applied sciences

MDPI

*Article*

# Delirium Prediction Using Machine Learning Interpretation Method and Its Incorporation into a Clinical Workflow

**Koutarou Matsumoto [1,2,*], Yasunobu Nohara [2,3], Mikako Sakaguchi [4], Yohei Takayama [2,4], Shota Fukushige [5], Hidehisa Soejima [2] and Naoki Nakashima [6]**

1 Biostatistics Center, Graduate School of Medicine, Kurume University, Kurume 830-0011, Japan
2 Institute for Medical Information Research and Analysis, Saiseikai Kumamoto Hospital, Kumamoto 861-4193, Japan
3 Big Data Science and Technology, Faculty of Advanced Science and Technology, Kumamoto University, Kumamoto 860-8555, Japan
4 Department of Nursing, Saiseikai Kumamoto Hospital, Kumamoto 861-4193, Japan
5 Department of Laboratory, Saiseikai Kumamoto Hospital, Kumamoto 861-4193, Japan
6 Medical Information Center, Kyushu University Hospital, Fukuoka 812-8582, Japan
* Correspondence: matsumoto_koutarou@kurume-u.ac.jp; Tel.: +81-942-35-3311

**Abstract:** Delirium in hospitalized patients is a worldwide problem, causing a burden on healthcare professionals and impacting patient prognosis. A machine learning interpretation method (ML interpretation method) presents the results of machine learning predictions and promotes guided decisions. This study focuses on visualizing the predictors of delirium using a ML interpretation method and implementing the analysis results in clinical practice. Retrospective data of 55,389 patients hospitalized in a single acute care center in Japan between December 2017 and February 2022 were collected. Patients were categorized into three analysis populations, according to inclusion and exclusion criteria, to develop delirium prediction models. The predictors were then visualized using Shapley additive explanation (SHAP) and fed back to clinical practice. The machine learning-based prediction of delirium in each population exhibited excellent predictive performance. SHAP was used to visualize the body mass index and albumin levels as critical contributors to delirium prediction. In addition, the cutoff value for age, which was previously unknown, was visualized, and the risk threshold for age was raised. By using the SHAP method, we demonstrated that data-driven decision support is possible using electronic medical record data.

**Keywords:** learning health system; machine learning interpretation method; Shapley additive explanation; delirium; machine learning

check for updates

**Citation:** Matsumoto, K.; Nohara, Y.; Sakaguchi, M.; Takayama, Y.; Fukushige, S.; Soejima, H.; Nakashima, N. Delirium Prediction Using Machine Learning Interpretation Method and Its Incorporation into a Clinical Workflow. *Appl. Sci.* **2023**, *13*, 1564. https://doi.org/10.3390/app13031564

Academic Editors: Chien-Yeh Hsu, Hsiu-An Lee, Chi-Ming Chu and Kuo-Chung Chu

Received: 11 November 2022
Revised: 31 December 2022
Accepted: 23 January 2023
Published: 25 January 2023

**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Learning health systems are self-learning transformation systems in which data accumulated in clinical practice are used to prepare evidence-based medicine [1]. As an example of digital technology's potential to advance a learning health system, a multicenter study involving 43 hospitals demonstrated a 44% reduction in all-cause central-line-associated bloodstream infections through a universal decolonization strategy in intensive care settings that used real-time patient data [2,3]. However, there are limited examples wherein medical support using real-world data has been structured as part of a hospital improvement activity. In this study, delirium was used as a case study for investigating the efficacy of learning health systems in real-world settings.

Delirium has been reported in 8–17% and 11–51% of emergency and surgical patients, respectively [4]. The occurrence of delirium in elderly patients is associated with mortality, institutionalization, dementia after discharge, and other poor outcomes. Therefore, the management of delirium during hospitalization is crucial [5].

Delirium has been reported globally, and many evaluation tools have been developed to identify patients at high risk of delirium at the time of admission, and subsequently implement preventative measures [6–14]. With the accumulation of large amounts of medical data and the development of machine learning techniques, research on machine learning prediction models for predicting delirium has been stimulated [15–26].

Machine learning techniques such as the decision tree ensemble do not need data preprocessing (neither normalization nor standardization), and can handle missing values without listwise deletion. Such models can use considerable data and capture nonlinearity and interaction. In our study on ischemic stroke patients, we confirmed that machine learning can be used to predict prognoses more accurately than using previously developed prognostic scores [27]. However, the "black box" problem, whereby recognizing processes and formulas is difficult, hinders the medical implementation of machine learning-based methods.

In a previous study, by focusing on the high prediction accuracy of machine learning technology and the disease management function of electronic clinical pathways, we developed a method for the management of patients with cerebral hemorrhage after stratifying the risk of developing aspiration pneumonia based on machine learning prediction results [28]. However, owing to the black-box nature of machine learning, the predictive contributing factors could not be well visualized, and the routine use of machine learning technology in clinical workflows was not possible.

Machine learning interpretation methods (ML interpretation methods), which have attracted research attention in recent years, can enable clinicians to interpret predictive models and provide appropriate feedback to analysts, promoting the development of data-driven medicine. Therefore, the use of ML interpretation methods for clinical data has been actively investigated [29–32]. However, because these methods have been developed recently, there are no examples of their operational incorporation into clinical workflows. Although there are previous studies that have also applied interpretation methods to models for predicting delirium [24–26], there are no examples of their implementation in clinical workflows.

Therefore, this study aims to build a data-driven decision support system by incorporating the predictive results of machine learning into daily clinical workflows, using a ML interpretation method. In other words, the objective of this study is to demonstrate that it is possible to develop a learning health system for delirium using an interpretation method.

## 2. Materials and Methods

The transparent reporting of a multivariable prediction model for individual prognosis or diagnosis was used in developing the prediction model [33].

### 2.1. Data Source and Transparency

We extracted clinical variables from clinical data accumulated through routine clinical practice in the hospital database of Saiseikai Kumamoto Hospital, and created an analytical database. The data, the methods used in the analysis, and the materials used to conduct the research will be made available to any researcher for the purpose of reproducing the results or replicating the procedure. Owing to the sensitive nature of the data collected for this study, only requests to utilization the dataset from qualified researchers trained in human subject confidentiality protocols will be sent by the corresponding author.

### 2.2. Study Design and Patients

We analyzed data from patients hospitalized at Saiseikai Kumamoto Hospital, Kumamoto City, Japan. This hospital is an emergency and critical care center in southern Japan. The study was approved by the institutional review board (approval no. 1072). Here, 61,181 consecutive patients hospitalized between December 2017 and February 2022 were studied. Of the 5792 patients excluded from this study, 2963 patients could not be traced, 357 had died within 24 h after admission, 2223 had delirium on admission, and 249 were

less than 18 years old. A total of 55,389 patients were considered in this analysis. Furthermore, for focusing on cases prone to delirium, patients in the analysis were categorized into two groups (Figure 1).

