

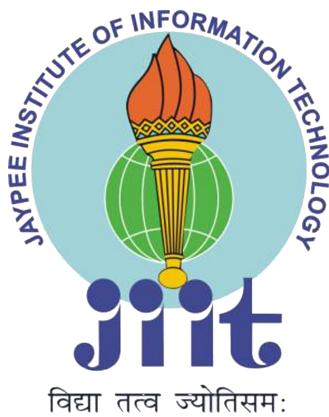
ENHANCEMENT OF BAG-OF-FEATURES METHOD FOR CLASSIFICATION OF HISTOPATHOLOGICAL IMAGES

Thesis submitted in fulfillment for the requirement of the degree of

DOCTOR OF PHILOSOPHY

By

**RAJU PAL
ENROLLMENT NO. 14403008**



Department of Computer Science Engineering and Information Technology

JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY

(Deemed to be University under Section 3 of UGC Act)

A-10, SECTOR-62, NOIDA, INDIA

June, 2019

© Copyright JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY

(Deemed to be University under Section 3 of UGC Act)

A-10, SECTOR-62, NOIDA, INDIA

June, 2019

ALL RIGHTS RESERVED

TABLE OF CONTENTS

	Page Number
INNER FIRST PAGE	i
TABLE OF CONTENTS	ii
DECLARATION BY SCHOLAR	vi
SUPERVISOR'S CERTIFICATE	viii
ACKNOWLEDGEMENT	x
ABSTRACT	xii
LIST OF ACRONYMS & ABBREVIATIONS	xiv
LIST OF FIGURES	xx
LIST OF TABLES	xxiv
CHAPTER-1	
INTRODUCTION	1
1.1 MOTIVATION	2
1.2 HISTOPATHOLOGICAL IMAGES ANALYSIS	3
1.3 CHALLENGES	4
1.4 SCOPE OF THE THESIS	5
CHAPTER-2	
LITERATURE REVIEW	9
2.1 INTRODUCTION	9
2.2 BAG-OF-FEATURES METHOD	12
2.2.1 FEATURE EXTRACTION	12
2.2.2 CODEBOOK CONSTRUCTION	15

2.2.3	FEATURE ENCODING	21
2.2.4	CLASSIFICATION	25
2.3	RESEARCH GAPS AND OBJECTIVES	26
2.3.1	RESEARCH GAPS	26
2.3.2	OBJECTIVES	26
2.4	SUMMARY	27

CHAPTER-3

GREY RELATIONAL BASED KEYPOINTS SELECTION	29	
3.1	INTRODUCTION	29
3.2	BAG-OF-FEATURES METHOD	30
3.3	GREY RELATIONAL ANALYSIS BASED KEYPOINTS SELECTION	34
3.4	EXPERIMENTAL RESULTS	36
3.4.1	DATASETS	36
3.4.2	PERFORMANCE ANALYSIS OF THE KEYPOINTS SELECTION	38
3.4.3	CLASSIFICATION RESULTS OF GKS BASED BOF METHOD	40
3.4.4	COMPARATIVE ANALYSIS OF GKS BASED BOF WITH STATE-OF-THE-ART METHODS	45
3.5	SUMMARY	47

CHAPTER-4

OPTIMAL CODEBOOK CONSTRUCTION	49	
4.1	INTRODUCTION	49
4.2	BIOGEOGRAPHY-BASED OPTIMIZATION ALGORITHM	50
4.2.1	IMPROVED BIOGEOGRAPHY-BASED OPTIMIZATION	54
4.2.2	SPIRAL BIOGEOGRAPHY-BASED OPTIMIZATION	57
4.3	OPTIMAL CODEBOOK CONSTRUCTION	59
4.4	EXPERIMENTAL RESULTS	61
4.4.1	PERFORMANCE ANALYSIS OF IBBO AND SBBO	61
4.4.2	PERFORMANCE ANALYSIS OF SBBO BASED HISTOPATHOLOGICAL IMAGE CLASSIFICATION	69
4.5	SUMMARY	74

CHAPTER-5

EFFICIENT CODEBOOK CONSTRUCTION	75
5.1 INTRODUCTION	75
5.2 GREY RELATIONAL ANALYSIS BASED BAG-OF-FEATURES	75
5.3 EXPERIMENTAL RESULTS	78
5.4 SUMMARY	84

CHAPTER-6

WEIGHTED TWO DIMENSIONAL VECTOR QUANTIZATION	85
6.1 INTRODUCTION	85
6.2 PRELIMINARIES	86
6.3 WEIGHTED TWO DIMENSIONAL VECTOR QUANTIZATION	92
6.4 EXPERIMENTAL RESULTS	95
6.4.1 PERFORMANCE ANALYSIS ON ADL DATASET	95
6.4.2 PERFORMANCE ANALYSIS ON BLUE HISTOLOGY DATASET	98
6.4.3 CLASSIFICATION ACCURACY	100
6.5 SUMMARY	101

CHAPTER-7

ENHANCED BAG-OF-FEATURES METHOD	103
7.1 INTRODUCTION	103
7.2 FEATURE EXTRACTION USING SIFT AND ORB	103
7.3 CODEBOOK CONSTRUCTION USING GRA	104
7.4 FEATURE ENCODING USING W2DVQ METHOD	105
7.5 CLASSIFICATION	105
7.6 EXPERIMENTAL RESULTS	106
7.7 SUMMARY	107

CHAPTER-8

CONCLUSION AND DIRECTION TO FUTURE WORK	109
8.1 RESEARCH CONTRIBUTION	109
8.2 DIRECTION TO FUTURE WORK	110

REFERENCES	111
LIST OF PUBLICATIONS	135
SYNOPSIS	Synopsis-1

DECLARATION BY THE SCHOLAR

I hereby declare that the work reported in the Ph.D. thesis entitled "**Enhancement of Bag-of-Features Method for Classification of Histopathological Images**" submitted at **Jaypee Institute of Information Technology, Noida, India**, is an authentic record of my work carried out under the supervision of **Dr. Mukesh Saraswat**. I have not submitted this work elsewhere for any other degree or diploma. I am fully responsible for the contents of my Ph.D. thesis.

(Signature of Scholar)

(Raju Pal)

Department of Computer Science Engineering and Information Technology,
Jaypee Institute of Information Technology, Noida, India.

Date:

SUPERVISOR'S CERTIFICATE

This is to certify that the work reported in the Ph.D. thesis entitled "**Enhancement of Bag-of-Features Method for Classification of Histopathological Images**", submitted by **Raju Pal** at **Jaypee Institute of Information Technology, Noida, India**, is a bonafide record of his original work carried out under my supervision. This work has not been submitted elsewhere for any other degree or diploma.

(Supervisor)

(Dr. Mukesh Saraswat)

Associate Professor,

Department of Computer Science Engineering and Information Technology,
Jaypee Institute of Information Technology, Noida, India.

Date:

ACKNOWLEDGEMENTS

I wish to express my deep gratitude to my supervisor Dr. Mukesh Saraswat. His precise and lucid thoughts have helped me to manage many responsibilities comfortably. He not only encouraged me to work but also gave his guidance, experience, constructive thoughts and took keen interest throughout the course of work and preparation of manuscript which had made me worth that I claim to be now. Working with him made me more independent and liable, which I believe is the most important aspect in anyone's career. I am indebted to him more than he recognize.

No appropriate words could be traced in the presently available lexicon to avouch the excellent guidance given by Prof. S. C. Saxena, Vice-Chancellor, Prof. Hari Om Gupta, Director, Prof. D. K. Rai, Dean Academic, Prof. R. C. Mittal, Dean Research, Dr. Prof. S. P. Purohit, Controller of Examinations, and Prof. Vikas Saxena, Head of Department, Jaypee Institute of Information Technology, Noida for their constant support and full co-operation. Moreover, I would like to thank my research committee members and faculty members of the Institute for their help and support.

I am thankful to the Science and Engineering Research Board, Department of Science & Technology, Government of India, New Delhi, India for funding this work as a part of the project (ECR/2016/000844). I am also thankful to Dr. J. C. Bansal, Dr. Harish Sharma, and Dr. Prashant Singh Rana for their suggestions and encouragement in problem domain had exceptionally enriched my growth. I am also thankful to Mr. Roop Singh Solanki to become part of the funded project.

I am extremely thankful to Dr. Vinay Lomash who has given his constant guidance, support and encouragement throughout this study. His expertise into the pathology, experience, vigorous efforts and co-operation had made me to complete this work. Without his unfailing support and belief in me, this thesis would not have been possible.

I am extremely thankful Dr. Charu Gandhi, Prof. Krishna Asawa, Dr. Anuj Bhardwaj, Dr. Pankaj Srivastava, Dr. Dinesh Bisht, Dr. Shashi Bhushan, Dr. A. K. Malik, Dr. Suryakant, Dr. Aruna Malik, and Dr. Aarti for their constant guidance, support and encouragement throughout this study. Their expertise, valuable suggestions and co-operation had made me complete this work.

I am sincerely thankful to my friends Mr. Himanshu Mittal, Mr. Ashish Kumar Tripathi, Mr. Ankur Kulhari, Mr. Avinash Pandey, Mr. Pushpendra Kumar Rajput for their scholastic guidance, suggestions and encouragement throughout the study duration. Along with them, I am also thankful to Mr. Sudhanshu Kulshrestha, Mrs. Anuradha Gupta, Mrs. Parul Agrawal, Mrs. Vartika Srivastva, Mrs. Bindu Verma and other colleagues for providing their kind helps and supports during the time of research.

I would also like to acknowledge the help rendered by Mr. Kirpal Singh, Mr. Abhishek Verma, Mr. Netra Pal, Mr. Gavendra Singh, Mr. Prashant Kumar, and Mr. Upendra Dubey for their day-to-day assistance and contribution, which made my work easier. It is my great privilege to mention the entire official staff for helping me in the official documentation work right from the time of my registration into Ph.D.

I would like to offer my most humble gratitude to my parents (Shri Brijmohan Pal and Smt. Yashoda Pal), my sister Rajni, my nephews Aarush and Aadwik, my wife Surbhi, and other family members for their unsurpassed love, care, endless patience and countless sacrifices which they made for me. Indeed, a plethora of words would not suffice to say what I owe to them.

And above all I am blessed by Lord Shri Salasar Balaji Maharaj for giving me the strength and perseverance to toil towards this accomplishment without any hardships.

Raju Pal

ABSTRACT

Image classification is a process by which a machine can categorize the images into a finite set of classes. This process is applied to many important applications like robotics, remote sensing, optical character recognition, and medical image analysis. Histopathological image classification is also one of the prime areas of medical imaging in which changes in the tissue images are studied. However, due to the complex texture of histopathological images, it is generally done manually by the pathologists. Manual analysis is a time consuming and biased process. Therefore, in literature, several histopathological image classification methods are presented, out of which bag-of-features is one of the popular image categorization methods. However, due to the various morphological variability, available in the histopathological images, the bag-of-features method also lack in the classification accuracy and needed to be enhanced. Therefore, in this work, enhanced bag-of-features method is designed and developed for efficient histopathological image classification.

The work of this thesis is divided into five folds. First, a new computationally efficient keypoint selection phase is introduced in the bag-of-features method. Secondly, two new variants of biogeography-based optimization algorithms are proposed and used to construct an optimal codebook. However, due to the high computation cost of biogeography-based optimization algorithms, the third part presents a new grey relational analysis based clustering method for obtaining the an efficient codebook. Fourth, a weighted two-dimensional vector quantization method has been presented for encoding the images. Finally, an enhance bag-of-features method for classification of histopathological images is designed and developed.

Keywords: Histopathological image classification, Bag-of-features, Keypoints selection, Codebook construction, Biogeography-based optimization, Feature encoding.

LIST OF ACRONYMS & ABBREVIATIONS

(Alphabetically)

ACO	Ant Colony Optimization
ADL	Animal Diagnostics Lab
AKM	Approximate K-Means
AzB	Alizarin Blue
BBO	Biogeography-based Optimization
BBO-M	Biogeography-based Optimization with Mutation Strategies
BBOPE	Biogeography-based Optimization with Partial Emigartion
BBOPI	Biogeography-based Optimization with Partial Immigartion
BBOTE	Biogeography-based Optimization with Total Emigartion
BBOTI	Biogeography-based Optimization with Total Immigartion
BC	Biocytin
BIRCH	Balanced Iterative Reducing and Clustering using Hierarchies
BOF	Bag-of-Features
BPLR	Boundary Preserving Dense Local Regions
BRIEF	Binary Robust Independent Elementary Features
BRISK	Binary Robust Invariant Scalable Keypoints
CCY	Carbocyanine
CEC	Congress of Evolutionary Computation
CNN	Convolutional Neural Networks
CT	Connective Tissue
CURE	Clustering using Representatives
DE	Differential Evolution
DE-TCR	Enhanced Differentail Evolution

DHCDC	Divisive Hierarchical Clustering with Diameter Criterion
DIVCLUST	Divisive clustering
DNN	Deep Neural Networks
DOG	Difference of Gaussians
DROP3	Reduction Techniques for Instance-based Learning
EA	Evoltionary Algorithm
EBR	Edge Based Regions
ECOC	Support Vector Machine Classifier using Error Correcting Codes
EL	Elastin
ES	Evolutionary Strategy
ET	Epithelial Tissue
FAST	Features from Accelerated Segment Test
FCM	Fuzzy C-Means
FCS	Fuzzy C-Shells
FLOG	Fan Laplacian of Gaussian
FNR	False Negative Rate
FPR	False Positive Rate
FREAK	Fast Retina Key-Point
FV	Fisher Vector
GKS	Grey Relational Analysis based Keypoints Selection
GLOH	Gradient Location and Orientation Histogram
GMM	Gaussian Mixture Model
GNN	Gaussian Naive Baise
GRA	Grey Relational Analysis
GRC	Grey Relational Coefficients
GRG	Grey Relational Grades
GSA	Gravitational Search Algorithm
GWO	Grey Wolf Optimization

H&E	Hematoxylin and Eosin
HIA	Histopathological Images Analysis
HIC	Histopathological Images Classification
HOG	Histogram of Oriented Gradients
HV	Hard Voting
IB3	Instance Based Learning
IBBO	Improved Biogeography-based Optimization
IBR	Intensity-based Regions
ICC	Immunocytochemistry
ICF	Iterative Case Filtering
IF	Immunofluorescence
IH	Iron Haematoxylin
IHC	Immunohistochemical
IKS	Iterative Keypoint Selection
ISI	Island Suitability Index
KCB	Kernel Codebook Encoding
LBG	Linde–Buzo–Gray Algorithm
LBP	Local Binary Pattern
LCC	Locality Coordinate Coding
LFC	Luxol Fast Blue, Cresyl Violet
LLC	Local constrained Linear Coding
LR	Logistic Regression
LSA	Localized Soft Assignment
LSHADE	Lightweight Success-History based Adaptive Differential Evolution Algorithm
LTC	Local Tangent-Based Coding
LVQ	Learning Vector Quantization
LX-BBO	Laplace Crossover Based Biogeography-Based Optimization
MB	Methylene Blue

MI-LXPM	Improved Real Coded Genetic Algorithm
MM	Multi-Modal
MSER	Maximally Stable Extremal Regions
MT	Muscle Tissue
NS	Non Separable
NT	Nervous Tissue
OMP	Orthogonal Matching Pursuit
ORB	Oriented Fast and Rotated Brief
PAS	Periodic Acid–Schiff
PCA	Principal Component Analysis
PDE	Partial Differential Equation
PSO	Particle Swarm Optimization
RAM	Random Access Memory
RBF	Radial Basis Function Kernel
RBM	Restricted Boltzmann Machines
RCC	Renal Cell Carcinoma
RE	Immigration Refusal
RET	Reticulin
RF	Random Forest
RGB	Red Green Blue
SA	Soft Assignment
SBBO	Spiral Biogeography-based Optimization
SGFNT	Significant Level
SHIRC	Simultaneous Sparsity Model for Histopathological Image Representation and Classi
SIFT	Scale-Invariant Feature Transform
SIV	Suitability Index Variable
SMO	Spider Monkey Optimization
SPC	Sparse Coding

SRC	Sparse Representation based Classifier
SSA	Salt Swarm Algorithm
SSAE	Stack Sparse Auto Encoder
SURF	Speeded Up Robust Features
SVC	Super Vector Coding
SVM	Support Vector Machine
TB	Toluidine Blue
TN	True Negative
TP	True Positive
TRI	Trichrome
UM	Unimodal
VG	Van Gieson
VLAD	Vector of Locally Aggregated Descriptors
VQ	Vector Quantization
W2DVQ	Weighted Two Dimensional Vector Quantization
WHP	Whipf's Polychrome
WOA	Whale Optimization Algorithm
WSI	Whole-Slide Image

LIST OF FIGURES

Figure Number	Caption	Page Number
1.1	Histopathological images	2
1.2	The work flow of the proposed research work	7
2.1	General work flow of an automated histopathological image classification method	9
2.2	Categorization of histopathological image classification techniques based on image representation	10
2.3	Visual description of voting based methods	23
3.1	The architecture of the BOF method	31
3.2	The keypoints detected by SURF in connective tissue and inflamed lung histopathological images	32
3.3	Flow chart of the enhanced BOF method	34
3.4	Healthy and inflammatory tissues of Kidney, Lung, and Spleen	37
3.5	Representative animal tissues from blue histology dataset at $40\times$ magnification level	38
3.6	Classification accuracy on validation set using the GKS method with different shrinking threshold values	39
3.7	Classification accuracy on validation set using the GKS method with different classifier on ADL and Blue histology datasets	41
3.8	The confusion matrices for the ADL dataset, generated by IB3, IKS1, IKS2, and GKS based classification methods	41
3.9	The confusion matrices for the Blue histology dataset, generated by IB3, IKS1, IKS2, and GKS based classification methods	42

3.10	Radar charts for average results obtained for SVM classifier on ADL dataset and Blue histology image dataset by considering F-1 score, sensitivity, specificity, and G-mean	45
4.1	Flow chart of the BOF method with modified codebook construction	50
4.2	Linear equilibrium model of an ecological system for a single island	51
4.3	The Fermat's Spiral Function in 2D and 3D with radius r (0 to 10) angle θ	58
4.4	3D map for multi-modal and composition based benchmark problems	63
4.5	Convergence trend of IBBO and SBBO with other considered Meta-heuristics standard benchmark problems	68
4.6	Convergence trend of IBBO and SBBO with other considered Meta-heuristics on CEC 2017 benchmark problems	69
4.7	The normalized confusion matrix for the blue histology tissue dataset for the five methods BOF, IB3, IKS2, BBO-M-BOF, and SBBO-BOF	70
4.8	Radar chart for average results obtained by SVM classifier on Blue histology dataset by considering F1 score, G-mean, sensitivity, and specificity	72
4.9	The overall classification accuracy of the considered methods for ADL dataset in form of bar chart	73
4.10	Radar chart for average results on Spleen dataset, Kidney dataset, and Lung dataset	73
5.1	The GRA based BOF method	76
5.2	Impact analysis of cut-off threshold value over classification accuracy on ADL and Blue histology datasets	80
5.3	Confusion matrices returned by SVM, SRC, SHIRC, BOF, IKS2, and GRA-BOF on Blue histology image dataset	81
5.4	Radar chart for average results on Spleen dataset, Kidney dataset, and Lung dataset	83
6.1	The overall procedure of the BOF method based on W2DVQ method	93
6.2	Radar chart for average results on Spleen dataset, Kidney dataset, and Lung dataset	98

6.3	Confusion Matrix returned by the BOF method using different combinations of features on Blue Histology dataset	98
6.4	Radar chart for average results obtained for SVM classifier on Blue histology dataset by considering F1 score, G-mean, sensitivity, and specificity	100
7.1	Developed histopathological image classification system	104
7.2	The confusion matrix returned by GRA-weighted-BOF method on Blue histology image dataset	106

LIST OF TABLES

Table Number	Caption	Page Number
2.1	Categorization of features extraction methods used in bag-of-features	13
2.2	Various codebook construction methods used in the bag-of-features	15
2.3	Various types of feature encoding methods	21
2.4	Popularly used classification methods for histopathological image classification	25
3.1	The different categories of histology image dataset and their staining methods	38
3.2	Number of selected keypoints returned by IB3, IKS1, IKS2, and GKS on considered datasets along with their average computational cost	40
3.3	Comparative analysis of the GKS based BOF method with other considered methods for ADL dataset in terms of various performance parameters	43
3.4	Comparative analysis of the new GKS based BOF method with other considered methods on blue histology tissue image dataset in terms of of various performance parameters	44
3.5	Classification performance of GKS based BOF method with other methods	46
4.1	Standard benchmark functions	61
4.2	Overview of CEC 2017 benchmark problems	62
4.3	Mean fitness values returned by IBBO, SBBO and other methods on the standard benchmark functions (F_1 to F_{10})	65
4.4	Mean fitness values returned by IBBO, SBBO and other methods on the standard benchmark functions (F_{11} to F_{20})	66
4.5	Mean fitness values returned by IBBO, SBBO, and other methods on the CEC 2017 benchmark problems	67

4.6	Mean ranking of considered algorithms on standard benchmark problems with 30, 60, and 90 dimensions using Friedman Test	68
4.7	Mean ranking of considered algorithms on CEC 2017 benchmark problems using Friedman Test	68
4.8	Comparative analysis of the SBBO-BOF based classification method	71
4.9	The confusion matrix for kidney image classification obtained by the SBBO-BOF method	72
4.10	The confusion matrix for lung image classification obtained by the SBBO-BOF method	72
4.11	The confusion matrix for spleen image classification obtained by the SBBO-BOF method	73
5.1	Confusion matrix generated by considered methods on ADL histopathological image dataset	80
5.2	Classification performance with other methods based on recall, specificity, precision, FPR, and accuracy	82
5.3	Classification performance with other methods based on recall, precision, F1 score, specificity, and average accuracy on Blue histology dataset	83
6.1	Confusion matrices returned by the BOF method on Kidney images with HV and W2DVQ encoding methods	96
6.2	Confusion matrices returned by the BOF method on Lung images with HV and W2DVQ encoding methods	96
6.3	Confusion matrices returned by the BOF method on Spleen images with HV and W2DVQ encoding methods	96
6.4	Comparative analysis of considered methods on ADL dataset in terms of various performance parameters	97
6.5	Comparative analysis of the considered methods on Blue Histology dataset in terms of various performance parameters	99
6.6	Comparative analysis of average classification accuracy the proposed W2DVQ method and other considered methods	101

7.1	The confusion matrix returned by GRA-weighted-BOF method on ADL histopathological image dataset	106
7.2	The performance comparison of various proposed BOF methods on ADL and Blue histology dataset	107

CHAPTER 1

INTRODUCTION

Image classification process categorizes various images into their corresponding classes. A large number of application areas uses image classification as one of the prime phases such as medical imaging, robotics, computer vision, and many more. However, due to the variations available in the images, a single classification process cannot be used for all the application areas. Out of many applications, medical imaging is of prime importance which directly influences the society and its automation helps for unbiased analysis of various diseases. Histopathological image analysis also is one of the important areas of medical imaging which is generally performed manually by the pathologists. It is a necessary part of drug development in the pharmaceutical companies and also used by pathologists for disease identification.

Histopathology is the investigation of tissues to identify the symptom of abnormality. This process is generally done by trained pathologists manually in the pathology labs. Generally, tissues are colorless, therefore, various staining methods are used to provide the colors. Among various staining methods, hematoxylin and eosin (H&E) staining is widely used. It provides a blue color to the nuclei and pink or red color to the cytoplasm. Due to the advancements in digital electronics, pathology labs are going to be transformed into the digital era [1]. The high-quality whole-slide image (WSI) scanners capture the images at very high resolutions from the tissue samples. Figure 1.1 depicts two sample histopathological images having healthy and inflamed tissues respectively. The digitization of tissue slides has substantially made accurate decisions for disease diagnosis [2]. The pathologists can view the tissue images on computer monitors but, still the analysis of the tissue sections is very time taking and biased due to the experience and knowledge of the pathologists. Therefore, there is a requirement to develop an automated method for histopathological image classification. It is a promising area that acquires the attention of the researchers [3]. These automated systems help the pathologists in the decision-making process for the diagnosis of the disease [4]. The improvement of computer

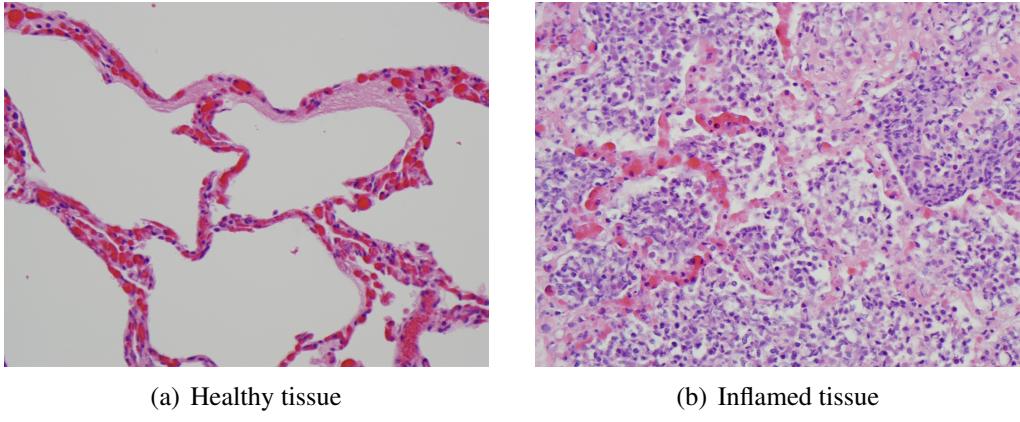


Figure 1.1: Histopathological images [2]

technologies and digitization of histopathological images further increase the need for such automated systems for the pathologists. Therefore, this work focuses on the development of significant automated methods for histopathological image classification.

This chapter presents the research motivation followed by a brief introduction to histopathological image analysis, the challenges of histopathological image classification, and the organization of the thesis.

1.1 Motivation

The past decade has shown an exponential increase in computational efficiency along with better image analysis techniques. These have insisted to develop the powerful machine learning algorithms for medical image analysis. Furthermore, the development of whole slide scanners for histopathological tissue slides, the digital histopathological images are used for analyses. However, due to the various variabilities and complexities available in the images, this analysis is performed manually by the expert pathologists. The manual histopathological image analysis is time taken process and also biased. The report generated by a pathologist is generally dependent on his knowledge and experience. Therefore, the automation of this task helps the pathologists to work efficiently with machine intelligence. Secondly, by automatically identifying the inflamed images, the observer variability can be eliminated and this can also improve the accuracy of disease diagnosis. However, the automation of the classification of histopathological images is a difficult process due to the morphological complexity and technical variations of histopathological images such as different organ types, variations of devices used, the ori-

tation of slides, and different staining methods [5]. Moreover, histopathological images contain an enormous density of data as compared to radiological and other imaging modalities. This makes the descriptor design task very complex which affects the performance of an automated image classification system [3]. Therefore, this work aims to develop an improved system to classify histopathological images and resolve constraints related with existing methods.

1.2 Histopathological Images Analysis

The histopathological image analysis is used in the development of drugs and the identification of the diseases. Regardless of the popularity of modern diagnostic tools in the biomedical field, histopathological image analysis is still performed manually. The manual analysis of these images takes lots of time and efforts of the dedicated pathologists. Some major problems of manual analysis of these images are enlisted below:

- **Scarcity of expert pathologists:** Due to the advancements in medical technology, pathology labs generate a lot of data every year and it is required to handle this data daily by the pathologists. Hence, there is a high demand for expert pathologists all over the world. In Australia, a shortage of pathologists impacts the international workforce as acknowledged by the advisory committee of the Australian Medical Workforce [6].
- **Variability and biased observations [3] [7]:** The experience and knowledge of different pathologists may vary which impact the opinion given by them for the same histopathological image. These variations may result in wrong treatment and sometimes delayed treatment.
- **Complexity and variability in biological structures among tissue slides [3]:** Histopathological images contain various morphological structures and complexities such as the use of different staining dyes, illumination variations at the time of capturing the images, different shapes and sizes of nuclei, and many more. It is very hard for the pathologists to analyze all these variabilities.
- **Consumption of time and efforts:** The manual analysis of a histopathological image requires at least three to four hours of sitting and evaluation of quantitative properties makes it a more complex task. Further, if the images are captured at a high magnification rate then more time is needed by the pathologists to analyze the whole slide.

To overcome the shortcomings mentioned above, the automated analysis of histopathological images may be used. An automated image analysis system can provide meaningful data more accurately and timely for the histopathological images. This reduces the workload on the pathologists and makes less human subjectivity. This process enhances the manual analysis process by generating reliable and fast results. Recent advancements in the computer vision and image analysis open up various opportunities to design and develop an automated image classification system for histopathological images. These systems help in making unbiased, efficient, and accurate analysis reports promptly and perform automated classification of different tissue types which can be accessible widely as a tool for research.

Therefore, the development of automated histopathological image classification methods has emerged as a significant research problem in medical imaging [3][2][8]. However, there are many problems associated with the development of such automated system. The following section discusses the various challenges in the automation of histopathological image classification.

1.3 Challenges

Though many histopathological image classification methods are available in the literature which show quite promising results, still there exists some challenges which affect the accuracy of the system. This needs the attention of the researchers for improvements so that the expectations of the pathologists can be fulfilled.

- 1. Variations in the color and illumination:** Histopathological images are the colorful tissue images after staining and captured by the whole slide scanners. The color provided by staining methods may vary as it is a manual process and similarly, there may be variations in the illumination light while capturing the images. These color and illumination variations may degrade the performance of the automated system significantly [9].
- 2. Different methods for staining:** Most of the works for histopathological image analysis have been done on tissue section images stained with some chemicals like immunohistochemical (IHC), immunofluorescence (IF), and hematoxylin & eosin (H&E). These staining procedures are very expensive and time-taking. The use of multiple staining methods for the preparation of histopathological images makes the analysis difficult for the automated methods [3] [10].

3. **Enormous density of data:** Another challenge in the automated histopathological image analysis is the huge density of the data as compared to other biomedical images like tomography, radiology, and other image modalities, which has to be contended by the automated methods. For example, the chest CT scan, captured on high resolution, consists of approximately 143 million pixels with $512 \times 512 \times 512$ spatial elements. On another hand, the biopsy tissue of the prostate, captured at $40\times$ resolution by a whole slide image scanner, consists of approximately 235 million pixels with $15,000 \times 15,000$ elements. Moreover, for one patient study, there is a requirement of 10 to 20 biopsy samples which results in generating a huge data of approximately 3 to 4 billion pixels. Therefore, unlike the automated methods proposed for radiology and other medical imaging, automated histopathological image analysis methods are usually built in such a way that they have to perform efficiently and accurately on the high density of data.
4. **Multimodal data fusion:** Histopathology is generally used to study the cancer disease in the patients but the diagnosis and prognosis of the cancer disease are very hard. If two patients are going through the same treatment procedure with the same disease may result in different outcomes. The reason for this difference may be patient specific or it may be due to the lack of related information between the progression of the disease and clinical aspects. There is a common consent between scientists and pathologists that the understanding of tumor visual morphology using automated image analysis methods along with the disease classification will result in better treatment and patient care. Therefore, multimodal data fusion has emerged as a relevant challenge for digital pathology labs for making the recommendation about the patient treatments.
5. **Maintenance of image data:** Due to the advancements in digital imaginary tools and image acquisition methods, a large amount of data is generated. The storage, registration, maintenance, and transmission of such a big volume of data is a challenging task. The spectral imaging data from the whole slide scanner makes the problem more complex.

1.4 Scope of the Thesis

For histopathology, computer vision, and biomedical scientists, the above-mentioned challenges will open various research problems and opportunities to develop new image analysis methods.

The pathologists and computer scientists work together in collaboration with microscopy and slide scanner vendors to build innovative and novel methods to solve various challenges of image analysis in digital pathology. Furthermore, in recent years, there is a rapid increase of histopathological digital images over the Internet and there is a need to organize them properly for better retrieval and analysis processes for the researchers. Therefore, an automated histopathological image classification system can be useful. In literature, several automated histopathological image classification methods exist which are based on the approaches like a graph, hashing, bag-of-features, and deep neural networks. The Bag-of-features method is one of the well-liked image classification methods and shows better performance when applied to histopathological images. However, there is a requirement for improvement in the different phases of the bag-of-features method. The main objective of this thesis is to design and develop an efficient histopathological image classification system based on the bag-of-features method. The main contribution of the thesis is five-fold. It first aims to develop an efficient keypoints selection method to find the relevant and discriminative features from histopathological images. Second, to reduce the effect of the dense region in the codebook construction phase of the bag-of-features method, a meta-heuristic based codebook construction method has been introduced. Third, a new and computationally efficient codebook construction method is designed and developed to find relevant visual words. Fourth, an efficient feature encoding method has to be designed and developed by incorporating the merits of two different features descriptors for the better image representation and finally, an enhanced bag-of-features method is developed by incorporating the newly introduced methods for the classification of histopathological images. The complete workflow of the thesis is shown in Figure 1.2. The rest of the material is arranged in the following chapters.

Chapter 2 surveys the existing methods for histopathological image classification along with a detailed literature survey presented for the different phases of the bag-of-features method. Based on the outcome of this study, research gaps are identified and documented along with the corresponding research objectives.

Chapter 3 deals with the problem of a large number of keypoints descriptors generated by feature extraction method for histopathological images. This chapter presents a new keypoints selection method based on a grey relation analysis for finding relevant and useful features. The

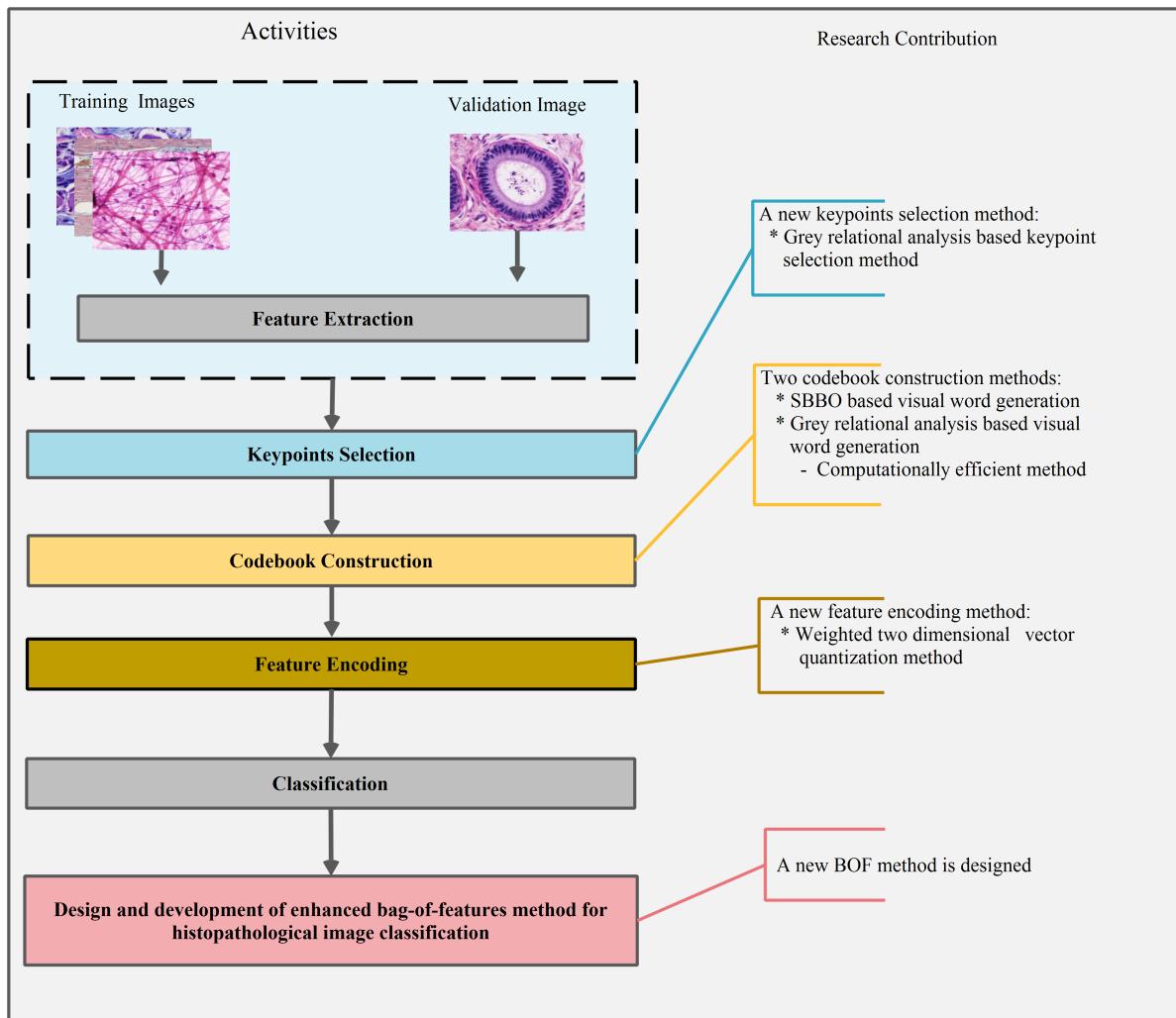


Figure 1.2: The work flow of the proposed research work

method is introduced before the codebook construction phase of the bag-of-features method. The new method is validated against other state-of-the-art keypoints selection methods. For a fair comparison, two standard histopathological image datasets are used.

Chapter 4 studies the biogeography-based optimization, as it has been used effectively to classify the medical images as compared to other nature-inspired methods available in the literature. However, sometimes it shows a slow convergence rate and poor precision due to its single feature migration property and poor mutation. Therefore, this chapter presents two variants of biogeography-based optimization, namely improved biogeography-based optimization and spiral biogeography-based optimization, to mitigate its demerits. With the experimental and statistical analysis over two benchmark sets, the new variants have been compared with recent meta-heuristic methods. The new methods have been used to develop an effective codebook in

the bag-of-features methods and reduce the effect of dense regions of histopathological images. The modified bag-of-features method is compared with other state-of-the-art methods over two standard histopathological image datasets.

Chapter 5 presents the grey relational analysis based codebook construction method and respective result analysis. The new codebook construction method is computationally efficient than the meta-heuristic based method of Chapter 4 for codebook construction.

Chapter 6 introduces a weighted two-dimensional vector quantization method based on the information of two different feature descriptors in the bag-of-features framework. The proposed method encodes the images in the feature encoding phase of bag-of-features using a weighted two-dimensional representation of two different features. The work is tested on two histopathological image datasets for the classification tasks.

Chapter 7 presents the developed automated histopathological image classification system using the above mentioned methods which identify the labels of histopathological images, acquired at $40\times$ magnification.

