

Predicting likelihood of in vivo chemotherapy response in canine lymphoma using ex vivo drug sensitivity and immunophenotyping data in a machine learning model

Zach Bohannon¹  | Raghavendra Sumanth Pudupakam¹ | Jamin Koo^{1,2,3} | Harrison Horwitz¹ | Josephine Tsang¹ | Amanda Polley¹ | Enyang James Han¹ | Elmer Fernandez¹ | Stanley Park¹ | Deanna Swartzfager¹ | Nicholas Seah Xi Qi¹ | Chantal Tu⁴ | Wendi Velando Rankin⁴ | Douglas H. Thamm⁵ | Hye-Ryeon Lee^{1,2} | Sungwon Lim^{1,2}

¹ImpriMed, Inc., Palo Alto, California

²ImpriMed Korea, Inc., Seoul, Republic of Korea

³Department of Chemical Engineering, Hongik University, Seoul, Republic of Korea

⁴SAGE Veterinary Centers, Dublin, California

⁵Flint Animal Cancer Center, Colorado State University, Fort Collins, Colorado

Correspondence

Sungwon Lim, CEO, ImpriMed, Inc., 4030 Fabian Way, Palo Alto, CA 94303, USA.
Email: sungwon@imprimedicine.com

Abstract

We report a precision medicine platform that evaluates the probability of chemotherapy drug efficacy for canine lymphoma by combining ex vivo chemosensitivity and immunophenotyping assays with computational modelling. We isolated live cancer cells from fresh fine needle aspirates of affected lymph nodes and collected post-treatment clinical responses in 261 canine lymphoma patients scheduled to receive at least 1 of 5 common chemotherapy agents (doxorubicin, vincristine, cyclophosphamide, lomustine and rabacfosadine). We used flow cytometry analysis for immunophenotyping and ex vivo chemosensitivity testing. For each drug, 70% of treated patients were randomly selected to train a random forest model to predict the probability of positive Veterinary Cooperative Oncology Group (VCOG) clinical response based on input variables including antigen expression profiles and treatment sensitivity readouts for each patient's cancer cells. The remaining 30% of patients were used to test model performance. Most models showed a test set ROC-AUC > 0.65, and all models had overall ROC-AUC > 0.95. Predicted response scores significantly distinguished ($P < .001$) positive responses from negative responses in B-cell and T-cell disease and newly diagnosed and relapsed patients. Patient groups with predicted response scores >50% showed a statistically significant reduction (log-rank $P < .05$) in time to complete response when compared to the groups with scores <50%. The computational models developed in this study enabled the conversion of ex vivo cell-based chemosensitivity assay results into a predicted probability of in vivo therapeutic efficacy, which may help improve treatment outcomes of individual canine lymphoma patients by providing predictive estimates of positive treatment response.

KEYWORDS

chemosensitivity, dog, lymphosarcoma, machine learning, precision medicine

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 ImpriMed, Inc. *Veterinary and Comparative Oncology* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Lymphoma is one of the most common cancers in dogs and represents the most common haematopoietic malignancy.¹ Dogs develop lymphoma at a higher incidence rate compared to people (114 per 100 000 dogs compared to 20 per 100 000 humans).¹ The most common form of canine lymphoma is the multicentric form, which manifests as generalized lymphadenopathy. Approximately 75% of canine lymphoma cases develop as multicentric lymphoma, with other forms being less common (6.9% alimentary and less than 5% mediastinal).² In terms of the immunophenotype, the B-cell subtype comprises 58% of newly diagnosed cases, and the T-cell subtype comprises 35%.³ Survival rates differ significantly between immunophenotypes—approximately 9 and 6 months for the most common B- and T-cell subtypes, respectively.¹

Chemotherapy is considered the treatment of choice, and the highest response rates and remission durations are achieved with multi-agent chemotherapy. Treatment regimens typically combine cyclophosphamide, doxorubicin, vincristine, and prednisone—collectively known as CHOP—which results in a clinical remission rate of 73% to 96% and overall survival time of 275 to 344 days for high-grade lymphomas.^{4,5} Recently, a novel antineoplastic drug called rabacfosadine (TANOVEA-CA1) has also become available; an overall response rate of 87% and median progression free interval of 122 days were reported for naïve canine multicentric lymphoma treated with single agent rabacfosadine.⁶ With the failure of first-line protocols or relapse, rescue treatments are considered, with lomustine being a key component of many rescue protocols.^{7,8} Rescue protocols typically result in lower response rates and shorter remission durations than first-line protocols.^{7,9,10} Treatment outcomes also differ between different subtypes of canine lymphoma, with most T-cell subtypes having poorer treatment response than B-cell subtypes.^{11,12}

The significant variance in treatment responses in various clinical scenarios implies a critical need to predict treatment response accurately, especially in patients with rare lymphoma subtypes or subtypes with low remission rates. Addressing this need may improve treatment outcomes by enabling clinicians to select the most effective drugs and exclude ineffective ones for each patient.

Cell-based *ex vivo* drug sensitivity assays have been widely studied as a precision medicine tool to recapitulate the tumour micro-environment *in vitro* and predict *in vivo* responses in human lymphoproliferative disorders.^{13–17} For canine lymphoma, Pawlak et al. reported an *in vitro* chemosensitivity assay that measures the cytotoxicity of various anticancer reagents in high-grade primary lymphoma cells.¹⁸ This previous work showed that drug sensitivity varies among individual patients and that a direct measurement of drug response in primary cancer cells is a potential predictor of actual response in the body. However, parameters derived from the outcome of a drug sensitivity assay alone can be insufficient when predicting *in vivo* response. Adding other phenotypic information such as the immunophenotype of a patient's cancer cells as determined by flow cytometry may enhance the predictive value of *ex vivo* drug sensitivity testing.^{18,19}

Individualized patient outcome modelling is another core feature of precision medicine. In human oncology, personalized predictive modelling has many clinical applications, especially those related to diagnostic and prognostic decisionmaking.²⁰ Machine learning is a particularly popular modern approach for predicting patient outcomes in human oncology,²¹ and machine learning strategies have been successfully applied to a variety of human cancers and treatment regimens.^{22–24} One of the most promising avenues of this research is related to direct prediction of drug response²⁵; however, despite the popularity and promise of predictive machine learning in human oncology, it remains relatively uncommon in veterinary oncology.

Clinical models are typically assessed strictly on predictive accuracy. However, the choice of output format is critical for maximizing clinical utility. Many traditional modelling approaches output some categorical value such as “good response” because categorical model outputs have a variety of convenient performance assessment methods.²⁶ However, these traditional models are not calibrated to accurately assess the likelihood of a good response, meaning a patient with a 55% likelihood of positive response and a patient with a 95% likelihood of positive response would both be classified as “positive responders”. Clearly, important information is lost in the conversion from likelihood scores to categorical reporting.²⁷ One less common modelling approach that represents major gains in clinical utility is to develop models that directly report the likelihood of positive response for a patient.^{20,21,27} This is especially important when there may be multiple treatment options available for a given case.

However, the development of probabilistic models that directly and accurately report likelihood scores is associated with significant challenges, such as more complex performance assessments.²⁸ Because probabilistic models report both a response and the likelihood of that response (eg, a 95% likelihood of positive response), many traditional performance metrics used for categorical models, such as sensitivity and specificity, must be extended to address this likelihood. For example, in categorical models, a model that predicts a positive response in a patient that fails to respond is simply wrong, whereas in probabilistic models, a model that predicts a 90% likelihood of positive response in a patient that fails to respond must be penalized more harshly than a model that predicts a 30% likelihood of positive response in the same patient. This difference leads to natural, if less well known, extensions of popular modelling metrics. Probabilistic models can be assessed using metrics such as the Brier score, which penalizes inaccuracy as in the example above, or calibration curves (as seen in Figure 1), which graph the predictions of a probabilistic model against the actual frequency of positive or negative outcomes in a dataset.²⁸ For a likelihood model to have good performance, among all patients predicted to have a 70% likelihood score, roughly 70% of them would have positive responses and 30% of them would have negative responses.

