National Technical University of Athens
School of Electrical and Computer Engineering
Division of Communications, Electronics and Information Systems

Development of Multimodal Deep Learning Models for Survival Prediction in Patients with Gliomas

Diploma Thesis

of

ALEXIOU ANNA

Supervisor: Nikita Konstantina
Professor NTUA

Athens, July 2025

National Technical University of Athens
School of Electrical and Computer Engineering
Division of Communications, Electronics and Information Systems

Development of Multimodal Deep Learning Models for Survival Prediction in Patients with Gliomas

---

Diploma Thesis

of

ALEXIOU ANNA

Supervisor: Nikita Konstantina
Professor NTUA

Approved by the examination committee on 4th July 2025.

(Signature)	*(Signature)*	*(Signature)*
Konstantina Nikita	Athanasios Voulodimos	Giorgos Stamou
Professor NTUA	Assistant Professor NTUA	Professor NTUA

Athens, July 2025

National Technical University of Athens
School of Electrical and Computer Engineering
Division of Communications, Electronics and Information Systems

DISCLAIMER OF ACADEMIC ETHICS AND INTELLECTUAL PROPERTY RIGHTS

Being fully aware of the implications of copyright laws, I expressly state that this diploma thesis, as well as the electronic files and source codes developed or modified in the course of this thesis, are solely the product of my personal work and do not infringe any rights of intellectual property, personality and personal data of third parties, or do not contain work / contributions of third parties for which the permission of the authors / beneficiaries is required and are not a product of partial or complete plagiarism, while the sources used are mentioned in the bibliographic references only and meet the rules of scientific citing. The points where I have used ideas, text, files and / or sources of other authors are clearly included in the text with the appropriate citation and the relevant complete reference is included in the bibliographic references section. I fully, individually and personally undertake all legal and administrative consequences that may arise in the event that it is proven, in the course of time, that this thesis or part of it does not belong to me because it is a product of plagiarism.

(Signature)

.....

Alexiou Anna
Graduate of Electrical and
Computer Engineering, National
Technical University of Athens

4th July 2025

Περίληψη

Ο καρκίνος του εγκεφάλου παραμένει μία από τις πιο επιθετικές και θανατηφόρες μορφές κακοήθειας, με σημαντική ετερογένεια και ελλιπή πρόγνωση. Η αξιόπιστη πρόβλεψη της επιβίωσης είναι κρίσιμη για τον σχεδιασμό εξατομικευμένης θεραπείας και τη βελτιστοποίηση της λήψης κλινικών αποφάσεων. Στην παρούσα εργασία προτείνεται ένα πλαίσιο βασισμένο στη βαθιά μάθηση, το οποίο ενσωματώνει ιστοπαθολογικές εικόνες και RNA sequencing δεδομένα, με στόχο την πρόβλεψη επιβίωσης σε ασθενείς με γλοιώματα διαφόρων βαθμών. Το σύνολο δεδομένων αντλήθηκε από τον Καρκινικό Γονιδιωματικό Άτλαντα (TCGA) και περιλάμβανε περιπτώσεις γλοιοβλαστώματος και γλοιώματος χαμηλότερου βαθμού. Αφού εφαρμόστηκαν τεχνικές προεπεξεργασίας, χρησιμοποιήθηκε το μοντέλο DeepSurv για την πρόβλεψη του χρόνου επιβίωσης. Το συνδυασμένο μοντέλο ενσωματώνει και τα δύο είδη δεδομένων και επέδειξε ανώτερη απόδοση από τα μονοτροπικά μοντέλα, όπως φάνηκε μέσω του δείκτη συμφωνίας (C-index), επιτυγχάνοντας τιμή **0.91**, έναντι **0.80** (μόνο εικόνες) και **0.89** (μόνο γονιδιακά δεδομένα). Επιπλέον, χρησιμοποιήθηκε η μέθοδος SurvLIME για την ερμηνεία των προβλέψεων, αναδεικνύοντας σημαντικά γονίδια και μονοπάτια που σχετίζονται με την πρόγνωση. Τα αποτελέσματα δείχνουν ότι ο συνδυασμός ετερογενών βιοϊατρικών δεδομένων με επεξηγήσιμα μοντέλα βαθιάς μάθησης ενισχύει την πρόβλεψη επιβίωσης και την υποστήριξη κλινικών αποφάσεων για ασθενείς με κακοήθεις εγκεφαλικούς όγκους.

Λέξεις-κλειδιά: Καρκίνος Εγκεφάλου, Γλοίωμα, Πρόβλεψη Επιβίωσης, Βαθιά Μάθηση, Ιστοπαθολογία, RNA-sequencing, DeepSurv, SurvLIME

Abstract

Brain cancer remains one of the most aggressive and lethal forms of cancer, with limited prognosis and high inter-patient variability. Reliable survival prediction is essential for personalized treatment planning and improved clinical decision-making. In this thesis, we present a deep learning-based framework that integrates histopathological whole-slide images (WSIs) and RNA-sequencing (RNA-seq) data to predict survival in patients with gliomas of varying grades. The dataset was derived from The Cancer Genome Atlas (TCGA) and included both glioblastoma and lower-grade glioma cases. For each data modality, we developed specialized preprocessing pipelines: patch-level feature extraction and attention-based aggregation for WSIs, and TPM normalization with dimensionality reduction for RNA-seq data. Deep autoencoders were used to compress each modality into a low-dimensional latent space, and the resulting representations were input to the DeepSurv model for survival prediction. The fused model, which fuses image- and gene-based features, achieved superior performance compared to unimodal models, as measured by concordance index (C-index), achieving a score of **0.91**, compared to **0.80** for the image-only model and **0.89** for the RNA-sequencing based model. We further employed SurvLIME to interpret the model's predictions, offering insights into biologically relevant genes and pathways associated with survival. Overall, this work demonstrates the value of multimodal data integration and explainable AI techniques for risk stratification in brain tumor patients.

Keywords: Brain Cancer, Glioma, Survival Prediction, Histopathology, RNA-sequencing, Deep Learning, DeepSurv, Multimodal Learning, Autoencoders, Explainable AI, SurvLIME

to my parents

Acknowledgements

I would first like to thank Professor Nikita Konstantina for supervising this thesis and for giving me the opportunity to carry it out in the The Biomedical Simulations and Imaging (BIOSIM) Laboratory. I also extend my special thanks to Ms. Maria Athanasiou for her guidance and the excellent collaboration we had. I would like to thank my parents and my sister for their guidance and moral support throughout all these years. Finally, I would like to thank my partner, Vaggelis, and my friends, as I would not be where I am today without their support.

Athens, July 2025

Anna Alexiou

Contents

List of Tables

List of Figures

Κεφάλαιο 1

Εκτενής Ελληνική Περίληψη

1.1 Εισαγωγή και Κίνητρο

Ο καρκίνος του εγκεφάλου, και ειδικότερα τα γλοιώματα, συγκαταλέγεται στις πιο επιθετικές και θανατηφόρες μορφές κακοήθειας, με εξαιρετικά χαμηλά ποσοστά επιβίωσης. Η αξιόπιστη πρόβλεψη του χρόνου επιβίωσης των ασθενών αυτών είναι ζωτικής σημασίας για τον σχεδιασμό εξατομικευμένων θεραπειών και τη βελτιστοποίηση της κλινικής αντιμετώπισης.

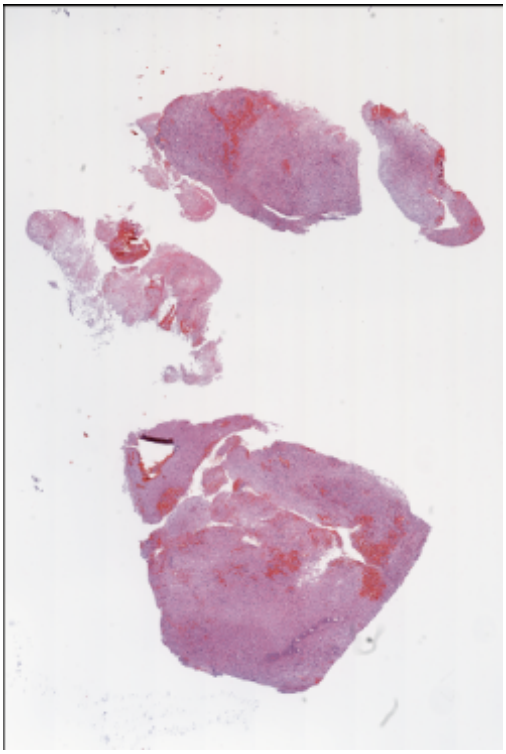
Στην παρούσα εργασία προτείνεται ένα πολυτροπικό σύστημα πρόγνωσης βασισμένο σε μεθόδους βαθιάς μάθησης, το οποίο συνδυάζει δύο ετερογενείς πηγές δεδομένων: αφενός τις ψηφιοποιημένες ιστοπαθολογικές εικόνες ολόκληρης διαφάνειας (Whole Slide Images, WSIs) εγκεφαλικών όγκων και αφετέρου τα προφίλ γονιδιακής έκφρασης (RNA-seq) των αντίστοιχων δειγμάτων.

Ο συνδυασμός των μορφολογικών χαρακτηριστικών του όγκου (από τις ιστολογικές εικόνες) με τα μοριακά δεδομένα (RNA) προσφέρει μια πιο ολοκληρωμένη περιγραφή της νόσου, με στόχο τη βελτίωση της ακρίβειας και της αξιοπιστίας της πρόγνωσης.

1.2 Περιγραφή της Μεθοδολογίας

1.2.1 Δεδομένα και Στόχος

Το σύνολο δεδομένων προήλθε από τον οργανισμό TCGA και περιλάμβανε συνολικά 583 περιπτώσεις γλοιοβλαστώματος (GBM) και γλοιωμάτων χαμηλότερου βαθμού (LGG), για τις οποίες ήταν διαθέσιμα τόσο ιστοπαθολογικά δείγματα όσο και δεδομένα γονιδιακής έκφρασης (RNA-seq). Από τα αντίστοιχα κλινικά δεδομένα ελήφθησαν πληροφορίες σχετικά με την επιβίωση των ασθενών και τον χρόνο επιβίωσης ή τον χρόνο έως την αποχώρηση από τη μελέτη.

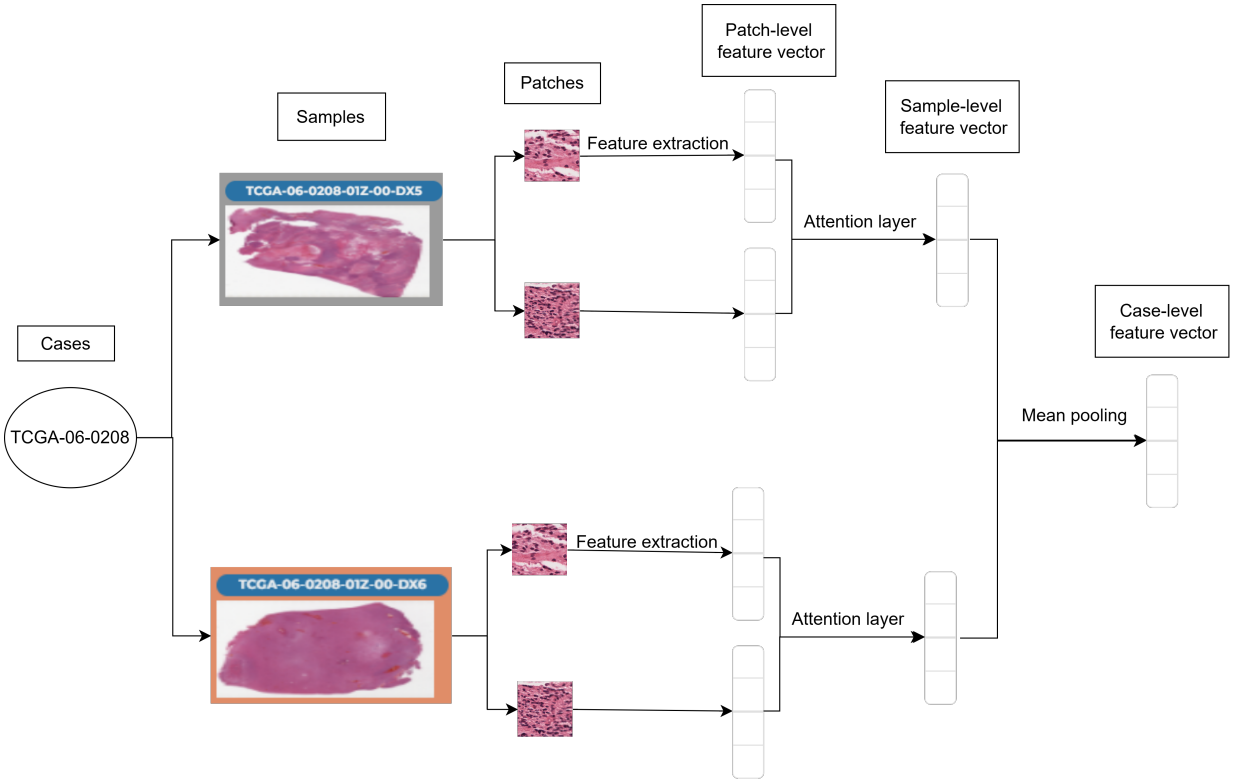


Σχήμα 1.1: Παράδειγμα ιστοπαθολογικής εικόνας από το τελικό σύνολο δεδομένων.

1.2.2 Επεξεργασία Ιστοπαθολογικών Δεδομένων

Αρχικά, εντοπίστηκαν οι περιοχές ιστού στις εικόνες και ακολούθησε εξαγωγή patches υψηλής ανάλυσης. Κάθε patch μετατράπηκε σε διάνυσμα χαρακτηριστικών μέσω ενός προεκπαιδευμένου CNN. Εξετάστηκαν τέσσερα διαφορετικά προεκπαιδευμένα μοντέλα: *KimiaNet, EfficientNet-B0, EfficientNet-B1, ResNet-50*.

Στη συνέχεια, εφαρμόστηκε μηχανισμός attention για την ανάδειξη των σημαντικότερων patches και τη μετάβαση από χαρακτηριστικά επιπέδου patch-level σε sample-level. Δεδομένου ότι κάθε ασθενής μπορεί να διέθετε πολλαπλά ιστολογικά δείγματα, εφαρμόστηκε mean pooling στα sample-level χαρακτηριστικά για την εξαγωγή ενός τελικού διανύσματος ανά ασθενή (case-level feature vector). Η διαδικασία απεικονίζεται συνοπτικά στο Σχήμα 1.2.



Σχήμα 1.2: Σχηματική απεικόνιση της προεπεξεργασίας των ιστοπαθολογικών εικόνων.

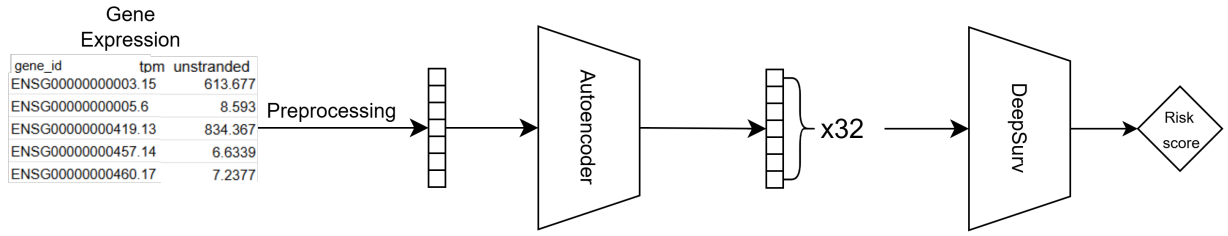
Πριν την εισαγωγή των χαρακτηριστικών στο μοντέλο επιβίωσης, εφαρμόστηκε συμπίεση των διανυσμάτων μέσω ενός autoencoder. Τα αρχικά διανύσματα περιείχαν περίπου 1000 features, ανάλογα με το εκάστοτε προεκπαιδευμένο μοντέλο, και εξετάστηκαν τέσσερα μεγέθη εξόδου: 32, 64, 128 και 256. Η καλύτερη απόδοση επιτεύχθηκε με μέγεθος 32.

Το συμπιεσμένο διάνυσμα για κάθε ασθενή εισήχθη στο μοντέλο επιβίωσης DeepSurv, μαζί με τα αντίστοιχα δεδομένα επιβίωσης (event, time until event). Η συνολική αρχιτεκτονική του πιπελινε των εικόνων παρουσιάζεται στο Σχήμα 1.2.

1.2.3 Επεξεργασία Δεδομένων RNA

Τα δεδομένα RNA-seq που χρησιμοποιήθηκαν ήταν ήδη κανονικοποιημένα με τη μέθοδο Transcripts per Million (TPM). Αρχικά, εφαρμόστηκε λογαριθμικός μετασχηματισμός, ενώ στη συνέχεια έγινε φιλτράρισμα με βάση τη χαμηλή διακύμανση, μειώνοντας τον αριθμό των γονιδίων από περίπου 60.000 σε 20.000.

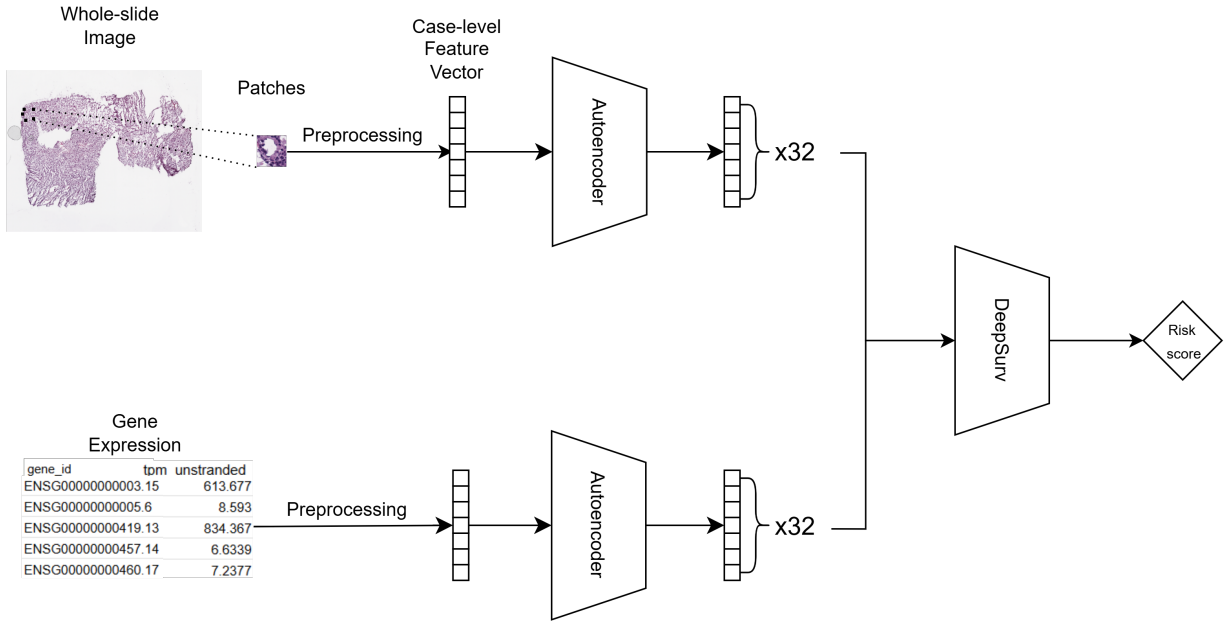
Ακολούθησε z-μετασχηματισμός και χρήση autoencoder για μείωση της διαστασιμότητας. Όπως και στην περίπτωση των εικόνων, το τελικό συμπιεσμένο διάνυσμα 32 διαστάσεων χρησιμοποιήθηκε ως είσοδος σε ανεξάρτητο μοντέλο DeepSurv. Η αρχιτεκτονική του μοντέλου που βασίζεται αποκλειστικά σε δεδομένα γονιδιακής έκφρασης παρουσιάζεται στο Σχήμα 1.3.



Σχήμα 1.3: Μοντέλο πρόβλεψης επιβίωσης βασισμένο αποκλειστικά σε δεδομένα RNA.

1.2.4 Συνδυαστικό Μοντέλο

Στο επόμενο στάδιο, αναπτύχθηκε ένα συνδυαστικό μοντέλο που ενσωματώνει χαρακτηριστικά τόσο από τις ιστολογικές εικόνες όσο και από τα γονιδιακής έκφρασης δεδομένα. Τα δύο συμπιεσμένα διανύσματα (32 διαστάσεων το καθένα) συνενώθηκαν (concatenation) δημιουργώντας ένα ενοποιημένο διάνυσμα 64 διαστάσεων, το οποίο χρησιμοποιήθηκε ως είσοδος στο μοντέλο DeepSurv. Το συνολικό pipeline του συνδυαστικού μοντέλου απεικονίζεται στο Σχήμα 1.4.



Σχήμα 1.4: Συνδυασμένο πολυτροπικό μοντέλο με ενοποιημένα χαρακτηριστικά εικόνας και RNA.

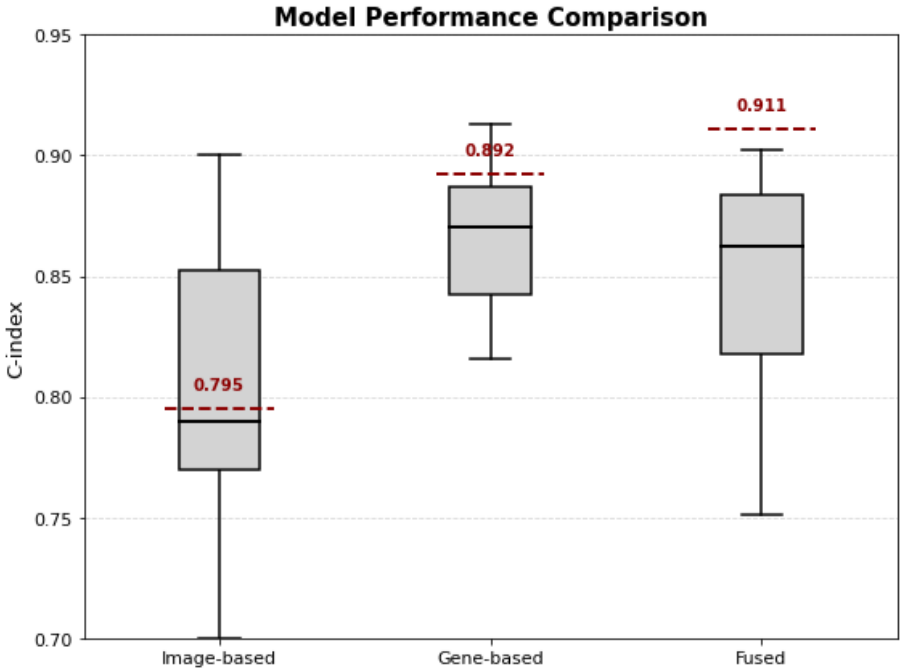
1.3 Αποτελέσματα και Συμπεράσματα

Στον Πίνακα 5.1 παρουσιάζεται η απόδοση του μοντέλου ιστοπαθολογικών εικόνων με διαφορετικούς προεκπαιδευμένους εξαγωγείς χαρακτηριστικών. Το μοντέλο KimiaNet υπερίσχυσε, επιτυγχάνοντας τις υψηλότερες τιμές C-index, και επιλέχθηκε για το συνδυαστικό μοντέλο.

Πίνακας 1.1: Cross-validation and final test set C-index for each pretrained feature extractor.