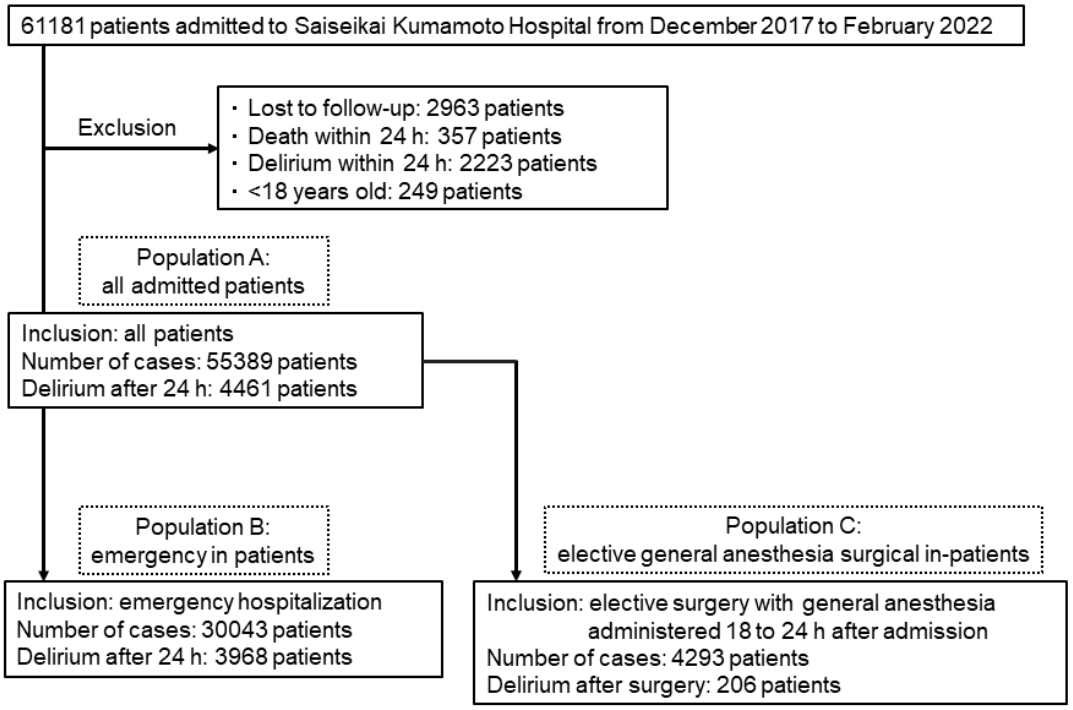



**Figure 1.** Flow chart of the patients.

### *2.3. Clinical Outcomes*

Incident delirium was defined with a positive confusion assessment method (CAM) [34] after 24 h of hospitalization. The occurrence of postoperative delirium was defined with a positive CAM after leaving the operating room. Patients were assessed daily by nurses trained by a specialized delirium subcommittee that included a psychiatrist.

### *2.4. Predictors*

All predictors were obtained prior to the onset of delirium. Predictive models were developed for the three populations subject to analysis, and various predictors were used for each (Tables 1, S1 and S2). In population A, which included all patients, 14 predictors were used. In population B, which included emergency hospitalization patients, 44 predictors were used. In population C, which included elective general anesthesia surgical patients, 23 predictors were used. When developing models for all patients, only basic patient information available on admission, such as age, gender, history of delirium, and treatment plan after admission was used. For emergency hospitalization and elective general anesthesia surgical patients, subject-specific predictors such as laboratory data, vital signs, and intraoperative information were used.

### *2.5. Development of Prediction Models*

Herein, eXtreme gradient boosting [XGBoost], a type of decision tree-based ensemble learning algorithm, was used to develop the prediction model [35]. This algorithm was selected because of its high prediction accuracy, its ability to capture non-linear relationships between outcomes and predictors, and its compatibility with the ML interpretation method used herein, as described later. The details of the machine learning programs are presented in Supplementary File S1.

**Table 1.** Predictor characteristics in analysis population A.

| Predictors | Overall n = 55,389 | No Delirium n = 50,928 | Delirium n = 4461 | *p* Value | % Missing |
|---|---|---|---|---|---|
| Predisposing risk factors | | | | | |
| Age, year | 73.0 (64.0–82.0) | 72.0 (64.0–81.0) | 83.0 (74.0–89.0) | <0.001 | 0.0 |
| Body mass index, kg/m$^2$ | 22.5 (20.0–25.1) | 22.7 (20.2–25.3) | 20.6 (18.2–23.2) | <0.001 | 1.1 |
| Men | 33,829 (61.1) | 31,474 (61.8) | 2355 (52.8) | <0.001 | 0.0 |
| Intake of benzodiazepine medications | 977 (1.8) | 812 (1.7) | 165 (3.8) | <0.001 | 4.4 |
| Intake of opioid medications | 896 (1.7) | 690 (1.4) | 206 (4.7) | <0.001 | 4.4 |
| Intake of steroid medications | 1829 (3.5) | 1573 (3.2) | 256 (5.9) | <0.001 | 4.4 |
| Dementia | 4907 (9.3) | 3450 (7.1) | 1457 (33.3) | <0.001 | 4.4 |
| Brain tissue disorder | 7028 (13.3) | 5950 (12.3) | 1078 (24.7) | <0.001 | 4.4 |
| Heavy drinker | 878 (1.7) | 701 (1.4) | 177 (4.0) | <0.001 | 4.4 |
| History of delirium | 1307 (2.5) | 802 (1.7) | 505 (11.5) | <0.001 | 4.4 |
| Emergency hospitalization | 30,043 (54.2) | 26,075 (51.2) | 3968 (88.9) | <0.001 | 0.0 |
| Use of ambulance | 19,189 (34.6) | 15,948 (31.3) | 3241 (72.7) | <0.001 | 0.0 |
| Room at hospitalization | | | | <0.001 | 0.0 |
| Bay of general ward | 16,139 (29.1) | 15,570 (30.6) | 569 (12.8) | | |
| Intensive care unit | 12,216 (22.1) | 9661 (19.0) | 2555 (57.3) | | |
| Private room of general ward | 27,034 (48.8) | 25,697 (50.5) | 1337 (30.0) | | |
| Schedule of treatment | | | | <0.001 | 0.0 |
| Catheterization | 7249 (13.1) | 6950 (13.6) | 299 (6.7) | | |
| Endoscopic treatment | 5032 (9.1) | 4792 (9.4) | 240 (5.4) | | |
| Preserved treatment | 23,554 (42.5) | 21,421 (42.1) | 2133 (47.8) | | |
| Surgery | 19,554 (35.3) | 17,765 (34.9) | 1789 (40.1) | | |

Data are expressed as median (interquartile range), number (%).

### *2.6. Handling of Missing Data*

For predictors with missing data, we used the sparsity-aware split-finding method in XGBoost, which treats the missing data as informative as they are.

### *2.7. Assessment of Prediction Models*

The discriminating ability of each model was assessed from the area under the receiver operating characteristic (AUROC) curve as well as that under the precision–recall curve (AUPRC), in addition to the sensitivity and specificity. The AUPRC was calculated because it is an insightful performance metric for imbalanced data such as those for infrequent outcome events [36]. The Youden index method was used to determine the optimal cut-off value that maximizes sensitivity and specificity. The calibration performance of each model was assessed using the calibration slope and intercept. The overall performance was assessed using the Brier score [37]. A stratified five-fold cross-validation was performed for internal validation, and the mean and standard deviation of each metric were calculated.

### *2.8. Confirmation of Predictive Contributors Using the Mchine Learning Interpretation Method*

We used Shapley additive explanations (SHAP), a model-agnostic method of machine learning, to visualize the relationship between predictors and delirium. SHAP is an excellent method for visualizing the contribution of predictors. Shapley values are guaranteed to be fairly distributed in cooperative game theory [38]. The SHAP algorithm is an additive feature attribution method that approximates each prediction $f(x)$ with $g(z')$, a linear function of the binary variables $z' \in \{0,1\}^M$ and the feature attribution values $\phi_i \in \mathbb{R}$, which are defined as follows:

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i, \quad (1)$$

where $M$ is the number of predictors.

