Utilizing Pattern Recognition Methods for Detecting the Adulteration of Glucose and Fructose in Honey

Saeed Faal¹, Mohammad Loghavi², Saadat Kamgar³*, Mohammad Hoseyn Raoufat⁴, Mohammad Taghi Golmakan⁵

1,3- Assistant professor, Biosystems Engineering Department, Shiraz University, Shiraz, Iran
2,4- Professor, Biosystems Engineering Department, Shiraz University, Shiraz, Iran
5- Associate professor, Food Science Engineering Department, Shiraz University, Shiraz, Iran

* Corresponding author (kamgar@shirazu.ac.ir)

Abstract
The aroma of honey is one of the important parameters in honey grading and that is depended on several factors, such as geographical origin, climate, botanical and environmental conditions. The aim of this study was the development and evaluation of an electronic nose as a new, fast and nondestructive method for detecting adulteration in honey. In this research, the ability of electronic nose as a non-destructive system for detecting honey adulteration with different percentages (pure, 20% syrup, 40% syrup, 60% syrup and 80% syrup) was investigated. The developed electronic nose consists of 8 metal oxide semiconductor sensors (MOS) to detect adultery in honey. After preprocessing the data obtained from the electronic nose the chemometric methods were utilized to classify different type of honey. Principle component analysis (PCA), hierarchical cluster analysis (HCA), linear discriminate analysis (LDA), were used to analyze the data obtained from electronic nose. Based on the results, the detection of adulteration was 98.4% of variance for PCA method, 99% accuracy for HCA method and 100% classification power by LDA method.

Keywords
Adulteration
Electronic Nose
Honey
Pattern Recognition
Sensor

Introduction
Food quality is a complex concept referring to multiple characteristics that make a food product acceptable or more desirable to eat. Important food quality aspects are safety, nutritional value, functionality, and aesthetics (color, texture, aroma, appearance). While the first three are subjectively quantifiable, the last has an even more important subjective component, which makes it more difficult to describe and/or quantify. Aroma, is a very important component of this subjective quality (Lammertyn et al., 2004).

Traditional analytical and quantitative techniques for aroma analysis include HPLC, GC with headspace sampling and GC-MS analysis with solid phase microextraction. Numerous reports exist on the aroma analysis of a wide range of food products with these techniques and they have proven to give very precise and reliable results. It has been proven that they give these techniques, however, involve a lot of sample preparation, are time consuming and can only be carried out in a specially equipped laboratory environment by well-trained operators. Next to a
chemical characterization, aroma analysis often also includes a sensory evaluation by both trained taste panels and consumer panels. This type of evaluation is important in classifying aroma characteristics according to human perception and consumer behavior. Evidently, this is a very subjective and variable evaluation, which involves a very costly and time-consuming procedure (Lammertyn et al., 2004).

Honey contains a viscous solution, as well as supersaturated sugar from flower nectars, which is collected and produced by the honeybee. According to the Iranian National Standard Rules, honey is defined as a pure substance, additives cannot be added to it. Due to the high demand of Iranian honey in the country as well as neighboring countries, this has led to an increase in the sale of honey from pure honey to adulterated honey. Adulteration with sweeteners is the most important issue to assess the authenticity of honey. Sweeteners that have been detected in honey are sugar syrup, corn molasses, sugar cane and sugar beet which is affected by the acid or enzyme. Various analytical techniques, including: isotopic (Padovan et al., 2003, Cabanero et al., 2006), chromatographic (Cordella et al., 2003; Morales et al., 2008) and thermal analysis (Cordella et al., 2005) have been implemented for the detection of honey adulteration. The strength of these methods in honey adulteration detection has been proven by numerous researches, however, they are time-consuming, destructive, and some of them expensive. Therefore, fast, non-destructive, and precise analytical methods are welcome to complement the existing techniques (Shafiee et al., 2016). One of the new techics for food quality control is the e-nose technic, e-nose is an instrument designed to emulate the sense of smell and discriminate among complex odors by means of an array of gas sensors (which respond to gases and vapors generated by the sample) and multivariate data analysis methods (Kiani et al., 2016).