Pretrained Model	**Mean CV C-index**	**Final Test C-index**
EfficientNet-B0	0.79	0.78
EfficientNet-B1	0.76	0.74
ResNet-50	0.79	0.73
KimiaNet	**0.80**	**0.80**

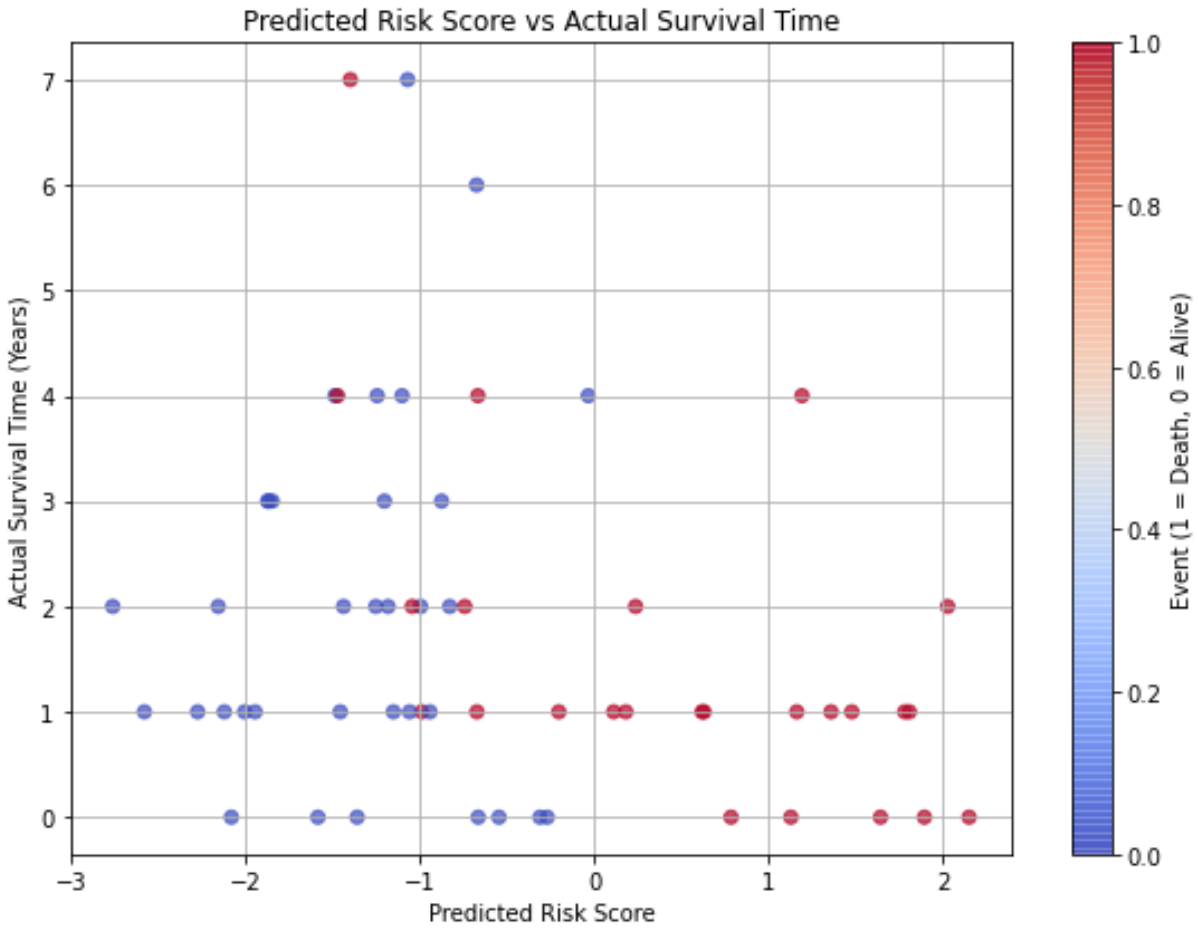
Στο Σχήμα 1.5 απεικονίζεται η συγκριτική απόδοση των τριών μοντέλων. Το συνδυαστικό μοντέλο σημείωσε την υψηλότερη επίδοση με C-index ίσο με 0.91.



Σχήμα 1.5: Σύγκριση δείκτη συμφωνίας (C-index) για κάθε μοντέλο.

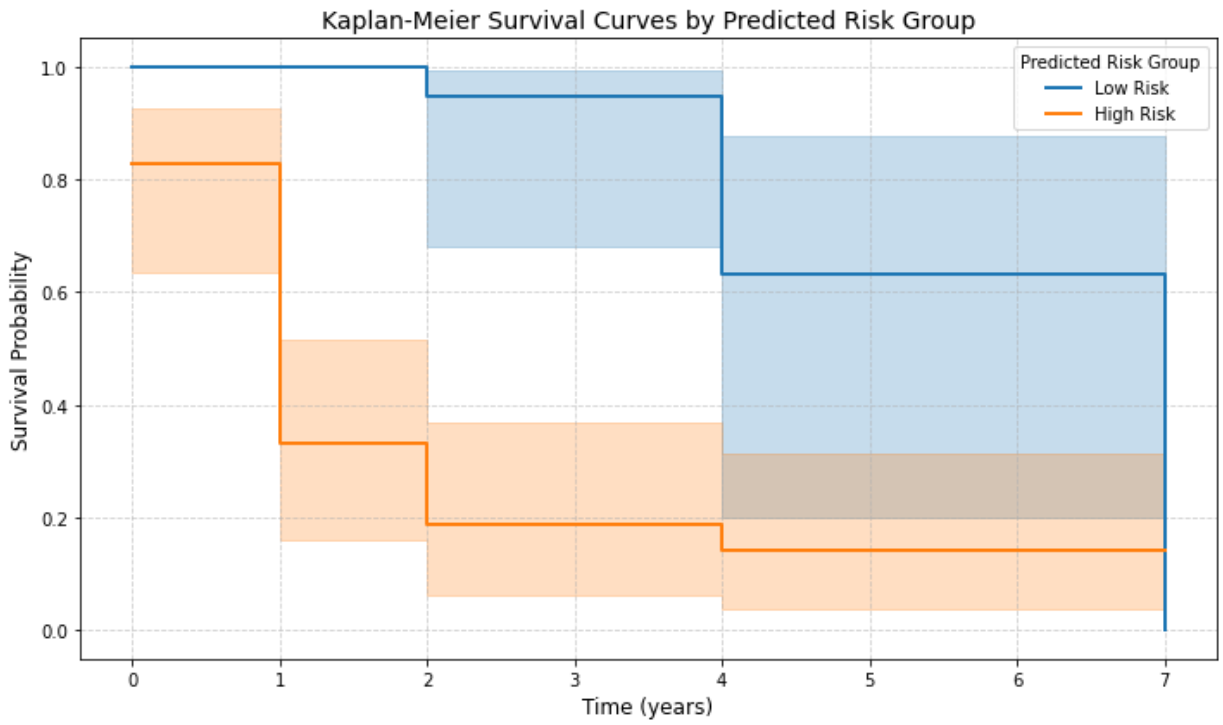
Για την περαιτέρω αξιολόγηση του συνδυαστικού μοντέλου χρησιμοποιήθηκαν καμπύλες επιβίωσης Kaplan–Meier, καθώς και η γραφική απεικόνιση της σχέσης μεταξύ προβλεπόμενου risk score και πραγματικού χρόνου επιβίωσης.

Στο Σχήμα 5.2, παρατηρείται σαφής αρνητική συσχέτιση μεταξύ κινδύνου και επιβίωσης, όπως αναμενόταν. Ωστόσο, εντοπίζονται περιοχές επικάλυψης, ιδιαίτερα σε ενδιάμεσες τιμές κινδύνου, υποδηλώνοντας περιορισμούς στην προγνωστική βεβαιότητα του μοντέλου.



Σχήμα 1.6: Σχέση μεταξύ προβλεπόμενων τιμών κινδύνου και πραγματικών χρόνων επιβίωσης.

Οι ασθενείς του συνόλου ελέγχου διαχωρίστηκαν σε ομάδες υψηλού και χαμηλού κινδύνου βάσει της διάμεσης τιμής του risk score. Οι καμπύλες επιβίωσης των δύο ομάδων παρουσιάζονται στο Σχήμα 1.7. Η διαφορά μεταξύ των καμπυλών είναι σημαντική, ιδίως στα πρώτα χρόνια μετά τη διάγνωση, επιβεβαιώνοντας την κλινική χρησιμότητα του μοντέλου. Ωστόσο, η σύγκλιση των καμπυλών μετά το τέταρτο έτος υποδεικνύει περιορισμούς στην πρόβλεψη μακροπρόθεσμης επιβίωσης.



Σχήμα 1.7: Καμπύλες επιβίωσης Kaplan–Meier ανά ομάδα κινδύνου.

1.4 Ερμηνευσιμότητα

Η κατανόηση των παραγόντων που συμβάλλουν στις προβλέψεις των μοντέλων αποκτά ολοένα και μεγαλύτερη σημασία για την υποστήριξη της εξατομικευμένης ιατρικής. Στο πλαίσιο αυτό, εφαρμόστηκε η μέθοδος SurvLIME, η οποία επέτρεψε την αποτίμηση της σημασίας των γονιδίων στο εκτιμώμενο ρίσκο επιβίωσης.

Η τεχνική βασίστηκε στη δημιουργία τοπικών γραμμικών μοντέλων γύρω από κάθε πρόβλεψη, εντοπίζοντας τα χαρακτηριστικά που επηρέασαν περισσότερο το αποτέλεσμα και αποδίδοντας αντίστοιχες τιμές σημαντικότητας (importance scores). Η μέθοδος εφαρμόστηκε αρχικά στο μοντέλο RNA-only και στη συνέχεια στο πολυτροπικό μοντέλο, επιτρέποντας την εξαγωγή των σημαντικότερων γονιδίων για κάθε περίπτωση.

Παρατηρήθηκε ότι τα δύο μοντέλα ανέδειξαν διαφορετικά γονίδια ως σημαντικότερα, γεγονός που καταδεικνύει τη συνεισφορά των μορφολογικών χαρακτηριστικών των εικόνων στη

διαδικασία πρόγνωσης. Εξετάστηκαν σε βάθος τρεις περιπτώσεις ασθενών:

- Ένας ασθενής με χαμηλό εκτιμώμενο κίνδυνο, ο οποίος κατέληξε κατά το δεύτερο έτος της περιόδου παρακολούθησης (Ασθενής 1).

- Ένας ασθενής με χαμηλό εκτιμώμενο κίνδυνο, ο οποίος δεν παρουσίασε κάποιο συμβάν και λογοκρίθηκε στο πέμπτο έτος της περιόδου παρακολούθησης (Ασθενής 2).

- Ένας ασθενής με υψηλό εκτιμώμενο κίνδυνο, ο οποίος κατέληξε κατά το πρώτο έτος της περιόδου παρακολούθησης (Ασθενής 3). Για την εξαγωγή βιολογικά ουσιαστικής ερμηνείας, οι τιμές σημαντικότητας των γονιδίων χαρτογραφήθηκαν σε γνωστά βιολογικά μονοπάτια (pathways) μέσω της βιβλιοθήκης gseapy. Με αυτόν τον τρόπο, υπολογίστηκαν συνολικά importance scores για κάθε pathway.

Μεταξύ των pathways που έλαβαν υψηλή αρνητική τιμή, αναδείχθηκαν εκείνα που σχετίζονται με τον κυτταρικό πολλαπλασιασμό και τη διηθητικότητα των όγκων, όπως:

- MET promotes cell motility (R-HSA-8875878)
- Basigin interactions (R-HSA-210991)

Τα ευρήματα αυτά ενισχύουν τη σύνδεση μεταξύ μοριακών μηχανισμών και πρόγνωσης επιβίωσης σε γλοιογενείς όγκους, και υποδεικνύουν τη χρησιμότητα επεξηγήσιμων τεχνικών τεχνητής νοημοσύνης στην ανάδειξη πιθανών θεραπευτικών στόχων. Τα αντίστοιχα γραφήματα που απεικονίζουν τα σημαντικότερα γονίδια και pathways παρατίθενται στο τέλος του παρόντος κειμένου.

1.5 Συμπεράσματα και Μελλοντικές Επεκτάσεις

Η παρούσα διπλωματική εργασία επικεντρώθηκε στην ανάπτυξη ενός μοντέλου πρόβλεψης επιβίωσης για ασθενείς με γλοιώματα, μέσω της συνδυαστικής αξιοποίησης ιστοπαθολογικών εικόνων και δεδομένων γονιδιακής έκφρασης. Η ολιστική προσέγγιση της νόσου μέσω διαφορετικών τύπων δεδομένων κρίθηκε αναγκαία, καθώς τα γλοιώματα παρουσιάζουν έντονη ετερογένεια τόσο σε μορφολογικό όσο και σε μοριακό επίπεδο.

Τα αποτελέσματα ανέδειξαν την υπεροχή του συνδυαστικού μοντέλου έναντι των επιμέρους (RNA-only ή μόνο εικόνας), επιτυγχάνοντας υψηλή ακρίβεια στην πρόγνωση της επιβίωσης. Επιπλέον, μέσω ερμηνευτικών τεχνικών, αναδείχθηκαν συγκεκριμένα γονίδια και μοριακά μονοπάτια με ισχυρή συσχέτιση προς τον εκτιμώμενο κίνδυνο, ενισχύοντας τη βιολογική αξιοπιστία των ευρημάτων.

Η εργασία αυτή υπογραμμίζει τη σημασία των πολυτροπικών μοντέλων στην ογκολογική πρόγνωση και θέτει τις βάσεις για τη μελλοντική ανάπτυξη εξατομικευμένων συστημάτων υποστήριξης απόφασης στην κλινική πράξη.

Προτεινόμενες μελλοντικές επεκτάσεις περιλαμβάνουν:

1. Ενσωμάτωση κλινικών μεταβλητών, όπως η ηλικία του ασθενούς, ο δείκτης λειτουργικότητας ή μοριακοί δείκτες, με στόχο τη βελτίωση της ακρίβειας και της κλινικής χρησιμότητας του μοντέλου.

2. Διερεύνηση προχωρημένων στρατηγικών συγχώνευσης των δεδομένων για την αποτελεσματικότερη μοντελοποίηση των αλληλεπιδράσεων μεταξύ των διαφορετικών μορφών πληροφορίας.

3. Εφαρμογή του μοντέλου σε εξωτερικά σύνολα δεδομένων και βιολογική επιβεβαίωση των σημαντικών μοριακών μονοπατιών που αναδείχθηκαν, ώστε να ενισχυθεί η μεταφραστική του αξία.

Η μεθοδολογία που προτείνεται μπορεί να λειτουργήσει ως αφετηρία για τη βελτίωση της πρόγνωσης και την προώθηση της εξατομικευμένης ιατρικής σε ασθενείς με όγκους του εγκεφάλου.

Chapter 2

The Brain and Gliomas: Clinical and Biological Foundations

2.1 Introduction to Brain Anatomy and Physiology

The human brain is a highly complex organ responsible for regulating perception, movement, cognition, and homeostasis. It contains approximately 86 billion neurons and a comparable number of glial cells [1]. Structurally, the brain is divided into the cerebrum, cerebellum, and brainstem, each mediating sensorimotor integration, coordination, and autonomic control [2].

Neurons are the primary signaling units, while glial cells, including astrocytes, oligodendrocytes, and microglia, support brain function through metabolic assistance, synaptic regulation, and immune defense. Astrocytes also maintain blood–brain barrier (BBB) integrity by regulating endothelial permeability and ion balance. [3]

The brain is protected by the skull, meninges, and cerebrospinal fluid. The BBB acts as a selective interface that prevents toxins from entering the brain while allowing essential nutrients to pass [4]. However, this barrier also impedes drug delivery in central nervous system diseases, including brain tumors [5].

2.2 Classification of Brain Tumors

Brain tumors are generally categorized as either primary, arising within the central nervous system, or secondary, originating from metastatic spread of cancers from other organs. Primary brain tumors are further classified by the World Health Organization (WHO) based on histopathological and molecular features [6]. Tumors are assigned a grade from I to IV, reflecting increasing levels of malignancy.

Gliomas, which arise from glial cells, are the most prevalent type of primary malignant brain tumors. These include astrocytomas, oligodendrogliomas, and ependymomas. Glioblastoma (GBM), also known as grade IV astrocytoma, is the most aggressive subtype, characterized by rapid proliferation, necrosis, and extensive invasion into surrounding tissue [7].

2.3 Diffuse Gliomas: Classification, Epidemiology, and Clinical Challenges

Lower-grade gliomas (WHO grade II and III), including diffuse astrocytomas and oligodendrogliomas, exhibit slower growth and occur more frequently in younger adults. Although they are less aggressive than GBM, LGGs often progress over time and share some of the same genetic and molecular alterations, including IDH mutations and 1p/19q codeletion. They remain clinically significant due to their variable prognosis and the potential for malignant transformation [8].

Glioblastoma accounts for approximately 14–15% of all primary brain tumors and more than 50% of malignant gliomas in adults [9]. The annual incidence of GBM is estimated at 3.2 per 100,000 population, with peak incidence in individuals aged 45–70 years. The tumor demonstrates a slight male predominance and occurs more frequently among individuals of Caucasian ethnicity [10].

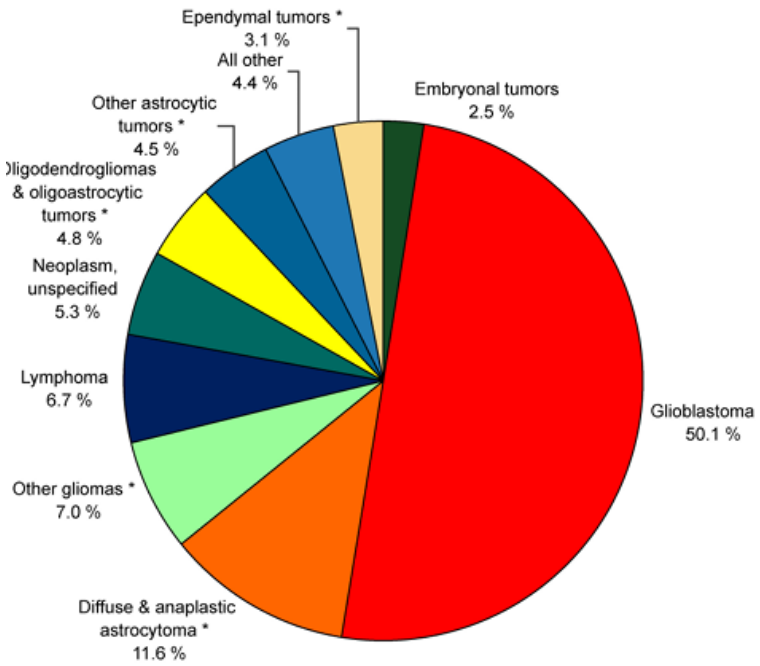


Figure 2.1: Distribution of malignant primary brain and other central nervous system tumors in adults. [9].

Despite aggressive multimodal therapy, including maximal safe surgical resection, radiotherapy, and temozolomide chemotherapy, the prognosis for GBM remains poor. Median survival is approximately 14 to 16 months, with fewer than 7% of patients surviving beyond five years post-diagnosis [11]. Recurrence is nearly universal and typically occurs within months following initial treatment. Recurrent tumors are often more resistant to conventional therapies [12].

In addition to its clinical severity, GBM imposes a substantial socioeconomic burden, owing to high healthcare utilization, progressive neurological decline, and the requirement for long-term supportive care and rehabilitation [13].

2.4 Diagnosis and Imaging Techniques

Clinical Presentation and Imaging

The diagnostic process for diffuse gliomas, encompassing both lower-grade gliomas (LGG) and glioblastoma (GBM), typically begins with a thorough clinical evaluation prompted by neurological symptoms. Common presenting signs include persistent headaches, seizures, cognitive dysfunction, and focal neurological deficits such as hemiparesis or aphasia. The nature and severity of these symptoms are closely related to the tumor's size, anatomical location, and associated mass effect or edema. While LGGs often develop slowly and may be discovered incidentally, GBMs tend to present more abruptly due to their rapid growth and necrotic core [14].

Magnetic Resonance Imaging (MRI) remains the cornerstone of both initial diagnosis and follow-up of gliomas. LGGs are typically seen as non-enhancing, T2-hyperintense lesions with poorly defined borders, whereas GBMs often demonstrate ring-enhancing lesions with central necrosis on T1-weighted post-contrast sequences. Additional sequences such as T2-weighted and FLAIR are critical in delineating tumor infiltration and peritumoral edema. Advanced modalities including magnetic resonance spectroscopy (MRS), diffusion tensor imaging (DTI), and positron emission tomography (PET) provide complementary insights into tumor metabolism, microstructural integrity, and therapeutic response [15, 16].

Histopathology and Digital Slide Analysis

Histopathological examination of biopsy or surgically resected tissue remains the gold standard for confirming glioma diagnosis. All diffuse gliomas are assessed microscopically for nuclear atypia, mitotic activity, necrosis, and microvascular proliferation. These features help distinguish lower-grade lesions from high-grade tumors. LGGs typically exhibit lower mitotic rates and a lack of necrosis, while GBMs show extensive mitoses, palisading necrosis, and abnormal vascular proliferation [6].

Hematoxylin and eosin (H&E)-stained sections are routinely used in this process. In modern workflows, these glass slides are digitized into whole-slide images (WSIs), enabling the integration of computational pathology. Artificial intelligence (AI) tools now assist in tumor segmentation, classification, and prediction of molecular subtypes, making histopathology a foundation for both diagnosis and machine learning-based research [17].

Molecular Profiling

Molecular diagnostics have become a central component of glioma classification, prognosis, and treatment planning. Key molecular markers include mutations in IDH1 or IDH2, MGMT promoter methylation, and 1p/19q co-deletion. The presence of an IDH mutation is associated with better prognosis and defines a major molecular subtype of gliomas. MGMT promoter methylation predicts improved response to temozolomide chemotherapy, especially in GBM. The 1p/19q co-deletion is a defining hallmark of oligodendrogliomas and correlates with favorable outcomes and chemosensitivity [8, 6].

2.5 Standard Treatment Strategies

Treatment approaches for diffuse gliomas differ significantly by grade and molecular profile. In LGGs, management may range from active surveillance in asymptomatic, low-risk patients to surgery followed by radiotherapy and chemotherapy in high-risk or progressive cases. In contrast, GBM treatment typically involves a more aggressive multimodal approach aimed at maximizing tumor control and prolonging survival despite the poor prognosis [18].

Surgical Resection

Surgical resection serves both diagnostic and therapeutic purposes in glioma care. In LGG, early and extensive resection has been associated with delayed malignant transformation and prolonged survival. For GBM, maximal safe resection is standard practice and has

been linked to improved outcomes. However, due to the infiltrative nature of gliomas and their proximity to eloquent brain regions, complete resection is often not feasible [19].

Radiation Therapy

Postoperative radiotherapy is a mainstay of treatment in both LGG and GBM. In LGGs, radiotherapy may be deferred in younger patients with favorable prognostic features, while it is routinely administered in high-risk or progressive cases. In GBM, radiotherapy typically consists of fractionated external beam radiation therapy totaling 60 Gy over six weeks. It targets both the tumor bed and surrounding tissue at risk for microscopic disease spread [20].

Chemotherapy

Temozolomide (TMZ) is the most commonly used chemotherapeutic agent in glioma treatment. In LGGs, chemotherapy may follow radiotherapy in high-risk patients or be used alone in select cases. In GBM, TMZ is administered concurrently with radiotherapy and continued as adjuvant therapy. The benefit of TMZ is significantly influenced by MGMT promoter methylation status, with methylated tumors demonstrating better responses [21].

Tumor Treating Fields (TTFields)

TTFields represent a novel adjunctive treatment modality currently approved for GBM. These wearable devices deliver alternating electric fields that disrupt mitotic spindle formation, thereby inhibiting tumor cell division. Clinical trials have shown modest improvements in progression-free and overall survival when TTFields are combined with maintenance temozolomide. This approach is not used for LGG at present due to a lack of evidence and lower aggressiveness of the disease [12].

Limitations and Treatment Resistance

Diffuse gliomas are notorious for their resistance to treatment. In LGG, disease progression and malignant transformation pose long-term management challenges. In GBM, recurrence is almost universal and occurs within months of treatment. Contributing factors include the blood-brain barrier, which limits drug delivery; tumor heterogeneity, which promotes adaptation and resistance; and glioma stem-like cells, which exhibit heightened resilience to therapy [14].

2.6 Challenges and Prognostic Factors

Both lower-grade and high-grade gliomas present distinct but overlapping challenges in their clinical management. While GBMs are acutely aggressive and rapidly fatal, LGGs pose long-term risks due to their unpredictable progression and potential for transformation.

Tumor Heterogeneity and Invasiveness

Intratumoral heterogeneity is a hallmark of diffuse gliomas. Genetic, epigenetic, and phenotypic variability exists not only between tumors but within individual tumors. In GBM, this leads to rapid development of treatment-resistant clones. LGGs also exhibit heterogeneity, though generally at a slower pace. Moreover, gliomas infiltrate adjacent normal brain tissue, complicating complete surgical resection and contributing to recurrence [22].

Therapy Resistance

Resistance to therapy is a multifactorial phenomenon in gliomas. It is driven by enhanced DNA repair mechanisms, activation of redundant signaling pathways, evasion of immune detection, and the presence of a treatment-resistant subpopulation of glioma stem-like cells. The blood-brain barrier further complicates systemic therapy by impeding drug penetration. These mechanisms are present in both LGGs and GBMs but are particularly prominent in

recurrent high-grade tumors [23].

