Review

Predictive Modeling for Spinal Metastatic Disease

Akash A. Shah [1],* and Joseph H. Schwab [2]

1   Department of Orthopaedic Surgery, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA
2   Department of Orthopaedic Surgery, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA;
    joseph.schwab@cshs.org
*   Correspondence: aashah@mednet.ucla.edu

Abstract: Spinal metastasis is exceedingly common in patients with cancer and its prevalence is expected to increase. Surgical management of symptomatic spinal metastasis is indicated for pain relief, preservation or restoration of neurologic function, and mechanical stability. The overall prognosis is a major driver of treatment decisions; however, clinicians' ability to accurately predict survival is limited. In this narrative review, we first discuss the NOMS decision framework used to guide decision making in the treatment of patients with spinal metastasis. Given that decision making hinges on prognosis, multiple scoring systems have been developed over the last three decades to predict survival in patients with spinal metastasis; these systems have largely been developed using expert opinions or regression modeling. Although these tools have provided significant advances in our ability to predict prognosis, their utility is limited by the relative lack of patient-specific survival probability. Machine learning models have been developed in recent years to close this gap. Employing a greater number of features compared to models developed with conventional statistics, machine learning algorithms have been reported to predict 30-day, 6-week, 90-day, and 1-year mortality in spinal metastatic disease with excellent discrimination. These models are well calibrated and have been externally validated with domestic and international independent cohorts. Despite hypothesized and realized limitations, the role of machine learning methodology in predicting outcomes in spinal metastatic disease is likely to grow.

Keywords: machine learning; spinal metastasis; orthopedic oncology; oncology

1. Introduction

The spinal column is the most common site of bony metastatic disease [1]. An estimated 40–70% of all patients with cancer develop spinal metastasis, with 5–10% developing metastatic epidural compression [1,2]. As the survival rates of most primary malignancies increase, the prevalence of symptomatic spinal metastatic disease is expected to increase. Patients with symptomatic spinal metastases may present with pain, radicular symptoms, mechanical instability, or neurologic decline.

The treatment of spinal metastases is largely palliative, with a focus on pain relief, preserving or restoring neurologic function, local oncologic control, and maintaining mechanical stability of the spinal column. Spinal metastases were historically treated with conventional external beam radiation therapy with or without corticosteroids. Patients with radiosensitive malignancies (e.g., multiple myeloma and prostate cancer) responded well, while relatively radioresistant malignancies did not (e.g., sarcoma and colorectal cancer) [3–5]. In 2005, Patchell and colleagues reported a randomized controlled trial comparing surgical decompression and radiation therapy in patients with metastatic epidural compression. They found that surgical decompression combined with radiation yielded improved survival, maintenance and recovery of ambulation, decreased opioid requirements, and improved neurologic outcomes compared to radiation therapy alone [6]. The rate of surgical management of spinal metastatic disease has increased, with evidence that palliative surgery is valued by patients and their families [7,8].

Surgical treatment involves decompression and/or instrumented fusion. There is a wide spectrum of invasiveness that surgical decompression may entail, ranging from minimally invasive laminotomies to open laminectomy and even corpectomy. Similarly, instrumented fusion may comprise open posterolateral fusion or percutaneous pedicle screw fixation. Surgical management carries with it considerable perioperative morbidity including incidental durotomy, wound infection, and venous thromboembolism [9]. It is thus of great utility to determine a patient's overall prognosis before deciding on a specific treatment strategy.

In this narrative review, we aim to describe the tools that have been developed to predict mortality in patients with extramedullary spinal metastasis. While there have been numerous studies reviewing spinal oncology and artificial intelligence in spinal surgery, we provide a focus on conventional and machine learning-driven methods for modeling mortality in spinal metastasis specifically. We first explain the NOMS decision framework that organizes the assessment of a patient with spinal metastatic disease into four categories. This framework hinges on the patient's prognosis. We detail the multiple prognostic systems proposed for predicting mortality in spinal metastatic disease; these have been largely developed with features derived from expert opinion or logistic regression. As machine learning methods have been utilized in spinal surgery with great success, machine learning-driven models have been built to predict mortality in spinal metastatic disease. We describe the development and validation of these models. While these models represent an important advance in our ability to predict valuable prognostic information for patients with spinal metastasis, they carry specific risks and limitations. Finally, we discuss future directions for how machine learning and artificial intelligence tools can continue to improve our ability to predict mortality in patients with spinal metastasis.

2. NOMS Framework

The NOMS decision framework was developed to guide decision making for the treatment of symptomatic spinal metastatic disease. It distills the assessment of spinal metastases into four considerations: neurologic function, oncologic control, mechanical stability, and systemic disease burden [10].

The neurologic considerations assess both clinical and imaging findings. Clinical findings include myelopathy and functional radiculopathy. Bilsky and colleagues report a validated magnetic resonance-based scoring system to evaluate the degree of epidural spinal cord compression. Utilizing axial T2-weighted images, the extent of epidural compression is graded into low-grade lesions without cord compression (grades 0 and 1a–c) or higher grade lesions with an increasing degree of cord compression (grades 2 and 3) [11]. The oncologic consideration is based on the expected clinical response to radiation and/or systemic therapy. Primary tumor histology is the most important predictor of response to conventional external beam radiation therapy. Patients with radiosensitive malignancies may be treated with conventional radiation therapy regardless of the degree of epidural compression, whereas those with less favorable primary tumor histology may require stereotactic radiosurgery with or without surgical decompression [2,10,11].

Mechanical instability of the spine is an indication for surgery regardless of the neurologic or oncologic assessments. Incorporating both clinical and radiologic features, the Spinal Instability Neoplastic Score (SINS) is a validated scoring system that was devised to identify patients with instability that may require surgical stabilization (e.g., posterolateral fusion and percutaneous pedicle screw fixation). The SINS comprises six parameters: location, mechanical pain, lytic versus blastic lesion, spinal alignment, vertebral body collapse, and involvement of the posterior elements. Patients with a low SINS (0–6) may not require surgical stabilization, whereas patients with a high SINS (13–18) would benefit from stabilization [12,13].

Perhaps the most important consideration for treatment is whether the patient can tolerate a proposed intervention. This is based on the patient's systemic medical comorbidities and overall tumor burden. The neurologic, oncologic, and mechanical considerations

are of significance; however, the patient should survive long enough to make perioperative morbidity and risks worthwhile. Similarly, we do not want to withhold treatment from a patient who would benefit from it.

There is significant variability between the complication profiles of radiation therapy, minimally invasive surgery, and open decompression and fusion. Surgery for spinal metastasis carries with it considerable perioperative morbidity; complications include wound infection, pain, neurologic deterioration, and venous thromboembolism [14–16]. Accurate prognosis is thus crucial when determining which treatment modalities carry an acceptable risk–benefit ratio and are in line with the patient's goals of care. Providing accurate estimates of mortality would be of great utility to patients and families, improving informed decision making.

