Pest Detection in Wheat by Integrating a Low-cost Electronic nose and Machine Learning Modelling

By

Yajing Ma

Supervisors

Dr. Nir Lipovetzky

Prof. Sigfredo Fuentes

Dr. Eden Tongson

Dr. Claudia Gonzalez Viejo

Submitted for the degree of

Master of Science (Computer Science)

at the University of Melbourne, Australia.

Student Number: 1075728

June, 2022

Abstract

Timely pest detection is very important for preventing the potential hazards of pests, traditional detection methods have been difficult to apply to today's modern agricultural system, although there have been a lot of automatic pest detection methods using advanced technologies, but they have the problems of high cost, difficult to deploy and hard to operate by non-professional personnel, thus it does not apply to non-commercial farmers.

This paper implements a cost-effective pest detection method that can be used for micro-farmers by using a low-cost electronic nose and machine learning modelling. This method was realized by conducting a controlled experiment and making measurements, we designed and conducted controlled experiment using wheat and an oat aphid as experimental materials, while measurements were made using E-nose and an open gas exchange system to acquire data sets used for model construction.

Artificial Neural Network (ANN) was used to develop three models including two classification models and one regression model. These classification models are able to classify the level of pest infestation based on E-nose measurements and the regression model can predict three physiological parameters including photosynthesis, stomatal conductance and transpiration based on E-nose measurements. Both classification models achieve high accuracy around 98% and the best one of them was confirmed through model evaluation and analysis to have no over-fitting or under-fitting problems. The regression model has the performance of overall correlation coefficient 0.79. Model evaluation were analyzed based on accuracy, Mean Squared Error (MSE) and correlation coefficient. These three models enable the method implemented in this study to detect pests efficiently and reliably.

Declaration

I certify that:

- This thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person where due reference is not made in the text.
- Where necessary I have received clearance for this research from the University's Ethics Committee and have submitted all required data to the Department.
- The thesis is 15517 words in length (excluding text in images, table, bibliographies and appendices).

Acknowledgements

I would like to thank you to the Digital Agriculture, Food and Wine Research group (DAFW) belonging to The School of Agriculture and Food, Faculty of Veterinary and Agricultural Sciences (FVAS).

Prof. Sigfredo Fuentes and Dr. Nir lipovetzky for their guidance, motivation, insightful comments and for sharing their knowledge through the entire process of research.

Dr Eden Tongson for her help in establishment of trial and training on gas exchange measurements.

Dr Claudia Gonzalez Viejo for her help in training on e-nose, data acquisition, analysis and machine learning modelling.

Finally, I am really thankful to my family who support me.

Table of Contents

1 Introduction

1.1 Background and Motivation

Agricultural pests are a group of insects that harm the growth and development of crops. They are one of the main factors causing agricultural disasters and can cause great losses to agricultural production and the economy. It can have serious effects on both yield and quality of crops, by consuming crops and producing secretions that affect plant physiological activities or spread plant diseases, pests affect the fruit quality of crops, thus reducing their market value, and also lead to the decline of crop yield or even cause extinction in severe cases. As the global population grows and the demand for food increases, maintaining stable food production has become an important long-term goal for the global community, and the damage caused by pests to food is a major threat to achieving this goal. Globally, pests have been reported to reduce agricultural yields by about 10-16% both before and at harvest, respectively [1]. On the other hand, the invasion of crops by pests can also have a great economic impact, many farmers have suffered economic losses from pest infestation and the high cost of pest management also poses a challenge for human society. According to the survey, the economic losses caused by pests have been underestimated because it is difficult to be estimated accurately and reliably [2].

Effective pest management can prevent the potential damage caused by pests and thus plays a vital role in maintaining the world's crop yields and reducing farmers' economic losses, and lack of early monitoring and management of some pests that are highly reproductive and disruptive can lead to devastating destruction, including infectious illnesses and worldwide agricultural pandemics [3].

Pest detection is the most critical part of pest management, it is the foundation of pest control. Reliable and timely pest detection can help farmers use more rational pest control schemes and reduce the potential damage of pests to the greatest extent. However, unreliable and incomplete information on pest density in farmlands obtained through detection can directly lead to ineffective pest control and even the use of wrong control strategies with more serious consequences. In today's agricultural society, the use of pesticides has become the dominant approach to pest control, due to its ease of implementation, readily availability, and effectiveness [4]. Nevertheless because of the difficulty of reliable and accurate pest detection over a large area of farmland for non-commercial farmers, especially micro-farmers, most of them spray entire fields with the same amount of chemicals. This usually leads to inappropriate use of pesticide doses because the level of pest infection generally varies within the farmland. As a result, excessive pesticides waste resources and may result in biodiversity loss [5], while insufficient amounts of pesticides can be ineffective in eliminating pests. More importantly, the abuse of pesticides, even if it only has a minor effect on agricultural products at present, can cause great harm to human and animal health in the long run. The main factor in farmers' misuse of pesticides as pest control measures is the limited availability of affordable and reliable pest detection methods for non-commercial farmers. Therefore, an important way to alleviate this problem is to come up with more effective, affordable, and reliable pest detection methods that will help most farmers, especially micro-farmers, to be able to apply the appropriate amount of pesticides based on accurate pest detection results.

The most traditional method of pest detection is manual identification [6], where farmers go out to their fields and check on crops and they usually determine the level of infestation by their own experience based on the appearance of the crops including color, shape, and overall condition of crops.

Although this method does not require any additional equipment or significant expenditure, it is not only time-consuming [6], labor-intensive, and inefficient, but also difficult to have a detailed and comprehensive examination in a large field. The rapid development of technology provides a solution to the limitations of traditional methods. Alternative to human visual inspection, digital vision combined with machine learning has attracted extensive attention in the research field of pest detection [4][7][8]. Besides, remote sensing technologies such as Near-Infrared Spectral sensors [9], thermal sensors [10], and fluorescence spectra [11] are also widely used to assist and optimize pest detection. These advanced approaches are quite effective and dependable, but typically require the use of expensive, non-portable technologies that are difficult to set up and are therefore only suitable for commercial grade large-scale farmland.

Electronic-nose, as a system of gas sensors, has been chosen for many plant-pest studies because of its ability to discriminate reliably among volatile organic compounds (VOCs) profiles [12] released from flowers, leaves, and fruits of plants. VOCs provide information that can be used to indicate the physiological state of the plant and as a basis for analysis of plant damage [13]. There are many types of e-noses on the market, most e-noses used in studies are commercial e-noses and are still expensive and complicated to use [14][15]. And so far, most of these studies on the use of e-nose have focused only on the changes in VOCs caused by pests [16][17], which does not directly contribute to pest detection.

Based on the above background description, we can see that although modern advanced technology has aided in the improvement of pest detection systems, a major challenge in the pest detection area is the trade-off between the dependability of pest detection methods and their cost. Hence modern farming requires a not only reliable but also affordable smart pest detection method, which can enable non-commercial farmers in making more reasonable and effective decisions in response to pest infestations, resulting in higher agricultural production and better profit. Furthermore, a solid and cost-effective pest detection method can also help to build a healthier agricultural system and the healthy development of agriculture not only generates certain economic benefits but also contributes considerably to the maintenance of ecological balance.

1.2 Aim and Objectives

This paper aims to implement an efficient, reliable, and affordable pest detection method by using a specific cost-effective e-nose consisting of nine gas sensors combined with state-of-the-art machine learning technology. Our pest detection method was achieved by developing two machine learning models based on the data collected through e-nose including a classification model to classify the pest infestation levels of crops and a regression model to predict physiological parameters (photosynthesis rate, stomatal conductance, and transpiration) on plants based on physiological measurements. This paper also evaluates our proposed pest detection method by evaluating the performance of two models.
Our goal was accomplished by achieving the following objectives:

- We firstly conducted controlled experiments to collect data by growing plants, introducing pests to the plants, and making measurements based on e-nose.
- We analyzed acquired data by carrying out an Analysis of Variance (ANOVA) and Tukey honestly significant difference (HSD) post hoc test to determine the ability of e-nose and physiological parameters to discriminate between plants with different degrees of pest infestation.
- We then developed two machine learning models by performing algorithmic selection, and neuron trimming, based on model evaluation parameters including accuracy and Mean Squared Error (MSE) after extracting and preprocessing the collected data.
- We derived conclusions for our research topic by evaluating the performance of our models and conducting an overfitting and underfitting analysis.

1.3 Thesis Outline

The main content and structure of this paper are as follows:
Chapter 2 Literature Review
In this chapter, we mainly review some existing research on pest detection. Based on different detection methods, they are classified into three groups for discussion that include: image-based pest detection, remote sensing detection, and e-nose-based detection. We also review and present the research contributions of the specific e-nose used in this study in other fields.

Chapter 3 Problem Statement

In this chapter, we describe the gap between the scope and effectiveness of existing pest detection methods and the complete and desired pest detection system required by modern agriculture. In addition, we also clarify the research questions of this paper.

Chapter 4 Data Collection

In this chapter, we demonstrate the implementation of our controlled experiments by elaborating on the experimental materials, experimental equipment we used, experimental settings, and experimental process in detail including the planting and measuring. Furthermore, we justify each of our experimental choices.

Chapter 5 Statistical Analysis

In this chapter, we describe the process of statistical analysis we did, which are ANOVA and *post hoc* tests. The purpose of the analysis is also clarified.

Chapter 6 Model Development

In this chapter, we start with data pre-processing by describing the detailed process and method, and then we introduce the model performance evaluation method used for model selection and explain the model construction process including algorithm selection, neuron trimming, and data division test. Finally, we expand on introducing the two optimal models we choose.

Chapter 7 Results

In this chapter, we present the results of statistical analysis for both e-nose data and physiological measurements of plants and demonstrate the results of different model algorithms tested during our model construction, and further evaluate the performance of our models.

Chapter 8 Discussion

In this chapter, we explicate and analyze important research findings from the results, and indicated the significance of these findings for pest detection. In addition, we also elaborated on the applicability and potential significance of the research results.

Chapter 9 Conclusion

In this chapter, we summarize the content of this research and describe the contributions and limitations of this research, to put forward the direction of future improvement and the potential value of this study in future work.

2 Literature Review

Many automated pest detection methods have been designed to enhance and help improve pest management in modern agriculture. These automatic methods can provide solutions for the limitations of traditional identification such as manual inspection and cumbersome counting of pests stuck on the sticky traps [18]. The enhanced sensitivity of modern digital tools, as well as improvements in data analysis techniques, have resulted in tremendous progress in automatic pest detection research in recent years.

These studies are generally carried out in two stages: 1) collection of different forms of data related to plants and pests including images, videos, thermal or physiological parameters, etc. 2) analysis of the data acquired. For the data analysis, most of the researchers adopted the state of the art technology, which is machine learning and deep learning[4][7][11]. Therefore we divided these studies into three categories mainly based on the data collection techniques they were based on: image-based automatic pest detection, remote sensing technology-based pest detection, and e-nose based pest detection.

2.1 Image-based automatic pest detection

Automatic detection through digital vision is a good alternative to human vision detection, and the images of plants are the most intuitive digital vision, which can provide a wealth of important information for pest detection instead of human observation. So there has been plenty of previous studies on pest detection which are based on plant images. Most of them are powered with machine learning (ML) or deep learning (DL) techniques to process the information extracted from the images [4][7][11].

Ebrahimi et al. have proposed an efficient pest detection method based on crop canopy images by incorporating an image processing technique with the Support Vector Machine (SVM) to detect parasites such as thrips, whitefly, housefly, and ant, which are prevalent in strawberry greenhouses [4]. The results of this study showed that it can successfully detect thrips based on plant images with less than 2.5% error. However, all images used in this

study were captured by mounting a digital camera on a mobile agricultural robot moved along the track. Therefore, despite the success of this work on the accuracy of detection, it has the problem of using unportable and expensive equipment for image collection, which makes the method unsuitable for non-commercial farmers.

A conceptually similar detection method has also been carried out in which white flies are automatically identified from the leaves based on the images of leaves by using various image processing techniques and SVM as a classifier [6]. This method can successfully classify leaf with white flies from leaf without white flies with 98% accuracy. Nevertheless, the method of image acquisition is manual photography, which greatly reduces the automaticity of the method and complicates the practical application as a large number of professionals are required.

In contrast to [6], another study put forward a fully automated pest detection method based on binocular stereo vision and image segmentation technique [8]. although this method implemented a complete pest management system by getting location information to guide the robot's automatic spray, a limitation of this study, similar to study [4], is that the images used were taken from a camera installed on a robot, which is very difficult to operate and deploy by non-specialists. And the high automation of the whole system is achieved on the basis of the high cost of equipment.

