Chapter 7

Classification of Cancer Images using CoM-based Feature Descriptor

The preceding chapters primarily concentrated on deep learning-based approaches for breast cancer classification of HBC and CMT. They have exceptional representation learning capabilities, but they require high compute resources and huge data to perform well. Therefore, this chapter addresses this limitation while also attempting to overcome the limitations of the machine learning approach in terms of incapability to handle uncertainty as well as sophisticated feature learning.

7.1 Introduction

Recent advancements in the digitization of histological images have made CAD possible, which can reduce the burden on pathologists by filtering out visibly benign areas to achieve better focus in difficult-to-diagnose cases. Correct diagnosis and treatment at early stages are crucial for the successful outcome of treatment modalities. Timely and accurate diagnosis of breast cancer in humans and mammary tumours in dogs seeks the

attention of researchers in AI. Recently, deep learning-based algorithms, particularly CNNs, have achieved remarkable success in the field of medical image analysis. In our previous study [216], we also utilized its exceptional representation ability by applying deep learning-based methods to achieve state-of-the-art accuracy for binary classification of HBC and CMT. The VGGNET16-based fused framework proposed in this study was able to differentiate between benign and malignant neoplasms with a classification accuracy of 97% for HBC and 93% for CMT. The classification accuracy achieved using the same framework in CMT was found to be less than in HBC, despite the strong similarities in histomorphological features. This might have been due to the small size of the CMTHis dataset, reflecting the reduced performance of deep learning-based approaches with smaller datasets. A major limiting factor for deep learning-based approaches is the requirement for huge amounts of labelled data. Although, in the field of medical imaging, acquiring a huge labelled dataset is difficult and expensive and becomes more complex in the case of rare diseases at an early stage.

Machine learning-based algorithms are a potential alternative to deep learning for medical imaging with smaller datasets. A major constraint of conventional machine learning techniques is the requirement of complex processing for the extraction of discriminatory features. Nevertheless, recent advancements in the field have gone some way towards addressing this problem. The application of feature extraction and selection simplifies the model by reducing training and execution times, input requirements, and computational cost. Furthermore, it not only improves the model compactness and transparency by removing insignificant features from the dataset but also facilitates interpretation. Therefore, feature extraction represents a critical challenge for automated cancer histopathology analysis as the performance of the machine learning model is highly influenced by feature quality. Several handcrafted feature descriptors are utilized in histopathological image analysis. Despite the significant performance of machine learning-based algorithms in histopathological image analysis, colour varia-

tion due to differences in staining protocols and illumination remains a serious concern. These variations significantly affect automated image analysis and interpretation, and various algorithms have been introduced recently to address them. Though various feature extraction modalities have been explored for cancer histopathological image analysis, to date, none have proved entirely convincing. Moreover, local feature descriptors like LBP are sensitive to uncertainty in the grey value of images. In addition, the performance of machine learning techniques is susceptible to uncertainty in the data. Now the concept of fuzzy learning is well established to address uncertainties in raw data and is often considered to solve problems associated with image processing. It also helps in designing robust and fault-tolerant classification methods that can help to deal with imprecise and uncertain data. Thus, the observations made from our studies, reported in previous chapters, directed us to find the alternative approach which is suitable for limited datasets with good representation learning ability. Noting that solutions to some of the most challenging problems of data science may lie in natural phenomena and the laws of science, the proposed framework exploits the potential of the popular concept of Center of Mass ("CoM") taken from physics. Having stated the issues of feature extraction in medical image analysis with limited and uncertain data, in this study, we propose a novel framework called Center of Mass Histopathology Pattern (CoMHisP), which integrates theoretical concepts of CoM and fuzzy theory with machine learning for feature extraction from histopathological images and effective classification.

7.2 Motivation and Significant Contribution

Feature extraction is one of the most important tasks in machine learning application for cancer image analysis due to the diversity of histological and morphological features. In histopathology, minor variations in the H&E staining protocols used by different laboratories can affect the classification accuracies of automated systems. Resolving this

issue requires stain normalization, and selecting an algorithm for stain normalization is a crucial task. Yet another variable in visible light images is illumination, as even slight differences in illumination conditions can affect algorithmic outcomes. Many studies have described the potential of local patterns and their variants, and this has inspired us to apply the illumination invariant descriptor for feature extraction. An illumination invariant descriptor based on CoM has also been used for face recognition without using any machine learning technique [235]. Recently, the deep learning framework has been widely applied in a variety of applications with high success rates. The reason for this success is the use of different filter sizes and numbers of layers. However, deep learning requires a large amount of data, and there is rarely enough labelled data available in the medical field. Hence, we also need to focus on the challenge of making valuable decisions with limited data using machine learning techniques. Since microscopic CMT images consist of multiple polymorphic nuclei and cells, which are variable in size and staining, we need to use a filter of optimal size so that we can capture valuable information in each pixel without losing its structural characteristics. We are therefore motivated to select a block size within an image that is neither too small nor too large, similar to the filters used by the deep learning framework. To address these issues, we propose a novel framework for automated cancer image analysis.

The major contributions of this chapter are summarized as follows:

1. A novel framework utilizing CoM-based illumination invariant feature extraction technique, which is used to compute angular features by considering pixel's neighbourhood, along with machine learning, is proposed for cancer diagnosis.

2. A new approach is designed to reduce block size of 4×4 to 3×3 for feature extraction. This reduction is somewhat similar to average pooling and contributes to enhanced performance.

3. To handle the uncertainty in data, we also utilize the concept of fuzzy theory with

SVM for classification, which presents an efficient and robust model.

4. The work is validated using novel histopathology data of canine tumour tissues, collected and interpreted under the supervision of experts at ICAR-Indian Veterinary Research Institute, Izatnagar, India.

5. To the best of our knowledge, this is the first work in the field to integrate illumination invariant feature extraction and fuzzy theory for CMT classification. Thus, this work presents a novel interdisciplinary approach to solve one of the most challenging problems of histopathological image interpretation in the veterinary and medical sciences.

6. To verify the proposed model, extensive experiments are performed and validated against different performance evaluation metrics.

7.3 Theoretical Background

In this section, basic definitions and preliminaries are discussed to aid understanding of the proposed framework.

7.3.1 Basic definitions

Definition 7.1 *(Fuzzy Sets): If A is a set of features denoted by $\boldsymbol{x}$, then a fuzzy set F in A is defined as an asset of ordered pairs:*

$$F = \{(x, \mu_F(x))|x \in A\} \tag{7.1}$$

where, $\mu_F(x)$ is a membership function (MF) for the fuzzy set F, which maps the element of A between 0 and 1 [236].

Definition 7.2 *Center of Mass: The center of mass (CoM) of an object O is defined as a point with position vector $r_{cm}^{\rightarrow}$ at which the object is balanced. Where there is a rigid*

body system of particles, it is the unique point at which the mass of the entire system is concentrated, and the weighted relative position of the mass distribution sums to 0*. The terms "center of gravity" and "center of mass" are used synonymously. Mathematically, it is defined for a rigid body system of n particles with mass* $\{m_1, m_2, \cdots, m_n\}$ *and position vector* $\{r_1, r_2, \cdots, r_n\}$ *as:*

$$\begin{aligned}(m_1 + m_2 + \cdots + m_n)\vec{r_{cm}} &= (m_1\vec{r_1} + m_2\vec{r_2} + \cdots + m_n\vec{r_n}) \\ \vec{r_{cm}} &= \frac{\sum_i^n m_i\vec{r_i}}{\sum_i^n m_i}\end{aligned} \tag{7.2}$$

7.3.2 Local patterns

A local pattern, also referred to as a micro-pattern, considers the variation of the sub-region in an image and efficiently presents information on the spatial structure in the form of discriminatory and robust features. There are numerous approaches, including LBP [54], fuzzy LBP [237], local discriminant direction binary pattern [238], high-order local ternary pattern [239], local tetra pattern [240], and many more [241]. Some of these are described below.

7.3.2.1 Local binary pattern (LBP)

The LBP operator is an invariant texture measure that is computationally efficient and has a high discrimination capability. The LBP value is calculated for a given center pixel x_c and its P number of neighbor pixels with a neighborhood radius R and a threshold value of T as follows:

$$LBP_{P,R}(x_c) = g_1(I(x_{p,r}), I(x_c)) \cdot 2^{p-1}|_{\{p=1,\cdots,P;r=1,\cdots,R\}} \tag{7.3}$$

$$g_1(I(x_{p,r}), I(x_c)) = \begin{cases} 1 & if(I(x_{p,r}) - I(x_c)) \geq T \\ 0 & else \end{cases} \tag{7.4}$$

7.3.2.2 Fuzzy LBP

LBP has some limitations such as gray-level uncertainty in digital images due to hard thresholding. To overcome these limitations, Keramidas et al. [237] proposed fuzzy LBP. For this two fuzzy sets $\tilde{V}$ and $\tilde{B}$ are defined, along with MFs $\mu_{\tilde{V}}(x)$ and $\mu_{\tilde{B}}(x)$ respectively. Here V is the set of all pixels p_x with gray value $g_x \in [g_c, g_{max}]$ and remaining pixels p_x with $g_x \in [0, g_c]$. Here, g_c is a reference gray value in the neighborhood pixels and $N = \{0, 1, 2, \cdots, n-1\}$ is defined as n neighborhood pixels. This can be illustrated by the following equations:

$$\tilde{V} \equiv \{< p_x, \mu_{\tilde{V}}(x) > |\ x \in N\} \tag{7.5}$$

$$\tilde{B} \equiv \{< p_x, \mu_{\tilde{B}}(x) > |\ x \in N\} \tag{7.6}$$

$$\mu_{\tilde{V}}(x) = \begin{cases} 1, & if g_x - g_c \leq T \\ \frac{T+g_x-g_c}{2T} & if\ |g_x - g_c| < T, T \neq 0 \\ 0, & if g_x - g_c \leq -T, T \neq 0 \\ 0, & if g_x - g_c < T, T \neq 0 \end{cases} \tag{7.7}$$

Similarly, MF $\mu_{\tilde{B}}$ for fuzzy set $\tilde{B}$ is:

$$\mu_{\tilde{B}}(x) = 1 - \mu_{\tilde{V}}(x) \tag{7.8}$$

Besides these MFs for fuzzy LBP, one more binary pattern code, termed as a contribution value C^{BP} is characterized in each neighbourhood pixel as follows:

$$C^{BP} = \prod_{X=0; z \in \{0,1\}}^{n-1} \mu_z(x) \tag{7.9}$$

$$C^{total} = \sum_{lbp=0}^{2^{n-1}} C^{BP} = 1 \tag{7.10}$$

However, fuzzy LBP too has disadvantages, for example manual setting of threshold

value and high computational cost.

