HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 1 (2024)

Deep Learning Framework for Diagnosis and Survival Prognosis of Central Nervous System Tumors

Diya Sreedhar[1] and Ashok David[#]

[1]Troy High School, USA
[#]Advisor

ABSTRACT

Diagnosis and prognosis of central nervous system (CNS) tumors rely on manual and algorithmic approaches that are subject to variations, inefficiencies and bias. With 5-year survival rates as low as 6%, timely tumor assessment is crucial to ensure patient health. Machine learning and deep learning techniques have been used over the past decade to reduce the assessment burden and augment physician diagnoses. Traditional machine learning models require manual feature extraction and engineering, introducing further variability of the results. On the other hand, automated deep learning models require complicated engineering and workflow implementations that need to be managed and updated for new datasets. This study endeavors to define a generalizable, two-phase neural network pipeline and user interface that can be applied to clinical assessment of any tumor with 3D MRI scans, allowing for improved personalization of patient diagnosis and treatment options. In the first diagnostic phase, radiomic features are used to predict tumor severity and extent. The outputs of this phase are then passed to a survival prognosis model along with the patients' clinical data to predict the 5-year overall survival rate. The diagnostic model achieved a F1-Score, a measure of classification accuracy, of 94% and the prognostic model achieved a risk-adjusted, time-dependent Harrell's Concordance Index score of 0.92, indicating the framework's generalizability to new datasets. The mobile app developed as part of this study offer ease of access to physicians and radiologists in reviewing predicted results and subsequent patient interactions.

Introduction

Diagnosis and intervention at early stages of central nervous system (CNS) cancers are associated with improved patient outcomes and increased chances of survival. Currently, magnetic resonance images (MRIs) remain the gold standard for non-invasive tumor diagnosis and pre-operative severity assessment (Meola et al, 2018). However, a global shortage of radiologists (Henderson, 2022; European Society of Radiologists, 2022), especially in many rural areas and developing countries, and a lack of access to advanced MRI analysis techniques (Frija, et al., 2021; Wuni et al 2021) is leading to suboptimal patient outcomes. Additionally, manual segmentation of MRI images is subject to observer bias, as different radiologists may analyze tumors with different methods (Maskell 2018).

With recent advances in artificial intelligence, researchers have leveraged deep learning for tumor segmentation and classification (Pei et al., 2020; Palson et al, 2022), while others have implemented traditional machine learning based approaches to predict patient survival from clinical data analysis (Moreau et al., 2020). Leveraging radiomic image features along with patient clinical and demographic data offers an integrated approach to assessing tumor severity and extent, while also accurately predicting the patient survival rate (Cousin et al, 2023).

Radiomics refers to the extraction of quantitative image features from medical image scans such as MRIs (Zhang et al, 2023). These image features, which are not visible to the naked eye, are often not considered in tumor diagnosis and survival prognosis. Radiomics techniques offer high-dimensional, artificially engineered features that can aid in improving the prediction accuracy and precision of machine learning models in clinical oncology (Xue et al, 2021). However, there are concerns around the variability and reliability of these features due to the inherent variance in MRI images, image processing techniques, model development methodology, and workflow implementation (Xue et al, 2021).

This study aims to address existing challenges prevalent in tumor severity assessment and patient survival prognosis by combining both tumor classification and patient survival prediction into a single unified deep learning workflow, where the radiomics features extracted from 3D MRI images are used to predict tumor extent, severity and overall survival probability. Developing an integrated, generalized deep learning workflow to perform both diagnosis and prognosis offers a simplified framework that can be tested and leveraged for multiple clinical imaging applications and extended to support oncological assessments for various cancers.

Methods

Data Acquisition and Patient Selection

This study used 3D MRI image data and deidentified patient clinical data from *The Cancer Imaging Archive's* (TCIA) *REpository for Molecular BRAin Neoplasia DaTa* (REMBRANDT) collection (Clark K et al., 2013). The REMBRANDT dataset consists of patient MRI scans in the *Digital Imaging and Communications in Medicine* (DICOM) format, and the associated patient's clinical data in Excel format. The REMBRANDT project was created at the *National Cancer Institute* and funded by the *Glioma Molecular Diagnostic Initiative*. The data was collected from 2004-2006 from two institutions, Thomas Jefferson University and Henry Ford Hospital and consists of 128 patients. 32 patients were excluded from this study as they had no clinical or demographic data to aid in survival prognosis. Of the remaining 96 patients, an additional 16 were excluded as they had no labeled MRIs scans. The final study cohort consisted of 80 patients that had both labeled MRI scans and associated clinical data.

Exploratory Data Analysis

An initial exploration of the clinical data associated with the REMBRANDT MRI images was conducted to assess their quality and usability. The data covers three CNS tumor types: Glioblastoma, Astrocytoma, and Oligodendroglioma. Patients range in age from 10 years to over 85 years. Other important treatment information is also included, such as type of radiation therapy received, patients' Karnofsky Performance Status (KPS), type of chemotherapy received, patient vital status (alive or deceased) at the end of the study, and the number of survival months from initial diagnosis. The dataset was split into training (50%), validation (25%), and test (25%) subsets.

Deep Learning Workflow

A two phase deep learning framework was designed and developed as part of this study. Prior to model training, the 3D MRI scans were preprocessed to normalize image size and pixel intensity. In the first phase of the deep learning framework, the normalized scans, along with the clinical data were used as inputs to train a Convolutional Neural Network (CNN) model that predicted tumor type, tumor grade and the Karnofsky Performance

Status (KPS). In the second phase, these predicted outputs from the classification model, along with the radiomic features and patient clinical data were fed into a survival prognosis model. The trained models were then deployed on the Amazon Web Services (AWS) cloud platform for inference using a Flask application server. An iOS mobile app and a web app were developed as two separate user interfaces, allowing a physician or radiologist to upload patient MRI scans and review predicted results. (Figure 1).

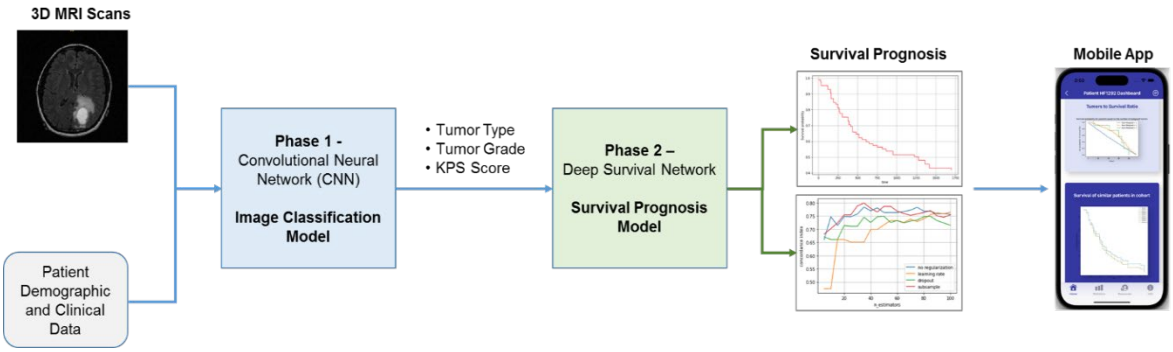


Figure 1. Overview of the Data Processing, Deep Learning and User Interface Pipeline