In addition to these automatic pest detection methods based on small datasets, a large and growing body of literature has focused on using large image databases to achieve more complex pest detection tasks, including the classification of healthy leaves, pest leaves, and disease leaves [18], and the detection and classification of many different pests [7], even the identification of 24-categories pest [19]. Due to a large amount of data, these methods typically use deep learning such as convolutional neural network (CNN) as the recognition algorithm and they achieve outstanding performance in obtaining high precision results on difficult pest detection tasks, but at the same time, the computational cost of image processing of these methods is extremely high, and the data sets used in these studies are either artificial images created in advanced research institute and laboratory, or collected by deploying a large number of devices such as a wireless image monitoring system comprised of several wireless imaging nodes [7]. Therefore, these studies mainly contribute to the factory-level farmland and are infeasible for micro-farmers.

In general, most image-based pest detection methods have two major limitations. First of all, large quantities of high-quality plant images are difficult and expensive to collect. In actual agriculture, most pests are extremely small in size, so the equipment needed for image

collection has to be advanced enough. And in order to obtain images containing more effective information, there are also high requirements on the angle and light when taking photographs. In addition, due to the diversity of the environment of plant growth, images captured will contain additional interference factors besides pests and plants, making the processing of the images a very complicated and challenging task.

Therefore, another important constraint on all these image-based pest detection methods discussed in this area is that they require additional complex feature extraction processes, which normally is a key factor affecting the effectiveness of these detection methods. For minimizing the noise contained in the information extracted from the image, these methods need to apply some feature extraction techniques to extract effective features from images for further model construction. Based on some basic image processing including image cropping, image segmentation, and image RGB extraction, the features extracted from images for pest detection are usually divided into traditional features and deep features [7]. Features extracted manually through image analysis [4] [6] are often too simple and less adaptive, making it difficult for plants with varying conditions and changing environments in real life. However, features extracted by deep learning [18] [19] usually require high computational costs and a convoluted extraction process.

2.2 Remote sensing technology-based pest detection

Providing the solution to the problem of high requirements of image collection and difficult feature extraction on image-based detection methods, some researchers have attempted to apply different remote sensing techniques to open up new directions for automatic pest detection methods.

Xu *et al.* carried out a study on the reflectance spectra of damaged tomato leaves caused by leaf miners and explored the wavelength that varies most in the changes of tomato leaf damage [9]. Thus, some spectral parameters were used to further determine the pest infestation level of the tomato. An encouraging result showed that the wavelength parameters obtained by near-infrared (NIR) measurements in this study could successfully classify the leaf-miner infestation severity of tomato leaves. Compared with most image-based pest detection methods, although there is no complex feature extraction process required in this study and usable features can be directly obtained through measurement, the measurement process requires the collection of tomato leaves, thus causing damage to the plant itself. This

method is destructive and does not have the sustainability characteristics required by modern agriculture and cannot be generalized.

Another study by Christoph *et al.* on wheat leaf rust detection before symptom occurrence based on fluorescence spectrum combined with SVM [11] also has similar dilemmas. They used a fluorescence spectrometer to take measurements on the wheat leaves that needed to be sampled. One of the advantages of fluorescence spectroscopy compared with reflection spectroscopy is that it can obtain physiological changes before symptoms of plant damage, and thus has the prospect of early detection. But currently available systems for measuring fluorescence spectrum are usually used for research purposes and are very expensive. Besides, the measured data are numerical data that reflect the health of the plant and can therefore be used directly as a feature, but because this method causes damage to the plants themselves, the benefits are outweighed when applied to large farms similar to [9].

Another remote sensing technology used for pest detection that is just as costly as fluorescence spectroscopy is thermal sensing. A study on tomato plants used thermal sensing techniques to propose a method for automatic detection of diseased plants based on depth, temperature, and color information extracted from thermal and visible light image data [10]. The experimental results of this method show that the accuracy of detection can be increased by adding thermal sensitivity information to the basic image information. Although this method reflects the potential of thermal sensing technology in pest detection, the use of the technology still lacks proof of independence. There are relatively few studies in the area of pest detection using only thermal sensing techniques due to the limitations of the effective information that thermal sensing technology can provide and the high cost.

The application of remote sensing technology to pest detection is still being explored. At present, Synthetic aperture radar (SAR) and lidar (Lidar), which have features that can be used to profile the characteristics of plants and their growth environment [20] [21], have also emerged as potential technologies for pest and disease monitoring. However, current research on them is limited only to changes in the structure and morphology of plants caused by pests, and direct application of these techniques for pest detection has not been achieved based on the weak relationship between the information these techniques can provide and the infested plants.

Therefore, the pest detection methods based on remote sensing technology are still in the conceptual stage, with the high cost and low applicability.

2.3 Electronic nose-based pest detection

One of the core basis of all pest detection methods is the changes in plants caused by pest infestation. In addition to the changes in plant leaf appearance, spectrum, structure, and morphology described above, another one that has attracted extensive attention of researchers is changing the volatile organic compounds (VOCs) emitted from plants. The distribution and amount of VOCs emitted by different plants at different phases of development varied substantially, and the difference in VOCs released by infested plants and non-infested plants is significant [12].

A conventional approach for identifying plant VOCs is gas chromatography/mass spectroscopy (GC/MS) [22], which is time-consuming, expensive, and requires experienced manpower to collect and analyze data [11] and it is a well-known analytical technique for the qualitative and quantitative analysis of plant volatiles [17]. In recent years, the electronic nose (E-nose) has gradually become a better alternative to GC/MS not only due to its highly sensitive identification of VOCs but also its convenience of use and non-invasiveness.

E-nose is an instrumental device consisting of an array of sensors that are sensitive to different types of gases [23] and used for reliable differentiation of VOCs profiles emitted by infested and un-infested plants, thus enabling further detection of pests on crops by monitoring the changes of VOCs on plants [12]. In the agricultural area, research into employing E-nose for pest identification is not new, Wu et al. investigated the ability and feasibility of E-nose in the detection of pest infection levels by conducting experiments to detect the infestation levels of pests in wheat grain with different humidity and to distinguish the pest species [14]. This method was able to successfully detect the presence of Red flour beetle in wheat and distinguish different infestation levels, but only at certain humidity levels, because the E-nose Alpha MOS FOX-3000 used in the study is particularly sensitive to humidity.

A new study in recent years used the same E-nose system as an alternative method to simulate biological olfactory processes for the detection of Citrus Tristeza Virus (CTV), a pathogen in Khasi Mandarin Orange plants [15], this method achieves 95.30% accuracy by using random forest classifier. Another E-nose called PEN2 (Airsense Analytics, Germany) consists of 10 sensors equipped with a headspace sampling device that was used to discriminate between the volatile characteristics released from healthy rice plants and those exposed to the different densities of Nilaparvata lugens adults [17]. A recognition model with

over 92.5% accuracy was developed by employing a three-layer back-propagation neural network in this study.

Although these E-noses have demonstrated their excellent performance and helped the methods in these studies achieved good results, which proves their reliability and shows clear and encouraging potential for application as an effective means of pest detection, they are all commercial-grade and therefore, although slightly less expensive than the GC/MS method, still necessitate expensive and elaborate experimental infrastructure and the participation of skilled professional operators, making them inaccessible to most farmers.

Driven by modern market demands, low-cost E-noses have been gradually developed and brought to the market by researchers in recent years. The affordable E-nose is regarded as the best choice to balance the reliability and the cost of discriminating VOCs on different objects with volatile gases. Recently, several studies have been conducted in different fields using a low-cost portable E-nose newly developed by the Digital Agriculture Food and Wine Group from the University of Melbourne [23].

So far, this specific E-nose has been used for different applications with remarkable achievements. These studies have mainly focused on assessing beverage quality, which has important implications for human health. Fuentes *et al.* carried out a field trial to use the low-cost E-nose to classify smoke contamination levels in wines made from grapevines exposed to smoke and coupled with a machine learning algorithm [24]. This study implemented five models, all of which achieved an overall accuracy of around 95%, and these models can provide winemakers and producers with valuable information to help them make better decisions when producing commercial wines by reducing smoke pollution. Because the study used E-nose, which is reliable, efficient, and cheap, the cost of the method has been greatly reduced, making it available to most winemakers and growers. A similar study used the same E-nose to implement a system that can evaluate the quality of beer by examining its aromatic profile [25]. The system consists of four machine learning models and can classify or predict beer fermentation levels, consumer preferences, acceptability, and physicochemical and colorimetric analysis with about overall 90% accuracy. Besides, the E-nose has also been used in predicting the presence of various aromas in coffee and identifying the intensity of coffees [26] with high accuracy.

These studies revealed the reliability and efficiency of the specific low-cost E-nose, thus providing a solid research basis for the development of a cost-effective pest detection method by using this E-nose for most farmers.

3 Problem Statement and Research Question

Many pest detection techniques have emerged as science and technology have progressed. Based on the review of existing research, we can see the existing pest detection technology is advanced, and the contribution to agriculture is significant, but most of them are only applicable to commercial large-scaled farmland. Consequently, in the field of pest detection, an important research gap is that few studies have proposed a cost-effective pest detection method. The trade-off between reliability and the cost of the method is a major challenge. However, it is worth noting that non-commercial farmers are an important part of modern agriculture. Therefore, from a broad perspective, the current pest detection system is incomplete, and an affordable and efficient pest detection method for non-commercial farmland needs to be developed to realize the desired pest detection system in the future. Based on the research gap and the current major challenges in the field described above, the research question proposed in this paper is whether an affordable and reliable pest detection method can be implemented by developing two models with high performance using a low-cost E-nose coupled with a machine learning modeling based on the model evaluation (accuracy and Mean Squared Error) to help growers with pest management.

The following sub-questions will be addressed in this study:

- · Would it be possible to classify the infestation levels on the plant (Low, Medium, High) using a low-cost e-nose and machine learning modeling?
- · Would it be possible to predict physiological parameters (photosynthesis rate, stomatal conductance, and transpiration) using a low-cost e-nose and machine learning modeling?

4 Data Collection

To collect the data needed for developing our models to answer our research question, we firstly conducted a controlled experiment. For providing sufficient reference or repeatable experimental basis for future field trials, this chapter will describe the experimental materials, experimental equipment, experimental settings, and experimental process in detail, and each experimental choice will also be justified.

4.1 Experimental materials

In the long term, this study aims to provide a pest detection method with a degree of universality suitable for most crops. Therefore, the selection of experimental plants and pests in this experiment is based on the principle of representativeness and universality. Based on

this principle, wheat was chosen as the plant material for our experiment since wheat is the most widely grown and important crop in the world [27]. And global wheat demand will grow by roughly 30% from its current production level of 642 million tonnes according to the Food and Agriculture Organization (FAO) [28].

The most common pest in wheat is aphids, which are highly destructive and reproductive. As long as wheat is infected by the aphid, without timely and effective detection and management, it will lead to wheat yield reduction or even extinction. Among the aphids associated with wheat and even other winter wheat, *Rhopalosiphum padi (R. padi)* is currently the prevalent and predominant species [29]. In recent years, it has been found in large numbers at all stages of wheat growth and development and it is the most frequently infested species in wheat growth [30]. Therefore, the pest material we choose for this experiment is *R. padi*, which was provided by laboratory cultures of Pest & Environmental Adaptation Research Group, School of Biosciences, The University of Melbourne, Australia.

4.2 Experimental equipment

Two devices were used to measure and collect data in this experiment: a newly developed Electronic nose by the Digital Agriculture, Food and Wine (DAFW) Group at the University of Melbourne's Faculty of Veterinary and Agricultural Sciences (FVAS) [26] and a Li-6400 XT open gas exchange system (Li-Cor Inc, Environmental Sciences, Lincoln, NE, USA).

The low-cost e-nose used in this experiment is lightweight and portable as shown in Figure 1. It consists of a printed circuit board (PCB) [23] with only 92 mm in diameter and an array of nine gas sensors sensitive to different gases, as well as a humidity sensor and a temperature sensor (AM2320, Guangzhou Aosong Electronics Co., Ltd., Guangzhou, China) to monitor the environmental condition while taking measurements.

Figure 1. The front part of E-nose consists of nine sensors

The model numbers of these nine sensors are MQ-3, MQ-4, MQ-7, MQ-8, MQ-135, MQ-136, MQ-137, MQ- 138, and MG811 (Henan Hanwei Electronics Co., Ltd, China) and their corresponding sensing gases and sensitivity ranges are shown in Table 1.

Table 1: Gas sensors integrated in the E-nose [23]

Sensor (Gas) *	Label/Model	Sensitivity
Alcohol	MQ3	0.5–10 mg L^{-1}
Methane	MQ4	200–10,000 ppm
Carbon monoxide	MQ7	20–2000 ppm
Hydrogen	MQ8	100–10,000 ppm
Ammonia/Alcohol/Benzene	MQ135	10–300 ppm/10–300 ppm/10–1000 ppm
Hydrogen Sulfide	MQ136	1–100 ppm
Ammonia	MQ137	5–200 ppm
Benzene/Alcohol/Ammonia	MQ138	10–1000 ppm/10–1000 ppm/10–3000 ppm
Carbon dioxide	MG811	350–10,000 ppm

The E-nose was programmed using Python (Python, Wilmington, DE, USA) to capture the data and the voltage output from each sensor is proportional to the amount of the particular gas [23] and this is because the e-nose records relative measurements of gases rather than absolute concentrations. The information in volts for each sensor, along with the temperature and humidity readings, is uploaded to a computer via Universal Serial Bus (USB), where it is automatically saved in a comma-separated values (.csv) file to simplify further analysis.

