

Arteriovenous fistula flow level prediction through ordinal classifier methods

Mario Garbelli¹, Luca Neri¹, and Francesco Bellocchio¹

¹ Fresenius Medical Care AG & Co. KGaA, Bad Homburg, Germany

Abstract

Arteriovenous fistula (AVF) is the conjunction of an artery and a venous, surgically created as point of access to blood stream in dialysis patients. The measure of blood flow in AVF (Qa) is an important indicator to assess the functioning of AVF and the consequent adequacy for effective dialysis treatment. The measurement of Qa requires time-consuming and costly procedures which are generally performed only for a subset of patients considered at risk of problems. Our hypothesis is that there is a set of information collected during dialysis treatment that can be exploited to estimate Qa. The objective of this study is twofold: analyze the accuracy of Qa estimation from routinely collected clinical and dialysis machine data variables and compare different approaches for this problem. We tackled this estimation as an ordinal classification problem. As in machine learning domain there are different approaches to include order information in multiclass classification, we explored pros and cons of such different methods.

Keywords

Ordinal Multiclass Classification, Hemodialysis, Arteriovenous fistula.

1. Introduction

Chronic kidney disease is an increasing long-term condition characterized by a deterioration of kidneys functioning. After a certain level of renal damage, the patient needs renal replacement therapy (RRT) to survive. There are two kinds of RRT: hemodialysis and peritoneal dialysis. The former (and most common) consists in re-moving fluid and waste products from the blood using a special filter called dialyzer. To take the blood from the patient and filter through the dialyzer and return it to the patient, the doctor needs to make access to the patient's blood vessel, called vascular access. There are several types of vascular access, arteriovenous fistula (AVF) is one associated with lower incidence of complication and longer patient's survival. It is a conjunction of an artery and a venous, surgically created, to exploit the blood pressure of the artery and guarantee an adequate blood flow through the dialyzer.

Continuous monitoring of AVF functioning is an important aspect of dialysis patient management [1] [2] as if the AVF doesn't work properly the alternative vascular access (typically, a central catheter) would increase the risk of complications and decrease the effectiveness of dialysis treatment.

The assessment of AVF blood flow (Qa) is an important prognostic test for evaluating the status of vascular access [1][2] and it is used to consider clinical procedures to reestablish the patency of AVF. This test can be performed in several way (doppler ultrasound, angiography, Body Thermal Monitor, ...) with different level of accuracy. These tests are time-consuming for clinical staff, operator dependent, and costly for healthcare providers, so they are limited to those patients with some signs of AVF malfunctioning.

On the other hand, most of the dialysis machines automatically collect, at each dialysis treatment, a set of data related to treatment blood flow and needles pressure which can be exploited to estimate the AVF blood flow. Furthermore, the information present in the patient health records, such as age,

HC@AIxIA 2022: 1st AIxIA Workshop on Artificial Intelligence For Healthcare
EMAIL: mario.garbelli@fmc-ag.com; luca.neri@fmc-ag.com; francesco.bellocchio@fmc-ag.com



© 2020 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

laboratory tests, comorbidities, and others, might influence the AVF blood flow. In this study we focused on the prediction of the AVF flow exploiting all this routinely collected data.

The value of AVF blood flow is generally categorized in clinical meaningful levels: < 525 ml/h (very low level), 525 ml/h - 925 ml/h (low level), > 925 ml/h (normal level).

Considering this categorization, AVF flow estimate can be seen as an ordinal classification problem, as there is a clear inherent order among the classes but not a well-defined numerical distance.

This kind of multiclassification can be tackled with different machine learning approaches. We have selected two kinds of ordinal classification methods: the first transforms ordinal targets into binary classification subtasks [3][4] and the second is based on a single deep neural network combined with a set of binary learning tasks [5]. The main difference between these two approaches is that the first one is generally less computational costly but suffers from classifier inconsistencies among the binary rankings [5].

In this study we also compare the ordinal classification methods with a model based on a standard regression method.

We already explored the application of machine learning approaches for the monitoring of AVF status in [6]. In that study we focused on the assessment of general AVF problem risk (any kind of event requiring physician intervention). Here, we focused on the prediction of a measurable variable that is well known and used by the physicians, simplifying the introduction of the model in clinical practice.

2. Method

2.1. Ordinal classification

Ordinal multi-classification problem consists in a task of predicting labels characterized by an ordinal scale. An ordinal classifier C maps each object $x^{[i]} \in X$ into an ordered set $l: X \rightarrow Y$, where $Y = \{r_1 < \dots < r_k\}$. The labels are characterized by the fact that it is possible to define an order among them but not a measure of distance.

Although, in Machine Learning there are many algorithms for general multi-classification problems, there are much less able to consider the order information.