Phase 1 - Tumor Classification

Convolutional Neural Network (CNN) models are a class of deep learning models used in radiology applications to segment and classify tumors. This class of models adaptively learns spatial hierarchies of radiomic image features through multiple building blocks or layers, such as convolution layers, pooling layers, and fully connected layers (Yamashita et al., 2018). ResNet, short for *Residual Network,* is a class of CNN models that have found application in medical classification tasks and are shown to perform better than other CNN models (Bressem et al., 2020). In this phase, a ResNet neural network class was constructed, and a forward propagation function was defined. For this study, seven different Resnet models were trained on the REMBRANDT dataset, and their performance was evaluated – Resnet-10, Resnet-18, Resnet-34, Resnet-50, Resnet-101, Resnet-152, and Resnet-200. The seven ResNet models were each trained for 50 epochs, or training instances, keeping the hyperparameters constant. The training loss, validation loss, accuracy, precision, F-1 score, and recall for each model and each training epoch were collected in arrays for plotting. All models were trained on a Google Colab Pro+ platform, with a V100 GPU, 84 GB RAM and 250GB disk storage space. Model training optimization was conducted using the cross-entropy loss function (Equation 1).

$$L = -\frac{i}{m}\sum_{i=1}^{m} y_i . \log\left(\hat{y}_i\right)$$

Equation 1. Cross-entropy loss function to measure the difference between predicted class probabilities and true class labels.

Accuracy for classification models is defined as the ratio of correctly predicted instances to the total number of instances. Although accuracy is commonly used to measure performance, it is not the best indicator when there is a skew in the dataset. To account for the skew in the REMBRANDT dataset, relevance metrics including *Precision*, *Recall* and *F1-Score* were calculated to assess model performance. Precision, also known as the positive predictive value, is defined as the fraction of relevant instances among the number of retrieved instances. Recall, also known as *Sensitivity*, is defined as the fraction of relevant instances that were retrieved.

F1-Score, also known as the harmonic mean of Precision and Recall (Equation 2) was used to measure model performance.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Equation 2. F1 Score, the harmonic mean of a model's precision and recall

Phase 2 - Survival Prognosis

Two machine learning models, Cox Proportional Hazards and Random Survival Forest and one deep learning model, Deep Survnet were trained on the REMBRANDT dataset for survival prediction. The most commonly used linear covariate regression survival model is the Cox Proportional Hazards model (George et al, 2014; Matsuo et al, 2018; Houwelingen et al, 2015). This was compared with the non-linear covariate Random Survival Forest model, which has been used in studies for overall survival prediction of cancer patients (Kalafi et al, 2019; Roshanaei et al, 2022; Amabale-Venkatesh et al, 2017). Deep Survnet was chosen to be the model for survival prognosis prediction that was integrated into the final framework since it yielded the best predictive performance.

To evaluate the prediction performance of survival models against right-censored data, in which the outcome is unknown, Harrell's and Uno's concordance index (c-index) was measured (Kremer, 2007). The Integrated Brier Score (IBS), a metric which is an extension of the mean squared error to account for right-censored data (Goldstein-Greenwood, 2021), was used to measure the overall performance of the models between specific time points. The IBS score (Equation 3) was calculated between the 10th and 90th percentile of data and a 5-fold cross validation technique was used to reduce overfitting. The models were tested against an unseen subset of the patient clinical data, and images taken from the test set and their predictions compared against the ground truth classifications of tumor grade, severity, and KPS provided by physicians at the Henry Ford Hospital.

$$Integrated\ Brier\ Score = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2$$

Equation 3. The integrated weighted squared distance between the estimated survival function and the empirical survival function.

Model Deployment for Inference

The tumor classification and the survival prognosis models were trained and evaluated on the Google Cloud platform. Google Colab, a Python Notebook interactive development environment was used to program the models. The trained models were saved as Pickle files (.pkl) for subsequent usage in generating predictions or inferences. The saved model files were then transferred to an Elastic Cloud Compute (EC2) instance on the AWS cloud platform. The Flask application server software was used to develop a model inference service that supported Application Programming Interfaces (API) calls to run the model prediction and return the results to the user.

User Interface Development

Two separate user interfaces were developed to allow physicians and patients to interact with the models. The first user interface is a web based application that can be accessed through a browser on a personal computer. This allows the technician to upload the MRI images and clinical data through a secure website for processing and prediction. The second user interface is a mobile iOS app that can be accessed through a tablet or phone. Physicians would be able to use this to view the prediction results and for interactions with the patients.

The frontend for the web app was designed and developed using ReactJS libraries and HTML5. Data exchange for image and clinical data uploads was handled through Representational State Transfer (REST) API calls, which allow two systems to exchange data securely and reliably over the internet using the REST architecture guidelines. The mobile app was developed using the Swift language and XCode software (Figure 2). Data was exchanged with the APIs in JSON format using the SwiftyJSON library. The API interactions in the mobile app were handled through the AlamoFire software library.

