OVERALL SURVIVAL PREDICTION OF GLIOMA PATIENTS USING GENOMICS

M. R. Navodini Wijethilake

208012F

Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree Master of Science by Research

Department of Computer Science & Engineering

University of Moratuwa

Sri Lanka

September 2021

DECLARATION

I, Navodini Wijethilake, declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature ***UOM Verified Signature*** Date: 23/09/2021

The above candidate has carried out research for the Masters thesis Dissertation under my supervision.

Name of Supervisor: Dr. Dulani Meedeniya

Signature of the Supervisor: Date: 23/09/2021

The above candidate has carried out research for the Masters thesis/Dissertation under my supervision.

Name of Supervisor: Dr. Charith Chithraranjan

Signature of the Supervisor: ***UOM Verified Signature*** 21

ABSTRACT

Overall survival prediction is a vital task that will lead for better patient management in clinical practise. Existing approaches mainly focus on imaging based survival prediction, which is non invasive, and thus, easier to be implemented at the initial diagnosis stages. However, the advancements in the DNA/RNA technologies has given access to genomic and transcriptomic profiles of the gliomas, that directly reflect the molecular level alterations. Thus, in this work we mainly focus on using transcriptomic profiles for survival prediction, an area that has not been widely analysed yet for survival prediction. We utilize the gene expression and mutation profiles, while augmenting the recent Artificial Intelligence approaches, such as deep probabilistic programming and multi task learning for prognosis prediction. Thereby we do not just focus on the application, we also contribute with novel learning paradigms to improve the classification task performances. Nonetheless, we also focus on proposing a novel loss function, since architectural wise the state of art performance has been achieved for classification tasks.

In addition, we also investigate ability to employ radiomics, for subtype classification, that is also associated with survival. Since subtypes mainly rely on the genomic alterations, we found it useful to focus on imaging features ability to predict prognosis of glioma.

Keywords: Survival Prediction; Prognosis; Deep Learning; Multi-task Learning; Deep Probabilistic Programming

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my supervisors Dr. Dulani Meedeniya and Dr. Charith Chithraranjan for the immense guidance provided to successfully finish this research. I'm extremely thankful for your patience, motivation, and immense knowledge given throughout the period. I firmly believe without your courageous support this would not be able to reach this stage.

I wish to thank Dr. Indika Perera for his valuable insights and guidance from the very beginning of this research. I would like to convey my gratitude to the entire staff of the Department of Computer Science and Engineering, both academic and non-academic for all their support given throughout the entire Masters course period. This research was supported by the University of Moratuwa Senate Research Grant. I would like to acknowledge the grant and other relevant parties who work hard to provide the required facilities to up bring the research facilities in Sri Lanka more specifically, by providing the financial support.

Nonetheless, I would like to thank my family for all the love and support.

Thank you!

LIST OF ABBREVIATIONS

GBM	Glioblastoma
DL	Deep Learning
ML	Machine Learning
IR	Importance Ranking
PCA	Principle Component Analysis
RFE	Recursive Feature Elimination
CC	Correlation Coefficient
RF	Random Forest
LR	Linear Regression
SVM	Support Vector Machine
XGB	eXtreme Gradient Boosting
ST	Survival Tree
ANN	Artificial Neural Network
LASSO	Least Absolute Shrinkage and Selection Operator
KM	Keplen Meier
TCGA	The Cancer Genome Atlas
CNN	Convolutional Neural Network
CGGA	Chinese Glioma Genome Atlas
OBTS	the Ohio Brain Tumor Study
GEO	Gene ExpressionOmnibus
CPTAC	Clinical Proteomic Tumor Analysis Consortium
TCIA	TheCancer Imaging Archive
LGG	Lower Grade Glioma
CoxPH	Cox Proportional Hazard model
CE	Cross-Entropy
FL	Focal Loss
LSCE	Label Smoothing Cross-Entropy
LSF	Label Smoothing Focal

normLSF	Normalized Label Smoothing Focal
Acc.	Accuracy
Prec.	Precision
Sens.	Sensitivity
MTL	Multi Task Learning

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES