dHealth 2020 – Biomedical Informatics for Health and Care
G. Schreier et al. (Eds.)

doi:10.3233/SHTI200071

Multivariable Risk Prediction of Dysphagia in Hospitalized Patients Using Machine Learning

Anna Maria LIENHART[a,1], Diether KRAMER[a], Stefanie JAUK[a], Markus GUGATSCHKA[b], Werner LEODOLTER[a] and Thomas SCHLEGL[c]
[a] *Styrian Hospitals Limited Liability Company (KAGes), Graz, Austria*
[b] *Division of Phoniatrics, Medical University of Graz, Austria*
[c] *University of Applied Sciences St. Pölten, Austria*

Abstract. Background: Dysphagia is a dysfunction of the swallowing act and is highly prevalent in acute post-stroke patients and patients with chronic neurological diseases. Dysphagia is associated with several potentially life threatening complications. Thus, an early identification and treatment could reduce morbidity and mortality rates. Objectives: The aim of the study was to develop a multivariable model predicting the individual risk of dysphagia in hospitalized patients. Methods: We trained different machine learning algorithms on the electronic health records of over 33,000 patients. Results: The tree-based Random Forest Classifier and Adaboost Classifier algorithms achieved an area under the receiver operating characteristic curve of 0.94. Conclusion: The developed models outperformed previously published models predicting dysphagia. In future, an implementation in the clinical workflow is needed to determine the clinical benefit.

Keywords. dysphagia, electronic health records, machine learning, predictive analytics

1. Introduction

1.1. Medical Background

Dysphagia is the difficulty or total incapability of swallowing and is prevalent in patients with neurologic diseases, diseases in the area of the throat, nose or ears and internal disorders [1,2]. Due to the association with various diseases, the occurrence of dysphagia is quite variable. 25% to 78% of all stroke patients and 27% to 30% of patients with head injury develop some degree of dysphagia [3–7]. In addition, other diseases of the central nervous system like Parkinson disease and Alzheimer disease or malignant diseases in the head neck region show a high incidence for dysphagia [2,7]. The severity of the disease, especially when chronic, presents another risk factor for the occurrence of dysphagia [2].

The mechanism of dysphagia is multifactorial [2]. Post-extubation dysphagia occurs in 59% of non-neurologic critically ill patients and duration of intubation correlates

[1] Corresponding Author: Anna Maria Lienhart, KAGes-Management, Informations- und Prozessmanagement, A-8010 Graz, Billrothgasse 18a, E-Mail: AnnaMaria.Lienhart2@klinikum-graz.at

positively with the severity of dysphagia [8,9]. In addition, chewing and swallowing is affected by a loss of muscle mass and strength reduction in elderly people, resulting in swallowing disorders in 10% to 20% of over 65 year olds [10]. Also, different drug components correlate with the occurrence of dysphagia such as antihypertensives, neuroleptics, antidepressants or antidementives [11–13].

Dysphagia is associated with several complications such as dehydration, malnutrition and aspiration leading to pneumonia [2,14,15]. The risk for pneumonia in the group of dysphagic patients is significantly higher than in non-dysphagic patients (29.7% vs. 3.7%) [8]. These conditions are leading to longer hospital stays, reduced level of independence and increased risk of mortality [14,15]. However, the large amount of risk factors and predispositions makes it complex to identify an individual's risk of dysphagia.

Early dysphagia screening and detection is associated with a reduced risk for aspiration pneumonia and disability [6,14,16]. Interventions to prevent individuals from aspiration pneumonia include personal assistance while eating (e.g. by a speech language pathologist), thickening of liquids, or artificial nutrition (nasogastral tube, parenteral nutrition). The ideal screening tool should therefore be a quick and non-invasive process with a focus on detecting the potential risk for dysphagia, as well as aspiration [5]. This would further clarify whether swallowing assessment is necessary or whether it is safe to feed the patient orally.

1.2. State-of-the-art of Dysphagia Diagnostic and Prediction

In clinical practice, various screening tools, scales and scores are used to detect and diagnose swallowing disorders as well as their possible consequences. Those methods presuppose the suspicion of a dysphagia and are based on additional clinical examinations (e.g. Dysphagia Outcome and Severity Scale [17], Timed Water Swallow test [18], Gugging Swallowing Screen [19], Standardized Swallowing Assessment [20]). A standardized screening and testing procedure for each hospitalized patient would lead to a large number of additional medical examinations and to an increased documentation effort.

