

Bulletin de la Dialyse à Domicile

Home Dialysis Bulletin (BDD)

International bilingual journal for the exchange of knowledge and experience in home dialysis

(English edition) (version française disponible à la même adresse)

Predicting the Need for Hemodialysis in Peritoneal Dialysis Patients with at Least One Episode of Peritonitis Using Artificial Intelligence

(Prédiction à l'aide de l'intelligence artificielle du passage en hémodialyse pour des patients en dialyse péritonéale présentant au moins une péritonite)

Alexander Watson ¹, Rémy Lozano ¹, Yannick Toussaint ²

¹ Ecole des Mines (Nancy, France), ² Laboratoire Lorrain de Recherche en Informatique et ses Applications (Nancy, France)

To cite: Watson A, Lozano R, Toussaint Y. Predicting the Need for Hemodialysis in Peritoneal Dialysis Patients with at Least One Episode of Peritonitis Using Artificial Intelligence. Bull Dial Domic [Internet]. [cited 2026 Aug. 29];9(3):115-26. doi: <https://doi.org/10.25796/bdd.v9i3.87120>

Summary

Transfer to hemodialysis is a major cause of permanent discontinuation of peritoneal dialysis (PD) treatment. Peritoneal infection is a common cause of discontinuation of peritoneal dialysis and permanent transfer to hemodialysis (HD). Accurately predicting the approximate date of transfer allows for the preparation of a vascular access under optimal conditions. Accurately predicting the date of transfer is therefore of clinical importance. The goal of this study is to develop an artificial intelligence model that predicts, based on a patient's history of peritoneal infection, whether the patient should be transferred to hemodialysis.

Our approach is based on a neural network learning method, specifically a Long Short-Term Memory (LSTM) recurrent neural network (RNN); the training set, representing 80% of the dataset—that is, 9,934 patient sequences—was used to optimize the model's parameters. The test set (10%, or 1,242 sequences) was kept strictly separate from the other sets and reserved exclusively for the final evaluation of the model's performance.

The monitoring of patients' conditions over time was based on peritoneal infections. Recurrent neural networks are particularly well-suited to this type of prediction based on time-series data. This study highlighted the predictive performance of such a model on this type of data. Introducing a bias into the training process helped limit false-negative predictions.

Our data, drawn from the French-Language Peritoneal Dialysis Registry (RDPLF) database, included 15,800 patients who had experienced at least one episode of peritonitis, with the aim of predicting the likelihood of transfer to hemodialysis following this complication. With our model, the false-negative rate—the proportion of patients who were transferred to HD within 6 months even though the model did not predict it—was 5.3%.

Keywords: peritoneal dialysis, artificial intelligence, neural network, peritonitis, transfer, hemodialysis

Résumé

Le transfert en hémodialyse est une cause importante d'arrêt définitif de traitement par dialyse péritonéale (DP). L'infection péritonéale est une cause fréquente de l'arrêt de la dialyse péritonéale et du transfert définitif en hémodialyse (HD). Une prédiction correcte de la date approximative de transfert permet de prévoir la préparation de l'abord vasculaire dans des conditions optimales. La prédiction correcte de la date de transfert revêt donc un intérêt clinique. Le but de ce travail est de construire un modèle d'intelligence artificielle qui prédit, étant donné l'historique d'infection péritonéale, si le patient doit être transféré en hémodialyse.

Notre approche repose sur une méthode d'apprentissage à base de réseau neuronal, un réseau neuronal récurrent (RNN) de type Long Short-Term Memory (LSTM) ; l'ensemble d'entraînement, représentant 80% du jeu de données exploité, soit 9 934 séquences patients, a été utilisé pour l'optimisation des paramètres du modèle. L'ensemble de test (10%, soit 1 242 séquences) a été quant à lui maintenu strictement isolé des autres ensembles et réservé exclusivement à l'évaluation finale des performances du modèle. Le suivi de l'état des patients dans le temps était composé des épisodes d'infections péritonéales. Les réseaux de neurones récurrents se montrent particulièrement adaptés à ce type de prédiction à partir de données évoluant dans le temps. Ce travail a permis de mettre en avant les performances de prédiction d'un tel modèle sur ce type de données. L'introduction d'un biais dans l'apprentissage a permis de limiter les prédictions de faux négatifs.

Nos données, issues de la base de données du Registre de Dialyse Péritonéale de Langue Française (RDPLF), contiennent 15800 patients qui ont présenté au moins une péritonite, dans le but de prédire la possibilité de transfert en hémodialyse à la suite de cette complication. Avec notre modèle le taux de faux négatifs, le taux de patients qui ont été transférés en HD dans les 6 mois alors que le modèle ne le prédisait pas, était 5.3%.

