Change in Pattern Recognition of Electronic Noses in Detecting Common Fungal Contaminants

Research Paper

Submitted to Kennesaw Mountain High School

by

DANIEL CHU

Kennesaw, GA

November 2017

Abstract

Indoor fungal contamination by mold species was a leading cause of SBS, which consisted of the symptoms occupants experience during time spent in a building. Electronic noses have had applications in many fields, including the detection of VOCs released by mold spores. However, the accuracy of electronic noses was lower than alternative methods, especially due to the varying employment of pattern recognition algorithms. The purpose of my quantitative, experimental study was to determine the pattern recognition algorithm with the best classification and detection of fungal contamination. I measured the proportion of correctly detected and classified samples by the Cyranose 320 electronic nose, using the treatments of the KNN, k-means, SVM, and CDA pattern recognition algorithms on grown samples of *Alternaria*, *Aspergillus*, *Cladosporium*, and *Penicillium*. A Cochran's Q test found statistical significance ($p = 0.0049$) between the classifications of the four algorithms. Post hoc McNemar's Chi-Square tests found statistical significance between the classifications of the k-means algorithm with the KNN ($p = 0.0046$), SVM ($p = 0.0433$), and CDA ($p = 0.0209$) algorithms and between the detection of the SVM and CDA algorithms ($p = 0.0014$). These results illustrated the k-means algorithm's low classification due to a poor first partition and SVM algorithm's low detection due to lower algorithm confidence. The findings of this study contributed to improvements of electronic nose accuracy in fungal detection. I recommended future research on additional pattern recognition algorithms and comprehensive simulation of fungal contamination to expand the scope and validity of this study.

Acknowledgements

I would like to thank my mentor, Dr. Hussein Abaza, for providing me this experience at Kennesaw State University under his guidance. I would also like to thank the lab directors, Mike Frailey and Lashaundra Fambro, for assisting with my experimentation and data collection.

Additionally, I express my gratitude towards Ms. Chelsea Sexton, Dr. Eden Hunt, and Dr. Kelly Ingle for maintaining and directing me through the research and internship program. I thank my peers for their feedback on this work and their continuous support.

Table of Contents

Chapter 1: Introduction

The Environmental Protection Agency (EPA, 1991) found that up to 30% of new and remodeled buildings have poor indoor air quality, which leaves occupants at an increased risk for the symptoms of Sick Building Syndrome (SBS). Specifically, SBS is the adverse and recurring symptoms, including headaches, nausea, irritation, fatigue, and coughing that occupants experience during their time spent in a building, that consequently hinder efficiency and increase absenteeism (Joshi, 2008; Khan & Karuppayil, 2012). A common cause of SBS is the presence of bioaerosols, most notably mold, a form of fungal contamination (Kuske et al., 2006; Schiffman, Wyrick, Gutierrez-Osuna, & Nagle, 2000). However, current contamination detection methods, including mycology, polymerase chain reaction, and chromatography-mass spectrometry, are often time-consuming, expensive, and off-site; as a result, researchers have turned towards detection using electronic noses, as they deliver quick, cheap, and conclusive information (Kuske et al., 2006; Lagod, Suchorab, Guz, & Sobczuk, 2017; Paolesse et al., 2006).

Statement of the Problem

Electronic noses provide an optimistic approach in the detection of molds over traditional methods (Schiffman et al., 2000). The accuracy of electronic noses is debatable, facing issues with low chemical species concentrations, false readings, inability to detect individual chemical compounds, and lack of common usage and methodology (Eusebio, Capelli, & Sironi, 2016; Gobbi & Locci, 2006; Lagod et al., 2017). Most overlook electronic noses as a detection method in favor of the more traditional and expensive gas chromatography-mass spectrometry (GC-MS), mycological analysis, and polymerase chain reaction (PCR) (Gobbi & Locci, 2006; Lagod et al., 2017).

Electronic noses are instruments that consist of incrementally different semiconductor gas sensors that form a sensor array, which responds to a wide variety of chemical species that pass over the array (Wilson & Baietto, 2009). Then, the sensory array sends data through pattern recognition algorithms to classify the exposure using the training data from previously sampled exposures (Mumyakmaz & Karabacak, 2015). The sensors across many electronic noses function similarly and produce precise readings; however, pattern recognition algorithms used by electronic noses vary among researchers (Eusebio et al., 2016).

The most simple and common pattern recognition algorithm used in electronic nose studies is the k-nearest neighbors (KNN) pattern recognition algorithm used with principal component analysis (PCA), which compresses and performs feature selection on sensor data (Schiffman et al., 2000; Tang, Lin, & Shyu, 2010). The KNN pattern recognition algorithm specialized in PCAs in which classes of chemical species have low variance (Huang et al., 2016; Intelligent Optical Systems, 2000; Wilson & Baietto, 2009). Most PCAs for fungal detection studies, including the studies of Lagod et al. (2017), Paolesse et al. (2006), and Schiffman et al. (2000), showed clustered classes of chemical species with high variance between classes. Additionally, the k-means, support vector machine (SVM), and canonical discriminant analysis (CDA) pattern recognition algorithms processed differently than the KNN pattern recognition algorithm and showed different pattern recognition capabilities for fungal contamination applications of electronic noses (Intelligent Optical Systems, 2000; Scott, James, & Ali, 2007). Incorrect usage of pattern recognition algorithms has been problem that affects electronic nose studies, and research should address this issue (Eusebio et al., 2016).

Purpose of the Study

A universal pattern recognition algorithm for the detection of indoor fungal contaminants remains undefined (Eusebio et al., 2016). Cagnasso et al. (2010), Lagod et al. (2017), Paolesse et al. (2006), and Schiffman et al. (2000) varied in this pattern recognition aspect of electronic nose detection. The purpose of this quantitative, experimental study is to establish a recommended pattern recognition algorithm for the detection and classification of common indoor fungal contaminants using an electronic nose. The researchers of this quantitative, experimental study can determine and conclude the most accurate pattern recognition algorithm used for electronic nose applications in mold detection based on measurements of classification and detection calculated as the proportion of correctly identified and classified exposures for each pattern recognition algorithm. The researchers will closely follow the procedures of previous case studies into electronic noses, including the studies of Kuske et al. (2006), Lagod et al. (2017), and Schiffman et al. (2000), but will control possible confounding variables, including building materials (Kuske et al., 2006) and contaminant species (Lagod et al., 2017), while also following an experimental design more suitable for the independent variable of pattern recognition and dependent variable of classification and detection accuracy. Each pattern recognition algorithm has a higher classification accuracy for certain types of data, including clustered classes with high variance between classes and binary classifications with only two classes; therefore, finding a suitable pattern recognition algorithm for common indoor molds may improve the accuracy of the electronic nose in fungal detection, allowing for a cheaper, quicker, and more portable method in comparison to GC-MS, PCR, and mycological analysis in improving indoor air quality and preventing SBS (Scott et al., 2007).

Research Questions

Q1. How does the classification of common indoor fungal contaminants vary for different electronic nose pattern recognition algorithms?

Q1.1. How does the classification of *Alternaria* vary for different electronic nose pattern recognition algorithms?

Q1.2. How does the classification of *Aspergillus* vary for different electronic nose pattern recognition algorithms?

Q1.3. How does the classification of *Cladosporium* vary for different electronic nose pattern recognition algorithms?

Q1.4. How does the classification of *Penicillium* vary for different electronic nose pattern recognition algorithms?

Q2. How does the detection of common indoor fungal contaminants vary for different electronic nose pattern recognition algorithms?

Q2.1. How does the detection of *Alternaria* vary for different electronic nose pattern recognition algorithms?

Q2.2. How does the detection of *Aspergillus* vary for different electronic nose pattern recognition algorithms?

Q2.3. How does the detection of *Cladosporium* vary for different electronic nose pattern recognition algorithms?

Q2.4. How does the detection of *Penicillium* vary for different electronic nose pattern recognition algorithms?

Hypotheses

RH. Different pattern recognition algorithms of an electronic nose will have different effects on an electronic nose's ability to detect and classify common fungal contaminants.

$H1_0$. There is no difference in the classification of common indoor fungal contaminants for different electronic nose pattern recognition algorithms.

$H1_a$. There is a difference in the classification of common indoor fungal contaminants for different electronic nose pattern recognition algorithms.

$H2_0$. There is no difference in the detection of common indoor fungal contaminants for different electronic nose pattern recognition algorithms.

$H2_a$. There is a difference in the detection of common indoor fungal contaminants for different electronic nose pattern recognition algorithms.

Significance of the Study

The significance of this quantitative, experimental study is to define a specific pattern recognition algorithm for the use of electronic noses under a controlled experimental design. Gobbi and Locci (2006), Lagod et al. (2017), Paolesse et al. (2006), and Schiffman et al. (2000) performed general exploratory analyses on the application of electronic noses in fungal contamination. Only Cagnasso et al. (2010), Hong et al. (2015), Schiffman et al. (2000), Tang et al. (2010), and Zhang et al. (2017) compared the different pattern recognition algorithms. Furthermore, some researchers did not focus directly on the factors of pattern recognition algorithms, fungal contaminants, and controlled experimental setup, thus reducing the validity of these studies (Cagnasso et al. 2010; Gobbi & Locci, 2006; Lagod et al., 2017; Paolesse et al., 2006; Schiffman et al., 2000). In this study, I aim to produce a controlled experiment between

multiple pattern recognition algorithms on a representative population of the exposures of indoor fungal contaminants to benefit the field of electronic noses.

Statistically significant research can further the application of electronic noses as mold detectors (Kuske et al., 2006). Rather than waiting for weeks to setup and obtain data from PCR, GC-MS, or mycological analysis (by which any potential fungal contamination will have grown considerably), researchers, inspectors, and property owners can perform repeatable, mobile tests in less than a minute (Gobbi & Locci, 2006; Lagod et al., 2017; Schiffman et al., 2000). Electronic noses will notify the presence of fungal contamination, allowing for earlier action against the mold, thus preventing many cases of SBS from occurring and impeding health, productivity, and attendance (Joshi, 2008).

Definition of Key Terms

Electronic nose. Electronic noses are instruments that use sensors and pattern recognition to determine the composition of airborne chemical species (Gardner & Bartlett, 1994).

Sick building syndrome (SBS). SBS is the set of symptoms experienced due to poor indoor air quality during one's time spent in a building (Joshi, 2008).

Volatile organic compounds (VOCs). VOCs are carbon-based compounds that exist as a vapor (EPA, 1991).

Class. Classes are a specific chemical species or chemical compound identified by the electronic nose (Intelligent Optical Systems, 2000).

Detection. Detection is the process of determining the presence of a mold (Schiffman et al., 2000).

Classification. Classification is the process of determining the species of a mold if mold is present (Schiffman et al., 2000).

Accuracy. Accuracy is the proportion of exposures in which the electronic nose correctly identifies the class of the unknown exposure; accuracy is equivalent to the quotient of correctly identified exposures over the total number of exposures (Schiffman et al., 2000).

Cross-validity. Cross-validity is the proportion of exposures of the training set in which the electronic nose correctly identifies the class of each training exposure using pattern recognition to determine how useful the training set is (Intelligent Optical Systems, 2000).

Exposure. An exposure is an instance of the electronic nose sampling the chemical species and recording the resultant data as a thirty-two dimension vector containing components of the relative change in resistance ($\Delta R/R_0$) of each of the thirty-two sensors of the electronic nose (Intelligent Optical Systems, 2000).

Principal component analysis (PCA). PCA is a data compression algorithm and resultant graphical representation that uses linear projections of correlated exposure vector variables to increase the variation between classes (Intelligent Optical Systems, 2000).

Pattern recognition. Pattern recognition is an algorithm of the electronic nose, including KNN, k-means, SVM, or CDA, that determines the class of identified samples by looking at sensor data from the sample and comparing it to sensor data from trained classes (Rahman, Charoenlarpnopparut, & Suksompon, 2016).

K-nearest neighbors (KNN). KNN is a pattern recognition algorithm that determines the class of an exposure by gathering the classes of the k nearest training exposures (Scott et al., 2007).

Support vector machine (SVM). SVM is a pattern recognition algorithm that determines the class of an exposure by optimizing hyperplanes between the classes of exposure vectors (Pardo & Sberveglieri, 2005).

K-means. K-means is a pattern recognition algorithm that determines the class of an exposure by partitioning the PCA using recursive class mean optimization (Scott et al., 2007).

Canonical discriminant analysis (CDA). CDA is a pattern recognition and data compression algorithm that determines the class of an exposure through linear combinations of the training exposure vector variables with the most correlation (Scott et al., 2007).

Summary

Poor indoor air quality in up to 30% of new and renovated buildings can cause the occupants to experience SBS (EPA, 1991). Electronic noses are a proposed method of detection for common fungal contaminants, a frequent cause of SBS (Schiffman et al., 2000). However, the accuracy of electronic noses falls behind other methods, as unregulated employment of pattern recognition hampers the nose's ability to detect and classify common fungal contaminants (Eusebio et al., 2016; Lagod et al., 2017). In this study, I seek to employ the KNN, k-means, SVM, and CDA pattern recognition algorithms under a controlled quantitative, experiment on a representative population of the exposures of common fungal contaminants. Conclusive data may help to define pattern recognition techniques to use, and, through improving the electronic nose, provide a quick, cheap, reusable, and mobile means of detecting and classifying fungal contamination. An effective pattern recognition algorithm will benefit the accuracy of future studies in the field. The next section discusses findings regarding electronic nose detection of SBS-causing fungal contaminants and shows promising potential for different pattern recognition algorithms.

Chapter 2: Literature Review

Sick building syndrome (SBS) is a widespread disease that causes occupants to experience symptoms during their time spent in a building (EPA, 1991; Joshi, 2008). A main cause of SBS is common indoor fungal contaminants, including the genera of *Aspergillus, Cladosporium*, *Penicillium,* and *Alternaria* (Khan & Karappuayil, 2012). Electronic noses are means of detecting fungal contaminants through their produced volatile organic compounds (VOCs) (Schiffman, Wyrick, Gutierrez-Osuna, & Nagle, 2000). Although electronic noses are faster and more efficient than traditional methods, they are less accurate (Lagod et al., 2017).

Electronic noses are instruments designed to detect airborne chemical species, one of which is the Cyranose 320 (Intelligent Optical Systems, 2000). The Cyranose 320 contains a sensory array to quantify data from chemical species (Wilson & Baietto, 2009). The Cyranose 320 sends the data to the principal component analysis (PCA) before arriving at a pattern recognition algorithm, including support vector machine (SVM), linear discriminant analysis (LDA), k-nearest neighbors (KNN), or k-means, to cross-validate the training set, which allows for more accurate use of the electronic nose; each algorithm has advantages and disadvantages in their uses, but their application to fungal detection remains ambiguous (Intelligent Optical Systems, 2000; Pardo & Sberveglieri, 2005; Scott, James, & Ali, 2007).

Sick Building Syndrome (SBS)

Sick building syndrome (SBS) is the health or comfort related symptoms experienced by occupants during their time spent in a building (Joshi, 2008). Occupants can experience SBS in a singular room, regionalized area, or an entire building (Joshi, 2008; Khan & Karappuayil, 2012). Sufferers of SBS acquire acute symptoms, including headaches, tiredness, induced asthma, inflammation, irritation, nausea, dizziness, allergies, and dryness, that taper off as an

occupant leaves the building (Joshi, 2008; Mui, Chan, Wong, & Hui, 2010; Schiffman et al., 2000). Additionally, SBS can have chronic consequences on occupant health, including an increased risk of acquiring cancers or allergies, infections from pathogens, and decreased cognition (Joshi, 2008). The EPA (1991) reported that 30% of new and remodeled buildings have poor indoor air quality that leads to SBS. These symptoms result in decreased worker efficiency and increased absenteeism that hamper productivity (Joshi, 2008).

Many indoor and outdoor contaminants induce SBS, and the symptoms of SBS are source dependent. Joshi (2008), Khan and Karappuayil (2012), and Mui et al. (2010) cite biological sources, including pollen, bacteria, viruses, and most notably, molds and their VOCs produced, as additional causes of SBS; the VOCs produced from these sources cause more of the acute, short-term effects, whereas the sources themselves can linger and cause sustained respiratory infections and weakened immune systems (Kuske et al., 2006; Lagod et al., 2017; Schiffman et al., 2000). Schiffman et al. (2000) further point out the strong correlation between fungal contamination and SBS in a study of the electronic nose detection of fungal and bacterial species responsible for contamination.

Indoor Fungal Contamination

Indoor fungal contamination includes all forms of mold growing on indoor surfaces; a wide variety of building abnormalities can lead to indoor fungal contamination, which can cause adverse health effects (Joshi, 2008; Khan & Karappuayil, 2012). Humidity is one of the dominant factors in fungal contamination; moisture from roof leaks, humidifiers, or storm damage facilitates mold growth on most surfaces (Khan & Karappuayil, 2012). Similarly, indoor objects, ventilation system quality, human activities, and outdoor molds also play a significant role in enabling fungal contamination (Mui et al., 2010).

There are several main genera involved in fungal contamination: *Aspergillus, Cladosporium, Penicillium,* and *Alternaria* (Khan & Karappuayil, 2012; Mui et al., 2010; Shelton, Kirkland, Flanders, & Morris, 2002). Kuske et al. (2006), Mui et al. (2010), Shelton et al. (2002), and the World Health Organization (WHO, 2009) analyzed properties of these species of mold. The WHO (2009) focused towards the symptoms of each mold; whereas, Kuske et al. (2006) studied the contamination of building materials by these species. Mui et al. (2010) and Shelton et al. (2002) sampled the presence of each species in buildings; while Mui et al. (2010) focused on the rooms within a single building, Shelton et al. (2002) covered tens of thousands of inspected buildings in the United States.

