A MACHINE LEARNING APPROACH TO ENHANCING FELINE INFECTIOUS PERITONITIS DIAGNOSIS

Dawn E Dunbar[a], Simon A Babayan[b], Diane D Addie[a], Margaret J Hosie[c] and William Weir[a]

[a]School of Veterinary Medicine, University of Glasgow, Glasgow, United Kingdom
[b]Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, United Kingdom
[c]MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

Abstract

Cats with feline infectious peritonitis (FIP) often demonstrate non-specific clinical signs which confound diagnosis and no single test can be used to confirm a diagnosis *ante mortem*. Currently, several tests need to be performed, with the results building a body of evidence that supports or refutes a diagnosis of FIP.

We applied machine learning to systematically analyse laboratory data generated as part of current diagnostic work-ups, with the aim of improving accuracy and reducing subjectivity. While humans can examine laboratory data, evaluate sets of parameters and provide interpretation, computers offer the capacity to evaluate high-dimensional datasets and recognise patterns instantaneously. In this project, we are exploiting these capabilities, applying a machine learning approach to improving FIP diagnostics, based on data generated from the current tests used for FIP diagnosis.

Laboratory data and clinical interpretations for a large number of suspected FIP cases (n=1,498) were randomly sampled to create ten subsets. Seven different machine learning algorithms were evaluated using these datasets. The models were trained according to expert clinical interpretations of the laboratory results and their accuracy was evaluated. A comparison of seven models revealed that the mean accuracy ranged from 95.4% (95% CI 93.6-97.2%) for a logistic regression model to 99.0% (95% CI 98.2-99.7%) for an XGBoost tree-based algorithm.

These results demonstrate that when machine learning is applied to high-quality laboratory datasets, samples from cats with a high index of suspicion of FIP can be accurately and efficiently identified. This approach, therefore, offers enormous potential in terms of standardising and improving diagnostics for this important disease of cats.

Introduction

Feline infectious peritonitis (FIP) is an important disease of felids. Disease is a consequence of an aberrant immune response to infection with feline coronavirus. FIP is described as a severe immune-mediated pyogranulomatous vasculitis or perivasculitis. Pathological changes observed include leakage of fluid into body cavities and commonly the development of pyogranulomatous lesions on internal organs. Together these two phenomena comprise the extremes of a continuum, with effusive disease at one end and non-effusive at the other. The broad spectrum of pathological changes leads to a gamut of changes in measures of the animals' haematological, biochemical and inflammatory status. Clinical presentation and signalment are helpful in establishing a list of differential diagnosis, which may include FIP. Currently, diagnosis of FIP involves building a body of evidence to support a diagnosis as there is no single diagnostic test pathognomonic for FIP. Signalment, clinical signs, serology, haematology and biochemistry are analysed for each case by an expert clinician and an interpretation is made, based upon experience and current diagnostic guidelines.

Complex diagnostic pathways such as those developed for FIP lend themselves to mathematical modelling and statistical analysis. Machine learning algorithms can process large amounts of data very rapidly and identify nuanced patterns in the data that may not be possible for a human to identify.

In this project we tested the hypothesis that machine learning algorithms can identify cats with a high likelihood of FIP in a series of cases as well as, if not better than, experienced infectious disease clinicians. Unfortunately, definitive *post mortem* diagnosis was not available for the retrospective dataset. Nevertheless, the results provide proof of concept for a future prospective study in which we will recruit patients during diagnostic work-up and perform routine follow-up and *post mortem* examinations where necessary.

Exploratory Analysis

The data we currently hold consist of 18 laboratory measured variables including those assessing pathogen exposure, anaemia, inflammatory and innate immune responses. Additional data include cat age, sex, breed details and limited information regarding clinical signs, if provided by the submitting clinician.

Figure 1 demonstrates correlations between 16 laboratory parameters. We observed that high anti-coronavirus antibody titres negatively correlated with A:G ratio and anaemia and positively correlated with AGP. As the cases from which these correlations are drawn comprise a mixture of FIP and non-FIP cases, cases in which the cats have diseases other than FIP confounds the results. If FIP cases alone were examined, we predict that stronger correlations would be detected.

