# CHEMICAL BIOMARKER PROFILES EXTRACTION FOR HONEYBEE PATHOGENS USING MACHINE LEARNING

**By**

**MUHAMMED MOYASAR ALAYOUBI**

Submitted to the Graduate School of Engineering and Natural Sciences

in partial fulfilment of

the requirements for the degree of Master of Science

Sabancı University

December 2021

CHEMICAL BIOMARKER PROFILES EXTRACTION FOR HONEYBEE PATHOGENS USING MACHINE LEARNING

APPROVED BY:



DATE OF APPROVAL: 17/12/2021

# ABSTRACT

## CHEMICAL BIOMARKER PROFILES EXTRACTION FOR HONEYBEE PATHOGENS USING MACHINE LEARNING

**MUHAMMED MOYASAR ALAYOUBI**

Computer Science, MSc Thesis, 2021

Thesis Supervisor: Asst. Prof. Christopher Mayack

Keywords: Biomarkers, chemical biomarkers, honeybee, Machine learning, pathogens, Metabolomics, varroa mite

In recent years, we have increased on our reliance upon honey bee pollination services yet bee health has been declining on a global scale. The decline in bee health is a complex multifactorial problem and it is caused by a number of interacting stressors. The stressors are mainly stemming from pesticide exposure, parasitic infections, poor nutrition, and loss of foraging habitat. However, how these stressors exactly interact to produce a synergistic decline in bee health remains elusive because previous studies have mainly focused on one or two stressors at a time using traditional experimental testing in the laboratory.

Here we utilize a systems biology approach that is a non-hypothesis data driven analysis. We integrate the exposome profile of 87 honey bee hives, sampled from rural to urban areas, with the abundance datasets of the 20 most common bee diseases to determine the specific interactions responsible for a decline in bee health. From this analysis, we have developed chemical biomarker libraries for 13 of the bee diseases that are able to predict whether a hive is infected or not. The biomarker libraries were validated using five different machine learning techniques that consistently demonstrated our chemical biomarker libraries can predict whether a hive is infected with a particular disease or not with roughly 85% accuracy, precision, sensitivity, selectivity, and recall. In addition, using a network analysis across the integrated datasets, we found that across the bee diseases there are five metabolite hubs that are suspected to be potential targets that are responsible for an increase in susceptibility of the honey bee to multiple infections or can explain how multiple infections lead to a synergistic decline in bee health mechanistically. Moreover, we identified a number of environmental pollutants that are highly toxic to humans, which are also associated with bee diseases and are linked to detoxification and oxidative stress response genes. Our findings suggest that not only can bees be used a bioindicators or sentinels for monitoring environmental quality for human

health, but the exposures themselves to the honey bees are likely to be a detriment to their health as well. These environmental exposures from polluted environments are likely another stressor that is negatively impacting bee health and their implications have yet to be fully recognized in the most recent decline in bee health.

From the systems biology analysis we provide chemical biomarkers that can be used as a possible rapid diagnostic tool such that beekeepers can change management practices to improve honey bee health before the colony collapses from parasitic infections. Novel stressors have been identified that are likely negatively impacting bee health and these are interacting with a multitude of exposures that are linked to an increase in disease prevalence in the honey bee hive. Collectively, our findings support the notion that the One Health paradigm is likely to be the most effective strategy for addressing the complexity of declining bee health and for improving it moving forward.

## ÖZET

MAKİNE ÖĞRENİMİNİ KULLANARAK BAL ARISI PATOJENLERİ İÇİN KİMYASAL BİYOMARKER PROFİLLERİNİN ÇIKARILMASI

**MUHAMMED MOYASAR ALAYOUBI**

Bilgisayar Bilimi , Yüksek Lisans Tezi, Haziran 2021

Tez Danışmanı: Dr.Öğr. Üyesi Christopher Mayack

Anahtar kelimeler: Biyobelirteçler, kimyasal biyobelirteçler, bal arısı, Makine öğrenimi, patojenler, Metabolomik, varroa akarı

Son yıllarda bal arısı tozlaşma hizmetlerine olan bağımlılığımız arttı, ancak arı sağlığı küresel ölçekte düşüyor. Arı sağlığındaki düşüş, karmaşık çok faktörlü bir sorundur ve bir dizi etkileşimli stres etkeninden kaynaklanır. Stresörler temel olarak pestisit maruziyetinden, paraziter enfeksiyonlardan, yetersiz beslenmeden ve yiyecek arama habitatının kaybından kaynaklanmaktadır. Bununla birlikte, bu stres etkenlerinin arı sağlığında sinerjik bir düşüş meydana getirmek için tam olarak nasıl etkileşime girdiği belirsizliğini koruyor çünkü önceki çalışmalar, laboratuvarda geleneksel deneysel testler kullanılarak aynı anda bir veya iki stres etkenine odaklanmıştı.

Burada, hipotez dışı veri odaklı bir analiz olan bir sistem biyolojisi yaklaşımı kullanıyoruz. Arı sağlığındaki düşüşten sorumlu spesifik etkileşimleri belirlemek için kırsal alanlardan kentsel alanlara örneklenen 87 bal arısı kovanının açıklayıcı profilini en yaygın 20 arı hastalığının bolluk veri setleriyle bütünleştiriyoruz. Bu analizden, bir kovanın enfekte olup olmadığını tahmin edebilen arı hastalıklarının 13'ü için kimyasal biyobelirteç kütüphaneleri geliştirdik. Biyobelirteç kitaplıkları, kimyasal biyobelirteç kitaplıklarımızın bir kovana belirli bir hastalık bulaşıp bulaşmadığını kabaca %85 doğruluk, kesinlik, duyarlılık, seçicilik ve geri çağırma ile tahmin edebildiğini tutarlı bir şekilde gösteren beş farklı makine öğrenme tekniği kullanılarak doğrulandı. Ek olarak, entegre veri setleri arasında bir ağ analizi kullanarak, arı hastalıkları boyunca, bal arısının birden fazla enfeksiyona duyarlılığının artmasından sorumlu olan veya nasıl olduğunu açıklayabilen potansiyel hedefler olduğundan şüphelenilen beş metabolit merkezi olduğunu bulduk. çoklu enfeksiyonlar, mekanik olarak arı sağlığında sinerjik bir düşüşe yol açar. Ayrıca, arı hastalıkları ile ilişkili ve detoksifikasyon ve oksidatif stres tepki genleriyle bağlantılı, insanlar için oldukça toksik olan bir dizi çevresel kirletici belirledik. Bulgularımız, arıların sadece insan sağlığı için çevre kalitesini izlemek için bir biyoindikatör veya gözcü olarak kullanılabileceğini değil, aynı zamanda bal arılarına maruz kalmanın da onların sağlığına zarar verme olasılığının yüksek olduğunu göstermektedir. Kirli ortamlardan kaynaklanan bu çevresel maruziyetler, muhtemelen arı sağlığını olumsuz etkileyen başka bir stres faktörüdür ve bunların etkileri, arı sağlığındaki en son düşüşte henüz tam olarak anlaşılmamıştır.

Sistem biyolojisi analizinden, olası bir hızlı teşhis aracı olarak kullanılabilecek kimyasal biyobelirteçler sağlıyoruz, öyle ki arıcılar, koloni parazit enfeksiyonlarından çökmeden

önce bal arısı sağlığını iyileştirmek için yönetim uygulamalarını değiştirebilirler. Arı sağlığını muhtemelen olumsuz yönde etkileyen yeni stres faktörleri tanımlanmıştır ve bunlar, bal arısı kovanındaki hastalık prevalansındaki artışla bağlantılı çok sayıda maruziyetle etkileşime girer. Toplu olarak, bulgularımız, Tek Sağlık paradigmasının, azalan arı sağlığının karmaşıklığını ele almak ve ilerlemeyi iyileştirmek için muhtemelen en etkili strateji olduğu fikrini desteklemektedir.

## ACKNOWLEDGEMENTS

First and foremost, I would want to praise and thank God, the Almighty, for his immeasurable gifts. , and who granted me the opportunity and the ability to learn, write and contribute to this work, truly my prayers, my sacrifice, my life, and my death, are all for Allah the Lord of all that Exists.

Apart from my contribution, the success of this thesis depends largely on the encouragement and guidance of many others.

I would like to show my greatest appreciation to Asst.Prof. Christopher Mayack as the thesis supervisor. Words absolutely betray me to express how truly grateful I am. Thank you does not begin to express my gratitude for his unwavering support and assistance. Every time I attend one of his meetings, I feel energized and inspired. The enormous kindness and patience that he showed me, reflects nothing but his great wisdom and pure heart. Without his encouragement and guidance this thesis would not have materialized.

Furthermore, I would like to extend a very special thanks to Asst.Prof. Berin Karaman Mayack her constant support and help not only helped to complete this work but also encouraged me to be better in other aspects of my personal life, she is a great teacher and a role model to me, this and any future success that I achieve will always carry her influence. I'm forever grateful to her.

I am also very grateful to my beloved family. They are the main reason and motivation that I was able to achieve anything in life. They always supported me in every decision I have ever made. Without their unique love, support, and efforts I wouldn't have achieved anything in life, Special thanks to my mother. Who unfortunately, lost her life during the completion of this work she had the biggest influence on any good aspects of myself. I pray that her soul is in heaven.

It would be impossible to list all their names, but several people also deserve my sincere and special thanks, all my professor and friends through the entirety of my academic journey have contributed into the completion of this work in their own special way from my professors and friends in Kadir Has to my professors and friends in Sabanci university. To all of them I say thank you from the bottom of my heart.

***In memory of my beloved Mother…***

## TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| ABPV | Acute Bee Paralysis Virus |
| ALPV | Aphid Lethal Paralysis Virus |
| BSRV | Big Sioux River Virus |
| BQCV | Black Queen Cell Virus |
| CBPV | Chronic Bee Paralysis Virus |
| DT | Decision Tree |
| DWV | Deformed Wing Virus |
| FAO | Food and Agriculture Organization |
| IAPV | Israeli Acute Paralysis Virus |
| KV | Kakugo Virus |
| KBV | Kashmir Bee Virus |
| LSV | Lake Sinai Virus |
| LV | Latent variable |
| LR | Logistic regression |
| ML | Machine Learning |
| NB | Naive Bayes |
| PLS | Partial least square |
| PCA | Principal component analysis |
| PCR | Principal Component Regression |
| SB | Sacbrood Virus |
| SMO | Sequential Minimal Optimization |
| SBPV | Slow Bee Paralysis Virus |
| SVM | Support vector machine |
| VDV | Varroa Destructor Virus |

# 1. INTRODUCTION

## 1.1 Bee importance, global bee health, and its current state

Bee pollinators, including honey bees, play a major role in global food production and sustaining biodiversity due to their pollination services and are now an integral part of sustaining domesticated crops and wild plants (Patel, 2021). They directly contribute to our wellbeing through the production of honey and other bee-related products such as: pollen, wax, propolis and royal jelly (Borycka et al, 2015).

Pollination services can be provided either by wild bee pollinators, or managed bee species, but modernized western agriculture has been heavily dependent on honey bees for their pollination services. With the world adoption of these current agricultural practices, their health is inextricably tied to our well-being, as we rely on them to feed an ever-increasing global population. Their pollination is linked to plant reproduction and biodiversity, which provides the foundation of many different ecosystems around the world, hence they serve as keystone species in many ecosystems (Breeze et al, 2011). Roughly 71 out of the 100 crops species that provide 90% of the global food supply are pollinated by bees. In the European Union, more than 80% of the crops are either partially or fully dependent on invertebrates for pollination, including many vegetables, fruits, and biofuel crops. Consequently, honey bee pollination services have an estimated global monetary value of 3.7 billion EUR (Kluser et al, 2010).

Despite their importance, over the last few years, honey bee populations have continued to decline due to a variety of complex factors, including those caused by human activity (Goulson et al, 2015). The bee health decline continues at an unsustainable rate, and we still have yet to identify the exact causes for the most recent decline in bee health. The major factors identified include global warming, pesticide exposure, habitat destruction, parasitic infections, and poor nutrition (Goulson et al, 2015). We now know that these stressors are likely to interact and cause a synergistic decline in global bee health (Collison et al 2016). However, we have yet to identify how different parasites are interacting with one another and how these interplays with interactions from other environmental stressors such as pesticide exposure. For example, parasites may change the level of susceptibility and mortality risk to environmental pollutants and pesticides,

while pesticide exposure may also increase the susceptibility to disease infections and mortality.

## 1.2 A systems biology approach to improve bee health

A systems biology approach differs from traditional hypothesis testing as it aims to understand the key interactions from multiple factors that are central to driving the entire system. This is in contrast to the traditional reductionist experimental approach where one variable at a time is isolated and tested. The advantages of using a systems biology approach is that it is a data driven approach in which inherently biased hypothesis testing can be avoided and it is best suited for complex multifactorial problems that have many variables involved in producing an effect or condition such as bee health (Siviter et al 2021). For an integrated systems biology approach, typically omic datasets are relied upon to characterize both known and previously unknown factors that could be driving the system and these are then integrated with other datasets that aim to measure the phenotype of interest (Hasin et al 2017). In our case, we aim to determine the significant predictors of bee health decline and to better characterize the key interacting factors that produce a synergistic decline in bee health. The typical outputs from a systems biology approach include a better understanding of the pathogenesis and etiology of the disease, characterizing biological pathways that are associated with the disease, biomarkers for rapid diagnoses of the disease and potential new drug targets for the treatment of the disease.

The relatively recent exposome paradigm aims to simplify the complexity of the environment factors that result in disease. Capturing the exposome is a holistic approach that allows for all documentation of known and previously unknown exposures, which is a tool that can be used to investigate additional environmental exposures that were not previously considered and could be leading to bee health decline (Rappaport et al 2010; Collison et al 2016). Under this framework, pathogen infections serve as an end point in which environmental exposures serve as predictors as well as a separate health factor that affects bees' overall morbidity and mortality.

Each non-genetic health factor can be thought of as an exposure event that can influence the phenotype of interest, in this case infection status of a bee disease. As a

result, the exposome is one of the most robust instruments for identifying the environmental factors that influence bee well-being. This is relevant now more than ever because environmental stressors are the main drivers in the global bee health decline, and learning how they affect bee health is critical to findings ways to improve honey bee health (Goulson et al, 2015).

Quantitative genetics theories in the biomedical community consider the role of both genetic and environmental influences in predicting continuous complex characteristics or phenotypes (such as cancer). However, these environmental factors are often overlooked and are underrepresented in many health studies (S. Dagnino, 2019). Recently, there has been an increased recognition to measure and calculate the many potential environmental causes that could be leading to health declines, as well as to forecast future declines of health (Wild, 2005); (A. Al-Chalabi, 2015); (Wild 2013).

A major challenge for beekeepers is having rapid diagnostic tools to assess the health of the bee such that they can change management practices to improve bee health before the colony collapses (Grozinger, 2020). Recently, it has also been highlighted that biomarkers are essential for improving bee health because they are objective indicators of bee health that can be applied across a variety of environments and across multiple bee species (M. M. Lopez-Uribe, 2020). What is used in today's market for diagnostic bee health tools are not enough to be considered an effective solution for this urgent problem (Moritz, 2010). Simultaneously, there is an increasing trend in medical and bioinformatics research for the extraction, discovery, and validation of biomarkers from high-dimensional omics data (Wang, 2019). Biomarkers are used in various fields of biology and medicine like personalized clinical treatment and early detection and treatment of illness. Unfortunately, the discovery of biomarkers is still a relatively new area of research in regards to bee health, and few studies in the literature have explored the possibility of using biomarkers for the detection of early disease in honeybees (Badiou-Bénéteau, 2012)

### 1.3 The use of biomarkers for rapid bee health diagnostics

While there is no single definition for a biomarker it is often defined as a biological measurement that acts as a stand-in for and, in the best-case scenario, acts as a predictor for a therapeutically meaningful endpoint or intermediate result that is more

difficult to detect. Although the terms "biomarker" and its equivalents, such as "surrogate marker" and "surrogate endpoint," have been around for decades, they are now much more extensively utilized. A search of publications indexed in PubMed for these terms reveals that interest in biomarkers skyrocketed in 2005 (Baker, 2005). Since 2010, there has been a significant surge in interest; over half of all articles published on the subject have appeared since 2014. Clinical biomarkers offer many advantages one of them is being easier to measure and less expensive than final clinical endpoints, as well as being able to be studied more often and for a shorter length of time. For example, it is far easier to monitor a patient's blood pressure than it is to evaluate left ventricular function with echocardiography (Devereux, 1983).

Biomarkers also enable clinical studies to be conducted with fewer participants than would otherwise be possible. For example, determining the impact of a medicine on blood pressure involves a small number of patients, perhaps 100 to 200, especially in a crossover design, and the trial can be finished in a year or two. To investigate the prevention of stroke mortality, a significantly bigger patient group would be required, crossover would be impossible, and the study would take many years (Kraus, 2018). Biomarkers can also help to circumvent the ethical issues that come with measuring clinical outcomes. Other advantages include, disease screening, diseases characterization, rule out, diagnose, stage, and monitor diseases, inform prognosis, individualize therapeutic interventions by monitoring responses to therapies or predicting outcomes in response to them, predict adverse drug reactions, predict and guide treatment of drug toxicity, and to identify cell types (Kraus, 2018); (Barbosa Jr, 2005).

Biomarkers are sometimes categorized according to the chain of causal mechanisms that may be linked to a number of susceptibility variables that activate a pathophysiological process. There may be more than one causal mechanism, more than one susceptibility factor, and more than one pathophysiological process for each ailment (P. Glasziou, 2008). This is especially true for complex disorders such as the decline in bee health (Liu et al 2014). The sickness or illness can occur as a direct outcome, or because of cumulative effects, or because of a final common pathway. Therefore, each fundamental pathophysiological process is thought to start with a molecular effect, which then leads to a chain of events at the cellular, tissue, and organ levels, resulting in the disease's signs and symptoms, each of the processes in the chain of pathophysiological events is matched by a biomarker or biomarkers that might be used

to track disease progression or changes in response to therapeutic intervention, as in the case of asthma (Szefler, 2012). The pharmacological level at which biomarkers arise can be used to classify them. The closer a biomarker is to the therapeutic or harmful impact, the better it is as a clinical endpoint indicator. A biomarker can be extrinsic to the individual, such as cigarette smoking as a lung cancer biomarker, or inherent to the individual, such as physical (symptoms and signs), psychological, or laboratory-based biomarkers (P. Glasziou, 2008).

The use of biomarkers for diagnosing, staging, or monitoring illness, as well as identifying the response to a treatment intervention, might be used to further classify them. They can also be classified based on the level of occurrence (molecular, cellular, tissue, organ) and whether they are related to susceptibility factors, main or secondary pathology, or illness consequences. The ultimate test of a biomarker is whether it accurately predicts the intended result under real-world situations. This is best investigated in well-designed randomized controlled clinical studies, and there have been instances where such trials have revealed that a hypothesized biomarker isn't valid.

The chain of events that leads from a disease's etiology to its clinical symptoms has several interconnections. A biomarker can be found at any stage along the chain that leads from the pathogenesis to clinical manifestations, including the molecular, cellular, tissue, organ, and whole-organ level. Similarly, a therapeutic intervention that alters such a biomarker might be positive therapy. A biomarker might be any observation that isn't the actual outcome.

Identifying biomarkers starts with understanding the pathophysiology of the disease and the factors associated with it; for example, biomarkers connected to pathways involved in the pathogenesis of heart failure appeared to be most suited for predicting and diagnosing the illness.

Next, biomarkers are identified by the mechanism by which the intervention affects the pathophysiology of the condition. Finally, the process should be correlated with the proposed biomarker. When looking for appropriate biomarkers of ageing, for example, it has been proposed that the following criteria be met: there should be a significant cross-sectional association between the biomarker and age, there should be a significant longitudinal shift in the same direction as the cross-sectional correlation, the individual differences should have a substantial level of consistency across time, and The rate of change in an aging biomarker should predict lifetime. (Ingram, 2001).

To understand complete effectiveness of a biomarker it is important to know which type of model best fits the disease condition. The Austin Bradford Hill guidelines (Phillips et al 2004) are a set of guidelines that may be utilized to determine if a biomarker and a clinical illness have a causal relationship. Biomarkers that follow these principles have a higher chance of being beneficial than those that don't. These recommendations include the following:

- strength: A strong link exists between a marker and an outcome, or between the treatment's effects on each, consistency: the link exists in various samples, in different places, in different situations, and at different periods,
- specificity: the marker is linked to a particular illness,
- temporality: changes in the marker and the outcome occur at the same time,
- gradient in biology: when you increase your exposure to an intervention, you get more results,
- plausibility: credible pathways link the marker, the disease's pathophysiology, and the intervention's mode of action,
- coherence: the link is consistent with the disease's and marker's natural histories,
- experimental evidence: an intervention produces effects that are consistent with the connection,
- analogy: we can infer a connection based on a comparable outcome.