The last **chapter** concludes the thesis and summarizes the research contributions and possible research directions for the future work.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The general framework of an automated histopathological image classification system, depicted in Figure 2.1, mainly composed of two phases, namely training phase and validation phase. The first phase is the training phase in which the sampled histopathological images are selected randomly from the database and converted into the feature vectors by an image representation method. The converted feature vectors along with the image labels are fed to the learning model (i.e., classifier) for training. The trained learning model is used in the validation phase in which a histopathological image is selected from the image database and is converted into a feature vector by applying the same image representation method. The obtained feature vector is passed to the trained learning model which predicts the label of the image. Therefore, in the classification of histopathological images, the image representation method plays an important role and the accuracy of a system is highly depends upon its performance. On the basis of image

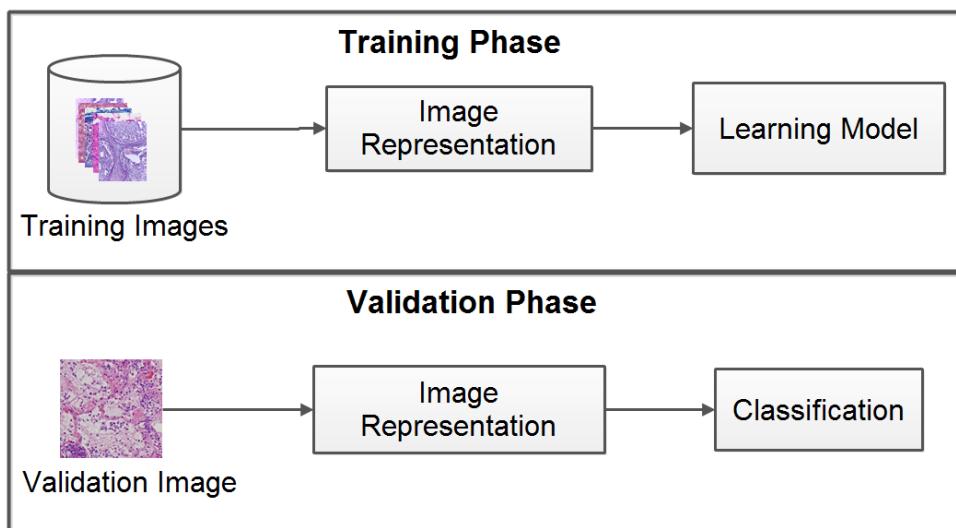


Figure 2.1: General work flow of an automated histopathological image classification method

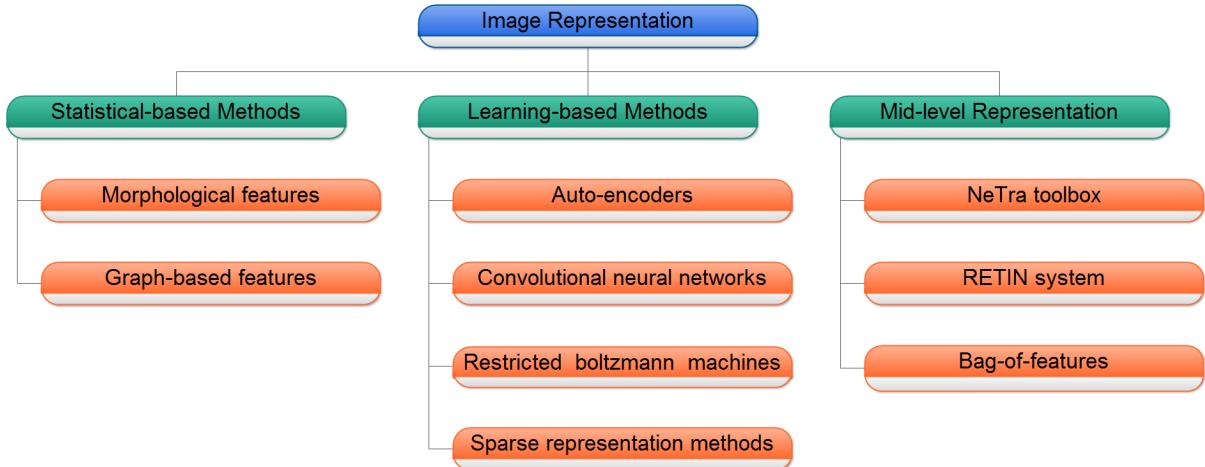


Figure 2.2: Categorization of histopathological image classification techniques based on image representation

representation techniques, the histopathological image classification methods can be divided into three categories, namely statistical-based methods, learning-based methods, and mid-level representation based methods as shown in Figure 2.2.

Statistics-based image representation methods extract the low-level or local features from the histopathological images based on pixel-level operations. The features provide enthralling details of the medical image contents for the classification task without the need for segmentation. A low-level feature represents an image region that is different from its instant vicinity [11]. In statistical-based methods, various characteristics of cells and nuclei are captured like shape, size, color, texture, and distribution of nuclei in small regions or patches [12]. These features are also called handcrafted features. Various customized feature extraction methods are developed and used to extract the handcrafted features from the images [13]. For example, to detect the vessel-like patterns in medical images, morphological features have been used [14] while to expose the spatial patterns or structures, various graph-based features have been used such as minimum spanning tree, Delaunay triangulation, query graphs, and others [15]. However, statistical-based image representation methods are not able to express the complex visual morphological structures in histopathological images [16] [17]. To overcome this, various learning-based methods have been used for image representation.

The learning-based image representation methods use different machine learning algorithms for automatic feature extraction from the images [18]. The features represent the images in a more meaningful and collective manner. These method includes auto-encoders [19], CNN (convolutional neural networks) [20], RBM (restricted Boltzmann machines) [21], and sparse

representation [22]. Cruz-Roa et al. [23] presented a deep learning model to automatically detect basal cell carcinoma. It includes an interpretation layer that is used to highlight the patterns for differentiating the normal and cancerous tissues. Furthermore, Arevalo et al. [24] used an unsupervised feature learning approach for the analysis of histopathological images. The local patches are represented using different methods like independent component analysis based on topography and reconstruction along with sparse auto-encoders, while for the global representation of the images convolutional neural network is used. Xu et al. [25] applied the stacked sparse auto-encoder to automatically detect the nuclei in breast cancer histopathological images. Further, a deep CNN is trained to learn the features for the segmentation of epithelial and stromal tissues from the breast cancer histopathological images [20]. On the contrary, Srinivas et al. [2] introduced a sparsity model for the histopathological images having multiple color channels. Each image is represented by sparse linear permutations of training samples, having channel-wise constraints. To handle different spatial locations of image objects, a locally robust variant of this method has also been presented. Further, Vu et al. [22] presented a method for automatic feature learning from the histopathological images having complex morphological structures. The learning is based on class-specific dictionaries having sparsity constraints. The learned dictionaries are further used for the classification and grading of various diseases via histopathological images. However, most of the learning-based methods for image representation are taking high computational cost [17] when applied to complex histopathological images. Moreover, The patch-selection based methods discussed in [2] and [22] are fast but usually failed in identifying inflamed or malignant samples when there is a less malignant area available in the images.

Mid-level feature extraction methods transform the local or low-level feature descriptors into a global representation which is considered as mid-level representations of the images [26]. As the global features are the descriptors defined by local features, they are generally close to the image-level information. There are mainly three steps to get the mid-level representation [26]: (i) local feature extraction, (ii) codebook construction, which is used to find the visual words from the feature set, (iii) encoding, which is generally used to represent each image into a histogram of visual words. Further, a classification algorithm is trained on these histograms. There are various mid-level image representation methods, available in the literature, and the most widely used methods are NeTra toolbox [27], RETIN system [28], and the bag-of-features (BOF) [29]. The first contribution made on the mid-level representation based

paradigm is the NeTra toolbox, which uses the unsupervised learning for the construction of the codebook defined over the color point descriptors and the Linde–Buzo–Gray (LBG) algorithm [30] for vector quantization. Further, the RETIN system [28] uses the self-organized maps [31] for generating codebook from Gabor feature vectors. The mid-level representation technique becomes more popular when it is applied in Video Google formalism. Sivic and Zisserman [32] used the bag-of-features method for the scene and object retrieval in which the features are extracted using scale-invariant feature transform (SIFT) and the codebook construction is done by K-means. Further, Csurka et al. [33] used this method for the categorization of images. Moreover, Caicedo et al. [29] used the bag-of-features method along with the kernel function for the categorization of histopathological images. Kumar et al. [34] validated that the bag-of-features representations outperform other representation methods for histopathological image classification. Therefore, in this work, bag-of-features methods are studied for histopathological image classification. The following section presents the state-of-art methods used in various phases of the BOF method for histopathological image classification.

2.2 Bag-of-Features Method

In literature, the BOF method has emanated as a convenient mechanism for histopathological image classification. It generally consists of four prime phases: (i) feature extraction, used to detect and represent the keypoints as feature descriptors, (ii) codebook construction, used to cluster the feature descriptors for generating the visual words, (iii) feature encoding in which each image is represented in a form of visual words, and (iv) classification phase which trains a classifier using the generated form of images along with the corresponding labels. The detailed survey of each phase of the bag-of-features method is presented in the following sections.

2.2.1 Feature Extraction

In the BOF method, each input image is represented as a pool of local features by a feature extraction method. The feature extraction phase detects features from an image and represents them as a feature descriptor.

The process of identification of local features, such as corners, blobs, edges, and interest points in an image, is called feature detection. A good feature detection method should generate relevant and distinct features. Once the features are detected from an image, it is converted

Table 2.1: Categorization of features extraction methods used in bag-of-features [44]

Category	Classification	Comments	Methods
Blob Detection (Interest Point)	Partial Differential Equations (PDE) based	Use the PDE on Gaussian scale spaces	DoG [45], Hessian-Laplacian [36], SIFT [37], SURF [46]
	Template based	Decision tree and binary comparison methods are used; Computationally fast	BRISK [39], FREAK [40], ORB [47]
Learned Features	Deep Features	High-level abstractions obtained from raw images	DNN [41], SSAE [42], CNN [43]

to a feature descriptor by taking the measurements from the region centered at the interest point. In the BOF method, feature extraction methods can be categorized into two categories, namely interest point based features and learned features. Interest point detection based methods are based on the analysis of corner points in different scale spaces. These methods can further be categorized into two types, namely partial differential equation (PDE) based methods and template-based methods. The PDE based methods use partial differentiation on Gaussian scale spaces, for example difference of Gaussian (DoG) [35], Hessian-Laplacian [36], SIFT [37], speeded-up robust features (SURF) [38], etc. Moreover, template-based methods use decision tree and binary comparison for extracting the features like binary robust invariant scalable keypoints (BRISK) [39], fast retina keypoints (FREAK) [40], and oriented features from accelerated segment test and rotated binary robust independent elementary features (ORB)[38]. The learned features are the deep features extracted from the images which are generally high-level abstraction obtained from raw images. Some of these methods available in literature are deep neural networks (DNN) [41], stacked sparse auto encoder (SSAE) [42], and CNN [43]. A brief overview of various feature extraction methods used for histopathological images are presented in Table 2.1 [44].

In this work, the interest point based feature detection methods are studied. An interest point in an image is a robustly detected and well-defined location, also known as keypoint which can be a corner point, an isolated point, and a point on a curve. In literature [48], most of the corner detection methods are used to find the interest points in an image. The Harris corner detector [49] detects the corner points more precisely by the use of first derivatives and considering the neighborhood information. These points are known as interest points as these are illumination and rotation invariant and robust to the noise. However, the interest point is not invariant to

large scales [50]. To overcome this, the Harris-laplace detector [51] has been introduced which considers different scales by the use of laplacian transform. However, the number of points returned by the Harris-laplace detector is very less. Therefore, another version of the Harris-laplace detector has been introduced with less strict criteria [52]. Furthermore, for the inclusion of affine-invariant patches the Harris-Affine detector can be used [53]. The detailed comparative analysis of these methods along with other corner detectors is provided by Mikolajczyk et al. [52].

Furthermore, SIFT [37], SURF [46], and ORB [47] features are the most widely used method for feature detection and description. SIFT is a method to detect and describe the low-level features from the digital images. First, it finds some interest points and then represents them in a quantitative manner known as descriptors. These descriptors are robust to scale, rotation, and illumination conditions. Irshad et al. [54] introduced an automated method for the detection of mitosis to assist the pathologists. The color spaces are distinguished by analyzing the combined pattern of texture features, namely SIFT along with the run-length and co-occurrence features. These features are then fed to the SVM (support vector machine) for the classification purpose. Further, Raza et al. [55] used scale invariant features in the BOF framework for the identification of subtypes in RCC (renal cell carcinoma). Caicedo et al. [29] also used the BOF method with SIFT for feature detection and description to classify histopathological images automatically. Raza et al. [35] studied and analyzed the rotation and scale in-variance behavior in the BOF method for the classification of histopathological images of RCC.

SURF [46] is a fast and efficient interest point detector method which uses the integral image for the convolutions. The method works in three phases, namely keypoint detection, description, and matching. Yang et al. [56] solved the problem of image stitching in biomedical research about the whole sections or larges areas. The proposed method considers the SURF method for feature extraction in a fast and efficient manner. Sanghavi and Agaian [57] proposed an automated method for the classification of histopathological images of prostate cancer based on the bag-of-features method. For feature extraction, a comparative analysis is presented between SURF and SIFT features based bag-of-features method. SURF returns higher sensitivity as compared to SIFT-based bag-of-features. Wand and Chen [58] addressed the problem of image alignment in X-ray images and tissue images and compared it with five other methods, namely TrackEM2, UnwarpJ, BUnwarpJ, mutual information, and SURF. The proposed method shows

Table 2.2: Various codebook construction methods used in the bag-of-features [63]

Category	Methods	Comments
Hierarchical methods	Agglomerative clustering, Mean shift	These methods can not be applied to large datasets or histology images due to high computational cost.
Partitional methods	K-means, FCM, GMM	Generates non-uniform coding (biased towards dense regions); Non robust; Optimal codebook size (K) is unambiguous.
Meta-heuristic based methods	PSO, GSA, WOA	Used to find optimal visual words based on some objective function defined over compactness and separation; Computationally expensive.

better results on both of the datasets.

ORB [47] uses oriented FAST [59] method as a high speed corner detector and the detected corners are represented by a variant of BRIEF descriptor, i.e., binary robust independent elementary features descriptor [60]. Davidson et al. [61] introduced a new method for stitching the images with the use of ORB descriptors, along with good enough transformation and local sensitivity hashing for fast processing. The proposed method is applied to study adaptive optics ophthalmoscopy automatically and efficiently. Further, Adel et al. [62] used ORB for extracting the features from microscopic images of Oral Epithelial Dysplasia and fed these features to SVM for the automated grading of disease.

Moreover, SIFT and SURF features are robust to scale, rotation, and illumination changes while ORB is a computationally fast method. However, due to complex morphological structures of histopathological images, these methods generate high dimensional feature vectors.

2.2.2 Codebook Construction

The second phase of the BOF method is the codebook construction in which the extracted features are divided into different clusters to find the visual words. The pool of visual words is known as a codebook. In literature, various clustering methods are used for codebook construction in BOF, mentioned in Table 2.2, which can be grouped into three classes, namely hierarchical methods, partitional methods, and meta-heuristic based methods.

- **Hierarchical Clustering:** In hierarchical clustering, the grouping of data at different similarity levels is performed which is schematized by a tree-like structure termed as a

dendrogram. Generally, the hierarchical splitting follows two approaches, namely divisive and agglomerative [64]. In divisive clustering, recursive hierarchical data splitting is performed in a top-down fashion to generate the clusters. All the data items belong to a single cluster initially and further divided into various small clusters until a terminating condition is satisfied or until each data item forms its cluster. Divisive clustering (DIVCLUS-T) [65] and divisive hierarchical clustering with diameter criterion (DHCDC) [66] are popular methods of this category. On the other side, the agglomerative clustering is performed in the bottom-up fashion where data points are merged hierarchically to produce clusters. Initially, each data item defines itself as a cluster which is further merged into bigger clusters, until a termination criterion is met or until a single cluster is formed consisting of all the data items. Some popular agglomerative methods are balanced iterative reducing and clustering using hierarchies (BIRCH) [67], chameleon [68], and clustering using representatives (CURE) [69].

Meijnen et al. [70] used the hierarchical clustering method for dividing the in-situ ductal carcinoma breast cancer images into two major groups and each major group is further divided into five sub-classes based on six markers. The proposed method is used to investigate gene expression profiling. Bhargava et al. [71] developed a system for automated spectroscopic analysis of tissue images. For the segmentation of tissue images, hierarchical clustering methods are used and the combination of tissue microarray analysis with Fourier transform of spectroscopic images are used for high throughput analysis. Furthermore, Pourahmad et al. [72] developed a framework for the automated grading of colorectal cancer with the use of hierarchical clustering and partitional clustering approaches. As per the experimental analysis, hierarchical clustering outperforms other considered clustering methods.

Generally, hierarchical methods employ a greedy approach and do not reconsider a data point after it has been assigned to a cluster. This results in lacking the capability of correcting the misclassified data point. Therefore, these methods lack robustness, especially in the case of noise and outliers. They do not intend to optimize an objective function while forming the clusters. Moreover, they perform poorly when clusters are overlapping. While generating clusters for a particular problem, knowledge about the number of clusters is required. Moreover, the formation of spherical clusters and the reversal of

the hierarchical structure are distorted [73]. Additionally, one of the major issues while clustering data through a hierarchical approach, especially on high-dimensional dataset like image, is the time complexity. The time complexity of hierarchical methods is very expensive which is approximately defined as $O(n^3)$ [64], where n corresponds to the number of data points. Hierarchical clustering methods, such as agglomerative clustering [74] and mean shift [75], cannot be applied to large datasets or histology images due to high computational cost. Therefore, to overcome this limitation, partitional clustering methods are preferred which are discussed in the following section.

- **Partitional Clustering:** It is relatively quite popular and preferred over the hierarchical clustering, especially for large datasets, due to its computational efficiency [76]. In this type of clustering method, the notion of similarity distance is usually used as the measurement parameters. Thus, partitional clustering is the process of dividing the data into clusters according to the considered objective function such that similar data points belong to one cluster and dissimilar data points in other clusters. To achieve this, the distance of each data point is measured from every cluster center and is allocated to that cluster which is nearest. Moreover, in partitional clustering, the general notion of the objective function is the minimization of the within-cluster similarity criteria which is usually computed by Euclidean distance. The objective function expresses the goodness of each formed cluster and returns the best representation from the generated clusters. Moreover, these methods surely assign data point with a cluster even if it is quite far from the respective cluster centroid which, sometimes, results in distortion of the cluster shapes or false results, especially in the case of noise or outlier [73]. There are two classes of partitional clustering methods, namely soft and hard clustering methods.

Soft clustering methods assign each data to either two or more clusters with a degree of belongingness (or membership) iteratively. The degree of belongingness illustrates the level of association among data more reasonably. Some popular methods are fuzzy c-means (FCM) [77], fuzzy c-shells (FCS) [78], and mountain method [79]. In FCM, each cluster is considered as a multidimensional hypersphere and defines the distance function accordingly. Mountain method uses the mountain function to find the cluster centers. Particularly, FCM is the most widely used and popular method of this approach.

Hard clustering methods partition the data into disjoint clusters according to the objec-

tive function. Generally, the objective function is the sum of squared Euclidean distance between data and associated centroid which is to be minimized. Usually, the center of the clustered data is considered as the centroid of the clusters in these methods. Moreover, in contrast to soft clustering, hard clustering assigns data to a single cluster only i.e., each data will have the degree of belongingness as either 0 or 1. One of the popular hard clustering methods is K-means [80]. In K-means based methods, the cluster center is cosidered as the mean of all the data points assigned to the corresponding cluster. This is iteratively continued until some defined convergence criterion is met. K-means method has the time complexity of $O(nkt)$ [64] where, t corresponds to the maximum iterations. This method is biased towards initial cluster centroids and usually gets trap into local minima. Moreover, solutions vary with the number of clusters.

Rueda et al. [81] used K-means clustering for codebook construction in the BOF method for the classification of magnetic resonance images of Alzheimer's disease. Cruz-Roa et al. [82] used the BOF method for the representation of visual content in histopathological images. K-means is used to find different visual words in the BOF method. Wiliem et al. [83] used K-means clustering for generating the codebook from SIFT-based descriptors and used them for the classification of immunofluorescence images of human epithelial cells. Avni et al. [84] generated the patch based visual words for the categorization of the X-ray images for automated retrieval. Further, Stanciu et al. [85] used the BOF method along with K-means for codebook generation for the diagnosis of liver fibrosis by excitation microscopy of two photons.

Therefore, various partitional clustering methods, used in the bag-of-features, are K-means, FCM, and GMM [86]. Some of them are discussed above. However, these methods are biased towards dense regions and may generate non-uniform coding.

- **Meta-heuristics based Clustering:** This category involves the use of meta-heuristic approaches to obtain optimal clusters, especially in the case of images. These approaches use random solutions that are updated according to defined optimality criteria known as objective function [87]. A meta-heuristic algorithm computes the objective function value by using the generated solutions. A meta-heuristic algorithm must be efficient so that the optimal solution can be reached. Broadly, the algorithms of this category belong to the class of optimization algorithms which correspond to a set of algorithms that can solve

computationally hard problems like NP-complete problems. An NP-complete problem is a problem for which there exists no deterministic algorithm that can provide the exact solution in polynomial time. There exist several optimization algorithms in the literature, however, no single algorithm is efficient in solving every problem as stated by No Free Lunch Theorem [88].

Researchers have proposed many such algorithms and their variants based on the inspiration from nature to provide an efficient and optimal solution. Over the last three decades, more than sixty meta-heuristic algorithms have been proposed. As these algorithms are inspired by nature, each algorithm mimics particular natural phenomena that may belong to evolutionary, physical, or biological. Researchers are continuously working to improve the existing algorithms and also proposing new algorithms that are giving competitive results as compared to the existing algorithms present in the literature. Two common aspects that are often found in these algorithms are exploration and exploitation [89]. Exploration represents the diversification in the search space wherein the existing solutions are updated to explore the search space. This helps in exploring the new solutions, prevents the stagnation problem, and responsible for achieving the global solution. The exploitation, which corresponds to the intensification of the current solution, performs the local search around the currently generated solutions. In this, the goal is to exploit the search space and responsible for convergence to the optimal solution. Generally, meta-heuristic algorithms may broadly be classified into two categories, namely evolutionary and swarm algorithms.

Evolutionary-based algorithms are based on evolution theories such as Darwin evolutionary theory. The evolutionary algorithms work on the principle of generating better individuals with the course of generation by combining the best individuals of the current generation. The genetic algorithm (GA) [90] is an evolutionary algorithm based on the evolution of natural species. It implements the exploration and exploitation through the mutation and crossover operators respectively. Another biological process based evolutionary algorithm is evolutionary strategy [91] which performs recombination and mutation with equal probability and uses multiple parents to accord an offspring. Differential evolution (DE) is another popular evolutionary algorithmic introduced by Storn et al. [92]. Simon [93] presented biogeography-based optimization (BBO) which is based on

the immigration and emigration of the species between the islands of natural biogeography. Baluja [94] proposed the probability-based incremental learning algorithm which manages only statics of the population rather than managing the complete population.

Swarm-based algorithms behave like the swarm of agents such as fishes or birds to achieve optimal results. Kennedy [95] proposed the particle swarm optimization (PSO) which mimics the swarm behavior of birds or fishes for searching the food. Ant colony optimization (ACO) imitates the ant's behavior for finding paths [96]. Gravitational search algorithm (GSA), introduced by Rashedi et al. [97], is an algorithm based on Newtonian laws of gravity and motion. Hosseini [98] proposed an intelligent water drop algorithm based on the flow of rivers, as rivers often follow the shortest path while flowing from source to destination. Further, Wang et al. [99] proposed the hybrid krill heard algorithm to overcome the problem of poor exploitation capability of the krill herd algorithm. Bansal et al. [100] introduced spider monkey optimization (SMO) that mimics the behavior of spider monkeys. Furthermore, Mirjalili [101] proposed an ant-lion based optimizer that imitates the behavior of ant-lions. Mirjalili [102] also introduced the moth-flame optimization, which simulates the death behavior of moths, in which the movement of the agent is based on the transverse orientation based navigation of moths. Some recent swarm algorithms are multi-verse optimizer [103], galaxy-based search algorithm [104], small-world optimization algorithm [105], and ray optimization [106].

The various meta-heuristic algorithms can provide promising results for unfolding the clustering problem. Generally, these clustering-based methods are better than other clustering methods in terms of independence from the initial parameter settings and return global optimal solution [107]. As these methods ensure a better solution, their use in clustering has been prominent in the literature. The basic approach to emerge meta-heuristics based clustering algorithm was introduced by Selim and Alsultan [108], using simulated annealing. After that, Bezdek et al. [109] presented a genetic algorithm based data clustering method that was the first evolutionary based method of data clustering. The first, swarm-based clustering method was developed by Lumer et al. [110] using ant colony optimization.

Meta-heuristic based clustering methods like BBO [111], GSA [112], and whale optimization algorithm (WOA) [113] are used to find optimal visual words based on some

Algorithm 2.1 Meta-heuristic based clustering method

Input: Initialize the number of clusters to be formed.

Output: Clustered data.

Choose a population of random individuals where, each individual corresponds to a cluster center.

while clustering condition is not satisfied or centers do not change **do**

 Calculate the fitness value of each solution through the considered objective function;

 Update the solution according to the meta-heuristic algorithm;

 Assign each data point to the nearest cluster center;

end while

Table 2.3: Various types of feature encoding methods [115, 114]

Type	Methods	Comments
Voting based	HV [32], SA [116], LSA [117], Salient Coding [118],	Based on the formation of histograms which represent the distribution of visual words
Reconstruction based	OMP [119], Sparse coding [120], LLC [121], LCC [122]	Each feature should be reconstructed by the visual words by applying some constraints and solving least-square optimization problem
Super vector based	LTC [123], SVC [124], Fisher Vector [125], VLAD [126]	Gaussian mixture model is used for estimating the feature distributions which contain means, covariance, and weights of Gaussian distributions

objective functions defined over compactness and separation. Mittal and Saraswat [112] proposed a modified version of the BOF method by generating the optimal visual words using GSA. The proposed methods is validated against the K-means clustering based BOF method for the categorization of tissue images. However, these methods have high computational cost when applied to complex histopathological images. The pseudo-code for a meta-heuristic based clustering method is presented in Algorithm 2.1.

2.2.3 Feature Encoding

The third phase of the BOF method is feature encoding in which each image is encoded in terms of visual words by a feature encoding method. Each image is encoded as a coding vector of size (k), where k represents the count of visual words. Based on the properties of the encoding method, they are divided into three main categories, namely voting, reconstruction, and super-vector based encoding methods [114]. The brief descriptions of these methods are also provided in Table 2.3.

- **Voting-based methods:** In the voting-based method, each feature descriptor cast their

vote for any of the codeword based on some rules or strategy. In response to these voting, a coding vector of k dimensions is created, which is known as a codeword. The collection of codewords is used to make the whole codebook. The most commonly used methods in this category are hard voting (Vector Quantization (VQ)) [32], kernel codebook encoding (KCB), soft assignment (SA) [116], localized soft assignment (LSA) [117], and salient coding [118].

For each feature vector f , the voting value for the visual word v can be calculated as a function of f , namely $c(i) = \phi(f)$. The formulation of $\phi(f)$ is different for each type of encoding method. In hard voting methods, each feature can cast the vote to its nearest codeword. Therefore, the formulation of $\phi(f)$ is defined as follows:

$$\phi(f) = \begin{cases} 1 & \arg \min_j \|f - v_j\|_2, \\ 0 & \text{otherwise,} \end{cases} \quad (2.1)$$

The VQ methods are computationally efficient but sometimes it may loss some information. As the resolution to this problem, soft coding methods are introduced which are also known as soft assignment based methods. In soft assignment technique, each feature must cast their vote to every visual word with some normalized weight (i.e., $\phi(f) = w_i$). This normalized weight (w_i) of i^{th} feature vector f for j^{th} visual word (v_j) is defined in Eq. (2.2).

$$w_i = \frac{\exp(-\beta \|f - v_j\|_2^2)}{\sum_{j=1}^k \exp(-\beta \|f - v_j\|_2^2)} \quad (2.2)$$

Furthermore, Liu et al. [117] proposed a variant of soft assignment methods, namely localized soft assignment in which a feature can only votes for its k -nearby codewords. The normalized weight formulation is modified by inducing an indication function ($I(f, v_j)$) for searching the k -nearest neighbor (N_k) members of f . The indication function and modified normalized weight is given in Eq. (2.3) and Eq. (2.4) respectively.

$$w_i = \frac{I(f, v_j) \exp(-\beta \|f - v_j\|_2^2)}{\sum_{j=1}^k \exp(-\beta \|f - v_j\|_2^2)} \quad (2.3)$$

$$I(f, v_j) = \begin{cases} 1 & v_j \in N_k(f), \\ 0 & \text{otherwise,} \end{cases} \quad (2.4)$$

Further, salient coding is another type of hard vector quantization method. Each feature calculates its weighted vote based on the difference between the nearest visual word and the other $k - 1$ nearest visual words [118]. The visual description of each voting based methods is depicted in Figure 2.3.

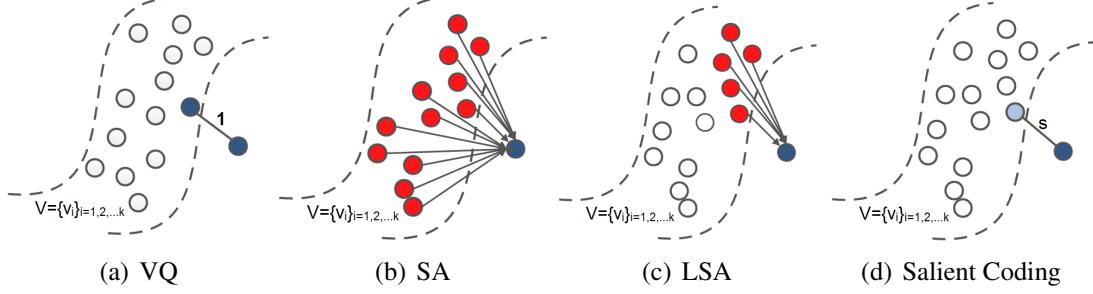


Figure 2.3: Visual description of voting based methods [114]

- **Reconstruction-based methods:** In reconstruction-based encoding methods, code c is formulated in the reconstruction or decoding perspective. The i^{th} input descriptor can be decoded from the code c with the objective of minimum reconstruction error. The objective is equated in Eq. (2.5) with a regularization component (Ψ).

$$\arg \min_c \| f - V_c \|_2^2 + \lambda \Psi(c), \quad (2.5)$$

where $\Psi(c)$ corresponds to the properties of c while λ is the balancing parameter between the reconstruction error and $\Psi(c)$.

Some prominent reconstruction methods are sparse coding (SPC) [120], orthogonal matching pursuit (OMP) [119], local coordinate coding (LCC) [121], and locality-constrained linear coding (LLC) [122].

OMP method constrains $\Psi(c)$ with l_0 - norm [127] which is defined in Eq. (2.6) and follows sparse encoding. However, the l_0 -norm is non-convex, therefore, this problem requires some heuristic approach for the solution.

$$\Psi(c) = \sqrt[0]{\sum_i f_i^0} \quad (2.6)$$

This non-convexity is relaxed by SPC method with the use of l_1 -norm [128] which results in obtaining global optimal solution. The modified constraint function of using l_1 -norm

is given in Eq. (2.7).

$$\Psi(c) = \sum_i f_i \quad (2.7)$$

In the SPC and OMP methods, locality is not defined theoretically, but it is measured based on the empirical analysis [121]. Therefore, Yu et al. [121] proposed a variant of SPC, namely LCC wherein locality of the encoding is encouraged instead of sparsity. Moreover, Wang et al. [122] introduced a faster version called LLC, for the large scale problems.

- **Super-vector based methods:** In super-vector based encoding methods, high order statistics are aggregated to obtain high dimensional encoding representation. The popular methods of this category are super vector coding (SVC) [124], local tangent-based coding (LTC) [123], vector of locally aggregated descriptors (VLAD) [126], and fisher vector (FV) [125]. In LTC [123], encoding is formulated in terms of manifold approximation and intrinsic dimensionality where, the non-linear feature manifold is approximated as a local linear function according to the Lipschitz smooth condition and this linear function consists of lower intrinsic dimensionality which is obtained through principal component analysis (PCA). Further, Zhou et al. [124] proposed a simplified version of LTC, known as SVC, in which vector quantization is employed instead of PCA. Moreover, another super-vector based encoding scheme is fisher vector which is based on fisher kernel [129] and leverages both generative and discriminative approaches. This method is used for the categorization of large scale images [125]. Further, Jegou et al. [126] proposed a complex and restricted version of FV which keeps only first-order statistics for encoding.

Wiliem et al. [83] proposed a new automated method for the recognition of human epithelial cells which represents each image into the codebook descriptor vectors having dual regions and feeds these descriptors to the nearest convex hull method for the classification. The codebook descriptors are generated by various encoding methods, namely vector quantization, sparse coding, and soft assignment. Zhou et al. [130] developed a learning model based on the multispectral sparse coding based on the convolutions for the classification of histology tissue sections. Further, Nayak et al. [21] also used sparse codes for finding the unique patches such as necrosis and viable tumor from the histology tissue sections.

In histopathological image analysis, voting-based encoding methods are widely used for the encoding of images. Nowakova et al. [131] presented a new method of vector quantization by the inclusion of fuzzy s-trees and fuzzy signatures to resolve the problem of medical image retrieval especially mammography images. Diamant et al. [132] used vector quantization based encoding method in the BOF and relevant visual word per task is selected based on mutual information for the automated classification of medical images. Han et al. [133] proposed a hierarchical vector quantization method to detect pulmonary nodules in CT images. The high-level and low-level VQ methods are responsible to detect lung segments and inter nodule candidate segments respectively. Another variant of VQ, namely learning vector quantization (LVQ) is presented by Dieterle et al. [134] for the evaluation of urinary nucleoside. The LVQ method has been compared with SVM and neural networks. Mattfeldt et al. [135] applied the LVQ method to classify prostate incidental carcinoma. Further, another version of vector quantization is introduced by Hipp et al. [136] which attains better performance as compared to standard VQ for the pattern recognition in pathological images.

2.2.4 Classification

A classifier is used to categorize or identify the image labels or components based on the features extracted from these images. The richness and quality of feature descriptors affect the performance of the classifiers. Histopathological image classification is a complex process for the machine or computer system which motivates the research community to study and analyze this field in perspective of machine intelligence.

Table 2.4: Popularly used classification methods for histopathological image classification

Classifier	Details	Applications
Random Forest	The prediction is done based on the majority voting of different decision trees	[38] [137] [138]
Logistic Regression	The score value is calculated based on the linear function for the prediction of the target class	[25] [139] [140]
Support Vector Machine	Data is classified by the defined hyperplane and new image is categorized based on hyperplane	[141] [142] [143]
Linear Discriminant Analysis	The separation between two or more classes is found based on the linear combination of features	[144] [145] [146]
Bayesian Classifier	Based on the conditional probability model where the classification is probabilistically related to the observed samples	[147] [148] [149] [150]

In the last phase of the BOF method, a classifier is trained using encoded images and their corresponding labels. After the training of a classifier, it is used to test the unknown encoded images. Table 2.4 enlists some of the popular classifiers used in the literature [3] along with their application to the histopathological image classification.

2.3 Research Gaps and Objectives

After reviewing the literature, following sections discuss the identified research gaps and research objectives.

2.3.1 Research Gaps

1. Histopathological images contain various morphological variabilities and structures, due to which high dimensional keypoint descriptors are generated. This makes codebook construction computationally expensive and prompts to generate irrelevant visual words. No method has been reported to select the prominent keypoints in the BOF method for histopathological images;
2. The generated visual words are biased towards the densest regions in descriptor space and due to which other informative regions could not be encoded;
3. Existing meta-heuristic based visual word generation methods are computationally expensive, and
4. Existing feature encoding methods consider only a single image feature to encode an image.

2.3.2 Objectives

After examining the research gaps, the following research objectives are defined:

1. A new keypoint selection method has to be designed and developed to find discriminative and relevant features for codebook construction;
2. An efficient meta-heuristic based codebook generation method has to be designed and developed to reduce the effect of dense regions of histopathological images;

3. A computationally efficient and effective codebook generation method has to be designed and developed for finding the relevant visual words, and
4. An efficient feature encoding method has to be designed and developed by incorporating the merits of two different features descriptors for the better image representation.

2.4 Summary

In this chapter, different types of image representation methods for histopathological image classification have been reviewed. The BOF method is one of the popular image representation techniques which performs better on histopathological images. There are four main phases of the BOF method, namely feature extraction, codebook construction, feature encoding, and classification. The popular feature extraction methods, like SIFT and SURF, are invariant to rotation, light illumination, and scaling but, these methods generate high dimensional feature vectors when applied on histopathological images. Therefore, there is a requirement of an efficient interest point selection method to reduce the high dimensional complexity of the extracted features. After feature extraction, the next phase in codebook construction, in which various clustering methods are used to find the visual words. In literature, mainly three types of clustering methods are available, namely hierarchical, partitional, meta-heuristic based clustering methods. In BOF, partitional clustering methods are widely used for the classification of histopathological images. However, these methods are generally biased towards the densest regions in the images. To overcome this limitation, various meta-heuristic based clustering methods are used. These methods sometimes stuck into local optima. Therefore, there is a requirement for a new meta-heuristic based clustering method. However, the codebook constructed from these methods is computationally expensive. Hence, it is required to make this step computationally efficient. After codebook construction, feature encoding methods are applied to represent each image in terms of visual words. For histopathological image classification, voting based encoding methods are widely used. In the next phase of BOF, the encoded images are fed to the classifier along with the image labels for training.

This chapter has also presented the research gaps obtained from literature survey and defined the corresponding research objectives of the thesis. The next chapter introduces a new efficient keypoints selection method in BOF for histopathological image classification.

CHAPTER 3

GREY RELATIONAL BASED KEYPOINTS SELECTION

3.1 Introduction

Histopathological image classification has grown spacially and essential for disease analysis. To automatically discriminate benign and malignant images, various classification and feature detection methods are presented in the literature [8]. Three types of features can be extracted from the images, namely hand-crafted features, mid-level features, and learned or high-level features. The hand-crafted features, also coined as local features, can be extracted using various popular feature extraction methods, like LBP (local binary pattern), SIFT, histogram of oriented gradient (HOG) [151], and SURF. These methods extract the local visual features from the images.

In the BOF method, features are extracted from images which are clustered using K-means to generate visual words. The occurrences of visual words in an image can be represented by the histogram which is used to train the classifier. However, this standard method of classification has following demerits [18]:

1. All the extracted keypoints are not relevant for the classification and may lead to computationally inefficient image representation.
2. For clustering, K-means is used which has its disadvantages like variance dependency, non-robust, and non-scalable.