Mots-clés : dialyse péritonéale, intelligence artificielle, réseau de neurones, péritonite, transfert, hémodialyse



Article licensed under a Creative Commons Attribution 4.0 International: <https://creativecommons.org/licenses/by/4.0/>

Copyright: les auteurs conservent le copyright.

Introduction

According to the RDPLF, 36% of the causes of permanent discontinuation of peritoneal dialysis treatment are related to a switch to hemodialysis. However, only 8% of patients in France have an arteriovenous fistula (AVF)[1] . This necessitates the initial use of a central catheter in the majority of cases due to the lack of a pre-existing vascular access[2] . This method significantly increases the risk of infection and the subsequent risk of vascular access loss[3] .

Reliably predicting the date of transition to hemodialysis is of clinical importance: it would allow for the patient's transfer to hemodialysis to be planned well in advance, providing sufficient time to create a functional arteriovenous fistula. The maturation time for an AVF is usually 3 weeks; however, in practice, the time required for preliminary examinations, the surgeon's availability, and operating room scheduling necessitate planning well in advance. According to nephrologists, the ability to predict the transition to hemodialysis 3 months in advance would provide the best chance of starting hemodialysis using a functional AVF and would therefore avoid the need for a central catheter.

A preventive approach consisting of routinely creating an arteriovenous fistula for all patients upon admission to peritoneal dialysis (PD) might seem like a simple solution to this problem. However, this option is not clinically desirable for two reasons. First, an arteriovenous fistula may have a limited lifespan and become nonfunctional after a few years, even without having been used. Second, although they remain rare, infectious or hemodynamic complications are always possible.

In France, the occurrence of one or more episodes of peritonitis accounts for 14% of transfers to hemodialysis[1] , often in emergency situations. An initial attempt to predict the transition to hemodialysis using a Random Forest model was unsuccessful, so it became necessary to employ more sophisticated techniques. Deep learning techniques using neural networks have proven effective in predicting complications for patients on PD; in particular, recurrent neural networks (RNNs) are used for time-series forecasting[4–6] . Our objective was therefore to train an RNN using data from patients on PD who had experienced at least one episode of peritonitis, and to evaluate its performance in predicting the risk of transfer to hemodialysis following peritonitis.

Materials and Methods

The data were extracted from the RDPLF database from centers in metropolitan France after being fully anonymized using a randomly generated hash key. All work was performed in Python using the following libraries: *pandas*, *scikit-learn*, *TensorFlow/Keras*, *Matplotlib*, and *seaborn*.

We used two tables from the RDPLF dataset: the “*patient*” table contained 40,952 patients, and the “*peritonitis*” table represented a subset of the “*patient*” table and included 15,800 patients with between 1 and 20 episodes of peritonitis.

After excluding missing data, the following variables, recorded for each time point, were included: the patient's dialysis center, sex, nephrology group, reason for treatment, previous treatment, previous type of PD, type of PD 90 days prior, transplant list status, peritonitis pathogen class, cause of peritonitis, treatment system and type, the number of support devices the patient uses, the

patient's total number of peritonitis episodes (at each time point), the Charlson score at initiation of PD, height, weight, the number of metabolic conditions, the number of treatment sessions, the age at the start of PD, and the patient's age at each time point.

Several variables were then added to the dataset. We therefore constructed the following variables: PD Duration, the length of time a patient has been on PD; Delta Days, indicating the number of days that have elapsed since the last episode of peritonitis; and 6-Month Infection Density, giving the number of episodes of peritonitis for the patient during the last 6 months on peritoneal dialysis.

We then constructed time series for each patient in the database. Since the patients' time series were not directly available in the database, they were constructed using the dates of the episodes of peritonitis as temporal markers. At the initial time point, all variables from the patient table were populated and retained at every point in the time series, as these variables are static by nature. Age is the only exception: it was recalculated at each point in the sequence by calculating the difference between the date of the corresponding peritonitis episode and the patient's date of birth. The variables that change over time are therefore primarily those from the peritonitis table, which means that the model bases its decisions largely on characteristics specific to the patients' peritonitis episodes. Since the number of peritonitis episodes varies from one patient to another, the sequences have varying lengths. As the maximum observed length was 20 episodes, this value was adopted as the fixed length for all sequences. *Post-padding* was applied to sequences of shorter length. The *post-padding* technique pads sequences shorter than 20 with zero values at the end of the sequence; these values are excluded from the calculation during the training phase using the masking index reserved for this purpose. The added values do not affect the error calculation during training, but they allow sequences of different lengths to be treated the same way by the model.