In this study, we present a machine learning approach for predicting treatment response in canine lymphoma. Rather than reporting discrete classifications of positive or negative response, we sought to use random forest models^{29,30} to accurately predict the likelihood of positive responses to individual chemotherapies. We

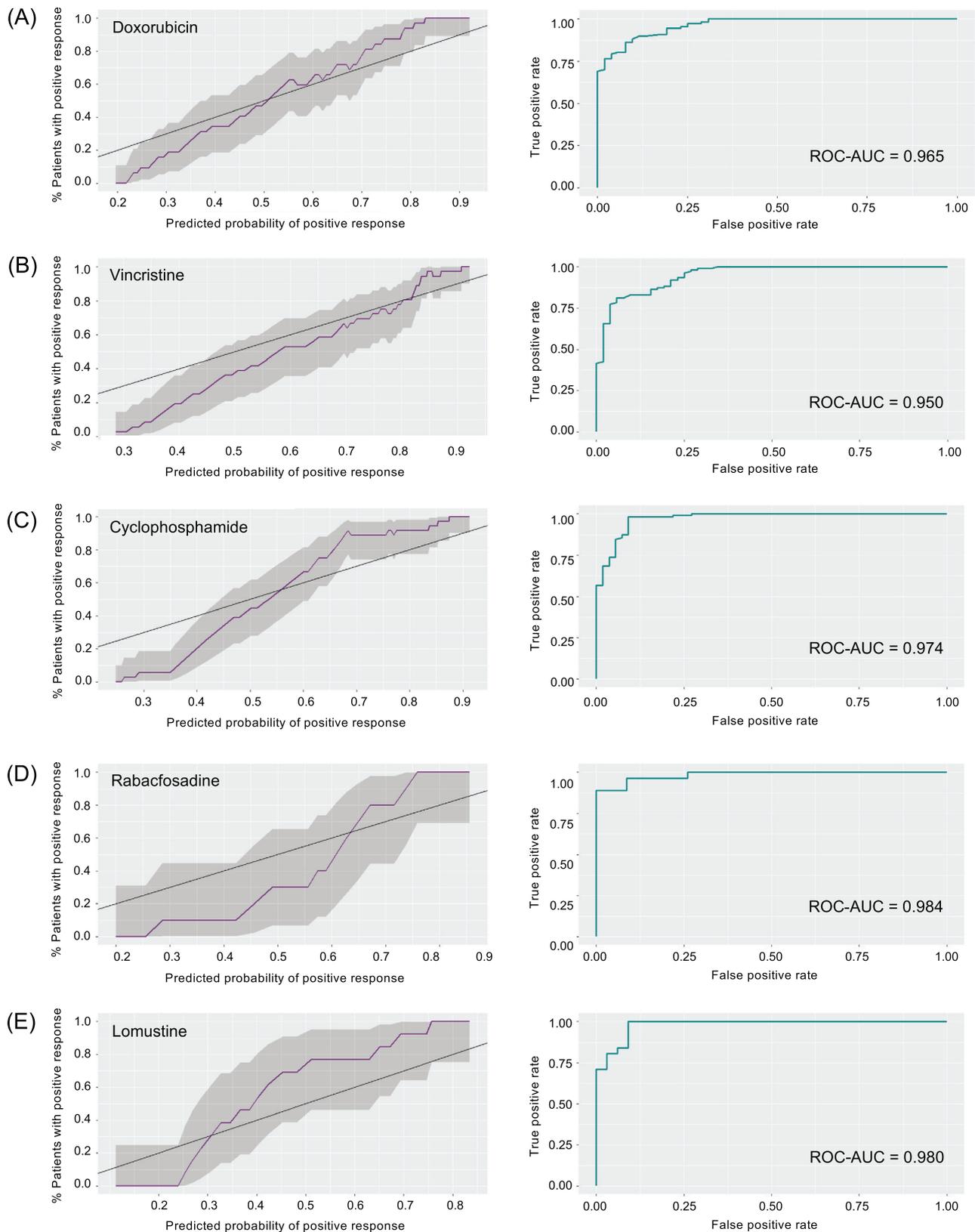


FIGURE 1 Performance metrics for probabilistic random forest models of drug responses. Sliding-window calibration graphs (left) of predicted probability of positive response vs percentage of patients experiencing a positive response (window size = 20% of sample size; purple line = prediction; shaded region = 95% prediction interval; black line = perfect calibration) and ROC curves (right) of probabilistic random forest model for A, doxorubicin; B, vincristine; C, cyclophosphamide; D, rabacfosadine and E, lomustine [Colour figure can be viewed at wileyonlinelibrary.com]

hypothesized that the combination of direct drug sensitivity measurements and flow cytometry data could accurately predict treatment outcomes in a machine learning model of treatment response in canine lymphoma.

2 | METHODS

2.1 | Case selection

The study included 261 canine patients with confirmed diagnosis of lymphoma who received at least one dose of the modelled drugs (Supplementary Table S1). All dog owners provided informed consent for participation in this study. Cell and molecular assays were conducted on fine needle aspiration (FNA) samples that were submitted to a clinical laboratory (ImpriMed, Inc., Palo Alto, CA) between April 2018 and October 2019. A sample was considered neoplastic if it met the following criteria: cytological or histological diagnosis of lymphoma plus flow cytometric findings that >80% of the lymphocytes had either B- or T-cell immunophenotype or >60% of the lymphocytes had a single phenotype and positive clonality by PCR for antigen receptor rearrangements. Data extracted from submitted medical records included patient age, weight, sex, clinical stage, date of diagnosis, date of initiation of treatment, treatment protocol, time to treatment response, time to disease progression and date of death.

2.2 | Flow cytometry

FNA samples from patient lymph nodes were shipped overnight in Transport Media (ImpriMed, Inc., Palo Alto, CA) with ice packs, and most samples arrived with high cell viability (Supplementary Figure S1). Flow cytometry was performed within 48 hours of sample collection. Directly conjugated antibody combinations consisting of fluorescein isothiocyanate (FITC), phycoerythrin (PE), alexa fluor 647 (AF647) and allophycocyanin (APC) were used for surface staining cell suspensions. When an abnormal lymphocyte population was identified on routine analysis, the following antibodies (with specific clones in parentheses) were included in the flow cytometry panel for immunophenotyping: anti-canine CD21 PE (CA2.1D6), CD3 FITC (CA17.2A12), CD4 PE (YKIX302.9), CD8 AF647 (YCATE55.9), CD5 PE (YKIX322.3), CD45 FITC (YKIX716.13), class II MHC APC (YKIX.334.2), CD34 FITC (1H6), and anti-human CD14 AF647 (TUK). The class II MHC antibody was purchased from Thermo Fisher Scientific (Waltham, MA), and all other antibodies from Bio-Rad Laboratories (Hercules, CA). Aliquots of cells at concentrations ranging from 0.2 to 1.0×10^6 cells/mL were labelled with antigen-specific antibodies or isotype controls at a concentration of 2 to 10 $\mu\text{g}/\text{mL}$ in phosphate-buffered saline (PBS) with 1% bovine serum albumin. Labelled cells were incubated at 4°C in the dark for 45 minutes and then washed repeatedly. Cytometric analysis was performed using Guava easyCyte 8HT (Luminex, Austin, TX) and FCS Express 6 (De Novo Software, Pasadena, CA) was used to analyse data. Lymphocyte populations were

gated to exclude dead cells based on forward scatter vs side scatter plots. When the antigen fluorescence was higher than that of an antibody isotype control, the antigen was considered positive.

