ARAB ACADEMY FOR SCIENCE, TECHNOLOGY
AND MARITIME TRANSPORT

College of Engineering and Technology

Computer Engineering Department

Staging of Clear Cell Renal Cell Carcinoma using Random Forest and Support Vector Machine

By

Doha Talaat Abd Al-Ghany Mohamed

B.Sc. Computer Engineering AASTMT, 2014

A thesis submitted to AASTMT in partial

Fulfillment of the requirements for the award of the degree of

MASTER OF SCIENCE

in

COMPUTER ENGINEERING

Supervisors

Dr. Fatma Mohamed Zada Dr. Rania Ahmed Kadry

Computer Engineering Department

College of Engineering and Technology

Arab Academy for Science and Technology and Maritime Transport

Alexandria

2020

ACKNOWLEDGMENT

This thesis is the culmination of my journey to master degree, which was just like climbing a high peak gradually accompanied with encouragement, hardship, trust, and frustration. When I found myself at top experiencing the feeling of fulfillment, I realized though only my name appears on the cover of this dissertation, a great many people including my family members, well-wishers, my friends, colleagues and various institutions have contributed to accomplish this huge task.

I thank all whom in one way or another contributed in the completion of this thesis. First, I give thanks to GOD for protection and power to finish this work.

I would like to express my sincere gratitude to my supervisor Dr. Fatma Zada for the continuous support of my thesis study and research, for her patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor and mentor for my master study.

Besides my supervisors, I would like to thank Dr. Farouk Hamoud and Dr. Mohamed Masoud, who helped me in the medical part of the thesis.

My sincere thanks and deepest gratitude goes to my second supervisor Dr. Rania kadry for her support, help, interest and valuable hints. Her insightful suggestions, comments and encouragement helped me in the time of writing and research of this thesis.

I would like to thank my Dad and Mom for their everlasting support and devotions that encouraged me to pursue my academic career. I'm grateful for their patience, attention and continuous understanding that helped me move forward and chase my dreams no matter how impossible they seem.

I owe thanks to my very special sisters, for showing faith in me and for their selfless love, care and dedicated efforts, which contributed a lot for completion of my thesis. Thanks for always being there for me encouraging me to be better.

I consider myself the luckiest in the world to have such a lovely and caring family, standing beside me with their love and unconditional support.

ABSTRACT

Kidney cancer is one of the deadliest types of cancer affecting the human body. It's regarded as the sixth most common type of cancer affecting men and the eighth affecting women. Early diagnosis of kidney cancer (renal cancer) can improve the survival rates for many patients. Renal cell carcinoma is the most common type of renal cancers which is subtyped into: clear cell renal cell carcinoma, papillary renal cell carcinoma, and chromophobe renal cell carcinoma. Renal cell carcinoma (RCC) accounts for 90% of kidney cancers. Although the exact cause of the kidney cancer is still unknown, early diagnosis can help patients get the proper treatment at the proper time.

Determining the clear cell renal cell carcinoma stage helps in receiving the appropriate treatment. The earlier the stage of the disease. The better the treatment is. Most of the previous work depends on clinical and medical ways in determining the ccRCC stage using the computed tomography devices.

In this thesis, an enhanced automated model is proposed for early detection and staging of clear cell renal cell carcinoma. The proposed model consists of three phases: segmentation, feature extraction, and classification. The first phase is image segmentation phase where blobs detection was used to segment the kidney lobes. Then the segmented kidneys images were fed to watershed algorithm to extract tumor from the kidney. The second phase is feature extraction phase where gray level co-occurrence matrix method was integrated with normal statistical method to extract the feature vectors from the segmented images. The last phase is the classification phase where the resulted feature vectors were introduced into random forest and support vector machines classifiers.

Experiments have been carried out to validate the effectiveness of the proposed model using TCGA-KRIC dataset which contains 228 CT scans for the four stages of ccRCC where 150 scans were used for learning and 78 for validation. The proposed model showed an outstanding improvement of 12.26% for accuracy and 13.49% for specificity.