Prognostic Biomarkers

A variety of molecular and clinical markers have prognostic relevance across the glioma spectrum. IDH mutations are associated with improved prognosis and are more frequent in LGGs. MGMT promoter methylation is predictive of temozolomide responsiveness and improved survival in GBM. The 1p/19q co-deletion is a favorable marker found in oligodendrogliomas. Additional factors such as patient age, functional status, tumor location, and extent of resection are also used for risk stratification and therapeutic planning [21, 24].

Recurrence and Lack of Curative Therapies

Recurrence remains a significant obstacle in glioma management. In GBM, recurrence is almost inevitable and typically occurs within a few months of completing initial therapy. Recurrent tumors are often more aggressive and resistant to further treatment. In LGG, recurrence and progression to higher-grade gliomas may occur over a period of years, underscoring the importance of long-term monitoring. At present, there are no curative options for diffuse gliomas, and research continues to explore immunotherapy, molecularly targeted therapies, and personalized treatment approaches [14, 12].

Chapter 3

Theoretical Background Information

3.1 Introduction

This chapter provides the necessary theoretical foundation to support the methods and models developed in this thesis. It introduces core concepts in machine learning and deep learning, with a focus on the architectures and techniques most relevant to biomedical data analysis and survival prediction tasks.

The chapter begins with a general overview of machine learning paradigms, including supervised and unsupervised learning, followed by an introduction to deep learning and its most commonly used architectures. Special attention is given to convolutional neural networks (CNNs) and transfer learning techniques, which play a key role in image-based feature extraction. Pre-trained models such as ResNet-50, EfficientNet, and KimiaNet are discussed due to their relevance in histopathological image analysis.

Next, we turn to the processing of biological data, specifically RNA-sequencing, and review how machine learning methods are applied to high-dimensional omics datasets. We then outline classical and modern approaches to survival analysis, including the DeepSurv model, which is used in this work.

The chapter also presents the Optuna framework for automated hyperparameter opti-

mization, and discusses various feature fusion strategies for integrating image and genomic data. Lastly, we introduce explainability methods such as LIME, and review evaluation metrics used to assess the performance of survival models.

Together, these sections aim to equip the reader with a clear understanding of the theoretical principles that inform the design and implementation of the proposed methodology.

3.2 Machine Learning Models

In 1956, a group of computer scientists introduced the foundational idea that computers could emulate human thought and reasoning. They proposed that "every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it" [25]. This idea marked the formal birth of the field of *artificial intelligence* (AI), dedicated to automating cognitive tasks traditionally performed by humans.

Within the broader scope of AI, machine learning (ML) and deep learning (DL) have emerged as core subfields that enable systems to learn from data and improve performance over time. ML algorithms leverage historical data to perform tasks such as prediction, classification, clustering, and dimensionality reduction. These algorithms iteratively refine their performance as more data become available, making them powerful tools across various industries [26].

For example, recommendation systems used by e-commerce platforms, social media, and news aggregators rely on ML to personalize content based on users' behavioral history. In the context of autonomous driving, ML models, often in combination with computer vision, are essential for object detection, decision-making, and navigation. In biomedical research, ML facilitates the analysis of high-dimensional, complex datasets, identifying nonlinear associations, interactions, latent structures, and subgroups that might be missed by traditional parametric statistical methods. Unlike classical models, which typically require

explicit assumptions and predefined relationships, ML algorithms are often more flexible and data-driven.

There are four primary types of learning in ML: supervised, unsupervised, semi-supervised, and reinforcement learning [26].

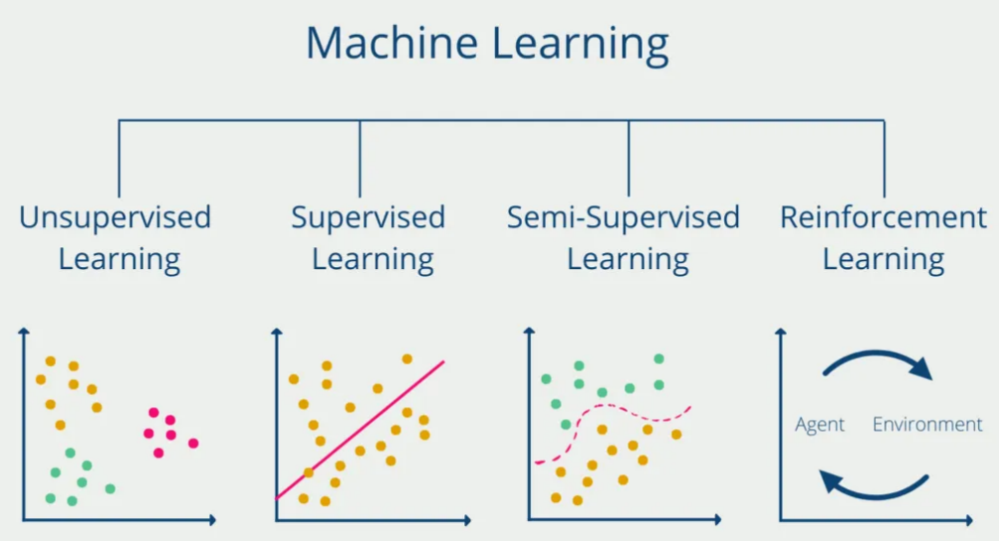


Figure 3.1: Types of Machine Learning

3.2.1 Supervised and Unsupervised Learning

The two most commonly used learning paradigms in machine learning are *supervised* and *unsupervised learning*, each addressing different types of problems based on the availability and structure of labeled data.

In **supervised learning**, the model is trained on a labeled dataset, where each input sample is paired with a corresponding target output. The objective is to learn a mapping from inputs to outputs that generalizes well to unseen data. Supervised learning is widely used in classification tasks (e.g., predicting cancer subtypes from histopathological images) and regression tasks (e.g., predicting patient survival time from genomic features). Common algorithms include support vector machines (SVMs), decision trees, random forests, and various types of neural networks [26].

On the other hand, **unsupervised learning** deals with unlabeled data, aiming to uncover hidden patterns, structures, or groupings within the dataset. Since no ground-truth labels are provided, the algorithm attempts to organize the data in a meaningful way based solely on its internal structure. This paradigm is commonly used for clustering (e.g., grouping patients based on gene expression profiles), dimensionality reduction (e.g., PCA, t-SNE, or autoencoders), and anomaly detection. Unsupervised learning is particularly useful in exploratory data analysis and in domains where labeling is expensive or infeasible.

These two learning paradigms serve as the foundation for many real-world machine learning applications, and more advanced approaches, such as semi-supervised learning and self-supervised learning, seek to combine their respective advantages. [27]

3.2.2 Introduction to Deep Learning

Deep learning is a subfield of machine learning that focuses on neural network architectures with multiple layers, enabling the automatic learning of hierarchical representations from data [28]. While traditional machine learning often relies on handcrafted features and domain-specific preprocessing, deep learning methods can learn both low-level and high-level abstractions directly from raw inputs.

The defining characteristic of deep learning models is the presence of multiple layers, typically composed of linear transformations followed by non-linear activation functions, which allow the network to progressively capture more complex patterns. These models are particularly effective for tasks involving unstructured data such as images, audio, and text.

Deep learning has demonstrated remarkable success in areas including computer vision, natural language processing, and biomedical data analysis [26]. In image-based tasks, convolutional neural networks (CNNs) are often employed due to their ability to capture spatial hierarchies and local dependencies. In other domains, architectures such as recurrent neural networks (RNNs), transformers, and autoencoders are commonly used.

The strength of deep learning lies in its scalability, flexibility, and ability to generalize

from large volumes of data. However, it often requires extensive computational resources and careful tuning of hyperparameters to achieve optimal performance.

3.3 Neural Networks

A foundational understanding of neural networks and deep learning models is essential for comprehending the concepts and methodologies presented in this work. Therefore, it is appropriate to first outline the fundamental theory of neural networks along with an overview of the main architectures commonly used in image processing.

3.3.1 Artificial Neural Networks

The term Artificial Neural Networks (ANNs) refers to a class of artificial intelligence algorithms inspired by the structure and information processing mechanisms of biological neural networks in the human brain. ANNs are composed of interconnected nodes, analogous to neurons, which are linked by weighted connections. By adjusting these weights through training, the network learns from input data via activation functions that govern the output of each node.

The concept of artificial neural networks was first introduced in 1943 by neurophysiologist Warren S. McCulloch and mathematician Walter Pitts in their seminal paper *A Logical Calculus of the Ideas Immanent in Nervous Activity* [29]. For several decades following their introduction, ANNs showed limited practical success. However, in recent years, the field has experienced remarkable progress due to the development of more sophisticated architectures, such as Convolutional Neural Networks (CNNs) [30, 31], as well as advancements in computational hardware, including powerful Graphics Processing Units (GPUs) and specialized accelerators like Tensor Processing Units (TPUs) [27]. These developments have significantly enhanced the performance and applicability of neural networks across a wide range of domains.

3.3.2 Network Architecture and Forward Propagation

An ANN typically consists of an input layer, one hidden layer, and an output layer. Each neuron in a given layer is connected to every neuron in the next layer through weighted connections. The forward propagation process involves computing a weighted sum of the inputs and applying an activation function to produce the neuron's output. Mathematically, the output $a_j^{(l)}$ of a neuron j in a layer l is given by:

$$a_j^{(l)} = \phi\left(\sum_i w_{ji}^{(l)} a_i^{(l-1)} + b_j^{(l)}\right) \tag{3.1}$$

where $w_{ji}^{(l)}$ are the weights, $b_j^{(l)}$ is the bias, and ϕ is the activation function.

3.3.3 Deep Neural Networks

The term Deep Neural Networks (DNNs) refers to a subclass of Artificial Neural Networks (ANNs) characterized by the presence of multiple hidden layers between the input and output layers [28]. This architectural depth enables DNNs to learn increasingly abstract and hierarchical representations of data, making them particularly effective for tasks involving high-dimensional inputs such as image and speech recognition [32], natural language processing [33], and biomedical data analysis [34].

In DNNs, each hidden layer applies a linear transformation followed by a nonlinear activation function, allowing the network to progressively extract higher-level features from the input. Compared to shallow networks, which contain only one or two hidden layers, DNNs are capable of modeling more intricate patterns and dependencies [26], although they also require more computational resources and are more susceptible to overfitting if not properly regularized.

The widespread adoption and success of DNNs have been facilitated by the availability of large-scale annotated datasets, substantial computational power through modern GPUs and TPUs, and key algorithmic advancements. These include improved weight initializa-

tion methods [35], adaptive optimization algorithms such as Adam [36], and regularization techniques like dropout [37]. Together, these developments have solidified deep learning as a foundational approach in fields ranging from computer vision and genomics to survival prediction in clinical settings.

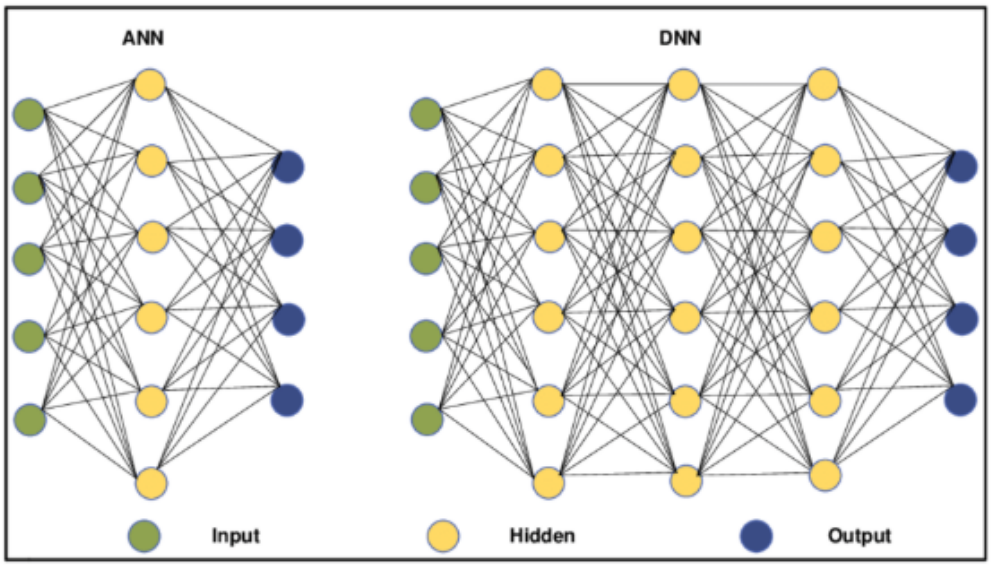


Figure 3.2: Visual comparison between a shallow Artificial Neural Network (ANN) and a Deep Neural Network (DNN) [38].

3.3.4 Activation Functions

To enable the network to capture complex, non-linear relationships in the data, activation functions are used after each neuron's linear transformation. Activation functions introduce non-linearity into the network, enabling it to learn complex patterns. Common choices include:

- **Sigmoid:** $\sigma(x) = \frac{1}{1+e^{-x}}$
- **Hyperbolic Tangent (tanh):** $\tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$
- **ReLU (Rectified Linear Unit):** $\text{ReLU}(x) = \max(0, x)$
- **Leaky ReLU:** Leaky $\text{ReLU}(x) = \max(\alpha x, x)$ where α is a small constant

ReLU and its variants are commonly used in deep networks due to their efficiency and ability to mitigate vanishing gradient issues [39].

3.3.5 Loss Functions and Optimization

Training a neural network involves adjusting its parameters to minimize the discrepancy between predictions and ground truth. This is quantified using loss functions. The most common loss functions are:

- **Mean Squared Error (MSE):** used for regression tasks.
- **Cross-Entropy Loss:** used for classification problems.

To minimize the loss function, optimization algorithms such as Stochastic Gradient Descent (SGD) or Adam [36] are employed. These algorithms update the network weights iteratively through the process of backpropagation.

3.3.6 Backpropagation and Learning

Backpropagation is the core algorithm used for training neural networks. It calculates the gradient of the loss function with respect to each weight using the chain rule of calculus, and then applies the optimizer to update the weights in the direction that reduces the loss.

3.3.7 Regularization and Overfitting

One of the main challenges in training neural networks is overfitting, where the model learns the training data too well and fails to generalize to unseen data. Several regularization techniques are commonly employed:

- **L1/L2 Regularization:** Add penalty terms to the loss function based on the magnitude of weights.

- **Dropout:** Randomly deactivates neurons during training to encourage redundancy and robustness [37].
- **Early Stopping:** Terminates training when the performance on a validation set starts to deteriorate.

3.3.8 Depth and Expressive Power

Deeper neural networks can model more complex functions but are also more prone to issues such as vanishing gradients and increased training times. The Universal Approximation Theorem shows that even a single hidden layer can approximate any continuous function, given sufficient neurons, but in practice, deeper architectures often yield better performance [40].

3.3.9 Convolutional Neural Networks(CNNs)

Convolutional Neural Networks (CNNs) were developed to address the growing complexity of data and the increasing size of datasets. They have become the foundational architecture for image processing tasks [20], replacing earlier neural networks that relied on fully connected layers, which were inefficient and computationally expensive due to their dense connectivity. CNNs are specifically designed to automatically learn patterns and features from images and other grid-like data through the use of convolutional and pooling layers. This hierarchical feature extraction makes them particularly effective for tasks such as image classification, object detection, and facial recognition [5].

The typical architecture of a Convolutional Neural Network [4] includes the following components:

- **Input layer**: Receives the raw image data.
- **Convolutional layers**: Apply filters (kernels) to extract local features.

- **Pooling layers**: Downsample the feature maps to reduce dimensionality and computational load.
- **Fully connected layers**: Perform high-level reasoning based on the extracted features.
- **Output layer**: Produces the final prediction.

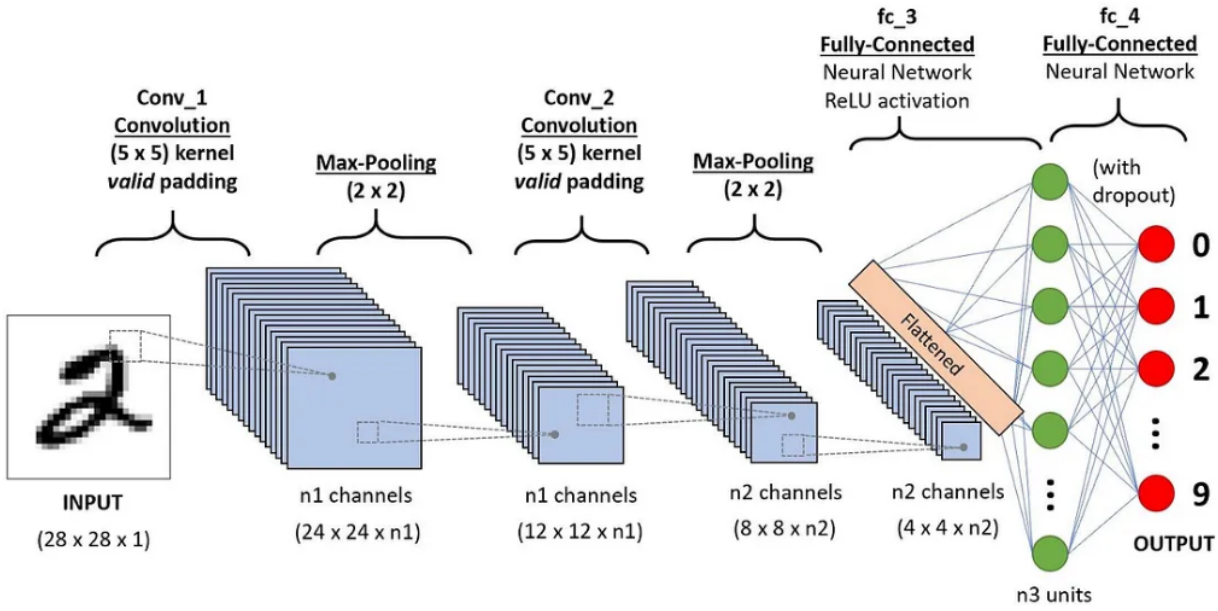


Figure 3.3: Example of a Convolutional Neural Network architecture.

The most important and unique parts of a CNN are the Convolution Layers, the Pooling Layers and the Fully connected Layers. The functionality and usage of those types of layers are analyzed below.

Convolutional Layer (or Kernel)

The convolutional layers of a CNN operate on the basis of a defined algorithm. Let h, w, and c represent the height, width, and number of channels of the input feature map, respectively, and let K denote the convolutional kernel (or filter), a matrix with dimensions

$k \times k$. A convolutional operation is applied to each $k \times k$ submatrix of the input with a sliding step determined by the stride S. For instance, if $S = 2$, the kernel is applied to every other $k \times k$ region, effectively reducing the spatial dimensions of the input by a factor of two. This downsampling is performed while attempting to preserve the most informative features, thanks to the learned values of the kernel.

Mathematically, the 2D convolution operation between an input feature map X and a kernel K can be expressed as:

$$Y(i,j) = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} K(m,n) \cdot X(i+m, j+n) \tag{3.2}$$

where $Y(i,j)$ is the output feature map at position (i,j), and (m,n) iterates over the elements of the kernel. This operation slides the kernel over the input feature map to compute localized feature responses.

When the number of input channels c is greater than 1, as in the case of RGB images ($c = 3$), the kernel K must also have the same depth. In such cases, a 3D kernel of shape $k \times k \times c$ is used to perform the convolution across all channels simultaneously. The outputs from all channels are summed to produce a single response per spatial location.

To control the spatial dimensions of the output, it is common to apply *padding* to the input. Padding involves adding rows and columns of predefined values (typically 0, 1, or the average of neighboring pixels) around the input feature map. This technique ensures that the output dimensions are not reduced after convolution.

Figure 3.4 illustrates the convolution process described above. The kernel K has dimensions 3×3 and depth 3, matching the depth of the input image. After the convolution is applied to each channel, the results are summed and passed through a non-linearity after adding a bias term. In this example, zero-padding of one row and column is applied around the input.

The main target of the convolution layer is to obtain high-level features from the input image. High-level features can be lines, edges, or blobs, but also more abstract patterns such

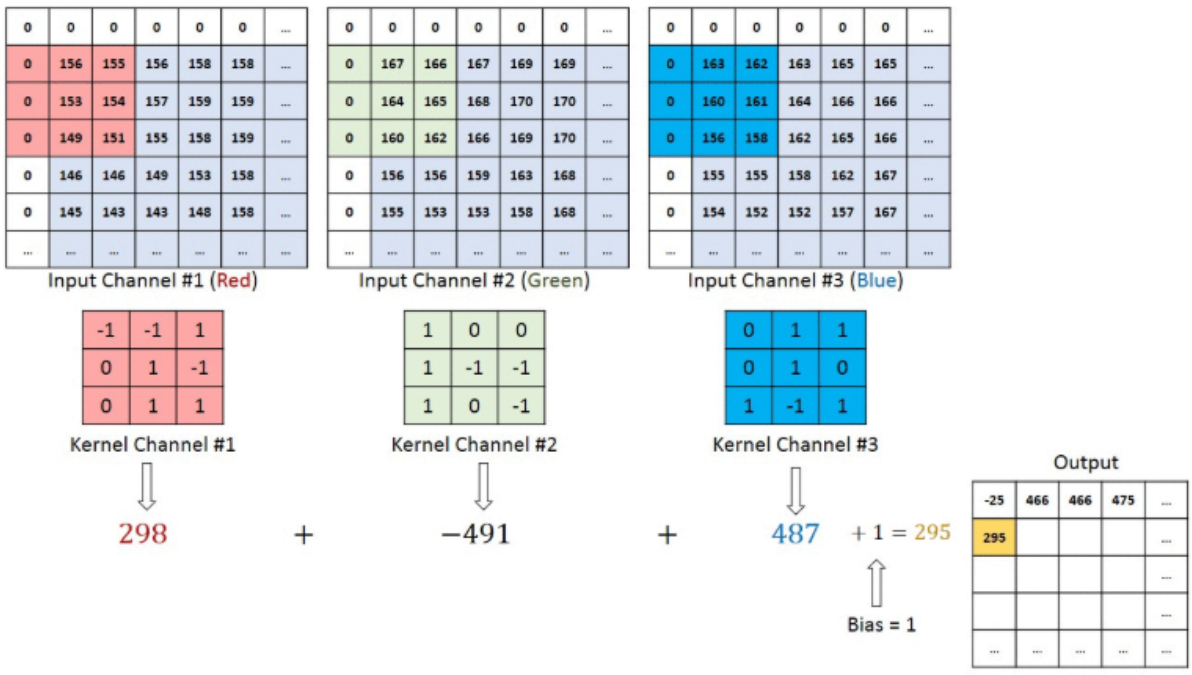


Figure 3.4: An example of a convolution operation.

as tires in cars or windows in buildings. In addition, the layer captures low-level features such as color gradients and orientation, which are essential for early visual understanding.

Pooling Layer

Pooling layers operate in a manner similar to convolutional layers but do not involve the application of a kernel for convolution. Instead, a pooling layer replaces each sub-region (typically a small square matrix) of the input with a single representative value. The specific value depends on the type of pooling operation used. The two most common types are:

- **Max Pooling**: Returns the maximum value from each sub-region.
- **Average Pooling**: Returns the average value of the elements in each sub-region.

Typically, the input to a pooling layer is the output of a preceding convolutional layer. The primary function of pooling is to reduce the spatial dimensions of the input, thereby

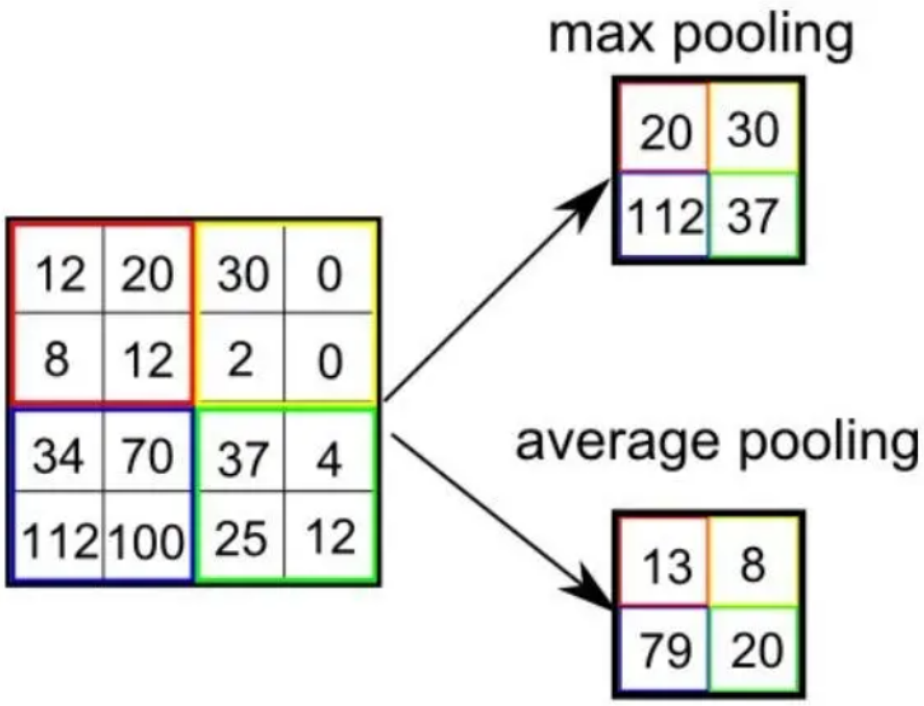


Figure 3.5: Example of a Pooling Layer architecture.

decreasing the computational and memory complexity of subsequent layers. Moreover, pooling helps in extracting dominant features that are invariant to small transformations such as rotation and translation. This invariance enhances the robustness of the model. Max pooling, in particular, is often used as a noise-suppressing operation, as it discards low-activation (and often noisy) values while preserving the most prominent features.