2.3 | Drug sensitivity testing

Ex vivo drug sensitivity assays were conducted on freshly isolated primary cells derived from the patients' FNA samples. Red blood cells (RBC) in the FNA samples were lysed in RBC Lysis Buffer (Thermo Fisher Scientific, Waltham, MA). The RBC-lysed samples were washed twice with PBS and resuspended in Optimum Culture Media (ImpriMed), which maintained high cell viability over the timespan of the assay (Supplementary Figure S2). Each well in a 384-well microtitre plate (Corning, Corning, NY) was seeded with 30 000 cells. The tested chemotherapeutic drugs were purchased from commercial vendors as follows: Selleckchem (Houston, TX) for doxorubicin, vincristine sulfate, and lomustine, and Niomech-IIT GmbH (Bielefeld, Germany) for mafosfamide cyclohexylamine. Chemically stable mafosfamide cyclohexylamine was selected as a cyclophosphamide analog, which does not require hepatic activation to produce its active metabolite 4-hydroxy-cyclophosphamide *in vitro*.³¹ Rabacfosadine (TANOVEA-CA1) was generously provided by the manufacturer (VetDC, Fort Collins, CO). All drugs were prepared in PBS containing 10% DMSO. Cells in the microtitre plates were treated in duplicate with the drug compounds in seven different concentrations per drug, covering approximately a 100 000-fold concentration range (1.3 nM to 100 μM), in addition to seven untreated control wells. The cells were incubated for 72 hours at 37°C in a 5% CO₂-charged incubator. After the incubation, cell viability was assessed using alamarBlue (Thermo Fisher Scientific) and the plates were read with the Spark multimode microplate reader (Tecan, Männedorf, Switzerland) following the manufacturer's instructions. The dose-response curves were analysed in GraphPad Prism 7 (GraphPad Software, San Diego, CA), which was used to generate IC₅₀ and AUC values. Maximum cytotoxicity was defined as the maximum percentage of cells killed in any well treated with a given drug.

2.4 | Response annotation

Patient responses were annotated using retrospective medical record review and heuristics derived from the VCOG Response Evaluation Criteria for Peripheral Nodal Lymphoma.³² For model training and testing, an automated response annotation approach was used. If a patient received multiple therapies in a combination regimen such as CHOP, response to each drug was assessed individually at the most proximal post-treatment visit, typically 1 to 2 weeks after administration.

Our weekly response annotations were further modified confidence weighting approach, which was an attempt to address confounding factors related to multidrug regimens. Because patient samples represent a "snapshot" of disease state, which changes over time as a result of both mutation and selective forces exerted by

chemotherapy, weekly responses were adjusted by a weighting factor scaled by the time elapsed since the sampling date. This annotation strategy was calibrated to assess the most likely “true” response within 90 days of the sample acquisition date. In cases of pre-treated patients (including relapse), responses were annotated both before and after the sampling date, which allowed us to effectively annotate patients who showed resistance to therapy before we received their samples.

The base selection equation was:

$$R = \max(\{r_{CR}, r_{PR}, r_{SD}, r_{PD}\})$$

where R represents the final annotated response for a drug, and each response category's weight was formulated as:

$$r_{\text{RECIST}} = \sum_{i=0}^n 1 - \frac{|t_i - 14|}{76}$$

where n is the number concordant weekly responses for a given drug in a patient and t_i is the time elapsed between the sample date and response i , with $t_0 = 90$ to ensure responses with no valid annotations were assigned a weight of 0. In cases of discordant annotations within a month of the sampling date, the later response was granted a further upward weighting adjustment to deemphasize transient responses. As an example, a patient that shows complete response after a single dose of cyclophosphamide immediately after sampling is weighted as a “more confident” response (ie, has a higher weighting value) than a patient who achieves complete response after 3 weeks of intermittent cyclophosphamide therapy, as in CHOP. Similarly, a patient who maintained durable complete response over several weeks of doxorubicin alone would be weighted as a “more confident” than a patient who only received doxorubicin once or twice over a period of months. The automated annotation was calibrated for a 95% concordance rate when compared to unanimous annotation consensus among three blinded independent reviewers (Raghavendra Sumanth Pudupakam, Hye-Ryeon Lee and Sungwon Lim).

2.5 | Predictive modelling

Models were trained and tested based on responses to single drugs. The selected drugs represented some of the most common cytotoxic drugs used for treatment of canine lymphoma and included doxorubicin, vincristine, cyclophosphamide, lomustine and rabacfosadine. Parameters included drug response variables (either area under the drug response curve or IC_{50} and maximum cytotoxicity³³) and the percentage of the sampled cell population expressing a variety of flow cytometry markers, including percent of lymphocytes, percent of large lymphocytes, forward scatter, side scatter, and percentages of cells expressing CD21, class II MHC, CD3, CD8 and CD34. All variables were formulated as continuous values, and the same input variables were used for each model unless otherwise noted.

Each drug response cohort was split into model training and testing cohorts using random selection, with 70% of samples being used for training and 30% being used for testing. All data were centred and scaled based on means and standard deviations derived from the training set. In samples missing data for four or fewer variables, data were imputed using a k-nearest neighbours approach with $k = 5$ and all nearest neighbours derived from the training set.³⁴

The pre-processed training data was used as input to tune and train drug-specific random forest models.²⁹ Tuning was performed using repeated cross-validation with 10 repeats of 4 random folds each and a log loss optimization function over 500 trees for each run. Tuning parameters included minimum leaf node size, number of variables assessed per split, and splitting criteria (maximum Gini coefficient³⁵ or random selection, otherwise known as extremely randomized trees²⁹). Expected model outputs were probability of a positive response to a specific drug. Variable importance was assessed using mean decrease in node impurity.³⁵

2.6 | Model performance assessment

Model performance was assessed by applying the tuned and trained models to the testing cohort for each drug and scored using a combination of ROC-AUC and Brier score. Because each model's training set had a different rate of positive responses, we also assessed test set performance using Brier skill score calibrated to a reference model of the positive response rate in the entire cohort for a given drug. Models were further assessed by visual inspection of their performance across the whole dataset using rolling calibration curves with binomial confidence intervals and boxplots.

2.7 | Time-to-response analysis

Response data for time to response analysis were collected from patient medical records. Time to complete response was assessed from the time of sample acquisition to the time of first documented complete response for a patient. Patients that did not achieve a documented complete response were censored at the time of last follow-up or death.

2.8 | Statistical methods

All machine learning models were developed in R 3.6.1. Data pre-processing was performed using the caret package.³⁶ The random forest model was generated using the ranger package.³⁰ Performance assessments were performed using the ROCR package³⁷ and custom R code. Time to complete response analyses were performed using R's survival package.³⁸ An unpaired t test was used to assess the statistical significance of the differences in the mean probability score among various patient subgroups.

2.9 | Cell line validation statement

No cell lines were used in this study.

3 | RESULTS

3.1 | Patient demographics

Table 1 presents the demographic information for all 261 dogs in the study. Intermediate to large-sized B-lymphocytes expressing high levels of class II MHC were the most common immunophenotype in the study. CD4-positive lymphocytes with low levels of class II MHC were the frequently identified immunophenotype among T-cell lymphomas.

