

Supervised Machine Learning Predictive Modeling of Survival Times of Malignant Glioma Patients

Erasmus Tetteh-Bator^{1*}, Lohuwa Mamudu², Seth K. Laryea³, Chris P. Tsokos⁴

¹Department of Mathematics & Statistical Sciences, Jackson State University, Jackson, MS, USA

²Department of Public Health, California State University, Fullerton, CA, USA

³Department of Computer Science, University of Ghana, Legon, Accra, Ghana

⁴Department of Mathematics & Statistics, University of South Florida, Tampa, FL, USA

Email: *erasmus.tetteh-bator@jsums.edu

How to cite this paper: Tetteh-Bator, E., Mamudu, L., Laryea, S.K. and Tsokos, C.P. (2026) Supervised Machine Learning Predictive Modeling of Survival Times of Malignant Glioma Patients. *Journal of Data Analysis and Information Processing*, **14**, 119-150.

<https://doi.org/10.4236/jdaip.2026.141007>

Received: December 22, 2025

Accepted: February 2, 2026

Published: February 5, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Gliomas are the most cancerous tumors arising from the brain, accounting for approximately 78% of all primary brain tumors. The progression of malignant gliomas significantly reduces both survival duration and quality of life, with long-term survival achieved by fewer than 5% of patients, with a significant portion facing a grim outlook within 1.5 years, despite standard treatment. To improve survival prediction and support clinical decision-making, this study analyzes real-world data from the SEER (Surveillance, Epidemiology, and End Results) database of the National Cancer Institute (NCI). In this study, we identified a well-defined probability distribution that characterizes the survival times of patients with Malignant Glioma (3-Parameter Generalized Pareto Probability Distribution) and derived its cumulative distribution function (CDF) and survival function, providing estimates of the probability of patient survival. This information proves invaluable to healthcare professionals and the families of patients, addressing questions raised about patients, such as the patient's survival probability given survival beyond a specified time, or the patient's risk or intensity of death (hazard) at time t after surviving to that point. This study developed an Artificial Neural Network (ANN)-based survival analysis model, with a binary classifier predicting patient survival less than 12 months or beyond, and a multiclass model predicting survival within the first 12 months, between 12 and 24 months, or beyond 24 months. Model performance is evaluated using multiple metrics and compared with other machine learning algorithm approaches.

Keywords

Gliomas, Supervised Machine Learning, Loss Function, Generalized Pareto Distribution, Artificial Neural Network

1. Introduction

Brain tumors stem from damaged genes that regulate cell division and repair processes. When these genes malfunction, cells can grow uncontrollably, forming tumors. While the immune system usually defends against abnormal cell growth, tumors can release substances that evade detection, allowing them to thrive despite internal and external barriers to their growth. The two categories of brain tumors are termed primary and metastatic. Primary brain tumors originate from brain tissues or their immediate surroundings. They are further classified into gliomas (composed of glial cells) and non-gliomas (developed on or in brain structures like nerves, blood vessels, and glands), with the potential to be either benign or malignant.

Glioma is a type of central nervous system (CNS) cancer affecting the glial cells. Gliomas are the most cancerous tumors arising from the brain with approximately 78% of all primary brain tumors being gliomas. According to the American Cancer Society, in 2023, in the United States, there will be an estimated 1,958,310 new cancer cases and 609,820 cancer deaths. That's approximately 5,365 new cases and 1,671 deaths every day, resulting in 4 cases and 1 death every minute [1]. In 2020, an estimated 180,047 people were living with brain and other nervous system cancer in the United States. Statistically, approximately 24,000 patients are diagnosed with new cases of primary malignant brain tumors diagnosed in the United States each year [2].

Neurons are the fundamental units of the nervous system, responsible for processing and transmitting information. Astrocytes, oligodendrocytes, and Ependymal cells play specific supportive roles, creating an environment conducive to neuron function. These three supportive cell types, collectively known as glial cells, can give rise to gliomas: brain tumors originating from either astrocytes, oligodendrocytes, or ependymal cells. The layout of the glial cells is shown in **Figure 1** below [3].

Unlike many other cancers that are staged, gliomas are classified by grades, indicating their level of aggressiveness as observed under a microscope and in terms of their molecular characteristics. This is because they do not metastasize outside the central nervous system. The World Health Organization (WHO) 2021 classification system categorizes gliomas into four grades, each representing a significant difference in aggressiveness. Grades range from I (least aggressive) to IV (most aggressive), based on factors like cell reproduction rate and the likelihood of tissue infiltration according to WHO guidelines [4] [5]. Diagnosing a brain tumor is a complex process that begins with a detailed medical history and physical

examination. Advanced imaging techniques like neuro exams, CT scans, MRI, fMRI, and PET scans play a crucial role. The MRI scan shows the tumor and inflammation around the tumor which pushes the brain into different directions against the bone structure, and since the brain can't extend, as a result patient experiences severe headaches. Other clinical symptoms will depend on the location of the tumor. It is possible that it could affect the patient's vision, smell, hearing, speech, motor control, among others, as shown in **Figure 2** below.

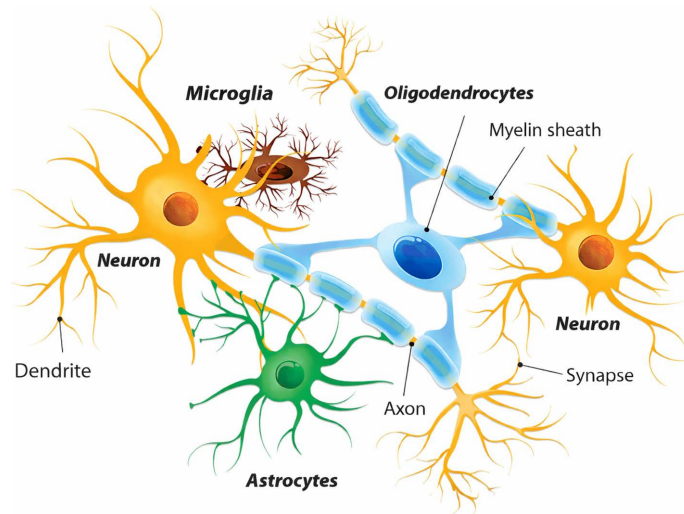


Figure 1. Simplified view of a neuron and glial cells.

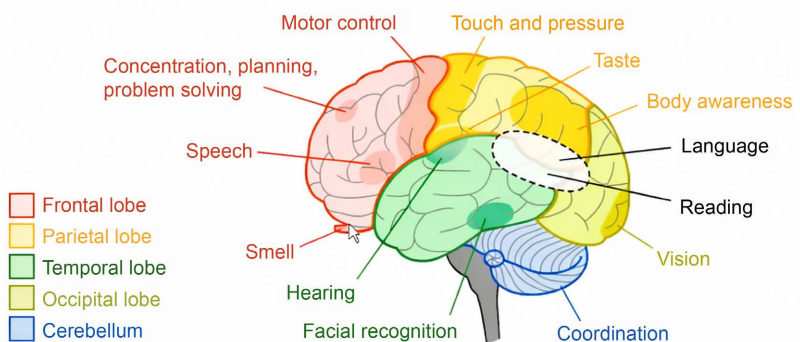
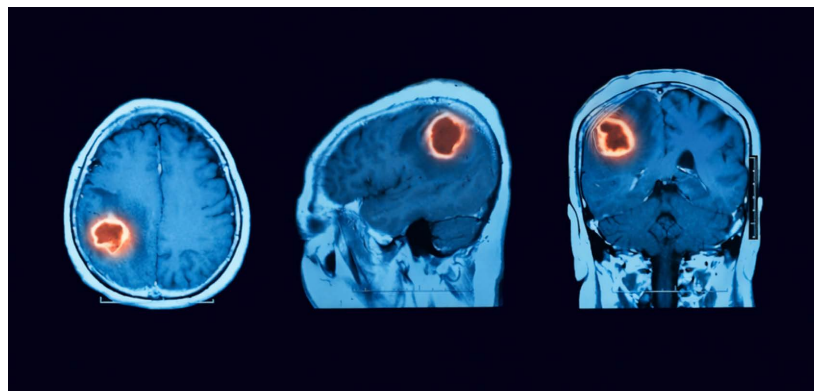


Figure 2. MRI scan showing tumor, locations and affected clinical symptoms.

The American Cancer Society estimated that there would be approximately 24,530 new cases of primary malignant brain tumors (including gliomas) diagnosed in the United States in 2021. However, this number of cases represents all primary brain tumors and includes various types and grades [6]. In 2022, a report by the Surveillance, Epidemiology, and End Results (SEER) Cancer Institute stated that Brain and other nervous system cancer represents 1.3% of all new cancer cases in the U.S. and ranked among the top 16 cancer diseases. A further projection by SEER indicates that in 2023, there will be 24,810 estimated new cases of brain and other nervous system cancer and an estimated 18,990 people will die of this disease. There is a sufficient increase compared with the 22,850 and 23,890 estimated new cases reported in 2015 and 2020, respectively. These figures are progressively increasing yearly and cannot be overlooked [6].

The latest advancement in tumor grading is the fifth edition of the World Health Organization (WHO) Classification of Tumors of the Central Nervous System (CNS)—WHO CNS5, 2021, shown in **Figure 3** below. This edition builds on the 2016 update, providing a more refined classification based on biological and molecular characteristics. It introduces new tumor types and subtypes, particularly in pediatric cases. These revisions offer clinicians a deeper insight into prognosis and tailored therapy for patients with specific CNS tumors. Additionally, they enhance the potential for more precise patient groups in clinical trials, facilitating the assessment of innovative treatments. The gliomas brain tumor classification by type and grade and molecular marker or profiling [7].

Grade \ Type	Grade I	Grade II	Grade III	Grade IV
	Low-grade		High-grade	
Astrocyte cell/tumor	Pilocytic Astrocytoma	Low-grade Astrocytoma	Anaplastic Astrocytoma	Astrocytoma & Glioblastoma
Oligodendrocyte cell/tumor		Low-grade Oligodendro glioma	Anaplastic Oligodendro glioma	
Molecular Marker or Profiling		1p/19q Co-deleted IDH-MUTANT		IDH-Mutant & IDH-Wildtype

Figure 3. Schematic diagram showing the gliomas brain tumor classification by type and grade and molecular profiling.

The concept of learning from data has been around for a long time and is the foundation of modern artificial intelligence (AI). Artificial neural networks (ANNs) combine probabilistic models from classical statistics to solve various classification and regression issues. The popularity of ANNs in health-related problems has increased rapidly over the past decades due to their capability to identify complex patterns in patient data. To addressing the most common questions raised by

healthcare professionals, patients and family members such as: What is it? How long does the patient have to live? and What are the chances? We aim to build a supervised machine learning algorithm, Artificial Neural Network, (ANN) to predict the survival time of malignant glioma (brain tumor) patients. The ANN-based survival analysis models are to predict the survival time of glioma patients, with a binary classifier predicting whether a patient will survive less than 12 months or beyond, and a multiclass model predicting survival within the first 12 months, between 12 and 24 months, or beyond 24 months. We contend that identifying a well-defined probability distribution that characterizes the survival times of patients with malignant glioma is a significant step toward accurately predicting survival probabilities. Using this, we can determine a patient's survival probability at or beyond a specified time, as well as their risk or intensity of death at a given time after surviving to that point. It is widely recognized that parametric analysis is more powerful in making decisions than its non-parametric analysis [8]. Additionally, other authors have noted that assuming an exponential distribution does not always works well for studying the survival of cancer-related cases [9] [10]. Additionally, we can assess the importance of various risk factors in predicting survival times. These insights offer crucial guidance for healthcare decision-making and provide answers to support questions raised by patients, and their families.

