Review

Artificial Intelligence-Driven Prediction Modeling and Decision Making in Spine Surgery Using Hybrid Machine Learning Models

Babak Saravi [1,2,3,*], Frank Hassel [2], Sara Ülkümen [1,2], Alisia Zink [2], Veronika Shavlokhova [4], Sebastien Couillard-Despres [3,5], Martin Boeker [6], Peter Obid [1] and Gernot Michael Lang [1]

1   Department of Orthopedics and Trauma Surgery, Medical Center-University of Freiburg, Faculty of Medicine, University of Freiburg, 79108 Freiburg, Germany; sara.uelkuemen@hotmail.de (S.Ü.); pm.obid@gmail.com (P.O.); gernotmichaellang@gmail.com (G.M.L.)
2   Department of Spine Surgery, Loretto Hospital, 79100 Freiburg, Germany; frank.hassel@rkk-klinikum.de (F.H.); alisia.zink@gmail.com (A.Z.)
3   Institute of Experimental Neuroregeneration, Spinal Cord Injury and Tissue Regeneration Center Salzburg (SCI-TReCS), Paracelsus Medical University, 5020 Salzburg, Austria; s.couillard-despres@pmu.ac.at
4   Department of Oral and Maxillofacial Surgery, University Hospital Heidelberg, 69120 Heidelberg, Germany; veronika.shavlokhova@med.uni-heidelberg.de
5   Austrian Cluster for Tissue Regeneration, 1200 Vienna, Austria
6   Intelligence and Informatics in Medicine, Medical Center Rechts der Isar, School of Medicine, Technical University of Munich, 81675 Munich, Germany; martin.boeker@tum.de
*   Correspondence: babak.saravi@jupiter.uni-freiburg.de

Abstract: Healthcare systems worldwide generate vast amounts of data from many different sources. Although of high complexity for a human being, it is essential to determine the patterns and minor variations in the genomic, radiological, laboratory, or clinical data that reliably differentiate phenotypes or allow high predictive accuracy in health-related tasks. Convolutional neural networks (CNN) are increasingly applied to image data for various tasks. Its use for non-imaging data becomes feasible through different modern machine learning techniques, converting non-imaging data into images before inputting them into the CNN model. Considering also that healthcare providers do not solely use one data modality for their decisions, this approach opens the door for multi-input/mixed data models which use a combination of patient information, such as genomic, radiological, and clinical data, to train a hybrid deep learning model. Thus, this reflects the main characteristic of artificial intelligence: simulating natural human behavior. The present review focuses on key advances in machine and deep learning, allowing for multi-perspective pattern recognition across the entire information set of patients in spine surgery. This is the first review of artificial intelligence focusing on hybrid models for deep learning applications in spine surgery, to the best of our knowledge. This is especially interesting as future tools are unlikely to use solely one data modality. The techniques discussed could become important in establishing a new approach to decision-making in spine surgery based on three fundamental pillars: (1) patient-specific, (2) artificial intelligence-driven, (3) integrating multimodal data. The findings reveal promising research that already took place to develop multi-input mixed-data hybrid decision-supporting models. Their implementation in spine surgery may hence be only a matter of time.

Keywords: mixed data; deep learning; deep neural networks; hybrid networks; multi-input; artificial intelligence; prediction; healthcare; machine learning; spine; degeneration

1. Introduction

Low back pain is one of the most frequently observed clinical conditions, and degenerative spine disease seems to be a leading driver of low back pain [1]. The global prevalence of low back pain increased from 377.5 million in 1990 to 577.0 million in 2017 [2].

The years lived with a disability increased globally from 42.5 million in 1990 to 64.9 million in 2017, representing an increase of 52.7%. Degenerative spinal disease is a common and impairing condition resulting in high socio-economic costs. Direct medical expenses spent on low-back pain doubled to 102 billion USD between 1997 and 2005, and the number of lumbar fusion procedures has quadrupled over the past 20 years, resulting in significantly increased healthcare costs [2].

Interestingly, the increase in performed surgeries is not directly proportional to improved patient outcomes. Impaired quality of life, persistent pain, and functional problems are reported in up to 40% of patients undergoing low back pain surgery and 20–24% undergoing revision surgeries [3,4]. Indications influencing the decision as to whether a patient should undergo surgery are not entirely based on guidelines but rather on discussions between the surgeon and patient, as well as the expertise and skills of the surgeon. Furthermore, there are no clear guidelines on surgical techniques for treating degenerative spinal diseases; as such, it remains unclear as to whether one treatment approach might perform better in particular cases than another. Overall, there is a considerable lack of data-driven decision-making in low back pain patients, which is particularly concerning when considering the global burden associated with low back pain.

Medical healthcare is driven by an incredible increase in the amount of data generated through various diagnostic tools and nodes within the healthcare systems. Patient data are the fundaments healthcare providers use to find the best fitting prognosis and diagnosis for each patient. Decisions are based on patterns across these datasets that guide towards the "right" diagnosis. Moreover, prognosis healthcare providers utilize these datasets to justify a specific treatment approach. Therefore, the correct interpretation of these datasets is crucial and directly impacts patient outcomes and the operations of healthcare systems.

Furthermore, improvements in treatment guidelines are mainly based on research that has been performed on such datasets. Researchers using these data might not be aware of the patterns hidden in their collected datasets. The process of finding patterns in large datasets which specifically fall under the category of big-data research is called data mining [5]. However, oftentimes, clinical researchers might not have profound knowledge in (bio)statistics personally, nor access to biostatisticians, to apply the best available tools to their datasets to extract all relevant pieces of information. Therefore, it is of high relevance that such datasets are made public and anonymized so that data scientists can use them and possibly determine these patterns using modern data-mining technologies.

The term "digital health" stands for the digitalization of healthcare data that was previously only assessed in an unproductive way through paper-based forms. New healthcare applications have become increasingly relevant and available. Such applications can range from mobile health applications, consumer techs, and telehealth for monitoring and guiding patients to precision medicine utilizing patient-specific data in artificial intelligence and bioinformatics models for individualized treatment approaches.

Machine learning is a subset of artificial intelligence and refers to computer techniques that allow complex tasks to be solved in a reproducible and standardized way. Machine learning combines biostatistics, mathematics, and computer science into one problem-solving pathway. One advantage is its efficiency and effectiveness, as the underlying programming code can be modified to enhance the accuracy of paths that solve a specific task. In this way, it can be more controllable, cost-efficient, and less error-prone than its "human template." Although the number of publications and citations in artificial-intelligence-related papers on healthcare topics is overwhelming, the technique is still at the beginning of its maturity. The industry highly supports the progress because of the great potential to improve medical research and clinical care, particularly as healthcare providers increasingly establish electronic health records in their institutions.

Predictive analysis with classical statistical techniques, such as regression models, applied on these datasets has been the gold standard to date. One may ask about the advantages of advanced machine learning techniques over simple regression analysis using widely available statistical software for predictive analysis. It is hard to draw a distinct

line indicating where basic statistical methods end and machine learning begins. It is often debated whether statistical techniques are somehow also considered as machine learning techniques, as in these cases, computers are using mathematical models to test a specific hypothesis. The primary differentiation might be the purpose of the application. Statistical methods, such as regression models, aim to find associations between independent variables (e.g., age, sex, body mass index) and dependent variables (e.g., patient-related outcome measures). Contrastingly, machine learning models also use statistical methods but aim to learn from training datasets, helping them to make more accurate predictions on the validation dataset so that the model can be reliably used on other independent datasets for predictive analysis. Hence, machine learning could be explained as focusing on predictive results, whereas simple statistical models analyze significant relationships. However, one further differentiation might be interpretability. The more complex a machine learning technique gets, the more accurate it can become at the cost of interpretability. For example, the lasso regression is a machine learning technique using regression analysis for feature selection and prediction. It has the advantage that it is not necessary to find the relevant independent variables first, as is required in linear regression modes. Its application is quite simple, and the interpretability is high. In contrast, deep learning, which is a subgroup of machine learning and will be discussed later, can get very complex but also very accurate; however, this comes at the cost of interpretability. The general principles of machine learning discussed in this review might help to differentiate between the most utilized approaches.

One significant barrier of machine learning applications is that reliable learning processes are very data-hungry. Machine learning is highly dependent on the premise that a large dataset is available. As computers cannot process visual and textual information the way human brains do, the algorithm needs to know what it is predicting or classifying in order to make decisions. When classification tasks need to be solved or specific areas need to be predicted, annotations are necessary. Data annotations help to make the input data understandable for computers. The task of the data scientist is to reliably label data such as text, audio, images, and video so it can be recognized by machine learning models and used to solve prediction and classification tasks. However, this process can be highly time-consuming, which might represent a major flaw in the implementation of machine learning algorithms. Non-accurate labeling will ultimately lead to inaccurate problem-solving. In previous research entitled "Deep Learning: A Critical Appraisal" [4], Marcus et al. proposed ten concerns associated with machine learning research, and data hungriness was listed as the top factor. He noted that "in problems where data are limited, deep learning often is not an ideal solution" [6]. Data-hungriness was also considered an unsolved problem in artificial intelligence (AI) research, described in Martin Ford's book "Architects of Intelligence: the Truth About AI From the People Building It" [7]. Most of the researchers interviewed in his book encourage the development of more data-efficient algorithms. Four pillars relevant for the implementation and interpretation of machine learning algorithms were described by Cutillo et al. based on discussions at the National Institute of Health (NIH) healthcare workshop in 2019 [8]. These were Trustworthiness, Explainability, Usability, Transparency, and Fairness.

An increase in data efficiency cannot be made feasible only by increasing the number of input samples but also by improving the machine learning architecture itself. One way to do this is to consider that different data types might contribute differently to the problem-solving task and that the connection between data types might also be relevant. Discussing the data dependency of machine learning algorithms and different hybrid models capable of processing different data types is, unfortunately, a research field that has not received the necessary attention yet. In particular, the translation of such hybrid algorithms to a clinical environment with real-world applications has not yet been reviewed. Our workgroup is currently investigating novel hybrid machine learning algorithms for applications in spine surgery. The search for comparable models in the literature while building the architecture revealed an unexpected lack of research in this field. This review

discusses hybrid algorithms for multimodal data processing and their implementation in spine research to close the gap currently available in the literature. The findings could help healthcare stakeholders plan and implement these promising algorithms in clinics.

2. Relevance of Digitalization in Spine Surgery

2.1. The Need for Structured Decision Making in Spine Surgery

A step towards precision and data-driven spine surgery can be achieved by meeting the significant requirement of developing informative outcomes in assessments. These include regular outcome assessments of patients, preferably utilizing digital app-based assessment forms, and the necessity to implement these outcomes as dependent variables in future risk assessment tools. Notably, the improvement of patient-related outcome measures (PROMs) should be the primary goal of decision-making. The value of such outcome measures is more critical in clinics than surrogate markers such as laboratory markers and classical clinical variables such as revision surgery, readmission, or absence of surgical infections. Our previous research has shown that patient-related outcome measures do not necessarily correlate with the factors a surgeon might consider relevant. For example, we could show that patient-related outcome measures were more correlated with the length of hospital stay than with postoperative complication rates [9]. Therefore, patient-related outcome measures should be an integral part of every predictive tool in spine surgery. In spine surgery, commonly utilized patient-related outcome measures include the Oswestry Disability Index, Core Outcome Measure Index (COMI), the eq-5D, SF-36 form, Numeric Rating Scale of pain, and the Visual Analogue Scale of pain, in addition to others [10]. Notably, Breakwell et al. reported in their publication entitled "Should we all go to the PROM? The first two years of the British Spine Registry" that a significant amount of PROMs forms were entered by the patients themselves [11]. Hence, an app-based tool transferring the results from the PROMs to a central database could be more time-efficient for spine surgeons. An additional benefit would be that outcomes could be compared considering all contributing institutes, and necessary quality-control steps could be performed in an early phase. This could also be very cost-efficient for healthcare institutes.

The integration of these patient-related outcome measures as dependent variables in clinical decision support tools would allow outcomes to be predicted during prospective follow-ups based on a set of several textual independent variables such as surgical technique, preoperative markers, as well as other data modalities such as imaging. This approach could reliably analyze large volumes of data based on previous data input and suggest next steps for treatment, flag potential problems, and enhance care team efficiency. Furthermore, this PROMs-including approach allows surgeons to discuss the possible outcome with patients and therefore improves the communication with patients. Contrastingly, a communication style in which the surgeon advises against surgery based on his subjective experience might lead to a negative surgeon-patient relationship. Such data-driven support tools might also be better to help surgeons communicate with patients.

2.2. Database Repositories for Machine Learning Applications in Spine Surgery

Databases are repositing data for future research. They are dedicated to housing data related to scientific research on a platform that can be restricted for access or publicly available. One often-used approach is to limit access to all contributors of the database. In this way, the database integrates a simple reward system: the contribution of data allows contributors to use the gathered data. Databases can collect and store a heterogeneous set of patient data and large datasets that fall under the category of big data. Usually, data in online medical databases are stored anonymously, maintaining that data cannot be linked to the patients' personal information. In such cases, radiological images can be stored with genetic and clinical data, all having a unique identification number linking the different datatypes of the case. These databases cover a wide range of data, including those related to cancer research, disease burden, nutrition and health, and genetics and the

environment. Researchers can apply for access to data based on the scope of the database and the application procedures required to perform relevant medical research.