There are several studies reporting the use of e-nose for quality and adulteration assessment of foods, aromatic plants and fruits such as olive oil (Oliveros et al., 2002; Melucci et al., 2016), milk (Yu et al., 2007), soy sauces (Gao et al., 2017), pumpkin (Zhou et al., 2017), honey (Lammertyn et al., 2004; Zakaria et al., 2011), saffron (Heidarbeigi et al., 2015; Kiani et al., 2017), tomato (Wang & Zhou, 2007), tea (Roy et al., 2012; Huo et al., 2014), apple (Ezhilan et al., 2018), coffee (Freitas et al., 2001), sesame oil (Hai & Wang, 2006) and tomato juice (Man et al., 2005). Therefore, the objective of this study was to assess the potential application of e-nose system for detection of honey adulterated by addition of sugar syrup.

Material and methods

All honey samples were collected from beekeepers of different provinces of southern region of Iran (Bushehr, Hormozgan, Khuzestan, Sistan and Baluchestan, Fars and Kerman), which are the main producers of Ziziphus honey. Adulterant solutions were produced by blending of the fructose-glucose mixtures 1:1 (w/w) were prepared by mixing fructose (HFCS 55, High fructose corn syrup) and glucose (DE 42, Dextrose equivalent) (Zar Fructose Company, Iran) completely. A set of honeys were subsampled and then adulterated with fructose-glucose adulterant solutions at four levels i.e. 20, 40, 60 and 80%. Finally, each group of samples was tested 20 times by the electronic nose.

The developed e-nose system consisted of a sample and sensor chamber, air circulation system, a set of
gas sensors a data collection system and suitable data preprocessing programs written in LabVIEW software (figure 1).

![Figure 1. The developed e-nose](image)

The basis of e-nose systems is based on gas sensors. Semiconductor sensors metal oxide semiconductor (MOS) due to its high chemical stability, high sensitivity and suitable for a wide range of food and agricultural products were used in the e-nose system.

For designing the sensor array, about 20 different MOS gas sensors were purchased and each of them was evaluated using different honey samples, and 8 sensors that showed a large difference among different honey samples were selected and placed inside the sensor chamber.

The sensor array consists of 8 different MOS gas sensors that consists of 6 MQ sensors fabricated in (HANWEI Electronics Co., Ltd., Henan, China) and 2 TGS sensors supplied by (Figaro Engineering, Inc., Osaka, Japan) (table 1). The circuit of both type of sensors are presented in Figure (2). According to the manufacturer's recommendations In order to ensure the correct functioning of the sensors, an hour before the experiments the sensors were turned on to achieve a steady state.

![Figure 2. The data sheet of gas sensor (a) MQ type (b) TGS type](image)

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Detect</th>
<th>Detection range (PPM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MQ-2</td>
<td>Methane, Butane, LPG, Smoke</td>
<td>200-5000</td>
</tr>
<tr>
<td>MQ-3</td>
<td>Alcohol, Ethanol, Smoke</td>
<td>0.05-10</td>
</tr>
<tr>
<td>MQ-4</td>
<td>Methane, CNG Gas</td>
<td>200-10000</td>
</tr>
<tr>
<td>MQ-5</td>
<td>Natural gas, LPG</td>
<td>200-10000</td>
</tr>
<tr>
<td>MQ-6</td>
<td>LPG, Butane gas</td>
<td>200-10000</td>
</tr>
<tr>
<td>MQ-136</td>
<td>Hydrogen Sulfide gas</td>
<td>1-200</td>
</tr>
<tr>
<td>TGS-2610</td>
<td>High sensitivity to LP and its component gases</td>
<td>500-10000</td>
</tr>
<tr>
<td>TGS-2620</td>
<td>For alcohol, toluene, xylene, other volatile organic vapors</td>
<td>500-5000</td>
</tr>
</tbody>
</table>