Fully Connected Layers

Fully connected (FC) layers, also known as dense layers, are typically used in the final stages of a Convolutional Neural Network. These layers receive flattened inputs, often feature maps resulting from convolutional and pooling layers, and perform high-level reasoning to produce the final output, such as class scores or regression values.

Mathematically, a fully connected layer performs a linear transformation followed by a non-linear activation:

$$\mathbf{y} = f(\mathbf{W} \cdot \mathbf{x} + \mathbf{b}) \tag{3.3}$$

where $\mathbf{x}$ is the input vector, $\mathbf{W}$ is the weight matrix, $\mathbf{b}$ is the bias vector, and f is a non-linear activation function (e.g., ReLU, sigmoid, or softmax). Every neuron in an FC layer is connected to all neurons in the previous layer, making it dense in connectivity but computationally expensive.

Although fully connected layers offer high expressiveness, they also introduce a large number of parameters and are prone to overfitting. In modern architectures, their use is sometimes minimized or replaced entirely by global average pooling layers, especially in lightweight or fully convolutional designs.

Modern Advances in Convolutional Neural Networks

In recent years, Convolutional Neural Networks (CNNs) have made substantial progress both in architectural design and overall performance. The availability of larger and more diverse datasets has played a crucial role in training more robust and generalizable models. Furthermore, techniques such as batch normalization, data augmentation, and the implementation of regularization methods have significantly improved the learning process, enhancing both generalization and robustness. The development of deeper network architectures has enabled CNNs to capture increasingly complex patterns and hierarchical features from input data. Additionally, the introduction of deep pre-trained models, such as ResNet [41], Inception [42], and EfficientNet [41], has transformed the field. These architectures allow researchers to leverage *transfer learning*, where models trained on large-scale datasets can be fine-tuned for different but related tasks, reducing training time and improving performance.

3.3.10 Transfer Learning and Pre-trained Models

Transfer learning is a powerful technique in deep learning that enables models trained on a large, generic dataset to be repurposed for a different but related task with limited data. This approach is especially advantageous in biomedical applications, where labeled data are often scarce or expensive to obtain. In the context of Convolutional Neural Networks (CNNs), transfer learning typically involves reusing a model that has been pre-trained on a large-scale dataset such as ImageNet [43]. The lower layers of these networks are capable of capturing general features such as edges, textures, and basic shapes, which are broadly useful across various visual recognition tasks.

By initializing a model with pre-trained weights instead of training from scratch, transfer learning reduces the computational cost and shortens the convergence time during training. Moreover, it improves performance in domains where the amount of training data is insufficient to support the full training of deep architectures. Fine-tuning, where selected layers of the pre-trained model are updated during training, allows the network to adapt its parameters to the specific characteristics of the target dataset while still benefiting from the knowledge encoded in the original weights.

In this work, transfer learning was applied by employing several widely used pre-trained CNN architectures, each with different design philosophies and parameter capacities. These models include ResNet-50, EfficientNet-B0 and B1, and KimiaNet, a model specifically designed for histopathological image analysis. The subsequent subsections provide a brief overview of each architecture and discuss their relevance and performance in the context of the current study.

3.3.11 ResNet-50

ResNet-50 is a 50-layer deep convolutional neural network introduced as part of the Residual Network (ResNet) family by He et al. [41]. The key innovation behind ResNet is the introduction of *residual connections*, or skip connections, which allow the network

to learn residual mappings instead of direct mappings. These connections help mitigate the vanishing gradient problem commonly encountered in very deep architectures, thereby enabling the effective training of networks with substantially more layers.

The ResNet-50 architecture consists of an initial convolutional layer followed by four stages, each containing multiple bottleneck residual blocks. These blocks comprise a series of convolutional layers with identity shortcut connections that bypass one or more layers. This design facilitates efficient gradient flow and improves convergence during training.

Due to its strong representational power and stability during optimization, ResNet-50 has become a widely adopted backbone in computer vision tasks, including those in the biomedical imaging domain. The architecture of ResNet-50 is shown in Figure 3.6, illustrating its residual block structure with skip connections that facilitate gradient flow and enable training of deeper networks.

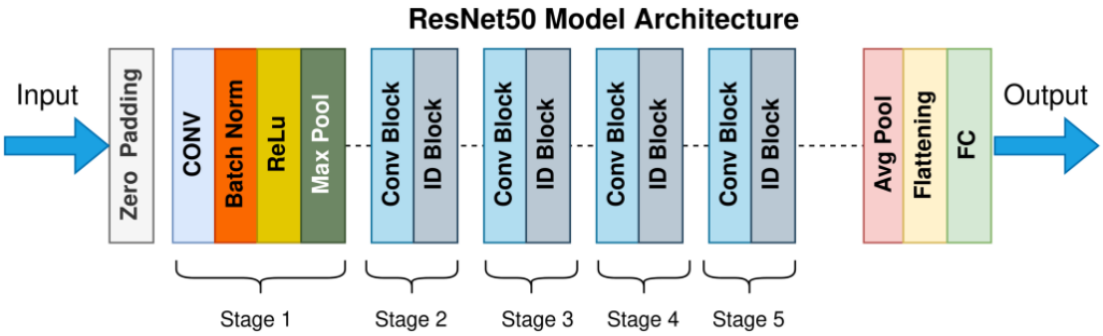


Figure 3.6: Schematic representation of the ResNet-50 architecture. Residual blocks with identity shortcut connections allow for deeper network training by mitigating the vanishing gradient problem. [44]

3.3.12 EfficientNet

EfficientNet is a family of convolutional neural networks introduced by Tan and Le [45], based on the idea of optimizing model performance through compound scaling. Unlike traditional approaches that scale network dimensions (depth, width, resolution) independently, EfficientNet employs a compound coefficient to uniformly scale all three dimensions in a balanced manner. This results in a family of models that achieve state-of-the-art accuracy

while maintaining high computational efficiency.

The baseline model, EfficientNet-B0, is discovered using neural architecture search and serves as the foundation for the scaled-up variants (B1, B2, etc.). Each successive variant increases the model capacity and input resolution, enabling better performance on more complex tasks. All EfficientNet models use a series of inverted bottleneck blocks with depthwise separable convolutions, which contribute to both parameter efficiency and representational power.

EfficientNet models have been widely adopted in computer vision due to their excellent trade-off between accuracy and computational cost. Their design makes them particularly suitable for applications where resource constraints are important, such as mobile or embedded systems, as well as in domains like medical imaging, where processing high-resolution data efficiently is essential. The architecture of EfficientNet is shown in Figure 3.7, showcasing the compound scaling approach and the use of inverted bottleneck MBConv blocks throughout the network.

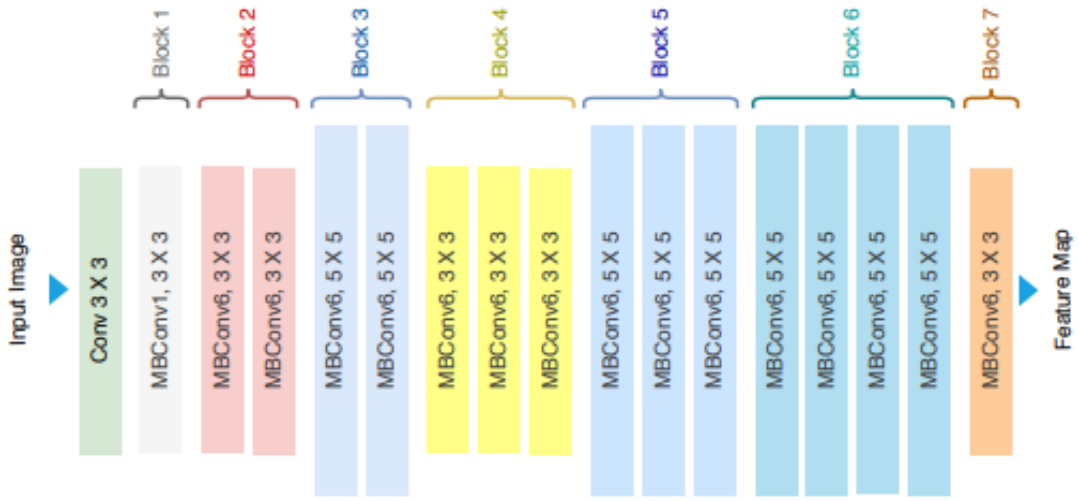


Figure 3.7: Overview of the EfficientNet architecture. The network scales depth, width, and resolution in a compound manner and uses MBConv blocks to balance efficiency and performance. [46]

3.3.13 DenseNet-121

DenseNet-121 is a deep convolutional neural network introduced by Huang et al. [47], known for its unique connectivity pattern where each layer receives input from all preceding layers. This dense connectivity leads to several advantages: improved gradient flow, enhanced feature reuse, and reduced parameter count compared to traditional architectures of similar depth. DenseNet-121 has been widely adopted in medical imaging tasks due to its efficiency and strong performance in capturing complex visual patterns. The architecture of DenseNet-121 is shown in Figure 3.8, highlighting its dense block structure and the flow of feature maps across layers.

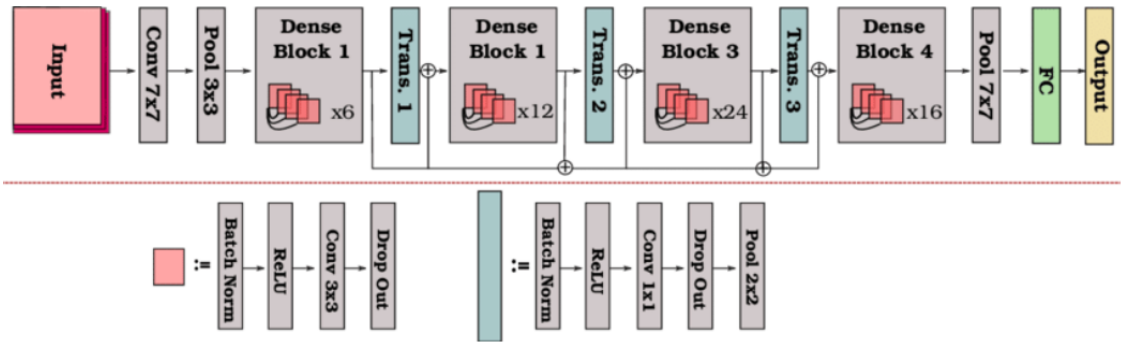


Figure 3.8: Schematic representation of the DenseNet-121 architecture. The network is composed of dense blocks with direct connections between all layers, interleaved with transition layers that reduce spatial dimensions. [47]

KimiaNet

KimiaNet is a customized version of DenseNet-121 fine-tuned specifically for histopathological image analysis. It was trained on the Kimia Path24 dataset [48], which consists of whole-slide histopathology images from a diverse set of tissue types. By leveraging domain-specific data, KimiaNet captures morphological features that are unique to histopathology, such as cellular architecture, texture, and staining variations.

Unlike standard DenseNet-121 models pre-trained on natural image datasets like ImageNet, KimiaNet is better aligned with the distributions and semantics of medical images.

This makes it particularly well-suited for downstream tasks in computational pathology, such as cancer detection, subtype classification, and image retrieval. Its success illustrates the importance of domain adaptation in transfer learning, especially when dealing with highly specialized data types like histological slides.

3.3.14 Encoder-Decoder Architectures and Autoencoders

Encoder-decoder architectures are a fundamental framework in modern deep learning, widely employed in tasks that involve transforming input data into structured or meaningful output formats. These models are composed of two main components: the *encoder*, which compresses the input into a lower-dimensional latent representation, and the *decoder*, which reconstructs or transforms this representation into the desired output. This separation between encoding and decoding stages enables efficient representation learning, as well as flexible adaptation to various modalities, including images, text, audio, and time-series data [49, 50]. The encoder is designed to extract the most informative features from the

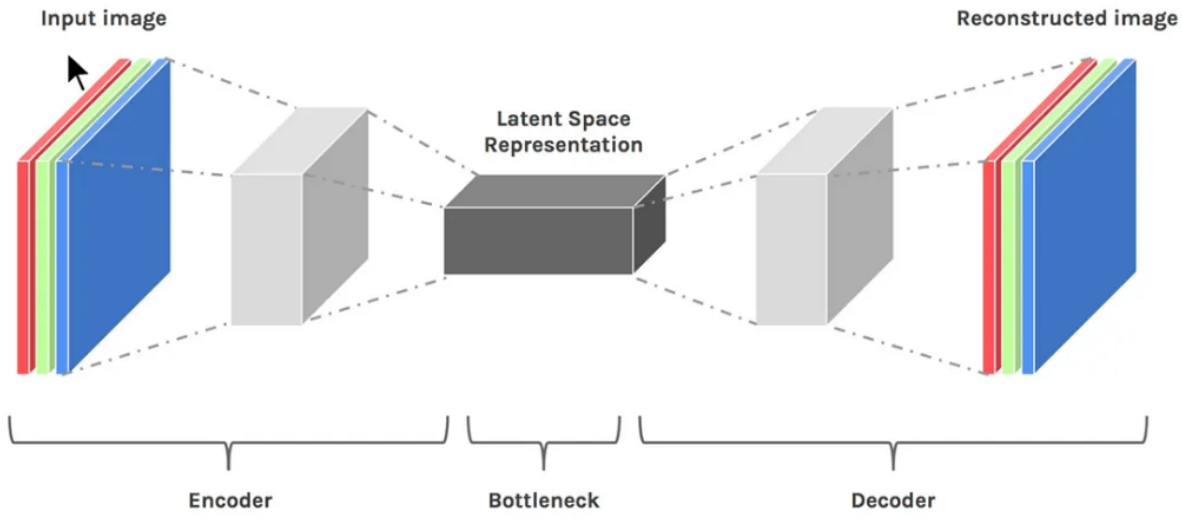


Figure 3.9: Example of an Autoencoder architecture.

input while filtering out irrelevant or noisy components. Depending on the data type, this can be achieved using different neural layers, such as convolutional layers for spatial data

(e.g., images), recurrent or transformer-based layers for sequential data (e.g., text or signals), or fully connected networks for tabular data. The resulting latent space captures a compact summary of the input and serves as a bottleneck that encourages generalization and information compression.

One of the most prominent uses of this architecture is in the form of *autoencoders*, which are trained to reconstruct the original input from its compressed representation. Autoencoders are typically trained in an unsupervised manner, minimizing a reconstruction loss (such as mean squared error) between the input and the output. The goal is not merely to memorize the data but to discover an efficient encoding that captures its underlying structure.

Autoencoders are widely used in a variety of applications, including dimensionality reduction, denoising, anomaly detection, and unsupervised feature learning. In many pipelines, they are employed as a preprocessing step to transform raw, high-dimensional input into compact, informative features that can be used for downstream tasks such as classification or survival analysis.

Several variants of autoencoders have been proposed to enhance their capabilities:

- **Denoising Autoencoders (DAEs)**: Trained to reconstruct clean inputs from noisy versions, improving robustness.
- **Sparse Autoencoders**: Encourage sparsity in the latent representation, leading to more interpretable features.
- **Variational Autoencoders (VAEs)**: Extend autoencoders into a probabilistic framework by learning distributions over the latent space, enabling generative modeling and sampling [51].

Beyond autoencoders, encoder–decoder models are widely applied in supervised learning tasks such as machine translation, image segmentation, text summarization, and time-series prediction. In these cases, the decoder is trained to generate task-specific outputs conditioned

on the encoded representation. This general architecture has also been adopted in more complex models, such as sequence-to-sequence transformers and U-Net for biomedical image segmentation.

Overall, encoder–decoder architectures, including autoencoders and their variants, provide a powerful and flexible framework for learning meaningful representations and performing complex data transformations across a wide range of domains.

3.4 RNA-Sequencing and Gene Expression Data

RNA sequencing (RNA-seq) is a widely used technology for profiling gene expression in biological samples [52].

It involves converting RNA molecules into complementary DNA (cDNA), sequencing the cDNA fragments, and mapping the resulting reads to a reference genome or transcriptome. This enables the quantification of transcript abundance across thousands of genes simultaneously [52, 53].

One commonly used unit for quantifying expression is Transcripts Per Million (TPM), which normalizes for both gene length and sequencing depth [54]. TPM values allow for within-sample and between-sample comparisons of gene expression levels, making them suitable for downstream statistical or machine learning analyses [55].

In this context, the term *unstranded* refers to a preprocessing mode in which the RNA strand information is not preserved during sequencing or alignment, simplifying the analysis when strand specificity is not essential [56].

Gene expression datasets derived from RNA-seq are typically high-dimensional, often containing tens of thousands of genes per sample. This high dimensionality presents challenges, including multicollinearity, sparsity, and overfitting. To overcome these issues, preprocessing steps such as normalization, transformation, and dimensionality reduction are commonly applied before applying statistical or machine learning models. RNA-seq data has been used

extensively in biomedical research to identify disease subtypes, uncover molecular biomarkers, and predict clinical outcomes [53].

3.5 Survival Analysis

Survival analysis refers to a set of statistical methods used to analyze the time until the occurrence of an event of interest, such as death, disease recurrence, or progression. A key challenge in this domain is the presence of censored data, where the outcome is not observed for all subjects during the study period. This phenomenon—referred to as censoring—occurs because not all individuals have experienced the event by the end of the observation period [57].

Several types of censoring may occur in practice:

(a) The patient has not yet experienced the event (e.g., relapse or death) by the time the study ends.

(b) The patient is lost to follow-up during the study period, making further observation impossible.

(c) A competing event occurs that precludes continued follow-up for the original event of interest.

In such cases, the recorded survival time reflects only a lower bound on the true time-to-event. This form of incomplete observation, where the event is expected to occur after the follow-up window, is known as **right censoring**. Conceptually, one can imagine the survival time as a timeline extending beyond the study horizon, the actual event lies beyond the point of observation. Other types of censoring include *left censoring* and *interval censoring*, though they are less frequently encountered in biomedical research.

Unlike standard regression techniques that assume fully observed outcomes, survival analysis models must jointly account for the time-to-event and the censoring indicator. This

requirement gives rise to specialized estimators and models, which form the foundation of both classical and modern survival analysis.

Survival analysis methods are generally used for three main purposes [58]:

- To **describe** the distribution of survival times within a group, using tools such as life tables, Kaplan–Meier estimators, survival functions, and hazard functions.
- To **compare** survival experiences between groups, often with statistical tests like the log-rank test.
- To **model** the relationship between covariates (categorical or continuous) and survival time, using regression-based techniques such as the Cox proportional hazards model, parametric survival models, or tree-based approaches like survival trees and random survival forests.

This section outlines the main classical survival models and introduces DeepSurv, a deep learning-based framework that enables flexible and data-driven survival modeling.

3.5.1 Classical Methods of Survival Analysis

Kaplan–Meier Estimator

The Kaplan–Meier (KM) estimator is a non-parametric method used to estimate the survival function from censored data [59]. It computes the probability of survival past a certain time point and generates a step-wise survival curve. The KM estimator is widely used for visualizing survival distributions and comparing groups (e.g., treated vs. control), often with the log-rank test [60].

While intuitive and widely applicable, the KM estimator does not account for covariates. This limits its use when individual-level characteristics are expected to influence survival outcomes [57].

Cox Proportional Hazards Model

The Cox proportional hazards model is a semi-parametric approach that incorporates covariates into the survival analysis [61]. The hazard function for an individual with covariates X is modeled as:

$$h(t|X) = h_0(t) \exp(\beta^\top X) \tag{3.4}$$

where $h_0(t)$ is the baseline hazard function, β is a vector of coefficients, and $\exp(\beta^\top X)$ models the relative risk. This formulation does not require a parametric form for $h_0(t)$, which makes the model flexible and interpretable.

However, the model relies on the **proportional hazards assumption**, which implies that the effect of covariates on the hazard is constant over time. Violations of this assumption, as well as the presence of complex non-linear interactions, limit the Cox model's performance, especially in high-dimensional or multi-modal biomedical datasets [62].

3.5.2 Deep Learning for Survival Analysis: The DeepSurv Model

To overcome the limitations of classical models, deep learning methods have been introduced into survival analysis. Among them, **DeepSurv** is one of the most widely adopted neural architectures for learning non-linear risk functions [63].

DeepSurv extends the Cox model by replacing the linear term $\beta^\top X$ with a learned function $f_\theta(X)$, represented by a fully connected neural network with parameters θ:

$$h(t|X) = h_0(t) \exp(f_\theta(X)) \tag{3.5}$$

This formulation maintains the interpretability of risk scores while enabling the modeling of complex, non-linear dependencies in the data. The network is trained by minimizing the negative log partial likelihood:

$$\mathcal{L}(\theta) = -\sum_{i \in \mathcal{D}} \left(f_\theta(X_i) - \log \sum_{j \in \mathcal{R}(T_i)} \exp(f_\theta(X_j)) \right) \tag{3.6}$$

where $\mathcal{D}$ is the set of uncensored cases, and $\mathcal{R}(T_i)$ is the risk set of individuals still at risk at time T_i.

Model Architecture and Training

DeepSurv typically consists of multiple dense layers with non-linear activations (e.g., ReLU or LeakyReLU). To avoid overfitting, especially in small sample settings, regularization techniques such as L2 weight decay, dropout, and early stopping are commonly applied. Training can be optimized using stochastic gradient descent (SGD) or its variants (e.g., Adam).

Because hyperparameters such as learning rate, hidden layer size, and number of layers significantly affect performance, DeepSurv is often combined with hyperparameter optimization frameworks such as Optuna.

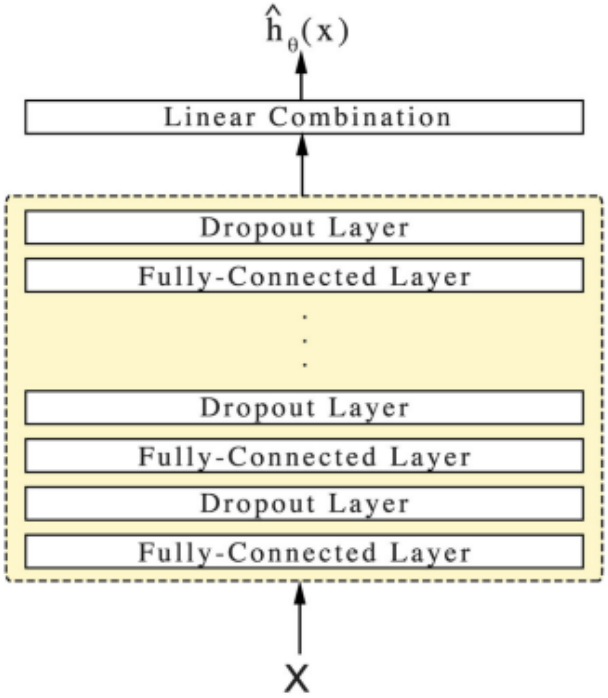


Figure 3.10: The Deepsurv architecture.

3.6 Optuna Hyperparameter Optimization Framework

Optuna is an open-source framework for automated hyperparameter optimization, widely used in machine learning and deep learning workflows [64]. It follows a define-by-run approach, allowing dynamic and flexible construction of the hyperparameter search space. Optuna is framework-agnostic, meaning it can be integrated with any model implementation, regardless of the underlying ML library.

To use Optuna, the user defines an *objective function* that wraps the model training and evaluation logic. A `trial` object is used within this function to suggest hyperparameter values, and the function returns a metric (e.g., accuracy, AUC, MSE) to be optimized. A `study` object manages the overall optimization process, including the direction (minimization or maximization), the number of trials, and storage options.

Optuna also supports advanced features such as pruning of unpromising trials, checkpointing, and resuming interrupted studies. Additionally, it provides visualization tools to analyze optimization history and hyperparameter importance.