For machine learning purposes, these data can also be labeled/annotated before being uploaded, allowing for utilization via data scientists. Although impactful machine learning models published to date might deal with a well-annotated dataset, the annotation process requires the necessary infrastructure, expertise, and resources, as it is very time-consuming depending on the number of data points. Considering the complexity of data annotation, crowdsourcing platforms are currently emerging. In this crowdsourcing model, the data are annotated by multiple crowdsourcing workers. One advantage is that the labeling can be checked against the consensus label using statistical parameters such as the inter-annotator agreement. Furthermore, this approach could lead to a more generalizable annotation style within the dataset. Therefore, the model might better predict future datasets coming from other workgroups. Crowdsourcing applications introduced were, for example, applications to database curation, the identification of medical terms in patient-authored texts, and the diagnosis of diseases from medical images [12–14]. Such platforms could also be applied by institutes in spine surgery. Although recent studies have shown that the accuracy performed by crowd workers is mainly similar to the individual annotation considering a given task, crowdsourcing is more resource-oriented and reliable [15–17]. The workflow of machine learning applications in spine surgery is shown in Figure 1.
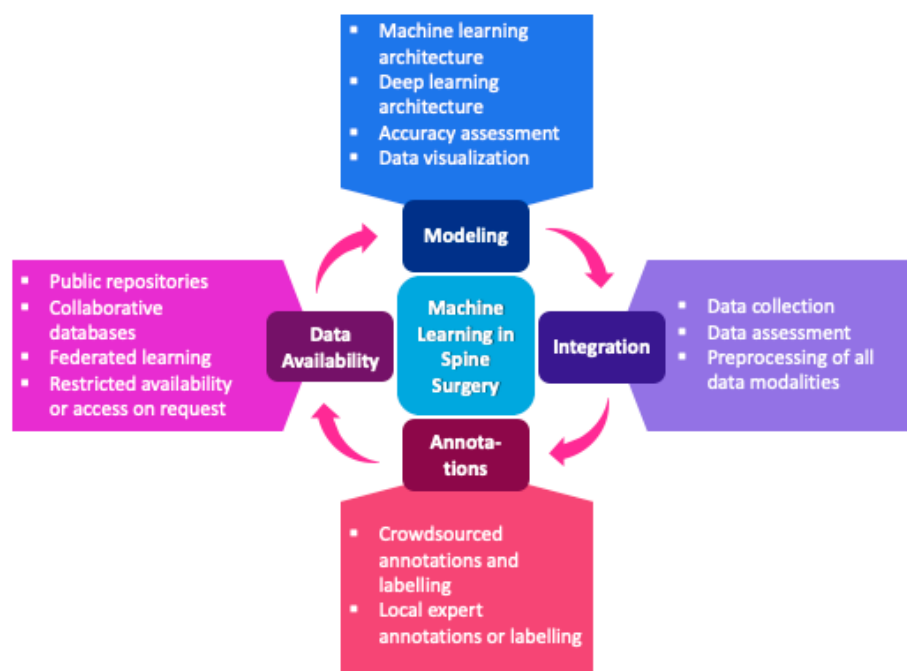
Figure 1. Workflow of machine learning applications in spine surgery.

Several databases are available that house an impressive number of global biomedical data. Notably, these repositories are regularly updated and extended using new image sets and data types provided by multiple institutions. Thus, they are often used by machine learning research studies, which is essential for progress in the field and exemplary for upcoming databases. For example, the GDC data portal [18] can provide RNA-sequencing, whole-genome, whole-exome sequencing, targeted sequencing, genotype, tissue and diagnostic slides, and ATAC-seq data. These data types could also be used as an input in hybrid machine learning models along with imaging and clinical data types to solve prediction tasks related to spinal oncology. Access to these platforms can be obtained from researchers but only for subsets of the whole dataset. The general principle of these platforms is that only data that will be used can be extracted. However, all mentioned databases do not contain data labeling and annotations. Considering that there can be vast amounts of data

depending on the research question, this might be a significant limitation for using these data for machine learning purposes. Several sources of public databases are accessible by anyone who wants to train and test their machine learning models. One such example is the Kaggle Dataset collection, which contains several algorithms and datasets in spine surgery [19,20]. These datasets are often used for competitions and training novel machine learning methods to determine whether they outperform existing models. This allows for a peer-review process as the algorithms are publicly available and commented on by other data scientists, validating the algorithm on the dataset provided and external datasets. However, since journal peer-reviewers may not have the resources to retest provided datasets with the algorithm code, often uploaded in GitHub repositories [21], such open peer-review processes meet crucial research goals, including validity, objectivity, and reliability. Furthermore, provided datasets and codes from the workgroups may not be available after some time. This represents a significant flaw in the assessment and development process of machine learning algorithms for healthcare applications. Publications are not the only relevant output of research; research data should also be considered. This is particularly true when considering that more accurate analysis pathways might not have developed when the study was conducted. This paradigm led to the emergence of data journals, such as Scientific Data from Nature [22] or GigaScience from Oxford Academic [23] in which the data can remain available for future analysis and validity assessments.

Notably, in surgical fields, such databases are still scarce. One of the largest and most intuitive databases in orthopedic surgery is the Osteoarthritis Initiative (OAI) database [24] from the National Institute of Health, which includes ten-year multi-center observational data of knee osteoarthritis cases. It includes DICOM images, clinical data, and laboratory data, and it is one of the few and most extensive repositories in orthopedic surgery capable of integrating multimodal data. Unfortunately, to the best of our knowledge, the only database that seems to include multimodal data in spine surgery is the Austrian Spinal Cord Injury Study [25]. The database contains longitudinal data on spinal cord injury cases in Austria and includes clinical data with patient-related outcome measures and imaging data. Other databases in spine surgery, which mainly include tabular clinical data, are the American College of Surgeon National Surgical Quality Improvement Project (ASC-NSQIP) database, the National Inpatient Sample (NIS) database, the Medicare and Private Insurance Database, the American Spine Registry, and the British Spine Registry [11,26,27]. The SORG ("Sorg Orthopaedic Research Group") has introduced the most recognized and cited predictive machine learning models, which can be accessed for free on their website [28]. They were already externally validated several times and include mortality prediction algorithms in spinal oncology, PROMs, and postoperative opioid use predictions after spine surgery, as well as discharge disposition for lumbar spinal surgery. Validation and external validation studies are both accessible on the website.

Another emerging field aiming to address the data handling problem in machine learning is privacy-first federated learning [29]. Federated learning [30,31] aims to train machine learning algorithms collaboratively without the need to transfer medical datasets. This approach would address the data governance and privacy politics, often limiting the use of medical data depending on the country where the research is conducted. Federated learning was extensively applied in mobile and edge device applications and is currently increasingly applied in healthcare environments [32,33]. It enables the assessment and development of models collaboratively using peer-review techniques without transferring the medical data out of the institutions where the data were obtained. Instead, machine learning training and testing take place on an institutional level, and only model architecture information and parameters are transferred between the collaborators. Recent studies have shown that machine learning models trained by Federated Learning can achieve similar accuracies to models that were implemented using central databases and are even superior to those processed on a single-institution-level [34,35]. Successful implementation of Federated Learning approaches could thus hold significant potential for enabling resource-oriented precision healthcare at a large scale, with external validation

to overcome selection bias in model parameters and to promote the optimal processing of patients' data by respecting the necessary governance and privacy policies of the participants [33]. Nevertheless, this approach still requires essential infrastructure and quality management processes to ensure that the applications perform well and do not impair healthcare processes or violate patient privacy rules.

Despite the advantages of Federated Learning, this method still has some disadvantages. For example, as described above, the integration of medical datasets in public databases could lead to more extensive research, and the investigation would not be limited to the collaborators. Furthermore, successful model training still depends on factors such as data labeling, data quality, bias, and standardization [36]. These issues would be better targeted when databases are accessible by more researchers and crowdfunding workers dealing with data annotation. This would be the case for both Federated and non-Federated Learning techniques. Appropriate protocols would be required, focusing on well-designed studies, standardized data extractions, standardized labeling and annotation of data, accuracy assessments and quality management, and regularly updated techniques to assess bias or failures. Considering this, Federated Learning would be a feasible approach to overcome data transfer limitations between institutions.

3. Hybrid Machine Learning Models for Classification and Prediction Tasks

3.1. Textual Data Conversion Methods for Deep Learning Approaches

Deep learning is a subset of machine learning utilizing artificial neural networks for information processing. Artificial neural networks were applied in various fields, including image analysis, natural language processing, and video or speech recognition [37]. The technique is based on information processing, similar to how humans process visual inputs. Artificial neurons are the units of the network, are organized in layers, and transmit the information from layer to layer depending on the predefined neural network architecture. Receptive fields, convolution kernels, and hierarchical feature abstraction are used in multiple layers to analyze data between the input and output layers [38]. The technique is used to process information with spatial or temporal dependencies, such as images where the spatial arrangement of pixels contains information about each image's content. Image processing through network layers is performed whereby the flow through the layers that are applying different mathematical functions leads to the desired pattern recognition. The architectures can be very heterogeneous, and the model is built depending on the task that needs to be solved. In healthcare data processing, these models are usually feedforward architectures where the information is processed from the input to the output layer. Tasks can, for example, be solved through classification or multi-classification models (i.e., for predicting categorical variables such as disease and health) or regression models (i.e., for predicting a continuous variable, such as a score) [39].

Deep neural networks have several benefits: finding hidden structures in the provided data, data augmentation, feature extraction, dimension reduction, optimum action selection in time-series data, and semi-supervised learning, including non-labeled data [40]. These neural networks are supposed to process imaging data; however, a large amount of healthcare data are textual (e.g., laboratory data, clinical data, genomic data). One way of using the strengths and advantages of convolutional neural networks to process such information is to rearrange tabular data into a 2-D shape that considers the relationship between the feature variables. Feature variables, in this case, are all variables that might be relevant for the task that has to be solved (e.g., outcome score prediction). To the best of our knowledge, four methods have been proposed for transforming non-image data into images shaped for use in convolutional neural networks. All transform the feature vector to a feature matrix using different transformation schemes. Ma et al. presented OmicsMapNet, a conversion method that arranges features based on annotations [41]. Hence, OmicsMapNet could be applied to RNA-Seq expression data of TCGA glioma samples. The functional hierarchical structure of the genes based on treemaps was used to construct 2-D images considering the function of genes. Consequently, gene functions could be learned for future functional

analysis of independent RNA-seq datasets [41]. The implementation of such conversions and inclusions of RNA-Seq data within hybrid models could allow researchers to analyze survival parameters (such as mortality or time to recurrence) in spinal oncology and better plan the follow-up of patients.

The second method was published by Sharma et al. in 2019 to convert genetic data and other textual data into images [40]. Their DeepInsight method uses similarity measurement or linear dimensionality reduction techniques, such as t-SNE or kernel principal component analysis, for data transformation [42]. The advantage of DeepInsight is that it can construct images where similar features are put together in neighboring elements, which is proposed to be beneficial for data processing through the convolutional neural network [40]. Another method was described by Baszir et al. in 2020, who introduced REFINED, which also considers similarities between features to generate 2-D feature maps in which the distance is reduced utilizing the Bayesian Metric Multidimensional Scaling Approach [43]. Both methods could have particular relevance for the analyses of gene phenotypes in spine research. For example, there is increasing evidence suggesting that there are genetic architectures of low back pain [44]. The use of more precise algorithms could help to classify patients according to their genetic architecture; in doing so, surgeons could plan therapeutical management according to the principles of precision medicine.

Recently, Kanber et al. [45] published another approach that compared five different conversion schemes for transforming sparse data into structured image sets. The transformation either applied for linear filling order strategies, keeping the initial ordering of the spatial feature intact (ASIS), random ordering with a randomized ordering of spatial features (RAND), and linear (SDIC) or circular (SDIC_c) orderings combined with mathematical calculations using the Pearson product-moment correlation coefficients to construct the image sets. They then compared the accuracy of the CNN application on the converted images to Random Forest decision tree classifier and the DeepInsight transformation scheme introduced by Sharma et al. [40] The highest accuracy of all mentioned transformation schemes was achieved by SDIC when applied to two public databases of textual data. The algorithm could be used to convert clinical information and surgery-specific parameters (such as operation time) into 2D shaped images which can be applied in deep learning models. This broadens the number of possible methods that could be applied to the initial dataset and allows for comparison between multiple techniques to obtain more precise prediction models in spine surgery clinics. Other published methods might not transform the textual data into images first. Still, they process textual information with imaging data in different ways to pass the layers of the CNN for classification or regression tasks. For example, the Dynamic Affine Feature Map Transform (DAFT), published by Pölsterl et al. in 2021, fuses information of high-dimensionality 3D MRI images and tabular data without first combining converted tabular features and images before feeding them into the network [46]. Specifically, DAFT rescales and shifts the feature maps of a convolutional layer conditional on a patients' tabular data in the CNN. Especially when considering the complexity and resource intensity of 3D images used in spine surgery (MRI, CT), this approach helps to effectively process such large datasets for prediction tests. Other published techniques were the processing of microscopic images, clinical data, and genomic data by Hao et al. (PAGE-Net) for survival analysis [47], and the processing of histological images and genomic biomarkers by Mobadersany et al. [48] Several other authors also used a multilayer perceptron (MLP) prior to the concatenation of image and tabular data before processing the information in the CNN ("early concatenation") [49–51]. The use of histological images and genomic biomarkers analyzed by these techniques could be routinely implemented in spine clinics. For example, blood from patients is usually taken for analysis before and after surgery on a routine basis. Meanwhile, the tissue taken in disc herniation surgery, for example, is considered biological waste and discarded. When more effective, cost-efficient, and fast diagnostic methods, such as in vivo reflectance confocal microscopy [52], are available in the future, the prediction could be made intraoperatively, and further management can be translated directly from the available data sets of each

patient ("precision medicine"). Overall, the extensive amount of research that has been per-formed within the last few years indicates the possibility of reliable textual data processing for multimodal hybrid deep learning models.