When the appropriate data is available most of these guidelines can be verified *in silico* using mathematical models and tests, particularly using methods and approaches of machine learning can be useful. Concepts like strength, consistency, specificity, temporality are permeated and can be consistently validated *in silico* using machine learning tools.

### 1.4 Machine learning for biomarker validation:

Data mining is one of the most important applications for Machine Learning (ML). People are prone to making errors during studies or while attempting to build links between different variables, this makes it harder for them to solve these challenges. Machine learning may frequently be used to solve these issues with

effectiveness, enhancing system efficiency and machine design (Valletta 2017,). Machine learning methods employ the same set of characteristics to describe every occurrence across every dataset. The characteristics might be continuous, categorical, or binary in nature. In contrast to unsupervised learning, where examples are provided with unknown labels (the corresponding correct outputs) (Jain et al 1999), supervised learning occurs when instances are given with known labels (Kotsiantis et al 2007). The process of learning a set of rules from cases, or more broadly, generating a classifier that can be used to generalize from new instances, is known as machine learning. Therefore, by partitioning the dataset into a training dataset and a test dataset, the biomarker features that are selected from statistical multivariate analysis can be tested (validated) *in silico*.

To employ machine learning, the dataset must first be gathered. If a qualified specialist is available, he or she can advise on which fields (attributes, characteristics) are the most useful. If not, the most basic strategy is "brute-force," which involves measuring everything accessible in the hopes of isolating the proper (informative, relevant) traits. A dataset gathered via the "brute-force" approach, on the other hand, is not immediately suited for induction. It contains noise and missing feature values in most situations, necessitating extensive pre-processing (Zhang et al 2002). Data preparation and preprocessing is the second phase. Researchers have a variety of options for dealing with missing data, depending on the circumstances. (Batista et al 2003)

One of the strategies employed to address these issues is feature selection. Feature selection is the process of identifying and eliminating as many unwanted and redundant qualities as possible. Feature selection is used not only to address problems that comes with noise, but also to deal with the difficulties of learning from excessively large datasets. In these datasets, feature selection is an optimization issue that aims to keep the mining quality, while reducing the data dimensions (Liu et al 2001). It decreases the amount of data in a dataset and allows a data mining algorithm to operate and perform successfully with very huge datasets. There are several methods for selecting samples from a huge dataset (Reinartz et al 2002). Decreasing the data's dimensionality, allows data mining algorithms to operate more quickly and effectively. The interdependence of various features has a negative impact on the accuracy of supervised machine learning classification models. This issue may be solved by creating new features from the existing feature set (Markovitch S, 2002). Feature

construction/transformation is the term for this approach. These newly discovered characteristics might contribute to the development of more succinct and accurate classifiers. Furthermore, the finding of significant characteristics leads to the created classifier's comprehensibility as well as a better grasp of the studied idea.

Another important step is deciding the exact learning algorithm to utilize. The classifier (which maps unlabeled cases to classes) becomes ready for typical usage once preliminary testing has deemed it adequate. Prediction accuracy (the proportion of right predictions divided by the total number of predictions) is frequently used metric to evaluate the classifier. There are at least three methods for calculating the accuracy of a classifier. One method is to divide the training set to several parts and devote two-thirds for training and the remaining third for performance estimation.

If the error rate evaluation is inadequate, an earlier step of the supervised ML process must be reexamined . In such cases, a number of reasons for the outcome must be considered: the use of features that are possibly not significant or not sufficiently related, or do not adequately describe the response for the problem, a larger training set is required, the disparity between the problem's dimensionality is too high (the curse of dimensionality), the method used is ineffective, or parameter adjustment is required. Data imbalance is also another issue that might affect the performance of the classifier (Japkowicz N., 2002).

On certain datasets, statistical comparisons of the accuracies of trained classifiers are frequently used for comparing supervised ML methods. In cases where we have a substantial amount of data we may sample several training sets of size N, run several different learning algorithms on each of them, then assess the difference in accuracy for each of the classifiers on a large test set. These differences' average is an estimate of the predicted difference in generalization error across all potential training sets of size N, and their variance is an estimate of the classifier's variance in the entire set. Next step is to test the null hypothesis, which is that the mean difference between the classifiers is zero using a paired t-test. A t-test analyses the t-statistic, t-distribution values, and degrees of freedom to determine statistical significance. An analysis of variance test could be used as an alternative to the t-test when comparing three or more means.

There are two sorts of failures that can occur throughout this test. The likelihood that the test would reject the null hypothesis incorrectly which is known as type I error. The other type of failure is that the null hypothesis is not rejected (i.e., there is a

"significant" difference, even though there is none). The type II mistake should be the most common of the two. However, in reality, we frequently have just one dataset of size N from which all estimations must be derived. Therefore, subsampling is used to create different sub-sets for training and testing, and the instances that were not sampled for training are utilized for testing. Regrettably, this contradicts the independence assumption that is required for adequate significance testing. This is problematic because the researcher must be able to manage Type I errors and know the likelihood of rejecting the null hypothesis mistakenly. To address this issue, several heuristic variations of the t-test have been devised (Dietterich, 1998).

The test result should be independent of the precise partitioning that comes from the randomization procedure because it would make it much easier to duplicate published experimental results. However, there is always some sensitivity to the partitioning employed in practice. To determine replicability, we must run the same test on the same data numerous times with various random partitioning — generally 10 times — and count the number of times that the results are the same. One of the most typical tasks that the so-called Machine learning systems is supervised classification. To verify the independence of the extracted biomarkers from the model that was used to extract them we need to test the quality of association between the biomarkers and the disease using a variety of machine learning algorithms that verify this association is valid and robust. For this purpose, a vast variety of Artificial Intelligence-based approaches have been used. These include logical techniques (Decision trees), perceptron-based techniques, and statistics-based techniques (naïve-Bayesian, and logistic regression).

### 1.4.1 Decision trees

Decision trees are classification trees that categorize instances based on the value of a feature Each node in a decision tree represents a feature in an instance to be classified, and each branch provides a value that the node could accept . Starting at the root node, instances are categorized and arranged depending on their feature values (Murthy, 1998). Because the difficulty of generating optimum binary decision trees, these decision problems are classified as NP (nondeterministic polynomial time). NP is a set of choice problems in which the problem instances contain proofs that can be

verified by a deterministic a Turing machine if the answer is "yes." In polynomial time . The feature that best separates the training data is the root node of the tree. Two methodologies for selecting which feature separates the training data the best are information gain and gini index. (Hunt E. 1966; Breiman L. 1984). The relief F method is another popular strategy for splitting the training data, which assesses each variable in the context of other variables, whereas measures such as information gain and gini index estimates each attribute separately (Kononenko, 1994). The multitude of options may sound promising. However, most of the research has found that there is no "one-size-fits-all" solution for the problem of accurately splitting data and much like other aspects of machine learning it is problem dependent (Murthy, 1998).

When determining which measurement to utilize, a comparison of specific methodologies may still be necessary to determine which method to use for a certain dataset. The method is then done once more for each partition of the divided data, building several sub-trees from the data until the training data is split into groups where the individuals are more similar within group than across groups. A decision tree, or any learned hypothesis h, is said to overfit training data if another hypothesis h′ exists that has a bigger error than h when tested on the training data but a smaller error when tested on the entire dataset. To avoid overfitting on the training data, decision tree algorithms can adopt one of two approaches: they can either halt the training procedure before the training data is fully fitted or they can trim the induced decision tree early. (Galathiya, 2012).

If two trees use the same tests and have the same prediction accuracy, the one with the fewer branches is typically selected. Pre-pruning the decision tree by not allowing it to grow to its full size is the simplest technique to avoid overfitting. Establishing a non-trivial termination condition for the feature quality measure, such as a threshold test, can help.

In decision tree classifiers, post-pruning methods are widely used to evaluate the performance of decision trees after they have been pruned using a validation set. Any node can be destroyed, and the most common training instance class sorted to it will be assigned to it. A comparison of well-known pruning strategies revealed that there is no one ideal pruning approach. Decision trees algorithm use the divide-and-conquer approach to achieve desired results. In computer science, divide-and-conquer is an algorithm design paradigm in which a problem is recursively broken down into two or

more sub-problems of the same or type, until these sub-problems become simple enough to be solved directly.

Even if the divide-and-conquer method is fast, in projects with hundreds of thousands of instances, efficiency is crucial. Sorting the instances on a numeric characteristic to get the optimal threshold, is the most time-consuming part. This can be sped up by determining potential numeric feature thresholds just once, thus turning the feature to discrete intervals, or by determining the threshold from a subset of the occurrences. Because they employ splits based on a single attribute at each internal node, decision trees are generally univariate. When it comes to issues that need diagonal partitioning of the instance space, most decision tree methods fall short. The instance space is divided orthogonally to one variable's axis and parallel to the other axes. As a result, following partitioning, the resultant areas are all hyperrectangles. However, there are a few techniques that may be used to avoid this (C. T. Yildiz, 2001).

### 1.4.2 Single layered perceptron

As stated before to verify the independence of the biomarkers from the model we tested the quality of association between the biomarkers and the disease using multiple algorithms, aside from decision trees we also used Single layered perceptron. The following is a basic description of a single layered perceptron: Perceptron computes the sum of weighted inputs, $\sum_i x_i w_i$ where $x_1$ through $x_i$ are input feature values and $w_1$ through $w_i$ are connection weights/prediction vector (usually real integers in the interval [-1, 1]), and the output goes through an adjustable threshold. If the sum is above the threshold, the output is 1 or else it is 0. The most typical technique to learn from a batch of training samples with the perceptron algorithm is to run it repeatedly over the training set until it finds the optimal prediction vector that can correctly predict all the training cases. The labels on the test set are then predicted using this prediction rule. The weights of the perceptron concept are updated as follows. If the weights are too low, the predicted value will be y' = 0 while the actual value y = 1, then each feature $x_i$ = 1, $w_i = w_i * \alpha$, where $\alpha$ is a number larger than 1, that is termed the promotion parameter. If the weights were excessively high, the predicted value will be y' = 1 and the actual value y = 0; as a result, it reduces the associated weight for each characteristic $x_i$ = 1 by setting $w_i = w_i * \beta$, where $\beta$ is a value between 0 and 1 termed the demotion

parameter. An exponential updating technique is where a relevant feature weight increases exponentially, while irrelevant feature weights decrease exponentially. As a result, it has been demonstrated experimentally that this updating mechanism can respond quickly to changes in the target function. The desired function (such as user preferences) does not remain constant over time. It is required to identify which previous training instances might be eliminated in order for a decision tree algorithm, for example, to adapt to these changes.

There are many algorithms created that are comparable to this one, such as the voted perceptron, which is a newer technique that accumulates more information during training and then uses this detailed knowledge to make better predictions about test data. The list of all prediction vectors that were created after each error is the information it tracks throughout training. The method counts the number of iterations it "survives" until the next error is made for each such vector and this count is referred to as the "weight" of the prediction vector. Afterwards the binary prediction of each of the prediction vectors is calculated and then all these predictions are aggregated using a weighted majority vote constituting a forecast. To summarize, perceptron-like linear algorithms excel at dealing with irrelevant information because of their greater time complexity. Single layer perceptron can be a highly advantageous algorithm when dealing with problems where there are numerous attributes, but only a handful of them are useful (Freund, 1999).

### 1.4.3 Naive Bayes classifiers

Naive Bayesian networks (NB) are very basic Bayesian networks made up of directed acyclic graphs with just one parent (representing the unseen node) and multiple children (corresponding to observed nodes) with a strong assumption of child nodes being independent of their parent. As a result, the independence model (Naive Bayes) is based on calculating the posterior probability of an instance being from a certain class given the set of features:

$$P(i|X) = \frac{P(X|i)P(i)}{P(X)} \qquad (1.4.3\text{-}1)$$

where:

$P(i|X)$: The posterior probability of class (i, target) given predictor (x, features).

$P(i|X)$: The prior probability of the class i.

$P(X|i)$: The probability of predictor given class.

$P(X)$: The prior probability of predictor.

When these probabilities for different classes are compared, the higher probability

suggests that the class label value is more likely to be of that actual label (if R > 1: predict I, otherwise, predict j where $R = \frac{P(i|X)}{P(j|X)}$). The Bayes classification process computes the probabilities $P(i|X)$ via a product operation, so it is especially vulnerable to being influenced by probabilities of 0. This may be avoided by using the Laplace estimator or m-estimate, which involves adding one to all numerators and multiplying the number of added ones by the number of added ones in the denominator (Cestnik, 1987).

Due to the strict child node independence assumption, Naive-Bayesian algorithms are sometimes considered to be largely inaccurate. However, on standard benchmark datasets,

The naive Bayes classifier has been compared against state-of-the-art algorithms for decision tree induction, instance-based learning, and rule induction. The findings revealed that this method outperforms other learning systems in several cases, even on datasets with significant feature dependencies (Domingos, 1997). Furthermore, the naive Bayes classifier is efficient in terms of training time because the model is in the form of a product, it may be transformed to a sum using logarithms, resulting in lower computing times. If a feature is numerical, it is usually discretized during data pre-processing, however if needed, the feature probabilities can be computed using the normal distribution.

### 1.4.4 Support vector machines

Support SVMs (Support Vector Machines) are the most recent out of all the supervised machine learning approaches. The concept of a "margin" on either side of a

hyperplane that separates two data classes is central to the SVM algorithm. It has been demonstrated that increasing the margin and therefore generating the biggest possible space between the separating hyperplane reduces the expected generalization error (Burges, 1998). For problems where data classes can be separated linearly, data points that lie on the edges of the optimum separating hyperplane are known as support vector points, and the solution is represented as a linear combination of only these points. Other data points are not considered. As a result, the number of features in the training data has no bearing on the SVM model complexity (the number of support vectors selected by the SVM learning method is usually modest). Consequently, SVMs are ideally adapted to learning problems involving a large number of features compared to the number of training cases.

Even though the greatest margin allows the SVM to choose among numerous candidate hyperplanes, because many datasets contain misclassified instances, the SVM may not be able to detect any separating hyperplane at all. This can be addressed by introducing positive slack variables on the constraints. Nonetheless, the majority of situations encountered with real-world training datasets, contain data that isn't separable and for which there isn't a hyperplane that successfully separates the positive and negative samples. In such cases, the data can be mapped onto higher-dimensional space and a defined separating hyperplane there may be a solution to the inseparability problem. A non-linear separation in the original input space will correlates to a linear separation in transformed feature space. Kernels are a sort of function that allows inner products to be computed directly in feature space without requiring the previously specified mapping. Once a hyperplane has been created, the kernel function is used to map new points into the feature space for classification (Genton, 2001). The kernel function defines the altered feature space in which the training set instances will be classified, hence choosing a proper kernel function is crucial for a proper analysis.

Estimating a range of alternative parameters and using cross-validation over the training set to identify the best one is a standard procedure. As a result, one of SVMs' drawbacks is their slow training speed. A SVM will work correctly if the kernel function is valid, even if the designer has no idea which training data features are employed in the kernel-induced transformed feature space . The Nth dimensional QP issue is used to train the SVM, where N is the number of samples in the training dataset. Solving this problem with normal QP methods necessitates huge matrix operations and

time-consuming numerical computations and is typically slow and inefficient for large datasets.

SMO (Sequential Minimal Optimization) is a simple technique that can address the SVM QP problem without needing any extra matrix storage or numerical QP optimization stages. The overall QP problem is decomposed into QP sub-problems via SMO. In most instances, two modified variants of SMO have been suggested as alternatives that are much faster than the original SMO (Keerthi, 2002). Finally, the SVM's training optimization problem must approach a global minimum, avoiding the local minimum that can occur with other search methods such as neural networks. However, because SVM approaches are binary, they must be reduced to a set of several binary classification problems in the event of a multi-class problem. While problematic, discrete data can still be analyzed, but proper rescaling is required using this approach.

## 1.4 Aim of the study

Here, we combine the relatively new exposome paradigm with a powerful non-hypothesis driven, unbiased, systems biology approach to address the complex multifactorial problem of honey bee health decline to unravel how environmental exposures interact with the 20 most common honey bee diseases responsible for the most recent decline in bee health. We used state-of-the-art LC-QTOF analysis to generate the exposome and identify 4631 exposures from 29 hives, across 10 different locations, spanning from rural to urban apiaries, sampled once per month, for three months. The exposome data was then integrated with the presence and abundance of the 20 most common diseases from molecular screening to identify the most significant exposure profile that is associated with the presence of a bee disease. We then validated the exposure profile comprised of chemical biomarkers using five different machine learning techniques to demonstrate that we can predict whether a hive is infected with a particular disease with roughly 80-90% accuracy, selectivity and sensitivity, recall, and precision.

# 2. MATERIALS AND METHODS

## 2.1 Overview of workflow

Across the three time points 29 hives were sampled and pooled together yielding 87 sampled hives total. The bee diseases of the sampled hives were screened using molecular techniques and the exposome of the beehives were generated using LC-QTOF analysis. This data was imported and combined into Excel (Microsoft) yielding 87 (Hives/rows/samples) by 4631 (compounds/columns/features) for each of the 11 diseases, where each row represents a hive, and each column represents an exposomic compound. We aimed to extract a set of diagnostic biomarkers (features/compounds). From a statistical perspective, a biomarkers suite is a subset of features, which if trained with a predictive model such as partial least square (PLS) and can discriminate between healthy and unhealthy hives for each disease with relatively high accuracy. The biomarker profiles selected were then validated using a 3-fold cross validation, twice, across five different machine learning models (Figure 1).

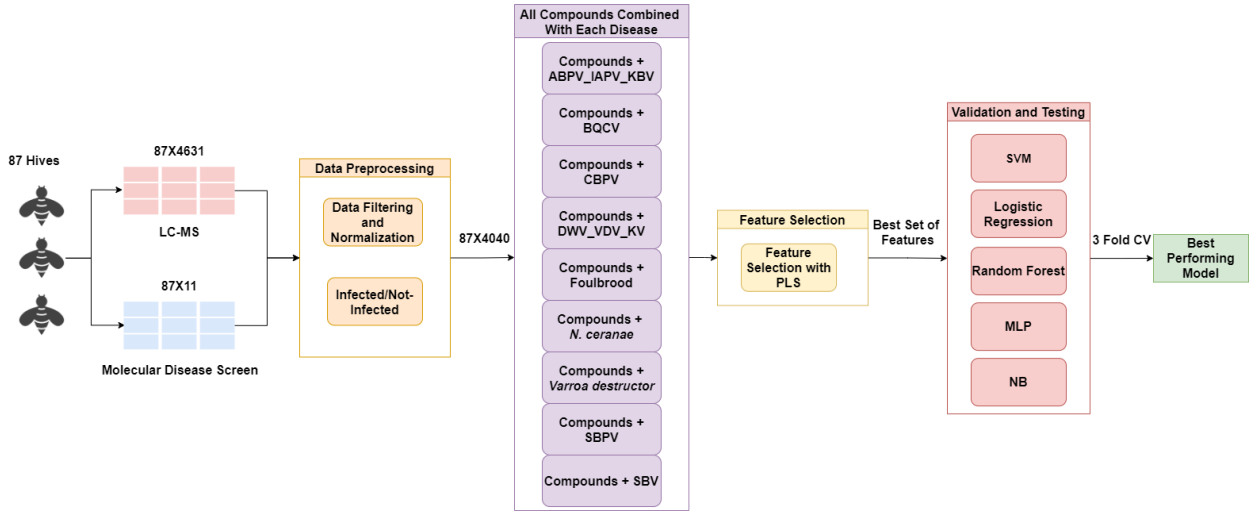



**Figure 1.** An overview of the analysis workflow. Starting with analysis for bee diseases using molecular screening techniques and the exposome of the hives using LC-QTOF. The exposome dataset was pre-processed, while from the quantitative diseases screening the hives were classified as infected or uninfected. These datasets were then integrated, and feature extraction was conducted using a Partial Least Square regression (PLS). These were then validated using five different machine learning algorithms.

## 2.2 Sample collection

Roughly 300 in-hive bees were sampled once a month for a total of 3 months, starting in June, from brood combs using an insect vacuum, from 30 hives, in 10 different geographical locations, in the Philadelphia, Pennsylvania, region. Apiaries were sampled across an urbanization gradient. The collected bees were placed in a sterile 50 mL Falcon tube that was immediately placed on dry ice. At the time of bee sampling the number of Varroa mites were quantified per 300 bees using the powdered sugar roll method (Macedo 2002). The collected bees were homogenized with 20 mL of RNase free water using a 50 mL tissue grinder (Fisher, Waltham, Massachusetts, USA). Two sub-samples of the bee homogenate were created, one was used for molecular disease screening and the other for LC-QTOF analysis. All samples were stored at -80 ˚C until further analysis.