3.7 Feature Fusion Techniques

In multimodal learning tasks, such as those involving gene expression profiles and histopathological images, the integration of heterogeneous data sources is a crucial step toward improving predictive performance. Feature fusion refers to the process of combining features derived from different modalities into a unified representation, enabling models to capture complementary information.

Feature fusion strategies are generally classified into three categories: early fusion (feature-level), intermediate fusion, and late fusion (decision-level). Early fusion involves the direct concatenation of feature vectors extracted from each modality into a single vector before feeding it into a learning algorithm [65]. This approach is simple and efficient, but may suffer from differences in data scale, modality-specific noise, and varying dimensionalities.

Intermediate fusion methods aim to learn joint representations through specialized architectures such as co-attention mechanisms, multimodal autoencoders, or neural tensor fusion layers [66]. These methods seek to model complex cross-modal relationships during feature learning, which can enhance the interaction between modalities but often require careful design and more computational resources.

Late fusion, on the other hand, combines the outputs (e.g., risk scores or class probabilities) of separate unimodal models through methods such as averaging, majority voting, or meta-learners [67]. This approach is modular and robust to modality-specific failure, but may miss fine-grained interactions between modalities.

In the context of survival analysis, where the objective is to model time-to-event data, feature fusion plays a critical role in capturing both morphological and molecular determinants of patient prognosis. Recent studies have demonstrated that combining features from histopathological images and genomic data can significantly improve survival prediction models for cancer patients [68, 69].

Choosing the appropriate fusion strategy depends on the nature of the modalities, the size of the dataset, and the complexity of the relationships between features. While simple concatenation is widely used due to its straightforward implementation, more advanced fusion techniques are actively being explored to better exploit multimodal information.

3.8 Explainability in Deep Learning

Deep learning models have shown remarkable success in a wide range of biomedical applications, including image classification, genomics, and survival prediction. However, their "black-box" nature raises significant concerns regarding interpretability, especially in critical domains such as healthcare, where transparency is essential for trust, validation, and clinical adoption [70].

Explainable AI (XAI) refers to a set of techniques designed to interpret and understand

the decision-making processes of complex models. In the context of deep learning, XAI methods aim to highlight which input features contribute most to a specific output, enabling insights into the model's behavior and promoting fairness, accountability, and reproducibility [71].

Among the most widely used model-agnostic approaches are LIME and SHAP, which provide local explanations by analyzing the influence of features on individual predictions.

3.8.1 Local Interpretable Model-agnostic Explanations (LIME)

LIME (Local Interpretable Model-agnostic Explanations) is an approach that approximates a complex model locally with an interpretable surrogate model, typically a linear regressor or decision tree [72]. For each prediction, LIME perturbs the input data around the instance of interest and observes how the black-box model's output changes. By fitting a simple, interpretable model to this neighborhood, LIME identifies which features have the most significant impact on the prediction.

In biomedical applications, LIME has been applied to tasks such as gene expression classification, histopathological image analysis, and survival modeling, where understanding local decision boundaries helps clinicians validate model predictions and identify relevant biological markers [73].

Despite its utility, LIME has limitations, including sensitivity to sampling noise and inconsistency across runs. Additionally, the quality of its explanations depends heavily on how well the surrogate model fits the local decision boundary.

SurvLIME A notable extension of LIME tailored to survival analysis is SurvLIME [74]. Traditional LIME is designed for models with scalar outputs, such as probabilities or risk scores. However, survival models often produce more complex outputs, such as time-dependent survival probabilities or hazard functions. SurvLIME adapts the LIME framework to this context by perturbing input features and observing the corresponding changes in sur-

vival risk estimates or predicted survival curves. It then fits a local interpretable model to approximate the behavior of the survival model around each instance.

This approach allows researchers to identify which features most influence the predicted survival outcomes at the individual level. SurvLIME has been used to interpret both classical and deep survival models, helping bridge the gap between model predictions and clinical understanding. It is particularly useful in high-stakes settings, such as oncology, where explaining why a patient is assigned a higher risk can guide treatment decisions and foster trust in AI-assisted prognostics.

3.9 Evaluation Metrics

Evaluating the performance of machine learning models is a critical component of the modeling process, as it provides insights into the model's generalization ability and helps guide decisions during development and optimization. Different types of problems, such as classification, regression, and survival prediction, require different evaluation metrics, each tailored to the structure of the outputs and the goals of the analysis.

3.9.1 General Evaluation Metrics in Machine Learning

In supervised learning, evaluation metrics vary depending on the nature of the target variable. For classification tasks, metrics such as accuracy, precision, recall, F1-score, and area under the Receiver Operating Characteristic curve (AUC-ROC) are commonly used [75]. For regression problems, mean squared error (MSE), mean absolute error (MAE), and the coefficient of determination (R^2) are typical choices [76].

3.9.2 Challenges in Survival Analysis

Survival analysis differs from traditional regression or classification in that it deals with time-to-event data, where some observations may be censored. This renders many standard

evaluation metrics inapplicable, as they assume fully observed outcomes. Consequently, specialized evaluation metrics are employed to assess the performance of survival models, accounting for both the time-to-event nature of the data and the presence of censored cases.

3.9.3 Concordance Index (C-index)

The most commonly used metric in survival analysis is the Concordance Index (C-index), which generalizes the AUC-ROC to censored data [77]. It measures the model's ability to correctly rank survival times: a higher predicted risk score should correspond to a shorter observed survival time. Formally, the C-index is defined as:

$$C = \frac{1}{|\mathcal{P}|} \sum_{(i,j)\in\mathcal{P}} I\left(\hat{r}_i > \hat{r}_j\right) \tag{3.7}$$

where $\mathcal{P}$ is the set of comparable pairs (i.e., pairs for which the order is known despite censoring), $\hat{r}_i$ is the predicted risk score for individual i, and I is the indicator function. A C-index of 1 indicates perfect concordance, while 0.5 corresponds to random guessing.

3.9.4 Integrated Brier Score (IBS)

The Brier score evaluates the accuracy of probabilistic predictions over time by comparing predicted survival probabilities with actual outcomes. The Integrated Brier Score (IBS) is the time-integrated version of the Brier score, taking into account censoring through inverse probability of censoring weights (IPCW) [78]:

$$\text{IBS} = \frac{1}{T} \int_0^T \text{BS}(t)\, dt \tag{3.8}$$

where $\text{BS}(t)$ is the Brier score at time t, and T is a pre-defined maximum follow-up time. Lower IBS values indicate better performance, with 0 being a perfect score.

3.9.5 Time-dependent ROC Curves

Time-dependent ROC analysis extends classical ROC curves to survival data by evaluating the true positive and false positive rates at specific time points [79]. For each time t, the sensitivity and specificity of the model in discriminating between individuals who experience the event before t and those who survive beyond it are computed. The area under the time-dependent ROC curve (AUC_t) provides a dynamic assessment of predictive performance.

3.9.6 Calibration Measures

In addition to discrimination (e.g., C-index), survival models should also be evaluated for *calibration*, the agreement between predicted and observed survival probabilities. Calibration curves compare predicted survival probabilities with empirical estimates, often using Kaplan-Meier curves within predicted risk strata [80].

In survival analysis, evaluation metrics must reflect both the discriminatory power of the model and its ability to correctly estimate survival probabilities in the presence of censored data. The Concordance Index, Integrated Brier Score, and time-dependent ROC analysis are the most frequently used tools for this purpose. A comprehensive evaluation involves multiple metrics to assess different aspects of model performance, such as ranking accuracy, probability estimation, and calibration.

Chapter 4

Methodology

4.1 TCGA Dataset

The primary dataset utilized in this study is derived from The Cancer Genome Atlas (TCGA), a comprehensive and publicly available repository of molecular and clinical data across a wide range of cancer types [81]. Specifically, this work focuses on patients diagnosed with primary brain tumors, including glioblastoma multiforme (GBM) and lower-grade glioma (LGG). These datasets provide multi-modal data comprising genomic measurements, histopathological whole-slide images, and detailed clinical information.

The TCGA repository offers standardized and curated data, making it particularly suitable for integrative analyses and machine learning applications. For each patient, clinical variables such as survival time, event status (i.e., alive or deceased), age at diagnosis, and tumor grade are available.

In this thesis, the data set was filtered to include only patients for whom all three data modalities, RNA-sequencing, histopathological images, and survival annotations, were available. Data preprocessing steps, including normalization and dimensionality reduction, are described in the following sections.

4.2 Data Overview

This study integrates two complementary data types: histopathological whole-slide images (WSIs) and RNA-sequencing gene expression profiles. These two modalities offer distinct insights into tumor biology, capturing both visual tissue architecture and intracellular molecular activity. Together, they provide a rich foundation for building multi-modal models for survival prediction.

RNA-Sequencing Data

The RNA-sequencing data provide a transcriptomic snapshot of the tumor's molecular profile. Each patient's sample contains expression levels of thousands of genes, quantified in terms of transcripts per million (TPM). These measurements allow us to infer biological pathways, signaling mechanisms, and gene activity patterns associated with disease progression. The RNA-seq data used in this study were collected from the TCGA project and matched to the same patients for whom histopathological images were available.

Histopathological Images

Histopathological images are digitized slides of tissue sections stained with hematoxylin and eosin (H&E), revealing important structural features such as cell morphology, tissue organization, and presence of necrosis or infiltration. In this work, we utilize H&E-stained slides from brain tumor patients in the TCGA database.

These high-resolution images, typically scanned at 40× magnification, are stored in the SVS format and may exceed several gigabytes in size.

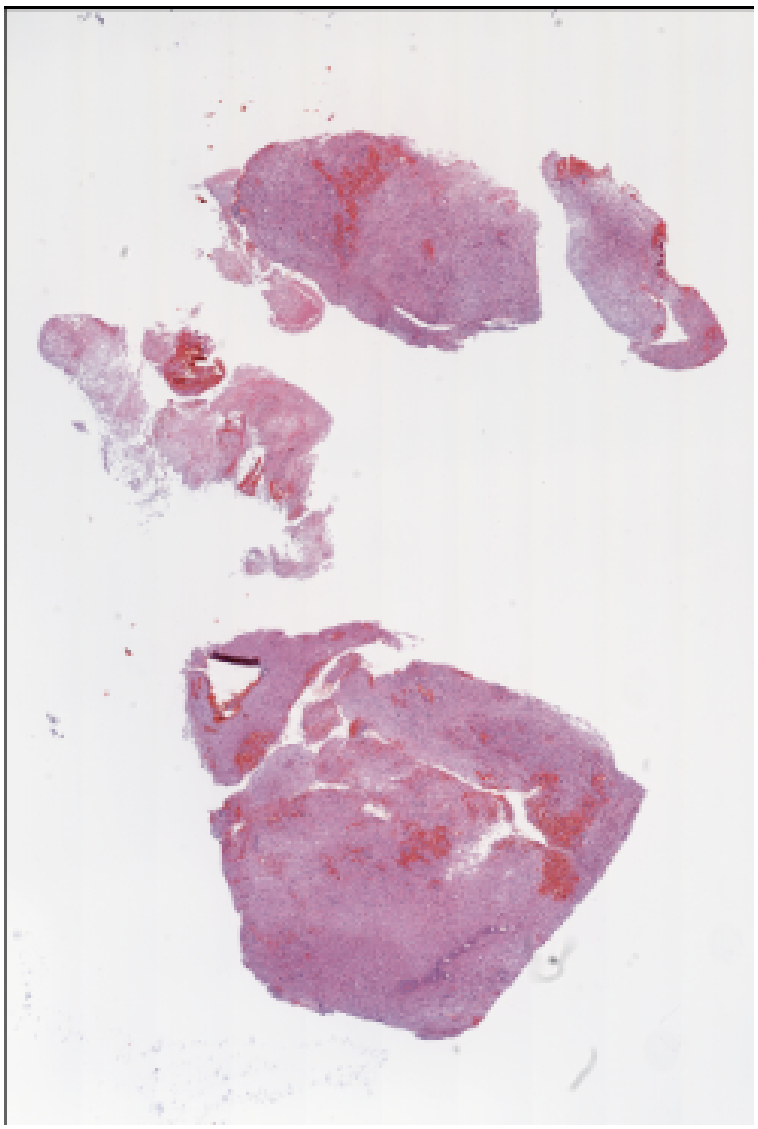


Figure 4.1: Example of an H&E-stained histopathological slide from a glioblastoma patient.

4.3 Preprocessing Steps

To ensure consistency across patients and extract meaningful representations from both data modalities, a series of preprocessing steps were applied prior to model development. These steps were tailored to the unique properties of histopathological images and RNA-sequencing data, enabling robust feature extraction and harmonized survival label computation. Each case was linked to clinical metadata obtained from TCGA's JSON records. The fields submitter_id, vital_status, days_to_death, and days_to_last_follow_up were used to

define survival outcomes as follows:

- **Event:** A binary indicator set to 1 if the patient was deceased (vital_status = Dead), and 0 otherwise.
- **Time until event:** The survival time in years, computed from diagnosis to either death (for uncensored events) or last follow-up (for censored observations).

This processing step ensured consistency between the image data and survival labels used for model training and evaluation.

4.3.1 Histopathological Image Preprocessing

To construct case-level representations from whole-slide histopathological images (WSIs), we implemented a multi-stage preprocessing pipeline. This pipeline included the parsing of clinical metadata, detection of tissue regions, extraction of high-resolution patches, and transformation of each patch into a fixed-length feature vector using deep convolutional neural networks (CNNs). The resulting vectors were aggregated into a single representation per patient, suitable for downstream survival analysis.

Tissue Mask Generation and Patch Extraction

To isolate informative tissue regions from WSIs, a color-based tissue detection algorithm was applied. Each slide was downsampled and converted from RGB to HSV color space. Tissue presence was inferred using Otsu's thresholding on both the RGB and saturation channels, followed by RGB-based minimum intensity filtering to suppress background noise.

Morphological operations (dilation and erosion) were then applied to refine the resulting binary tissue masks, removing isolated pixels and smoothing tissue contours.

Using the final tissue mask, up to 1000 patches of size 224×224 pixels were extracted per WSI at the highest available resolution. Patch coordinates were recorded for downstream analysis. Only patches that passed both a tissue coverage threshold and a contrast-based

filter were retained, ensuring that low-content or background regions were excluded from feature extraction.

Feature Extraction Using CNN Backbones

To embed each patch into a lower-dimensional representation, we used several pretrained CNNs as feature extractors. The tested architectures included EfficientNet-B0, B1 [45], ResNet-50 [41], and KimiaNet [48]. All models were used without their classification heads. Input patches were normalized with ImageNet statistics and passed through the convolutional backbone. Features were then extracted from the final convolutional layers using global average pooling (GAP).

Attention-Based Aggregation at the Sample Level

For each sample (i.e., whole-slide image), the extracted patch-level feature vectors were further processed using a self-attention mechanism to emphasize the most informative regions. Specifically, scaled dot-product attention was computed across all patch features within the same sample:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \tag{4.1}$$

where Q, K, and V denote the matrices of patch features and d_k is the dimensionality of the feature vectors. In this implementation, $Q = K = V$, allowing the model to learn contextual relationships among patches.

The attention-weighted patch features were subsequently averaged to obtain a single aggregated vector per sample. Additionally, prior to attention computation, dimensions that were empirically found to be consistently uninformative were removed to reduce noise and improve interpretability.

Mean Pooling Across Samples to Form Case-Level Representations

Each patient case may consist of multiple WSIs (samples). Following attention-based aggregation at the sample level, the resulting sample vectors were combined using mean pooling to form a unified case-level representation:

$$\mathbf{f}_{\text{case}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{f}_{\text{sample}_i} \tag{4.2}$$

where N is the number of samples associated with the patient. This final case-level feature vector served as the input for the autoencoder architecture that was used after.

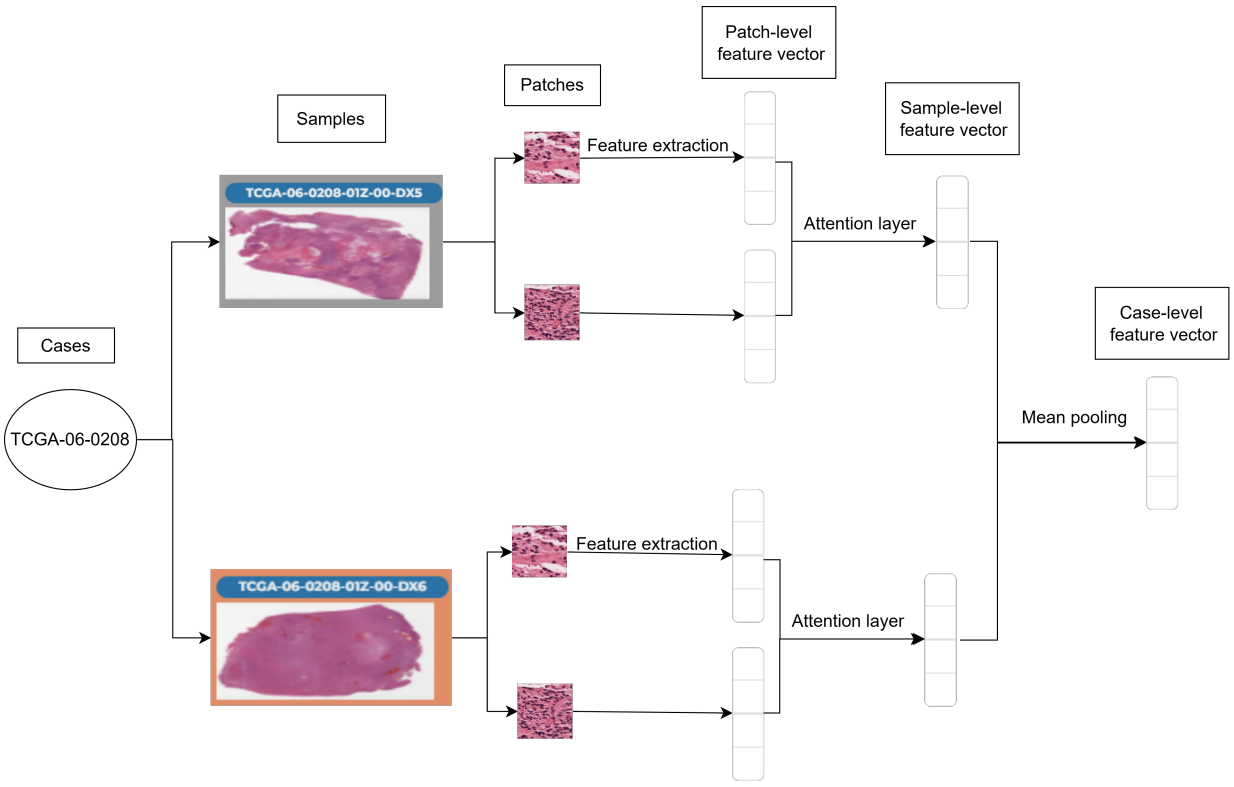


Figure 4.2: Overview of the histopathological image preprocessing pipeline. Tissue masks are generated from WSIs, followed by patch extraction, feature encoding, and aggregation.

4.3.2 RNA-sequencing Data Preprocessing

RNA-sequencing (RNA-seq) data was processed to generate standardized numerical representations of gene expression for each patient. In contrast to histopathological image processing, RNA-seq preprocessing was more direct and computationally efficient. The primary goal was to extract transcript abundance measurements in the form of TPM (Transcripts Per Million) values, followed by appropriate normalization and dimensionality reduction steps to enable robust survival modeling.

Metadata Parsing and File Matching

RNA-seq files were organized according to patient identifiers provided in TCGA metadata files. Using the JSON records, the submitter_id and associated file_name fields were extracted. Only files with the .tsv extension and containing the tpm_unstranded column were retained for analysis.

A patient-to-file mapping was constructed to ensure consistent and accurate retrieval of the correct RNA-sequencing data for each case. This mapping allowed automated loading of RNA-seq measurements and alignment of molecular data with corresponding survival labels and histopathological samples.

Transformation and Filtering

For each matched RNA-seq file, expression values from the tpm_unstranded column were extracted. To prepare the RNA-seq data for survival modeling, the TPM values underwent the following transformation and filtering steps:

- **Log transformation:** TPM values were log-transformed to reduce right-skewness and stabilize variance across gene expression levels.
- **Imputation:** Missing values, if present, were imputed using the median expression value of the corresponding gene across the cohort.

- **Low-variance filtering:** Genes with variance below a fixed threshold (0.1) were removed to eliminate uninformative features and reduce dimensionality.

- **Z-score normalization:** The remaining gene expression values were standardized across patients to have zero mean and unit variance.

The final preprocessed RNA-seq dataset consisted of 583 samples in rows and nearly 20,000 selected genes as columns, with all values normalized and suitable for downstream modeling. The submitter_id column was retained to enable alignment with clinical and image-based data, as well as for cross-modal feature fusion.

4.4 Histopathological Image-based model

Following preprocessing and feature extraction of histopathological images, a complete analytical pipeline was constructed to perform survival prediction using image-derived representations. This framework incorporated dimensionality reduction via autoencoders and survival modeling using the DeepSurv architecture, whose theoretical foundations were discussed in previous chapters.

4.4.1 Dimensionality Reduction with Autoencoders

The extracted case-level features, obtained through attention-based aggregation of patch-level CNN representations, were typically high-dimensional. To reduce noise, alleviate overfitting, and improve interpretability, we employed a fully connected autoencoder to project the input features into a lower-dimensional latent space.

The autoencoder architecture consisted of an encoder and a decoder with symmetrical fully connected layers. The encoder compressed the input into a bottleneck layer of predefined size, while the decoder attempted to reconstruct the original input. The model was trained using mean squared error (MSE) loss and optimized with the Adam optimizer.

Several bottleneck sizes were evaluated (32, 64, 128, 256) to identify the optimal latent dimensionality for survival prediction. After cross-validation and downstream evaluation with DeepSurv, the bottleneck size of 32 consistently yielded the best performance in terms of validation concordance index. Therefore, it was selected as the final dimensionality for all subsequent survival modeling.

Feature standardization was applied using statistics computed from the training set, and the same transformation was applied to the test set. The encoded representations with 32 dimensions were saved and used as input to DeepSurv and multimodal fusion experiments.

4.5 RNA-sequencing based model

A parallel analytical pipeline was developed for the RNA-sequencing data, following the same structure as described in the histopathological Image-based model. As previously discussed, the pipeline included dimensionality reduction via an autoencoder, followed by survival modeling using DeepSurv, with hyperparameter optimization performed using Optuna.

4.5.1 Dimensionality Reduction with Autoencoders

The input to the model consisted of preprocessed and normalized TPM (Transcripts Per Million) gene expression values, as described in Section 4.3.2. Each patient was initially represented by tens of thousands of genes, which were filtered to retain only those exhibiting sufficient variance across the dataset.

To reduce dimensionality and denoise the high-dimensional RNA-sequencing data, a deep fully connected autoencoder was implemented. The architecture used for RNA-sequencing was deeper than the one used for image-derived features, incorporating additional hidden layers, batch normalization, and dropout to better handle the increased complexity and potential noise inherent in transcriptomic data.

Several bottleneck sizes were evaluated (32, 64, 128, 256) to identify the optimal latent representation for survival prediction. Based on cross-validation results and subsequent evaluation with DeepSurv, a bottleneck size of 32 yielded the highest validation concordance index and was therefore selected as the final dimensionality for all downstream modeling.

Feature standardization was applied using statistics from the training set, and the same transformation was consistently applied to the test set. The resulting 32-dimensional encoded representations were used as input to the DeepSurv model and for multimodal integration with image-based features.

4.6 Framework for Histopathological Image and RNA-sequencing data fusion

To leverage complementary information from both imaging and transcriptomic data, a feature-level fusion strategy was implemented. After independently extracting and reducing the dimensionality of each modality using autoencoders, the resulting latent representations were concatenated to form a unified feature vector for each case.

4.6.1 Feature Concatenation

Let $\mathbf{z}_{\text{image}} \in R^{d_1}$ denote the bottleneck representation derived from histopathological images and $\mathbf{z}_{\text{gene}} \in R^{d_2}$ the corresponding representation from RNA-sequencing data. The fused feature vector $\mathbf{z}_{\text{fused}}$ was constructed via:

$$\mathbf{z}_{\text{fused}} = [\mathbf{z}_{\text{image}} \parallel \mathbf{z}_{\text{gene}}] \in R^{d_1+d_2} \tag{4.3}$$

where $\parallel$ denotes vector concatenation. Each modality was independently standardized using training set statistics prior to fusion.

