Barking Up The Right Tree: Explainable Machine Learning for Predicting Pet Health Outcomes

Christian Debes
Spryfox GmbH
Darmstadt, Germany
christian.debes@spryfox.de

Eike Mentzendorff
Spryfox GmbH
Darmstadt, Germany
eike.mentzendorff@spryfox.de

Johannes Wowra
Spryfox GmbH
Darmstadt, Germany
johannes.wowra@spryfox.de

Aliya McCullough
Fetch, Inc.
New York City, USA
aliya.wilsonmccullough@fetchpet.com

Audrey Ruple
Virginia Tech
Blackburg, USA
aruple@vt.edu

Abstract—**Predicting pet health outcomes holds significant value for improving veterinary care. In this paper, we present a machine learning-based approach using LightGBM (LGBM) to predict 50 health outcomes in dogs. LGBM demonstrated strong predictive performance with an average AUC of 80.48% ± 0.61% while significantly reducing training time compared to other tree-based models. The reduced training time enabled rapid iteration and real-time adjustments with veterinary experts, facilitating improvements in feature representation and disease grouping. We analyze feature importance to provide insights into disease prediction patterns, focusing on thyroid disorders, Cushing's syndrome, injuries, and tooth abnormalities. The explainability of the model confirms known clinical relationships for chronic diseases and highlights complex feature interactions for multifactorial disease processes. These insights increase model transparency and enhance trust among veterinary researchers and practitioners.**

Index Terms—**Machine Learning, Disease Prediction, Explainable ML**

I. INTRODUCTION

Machine learning (ML) has emerged as a powerful tool in healthcare, with applications ranging from human disease diagnosis to treatment optimization and risk stratification. Veterinary medicine has started to benefit from similar advances [1], particularly in the prediction and early diagnosis of health conditions in companion animals [2]. Accurate prediction of pet health outcomes can lead to improved clinical decision-making, better resource allocation, and more targeted preventive care.

Explainable machine learning is particularly critical in health-related fields [3], [4]. Unlike traditional black-box models, interpretable models provide insights into why a certain prediction was made, allowing clinicians to evaluate the biological plausibility and reliability of the model's output. In veterinary medicine, understanding the relationship between clinical features (such as breed and age) and health outcomes can enhance the confidence of veterinarians and pet owners, supporting better treatment decisions. [5]

The increasing complexity of health datasets further underscores the importance of explainability. Pet health data is inherently multi-modal, combining demographic factors (age, breed, sex), environmental influences (climate, geographical area, access to care), and historical health records. Machine learning models need to account for these complex relationships while maintaining interpretability. Tree-based models, particularly gradient-boosting frameworks like LightGBM, have shown strong potential for handling structured data with mixed feature types [6]. LightGBM's ability to efficiently capture feature interactions makes it particularly well-suited for pet health outcome prediction.

The global pet care market was valued at 179.4 billion dollars in 2020 and is expected to grow to 241.1 billion dollars by 2026 [7]. As more families consider pets as part of the family unit, investment in pet healthcare is increasing. This trend is reflected in the growing pet insurance market, which is forecasted to expand from 4.5 billion to 16.8 billion dollars by 2030 [8]. Understanding and predicting pet health outcomes is therefore becoming increasingly valuable for improving veterinary care and optimizing insurance policies.

Predicting health outcomes in pets presents specific challenges. Disease presentation varies across breeds, environmental factors influence health patterns, and veterinary records are often incomplete or inconsistent. A robust model must account for these variations while providing interpretable insights into the underlying factors driving health outcomes.

To address these challenges, we explore the use of tree-based models, specifically LightGBM. LightGBM's leaf-wise growth strategy and histogram-based algorithm allow for faster training times without compromising predictive accuracy. We focus on interpreting LGBM's feature importance to provide biological and medical insights, linking predictive patterns to established veterinary knowledge — a key step in establishing trust among veterinary researchers, practitioners, insurers, and pet owners.