## 2.3 Disease screening methods

The disease screening methods were followed from (Mayack, 2022). All screening methods started with 150 µl of bee homogenate. Briefly, RNA was extracted using a Qiagen RNA mini kit (Qiagen, Hilden, Germany) following the manufacturer's instructions and the 14 most common honey bee viruses (Chronic Bee Paralysis Virus (CBPV), the Deformed Wing Virus (DWV), *Varroa* Destructor Virus (VDV), and Kakugo Virus (KV) complex, Black Queen Cell Virus (BQCV), Acute Bee Paralysis Virus (ABPV), Israeli Acute Paralysis Virus (IAPV), and Kashmir Bee Virus (KBV) complex, Slow Bee Paralysis Virus (SBPV), Sacbrood Virus (SB), Aphid Lethal Paralysis Virus (ALPV), Lake Sinai Virus (LSV), and Big Sioux River Virus (BSRV)) were screened using the MLPA BeeDoctor Kit (MRC-Holland, Amsterdam, the Netherlands) (Smet., 2012). The MLPA product was visualized using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, USA). Technical duplicates were run for each sample and then averaged. The 2100 Bioanalyzer Expert Software (Agilent Technologies, Santa Clara, USA) was used to carry out the semi-quantification of each virus relative to the *RPL8* reference gene.

American and European Foulbrood were screened for, following the manufacturer's instructions, using the respective BactoReal qPCR assay kit (InGenetix, Vienna, Austria). The extracted DNA was measured with a nanodrop and standardized to a concentration of 5 ng/µL for qPCR assays. Technical duplicates were run and averaged for each sample. Relative quantification was accomplished by using a fit-point analysis, Cp values were used to determine the relative abundance of foulbrood present in each sample. DNA extraction methods for American Foulbrood were followed from (Graaf, et al. 2013). DNA extraction methods for European Foulbrood were followed from (Forsgren et al. 2013).

*Nosema* quantification was based on methods adapted from the HBRC phenol chloroform method (Hamiduzzaman, 2010). DNA from each sample was standardized to 5 ng/µl for the PCR assay. Technical triplicates were run and averaged. Methods for stonebrood screening were adopted from Nasri et al., (2015) and for chalkbrood (*Ascosphaera apis*) screening was adopted from Jensen et al., (2013) (Nasri, 2015). All samples were standardized to 5 ng/µL for template DNA and 25 µL duplex PCR reactions were carried out with *Rps5* reference gene for semi-quantification. The Beta-tubulin gene target was used for both *Aspergillus fumigatus* and *Aspergillus flavus* screening. All PCR products were visualized on a 1% agarose gel electrophoresis run at 100 v for 45 min (New England BioLabs, Ipswich, MA). Technical duplicates were run for all stone and chalkbrood samples. The semi-quantitative analysis for each screening method was carried out using ImageJ (Abramoff 2004).

## 2.4 Exposome methods

### 2.4.1 Sample preparation

Bees were transferred to a 50-mL plastic tube and further extracted according to a previously described modified QuEChERS procedure (Mullin, 2010). To each sample a 27-mL portion of extraction solution (55 % acetonitrile, 44 % Milli-Q water, 1 % glacial acetic acid) was added before high-speed homogenization for 1 min in a Magic Bullet blender. To each sample was then added 6 g anhydrous magnesium sulfate ($MgSO_4$) and 1.5 g anhydrous sodium acetate (NaAc). After 1 min of agitation, tubes were centrifuged for 5 min at 1600 rpm. A 1-mL portion of the supernatant was

transferred to a 2-mL plastic centrifuge tube containing 0.05 g of primary secondary amine (PSA), 0.05 g C18, and 0.15 g $MgSO_4$. Tubes were vortexed for 1 min and centrifuged for 5 min at 1600 rpm. Half of the resulting supernatant was gently heated in a sand bath, reconstituted in 50 μL of acetonitrile, and transferred to an autosampler vial containing an insert for analysis via Q-TOF LC/MS.

### 2.4.2 LC-QTOF analysis

Analysis via Q-TOF LC/MS was performed using an Agilent ZORBAX Eclipse Plus C18 column (2.1 x 50 mm, 1.8 μm, P.N. 959757-902) and an Agilent Technologies 6545 Q-TOF LC/MS. Mobile phases were 0.1 % formic acid in water v/v (mobile phase A) and 0.1% formic acid in acetonitrile v/v (mobile phase B). The injection volume was 1 μL. The linear gradient
was set to change from 95% mobile phase A to 5% mobile phase A over 50 min with a flow rate of 0.300 mL/min. Samples were run in electrospray ionization positive and negative modes. Additional parameters were as follows: nebulizer gas 15 psi, capillary voltage 3500 V, acquisition range 100-1700 m/z, acquisition speed of 1 spectrum per second. A solvent blank was run after every 15 samples.

## 2.5 Chemical biomarker analyses

For each disease, the hive was classified as healthy or unhealthy based on the relative pathogen load measured. For *Varroa* mites a threshold of 9 mites or above per 300 bees was considered as an unhealthy infested hive as this is the first threshold established that is used to determine if beehives need to be treated with miticides (Macedo 2002). For DWV anything lower than 1 on the relative quantification scale was healthy because it is known that many hives have low levels of DWV that are asymptomatic around the world (Wilfert 2016). For *N. ceranae* and the rest of the viruses if the hives contained no infection, then they were a healthy hive. LSV, BSRV, American Foulbrood, *Nosema apis*, Stonebrood, and Chalkbrood were screened for, but due to lack of prevalence were excluded from the final analysis. We did not find a high level of multicollinearity across time points (r = 0.47, 0.46, and 0.45, respectively) and

so to increase statistical power we pooled hive data across the time points. However, across the three time points, one hive collapsed at the second time point, so this hive was removed from the analysis to maintain a balanced design across time points, yielding a total of 87 sampled hives for the final analysis.

The exposome data was not normal so it was $\log_2$ transformed and exported to an Excel sheet with 87 rows representing the sampled hives and 4631 columns representing the different compounds from the exposome LC-QTOF analysis. For dimensionality reduction we used a Partial Least Squares (PLS) regression across each disease. We then used a 3-fold cross validation, two times, for assessing the quality of the PLS regression model. Then, the best model's regression coefficients were extracted. These regression coefficients indicated the amount of association between each exposome compound to the respective disease. A low absolute value of the regression coefficient indicated a low association between the exposome compound and the disease of interest. We discarded the exposome compounds with the lowest association indicated by the regression coefficients. The optimal PLS regression model selected was the one that had the lowest means square error (MSE).

A Principal Component Regression (PCR) regression model was also used as an analysis method. PLS and PCR both use multiple linear regression, which means they create a linear model with the formula Y=XB+E. In statistical terms, the predictors are X, and the response is Y. X is the collection of compounds in our research, and Y is the amount of viral load we wish to predict. E represents the amount of error around the predictions from the model. If the matrix X contains highly correlated data, then this may obfuscate the variation we wish to detect, namely disease load variations. In PCR, a linear transformation W transforms a set of measurements of X into an equivalent set X' = XW, resulting in linear independence for all new compounds (which are the principal components). X’ represents the factor scores. In PCR, the linear transformation reduces the covariance between the rows of X' to a minimum. That is, just the features from the data are used in this procedure, not the response values.

The main distinction between PCR and PLS regression is this that PLS works by maximizing the covariance between Y and X' and this is based on obtaining a suitable linear transformation. In other words, PLS considers both spectra and response values, overcoming some of the limitations of PCR in the process. As a result, the PLS analysis method has become a gold standard in current bioinformatic analyses.

Cross validation with the coefficient of determination $R^2$ and the mean squared error are used to assess the quality of the model's calibration (MSE). The most common cross-validation method is to divide the data into a few groups, leave one out, then fit a PLS model to the remaining groups (Refaeilzadeh, 2009). After that, the model is used to forecast the values of the group that was left out. This fundamental procedure is repeated until all samples have been predicted. Cross-validation scores are then calculated by averaging the measures.

The variable selection of the PLS analysis was performed using the following steps:

1. Quantify the quality of the PLS regression using the entire dataset, such as with cross-validation or prediction data.
2. Take the regression coefficients from the best model. Each regression coefficient links each variable to the response in a unique way. A low absolute value of the regression coefficient indicates a low correlation between a certain variable and the response of interest.
3. Discard the variable with the lowest correlation. These are the variables that often degrade the model's quality, thus by basically rejecting them, we anticipate enhancing the metrics associated with our prediction or cross-validation.

Setting a threshold and removing any variables whose regression coefficients (absolute value) fall below that threshold is one technique to exclude the lowest correlation variables. However, this strategy is very dependent on the threshold chosen, which is set subjectively or necessitates a trial-and-error approach. Another less subjective technique is to remove one variable at a time (the one with the lowest absolute value of the related regression coefficient) and rebuild the prediction model. The technique is iterated until the MSE (means square error) of the prediction or cross-validation falls below a certain threshold. The halting condition for the optimization process is that eliminating more variables will result in a worse prediction at some point.

Alternatively, one might iteratively delete a predetermined number of variables, then verify for which number of variables the MSE is minimized. The procedure is independent of the threshold's subjective selection and can be used to any data set without modification. One disadvantage of this strategy is that the processing time can be slow for large datasets. In our research the quality of the PLS regression model was assessed using two rounds of 3-fold cross validation. We split the data into thirds and then trained the model with two thirds of the dataset and tested it on the last third, this

operation was repeated three times for each round, during each repeat, a different third of the data was chosen for testing and training the model. The regression coefficients of the best model were then retrieved. The extent of correlation between each exposome chemical and the bee disease was revealed by the regression coefficients. The exposome chemicals having the weakest correlation, as shown by the regression coefficients, were eliminated. Consequently, the PLS regression model with the lowest mean square error (MSE) was chosen as the best prediction model.

### 2.6 Validation of biomarkers using machine learning

Five different machine learning techniques were used to evaluate the biomarker profiles (predictive models) found from the PLS regression analysis having the strongest relationship with each bee disease. We assessed the accuracy and predictability of the biomarker profiles to see if they could properly distinguish a healthy and unhealthy hive. Logistic Regression (LR), Support Vector Machine (SVM), Random Forest classifier (RF), Multi Layered Perceptron (MLP) and Naive Bayes classifier (NB) were implemented using sklearn (Pedregosa, 2011). Because of the high sparsity and large disparity between our dataset dimensions, I relied on the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve, precision, the F-1 score, and recall as measures of the biomarker libraries' performance.

Across all bee diseases, there was a total of 1018 unique chemical biomarker compounds identified. Using a python script[1], I automated a search to collect additional information about the chemical biomarkers from the PubChem database, out of the 1018 biomarkers we found more detailed information for 626 biomarkers. For the 626 biomarkers, I retrieved classification information about the chemical itself, its role, if any, in metabolic pathways, the honey bee genes that they interact with and the biological functions these are associated with, and lastly the human diseases that are related to these chemical biomarkers, that are coming from the local environment, as honey bees can serve as sentinels for surveying environmental quality.

[1] https://github.com/MHDalayoubi/Biomarkers_extraction_code.git

### 2.7 Pathway analysis

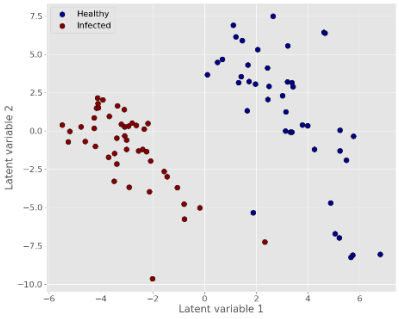
Out of the 626 chemical biomarkers, only 72 of them contained information about the honey bee metabolic pathways that they are involved in. A network analysis was conducted across the honey bee diseases to display which metabolic pathways are related to one another that are associated with one or more diseases. From this analysis we aim to understand possible interactions from multiple stressors that might lead to a synergistic decline in bee health. In addition, using the PubChem database (Kim S, 2021) chemical biomarkers identified to have relationships with genes of the honey bee ontology with *Drosophila* were identified and a network analysis was conducted. The function of the interacting genes was identified using a PANTHER GO analysis (Huaiyu Mi, 2021).

### 2.8 Human health environmental monitoring analysis

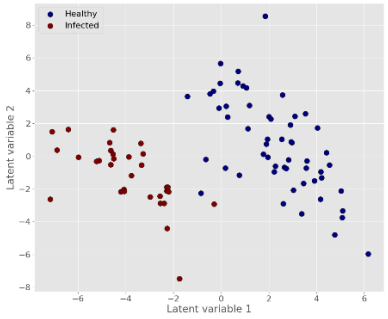
Furthermore, we noticed that many of the compounds that were found to be related to the diseases under study were also related to diseases found in humans and other organisms which could be an indication that bees can be used for monitoring environmental pollution. All chemicals with various human disease and disorders were identified and presented in a table. This classification procedure also included whether the chemical biomarker can be classified as a pesticide, pollutant, or pharmaceutical drug, or a derivative from these.

## 3. RESULTS

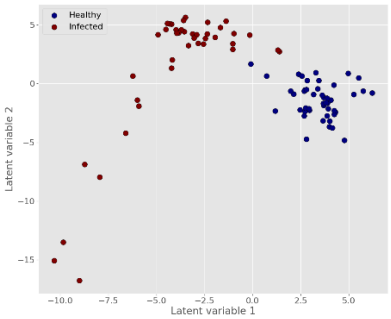
The PLS analysis reveals distinct and tight clustering for healthy and unhealthy hives based on infection load for each of the bee diseases (Figure 2).


Healthy
Infected
Latent variable 2
Latent variable 1

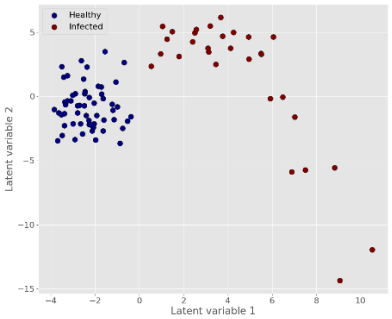
A


Healthy
Infected
Latent variable 2
Latent variable 1

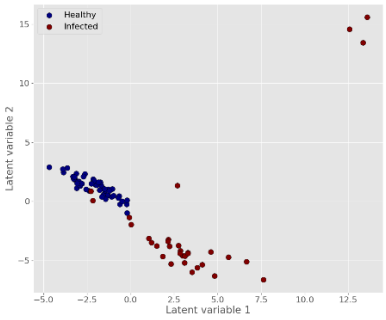
B


Healthy
Infected
Latent variable 2
Latent variable 1

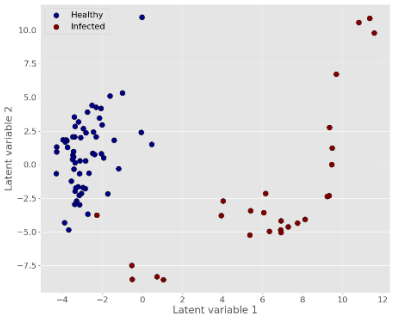
C


Healthy
Infected
Latent variable 2
Latent variable 1

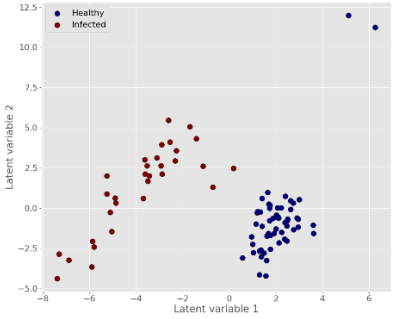
D


Healthy
Infected
Latent variable 2
Latent variable 1

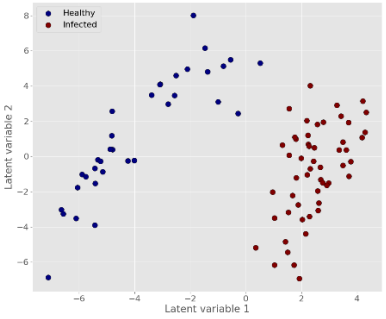
E


Healthy
Infected
Latent variable 2
Latent variable 1

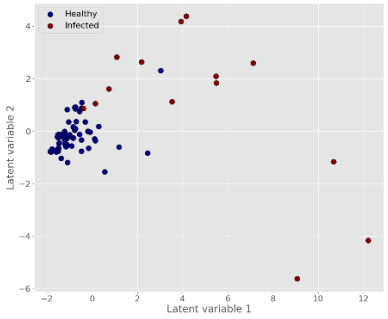
F


Healthy
Infected
Latent variable 2
Latent variable 1

G


Healthy
Infected
Latent variable 2
Latent variable 1

H




I

**Figure 2.** The output of the PLS analysis on the exposome data for the following bee diseases: (A) Black Queen Cell Virus (BQCV), (B) Chronic Bee Paralysis Virus (CBPV), (C) the Deformed Wing Virus, Varroa Destructor Virus and the Kakugo Virus complex (DWV_VDV_KV), (D) European Foulbrood, (E) *N. ceranae*, (F) *Varroa destructor*, (G) Slow Bee Paralysis Virus (SBPV), (H) Sac Brood Virus (SBV) and the (I) Acute Bee Paralysis Virus, Israelei Acust Bee Paralysis Virus, and the Kashmir Bee Virus complex (ABPV_IAPV_KBV). Tight clustering indicates that the latent variables contain chemical biomarkers that are uniquely associated with each state of the honey bee hive health status, either healthy or unhealthy.

From this analysis a total number of 36, 308, 203, 454, 285, 398, 342, 148, and 390 biomarkers constituted a biomarker profile as predictors for ABPV_IAPV_KBV, BQCV, CBPV, DWV_VDV_KV, SBPV, SBV, European Foulbrood, *N. ceranae*, and *Varroa* mites, respectively. The average of the two 3-fold cross validations of each biomarker profile across the 5 different machine learning models demonstrate a range of sensitivity and selectivity (AUC), accuracy (F1-score), recall, and precision from 0.77 to 0.93. The highest performing model across all of the diseases was SVM with 0.94 AUC and the lowest was RF with a 0.72 accuracy score (Table 1).

**Table 1.** The machine learning validation results for each bee disease. A total of 5 different machine learning models, LR, SVM, RF, MLP, and NB, were used for the training and testing of the biomarker profiles, which is shown in the column on the right. Averages within and across the machine learning models are presented and when averaged for all diseases is about an 80 - 90% level of selectivity, sensitivity, accuracy, recall, and precision as indicated by the AUC, F1-score, recall, and precision score, respectively, in predicting the respective bee disease. AUC is comprised of sensitivity and selectivity performance of the model. The precision is the ratio of tp / (tp + fp) where tp is the number of true positives and fp is the number of false positives. Recall is the ratio of tp / (tp + fn) where tp is the number of true positives and fn the number of false negatives. The F1-score can be interpreted as a weighted harmonic mean of the precision and recall, where an F-1 score reaches its best value at 1 and worst score at 0.

The F1-score weights recall more than precision by a factor of beta.

| | | ABPV_IAPV_KBV | BQCV | CBPV | DWV_VDV_KV | SBPV | SBV | European Foulbrood | *N. ceranae* | *Varroa destructor* | AVG |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LR | AUC | 0.89 | 0.9 | 0.91 | 0.83 | 0.85 | 0.8 | 0.83 | 0.87 | 0.92 | 0.87 |
| | Precision | 0.86 | 0.82 | 0.84 | 0.75 | 0.82 | 0.73 | 0.77 | 0.80 | 0.86 | 0.81 |
| | Recall | 0.88 | 0.80 | 0.83 | 0.74 | 0.80 | 0.72 | 0.76 | 0.80 | 0.84 | 0.80 |
| | F1-score | 0.87 | 0.80 | 0.83 | 0.74 | 0.80 | 0.71 | 0.75 | 0.80 | 0.84 | 0.79 |
| SVM | AUC | 0.89 | 0.95 | 0.96 | 0.96 | 0.95 | 0.95 | 0.93 | 0.87 | 0.98 | 0.94 |
| | Precision | 0.87 | 0.89 | 0.89 | 0.89 | 0.91 | 0.91 | 0.85 | 0.81 | 0.93 | 0.89 |
| | Recall | 0.87 | 0.87 | 0.88 | 0.89 | 0.91 | 0.90 | 0.85 | 0.80 | 0.93 | 0.88 |
| | F1-score | 0.86 | 0.87 | 0.88 | 0.89 | 0.91 | 0.89 | 0.85 | 0.80 | 0.92 | 0.87 |
| RF | AUC | 0.93 | 0.78 | 0.85 | 0.84 | 0.77 | 0.69 | 0.79 | 0.85 | 0.89 | 0.82 |
| | Precision | 0.84 | 0.71 | 0.83 | 0.76 | 0.71 | 0.65 | 0.77 | 0.83 | 0.81 | 0.77 |
| | Recall | 0.87 | 0.68 | 0.80 | 0.75 | 0.68 | 0.65 | 0.71 | 0.82 | 0.78 | 0.75 |
| | F1-score | 0.83 | 0.68 | 0.79 | 0.75 | 0.64 | 0.62 | 0.64 | 0.81 | 0.75 | 0.72 |
| MLP | AUC | 0.81 | 0.93 | 0.92 | 0.97 | 0.92 | 0.87 | 0.88 | 0.93 | 0.94 | 0.91 |
| | Precision | 0.84 | 0.87 | 0.85 | 0.92 | 0.84 | 0.81 | 0.82 | 0.86 | 0.90 | 0.86 |
| | Recall | 0.86 | 0.84 | 0.84 | 0.91 | 0.833 | 0.79 | 0.81 | 0.86 | 0.89 | 0.85 |
| | F1-score | 0.84 | 0.84 | 0.84 | 0.91 | 0.83 | 0.79 | 0.80 | 0.85 | 0.89 | 0.84 |
| NB | AUC | 0.91 | 0.92 | 0.96 | 0.96 | 0.93 | 0.85 | 0.92 | 0.96 | 0.94 | 0.93 |
| | Precision | 0.92 | 0.90 | 0.95 | 0.91 | 0.87 | 0.84 | 0.87 | 0.92 | 0.89 | 0.90 |
| | Recall | 0.91 | 0.89 | 0.95 | 0.90 | 0.85 | 0.81 | 0.86 | 0.91 | 0.89 | 0.89 |
| | F1-score | 0.91 | 0.89 | 0.95 | 0.90 | 0.85 | 0.81 | 0.86 | 0.91 | 0.88 | 0.88 |
| | AVG AUC | 0.89 | 0.90 | 0.92 | 0.91 | 0.88 | 0.83 | 0.87 | 0.90 | 0.93 | |
| | AVG Precision | 0.87 | 0.84 | 0.87 | 0.85 | 0.83 | 0.79 | 0.82 | 0.84 | 0.88 | |
| | AVG Recall | 0.88 | 0.82 | 0.86 | 0.84 | 0.81 | 0.77 | 0.80 | 0.84 | 0.87 | |
| | AVG F1-score | 0.86 | 0.82 | 0.86 | 0.84 | 0.80 | 0.77 | 0.78 | 0.83 | 0.86 | |
| | Number of biomarkers | 36 | 308 | 203 | 454 | 285 | 398 | 342 | 148 | 390 | |

The majority of chemical biomarkers are unique to a given honey bee diseases while others are in common with two or more diseases. A total of 11 chemical biomarkers are shared by all bee diseases (Figure 3).

