

Classification of Brain Metastases Prognostic Groups Utilizing Artificial Neural Network Approaches

George Rodrigues ¹, Glenn Bauman ², Ben J. Slotman ³, Frank Lagerwaard ⁴

1. Department of Radiation Oncology, London Regional Cancer Program, London, Ontario, CA; Schulich School of Medicine & Dentistry, Western University, London, Ontario, CA 2. Department of Radiation Oncology, London Regional Cancer Program, London Health Sciences Centre, London, ON, CANADA 3. Department of Radiation Oncology, VU University Medical Center 4. Radiation Oncology Department, VU University Medical Center

✉ **Corresponding author:** George Rodrigues, george.rodrigues@lhsc.on.ca
Disclosures can be found in Additional Information at the end of the article

Abstract

Objective: The purpose of this investigation is to explore the performance of an artificial neural network (ANN) based prognostic index compared to traditional logistic regression (LR) modeling and other published prognostic indices (PI) in classifying survival among patients with brain metastases treated with stereotactic radiotherapy.

Methods: A database of 460 patients having received either stereotactic radiosurgery or fractionated stereotactic radiation therapy brain radiotherapy was utilized and divided into three sub-databases for ANN/LR analysis: a testing dataset, n=276 (65%); a cross-validation dataset for training, n=69 (15%); and a validation dataset, n=115 (25%). The primary endpoint of survival was classified into one of three categories: unfavorable survival (<two months), intermediate survival (two to six months) and favorable survival (>six months) endpoint classifications. ANNs were optimized in terms of model structure, complexity, and a cost optimization algorithm and then compared to both LR and published PIs in terms of classification accuracy (CA) and total major misclassification rates (TMMR) according to the three category survival scheme.

Results: CA and TMMR for the nine published PIs for the total database (n=460) ranged from 34-53% and 4-11%, respectively. Both the LR and ANN approaches (in the validation database) were over 10% superior to the best existing PI system in terms of CA (LR/ANN 62.6%, published prognostic indices 27-49%) with a similar rate of TMMR (LR 7.8%, ANN 6.1%, published prognostic indices 2-17%).

Conclusions: While a modest improvement over published PI was noted, use of various ANN model structures, nodal complexity, and cost function optimization algorithms did not lead to a significant improvement in survival classification when compared to LR.

Received 05/13/2013
Review began 05/13/2013
Published 05/14/2013

© Copyright 2013

Rodrigues et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 3.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Categories: Radiation Oncology, Epidemiology/Public Health

Keywords: radiosurgery, artificial neural network, classification, brain metastasis, radiation oncology, prognostic index

Introduction

How to cite this article

Rodrigues G, Bauman G, Slotman B J., et al. (May 14, 2013) Classification of Brain Metastases Prognostic Groups Utilizing Artificial Neural Network Approaches. Cureus 5(5): e115. DOI 10.7759/cureus.115

Validated prognostic factors and indices can be used to assist the clinician in patient counseling and treatment decision-making. Additionally, such indices can support the conduct of prospective clinical trials by defining patient eligibility and stratification criteria. Multiple prognostic factors have been shown to be related to patient survival in the context of brain metastases which include: performance status, extracranial disease, age, controlled primary, primary site, interval between primary disease and brain metastases, number/volume of brain metastases, and clinical response to steroids [1-11]. The Radiation Therapy Oncology Group (RTOG) Recursive Partitioning Analysis (RPA) brain metastases prognostic index is the oldest system currently in use [2, 12-16]. However, the utility of the system has been limited by the large relative proportion of patients within the intermediate-risk group, as has been previously highlighted by several investigators [17-18].

Other systems have been subsequently developed using different combinations of the previously listed prognostic factors. These include the Score Index for Radiosurgery – SIR [4], Rotterdam scale – RDAM [3], Basic Score for Brain Metastases – BSBM [5], Golden Grading System - GGS [8], Graded Prognostic Assessment – GPA [6, 10], Disease Specific GPA - DS-GPA [9] and the German I and II scales developed by Rades et al. [7, 11]. A recent systematic review of all published systems was not able to definitively identify a superior system [1]; however, a recent neural network analysis suggested that the newly developed RTOG Graded Prognostic Assessment (GPA) system may have some advantage in prognostic utility in the context of WBRT patient populations [19].

Creation and assessment of prognostic factors and indices traditionally involves the use of Cox proportional hazards and logistic regression modeling approaches for survival time and categorical event endpoints, respectively. Artificial neural networks (ANN) consist of a set of multivariable approaches that mimic networks of biological neurons [20]. Layers of nodes (input, hidden, and output) are interconnected with weighed connection lines (connecting all nodes from one layer to another) in order to form the non-linear computational structure of the ANN (Figure 1). During ANN training with patient data (a training dataset that includes prognostic factors and study endpoints), the initial weights of the connection lines that input into the various nodes are altered using one of many available cost-function optimization procedures against a separate cross-validation dataset. Once training is completed, the trained ANN can be assessed using a third independent validation dataset in order to report on ANN prognostic ability (Figure 2).