The schematic drawing of the olfactory machine system is shown in Figure (3), which shows how the system works. The measurement process in the experiments is divided into two phases: 1) saturation and measurement; and 2) purification.
About 250 g of sample was poured inside a sample chamber and its lid was sealed with a cork to prevent the outflow of gases from the honey space. Regarding the sampling step, the air flowed through the various paths controlled by a computer program. The sensor chamber is cleaned with dry air before the honey head space gasses were guided into the sensor chamber. At this stage, the pumps 1, 2 and 3 and the valves 1, 2 and 4 are switched on and the solenoid valve 3 is off. At this stage, the dry air passes through the sensor and sample chamber and cleans out the two compartments from any scent of aromas and eventually exhausts from the outlet. This takes about 600 s. In the next step, the pumps 2 and 3 are still on, but the pump 1 was turned off, as well as the electric switches 2 and 3 were switched on and the electric valves 1 and 4 were turned off. At this stage, the aroma of honey was saturated and circulates from the sensor chamber and the sensors are exposed to the honey head space gasses and the changes in the output of the sensors were stored. The time required to reach the steady state was 550 s, and this time was considered as the response time. After this phase, the sensor and sample chambers that were saturated with honey aroma were cleaned for 600 s using the dry air, to return the sensor response back to the baseline. The output of the sensors was collected using the Arduino Mega 2560 microcontroller, and the LabVIEW 2017 software was used to connect the microcontroller to the computer for data storage and preprocessing.

In this research, the Arduin mega 2560 was used to communicate between sensors and convert analogue data to digital output. The Mega 2560 Arduino board has a microcontroller board based on ATmega 2560. This board has a total of 54 digital PINs which fifteen of them can be used as pulse width modulation (PWM) outputs, as well as 16 analog inputs. The Mega 2560 board includes features that allow communication with the computer, other Arduino boards or other microcontrollers. The data collected from the sensors was sent to LabVIEW by the Arduino board. LabVIEW is a graphical programming language that is widely used for applications in military, educational, and laboratory industries as a standard model for data collection and analysis, as well as a tool for controlling and simulating systems. LabVIEW is a graphical programming language that is widely used for applications in military,
Utilizing Pattern Recognition Methods for Detecting the ... 423

Educational, and laboratory industries as a standard model for data collection and analysis, as well as a tool for controlling and simulating systems. This program has two basic parts, first part is the software interface and operator, and the other part is the block diagram, which is the location of the codes and symbols that is the environment of programming. Figure (4) shows the codes written in the LabVIEW software for communicating the information between the e-nose and computer for analysis.

**Figure 4.** Codes written in the LabVIEW software

Signal preprocessing is used for extraction of relevant data from the obtained responses and also for preparation of the data for multivariate pattern analysis. The major aspects of this preprocessing are: (a) baseline identification and manipulation/determination, (b) compression, (c) normalization (Pearce et al., 2003).

\[ y_s(t) = \frac{x_s(t) - x_s(0)}{x_s(0)} \]  

(1)

Differential, relative and fractional techniques are 3 different techniques for baseline manipulation. Fractional method is also widely used for MOS chemo-resistors. In this method, the baseline \( x_i(0) \) is first subtracted from the sensor response \( x_i(t) \) and then divided into the baseline. Fractional measurements are not only dimensionless but also normalized since the resulting response \( y_s(t) \) is a per-unit change with respect to the baseline, which compensates for sensors that have intrinsically large (or small) response levels. Fractional method was used in the current study (Sanaeifar et al., 2016).

Compression is a preprocessing stage in which the response of each sensor array is utilized as a feature vector or a fingerprint by decreasing the number of descriptors. In this study, the maximum response value for each sensor was individually extracted and analyzed. Normalization is the final stage of preprocessing which is applied to operate on the sensors signals to compensate for sample-to-sample variations due to the change in analytic concentration and drift in the sensors. On the other hand, normalization operates across the entire database for a single sensor (e.g., the complete history of each sensor), and is generally employed to compensate for differences in sensor scaling. In what follows, we
will denote it by $x^{(k)}_s$ which is the response of sensor “s” to the k-th example in the database (Eq. 2). In Sensor normalization, the range of values for each individual sensor is set to $[1, -1]$ (Sanacifar et al., 2016). Figure (5) shows the sensor responses for honey sample.

$$y^{(k)}_s = \frac{x^{(k)}_s - \text{min}_{x^{(k)}_s}[x^{(k)}_s]}{\text{max}_{x^{(k)}_s}[x^{(k)}_s] - \text{min}_{x^{(k)}_s}[x^{(k)}_s]}$$

After preprocessing the data principle component analysis (PCA), hierarchical clustering analysis (HCA) and linear discriminant analysis (LDA) studied in order to create the models and to discriminate between different honey samples through the sensor array responses of e-nose.