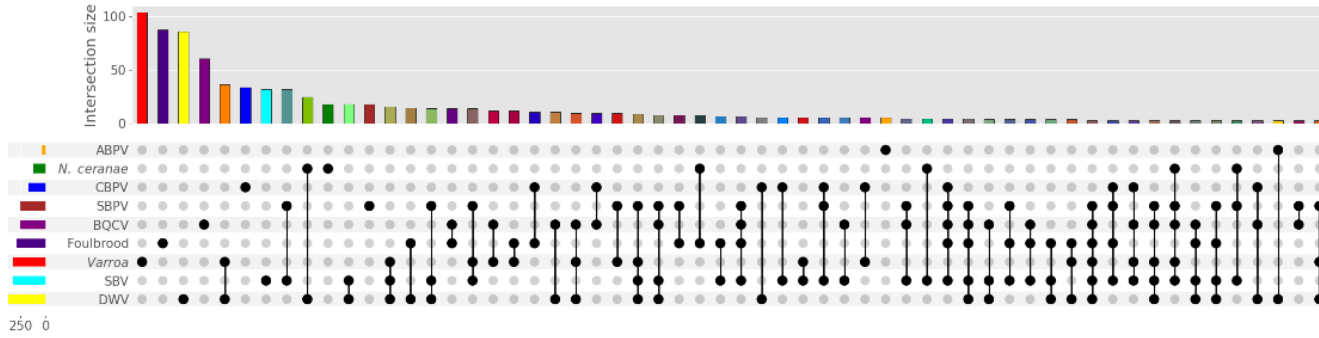



**Figure 3.** The distribution of the chemical biomarker overlap is displayed with each bar color corresponding to a bee disease or a combination of them. The amount of overlap and with which disease is indicated by the lines that connect the black dots. If there is no line and only a black dot, then these chemical biomarkers are unique to this disease which is indicated by the panel to the left. The number of chemical biomarkers within each library, for each disease, is indicated by the horizontal bars to the left of the bee disease label. For clarity, only chemical biomarkers with overlap with three or fewer diseases are displayed.

Furthermore, from the metabolic pathway analysis we identified 43 chemical biomarkers that are significantly associated with a bee disease and have an interaction with one or more other metabolic pathways. From this analysis, Ethanolamine was the top hub metabolite, which was associated with a SBPV and SBV infection and this

metabolite was connected to almost all other metabolites in the network that are significantly associated with the other diseases measured. Therefore, it has the highest centrality and connectedness within the network. Other notably hub metabolites include Lithocholic acid, Formic acid, Oleic acid, Guanidineacetic acid, Eicosapentaenoic acid (EPA), and Cortisone, which are mostly associated with *Varroa* mites and multiple viral infections (Figure 4, Table 2).

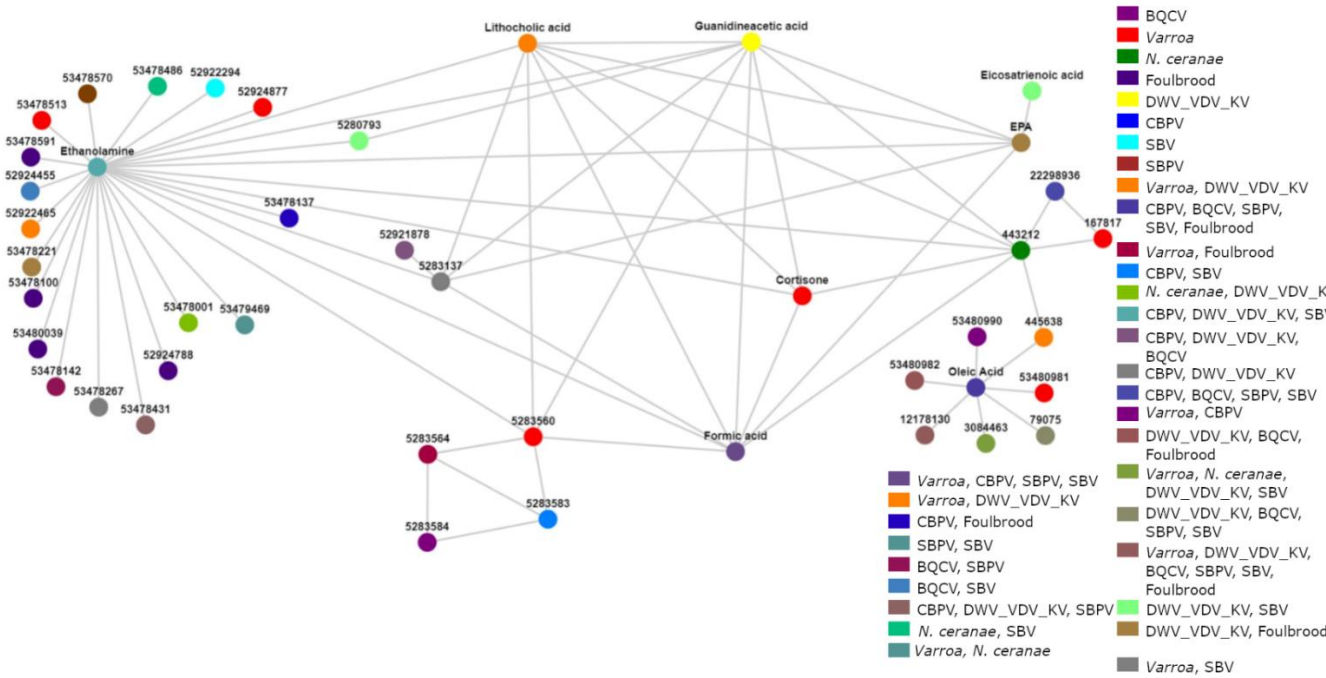



**Figure 4.** Each node represents a chemical biomarker that is a metabolite and part of a known metabolic pathway. An edge indicates that the chemical biomarkers consists of two metabolic pathways that interact with one another. The color of the node indicates the disease or diseases that the chemical biomarker is significantly associated with. If chemical names were unknown in the Pubchem database, then the CAS number of the metabolite is used instead as the identifying information.

**Table 2.** List of chemicals from the biomarker libraries that were found to be affecting the same biological pathway. The compound name or unique identifying code is presented along with the diseases that it is associated with in terms of abundance.

| Compound Name | Disease |
|---|---|
| 445638 | DWV_VDV_KV, *Varroa* |
| 443212 | *N. ceranae* |
| Oleic Acid | SBPV, Foulbrood, BQCV, CBPV, SBV |
| 5283564 | *Varroa*, Foulbrood |
| 5283560 | *Varroa* |
| 5283583 | SBV, CBPV |

| | |
|---|---|
| 5283584 | BQCV |
| 53478001 | *DWV_VDV_KV, N. ceranae* |
| Ethanolamine | DWV_VDV_KV, SBV, CBPV |
| 52921878 | DWV_VDV_KV, CBPV, BQCV |
| 5283137 | DWV_VDV_KV, CBPV |
| 22298936 | SBPV, SBV, CBPV, BQCV |
| 167817 | *Varroa* |
| 53480990 | *Varroa*, CBPV |
| 53480982 | DWV_VDV_KV, BQCV, Foulbrood |
| 3084463 | DWV_VDV_KV, Varroa, SBV, N. ceranae |
| 79075 | DWV_VDV_KV, SBV, SBPV, BQCV |
| 12178130 | DWV_VDV_KV, *Varroa*, SBPV, Foulbrood, BQCV, SBV |
| 53480981 | *Varroa* |
| Eicosatrienoic acid | DWV_VDV_KV, SBV |
| EPA | DWV_VDV_KV, Foulbrood |
| 5280793 | DWV_VDV_KV, SBV |
| Guanidineacetic acid | DWV_VDV_KV |
| 53478267 | *Varroa*, SBV |
| Formic acid | SBPV, *Varroa*, CBPV, SBV |
| Lithocholic acid | DWV_VDV_KV, *Varroa* |
| 52922294 | SBV |
| 52922465 | DWV_VDV_KV, *Varroa* |
| 53478100 | Foulbrood |
| 53478137 | CBPV, Foulbrood |
| 53478142 | SBPV, BQCV |
| 52924455 | SBV, BQCV |
| 52924877 | Varroa |
| 53478221 | DWV_VDV_KV, Foulbrood |
| 52924788 | Foulbrood |
| 53478431 | DWV_VDV_KV, CBPV, SBPV |
| 53478486 | *N. ceranae*, SBV |
| 53478513 | *Varroa* |
| 53479469 | SBPV, SBV |
| 53478570 | *Varroa, N. ceranae* |
| 53478591 | Foulbrood |
| 53480039 | Foulbrood |
| Cortisone | *Varroa* |

In addition, according to PubChem database there were 23 chemical biomarkers associated with bee diseases that interacted with 40 genes that have a known biological function. Diethylinitrosamine had the most interactions, but whether this chemical

biomarker is associated with an upregulation or downregulation of these genes remains unknown. Formic acid, Oleic acid, EPA, and Lithocholic acid are not only connected to multiple metabolic pathways demonstrated from the metabolite pathway analysis, but they also are affecting or upregulating a number of genes as well. Styrenes and polystyrenes are known to be associated with toxic microplastics and they are connected with the downregulation with three different genes, while Polychlorinated Biphenyls, the second largest hub chemical biomarker is known to be a toxic pollutant that remains stable in the environment for a long time. The Polychlorinated Biphenyls are associated with an upregulation of a number of detoxification, oxidative stress, autophagy regulation genes. In general, the genes listed are involved with cellular aerobic respiration, oxidative stress, aging, translation, and autophagy regulation (Figure 5, Table 3).

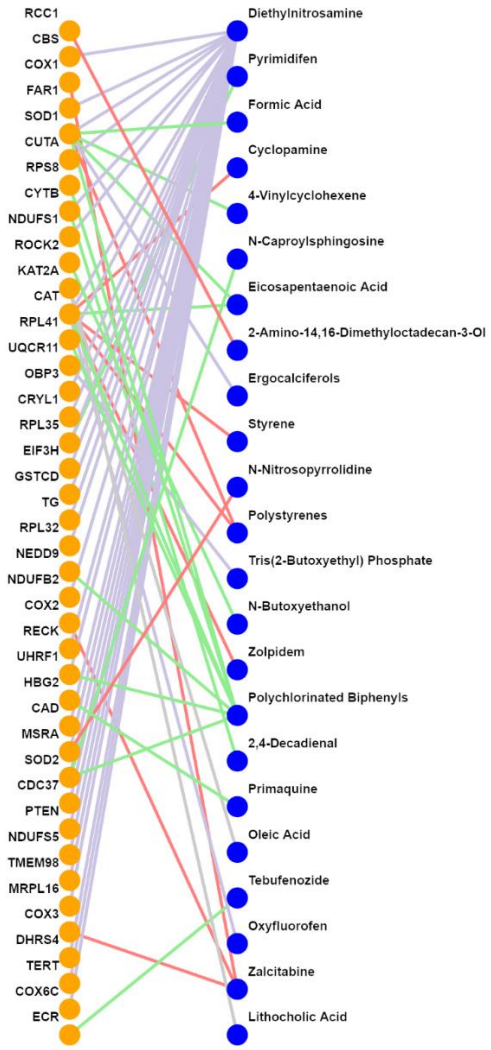



**Figure 5.** A digraph that shows the chemical biomakers (blue nodes) that affect the expression of honey bee genes (orange nodes) as indicated by the edges in the network. Grey edges indicate that the effect is unknown. Green edges indicate that there is an upregulation of gene expression while the red edges indicate there is a downregulation in gene expression. Full gene names can be found in the GO analysis enrichment table 3.

**Table 3.** A list of the genes that are interacting with the chemical biomarkers with their associated biological functions. The biological functions were determined based on a

gene ontology enrichment analysis using the PANTHER database.

| Gene abbreviation | Gene Name | PANTHER GO-slim Molecular Function | GO ID |
|---|---|---|---|
| CAD | Carbamoyl-Phosphate Synthetase 2 | DNA-binding transcription factor activity | GO:0003700 |
| | Carbamoyl-Phosphate Synthetase 3 | RNA polymerase II transcription regulatory region sequence-specific DNA binding | GO:0000977 |
| CAT | Catalase | oxidoreductase activity, acting on peroxide as acceptor | GO:0016684 |
| | Catalase | heme binding | GO:0020037 |
| Cbs | Cystathionine beta-synthase | anion binding | GO:0043168 |
| | Cystathionine beta-synthase | heterocyclic compound binding | GO:1901363 |
| | Cystathionine beta-synthase | organic cyclic compound binding | GO:0097159 |
| | Cystathionine beta-synthase | small molecule binding | GO:0036094 |
| CDC37 | Hsp90 co-chaperone Cdc37 | heat shock protein binding | GO:0031072 |
| | Hsp90 co-chaperone Cdc38 | unfolded protein binding | GO:0051082 |
| | Hsp90 co-chaperone Cdc39 | chaperone binding | GO:0051087 |
| COX1 | Cytochrome c oxidase subunit 1 | proton transmembrane transporter activity | GO:0015078 |
| | Cytochrome c oxidase subunit 1 | electron transfer activity | GO:0009055 |
| COX2 | Cytochrome c oxidase subunit 2 | proton transmembrane transporter activity | GO:0015078 |
| | Cytochrome c oxidase subunit 2 | electron transfer activity | GO:0009055 |
| COX3 | Cytochrome c oxidase subunit 3 | electron transfer activity | GO:0009055 |
| | Cytochrome c oxidase subunit 3 | active transmembrane transporter activity | GO:0022804 |
| ECR | Ecdysone receptor | RNA polymerase II cis-regulatory region sequence-specific DNA binding | GO:0000978 |
| | Ecdysone receptor | DNA-binding transcription factor activity, RNA polymerase II-specific | GO:0000981 |
| EIF3H | Eukaryotic translation initiation factor 3 subunit H | translation initiation factor activity | GO:0003743 |
| | Eukaryotic translation initiation factor 3 subunit H | metallopeptidase activity | GO:0008237 |
| MRPL16 | 39S ribosomal protein L16, mitochondrial | structural constituent of ribosome | GO:0003735 |
| | 39S ribosomal protein L16, mitochondrial | rRNA binding | GO:0019843 |
| MSRA | Peptide methionine sulfoxide reductase | oxidoreductase activity, acting on a sulfur group of donors | GO:0016667 |
| PDHB | Pyruvate dehydrogenase E1 component subunit beta | oxidoreductase activity | GO:0016491 |
| PTEN | IP16020p | phosphatase activity | GO:0016791 |
| RECK | FI21605p1 | protease binding | GO:0002020 |
| | FI21605p2 | endopeptidase inhibitor activity | GO:0004866 |
| | FI21605p3 | endopeptidase activity | GO:0004175 |
| RPL35 | 60S ribosomal protein L35 | structural constituent of ribosome | GO:0003735 |
| | 60S ribosomal protein L36 | mRNA binding | GO:0003729 |
| RPS8 | 40S ribosomal protein S8 | structural constituent of ribosome | GO:0003735 |
| SOD1 | Superoxide dismutase [Cu-Zn] | oxidoreductase activity | GO:0005507 |
| | Superoxide dismutase [Cu-Zn] | copper ion binding | GO:0005507 |
| SOD2 | Superoxide dismutase [Mn], mitochondrial | oxidoreductase activity | GO:0016491 |
| | Superoxide dismutase [Mn], mitochondrial | transition metal ion binding | GO:0046914 |
| TG | TGc domain-containing protein | catalytic activity, acting on a protein | GO:0140096 |
| | TGc domain-containing protein | transferase activity, transferring acyl groups | GO:0016746 |

| Gene abbreviation | Gene Name | PANTHER GO-slim Biological Process: | GO ID |
|---|---|---|---|
| ATP6 | ATP synthase membrane subunit 6 | ATP synthesis coupled proton transport | GO:0015986 |
| CAD | Carbamoyl-Phosphate Synthetase 2 | embryo development | GO:0009790 |
| | Carbamoyl-Phosphate Synthetase 3 | transcription by RNA polymerase II | GO:0006366 |
| | Carbamoyl-Phosphate Synthetase 4 | regulation of transcription by RNA polymerase II | GO:0006366 |
| | Carbamoyl-Phosphate Synthetase 5 | anterior/posterior pattern specification | GO:0009952 |
| | Carbamoyl-Phosphate Synthetase 6 | cell differentiation | GO:0030154 |
| | Carbamoyl-Phosphate Synthetase 7 | animal organ morphogenesis | GO:0030154 |
| CAT | Catalase | response to oxygen-containing compound | GO:1901700 |
| | Catalase | cellular catabolic process | GO:0044248 |
| | Catalase | response to inorganic substance | GO:0010035 |
| | Catalase | response to oxidative stress | GO:0006979 |
| Cbs | Cystathionine beta-synthase | sulfur compound metabolic process | GO:0006790 |
| | Cystathionine beta-synthase | alpha-amino acid metabolic process | GO:1901605 |
| | Cystathionine beta-synthase | cellular biosynthetic process | GO:0044249 |
| CDC37 | Hsp90 co-chaperone Cdc37 | protein folding | GO:0006457 |
| | Hsp90 co-chaperone Cdc38 | protein stabilization | GO:0050821 |
| COX1 | Cytochrome c oxidase subunit 1 | respiratory electron transport chain | GO:0022904 |
| | Cytochrome c oxidase subunit 2 | proton transmembrane transport | GO:1902600 |
| | Cytochrome c oxidase subunit 3 | aerobic respiration | GO:0009060 |
| COX2 | Cytochrome c oxidase subunit 2 | ATP synthesis coupled proton transport | GO:0015986 |
| | Cytochrome c oxidase subunit 3 | ATP synthesis coupled electron transport | GO:0042773 |
| COX3 | Cytochrome c oxidase subunit 3 | aerobic respiration | GO:0009060 |
| ECR | Ecdysone receptor | negative regulation of transcription by RNA polymerase II | GO:0000122 |
| | Ecdysone receptor | transcription by RNA polymerase II | GO:0006366 |
| | Ecdysone receptor | positive regulation of transcription by RNA polymerase II | GO:0045944 |
| | Ecdysone receptor | cell differentiation | GO:0030154 |
| MSRA | Peptide methionine sulfoxide reductase | cellular response to oxidative stress | GO:0034599 |
| PDHB | Pyruvate dehydrogenase E1 component subunit beta | acyl-CoA metabolic process | GO:0006637 |
| | Pyruvate dehydrogenase E1 component subunit beta | monocarboxylic acid metabolic process | GO:0032787 |
| | Pyruvate dehydrogenase E1 component subunit beta | purine ribonucleotide biosynthetic process | GO:0009152 |
| | Pyruvate dehydrogenase E1 component subunit beta | amide biosynthetic process | GO:0043604 |
| PTEN | IP16020p | dephosphorylation | GO:0016311 |
| RECK | FI21605p1 | extracellular matrix organization | GO:0030198 |
| | FI21605p2 | proteolysis | GO:0006508 |
| | FI21605p3 | negative regulation of endopeptidase activity | GO:0010951 |
| ROCK2 | FI21605p4 | cellular protein metabolic process | GO:0044267 |
| RPL35 | 60S ribosomal protein L35 | maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA) | GO:0000463 |
| RPS8 | 40S ribosomal protein S8 | maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA) | GO:0000462 |
| SOD1 | Superoxide dismutase [Cu-Zn] | cellular response to oxidative stress | GO:0034599 |
| | Superoxide dismutase [Cu-Zn] | cellular response to oxygen-containing compound | GO:1901701 |
| | Superoxide dismutase [Cu-Zn] | cellular metabolic process | GO:0044237 |
| | Superoxide dismutase [Cu-Zn] | response to toxic substance | GO:0009636 |
| | Superoxide dismutase [Cu-Zn] | response to inorganic substance | GO:0010035 |
| TG | TGc domain-containing protein | cellular protein modification process | GO:0006464 |

| Gene abbreviation | Gene Name | PANTHER GO-slim Cellular Component: | GO ID |
|---|---|---|---|
| ATP6 | ATP synthase membrane subunit 6 | proton-transporting ATP synthase complex | GO:0045259 |
| CAD | Carbamoyl-Phosphate Synthetase 2 | nucleus | GO:0005634 |
| CAT | Catalase | cytoplasm | GO:0005737 |
| CBS | GRIP domain-containing protein | vacuole | GO:0005773 |
| | GRIP domain-containing protein | Golgi apparatus | GO:0005794 |
| | GRIP domain-containing protein | plasma membrane | GO:0005886 |
| Cbs | Cystathionine beta-synthase | cytoplasm | GO:0005737 |
| CDC37 | Hsp90 co-chaperone Cdc37 | cytoplasm | GO:0005737 |
| CYTB | Cytochrome b | intrinsic component of membrane | GO:0031224 |
| EIF3H | Eukaryotic translation initiation factor 3 subunit H | polysome | GO:0005844 |
| | Eukaryotic translation initiation factor 3 subunit H | cytosolic small ribosomal subunit | GO:0022627 |
| | Eukaryotic translation initiation factor 3 subunit H | eukaryotic translation initiation factor 3 complex | GO:0005852 |
| GSTCD | Glutathione S-transferase C-terminal domain-containing protein homolog | cytoplasm | GO:0005737 |
| MSRA | Peptide methionine sulfoxide reductase | cytoplasm | GO:0005737 |
| ND6 | NADH-ubiquinone oxidoreductase chain 6 | mitochondrion | GO:0005739 |
| PTEN | IP16020p | cytosol | GO:0005829 |
| RECK | FI21605p1 | plasma membrane | GO:0005886 |
| RPL32 | 60S ribosomal protein L32 | cytosolic large ribosomal subunit | GO:0022625 |
| RPL35 | 60S ribosomal protein L35 | cytosolic large ribosomal subunit | GO:0022625 |
| RPS8 | 40S ribosomal protein S8 | cytosolic small ribosomal subunit | GO:0022627 |
| SOD1 | Superoxide dismutase [Cu-Zn] | extracellular space | GO:0005615 |

The 62 chemical biomarkers identified are associated with many different human health disorders as well. These associations include both mental and physical health disorders. For example, some of the disorders listed include: neurological disorders, pain, drug overdoses, fever, seizures, influenza, pneumonia, cancer, cardiovascular diseases, genetic disorders and infertility (Table A1).