2. Methodology and Materials

2.1. Data Description

The data used in this research are from the SEER (Surveillance, Epidemiology, and End Results) database from the NCI (National Cancer Institute). The Surveillance, Epidemiology, and End Results (SEER) data from the National Cancer Institute is a population-based cancer registry of the US population [11]. The data obtained from the SEER database constituted a total of 59,123 patients with malignant, primary tumors of the brain and the central nervous system, diagnosed between 2010 to 2020 from 17 different registries across the United States. The data was preprocessed and cleaned to eliminate redundancies (unknown, blanks, not documented) and missing data in at least one of the risk factors. The resulting data set had 21,430 patients with malignant gliomas (brain tumors). The data for analysis comprised the survival time (the time up to which the patient survived the disease from diagnosis to the nearest month) and survival status (dead or alive) of the patients, along with 14 risk factors of the patient shown in **Figure 4** below.

2.2. Artificial Neural Network (ANN)

An Artificial Neural Network (ANN) is a computational model inspired by biological neural networks in the human brain. ANNs are designed to recognize patterns, learn from data, and make decisions or predictions. They consist of layers of interconnected nodes (neurons) that work together to solve complex problems [12].

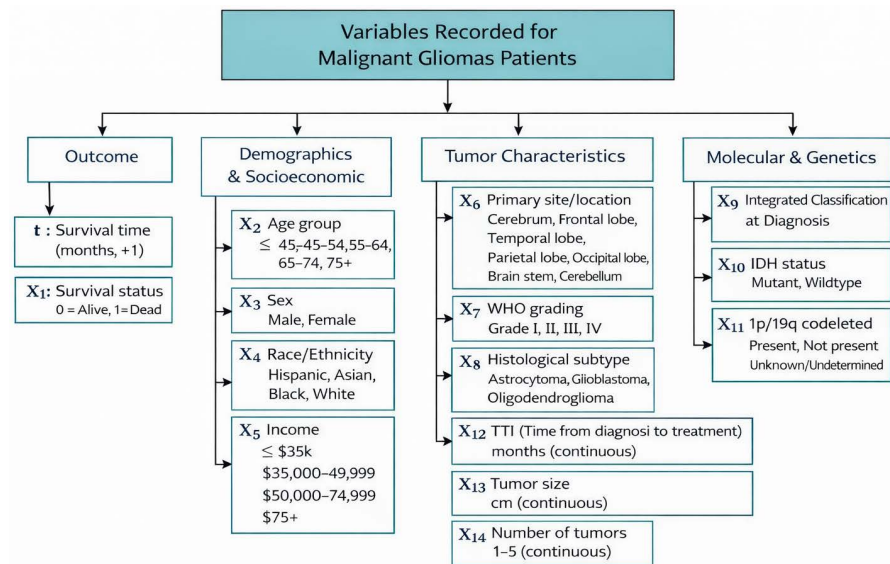


Figure 4. Variables recorded for malignant gliomas (primary brain tumor) patients.

Artificial Neural Networks have become extremely popular across a wide range of fields due to their ability to learn from data, recognize patterns, and make accurate predictions. Their impact is evident in everyday applications, from healthcare to entertainment, including image and speech recognition, natural language processing, autonomous driving, game playing, and financial forecasting. ANNs have revolutionized many areas by providing powerful tools for solving complex problems that are difficult to address with traditional algorithms [13].

ANN's adaptability and capacity to learn from data through training enable them to extract relevant features from raw data. ANNs improve their performance as they are exposed to more data, adjusting their internal parameters (weights and biases) to minimize errors. This ability makes them efficient for handling large datasets and unseen data, rendering them excellent at generalizing [14]. ANNs are versatile and can be applied to a wide range of tasks, including classification, regression, clustering, and even generative tasks. Moreover, they are largely free from the rigid statistical assumptions required by traditional models, such as linearity, normality, and homoscedasticity, making them more adaptable to diverse and complex data patterns [15].

Artificial Neural Networks (ANNs) have been applied to survival analysis for several decades, evolving from early exploratory studies to sophisticated deep learning models. One of the pioneering works in this area was by Faraggi and Simon in 1995, who demonstrated that ANNs could effectively model relationships in survival data, introducing methods for handling censored data [16]. Building on this, Biganzoli *et al.* in 1998 proposed a partial logistic regression approach using feedforward neural networks, improving the modeling of survival probabilities in censored data [17]. Further advancements were made by Lisboa *et al.* in 2003, who introduced a neural network framework for survival analysis, highlighting its application in medical prognostics and emphasizing the flexibility of

ANNs in handling non-linear relationships [18]. In recent years, the application of ANNs in survival analysis has seen significant advancements with the integration of deep learning techniques. Ching *et al.* in 2018 developed Cox-nnet, an ANN-based method tailored for high-dimensional omics data, demonstrating its effectiveness in cancer prognosis [19]. Similarly, Luck *et al.* in 2017 presented a comprehensive approach to applying deep learning to survival analysis, introducing architectures Artificial neural networks to handle time-dependent covariates [20]. Katzman *et al.* in 2018 introduced DeepSurv, a deep neural network implementation of the Cox proportional hazards model, which showed substantial improvements in modeling complex interactions and provided personalized treatment recommendations [21]. Another notable advancement was made by Lee *et al.* in 2018, who introduced Dynamic-DeepHit, a novel deep learning framework for dynamic survival analysis with competing risks, capable of handling the dynamic aspect of time-to-event data [22]. These advancements underscore the evolution of ANNs in survival analysis, from early studies that established their potential to contemporary models that leverage deep learning to handle complex, high-dimensional data. This progression has significantly enhanced the ability to provide accurate and personalized predictions, particularly in healthcare and medical research, making ANNs an invaluable tool in modern survival analysis.

2.2.1. Network Architecture: Multi-Layer Neural Network

A multi-layer neural network, also known as a multi-layer perceptron (MLP), is a type of artificial neural network that consists of multiple layers of neurons. It is composed of an input layer, one or more hidden layers, and an output layer, with neurons in each layer connected to neurons in adjacent layers through weighted connections and biases. The network learns by adjusting these weights and biases during training to minimize the error in its predictions.

Through the feedforward process, data is propagated through the network from the input layer to the output layer. Each neuron in a layer receives inputs from the previous layer, computes the weighted sum of inputs, applies the activation function, and passes the result to the neurons in the next layer. The key components and composition of a multi-layer neural network are as follows [12] [23]:

1. Input Layer: The input layer consists of neurons that receive input data. Each neuron in this layer corresponds to a feature in the input data. Let

$\mathbf{x} = [x_1, x_2, \dots, x_p]$ be the input vector.

2. Hidden Layers: Hidden layers are the intermediate layers between the input layer and the output layer. A multi-layer neural network can have one or more hidden layers. Neurons: A neuron j in a neural network can be represented mathematically as follows:

$$\begin{aligned} z_j &= \sum_{i=1}^n w_{ij} \cdot x_i + b_j \\ a_j &= f(z_j) \end{aligned} \quad (1)$$

where: z_j is the weighted sum of inputs to neuron j , x_i is the input from

neuron i in the previous layer, w_{ij} is the weight associated with the connection between neuron i and neuron j , b_j is the bias term for neuron j , f is the activation function, and h_j is the output of neuron j after applying the activation function.

Connections: The weights between neurons in adjacent layers are represented as a weight matrix $W^{(l)}$, where each element $w_{ij}^{(l)}$ represents the weight of the connection from neuron i in layer $l-1$ to neuron j in layer l . The weights determine the strength and direction of the influence between neurons, while the biases allow the network to shift the activation function.

Each neuron applies an activation function to its weighted input sum plus bias.

Common activation functions include: Sigmoid: $f(z) = \frac{1}{1+e^{-z}}$

Tanh: $f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$

ReLU (Rectified Linear Unit): $f(z) = \max(0, z)$

Softmax: $f(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$ (used in the output layer for multi-class classification)

The network learns from data through a process called training. During training, the network adjusts the weights of the parameters to minimize a loss function, J using an optimization algorithm such as gradient descent and backpropagation.

The weight update rule for gradient descent can be expressed as:

$$w_{ij}^{(l)} = w_{ij}^{(l)} - \alpha \frac{\partial J}{\partial w_{ij}^{(l)}}. \quad (2)$$

where α is the learning rate.

Backpropagation involves computing the gradients of the loss function with respect to the weights using the chain rule and updating the weights accordingly. The loss function $J(\mathbf{w}, \mathbf{b})$ measures the difference between the predicted output \hat{y} and the true output y . The choice of the loss function depends on the specific task the neural network is being trained for. For classification tasks, where the goal is to predict class labels, the Cross-Entropy Loss (also known as the log-loss) is commonly used. The form of the cross-entropy loss depends on whether the classification is binary or multi-class. For binary classification, where the output is either 0 or 1, the binary cross-entropy loss function is defined as:

$$J(\mathbf{w}, \mathbf{b}) = -\frac{1}{m} \sum_{i=1}^m [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]. \quad (3)$$

where: y_i is the true label (0 or 1) for the i -th training example, \hat{y}_i is the predicted probability that the output is 1 for the i -th training example.

For multi-class classification, where there are K possible classes, the categorical cross-entropy loss function is defined as:

$$J(\mathbf{w}, \mathbf{b}) = -\frac{1}{m} \sum_{i=1}^m \sum_{k=1}^K y_{i,k} \log(\hat{y}_{i,k}). \quad (4)$$

where: K is the number of classes, $y_{i,k}$ is a binary indicator (0 or 1) if class label k is the correct classification for the i -th training example, $\hat{y}_{i,k}$ is the predicted probability that the i -th training example belongs to class k . During training, we need to specify the optimization algorithm to use. For example, RSM prop, Stochastic Gradient Descent, Adam) and the rationale behind its choice. Epochs and Batches: During training, the dataset is divided into batches, and the network goes through multiple epochs. The weight updates are typically performed after processing each batch. Detail the training procedure, including the number of epochs, batch size, and any techniques used to prevent overfitting such as dropout, early stopping, regularization.

3. Output Layer: The output layer consists of neurons that produce the final output of the network. The number of neurons in the output layer depends on the task. For example, in a classification task with K classes, the output layer will have K neurons and for a network with L layers, the final output is computed as:

$$z_j^{(L)} = \sum_{i=1}^{n^{(L-1)}} w_{ij}^{(L)} a_i^{(L-1)} + b_j^{(L)} \quad (5)$$

$$y = a^{(L)} = f(z^{(L)}).$$

4. Evaluation: After training, the network is evaluated on a separate validation or test dataset to assess its performance. Based on the problem you are tackling, we define the metrics used to evaluate the model's performance, such as accuracy, precision, recall, F1 score, ROC-AUC for classification, and RMSE, MAE for regression. A simplified artificial neural network architecture with the description above is shown in **Figure 5**.

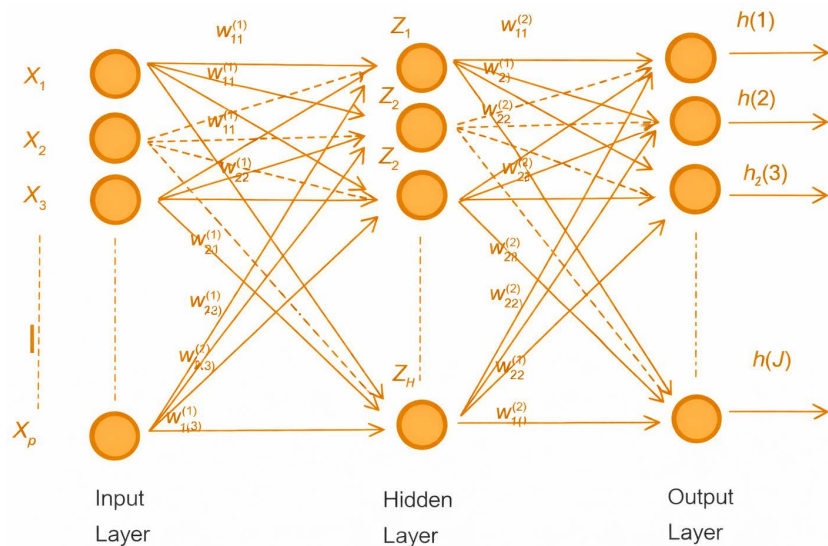


Figure 5. Artificial neural network architecture.

