Handcrafted and Deep Features for Biomedical Image Retrieval and Classification Applications

A thesis submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Deepamoni Mahanta

Registration No. TZ202258 of 2020

Department of Electronics and Communication Engineering

School of Engineering, Tezpur University

Tezpur, Assam, India - 784028

February, 2025

Chapter 6

Conclusions and Future scope

This chapter summarizes the key findings from this thesis's primary contributions and outlines promising directions for future research.

6.1 Concluding remarks

The key goal of this research has been to develop (i) Handcrafted texture feature descriptors for biomedical image retrieval applications and (ii) Developing only DL-based, and combined handcrafted and DL-based frameworks for biomedical image classification applications. The conclusions drawn from the investigations presented in this thesis are summarized chapter by chapter.

1. The scope of the biomedical image retrieval problem was explored in the introduction, and the importance of texture feature extraction within a CBIR framework was highlighted. The benefits of local pattern-based approaches and the key associated challenges were addressed and elaborated upon. Additionally, the impact of ML and DL in computer-aided analysis for various diseases, which includes COVID-19 diagnosis and skin cancer detection, was discussed. A discussion on handcrafted features and deep features in the context of biomedical images was also provided. Based on these discussions, the research problems were identified, and the significant contributions of this thesis were outlined.

2. Chapter 2 presents a literature review of various image feature descriptors, focusing on local pattern-based texture descriptors. This chapter reviews the primary LBP operator and its numerous extensions, along with

bit plane-based descriptors and those utilizing multiscale feature extraction techniques. From this review, it became evident that there is a pressing need to investigate local pattern-based texture descriptors that capture fine details and extract features at multiple scales. A detailed description of various DL-based techniques to diagnose COVID-19 employing CXR pictures is provided. Additionally, several recent studies that combine handcrafted and deep features are examined. The chapter concludes with a comprehensive review of DL-based approaches for SC detection using dermoscopic images. Upon reviewing the literature for both, COVID-19 identification using CXR images and skin lesion categorization using dermoscopic images, it is evident that DL-based techniques still require improvement in several key areas.

3. Chapter 3 introduces two local pattern-based descriptors that use the concept of decomposition of images into bit planes and utilize novel arbitrary sampling structures.

    - In the first descriptor, MS-LBASP, the input image is initially down-sampled at various scales using bi-cubic interpolation to address potential limitations of features extracted at the original scale. These images are decomposed into eight-bit planes to capture a wide range of image details, from coarse to fine. Three novel 2D arbitrary sampling structures and the conventional circular sampling technique are then employed to encode these BPs, enabling the capture of both uniform and non-uniform image textures. It has been shown that our arbitrary sampling structures use more arbitrary angular fluctuations in order to capture more information and irregular textures present in the images. The relationship between the fused local BP encoded values and the raw pixel intensity values of the input multi-scale image is calculated by considering both sign and magnitude differences. Each multi-scale image is split into four equal-dimensioned, separate patches to enhance spatial information. The histograms of the pattern maps are computed for each patch and combined, resulting in a higher feature dimension. To reduce this dimensionality, the feature maps are quantized into 16 levels. This descriptor has been shown to outperform the most relevant bit plane-based descriptors regarding retrieval efficiency across the NEMA-CT, TCIA-CT, and York-MRI image databases. For NEMA-CT dataset (at Tm=40), with regard to (%ARP and %ARR), the MS-LBASP enhances upon LBP, LBDP, LDEP, LWP, LBPANDP, LBPDAP and CSLBCoP, descriptors by (6.22%,7.83%), (2.39%, 3.74 %), (11.39%, 10.62 %), (8.92%, 9.49 %), (0.87%, 0.81 %), (1.29%, 1.18

%), (3.39%, 4.60 %), (3.55%,4.03%),(2.68%,2.56%), (4.58%,5.10%) and (0.60%,0.78%) respectively.