3.2. Multi-Input Mixed Data Deep Learning Models

The first applications performed on healthcare datasets were mainly focused on image data, such as the classification of skin cancer types, diabetic retinopathy, or pneumonitis in chest x-rays [53,54]. Multiple features in a dataset might contribute to a specific outcome of interest. These features are often not only relevant within a particular data type. A standard convolutional neural network architecture can consider multiple factors for prediction tasks while evaluating the impact of every factor on the target variable. Combining various data types for deep learning algorithms, which can require multiple inputs, can be seen as a multi-input mixed data deep learning approach based on a hybrid machine learning model pre-processing multiple data types. However, different data types can often not be processed within a single CNN. As discussed in the previous section, tabular data can be converted into image sets utilizing various techniques. The converted dataset can then be fed into the neural network via a separate input along with the non-converted dataset. The information can then be concatenated for feature processing and computing the prediction of these inputs. This approach is illustrated in Figure 2.
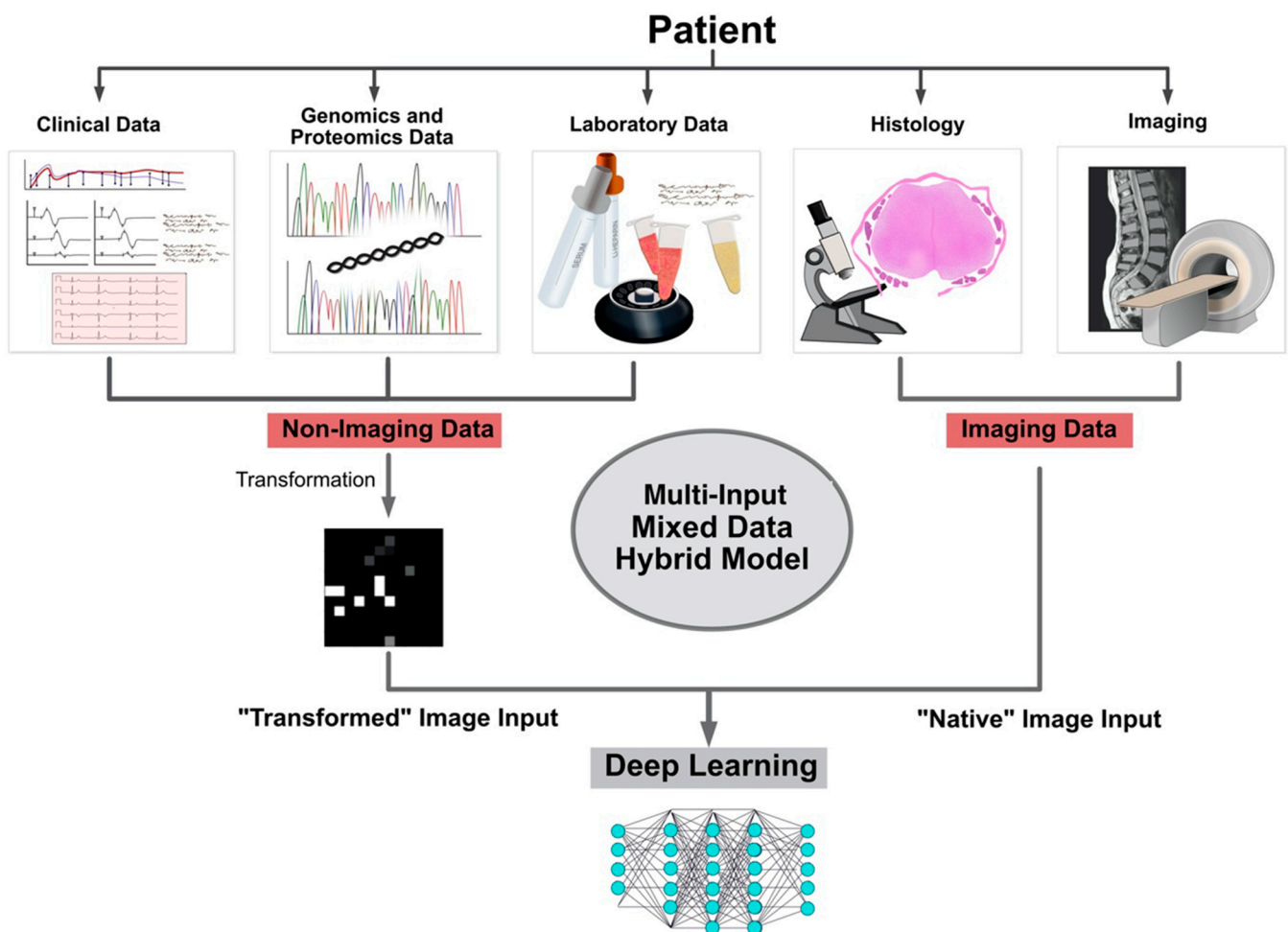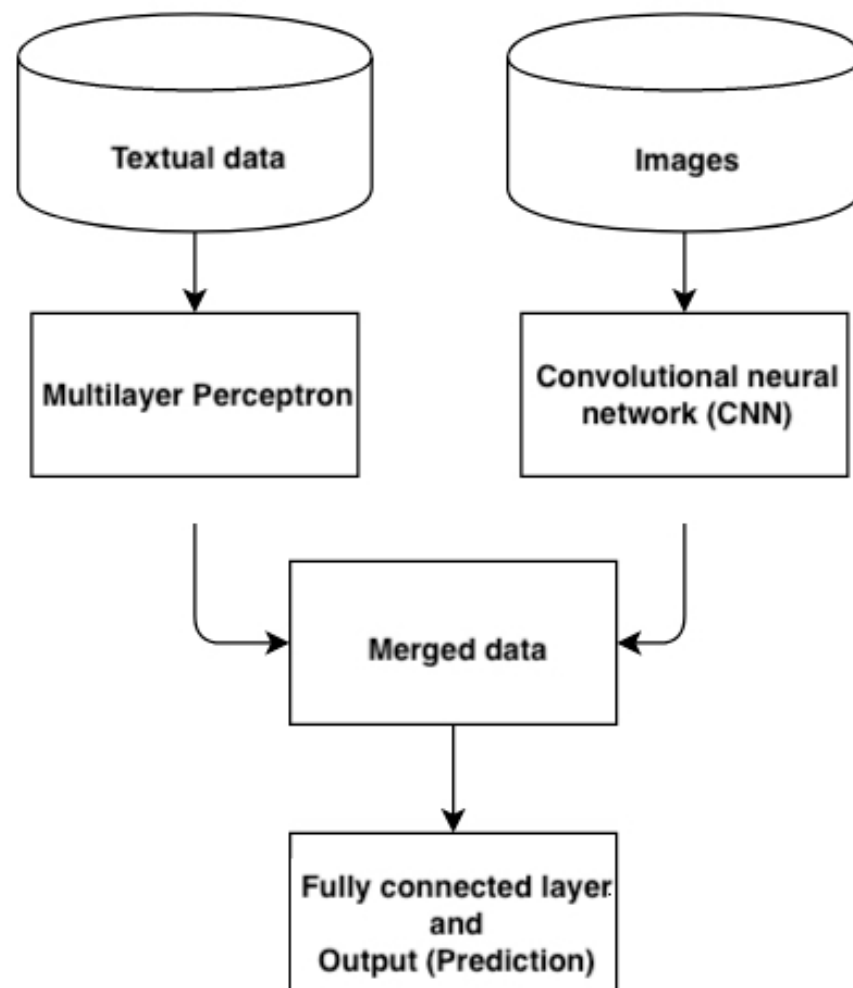

Figure 2. Illustration of multi-input mixed data model application in clinics.

The inputs used can be images, audio, text, and videos, for example. They can have multiple dimensions and are not restricted to specific variable types (e.g., continuous or categorical variables). This would allow surgeons to also implement text, audio, and videos

from patients' clinical data obtained during spine examinations in such models. Usually, the textual variables are normalized and scaled in pre-processing steps for better handling in the CNN. There are typically the following primary strategies for concatenating the multiple inputs [54]: 1. early concatenating strategies where the concatenation is performed in the input layer; 2. intermediate concatenating strategies, where the concatenation is also performed at the input layer, but backpropagation is used to propagate loss from the prediction model to the feature extracting networks; 3. concatenating strategies where the concatenation is performed at the output layers of different CNN branches. However, the pre-processing steps of converting the textual data into images for multi-input models are highly variable. The combination of both is currently not reported in the literature to the best of our knowledge. Textual inputs which have not been converted with techniques shown in the previous section could then be handled through a separate input along with the CNN input of the images. This can be, for example, done via Multilayer Perceptron, which handles one input before concatenation [55]. The Perceptron is able to solve classification tasks by the Stochastic Gradient Descent, which is used to minimize the distance between misclassified points and the decision boundary, and an activation function. Multilayer perceptron, a feedforward algorithm, can have multiple hidden layers and can also handle non-linear data. The structure of such an approach is shown in Figure 3.

Figure 3. Illustration of multi-input mixed data architecture using two separate inputs, handled by a multilayer perceptron and a convolutional neural network before concatenation and outcome prediction.

It is also possible to construct random forests, support vector machines, and variations of CNN models which can handle multiple inputs. For example, Li et al. introduced a concatenation framework capable of handling various data types enabling shortcut connections to the fully-connected layer, which is then directly fed into the output layer for predictions [56]. They reported satisfying accuracies across multiple datasets. This would allow researchers to integrate clinical data, radiological images, and genetic architectures from patients obtained in diagnostic examinations. Additional efforts were made by authors combining deep learning models with the help of ensemble methods using ultrasound images and x-ray images as data inputs [57,58]. Vasile et al., for example, used x-ray images, symptoms, and clinical and biological variables within an ensemble of deep learning models to predict the severity of COVID-19 diagnosis [57]. In combination with imaging techniques such as in vivo or ex vivo reflectance confocal microscopy or ultrasound imaging, this multimodal approach could solve prediction tasks in real-time, such as chairside applications [52,59]. Furthermore, Yuan et al. recently introduced a general architecture for Hybrid deep neural networks supporting mixed inputs reporting that the Hybrid model reached higher accuracies with classical MLP and CNN models [60]. Notably, the majority of studies published to date using multi-input models applied early concatenating strategies [54]. The techniques applied in these studies ranged from simply concatenating image and clinical features [61–63], applying dimensionality reduction techniques before early concatenation [64], to medical image feature extraction using automated or manual feature extraction methods before concatenation with textual data [65–67]. Interestingly, the fourth possible strategy of converting textual data into images for feeding them into multi-input models is not published to date, to the best of our knowledge. Our workgroup is currently constructing a hybrid multi-input mixed data model that converts the textual data of spine surgery patients into image datasets before feeding them into the hybrid model. It would be highly interesting to evaluate whether this strategy would result in higher accuracy than using the tabular data or imaging data of patients alone. This would open the door for an algorithm that could be easily integrated into clinical software for supported decision-making.

4. Available Artificial Intelligence-Based Models and Classical Statistical Prediction Models Utilized in Spine Surgery

References in this section were researched in Pubmed (Medline) and Web of Science utilizing the following search terms connected by Boolean operators: ("spine surgery") AND ("machine learning" OR "artificial intelligence") using both "MeSH terms" and "All fields" searches. An additional search in Google Scholar was conducted for grey literature. Furthermore, reference lists of the extracted works were screened.

Artificial intelligence-based approaches in spine surgery can help to improve the accuracy currently reported using traditional statistical modeling. Our literature search, which aimed to find predictive studies (prediction of clinically relevant outcomes, such as PROMs, complications, and mortality) utilizing machine learning methods in spine research, revealed that most of the work on this topic has been completed within the last few years. A list of these spine surgery outcome prediction algorithms published to date is shown in Table 1. Seventeen studies predicted PROMs [68–84], fifteen predicted complications [85–98], seven examined the discharge disposition after spine surgery [99–105], seven predicted the length of hospital stay [81,87,92,102,105–107], six predicted readmission or re-herniation after spine surgery [100,108–112], four predicted mortality rates [113–117], three predicted prolonged opioid use after spine surgery [118–120], two predicted the need for blood transfusions [121,122], one focused on the duration until "return to work" [123], one focused on 3D spinal alignment [124], and one predicted the future fracture probability after spine surgery [125]. Twenty-six of the studies included artificial neural networks within their prediction models, whereas thirty-seven utilized machine learning methods

without the use of artificial neural networks. None of the studies applied multi-input mixed data models for prediction modelling.