3. Prognostic Scoring Systems

3.1. Modeling Approaches

The goal of a predictive model is to determine the relationship between existing data and a future outcome. Most prognostic models predicting survival for spinal metastatic disease have done so with regression techniques (e.g., logistic regression). These models employ methods that are grounded in statistical inference. As an example, statistical inference may be employed to identify risk factors independently associated with one-year mortality; the odds ratios of these factors would then be weighted and used to build an equation that determines the risk of mortality. Such models are optimized for interpretability—that is, identifying which features drive mortality. In contrast, machine learning models are optimized for prediction. Spanning a vast array of learning algorithms (e.g., gradient boosting and Bayes Point Machine), machine learning models interrogate relationships between multiple variables in large datasets with the purpose of maximizing predictive performance. They often utilize complex equations to achieve this end, limiting their interpretability [17].

3.2. Classical or Regression-Driven Scores

Despite the recognition that patient survival is a key consideration in the planning of treatment for spinal metastatic disease, physicians have historically relied on clinical judgment to predict prognosis. Survival predictions by oncologists, radiation oncologists, and surgeons have been consistently determined to be inaccurate and overly optimistic [18–22]. Spine surgeons tend to overestimate life expectancy, leading to invasive procedures with greater associated perioperative morbidity [22]. There has thus been a significant effort to predict prognosis in patients with spinal metastatic disease in a data-driven manner with greater accuracy.

Tokuhashi and colleagues published a series of 64 patients with operatively treated spinal metastatic disease, utilizing 6 preoperative features to develop a scoring system to determine survival: performance status, the number of extraspinal bony metastases, the number of spinal metastases, the presence of visceral metastasis, primary tumor histology, and neurologic deficit. Out of a total score of 12, patients with a score ≤ 5 survived 3 months or less on average, whereas those with a score ≥ 9 survived 12 months or more [23]. They recommend that patients with a score ≤ 5 undergo palliative treatment, while those scoring ≥ 9 undergo surgical excision. In 2005, the authors reported an updated scoring system where they utilized the same 6 features in 246 patients composed of those treated conservatively and operatively; however, they expanded the primary tumor histology score to 0–5, hoping to provide more prognostic value to lesion biology. Out of a total score of 15, a score ≤ 8 is associated with survival < 6 months, with the authors recommending conservative treatment. A score between 9 and 11 is associated with survival > 6 months, with recommended palliative surgery. A score ≥ 12 is associated with survival > 1 year and excisional surgery is recommended by the authors [24].

In 1995, Bauer and Wedin assessed survival after surgery in 153 patients with extremity metastases and 88 with spinal metastases. They found five positive prognostic factors for

1-year survival: the absence of visceral metastasis, the absence of pathologic fracture, solitary spinal metastasis, no lung cancer, and primary tumor histology of breast, kidney, lymphoma, or multiple myeloma. The greater the number of positive prognostic factors, the higher the probability of 1-year survival [25]. Leithner and colleagues proposed a modified Bauer score in which they removed pathologic fracture from the list of prognostic features [26]. For the modified Bauer score, the authors recommend non-operative management, dorsal palliative surgery, or ventral–dorsal excisional surgery depending on the number of positive prognostic factors.

Spurred by the growing popularity of wide excisional resection for spinal metastases, Tomita and colleagues developed a prognostic scoring system for patients to guide surgical strategy. They included primary tumor histology, the presence of visceral metastasis, and the number of spinal/extraspinal bony metastases to generate a prognostic score ranging from 2 to 10. The lower the score, the more aggressive the recommended surgical strategy. A score of 2 or 3 was associated with a recommendation of wide or marginal excision, with the treatment goal of long-term local control, while a score of 4 or 5 yielded a recommendation of marginal or intralesional excision, with the goal of middle-term local control. A score of 6 or 7 led to a recommendation of palliative surgery, with the treatment goal of short-term palliation. Patients with scores ≥ 8 were recommended to undergo supportive care [27].

In 2005, van der Linden and colleagues published a prospectively randomized radiation therapy trial including 342 patients with symptomatic spinal metastases without neurologic deficit. They proposed a scoring system based on three prognostic factors: performance status, primary tumor histology, and the presence of visceral metastasis. Patients with ≤ 3 out of 6 points had a median survival of 3 months, 4 or 5 points was associated with a median survival of 9 months, and 6 points was associated with a median survival of 18.7 months [28].

Katagiri and colleagues reported a prognostic scoring system in 2005 for patients with skeletal metastasis who underwent non-operative or operative management. Similar to Bauer and Wedin, their cohort of 350 patients included both spinal and extremity lesions. They identify performance status, multiple skeletal metastases, the presence of visceral metastasis, and prior systemic therapy as important prognostic factors in their scoring system out of 8 points [29]. In 2014, they published a revised score with three major changes. They modified the primary histology breakdown to slow-growth, moderate-growth, and rapid-growth lesions based on the molecular characterization of the primary lesion (e.g., hormone-dependent versus hormone-independent lung cancer with or without molecularly targeted drugs) and refined the visceral metastasis factor as nodular/cerebral metastasis or disseminated metastasis. Additionally, they included laboratory data (e.g., C-reactive protein and serum albumin). The higher the number of points out of 10, the lower the survival [30].

With 307 patients who underwent surgery for spinal metastasis across 4 centers, Ghori and colleagues developed the New England Spinal Metastasis Score (NESMS). Feeling that the modified Bauer score inadequately assessed the general health status of patients, the authors included ambulatory status and serum albumin levels in addition to the modified Bauer score in their scoring system. The NESMS was externally validated on a cohort of operatively managed patients as well as separately validated on a cohort of non-operatively managed patients [31,32]. Additionally, a prospective analysis found that the NESMS shows superior predictive capability for 1-year mortality compared to the Tokuhashi, Tomita, and SINS systems [33].

In 2016, the SORG Orthopaedic Research Group published a survival algorithm for patients with operatively managed spinal metastatic disease. With 649 patients, they identified the following prognostic factors: age ≥ 65 years, performance status, primary tumor histology, multiple spinal metastases, visceral metastasis, prior systemic therapy, white blood cell count $\geq 11,000/\mu L$, and hemoglobin ≤ 10 g/dL. Based on the number of points out of 12, the authors divided patients into good, intermediate, and poor prognosis groups, with associated survival probabilities of 30, 90, and 365 days. To provide a more

user-friendly tool for clinicians, they additionally built a nomogram that converts each of the aforementioned factors into a score. The points are summed and translated into a patient-specific risk of survival at 30, 90, and 365 days [34]. The SORG nomogram was externally validated for 90-day and 1-year mortality on an independent cohort of 100 patients, outperforming the Tokuhashi score, Tomita score, modified Bauer score, and NESMS [35]. The nomogram was additionally externally validated for 30-day and 90-day survival after surgery, outperforming eight other scoring systems [36]. A recent systematic review similarly found that the SORG nomogram performs the best out of 17 scoring systems at predicting 90-day and 1-year mortality [37]. A summary of the discussed prognostic scoring systems is provided in Table 1.

Table 1. Conventional prognostic scoring systems.