Zhou et al. [21] developed the bedside scoring model SSG-OD on prospectively collected data from 395 consecutive post-cardiac surgery patients (including 103 dysphagic patients). Univariate and multivariate logistic analyses were performed in order to identify independent predictors for dysphagia. The final SSG-OD model identified patients at risk for dysphagia based on three predictors: gastric intubation, sedative drug use duration and occurrence of stroke. The authors reported a sensitivity of 68.5% and a specificity of 89.0% of the model.

Another risk score for dysphagia after cardiac surgery is the RODICs score [22]. Grimm et al. collected patient-specific characteristics, intraoperative variables and postoperative outcomes from 1,314 patients undergoing heart surgery including 115 patients with dysphagia. The 38-point RODICS score comprises seven patient-specific characteristics and perioperative factors like body mass index, chronic lung disease, postoperative ventilation >24 hours. The score achieved an area under the receiver operating characteristic curve (AUROC) of 0.75 in the test data set.

A score for predicting persistent dysphagia (PreDyScore) was developed by Gandolfo et al. [4] performing a multivariate logistic regression on 249 post-stroke patients (including 94 dysphagic patients). The PreDyScore, which represents a combination of body mass index and modified Rankin Scale [23], had a sensitivity of 67.0% and a specificity of 95.7% (AUROC 0.79) in the test data set.

1.3. Objective

The aim of this study was to develop a predictive model to identify patients with an increased risk for dysphagia at an early state of hospitalization. This could enable the appropriate diagnostic, preventive and therapeutic steps at an earlier stage. Another goal of the study was not to increase the documentation effort for the health professionals, and thus develop a model based on routinely documented electronic health records (EHR) only.

2. Materials and Methods

The development and implementation of the study received approval from the Ethics Committee of the Medical University of Graz (30-146 ex 17/18). We used the TRIPOD statement [24] as guideline for developing, validating and reporting the model.

2.1. Data

For this study, routine clinical data of the Hospital Information System (HIS) (openMEDOCS) of Steiermärkische Krankenanstaltengesellschaft m. b. H. (KAGes) was used. In Styria, the KAGes covers about 90% of all hospital beds, which leads to an access of about 2 million longitudinal patient histories. The HIS of KAGes is based on IS-H/i.s.h.med information system and implemented on SAP platforms.

The predicted outcome was defined as having an ICD-10 coded diagnosis for dysphagia (R13) or aspiration pneumonia (J69), which is often preceded by dysphagia [25], during the recent hospital stay. In addition, we included nursing diagnoses of swallowing disorder in the predicted outcome.

A retrospective data set of hospitalized patients was extracted from the HIS. This set of data included all in-patients with the defined outcome in the period from January 1st, 2011 to October 31st, 2019. Patients under the age of 18 years were excluded. The data extraction resulted in a dysphagia sample of 12,068 patients. As a control group, we randomly selected 21,716 patients without dysphagia. The final cohort included 33,784 patients.

The feature set (n = 886) consisted of routinely stored EHRs extracted for the defined cohort (summarized in Table 1). ICD-10 codes from chronic diseases like diabetes and hypertonia were used without exclusion, any other ICD-10 code was only included from the last three years before first dysphagia occurrence. The ICD-10 codes were additionally grouped into related chapters e.g. I10_I15. Laboratory data were only included from the last 30 days relative to first dysphagia diagnose. Likewise, medication data for chronical illness were included for the whole time period, and acute ones from the last years only. Missing values in the nursing assessment were imputed by last observation carried forward method.

Table 1. Feature set for the prediction of dysphagia from EHRs with examples for the seven feature groups.

Data type	Description	n
Demographic Data	age, gender	28
Diagnosis Codes	ICD-10 Codes, Groups of ICD-10 Codes	286
Procedures Codes	examinations and procedures: MRI, CT	103
Laboratory Data	thrombocytes, creatinine	190
Nursing Protocols	body mass index, movement disorders	92
Administrative Data, Indices	Charlson Comorbitity Index, number of hospital stays	25
Medication	medication associated with dysphagia	162

2.2. Methods

The data set was split into a training data set, with 80% of the cases (n = 27,027) and a test data set comprising 20% of the cases (n = 6,757).