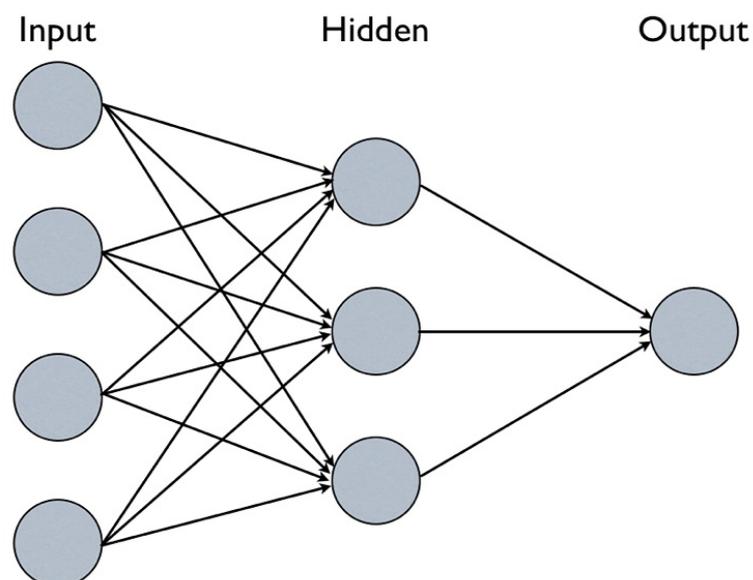


FIGURE 1: Single Layer Perceptron Artificial Neural Network Example Demonstrating Input, Hidden, and Output Nodes and Node Interconnections

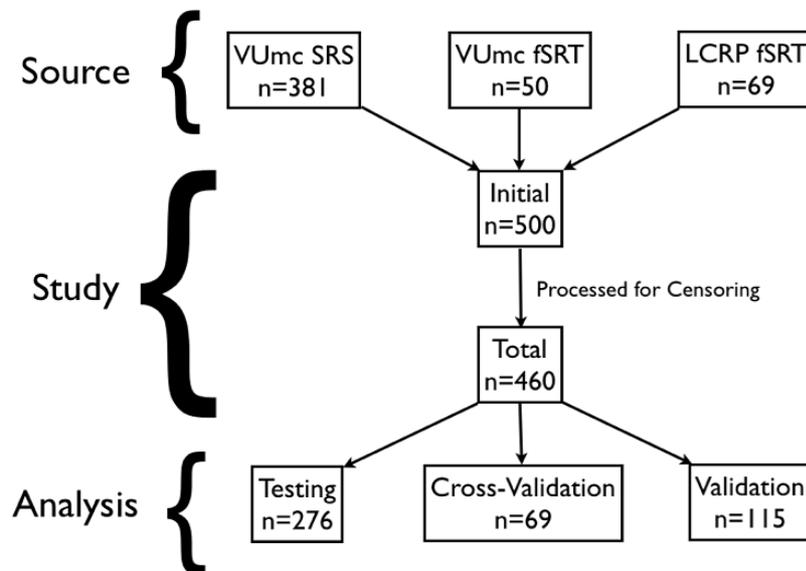


FIGURE 2: Study Databases

Study Databases. VUmc = VU University Medical Centre, LCRP = London Regional Cancer Program, SRS = stereotactic radiosurgery, fSRT = fractionated stereotactic radiation therapy.

ANN approaches have three main advantages over other techniques including: [1] not requiring an assumption of proportional hazard risk to baseline, [2] utilization of non-linear (and non-parametric) associations, and [3] all interactions between input factors and hidden layers are allowed (i.e. model complexity). Despite these advantages, several disadvantages with the use of ANNs have been noted including: the possible over-fitting of testing datasets leading to inferior validation predictive power, the complexity of ANN models which may hinder translation into clinical applications, and non-robust weighting cost-function optimization leading to local minima solutions. ANN approaches have been reported in the medical literature in the prediction of a variety of endpoints [20-23].

Given the lack of an ideal validated prognostic index to reliably and accurately classifying patients into distinct prognostic groups relevant to clinical decision-making, the purpose of this investigation is to explore the utility of an ANN-based index compared to traditional logistic regression modeling. Both ANN and logistic regression models are compared to published prognostic indices in order to assess the potential improvement in patient classification into distinct prognostic groups.