The SHAP value $\phi_i$ is defined as follows:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(M - |S| - 1)!}{M!} [f_x(S \cup \{i\}) - f_x(S)], \quad (2)$$

where $f$ is the model, $N$ is the set of all input variables, and $S$ is a subset of set $N$ excluding variable $i$. The importance of each variable visualized in the SHAP summary plots is defined as follows:

$$FeatureImportance_i = \frac{1}{N} \sum_{j=1}^{N} \left| \phi_i^{(j)} \right|, \quad (3)$$

where $j$ is the subscript of the case number. The XGBoost algorithm was used to visualize Shapley values because of its high affinity to the SHAP method.

### 2.9. Statistical Methods

The differences in baseline characteristics and clinical data were compared using the $\chi^2$ test, Fisher's exact test, and the Mann–Whitney U test, as appropriate. Two-sided probability values less than 0.05 were considered statistically significant. All statistical analyses were conducted using the R statistical package (http://www.r-project.org/, version 4.2.0).

## 3. Results

### 3.1. Patient Characteristics

The median age (and IQR) of the 55,389 patients was 73.0 (64.0–82.0) years, and 33,829 (61.1%) of them were men. Delirium occurring 24 h or more after admission was reported among 4461 (8.1%) of the 55,389 patients in population A, and among 3968 (13.2%) of the 30,043 patients in population B. Postoperative delirium occurred in 206 (4.8%) of 4293 elective general anesthesia surgical patients. Table 1 lists the differences in predictors with and without delirium in population A; all the predictors differed significantly. Similarly, on examining the differences in population B, significant differences were observed in all predictors except laboratory data potassium and total bilirubin (Table S1). In population C, many predictors, except brain tissue disorder, heavy drinking, central venous port insertion, physical restraints, and the numerical rating scale, which is a pain rating scale, differed significantly (Table S2). The missing ratio of predictors was less than 10% for all the predictors except the Glasgow Coma Scale (Tables 1, S1 and S2).

### 3.2. Performance of Delirium Prediction Models Based on XGBoost

The prediction performance of the model based on XGBoost after cross-validation is presented in Table 2. For the discrimination performance of the prediction model, the AUROC was 0.852 for population A, 0.806 for population B, and 0.794 for population C. Similarly, the AUPRC calculated to assess the ability to adequately identify imbalanced events was 0.329 in population A, 0.365 in population B, and 0.177 in population C. Furthermore, the calibration slope was 1.013 for population A, 1.102 for population B, and 1.032 for population C. Discrimination and calibration performances were generally good.

**Table 2.** Predictive performance of prediction models.

| Analysis Population | Algorithm | Discrimination | | | | Calibration | | Overall |
|---|---|---|---|---|---|---|---|---|
| | | Sensitivity | Specificity | AUROC | AUPRC | Slope | Intercept | Brier Score |
| A, n = 55,389 | XGBoost | 0.838 (0.015) | 0.721 (0.022) | 0.852 (0.005) | 0.329 (0.015) | 1.013 (0.038) | 0.001 (0.049) | 0.062 (0.002) |
| B, n = 30,043 | | 0.783 (0.027) | 0.690 (0.015) | 0.806 (0.006) | 0.365 (0.014) | 1.102 (0.017) | 0.002 (0.017) | 0.098 (0.001) |
| C, n = 4293 | | 0.767 (0.084) | 0.709 (0.040) | 0.794 (0.033) | 0.177 (0.052) | 1.032 (0.347) | −0.047 (0.306) | 0.044 (0.004) |

The mean and standard deviation of each metric after stratified 5 fold cross-validation are shown.

### *3.3. Interpretation of Predictors*

The SHAP summary plots visualized age, body mass index (BMI), dementia, and history of delirium as key predictors in all analyzed populations (Figures 2–4). Other predictors specific to each population were visualized. We identified SHAP dependence plots to visualize the relationship between the BMI and albumin levels with delirium, which were not focused on in the pre-revision risk score used at the acute care hospital in this study but made a large contribution to prediction (Figures 5 and 6). Furthermore, to review the reference value of age in the pre-revised risk score, the relationship between age and delirium was similarly visualized (Figure 7). The results revealed a sharp increase in risk when BMI was near 20–25 and albumin levels were near 3.0–3.5. The risk of delirium increased as both values decreased. The risk of delirium increased rapidly in the age group of 70–75 years.

### *3.4. Review and Use of Analysis*

An expert panel including a psychiatrist reviewed the results of this analysis and consented to add the BMI and albumin levels as new risk factors. Furthermore, they increased the risk threshold from age 60 to 70 years (Figure 8). The revised risk score can lower the false positive and false negative rates by 4.66% and 0.02%, respectively. A novel clinical pathway was introduced on 3 August 2022. The pathway includes the revised risk score and delirium-specific care content.

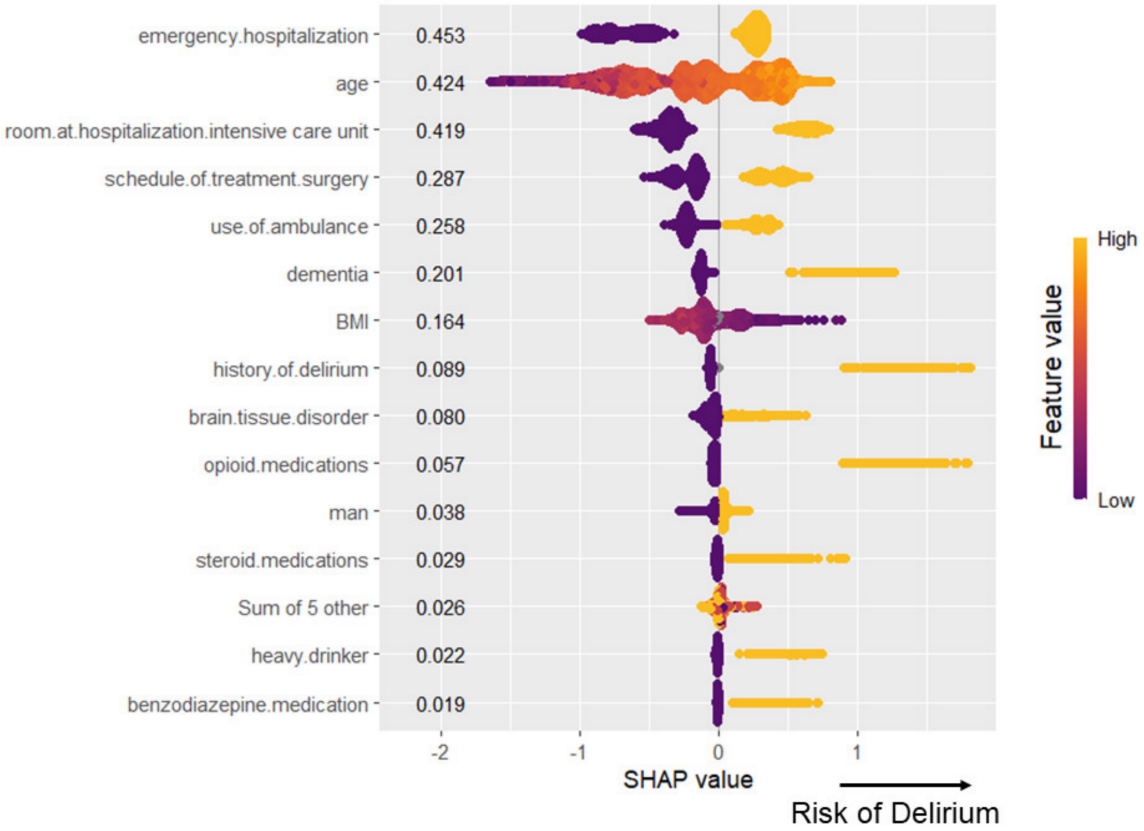



**Figure 2.** SHAP summary plot for Population A. SHAP summary plot of a 14-predictor XGBoost model. Each dot denotes a patient. Although there are 14 predictors, the one-hot encoding process results in 19 features when inputting into the XGBoost model. The top 15 predictive contributing features are shown. The x-axis shows the SHAP value attributed to the predictors. Higher SHAP values represent higher risk of delirium due to predictors.