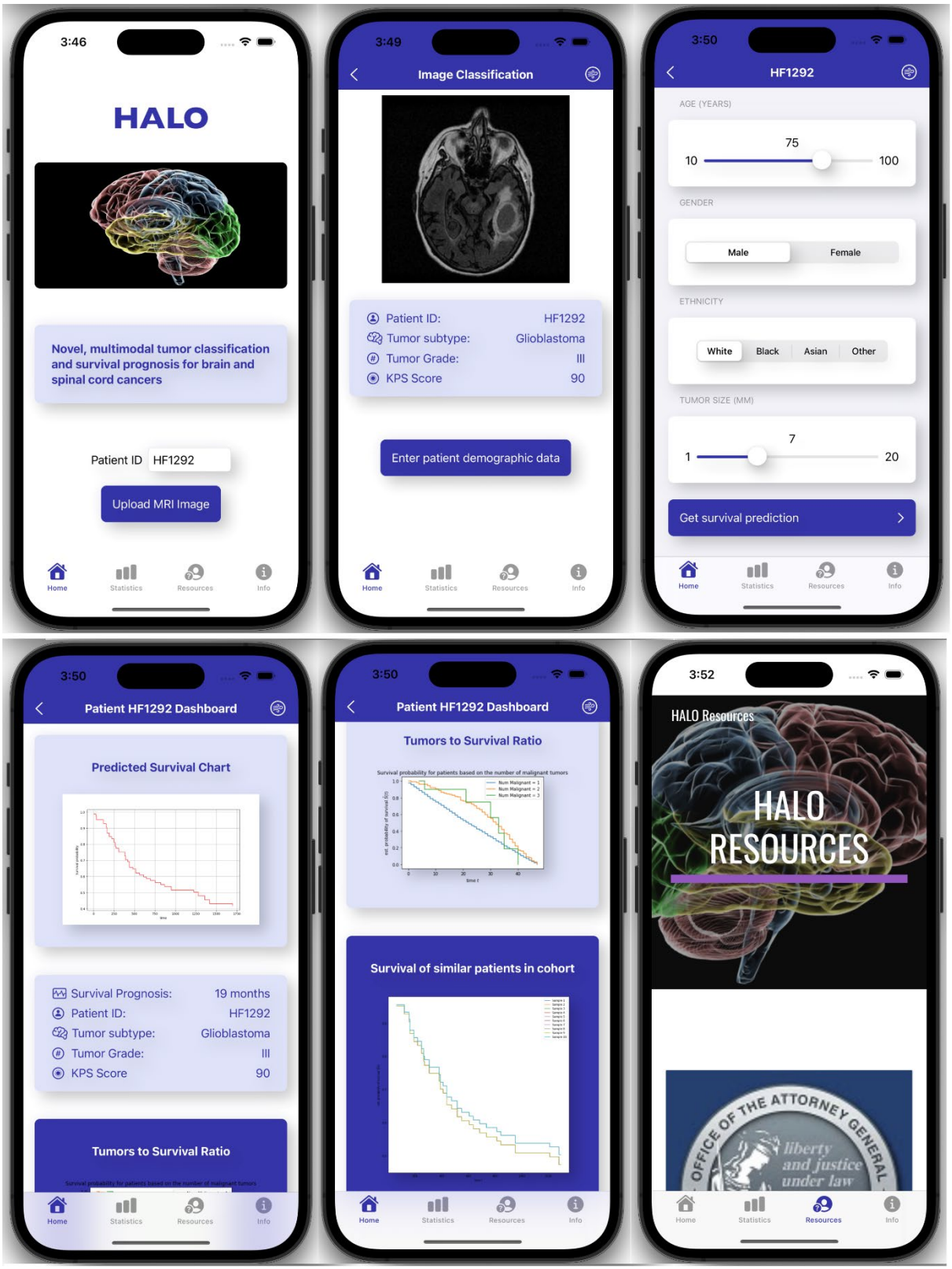


Figure 2. iOS Mobile App Interface Showing Predicted Diagnostic and Prognostic Results

Results

Phase 1 - Tumor Classification

All seven trained ResNet classification models were used to run inference against the test dataset of twenty patients and results were documented. The Accuracy, Precision, Recall and F1-Score were calculated to evaluate model performance (Table 1). The ResNet-18 model was observed to have the highest F1-Score of 0.94 and was selected as the final model for inference.

Table 1. Model Performance Metrics of ResNet Classification Models on the Test Dataset

Model	Accuracy	Precision	Recall	F1-Score
Resnet-10	0.64	0.52	0.53	0.5249
Resnet-18	0.94	0.94	0.95	0.9449
Resnet-34	0.72	0.88	0.9	0.8898
Resnet-50	0.75	0.89	0.91	0.8998
Resnet-101	0.78	0.67	0.69	0.6798
Resnet-152	0.52	0.42	0.42	0.4200
Resnet-200	0.86	0.92	0.94	0.9299

A multiclass confusion matrix was constructed from the predicted results of the ResNet-18 model for each of the three prediction classes, tumor type (Figure 3a), tumor grade (Figure 3b) and KPS (Figure 3c). It was observed that the classification model accurately predicted the Glioblastoma tumors in 100% of the test cases, followed by Astrocytoma (73%) and Oligodendroglioma (50%). The model predicted tumor Grade 4 accurately in 100% of the test cases, Grade 3 in 67% of the test cases, and Grade 2 in 67% of the test cases. The predicted KPS was most accurate for KPS-100 at 90%, followed by KPS-90 at 67% and KPS-80 at 67%.

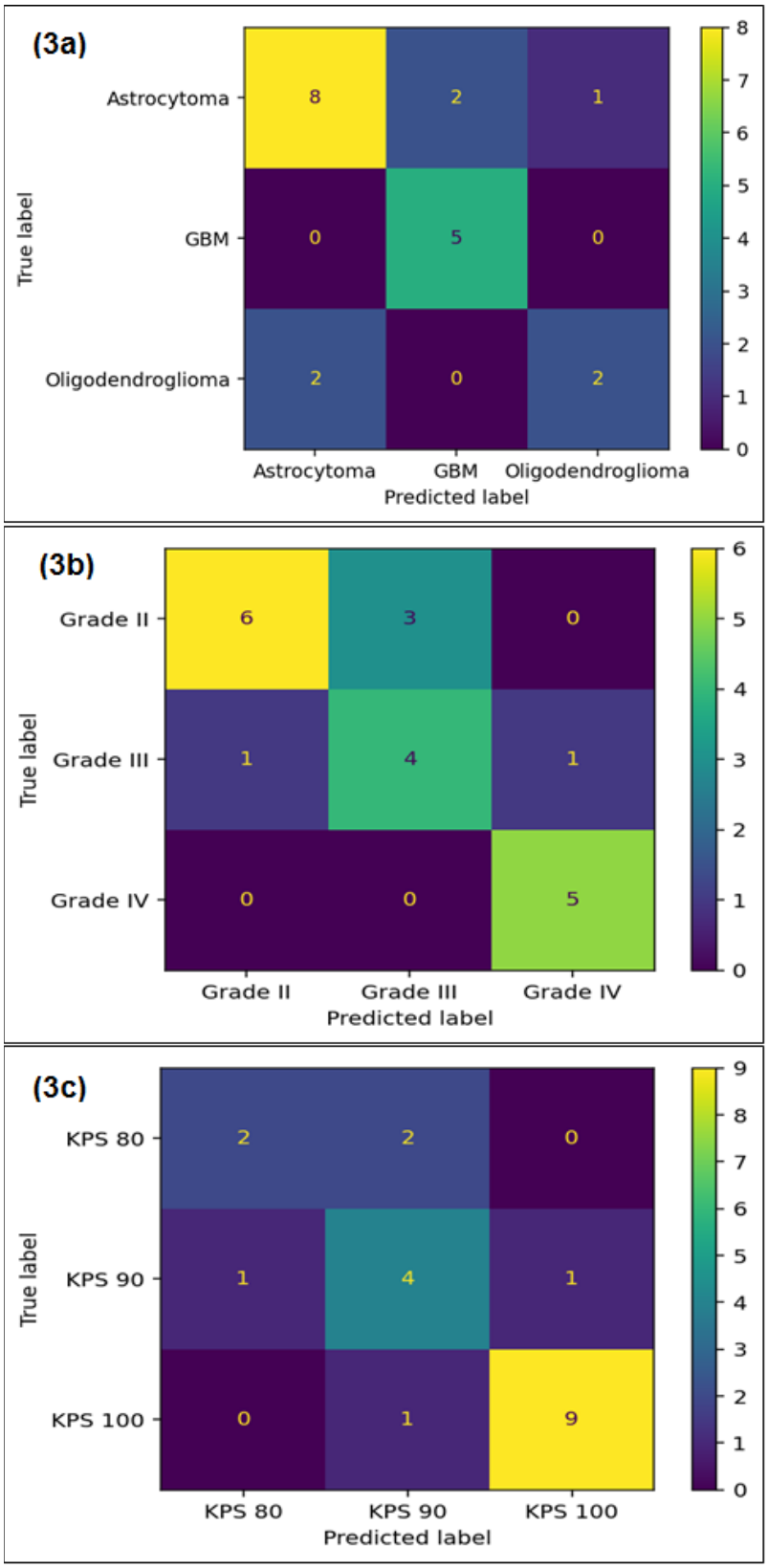


Figure 3. ResNet-18 multiclass confusion matrix for predicted class labels

Phase 2 - Overall Survival Prognosis

Kaplan-Meier curves were generated for each significant independent variable to assess its impact on patient survival. Figure 7 shows patient survival probability over time based on KPS scores and chemotherapy. Lower KPS predicts reduced survival time compared to higher KPS scores (Figure 4a). Similarly, patients who underwent chemotherapy had better survival than those who did not (Figure 4b).