Materials And Methods

Source databases

A retrospective review was performed on two institutional databases of 500 patients diagnosed with oligometastatic brain metastatic disease. Patients received either stereotactic radiosurgery (SRS, n=381 with one to three brain metastases) or fractionated stereotactic radiation therapy (fSRT, n=119, one to six brain metastases) between 2002 and 2011. This database contained pretreatment information (including derived risk stratification categories for all nine published systems), treatment details, and outcome information, including the primary endpoint of overall survival. Patients were treated at one of two cancer sites: London Regional Cancer Program (LRCP, London, ON, n=69 fSRT patients) or at VU Medical Centre (VUMC, Amsterdam, The Netherlands, n=381 SRS patients, plus n=50 fSRT patients). Institutional ethics approval was obtained for this joint database analysis.

Full treatment details for both SRS and fSRT approaches have been extensively published in the medical literature [24-27]. Pooling of fSRT and SRS into a joint database for this investigation was performed after confirming that treatment assignment (fSRT vs. SRS) was not a significant predictor for overall survival. This analysis was performed in the context of a propensity-score matched pair analysis (accepted for publication in *Radiotherapy and Oncology*).

Endpoint definitions and ANN study databases

The source databases described above were merged into a common database for further statistical analyses. Individual survival time was recoded into one of three discrete categories, as suggested by Nieder, et al. [18], as no other internationally recognized prognostic categories exist in the medical literature. These three categories consisted of: unfavorable survival (< two months); intermediate survival (two to six months); and favorable survival (> six months) endpoint classifications. In order to avoid the issue of incomplete follow-up of censored patients with less than six months of follow-up, all patients (whether censored or not) treated in the six months prior to the final database update were removed from the database (n=460 patients remaining). These remaining patients were then randomly assigned into one of three study databases (testing dataset n=276 (65%), cross-validation dataset for ANN training n=69 (15%), and an ANN validation dataset n=115 (25%).

Control analyses

Descriptive and operating characteristic (OC) statistics (Figure 3) were calculated for each of the published prognostic indices for each of the three databases (testing, cross-validation, and validation) and the complete study database (n=460). Most prognostic indices could be calculated for all study patients, except for the DS-GPA 412/460 (90%) and the RDM 362/460 (79%) due to missing information from the database (non DS-GPA tumor sites and steroid response information). Missing information was found to be distributed relatively equally between all study databases. The first calculated OC statistic was prognostic index accuracy (rate of correct classification equals high-risk category predicting survival < two months, plus intermediate-risk category/categories predicting survival two to six months, plus low-risk classification predicting survival > six months, Figure 3). The second OC statistic was the total major misclassification rate defined as proportion of all patients that are misclassified into the opposite survival group (i.e. high-risk category patients surviving greater than six months, plus low-risk category patients surviving less than two months, Figure 3).

		Survival Classification			
		< 2 months (unfavorable)	2-6 months (intermediate)	> 6 months (favorable)	
Prognostic Index Classification	high-risk	a	b	c	Classification Accuracy: $(a+e+i) / \sum a-i$ Total Major Misclassification Rate: $(c+g) / \sum a-i$
	intermediate-risk	d	e	f	
	low-risk	g	h	i	

FIGURE 3: Confusion Matrix for Calculation of Classification Accuracy and Total Major Misclassification Rate

A backward elimination logistic regression (LR) analysis using the constituent prognostic factors of the published prognostic indices (primary class/site, presence of systematic metastases, performance status, age, interval between primary diagnosis and brain metastases presentation, volume/number brain metastases, and active primary) as input variables to predict for categorical survival group (favorable, intermediate, and unfavorable survival endpoint groups) was performed using NeuroSolutions 6.1 modeling software (NeuroDimension Inc., Gainesville FL, USA). Both classification accuracy and total major misclassification rate for the logistic regression were calculated for all three study databases (testing, cross-validation, and validation) and for the total study database.

Artificial neural network analyses

Artificial Neural Network (ANN) analyses were performed using the Express Builder functionality of the NeuroSolutions software in which multiple ANN constructs can be assessed in parallel to identify superior ANN approaches. ANN analysis differs from traditional regression approaches given the fact that multiple ANN structures and optimization approaches need to be explored in order find the ideal ANN solution. All steps described herein will utilize the classification accuracy metric to adjudicate superior ANN (or LR) strategies. Receiver operator curves were considered not to be appropriate for this analysis given the classification of a three-category endpoint. The investigation of ANN approaches was defined as an a-priori sequence of steps, which are described below:

Step 1 – ANN Structure Assessment: Three unique ANN structures (One-layer Perceptron, Two-layer Perceptron, and Probabilistic Neural Network) were assessed against the LR approach to assess the ideal ANN structure for further node structure and cost optimization algorithm. These three structures were assessed due to their potential utility in the assessment of smaller datasets (<1000 data-points). In addition to the more familiar perceptron ANN approaches, the Probabilistic Neural Network (PNN) approach was also used given the known advantage of efficient training of smaller datasets. The one-layer perceptron utilized nine input nodes, five hidden nodes (to avoid over-fitting), and one output node with Levenberg-

Marquardt (LM) optimization (MLP1-951-LM). The two-layer perceptron utilized nine input nodes, six and then three hidden nodes, and one output node with LM optimization (MLP2-9631-LM). The PNN used a knowledge map approach with nine input nodes, 235 internal nodes, and one output node. Perceptron-based and PNN optimization was limited to 100 iterations and three epochs, respectively.