Study	Number	Treatment	Location	Features	Points	Model
Tokuhashi et al., 1990 [23]	64	Operative	Spine	Performance status, the number of bony lesions, the number of spinal lesions, visceral lesions, primary histology, and neurologic deficit	12	None
Tokuhashi et al., 2005 [24]	264	Operative and non-operative	Spine	Performance status, the number of bony lesions, the number of spinal lesions, visceral lesions, primary histology, and neurologic deficit	15	None
Bauer and Wedin, 1995 [25]	241	Operative	Spine and extraspinal region	The number of spinal lesions, visceral lesions, primary histology, and pathologic fracture	5	Multivariate logistic regression
Leithner et al., 2008 [26]	69	Operative	Spine	The number of spinal lesions, visceral lesions, and primary histology	4	Multivariate logistic regression
Tomita et al., 2001 [27]	67	Operative and non-operative	Spine	The number of bony lesions, visceral lesions, and primary histology	10	None
van der Linden et al., 2005 [28]	342	Non-operative	Spine	Performance status, visceral lesions, and primary histology	6	Cox proportional hazards
Katagiri et al., 2005 [29]	350	Operative and non-operative	Spine and extraspinal region	Performance status, the number of bony lesions, visceral lesions, primary histology, and prior systemic therapy	8	Cox proportional hazards
Katagiri et al., 2014 [30]	808	Operative and non-operative	Spine and extraspinal region	Performance status, the number of bony lesions, visceral lesions, primary histology, prior systemic therapy, and laboratory data (C-reactive protein, lactate dehydrogenase, albumin, calcium, and bilirubin levels, and platelet count)	10	Cox proportional hazards

Table 1. *Cont.*

Study	Number	Treatment	Location	Features	Points	Model
Ghori et al., 2015 [38]	307	Operative	Spine	Modified Bauer score, performance status, and albumin level	3	Multivariate logistic regression
Paulino Pereira et al., 2016 [30]	649	Operative	Spine	Performance status, the number of spinal lesions, visceral lesions, primary histology, prior systemic therapy, age, white blood cell count, and hemoglobin level	12	Cox proportional hazards

Primary tumor histology and the presence of visceral lesions were most frequently identified as being important for prognosis; they were included in 8 of the 10 aforementioned studies. Pretreatment performance status was included in six studies. Table 2 details the most commonly included features in conventional prognostic models.

Table 2. Features included in conventional prognostic scoring systems.

Feature	Frequency	References
Primary tumor histology and the presence of visceral metastases	8	[23–30,34]
Performance status	6	[23,24,28–30,34]
The number of spinal metastases	5	[23–26,34]
The number of bony metastases	5	[23,24,27,29,30]
Prior systemic therapy	3	[29,30,34]
Serum albumin level	2	[30,38]
Neurologic deficit	2	[23,24]
Pathologic fracture	1	[25]
Age, WBC count, and hemoglobin level	1	[34]
Abnormal C-reactive protein, lactate dehydrogenase, calcium, and bilirubin levels and platelet count	1	[30]

3.3. Machine Learning-Driven Scoring Systems

The original Tokuhashi, revised Tokuhashi, and Tomita studies do not report utilizing statistical modeling to derive the prognostic factors that they included in their scoring systems [23,24,27]. The remaining prognostic scoring systems between 1990 and 2016 for predicting survival in patients with spinal metastatic disease were largely developed with Cox proportional hazards or multivariable logistic regression to identify prognostic factors and predict short- or long-term mortality [25,26,28–30,34,38]. Although regression techniques are often successful in assessing the association between explanatory features and outcomes, they are less suited for prediction. Machine learning methods have immensely grown in popularity in the past decade due to their ability to better capture complex non-linear relationships and factor–factor interactions compared to traditional regression techniques [39,40]. Machine learning algorithms have been shown to successfully predict clinical outcomes across diverse clinical disciplines, often outperforming logistic regression [41–43]. Additionally, machine learning modeling allows for the assessment of model discrimination, model calibration, and decision curve analysis—important metrics for determining the utility of a reported model [44].

The SORG Orthopaedic Research Group performed the first published machine learning-driven analysis for the prediction of survival after operatively managed spinal

metastasis in 2016. They developed a boosting algorithm that performed better at all time points for the prediction of 30-day, 90-day, and 1-year survival in the training dataset but performed slightly worse than the nomogram in the testing dataset [34]. They constructed the model with factors shown to be independently associated with survival: age ≥65 years, performance status, primary tumor histology, multiple spinal metastases, visceral metastasis, prior systemic therapy, white blood cell count ≥ 11,000/μL, and hemoglobin ≤ 10 g/dL.

With 1790 patients from the American College of Surgeons National Surgical Quality Improvement Program, the SORG group built a Bayes Point Machine algorithm for the prediction of 30-day mortality after surgery for spinal metastasis. This machine learning algorithm performed well, with an area under the receiver-operating characteristic curve (AUROC) of 0.782, and was well calibrated. Preoperative patient features used for algorithm development included serum albumin level, performance status, white blood cell count, hematocrit, alkaline phosphatase level, American Society of Anesthesiologists class, and location within the spine. The Bayes Point Machine was incorporated into a web application and deployed as an open-access tool for clinicians, the first such web-based application for the prediction of survival after spinal metastasis [45].

The SORG group then transitioned to building machine learning-driven models with high-quality institutional cohorts. With 732 patients from 2 large academic medical centers in Massachusetts, Karhade and colleagues developed stochastic gradient boosting algorithms that accurately predicted 90-day and 1-year mortality for operatively managed spinal metastatic disease. Features important for both models included serum albumin level, primary tumor histology, functional status, absolute lymphocyte count, alkaline phosphatase level, and creatinine level. Both models were well calibrated with excellent discrimination, with AUROC values of 0.83 and 0.89 for 90-day mortality and 1-year mortality, respectively. The discriminations of the stochastic gradient boosting models for 90-day and 1-year mortality were greater than those of the following scores: original Tokuhashi, revised Tokuhashi, Tomita, modified Bauer, van der Linden, Katagiri, NESMS, and SORG nomogram. The authors also reported a web-based calculator where users input patient feature values and receive a probability of 90-day or 1-year mortality [46]. These models were externally validated multiple times, first with a single independent cohort spanning patients from 2003 and 2016 in a tertiary-care medical center in Maryland [47]. They were externally validated again by Bongers and colleagues with a contemporary cohort composed of 200 patients from a tertiary-care center in New York between 2014 and 2016, performing excellently and with good calibration [48]. Despite taking important steps toward the generalizability of these models, both of these external validation studies were performed in the same geographic region of the United States that the SORG algorithms were developed in. Shah and colleagues successfully externally validated these algorithms with a geographically distinct cohort from a tertiary-care center in California. Furthermore, they showed that the algorithms perform accurately in a contemporary cohort subgroup treated in 2015 or later [49]. Additionally, Yang and colleagues reported successful international external validation of the SORG algorithms with a Taiwanese cohort of 427 patients [50]. Zhong and colleagues also reported successful external validation of the SORG algorithms in a cohort of 150 patients with lung-derived spinal metastasis [51].