Li-6400 XT open gas exchange system (Figure 2) was used to measure three plant physiological parameters:

- Photosynthesis (*A*; μmol CO_2 m^{-2} s^{-1})
- Stomatal Conductance (*gs*; mol H_2O m^{-2} s^{-1}),

- Transpiration (E; mmol H_2O m^{-2} s^{-1})

Figure 2. Li-6400 XT open gas exchange system

4.3 Experimental settings

Three treatments with different levels of pest infection were designed in our experiment: (i) low level of aphids infestation, (ii) medium level of aphids infestation, and (iii) high level of aphids infestation, and one control was also designed in our experiment: (iv) aphids-free. The specific number of aphids in each treatment will be determined by counting the number of tillers of wheat prior to aphids introduction, which will be explained later in the experimental process. In order to reduce the bias of data collected through the controlled experiment, we set a repeating group for each treatment, that is, each treatment was applied to two groups, so as to improve the reliability of our experimental data. Similarly, there are also two groups for control.

Experimental set-ups and their corresponding labels:

- Two groups of treatment with low-density of aphids: T1-1, T1-1
- Two groups of treatment with medium-density of aphids: T2-1, T2-2
- Two groups of treatment with high-density of aphids: T3-1, T3-2
- Two groups of control with no aphids: CO-1, CO-2

Each group contains 4 adjacent wheat plants and our total number of samples is 32.

Samples: (3 treatments + 1 control) ∗ 2 groups ∗ 4 wheats = 32 wheats

Our experiment was carried out in a growing chamber (Biosciences Glasshouse Complex, The University of Melbourne, Parkville, VIC Australia) with automatic temperature control of 20-25°C and 16 hours of daylight / 8 hours of dark. A general overhead view of the laboratory setup is shown in figure 3.

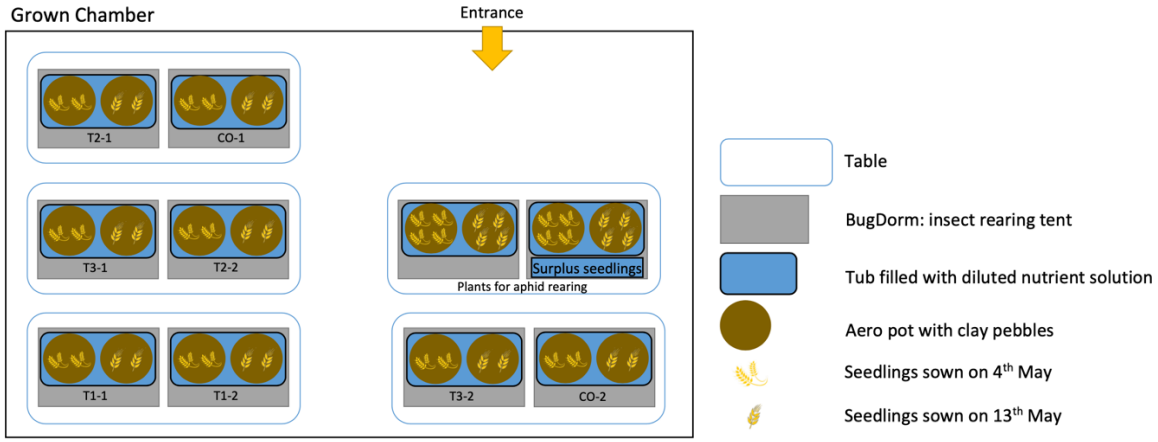

Figure 3. Experimental settings in the grown chamber.

Each group consisted of an insect rearing tent (BugDorm, Australian Entomological Supplies Pty. Ltd., South 116 Murwillumbah, NSW Australia) containing a tub filled with diluted nutrient solution, and two aero pots (Anti-Spiral Pot, Garden City Plastics, Dandenong South, VIC Australia) planted with two wheat plants was placed in each tub (Figure 4). In addition to 8 groups of 32 wheat samples, we also planted some additional plants in two set-ups for aphids rearing. The positions of each group were arranged by random number extraction.

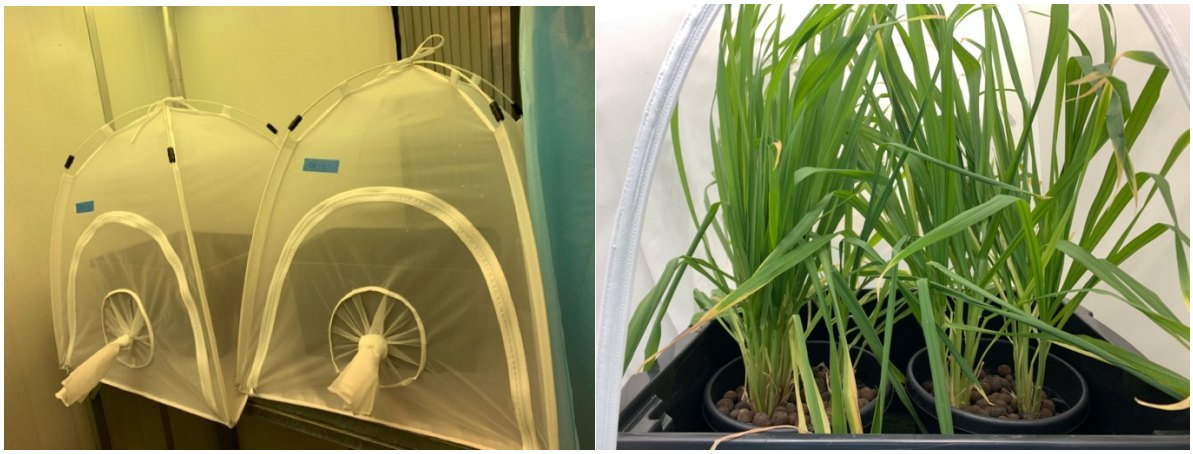

Figure 4 Insect rearing tent (left) containing a tub with two aero pots planted with two wheat plants (right) each.

4.4 Experimental process

The experiment is mainly divided into three parts to acquire data:

Step1: Grow plants for aphid infection.

In order to minimize the interference and influence of external factors on wheat samples before aphids infection and ensure the uniformity among the experimental groups, the wheat in this experiment was grown from seeds, which were first sterilized and placed in a culture dish in the fridge in dark at a temperature of 4 °C for 48 hours, then they were transferred to a light environment at normal room temperature (17-25 °C) for 72 hours. The germinated seeds were next transferred to Jiffy peat pellets that have been hydrated and expanded in warm water, and all peat pellets containing germinated seeds were placed in germination trays with a vented dome (Figure 5). This step is to allow the seedlings to be transplanted until they are large enough to avoid transplant shock, which can result from destroying the delicate root system of the plant. Jiffy Peat Pellets were used due to it consists of compressed peat, wood pulp, and coir pitch pellets and can offer an ideal growing media for seeds.

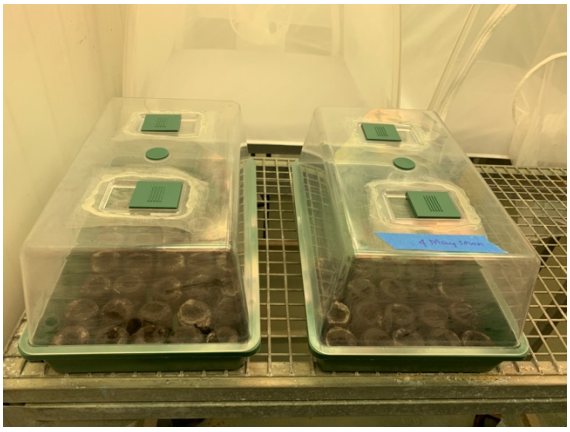

Figure 5. Germination trays with vented dome placed Jiffy peat pellets with seeds

Transplanting was carried out when the seedlings reached the two-leaf stage in the germination tray. We carefully transferred every two wheat seedlings to an aero-pot filled with water-soaked clay pebbles, then we placed a set of two pots into a tub in aphids rearing tent according to the experimental setup, and finally filled each tub with 5L diluted nutrient solution and closed the opening of the tent to allow wheat to grow in an ideal safe environment.

The wheat planting method we use is hydroponics, which is a method of growing plants in water containing nutrient solutions instead of using soil [31]. Hydroponics prevents the occurrence of a variety of malignant diseases transmitted by soil-borne pathogens and cuts off one of the most important channels of pest and disease transmission, thus providing more ideal growing conditions for wheat and improving its survival rate [32]. Furthermore, hydroponics can ease the frequency of watering, thereby reducing their exposure to other pests.

During the whole process of plant growth, we performed a weekly routine inspection on the plant, to monitor the growth of the plant, and supplement or replace the nutrient solution appropriately according to the observed plant status and nutrient solution consumption.

Step2: Aphids infestation trial

To obtain sufficient quantities of aphids for the infestation trial, we first multiplied oat aphids (*R. padi*) by introducing them to healthy wheat grown in two extra experimental set-ups under the same conditions as other set-ups.

The national economic threshold (ET) is used as a criterion to determine the number of aphids that need to be introduced for different treatments. ET is a disease density index that refers to a reasonable time for disease control. That is when the pest or disease reaches this density, control measures should be taken to prevent the disease population from increasing in density and reaching economic damage levels [33]. For Australian winter cereals, the economic threshold for oat aphids is 10-20 aphids per tiller on 50% of tillers [34].

Therefore, according to the national ET, the aphid load standard for three treatments in our experiment is:

- Treatment with low-density of aphids: 5 aphids per tiller in 50% of tillers
- Treatment with medium-density of aphids: 10 aphids per tiller in 50% of tillers
- Treatment with high-density of aphids: 17 aphids per tiller in 50% of tillers

Based on the standard, we manually counted the number of tillers per wheat and calculated the exact number of aphids per wheat needed according to their treatment levels. When the wheat grew to about the stem elongation stage, a specific number of aphids randomly selected from the multiplied population were introduced to the wheat plants in each treatment using a natural brush to start the aphid infestation. And in the process, we made sure that the controls were isolated.

Step3: Make measurements.

Two measurements were scheduled at different times throughout the experiment. A baseline measurement was conducted before the aphid's introduction, and a post-infection measurement was taken 5 days after the aphids were transferred to wheat plants. For each measurement, we had the same process by using E-nose and Licor.

The E-nose as a gas sensor has the same potential problems as other measuring instruments, which can be more or less error-prone after being used for a period of time due to surrounding factors such as humidity and movement. Therefore, in order to ensure the consistency of each measurement and minimize the bias caused by the equipment, we take E-nose outside the tent for calibration in the natural grown chamber environment for 30 seconds every time before and after measuring wheat. Data was recorded by placing E-nose on the top of wheat plants inside the tent for one minute, and repeated four times for each experimental group (each tent) including two times were measured on each pot. To keep E-nose under the

volatile compounds emitted by the plants as much as possible, and avoid affecting by the gases outside the tent, we opened the tent just wide enough for the e-nose to fit in.

The physiological status of wheat was measured by LiCOR to obtain relevant data. Physiological measurements were taken 12 times in total for each group, and repeated three times on different leaves of each wheat.

5 Statistical Analysis

We performed statistical analysis on the physiological and E-nose data acquired to gain basic information about the relationship between the result of measurement and different wheat sample groups. For simplicity, we use the control, low, medium, and high to represent four wheat sample groups including three treatments with different levels of infection and one control when introducing and describing our analysis. Analysis of Variance (ANOVA) and a post hoc test was conducted to examine if E-nose and physiological measurements can discriminate across wheat groups using XLSTAT v.2020.3.1 (Addinsoft, New York, NY, USA) in this study.

ANOVA was adopted in our analysis because it is the most flexible, effective, and common method for analyzing data from controlled experiments, and the computer software that provides the ANOVA is readily available, making this analysis easy accessible, and time-saving to perform. The aim of conducting ANOVA is to determine whether or not there is a statistically significant difference ($p < 0.05$) between the means of the output of nine gas sensors from E-nose for four sample groups including control, low, medium, and high. Similarly, physiological measurements also were analyzed using ANOVA to determine whether the mean values for photosynthesis, stomatal conductance, and transpiration differed between four sample groups.

Based on the result of ANOVA, we also conducted a Tukey honestly significant difference (HSD) *post hoc* test ($a= 0.05$) to obtain more information about the relationships between different sample groups. The statistically significant results of ANOVA can only show that not all wheat groups are equal in means. However, the results do not specify which group difference is significant for the mean.