To overcome the above mentioned demerits, various methods have been proposed for the selection of relevant keypoints [152] [153]. Dorko and Schmid [153] enhanced the classification performance by applying the SVM on various groups of descriptors to select the most relevant groups. The descriptor vector is divided into groups by the Gaussian mixture model (GMM). Moreover, there are the three best known instance selection algorithms known as IB3 (instance

based selection) [154], DROP3 (reduction techniques for instance-based learning) [155], and ICF (iterative case filtering) [152] with the computational cost of $O(n^2 \log_2 n)$, $O(n^3)$, and $O(n^2)$ respectively. Here n denotes the instances. Hence, IB3 is an efficient instance selection algorithm which is further used as a keypoints selection method [156]. However, these methods do not perform well on high dimensional image datasets such as histopathological image datasets. Recently, Lin et al. [156] proposed two iteration keypoints selection methods to improve the bag-of-features but these methods also work on Euclidean distance as similarity measures. On the other hand, histopathological images are very complex having a huge number of keypoints for a single image and there is no method available in the literature for keypoints selection of histopathological images. However, the methods proposed by Lin et. al [156] shows good performance on Caltech image datasets which are less complex as compared to histopathological images and the same method fails to improve the performance to classify the histopathological images. Therefore, an efficient keypoints selection technique based on the grey relational analysis (GRA) is introduced in this chapter which selects comparatively more relevant keypoints. Further, the modified GRA based BOF method is used to classify the histopathological images. The next section presents the BOF approach for histopathological image classification (HIC) followed by the new grey relational analysis based BOF method.

3.2 Bag-of-Features Method

The BOF method is one of the convenient mechanisms for histopathological image classification. It generally consists of five phases as shown in Figure 3.1: (i) Setup the image data-store in which images are organized into predefined categories and partitioned into training and validation subsets, (ii) Extract the texture features or keypoints using feature extraction method, (iii) Cluster the keypoints to generate the visual words, (iv) Encode each image as the histogram of visual words, and (v) Train the classifier using these histograms and corresponding image labels. Finally, the images from the validation set are fed to the trained classifier without a label to predict their labels. Mathematically, the BOF method can be described as follows:

- 1. Dataset partition:** Let $C = \{c_1, c_2, \dots, c_i, \dots, c_n\}$ is a set of n classes. Each class c_i is associated with a set of images. The image dataset is divided into two parts. One is a training set on which the classifier is trained and the other is a validation set, which is used to validate the trained classifier. The training set of N images is prepared by

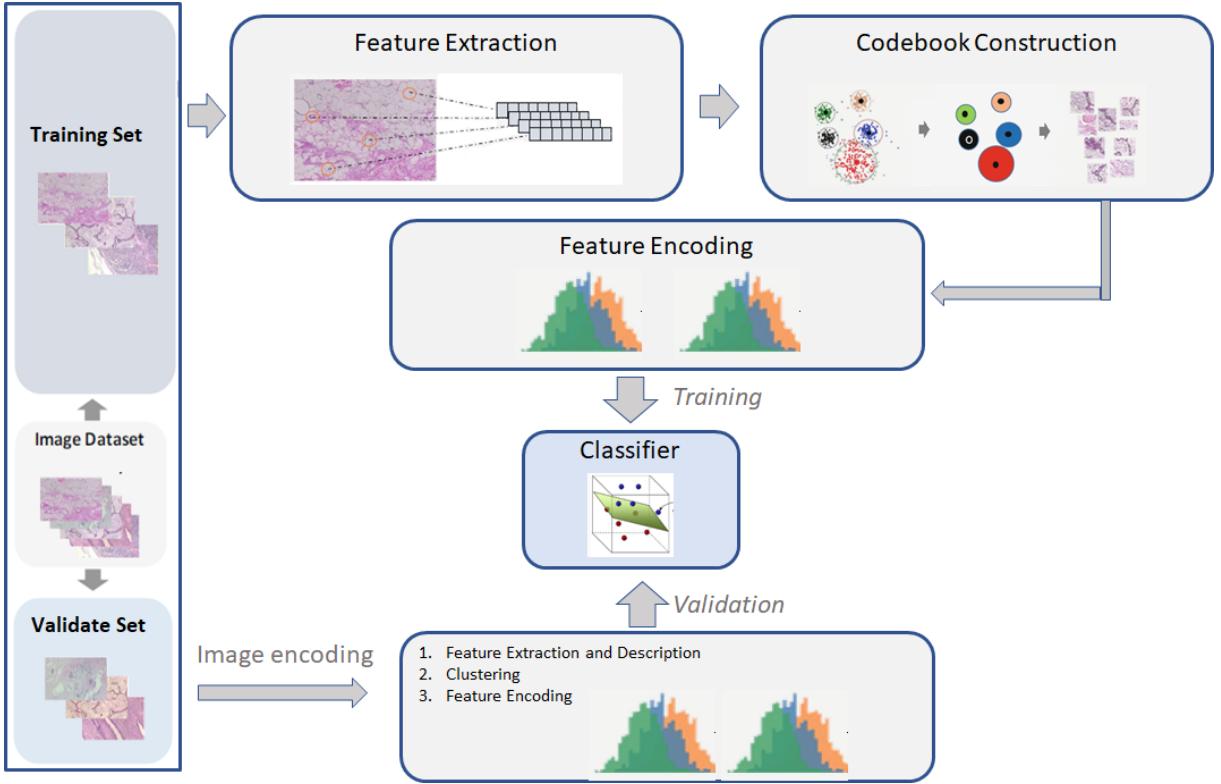


Figure 3.1: The architecture of the BOF method

randomly selecting M_i images from each class c_i which is also given by Eq. (3.1). The remaining images of the classes are considered as a part of validation set.

$$N = \sum_{i=1}^n M_i \quad (3.1)$$

2. Feature extraction: Extract the keypoints from all N images of training set using a feature extraction method like, SURF [46], FREAK [40], SIFT [37], and HOG [157]. Let X is a set of keypoints, defined as Eq. (5.1).

$$X = [F_1, F_2, \dots, F_N]^T \quad (3.2)$$

where, F_i is a matrix of P keypoints for the i^{th} image, defined over d -dimensional space

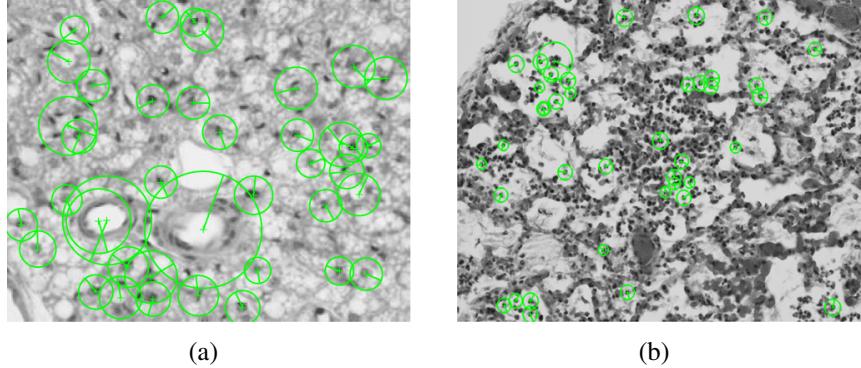


Figure 3.2: The keypoints detected by SURF in (a) connective tissue image and (b) inflamed lung tissue images

and is given by Eq (3.3).

$$F_i = \begin{Bmatrix} f_{11} & f_{12} & f_{13} & \dots & f_{1d} \\ f_{21} & f_{22} & f_{23} & \dots & f_{2d} \\ \dots & \dots & \dots & \dots & \dots \\ f_{P1} & f_{P2} & f_{P3} & \dots & f_{Pd} \end{Bmatrix} \quad (3.3)$$

Figure 3.2 shows representative keypoints detected by one of the feature extraction method i.e., SURF from two images, randomly taken from the two considered histopathological image datasets. Each image is first converted to grayscale then SURF detector is used to find the predefined number of keypoints from these images. In the figure, only 40 strong keypoints are depicted for simplicity and visualization.

3. **Codebook construction:** Create visual words by iteratively grouping the extracted descriptor vector X into n -mutually exclusive clusters. Each cluster can have any number of keypoints based on the similitude of the intensity values of pixels in an image with the extracted keypoints. For the same, K -means algorithm is used and the returned cluster centers are represented as visual words.
4. **Encoding:** Encode each image into a histogram ($H_j = \{H_{j1}, H_{j2}, \dots, H_{jn}\}$ for $j = 1, 2, \dots, N$), representing the visual word occurrences in each image which is given by Eq. (3.4).

$$H_{jk} = \sum_{i=1}^P \mu_{ik}(j) \text{ where } k = 1, 2, \dots, n \quad (3.4)$$

$$\mu_{ik} = \begin{cases} 1 & \text{if } \|v_k - f_i\| \leq \|v_s - f_i\| \text{ for } s = 1, 2, \dots, n, \\ 0 & \text{otherwise} \end{cases} \quad (3.5)$$

where, P represents the number of keypoints and $\mu_{ik}(j)$ is 1 when any visual word (v_k) is close to any keypoint f_i in the image. This method is also known as vector quantization.

5. **Classification:** Each histogram (H_j) along with its annotation is used to train the classifier for the image classification task. Once the classifier is trained, it is tested to predict the label of images provided in the validation set. Each validation image is represented as the histogram as discussed above and fed to the classifier without a label. Based on the returned label by the classifier, its accuracy is measured.

In the feature extraction phase of the BOF method, a feature detection and representation method is used to find the keypoints in the images. These keypoints are then represented as the descriptor vectors which are further used for codebook construction. Out of many feature extraction methods, SURF is the fastest method because it uses box filters for the convolution of images and converts each image as the integral image. It extracts the texture features from the images [46]. Moreover, SURF is a resolution invariant feature detector, hence images of different resolutions do not have any impact on the classification performance. This property of SURF helps to analyze the histopathological images, having different resolutions (e.g., 10x, 20x, 40x) [158]. The interest points in the images are detected using the Hessian matrix approximation. SURF also shows good performance over other alternatives like SIFT [159]. Therefore, this work uses SURF to extract a set of keypoints (X) from N training images.

Generally, SURF extracts a large number of keypoints due to the complex texture of histopathological images. This reduces the efficiency of visual vocabulary generation [156]. Furthermore, all of the detected keypoints are not necessary for image classification and annotation [156]. Hence, an efficient keypoints selection method is required for the acquisition of relevant keypoints that can improve the speed and efficiency of the BOF method. Some of the popular keypoints selection techniques are IB3 [154] and iterative keypoints selection (IKS1, IKS2) [156]. IB3 is an efficient instance selection method with high space complexity while IKS1 and IKS2 are the keypoints selection methods that are used to find representative keypoints from the images. IKS1 and IKS2 are differed by their initial representative keypoints selection meth-

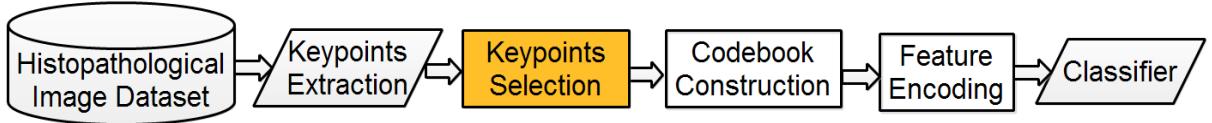


Figure 3.3: Flow chart of the enhanced BOF method

ods. In IKS1, representative keypoints are selected randomly while in IKS2, cluster centers are considered as representative keypoints. The remaining keypoints are eliminated based on their Euclidean distances from the selected representative keypoints. However, Euclidean distance similarity measure is computationally expensive for high dimensional data. Chang et al. [160] has shown that computational cost of Grey relational analysis (GRA) [161] based similarity measure is better than the Euclidean distance based similarity. Therefore, in this work, a new GRA based keypoints selection (GKS) method is introduced to reduce the number of keypoints before feeding them into the next phase of the BOF method i.e., codebook construction. The modified flow of the BOF method is depicted in Figure 3.3. The next subsection provides a detailed description of the new grey relation analysis based keypoint selection method.

3.3 Grey Relational Analysis based Keypoints Selection

The GKS method uses the concept of grey relational analysis for finding the similarity between the keypoints. GRA [161] is a part of Grey system theory which is used to examine the similarity between data tuples based on geometrical mathematics [162]. It conforms to four basic principles in the dataset, i.e., proximity, normality, symmetry, and entirety [163]. In GRA, the similarity between a reference tuple and the remaining tuples for a given data is computed by Grey relational grades (GRGs) whose value lies between 0 and 1. For any data tuple, if GRG is close to 1, then it is highly similar to the reference tuple while the dissimilarity will be signified if GRG value is close to 0 [160].

Therefore, the new keypoints selection method uses GRA to eliminate similar keypoints from the feature descriptor, generated by SURF. The new GKS method has the following steps:

1. Cluster the keypoints into K clusters using approximate K-means (AKM) algorithm [80].
2. Make the cluster centers as the member of selected keypoints set and also consider them as reference points for the computation of GRGs for the remaining keypoints.
3. Compute the GRG values between the reference point and the keypoints lying within the

corresponding cluster. The mathematical formulation of GRG computation is described below.

Let $X_o = X_{r1}, X_{r2}, \dots, X_{ri}, \dots, X_{rn}$ be a set of n reference points. The elements in X_o are of the form $X_{ri} = \langle X_{ri}(1), X_{ri}(2), \dots, X_{ri}(u) \rangle$, where u corresponds to the dimension of the extracted keypoint. Similarly, let $X_c = X_{c1}, X_{c2}, \dots, X_{cm}$ be a set of $m = P - n$ remaining keypoints considered as comparative keypoints where, each element in X_c can be denoted as $X_{cj} = \langle X_{cj}(1), X_{cj}(2), \dots, X_{cj}(u) \rangle$. The GRG value of each keypoint in X_c is given by Eq. (5.3) [160].

$$GRG(X_{oi}, X_{cj}) = \sum_{t=1}^u [\alpha_i(t) \cdot GRC(X_{oi}(t), X_{cj}(t))] \quad (3.6)$$

where, GRC is the grey relational coefficients and $\alpha_i(t) = \frac{1}{u}$ is the weighting factor of GRC. The GRC value, between i^{th} keypoint of X_o and j^{th} keypoint of X_c at u^{th} datum, belonging to the i^{th} cluster only is given by Eq. (5.2) [160].

$$GRC(X_{oi}(u), X_{cj}(u)) = \frac{\min_{ij} \Delta_{ij}(u) + \xi \max_{ij} \Delta_{ij}(u)}{\Delta_{ij}(u) + \xi \max_{ij} \Delta_{ij}(u)}, \quad (3.7)$$

where, $\xi \in (0, 1]$ is a random number to control the constancy between $\max_{ij} \Delta_{ij}(u)$ and $\min_{ij} \Delta_{ij}(u)$. $\Delta_{ij}(u)$ is computed by $|X_{oi}(u) - X_{cj}(u)|$ for $i = 1, 2, \dots, n$, $j = 1, 2, \dots, c$.

4. In every cluster, the above computation is performed to find the highly similar points with cluster center and eliminate $s\%$ of the keypoints from each cluster whose GRG values are higher, in their corresponding cluster. Here, s is termed as shrinking threshold.
5. Repeat the steps 1 to 4 till the remaining keypoints are greater than K and add the last set (having K points only) of cluster centers to the selected keypoints set.
6. Use the selected keypoints set as input to the next phase of BOF i.e., codebook construction.

After finding the optimum keypoints from the new GKS method, the codebook construction phase of BOF (as described in Section 3.2) is performed which uses K-means clustering to generate various visual words. Further, the frequencies of each visual word in the images are

represented by histograms. These histograms along with the corresponding image labels are given to SVM for training which is further used for image classification.

3.4 Experimental Results

The performance of the GKS method is analyzed in three phases on two histopathological image datasets. First, it is compared with the state-of-the-art keypoints selection methods in Section 3.4.2. Second, the results of the GKS based BOF method for classifying histopathological images are presented in Section 3.4.3. In the third phase, the performance of the new classification method has been analyzed against three state-of-the-art classification methods in Section 3.4.4.

3.4.1 Datasets

Two standard histopathological image datasets are considered for the classification task, namely ADL histopathological image dataset and Blue histology image dataset which are described below.

- *ADL histopathological image dataset [2]*: This dataset contains images of three bovine organs, namely Kidney, Lung, and Spleen. These images are further divided into two subcategories: healthy and inflammatory. Sample images are depicted in Figure 3.4. Each image is stained with hematoxylin and eosin (H&E) dye. The inflammatory tissue images, having some specific white blood cells, are used to identify the reason and duration of the infectious diseases. The presence of eosinophil cells in tissue indicates the inflammations such as parasites, allergic infections, or bacteria. There are two types of infections in the organs, acute and chronic, which can be recognized by the neutrophils cells and lymphocytes cells respectively in the tissues. From Figure 3.4, it can be observed that in the inflamed lung, the alveoli are not clear as healthy lungs have, but it is filled with bluish-purple inflammatory cells. Similarly, inflammation in other organs is also indicated by dark blue nuclei. This dataset contains 120 images per organ. This dataset is taken from Animal Diagnostics Lab (ADL), Pennsylvania State University. The ground truth label for each image is assigned by ADL pathologists by performing manual segmentation and detection of healthy and inflammatory tissues. These tissue images have special biological structures with different conditions. The infectious tissue is generally an indication of the transferable disease.

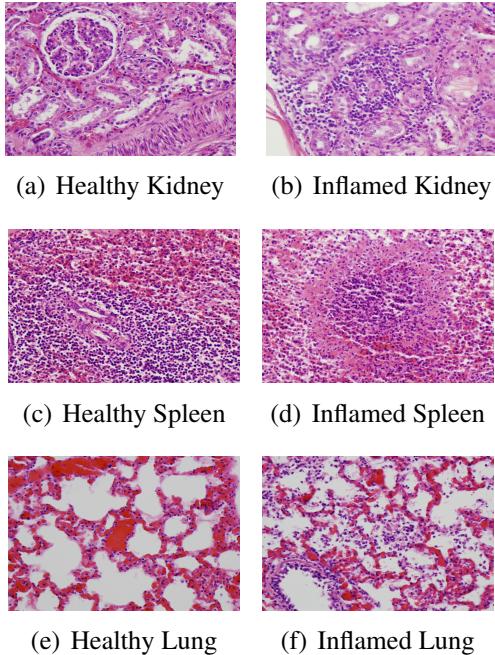


Figure 3.4: Ground truth labels for healthy and inflammatory tissues of Kidney, Lung, and Spleen [2]

- *Blue histology image dataset [164]:* This dataset consists images of four tissues, namely epithelium, connective, muscular, and nervous. Epithelium tissues include those cells that cover the organ surfaces like skin surfaces and digestive tract inner lining. The cells are linked together by the epithelial layer to provide a fence between the organ and the external environment. Epithelium tissues help to protect organs from laceration, fluid loss, and microbe. Connective tissues are gristly tissues made up of an extracellular matrix. These tissues provide shape to organs and bind them in place. Examples of connective tissues include ligament, blood, tendon, and areolar tissues. Muscle tissues are the active contractile tissues of the organ formed by muscle cells. These tissues generate a force that results in movement within internal organs or locomotion. Nervous tissues consist of different nerve cells having an axon, which are responsible for sending an activation signal to the subsequent cell. These tissues are the component of the brain and spinal cord nervous system [165]. Figure 3.5 shows sample images taken from each type of tissue images. Each image category contains 101 tissue images. Moreover, various staining methods are used to provide colors for each category tissue images and the brief description of staining methods along with tissue category is depicted in Table 3.1.

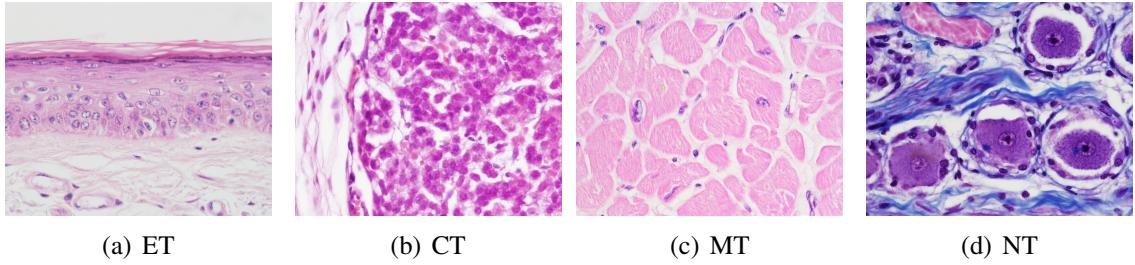


Figure 3.5: Representative animal tissues from blue histology dataset at $40\times$ magnification level [164]. Here, CT: Connective Tissue, ET: Epithelial Tissue, MT: Muscle Tissue, and NT: Nervous Tissue.

Table 3.1: The various tissue categories of blue histology dataset with their staining methods

S. No.	Tissue Category	Staining Method	Image count
A.	Connective	H&E ¹ , TRI ² , EL ³ , TB ⁴ , RET ⁵ , CCY ⁶	101
B.	Epithelial	H&E, VG ⁷ , MB ⁸ , PAS ⁹ /H&E	101
C.	Muscle	H&E, WHP ¹⁰ , IH ¹¹ , AzB ¹²	101
D.	Nervous	H&E, BC ¹³ , ICC ¹⁴ , VG, LFC ¹⁵ , H&E/MB	101

1: Hematoxylin and eosin 2: Trichrome 3: Elastin 4: Toluidine blue 5: Reticulin 6: Carbocyanine 7: Van gieson 8: Methylene blue 9: Periodic acid–schiff 10: Whipf’s polychrome 11: Iron haematoxylin 12: Alizarin blue 13: Biocytin 14: Immunocytochemistry 15: Luxol fast blue, Cresyl violet

3.4.2 Performance Analysis of the Keypoints Selection

The performance of the GKS method is evaluated against three other methods, namely IB3 [154], IKS1 [156], and IKS2 [156]. IB3 is an instance selection algorithm introduced by Aha et al. [154]. The IB3 algorithm is the extension of the nearest neighbour algorithm with high space complexity which is reduced by selective utilization filtering. The training set is considered as a union of closed hyper-curves whose instances are selected by bounded and fixed continuous distribution. Firstly, in IB3, m different keypoints are clustered into k groups that represent classes. Secondly, keypoints selection is performed by selecting only representative keypoints. A random acceptable instance I is selected and added to a set CD . If it has a different class than I , then its nearest acceptable instance from the training set is selected. The acceptability is given by the following confidence interval.

$$\frac{p + (z^2/2m) \pm z\sqrt{[p(p-1)/m] + (z^2/2m^2)}}{1 + (z^2/m)} \quad (3.8)$$

where, p is the accuracy, z is a confidence factor, and m is classification trials per instance at the time of selection. Due to its simplicity, it has been treated as a baseline algorithm for the analysis of the new GKS. The other two methods, IKS1 and IKS2, find the representative

keypoints from the images using iterative keypoints selection (IKS) method. These methods have different initialization procedure. IKS1 selects the initial keypoints randomly while IKS2 uses cluster centroids returned by K-means algorithm as initial keypoints. After the initialization of keypoints, the other keypoints are eliminated if their distances from representative keypoints are less than a predefined threshold. The parameter settings for all the considered algorithms are taken from their respective literature [156] [154].

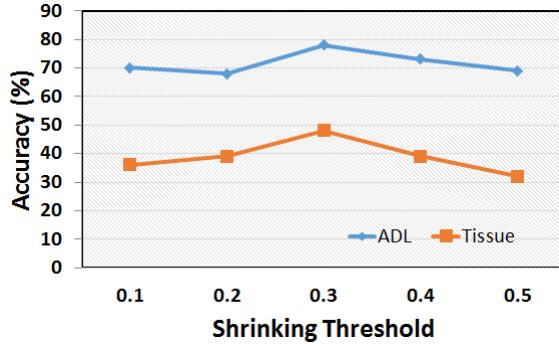


Figure 3.6: Classification accuracy on validation set using the GKS method with different shrinking threshold values

Moreover, the GKS method uses a shrinking threshold to eliminate similar points from the clusters. In this work, its value is empirically set to 0.3 using its effect on classification accuracy on validation images. To visualize the same, Figure 3.6 shows the classification efficiency on the validation images of two considered datasets for different shrinking threshold values. It can be observed from the figure that the classification accuracy on ADL and Blue histology datasets are higher at the shrinking threshold 0.3. The other parameter in the new GKS is the number of clusters for approximate k-means which is also set empirically to 1000.

The performance of the GKS method is measured in terms of a number of selected keypoints and average computation time taken by the considered algorithms. Table 3.2 depicts the total number of extracted keypoints from SURF and selected keypoints by the proposed and the considered algorithms over two datasets. The percentage of the eliminated keypoints is also mentioned for each algorithm on the different datasets in parenthesis. Table 3.2 also depicts the average computational time taken by the different algorithms. From the table, it can be observed that the IB3 algorithm eliminates 85% and 64% keypoints from ADL and Blue histology datasets respectively. However, it consumes more computational cost as its complexity is $O(n^2 \log_2 n)$ [154]. From Table 3.2, it can be observed that IKS1 and IKS2 methods eliminate almost similar amount of keypoints (41% & 44%) for Blue histology dataset. However, for ADL

Table 3.2: Number of selected keypoints returned by IB3, IKS1, IKS2, and GKS on considered datasets along with their average computational cost. The best results are in bold.

Methods	Number of keypoints		Average computational time (in hours)	
	ADL	Blue histology	ADL	Blue histology
Without keypoints selection	4177920	158720		
IB3	626688 (85%)	57139 (64%)	85.71	15
IKS1	1100000 (74%)	65948 (59%)	40	7
IKS2	2003000 (52%)	70312 (56%)	8	3.5
GKS	203000 (95%)	51000 (68%)	4	1

dataset, the reduction rate of IKS1 (74%) is higher than IKS2. As far as time complexities are a concern, both the methods take lesser time than IB3. However, the time complexity of IKS2 is $O(n \log k n)$ which is better than IKS1 whose complexity is $O(n^2)$, where k is the number of clusters. As compared to the algorithms mentioned above, the new GKS method shows the best reduction rate along with an efficient computational cost. The GKS method eliminates 95% and 68% keypoints from ADL and Blue histology datasets respectively. The time complexity of the GKS method is similar to IKS2 i.e., $O(n \log k n)$. However, the GKS method uses approximate K-means and GRA which take lesser time than K-means and Euclidean distance similarity measure, used by IKS2 respectively. This difference can be visualized from the average time taken, as mentioned in Table 3.2.

3.4.3 Classification Results of GKS based BOF Method

In this section, the efficiency of GKS for keypoints selection is validated through the BOF method for classifying the histopathological images. To make the training set, 30 images per category are selected and the remaining images in each category are added to the validation set. The size of the codebook is empirically set to 500 for visual word generation. Moreover, the performance of GKS based BOF method is analyzed using four different classifiers, namely SVM, logistic regression (LR), random forest (RF), and Gaussian naive Bayes (GNN) classifiers. Figure 3.7 shows the classification accuracy returned by the new method with different classifiers on ADL and Blue histology datasets. From the figure, it can be visualized that the new method performs better when the SVM classifier is used. Hence, for further analysis, SVM is used as the classifier in the enhanced BOF method.

For the classification of images using histograms, SVM classifier using error correcting

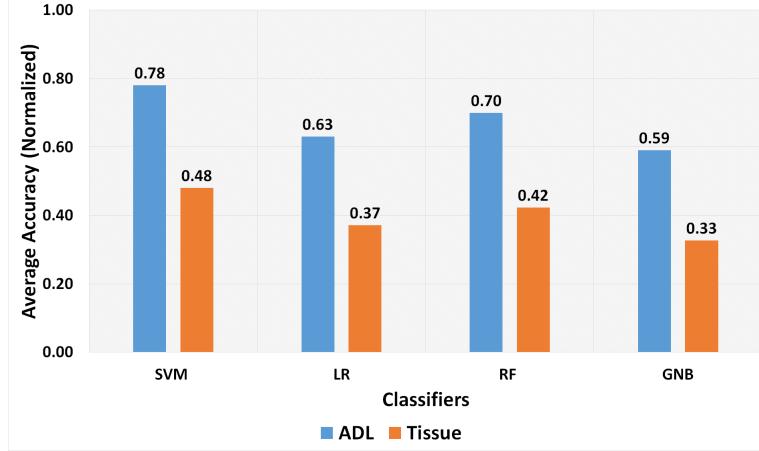


Figure 3.7: Classification accuracy on validation set using the GKS method with different classifier on ADL and Blue histology datasets

codes (ECOC) [166] is used. ECOC is an efficient method to handle multi-class classification problems and is based on aggregating of the binary classifiers. Each considered binary classifier is independent. Efficient selection of kernel function is also desirable for better classification results. In this work, the χ^2 -kernel function is used instead of linear-kernel function due to its higher performance [167]. Moreover, tenfold cross-validation is applied to prevent the over-fitting problem. The random search is also used for hyperparameter tuning which uses uniformly distributed random values and finds the optimal combination in the parameter space.

IB3		Predicted					
		KI	KN	LI	LN	SI	SN
Actual	KI	0.35	0.15	0.05	0	0	0.5
	KN	0.8	0.1	0	0	0	0.25
	LI	0.1	0.25	0.4	0.15	0	0
	LN	0	0.55	0.1	0.5	0.2	0
	SI	0	0.1	0	0.5	0.1	0.65
	SN	0	0	0	0	0.05	0.2

(a) IB3

IKS1		Predicted					
		KI	KN	LI	LN	SI	SN
Actual	KI	0.6	0.35	0	0.05	0	0
	KN	0.15	0.6	0	0.1	0.05	0.1
	LI	0.05	0.15	0.4	0.3	0.1	0
	LN	0	0	0	1	0	0
	SI	0	0	0	0	0.6	0.4
	SN	0	0.05	0.05	0	0.05	0.85

(b) IKS1

IKS2		Predicted					
		KI	KN	LI	LN	SI	SN
Actual	KI	0.8	0.15	0.05	0	0	0
	KN	0.35	0.65	0	0	0	0
	LI	0.1	0.25	0.5	0.15	0	0
	LN	0	0	0	1	0	0
	SI	0	0.1	0	0	0.25	0.65
	SN	0	0	0	0	0.05	0.95

(c) IKS2

GKS		Predicted					
		KI	KN	LI	LN	SI	SN
Actual	KI	0.85	0.15	0	0	0	0
	KN	0.05	0.8	0	0	0.15	0
	LI	0.2	0.05	0.5	0.1	0.1	0.05
	LN	0	0	0	1	0	0
	SI	0	0.1	0	0	0.7	0.2
	SN	0	0.05	0	0	0.1	0.85

(d) GKS

Figure 3.8: The confusion matrices for the ADL dataset, generated by (a) IB3, (b) IKS1, (c) IKS2, and (d) GKS based classification methods. Here, KI: Kidney-Inflamed, KN: Kidney-Normal, LI: Lung-Inflamed, LN: Lung-Normal, SI: Spleen-Inflamed, and SN: Spleen-Normal.

IB3		Predicted			
		CT	ET	MT	NT
Actual	CT	0.25	0.3	0.4	0.05
	ET	0.25	0.2	0.35	0.2
	MT	0.35	0.1	0.1	0.25
	NT	0.4	0.3	0.3	0.1

(a) IB3

IKS1		Predicted			
		CT	ET	MT	NT
Actual	CT	0.45	0.1	0.4	0.05
	ET	0.05	0.5	0.35	0.1
	MT	0.35	0.1	0.3	0.25
	NT	0.4	0.3	0.1	0.2

(b) IKS1

IKS2		Predicted			
		CT	ET	MT	NT
Actual	CT	0.5	0.2	0.3	0
	ET	0.35	0.3	0.25	0.1
	MT	0.15	0.05	0.7	0.1
	NT	0.5	0.1	0.2	0.2

(c) IKS2

GKS		Predicted			
		CT	ET	MT	NT
Actual	CT	0.75	0.1	0.15	0
	ET	0.45	0.35	0.2	0
	MT	0.2	0.05	0.75	0
	NT	0.6	0.15	0.2	0.05

(d) GKS

Figure 3.9: The confusion matrices for the Blue histology dataset, generated by (a) IB3, (b) IKS1, (c) IKS2, and (d) GKS based classification methods.

Figure 3.8 and 3.9 show the confusion matrices, generated by each considered method over ADL and Blue histology datasets respectively. The confusion matrices for ADL dataset show that IB3 based BOF method does not perform well on any of the classes, although it eliminates a significant amount of keypoints as shown in Table 3.2. That's means, it does not select the prominent keypoints. The performance of both the IKS1 and IKS2 is far better than IB3 for all the classes. However, IKS2 is slightly better than IKS1 for identifying kidney inflamed (KI) class and spleen normal (SN) class images. The performance of the new GKS based BOF method is tremendous in identifying the images of all the classes more accurately. It identifies all the lung-normal images from the ADL dataset while the worst performance is observed in the case of the lung-inflamed class which is also the best as compared to other methods. Likewise, Figure 3.9 shows the confusion matrices for the Blue histology dataset, returned by IB3, IKS1, IKS2, and GKS based BOF methods. It can be seen from the figure that a classification accuracy of 75% is returned by the GKS based BOF method for connective and muscle tissues which is better as compared to other methods. For epithelial tissue, IKS1 shows slightly better classification accuracy than the new method. For nervous tissue, IKS1 and IKS2 based methods outperform GKS and classify the images with equal accuracy. Similar to ADL dataset, IB3 does not perform well for the Blue histology dataset too.

Table 3.3: Comparative analysis of the new GKS based BOF method with other considered methods for ADL dataset. The best results are in bold.

Category	Parameters	IB3	IKS1	IKS2	GKS
Kidney - inflammation	Recall	33	60	80	85
	Precision	28	75	64	77
	F-measure	30	67	71	81
	Specificity	82	96	91	95
Kidney - normal	Recall	9	60	65	80
	Precision	9	52	57	70
	F-measure	9	56	60	74
	Specificity	79	89	90	93
Lung - inflammation	Recall	44	40	50	50
	Precision	73	89	91	100
	F-measure	55	55	65	67
	Specificity	97	99	99	100
Lung - normal	Recall	37	99	99	100
	Precision	43	69	87	91
	F-measure	40	82	93	95
	Specificity	86	91	97	98
Spleen - inflammation	Recall	7	60	25	70
	Precision	29	75	83	67
	F-measure	12	67	38	68
	Specificity	95	96	99	93
Spleen - normal	Recall	80	85	95	85
	Precision	13	63	59	77
	F-measure	22	72	73	81
	Specificity	76	90	87	95
Average Accuracy (%)		27	68	69	78

To analyze the results of confusion matrices quantitatively, recall, precision, F1-measure, specificity, and average accuracies are measured and depicted in Table 3.3 and 3.4 for ADL and Blue histology datasets respectively. From Table 3.3, it can be stated that GKS outperforms the other methods for almost all the parameters. Furthermore, the average classification accuracy of GKS on ADL dataset is 78% which is higher than other considered state-of-the-art methods, i.e., IB3, IKS1, and IKS2 which give 27%, 68%, and 69% accuracy respectively. Likewise, the new method also shows the best performance for all the tissue classes of Blue histology dataset with F1-measures equals to 65%, 50%, and 42% for muscle, epithelial, and connective respectively except nervous tissue where IKS2 shows better results. Moreover, the overall accuracy of the new method for the Blue histology dataset is 48% while IB3, IKS1, and IKS2 return 17%, 36%, and 43% accuracy respectively. However, the accuracy on the Blue histology dataset is not up to the mark due to the lots of staining variations available in the images of the Blue histology

Table 3.4: Comparative analysis of the new GKS based BOF method with other considered methods on blue histology tissue image dataset. The best results are in bold.

Category	Parameters	IB3	IKS1	IKS2	GKS
Muscle Tissue	Recall	13	30	70	75
	Precision	9	26	48	58
	F-measure	10	28	57	65
	Specificity	66	72	75	82
Connective Tissue	Recall	25	45	50	75
	Precision	20	36	33	38
	F-measure	22	40	40	50
	Specificity	66	73	67	58
Epithelial Tissue	Recall	20	34	30	35
	Precision	22	50	46	54
	F-measure	21	40	36	42
	Specificity	75	83	88	90
Nervous Tissue	Recall	9	18	20	5
	Precision	17	33	50	100
	F-measure	12	25	29	10
	Specificity	82	87	93	100
Average Accuracy (%)		17	36	43	48

dataset as depicted in Table 3.1. Especially, in the nervous tissue, LFC staining images are very much different from nervous tissue images. Therefore, its performance is degraded in all the methods.

From the results, it can be stated that the classification accuracy of the GKS based BOF method is better than the other considered methods. The baseline algorithm (IB3) gives poor performance in all scenarios as it filters out a large number of keypoints including the relevant ones. This reduces the size but also degrades the classification performance. IKS2 performs better than IKS1 as it starts with multiple reference points together and applies the reduction phase cluster-wise to reduce the overall training set. Therefore, IKS2 is fast and efficient than IKS1. In the new GKS method, the use of Grey relational analysis based similarity measure and approximate K-means make it faster and efficient. As the number of keypoints is reduced, the number of visual words is also reduced in the GKS based BOF method.

However, accuracies of the classifiers may not be the appropriate performance metric if the number of inflamed images is larger than the number of normal images. For example, if there are 1000 histopathological images in which 995 images are normal and 5 images are inflamed. If the proposed method classifies all images as inflamed, then the accuracy would be 99.5%. However, the method miss-classify all the normal cases. To mitigate this, G-mean ($\sqrt{TP * TN}$)

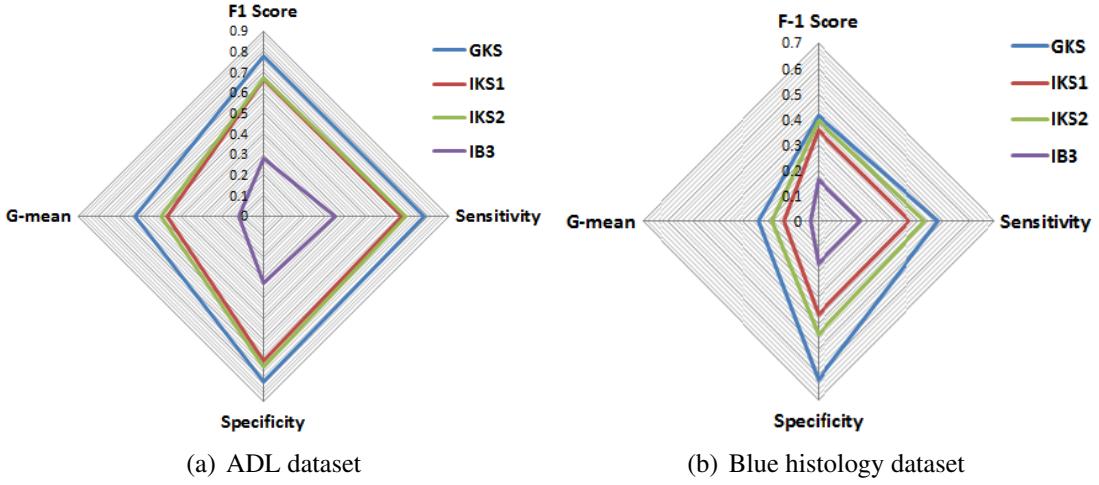


Figure 3.10: Radar charts for average results obtained for SVM classifier on a) ADL dataset and b) Blue histology image dataset by considering F-1 score, sensitivity, specificity, and G-mean

is considered as the performance metric which includes true positives (TP) and true negatives (TN) in a product. Furthermore, the performance of IB3, IKS1, IKS2, and GKS methods are also analyzed using radar charts which are shown in Figure 3.10 that depict four evaluation criteria, namely F1 score, sensitivity, specificity, and G-mean which resulting in four-sided shape. The method with a maximum area and symmetrical shape perform better than others. From the figure, it can be observed that the GKS based BOF method achieves better results among all the other methods in four considered measures. Therefore, it can be stated that the new keypoints selection in the BOF method outperforms the other keypoints selection methods and may be applied for histopathological image classification.

