

An artificial neural network-based model to predict chronic kidney disease in aged cats

Vincent Biourge¹ | Sebastien Delmotte² | Alexandre Feugier¹ |
Richard Bradley³ | Molly McAllister⁴ | Jonathan Elliott⁵

¹Royal Canin, Research Center, Aimargues, France

²Mad-environnement, Nailloux, France

³Waltham Pet Science Institute, Waltham on the Wolds, Leicestershire, United Kingdom

⁴Banfield Pet Hospitals, Vancouver, Washington

⁵The Royal Veterinary College, London, United Kingdom

Correspondence

Vincent Biourge, Royal Canin, CS10309,
30470 Aimargues, France.
Email: vincent.biourge@royalcanin.com

Funding information

Royal Canin SAS, Subsidiary of Mars, Inc.

Abstract

Background: Chronic kidney disease (CKD) frequently causes death in older cats; its early detection is challenging.

Objectives: To build a sensitive and specific model for early prediction of CKD in cats using artificial neural network (ANN) techniques applied to routine health screening data.

Animals: Data from 218 healthy cats ≥ 7 years of age screened at the Royal Veterinary College (RVC) were used for model building. Performance was tested using data from 3546 cats in the Banfield Pet Hospital records and an additional 60 RCV cats—all initially without a CKD diagnosis.

Methods: Artificial neural network (ANN) modeling used a multilayer feed-forward neural network incorporating a back-propagation algorithm. Clinical variables from single cat visits were selected using factorial discriminant analysis. Independent sub-models were built for different prediction time frames. Two decision threshold strategies were investigated.

Results: Input variables retained were plasma creatinine and blood urea concentrations, and urine specific gravity. For prediction of CKD within 12 months, the model had accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of 88%, 87%, 70%, 53%, and 92%, respectively. An alternative decision threshold increased specificity and PPV to 98% and 87%, but decreased sensitivity and NPV to 42% and 79%, respectively.

Conclusions and Clinical Importance: A model was generated that identified cats in the general population ≥ 7 years of age that are at risk of developing CKD within

Abbreviations: AI, artificial intelligence; ANN, artificial neural network; AUC, area under the curve; BUN, blood urea nitrogen; CKD, chronic kidney disease; FDA, factorial discriminant analysis; FGF23, fibroblast growth factor 23; FPR, false positive rate; GFR, glomerular filtration rate; IRIS, International Renal Interest Society; MLP, multilayers perception; NPV, negative predictive value; PPV, positive predictive value; PTH, parathyroid hormone; ROC, receiver operating curve; RVC, Royal Veterinary College; SDMA, symmetric dimethylarginine; USG, urine specific gravity.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2020 The Authors. *Journal of Veterinary Internal Medicine* published by Wiley Periodicals LLC on behalf of American College of Veterinary Internal Medicine.

12 months. These individuals can be recommended for further investigation and monitoring more frequently than annually. Predictions were based on single visits using common clinical variables.

KEYWORDS

artificial intelligence, CKD modeling, prediction tool, prevention, senior health check

1 | INTRODUCTION

Chronic kidney disease (CKD) is a progressive, heterogeneous syndrome afflicting older cats. It is characterized by persistent azotemia (assessed most commonly by plasma creatinine concentration) in association with decreased urinary concentrating ability.¹ Chronic kidney disease results from loss of functioning nephrons, leading to a decrease in total glomerular filtration rate (GFR).² Nephron loss can result from multiple continuous or intermittent primary disease processes²; which trigger inflammation and repair processes leading to interstitial fibrosis,³ that is closely linked to decline in kidney function.⁴ Adaptation of the remaining functioning nephrons leads to their hypertrophy, accompanied by glomerular capillary hypertension and hyperfiltration.⁵ Such intrinsic nephron damage may lead to further loss of functioning nephrons in the absence of any primary disease process.⁶

Chronic kidney disease is a frequent cause of death in cats >5 years of age,⁷ and is a reason why routine annual health screening assessing kidney function should be common practice for senior cats.⁸ However, early detection of decreased functioning renal mass is challenging because the relationship between GFR and surrogate markers of GFR is exponential; when GFR decreases from normal, the change in plasma concentration of the surrogate marker is small.⁹ Second, the adaptation of remaining functioning nephrons tends to limit the decrease in total kidney GFR.¹⁰

Techniques dependent on artificial intelligence (AI) can be efficient in predicting a large variety of outcomes, particularly in the field of medicine, and have capabilities offering advantages over traditional statistical methodologies of multivariate regression.¹¹ Artificial neural networks (ANN) in particular have great capacity for prediction of complex and nonlinear processes coupled with ease and flexibility of implementation.^{12,13}

Our objective was to build a sensitive and specific model for early prediction of CKD in cats using AI mathematical tools and 2 populations of cats; 1 small but well characterized and a second large but less well-defined.

2 | METHODS

2.1 | Datasets sourced from clinical studies

Data for the development of the diagnostic algorithm were extracted from the database of the Feline Kidney Research Clinic at the Royal Veterinary College (RVC database), which contains the data of cats recruited to studies conducted there since 1995. These were cats that

owners had believed to be healthy and had been recruited to be screened for kidney disease and followed longitudinally at 2 clinics in central London (UK), the Beaumont Sainsbury Animals' Hospital (BSAH), Royal Veterinary College, Camden, and the People's Dispensary for Sick Animals (PDSA), Bow.

A first dataset (RVC1) was extracted from the RVC database from cats added between 1995 and November 2013 and used for prediction algorithm development. A second dataset (RVC2) later was extracted for validation purposes and consisted of (1) healthy cats added to the database between November 2013 and April 2016 and (2) cats that were in the database before November 2013 but were not eligible for RVC1 because they only reached 18 months of follow-up after November 2013. Some of the cats in the RVC1 and RVC2 groups contributed to other studies reported by the RVC.^{4,10,14-19}

To be included in either the RVC1 or RVC2 dataset, cats had to be ≥ 7 years of age and considered healthy based on history, physical examination, blood biochemistry and urine screenings. Where available, blood pressure measurements (standardized Doppler method, Parks Electronic Doppler Model 811B, Perimed UK, Bury St Edmunds, UK)²⁰ were included in algorithm development but were not used to determine health status. These datasets encompassed a large set of numerical and nonnumerical variables per time point that described the cat's environment,²¹ signalment, clinical examination findings, vaccination history, packed cell volume, plasma biochemistry, urinalysis, and biomarkers such as parathyroid hormone (PTH) and fibroblast growth factor 23 (FGF-23).

For cats that remained nonazotemic to be included in the datasets, at least 1 follow-up visit had to have occurred ≥ 540 days (18 months) after the initial screen at which the cat was examined, and blood and urine samples must have been collected to assess renal function. The datasets also included data from any follow-up visits between the initial screen and 540 days. Similarly, the dataset contained all data from multiple visits by cats documented as azotemic (plasma creatinine concentration ≥ 2 mg/dL) after the initial screen.

All cats were categorized as either becoming azotemic after the initial health screen (cases) or remaining nonazotemic over at least 18 months (controls). A single group of veterinary-qualified staff from the RVC reviewed the records from baseline to agree on the initial classification of cats as healthy and from follow-up visits to verify the diagnosis of CKD.