It has been postulated that a patient is likely to benefit from surgical intervention for spinal metastatic disease if he/she is likely to live longer than 3 months [6,52]. Three months has thus been a survival time point that most of the classical and machine learning models have aimed to predict. With the advent of minimally invasive techniques that may be associated with less perioperative morbidity, the survival required to benefit from surgery may be less than this original cut-off. The SORG group thus sought to predict survival at an earlier time point. With an institutional cohort of 3001 patients with spinal metastasis treated non-operatively or operatively, they developed a machine learning model predicting 6-week mortality. The elastic net penalized logistic regression model was well calibrated, with excellent discrimination (AUROC: 0.85). They found that the

most important features for prediction were serum albumin level, primary tumor histology, absolute lymphocyte count, the number of spinal metastases, and functional status. This algorithm was externally validated with 1303 patients from 4 independent external cohorts from the United States and Taiwan [53]. A summary of the machine learning studies for spinal metastatic disease is provided in Table 3.

Table 3. Machine learning-driven prognostic tools.

Study	Number	Treatment	Features	Model
Paulino Pereira et al., 2016 [34]	649	Operative	Performance status, the number of spinal lesions, visceral lesions, primary histology, prior systemic therapy, age, WBC count, and hemoglobin level	Boosting regression
Karhade et al., 2019 [45]	1790	Operative	Performance status, ASA class, albumin level, WBC count, hematocrit, alkaline phosphatase level, and spinal location	Bayes Point Machine
Karhade et al., 2019 [46]	732	Operative	Performance status, visceral lesions, primary histology, BMI, creatinine level, alkaline phosphatase level, albumin level, platelet count, absolute lymphocyte count, hemoglobin level, INR, neutrophil–lymphocyte ratio, and platelet–lymphocyte ratio	Stochastic gradient boosting
Karhade et al., 2022 [53]	3001	Operative and non-operative	Performance status, the number of spinal lesions, visceral lesions, brain lesions, primary histology, albumin level, absolute lymphocyte count, WBC count, and alkaline phosphatase level	Elastic net penalized logistic regression

Performance status was the most commonly employed feature in the SORG algorithms, included in all four models. The frequency of features viewed as important for the machine learning-driven models are detailed in Table 4.

Table 4. Features included in machine learning-driven prognostic scoring systems.

Feature	Frequency	References
Performance status	4	[34,45,46,52]
Primary tumor histology and the presence of visceral metastases	3	[34,46,52]
Serum albumin and alkaline phosphatase level	3	[45,46,53]
WBC count	3	[34,45,53]
The number of spinal metastases	2	[34,53]
Absolute lymphocyte count	2	[46,53]
Hemoglobin level	2	[34,46]
Hematocrit, spinal region, and ASA class	1	[45]
Creatinine level, platelet count, neutrophil–lymphocyte ratio, platelet–lymphocyte ratio, and body mass index	1	[46]
Prior systemic therapy and age	1	[34]

Many of the same features are included in the prognostic models built with conventional statistics and those built with machine learning algorithms: primary tumor histology, the presence of visceral metastases, the number of spinal metastases, serum albumin level, white blood cell count, hemoglobin level, and prior systemic therapy. The SORG machine

learning models did not include the number of bony metastases, neurologic symptoms, or the presence of pathologic fracture. Additionally, both modeling approaches employed discrete features; there were no included continuous variables or imaging variables. The machine learning models included a greater number of features on average per model compared to the conventional models (8.8 features versus 4.7 features), primarily due to an increased number of laboratory values included. Earlier models built with traditional statistics attempted to distil the complex interplay of features that influence prognosis into a small number of features, with the aim of simplifying the features that physicians had to consult to receive accurate and interpretable information about prognosis. In contrast, a greater number of features tend to optimize the performance of machine learning models. The higher dimensionality of inputs maximizes predictive performance, often at the expense of interpretability.

Machine learning-driven analyses to predict survival for spinal metastatic disease represent a major advance in the ability of clinicians to accurately assess prognosis. These algorithms perform very well, with excellent discrimination, and are well calibrated across multiple time points—successfully predicting survival at 30 days, 6 weeks, 90 days, and 1 year. By quantifying risk with more granularity than simply "low-risk" or "high-risk", these tools can facilitate a data-driven discussion of risks and benefits, with subsequently improved preoperative patient counseling. Additionally, the machine learning analyses were shown to be internally valid, with the reporting of key metrics for the validation of clinical prediction models: model discrimination, model calibration, and decision curve analysis. Reporting these metrics is crucial to adequately assess the clinical utility of any predictive algorithm. Finally, the lack of external validation is a pervasive problem in the machine learning literature—significantly limiting the generalizability of these models [54]. The SORG machine learning models have been externally validated on multiple occasions with contemporary, geographically distinct domestic and international cohorts. The classical regression-based studies generate a total score based on the presence or absence of prognostic factors and recommend treatment based on which prognostic group the patient falls into. The SORG machine learning algorithms from 2019 onward provide web-based risk calculators that generate the patient-specific probability of mortality. These risk calculators allow clinicians to input features and receive the patient-specific probability of mortality as an output in a user-friendly digital interface.

Clinical data that are readily available in patient charts comprise the majority of input features used to develop these models. These include features such as primary tumor histology, the number of spinal metastases, the presence of visceral metastasis, and laboratory values. While primary tumor histology reflects the molecular characteristics of specific malignancies, molecular markers or gene expression data are yet to be incorporated into machine learning models for spinal metastasis mortality. Similarly, although deep learning methods have been applied to imaging for spinal metastasis, imaging data that are more detailed than the number of spinal metastases have not yet been included in prognostic models for spinal metastasis [55,56]. Additionally, patient-reported outcomes have not been utilized in reported models. Expanding the sources of input data to include molecular markers, imaging characteristics, and patient-reported outcomes is likely to improve predictive performance.

4. Assessing Machine Learning Algorithms

It is critical to determine the quality and validity of the proposed clinical prediction models. Let us consider internal validity and external validity separately. Internal validity refers to the validity of a model's performance on the cohort for which it was developed (i.e., the derivation cohort). An important characteristic of a clinical prediction tool is not only how reproducible it is but also how applicable it is to an independent population. While internal validity is a measure of the reproducibility of a model, external validity is a measure of generalizability [57]. Despite being critical to establish the utility of prediction tools, external validation studies are rare in the machine learning literature [54].

Aiming to provide rigorous assessment of the validity of a predictive model, Steyerberg and Vergouwe proposed reporting and evaluating four metrics: discrimination, the calibration slope, the calibration intercept, and decision curve analysis [57]. An accurate model should reliably discriminate between those who develop an outcome and those who do not. Discrimination is typically measured with the AUROC, also known as the concordance statistic or c-statistic. The AUROC represents the probability that the model can distinguish between a patient who develops an outcome and one who does not. An AUROC of 1 represents perfect discrimination and an AUROC of 0.5 indicates random prediction by the model [57].