3.4.4 Comparative Analysis of GKS based BOF with State-of-the-Art Methods

The performance of the GKS based BOF method is also compared with the three state-of-the-art methods for ADL histopathological image dataset, namely WND-CHARM [168], SRC [169], and SHIRC [2] in terms of recall, specificity, precision, false negative rate (FNR), average accuracy, and F1-score. Shamir et al. [168] introduced a method for the analysis of biological images in which image content features are detected from the raw images and selected informative feature descriptors are used to train the classifier. In the sparse representation-based classification (SRC) method [169], RGB images are represented by a single luminance channel and this representation is used to train the classifier. Moreover, this work is further extended to three color channels and known as a multi-channel simultaneous sparsity model (SHIRC) [2].

Table 3.5: Classification performance of GKS based BOF method with other state-of-the-art methods

Organ	Algorithms	Recall	Specificity	Precision	FNR	F1-score	Avg. Accuracy (%)
Kidney	WND-CHARM	0.690	0.720	0.710	0.280	0.700	71.0
	SRC	0.875	0.750	0.778	0.250	0.825	81.3
	SHIRC	0.825	0.833	0.832	0.167	0.828	82.9
	BOF	0.870	0.650	0.731	0.350	0.826	80.0
	GKS	0.950	0.890	0.888	0.110	0.879	88.0
Lung	WND-CHARM	0.725	0.626	0.705	0.374	0.791	75.7
	SRC	0.880	0.765	0.750	0.235	0.737	74.5
	SHIRC	0.750	0.850	0.833	0.150	0.791	80.0
	BOF	0.730	0.750	0.745	0.250	0.737	74.0
	GKS	0.888	0.860	0.863	0.140	0.871	87.0
Spleen	WND-CHARM	0.512	0.873	0.800	0.128	0.640	69.2
	SRC	0.708	0.792	0.773	0.208	0.740	75.0
	SHIRC	0.650	0.883	0.848	0.117	0.742	76.7
	BOF	0.880	0.530	0.652	0.470	0.749	70.5
	GKS	0.750	0.880	0.862	0.120	0.804	81.5

This method is also analyzed and validated on ADL histopathological images. Table 3.5 shows the results of each considered method on the various performance parameters, namely recall, specificity, precision, false negative rate (FNR), and F1-score to identify the inflamed images of each organ in ADL dataset. Recall and specificity are the two key statistics to validate the performance of classification in medical diagnosis. Recall is the probability to identify diseased images correctly, while specificity returns the probability of identifying the healthy images correctly. In histopathological image analysis, it is always important to identify inflamed images with higher accuracy. From Table 3.5, it can be noticed that the new GKS method has high recall values of 95%, 88.8%, and 75% for Kidney, Lung, and Spleen organs respectively. Moreover, the true negative rates returned by the GKS method are 89%, 86%, and 88% for Kidney, Lung, and Spleen organs respectively. Hence, it can be stated that the GKS method also identifies healthy images more accurately as compared to the other considered methods. Further, the GKS based BOF method also attains high average accuracy, precision, and F-1 score. The results have also been analyzed on the FNR which can be defined as the rate of identifying inflamed images as healthy images. It is very dangerous in medical diagnosis and it should be minimized. The GKS method has the lowest FNR of 11% and 14% on Kidney and Lung organ images respectively. However, for Spleen organ images SHIRC outperforms the GKS method in terms of FNR.

3.5 Summary

In this chapter, a new GKS method of keypoints selection has been presented which improves the efficiency of the bag-of-features method. The method uses Grey relational analysis and approximate K-means for the elimination of irrelevant and similar keypoints. Furthermore, the new keypoint selection method has been incorporated in the BOF method to reduce its computational complexity. Moreover, the support vector machine is used to train and classify the images into respective categories. The GKS based BOF method is tested on two complex histopathological image datasets. The simulation results and classification accuracy show the efficacy of the new method on the other considered methods.

The next chapter introduces an optimal codebook construction method to be used in the BOF for histopathological image classification.

CHAPTER 4

OPTIMAL CODEBOOK CONSTRUCTION

4.1 Introduction

The effectiveness of the BOF method is highly dependent on the formed visual vocabulary which is generated by the clustering of feature descriptors in the codebook construction phase. Clustering, an optimization problem, is a method to assign all the data points to certain clusters by using minimum cohesion or maximum separation. For the same, K-means clustering is generally employed to cluster feature vectors as it is computationally fast. However, it may generate biased centers when applied to high dimensional descriptor space extracted from tissue images. Therefore, there is a requirement of generating optimal visual words that are not biased and this can be achieved with the use of optimization algorithms that can work well for the clustering problem. In literature, various meta-heuristic methods have exhibited better performance for clustering [170] [171]. Out of these, biogeography-based optimization (BBO) [93] is one of the popular methods based on island biogeography and is used effectively in clustering problems and computer vision applications [172]. However, BBO also suffers from some demerits like single feature migration property, poor population diversity, and sticking into local optima sometimes [173]. To mitigate these demerits, two new variants of BBO, namely improved biogeography-based optimization (IBBO) and spiral biogeography-based optimization (SBBO) are introduced in this chapter which is described in the following sections. Further, these variants are used to generate the optimal visual vocabulary in the BOF method which is used to classify the histopathological images. The modified BOF method with the optimal codebook construction method is shown in Figure 4.1.

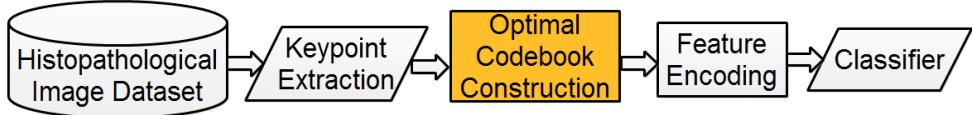


Figure 4.1: Flow chart of the enhanced BOF method with optimal codebook construction

4.2 Biogeography-based Optimization Algorithm

BBO [93] is an evolutionary optimization algorithm inspired by the mathematical model of island biogeography [174]. The island biogeography model discusses the species migration among islands, new species generation, and species extinction. Moreover, the islands with good features share the information with the bad islands to improve them. These features are called SIVs (suitability index variables) such as temperature, diversity, land area, rainfall, etc. In BBO, each island represents an individual of the population and SIVs are considered as independent variables. The cost of each is represented by ISI (island suitability index) which is a function of SIVs. Let there are n islands and d SIVs, then the fitness (ISI) of i^{th} island can be represented by Eq. (4.1).

$$ISI_i = f(SIV_1, SIV_2, \dots, SIV_{d-1}, SIV_d) \quad i = 1, 2, 3, \dots, n \quad (4.1)$$

where, f is the objective function defined over d dimensions.

To control the movement of species between the islands, two migration rates, namely immigration (λ) and emigration (μ) rates are defined. Both λ and μ are the functions of species count in the island and can be defined as Eq. (4.2) and (4.3) respectively.

$$\lambda_i = I * \left(1 - \frac{S_i}{S_{max}}\right) \quad (4.2)$$

$$\mu_i = E * \left(\frac{S_i}{S_{max}}\right) \quad (4.3)$$

where, I and E are the maximum values of immigration and emigration rates respectively and both are generally set to 1; S_i is the species count of the i^{th} island at current time and space for $i = 1, 2, 3, \dots, n$. The graphical model of the species movement in a single island is depicted in Figure 4.2. If the island contains zero species, then the immigration rate (λ) is maximum which is denoted by $\lambda = I$ and it decreases as the species count increases in the island. When the island achieves the maximum possible species count (S_{max}), the λ becomes zero and emigration rate

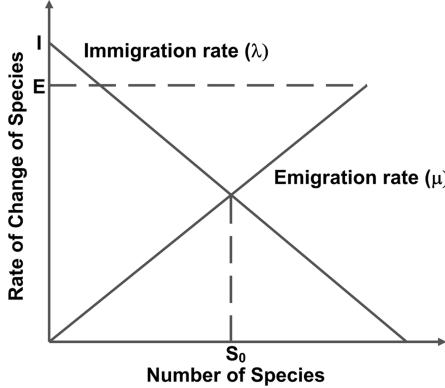


Figure 4.2: Linear equilibrium model of an ecological system for a single island [93]

(μ) becomes maximum ($\mu = E$). Hence, more number of species will leave the island. As the number of species reaches to zero, the emigration of species is not possible and μ will become zero. In the figure, S_0 denotes the equilibrium species count, which is achieved at $\lambda = \mu$.

BBO initializes the population of individuals in the specified continuous or discrete domain and creates a new population by modifying the existing individuals with the help of three operators, namely elitism, migration, and mutation. BBO follows the elitism property by moving one or more good solutions from the current population to the next generation which perpetuate the quality of solutions in subsequent generations.

Migration Operator: Migration is the process in which high ISI solutions, also known as emigrating islands, share its features to low ISI solutions, termed as immigrating islands. Immigrating islands are selected by island modification probability (P_{modify}) and immigration rate (λ) is used to select some of their features which are to be replaced by corresponding features of emigrating islands. Further, the emigration rate (μ) is used by the roulette wheel selection method to select more than one emigrating islands for the same. This process of migration generates new islands that may be selected for the next generation. Algorithm 4.1 presents the detailed description of migration process.

Mutation Operator: Mutation is the abrupt extreme alteration in the island due to some calamitous or catastrophic events. It increases the diversity in the islands. In BBO, mutation is performed by replacing some of the SIVs of each island with randomly generated SIVs. These SIVs are chosen on the basis of the mutation rate (σ) of their corresponding islands. The

Algorithm 4.1 Migration Operator [175]

Input: A population P of n islands. Each island has d suitability index variables (SIVs).

Let P_{modify} is the island modification probability.

Output: Modified population P .

```
for i = 1 to n do
    if rand() < Pmodify then
        continue;
    end if
    for j = 1 to d do
        if rand() < λi then
            Select emigrating island,  $island_k$ , according to  $μ_i$  using roulette wheel selection;
             $island_i(SIV_j) \leftarrow island_k(SIV_j)$ 
        end if
    end for
end for
```

mutation rate $σ_i$ of i^{th} island is calculated by using Eq. (4.4).

$$σ_i = mu_{max} * (1 - \frac{P_i}{P_{max}}) \quad (4.4)$$

where, P_i is the probability of species on i^{th} island at the current time, mu_{max} is the maximum mutation probability, and P_{max} is the maximum probability. Algorithm 4.2 [175] demonstrates the mutation process of BBO. Moreover, the process of BBO Algorithm has been presented in Algorithm 4.3 [175].

Algorithm 4.2 Mutation Operator [175]

Input: Let population P has m islands ($island_i$, $i = 1, 2, \dots, n$) and $σ_i$ is the mutation rate of i^{th} island.

Output: Modified population P

```
for i = 1 to n do
    for j = 1 to d do
        if σi > rand() then
            Generate a feasible random value R in solution space;
             $island_i(SIV_j) \leftarrow R$ 
        end if
    end for
end for
```

Sometimes it has been observed that BBO may trap into local optimum and shows slow convergence behavior due to poorly generated population diversity by its mutation and migration

Algorithm 4.3 Biogeography-based Optimization [175]

Input: A population P has n islands ($island_i$, $i = 1, 2, \dots, n$)

Output: An optimal island.

Measure the value of ISI (fitness) for each island

while stopping condition **do**

Determine the λ_i and μ_i for each island $island_i$;

Perform migration;

Perform mutation;

Replace one or more worst islands by the previous best islands to perform elitism;

end while

Select the best island with high ISI.

operators [176]. Therefore, researchers have developed different variants of BBO by modifying its migration and mutation operators to enhance the performance. Due et al. [177] hybridized BBO with immigration refusal (RE) and evolutionary strategy (ES) to escape from local optimum. Immigration refusal has been used for migration between the individuals while ES has been used to mutate the individuals. Moreover, Gong et al. [178] introduced a real-coded BBO which represents each solution by real parameters in the continuous domain. Further, to enhance the exploration of the algorithm, they introduced three novel mutation operators, namely levy, Cauchy, and Gaussian. Ma et al. [179] improved the exploration phase of BBO by introducing a blended migration operator and solved the constrained optimization problem. Further, Lohokare et al. [180] introduced a memetic technique to refine the search capability of BBO in which a modified DE has been used with mutation operators to find global optimum in search space. Moreover, Ma et al. [181] introduced four variants of the migration process, namely BBO with total immigration (BBOTI), BBO with partial immigration (BBOPI), BBO with total emigration (BBOTE), and BBO with partial emigration (BBOPE). The migration process of BBOPI uses immigration rate (λ) to immigrate an island feature followed by the selection of the emigrating islands. However, BBOTI uses λ for immigration of a whole island followed by the selection of the emigrating islands. In contrast to this, BBOTE and BBOPE choose emigrating islands followed by the immigrating island. Therefore, BBOTI and BBOPI use the immigration rate to decide whether migration will occur or not, while BBOTE and BBOPE use the emigration rate for the same. Simon et al. [182] proposed the rotationally invariant migration operator to enhance the performance for non-separable functions. Further, they used a gradient descent operator to overcome the weakness of premature convergence. Moreover, for the real

world problems, relying on the constrained boundary, they introduced a global grid search and boundary search to cover the search space systematically. Re-initialization and restart procedures have also been introduced. Gong et al. [183] introduced a hybrid version of BBO and DE in which exploitation is done by BBO while DE performs the exploration to generate an optimal solution. Further, Niu et al. [184] introduced various mutation strategies for BBO (BBO-M) to maintain the diversity and solutions quality of the population. The modification incorporates the DE with chaos theory in mutation operator. Lim et al. [173] modified the mutation operator of BBO by replacing it with the Tabu search which maintains the population diversity and quality of the population. Further, Al-Roomi et al. [185] proposed a novel variant of BBO for improving the exploration capability by hybridizing it with simulated annealing (SA). They used five different cooling strategies and metropolis criteria of SA to select superior migrated islands. Garg and Deep [186] proposed a Laplace crossover based BBO, namely LX-BBO, to solve continuous function optimization problems. In addition to above mentioned variants, several other modifications of BBO have also been proposed in literature [187], [188], [189], [190], [191], [172]. However, most of the variants of BBO either modified migration operator to increase the exploitation capabilities or mutation operator for exploration. In general, migration operator follows the single feature migration property which generates poor performance for non-separable functions [93], [192], [182]. Moreover, due to poor population diversity, generated by mutation operator increases the chance to converge into local optimum [93], [179], [181].

To overcome the above-mentioned limitations, this work introduces two new variants of BBO known as improved BBO (IBBO) and spiral BBO (SBBO) which are discussed in subsequent sections.

4.2.1 Improved Biogeography-based Optimization

The existing BBO, as described in the previous section, has two main flaws. The first one is that migration operator follows the single feature migration property i.e., each SIV is handled individually and only a few are updated. This property works well for separable functions in which all the variables are independent. However, a non-separable function contains the inter-dependent variables and that's why the single feature migration property of BBO algorithms on these functions shows poor performance [182], [193]. Secondly, poor population diversity is

generated by mutation operator because it replaces some of the selected SIVs of an island with random SIVs. Since, the generated SIVs take random values, the quality of the mutated islands is not guaranteed and quite often solution may trap into local optimum [173], [179], [181]. Therefore, in the new improved BBO (IBBO) algorithm, both the migration and mutation operators are modified. The details description of improved operators is provided below. However, the sequence of these operators is similar to the original BBO as described in Algorithm 4.3.

4.2.1.1 Improved Migration Operator

The proposed migration operator considers two modifications in the basic migration operator. First, it generates a new island by modifying all the features of the immigrating island, irrespective of original BBO where only limited features are modified. Second, it also uses the best individual along with other islands as emigrating islands while in the original BBO, the best island is not explicitly selected for migration.

Similar to BBO, the immigrating islands are selected with the help of modification probability P_{modify} . If this probability is greater than a random number, then the individual is selected as immigrating islands for modification. Further, the emigrating islands are selected with the help of roulette wheel selection based on μ i.e., emigration rate. Let each island has d number of SIVs. Now, d number of random numbers are generated. If a generated random number is less than the immigration rate (λ) then corresponding SIV of immigrating island is replaced by SIV of emigrating island otherwise, selected SIV takes its value from the corresponding best island SIV. This way, all the features of immigrating islands are modified.

In the proposed migration method, the replacement of all SIVs in place of limited SIVs increases exploration capability. Moreover, the use of the best solution increases the chance of generating an optimal result with better convergence rate. Therefore, it also shows efficient results for non-separable functions. The detailed improved migration process is also depicted in Algorithm 4.4.

4.2.1.2 Improved Mutation Operator

Mutation is one of the important operators of BBO which is used to create diversity in the population and helps to explore the search space. However, the mutation operator of BBO only modifies some of the SIVs of randomly selected islands due to which it shows poor diversity in

Algorithm 4.4 Improved Migration Operator

Input: A population P of n islands. Each island has d suitability index variables (SIVs).

Let P_{modify} is the island modification probability.

Output: Modified population P .

for $i = 1$ to n **do**

if $rand() < P_{modify}$ **then**

 continue;

end if

for $j = 1$ to d **do**

if $rand() < \lambda_i$ **then**

 Select emigrating island, $island_k$, according to μ_i using roulette wheel selection;

$island_i(SIV_j) \leftarrow island_k(SIV_j)$

else

$island_i(SIV_j) \leftarrow island_{best}(SIV_j)$

end if

end for

end for

the population.

The proposed method introduces a biased mutation operator based on a random walk and step size. In the biased mutation operator, all the SIVs of randomly selected poor islands are modified. Let m_i is the mutation rate of i^{th} island which is used to select it for mutation. The mutated island for next iteration ($t + 1$) is generated by Eq. (4.5).

$$island_i(t + 1) = island_i(t) + s_i * r(t), i = 1, 2, \dots, n \quad (4.5)$$

where, $r \in (0, 1)$ is randomly generated number. Further, s_i denotes the step size of i^{th} island for a random walk. The value of step size (s) plays a major role to explore the solution space. The islands with very large step size (s) may produce out of bound SIVs which is unacceptable. Moreover, if the step size taken by an island is too small then it will be incapable to explore the search space. Therefore, it is desirable to use a balanced step size for improving the population diversity. The proposed mutation operator obtains the step sizes by differencing the randomly generated permutations of the existing population.

Let, population P has n islands ($island_i$), as represented by Eq. (4.6), and $P_k^n(i)$ represents the randomly generated k^{th} permutation of i^{th} island. The proposed step size (s_i) for i^{th} island

is calculated by Eq. (4.7).

$$P = \{island_1, island_2, \dots, island_{n-1}, island_n\} \quad (4.6)$$

$$s_i = P_1^n(i) - P_2^n(i), \quad i = 1, 2, \dots, n \quad (4.7)$$

The difference between two randomly generated permutations of an island for the current population gives a balanced and random step size. Therefore, the new mutation operator performs random walk with biased behavior which reduces the chances to trap into local optimum and increases population diversity. The detailed process of improved mutation operator is also depicted in Algorithm 4.5.

Algorithm 4.5 Improved Mutation Operator

Input: Let population P has n islands ($island_i$, $i = 1, 2, \dots, n$) and σ_i is the mutation rate of i^{th} island.

Output: Modified population P

for $i = 1$ to n **do**

if $\sigma_i > rand()$ **then**

 Randomly generate two permutation ($P_1^n(i)$ and $P_2^n(i)$) of i^{th} island;

$s_i = P_1^n(i) - P_2^n(i)$

$island_i(t+1) = island_i(t) + s_i * r(t)$

 Check whether $island_i(t+1)$ is feasible or not;

 If not then map to the original bounds;

end if

end for

4.2.2 Spiral Biogeography-based Optimization

The existing BBO, as described in Section 4.2 has poor population diversity which is generated by a mutation operator because it replaces some of the selected SIVs of an island with random SIVs. Since the generated SIVs take only random values, the qualities of the mutated islands are not guaranteed and quite often solution may trap into local optimum [179, 173]. Therefore, in the new spiral BBO (SBBO) algorithm, the mutation operator is modified by introducing a spiral search phase in mutation along with random search. The rest of the steps of SBBO remains similar to the original BBO as described in Algorithm 4.3.

The new mutation operator generates new SIVs by using two searches; (i) spiral search and

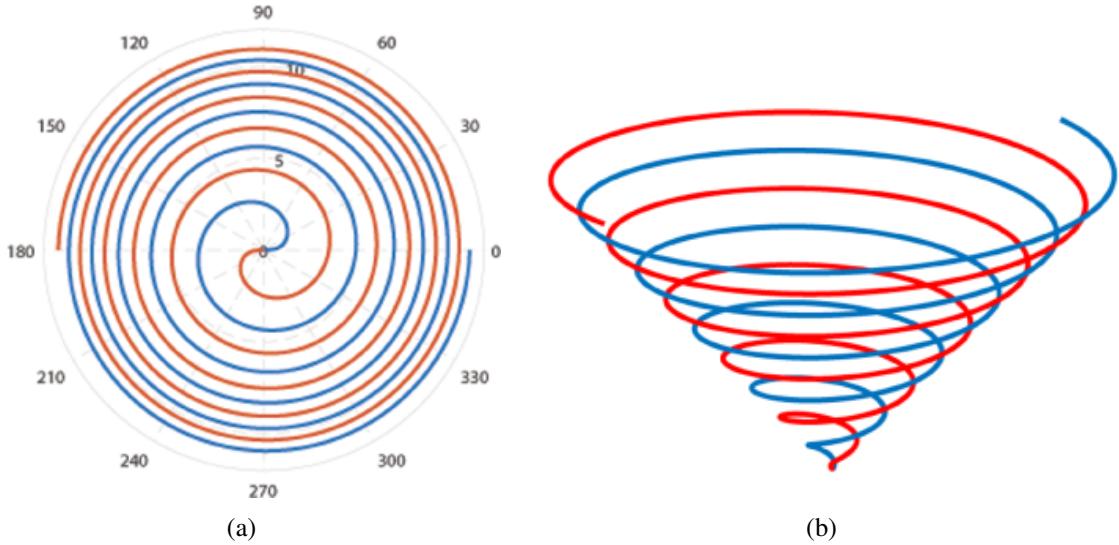


Figure 4.3: The Fermat's Spiral Function in (a) 2D and (b) 3D with radius r (0 to 10) and angle θ

(ii) random search, which is explained in the following sections.

4.2.2.1 Spiral Search

In this phase of mutation operator, some of the SIVs for the new island are searched in a spiral trajectory, as defined by Fermat's spiral function, around the best island so far as shown in Figure 4.3 for a two and three-dimensional spaces with radius r and angle θ . In the figure, red and blue spirals correspond to negative and positive r respectively for any given positive value of θ . The generated spiral trajectory looks like tornadoes and hurricanes. In the BBO, new SIVs are generated using the concept of Fermat's spiral. Fermat's spiral is a type of Archimedean spiral, also known as a parabolic spiral, and is developed by Fermat et al. [194]. Mathematically, it can be represented by Eq. (4.8).

$$r = \pm \theta^{1/2}, \quad (4.8)$$

where, θ is an angle of rotation and r is a radius of convergence. In polar coordinates, it can be represented as Eq. 4.9.

$$r^2 = a^2\theta \quad (4.9)$$

here, a is a constant. Further, in a parametric equation, it can be presented as Eq. (4.10) along x-axis.

$$x = \text{sgn}(a) \cdot |a| \cdot \theta^{(1/2)} \cdot \cos \theta \quad (4.10)$$

where, $sgn(a)$ corresponds to the sign function used to extract the sign of a real number. This paper uses the similar movement of species in mutation operator around the best island so far and is defined mathematically by Eq. (4.11).

$$Island_i(t+1) = Island_{Best}(t) + sgn(R_i(t)) \cdot |R_i(t)| \cdot l^{(1/2)} \cdot \cos(2 * \pi * l) \quad (4.11)$$

where, $R_i(t)$ is the absolute distance between the best island ($Island_{Best}$) and i^{th} island at t^{th} iteration and l is randomly generated by Eq. (4.12).

$$l = 1 + t * rand() / max_{iteration} \quad (4.12)$$

4.2.2.2 Random Search

The population diversity in BBO can be increased by random search by generating the new SIVs using the randomly selected islands. The mathematical model for random search is given in Eq. (4.13).

$$Island_i(t+1) = Island_{random}(t) - R_i(t) \quad (4.13)$$

where, $Island_{random}(t)$ is any random island at iteration t , $R_i(t)$ is a residual distance from the randomly selected island and is given by Eq. (4.14).

$$R_i(t) = \|rand() * Island_{random}(t) - Island_i(t)\| \quad (4.14)$$

here, $rand()$ is a random function for generating the random values between 0 & 1 and $Island_i$ is the i^{th} individual in the population. The complete improved mutation operator is also presented in Algorithm 4.6.

4.3 Optimal Codebook Construction

The new variant of BBO, namely IBBO and SBBO are used to generate the optimal visual words in the BOF method. The BOF method is modified by introducing a new clustering method in place of K-means. The new clustering method uses IBBO/SBBO to find optimal cluster centers. The nearby features of the cluster centers are considered as the optimal visual words. The complete process of the modified BOF method is described below.

Algorithm 4.6 Improved Mutation Operator

Input: Let population P has n islands ($island_i, i = 1, 2, \dots, n$).

Output: Modified population P

for $i = 1$ to n **do**

if $rand() \leq 0.5$ **then**

 Update the $Island_i$ using Eq. (4.11);

else

 Select a random island $Island_{random}$;

 Update the $Island_i$ using Eq. (4.13);

end if

end for

1. The D -dimensional feature vectors are extracted from the collection of histopathological images.
2. Optimal set of visual words are computed using the following IBBO/SBBO based clustering methods.
 - (a) Let the set F of local features are $\{f_1, f_2, \dots, f_N\}$ where $f_n \in \mathbb{R}^D$. Apply K-means on the set F to find the K cluster centers $\{c_1, c_2, \dots, c_K\}$ where, $c_k \in \mathbb{R}^D$.
 - (b) Initialize the population of M islands. Each island is a vector of K features that are nearby to the centroids returned K-means.
 - (c) Calculate the fitness of each island by considering the following objective function

$$Minimize \quad \delta(f_m, c_k) = \sum_{i=1}^N \sum_{j=1}^K \| f_m - c_k \|^2 \quad (4.15)$$

where, δ represents the compactness of the clusters and the objective is to minimize the compactness for better clustering.

- (d) Determine better solutions to perform elitism.
 - (e) Improve the quality of the population by performing migration and mutation operators of IBBO/SBBO.
 - (f) Repeat Steps (c) to (e) until the termination criteria are not achieved.
 - (g) Determine the best island in the last iteration.
3. The nearby features of the best island are considered as visual words.

Table 4.1: Summary of standard benchmark problems

Sr. No.	Function Name	Range	Optimal Value	Optimal Position Values	Category
1.	Ackley	[-35 to 35]	0	(0,...,0)	MM, NS
2.	Alpine	[-10 to 10]	0	(0,...,0)	MM, S
3.	Brown	[-5 to 5]	0	(0,...,0)	UM, NS
4.	Levy	[-10 to 10]	0	(1,1,...,1)	MM
5.	New Schwefel	[-500 to 500]	0	(420.9687,..., 420.9687)	MM
6.	Pathological	[-100 to 100]	0	(0,...,0)	MM, NS
7.	Penalty1	[-50 to 50]	0	(1,1,...,1)	MM, NS
8.	Penalty2	[-50 - 50]	0	(1,1,...,1)	MM, NS
9.	Powell's First Singular	[-4 - 5]	0	(0,0,...,0)	UM, NS
10.	Powell's Second Singular	[-4 - 5]	0	(0,0,...,0)	UM, NS
11.	Powell Sum	[-1 - 1]	0	(0,0,...,0)	UM, S
12.	Quartic	[-1.28 - 1.28]	0	(0,0,...,0)	UM, S
13.	Rastrigin	[-5.12 - 5.12]	0	(0,0,...,0)	MM
14.	Rotated Hyper-Ellipsoid	[-65.536 - 65.536]	0	(0,0,...,0)	UM
15.	Schwefel	[-512 - 512]	-12965.5	(420.9687,..., 420.9687)	MM, S
16.	Schwefel3	[-10 - 10]	0	(0,0,...,0)	MM, NS
17.	Sphere	[-5.12 - 5.12]	0	(0,0,...,0)	UM, S
18.	Step	[-100 - 100]	0	(0,...,0)	UM, S
19.	Sum Squares	[-10 - 10]	0	(0,0,...,0)	UM, S
20.	Trigonometric	[0 - π]	0	(0,0,...,0)	MM, NS

4. Each image is encoded by the optimal set of visual words in the form of a histogram.
5. The encoded histograms along with the image annotations are used to train the SVM classifier. Finally, the trained SVM classifier is tested on the validation dataset.

4.4 Experimental Results

This section presents the results of IBBO and SBBO methods along with the IBBO/SBBO based BOF method for histopathological image classification. Section 4.4.1 analyzes the effectiveness of IBBO and SBBO on various benchmark functions while experimental results of histopathological image classification on ADL and Blue histology datasets are presented in Section 4.4.2.

4.4.1 Performance Analysis of IBBO and SBBO

The performance of newly introduced IBBO and SBBO have been analyzed on 20 representative benchmark functions as listed in Table 4.1 [89] [195] [196] and 29 bound-constrained CEC 2017 benchmark problems [197] as depicted in Table 4.2. The performance has been evaluated and statistically analyzed using three performance parameters, namely mean fitness value, rank value using Friedman test, and convergence.

Each considered benchmark function of Table 4.1 belongs to either unimodal (UM) or multi-

Table 4.2: Overview of CEC 2017 benchmark problems

S.No.	Function	Optimal Value	Characteristic
C_1	SR Bent Cigar Function	100	Unimodal
C_2	SR Sum of Different Power Function	200	
C_3	SR Zakharov Function	300	
C_4	SR Rosenbrock's Function	400	Multi-modal
C_5	SR Rastrigin's Function	500	
C_6	SR Expanded Scaffer's F6 Function	600	
C_7	SR Lunacek B_Rastrigin Function	700	
C_8	SR Non-Continuous Rastrigin's Function	800	
C_9	SR Levy Function	900	
C_{10}	SR Schwefel's Function	1000	
C_{11}	HF 1	1100	Hybrid
C_{12}	HF 2	1200	
C_{13}	HF 3	1300	
C_{14}	HF 4	1400	
C_{15}	HF 5	1500	
C_{16}	HF 6	1600	
C_{17}	HF 6	1700	
C_{18}	HF 6	1800	
C_{19}	HF 6	1900	
C_{20}	HF 6	2000	
C_{21}	CF 1 (N=3)	2100	Composition
C_{22}	CF 2 (N=3)	2200	
C_{23}	CF 3 (N=4)	2300	
C_{24}	CF 4 (N=4)	2400	
C_{25}	CF 5 (N=5)	2500	
C_{26}	CF 6 (N=5)	2600	
C_{27}	CF 7 (N=6)	2700	
C_{28}	CF 8 (N=6)	2800	
C_{29}	CF 9 (N=3)	2900	
C_{30}	CF 10 (N=3)	3000	

SR: Shifted and Rotated; HF: Hybrid Function; CF: Composite Function

modal (MM) class and/or separable (S) or non-separable (NS) class. Table 4.1 also depicts the search space, optimal value, category, the optimal solution value for each benchmark function. Moreover, the functions of CEC 2017 contain novel basic problems, composing test problems, rotated trap problems, graded level linkages, and many other problems. These benchmark problems show several novel features like composing 3, 4, 5, and 6 test problems by considering features from various problems and rotated trap problems [198]. In CEC 2017, 30 problems were defined in which three problems are unimodal (C_1-C_3), seven problems are multi-modal (C_4-C_{10}), ten problems are hybrid ($C_{11}-C_{20}$), and ten are composition based problems ($C_{21}-C_{30}$). Every function contains the search space $[-100, 100]^D$ where D is the dimension. Function C_2 is excluded from the competition due to its unstable behavior. Figure 4.4 shows the 3D map of two multi-modal (C_5 and C_6) and two composition (C_{22} and C_{28}) functions respectively. From the figure, it can be visualized that functions C_5 and C_6 contain multiple local optima and global optimum is not near to the second better local optimum. Further, C_{22} and C_{28} have different

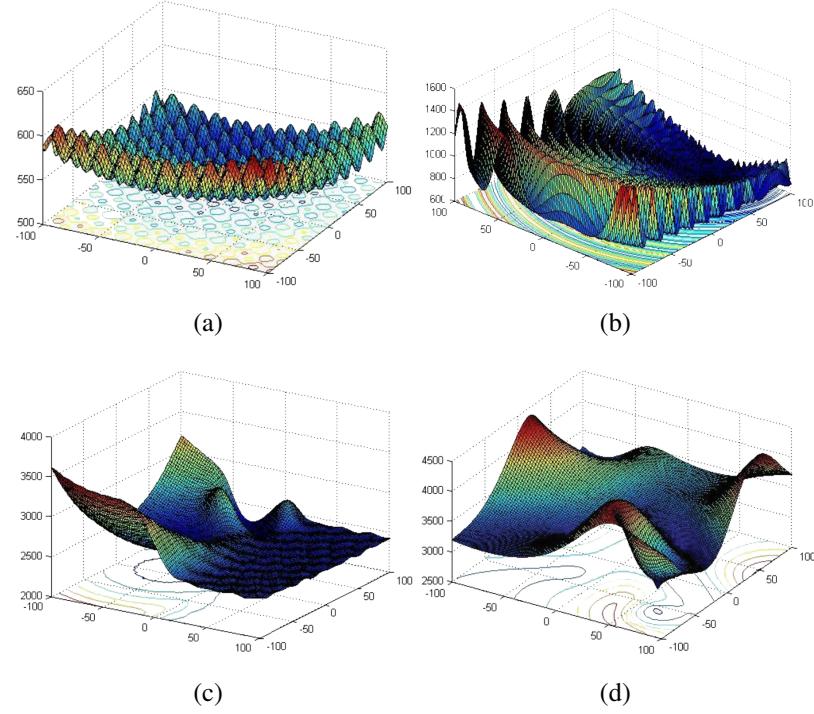


Figure 4.4: 3D map for (a) C_5 , (b) C_6 , (c) C_{22} , and (d) C_{28} benchmark problems of CEC 2017 [197]

properties around different local optima due to its composite nature.

For the comparative analysis of SBBO and IBBO, one variant of EA, namely LSHADE (differential evolution based on success history and population size) [199], two variants of BBO, namely BBO-M [184], LX-BBO [186], and four variants of swarm-based approaches, namely SSA [200], GWO (grey wolf optimizer) [201], ckGSA (chaotic GSA) [202], and WOA (whale optimization Algorithm) [203] have been simulated along with the BBO algorithm. The parameter settings of all the considered algorithms have been set according to their literature except population size (N), number of dimensions (D), and number of iterations (itr) which are kept to 100, 50, and 1000 respectively. To minimize the interference, the simulations have been executed 30 times.

Table 4.3 & 4.4 depict the mean fitness values for 30, 60, and 90 dimensions, generated by each existing and new methods on standard benchmark functions. Table 4.3 shows the comparative view for the functions F_1 to F_{10} while, F_{11} to F_{20} are depicted in Table 4.4. In the tables, bold entries indicate the best results. It is envisioned from the table that IBBO and SBBO perform much superior to the existing methods, however between the two, SBBO performs the best. Moreover, Table 4.5 depicts the comparison of the IBBO and SBBO with considered algorithms on unimodal, multi-modal, hybrid, and composite benchmark problems

of CEC 2017 in terms of mean fitness values. From the table, it can be stated that mean fitness values returned by SBBO are better than the other algorithms for 23 problems out of 29. For the unimodal problem C_1 , BBO-M shows better mean fitness value. However, for the same problem, the second best mean fitness value is returned by SBBO. Similarly, for the multi-modal and composite benchmark problems, SBBO outperforms on all the problems. Moreover, for hybrid benchmark problems, SBBO shows better results except C_{11} , C_{13} , C_{15} , and C_{19} where BBO-M and LX-BBO perform better. From Table 4.5, it can be seen that IBBO does not perform adequately for CEC 2017 functions

To statistically validate the results presented in Table 4.3 to 4.5, Friedman's test [204] has been conducted among IBBO, SBBO and considered algorithms. Friedman's test is a non-parametric statistical test having two hypotheses, H_0 (null hypothesis) and H_1 (alternate hypothesis). It has been performed by considering the mean fitness values over 30 runs for standard benchmark functions and CEC 2017 benchmark problems and results are presented in Tables 4.6 & 4.7. The null hypothesis H_0 refers that there is no significant difference among the samples and the alternative hypothesis signifies that all samples are significantly different. In Table 4.6, the p-values are $7.45e - 16$, $8.58e - 16$, and $4.56e - 10$ for 30, 60, and 90 dimensions respectively which are less than the considered significance level i.e., $\alpha = 0.05$ that indicates that the null hypothesis (H_0) is rejected and obtained results are significantly different. Similarly, in Table 4.7, the Friedman's test returns a p-value to $5.38e - 36$. The lower p-value signifies that the evidence is against H_0 . Furthermore, both the tables depict the ranking obtained by Friedman's test in which SBBO has been ranked first in both the benchmark functions as it has the minimum mean ranking value. However, IBBO stands second for standard benchmark functions but, it fails for CEC 2017 problems. Thus, it is validated that SBBO is significantly different and comparatively superior to the other considered meta-heuristic algorithms for both the standard and CEC 2017 benchmark problems and IBBO only performs better for standard benchmark functions. Therefore, it can be stated that IBBO shows limitations for shifted and rotated problems like, CEC 2017 benchmark functions and does not work well for constrained optimization functions and real-world problems.

