ARC Journal of AI in Clinical Practice

Volume 01, Issue 01
www.arc-ai-clinical.com

Interpretable Machine Learning for Predicting Discharge Outcomes in Rehabilitation Settings

Mahsa Pahlevani[1], Majid Taghavi [1,2], Peter Vanberkel[1*]

[1]Department of Industrial Engineering, Dalhousie University, 5269 Morris Street, Halifax, B3H 4R2 NS Canada.

[2]Sobey School of Business, Saint Mary's University, 923 Robie, Halifax, B3H 3C3 NS Canada.

***Corresponding Author's:** Peter Vanberkel, E-mail: peter.vanberkel@dal.ca

Received:23/02/2026; Revised: 27/03/2026; Accepted: 04/04/2026; Published: 11/04/2026

ABSTRACT

Rehabilitation discharge planning is a complicated clinical procedure that is determined by various patient and care-related variables. Interpretable machine learning has the potential to aid such decisions by making correct and clear predictions with the help of routinely gathered clinical data. This study aimed to develop and compare interpretable machine learning models for predicting discharge outcomes, with a focus on identifying patients at risk of non-home discharge. A retrospective observational analysis was conducted using an open clinical dataset of neurological and orthopedic rehabilitation patients. After removing duplicate records, 8,468 unique cases were included. The original multi-class discharge variable was converted into a binary outcome of home versus non-home discharge. Data preprocessing involved the removal of fully missing variables, imputation of remaining missing values, encoding of categorical variables, and exclusion of outcome-related features to prevent data leakage. Logistic Regression and Decision Tree models were trained using an 80:20 stratified train-test split with class weighting. The Decision Tree model demonstrated superior performance, achieving an accuracy of 0.906 and an ROC-AUC of 0.962, compared to 0.760 and 0.816 for Logistic Regression. Notably, recall for non-home discharge improved from 0.686 to 0.932. Feature importance analysis indicated that prediction was driven by a small number of variables, with minimal contribution from demographic factors. Interpretable machine learning shows strong potential for predicting discharge outcomes in rehabilitation. The Decision Tree model provided both high predictive performance and a transparent decision-making structure, supporting its potential role in clinical decision support.

Keywords: rehabilitation, discharge outcome prediction, interpretable machine learning, clinical decision support, decision tree

1. Introduction

In recent years, the artificial intelligence (AI) and machine learning (ML) integration in healthcare has been growing at an impressive pace due to the growing access to digital health data and the development of new computational approaches. They are also becoming widely implemented to aid clinical decision-making, streamline work processes, and enhance patient outcomes in a variety of healthcare environments (Saxena and Chandra, 2021). Specifically, predictive analytics has gained a new avenue in clinical practice due to the capacity of machine learning models to detect patterns in large and complex data sets. There is also an increasing interest in using this data to help make more informed and consistent clinical decisions as healthcare systems continue to produce large volumes of routine data. Recent publications emphasize that AI and ML can not only be used in the process of automation, but they can also be used in the process of clinical reasoning and research augmentation (Rubinger et al., 2023). Nevertheless, the practice of these technologies is not evenly adopted, in part because of the issues with interpretability, reliability, and clinical applicability. Although predictive models can be very efficient, to make them a part of everyday care, they have to offer healthcare professionals insights that are easily readable and can be acted upon. Although these improvements have been made, AI implementation in the healthcare field must be cautiously designed to consider ethical, operational, and clinical limitations to make it a safe and efficient tool (Bhardwaj, 2022).

The role of data-driven decision-making in healthcare has become an increasingly popular topic, particularly when it comes to the management of patients, as well as care planning. Systematic reviews show that predictive modeling can allow making more timely and correct decisions and eventually lead to better outcomes and the use of resources (Lyu, 2025). Likewise, it has been recently shown that data-driven solutions can be used to improve patient care by detecting risk patterns and shaping clinical pathways (Adeniran et al., 2024). These changes are especially applicable to rehabilitation contexts, whereby patient progression might differ greatly, and choices have to consider various clinical and functional variables. Discharge planning is a very important aspect of rehabilitation, and it involves the coordination of healthcare professionals, patients, and caregivers. Smooth discharge decisions are critical in providing continuity of care and minimizing the chances of negative outcomes. It has been demonstrated that collaborative discharge planning can enhance the quality of decisions, which can be made by considering a variety of views, such as those of patients and families (Gledhill et al., 2023). Simultaneously, the discharge decisions are affected by a wide variety of clinical, organizational, and social factors, which makes them always complex and even inconsistent (Ward-Stockham et al., 2024).