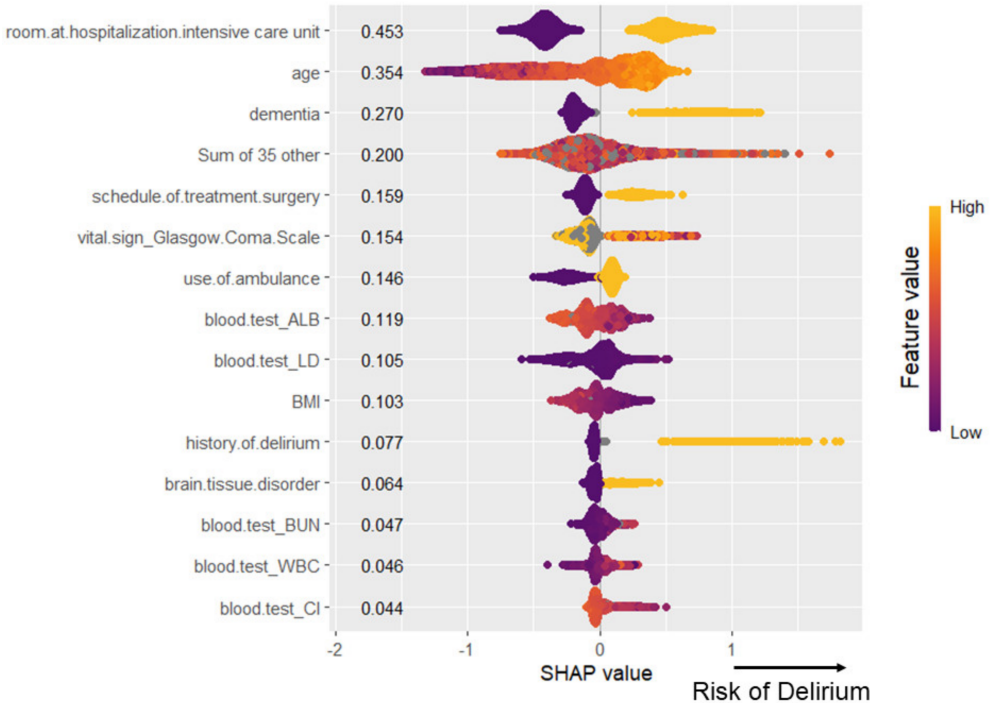



**Figure 3.** SHAP summary plot for Population B. SHAP summary plot of a 44-predictor XGBoost model. Each dot denotes a patient. Although there are 44 predictors, the one-hot encoding process results in 49 features when inputting into the XGBoost model. The top 15 predictive contributing features are shown. The x-axis shows the SHAP value attributed to the predictors. Higher SHAP values represent higher risk of delirium due to predictors.

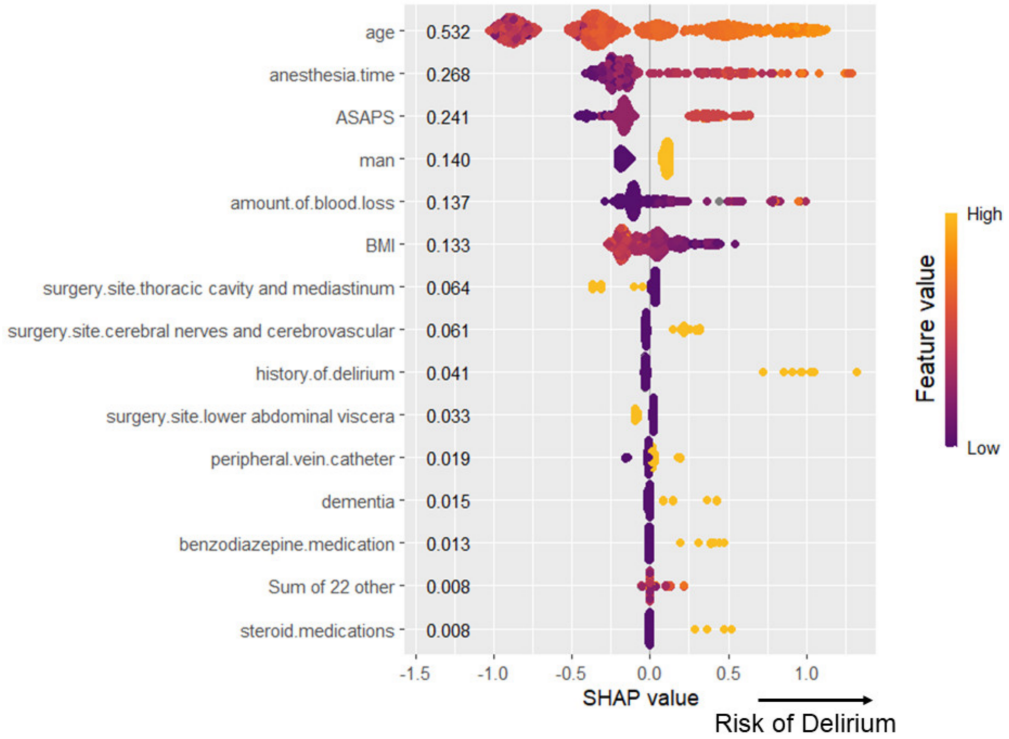



**Figure 4.** SHAP summary plot for Population C. SHAP summary plot of a 23-predictor XGBoost model. Each dot denotes a patient. Although there are 23 predictors, the one-hot encoding process results in 34 features when inputting into the XGBoost model. The top 15 predictive contributing features are shown. The x-axis shows the SHAP value attributed to the predictors. Higher SHAP values represent higher risk of delirium due to predictors.

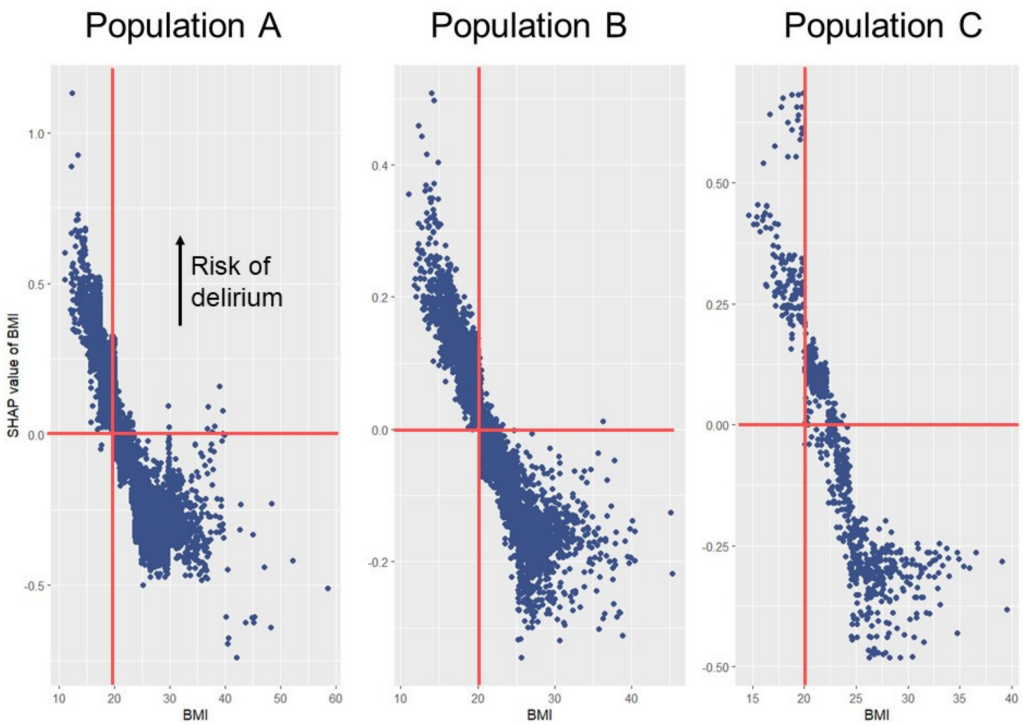



**Figure 5.** SHAP dependence plot for BMI. Each dot denotes a patient. The x-axis shows the BMI of the patients, and the y-axis shows the SHAP value attributed to the BMI. Higher SHAP values represent higher risk of delirium due to BMI.

