**Title and Hypothesis (1000 characters maximum)**

Application of machine learning methods to predict antimicrobial resistant UTIs in dogs.

Hypothesis

Patient risk factors and clinical data can be incorporated into a machine learning model that can predict the likelihood of an antimicrobial resistant organism as the causative agent in dogs with urinary tract infections. These models will have high diagnostic performance as determined by the area under the receiver operator characteristic curve > 0.9.

Specific Aims (2000 characters maximum)

A machine learning model is a tool trained with artificial intelligence algorithms to recognize patterns within datasets to accomplish tasks or make predictions about subjects of interest. The objective of this study is to train a model to predict the presence of antimicrobial resistance (AMR) in dogs with bacterial urinary tract infections (UTIs). Two models will be trained, the first to predict resistance to the first line treatment for bacterial lower UTIs, amoxicillin, and the second model will be trained to predict multidrug resistance, classified as resistance to three or more antimicrobial drug classes. Retrospective data from UC Davis’ Veterinary Teaching Hospital (VMTH) will be utilized in this study. Training data will be curated and supervised machine learning (ML) methods will be utilized to identify patterns in patient variables including signalment, prior antibiotic therapy, duration of hospitalization, urinary catheterization, and clinical data such as urinalysis to predict drug resistance classification.

Specific aim 1***:*** Develop a machine learning model to predict the presence of an amoxicillin resistant organism as the causative agent of a UTI.

Specific aim 2: Develop a machine learning model to predict the presence of a multiple drug resistant organism as the causative agent of a bacterial UTI.

Project Plan Significance (500 characters maximum)

With antimicrobial resistant infections becoming increasingly diagnosed in veterinary medicine, tools to rapidly identify these patients are needed. Identifying those at highest risk for RIs can guide clinicians to further diagnostics and therapy. It may also aid in identifying dogs with low risk of a RI and prevent overly broad antimicrobial coverage. The proposed project addresses the critical need for a system that informs clinicians of likelihood of resistance prior to initial therapy.

Innovation (500 characters maximum)

This project utilizes artificial intelligence as a novel method to predict antimicrobial resistance in animals with infections, using canine UTIs as a use case. Bacterial cultures determine antimicrobial resistance, but turnaround times can be up to one week. This research uses machine learning algorithms as a diagnostic tool to predict antimicrobial therapy before culture results are available.

**Approach. Must include Rationale and Methods, Potential Problems and Alternatives, Experimental Rigor (statistics, validation of reagents, sample size, etc). 8000 characters maximum.**

Rationale:

As antibiotics have improved over time, microbes have also developed mechanisms to extend their survival, becoming a pressing global matter. In 2019, it is estimated that 1.27 million people died as a direct result of antimicrobial resistant infections, surpassing deaths caused by HIV/AIDS (864,000) and malaria (643,000) (Thompson, 2022). In veterinary medicine, there is a looming concern regarding the rise of antibiotic resistance in companion animals, with drug-resistant urinary tract infections (UTIs) being a troubling condition (Dall, 2016). Urinary tract infections are a common diagnosis that warrants prescribing a course of antibiotic therapy. Published veterinary antimicrobial use guidelines recommend amoxicillin as a first line antimicrobial for empiric treatment of uncomplicated lower UTIs (Blondeau et al., 2019). However, there is evidence of increasing antimicrobial resistance, which may complicate the veterinarian’s decision on empiric antimicrobial use (Dall, 2016). One study surveying urine from both healthy dogs and dogs with cystitis for multidrug resistant (MDR) bacteria, defined as resistance to 3 or more drug classes, found that 65.9% (29/44) of dogs with positive urine cultures in the cystitis group contained MDR isolates (Camargo et al., 2019). Alarmingly, 25% (10/41) of the healthy control dogs had positive urine cultures and 7 of those 10 dogs also were carrying an MDR bacteria (Camargo et al., 2019). In that population, 54% were determined to be infected with bacteria resistant to amoxicillin while 25% had resistance to amoxicillin with clavulanic acid (Clavamox) (Camargo et al., 2019). A bacterial urine culture and subsequent antimicrobial susceptibility testing is the gold-standard test to determine antimicrobial susceptibility patterns of the infecting bacteria. However, the turnaround time is 3-7 days. This creates scenarios in which empiric antibiotic therapy is started prior to receiving the culture results and thus potentially undertreating or perpetuating resistance with inappropriately broad antimicrobial use. This proposed study aims to address the need for a tool that provides rapid and actionable information to direct clinical decision making regarding empiric antimicrobial choice while awaiting confirmatory urine culture results in dogs with UTIs. The objective is to develop machine learning models that can predict the presence of antimicrobial resistant UTIs using data that is available on the first day of presentation to the hospital. Certain patient features have been discovered that are significantly associated with the diagnosis of antimicrobial resistant UTIs. These include treatment with antimicrobials in the preceding three months, occurrence and length of hospitalization, urinary catheterization, systemic diseases such as diabetes mellitus, chronic renal failure, hyperthyroidism, and Cushing’s disease, and therapy with immunosuppressive medications such as corticosteroids, cyclosporine, or chemotherapy for neoplasia (Bernardin et al., 2013; Byron, 2011; Johnstone, 2019). The etiology of infection also alters the risk of antimicrobial resistance, as some organisms have higher rates of resistance than others. According to the American Veterinary Medical Association (2020), the pathogens of greatest concern are *Staphylococcus* spp, *Escherichia coli, Klebsiella* spp, *Enterococcus* spp, and *Pseudomonas* spp.

Machine learning (ML) has been successfully applied in veterinary medicine to make predictions and diagnoses from patient data. An artificial intelligence algorithm predicting the development of chronic kidney disease in cats was successful in doing so up to 2 year prior to a traditional blood work based diagnosis (Bradley et al., 2019). Further, a model has been found to be accurate, sensitive, and specific in screening dogs for Addison’s disease using patient blood work data (Gilor, Reagan, & Reagan, 2019). These examples highlight the utility of machine learning methods in detecting patterns associated with a disease process. Integration of patient features and machine learning models may allow stratification of patients into those that are high or low risk for the presence of an antimicrobial resistant UTI.

Methods:

Data that will be used to train the machine learning model will be obtained from UC Davis’ VMTH electronic records found on the Veterinary Medical & Administrative Computer System (VMACS). Medical records will be evaluated over a ten year time period, 2011 to 2021, in reverse chronological order. Dogs will be included in the study if a sterile urine specimen obtained by cystocentesis was positive for bacterial growth and corresponding antimicrobial susceptibility results are available for review. Dogs will be excluded if the attending clinician elected not to treat the bacteriuria with antimicrobials to exclude those with a diagnosis of subclinical bacteriuria. Additional inclusion criteria includes the availability of a urinalysis in the 24 hours before urine culture was performed. Only the first occurrence of a positive urine culture for an individual will be included to ensure independence. Bacterial species identification and quantification will be recorded. Antimicrobial susceptibility panels appropriate for the species (gram negative or gram positive) will be recorded and an interpretation of susceptible, resistant, intermediate, or no interpretation possible will be recorded based on the Clinical Laboratory Standards Institute (CLSI) bacterial cut points. The following patient data will be collected, if available, for each enrolled dog to be used as features within the ML models:

Age, sex, number of presentations to veterinarian in preceeding 30 days, hospitalized in preceeding 30 days (yes/no), duration of hospitalization in days, urine catheter placement in the preecding 30 days, length of urinary catheterization in days, underlying conditions (diabetes mellitus, hyperadrenocorticism, chronic kidney disease, hyperthyroidism), non-infection related urinary tract disease (congenital abnormality, incontinence, cystoliths, urinary implants), therapy with immunosuppressive drugs (corticosteroids, cyclosporine, or chemotherapy), antibiotic therapy in the prior three months (yes/no), duration of antibiotic therapy, class of antibiotic therapy. Data features related to urine analysis will also be collated including urine specific gravity, urine color, urine clarity, urine pH, urine protein concentration (mg/dL), urine glucose (mg/dL), number of white blood cells and red blood cells per high powered field, and bacterial morphology on sediment analysis. Location of infection within the urinary tract will be recorded as lower or upper.