## 4. DISCUSSION

Using a systems biology approach that integrated the exposome of beehives with disease abundance and prevalence we are able to identify a custom chemical biomarker library that is able to predict common bee diseases with around 85% accuracy, sensitivity, selectivity, precision, and recall that is derived from the exposome of the bee hive. The number of chemical biomarkers within a library of a disease range from 36 to 454, which suggests that more biomarkers do not necessarily increase the library's performance in predicting whether the hive is infected or uninfected with a particular bee disease. The relative consistency in performance across the five different machine learning algorithms highlights the fact that these chemical biomarkers are making relatively robust predictions. In summary, the performance of the chemical biomarkers is comparable to previous study derived from lab experiments focusing on predicting nutritional stress in bumble bees (Wang et al 2019). The majority bee health studies that propose potential biomarkers do not study their strength of association across varying levels of the phenotype sampled from the natural environment (Wild et al 2005). Although the performance of these chemical biomarker libraries is promising based on

the machine learning validation, especially given that these significant associations are drawn from bee samples collected from the field, these chemical biomarkers could be further validated empirically to determine if there is a cause-and-effect relationship between the bee disease and the chemical biomarker. From this study, we are unable to disentangle whether the change in the relative abundance of the chemical biomarker is due to the bee disease infection or whether chemical biomarker itself is causing the honey bee to be more prone to a particular bee disease. The majority of these appear to be an environmental exposure as opposed to an endogenous change in the honey bee's metabolite profile, which is also captured with an exposomic analysis (Mayack, 2022). Most of the chemical biomarkers identified have not been previously associated with a decline in bee health or bee diseases previously revealing novel connections and underlying mechanisms that are likely driving the most recent decline in bee health.

We could develop chemical biomarker libraries for only 13 out of the 20 common bee diseases, but the low prevalence of the other diseases is the reason that they could not be included in the analysis. Therefore, we assume that these other bee diseases are not playing a major role in the most recent decline in bee health. This assumption is substantiated by previous findings as well, based on the relative prevalence of these bee diseases in honey bees around the world (Genersch, 2010). In general, Varroa mites and Deformed Wing Virus are known to be globally distributed with a high level of prevalence (Wilfert et al 2016), this is also the case for *N. ceranae*, but this is in contrast to *N. apis*, which has largely been replaced by *N. ceranae* (Chen, 2012). In recent years Israeli Acute Paralysis virus was once thought to be a major driver in bee colony collapses, but it is now rarely detected in unhealthy bee hives (Molineri, 2017).

Most chemical biomarkers are unique to each disease, while others have overlap with two or more bee diseases. This suggests that certain general biomarkers may be selected that might indicate that a bee hive is sick in general with many diseases. On the other hand, most chemical biomarkers are specific for one disease only increasing their utility in diagnosing a particular bee disease and reducing the chances of producing false positives. For example, *Varroa* mites are known to vector Deformed Wing Virus and therefore many times they co-occur in the bee hive environment (Adler et al. 2021), despite this over 100 chemical biomarkers are unique to a Varroa mite infestation and do not overlap with Deformed Wing Virus chemical biomarkers. The overlap between different bee diseases could signify that the change in the relative abundance of this

chemical may play a critical role in the increased susceptibility to multiple infections. However, further research is needed to determine the connection between the biological effect of the chemical and its relation to increased susceptibility to multiple infections.

How the chemical exposures are specifically interacting with bee diseases were previously unknown even though we know multiple stressors have been key in causing a decline in bee health (Botías et al 2021). Based on our metabolic pathway network analysis Ethanolamine was clearly the top hub metabolite suggesting that the regulation of this chemical could be key in maintaining bee health in general. Sac Brood Virus and Slow Bee Paralysis Virus, which are significantly associated with the dysregulation of this chemical abundance, may therefore act as a catalyst for other bee infections and consequently result in a synergistic decline in bee health. In the honey bee ethanolamine is involved in amino acid metabolism and glycan biosynthesis. Ethanolamine is a metabolite of the amino acid serine and can be produced from the breakdown of the phospholipid membrane, which in turn can act as a nitrogen source for bacteria living in the honey bee gut. Ethanolamine can increase in the gut due to the high turnover of bacteria and host cells (Ellegaard 2019) and this is interesting to note because *Nosema* infected honey bees had higher levels of serine in comparison to uninfected bees (Wang 1970). Pathogenic bacteria have been shown to utilize ethanolamine in the mammalian gut (Tsoy et al 2009), but the exact function in the honey bee remains unknown. Nonetheless, taken together it is plausible that ethanolamine production may serve as a reliable biomarker for when the honey bee is in infected state. Ethanolamine is also known to vary based on pollen diet that is fed to bee larvae (Singh and Singh 1996), so the amount of it is likely to vary in time and CDP-ethanolamine which is responsible for the *de novo* synthesis of phosphatidylethanolamine, is found in lower levels in the bee brain of bees infected with *Varroa* mites (Wu et al 2017).

A number of acids were also considered to be hub metabolites that are linked to other metabolic pathways. Fatty acids such as oleic acid and formic acid are used to treat *Varroa* and tracheal mites in honey bees (Vilarem et al 2021). Therefore, the higher abundance of this chemical could be due to the beekeepers using these treatments as a management strategy once *Varroa* is detected visually in the hive. These are also known as plant derived acids and as a result are used as organic beekeeping treatment for these mites (Genath, 2021). Oleic acid levels also increase from dead honey bees and this acts as a pheromone triggering bees to perform hygienic behavior and remove

dead and dying bees from the hive (McAfee et al 2018). Eicosapentaenoic acid is an omega fatty acid and an omega 6:3 ratio of 5 is known to be detrimental to honey bee associative learning (Arien et al. 2018). Therefore, multiple infections such as with European Foulbrood and DWV, which is what EPA is associated with, may disrupt this omega fatty acid balance and also have an effect of the cognitive functioning of the honey bee.

Lithocholic acid on the other hand is a toxic bile acid and is a known carcinogen (Sivcev et al 2020). Guanidine acetic acid is be added to animal feed for fattening before slaughter (Khakran et al 2018) and it also the pre-cursor to creatine, which aids in muscle energy recovery and the buildup of muscle mass (Antonio et al 2021). Cortisone is a well-known anti-inflammatory steroid hormone (Martínez-Urbistondo et al 2021), but connections of these metabolites to bee health are currently unknown.

Diethylnitrosamine (DEN) is a toxic material to mammals that can cause liver necrosis and carcinomas (Singh et al 2021). Here we show that it has a number of effects on honey bee genes, which many are involved in detoxification and oxidative stress pathways. Diethylnitrosamine is extensively found in milk products, meat products, soft drinks, alcoholic beverages, and tobacco smoke (Owumi et al 2021). Although not previously recognized, this pollutant may have severe health impacts for honey bees as well increasing its susceptibility to bee diseases and increased mortality from multiple stressors. Polychlorinated Biphenyls (PCBs) are another toxic pollutant that is a known carcinogen and has a multitude of detrimental human health effects (Christensen et al 2021), it is an extremely stable compound and is therefore pervasive throughout the natural environment (Gabryszewska et al 2020). PCBs are commonly used as dielectric and refrigerant fluids in electrical appliances, paints, and plasticizers (Sari et al 2021). Previous studies show that PCB exposure is regularly detected in honey bee hives and the workers have higher concentrations than the bee products within the hives. However, previously honey bees were only considered as bioindicators of the environment for human health purposes (Ćirić et al 2021). Our results suggest that they are having a negative impact on bee health as well as PCBs relate to 4 different upregulated genes associated with detoxification and oxidative stress. These toxic pollutants are having a relatively higher impact on gene regulation in comparison to the other chemical biomarkers identified, which suggests that it would be interesting to investigate their effects on bee health in more detail in the future. With this said, honey bees in this context are also valuable sentinels or bioindicators of environmental quality

as well. This is reflected by the vast number of associations of human health diseases that were made from the Pubchem database with the different chemical biomarkers we identified to be associated with bee diseases. If these biomarkers were to be screened in the future to assess bee health at the same time it could be used to monitor environmental quality for human health as well.

In summary, using a systems biology analysis in which the exposome data was integrated with disease abundance and prevalence datasets to look for significant interactions among them we were able to identify chemical biomarker libraries that are significantly associated with common bee diseases, which are known to play a role in the most recent decline in bee health. We identified potential metabolic pathway interactions across the bee diseases in which putative targets could be investigated further to determine if they result in a synergistic decline in bee health. Furthermore, we identified new environmental exposures and metabolic pathways that are associated with honey bee diseases that have not been previously linked with one another. Toxic environmental pollutants to humans appear to also have a potential damaging effect on bee health and as we consider these exposures in terms of environmental quality this supports the notion that a holistic and optimal framework to be applied is a One Health approach which is creating and executing programs, policies, laws, and research in which diverse sectors communicate and work together to improve bee health. (M. M. Lopez-Uribe, 2020).

## 5. BIBLIOGRAPHY


A. Al-Chalabi, N. P. (2015). Mapping the Human Exposome Without It, How Can We Find Environmental Risk Factors for ALS? *Epidemiology 26*, 821-823.

A.J. McMenamin, E. G. (2008). Bee colony losses and associated viruses . *Curr.Opin.Insect.sci 8*, 121-129.

Adler, L. S. (2021). Floral traits affecting the transmission of beneficial and pathogenic pollinator-associated microbes. *Current Opinion in Insect Science 44*, 1-7.

Antonio, J. D. (2021). Common questions and misconceptions about creatine supplementation: what does the scientific evidence really show?. *Journal of the International Society of Sports Nutrition 18, no. 1* , 1-17.

Aronson, J. K. (2005). *British Journal of Clinical Pharmacology, 59*, 491-494.

Badiou-Bénéteau, A. S.-L. (2012). Development of biomarkers of exposure to xenobiotics in the honey bee Apis mellifera: application to the systemic insecticide thiamethoxam. *Ecotoxicology and environmental safety 82* , 22-31.

Baker, M. (2005). In biomarkers we trust? *Nat Biotechnol 23*, 297–304.

Barbosa Jr, F. J.-S. (2005). A critical review of biomarkers used for monitoring human exposure to lead: advantages, limitations, and future needs. *Environmental health perspectives 113*, 1669-1674.

Batista, G. &. (2003). An Analysis of Four Missing Data Treatment Methods for Supervised Learning. *Applied Artificial Intelligence vol. 17*, 519-533.

Borycka et al, D. G.-L. (2015). Antioxidant and antibacterial properties of commercial bee pollen products. *Journal of Apicultural Research*, 491-502.

Botías, C. J. (2021). Multiple stressors interact to impair the performance of bumblebee Bombus terrestris colonies. *Journal of Animal Ecology 90, no. 2*, 415-431.

Bouckaert, R. (2003). Choosing between two learning algorithms based on calibrated tests. *Proc 20th Int Conf on Machine Learning*, 51-58.

Breeze et al, T. D. (2011). Pollination services in the UK: How important are honeybees? *Agriculture, Ecosystems & Environment*, 137-143.

Breiman L., F. J. (1984). Classification and Regression Trees. *Wadsforth International Group*.

Burges, C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery 2(2)*, 1-47.

C. P. Wild, A. S. (2013). Measuring the exposome: A powerful basis for evaluating environmental exposures and cancer risk. *Environ. Mol. Mutagen. 54*, 480-499.

C. T. Yildiz, E. A. (2001). Omnivariate decision trees. *IEEE Transactions on Neural Networks vol. 12, no. 6*, 1539-1546.

Cestnik, B. K. (1987). Assistant 86: A knowledge elicitation tool for sophisticated users. *Proceedings of the Second European Working Session on Learning*, 31-45.

Chen, Y.-W. W.-P.-H.-F. (2012). Nosema ceranae infection intensity highly correlates with temperature. *Journal of invertebrate pathology 111, no. 3*, 264-267.

Christensen, K. L. (2021). The role of epidemiology studies in human health risk assessment of polychlorinated biphenyls. *Environmental Research 194*.

Ćirić, J. D. (2021). Honey bees and their products as indicators of environmental element deposition. *Biological Trace Element Research 199, no. 6*, 2312-2319.

Devereux, R. B. (1983). Left ventricular hypertrophy in patients with hypertension: importance of blood pressure response to regularly recurring stress. *Circulation 68, no. 3*, 470-476.

Dietterich, T. G. (1998). Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation 10(7)*, 1895–1924.

Domingos, P. &. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning 29*, 103-130.

E. Collison, H. H. (2016). Interactive effects of pesticide exposure and pathogen infection on bee health - a critical analysis. *Biol. Rev. 91*, 1006-1019.

Freund, Y. S. (1999). Large Margin Classification Using the Perceptron Algorithm. *Machine Learning 37*, 277-296.

Gabryszewska, M. B. (2020). Impact of municipal and industrial waste incinerators on PCBs content in the environment. *PloS one 15, no. 11*.

Galathiya, A. G. (2012). Improved decision tree induction algorithm with feature selection, cross validation, model complexity and reduced error pruning. *International Journal of Computer Science and Information Technologies, 3(2)*, 3427-3431.

Genath, A. H. (2021). Influence of formic acid treatment on the proteome of the ectoparasite Varroa destructor. *PloS one 16, no. 10*.

Genersch. (2010). Honey bee pathology: current threats to honey bees and beekeeping. *Applied microbiology and biotechnology 87(1)*, 87-97.

Genton, M. (2001). Classes of Kernels for Machine Learning: A Statistics Perspective. *Journal of Machine Learning Research 2*, 299-312.

Goulson et al, D. N. (2015). Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science 347 (6229)*.

Grozinger, C. M. (2020). Improving bee health through genomics. *Nature Reviews Genetics 21, no. 5*, 277-291.

Hamiduzzaman, M. G.-N. (2010). A multiplex PCR assay to diagnose and quantify Nosema infections in honey bees (Apis mellifera). *Journal of Invertebrate Pathology 105*, 151-155.

Hasin, Y. S. (2017). Multi-omics approaches to disease. *Genome biology 18(1)*, 1-15.

Huaiyu Mi, D. E.-P. (2021). PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive AP. *Nucleic Acids Research, Volume 49, Issue D1*, D394–D403.

Hunt E., M. J. (1966). Experiments in Induction. *New York, Academic Press*.

Ingram, D. K. (2001). *Experimental Gerontology, 36*.

Jain, A. M. (1999). Data clustering: A review,. *ACM Computing Surveys*, 264-232.

Japkowicz N., S. S. (2002). he Class Imbalance Problem: A Systematic Study. *Intelligent Data Analysis Volume 6, Number 5*.

John Joseph Valletta, C. T. (2017,). Applications of machine learning in animal behaviour studies,. *Animal Behaviour,Volume 124,*, 203-220.

Keerthi, S. ,. (2002). Convergence of a Generalized SMO Algorithm for SVM Classifier Design. *Machine Learning 46*, 351-360.

Kim S, C. J. (2021). PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res. 49(D1)*, D1388–D1395.

Kluser et al, S. P.-P. (2010). Global honey bee colony disorders and other threats to insect pollinators.

Kononenko, I. L. (1994). Estimating attributes: analysis and extensions of Relief. *Machine Learning: ECML-94*, 171–182.

Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering 160, no. 1*, 3-24.

Kraus, V. B. (2018). Biomarkers as drug development tools: discovery, validation, qualification and use. *Nature Reviews Rheumatology 14, no. 6* , 354-362.

Liu, H. a. (2001). Instance Selection and Constructive Data Mining. *Data Mining and Knowledge Discovery 4*, 315–344.

Liu, R. X. (2014). Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers. *Medicinal research reviews 34, no. 3*, 455-478.

Lopez-Uribe, M. R.-F. (2020). Defining Pollinator Health: A Holistic Approach Based on Ecological, Genetic, and Physiological Factors. *Lewin, H.A., Roberts, R.M. (Eds.). Annual Review of Animal Biosciences, Vol 8, 2020. Annual Reviews, Palo Alto*, 269-294.

M. D. Abramoff, P. J. (2004). Image processing with ImageJ. *Biophotonics International 11*, 36-42.

M. M. Lopez-Uribe, V. A.-F. (2020). in Annual Review of Animal Biosciences. *(Annual Reviews, Palo Alto, 2020), vol. 8*, 269-294.

Markovitch S, R. D. (2002). Feature Generation Using General Construction Functions. *Machine Learning 49*, 59-98.

Martínez-Urbistondo, D. R. (2021). Early combination of tocilizumab and corticosteroids: An upgrade in anti-inflammatory Therapy for severe coronavirus disease (COVID). *Clinical Infectious Diseases 72, no. 9*, 1682-1683.

Mayack, C. (2022). Environmental exposures associated with honey bee health. *Chemosphere 131948* .

Molineri, A. P.-C. (2017). Prevalence of honey bee (Apis mellifera) viruses in temperate and subtropical regions from Argentina. *Revista Argentina de Microbiología 49*, 166-173.

Moritz, R. F. (2010). Research strategies to improve honeybee health in Europe. *Apidologie 41, no. 3* , 227-242.

Mullin, e. a. (2010). High Levels of Miticides and Agrochemicals in North American Apiaries: Implications for Honey Bee Health. *Plos One 5, 19*.

Murthy. (1998). Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey. *Data Mining and Knowledge Discovery 2*, 345–389.

Nasri, T. e. (2015). PCR-RFLP on β-Tubulin gene for rapid identification of the most clinically important species of Aspergillus. *Journal of microbiological methods 117*.

Owumi, S. E. (2021). Leaf paste of Telfairia occidentalis favourably modulates deleterious effects associated with exposure to diethylnitrosamine in male Wistar rats. *Journal of Complementary and Integrative Medicine* .

P. A. Macedo, J. W. (2002). Using inert dusts to detect and assess varroa infestations in honey bee colonies. *Journal of Apicultural Research 41*, 3-7.

P. Glasziou, L. I. (2008). Evidence-based medical monitoring:From principles to practice. *Oxford:Blackwell Publishing.*, 48-62.

Patel, V. P. (2021). Why bees are critical for achieving sustainable development. *Ambio 50*, 49–59.