2.2 | Dataset from a large group of primary care clinics

To test the model performance on data collected by primary care practitioners in a clinical service setting, a subset of domestic short,

medium and long-haired cats visiting Banfield Pet Hospitals (Vancouver, Washington) between January 1995 and June 2016 was identified. Data extracted from 6.5 million electronic medical records included reproductive status, breed, age and laboratory results at each visit. From those records, 24 497 CKD cats were identified by either a formal recorded diagnosis of CKD or by at least 2 CKD-suggestive data points from the following list: plasma creatinine concentration above normal, urine specific gravity (USG) below normal, and “CKD,” “azotemic,” “Royal Canin Veterinary diet Renal,” or “Hill’s prescription diet k/d” in the medical notes. A first set of filters was applied to retain CKD cats presented in a Banfield Hospital as healthy before the onset of CKD and for which there were at least 2 visits with a measurement of plasma creatinine concentration as well as to remove outliers, yielding 1855 CKD cats. A second set of filtering criteria then was applied to ensure that data were suitable for modeling, including age between 7 and 22 years, and at least 2 visits at which plasma creatinine and blood urea nitrogen (BUN) concentration and USG were measured, yielding 1510 CKD cats.

For this Banfield dataset, cases were defined by a clinical diagnosis of CKD made using the equipment, training and guidelines in place for each veterinarian at the time of diagnosis. No allowance was made for the alteration of diagnosis at a subsequent visit, contrary medical notes, or the possibility of a different diagnosis on the basis of subsequent clinical practices or guidelines, such as the International Renal Interest Society (IRIS) guidelines.

Controls were cats that had not been diagnosed with CKD and had a further 2 years of visits beyond the last data point provided to the model, during which they remained free of CKD. They were not specifically healthy cats and could have been diagnosed with any other illness. The medical notes were searched electronically to remove any control cats that seemed likely to have received a tentative diagnosis of CKD that was not formally recorded.

2.3 | Clinical assays

2.3.1 | RVC1 and RVC2 datasets

Owners were asked to withhold food for 8 hours before blood samples by jugular venipuncture were collected into lithium heparin and plain tubes. Heparinized plasma was used for immediate biochemical

analysis (Idexx Laboratories, Wetherby, West Yorkshire, UK). A urine sample collected by cystocentesis was required for study eligibility. Urinalysis included measurement of specific gravity, pH (HI 9224 pH meter, Hanna Instruments, Leighton Buzzard, UK), dipstick chemistry analysis (Multistix Urine Chemistry Reagent Strips, Bayer Diagnostics, Newbury, Berks, UK) and microscopic examination of sediment. Residual samples were centrifuged (Mistral 3000, Sanyo-Gallenkamp, Leics, UK) at 4°C for 10 minutes, and stored at –80°C for later batched measurement of urine protein-to-creatinine ratio and urine albumin-to-creatinine ratio. Serum PTH and FGF-23 measurements were not available for enough cats for model building.

2.3.2 | Banfield dataset

The clinical assays used for the Banfield database were dictated by the procedures of the Banfield Pet Hospitals at the time of the cat visits.

2.4 | Model development

Modeling was conducted in 3 phases. Phase 1 (2014) was the initial development of the model and its predictive algorithms using the RVC1 dataset. During this phase, the model was designed, the input variables were selected and a first run of training and validation of the model was undertaken. The performance of the initial model was tested in phase 2 (2016) using the Banfield and RVC2 datasets as new, independent data. Phase 3 (2016) consisted of a new cycle of training and validation keeping the same input factors (creatinine, BUN, and USG) but reinitializing and optimizing the model using all 3 datasets (RVC1, RVC2, and Banfield). Table 1 provides a summary of modeling terminology.

2.5 | Model design and dataset building

All cats had at least 2 visits with known outcomes of negative or positive CKD status, these being the initial visit at which all enrolled cats were required to have a negative CKD status, and at least 1 follow-up

TABLE 1 Terminology

Model	A global algorithm developed with artificial intelligence tools that allowed the outcome of CKD to be predicted based on a limited set of data
Submodel	A model that was developed to predict occurrence of CKD within each time period studied (ie, 0, 3, 6, 9, 12, or more than 12 months +). All the submodels together form the final model
Model building	Development of the model based on a subset of individual-level data that encompassed a range of clinical and laboratory parameters in cats with known CKD outcomes
Model validation	The process by which the accuracy of the model in correctly predicting CKD status was tested using a different subset of data that is, independent data from different cats to those used for the model building
Phase	Each phase of the model building represented the application of a given dataset or combination of datasets to either build or validate the model