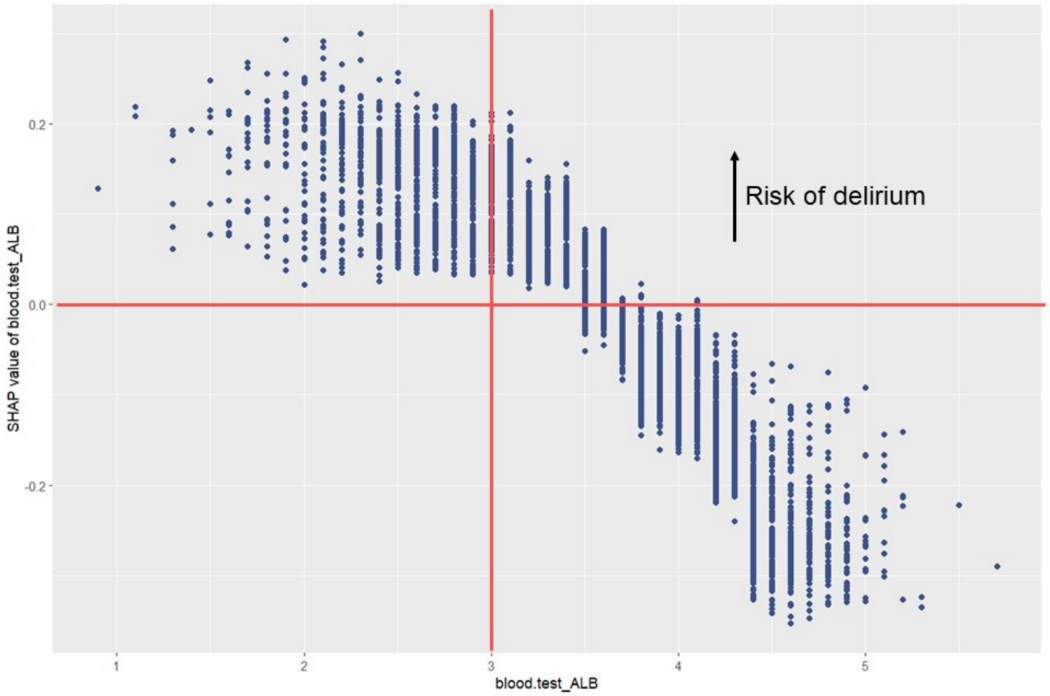



**Figure 6.** SHAP dependence plot for albumin levels in Population B. Each dot denotes a patient. The x-axis shows the albumin levels, and the y-axis shows the SHAP values attributed to the albumin levels. Higher SHAP values represent higher risk of delirium due to albumin levels.

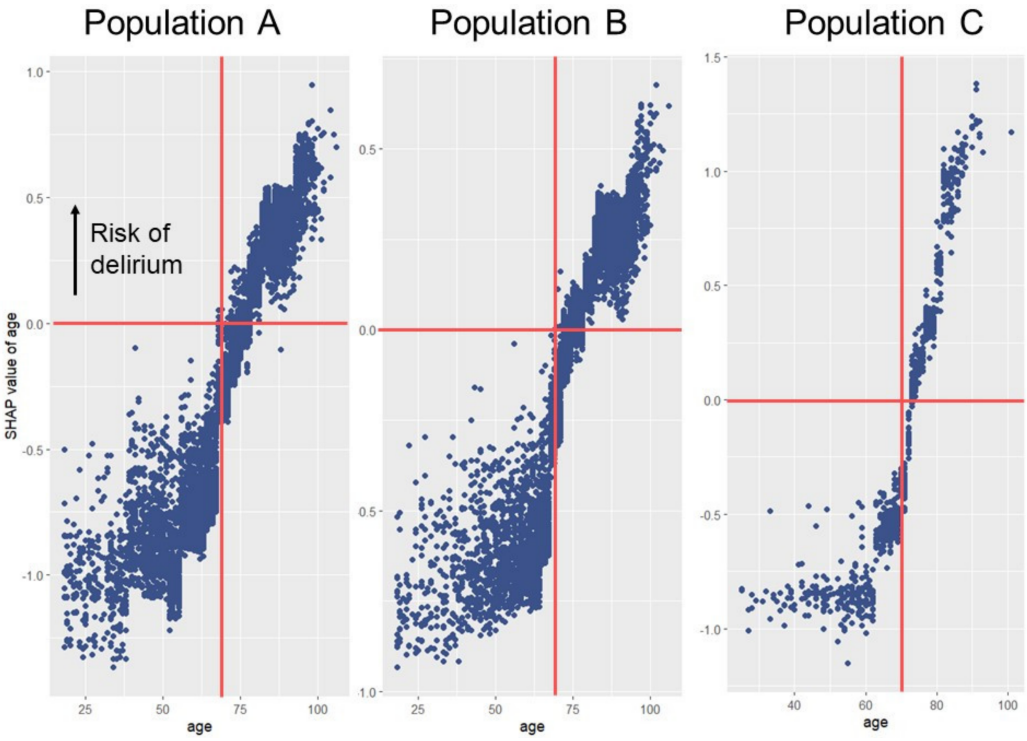



**Figure 7.** SHAP dependence plot for age. Each dot denotes a patient. The x-axis is the age of patients, and the y-axis shows the SHAP values attributed to the age. Higher SHAP values represent higher risk of delirium due to age.