Categorical data was encoded, which resulted in a binary representation for each category of the feature. Numerical features were scaled between 0 and 1. The data preparation led to 1,783 features.

We applied the following classification algorithms on the given learning task: Random Forest Classifier (RF), AdaBoost Classifier, Logistic Regression (LR), Support Vector Machine (SVM) / C-Support Vector Classification (SVC), and K−Nearest Neighbor Classifier (KNN). Furthermore, we applied a regularized linear model with stochastic gradient descent (SGD) learning. All models were trained by supervised learning.

Tree-based machine learning algorithms are scale-invariant [27], hence we trained the RF and AdaBoost on the unscaled and unencoded data. All other methods were trained on the scaled and encoded data only.

As some of the used algorithms like SVM / SVC and LR are sensible to unbalanced data, we incorporated the weights of the classes in order to get higher weights on the minority class and lower weights on the majority class (class_weight = {0:1, 1:2}).

We trained the models on the training data set with a stratified 10-fold cross-validation, preserving the relative class distributions. To determine the optimal setting of hyperparameters for each classification algorithm we performed an exhaustive grid search. Hence, a n-dimensional grid was explored for parameter search. We then selected the best performing setting for each method reviewing all performance parameters with a focus on the AUROC (Table 3). Finally, the selected models were evaluated on the held-out test data set. Sensitivity and specificity were computed using the closest topleft method.

Data preparation and main data preprocessing were performed in R. Data normalization and modelling were executed using the programming language Python and the scikit-learn library [26].

3. Results

3.1. Data Characteristics

We used descriptive statistics for a first evaluation of the data. Relevant factors and diseases associated with dysphagia are presented in Table 2. The dysphagia cohort tends to be older and to have a lower body mass index than the cohort without dysphagia.

Furthermore, the patients with dysphagia had similar comorbidities as described in literature.

Table 2. Descriptive statistics for the included cohorts, with a focus on relevant factors and diseases associated with dysphagia described in literature.

	Dysphagia (n = 12,068)		**No Dysphagia (n = 21,716)**	
Age, years [a]	74 (63-83)		67 (53-78)	
BMI [a]	24.0 (21.1-27.3)		26.1 (23.6-29.1)	
	n	**%**	**n**	**%**
Comorbidities [b]				
Cerebral infarction (I63)	1,935	16.0	999	4.6
Stroke, not specified as haemorrhage or infarction (I64)	378	3.1	182	0.8
Intracranial injury (S06)	420	3.5	317	1.5
Parkinson disease (G20)	1,057	8.8	416	1.9
Alzheimer disease (G30)	1,167	9.7	436	2.0
Malignant neoplasm of oesophagus (C15)	1,071	8.9	23	0.1
Other chronic obstructive pulmonary disease (J44)	1,563	13.0	2,121	9.8

Note: [a] median and interquartile range; [b] absolute frequencies and column percentages

3.2. Modelling and Performance

The model performances of the different machine learning models on the test data set are listed in Table 3. Figure 1 illustrates the receiver operating characteristic (ROC) curves of all methods including ROC confidence intervals computed with the DeLong method [28].

Comparing the models, the tree-based methods RF and AdaBoost achieved the highest AUROC of 0.94. The RF model had a sensitivity and specificity of 0.88. The AdaBoost model performed quite similar, with a sensitivity of 0.88 and a specificity of 0.89.

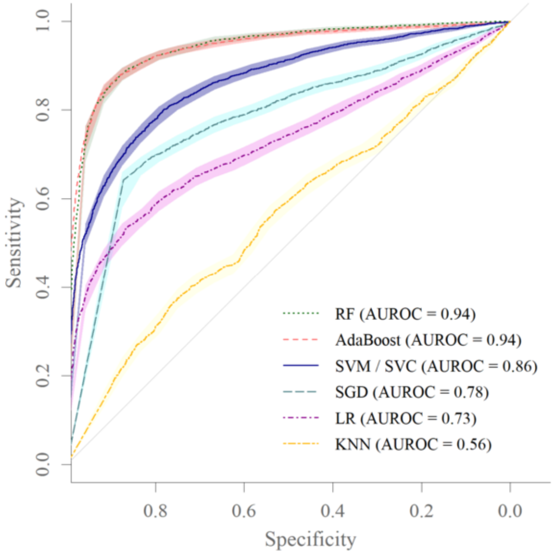

Figure 1. Receiver operating characteristic (ROC) curves of the model performances on the x-set with the corresponding area under the ROC curve (AUROC) values.