One of the most prominent applications of machine learning in spine surgery for prediction modeling was performed by Ogink et al. in 2019 [103]. They used the American College of Surgeons National Surgical Quality Improvement Program database to predict the non-home discharge after lumbar spinal stenosis surgery applying a convolutional neural network. They reported an area under the curve (a measure of accuracy in diagnostic tests, ranging from 0 (inaccurate) to 1 (perfect accuracy)) of 0.74, which was confirmed by another study validating the model [105]. Although they only used non-imaging data for their prediction, the results reveal the potential of machine learning approaches in predicting spine surgery outcomes. Another study published by Khan et al. in 2021 applied several machine-learning algorithms, including a classification tree, support vector machine, partial least squares, generalized boosted models, and multivariable adaptive regression splines to predict the health-related quality of life using the SF-36 form in patients who underwent surgery for degenerative cervical myelopathy. The results achieved an area under the curve of up to 0.78 for the multivariable adaptive regression spline, indicating a good degree of accuracy [74]. However, the sample size was small, with 130 samples for the training set. Moreover, no multi-input model was applied to handle patient data. Nevertheless, the approach reached considerable accuracy considering that tabular data was used solely. Varghese et al. reported an accuracy of 99% when using a random forest regression model to predict the pull-out strength of pedicle screws in osteoporotic and normal bone considering the density, insertion depth, and insertion angle feature variables [126]. This highlights that simple non-neural network-based modeling of simple tabular data can also lead to high accuracy in prediction tasks.

Hoffman et al. applied traditional statistical techniques (multivariate linear regression) and machine learning techniques (support vector machines) to predict the Oswestry disability index (ODI) and the modified Japanese Orthopaedic Association questionnaire [71]. They reported that both outcome measures could be more accurately predicted using a support vector machine compared to the t-test statistical method. Support vector machines seem to be a well-performing approach in spine surgery when no imaging data are processed because it performs well in cases where the sample sizes are limited and the number of features is large [127]. However, the predictability may be a severe limitation in cases where the sample size is small. In these cases, the classical statistical approach might be better suited [128].

One advantage of deep learning techniques over simple machine learning techniques is that they allow for in-depth processing of imaging data, which can lead to more accurate predictions. Kim et al., for example, applied artificial neural networks to predict complications following posterior lumbar spine fusion [91]. They reported that artificial neural networks outperformed the American Society of Anesthesiology classification and logistic regression as a classical statistical technique in predictive accuracy for several complication types. Several open-access web applications have been introduced and allow surgeons to predict the prolonged opioid prescription [119,129], postoperative failure [98], in-hospital and postoperative mortality [113,114,116], and discharge disposition [101,103,104] using a simple online application platform, where predefined variables can be filled in. They allow for external validation and might be improved in the future to be more precise based on increasing data collection, which is currently being seen in spine surgery.

Table 1. List of spine surgery outcome prediction studies. ANN: artificial neural networks; DTL: decision tree learning; LOR: logistic regression; elastic-net-LOR: elastic-net penalized logistic regression; MLOR: multivariate logistic regression; MLIR: multivariate linear regression; MARS: multivariable adaptive regression splines; MLP: multilayer perceptron; LASSO: least absolute shrinkage and selection operator; k-NN: k-nearest neighbor; BN: Bayesian network; EL: ensemble learning; NB: Naïve Bayes; RF: random forest; GLM: generalized linear model; GLMnet: elastic-net generalized linear model; GBM: gradient boosting machine; GAM: generalized additive model; PLS: partial least squares; XGBoost: extreme gradient boosting; NLP: natural language processing; pLDA: penalized linear discriminant analysis; RBF: radial basis function network; SVM: support vector machine; SVR: support vector regression; SGB: stochastic gradient boost; PROMs: patient-reported outcome measures; LOS: length of hospital stayS.

Author (Year)	Number of Datapoints	Algorithm	Intervention/Diagnosis	Outcome
Aldebeyan et al. (2016) [99]	15.092	MLOR	lumbar spine fusion surgery	discharge disposition
Andre et al. (2020) [85]	60	ANN	lumbar decompression	complications
Arvind et al. (2018) [86]	20.879	ANN, LOR, RF, SVM	cervical discectomy and fusion	complications
Babaee et al. (2018) [69]	480	MLP, RBF, LOR	posterior spinal fusion surgery	PROMs
Bekelis et al. (2014) [87]	2732	MLOR	corpectomy; spinal fusion	complications; LOS
Berjano et al. (2021) [68]	1243	RF	spinal lumbar arthrodesis	PROMs
Dong et al. (2021) [121]	152	SVM, DTL, MLP, NB, k-NN, RF	spinal fusion	blood transfusion
Durand et al. (2018) [122]	1029	RF, DTL	spinal deformity	blood transfusion
Finkelstein et al. (2021) [70]	122	LASSO, bootstrapping	spinal decompression/fusion surgery	PROMs
Goyal et al. (2019) [100]	59.145	ANN, GLM, GLMnet, GBM, NB, pLDA	spinal fusion	discharge disposition; readmission
Han et al. (2019) [88]	1.106.234	LASSO-R; LOR	spine surgery (various diagnoses and procedures)	complications
Harada et al. (2021) [130]	2630	XGBoost	lumbar microdiscectomy	disc re-herniation
Hoffmann et al. (2015) [71]	27	MLIR; SVR	cervical spondylotic myelopathy	PROMs
Hopkins et al. (2019) [108]	23.263	ANN	spinal fusion surgery	30-day hospital readmission
Hu et al. (2022) [118]	1316	SORG-algorithm (SGB)	lumbar disc herniation	prolonged postoperative opioid prescription
Janssen et al. (2021) [72]	77	RF	lumbar spinal fusion	PROMs
Kalagara et al. (2018) [109]	26.869	GBM	lumbar laminectomy	readmission
Karhade et al. (2019) [131]	2737	ANN, elastic-net-LOR, SGB, SVM, RF	cervical discectomy and fusion	sustained opioid prescription
Karhade et al. (2018) [101]	26.364	ANN, BN, DTL, SVM	lumbar disc surgery	discharge disposition
Karhade et al. (2019) [115]	1053	ANN, elastic-net-LOR, SGB, SVM, RF	spinal epidural abscess	in-hospital and 90-day post-charge mortality
Karhade et al. (2019) [114]	1790	ANN, DTL, BN, SVM	spinal metastasis surgery	30-day-mortality

Table 1. *Cont.*

Author (Year)	Number of Datapoints	Algorithm	Intervention/Diagnosis	Outcome
Karhade et al. (2019) [113]	732	ANN, elastic-net-LOR, SGB, SVM, RF	spinal metastatic disease management	90-day-mortality and 1-year-mortality
Karnuta et al. (2019) [102]	3807	NB	lumbar spinal fusion	discharge disposition and LOS
Karhade et al. (2021) [89]	1035	XGBoost (NLP algorithm)	anterior lumbar spine surgery	complications
Karhade et al. (2022) [110]	708	XGBoost (NLP algorithm)	posterior lumbar fusion	readmission
Khan et al. (2021) [74]	193	SVM, GAM (LogitBoost), MARS (earth), GBM, DTL, RF, LOR, PLS	degenerative cervical myelopathy	PROMs
Khan et al. (2021) [73]	757	SVM, GAM (LogitBoost), MARS (earth), GBM, DTL, RF, LOR, PLS	degenerative cervical myelopathy	PROMs
Khor et al. (2018) [75]	1583	MLOR	lumbar spine surgery	PROMs
Kim et al. (2018) [90]	4073	ANN, LOR	adult spinal deformity	complications
Kim et al. (2018) [91]	22.629	ANN, LOR	lumbar spine fusion	complications
Kuo et al. (2018) [127]	532	ANN, SVM, DTL, BN	spinal fusion	cost prediction
Kuris et al. (2021) [111]	63.533	ANN	posterior lumbar interbody fusion	readmission
Lewandrowski et al. (2020) [76]	383	ANN, LOR	lumbar spinal decompression	PROMs
Li et al. (2021) [132]	385	LOR, GBM, XGBoost, RF, DTL, MLP	osteoporotic vertebral compression fracture	bone cement leakage
Maki et al. (2021) [133]	478	GBM, XGBoost, RF, LOR	cervical ossification of the posterior longitudinal ligament	PROMs
Massaad et al. (2022) [92]	484	k-means clustering analysis, LOR	spinal metastases surgery	complications, LOS, mortality
McGirt et al. (2015) [77]	1803	BN, LOR	lumbar spine surgery	PROMs
Merali et al. (2019) [78]	757	ANN, LOR, DTL, RF, SVM	degenerative cervical myelopathy	PROMs
Nunes et al. (2022) [112]	215.999	ANN, Cox-Regression, XGBoost, DTL, NB, RF	thoracolumbar fractures	30-day readmission
Ogink et al. (2019) [104]	9338	ANN, DTL, BN, SVM	Spondylolisthesis	discharge disposition
Ogink et al. (2019) [103]	28.600	ANN, DTL, BN, SVM	lumbar spinal stenosis	discharge disposition
Oh et al. (2017) [79]	234	DTL	adult spinal deformity	PROMs
Papić et al. (2016) [123]	153	DTL, SVM, MLP	lumbar microdiscectomy	return to work
Pasha et al. (2021) [124]	371	EL	adult idiopathic scoliosis	3D spinal alignment
Passias et al. (2018) [134]	101	DTL	cervical deformity surgery	distal junctional kyphosis
Pedersen et al. (2020) [80]	1968	ANN, DTL, RF, SVM	lumbar disc herniation	PROMs
Ratliff et al. (2016) [93]	279.135	LASSO, LOR	spine surgery (various diagnoses and procedures)	complications
Russo et al. (2021) [106]	1516	MLOR, LASSO	cervical discectomy	LOS

Table 1. *Cont.*

Author (Year)	Number of Datapoints	Algorithm	Intervention/Diagnosis	Outcome
Shah et al. (2019) [98]	367	ANN, RF, SVM, SGB, elastic-net-LOR	spinal epidural abscess	complications
Shah et al. (2021) [116]	298	SORG-algorithm (SGB)	spinal metastasis surgery	90-day and 1-year mortality
Shah et al. (2021) [94]	6822	LOR, RF, GBM, XGBoost	posterior cervical spinal fusion	complications
Siccoli et al. (2019) [81]	635	ANN, RF, XGBoost, DTL, GLM, k-NN	lumbar spinal stenosis	PROMs, reoperations, LOS
Staartjes et al. (2019) [82]	422	ANN, LOR	lumbar discectomy	PROMs
Staartjes et al. (2022) [83]	817	GLM, elastic-net-LOR, k-NN	lumbar spinal fusion	PROMs
Stopa et al. (2019) [105]	144	ANN	lumbar disc surgery	discharge disposition, LOS
Veeramani et al. (2022) [95]	54.502	ANN, LOR, MVR, DTL, RF, GBM, XGBoost	anterior cervical discectomy and fusion	complications
de Vries et al. (2021) [125]	7578	ANN, RF, Cox-regression	fracture patients with osteopenia and osteoporosis	future fracture
Wang et al. (2021) [96]	13.500	XGBoost	posterior lumbar fusion	complications
Wang et al. (2021) [135]	184	SVM	posterior laminectomy and fusion with cervical myelopathy	complications
Wong et al. (2020) [136]	1164	SVM	anterior cervical discectomy and fusion	complications
Wirries et al. (2021) [84]	60	ANN	lumbar disc herniation	PROMs
Yang et al. (2021) [117]	427	SORG-algorithm (SGB)	spinal metastasis surgery	90-day and 1-year mortality
Zhang et al. (2021) [107]	1281	LOR, DTL, RF, XGBoost, GM	spinal fusion surgery	LOS
Zhang et al. (2020) [120]	19.317	ANN, LASSO, LOR, RF, SGB	thoracic or lumbar spine surgery (low back pain)	prolonged opioid use

One of the main advantages of using deep learning algorithms is that training can be performed without predefining the variables and features that need to be included. This is especially advantageous for high-volume, multidimensional, and complex data types, including genomic or sequencing data, which require high computational resources and time-consuming annotations. These also require experts for feature selection or feature engineering [137]. However, images can also be taken in 3D, such as computed tomography scans or magnetic resonance images, and would need extensive pre-processing or complex strategies for multi-input processing [19]. One way to deal with such multi-slice data is to split them into multiple 2D images per patient or apply 3D neural networks for multi-input mixed-data hybrid data processing to overcome the applicable selection bias [138]. Notably, the generalizability of images is limited because the data are obtained through different methods (e.g., different MRI models), which might be particularly problematic when obtaining genomic data that are also dependent on the platform [139]. Especially for genomic data, for example, to classify cell types of intervertebral discs based on sequencing data, the large number of data points are better processed through more time-efficient methods. Combining the findings from such datasets with histopathological data can also be used for survival predictions during cancer research, which may also extend to many spinal diseases [140]. Furthermore, artificial intelligence can be applied to assess molecular

markers of spine tumors, predicting the survival of primary spine tumors or metastatic recurrence rates [140]. In summary, the impressive amount of research performed on the applicability of artificial intelligence in clinically highly relevant prediction tasks is opening the door for a new kind of decision-making, which is currently on the rise in spine surgery.