There are mainly two classes of algorithms for this kind of problem. The first [3] exploits the use of a set of $K - 1$ binary classifiers. In our experiments, we used a XGBoost model for generating the classifiers. We call this model BINCM.

The estimation of the first and the last class depends on a single classifier, $P(Qa < 525)$ is given by the first classifier and $P(Qa \geq 925)$ is given by the last classifier. The probability of the middle class is computed as $P(Qa \geq 525) \cdot (1 - P(Qa \geq 925))$.

This approach suffers from the possible inconsistency of the predicted probabilities [5]. This inconsistency, since the predictions for individual binary tasks may disagree, can lead to a contradictory result for which,

$$P(Qa \geq 925) > P(525 < Qa < 925) \text{ and } P(Qa \geq 925) < (P(Qa \geq 525)).$$

The second approach, called rank-consistent ordinal regression framework (CORN), is based on a neural network with an output layer that implements a set of $K - 1$ binary learning tasks associated with the ranks [5][7].

In this study we applied the method described in [5] and we summarize it in the following.

Given a dataset $D = \{x^{[i]}, y^{[i]}\}_{i=1}^N$, a rank $y^{[i]}$ is transformed into $K - 1$ binary labels $y_1^{[i]}, \dots, y_{k-1}^{[i]}$ such that $y_k^{[i]} \in \{0,1\}$ and indicates if $y^{[i]}$ is greater than r_k . To ensure classifiers consistency, the model estimates a series of conditional probabilities such that the output of the k -th binary task $f_k(x^{[i]})$ represents the conditional probability

$$f_k(x^{[i]}) = P(y^{[i]} > r_k | y^{[i]} > r_{k-1}), \text{ where } \{y^{[i]} > r_k\} \subseteq \{y^{[i]} > r_{k-1}\}$$

The unconditional probabilities can be derived through the chain rule for probabilities to the model outputs:

$$P(y^{[i]} > r_k) = \prod_{j=1}^k f_j(x^{[i]})$$

Since $\forall j 0 \leq f(x^{[i]}) \leq 1$, we have

$$P(y^{[i]} > r_1) \geq P(y^{[i]} > r_2) \geq \dots \geq P(y^{[i]} > r_{k-1})$$

This guarantees the rank consistency among the $k - 1$ binary tasks.

To evaluate the benefits of using ordinal classifier approaches than a classical regression approach, we used as benchmark an XGBoost regressor to estimate the exact value of Qa and, after that, we map these values to the three classes (very low flow, low flow, normal flow) using the thresholds 525 ml/h and 925 ml/h. We call this model REGM.

The accuracy comparison among the three methods was assessed computing mean absolute error (MAE), precision and recall, F1-score and confusion matrix

2.2. Dataset

Our dataset was composed of 46,292 Qa measurements referred to 5,940 hemodialysis patients from four different European countries (Czech Republic, Portugal, Slovakia, Spain) and collected between 2015 and 2022. The Qa distribution was shown in Figure 1.

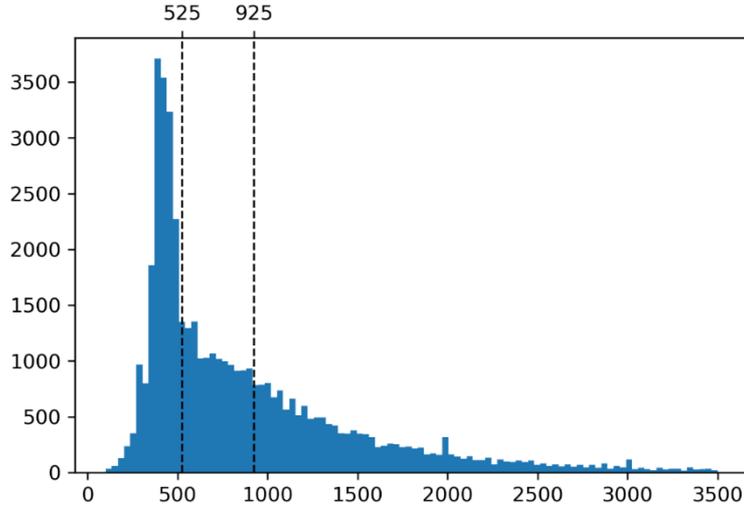


Figure 1: Qa measures distribution

Written informed consent for statistical analysis was obtained from all the patients. Qa measurements are mainly obtained through dilution technique (thermal 93.34% and other 6.27%) and only a few measures (0.39%) were obtained through doppler ultrasound.

The input dataset was composed of 49 parameters that characterize patient, AVF and treatments routinely recorded through Fresenius Medical Care clinical database (EuCIID®). The input variables were computed collecting the information in the 90 days before the Qa measure.

70% of the dataset was randomly selected to train each model, 10% was used as validation set (for CORN) and the remaining 20% was used as test set.