7.3.3 Theoretical analysis of center of mass

A rigid body is viewed as an agglomeration of particles in which the distance between each pair of particles remains invariant and where the particles are held together by cohesive forces such that action and reaction between any two particles are equal but working in opposite directions. Introducing CoM simplifies the problem by representing a collection of mass as one at their common center of mass. CoM is defined mathematically in definition 7.2. Any external force applied to CoM will cause the body to move in the direction of force without any rotational movement or torque. Moreover, if it undergoes any displacement, the particles within it will experience similar displacement. A few examples are shown in Fig. 7.1.

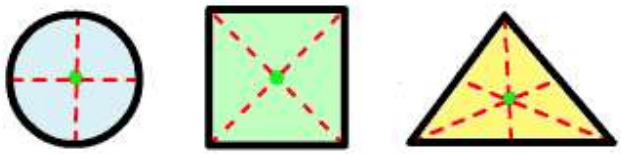


Figure 7.1: Green dot represents center of mass of each shape.

It is helpful to consider the two-particle system (i.e. discrete mass) with mass m_1 and m_2 in a 2-dimensional space, in a static position with reference to origin O and position vectors $\vec{r_1} \equiv (x_1, y_1)$ and $\vec{r_2} \equiv (x_2, y_2)$, respectively. Then, the complete system can be replaced by a single point with position vector $\vec{r_{cm}} \equiv (x_{cm}, y_{cm})$, around which the whole distribution of mass is balanced. Further, CoM can be derived by considering the total force $\vec{F^1}$ and $\vec{F^2}$ acting on m_1 and m_2, respectively. $\vec{F^{12}}$ and $\vec{F^{21}}$ are internal forces acting on m_1 due to m_2 and vice versa. $\vec{F^{ext1}}$ and $\vec{F^{ext2}}$ are external forces experienced respectively by these two particles. We can then derive the position of

CoM as follows:

$$\vec{F^1} = \vec{F^{ext1}} + \vec{F^{12}} \tag{7.11}$$

$$\vec{F^2} = \vec{F^{ext2}} + \vec{F^{21}} \tag{7.12}$$

As per Newton's 3^{rd} law of motion within the system of particles, we get:

$$\vec{F^{12}} = -\vec{F^{21}} \Rightarrow \vec{F^{12}} + \vec{F^{21}} = 0 \tag{7.13}$$

Then, the total force acting on the system $\vec{F}$ is computed using (7.11) to (7.13) as:

$$\vec{F} = \vec{F^1} + \vec{F^2} \tag{7.14}$$

$$\vec{F} = \vec{F^{ext1}} + \vec{F^{ext2}} \tag{7.15}$$

$$\vec{F} = \frac{d}{dt}\left(m_1 \cdot \frac{d\vec{r_1}}{dt} + m_2 \cdot \frac{d\vec{r_2}}{dt}\right) \tag{7.16}$$

$$\therefore \vec{F} = \frac{d^2}{dt^2}\left(m_1 \cdot \vec{r_1} + m_2 \cdot \vec{r_2}\right) \tag{7.17}$$

Multiplying total mass of a system $(m_1 + m_2)$ in numerator and denominator in the right side of (7.17), we get:

$$(m_1 + m_2)\frac{d^2\vec{r_{cm}}}{dt^2} = (m_1 + m_2)\frac{d^2}{dt^2}\left(\frac{m_1 \cdot \vec{r_1} + m_2 \cdot \vec{r_2}}{m_1 + m_2}\right) \tag{7.18}$$

$$\frac{d^2\vec{r_{cm}}}{dt^2} = \frac{d^2}{dt^2}\left(\frac{m_1 \cdot \vec{r_1} + m_2 \cdot \vec{r_2}}{m_1 + m_2}\right) \tag{7.19}$$

$$\vec{r_{cm}} = \frac{m_1 \cdot \vec{r_1} + m_2 \cdot \vec{r_2}}{m_1 + m_2} \tag{7.20}$$

It is to be noted that if the system of particles is in a stationary state, then only the internal force will act on CoM, with position vector $\vec{r_{cm}}$, and the resultant force will be directed towards CoM. Further, on the basis of the proposition discussed in

[242], each pixel of the image can be treated as a charged particle in space. Thus, we can extend the concept of CoM for multiple particle systems in a stationary state for feature extractions in histopathology images where each individual pixel acts as a particle mass. The detailed description of feature extraction is presented in Section 7.4.

7.3.4 $l_1 - norm$ computation

The length of a vector is calculated using $l_1 - norm$ and the notation of $l_1 - norm$ of vector $\vec{a}$ is written as $\|a\|_1$. It is computed as the sum of absolute vector values. The length of a given vector or its magnitude is normally used as a regularization method in machine learning. It is also used for broader vector and matrix operations. Its graphical representation is shown in Fig. 7.2. It is explained for the given two vectors of n dimensions $\vec{a} \equiv (a_1, a_2, \cdots, a_n)$ and $\vec{b} \equiv (b_1, b_2, \cdots, b_n)$. $L_1(a, b)$ norm is calculated by the following equation:

$$L_1(\vec{a}, \vec{b}) = (\sum_{i=1}^{n} |a_i - b_i|) \tag{7.21}$$

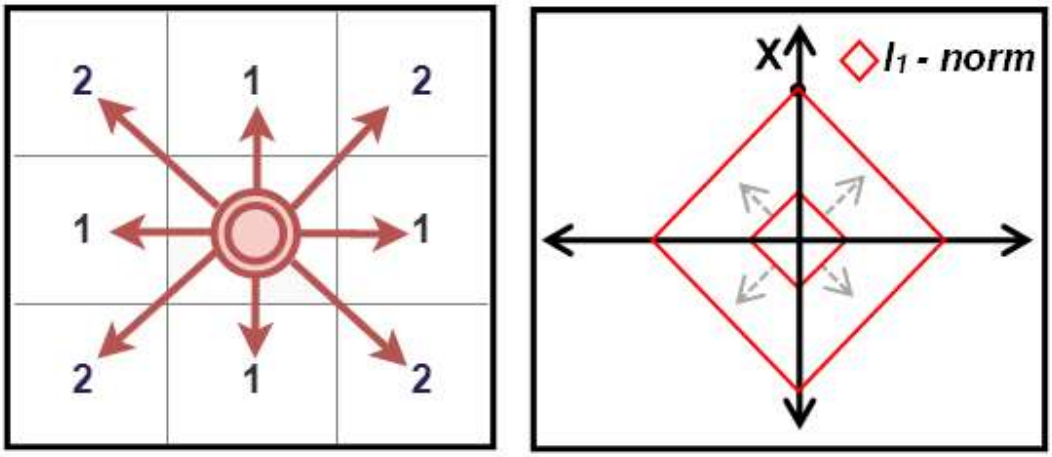


Figure 7.2: l_1-norm representation.

7.3.5 Principal component analysis

Principal Component Analysis (PCA) is widely used for dimension reduction in high dimensional datasets while preserving trends and patterns. It is useful for extracting valuable information from a confused dataset, and its simple and non-parametric nature makes it a standard tool in modern and big data analysis in diverse applications [243]. It provides a roadmap to reduce the high dimensional complex data into low dimensional space and extract hidden and simplified structures without having prior knowledge about its class or phenotype.