5. Future Perspectives and Limitations

The combination of different datatypes allows for a multi-perspective view of patients' data. This approach can reveal more information than the inclusion of one datatype alone. The clinical information is especially relevant for the interpretation of radiological images. Not having access to the laboratory or clinical data has been shown to significantly impact the interpretability of radiological images [141,142]. This was demonstrated in a survey where most radiologists reported that the availability of clinical information highly affected their reports [143]. Thus, this can also be assumed to be relevant for artificial intelligence models simulating human behavior. Notably, the volume of workload in surgery is very high. As such, when considering that spine surgeons often have to perform interpretation tasks during night shifts or at times of high workloads, the interpretation of multiple data types can be prone to errors. An automated assessment could help integrate an alert system for spine surgeons who can then take further care of their patients in cases whereby artificial intelligence indicates some sort of attention regarding the prediction of outcomes.

The growth in the number of publications focusing on deep learning for images is enormous, whereas hybrid models are only just beginning to grow [54]. Several factors have impacted their development and implementation. These factors range from data-sharing limitations in healthcare institutions currently to the integration of machine learning algorithms in clinical settings considering the "good machine learning principles" [8]. It would be advantageous to consider these limitations before the implementation is planned. Following this strategy may allow healthcare specialists to learn from previously reported difficulties in the implementation phase. Consequently, the implementation will be more time-efficient and resource-oriented. Hence, the full potential of machine learning applications in healthcare settings can be maximized while avoiding problems that may arise due to the inherent privacy governance in healthcare [8].

Reliable data and algorithms are necessary but insufficient for implementing machine learning techniques in clinical settings. The application of machine learning requires comprehension and assessment of its implications to clinicians, patients, and other nodes in the healthcare system while developing the algorithms in real-time based on patient data in rapidly changing clinical environments. Increased collaboration among researchers and healthcare providers is needed regarding the development of machine learning workflow and data training, with outputs reflecting the needs of patients, to ensure that these systems are feasible, trustworthy, and usable in clinical settings. Providing datasets in repositories is highly encouraged. However, even if data deposition is made mandatory for research performed in spine surgery, there are several challenges to making the data in these repositories useful for machine learning and deep learning tasks. One must obtain ethics approval which, depending on the relevant government policies, may be a somewhat complicated process. Then, proper and careful labeling of the data is a critical task that can limit usability when not performed in a standardized way by experts [144]. As such, in cases whereby the researcher could not validate the labeling, the "failure" in data labeling and annotations would still exist in the database.

Furthermore, imaging plays a crucial role in spine surgery; however, processing 3D images such as MRI and CT can be very time-consuming. In addition, different machines used to obtain the images might introduce selection bias [145]. Consequently, this can affect the translation of data to other institutes. In this case, it is particularly warranted to provide multi-center data in such repositories or combine similar datasets from different institutes utilizing standardized techniques. The more data the algorithm processes, the more applicable it can be regarding unseen data. Notably, unsupervised machine learning techniques do not need any labeling of data [146]. Instead, these techniques identify

patterns in the dataset through dimension reduction techniques and are mainly exploratory. However, they still do not perform well on non-qualitative small datasets and require more computational power than large datasets. Overall, they can be more complex than supervised learning methods and are mainly intended for clustering tasks, such as disease taxonomy, based on their pathophysiology [147]. In this case, the validation is limited as there are no labels and, therefore, no "ground truths" to confirm the obtained results of the task performed by the unsupervised learning algorithm [148]. Thus, experts need to validate the performance afterward, which can also be time-consuming.

Finally, data in spine surgery is collected at an institutional level. The anonymization and transfer of large datasets might require specialist infrastructure, along with a dedicated team of data scientists to handle such large data volumes [149] Implementation of repositories that lower the barriers for large volume data transfer, such as 3D images from clinics to the database without compromising data privacy, might increase the efficiency of data transfer and thus could help increase the amount of data provided by surgeons [26].

Furthermore, researchers developing machine learning algorithms must also consider model updates when larger or novel datasets are made publicly available or if algorithms need to be improved based on more recent research. As a result, continuous monitoring of machine learning applications becomes a necessity. This can result in high maintenance costs and can also be time-consuming. Failures to comply can have far-reaching consequences for the patients. After implementing such algorithms, continuous monitoring can ensure that the algorithms are working as expected, disregarding the data type that is used. However, as in other automated applications, such as self-driving cars, the applications should not be the only pillar decisions are based on [150,151]. Clinicians should consider these applications as decision-supporting tools, which should be questioned every time they are applied. Finally, the workflow should be explainable to maintain safety, which is particularly complicated in complex machine learning models involving artificial neural networks. While these uprising decision-supporting techniques can add significant value to healthcare systems, they still have substantial challenges that need to be considered by all clinicians aiming to implement these in their daily practices. Poorly implemented tools that generate unnecessary alerts can not only be a threat to patients and lead to worse outcomes, but they could also significantly increase healthcare costs.

Furthermore, accurate models require large data sets to sufficiently perform in various clinical environments and enable them to generalize well to new and unseen data [36]. As such, conclusions drawn from machine learning in healthcare applications are dependent on the quality of the training dataset with which the algorithm was initially trained. Hence, the deposition of research data can be considered an essential step in research as this ensures that models can be tested against new unseen datasets for external validation before evaluating whether the applications prove themself in clinical settings. As computer technologies develop, machine learning processes can be tested against a vast amount of high-quality medical data to evaluate outcome predictions. While the application of machine learning in healthcare is still in an early phase, the lack of data preventing its integration and implementation into the health care system might slow down the development process. Encouraging data deposition by researchers helps to take advantage of such algorithms towards patient care in the near future. Hence, we encourage researchers to make their data available publicly, particularly when the data set is multimodal, considering that such datasets are currently scarce regarding spine surgery specifically. This severely limits the progress of machine learning implementation in spine surgery.

6. Conclusions

As artificial-intelligence-based decision-making tools develop, the availability of databases and modern hybrid algorithms can accomplish intricate and highly complex tasks to help clinicians with their daily practice. The various models published to date can assist spine surgeons in predicting the outcomes of treatments, estimating the probability of failure, and detecting disease patterns in multimodal data. Implementing "good ma-

chine learning principles" and strengthening cooperation between healthcare providers and industries will be beneficial for the use of modern machine learning algorithms in clinics. The integration of multimodal data in novel machine learning hybrid models might better reflect patient information, and more research in this field is highly warranted. Data repositories containing data from different institutes can help researchers develop such algorithms better. Various techniques, including Federated Learning and crowdsourcing, can be beneficial for an unbiased implementation. However, although these algorithms might be developed using large-scale data, they need to be still questioned and only considered as decision-supporting tools in clinics. Continuous updates are necessary that integrate new data and research into existing algorithms. Following these recommendations could hasten the development process and lead to a safer integration of artificial intelligence in healthcare environments. This narrative literature review highlighted the benefits of machine learning implementations in spine surgery, focusing on multimodal data. The findings reveal promising prior research to develop multi-input mixed-data hybrid decision-supporting models. As such, their implementation into clinics seems to be only a matter of time.

Author Contributions: Conceptualization, F.H., G.M.L., S.C.-D. and B.S.; methodology, B.S., S.Ü. and A.Z.; validation, F.H., G.M.L., S.C.-D., M.B., V.S. and P.O.; formal analysis, G.M.L., F.H., S.C.-D., V.S., M.B. and P.O., investigation, B.S., S.Ü. and A.Z.; resources, F.H., G.M.L. and S.C.-D.; data curation, B.S., V.S., S.Ü. and A.Z.; writing—original draft preparation, B.S. and S.Ü.; writing—review and editing, F.H., G.M.L., M.B., S.C.-D., V.S. and P.O.; visualization, B.S. and S.Ü.; supervision, F.H., G.M.L. and S.C.-D.; project administration, F.H., G.M.L. and S.C.-D.; funding acquisition, F.H., G.M.L. and B.S. All authors have read and agreed to the published version of the manuscript.

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

References

1. Saravi, B.; Li, Z.; Lang, C.N.; Schmid, B.; Lang, F.K.; Grad, S.; Alini, M.; Richards, R.G.; Schmal, H.; Südkamp, N.; et al. The Tissue Renin-Angiotensin System and Its Role in the Pathogenesis of Major Human Diseases: Quo Vadis? *Cells* **2021**, *10*, 650. [CrossRef] [PubMed]
2. Wu, A.; March, L.; Zheng, X.; Huang, J.; Wang, X.; Zhao, J.; Blyth, F.M.; Smith, E.; Buchbinder, R.; Hoy, D. Global Low Back Pain Prevalence and Years Lived with Disability from 1990 to 2017: Estimates from the Global Burden of Disease Study 2017. *Ann. Transl. Med.* **2020**, *8*, 299. [CrossRef] [PubMed]
3. Archer, K.R.; Coronado, R.A.; Haug, C.M.; Vanston, S.W.; Devin, C.J.; Fonnesbeck, C.J.; Aaronson, O.S.; Cheng, J.S.; Skolasky, R.L.; Riley, L.H.; et al. A Comparative Effectiveness Trial of Postoperative Management for Lumbar Spine Surgery: Changing Behavior through Physical Therapy (CBPT) Study Protocol. *BMC Musculoskelet. Disord.* **2014**, *15*, 325. [CrossRef] [PubMed]
4. Martin, B.I.; Mirza, S.K.; Comstock, B.A.; Gray, D.T.; Kreuter, W.; Deyo, R.A. Reoperation Rates Following Lumbar Spine Surgery and the Influence of Spinal Fusion Procedures. *Spine* **2007**, *32*, 382–387. [CrossRef]
5. Mallappallil, M.; Sabu, J.; Gruessner, A.; Salifu, M. A Review of Big Data and Medical Research. *SAGE Open Med.* **2020**, *8*, 2050312120934839. [CrossRef] [PubMed]
6. Marcus, G. Deep Learning: A Critical Appraisal. *arXiv* **2018**, arXiv:1801.00631. [CrossRef]
7. Ford, M. *Architects of Intelligence: The Truth about AI from the People Building It*; Packt Publishing: Birmingham, UK, 2018; ISBN 978-1-78913-126-0.
8. Cutillo, C.M.; Sharma, K.R.; Foschini, L.; Kundu, S.; Mackintosh, M.; Mandl, K.D. Machine Intelligence in Healthcare—Perspectives on Trustworthiness, Explainability, Usability, and Transparency. *NPJ Digit. Med.* **2020**, *3*, 47. [CrossRef]
9. Saravi, B.; Lang, G.; Ülkümen, S.; Südkamp, N.; Hassel, F. Case-Matched Radiological and Clinical Outcome Evaluation of Interlaminar versus Microsurgical Decompression of Lumbar Spinal Stenosis. Meeting Abstract. German Congress of Orthopaedics and Traumatology (DKOU 2021). 2021. Available online: https://doi.org/10.3205/21DKOU024 (accessed on 11 March 2022).
10. Finkelstein, J.A.; Schwartz, C.E. Patient-Reported Outcomes in Spine Surgery: Past, Current, and Future Directions. *J. Neurosurg. Spine* **2019**, *31*, 155–164. [CrossRef]

11. Breakwell, L.M.; Cole, A.A.; Birch, N.; Heywood, C. Should We All Go to the PROM? The First Two Years of the British Spine Registry. *Bone Jt. J.* **2015**, *97*, 871–874. [CrossRef]

12. MacLean, D.L.; Heer, J. Identifying Medical Terms in Patient-Authored Text: A Crowdsourcing-Based Approach. *J. Am. Med. Inform. Assoc.* **2013**, *20*, 1120–1127. [CrossRef]

13. Warby, S.C.; Wendt, S.L.; Welinder, P.; Munk, E.G.S.; Carrillo, O.; Sorensen, H.B.D.; Jennum, P.; Peppard, P.E.; Perona, P.; Mignot, E. Sleep-Spindle Detection: Crowdsourcing and Evaluating Performance of Experts, Non-Experts and Automated Methods. *Nat. Methods* **2014**, *11*, 385–392. [CrossRef]

14. Mavandadi, S.; Dimitrov, S.; Feng, S.; Yu, F.; Sikora, U.; Yaglidere, O.; Padmanabhan, S.; Nielsen, K.; Ozcan, A. Distributed Medical Image Analysis and Diagnosis through Crowd-Sourced Games: A Malaria Case Study. *PLoS ONE* **2012**, *7*, e37245. [CrossRef] [PubMed]

15. Crump, M.J.C.; McDonnell, J.V.; Gureckis, T.M. Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLoS ONE* **2013**, *8*, e57410. [CrossRef]

16. Bartneck, C.; Duenser, A.; Moltchanova, E.; Zawieska, K. Comparing the Similarity of Responses Received from Studies in Amazon's Mechanical Turk to Studies Conducted Online and with Direct Recruitment. *PLoS ONE* **2015**, *10*, e0121595. [CrossRef]

17. Wang, C.; Han, L.; Stein, G.; Day, S.; Bien-Gund, C.; Mathews, A.; Ong, J.J.; Zhao, P.-Z.; Wei, S.-F.; Walker, J.; et al. Crowdsourcing in Health and Medical Research: A Systematic Review. *Infect. Dis. Poverty* **2020**, *9*, 8. [CrossRef] [PubMed]