3. Results

In Figures 2, 3, and 4 we reported the confusion matrix with the percentage of labels correctly and incorrectly classified for the three classes for: REGM, BINCM and CORN respectively. The overall MAE was 0.37, 0.27 and 0.36.

The results show that both ordinal classification models performed better than the benchmark in terms of overall MAE, and BINCM outperformed CORN (0.27 Vs. 0.36). It is interesting to note that when we looked at the best recall for single class, we found that the best one for “very low” and “low” flow is still BINCM but for normal flow the best one is REGM. On the other hand, the best one in terms of precision for “very low” flow is REGM and for “low” and “normal” flow it is BINCM.

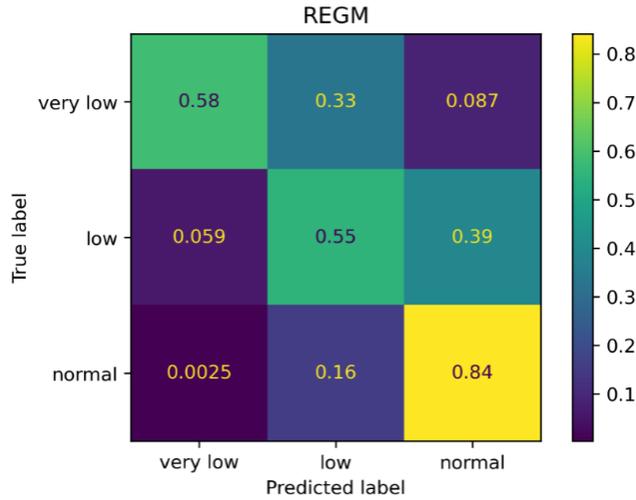


Figure 2: Confusion matrix (CM) for REGM. The diagonal elements of CM represent the number of right matches between predicted and actual (true) label, while off-diagonal elements are those that are mislabeled. The CM values are normalized over actual label (by row)

Table 1

Accuracy performance metric for REGM

	precision	recall	F1-score	MAE	support
very low	0.93	0.58	0.72	0.50	3578
low	0.45	0.55	0.49	0.45	2470
normal	0.68	0.84	0.75	0.16	3211
overall	0.71	0.66	0.67	0.37	9259

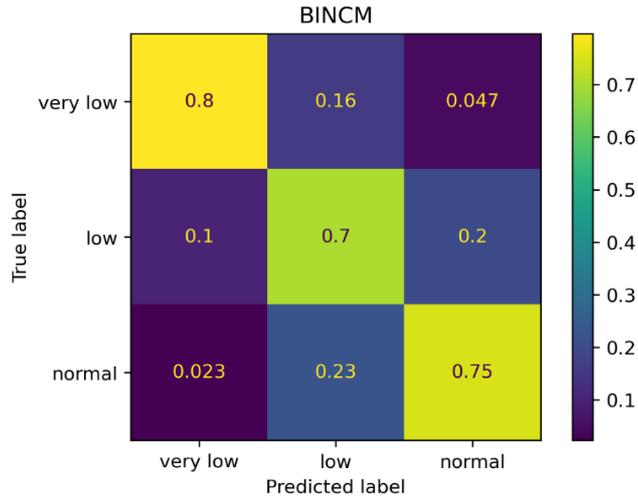


Figure 3: Confusion matrix for BINCM

Table 2

Accuracy performance metric for BINCM

	precision	recall	F1-score	MAE	support
very low	0.9	0.8	0.84	0.25	3578
low	0.57	0.7	0.63	0.30	2470
normal	0.79	0.75	0.77	0.27	3211
overall	0.77	0.76	0.76	0.27	9259

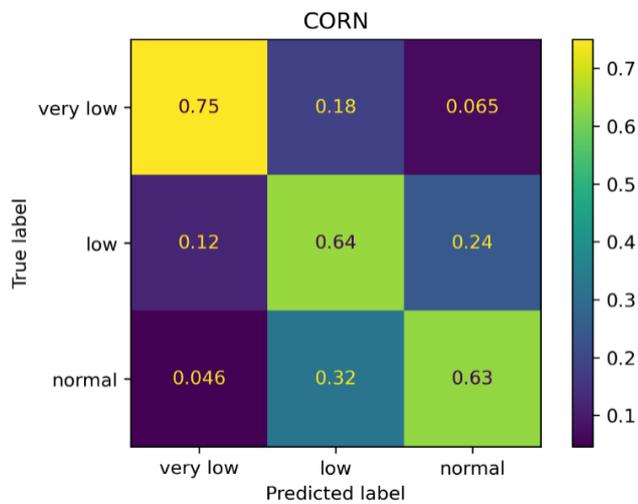


Figure 4: Confusion matrix for BINCM

Table 3

Accuracy performance metric for CORN

	precision	recall	F1-score	MAE	support
very low	0.86	0.75	0.80	0.32	3578
low	0.48	0.64	0.55	0.36	2470
normal	0.71	0.63	0.67	0.41	3211
overall	0.71	0.68	0.69	0.36	9259

4. Discussion

In this exploratory paper we studied the estimation of AVF blood flow exploiting routinely collected clinical data extracted from a clinical database. AVF blood flow assessment is critically important to assess the status of the fistula and evaluate preemptive intervention to avoid AVF failure, that may lead to the risk of complications.