The convergence behavior of an evolutionary algorithm is also a major component to be analyzed. Therefore, the convergence trends of the IBBO and SBBO have also been examined and compared with existing algorithms by plotting the convergence graph for standard benchmark problems and CEC 2017 benchmark problems. Figure 4.5 & 4.6 depict the convergence trends

Table 4.3: Mean fitness values returned by IBBO, SBBO and other methods on the standard benchmark functions (F_1 to F_{10})

Func.	Dims.	BBO	BBO-M	GWO	ckGSA	LXBBO	WOA	LSHADE	SSA	IBBO	SBBO
F1	30	1.38e-01	1.260	1.440	1.300	4.71e-01	4.24e-01	1.98e-01	1.78e-01	0	0
	60	4.96e-01	1.930	3.110	2.800	8.26e-01	7.43e-01	5.71e-01	3.45e-03	2.42e-03	
	90	8.84e-01	2.490	3.780	3.400	1.070	9.63e-01	1.090	9.81e-01	8.63e-02	6.04e-02
F2	30	5.02e-04	2.69e-03	8.50e-01	7.65e-01	9.76e-03	8.78e-03	4.99e-04	4.49e-04	0	0
	60	5.10e-01	3.04e-01	3.12e-01	2.81e-01	1.820	1.640	1.140	1.030	0	0
	90	3.840	8.49e-01	5.81e-01	5.23e-01	5.260	4.730	7.480	6.730	0	0
F3	30	0	5.60e-14	0	0	5.63e-03	5.07e-03	0	0	0	0
	60	4.84e-02	1.46e-03	0	0	4.050	3.650	1.77e-01	1.59e-01	0	0
	90	1.880	9.77e-01	7.79e-02	7.01e-02	2.26e01	2.03e01	4.490	4.040	0	0
F4	30	2.45e-03	1.00e-10	0	0	8.00e-03	7.20e-03	6.08e-04	5.47e-04	0	0
	60	9.95e-01	1.660	1.60e01	1.44e01	1.020	9.18e-01	1.710	1.540	0	0
	90	4.980	5.43e-01	5.94e01	5.35e01	2.750	2.480	9.460	8.510	0	0
F5	30	1.25e01	5.40e-02	2.12e02	1.91e02	9.95e02	8.96e02	1.77e01	1.59e01	3.540	2.480
	60	2.09e02	1.50e-04	8.81e03	7.93e03	4.55e03	4.10e03	5.51e02	4.96e02	4.14e01	2.90e01
	90	8.85e02	2.54e-04	1.72e04	1.55e04	8.73e03	7.86e03	2.54e03	2.29e03	1.57e02	1.10e02
F6	30	2.670	2.980	4.410	3.970	5.050	4.550	2.880	2.590	2.760	1.930
	60	7.560	6.860	1.33e01	1.20e01	1.29e01	1.16e01	8.500	7.650	6.830	4.780
	90	1.45e01	2.05e01	2.34e01	2.11e01	2.19e01	1.97e01	1.70e01	1.53e01	1.12e01	7.840
F7	30	1.90e-03	3.33e-10	1.150	1.040	2.33e-02	2.10e-02	1.10e-02	9.90e-03	0	0
	60	8.820	3.70e01	1.11e01	9.990	2.840	2.560	1.50e01	1.35e01	1.05e-02	7.35e-03
	90	6.07e01	3.71e06	4.46e01	4.01e01	9.240	8.320	3.51e03	3.16e03	1.17e-01	8.19e-02
F8	30	5.51e-01	2.42e-09	2.920	2.630	1.650	1.490	9.54e-01	8.59e-01	0	0
	60	6.85e01	9.92e01	1.46e02	1.31e02	1.38e01	1.24e01	2.79e03	2.51e03	1.77e-01	1.24e-01
	90	6.94e04	8.24e06	1.49e06	1.34e06	4.46e01	4.01e01	4.27e05	3.93e05	1.390	9.73e-01
F9	30	7.17e-01	2.17e01	5.420	4.880	8.40e-02	7.56e-02	1.300	1.170	4.77e-01	3.34e-01
	60	4.31e01	5.13e03	5.97e01	5.37e01	1.45e01	1.31e01	1.24e02	1.12e02	5.100	3.570
	90	2.39e02	1.79e04	2.64e02	2.38e02	4.09e01	3.68e01	4.61e02	4.15e02	1.75e01	1.23e01
F10	30	8.40e-03	8.45e-06	0	2.140	1.930	3.15e-03	3.00e-03	0	0	0
	60	3.63e01	4.080	1.43e02	1.29e02	7.14e01	6.43e01	8.46e01	8.46e01	0	0
	90	5.97e02	3.74e03	1.02e04	9.18e03	3.26e02	2.93e02	1.55e03	1.40e03	0	0

Table 4.4: Mean fitness values returned by IBBO, SBBO and other methods on the standard benchmark functions (F_{11} to F_{20})

Func.	Dims.	BBO	BBO-M	GWO	ckGSA	LXBBO	WOA	LSHADE	SSA	IBBO	SBBO
F11	30	0	4.44e-25	0	0	4.54e-57	4.09e-57	0	0	0	0
	60	9.22e-38	1.78e-07	0	0	1.13e-24	1.02e-24	4.97e-29	4.47e-29	0	0
	90	4.33e-12	3.55e-02	0	0	2.28e-21	2.05e-21	6.66e-15	5.99e-15	0	0
F12	30	0	5.40e-27	0	0	4.56e-15	4.10e-15	0	0	0	0
	60	4.81e-04	3.36e-04	0	0	5.68e-05	5.11e-05	5.63e-04	5.07e-04	0	0
	90	9.53e-02	1.510	0	0	2.03e-03	1.83e-03	1.70e-01	1.53e-01	0	0
F13	30	0	1.26e02	4.530	4.080	1.220	1.100	0	0	0	0
	60	1.600	4.48e02	7.52e01	6.77e01	1.75e01	1.58e01	4.920	4.430	0	0
	90	2.30e01	8.33e02	2.28e02	2.05e02	5.40e01	4.86e01	3.13e01	2.82e01	0	0
F14	30	4.47e01	2.23e-37	1.11e03	9.99e02	1.48e02	1.33e02	9.18e01	8.26e01	6.30e-01	4.41e-01
	60	3.21e03	1.06e01	6.46e04	5.81e04	2.25e03	2.03e03	6.81e03	6.13e03	6.71e01	4.70e01
	90	2.80e04	9.34e03	3.23e05	2.91e05	8.85e03	7.97e03	5.44e04	4.90e04	7.42e02	5.19e02
F15	30	1.20e02	1.70e03	5.54e02	4.99e02	9.25e01	8.33e01	2.21e01	1.99e01	5.090	3.560
	60	2.10e02	8.05e03	3.96e03	3.56e03	4.16e03	3.74e03	5.11e02	4.60e02	3.05e01	2.14e01
	90	1.18e03	1.73e04	7.74e03	6.97e03	9.43e03	8.49e03	2.42e03	2.18e03	1.41e02	9.87e01
F16	30	0	2.46e-06	2.11e01	1.90e01	7.44e-02	6.70e-02	5.28e-03	4.75e-03	0	0
	60	2.000	7.31e-01	9.94e01	8.95e01	5.460	4.910	4.070	3.660	0	0
	90	1.08e01	3.00e01	2.22e02	2.00e02	1.46e01	1.31e01	1.98e01	1.78e01	0	0
F17	30	9.62e-04	6.11e-09	1.820	1.640	1.510	1.360	1.40e-01	1.26e-01	0	0
	60	1.93e-01	3.34e-03	2.18e01	1.96e01	2.49e-01	2.24e-01	7.73e-01	6.96e-01	0	0
	90	2.960	2.060	7.60e01	6.84e01	1.060	9.54e01	7.770	6.990	0	0
F18	30	3.00e-01	0	1.13e02	1.02e02	1.900	1.710	2.400	2.160	2.00e-01	1.40e-01
	60	2.87e01	0	5.10e02	4.59e02	4.25e01	3.83e01	5.78e01	5.20e01	6.700	4.690
	90	1.28e02	1.01e02	1.05e03	9.45e02	1.06e02	9.54e01	2.56e02	2.30e02	2.34e01	1.64e01
F19	30	9.63e-03	6.33e-10	1.820	1.640	1.510	1.360	1.40e-01	1.26e-01	0	0
	60	5.12e01	2.36e-01	6.76e02	6.08e02	4.16e01	3.74e01	1.42e02	1.28e02	0	0
	90	6.65e02	2.04e02	4.51e03	4.06e03	1.90e02	1.71e02	1.42e03	1.28e03	0	0
F20	30	2.430	9.02e-04	3.01e01	2.71e01	1.47e01	1.32e01	2.350	2.120	0	0
	60	1.76e02	3.30e-01	4.170	3.750	1.70e03	1.53e03	3.15e02	2.84e02	0	0
	90	1.30e03	1.29e01	2.16e01	1.94e01	6.23e03	5.61e03	1.88e03	1.69e03	0	0

Table 4.5: Mean fitness values returned by IBBO, SBBO, and other methods on the CEC 2017 problems

ID	BBO	BBO-M	GWO	cKGSA	LX-BBO	SSA	WOA	LSHADE	IBBO	SBBO
C_1	6.20e09	5.04e06	4.88e09	2.05e10	5.51e07	1.02e10	6.32e08	1.14e11	7.82e10	1.61e07
C_3	1.38e05	9.33e05	8.61e04	1.81e05	2.02e05	2.42e05	1.77e05	3.12e05	4.62e05	3.94e04
C_4	1.19e03	5.52e02	8.57e02	4.18e03	1.05e03	4.28e03	9.44e02	1.68e04	1.68e04	5.1e02
C_5	8.20e02	8.26e02	6.97e02	8.39e02	3.95e02	6.83e02	9.57e02	1.39e03	1.15e03	5.77e02
C_6	6.13e02	1.62e03	6.12e02	6.65e02	2.60e03	5.56e03	6.82e02	7.69e02	6.69e02	6.00e02
C_7	1.25e03	3.17e03	1.10e03	1.37e03	2.19e03	3.18e03	1.73e03	8.55e03	2.76e03	8.21e02
C_8	1.14e03	3.01e03	1.00e03	1.16e03	1.13e03	6.11e03	1.26e03	6.17e03	1.45e03	8.75e02
C_9	4.72e03	3.74e03	5.69e03	1.20e04	1.52e03	2.10e04	2.71e04	8.06e04	3.50e04	1.78e03
C_{10}	9.94e03	5.58e03	7.39e03	8.20e03	3.02e04	4.85e03	1.10e04	5.17e04	1.36e04	4.67e03
C_{11}	3.47e03	1.20e03	3.28e03	1.95e04	4.29e03	3.01e04	2.24e03	9.29e04	4.51e04	1.73e03
C_{12}	2.86e08	1.21e07	4.68e08	5.30e09	1.29e08	2.64e09	5.57e08	9.11e10	2.27e10	1.97e06
C_{13}	7.73e05	1.65e05	1.63e08	3.75e07	6.69e05	5.97e07	6.57e06	2.33e09	9.93e09	9.71e07
C_{14}	1.91e06	7.19e05	7.42e05	7.11e06	3.08e05	4.55e06	2.47e06	8.78e06	3.78e07	2.36e05
C_{15}	8.01e04	5.19e04	3.31e07	1.89e07	1.33e04	5.83e07	6.26e05	2.08e08	3.40e09	2.42e04
C_{16}	3.53e03	5.37e03	2.91e03	4.68e03	8.33e03	5.42e03	5.48e03	2.91e03	6.60e03	2.81e03
C_{17}	2.89e03	2.34e03	2.79e03	3.80e03	9.25e03	6.37e03	4.25e03	4.72e03	1.15e04	2.24e03
C_{18}	6.35e06	3.61e06	3.78e06	5.94e06	7.31e06	4.80e06	1.79e07	5.49e07	1.60e08	1.54e06
C_{19}	6.48e04	2.65e04	4.85e06	2.89e05	1.23e04	2.14e05	6.55e06	4.23e08	1.18e09	1.94e04
C_{20}	3.08e03	4.31e03	2.84e03	3.74e03	8.28e03	6.37e03	3.71e03	9.31e03	4.58e03	2.61e03
C_{21}	2.63e03	5.25e03	2.49e03	2.87e03	6.23e03	2.87e03	2.95e03	8.03e03	2.95e03	2.39e03
C_{22}	1.25e04	6.49e03	8.88e03	1.21e04	1.20e04	9.19e04	1.25e04	5.10e04	1.53e04	4.66e03
C_{23}	3.10e03	5.03e03	2.93e03	4.89e03	4.12e03	8.82e03	3.63e03	9.36e03	3.61e03	2.79e03
C_{24}	3.30e03	5.21e03	3.13e03	4.59e03	4.33e03	6.44e03	3.70e03	4.38e03	3.79e03	3.00e03
C_{25}	3.53e03	4.12e03	3.42e03	4.78e03	6.60e03	9.91e03	3.39e03	6.38e04	1.42e04	2.89e03
C_{26}	7.49e03	8.65e03	6.00e03	1.24e04	9.70e03	2.17e04	1.34e04	6.99e03	1.30e04	5.00e03
C_{27}	3.56e03	3.27e03	3.55e03	8.07e03	3.61e03	8.35e03	4.42e03	3.60e03	4.81e03	3.26e03
C_{28}	3.82e03	4.48e03	9.32e03	5.94e03	4.91e03	6.07e03	4.04e03	3.89e03	1.06e04	3.38e03
C_{29}	4.28e03	5.37e03	5.51e03	1.18e04	4.35e03	1.19e04	7.91e03	3.92e03	5.05e04	3.93e03

of all the algorithms for 200, 400, 600, 800, and 1000 iterations on representative functions. The y-axes of all the convergence graphs shows the corresponding best fitness values and are represented in a logarithmic scales for standard benchmark problems and on linear scale for CEC 2017 benchmark problems. From the figures, it can be visualized that IBBO performs good for standard benchmark problems but it fails to show good behavior on CEC 2017 problems. However, SBBO has better convergence rate for almost all the unimodal, multi-modal, hybrid, and composite functions as compared to existing algorithms. The exploration capability of SBBO is very good among all the other considered state-of-the-art methods as it converges gradually and finally optimum solution is achieved. Moreover, the convergence trend towards optimum is relatively more consistent. From all the presented results, it is elicited that the new SBBO is robust against local optimum and has attained a good balance between global and local search.

Table 4.6: Mean ranking of considered algorithms on standard benchmark problems with 30, 60, and 90 dimensions using Friedman Test

Rank	30 dims.		60 dims.		90 dims.			
	Algorithm	Mean Rank	Algorithm	Mean Rank	Algorithm	Mean Rank		
1	SBBO	3.3272	SBBO	2.6618	SBBO	2.1294		
2	IBBO	4.5816	IBBO	3.6653	IBBO	2.9322		
3	BBO	5.2184	BBO	4.1747	BBO	3.3398		
4	CKGSA	6.0000	CKGSA	4.8000	CKGSA	3.8400		
5	GWO	6.0728	GWO	4.8582	GWO	3.8866		
6	WOA	6.1456	WOA	4.9165	WOA	3.9332		
7	SSA	6.1640	BBO-M	4.9312	BBO-M	3.9450		
8	BBO-M	6.2728	SSA	5.0182	SSA	4.0146		
9	LSHADE	6.5272	LSHADE	5.2218	LSHADE	4.1774		
10	LXBBO	7.5640	LXBBO	6.0512	LXBBO	4.8410		
		p-value= 7.45e - 16			p-value= 8.58e - 16			p-value= 4.56e - 10

Table 4.7: Mean ranking of considered algorithms on CEC 2017 benchmark problems using Friedman Test

Rank	Algorithm	Mean Rank Value
1	SBBO	1.93
2	BBO-M	3.70
3	GWO	3.93
4	LX-BBO	4.70
5	BBO	4.12
6	cKGSA	6.97
7	WOA	6.97
8	SSA	7.92
9	LSHADE	8.57
10	IBBO	9.52

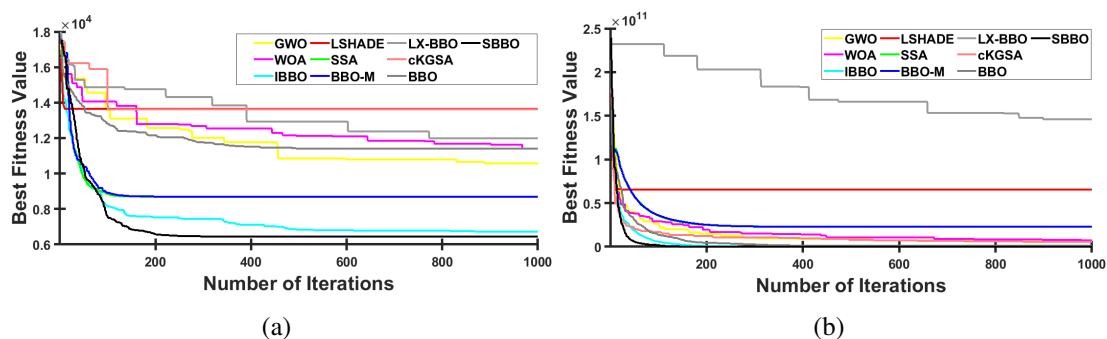


Figure 4.5: Convergence trend of SBBO with other considered Meta-heuristics on standard benchmark problems, (a) Rastrigin (F_{13}) and (b) New Schwefel (F_5)

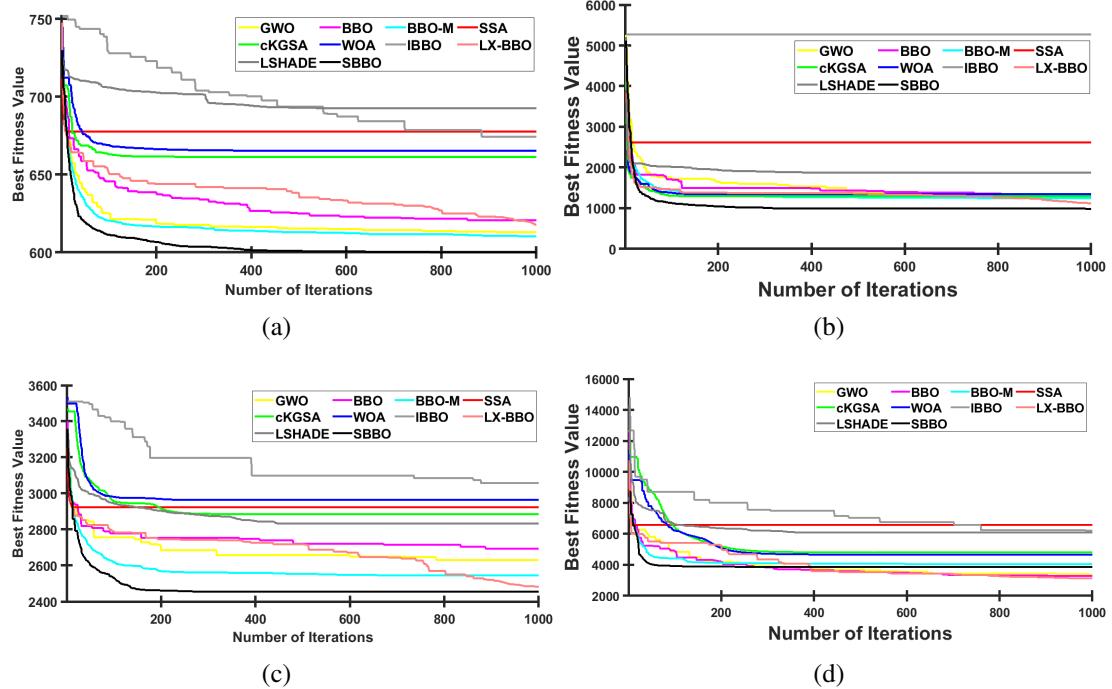


Figure 4.6: Convergence trend of SBBO with other considered Meta-heuristics on CEC 2017 benchmark problems, (a) SR Expanded Scaffer's Function (C_6), (b) HF 6 ($N=6$) (C_7), (c) CF 1 ($N = 3$) (C_{21}), and (d) CF 5 ($N = 5$) (C_{25})

4.4.2 Performance Analysis of SBBO based Histopathological Image Classification

Since SBBO performs better than IBBO on various benchmark functions, it is used in the BOF to find the optimal visual words. The performance of SBBO based BOF method to classify histopathological images has been tested on two histopathological image datasets, namely the Blue histology dataset and ADL dataset as discussed in Section 3.4.1. Both the datasets are partitioned into two sets, namely training and validation sets using stratified random sampling.

4.4.2.1 Result Analysis on Blue Histology Dataset

The SBBO based BOF method is compared with two state-of-the-art methods, namely IB3 [154] and IKS2 [156] with original BOF [29] and BBO-M-BOF based classification methods. Further, IKS2 is used to find out the representative keypoints from the image set using iterative keypoint selection (IKS) method. The keypoints are eliminated if their distances from the representative keypoints are less than a predefined distance threshold. The distance threshold for IKS2 is taken as 0.57. Moreover, the value of K for IKS2 is 3 as given in Lin et al. [156].

Figure 4.7 shows the confusion matrix for all the considered methods over blue histology

BOF		Predicted			
Actual		CT	ET	MT	NT
	CT	0.75	0.1	0.15	0
	ET	0.45	0.35	0.2	0
	MT	0.2	0.05	0.75	0
	NT	0.6	0.15	0.2	0.05

IB3		Predicted			
Actual		CT	ET	MT	NT
	CT	0.35	0.2	0.3	0.15
	ET	0.04	0.6	0.2	0.16
	MT	0.35	0.1	0.3	0.25
	NT	0.35	0.25	0.05	0.35

IKS2		Predicted			
Actual		CT	ET	MT	NT
	CT	0.85	0.05	0.1	0
	ET	0.36	0.35	0.2	0.04
	MT	0.15	0.05	0.75	0.05
	NT	0.6	0.15	0.2	0.05

BBO-M-BOF		Predicted			
Actual		CT	ET	MT	NT
	CT	0.86	0.04	0.08	0.02
	ET	0.11	0.64	0.16	0.09
	MT	0.05	0.06	0.78	0.11
	NT	0.45	0.15	0.15	0.25

SBBO-BOF		Predicted			
Actual		CT	ET	MT	NT
	CT	0.69	0.12	0.09	0.09
	ET	0.03	0.9	0	0.07
	MT	0	0.19	0.81	0
	NT	0.161	0.29	0.07	0.49

Figure 4.7: The normalized confusion matrix for the blue histology tissue dataset for the five methods BOF, IB3, IKS2, BBO-M-BOF, and SBBO-BOF.

dataset. In the figure, CT, ET, MT, and NT represent connective tissue, epithelium tissue, muscular tissue, and nervous tissue images respectively. The predicted class is given along the x-axis. The diagonal entries in the confusion matrices indicate the ratio of correct classification. The off-diagonal entries give the ratio of misclassification. From the figure, it can be visualized that the connective tissue images are correctly identified by BOF, IB3, IKS2, BBO-M-BOF, and SBBO-BOF based methods with an accuracy of 75%, 35%, 85%, 86%, and 69% respectively. Although BOF, IKS2, and BBO-M-BOF show better performance in the case of connective tissue images, they show degraded performance to classify the rest of the image categories. From the figure, it can be illustrated that the SBBO-BOF method identifies the rest of the image classes with the accuracy of 90%, 81%, and 49% which are the best among other considered classification methods. This leads to an overall accuracy of SBBO-BOF based method to 72.23% while BOF, IB3, IKS2, and BBO-M-BOF methods give only 47.5%, 40%, 50.63%, and 63.25% accuracy. Therefore, based on the results, it can be stated the SBBO-BOF method outperforms.

Moreover, the comparative analysis of various performance matrices like recall, precision, F1-measure, and specificity are depicted in Table 4.8. The SBBO-BOF method shows better results for almost all the considered parameters. For muscle tissue class, it shows the best F1-measure equals to 0.82 followed by connective, epithelial, and nervous tissues respectively.

Table 4.8: Comparative analysis of the SBBO-BOF based classification method

Category	Parameters	BOF	IB3	IKS2	BBO-M-BOF	SBBO-BOF
Muscle Tissue	Recall	0.750	0.300	0.750	0.750	0.810
	Precision	0.577	0.353	0.600	0.667	0.835
	F-measure	0.652	0.324	0.667	0.719	0.822
	Specificity	0.817	0.817	0.831	0.870	0.947
Connective Tissue	Recall	0.750	0.350	0.850	0.850	0.697
	Precision	0.375	0.321	0.434	0.585	0.783
	F-measure	0.500	0.335	0.574	0.696	0.738
	Specificity	0.583	0.753	0.624	0.797	0.937
Epithelial Tissue	Recall	0.350	0.600	0.368	0.684	0.900
	Precision	0.538	0.522	0.583	0.719	0.600
	F-measure	0.424	0.558	0.452	0.677	0.720
	Specificity	0.900	0.817	0.917	0.917	0.800
Nervous Tissue	Recall	0.050	0.350	0.050	0.200	0.485
	Precision	1.000	0.385	0.357	0.532	0.754
	F-measure	0.095	0.366	0.088	0.340	0.590
	Specificity	1.000	0.813	0.969	0.927	0.946
	Overall Accuracy	47.5	40	50.63	63.25	72.23

The overall accuracy of the new classification method increases by more than 10% as compared to BBO-M-BOF based classification method which has the next highest accuracy. The results validate that the new SBBO-BOF classification method outperforms the other methods.

Furthermore, a comparison of the F1 score, G-mean, sensitivity, and specificity is represented using the radar chart as shown in Figure 4.8. The radar chart shows that the method with the largest area and symmetrical shape performs better than others. From the figure, it is visualized that for muscle tissue, the area covered by the SBBO-BOF method is larger than other methods on all four considered aspects while for other tissue classes, the covered area is larger as compared with the other methods. These results validate the effectiveness of the SBBO-BOF method in the image classification problem for the Blue histology dataset.

4.4.2.2 Result Analysis on ADL dataset

The classification efficiency of the new method has been observed for all the three different types of organ images i.e., Kidney, Lung, and Spleen in the ADL dataset. The results of the SBBO-BOF method have been compared with three state-of-the-art methods, namely SVM [205], SRC [169], and SHIRC which are already analyzed in the literature on ADL dataset by Srinivas et al. [2]. SVM is used for decision making in the classification process. For feature extraction, the state-of-the-art method, namely WND-CHARM [13] is used. Further, SRC is the classification method with a single channel sparse representation of RGB images. Moreover, SHIRC is the extension of the SRC approach and is known as a simultaneous sparsity

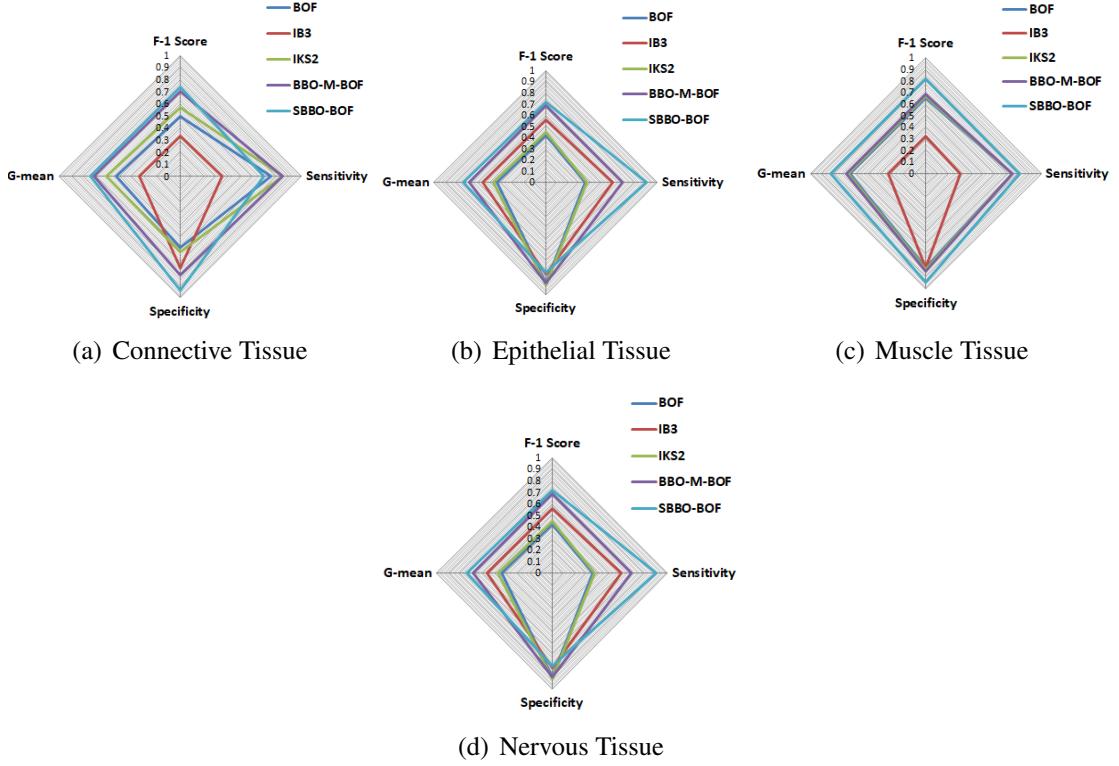


Figure 4.8: Radar chart for average results obtained by SVM classifier on Blue histology dataset.

Table 4.9: The confusion matrix for kidney image classification obtained by the SBBO-BOF method

Class	Healthy	Inflammatory	Method
Healthy	0.6925	0.3075	SVM
	0.875	0.125	SRC
	0.825	0.175	SHIRC
	0.89	0.11	SBBO-BOF
Inflammatory	0.2812	0.7188	SVM
	0.25	0.75	SRC
	0.1667	0.8333	SHIRC
	0.175	0.875	SBBO-BOF

Table 4.10: The confusion matrix for lung image classification obtained by the SBBO-BOF method

Class	Healthy	Inflammatory	Method
Healthy	0.8875	0.1125	SVM
	0.725	0.275	SRC
	0.75	0.25	SHIRC
	0.92	0.08	SBBO-BOF
Inflammatory	0.372	0.6238	SVM
	0.2417	0.7853	SRC
	0.15	0.85	SHIRC
	0.15	0.85	SBBO-BOF

model for multi-channel histopathological images which designs three color dictionaries for RGB channels.

Tables 4.9 to 4.11 show the confusion matrices for all the considered methods over three datasets as mentioned above. From the tables, it can be observed that the SBBO-BOF method identifies the healthy kidney, lung, and spleen images with an accuracy of 89%, 92%, and 82.8% respectively. Likewise, the accuracy of 87.5%, 85%, and 90% is observed for inflamed kidney, lung, and spleen images. All the accuracy is better than considered methods. Moreover, the overall average accuracy of the SBBO-BOF method is 87% followed by SHIRC method with an average accuracy of 79.86%. The average accuracy of all other considered methods is

Table 4.11: The confusion matrix for spleen image classification obtained by the SBBO-BOF method

Class	Healthy	Inflammatory	Method
Healthy	0.5112	0.488	SVM
	0.7083	0.2917	SRC
	0.65	0.35	SHIRC
	0.828	0.172	SBBO-BOF
Inflammatory	0.1275	0.8725	SVM
	0.2083	0.7917	SRC
	0.1167	0.8833	SHIRC
	0.1	0.9	SBBO-BOF

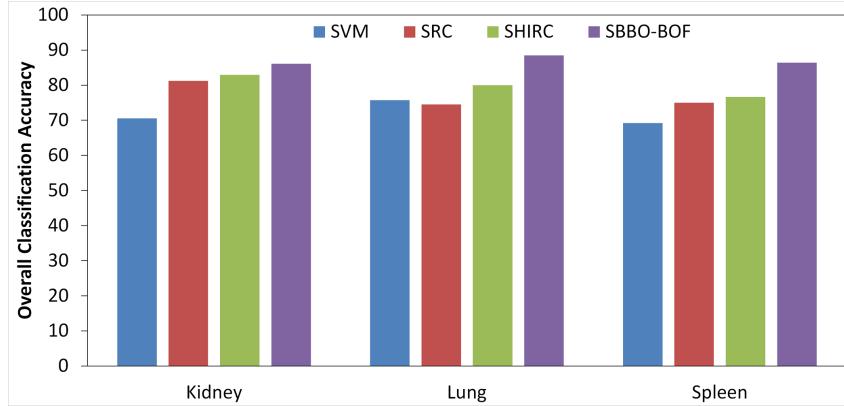


Figure 4.9: The overall classification accuracy of the considered methods for ADL dataset in form of bar chart

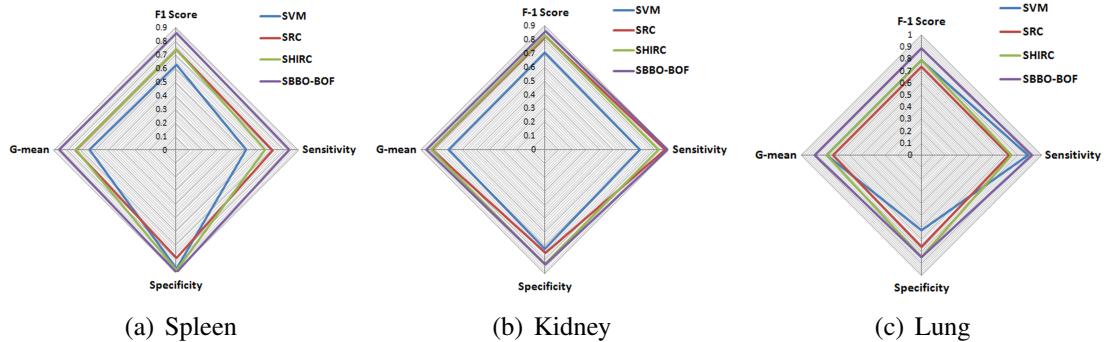


Figure 4.10: Radar chart for average results on (a) Spleen dataset, (b) Kidney dataset, and (c) Lung dataset

depicted in Figure 4.9. Furthermore, a comparison of the F1 score, G-mean, sensitivity, and specificity is also represented using radar charts as shown in Figure 4.10. These figures also show the effectiveness of the SBBO-BOF method over other methods.

4.4.2.3 Computational Complexity of SBBO based BOF Method

The computational cost of the SBBO-BOF method has also been measured asymptotically and described below:

- The computational cost of K-means is $O(N^2)$ [206], where N refers to the descriptor

count.

- For the calculation of fitness value, the sum of squared Euclidean distances between the cluster centers and the corresponding descriptors are considered. Hence, the computational cost of the fitness calculation becomes $O(N \times C)$, where C represents the cluster count.
- Since, the SBBO only enhances the mutation operator of BBO, therefore its computational cost will be similar to BBO which is $O(P^2)$ [99], where P represents the size of the population.

Therefore, the overall computational complexity of the new SBBO-BOF method can be given as $O(N^2 + N \times C + P^2)$. As the descriptor count (N) for the histopathological images is much larger than the population size (P), therefore, the worst case time complexity of the SBBO-BOF method is $O(N^2)$.

4.5 Summary

This chapter presents two new variants of BBO, namely IBBO and SBBO and analyze their performance on standard and CEC 2017 benchmark problems. IBBO performs well for standard benchmark problems but it fails to optimize the CEC 2017 benchmark problems. However, SBBO outperforms the other considered algorithms on both types of benchmark problems. Therefore, SBBO is further applied to find the optimal visual words in the BOF method and the modified SBBO based BOF method (SBBO-BOF) has been tested and validated on two histopathological image datasets for the classification of tissue images. The SBBO-BOF method outperforms other image classification methods in terms of average accuracy, recall, precision, and F1-measure. The results depict that the SBBO-BOF method attains a high average accuracy of 72.23% and 87% for blue histology dataset and ADL dataset respectively. However, the SBBO-BOF codebook construction method is computationally expensive. Therefore, in the next chapter, a computationally efficient grey relational analysis based codebook construction method is introduced.

CHAPTER 5

EFFICIENT CODEBOOK CONSTRUCTION

5.1 Introduction

The codebook construction method using SBBO as discussed in Chapter 4 shows significant performance for the classification of histopathological images using the BOF method. However, these methods are computationally expensive. Moreover, it is observed that the GRA based keypoint selection method produces significant features due to effective GRG similarity measure and is computationally efficient as discussed in Chapter 3. Therefore, in this chapter, a new Grey relational analysis based BOF method (GRA-BOF) is introduced which improves the efficiency of the codebook construction step of the standard BOF method. The method uses a Grey relational analysis for similarity measure in the clustering of the feature descriptors.

5.2 Grey Relational Analysis based Bag-of-Features

As discussed in Chapter 4, K-means is not an efficient clustering method when applied on histopathological images having a large number of feature vectors. However, meta-heuristics based clustering methods show effective results but, they are not computationally efficient. Therefore, there is a need for a computationally efficient codebook construction method.

The computational cost of K-means depends upon the distance calculations using Euclidean similarity measure. The size of the distance matrix is generally $O(n \times K)$, where, n denotes feature count and K denotes the number of visual words. For large n , the Euclidean similarity becomes computationally expensive in terms of time and space. In literature, it has been proved that Grey relation analysis based similarity measure is computationally efficient than Euclidean similarity [160]. Therefore, in this chapter, a new clustering algorithm has been introduced to find the most representative and relevant visual words in the BOF method. The overall flow

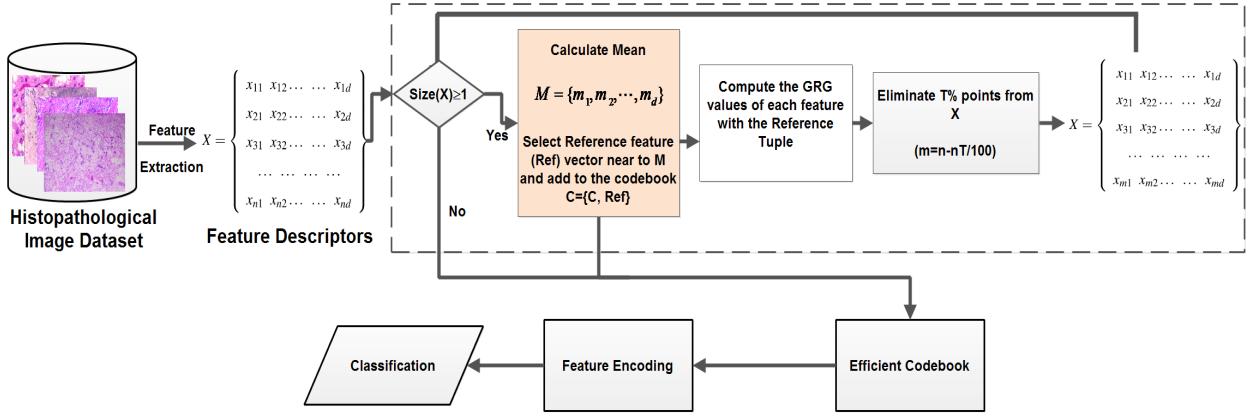


Figure 5.1: The GRA based BOF method

of the new method is depicted in Figure 5.1 and a detailed description of the same is provided below.

1. Generate random training and validation sets from the histopathological image dataset.
2. Extract the features from training images to generate a set of feature descriptors (X) having n features of d dimensions as shown in Eq. (5.1)

$$X = \left\{ \begin{array}{cccc} x_{11} & x_{12} & \dots & x_{1d} \\ x_{21} & x_{22} & \dots & x_{2d} \\ x_{31} & x_{32} & \dots & x_{3d} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nd} \end{array} \right\} \quad (5.1)$$

3. Calculate the mean of X and find the feature vector which is nearby to the mean vector. Select this as a reference feature vector (Ref) which is considered as the first visual word and add this to the final codebook (C).
4. All other feature vectors in X are considered as the comparative sequences and similarity between all the comparative sequences with the reference sequence is measured using grey relational analysis. Grey relational similarity measure is a two-step process which is as follows:
 - To measure the closeness between feature sequences a grey relational coefficient

matrix (γ) is computed by using Eq. (5.2).

$$\gamma(Ref(d), X_c(d)) = \frac{\min \Delta(d) + \xi \max \Delta(d)}{\Delta(d) + \xi \max \Delta(d)}, \quad (5.2)$$

where, $\Delta(d) = |Ref(d) - X_c(d)|$, for $c = 1, \dots, n$ ($c \neq id(Ref)$), $d = 1, \dots, D$ and $\xi \in (0, 1]$. ξ is the random variable to control the variation between $\min \Delta(d)$ and $\max \Delta(d)$.