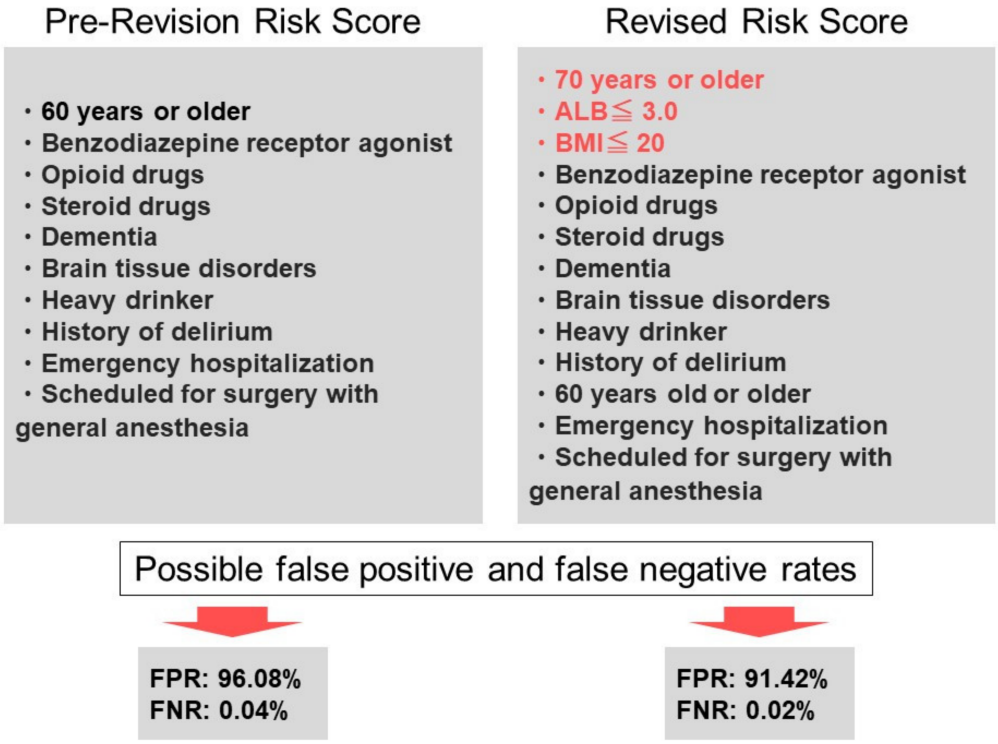



**Figure 8.** Risk score revision details.

## 4. Discussion

The major findings of this study are as follows. The discriminative performance and calibration of the XGBoost-based model were satisfactory. The visualization of SHAP highlighted the effect of the BMI and albumin levels on delirium and the cut-off values

of age (Figures 5–7). These factors have previously received limited attention, and their values were reflected in the revision of risk scores.

Although the populations and predictors used in previous studies on the prediction of delirium using machine learning differ from those used in the present study, they exhibited higher AUROCs of 0.78–0.95 [15–26]. However, most studies were conducted on Western subjects, and the results may differ for Asians. The current study yielded a certain value because of the numerous cases involving Asians. Because Japan has an aging population, many of its hospitalized patients are senior citizens. Therefore, risk factors for delirium may not necessarily match those of other countries. Comparing the risk factors presented in previous studies with those extracted in the current study, many factors such as age, functional status, cognitive impairment, and dementia are common [4–26,39–41]. However, reviewing the predictive contributing factors visualized in SHAP revealed that low BMI and albumin levels were risk factors for delirium (Figures 5 and 6). This result may reflect a low nutritional level associated with thinness, i.e., frailty. Thus, the development and utilization of individualized predictive models that reflect the unique domestic situation may exhibit certain benefits.

In a previous study, we focused on the high prediction accuracy of machine learning and the disease management capability of electronic clinical pathways, and confirmed a deterrent effect on aspiration pneumonia by revising care items such as oral care in the electronic pathways for hemorrhagic stroke patients, using data analysis [28]. The combination of machine learning and electronic clinical pathways could be used to realize a learning health system; however, the black box nature of machine learning hinders its application. The ML interpretation method can enable experts to interpret the effect of various predictors of delirium and allow health care providers to make data-driven decisions. Figure 9 shows a schematic of a series of clinical workflows that incorporate the results of analyses based on ML interpretation methods. The analyst submits analysis results obtained using the interpretation method in an expert meeting on delirium. Specifically, the expert meeting is held based on the results, and if the results are judged to have clinical value, they are used in actual clinical practice. Note that an expert meeting including a psychiatrist was conducted to validate the results of the analysis. SHAP and other ML interpretation methods are still in the nascent stage of development, and the interpretation of visualized results should be kept in mind. These results should only be used to detect anomalies in the input data or to understand the effect of predictors. Against this background, this study was limited to revising the risk score with reference to the results of the ML interpretability method.

However, revising the risk score alone cannot accurately reflect the complex interactions among predictors and cannot improve the prediction accuracy. The false-positive rate would remain high even with a revised risk score. Therefore, predictive models based on machine learning must be implemented in the future; for example, in patient information systems for electronic medical records. Furthermore, a system should be constructed to initiate measures against delirium based on the predictive values for each case output of the models. Studies focused on the prospective implementation of machine learning-based delirium prediction have exhibited consistency with independent expert ratings; however, calibration challenges have been reported [42]. In addition, the predictive performance of machine learning predictive models has been reported to deteriorate across hospitals [43], leaving challenges for the development and implementation of generic machine learning predictive models that can be used across different hospitals. To spread the implementation of medical diagnosis support systems using machine learning prediction models, it is necessary to seriously consider whether prediction models specific to each hospital should be developed by each hospital, or whether prediction models should be developed using integrated data from different hospitals after establishing a statistical mechanism to eliminate inter-institutional bias. Furthermore, to implement predictive models using machine learning, in addition to the challenges of the predictive models, operational problems such as the timing of data acquisition should be resolved. Therefore, further research is necessary.

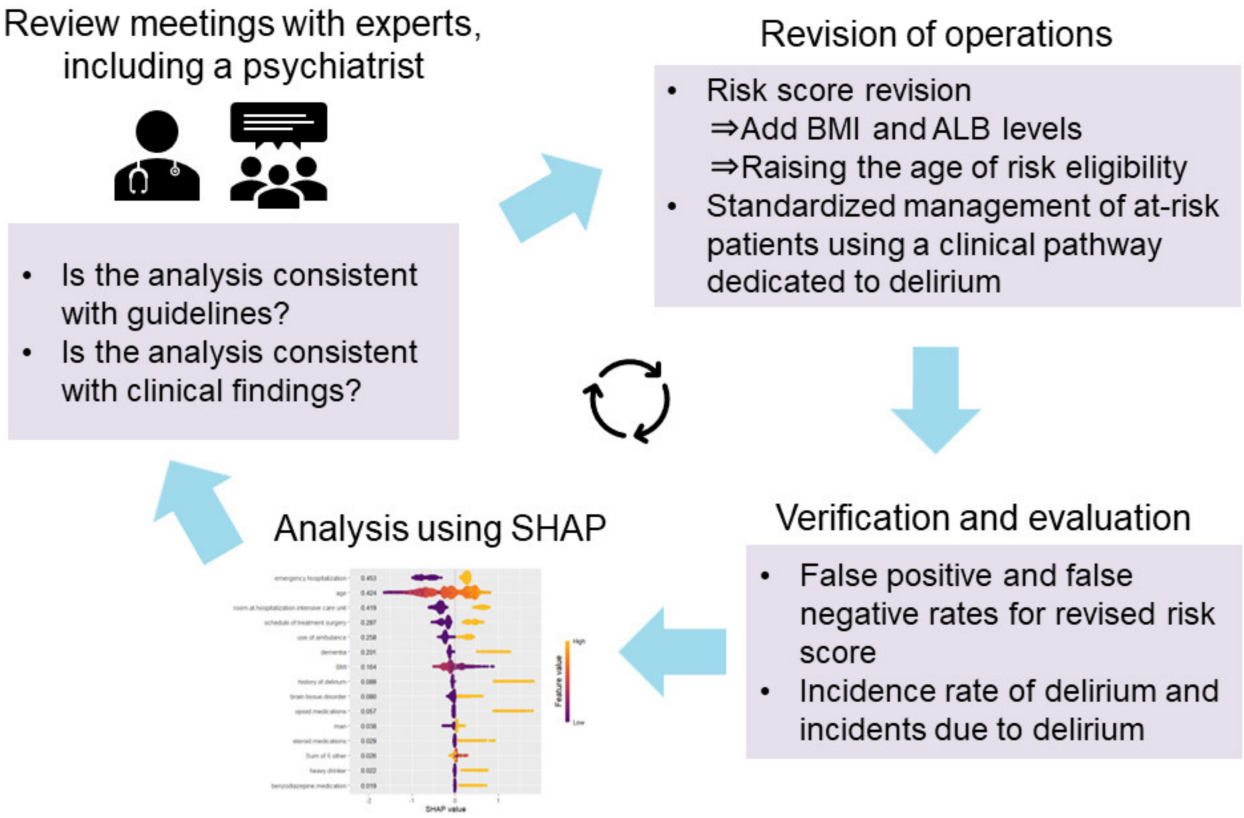



**Figure 9.** Model case for a learning health system focused on delirium.

This study had several limitations. First, patients from a single center were selected, which resulted in potential selection bias. Second, delirium incidence was low, and data including such unbalanced events may affect model development. Third, although the prediction models were internally validated, external validation was not performed, owing to limited time. Therefore, the validity of the prediction accuracy was not assured. Finally, because this study was conducted on Japanese subjects, the results of the analysis should be assessed in other social settings and ethnic groups.