This paper is structured as follows. Section 2 describes the available dataset as well as all necessary preprocessing and

This work has been funded by Fetch Forward, a subsidiary of Fetch, Inc.

feature extraction steps. It also introduces the optimization problem, the target variables and the LGBM. In Section 3 the performance evaluation - both in terms of AUC as well as computational complexity - for LGBM in comparison with other tree-based methods is shown. A discussion on feature importances follows where we focus on four different disease outcomes: Thyroid disorder, Tooth abnormalities, Injuries and Cushing's Syndrome. The paper finishes with a conclusion and an outlook.

II. METHODS

A. Data and Preprocessing

This work is based on a unique real-world dataset provided by the pet insurance company Fetch Inc. Fetch Inc. has active policies in the US and Canada and over the course of the past 17 years has insured 897,650 dogs with 2,693,872 annual policies and 3,251,433 claims. Fetch's dataset provides a unique opportunity to analyze long-term health outcomes across a diverse dog population of dogs. It is a rich dataset that includes information on the dog's age, sex, neuter status, breed, detailed disease history and - through linkage with the postal code - information on the environment in which the dog is living, including climate data (average temperature, precipitation, number of very hot and very cold days per year) as well as human census data such as population density and median household income [9], [10]. The disease history data was collected through insurance claims submitted by dog owners and has been fully de-identified for privacy.

1) Initial Datasets: We combined two distinct datasets: a policy dataset and a claims dataset. The policy dataset contains information about pet insurance policies, including demographic details of the insured dogs (age, sex, breed), enrollment date, and coverage type. The claims dataset includes detailed medical records, such as diagnosis codes, treatment details, and associated costs. These datasets were merged based on a unique pet ID, allowing us to align demographic and medical information for each individual pet.

2) Preprocessing: In order to prepare the dataset for model training a number of preprocessing steps have been taken. First, only dogs insured before the age of 12 months were included to ensure a complete medical history. Second, only dogs with an insured time of at least 12 months were taken into consideration, the reason being that the model at its core predicts disease probabilities for the upcoming year. To train a model, at least one year of data thus needs to be available. Third, for dogs who were insured for multiple years a splitting into multiple training samples is performed to increase the training dataset. Finally, missing data were handled using a combination of mean and mode imputation, depending on the data type and distribution. Records with incomplete or contradictory information were excluded from the dataset. After preprocessing, 543,433 dogs with 1,616,766 policies and 1,765,432 claims are used for model training.

3) Feature Engineering and Transformation: The preprocessed and combined dataset was further used to generate target variables and meaningful input features for the model as described in the following

B. Target Variables

Our task at hand is to predict the occurrence of future diseases. In the raw dataset, diseases are coded via condition codes from which there are 1049, many of which are rare. In order to construct meaningful target variables, a grouping of condition codes into broader disease categories had to happen. This was done based on clinical similarity. Seventy-nine disease categories were generated. As some of these categories still had a too low number for reliable prediction only the 50 categories with the highest sample sizes were chosen for training. For these disease categories it is guaranteed that at least 5000 samples end up in the training process.

Breed-related features

The raw dataset includes more than 500 breeds, many of which are rare. Following the same methodology as in [2] we grouped breed based on genetic similarity [11]. Three breed groupings were performed and run in parallel. In addition to the breed/breed group we also used so-called breed characteristics which map variables such as trainability, demeanor, size or coat length to a breed. It is known that for dogs the actual breed and its characteristics is strongly related to the probability of having certain diseases. For example, degenerative joint diseases are known to be more prevalent in German Shepherds [12] whereas Osteosarcoma is more prevalent in some large breed dogs such as Irish Wolfhounds and Saint Bernards [13].

Disease-related features

It is known that the previous disease history has a strong influence on future diseases. To code the past disease history we consider the same disease grouping described in the target variable section above. One binary variable per disease category is constructed with one-hot-encoding to describe the presence or absence of that disease in the dogs history.

The feature set used for training is summarized in Table I.

C. LightGBM and Tree-Based Methods

LightGBM (LGBM) is a gradient-boosting framework based on decision trees. It is well-suited for structured data with mixed feature types, offering fast training and strong predictive performance. Unlike traditional tree-based models that grow level-wise, LGBM uses a leaf-wise growth strategy. In a level-wise approach, all leaves at a certain depth are expanded simultaneously, which can lead to redundancy and increased computation time. In contrast, LGBM expands only the leaf that reduces the loss function the most, resulting in faster convergence and better handling of large datasets.