2.2.2. Implementation of the ANN Model

We used the data described under the data description, which are patients with

malignant gliomas (brain tumors) along with 14 risk factors. The continuous risk factors were normalized, and categorical variables were one-hot encoded. The data was then split into a training set (90%), a validation set (7%), and a test set (3%). Below is the structure of the data and the algorithm flowchart shown in **Figure 6**. No data balancing techniques were applied during preprocessing. All class distributions in the validation and test sets reflect the original dataset to ensure realistic performance evaluation.

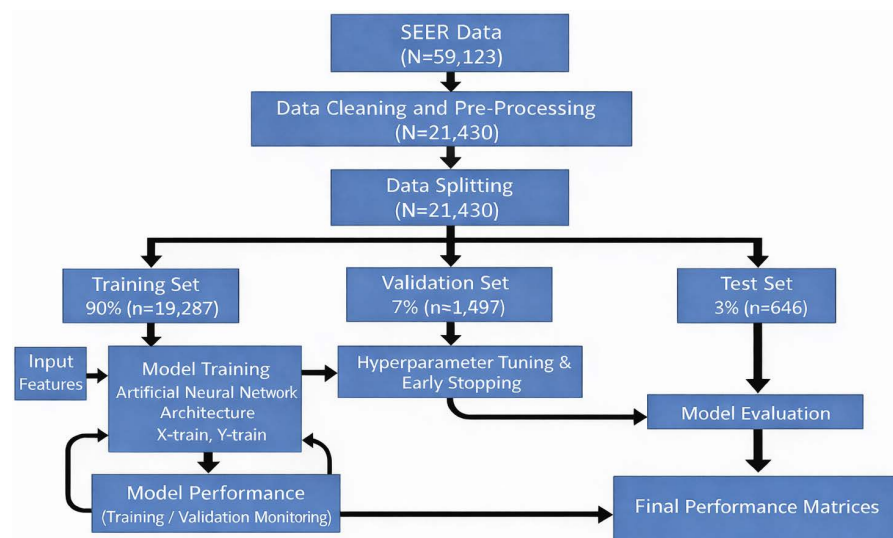


Figure 6. Schematic diagram showing the data structure and algorithm flowchart.

2.2.3. The Art of Model Training

Building an Artificial Neural Network (ANN) model is typically an empirically iterative process. This involves experimenting with different model architectures and hyperparameters through training data by iteratively updating their parameters to minimize a defined loss function, an approach known as empirical risk minimization.

In the art of model training, we went through the crucial processes to find an optimum neural network model based on performance for our objective. We built an ANN binary classifier to predict the survival time of patients with malignant gliomas. The survival time was split into two categories using the mean survival time of $11.73 \approx 12$ months. The classifier aimed to distinguish between patients who could survive beyond 12 months and those whose survival would be within 12 months. To develop this binary classifier, we considered various configurations and options through a trial-and-error approach. The configurations explored are detailed below:

Model Architecture: We experimented with different model architectures, including deeper or wider networks. This involves adjusting the number of hidden layers and the number of units (neurons) in each layer. The goal was to find a balance that allowed the model to learn complex patterns without overfitting.

Weight Initialization: Different weight initialization techniques were tested to en-

sure that the initial weights of the model were set optimally. *Activation Functions:* For the hidden layers, we tried different activation functions such as Tanh and ReLU. The choice of activation function affects how the neural network learns and represents complex patterns in the data. For the output layer, we used the sigmoid activation function to produce probabilities for the binary classification task. *Regularization:* To prevent overfitting, we applied regularization techniques. L2 regularization (weight decay) and dropout regularization were particularly useful. L2 regularization adds a penalty to the loss function proportional to the square of the weights, discouraging overly complex models. Dropout regularization involves randomly setting a fraction of the neurons to zero during each training step, which helps to prevent the network from becoming too reliant on any single neuron. *Hyperparameter Tuning:* We performed hyperparameter tuning to find the best combination of hyperparameters for training our dataset. This included experimenting with different learning rates, batch sizes, numbers of epochs, activation functions, and early stopping. *Learning rate:* controls how much the model's weights are adjusted during each iteration. The learning rate determines the size of the steps the model takes while optimizing. *Batch Size:* We experimented with different batch sizes to find the optimal batch size that balances training speed and model performance. The batch size used during training was adjusted to balance the trade-off between training speed and generalization. *Epochs and Early Stopping:* Training was conducted over multiple epochs, and early stopping was implemented with patience values. Early stopping monitors the model's performance on the validation set and stops training if there is no improvement after a certain number of epochs. *Algorithm Optimization:* When it comes to algorithm optimization, there are several optimizers available, such as Momentum, RMS Prop, and Adam, among others. For our model, we decided to use the Adam optimizer. Adam is a combination of Momentum and RMS Prop and is one of the most commonly used optimizers in neural network models due to its efficiency and effectiveness in handling sparse gradients on noisy problems. Adam computes adaptive learning rates for each parameter, which helps in achieving faster convergence and better performance. *Loss Function:* Binary cross-entropy loss was used as the loss function. This is appropriate for binary classification tasks as it measures the performance of a classification model whose output is a probability value between 0 and 1. For the multi-class model, we employed the softmax activation function in the output layer and used categorical cross entropy as the loss function.

These steps collectively ensured that our model was well-tuned and capable of delivering high performance in predicting the survival time of the patients. **Table 1** and **Table 2** present the process described in the ANN model building and implementation.

After running different versions of the ANN models shown in **Table 1** and **Table 2**, our proposed model (best model selected) for both binary and multi-class classification is a 4-layer architecture with 3 hidden layers [64, 32, 16]. The model

parameters are as follows: dropout = 0.2, learning rate = 0.001, optimizer = Adam, batch size = 32, epochs = 50, and patience = 10.

Table 1. ANN models architecture and specifications.

ANN Models	Hidden Layers Architecture	Activation Function (Hidden Layer)	Activation Function (Output Layer)	Loss Function
Binary Classifier ($<12 = 0, \geq 12 = 1$)	[128, 64, 32, 16]	Tanh, ReLU	Sigmoid (Dense = 3)	Binary Cross Entropy
Multi-Layer Classifier ($<12 = 0, \geq 12 = 1, \geq 24 = 2$)	[128, 64, 32, 16]	Tanh, ReLU	Softmax (Dense = 3)	Categorical Cross Entropy

Table 2. ANN Models hyperparameter tuning/setting.

ANN Models	Dropout	Batch Size	Learning Rate	Epoch	Early Stopping
Binary Classifier ($<12 = 0, \geq 12 = 1$)	[0.2, 0.3, 0.4]	[64, 32]	0.1 0.01 0.001	[10, 20, 30, 50, 100, 200]	[5, 10, 20, 30, 50]
Multi-Layer Classifier ($<12 = 0, \geq 12 = 1, \geq 24 = 2$)	[0.2, 0.3, 0.4]	[64, 32]	0.1 0.01 0.001	[10, 20, 30, 50, 100, 200]	[5, 10, 20, 30, 50]

2.3. Random Forest (RF)

Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. The process involves bootstrap sampling, where the training dataset is randomly sampled with replacement to create multiple subsets. For each subset, a decision tree is built using a random subset of features at each split. The predictions of all the trees are then combined through voting (for classification) or averaging (for regression) to produce the final output.

2.4. Logistic Regression

Logistic regression as a binary classification model can be likened to an Artificial Neural Network (ANN) with one hidden layer. It predicts the probability of a binary outcome using a logistic function, similar to an ANN's input, hidden, and output layers. The process involves defining the model, estimating coefficients using maximum likelihood estimation, and making predictions by converting inputs to probabilities and applying a threshold.

2.5. Gradient Boosting Machine (GBM)

Gradient Boosting Machine (GBM) is an ensemble learning technique used for classification that builds models sequentially, with each new model correcting the errors of the previous ones, combining the strengths of multiple weak learners,

usually decision trees. The process begins with an initial weak model, such as a small decision tree. The residual errors of the model are then calculated, and a new model is fitted to these residuals. This new model is added to the ensemble, adjusting the predictions accordingly. The steps of calculating residuals, fitting new models, and updating the ensemble are repeated for a specified number of iterations or until convergence.

2.6. K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) is a non-parametric, instance-based learning algorithm that classifies a data point based on the classification of its neighbors. The process begins with distance calculation, where the distance between the input data point and all points in the training set is computed, commonly using Euclidean distance. The K closest neighbors to the input data point are then selected. For classification tasks, the class of the input data point is determined by a majority vote of its neighbors.

3. Results

3.1. Model Performance and Evaluation

Our proposed ANN-based survival analysis models aim to predict the survival time of glioma patients. A confusion matrix is a table used to evaluate the performance of a classification algorithm. It is especially useful for understanding the performance of a model in binary and multiclass classification tasks. The matrix compares the actual target values with those predicted by the machine learning model. The structure of the confusion matrix for a multi-class classification typically does not differ much in meaning from the content presented above. **Figure 7** below shows the confusion matrix of the ANN BC Model classifier.

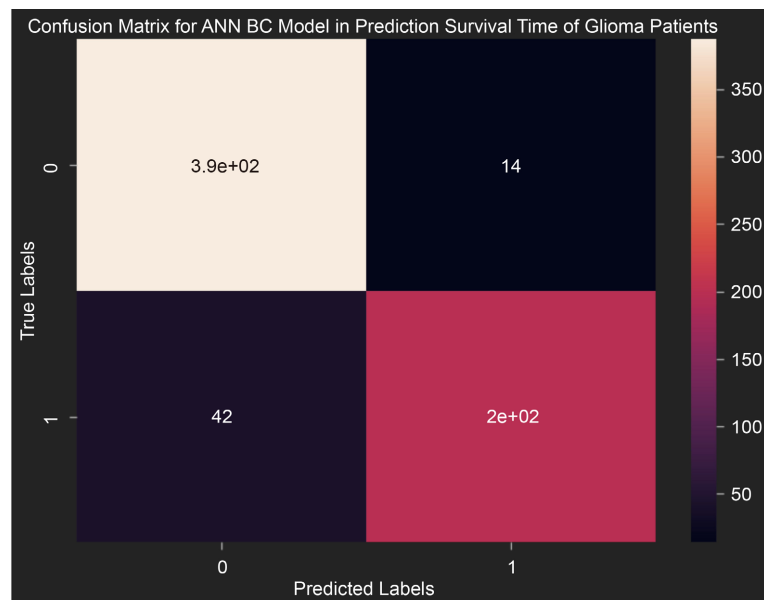


Figure 7. Confusion matrix for ANN BC model.