Each of the common genera produces a varied set of symptoms. *Aspergillus* infections are the most aggressive of the molds, causing many complications in immunocompromised individuals; in addition, *Alternaria* can cause chronic sinusitis in response to *Aspergillus* infection or presence (WHO, 2009). The most common genera associated with allergies are *Aspergillus, Cladosporium,* and *Penicillium*, but these allergies only develop in a small proportion of the population exposed to molds (WHO, 2009). *Aspergillus* and *Penicillium* produce neurotoxins, but, as of 2009, they have not shown correlation with nervous system symptoms (WHO, 2009). *Aspergillus* and *Penicillium* also produce mycotoxins; however, *Penicillium* produces relatively harmless mycotoxins including penicillin, an extremely powerful antibiotic, while *Aspergillus* produces several carcinogenic mycotoxins (WHO, 2009).

Aspergillus species tend to colonize inorganic materials and modified wood products, *Cladosporium* species tend to colonize wooden building materials, and *Penicillium* species tend to colonize wooden building materials and modified wood products (Khan & Karappuayil, 2012). Kuske et al. (2006) were successfully able to discriminate the mold contamination of

Aspergillus versicolor, Penicillium chrysogeum, Penicillium aurantiogriseum, and *Cladosporium sphaerospermum* species on plasterboard, particleboard, oriented strain board, and wallpaper.

Mui et al. (2010) sampled the presence of mold in a Hong Kong office building using mycological analysis. Mui et al. (2010) found *Aspergillus* in 50.0% of rooms, *Cladosporium* in 48.8% of rooms, and *Penicillium* in 56.1% of rooms. However, the procedure involved sampling five hundred liters of air for only colony forming units, which can result from a single spore, so the actual rates of fungal contamination are lower; additionally, the building boasted an excellent air quality classification during previous inspections (Mui et al., 2010).

Shelton et al. (2002) compiled 12,026 fungal air samples from official air quality inspections of buildings across the United States. Shelton et al. (2002) found that buildings in the Southeast United States experience higher mold concentrations during the fall, which is comparable with the national trend. Indoor presence of fungal genera for the Southeast in the fall include *Penicillium* (89%), *Cladosporium* (89%), and *Aspergillus* (73%), and indoor concentrations of fungal genera for the Southeast in the fall include *Aspergillus* (400 Colony Forming Units (CFU)/m^3), *Cladosporium* (330 CFU/m^3), and *Penicillium* (360 CFU/m^3) (Shelton et al., 2002). However, similar to Mui et al. (2010), Shelton et al. (2002) likely also showed skewed data as most of the buildings prompted inspections due to visible mold or symptoms associated with SBS. Therefore, the actual presence of fungal genera is likely lower than the reported quantities. Nevertheless, Kuske et al. (2006), Mui et al. (2010), Shelton et al. (2002), and the World Health Organization (WHO, 2009) show agreement in determining *Alternaria*, *Aspergillus*, *Cladosporium*, and *Penicillium* as the main genera involved in indoor fungal contamination and resultant cases of SBS.

Volatile Organic Compounds (VOCs)

Volatile organic compounds (VOCs) are carbon-based compounds that exist as a vapor (EPA, 1991). Various VOCs can contribute to indoor pollution, leading to poor indoor air quality and SBS (Khan & Karappuayil, 2012; Lagod et al., 2017). Smoking, perfumes, temperature, humidity, ventilation, air conditioning systems, cleaning products, and building renovation are indoor abiotic sources of VOCs (Khan & Karappuayil, 2012; Mumyakmaz & Karabacak, 2015). Microbial VOCs (MVOCs) are a subclass of VOCs that are released during mold development and metabolism; The MVOCs and mycotoxins that are released from molds can escape through surfaces that are not accessible to spores and cause symptoms that are independent of the spores themselves, including headaches, irritation, and nausea (Khan & Karappuayil, 2012; Lagod et al., 2017; Schiffman et al., 2000).

Specific VOCs indicate microbial growth, and their presence correlates with fungal contamination (Schiffman et al., 2000). Individual VOCs cannot indicate contamination, but rather, combinations and patterns of different MVOCs can signal mold contamination (Kuske et al., 2006). Gobbi and Locci (2006), Kuske et al. (2006), and Schiffman et al. (2000) each used an electronic nose in the detection of VOCs, with each group reporting a near perfect classification accuracy. Rahman et al. (2016) supported the argument that electronic noses can detect concentrations of VOCs that are well below the human detection threshold, whereas Lagod et al. (2017) experimentally concluded that low MVOC concentrations are ambiguous and hinder the electronic nose's ability to detect and classify common fungal contaminants.

Detection of Molds

Mold detection methods are similar to MVOC and VOC detection methods and can also rely on the detection of MVOCs and VOCs themselves (Schiffman et al., 2000). Gas

chromatography-mass spectroscopy (GC-MS) detection of individual VOCs can indicate the compositions of different mold species with high precision, whereas an electronic nose can only give general results about the entire chemical profile of VOCs (Gobbi & Locci, 2006; Pan, Zhan, Zhu, & Tu, 2014). Other methods include mycological analysis, which is a combination of the macroscopic investigation of visible mold and the microscopic investigation of individual spores; however, this process is also expensive and especially time-consuming in allowing cultures to grow over weeks as compared to the seconds required for electronic nose usage, but mycological analysis provides more detailed, accurate results about the concentration and species of the mold contamination than the identifications by the electronic nose (Lagod et al., 2017; Paolesse et al., 2006). Polymerase chain reaction (PCR) is another method of mold detection; PCR involves the extraction of mitochondrial DNA from microbial species, and PCR specifically analyzes portions of the DNA to classify the sample of mold (Khan & Karappuayil, 2012; Lagod et al., 2017). PCR similarly has a time-consuming and expensive preparation process (Schiffman et al., 2000). Ultimately, electronic noses provide a quicker, cheaper, and mobile means of mold detection in comparison to traditional methods (Schiffman et al., 2000).

Using GC-MS and mycological analysis, Gobbi and Locci (2006), Pan et al. (2014), and Paolesse et al. (2006) confirmed the electronic nose's ability to detect species of fungal contaminants with 96.4%, 85.3%, and 96.6% classification accuracy, respectively. Schiffman et al. (2000) and Kuske et al. (2006) obtained an electronic nose classification accuracy of 80-85% and 90-95%, respectively. Schiffman et al. (2000) and Lagod et al. (2017) showed a decrease in electronic nose classification accuracy with a decrease in VOC and MVOC concentrations. Gobbi and Locci (2006), Kuske et al. (2006), Pan et al. (2014), and Schiffman et al. (2000) incubated and colonized mold species on agar plates using malt extract agar (MEA) at various

concentrations for various time periods that ranged from several days to a month; however, this procedure generally occurred with the intent to improve the training set of the electronic nose for rapid and more accurate usage outside the laboratory. Cagnasso et al. (2010), Lagod et al. (2017), and Paolesse et al. (2006) avoided the incubation and growth of fungal contaminant cultures and directly sampled the fungal contamination of materials or buildings in the training process of the electronic nose.

Electronic Nose

An electronic nose is an instrument used for machine olfaction. According to Capelli and Sironi (2016), the most popular and critical definition of an electronic nose is of Gardner and Bartlett (1994, p. 213): "An instrument, which comprises an array of electrochemical sensors with partial specificity and an appropriate pattern-recognition system, capable of recognizing simple or complex odors." Electronic noses consist of sensory array, data processing, and pattern recognition components (Gardner & Bartlett, 1994; Intelligent Optical Systems, 2000; Kuske et al., 2006; Scott et al., 2007).

Electronic noses began in the 1920's with observations of electrical current interactions with aromas, with official technology releasing in the 1980's (Wilson & Baietto, 2009). Electronic noses now measure mixtures of organic exposures without identifying the individual species within the exposure (Wilson & Baietto, 2009). The electronic nose has applications in pollution, wastes, soil, alcohol, crop cultivation, spacecraft, houses, drugs, explosives, foods, and diseases (Gobbi & Locci, 2006; Scott et al., 2007; Wilson & Baietto, 2009). The EOS series is a common brand of electronic noses used by Cagnasso et al. (2010), Eusebio et al. (2016), Falasconi et al. (2012), and Gobbi and Locci (2006) with different versions, including the EOS101, EOS835, and EOS8507 electronic noses. The Cyranose 320 has a similar electronic

nose function and uses carbon-black polymer sensors that also function by changing resistances to VOCs, which is similar to the metal oxide sensors of the EOS electronic noses (Falasconi et al., 2012; Intelligent Optical Systems, 2000; Wilson & Baietto, 2009).

The Cyranose 320 is an electronic nose from Sensigent, which uses intellectual property from Smith's Group Plc (Intelligent Optical Systems, 2000). The Cyranose 320 samples chemicals by using a sampling inlet with a two-inch needle, a purge inlet with an air filter, and an exhaust port (Intelligent Optical Systems, 2000). The Cyranose 320 has a sensor array that consists of thirty-two carbon black-polymer sensors; each sensor responds to a specific group of chemical compounds (Falasconi et al., 2012; Intelligent Optical Systems, 2000; Wilson & Baietto, 2009). The sampling period begins with the purge inlet, which cleans the sensor array using filtered air (Intelligent Optical Systems, 2000). The sampling inlet then sends one or two volumes of the sample to the sensor array (Intelligent Optical Systems, 2000). Finally, the purge inlet removes the sample by reintroducing filtered air over the sensor array (Intelligent Optical Systems, 2000). The Cyranose 320 holds up to four methods (combinations of settings); each method allows the storage of up to six training classes of chemical compounds; each training class further allows the storage of up to ten exposures each (Intelligent Optical Systems, 2000). The PCNose and CDAnalysis programs can create custom methods for the Cyranose 320, which allows for the adjustments of settings, including sampling time, purging time, sensor array temperature, sensor activity, normalization, preprocessing, pattern recognition, and cross-validation, that will allow the Cyranose 320 to calibrate itself for many purposes (Intelligent Optical Systems, 2000). The electronic nose trains itself by identifying features in the sensor data for each class of chemical compounds (Intelligent Optical Systems, 2000). The Cyranose 320 can then classify unknown chemical species by gathering the sensor features of the unknown

species and comparing these features with the features of trained classes by using pattern recognition algorithms to determine the most probable class of the unknown species (Intelligent Optical Systems, 2000). The Cyranose 320 utilizes innovations in its sensors as well as data preprocessing and normalization to accurately collect data and prepare for pattern recognition (Intelligent Optical Systems, 2000).

Data Collection and Preparation

Sensors are an essential part of data collection in electronic noses; electronic nose brands use many diverse types of sensors (Wilson & Baietto, 2009). Most electronic noses use electrochemical sensors, as they are cheap to implement and provide quantifiable electrical data, including current, voltage, and most resistance, that can function easily with pattern recognition algorithms (Wilson & Baietto, 2009). The carbon black-polymer sensors of the Cyranose 320 are a subset of electrochemical sensors, which include metal oxide sensors that function similarly in the measurement of electrical data (Wilson & Baietto, 2009). Other sensor types include piezoelectric sensors, which rely on the mass of the sensor and the change in resonance frequency, and optical fiber sensors, which detect changes in the color of light that hits the sensor (Gobbi & Locci, 2006). In comparison with other electrochemical sensors, carbon black-polymer sensors are quicker, cheaper, and more sensitive, but carbon black-polymer sensors are also more susceptible to humidity and long-term damage (Wilson & Baietto, 2009).

The carbon black-polymer sensors of the Cyranose 320 are active while the Cyranose 320 is running (Intelligent Optical Systems, 2000). The Cyranose 320 records the sensors' resistances following the initial purge stage and measures, in arbitrary units, the change in the sensors' resistances during the sampling period relative to the initial resistances ($\Delta R/R_0$) (Intelligent Optical Systems, 2000). The resistance of a sensor commonly increases as chemical

compounds accumulate on the surface of the sensor (Intelligent Optical Systems, 2000; Wilson & Baietto, 2009).

The Cyranose 320 stores each $\Delta R/R_0$ value for each exposure as a thirty-two-dimension vector (Intelligent Optical Systems, 2000). The Cyranose 320 first normalizes each vector by using $(\Delta R/R_0)_i / \Sigma|(\Delta R/R_0)|_j$, or normalization one, and later auto-scales each vector in order to weigh each sensor's data evenly and reduce the effect of vector magnitude that is caused by changes in the concentration of the sample (Intelligent Optical Systems, 2000). The Cyranose 320 finally sends the processed thirty-two-dimension vectors to feature recognition, most commonly a PCA, to reduce the number of dimensions of the vector, which allows for easier pattern recognition and visual representation of the vectors (Intelligent Optical Systems, 2000; Wilson & Baietto, 2009).

Principal Component Analysis (PCA)

Principal component analysis (PCA) is a process used by an electronic nose in compressing data by maximizing the features, or variations, between training classes (Intelligent Optical Systems, 2000; Schiffman et al., 2000; Scott et al., 2007; Wilson & Baietto, 2009). The PCA operates by projecting correlated variables of the vector using linear combinations into another orthogonal feature space to reduce the correlation and increase the variation between classes of chemical compounds (Scott et al., 2007). The resulting two or three variables (depending on settings) of the vector are principal components; principal components encapsulate all the variation of the training classes, which allows for the visual analysis of class separation (Intelligent Optical Systems, 2000; Scott et al., 2007).

The PCA displays the vectors in two or three-dimensional plots as color-coded and labeled points, which are based on class, to visualize the data (Cagnasso et al., 2010; Intelligent

Optical Systems, 2000; Kuske et al., 2006; Lagod et al., 2017; Paolesse et al., 2006). The PCA provides exploratory analysis of the training set to determine if the training samples are suitable for the detection and classification of new samples (Cagnasso et al., 2010; Gobbi & Locci, 2006; Intelligent Optical Systems, 2000). Subjectively, visibly separated classes concentrated around class centroids can imply training classes that have stronger validity (Intelligent Optical Systems, 2000).

The Cyranose 320 offers all the PCA abilities that allow for the detection of common fungal contaminants (Intelligent Optical Systems, 2000). The graphs of the PCAs of several listed studies favor the application of electronic noses in the detection of molds. Schiffman et al. (2000) could visually maximize variation between ten different classes of microbial species into distinct groups of four exposures. Lagod et al. (2017) generated a PCA without normalization, which showed a moderately strong gradient through exposures based on mold concentration. Kuske et al. (2006) showed distinct, moderately separated classes in the PCA based off many exposures of combinations of mold species and building materials. Paolesse et al. (2006) were also able to form a graph of a PCA of moderately separated classes of mold species. These subjective analyses point to a mostly favorable usage of electronic noses for fungal contaminant training sets. Furthermore, the cross-validation of the training set of pattern recognition algorithms, including KNN, can further support electronic nose usage by objectively determining the effectiveness, or cross-validity, of a training set by recording the reclassification of each exposure in the training set (Intelligent Optical Systems, 2000).

K-Nearest Neighbors (KNN)

K-nearest neighbors (KNN) is a pattern recognition algorithm that enables electronic noses to detect and identify samples of chemical species (Intelligent Optical Systems, 2000).

The KNN pattern recognition algorithm is the simplest pattern recognition algorithm used by electronic noses; researchers can also easily implement the KNN pattern recognition algorithm, but it requires prior supervision (training data) (Scott et al., 2007). The KNN pattern recognition algorithm functions by plotting the vector information of an identified sample after preprocessing, normalization, and PCA linear transformations have occurred; the algorithm then determines the k nearest vectors of training exposures on the graph of a PCA (hence the name, k-nearest neighbors) (Scott et al., 2007). The training class with the highest number of exposures of the group of the k nearest exposures then becomes the classification of the identified sample (Intelligent Optical Systems, 2000; Scott et al., 2007). Researchers commonly use the 1NN pattern recognition algorithm as it is the simplest of all the KNN pattern recognition algorithms, and the 1NN requires the least amount of time to complete as increasing the value of k exponentially increases the time necessary for computation; as stated earlier, the 1NN pattern recognition algorithm involves finding the nearest exposure to the identified sample and classifying the sample as the class of the exposure (Gobbi & Locci, 2006; Scott et al., 2007). The 1NN pattern recognition algorithm produces rigid boundaries between classes and is subject to weaknesses for outlying data and certain training sets than the 2NN pattern recognition algorithm or higher order KNN pattern recognition algorithms can nullify the effect of (Scott et al., 2007). Additionally, the KNN pattern recognition algorithm, or any other pattern recognition algorithm, can determine the quality of a training sample through cross-validation; cross-validation involves the removal of a single exposure from the training set and the temporary establishment of the exposure as an identification sample, which is then repeated for every exposure in the training set (Intelligent Optical Systems, 2000).

Schiffman et al. (2000) compared the KNN pattern recognition algorithm with a least squares pattern recognition algorithm. The KNN pattern recognition algorithm achieved a 4-11% higher class discrimination, or classification accuracy, than the least squares pattern recognition algorithm for the detection of three VOCs at low concentrations between 0.002% and 0.050%. Furthermore, the KNN pattern recognition algorithm achieved an average of 4% higher class discriminations, or classification accuracies, than the least squares pattern recognition algorithm for the detection of three fungal species over alternating days during a seven-day incubation period; when Schiffman et al. introduced two additional fungal species on two different media, the KNN pattern recognition algorithm continued to achieve an average of 10% higher class discrimination, or classification accuracy, than the least squares algorithm over alternating days during a fifteen-day incubation period (2000). Cagnasso et al. (2010) compared the 1NN pattern recognition algorithm with the SVM pattern recognition algorithm; the 1NN pattern recognition algorithm only outperformed the SVM pattern recognition algorithm in the detection of spoilage in apple juice, and the 1NN pattern recognition algorithm performed much less successfully than the SVM pattern recognition algorithm in the detection of spoilage in pear and orange juice. Gobbi and Locci (2006) found that the 1NN pattern recognition algorithm successfully detected fungal contamination of maize in twenty-seven out of twenty-eight trials. Tang, Lin, and Shyu (2010) used various KNN pattern recognition algorithm combinations in classifying methanol, ethanol, and acetone. The KNN pattern recognition algorithm alone was one of the weaker algorithms, correctly classifying only 94-95% of the sample, while the KNN pattern recognition algorithm used in conjunction with the LDA algorithm correctly classified 97-100% of the samples (Tang et al., 2010). The KNN pattern recognition algorithm used in conjunction with PCA only classified 26-49% of the samples due to the PCA's nature to

distribute all data points rather than clustering training classes (Tang et al., 2010). The KNN pattern recognition algorithm alone performed the most accurately with k = 3, whereas the KNN pattern recognition algorithm used with PCA performed the most accurately with k = 1 (Tang et al., 2010). Overall, Gobbi and Locci (2006) and Schiffman et al. (2000) showed moderate favorability towards the KNN pattern recognition algorithm, whereas Cagnasso et al. (2010) and Tang et al. (2010) showed quite the opposite. The Cyranose 320 offers the KNN pattern recognition algorithm with customizable k, but the electronic nose also has access to other pattern recognition algorithms, including k-means, SVM, and CDA (Intelligent Optical Systems, 2000).