Since diseases are not mutually exclusive, the task is framed as a set of 50 independent binary classification problems rather than a multi-class problem. Each disease is predicted independently using a separate binary classifier.

TABLE I
SUMMARY OF DATASET CHARACTERISTICS.

Dataset	Records	Key Variables
Policy Dataset	1,616,766 policies	Age, Breed Groups, Country, Environmental features Residential Features
Claims Dataset	1,765,432 claims	Target variables (50 disease groups)
Final Combined Dataset	2,894,787 rows	Combined Variables

The objective function for predicting multiple diseases simultaneously is defined as:

$$L(\theta) = \frac{1}{K} \sum_{k=1}^{K} \sum_{i=1}^{N_k} l(y_{i,k}, f(x_{i,k}; \theta)) + \Omega(f) \tag{1}$$

where K is the number of diseases, N_k is the number of samples for disease k, $y_{i,k}$ is the true label, $f(x_{i,k}; \theta)$ is the model prediction, and $\Omega(f)$ is the regularization term defined as:

$$\Omega(f) = \lambda \|w\|^2 + \gamma T \tag{2}$$

where w represents the leaf weights, T is the number of leaves, and λ, γ are regularization parameters.

Feature importance in LGBM is computed using the total reduction in the loss function caused by each feature split:

$$I_j = \sum_{t \in T_j} \Delta L_t \tag{3}$$

where I_j is the importance of feature j, T_j is the set of trees where feature j is used for splitting, and ΔL_t is the reduction in loss from the split.

III. RESULTS AND DISCUSSION

A. Model Performance

We evaluated the performance of LightGBM using five-fold cross-validation. The primary metric used for evaluation was the area under the receiver operating characteristic curve (AUC). LightGBM achieved an average AUC of 80.48% ± 0.61% across all disease categories.

A comparison of AUC values and training times for LightGBM and the baseline models (XGBoost, Random Forest and Naive Bayes) is provided in Table II. LightGBM exhibited similar predictive performance to XGBoost and Random Forest while significantly reducing training time.

TABLE II
COMPARISON OF MODEL PERFORMANCE.

Model	Mean AUC	Training Time (min)
LightGBM	80.48% ± 0.61%	5.01
XGBoost	80.18% ± 0.69%	68.98
Random Forest	73.76% ± 0.81%	20.31
Naive Bayes	70.76% ± 1.06%	2.71

The individual AUCs per disease group are shown in Table III. We observe a span of AUC values, ranging from diseases with high predictability and AUC values around 95% to diseases with lower predictability, such as injuries or intoxication.

B. Example Cases

We analyzed the feature importance plots for individual diseases to provide deeper insight into the model behavior.

TABLE III
SUMMARY OF TARGET VARIABLES.

Disease	AUC
Adrenal insufficiency	90,12
Anal gland disorders	78,05
Anxiety or phobia	79,4
Arthritis	90,17
Behavioral disorders	82,5
Blood cancers	87,16
Cancerous tumors	88,96
Conformational disorders related to the skeleton	88,61
Cushing's syndrome	94,69
Diabetes	94,19
Digestive disorders	76,65
Disc diseases	88,51
Ear disorders	77,99
Eye abnormalities	86,02
Eyelid abnormalities	87,48
Foreign body ingestion	75,08
Gait abnormalities	73,4
Gastrointestinal disorders	75,82
Gastrointestinal nervous system disorders	71,53
Heart disorders	85,2
Immune disorders	78,2
Infectious diseases	73,65
Inflammation	74,32
Inflammation of the eyes	79,4
Injuries	69,97
Internal parasites	81,47
Intoxication or poisoning	70,82
Itching	80,59
Kidney disorders	82,41
Knee injuries	83,83
Leg injuries	73,53
Lethargy	68,7
Liver disorders	84,65
Mass lesions	77,5
Pain disorders	77,3
Periodontal diseases	84,38
Preventive	77,48
Respiratory infections	75,47
Seizures	80,06
Skin disorders	82,77
Skin infections	77,87
Soft tissue injuries	72,96
Surgical	85,32
Thyroid disorders	89,35
Tooth abnormalities	71,22
Treatment	77,45
Unspecified allergies	85,65
Urinary incontinence	86,16
Urinary tract infections	76,97
Vomiting and diarrhea	72,87