3.2 | Optimum model parameters varied despite relatively consistent optimum input variables

Model tuning parameters were broadly comparable across drugs, but there were slight differences between models. All models showed the best tuning performance with minimum leaf node size equal to 1 and the number of variables per split equal to 2. The doxorubicin and cyclophosphamide models were most performant when splitting criteria were based on the maximum Gini coefficient, whereas the other three models were optimized using the extremely randomized trees approach, which randomly selects splitting variables among a selected subset.²⁹ Most of the tested drugs showed better predictive power when the drug response was formulated as a combination of IC₅₀ and maximum cytotoxicity rather than the area under the drug response curve. The only exception was cyclophosphamide, which

TABLE 1 Patient characteristics

Parameter	Total study population (N = 261)	Doxorubicin study population (N = 159)	Vincristine study population (N = 163)	Cyclophosphamide study population (N = 166)	Rabacfosadine study population (N = 50)	Lomustine study population (N = 64)
Age						
Median ± SD	9 ± 3.1	9 ± 3.1	9 ± 3.1	9 ± 3.1	9 ± 2.7	9 ± 3.2
Range	1 to 16 years	1 to 16 years	1 to 16 years	1 to 16 years	3 to 15 years	2 to 16 years
Weight						
Median ± SD	27 ± 11.9	27 ± 12.4	27 ± 11.4	27 ± 11.7	25.5 ± 12.9	27 ± 11.5
Sex						
Male	18 (6.9%)	13 (8.1%)	10 (6.1%)	9 (5.4%)	2 (4.0%)	5 (7.8%)
Female	8 (3.0%)	5 (3.1%)	4 (2.4%)	4 (2.4%)	2 (4.0%)	5 (7.8%)
Male neutered	127 (48.6%)	81 (50.9%)	79 (48.4%)	79 (47.5%)	24 (48.0%)	28 (43.7%)
Female spayed	97 (37.1%)	53 (33.3%)	62 (38.0%)	66 (39.7%)	19 (38.0%)	21 (32.8%)
Unknown	11 (4.2%)	7 (4.4%)	8 (4.9%)	8 (4.8%)	3 (6.0%)	5 (7.8%)
Clinical stage						
2	6 (2.2%)	3 (1.9%)	2 (1.2%)	1 (0.6%)	0 (0.0%)	2 (3.1%)
3	97 (37.1%)	60 (37.7%)	62 (38.0%)	61 (36.7%)	16 (32.0%)	27 (42.2%)
4	48 (18.3%)	27 (17.0%)	26 (16.0%)	26 (15.7%)	15 (30.0%)	7 (10.9%)
5	15 (5.7%)	7 (4.4%)	6 (3.7%)	6 (3.6%)	3 (6.0%)	1 (1.6%)
NS ^a	95 (36.3%)	62 (39%)	67 (41.1%)	72 (43.4%)	16 (32.0%)	27 (42.2%)
a	101 (61.2%)	66 (68%)	57 (59.4%)	56 (59.6%)	25 (73.5%)	30 (81%)
b	34 (20.6%)	13 (13.4%)	17 (17.7%)	17 (18.0%)	8 (23.5%)	4 (10.8%)
NSS ^b	30 (18.2%)	18 (18.6%)	22 (22.9%)	21 (22.3%)	1 (3.0%)	3 (8.1%)
Immunophenotype						
B	171 (65.5%)	103 (64.7%)	100 (61.3%)	101 (60.8%)	35 (70.0%)	29 (45.3%)
T	46 (17.6%)	27 (16.9%)	32 (19.6%)	29 (17.4%)	8 (16.0%)	26 (40.6%)
Others ^c	30 (11.4%)	17 (10.6%)	19 (11.6%)	23 (13.8%)	2 (4.0%)	7 (10.9%)
N ^d	14 (5.3%)	12 (7.5%)	12 (7.3%)	13 (7.8%)	5 (10.0%)	2 (3.1%)

^aNot staged.

^bNot substaged.

^cDual immunophenotype, inconclusive, non-B/non-T, and immature phenotypes.

^dNot immunophenotyped.

showed significantly better performance when using the area under the drug response curve.

3.3 | Model performance

Test set performance was measured using the area under the receiver operating characteristic curve (ROC-AUC) for positive response (clinical response or partial response), Brier score, and Brier skill score. The doxorubicin predictive model showed the best performance with a test set ROC-AUC of 0.702, Brier score of 0.188, and Brier skill score of 0.09. The cyclophosphamide model was similarly performant with an ROC-AUC of 0.697 but a less performant Brier score of 0.21 and Brier skill score of 0.05. Vincristine had an ROC-AUC of 0.603, Brier score of 0.239, and Brier skill score of 0.02. Rabacfosadine had an ROC-AUC of 0.714, Brier score of 0.248, and Brier skill score of 0.008. Lomustine had an ROC-AUC of 0.63, Brier score of 0.246, and Brier skill score of 0.032.

Most models showed good calibration across the entire dataset, although the models tended to underforecast in samples with lower probabilities of positive response (Figure 1, left). When models were back-applied to the whole dataset for each drug, the ROC-AUC was universally above 0.95, indicating high resolution (Figure 1, right). Figure 2 indicates high sharpness in the drug response forecasts as well. Finally, back-calculated Brier scores were also similarly low, with all models below 0.095.

3.4 | Predictive scores accurately distinguish between positive and negative responses among different immunophenotypes and relapse status

We next assessed the model performance by analysing the distribution of the predictive scores, namely probabilities of positive response for each drug, among different biological subgroups. The results demonstrate a clear difference in the mean predictive score between the patients with positive response (complete response or partial response) and negative response (stable disease or progressive disease). *P* values were less than .001 for all drugs when the number of patients was larger than 10 (Figure 2, left). Our models performed well for both B- and T-cell subtypes; distributions were the most dichotomized for rabacfosadine. The same level of significance was observed when comparing the distribution and mean predictive scores of naïve and relapsed patients for each drug (Figure 2, right). Notably, the dichotomization was more pronounced among the relapsed patients.

3.5 | Time to complete response is significantly better for patients with higher prediction scores

We also analysed time to complete response for each model when samples were dichotomized into patients with predictive scores >50% and patients with predictive scores <50% (Figure 3). The doxorubicin,

vincristine and cyclophosphamide models all showed significant differences in time to complete responses (log-rank $P < .05$), with patients with higher scores responding more quickly. The rabacfosadine and lomustine models did not show significant differences in time to complete response (log-rank $P = .2$ for rabacfosadine and $.3$ for lomustine). However, we note that both the rabacfosadine and lomustine datasets contained significantly smaller numbers of patients than the other three drugs.