There is an indication that the process of discharge can be highly unpredictable and might fail to meet uniform standards, thus impacting the efficiency and patient outcomes (Luther et al., 2019). The negative outcomes of inadequate discharge planning, including unplanned readmissions, have led to the need to increase the systematic and dependable decision-support mechanisms (Considine et al., 2020). These issues highlight the possible importance of predictive tools that can help clinicians to determine patients who might need further support or other discharge routes. Machine learning has been suggested as one of the promising solutions in helping with clinical decision-making in such situations. The analysis of historical data allows the ML models to define patterns related to certain outcomes and make probabilistic predictions to make care decisions (Adlung et al., 2021). The findings of previous studies in the related areas have shown that machine learning methods could identify the trends in regular behavior and deviation, implying their potential to be used in patient trajectories and outcomes modeling (Chifu et al., 2018). The more recent developments in data mining and clustering methods also indicate that these strategies can be used to identify the meaningful trends in complex data (Chifu et al., 2022).

Although such advances have been made, there are a number of gaps in the existing literature. Most of the literature is aimed at maximizing the predictive performance without sufficient consideration of interpretability, which restricts their application in clinical practice. Also, few studies have been done to use interpretable models on real-world rehabilitation data, especially in predicting discharge outcomes. The necessity of both accurate and transparent models is particularly significant in clinical practice, where the decisions should be supported and comprehended by healthcare providers.

To fill those gaps, the current study will come up with and test interpretable machine learning models to predict discharge outcomes in rehabilitation settings. The study uses a clinical dataset in the real-world setting to compare a baseline statistical model to a rule-based machine learning method in terms of their predictive performance and interpretability. This is not only to determine the accuracy of the model, but also to determine the important variables related to the discharge outcomes, thus coming up with information that can be used in making clinical decisions.

2. Methodology

2.1 Research Design

The research design was a retrospective observational design to develop and test predictive models on discharge outcomes in rehabilitation contexts. The main goal was to evaluate the ability of the routinely available clinical variables on admission to predict discharge disposition in simple and interpretable machine learning methods. The main focus was on the methodological transparency, reproducibility, and clinical applicability instead of the complexity of the models. The two models that were chosen to offer a compromise between the statistical baseline performance and interpretability are Logistic Regression and Decision Tree.

2.2 Data Source

The data that was used in this research were sourced from an open-access repository (Buscarini et al., 2024). It consists of anonymized clinical and demographic data of patients admitted to neurological and orthopedic rehabilitation units. The data set consists of variables in the form of patient characteristics, clinical condition and administrative coding. It was strictly selected to assist predictive modeling of discharge results, especially home discharge and alternative discharge pathways.

2.3 Study Population and Outcome Definition

The first dataset comprised 10,520 records of patients. Redundant records were found and eliminated, leaving an ultimate analytical cohort of 8,468 distinct patients. The main outcome measure was discharge disposition, which was originally measured as a multi-class variable. To conduct this study, it was converted to a binary variable: home discharge/non-home discharge. This change can be seen as a clinically relevant decision-making situation in which it is of paramount importance to differentiate between patients who need further intervention.

2.4 Data Preprocessing

Preprocessing of data was done in an organized and reproducible fashion. Variables that were totally missing were not analyzed. Missing values that were not eliminated had been handled through median imputation of numerical variables and mode imputation of categorical variables.

One-hot encoding was used to encode categorical variables with low cardinality, whereas the high-cardinality variables were coded as numerical values in order to prevent dimensionality inflation. Also, to avoid data leakage, variables that directly encode the outcome were eliminated. Duplicate records were eliminated before analysis, and all preprocessing was done in the same way throughout the dataset.

2.5 Feature Selection and Representation

There was no clearly specified feature selection algorithm before modeling. Rather, all the available variables following the process of preprocessing were maintained to maintain the integrity of the routinely collected clinical data. The importance of the features was then determined by the Decision Tree model to determine the most significant predictors. This method provides the opportunity to interpret post hoc, without the pre-selection bias.

