CERTIFICATE

It is certified that the work contained in the thesis titled ***"Artificial Intelligence Driven Techniques for Cancer Classification using Histopathological Images"*** by ***Abhinav*** has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all requirements of Comprehensive Examination, Candidacy, and SOTA for the award of Ph.D. Degree.

Co-Supervisor

Dr. Lakshmanan Kailasam

Assistant Professor

Department of Computer Science and Engineering,

Indian Institute of Technology (BHU) Varanasi,

Uttar Pradesh, INDIA 221005.

Supervisor

Prof. Sanjay Kumar Singh

Professor and HOD

Department of Computer Science and Engineering,

Indian Institute of Technology (BHU) Varanasi,

Uttar Pradesh, INDIA 221005.

DECLARATION BY THE CANDIDATE

I, ***Abhinav***, certify that the work embodied in this Ph.D. thesis is my own bonafide work carried out by me under the supervision of ***Prof. Sanjay Kumar Singh*** from ***July 2017*** to ***December 2021*** at ***Department of Computer Science and Engineering***, Indian Institute of Technology (BHU) Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, *etc.* reported in journals, books, magazines, reports, dissertations, theses, *etc.*, or available at websites and have not included them in this thesis and have not cited as my own work.

Date:

Place: Varanasi

Abhinav

(Abhinav)

CERTIFICATE BY THE SUPERVISOR

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

(Prof. Sanjay Kumar Singh)
Professor and HOD
Dept. of Computer Science and Engineering
Indian Institute of Technology (BHU) Varanasi

पर्यवेक्षक/Supervisor
संगणक विज्ञान एवं अभियांत्रिकी विभाग
Dept. of Computer Sc. & Engg.
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(काशी हिन्दू विश्वविद्यालय)
(Banaras Hindu University)
वाराणसी/Varanasi-221005

Signature of Head of the Department

21/12/21

आचार्य व विभागाध्यक्ष
Professor & Head
संगणक विज्ञान एवं अभियांत्रिकी विभाग
Department of Computer Sc. & Engg
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(काशी हिन्दू यूनिवर्सिटी)
(Banaras Hindu University)

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: Artificial Intelligence Driven Techniques for Cancer Classification using Histopathological Images

Name of the Student: Abhinav

Copyright Transfer

The undersigned hereby assigns to the Indian Institute of Technology (Banaras Hindu University) Varanasi all rights under copyright that may exist in and for the above thesis submitted for the award of the *Doctor of Philosophy*.

Date: 20.12.2021

Place: Varanasi

(Abhinav)

This thesis is dedicated to my beloved family.

For their endless love, support and encouragement

ACKNOWLEDGEMENT

Apart from the efforts of ours, the success of any work depends largely on the encouragement and guidelines of many others. I take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this work. First of all, I would like to thank the Almighty for his unlimited favours, endless mercies, and choicest blessings bestowed upon me that helped me to come across all the way and imparadise.

At this stage, I owe my indebtedness to my supervisor, Prof. Sanjay Kumar Singh, Head of the Department, Computer Science and Engineering. He has been the torchbearer of my Ph.D. work. I express a deep sense of reverence to him for his invaluable guidance, constant encouragement, constructive suggestions, sagacious advice and keen supervision during the entire duration of my Ph.D. To have such an ingenuous perfectionist as my guide was a privilege for me as he gave new dimensions to my knowledge. A formal statement of acknowledgement will hardly meet the sense in the matter of expression of enormous debt I owe to him.

I further express my thankfulness to Dr Sonal and Dr Sameer Shrivastava, Department of Biotechnology, IVRI, for their prudent suggestions and for the kind consideration and cooperation which they extended whenever I approached them. A mere thanks and words of appreciation cannot compensate for their wholehearted support, incessant efforts, timely help, valuable suggestions, cordial encouragement, and generosity offered.

I would also like to take this opportunity to express my sincere regards to RPEC members Prof. Subir Das and Dr. Ravi Shankar Singh for their constant support and encouragement. I am also thankful to my co-supervisor Dr. Lakshmanan Kailasam, for the guidance and generosity offered. Regards are also due to all faculty members and staff of the Computer Science and Engineering Department for their kind support

and the necessary facilities provided. Further, I extend my special gratitude to my colleagues and friends, Dr. Anshul, Dr. Rishav Singh, Dr. Sandeep Sambhaji Udamle, Miss. Vandana, Mr. Ritesh Sharma, and Mr. Aneesh G Nath who supported me throughout my thesis work and provided technical, moral, and emotional support during my research work. I want to express my deepest gratitude to them for providing moral support throughout my thesis work. I have also spent some of the craziest and most memorable moments with them.

From the core of my heart, I thank my family for all that they have done in bringing me to this platform, and no words can describe my feeling towards them. My parents have been the driving force behind me, taking all pains to illumine my path and revive my energy. Their boundless love and encouragement are and always will be a constant source of motivation for me in my life. The constant encouragement, love and affection of my family have always been a reservoir of the impetus for me.