As AVF blood flow is generally categorized in three medical meaningful classes (very low when less than 525 ml/h, low when between 525 and 925 ml/h and normal when greater than 925 ml/h), we tackled the problem as an ordinal classification task.

There are several approaches in machine learning domain for ordinal classification, we focused on two techniques: ordinal classification with K-1 binary classifiers and CORN. We used a popular regression model (XGBoost) with post classification as benchmark.

The objective of this study can be summarized in two points: first, evaluate the accuracy of AVF blood estimation to assess its possible medical use and second, compare different Machine Learning approaches for this problem.

Regarding the first point, the preliminary results are very promising. The ability to detect fistula with very low flow with a precision of 0.9 (with BINCM) is a remarkable result per se. It should be also noted that 7.3 % of the missed “very low flow” AVFs are predicted as “low flow” and only 2.7% are predicted as normal flow. This suggests that the model could be used to build an alert system to warn physicians about AVFs in critical situations. It might substantially improve the AVF surveillance process, referring the patient for further investigation when it is really needed and potentially decrease the risk of AVF failure.

Regarding the comparison among different approaches, the results are interesting as well. If we focus on the overall MAE and F1-score, it seems that the possible inconsistency of probabilities classes might be an acceptable tradeoff for our problem as those metrics are much better for the BINCM than CORN (0.27 and 0.76 Vs. 0.36 and 0.69). Although CORN results less accurate than BINCM, it can be a useful tool for the application where it is important to show to the physician the single probability of the three classes. In this case, the inconsistency of the probabilities would be difficult to understand.

It is also important to consider that if we focus on the performance of the models in the single classes, the best model is not always the same. This suggests that those approaches have different peculiarities that can be exploited to create a more robust and effective medical support decision system.

Considering the promising preliminary results, we see different next steps for this study, for example: a deep feature analysis to explore the contribution of each variable to the prediction; evaluating further ordinal classification methods; discuss with clinicians for a possible use case of the model in real clinic setting and how integrate it in the clinical system and in clinical practice.

5. References

- [1] C.E. Lok, T.S. Huber, T. Lee, S. Shenoy, A.S. Yevzlin, K. Abreo, M. Allon, A. Asif, B.C. Astor, M.H. Glickman, et al., KDOQI Clinical Practice Guideline for Vascular Access: 2019 Update. *Am. J. Kidney Dis.*, 75, S1–S164, (2020)..
- [2] M. Gallieni, M. Hollenbeck, N. Inston, M. Kumwenda, S. Powell, J. Tordoir, J. Al Shakarchi, P. Berger, D. Bolignano, D. Cassidy, et al., Clinical practice guideline on peri- and postoperative care of arteriovenous fistulas and grafts for haemodialysis in adults. *Nephrol. Dial. Transplant*, 34 (Suppl. 2), II1–II42, (2019).
- [3] F. Eibe and M. Hall. A simple approach to ordinal classification. In *Proc 12th European Conference on Machine Learning*, Freiburg, Germany, pages 145--156. Springer, (2001)
- [4] L. Li, H.T. Lin, Ordinal regression by extended binary classification, *Advances in Neural Information Processing Systems*, pp. 865-872, (2007).
- [5] X. Shi, W. Cao, S. Raschka, Deep Neural Networks for Rank-Consistent Ordinal Regression Based On Conditional Probabilities, *CoRR*, <https://arxiv.org/abs/2111.08851>, (2021)
- [6] R. Peralta, M. Garbelli, F. Bellocchio, P. Ponce, S. Stuard, M. Lodigiani, J. Fazendeiro Matos, Ribeiro, M. Nikam, M. Botler, E. Schumacher, D. Brancaccio, L. Neri, Development and

Validation of a Machine Learning Model Predicting Arteriovenous Fistula Failure in a Large Network of Dialysis Clinics. *Int J Environ Res Public Health*. 2021 Nov 24;18(23):12355. doi: 10.3390/ijerph182312355. PMID: 34886080; PMCID: PMC8656573.

- [7] W. Cao, V. Mirjalili, S. Raschka, Rank Consistent Ordinal Regression for Neural Networks with Application to Age Estimation. *Pattern Recognition Letters*. <https://doi.org/10.1016/j.patrec.2020.11.008>, (2020).