18. GDC Data Portal. Available online: Https://Portal.Gdc.Cancer.Gov (accessed on 11 March 2022).

19. Phan, N.N.; Chattopadhyay, A.; Lu, T.-P.; Tsai, M.-H. Leveraging Well-Annotated Databases for Deep Learning in Biomedical Research. *Transl. Cancer Res. TCR* **2020**, *9*, 7682–7684. [CrossRef] [PubMed]

20. Kaggle. Available online: https://www.kaggle.com/ (accessed on 11 March 2022).

21. GitHub. Available online: https://github.com (accessed on 11 March 2022).

22. Nature Scientific Data. Available online: https://www.nature.com/sdata (accessed on 11 March 2022).

23. Oxford Academic GigaScience. Available online: https://academic.oup.com/gigascience (accessed on 11 March 2022).

24. Osteoarthritis Initiative (OAI) Database. Available online: https://nda.nih.gov/oai (accessed on 11 March 2022).

25. Austrian Spinal Cord Injury Study. Available online: https://www.ascis.at (accessed on 11 March 2022).

26. Sebastian, A.S. Database Research in Spine Surgery. *Clin. Spine Surg. Spine Publ.* **2016**, *29*, 427–429. [CrossRef] [PubMed]

27. Asher, A.L.; Knightly, J.; Mummaneni, P.V.; Alvi, M.A.; McGirt, M.J.; Yolcu, Y.U.; Chan, A.K.; Glassman, S.D.; Foley, K.T.; Slotkin, J.R.; et al. Quality Outcomes Database Spine Care Project 2012–2020: Milestones Achieved in a Collaborative North American Outcomes Registry to Advance Value-Based Spine Care and Evolution to the American Spine Registry. *Neurosurg. Focus* **2020**, *48*, E2. [CrossRef]

28. SORG. Available online: https://sorg.mgh.harvard.edu (accessed on 11 March 2022).

29. Sadilek, A.; Liu, L.; Nguyen, D.; Kamruzzaman, M.; Serghiou, S.; Rader, B.; Ingerman, A.; Mellem, S.; Kairouz, P.; Nsoesie, E.O.; et al. Privacy-First Health Research with Federated Learning. *NPJ Digit. Med.* **2021**, *4*, 132. [CrossRef]

30. Yang, Q.; Liu, Y.; Chen, T.; Tong, Y. Federated Machine Learning: Concept and Applications. *ACM Trans. Intell. Syst. Technol.* **2019**, *10*, 1–19. [CrossRef]

31. Li, T.; Sahu, A.K.; Talwalkar, A.; Smith, V. Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Processing Mag.* **2020**, *37*, 50–60. [CrossRef]

32. Kairouz, P.; McMahan, H.B.; Avent, B.; Bellet, A.; Bennis, M.; Nitin Bhagoji, A.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. Advances and Open Problems in Federated Learning. *FNT Mach. Learn.* **2021**, *14*, 1–210. [CrossRef]

33. Rieke, N.; Hancox, J.; Li, W.; Milletarì, F.; Roth, H.R.; Albarqouni, S.; Bakas, S.; Galtier, M.N.; Landman, B.A.; Maier-Hein, K.; et al. The Future of Digital Health with Federated Learning. *NPJ Digit. Med.* **2020**, *3*, 119. [CrossRef] [PubMed]

34. Li, W.; Milletarì, F.; Xu, D.; Rieke, N.; Hancox, J.; Zhu, W.; Baust, M.; Cheng, Y.; Ourselin, S.; Cardoso, M.J.; et al. Privacy-Preserving Federated Brain Tumour Segmentation. In *International Workshop on Machine Learning in Medical Imaging*; Suk, H.-I., Liu, M., Yan, P., Lian, C., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 133–141.

35. Sheller, M.J.; Reina, G.A.; Edwards, B.; Martin, J.; Bakas, S. Multi-Institutional Deep Learning Modeling Without Sharing Patient Data: A Feasibility Study on Brain Tumor Segmentation. In *International MICCAI Brainlesion Workshop*; Springer: Cham, Switzerland, 2018; pp. 92–104.

36. Wang, F.; Casalino, L.P.; Khullar, D. Deep Learning in Medicine—Promise, Progress, and Challenges. *JAMA Intern. Med.* **2019**, *179*, 293. [CrossRef] [PubMed]

37. Bohr, A.; Memarzadeh, K. The Rise of Artificial Intelligence in Healthcare Applications. In *Artificial Intelligence in Healthcare*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 25–60. ISBN 978-0-12-818438-7.

38. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. *J. Big Data* **2021**, *8*, 53. [CrossRef]

39. Dreiseitl, S.; Ohno-Machado, L. Logistic Regression and Artificial Neural Network Classification Models: A Methodology Review. *J. Biomed. Inform.* **2002**, *35*, 352–359. [CrossRef]

40. Sharma, A.; Vans, E.; Shigemizu, D.; Boroevich, K.A.; Tsunoda, T. DeepInsight: A Methodology to Transform a Non-Image Data to an Image for Convolution Neural Network Architecture. *Sci. Rep.* **2019**, *9*, 11399. [CrossRef]

41. Ma, S.; Zhang, Z. OmicsMapNet: Transforming Omics Data to Take Advantage of Deep Convolutional Neural Network for Discovery. *arXiv* **2018**, arXiv:1804.05283.

42. van der Maaten, L.; Hinton, G. Visualizing High-Dimensional Data Using t-Sne. 4. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.

43. Bazgir, O.; Zhang, R.; Dhruba, S.R.; Rahman, R.; Ghosh, S.; Pal, R. Representation of Features as Images with Neighborhood Dependencies for Compatibility with Convolutional Neural Networks. *Nat. Commun.* **2020**, *11*, 4391. [CrossRef]

44. Freidin, M.B.; Tsepilov, Y.A.; Palmer, M.; Karssen, L.C.; Suri, P.; Aulchenko, Y.S.; Williams, F.M.K.; CHARGE Musculoskeletal Working Group. Insight into the Genetic Architecture of Back Pain and Its Risk Factors from a Study of 509,000 Individuals. *Pain* **2019**, *160*, 1361–1373. [CrossRef] [PubMed]

45. Kanber, B. Sparse Data to Structured Imageset Transformation. *arXiv* **2020**, arXiv:2005.10045.

46. Pölsterl, S.; Wolf, T.N.; Wachinger, C. Combining 3D Image and Tabular Data via the Dynamic Affine Feature Map Transform. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2021; Volume 12905, pp. 688–698. [CrossRef]

47. Hao, J.; Kosaraju, S.C.; Tsaku, N.Z.; Song, D.H.; Kang, M. PAGE-Net: Interpretable and Integrative Deep Learning for Survival Analysis Using Histopathological Images and Genomic Data. *Pac. Symp. Biocomput.* **2020**, *25*, 355–366. [PubMed]

48. Mobadersany, P.; Yousefi, S.; Amgad, M.; Gutman, D.A.; Barnholtz-Sloan, J.S.; Velázquez Vega, J.E.; Brat, D.J.; Cooper, L.A.D. Predicting Cancer Outcomes from Histology and Genomics Using Convolutional Networks. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E2970–E2979. [CrossRef]

49. El-Sappagh, S.; Abuhmed, T.; Riazul Islam, S.M.; Kwak, K.S. Multimodal Multitask Deep Learning Model for Alzheimer's Disease Progression Detection Based on Time Series Data. *Neurocomputing* **2020**, *412*, 197–215. [CrossRef]

50. Li, S.; Shi, H.; Sui, D.; Hao, A.; Qin, H. A Novel Pathological Images and Genomic Data Fusion Framework for Breast Cancer Survival Prediction. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2020**, *2020*, 1384–1387. [CrossRef]

51. Spasov, S.; Passamonti, L.; Duggento, A.; Liò, P.; Toschi, N. Alzheimer's Disease Neuroimaging Initiative A Parameter-Efficient Deep Learning Approach to Predict Conversion from Mild Cognitive Impairment to Alzheimer's Disease. *Neuroimage* **2019**, *189*, 276–287. [CrossRef]

52. Shavlokhova, V.; Vollmer, M.; Vollmer, A.; Gholam, P.; Saravi, B.; Hoffmann, J.; Engel, M.; Elsner, J.; Neumeier, F.; Freudlsperger, C. In Vivo Reflectance Confocal Microscopy of Wounds: Feasibility of Intraoperative Basal Cell Carcinoma Margin Assessment. *Ann. Transl. Med.* **2021**, *9*, 1716. [CrossRef]

53. Stead, W.W. Clinical Implications and Challenges of Artificial Intelligence and Deep Learning. *JAMA* **2018**, *320*, 1107. [CrossRef]

54. Huang, S.-C.; Pareek, A.; Seyyedi, S.; Banerjee, I.; Lungren, M.P. Fusion of Medical Imaging and Electronic Health Records Using Deep Learning: A Systematic Review and Implementation Guidelines. *NPJ Digit. Med.* **2020**, *3*, 136. [CrossRef]

55. Manning, T.; Sleator, R.D.; Walsh, P. Biologically Inspired Intelligent Decision Making: A Commentary on the Use of Artificial Neural Networks in Bioinformatics. *Bioengineered* **2014**, *5*, 80–95. [CrossRef] [PubMed]

56. Li, Y.; Zhang, T.; Liu, Z.; Hu, H. A Concatenating Framework of Shortcut Convolutional Neural Networks. *arXiv* **2017**, arXiv:1710.00974.

57. Vasile, C.M.; Udriștoiu, A.L.; Ghenea, A.E.; Popescu, M.; Gheonea, C.; Niculescu, C.E.; Ungureanu, A.M.; Udriștoiu, Ș.; Drocaș, A.I.; Gruionu, L.G.; et al. Intelligent Diagnosis of Thyroid Ultrasound Imaging Using an Ensemble of Deep Learning Methods. *Medicina* **2021**, *57*, 395. [CrossRef] [PubMed]

58. Udriștoiu, A.L.; Ghenea, A.E.; Udriștoiu, Ș.; Neaga, M.; Zlatian, O.M.; Vasile, C.M.; Popescu, M.; Țieranu, E.N.; Salan, A.-I.; Turcu, A.A.; et al. COVID-19 and Artificial Intelligence: An Approach to Forecast the Severity of Diagnosis. *Life* **2021**, *11*, 1281. [CrossRef]

59. Shavlokhova, V.; Sandhu, S.; Flechtenmacher, C.; Koveshazi, I.; Neumeier, F.; Padrón-Laso, V.; Jonke, Ž.; Saravi, B.; Vollmer, M.; Vollmer, A.; et al. Deep Learning on Oral Squamous Cell Carcinoma Ex Vivo Fluorescent Confocal Microscopy Data: A Feasibility Study. *J. Clin. Med.* **2021**, *10*, 5326. [CrossRef]

60. Yuan, Z.; Jiang, Y.; Li, J.; Huang, H. Hybrid-DNNs: Hybrid Deep Neural Networks for Mixed Inputs. *arXiv* **2020**, arXiv:2005.08419.

61. Kharazmi, P.; Kalia, S.; Lui, H.; Wang, Z.J.; Lee, T.K. A Feature Fusion System for Basal Cell Carcinoma Detection through Data-Driven Feature Learning and Patient Profile. *Ski. Res. Technol.* **2018**, *24*, 256–264. [CrossRef]

62. Yap, J.; Yolland, W.; Tschandl, P. Multimodal Skin Lesion Classification Using Deep Learning. *Exp. Dermatol.* **2018**, *27*, 1261–1267. [CrossRef]

63. Purwar, S.; Tripathi, R.K.; Ranjan, R.; Saxena, R. Detection of Microcytic Hypochromia Using Cbc and Blood Film Features Extracted from Convolution Neural Network by Different Classifiers. *Multimed. Tools Appl.* **2020**, *79*, 4573–4595. [CrossRef]

64. Nie, D.; Lu, J.; Zhang, H.; Adeli, E.; Wang, J.; Yu, Z.; Liu, L.; Wang, Q.; Wu, J.; Shen, D. Multi-Channel 3D Deep Feature Learning for Survival Time Prediction of Brain Tumor Patients Using Multi-Modal Neuroimages. *Sci. Rep.* **2019**, *9*, 1103. [CrossRef]

65. Hyun, S.H.; Ahn, M.S.; Koh, Y.W.; Lee, S.J. A Machine-Learning Approach Using PET-Based Radiomics to Predict the Histological Subtypes of Lung Cancer. *Clin. Nucl. Med.* **2019**, *44*, 956–960. [CrossRef] [PubMed]

66. Bhagwat, N.; Viviano, J.D.; Voineskos, A.N.; Chakravarty, M.M. Alzheimer's Disease Neuroimaging Initiative Modeling and Prediction of Clinical Symptom Trajectories in Alzheimer's Disease Using Longitudinal Data. *PLoS Comput. Biol.* **2018**, *14*, e1006376. [CrossRef] [PubMed]

67. Liu, J.; Chen, Y.; Lan, L.; Lin, B.; Chen, W.; Wang, M.; Li, R.; Yang, Y.; Zhao, B.; Hu, Z.; et al. Prediction of Rupture Risk in Anterior Communicating Artery Aneurysms with a Feed-Forward Artificial Neural Network. *Eur. Radiol.* **2018**, *28*, 3268–3275. [CrossRef] [PubMed]