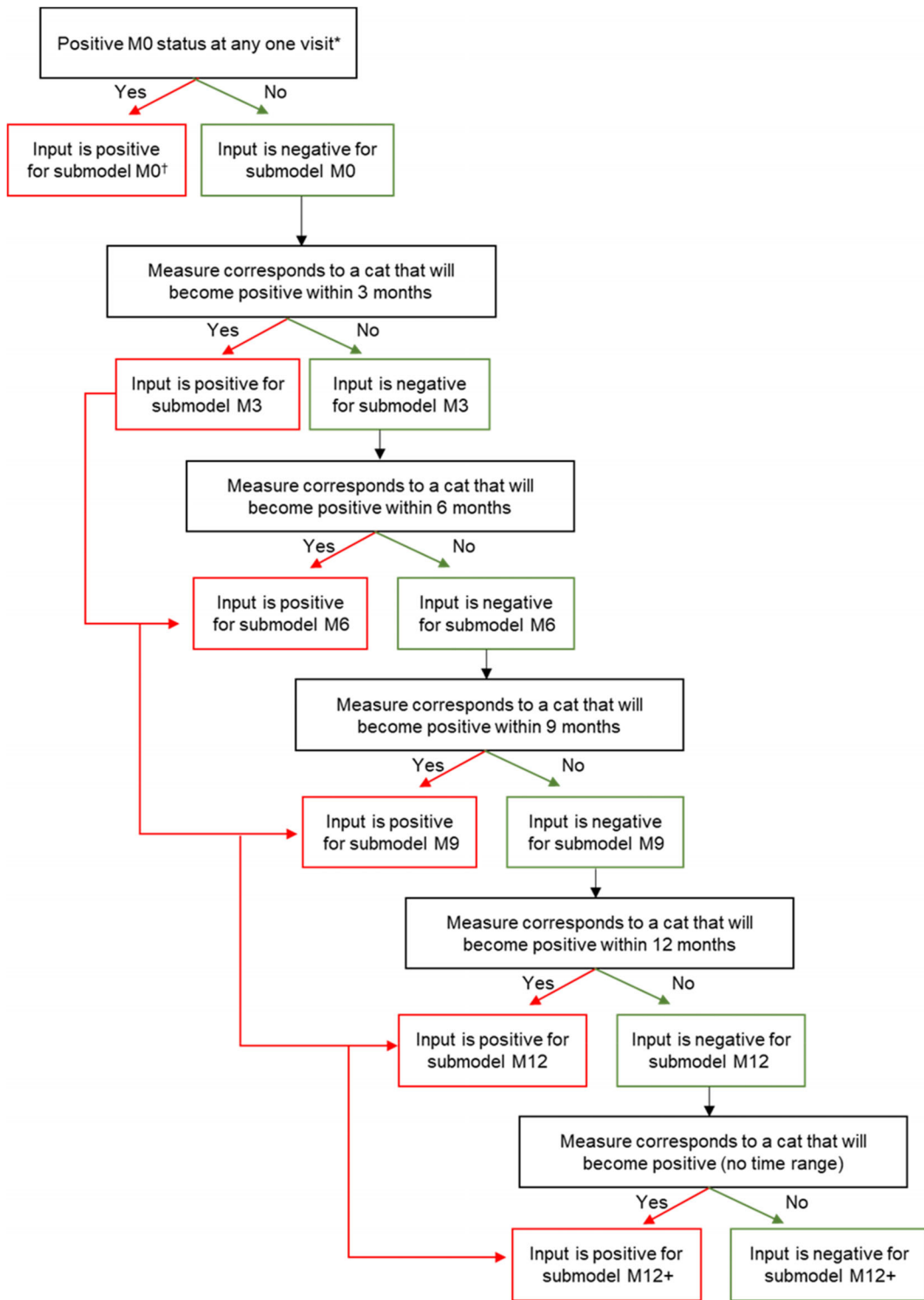


FIGURE 1 Design of datasets for time-delineated submodels. The chart shows the flow of data into time-delineated submodels with respect to a single visit. A cat had a separate M0 status for each of its visit, which were analyzed independently. A given visit was associated with a chronic kidney disease (CKD) status (positive or negative) for each of the submodels M0, M3, M6, M12 or M12+, depending on the time interval in months between the visit and the date when the cat was diagnosed or not with CKD. Cats diagnosed with CKD at or before the date of any visit were included as positive for the associated M0 submodel but excluded from other submodels, because unlike M0 these were designed to predict a future not a present status. *M0 status at the initial visit was negative for all cats because this was an eligibility criterion. †Once a cat was positive at any of the visits it was automatically recorded as positive for all subsequent visits and was no longer followed up that is, it no longer provided input for submodels other than M0

visit. For each visit, a dataset was built a posteriori by coupling disease status at the time of the visit, the set of variables measured from the samples taken, and the visit date. Each dataset was used as input to create 6 submodels that could use measures from a single visit to predict the current (submodel M0) or future occurrence of CKD within 3, 6, 9, 12, or >12 months (submodels M0, M3, M6, M9, M12, and M12+, respectively; Figure 1). The CKD status of a cat at the time of a visit was termed the M0 status. Every visit was considered separately and had an independent M0 status in order to establish a prediction tool based on a single visit. The input data used to build the submodels therefore were visits and not the cat's history. Once a cat became CKD positive at any visit, it was automatically entered into the model with a positive status for all subsequent visits, even if any of the following submodels predicted a negative output (Figure 1).

2.5.1 | Initial selection of variables

The raw dataset was cleaned from 116 to 16 variables by excluding 55 with >40% missing data, all remaining 36 qualitative variables because of their minimal relevance to CKD or paucity of abnormal findings, and 9 blood variables that were not consistently present with the same other variables to constitute complete sets of measures. Variables remaining were: age, creatinine, USG, chloride, total plasma protein, phosphate, total plasma calcium, albumin, globulin, urea, alanine transaminase, alkaline phosphatase, bilirubin, cholesterol, sodium, and potassium. Factor discriminant analysis (FDA) with a threshold fixed to $|0.5|$ was used for variable selection from these 16 variables.^{22,23} It was applied separately on each submodel dataset to ensure that the best predictors for each time range were retained.

2.5.2 | Artificial neural network modeling

The ANN modeling was undertaken using a multilayer feed-forward neural network, so-called multilayers perceptron (MLP), incorporating a back-propagation algorithm.^{24,25} In the first step of the model building, a set of input/output vector pairs was presented to the network for the "training" process. For each input vector, the neural network model calculated an output vector, and by comparing this output vector with the actual output vector, an error term for the outputs of all hidden and output neurons was derived. The weights and biases were updated using this error term, and the procedure was repeated in order to minimize the error. A 10-fold cross-validation approach was used with 5 repetitions for each submodel, and the 20 best ANNs were selected each time to form an "ensemble model." During the training step, the internal parameters of the MLP were tuned simultaneously in a full factorial design that tested all combinations for each submodel.

These parameters were the number of hidden layers, the number of neurons in the hidden layer and the decay. There was 1 hidden layer, neuron numbers were tested from 2 to 30, and decay was set to vary from 0.001 to 0.1. Receiver operator curves (ROCs) were

generated for the validation datasets by plotting the true positive rate, or sensitivity, against the false positive rate (equal to 1-specificity) across various thresholds. Optimal parameter values were selected based on the area under the curve (AUC) of the ROCs as the measure of model accuracy.

Because the prediction was based on a single visit, the final calculation applied the best submodels on all the initial data based on one visit from each cat. The process was repeated using the different visits for each cat, and the average values for each variable (sensitivity, specificity, PPV, and NPV) are reported.

2.6 | Model performance

Performance of the model in each phase was based on the repeated application of the submodels on 1 selected visit by a cat. For each submodel, the percentage of correct to incorrect answers was calculated and the algorithm optimized the model to the lowest rate of errors.

Two strategies were used to set decision thresholds for determining the predicted state of a cat from the probability output of the model. Strategy 1 was to select the optimal threshold for ensuring both high specificity and sensitivity using the Youden's index.²⁶ Strategy 2 was to define the threshold that would ensure a high specificity corresponding to the highest positive predictive value (PPV), with the trade-off of a potentially low sensitivity. This approach aimed to decrease false positives at the risk of potentially increasing false negatives.

The NPV and PPV were dependent upon CKD prevalence, and differed between submodels because they were based upon visits not cats. The NPVs and PPVs also were calculated for the CKD prevalence normalized to 10%, 15%, 20%, and 30% for all submodels.

The independence of forward predictive submodels from each other meant it was possible to have a positive prediction within 1 time frame followed by a negative prediction within any later time frame. Such disagreements between submodels are termed incoherences. The rate of incoherence between submodels was calculated for the final model (ie, the proportion of cases predicted by any of the submodels that were not confirmed by ≥ 1 of the following submodels).

2.7 | Modeling software

All calculations were carried using R software (v 3.3.3, open-source software; <https://cran.r-project.org/bin/windows/base/old/3.3.3/>) with the MASS (Discriminant analysis, <https://cran.r-project.org/package=MASS>), ADE4 (Multivariate analysis (FDA), <https://cran.r-project.org/package=ade4>), CARET (Data modeling, <https://cran.r-project.org/package=caret>), pROC (ROC curves, <https://cran.r-project.org/package=pROC>), and NNET (ANN, <https://cran.r-project.org/web/packages/nnet/index.html>) libraries dedicated to modeling and machine learning.

2.8 | Study conduct

The Ethics and Welfare Committee of the RVC and the Royal Canin ethics committee approved the clinic protocols for screening healthy cats for CKD in the RVC datasets. Samples were collected and stored with the informed consent of the cats' owners.