2.6 Model Development

Two comparative models were designed, and these include Logistic Regression and Decision Tree. The chosen baseline model was Logistic Regression because it is interpretable, and it is widely used in clinical research; the Decision Tree model was chosen because of its capability to model non-linear relationships as well as the provision of rule-based and easily interpretable decision structures. Class weights were also added in the model training to deal with the imbalance in classes in the outcome variable. The model hyperparameters were selected sparingly in order to minimize the chance of overfitting, and there were certain restrictions on tree depth and minimum sample sizes in the Decision Tree model.

2.7 Model Evaluation

To maintain the distribution of the outcome classes, an 80:20 stratified split was used to separate the dataset into training and testing sets. Various measures were used to assess model performance on the test set, such as accuracy, area under the receiver operating characteristic curve (ROC-AUC), recall (sensitivity) of both classes, precision, and F1-score. Recall was specifically targeted in the non-home discharge group because there are more patients who need extra attention and hence have more clinical significance.

3. Results

3.1 Cohort Characteristics

The final analysis used 8,468 unique patients following the elimination of 2,052 duplicates. The sample was mostly older, and the mean age of the sample was 71.27 years (SD 12.33) with a median of 73.98 years, as the population of rehabilitation centers is characterized by the elderly. In this study, home discharge was seen in 6,700 patients (79.1%), and 1,768 patients (20.9%) needed other discharge plans. The gender variable (Gender) was distributed quite equally, with 3,734 and 4,734 patients in codes 1 and 2, respectively. The description of the characteristics of the study population at the baseline is summarized in Table 1.

Table 1. *Cohort characteristics and discharge outcomes*

Characteristic	Value
Number of patients	8,468
Number of variables after cleaning	82
Duplicate records removed	2,052
Mean age	71.27
Standard deviation of age	12.33
Median age	73.98
Home discharge	6,700
Non-home discharge	1,768
Gender code 1	3,734
Gender code 2	4,734

The discharge outcomes are depicted in Figure 1 that shows the imbalance in the classes towards home discharge. This imbalance was directly corrected with model training via class-weighted learning.

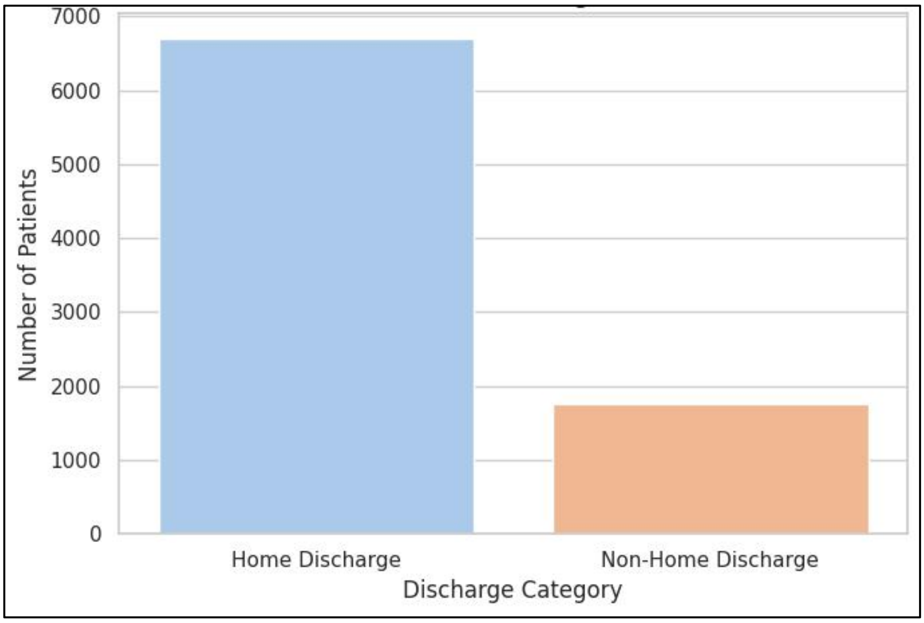

Figure 1. *Distribution of Discharge Outcomes in the Study Cohort*

3.2 Model Performance

Table 2 shows the predictive performance of the models. The Decision Tree model was shown to be more effective in all evaluation measures than the Logistic Regression. The Decision Tree obtained an accuracy of 0.906, ROC-AUC of 0.962, which means that the Decision Tree is very discriminative. Conversely, the accuracy and ROC-AUC of Logistic Regression were 0.760 and 0.816, respectively.