Other Pattern Recognition Algorithms

The k-means pattern recognition algorithm minimizes distances in the PCA of exposures from a class to the centroid of the class through unsupervised recursive partitions of the training set of exposures into k sections (Intelligent Optical Systems, 2000; Jain, Murty, & Flynn, 1999; Scott et al., 2007). The k-means pattern recognition algorithm is popular, simple, and easily implemented, but the pattern recognition algorithm faces a potential weakness in the first partition; a poorly selected first partition may result in the incorrect convergence of the recursion, which decreases the potential accuracy of electronic noses (Jain et al., 1999; Scott et al., 2007). Running time for the k-means pattern recognition algorithm is proportional to the number of classes, partitions, and exposures, but the k-means pattern recognition algorithm is generally faster than other pattern recognition algorithms (Jain et al., 1999; Scott et al., 2007). However, the k-means pattern recognition algorithm only works well for datasets with isotropic clusters (Jain et al., 1999).

The SVM pattern recognition algorithm operates off statistical learning theory by optimizing class bounds using inequalities and through maximizing hyperplane margins (distances from hyperplanes to the nearest point of separated classes) (Intelligent Optical Systems, 2000; Pardo & Sberveglieri, 2005; Zhang, Liu, & Deng, 2017). The SVM pattern recognition algorithm is a new algorithm that benefits from lower analysis times and higher efficiencies (Huang et al., 2016). The SVM pattern recognition algorithm specializes in binary classifications, which renders the algorithm less accurate for multi-class classifications; on average, the SVM pattern recognition algorithm had 12-20% lower classification rates, or classification accuracies, for detection of formaldehyde, toluene, and carbon monoxide in comparison to a modified multi-class SVM pattern recognition algorithm (Huang et al., 2016). Rahman et al. (2016) found similar results regarding the weaknesses of the SVM pattern recognition algorithm in multi-class detections. Furthermore, a low variance of the principal components can lower the classification and detection accuracies of the SVM pattern recognition algorithm (Pardo & Sberveglieri, 2005). Cagnasso et al. (2010) found that the SVM pattern recognition algorithm performed more accurately than the 1NN pattern recognition algorithm in the detection of microbes in fruit juices by up to 6% in classification accuracy. Conversely, Zhang et al. (2017) found that the SVM pattern recognition algorithm performed up to 8% worse in classification accuracy than the LDA algorithm on several tests of classification of chemical species. Hong, Wang, and Qi (2015) found similar results of the SVM pattern recognition algorithm, which performed 2-13% worse in terms of classification accuracy than the nearly perfect LDA algorithm in the detection of tomato juices. Overall, these studies find the SVM pattern recognition algorithm to be more accurate in comparison to the KNN pattern recognition

algorithm, but the SVM pattern recognition algorithm lacks in accuracy for multi-class classifications in comparison to the LDA pattern recognition algorithm.

The CDA pattern recognition algorithm is another pattern recognition algorithm that is offered by the Cyranose 320; the CDA pattern recognition algorithm is a multi-class subset of the LDA algorithm (Intelligent Optical Systems, 2000). The LDA algorithm offers capabilities of PCA in reducing variable dimensions and maximizing interclass variation while minimizing the variation between the exposures of a single class (Scott et al., 2007; Tang et al., 2010; Zhang et al., 2017). Several consider the pattern recognition and classification abilities of the LDA algorithm to be simple, quick, and superior to the KNN pattern recognition algorithm, SVM pattern recognition algorithm, and many other pattern recognition algorithms for larger class sizes and amounts of classes, while also being able to enhance other pattern recognition algorithms using the LDA algorithm's data compression, which can outperform the compression of PCA (Hong et al., 2015; Schiffman et al., 2000; Scott et al., 2007; Tang et al., 2010; Zhang et al., 2017). These studies show the LDA algorithm and its CDA pattern recognition algorithm subcomponent as accurate pattern recognition algorithms for multi-class electronic nose classification (Hong et al., 2015; Scott et al., 2007; Tang et al., 2010).

Summary

Electronic noses can detect and classify common indoor contaminants in the prevention of sick building syndrome (SBS), which is the widespread symptoms experienced during an occupant's time spent in a building. Electronic noses are sensor instruments that are used for machine olfaction. However, Khan and Karappuayil (2012) and Lagod et al. (2017) have shown somewhat limited accuracy for electronic nose detection in comparison with gas chromatography-mass spectrometry, mycological analysis, and polymerase chain reaction.

Studies have been inconsistent with the pattern recognition algorithm and molds used for detection. Additionally, studies of American brands, including the Cyranose 320, have been sparse for this subject in comparison to the numerous studies of European brands, including the EOS series. Eusebio et al. (2016) and Marco (2014) have cited the inconsistencies between the electronic nose studies and wish to define a common standard for the pattern recognition algorithms used across studies.

Chapter 3: Research Method

Occupants of up to 30% of new and remodeled buildings faced poor indoor air quality that leads to SBS (EPA, 1991). Electronic noses provided an optimistic approach over other traditional methods in the detection of molds that caused SBS, but the accuracy of electronic noses was debatable, facing issues with low chemical species concentrations, false readings, inability to detect individual chemical compounds, and lack of common usage and methodology, which included the varying usage of pattern recognition by electronic noses and electronic nose studies (Eusebio et al., 2016; Gobbi & Locci, 2006; Lagod et al., 2017; Schiffman et al., 2000). The purpose of this study was to establish a recommended pattern recognition algorithm for the electronic nose detection and classification of common indoor fungal contaminants to answer the question(s): "How does the classification (and detection) of common indoor fungal contaminants vary for different electronic nose pattern recognition algorithms?" The corresponding hypothesis that I analyzed through this methodology follows: "Different pattern recognition algorithms of an electronic nose will have different effects on an electronic nose's ability to detect and classify common fungal contaminants." Previous literature has differed on the classification and detection accuracies of each pattern recognition algorithm. I addressed these issues with the methodology and have explained my decisions below.

Research Method and Design

A quantitative, experimental study took place to determine the accuracy of an electronic nose in the detection of common fungal contaminants. Quantitative data, or numerical data in the form of the proportion of correctly classified exposures, was an appropriate measure for the classification accuracy and detection accuracy of the electronic nose, as used by Cagnasso et al. (2010), Kuske et al. (2006), and Schiffman et al. (2000). The experimental design was appropriate for the study as the treatments of different pattern recognition algorithms were independently applicable to all members of the sample due to the assumption that the use of one pattern recognition algorithm to identify an exposure does not affect the use of another pattern recognition algorithm to identify the same exposure (Intelligent Optical Systems, 2000). The experimental design controlled confounding variables and purposely altered an independent variable, which allowed the researchers to draw conclusions regarding causation. Due to the nature of the treatments, I was able to apply the four independent treatments on each member of the sample rather than randomly assigning one treatment to each member of the sample.

The variables of this study closely followed those of the studies of Hong et al. (2015) and Zhang et al. (2017) but differed in the employment of controls in measuring the effect of pattern recognition algorithms. The independent variable of this study was the pattern recognition algorithm applied to each exposure: KNN, SVM, CDA, and k-means. The dependent variables were the classification accuracy (correct class or incorrect class) and detection accuracy (detection of contamination or no detection of contamination) of the electronic nose. I measured the dependent variables as the proportion of correctly classified and proportion of correctly detected exposures from the electronic nose. These variables were appropriate in this study's

purpose to determine an algorithm with the highest classification and detection accuracies, as used in Hong et al. (2015), Schiffman et al. (2000), and Zhang et al. (2017).

Population and Sample

The population of the research was the exposures of fungal contaminants by electronic nose sensors. An exposure consisted of a multiple dimension vector that contained components of the relative change in the resistances of each of the electronic nose sensors ($\Delta R/R_0$) (Intelligent Optical Systems, 2000). The population was appropriate based on the treatments of the KNN, k-means, SVM, and CDA pattern recognition algorithms that classified each vector of the sampling data.

The sample of this study consisted of the two hundred exposures, with fifty exposures coming from each of *Aspergillus*, *Cladosporium*, *Penicillium*, and *Alternaria* genera, of fungal contaminants that satisfied a Cochran's Q test. The sample was not completely representative of the population of all exposures of all fungal contaminants as I excluded rarer fungal contaminants from the study due to the difficult acquisition of these genera and biosafety concerns for toxic genera. Rarer fungal contaminants included *Acremonium* (9% of contamination cases), *Candida* (4% of contamination cases), *Stachybotrys* (4% of contamination cases), and *Trichoderma* (11% of contamination cases) (Shelton et al., 2002). A convenience sample utilized the nearest resources or population as members of a sample, which allowed for cheaper and quicker research while limiting the generalizability of the data due to potential confounding variables. The sample remained a convenience sample as a true random sample was impossible to obtain, but the external validity remained intact as these genera resulted in the majority of fungal contamination cases. Furthermore, the exposures in the sample were representative of most of the exposures of the population of all exposures as most electronic

noses used the relative change in the resistances of individual sensors as components of a multi-dimensional vector that the Cyranose 320 sent to a pattern recognition algorithm. The only potential difference between the exposures of this study's sample and the exposures of the population of all fungal exposures was the dimensionality. The thirty-two dimensions of the Cyranose 320 exposures proved sufficient in comparison to other studies, including Gobbi and Locci (2006), Kuske et al. (2006), Paolesse et al. (2006), and Schiffman et al. (2000), which used six-dimensional, twelve-dimensional, eight-dimensional, and fifteen-dimensional (number of sensors) exposures, respectively, and received training sets with high cross-validation. The dimensionality of each vector had no major effect on pattern recognition, given that there were enough dimensions from sensors that allowed for feature recognition by pattern recognition algorithms.

Materials and Instruments

I utilized the Cyranose 320, which functioned off carbon black-polymer sensors. Based on descriptions by Wilson and Baietto (the 2009), the functionality of the Cyranose 320 sensors that I used is similar to that of the electrochemical sensors present in most common electronic nose brands. The Cyranose 320 also provided the machine olfaction features present in most common electronic nose brands: pattern recognition, data preprocessing, and controlled sensory array setup (Intelligent Optical Systems 2000; Wilson & Baietto, 2009). The Cyranose 320 functioned with the PCNose computer program to transfer data between the electronic nose and a personal computer; the CDAnalysis computer program performed in-depth analysis of the data retrieved by the PCNose program (Intelligent Optical Systems, 2000). Each of these computer programs ran on a laptop installed with Windows ten.

To increase the replicability and parallel-forms reliability of this study, I used potato dextrose agar (1% potato extract, 5% dextrose, 4% agar, and 1% tartaric acid in a colloid) plates, which allowed for the isolation and cultivation of fungal contaminants. The agar plates had a diameter of ninety millimeters and a height of fifteen millimeters. BBL™ CultureSwab™ swabs sampled various surfaces for microbes and inoculated the potato dextrose agar plates. The BBL™ CultureSwab™ swabs allowed the storage of the air vent samples in a sealed enclosure, which prevented bacterial and foreign fungal contamination from occurring. A ProgRes CapturePro 2.5 microscope camera analyzed the species of each cultivated agar plate at ten, forty, one hundred, and four hundred magnifications as it was the only microscope available on the campus. To increase the reliability of this study, I swabbed cultivated mold samples from the potato dextrose agar plate onto a twenty-five by seventy-five-millimeter plastic microscope slide that contained a drop of distilled water to adhere a one-centimeter diameter circular cover slip over the slide. I used a fungal identification manual by Navi, Bandyopadhyay, Hall, and Bramel-Cox (1999) to classify the species of the fungus on the slide. Navi et al. (1999) compiled micrographs of various fungal species, including those of the *Alternaria, Aspergillus, Cladosporium,* and *Penicillium* genera of this study. I used five hundred milliliter airtight clear containers, utilized in a study by Kuske et al. (2006), with a coverable five-millimeter circular hole drilled into the center of the lid of each container that captured VOCs produced by exposed fungal contaminant samples, which allowed for the simulation of accumulation of VOCs in a room that contained fungal accumulation and allowed for the sampling of the electronic nose through a two-inch needle. Most of these materials were replaceable with similar substitutes, which kept the reliability of this study high. This was true for the BBL™ CultureSwab™ swabs and ProgRes CapturePro 2.5, which were the most convenient materials already present in my

laboratory and in most biology laboratories across the world. Finally, I used an excel sheet to store data and conduct statistical analyses. Availability and specificity of these materials increased the reliability of this study.

Operational Definition of Variables

Classification accuracy. Classification accuracy is the proportion of exposures of a specific mold species in which the electronic nose, using the identification setting and a training set, correctly displays the species of the mold; I calculated it as the quotient of the total number of correctly classified exposures out of the total number of exposures (Schiffman et al., 2000).

Detection accuracy. Detection accuracy is the proportion of exposures of a specific mold species in which the electronic nose, using the identification setting and a training set, correctly displays any species of the mold; I calculated it as the quotient of the total number of correctly detected exposures out of the total number of exposures (Schiffman et al., 2000).

Cross-validity. Cross-validity is the proportion of exposures specific mold species in the training set in which the electronic nose correctly uses pattern recognition algorithms to determine how useful the training set is; I found it by running the pattern recognition algorithm for the training set by removing only one exposure at a time and identifying the class of the exposure for the entire training set; I calculated it as the quotient of the total number of correctly classified exposures in the training set out of the total number of exposures in the training set (Intelligent Optical Systems, 2000).

Data Collection

For the first phase of research preparation, I used BBL™ CultureSwab™ swabs to sample a hall vent and a bathroom vent of various buildings on a college campus that potentially contained indoor mold contamination to generate the supply of molds. There was a concern for

the reliability of this study as I had sampled surfaces in a specific geographic location. However, spores varied little based on geographic location as most genera have a wide geographic range, which kept the external validity of this study high. Therefore, the source of the mold samples was insignificant in this test, allowing for the increased reliability of this study. Spores tended to accumulate around air vents at higher quantities, which allowed for a more representative sample of indoor fungal contamination. Furthermore, the inspection of air vents provided a more complete representation of rooms in comparison to the inspection of open-air agar plates placed in single locations of a room which captured huge variations in specimens (Kuhn & Ghannoum, 2003).

I transferred each swabbed sample to an individual potato dextrose agar plate by gently rubbing the BBL™ CultureSwab™ swab onto the potato dextrose agar in a zigzag pattern, and I cultivated the potato dextrose agar plates over seven days at room temperature, which simulated growth of indoor fungal contaminants in the temperature of a normal room. The temperature of the location ranged between seventy and seventy-five degrees Fahrenheit, which was representative of normal room temperature. The temperature of the room prevented the spores of thermophilic and psychrophilic molds from activating, which often did not pose a threat of indoor contamination. However, this did pose a reliability issue with laboratories of different temperatures, but incubators and refrigerators were usable instruments that could have accounted for this issue in other laboratories. Using a marker, I labeled each potato dextrose agar plate from one through ten based off the used BBL™ CultureSwab™ swab that inoculated the plate. I left each potato dextrose agar plate upside-down, which reduced the possibility of outside contamination and the effect of condensation on the surface of the potato dextrose agar solution, thus increasing the internal validity of this study. I classified each sample present on the potato

dextrose agar plates through mycological analysis by using ProgRes CapturePro 2.5 microscope observations and comparing them with images from the fungal identification manual by Navi et al. (1999). I separated each species (*Aspergillus, Cladosporium, Penicillium,* and *Alternaria* species) and, using BBL™ CultureSwab™ swabs, inoculated each species onto new, separate potato dextrose agar plates, labeled "Asp", "Alt", "Cld", and "Pen", for the second phase of research. I inoculated the second batch of potato dextrose agar plates by using BBL™ CultureSwab™ swabs to gather the spores of mold from the first samples and to gently disperse the spores onto the new agar plates in a circular pattern that covered the entire plate. This ensured the reliability of this study as the potato dextrose agar plates had an even layer of mold.

For the second stage of research preparation, I cultivated the inoculated potato dextrose agar plates at room temperature for seven days with visual observations on varying days–using the fungal identification manual by Navi et al. (1999)–which ensured that environmental bacteria , environmental molds, or mold contaminants from other samples did not contaminate the potato dextrose agar plates. On day seven, I opened each of the four potato dextrose agar plates and set down the exposed potato dextrose agar plates, with each on the bottom of a prepared five hundred milliliter container. The potato dextrose agar plates sat for thirty minutes, allowing for VOC accumulation, which simulated the accumulation of VOCs in rooms with actual fungal contamination and increased the external validity and reliability of this study.

Prior to sampling, I calibrated the Cyranose 320 by allowing filtered air to flow through the electronic nose's sensor array for the five minutes before the potato dextrose agar plates had sat for a total of thirty minutes. The calibration reduced a possible confounding variable by removing chemical species that may have accumulated on the sensors while the Cyranose 320 was inactive (Intelligent Optical Systems, 2000). Though calibration may have reduced

reliability for electronic noses without this function, this step was necessary to maintain the internal validity of this study. Following the initial thirty minutes, the Cyranose 320 sampled each airtight container by using its two-inch sampling needle, which I inserted through the five-millimeter hole in the lid of each container. I repeated this sampling ten times using the electronic nose's train function to create a training class. While the Cyranose 320 was not sampling, I sealed off the holes of each container with a one-inch long piece of adhesive tape to prevent any VOCs from escaping, which kept the data consistent and blocked any spores or toxins from escaping and becoming a health hazard. To increase the reliability of this study, the Cyranose 320 used the following recommended settings when sampling each airtight container: high baseline purge for three seconds, high sample draw one for eight seconds, high first air intake purge for two seconds, high second sample gas purge for two seconds, activated digital filtering, substrate heating at 42.0^{o}C, auto-scaling preprocessing, normalization one, medium identification quality), and activated all sensors (Intelligent Optical Systems, 2000). The high pump speed accelerated the sampling process for each exposure, which allowed for shorter purge and intake periods. A constant normalization, preprocessing, substrate heat, digital filtering, and sensor setting kept the exposures controlled for many variables, thus increasing the internal validity of this study in addition to the increased reliability of this study. I exported these training classes from the Cyranose 320 to a laptop. I ran each pattern recognition algorithm on the training set on the CDAnalysis program to determine the cross-validation of the training set for each pattern recognition algorithm.