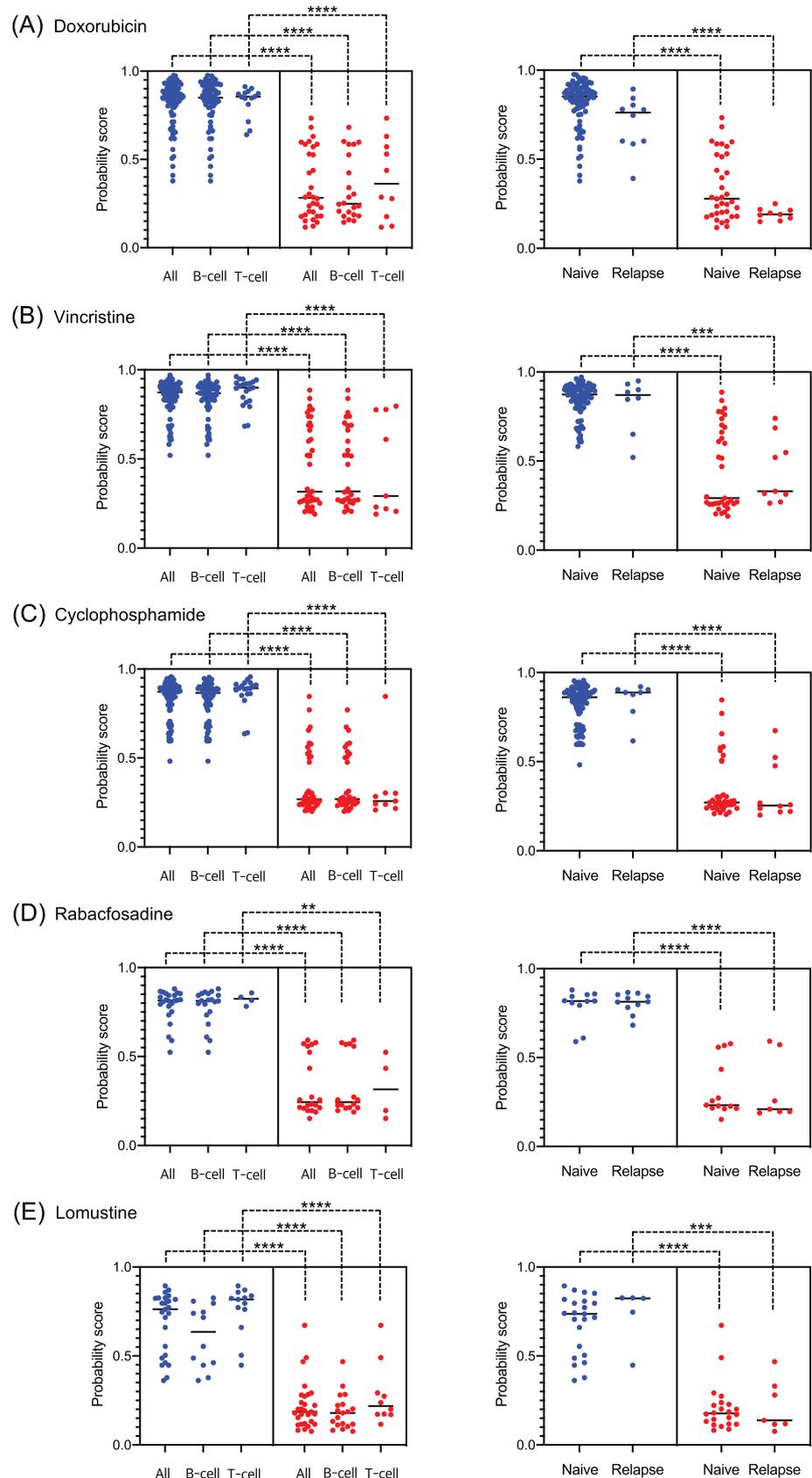
4 | DISCUSSION

In this study, we showed that drug sensitivity data and flow cytometry data acquired from primary canine lymphoma samples were sufficient to generate accurate probabilistic models of positive responses to individual chemotherapeutic agents. This finding addresses the current need for predictive models of treatment response to minimize the treatment burden in canine lymphoma. Our modelling approach features a variety of strengths, including the ability to informatively predict the likelihood of positive treatment response in individual canine lymphoma patients as well as close association with improved time to complete response in patients with positive predictions.

We modelled *in vivo* drug responses to five common chemotherapeutic agents that have direct cytotoxic effects in canine lymphoma cells: doxorubicin, vincristine, cyclophosphamide, lomustine, and rabacfosadine.^{1,6} Model performance was evaluated with a variety of measures, including ROC-AUC, representing a model's ability to discriminate between positive and negative responses; Brier score, a proper scoring function that represents the error between probabilistic predictions and actual outcomes; and Brier skill score, which compares the Brier score for a drug response model to the prevalence of positive responses in our data set. Test set ROC-AUC was above 0.6 for all drugs, and overall ROC-AUC was above 0.95 for all drugs, indicating good discrimination between positive and negative responses in our data set.³⁹ Brier scores varied between drugs, but because Brier score is dependent on prevalence of positive responses, we also evaluated Brier skill score to normalize all Brier scores to the prevalence of positive responses for a given drug in our data set. Brier skill scores varied, but all models showed positive Brier skill scores, indicating improved predictive value over positive response prevalence alone.

Probabilistic predictions of treatment response and survival in human oncology represent powerful tools to improve patient outcomes and have shown success in a wide variety of diseases.⁴⁰⁻⁴² We have applied the concept of probabilistic treatment response prediction to canine lymphoma. However, the identification of relevant features is a challenge in any biological model. Direct measurements of treatment response in primary tumour samples are almost certainly relevant to treatment response in patients.¹⁸ However, *ex vivo* assessments of chemosensitivity may not fully capture the behaviour of drugs or cells *in vivo*, which may render these measures insufficient to fully predict individual treatment responses alone. Because many immunophenotypic subtypes with diverse treatment response profiles have been described in canine lymphoma,⁴³ we sought to use

FIGURE 2 Discrimination between positive and negative scores among patient subpopulations. Comparison of the predictive probability scores for A, doxorubicin; B, vincristine; C, cyclophosphamide; D, rabacfosadine and E, lomustine across the patient subgroups (B- vs T-cell patients [left] and naïve vs relapse patients [right]) showing response positive responses (blue; complete response or partial response) and negative responses (red; stable disease or progressive disease). Lines indicate unpaired *t*-tests, and asterisks represent significance levels (*****P* < .0001; ****P* < .001; ***P* < .01) [Colour figure can be viewed at wileyonlinelibrary.com]



common immunophenotypic markers as additional variables for our models.

Although we used the same immunophenotypic markers for each drug model in this study, response prediction values varied

significantly between drugs for each patient, which implies significant differences in variable importance. Drug sensitivity parameters were highly important for doxorubicin, vincristine, and rabacfosadine models, whereas flow cytometry variables were more important in the