The predictions indicate the following: A True Negative (TN) count of 390 indicates that the model correctly predicted that 390 patients would survive less than 12 months. The False Positive (FP) count of 14 means the model predicted that 14 patients would survive 12 months or more, but they actually survived less than 12 months. A False Negative (FN) count of 42 indicates that the model predicted that 42 patients would survive less than 12 months, but they actually survived 12 months or more. Lastly, the True Positive (TP) count of 200 means the model correctly predicted that 200 patients would survive 12 months or more. The overall accuracy of the ANN BC Model is 0.9126; (91.26%). The incorrectly predicted (error) survival times by the model is due the False Positive and False Negative. Type I Error (False Positive) was predicting survival of 12 months or more when the actual survival is less than 12 months. In the model, this corresponds to the 14 false positives. Type II Error (False Negative) was predicting survival of less than 12 months when the actual survival is 12 months or more. In the model, this corresponds to the 42 false negatives.

The implications of false positives and false negatives in predictive models for patient survival can profoundly impact both patients and their families. False positives, where the model wrongly predicts longer survival times, can lead to unjustified hope, only to be shattered when reality sets in, causing emotional and psychological distress. Conversely, false negatives, where shorter survival times are predicted when patients actually have longer lifespans, can instigate unnecessary anxiety and stress due to a mistaken poor prognosis. In medical contexts, often false negatives are considered more critical because missing a diagnosis can have immediate life-threatening consequences. However, both types of errors are critical to minimize to ensure accurate patient care and effective allocation of resources. Research has shown that the clinical consequences of false positives and false negatives are rarely symmetric, and clinicians may prioritize sensitivity over specificity in high-stakes medical decision making to reduce the likelihood of dangerous missed diagnoses [24] [25]. Additionally, screening studies highlight that false positives can lead to unnecessary procedures and psychological distress, while false negatives may result in delayed treatment or false reassurance. Cost-sensitive approaches in machine learning further underscore the need to weight these errors differently based on their real-world impact [26].

The ANN multiclass model predicts whether a patient will survive within the first 12 months, between 12 and 24 months, or beyond 24 months shown in **Figure 8**. The diagonal elements of the matrix represent correct predictions, with 246 instances correctly predicted for patients surviving less than 12 months (class 0), 156 instances for those surviving 12 - 24 months (class 1), and 61 instances for those surviving more than 24 months (class 2). The off-diagonal elements indicate misclassifications. Model performance insights indicate that the high diagonal values suggest reasonably good accuracy, particularly for patients surviving less than 12 months and those surviving 12 - 24 months. However, the model tends to confuse patients surviving more than 24 months with those surviving 12 - 24

months more often than with those surviving less than 12 months. The overall accuracy of the ANN multiclass model is 0.8134 (81.34%).

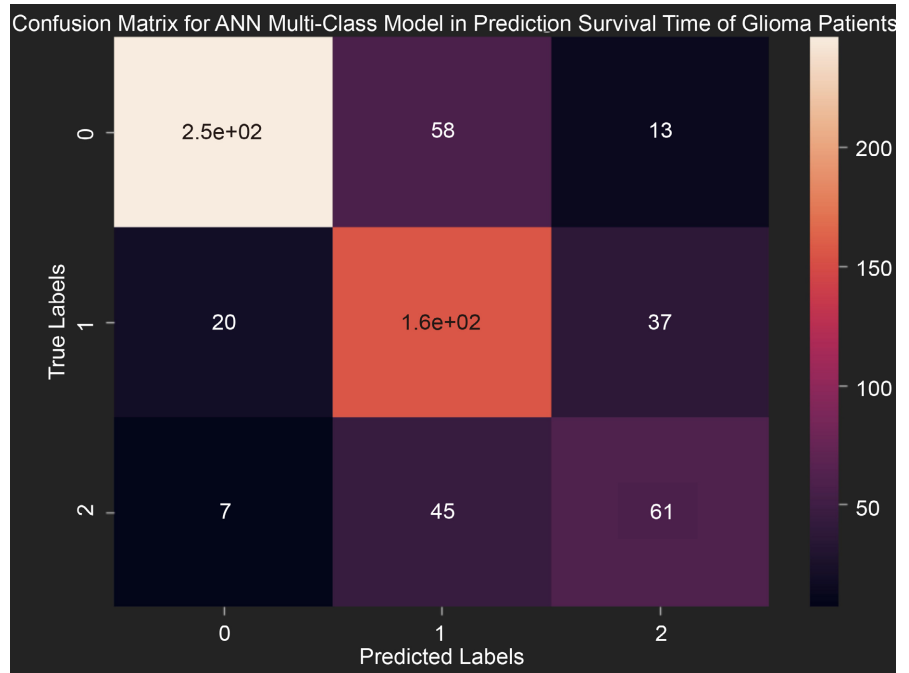


Figure 8. Confusion matrix for ANN multi-class model.

3.2. Model Loss

Figure 9 shows the training and validation loss from the ANN model, where a lower loss value indicates better model performance. Both the training and validation losses decrease rapidly, indicating that the model is learning and improving its performance on both datasets. The validation loss levels off while the training loss continues to decrease slightly, suggesting no risk of overfitting. The minimal gap between training and validation loss demonstrates the model's strong generalization ability on unseen data. As the validation loss has stopped improving significantly while the training loss continues to decrease, it may be an optimal point to stop training to prevent overfitting. Overall, the plot indicates that the ANN model has learned effectively, with both training and validation loss decreasing over the epochs.

3.3. Models Performance Metrics

We considered specificity, sensitivity, the F1 score, and the area under the receiver operating characteristic curve (AUC) to provide a comprehensive evaluation. Sensitivity measures the proportion of true positive instances among the actual positive instances, while specificity measures the proportion of true negative instances among the actual negative instances. The F1 score, which is the harmonic mean of precision and recall, provides a single metric that balances both concerns. The AUC offers an aggregate measure of performance across all classification thresh-

olds, with a higher AUC indicating better model performance. By presenting these metrics, we aim to compare the ANN model to other supervised machine learning models built on the same dataset and highlight their capacities.

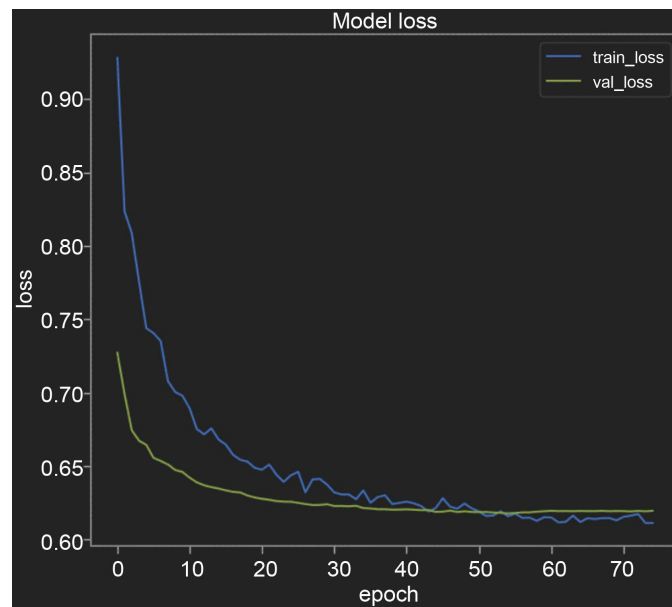


Figure 9. Evaluation the ANN BC model loss.

Table 3. Performance metrics summary of the models.

Models	Accuracy	Sensitivity	Specificity	F1 Score	AUC
ANN BC	0.9126	0.8250	0.9651	0.8761	0.90
RF	0.8908	0.7686	0.9642	0.8416	0.87
Logistic	0.8574	0.7215	0.9391	0.7917	0.83
GBM	0.8515	0.7029	0.9268	0.7834	0.82
KNNs	0.8425	0.7143	0.9342	0.7706	0.81
ANN MC	0.8134	0.6827	0.8611	0.6972	0.85

The performance metrics comparison of models reveals distinct strengths and weaknesses across different criteria, as shown in **Table 3**. In terms of accuracy, the Artificial Neural Network Binary Classifier (ANN BC) leads with a score of 0.9126, followed by the Random Forest (RF) at 0.8908, indicating both models make the most correct predictions. Logistic Regression, Gradient Boosting Machine (GBM), and K-Nearest Neighbors (KNN) show moderate accuracy, while the Artificial Neural Network Multi-Class (ANN MC) lags behind. ANN BC again performs best with a sensitivity of 0.8250. RF follows with 0.7686, while Logistic Regression and KNN are slightly lower but still competitive.

For specificity, ANN BC and RF are almost tied, with scores of 0.9651 and 0.9642, respectively, indicating they are excellent at correctly identifying negatives.

Logistic Regression, KNN, and GBM also perform well, though slightly lower, and ANN MC has the lowest specificity.

In terms of the F1 score, which balances precision and recall, ANN BC stands out with a score of 0.8761. RF follows with 0.8416, showing good balance. Logistic Regression, GBM, and KNN have moderate F1 scores, while ANN MC has the lowest, indicating the least balance between precision and recall, which is understandable because it handles three classifications as compared to the others with binary classifications.

For the AUC (Area Under the Curve), which measures the model's overall ability to distinguish between classes, ANN BC leads with 0.90, demonstrating superior discriminative ability and the best overall performance in distinguishing between classes. RF follows with 0.87, and while Logistic Regression and GBM are close, ANN MC performs better with 0.85, becoming better than Logistic Regression and GBM, KNN's, despite its lower performance in other areas. The Gradient Boosting Machine (GBM) has an AUC of 0.82, closely followed by the K-Nearest Neighbors (KNN) model with an AUC of 0.81, both indicating moderate performance.

The ROC curves and AUC values for the different models are displayed in **Figure 10**.

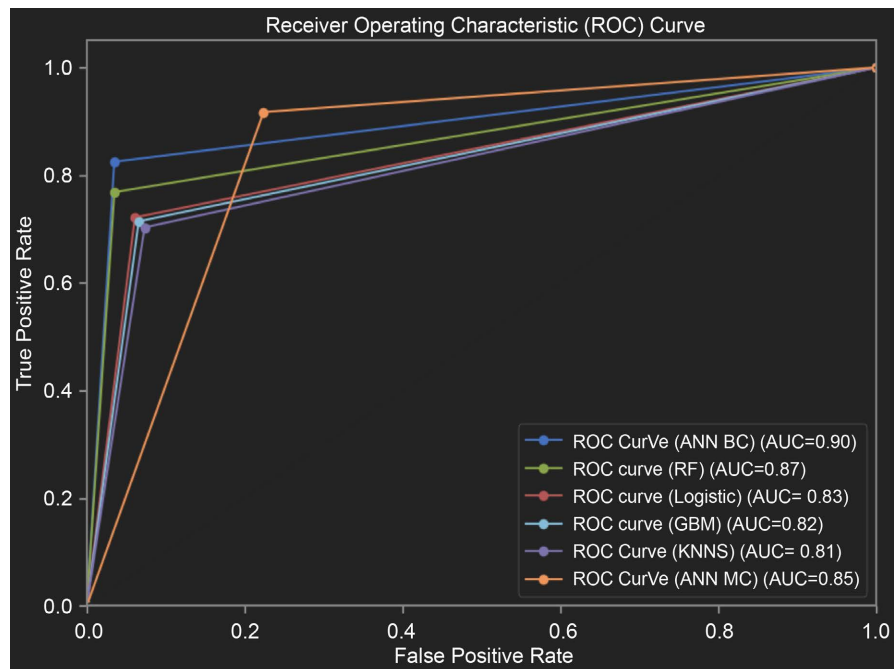


Figure 10. The ROC curves and AUC values for all models.

Overall, the ANN BC model consistently outperforms others across most metrics, making our built-proposed model the best choice model. The other models show varying degrees of performance, with the random forest (RF) performing relatively well and could be considered a reliable alternative.