3 | RESULTS

3.1 | Composition of datasets

Data from 218 RVC cats (RVC1 dataset) were used to design and build the initial model in phase 1, and data from 60 RVC2 cats and 3486 Banfield cats were used to validate the model in phase 2 as an independent dataset (Table 2). Visits were excluded due to missing data (701 from RCV1 dataset) and for cats <7 years of age (31 from RCV1; 9835 from Banfield), leaving a total of 10 576 visits for analysis—672, 60 and 9844 from RCV1, RCV2, and Banfield datasets, respectively (Tables 2 and 3). The baseline characteristics of cats are summarized in Table 4. Most cats were neutered, the overall proportion of female cats ranged from 51% to 56%, mean age ranged from 11.1 years in the Banfield data set to 13.2 years in the RCV1 dataset.

Over the study period, 76% of cats in the RVC1 dataset and 57% of Banfield cats remained nonazotemic, corresponding to an overall prevalence of visits with a positive CKD for the model building of 24% and 43%, respectively. Of the cats included in phase 1 of the modeling (RVC1 dataset), 11.5% showed azotemia within 6 months and 15% at 12 months. In phase 3 of the modeling, in which the same model design was used to regenerate the model by fresh training and validation using a combined dataset of RVC1 plus RCV2 plus Banfield, 11.4% of included cats showed azotemia within 6 months and 24% at 12 months.

Table 3 shows the total number of visits made by cats with and without CKD that contributed to the modeling (negative controls and positive cases, respectively). Table 3 also shows the number of visits

with positive or negative status that formed the input for each submodel. The prevalence of CKD for each submodel is presented in Supporting Information Table S1.

3.2 | Input variables selection

For each of the time-delineated submodels (M0-M12+) the same 3 variables had a discriminant weight $\geq |0.5|$ on FDA: creatinine, USG and urea (Figure 2); these were retained as input variables for the predictive model. Using the final dataset for phase 3 (RVC1 plus Banfield plus RVC2), the correlation between creatinine and USG as well as the correlation between urea and urine SG varied between -0.2 and -0.3 (Figure 2). The correlation between creatinine and urea was approximately 0.54. These correlations were low enough to exclude potential modeling problems associated with covariation of predictors.

3.3 | Model performances

The performance of the model in each phase of development is summarized in Tables 5 and 6. The ROCs illustrating the accuracy, sensitivity and specificity of submodel M12 are shown in Figure 3.

When the initial model from phase I (RVC1 dataset) was established, specificity and sensitivity were optimized (strategy 1). The submodels had specificities ranging from 0.72 to 0.93 and sensitivities from 0.83 to 0.93 (Table 5). The negative predictive values (NPVs) were close to 1 and the PPVs were approximately 70%. A second strategy was tested optimizing specificity and, in doing so PPV, but at the cost of decreasing sensitivity and NPV (Table 6).

In phase 2, applying the Banfield and RCV2 datasets to the initial model was associated with high accuracy and sensitivity but a decrease in specificity and PPV for both strategies but especially strategy 1, which had an accuracy of 0.82-0.89, sensitivity of 0.93-0.97, specificity of 0.23-0.65 and PPV of 0.27-0.62 (range across submodels, Table 5).

TABLE 2 Numbers of cats and visits used to create and validate the model

Model phase/dataset	Cat status	Cats	Visits
Phase 1/RVC1	Controls, n	166	447
	Cases, n	52	225
	Prevalence of CKD	24%	33%
Phase 2/Banfield + RVC2	Controls, n	2025	4981
	Cases, n	1521	4923
	Prevalence of CKD	43%	50%
Phase 3/Banfield + RVC1 + RCV2	Controls, n	2191	5428
	Cases, n	1573	5148
	Prevalence of CKD	42%	49%

Notes: Controls were cats remaining free from chronic kidney disease for 18 months after the initial qualifying visit. Cases were cats that were diagnosed with chronic kidney disease at visits after the qualifying visit.

Abbreviations: CKD, chronic kidney disease; RVC, Royal Veterinary College.

TABLE 3 Number of control (negative) and case (positive) cat visits used to build and test (phase 1 and 3) and test (phase 2) the submodels

Submodel	Phase 1		Phase 2		Phase 3	
	Negative control	Positive case	Negative control	Positive case	Negative control	Positive case
M0	521	151	8314	890	8835	1041
M3	514	7	8109	205	8623	212
M6	490	31	7899	415	8389	446
M9	471	50	7598	716	8069	766
M12	456	65	7350	964	7806	1029
M12+	447	74	4281	4033	4728	4107

Notes: Submodel M0 was based on each clinic visit assessing the status of a cat with respect to a diagnosis of chronic kidney disease (CKD). Submodels M3-M12+ were for the prediction of the CKD status of negative cats in time frames out to 3, 6, 9, 12 months, and more than 12 months after M0.

TABLE 4 Baseline characteristics of cats in the datasets used for modeling

	RVC1	RVC2	Banfield
Total number of cats, n	218	60	3486
Intact males, n (%)	2 (0.92)	1 (1.67)	4 (0.11)
Neutered males, n (%)	94 (43.1)	28 (46.7)	1705 (48.9)
Intact females, n (%)	1 (0.46)	0	16 (0.11)
Spayed females, n (%)	121 (55.5)	31 (51.7)	1761 (50.5)
Domestic short-hair, n (%)	169 (77.5)	45 (75.6)	NR
Domestic long-hair, n (%)	140 (6.4)	11 (17.6)	NR
Pure breed, n (%)	35 (16.1)	4 (6.8)	0
Age, mean (SD) years	13.2 (3.9)	12.6 (2.5)	11.1 (2.9)

Abbreviation: NR, not recorded.

When the model was updated in phase 3 using the Banfield, RVC1 and RVC2 datasets, the sensitivity for strategy 1 was generally slightly lower than for phases 1 and 2 and the NPV was similar (Table 5). Specificity and PPV were intermediate. Strategy 2 in phase 3 improved specificity to levels similar to those for strategy 2 in phase 1 and PPV was higher than for phase 2 (Table 6). Sensitivity was decreased compared with other phases and a modest decrease in NPV was noted.

The accuracy of the model in all phases was close to 90% with the exception of submodel M12+, for which accuracy was closer to 80% (Tables 5 and 6). In the final model, the rate of incoherence between submodels was 3.2%. The values (ie, model output) of incoherent observations were very close to the threshold value for the model cut-off.

The range of NPVs and PPVs associated with CKD prevalence fixed to between 10% and 30% are shown in Supporting Information Tables S2 (strategy 1) and S3 (strategy 2).

4 | DISCUSSION

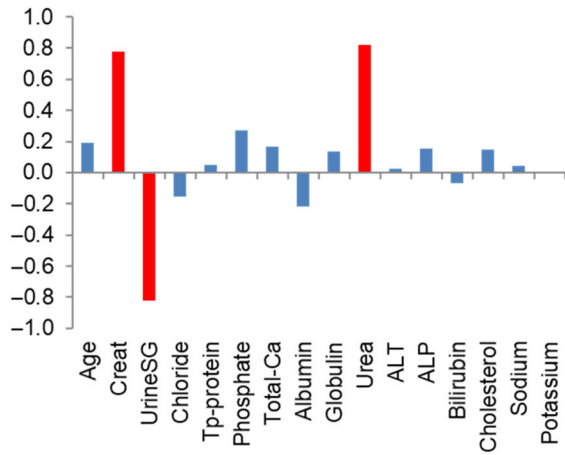
A model was developed that identifies cats ≥ 7 years of age that are at risk of developing azotemic CKD within 12 months, based on laboratory

variables very familiar to veterinarians in general practice. The work is distinct from classical CKD prediction research in cats^{10,19,21,27} owing to its use of ANN and the large size of the population for model validation and refinement.