Good discriminatory capability is necessary but not sufficient for overall model performance. In addition to discriminating an event from a non-event, a model's predictions should align with the observed outcomes within the study population. For every 100 patients for which a model predicts a 90% probability of 1-year survival, 90 patients should be alive in one year. Calibration captures this characteristic of a model; it is a measure of the agreement between the model predictions and the observed outcomes. Calibration is assessed with the calibration intercept and the calibration slope. The calibration intercept is a measure of the degree to which a model overestimates or underestimates the outcome of interest. A perfect model has a calibration intercept of 0. If a model predicting survival for spinal metastatic disease has a positive calibration intercept, it overestimates survival; an intercept value less than 0 means that the model underestimates survival. The calibration slope is a measure of how extreme a model's predictions are [57,58]. A perfect model has a calibration slope of 1.

While discrimination and calibration assess the accuracy of a model's predictions, they do not determine the clinical usefulness of a model. First introduced by Vickers and colleagues, decision curve analysis assesses clinical strategies by evaluating their net benefit across different probability thresholds [59]. The net benefit accounts for the potential benefit or harm from predictions issued by a model. It is a measure of the net true positives for a given threshold probability. A net benefit of 0.05 for a given threshold probability means that applying the model to 100 patients allows for the identification of 5 extra true positives without increasing the false positive rate [57,60]. The decision curve is drawn with the threshold probability on the x-axis and the net benefit on the y-axis. A model's performance is compared to the default strategies of changing management for all or no patients [44]. The decision curve is thus a graphical depiction of the trade-off between true positive and false positive predictions for a model [60]. Decision curves for different models can be compared to determine which models maximize the net benefit. The decision curves for the SORG machine learning models predicting 90-day and 1-year mortality in an independent cohort of surgically treated patients with spinal metastasis are shown in Figure 1.
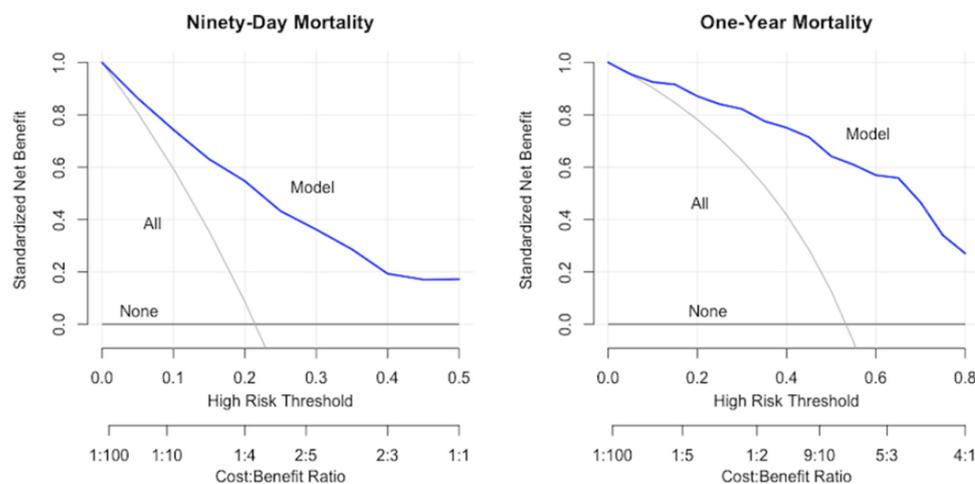

Figure 1. Decision curves for SORG models predicting 90-day and 1-year mortality after surgery for spinal metastasis. Figure obtained with permission from authors of Shah et al., 2021 [49].

5. Future Directions and Limitations of Artificial Intelligence

The scope of artificial intelligence in orthopedic surgery has expanded considerably in recent years, growing to include both unsupervised learning and natural language processing algorithms. As clinical datasets grow in both number and dimensionality, machine learning-driven algorithms will likely be utilized even more rapidly in orthopedic surgery and spinal surgery [61–63].

Advances in the molecular and immunohistochemical characterization of malignancies have led to an increase in biological data that should be incorporated into machine learning algorithms. Additionally, imaging data from computed tomography and magnetic resonance imaging studies should be incorporated into prognostic models. Other disciplines of machine learning such as natural language processing and unsupervised learning models may improve the prediction of outcomes; these remain underutilized for the prediction of outcomes in spinal metastatic disease. The clinical utility of machine learning algorithms would be enhanced by incorporating them into the electronic health record (EHR). Although care must be taken to protect patient confidentiality, embedding machine learning-driven risk calculators into the EHR would allow for the automated collection of relevant data, with the ability to continually self-update with prospectively collected data.

Despite the successful published applications of machine learning methodology for spinal metastatic disease, there are multiple potential limitations that must be addressed as the uptake of this technology increases. Machine learning methods are optimized for prediction but not necessarily for explanation. The individual effects of factors employed by advanced machine learning methods on the outcome of interest are not straightforward to interpret in the way that regression coefficients for traditional statistical modeling are—leading to a "black box" phenomenon.

Additionally, machine learning methods are prone to overfitting the data on which they are developed. It is thus crucial that robust internal and external validation strategies be employed for all predictive algorithms, with standardized reporting of key validation metrics (i.e., discrimination, calibration, and decision curve analysis). Prospective testing of these models is key. Finally, machine learning models perform well due to their ability to detect patterns within a dataset. As a result, the quality of the data source determines the quality of the models built with those data. Missing or inaccurate data can severely impact the performance of machine learning models. Furthermore, machine learning models may have the unintended consequence of propagating biases. Existing disparities in cohort creation and treatment strategies may be exacerbated, potentially causing harm to underrepresented groups such as ethnic minorities or those of lower socioeconomic status [64].

The most significant barrier to the clinical implementation of prognostic models in spinal metastatic disease is the lack of prospective randomized trials that interrogate the effectiveness of machine learning-driven predictions. While successful external validation studies improve the argument for generalizability, prospective randomized trials remain the gold standard for proving that prognostic tools yield improved outcomes for patients. Prospective trials must be designed to test the reported models.

6. Conclusions

The considerable improvements in data science must continue to be applied to the management of spinal metastatic disease. Symptomatic spinal metastatic disease is an important and growing clinical entity that requires a data-driven approach. Prognosis is a fundamental data point that patients and their families rely on to determine the next step in management. Patients with spinal metastatic disease are typically older patients with multiple medical comorbidities; even with improvements in perioperative care, surgical intervention in these patients has a considerable complication profile. Setting realistic expectations is a crucial component appropriately informing the risk–benefit calculation that patients and their families must make when considering treatment options. Machine