The categorical variables in the dataset required specific encoding. Since recurrent neural networks cannot directly process categorical variables, a vector representation of these variables was implemented. This step applies to all categorical variables in the training dataset, such as the *PreTypDP* variable, which encodes the type of peritoneal dialysis initially performed (mixed PD (Association of APD+CAPD), CAPD, daily APD, IDP). An independent categorical encoding (*label encoding*) was first applied to each categorical variable, assigning a unique integer to each of its categories (*Table I*), with an offset of 1 in order to reserve the index 0 for *padding*, which is used to align sequences of variable length. The index of each variable value was then converted into a vector representation. The Word2Vec algorithm, in its Skip-gram configuration, was thus trained exclusively on sequences derived from the training data to avoid any information leakage. The Skip-gram model thus learns, for each token (index), to predict the tokens present in a 5-token context window (using 5 neighboring tokens), thereby capturing the co-occurrence relationships between different variable values within patients' clinical trajectories. Each categorical value is thus represented by a dense vector of dimension 50 (*Table II*). This dimension was chosen to allow for a sufficiently differentiated representation of the categories present in the data without making them too large during model training. Once initialized, the vector representations of the categorical variable values are fed as a weight matrix into the input layer of the neural network (*embedding*). These weights evolve during model training to adjust the representation of the categories (*fine-tuning*).

Table I. Example of categorical encoding for the PreTypDP variable

Value	Token (Index)
No data (padding/mask)	0
Mixed PD (CC & CA)	1
Daily APD	2
CAPD	3
DPI	4

Table II. Dimensions by value for PreTypDP

Values	Dim 1	Dim 2	Dim 3	Dim 4	...	Dim 50
Mixed PD (CAPD & APD)	-0.0593	0.0272	0.0170	0.0284	...	0.0082
Daily APD	-0.0425	0.0217	-0.0088	0.0080	...	0.0034
CAPD	0.0046	-0.0024	0.0120	0.0182	...	0.0103
IAPD	0.0156	-0.0190	-0.0004	0.0069	...	-0.0190

CAPD: continuous ambulatory peritoneal dialysis; APD: automated peritoneal dialysis; IPD: intermittent automated peritoneal dialysis; PreTypDP: initial PD type (CAPD, APD or mixed); mixed PD (CC and CA): a combination of CAPD and APD CAPD: continuous ambulatory peritoneal dialysis; APD: automated peritoneal dialysis; IPD: intermittent automated peritoneal dialysis; PreTypDP: initial PD type (CAPD, APD or mixed); mixed PD (CC and CA): a combination of CAPD and APD

A binary target variable was introduced to supervise the model's training. At each point in the sequence—that is, for each episode of peritonitis—this variable indicates whether the patient underwent hemodialysis within 6 months of the episode in question. It takes the value 1 if the patient underwent HD within this time window and 0 otherwise. Thus, a positive prediction made during a peritonitis episode allows, in the best-case scenario, for the transition to HD to be anticipated up to 6 months in advance, providing sufficient time to plan for fistula creation.

For our model architecture, we selected an LSTM-type RNN ([4,7]). This model relies on two distinct input streams (Figure 1): previously standardized numerical variables and categorical variables in the form of 50-dimensional vectors. These two streams are concatenated at each time step to form a state vector, which serves as the input to the recurrent layer. The sequence is then processed by the LSTM layer, which produces a hidden state at each time step and thus preserves all intermediate temporal information. To limit overfitting, we applied L2 regularization to this layer—which introduces the sum of the squares of the weights as a penalty term—and *dropout*, which randomly freezes a fixed proportion of the input weights at each time step. The LSTM layer is followed by a dense layer with a hyperbolic tangent activation function (Figure 1). *Dropout* and an L2 regularization factor are again applied to this dense layer (Figure 1). The output layer is a fully connected layer followed by a sigmoid activation function to obtain an output probability of transitioning to HD at each time step of the input sequence (Figure 1). Thus, each prediction incorporates all available information up to the current time step, including the hidden states propagated by the LSTM layer from previous time steps.