3.4. Relative Importance

The relative importance of risk factors was estimated using an artificial neural network-based connection-weight approach. After training the ANN model, the contribution of each input variable was quantified by aggregating the absolute values of the connection weights linking input neurons to the hidden layer and subsequently to the output node. These aggregated weights reflect the relative influence of each predictor on the model's output. The resulting values were normalized to sum to 100%, yielding relative importance scores that were used to rank the top contributing risk factors. **Figure 11** shows the rank of the top 10 risk factors by their relative importance in contributing to the survival time of glioma patients.

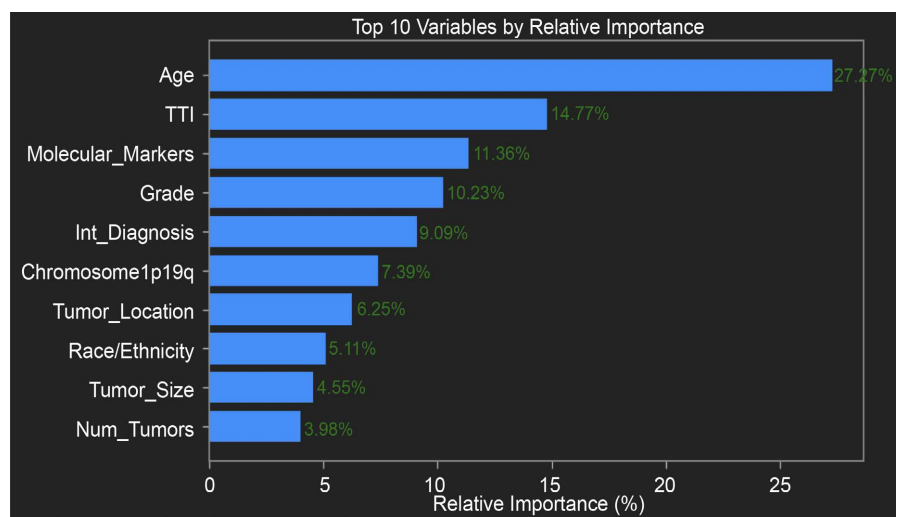


Figure 11. Ranking of risk factors based on relevance importance.

The most significant factor is Age, with a relative importance of 27.27%, indicating that the age of the patient is the strongest predictor of survival time. This aligns with medical literature, which often highlights that older age is associated with poorer prognosis and less survival time in glioma patients due to the body's decreased ability to tolerate aggressive treatment and the presence of comorbidities. Thereby explicitly recognizes age at diagnosis as a key clinical prognostic indicator to tumor biology, molecular subtypes, and survival outcomes [27] [28]. The second most important variable is TTI (Time to Treatment Initiation) at 14.77%. This underscores the critical role of timely intervention in improving survival outcomes, as patients who start treatment early generally have longer survival times than those who delay. Early treatment may limit tumor progression and improve the effectiveness of therapeutic interventions [29]. Molecular Markers, contributing 11.36%, are also crucial, reflecting the growing recognition of genetic and molecular profiling in tailoring personalized treatment plans and predicting patient outcomes. The Grade of the tumor, with a 10.23% importance, is another significant factor, as higher-grade tumors are typically more aggressive

and associated with shorter survival times. This is coherent with finding that molecular profiling refines survival prediction beyond histology and that tumor grade remains a strong determinant of aggressiveness and survival, which forms the global clinical standard for glioma diagnosis and prognosis [30]. Chromosome 1p/19q codeletion status, contributing 7.39%, is a well-documented prognostic marker in glioma, with studies showing that patients with this genetic alteration often have better responses to treatment and longer survival. A study emphasized the prognostic significance of 1p/19q codeletion in glioma. The ninth and tenth top risk factors are tumor size (4.55%), indicating the extent of disease burden, and Number of tumors (3.98%), which could complicate treatment and prognosis [31]. These findings are supported by various studies and medical literature, which consistently highlight the multifactorial nature of glioma prognosis. For instance, a study by Lacroix *et al.* found that younger age, smaller tumor size, and certain molecular markers were associated with improved survival [32]. We are able to identify top 10 significant risk factors which contributes to the ANN model in predicting the survival times of patients with Malignant Gliomas.

3.5. Generalized Pareto Distribution for Survival Analysis of Patients with Malignant Gliomas

We identified a well-defined probability distribution that accurately characterizes the behavior of the survival times of malignant glioma patients as the 3-Parameter Generalized Pareto Probability Distribution. The use of the 3-parameter Generalized Pareto distribution in generating these survival probabilities provides a robust method for modeling the extreme values and tail behavior of survival times across the different significant identified risk factors. Before proceeding to perform the parametric analysis of the survival times of patients with Malignant Gliomas, we sought to determine if there was a difference in survival times based on gender (*i.e.*, male and female) using the log-rank test [33].

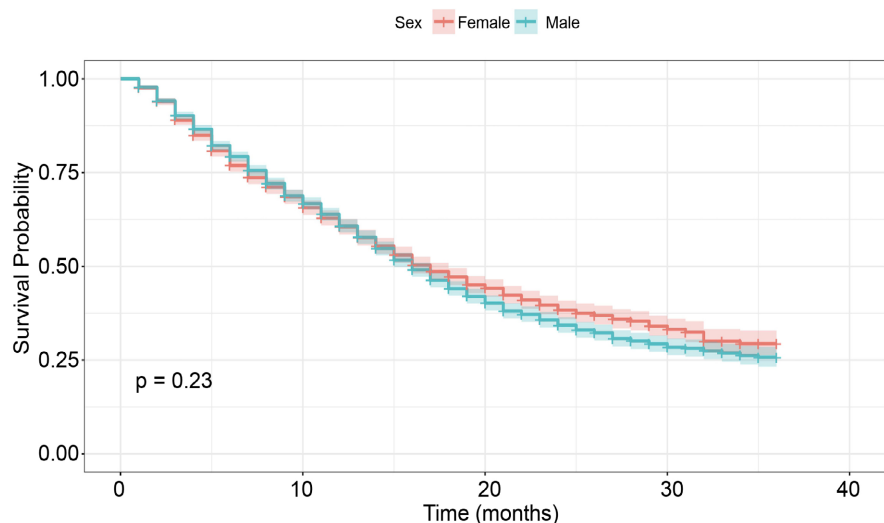


Figure 12. Log-rank test for difference in survival time of gender.

From **Figure 12**, the log-rank test yielded a large p-value of 0.23, indicating a failure to reject the null hypothesis (*i.e.* $H_0 : \mu_M = \mu_F$) that there is no difference in survival times between males and females. Since there is no discernible difference in survival times between males and females, we proceeded with the parametric analysis without stratification based on gender. Through the parametric analysis, we determined that the probability distribution that accurately characterizes the behavior of the survival time of patients with Malignant Gliomas follows a three-parameter General Pareto Probability Distribution.

Given the real-world nature of the data, there is a need to validate the choice of the distribution. The statistical analysis emphasized empirical goodness-of-fit and tail behavior to assess distributional adequacy for real-world data containing extreme observations. The empirical cumulative distribution function (ECDF)-based goodness-of-fit tests were employed as the primary comparative statistical criteria by evaluating agreement between the empirical cumulative distribution function and the corresponding theoretical cumulative distribution function. The Kolmogorov-Smirnov goodness-of-fit test is a statistical test used to determine if a set of observations is drawn from a hypothesized continuous probability distribution. It is based on the empirical cumulative distribution function (ECDF), which is a step function that jumps up by $\frac{1}{n}$ at each observed data point. Given a dataset $T = \{t_1, t_2, \dots, t_n\}$ from a continuous distribution with a cumulative distribution function (CDF), the ECDF is defined as:

$$F_n(t) = \frac{1}{n}(\text{number of observations} \leq t).$$

The Kolmogorov-Smirnov statistic is based on the largest vertical difference between the CDF and the ECDF, and is defined as:

$$D_n = \max \{ |F_n(t) - F(t)| \}.$$

Under the null hypothesis that the data follow the specified distribution, the distribution of the Kolmogorov-Smirnov statistic is known. The hypothesis is rejected at the α level of significance if the calculated statistic is greater than the critical value of the theoretical distribution. Anderson-Darling test is a useful tool for assessing the goodness-of-fit of a given dataset to a specified probability distribution, especially when the tails of the distribution are of particular interest. The test is based on the comparison of the observed cumulative distribution function (CDF) to the expected CDF of the specified distribution. This test places more emphasis on the tails of the distribution than the Kolmogorov-Smirnov test. The Anderson-Darling statistic, denoted by

$$A^2 = -n - \sum_{i=1}^n \frac{2i-1}{n} \left[\ln F(t_i) + \ln(1 - F(t_{n-i+1})) \right],$$

is calculated as the weighted sum of the squared differences between the observed cumulative distribution function and the expected cumulative distribution function. The weights are chosen based on the inverse of the variance of the expected

distribution. The statistic A^2 is then compared to the critical value of the theoretical distribution under the null hypothesis that the data follow the specified distribution. The chi-squared statistic is calculated under the null hypothesis that the data follow the specified distribution. The chi-squared goodness-of-fit test is a statistical test that the expected frequencies are calculated using the cumulative distribution function of the specified distribution. The resulting statistic follows a chi-squared distribution with $k - 1$ degrees of freedom, where k is the number of cells. There is no optimal choice for the number of cells. However, each cell is required to contain at least five data points. Given a dataset $T = \{t_1, t_2, \dots, t_n\}$ with cumulative distribution function $F(t)$, the chi-squared statistic denoted by χ^2 is defined as,

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i},$$

where O_i is the observed frequency for cell i , and E_i is the expected frequency for cell i calculated as

$$E_i = F(t_2) - F(t_1),$$

where t_1 and t_2 are the limits for cell i . The chi-squared statistic is calculated under the null hypothesis that the data follow the specified distribution. The hypothesis is rejected at the level of significance if the statistic χ^2 is greater than the critical value $\chi_{0.05, k-1}^2$.

Figure 13 depicts the identified probability distribution of survival time among malignant glioma patients.

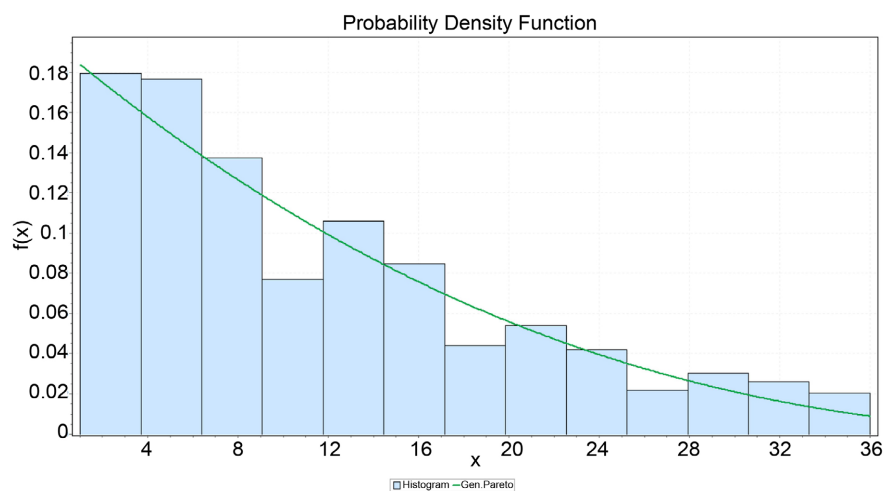


Figure 13. 3-P generalized pareto distribution of survival time of malignant gliomas.

Table 4 shows three different goodness-of-fit tests of the 3p-General Pareto probability distribution of the survival times of the Malignant Glioma patients given by the Kolmogorov-Smirnov, Anderson-Darling, and Chi-square test. The test results revealed p-values exceeding 0.05, indicating a failure to reject the null hypothesis, H_0 : the probability distribution follows the 3p-Generalised Pareto

Probability Distribution. This suggests that the selected probability distribution offers a robust fit for the data.