Step 2 – ANN Node Alteration: After selection of the best performing Step 1 ANN based on the classification accuracy of all ANNs tested, the nodal structure of the ANN in question was altered in order to explore whether or not the addition or subtraction of nodes improve the classification accuracy of the ANN approach. For perceptron-based ANNs, the number of hidden nodes was changed to plus or minus two nodes (i.e. if five hidden nodes are used in step one and found to be ideal, a range of hidden nodes from three to seven will be tested in step two).

Step 3 – Cost Optimization Comparison: A series of available cost optimization algorithms were used to assess whether or not the choice of algorithm used can impact model performance. The following algorithms were assessed: Levenberg-Marquardt (LM), Conjugate Gradient (CG), Delta Bar Delta (DBD), Quickprop (QUICK), Step (STEP), and Momentum (MOM).

Step 4 – Final Optimization: This final step will be utilized if the ANN approach is found to be superior to the baseline LR approach. The final ANN will be subjected to both a rotating 10-fold cross validation to assess average ANN classification accuracy. If the ANN approach is not found to be superior to the baseline LR approach, then the step four analyses will not be performed.

Results

Database composition

Descriptive statistics for the total study database (n=460) as well as the three investigative databases (testing, cross-validation, and validation) are summarized in Supplementary Table 1.

Variable	Category	Total (n=460)
Data source		
	fSRT	98 (21.3%)
	SRS	362 (78.7%)
Systemic metastases		
	No	226 (49.1%)
	Yes	234 (50.9%)
Age	Mean	61.4 years
Interval from diagnosis to brain metastases treatment	Median	293 days
Volume of brain metastases		
	Mean	8.35 cc
	Range	0.03 - 151.5 cc

Number of brain metastases		
	Median	2
	Range	1-6
Primary active		
	No	239 (52.0%)
	Yes	221 (48.0%)
Tumor classification		
	Lung	263 (57.2%)
	Breast	48 (10.4%)
	Kidney	36 (7.8%)
	Melanoma	31 (6.7%)
	Colorectal	33 (7.2%)
	Ovary	16 (3.5%)
	Other	26 (5.7%)
	Esophageal	7 (1.5%)
WHO PS		
	0	94 (20.4%)
	1	266 (57.8%)
	2	88 (19.1%)
	3	11 (2.4%)
	4	1 (0.2%)
Death		
	No	56 (12.2%)
	Yes	404 (87.8%)
Survival	Median	211 days
Survival classification		
	Unfavorable (<2M)	75 (16%)
	Intermediate (2-6M)	254 (55%)
	Favorable (>6M)	131 (29%)

TABLE 1: Descriptive Statistics for the Three Study Databases and the Total

Database.

fSRT = Fractionated Stereotactic Radiation Therapy; SRS = Stereotactic Radiation Therapy; cc = cubic centimeters; WHO = World Health Organization; PS = Performance Status; M = months

In terms of the total database, the majority of patients were treated with SRS (78.7%). Mean age was 61.4 years with a median primary tumor to brain metastases diagnostic time interval of 293 days. Over 97% of patients had a World Health Organization performance status level of 0-2. Median number of brain metastases was two (range: one to six) with a mean total volume of 8.35 cc (range 0.03-151.5cc). Primary cancer was lung in 57.2% of cases with breast cancer being the second most commonly diagnosed cancer in 10.4%. The primary cancer was active at the time of brain metastases treatment in 52% of cases and other systemic metastases were present in 50.9%. Comparison of the total database with each of the smaller investigative databases did not highlight any significant difference between the groups.

Prognostic index classification for all published indices is summarized in [Supplementary Table 2](#).

Variable	Category	Total (n=460)
RTOG RPA Group [2]		
	1	100 (21.7%)
	2	286 (62.2%)
	3	74 (16.1%)
BSBM Group [4]		
	0	35 (7.6%)
	1	137 (29.8%)
	2	176 (38.3%)
	3	112 (24.4%)
SIR Group [3]		
	1	40 (8.7%)
	2	351 (76.3%)
	3	69 (15.0%)
GPA Group [6, 10]		
	1	108 (23.5%)
	2	301 (65.4%)
	3	32 (7.0%)
	4	19 (4.1%)

GGS Group [8]		
	0	122 (26.2%)
	1	194 (42.2%)
	2	116 (25.2%)
	3	28 (6.1%)
DSGPA Group [9] n=48 missing		
	1	87 (21.1%)
	2	235 (57.0%)
	3	47 (11.4%)
	4	213 (10.4%)
RDAM [3] n=98 missing		
	1	33 (9.1%)
	2	253 (69.9%)
	3	76 (21.0%)
RADES I Group [7]		
	1	50 (10.9%)
	2	80 (17.4%)
	3	204 (44.4%)
	4	126 (27.4%)
RADES II Group [11]		
	1	37 (8.0%)
	2	157 (34.1%)
	3	266 (57.8%)

TABLE 2: Prognostic Index Classification Distribution.