learning methods have built on the important work of classical prognostic algorithms, allowing for the accurate prediction of mortality in spinal metastatic disease. While these validated tools provide patient-specific risk of mortality, they do not and cannot tell clinicians how to treat a patient. It is not within the purview of models to determine treatment for a patient. How to treat symptomatic spinal metastasis is challenging and requires consideration of both clinical and psychosocial components. Shared decision making between patients, families, and their physicians is required to determine a treatment plan. Machine learning-driven algorithms can facilitate improved shared decision making by providing an accurate assessment of prognosis.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sciubba, D.M.; Petteys, R.J.; Dekutoski, M.B.; Fisher, C.G.; Fehlings, M.G.; Ondra, S.L.; Rhines, L.D.; Gokaslam, Z.L. Diagnosis and management of metastatic spine disease: A review. *J. Neurosurg. Spine* **2010**, *13*, 94–108. [CrossRef] [PubMed]
2. Barzilai, O.; Laufer, I.; Yamada, Y.; Higginson, D.S.; Schmitt, A.M.; Lis, E.; Bilsky, M.H. Integrating evidence-based medicine for treatment of spinal metastases into a decision framework: Neurologic, oncologic, mechanical stability, and systemic disease. *J. Clin. Oncol.* **2017**, *35*, 2419–2427. [CrossRef] [PubMed]
3. Rades, D.; Fehlauer, F.; Stalphers, L.J.A.; Wildfang, I.; Zschenker, O.; Schild, S.E.; Schmoll, H.J.; Karstens, J.H.; Alberti, W. A prospective evaluation of two radiotherapy schedules with 10 versus 20 fractions for the treatment of metastatic spinal cord compression: Final results of a multicenter study. *Cancer* **2004**, *101*, 2687–2692. [CrossRef] [PubMed]
4. Mizumoto, M.; Harada, H.; Asakura, H.; Hashimoto, T.; Furatani, K.; Hashii, H.; Murata, H.; Takagi, T.; Katagiri, H.; Takahashi, M.; et al. Radiotherapy for patients with metastases to the spinal column: A review of 603 patients at Shizuoka Cancer Center Hospital. *Int. J. Radiat. Oncol. Biol. Phys.* **2011**, *79*, 208–213. [CrossRef] [PubMed]
5. Gerszten, P.C.; Mendel, E.; Yamada, Y. Radiotherapy and radiosurgery for metastatic spine disease: What are the options, indications, and outcomes? *Spine* **2009**, *34 (22 Suppl)*, S78–S92. [CrossRef]
6. Patchell, R.A.; Tibbs, P.A.; Regine, W.F.; Payne, R.; Saris, S.; Kryscio, R.J.; Mohiuddin, M.; Young, B. Direct decompressive surgical resection in the treatment of spinal cord compression caused by metastatic cancer: A randomised trial. *Lancet* **2005**, *366*, 643–648. [CrossRef] [PubMed]
7. Yoshihara, H.; Yoneoka, D. Trends in the surgical treatment for spinal metastasis and the in-hospital patient outcomes in the United States from 2000 to 2009. *Spine J.* **2014**, *14*, 1844–1849. [CrossRef] [PubMed]
8. Fujibayashi, S.; Neo, M.; Miyaki, K.; Nakayama, T. The value of palliative surgery for metastatic spinal disease: Satisfaction of patients and their families. *Spine J.* **2010**, *10*, 42–49. [CrossRef] [PubMed]
9. Lau, D.; Leach, M.R.; Than, K.D.; Ziewacz, J.; La Marca, F.; Park, P. Independent predictors of complication following surgery for spinal metastasis. *Eur. Spine J.* **2013**, *22*, 1402–1407. [CrossRef]
10. Laufer, I.; Rubin, D.G.; Lis, E.; Cox, B.W.; Stubblefield, M.D.; Yamada, Y.; Bilsky, M.H. The NOMS framework: Approach to the treatment of spinal metastatic tumors. *Oncologist* **2013**, *18*, 744–751. [CrossRef]
11. Bilsky, M.H.; Laufer, I.; Fourney, D.R.; Groff, M.; Schmidt, M.H.; Varga, P.P.; Vrionis, F.D.; Yamada, Y.; Gerszten, P.C.; Kuklo, T.R. Reliability analysis of the epidural spinal cord compression scale. *J. Neurosurg. Spine* **2010**, *13*, 324–328. [CrossRef] [PubMed]
12. Fisher, C.G.; DiPaola, C.P.; Ryken, T.C.; Bilsky, M.H.; Shaffrey, C.I.; Berven, S.H.; Harrop, J.S.; Fehlings, M.G.; Boriani, S.; Chou, D.; et al. A novel classification system for spinal instability in neoplastic disease: An evidence-based approach and expert consensus from the Spine Oncology Study Group. *Spine* **2010**, *35*, E1221–E1229. [CrossRef] [PubMed]
13. Fourney, D.R.; Frangou, E.M.; Ryken, T.C.; Dipaola, C.P.; Shaffrey, C.I.; Berven, S.H.; Bilsky, M.H.; Harrop, J.S.; Fehlings, M.G.; Boriani, S.; et al. Spinal instability neoplastic score: An analysis of reliability and validity from the Spine Oncology Study Group. *J. Clin. Oncol.* **2011**, *29*, 3072–3077. [CrossRef] [PubMed]
14. Sebaaly, A.; Shehid, D.; Boubez, G.; Zairi, F.; Kanhonou, M.; Yuh, S.; Wang, Z. Surgical site infection in spinal metastasis: Incidence and risk factors. *Spine J.* **2018**, *18*, 1382–1387. [CrossRef] [PubMed]
15. Paulino Pereira, N.R.; Ogink, P.T.; Groot, O.Q.; Ferrone, M.L.; Hornicek, F.J.; van Dijk, C.N.; Bramer, J.A.M.; Schwab, J.H. Complications and reoperations after surgery for 647 patient with spine metastatic disease. *Spine J.* **2019**, *19*, 144–156. [CrossRef] [PubMed]
16. Groot, O.Q.; Ogink, P.T.; Paulino Pereira, N.R.; Ferrone, M.L.; Harris, M.B.; Lozano-Calderon, S.L.; Schoenfeld, A.J.; Schwab, J.H. High risk of symptomatic venous thromboembolism after surgery for spine metastatic bone lesions: A retrospective study. *Clin. Orthop. Relat. Res.* **2019**, *477*, 1674–1686. [CrossRef] [PubMed]
17. Bzdok, D.; Altman, N.; Krzywinski, M. Statistics versus machine learning. *Nat. Methods* **2018**, *15*, 233–234. [CrossRef] [PubMed]

18. Vigano, A.; Dorgan, M.; Bruera, E.; Suarez-Almazor, M. The relative accuracy of the clinical estimation of the duration of life for patients with end of life cancer. *Cancer* **1999**, *86*, 170–176. [CrossRef]

19. Chow, E.; Harth, T.; Hruby, G.; Finkelstein, J.; Wu, J.; Danjoux, C. How accurate are physicians' clinical predictions of survival and the available prognostic tools in estimating survival times in terminally il cancer patients? A systematic review. *Clin. Oncol. (R Coll Radiol).* **2001**, *13*, 209–218.