68. Berjano, P.; Langella, F.; Ventriglia, L.; Compagnone, D.; Barletta, P.; Huber, D.; Mangili, F.; Licandro, G.; Galbusera, F.; Cina, A.; et al. The Influence of Baseline Clinical Status and Surgical Strategy on Early Good to Excellent Result in Spinal Lumbar Arthrodesis: A Machine Learning Approach. *JPM* **2021**, *11*, 1377. [CrossRef]

69. Babaee, M.; Soleimani, P.; Zali, A. A Predictive Model for Assessment of Successful Outcome in Posterior Spinal Fusion Surgery. *Int. Clin. Neurosci. J.* **2018**, *4*, 143–151. [CrossRef]

70. Finkelstein, J.A.; Stark, R.B.; Lee, J.; Schwartz, C.E. Patient Factors That Matter in Predicting Spine Surgery Outcomes: A Machine Learning Approach. *J. Neurosurg. Spine* **2021**, *35*, 127–136. [CrossRef]

71. Hoffman, H.; Lee, S.I.; Garst, J.H.; Lu, D.S.; Li, C.H.; Nagasawa, D.T.; Ghalehsari, N.; Jahanforouz, N.; Razaghy, M.; Espinal, M.; et al. Use of Multivariate Linear Regression and Support Vector Regression to Predict Functional Outcome after Surgery for Cervical Spondylotic Myelopathy. *J. Clin. Neurosci.* **2015**, *22*, 1444–1449. [CrossRef]

72. Janssen, E.R.; Osong, B.; van Soest, J.; Dekker, A.; van Meeteren, N.L.; Willems, P.C.; Punt, I.M. Exploring Associations of Preoperative Physical Performance With Postoperative Outcomes After Lumbar Spinal Fusion: A Machine Learning Approach. *Arch. Phys. Med. Rehabil.* **2021**, *102*, 1324–1330.e3. [CrossRef]

73. Khan, O.; Badhiwala, J.H.; Akbar, M.A.; Fehlings, M.G. Prediction of Worse Functional Status After Surgery for Degenerative Cervical Myelopathy: A Machine Learning Approach. *Neurosurgery* **2021**, *88*, 584–591. [CrossRef]

74. Khan, O.; Badhiwala, J.H.; Witiw, C.D.; Wilson, J.R.; Fehlings, M.G. Machine Learning Algorithms for Prediction of Health-Related Quality-of-Life after Surgery for Mild Degenerative Cervical Myelopathy. *Spine J.* **2021**, *21*, 1659–1669. [CrossRef]

75. Khor, S.; Lavallee, D.; Cizik, A.M.; Bellabarba, C.; Chapman, J.R.; Howe, C.R.; Lu, D.; Mohit, A.A.; Oskouian, R.J.; Roh, J.R.; et al. Development and Validation of a Prediction Model for Pain and Functional Outcomes After Lumbar Spine Surgery. *JAMA Surg.* **2018**, *153*, 634–642. [CrossRef] [PubMed]

76. Lewandrowski, K.-U.; Muraleedharan, N.; Eddy, S.A.; Sobti, V.; Reece, B.D.; Ramírez León, J.F.; Shah, S. Artificial Intelligence Comparison of the Radiologist Report With Endoscopic Predictors of Successful Transforaminal Decompression for Painful Conditions of the Lumber Spine: Application of Deep Learning Algorithm Interpretation of Routine Lumbar Magnetic Resonance Imaging Scan. *Int. J. Spine Surg.* **2020**, *14*, S75–S85. [CrossRef] [PubMed]

77. McGirt, M.J.; Sivaganesan, A.; Asher, A.L. Prediction Model for Outcome after Low-Back Surgery: Individualized Likelihood of Complication, Hospital Readmission, Return to Work, and 12-Month Improvement in Functional Disability. *Neurosurg. Focus.* **2015**, *39*, E13. [CrossRef]

78. Merali, Z.G.; Witiw, C.D.; Badhiwala, J.H.; Wilson, J.R.; Fehlings, M.G. Using a Machine Learning Approach to Predict Outcome after Surgery for Degenerative Cervical Myelopathy. *PLoS ONE* **2019**, *14*, e0215133. [CrossRef] [PubMed]

79. Oh, T.; Scheer, J.K.; Smith, J.S. Potential of Predictive Computer Models for Preoperative Patient Selection to Enhance Overall Quality-adjusted Life Years Gained at 2-Year Follow-up: A Simulation in 234 Patients with Adult Spinal Deformity. *Neurosurg. Focus.* **2017**, *43*, E2. [CrossRef]

80. Pedersen, C.F.; Andersen, M.Ø.; Carreon, L.Y.; Eiskjær, S. Applied Machine Learning for Spine Surgeons: Predicting Outcome for Patients Undergoing Treatment for Lumbar Disc Herniation Using PRO Data. *Glob. Spine J.* **2020**, *11*, 219256822096764. [CrossRef]

81. Siccoli, A.; de Wispelaere, M.P.; Schröder, M.L. Machine Learning– Based Preoperative Predictive Analytics for Lumbar Spinal Stenosis. *Neurosurg. Focus* **2019**, *46*, 5. [CrossRef]

82. Staartjes, V.E.; de Wispelaere, M.P.; Vandertop, W.P. Deep Learning-Based Preoperative Predictive Analytics for Patient-Reported Outcomes Following Lumbar Discectomy: Feas9ibility of Center-Specific Modeling. *Spine J.* **2019**, *19*, 853–861. [CrossRef]

83. Staartjes, V.E.; Stumpo, V.; Ricciardi, L.; Maldaner, N.; Eversdijk, H.A.J.; Vieli, M.; Ciobanu-Caraus, O.; Raco, A.; Miscusi, M.; Perna, A.; et al. FUSE-ML: Development and External Validation of a Clinical Prediction Model for Mid-Term Outcomes after Lumbar Spinal Fusion for Degenerative Disease. *Eur. Spine J.* **2022**. [CrossRef]

84. Wirries, A.; Geiger, F.; Hammad, A.; Oberkircher, L.; Blümcke, I.; Jabari, S. Artificial Intelligence Facilitates Decision-Making in the Treatment of Lumbar Disc Herniations. *Eur. Spine J.* **2021**, *30*, 2176–2184. [CrossRef]

85. André, A.; Peyrou, B.; Carpentier, A.; Vignaux, J.-J. Feasibility and Assessment of a Machine Learning-Based Predictive Model of Outcome After Lumbar Decompression Surgery. *Glob. Spine J.* **2020**, *11*, 219256822096937. [CrossRef] [PubMed]

86. Arvind, V.; Kim, J.S.; Oermann, E.K. Predicting Surgical Complications in Adult Patients Undergoing Anterior Cervical Discectomy and Fusion Using Machine Learning. *Neurospine* **2018**, *15*, 329–337. [CrossRef] [PubMed]

87. Bekelis, K.; Desai, A.; Bakhoum, S.F.; Missios, S. A Predictive Model of Complications after Spine Surgery: The National Surgical Quality Improvement Program (NSQIP) 2005–2010. *Spine J.* **2014**, *14*, 1247–1255. [CrossRef]

88. Han, S.S.; Azad, T.D.; Suarez, P.A.; Ratliff, J.K. A Machine Learning Approach for Predictive Models of Adverse Events Following Spine Surgery. *Spine J.* **2019**, *19*, 1772–1781. [CrossRef] [PubMed]

89. Karhade, A.V.; Bongers, M.E.R.; Groot, O.Q.; Cha, T.D.; Doorly, T.P.; Fogel, H.A.; Hershman, S.H.; Tobert, D.G.; Srivastava, S.D.; Bono, C.M.; et al. Development of Machine Learning and Natural Language Processing Algorithms for Preoperative Prediction and Automated Identification of Intraoperative Vascular Injury in Anterior Lumbar Spine Surgery. *Spine J.* **2021**, *21*, 1635–1642. [CrossRef] [PubMed]

90. Kim, J.S.; Arvind, V.; Oermann, E.K. Predicting Surgical Complications in Patients Undergoing Elective Adult Spinal Deformity Procedures Using Machine Learning. *Spine Deform.* **2018**, *6*, 762–770. [CrossRef] [PubMed]

91. Kim, J.S.; Merrill, R.K.; Arvind, V.; Kaji, D.; Pasik, S.D.; Nwachukwu, C.C.; Vargas, L.; Osman, N.S.; Oermann, E.K.; Caridi, J.M.; et al. Examining the Ability of Artificial Neural Networks Machine Learning Models to Accurately Predict Complications Following Posterior Lumbar Spine Fusion. *Spine* **2018**, *43*, 853–860. [CrossRef]

92. Massaad, E.; Williams, N.; Hadzipasic, M.; Patel, S.S.; Fourman, M.S.; Kiapour, A.; Schoenfeld, A.J.; Shankar, G.M.; Shin, J.H. Performance Assessment of the Metastatic Spinal Tumor Frailty Index Using Machine Learning Algorithms: Limitations and Future Directions. *Neurosurg. Focus* **2021**, *50*, E5. [CrossRef]

93. Ratliff, J.K.; Balise, R.; Veeravagu, A. Predicting Occurrence of Spine Surgery Complications Using "Big Data" Modeling of an Administrative Claims Database. *J. Bone Jt. Surg.* **2016**, *98*, 824–834. [CrossRef]

94. Shah, A.A.; Devana, S.K.; Lee, C.; Bugarin, A.; Lord, E.L.; Shamie, A.N.; Park, D.Y.; van der Schaar, M.; SooHoo, N.F. Machine Learning-Driven Identification of Novel Patient Factors for Prediction of Major Complications after Posterior Cervical Spinal Fusion. *Eur. Spine J.* **2021**, *11*, 1–8. [CrossRef]

95. Veeramani, A.; Zhang, A.S.; Blackburn, A.Z.; Etzel, C.M.; DiSilvestro, K.J.; McDonald, C.L.; Daniels, A.H. An Artificial Intelligence Approach to Predicting Unplanned Intubation Following Anterior Cervical Discectomy and Fusion. *Glob. Spine J.* **2022**, *11*, 21925682211053590. [CrossRef]

96. Wang, K.Y.; Ikwuezunma, I.; Puvanesarajah, V.; Babu, J.; Margalit, A.; Raad, M.; Jain, A. Using Predictive Modeling and Supervised Machine Learning to Identify Patients at Risk for Venous Thromboembolism Following Posterior Lumbar Fusion. *Glob. Spine J.* **2021**, *11*, 21925682211019360. [CrossRef]

97. Wong, D.J.N.; Oliver, C.M.; Moonesinghe, S.R. Predicting Postoperative Morbidity in Adult Elective Surgical Patients Using the Surgical Outcome Risk Tool (SORT). *Br. J. Anaesth.* **2017**, *119*, 95–105. [CrossRef] [PubMed]

98. Shah, A.A.; Karhade, A.V.; Bono, C.M. Development of a Machine Learning Algorithm for Prediction of Failure of Nonoperative Management in Spinal Epidural Abscess. *Spine J.* **2019**, *19*, 1657–1665. [CrossRef]

99. Aldebeyan, S.; Aoude, A.; Fortin, M. Predictors of Discharge Destination after Lumbar Spine Fusion Surgery. *Spine* **2016**, *41*, 1535–1541. [CrossRef] [PubMed]

100. Goyal, A.; Ngufor, C.; Kerezoudis, P. Can Machine Learning Algorithms Accurately Predict Discharge to Nonhome Facility and Early Unplanned Readmissions Following Spinal Fusion? Analysis of a National Surgical Registry. *J. Neurosurg. Spine* **2019**, *31*, 568–578. [CrossRef] [PubMed]

101. Karhade, A.V.; Ogink, P.; Thio, Q. Development of Machine Learning Algorithms for Prediction of Discharge Disposition after Elective Inpatient Surgery for Lumbar Degenerative Disc Disorders. *Neurosurg. Focus.* **2018**, *45*, E6. [CrossRef]

102. Karnuta, J.M.; Golubovsky, J.L.; Haeberle, H.S. Can a Machine Learning Model Accurately Predict Patient Resource Utilization Following Lumbar Spinal Fusion? *Spine J.* **2019**, *20*, 329–336. [CrossRef]

103. Ogink, P.T.; Karhade, A.V.; Thio, Q.C.B.S.; Gormley, W.B.; Oner, F.C.; Verlaan, J.J.; Schwab, J.H. Predicting Discharge Placement after Elective Surgery for Lumbar Spinal Stenosis Using Machine Learning Methods. *Eur. Spine J.* **2019**, *28*, 1433–1440. [CrossRef] [PubMed]

104. Ogink, P.T.; Karhade, A.V.; Thio, Q.C.B.S. Development of a Machine Learning Algorithm Predicting Discharge Placement after Surgery for Spondylolisthesis. *Eur. Spine J.* **2019**, *28*, 1775–1782. [CrossRef]

105. Stopa, B.M.; Robertson, F.C.; Karhade, A.V.; Chua, M.; Broekman, M.L.D.; Schwab, J.H.; Smith, T.R.; Gormley, W.B. Predicting Nonroutine Discharge after Elective Spine Surgery: External Validation of Machine Learning Algorithms. *J. Neurosurg. Spine* **2019**, *31*, 742–747. [CrossRef] [PubMed]