## 5. Conclusions

Although research on machine learning predictive models using ML interpretation methods is on the rise, studies showing the sequence of events leading up to actual implementation in clinical practice are rare. In this study, machine learning models were developed for predicting delirium for numerous Japanese cases. The models displayed high degrees of discrimination and calibration performance. A risk score was revised using the ML interpretation method, and an example of incorporating the analysis results of electronic health record data into a clinical workflow was presented.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app13031564/s1, Table S1: Predictor characteristics in analysis population B; Table S2: Predictor characteristics in analysis population C; Supplementary File S1: R programs; Supplementary File S2: TRIPOD checklist.

**Author Contributions:** Conceptualization, N.N. and H.S.; methodology, Y.N. and K.M.; investigation, M.S., Y.T. and S.F.; data curation, K.M.; writing—original draft preparation, K.M.; writing—review and editing, N.N. and H.S.; supervision, N.N. and H.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by the Grants-in-Aid for Scientific Research (KAKENHI; Grant number: 22K17336, 22H03328) from the Japanese Ministry of Education, and the MHLW Program (Grant number: JP21AC1002) from the Japanese Ministry of Health, Labour and Welfare.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Saiseikai Kumamoto Hospital (Approval No. 1072 and 27 April 2022).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the sensitive nature of the data collected for this study.

**Acknowledgments:** We thank the staff in Saiseikai Kumamoto Hospital, who helped us in collecting the data, for their help.

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

## References

1. Olsen, L.A.; Aisner, D.; McGinnis, J.M. *The Learning Healthcare System: Workshop Summary*; National Academies Press: Washington, DC, USA, 2007.
2. McGinnis, J.M.; Fineberg, H.V.; Dzau, V.J. Advancing the Learning Health System. *N. Engl. J. Med.* **2021**, *385*, 1–5. [CrossRef] [PubMed]
3. Platt, R.; Harvard Pilgrim Health Care Institute; Huang, S.S.; Perlin, J.B. A win for the learning health system. In *NAM Perspect*; National Academy of Medicine: Washington, DC, USA, 2013; Volume 3. [CrossRef]
4. Inouye, S.K.; Westendorp, R.G.; Saczynski, J.S. Delirium in elderly people. *Lancet* **2014**, *383*, 911–922. [CrossRef] [PubMed]
5. Witlox, J.; Eurelings, L.S.M.; de Jonghe, J.F.M.; Kalisvaart, K.J.; Eikelenboom, P.; van Gool, W.A. Delirium in elderly patients and the risk of postdischarge mortality, institutionalization, and dementia. *JAMA* **2010**, *304*, 443–451. [CrossRef] [PubMed]
6. Inouye, S.K.; Viscoli, C.M.; Horwitz, R.I.; Hurst, L.D.; Tinetti, M.E. A predictive model for delirium in hospitalized elderly medical patients based on admission characteristics. *Ann. Intern. Med.* **1993**, *119*, 474–481. [CrossRef] [PubMed]
7. Pompei, P.; Foreman, M.; Rudberg, M.A.; Inouye, S.K.; Braund, V.; Cassel, C.K. Delirium in hospitalized older persons: Outcomes and predictors. *J. Am. Geriatr. Soc.* **1994**, *42*, 809–815. [CrossRef]
8. O'Keeffe, S.T.; Lavan, J.N. Predicting delirium in elderly patients: Development and validation of a risk-stratification model. *Age Ageing* **1996**, *25*, 317–321. [CrossRef]
9. Pendlebury, S.T.; Lovett, N.; Smith, S.C.; Cornish, E.; Mehta, Z.; Rothwell, P.M. Delirium risk stratification in consecutive unselected admissions to acute medicine: Validation of externally derived risk scores. *Age Ageing* **2016**, *45*, 60–65. [CrossRef]
10. Rudolph, J.L.; Doherty, K.; Kelly, B.; Driver, J.A.; Archambault, E. Validation of a delirium risk assessment using electronic medical record information. *J. Am. Med. Dir. Assoc.* **2015**, *17*, 244–248. [CrossRef]
11. de Wit, H.A.J.M.; Winkens, B.; Gonzalvo, C.M.; Hurkens, K.P.G.M.; Mulder, W.J.; Janknegt, R.; Verhey, F.R.; van der Kuy, P.-H.M.; Schols, J.M.G.A. The development of an automated ward independent delirium risk prediction model. *Pharm. Weekbl.* **2016**, *38*, 915–923. [CrossRef]
12. Solà-Miravete, E.; López, C.; Martínez-Segura, E.; Adell-Lleixà, M.; Juvé-Udina, M.E.; Lleixà-Fortuño, M. Nursing assessment as an effective tool for the identification of delirium risk in older in-patients: A case-control study. *J. Clin. Nurs.* **2017**, *27*, 345–354. [CrossRef]
13. Douglas, V.C.; Hessler, C.S.; Dhaliwal, G.; Betjemann, J.P.; Fukuda, K.A.; Alameddine, L.R.; Lucatorto, R.; Johnston, S.C.; Josephson, S.A. The AWOL tool: Derivation and validation of a delirium prediction rule. *J. Hosp. Med.* **2013**, *8*, 493–499. [CrossRef]
14. Brown, E.G.; Josephson, S.A.; Anderson, N.; Reid, M.; Lee, M.; Douglas, V.C. Predicting inpatient delirium: The AWOL delirium risk-stratification score in clinical practice. *Geriatr. Nurs.* **2017**, *38*, 567–572. [CrossRef]
15. Wong, A.; Young, A.T.; Liang, A.S.; Gonzales, R.; Douglas, V.C.; Hadley, D. Development and validation of an electronic health record-based machine learning model to estimate delirium risk in newly hospitalized patients without known cognitive impairment. *JAMA Netw. Open* **2018**, *1*, e181018. [CrossRef]
16. Corradi, J.P.; Thompson, S.; Mather, J.F.; Waszynski, C.M.; Dicks, R.S. Prediction of Incident Delirium Using a Random Forest classifier. *J. Med. Syst.* **2018**, *42*, 261. [CrossRef]
17. Davoudi, A.; Ozrazgat-Baslanti, T.; Ebadi, A.; Bursian, A.C.; Bihorac, A.; Rashidi, P. Delirium Prediction using Machine Learning Models on Predictive Electronic Health Records Data. In Proceedings of the 2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE), Washington, DC, USA, 23–25 October 2017; pp. 568–573. [CrossRef]
18. Kramer, D.; Veeranki, S.; Hayn, D.; Quehenberger, F.; Leodolter, W.; Jagsch, C.; Schreier, G. Development and validation of a multivariable prediction model for the occurrence of delirium in hospitalized gerontopsychiatry and internal medicine patients. *Stud. Health Technol. Inform.* **2017**, *236*, 32–39.