Pedregosa, F. a. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 2825-2830.

Phillips, C. V. (2004). The missed lessons of sir Austin Bradford Hill. *Epidemiologic Perspectives & Innovations 1, no. 1*, 1-5.

Refaeilzadeh, P. L. (2009). "Cross-validation.". *Encyclopedia of database systems 5*, 532-538.

S. Dagnino, A. M. (2019). Unraveling the Exposome: A Practical View. *Springer International Publishing, ed. 1*, 447.

S. M. Rappaport, M. T. (2010). Environment and Disease Risks. *Science 330*.

Sari, M. F. (2021). Levels of polychlorinated biphenyls (PCBs) in honeybees and bee products and their evaluation with ambient air concentrations. *Atmospheric Environment 244*.

Singh, D. D. (2021). Amelioration of diethylnitrosamine (DEN) induced renal oxidative stress and inflammation by Carissa carandas embedded silver nanoparticles in rodents. *Toxicology Reports 8*, 636-645.

Sivcev, S. B. (2020). Lithocholic acid inhibits P2X2 and potentiates P2X4 receptor channel gating. *The Journal of steroid biochemistry and molecular biology 202* .

Siviter, H. B. (2021). Agrochemicals interact synergistically to increase bee mortality. *Nature, 596(7872)*, 389-392.

Smet., L. D. (2012). BeeDoctor, a Versatile MLPA-Based Diagnostic Tool for Screening Bee Viruses. *PLoS ONE 7, e47953*.

Szefler, S. J. (2012). Asthma outcomes: biomarkers. *Journal of Allergy and Clinical Immunology 129, no. 3*, S9-S23.

T.Reinartz. (2002). A Unifying View on Instance Selection,. *Data Mining and Knowledge Discovery 6*, 191–210.

Vilarem, C. V. (2021). Varroa destructor from the Laboratory to the Field: Control, Biocontrol and IPM Perspectives—A Review. *Insects 12, no. 9* , 800.

Wang, L. M. (2019). Metabolomics-based biomarker discovery for bee health monitoring: A proof of concept study concerning nutritional stress in Bombus terrestris. . *Scientific reports, 9(1)*, 1-11.

Wild, C. P. (2005). Complementing the genome with an “exposome”: the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiology Biomarkers & Prevention 14, 1847*.

Wilfert, e. a. (2016). Deformed wing virus is a recent global epidemic in honeybees driven by Varroa mites. *Science 351*, 594-597.

Wilfert, L. G.-H. (2016). Deformed wing virus is a recent global epidemic in honeybees driven by Varroa mites. *Science 351, no. 6273*, 594-597.

Zhang, S. Z. (2002). Data Preparation for Data Mining. *Applied Artificial Intelligence Volume 17*, 375 - 381.

## 6. APPENDIX

**Table 1A.** List of chemical biomarkers associated with bee diseases and their relationship with human diseases. The data were extracted from the PubChem database. The CID refers to the chemical identification number, while the compound full name is presented to the right. If more than one human disease or disorder corresponds to a chemical compound then it is listed multiple times. The direct evidence refers to how the association was identified, if the chemical is a biomarker for the disease, involved in an underlying mechanism of the disease or is a known therapeutic drug for treating the disease.

| CID | Compound | Disease | Direct Evidence |
|---|---|---|---|
| 443118 | acetylleucyl-leucyl-norleucinal | Seizures | marker/mechanism |
| 65028 | Oseltamivir | Influenza, Human | therapeutic |
| 65028 | Oseltamivir | Neurologic Manifestations | marker/mechanism |
| 65028 | Oseltamivir | Orthomyxoviridae Infections | therapeutic |
| 65028 | Oseltamivir | Stevens-Johnson Syndrome | marker/mechanism |
| 45789647 | methylone | Chemical and Drug Induced Liver Injury | marker/mechanism |
| 45789647 | methylone | Fever | marker/mechanism |
| 45789647 | methylone | Substance-Related Disorders | marker/mechanism |
| 14707 | Phenazocine | Dizziness | marker/mechanism |
| 14707 | Phenazocine | Headache | marker/mechanism |
| 14707 | Phenazocine | Nausea | marker/mechanism |
| 14707 | Phenazocine | Pain, Postoperative | therapeutic |
| 14707 | Phenazocine | Vertigo | marker/mechanism |
| 14707 | Phenazocine | Vomiting | marker/mechanism |
| 14707 | Phenazocine | Xerostomia | marker/mechanism |
| 5283137 | Thromboxane B2 | Brain Injuries | marker/mechanism |
| 5283137 | Thromboxane B2 | Diabetes Mellitus, Experimental | marker/mechanism |

| | | | |
|---|---|---|---|
| 5283137 | Thromboxane B2 | Hypertension | marker/mechanism |
| 5283137 | Thromboxane B2 | Hypertension, Pulmonary | marker/mechanism |
| 5283137 | Thromboxane B2 | Mucocutaneous Lymph Node Syndrome | marker/mechanism |
| 5283137 | Thromboxane B2 | Nephrosis | marker/mechanism |
| 7501 | Latex | Anaphylaxis | marker/mechanism |
| 7501 | Latex | Dermatitis, Allergic Contact | marker/mechanism |
| 7501 | Latex | Pneumonia | marker/mechanism |
| 7501 | Styrene | Acute Disease | marker/mechanism |
| 7501 | Styrene | Breast Neoplasms | marker/mechanism |
| 7501 | Styrene | Chemical and Drug Induced Liver Injury | marker/mechanism |
| 7501 | Styrene | Extravasation of Diagnostic and Therapeutic Materials | marker/mechanism |
| 7501 | Styrene | Eye Diseases | marker/mechanism |
| 7501 | Styrene | Gastrointestinal Diseases | marker/mechanism |
| 7501 | Styrene | Hearing Loss | marker/mechanism |
| 7501 | Styrene | Hyperplasia | marker/mechanism |
| 7501 | Styrene | Lung Injury | marker/mechanism |
| 7501 | Styrene | Lung Neoplasms | marker/mechanism |
| 7501 | Styrene | Lymphoma, Non-Hodgkin | marker/mechanism |
| 7501 | Styrene | Mammary Neoplasms, Animal | marker/mechanism |
| 7501 | Styrene | Neuroectodermal Tumors, Primitive | marker/mechanism |
| 7501 | Styrene | Pneumonia | marker/mechanism |
| 7501 | Styrene | Poisoning | marker/mechanism |
| 7501 | Styrene | Respiratory Tract Diseases | marker/mechanism |
| 7501 | Styrene | Skin Diseases | marker/mechanism |
| 7501 | styrofoam | Hemolysis | marker/mechanism |
| 20299 | tridihexethyl | Colonic Diseases, Functional | therapeutic |
| 20299 | tridihexethyl | Color Vision Defects | marker/mechanism |
| 4201 | Minoxidil | Acromegaly | marker/mechanism |
| 4201 | Minoxidil | Acute Kidney Injury | marker/mechanism |
| 4201 | Minoxidil | Ageusia | marker/mechanism |
| 4201 | Minoxidil | Alopecia | therapeutic |
| 4201 | Minoxidil | Alopecia Areata | therapeutic |
| 4201 | Minoxidil | Anorexia | marker/mechanism |
| 4201 | Minoxidil | Autonomic Nervous System Diseases | marker/mechanism |
| 4201 | Minoxidil | Cardiac Tamponade | marker/mechanism |
| 4201 | Minoxidil | Cardiomegaly | marker/mechanism |
| 4201 | Minoxidil | Cardiomyopathies | marker/mechanism |
| 4201 | Minoxidil | Cardiovascular Diseases | marker/mechanism |
| 4201 | Minoxidil | Chest Pain | marker/mechanism |
| 4201 | Minoxidil | Color Vision Defects | marker/mechanism |
| 4201 | Minoxidil | Dizziness | marker/mechanism |
| 4201 | Minoxidil | Edema | marker/mechanism |
| 4201 | Minoxidil | Erectile Dysfunction | therapeutic |

| | | | |
|---|---|---|---|
| 4201 | Minoxidil | Fatigue | marker/mechanism |
| 4201 | Minoxidil | Fever | marker/mechanism |
| 4201 | Minoxidil | Headache | marker/mechanism |
| 4201 | Minoxidil | Hearing Loss, Bilateral | marker/mechanism |
| 4201 | Minoxidil | Heart Diseases | marker/mechanism |
| 4201 | Minoxidil | Heart Failure | marker/mechanism |
| 4201 | Minoxidil | Heart Failure | therapeutic |
| 4201 | Minoxidil | Hemorrhage | marker/mechanism |
| 4201 | Minoxidil | Hirsutism | marker/mechanism |
| 4201 | Minoxidil | Hypertension | therapeutic |
| 4201 | Minoxidil | Hypertension, Malignant | therapeutic |
| 4201 | Minoxidil | Hypertension, Pulmonary | therapeutic |
| 4201 | Minoxidil | Hypertension, Renal | therapeutic |
| 4201 | Minoxidil | Hypertrichosis | marker/mechanism |
| 4201 | Minoxidil | Hypertrophy, Left Ventricular | marker/mechanism |
| 4201 | Minoxidil | Hypotension | marker/mechanism |
| 4201 | Minoxidil | Hypotension, Orthostatic | marker/mechanism |
| 4201 | Minoxidil | Inflammation | marker/mechanism |
| 4201 | Minoxidil | Ischemia | therapeutic |
| 4201 | Minoxidil | Myocardial Infarction | marker/mechanism |
| 4201 | Minoxidil | Necrosis | marker/mechanism |
| 4201 | Minoxidil | Necrosis | therapeutic |
| 4201 | Minoxidil | Optic Neuritis | marker/mechanism |
| 4201 | Minoxidil | Otitis Externa | marker/mechanism |
| 4201 | Minoxidil | Pain | marker/mechanism |
| 4201 | Minoxidil | Pericardial Effusion | marker/mechanism |
| 4201 | Minoxidil | Pericarditis | marker/mechanism |
| 4201 | Minoxidil | Pleural Effusion | marker/mechanism |
| 4201 | Minoxidil | Polymyalgia Rheumatica | marker/mechanism |
| 4201 | Minoxidil | Proteinuria | marker/mechanism |
| 4201 | Minoxidil | Pruritus | marker/mechanism |
| 4201 | Minoxidil | Renal Insufficiency | marker/mechanism |
| 4201 | Minoxidil | Renal Insufficiency | therapeutic |
| 4201 | Minoxidil | Retinitis | marker/mechanism |
| 4201 | Minoxidil | Tachycardia | marker/mechanism |
| 4201 | Minoxidil | Weight Loss | marker/mechanism |
| 73402 | eburicoic acid | Glioblastoma | therapeutic |
| 5359272 | Levorphanol | Catalepsy | marker/mechanism |
| 5359272 | Levorphanol | Hyperalgesia | therapeutic |
| 5359272 | Levorphanol | Pain | therapeutic |
| 5359272 | Levorphanol | Substance-Related Disorders | marker/mechanism |
| 5359272 | Levorphanol | Substance Withdrawal Syndrome | therapeutic |
| 68712 | Rilmenidine | Arrhythmias, Cardiac | therapeutic |
| 68712 | Rilmenidine | Bradycardia | marker/mechanism |
| 68712 | Rilmenidine | Hypertension | therapeutic |

| | | | |
|---|---|---|---|
| 68712 | Rilmenidine | Hypotension | marker/mechanism |
| 5639 | urapidil | Bronchial Spasm | therapeutic |
| 5639 | urapidil | Dyspnea | therapeutic |
| 5639 | urapidil | Hypertension | therapeutic |
| 5639 | urapidil | Hypotension | marker/mechanism |
| 5639 | urapidil | Postoperative Complications | therapeutic |
| 5639 | urapidil | Pre-Eclampsia | therapeutic |
| 5639 | urapidil | Tachycardia | therapeutic |
| 5280793 | Ergocalciferols | Arteriosclerosis | marker/mechanism |
| 5280793 | Ergocalciferols | Atherosclerosis | marker/mechanism |
| 5280793 | Ergocalciferols | Azotemia | marker/mechanism |
| 5280793 | Ergocalciferols | Familial Hypophosphatemic Rickets | therapeutic |
| 5280793 | Ergocalciferols | Gastrointestinal Hemorrhage | marker/mechanism |
| 5280793 | Ergocalciferols | Hepatitis C | therapeutic |
| 5280793 | Ergocalciferols | Hypercalcemia | marker/mechanism |
| 5280793 | Ergocalciferols | Hypercalciuria | marker/mechanism |
| 5280793 | Ergocalciferols | Hypoparathyroidism | therapeutic |
| 5280793 | Ergocalciferols | Juvenile osteoporosis | therapeutic |
| 5280793 | Ergocalciferols | Muscular Diseases | therapeutic |
| 5280793 | Ergocalciferols | Nephrocalcinosis | marker/mechanism |
| 5280793 | Ergocalciferols | Osteomalacia | therapeutic |
| 5280793 | Ergocalciferols | Osteonecrosis | therapeutic |
| 5280793 | Ergocalciferols | Ulcer | marker/mechanism |
| 5921 | Diethylnitrosamine | Adenoma | marker/mechanism |
| 5921 | Diethylnitrosamine | Adenoma, Bile Duct | marker/mechanism |
| 5921 | Diethylnitrosamine | Adenoma, Liver Cell | marker/mechanism |
| 5921 | Diethylnitrosamine | Aneuploidy | marker/mechanism |
| 5921 | Diethylnitrosamine | Bile Duct Diseases | marker/mechanism |
| 5921 | Diethylnitrosamine | Body Weight Changes | marker/mechanism |
| 5921 | Diethylnitrosamine | Breast Neoplasms | marker/mechanism |
| 5921 | Diethylnitrosamine | Carcinogenesis | marker/mechanism |
| 5921 | Diethylnitrosamine | Carcinoma | marker/mechanism |
| 5921 | Diethylnitrosamine | Carcinoma, Hepatocellular | marker/mechanism |
| 5921 | Diethylnitrosamine | Carcinoma, Renal Cell | marker/mechanism |
| 5921 | Diethylnitrosamine | Carcinoma, Squamous Cell | marker/mechanism |
| 5921 | Diethylnitrosamine | Cardiomyopathies | marker/mechanism |
| 5921 | Diethylnitrosamine | Cell Transformation, Neoplastic | marker/mechanism |
| 5921 | Diethylnitrosamine | Chemical and Drug Induced Liver Injury | marker/mechanism |
| 5921 | Diethylnitrosamine | Chemical and Drug Induced Liver Injury | therapeutic |
| 5921 | Diethylnitrosamine | Cholangiocarcinoma | marker/mechanism |
| 5921 | Diethylnitrosamine | Chromosome Breakage | marker/mechanism |
| 5921 | Diethylnitrosamine | Colonic Neoplasms | marker/mechanism |
| 5921 | Diethylnitrosamine | Cystadenoma | marker/mechanism |
| 5921 | Diethylnitrosamine | Diabetes Mellitus, Type 2 | marker/mechanism |
| 5921 | Diethylnitrosamine | Esophageal Achalasia | marker/mechanism |

| | | | |
|---|---|---|---|
| 5921 | Diethylnitrosamine | Esophageal Neoplasms | marker/mechanism |
| 5921 | Diethylnitrosamine | Fatty Liver | marker/mechanism |
| 5921 | Diethylnitrosamine | Fibrosis | marker/mechanism |
| 5921 | Diethylnitrosamine | Flushing | marker/mechanism |
| 5921 | Diethylnitrosamine | Gastritis | marker/mechanism |
| 5921 | Diethylnitrosamine | Gastrointestinal Diseases | marker/mechanism |
| 5921 | Diethylnitrosamine | Genomic Instability | marker/mechanism |
| 5921 | Diethylnitrosamine | Glucose Intolerance | marker/mechanism |
| 5921 | Diethylnitrosamine | Hemangioma | marker/mechanism |
| 5921 | Diethylnitrosamine | Hepatic Insufficiency | marker/mechanism |
| 5921 | Diethylnitrosamine | Hepatitis | marker/mechanism |
| 5921 | Diethylnitrosamine | Hepatoblastoma | marker/mechanism |
| 5921 | Diethylnitrosamine | Hepatomegaly | marker/mechanism |
| 5921 | Diethylnitrosamine | Hyperplasia | marker/mechanism |
| 5921 | Diethylnitrosamine | Inflammation | marker/mechanism |
| 5921 | Diethylnitrosamine | Kidney Diseases | marker/mechanism |
| 5921 | Diethylnitrosamine | Kidney Neoplasms | marker/mechanism |
| 5921 | Diethylnitrosamine | Liver Cirrhosis | marker/mechanism |
| 5921 | Diethylnitrosamine | Liver Cirrhosis, Experimental | marker/mechanism |
| 5921 | Diethylnitrosamine | Liver Diseases | marker/mechanism |
| 5921 | Diethylnitrosamine | Liver Neoplasms | marker/mechanism |
| 5921 | Diethylnitrosamine | Liver Neoplasms, Experimental | marker/mechanism |
| 5921 | Diethylnitrosamine | Lung Neoplasms | marker/mechanism |
| 5921 | Diethylnitrosamine | Metaplasia | marker/mechanism |
| 5921 | Diethylnitrosamine | Micronuclei, Chromosome-Defective | marker/mechanism |
| 5921 | Diethylnitrosamine | Necrosis | marker/mechanism |
| 5921 | Diethylnitrosamine | Neoplasm Metastasis | marker/mechanism |
| 5921 | Diethylnitrosamine | Neoplasms, Experimental | marker/mechanism |
| 5921 | Diethylnitrosamine | Nephrosis | marker/mechanism |
| 5921 | Diethylnitrosamine | Neurodegenerative Diseases | marker/mechanism |
| 5921 | Diethylnitrosamine | Neurologic Manifestations | marker/mechanism |
| 5921 | Diethylnitrosamine | Non-alcoholic Fatty Liver Disease | marker/mechanism |
| 5921 | Diethylnitrosamine | Nose Neoplasms | marker/mechanism |
| 5921 | Diethylnitrosamine | Obesity | marker/mechanism |
| 5921 | Diethylnitrosamine | Ovarian Neoplasms | marker/mechanism |
| 5921 | Diethylnitrosamine | Pancreatic Neoplasms | marker/mechanism |
| 5921 | Diethylnitrosamine | Papilloma | marker/mechanism |
| 5921 | Diethylnitrosamine | Precancerous Conditions | marker/mechanism |
| 5921 | Diethylnitrosamine | Renal Insufficiency | marker/mechanism |
| 5921 | Diethylnitrosamine | Stomach Neoplasms | marker/mechanism |
| 5921 | Diethylnitrosamine | Thyroid Neoplasms | marker/mechanism |
| 5921 | Diethylnitrosamine | Tongue Neoplasms | marker/mechanism |
| 5921 | Diethylnitrosamine | Urinary Bladder Neoplasms | marker/mechanism |
| 5921 | Diethylnitrosamine | Weight Loss | marker/mechanism |
| 5283349 | 2,4-decadienal | Adenocarcinoma | marker/mechanism |

| | | | |
|---|---|---|---|
| 5283349 | 2,4-decadienal | Congenital Abnormalities | marker/mechanism |
| 5283349 | 2,4-decadienal | Lung Neoplasms | marker/mechanism |
| 7047 | quinoline | Carcinoma, Hepatocellular | marker/mechanism |
| 7047 | quinoline | Hemangioendothelioma | marker/mechanism |
| 7047 | quinoline | Hemangiosarcoma | marker/mechanism |
| 7047 | quinoline | Liver Neoplasms | marker/mechanism |
| 7047 | quinoline | Neoplasm Metastasis | marker/mechanism |
| 4342 | Acecainide | Drug-Related Side Effects and Adverse Reactions | marker/mechanism |
| 4342 | Acecainide | Tachycardia, Paroxysmal | therapeutic |
| 4342 | Acecainide | Tachycardia, Ventricular | marker/mechanism |
| 4342 | Acecainide | Torsades de Pointes | marker/mechanism |
| 4342 | Acecainide | Ventricular Fibrillation | marker/mechanism |
| 4342 | Acecainide | Ventricular Premature Complexes | therapeutic |
| 10436839 | NH 125 | Mammary Neoplasms, Experimental | therapeutic |
| 65015 | plerixafor | Glioblastoma | therapeutic |
| 65015 | plerixafor | Reflex, Abnormal | therapeutic |
| 442972 | cyclopamine | Abnormalities, Drug-Induced | marker/mechanism |
| 442972 | cyclopamine | Cleft Lip | marker/mechanism |
| 442972 | cyclopamine | Cleft Palate | marker/mechanism |
| 442972 | cyclopamine | Craniofacial Abnormalities | marker/mechanism |
| 442972 | cyclopamine | Edema | marker/mechanism |
| 442972 | cyclopamine | Holoprosencephaly | marker/mechanism |
| 442972 | cyclopamine | Jaw Abnormalities | marker/mechanism |
| 442972 | cyclopamine | Lung Neoplasms | therapeutic |
| 442972 | cyclopamine | Prenatal Injuries | marker/mechanism |
| 3219 | emedastine | Bacterial Infections | therapeutic |
| 3219 | emedastine | Chorea | marker/mechanism |
| 3219 | emedastine | Hypothermia | therapeutic |
| 129695 | 1-naphthylacetylspermine | Anhedonia | therapeutic |
| 129695 | 1-naphthylacetylspermine | Cocaine-Related Disorders | therapeutic |
| 129695 | 1-naphthylacetylspermine | Depressive Disorder | therapeutic |
| 129695 | 1-naphthylacetylspermine | Nerve Degeneration | therapeutic |
| 4456 | Neostigmine | Abdominal Pain | marker/mechanism |
| 4456 | Neostigmine | Airway Obstruction | marker/mechanism |
| 4456 | Neostigmine | Amnesia | therapeutic |
| 4456 | Neostigmine | Apnea | marker/mechanism |
| 4456 | Neostigmine | Apnea | therapeutic |
| 4456 | Neostigmine | Blepharoptosis | therapeutic |
| 4456 | Neostigmine | Blister | therapeutic |
| 4456 | Neostigmine | Bradycardia | marker/mechanism |
| 4456 | Neostigmine | Cardiomegaly | therapeutic |
| 4456 | Neostigmine | Constipation | therapeutic |
| 4456 | Neostigmine | Dermatitis, Irritant | therapeutic |
| 4456 | Neostigmine | Diarrhea | marker/mechanism |