PCA reduces data by applying geometrical projection on the dataset and mapping to low dimensions, called principal components (PCs) [244]. It considers the minimum number of PCs to find the best summary of the dataset. In other words, PCA maps n-dimensional feature space to k-dimensional features, such that $k < n$. PCA reduces the perpendicular distance between PCs and data points to minimize the redundancy of data under the condition of minimal loss of information. It is based on a few assumptions, for example, linearity; large variances and PCs are orthogonal. Another major objective of PCA is to find the linearly independent dimensions to represent data without losing information. It helps to reconstruct new feature space with minimal reconstruction loss. The basic steps of PCA for a given dataset $X \equiv (x_i, y_i)_{i=1}^{N}$ with n dimensions and N instances are as follows:

1. Compute the mean value μ of N data instances.

$$\mu = \frac{1}{N} \sum_{i=1}^{N} (x_i) \tag{7.22}$$

2. Compute the covariance matrix R.

$$R = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T \tag{7.23}$$

3. The projection variance $\sigma_p^2(U)$ is subject to the unit vector constraint and can be minimized using objective function J defined with the help of Lagrange's multiplier:

$$\begin{aligned} J &= U^T RU + \lambda(1 - U^T U) \\ &\Rightarrow \bigtriangledown_U J = 2RU - 2\lambda U \\ &\Rightarrow \bigtriangledown_U J = 0 \\ &\therefore RU = \lambda U \end{aligned} \tag{7.24}$$

where λ is an eigenvalue or corresponding feature value of covariance matrix R, and U is the feature matrix consisting of the corresponding feature vectors u_i of feature values λ, arranged in descending order.

4. The cumulative variance contribution rate of the first k-row principal elements is computed by using feature values and feature vectors.

5. Data is transformed from X to $X^{'}$, which realizes the linear transformation from the n dimension to the k dimension such that $k < n$ by following:

$$Pr = U_k \tag{7.25}$$

$$X^{'} = Pr \cdot X \tag{7.26}$$

where Pr is the feature matrix of the corresponding feature vectors of the first k-row feature values.

7.3.6 Support vector machines

The SVM algorithm introduced by Vapnik [245] is a popular machine learning algorithm with strong theoretical foundations, and experimentally validated on different pattern

recognition problems. In this work we mainly concentrate on the binary classification problem in the n dimensional space R^n. Data is represented by $X \equiv (x_i, y_i)_{i=1}^{l}$ where $(x_i) \in R^n$ denotes an input sample with class label $y_i = \{-1, +1\}$. An overview of classifiers is also presented with respect to the binary classification problem. The main objective of linear SVM is to construct a pair of parallel hyperplanes between the samples of binary class:

$$P1 : w^T \cdot x + b = +1 \tag{7.27}$$

$$P2 : w^T \cdot x + b = -1 \tag{7.28}$$

where $b \in R$ is a bias term and $w \in R^n$ is a normal vector of hyperplane. As per the structural risk minimization principle, the risk bound is minimized, and separating hyperplanes are obtained by formulating the optimization problem as a quadratic programming problem (QPP) as shown in (7.29).

$$min_{w,b,\xi} \frac{1}{2} \|w\|^2 + c \sum_{1}^{N} \xi_i \tag{7.29}$$

$$such\ that\ y_i(w^T x_i + b) \geq 1 - \xi_i \;\; \xi_i \geq 0 \quad \forall i \in \{1, ..., N\}$$

$\|.\|$ represents the l_2- norm and ξ_i is a slack variable which indicates misclassification error in the i^{th} input sample; $c > 0$ is a regularization factor that balances the trade-off between the training error and generalization ability of the model. The dual QPP formulation [246] of (7.29) is illustrated as follows:

$$maximize_\alpha \sum_{i=1}^{N} \alpha_i + \frac{1}{2} \sum_{j=1}^{N} \sum_{i=1}^{N} y_j y_i \langle x_j x_i \rangle \alpha_j \alpha_i \tag{7.30}$$

$$subject\ to \sum_{i=1}^{N} \alpha_i y_i = 0, \ and \quad \forall i \in [1, N] \quad 0 \leq \alpha_i \leq 1$$

In the above formulation, $\alpha \in R^N$ is termed as a Lagrangian multiplier; test sample x is classified as -1 or +1 by solving dual QPP in (7.30) and following the decision function $\varphi(x)$:

$$\varphi(x) = sign(w^{\star T} \cdot x + b^{\star}) \tag{7.31}$$

where $(w^{\star}, b^{\star})$ and $\alpha_i^{\star}$ are the optimal solutions to the problem in (7.29) and (7.30), respectively.

7.3.7 Fuzzy SVM

Many real-world applications involve large datasets, which may have uncertainty in terms of noise, outliers, and class imbalance. Furthermore, some data may not be as important as other data for constructing a separation hyperplane or have a different distribution. To overcome these issues, researchers have applied the concept of fuzzy theory with SVM and other machine learning techniques [247, 248]. For this work, fuzzy SVM based classifier FSVM-WD [249] is considered due to its efficient computation and lower complexity compared to other fuzzy variants of SVM. Normally, the dataset for binary classification is represented as $(x_i, y_i)_1^N$, where $x_i \in R^n$ with $y_i \in \{-1, 1\}$. However, for using the fuzzy SVM, the dataset will also include a membership value Υ_i with respect to the i^{th} sample. Data distribution is analyzed prior to assigning different penalty parameters to different instances as per their significance and contribution. The larger the Υ_i value, the more valuable the given instance. SVM with fuzzy membership Υ_i, known as fuzzy SVM, and thus the QPP problem (7.29) with kernel $\phi(x_i)$ is reformulated as:

$$min_{w,b,\xi} \frac{1}{2} \|w\|^2 + c \sum_{1}^{N} \Upsilon_i \xi_i \tag{7.32}$$

$$such\, that\ y_i(w^T \cdot \phi(x_i) + b) \geq 1 - \xi_i \ \ \xi_i \geq 0 \quad \forall i \in \{1, ..., N\}$$

Similarly, (7.30) is modified with kernel as follows:

$$maximize_{\alpha} \sum_{i=1}^{N} \alpha_i + \frac{1}{2} \sum_{j=1}^{N} \sum_{i=1}^{N} y_j y_i \phi(x_j)\phi(x_i)\alpha_j \alpha_i \tag{7.33}$$
$$subject\ to \sum_{i=1}^{N} \alpha_i y_i = 0, and \quad \forall i \in [1, N] \quad 0 \leq \alpha_i \leq \Upsilon_i c$$

Further, it should satisfy the Karush-Kuhn-Tucker(KKT) condition [250] as follows:

$$\forall i;\ \alpha_i(y_i(\phi(x_i) \cdot w + b) + \xi_i - 1) = 0$$
$$\forall i; (\Upsilon c - \alpha_i)\xi_i = 0 \tag{7.34}$$

Finally, optimal $w^\star, b^\star$ and $\alpha^\star$ values are computed and used for sample classification. For FSVM-WD, we choose the RBF kernel and use the concept of k-nearest neighbour distance to estimate the relative density within the class. The value of K is taken as $\sqrt{N}$, where N is the number of training samples. Initially, for each sample, K^{th} nearest neighbour is searched, and the distance d_{jk} is noted and used to evaluate the significance of each sample. Higher-density instances will get a higher value, and lower-density instances will get a lower value. The relative density ρ between two instances (x_i, x_j) is calculated by taking the ratio of the reciprocal of distance:

$$\rho(x_i, x_j) = \frac{d_{jk}}{d_{ik}} \tag{7.35}$$

7.3.7.1 Fuzzy membership function

The membership function $\Upsilon(x_i)$ for any sample x_i is defined as the reciprocal of the distance d_{ik}, which is measured as the distance between the i^{th} instance and its k^{th}

nearest neighbours within the same class. Samples in the high-density region are considered as more informative and thus assigned high $\Upsilon(x_i)$ value. Those that lie far from the high-density region are assigned a low value and considered as outliers, noise, etc. To discard the impact of data distribution scale on the decision, a normalized fuzzy member is defined as:

$$\Upsilon(x_i) = \frac{\frac{1}{d_{ik}}}{\sum_{j=1}^{n_c} \frac{1}{d_{jk}}} \tag{7.36}$$

where n^c is the number of instances of the class under which x_i drops in. The FSVM-WD algorithm [249] is presented in the Algorithm 7.1.

Algorithm 7.1: FSVM-WD

Input: Training samples: $\Xi \equiv \{(x_i, y_i)\}_{i=1}^{N_F}$ penalty-factor c, and kernel parameter (γ)

Output: Fuzzy SVM decision function- φ

1 Split Ξ into two sets, Ξ^- only contain malignant (negative) instances, and Ξ^+ only contain benign (positive) instances.

2 Calculate the number of instances in Ξ^- and Ξ^+, then record these values as F^- and F^+, respectively, where $F^- + F^+ = N_F$.

3 Calculate κ for Ξ^- and Ξ^+ classes, respectively, where $\kappa^- = \lceil \sqrt{F^-} \rceil$, $\kappa^+ = \lceil \sqrt{F^+} \rceil$.

4 For every instance x_i^+ in Ξ^+, calculate the distance d_{i+} to the $\kappa^+ th$ nearest neighbors in Ξ^+ and similarly for each instance x_j^- and denote it as d_{j-}.

5 For every instance x_i in Ξ, compute the relative density by using (7.35), and then compute its fuzzy membership value $\Upsilon(x_i)$ as defined in (7.36).

6 Fuzzy SVM is trained by (7.32) with given parameters c and γ. Get the value of decision function $\psi(x)$ as a class label.

7.4 Proposed Methodology

This section gives a detailed description of the proposed framework for CMT classification. The proposed framework uses novel feature extraction techniques for the CMTHis dataset and an FSVM-WD classifier for binary classification. This classification model is useful because of its low complexity and ability to overcome data uncertainty in a relatively short time. As shown in Fig. 7.3, the framework includes stain normalization, feature extraction, dimensionality reduction, and classification as sub-tasks.