20. Chow, E.; Davis, L.; Panzarella, T.; Hayter, C.; Szumacher, E.; Loblaw, A.; Wong, R.; Danjoux, C. Accuracy of survival prediction by palliative radiation oncologists. *Int. J. Radiat. Oncol. Biol. Phys.* **2005**, *61*, 870–873. [CrossRef]

21. Razvi, Y.; Chang, S.; Zhang, L.; Tsao, M.; Barnes, E.; Danjoux, C.; Sousa, P.; Zaki, P.; McKenzie, E.; Lam, H.; et al. Are we better a decade later in the accuracy of survival prediction by palliative radiation oncologists? *Ann. Palliat. Med.* **2019**, *8*, 150–158. [CrossRef] [PubMed]

22. Thio, Q.C.B.S.; Paulino Pereira, N.R.; Palthe, O.V.W.; Sciubba, D.M.; Bramer, J.A.M.; Schwab, J.H. Estimating survival and choosing treatment for spinal metastases: Do spine surgeons agree with each other? *J. Orthop.* **2021**, *28*, 134–139. [CrossRef] [PubMed]

23. Tokuhashi, Y.; Matsuzaki, H.; Toriyama, S.; Kawano, H.; Ohsaka, S. Scoring system for the preoperative evaluation of metastatic spine tumor prognosis. *Spine* **1990**, *15*, 1110–1113. [CrossRef] [PubMed]

24. Tokuhashi, Y.; Matsuzaki, H.; Oda, H.; Oshima, M.; Ryu, J. A revised scoring system for preoperative evaluation of metastatic spine tumor prognosis. *Spine* **2005**, *30*, 2186–2191. [CrossRef] [PubMed]

25. Bauer, H.C.F.; Wedin, R. Survival after surgery for spinal and extremity metastases. Prognostication in 241 patients. *Acta Orthop. Scand.* **1995**, *66*, 143–146. [CrossRef] [PubMed]

26. Leithner, A.; Radl, R.; Gruber, G.; Hochegger, M.; Leithner, K.; Welkerling, H.; Rehak, P.; Windhager, R. Predictive value of seven preoperative prognostic scoring systems for spinal metastases. *Eur. Spine J.* **2008**, *17*, 1488–1495. [CrossRef] [PubMed]

27. Tomita, K.; Kawahara, N.; Kobayashi, T.; Yoshida, A.; Murakami, H.; Akamaru, T. Surgical strategy for spinal metastases. *Spine* **2001**, *26*, 298–306. [CrossRef] [PubMed]

28. van der Linden, Y.M.; Dijkstra, S.P.D.S.; Vonk, E.J.A.; Marijnen, C.A.M.; Leer, J.W.H.; Dutch Bone Metastasis Study Group. Prediction of survival in patients with metastases in the spinal column: Results based on a randomized trial of radiotherapy. *Cancer* **2005**, *103*, 320–328. [CrossRef] [PubMed]

29. Katagiri, H.; Takahashi, M.; Wakai, K.; Sugiura, H.; Kataoka, T.; Nakanishi, K. Prognostic factors and a scoring system for patients with skeletal metastasis. *J. Bone Jt. Surg. Br.* **2005**, *87*, 698–703. [CrossRef]

30. Katagiri, H.; Okada, R.; Takagi, T.; Takahashi, M.; Murata, H.; Harada, H.; Nishimura, T.; Asakura, H.; Ogawa, H. New prognostic factors and scoring system for patients with skeletal metastasis. *Cancer Med.* **2014**, *3*, 1359–1367. [CrossRef]

31. Goodwin, C.R.; Schoenfeld, A.J.; Abu-Bonsrah, N.A.; Garzon-Muvdi, T.; Sankey, E.W.; Harris, M.B.; Sciubba, D.M. Reliability of a spinal metastasis prognostic score to model 1-year survival. *Spine J.* **2016**, *16*, 1102–1108. [CrossRef] [PubMed]

32. Shi, D.D.; Chen, Y.-H.; Lam, T.C.; Leonard, D.; Balboni, T.A.; Schoenfeld, A.; Skamene, S.; Cagney, D.N.; Chi, J.H.; Cho, C.H.; et al. Assessing the utility of a prognostication model to predict 1-year mortality in patients undergoing radiation therapy for spinal metastases. *Spine J.* **2018**, *18*, 935–940. [CrossRef] [PubMed]

33. Schoenfeld, A.J.; Ferrone, M.L.; Blucher, J.A.; Agaronnik, N.; Nguyen, L.; Tobert, D.G.; Balboni, T.A.; Schwab, J.H.; Shin, J.H.; Sciubba, D.M.; et al. Prospective comparison of the accuracy of the New England Spinal Metastasis Score (NESMS) to legacy scoring systems in prognosticating outcomes following treatment of spinal metastases. *Spine J.* **2022**, *22*, 39–48. [CrossRef]

34. Paulino Pereira, N.R.; Janssen, S.J.; van Dijk, E.; Harris, M.B.; Hornicek, F.J.; Ferrone, M.L.; Schwab, J.H. Development of a prognostic survival algorithm for patients with metastatic spine disease. *J. Bone Jt. Surg. Am.* **2016**, *98*, 1767–1776. [CrossRef] [PubMed]

35. Paulino Pereira, N.R.; Mclaughlin, L.; Janssen, S.J.; van Dijk, C.N.; Bramer, J.A.M.; Laufer, I.; Bilsky, M.H.; Schwab, J.H. The SORG nomogram accurately predicts 3- and 12-months survival for operable spine metastatic disease: External validation. *J. Surg. Oncol.* **2017**, *115*, 1019–1027. [CrossRef]

36. Ahmed, A.K.; Goodwin, C.R.; Heravi, A.; Kim, R.; Abu-Bonsrah, N.; Sankey, E.; Kerekes, D.; De la Garza Ramos, R.; Schwab, J.; Sciubba, D.M. Predicting survival for metastatic spine disease: A comparison of nine scoring systems. *Spine J.* **2018**, *18*, 1804–1814. [CrossRef]

37. Smeijers, S.; Depreitere, B. Prognostic scores for survival as decisional support for surgery in spinal metastases: A performance assessment systematic review. *Eur. Spine J.* **2021**, *30*, 2800–2824. [CrossRef] [PubMed]

38. Ghori, A.K.; Leonard, D.A.; Schoenfeld, A.J.; Saadat, E.; Scott, N.; Ferrone, M.L.; Pearson, A.M.; Harris, M.B. Modeling 1-year survival after surgery on the metastatic spine. *Spine J.* **2015**, *15*, 2345–2350. [CrossRef] [PubMed]

39. Chen, J.; Asch, S. Machine learning and prediction in medicine—Beyond the peak of inflated expectations. *N. Engl. J. Med.* **2017**, *376*, 2507–2509. [CrossRef]

40. Hashimoto, D.; Rosman, G.; Rus, D.; Meireles, O. Artificial intelligence in surgery: Promises and perils. *Ann. Surg.* **2018**, *268*, 70–76. [CrossRef]

41. Esteva, A.; Kuprel, B.; Novoa, R.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [CrossRef]

42. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.C.; Wu, D.; Narayanaswamy, A.; Venugopalan, S.; Widner, K.; Madams, T.; Cuadros, J.; et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **2016**, *316*, 2402–2410. [CrossRef] [PubMed]

43. Menden, M.; Iorio, F.; Garnett, M.; McDermott, U.; Benes, C.H.; Ballester, P.J.; Saez-Rodriguez, J. Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS ONE* **2013**, *8*, e61318. [CrossRef] [PubMed]