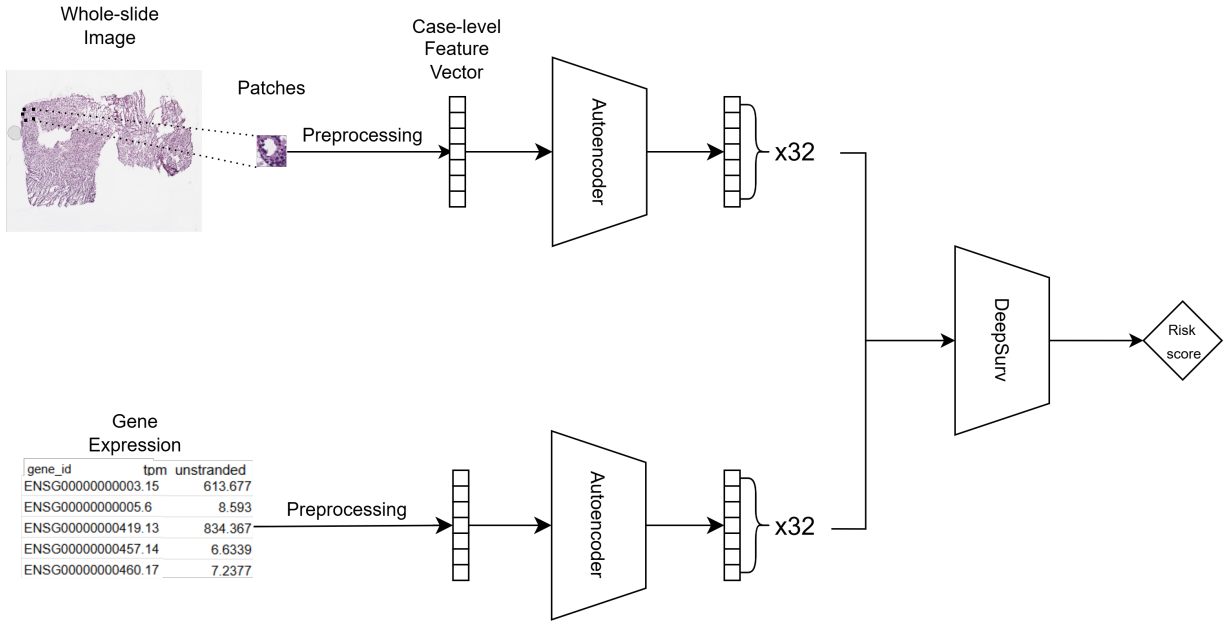


Figure 4.3: Fused Data Framework

4.7 Common Survival Modeling Strategy

4.7.1 Data Splitting and Stratification

For all three survival prediction models—image-based, RNA-seq based, and the fused model—the same data splitting and stratification strategy was applied to ensure robust evaluation and fair comparison. Initially, 10% of the data was held out as an independent test set, never used during training or hyperparameter tuning. The remaining 90% was used for training and validation within a 10-fold stratified cross-validation framework. In each fold, a further 10% of the training data was set aside as a validation set to support early stopping and performance monitoring.

Stratification was applied both during the initial test split and throughout the cross-validation procedure to preserve the distribution of key clinical variables across all subsets. Specifically, stratification was based on tumor grade, patient age, and survival time, ensuring balanced representation of prognostic factors. This approach helps reduce sampling bias and

enables consistent model comparison across different data modalities.

4.7.2 Hyperparameter Optimization

All three survival prediction models—image-based, RNA-seq based, and the fused model—employ the same downstream modeling strategy once feature vectors have been extracted. In each case, the low-dimensional representations, are used as input to a DeepSurv model [63]. As discussed in the previous chapter, DeepSurv is a deep neural network that generalizes the Cox proportional hazards model by learning a nonlinear risk function from covariates.

Each model variant was trained using the same pipeline: 10-fold stratified cross-validation on the training set, coupled with hyperparameter optimization via Optuna [64], a Bayesian optimization framework for efficient automated search. The following hyperparameter space was defined:

- **Number of hidden layers:** 1 to 3
- **Nodes per hidden layer:** 16 to 128
- **Learning rate:** log-uniform in the range 10^{-6} to 10^{-3}
- **Learning rate decay:** 0 to 0.01
- **Momentum:** 0.8 to 0.95
- **L2 regularization coefficient:** 0.001 to 1.0
- **Dropout rate:** 0.0 to 0.5

The objective function maximized the average concordance index (C-index) across folds. Early stopping and validation-based monitoring were used during training to prevent overfitting. The best-performing configuration identified by Optuna was retained and evaluated on the hold-out test set. This consistent modeling framework enables fair comparison across

modalities and isolates the contribution of each data source to survival prediction performance.

4.8 Interpretability

4.8.1 Interpretability of the RNA-sequencing based Model

For the interpretability analysis of the RNA-seq based model, SurvLIME was applied by perturbing the original gene expression values (prior to any transformation). To evaluate the model's behavior in a clinically meaningful and representative manner, we selected three test cases for SurvLIME analysis:

- One patient assigned a high risk score that experienced a death event in the first year of the observation period,
- One patient assigned a low risk score that was right-censored at year 5 of the observation period without experiencing any event,
- One patient assigned a low risk score that experienced a death event in the second year of the observation period.

These cases were chosen to capture both successful and erroneous predictions. The interpretability pipeline described below was applied separately to each of the three selected cases.

Specifically, for each test case, we generated 100 synthetic variants by randomly perturbing subsets of gene expression values. Each perturbed sample was then passed through the entire survival prediction pipeline, including log transformation, Z-score normalization, dimensionality reduction via autoencoder, and finally DeepSurv, to obtain a corresponding risk score.

To estimate gene-level importance, we fitted a linear regression model using the perturbed gene values as input features and the resulting DeepSurv risk scores as the target variable.

Since the DeepSurv model outputs higher scores for worse prognosis, we transformed the predicted risk scores by multiplying them by -1 prior to fitting the regression model. This transformation allowed us to interpret the SurvLIME coefficients in terms of survival rather than risk:

- A **positive importance score** indicates that the gene contributes to *higher survival probability* (i.e., it is protective).
- A **negative importance score** indicates that the gene is associated with *lower survival probability* (i.e., it is unprotective).

This interpretation convention was applied consistently across all cases.

The learned regression coefficients were used as importance scores, quantifying the contribution of each gene to the model's prediction for the selected case. The 30 most influential genes, ranked by the absolute value of their coefficients, were visualized to highlight those with the strongest local effect on the survival prediction.

4.8.2 Interpretability of the Fused Model

To assess how the inclusion of histopathological image features influences survival predictions, we extended the SurvLIME analysis to the fused model. In this combined architecture, features extracted from H&E-stained whole-slide images using a pretrained convolutional neural network (KimiaNet) were concatenated with encoded gene expression features from the RNA-seq based pipeline. The resulting fused feature vector was then fed into DeepSurv for risk prediction.

For interpretability, SurvLIME was applied by perturbing only the original gene expression values, while keeping the image-derived features fixed. Each perturbed sample was processed through the full multimodal pipeline, including autoencoder compression and DeepSurv inference, to generate a new risk score. A linear regression model was then fitted to estimate the local importance of each gene feature.

Chapter 5

Results

This chapter presents the experimental results obtained from the developed survival prediction models. Three distinct pipelines were evaluated: the histopathological image-based model, the RNA-sequencing based model, and the fused model that combines both data types.

Each model was trained and validated using 10-fold cross-validation, and final performance was assessed on a held-out test set. The main evaluation metric employed was the Concordance Index (C-index), which quantifies the agreement between predicted and actual survival rankings.

Additionally, we examine the correlation between predicted risk scores and actual survival times, visualize survival stratification using Kaplan–Meier curves, and provide interpretability analyses using SurvLIME to identify key predictive features and pathways in each modality. The results collectively demonstrate the predictive value of each data modality and the added benefit of multimodal fusion.

5.1 Performance Evaluation of Investigated Feature Extractors for the Histopathological Image-based Model

The concordance index (C-index) was used to evaluate model performance in terms of survival prediction. Initially, four different pre-trained CNN backbones were compared for histopathological image feature extraction. Table 5.1 summarizes the mean C-index across 10-fold cross-validation as well as the final test set C-index for each backbone.

Table 5.1: Cross-validation and final test set C-index for each pretrained feature extractor.

Pretrained Model	**Mean CV C-index**	**Final Test C-index**
EfficientNet-B0	0.79	0.78
EfficientNet-B1	0.76	0.74
ResNet-50	0.79	0.73
KimiaNet	**0.80**	**0.80**

Among the tested backbones, **KimiaNet** achieved the highest average performance and generalization ability. Given its domain-specific training on histopathological slides, KimiaNet was selected as the image feature extractor for the remainder of the study.

5.2 Obtained Performance using Different Data Modalities

Following the model selection, three pipelines were developed and compared: an image-only model using KimiaNet features, an RNA-seq based model using gene expression data, and a fused model combining both data types through feature concatenation. Figure 5.1 presents the C-index distributions across 10-fold cross-validation, along with the dashed lines indicating test set performance.

While the **fused model** achieved the highest test set C-index (0.91), the **RNA-seq based model** demonstrated slightly superior median performance across cross-validation

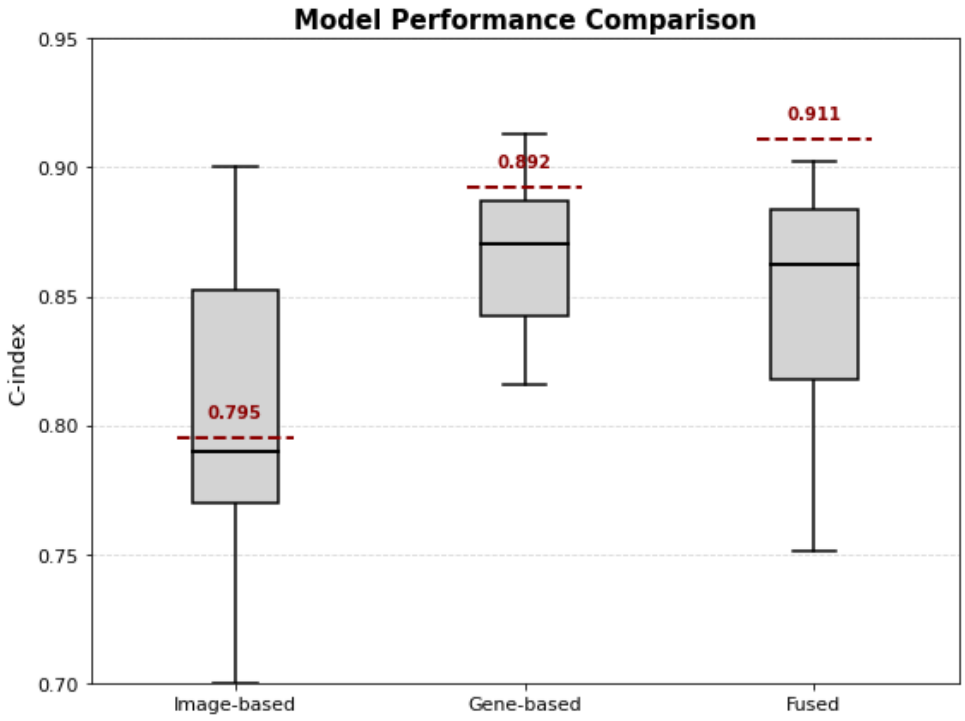


Figure 5.1: C-index distributions across 10-fold cross-validation for the three models. Dashed lines represent the final test set performance.

folds. This indicates that gene expression features alone carry strong prognostic value and generalize well across different subsets of the data.

The image-based model showed higher variance and lower overall C-index, which is expected due to the complex nature of histopathological images and their sensitivity to intratumoral heterogeneity and tissue artifacts. Nevertheless, the superior performance of the fused model on the test set confirms that integrating complementary modalities can improve overall generalization and robustness in survival prediction.

5.2.1 Comparison with State-of-the-Art Models

To evaluate the effectiveness of our proposed fused approach, we compared it against state-of-the-art models combining histopathology and genomic data, as well as relevant unimodal baselines (Table5.2). Our model demonstrated a C-index of 0.91 on the TCGA GBM + LGG cohort—a notable improvement over the image-only baseline (0.80) and even sur-

passing the RNA-seq based model (0.89). In comparison, Chen etal. (2020) using Pathomic Fusion achieved 0.826 on TCGA-GBM by integrating WSI and genomic data [82]. Fan etal. (2023) applied a Convolutional MIL framework on WSIs only, with a C-index of 0.67 [83]. Steyaert etal. (2022) combined TCGA and CPTAC data (WSI + genomic) and reached 0.822 [84]. The integrative PAGE-Net model (Hao etal., 2020) reported a C-index of approximately 0.702 on TCGA-GBM, outperforming its image-only (0.509) and Cox-PASNet genomic (0.640) variants [85]. In broader cancer analysis, Chen etal. (2022) reported a 0.821 C-index for LGG using pan-cancer integrative modeling [86], while Zhang etal. (2024) demonstrated a C-index of 0.849 on TCGA-LGG with CATFusion, a cross-attentional pan-cancer framework [87]. These results underscore the superior performance of our explainable fused model, confirming that fusing WSI and genomic features—when implemented with effective architectures—leads to substantial improvements in survival prediction for glioma patients.

Table 5.2: Comparison of the proposed model with state-of-the-art survival prediction models.

Model	**Dataset**	**C-index**
Proposed Fused Model	TCGA GBM, LGG (WSI + Genomic)	**0.91**
Image-based	TCGA (WSI)	0.80
RNA-seq based	TCGA (Genomic)	0.89
Chen et al. (2020) [82]	TCGA-GBM (WSI + Genomic)	0.826
Fan et al. (2023) [83]	TCGA (WSI)	0.67
Steyaert et al. (2022) [84]	TCGA, CPTAC (WSI + Genomic)	0.822
Hao et al. (2020) [85]	TCGA GBM (WSI + genomic)	0.70
Chen et al. (2022) [86]	TCGA LGG (WSI + genomic)	0.821
Zhang et al. (2024) [87]	TCGA LGG (WSI + genomic)	0.849

5.3 Evaluation of Fused Model

5.3.1 Predicted Risk Scores vs. Survival Time

Figure 5.2 illustrates the relationship between predicted risk scores from the fused model and actual survival times. As expected, a general inverse trend is observed: patients with

higher predicted risk tend to have shorter survival durations. This negative correlation confirms that the model captures clinically meaningful information when ranking patients by risk.

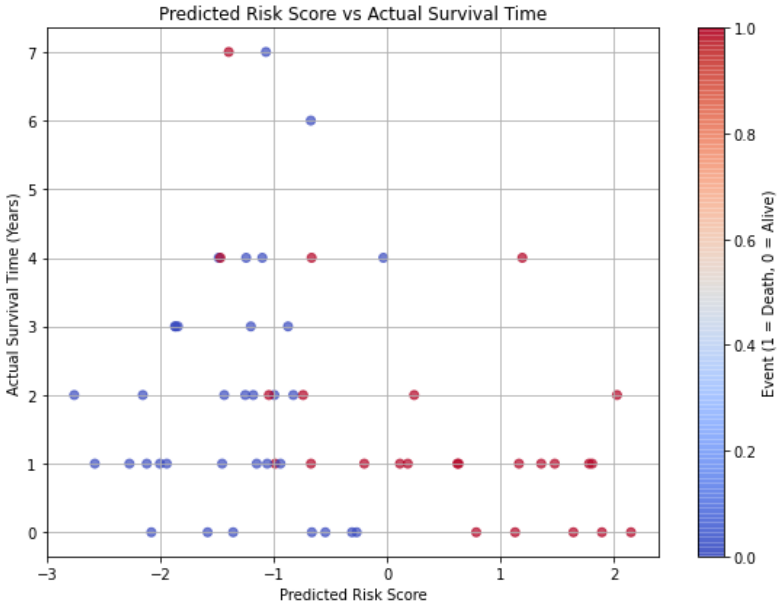


Figure 5.2: Predicted risk scores vs. actual survival times for the fused model. A negative correlation confirms the model's ability to rank patients according to survival risk.

Nonetheless, some degree of overlap is evident, particularly among patients with intermediate risk scores, highlighting regions where the model's predictive certainty is lower. This variability may reflect underlying biological heterogeneity or limitations in the representational capacity of the fused feature space.

5.3.2 Kaplan–Meier Curves by Predicted Risk

To evaluate the clinical utility of the predicted risk scores, patients in the test set were stratified into low- and high-risk groups using the median predicted risk score from the fused model as a threshold. Figure 5.3 presents the Kaplan–Meier survival curves for both groups.

As expected, the high-risk group exhibited significantly reduced survival durations compared to the low-risk group, indicating that the model's risk stratification aligns with clinically relevant outcomes. However, the two curves intersect around year four, suggesting

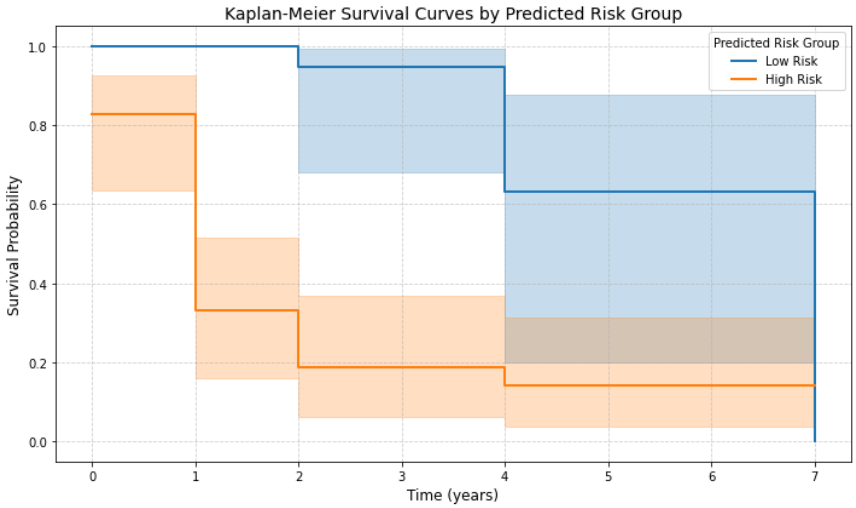


Figure 5.3: Kaplan–Meier survival curves for low- and high-risk groups (defined by the median predicted risk from the fused model).

diminished discriminative ability at longer time horizons. This could be attributed to late-occurring adverse events in patients initially predicted as low-risk, or to limitations in the model's long-term predictive capacity. Future work could incorporate time-dependent AUC analysis or alternative stratification schemes to further investigate this phenomenon.

5.4 Interpretability

5.4.1 Pathway Interpretations of RNA-sequencing based Model and Fused Model

To move beyond individual gene-level insights and explore broader biological mechanisms, we selected the top 1000 genes ranked by their absolute SurvLIME importance scores. These genes were mapped to curated biological pathways using the Reactome database [88]. For each pathway, we computed a pathway-level importance score by averaging the importance values of all constituent genes associated with that pathway.

This analysis enabled us to identify the 20 most influential pathways contributing to the model's survival risk predictions. These results offer a biologically meaningful context

for understanding the collective impact of gene sets rather than isolated genes. The top pathways are visualized in the figures presented below.

Subsequently, we repeated the pathway-level analysis using the gene importance scores derived from the multimodal setting. Interestingly, the set of top-ranked pathways shifted when image features were introduced, suggesting that morphological information can modulate the model's reliance on certain molecular pathways. This result underscores the complementary nature of imaging and transcriptomic modalities in capturing survival-relevant patterns in glioma.

5.4.2 Biological Interpretation of Important Pathways Across Cases

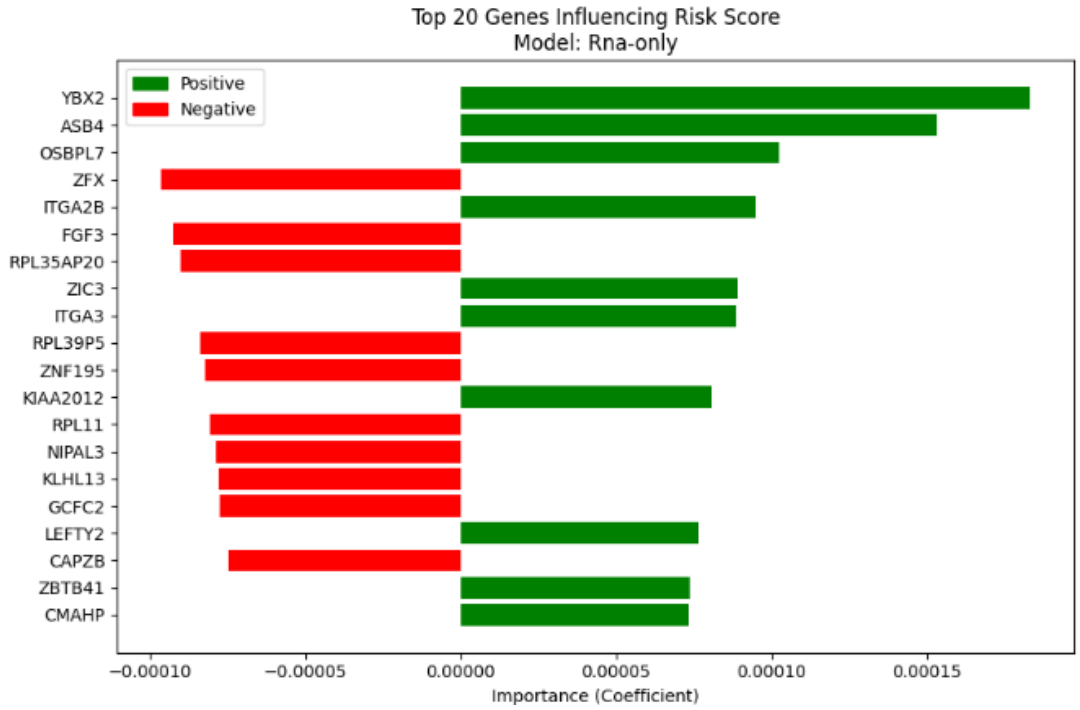
To further interpret the internal decision mechanisms of the models, we analyzed the most important biological pathways identified by SurvLIME for each of the three representative cases. This analysis highlights how the model's focus shifts across scenarios and data modalities.

Patient assigned a low risk score that experienced a death event in the second year of the observation period (Patient 1) In the RNA-seq based model, the top pathways included *MAP2K and MAPK activation* (R-HSA-5674135) and *Signaling by BRAF and RAF1 fusions* (R-HSA-6802952), both exhibiting negative importance scores. These are central components of the MAPK cascade, a signaling axis frequently dysregulated in glioblastoma and associated with enhanced cell proliferation, survival, and therapy resistance [89, 90]. Their contribution to increased predicted risk aligns with their oncogenic roles.

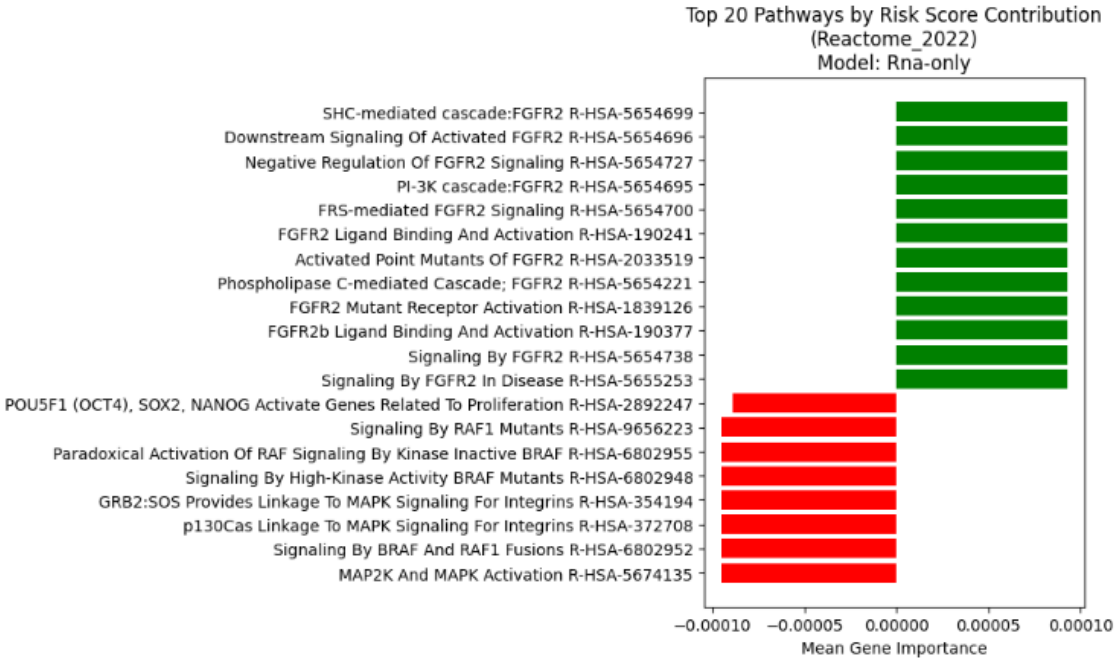
Interestingly, the pathway *p130Cas linkage to MAPK signaling for integrins* (R-HSA-372708) showed a positive importance score, suggesting a protective effect in this case, potentially reflecting context-specific modulation of integrin signaling [91].