Table 4. Goodness-of-fit test of the 3-P general pareto probability distribution.

Type of Test	P-value
Kolmogorov-Smirnov	0.4019
Anderson-Darling	0.4781
chi-squared	0.2026

The Generalized Pareto Distribution (GPD) is a statistical distribution used to model extreme events or values that exceed a certain threshold. It is a flexible distribution commonly employed in risk analysis and extreme value theory. The GPD is characterized by three parameters: location μ , scale σ , and shape ξ parameters. Given the survival time of Malignant Gliomas, t as a random variable, then the pdf of the 3p-General Pareto probability distribution is given by,

$$f(t | \mu, \sigma, \xi) = \begin{cases} \frac{1}{\sigma} \left(1 + \frac{\xi(x - \mu)}{\sigma} \right)^{-1/\xi - 1}, & \text{if } \xi \neq 0 \\ \frac{1}{\sigma} \exp\left(-\frac{x - \mu}{\sigma}\right), & \text{if } \xi = 0 \end{cases} \tag{6}$$

where ξ is the shape parameter, which governs the tail behavior of the distribution, $\sigma > 0$ denotes continuous scale parameter, which determines the spread or width of the distribution, μ represents the continuous location parameter, which shifts the distribution horizontally. To estimate the three parameters μ, σ, ξ , we apply the maximum likelihood estimation (MLE) procedure. MLE estimates the parameters of the probability distribution by maximizing the likelihood function. MLE is the most widely used parameter estimation method due to its robustness compared to other traditional methods such as the method of moment estimation [34] [35]. To compute the MLE estimators, we first define the likelihood function

$L(\mu, \sigma, \xi | t_1, t_2, \dots, t_n)$ which is the product of the individual probability densities for each data point which can be expressed as follows:

we simplified this expression further by expanding the logarithm term and taking the summation over the observed data points t_1, t_2, \dots, t_n .

$$\begin{aligned} L(\mu, \sigma, \xi | t_1, t_2, \dots, t_n) &= \prod_{i=1}^n f(t_i | \mu, \sigma, \xi) \\ &= \prod_{i=1}^n \left[\frac{1}{\sigma} \left(1 + \frac{\xi(t_i - \mu)}{\sigma} \right)^{-1/\xi - 1} \right] \end{aligned} \tag{7}$$

Now, we take the log of the likelihood function in Equation (7), given by

$$\begin{aligned} \ln L &= \ln L(t_i | \mu, \sigma, \xi) \\ \ell(\mu, \sigma, \xi | t_1, t_2, \dots, t_n) &= \sum_{i=1}^n \log \left[\frac{1}{\sigma} \left(1 + \frac{\xi(t_i - \mu)}{\sigma} \right)^{-\frac{1}{\xi} - 1} \right] \end{aligned} \tag{8}$$

We can expand the logarithm term using the properties of logarithms:

$$\log \left[\frac{1}{\sigma} \left(1 + \frac{\xi(t_i - \mu)}{\sigma} \right)^{\frac{1}{\xi} - 1} \right] = \log \left(1 + \frac{\xi(t_i - \mu)}{\sigma} \right)^{\frac{1}{\xi} - 1} - \log \sigma$$

Therefore, the log-likelihood function becomes:

$$\begin{aligned} \ell(\mu, \sigma, \xi | t_1, t_2, \dots, t_n) &= \sum_{i=1}^n \left[\log \left(1 + \frac{\xi(t_i - \mu)}{\sigma} \right)^{\frac{1}{\xi} - 1} - \log \sigma \right] \\ &= \sum_{i=1}^n \left[-\left(\frac{1}{\xi} + 1 \right) \log \left(1 + \frac{\xi(t_i - \mu)}{\sigma} \right) - \log \sigma \right] \\ \ell(\mu, \sigma, \xi | t_1, t_2, \dots, t_n) &= -\left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^n \log \left(1 + \frac{\xi(t_i - \mu)}{\sigma} \right) - n \log \sigma \quad (9) \end{aligned}$$

Differentiate the Log-Likelihood Function, by taking partial derivatives of the log-likelihood function with respect to each parameter:

$$\ell(\mu, \sigma, \xi | x_1, x_2, \dots, x_n) = -\left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^n \log \left(1 + \frac{\xi(x_i - \mu)}{\sigma} \right) - n \log \sigma$$

a. Partial derivative with respect to μ :

$$\frac{\partial \ell}{\partial \mu} = -\left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^n \frac{\frac{\xi}{\sigma}}{1 + \frac{\xi(x_i - \mu)}{\sigma}}$$

b. Partial derivative with respect to σ :

$$\frac{\partial \ell}{\partial \sigma} = -\left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^n \frac{\frac{\xi(x_i - \mu)}{\sigma}}{\sigma \left(1 + \frac{\xi(x_i - \mu)}{\sigma} \right)} - \frac{n}{\sigma}$$

c. Partial derivative with respect to ξ :

$$\frac{\partial \ell}{\partial \xi} = \left(\frac{1}{\xi^2} + \frac{1}{\xi} \right) \sum_{i=1}^n \log \left(1 + \frac{\xi(x_i - \mu)}{\sigma} \right) - \sum_{i=1}^n \frac{x_i - \mu}{\sigma \left(1 + \frac{\xi(x_i - \mu)}{\sigma} \right)}$$

Next, we set each partial derivative equal to zero and solve the resulting equations to find the maximum likelihood estimates (MLEs) of μ , σ , and ξ .

$$\frac{\partial \ell}{\partial \mu} = 0, \quad \frac{\partial \ell}{\partial \sigma} = 0, \quad \frac{\partial \ell}{\partial \xi} = 0$$

This step may involve solving a system of equations. We Verify with Second derivatives by checking the second derivatives to ensure that the estimates are indeed maxima.

$$\frac{\partial^2 \ell}{\partial \mu^2}, \quad \frac{\partial^2 \ell}{\partial \sigma^2}, \quad \frac{\partial^2 \ell}{\partial \xi^2}$$

Solving these equations of the partial derivatives simultaneously for μ , σ , and ξ will give us the maximum likelihood estimates (MLEs) of the parameters

which are given in **Table 5** below.

Table 5. Parameter estimates for the three-parameter general pareto probability distribution.

Location ($\hat{\mu}$)	Scale ($\hat{\sigma}$)	Shape ($\hat{\xi}$)
0.6618	14.395	-0.2918

We substitute the parameter estimates in **Table 5** into Equation (6) to obtain the estimated probability density function (pdf) of the 3p-General Pareto Probability Distribution of the Survival Time of Patients with Malignant Gliomas, given by:

$$f(t) = \begin{cases} \frac{1}{14.395} \left(1 + \frac{-0.2918(t-0.6618)}{14.395} \right)^{-1/(-0.2918)-1}, & \text{if } \xi \neq 0 \\ \frac{1}{14.395} \exp\left(-\frac{t-0.6618}{14.395}\right), & \text{if } \xi = 0 \end{cases} \quad (10)$$

The above-estimated probability density function (pdf) findings that the survival times of Malignant Gliomas patients data follows 3p-General Pareto Probability Distribution can ensure efficient and accurate analysis of their survival times. Given the pdf in Equation (10), we find the cumulative density function (CDF) by taking the integral of the pdf with respect to the random variable t , given by:

$$\begin{aligned} F_T(t) &= P(t \leq T) \\ F_T(t; \mu, \sigma, \xi) &= \int_0^t f_T(s | \mu, \sigma, \xi) ds \\ &= \int_0^t \frac{1}{\sigma} \left(1 + \frac{\xi(x-\mu)}{\sigma} \right)^{-1/\xi-1} ds \end{aligned} \quad (11)$$

To solve this integral, you can use the substitution method.

Let $u = 1 + \frac{\xi(s-\mu)}{\sigma}$, then $du = \frac{\xi}{\sigma} ds$, which implies $ds = \frac{\sigma}{\xi} du$.

Also, when $s = 0$, $u = 1$, and when $s = t$, $u = 1 + \frac{\xi(t-\mu)}{\sigma}$.

Substituting these into the integral, we get:

$$\begin{aligned} F_T(t; \mu, \sigma, \xi) &= \int_1^{1+\frac{\xi(t-\mu)}{\sigma}} \frac{1}{\sigma} u^{-1/\xi-1} \left(\frac{\sigma}{\xi} \right) du \\ &= \frac{1}{\xi} \int_1^{1+\frac{\xi(t-\mu)}{\sigma}} u^{-1/\xi-1} du \\ &= \frac{1}{\xi} \left[\frac{u^{-1/\xi}}{-1/\xi} \right]_1^{1+\frac{\xi(t-\mu)}{\sigma}} \\ &= - \left[u^{-1/\xi} \right]_1^{1+\frac{\xi(t-\mu)}{\sigma}} \\ &= - \left(1 + \frac{\xi(t-\mu)}{\sigma} \right)^{-1/\xi} + 1 \end{aligned} \quad (12)$$

Therefore, the cumulative distribution function (CDF) of the Generalized Pareto Distribution is given by:

$$F_T(t; \mu, \sigma, \xi) = 1 - \left(1 + \frac{\xi(t - \mu)}{\sigma} \right)^{-1/\xi} \quad (13)$$

Substituting the parameter estimates given in **Table 5**, into Equation (13), then the estimated cumulative distribution function (CDF) of a 3-parameter Generalized Pareto Distribution is given by:

$$F_T(t; \mu, \sigma, \xi) = 1 - \left(1 + \frac{-0.2918(t - 0.6618)}{14.395} \right)^{-1/(-0.2918)} \quad (14)$$

This CDF is useful in determining that the probability of a given random observation (survival time, t) of a patient would be less than or equal to some value T . In other words, we can estimate the probability or likelihood of a patient with malignant gliomas surviving up to time t . **Figure 14** shows the Cumulative Distribution Function (CDF) plot for the survival times of Malignant Gliomas patients data. By examining this plot, we can estimate that the probability that a patient with Malignant Gliomas survives up to time $t = 12$ (1 year) and 25 months (2 years, 1 month) is approximately 0.6 (60%) and 0.9 (90%), respectively.

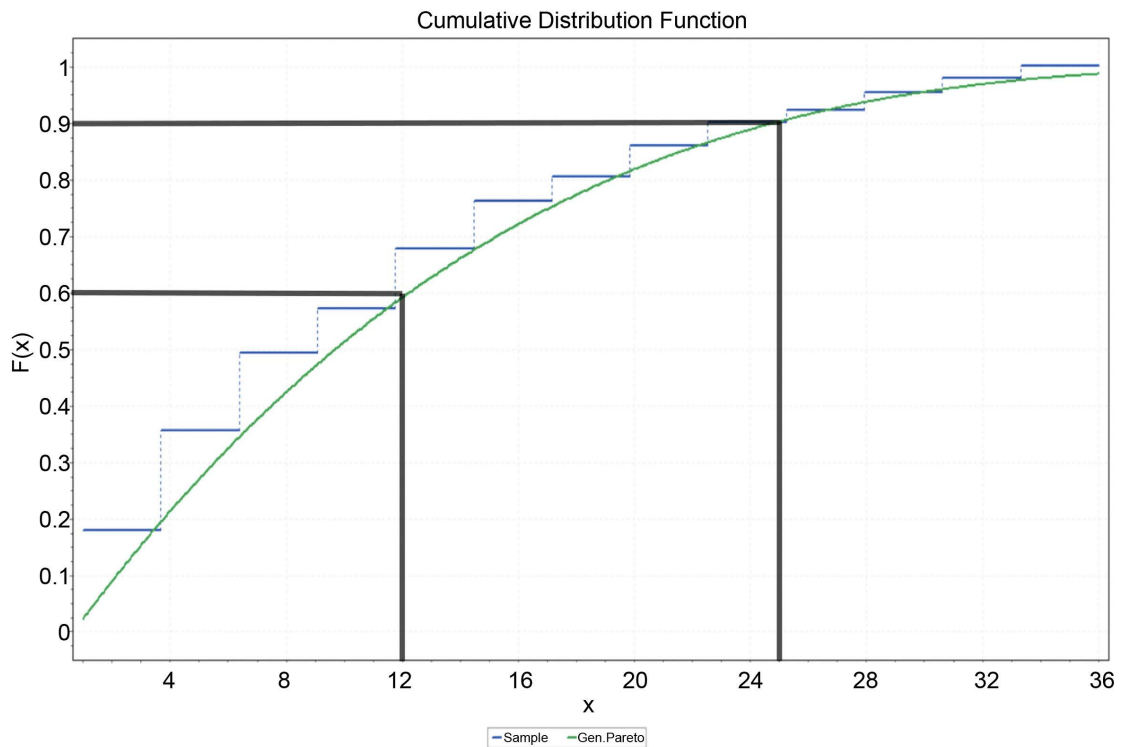


Figure 14. Cumulative distribution function plot for the survival time of MG.

