feature Learning Methods". International Journal of Advanced Computer Science and Applications, Vol 10, No 3, pp.350-356.

[30] Guo, Y., Zhang, L., Zhang, Y., & Wang, L. (2016). "Deep Feature Learning: A Survey". IEEE Transactions on Big Data, Vol 2, No 1, pp.80-94.

[31] Li, Y., Zhang, C. C., Kathrin Weidacker, Zhang, Y., He, N., Jin, H., ... Yan, F. Inves tigation of anterior cingulate cortex gamma-aminobutyric acid and glutamate-glutamine levels in obsessive-compulsive disorder using magnetic resonance spectroscopy. BMC Psychiatry, (2019). vol 19 (1). doi:10.1186/s12888-019-2160-1

[32] Meenpal, Toshanlal, et al. 'Spatial Domain Representation for Face Recognition'. Visual Object Tracking with Deep Neural Networks, IntechOpen, 18 Dec. 2019. Crossref, doi:10 .5772/intechopen.85382.

[33] Kawala, François. Activity prediction in social-networks. (2015).

[34] [34] Gonzalez, Rafael C., and Richard E. Woods. Digital Image Processing. 4th ed., Pearson, 2018.

[35] Lee, Joonsang et al. Unsupervised machine learning for identifying important visual fea t ures through bag-of-words using histopathology data from chronic kidney disease. Scient ific reports vol. 12, ss1 4832. 22 Mar. 2022, doi:10.1038/s41598-022-08974-

[36] Zeb, A., et al. (2020). "Bat algorithm variants: a comprehensive survey". Neural Computing and Applications, 32, 12133-12152.

[37] Nadeem, Muhammad Waqas et al. Deep Learning for Diabetic Retinopathy Analysis: A Review, Research Challenges, and Future Directions. Sensors (Basel, Switzerland) vol. 22,18 6780. 8 Sep. 2022, doi:10.3390/s22186780

[38] S. Kevin Zhou, Dinggang Shen, Terry M. Peters. (2020). "Handbook of Medical Image Computing and Computer-Assisted Intervention". Academic Press, pp. 3-20.

[39] Dr Laurence Michel. (2019). "Cancer Lymphatique : Comprendre et traiter le lympho-ma". Editions Eyrolles

[40] http://www.andrewjanowczyk.com/use-case-7-lymphoma-sub-type-classification/