**Specific aim 1:Establish a machine learning model to predict the presence of an amoxicillin resistant organism as the causative agent of a UTI.** This model will aim to identify patients at high risk of amoxicillin resistant UTIs. As the first line antimicrobial, it is critical to know when it would be appropriate to treat with this antimicrobial or escalate to higher tier antimicrobial. Each dog will be classified as amoxicillin resistant or susceptible based on CLSI cutpoints. Organisms with intrinsic resistance such as *Pseudomonas*, in which antimicrobial susceptibility is not performed will be recorded as resistant, as amoxicillin would not be an effective therapy for these dogs. A binary classification model will be trained and tested based on the data collected.

Specific aim 2: **Develop a machine learning model to predict the presence of a multiple drug resistant organism as the causative agent of a bacterial UTI.** A multi-class classification model will be used to predict patients at high risk for multidrug resistant UTIs. Each dog will be classified into three exclusive categories. 1. dog with MDR UTI 2. dog with UTI and presence of antimicrobial resistance (not MDR) and 3. dog without any resistance and UTI. MDR will be defined as resistance to three or more antimicrobial drug classes.

Experimental Rigor (statistics, validation of reagents, sample size, etc):

The estimated sample size needed for machine learning methods is 50 cases for each feature. With 22 patient features included, 1100 dogs will be enrolled. Based on VMTH data from 2019-2020, 918 urine cultures were performed and amoxicillin resistance was noted in 42% of cases. Thus, adequate patient data is available to achieve this sample size with relative equality between class sizes.

Supervised machine learning methods will be employed, with patient classes being assigned based on a medical record review as outlined above. One-hot encoding will be utilized to transform categorical values. A random subset of 80% of the data will be included in the training set. Python Scikit-learn library will be employed to compare performance of a panel of ML classification models (support vector machine, random forest, and gradient boosting) as compared to logistic regression. The model with the highest performance based on accuracy and a receiver operating characteristic (ROC) curve on the test set will be selected for each aim. Once the model is created, it will be tested on the remaining 20% of the data to determine the model performance. The sensitivity, specificity, and accuracy of the tuned model will be calculated for each aim. For each model, we will explore the feature importance to understand features (predictors) that drive the presence of AMR and MDR in dogs with UTI.

Potential Problems and Alternatives:

Given the retrospective nature of this analysis there is a risk of incomplete patient history, creating bias in the data set. This study is designed to include patient features that have been associated with the development of antimicrobial resistant infections, or are thought to be associated. There may be patient features that have not been assessed that may be strongly associated with antimicrobial resistance that were overlooked in this study, therefore, decreasing the robustness of the models. Furthermore, it is possible there may not be patterns in the patient data that are predictive of resistance and model accuracy may not be clinically relevant.

Trimmed down version 7638 characters

Rationale:

Over time, microbes have developed mechanisms to survive in the face of antimicrobials, becoming a pressing global matter. In 2019, 1.27 million people died because of antimicrobial resistant (AMR) infections, surpassing deaths caused by HIV and malaria (Thompson, 2022). In veterinary medicine, there is a looming concern regarding the rise of antibiotic resistance, with AMR UTIs being a troubling condition (Dall, 2016). Urinary tract infections are common and warrant antibiotic therapy. Veterinary antimicrobial use guidelines recommend amoxicillin as a first line antimicrobial for empiric treatment of uncomplicated lower UTIs (Weese, 2021). However, there is increasing AMR, which may complicate the veterinarian’s decision on empiric treatment (Dall, 2016). One study assessing healthy dogs and dogs with cystitis found that 65.9% of dogs with positive urine cultures in the cystitis group had multidrug resistant (MDR) infections with half being resistant to amoxicillin (Camargo, 2019). Alarmingly, 25% of the healthy control dogs were carrying bacteria in their urine, and 70% were MDR (Camargo, 2019). A bacterial urine culture and subsequent antimicrobial susceptibility testing is the gold-standard to determine antimicrobial susceptibility patterns of infecting bacteria. However, the turnaround time is 3-7 days. This creates scenarios in which empiric antibiotic therapy is started prior to receiving results and thus potentially undertreating or perpetuating resistance with inappropriately broad antimicrobial use. This study aims to create tools to provide rapid and actionable information to direct clinical decisions regarding empiric antimicrobial choice for dogs with UTIs. The objective is to develop ML models that predict the presence of AMR UTIs using data available at the point of care. Patient features have been discovered that are associated with the diagnosis of AMR UTIs. These features include treatment with antimicrobials in the preceding three months, occurrence and length of hospitalization, urinary catheterization, systemic diseases such as diabetes mellitus, chronic renal failure, hyperthyroidism, and Cushing’s disease, and therapy with immunosuppressive medications such as corticosteroids, cyclosporine, or chemotherapy for neoplasia (Bernardin, 2013; Byron, 2011; Johnstone, 2019). The etiology of infection also alters the risk of antimicrobial resistance, as some organisms have higher rates of resistance than others.

Machine learning (ML) has been successfully applied in veterinary medicine to make predictions and diagnoses from patient data. An artificial intelligence algorithm predicts chronic kidney disease in cats up to 2 years before traditional blood work (Bradley, 2019). Further, a ML model was trained and found to be highly accurate in screening dogs for Addison’s disease using patient blood work data (Reagan, 2019). These examples highlight the utility of ML methods in detecting patterns associated with a disease process. Integration of patient features and ML models may allow stratification of patients into those that are high or low risk for the presence of an AMR UTI.

Methods:

Data will be obtained from UC Davis’ VMTH electronic records. Medical records will be evaluated starting in Jan 2022 and working in reverse chronological order until a sample size of 1100 is achieved. Inclusion criteria includes:

·Positive bacterial urine culture and antimicrobial susceptibility, obtained by cystocentesis.

· UTI was not treated with antimicrobials by attending clinician, to exclude those with subclinical bacteriuria.

·Urinalysis performed within 24 hours of urine culture.

Only the first occurrence of a positive urine culture for an individual will be included to ensure independence. Bacterial identification, quantification and antimicrobial susceptibility will be recorded. An interpretation of susceptible, resistant, intermediate, or no interpretation possible will be recorded based on the Clinical Laboratory Standards Institute bacterial cut points.

The following patient data will be collected for each dog:

·Age

·Sex

·In the preceding 30 days:

o Number of veterinary visits

o Hospitalization (y/n)

o Duration of hospitalization

o Urine catheter placement

o Duration of urinary catheterization

·In preceding 90 days

o Antimicrobial therapy

o Duration of therapy

o Antimicrobial class

·Underlying conditions

o Diabetes mellitus

o Hyperadrenocorticism

o Chronic kidney disease

o Hyperthyroidism

o Other urinary tract disease (congenital abnormality, incontinence, cystoliths, urinary implants)

o Therapy with immunosuppressive drugs (corticosteroids, cyclosporine, or chemotherapy)

·Urinalysis (specific gravity, urine color, urine clarity, urine pH, urine protein concentration (mg/dL), urine glucose (mg/dL), white and red blood cells per high powered field, and bacterial morphology)

·Location of infection within the urinary tract

o Lower

o Upper

**Specific aim 1:Establish a ML model to predict the presence of an amoxicillin resistant organism as the causative agent of a UTI.** Dogs will be classified as amoxicillin resistant or susceptible. A binary classification model will be utilized to predict which dogs have amoxicillin resistant UTIs.

**Specific aim 2:** **Develop a ML model to predict the presence of a MDR organism as the causative agent of a UTI.** A multi-class classification model will be utilized to predict the level of resistance of a UTI. Each dog will be classified into three exclusive categories. 1. MDR isolate 2. presence of antimicrobial resistance, but not MDR and 3. Pan-susceptible isolate.