From the practical significance of this study, the results of ANOVA can only show whether our measurements have the discrimination ability to distinguish between different wheat groups, but they cannot assess the strength of their ability to discriminate, such as, between the control group and the low level of infestation group, two physiologically similar groups in all wheat groups. Therefore, we performed a post hoc test on our measurements to further explore the relationships between different pairwise groups.

6 Model development

Machine learning modeling was performed based on an artificial neural network (ANN) using a customized code written in Matlab® R2020a to develop two models with different purposes in this study: a classification model was constructed to classify the level of infestation on wheat (control, low, medium, high) and a regression model was developed to predict the physiological parameters including photosynthesis, stomatal conductance, and transpiration. The first model is the core of the pest detection method implemented in this study, and its results can be used as an indication to inform farmers whether and to what extent corresponding pest control measures are to be taken, providing them with the most important pest detection information. The second model provides additional plant details to help farmers make more rational decisions about pest management. It also provides the basis for the possibility of using e-nose instead of LiCOR, which is complex and expensive to operate, to obtain the physiological status of plants.

6.1 Data pre-processing

The quantity and quality of data sets play an important role in model construction. Data and its features determine the upper limit of machine learning performance. Therefore, we preprocessed the raw datasets before model training.

E-nose is data logging every 500ms by using the microcontroller with an on-board ADC, which means that E-nose reads data twice per second. Since we have a total of 2 minutes of measurement time including 30 seconds of calibration before and after each measurement and 1 minute of measuring on the top of the wheat. The amount of data obtained per measurement is around 240. And each group repeats 4 measurements, which though provides us with a large enough dataset, but that dataset suffers from poor stability and the amount of data varies

between groups leading to data imbalance. To increase the stability and reliability of the data set and build a better model, we extracted data based on the raw data from E-nose.
A supervised code written in MATLAB (MathWorks Inc., Natick, MA, USA) was used to help us with data extraction. By reading all the data for each measurement, the code automatically recognizes the e-nose output and plots a curvy figure. By observing the figure, we then manually select a starting and ending point to determine the most stable continuous range of measurements, the code subsequently automatically subdivides the selected data into ten equidistant segments and computes an average for each segment, thus outputting 10 averages as the result of the data extraction. This code can simultaneously batch read and process multiple groups of data, and output all the results into a CSV file, each group of output data is separated by a blank line.

After conducting data extraction of E-nose measurements, we also normalized both E-nose and physiological data before building the regression model. Normalization can make each variable in the same range or have the same variance, reducing bias caused by different dimensions, self-variation, or large numerical differences across data. These pre-processed E-nose data were used as inputs for both classification and regression model construction. Since each set of an average of nine sensors extracted was taken as an input, thus, there were 40 inputs in total for each group of four measurements containing 10 averages each.
For the classification model, we use four infestation levels (control, low, medium, high) as targets, and the label values corresponding to each instance were realized by assigning the number 1 or 0 to each level, which means that the corresponding actual infestation level of an instance is set as 1, and all other labels are set as 0.

For the regression model, three physiological measurements including photosynthesis, stomatal conductance, and transpiration were used as multi-targets for model development, one problem with the data set used to build the regression model is that the amount of inputs and targets data does not match. There were 12 physiological measurements in each group and 40 E-nose measurements as we described before, there was a large quantitative difference. Therefore, data matching was carried out as another data pre-processing before regression model construction. We performed the simplest method of data matching, which is to copy and paste the physiological measurements to match the E-nose data. This method can minimize the problem of data authenticity caused by data matching and the purpose of this

study is to detect and report the most realistic state of wheat to detect pests, the authenticity of data is taken as the priority in this paper.

The pre-processing of the raw data maximized the correctness, completeness, and consistency of the dataset used to further construct our model, providing a robust basis for model development.

6.2 Model evaluation method

Data pre-processing is the basis of model training, while the evaluation method is the cornerstone of model selection. In this study, to build the best model, we trained multiple models on the dataset, and different model evaluation methods were used as a standard to compare model performance.

Model evaluation, as an integral part of the machine learning field, is crucial to solving practical problems. Only applying evaluation methods that match application scenarios can better solve research problems. Therefore, it is necessary before model construction to understand the significance of different evaluation indexes and choose suitable ones based on our research questions. This section will introduce the model evaluation methods used in this study.

Each evaluation index has its meaning, most of the evaluation indexes can only reflect a part of the overall performance of the model. If the evaluation indicators are not properly applied, not only the problems of the model itself cannot be found, but also the wrong conclusions will be drawn. A good model evaluation scheme can provide good feedback during model construction to help adjust and optimize the model. Hence, we avoid considering a single one-sided index in the selection of the evaluation method, due to the results obtained in this way usually do not have sufficient reference significance.

This study adopted a combination of different evaluation methods as the model evaluation criteria. Accuracy, Mean Squared Error (MSE), and Correlation Coefficient (R) was used as the main evaluation indexes, as well as Confusion Matrix and Learning Curve were also used as our evaluation tools to form a comprehensive and solid model evaluation system. The model development was based on this evaluation scheme and we also applied it as a model analysis method to evaluate the performance of two models we developed to conclude our research questions.

Accuracy is the simplest and most intuitive evaluation metric in classification problems, but it has a limitation that when the sample proportions of different categories are very unbalanced, the sample class with a large proportion tends to be the dominant class affecting the accuracy. And accuracy can be the best evaluation metric for our classification model since the sample proportion of the four classes (control, low, medium, high) in the data set of this study is relatively balanced. Accuracy is the percentage of the total sample that is predicted correctly. Formula:

$$Accuracy = \frac{The\ number\ of\ instances\ that\ are\ correctly\ predicted}{The\ total\ number\ of\ instances}$$

The magnitude of the accuracy value represents the strength of the model's overall classification ability, and the accuracy value is proportional to the model's classification ability. Accuracy was used as the main evaluation parameter for our classification model.

The MSE is the expected value reflecting the difference between the estimated value and the real value and measures the degree to which the predicted value matches the real value. It is a statistical measure and loss function commonly used in machine learning models such as linear regression. It is also often used to evaluate the deviation of data and the performance of the fitting model. The statistical parameter is the mean of the sum of squares of the errors at the corresponding points between the predicted data and the true data. Formula:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(actual - prediction)^2$$

Contrary to accuracy, MSE is negatively correlated with model performance because it reflects error. Therefore, when the value of MSE is smaller, it indicates that the model has better performance. MSE was used as the main evaluation index for both our classification and regression models.

The correlation coefficient is a measure of the magnitude of the correlation between two variables. It reflects the degree of explanation of predicted value to real value and is an indicator of model explanatory power. Formula:

$$R(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var[X]Var[Y]}}$$

Where $Cov(X, Y)$ is the covariance of X and Y, $Var[X]$ is the variance of X, and $Var[Y]$ is the variance of Y.

When the correlation coefficient is 1, it indicates that the positive similarity of the changes in two variables is the greatest, that is, the perfect positive correlation. As their correlation coefficient decreases, the similarity of the changes among the two variables becomes smaller. When the correlation coefficient is 0, there is no similarity in the change process of the two variables, that is, the two variables are irrelevant.

When the correlation coefficient continues to decrease and is less than 0, the two variables begin to show negative similarities. As the correlation coefficient continues to decrease, the negative similarity will gradually increase. When the correlation coefficient decreases to -1, it indicates that the negative similarity of the changes of the two variables is the largest, that is, the perfect negative correlation. Therefore, the closer the correlation coefficient is to -1 or 1, the higher the correlation will be. When it is 0, there is almost no correlation. The correlation coefficient was used as the main evaluation metric of our regression model.

In addition to the three intuitive evaluation parameters described above, we also used some visual display graphics as evaluation tools for our model. Although accuracy can well reflect the overall performance of the classification model, it cannot meet the specific task requirements, such as evaluating the performance of the model that classifies a specific class from all other classes. Therefore, to analyze the performance of the model more comprehensively, we used the confusion matrix as an important evaluation tool.

The confusion matrix is a visualization of the classification results for each category that compares them with the real results. Through the confusion matrix, we can see the correct or incorrect identification of each class of sample. It provides a more detailed picture of how "good or bad" the model is than accuracy.

In addition, the confusion matrix is one of the most widely utilized approaches in machine learning model analysis since it was resistant to any data distribution. It provides us with detailed information about model errors, which can help us analyze the pest detection methods we implement from different directions based on the model performance.

The confusion matrix for binary classification is shown in Table 2.

Table 2: Confusion matrix of binary classification

	Predict		
Real		Positive	Negative
	Positive	TP	FN
	Negative	FP	TN

- True Positive (TP): The number of samples for which the real label is positive and also predicted as positive correctly.
- False Negative (FN): The number of samples for which the real label is positive but predicted as negative incorrectly.
- False Positive (FP): The number of samples for which the real label is negative but predicted as positive incorrectly.
- True Negative (TN): The number of samples for which the real label is negative and predicted as negative correctly.

This kind of confusion matrix is the basis for computing two important evaluation indexes: Precision and Recall.

Precision refers to the percentage of the sample that the model predicts to be positive is actually positive:

$$Precision = \frac{TP}{TP + FP}$$

In our four-class classification problem, FP is the sum of the number of the other three types incorrectly classified as positive, as shown in table 3, it is the sum of the pink grid values for each row. In contrast to accuracy, which is the overall accuracy of the classifier, precision is the accuracy based on a particular class.

Recall refers to the ratio of the number of samples predicted to be positive among all the real positive samples:

$$Recall = \frac{TP}{TP + FN}$$

In general, the higher Recall is, the more positive samples are correctly predicted by the model, and the better the model performance is.

The confusion matrix used in this paper is a modified version based on the concept of binary classification confusion matrix (See Table. 3), which applies to our four-class classification problem. More importantly, it can intuitively display precision and recall for each class. And in this matrix, the class being analyzed is positive and all other classes are negative.

Table 3: Confusion matrix of classification model for classifying aphids infestation level

Predicted Class	Control	T	F	F	F	*Precision1*
	Low	F	T	F	F	*Precision2*
	Medium	F	F	T	F	*Precision3*
	High	F	F	F	T	*Precision4*
		Recall1	*Recall2*	*Recall3*	*Recall4*	*Accuracy*
		Control	Low	Medium	High	
		Real Class				

In machine learning analysis, even based on all the above evaluation parameters, it cannot be concluded that the model has a good performance. Since overfitting is the most common problem in machine learning modeling [35], especially when modeling small datasets using powerful machine learning methods, thus, overfitting and underfitting analysis must be included in a thorough and comprehensive evaluation method. Learning Curve was used as an evaluation tool in this study for over- and under-fitting analysis. The learning curve is generally a curve showing the change in Accuracy or MSE of the training and validation sets as the number of samples trained increases. The learning curve can show the general trend of the model and indicate the time when the stable state of the model occurs.

For the learning curve used in this study, the X-axis is the number of epochs of training, and the Y-axis is MSE. This is because the number of training as the X-axis of a small data set does not have much reference significance, and using the number of training epochs can better explain the training status of the model. When the validation set reaches the lowest point, the model performs best. By observing the ordinate distance between the validation set

and the training set in the learning curve, it can be seen whether the model has over-fitting phenomenon.

In general, the model evaluation method used in this paper is adapted and designed based on traditional machine learning model evaluation. By combining different evaluation metrics and tools, we realized a comprehensive and robust evaluation strategy suitable for the development of our models.

6.3 Modeling

Several supervised machine learning models were developed based on artificial neural networks (ANN). The construction of our models mainly includes algorithm selection, neuron trimming and data set division test. The three main processes for our model construction were all performed by using a customized code written in Matlab® R2020a.

Different machine learning algorithms have their own characteristics, and for different data sets, the most suitable algorithm is usually different, and the performance of different algorithms on the same data set varies greatly. So it is very necessary to test different algorithms on data sets when building models. A total of seventeen different training algorithms were tested through a loop of the code (See Table. 4). They are grouped into three categories according to the main function type used, including Backpropagation with Jacobian derivatives, Backpropagation with gradient derivatives and Supervised weight and bias training functions.

Table 4: Algorithms used and description of the main function type and abbreviations [39]

Main Function Type	Algorithm	**Abbreviation**
Backpropagation with Jacobian derivatives	Levenberg Marquardt	LM
	Bayesian Regularization	BR
Backpropagation with gradient derivatives	Broyden, Fletcher, Goldfarb, and Shanno quasi-Newton	BFGS
	Conjugate gradient with Powell-Beale restarts	PB
	Conjugate gradient with Fletcher-Reeves updates	FR
	Conjugate gradient with Polak-Ribiere updates	PR
	Gradient descent backpropagation	GD
	Gradient descent with adaptive learning rate	GDLR
	Gradient descent with momentum	GDM
	Gradient descent with momentum and adaptive learning rate	GDMLR
	One step secant	OSS
	Resilient backpropagation	RPROP
	Scaled conjugate gradient	SCG
Supervised weight and bias training functions	Batch training with weight and bias learning rate	BLR
	Cyclical order weight and bias	CO
	Random order weight and bias	RO
	Sequential order weight and bias	SO

In addition to algorithms, the number of neurons is an important parameter of neural network learning, the number of neurons in the input layer equals the number of features and finding the correct number of neurons in the hidden layer is critical to the training of the model. Too few neurons can result in under-fitting, because the network cannot deep learn properly, however, too many neurons can easily lead to overfitting of the model because the network will capture a lot of noise due to overlearning and thus fail to generalize. Therefore, the right number of neurons must be selected to ensure that the model can be well trained.