- The GRC values are used to compute Grey relational grades (GRGs) which is given by Eq. (5.3)

$$\Gamma(Ref, X_c) = \sum_{d=1}^D [\alpha(d) \cdot \gamma(Ref(d), X_c(d))] \quad (5.3)$$

where, $\alpha(d)$ is the weighting factor which generally chooses $\alpha_i(d) = 1/p$ for all D . The values of GRC and GRG must always be between 0 and 1. The resulting GRG vector is given by Eq. (5.4).

$$GRG = \left\{ \begin{array}{c} g_1 \\ g_2 \\ g_3 \\ \dots \\ g_c \end{array} \right\} \quad (5.4)$$

where, $c = n - 1$, and $GRG \in (0, 1)$. Larger values of GRGs refer to the high similarity of the comparative sequence with the reference sequence.

5. Eliminate $T\%$ comparative feature vectors having GRG values close to 1 because these features are more similar to the reference feature vector and are not relevant for the code-book construction.
6. Update the set X using Eq. (5.6) which only contains m feature vectors given by Eq. (5.5)

$$m = n - \frac{n \times T}{100} \quad (5.5)$$

$$X = \left\{ \begin{array}{cccccc} x_{11} & x_{12} & \dots & \dots & x_{1d} \\ x_{21} & x_{22} & \dots & \dots & x_{2d} \\ x_{31} & x_{32} & \dots & \dots & x_{3d} \\ \dots & \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & \dots & x_{md} \end{array} \right\} \quad (5.6)$$

7. Repeat Step 3 to 6 until only one feature is left in X .
8. The updated set C in step 3 is considered as the cluster centers or visual words.
9. Represent each image in the training set in the histogram of these visual words.
10. Give the histogram along with the corresponding annotations to the SVM classifier for training.
11. Pick any image from the validation set and represent it into the histogram in the same way as discussed above and feed it to SVM for predicting its label.

The process of clustering algorithm using GRA is also given in Algorithm 5.1.

Algorithm 5.1 The GRA based clustering method

Input: A set of feature vectors, known as descriptor D , having n strong feature vectors and a cut-off threshold (T)

Output: Reduced descriptor S having m feature vectors ($m < n$)

while ($\text{size}(D) > 1$) **do**

 Select a reference vector (R) near to the mean of D ;

$D = D - R$

 Calculate GRGs for each feature vector in D with R (Eq. (5.3));

 Sort the D according the GRG values in descending order and delete the first $T\%$ feature vectors from D ;

 Update vector $S = [S \ R]$;

end while

The resulting set S is considered as the cluster centers or visual words.

5.3 Experimental Results

The efficacy of the GRA-BOF method against other considered state-of-the-art methods, it has been tested on ADL and Blue histology histopathological image datasets. For comparative analysis, two BOF based classification methods, standard BOF and IKS2-BOF and three

state-of-the-art classification methods, SVM, SRC, and SHIRC, have been considered. A brief description of these methods is provided below.

- *IKS2-BOF*: Iterative keypoints selection methods [156] reduce the complexity of vector quantization in the BOF method by selecting some representative keypoints from the images. The representative keypoints can be generated either randomly (IKS1) or using K-means (IKS2). The performance of the IKS2-BOF classification method is better than the IKS1-BOF method on natural image datasets. Hence, in this chapter, the IKS2-BOF method is considered and evaluated on the histopathological image dataset for the comparative analysis against the new method. The parameter settings for the simulation is taken from the respective literature.
- *SVM*: SVM is one the state-of-the-art classification method discussed by Srinivas et. al [2]. It extracts the features using a well-known feature extraction method for histopathological images, namely WND-CHARM [13] and classification, support vector machine is used. The simulation results for this method on the ADL dataset is taken from Srinivas et al. [2].
- *SRC*: Sparse representation based classification methods are initially used for face recognition applications[207].After that, it is also widely used in medical images [2]. It is a single luminance channel based sparse representation of RGB images which is further used for classification. The simulation results for this method on the ADL dataset is taken from [2].
- *SHIRC*: SHIRC is the multi-channel sparsity model which is an extension of the standard SRC approach for three color channels. It is used for the classification of histopathological images. The simulation results for this method on the ADL dataset is taken from Srinivas et al. [2].

The main parameter in the GRA-BOF is the cut-off threshold (T) which is used to eliminate similar points. Figure 5.2 shows the classification accuracy obtained by the GRA-BOF method using SVM on ADL and Blue histology datasets. From the figure, it can be observed that for different values of cut-off thresholds there is a slight variation in the obtained accuracy over each dataset. From the results, the cut-off threshold is taken 2% for ADL histopathological dataset

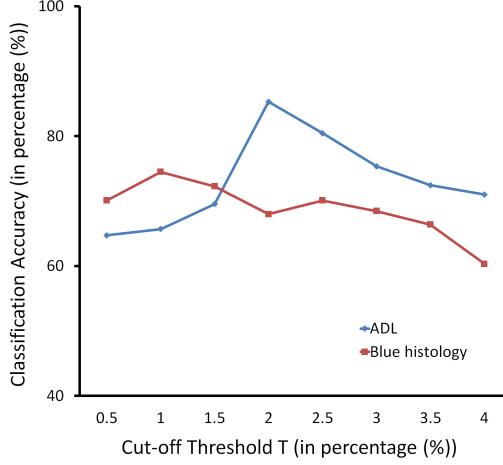


Figure 5.2: Impact analysis of cut-off threshold (T) value over classification accuracy on ADL and Blue histology datasets

while for blue histology dataset 1% cut-off threshold value is considered. Moreover, the new GRA-BOF method uses SVM for the classification with 10-fold cross-validation to overcome the issue of over-fitting. The selection of appropriate kernel function is very important, therefore χ^2 RBF kernel is used in this chapter as it shows better performance in Lin et. al [156]. For each trial, the images for training and validation sets are randomly selected from the dataset.

Table 5.1: Confusion matrix generated by considered methods on ADL histopathological image dataset

Class	Method	Kidney		Lung		Spleen	
		Normal	Inflamed	Normal	Inflamed	Normal	Inflamed
Normal	SVM	0.69	0.28	0.89	0.37	0.51	0.13
	SRC	0.88	0.25	0.73	0.24	0.71	0.21
	SHIRC	0.82	0.17	0.75	0.15	0.65	0.12
	BOF	0.82	0.31	0.72	0.25	0.55	0.23
	IKS2	0.85	0.23	0.83	0.30	0.68	0.13
	GRA-BOF	0.89	0.09	0.92	0.06	0.75	0.11
Inflamed	SVM	0.31	0.72	0.11	0.62	0.49	0.87
	SRC	0.13	0.75	0.27	0.78	0.29	0.79
	SHIRC	0.18	0.83	0.25	0.85	0.35	0.88
	BOF	0.18	0.69	0.28	0.75	0.45	0.77
	IKS2	0.15	0.77	0.17	0.7	0.32	0.87
	GRA-BOF	0.11	0.91	0.08	0.94	0.25	0.89

The statistics of the classification performance are presented by the confusion matrices returned by each method to classify each organ image. These confusion matrices are shown in Table 5.1 for kidney, lung, and spleen of ADL histopathological image datasets. The rows of the confusion matrices represent the predicted output of the classifier and columns refer to the actual class of the test images. The confusion matrices show the average results achieved from 10 trials. Figure 5.3 shows the confusion matrices returned by each method on blue histology image dataset. From Table 5.1 and Figure 5.3, it can be observed that SVM, trained directly on

SVM		Predicted			
		CT	ET	MT	NT
Actual	CT	0.45	0.1	0.4	0.05
	ET	0.15	0.4	0.35	0.1
	MT	0.35	0.1	0.3	0.25
	NT	0.4	0.3	0.1	0.2

SHIRC		Predicted			
		CT	ET	MT	NT
Actual	CT	0.55	0.1	0.05	0.3
	ET	0.15	0.7	0.15	0
	MT	0.1	0.05	0.7	0.1
	NT	0.3	0.05	0.1	0.55

IKS2		Predicted			
		CT	ET	MT	NT
Actual	CT	0.71	0.1	0.15	0.04
	ET	0.47	0.33	0.2	0
	MT	0.2	0.05	0.73	0.02
	NT	0.4	0.15	0.2	0.25

SRC		Predicted			
		CT	ET	MT	NT
Actual	CT	0.65	0.1	0.15	0.1
	ET	0.15	0.7	0.15	0.2
	MT	0.1	0.05	0.75	0.1
	NT	0.37	0.05	0.1	0.48

BOF		Predicted			
		CT	ET	MT	NT
Actual	CT	0.75	0.1	0.15	0
	ET	0.45	0.35	0.2	0
	MT	0.2	0.05	0.75	0
	NT	0.6	0.15	0.2	0.05

GRA-BOF		Predicted			
		CT	ET	MT	NT
Actual	CT	0.8	0.17	0.02	0.01
	ET	0.1	0.78	0.1	0.02
	MT	0.15	0.05	0.75	0.05
	NT	0.1	0.2	0.05	0.65

Figure 5.3: Confusion matrices returned by (a) SVM, (b) SRC, (c) SHIRC, (d)BOF, (e)IKS2 (f) GRA-BOF on Blue histology image dataset

features extracted by the WND-CHARM method, does not perform well for any of the classes. It means the relevant features are not used for training. The performance of other state-of-the-art methods is far better than this method. The performance of the GRA-BOF method is tremendous in identifying the images of all the classes more accurately. It identifies almost all the inflamed lung images from the ADL dataset while the worst performance is observed in the case of the normal spleen image class which is also the best as compared to other methods. While, in the case of blue histology dataset, the GRA-BOF method identifies all four types of tissues in a more efficient manner as compared to other considered methods.

In medical image classification, pathologists are more interested in true negatives (i.e., the accuracy of identifying the inflamed images correctly). It can be illustrated from Table 5.1 that the GRA-BOF method identifies the inflamed test images more accurately than other considered competitive methods with the accuracy of 91%, 94%, and 89% for kidney, lung, and spleen organs respectively. Moreover, it also maintains high classification accuracy to categorize normal test images. The GRA-BOF method classifies the inflamed tissue images of each organ much more accurately than normal tissue images. Hence, the probability of normal image wrongly identified as the inflammatory image (i.e., false alarm rate) is very less for the new method while

SRC, BOF, and IKS2 have high false alarm rates on kidney images. Moreover, SVM and IKS2 also show the high false alarm rates for lung organ images. The false negatives to detect the disease is comparatively higher in the case of the GRA-BOF (91.3%) method which is the most relevant metric for the pathologists. For the blue histology image dataset, it can be observed from the Figure 5.3 that the GRA-BOF method returns highest true positives for the identification of all the tissue types. However, the identification of nervous tissue image is a challenging task due to the variety of staining used in it. The GRA-BOF method identifies the nervous tissue test image with an accuracy of 65% while the identification rate of other considered methods for nervous tissue is below 50% except SHIRC. Moreover, for connective and epithelial tissues, the GRA-BOF method identifies 80% and 78% of test images respectively. Therefore, from the study of confusion matrices returned by all the classification methods, it can be stated that the GRA-BOF method shows consistent and the best results for all the considered organ images.

Table 5.2: Classification performance with other methods based on recall, specificity, precision, FPR, and accuracy on ADL dataset

Organ	Algorithms	Recall	Specificity	Precision	FPR	Accuracy
Kidney	SVM	0.69	0.72	0.71	0.28	0.71
	SRC	0.88	0.75	0.78	0.25	0.81
	SHIRC	0.83	0.83	0.83	0.17	0.83
	BOF	0.82	0.69	0.73	0.31	0.76
	IKS2-BOF	0.85	0.77	0.79	0.23	0.81
	GRA-BOF	0.89	0.91	0.91	0.09	0.90
Lung	SVM	0.89	0.63	0.70	0.37	0.76
	SRC	0.73	0.76	0.75	0.24	0.75
	SHIRC	0.75	0.85	0.83	0.15	0.80
	BOF	0.72	0.75	0.74	0.25	0.74
	IKS2-BOF	0.83	0.70	0.73	0.30	0.77
	GRA-BOF	0.92	0.94	0.94	0.06	0.93
Spleen	SVM	0.51	0.87	0.80	0.13	0.69
	SRC	0.71	0.79	0.77	0.21	0.75
	SHIRC	0.65	0.88	0.85	0.12	0.77
	BOF	0.55	0.77	0.71	0.23	0.66
	IKS2-BOF	0.68	0.87	0.84	0.13	0.78
	GRA-BOF	0.75	0.89	0.87	0.11	0.82

Moreover, the various performance parameters are used to analyze the efficacy of the GRA-BOF method such as recall, specificity, precision, false positive rate (FPR), and accuracy. Table 5.2 depicts the computed values of these parameters on the kidney, lung, and spleen datasets while Table 5.3 shows the results for blue histology dataset. From the tables, it can be stated that the GRA-BOF method outperforms other methods over recall, specificity, precision, recall, FPR, and accuracy. The classification accuracies of the GRA-BOF on kidney, lung, and spleen datasets are 90%, 93%, and 82% respectively which is higher than the other considered state-

Table 5.3: Classification performance with other methods based on recall, precision, F1 score, specificity, and average accuracy on Blue histology dataset

Category	Parameters	SVM	SRC	SHIRC	BOF	IKS2	GRA-BOF
CT	Recall	0.450	0.650	0.550	0.750	0.710	0.800
	Precision	0.512	0.512	0.500	0.375	0.399	0.696
	F1 Score	0.573	0.573	0.524	0.500	0.511	0.744
	Specificity	0.806	0.806	0.814	0.583	0.643	0.883
ET	Recall	0.400	0.583	0.700	0.350	0.330	0.780
	Precision	0.778	0.778	0.778	0.538	0.524	0.650
	F1 Score	0.667	0.667	0.737	0.424	0.405	0.709
	Specificity	0.933	0.933	0.932	0.900	0.900	0.860
MT	Recall	0.300	0.750	0.737	0.750	0.730	0.751
	Precision	0.652	0.652	0.700	0.577	0.570	0.815
	F1 Score	0.698	0.698	0.718	0.652	0.640	0.781
	Specificity	0.875	0.875	0.900	0.817	0.817	0.943
NT	Recall	0.200	0.480	0.550	0.050	0.250	0.650
	Precision	0.545	0.545	0.579	1.000	0.806	0.890
	F1 Score	0.511	0.511	0.564	0.095	0.382	0.751
	Specificity	0.875	0.875	0.864	1.000	0.980	0.973
	Average Accuracy	33.8	61.4	63.3	47.5	50.5	74.5

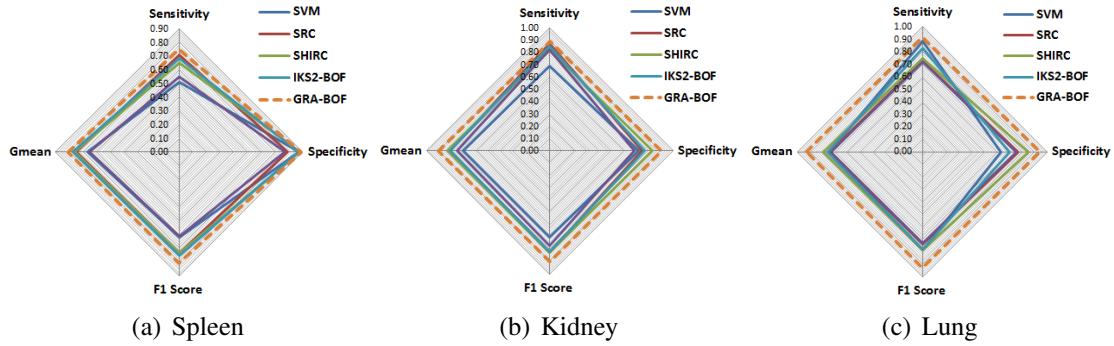


Figure 5.4: Radar chart for average results on (a) Spleen dataset, (b) Kidney dataset, and (c) Lung dataset

of-art methods. The overall accuracy on the ADL dataset is 88.3% which is better than other considered methods. Further, the average classification accuracy of the GRA-BOF method on the blue histology dataset is 74.5%, which is highest among other considered methods.

Moreover, a comparison of the F1 score, G-mean, sensitivity, and specificity is represented using the radar chart in Figure 5.4. The radar chart shows that the method with the largest area and symmetrical shape performs better than other. From the figure, it can be illustrated that the GRA-BOF method shows better results than all other methods. These results validate that the efficacy of the GRA-BOF method for histopathological image classification.

5.4 Summary

This chapter introduced a new Grey relational analysis based BOF method (GRA-BOF) which improves the efficiency of the codebook construction phase of the standard BOF method. The method uses a Grey relational analysis for similarity measure in the codebook construction phase of BOF. The GRA-BOF method has been validated on three datasets, Kidney, Lung, and Spleen of ADL histopathological image dataset along with the blue histology tissue image dataset. The average accuracy of the GRA-BOF method is 88.3% and 74.5% on ADL and Blue histology datasets respectively which are the highest among other state-of-the-art methods. The classification results of the GRA-BOF method has also been analyzed using confusion matrix, precision, recall, G-mean, F1 score, and radar charts. The experimental results validate that the GRA-BOF method outperforms the other considered methods for histopathological image classification.

The next chapter discusses the vector quantization phase of BOF and presents a new efficient vector quantization method in the BOF framework.

CHAPTER 6

WEIGHTED TWO DIMENSIONAL VECTOR QUANTIZATION

6.1 Introduction

Feature encoding is a third phase of the BOF method which represents the images in terms of a codebook. This process affects the accuracy and speed of the image classification process. In the last decade, various feature encoding methods have been presented in the literature. These methods can broadly be classified into three types, namely voting-based, reconstruction-based, and super vector-based methods.

Voting-based methods encode the images in terms of histograms which describe the occurrences features in the bins of visual words. A most popular voting-based method is the hard assignment (HA) [33] method which allocates the features to the nearest visual word. However, for attaining better performance, the codebook size should be large [208]. To mitigate this problem, the HA method is extended by using GMM based assignment and known as the Fisher vector (FV) [208] or fisher kernel. FV performs well for small size codebooks. However, there is an ambiguity in assigning the features with continuous values to the discrete visual words [209]. Therefore, KCB and SA [209] methods have been presented to overcome the ambiguity problem. In these methods, each feature vote for all the visual words based on the Gaussian kernel function is applied on Euclidean distance between the visual words and the considered feature. However, voting-based methods are not able to reconstruct the features after representation which may be important in some applications [124] [125].

Reconstruction-based feature encoding methods can reconstruct the features from the codewords. In literature, various encoding methods have been existed based on a reconstruction-based approach. Yang et al. [120] proposed a sparse coding method which uses the codewords sparsely to represent the features. For better image classification, sparse coding can be combined with SPM (spatial pyramid matching) and maximum pooling method to be used with a

linear SVM. Yang et al [210] have shown that sparse coding performs better for high dimensional representations by using the feature space with manifolds. Furthermore, Yu and Zhang [123] proposed LCC method which considers local manifold learning to incorporate locality in reconstruction. However, these methods are computationally expensive. To overcome this, Wang et al. [122] proposed a fast and efficient implementation of LCC known as LLC. It uses the locality constraints to describe each feature into their local coordinate system. Further, Gao et al. [211] proposed Laplacian sparse coding to increase the robustness of sparse encoding methods.

Furthermore, super vector-based encoding methods use higher order statistics and aggregate them to show high dimensional representations [123]. Zhou et al. [124] proposed a super vector coding (SVC) method which reduces the reconstruction error. It is an extension of the vector quantization method applied to high dimensional descriptor space. For the practical exploitation of SVC based high dimensional methods, Jegou et al. [126] used the dimensionality reduction along with the indexing method in the proposed vector of locally aggregated descriptors method (VLAD) which takes less memory and achieved better performance. The VLAD method is further enhanced using image similarity measure in vectors of locally aggregated tensors (VLAT) method proposed by Picard and Gosselin [212]. Further, Zhao et al. [213] used the feature maps to find the vector difference between the high dimensional feature vectors and used the vector difference to perform encoding.

All the above-discussed feature encoding methods are designed and developed in different scenarios and applications. For histopathological image classification using BOF, mainly voting based methods have been used in the literature [29] [85] [142] in which each extracted feature casts its vote to a particular visual word only. However, this limits the feature encoding method to consider only a single feature due to which less information is incorporated in the representation and hence, affects the classification performance. Therefore, this chapter introduces a new weighted two-dimensional vector quantization method in BOF to mitigate this issue. The efficiency of the proposed weighted two-dimensional vector quantization method has been tested and statistically validated for the classification of histopathological images.

6.2 Preliminaries

In this section, three popular local feature extraction methods, used in this work, are briefed.

1. Scale Invariant Feature Transform: SIFT detects and describes the low-level features from the digital images. It finds some interest points and then represents them in a quantitative manner known as descriptors. These descriptors are invariant to scale, rotation, and illumination conditions. The process of the SIFT descriptor calculation is discussed below.

- *Approximation of interest point locations:* In this phase, those locations and scales are detected from the image which can be identifiable in different viewpoints of the same object. This is achieved by applying the Gaussian filtering (i.e., the scale space function (G)) on the image and defined as follows:

$$G(i, j, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(i^2+j^2)/2\sigma^2} \quad (6.1)$$

$$A(i, j, \sigma) = G(i, j, \sigma) * I(i, j) \quad (6.2)$$

where, $I(i, j)$ is the image, $G(i, j, \sigma)$ is a Gaussian function defined over σ scale, and $*$ represents the convolution operator. To detect the stable locations of interest points, difference of Gaussian's has been computed between two images with different scales (e.g., one scale is l times other) and can be defined as follows:

$$DoG(i, j, \sigma) = A(i, j, l\sigma) - A(i, j, \sigma) \quad (6.3)$$

The resulting stack of DoG images is called a scale space pyramid. The extrema values (i.e., maxima or minima) of interest points in the scale space pyramid is known as keypoints. To identify interest points as a keypoints, its value is compared with 9 neighboring interest points on two different scales (one up and one down in the scale space pyramid) and 8 neighboring interest points on the same scale. The points are said to be extrema if this point contains maximum or minimum value among all other comparative points.

- *Refining interest point locations:* All the localized keypoints are not relevant due to their low contrast or the keypoints may reside on the edges which are not considered as the salient points. Therefore, these points should be discarded. In SIFT methods,

second order derivative Hessian matrix is used to reveal the local information.

$$Hes = \begin{bmatrix} DoG_{ii} & DoG_{ij} \\ DoG_{ij} & DoG_{jj} \end{bmatrix} \quad (6.4)$$

The surface of DoG_{ij} contains the maximal and minimal curvatures which can be identified by the eigenvalues of Hes . The elimination of extrema is performed based on poor localization. For a keypoint to be survived, the ratio between the smallest and largest eigenvectors should be less than a threshold.

- *Orientation assignment:* Once the keypoints locations are determined and refined, these must be described with the help of their orientations and magnitudes. This helps in keypoints to become rotation invariant. The approach of finding the keypoint orientations are as follows:

- The Gaussian smoothed image ($A(i, j, \sigma)$) is used to find out the magnitude (m) and orientation (θ) of each keypoint at location (i, j)

$$m(i, j) = \sqrt{(A(i+1, j) - A(i-1, j))^2 + (A(i, j+1) - A(i, j-1))^2} \quad (6.5)$$

$$\theta(i, j) = \tan^{-1}((A(i, j+1) - A(i, j-1)) / (A(i+1, j) - A(i-1, j))) \quad (6.6)$$

- An orientation histogram is created with the help of sample points (from a small region of size $(n \times n)$) gradient orientations.
- Analyze the histogram and find the highest frequency along with other 80% of the highest frequency and use these to assign the orientation of the keypoint.
- Multiple histograms may be assigned some keypoints.
- The peak position is interpolated by fitting the parabola to peak values of three histograms.
- *Descriptors calculation for each keypoint:* After creating the histogram, each keypoint is represented by a vector. If the histogram contains b bins then the size of the descriptor would be bn^2 . For example, if 4×4 region is considered for histogram creation and each histogram contains 8 bins then the size of the descriptor vector

is 128. These descriptors are rotation invariant and the level of scale in-variance is adjusted by changing the considered window size.

Irshad et al. [54] presented an automated method to detect mitosis to assist the pathologists. The color spaces are distinguished by analyzing the combined pattern of texture features, namely SIFT, co-occurrence features, and run-length features. These features are used for the classification with the help of SVM. Further, Raza et al. [55] used scale invariant features in the BOF method for the classification of subtypes in renal cell carcinoma. Caicedo et al. [29] also used the BOF method with SIFT for feature detection and description to classify histopathological images automatically. Raza et al. [35] studied and analyzed the rotation and scale in-variance behavior in the bag-of-features method for the classification of histopathological images of RCC.

2. Speeded-up Robust Features: SURF [46] is a fast and efficient interest point detector and descriptor method which uses the integral image for the convolutions. The method works in three phases, namely keypoint detection, description, and matching. A brief description of each phase is given below.

- *Keypoint detection:* SURF is a blob detector and uses the integral images for fast processing. To find the keypoints, it uses the Hessian matrix. A keypoint is selected if the determinant of the Hessian matrix is maximal around that keypoint. For an image I , the Hessian matrix $HM(p, q, \sigma)$ at scale σ is given by Eq. (6.7).

$$HM(p, q, \sigma) = \begin{pmatrix} CSS_{pp}(p, q, \sigma) & CSS_{pq}(p, q, \sigma) \\ CSS_{qp}(p, q, \sigma) & CSS_{qq}(p, q, \sigma) \end{pmatrix} \quad (6.7)$$

Further, as the keypoints may be detected at different scales, box filters with different sizes are used for the implementation of scale spaces. In SURF, 9×9 filter is used to obtain the lowest level at scale $\sigma = 1.2$ for identifying the blob-response maps having a high spatial resolution.

- *Description of keypoints:* The objective of a descriptor is to represent every keypoint by their neighborhood pixels intensity distribution. The descriptor size affects the computation cost and keypoint matching accuracy. Short descriptors may give more false positives but, these are more robust. SURF calculates the descriptor vec-

tor of every keypoints in two phases, namely orientation assignment and descriptor component. In the orientation assignment phase, a circular region of radius 6σ is considered to compute the Haar wavelet responses which are further weighted keypoint's Gaussian function and plotted in 2D space. Then, the responses within a sliding window of $\pi/3$ size are summed up to calculate the dominant orientation. The size of the orientation window affects the robustness and resolution of the descriptor. In SURF, the circular region is further divided into 4×4 sub-cells and for each cell, Haar wavelet outcomes are computed and weighted with the Gaussian function. The orientation window size is considered as 20σ .

- *Matching:* Finally, matching pairs are found by evaluating the descriptors attained from different image scales.

Yang et al. [56] solved the problem of image stitching in biomedical research concerning the whole sections or large areas. The proposed method considers SURF method for feature extraction in a fast and efficient manner. Sanghavi and Agaian [57] proposed an automated method for the classification of histopathological images of prostate cancer based on the bag-of-feature method. For feature extraction, a comparative analysis is presented between SURF and SIFT features based bag-of-features method. SURF returns higher sensitivity as compared to SIFT-based bag-of-features. Wand and Chen [58] addressed the problem of image alignment in medical or clinical diagnosis. The authors have considered two types of medical images, namely tissue images, and X-ray images and compared with five other methods, namely TrackEM2, UnwarpJ, BUnwarpJ, mutual information, and SURF. The proposed method shows better results on both of the datasets.

3. **Oriented FAST and Rotated BRIEF:** ORB [47] uses oriented FAST [59] method as a high speed corner detector and rotated BRIEF [60] as a descriptor. The details of these two sub-components are given as follows:

- *Keypoint detection using oriented FAST:* The FAST keypoints are detected from a circular ring centered at the keypoints with radius 9. However, it attains larger responses with the edges and detection of corners are not possible. Therefore, FAST keypoints are sorted based on the Harris corner parameter and selects only the top N keypoints. Further, FAST features are also calculated at each level of scale pyramid

to generate multi-scale features.

- *Orientation assignment:* The corner orientation is assigned by the intensity centroid approach. The orientation may be imputed by a vector having a corner's intensity from the center keypoint. Mathematically, the moment (m) of a small region can be defined in Eq. (6.8) which is further used to find the centroid (C) as given in Eq. (6.9).

$$m_{ab} = \sum_{p,q} p^a q^b I(p,q) \quad (6.8)$$

$$C = \left(\frac{m_{10}}{m_{01}}, \frac{m_{00}}{m_{00}} \right) \quad (6.9)$$

Now, a vector (\overrightarrow{OC}) is constructed from corner's center (O), computed centroid (C) and the orientation (ϕ) assigned to this region and defined in Eq. (6.10).

$$\phi = \begin{cases} \tan^{-1} \frac{m_{01}}{m_{10}} & \text{if } \text{color} = \text{bright} \\ \tan^{-1} \frac{m_{01}}{m_{10}} + 180^\circ & \text{if } \text{color} = \text{dark} \end{cases} \quad (6.10)$$

However, the color of the corner is ignored in orientation assignment. A simplified version of orientation of the region is given in Eq. (6.11).

$$\phi = \arctan(m_{01}, m_{10}) \quad (6.11)$$

- *Keypoint descriptor:* For generating the descriptor vector of a keypoint, a variant of BRIEF has been used which is an improvement of original BRIEF [60] by incorporating learning step to make it rotational invariant. The BRIEF method works on binary patches which are obtained by Eq. (6.12).

$$\tau(R; p, q) = \begin{cases} 1 & R(p) < R(q) \\ 0 & R(p) \geq R(q) \end{cases} \quad (6.12)$$

where, $R(p)$ is the intensity value of R at location p . Based on n binary operations,

a feature vector (FV) is defined and is given by Eq. (6.13).

$$FV_n(R) = \sum_{1 \leq j \leq n} 2^{j-1} \tau(R; p_j, q_j) \quad (6.13)$$

Generally, the length of the feature vector (n) is considered as 256 and for smoothing of the patches integral images have been used.

Further, to make BRIEF invariant to rotation, a more efficient method, namely steer BRIEF has been presented which defines the feature vector (G) by considering the patch (R) and its steered version of rotation matrix (S_ϕ). It is given by Eq. (6.14).

$$G_n(R, \phi) = FV_n(R)|(p_j, q_j) \in S_\phi \quad (6.14)$$

where, S_ϕ is calculated as $S_\phi = Rot_\phi S$. Rot_ϕ is the rotation matrix and S is a $2 \times n$ matrix, given in Eq. (6.15).

$$S = \begin{pmatrix} p_1, p_2, \dots, p_n \\ q_1, q_2, \dots, q_n \end{pmatrix} \quad (6.15)$$

6.3 Weighted Two Dimensional Vector Quantization

The new weighted two-dimensional vector quantization (W2DVQ) based BOF method contains four main phases, namely feature extraction, codebook construction, W2DVQ based feature encoding, and classification. The overall process is also depicted in Figure 6.1 and step by step description of each phase is discussed below.

1. *Feature extraction:* Local features are extracted from the images which are arranged in two sets, namely training set and validation set. For feature extraction, any two feature detection and descriptor methods (FE_1 and FE_2) as explained in Section 6.2 are used. The two feature extraction methods FE_1 and FE_2 generate two different feature vectors X and Y respectively, each having its own attributes.

$$X = \{x_1, x_2, \dots, x_N\} \in \mathbb{R}^{D_1 \times N} \quad (6.16)$$

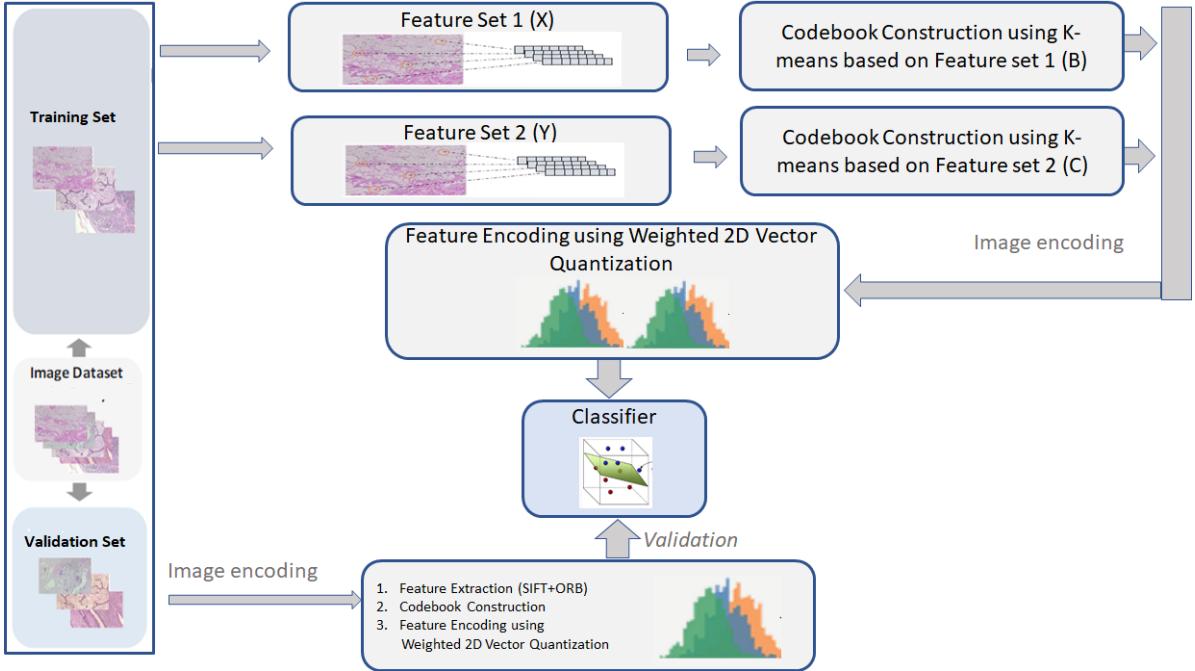


Figure 6.1: The overall procedure of the BoF method based on W2DVQ method

$$Y = \{y_1, y_2, \dots, y_M\} \in \mathbb{R}^{D_2 \times M} \quad (6.17)$$

where, x_i and y_j are the i^{th} and j^{th} features, extracted by the local feature extraction methods FE_1 and FE_2 respectively. D_1 and D_2 are the dimensions of each feature, generated by FE_1 and FE_2 respectively.

2. *Codebook construction:* Due to the complexities of histopathological images, the extracted descriptor sets are large in sizes. Moreover, all the features in the descriptor sets are not relevant for the automated classification. Therefore, these descriptor sets are clustered using K-means clustering method which generates cluster heads. These cluster heads are used as the most representative features of an image and known as visual words. The collection of these visual words is called codebook or visual vocabulary. As the proposed method uses two descriptor sets (X and Y), therefore, two codebooks (\mathcal{B} and \mathcal{C}) are constructed separately, each of size K , which are further used in encoding phase.

$$\mathcal{B} = \{b_1, b_2, \dots, b_K\} \in \mathbb{R}^{D_1 \times N} \quad (6.18)$$

$$\mathcal{C} = \{c_1, c_2, \dots, c_K\} \in \mathbb{R}^{D_2 \times M} \quad (6.19)$$

where, b_i and c_j are the i^{th} visual words, defined over X and Y respectively.

3. *W2DVQ based feature encoding:* Once the codebooks are generated, they are used to encode the training and validation images. In this works, a new encoding method, namely two-dimensional vector quantization is introduced which is an extension of voting-based feature encoding. In voting-based encoding methods, each descriptor directly votes for the visual word using a specific strategy. In the W2DVQ method, a *two*-dimensional code s of size $K \times K$ is constructed to represent the votes of the whole codebook where, K corresponds to the size of the codebook.

The voting value of the code ($s(i, j)$) is measured by a weighted function of both descriptors (X and Y) using the following equation.

$$\forall i \ s(i, j)_{i=1, \dots, K} = \alpha \cdot \phi_1(i) + (1 - \alpha) \cdot \phi_2(j), \ j = 1, 2, \dots, K \quad (6.20)$$

where, $\alpha \in (0, 1)$ is a weighting factor. The $\phi_1(i)$ and $\phi_2(j)$ are defined in Eq. (6.21) and Eq. (6.22) respectively.

$$\phi_1(i) = \begin{cases} 1 & \text{if } i = \underset{j}{\operatorname{argmin}}(\|x - b_j\|_2), \ i = 1, 2, \dots, K \\ 0 & \text{otherwise} \end{cases} \quad (6.21)$$

$$\phi_2(j) = \begin{cases} 1 & \text{if } j = \underset{l}{\operatorname{argmin}}(\|y - c_l\|_2), \ l = 1, 2, \dots, K \\ 0 & \text{otherwise} \end{cases} \quad (6.22)$$

4. *Classification:* The last step of the W2DVQ based BOF method is classification. After encoding of each image by the W2DVQ method, the encoded images along with their labels are given to support vector machine classifier for the training. Once the model is trained, the similar steps (1 to 3) are applied to the validation images and the encoded images are passed to a trained classifier to predict the labels of the images.

6.4 Experimental Results

The efficacy of the new method is analyzed for classifying histopathological images. For the same, two standard histopathological image datasets are considered, namely ADL dataset and Blue histology dataset. For comparative analysis, voting based encoding method, namely hard voting or vector quantization has been considered and for feature extraction, three methods, namely SIFT, SURF, and ORB have been used. The codebook size is empirically set to 500 for visual word generation. For classification of images SVM classifier is used [214] is used.

6.4.1 Performance Analysis on ADL Dataset

ADL dataset contains the images of three organs and each organ contains inflamed and healthy histopathological images of these organs. This chapter used a binary classifier to identify these images using the BOF method with VQ and W2DVQ method. Table 6.1, 6.2, and 6.3 depict the confusion matrices returned by the BOF method with VQ and W2DVQ encoding methods on Kidney, Lung, and Spleen images of ADL dataset respectively. In the tables, the methods with one feature extractor use the VQ encoding technique and the methods with a combination of two feature extractors use W2DVQ encoding technique. These tables represent classification and miss-classification rates returned by the BOF method on each organ image dataset by considering a combination of three feature extraction methods. From these tables, it can be observed that the combination of two features in encoding methods performs better than single feature based encoding methods. In this chapter, the three combinations have been formed using SIFT, SURF, and ORB and used in the W2DVQ method in BOF. Table 6.1 shows the results for Kidney images and it can be seen that combination SIFT and ORB performs well in identifying healthy images while the combination of ORB and SURF gives the better accuracy in identifying the inflamed images of the kidney. Similarly, Table 6.2 & 6.3 show that the combination of SIFT and ORB features outperforms other combinations and achieve better accuracy in identifying healthy as well as inflamed images of Lung and Spleen.