7.4.1 Stain normalization

Variations in staining procedures used in histopathological slides affect the image quality, influencing feature extraction and classification algorithms. In order to take account of variations in staining techniques, stain normalization was done as per Macenko et al. [73], which is described in Chapter 3.

7.4.2 Feature extractor- CoMHisP

This section explains how an image can be treated as a mass of particles and how the effect of illumination can be minimized by choosing the CoM of an image as a feature [235]. We also describe its illumination invariant characteristics mathematically and detail the process followed for feature engineering. We have already described the basic concepts of CoM in Section 7.3.3.

7.4.2.1 Illumination invariant features

An image can be treated as a collection of mass particles where each individual pixel acts as a charged particle with intensity $I(x, y)$ as mass m and with corresponding spatial position vector $\vec{r_i} \equiv (x_i, y_i)$. Thus, pixels are the smallest addressable unit and act similarly to atoms. We consider a window of size 3×3 in the local neighbourhood of a pixel within an image. The total number of mass particles is 9. Thus, we can

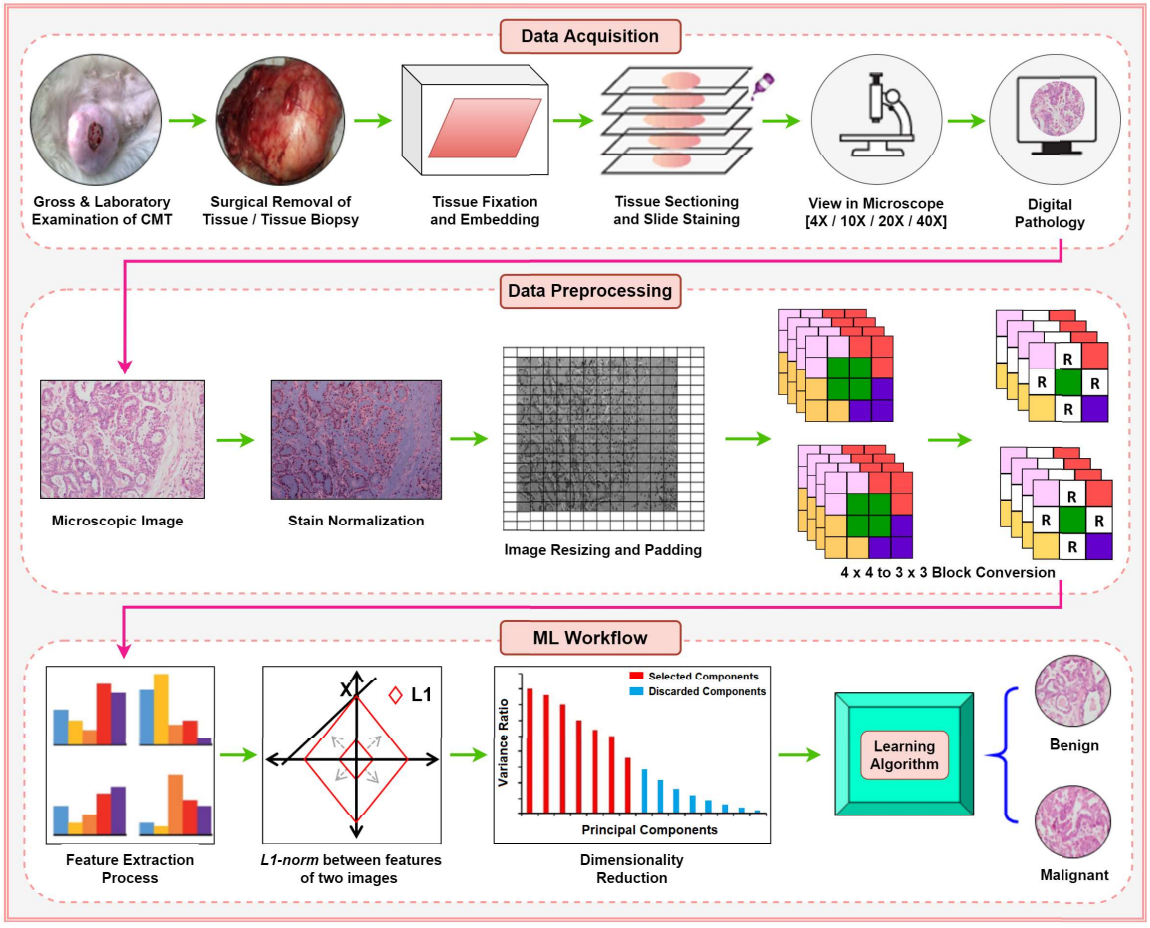


Figure 7.3: Overview of the proposed framework-CoMHisP.

rewrite (7.2) as:

$$\begin{aligned} \vec{r_{cm}} &= \frac{I_1\vec{r_1} + I_2\vec{r_2}, \cdots, I_9\vec{r_9}}{I_1 + I_2 + \cdots + I_9} \\ \vec{r_{cm}} &= \frac{\sum_{i=1}^{9} I_i\vec{r_i}}{I} \end{aligned} \tag{7.37}$$

Generally, local patterns follow the Retinex theory [251] shown in (7.38). Based on this, Horn [252] introduced an illumination reflectance model:

$$I(i,j) = R(i,j) \cdot L(i,j) \quad \forall i,j \tag{7.38}$$

where (i, j) are spatial coordinates in the image, and $L(i, j)$ and $R(i, j)$ are its illuminance and reflectance components, respectively. It has been shown that in a small neighbourhood, L can be approximated as a constant because it hardly varies across the spatial region, and R is considered as an illumination invariant feature of an image. The proposed method follows a division method for maintaining the ratio between CoM and the center of a neighbourhood. For two adjacent pixels (i, j) and $(i + 1, j)$, the intensities ratio is written as:

$$\frac{I(i+1,j)}{I(i,j)} \approx \frac{R(i+1,j)}{R(i,j)} \tag{7.39}$$

The illumination invariant characteristics are proved by stating theorem [235] as follows:

Theorem 7.1 *For a given image I with position vector (x, y) in the specified neighbourhood of size 3×3, the arc-tangent of the disparity between the neighbourhood's center pixel and CoM is an invariant feature with respect to an illumination.*

Proof: For this we consider the origin of the image space as o, the center pixel of neighborhood as $P \equiv (p_x, p_y)$ and CoM as $C \equiv (c_x, c_y)$. The gradient of CoM is represented by $G \equiv (\Delta G_x, \Delta G_y)$. The following equations show the calculation of the gradient of CoM in the x and y directions:

$$\begin{aligned} \Delta G_x &= [(c_x - o) - (p_x - o)] \times \sum I(x, y) \\ \Delta G_y &= [(c_y - o) - (p_y - o)] \times \sum I(x, y) \end{aligned} \tag{7.40}$$

The magnitude of G and the angle of the gradient are computed by using (7.41), and (7.42).

$$G = \sqrt{\Delta G_x^2 + \Delta G_y^2} \tag{7.41}$$

$$\phi = tan^{-1} \frac{(\Delta G_y)}{(\Delta G_x)} \tag{7.42}$$

By using (7.37), (7.40), and (7.42) we can write:

$$\phi = tan^{-1}\frac{(c_y - p_y)}{(c_x - p_x)} \tag{7.43}$$

$$\phi = tan^{-1}\left(\frac{(\sum_{i=1}^{9} I_i \cdot y_i) - p_y \cdot (\sum_{i=1}^{9} I_i)}{(\sum_{i=1}^{9} I_i \cdot x_i) - p_x \cdot (\sum_{i=1}^{9} I_i)}\right) \tag{7.44}$$

Thus, by using (7.38) we can rewrite (7.44):

$$\phi = tan^{-1}\left(\frac{(\sum_{i=1}^{9} R_i L_i \cdot y_i) - p_y \cdot (\sum_{i=1}^{9} R_i L_i)}{(\sum_{i=1}^{9} R_i L_i \cdot x_i) - p_x \cdot (\sum_{i=1}^{9} R_i L_i)}\right) \tag{7.45}$$

We adopt the assumption that the luminance component L varies very slowly within the 3×3 neighborhood, which is widely accepted. We can infer that $L_i \approx L_{i+1} \approx L$ and it will be canceled out due to its presence in both numerator and denominator. Subsequently, we can rewrite (7.45) as follows:

$$\phi = tan^{-1}\left(\frac{(\sum_{i=1}^{9} R_i \cdot y_i) - p_y \cdot (\sum_{i=1}^{9} R_i)}{(\sum_{i=1}^{9} R_i \cdot x_i) - p_x \cdot (\sum_{i=1}^{9} R_i)}\right) \tag{7.46}$$

□

It is already known that the reflectance component R of any pixel within the image is an illumination invariant feature. Hence, it is also proved that ϕ is also an illumination invariant feature. The gradient of CoM with the neighbourhood center preserves the characteristics of local edges and makes it superior to others.

7.4.2.2 Features extraction

On the basis of the aforementioned concept, features were extracted from the CMTHis dataset. Data specifics are provided in Section 7.5. Prior to feature extraction, the following steps were followed.