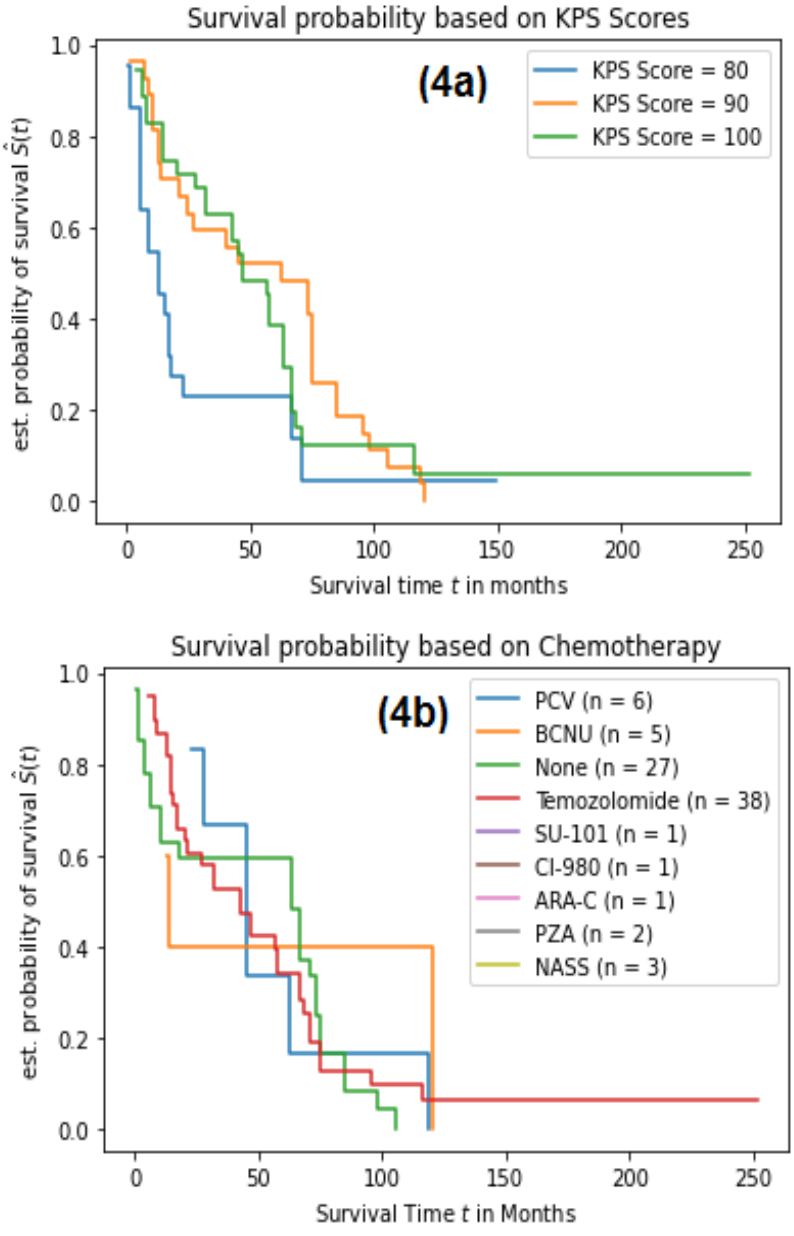


Figure 4. Kaplan-Meier plot of overall survival probability

The Deep Survnet model was identified to be the best performing survival prognosis model with the highest Harrell's Concordance Index score and the lowest time-dependent Integrated Brier Score among the three model types tested (Table 2).

Table 2. Comparative Performance Scores of Survival Prognosis Models

Model	Harrell's C Concordance Index	Uno's C Concordance Index	Integrated Brier Score
Random Survival Forest	0.91	0.88	0.189
Cox Proportional Hazards	0.73	0.71	0.175
Deep Survnet	0.92	0.93	0.0577

The time-dependent Area under the ROC Curve (AUC) shows the accuracy of survival prediction over time. Deep Survnet surpassed the Cox Proportional Hazards model and the Random Survival Forest model with a mean AUC of 0.95 (Figure 5).

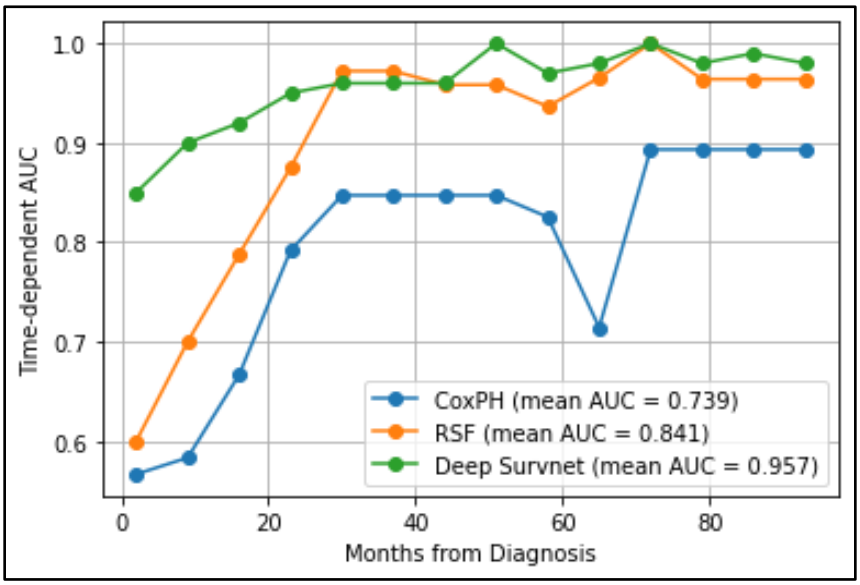


Figure 5. Time-dependent AUC comparison of survival prognosis models

Discussion

Tumor Severity and Extent

It was observed that the model predicted Glioblastoma tumors most accurately followed by Astrocytoma. This could be due to the fact that Glioblastoma is the most aggressive type of Astrocytoma, both tumors affecting glial cells called astrocytes (Appin et al, 2013). Similarly, Grade 4 was most accurately predicted by the model, potentially due to the extent of tumor growth in the MRI scans.

Class Weightage

The REMBRANDT dataset was skewed across all predictive classification classes, with 82% of the clinical data representing Caucasian patients, 55% of the Axial FLAIR MRI scans corresponding to the Astrocytoma tumor subtype, and 42% of the cases having Grade 2 tumor severity. Initial model training and performance evaluation conducted without balancing class weights yielded model performance of less than 60% accuracy for all model variants. To ensure the estimator was equally informed of all classes, the data had to be balanced to assign specific importance to classes that were under-represented in the data distribution (Table 3). Adjusting the class weights to balance the dataset resulted in model accuracy of over 90%. Class weights for each class were calculated using the equation $w_j = \frac{n}{Kn_j}$, where w_j indicates the weights of the classes, K the total number of classes, n the number of observations, and n_j the number of observations in each class.