Table 3. Modelling results for different machine learning methods predicting the risk for dysphagia in hospitalized patients.

Model (Pythonmodule)	Parameter	AUROC	Acc.	Sens.	Spec.	Prec.
RF (ensemble)	n_estimator = 500, max_features = √886	**0.94**	0.88	0.88	0.88	0.80
AdaBoost (ensemble)	base_estimator = DecisionTreeClassifier, max_depth = 4, n_estimator = 400, learning_rate = 0.03	**0.94**	0.88	0.88	0.89	0.81
SVM / SVC (svm)	probability = True	0.86	0.79	0.80	0.78	0.68
SGD (linear_model)	loss = ’log’, alpha = 0.00001, max_iter = 1000, average = True	0.78	0.77	0.70	0.81	0.67
LR (linear_model)	max_iter = 1000, multi_class = ‘ovr’, solver = ‘sag’	0.73	0.71	0.62	0.76	0.59
KNN (neighbors)	algorithm = ’ball_tree’, leaf_size = 10, n_neighbors = 8, weights = ’distance’	0.56	0.55	0.55	0.55	0.41

Note: AUROC: Area under the receiving operating curve; Acc.: Accuracy; Sens.: Sensitivity; Spec.: Specificity; Prec.: Precision; RF and AdaBoost were trained and tested on the data set with 886 features. SVM / SVC, SGD, LR and KNN were trained and tested on the scaled and encoded data set with 1,783 features.

4. Discussion

The present study describes the development of a multivariable prediction model for the occurrence of dysphagia in hospitalized patients. We used the EHRs of more than 33,000 patients for the training process. Compared to previously published prediction models for dysphagia our cohort included a larger patient population and feature set with more than 800 features.

Our best performing models were the tree-based RF and Adaboost, achieving an AUROC of 0.94, which outperforms already published models [4, 23, 24]. However, these results are only to some extent comparable to our model. Previous prediction models did not use machine learning methods and were thus developed with less variables. Furthermore, our prediction model is not limited on a single cohort of patients such as those with stroke [4] and is thus more generally applicable in hospitalized patients. Some of the already established risk prediction models are based on additional clinical examinations which come along with extra effort for clinicians and nurses to obtain this data. Our model uses already documented EHR from the HIS openMEDOCS for dysphagia risk prediction only.

4.1. Limitations

In general, all limitations for the retrospective use of EHRs apply to this study as well. In case of missing data in the EHR, the model will have difficulties to predict a dysphagia. Ongoing evaluation and when indicated an adaption of the model to the missing values might be necessary. If dysphagia or aspiration pneumonia is neither ICD-10 coded nor mentioned in the discharge letter or nursing diagnoses, patients might be included in the control group which might influence the performance of the model.

4.2. Future work

Further machine learning algorithms, especially deep learning methods, should be applied on our data set. It should be determined, whether they can outperform the models in this paper. Additionally, further feature selection methods should be examined to reduce the number and dimensions in the preprocessed data set, which might improve the performance for algorithms such as SVM / SVC, LR and KNN.

While there are little barriers in training prediction models based on machine learning, the application into the clinical workflow still presents a challenge in predictive health care analytics. We were among the first to implement machine learning models for delirium [29] und ICU prediction [30] that are still in use by health care personal in their workflows. For dysphagia prediction, there is also a need to evaluate the model in a clinical setting in order to determine the clinical benefit of the prediction. Depending on the results during real-time prediction and feedback from health care professionals, we will consider implementing the machine learning based prediction in the hospital information system in KAGes.

4.3. Conclusion

We developed a multivariable risk classification model to predict the risk for dysphagia in hospitalized patients. The model was trained with EHRs using machine learning. The results on the test data set outperformed previous reported models in literature with an AUROC of 0.94. Further evaluation during real-time prediction is needed in order to determine the benefit in clinical practice.

References

[1] K.W. Altman, G.-P. Yu, and S.D. Schäfer, Consequence of Dysphagia in Hospitalized Patient, *Arch Otolaryngol Head Neck Surgery*. **8** (2010) 784–789.