The data collection stage followed immediately after the creation of the training set. I first formed the sample of the population by using the Cyranose 320's identify function with the same electronic nose settings and procedures as I had used with the previous exposures of the

training set. I sampled each airtight container sixty times, which created a sample of 240 exposures. I exported the exposures from the Cyranose 320 to a laptop. On the PCNose program, I created a PCA of the training set and every identification exposure before removing the ten most outlying exposures per identification class to normalize data and improve the validity by negating the effect of outliers. The sample size reduced to two hundred, which satisfied the conditions for a Cochran's Q test. The Cochran's Q test was an appropriate statistical test as it searched for statistically significant differences between multiple treatments applied to the same sample members (Laerd Statistics, n.d.). I determined the ten most outlying exposures per identification class by sorting the Euclidian distances, or distance between one vector and all other vectors, of each exposure and removing the ten exposures with the highest Euclidian distances before exporting the sample to the CDAnalysis program.

Data Processing

Following the creation of the sample, I applied the KNN, SVM, CDA, and k-means pattern recognition algorithm treatment to every member of the sample of the CDAnalysis program and recorded whether the pattern recognition algorithm correctly identified the class of the exposure. If the pattern recognition algorithm determined the class of the exposure to be the same as the same class, or genus of mold, from which the Cyranose 320 sampled the exposure, then the classification for the exposure was correct. I recorded the correct and incorrect classifications for all exposures from the sample of two hundred exposures and calculated the proportion of the classifications that were correct. If the pattern recognition algorithm determined any class, or genus of mold, for the exposure, then the detection for the exposure was correct. I kept this data in an excel sheet for every exposure of the sample. I calculated the

classification and detection accuracies as operationally defined earlier, which allowed for me to perform statistical analysis on the hypotheses.

Data Analysis

I ran a Cochran's Q test on the obtained classification data on an excel sheet. The Cochran's Q test assumed a nominal independent variable of three or more treatments and dichotomous (binary) dependent variable, which were representative of the four independent variable treatments (KNN, SVM, CDA, and k-means pattern recognition algorithms) and two dependent variable outcomes (correctly classified and incorrectly classified) (Laerd Statistics, n.d.). The Cochran's Q test also required an assumption of a large enough sample size for statistically significant differences to exist between proportions due to actual differences in the treatments rather than random chance (Laerd Statistics, n.d.). I expected 75-95% classification rate, which prompted a sample size of about two hundred per experimental treatment group. The final assumption of Cochran's Q test called for a randomized sample of the population (Laerd Statistics, n.d.). Due to the use of a convenience sample, it was not possible to meet the assumption of a randomized sample as a randomized sample was impractical and nearly impossible to have obtained. However, the sample was representative of the population as I used four common mold genera, so the external validity of this study remained high.

The null hypotheses stated that there is no difference in the classification (or detection) of common indoor fungal contaminants for different electronic nose pattern recognition algorithms: $\pi_{KNN} = \pi_{k\text{-means}} = \pi_{SVM} = \pi_{CDA}$ where π is the classification (or detection) accuracy. The alternative hypotheses stated that there is a difference in the classification (or detection) of common indoor fungal contaminants for different electronic nose pattern recognition algorithms: $\pi_a \neq \pi_b$; $1 \leq a, b \leq 4$ where π is the classification (or detection) accuracy and a and b are pattern

recognition aglgorithms. I established the significance level at the commonly accepted α level of 0.05 and used three degrees of freedom when establishing the chi-squared statistic. The rejection of the null hypotheses would have resulted in the post hoc performance of six pairwise McNemar's Chi-Square tests between combinations of two treatments to locate significant differences between the accuracies of specific pairs of treatments. The assumptions of McNemar's Chi-Square tests included a dichotomous independent variable (two pattern recognition algorithms) and a dichotomous dependent variable ("correct" or "incorrect") (Laerd Statistics, n.d.). The McNemar's Chi-Square test also assumed a randomized sample, which I covered in Cochran's Q test.

Assumptions

Assumptions of studies were factors that were out of control but still justified by the researcher (Simon, 2011). A research assumption of this study required that the lab-grown molds represented actual fungal contamination in buildings. This assumption required that no bacteria or foreign fungal species in the laboratory contaminated the potato dextrose agar plates. Acumedia (2017), a producer of potato dextrose agar, reported that tartaric acid present in the solution decreased the pH, which inhibited bacterial growth. Furthermore, the potato dextrose agar plates remained sealed during cultivation, and I inoculated the potato dextrose agar plates and took slide samples in a sterile Laminar flow hood in the BSL-2 laboratory, which generated airflow to prevent the contamination of the potato dextrose agar plates. The potato dextrose agar simulated an environment similar to that of damp, unpolished wood, for which actual fungal contamination would have occurred on. Therefore, this assumption remained accurate, which increased the external validity of this study.

Another major assumption was the creation of a valid training set of training exposures. The accumulation of VOCs and MVOCs in airtight containers and controlled sampling by the electronic nose through the use of the same sampling procedures and Cyranose 320 settings ensured a stronger training set. This assumption later verified by a subjectively high cross-validation of the training set by each pattern recognition algorithm.

An additional basic assumption included the proper functioning of the Cyranose 320. Prior to experimentation, I sampled several volatile chemicals using the Cyranose 320 and PCNose and subjectively inferred that the electronic nose was functioning correctly. A final assumption was that the humidity of each airtight container was the same. Eusebio et al. (2016) found that humidity had a major impact on an electronic nose's detection of chemical species. The assumption remained intact as I handled each airtight container with the same procedures and exposed each container to the same open air prior to inserting the open potato dextrose agar plates.

Limitations

Limitations of studies were factors that were out of control for the researcher confined their effects (Simon, 2011). The use of electronic noses was relatively limited due to the initial relatively inflated cost of manufacturing, though the applications of an electronic nose proved to be potentially cost-effective. I only used a single Cyranose 320 electronic nose in this study due to limits on funding and resources for obtaining the electronic noses and the redundancy of using multiple electronic noses in practical aspects, thus limiting the external validity of this study. Sensigent has manufactured Cyranose 320 electronic noses with a patented design, thus greatly reducing the variability, minimizing effects on the external validity, and increasing the reliability of this study (Intelligent Optical Systems, 2000). Furthermore, many other electronic nose

brands contained similar electrochemical sensors and similar, if not exact, pattern recognition algorithms for classifying samples (Wilson & Baietto, 2009). Thus, the external validity of this study remained high.

The cultivation of mold faced several limitations in the methodology. I only cultivated a single agar plate of each mold genera due to the limited resources including the availability of the potato dextrose agar solution. The use of a single plate did not have significant effects on the validity of this study as the mutation rate of molds remained insignificant; the mutation rate of *Aspergillus* spores was roughly one in sixty thousand, and similar values existed for other mold genera (Baracho & Baracho, 2003). Thus, the mold samples were representative of the general population of common fungal contaminants, which increased the external validity of this study.

Delimitations

Assumptions of studies were factors that were controllable and justifiable by the researcher (Simon, 2011). In addition to the limitations imposed on the general cultivation of mold, I narrowed the specific genera cultivated. Eusebio et al. (2016), Kuske et al. (2006), and Lagod et al. (2017) cited building materials, mold concentration, age, and humidity as possible confounding factors to electronic nose detection. I cultivated the molds in the controlled habitat of a potato dextrose agar plate to reduce these confounding factors. I only sampled the four genera, *Aspergillus, Cladosporium, Penicillium,* and *Alternaria*, in the study as they were the most common SBS-causing fungal contaminants that were present in buildings (Shelton et al., 2000). The limitation of genera left a higher reliability in the reproduction of this study.

Another delimitation of this study was the treatments used. I used the KNN, CDA, SVM, and k-means pattern recognition algorithms. These pattern recognition algorithms were available by most electronic noses and programmable by statistical software if they were not available

(Scott et al., 2006). Furthermore, these pattern recognition algorithms performed well in the studies analyzed in the previous chapter.

I imposed a further delimitation on the collected data. When identifying a chemical species, the Cyranose 320 outputted the confidence of its identification of the class of the chemical species proportionally to the number of stars, ranging from one through five (Intelligent Optical Systems, 2000). I only considered the accuracy of the Cyranose 320 identifications since including the confidence of each identification would not have met the requirements for a Cochran's Q test. Furthermore, the incorporation of identification confidence would have decreased the reliability of this study for others when using electronic noses without this feature, so this delimitation was necessary for this study.

Ethical Assurances

I handled the mold with preemptive caution during inoculation, growth, exposure, and disposal. Researchers wore laboratory clothing, gloves, surgical masks, and goggles while preparing samples and collecting data. The data collection occurred in the Kennesaw State University Marietta Campus biology laboratory, which was a certified biosafety level two laboratory. I submitted safety and precaution procedures to an institutional review board prior to data collection. I thoroughly sanitized the potato dextrose agar plates containing the fungal samples and removed the remaining potato dextrose agar solution following data collection. I disposed of the BBL™ CultureSwab™ swabs and potato dextrose agar plates following the conclusion of data collection.

Summary

This quantitative, experimental study aimed to determine the effects of pattern recognition algorithms on an electronic nose's ability to detect common fungal contaminants to

improve the limited accuracy of electronic noses. I measured the classification accuracy of each treatment based on the correctness of the pattern recognition algorithm's classification of each of two hundred exposures in the sample of the population. I prepared mold samples in agar plates in two rounds of cultivation. The Cyranose 320 sampled an accumulation of VOCs produced by the molds to generate exposures. I applied the treatments of the KNN, SVM, CDA, and k-means pattern recognition algorithms to each exposure on the CDAnalysis software to collect data. I used a Cochran's Q test to search for significant results from the use of different pattern recognition algorithms and subsequent McNemar's Chi-Square tests if significant results existed. The study faced external validity limitations based on the electronic nose, cultivation process, and molds. I delimited the treatments used and data measured to narrow the scope of the study towards its purpose. I attempted to limit confounding variables by taking measures to control humidity and training sets. These measures allowed the analysis and application of the results to external applications in chapters four and five.

Chapter 4: Findings

Electronic noses have shown significant capability in the detection of common fungal contaminants, but they still faced setbacks, including false readings, susceptibility to humidity, and varied sensor and pattern recognition usage (Eusebio et al., 2016; Lagod et al., 2017; Schiffman et al., 2000). The purpose of this study was to establish a recommended pattern recognition algorithm for the detection and classification of common indoor fungal contaminants using an electronic nose. In a controlled laboratory environment, I utilized the Cyranose 320 to sample various mold genera grown on potato dextrose agar plates. I processed the resulting exposures through several pattern recognition algorithms to experimentally fulfill the purpose of this study.

Results

Following the collection of samples, I exported the data from the Cyranose 320 to the PCNose program to convert the exposures into a data file readable by the CDAnalysis program. I ran the pattern recognition algorithms on the CDAnalysis program and extracted the identification data from the CDAnalysis program into an excel spreadsheet. I placed the raw data in a two-way table organized by pattern recognition algorithm and exposure number, with each individual cell listing the output of the pattern recognition algorithm in Appendix A. I processed these outputs into dichotomous correct and incorrect classifications and true and false detections, which allowed for the completion of a Cochran's Q test and McNemar's Chi-Square tests to assess my hypotheses. I further organized the raw data by mold genera, which allowed the completion of further Cochran's Q tests and McNemar's Chi-Square tests to analyze my sub-questions.

Summary data. Table 1 in Appendix A listed the marginal relative frequencies of the classifications of exposures of each pattern recognition algorithm. All the pattern recognition algorithms correctly classified *Aspergillus* with the highest accuracy of 96-98%, followed by *Penicillium* with 80-90% classification accuracy. The k-means and SVM pattern recognition algorithms classified *Cladosporium* the worst, at only a 38% and 58% classification accuracy, respectively. The KNN and CDA pattern recognition algorithms classified *Alternaria* the worst, at only a 48% and 46%, respectively. The pattern recognition algorithms overall had lower frequencies of *Alternaria* and *Cladosporium* classifications and higher frequencies of *Aspergillus* and *Penicillium* classifications. The pattern recognition algorithms misclassified *Alternaria* and *Cladosporium* exposures as *Aspergillus* or *Penicillium* in an average of 21% of those exposures. Only the SVM and CDA pattern recognition algorithms failed to detect 14% and 2.5% of the

exposures, respectively. The KNN and k-means pattern recognition algorithms produced a 92.5% cross-validity for the training set. The SVM pattern recognition algorithm performed its own cross-validation, which achieved a cross-validity of 96.9%. The CDA pattern recognition algorithm had a much lower cross-validity of only 65%.

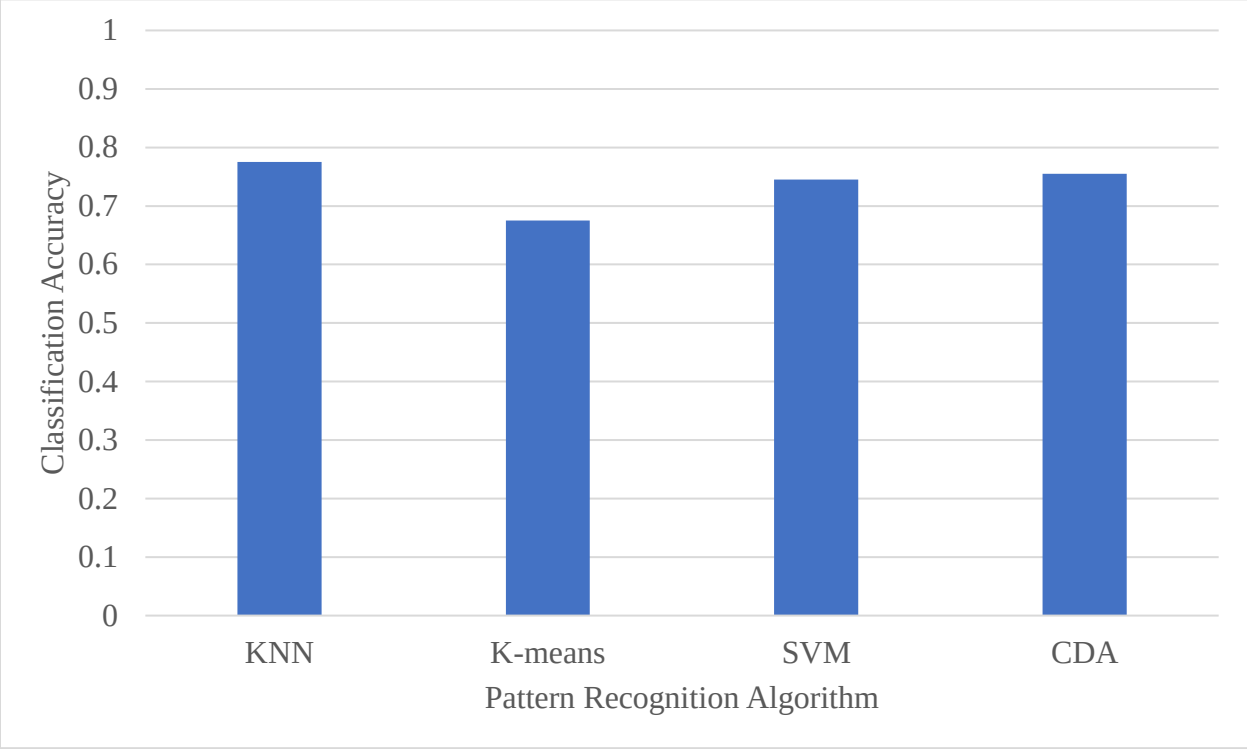

Figure 1. Bar chart displaying the classification accuracy of the KNN, K-means, SVM, and CDA pattern recognition algorithms.

Classification accuracy. The KNN pattern recognition algorithm had the highest classification accuracy, correctly classifying 155 out of 200 exposures, followed by the CDA pattern recognition algorithm (151 out of 200), SVM pattern recognition algorithm (149 out of 200), and k-means pattern recognition algorithm (135 out of 200). This small variation was much more significant when looking at exposures individually. Of the two hundred exposures, 112 had correct classifications by all the pattern recognition algorithms and twenty-one had correct classifications by none of the pattern recognition algorithms. Additionally, twenty-four exposures had only one correct classification, eleven exposures had two correct classifications,

and thirty-two exposures had three correct classifications. This variation significantly contributed to the result of the Cochran's Q test, which objectively analyzed the question: "How does the classification of common indoor fungal contaminants vary for different electronic nose pattern recognition algorithms?"

To perform the Cochran's Q test, I satisfied the assumptions of a nominal independent variable of four pattern recognition algorithms, a dichotomous dependent variable of correct or incorrect classifications, and a reasonably large sample size of two hundred exposures. The final assumption required a randomized sample, for which I did not meet. I minimized its effect on validity by creating a representative convenience sample, which consisted of several genera of fungal contaminant exposures, and justifying the difficulty of obtaining a truly random sample. The Cochran's Q test for the classification accuracy of pattern recognition algorithms yielded a Q-statistic of 12.849, which was greater than the critical chi-squared value of 7.815 obtained from a 0.05 significance level and three degrees of freedom. Due to this statistically significant outcome, I rejected the null hypothesis of no difference in the classification accuracy between pattern recognition algorithms. The data further supported the alternative and research hypotheses, which showed that different pattern recognition algorithms have different effects on classification accuracy.