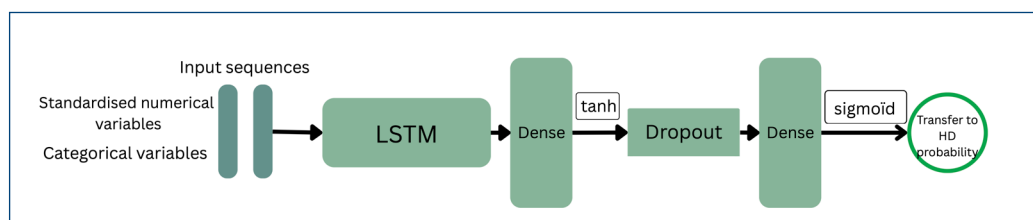


Figure 1. Model architecture

Before training the model, we partitioned the dataset into three distinct patient sets while maintaining the same proportion of patients transferred to HD in each set. The training set, representing 80% of the dataset used—or 9,934 patient sequences—was used to optimize the model parameters. The validation set (10%, or 1,242 patient sequences) was used to monitor the progress of the training. The test set (10%, or 1,242 patient sequences) was kept strictly separate from the other two sets and reserved exclusively for the final evaluation of the model's performance. Finally, the partition was constructed to ensure sufficient representation of patients who underwent hemodialysis within each of the three sets.

The model was then trained using this partition of the dataset. The number of epochs—once the entire training set had been processed—was set to 50. During each epoch, the model's weights were optimized using the Adam stochastic gradient descent algorithm. The gradient was calculated using a loss function that measured the difference between the model's predictions on the training set and the true values in that set. We selected weighted binary cross-entropy (WBCE) as the loss function[8,9] . It is based on the standard binary cross-entropy (BCE) but addresses the imbalance in the predicted categories, since the transition to HD is associated with only a single point in the sequence for the majority of patients. It also penalizes errors in predicting the transition to HD more heavily.

$$BCE = -p_1 \times \log(p_1) - p_0 \times \log(p_0)$$

$$WBCE = BCE \times [y_{\text{true}} \times 3.88 + (1 - y_{\text{true}})]$$

Where p_1 represents the probability of transition to HD, $p_0 = 1 - p_1$ et y_{true} , and y_{true} represents the actual value of the target variable. The term 3.88 is the ratio of the number of negative examples to the number of positive examples in the training set.

During training, the evaluation metric was the area under the ROC curve (AUC, Area Under the Curve)[10] , which reflects the model's discriminative power—that is, its ability to distinguish between the two classes. We also added an early stopping mechanism that prematurely terminated training after five epochs without any improvement in the AUC. The model's performance was also evaluated graphically through its calibration.

The choice of hyperparameters (*Table III*) was made empirically. We trained the model multiple times with different hyperparameter values, averaging the results across five training runs to ensure reliability. The dropout values represent the proportion of weights frozen in the corresponding layer of the model (*Table III, Figure 1*). The regularization factors apply to the LSTM layer and to the dense layer following the LSTM.

↓ *Table III. Selected hyperparameter values for model training*

Hyperparameters	Selected value
LSTM dropout	0.1
LSTM L2 regularization factor	0.001
L2 regularization factor for the dense layer	0.001
Dropout after the dense layer	0.1
Adam optimizer learning rate	0.001

Results

Training was performed over five successive independent iterations. The final predictions on the test set were obtained by calculating the arithmetic mean of the model's output probabilities.

For each patient, the model returned a sequence of probabilities of transition to HD following each episode of peritoneal infection. We also evaluated the model's predictions using different decision thresholds for transition to HD (Table IV). Thus, any predicted probability above the threshold was counted as a transition to HD.

Table IV. Results obtained for the different criteria based on the chosen loss function and threshold

Loss Function	WBCE	WBCE	BCE	BCE
Decision Threshold	0.5	Youden (0.35)	0.5	Youden (0.35)
Criteria				
AUC	0.946	0.946	0.949	0.949
Sensitivity	0.881	0.947	0.637	0.910
Specificity	0.842	0.802	0.956	0.856

WBCE: Weighted Binary Cross-Entropy

BCE: Binary cross-entropy

AUC: Area Under the Curve, area under the ROC curve

Youden: optimal decision threshold as defined by Youden

The Youden index is a metric derived from the ROC curve (Figure 2 and Figure 3) used to determine the optimal decision threshold for a binary classification model[11]. It is defined as the sum of sensitivity and specificity, minus 1: $J = \text{sensitivity} + \text{specificity} - 1$, which is equivalent to maximizing the difference between the true positive rate and the false positive rate. It ranges from 0 to 1: a value of 0 indicates performance equivalent to a random classifier, while a value of 1 corresponds to perfect classification. The optimal threshold is the one that maximizes this index on the ROC curve, thus offering the best trade-off between sensitivity and specificity.