Table 3. Distribution of Classes in REMBRANDT Data and Balanced Weights used for Model Training

Class Category	Class Labels	Distribution of Class in Dataset	Balanced Weight Used for Training
Tumor Type	Oligodendroglioma	19%	0.811320755
	Astrocytoma	55%	0.452830189
	Glioblastoma	26%	0.735849057
Tumor Grade	Grade 2	42%	0.58490566
	Grade 3	32%	0.679245283
	Grade 4	26%	0.735849057

Tumor Classification

Classification model performance is dependent on optimal tuning of the deep learning model's hyperparameters, especially when the data is imbalanced (Zhang et al. 2022). The batch size (Figure 6a), learning rate (Figure 6b) and weight decay (Figure 6c)hyperparameters were tuned in turn over a period of five trials, and the most optimal value of each hyperparameter was chosen for model development. A batch size of 32, a learning rate of 0.0001, and a weight decay of 0.0001 were identified as optimal as the validation loss was lowest and F1 Score the highest for these hyperparameter values. The cross-entropy loss validation results show that ResNet-18 (Figure 7b) and ResNet-200 (Figure 7g) produce the most optimal fit. ResNet-18 performed better than ResNet-200 on unseen test data as measured by statistical metrics and was chosen as the final model for deployment (Table 1).

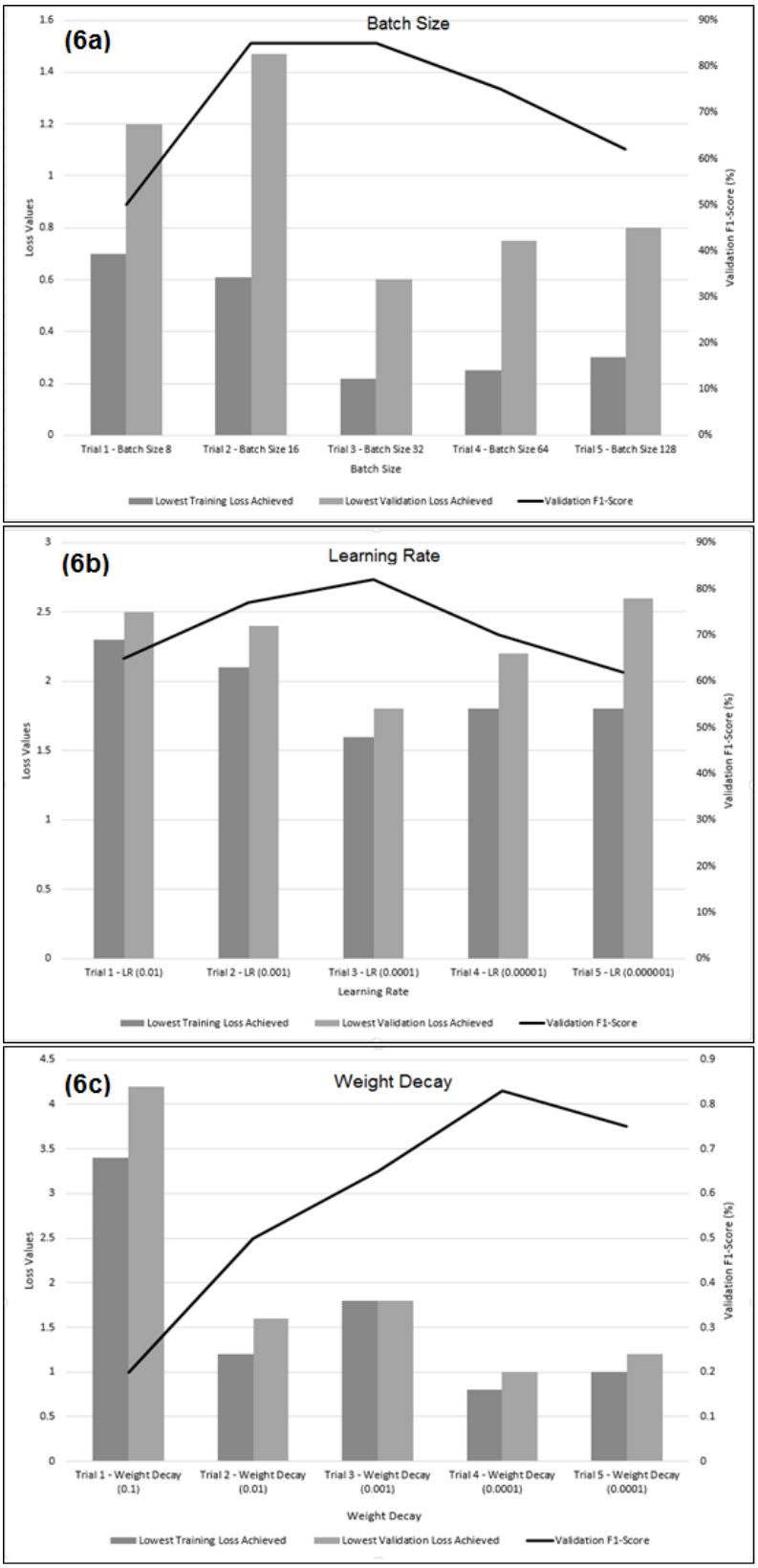
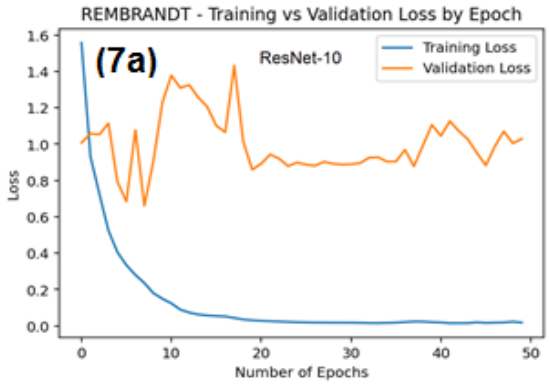
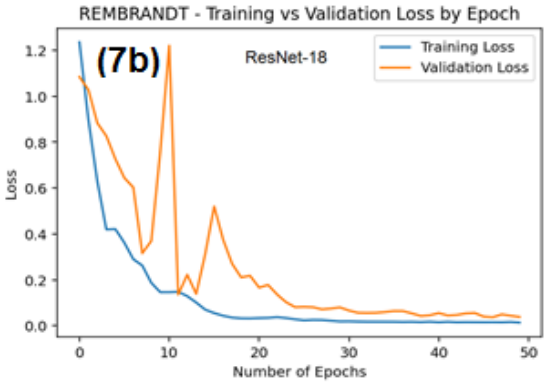
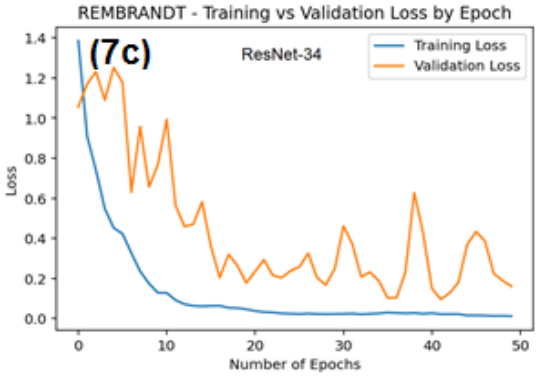
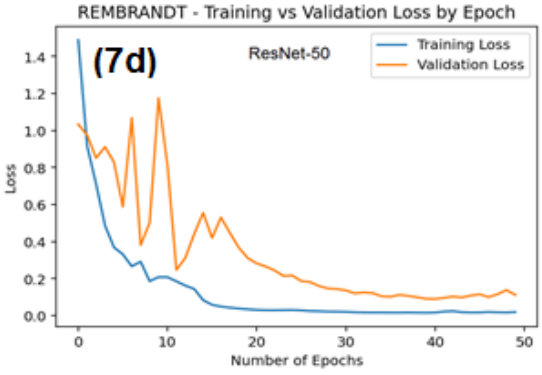


