Chapter 8

Conclusion and Future Directions

This chapter summarizes the key findings from the contributions in this thesis. Furthermore, it suggests promising future directions for research in the field of histopathological image classification.

8.1 Conclusion

A transformation in pathology is being witnessed as a result of the widespread adoption of digital pathology, which presents new horizons in the application of AI for cancer histopathology. Accurate identification and classification of benign and malignant tumour tissues is a vital clinical task, and automated procedures can substantially speed up the diagnosis, minimize the errors and contribute to better healthcare and treatment. Despite the enthusiasm and outstanding results that have been revealed so far, there are significant bottlenecks that prevent AI technologies from being easily implemented in cancer histopathology. The first bottleneck is the availability of well-annotated datasets of good quality and a substantial number of histopathology images. CMTs are an important neoplastic disease of dogs, and present diverse histological subtypes, requiring intense efforts of trained veterinary pathologists for diagnosis. Thus, CAD of CMTs is much needed, but to date, due to the lack of a publicly available dataset of CMT histopathological images, no studies have focussed so far in this direction. Availability of datasets consisting of well-annotated cancer histopathology images opens up new avenues for developing algorithms with higher accuracy and scope of application. Therefore, in this study, we have introduced a dataset of histopathological images of CMT, which is also an excellent model for HBC.

The research in this thesis was essentially focused on designing CAD for cancer histopathology classification using AI techniques. Further, the approaches proposed in this study have been evaluated on publicly available BreakHis datasets. The present study demonstrated effective solutions for the classification of HBC and CMTs with high accuracy and solves the concerns associated with deep learning-based approaches such as the requirement of large datasets, data security and privacy issues, as well as limited applicability on edge devices. In this study, we also presented an integrative approach of multidisciplinary concepts in the form of a novel feature descriptor and its classification by addressing the limitation of the machine learning technique. The present study addresses several bottlenecks in the application of AI-based approaches for cancer histopathology. The main concerns addressed and the major contributions of the present study can be summarized as:

1. Main bottleneck in the development of AI-based algorithms for cancer histopathology is the availability of larger well-labelled datasets. Before this study, to the best of our knowledge, no study was attempted for CAD of CMTs due to the non-availability of the CMT histopathology image dataset. In this study, we present a dataset of CMT histopathology images for the first time. The dataset represents well-annotated histopathological images collected from clinical cases of CMTs. Creation of the CMT dataset in this study was the first step towards the development of the framework for CAD of CMTs that will provide an opportunity for AI researchers to work in this area. Since CMTs are considered as excellent models for HBC, the dataset can be utilized in future for developing advanced algorithms

and programs for breast cancer histopathology as well.

2. For the creation of a robust and trustworthy feature set, a framework based on VGGNet-16, a popular CNN, has been proposed in this study. For binary classification of H&E stained cancer histopathology images, the proposed framework featured a VGGNet-16 fused model with SVM and RF classifiers. A novel deep feature extractor based on VGGNet-16 is presented and the features extracted from this were fed into a learning model for further classification. The model was tested on two datasets- a standard HBC dataset (BreakHis) and the new CMT dataset (CMTHis). Moreover, the effects of data augmentation, stain normalization, and magnification on the performance of the proposed framework were also analyzed.

3. Large-scale collection of data is critical for the effectiveness of the model, but issues related to data security and privacy makes data owners reluctant for contributing sensitive information in machine learning systems that contain multiple parties. This study also addresses such concerns related to confidential input data during training or inference, as well as to the sharing of a trained model. To address these, an efficient and secure cancer diagnostic framework for histopathological image classification is proposed by utilizing both differential privacy and secure multiparty computation. The model proposed in this study efficiently achieves the trade-off between privacy and model performance and reduces the risk of medical data leakage. Differential privacy, secure multiparty computation, deep learning and transfer learning are combined together for solving histopathology cancer classification while maintaining data privacy and model security.