One of the most significant improvements was the increased capability of the model to detect non-discharged home patients. The non-home discharge was also significantly higher in the recall of 0.686 (Logistic Regression) to 0.932 (Decision Tree). This is a clinically significant improvement because it is important to identify such patients early to plan discharge and resources.

Table 2. *Model performance comparison*

Model	Accuracy	ROC AUC	Recall (Home)	Recall (non-Home)	Precision (Home)	Precision (non-Home)	F1 Score
Logistic Regression	0.760	0.816	0.780	0.686	0.904	0.452	0.837
Decision Tree	0.906	0.962	0.899	0.932	0.980	0.708	0.938

Figure 2 indicates the ROC of the two models, and at all thresholds, the Decision Tree always dominates the Logistic Regression curve, which once again supports its discriminative superiority.

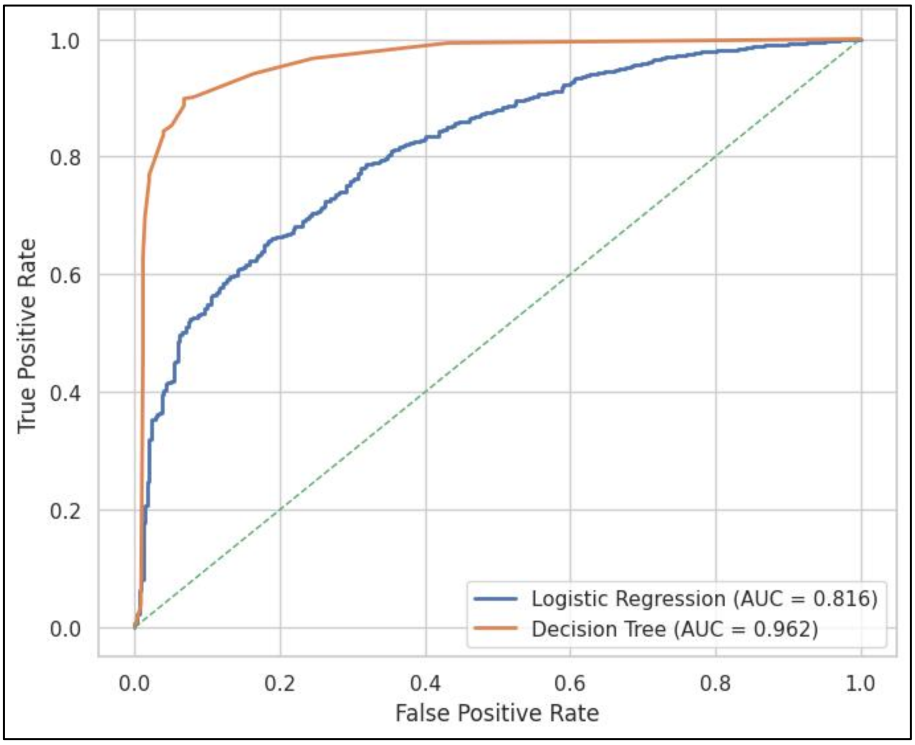

Figure 2. *Receiver Operating Characteristic Curves for Logistic Regression and Decision Tree Models*

3.3 Feature Importance and Model Interpretability

The analysis of the feature importance showed that a small group of variables was used to build the predictive model, and the concentration of importance was high among the top predictors. COD 80 was noted to be the most influential variable, with a total impact of about 67.2%. This was succeeded by COD 65 (15.9%) and COD 8010 (9.6), meaning that few clinical features are largely the determinants of discharge outcomes. Demographic variables, including gender and pension status, on the other hand, had little contribution to the model, so it is possible that the predictions are largely defined by clinical features and less by demographic ones.