(Abhinav)

Contents

List of Figures

List of Tables

Abbreviations

Abbreviation	Description
AI	Artificial Intelligence
AUC	Area Under Curve
CAD	Computer-Aided Diagnosis
CMT	Canine Mammary Tumour
CNN	Convolutional Neural Network
CoM	Center of Mass
DP	Differential Privacy
DPSGD	Differentially Private Stochastic Gradient Descent
FC	Fully Connected
GPU	Graphics Processing Unit
HBC	Human Breast Cancer
H&E	Hematoxylin and Eosin
IoT	Internet of Thing
LBP	Local Binary Patterns
MPC	Multi-party Computation
OD	Optical Density
RF	Random Forest
ROI	Region of Interest
SD	Standard Deviation

Abbreviation	**Description**
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SMPC	Secure Multi-party Computation

List of Symbols

Symbol	Description
ϱ	Angular feature vector
C_R	Compression rate
d_{jk}	Distance of j^{th} sample to the k^{th}neighbour
λ	Eigen value
K_i	Feature vector of i^{th} sample image
$\vec{F}$	Force acting on the system
g_t	Gradient
C_u	Gradient norm bound
G	Gradient of Center of Mass
g_x	Grey value of pixel
$L(i,j)$	Illuminance component
α	Lagrangian multiplier
γ_t	Learning rate
$L_1()$	l_1-norm
m_i	Mass of i^{th} particle
μ	Mean value
Υ_i	Membership value of i^{th} sample
t_h	Number of corrupted participants
p	Number of neighbour pixel

Symbol	**Description**
σ	Noise scale
j	Objective function
C_{dws}	Overall operation cost for depth-wise convolutions
$\vec{r}_{cm}$	Position vector of Center of Mass
$h(.)$	Pooling function
g_c	Reference gray value
$R(i,j)$	Reflectance component
$\rho()$	Relative density between two instances
C_{std}	Standard convolutional cost
s_f	Sensitivity
ξ_i	Slack variable for the i^{th} sample
T	Threshold value
D_i	Tissue images corresponding to i^{th} patient
γ	Width multiplier

Preface

Cancer is among the most perplexing ailments that has tormented humans for much of our civilization. In both humans and dogs, tumours of the mammary gland are most common among females. Both human breast cancer (HBC) and canine mammary tumours (CMTs) are associated with high mortality rates and poor prognosis. Moreover, the disease is similar in both the species and hence, CMTs are considered excellent models for HBC studies. Timely classification of cancers is crucial for deciding the most appropriate therapeutic regime and prognosis. However, many a times, there is a delay in diagnosis due to the paucity of trained oncopathologists. 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 errors and contribute to better healthcare management. Thus, the development of automated computer-aided diagnosis (CAD) systems is a priority research area for image enhancement, feature extraction, and classification of cancer histopathological images.

The increasing adoption of deep learning and machine learning across healthcare domains, together with the availability of highly characterized cancer datasets, has accelerated research into the utility of artificial intelligence in the analysis of the complex biology of cancer. While early results are promising, this is a rapidly evolving field with new knowledge emerging in both cancer biology and deep learning. Deep learning-based approaches are recently being explored for analyzing histopathological images of HBC. However, so far, due to the lack of any publicly available CMT database, no studies

have focused on the automated classification of CMTs. The existence of heterogeneous and diverse types of CMTs and the paucity of skilled veterinary pathologists justify the need for automated diagnosis. Thus, the study was proposed with the aim to develop a framework for CAD of CMTs with high accuracy. To the best of our knowledge, for the first time, a dataset of CMT histopathological images (CMTHis) was introduced in this study. Further, a VGGNet-16 as a feature extractor was proposed, wherein a FC layer was removed and experimented with various classifiers.

As the success of deep learning depends on the amount of data available, the large-scale collection of data is critical for the effectiveness of the model. However, privacy and security concerns often prevent data owners from contributing sensitive information in machine learning systems that contain multiple parties. Thus, another objective of this study was to address such concerns related to confidential input of 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 in this study by utilizing both differential privacy and secure multi-party computation.

Another area of concern addressed in this study is the poor applicability of deep learning-based approaches on edge devices. Moreover, deep learning-based approaches are computationally expensive and have huge parameters, which makes them less affordable for edge devices. Recently in the edge computing world, providing a low-cost solution has become highly valuable. To the best of our knowledge, this is the first study to develop a model for histopathological image classification that works effectively on edge devices. The proposed model was tested on a Raspberry Pi and three different smartphones to demonstrate its efficiency on a lightweight processor. The model proposed herewith speeds up histopathological image classification significantly without sacrificing accuracy, allowing it to compete with other models for applications on edge devices.

The requirement for massive amounts of labelled data is a major limiting factor for deep learning-based approaches. In the field of medical imaging, such large labelled datasets are difficult to acquire. 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. Traditional machine learning algorithms have limited ability to reveal the most sophisticated features of cancer histopathological images, but their robustness and fault tolerance can be enhanced by using fuzzy modelling to capture the uncertainty in image data. Therefore, in this study, a novel CoMHisP framework based on a fuzzy support vector machine with within-class density information was also proposed. It utilizes a novel feature extraction technique by optimizing the block size to extract image micro-patterns and computing centre of mass (CoM) for each pixel to extract feature vectors. 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.

The approaches, proposed in this study, have been evaluated on publicly available BreakHis dataset and newly introduced CMTHis dataset. 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.