1) Thyroid Disorders: Thyroid disorders, specifically hypothyroidism is a common endocrine conditions in dogs. It is particularly prevalent in certain breeds such as Golden Retrievers and Doberman Pinschers [14], [15]. On this category, LightGBM achieved an AUC of approximately 89%. As can be seen in Figure 1 the most important features were age, skin disorders, unspecified allergies, gait abnormalities, thyroid disorders and seizures. The model also identified an influence of breed, with Doberman Pinschers and Golden Retrievers showing higher probabilities for thyroid disorders, consistent with existing veterinary findings.

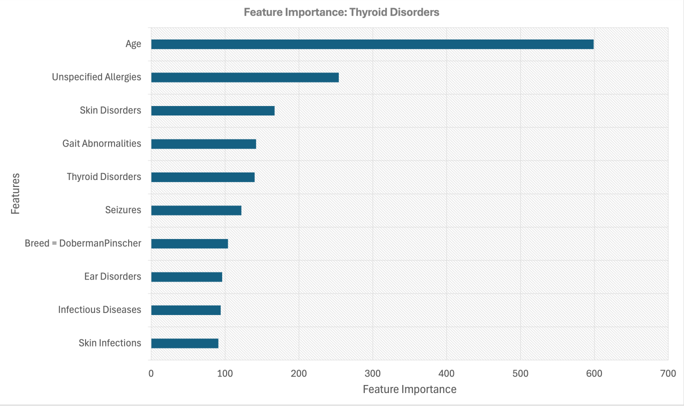

Fig. 1. Feature importance plot for Thyroid Disorders.

2) Cushing's Syndrome: Cushing's syndrome, also known as hyperadrenocorticism, is caused by excessive production of cortisol. It can lead to symptoms such as increased thirst, weight gain, and skin issues. LightGBM achieved an AUC greater than 90% for Cushing's syndrome. Age was the dominant predictive feature, followed by pre-existing Cushing's syndrome, thyroid disorders and urinary tract infection. The predictive pattern aligns with known clinical manifestations of Cushing's syndrome [16].

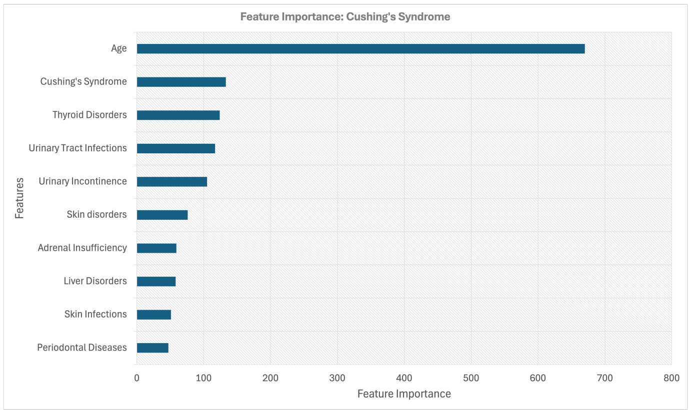

Fig. 2. Feature importance plot for Cushing's Syndrome.

3) Injuries: Injury prediction proved to be more challenging, with an AUC of approximately 69%. Age was the only significant predictor, while other demographic and medical features showed low importance. The low predictive performance for injuries reflects the stochastic nature of injury events and the absence of clear clinical predictors.

4) Tooth Abnormalities: Tooth abnormalities, including malocclusions, enamel defects, and periodontal disease, are influenced by both genetic and environmental factors. Dental health is often linked to the overall health of the dog, extent of dental home care, characteristics of the diet fed and chewing behavior [17], [18]. Tooth abnormality prediction had a low AUC, consistent with the complexity and variability of dental issues. The top predictors were previous tooth abnormalities, injuries, vomiting, and diarrhea — but all had low individual contribution.