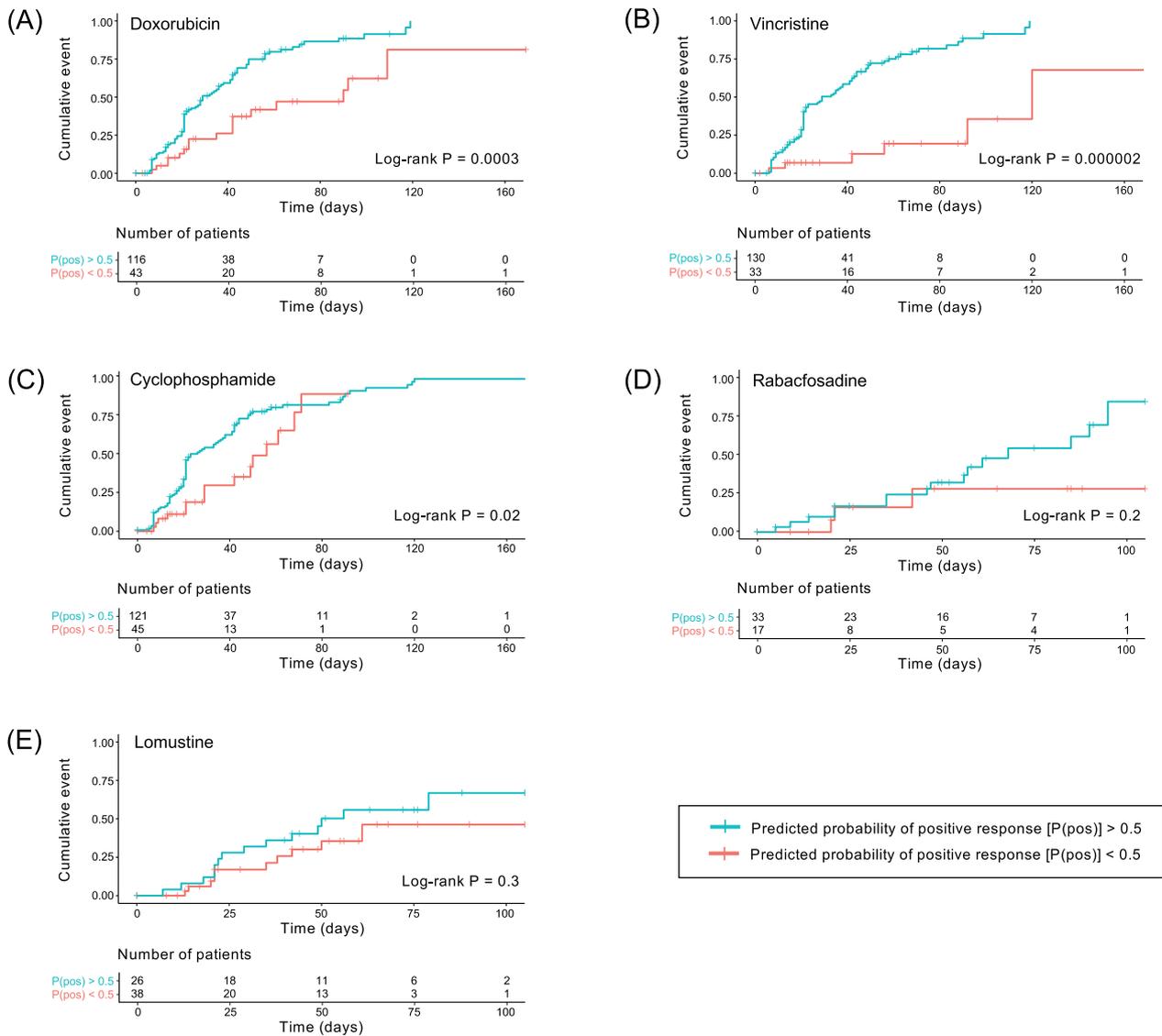


FIGURE 3 Cumulative incidence curves of time to complete response. All patients received at least one dose of A, doxorubicin; B, vincristine; C, cyclophosphamide; D, rabacfosadine and E, lomustine after the ex vivo chemosensitivity assay sampling date. Patient groups were dichotomized at 50% predicted probability of positive response. The bottom tables show number of patients who have not achieved complete response or been censored at a given time point [Colour figure can be viewed at wileyonlinelibrary.com]

cyclophosphamide and lomustine models. However, the importance of IC_{50} vs maximum cytotoxicity varied depending on the model in question. Percentage of CD21+ cells was also commonly in the top three most important variables, and it was particularly important for the cyclophosphamide and lomustine models. Although our cyclophosphamide model was optimized by using AUC of the drug response curve rather than IC_{50} and maximum cytotoxicity as variables, that variable was still only of moderate importance in the final random forest. This may indicate that the AUC-based cyclophosphamide model's superior predictive accuracy was related to a reduction of treatment-response variables among the total variable set rather than any predictive superiority of AUC over IC_{50} and maximum cytotoxicity. This difference seen in the cyclophosphamide model may also be related to our response annotation strategy since

cyclophosphamide frequently occurs later in a variety of combination chemotherapy regimens (eg, after doses of doxorubicin and vincristine in CHOP).

The cases that were modelled in this study were broadly representative of the cases seen in community practice, with similar distributions of immunophenotypes and disease stage at presentation.¹ We do note some differences between each individual drug cohort, however. Most notably, patients treated with lomustine were more likely to have T-cell disease (40% vs 16%-20% in the other treatment cohorts; Table 1), which is representative of lomustine-based regimens being a common alternative to CHOP chemotherapy in this immunophenotype.^{7,44,45} Patients treated with rabacfosadine in our study were also more likely to have stage 4 or higher disease (36% vs 19%-24% in other treatment cohorts; Table 1). This prevalence of

high-stage disease in our rabacfosadine cohort translated to a positive response rate lower than that reported in clinical trials (54% vs 74%-87%).^{6,9}

A core assumption made in this study is the decision to ascribe responses to individual drugs, even in the case of multi-agent chemotherapies. Although we recognize that there are likely significant interactions between drugs in multi-agent regimens, the strength of these synergistic or antagonistic interactions remains to be fully understood.^{46,47} We have attempted to mitigate the effects of this assumption using a confidence weighting approach that is heavily influenced by rapid changes in response status, whether positive or negative. It is difficult to derive any meaningful information from sustained responses across multiple drugs in a chemotherapy regimen (ie, there is little meaningful information regarding response to cyclophosphamide if a patient is in complete response, receives cyclophosphamide, and remains in complete response), but the combination of weekly or biweekly response assessments and a confidence weighting scale does allow our method to capture any sudden changes in response status, such as a patient with complete response to CHOP after receiving vincristine and doxorubicin suddenly experiencing progressive disease after receiving cyclophosphamide. Despite this, we do note that it is especially difficult to fully dissect the role of individual therapeutic agents in survival analyses for patients treated with multi-agent regimens.

Although our modelling approach reliably distinguished between positive and negative responses and showed clear advantages in time to complete response for drugs with larger sample sizes, we note that there were differences between test set performance depending on the drug in question. This may indicate that drugs with different mechanisms of action necessitate different measures of treatment response or secondary model variables or it may be related to a confounding factor between drugs that are commonly combined (eg, cyclophosphamide is often administered in the second week of a CHOP protocol, which may influence *in vivo* drug sensitivity in a way that is not captured by individual drug sensitivity measurements). The incorporation of additional variables, such as genomic sequencing data or interaction terms between combination therapies, into our models may increase predictive performance. For example, certain breed-specific mutations, such as MDR1, can have significant impacts on chemotherapy efficacy and patient outcomes.⁴⁸ These data are not currently captured in our models, and their addition may enhance the predictive accuracy of the models presented here. Furthermore, the models of rabacfosadine and lomustine showed relatively good performance by a variety of metrics and clearly dichotomized positive and negative treatment responses in all the tested subgroups, but there was no significant difference in time to complete response. This observation may be the result of relatively small sample sizes for these drugs in our study, and more patients may be necessary to completely model treatment responses in these patients.

Our results show that drug sensitivity parameters and common flow cytometry markers used for immunophenotyping are sufficient to develop probabilistic models of chemotherapy response in canine lymphoma. It is highly likely that larger sample sizes and the inclusion

of other variables may enhance the predictive power of the models described here. We believe that these and other predictive models of treatment outcome may empower veterinarians to make personalized therapy recommendations for canine lymphoma, which may lead to reduced treatment burden and increased survival.