Higher category number reflects lower risk. RTOG RPA = Radiation Therapy Oncology Group Recursive Partitioning Analysis; BSBM = Basic Score for Brain Metastases; SIR = Score Index for Radiosurgery; GPA = Graded Prognostic Assessment; GGS = Golden Grading System; DSGPA = Disease-Specific GPA; RDAM = Rotterdam System; RADES I = First Rades et al. scale circa 2008; RADES II = Second Rades et al. scale circa 2011

As noted by previous investigators, the intermediate-risk category/categories contained the largest number of patients (34-77% with 8/9 indices >50%). In terms of the most commonly used RTOG RPA system, 100 patients (21.7%), 286 patients (62.2%), and 74 patients (16.1%)

were in the high-risk, intermediate-risk, and low-risk categories, respectively. A total of 404/460 (87.8%) of patients died during follow-up with a median actuarial survival for all patients of 211 days. Survival classification was unfavorable (< two months) in 75 patients (16%), intermediate (two to six months) in 254 patients (55%), and favorable (> six months) in 131 patients (29%).

Traditional prognostic index assessment

Classification accuracy (CA) and total major misclassification rates (TMMR) for all nine published prognostic indices are summarized in Table 3 for all study databases. In terms of the total database, classification accuracy ranged from 34% (DS-GPA) to 53% (RADES II). Total major misclassification rates ranged from 4% (RDAM and BSBM) to 11% (GPA).

Database	Statistic	RTOG RPA	BSBM	SIR Group	GPA Group	GGG Group	DS-GPA Group	RDAM	RADES I	RADES II
Testing (n=276)	CA	110/276 (40%)	121/276 (44%)	106/276 (38%)	94/276 (34%)	129/276 (47%)	89/244 (36.5%)	89/222 (40%)	124/276 (45%)	152/276 (55%)
	TMMR	17/276 (6%)	9/276 (3%)	10/276 (3.6%)	22/276 (8%)	13/276 (5%)	20/244 (8%)	6/222 (2.7%)	13/276 (4.7%)	20/276 (7%)
Cross-Validation (n=69)	CA	24/69 (35%)	23/69 (33%)	22/69 (32%)	22/69 (32%)	32/69 (46%)	19/62 (31%)	17/52 (33%)	26/69 (38%)	34/69 (49%)
	TMMR	7/69 (10%)	4/69 (6%)	8/69 (12%)	11/69 (16%)	6/69 (9%)	7/62 (11%)	3/52 (6%)	5/69 (7%)	9/69 (13%)
Validation (n=115)	CA	38/115 (33%)	46/115 (40%)	38/115 (33%)	31/115 (27%)	45/115 (39%)	32/106 (30%)	34/88 (39%)	49/115 (43%)	56/115 (49%)
	TMMR	13/115 (11%)	7/115 (6%)	6/115 (5%)	19/115 (17%)	2/115 (2%)	10/106 (9%)	4/88 (5%)	8/115 (7%)	11/115 (10%)
Total (n=460)	CA	172/460 (37%)	190/460 (41%)	166/460 (36%)	141/460 (31%)	206/460 (45%)	140/412 (34%)	140/362 (39%)	205/460 (45%)	242/460 (53%)
	TMMR	37/460 (8%)	20/460 (4%)	24/460 (5%)	51/460 (11%)	21/460 (5%)	37/412 (9%)	13/362 (4%)	26/460 (6%)	40/460 (9%)

TABLE 3: Classification Accuracy and Total Major Misclassification Rates for Published Prognostic Indices.

CA = Classification Accuracy; TMMR = Total Major Misclassification Rate; RTOG RPA = Radiation Therapy Oncology Group Recursive Partitioning Analysis; BSBM = Basic Score for Brain Metastases; SIR = Score Index for Radiosurgery; GPA = Graded Prognostic Assessment; GGS = Golden Grading System; DSGPA = Disease-Specific GPA; RDAM = Rotterdam System; RADES I = First Rades et al. scale circa 2008; RADES II = Second Rades et al. scale circa 2011

Artificial neural network assessment

The CA and TMMR (for the testing, cross-validation, and validation) metrics for all three steps of the ANN assessment is summarized in Table 4.