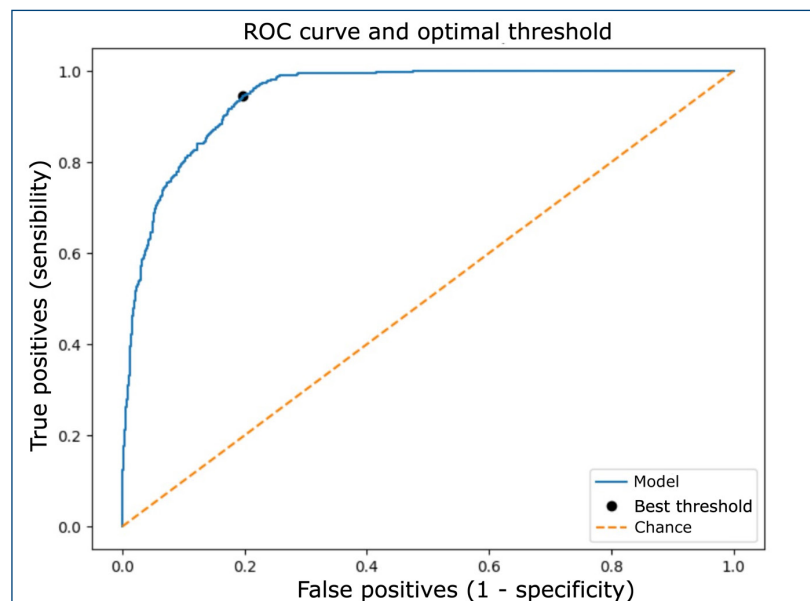


Figure 2. ROC curve for the model using ECBP

The point on the curve above indicates the true positive and false positive rates for the optimal Youden threshold.

The ROC (*Receiver Operating Characteristic*) curve (*Figure 2 and Figure 3*) is constructed by plotting, for each possible decision threshold, the true positive rate (sensitivity) on the y-axis against the false positive rate ($1 - \text{specificity}$) on the x-axis. The thresholds tested correspond to the set of probabilities of requiring hemodialysis returned by the model. A diagonal line with the equation $y = x$, passing through the origin, is drawn to represent the performance of a random classifier, for which the true positive rate is equal to the false positive rate regardless of the threshold considered. An ROC curve lying below this diagonal would indicate performance worse than random, as the model would then produce more false positives than true positives for the thresholds in question. The goal is therefore to obtain a ROC curve as far as possible from this diagonal, toward the upper-left corner of the graph.