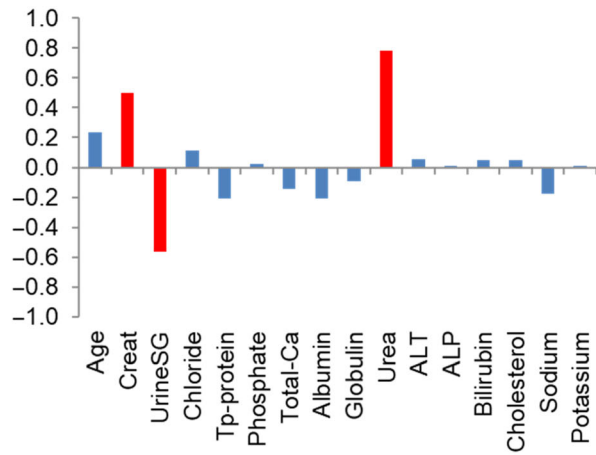
Artificial neural network modeling is a powerful tool that can harness the value of large and complex datasets and that, in the era of "big data," is of increasing interest in medical diagnostics (imaging and histopathology in particular) and prognostics.^{28,29} Disease predictions from ANN models are not limited by an understanding of the associated pathophysiology because they are based purely on mathematical relationships. Modeling by ANN is well established in human medicine as a tool to define populations at risk and to assist diagnosticians,¹² particularly in cancer, neurology and cardiology, where early diagnosis can impact leading causes of death.²⁸ The high morbidity and mortality rate of CKD in cats, combined with the difficulty of early diagnosis, suggest that ANN could be helpful in feline renal medicine. The first ANN model for predicting disease development in cats was a model to predict CKD.³⁰ The model presented here validates this approach.

The model derived in our study had an accuracy (percentage of true positives plus true negatives) of approximately 90%. When the model was set to interpret the computer simulations using a strategy that maximized both sensitivity and specificity (strategy 1), it was able to identify 87% of cats that would develop CKD within 12 months, and 70% of the cats that would not develop CKD within this time frame on the basis of plasma creatinine and BUN concentrations and USG at a single visit. The corresponding PPV (53%) meant that at least 1 in every 2 cats "positive" by this screening would develop azotemic CKD within 12 months. This strategy is appropriate if it is considered more important to correctly identify cats that will not develop CKD within the 12 months (ie, high NPV) than to correctly identify cats that will develop CKD. Also, the PPV of the model developed here will be calculably higher in older populations of cats because the prevalence of CKD increases,⁷ which may inform the owner or veterinarian as to its relative merits for an individual cat. An overall effect of increasing PPV with increasing CKD prevalence was identified when submodels were normalized to fixed prevalences of 10%-30% (Supporting Information Tables S2 and S3). The sensitivity of the model for predictions beyond 12 months was decreased, but the specificity increased.

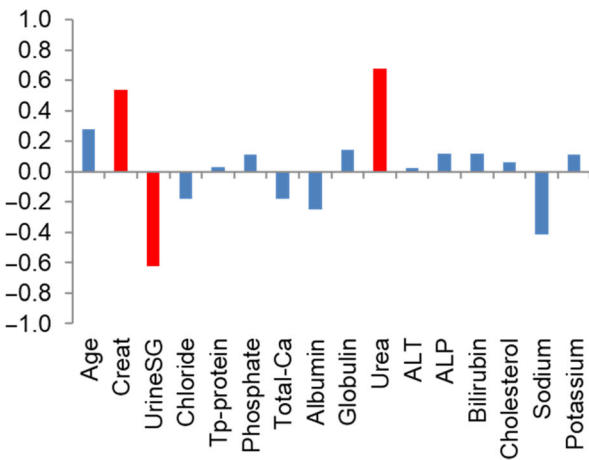
(A) Submodel M0



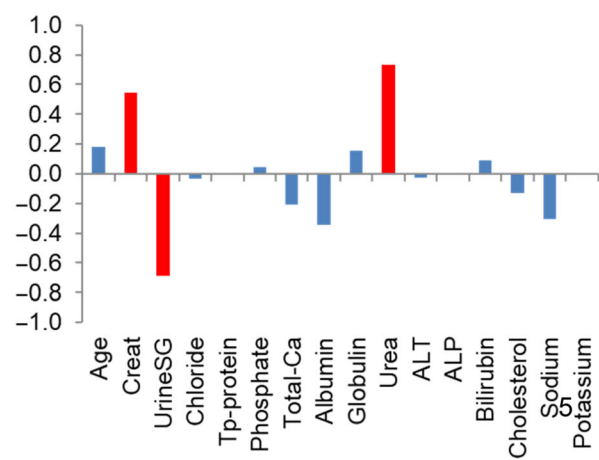
(B) Submodel M3



(C) Submodel M6



(D) Submodel M9



(E)

Submodel M12

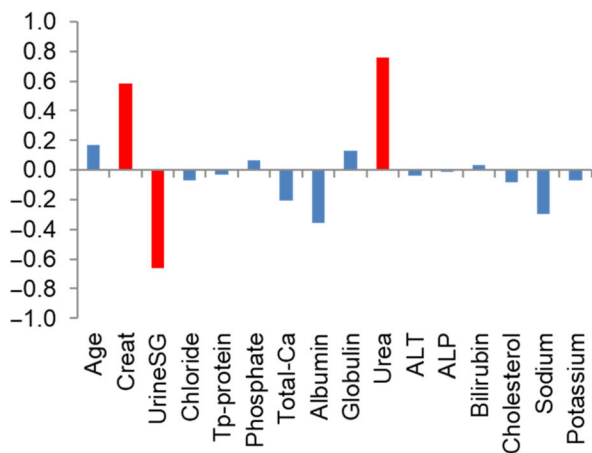


FIGURE 2 Input variables selection using factorial discriminant analysis. The figure shows the discriminant weight of each input variable on the model response of positive or negative for chronic kidney disease. The highest values are shown in red and are the variables retained in the model. M0-M12 are the submodels corresponding to the screening visit and prediction time frames of 3, 6, 9, and 12 months. ALP, alkaline phosphatase; ALT, alanine transaminase; creat, plasma creatinine; Tp-protein, total plasma protein; UrineSG, urine specific gravity

Modeling phase	Performance measure	M0	M3	M6	M9	M12	M12+
Phase 1	Accuracy	0.97	0.97	0.88	0.86	0.87	0.87
	Sensitivity	0.92	0.83	0.93	0.91	0.91	0.93
	Specificity	0.90	0.93	0.80	0.81	0.78	0.72
	PPV	0.72	0.77	0.66	0.70	0.70	0.66
	NPV	0.95	0.89	0.98	0.97	0.96	0.97
Phase 2	Accuracy	0.89	0.89	0.88	0.88	0.87	0.82
	Sensitivity	0.93	0.94	0.96	0.96	0.97	0.94
	Specificity	0.65	0.59	0.39	0.42	0.23	0.23
	PPV	0.29	0.30	0.27	0.33	0.31	0.62
	NPV	0.98	0.97	0.98	0.97	0.96	0.76
Phase 3	Accuracy	0.91	0.89	0.88	0.88	0.88	0.82
	Sensitivity	0.90	0.87	0.86	0.86	0.87	0.71
	Specificity	0.81	0.77	0.74	0.73	0.70	0.81
	PPV	0.44	0.45	0.47	0.52	0.53	0.85
	NPV	0.98	0.95	0.94	0.93	0.92	0.64

Notes: Submodel M0 was based on each clinic visit assessing the status of a cat with respect to a diagnosis of chronic kidney disease (CKD). Submodels M3-M12+ were for the prediction of the CKD status of negative cats in time frames out to 3, 6, 9, 12 months, and more than 12 months after M0. Abbreviations: AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value.