[2] M. Prosiegel, and S. Weber, Dysphagie: Diagnostik und Therapie. Ein Wegweiser für kompetentes Handeln, 3. Auflage, Springer, Berlin, 2018.

[3] R. Martino, N. Foley, S. Bhogal, N. Diamant, M. Speechley, and R. Teasell, Dysphagia After Stroke: Incidence, Diagnosis, and Pulmonary Complications, *Stroke*. **36** (2005) 2756–2763. doi:10.1161/01.STR.0000190056.76543.eb.

[4] C. Gandolfo, S. Sukkar, M.G. Ceravolo, F. Cortinovis, C. Finocchi, R. Gradaschi, P. Orlandoni, N. Reale, S. Ricci, D. Vassallo, and A. Zini, The predictive dysphagia score (PreDyScore) in the short- and medium-term post-stroke: a putative tool in PEG indication, *Neurol Sci*. **40** (2019) 1619–1626. doi:10.1007/s10072-019-03896-2.

[5] G. Mann, G.J. Hankey, and D. Cameron, Swallowing Disorders following Acute Stroke: Prevalence and Diagnostic Accuracy, *Cerebrovasc Dis*. **10** (2000) 380–386. doi:10.1159/000016094.

[6] M. Al-Khaled, C. Matthis, A. Binder, J. Mudter, J. Schattschneider, U. Pulkowski, T. Strohmaier, T. Niehoff, R. Zybur, J. Eggers, J.M. Valdueza, G. Royl, and for QugSS II Group, Dysphagia in Patients with Acute Ischemic Stroke: Early Dysphagia Screening May Reduce Stroke-Related Pneumonia and Improve Stroke Outcomes, *Cerebrovasc Dis*. **42** (2016) 81–89. doi:10.1159/000445299.

[7] C. Takizawa, E. Gemmell, J. Kenworthy, and R. Speyer, A Systematic Review of the Prevalence of Oropharyngeal Dysphagia in Stroke, Parkinson's Disease, Alzheimer's Disease, Head Injury, and Pneumonia, *Dysphagia*. **31** (2016) 434–441. doi:10.1007/s00455-016-9695-9.

[8] M.J. Kim, Y.H. Park, Y.S. Park, and Y.H. Song, Associations Between Prolonged Intubation and Developing Post-extubation Dysphagia and Aspiration Pneumonia in Non-neurologic Critically Ill Patients, *Ann Rehabil Med*. **39** (2015) 763. doi:10.5535/arm.2015.39.5.763.

[9] S.A. Skoretz, T.M. Yau, J. Ivanov, J.T. Granton, and R. Martino, Dysphagia and Associated Risk Factors Following Extubation in Cardiovascular Surgical Patients, *Dysphagia*. **29** (2014) 647–654.