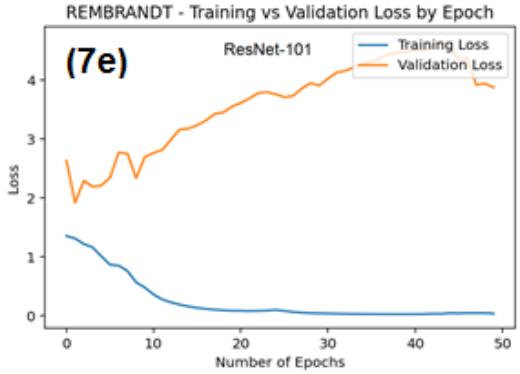
Figure 6. Hyperparameter tuning and optimization trials

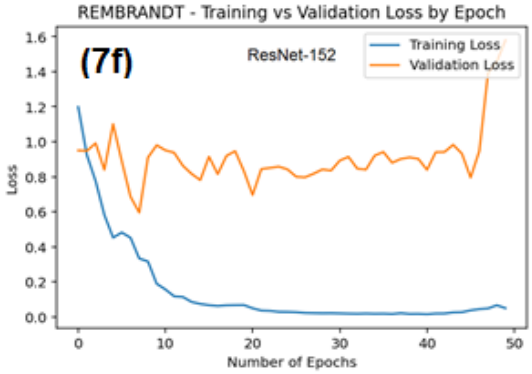

REMBRANDT - Training vs Validation Loss by Epoch
(7a)
ResNet-10
Training Loss
Validation Loss
Loss
Number of Epochs


REMBRANDT - Training vs Validation Loss by Epoch
(7b)
ResNet-18
Training Loss
Validation Loss
Loss
Number of Epochs


REMBRANDT - Training vs Validation Loss by Epoch
(7c)
ResNet-34
Training Loss
Validation Loss
Loss
Number of Epochs


REMBRANDT - Training vs Validation Loss by Epoch
(7d)
ResNet-50
Training Loss
Validation Loss
Loss
Number of Epochs


REMBRANDT - Training vs Validation Loss by Epoch
(7e)
ResNet-101
Training Loss
Validation Loss
Loss
Number of Epochs


REMBRANDT - Training vs Validation Loss by Epoch
(7f)
ResNet-152
Training Loss
Validation Loss
Loss
Number of Epochs

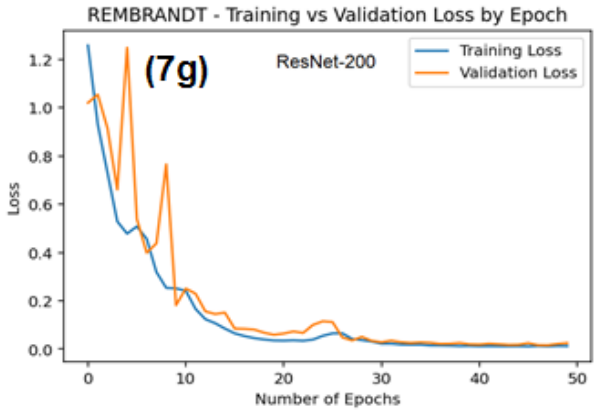

REMBRANDT - Training vs Validation Loss by Epoch
(7g)
ResNet-200
Training Loss
Validation Loss
Loss
Number of Epochs

Figure 7. Training and validation losses of ResNet models over a 50 epoch training period

AdamW, the adaptive optimizer with weight decay was used instead of the standard *Adam* optimizer, as it has been shown to result in better model generalization (Loshchilov, 2019). Early stopping, a form of regularization, was used to avoid overfitting the training dataset. This was accomplished by tracking the validation loss throughout the training process using the Pytorch logging function. The training process was stopped if the validation loss did not decrease for 10% of the total number of training epochs (Figure 8). It can be noticed that even though the total number of training epochs was 50, the validation loss reached its lowest at epoch 20. Since the goal is to minimize the validation loss for greater accuracy, the training was stopped at epoch 20, as continuing the training would not yield better accuracy. On the contrary, the model shows signs of overfitting and increasing validation loss beyond epoch 20.

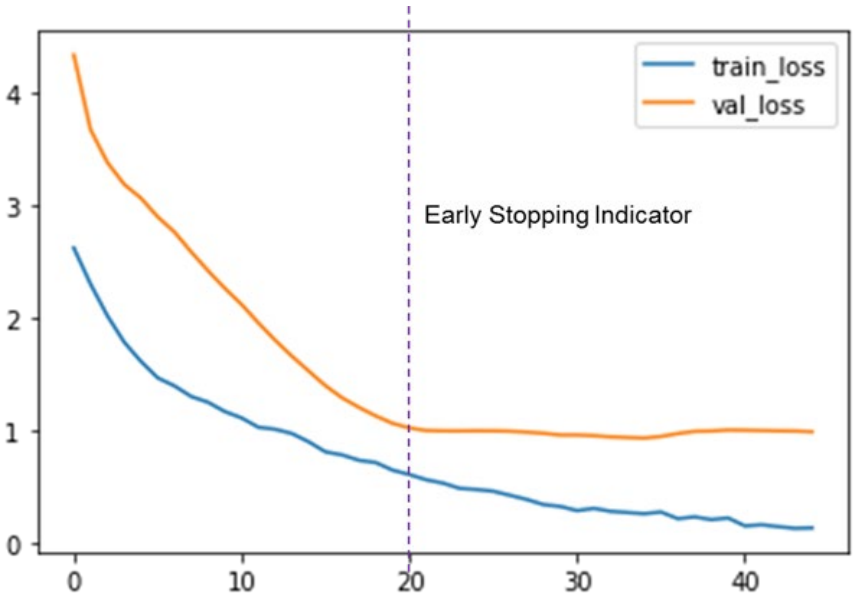


Figure 8. Optimizing validation loss using early stopping on a 50 epoch sample training iteration

Survival Prognosis

In all metrics of performance (concordance indexes, Integrated Brier Score, and time-dependent AUC), Deep Survnet outperformed the Cox Proportional Hazards and Random Survival Forest models. Deep Survnet uses the backward elimination procedure technique, in which the model gradually eliminates all the features that are irrelevant to prediction and automatically identifies the features with the most predictive power (Wang et al, 2020). Doing so helps Deep Survnet control false discovery rates during training, so it does not focus on variables with minimal predictive power.