- The second descriptor, LB-3D-OACSP, was developed to address certain limitations of 2D arbitrary patterns by encoding image bit planes using multi-directional 3D arbitrary and circular-shaped patterns. Starting with the input image, a 2D circular symmetric Gaussian filter bank produces three Gaussian-filtered pictures, which are then decomposed into eight binary BPs. LB-3D-OACSP encodes inter-scale geometrical details across the BPs using 3D arbitrary and 3D circular scanning patterns in four directions. The resulting pattern image is partitioned into four equal-dimensioned, separate patches, from which features are extracted. Only 'uniform' patterns are computed and considered to reduce feature dimensionality. The LB-3D-OACSP's superior performance is ascribed to its application of multiscale pictures and the encoding of inter-scale bit planes. Despite having fewer dimensions, it demonstrates superior retrieval performance compared to many well-known descriptors, including bit plane-based and 3D descriptors, when tested on the NEMA-CT, TCIA-CT, and York-MRI image databases. For NEMA-CT dataset (at Tm=40), with regard to (%ARP and %ARR), the LB-3D-OACSP enhances upon LBP, LBDP, LDEP, LWP, LBPANDP, LBPDAP, CSLBCoP, SS-3D-LTP, 3D-LTCoP, and 3D-LOZFP descriptors by (6.86%,8.67%), (2.30%,4.55%), (12.06%,11.49%), (9.57%,10.35 %), (1.46%,1.6%), (1.90%,1.97%), (4.00%,5.42%), (3.55%,4.03%), (2.68%,2.56%), and (4.58%,5.10%) respectively.

4. In Chapter 4, two classification frameworks for automated diagnosis of COVID-19 from CXRs have been developed using DL-based techniques.

- Given the limited availability of openly accessible images of COVID-19-infected individuals, which can hinder the training of DL models, GAN-based data augmentation was employed in the first work, to address this challenge. Before augmentation, the input training images were pre-processed using CLAHE to enhance their quality. In this study, five pre-trained CNN models— Inception-v3, ResNet-50, VGG-19, DenseNet-201, and MobileNet-v2—were fine-tuned, and features were extracted from these models. To enrich the diversity of the features, the features extracted from two different models were combined, and these hybrid features were then classified using an SVM classifier. The blend of Inception-V3 and VGG-19 features among different combinations tested yielded the best results across three CXR image datasets. This framework demonstrated superior classification accuracy contrasted with several pioneering techniques for 3-class categorization in detecting COVID-19. For curated dataset (dataset 1), this framework has achieved sensitivity, specificity, precision, F1-Score ,and accuracy of 99.80%, 100%, 100%, 99.90%, and 99.47% respectively for COVID-19 detection.

- DL techniques typically extract features from raw spatial images but cannot often capture extremely fine details, which are crucial in biomedical image analysis. To address this limitation, in our second work, the proposed model combines deep features extracted from raw spatial images and with those derived from bit plane-based pattern maps to enhance the classification of COVID-19, pneumonia, and normal cases. This combination improves discrimination by complementing the strengths of each feature set. The framework leverages multiscale information from each BP and inter-scale features to generate the final BP-based pattern maps. Data augmentation uses six affine transformation schemes, and deep features are extracted using a fine-tuned DenseNet-201 model. The framework is evaluated on three publicly available CXR image datasets, and the results demonstrate superior classification accuracy compared to most existing methods for three-class classification of these conditions. For curated dataset (dataset 1), this framework has achieved a sensitivity, specificity, precision, F1-Score, and accuracy of 100%, 100%, 100%, 100%, and 100% respectively for three class categorization of COVID-19, Normal and Pneumonia.

5. In Chapter 5, DL-based techniques have been used to develop classification framework for automated diagnosis of SLs from dermoscopic images.