PCA is a statistical technique that is used for extracting information from a multivariate dataset. It transforms the original, measured variables into new uncorrelated variables called principal components (PCs), which retain the information present in the original data as much as possible. In this work, the PCs were selected to contain the maximum variance in the sensor array signals and subsequently as a sensor selection method. The first PC is oriented to explain as much variance in the data as possible and the second PC explains the next largest variance in the data (Kiani et al., 2016).

The HCA is one of the well-known unsupervised clustering methods and providing a succinct graphical representation of how well each object lies within its cluster, known as a dendrogram (Kiani et al., 2016).

LDA is one of the most widely used classification procedure. LDA, as a supervised method, has been used for feature extraction and variable selection in a dataset (Tudu et al., 2008). In this research, the past V3 software was used to analyze PCA and LDA also, R software was used for HCA analysis.
LDA is a technique that directly maximizes class separately; generating projections where the examples of each class, form compact clusters and the different clusters are far from each other (Sanacifar et al., 2016).

**Result and discussion**

**PCA test results**

PCA score plot was utilized to investigate clustering of data points within the multi-dimensional space of different samples. The level of adulteration was evaluated in 5 levels of 0, 20, 40, 60 and 80%. In the statistical analysis, 20 replicates were used for each sample using the e-nose system and a total of 100 samples were evaluated for evaluation. The first and second PCs (PC1-PC2) account for 98.4% of data variance for the original dataset (PC1=82.36% and PC2=16.04%) which means that the differences existing among honey samples along the first two axes are more significant. The PCA results are illustrated in Figure (6). That shows the clear discrimination between the samples except between pure honey and 20% syrup because of their similar aroma strengths.

**HCA results**

To check the PCA results, the HCA was applied to analyze the preprocessed data. The results of analysis can be displayed graphically using a tree diagram, known as a dendrogram, which shows all the steps in the hierarchical procedure in this method, all classes are initially connected and then categorized according to the distance between the data. Figure (7) shows the dendrogram of honey samples that were successfully discriminated based on their aroma strength. As depicted in Figure (7), HCA can successfully discriminate honey samples in to 5 groups as it can be seen the pure honey, 20 and 40% syrup are classified in first branch and 60 and 80% syrup classified in second branches based on their aroma strength. By comparing the PCA and HCA results it can be concluded that pure honey and 20% syrup are more similar and from the Figure (7) it can be seen that one of the pure honey samples miss classified in to 20% syrup branches that indicates these two classes have the close aroma strength and the close aromatic quality. All the honey samples were discriminated based on their aroma with 99% accuracy.
LDA analysis results
The response of the sensors after applying the fractional preprocessing method was considered as input in the linear discrimination analysis method. LDA analysis results are shown in Figure (8). This Figure (8) represents analysis results on a two-dimensional plane, linear discriminant 1 (LD1) and linear discriminant 2 (LD2). The results obtained by LDA plot, provided a perfect classification. The first two LDs (LD1–LD2) account for 95.75% of data variance for the original dataset (LD1=74.54% and LD2=21.21%) which means that the differences existing among honey samples along the first two axes are more significant.

As the Figure (8) shows, all the honey samples are completely separated and as expected from the PCA and HCA results the pure honey and 20% syrup groups are very close to each other. The confusion matrix of LDA is given in Table (2). As shown in Table (2), in the LDA method, the 100% discrimination between pure honey and adulterated honey is evident. Yu et al. (2007) examined the ability of the e-nose to detect milk adulteration. They found that milk adulteration using linear discriminant analysis (LDA) were better than principle component analysis (PCA).
Table 2. Confusion matrix of LDA analysis

<table>
<thead>
<tr>
<th>Samples</th>
<th>Pure honey</th>
<th>20% syrup</th>
<th>40% syrup</th>
<th>60% syrup</th>
<th>80% syrup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pure honey</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20% syrup</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>40% syrup</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>60% syrup</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>80% syrup</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
</tbody>
</table>