[10] J.A.Y. Cichero, Age-Related Changes to Eating and Swallowing Impact Frailty: Aspiration, Choking Risk, Modified Food Texture and Autonomy of Choice, *Geriatrics (Basel).* **3** (2018). doi:10.3390/geriatrics3040069.
[11] B. Marliani, Medikamentenassoziierte Schluckstörungen, (2014) 4.
[12] C. Schwemmle, M. Jungheim, S. Miller, D. Kühn, and M. Ptok, Medikamenteninduzierte Dysphagien: Ein Überblick, *HNO.* **63** (2015) 504–510. doi:10.1007/s00106-015-0015-8.
[13] D. Jaspersen, Drug-Induced Oesophageal Disorders: Pathogenesis, Incidence, Prevention and Management, *Drug Safety.* **22** (2000) 237–249. doi:10.2165/00002018-200022030-00007.
[14] N.J. Donovan, S.K. Daniels, J. Edmiaston, J. Weinhardt, D. Summers, and P.H. Mitchell, Dysphagia Screening: State of the Art: Invitational Conference Proceeding From the State-of-the-Art Nursing Symposium, International Stroke Conference 2012, *Stroke.* **44** (2013). doi:10.1161/STR.0b013e3182877f57.
[15] V. Guyomard, R.A. Fulcher, O. Redmayne, A.K. Metcalf, J.F. Potter, and P.K. Myint, Effect of Dysphasia and Dysphagia on Inpatient Mortality and Hospital Length of Stay: A Database Study: EFFECT OF DYSPHASIA AND DYSPHAGIA ON STROKE OUTCOMES, *Journal of the American Geriatrics Society.* **57** (2009) 2101–2106. doi:10.1111/j.1532-5415.2009.02526.x.
[16] C. Palli, S. Fandler, K. Doppelhofer, K. Niederkorn, C. Enzinger, C. Vetta, E. Trampusch, R. Schmidt, F. Fazekas, and T. Gattringer, Early Dysphagia Screening by Trained Nurses Reduces Pneumonia Rate in Stroke Patients: A Clinical Intervention Study, *Stroke.* **48** (2017) 2583–2585. doi:10.1161/STROKEAHA.117.018157.
[17] K.H. O'Neil, M. Purdy, J. Falk, and L. Gallo, The Dysphagia Outcome and Severity Scale, *Dysphagia.* **14** (1999) 139–145. doi:10.1007/PL00009595.
[18] K.M. Nathadwarawala, J. Nicklin, and C.M. Wiles, A timed test of swallowing capacity for neurological patients., *J Neurol Neurosurg Psychiatry.* **55** (1992) 822–825.
[19] M. Trapl, P. Enderle, M. Nowotny, Y. Teuschl, K. Matz, A. Dachenhausen, and M. Brainin, Dysphagia Bedside Screening for Acute-Stroke Patients: The Gugging Swallowing Screen, *Stroke.* **38** (2007) 2948–2952. doi:10.1161/STROKEAHA.107.483933.
[20] L. Perry, Screening swallowing function of patients with acute stroke. Part one: identi®cation, implementation and initial evaluation of a screening tool for use by nurses, *Journal of Clinical Nursing.* (2001) 11.
[21] X.-D. Zhou, W.-H. Dong, C.-H. Zhao, X.-F. Feng, W.-W. Wen, W.-Y. Tu, M.-X. Cai, T.-C. Xu, and Q.-L. Xie, Risk scores for predicting dysphagia in critically ill patients after cardiac surgery, *BMC Anesthesiol.* **19** (2019) 7. doi:10.1186/s12871-019-0680-3.
[22] J.C. Grimm, J.T. Magruder, R. Ohkuma, S.P. Dungan, A. Hayes, A.K. Vose, M. Orlando, M.S. Sussman, D.E. Cameron, and G.J.R. Whitman, A Novel Risk Score to Predict Dysphagia After Cardiac Surgery Procedures, *The Annals of Thoracic Surgery.* **100** (2015) 568–574.
[23] J.C. van Swieten, P.J. Koudstaal, M.C. Visser, H.J. Schouten, and J. van Gijn, Interobserver agreement for the assessment of handicap in stroke patients., *Stroke.* **19** (1988) 604–607.
[24] G.S. Collins, J.B. Reitsma, D.G. Altman, and K. Moons, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement, *BMC Medicine.* **13** (2015) 1. doi:10.1186/s12916-014-0241-z.
[25] S.E. Langmore, M.S. Terpenning, A. Schork, Y. Chen, J.T. Murray, D. Lopatin, and W.J. Loesche, Predictors of Aspiration Pneumonia: How Important Is Dysphagia?, *Dysphagia.* **13** (1998) 69–81.
[26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research.* **12** (2011) 2825−2830.
[27] S. Raschka, and V. Mirjalili, Machine Learning mit Python und Scikit-Learn und TensorFlow: das umfassende Praxis-Handbuch für Data Science, Deep Learning und Predictive Analytics, 2., aktualisierte und erweiterte Auflage, mitp, Frechen, 2018.
[28] E.R. DeLong, D.M. DeLong, and D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, *Biometrics.* **44** (1988) 837–845.
[29] D. Kramer, S. Veeranki, D. Hain, F. Quehenberger, W. Leodolter, C. Jagsch, and G. Schreier, Development and Validation of a Multivariable Prediction Model for the Occurrence of Delirium in Hospitalized Gerontopsychiatry and Internal Medicine Patients, *Studies in Health Technology and Informatics.* (2017) 32–39. doi:10.3233/978-1-61499-759-7-32.
[30] S. Jauk, D. Kramer, G. Stark, K. Hasiba, W. Leodolter, S. Schulz, and J. Kainz, Development of a Machine Learning Model Predicting an ICU Admission for Patients with Elective Surgery and Its Prospective Validation in Clinical Practice, *Studies in Health Technology and Informatics.* (2019) 173–177. doi:10.3233/SHTI190206.