In the REMBRANDT dataset, the occurrence of the significant event (death of the patient) spans a duration from within a month of diagnosis, to over 120 months after initial diagnosis. The survival prediction model performs best between the 10th and 80th percentile of this duration span. That is, the Brier Score demonstrated the worst model performance when the prediction was for patients who were alive for less than a year or more than eight years after diagnosis. This is expected behavior, since patients within the 10th percentile do not contain several predictive features such as the chemotherapy and radiation therapy used, or whether there was a recurrence in the tumor, all of which are significant predictors of survival. Similarly, patients alive for

over eight years offer minimal distinct features. The model performs best in the desired 1 - 5 year survival duration, which is regarded as the most appropriate prediction window for survival assessment, treatment and palliative care programs (Liu et al, 2017; Sarveazad et al. 2018).

A cohort analysis was performed to compare survival prediction for patients with similar characteristics (Figure 9). Five patients with the same tumor type (Oligodendroglioma), and similar demographic background (females between the ages 60 - 70) were chosen as part of a cohort. However, their treatment plans were different, leading to differing survival outcomes. As expected, Patient 1, who had not received chemotherapy or surgery showed a lower probability of survival compared to Patient 5, who had surgery, chemotherapy and radiation therapy treatments. It can also be seen that Patient 3 and Patient 5 who had received chemotherapy show very similar survival probabilities. This novel technique could allow physicians to conduct treatment scenario analysis on new patients to understand how different treatment options could impact overall patient survival.

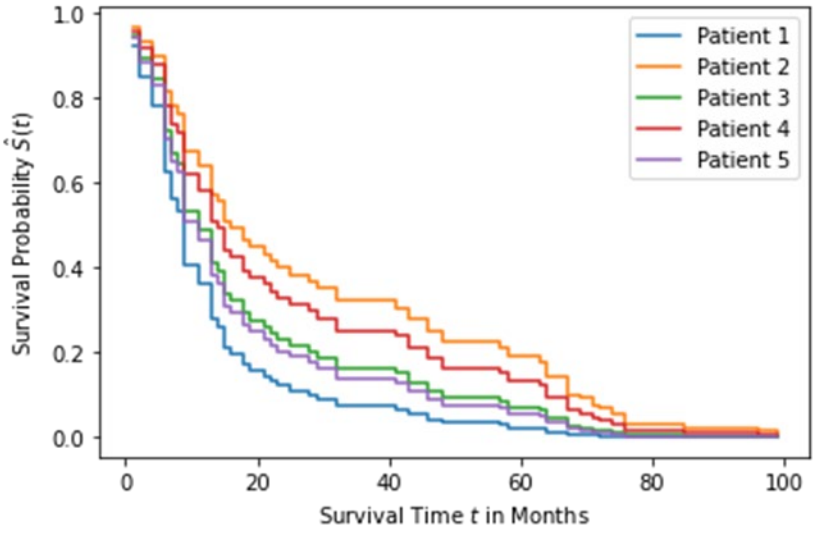


Figure 9. Comparative survival analysis of patient cohorts with similar tumor characteristics

Conclusion

This study leveraged an ensemble of deep learning models to diagnose CNS tumor severity and extent, and predict the patient's 5-year overall survival rate. The diagnostic classification model was able to achieve a high level of Precision and Recall, indicating its generalizability on unseen CNS tumor datasets. The prognostic survival prediction model achieved a high AUC score, also showing that it is able to predict survival rates for patients with different demographic and treatment backgrounds. The study and model performance was constrained by the limited amount of patient data and the sparse nature of available clinical information in the REMBRANDT dataset. However, due to the reusable nature of this deep learning framework, it can be extended to cover other cancer types and subtypes. As shown in this study, model performance reaching medically acceptable percentages can be achieved even on moderate datasets. Future analysis includes leveraging other brain and CNS cancer studies such as The Cancer Genomic Atlas (TCGA), UPenn-GBM study, and clinical data from the SEER cancer database to improve the prediction accuracy of the neural network models.

Limitations

The study was constrained to a cohort of 80 cases that had both imaging and corresponding clinical data, resulting in a limited dataset for neural network model training. Although the REMBRANDT study included patient data from two institutions, the cases from Thomas Jefferson University had no associated clinical data and was excluded from the study. The model was hence trained on a comparatively small dataset from a single institution, which may limit its generalizability to unseen data from other institutions.

Acknowledgments

I would like to thank my advisor for their support and guidance throughout this research study.

References

Alzubaidi, L., et al. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data. https://doi.org/10.1186%2Fs40537-021-00444-8

Ambale-Venkatesh et al. (2017). Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis. Circulation Research. https://doi.org/10.1161/CIRCRESAHA.117.311312

Amin, J. et al. (2020). Brain tumor classification based on DWT fusion of MRI sequences using convolutional neural network. Pattern Recognition Letters. https://doi.org/10.1016/j.patrec.2019.11.016

Appin, Christina L. et al. (2013). Glioblastoma with Oligodendroglioma Component (GBM-O): Molecular Genetic and Clinical Characteristics. Brain Pathology. https://doi.org/10.1111/bpa.12018

Bressem, Keno K. et al. (2020). Comparing different deep learning architectures for classification of chest radiographs. Scientific Reports. https://doi.org/10.1038/s41598-020-70479-z

Cancer Stat Facts: Brain and Other Nervous System Cancer. (2022). Accessed April 10, 2023. https://seer.cancer.gov/statfacts/html/brain.html

Chen, S. et al. (2019). Med3D: Transfer Learning for 3D Medical Image Analysis. arXiv. https://doi.org/10.48550/arXiv.1904.00625

Clark, K. et al. (2013). The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. Journal of Digital Imaging. https://doi.org/10.1007/s10278-013-9622-7

Cousin, F. et al. (2023). Radiomics and Delta-Radiomics Signatures to Predict Response and Survival in Patients with Non-Small-Cell Lung Cancer Treated with Immune Checkpoint Inhibitors. Cancers. https://doi.org/10.3390/cancers15071968

European Society of Radiologists. (2022). Attracting the next generation of radiologists: a statement by the European Society of Radiology (ESR). Insights Imaging. https://doi.org/10.1186/s13244-022-01221-8

Ewart, M. H. et al. (2021). A Comparison of Magnetic Resonance Imaging Methods to Assess Multiple Sclerosis Lesions: Implications for Patient Characterization and Clinical Trial Design. Diagnostics. https://doi.org/10.3390%2Fdiagnostics12010077

Frija, G. et al. (2021). How to improve access to medical imaging in low- and middle-income countries? eClinicalMedicine. https://doi.org/10.1016/j.eclinm.2021.101034

Gerds, Thomas A. et al. “Estimating a Time-Dependent Concordance Index for Survival Prediction Models with Covariate Dependent Censoring”, Statistics in Medicine, Nov 2012, https://doi.org/10.1002/sim.5681