Based on the size of our data set, we carried out neuron trimming with different neuron numbers including 5, 7, and 10 neurons on the two algorithms that performed best after testing all algorithms. We used 10 as the maximum number of neurons because, in related studies of similar size to our dataset [23-26], they all used 10 as the maximum number of neurons and got good results. Therefore, based on experience with research basis, we set up these three neuron number options.

The number of data sets used for training the model is also an important factor affecting the performance of the model because enough data sets can provide the basis for the network to obtain more information so that the model can get better training, but too small data sets often lead to the network difficult to capture enough information. The data split process is normally used to measure the performance of machine learning algorithms when they are used to make predictions. Using the same algorithm and neurons but with different data partitioning methods, the results of the model will be different.

The same random data splitting was used for both algorithm testing and neuron trimming. For algorithms with no validation set, we divide the data into 70% training set and 30% test set. For algorithms with the validation set, we divide the total data into 70% training sets, 15% validation sets, and 15% test sets. This is one of the most common data partitioning methods in machine learning modeling.

In order to further understand the impact of different data sets on the partition on model performance and to have a more adequate understanding of the model we constructed, after algorithm selection and neural trimming, we also tested another data partitioning method based on their results, it is that the data is randomly divided into 60% of the training set and 40% test set, or 60% of the training set, 20% validation set and 20% test set.

After the algorithm selection, neuron trimming, and data set division tests based on the model evaluation method described in the previous section, several models with different purposes were developed.

Classification models were developed using pre-processed E-nose data as inputs (MQ3: alcohol, MQ4: methane, MQ7: carbon monoxide, MQ8: Hydrogen, MQ135: ammonia/alcohol/benzene, MQ136: Hydrogen sulfide, MQ137: ammonia, MQ138: benzene/alcohol/ammonia, MG811: carbon dioxide) and the level of pest infestation as targets to classify the level of aphids infestation on wheat (control, low, medium, high). Considering the small time interval between the baseline measurement and the first measurement after pest infection, the physiological state of plants without aphids changed very little, so the difference between the plant status at the baseline and the measurement after the first infection was very small. We attempted to add all the measurements of the baseline as control samples to the total data set for model training. This is to increase the number of data sets to try to build better models. Therefore, there are two models for

classifying pest infestation levels using different inputs. The first model used only E-nose data after 5 days of pest infection as inputs, and the second model used both the baseline E-nose measurements and E-nose data after 5 days together as inputs. Both models (See Figure. 6) were developed based on an algorithm that performs best in the algorithm testing, which is the Bayesian Regularization training algorithm. A two-layers feedforward network with a hidden layer and an output layer was used in both two classification models and the hidden layer with 10 neurons used the tan-sigmoid function and the SoftMax function was used in the output layer. Both models were selected mainly based on accuracy, using a randomly divided 70% training set and 30% test set.

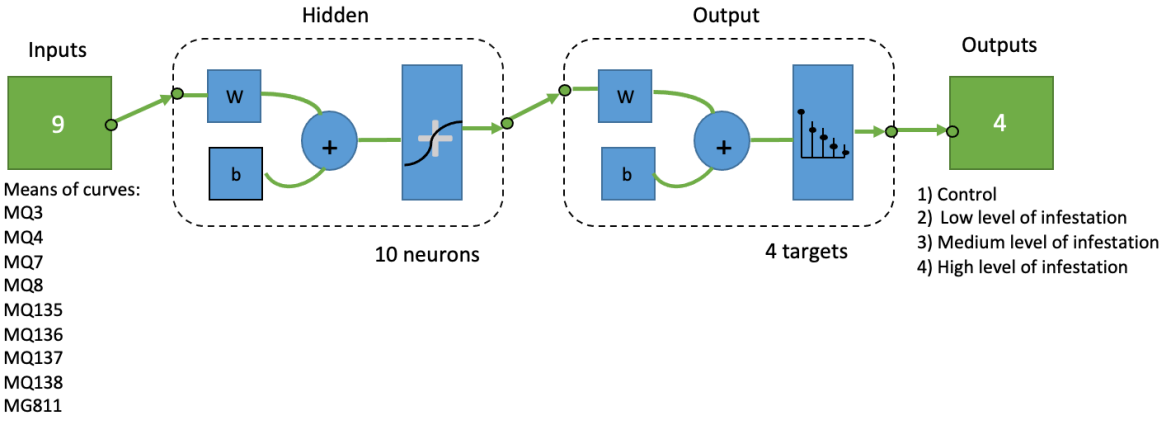

Figure 6: The Artificial neural network structure of two best classification models

For the regression model used to predict the photosynthesis rate, transpiration, and stomatal conductance, the E-nose data after 5 days of aphids introduction with E-nose baseline measurements together were used as inputs (MQ3: alcohol, MQ4: methane, MQ7: carbon monoxide, MQ8: Hydrogen, MQ135: ammonia/alcohol/benzene, MQ136: Hydrogen sulfide, MQ137: ammonia, MQ138: benzene/alcohol/ammonia, MG811: carbon dioxide) and three plant physiological parameters (photosynthesis, transpiration, and stomatal conductance) were used as targets. Both input and target data were normalized from -1 to 1 before training. The best regression model to predict the photosynthesis rate, transpiration, and stomatal conductance was developed using the Bayesian regularization algorithm. Similar to classification modeling, we also tried to train the model by adding E-nose baseline data to the data measured after the infection as inputs, and the effect was significant in regression model fitting. The regression model based only on the data measured after aphids infestation fit performed poorly, and the model trained after adding baseline data became the model that

could be used in this study. Therefore, the best regression model used both the baseline E-nose measurements and E-nose data after 5 days together as inputs and the physiological measurements as targets and it was developed based on the Bayesian Regularization training algorithm selected through algorithm testing. The model is a two-layer feedforward network (See Figure. 7), which consists of a hidden layer with 10 neurons and an output layer and tan-sigmoid function and linear transfer function were used in the hidden layer and the output layer respectively. The training of the model is also based on randomly divided 70% of the training set and 30% of the test set using R and MSE as the main evaluation metrics.

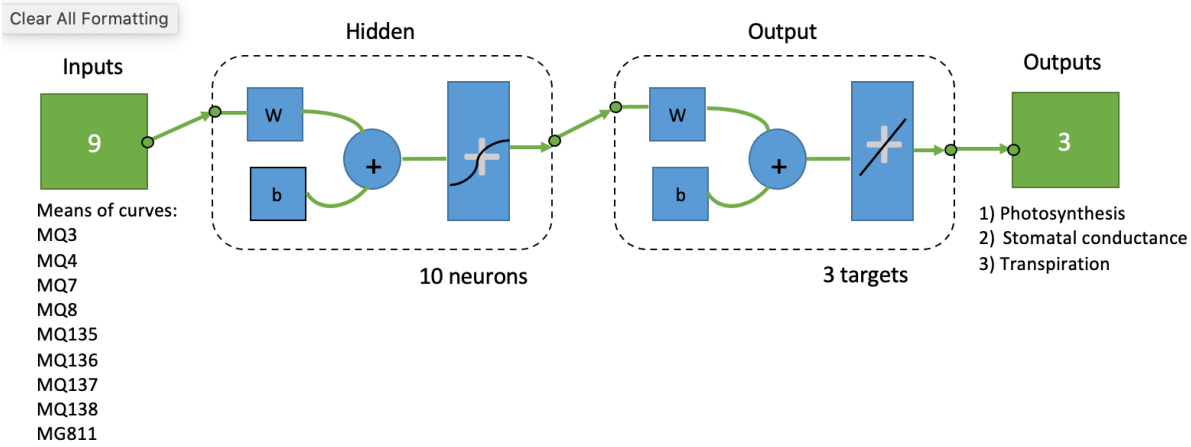

Figure 7: The Artificial neural network structure of the best regression model

A total of three models were developed through our model building strategy of this study, including two classification models with different inputs and a regression model.

7 Results

ANOVA was performed to compare the effect of different levels of aphids infestation in wheat (control, low, medium, high) on the output of nine sensors corresponding to different gases after 5 days of aphids infestation. The result as shown in Table 5 revealed that there were statistically significant differences in both MQ3, MQ7, MQ8, MQ135, MQ136, MQ137, MQ138, and MG811 between different wheat groups ($p < 0.05$) at 5 days after aphids infestation. Among all sensors, only MQ4 (Methane) between different groups was not statistically significant ($p = 0.286 > 0.05$). Some patterns can be seen in all sensors outputs for different groups. For all sensors, the medium group was significantly higher than the other three wheat groups (control, low, high) except for MG811 (Carbon dioxide). The

medium was significantly the smallest for MG811 among four wheat groups (1.358 Volts). The high group was significantly lower than the other three wheat groups for all sensors except for MQ8 (Hydrogen) and MG811. For MQ8, control, Low and high were significantly the same and the high group was significantly higher than the other three wheat groups for MG811. high group was significantly higher than the other three wheat groups for MG811.

Table 5: The ANOVA results for E-nose data measured after 5 days of aphids infestation

Sensor	MQ3	MQ4	MQ7	MQ8	MQ135	MQ136	MQ137	MQ138	MG811
Gas	Alcohol	Methane	Carbon monoxide	Hydrogen	Ammonia /Alcohol /Benzene	Hydrogen Sulfide	Ammonia	Benzene /Alcohol /Ammonia	Carbon dioxide
Control	0.523	1.145	0.223	0.039	0.026	0.000	0.084	0.054	1.413
Low	0.559	1.151	0.235	0.039	0.030	0.000	0.081	0.057	1.370
Medium	0.736	1.143	0.275	0.041	0.037	0.002	0.106	0.070	1.358
High	0.471	1.136	0.198	0.039	0.022	0.000	0.079	0.051	1.418
Pr > F(Model)	<0.0001	0.286	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Significant	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pr > F(Sample)	<0.0001	0.286	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Significant	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

The effect of different levels of aphids infestation in wheat (control, low, medium, high) on the physiological state of wheat (photosynthesis, stomatal conductance, transpiration) at baseline and after 5 days of aphids introduction was also compared by conducting ANOVA. Table 6 presents and compares ANOVA results of photosynthesis, stomatal conductance, and transpiration measured at baseline and five days after infection for different wheat groups. It can be seen from the data in Table 6 that there were non-significant differences in both measurements of photosynthesis, stomatal conductance, and transpiration between different wheat groups ($p > 0.05$) at baseline. For the measurements taken five days after aphids introduction, the mean value of stomatal conductance and transpiration were significantly different between different wheat groups ($p < 0.05$), but there was no statistically significant difference in mean photosynthesis between four wheat groups ($p = 0.07 > 0.5$).

For any physiological parameters measured 5 days after aphids introduction, the measurements of the control group were significantly higher than the other three wheat

groups (low, medium, high). For physiological measurements at baseline, the high group was significantly lower than the other three wheat groups (control, low, medium). Although there was no statistically significant difference in the photosynthesis after 5days of infestation, the p-value dramatically decreased from the baseline (p = 0.924) to the measurement after 5 days of infection (p = 0.070) and was slightly above 0.05.

Table 6: The ANOVA results for physiological measurements of baseline measurements and measurements after 5 days of aphids infestation

	Baseline Measurements			Measurements taken 5 days after aphids introduction		
Sample	Photosynthesis (μmol CO_2 m^{-2} s^{-1})	Stomatal Conductance (mol H_2O m^{-2} s^{-1})	Transpiration (mmol H_2O m^{-2} s^{-1})	Photosynthesis (μmol CO_2 m^{-2} s^{-1})	Stomatal Conductance (mol H_2O m^{-2} s^{-1})	Transpiration (mmol H_2O m^{-2} s^{-1})
Control	7.847	0.125	1.833	7.863	0.234	3.305
Low	8.243	0.148	1.967	6.457	0.124	2.265
Medium	7.815	0.138	1.973	6.203	0.105	1.950
High	7.737	0.116	1.729	5.669	0.102	1.984
Pr > F(Model)	0.924	0.476	0.585	0.070	<0.0001	<0.0001
Significant	No	No	No	No	Yes	Yes
Pr > F(Sample)	0.924	0.476	0.585	0.070	<0.0001	<0.0001
Significant	No	No	No	No	Yes	Yes

Based on the results of ANOVA, we carried out a Tukey honestly significant difference (HSD) post hoc test for both E-nose measurements and physiological measurements. The ANOVA results provided us with basic information on the statistically overall significant differences between groups, while further Tukey tests were performed on all pairwise comparisons among four wheat groups (control, low, medium, high).