Moreover, a more detailed comparative analysis of the considered methods has been performed in terms of sensitivity, specificity, precision, false negative rate, and accuracy. In histopathological image classification, sensitivity plays an important role as it is the rate at which inflamed images are identified. It can be observed from the Table 6.4 that the combination of SIFT and SURF features attain higher sensitivity of 84.4% followed by the combination

Table 6.1: Confusion matrices returned by the BOF method on Kidney images with HV and W2DVQ encoding methods

Class	Healthy	Inflamed	Method
Healthy	0.67	0.28	SIFT
	0.65	0.29	ORB
	0.64	0.30	SURF
	0.78	0.17	SIFT+ORB
	0.76	0.51	SIFT+SURF
	0.69	0.17	ORB+SURF
Inflamed	0.33	0.72	SIFT
	0.35	0.71	ORB
	0.36	0.70	SURF
	0.22	0.82	SIFT+ORB
	0.14	0.49	SIFT+SURF
	0.31	0.83	ORB+SURF

Table 6.2: Confusion matrices returned by the BOF method on Lung images with HV and W2DVQ encoding methods

Class	Healthy	Inflamed	Method
Healthy	0.74	0.27	SIFT
	0.73	0.28	ORB
	0.72	0.29	SURF
	0.82	0.17	SIFT+ORB
	0.79	0.19	SIFT+SURF
	0.79	0.26	ORB+SURF
Inflamed	0.26	0.73	SIFT
	0.27	0.72	ORB
	0.28	0.71	SURF
	0.18	0.83	SIFT+ORB
	0.21	0.81	SIFT+SURF
	0.21	0.74	ORB+SURF

Table 6.3: Confusion matrices returned by the BOF method on Spleen images with HV and W2DVQ encoding methods

Class	Healthy	Inflamed	Method
Healthy	0.59	0.25	SIFT
	0.58	0.24	ORB
	0.57	0.25	SURF
	0.76	0.21	SIFT+ORB
	0.72	0.22	SIFT+SURF
	0.67	0.22	ORB+SURF
Inflamed	0.41	0.75	SIFT
	0.42	0.76	ORB
	0.43	0.75	SURF
	0.24	0.79	SIFT+ORB
	0.28	0.78	SIFT+SURF
	0.33	0.78	ORB+SURF

Table 6.4: Comparative analysis of considered methods on ADL dataset in terms of various performance parameters

Organ	Algorithms	Sensitivity	Specificity	Precision	FNR	Accuracy
Kidney	SIFT	0.670	0.720	0.705	0.330	0.695
	ORB	0.650	0.710	0.691	0.350	0.680
	SURF	0.640	0.700	0.681	0.360	0.670
	SIFT+ORB	0.780	0.828	0.821	0.220	0.804
	SIFT+SURF	0.844	0.490	0.598	0.156	0.658
	ORB+SURF	0.690	0.830	0.802	0.310	0.760
Lung	SIFT	0.740	0.730	0.733	0.260	0.735
	ORB	0.730	0.720	0.723	0.270	0.725
	SURF	0.720	0.710	0.713	0.280	0.715
	SIFT+ORB	0.820	0.830	0.828	0.180	0.825
	SIFT+SURF	0.790	0.810	0.806	0.210	0.800
	ORB+SURF	0.790	0.740	0.752	0.210	0.765
Spleen	SIFT	0.590	0.750	0.702	0.410	0.670
	ORB	0.580	0.760	0.707	0.420	0.670
	SURF	0.570	0.750	0.695	0.430	0.660
	SIFT+ORB	0.760	0.790	0.784	0.240	0.775
	SIFT+SURF	0.720	0.780	0.766	0.280	0.750
	ORB+SURF	0.670	0.780	0.753	0.330	0.725

of SIFT and ORB features as 78%. Whereas, for Lung and Spleen images the combination of SIFT and ORB features perform better in identifying inflamed images and return high sensitivity of 82% and 76% respectively. Similarly, specificity is the rate of identifying the healthy images which are healthy and precision is the rate of identifying total predicted healthy images by the classifier. The combination of SIFT and ORB features outperform in terms of precision and specificity while applying on Lung and Spleen images. However, for Kidney images, the combination of ORB and SURF features perform better in terms of specificity while the combination of SIFT and SURF features perform better in terms of precision.

Further, the FNR in medical diagnosis gives the assurance of the presence of disease while actually, it was absent. Therefore, the combination of SIFT and ORB features based the BOF method performs better on Lung and Spleen images as it returned minimum FNR while the combination of SIFT and SURF features based the BOF method perform better on Kidney images. For better visualization of the performance of different combinations of features in encoding methods, radar charts have also been constructed on four parameters, namely F1-score, sensitivity, specificity, and G-mean. These parameters must have high values for better performance. Therefore, the method with larger area coverage on radar charts is considered

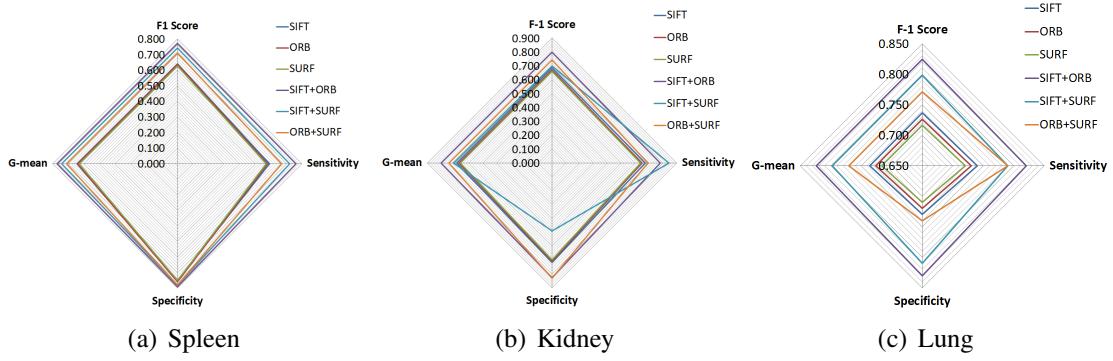


Figure 6.2: Radar chart for average results on (a) Spleen dataset, (b) Kidney dataset, and (c) Lung dataset

CT	0.84	0.05	0	0.09
ET	0.1	0.73	0.05	0.12
MT	0.05	0.09	0.78	0.08
NT	0.03	0.05	0.23	0.69
CT	ET	MT	NT	

CT	0.7	0.1	0.2	0
ET	0.1	0.75	0.05	0.1
MT	0.1	0.15	0.6	0.15
NT	0.3	0.05	0.05	0.6
CT	ET	MT	NT	

CT	0.7	0.1	0.1	0.1
ET	0.01	0.69	0	0.3
MT	0	0.15	0.7	0.15
NT	0.25	0.1	0	0.65
CT	ET	MT	NT	

(a) SIFT+ORB (b) SIFT+SURF (c) SURF+ORB

Figure 6.3: Confusion Matrix returned by the BOF method using different combinations of features on Blue Histology dataset.

better. Figure 6.2 depicts the radar chart of each considered method on Kidney, Lung, Spleen images. It can be observed that for all three organ images, the combination of the SIFT and ORB based method covers the larger area, therefore it can be considered as the better combination for the W2DVQ feature encoding method on ADL dataset.

6.4.2 Performance Analysis on Blue Histology Dataset

The blue histology dataset contains four types of tissue images stained with various staining methods. This diversity in the staining methods make these images more complex and challenging for the automated classification. For the training and validation, multi-class SVM has been used. The confusion matrices returned by the classifier on dual feature based vector quantization method are depicted in Figure 6.3. The diagonal elements in the figures represent the true positives i.e., the rate of correctly identifying the tissue image. It can be seen that the combination of SIFT and ORB features based method attained high accuracy in identifying connective, muscle, and nervous tissues. For epithelial tissue, the combination of SIFT and SURF feature-based method has a good identification rate i.e., 75%.

Moreover, for detailed analysis, a comparative analysis based on various performance pa-

Table 6.5: Comparative analysis of the considered methods on Blue Histology dataset in terms of various performance parameters

Category	Algorithms	Sensitivity	Specificity	Precision	FNR	F1 Score
CT	SIFT	0.84	0.75	0.54	0.16	0.66
	ORB	0.80	0.68	0.46	0.20	0.58
	SURF	0.70	0.75	0.48	0.30	0.57
	SIFT+ORB	0.86	0.95	0.85	0.14	0.85
	SIFT+SURF	0.70	0.83	0.58	0.30	0.64
	ORB+SURF	0.70	0.91	0.73	0.30	0.71
ET	SIFT	0.64	0.84	0.58	0.36	0.61
	ORB	0.53	0.90	0.63	0.47	0.57
	SURF	0.63	0.87	0.60	0.37	0.62
	SIFT+ORB	0.76	0.94	0.80	0.24	0.78
	SIFT+SURF	0.75	0.90	0.71	0.25	0.73
	ORB+SURF	0.69	0.88	0.66	0.31	0.68
MT	SIFT	0.55	0.95	0.79	0.45	0.65
	ORB	0.75	0.85	0.63	0.25	0.68
	SURF	0.65	0.92	0.72	0.35	0.68
	SIFT+ORB	0.78	0.91	0.74	0.22	0.76
	SIFT+SURF	0.60	0.90	0.67	0.40	0.63
	ORB+SURF	0.70	0.97	0.88	0.30	0.78
NT	SIFT	0.31	0.93	0.56	0.69	0.40
	ORB	0.05	0.95	0.25	0.95	0.08
	SURF	0.35	0.92	0.58	0.65	0.44
	SIFT+ORB	0.69	0.90	0.72	0.31	0.70
	SIFT+SURF	0.60	0.92	0.71	0.40	0.65
	ORB+SURF	0.65	0.82	0.54	0.35	0.59

rameters such as sensitivity, specificity, precision, false negative rate, and F1 score have also been performed and results are represented in Table 6.5. The results are enlisted separately for each category of tissue images. From the table, it can be stated that for the connective and epithelial tissues, the combination of SIFT and ORB features performs better on all considered performance parameters. However, for muscle tissues, the combination of ORB and SURF features perform better in terms of precision, specificity, and F-1 score, and for nervous tissues, ORB with vector quantization perform better in terms of specificity. However, for these tissue images, the SIFT and ORB features perform better in terms of sensitivity. Furthermore, the radar charts have been depicted in Figure 6.4 for each category of tissue images. It can be stated that the area covered by SIFT and ORB based encoding method is larger on every category of tissue images.

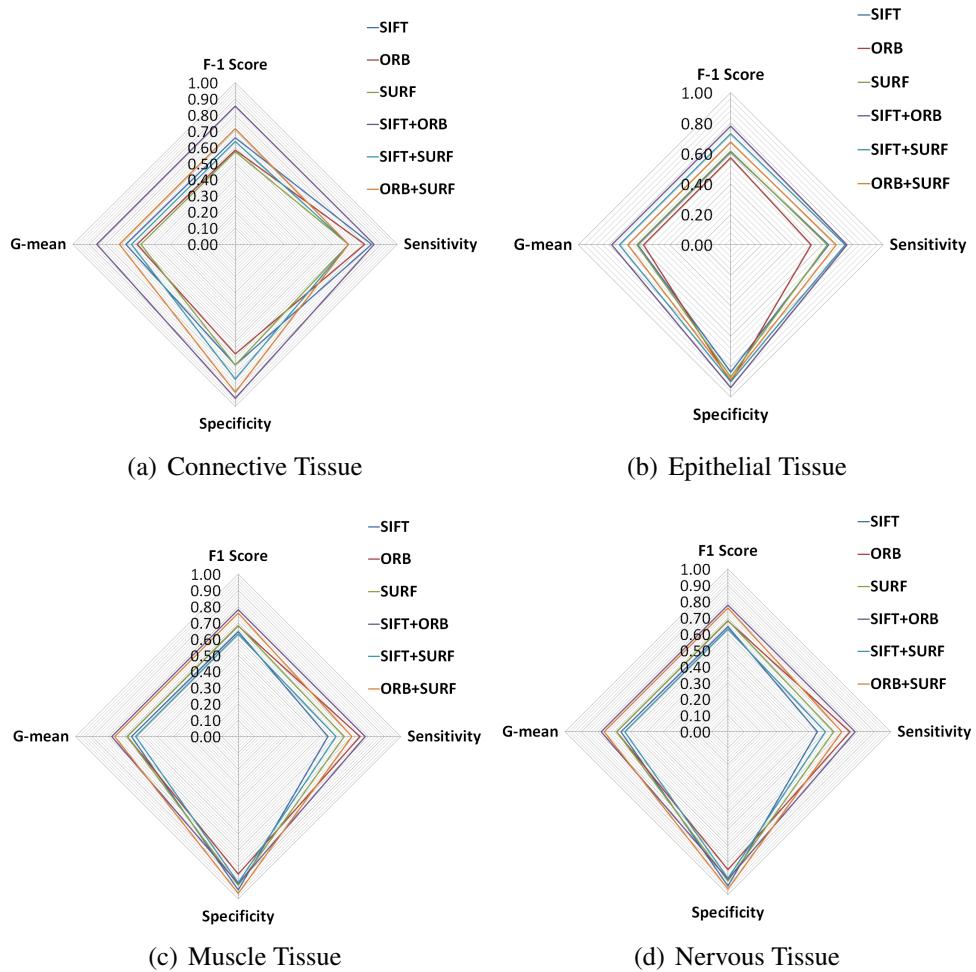


Figure 6.4: Radar chart for average results obtained for SVM classifier on Blue histology dataset

6.4.3 Classification Accuracy

Table 6.6 depicts the average classification accuracy achieved by the W2DVQ method and compared with the existing VQ method over the considered histopathological datasets. It can be observed from the table that each pair of feature descriptors achieves superior average accuracy than the single feature descriptor over both the datasets. On the ADL dataset, the SIFT+ORB pair achieves the highest average accuracy i.e., 80.13% followed by the ORB+SURF pair which reported 75% average accuracy while the SIFT+SURF pair returned 73.59%. Similarly, the same trend can be observed for the Blue histology dataset where the SIFT+ORB pair comes first.

From the results, it can be observed that the single feature based vector quantization methods are not adequate as compared to weighted two-dimensional vector quantization methods for feature encoding phase. The reason behind the better performance of the later technique is that

Table 6.6: Comparative analysis of average classification accuracy the proposed W2DVQ method and other considered methods (in percentage)

Features	Encoding	ADL	Blue histology
SIFT	HV	70	60
ORB	HV	69.16	53.16
SURF	HV	68.16	58.22
SIFT+ORB	2DVQ	80.13	77.13
SIFT+SURF	2DVQ	73.59	66.25
ORB+SURF	2DVQ	75	68.5

SVM can be trained by using two different features, having different qualities. For example, SIFT produces better results when images of different scales are to be handled [215], SURF works well for noisy and blurred images [215], and ORB performs better on images which are rotated and noisy [216]. Thus, the concept of the weighted two-dimensional vector quantization based feature encoding method provides better results. However, from the above quantitative and graphical analysis, it can be stated that the combination of SIFT and ORB features based encoding method performs better in both the considered datasets.

6.5 Summary

This work presented a new weighted two-dimensional vector quantization based feature encoding method in the BOF framework. The proposed method encodes the images in the feature encoding phase of bag-of-features using a weighted two-dimensional representation of SIFT and ORB features. The work is tested on two histopathological image datasets for the classification task. The performance is analyzed using various parameters such as sensitivity, specificity, precision, FNR, F1 Scores, and accuracy. Radar charts are also studied for the analysis of the proposed method. The experimental results show that the introduced method outperforms the other hard voting based methods.

The next chapter presents the use of the modifications, presented in various chapters to design an enhanced BOF method to be used for histopathological image classification.

CHAPTER 7

ENHANCED BAG-OF-FEATURES METHOD

7.1 Introduction

The output of this research is an automated histopathological image classification system based on the enhanced BOF method. The block diagram of the developed system is shown in Figure 7.1. This system consists of four phases (i) Feature extraction which uses SURF and ORB methods to extract the features from histopathological images, (ii) the features are clustered using GRA-BOF method as discussed in Section 5.2 to generate two codebooks, (iii) each image is encoded using the two dimensional vector quantization method, as discussed in Section 6.3, and (iv) the encoded images along with the labels are used to train the classifier which is further used to predict the labels of validated images. The detail description of each phase of the system is presented in the following section.

7.2 Feature Extraction using SIFT and ORB

The first phase of the new system is the feature extraction. For the same, the histopathological image data set is divided into training and validation sets and the feature extraction methods are used to extract the features from these images. The combination of SIFT and ORB outperforms other feature extraction methods as discussed in Chapter 6. Therefore, both methods are applied for extracting the features from the images. The two feature extraction methods SIFT and ORB generate two different feature vectors X and Y respectively, each having its own attributes and are defined by Eq. (7.1) and (7.2)

$$X = \{x_1, x_2, \dots, x_N\} \in \mathbb{R}^{D_1 \times N} \quad (7.1)$$

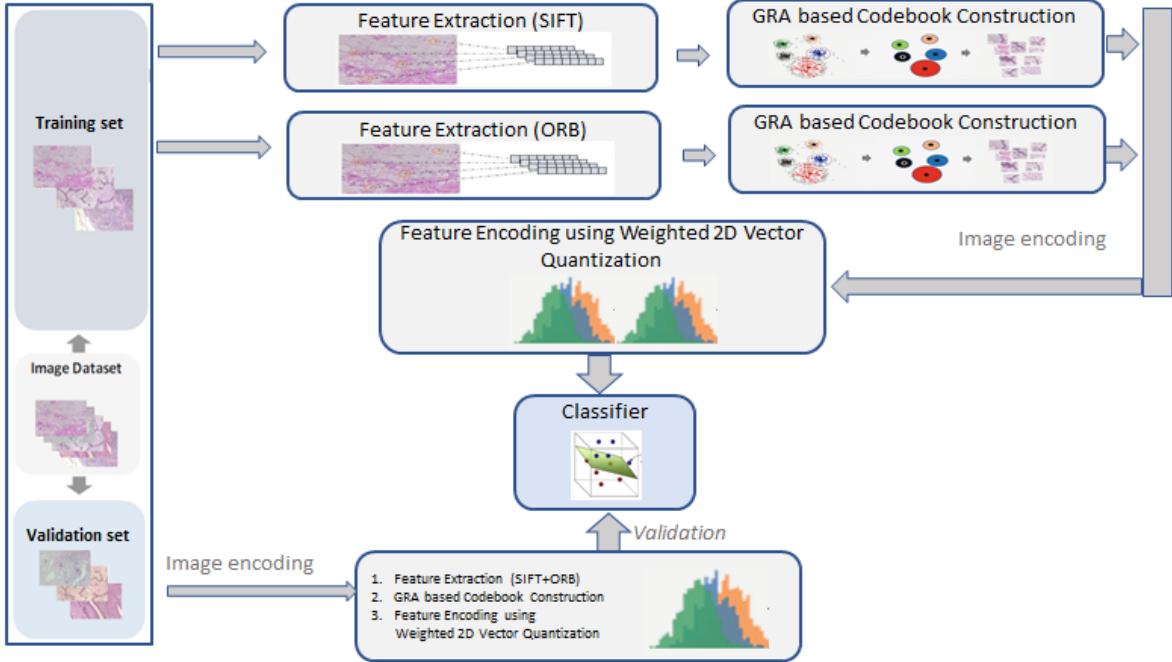


Figure 7.1: Developed histopathological image classification system

$$Y = \{y_1, y_2, \dots, y_M\} \in \mathbb{R}^{D_2 \times M} \quad (7.2)$$

where, x_i and y_j are the i^{th} and j^{th} features and D_1 and D_2 are the dimensions of each feature, generated by SIFT and ORB respectively.

Due to the complexity of histopathological images the resulting feature or descriptor vectors are large in size and high dimensional. Therefore, there is a requirement to find a more representative feature vector, which is performed in the codebook construction phase.

7.3 Codebook Construction using Grey Relational Analysis

In this phase, clustering is performed on extracted features to find promising and representative features known as visual words. In the proposed system, Grey relational analysis based clustering method is used to generate the visual words. The GRA based clustering method is computationally efficient and generates most representative visual words. For better classification results, two different codebooks \mathcal{B} and \mathcal{C} are constructed separately, each of

size K , which are further used in encoding phase.

$$\mathcal{B} = \{b_1, b_2, \dots, b_K\} \in \mathbb{R}^{D_1 \times N} \quad (7.3)$$

$$\mathcal{C} = \{c_1, c_2, \dots, c_K\} \in \mathbb{R}^{D_2 \times M} \quad (7.4)$$

where, b_i and c_j are the i^{th} visual words, defined over X and Y respectively

7.4 Feature Encoding using W2DVQ Method

Feature encoding is the third phase of the proposed system in which each image is encoded using the new weighted two-dimensional vector quantization method as discussed in Section 6.3. In the W2DVQ method, a 2-dimensional code s of size $K \times K$ is constructed to represent the votes of the whole codebook where K corresponds to the size of the codebook.

The voting value of the code $(s(i, j))$ is measured by a weighted function of both descriptors (X and Y) using the following equation.

$$\forall i \ s(i, j)_{i=1, \dots, K} = \alpha \cdot \phi_1(i) + (1 - \alpha) \cdot \phi_2(j), \ j = 1, 2, \dots, K \quad (7.5)$$

where, α is a weighting factor between $(0, 1)$. The $\phi_1(i)$ and $\phi_2(j)$ are defined in Eq. (7.6) and Eq. (7.7) respectively.

$$\phi_1(i) = \begin{cases} 1 & \text{if } i = \underset{j}{\operatorname{argmin}}(\|x - b_j\|_2), \ i = 1, 2, \dots, K \\ 0 & \text{otherwise} \end{cases} \quad (7.6)$$

$$\phi_2(j) = \begin{cases} 1 & \text{if } j = \underset{l}{\operatorname{argmin}}(\|y - c_l\|_2), \ l = 1, 2, \dots, K \\ 0 & \text{otherwise} \end{cases} \quad (7.7)$$

7.5 Classification

The last step of the proposed system is the classification. After encoding of each image by the W2DVQ method, the encoded images along with their labels are given to support vector

Table 7.1: The confusion matrix returned by GRA-weighted-BOF method on ADL histopathological image dataset

Organ	Class	Healthy	Inflamed
Kidney	Healthy	0.95	0.05
	Inflamed	0.15	0.85
Lung	Healthy	0.93	0.07
	Inflamed	0.03	0.97
Spleen	Healthy	0.91	0.09
	Inflamed	0.13	0.87

machine classifier for the training. Once the model is trained, the similar steps are applied to the validation images and the encoded images are passed to a trained classifier to predict the label of the image.

7.6 Experimental Results

The ADL and Blue histology image datasets are used to validate the efficacy of the newly introduced enhanced BOF method, called GRA-weighted-BOF which uses SIFT and ORB methods for feature extraction, GRA-BOF method for codebook construction and W2DVQ method for image encoding.

Table 7.1 and Figure 7.2 depict the confusion matrices returned by the GRA-weighted-BOF method on ADL and Blue histology datasets respectively. The rows in the confusion matrix denote the actual labels and columns represent the predicted labels by the classifier. From Table 7.1, it can be visualized that the rate of identifying inflamed lung images are 97% which is higher than the other two organ identification rates which are between 87% and 85% for spleen and kidney respectively. Similarly, from Figure 7.2, it can be observed that the identification

GRA-weighted BOF		Predicted			
		CT	ET	MT	NT
Actual	CT	0.8	0.17	0.02	0.01
	ET	0.1	0.8	0.08	0.02
	MT	0.15	0.05	0.75	0.05
	NT	0.1	0.1	0.05	0.75

Figure 7.2: The confusion matrix returned by GRA-weighted-BOF method on Blue histology image dataset

rate of the GRA-weighted-BOF method is 80% for connective and epithelial tissues while for

Table 7.2: The performance comparison of various proposed BOF methods on ADL and Blue histology dataset

S.No.	BOF Approach	ADL	Blue Histology
		Dataset	Dataset
1.	GKS-BOF	78	48
2.	SBBO-BOF	87	72.23
3.	GRA-BOF	88.3	74.5
4.	Weighted-BOF	79	74
5.	GRA-Weighted-BOF	91.2	77.5

both the muscle and nervous tissues it gives 75% identification rate.

Moreover, Table 7.2 shows the comparison of all the proposed methods over considered datasets and it can be observed that on the GRA-weighted-BOF method returned 91.2% and 77.5% average accuracy on ADL and Blue histology datasets respectively which are the highest among all other proposed methods. This validates that the proposed system outperforms the existing methods for histopathological image classification.

7.7 Summary

In this chapter, a new enhanced BOF method is developed based on the various new contributions introduced in this work for the classification of histopathological images. The new method first extracts the SIFT and ORB features from the training images and develop two codebooks using GRA based clustering method. Further, these codebooks are used by two-dimensional vector quantization method to encode each image. The encoded images are fed to the classifier along with their annotations for training. Once the classifier is trained, it is used to predict the labels of validation images which are passed to the classifier in encoded form without labels. The above described enhanced BOF method has been tested and validated on ADL and Blue histology image datasets. The experiential results show the efficacy of the newly introduced histopathological image classification system.

CHAPTER 8

CONCLUSION AND DIRECTION TO FUTURE WORK

In this thesis, an attempt has been made to identify the problems concerning histopathological image classification. Color variations, illumination variations, and presence of artifacts/noise are the major problems encountered in the analysis of histopathological images. In this work, some inherent drawbacks of existing automated classification methods used for the analysis of tissue section images are studied. The outcome of this study motivated to develop a histopathological image classification system based on the bag-of-features method.

This work introduces an efficient bag-of-features method for histopathological image classification. The main contribution of this work is divided into five folds. First, as the number of keypoints detected from histological images is of high dimensions, a new Grey relational analysis based keypoints selection technique is presented in the BOF method. Second, to overcome the limitations of the K-means based codebook construction algorithm in the BOF method, an efficient spiral biogeography-based optimization (SBBO) based BOF method has been presented. Third, to make the codebook construction computationally efficient, a new Grey relational analysis based method (GRA-BOF) has been introduced. Fourth, to take the merits of the two different feature descriptors, a new weighted two-dimensional vector quantization method has been presented. Finally, an efficient histopathological image classification system is designed which is based on the enhanced bag-of-features method. The developed system outperforms the other considered histopathological image classification methods.

8.1 Research Contribution

- The bag-of-features method for histopathological images is enhanced by incorporating a new Grey relational analysis based keypoints selection phase which reduces the extracted high dimensional features by 95% and 48% from the ADL and Blue histology datasets respectively. This also increases the respected classification accuracy by 14% and 11%.

- To overcome the limitations of k-means based codebook construction algorithm in the BOF method, an efficient spiral biogeography-based optimization (SBBO) has been presented which has been given the best mean ranking of 1.93 by Friedman test. The SBBO based codebook construction algorithm in the BOF method produces 72.23% and 87% accuracy for Blue histology and ADL datasets respectively. However, the method is computationally expensive.
- To make the codebook construction computationally efficient, a new Grey relational analysis based method (GRA-BOF) has been presented which improves the efficiency of codebook construction step of the standard BOF method and gives the accuracy of 74.5% and 88.3% for Blue histology and ADL datasets respectively.
- To take the merits of two different feature descriptors, a new weighted two-dimensional vector quantization method has been introduced which gives the classification accuracy of 80.13% and 77.13% for Blue histology and ADL datasets respectively using SIFT and ORB features.
- Finally, a new and efficient bag-of-features model is presented using GRA based codebook construction and weighted 2D vector quantization methods. The presented model gives the classification accuracy of 77.5% and 91.2% for Blue histology and ADL datasets respectively which is the best among other methods.

8.2 Direction to Future Work

Some directions for future extension of this work are given below.

- The developed BOF method can be used to analyze the different structures of histopathological and cytological images such as white blood cells, hair follicles, and others.
- The proposed keypoints selection method may be applied to other data sets such as microarrays data sets.
- The developed classification system may be explored for solving other real-world pattern recognition problems.
- To reduce the effect of illumination and staining variations, some pre-processing steps may be incorporated to increase the classification accuracy of the system.

REFERENCES

- [1] Stathonikos N. , Veta M. , Huisman A. , and Diest P. J. van , “*Going fully digital: Perspective of a dutch academic pathology lab,*” Journal of Pathology Informatics, vol. 4, pp. 114–129, 2013.
- [2] Srinivas U. , Mousavi H. S. , Monga V. , Hattel A. , and Jayarao B. , “*Simultaneous sparsity model for histopathological image representation and classification,*” IEEE Transactions on Medical Imaging, vol. 33, pp. 1163–1179, 2014.
- [3] Gurcan M. N. , Boucheron L. E. , Can A. , Madabhushi A. , Rajpoot N. M. , and Yener B. , “*Histopathological image analysis: A review,*” IEEE Reviews in Biomedical Engineering, vol. 2, pp. 147–171, 2009.
- [4] McCann M. T. , Ozolek J. A. , Castro C. A. , Parvin B. , and Kovacevic J. , “*Automated histology analysis: Opportunities for signal processing,*” IEEE Signal Processing Magazine, vol. 32, pp. 78–87, 2014.
- [5] Basavanhally A. N. , Ganesan S. , Agner S. , Monaco J. P. , Feldman M. D. , Tomaszewski J. E. , Bhanot G. , and Madabhushi A. , “*Computerized image-based detection and grading of lymphocytic infiltration in her2+ breast cancer histopathology,*” IEEE Transactions on Biomedical Engineering, vol. 57, pp. 642–653, 2009.
- [6] *Review of national cancer control activity in Australia - Ch 07: Diagnostic and assessment services.* Australian Government Cancer Australia, 2010.
- [7] Demir C. and Yener B. , “*Automated cancer diagnosis based on histopathological images: a systematic survey,*” Rensselaer Polytechnic Institute, Department of Computer Science, Tech. Rep., 2005.

- [8] Saraswat M. , Arya K. , and Sharma H. , “*Leukocyte segmentation in tissue images using differential evolution algorithm*,” Swarm and Evolutionary Computation, vol. 11, pp. 46–54, 2013.
- [9] Magee D. , Treanor D. , Crellin D. , Shires M. , Mohee K. , and Quirke P. , “*Colour normalisation in digital histopathology images*,” in Proc. of Optical Tissue Image Analysis in Microscopy, Histopathology and Endoscopy, United Kingdom, pp. 1–12, 2009.
- [10] Ong S. H. , Jin X. C. , Jayasooriah , and Sinniah R. , “*Image analysis of tissue sections*,” Computers in Biology and Medicine, vol. 26, pp. 269–279, 1996.
- [11] Tuytelaars T. , Mikolajczyk K. *et al.*, “*Local invariant feature detectors: a survey*,” Foundations and Trends® in Computer Graphics and Vision, vol. 3, pp. 177–280, 2008.
- [12] Dundar M. M. , Badve S. , Bilgin G. , Raykar V. , Jain R. , Sertel O. , and Gurcan M. N. , “*Computerized classification of intraductal breast lesions using histopathological images*,” IEEE Transactions on Biomedical Engineering, vol. 58, pp. 1977–1984, 2011.
- [13] Orlov N. , Shamir L. , Macura T. , Johnston J. , Eckley D. M. , and Goldberg I. G. , “*Wnd-charm: Multi-purpose image classification using compound image transforms*,” Pattern Recognition Letters, vol. 29, pp. 1684–1693, 2008.
- [14] Zana F. and Klein J.-C. , “*Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation*,” IEEE Transactions on Image Processing, vol. 10, pp. 1010–1019, 2001.
- [15] Ozdemir E. and Gunduz-Demir C. , “*A hybrid classification model for digital pathology using structural and statistical pattern recognition*,” IEEE Transactions on Medical Imaging, vol. 32, pp. 474–483, 2012.
- [16] Qureshi H. , Rajpoot N. , Nattkemper T. W. , and Hans V. , “*A robust adaptive wavelet-based method for classification of meningioma histology images*,” in Proc. of Workshop on Optical Tissue Image Analysis in Microscopy, Histology, and Endoscopy, MICCAI, London, pp. 31–42, 2009.

- [17] Gutiérrez R. , Rueda A. , and Romero E. , “*Learning semantic histopathological representation for basal cell carcinoma classification*,” in Proc. of SPIE, Medical Imaging 2013: Digital Pathology, Florida, United States, pp. 48–54, March 2013.
- [18] Zheng Y. , Jiang Z. , Xie F. , Zhang H. , Ma Y. , Shi H. , and Zhao Y. , “*Feature extraction from histopathological images based on nucleus-guided convolutional neural network for breast lesion classification*,” Pattern Recognition, vol. 71, pp. 14–25, 2017.
- [19] Arévalo J. , Cruz-Roa A. *et al.*, “*Histopathology image representation for automatic analysis: A state-of-the-art review*,” Revista Med, vol. 22, pp. 79–91, 2014.
- [20] Xu J. , Luo X. , Wang G. , Gilmore H. , and Madabhushi A. , “*A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images*,” Neurocomputing, vol. 191, pp. 214–223, 2016.
- [21] Nayak N. , Chang H. , Borowsky A. , Spellman P. , and Parvin B. , “*Classification of tumor histopathology via sparse feature learning*,” in IEEE International Symposium on Biomedical Imaging, California, USA, pp. 410–413, June, 2013.
- [22] Vu T. H. , Mousavi H. S. , Monga V. , Rao G. , and Rao U. A. , “*Histopathological image classification using discriminative feature-oriented dictionary learning*,” IEEE Transactions on Medical Imaging, vol. 35, pp. 738–751, 2016.
- [23] Cruz-Roa A. A. , Ovalle J. E. A. , Madabhushi A. , and Osorio F. A. G. , “*A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection*,” in Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention, Nagoya, Japan, pp. 403–410, Sep. 2013.
- [24] Arevalo J. , Cruz-Roa A. , Arias V. , Romero E. , and González F. A. , “*An unsupervised feature learning framework for basal cell carcinoma image analysis*,” Artificial Intelligence in Medicine, vol. 64, pp. 131–145, 2015.
- [25] Xu J. , Xiang L. , Liu Q. , Gilmore H. , Wu J. , Tang J. , and Madabhushi A. , “*Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images*,” IEEE Transactions on Medical Imaging, vol. 35, pp. 119–130, 2016.

- [26] Boureau Y.-L. , Bach F. , LeCun Y. , and Ponce J. , “*Learning mid-level features for recognition*,” in Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, California, pp.2559–2566, June, 2010.
- [27] Ma W.-Y. and Manjunath B. S. , “*Netra: A toolbox for navigating large image databases*,” Multimedia Systems, vol. 7, pp. 184–198, 1999.
- [28] Fournier J. , Cord M. , and Philipp-Foliguet S. , “*Retin: A content-based image indexing and retrieval system*,” Pattern Analysis & Applications, vol. 4, pp. 153–173, 2001.
- [29] Caicedo J. C. , Cruz A. , and Gonzalez F. A. , “*Histopathology image classification using bag of features and kernel functions*,” in Proc. of Conference on Artificial Intelligence in Medicine in Europe, Verona, Italy, pp. 126–135, July, 2009.
- [30] Linde Y. , Buzo A. , and Gray R. , “*An algorithm for vector quantizer design*,” IEEE Transactions on Communications, vol. 28, pp. 84–95, 1980.
- [31] Kohonen T. , “*Self-organized formation of topologically correct feature maps*,” Biological Cybernetics, vol. 43, pp. 59–69, 1982.
- [32] Sivic J. and Zisserman A. , “*Video google: A text retrieval approach to object matching in videos*,” in IEEE International Conference on Computer Vision, France, pp. 1470–1478,, Oct., 2003.
- [33] Csurka G. , Dance C. , Fan L. , Willamowski J. , and Bray C. , “*Visual categorization with bags of keypoints*,” in Proc. of Workshop on Statistical Learning in Computer Vision, Prague, pp. 1–2, May, 2004.
- [34] Kumar M. D. , Babaie M. , Zhu S. , Kalra S. , and Tizhoosh H. , “*A comparative study of cnn, bovw and lbp for classification of histopathological images*,” arXiv preprint arXiv:1710.01249, 2017.
- [35] Raza S. H. , Parry R. M. , Moffitt R. A. , Young A. N. , and Wang M. D. , “*An analysis of scale and rotation invariance in the bag-of-features method for histopathological image classification*,” in International Conference on Medical Image Computing and Computer-Assisted Intervention, Toronto, Canada, pp. 66–74, 2011.

- [36] Ørting S. N. , Petersen J. , Thomsen L. H. , Wille M. M. , and Bruijne M. de , “*Detecting emphysema with multiple instance learning*,” in Proc. of International Symposium on Biomedical Imaging, Washington, United States, pp. 510–513, 2018.
- [37] Lowe D. G. , “*Distinctive image features from scale-invariant keypoints*,” International Journal of Computer Vision, vol. 60, pp. 91–110, 2004.
- [38] Spanhol F. A. , Oliveira L. S. , Petitjean C. , and Heutte L. , “*A dataset for breast cancer histopathological image classification*,” IEEE Transactions on Biomedical Engineering, vol. 63, pp. 1455–1462, Jul 2016.
- [39] Leutenegger S. , Chli M. , and Siegwart R. , “*Brisk: Binary robust invariant scalable keypoints*,” in Proc. of IEEE International Conference on Computer Vision, Tokyo, Japan, pp. 2548–2555, May, 2011.
- [40] Alahi A. , Ortiz R. , and Vandergheynst P. , “*Freak: Fast retina keypoint*,” in Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Rhode Island, England, pp. 510–517, June, 2012.
- [41] Bar Y. , Diamant I. , Wolf L. , Lieberman S. , Konen E. , and Greenspan H. , “*Chest pathology detection using deep learning with non-medical training*,” in Proc. of IEEE International Symposium on Biomedical Imaging, New York, United States, pp. 294–297, 2015.
- [42] Xu Q. , Varadarajan S. , Chakrabarti C. , and Karam L. J. , “*A distributed canny edge detector: algorithm and fpga implementation*,” IEEE Transactions on Image Processing, vol. 23, pp. 2944–2960, 2014.
- [43] Cruz-Roa A. A. , Ovalle J. E. A. , Madabhushi A. , and Osorio F. A. G. , “*A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection*,” in Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention, Nagoya, Japan, pp. 403-410, 2013.
- [44] Li Y. , Wang S. , Tian Q. , and Ding X. , “*A survey of recent advances in visual feature detection*,” Neurocomputing, vol. 149, pp. 736–751, 2015.