106. Russo, G.S.; Canseco, J.A.; Chang, M.; Levy, H.A.; Nicholson, K.; Karamian, B.A.; Mangan, J.; Fang, T.; Vaccaro, A.R.; Kepler, C.K. A Novel Scoring System to Predict Length of Stay After Anterior Cervical Discectomy and Fusion. *J. Am. Acad. Orthop. Surg.* **2021**, *29*, 758–766. [CrossRef]

107. Zhang, A.S.; Veeramani, A.; Quinn, M.S.; Alsoof, D.; Kuris, E.O.; Daniels, A.H. Machine Learning Prediction of Length of Stay in Adult Spinal Deformity Patients Undergoing Posterior Spine Fusion Surgery. *J. Clin. Med.* **2021**, *10*, 4074. [CrossRef] [PubMed]

108. Hopkins, B.S.; Yamaguchi, J.T.; Garcia, R. Using Machine Learning to Predict 30-Day Readmissions after Posterior Lumbar Fusion: An NSQIP Study Involving 23,264 Patients [Published Ahead of Print. *Neurosurg. Spine* **2019**, *32*, 399–406. [CrossRef] [PubMed]

109. Kalagara, S.; Eltorai, A.E.M.; Durand, W.M.; DePasse, J.M.; Daniels, A.H. Machine Learning Modeling for Predicting Hospital Readmission Following Lumbar Laminectomy. *J. Neurosurg. Spine* **2018**, *30*, 344–352. [CrossRef]

110. Karhade, A.V.; Lavoie-Gagne, O.; Agaronnik, N.; Ghaednia, H.; Collins, A.K.; Shin, D.; Schwab, J.H. Natural Language Processing for Prediction of Readmission in Posterior Lumbar Fusion Patients: Which Free-Text Notes Have the Most Utility? *Spine J.* **2022**, *22*, 272–277. [CrossRef]

111. Kuris, E.O.; Veeramani, A.; McDonald, C.L.; DiSilvestro, K.J.; Zhang, A.S.; Cohen, E.M.; Daniels, A.H. Predicting Readmission After Anterior, Posterior, and Posterior Interbody Lumbar Spinal Fusion: A Neural Network Machine Learning Approach. *World Neurosurg.* **2021**, *151*, e19–e27. [CrossRef]

112. Nunes, A.A.; Pinheiro, R.P.; Costa, H.R.T.; Defino, H.L.A. Predictors of Hospital Readmission within 30 Days after Surgery for Thoracolumbar Fractures: A Mixed Approach. *Health Plan. Manag.* **2022**, *11*, 2192568220969373. [CrossRef]

113. Karhade, A.V.; Thio, Q.C.B.S.; Ogink, P.T. Predicting 90-Day and 1-Year Mortality in Spinal Metastatic Disease: Development and Internal Validation. *Neurosurgery* **2019**, *85*, 671–681. [CrossRef]

114. Karhade, A.V.; Thio, Q.C.B.S.; Ogink, P.T. Development of Machine Learning Algorithms for Prediction of 30-Day Mortality after Surgery for Spinal Metastasis. *Neurosurgery* **2019**, *85*, 83–91. [CrossRef]

115. Karhade, A.V.; Shah, A.A.; Bono, C.M. Development of Machine Learning Algorithms for Prediction of Mortality in Spinal Epidural Abscess. *Spine J.* **2019**, *19*, 1950–1959. [CrossRef] [PubMed]

116. 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]

117. Yang, J.-J.; Chen, C.-W.; Fourman, M.S.; Bongers, M.E.R.; Karhade, A.V.; Groot, O.Q.; Lin, W.-H.; Yen, H.-K.; Huang, P.-H.; Yang, S.-H.; 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] [PubMed]

118. Hu, M.-H. A Machine Learning Algorithm for Predicting Prolonged Postoperative Opioid Prescription after Lumbar Disc Herniation Surgery. An External Validation Study Using 1,316 Patients from a Taiwanese Cohort. *Spine J.* **2022**. [CrossRef] [PubMed]

119. Karhade, A.V.; Ogink, P.T.; Thio, Q.C.B.S. Development of Machine Learning Algorithms for Prediction of Prolonged Opioid Prescription after Surgery for Lumbar Disc Herniation. *Spine J.* **2019**, *19*, 1764–1771. [CrossRef] [PubMed]

120. Zhang, Y.; Fatemi, P.; Medress, Z.; Azad, T.D.; Veeravagu, A.; Desai, A.; Ratliff, J.K. A Predictive-Modeling Based Screening Tool for Prolonged Opioid Use after Surgical Management of Low Back and Lower Extremity Pain. *Spine J.* **2020**, *20*, 1184–1195. [CrossRef] [PubMed]

121. Dong, S.; Li, W.; Tang, Z.-R.; Wang, H.; Pei, H.; Yuan, B. Development and Validation of a Novel Predictive Model and Web Calculator for Evaluating Transfusion Risk after Spinal Fusion for Spinal Tuberculosis: A Retrospective Cohort Study. *BMC Musculoskelet. Disord.* **2021**, *22*, 825. [CrossRef]

122. Durand, W.M.; DePasse, J.M.; Daniels, A.H. Predictive Modeling for Blood Transfusion After Adult Spinal Deformity Surgery: A Tree-Based Machine Learning Approach. *Spine* **2018**, *43*, 1058–1066. [CrossRef]

123. Papić, M.; Brdar, S.; Papić, V. Return to Work after Lumbar Microdiscectomy—Personalizing Approach through Predictive Modeling. *Stud Health Technol. Inform.* **2016**, *224*, 181–183.

124. Pasha, S.; Shah, S.; Newton, P.; Harms Study Group. Machine Learning Predicts the 3D Outcomes of Adolescent Idiopathic Scoliosis Surgery Using Patient-Surgeon Specific Parameters. *Spine* **2021**, *46*, 579–587. [CrossRef]

125. de Vries, B.C.S.; Hegeman, J.H.; Nijmeijer, W.; Geerdink, J.; Seifert, C.; Groothuis-Oudshoorn, C.G.M. Comparing Three Machine Learning Approaches to Design a Risk Assessment Tool for Future Fractures: Predicting a Subsequent Major Osteoporotic Fracture in Fracture Patients with Osteopenia and Osteoporosis. *Osteoporos. Int.* **2021**, *32*, 437–449. [CrossRef] [PubMed]

126. Varghese, V.; Krishnan, V.; Kumar, G.S. Evaluating Pedicle-Screw Instrumentation Using Decision-Tree Analysis Based on Pullout Strength. *Asian Spine J.* **2018**, *12*, 611–621. [CrossRef] [PubMed]

127. Kuo, C.-Y.; Yu, L.-C.; Chen, H.-C. Comparison of Models for the Prediction of Medical Costs of Spinal Fusion in Taiwan Diagnosisrelated Groups by Machine Learning Algorithms. *Healthc. Inform. Res.* **2018**, *24*, 29. [CrossRef] [PubMed]

128. Chang, M.; Canseco, J.A.; Nicholson, K.J.; Patel, N.; Vaccaro, A.R. The Role of Machine Learning in Spine Surgery: The Future Is Now. *Front. Surg.* **2020**, *7*, 54. [CrossRef] [PubMed]

129. Karhade, A.V.; Schwab, J.H.; Bedair, H.S. Development of Machine Learning Algorithms for Prediction of Sustained Postoperative Opioid Prescriptions after Total Hip Arthroplasty. *J. Arthroplast.* **2019**, *34*, 2272–2277. [CrossRef]

130. Harada, G.K.; Siyaji, Z.K.; Mallow, G.M.; Hornung, A.L.; Hassan, F.; Basques, B.A.; Mohammed, H.A.; Sayari, A.J.; Samartzis, D.; An, H.S. Artificial Intelligence Predicts Disk Re-Herniation Following Lumbar Microdiscectomy: Development of the "RAD" Risk Profile. *Eur. Spine J.* **2021**, *30*, 2167–2175. [CrossRef]

131. Karhade, A.V.; Ogink, P.T.; Thio, Q.C.B.S. Machine Learning for Prediction of Sustained Opioid Prescription after Anterior Cervical Discectomy and Fusion. *Spine J.* **2019**, *19*, 976–983. [CrossRef]

132. Li, W.; Wang, J.; Liu, W.; Xu, C.; Li, W.; Zhang, K.; Su, S.; Li, R.; Hu, Z.; Liu, Q.; et al. Machine Learning Applications for the Prediction of Bone Cement Leakage in Percutaneous Vertebroplasty. *Front. Public Health* **2021**, *9*, 812023. [CrossRef]

133. Maki, S.; Furuya, T.; Yoshii, T.; Egawa, S.; Sakai, K.; Kusano, K.; Nakagawa, Y.; Hirai, T.; Wada, K.; Katsumi, K.; et al. Machine Learning Approach in Predicting Clinically Significant Improvements After Surgery in Patients with Cervical Ossification of the Posterior Longitudinal Ligament. *Spine* **2021**, *46*, 1683–1689. [CrossRef]

134. Passias, P.G.; Vasquez-Montes, D.; Poorman, G.W. Predictive Model for Distal Junctional Kyphosis after Cervical Deformity Surgery. *Spine J.* **2018**, *18*, 2187–2194. [CrossRef]

135. Wang, H.; Tang, Z.-R.; Li, W.; Fan, T.; Zhao, J.; Kang, M.; Dong, R.; Qu, Y. Prediction of the Risk of C5 Palsy after Posterior Laminectomy and Fusion with Cervical Myelopathy Using a Support Vector Machine: An Analysis of 184 Consecutive Patients. *J. Orthop. Surg. Res.* **2021**, *16*, 332. [CrossRef] [PubMed]

136. Wong, A.Y.L.; Harada, G.; Lee, R.; Gandhi, S.D.; Dziedzic, A.; Espinoza-Orias, A.; Parnianpour, M.; Louie, P.K.; Basques, B.; An, H.S.; et al. Preoperative Paraspinal Neck Muscle Characteristics Predict Early Onset Adjacent Segment Degeneration in Anterior Cervical Fusion Patients: A Machine-learning Modeling Analysis. *J. Orthop. Res.* **2021**, *39*, 1732–1744. [CrossRef]

137. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef]

138. Hassanzadeh, T.; Essam, D.; Sarker, R. 2D to 3D Evolutionary Deep Convolutional Neural Networks for Medical Image Segmentation. *IEEE Trans. Med. Imaging* **2021**, *40*, 712–721. [CrossRef] [PubMed]

139. Koumakis, L. Deep Learning Models in Genomics; Are We There Yet? *Comput. Struct. Biotechnol. J.* **2020**, *18*, 1466–1473. [CrossRef]

140. Massaad, E.; Fatima, N.; Hadzipasic, M.; Alvarez-Breckenridge, C.; Shankar, G.M.; Shin, J.H. Predictive Analytics in Spine Oncology Research: First Steps, Limitations, and Future Directions. *Neurospine* **2019**, *16*, 669–677. [CrossRef] [PubMed]

141. Leslie, A.; Jones, A.J.; Goddard, P.R. The Influence of Clinical Information on the Reporting of CT by Radiologists. *Br. J. Radiol.* **2000**, *73*, 1052–1055. [CrossRef]

142. Cohen, M.D. Accuracy of Information on Imaging Requisitions: Does It Matter? *J. Am. Coll. Radiol.* **2007**, *4*, 617–621. [CrossRef]

143. Boonn, W.W.; Langlotz, C.P. Radiologist Use of and Perceived Need for Patient Data Access. *J. Digit Imaging* **2009**, *22*, 357–362. [CrossRef]

144. Austin, C.C.; Brown, S.; Fong, N.; Humphrey, C.; Leahey, A.; Webster, P. Research Data Repositories: Review of Current Features, Gap Analysis, and Recommendations for Minimum Requirements. *IASSIST Q.* **2016**, *39*, 24. [CrossRef]

145. Antani, S.; Long, L.R.; Thoma, G.R. Content-Based Image Retrieval for Large Biomedical Image Archives. *Stud. Health Technol. Inform.* **2004**, *107*, 829–833. [PubMed]

146. Raza, K.; Singh, N.K. A Tour of Unsupervised Deep Learning for Medical Image Analysis. *Curr. Med. Imaging* **2021**, *17*, 1059–1077. [CrossRef] [PubMed]

147. National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease. In *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*; The National Academies Collection: Reports funded by National Institutes of Health; National Academies Press (US): Washington, DC, USA, 2011; ISBN 978-0-309-22222-8.

148. Caballé-Cervigón, N.; Castillo-Sequera, J.L.; Gómez-Pulido, J.A.; Gómez-Pulido, J.M.; Polo-Luque, M.L. Machine Learning Applied to Diagnosis of Human Diseases: A Systematic Review. *Appl. Sci.* **2020**, *10*, 5135. [CrossRef]

149. Machluf, Y.; Tal, O.; Navon, A.; Chaiter, Y. From Population Databases to Research and Informed Health Decisions and Policy. *Front. Public Health* **2017**, *5*, 230. [CrossRef] [PubMed]

150. Hengstler, M.; Enkel, E.; Duelli, S. Applied Artificial Intelligence and Trust—The Case of Autonomous Vehicles and Medical Assistance Devices. *Technol. Forecast. Soc. Change* **2016**, *105*, 105–120. [CrossRef]

151. Wang, W.; Siau, K. Trusting Artificial Intelligence in Healthcare. In Proceedings of the AMCIS, AMCIS 2018, New Orleans, LA, USA, 16–18 August 2018.