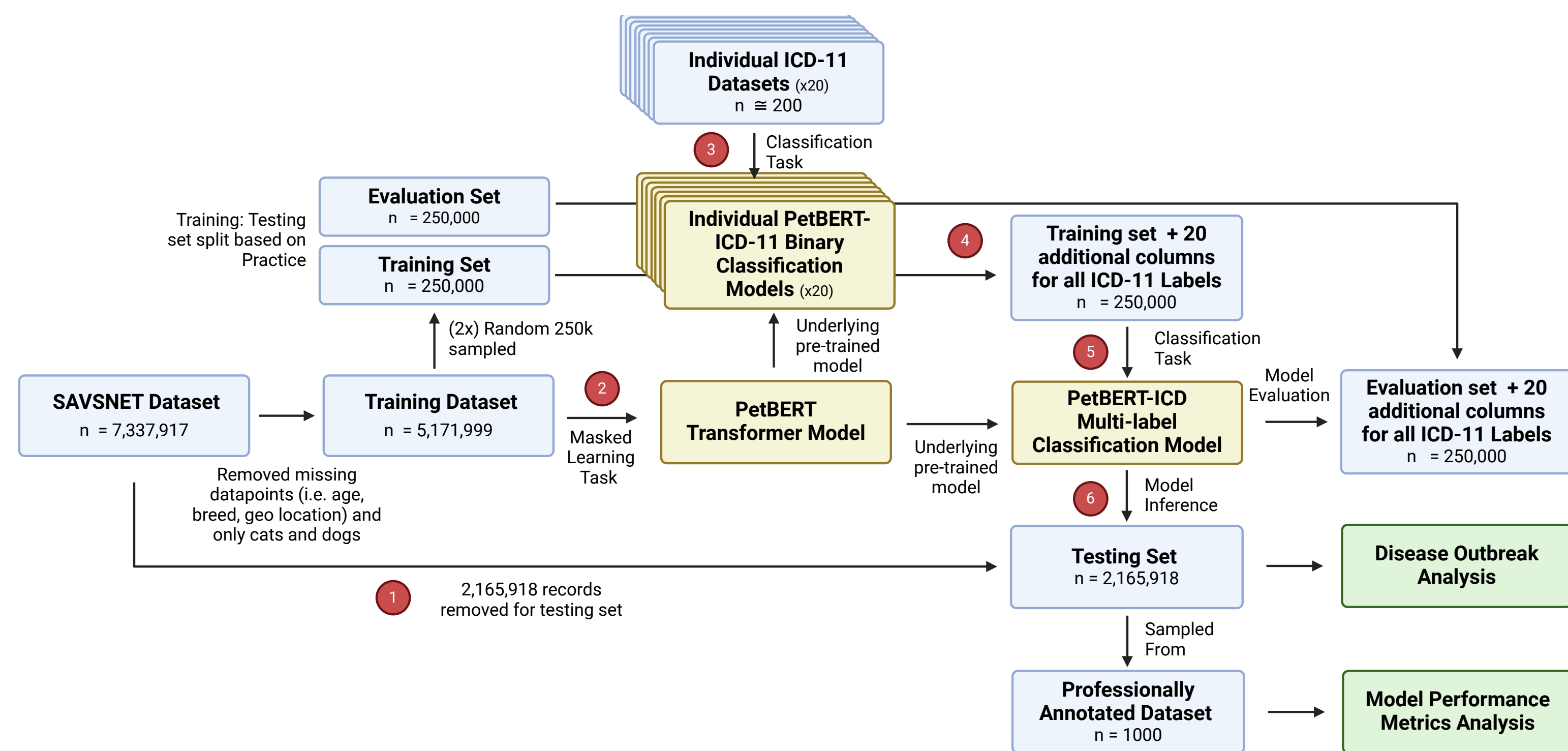


INTRODUCTION

First-opinion veterinary EHRs are the initial reporting point for clinical syndromes in animals and therefore have the potential to surface disease signals. These records however are often lack specific detail and can be fraught with inaccuracy even when annotated by clinicians. Automated annotation of records requires tools able to cope with negation, poor grammar, spelling and abbreviations. Large language models offer scope to overcome these challenges to label EHR for the presence of clinical syndromes for use in national disease surveillance. A recent outbreak of gastrointestinal disease in UK dogs allowed evaluation of the performance of such a model compared to clinician-derived annotations.