1. **Image resizing and optimal block selection:** The dataset comprises images of size 1600 × 1200. To make the process computationally efficient for feature extraction, images were resized to 128 × 128. Further, 8 blocks were considered, and for each block, we used 4 × 4 window size with normalized intensity. Subsequently, these windows were converted to 3 × 3 using a specific method to obtain the minute details or micro-patterns of the images. Before the following steps are undertaken, the image is converted to grayscale. The reason for selecting a 3 × 3 window is that it satisfies the Retinex theory that is already explained and proved for illumination invariant features. Additionally, the most popular deep architecture uses different filter sizes to extract features from the images. Moreover, the presence of non-uniformly distributed nuclei in cancer images also motivated us to choose a window size of 4 × 4 for each pixel considered as a center for this block, which is further converted to 3 × 3 along with appropriate zero padding as per requirement and stride 1. The conversion of pixels from 4 × 4 to 3 × 3 as shown in Fig. 7.3 is performed by the following equations:

$$I_c^n = (I_c + I_6 + I_9 + I_{10})/4 \tag{7.47}$$

$$I_1^n = (I_1 + I_2 + I_5)/3 \tag{7.48}$$

$$I_2^n = (I_2 + I_3)/2 \tag{7.49}$$

$$I_3^n = (I_3 + I_4 + I_7) \tag{7.50}$$

$$I_4^n = (I_1 + I_5)/2 \tag{7.51}$$

$$I_5^n = (I_7 + I_{11})/2 \tag{7.52}$$

$$I_6^n = (I_8 + I_{12} + I_{13})/3 \tag{7.53}$$

$$I_7^n = (I_{13} + I_{14})/2 \tag{7.54}$$

$$I_8^n = (I_{11} + I_{14} + I_{15})/3 \tag{7.55}$$

Algorithm 7.2: Center of Mass Computation

Input: (Histopathological Image I, r)

Output: G_x, G_y

$I \leftarrow RGBTOGRAY\ (I)$

img $= RESIZE\ (I, (128 \times 128)$

img $= ZEROPADDING$ ▷ 1 column on left, 1 row on top, 2 columns on right and 2 rows on the bottom row of the matrix

for *each pixel in img* **do**

  Select a 4×4 window

  Normalizing each *pixel* intensity of the 4×4 window

  3×3 windows $\leftarrow$ 4×4 windows($pixel$) ▷ Converting 4×4 windows to 3×3 using equations (7.47)-(7.55)

  for *each* 3×3 *window* **do**

    $\vec{x} \leftarrow$ PIXELS($pixels$, r) ▷ r defines the neighborhood pixels of the center pixel in 3×3 window (r=1, for 3×3 window)

    sum $=$ SUM $(\vec{x})$

    C_x, $C_y \leftarrow \vec{x}$

    P_x, $P_y \leftarrow$ POSITION($centerpixel$)

    O_x, $O_y \leftarrow$ Origin

    Gx $= (C_x - O_x)$ - $((P_x - O_x)\ \times$ sum$)$

    Gy $= (C_y - O_y)$ - $((P_y - O_y)\ \times$ sum$)$

  end

end

2. **CoM with angular features:** The newly converted 3×3 block of image pixels, where each pixel's intensity acts as a mass particle, and these 9 pixels represent the system of mass particles. CoM of this system is computed with the help of (7.37) where $I = \sum_{i=1}^{9} I_i$ is the sum of all particle intensities, with each pixel having the position vector $\vec{r_i} \equiv (x_i, y_i)$. Similarly, we compute CoM in the X and

Algorithm 7.3: Feature Vector Extraction

Input: G_x, G_y

Output: theta (θ)

$array_1$, $array_2 \leftarrow [0]$
for *m in (128/block size)* **do**
  for *n in (128/block size)* **do**
    S_x, $S_y \leftarrow 0$ ▷ block size = 8
    $S_x = \text{SUM}((block(G_x))$
    $S_y = \text{SUM}((block(G_y))$
    $array_1 \leftarrow S_x$
    $array_2 \leftarrow S_y$
  end
end
for k *in* $array_1$ **do**
  $\theta = tan^{-1}(\frac{array_2[k]}{array_1[k]})$
end

Algorithm 7.4: Distance between two vectors

Input: $\vec{u}$, $\vec{v} \in \theta$

Output: Distance (X)

$E \leftarrow |\vec{u}|$
ds $\leftarrow 0$
for i *in* $\{1, 2, \cdots, E\}$ **do**
  ds $\leftarrow$ ds + $\min(|\vec{u_i} - \vec{v_i}|, 2\pi - |\vec{u_i} - \vec{v_i}|)$
end
X $\leftarrow$ ds

Y direction denoted as c_x and c_y for each pixel of the 128×128 image. Further, the gradient of CoM (c_x, c_y) along the X and Y axes are computed, respectively, with the help of (7.40), which denotes the two immediate feature matrices, each

of size 128×128. Afterwards, these matrices are divided into 8×8 blocks, and for each block, net ΔG_x and net ΔG_y are computed, which results in each case in a 256 feature vector. Then, angular features, i.e. angle of the gradient, are computed by using (7.42). Finally, we have 256 features for each image. Similar operations are applied on both the training and the test dataset to obtain the features. The L_1 norm among these feature space, which is the minimum angle distance difference between the angular feature vectors ϱ_1 and ϱ_2 of two instances, is computed as follows:

$$D(\varrho_1, \varrho_2) = \sum_{i=1}^{256} min(|\rho_{1i} - \rho_{2i}|, 2\pi - |\rho_{1i} - \rho_{2i}|) \tag{7.56}$$

The complete process of feature engineering is explained in Algorithms 7.2, 7.3, and 7.4.

7.4.2.3 Dimensionality reduction and classification

The high dimensionality of the feature space is reduced by using PCA, as described in Section 7.3.5, and classified into malignant & benign categories using FSVM-WD, SVM, and RF classifiers.

7.5 Experimental Results and Analysis

This section presents experimental results and comparative analysis to illustrate the superiority of the proposed method over existing methods.

7.5.1 Experimental setup

The framework described in Section 7.4 was implemented in Python 3.7 on an Intel Core i7-3770 CPU, clocked at 3.40GHz with 4-cores and 12 GB RAM. The performance of

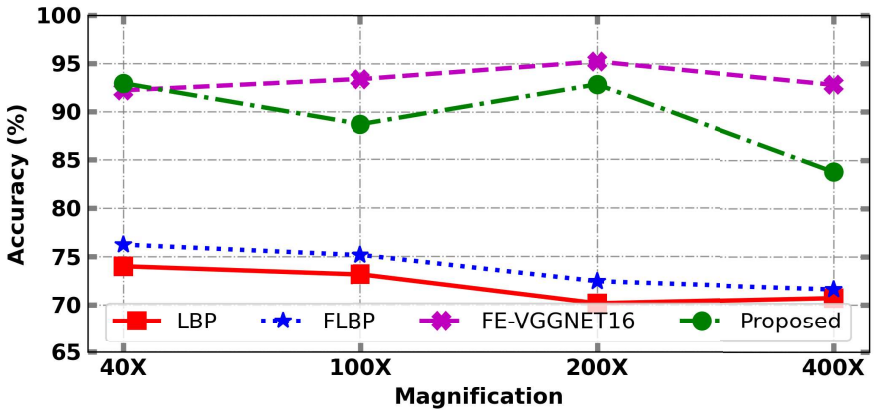
DL based models developed in previous studies, as detailed in preceding chapters was found to be lower on CMTHis dataset with smaller number of images, as compared to BreakHis dataset, which comprise of large number of images. As the study was conducted with the aim of developing a model which works well on small size datasets as well, thus the performance of the framework was first evaluated on a standard BreakHis data set and thereafter it was evaluated thoroughly on CMTHis dataset. The proposed framework was applied to the CMTHis dataset, which was randomly divided into 70% (31 patients) for training and 30% for testing (13 patients). The split protocol described in chapter 3 was used to create five-folds, and the reported results represent the average from these. To validate the efficacy and superiority of our proposed framework, we used several window sizes – $3 \times 3, 4 \times 4, 5 \times 5, 6 \times 6, 7 \times 7, 9 \times 9$ & 11×11, and block sizes of 8×8 & 16×16. After extensive experimental analysis, we found that the window size of 4×4 converted to 3×3 and block size of 8×8 represented the best possible combination. We compared our proposed framework (CoMHisP) with three state-of-the-art classification algorithms – SVM with RBF kernel, RF, and FSVM-WD. After applying PCA, feature scaling was carried out along axis=0, and hyperparameters of classifiers were tuned using grid-search. The proposed classification model is evaluated in terms of performance metrics, accuracy, and AUC.