19. Mufti, H.N.; Hirsch, G.M.; Abidi, S.R.; Abidi, S.S.R. Exploiting machine learning algorithms and methods for the prediction of agitated delirium after cardiac surgery: Models development and validation study. *JMIR Public Health Surveill.* **2019**, *7*, e14993. [CrossRef]
20. Veeranki, S.P.K.; Hayn, D.; Jauk, S.; Quehenberger, F.; Kramer, D.; Leodolter, W.; Schreier, G. An improvised classification model for predicting delirium. *Stud. Health Technol. Inform.* **2019**, *264*, 1566–1567. [CrossRef]
21. Chua, S.J.; Wrigley, S.; Hair, C.; Sahathevan, R. Prediction of delirium using data mining: A systematic review. *J. Clin. Neurosci.* **2021**, *91*, 288–298. [CrossRef]
22. Oosterhoff, J.H.F.; Karhade, A.V.; Oberai, T.; Franco-Garcia, E.; Doornberg, J.N.; Schwab, J.H. Prediction of postoperative delirium in geriatric hip fracture patients: A clinical prediction model using machine learning algorithms. *Geriatr. Orthop. Surg. Rehabil.* **2021**, *12*, 21514593211062277. [CrossRef]
23. Hur, S.; Ko, R.-E.; Yoo, J.; Ha, J.; Cha, W.C.; Chung, C.R. A machine learning-based algorithm for the prediction of intensive care unit delirium (pride): Retrospective study. *JMIR Public Health Surveill.* **2021**, *9*, e23401. [CrossRef]
24. Bishara, A.; Chiu, C.; Whitlock, E.L.; Douglas, V.C.; Lee, S.; Butte, A.J.; Leung, J.M.; Donovan, A.L. Postoperative delirium prediction using machine learning models and preoperative electronic health record data. *BMC Anesthesiol.* **2022**, *22*, 8. [CrossRef] [PubMed]
25. Zhang, Y.; Wan, D.H.; Chen, M.; Li, Y.L.; Ying, H.; Yao, G.L.; Liu, Z.L.; Zhang, G.M. Automated machine learning-based model for the prediction of delirium in patients after surgery for degenerative spinal disease. *CNS Neurosci. Ther.* **2022**, *29*, 282–295. [CrossRef] [PubMed]
26. Liu, S.; Schlesinger, J.J.; McCoy, A.B.; Reese, T.J.; Steitz, B.; Russo, E.; Koh, B.; Wright, A. New onset delirium prediction using machine learning and long short-term memory (LSTM) in electronic health record. *J. Am. Med. Inform. Assoc.* **2022**, *30*, 120–131. [CrossRef] [PubMed]
27. Matsumoto, K.; Nohara, Y.; Soejima, H.; Yonehara, T.; Nakashima, N.; Kamouchi, M. Stroke prognostic scores and data-driven prediction of clinical outcomes after acute ischemic stroke. *Stroke* **2020**, *51*, 1477–1483. [CrossRef]
28. Matsumoto, K.; Nohara, Y.; Wakata, Y.; Yamashita, T.; Kozuma, Y.; Sugeta, R.; Yamakawa, M.; Yamauchi, F.; Miyashita, E.; Takezaki, T.; et al. Impact of a learning health system on acute care and medical complications after intracerebral hemorrhage. *Learn. Health Syst.* **2021**, *5*, e10223. [CrossRef]
29. Zhao, Q.-Y.; Wang, H.; Luo, J.-C.; Luo, M.-H.; Liu, L.-P.; Yu, S.-J.; Liu, K.; Zhang, Y.-J.; Sun, P.; Tu, G.-W.; et al. Development and validation of a machine-learning model for prediction of extubation failure in intensive care units. *Front. Med.* **2021**, *8*, 676343. [CrossRef]
30. Wang, K.; Tian, J.; Zheng, C.; Yang, H.; Ren, J.; Liu, Y.; Han, Q.; Zhang, Y. Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP. *Comput. Biol. Med.* **2021**, *137*, 104813. [CrossRef]
31. Tseng, P.-Y.; Chen, Y.-T.; Wang, C.-H.; Chiu, K.-M.; Peng, Y.-S.; Hsu, S.-P.; Chen, K.-L.; Yang, C.-Y.; Lee, O.K.-S. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit. Care* **2020**, *24*, 478. [CrossRef]
32. Hathaway, Q.A.; Roth, S.M.; Pinti, M.V.; Sprando, D.C.; Kunovac, A.; Durr, A.J.; Cook, C.C.; Fink, G.K.; Cheuvront, T.B.; Grossman, J.H.; et al. Machine-learning to stratify diabetic patients using novel cardiac biomarkers and integrative genomics. *Cardiovasc. Diabetol.* **2019**, *18*, 78. [CrossRef]
33. Localio, A.R.; Stack, C.B. TRIPOD: A New Reporting Baseline for Developing and Interpreting Prediction Models. *Ann. Intern. Med.* **2015**, *162*, 73–74. [CrossRef]
34. Inouye, S.K.; Van Dyck, C.H.; Alessi, C.A.; Balkin, S.; Siegal, A.P.; Horwitz, R.I. Clarifying confusion: The confusion assessment method. A new method for detection of delirium. *Ann. Intern. Med.* **1990**, *113*, 941–948. [CrossRef]
35. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
36. Ozenne, B.; Subtil, F.; Maucort-Boulch, D. The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *J. Clin. Epidemiol.* **2015**, *68*, 855–859. [CrossRef]
37. Steyerberg, E.W.; Vickers, A.J.; Cook, N.R.; Gerds, T.; Gonen, M.; Obuchowski, N.; Pencina, M.J.; Kattan, M.W. Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology* **2010**, *21*, 128–138. [CrossRef]
38. Lundberg, S.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 4765–4774.
39. Kalimisetty, S.; Askar, W.; Fay, B.; Khan, A. Models for Predicting Incident Delirium in Hospitalized Older Adults: A Systematic Review. *J. Patient-Cent. Res. Rev.* **2017**, *4*, 69–77. [CrossRef]
40. Chen, X.; Lao, Y.; Zhang, Y.; Qiao, L.; Zhuang, Y. Risk predictive models for delirium in the intensive care unit: A systematic review and meta-analysis. *Ann. Palliat. Med.* **2021**, *10*, 1467. [CrossRef]
41. Ruppert, M.M.B.; Lipori, J.B.; Patel, S.; Ingersent, E.; Cupka, J.B.; Ozrazgat-Baslanti, T.; Loftus, T.; Rashidi, P.; Bihorac, A.M. ICU Delirium-Prediction Models: A Systematic Review. *Crit. Care Explor.* **2020**, *2*, e0296. [CrossRef]

42. Jauk, S.; Kramer, D.; Großauer, B.; Rienmüller, S.; Avian, A.; Berghold, A.; Leodolter, W.; Schulz, S. Risk prediction of delirium in hospitalized patients using machine learning: An implementation and prospective evaluation study. *J. Am. Med. Inform. Assoc.* **2020**, *27*, 1383–1392. [CrossRef]
43. Sun, H.; Depraetere, K.; Meesseman, L.; Silva, P.C.; Szymanowsky, R.; Fliegenschmidt, J.; Hulde, N.; von Dossow, V.; Vanbiervliet, M.; De Baerdemaeker, J.; et al. Machine learning-based prediction models for different clinical risks in different hospitals: Evaluation of live performance. *J. Med. Internet Res.* **2022**, *24*, e34295. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.