| | | | |
|---|---|---|---|
| 4456 | Neostigmine | Dizziness | marker/mechanism |
| 4456 | Neostigmine | Heart Block | marker/mechanism |
| 4456 | Neostigmine | Hyperalgesia | therapeutic |
| 4456 | Neostigmine | Hyperglycemia | marker/mechanism |
| 4456 | Neostigmine | Hypertension | marker/mechanism |
| 4456 | Neostigmine | Hypotension | marker/mechanism |
| 4456 | Neostigmine | Hypotension | therapeutic |
| 4456 | Neostigmine | Hypotension, Orthostatic | therapeutic |
| 4456 | Neostigmine | Hypothermia | marker/mechanism |
| 4456 | Neostigmine | Labor Pain | therapeutic |
| 4456 | Neostigmine | Long QT Syndrome | marker/mechanism |
| 4456 | Neostigmine | Muscle Weakness | therapeutic |
| 4456 | Neostigmine | Muscular Diseases | therapeutic |
| 4456 | Neostigmine | Myasthenia Gravis | therapeutic |
| 4456 | Neostigmine | Nausea | marker/mechanism |
| 4456 | Neostigmine | Pain | marker/mechanism |
| 4456 | Neostigmine | Pain | therapeutic |
| 4456 | Neostigmine | Pain, Postoperative | therapeutic |
| 4456 | Neostigmine | Paralysis | marker/mechanism |
| 4456 | Neostigmine | Paralysis | therapeutic |
| 4456 | Neostigmine | Postoperative Complications | marker/mechanism |
| 4456 | Neostigmine | Respiratory Paralysis | marker/mechanism |
| 4456 | Neostigmine | Seizures | marker/mechanism |
| 4456 | Neostigmine | Spasm | marker/mechanism |
| 4456 | Neostigmine | Sphincter of Oddi Dysfunction | marker/mechanism |
| 4456 | Neostigmine | Tachycardia | therapeutic |
| 4456 | Neostigmine | Tremor | marker/mechanism |
| 4456 | Neostigmine | Trismus | therapeutic |
| 4456 | Neostigmine | Urinary Retention | therapeutic |
| 446284 | Eicosapentaenoic Acid | Adenomatous Polyposis Coli | therapeutic |
| 446284 | Eicosapentaenoic Acid | Anemia, Hemolytic | marker/mechanism |
| 446284 | Eicosapentaenoic Acid | Arrhythmias, Cardiac | therapeutic |
| 446284 | Eicosapentaenoic Acid | Calcinosis | therapeutic |
| 446284 | Eicosapentaenoic Acid | Cardiovascular Diseases | therapeutic |
| 446284 | Eicosapentaenoic Acid | Coronary Disease | therapeutic |
| 446284 | Eicosapentaenoic Acid | Depressive Disorder | therapeutic |
| 446284 | Eicosapentaenoic Acid | Dyslipidemias | therapeutic |
| 446284 | Eicosapentaenoic Acid | Glucose Intolerance | therapeutic |
| 446284 | Eicosapentaenoic Acid | Hypertension | therapeutic |
| 446284 | Eicosapentaenoic Acid | Hypertriglyceridemia | therapeutic |
| 446284 | Eicosapentaenoic Acid | Insulin Resistance | therapeutic |
| 446284 | Eicosapentaenoic Acid | Intestinal Neoplasms | therapeutic |
| 446284 | Eicosapentaenoic Acid | Micronuclei, Chromosome-Defective | therapeutic |
| 446284 | Eicosapentaenoic Acid | Pain | therapeutic |
| 446284 | Eicosapentaenoic Acid | Parkinson Disease | therapeutic |

| | | | |
|---|---|---|---|
| 446284 | Eicosapentaenoic Acid | Poisoning | marker/mechanism |
| 446284 | Eicosapentaenoic Acid | Raynaud Disease | therapeutic |
| 446284 | Eicosapentaenoic Acid | Schizophrenia | therapeutic |
| 446284 | Eicosapentaenoic Acid | Stomach Neoplasms | therapeutic |
| 446284 | Eicosapentaenoic Acid | Thrombosis | therapeutic |
| 700 | Ethanolamine | Airway Obstruction | marker/mechanism |
| 700 | Ethanolamine | Eye Injuries | marker/mechanism |
| 4642 | oxyphencyclimine | Lipidoses | marker/mechanism |
| 4642 | oxyphencyclimine | Lysosomal Storage Diseases | marker/mechanism |
| 8814 | 4-tert-octylphenol | Embryo Loss | marker/mechanism |
| 8814 | 4-tert-octylphenol | Fatty Liver | marker/mechanism |
| 8814 | 4-tert-octylphenol | Feminization | marker/mechanism |
| 8814 | 4-tert-octylphenol | Infertility, Male | marker/mechanism |
| 8814 | 4-tert-octylphenol | Testicular Diseases | marker/mechanism |
| 8814 | 4-tert-octylphenol | Weight Gain | marker/mechanism |
| 8814 | 4-tert-octylphenol | Weight Loss | marker/mechanism |
| 12433 | dibromoacetic acid | Immune System Diseases | marker/mechanism |
| 24066 | Zalcitabine | Acquired Immunodeficiency Syndrome | therapeutic |
| 24066 | Zalcitabine | AIDS-Related Complex | therapeutic |
| 24066 | Zalcitabine | Anemia | marker/mechanism |
| 24066 | Zalcitabine | Cardiomyopathies | marker/mechanism |
| 24066 | Zalcitabine | Chemical and Drug Induced Liver Injury | marker/mechanism |
| 24066 | Zalcitabine | Demyelinating Diseases | marker/mechanism |
| 24066 | Zalcitabine | Diarrhea | marker/mechanism |
| 24066 | Zalcitabine | Disease Models, Animal | marker/mechanism |
| 24066 | Zalcitabine | Drug-Related Side Effects and Adverse Reactions | marker/mechanism |
| 24066 | Zalcitabine | Fatty Liver | marker/mechanism |
| 24066 | Zalcitabine | Hearing Loss, Bilateral | marker/mechanism |
| 24066 | Zalcitabine | Hearing Loss, Sensorineural | marker/mechanism |
| 24066 | Zalcitabine | HIV Infections | therapeutic |
| 24066 | Zalcitabine | Hyperalgesia | marker/mechanism |
| 24066 | Zalcitabine | Lymphoma | marker/mechanism |
| 24066 | Zalcitabine | Lymphoma, AIDS-Related | therapeutic |
| 24066 | Zalcitabine | Mitochondrial Diseases | marker/mechanism |
| 24066 | Zalcitabine | Mitochondrial encephalopathy | marker/mechanism |
| 24066 | Zalcitabine | Neoplasms, Experimental | marker/mechanism |
| 24066 | Zalcitabine | Nerve Degeneration | marker/mechanism |
| 24066 | Zalcitabine | Neuralgia | marker/mechanism |
| 24066 | Zalcitabine | Neurotoxicity Syndromes | marker/mechanism |
| 24066 | Zalcitabine | Pain | marker/mechanism |
| 24066 | Zalcitabine | Pancytopenia | marker/mechanism |
| 24066 | Zalcitabine | Paresis | marker/mechanism |
| 24066 | Zalcitabine | Paresthesia | marker/mechanism |
| 24066 | Zalcitabine | Peripheral Nervous System Diseases | marker/mechanism |

| | | | |
|---|---|---|---|
| 24066 | Zalcitabine | Thymus Neoplasms | marker/mechanism |
| 24066 | Zalcitabine | Tinnitus | marker/mechanism |
| 24066 | Zalcitabine | Vertigo | marker/mechanism |
| 5283560 | sphingosine 1-phosphate | Neoplasm Invasiveness | marker/mechanism |
| 5283560 | sphingosine 1-phosphate | Scleroderma, Systemic | marker/mechanism |
| 5283560 | sphingosine 1-phosphate | Wounds and Injuries | marker/mechanism |
| 8078 | Cyclohexane | Cardiovascular Diseases | marker/mechanism |
| 8078 | Cyclohexane | Coronary Disease | marker/mechanism |
| 8078 | Cyclohexane | Myocardial Infarction | marker/mechanism |
| 8078 | Cyclohexane | Poisoning | marker/mechanism |
| 8078 | Cyclohexane | Pulmonary Disease, Chronic Obstructive | marker/mechanism |
| 55473 | loxtidine | Stomach Neoplasms | therapeutic |
| 5732 | Zolpidem | Amnesia | marker/mechanism |
| 5732 | Zolpidem | Anxiety Disorders | therapeutic |
| 5732 | Zolpidem | Arrhythmias, Cardiac | marker/mechanism |
| 5732 | Zolpidem | Ataxia | marker/mechanism |
| 5732 | Zolpidem | Brain Injuries | therapeutic |
| 5732 | Zolpidem | Cognition Disorders | marker/mechanism |
| 5732 | Zolpidem | Coma | marker/mechanism |
| 5732 | Zolpidem | Delirium | marker/mechanism |
| 5732 | Zolpidem | Depressive Disorder | therapeutic |
| 5732 | Zolpidem | Disorders of Excessive Somnolence | marker/mechanism |
| 5732 | Zolpidem | Drug Overdose | marker/mechanism |
| 5732 | Zolpidem | Hallucinations | marker/mechanism |
| 5732 | Zolpidem | Hepatic Encephalopathy | marker/mechanism |
| 5732 | Zolpidem | Hepatomegaly | marker/mechanism |
| 5732 | Zolpidem | Hypokalemia | marker/mechanism |
| 5732 | Zolpidem | Hypoxia, Brain | therapeutic |
| 5732 | Zolpidem | Learning Disabilities | marker/mechanism |
| 5732 | Zolpidem | Lethargy | therapeutic |
| 5732 | Zolpidem | Long QT Syndrome | marker/mechanism |
| 5732 | Zolpidem | Movement Disorders | marker/mechanism |
| 5732 | Zolpidem | Movement Disorders | therapeutic |
| 5732 | Zolpidem | Parasomnias | marker/mechanism |
| 5732 | Zolpidem | Persistent Vegetative State | therapeutic |
| 5732 | Zolpidem | Psychomotor Disorders | marker/mechanism |
| 5732 | Zolpidem | Psychoses, Substance-Induced | marker/mechanism |
| 5732 | Zolpidem | REM Sleep Behavior Disorder | marker/mechanism |
| 5732 | Zolpidem | Seizures | marker/mechanism |
| 5732 | Zolpidem | Seizures | therapeutic |
| 5732 | Zolpidem | Sleep Initiation and Maintenance Disorders | therapeutic |
| 5732 | Zolpidem | Sleep Wake Disorders | therapeutic |
| 5732 | Zolpidem | Somnambulism | marker/mechanism |

| | | | |
|---|---|---|---|
| 5732 | Zolpidem | Substance-Related Disorders | marker/mechanism |
| 5732 | Zolpidem | Substance Withdrawal Syndrome | marker/mechanism |
| 5732 | Zolpidem | Torsades de Pointes | marker/mechanism |
| 60149 | sertindole | Basal Ganglia Diseases | marker/mechanism |
| 60149 | sertindole | Catalepsy | marker/mechanism |
| 60149 | sertindole | Hyperkinesis | therapeutic |
| 60149 | sertindole | Long QT Syndrome | marker/mechanism |
| 60149 | sertindole | Schizophrenia | therapeutic |
| 60149 | sertindole | Sleep Initiation and Maintenance Disorders | therapeutic |
| 6378383 | N-desmethyltamoxifen | Vision Disorders | marker/mechanism |
| 8956 | 20-alpha-Dihydroprogesterone | Diabetic Neuropathies | therapeutic |
| 8956 | 20-alpha-Dihydroprogesterone | Peripheral Nervous System Diseases | therapeutic |
| 8956 | 20-alpha-Dihydroprogesterone | Weight Loss | therapeutic |
| 8117 | diethylene glycol | Acidosis | marker/mechanism |
| 8117 | diethylene glycol | Chemical and Drug Induced Liver Injury | marker/mechanism |
| 8117 | diethylene glycol | Kidney Diseases | marker/mechanism |
| 8117 | diethylene glycol | Kidney Tubular Necrosis, Acute | marker/mechanism |
| 8117 | diethylene glycol | Poisoning | marker/mechanism |
| 6557 | isoprene | Breast Neoplasms | marker/mechanism |
| 6557 | isoprene | Kidney Neoplasms | marker/mechanism |
| 6557 | isoprene | Liver Neoplasms, Experimental | marker/mechanism |
| 6557 | isoprene | Lung Neoplasms | marker/mechanism |
| 6557 | isoprene | Mammary Neoplasms, Animal | marker/mechanism |
| 6557 | isoprene | Mammary Neoplasms, Experimental | marker/mechanism |
| 6557 | isoprene | Neoplasm Invasiveness | marker/mechanism |
| 6557 | isoprene | Sleep Apnea, Obstructive | marker/mechanism |
| 6557 | isoprene | Stomach Neoplasms | marker/mechanism |
| 6557 | isoprene | Testicular Neoplasms | marker/mechanism |
| 54739 | imazaquin | Respiratory Sounds | marker/mechanism |
| 12130 | N-nitroso(di-n-propyl)amine | Carcinogenesis | marker/mechanism |
| 12130 | N-nitroso(di-n-propyl)amine | Neoplasms, Experimental | marker/mechanism |
| 688009 | oxilofrine | Hypotension, Orthostatic | therapeutic |
| 9403 | estradiol 17 beta-cypionate | Anemia, Aplastic | marker/mechanism |
| 9403 | estradiol 17 beta-cypionate | Anorexia | marker/mechanism |
| 9403 | estradiol 17 beta-cypionate | Lethargy | marker/mechanism |
| 5283387 | oleylamide | Amnesia | therapeutic |
| 5283387 | oleylamide | Seizures | therapeutic |
| 128423 | alpha-ethyl, alpha-methyl-thiobutyrolactone | Seizures | therapeutic |
| 409805 | NSC 23766 | Acute Lung Injury | therapeutic |
| 409805 | NSC 23766 | Cardiotoxicity | therapeutic |
| 409805 | NSC 23766 | Ventricular Dysfunction, Left | therapeutic |

| | | | |
|---|---|---|---|
| 4908 | Primaquine | Acute Kidney Injury | marker/mechanism |
| 4908 | Primaquine | Anemia | marker/mechanism |
| 4908 | Primaquine | Anemia, Hemolytic | marker/mechanism |
| 4908 | Primaquine | Angioedema | marker/mechanism |
| 4908 | Primaquine | Confusion | marker/mechanism |
| 4908 | Primaquine | Depressive Disorder | marker/mechanism |
| 4908 | Primaquine | Drug Eruptions | marker/mechanism |
| 4908 | Primaquine | Drug-Related Side Effects and Adverse Reactions | marker/mechanism |
| 4908 | Primaquine | Headache | marker/mechanism |
| 4908 | Primaquine | Hematologic Diseases | marker/mechanism |
| 4908 | Primaquine | Hemolysis | marker/mechanism |
| 4908 | Primaquine | Long QT Syndrome | marker/mechanism |
| 4908 | Primaquine | Malaria | therapeutic |
| 4908 | Primaquine | Malaria, Vivax | therapeutic |
| 4908 | Primaquine | Methemoglobinemia | marker/mechanism |
| 4908 | Primaquine | Pruritus | marker/mechanism |
| 4908 | Primaquine | Urination Disorders | marker/mechanism |
| 5281149 | falcarinol | Brain Ischemia | therapeutic |
| 5281149 | falcarinol | Hyperglycemia | therapeutic |
| 5281149 | falcarinol | Reperfusion Injury | therapeutic |
| 5282944 | 9-hydroxy-10,12-octadecadienoic acid | Hyperalgesia | marker/mechanism |
| 13591 | N-Nitrosopyrrolidine | Carcinogenesis | marker/mechanism |
| 13591 | N-Nitrosopyrrolidine | Neoplasms, Experimental | marker/mechanism |
| 5283314 | 4-hydroxy-2-hexenal | Necrosis | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Adenocarcinoma | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Adenoma | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Amyotrophic Lateral Sclerosis | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Aneuploidy | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Asthma | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Autism Spectrum Disorder | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Birth Weight | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Breast Neoplasms | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Cardiovascular Diseases | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Chemical and Drug Induced Liver Injury | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Cognition Disorders | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Colorectal Neoplasms | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Death | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Dementia | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Depressive Disorder | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Developmental Disabilities | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Diabetes Mellitus | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Diabetes Mellitus, Type 2 | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Endometriosis | marker/mechanism |

| | | | |
|---|---|---|---|
| 6636 | Polychlorinated Biphenyls | Epilepsy, Reflex | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Eye Diseases | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Fatty Liver | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Fetal Growth Retardation | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Focal Nodular Hyperplasia | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Glucose Metabolism Disorders | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Hearing Loss | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Hyperplasia | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Hypertension | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Hypertriglyceridemia | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Infertility, Female | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Infertility, Male | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Inflammation | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Intellectual Disability | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Ischemic Stroke | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Learning Disabilities | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Liver Neoplasms | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Liver Neoplasms, Experimental | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Lymphoma, Follicular | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Lymphoma, Large B-Cell, Diffuse | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Lymphoma, Non-Hodgkin | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Memory Disorders | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Metabolic Syndrome | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Motor Disorders | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Motor Skills Disorders | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Nervous System Diseases | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Obesity | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Overweight | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Parkinson Disease | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Poisoning | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Precancerous Conditions | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Precursor Cell Lymphoblastic Leukemia-Lymphoma | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Pregnancy Complications | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Prenatal Exposure Delayed Effects | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Prostatic Neoplasms | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Sex Chromosome Aberrations | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Stroke | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Testicular Neoplasms | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Thyroid Diseases | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Uterine Neoplasms | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Vascular Diseases | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Weight Loss | marker/mechanism |
| 6636 | Polychlorinated Biphenyls | Wilms Tumor | marker/mechanism |
| 284 | Carboxylic Acids | Abnormalities, Drug-Induced | marker/mechanism |

| | | | |
|---|---|---|---|
| 284 | Carboxylic Acids | Hyperalgesia | therapeutic |
| 284 | formic acid | Death | marker/mechanism |
| 11976 | 4-methoxy-3-phenylenediamine | Adenoma | marker/mechanism |
| 11976 | 4-methoxy-3-phenylenediamine | Breast Neoplasms | marker/mechanism |
| 11976 | 4-methoxy-3-phenylenediamine | Carcinoma | marker/mechanism |
| 11976 | 4-methoxy-3-phenylenediamine | Thyroid Neoplasms | marker/mechanism |
| 6540 | tris(2-butoxyethyl) phosphate | Bradycardia | marker/mechanism |
| 6540 | tris(2-butoxyethyl) phosphate | Congenital Abnormalities | marker/mechanism |
| 6540 | tris(2-butoxyethyl) phosphate | Edema | marker/mechanism |
| 6540 | tris(2-butoxyethyl) phosphate | Embryo Loss | marker/mechanism |
| 6540 | tris(2-butoxyethyl) phosphate | Fetal Death | marker/mechanism |
| 6540 | tris(2-butoxyethyl) phosphate | Growth Disorders | marker/mechanism |
| 8133 | n-butoxyethanol | Acidosis | marker/mechanism |
| 8133 | n-butoxyethanol | Coronary Thrombosis | marker/mechanism |
| 8133 | n-butoxyethanol | Hemangiosarcoma | marker/mechanism |
| 8133 | n-butoxyethanol | Hemolysis | marker/mechanism |
| 8133 | n-butoxyethanol | Hypoxia | marker/mechanism |
| 8133 | n-butoxyethanol | Inflammation | marker/mechanism |
| 8133 | n-butoxyethanol | Teratogenesis | marker/mechanism |
| 3202 | Edrophonium | Blepharoptosis | therapeutic |
| 3202 | Edrophonium | Bradycardia | marker/mechanism |
| 3202 | Edrophonium | Chest Pain | marker/mechanism |
| 3202 | Edrophonium | Drug-Related Side Effects and Adverse Reactions | therapeutic |
| 3202 | Edrophonium | Esophageal Motility Disorders | marker/mechanism |
| 3202 | Edrophonium | Esophageal Spasm, Diffuse | marker/mechanism |
| 3202 | Edrophonium | Heart Arrest | marker/mechanism |
| 3202 | Edrophonium | Hypotension | therapeutic |
| 3202 | Edrophonium | Muscle Weakness | therapeutic |
| 3202 | Edrophonium | Myasthenia Gravis | therapeutic |
| 3202 | Edrophonium | Neuromuscular Diseases | therapeutic |
| 3202 | Edrophonium | Ocular Motility Disorders | marker/mechanism |
| 3202 | Edrophonium | Paralysis | therapeutic |
| 3202 | Edrophonium | Tachycardia, Paroxysmal | therapeutic |
| 3202 | Edrophonium | Tachycardia, Supraventricular | therapeutic |
| 3202 | Edrophonium | Tachycardia, Ventricular | marker/mechanism |
| 4158 | Methylphenidate | Abdominal Pain | marker/mechanism |