4. Another bottleneck of deep learning-based approaches is poor applicability on edge devices because of requirements of high computation and huge parameters. In this study, to the best of our knowledge, for the first time, a model was de-

veloped for cancer histopathology that works effectively on edge devices. MobiHisNet, an efficient and lightweight deep learning model developed in this study, uses a series of depth-wise separable convolutions to reduce computational parameters. The model was quantized with FP32, FP16, and INT8. The model proposed here is also tested on a Raspberry Pi and three different smartphones to prove its performance on a lightweight processor. The model proposed herewith speeds up histopathological image classification significantly without sacrificing accuracy and allowing it to compete with other models for applications on edge devices. The model demonstrated the superior performance of MobiHisNet on edge devices in terms of higher accuracy, lesser complexity, and lesser memory requirements.

5. The requirement of massive amounts of data, which is generally not available in most cases, is another stumbling block for the applicability of deep learning-based approaches for cancer histopathology. Machine learning-based algorithms are a potential alternative to deep learning for medical imaging with smaller datasets. However, issues such as the requirement of complex processing for feature extraction makes machine learning techniques less popular. Thus, this work presents an interdisciplinary approach by integrating illumination invariant feature extraction, CoM and fuzzy theory for efficient feature engineering. The study presents a novel feature extraction technique by optimizing the block size to extract image micro-patterns and computing CoM for each pixel to extract feature vectors. For making the model more robust and to handle uncertainties in the image data, the concept of fuzzy theory along with SVM is applied in this study. The proposed framework achieves a classification accuracy of 97.25% using an FSVM-WD classifier, outperforming both traditional machine learning and deep FE-VGGNET16-based feature descriptors for CMTs. 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.

8.2 Future Directions

It is worth mentioning that the healthcare industry revolution has created a plethora of opportunities in histopathology cancer diagnosis. The following are promising future directions that can be explored based on the research work presented in this thesis.

1. In the near future, we can extend this with other potential deep frameworks with an attention mechanism. We can also integrate this framework with other private prediction techniques. Another possibility is to extend this framework for the multi-class classification. This framework may be generalized for other applications as well in the medical field.

2. The quality of features in a classification model plays a crucial role. Initially, we utilized the exceptional representation learning ability of deep learning in our study. We can incorporate feature selection before classification. Since feature selection is a multi modal problem in search space, we can consider it in the future learning process. Further, the whole feature extraction and classification problem can be reformulated as a "multi-modal multitask learning" to get an efficient framework for CAD which is still unexplored in histopathological cancer image classification problem.

3. This study mainly focuses on valuable feature extraction from complex tissue images and their classification by designing efficient models either focusing on CNN or other ML-based approaches under different constraints. For this, interdisciplinary concepts are utilized to overcome the limitations of huge data sets and

the high compute resource requirements of deep models. Further, security concerns have been addressed as well as a compressed model requirement for smart health technology was also presented. But we can view the problem after feature extraction from another perspective of optimization, where we can formulate the feature selection and classification as multi-objective optimization, which is still in the exploration phase in medical image analysis.

4. This study only consider histopathological image data for analysis. It will be quite interesting and challenging to use "multi-modality information fusion" for histopathological cancer diagnosis. In this case, we can also analyze the genomic data of the same patient along with the features of a histopathological image extracted by proposed CoMHisP or deep features for classification. In such cases, it would be advantageous to explore multitask learning as a multi-objective optimization problem.

5. Future studies should focus on how to adapt the proposed framework to solve the multi-class problem efficiently by designing a more robust membership function. Moreover, fuzzy relations can be investigated in the proposed CoMHisp.

6. In the near future, it would be interesting to develop a more efficient and robust algorithm for edge-based histopathological image classification by considering component-wise computational costs as well as energy consumption.

7. Availability of labelled data is one of the major challenges in developing a classification model; thus, an unsupervised technique can be adopted at the first stage. Deep active learning is still unexplored in this area that will also help in the auto-annotation of data and reduce the burden on pathologists for manual data labelling.

8. The proposed approaches are generalized so that they can be used for solving different medical image classifications as well. From the perspective of secure

modelling, our proposal is based on the assumption that server is honest, however in some cases it may be untrue. Hence, we can be made more secure and efficient model by utilizing the advanced security techniques like Knowledge distillation with federated learning can be explored for the secure classification of histopathological cancer images.