As shown in Figure 8 and Figure 9, we use a compact letter display (CLD) with a stacked histogram to show the results. CLD is a very effective method for reporting pairwise comparisons among treatments. The letters in CLD indicate significant differences between groups. Within each output of the sensor, the same letter means there is no significant difference between groups, while different letters mean there is a significant difference between groups.

Since E-nose measurements of MQ4 and physiological measurements of photosynthesis were not statistically different between groups according to ANOVA results, Tukey's HSD test was not used for them and was represented by NS in the figure.

The four groups (control, low, medium, high) in this study form a total of 6 pairwise comparisons, which are control-low, control-medium, control-high, low-medium, low-high and medium-high. From the result of the post hoc test on E-nose measurements taken after 5 days of aphids introduction as shown in Figure 8, we can see some patterns for this comparison between 6 pairwise groups of the nine sensors.

We found that the control group was significantly different from the low group for only MQ135 and MG811 and from the high group for only MQ7 and MQ135. There were statistically significant differences in any sensors between medium and control, or between medium and high and the medium group was also significantly different from the low for any sensors, except for MG811. It is worth noting that even though there were statistical differences between medium and control, or between medium and high for all sensors, there were no statistical differences between the control and high group for most sensors.

In addition, we also observed that only for the measurements of MQ8, MQ136, MQ137, and MQ138, there were statistically significant differences between medium and the other three groups respectively (medium-control, medium-low, medium-high), while there is no statistically significant difference between all pairwise comparisons of the other three groups (control-low, control-high, low-high), and only in these four sensors among all the sensors, there was no statistical difference between low and high.

In addition to the findings from the comparative perspective of different wheat groups based on the result, there were also some surprising results from the sensor perspective. The most important one is that all six pairwise comparisons between the four groups were statistically different only for MQ135 (Ammonia/Alcohol/Benzene), that is, there were statistically significant differences in MQ135 both between control and low, between control and medium, between control and high, between low and medium, between low and high and between medium and high. Apart from MQ135, all pairwise comparisons among low, medium, and high in MQ3, and MQ7 were the statistically significant difference and we can also find that the largest value is from sensors MG811 (), followed by MQ4 (Methane), MQ3 (Alcohol), and MQ7 (Carbon monoxide).

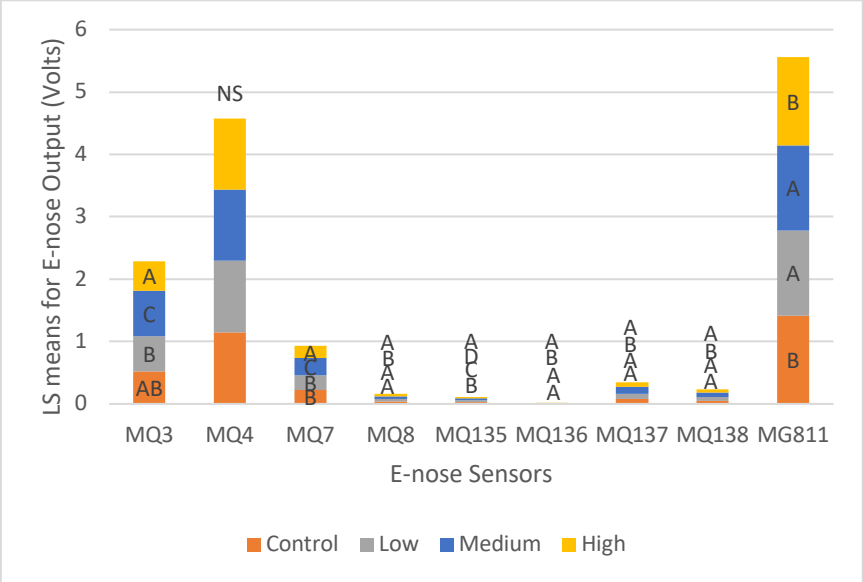

Figure 8: Tukey (HSD) post hoc results for all pairwise comparisons of E-nose outputs for all sensors at 5 days after aphids introduction

The results of Tukey's HSD test on physiological parameters are shown in Figure 9, we found that the control group was significantly different from all the other groups (low, medium, high) respectively for both stomatal conductance and transpiration, but the result showed non-significant differences between all pairwise comparisons of the other three groups (low-medium, low-high, medium-high) for both stomatal conductance and transpiration. Therefore, it indicates that the statistically significant difference between groups shown in ANOVA results is mainly due to the statistically significant difference between the control and the other groups.

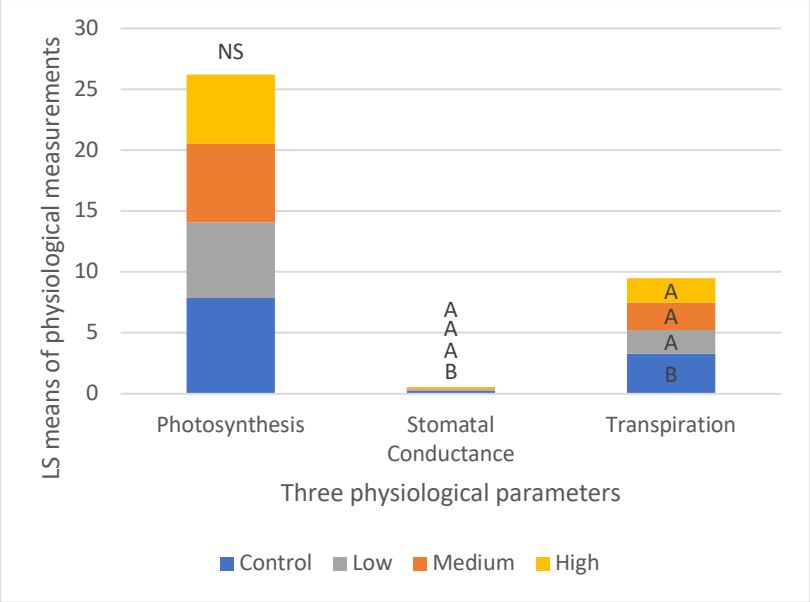

Figure 9: Tukey (HSD) post hoc results for all pairwise comparisons of physiological measurements at 5 days after aphids introduction

Several models were trained in the model development and the three best models were selected based on the results and performance of all models. Table 7 shows the statistical results of seventeen models developed using different algorithms from mainly different three function types, and these models all used measurements taken five days after E-nose infection as inputs and the level of infestation as targets, and they all trained with the same data division method using 10 neurons in the hidden layer. It can be observed that the two models with the best performance were developed using Levenberg-Marquardt and Bayesian Regularization algorithms and they all are from Backpropagation with Jacobian derivatives. The worst-performing model was trained using Sequential order weight and bias algorithm belonging to Supervised weight and bias training functions with only 50.40% overall accuracy. Most models trained using the algorithms of the Backpropagation with gradient derivatives have a good performance, with an overall accuracy of about 87% except for Gradient descent backpropagation and Gradient descent with momentum.

Table 7: Statistical results of classification models developed using seventeen different algorithms.

	Neuron number: 10				
	Data Division: 70% training set, 30% (15% validation set, 15% testing set)				
Main Function Type	**Algorithms**	**Training**	**Validation**	**Testing**	**Overall**
Backpropagation with Jacobian derivatives	LM	15.00%	88.10%	90.50%	93.90%
	BR	99.50%	-	86.90%	95.70%
Backpropagation with gradient derivatives	BFGS	87.20%	88.10%	85.70%	87.10%
	PB	90.80%	88.10%	88.10%	90.00%
	FR	89.30%	81.00%	90.50%	88.20%
	PR	83.20%	100%	88.10%	86.40%
	GD	51.00%	54.80%	54.80%	52.10%
	GDLR	85.70%	92.90%	83.30%	86.40%
	GDM	58.20%	50.00%	64.30%	57.90%
	GDMLR	88.30%	85.70%	90.50%	88.20%
	OSS	87.80%	83.30%	83.30%	86.40%
	RPROP	92.30%	83.30%	90.50%	90.70%
	SCG	92.90%	88.10%	71.40%	88.90%
Supervised weight and bias training functions	BLR	55.10%	61.90%	54.80%	56.10%
	CO	91.80%	90.50%	85.70%	90.70%
	RO	88.80%	71.40%	78.60%	84.60%
	SO	48.50%	-	54.80%	50.40%

Based on the performance results of 17 algorithms, we further conduct neuron trimming and data division tests for the two algorithms with the best performance, and the results are shown in Table 8. Inputs to all models were measurements taken five days after infestation. We can see that when using the same randomly divided 70% training set, 15% verification set, and 15% test set, the Bayesian Regularization algorithm always performs better than Levenberg-Marquardt even with different numbers of neurons, and better models can be trained with 10 neurons than with 5 and 7. Levenberg-Marquardt performs slightly better than Bayesian Regularization only when using a smaller set of data (60% training set). However, it is clear from the table that when we use 60% of the training set, the trained model performs worse than when we use 70% of the training set for both algorithms. The model with the best performance was obtained by using the Bayesian Regularization algorithm based on a 70% training set and 30% test set with 10 neurons in the hidden layer.

Table 8: Statistical results of models trained using algorithms LM or BR with different numbers of neurons (5,7,10) and data division methods. Inputs to all models were measurements taken five days after infestation.

	Algorithm	Levenberg-Marquardt			Bayesian Regularization		
Neurons	Stage	Data Division	Accuracy	Error	Data Division	Accuracy	Error
10	Training	70%	99.00%	1.00%	70%	100.00%	0.00%
	Validation	15%	95.20%	4.80%	15%	-	-
	Testing	15%	90.50%	9.50%	15%	91.70%	8.30%
	Overall		97.10%	2.90%		97.50%	2.50%
7	Training	70%	96.90%	3.10%	70%	100.00%	0.00%
	Validation	15%	81.00%	19.00%	15%	-	-
	Testing	15%	95.20%	4.80%	15%	88.10%	11.90%
	Overall		94.30%	5.70%		96.40%	3.60%
5	Training	70%	97.40%	2.60%	70%	100%	0.00%
	Validation	15%	95.20%	4.80%	15%	-	-
	Testing	15%	90.50%	9.50%	15%	89.3%	10.7%
	Overall		96.10%	3.90%		96.80%	3.20%
10	Training	60%	98.20%	1.80%	60%	100.00%	0.00%
	Validation	20%	91.10%	8.90%	20%	-	-
	Testing	20%	92.20%	7.80%	20%	88.40%	11.60%
	Overall		95.70%	4.30%		95.40%	4.60%

Table 9 shows the results for eight models developed using Bayesian Regularization or Levenberg-Marquardt algorithm and different neuron numbers (5,7,10), the results were obtained by using baseline measurements together with measurements taken five days after an infestation as inputs to all models, which is the main difference from previous eight models. Compared with the data in Table 8, it is obvious that the models have a better performance after increasing the number of data sets no matter of different neurons, different algorithms, or data division methods. As the number of neurons decreases, the overall accuracy of the model gradually decreases. 10 is still the best number of neurons for training models, and Bayesian Regularization performs better than Levenberg-Marquardt when using 10 or 7 neurons. The best model was developed by also using the Bayesian Regularization algorithm based on a 70% training set and 30% test set with 10 neurons in the hidden layer. Therefore, the neural network structure of the best two models is the same even with different inputs.

Table 9: Statistical results of models trained using algorithms LM or BR with different numbers of neurons (5,7,10) and data division methods. . Inputs to all models were baseline measurements together with measurements taken five days after infestation.

	Algorithm	Levenberg-Marquardt			Bayesian Regularization		
Neurons	Stage	Data Division	Accuracy	Error	Data Division	Accuracy	Error
10	Training	70%	99.00%	1.00%	70%	99.80%	0.20%
	Validation	15%	96.70%	3.30%	15%	—	—
	Testing	15%	94.40%	5.60%	15%	96.70%	3.30%
	Overall		98.00%	2.00%		98.80%	1.20%
7	Training	70%	99.00%	1.00%	70%	100.00%	0.00%
	Validation	15%	97.80%	2.20%	15%	—	—
	Testing	15%	91.10%	8.90%	15%	93.30%	6.70%
	Overall		97.70%	2.30%		98.00%	2.00%
5	Training	70%	98.30%	1.70%	70%	98.80%	1.20%
	Validation	15%	97.80%	2.20%	15%	—	—
	Testing	15%	93.30%	6.70%	15%	93.90%	6.10%
	Overall		97.50%	2.50%		97.30%	2.70%
10	Training	60%	97.20%	2.80%	60%	100.00%	0.00%
	Validation	20%	96.70%	3.30%	20%	—	—
	Testing	20%	96.70%	3.30%	20%	92.90%	7.10%
	Overall		97.00%	3.00%		97.20%	2.80%

Table 10 shows the results of the two best classification models selected using measurements taken 5 days after an infestation as inputs (Model 1), or using baseline measurements together with measurements taken 5 days after an infestation as inputs (Model 2) to classify the level of aphids infestation (control, low, medium, high). Both models are built using 10 neurons. As can be seen from the table, both models have high overall accuracy around 98%, and when the amount of data increases from 280 to 600, model 2 (98.8%) shows higher overall accuracy than model 1 (97.5). The accuracy for the training of the two models is extremely high and similar, Model 2 mainly performs better than Model 1 on the test set. For model 2, the MSE of both the test set and the training set were very small and close, so there were no over-fitting or under-fitting problems. However, for model 1, there is a certain gap between the performance of the test set and training set, the under-fitting problem was further analyzed using an evaluation tool.