Now, given the Cumulative Distribution Function (CDF), of 3-parameter Generalized Pareto Distribution in Equation (14), we obtained the survival function $\hat{S}(t)$ of the 3P-parameter Generalized Pareto Distribution given by

$$\begin{aligned}
 \hat{S}(t) &= P(t \geq T) \\
 &= 1 - F_T(t) \\
 &= 1 - F_T(t; \mu, \sigma, \xi) \\
 &= 1 - \left[1 - \left(1 + \frac{\xi(t - \mu)}{\sigma} \right)^{-1/\xi} \right]
 \end{aligned}
 \tag{15}$$

Substituting the parameter estimates given in **Table 5**, into Equation (15), then the estimated survival function, $\hat{S}(t)$ of a 3P-parameter Generalized Pareto Distribution is given by:

$$\hat{S}(t) = 1 - \left[1 - \left(1 + \frac{-0.2918(t - 0.6618)}{14.395} \right)^{-1/(-0.2918)} \right]
 \tag{16}$$

The survival function of 3P-parameter Generalized Pareto Distribution can be used to estimate the probability that a patient diagnosed with malignant glioma would survive beyond a specified time T ; thus, $P(t > T)$.

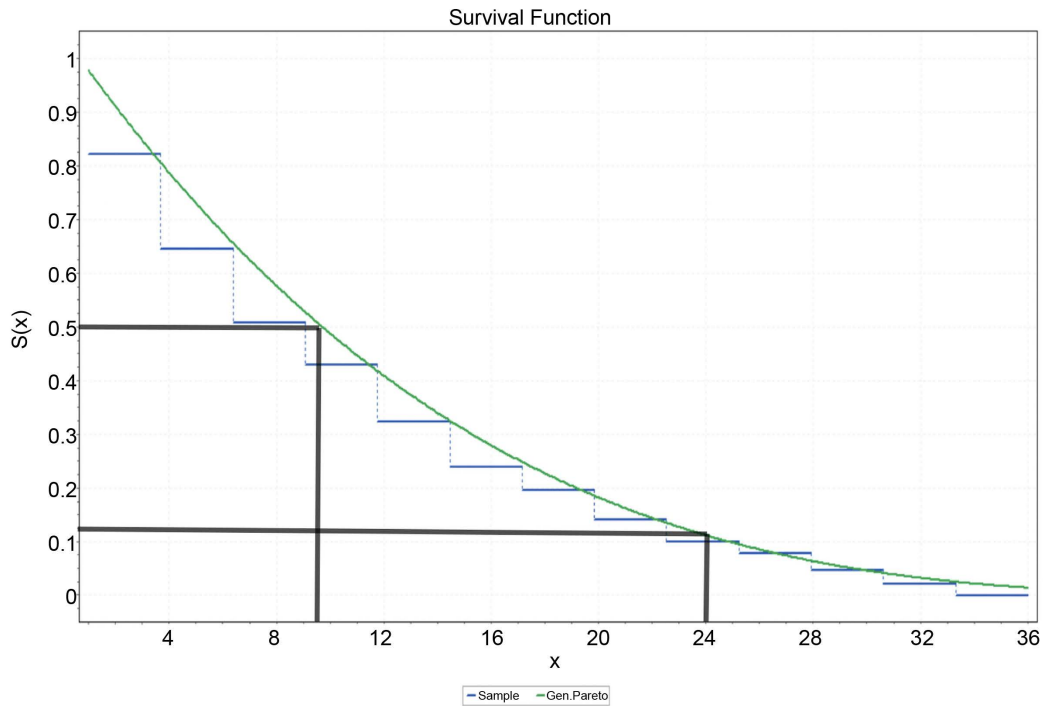


Figure 15. Survival estimate for the survival time of malignant glioma patients.

Figure 15 displays the estimate of the survival function $\hat{S}(t)$ of the survival times of malignant glioma patients. As expected, we can see that the survival function is decreasing and approaching approximately zero beyond $t = 36$ months. Thus, no individual is likely to survive beyond this time. In other estimates, the probability that a patient survives beyond 9 to 10 months and beyond 24 months is approximately 50% and 12.5%, respectively.

Given that age was identified as the most significant risk factor in predicting

the survival time of glioma patients, we used this distribution and its estimated parameters to plot the survival probabilities of glioma patients categorized by age groups. This approach can also be applied to other significant risk factors to analyze their impact on the survival probabilities and hazard rates of the patients. **Figure 16** shows smoothed survival curves for glioma patients categorized by age group, compared against a baseline curve, where each curve indicates the probability of survival over time for different age groups with survival probabilities generated using the 3-parameter Generalized Pareto distribution.

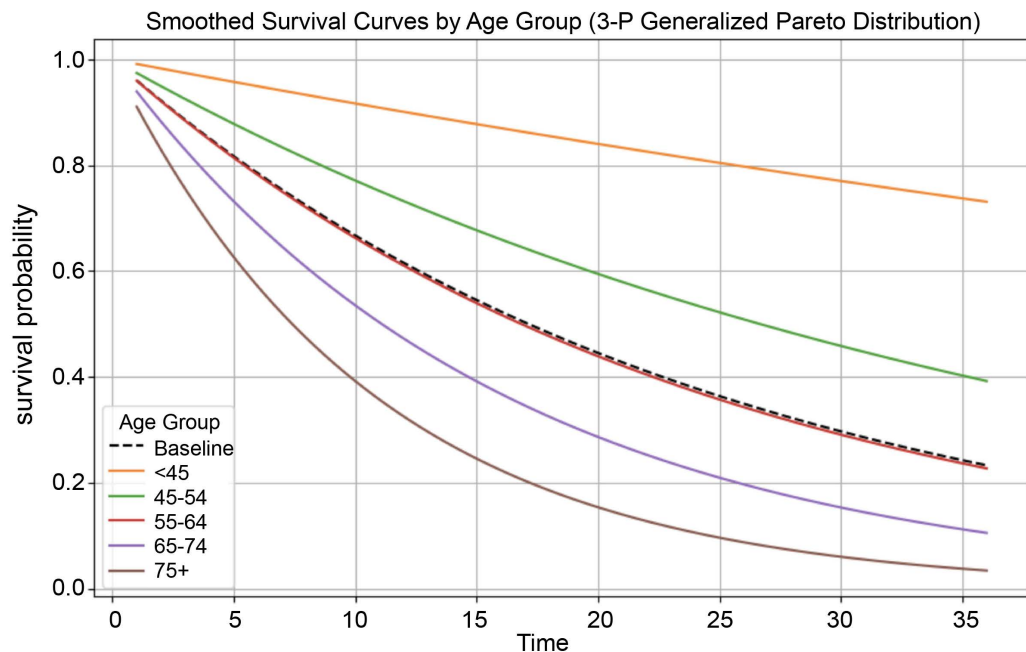


Figure 16. Smoothed survival curves for glioma patients categorized by age group.

From the plot, it is evident that younger glioma patients (<45) have the best survival rates, which gradually decrease with age. This trend is consistent with medical literature, where younger patients generally have better overall health, fewer comorbidities, and a better ability to tolerate aggressive treatments. Older patients, especially those over 75, have the poorest survival rates, likely due to reduced physiological resilience, more aggressive tumor biology, and the presence of other health conditions that can complicate treatment. The 45 - 54 age group has a lower survival probability compared to the under-45 group but still shows relatively better survival rates. The 55 - 64 age group shows a further decrease in survival probability, indicating a decline in survival rates as age increases. 65 - 74 age group has a significantly lower survival probability, reflecting the impact of advancing age on survival. It also highlights the importance of age-specific approaches in the management and treatment planning for glioma patients. The use of the 3-parameter Generalized Pareto distribution in generating these survival probabilities provides a robust method for modeling the extreme values and tail behavior of survival times across different age groups.

3.6. Comparison of the 3-Parameter Generalized Pareto Probability Distribution with the Kaplan Meier Estimation of the Survival Function

Parametric survival analysis revealed that the survival times of patients with malignant gliomas follow the three-parameter Generalized Pareto Probability Distribution. In parallel, we performed a non-parametric analysis using the Kaplan-Meier estimator to determine the survival probability of these patients. We then compared the survival probability estimates from the 3-parameter Generalized Pareto distribution with those from the Kaplan-Meier estimator. The Kaplan-Meier (KM) estimate, introduced by Edward L. Kaplan and Paul Meier in 1958, is a non-parametric method used to estimate survival probabilities over time. The KM estimate, often referred to as the product-limit estimator, calculates the probability of survivorship at a given time, known as survival time. Kaplan-Meier (KM) estimator is widely used in survival analysis for assessing recovery rates, death probabilities, and treatment effectiveness due to its ease of use in estimation, however it falls short compared to parametric survival analysis in decision-making power [36].

Table 6. Kaplan-Meier ($\hat{S}_{KM}(t)$) vs parametric (3-Parameter generalized pareto, $\hat{S}_{GP}(t)$) estimates of survival probabilities.

t_i	$\hat{S}_{KM}(t)$	$\hat{S}_{GP}(t)$	t_i	$\hat{S}_{KM}(t)$	$\hat{S}_{GP}(t)$
1.00	0.977	0.986	19.00	0.433	0.203
2.00	0.941	0.946	20.00	0.418	0.198
3.00	0.897	0.852	21.00	0.398	0.186
4.00	0.859	0.804	22.00	0.388	0.158
5.00	0.816	0.750	23.00	0.373	0.137
6.00	0.783	0.658	24.00	0.360	0.125
7.00	0.748	0.621	25.00	0.349	0.103
8.00	0.717	0.585	26.00	0.342	0.093
9.00	0.687	0.557	27.00	0.329	0.089
10.00	0.663	0.489	28.00	0.323	0.083
11.00	0.635	0.429	29.00	0.313	0.063
12.00	0.607	0.402	30.00	0.304	0.061
13.00	0.577	0.397	31.00	0.299	0.039
14.00	0.550	0.354	32.00	0.286	0.034
15.00	0.522	0.329	33.00	0.282	0.028
16.00	0.496	0.296	34.00	0.275	0.025
17.00	0.473	0.264	35.00	0.272	0.022
18.00	0.454	0.224			

Table 6 presents the survival times and corresponding survival probabilities from both methods. The probability estimates from the 3-parameter Generalized Pareto survival function are generally lower compared to those from the Kaplan-Meier estimator. At earlier time points, the two methods produce similar survival probabilities, indicating strong agreement in regions with abundant observed events. As time increases, systematic divergence is observed, with the parametric GPD yielding lower survival probabilities than the Kaplan-Meier estimator. This divergence reflects fundamental methodological differences between the two approaches. The Kaplan-Meier estimator provides an empirical, stepwise estimate of the survival function based solely on observed event times and does not impose distributional assumptions. In contrast, the 3-parametric GPD model yields smooth survival estimates based on an assumed distributional form, allowing for more stable estimation and tail extrapolation. Therefore, it should be prioritized as the primary alternative for analyzing cancer survivorship data. The Non-parametric methods are only strongly recommended when there is no available parametric probability distribution for the dataset and are particularly noticeable with limited sample sizes. Accordingly, the parametric GPD model provides a structured representation of long-term survival trends under validated assumptions, while the Kaplan-Meier estimator remains an important non-parametric benchmark for descriptive survival analysis. The two methods should therefore be viewed as complementary rather than hierarchical, with their use guided by the analytical objective and the underlying data characteristics.