Table 3. *Most influential predictors of discharge outcome*

Feature	Importance
COD 80	0.672
COD 65	0.159
COD 8010	0.096
COD 62	0.040
COD 85	0.024
COD 6701	0.006
Gender	0.002
GGPensione	0.001
COD 38	0.000
COD 76	0.000

Figure 3 presents the visual representation of the dominance of several key variables, in which the sharp drop in the importance of the variables after the top ones is seen. The number of variables that were significant to the model was limited to 10, which is an indication of the sparsity and interpretability of the model.

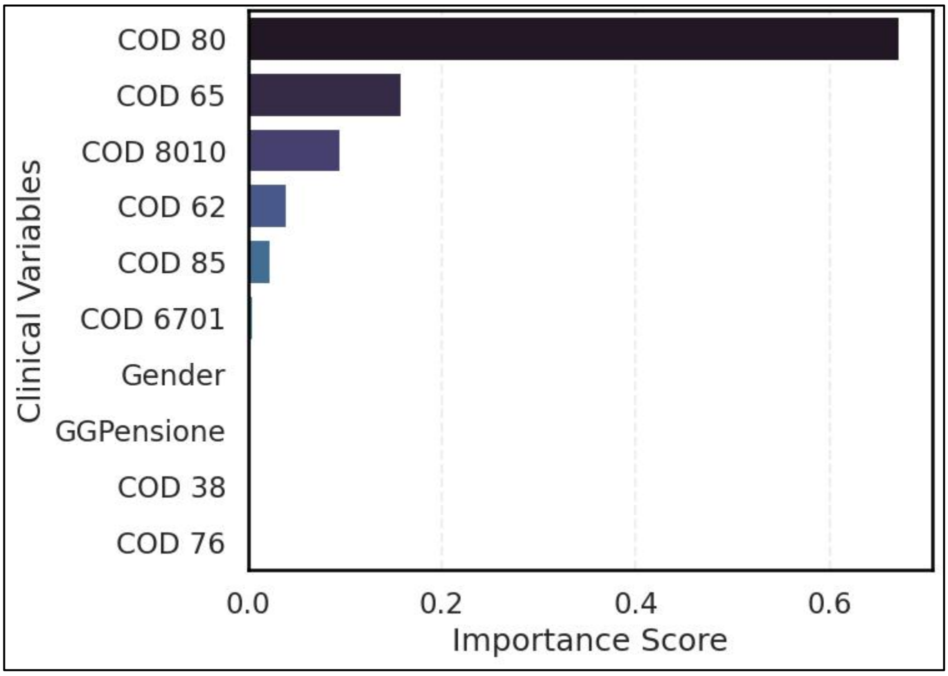

Figure 3. *Feature Importance from the Decision Tree Model for Discharge Outcome Prediction*

4. Discussion

The results indicate that, based on routine clinical data, predicting discharge outcomes in a rehabilitation setting with high accuracy is possible using a set of interpretable machine learning models. The Decision Tree model performed better than the Logistic Regression in all measures, especially discrimination and non-home discharge recall. This implies that the relationship between discharge decisions is complex and non-linear and is not captured by linear methods, but by rule-based methods. One of the outcomes was that the identification of patients not discharged home was significantly improved. This is also clinically significant because such patients usually need more care plans, coordination, and resource distribution. To the extent that a better recall means that the model is more appropriate when it comes to identifying patients who do not follow the anticipated discharge route. The importance analysis of features further demonstrated that a few variables contributed to most of the predictive power, and the demographic factors made little contribution. This focus indicates that a particular set of clinical indicators is central to the discharge outcomes and, therefore, simplified and interpretable decision-support instruments are possible.