TABLE 5 Performance criteria of submodels in the 3 modeling phases based on a decision threshold set to maximize both sensitivity and specificity (strategy 1)

Modeling phase	Performance measure	M0	M3	M6	M9	M12	M12+
Phase 1	Accuracy	0.97	0.97	0.88	0.86	0.87	0.87
	Sensitivity	0.81	0.72	0.64	0.59	0.64	0.64
	Specificity	0.98	0.98	0.97	0.97	0.96	0.96
	PPV	0.89	0.91	0.89	0.91	0.91	0.95
	NPV	0.89	0.86	0.83	0.81	0.81	0.81
Phase 2	Accuracy	0.89	0.89	0.88	0.88	0.87	0.82
	Sensitivity	0.90	0.90	0.85	0.81	0.87	0.70
	Specificity	0.75	0.73	0.76	0.80	0.66	0.80
	PPV	0.34	0.37	0.44	0.52	0.46	0.82
	NPV	0.97	0.97	0.95	0.93	0.93	0.66
Phase 3	Accuracy	0.91	0.89	0.88	0.88	0.88	0.82
	Sensitivity	0.51	0.35	0.41	0.45	0.42	0.46
	Specificity	0.95	0.98	0.98	0.98	0.98	0.97
	PPV	0.65	0.78	0.81	0.84	0.87	0.96
	NPV	0.91	0.86	0.84	0.82	0.79	0.53

Notes: Submodel M0 was based on each clinic visit assessing the status of a cat with respect to a diagnosis of chronic kidney disease (CKD). Submodels M3-M12+ were for the prediction of the CKD status of negative cats in time frames out to 3, 6, 9, 12 months, and more than 12 months after M0. Abbreviations: AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value.

TABLE 6 Performance criteria of submodels in the 3 modeling phases based on a decision threshold set to maximize specificity and PPV (strategy 2)

This model is a tool for prediction, not a diagnostic algorithm. Its value is in identifying cats at higher risk of developing CKD in the future (or, in the case of submodel M0, a greater likelihood of having current CKD), so that they can be recommended for closer monitoring than an annual health check (ie, every 3-6 months), and be considered

for further investigations to determine the presence of any underlying disease leading to loss of functioning nephrons. It has the potential to increase the likelihood that CKD is diagnosed early, supporting more timely preventive medicine, especially in the older cat population. Although not intended for research purposes, it is possible that the

model also could help identify higher-risk cats for assessing the efficacy of interventions for early kidney disease in randomized controlled clinical trials.

The ability of the model to make accurate predictions at the time of presentation (M0), suggests that apparently healthy cats with a positive prediction of CKD might be at high risk of suffering from early renal disease, despite having plasma creatinine and BUN concentrations within

the normal reference ranges as well as a USG slightly >1.035. The model might therefore help veterinarians to identify cats with combinations of laboratory results close to but not exceeding normal limits, and potentially allow them to make an earlier diagnosis. Importantly, the model was fitted to cats on the basis of a single presentation, so a veterinarian would not need to have historical data for an accurate prediction. Of course, serial monitoring is an important part of early detection of chronic disease in veterinary patients and cannot be replaced by results from a single visit. The high NPV supports the expert-recommended annual frequency of health examinations for a cat with a “negative” prediction and no sign of kidney impairment or other risk factors,^{31,21} although no test is perfect and practitioners should be aware that some negative cats might develop CKD within a year.³¹

Strategy 2 for interpreting the model output favored specificity and a high PPV, making it more appropriate if there is an emphasis on not erroneously predicting that a cat will develop CKD (ie, limiting false positives) at the cost of missing some cats that will develop CKD (increase in false negative rate). On this basis, the model correctly identified at least 8 of 10 positive cats that would develop azotemia within 12 months. The decreased sensitivity meant that only 4 of 10 cats at risk would be detected, and the decrease in NPV to 79% meant that 2 in 10 cats predicted to be negative would be false negatives. The choice of strategy for setting the decision threshold should be guided by the purpose of the test and the population characteristics of the cats being screened. It may be more important for a screening test that predicts the future development of a disease to favor the detection of potential positives over false negatives, especially where the recommended action for positive individuals is not harmful, and in the case of more frequent health checks, is relatively inexpensive.

Phase 1 of the modeling process designed the model using well characterized cats from a clinical research database (RVC1 dataset). Despite the modest size of the dataset (218 cats), the model developed from it was robust, retaining high accuracy when presented with Banfield and RVC2 data from 3546 cats in phase 2. Artificial neural network modeling is very much dependent on the dataset on which it has been trained. The decrease in sensitivity observed during phase 2 therefore was expected. It might have been because of the difference

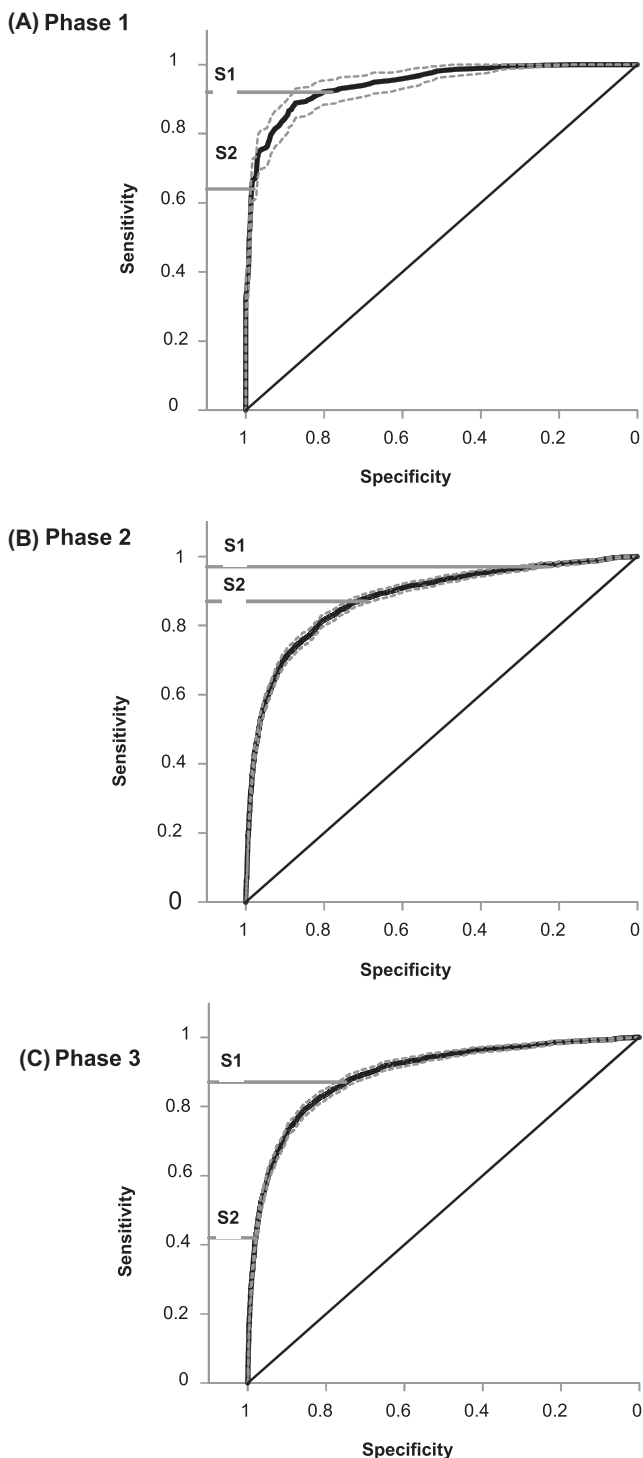


FIGURE 3 Legend on next column.