By employing the effective CycleGAN approach to the training samples, we were able to enrich the dataset with more training images which in turn reduced the risk of overfitting, which can occur when working with a limited training set. This framework modified three pre-trained CNNs namely, DarkNet-53, ResNet-50, and MobileNet-V2 to enhance robust feature extraction and accurate discrimination, followed by fine-tuning on skin lesion images. An ensemble strategy was employed to improve classification performance, combining predictions from the all the architectures to reach a consensus rather than relying on a single network's output. A majority voting technique was used to merge the individual decisions from each CNN model, resulting in a highly accurate final decision. The efficacy of this method was tested on two challenging SL datasets, ISIC 2018 and ISIC 2019. Comparative analyses demonstrated that the proposed ensemble model outperformed existing methodologies on both datasets. For ISIC 2018 dataset, our proposed framework is able to achieve sensitivity, specificity, precision, F1-Score, and accuracy of 95.18%, 99.12%, 96.88%, 96.02%, and 98.79% respectively for seven class categorization of skin lesions.

6.2 Future scope of research

While the biomedical image classification framework presented in this dissertation has shown promising results, there are still opportunities for improvement that could boost the effectiveness of future categorization tasks. Based on our findings, we have identified several potential directions for future research, which are outlined below:

1. Use of optimized feature selection techniques:

   Feature selection (FS) is crucial in ML and computer vision, particularly medical imaging. When extracting features from raw images, a wide range of patterns is generated, many of which are optional for the classification task. These irrelevant features can mislead classifiers and degrade overall performance. Implementing FS techniques aims to identify and retain the most significant features, thereby improving accuracy and reducing execution time. Research into the use of DL frameworks with more enhanced optimized FS strategies for COVID-19 diagnosis and skin lesion categorization problems, is an intriguing area to explore.

2. Use of improved semantic segmentation techniques:

   Semantic segmentation is an elementary and complex task in digital image examination. It involves the pixel-level labelling regions within an image that share standard high-level semantics. This process clusters parts of an image that belong to the same object class, effectively grouping similar elements. Timely detection and precise delimitation of lesion boundaries are crucial for the clinical treatment of SLs, particularly for accurately localizing cancerous regions. Accurate segmentation of SLs is essential for effective identification and classification using computational methods. Melanoma, the most dangerous form of SC, poses a notable challenge for dermatologists owing to the low contrast between the lesion and adjoining skin. Therefore, precise lesion boundary segmentation is vital for accurately locating lesions in dermoscopic images and diagnosing various skin lesions. SL segmentation is challenging due to several factors, including variations in skin tone, uneven lighting, partial obstructions from hair, low contrast between the lesion and surrounding skin, and the presence of freckles or gauze, which can be mistaken for lesions. These challenges significantly impact the accuracy of segmentation results, affecting all subsequent image analysis operations, particularly the classification of the object. It would be fascinating to look at better semantic segmentation techniques that can be included into our frameworks to make them more resilient to various restrictions. This would allow us to obtain better classification performance in the context of actual user needs .

3. Use of improved diffusion models:

   In recent years, generative modelling has made significant strides, driven by the development of advanced neural network architectures capable of learning and generating complex data distributions. These innovations have transformed fields such as image synthesis and opened new avenues for research and practical applications in computer vision and biomedical imaging. GANs introduced the adversarial training framework, which allows them to excel at fitting and reproducing narrow, unimodal distributions. However, they face challenges with more complex, multi-modal data. Despite issues like mode collapse and unstable training dynamics, GANs have set a new standard for creating realistic, high-quality images. While these models are already impressive, there remains considerable potential for further improvement beyond the existing state-of-the-art. Diffusion models are generative models that create samples by learning to remove noise from data points incrementally. Recent studies indicate that diffusion models surpass GANs in high-resolution image generation tasks, offering superior performance without issues like mode collapse and unstable training. They have also achieved remarkable results in conditional image generation. Stable diffusion models, a recent advancement, use iterative refinement techniques to generate high-resolution, semantically consistent images, addressing some of the limitations of GANs. However, despite their success, these models have inherent drawbacks, such as needing to be more computationally intensive and time-consuming. It would be interesting to look for improved diffusion models that can be combined with our frameworks in order to boost the overall classification performance.