Table 1

Joint Frequencies of Pattern Recognition Classifications and McNemar's Test Statistics

	Combinations of Pattern Recognition Algorithms					
Measurement	KNN/K-M	KNN/SVM	KNN/CDA	K-M/SVM	K-M/CDA	SVM/CDA
Correct/Correct	120	140	144	118	119	138
Correct/Incorrect	35	15	11	17	16	11
Incorrect/Correct	15	9	7	31	32	13
Incorrect/Incorrect	30	36	38	34	33	38
Degrees of Freedom	1	1	1	1	1	1
Significance Level	0.05	0.05	0.05	0.05	0.05	0.05
Critical Value	3.84	3.84	3.84	3.84	3.84	3.84

Test Statistic	8.00	1.50	0.89	4.08	5.33	0.17
p-value	0.0047	0.2207	0.3455	0.0434	0.0210	0.6801

Note. A value of 35 for Correct/Incorrect and KNN/K-M represents that the KNN pattern recognition algorithm correctly classified thirty-five exposures that the k-means pattern recognition algorithm failed to classify.

McNemar's Chi-Square tests. I followed up with multiple post hoc McNemar's Chi-Square tests to locate specific statistically significant differences between pairs of pattern recognition algorithms. The McNemar's Chi-Square test was appropriate in the analysis of the significant differences found by Cochran's Q test. To perform the McNemar's Chi-Square tests, I satisfied the assumptions of a dichotomous independent variable of two pattern recognition algorithms, a dichotomous dependent variable of correct or incorrect classifications, and reasonably large joint frequencies of discordant exposures that were between eighteen and fifty. I failed to meet the randomized sample assumption, but I limited its effects as discussed earlier.

In Table 1, the obtained McNemar's Chi-Square test statistics of 8.000, 4.083, and 5.333 for the pairs of the k-means and KNN, k-means and SVM, and k-means and CDA pattern recognition algorithms, respectively, were greater than the chi-square critical value of 3.814 from a 0.05 significance level and one degree of freedom, while the obtained McNemar's Chi-Square test statistics of 1.500, 0.889, and 0.167 for the pairs of the KNN and SVM, KNN and CDA, and SVM and CDA pattern recognition algorithms, respectively, were less than the chi-square critical value. I rejected the null hypothesis of no difference in the classification accuracy between the pairs of the k-means and KNN, k-means and SVM, and k-means and CDA pattern recognition algorithms. I failed to reject the null hypothesis of no difference in the classification accuracy between the pairs of the KNN and SVM, KNN and CDA, and SVM and CDA pattern recognition algorithms.

Table 2

Joint Frequencies of Correct Detections by Genera and Pattern Recognition Algorithm

	Actual Exposure Class			
Pattern Recognition Algorithm	Alternaria	Aspergillus	Cladosporium	Penicillium
KNN	24	24	32	23
K-means	48	49	48	48
SVM	39	19	29	35
CDA	44	43	40	45

Note. There are fifty total exposures for each joint frequency.

Alternaria, Aspergillus, Cladosporium, and Penicillium. The results addressed the four sub-questions regarding classification accuracy: "How does the classification of *Alternaria, Aspergillus, Cladosporium,* and *Penicillium* vary for different electronic nose pattern recognition algorithms?" The SVM, k-means, KNN, and CDA pattern recognition algorithms had the highest classification accuracy with thirty-four, forty-nine, thirty-nine, and forty-five out of fifty identified exposures for the classification of *Alternaria, Aspergillus, Cladosporium, and Penicillium*, respectively.

I performed a Cochran's Q test to answer each sub-question, and I met the assumptions made in the first Q test, though with a lower external validity due to the reduced sample size of only fifty exposures per genera. The obtained Q-statistics of 8.44 and 26.192 for *Alternaria* and *Cladosporium*, respectively, were greater than the chi-squared critical value of 7.815 for a 0.05 significance level and three degrees of freedom, while the obtained Q-statistics of 1.286 and 6.462 for *Aspergillus* and *Penicillium*, respectively, were less than the chi-squared critical value. I rejected the null hypothesis of no difference in the classification accuracy between pattern recognition algorithms for *Alternaria* and *Cladosporium* and performed post hoc McNemar's Chi-Square tests to locate significant differences between pattern recognition algorithms. I failed

to reject the null hypothesis of no difference in the classification accuracy between pattern recognition algorithms for *Aspergillus* and *Penicillium*.

Due to the decreased quantity of discordant joint frequencies, a binomial distribution modeled McNemar's Chi-Square tests. Tables 1 and 2 in Appendix B showed statistical significance for a difference in the classification accuracy between the pairs of the KNN and SVM (p = 0.0078) and the SVM and CDA (p = 0.0039) pattern recognition algorithms for *Alternaria* and the pairs of the KNN and k-means (p = 0.0054), KNN and SVM (p = 0.002), k-means and CDA (p = 0.0025), and SVM and CDA (p = 0.0313) pattern recognition algorithms for *Cladosporium*.

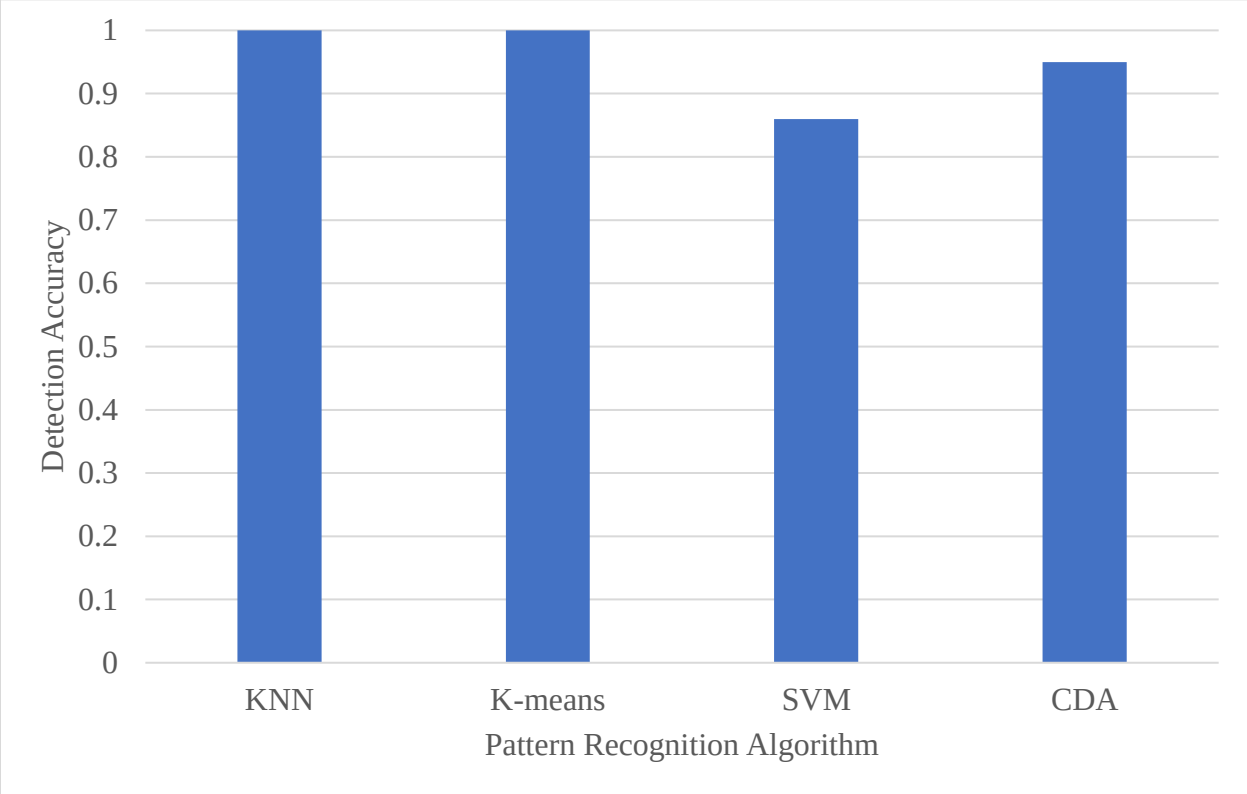

Figure 2. Bar chart displaying the detection accuracy of the KNN, K-means, SVM, and CDA pattern recognition algorithms.

Detection accuracy. The KNN and k-means pattern recognition algorithms detected fungal species in every exposure due to the nature of the calculations behind these pattern

recognition algorithms, so I excluded these pattern recognition algorithms from any statistical tests. The SVM pattern recognition algorithm detected fungal contaminants in 172 out of 200 exposures, and the CDA pattern recognition algorithm detected fungal contaminants in 190 out of 200 exposures**.** Both correctly detected 165 exposures and incorrectly detected three exposures. The SVM pattern recognition algorithm correctly classified an additional seven exposures, and the CDA pattern recognition algorithm correctly classified an additional twenty-five exposures. I performed a McNemar's Chi-Square test with previously justified assumptions to answer the question: "How does the detection of common indoor fungal contaminants vary for different electronic nose pattern recognition algorithms?" The obtained test static of 10.125 was greater than the critical value of 3.841 for a 0.05 significance level and one degree of freedom. I rejected the null hypothesis of no difference in the classification accuracy between pattern recognition algorithms in favor of the alternative hypothesis, which showed that there is a difference in the detection accuracy between the SVM and CDA pattern recognition algorithms.

Table 3

Joint Frequencies of Pattern Recognition Detections and McNemar's Test Statistics (SVM/CDA)

	Actual Exposure Class				
Measurement	Alternaria	Aspergillus	Cladosporium	Penicillium	All Classes
Detect/Detect	35	48	37	45	165
Detect/Fail	1	0	5	1	7
Fail/Detect	12	2	8	3	25
Fail/Fail	2	0	0	1	3
Significance Level	0.05	0.05	0.05	0.05	0.05
p-value	0.0034	0.5000	0.5811	0.6250	0.0014

Note. A value of 1 for Detect/Fail for Alternaria represents that the SVM pattern recognition algorithm correctly detected one exposure that the CDA pattern recognition algorithm failed to detect.

Alternaria, Aspergillus, Cladosporium, and Penicillium. The CDA pattern recognition algorithm had a higher detection accuracy than the SVM pattern recognition algorithm for all genera of fungal contaminants. Both detected fungal contaminants in thirty-five, forty-eight,

thirty-seven, and forty-five out of fifty exposures and failed to detect fungal contaminants in two, zero, zero, and one out of fifty exposures for *Alternaria*, *Aspergillus*, *Cladosporium*, and *Penicillium* genera, respectively. The SVM pattern recognition algorithm correctly detected an additional one, zero, five, and one out of fifty exposures, and the CDA pattern recognition algorithm correctly detected an additional twelve, two, eight, and three out of fifty exposures for *Alternaria*, *Aspergillus*, *Cladosporium*, and *Penicillium* genera, respectively.

Due to the decreased quantity of discordant joint frequencies, I used a binomial distribution to model McNemar's Chi-Square tests. Table 3 listed the results of each test, showing statistical significance for a difference in the detection accuracy between the SVM and CDA pattern recognition accuracies for only *Alternaria* ($p = 0.003418$). I rejected the null hypothesis of no difference in the detection accuracy in favor of the alternative hypothesis, which showed that there is a difference in the detection accuracy between the SVM and CDA pattern recognition algorithms for *Alternaria*. I failed to reject the null hypotheses of no difference in the detection accuracy between the SVM and CDA pattern recognition algorithms for *Aspergillus*, *Cladosporium*, and *Penicillium*.

Evaluation of Findings

The data of this study reflected the design and limitations of the methodology. I successfully carried out the planned data collection procedures using the obtained materials and instruments under controlled conditions. I performed Cochran's Q tests and McNemar's Chi-Square tests to analyze the data and draw conclusions regarding my hypotheses and questions. I found statistically significant differences between pattern recognition algorithms for both the classification and detection accuracy of common fungal contaminants.

There did not exist a statistically significant difference between the classification accuracies of any of the pairs of the KNN, SVM, and CDA pattern recognition algorithms as warranted by the literature. The classification accuracy of the KNN algorithm represented the results of Gobbi and Locci (2006) and Schiffman et al. (2000) as a capable pattern recognition algorithm, though with a lower classification accuracy due to drift variation. However, the classification accuracy of the KNN pattern recognition algorithm did contradict results from the studies of Cagnasso et al. (2010) and Tang et al. (2010), who concluded that the SVM and LDA pattern recognition algorithms were superior in classification accuracy in comparison to the KNN pattern recognition algorithm. In terms of classification accuracy, the SVM pattern recognition algorithm performed similarly to the CDA pattern recognition algorithm. This closely followed the results of the studies of Hong et al. (2015) and Zhang et al. (2017) of the SVM pattern recognition algorithm performing the same or only a few percentage points worse than the LDA pattern recognition algorithm. There was little previous literature comparing the k-means algorithm to other pattern recognition algorithms. Overall, there was a fair agreement between the data of the pattern recognition algorithms and previous studies.

The study faced the limitations previously accounted for in the methodology. The use of exposures on lab-grown molds decreased the construct validity of this study due to the limitation on fungal genera and environmental factors, but the use of commonly found fungal contaminants allowed the external validity of this study to remain high. I assumed the functionality of the Cyranose 320 and cross-validity of the training set, which held true for all the exposures and pattern recognition algorithms, preventing the effects of confounding variables and increasing internal validity of this study. The restriction of diversity in electronic noses and mold cultures did hamper the content validity of this study as I only sampled four common genera and used

only one electronic nose brand, but I still achieved higher external validity of this study due to similarities between various electronic nose designs and samples of mold. None of these limitations seemed to have significant effects on my significant conclusion drawn from the experimentation.

An unexpected limitation did play a noticeable role in potentially numerically decreasing the classification accuracy of all pattern recognition algorithms, though with few effects on the conclusions of this study. A drift in the principal components of exposures plotted on the PCA for each genus was visible in all the mold genera. The *Alternaria* exposures experienced a downward drift in the second principal component in later exposures (Appendix C, Figure 1). The *Aspergillus* exposures experienced a rightward drift in the first principal component (Appendix C, Figure 2). The *Cladosporium* and *Penicillium* exposures experienced a downward, followed by a rightward drift in the second, then first principal components (Appendix C, Figure 3; Appendix C, Figure 4). The training set exposures also experienced a drift for the *Cladosporium* and *Penicillium* classes in the rightward and downward directions for the first and second principal components, respectively (Appendix C, Figure 5).

The drift did not seem to have much of an impact on the KNN, k-means, and SVM pattern recognition algorithms due to their high cross-validities of 92-97%. The cross-validity of the CDA pattern recognition algorithm was much lower, but this likely resulted from the misclassifications between *Aspergillus* and *Alternaria* rather than the drift in *Cladosporium* and *Penicillium*. The drifts did increase variation within classes, which may have resulted in the first and last ten exposures of each genus having more misclassifications and misdetections than the middle thirty (Appendix A, Table 1). However, this trend was visible in all pattern recognition algorithms and therefore was not a confounding variable in the purpose of this study (Appendix

A, Table 1). The individual classification and detection accuracies may have had lower external validities, but that was not the focus of this study.

These results expanded on the pattern recognition aspect of exploratory analyses by Gobbi and Locci (2006), Kuske et al. (2006), Lagod et al. (2017), Pan et al. (2014), Paolesse et al. (2006), and Schiffman et al. (2000) on the electronic nose detection of various genera of fungal contamination grown on various media. The results specifically targeted the most common genera of fungal contaminants, as outlined by Shelton et al. (2000) and WHO (2009), that were responsible for the symptoms of SBS, which directly affected human health (Joshi, 2008). Furthermore, the results of this study were applicable to the use of the Cyranose 320, which has lacked research in its capability to detect and classify fungal contamination (Intelligent Optical Systems, 2000). Additionally, the results were applicable to pattern recognition algorithms widely available to many electronic noses (Wilson & Baietto, 2009).

Summary

The data collected from this study closely followed the outlined procedures and limitations. The Cochran's Q test produced a significant difference between the classification accuracies of pattern recognition algorithms, with post hoc McNemar's Chi-Square tests showing significant differences between the classification accuracies of the k-means pattern recognition algorithm with each of the KNN, SVM, and CDA pattern recognition algorithms. Sub-analyses showed further statistical significance between pattern recognition algorithms for the classification *Alternaria* and *Cladosporium* exposures. A McNemar's Chi-Square test for detection accuracy showed a significant difference between the detection accuracies of the SVM and CDA pattern recognition algorithms.

The data analyses contained limitations previously predicted, addressed, and justified before data collection. An additional limitation of exposure drift appeared during data collection but showed no significant effect on the conclusions of the statistical analyses. The conclusions followed the results of most of the previous studies on electronic nose fungal detection.

Chapter 5: Implications, Recommendations, and Conclusions

The symptoms of SBS that occupants experience in up to 30% of buildings correlate with mold, a form of fungal contamination (EPA, 1991; Khan & Karuppayil, 2012). Electronic noses provide a potential means of mold detection, delivering quick, cheap, and conclusive information while boasting numerous additional applications (Lagod et al., 2017). However, electronic have poor accuracy due to the lack of a common standard for sensors, data processing, and pattern recognition (Eusebio et al., 2016). The purpose of this quantitative, experimental study was to determine the pattern recognition algorithm with the highest accuracy for the detection of common fungal contaminants. Literature has varied on the effectiveness of the pattern recognition algorithms of this study, which prompted further background research, data collection, and analysis.

I used a Cyranose 320 electronic nose in the detection of lab-grown mold samples under a controlled environment. I applied different pattern recognition algorithms on each exposure to collect data for the research questions. I analyzed the classification and detection accuracies of each pattern recognition algorithm using Cochran's Q tests and McNemar's Chi-Square tests and obtained statistically significant differences between the accuracies of each pattern recognition algorithm. However, the limitations, including the simulation of fungal contamination and the drift in the data collection, have decreased the applicability of this study's results. This chapter further analyzes the data and results before drawing conclusions.

Implications

The results and analysis of this study aimed to address the proposed research questions and their resulting hypotheses. The findings showed fair support for the research hypotheses of "Different pattern recognition algorithms of an electronic nose will have different effects on an electronic nose's ability to classify common fungal contaminants" and "Different pattern recognition algorithms of an electronic nose will have different effects on an electronic nose's ability to detect common fungal contaminants."