7.5.2 Evaluation of the performance of CoMHisP in magnification dependent model

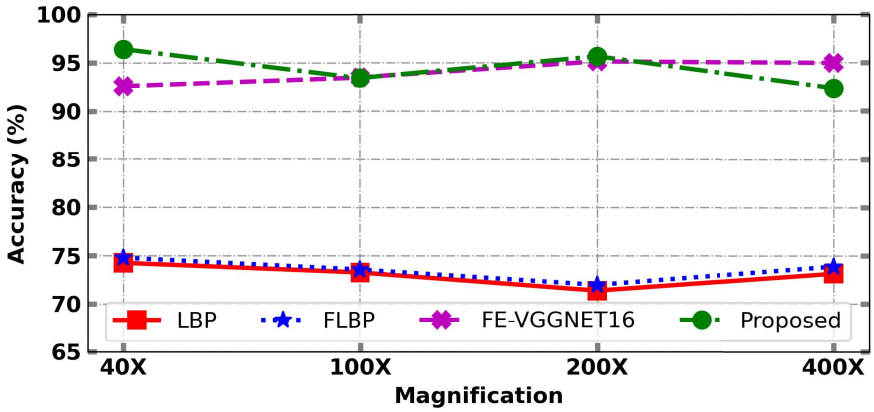
The framework's performance was initially evaluated using standard BreakHis dataset, which contains a large number of images of HBC. The performance of the model on BreakHis dataset is presented as mean classification accuracies for the differentiation of malignant and benign breast cancers. The highest classification accuracy of 96.89% was obtained at 40X magnification with FSVM-WD classifier. The results are presented in Figure 7.4. Once the performance of the model was established on the BreakHis dataset,

Table 7.1: Performance of Different Feature Descriptors with Classifiers

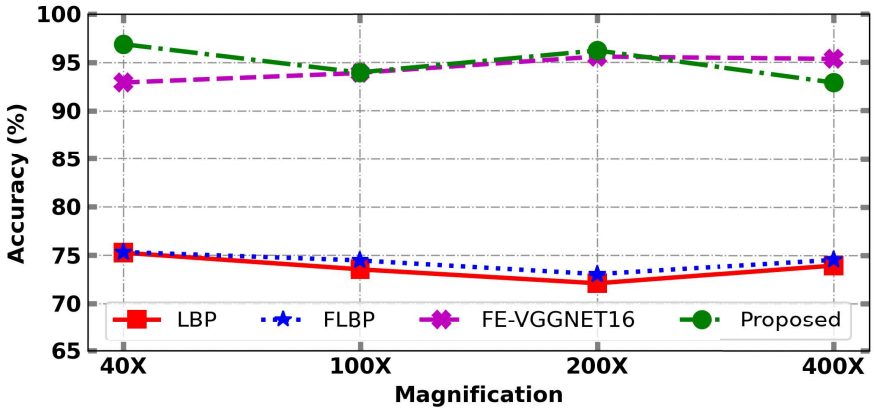
Descriptor	Classifier	Without Stain Normalization								With Stain Normalization							
		Magnification Factors								Magnification Factors							
		40X		100X		200X		400X		40X		100X		200X		400X	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
LBP	SVM	75.90	4.75	75.10	3.52	74.52	3.80	73.25	4.28	77.43	3.75	76.82	3.22	75.28	4.02	74.72	4.87
	RF	75.20	3.27	74.05	2.95	73.58	2.58	72.20	3.42	76.12	3.22	75.98	2.52	74.05	2.58	73.70	3.70
	FSVM-WD	77.70	4.10	77.58	3.15	75.22	3.13	74.92	3.24	78.58	3.84	78.18	2.92	76.92	3.14	75.82	3.24
Fuzzy LBP	SVM	77.58	3.28	77.20	2.85	75.82	2.98	74.82	4.12	78.50	2.45	77.82	3.15	76.90	2.67	76.15	3.62
	RF	76.15	2.95	75.10	3.42	75.22	2.52	73.62	3.54	78.16	3.22	77.30	2.28	76.24	2.55	74.72	3.52
	FSVM-WD	78.40	3.42	78.18	3.10	76.90	2.74	75.54	3.82	80.45	2.74	79.22	2.82	78.27	2.34	77.54	3.80
FE-VGGNET16	SVM	89.09	4.10	89.15	3.40	85.44	2.52	81.87	0.69	81.15	2.97	82.58	3.13	77.58	3.57	79.07	3.47
	RF	91.27	2.93	91.41	3.00	86.13	1.31	81.63	1.56	84.68	3.25	84.37	2.97	81.96	4.68	81.62	3.16
	FSVM-WD	91.15	3.10	92.18	3.28	85.88	3.62	82.70	3.26	83.10	2.62	83.88	3.20	79.76	3.67	80.74	4.22
Proposed	SVM	95.09	2.54	93.80	2.65	93.05	2.72	90.24	3.84	96.95	1.92	95.47	2.62	95.28	2.28	92.30	2.98
	RF	93.25	2.82	91.60	3.12	91.10	2.78	87.58	4.72	95.90	2.80	94.15	2.40	94.25	2.30	91.70	3.52
	FSVM-WD	95.77	2.20	94.98	2.48	94.15	2.14	92.40	3.28	97.25	1.80	96.10	2.15	95.98	2.40	93.76	3.24



(a) RF without Stain Normalization



(b) SVM without Stain Normalization



(c) FSVM-WD without Stain Normalization

Figure 7.4: Performance of the proposed CoMHisP feature extractor versus other state-of-the-art feature descriptors on the BreakHis dataset without stain normalization.

the model was evaluated in detail on CMTHis dataset, which includes histopathological images of CMT captured across different magnifications. The results are presented as mean classification accuracies for the differentiation of malignant and benign CMTs.

7.5.2.1 Effect of different classifiers on the performance of CoMHisP framework

In this study, we proposed the CoMHisP framework for feature extraction from the histopathological images, and the classification is achieved using various classifiers, namely SVM, RF and, FSVM-WD. To neutralize the variations in the staining patterns, a stain normalization algorithm was applied. As evident from the Table 7.1, the application of the CoMHisP framework resulted in the highest classification accuracies with FSVM-WD, ranging from 93.76 ($\pm$3.24) to 97.25 ($\pm$1.80)% across different magnifications after stain normalization.

Fig. 7.5 represents classification accuracies achieved by applying the proposed

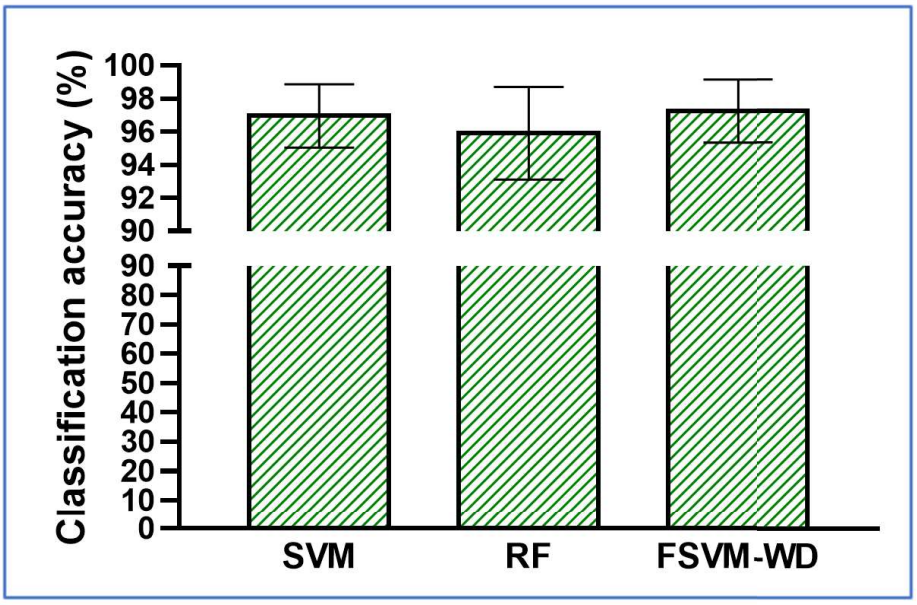


Figure 7.5: Bar diagram depicting classification accuracies achieved by applying the proposed CoMHisP feature extractor and different classifiers on CMTHis data after stain normalization at 40X magnification.

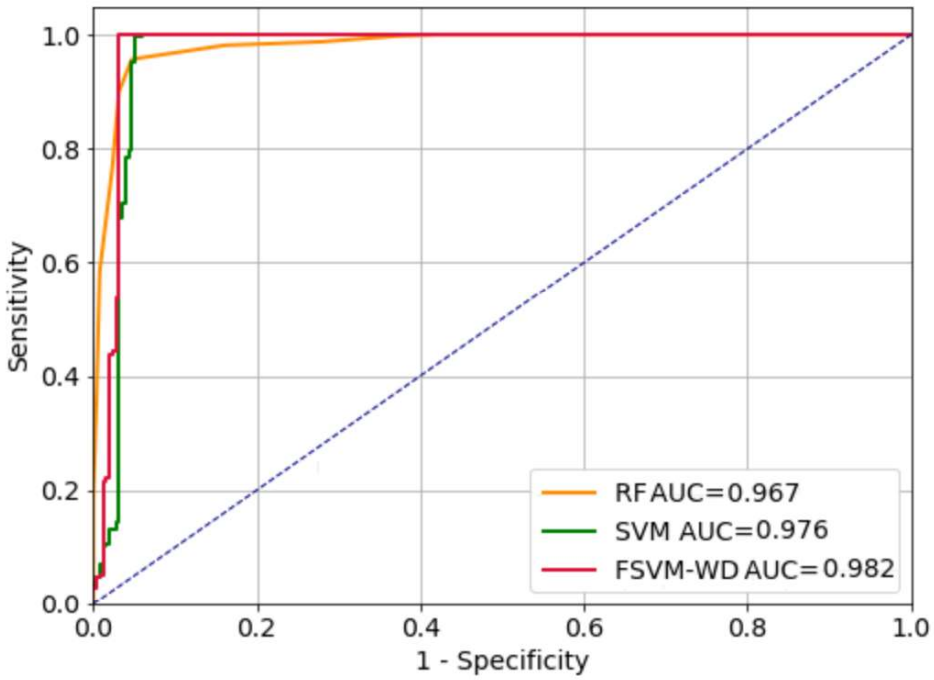


Figure 7.6: ROC curve analysis showing AUC of the CoMHisP framework using CMTHis dataset with different classifiers at 40X magnification.