Table 10: Accuracy, Error and MSE for two machine learning classification models with different inputs developed using Bayesian Regularization algorithm to classify the level of pest infestation.

10 neurons				
Model 1: Inputs: Day 5 only				
Stage	**Samples**	**Accuracy**	**Error**	**MSE**
Training	196	100%	0%	<0.01
Testing	84	91.70%	8.30%	0.04
Overall	280	97.50%	2.50%	-
Model 2: Inputs: Baseline + Day 5				
Training	420	99.80%	0.20%	<0.01
Testing	180	96.70%	3.30%	0.02
Overall	600	98.80%	1.20%	-

Precision and recall results of the model 1 are shown in the confusion matrix in Figure 10 (a), The precision and recall of both four classes (control, low, medium, high) of Model 1 are very high, and the precision of the control group is 100%, so a very important advantage of this model is that none of the actual infected wheat samples were misclassified as uninfected samples. For Low, Medium, and High, their precision and recall are relatively balanced. We can also see that some of the control samples were misclassified as low- or medium-level infected samples.

Figure 10 (b) shows the change of MSE of Model 1 as the number of training epochs increases, from the curve, it can be seen that the overall training status of the model is relatively stable, and the number of epochs 90 can be taken as the cut-off point. Before epoch 90, There was not much change in the model's performance and the MSE of the training set and test set are both high but the gap is small. After epoch 90, the MSE of the model training set began to decrease significantly, while the test set does not decrease together with it and with the increase of the number of epochs, the MSE gap between the training set and the test set gradually increased until the model's training set performance reaches its optimal point. Therefore, we found that the model presents a certain over-fitting.

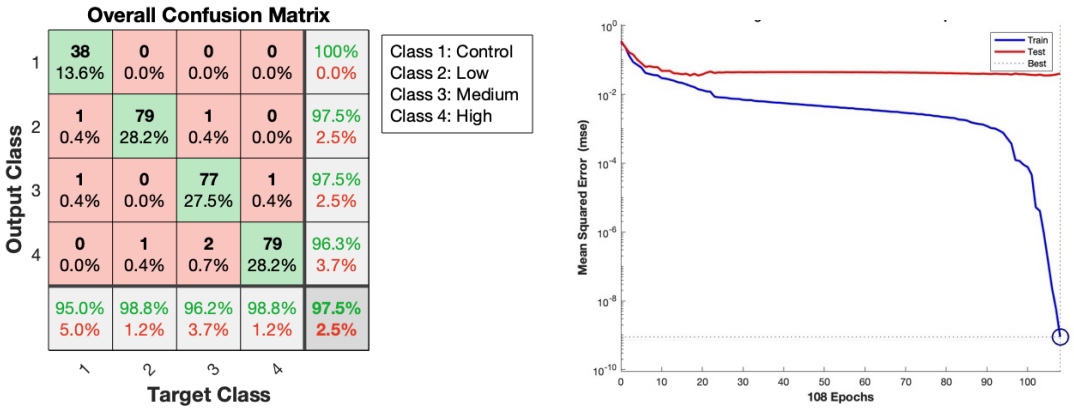

Figure 10: (a) Confusion matrix and (b) Learning curve for machine learning classification model developed based on Bayesian Regularization algorithm and using measurements taken 5 days after infestation as inputs.

For Model 2, due to the addition of a large number of control samples to the dataset, which leads to the problem of the unbalanced dataset and the explanatory power of accuracy on the model performance will be limited. Therefore, the results of precision and recall as shown in Figure 11(a) are crucial for the analysis of this model. The precision and recall for all four sample groups are high, while the precision of control and low are all 100%. Therefore, model 2 can not only avoid the misclassification of actually infected wheat as uninfected plants but also avoid the misclassification of medium- and high-level infected plants as low-level infected plants. Besides, we can also see that the precision of medium-level samples is relatively low at 94%, the model tends to misclassify the other three levels of infection (control, low, high) as medium-level infestation.

As can be seen from the learning curve of the model 2 (Figure 11 (a)), the training process of the model has good stability, and the training and test sets of the model remain stable without any tendency to increase the gap between them after the training set reaches the optimal point of the model performance. Therefore, based on all the analyses there were no indications of over-fitting or under-fitting of model 2.

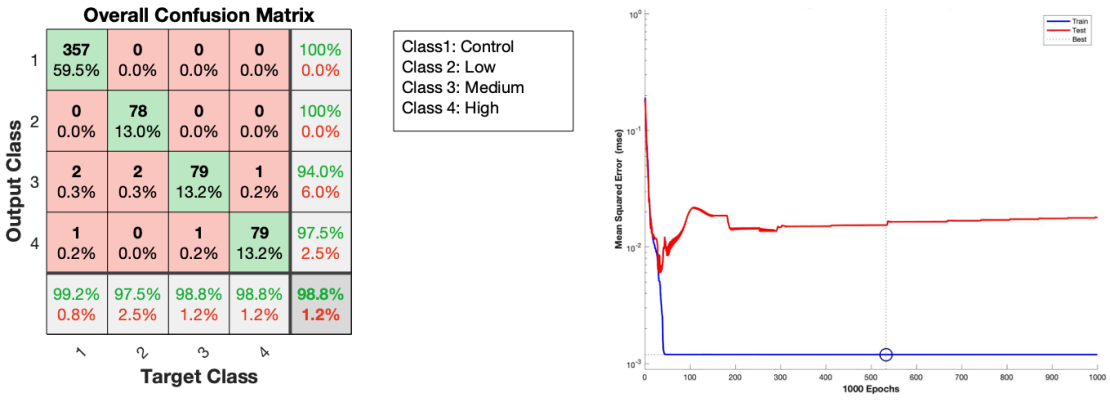

Figure 11: (a) Confusion matrix and (b) Learning curve for machine learning classification model developed based on Bayesian Regularization algorithm and using baseline measurements together with measurements taken 5 days after infestation as inputs.

In the case of the regression model, although we used the same process of building classification models, we only selected part of the results for demonstration and only a brief description will be given because most of the regression models built had poor performance and had no reference value to answer the research questions in this paper.

Among seventeen algorithms in the algorithm selection, except for the three algorithms shown in Table 11, other algorithms only have a correlation coefficient of about or lower than 0.50. Besides, we tried to construct the model using only E-nose measurements five days after infection as the inputs, but all the models trained by 17 algorithms had only correlation coefficient values of around 0.50. Therefore, for all models demonstrated in table 11, baseline measurement of E-nose was used together with the measurement at 5 days after infection as inputs, and physiological measurements at 5 days after infection were used as targets and they were all constructed using 10 neurons based on a randomly divided data set (70% training set, 15% validation set and 15% testing set). From Table 11, it can be seen that the two best-performing models also used Bayesian Regularization and Levenberg-Marquardt algorithm. Among them, the correlation coefficient of Bayesian Regularization (0.78) was slightly higher than that of Levenberg-Marquardt (0.76). Although BFGS performs better than most other algorithms, it still has an overall correlation coefficient of only 0.67.

Table 11: The results of three regression models developed based on different algorithms.

Neuron number: 10				
Data Division: 70% training set, 30% (15% validation set, 15% testing set)				
Algorithms	**Training**	**Validation**	**Testing**	**Overall**
LM	0.78	0.76	0.67	0.76
BR	0.86	-	0.57	0.78
BFGS	0.69	0.63	0.64	0.67

Table 12 shows the detailed results of performance of the regression model we used to predict physiological parameters. It can be seen that the overall correlation coefficient of the model is 0.79 and the slope is 0.68, while the correlation coefficient of the training set is 0.85 and that of the testing set is 0.64.

Table 12: Summary of Correlation coefficient, Slope and MSE for the machine learning regression model developed using Bayesian Regularization algorithm to predict the physiological parameters (photosynthesis, stomatal conductance and transpiration)

10 neurons				
Inputs: Baseline + Day 5				
Stage	**Samples**	**R**	**Slope**	**MSE**
Training	420	0.85	0.72	0.1
Testing	180	0.64	0.57	0.2
Overall	600	0.79	0.68	-

The relationship between E-nose measurements and physiological parameters was visualized as shown in Figure 12. It presents the problem in the regression model, even though there was a clear relationship between E-nose measurements and physiological values, we can see that there are many outliers in the dataset that are far from the fitting line and therefore difficult to fit.

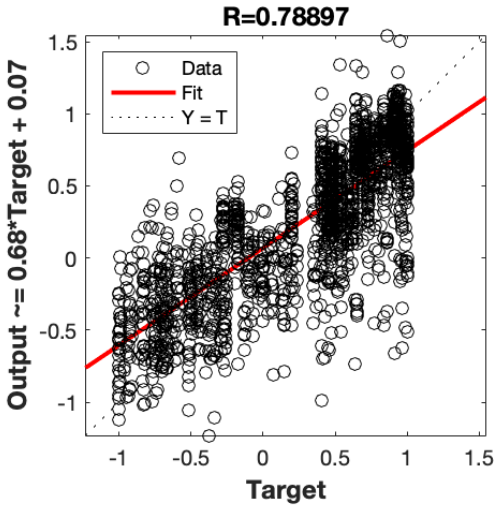

Figure 12: The regression plot for the machine learning model developed using Bayesian Regularization algorithm to predict the physiological parameters (photosynthesis, stomatal conductance and transpiration)

8 Discussion

8.1 Effects of pest infection on different gases and physiological states

Significant differences of most of sensors on different sample groups in the E-nose measurements can indicate the impact of pest infestation on VOCs in plants. Except MQ4, which showed no significant difference between groups, all the other eight sensors could be an indicator for the different gas changes of plants with different infection levels to a certain extent. Therefore, the results confirm that E-nose can be used as a good alternative to the expensive traditional GC/MS method for VOCs identification in plants [22]. This also validates its ability to identify patterns of gas change that have been used by other studies [12]. Furthermore, the results of Tukey's HSD post hoc test shows that the sensitivity of each sensor to pest infection is different, which further indicates that each gas or combination of gases emitted by plants is affected to a different extent by pest infestation. The results could provide evidence for future studies on the interaction between pests and plants. Some gases that are insensitive to pest-plant interaction will not be of great research significance in the future, such as MQ4 (Methane) with no significant differences between groups, while some gases or gas compositions that are extremely sensitive to the interaction between pests and plants can be used as an important factor in future research, such as MQ135 (Ammonia/Alcohol/Benzene) with significant differences between all pairwise groups.

In addition, there was no significant difference in baseline physiological measurements between different groups, indicating that plants growing in the same environment without pest infection had similar physiological states, which also provided a guarantee for the starting conditions of the aphids infestation trial in this study.

After the introduction of pests, it is reasonable that there was no significant difference in photosynthesis between different groups. This is because when plants are attacked by pests, they will automatically adjust or make compensation on the photosynthetic rate to maintain its level [36]. And the change in stomatal conductance and transpiration of plants in response to pest infestation has been described in other studies [37] . When plants are disturbed by the external factors, stomatal conductance and transpiration will decrease.

Although there were significant differences in stomatal conductance and transpiration between different groups, it could be seen from the results of post hoc that the difference was actually caused by significant differences between the control group and other groups, while there were no significant differences between the three treatment groups, suggesting that they can be sensitive to infected and non-infected pests, but did not capable of strong discrimination ability to differentiate among all infected plants.

8.2 Machine learning models for pest detection

Two classification models and a regression model were constructed in this study as the cornerstone of our pest detection method. The two classification models which use E-nose data as inputs and aphids infestation level as targets have high accuracy in classifying the level of infestation on wheat. Both models could be used in the future for early detection of pests. In addition, the evaluation results of the models show that the high precision of the two models in the control group ensured that they would not misclassify infected plants as non-infected pests, which greatly improves the practical application value and generalization of the model. Although the model 1 has the phenomenon of over-fitting, after increasing the datasets by adding the baseline data, model 2 does not show the signs of over-fitting, which indicates that increasing the amount of data sets can be an effective way to solve the over-fitting problem of model 1 in the future. Furthermore, the classes of data set of Model 2 is unbalanced, but it still has a good model performance, which ensures that the model can

accommodate unbalanced dataset. Therefore, Model 2 is a perfect model for classifying pest infestation and thus helping to detect pests .