Experimental Rigor:

The estimated sample size needed for machine learning methods is 50 cases for each feature (Steyerberg, 2000). With 22 patient features included, 1100 dogs will be enrolled. Based on VMTH data from 2019-2020, 918 urine cultures were performed, and amoxicillin resistance was noted in 42% of cases. Thus, adequate patient data is available to achieve this sample size with relative equality between class sizes.

Supervised machine learning methods will be employed, with patient classes being assigned based on a medical record review as outlined above. One-hot encoding will be utilized to transform categorical values. A random subset of 80% of the data will be included in the training set. Python Scikit-learn library will be employed to compare performance of a panel of ML classification models (support vector machine, random forest, and gradient boosting) as compared to logistic regression. The model with the highest performance based on accuracy and a receiver operating characteristic curve on the test set will be selected for each aim. Once the model is created, it will be tested on the remaining 20% of the data to determine the model performance. The sensitivity, specificity, and accuracy of the tuned model will be calculated for each aim. For each model, we will explore the feature importance to understand features (predictors) that drive the presence of AMR and MDR in dogs with UTI.

Potential Problems and Alternatives:

Given the retrospective nature of this analysis there is a risk of incomplete patient history, creating bias in the data set. This study is designed to include patient features that have been associated with the development of AMR infections, or are thought to be associated. There may be patient features that have not been assessed that may be strongly associated with antimicrobial resistance that were overlooked in this study, therefore, decreasing the robustness of the models. Furthermore, it is possible there may not be patterns in the patient data that are predictive of resistance and model accuracy may not be clinically relevant.

**Citations**. Properly cite all sources indicated in your research proposal.

American Veterinary Medical Association. (2020). Antimicrobial-resistant pathogens affecting animal health. American veterinary medical association.

Bradley, R., Tagkopoulos, I., Kim, M., Kokkinos, Y., Panagiotakos, T., Kennedy, J., ... & Elliott, J. (2019). Predicting early risk of chronic kidney disease in cats using routine clinical laboratory tests and machine learning. Journal of veterinary internal medicine, 33(6), 2644-2656.

Byron, J. K. (2011). Resistant urinary tract infections. DVM360.

Dall, C. (2016). Antibiotic resistance in pets: an increasing problem. Center for infectious disease and research and policy.

Hernandez, J., Bota, D., Farbos, M., Bernardin, F., Ragetly, G., & Médaille, C. (2014). Risk factors for urinary tract infection with multiple drug-resistant Escherichia coli in cats. Journal of feline medicine and surgery, 16(2), 75-81.

Johnstone, T. (2020). A clinical approach to multidrug-resistant urinary tract infection and subclinical bacteriuria in dogs and cats. New Zealand Veterinary Journal, 68(2), 69-83.

Reagan, K. L., Reagan, B. A., & Gilor, C. (2020). Machine learning algorithm as a diagnostic tool for hypoadrenocorticism in dogs. Domestic Animal Endocrinology, 72, 106396.

Steyerberg, E. W., Eijkemans, M. J., Harrell Jr, F. E., & Habbema, J. D. F. (2000). Prognostic modeling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Statistics in medicine, 19(8), 1059-1079.

Thompson, T. (2022). The staggering death toll of drug-resistant bacteria. Nature.

Weese, J. S., Blondeau, J., Boothe, D., Guardabassi, L. G., Gumleyg, N., Papich, M., ... & Sykes, J. (2021). International Society for Companion Animal Infectious Diseases (ISCAID) guidelines for the diagnosis and management of bacterial urinary tract infections in dogs and cats. Journal of Japanese association of veterinary nephrology and urology, 13(1), 46-63.

Yamanaka, A. R., Hayakawa, A. T., Rocha, Í. S., Dutra, V., Souza, V. R., Cruz, J. N., ... & Nakazato, L. (2019). The occurrence of multidrug resistant bacteria in the urine of healthy dogs and dogs with cystitis. Animals, 9(12), 1087.