CoMHisP feature extractor and different classifiers on CMTHis data at 40X magnification after stain normalization. From the results, it can be concluded that the FSVM-WD, an extension of SVM, is a useful classifier for benign and malignant classification of histopathological images. Performance of the proposed framework was also evaluated using ROC curve analysis (Fig. 7.6). This revealed that the AUC of the CoMHisP framework, using different classifiers at 40X magnification, ranged from 0.967 to 0.982, reflecting the high accuracy of the proposed framework in differentiating between malignant and benign CMTs.

7.5.2.2 Effect of magnification on the performance of CoMHisP

The proposed framework was independently applied to each of the four magnifications available in the CMTHis dataset. The results obtained by different classifiers varied with the magnification factor. The proposed CoMHisP feature descriptor, irrespective

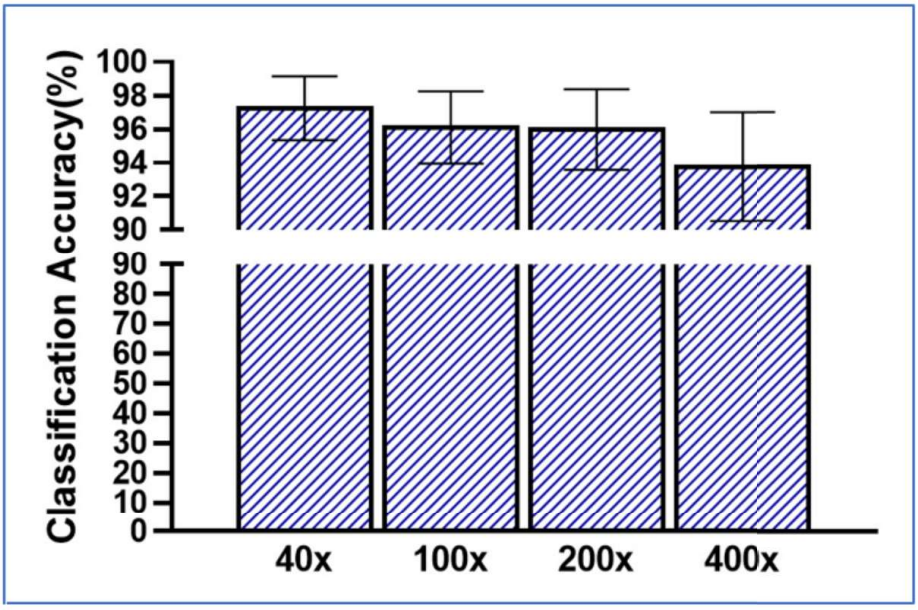
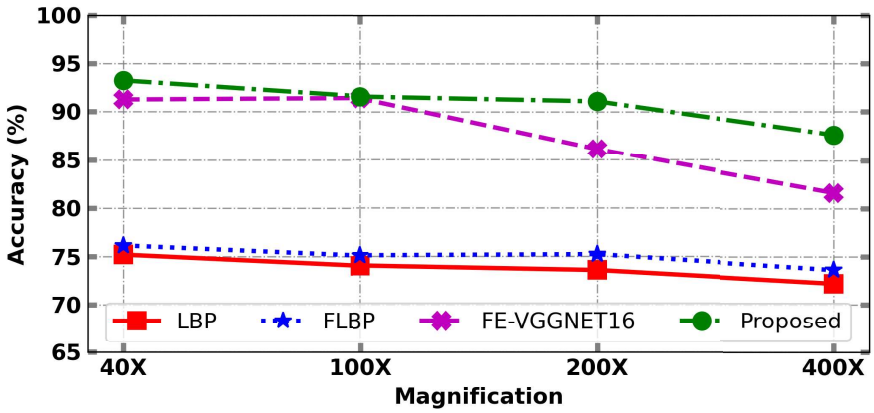


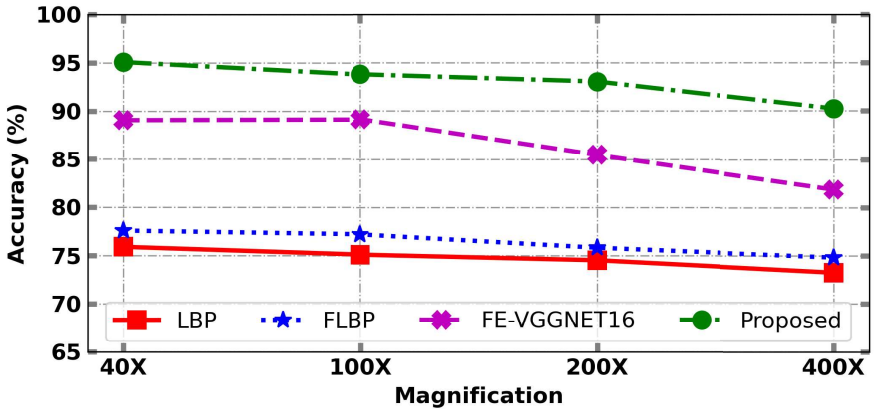
Figure 7.7: Bar diagram depicting performance of the proposed CoMHisP framework with FSVM-WD classifier at different magnifications.

of the classifier applied, showed the highest classification accuracies at 40X magnification, and the lowest accuracy was observed at 400X magnification. All the classifiers showed higher accuracies at lower magnification. As shown in Table 7.1, the CoMHisP feature descriptor performed best at 40X magnification using the FSVM-WD classifier, producing a classification accuracy of 97.25 ($\pm$1.80)% for differentiation of malignant and benign CMTs. As demonstrated in Fig. 7.7, classification accuracies (on CMTHis data after stain normalization) using the proposed framework decreased as the magnification factor increased.

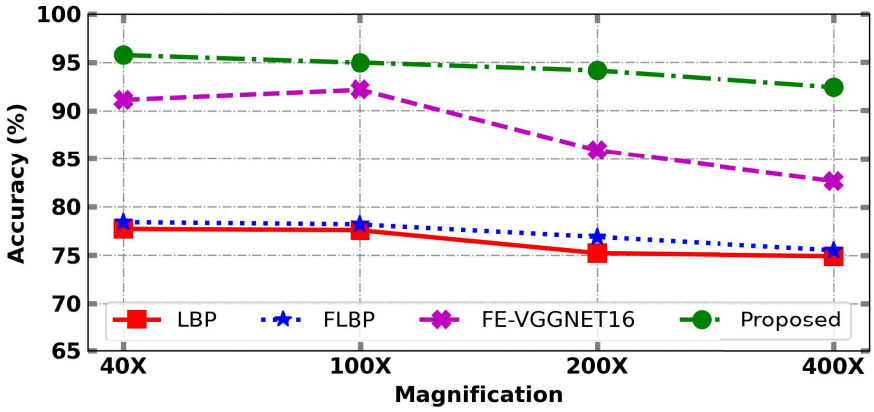
Magnification also affects manual image interpretation by a pathologist, who first examines the stained slides under lower magnification and then utilizes higher magnification for fine-tuning of results in the areas of interest. The proposed framework outperformed at lower magnification as compared to higher magnifications. This may be because, at lower magnification, a larger region of interest (ROI) is available for feature extraction as compared to higher magnifications. Similar results have been