FIGURE 3 Receiver operator curve (ROC) curves of the submodel M12 for the 3 modeling phases. M12 is the submodel for predicting the development of CKD in a time frame of 12 months. A, Phase 1 consisted of the design of the model from the RCV1 dataset, including a first run of training and validation. B, In phase 2 the model was tested using the Banfield plus RVC2 datasets as independent data. C, Phase 3 consisted of a new cycle of training and validation keeping the same model design but using all 3 datasets (RVC1 plus RVC2 plus Banfield). The ROCs show all sensitivity/specificity pairs across the full range of potential thresholds, thereby encompassing both strategies for decision threshold setting. The dashed lines delineate the 95% confidence intervals obtained using bootstrapping. S1 is the decision threshold for strategy 1, which aimed to maximize both specificity and sensitivity. S2 is the decision threshold for strategy 2, which aimed for a high specificity corresponding to the highest positive predictive value. RVC, Royal Veterinary College

in the way CKD was diagnosed between the 2 institutions, demographic differences, or other factors that remain to be determined.

Another ANN model for CKD prediction in cats had a sensitivity of 63% and specificity of 99% to predict CKD 12 months before diagnosis.³⁰ Their model was different because it relied upon changes over time in the plasma creatinine and BUN concentrations and USG. Our model was designed to study any new case on the basis of a single visit, without a medical history, making it practical for a veterinarian who does not have easy access to past biochemistry test results. The previous model³⁰ also included age, which was not retained in our model on the basis of the FDA results, probably because most of the healthy cats screened by the RVC were ≥ 9 years of age. Both models used large datasets from clinical practice, but we have demonstrated that a model can also be created from a much smaller dataset of well-characterized cats from clinical studies. The previous model gave predictions of CKD up to 2 years before diagnosis.³⁰ We focused on a 12-month predictive window, which fits with the annual frequency of veterinary visits for routine vaccination and associated health screening.

Other studies have used a combination of traditional univariate and multivariate logistical regression to identify clinical variables associated with subsequent development of azotemia.¹⁹ Of their 2 models, the 1 with the higher sensitivity identified 83% of their population of cats that would in fact develop azotemia with a specificity of 47%, based on data from the research setting only. Although specificity was relatively low in phase 2 of the modeling reported here, additional training and validation in phase 3 increased it to at least 70%. These differences in approaches and results suggest that the ANN model may be more suitable for a screening test.

The 2 populations used for the ANN modeling encompassed a diverse group of cats over >20 years, during which time health care practices have varied. The size and relevance of a dataset from general practice were considered to outweigh any disadvantages of this approach. Although the model's accuracy was lower for longer predictive intervals, it remained high when both RVC and Banfield data were used. A specific strength of the Banfield dataset was the inclusion of control cats with medical conditions other than CKD. Any mass screening tool for clinical practice must have predictive value in client-owned cats presented to veterinary practice for different reasons and with a range of preexisting conditions. The prevalence of CKD in the Banfield population was considerably higher than in the RCV1 dataset despite the Banfield cats being generally younger, but it is not outside of reported ranges.^{32,33} It is plausible that CKD is more likely to be detected if cats are being monitored for plasma creatinine and BUN concentrations and USG simultaneously, which was required for our model. Furthermore, the ability of the RVC model to predict CKD in the Banfield cat population supports its relevance.

Of the 16 variables in the cleaned dataset, plasma creatinine and BUN concentrations and USG were clearly the most appropriate input variables by FDA. These are classical markers of renal disease, routinely used in clinical practice to diagnose CKD and are easy to measure from readily available samples. Additional biomarkers, including GFR,⁹ serum symmetric dimethylarginine (SDMA) concentration^{27,34} and the phosphaturic hormone FGF-23¹⁰ potentially could improve predictions, but were not evaluated here.³⁵ Although GFR is recognized as the gold standard measure for the early detection of CKD, it cannot be used for screening senior cats.^{9,36} In a

small retrospective study in colony cats, SDMA was shown to be a sensitive and specific biomarker for early diagnosis of CKD up to 4 years in advance of abnormal serum creatinine concentration suggesting impaired renal function.²⁷ The utility of both SDMA and FGF-23 should be studied more extensively to test their rigor in the field.¹⁰

Further development of the ANN modeling used in our study is possible. Although the dataset used to build the model was very large and diverse, it still may contain nuances unique to this dataset. This issue can be overcome by continuously training the model on new datasets collected from the field. Only visits without missing data were used to ensure the robustness of the model. It could be updated using statistical techniques to generate missing data, in order to assess its predictive strength when data values exist for only 1 or 2 of the selected variables. The clinical utility of the model also could be improved if sufficient data was available to extend the prediction time frame beyond 18 months. The rate of incoherence between submodels was 3.2%. This rate is very low and inherent to the use of independent submodels. Moreover, when a negative prediction appeared to contradict an earlier positive prediction, the variables were very close to the thresholds of the submodels, indicating that those individuals were at the margin of being reported as positive. Reporting an estimate of closeness to the threshold along with the binary outcome might be useful to clinicians.

In summary, ANN techniques have been used to generate a model that can identify individuals in the general population of cats ≥ 7 years of age that are at risk of developing CKD within 12 months, and that therefore will benefit most from more frequent clinical follow-up to maximize the opportunities for early CKD diagnosis and intervention. Predictions were based on the widely used variables of plasma creatinine and BUN concentrations and USG from a single visit, making the model highly applicable to clinical practice without the need for historical data. The model also may help improve the selection of older cats suitable for prospective clinical trials to investigate interventions aimed at delaying the onset of azotemia.

ACKNOWLEDGMENT

The authors thank Dr Susana Ryan for her help reviewing and editing this article.

CONFLICT OF INTEREST DECLARATION

Vincent Biourge, Alexandre Feugier are employees of Royal Canin SAS, Richard Bradley is employee of WCPN, Molly McAllister is an employee of Banfield Vet Hospital. RC, WCPN, and Banfield Vet Hospital are subsidiaries of Mars Inc. Sebastien Delmotte is a consultant for RC. The kidney clinic of the Royal Veterinary College is supported in part by Royal Canin SAS.

OFF-LABEL ANTIMICROBIAL DECLARATION

Authors declare no off-label use of antimicrobials.