METHODS

MODEL TRAINING



1. SAVSNET Dataset (n~7.3M) is split into train (~5.1M) and test (~2.1M) sets
2. The training set enters into a fine-tuning Masked language modelling (MLM) and next sentence prediction (NSP) task to create the underlying pre-trained language model, PetBERT
3. 20 binary classification models are trained on each of high-level ICD-11 codings based on human-annotated samples from the training set
4. 20 Binary classifiers are applied to a random (x2) 250k set as a one-hot encoding (each narrative has 20 columns)
5. PetBERT-ICD multi-label classifier is trained on 250k dataset using PetBERT
6. PetBERT-ICD model is applied to the testing set for downstream tasks

OUTBREAK DETECTION

$$q_{i,t} = \frac{1}{M} \sum_{m=1}^M I(S_{i,t}^{(m)} > l)$$

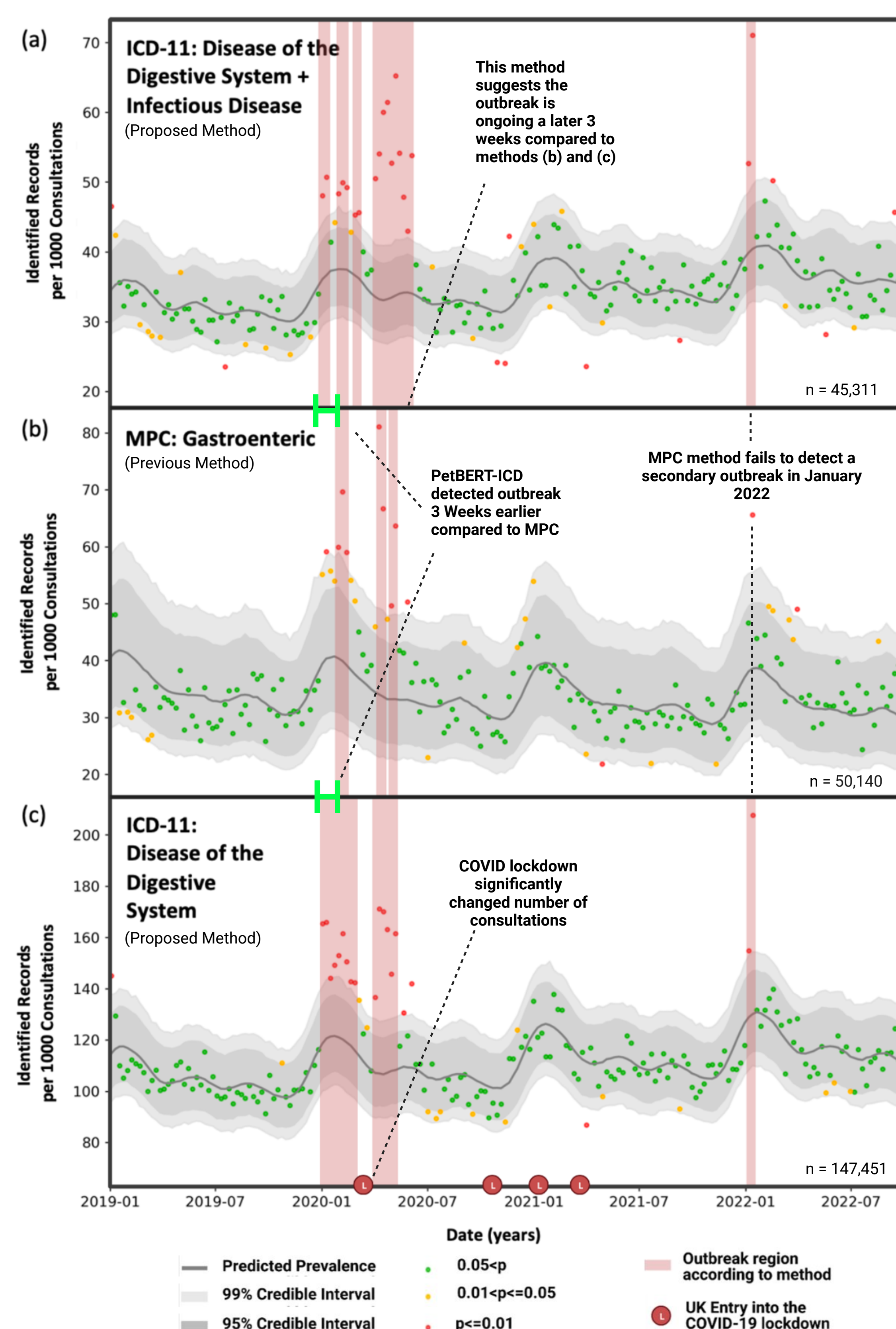
- **Response Variable:** Binary variable (1: target, 0: other cases)
- **Bayesian Inference:** fit spatiotemporal mixed effects regression model.
- **Fixed Effect:** Linear regression to model predictors.
- **Random Effect:** Considers unexplained changes using stochastic process ($S_{i,t}$).
- **Threshold (l):** Defined a value indicating potential outbreak
- **Predictive Probabilities (q):** Calculated on PetBERT-ICD output
- **Markov Chain Monte Carlo:** Generate posterior samples ($S_{(1)it}, S_{(2)it}, \dots, S_{(M)it}$)
- **Posterior Samples:** Uncertainty quantification and inference about random effects (Hale et al. 2019).