44. Karhade, A.V.; Shin, J.H.; Schwab, J.H. Prognostic models for spinal metastatic disease: Evolution of methodologies, limitations, and future opportunities. *Ann. Transl. Med.* **2019**, *7*, 219. [CrossRef] [PubMed]

45. Karhade, A.V.; Thio, Q.C.B.S.; Ogink, P.T.; Shah, A.A.; Bono, C.M.; Oh, K.S.; Saylor, P.J.; Schoenfeld, A.J.; Shin, J.H.; Harris, M.B.; et al. Development of machine learning algorithms for prediction of 30-day mortality after surgery for spinal metastasis. *Neurosurgery* **2019**, *85*, E83–E91. [CrossRef] [PubMed]

46. Karhade, A.V.; Thio, Q.C.B.S.; Ogink, P.T.; Bono, C.M.; Ferrone, M.L.; Oh, K.S.; Saylor, P.J.; Schoenfeld, A.J.; Shin, J.H.; Harris, M.B.; et al. Predicting 90-day and 1-year mortality in spinal metastatic disease: Development and internal validation. *Neurosurgery* **2019**, *85*, E671–E681. [CrossRef] [PubMed]

47. Karhade, A.V.; Ahmed, A.K.; Pennington, Z.; Chara, A.; Schilling, A.; Thio, Q.C.B.S.; Ogink, P.T.; Sciubba, D.M.; Schwab, J.H. External validation of the SORG 90-day and 1-year machine learning algorithms for survival in spinal metastatic disease. *Spine J.* **2020**, *20*, 14–21. [CrossRef] [PubMed]

48. Bongers, M.E.R.; Karhade, A.V.; Villavieja, J.; Groot, O.Q.; Bilsky, M.H.; Laufer, I.; Schwab, J.H. Does the SORG algorithm generalize to a contemporary cohort of patients with spinal metastases on external validation. *Spine J.* **2020**, *20*, 1646–1652. [CrossRef]

49. Shah, A.A.; Karhade, A.V.; Park, H.Y.; Sheppard, W.L.; Macyszyn, L.J.; Everson, R.G.; Shamie, A.N.; Park, D.Y.; Schwab, J.H.; Hornicek, F.J. Updated external validation of the SORG machine learning algorithms for prediction of ninety-day and one-year mortality after surgery for spinal metastasis. *Spine J.* **2021**, *21*, 1679–1686. [CrossRef]

50. Yang, J.; Chen, C.; Fourman, M.S.; Bongers, M.E.R.; Karhade, A.V.; Groot, O.Q.; Lin, W.; Yen, H.; Huang, P.; Yang, S.; et al. International external validation of the SORG machine learning algorithms for predicting 90-day and one-year survival of patients with spine metastases using a Taiwanese cohort. *Spine J.* **2021**, *21*, 1670–1678. [CrossRef]

51. Zhong, G.; Cheng, S.; Zhou, M.; Xie, J.; Xu, Z.; Lai, H.; Yan, Y.; Xie, Z.; Zhou, J.; Xie, X.; et al. External validation of the SORG machine learning algorithms for predicting 90-day and 1-year survival of patients with lung cancer-derived spine metastases: A recent bi-center cohort from China. *Spine J.* **2023**, *23*, 731–738. [CrossRef] [PubMed]

52. Fehlings, M.G.; Nater, A.; Tetreault, L.; Kopjar, B.; Arnold, P.; Dekutoski, M.; Finkelstein, J.; Fisher, C.; France, J.; Gokaslan, Z.; et al. Survival and clinical outcomes in surgically treated patients with metastatic epidural spinal cord compression: Results of the prospective multicenter AOSpine study. *J. Clin. Oncol.* **2016**, *34*, 268–276. [CrossRef]

53. Karhade, A.V.; Fenn, B.; Groot, O.Q.; Shah, A.A.; Yen, H.; Bilsky, M.H.; Hu, M.; Laufer, I.; Park, D.Y.; Sciubba, D.M.; et al. Development and external validation of predictive algorithms for six-week mortality in spinal metastasis using 4,304 patients from five institutions. *Spine J.* **2022**, *22*, 2033–2041. [CrossRef]

54. Alsoof, D.; McDonald, C.; Kuris, E.; Daniels, A. Machine learning for the orthopaedic surgeon: Uses and limitations. *J. Bone Jt. Surg. Am.* **2022**, *104*, 1586–1594. [CrossRef] [PubMed]

55. Wang, D.; Sun, Y.; Tang, X.; Liu, C.; Liu, R. Deep learning-based magnetic resonance imaging of the spine in the diagnosis and physiological evaluation of spinal metastases. *J. Bone Oncol.* **2023**, *40*, 100483. [CrossRef] [PubMed]

56. Motohashi, M.; Funauchi, Y.; Adachi, T.; Fujioka, T.; Otaka, N.; Kamiko, Y.; Okada, T.; Tateishi, U.; Okawa, A.; Yoshii, T.; et al. A new deep learning algorithm for detecting spinal metastases on computed tomography images. *Spine* **2024**, *49*, 390–397. [CrossRef] [PubMed]

57. Steyerberg, E.; Vergouwe, Y. Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *Eur. Heart J.* **2014**, *35*, 1925–1931. [CrossRef] [PubMed]

58. Steyerberg, E.; Vickers, A.; Cook, N.; Gerds, T.; Gonen, M.; Obuchowski, N.; Pencina, M.J.; Kattan, M.W. Assessing the performance of prediction models: A framework for some traditional and novel measures. *Epidemiology* **2010**, *21*, 128–138. [CrossRef] [PubMed]

59. Vickers, A.; Elkin, E. Decision curve analysis: A novel method for evaluating prediction models. *Med. Decis. Mak.* **2006**, *26*, 565–574. [CrossRef]

60. Piovani, D.; Sokou, R.; Tsantes, A.G.; Vitello, A.S.; Bonovas, S. Optimizing clinical decision making with decision curve analysis: Insights for clinical investigators. *Healthcare* **2023**, *11*, 2244. [CrossRef]

61. Katsos, K.; Johnson, S.E.; Ibrahim, S.; Bydon, M. Current applications of machine learning for spinal cord tumors. *Life* **2023**, *13*, 520. [CrossRef] [PubMed]

62. Yagi, M.; Yamanouchi, K.; Fujita, N.; Funao, H.; Ebata, S. Revolutionizing spinal care: Current applications and future directions of artificial intelligence and machine learning. *J. Clin. Med.* **2023**, *12*, 4188. [CrossRef] [PubMed]

63. Kuijten, R.H.; Zijlstra, H.; Groot, O.Q.; Schwab, J.H. Artificial intelligence and predictive modeling in spinal oncology: A narrative review. *Int. J. Spine Surg.* **2023**, *17*, S45–S56. [CrossRef] [PubMed]

64. Gianfrancesco, M.; Tamang, S.; Yazdany, J.; Schmajuk, G. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern. Med.* **2018**, *178*, 1544–1547. [CrossRef] [PubMed]