| | | | |
|---|---|---|---|
| 4158 | Methylphenidate | Akinetic Mutism | marker/mechanism |
| 4158 | Methylphenidate | Amnesia | marker/mechanism |
| 4158 | Methylphenidate | Amphetamine-Related Disorders | therapeutic |
| 4158 | Methylphenidate | Anemia | marker/mechanism |
| 4158 | Methylphenidate | Anorexia | marker/mechanism |
| 4158 | Methylphenidate | Anxiety Disorders | marker/mechanism |
| 4158 | Methylphenidate | Arrhythmias, Cardiac | marker/mechanism |
| 4158 | Methylphenidate | Arteritis | marker/mechanism |
| 4158 | Methylphenidate | Ataxia | marker/mechanism |
| 4158 | Methylphenidate | Athetosis | marker/mechanism |
| 4158 | Methylphenidate | Athetosis | therapeutic |
| 4158 | Methylphenidate | Attention Deficit and Disruptive Behavior Disorders | therapeutic |
| 4158 | Methylphenidate | Attention Deficit Disorder with Hyperactivity | marker/mechanism |
| 4158 | Methylphenidate | Attention Deficit Disorder with Hyperactivity | therapeutic |
| 4158 | Methylphenidate | Autistic Disorder | therapeutic |
| 4158 | Methylphenidate | Basal Ganglia Diseases | marker/mechanism |
| 4158 | Methylphenidate | Bipolar Disorder | marker/mechanism |
| 4158 | Methylphenidate | Bipolar Disorder | therapeutic |
| 4158 | Methylphenidate | Bradycardia | marker/mechanism |
| 4158 | Methylphenidate | Brain Infarction | marker/mechanism |
| 4158 | Methylphenidate | Brain Injuries | therapeutic |
| 4158 | Methylphenidate | Cardiomyopathies | marker/mechanism |
| 4158 | Methylphenidate | Cardiomyopathy, Dilated | marker/mechanism |
| 4158 | Methylphenidate | Cataplexy | therapeutic |
| 4158 | Methylphenidate | Catatonia | marker/mechanism |
| 4158 | Methylphenidate | Cerebral Arterial Diseases | marker/mechanism |
| 4158 | Methylphenidate | Cerebral Infarction | marker/mechanism |
| 4158 | Methylphenidate | Chemical and Drug Induced Liver Injury | marker/mechanism |
| 4158 | Methylphenidate | Child Behavior Disorders | therapeutic |
| 4158 | Methylphenidate | Child Development Disorders, Pervasive | therapeutic |
| 4158 | Methylphenidate | Chorea | marker/mechanism |
| 4158 | Methylphenidate | Chorea | therapeutic |
| 4158 | Methylphenidate | Cocaine-Related Disorders | marker/mechanism |
| 4158 | Methylphenidate | Cocaine-Related Disorders | therapeutic |
| 4158 | Methylphenidate | Cognition Disorders | marker/mechanism |
| 4158 | Methylphenidate | Cognition Disorders | therapeutic |
| 4158 | Methylphenidate | Conduct Disorder | therapeutic |
| 4158 | Methylphenidate | Confusion | marker/mechanism |
| 4158 | Methylphenidate | Consciousness Disorders | therapeutic |
| 4158 | Methylphenidate | Conversion Disorder | marker/mechanism |
| 4158 | Methylphenidate | Coronary Vasospasm | marker/mechanism |
| 4158 | Methylphenidate | Depressive Disorder | marker/mechanism |

| | | | |
|---|---|---|---|
| 4158 | Methylphenidate | Depressive Disorder | therapeutic |
| 4158 | Methylphenidate | Disorders of Excessive Somnolence | marker/mechanism |
| 4158 | Methylphenidate | Disorders of Excessive Somnolence | therapeutic |
| 4158 | Methylphenidate | Disruptive, Impulse Control, and Conduct Disorders | marker/mechanism |
| 4158 | Methylphenidate | Dizziness | marker/mechanism |
| 4158 | Methylphenidate | Drug-Related Side Effects and Adverse Reactions | marker/mechanism |
| 4158 | Methylphenidate | Dyskinesia, Drug-Induced | marker/mechanism |
| 4158 | Methylphenidate | Dystonia | marker/mechanism |
| 4158 | Methylphenidate | Echolalia | marker/mechanism |
| 4158 | Methylphenidate | Enuresis | marker/mechanism |
| 4158 | Methylphenidate | Eosinophilia | marker/mechanism |
| 4158 | Methylphenidate | Epilepsy, Tonic-Clonic | marker/mechanism |
| 4158 | Methylphenidate | Erectile Dysfunction | therapeutic |
| 4158 | Methylphenidate | Fatigue | marker/mechanism |
| 4158 | Methylphenidate | Fatigue | therapeutic |
| 4158 | Methylphenidate | Feeding and Eating Disorders | marker/mechanism |
| 4158 | Methylphenidate | Hallucinations | marker/mechanism |
| 4158 | Methylphenidate | Headache | marker/mechanism |
| 4158 | Methylphenidate | Heart Arrest | marker/mechanism |
| 4158 | Methylphenidate | Hemiplegia | marker/mechanism |
| 4158 | Methylphenidate | Hepatitis, Autoimmune | marker/mechanism |
| 4158 | Methylphenidate | Hepatoblastoma | marker/mechanism |
| 4158 | Methylphenidate | Huntington Disease | marker/mechanism |
| 4158 | Methylphenidate | Hyperkinesis | marker/mechanism |
| 4158 | Methylphenidate | Hyperkinesis | therapeutic |
| 4158 | Methylphenidate | Hypertension | marker/mechanism |
| 4158 | Methylphenidate | Hypertension, Pulmonary | marker/mechanism |
| 4158 | Methylphenidate | Hypotension, Orthostatic | marker/mechanism |
| 4158 | Methylphenidate | Jaundice | marker/mechanism |
| 4158 | Methylphenidate | Leukopenia | marker/mechanism |
| 4158 | Methylphenidate | Livedo Reticularis | marker/mechanism |
| 4158 | Methylphenidate | Liver Neoplasms | marker/mechanism |
| 4158 | Methylphenidate | Memory Disorders | marker/mechanism |
| 4158 | Methylphenidate | Motor Skills Disorders | therapeutic |
| 4158 | Methylphenidate | Movement Disorders | marker/mechanism |
| 4158 | Methylphenidate | Muscle Rigidity | marker/mechanism |
| 4158 | Methylphenidate | Muscular Diseases | therapeutic |
| 4158 | Methylphenidate | Myocardial Infarction | marker/mechanism |
| 4158 | Methylphenidate | Myoclonus | marker/mechanism |
| 4158 | Methylphenidate | Narcolepsy | marker/mechanism |
| 4158 | Methylphenidate | Narcolepsy | therapeutic |
| 4158 | Methylphenidate | Nausea | marker/mechanism |
| 4158 | Methylphenidate | Neurobehavioral Manifestations | therapeutic |

| | | | |
|---|---|---|---|
| 4158 | Methylphenidate | Neuroleptic Malignant Syndrome | marker/mechanism |
| 4158 | Methylphenidate | Obsessive-Compulsive Disorder | marker/mechanism |
| 4158 | Methylphenidate | Ocular Motility Disorders | marker/mechanism |
| 4158 | Methylphenidate | Pain | therapeutic |
| 4158 | Methylphenidate | Pericarditis | marker/mechanism |
| 4158 | Methylphenidate | Peripheral Vascular Diseases | marker/mechanism |
| 4158 | Methylphenidate | Pica | marker/mechanism |
| 4158 | Methylphenidate | Prenatal Exposure Delayed Effects | marker/mechanism |
| 4158 | Methylphenidate | Psychomotor Agitation | marker/mechanism |
| 4158 | Methylphenidate | Psychomotor Agitation | therapeutic |
| 4158 | Methylphenidate | Psychomotor Disorders | therapeutic |
| 4158 | Methylphenidate | Psychoses, Substance-Induced | marker/mechanism |
| 4158 | Methylphenidate | Psychotic Disorders | therapeutic |
| 4158 | Methylphenidate | Raynaud Disease | marker/mechanism |
| 4158 | Methylphenidate | Reflex, Abnormal | marker/mechanism |
| 4158 | Methylphenidate | Seizures | marker/mechanism |
| 4158 | Methylphenidate | Serotonin Syndrome | marker/mechanism |
| 4158 | Methylphenidate | Sexual Dysfunctions, Psychological | marker/mechanism |
| 4158 | Methylphenidate | Sleep Bruxism | marker/mechanism |
| 4158 | Methylphenidate | Sleep Initiation and Maintenance Disorders | marker/mechanism |
| 4158 | Methylphenidate | Sleep Wake Disorders | marker/mechanism |
| 4158 | Methylphenidate | Somatoform Disorders | marker/mechanism |
| 4158 | Methylphenidate | Somatoform Disorders | therapeutic |
| 4158 | Methylphenidate | Somnambulism | marker/mechanism |
| 4158 | Methylphenidate | Speech Disorders | marker/mechanism |
| 4158 | Methylphenidate | Stereotypic Movement Disorder | marker/mechanism |
| 4158 | Methylphenidate | Stuttering | marker/mechanism |
| 4158 | Methylphenidate | Substance Abuse, Intravenous | marker/mechanism |
| 4158 | Methylphenidate | Substance-Related Disorders | marker/mechanism |
| 4158 | Methylphenidate | Substance Withdrawal Syndrome | marker/mechanism |
| 4158 | Methylphenidate | Substance Withdrawal Syndrome | therapeutic |
| 4158 | Methylphenidate | Tachycardia | marker/mechanism |
| 4158 | Methylphenidate | Tachycardia, Ventricular | marker/mechanism |
| 4158 | Methylphenidate | Thrombocytosis | marker/mechanism |
| 4158 | Methylphenidate | Tic Disorders | marker/mechanism |
| 4158 | Methylphenidate | Tic Disorders | therapeutic |
| 4158 | Methylphenidate | Tics | marker/mechanism |
| 4158 | Methylphenidate | Torticollis | marker/mechanism |
| 4158 | Methylphenidate | Tourette Syndrome | marker/mechanism |
| 4158 | Methylphenidate | Tremor | marker/mechanism |
| 4158 | Methylphenidate | Trichotillomania | marker/mechanism |
| 4158 | Methylphenidate | Vasculitis, Central Nervous System | marker/mechanism |
| 4158 | Methylphenidate | Ventricular Dysfunction, Left | marker/mechanism |
| 4158 | Methylphenidate | Ventricular Fibrillation | marker/mechanism |

| | | | |
|---|---|---|---|
| 4158 | Methylphenidate | Vertigo | marker/mechanism |
| 4158 | Methylphenidate | Vomiting | marker/mechanism |
| 4158 | Methylphenidate | Xerostomia | marker/mechanism |
| 445639 | Oleic Acid | Acute Lung Injury | marker/mechanism |
| 445639 | Oleic Acid | Breast Neoplasms | therapeutic |
| 445639 | Oleic Acid | Cognition Disorders | marker/mechanism |
| 445639 | Oleic Acid | Fatty Liver | marker/mechanism |
| 445639 | Oleic Acid | Hepatitis, Animal | marker/mechanism |
| 445639 | Oleic Acid | Hypoxia | marker/mechanism |
| 445639 | Oleic Acid | Insulin Resistance | marker/mechanism |
| 445639 | Oleic Acid | Lung Injury | marker/mechanism |
| 445639 | Oleic Acid | Non-alcoholic Fatty Liver Disease | marker/mechanism |
| 445639 | Oleic Acid | Pulmonary Edema | marker/mechanism |
| 445639 | Oleic Acid | Respiratory Distress Syndrome | marker/mechanism |
| 9903 | Lithocholic Acid | Birth Weight | marker/mechanism |
| 9903 | Lithocholic Acid | Chemical and Drug Induced Liver Injury | marker/mechanism |
| 9903 | Lithocholic Acid | Cholangitis | marker/mechanism |
| 9903 | Lithocholic Acid | Cholestasis | marker/mechanism |
| 9903 | Lithocholic Acid | Cholestasis, Intrahepatic | marker/mechanism |
| 9903 | Lithocholic Acid | Colonic Neoplasms | marker/mechanism |
| 9903 | Lithocholic Acid | Diabetes Mellitus, Experimental | therapeutic |
| 9903 | Lithocholic Acid | Diabetes Mellitus, Type 2 | therapeutic |
| 9903 | Lithocholic Acid | Fatty Liver | marker/mechanism |
| 9903 | Lithocholic Acid | Fetal Weight | marker/mechanism |
| 9903 | Lithocholic Acid | Glucose Intolerance | therapeutic |
| 9903 | Lithocholic Acid | Hyperplasia | marker/mechanism |
| 9903 | Lithocholic Acid | Liver Cirrhosis | marker/mechanism |
| 9903 | Lithocholic Acid | Necrosis | marker/mechanism |
| 9903 | Lithocholic Acid | Neoplasm Metastasis | therapeutic |
| 9903 | Lithocholic Acid | Precancerous Conditions | marker/mechanism |
| 9903 | Lithocholic Acid | Prenatal Exposure Delayed Effects | marker/mechanism |
| 3341 | Fenoldopam | Acute Kidney Injury | therapeutic |
| 3341 | Fenoldopam | Arteritis | marker/mechanism |
| 3341 | Fenoldopam | Bradycardia | marker/mechanism |
| 3341 | Fenoldopam | Cardiac Output, High | marker/mechanism |
| 3341 | Fenoldopam | Hemorrhage | marker/mechanism |
| 3341 | Fenoldopam | Hypertension | therapeutic |
| 3341 | Fenoldopam | Hypotension | marker/mechanism |
| 3341 | Fenoldopam | Necrosis | marker/mechanism |
| 3341 | Fenoldopam | Peripheral Arterial Disease | marker/mechanism |
| 3341 | Fenoldopam | Polyarteritis Nodosa | marker/mechanism |
| 5283324 | 2-octenal | Neurotoxicity Syndromes | marker/mechanism |
| 3114 | Disopyramide | Acute Kidney Injury | marker/mechanism |
| 3114 | Disopyramide | Arrhythmias, Cardiac | marker/mechanism |
| 3114 | Disopyramide | Arrhythmias, Cardiac | therapeutic |

| | | | |
|---|---|---|---|
| 3114 | Disopyramide | Arthralgia | marker/mechanism |
| 3114 | Disopyramide | Atrial Fibrillation | therapeutic |
| 3114 | Disopyramide | Atrial Flutter | therapeutic |
| 3114 | Disopyramide | Atrial Premature Complexes | therapeutic |
| 3114 | Disopyramide | Atrioventricular Block | marker/mechanism |
| 3114 | Disopyramide | Bradycardia | marker/mechanism |
| 3114 | Disopyramide | Bundle-Branch Block | marker/mechanism |
| 3114 | Disopyramide | Cardiac Complexes, Premature | therapeutic |
| 3114 | Disopyramide | Cardiomegaly | marker/mechanism |
| 3114 | Disopyramide | Cardiomyopathy, Hypertrophic | therapeutic |
| 3114 | Disopyramide | Chemical and Drug Induced Liver Injury | marker/mechanism |
| 3114 | Disopyramide | Cholestasis | marker/mechanism |
| 3114 | Disopyramide | Cholestasis, Intrahepatic | marker/mechanism |
| 3114 | Disopyramide | Coma | marker/mechanism |
| 3114 | Disopyramide | Death, Sudden | marker/mechanism |
| 3114 | Disopyramide | Disseminated Intravascular Coagulation | marker/mechanism |
| 3114 | Disopyramide | Dizziness | marker/mechanism |
| 3114 | Disopyramide | Drug-Related Side Effects and Adverse Reactions | marker/mechanism |
| 3114 | Disopyramide | Erectile Dysfunction | marker/mechanism |
| 3114 | Disopyramide | Erythema Nodosum | marker/mechanism |
| 3114 | Disopyramide | Eye Diseases | marker/mechanism |
| 3114 | Disopyramide | Fatigue | marker/mechanism |
| 3114 | Disopyramide | Fetal Diseases | therapeutic |
| 3114 | Disopyramide | Heart Arrest | marker/mechanism |
| 3114 | Disopyramide | Heart Block | marker/mechanism |
| 3114 | Disopyramide | Heart Block | therapeutic |
| 3114 | Disopyramide | Heart Diseases | marker/mechanism |
| 3114 | Disopyramide | Heart Failure | marker/mechanism |
| 3114 | Disopyramide | Hypertension | marker/mechanism |
| 3114 | Disopyramide | Hypoglycemia | marker/mechanism |
| 3114 | Disopyramide | Hypotension | marker/mechanism |
| 3114 | Disopyramide | Jaundice, Obstructive | marker/mechanism |
| 3114 | Disopyramide | Liver Diseases | marker/mechanism |
| 3114 | Disopyramide | Long QT Syndrome | marker/mechanism |
| 3114 | Disopyramide | Muscle Cramp | marker/mechanism |
| 3114 | Disopyramide | Myocardial Infarction | therapeutic |
| 3114 | Disopyramide | Neuromuscular Diseases | marker/mechanism |
| 3114 | Disopyramide | Pain | marker/mechanism |
| 3114 | Disopyramide | Paresthesia | marker/mechanism |
| 3114 | Disopyramide | Peripheral Nervous System Diseases | marker/mechanism |
| 3114 | Disopyramide | Pulmonary Edema | marker/mechanism |
| 3114 | Disopyramide | Pyelonephritis | marker/mechanism |
| 3114 | Disopyramide | Renal Insufficiency | marker/mechanism |
| 3114 | Disopyramide | Respiration Disorders | marker/mechanism |

| | | | |
|---|---|---|---|
| 3114 | Disopyramide | Seizures | marker/mechanism |
| 3114 | Disopyramide | Shock, Cardiogenic | marker/mechanism |
| 3114 | Disopyramide | Sinoatrial Block | marker/mechanism |
| 3114 | Disopyramide | Sinus Arrest, Cardiac | marker/mechanism |
| 3114 | Disopyramide | Syncope | marker/mechanism |
| 3114 | Disopyramide | Tachycardia | therapeutic |
| 3114 | Disopyramide | Tachycardia, Atrioventricular Nodal Reentry | therapeutic |
| 3114 | Disopyramide | Tachycardia, Paroxysmal | therapeutic |
| 3114 | Disopyramide | Tachycardia, Sinus | marker/mechanism |
| 3114 | Disopyramide | Tachycardia, Supraventricular | therapeutic |
| 3114 | Disopyramide | Tachycardia, Ventricular | marker/mechanism |
| 3114 | Disopyramide | Tachycardia, Ventricular | therapeutic |
| 3114 | Disopyramide | Takotsubo Cardiomyopathy | marker/mechanism |
| 3114 | Disopyramide | Torsades de Pointes | marker/mechanism |
| 3114 | Disopyramide | Urinary Retention | marker/mechanism |
| 3114 | Disopyramide | Ventricular Dysfunction | marker/mechanism |
| 3114 | Disopyramide | Ventricular Dysfunction, Left | marker/mechanism |
| 3114 | Disopyramide | Ventricular Dysfunction, Left | therapeutic |
| 3114 | Disopyramide | Ventricular Fibrillation | marker/mechanism |
| 3114 | Disopyramide | Ventricular Fibrillation | therapeutic |
| 3114 | Disopyramide | Ventricular Flutter | marker/mechanism |
| 3114 | Disopyramide | Ventricular Premature Complexes | marker/mechanism |
| 3114 | Disopyramide | Ventricular Premature Complexes | therapeutic |
| 3114 | Disopyramide | Vision Disorders | marker/mechanism |
| 3114 | Disopyramide | Xerostomia | marker/mechanism |
| 6437074 | SQ 29548 | Bradycardia | therapeutic |
| 6437074 | SQ 29548 | Hypertension | therapeutic |