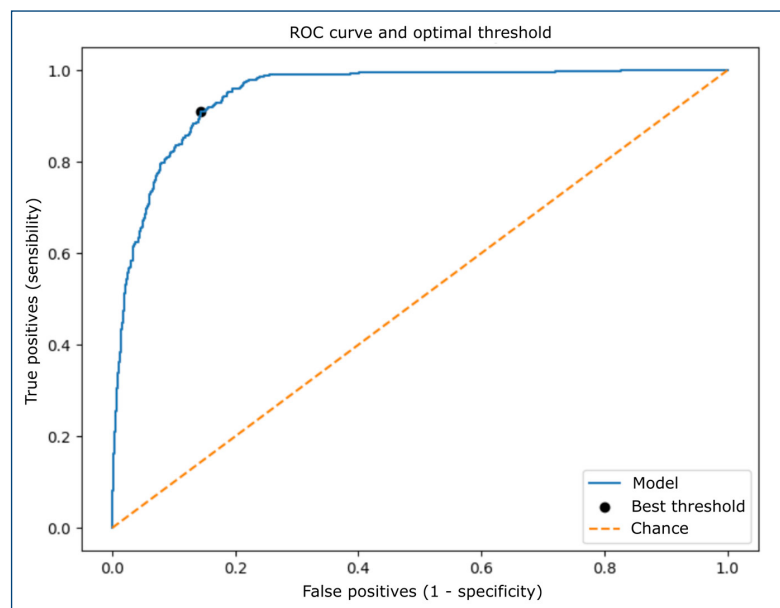


Figure 3. ROC curve for the model with ECB

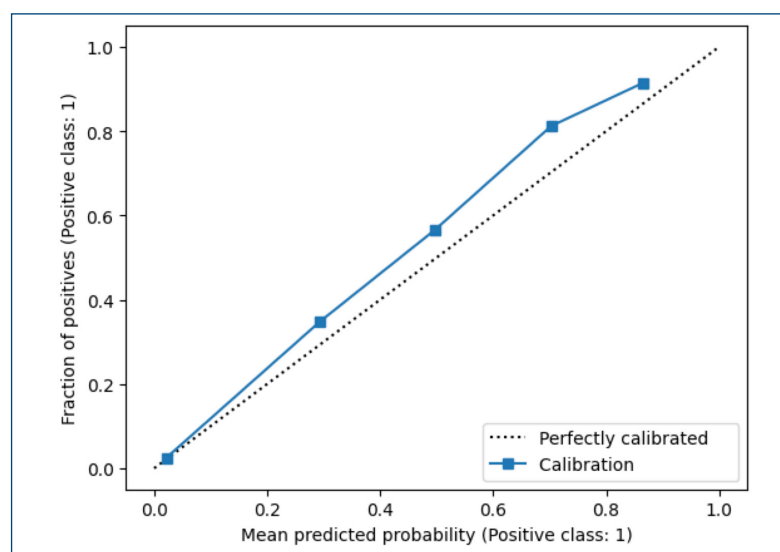


Figure 4. Calibration curve for the model with ECB

The calibration curve (*Figure 4 and Figure 5*) is a graphical representation that allows us to compare the probabilities predicted by the model with the actual observed frequencies of

hemodialysis in the test set. The predictions are divided into five quantiles, and for each quantile, the average predicted probability is plotted against the corresponding observed frequency. The diagonal line $y = x$ represents perfect calibration, where the predicted probabilities coincide exactly with the observed frequencies. A deviation toward the lower right of the graph indicates that the model overestimates the risk of starting hemodialysis, while a deviation toward the upper left indicates an underestimation of this risk.

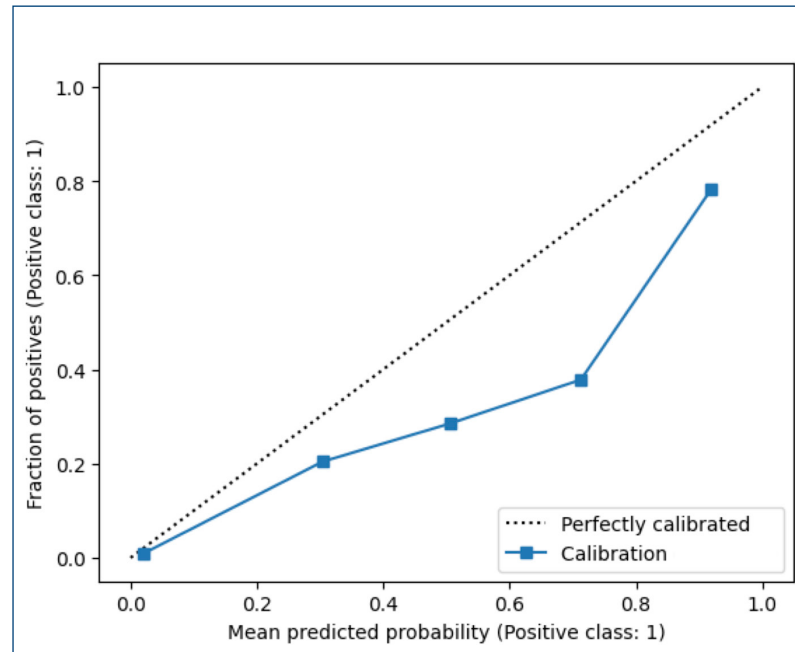


Figure 5. Calibration curve for the model using ECBP

Discussion

The onset of peritonitis during peritoneal dialysis carries a significant risk of transfer to hemodialysis, but this risk depends on both patient characteristics and the nature of the infection. This study demonstrated that, after training a neural network, it is possible to predict the need for transfer to hemodialysis following a peritoneal infection with sufficient reliability to make its clinical application feasible and to aid in medical decision-making.

One of the priorities during model training was to limit any risk of overfitting, which would impair the model's ability to generalize its predictions. We therefore used *dropout and early stopping* techniques to improve generalization and L2 regularization to limit the model's sensitivity to noise (*Table III*).

The choice of the WBCE loss function is justified by the desire to reduce the number of false negatives at the expense of a higher number of false positives. According to *Table IV*, sensitivity increases when WBCE is used instead of BCE, regardless of the threshold chosen. In contrast to sensitivity, specificity decreases when WBCE is used (*Table IV*).

During training, the selected decision threshold was the Youden optimal threshold when it was less than 0.5 and 0.5 otherwise. Using a threshold below 0.5 amounts to considering that a moderate predicted probability is sufficient to trigger a prediction of transition to HD, thereby making the model more sensitive to small variations in probability. This choice is consistent with the previously stated objective of minimizing false negatives: by lowering the decision threshold, the model prioritizes the detection of positive cases, again at the expense of a higher false-positive rate (a false-positive rate of 15.8% with a threshold of 0.5 versus 19.8% with the Youden threshold, for WBCE; *Table IV*).

The performance of a binary classification model can be evaluated along two complementary dimensions: discrimination and calibration[12,13] .