George, B. et al. (2014). Survival analysis and regression models. Journal of Nuclear Cardiology. https://doi.org/10.1007/s12350-014-9908-2

Goldstein-Greenwood, J. A Brief on Brier Scores. University of Virginia Library Research Data Services + Sciences. Accessed Jul 31, 2023. https://data.library.virginia.edu/a-brief-on-brier-scores/

Henderson, M., Radiology Facing a Global Shortage", Accessed September 15, 2023, https://www.rsna.org/news/2022/may/Global-Radiologist-Shortage

Houwelingen, Hans C. and Putter, H. (2015). Comparison of stopped Cox regression with direct methods such as pseudo-values and binomial regression. Lifetime Data Analysis. https://doi.org/ 10.1007/s10985-014-9299-3

Cho, H. and Park, H. (2017). Classification of low-grade and high-grade glioma using multi-modal image radiomics features. IEEE Xplore. https://doi.org/10.1109/EMBC.2017.8037508

In, J and Lee, D. K. (2018). Survival analysis: Part I — analysis of time-to-event", Korean Journal of Anesthesiology. https://doi.org/10.4097%2Fkja.d.18.00067

Kalafi E Y. et al. (2023). Machine Learning and Deep Learning Approaches in Breast Cancer Survival Prediction Using Clinical Data. Folia Biologica, Accessed September 16, 2023, https://pubmed.ncbi.nlm.nih.gov/32362304/

Kathleen, A. C. 2022. Annual report to the nation on the status of cancer, part 1: National cancer statistics Interpretation. Accessed March 15, 2023. https://academic.oup.com/jncimono/article/2014/49/145/904712

Klaveren, D., et al. (2014). Assessing discriminative ability of risk models in clustered data. BMC Medical Research Methodology. https://doi.org/10.1186/1471-2288-14-5

Liu, F. et al. (2017). Epidemiology and survival outcome of breast cancer in a nationwide study. Oncotarget. https://doi.org/10.18632/oncotarget.15207

Loshchilov, I., and Hutter, F. (2019). Decoupled Weight Decay Regularization. arXiv. https://doi.org/10.48550/arXiv.1711.05101

Mariotto, Angela B. et al. (2014). Cancer Survival: An Overview of Measures, Uses, and Interpretation. JNCI Monographs. https://doi.org/10.1093/jncimonographs/lgu024

Matsuo, K. et al. (2018). Survival outcome prediction in cervical cancer: Cox models vs deep-learning model. American Journal of Obstetrics and Gynecology. https://doi.org/10.1016/j.ajog.2018.12.030

Meola, A. et al. (2018). Gold Nanoparticles for Brain Tumor Imaging: A Systematic Review. Frontiers in Neurology. https://doi.org/10.3389/fneur.2018.00328

Moreau, J. T. et al. (2020). Individual-patient prediction of meningioma malignancy and survival using the Surveillance, Epidemiology, and End Results database. npj Digital Medicine. https://doi.org/10.1038/s41746-020-0219-5

Palsson, S. et al. (2022). Predicting survival of glioblastoma from automatic whole-brain and tumor segmentation of MR images. Scientific Reports. https://doi.org/10.1038/s41598-022-19223-3

Painuli D., et al. (2022). Recent advancement in cancer diagnosis using machine learning and deep learning techniques: A comprehensive review. Computers in Biology and Medicine. https://pubmed.ncbi.nlm.nih.gov/35551012/

Pei, L. et al. (2020). Context aware deep learning for brain tumor segmentation, subtype classification, and survival prediction using radiology images. Scientific Reports. https://doi.org/10.1038/s41598-020-74419-9

Raykar, V. C., et al. (2007). On Ranking in Survival Analysis: Bounds on the Concordance Index. https://proceedings.neurips.cc/paper/2007/file/33e8075e9970de0cfea955afd4644bb2-Paper.pdf

Rezaianzadeh, A. et al. (2017). The overall 5-year survival rate of breast cancer among Iranian women: A systematic review and meta-analysis of published studies. Breast Disease. https://doi.org/10.3233/bd-160244

Roshanei, G. et al. (2022). Factors affecting the survival of patients with colorectal cancer using random survival forest. Journal of gastrointestinal cancer. https://doi.org/10.1007/s12029-020-00544-3

Sarveazad, A. et al. (2018). 5-Year Survival Rates and Prognostic Factors in Patients with Synchronous and Metachronous Breast Cancer from 2010 to 2015. Asian Pacific journal of cancer prevention. https://doi.org/10.31557/apjcp.2018.19.12.3489

Swinburne, N. C. et al. (2019). Machine learning for semi-automated classification of glioblastoma, brain metastasis and central nervous system lymphoma using magnetic resonance advanced imaging. Annals of Translational Medicine. https://doi.org/10.21037%2Fatm.2018.08.05

University of Virginia Library Research Data Services + Sciences. (2022). A Brief on Brier Scores. Accessed on July 8, 2023. https://data.library.virginia.edu/a-brief-on-brier-scores/

Vagvala, S., et al. (2022). Imaging diagnosis and treatment selection for brain tumors in the era of molecular therapeutics. Cancer Imaging. https://doi.org/10.1186/s40644-022-00455-5

Walter K. K. (2007). Concordance for Survival Time Data: Fixed and Time-Dependent Covariates and Possible Ties in Predictor and Time. Accessed March 20, 2023. https://www.mayo.edu/research/documents/biostat-80pdf/doc-10027891

Wang, J. et al. (2020). SurvNet: A Novel Deep Neural Network for Lung Cancer Survival Analysis With Missing Values. Frontiers in Oncology. https://doi.org/10.3389/fonc.2020.588990

Wu, W. et al. (2020). An Intelligent Diagnosis Method of Brain MRI Tumor Segmentation Using Deep Convolutional Neural Network and SVM Algorithm. Computational and Mathematical Methods in Medicine. https://doi.org/10.1155%2F2020%2F6789306

Wuni, A. et al. (2020). Developing a policy framework to support role extension in diagnostic radiography in Ghana. Journal of Medical Imaging and Radiation Sciences. https://doi.org/10.1016/j.jmir.2020.09.013

Xue, C. et al. (2021). Radiomics feature reliability assessed by intraclass correlation coefficient: a systematic review. Quantitative Imaging in Medicine and Surgery. https://doi.org/10.21037/qims-21-86

Yamashita, R. et al. (2018). Convolutional neural networks: an overview and application in radiology. Insights into Imaging. https://doi.org/10.1007/s13244-018-0639-9

Yan, P. et al. (2016). Accuracy of conventional MRI for preoperative diagnosis of intracranial tumors: A retrospective cohort study of 762 cases. International Journal of Surgery. https://doi.org/10.1016/j.ijsu.2016.10.023

Zhang, Z. et al. (2023). Bayesian inference for Cox proportional hazard models with partial likelihoods, nonlinear covariate effects and correlated observations. Statistical methods in medical research. https://doi.org/10.1177/09622802221134172

Zhang, Y. et al. (2023). Artificial intelligence-driven radiomics study in cancer: the role of feature engineering and modeling. Military Medical Research. https://doi.org/10.1186/s40779-023-00458-8