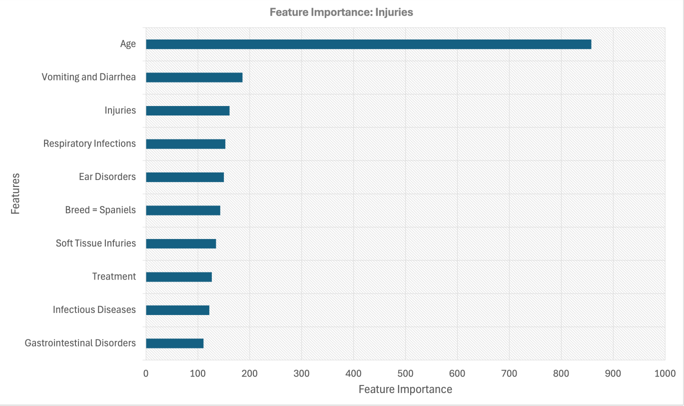

Fig. 3. Feature importance plot for Injuries.

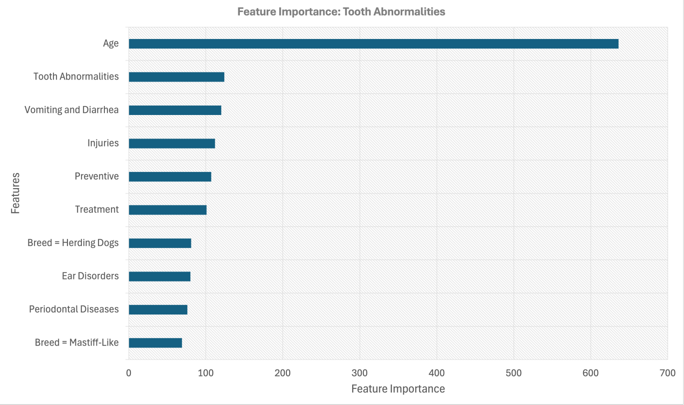

Fig. 4. Feature importance plot for Tooth Abnormalities.

C. *Explainability and Communication with Veterinary Experts*

The explainability of LightGBM's predictions proved valuable when communicating results with veterinary researchers and practitioners. Both the AUC and the feature importance plots provided meaningful insights that aligned with clinical expectations.

The AUC values helped to validate the model's ability to predict certain diseases. For instance, chronic diseases, such as diabetes and arthritis, were expected to be highly predictable once this disease is present — a pattern confirmed by the high AUC values for these conditions. Conversely, conditions like injuries, which are more influenced by random factors and environmental events, showed lower AUC values, reflecting the expected unpredictability of these outcomes.

Feature importance plots provided further clinical insight by revealing which factors the model considered most influential.

For chronic diseases like diabetes and arthritis, the plots showed age and the presence of the disease itself as dominant feature, which aligns with existing veterinary knowledge. For more complex diseases with multifactorial causes — such as gastrointestinal disorders or immune-related diseases — the feature importance plots highlighted the contribution of multiple factors, including co-morbidities and environmental influences.

This level of explainability increased trust in the model among veterinary professionals, as the patterns identified by the model were consistent with clinical experience. Understanding why the model made certain predictions allowed researchers and practitioners to validate its output and integrate it more confidently into their clinical decision-making processes.

IV. CONCLUSION

LGBM proved to be a powerful yet interpretable classifier for disease prediction in this dataset. The decrease in training time by approximately 91% allows for faster retraining, hyperparameter tuning, and real-time testing of new feature representations. This enabled rapid iteration with veterinary epidemiologists, allowing real-time adjustments to breed and disease grouping and feature representation.

The use of feature importance plots provided intuitive insights into the model's behavior, reinforcing trust among veterinary experts. Partial dependence plots offered additional interpretability by showing how individual features influenced predictions. Future work will focus on improving prediction for complex, multifactorial diseases and expanding the dataset to include a broader range of breeds and clinical conditions.

REFERENCES

[1] A. N. Paynter, M. D. Dunbar, K. E. Creevy, and A. Ruple, "Veterinary big data: when data goes to the dogs," *Animals*, vol. 11, no. 7, p. 1872, 2021.

[2] C. Debes, J. Wowra, S. Manzoor, and A. Ruple, "Predicting health outcomes in dogs using insurance claims data," *Scientific Reports*, vol. 13, no. 1, p. 9122, 2023.

[3] M. A. Ahmad, C. Eckert, and A. Teredesai, "Interpretable machine learning in healthcare," in *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics*, 2018, pp. 559–560.