The AUC is a model selection criterion[7,14] . The higher the AUC, the further the ROC curve (*Figure 2 and Figure 3*) deviates from the diagonal line of chance, indicating better discriminative performance of the model. Formally, the AUC can be interpreted as the probability that the predicted score for a positive example is higher than that predicted for a negative example[15] . Discrimination measures the model's ability to effectively distinguish between the two classes. In our case, the AUC reaches 0.946 and 0.949 (depending on the configuration; *Table IV*), values close to 1, indicating excellent discriminatory power.

Calibration, on the other hand, assesses the alignment between the probabilities predicted by the model and the frequencies observed in the actual data[16] . The curve associated with the BCE configuration (*Figure 4*) shifts toward the upper left, indicating that the model underestimates the risk of transition to HD, which results in a high false-negative rate. Conversely, the curve associated with the WBCE configuration (*Figure 5*) shifts toward the lower right, reflecting an overestimation of risk, which manifests as a higher false-positive rate. This deviation indicates that the model overestimates the probabilities of transition to HD compared to actual observations. In other words, the model tends to be overly confident in its positive predictions[17] , or to favor false-positive errors (patients incorrectly classified as requiring HD) over false-negative errors (patients requiring hemodialysis but not identified). However, this bias helps minimize the false-negative rate (11.9% with WBCE versus 36.3%, when the decision threshold is set at 0.5; *Table IV*), a priority in our clinical context, where failing to identify patients requiring hemodialysis is more critical than erroneously including patients who do not require it.

Limitations of this model become apparent through the partitioning of the data. In particular, patients are randomly assigned—regardless of their center of origin—across the three datasets used. This means that the model can learn the specific practices of each center regarding PD treatment and the transition of patients to HD. This was a deliberate choice, as the predictions also adapt to the standard practices of the various centers.

Conclusion

Artificial intelligence and machine learning techniques have proven effective in predicting the transition to hemodialysis (HD) for patients on peritoneal dialysis following one or more episodes of peritonitis. The model's training strategy aimed to minimize the false-negative rate. Nevertheless, a balance had to be struck between a confident prediction by the model and an accurate representation of reality through that prediction.

This approach paves the way for the development of an application—to be tested through external validation—to determine whether it can reduce the rate of unscheduled transfers and the risk of having use a central catheter.

Authors' Contributions

All authors contributed to the study design.

AW: writing, model development, analysis of results. RL: writing, model development, analysis of results.

YT: advice on model development, analysis of results, and writing.

Ethical Considerations and Patient Consent

This study was conducted in accordance with the ethical principles of the Declaration of Helsinki. Informed consent was not required for this study because it involved retrospective data from a registry for which patients had previously provided authorization for inclusion; patient and institutional data were fully anonymized and analyzed retrospectively. Patient and center identifiers were anonymized in advance using a randomly generated hash key.

Funding

No specific funding was received for this work.

Acknowledgments

The authors thank Dr. Christian Verger (RDPLF) and Dr. Jacques Chanliau (RDPLF) for their assistance throughout this study, and Dr. Emmanuel Fabre for the anonymized export of data from the RDPLF database.

Conflicts of Interest

The authors declare that they have no conflicts of interest related to this work.

Data availability

The datasets generated and analyzed in this study are available from the authors and the RDPLF upon reasonable request.

Statement on Artificial Intelligence

The authors declare that this manuscript is the result of their own original work. No artificial intelligence tools or applications were used other than those designed by the authors themselves—which were the subject of this study—for data analysis and the generation of results.

ORCID iDs

Alexander Watson: <https://orcid.org/0009-0008-5143-7882>

Rémy Lozano: <https://orcid.org/0009-0005-7138-136X>

Yannick Toussaint: <https://orcid.org/0009-0005-0507-7204>

References

1. Verger C, Fabre E, Veniez G, Padernoz MC. Synthetic 2018 data report of the French Language Peritoneal Dialysis and Home Hemodialysis Registry (RDPLF) Bull Dial Domic Internet. 2019 Apr 10;2(1):1–10.
<https://doi.org/10.25796/bdd.v2i1.19093>