4. Conclusions

The endeavor to enhance the survival time, or prolong life, of patients with malignant glioma (MG) has prompted the adoption of various research approaches and methodologies. In this study, we built supervised machine learning ANN-based survival analysis models to predict the survival times of malignant glioma patients. We aim to answer the question of how long a patient has to live by predicting whether patients will survive up to 12 months (1 year), between 12 to 24 months (up to 2 years), or beyond 24 months (2 years). The proposed ANN model can identify the top 10 significant risk factors in predicting the survival times of the patients.

We identified a well-defined probability distribution that characterizes the survival times of malignant glioma patients. Subsequently, we derived its cumulative distribution function (CDF) and survival function, which provided estimates of the probability of patient survival. Additionally, we estimated the proportion of survival times using the commonly employed Kaplan-Meier (KM) estimator for survivorship analysis. We then compared and contrasted these results with the survival probability estimates obtained from our identified distribution. This information is invaluable to healthcare professionals and the families of patients, addressing questions about the chances of survival of the patient up to or beyond a certain specified survival time t , post-diagnosis. Specifically, we answer the ques-

tion of the probability of patient survival up to or beyond time t . Considering that age is the most significant risk factor influencing survival predictions, we analyzed the survival probabilities of the survival times across different age groups.

This study is limited by class imbalance in the outcome variable. To preserve real-world prevalence, no data balancing techniques were applied, which may have contributed to uneven error rates, particularly a higher number of false negatives. Future research may evaluate cost-sensitive or resampling approaches to enhance minority-class detection.

Our proposed ANN-based survival analysis models serve as efficient alternatives to conventional survival analysis models, offering enhanced predictive power. The identification of the 3-Parameter Generalized Pareto Probability Distribution stands out as superior, providing a higher degree of accuracy in estimating the survival probability of patients diagnosed with malignant glioma compared to the non-parametric Kaplan-Meier method.

Acknowledgements

Sincere thanks to the members of the research team, Prof. Chris P. Tsokos, a mentor/advisor, for his professional guide and special thanks to managing editor *Hellen XU* for a rare attitude of high quality.

Funding

This research is protected by the University of South Florida TTO.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] American Cancer Society (2023) Cancer Statistics Center. Explore Cancer Statistics. <https://cancerstatisticscenter.cancer.org/#/>
- [2] National Cancer Institute. SEER Cancer Stat Facts: Brain and Other Nervous System Cancer. <https://seer.cancer.gov/statfacts/html/brain.html>
- [3] Li, X.Y. (2018) Investigate the Impact of Traumatic Brain Injury (TBI) on Glia Neighborhoods of Neuron.
- [4] Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., *et al.* (2021) The 2021 WHO Classification of Tumors of the Central Nervous System: A Summary. *Neuro-Oncology*, **23**, 1231-1251. <https://doi.org/10.1093/neuonc/noab106>
- [5] Reinhardt, A., Stichel, D., Schrimpf, D., Sahm, F., Korshunov, A., Reuss, D.E., *et al.* (2018) Anaplastic Astrocytoma with Piloid Features, a Novel Molecular Class of IDH Wildtype Glioma with Recurrent MAPK Pathway, CDKN2A/B and ATRX Alterations. *Acta Neuropathologica*, **136**, 273-291. <https://doi.org/10.1007/s00401-018-1837-8>
- [6] American Cancer Society (2021) Cancer Statistics Center. <http://cancerstatisticscenter.cancer.org>
- [7] Louis, D.N., Aldape, K., Brat, D.J., Capper, D., Ellison, D.W., Hawkins, C., *et al.*

- (2016) cIMPACT-NOW (the Consortium to Inform Molecular and Practical Approaches to CNS Tumor Taxonomy): A New Initiative in Advancing Nervous System Tumor Classification. *Brain Pathology*, **27**, 851-852. <https://doi.org/10.1111/bpa.12457>
- [8] Feigl, P. and Zelen, M. (1965) Estimation of Exponential Survival Probabilities with Concomitant Information. *Biometrics*, **21**, 826-838. <https://doi.org/10.2307/2528247>
- [9] Xu, Y. and Tsokos, C.P. (2012) Probabilistic Survival Analysis Methods Using Simulation and Cancer Data. *Problems of Nonlinear Analysis in Engineering Systems*, **18**, 47-59.
- [10] Xu, Y., Kepner, J. and Tsokos, C.P. (2011) Identify Attributable Variables and Interactions in Breast Cancer. *Journal of Applied Sciences*, **11**, 1033-1038. <https://doi.org/10.3923/jas.2011.1033.1038>
- [11] Data Sources (2022) Surveillance, Epidemiology, and End Results (SEER) 17 Registries. National Cancer Institute.
- [12] Goodfellow, I., Bengio, Y. and Courville, A. (2016) Deep Learning. MIT Press.
- [13] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) Imagenet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, **25**.
- [14] Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K. and Kuksa, P. (2011) Natural Language Processing (Almost) from Scratch. *Journal of Machine Learning Research*, **12**, 2493-2537.
- [15] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D. and Riedmiller, M. (2013) Playing Atari with Deep Reinforcement Learning. arXiv preprint arXiv:1312.5602.
- [16] Faraggi, D. and Simon, R. (1995) A Neural Network Model for Survival Data. *Statistics in Medicine*, **14**, 73-82. <https://doi.org/10.1002/sim.4780140108>
- [17] Biganzoli, E., Boracchi, P., Mariani, L. and Marubini, E. (1998) Feed Forward Neural Networks for the Analysis of Censored Survival Data: A Partial Logistic Regression Approach. *Statistics in Medicine*, **17**, 1169-1186. [https://doi.org/10.1002/\(sici\)1097-0258\(19980530\)17:10<1169::aid-sim796>3.0.co;2-d](https://doi.org/10.1002/(sici)1097-0258(19980530)17:10<1169::aid-sim796>3.0.co;2-d)
- [18] Lisboa, P.J., Taktak, A.F.G., Santos, J.M. and Ratanachina, A. (2003) A Neural Network Approach for Survival Analysis. *Proceedings of the 2003 IEEE International Conference on Neural Networks and Signal Processing*, 222-225.
- [19] Ching, T., Zhu, X. and Garmire, L.X. (2018) Cox-nnet: An Artificial Neural Network Method for Prognosis Prediction of High-Throughput Omics Data. *PLOS Computational Biology*, **14**, e1006076. <https://doi.org/10.1371/journal.pcbi.1006076>
- [20] Luck, M., Sylvain, T., Cardinal, H., Lodi, A. and Bengio, Y. (2017) Deep Learning for Survival Analysis. *Proceedings of the 2nd Machine Learning for Healthcare Conference*, 182-197.
- [21] Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T. and Kluger, Y. (2018) DeepSurv: Personalized Treatment Recommender System Using a Cox Proportional Hazards Deep Neural Network. *BMC Medical Research Methodology*, **18**, Article No. 24. <https://doi.org/10.1186/s12874-018-0482-1>
- [22] Lee, C., Zame, W.R., Yoon, J. and van der Schaar, M. (2018) Dynamic-DeepHit: A Deep Learning Approach for Dynamic Survival Analysis with Competing Risks. *Proceedings of the 35th International Conference on Machine Learning*, 2885-2894.
- [23] Nielsen, M. (2015) Neural Networks and Deep Learning. Determination Press.

- [24] Lyons, P.G., Mathis, M.R. and Iwashyna, T.J. (2021) Predictions with a Purpose: Elevating Standards for Clinical Prediction Models. *Annals of the American Thoracic Society*, **18**, 1616-1619.
- [25] Sidebotham, D. and Barlow, C.J. (2022) False-Positive and False-Negative Risks for Individual Multicentre Trials in Critical Care. *BJA Open*, **1**, 100003. <https://doi.org/10.1016/j.bjao.2022.100003>
- [26] Ling, C.X. and Sheng, V.S. (2011) Cost-Sensitive Learning and the Class Imbalance Problem. In: Sammut, C. and Webb, G.I., Eds., *Encyclopedia of Machine Learning*, Springer, 231-235.
- [27] Scott, J.G., Bauchet, L., Fraum, T.J., *et al.* (2012) Recursive Partitioning Analysis of Prognostic Factors for Glioblastoma Patients Aged 70 Years or Older. *Journal of Clinical Oncology*, **30**, 1195-1201.
- [28] Ostrom, Q.T., Bauchet, L., Davis, F.G., Deltour, I., Fisher, J.L., Langer, C.E., *et al.* (2014) The Epidemiology of Glioma in Adults: A “State of the Science” Review. *Neuro-Oncology*, **16**, 896-913. <https://doi.org/10.1093/neuonc/nou087>
- [29] Blumenthal, D.T., Gorlia, T., Gilbert, M.R., *et al.* (2014) Is More Better? The Impact of Time from Surgery to Radiotherapy on Survival in Glioblastoma. *Neuro-Oncology*, **16**, 353-361.
- [30] Louis, D.N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W.K., *et al.* (2016) The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A Summary. *Acta Neuropathologica*, **131**, 803-820. <https://doi.org/10.1007/s00401-016-1545-1>
- [31] van den Bent, M.J., Carpentier, A.F., Brandes, A.A., *et al.* (2005) Adjuvant Chemotherapy for Newly Diagnosed Anaplastic Oligodendroglioma: Long-Term Follow-Up of EORTC Trial 26951. *Journal of Clinical Oncology*, **23**, 8146-8156.
- [32] Lacroix, M., Abi-Said, D., Fournay, D.R., Gokaslan, Z.L., Shi, W., DeMonte, F., *et al.* (2001) A Multivariate Analysis of 416 Patients with Glioblastoma Multiforme: Prognosis, Extent of Resection, and Survival. *Journal of Neurosurgery*, **95**, 190-198. <https://doi.org/10.3171/jns.2001.95.2.0190>
- [33] Peto, R. and Peto, J. (1972) Asymptotically Efficient Rank Invariant Test Procedures. *Journal of the Royal Statistical Society. Series A (General)*, **135**, 185-207. <https://doi.org/10.2307/2344317>
- [34] Chambers, R.L., Steel, D.G., Wang, S. and Welsh, A. (2012). Maximum Likelihood Estimation for Sample Surveys. CRC Press. <https://doi.org/10.1201/b12038>
- [35] Calitz, F. (1973) Maximum Likelihood Estimation of the Parameters of the Three-Parameter Lognormal Distribution—A Reconsideration. *Australian Journal of Statistics*, **15**, 185-190. <https://doi.org/10.1111/j.1467-842x.1973.tb00199.x>
- [36] Kaplan, E.L. and Meier, P. (1958) Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, **53**, 457-481. <https://doi.org/10.1080/01621459.1958.10501452>