Such results are in line with the previous studies that prove the importance of machine learning in discharge prediction and clinical decision support. Previous studies have demonstrated that machine learning models could be useful in facilitating discharge decision-making in intensive care by describing patient patterns that are not reflected by conventional methods (McWilliams et al., 2019). In the same way, interpretable machine learning models have been used to predict discharge disposition in post-stroke populations with meaningful performance and transparency (Cho et al., 2019). Machine learning has been applied in rehabilitation settings to forecast clinical outcomes based on routinely collected data, especially when patient trajectories are variable to the model (Harari et al., 2020). The explainable AI strategies have also shown that they can identify important predictors without losing clinical interpretability, which is a prerequisite to successful implementation (Gandolfi et al., 2022). Recent literature highlights that interpretable models have the potential to facilitate clinical reasoning by delivering outputs that are comprehensible and implementable (Petrović et al., 2025). The same strategies have been employed to estimate the discharge destination out of inpatient rehabilitation, which underscores the possibility of early recognition of patients who need alternative care pathways (Anderson et al., 2024). Furthermore, explainable models have been used to monitor patient progression throughout their admission to the discharge phase, which strengthens their use in aiding decision-making throughout the care continuum (Gkantzios et al., 2023). Disagreement in decision-making of discharge can lead to variability in model performance, and the lack of consistency in human annotations has been demonstrated to impact AI-driven clinical predictions in such settings (Sylolypavan et al., 2023).

The findings indicate that the discharge planning can be assisted with the help of interpretable machine learning models that allow to make early and accurate predictions related to patient outcomes. By collecting risk information on patients who will not be discharged home, clinicians can implement specific interventions, provide multidisciplinary care, and better distribute resources. The fact that only a few important variables were relied on further points towards the possibility of an actual development of practical decision-support tools with little to no complex or high-dimensional data. Interpretability of the Decision Tree model is especially important to clinical use, where trust and adoption require transparency. The concept of models that offer clear decision pathways can support the process of communication between healthcare professionals and shared decision-making with families and patients. Interpretable machine learning, as such, provides a practical solution to the incorporation of information-driven knowledge into everyday clinical practice.

There are a number of limitations that ought to be taken into consideration. It has been based on a single dataset, which can be a limitation to the generalizability of the results to other settings. Coded variables that are not clinically interpreted in detail also limit the possibility of direct translation of the results into certain clinical actions. Besides this, the binary outcome oversimplifies a more complicated discharge process and might not be able to account for all pertinent pathways. The other weakness is associated with the possible discrepancies in labeling outcomes. Future studies need to be directed towards external validation in various populations of rehabilitation and healthcare systems. The models can be enhanced and made more relevant by adding other clinical and social variables. Future research is also required to determine the practical effectiveness of the implementation of such models in clinical practice, and especially the process of decision-making and patient outcomes.

5. Conclusion

To predict the outcome of discharge accurately in rehabilitation, the methods used must be effective and also clinically interpretable. The current results demonstrate that systematically gathered clinical information can be used to make significant predictions of home and non-home discharge, and the Decision Tree model is more effective than the Logistic Regression based on all the principal evaluation criteria. Its greater success in recognizing the patients who will not go home is especially applicable in early discharge planning, multidisciplinary coordination, and allocation of resources. The findings also suggest that the few variables were used to predict, and demographic factors did not have a significant influence on model prediction. This justifies the possible merit of clear, narrow-purpose decision-support instruments capable of aiding clinicians without having to depend on excessively complicated modeling plans. The results must also be understood in the background of one data set and simplified outcomes. Interpretable machine learning seems to provide an effective and clinically viable method of predicting discharge outcomes in the rehabilitation environment. Such models remain to be externally validated and evaluated in the future before they can be regarded as a routine tool.