- [45] Polakowski W. E. , Cournoyer D. A. , Rogers S. K. , DeSimio M. P. , Ruck D. W. , Hoffmeister J. W. , and Raines R. A. , “*Computer-aided breast cancer detection and diagnosis of masses using difference of gaussians and derivative-based feature saliency*,” IEEE Transactions on Medical Imaging, vol. 16, pp. 811–819, 1997.
- [46] Bay H. , Ess A. , Tuytelaars T. , and Van Gool L. , “*Speeded-up robust features (surf)*,” Computer Vision and Image Understanding, vol. 110, pp. 346–359, 2008.
- [47] Rublee E. , Rabaud V. , Konolige K. , and Bradski G. R. , “*Orb: An efficient alternative to sift or surf.*” in Proc. of IEEE International Conference on Computer Vision Workshops, Barcelona, Spain, pp. 2564–2571, 2011.
- [48] Durand T. , Thome N. , Cord M. , and Avila S. , “*Image classification using object detectors*,” in Proc. of IEEE International Conference on Image Processing, Melbourne, Australia, pp. 4340–4344, 2013.
- [49] Harris C. G. , Stephens M. *et al.*, “*A combined corner and edge detector*.” in Proc. of Alvey Vision Conference, Manchester, UK, vol. 15, pp. 10–5244, 1988.
- [50] Schmid C. , Mohr R. , and Bauckhage C. , “*Evaluation of interest point detectors*,” International Journal of Computer Vision, vol. 37, pp. 151–172, 2000.
- [51] Mikolajczyk K. and Schmid C. , “*An affine invariant interest point detector*,” in Proc. of European Conference on Computer Vision, Copenhagen, Denmark, pp. 128–142, May 2002.
- [52] Mikolajczyk K. and Schmid C. , “*A performance evaluation of local descriptors*,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, pp. 1615–1630, 2005.
- [53] Mikolajczyk K. and Schmid C. , “*Scale & affine invariant interest point detectors*,” International Journal of Computer Vision, vol. 60, pp. 63–86, 2004.
- [54] Irshad H. , Jalali S. , Roux L. , Racoceanu D. , Hwee L. J. , Le Naour G. , and Capron F. , “*Automated mitosis detection using texture, sift features and hmax biologically inspired approach*,” Journal of Pathology Informatics, vol. 4, pp. 1–7, 2013.

- [55] Raza S. H. , Parry R. M. , Sharma Y. , Chaudry Q. , Moffitt R. A. , Young A. , and Wang M. D. , “*Automated classification of renal cell carcinoma subtypes using bag-of-features*,” in Proc. of Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, Argentina, pp. 6749–6752, 2010.
- [56] Yang F. , Deng Z.-S. , and Fan Q.-H. , “*A method for fast automated microscope image stitching*,” Micron, vol. 48, pp. 17–25, 2013.
- [57] Sanghavi F. M. and Agaian S. S. , “*Automated classification of histopathology images of prostate cancer using a bag-of-words approach*,” in Proc. of Mobile Multimedia/Image Processing, Security, and Applications, Maryland, United States, pp. 98690T, 2016.
- [58] Wang C.-W. and Chen H.-C. , “*Improved image alignment method in application to x-ray images and biological images*,” Bioinformatics, vol. 29, pp. 1879–1887, 2013.
- [59] Tuzel O. , Porikli F. , and Meer P. , “*Region covariance: A fast descriptor for detection and classification*,” in Proc. of European conference on computer vision, Graz, Austria, pp. 589–600, 2006.
- [60] Calonder M. , Lepetit V. , Ozuysal M. , Trzcinski T. , Strecha C. , and Fua P. , “*Brief: Computing a local binary descriptor very fast*,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, pp. 1281–1298, 2011.
- [61] Davidson B. , Kalitzeos A. , Carroll J. , Dubra A. , Ourselin S. , Michaelides M. , and Bergeles C. , “*Fast adaptive optics scanning light ophthalmoscope retinal montaging*,” Biomedical Optics Express, vol. 9, pp. 4317–4328, 2018.
- [62] Adel D. , Mounir J. , El-Shafey M. , Eldin Y. A. , El Masry N. , AbdelRaouf A. , and Elhamid I. S. A. , “*Oral epithelial dysplasia computer aided diagnostic approach*,” in Proc. of International Conference on Computer Engineering and Systems, Cairo, Egypt, pp. 313–318, 2018.
- [63] Jurie F. and Triggs B. , “*Creating efficient codebooks for visual recognition*,” in Proc. of IEEE International Conference on Computer Vision, 2005, California, United States, pp. 604–610, 2005.

- [64] Saxena A. , Prasad M. , Gupta A. , Bharill N. , Patel O. P. , Tiwari A. , Er M. J. , Ding W. , and Lin C. T. , “*A review of clustering techniques and developments*,” Neurocomputing, vol. 267, pp. 664–681, 2017.
- [65] Chavent M. , Lechevallier Y. , and Briant O. , “*Divclus-t: A monothetic divisive hierarchical clustering method*,” Computational Statistics & Data Analysis, vol. 52, pp. 687–701, 2007.
- [66] Guénoche A. , Hansen P. , and Jaumard B. , “*Efficient algorithms for divisive hierarchical clustering with the diameter criterion*,” Journal of Classification, vol. 8, pp. 5–30, 1991.
- [67] Zhang T. , Ramakrishnan R. , and Livny M. , “*Birch: an efficient data clustering method for very large databases*,” in Proc. of ACM Sigmod Record, pp. 103 – 114, 1996.
- [68] Karypis G. , Han E.-H. , and Kumar V. , “*Chameleon: Hierarchical clustering using dynamic modeling*,” Computer, vol. 32, pp. 68–75, 1999.
- [69] Guha S. , Rastogi R. , and Shim K. , “*Cure: an efficient clustering algorithm for large databases*,” in ACM Sigmod Record, vol. 27, pp. 73–84, 1998.
- [70] Meijnen P. , Peterse J. , Antonini N. , Rutgers E. T. , and Van De Vijver M. , “*Immuno-histochemical categorisation of ductal carcinoma in situ of the breast*,” British Journal of Cancer, vol. 98, p. 137, 2008.
- [71] Bhargava R. , Fernandez D. C. , Hewitt S. M. , and Levin I. W. , “*High throughput assessment of cells and tissues: Bayesian classification of spectral metrics from infrared vibrational spectroscopic imaging data*,” Biochimica et Biophysica Acta (BBA)-Biomembranes, vol. 1758, pp. 830–845, 2006.
- [72] Pourahmad S. , Pourhashemi S. , and Mohammadianpanah M. , “*Colorectal cancer staging using three clustering methods based on preoperative clinical findings*,” Asian Pacific Journal of Cancer Prevention, vol. 17, pp. 823–827, 2016.
- [73] Xu R. and Wunsch D. C. , “*Survey of clustering algorithms*,” IEEE Transactions on Neural Networks, vol. 16, pp. 645–678, 2005.
- [74] Leibe B. and Schiele B. , “*Interleaving object categorization and segmentation*,” in *Cognitive Vision Systems*. Springer, pp. 145–161, 2006.

- [75] Georgescu B. , Shimshoni I. , and Meer P. , “*Mean shift based clustering in high dimensions: A texture classification example.*” in Proc. of International Conference on Computer Vision, Beijing, China, pp. 456–462, 2003.
- [76] Bouguettaya A. , Yu Q. , Liu X. , Zhou X. , and Song A. , “*Efficient agglomerative hierarchical clustering,*” Expert Systems with Applications, vol. 42, pp. 2785–2797, 2015.
- [77] Bezdek J. C. , “*Cluster validity with fuzzy sets,*” Journal of Cybernetics, vol. 3, pp. 58–73, 1973.
- [78] Dave R. N. and Bhaswan K. , “*Adaptive fuzzy c-shells clustering and detection of ellipses,*” IEEE Transactions on Neural Networks, vol. 3, pp. 643–662, 1992.
- [79] Yager R. R. and Filev D. P. , “*Approximate clustering via the mountain method,*” IEEE Transactions on Systems, Man, and Cybernetics, vol. 24, pp. 1279–1284, 1994.
- [80] Wang J. , Wang J. , Ke Q. , Zeng G. , and Li S. , “*Fast approximate k k-means via cluster closures,*” in *Multimedia Data Mining and Analytics.* Springer, pp.373–395, 2015.
- [81] Rueda A. , Arevalo J. , Cruz A. , Romero E. , and González F. A. , “*Bag of features for automatic classification of alzheimer’s disease in magnetic resonance images,*” in Proc. of Iberoamerican Congress on Pattern Recognition, Buenos Aires, Argentina, pp. 559–566, 2012.
- [82] Cruz-Roa A. , Caicedo J. C. , and González F. A. , “*Visual pattern mining in histology image collections using bag of features,*” Artificial Intelligence in Medicine, vol. 52, pp. 91–106, 2011.
- [83] Wiliem A. , Wong Y. , Sanderson C. , Hobson P. , Chen S. , and Lovell B. C. , “*Classification of human epithelial type 2 cell indirect immunofluorescence images via codebook based descriptors,*” in Proc. of Workshop on Applications of Computer Vision, Clearwater, Florida, pp. 95–102, 2013.
- [84] Avni U. , Greenspan H. , Konen E. , Sharon M. , and Goldberger J. , “*X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words,*” IEEE Transactions on Medical Imaging, vol. 30, pp. 733–746, 2010.

- [85] Stanciu S. G. , Xu S. , Peng Q. , Yan J. , Stanciu G. A. , Welsch R. E. , So P. T. C. , Csucs G. , and Yu H. , “*Experimenting liver fibrosis diagnostic by two photon excitation microscopy and bag-of-features image classification*,” *Scientific Reports*, vol. 4, pp. 4636–4656, Apr 2014.
- [86] Saygili A. , Uysal G. , and Bilgin G. , “*Comparative analysis of codeword representation by clustering methods for the classification of histological tissue types*,” in Proc. of International Conference on Machine Vision, Barcelona, Spain, pp. 98750U, 2015.
- [87] Zhang Y. , Huang D. , Ji M. , and Xie F. , “*Image segmentation using pso and pcm with mahalanobis distance*,” *Expert Systems with Applications*, vol. 38, pp. 9036–9040, 2011.
- [88] Wolpert D. H. and Macready W. G. , “*No free lunch theorems for optimization*,” *IEEE Transactions on Evolutionary Computation*, vol. 1, pp. 67–82, 1997.
- [89] Dixit M. , Upadhyay N. , and Silakari S. , “*An exhaustive survey on nature inspired optimization algorithms*,” *International Journal of Software Engineering and Its Applications*, vol. 9, pp. 91–104, 2015.
- [90] Maulik U. and Bandyopadhyay S. , “*Genetic algorithm-based clustering technique*,” *Pattern Recognition*, vol. 33, pp. 1455–1465, 2000.
- [91] Babu G. P. and Murty M. N. , “*Clustering with evolution strategies*,” *Pattern Recognition*, vol. 27, pp. 321–329, 1994.
- [92] Storn R. and Price K. , “*Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces*,” *Journal of Global Optimization*, vol. 11, pp. 341–359, 1997.
- [93] Simon D. , “*Biogeography-based optimization*,” *IEEE Transactions on Evolutionary Computation*, vol. 12, pp. 702–713, 2008.
- [94] Dasgupta D. and Michalewicz Z. , *Evolutionary algorithms in engineering applications*. Springer Science & Business Media, 2013.
- [95] Kennedy J. , “*Particle swarm optimization*,” in *Encyclopedia of Machine Learning*. Springer, 2011.

- [96] Dorigo M. , Birattari M. , and Stützle T. , “*Ant colony optimization*,” IEEE Computational Intelligence Magazine, vol. 1, pp. 28–39, 2006.
- [97] Rashedi E. , Nezamabadi-Pour H. , and Saryazdi S. , “*Gsa: a gravitational search algorithm*,” Information Sciences, vol. 179, pp. 2232–2248, 2009.
- [98] Shah-Hosseini H. , “*The intelligent water drops algorithm: a nature-inspired swarm-based optimization algorithm*,” International Journal of Bio-Inspired Computation, vol. 1, pp. 71–79, 2009.
- [99] Wang G.-G. , Gandomi A. H. , Alavi A. H. , and Hao G.-S. , “*Hybrid krill herd algorithm with differential evolution for global numerical optimization*,” Neural Computing and Applications, vol. 25, pp. 297–308, 2014.
- [100] Bansal J. C. , Sharma H. , Jadon S. S. , and Clerc M. , “*Spider monkey optimization algorithm for numerical optimization*,” Memetic Computing, vol. 6, pp. 31–47, 2014.
- [101] Mirjalili S. , “*The ant lion optimizer*,” Advances in Engineering Software, vol. 83, pp. 80–98, 2015.
- [102] Mirjalili S. , “*Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm*,” Knowledge-Based Systems, vol. 89, pp. 228–249, 2015.
- [103] Mirjalili S. , Mirjalili S. M. , and Hatamlou A. , “*Multi-verse optimizer: a nature-inspired algorithm for global optimization*,” Neural Computing and Applications, vol. 27, pp. 495–513, 2016.
- [104] Shah-Hosseini H. , “*Principal components analysis by the galaxy-based search algorithm: a novel metaheuristic for continuous optimisation*,” International Journal of Computational Science and Engineering, vol. 6, pp. 132–140, 2011.
- [105] Du H. , Wu X. , and Zhuang J. , “*Small-world optimization algorithm for function optimization*,” in Proc. of Springer International Conference on Natural Computation, China, pp. 264–273, 2006.
- [106] Kaveh A. and Khayatazad M. , “*A new meta-heuristic method: ray optimization*,” Computers & Structures, vol. 112, pp. 283–294, 2012.

- [107] Jose-Garcia A. and Gómez-Flores W. , “*Automatic clustering using nature-inspired metaheuristics: A survey*,” Applied Soft Computing, vol. 41, pp. 192–213, 2016.
- [108] Selim S. Z. and Alsultan K. , “*A simulated annealing algorithm for the clustering problem*,” Pattern Recognition, vol. 24, pp. 1003–1008, 1991.
- [109] Bezdek J. C. , Boggavarapu S. , Hall L. O. , and Bensaid A. , “*Genetic algorithm guided clustering*,” in Proc. of IEEE Conference on World Congress on Computational Intelligence, USA, pp. 34–39, 1994.
- [110] Langham A. E. and Grant P. , “*Using competing ant colonies to solve k-way partitioning problems with foraging and raiding strategies*,” in Proc. of Springer European Conference on Artificial Life, Switzerland, pp. 621–625, 1999.
- [111] Pal R. and Saraswat M. , “*A new bag-of-features method using biogeography-based optimization for categorization of histology images*,” International Journal of Information Systems & Management Science, vol. 1, pp. 1–6, 2018.
- [112] Mittal H. and Saraswat M. , “*Classification of histopathological images through bag-of-visual-words and gravitational search algorithm*,” in Proc. of Soft Computing for Problem Solving, pp. 231–241. Springer, 2019.
- [113] Mostafa A. , Hassanien A. E. , Houseni M. , and Hefny H. , “*Liver segmentation in mri images based on whale optimization algorithm*,” Multimedia Tools and Applications, vol. 76, pp. 24 931–24 954, 2017.
- [114] Peng X. , Wang L. , Wang X. , and Qiao Y. , “*Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice*,” Computer Vision and Image Understanding, vol. 150, pp. 109–125, 2016.
- [115] Huang Y. , Wu Z. , Wang L. , and Tan T. , “*Feature coding in image classification: A comprehensive study*,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, pp. 493–506, 2014.
- [116] Gemert J. C. van , Veenman C. J. , Smeulders A. W. M. , and Geusebroek J.-M. , “*Visual word ambiguity*,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, pp. 1271–1283, Jul 2010.

- [117] Liu L. , Wang L. , and Liu X. , “*In defense of soft-assignment coding*,” in Proc. of International Conference on Computer Vision, Tokyo, Japan, pp. 2486–2493, Nov 2011.
- [118] Huang Y. , Huang K. , Yu Y. , and Tan T. , “*Salient coding for image classification*,” in Proc. of IEEE Interantional Conference on Computer Vision and Pattern Recognition, Colorado, USA, June, pp. 1753–1760, 2011.
- [119] Tropp J. A. and Gilbert A. C. , “*Signal recovery from random measurements via orthogonal matching pursuit*,” IEEE Transactions on Information Theory, vol. 53, pp. 4655–4666, Dec 2007.
- [120] Yang J. , Yu K. , Gong Y. , Huang T. S. *et al.*, “*Linear spatial pyramid matching using sparse coding for image classification*.” in Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Florida, United States, pp. 179–1801, 2009.
- [121] Yu K. , Zhang T. , and Gong Y. , “*Nonlinear learning using local coordinate coding*,” in Proc. of Advances in neural information processing systems, Vancouver, Canada, pp. 2223–2231, 2009.
- [122] Wang J. , Yang J. , Yu K. , Lv F. , Huang T. , and Gong Y. , “*Locality-constrained linear coding for image classification*,” in Proc. of IEEE computer society conference on computer vision and pattern recognition, California, United States, pp. 3360–3367, 2010.
- [123] Yu K. and Zhang T. , “*Improved local coordinate coding using local tangents*,” in Proc. of International Conference on Machine Learning, Haifa, Israel, pp. 1–8, 2010.
- [124] Zhou X. , Yu K. , Zhang T. , and Huang T. S. , “*Image classification using super-vector coding of local image descriptors*,” in Proc. of European Conference on Computer Vision, Crete, Greece, pp. 141–154, 2010.
- [125] Perronnin F. , Sánchez J. , and Mensink T. , “*Improving the fisher kernel for large-scale image classification*,” in *Proc. of European Conference on Computer Vision, Crete, Greece*, pp.143–156. Springer Berlin Heidelberg, 2010.

- [126] Jegou H. , Perronnin F. , Douze M. , Sanchez J. , Perez P. , and Schmid C. , “*Aggregating local image descriptors into compact codes*,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, pp. 1704–1716, Sep 2012.
- [127] Bruckstein A. M. , Donoho D. L. , and Elad M. , “*From sparse solutions of systems of equations to sparse modeling of signals and images*,” SIAM Review, vol. 51, pp. 34–81, 2009.
- [128] Lee H. , Battle A. , Raina R. , and Ng A. Y. , “*Efficient sparse coding algorithms*,” in Proc. of Advances in neural information processing systems, Vancouver, Canada, pp. 801–808, 2007.
- [129] Jaakkola T. and Haussler D. , “*Exploiting generative models in discriminative classifiers*,” in Proc. of Advances in Neural Information Processing Systems, Massachusetts, pp. 487–493, 1999.
- [130] Zhou Y. , Chang H. , Barner K. , Spellman P. , and Parvin B. , “*Classification of histology sections via multispectral convolutional sparse coding*,” in Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Columbus, Ohio, pp. 3081–3088, Jun 2014.
- [131] Nowaková J. , Prílepková M. , and Snášel V. , “*Medical image retrieval using vector quantization and fuzzy s-tree*,” Journal of Medical Systems, vol. 42, pp. 1–18, Apr 2018.
- [132] Diamant I. , Klang E. , Amitai M. , Konen E. , Goldberger J. , and Greenspan H. , “*Task-driven dictionary learning based on mutual information for medical image classification*,” IEEE Transactions on Biomedical Engineering, vol. 64, pp. 1380–1392, 2017.
- [133] Han H. , Li L. , Han F. , Song B. , Moore W. , and Liang Z. , “*Fast and adaptive detection of pulmonary nodules in thoracic CT images using a hierarchical vector quantization scheme*,” IEEE Journal of Biomedical and Health Informatics, vol. 19, pp. 648–659, Mar 2015.
- [134] Dieterle F. , Musller-Hagedorn S. , Liebich H. M. , and Gauglitz G. , “*Urinary nucleosides as potential tumor markers evaluated by learning vector quantization*,” Artificial Intelligence in Medicine, vol. 28, pp. 265–279, July, 2003.

- [135] Mattfeldt T. , Trijic D. , Gottfried H.-W. , and Kestler H. A. , “*Classification of incidental carcinoma of the prostate using learning vector quantization and support vector machines*,” *Analytical Cellular Pathology*, vol. 26, pp. 45–55, 2004.
- [136] Balis U. , Hipp J. , Cheng J. , Toner M. , and Tompkins R. , “*Spatially invariant vector quantization: A pattern matching algorithm for multiple classes of image subject matter including pathology*,” *Journal of Pathology Informatics*, vol. 2, p. 13, 2011.
- [137] Sommer C. , Fiaschi L. , Hamprecht F. A. , and Gerlich D. W. , “*Learning-based mitotic cell detection in histopathological images*,” in Proc. of the International Conference on Pattern Recognition, Tsukuba, Japan, pp. 2306–2309, 2012.
- [138] Cruz-Roa A. , Basavanhally A. , González F. , Gilmore H. , Feldman M. , Ganesan S. , Shih N. , Tomaszewski J. , and Madabhushi A. , “*Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks*,” in Proc. of SPIE, Medical Imaging 2014: Digital Pathology, California, USA, pp. 1-15, Mar, 2014.
- [139] Hou L. , Samaras D. , Kurc T. M. , Gao Y. , Davis J. E. , and Saltz J. H. , “*Patch-based convolutional neural network for whole slide tissue image classification*,” in Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Nevada, United States, pp. 2424–2433, June, 2016.
- [140] Wang H. , Cruz-Roa A. , Basavanhally A. , Gilmore H. , Shih N. , Feldman M. , Tomaszewski J. , Gonzalez F. , and Madabhushi A. , “*Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features*,” *Journal of Medical Imaging*, vol. 1, p. 034003, Oct 2014.
- [141] Wang S. , Lu S. , Dong Z. , Yang J. , Yang M. , and Zhang Y. , “*Dual-tree complex wavelet transform and twin support vector machine for pathological brain detection*,” *Applied Sciences*, vol. 6, p. 169, June 2016.
- [142] Zhang Y.-D. , Chen S. , Wang S.-H. , Yang J.-F. , and Phillips P. , “*Magnetic resonance brain image classification based on weighted-type fractional fourier transform and nonparallel support vector machine*,” *International Journal of Imaging Systems and Technology*, vol. 25, pp. 317–327, Nov. 2015.

- [143] Doyle S. , Agner S. , Madabhushi A. , Feldman M. , and Tomaszewski J. , “*Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features*,” in IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Paris, France, pp. 496–499, May, 2008.
- [144] Sertel O. , Kong J. , Catalyurek U. V. , Lozanski G. , Saltz J. H. , and Gurcan M. N. , “*Histopathological image analysis using model-based intermediate representations and color texture: Follicular lymphoma grading*,” Journal of Signal Processing Systems, vol. 55, pp. 169–183, May 2008.
- [145] Sertel O. , Kong J. , Lozanski G. , Shanaah A. , Catalyurek U. , Saltz J. , and Gurcan M. , “*Texture classification using nonlinear color quantization: Application to histopathological image analysis*,” in Proc. of International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, pp. 597–600, Mar 2008.
- [146] Esgiar A. N. , Naguib R. , Sharif B. , Bennett M. , and Murray A. , “*Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa*,” IEEE Transactions on Information Technology in Biomedicine, vol. 2, pp. 197–203, 1998.
- [147] Huang P.-W. and Lee C.-H. , “*Automatic classification for pathological prostate images based on fractal analysis*,” IEEE Transactions on Medical Imaging, vol. 28, pp. 1037–1050, Jul 2009.
- [148] Zhou X. , Wang S. , Xu W. , Ji G. , Phillips P. , Sun P. , and Zhang Y. , “*Detection of pathological brain in MRI scanning based on wavelet-entropy and naive bayes classifier*,” in *Bioinformatics and Biomedical Engineering*. Springer International Publishing, pp. 201–209, 2015.
- [149] Naik S. , Doyle S. , Agner S. , Madabhushi A. , Feldman M. , and Tomaszewski J. , “*Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology*,” in Proc. of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Paris, France, pp. 284-287, May 2008.

- [150] Doyle S. , Feldman M. , Tomaszewski J. , and Madabhushi A. , “*A boosted bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies*,” IEEE Transactions on Biomedical Engineering, vol. 59, pp. 1205–1218, May 2012.
- [151] Dalal N. and Triggs B. , “*Histograms of oriented gradients for human detection*,” in Proc. of International Conference on Computer Vision & Pattern Recognition, California, United States, pp. 886–893, vol. 1, 2005.
- [152] Brighton H. and Mellish C. , “*Advances in instance selection for instance-based learning algorithms*,” Data Mining and Knowledge Discovery, vol. 6, pp. 153–172, 2002.
- [153] Dorkó G. and Schmid C. , “*Selection of scale-invariant parts for object class recognition*.” in Proc. of International Conference on Computer Vision, Beijing, China, pp. 634–640, 2003.
- [154] Aha D. W. , Kibler D. , and Albert M. K. , “*Instance-based learning algorithms*,” Machine Learning, vol. 6, pp. 37–66, 1991.
- [155] Wilson D. R. and Martinez T. R. , “*Reduction techniques for instance-based learning algorithms*,” Machine Learning, vol. 38, pp. 257–286, 2000.
- [156] Lin W.-C. , Tsai C.-F. , Chen Z.-Y. , and Ke S.-W. , “*Keypoint selection for efficient bag-of-words feature generation and effective image classification*,” Information Sciences, vol. 329, pp. 33–51, 2016.
- [157] Albiol A. , Monzo D. , Martin A. , Sastre J. , and Albiol A. , “*Face recognition using hog-ebgm*,” Pattern Recognition Letters, vol. 29, pp. 1537–1543, 2008.
- [158] Xu Z. G. , Chen C. , and Liu X. H. , “*An efficient view-point invariant detector and descriptor*,” in Advanced Materials Research, vol. 659. Trans Tech Publ, 2013, pp. 143–148.
- [159] Juan L. and Gwun O. , “*A comparison of sift, pca-sift and surf*,” International Journal of Image Processing, vol. 3, pp. 143–152, 2009.
- [160] Chang K. , Lee R. , Wen C. , and Yeh M. , “*Comparison of similarity measures for clustering electrocardiogram complexes*,” in Proc. of Conference on Computers in Cardiology, Lyon, France, pp. 759–762, 2005.

- [161] Julong D. , “*Introduction to grey system theory*,” The Journal of Grey System, vol. 1, pp. 1–24, 1989.
- [162] Sallehuddin R. , Shamsuddin S. M. H. , and Hashim S. Z. M. , “*Application of grey relational analysis for multivariate time series*,” in Proc. of International Conference on Intelligent Systems Design and Applications, Kaohsiung, Taiwan, pp. 432–437, 2008.
- [163] Wang C. , Chen S.-F. , and Yuen M. M. F. , “*Fuzzy part family formation based on grey relational analysis*,” The International Journal of Advanced Manufacturing Technology, vol. 18, pp. 128–132, 2001.
- [164] “*Blue histology - epithelia and glands*,” <http://www.lab.anhb.uwa.edu.au/mb140/corepages/epithelia/epithel.htm>, (Accessed on 13/04/2018).
- [165] Eurell J. A. and Frappier B. L. , *Dellmann’s textbook of veterinary histology*. John Wiley & Sons, 2013.
- [166] Ali S. and Madabhushi A. , “*An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery*,” IEEE Transactions on Medical Imaging, vol. 31, pp. 1448–1460, 2012.
- [167] Jiang Y.-G. , Yang J. , Ngo C.-W. , and Hauptmann A. G. , “*Representations of keypoint-based semantic concept detection: A comprehensive study*,” IEEE Transactions on Multimedia, vol. 12, pp. 42–53, 2010.
- [168] Shamir L. , Orlov N. , Eckley D. M. , Macura T. , Johnston J. , and Goldberg I. G. , “*Wnfchrm—an open source utility for biological image analysis*,” Source Code for Biology and Medicine, vol. 3, p. 13, 2008.
- [169] Wright J. , Yang A. Y. , Ganesh A. , Sastry S. S. , and Ma Y. , “*Robust face recognition via sparse representation*,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, pp. 210–227, 2009.
- [170] Bong C. and Rajeswari M. , “*Multiobjective clustering with metaheuristic: current trends and methods in image segmentation*,” IET Image Processing, vol. 6, pp. 1–10, 2012.

- [171] Ahmed H. , Shedeed H. A. , Hamad S. , and Tolba M. F. , “*On combining nature-inspired algorithms for data clustering*,” in *Proc. of Handbook of Research on Machine Learning Innovations and Trends*, pp. 826–855. IGI Global, 2017.
- [172] Ma H. , Simon D. , Siarry P. , Yang Z. , and Fei M. , “*Biogeography-based optimization: a 10-year review*,” *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 1, pp. 391–407, 2017.
- [173] Lim W. L. , Wibowo A. , Desa M. I. , and Haron H. , “*A biogeography-based optimization algorithm hybridized with tabu search for the quadratic assignment problem*,” *Computational Intelligence and Neuroscience*, vol. 2016, p. 27, 2016.
- [174] Kingsland S. , “*The theory of island biogeography*,” *Journal of the History of Biology*, vol. 35, pp. 178–179, 2002.
- [175] Ma H. , “*An analysis of the equilibrium of migration models for biogeography-based optimization*,” *Information Sciences*, vol. 180, pp. 3444–3464, 2010.
- [176] Pal R. and Saraswat M. , “*Improved biogeography-based optimization*,” *International Journal of Advanced Intelligence Paradigms*, vol. (In Press), 2017.
- [177] Du D. , Simon D. , and Ergezer M. , “*Biogeography-based optimization combined with evolutionary strategy and immigration refusal*,” in *Proc. of IEEE International Conference on Systems, Man and Cybernetics*, Texas, USA, pp. 997–1002, 2009.
- [178] Gong W. , Cai Z. , Ling C. X. , and Li H. , “*A real-coded biogeography-based optimization with mutation*,” *Applied Mathematics and Computation*, vol. 216, pp. 2749–2758, 2010.
- [179] Ma H. and Simon D. , “*Blended biogeography-based optimization for constrained optimization*,” *Engineering Applications of Artificial Intelligence*, vol. 24, pp. 517–525, 2011.
- [180] Lohokare M. , Pattnaik S. S. , Panigrahi B. K. , and Das S. , “*Accelerated biogeography-based optimization with neighborhood search for optimization*,” *Applied Soft Computing*, vol. 13, pp. 2318–2342, 2013.

- [181] Ma H. , Simon D. , Fei M. , and Xie Z. , “*Variations of biogeography-based optimization and markov analysis*,” *Information Sciences*, vol. 220, pp. 492–506, 2013.
- [182] Simon D. , Omran M. G. , and Clerc M. , “*Linearized biogeography-based optimization with re-initialization and local search*,” *Information Sciences*, vol. 267, pp. 140–157, 2014.
- [183] Gong W. , Cai Z. , and Ling C. X. , “*De/bbo: a hybrid differential evolution with biogeography-based optimization for global numerical optimization*,” *Soft Computing*, vol. 15, pp. 645–665, 2010.
- [184] Niu Q. , Zhang L. , and Li K. , “*A biogeography-based optimization algorithm with mutation strategies for model parameter estimation of solar and fuel cells*,” *Energy Conversion and Management*, vol. 86, pp. 1173–1185, Oct 2014.
- [185] Al-Roomi A. R. and El-Hawary M. E. , “*Metropolis biogeography-based optimization*,” *Information Sciences*, vol. 360, pp. 73–95, 2016.
- [186] Garg V. and Deep K. , “*Performance of laplacian biogeography-based optimization algorithm on CEC 2014 continuous optimization benchmarks and camera calibration problem*,” *Swarm and Evolutionary Computation*, vol. 27, pp. 132–144, Apr 2016.
- [187] Saremi S. , Mirjalili S. , and Lewis A. , “*Biogeography-based optimisation with chaos*,” *Neural Computing and Applications*, vol. 25, pp. 1077–1097, 2014.
- [188] Feng Q. , Liu S. , Zhang J. , Yang G. , and Yong L. , “*Biogeography-based optimization with improved migration operator and self-adaptive clear duplicate operator*,” *Applied intelligence*, vol. 41, pp. 563–581, 2014.
- [189] Zheng Y.-J. , Ling H.-F. , Wu X.-B. , and Xue J.-Y. , “*Localized biogeography-based optimization*,” *Soft Computing*, vol. 18, pp. 2323–2334, 2014.
- [190] Guo W. , Wang L. , Ge S. S. , Ren H. , and Mao Y. , “*Drift analysis of mutation operations for biogeography-based optimization*,” *Soft Computing*, vol. 19, pp. 1881–1892, 2015.
- [191] Bansal N. , Kumar S. , and Tripathi A. , “*Application of artificial bee colony algorithm using hadoop*,” in Proc. of International Conference on Computing for Sustainable Global Development, Delhi, India, pp. 3615–3619, 2016.

- [192] Gong W. , Cai Z. , and Ling C. X. , “*De/bbo: a hybrid differential evolution with biogeography-based optimization for global numerical optimization*,” Soft Computing, vol. 15, pp. 645–665, 2010.
- [193] Chen X. , Tianfield H. , Du W. , and Liu G. , “*Biogeography-based optimization with covariance matrix based migration*,” Applied Soft Computing, vol. 45, pp. 71–85, 2016.
- [194] Mahoney M. S. , *The mathematical career of Pierre de Fermat, 1601-1665*. Princeton University Press, 1994.
- [195] Simon D. , *Evolutionary optimization algorithms*. John Wiley & Sons, pp. 1–772, 2013.
- [196] Jamil M. and Yang X.-S. , “*A literature survey of benchmark functions for global optimisation problems*,” International Journal of Mathematical Modelling and Numerical Optimisation, vol. 4, pp. 150–194, 2013.
- [197] Wu G. , Mallipeddi R. , and Suganthan P. , “*Problem definitions and evaluation criteria for the cec 2017 competition on constrained real-parameter optimization*,” National University of Defense Technology, Changsha, Hunan, PR China and Kyungpook National University, Daegu, South Korea and Nanyang Technological University, Singapore, Technical Report, 2016.
- [198] Brest J. , Maučec M. S. , and Bošković B. , “*Single objective real-parameter optimization: Algorithm jso*,” in Proc. of IEEE congress on evolutionary computation, San Sebastian, Spain, pp. 1311–1318, June, 2017.
- [199] Mohamed A. W. , Hadi A. A. , Fattouh A. M. , and Jambi K. M. , “*LSHADE with semi-parameter adaptation hybrid with CMA-ES for solving CEC 2017 benchmark problems*,” in Proc. of IEEE Congress on Evolutionary Computation, Saint Sebastian, Spain, pp. 145-152, June, 2017.
- [200] Mirjalili S. , Gandomi A. H. , Mirjalili S. Z. , Saremi S. , Faris H. , and Mirjalili S. M. , “*Saltp swarm algorithm: A bio-inspired optimizer for engineering design problems*,” Advances in Engineering Software, vol. 114, pp. 163–191, 2017.
- [201] Mirjalili S. , Mirjalili S. M. , and Lewis A. , “*Grey wolf optimizer*,” Advances in Engineering Software, vol. 69, pp. 46–61, 2014.

- [202] Mittal H. , Pal R. , Kulhari A. , and Saraswat M. , “*Chaotic kbest gravitational search algorithm (ckgsa)*,” in Proc. of International Conference on Contemporary Computing, Noida, India, pp. 1–6, 2016.
- [203] Mirjalili S. and Lewis A. , “*The whale optimization algorithm*,” Advances in Engineering Software, vol. 95, pp. 51–67, 2016.
- [204] Theodorsson-Norheim E. , “*Friedman and quade tests: Basic computer program to perform nonparametric two-way analysis of variance and multiple comparisons on ranks of several related samples*,” Computers in Biology and Medicine, vol. 17, pp. 85–99, 1987.
- [205] Chapelle O. , Haffner P. , and Vapnik V. N. , “*Support vector machines for histogram-based image classification*,” IEEE Transactions on Neural Networks, vol. 10, pp. 1055–1064, 1999.
- [206] Pakhira M. K. , Bandyopadhyay S. , and Maulik U. , “*Validity index for crisp and fuzzy clusters*,” Pattern Recognition, vol. 37, pp. 487–501, 2004.
- [207] Wagner A. , Wright J. , Ganesh A. , Zhou Z. , Mobahi H. , and Ma Y. , “*Toward a practical face recognition system: Robust alignment and illumination by sparse representation*,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, pp. 372–386, 2012.
- [208] Perronnin F. and Dance C. , “*Fisher kernels on visual vocabularies for image categorization*,” in Proc. on IEEE conference on computer vision and pattern recognition, Minneapolis, United States, pp. 1–8, 2007.
- [209] Van Gemert J. C. , Veenman C. J. , Smeulders A. W. , and Geusebroek J.-M. , “*Visual word ambiguity*,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, pp. 1271–1283, 2010.
- [210] Yang J. , Yu K. , and Huang T. , “*Efficient highly over-complete sparse coding using a mixture model*,” in Proc. of European Conference on Computer Vision, Crete, Greece, pp. 113–126, 2010.

- [211] Gao S. , Tsang I. W.-H. , Chia L.-T. , and Zhao P. , “*Local features are not lonely–laplacian sparse coding for image classification*,” in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, California, USA, pp. 3555–3561, 2010.
- [212] Picard D. and Gosselin P.-H. , “*Improving image similarity with vectors of locally aggregated tensors*,” in Proc. of IEEE International Conference on Image Processing, Brussels, Belgium, pp. 669–672, 2011.
- [213] Zhao X. , Yu Y. , Huang Y. , Huang K. , and Tan T. , “*Feature coding via vector difference for image classification*,” in Proc. of IEEE International Conference on Image Processing, Vienna, Austria, pp. 3121–3124, Apr. 2012.
- [214] Hacioglu K. and Ward W. , “*Question classification with support vector machines and error correcting codes*,” in Proc. of the Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, USA, pp. 28–30, 2003.
- [215] Panchal P. , Panchal S. , and Shah S. , “*A comparison of sift and surf*,” International Journal of Innovative Research in Computer and Communication Engineering, vol. 1, pp. 323–327, 2013.
- [216] Işık Ş. , “*A comparative evaluation of well-known feature detectors and descriptors*,” International Journal of Applied Mathematics, Electronics and Computers, vol. 3, pp. 1–6, 2014.

LIST OF PUBLICATIONS

- **Published**

- International Journals

1. R. Pal and M. Saraswat, “Histopathological image classification using enhanced bag-of-Feature with spiral biogeography-based optimization”, *Applied Intelligence*, 2019. (Online available, doi: <https://doi.org/10.1007/s10489-019-01460-1>)
2. R. Pal and M. Saraswat, “Grey relational analysis based keypoint selection in bag-of-Features for histopathological image classification”, *Recent Patents on Computer Science*, vol. 12, pp. 1-6, 2018.
3. R. Pal and M. Saraswat, “Improved biogeography-based optimization”, *International Journal of Advanced Intelligence Paradigms*, 2017. (In press)

- International Conferences

1. R. Pal and M. Saraswat, “A new bag-of-feature method using biogeography-based optimization for categorization of histology images”, in proc. of *International Conference on Computers and Management (ICCM)*, Delhi, India, pp. 155-160, 2018.
2. R. Pal and M. Saraswat, “Enhanced bag-of-features using AlexNet and improved biogeography-based optimization for histopathological image analysis”, in proc. of *IEEE International Conference on Contemporary Computing (IC3)*, Noida, India, pp. 1-6, 2018.
3. R. Pal and M. Saraswat, “Data clustering using enhanced biogeography-based optimization”, in proc. of *IEEE International Conference on Contemporary Computing (IC3)*, Noida, India, pp. 1-6, 2017.

• **Under Review**

1. R. Pal and M. Saraswat, “Efficient Bag-Of-Features using grey relational analysis for Histopathology Image classification”, *Computers in Biology and Medicine*.
2. R. Pal and M. Saraswat, “A New Weighted Two Dimensional Vector Quantization Encoding Method in Bag-of-Features for Histopathological Image Classification”, *International Journal of Intelligent Information and Database Systems*.