2. Cheikh Hassan HI, Murali KM, Chen JHC, Mullan J. Back-up arteriovenous fistula in peritoneal dialysis patients: a retrospective cohort study of long-term meaningful outcomes. *Intern Med J.* 2026 May 23;imj.70493. doi: <https://doi.org/10.1111/imj.70493> PubMed PMID: 42175666.
3. Raskin D, Vachharajani TJ, Partovi S, Khan A, Lyden SP, Kirksey L. Value-Based Model for Vascular Access Management in the End-Stage Kidney Disease Population. *Semin Dial.* 2026 Jan;39(1–2):15–25. doi: <https://doi.org/10.1111/sdi.70024> PubMed PMID: 41957978; PubMed Central PMCID: PMC13102636.
4. Hsu FY, Hwang RH, Tsai MH, Wang JT. Predicting Peritoneal Dialysis Failure Within the Next Three Months Based on Deep Learning and Important Features Analysis. *Information.* 2024 Dec 5;15(12):776. doi: <https://doi.org/10.3390/info15120776>
5. Tangri N, Ansell D, Naimark D. Predicting technique survival in peritoneal dialysis patients: comparing artificial neural networks and logistic regression. *Nephrol Dial Transplant.* 2008 Apr 3;23(9):2972–81. doi: <https://doi.org/10.1093/ndt/gfn187> PubMed PMID: 18441002; PubMed Central PMCID: PMC2517147.
6. Bai Q, Tang W. Artificial intelligence in peritoneal dialysis: general overview. *Ren Fail.* 2022 Dec 31;44(1):682–7. doi: <https://doi.org/10.1080/0886022X.2022.2064304> PubMed PMID: 35469549; PubMed Central PMCID: PMC9045776.
7. Ioannou GN, Tang W, Beste LA, Tincopa MA, Su GL, Van T, et al. Assessment of a Deep Learning Model to Predict Hepatocellular Carcinoma in Patients With Hepatitis C Cirrhosis. *JAMA Netw Open.* 2020 Sep 1;3(9):e2015626. doi: <https://doi.org/10.1001/jamanetworkopen.2020.15626> PubMed PMID: 32870314; PubMed Central PMCID: PMC7489819.
8. Sugino T, Kawase T, Onogi S, Kin T, Saito N, Nakajima Y. Loss Weightings for Improving Imbalanced Brain Structure Segmentation Using Fully Convolutional Networks. *Healthcare.* 2021 Jul 26;9(8):938. doi: <https://doi.org/10.3390/healthcare9080938> PubMed PMID: 34442075; PubMed Central PMCID: PMC8393549.
9. Yeung M, Sala E, Schönlieb CB, Rundo L. Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Comput Med Imaging Graph.* 2022 Jan;95:102026. doi: <https://doi.org/10.1016/j.compmedimag.2021.102026> PubMed PMID: 34953431; PubMed Central PMCID: PMC8785124.
10. Fang J, Xu Y, Zhao Y, Yan Y, Liu J, Liu J. Weighing features of lung and heart regions for thoracic disease classification. *BMC Med Imaging.* 2021 Dec;21(1):99. doi: <https://doi.org/10.1186/s12880-021-00627-y> PubMed PMID: 34112095; PubMed Central PMCID: PMC8194196.
11. Ruopp MD, Perkins NJ, Whitcomb BW, Schisterman EF. Youden Index and Optimal Cut-Point Estimated from Observations Affected by a Lower Limit of Detection. *Biom J.* 2008 Jun;50(3):419–30. doi: <https://doi.org/10.1002/bimj.200710415>
12. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the Performance of Prediction Models: A Framework for Traditional and Novel Measures. *Epidemiology.* 2010 Jan;21(1):128–38. doi: <https://doi.org/10.1097/EDE.0b013e3181c30fb2>
13. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat Med.* 2014 Feb 10;33(3):517–35. doi: <https://doi.org/10.1002/sim.5941>
14. Whitney HM, Drukker K, Giger ML. Performance metric curve analysis framework to assess impact of the decision variable threshold, disease prevalence, and dataset variability in two-class classification. *J Med Imaging.* 2022 May 31;9(03). doi: <https://doi.org/10.1117/1.JMI.9.3.035502>
15. Li J. Area under the ROC Curve has the most consistent evaluation for binary classification. Qin H, editor. *PLOS ONE.* 2024 Dec 23;19(12):e0316019. doi: <https://doi.org/10.1371/journal.pone.0316019>
16. Friesacher HR, Engkvist O, Mervin L, Moreau Y, Arany A. Achieving well-informed decision-making in drug discovery: a comprehensive calibration study using neural network-based structure-activity models. *J Cheminformatics.* 2025 Mar 5;17(1):29. doi: <https://doi.org/10.1186/s13321-025-00964-y>

17. Zhang F, Dvornek N, Yang J, Chapiro J, Duncan J. Layer Embedding Analysis in Convolutional Neural Networks for Improved Probability Calibration and Classification. IEEE Trans Med Imaging. 2020 Nov;39(11):3331–42. doi: <https://doi.org/10.1109/TMI.2020.2990625>