(a) RF without Stain Normalization

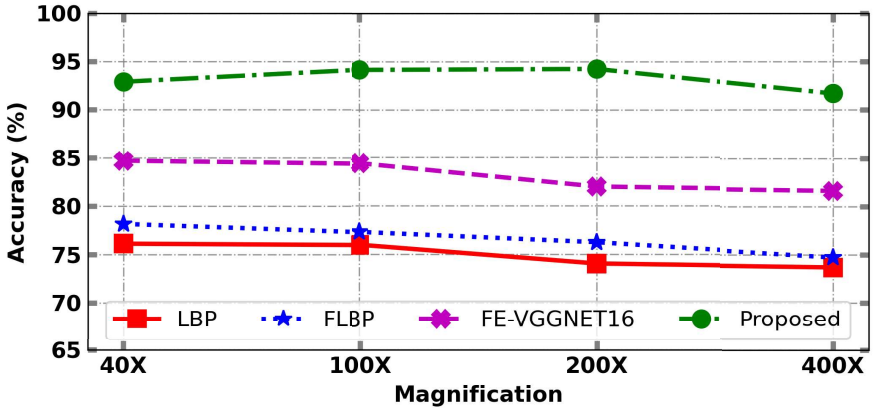


(b) SVM without Stain Normalization

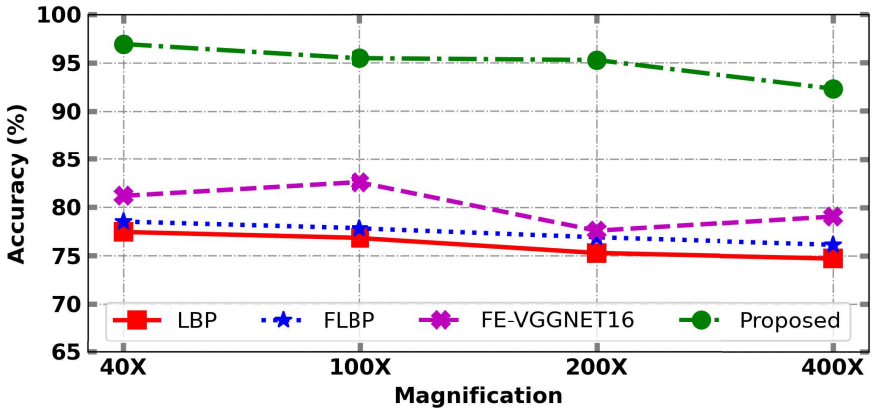


(c) FSVM-WD without Stain Normalization

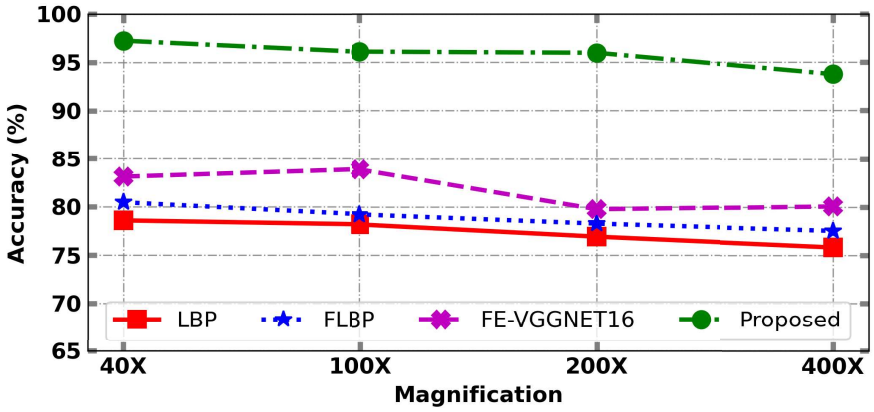
Figure 7.8: Performance of the proposed CoMHisP feature extractor versus other state-of-the-art feature descriptors on the CMTHis dataset without stain normalization.



(a) RF with Stain Normalization



(b) SVM with Stain Normalization



(c) FSVM-WD with Stain Normalization

Figure 7.9: Performance of the proposed CoMHisP feature extractor versus other state-of-the-art feature descriptors on the CMTHis dataset with stain normalization.

reported by Spanhol et al. [6].

Notably, the higher accuracies achieved at lower magnification imply that even images captured using simple microscopes providing lower magnifications will work extraordinarily well with the proposed algorithm. This is an additional advantage for low-cost clinical settings.

7.5.2.3 Comparison of the performance of the proposed CoMHisP feature extractor with other state-of-the-art feature descriptors

We compared the performance of the proposed CoMHisP feature extractor with other state-of-the-art local feature descriptors, namely LBP and fuzzy LBP. The comparison was also carried out using FE-VGGNET16, a promising deep ConvNet-based feature extractor for automated image recognition [216]. The performances of individual frameworks were compared for the classification accuracies achieved using different classifiers at different magnifications on the CMTHis dataset. The results are compiled in Table 7.1. The application of the FSVM-WD classifier resulted in better accuracies in comparison to the SVM and RF classifiers using the CoMHisP, LBP, and fuzzy LBP feature extractors. However, for FE-VGGNET16, better results were obtained using the RF classifier. It can therefore be concluded that the FSVM-WD classifier best suits the CoMHisP framework. As shown in Fig. 7.8 and 7.9, the proposed framework clearly outperformed other feature descriptors at all magnifications, with SVM, FSVM-WD, or RF classifiers. The highest classification accuracy (97.25 $\pm$1.80%), achieved with the FSVM-WD classifier, clearly establishes the superiority of the proposed framework over other state-of-the-art approaches for feature extraction on datasets with smaller sizes. Furthermore, the proposed framework is simple and easy to use compared to some other existing frameworks.

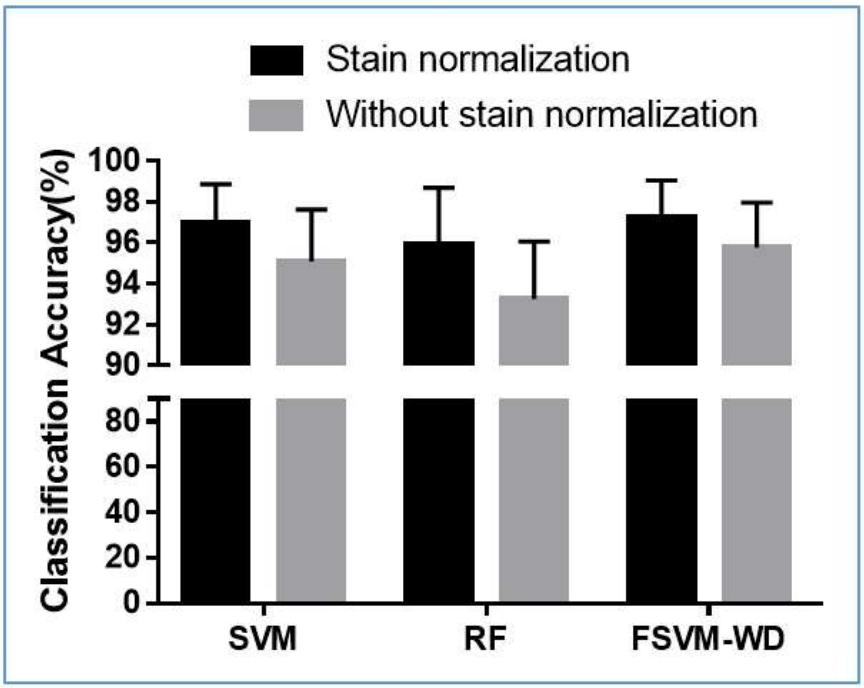


Figure 7.10: Effect of stain normalization on the performance of the proposed framework.

7.5.2.4 Effect of stain normalization on the performance of CoMHisP and other feature descriptors

Differences in the staining protocols at various laboratories pose a major bottleneck in the reproducibility, and large-scale application of algorithms for histopathological image analysis [241, 243]. To minimize these staining variations, various stain normalization algorithms have been proposed by different researchers. The stain normalization algorithm described in Section 7.4 was used in this study. The classification accuracies of the framework were compared with and without the application of the stain normalization algorithm. As can be seen from Table 7.1, and Fig. 7.10, stain normalization was found to increase the classification accuracies of the algorithms used in the proposed framework and other conventional algorithms, namely LBP and fuzzy LBP used in this study. Representative H&E-stained histopathological images from the CMTHis dataset with and without stain normalization are shown in Chapter 3.

However, stain normalization significantly reduced classification accuracy using FE-VGGNET16 as a feature extractor. This can be explained by the fact that the most commonly used stain normalization methods have been developed based on diverse approaches to machine learning. Thus, stain normalization algorithms may not always work well for deep learning-based schemes such as ConvNets, which have the capability to effectively learn in the presence of colour variation [216, 159].

7.5.3 Effect of classifiers and stain normalization on the performance of the proposed framework in a magnification-independent model

The performance of the proposed CoMHisP framework was also evaluated using a magnification-independent model, in which images were classified independent of their magnification. The classification accuracies achieved after the application of various classifiers with and without stain normalization were compared, as shown in Table 7.2. The highest classification accuracy was achieved using the proposed CoMHisP framework with the FSVM-WD classifier, followed by SVM. In the magnification-independent

Table 7.2: Performance of Magnification-Independent Model using the CoMHisP Framework.

Feature Extractor	**Classifier**	**Magnification Independent**			
		With Stain Normalization		Without Stain Normalization	
		Mean	SD	Mean	SD
FE-VGGNET16	SVM	84.78	3.87	85.64	3.74
	RF	86.70	2.90	88.66	2.68
	FSVM-WD	87.10	2.87	88.98	3.25
Proposed	SVM	94.58	2.78	93.18	2.55
	RF	94.16	2.48	92.85	3.27
	FSVM-WD	95.18	2.84	93.65	2.78

model, the classification accuracy was found to be 95.18 (±2.84)%, which was slightly less as compared to the accuracy of 97.25 (±1.80)% achieved at 40X magnification in the magnification-dependent model. The application of stain normalization improved the performance of the proposed framework to a level similar to that in the magnification-dependent model. The magnification-independent model also has the advantage of reduced processing time and complexity.

7.6 Summary

Machine learning is a powerful technique for biomedical image analysis that is widely accepted across different research communities. The performance of machine learning is highly dependent on the feature extraction and selection model. This chapter has proposed a novel framework, CoMHisP, for the extraction of features from cancer histopathological images. The proposed method was assessed in relation to novel techniques such as reducing block size for extraction of micro-patterns, use of CoM theory for feature vector extraction, and use of PCA for dimensionality reduction. Performance of the proposed framework was evaluated using different classifiers, namely SVM, RF, and FSVM-WD. The use of the FSVM-WD classifier with the proposed framework achieved the best results. Performance increased with stain normalization and was influenced by a magnification factor, producing a top mean classification accuracy of 97.25% (±1.80) at 40X magnification. Quantitatively, the proposed framework significantly outperformed the competitor methods LBP and fuzzy LBP, as well as the deep learning-based FE-VGGNET16 feature extractor. Moreover, the proposed methodology is simpler and requires less processing time than other methods, and consequently does not require high-performance computers. It performs extraordinarily well at lower magnification, which is an additional advantage for providing diagnosis using simple microscopes in low-cost clinical settings. To the best of our knowledge, this study is the first to test novel feature descriptors for machine learning-based CAD of CMTs. It

is worth repeating that the proposed framework can be expected to give superior results for CAD of HBC, owing to similarities in the histomorphological features of CMT and HBC.