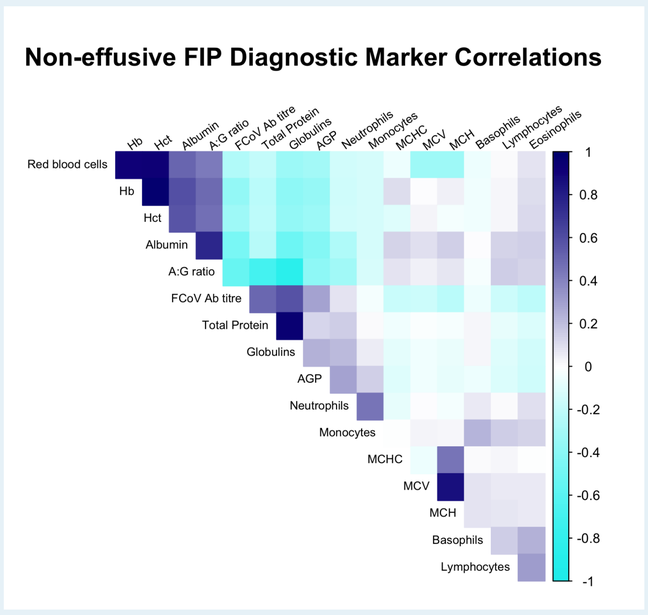

Figure. 1. Sixteen measurements are displayed with the level of correlation they show to each other depicted by the colour and intensity of colour in the corresponding squares. Two variables, normoblasts and band neutrophils, were omitted as the results were negligible.

We used principle component analysis (PCA) to reduce the dimensions of the dataset in order to visualise variation in the data. The first two principle components combined explain 44% of the variation in the data and by labelling the data points by colour we observed a decision boundary separating cases with and without FIP (Figure 2). This exploratory analysis indicated that machine learning algorithms would likely perform well with these data, given the acknowledged excellent performance of this unsupervised mathematical analysis. Random sampling was used to prepare ten small subsets of data for model building. Within the model generation there is a further step called 10-fold cross validation in which the model itself randomly subsets the groups of data and runs these subsets through the algorithm. These cross-validation folds were used to compute in-model accuracy.

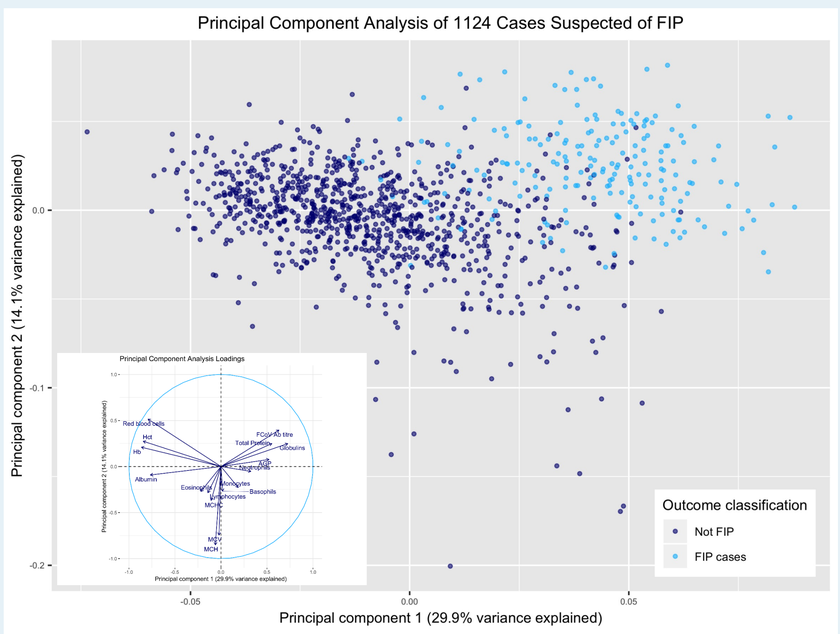

Figure. 2. Principal component analysis of 1,498 cases submitted to the laboratory for non-effusive FIP investigation. 18 variables were compressed to form new variables which explain the variation in the data in a more concise fashion. The data shown are plotted on principle components 1 and 2, where overall these two new variables explain 44% of the variation observed in the entire dataset. The inset on the lower left shows the influence of the original data on the new variables, the longer arrows show greater influence on the new principal components and the direction of the arrow shows the correlation with the original variable.

Results

We performed supervised machine learning using seven different types of mathematical algorithm. Ten models were built for each algorithm using a randomly sampled data subset. The accuracy measures from the models are shown in Figure 3. All models performed with greater than 85% accuracy, with the two tree based (Random Forest and Xtreme Gradient Boosting) algorithms having the highest mean accuracies. Accuracy measures alone cannot be used to assess the viability of a machine learning model. Rigorous independent testing of the models using data not seen before further demonstrates the reliability of the model to correctly predict the outcome of each case. Additionally, we were able to observe the details of each case, especially where the case has been misclassified by the algorithm; this allows us to interrogate the models and identify where improvements can be made. Performance data is shown in Table 1.