[4] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang, "Xai—explainable artificial intelligence," *Science robotics*, vol. 4, no. 37, p. 7120, 2019.

[5] B. M. Vonholdt, J. P. Pollinger, K. E. Lohmueller, E. Han, H. G. Parker, P. Quignon, J. D. Degenhardt, A. R. Boyko, D. A. Earl, A. Auton *et al.*, "Genome-wide snp and haplotype analyses reveal a rich history underlying dog domestication," *Nature*, vol. 464, no. 7290, pp. 898–902, 2010.

[6] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems*, vol. 30, 2017.

[7] "StrategyR, pet care - global market trajectory & analytics," https://www.strategyr.com/market-report-pet-care-forecasts-global-industry-analysts-inc.asp, accessed: 2022-01-12.

[8] "Allied market research, pet insurance market by policy coverage (accident only, accident and illness, and others), by animal type (dogs, cats, others), by sales channel (agency, broker, and others): Global opportunity analysis and industry forecast, 2021–2030," https://www.alliedmarketresearch.com/pet-insurance-market, accessed: 2022-01-12.

[9] I. Papajová, J. Pipiková, J. Papaj, and A. Čižmár, "Parasitic contamination of urban and rural environments in the slovak republic: dog's excrements as a source," *Helminthologia*, vol. 51, no. 4, pp. 273–280, 2014.

[10] B. M. McCoy, L. Brassington, K. Jin, G. A. Dolby, S. Shrager, D. Collins, M. Dunbar, A. Ruple, and N. Snyder-Mackler, "Social determinants of health and disease in companion dogs: a cohort study from the dog aging project," *Evolution, Medicine, and Public Health*, vol. 11, no. 1, pp. 187–201, 2023.

[11] A. Vaysse, A. Ratnakumar, T. Derrien, E. Axelsson, G. Rosengren Pielberg, S. Sigurdsson, T. Fall, E. H. Seppälä, M. S. Hansen, C. T. Lawley *et al.*, "Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping," *PLoS genetics*, vol. 7, no. 10, p. e1002316, 2011.

[12] G. K. Smith, P. D. Mayhew, A. S. Kapatkin, P. J. McKelvie, F. S. Shofer, and T. P. Gregor, "Evaluation of risk factors for degenerative joint disease associated with hip dysplasia in german shepherd dogs, golden retrievers, labrador retrievers, and rottweilers," *Journal of the American Veterinary Medical Association*, vol. 219, no. 12, pp. 1719–1724, 2001.

[13] J. L. Kelsey, A. S. Moore, and T. Glickman, "Epidemiologic studies of risk factors for cancer in pet dogs," *Epidemiologic reviews*, vol. 20, no. 2, pp. 204–217, 1998.

[14] M. Dixon, S. Reid, and C. Mooney, "Epidemiological, clinical, haematological and biochemical characteristics of canine hypothyroidism," *Veterinary record*, vol. 145, no. 17, pp. 481–487, 1999.

[15] V. Travail, C. Fernandez Sanchez, J. M. Costo, N. Valentine, M. Conroy, V. Lee, D. Bouziopoulos, K. Bateman, E. Gatehouse, J. Cruzado-Perez *et al.*, "Assessment of the likelihood of hypothyroidism in dogs diagnosed with and treated for hypothyroidism at primary care practices: 102 cases (2016-2021)," *Journal of Veterinary Internal Medicine*, vol. 38, no. 2, pp. 931–941, 2024.

[16] I. Schofield, D. Brodbelt, S. Niessen, D. Church, R. Geddes, and D. O'Neill, "Frequency and risk factors for naturally occurring cushing's syndrome in dogs attending uk primary-care practices," *Journal of Small Animal Practice*, vol. 63, no. 4, pp. 265–274, 2022.

[17] E. I. Logan, "Dietary influences on periodontal health in dogs and cats," *Veterinary Clinics: Small Animal Practice*, vol. 36, no. 6, pp. 1385–1401, 2006.

[18] B. A. Niemiec, J. Gawor, S. Tang, A. Prem, and J. A. Krumbeck, "The bacteriome of the oral cavity in healthy dogs and dogs with periodontal disease," *American journal of veterinary research*, vol. 83, no. 1, pp. 50–58, 2022.