Discrimination accuracy % | 100 | 100 | 100 | 100 | 100 |

Conclusion

In this research, the portable e-nose system was built on the basis of metal oxide semiconductor (MOS) sensors. The e-nose, in contrast to conventional technologies, such as high-performance liquid chromatography (HPLC) gas chromatography (GC), which determines the aroma characteristics and components of each substance, do not have the problem of high cost, the need for an expert operator for running the systems and the laborious preparation of samples. Based on the results, the presented e-nose is a reliable tool for recording changes between different levels of adulterated honey and it’s easy to use. PCA, HCA and LDA methods were evaluated to evaluate the ability of the e-nose to detect levels of adulteration in honey. The results included 98.4% of the variance for the PCA method and 99% accuracy classification for the HCA method and 100% classification accuracy for LDA analysis. Also, from all the analytical methods employed in this study, it can be concluded that by increasing the amount of adulteration, the similarity of the pure honey with the adulterated honeys is reduced that is evidence of the ability of the e-nose to distinguish between different honey and the ability to detect adulteration in honey with high precision.

References


استفاده از روش‌های تشخیص الگو در آنالیز داده‌های حسگر بویایی جهت تحلیل تقلب گل‌کوز و فروکتوز در عسل

سعید فعال ۱، محمد لغوی ۲، سعادت کامگار ۳، محمدحسن روفت ۴، محمدتقی گلکنانی ۵

۱- استادیار، بخش مهندسی بیوشیمی، دانشکده کشاورزی، دانشگاه شیراز، شیراز، ایران (kamgar@shirazu.ac.ir)
۲- استاد، بخش مهندسی بیوشیمی، دانشکده کشاورزی، دانشگاه شیراز، شیراز، ایران
۳- دانشیار، بخش مهندسی علوم و صنایع غذایی، دانشکده کشاورزی، دانشگاه شیراز، شیراز، ایران
۴- دانشیار، بخش مهندسی علوم و صنایع غذایی، دانشکده کشاورزی، دانشگاه شیراز، شیراز، ایران
۵- استاد وزیر، بخش مهندسی علوم و صنایع غذایی، دانشکده کشاورزی، دانشگاه شیراز، شیراز، ایران

چکیده

راهنمای عمل‌کننده از پارامترهای مهم در طبقه‌بندی عمل به‌حساب می‌آید و بی‌آن از یکی به گل‌های مختلف، موقعيت جغرافیایی و ترکیبات تشکیل دهنده عمل می‌توانند افزایش یابد. هدف از این تحقیق توصیف و ارزیابی یک سامانه بویایی روشن آمیزی این منطوق ارتباط بین گازهای سر فضای عمل‌های با درصد‌های مختلف تقلب (خالص)، ۲۰ درصد شیرین، ۶۰ درصد شیرین و ۸۰ درصد شیرین) مورد ارزیابی قرار گرفت. سامانه بویایی مدل‌شده شامل ۸ عدد حسگر نیمه‌هادی اکسید فلزی برای گزارش اطلاعات موجود در گازهای فضائی قرار داشت. بعد از بردارش داده‌های مورد ارزیابی مدل‌های تشخیص الگو جهت شناسایی تقلب مورد استفاده قرار گرفتند. اجزای اصلی سامانه تحلیلی حاوی داده آلگوریتم‌های تشخیص الگو و (LDA) تحلیل داده می‌باشد. تحلیل مؤلفه‌های اصلی (PCA)، تجزیه خوشه‌ای (HCA)، تکنیک کندنده خطی (KNN) و روش‌های دقت تقلب شامل ۹۸/۸ درصد واریانس به روش HCA و ۹۹ درصد دقت طبقه‌بندی به روش LDA. با توجه به نتایج بدست‌آمده، طبقه‌بندی به روش LDA بود. نتایج نشان داد سامانه بویایی خودکار و قابل اطمینان در تشخیص تقلب عمل است.

واژه‌کلیدی: تشخیص الگو، تقلب، حسگر، عمل، ماهیت بویایی