ACKNOWLEDGEMENTS

The rabacfosadine used in this study was generously provided by Steven Roy (VetDC). We would also like to thank Tariq Shah (Oncologize) for his assistance with sample enrollment. We also thank the following doctors for providing the samples used in this study: Dr. Chelsea Tripp (Bridge Animal Referral Center, WA), Dr. Kathy Mitchener (Southwind Animal Hospital, TN), Dr. Conor McNeill (Hope Advanced Veterinary Center, VA), Dr. Kimberly Freeman (Veterinary Cancer and Surgery Specialists, OR), Dr. Rebecca Regan (SAGE Centers, CA), Dr. Rebecca Newman (BluePearl Pittsburgh Veterinary Specialty and Emergency Center, PA), Dr. Jennifer Baez (Center for Animal Referral and Emergency Services, PA), Dr. Kai-Biu Shiu (VCA Emergency and Specialty Center, WI), Dr. Heidi Ward (Gulf Coast Veterinary Oncology, FL), Dr. Stephen Atwater (VCA Encina Veterinary Medical Center, CA), Dr. Julie Bulman-Fleming (Veterinary Cancer Group of Orange County, CA), Dr. Krystal Harris (Central Texas Veterinary Specialty and Emergency Hospital, TX), Dr. Luminita Sarbu (Veterinary Oncology Center, WA), Dr. Kevin Choy (BluePearl Seattle Veterinary Specialists, WA), Dr. Michael Kiselow (SAGE Centers, CA), Dr. Sami Al-Nadaf (Pet Specialists of Monterey, CA), Dr. Carrie DeRegis (Pieper Memorial Veterinary Center, CT), Dr. Sharon Shor (BluePearl Veterinary Partners Tacoma, WA), Dr. M.J. Hamilton (Private Veterinary Specialties, NJ), Dr. Matt Dowling (VCA Northwest Veterinary Specialists, OR), Dr. Christine Swanson (BluePearl Grand Rapids, MI), Dr. Bridget Urie (BluePearl Pittsburgh Veterinary Specialty and Emergency Center, PA), Dr. Rachael Gaeta (BluePearl Veterinary Specialty Center, DE), Dr. Steve Shaw (SAGE Centers, CA), Dr. Emily Manor (VCA Advanced Veterinary Care Center, IN), Dr. Bonnie Smith (Veterinary Specialty Services, MO), Dr. Candace Pagano (Summit Veterinary Referral Center, WA), Dr. Cecilia Robot (VCA Emergency and Specialty Center, WI), Dr. Todd Erfourth (BluePearl Pittsburgh Veterinary Specialty and Emergency Center, PA), Dr. Amelia Keith (Southern Colorado Veterinary Internal Medicine, CO), Dr. Macon Miles (Southern Colorado Veterinary Internal Medicine, CO), Dr. Catie McDonald (VCA Northwest Veterinary Specialist, OR), Dr. Naoko Sogame (SAGE Centers, CA), Dr. Bryan Marker (SAGE Centers, CA), Dr. Richard Segaloff (VCA South Shore, MA), Dr. Erin Cletzer (Veterinary Specialty Services, MO), Dr. Breann Sommer (VCA Emergency and Specialty Center, WI), Dr. Martin Crawford-Jakubiak (SAGE Centers, CA), and Dr. Lillie Davis (Metropolitan Veterinary Associates, PA).

CONFLICT OF INTEREST

All authors except Chantal Tu, Wendi Velando Rankin and Douglas H. Thamm are employees of ImpriMed, Inc. Deanna Swartzfager is an independent contractor for ImpriMed, Inc. Wendi Velando Rankin and Douglas H. Thamm are shareholders and members of the scientific advisory board of ImpriMed, Inc.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are not publicly available due to privacy restrictions.

ORCID

Zach Bohannan  <https://orcid.org/0000-0003-2829-9668>

REFERENCES

- Zandvliet M. Canine lymphoma: a review. *Vet Q*. 2016;36(2):76-104. <https://doi.org/10.1080/01652176.2016.1152633>.
- Ponce F, Marchal T, Magnol JP, et al. A morphological study of 608 cases of canine malignant lymphoma in France with a focus on comparative similarities between canine and human lymphoma morphology. *Vet Pathol*. 2010;47(3):414-433. <https://doi.org/10.1177/0300985810363902>.
- Valli VE, San Myint M, Barthel A, et al. Classification of canine malignant lymphomas according to the World Health Organization criteria. *Vet Pathol*. 2011;48(1):198-211. <https://doi.org/10.1177/0300985810379428>.
- Rao S, Lana S, Eickhoff J, et al. Class II major histocompatibility complex expression and cell size independently predict survival in canine B-cell lymphoma. *J Vet Intern Med*. 2011;25(5):1097-1105. <https://doi.org/10.1111/j.1939-1676.2011.0767.x>.
- Curran K, Thamm DH. Retrospective analysis for treatment of naïve canine multicentric lymphoma with a 15-week, maintenance-free CHOP protocol. *Vet Comp Oncol*. 2016;14(S1):147-155. <https://doi.org/10.1111/vco.12163>.
- Saba CF, Clifford C, Burgess K, et al. Rabacfosadine for naïve canine intermediate to large cell lymphoma: efficacy and adverse event profile across three prospective clinical trials. *Vet Comp Oncol* Published online April 28. 2020;1-7. <https://doi.org/10.1111/vco.12605>.
- Saba CF, Thamm DH, Vail DM. Combination chemotherapy with L-asparaginase, lomustine, and prednisone for relapsed or refractory canine lymphoma. *J Vet Intern Med*. 2007;21(1):127-132. [https://doi.org/10.1892/0891-6640\(2007\)21\[127:ccwlla\]2.0.co;2](https://doi.org/10.1892/0891-6640(2007)21[127:ccwlla]2.0.co;2).
- Tanis J-B, Mason SL, Maddox TW, et al. Evaluation of a multi-agent chemotherapy protocol combining lomustine, procarbazine and prednisolone (LPP) for the treatment of relapsed canine non-Hodgkin high-grade lymphomas. *Vet Comp Oncol*. 2018;16(3):361-369. <https://doi.org/10.1111/vco.12387>.
- Saba CF, Vickery KR, Clifford CA, et al. Rabacfosadine for relapsed canine B-cell lymphoma: efficacy and adverse event profiles of 2 different doses. *Vet Comp Oncol*. 2018;16(1):E76-E82. <https://doi.org/10.1111/vco.12337>.
- Cawley JR, Wright ZM, Meleo K, et al. Concurrent use of rabacfosadine and L-asparaginase for relapsed or refractory multicentric lymphoma in dogs. *J Vet Intern Med*. 2020;34(2):882-889. <https://doi.org/10.1111/jvim.15723>.
- Deravi N, Berke O, Woods JP, Bienzle D. Specific immunotypes of canine T cell lymphoma are associated with different outcomes. *Vet Immunol Immunopathol*. 2017;191:5-13. <https://doi.org/10.1016/j.vetimm.2017.07.008>.
- Valli VE, Kass PH, Myint MS, Scott F. Canine lymphomas: association of classification type, disease stage, tumor subtype, mitotic rate, and treatment with survival. *Vet Pathol*. 2013;50(5):738-748. <https://doi.org/10.1177/0300985813478210>.
- Frismantas V, Dobay MP, Rinaldi A, et al. Ex vivo drug response profiling detects recurrent sensitivity patterns in drug-resistant acute lymphoblastic leukemia. *Blood*. 2017;129(11):e26-e37. <https://doi.org/10.1182/blood-2016-09-738070>.
- Dietrich S, Oleś M, Lu J, et al. Drug-perturbation-based stratification of blood cancer. *J Clin Invest*. 2018;128(1):427-445. <https://doi.org/10.1172/JCI93801>.
- Snijder B, Vladimer GI, Krall N, et al. Image-based ex-vivo drug screening for patients with aggressive haematological malignancies: interim results from a single-arm, open-label, pilot study. *Lancet Haematol*. 2017;4(12):e595-e606. [https://doi.org/10.1016/S2352-3026\(17\)30208-9](https://doi.org/10.1016/S2352-3026(17)30208-9).
- Kurtz SE, Eide CA, Kaempf A, et al. Molecularly targeted drug combinations demonstrate selective effectiveness for myeloid- and lymphoid-derived hematologic malignancies. *Proc Natl Acad Sci U S A*. 2017;114(36):E7554-E7563. <https://doi.org/10.1073/pnas.1703094114>.
- Blom K, Nygren P, Larsson R, Andersson CR. Predictive value of ex vivo chemosensitivity assays for individualized cancer chemotherapy: a meta-analysis. *SLAS Technol*. 2017;22(3):306-314. <https://doi.org/10.1177/2472630316686297>.
- Pawlak A, Obmińska-Mrukowicz B, Zbyryt I, Rapak A. In vitro drug sensitivity in canine lymphoma. *J Vet Res*. 2016;60(1):55-61. <https://doi.org/10.1515/jvetres-2016-0009>.
- Hernández P, Gorrochategui J, Primo D, et al. Drug discovery testing compounds in patient samples by automated flow cytometry. *SLAS Technol*. 2017;22(3):325-337. <https://doi.org/10.1177/2472630317700346>.
- Shipe ME, Deppen SA, Farjah F, Grogan EL. Developing prediction models for clinical use using logistic regression: an overview. *J Thorac Dis*. 2019;11(suppl 4):S574-S584. <https://doi.org/10.21037/jtd.2019.01.25>.
- Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform*. 2007;2:59-77.
- Deist TM, Dankers FJWM, Valdes G, et al. Machine learning algorithms for outcome prediction in (chemo)radiotherapy: an empirical comparison of classifiers. *Med Phys*. 2018;45(7):3449-3459. <https://doi.org/10.1002/mp.12967>.
- deAndrés-Galiana EJ, Fernández-Martínez JL, Luaces O, et al. On the prediction of Hodgkin lymphoma treatment response. *Clin Transl Oncol*. 2015;17(8):612-619. <https://doi.org/10.1007/s12094-015-1285-z>.
- Biccler JL, Eloranta S, de Nully Brown P, et al. Optimizing outcome prediction in diffuse large B-cell lymphoma by use of machine learning and Nationwide lymphoma registries: a Nordic lymphoma group study. *JCO Clin Cancer Inform*. 2018;2(1):1-13. <https://doi.org/10.1200/CCI.18.00025>.
- Ali M, Aittokallio T. Machine learning and feature selection for drug response prediction in precision oncology applications. *Biophys Rev*. 2018;11(1):31-39. <https://doi.org/10.1007/s12551-018-0446-z>.
- Richter AN, Khoshgoftaar TM. A review of statistical and machine learning methods for modeling cancer risk using structured clinical data. *Artif Intell Med*. 2018;90:1-14. <https://doi.org/10.1016/j.artmed.2018.06.002>.
- Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. *Int J Med Inf*. 2008;77(2):81-97. <https://doi.org/10.1016/j.ijmedinf.2006.11.006>.
- Dankers FJWM, Traverso A, Wee L, Van Kuijk SMJ. Prediction modeling methodology. In: Kubben P, Dumontier M, Dekker A, eds. *Fundamentals of Clinical Data Science*. Cham, Switzerland: Springer; 2019. Accessed September 4, 2020. <http://www.ncbi.nlm.nih.gov/books/NBK543534/>.
- Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn*. 2006;63(1):3-42. <https://doi.org/10.1007/s10994-006-6226-1>.
- Wright MN, Ziegler A. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J Stat Softw*. 2017;77(1):1-17. <https://doi.org/10.18637/jss.v077.i01>.
- Pette M, Gold R, Pette DF, Hartung HP, Toyka KV. Mafosfamide induces DNA fragmentation and apoptosis in human T-lymphocytes. A possible mechanism of its immunosuppressive action. *Immunopharmacology*. 1995;30(1):59-69. [https://doi.org/10.1016/0162-3109\(95\)00005-e](https://doi.org/10.1016/0162-3109(95)00005-e).