6. References

1. Adeniran, I. A., Efunniyi, C. P., Osundare, O. S., & Abhulimen, A. O. (2024). Data-driven decision-making in healthcare: Improving patient outcomes through predictive modeling. Engineering Science & Technology Journal, 5(8), 1-9.
2. Adlung, L., Cohen, Y., Mor, U., & Elinav, E. (2021). Machine learning in clinical decision making. Med, 2(6), 642-665.
3. Anderson, H., Polovneff, A., & Durand, M. (2024). Using Machine Learning to Predict Discharge Destination from Inpatient Rehabilitation. Archives of Physical Medicine and Rehabilitation, 105(4), e48.
4. Bhardwaj, A. (2022). Promise and provisos of artificial intelligence and machine learning in healthcare. Journal of Healthcare Leadership, 113-118.
5. Buscarini, L., Romano, P., Cocco, E. S., Damiani, C., Pournajaf, S., Franceschini, M., & Infarinato, F. (2024). Clinical Dataset for Artificial Intelligence-Driven Predictive Modeling for Home Discharge in Neurological and Orthopedic Conditions [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10991206
6. Chifu, E. S., Chifu, V. R., Pop, C. B., Vlad, A., & Salomie, I. (2018, September). Machine Learning based technique for detecting daily routine and deviations. In 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP) (pp. 183-189). IEEE.
7. Chifu, V. R., Pop, C. B., Rancea, A. M., Morar, A., Cioara, T., Antal, M., & Anghel, I. (2022). Deep learning, mining, and collaborative clustering to identify flexible daily activities patterns. Sensors, 22(13), 4803.
8. Cho, J., Alharin, A., Hu, Z., Fell, N., & Sartipi, M. (2019, December). Predicting post-stroke hospital discharge disposition using interpretable machine learning approaches. In 2019 IEEE International Conference on Big Data (Big Data) (pp. 4817-4822). IEEE.
9. Considine, J., Berry, D., Sprogis, S. K., Newnham, E., Fox, K., Darzins, P., ... & Street, M. (2020). Understanding the patient experience of early unplanned hospital readmission following acute care discharge: a qualitative descriptive study. BMJ open, 10(5), e034728.
10. Gandolfi, M., Galazzo, I. B., Pavan, R. G., Cruciani, F., Vale, N., Picelli, A., ... & Menegaz, G. (2022). eXplainable AI allows predicting upper limb rehabilitation outcomes in sub-acute stroke patients. IEEE Journal of Biomedical and Health Informatics, 27(1), 263-273.
11. Gkantzios, A., Kokkotis, C., Tsiptsios, D., Moustakidis, S., Gkartzonika, E., Avramidis, T., ... & Vadikolias, K. (2023). From admission to discharge: predicting national institutes of health stroke scale progression in stroke patients using biomarkers and explainable machine learning. Journal of Personalized Medicine, 13(9), 1375.

12. Gledhill, K., Bucknall, T. K., Lannin, N. A., & Hanna, L. (2023). The role of collaborative decision-making in discharge planning: Perspectives from patients, family members and health professionals. Journal of clinical nursing, 32(19-20), 7519-7529.
13. Harari, Y., O'Brien, M. K., Lieber, R. L., & Jayaraman, A. (2020). Inpatient stroke rehabilitation: prediction of clinical outcomes using a machine-learning approach. Journal of neuroengineering and rehabilitation, 17(1), 71.
14. Luther, B., Wilson, R. D., Kranz, C., & Krahulec, M. (2019). Discharge processes: What evidence tells us is most effective. Orthopaedic nursing, 38(5), 328-333.
15. Lyu, G. (2025). Data-driven decision making in patient management: a systematic review. BMC medical informatics and decision making, 25(1), 239.
16. McWilliams, C. J., Lawson, D. J., Santos-Rodriguez, R., Gilchrist, I. D., Champneys, A., Gould, T. H., ... & Bourdeaux, C. P. (2019). Towards a decision support tool for intensive care discharge: machine learning algorithm development using electronic healthcare data from MIMIC-III and Bristol, UK. BMJ open, 9(3), e025925.
17. Petrović, I., Njegovan, S., Tomašević, O., Vlahović, D., Rajić, S., Živanović, Ž., ... & Jorgovanović, N. (2025). Dynamic, Interpretable, Machine Learning–Based Outcome Prediction as a New Emerging Opportunity in Acute Ischemic Stroke Patient Care: A Proof-of-Concept Study. Stroke Research and Treatment, 2025(1), 3561616.
18. Rubinger, L., Gazendam, A., Ekhtiari, S., & Bhandari, M. (2023). Machine learning and artificial intelligence in research and healthcare. Injury, 54, S69-S73.
19. Saxena, A., & Chandra, S. (Eds.). (2021). Artificial intelligence and machine learning in healthcare. Singapore: Springer.
20. Sylolypavan, A., Sleeman, D., Wu, H., & Sim, M. (2023). The impact of inconsistent human annotations on AI driven clinical decision making. NPJ Digital Medicine, 6(1), 26.
21. Ward-Stockham, K., Omonaiye, O., Darzins, P., Kitt, C., Newnham, E., Taylor, N. F., & Considine, J. (2024). Understanding the influences on hospital discharge decision-making from patient, carer and staff perspectives. BMC Health Services Research, 24(1), 1097.