Classification. The Cochran's Q-test found a statistically significant ($p = 0.0049$) difference between the classification accuracies of the KNN, k-means, SVM, and CDA pattern recognition algorithms for the research question "How does the classification of common indoor fungal contaminants vary for different electronic nose pattern recognition algorithms?" Due to this statistically significant outcome, I rejected the null hypothesis of "There is no difference in the classification of common indoor fungal contaminants for different electronic nose pattern recognition algorithms" in favor of the alternative hypothesis of "There is a difference in the classification of common indoor fungal contaminants for different electronic nose pattern recognition algorithms."

Based on the obtained data, the k-means pattern recognition algorithm performs worse than the KNN, SVM, and CDA pattern recognition algorithms in the classification of common fungal contaminants. McNemar's Chi-Square tests produced a statistically significant difference for the classification accuracies between three pairs of pattern recognition algorithms: KNN and k-means ($p = 0.0047$), SVM and k-means ($p = 0.0434$), and CDA and k-means ($p = 0.0210$). The k-means pattern recognition algorithm had a poor classification accuracy, especially for *Cladosporium* exposures for which it detected only nineteen out of fifty (Table 3). This was

representative of the weaknesses in the k-means algorithm discussed by Scott et al. (2007). The k-means pattern recognition algorithm likely performed a poor first partition that captured an outlying exposure, such as the fifth *Alternaria* training exposure, which resulted in an incorrect convergence of the following partitions used in the algorithm (Scott et al., 2007). This poor partition may have caused the k-means pattern recognition algorithm to misclassify the *Alternaria* and the *Cladosporium* exposures between the *Alternaria* training exposures, thus reducing the classification accuracy. The SVM pattern recognition algorithm utilizes hyper-dimensional confidence levels, so the outliers in the training set did not have a major effect on classifications (Appendix C, Figure 6; Zhang et al., 2017). The CDA pattern recognition algorithm similarly considered outliers when maximizing class variation (Appendix C, Figure 7; Intelligent Optical Systems, 2000). The KNN pattern recognition algorithm likely used a k greater than one to nullify the effect of the singular outliers when counting neighbors (Appendix C, Figure 5; Gobbi & Locci, 2006). These are likely causes for the k-means pattern recognition algorithm performing less accurately than the other pattern recognition algorithms.

Detection. The McNemar's Chi-Square test produced a statistically significant ($p = 0.0014$) difference for the classification accuracies between the SVM and CDA pattern recognition algorithms for the research question "How does the classification of common indoor fungal contaminants vary for different electronic nose pattern recognition algorithms?" Due to this statistically significant outcome, I rejected the null hypothesis of "There is no difference in the classification of common indoor fungal contaminants for different electronic nose pattern recognition algorithms" in favor of the alternative hypothesis of "There is a difference in the classification of common indoor fungal contaminants for different electronic nose pattern recognition algorithms."

Based on the obtained data, the SVM pattern recognition algorithm performs worse than the CDA pattern recognition algorithms in the detection of common fungal contaminants. The SVM and CDA pattern recognition algorithms both rely heavily on confidence levels of each training class when identifying exposures (Intelligent Optical Systems, 2000). The CDA pattern recognition algorithm maximizes the variation between classes while the SVM pattern recognition algorithm does not (Hong et al., 2015; Scott et al., 2007). Resultingly, the CDA pattern recognition algorithm detects and classifies a larger range of exposures on the PCA than the SVM pattern recognition algorithm (Appendix C, Figure 6; Appendix C, Figure 7). The KNN and k-means pattern recognition algorithms do not rely on confidence and classify every exposure based on rigid partitions of classes in the PCA graph (Scott et al., 2007).

Literature and applications. The classification accuracy of the KNN algorithm emulated the results from the studies of Gobbi and Locci (2006) and Schiffman et al. (2000) as a numerically superior pattern recognition algorithm, but the algorithms were all less accurate due to the drift in the exposures. The classification accuracies of the SVM and CDA pattern recognition contradicted the results from the studies of Cagnasso et al. (2010) and Tang et al. (2010), who alternatively concluded that the SVM and LDA pattern recognition algorithms had superior classification accuracies in comparison to the KNN pattern recognition algorithm. This could have attributed from Cagnasso et al. (2010) only warranting a binary classification, which was a strength of the SVM algorithm in comparison to the weaker multi-class classification of this study and the study of Tang et al. (2010) using the LDA algorithm in conjunction with the KNN pattern recognition algorithm. The SVM and CDA pattern recognition algorithms also had similar classification accuracies, which followed the results of the studies of Hong et al. (2015) and Zhang et al. (2017).

The limitations of this study restrict the application of results. The use of the Cyranose 320 inhibits the results to electronic noses with carbon-black polymer sensors as other sensors provide different forms of measurement that are not generalizable from the R/R_0 measurements of the Cyranose 320 sensors (Scott et al., 2007). Regulations on potentially hazardous biological agents prevented the creation of exposures on a representative sample of all mold genera. Most importantly, an unpredicted limitation of a drift in exposures affected all pattern recognition algorithms. Though having no effect on my conclusions between pattern recognition algorithms, the drift does impede applications of the numerical classification and detection accuracies of individual pattern recognition algorithms.

There were several possible causes of this drift. Intelligent Optical Systems (2000) noted that the overuse of the Cyranose 320 would have resulted in the accumulation of chemical species, though this was unlikely as we had purchased the Cyranose 320 only three months prior to experimentation with minimal prior usage. Another possible cause deduced from the results of the study by Eusebio et al. (2016) on humidity. The accumulation of water vapor from cellular respiration over time may have increased the sensors' abilities to detect chemical species, thus shifting the later exposures, though the rate of humidity increase was questionable, which leaves this as an unlikely cause of the shift. A final cause deduced from the results of the study by Lagod et al. (2017) on MVOC concentration. The sensors' inability to detect MVOCs below certain thresholds may have caused a shift in the exposures if the MVOC concentrations increased significantly during the experimentation period, which was a significant fifteen minutes following the initial thirty-minute buildup of MVOCs in the airtight containers, which leaves this as a potential cause of the drift in exposures (Lagod et al., 2017).

The limitation of a convenience sample of mold exposures prompted the assumption of a representative sample of the population. This assumption did not strongly affect the results as WHO (2009) and Shelton et al. (2000) determined that *Alternaria*, *Aspergillus*, *Cladosporium*, and *Penicillium* were the four most common fungal genera responsible for SBS. However, there remains caution in the application of these results for the potential of other confounding variables. Further delimitations on the pattern recognition algorithms themselves also limit the use of the results. The data only applies to the KNN, k-means, SVM, and CDA pattern recognition algorithms and cannot apply in the comparison of other forms of pattern recognition. Furthermore, the data only pertain to the classification of an exposure rather than the confidence of the classification, which is a feature of the Cyranose 320 and other electronic noses (Intelligent Optical Systems, 2000; Wilson & Baietto, 2009)

Despite the limitations, assumptions, and delimitations of the results of this study, the results still greatly contribute to the utilization of electronic noses in the detection of fungal contamination. The KNN, SVM, and CDA pattern recognition algorithms can identify and classify fungal VOCs with fair accuracy, and the k-means pattern recognition algorithm can similarly perform these functions, though with caution for poor initial partitions. These findings on pattern recognition algorithms can assist in minimizing inaccuracy, which was a significant weakness faced by electronic noses (Schiffman et al., 2000). The elimination of this weakness may allow electronic noses to become a cheaper, quicker, and more mobile means of mold detection than the traditional methods of polymerase chain reaction (PCR), gas chromatography-mass spectrometry (GC-MS), and mycological analysis (Paolesse et al., 2006). The use of electronic noses may prevent numerous cases of SBS caused by the growth of fungal contaminants, including those of *Alternaria, Aspergillus, Chaetomium, Cladosporium, Fusarium,*

Penicillium, Stachybotrys, and *Trichoderma* genera (Shelton et al., 2002). Recommendations for further research in pattern recognition algorithms in the detection of common fungal contaminants can hasten the progress of this application of machine olfaction.

Recommendations

Several aspects of this study leave room for additional research. There was a reduction in the external validity of this study due to the decreased applicability of the results, which stemmed from the design. The simulation of fungal contamination was insufficient in representing contamination in actual buildings. Given more experience and resources, future researchers should use fungal contamination on building surfaces or construction materials, including the design of the study by Lagod et al. (2017) under a controlled experiment or the experimental design of the study by Kuske et al. (2006) on a larger scale. A further resource constraint was the use of electronic noses. With sufficient funds, researchers should consider other electronic nose brands, such as the Sacmi EOS series used studies by Cagnasso et al. (2010) and Falasconi et al. (2012). This can further expand into different types of sensors, including the metal oxide sensors of the FOX 2000 electronic nose (Wilson & Baietto, 2009). These recommendations can increase the applicability and external validity of the findings.

The delimitations that I imposed on this study were resource and convenience based, and I justified their effects on the scope of this study. The CDAnalysis program provides three additional pattern recognition algorithms as treatments: soft independent modeling of class analogy (SIMCA), hierarchal cluster analysis (HCA), and single class (Intelligent Optical Systems, 2000). There is little prior research on those pattern recognition algorithms for electronic noses, so I recommend that future researchers should explore into the accuracies of these treatments. I also recommend further research into variations of these pattern recognition

algorithms for mold detection, including fuzzy-KNN or different values of k as seen in studies by Kuske et al. (2006) and Tang et al. (2010), respectively. Researchers with stronger computer science backgrounds may also explore the use of artificial neural networks, including multilayer perceptron, probabilistic neural network, and radial basis function neural network, for machine learning (Scott et al., 2007). Additional pattern recognition techniques can strengthen or weaken these findings.

I also recommend research on different samples of exposures. The results of the study suggest that the pattern recognition algorithms would have had higher accuracies if I had excluded the *Alternaria* or *Cladosporium* exposures or introduced exposures of new genera. Research regulations prevented me from using some of the rarer and more dangerous genera, including *Acremonium*, *Fusarium*, and *Stachybotrys* (black mold), which are the causes of the most significant symptoms of SBS (Shelton et al., 2002; WHO, 2009). The length, temperature, and filtering of each exposure are also alterable in the Cyranose 320 methods (Intelligent Optical Systems, 2000). Researchers can also increase the sample size to thousands of exposures with automation of exposure creation to further validate the results of this study and give more power to the statistical tests and conclusions. Another option proposes an increase in the number of exposures in the training set to increase the precision of the pattern recognition algorithms (Intelligent Optical Systems, 2000)

A final recommendation for this study addresses the drift in exposures and its effect on the internal validity of the findings. The apparent drift in the exposures may have affected my data by reducing the classification and detection accuracies of the pattern recognition algorithms. Future research should look towards increasing the concentration of MVOCs by extending the period for VOC buildup, decreasing the container size, or increasing the quantity of mold

through larger or multiple potato dextrose agar plates to decrease the effect of drift as a confounding variable.

Conclusions

This study showed a statistically significant difference in the classification and detection accuracies between different pattern recognition algorithms. The KNN, SVM, and CDA pattern recognition algorithms performed as predicted in previous literature. The k-means pattern recognition algorithm faced classification limitations due to a poor first partition, and the SVM pattern recognition algorithm faced detection limitations due to its lower confidence in classifying exposures. Limitations in the simulation of fungal contamination and the drift of exposures decreased the applicability of these results, but they still lead towards valid conclusions. The conclusion of the KNN, SVM, and CDA pattern recognition algorithms over the k-means pattern recognition algorithm contributes to a general increase in the accuracy of electronic noses, which allows for realistic applications in mold detection (Lagod et al., 2017).

Further research into pattern recognition algorithms in the detection of fungal contaminants can hasten progress in machine olfaction. The authentic simulation of fungal contamination and utilization of multiple electronic noses can increase the external validity of research. Improvements to the sample, including more potato dextrose agar plates, fungal contaminant species, number of exposures, and exposure settings, can the improve validity of this study as well. Future studies could also improve internal validity through increasing VOC accumulation to reduce drift and expand the scope of the research through adding additional pattern recognition algorithms as treatments. Overall, the advancing research on electronic noses for fungal detection has the potential to provide immense health and safety benefits to a considerable proportion of the worldwide population.

References

Acumedia. (2017, March 29). *Potato dextrose agar, product information*. Retrieved October 13, 2017 from http://foodsafety.neogen.com/pdf/acumedia_pi/7149_pi.pdf

Baracho, M. S., & Baracho, I. R. (2003). An analysis of the spontaneous mutation rate measurement in filamentous fungi. *Genetics and Molecular Biology*, *26*(1), 83-87. doi:10.1590/S1415-47572003000100014

Cagnasso, S., Falasconi, M., Previdi, M. P., Franceschini, B., Cavalieri, C., Sberveglieri, V., & Rovere, P. (2010). Rapid screening of Alicyclobacillus acidoterrestris spoilage of fruit juices by electronic nose: A confirmation study. *Journal of Sensors, 2010*(11), 1-9. doi:10.1155/2010/143173

Environmental Protection Agency (1991). *Indoor air facts no.4 (revised) sick building syndrome. Retrieved fr*om https://www.epa.gov/indoor-air-quality-iaq/indoor-air-facts-no-4-sick-building-syndrome

Eusebio, L., Capelli, L., & Sironi, S. (2016). Electronic nose testing procedure for the definition of minimum performance requirements for environmental odor monitoring. *Sensors, 16*(1), 1548-1564. doi:10.3390/s16091548

Falasconi, M., Concina, I., Gobbi, E., Sberveglieri, V., Pulvirenti, A., & Sberveglieri, G. (2012). Electronic nose for microbiological quality control of food products, *International Journal of Electrochemistry, 2012*(12), 1-13. doi:10.1155/2012/715763

Gardner, J. W. & Bartlett, P. N. (1994). A brief history of electronic noses. *Sensors and Actuators B, 18*(1), 211-220. Retrieved from https://www.researchgate.net/publication/223387237

Gobbi, E., & Locci, R. (2006). An electronic nose to evaluate fungal contamination. *Gjomal Furlan des Siencis, 7*(1), 71-82. Retrieved from http://www.siencis-par-furlan.net/wp-content/uploads/Gobbi.Nose_.pdf

Hong, X., Wang, J., & Qi, G. (2015). Comparison of semi-supervised and supervised approaches for classification of e-nose datasets: Case studies of tomato juices. Chemometrics and Intelligent Laboratory Systems, *2015*(146), 457-463. doi:10.1016/j.chemolab.2015.07.001

Huang, T., Jia, P., He, P., Duan, S., Yan, J., & Wang, L. (2016). A novel semi-supervised method of electronic nose for indoor pollution detection trained by m-s4vms. *Sensors, 16*(1), 1462-1479. doi:10.3390/s16091462

Intelligent Optical Systems. (2000). *The Cyranose 320 eNose: User manual* (6th ed.). Baldwin Park, CA.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys, 31*(3). doi:10.1.1.18.2720

Joshi, S. M. (2008). The sick building syndrome. *Indian Journal of Occupational and Environmental Medicine, 12*(2), 61-64. doi:10.4103/0019-5278.43262

Khan, H. A., & Karappuayil, S. M. (2012). Fungal pollution of indoor environments and its management. *Saudi Journal of Biological Sciences, 2012*(19), 405-426. doi:10.1016/j.sjbs.2012.06.002

Kuhn, D. M., & Ghannoum, M. A. (2003). Indoor mold, toxigenic fungi, and Stachybotrys chartarum: infectious disease perspective. *Clinical Microbiology Reviews, 16*(1), 144-172. doi:10.1128/CMR.16.1.144-172.2003

Kuske, M., Padilla, M., Romain, A.C., Nicolas, J., Rubio, R., & Marco, S. (2006). Detection of diverse mould species growing on building materials by gas sensor arrays and pattern recognition. *Sensors and Actuators B, 2006*(119), 33-40. doi:10.1016/j.snb.2005.02.059

Laerd Statistics. (n.d.). Cochran's Q test using SPSS Statistics. *Lund Research.* Retrieved October 16, 2017, from https://statistics.laerd.com/spss-tutorials/cochrans-q-test-in-spss-statistics.php

Lagod, G., Suchorab, Z., Guz, L., & Sobczuk, H. (2017). Classification of buildings mold threat using electronic nose. *American Institute of Physics*, *1866*(1), 1-5. doi:10.1063/1.4994478

Marco, S. (2014). The need for external validation in machine olfaction: Emphasis on health-related applications. *Analytical and Bioanalytical Chemistry*, *406*(4), 3941-3956. doi:10.1007/s00216-014-7807-7

Mui, K. W., Chan, W. Y., Wong, L. T., & Hui, P. S. (2010). Scoping indoor airborne fungi in an excellent indoor air quality office building in Hong Kong. *Building Services Engineering Research and Technology, 31*(2), 191-199. doi:10.1177/0143624409359752

Mumyakmaz, B., & Karabacak, K. (2015). An e-nose-based indoor air quality monitoring system: Prediction of combustible and toxic gas concentrations. *Turkish Journal of Electrical Engineering & Computer Sciences, 2015*(23), 729-740. doi:10.3906/elk-1304-210

Navi, S. S., Bandyopadhyay, R., Hall, A. J., Bramel-Cox, P. J. (1999). A pictorial guide for the identification of mold fungi on sorghum grain. *Natural Resources Institute*. Retrieved from http://oar.icrisat.org/1948/1/A_Pictorial_Guide_for_the_Identification.pdf

Pan, L., Zhang, W., Zhu, N., Mao, S., & Tu, K. (2014). Early detection and classification of pathogenic fungal disease in post-harvest strawberry fruit by electronic nose and gas chromatography-mass spectrometry. *Food Research International, 62*(8), 162-168. doi:10.1016/j.foodres.2014.02.020

Paolesse, R., Alimelli, A., Martinelli, E., Natale, D., Amico, A. D., D'Egidio, M. G., . . ., Fanelli, C. (2006). Detection of fungal contamination of cereal grain samples by an electronic nose. *Sensors and Actuators B, 2006*(119), 425-430. doi:10.1016/j.snb.2005.12.047

Pardo, M., & Sberveglieri, G. (2005). Classification of electronic nose data with support vector machines. *Sensors and Actuators B, 107*(6), 730-737. doi:10.1016/j.snb.2004.12.005

Rahman, M., Charoenlarpnopparut, C., & Suksomgpong, P. (2016). Classification and pattern recognition algorithms applied to e-nose. International Conference on Electrical Information and Communication Technology, Khunla, 2015. Piscataway, NJ: IEEE.