The regression model using e-nose data as input and physiological measurement as targets used to predict physiological parameters has the performance of overall correlation coefficient 0.79. Although this model cannot be used directly for pest detection, it can provide us with additional useful information about the state of plants. Such model used to monitor plant status has also been proposed in other studies [38] and has achieved good model performance based on a larger data set than the one we used. In contrast, it shows that it is possible to improve the regression model by increasing the data set. Compared with LiCOR, which is used to measure plant physiological parameters directly, E-nose is cheaper, more portable, and more user-friendly, so this model also provides a good demonstration of the potential of using E-nose as an alternative method for monitoring photosynthesis, stomatal conductance and transpiration of plant. Therefore, the three models constructed in this paper can be used to monitor plant status and classify pest infestation levels, which can be a potential cost-effective method for a more comprehensive pest management system in the future.

9 Conclusion

This study presented a novel pest detection method by integrating a low-cost E-nose and machine learning modeling, our research questions can be answered from the analysis of three developed models. By designing controlled experiments, making measurements, building models and analyzing results, we can conclude that it is possible to predict the levels of pest infestation using a low-cost E-nose combined with machine learning, and it is also possible to predict the photosynthesis, stomatal conductance and transpiration in wheat using a low-cost E-nose combined with machine learning when sufficient data sets are available. In general, an affordable, reliable and effective pest detection method was implemented by developing three models with high performance using a low-cost E-nose powered by machine learning modelling based on the model evaluation.

The contribution of this paper is mainly divided into three points:

- The E-nose used in this paper is a rare low-cost and portable E-nose in the market, and the research on this kind of e-nose is very limited. Therefore, this paper

successfully explores the potential application value of the e-nose by using it in the agricultural field and achieving good results.

- The cost-effective method presented in this study can be used for most farmers including micro-farmers to detect pests by integrating it with a automated pest management systems, so as to fill the gap of the lack of economical and efficient pest detection methods in the agricultural field.
- The three models developed in this paper can be a good basis for more complex studies related to pest-plant interaction and the regression model could also be used for monitoring physiological status of plants on other studies.

Although this study has great contributions, it has certain limitations:

- The classification on the level of pest infestation that includes only three levels is relatively simple, and infestations can be more complicated in real life, especially on large farmlands.
- The study used only one species of plant and pest, respectively, so it is unclear whether it has the ability to be applied to a wider range of other plants and pests.
- Plant growth environment is usually a lot more complicated in nature, in addition to pests, diseases are likely or other external factors affect the development of plants, and in our controlled experiment, pest is the only factor that affect plants, so this study is the lack of field trials to further verify the effectiveness of the method.

Future direction

- In the future, this method can be applied to different plants and pests to verify the generalization ability of this method. Besides the classification of pest infestation, we can further explore the classification of different kinds of pests that cause plant damage.
- Since the pest infection trial in this experiment is based on the number of pests counted, we can try to predict the number of pests by collecting more data sets in the future, so as to obtain a more accurate pest detection result.
- Based on this method, field trial can be carried out in the future. In large farms, we can set samples and collect sample data, and use the same process of this method to analyze the data, so as to create a more robust and economic pest detection method.

References

[1] Bradshaw, C. J., Leroy, B., Bellard, C., Roiz, D., Albert, C., Fournier, A., ... & Courchamp, F. (2016). Massive yet grossly underestimated global costs of invasive insects. *Nature communications*, *7*(1), 1-8.

[2] Sharma, S., Kooner, R., & Arora, R. (2017). Insect pests and crop losses. In *Breeding insect resistant crops for sustainable agriculture* (pp. 45-66). Springer, Singapore.

[3] Bellotti, A., Herrera Campo, B. V., & Hyman, G. (2012). Cassava production and pest management: present and potential threats in a changing environment. *Tropical Plant Biology*, *5*(1), 39-72.

[4] Ebrahimi, M. A., Khoshtaghaza, M. H., Minaei, S., & Jamshidi, B. (2017). Vision-based pest detection based on SVM classification method. *Computers and Electronics in Agriculture*, *137*, 52-58.

[5] Mahmood, I., Imadi, S. R., Shazadi, K., Gul, A., & Hakeem, K. R. (2016). Effects of pesticides on environment. In *Plant, soil and microbes* (pp. 253-269). Springer, Cham.

[6] Gondal, M. D., & Khan, Y. N. (2015). Early pest detection from crop using image processing and computational intelligence. *FAST-NU Research Journal*, *1*(1), 59-68.

[7] Rustia, D. J. A., Chao, J. J., Chiu, L. Y., Wu, Y. F., Chung, J. Y., Hsu, J. C., & Lin, T. T. (2021). Automatic greenhouse insect pest detection and recognition based on a cascaded deep learning classification method. *Journal of Applied Entomology*, *145*(3), 206-222.

[8] Li, Y., Xia, C., & Lee, J. (2009, July). Vision-based pest detection and automatic spray of greenhouse plant. In *2009 IEEE international symposium on industrial electronics* (pp. 920-925). IEEE.

[9] Xu, H. R., Ying, Y. B., Fu, X. P., & Zhu, S. P. (2007). Near-infrared spectroscopy in detecting leaf miner damage on tomato leaf. *Biosystems Engineering*, *96*(4), 447-454.

[10] Raza, S. E. A., Prince, G., Clarkson, J. P., & Rajpoot, N. M. (2015). Automatic detection of diseased tomato plants using thermal and stereo visible light images. *PloS one*, *10*(4), e0123262.

[11] Römer, C., Bürling, K., Hunsche, M., Rumpf, T., Noga, G., & Plümer, L. (2011). Robust fitting of fluorescence spectra for pre-symptomatic wheat leaf rust detection with support vector machines. *Computers and Electronics in Agriculture*, *79*(2), 180-188.

[12] Laothawornkitkul, J., Moore, J. P., Taylor, J. E., Possell, M., Gibson, T. D., Hewitt, C. N., & Paul, N. D. (2008). Discrimination of plant volatile signatures by an electronic nose: a potential technology for plant pest and disease monitoring. *Environmental science & technology*, *42*(22), 8433-8439.

[13] Zhou, B., & Wang, J. (2011). Discrimination of different types damage of rice plants by electronic nose. *Biosystems engineering*, *109*(4), 250-257.

[14] Wu, J., Jayas, D. S., Zhang, Q., White, N. D. G., & York, R. K. (2013). Feasibility of the application of electronic nose technology to detect insect infestation in wheat. *Canadian Biosystems Engineering*, *55*.

[15] Hazarika, S., Choudhury, R., Montazer, B., Medhi, S., Goswami, M. P., & Sarma, U. (2020). Detection of citrus tristeza virus in mandarin orange using a custom-developed electronic nose system. *IEEE Transactions on Instrumentation and Measurement*, *69*(11), 9010-9018.

[16] Lan, Y. B., Zheng, X. Z., Westbrook, J. K., Lopez, J., Lacey, R., & Hoffmann, W. C. (2008). Identification of stink bugs using an electronic nose. *Journal of Bionic Engineering*, *5*(1), 172-180.

[17] Zhou, B., & Wang, J. (2011). Use of electronic nose technology for identifying rice infestation by Nilaparvata lugens. *Sensors and Actuators B: Chemical*, *160*(1), 15-21.

[18] Devi, M. S., & Roy, K. (2017). Comparable study on different coloured sticky traps for catching of onion thrips, Thrips tabaci Lindeman. *Journal of Entomology and Zoology Studies*, *5*(2), 669-671.

[19] Jiao, L., Dong, S., Zhang, S., Xie, C., & Wang, H. (2020). AF-RCNN: An anchor-free convolutional neural network for multi-categories agricultural pest detection. *Computers and Electronics in Agriculture*, *174*, 105522.

[20] Capodici, F., Maltese, A., Ciraolo, G., D'Urso, G., & La Loggia, G. (2017). Power sensitivity analysis of multi-frequency, multi-polarized, multi-temporal SAR data for soil-vegetation system variables characterization. *Remote Sensing*, *9*(7), 677.

[21] Pichierri, M., Hajnsek, I., Zwieback, S., & Rabus, B. (2018). On the potential of Polarimetric SAR Interferometry to characterize the biomass, moisture and structure of agricultural crops at L-, C-and X-Bands. *Remote Sensing of Environment*, *204*, 596-616.

[22] Ghaffari, R., Laothawornkitkul, J., Iliescu, D., Hines, E., Leeson, M., Napier, R., ... & Taylor, J. E. (2012). Plant pest and disease diagnosis using electronic nose and support vector machine approach. *Journal of plant diseases and protection*, *119*(5), 200-207.

[23] Viejo, C. G., Fuentes, S., Godbole, A., Widdicombe, B., & Unnithan, R. R. (2020). Development of a low-cost e-nose to assess aroma profiles: An artificial intelligence application to assess beer quality. *Sensors and Actuators B: Chemical*, *308*, 127688.

[24] Fuentes, S., Summerson, V., Gonzalez Viejo, C., Tongson, E., Lipovetzky, N., Wilkinson, K. L., ... & Unnithan, R. R. (2020). Assessment of smoke contamination in grapevine berries and taint in wines due to bushfires using a low-cost E-nose and an artificial intelligence approach. *Sensors*, *20*(18), 5108.

[25] Gonzalez Viejo, C., & Fuentes, S. (2020). Low-Cost Methods to Assess Beer Quality Using Artificial Intelligence Involving Robotics, an Electronic Nose, and Machine Learning. *Fermentation*, *6*(4), 104.

[26] Gonzalez Viejo, C., Tongson, E., & Fuentes, S. (2021). Integrating a low-cost electronic nose and machine learning modelling to assess coffee aroma profile and intensity. *Sensors*, *21*(6), 2016.

[27] Deutsch, C. A., Tewksbury, J. J., Tigchelaar, M., Battisti, D. S., Merrill, S. C., Huey, R. B., & Naylor, R. L. (2018). Increase in crop losses to insect pests in a warming climate. *Science*, *361*(6405), 916-919.

[28] Sharma, I., Tyagi, B. S., Singh, G., Venkatesh, K., & Gupta, O. P. (2015). Enhancing wheat production-A global perspective. *Indian Journal of Agricultural Sciences*, *85*(1), 3-13.

[29] Hesler, L. S. (2005). Resistance to Rhopalosiphum padi (Homoptera: Aphididae) in three triticale accessions. *Journal of Economic Entomology*, *98*(2), 603-610.

[30] Savaris, M., Lampert, S., Salvadori, J. R., Lau, D., Pereira, P. D. S., & Smaniotto, M. A. (2013). Population growth and damage caused by Rhopalosiphum padi (L.)(Hemiptera, Aphididae) on different cultivars and phenological stages of wheat. *Neotropical Entomology*, *42*(5), 539-543.

[31] Tripp, T. (2014). *Hydroponics advantages and disadvantages: Pros and cons of having a hydroponic garden*. Speedy Publishing LLC.

[32] Nalwade, R., & Mote, T. (2017, May). Hydroponics farming. In *2017 International Conference on Trends in Electronics and Informatics (ICEI)* (pp. 645-650). IEEE.

[33] Plant, R. E. (1986). Uncertainty and the economic threshold. *Journal of economic entomology*, *79*(1), 1-6.

[34] McDonald, G.; Umina, P.; Hangartner, S. Oat Aphid Rhophalosiphum padi. . Availabe online: https://cesaraustralia.com/pestnotes/aphids/oat-aphid/ (accessed on 25 November 2020).

[35] Dietterich, T. (1995). Overfitting and undercomputing in machine learning. *ACM computing surveys (CSUR)*, *27*(3), 326-327.

[36] Zhao, W., Liu, L., Shen, Q., Yang, J., Han, X., Tian, F., & Wu, J. (2020). Effects of water stress on photosynthesis, yield, and water use efficiency in winter wheat. *Water*, *12*(8), 2127.

[37] Francesconi, S., & Balestra, G. M. (2020). The modulation of stomatal conductance and photosynthetic parameters is involved in Fusarium head blight resistance in wheat. *PloS one*, *15*(6), e0235482.

[38] Fuentes, S., Tongson, E., Unnithan, R. R., & Gonzalez Viejo, C. (2021). Early Detection of Aphid Infestation and Insect-Plant Interaction Assessment in Wheat Using a Low-Cost Electronic Nose (E-Nose), Near-Infrared Spectroscopy and Machine Learning Modeling. *Sensors*, *21*(17), 5948.

[39] Gonzalez Viejo, C., Torrico, D. D., Dunshea, F. R., & Fuentes, S. (2019). Development of artificial neural network models to assess beer acceptability based on sensory properties using a robotic pourer: A comparative model approach to achieve an artificial intelligence system. Beverages, 5(2), 33.