RESULTS

DISEASE OUTBREAK: Canine Enteric Coronavirus outbreak 2020

- The SAVSNET group identified a severe disease outbreak in dogs caused by canine enteric coronavirus, with primary symptoms being **vomiting** and **severe diarrhoea** from December 2019 to March 2020 (Radford et. al 2021).
- The existing method used to identify the outbreak by SAVSNET involves appending a single 'main presenting complaints' (MPC) label to each consultation, selecting from a limited set of ten labels by the attending clinician.
- Clinician-assigned labels at point-of-care are historically prone to error (Approximately 39% of our records have an inappropriate MPC)

Possible Disease Outbreak = Two consecutive weeks where the values exceed the 99% confidence interval (Red areas show the outbreak period according to the respected method)



- PetBERT-ICD dual labelling system:** Multi-label nature permits for specifying multi-syndromic disease conditions to be picked up, providing a higher resolution of outbreak capture
- Previous MPC-based method:** collated by point-of-care clinician labels
- PetBERT-ICD single labelling system**

Key Takeaways:

- Both single and dual labelling systems by PetBERT-ICD outperforms MPC strategy by 3 weeks
- Single PetBERT-ICD found ~3x many records compared to MPC
- Confidence intervals around PetBERT-ICD are more robust compared to MPC
- A second possible outbreak in 2022 was not detected by MPC method



Example clinical narratives annotated by PetBERT-ICD

Example Narrative

owner no concerns. defecating, urinating, drinking, eating.no vomiting/diarrohea. clinical exam: nothing abnormal detected. Given tricat and felv vac booster.

. mass betwenn blade infested,possible cyst/wart.....ab given and remove if given too much problem. heart murmur increasing from grade 2 to grade 4/6.advice to start vetmedin (have a syncope few days ago). check in 1 month. Next appointment in 1 month.

History: diar over weekend but also lots of urine passed in house. Clinical Exam: bright, eating, no sign ofdehydration although dif to examine mouth, temp 101.2, tender abd, no blood on therm. Differential Diagnosis:? dietary indiscretion +/- UTI. Plan:req urine sample, start on abs and yumpro, hills id. Owners on hols~~~pay on return

Model Output

[02] Neoplasm
[11] Disease of the Circulatory System

[01] Certain infectious or parasitic diseases
[13] Diseases of the Digestive System
[16] Diseases of the genitourinary system

DATASET

8 Million Veterinary Consultation Notes since 2014
~1 Million new records per year

Date of Birth	Owner Postcode	Practice ID	Free-text Narrative based on events during consultation
Species	Breed	Consult Date	
Insurance Status	Neuter Status	MPC	

CONCLUSION

PetBERT is a large language model trained on over 500 million words from UK veterinary clinicians. PetBERT-ICD is the automated ICD-11 tool for veterinary free-text EHRs, achieving over 83% f1 accuracy across 20 labels. PetBERT-ICD can detect disease outbreaks upto three weeks earlier than current clinician-assigned (MPC) methods, providing a novel method for mass surveillance of the UK companion animal population and will now be continuously applied for proactive outbreak detection.

REFERENCES

1. Hale, A. C. et al. A real-time spatio-temporal syndromic surveillance system with application to small companion animals. Sci. Reports 2019 9:1 9, 1–14, DOI: 10.1038/s41598-019-53352-6 (2019).
2. Radford, A. D. et al. Outbreak of Severe Vomiting in Dogs Associated with a Canine Enteric Coronavirus, United Kingdom. Emerg. Infect. Dis. 27, 517, DOI: 10.3201/EID2702.202452 (2021)