Schiffman, S. S., Wyrick, D. W., Gutierrez-Osuna, R., & Nagle, H. T. (2000). Effectiveness of an electronic nose for monitoring bacterial and fungal growth. *International Symposium on Olfaction and Electronic Nose*, 173-180. Retrieved from https://www.researchgate.net/publication/253885081

Scott, S. M., James, D., & Ali, Z. (2007). Data analysis for electronic nose systems. *Microchimica Acta, 2007*(156), 183-207. doi:10.1007/s00604-006-0623-9

Shelton, B. G., Kirkland, K. H., Flanders, W. D., & Morris, G. K. (2002). Profiles of airborne fungi in buildings and outdoor environments in the United States. *Applied and Environmental Microbiology 2002, 68*(4), 1743–1753. doi:10.1128/AEM.68.4.1753.2002

Simon, M. K. (2011). *Dissertation and scholarly research: Recipes for success* (2011 Ed.). Seattle, WA: Dissertation Success, LLC.

Tang, K., Lin, Y., & Shyu, J. (2010). A local weighted nearest neighbor algorithm and a weighted and constrained least-squared method for mixed odor analysis by electronic nose systems. *Sensors, 10*(11), 10467-10483. doi:10.3390/s01110467

Wilson, A. D., & Baietto, M. (2009). Applications and advances in electronic-nose technologies. *Sensors, 9*(7), 5099-5148. doi:10.3390/s90705099

World Health Organization. (2009). WHO guidelines for indoor air quality: Dampness and mold. *World Health Organization*. Retrieved from http://www.euro.who.int/__data/assets/pdf_file/0017/43325/E92645.pdf

Zhang, L., Liu, Y., & Deng, P. (2017). Odor recognition in multiple e-nose systems with cross-domain discriminative subspace learning. *IEEE Transactions on Instrumentation and Measurement, 66*(7), 1679-1692. doi:10.1109/TIM.2017.2669818

Appendix A: Raw Data and Summary Data

Table 1

Classifications of Pattern Recognition Algorithms by Exposure and Class

	Pattern Recognition Algorithm			
Alternaria	KNN	K-means	SVM	CDA
1	Penicillium	Penicillium	Penicillium	Penicillium
2	Penicillium	Penicillium	Undetectable	Undetectable
3	Penicillium	Penicillium	Undetectable	Undetectable
4	Penicillium	Penicillium	Undetectable	Penicillium
5	Cladosporium	Penicillium	Cladosporium	Cladosporium
6	Alternaria	Alternaria	Alternaria	Cladosporium
7	Cladosporium	Alternaria	Cladosporium	Cladosporium
8	Cladosporium	Alternaria	Undetectable	Cladosporium
9	Cladosporium	Alternaria	Undetectable	Cladosporium
10	Cladosporium	Alternaria	Undetectable	Cladosporium
11	Penicillium	Penicillium	Undetectable	Penicillium
12	Alternaria	Alternaria	Alternaria	Alternaria
13	Alternaria	Alternaria	Alternaria	Alternaria
14	Alternaria	Alternaria	Alternaria	Alternaria
15	Alternaria	Alternaria	Alternaria	Alternaria
16	Alternaria	Alternaria	Alternaria	Alternaria
17	Cladosporium	Alternaria	Alternaria	Alternaria
18	Alternaria	Alternaria	Alternaria	Alternaria
19	Penicillium	Penicillium	Undetectable	Penicillium
20	Cladosporium	Aspergillus	Undetectable	Cladosporium
21	Cladosporium	Alternaria	Undetectable	Cladosporium
22	Alternaria	Alternaria	Alternaria	Alternaria
23	Cladosporium	Alternaria	Undetectable	Cladosporium
24	Alternaria	Alternaria	Alternaria	Alternaria
25	Cladosporium	Cladosporium	Cladosporium	Cladosporium
26	Alternaria	Alternaria	Alternaria	Alternaria
27	Penicillium	Penicillium	Undetectable	Penicillium
28	Alternaria	Cladosporium	Alternaria	Alternaria
29	Alternaria	Cladosporium	Alternaria	Alternaria
30	Alternaria	Alternaria	Alternaria	Alternaria
31	Aspergillus	Aspergillus	Alternaria	Aspergillus
32	Alternaria	Alternaria	Alternaria	Alternaria
33	Aspergillus	Aspergillus	Alternaria	Aspergillus
34	Alternaria	Alternaria	Alternaria	Alternaria
35	Penicillium	Penicillium	Undetectable	Penicillium
36	Aspergillus	Alternaria	Alternaria	Alternaria
37	Alternaria	Aspergillus	Alternaria	Aspergillus
38	Alternaria	Aspergillus	Alternaria	Alternaria
39	Alternaria	Alternaria	Alternaria	Alternaria

40	Alternaria	Aspergillus	Alternaria	Alternaria
41	Aspergillus	Aspergillus	Alternaria	Alternaria
42	Alternaria	Aspergillus	Alternaria	Aspergillus
43	Cladosporium	Penicillium	Undetectable	Penicillium
44	Aspergillus	Aspergillus	Alternaria	Aspergillus
45	Alternaria	Alternaria	Alternaria	Alternaria
46	Alternaria	Aspergillus	Alternaria	Alternaria
47	Alternaria	Aspergillus	Alternaria	Alternaria
48	Aspergillus	Aspergillus	Alternaria	Aspergillus
49	Aspergillus	Aspergillus	Alternaria	Aspergillus
50	Alternaria	Alternaria	Alternaria	Alternaria

Aspergillus	KNN	K-means	SVM	CDA
1	Penicillium	Penicillium	Undetectable	Penicillium
2	Alternaria	Aspergillus	Aspergillus	Alternaria
3	Aspergillus	Aspergillus	Aspergillus	Aspergillus
4	Aspergillus	Aspergillus	Aspergillus	Aspergillus
5	Aspergillus	Aspergillus	Aspergillus	Aspergillus
6	Aspergillus	Aspergillus	Aspergillus	Aspergillus
7	Aspergillus	Aspergillus	Aspergillus	Aspergillus
8	Aspergillus	Aspergillus	Aspergillus	Aspergillus
9	Aspergillus	Aspergillus	Aspergillus	Aspergillus
10	Aspergillus	Aspergillus	Aspergillus	Aspergillus
11	Aspergillus	Aspergillus	Aspergillus	Aspergillus
12	Aspergillus	Aspergillus	Aspergillus	Aspergillus
13	Aspergillus	Aspergillus	Aspergillus	Aspergillus
14	Aspergillus	Aspergillus	Aspergillus	Aspergillus
15	Aspergillus	Aspergillus	Aspergillus	Aspergillus
16	Aspergillus	Aspergillus	Aspergillus	Aspergillus
17	Aspergillus	Aspergillus	Aspergillus	Aspergillus
18	Aspergillus	Aspergillus	Aspergillus	Aspergillus
19	Aspergillus	Aspergillus	Aspergillus	Aspergillus
20	Aspergillus	Aspergillus	Aspergillus	Aspergillus
21	Aspergillus	Aspergillus	Aspergillus	Aspergillus
22	Aspergillus	Aspergillus	Aspergillus	Aspergillus
23	Aspergillus	Aspergillus	Aspergillus	Aspergillus
24	Aspergillus	Aspergillus	Aspergillus	Aspergillus
25	Aspergillus	Aspergillus	Aspergillus	Aspergillus
26	Aspergillus	Aspergillus	Aspergillus	Aspergillus
27	Aspergillus	Aspergillus	Aspergillus	Aspergillus
28	Aspergillus	Aspergillus	Aspergillus	Aspergillus
29	Aspergillus	Aspergillus	Aspergillus	Aspergillus
30	Aspergillus	Aspergillus	Aspergillus	Aspergillus
31	Aspergillus	Aspergillus	Aspergillus	Aspergillus
32	Aspergillus	Aspergillus	Aspergillus	Aspergillus
33	Aspergillus	Aspergillus	Aspergillus	Aspergillus
34	Aspergillus	Aspergillus	Aspergillus	Aspergillus

35	Aspergillus	Aspergillus	Aspergillus	Aspergillus
36	Aspergillus	Aspergillus	Aspergillus	Aspergillus
37	Aspergillus	Aspergillus	Aspergillus	Aspergillus
38	Aspergillus	Aspergillus	Aspergillus	Aspergillus
39	Aspergillus	Aspergillus	Aspergillus	Aspergillus
40	Aspergillus	Aspergillus	Aspergillus	Aspergillus
41	Aspergillus	Aspergillus	Aspergillus	Aspergillus
42	Aspergillus	Aspergillus	Aspergillus	Aspergillus
43	Aspergillus	Aspergillus	Undetectable	Aspergillus
44	Aspergillus	Aspergillus	Aspergillus	Aspergillus
45	Aspergillus	Aspergillus	Aspergillus	Aspergillus
46	Aspergillus	Aspergillus	Aspergillus	Aspergillus
47	Aspergillus	Aspergillus	Aspergillus	Aspergillus
48	Aspergillus	Aspergillus	Aspergillus	Aspergillus
49	Aspergillus	Aspergillus	Aspergillus	Aspergillus
50	Aspergillus	Aspergillus	Aspergillus	Aspergillus

Cladosporium	KNN	K-means	SVM	CDA
1	Penicillium	Penicillium	Alternaria	Undetectable
2	Penicillium	Penicillium	Penicillium	Penicillium
3	Cladosporium	Penicillium	Undetectable	Cladosporium
4	Cladosporium	Penicillium	Cladosporium	Cladosporium
5	Cladosporium	Penicillium	Cladosporium	Cladosporium
6	Cladosporium	Penicillium	Cladosporium	Cladosporium
7	Cladosporium	Alternaria	Cladosporium	Cladosporium
8	Cladosporium	Penicillium	Cladosporium	Cladosporium
9	Cladosporium	Penicillium	Cladosporium	Cladosporium
10	Cladosporium	Alternaria	Cladosporium	Cladosporium
11	Cladosporium	Alternaria	Cladosporium	Cladosporium
12	Cladosporium	Penicillium	Undetectable	Cladosporium
13	Cladosporium	Penicillium	Cladosporium	Cladosporium
14	Cladosporium	Penicillium	Undetectable	Cladosporium
15	Cladosporium	Penicillium	Cladosporium	Cladosporium
16	Cladosporium	Penicillium	Cladosporium	Cladosporium
17	Cladosporium	Cladosporium	Cladosporium	Cladosporium
18	Cladosporium	Penicillium	Cladosporium	Cladosporium
19	Cladosporium	Cladosporium	Cladosporium	Cladosporium
20	Cladosporium	Alternaria	Undetectable	Cladosporium
21	Cladosporium	Penicillium	Cladosporium	Cladosporium
22	Cladosporium	Alternaria	Cladosporium	Cladosporium
23	Cladosporium	Cladosporium	Cladosporium	Cladosporium
24	Cladosporium	Cladosporium	Cladosporium	Cladosporium
25	Cladosporium	Cladosporium	Cladosporium	Cladosporium
26	Cladosporium	Cladosporium	Cladosporium	Cladosporium
27	Penicillium	Cladosporium	Alternaria	Undetectable
28	Alternaria	Cladosporium	Alternaria	Alternaria
29	Cladosporium	Aspergillus	Cladosporium	Cladosporium

30	Alternaria	Cladosporium	Alternaria	Alternaria
31	Cladosporium	Aspergillus	Cladosporium	Cladosporium
32	Cladosporium	Alternaria	Undetectable	Alternaria
33	Cladosporium	Cladosporium	Cladosporium	Cladosporium
34	Penicillium	Cladosporium	Penicillium	Cladosporium
35	Cladosporium	Cladosporium	Cladosporium	Cladosporium
36	Alternaria	Cladosporium	Alternaria	Alternaria
37	Alternaria	Alternaria	Alternaria	Undetectable
38	Alternaria	Cladosporium	Alternaria	Alternaria
39	Cladosporium	Cladosporium	Cladosporium	Cladosporium
40	Cladosporium	Cladosporium	Alternaria	Cladosporium
41	Alternaria	Alternaria	Alternaria	Cladosporium
42	Cladosporium	Cladosporium	Cladosporium	Cladosporium
43	Cladosporium	Penicillium	Cladosporium	Cladosporium
44	Aspergillus	Alternaria	Alternaria	Undetectable
45	Cladosporium	Alternaria	Undetectable	Alternaria
46	Cladosporium	Alternaria	Alternaria	Alternaria
47	Cladosporium	Alternaria	Undetectable	Alternaria
48	Cladosporium	Cladosporium	Cladosporium	Cladosporium
49	Cladosporium	Alternaria	Undetectable	Alternaria
50	Cladosporium	Cladosporium	Cladosporium	Cladosporium

Penicillium	KNN	K-means	SVM	CDA
1	Penicillium	Penicillium	Undetectable	Undetectable
2	Penicillium	Penicillium	Penicillium	Undetectable
3	Penicillium	Penicillium	Penicillium	Penicillium
4	Penicillium	Penicillium	Penicillium	Penicillium
5	Penicillium	Penicillium	Penicillium	Penicillium
6	Penicillium	Penicillium	Penicillium	Penicillium
7	Penicillium	Penicillium	Penicillium	Penicillium
8	Penicillium	Penicillium	Penicillium	Penicillium
9	Penicillium	Penicillium	Penicillium	Penicillium
10	Penicillium	Penicillium	Penicillium	Penicillium
11	Penicillium	Penicillium	Penicillium	Penicillium
12	Penicillium	Penicillium	Penicillium	Penicillium
13	Penicillium	Penicillium	Penicillium	Penicillium
14	Penicillium	Penicillium	Penicillium	Penicillium
15	Penicillium	Penicillium	Penicillium	Penicillium
16	Penicillium	Penicillium	Penicillium	Penicillium
17	Penicillium	Penicillium	Penicillium	Penicillium
18	Penicillium	Penicillium	Penicillium	Penicillium
19	Penicillium	Penicillium	Penicillium	Penicillium
20	Penicillium	Penicillium	Penicillium	Penicillium
21	Penicillium	Penicillium	Penicillium	Penicillium
22	Penicillium	Penicillium	Penicillium	Penicillium
23	Penicillium	Penicillium	Penicillium	Penicillium
24	Penicillium	Penicillium	Penicillium	Penicillium

25	Penicillium	Penicillium	Penicillium	Penicillium
26	Penicillium	Penicillium	Penicillium	Penicillium
27	Penicillium	Penicillium	Penicillium	Penicillium
28	Penicillium	Penicillium	Penicillium	Penicillium
29	Penicillium	Penicillium	Penicillium	Penicillium
30	Penicillium	Penicillium	Penicillium	Penicillium
31	Penicillium	Penicillium	Penicillium	Penicillium
32	Penicillium	Penicillium	Penicillium	Penicillium
33	Penicillium	Penicillium	Penicillium	Penicillium
34	Alternaria	Cladosporium	Alternaria	Alternaria
35	Penicillium	Penicillium	Penicillium	Penicillium
36	Penicillium	Penicillium	Penicillium	Penicillium
37	Penicillium	Penicillium	Penicillium	Penicillium
38	Alternaria	Cladosporium	Alternaria	Alternaria
39	Penicillium	Penicillium	Undetectable	Penicillium
40	Cladosporium	Cladosporium	Cladosporium	Penicillium
41	Cladosporium	Cladosporium	Cladosporium	Penicillium
42	Alternaria	Cladosporium	Alternaria	Alternaria
43	Penicillium	Penicillium	Penicillium	Penicillium
44	Cladosporium	Cladosporium	Cladosporium	Penicillium
45	Penicillium	Penicillium	Penicillium	Penicillium
46	Penicillium	Penicillium	Penicillium	Penicillium
47	Penicillium	Penicillium	Penicillium	Penicillium
48	Penicillium	Penicillium	Penicillium	Penicillium
49	Penicillium	Penicillium	Undetectable	Penicillium
50	Penicillium	Cladosporium	Undetectable	Penicillium
Correct Classifications	155	135	149	151
Correct Detections	200	200	172	190
Cross-Validity	92.5%	92.5%	96.9%	65.0%

Note. KNN and k-means do not produce undetectable classifications.

Table 2

Marginal Relative Frequencies of Pattern Recognition Algorithms by Exposure Genera

	Actual Exposure Class			
Classified As	Alternaria	Aspergillus	Cladosporium	Penicillium
KNN				
Alternaria	0.48	0.02	0.12	0.06
Aspergillus	0.14	0.96	0.02	0.00
Cladosporium	0.22	0.00	0.78	0.06
Penicillium	0.16	0.02	0.08	0.88
Undetectable	0.00	0.00	0.00	0.00
Total	1.00	1.00	1.00	1.00
K-means				
Alternaria	0.48	0.00	0.26	0.00
Aspergillus	0.26	0.98	0.04	0.00
Cladosporium	0.06	0.00	0.38	0.14
Penicillium	0.20	0.02	0.32	0.86
Undetectable	0.00	0.00	0.00	0.00
Total	1.00	1.00	1.00	1.00
SVM				
Alternaria	0.64	0.00	0.22	0.09
Aspergillus	0.00	0.96	0.00	0.00
Cladosporium	0.06	0.00	0.58	0.06
Penicillium	0.02	0.00	0.04	0.80
Undetectable	0.28	0.04	0.16	0.08
Total	1.00	1.00	1.00	1.00
CDA				
Alternaria	0.46	0.02	0.18	0.06
Aspergillus	0.14	0.96	0.02	0.00
Cladosporium	0.20	0.00	0.70	0.00
Penicillium	0.14	0.02	0.10	0.90
Undetectable	0.06	0.00	0.00	0.04
Total	1.00	1.00	1.00	1.00

Note. KNN and k-means do not produce undetectable classifications.

Appendix B: Statistical Analyses

Table 1

Joint Frequencies of Pattern Recognition Classifications and McNemar's Test Statistics (Altern.)