Optimization Step	Model	Testing (n=276)		Cross Validation (n=69)		Validation (n=115)	
		CA	TMMR	CA	TMMR	CA	TMMR
Step One	LR*	165 (59.8%)	23 (8.3%)	37 (53.6%)	8 (11.6%)	72 (62.6%)	9 (7.8%)
	MLP1-951-LM**	170 (61.6%)	19 (6.9%)	40 (58.0%)	6 (8.7%)	72 (62.6%)	7 (6.1%)
	PNN	186 (67.4%)	24 (8.7%)	39 (56.5%)	10 (14.5%)	68 (59.1%)	9 (7.8%)
	MLP2-9631-LM	156 (56.5%)	39 (14.1%)	37 (53.6%)	14 (20.3%)	70 (60.9%)	16 (13.9%)
Step Two	LR (Same as Step One LR)*	165 (59.8%)	23 (8.3%)	37 (53.6%)	8 (11.6%)	72 (62.6%)	9 (7.8%)
	MLP1-951-LM* **	170 (61.6%)	19 (6.9%)	40 (58.0%)	6 (8.7%)	72 (62.6%)	7 (6.1%)
	MLP1-931-LM	164 (59.4%)	16 (5.8%)	39 (56.5%)	8 (11.6%)	64 (55.7%)	10 (8.7%)
	MLP1-941-LM	168 (60.9%)	22 (8.0%)	40 (58.0%)	6 (8.7%)	65 (56.5%)	7 (6.1%)
	MLP1-961-LM	153 (55.4%)	35 (12.7%)	39 (56.5%)	10 (14.5%)	67 (58.3%)	18 (15.7%)
	MLP1-971-LM	189 (68.5%)	16 (5.8%)	38 (55.1%)	7 (10.1%)	60 (52.1%)	12 (10.4%)
Step Three	LR (Same as Step One LR)*	165 (59.8%)	23 (8.3%)	37 (53.6%)	8 (11.6%)	72 (62.6%)	9 (7.8%)
	MLP1-951-LM*	170 (61.6%)	19 (6.9%)	40 (58.0%)	6 (8.7%)	72 (62.6%)	7 (6.1%)
	MLP1-951-CG	165 (59.8%)	21 (7.6%)	39 (56.5%)	6 (8.7%)	69 (60.0%)	7 (6.1%)
	MLP1-951-DBD	168 (60.9%)	22 (8.0%)	38 (55.1%)	6 (8.7%)	70 (60.9%)	11 (9.6%)
	MLP1-951-QUICK	163 (59.1%)	23 (8.3%)	40 (58.0%)	6 (8.7%)	67 (58.3%)	12 (10.4%)
	MLP1-951-STEP	158 (57.3%)	20 (7.2%)	37 (53.6%)	7 (10.1%)	72 (62.6%)	10 (8.7%)
	MLP1-951-MOM	163 (59.1%)	21 (7.6%)	39 (56.5%)	6 (8.7%)	68 (59.1%)	10 (8.7%)

TABLE 4: Classification Accuracy and Total Major Misclassification Rate for Logistic Regression and Selected ANN Optimization.

CA = Classification Accuracy; TMMR = Total Major Misclassification Rate; LR = Logistic Regression; MLP1 = One Hidden Layer Multilayer Perceptron; MLP2 = Two Hidden Layers Multilayer Perceptron; PNN = Probabilistic Neural Network; LM = Levenberg Marquardt Optimization; STEP = Step Optimization; CG = Conjugate Gradient Optimization; DBD = Delta Bar Delta Optimization; QUICK = Quickprop Optimization; MOM = Momentum Optimization

The logistic regression model was found to have a CA and TMMR of 62.6% and 7.8%, respectively. The first step of the ANN optimization demonstrated that the one-layer perceptron ANN (MLP1-951-LM) had an identical CA (to the LR approach) of 62.6% with a slightly improved TMMR of 6.1%. Optimization procedures in step two (variable hidden nodes) and step three (cost optimization algorithm) did not further improve the CA and TMMR rates compared to LR. Step four was not conducted due to the fact that the final ANN approach (MLP1-951-LM) did not demonstrate a superior CA rate to the traditional LR approach. Both the LR and ANN approaches (in the validation database) were over 10% superior to the best existing prognostic system in terms of CA (LR/ANN 62.6%, published prognostic indices 27-49%) with a similar rate of TMMR (LR 7.8%, ANN 6.1%, published prognostic indices 2-17%).

Discussion

Classification of brain metastases patients into different prognostic groups is an important clinical and research endeavor due to the potential impact to patient treatment and clinical trial design. No superior system has been described in the literature; however, the RTOG RPA system is the most commonly used index and has previously undergone extensive validation. The development of these prognostic indices were primarily driven with statistical methodologies that attempted to identify relatively small patient groups (i.e. good and poor prognosis) with extreme survival characteristics leaving the majority of patients in intermediate categories. No prognostic indices have been developed with a-priori defined clinically relevant survival categories; however, these categories have been recently suggested in the medical literature and were used in this report [18].