INSTITUTIONAL ANIMAL CARE AND USE COMMITTEE (IACUC) OR OTHER APPROVAL DECLARATION

The Royal Canin and the Royal Veterinary College ethical committees approved the protocols used to collect part of the data necessary for

this study. Owners of cats presented at Banfield clinics signed an inform consent allowing the use of data from their cat for research purpose.

HUMAN ETHICS APPROVAL DECLARATION

Authors declare human ethics approval was not needed for this study.

ORCID

Vincent Biourge  <https://orcid.org/0000-0002-2237-6397>

Richard Bradley  <https://orcid.org/0000-0001-7508-7102>

REFERENCES

- Polzin DJ. Chronic kidney disease in small animals. *Vet Clin North Am Small Anim Pract.* 2011;41:15-30.
- Brown CA, Elliott J, Schmiedt CW, Brown SA. Chronic kidney disease in aged cats: clinical features, morphology, and proposed pathogenesis. *Vet Pathol.* 2016;53:309-326.
- Lawson J, Elliott J, Wheeler-Jones C, Syme H, Jepson R. Renal fibrosis in feline chronic kidney disease: known mediators and mechanisms of injury. *Vet J.* 2015;203:18-26.
- Chakrabarti S, Syme HM, Brown CA, Elliott J. Histomorphometry of feline chronic kidney disease and correlation with markers of renal dysfunction. *Vet Pathol.* 2013;50:147-155.
- Brown SA, Crowell WA, Brown CA, Barsanti JA, Finco DR. Pathophysiology and management of progressive renal disease. *Vet J.* 1997;154:93-109.
- Hostetter TH, Rennke HG, Brenner BM. Compensatory renal hemodynamic injury: a final common pathway of residual nephron destruction. *Am J Kidney Dis.* 1982;1:310-314.
- O'Neill DG, Church DB, McGreevy PD, et al. Longevity and mortality of cats attending primary care veterinary practices in England. *J Feline Med Surg.* 2015;17:125-133.
- Paepe D, Verjans G, Duchateau L, Piron K, Ghys L, Daminet S. Routine health screening: findings in apparently healthy middle-aged and old cats. *J Feline Med Surg.* 2013;15:8-19.
- Finch N. Measurement of glomerular filtration rate in cats: methods and advantages over routine markers of renal function. *J Feline Med Surg.* 2014;16:736-748.
- Finch NC, Geddes RF, Syme HM, Elliott J. Fibroblast growth factor 23 (FGF-23) concentrations in cats with early nonazotemic chronic kidney disease (CKD) and in healthy geriatric cats. *J Vet Intern Med.* 2013;27:227-233.
- Dayhoff JE, DeLeo JM. Artificial neural networks: opening the black box. *Cancer.* 2001;91:1615-1635.
- Amato F, López A, Peña-Méndez EM, Vañhara P, Hampl A, Havel J. Artificial neural networks in medical diagnosis. *J Appl Biomed.* 2013;11:47-58.
- Vapnik V. *Statistical Learning Theory.* New York, NY: John Wiley & Sons; 1998 ISBN: 978-0-471-03003-4.
- van den Broek DH, Chang YM, Elliott J, Jepson RE. Chronic kidney disease in cats and the risk of total hypercalcemia. *J Vet Intern Med.* 2017;31:465-475.
- Barber PJ, Elliott J. Feline chronic renal failure: calcium homeostasis in 80 cases diagnosed between 1992 and 1995. *J Small Anim Pract.* 1998;39:108-116.
- Elliott J, Barber PJ. Feline chronic renal failure: clinical findings in 80 cases diagnosed between 1992 and 1995. *J Small Anim Pract.* 1998;39:78-85.
- Elliott J, Syme HM, Reubens E, Markwell PJ. Assessment of acid-base status of cats with naturally occurring chronic renal failure. *J Small Anim Pract.* 2003;44:65-70.
- Geddes RF, Biourge V, Chang Y, Syme HM, Elliott J. The effect of moderate dietary protein and phosphate restriction on calcium-phosphate homeostasis in healthy older cats. *J Vet Intern Med.* 2016;30:1690-1702.
- Jepson RE, Brodbelt D, Vallance C, Syme HM, Elliott J. Evaluation of predictors of the development of azotemia in cats. *J Vet Intern Med.* 2009;23:806-813.
- Syme HM, Barber PJ, Markwell PJ, Elliott J. Prevalence of systolic hypertension in cats with chronic renal failure at initial evaluation. *J Am Vet Med Assoc.* 2002;220:1799-1804.
- Finch NC, Syme HM, Elliott J. Risk factors for development of chronic kidney disease in cats. *J Vet Intern Med.* 2016;30:602-610.
- Fisher RA. The use of multiple measurements in taxonomic problems. *Ann Eugen.* 1936;7:179-188.
- Friedman JH. Regularized discriminant analysis. *J Am Stat Assoc.* 1989;84:165-175.
- Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagation errors. *Nature.* 1996;323:533-536.
- Hagan M, Demuth HB, Beale M. *Neural Network Design.* Boston, MA: PWS Publishing Company; 1996.
- Youden WJ. Index for rating diagnostic tests. *Cancer.* 1950;3:32-35.
- Hall JA, Yerramilli M, Obare E, Yerramilli M, Jewell DE. Comparison of serum concentrations of symmetric dimethylarginine and creatinine as kidney function biomarkers in cats with chronic kidney disease. *J Vet Intern Med.* 2014;28:1676-1683.
- Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol.* 2017;2:230-243.
- Karako K, Chen Y, Tang W. On medical application of neural networks trained with various types of data. *Biosci Trends.* 2019;12:553-559.
- Bradley R, Tagkopoulos I, Kim M, et al. Predicting early risk of chronic kidney disease in cats using routine clinical laboratory tests and machine learning. *J Vet Intern Med.* 2019;33:2644-2656.
- Pittari J, Rodan I, Beekman G, et al. American association of feline practitioners. Senior care guidelines. *J Feline Med Surg.* 2009;11:763-778.
- Brown SA, Brown CA. Single-nephron adaptations to partial renal ablation in cats. *Am J Physiol.* 1995;269:R1002-R1008.
- Marino CL, Lascelles BD, Vaden SL, et al. Prevalence and classification of chronic kidney disease in cats randomly selected from four age groups and in cats recruited for degenerative joint disease studies. *J Feline Med Surg.* 2014;16:465-472.
- Braff J, Obare E, Yerramilli M, Elliott J, Yerramilli M. Relationship between serum symmetric dimethylarginine concentration and glomerular filtration rate in cats. *J Vet Intern Med.* 2014;28:1699-1701.
- Sparkes AH, Caney S, Chalhoub S, et al. ISFM consensus guidelines on the diagnosis and management of feline chronic kidney disease. *J Feline Med Surg.* 2016;18:219-239.
- Finch NC, Syme HM, Elliott J. Development of an estimated glomerular filtration rate formula in cats. *J Vet Intern Med.* 2018;32:1970-1976.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Biourge V, Delmotte S, Feugier A, Bradley R, McAllister M, Elliott J. An artificial neural network-based model to predict chronic kidney disease in aged cats. *J Vet Intern Med.* 2020;34:1920-1931. <https://doi.org/10.1111/jvim.15892>