In the fused model, the influence of molecular features was altered. The top-ranked

pathway was *Synthesis of bile acids and bile salts* (R-HSA-192105), which does not have a known role in brain tumors and may reflect incidental metabolic variation. Notably, the *p130Cas linkage to MAPK signaling for integrins* pathway reappeared, but with a negative importance score, indicating a shift toward an hazardous interpretation in the presence of histopathological features. Another high-ranking pathway, *COPI-dependent Golgi-to-ER retrograde traffic* (R-HSA-6811434), has been implicated in stress responses and protein homeostasis and may reflect tumor adaptation mechanisms [92].

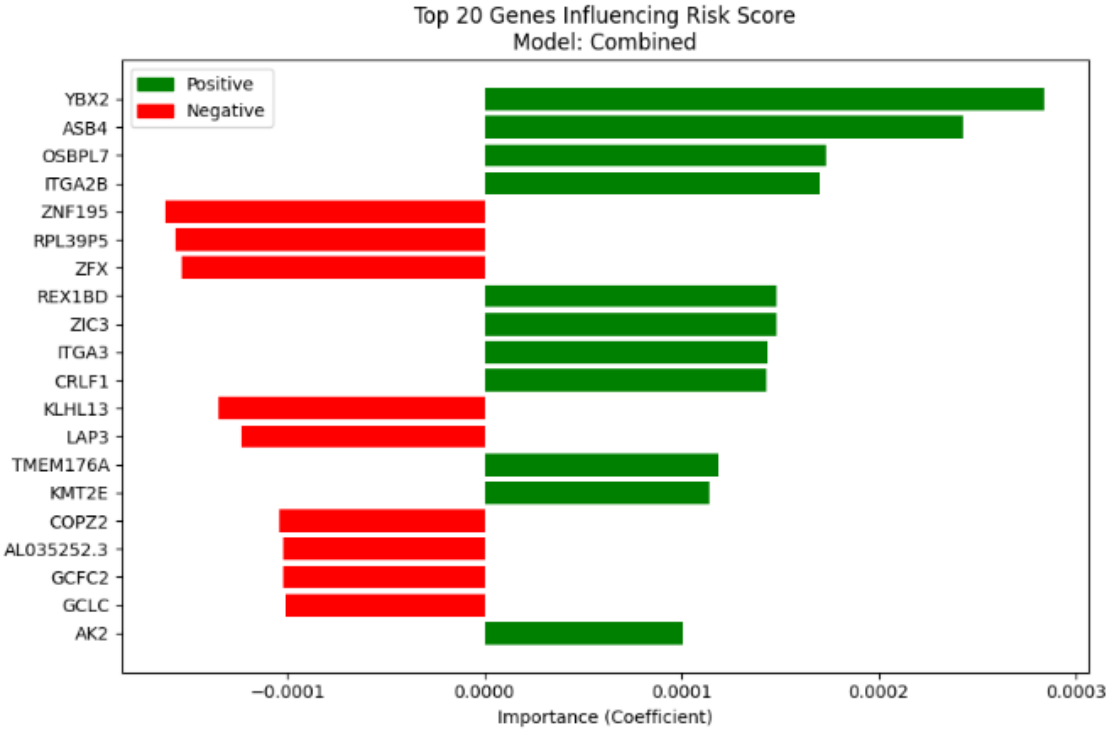


(a) Top genes (RNA-seq based model)

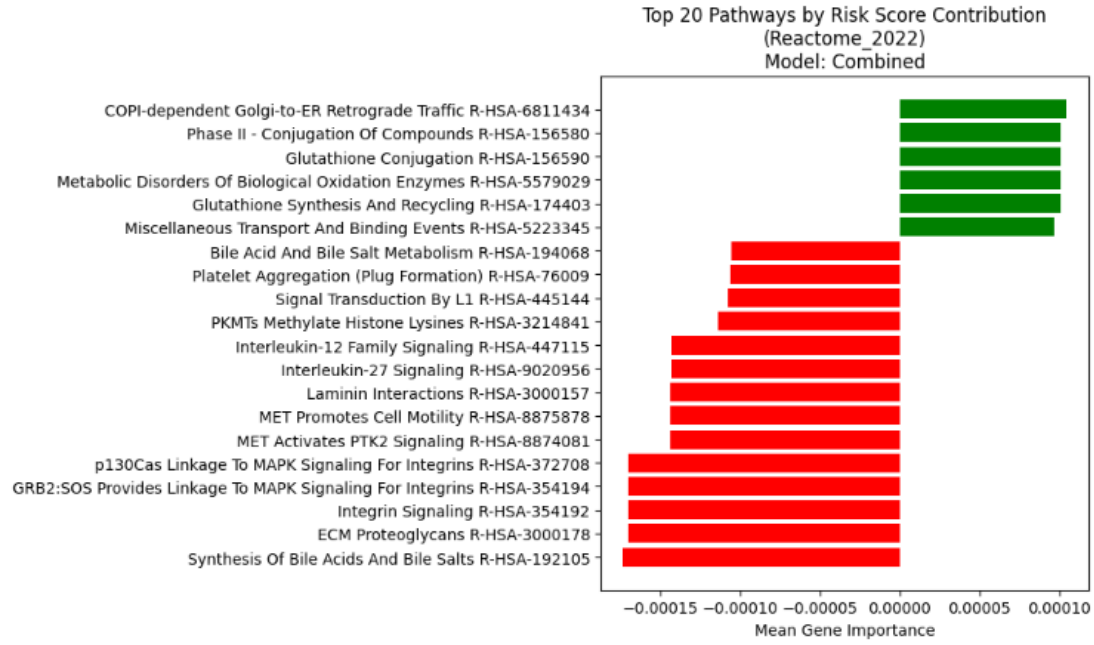


(b) Top pathways (RNA-seq based model)

Figure 5.4: SurvLIME-based gene and pathway importance for the RNA-seq based model in patient 1.



(a) Top genes (fused model)



(b) Top pathways (fused model)

Figure 5.5: SurvLIME-based gene and pathway importance for the fused model in patient 1.

Patient assigned a low risk score that was right-censored at year 5 of the observation period without experiencing any event (Patient 2) In both the RNA-seq based and fused models, highly protective pathways included *CD28 dependent PI3K/Akt signaling* (R-HSA-389357) and *CD28 co-stimulation* (R-HSA-389356), which are immune-related pathways involved in T-cell activation and survival [93]. Their association with reduced predicted risk may indicate a favorable immune surveillance state in this patient.

The *NIK to noncanonical NF-κB signaling* pathway (R-HSA-5676590) also ranked among the top features in both models. Although this pathway has been linked to inflammatory responses and tumor progression [94], its positive importance here suggests a protective role, possibly reflecting its complex regulatory functions.

Notably, the pathway *Signal transduction by L1* (R-HSA-445144) was the only negative pathway among the top-ranked features in the RNA-seq based model and one of two in the fused model. L1CAM, the protein encoded by this pathway, is known to promote glioma invasion and is associated with poor prognosis [95].

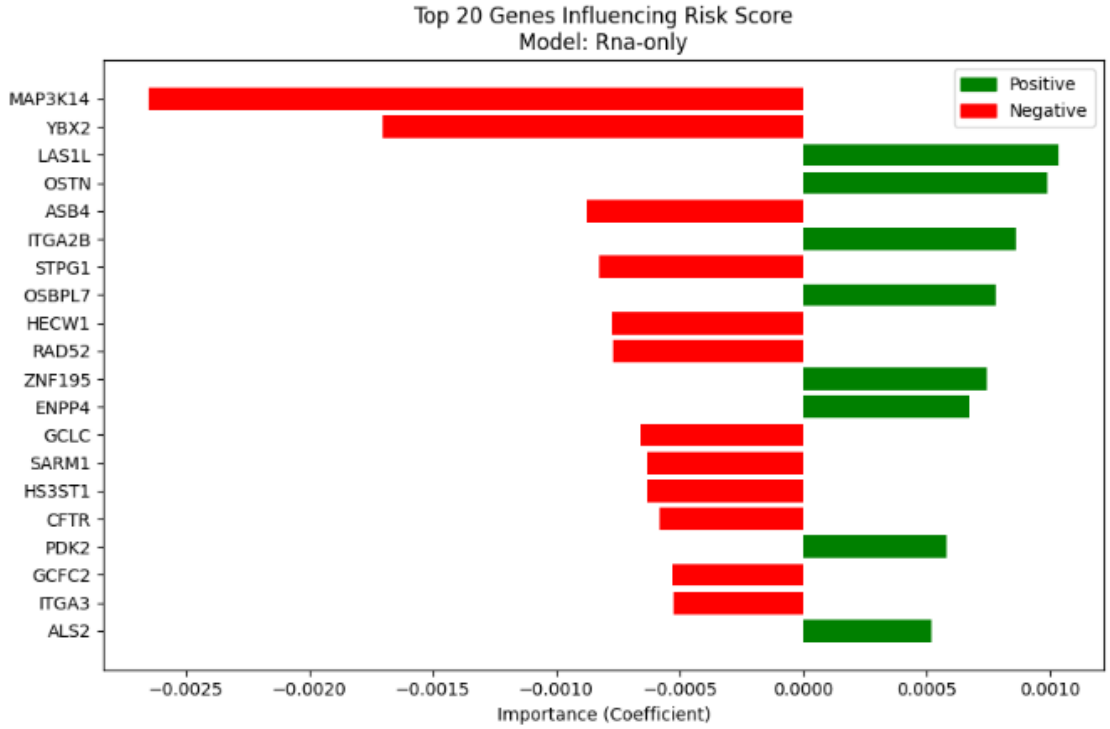


Figure 5.6: Top genes (RNA-seq based model) in patient 2.

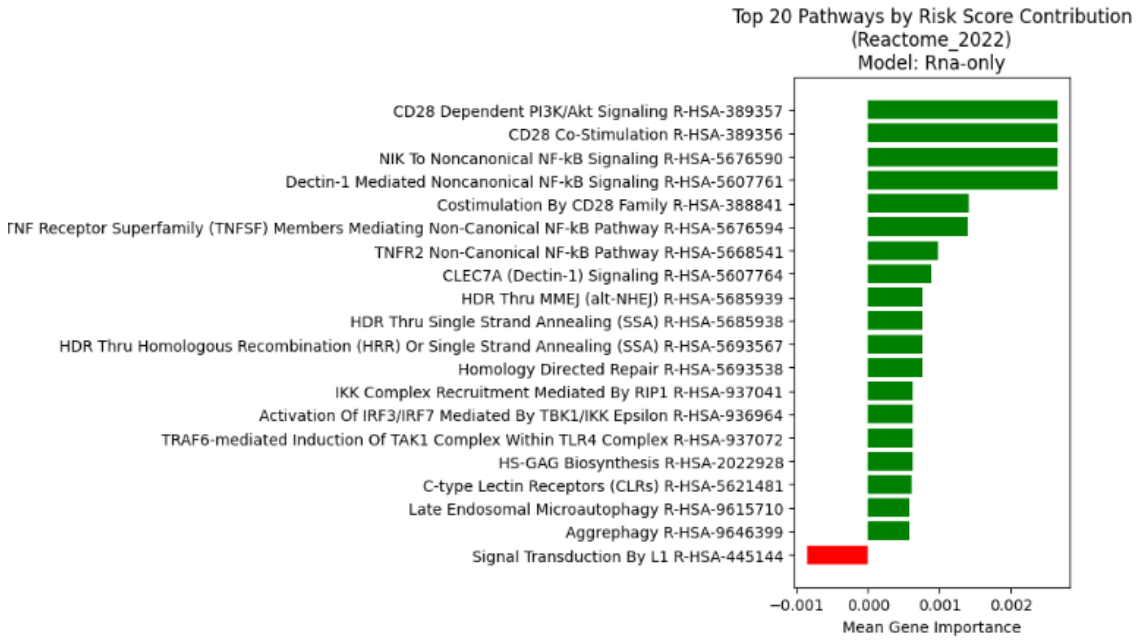


Figure 5.7: Top pathways (RNA-seq based model) in patient 2.

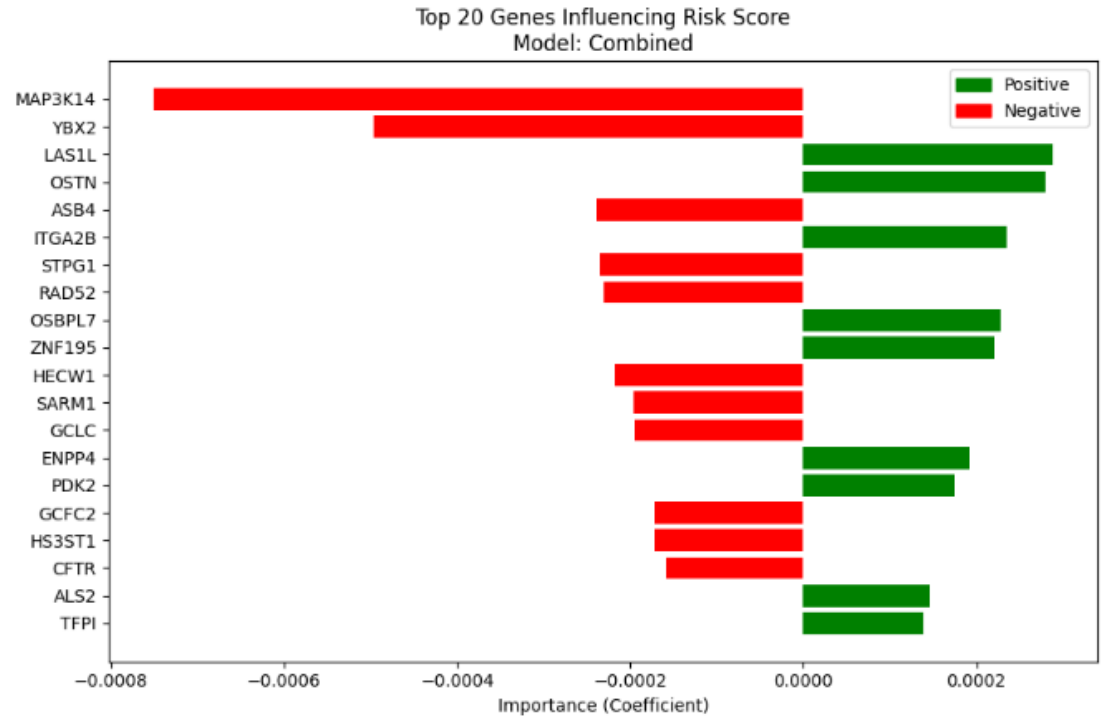


Figure 5.8: Top genes (fused model) in patient 2.

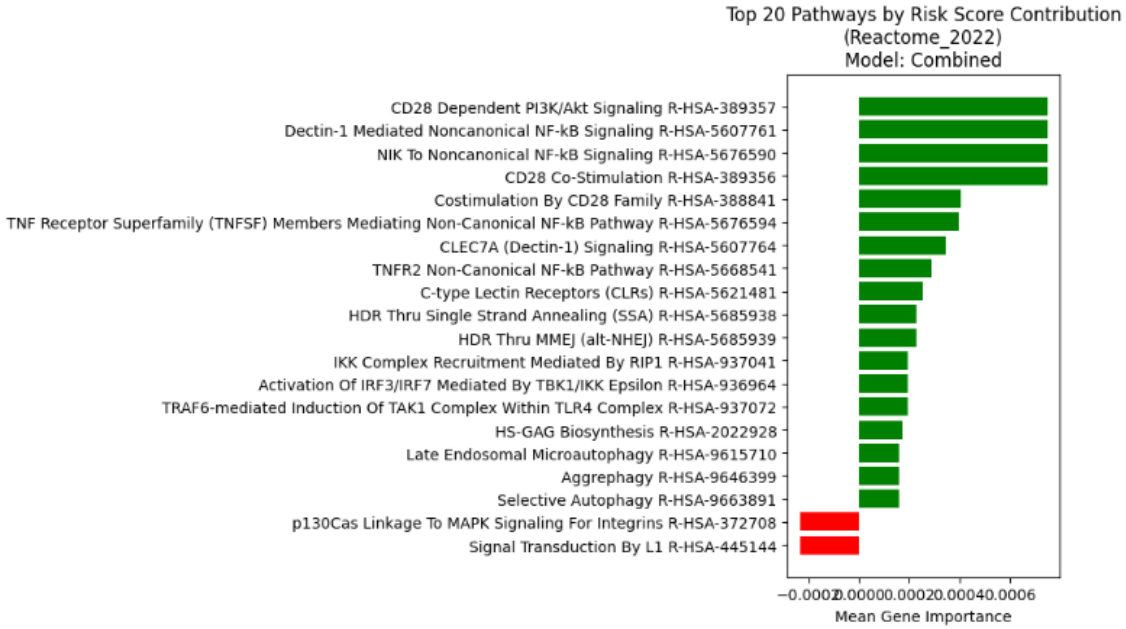


Figure 5.9: Top pathways (fused model) in patient 2.

Patient assigned a high risk score that experienced a death event in the first year of the observation period (Patient 3) This case corresponds to a high-risk patient who experienced an event and was correctly classified by both models. In the RNA-seq based model, the most hazardous pathways (i.e., with negative importance scores) were *Intrinsic Pathway for Apoptosis* (R-HSA-109606) and *TP53 regulates metabolic genes* (R-HSA-5628897). While both pathways are canonically tumor-suppressive, promoting programmed cell death and regulating metabolism, their negative scores here likely reflect reduced activity in this patient, consistent with hallmark glioblastoma features such as impaired apoptosis and disrupted p53 signaling [96].

The most protective pathway in the RNA-seq based model was *Reduction of cytosolic Ca^{++} levels* (R-HSA-418359). Although calcium signaling is highly context-dependent, reductions in intracellular calcium have been associated with decreased glioma cell proliferation and enhanced differentiation, potentially explaining the pathway's protective role [97].

In the fused model, the inclusion of histopathological features shifted importance toward pathways associated with tumor invasion and immune modulation. Among the most haz-

ardous features were *MET promotes cell motility* (R-HSA-8875878) and *Basigin interactions* (R-HSA-210991), both involved in enhancing glioma cell migration and invasiveness. MET signaling is a key driver of cell motility and therapeutic resistance [98], while basigin (CD147) facilitates extracellular matrix degradation and angiogenesis [99].

The most protective pathway in the fused model was *TNF receptor superfamily (TNFSF) members mediating non-canonical NF-κB signaling* (R-HSA-5676594). Although this pathway is often linked to pro-tumor inflammation, its protective interpretation here may indicate activation of anti-tumor immune responses in specific cellular contexts [94].

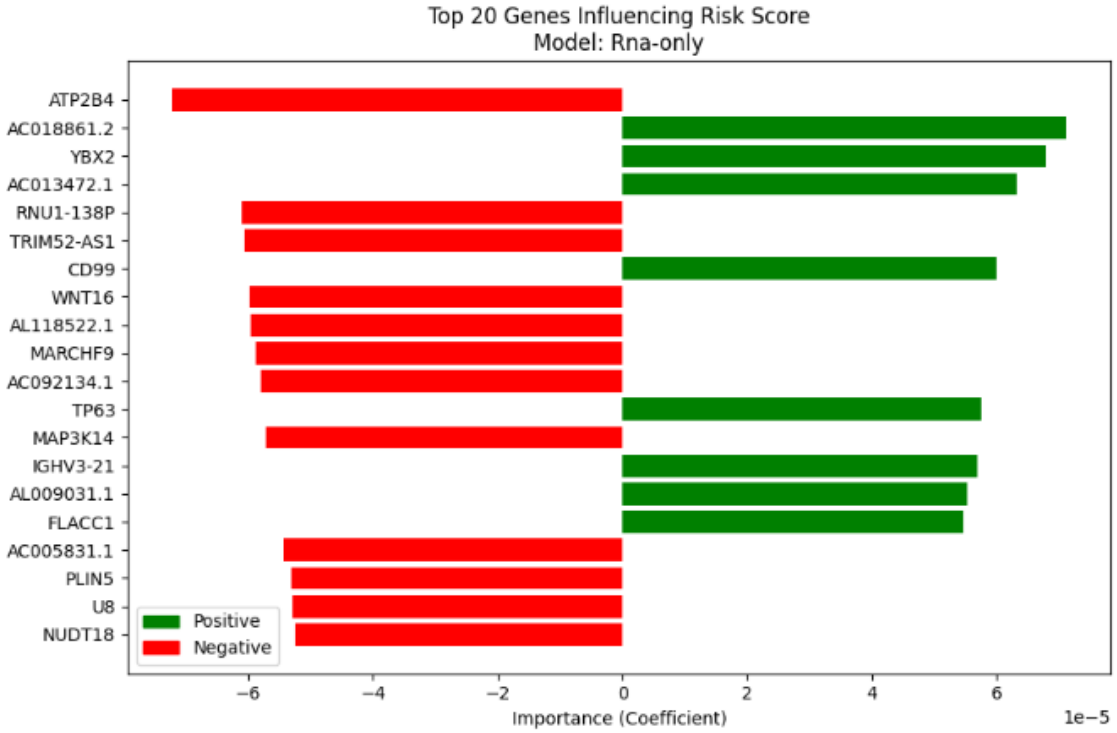


Figure 5.10: Top genes (RNA-seq based model) in patient 3.

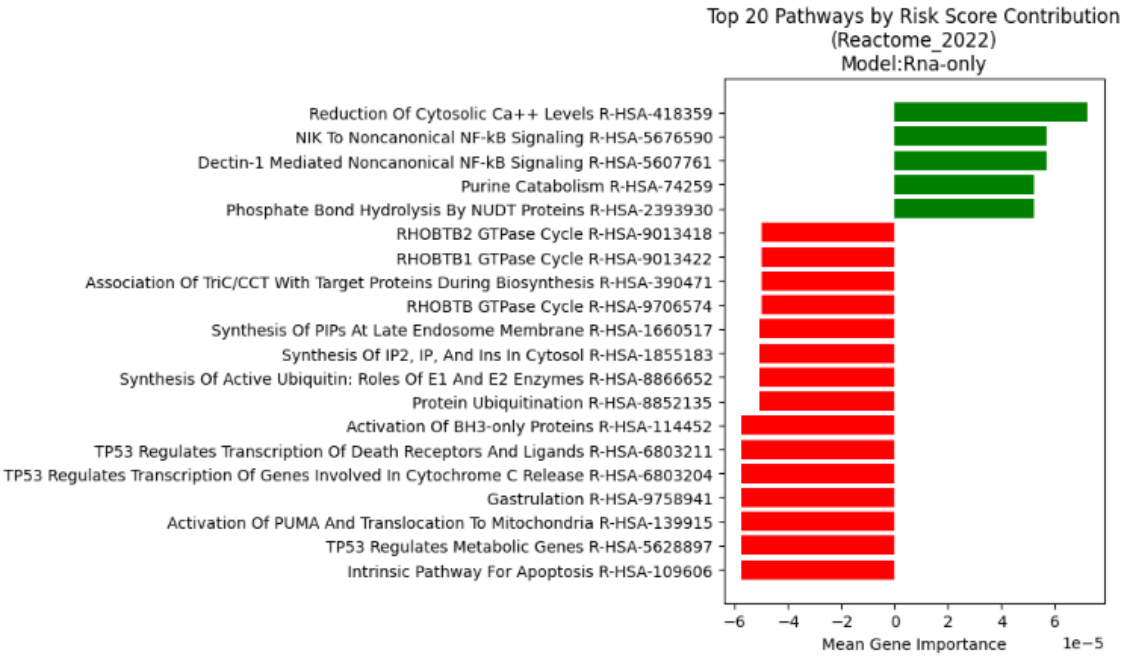


Figure 5.11: Top pathways (RNA-seq based model) in patient 3.