Our investigation was not able to identify an ANN approach with superior CA to a more traditional LR statistical approach. This finding of equivalent utility between ANN and non-ANN approaches has been previously observed by Sargent et al. [20]. In this systematic review of 28 ANN vs. LR/Cox regression analyses of larger datasets (>200 patients) within the medical literature, half of the comparisons that were reviewed demonstrated equivalence between ANN and LR/Cox regression analyses. This effect becomes exaggerated as N becomes large (i.e. > 5000 patients) where 87.5% of comparisons were equivalent between ANN and LR/Cox. Therefore, this analysis re-confirms the fact that using complex modeling does not necessarily lead to improvements in predictive power over more traditional techniques, such as LR and Cox proportional hazards regression analyses.

Both ANN and LR approaches were superior to published prognostic indices in terms of CA in the order of 10% or greater. This is likely due to the fact that both the ANN and LR approaches were specifically optimized (using the cross-validation dataset) to categorize patients into one of three distinct prognostic groups (unfavorable < two months; intermediate two to six months; favorable > six months). None of the published prognostic systems were similarly optimized leading to inferior CA rates when compared to ANN and LR. The ANN approach demonstrated a small TMMR difference of less than 2% compared to the baseline LR statistical approach; however, both models had TMMR similar to published prognostic indices. Given this small difference in TMMR performance, which was similar to other indices in terms of absolute TMMR rates, an effective argument to adopt a complex ANN approach for patient

categorization cannot be effectively made. If internationally accepted categories, either identical or similar to those proposed in this manuscript, can be generally agreed upon and adopted, further research can be performed to create a new ideal prognostic index or to further refine existing scales. We would recommend the initial use of an LR approach to predict survival categories, as this investigation did not find evidence of complex interactions between prognostic variables requiring non-linear statistical techniques that are used in ANN optimization.

This investigation has several limitations relating to the database utilized that includes: a SRS/fSRT patient population with no external validation dataset. Also, the possibility of superior ANN optimization approaches and model structures need to be acknowledged. Future investigation in this area should include: integration of novel prognostic factors (e.g. genetic analysis or imaging-based parameters) into existing or new prognostic indices, survival classification standardization to drive future research and clinical care, and investigations into broader patient populations (e.g. WBRT, neurosurgery) to assess study generalizability. Additionally, the prediction of long-term survival (>one year) needs to be explored further using appropriate patient dataset(s) with sufficient median survival and statistical power to draw robust conclusions.

Conclusions

The use of various ANN model structures, nodal complexity, and cost function optimization algorithms did not lead to a significant improvement in survival classification when compared to a LR approach. Both ANN and LR approaches were superior in terms of CA but not TMMR when compared to traditional brain metastases prognostic indices.

Additional Information

Disclosures

Human subjects: Consent was obtained by all participants in this study. The University of Western Ontario and VU Amsterdam University issued approval # UWO Ethics 11164. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** The VUmc Amsterdam has research agreements with Varian Medical Systems and Brainlab AG, and the LRPC has been awarded a TomoTherapy Research Partnership Grant for partial funding of an ongoing phase 2 clinical trial. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Rodrigues G, Bauman G, Palma D, et al: Systematic review of brain metastases prognostic indices. *Pract Radiat Oncol* . 2012, [10.1016/j.prro.2012.04.001](https://doi.org/10.1016/j.prro.2012.04.001)
2. Gaspar L, Scott C, Rotman M, et al: Recursive partitioning analysis (RPA) of prognostic factors in three Radiation Therapy Oncology Group (RTOG) brain metastases trials. *Int J Radiat Oncol Biol Phys*. 1997, 37:745-51.
3. Lagerwaard FJ, Levendag PC, Nowak PJ, et al: Identification of prognostic factors in patients with brain metastases: a review of 1292 patients. *Int J Radiat Oncol Biol Phys* . 1999, 43:795-803.
4. Weltman E, Salvajoli JV, Brandt RA, et al: Radiosurgery for brain metastases: a score index for predicting prognosis. *Int J Radiat Oncol Biol Phys* . 2000, 46:1155-61.
5. Lorenzoni J, Devriendt D, Massager N, et al: Radiosurgery for treatment of brain metastases:

- estimation of patient eligibility using three stratification systems. *Int J Radiat Oncol Biol Phys*. 2004, 60:218-24.
6. Sperduto PW, Berkey B, Gaspar LE, Mehta M, Curran W: A new prognostic index and comparison to three other indices for patients with brain metastases: an analysis of 1,960 patients in the RTOG database. *Int J Radiat Oncol Biol Phys* . 2008, 70:510-14.
 7. Rades D, Dunst J, Schild SE: A new scoring system to predicting the survival of patients treated with whole-brain radiotherapy for brain metastases. *Strahlenther Onkol*. 2008, 184:251-55.
 8. Golden DW, Lamborn KR, McDermott MW, et al: Prognostic factors and grading systems for overall survival in patients treated with radiosurgery for brain metastases: variation by primary site. *J Neurosurg* . 2008, 109:77-86.
 9. Sperduto PW, Chao ST, Sneed PK, et al: Diagnosis-specific prognostic factors, indexes, and treatment outcomes for patients with newly diagnosed brain metastases: a multi-institutional analysis of 4,259 patients. *Int J Radiat Oncol Biol Phys* . 2010, 77:655-61.
 10. Sperduto PW, Kased N, Roberge D, et al: Summary report on the graded prognostic assessment: an accurate and facile diagnosis-specific tool to estimate survival for patients with brain metastases. *J Clin Oncol*. 2012, 30:419-25.
 11. Rades D, Dziggel L, Haatanen T, et al: Scoring systems to estimate intracerebral control and survival rates of patients irradiated for brain metastases. *Int J Radiat Oncol Biol Phys*. 2011, 80:1122-1127.
 12. Gaspar LE, Scott C, Murray K, Curran W: Validation of the RTOG recursive partitioning analysis (RPA) classification for brain metastases. *Int J Radiat Oncol Biol Phys*. 2000, 47:1001-06.
 13. Nieder C, Nestle U, Motaref B, Walter K, Niewald M, Schnabel K: Prognostic factors in brain metastases: should patients be selected for aggressive treatment according to recursive partitioning analysis (RPA) classes? . *Int J Radiat Oncol Biol Phys*. 2000, 46:297-302.
 14. Fleckenstein K, Hof H, Lohr F, Wenz F, Wannemacher M: Prognostic factors for brain metastases after whole brain radiotherapy. Data from a single institution. *Strahlenther Onkol*. 2004, 180:268-73.
 15. Paek SH, Audu PB, Sperling MR, Cho J, Andrews DW: Reevaluation of surgery for the treatment of brain metastases: review of 208 patients with single or multiple brain metastases treated at one institution with modern neurosurgical techniques. *Neurosurgery*. 2005, 56:1021-34.
 16. Saito EY, Viani GA, Ferrigno R, et al: Whole brain radiation therapy in management of brain metastasis: results and prognostic factors. *Radiat Oncol* . 1:20.
 17. Lagerwaard FJ, Levendag PC: Prognostic factors in patients with brain metastases. 2001, 11:27-46.
 18. Nieder C, Mehta MP: Prognostic indices for brain metastases--usefulness and challenges. *Radiat Oncol*. 2009, 4:10.
 19. Viani GA, Bernardes da Silva LG, Stefano EJ: Prognostic indexes for brain metastases: which is the most powerful?. *Int J Radiat Oncol Biol Phys*. 2012, 83:325-30.
 20. Sargent DJ: Comparison of artificial neural networks with other statistical approaches: results from medical data sets. *Cancer*. 2001, 91:1636-42.
 21. Su M, Miften M, Whiddon C, Sun X, Light K, Marks L: An artificial neural network for predicting the incidence of radiation pneumonitis. *Med Phys*. 2005, 32:318-25.
 22. Zhu J, Zhu XD, Liang SX, et al: Prediction of radiation induced liver disease using artificial neural networks. *Jpn J Clin Oncol*. 2006, 36:783-88.
 23. Pella A, Cambria R, Riboldi M, et al: Use of machine learning methods for prediction of acute toxicity in organs at risk following prostate radiotherapy. *Med Phys*. 2011, 38:2859-67.
 24. Rodrigues G, Eppinga W, Lagerwaard F, et al: A pooled analysis of arc-based image-guided simultaneous integrated boost radiation therapy for oligometastatic brain metastases. *Radiat Oncol*. 2012, 102:180-86.
 25. Lagerwaard FJ, van der Hoorn EA, Verbakel WF, Haasbeek CJ, Slotman BJ, Senan S: Whole-brain radiotherapy with simultaneous integrated boost to multiple brain metastases using volumetric modulated arc therapy. *Int J Radiat Oncol Biol Phys* . 2009, 75:253-59.
 26. Bauman G, Yartsev S, Fisher B, et al: Simultaneous infield boost with helical tomotherapy for patients with 1 to 3 brain metastases. *Am J Clin Oncol* . 2007, 30:38-44.
 27. Rodrigues G, Yartsev S, Yaremko B, et al: Phase I trial of simultaneous in-field boost with

helical tomotherapy for patients with one to three brain metastases. *Int J Radiat Oncol Biol Phys.* 2011, 80:1128-33.