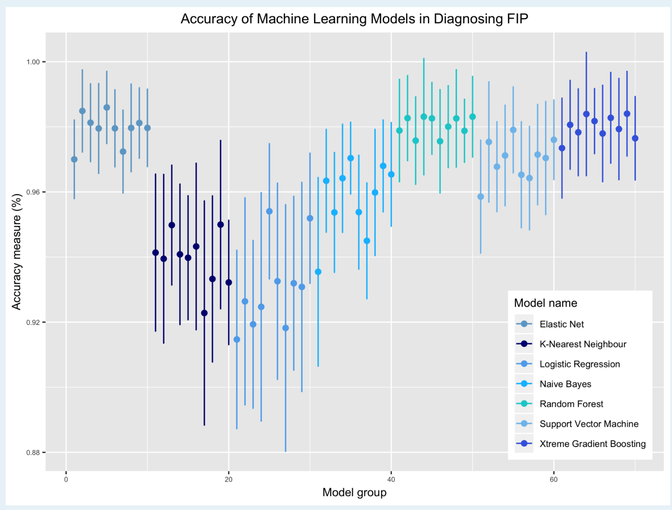

Figure. 3. Mean accuracy measures are displayed for 10 distinct subsets of FIP cases. Logistic regression is the least elastic of the algorithms and performed least well. Both Xtreme Gradient Boosting and Random Forest models performed very well, these are the most elastic, with Xtreme Gradient Boosting best overall.

	Elastic Net	K – Nearest Neighbour	Logistic Regression	Naïve Bayes	Random Forest	Support Vector Machine	Xtreme Gradient Boosting
Model Accuracy (%)	98.0	94.0	93.3	95.4	97.4	96.1	97.8
Test Accuracy (%)	97.8	93.3	94.4	97.8	96.7	98.9	100
True Positives	43	41	44	43	42	44	45
False Positives	2	4	1	2	3	1	0
False Negatives	0	2	4	0	0	0	0
True Negatives	45	43	44	45	45	45	45

Table.1. Illustration of performance of models on one subset (subset 6). Comparing models internal accuracy measure to accuracy using an unseen test data.

Discussion

Exploratory methods such as PCA, detailed earlier, provide evidence that our primary diagnostic work-up for cases suspected of having FIP contains valuable data. In addition, machine learning algorithms can be used to extract the maximum knowledge from the data and in turn enhance our diagnostic capabilities. The complex relationships observed in multi-dimensional datasets such as this can lead to the loss of information to human observers among the "noise". However, machine learning algorithms can assess all of the information represented in high-dimensional datasets and provide a usable interpretation without the loss of information. The models provide us with useful information which can be used to help improve the performance of the algorithms. Feature selection can be used to remove the variables that contribute least to the model's outcomes and improve the model performance. Two methods are available to select the features to remove. Figure 4 shows the variable importance (Gini index) for each variable in a KNN model; variables with low importance can be removed. Secondly, biological knowledge allows us to make informed decisions on variables to remove. For example, since six variables assess anaemia, these can be reduced to include the single most informative measure and redundant variables can be removed.

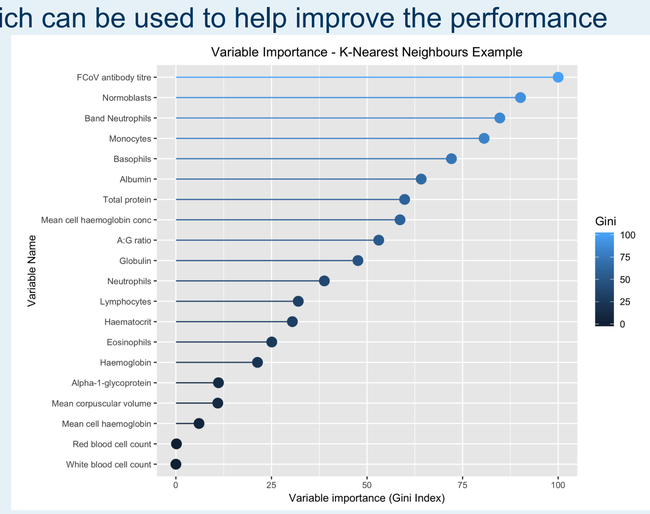

Figure.4. Variable importance, measured in Gini index. A:G = albumin:globulin ratio.

CONCLUSION & FUTURE WORK

Evidence-based medicine has been embraced in veterinary medicine for decades and the use of machine learning to enhance the diagnosis of disease in animals is an extension of that principle. We demonstrated that this approach performs well against our expert clinical diagnosis, however the lack of confirmation of disease status is a current limitation and so we propose to conduct a prospective longitudinal survival study over the next 3 years. We will collect data from diagnosis, follow-up and *post mortem* and use this information to build a working diagnostic tool, using a training-set of confirmed cases of FIP.

FUNDING

LUCA FUND for FIP Research

University of Glasgow | School of Veterinary Medicine

PRINTED BY SCIENCEPOSTERS .co.uk