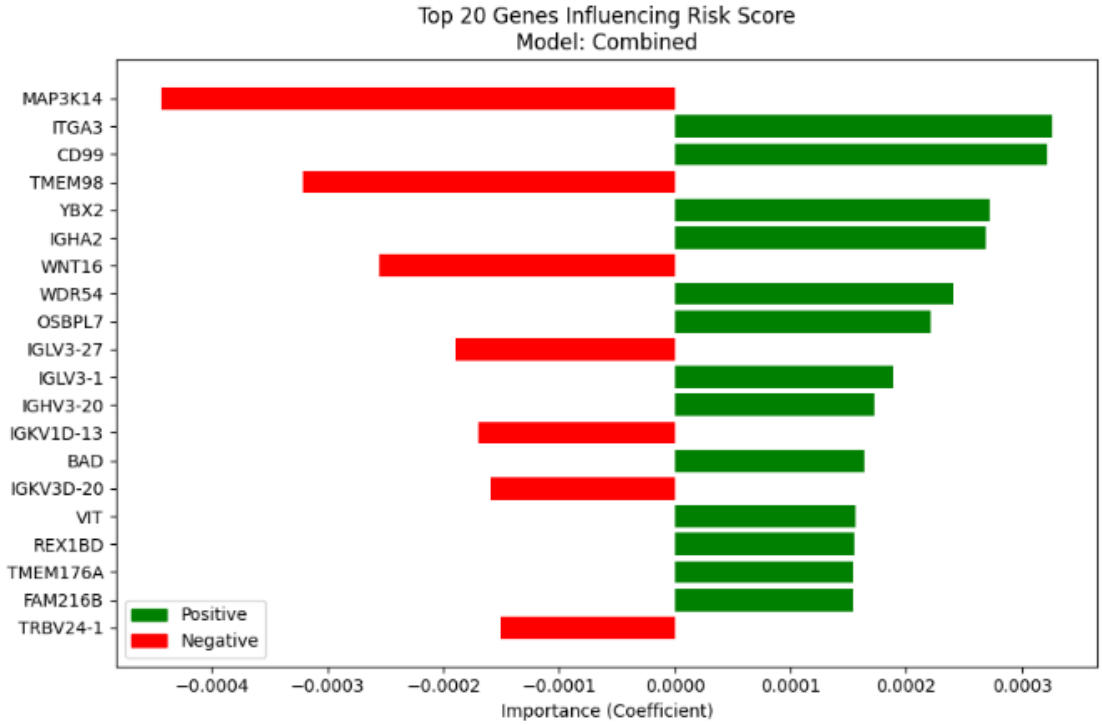


Figure 5.12: Top genes (fused model) in patient 3.

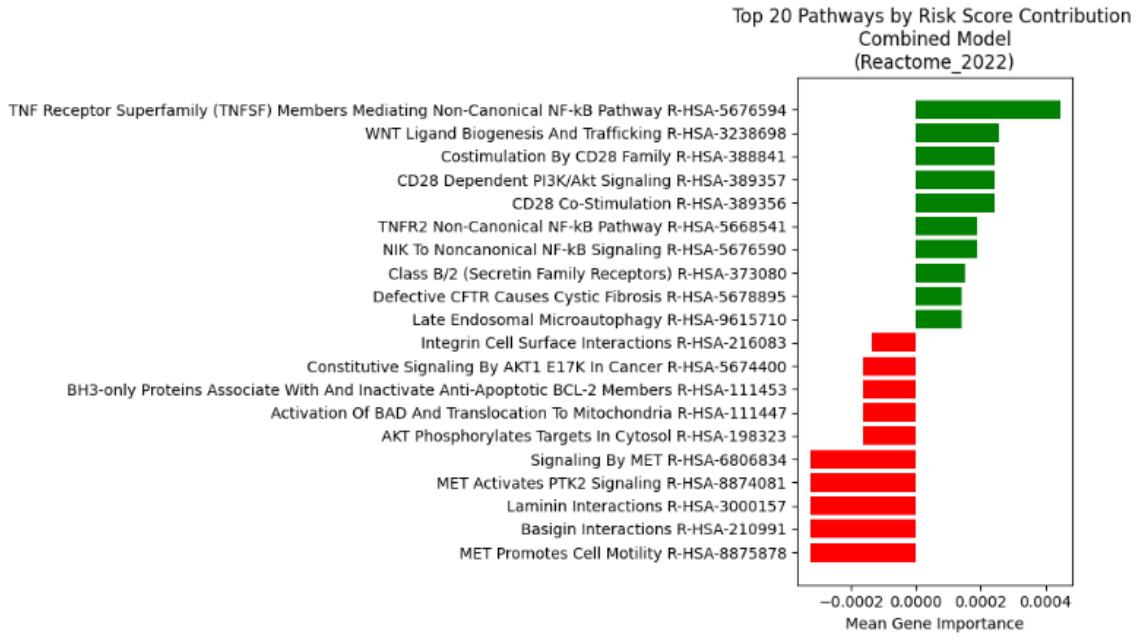


Figure 5.13: Top pathways (fused model) in patient 3.

Chapter 6

Conclusion

The primary objective of this thesis was to develop and evaluate a multimodal deep learning framework for survival prediction in glioblastoma patients by integrating histopathological whole-slide images and RNA-sequencing data. We constructed parallel pipelines for each data modality that included preprocessing, dimensionality reduction via autoencoders, and survival modeling using DeepSurv. The pipelines were subsequently fused at the feature level to assess whether combining morphological and molecular information could enhance prognostic accuracy.

Comprehensive experiments demonstrated that the gene expression model achieved high predictive performance across cross-validation and generalization to unseen test cases. Although the image-based model exhibited higher variance, it still captured meaningful prognostic patterns. Most notably, the fused model achieved the highest test set concordance index, suggesting that the integration of transcriptomic and histological features provides complementary information for survival prediction.

In addition to performance evaluation, we incorporated model interpretability via SurvLIME, enabling case-wise and pathway-level explanations for model decisions. This allowed us to investigate the biological relevance of important features and provided insight into the mechanisms associated with high- and low-risk profiles.

Overall, this work contributes to the growing field of multimodal learning in computational oncology by proposing a modular and interpretable survival prediction framework applicable to large-scale cancer datasets.

6.1 Future Work

While the framework demonstrated promising results, several directions remain open for future research. Incorporating clinical variables such as patient age, performance status, or molecular markers could further improve performance and clinical utility. More advanced fusion strategies may also enhance interaction modeling between modalities. Finally, applying the model to external datasets and validating the identified pathways biologically could strengthen its translational relevance.

In summary, this thesis presents a robust and interpretable deep learning pipeline for survival prediction in glioblastoma, offering a foundation for future developments in multimodal modeling for cancer prognosis.

List of Abbrevations

AI	Artificial Intelligence
AUC	Area Under the Curve
C-index	Concordance Index
CNN	Convolutional Neural Network
DL	Deep Learning
GSEA	Gene Set Enrichment Analysis
Gy	Gray
H&E	Hematoxylin and Eosin
ML	Machine Learning
MLP	Multilayer Perceptron
mRNA	Messenger Ribonucleic Acid
RNA	Ribonucleic Acid
RNA-seq	RNA Sequencing
TCGA	The Cancer Genome Atlas
TMZ	Temozolomide
TPM	Transcripts Per Million
TTFields	Tumor Treating Fields
UMAP	Uniform Manifold Approximation and Projection
WSI	Whole Slide Image
WSIs	Whole Slide Images

Bibliography

[1] Francisco A. C. Azevedo and et al. "Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain". In: *Journal of Comparative Neurology* 513.5 (2009), pp. 532–541. DOI: 10.1002/cne.21974.

[2] NCBI Bookshelf. *Major Structures and Functions of the Brain*. National Center for Biotechnology Information, 2024. URL: https://www.ncbi.nlm.nih.gov/books/NBK234157/.

[3] Clare Howarth. "The contribution of astrocytes to the regulation of cerebral blood flow". In: *Frontiers in Neuroscience* 8 (2014), p. 103. DOI: 10.3389/fnins.2014.00103.

[4] Praveen Ballabh, Armin Braun, and Maiken Nedergaard. "The blood-brain barrier: an overview". In: *Neurobiology of Disease* 16.1 (2004), pp. 1–13. DOI: 10.1016/j.nbd.2003.12.016.

[5] Hania Kadry, Badr Noorani, and Luca Cucullo. "A blood–brain barrier overview on structure, function, impairment, and biomarkers of integrity". In: *Fluids and Barriers of the CNS* 17 (2020), p. 69. DOI: 10.1186/s12987-020-00230-3.

[6] David N Louis et al. "The 2021 WHO classification of tumors of the central nervous system: a summary". In: *Neuro-Oncology* 23.8 (2021), pp. 1231–1251.

[7] Beata Kaminska et al. “Glioblastoma multiforme: a review of its epidemiology, pathogenesis, and molecular subtypes”. In: *Biomedical Reports* 6 (2017), pp. 3–11. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5563115/.

[8] Cancer Genome Atlas Research Network. “Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas”. In: *New England Journal of Medicine* 372.26 (2015), pp. 2481–2498.

[9] Quinn T Ostrom et al. “CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2014–2018”. In: *Neuro-oncology* 23 (2021), pp. iii1–iii105.

[10] Quinn T Ostrom et al. “CBTRUS Statistical Report: Primary brain and other central nervous system tumors diagnosed in the United States in 2011–2015”. In: *Neuro-Oncology* 20.suppl_4 (2018), pp. iv1–iv86. DOI: 10.1093/neuonc/noy131.

[11] David R Johnson and Brian P O’Neill. “Glioblastoma survival in the United States before and during the temozolomide era”. In: *Journal of Neuro-Oncology* 107.2 (2012), pp. 359–364. DOI: 10.1007/s11060-011-0749-4.

[12] Roger Stupp et al. “Maintenance therapy with tumor-treating fields plus temozolomide vs temozolomide alone for glioblastoma: a randomized clinical trial”. In: *JAMA* 318.23 (2017), pp. 2306–2316.

[13] Neil G Burnet et al. “Years of life lost (YLL) from cancer is an important measure of population burden—and should be considered when allocating research funds”. In: *British Journal of Cancer* 92.2 (2005), pp. 241–245. DOI: 10.1038/sj.bjc.6602321.

[14] Michael Weller et al. “EANO guideline for the diagnosis and treatment of anaplastic gliomas and glioblastoma”. In: *The Lancet Oncology* 15.9 (2014), e395–e403.

[15] Patrick Y Wen et al. “Response Assessment in Neuro-Oncology Clinical Trials”. In: *Journal of Clinical Oncology* 35.21 (2017), pp. 2439–2449. DOI: 10.1200/JCO.2017.72.7511. URL: https://doi.org/10.1200/JCO.2017.72.7511.

[16] Soonmee Cha. "Update on brain tumor imaging: from anatomy to physiology". In: *AJNR American Journal of Neuroradiology* 27.3 (2006), pp. 475–487. URL: https://www.ajnr.org/content/27/3/475.

[17] Daisuke Komura and Seiya Ishikawa. "Machine learning methods for histopathological image analysis". In: *Computational and Structural Biotechnology Journal* 16 (2018), pp. 34–42.

[18] Roger Stupp et al. "Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma". In: *New England Journal of Medicine* 352.10 (2005), pp. 987–996.

[19] Michel Lacroix et al. "A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival". In: *Journal of Neurosurgery* 95.2 (2001), pp. 190–198.

[20] Michael Weller et al. "Radiation therapy for glioblastoma: American Society of Clinical Oncology clinical practice guideline". In: *Journal of Clinical Oncology* 32.31 (2014), pp. 4127–4136. DOI: 10.1200/JCO.2014.55.0135.

[21] Monika E Hegi et al. "MGMT gene silencing and benefit from temozolomide in glioblastoma". In: *New England Journal of Medicine* 352.10 (2005), pp. 997–1003.

[22] Anoop P Patel et al. "Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma". In: *Science* 344.6190 (2014), pp. 1396–1401.

[23] Frank Eckerdt and Leonidas C. Platanias. "Emerging Role of Glioma Stem Cells in Mechanisms of Therapy Resistance". In: *Cancers* 15.13 (2023), p. 3458. DOI: 10.3390/cancers15133458.

[24] Hai Yan et al. "IDH1 and IDH2 mutations in gliomas". In: *New England Journal of Medicine* 360.8 (2009), pp. 765–773.

[25] John McCarthy et al. "A proposal for the Dartmouth summer research project on artificial intelligence". In: *AI Magazine*. Vol. 27. 4. 2006, pp. 12–14.

[26] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[27] Norman P Jouppi et al. "In-datacenter performance analysis of a tensor processing unit". In: *Proceedings of the 44th Annual International Symposium on Computer Architecture* (2017), pp. 1–12.

[28] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *Nature* 521.7553 (2015), pp. 436–444.

[29] Warren S McCulloch and Walter Pitts. "A logical calculus of the ideas immanent in nervous activity". In: *The bulletin of mathematical biophysics* 5.4 (1943), pp. 115–133.

[30] Yann LeCun et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

[31] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.

[32] Li Deng and Dong Yu. "Deep learning: methods and applications". In: *Foundations and Trends in Signal Processing* 7.3–4 (2014), pp. 197–387. DOI: 10.1561/2000000039.

[33] Tom Young et al. "Recent trends in deep learning based natural language processing". In: *IEEE Computational Intelligence Magazine* 13.3 (2018), pp. 55–75. DOI: 10.1109/MCI.2018.2840738.

[34] Andre Esteva et al. "A guide to deep learning in healthcare". In: *Nature Medicine* 25.1 (2019), pp. 24–29. DOI: 10.1038/s41591-018-0316-z.

[35] Xavier Glorot and Yoshua Bengio. "Understanding the difficulty of training deep feedforward neural networks". In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*. Vol. 9. JMLR Workshop and Conference Proceedings, 2010, pp. 249–256.

[36] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[37] Nitish Srivastava, Geoffrey Hinton, et al. "Dropout: A simple way to prevent neural networks from overfitting". In: *Journal of machine learning research* 15.1 (2014), pp. 1929–1958.

[38] Sheraz Aslam, Herodotos Herodotou, and Syed Muhammad Mohsin. "A typical architecture of ANN and DNN". In: *A Survey on Deep Learning Methods for Power Load and Renewable Energy Forecasting in Smart Microgrids* (2021). Figure from ResearchGate.

[39] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. "Deep sparse rectifier neural networks". In: (2011), pp. 315–323.

[40] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. "Multilayer feedforward networks are universal approximators". In: *Neural networks* 2.5 (1989), pp. 359–366.

[41] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778.

[42] Christian Szegedy et al. "Going deeper with convolutions". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 1–9.

[43] Jia Deng et al. "ImageNet: A large-scale hierarchical image database". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2009), pp. 248–255.

[44] Keiran Wang. *The Annotated ResNet-50*. `https://towardsdatascience.com/the-annotated-resnet-50-a7c402a25c64`. Accessed: 2025-06-29. 2021.

[45] Mingxing Tan and Quoc V Le. "EfficientNet: Rethinking model scaling for convolutional neural networks". In: *International Conference on Machine Learning (ICML)*. PMLR. 2019, pp. 6105–6114.

[46] Tashin Ahmed and Noor Hossain Nuri Sabab. "Classification and Understanding of Cloud Structures via Satellite Images with EfficientUNet". In: *SN Computer Science* 3.1 (2022), p. 99. DOI: 10.1007/s42979-021-00981-2.

[47] Gao Huang et al. "Densely connected convolutional networks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 4700–4708.

[48] Sameer Kalra et al. "Pan-cancer image-based detection of clinically actionable genetic alterations". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2020, pp. 954–955.

[49] Kyunghyun Cho et al. "Learning phrase representations using RNN encoder-decoder for statistical machine translation". In: *arXiv preprint arXiv:1406.1078* (2014).

[50] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. "Sequence to sequence learning with neural networks". In: *Advances in Neural Information Processing Systems* 27 (2014).

[51] Diederik P Kingma and Max Welling. "Auto-encoding variational Bayes". In: *arXiv preprint arXiv:1312.6114* (2013).

[52] Zhong Wang, Mark Gerstein, and Michael Snyder. "RNA-Seq: a revolutionary tool for transcriptomics". In: *Nature Reviews Genetics* 10.1 (2009), pp. 57–63. DOI: 10.1038/nrg2484.

[53] Dhrithi Deshpande et al. "RNA-seq data science: From raw data to effective interpretation". In: *arXiv preprint* arXiv:2010.02391 (2020).

[54] Bo Li and Colin N. Dewey. "RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome". In: *BMC Bioinformatics* 12 (2011), p. 323. DOI: 10.1186/1471-2105-12-323.

[55] G. Paul Wagner, Laura J. Kin, and Michael Lynch. "Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols". In: *Bioinformatics* 31.18 (2015), pp. 2905–2907. DOI: 10.1093/bioinformatics/btv279.

[56] Ana Conesa, Pedro Madrigal, and et al. "A survey of best practices for RNA-seq data analysis". In: *Genome Biology* 17 (2016), p. 13. DOI: 10.1186/s13059-016-0881-8.

[57] David G. Kleinbaum and Mitchel Klein. *Survival Analysis: A Self-Learning Text*. 3rd. Springer, 2012. DOI: 10.1007/978-1-4419-6646-9.

[58] John P Klein and Melvin L Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd. New York, NY: Springer, 2003.

[59] Edward L. Kaplan and Paul Meier. "Nonparametric estimation from incomplete observations". In: *Journal of the American Statistical Association* 53.282 (1958), pp. 457–481. DOI: 10.1080/01621459.1958.10501452.

[60] Nathan Mantel. "Evaluation of survival data and two new rank order statistics arising in its consideration". In: *Cancer Chemotherapy Reports* 50.3 (1966), pp. 163–170.

[61] David R. Cox. "Regression models and life-tables". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 34.2 (1972), pp. 187–202.

[62] Hemant Ishwaran et al. "Random survival forests". In: *The Annals of Applied Statistics* 2.3 (2008), pp. 841–860. DOI: 10.1214/08-AOAS169.

[63] Jared L Katzman et al. "DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network". In: *BMC Medical Research Methodology* 18.1 (2018), p. 24.

[64] Takuya Akiba et al. "Optuna: A next-generation hyperparameter optimization framework". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2019, pp. 2623–2631.

[65] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. "Multimodal machine learning: A survey and taxonomy". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41.2 (2018), pp. 423–443.

[66] Dinesh Ramachandram and Graham W Taylor. “Deep multimodal learning: A survey on recent advances and trends”. In: *IEEE Signal Processing Magazine* 34.6 (2017), pp. 96–108.

[67] Valerio Guarrasi et al. “A Systematic Review of Intermediate Fusion in Multimodal Deep Learning for Biomedical Applications”. In: *arXiv preprint arXiv:2408.02686* (2024).

[68] Parsa Mobadersany et al. “Predicting cancer outcomes from histology and genomics using convolutional networks”. In: *Proceedings of the National Academy of Sciences* 115.13 (2018), E2970–E2979.

[69] Anika Cheerla and Olivier Gevaert. “Deep learning with multimodal representation for pancancer prognosis prediction”. In: *Bioinformatics* 35.14 (2019), pp. i446–i454.

[70] Andreas Holzinger et al. “What do we need to build explainable AI systems for the medical domain?” In: *arXiv preprint arXiv:1712.09923* (2017).

[71] Diogo Carvalho, Eduardo Pereira, and Jaime Cardoso. “Machine learning interpretability: A survey on methods and metrics”. In: *Electronics* 8.8 (2019), p. 832.

[72] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. ““Why should I trust you?” Explaining the predictions of any classifier”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, pp. 1135–1144.

[73] Scott M Lundberg et al. “From local explanations to global understanding with explainable AI for trees”. In: *Nature Machine Intelligence* 2.1 (2020), pp. 56–67.

[74] Mikhail Kovalev et al. “Interpreting survival analysis models with SurvLIME”. In: *Machine Learning* 110.8 (2021), pp. 2125–2150.

[75] Tom Fawcett. “An introduction to ROC analysis”. In: *Pattern Recognition Letters* 27.8 (2006), pp. 861–874.

[76] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd. New York: Springer Science & Business Media, 2009.

[77] Frank E Harrell et al. "Evaluating the yield of medical tests". In: *JAMA* 247.18 (1982), pp. 2543–2546.

[78] Erika Graf et al. "Assessment and comparison of prognostic classification schemes for survival data". In: *Statistics in Medicine* 18.17-18 (1999), pp. 2529–2545.

[79] Patrick J Heagerty, Thomas Lumley, and Margaret S Pepe. "Time-dependent ROC curves for censored survival data and a diagnostic marker". In: *Biometrics* 56.2 (2000), pp. 337–344.

[80] Hans C. van Houwelingen. "Validation, calibration, revision and combination of prognostic survival models". In: *Statistics in Medicine* 19.24 (2000), pp. 3401–3415.

[81] John N Weinstein et al. "The Cancer Genome Atlas Pan-Cancer analysis project". In: *Nature Genetics* 45.10 (2013), pp. 1113–1120.

[82] Richard J. Chen et al. "Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis". In: *IEEE Transactions on Medical Imaging* 41.3 (2022), pp. 757–770. DOI: 10.1109/TMI.2021.3118433.

[83] Ziyue Fan et al. "Deep Learning-based Survival Prediction for Brain Tumors using Convolutional MIL and Aggregation Modules". In: *Frontiers in Oncology* 13 (2023), p. 1182491. DOI: 10.3389/fonc.2023.1182491.

[84] Sandra Steyaert et al. "Multimodal data fusion of adult and pediatric brain tumors with deep learning". In: *Scientific Reports* 12 (2022), p. 17399. DOI: 10.1038/s41598-022-21699-3.

[85] Jie Hao et al. "PAGE-Net: Interpretable and Integrative Deep Learning for Survival Analysis Using Histopathological Images and Genomic Data". In: *NPJ Digital Medicine* 3.1 (2020), pp. 1–11. DOI: 10.1038/s41746-020-0225-1.

[86] Richard J. Chen et al. "Pan-cancer integrative histology-genomic analysis via multimodal deep learning". In: *Cancer Cell* 40.8 (2022), 865–878.e7. DOI: 10.1016/j.ccell.2022.07.002.

[87] Xinyu Zhang et al. "CATFusion: Cross-Attentional Multimodal Fusion for Pan-Cancer Survival Prediction". In: *IEEE Journal of Biomedical and Health Informatics* (2024). In Press. DOI: 10.1109/JBHI.2024.3365432.

[88] Bijay Jassal et al. "The Reactome pathway knowledgebase". In: *Nucleic Acids Research* 48.D1 (2020), pp. D498–D503. DOI: 10.1093/nar/gkz1031.

[89] AS Dhillon et al. "MAP kinase signalling pathways in cancer". In: *Oncogene* 26 (2007), pp. 3279–3290. DOI: 10.1038/sj.onc.1210421.

[90] Roger Stupp, Warren P Mason, Martin J van den Bent, et al. "Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma". In: *New England Journal of Medicine* 352.10 (2005), pp. 987–996. DOI: 10.1056/NEJMoa043330.

[91] Patrizia De Filippi and Paola Defilippi. "p130Cas: a versatile scaffold in signaling networks". In: *Trends in Cell Biology* 16.5 (2006), pp. 257–263. DOI: 10.1016/j.tcb.2006.03.003.

[92] Yuan Zhen et al. "COPI-mediated retrograde trafficking and its role in cell signaling". In: *BioMed Research International* 2017 (2017), pp. 1–9. DOI: 10.1155/2017/9736071.

[93] CE Rudd. "CD28: a co-stimulatory receptor for T-cell activation". In: *Nature Reviews Immunology* 9 (2009), pp. 329–336. DOI: 10.1038/nri2546.

[94] Shao-Cong Sun. "Non-canonical NF-κB signaling pathway". In: *Cell Research* 21.1 (2011), pp. 71–85. DOI: 10.1038/cr.2010.177.

[95] Elena Rodriguez et al. "L1 cell adhesion molecule in cancer: a systematic review on its diagnostic and therapeutic potential". In: *International Journal of Molecular Sciences* 18.10 (2017), p. xxx.

[96] J Liu et al. "The role of p53 in metabolism and cancer". In: *Cold Spring Harbor Perspectives in Medicine* 9.6 (2019), a031955. DOI: 10.1101/cshperspect.a031955.

[97] Emmanuel Terrié et al. "Store-operated calcium channels control proliferation and self-renewal of cancer stem cells from glioblastoma". In: *Cell Reports* 19.8 (2017), pp. 1568–1579. DOI: 10.1016/j.celrep.2017.04.052.

[98] Simon Bougnaud et al. "HGF/MET signaling in malignant brain tumors: molecular mechanisms and therapeutic strategies". In: *International Journal of Molecular Sciences* 21.20 (2020), p. 7546. DOI: 10.3390/ijms21207546.

[99] Ibtissam Marchiq, Jean Albrengues, Cristine Pasquali, et al. "Knockout of basigin (CD147) in glioma cells does not reduce MMP induction but impairs angiogenesis via VEGF downregulation". In: *Oncotarget* 6.24 (2015), pp. 24636–24648. DOI: 10.18632/oncotarget.4323.