32. Vail DM, Michels GM, Khanna C, Selting KA, London CA. Response evaluation criteria for peripheral nodal lymphoma in dogs (v1.0)—a veterinary cooperative oncology group (VCOG) consensus document. *Vet Comp Oncol*. 2010;8(1):28-37. <https://doi.org/10.1111/j.1476-5829.2009.00200.x>.
33. Jang IS, Neto EC, Guinney J, Friend SH, Margolin AA. Systematic assessment of analytical methods for drug sensitivity prediction from cancer cell line data. *Pac Symp Biocomput*. Published online. 2014;19:63-74.
34. Beretta L, Santaniello A. Nearest neighbor imputation algorithms: a critical evaluation. *BMC Med Inform Decis Mak*. 2016;16(suppl 3):74. <https://doi.org/10.1186/s12911-016-0318-z>.
35. Breiman L. Random Forests. *Mach Learn*. 2001;45(1):5-32. <https://doi.org/10.1023/A:1010933404324>.
36. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw*. 2008;28(1):1-26. <https://doi.org/10.18637/jss.v028.i05>.
37. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCr: visualizing classifier performance in R. *Bioinformatics*. 2005;21(20):3940-3941. <https://doi.org/10.1093/bioinformatics/bti623>.
38. Therneau TM, Grambsch PM, eds. *Modeling Survival Data: Extending the Cox Model*. New York, NY: Springer; 2000. Accessed July 19, 2020. <https://CRAN.R-project.org/package=survival>.
39. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin Chem*. 2008;54(1):17-23. <https://doi.org/10.1373/clinchem.2007.096529>.
40. Berlow NE, Rikhi R, Geltzeiler M, et al. Probabilistic modeling of personalized drug combinations from integrated chemical screen and molecular data in sarcoma. *BMC Cancer*. 2019;19(1):593. <https://doi.org/10.1186/s12885-019-5681-6>.
41. Comandante-Lou N, Khaliq M, Venkat D, Manikkam M, Fallahi-Sichani M. Phenotype-based probabilistic analysis of heterogeneous responses to cancer drugs and their combination efficacy. *PLoS Comput Biol*. 2020;16(2):e1007688. <https://doi.org/10.1371/journal.pcbi.1007688>.
42. McKenna MT, Weis JA, Brock A, Quaranta V, Yankeelov TE. Precision medicine with imprecise therapy: computational modeling for chemotherapy in breast cancer. *Transl Oncol*. 2018;11(3):732-742. <https://doi.org/10.1016/j.tranon.2018.03.009>.
43. Vail DM, Thamm DH, Liptak JM, eds. Hematopoietic tumors. *Withrow MacEwens Small Animal Clinical Oncology*. 6th Philadelphia, PA: Saunders. Published online. 2019;688-772. <https://doi.org/10.1016/B978-0-323-59496-7.00033-5>.
44. Brown PM, Tzannes S, Nguyen S, White J, Langova V. LOPP chemotherapy as a first-line treatment for dogs with T-cell lymphoma. *Vet Comp Oncol*. 2018;16(1):108-113. <https://doi.org/10.1111/vco.12318>.
45. Morgan E, O'Connell K, Thomson M, Griffin A. Canine T cell lymphoma treated with lomustine, vincristine, procarbazine, and prednisolone chemotherapy in 35 dogs. *Vet Comp Oncol*. 2018;16(4):622-629. <https://doi.org/10.1111/vco.12430>.
46. Palmer AC, Chidley C, Sorger PK. A curative combination cancer therapy achieves high fractional cell killing through low cross-resistance and drug additivity. *Elife*. 2019;8:e50036. <https://doi.org/10.7554/eLife.50036>.
47. Palmer AC, Sorger PK. Combination cancer therapy can confer benefit via patient-to-patient variability without drug additivity or synergy. *Cell*. 2017;171(7):1678-1691.e13. <https://doi.org/10.1016/j.cell.2017.11.009>.
48. Gramer I, Kessler M, Geyer J. Determination of MDR1 gene expression for prediction of chemotherapy tolerance and treatment outcome in dogs with lymphoma. *Vet Comp Oncol*. 2015;13(4):363-372. <https://doi.org/10.1111/vco.12051>.

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: Bohannan Z, Pudupakam RS, Koo J, et al. Predicting likelihood of in vivo chemotherapy response in canine lymphoma using ex vivo drug sensitivity and immunophenotyping data in a machine learning model. *Vet Comp Oncol*. 2021;19:160–171. <https://doi.org/10.1111/vco.12656>