	Combinations of Pattern Recognition Algorithms					
Measurement	KNN/K-M	KNN/SVM	KNN/CDA	K-M/SVM	K-M/CDA	SVM/CDA
Correct/Correct	16	24	20	18	16	23
Correct/Incorrect	8	0	4	6	8	9
Incorrect/Correct	8	8	3	14	7	0
Incorrect/Incorrect	18	18	23	12	19	18
Significance Level	0.05	0.05	0.05	0.05	0.05	0.05
p-value	1.0000	0.0078	1.0000	0.1153	1.0000	0.0039

Note. A value of 16 for Correct/Incorrect and KNN/K-M represents that the KNN pattern recognition algorithm correctly classified sixteen *Alternaria* exposures that the k-means pattern recognition algorithm failed to classify.

Table 2

Joint Frequencies of Pattern Recognition Classifications and McNemar's Test Statistics (Clad.)

	Combinations of Pattern Recognition Algorithms					
Measurement	KNN/K-M	KNN/SVM	KNN/CDA	K-M/SVM	K-M/CDA	SVM/CDA
Correct/Correct	13	29	34	12	14	29
Correct/Incorrect	26	10	5	7	5	0
Incorrect/Correct	6	0	1	17	21	6
Incorrect/Incorrect	5	11	10	14	10	15
Significance Level	0.05	0.05	0.05	0.05	0.05	0.05
p-value	0.0054	0.0020	0.2188	0.0639	0.0025	0.0313

Note. A value of 13 for Correct/Incorrect and KNN/K-M represents that the KNN pattern recognition algorithm correctly classified thirteen *Cladosporium* exposures that the k-means pattern recognition algorithm failed to classify.

Appendix C: Principal Component Analyses

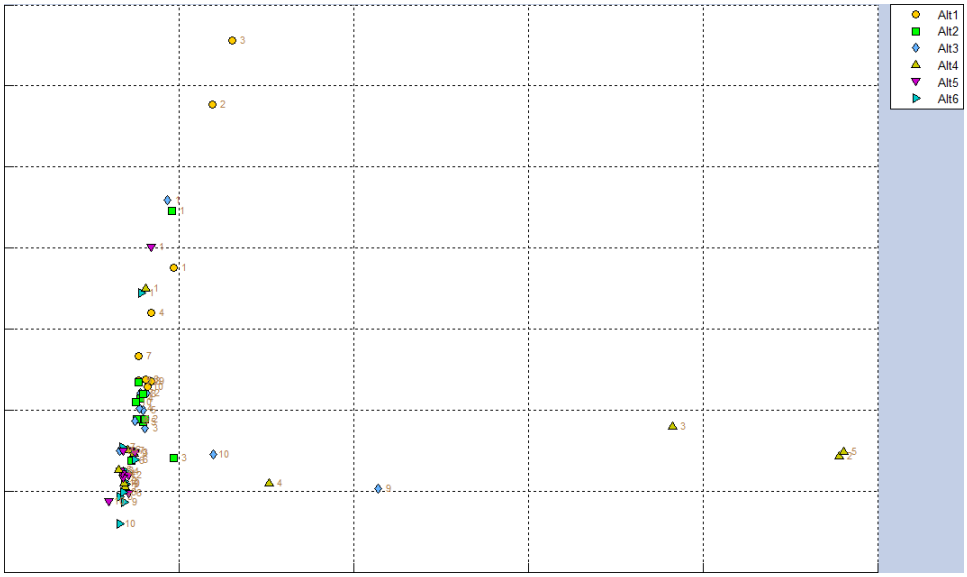

Figure 1. Principal component analysis (PCA) graph of the sixty unnormalized *Alternaria* exposures based on the two principal components with the highest variance. The exposures begin with the numbered yellow circles for "Alt1" as the first ten exposures and end with the numbered blue triangles for "Alt6" as the last ten exposures.

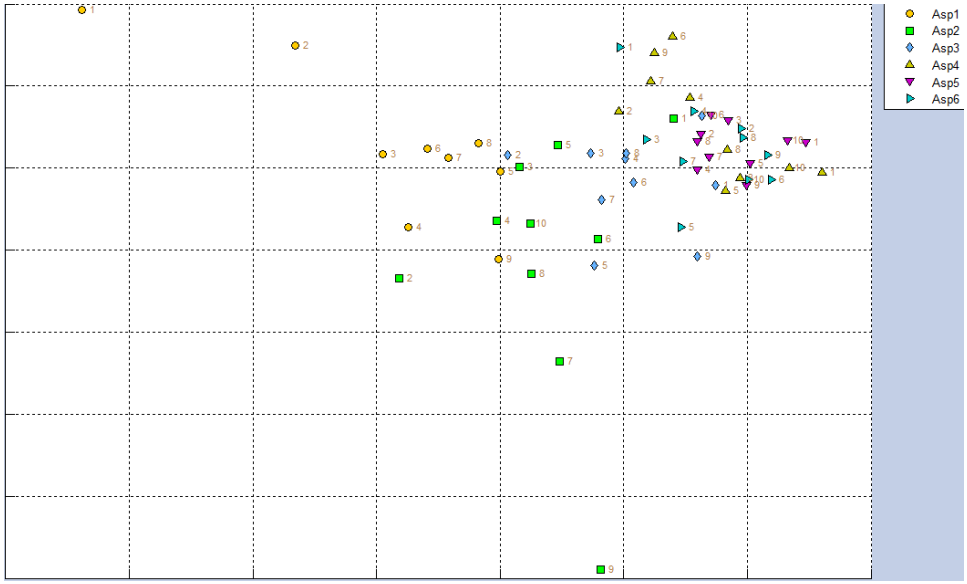

Figure 2. Principal component analysis (PCA) graph of the sixty unnormalized *Aspergillus* exposures based on the two principal components with the highest variance. The exposures begin with the numbered yellow circles for "Asp1" as the first ten exposures and end with the numbered blue triangles for "Asp6" as the last ten exposures.

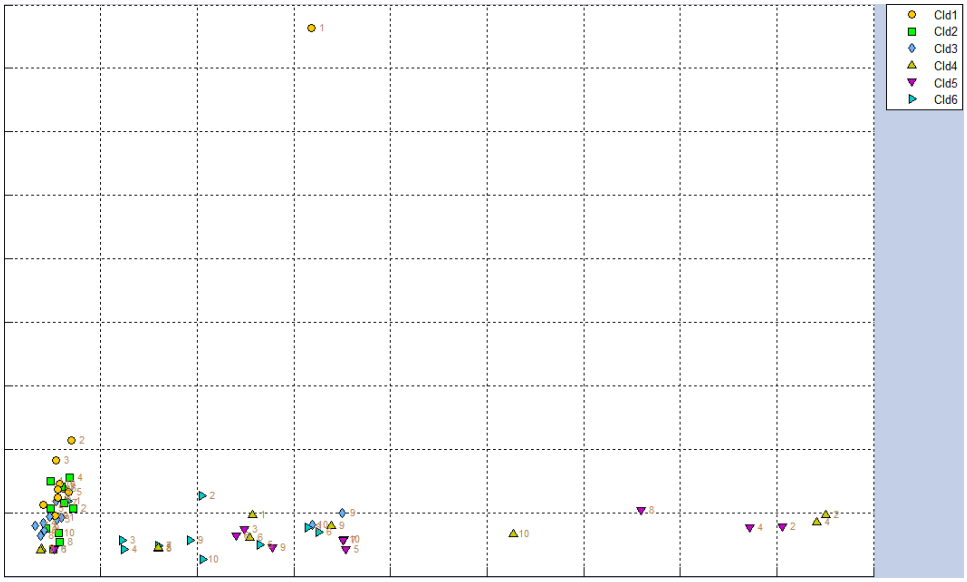

Figure 3. Principal component analysis (PCA) graph of the sixty unnormalized *Cladosporium* exposures based on the two principal components with the highest variance. The exposures begin with the numbered yellow circles for "Cld1" as the first ten exposures and end with the numbered blue triangles for "Cld6" as the last ten exposures.

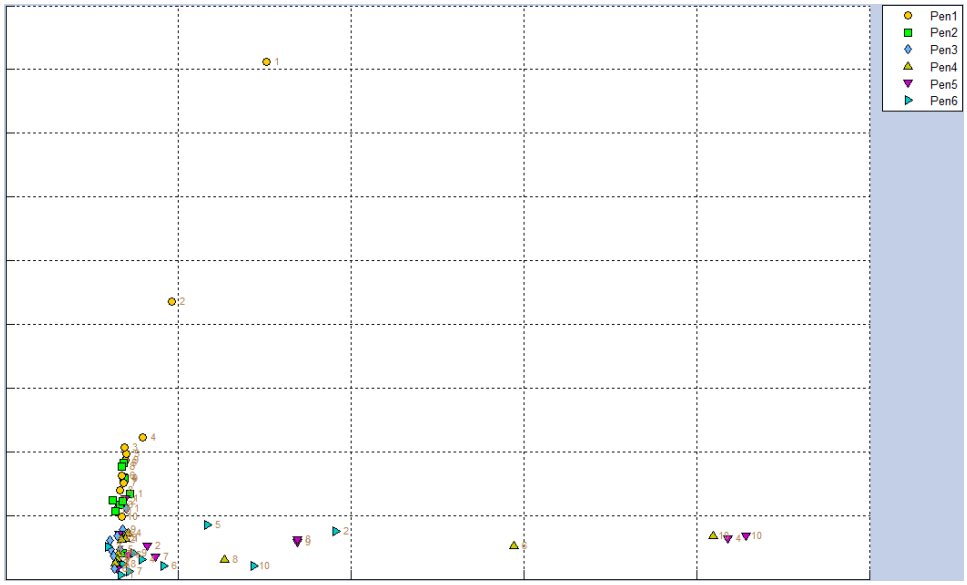

Figure 4. Principal component analysis (PCA) graph of the sixty unnormalized *Penicillium* exposures based on the two principal components with the highest variance. The exposures begin with the numbered yellow circles for "Pen1" as the first ten exposures and end with the numbered blue triangles for "Pen6" as the last ten exposures.

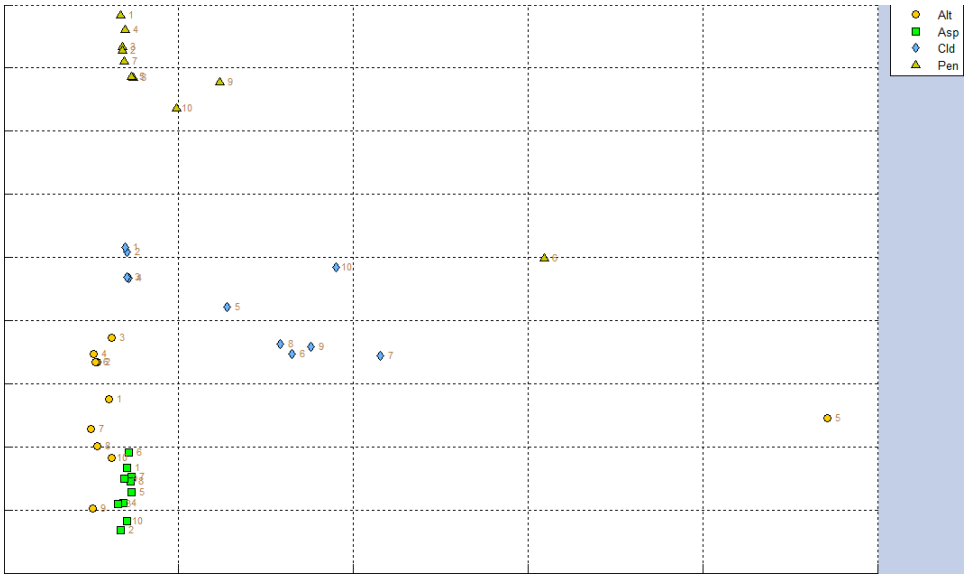

Figure 5. Principal component analysis (PCA) graph of the ten training exposures from each class based on the two principal components with the highest variation. The exposures begin with the plots numbered "1" and end with the plots numbered "10"

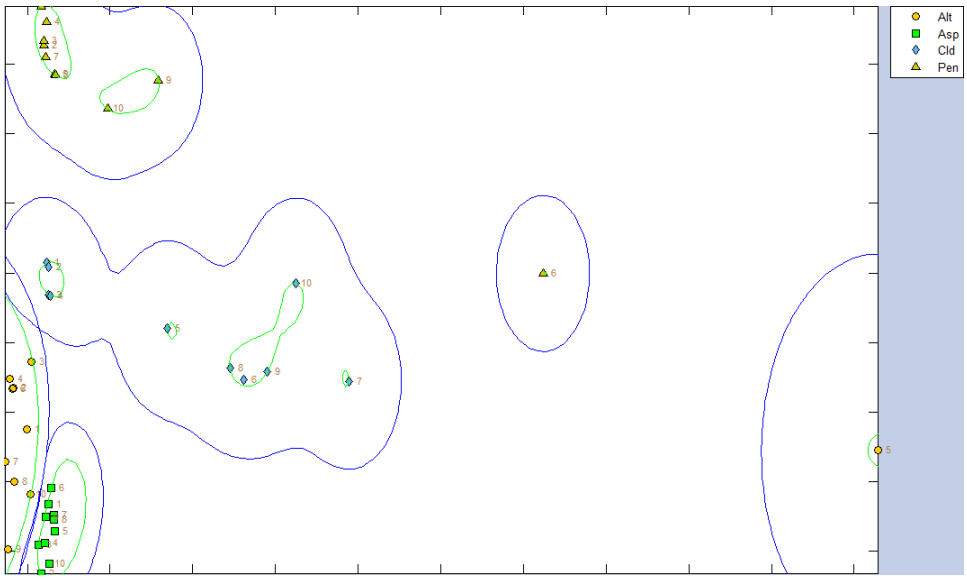

Figure 6. Principal component analysis (PCA) graph of the SVM pattern recognition algorithm of the ten training exposures from each class based on the two principal components with the highest variation. The exposures begin with the plots numbered "1" and end with the plots numbered "10".

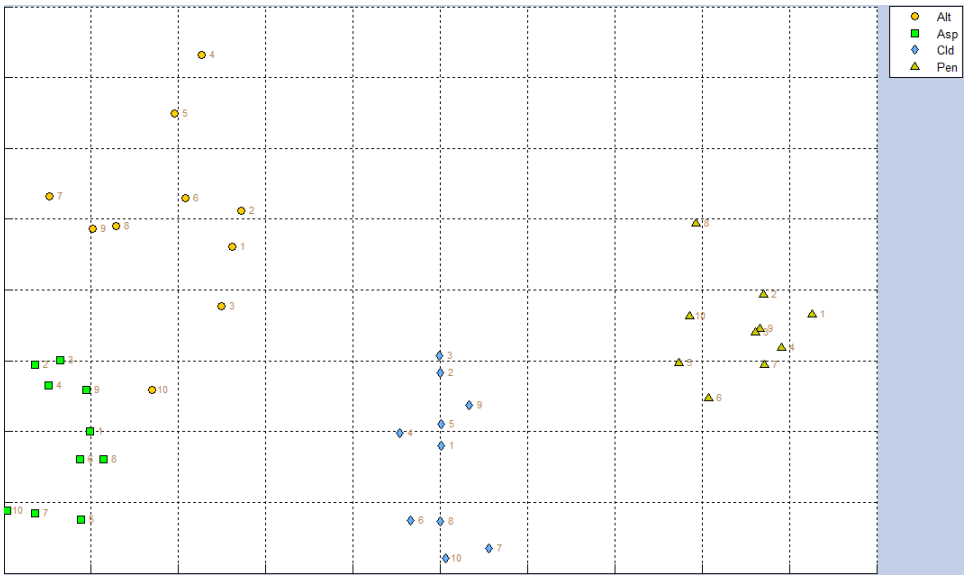

Figure 7. Principal component analysis (PCA) graph of the CDA pattern recognition algorithm of the ten training exposures from each class based on the two principal components with the highest variation. The exposures begin with the plots numbered "1" and end with the plots numbered "10".

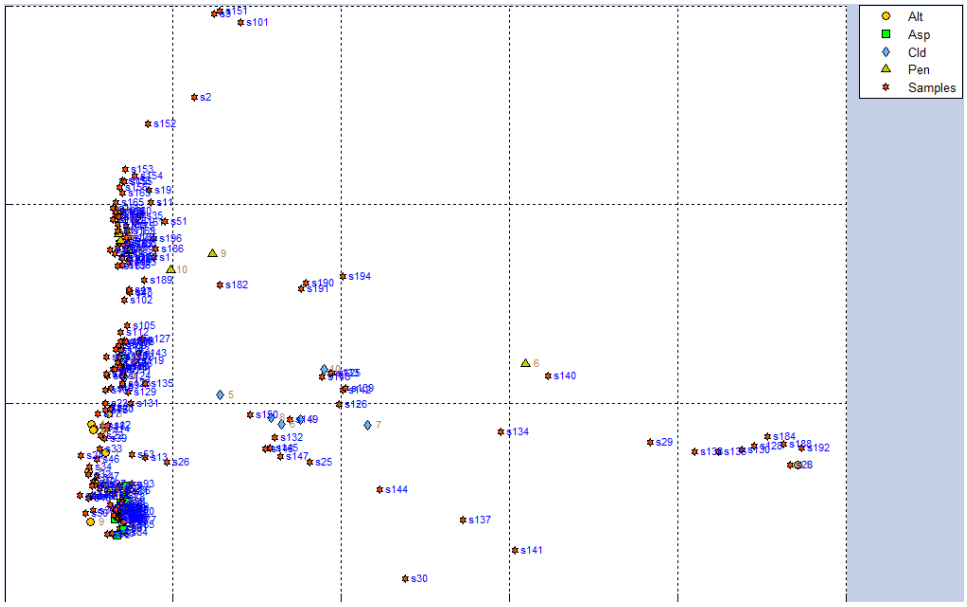

Figure 8. Principal component analysis (PCA) graph of the ten training exposures and fifty normalized identification exposures from each class based on the two principal components with the highest variation. The training exposures begin with the plots numbered "1" and end with the plots numbered "10". The identification exposures begin with the "Sample" plot numbered "1" and end with the plot numbered "200" with 1-50 representing *Alternaria*, 51-100 representing *Aspergillus*, 101-150 representing *Cladosporium*, and 151-200 representing *Penicillium*.