HUMAN ODOUR DETECTION APPROACH USING MACHINE LEARNING

AHMED QUSAY SABRI

THESIS SUBMITTED IN FULLFILLMENT FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF COMPUTING AND INFORMATICS
UNIVERSITI MALAYSIA SABAH
2019

UMS
UNIVERSITI MALAYSIA SABAH

UNIVERSITI MALAYSIA SABAH

BORANG PENGESAHAN STATUS TESIS

JUDUL: **HUMAN ODOUR DETECTION APPROACH USING MACHINE LEARNING**

IJAZAH: **DOKTOR FALSAFAH (SAINS KOMPUTER)**

Saya **AHMED QUSAY SABRI**, Sesi **2014-2019**, mengaku membenarkan tesis Doktoral ini disimpan di Perpustakaan Universiti Malaysia Sabah dengan syarat-syarat kegunaan seperti berikut:-

1. Tesis ini adalah hak milik Universiti Malaysia Sabah
2. Perpustakaan Universiti Malaysia Sabah dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (/):

☐	SULIT	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA 1972)
☐	TERHAD	(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)
☐	TIDAK TERHAD	

Disahkan Oleh,

AHMED QUSAY SABRI
DI1411003A

NURAZLYNNE MOHD. JOHAN @ JACKLYNE
PUSTAKAWAN
UNIVERSITI MALAYSIA SABAH
(Pendaftaran Pustakawan)

Tarikh : 30hb September 2019

(Prof. Madya Dr. Rayner Alfred)
Penyelia Utama

( )
Penyelia Bersama

DECLARATION

I hereby declare that the material in this thesis is my own except for quotations, excerpts, equations, summaries and references, which have been dully acknowledged.

30th September 2019

Ahmed Qusay Sabri
DI1411003A

CERTIFICATION

NAME	:	**AHMED QUSAY SABRI**
MATRIC NO.	:	**DI1411003A**
TITLE	:	**HUMAN ODOUR DETECTION AND RECOGNITION USING MACHINE LEARNING**
DEGREE	:	**DOCTOR OF PHILOSOPHY (COMPUTER SCIENCE)**
VIVA DATE	:	**17 MAY 2019**

CERTIFIED BY:

Signature

SUPERVISION COMMITTEE

CHAIRPERSON

ASSOC. PROF. DR. RAYNER ALFRED _____

COMMITTEE MEMBERS

1. ASSOC. PROF. TS. DR. MOHD HANAFI AHMAD HIJAZI _____
2. DR. LAU HUI KENG _____

ACKNOWLEDGMENT

"تِلْكَ الدَّارُ الْآخِرَةُ نَجْعَلُهَا لِلَّذِينَ لَا يُرِيدُونَ عُلُوًّا فِي الْأَرْضِ وَلَا فَسَادًا وَالْعَاقِبَةُ لِلْمُتَّقِينَ" (83 القصص)

بسم الله والصلاة والسلام على سيدنا محمد , هذا عمل خالص لوجه الله , ارجوا فيه مرضاة الله
اهداء رسالة دكتوراة في الذكاء الاصطناعي
اهدي هذا العمل الى سيدي رسول الله محمد صل الله عليه وسلم , جعلني الله من اتباعه في الدنيا و الاخرة وذريتي

الى ابي رحمه الله العميد الركن قصي صبري عبدالقادر
الى جدتي رحمها الله نعمت محمود خليل
الى امي حفظها الله ربلب ابراهيم محمود
الى امي (خالتي) حفظها الله رحاب ابراهيم محمود
الى ام ندى
الى ندى احمد قصي
الى داليا احمد قصي
الى ام زهراء
الى زينب قصي صبري و احمد فريد عبدالحميد
الى عمر قصي صبري
الى عدي قصي صبري
الى نعمت قصي صبري و عدي حاتم هاشم
الى زبيدة قصي صبري
الى العميد الركن منذر ابراهيم (رحمه الله) , حارث ابراهيم محمود , اللواء الدكتور الركن محمد ابراهيم محمود , محد ابراهيم محمود
الى محمد قحطان جميل و احمد قحطان جميل و عدي محمود عودة
الى شريفة كرم

I would like to express my deepest appreciation to my supervisor, Dr. Rayner Alfred for all his advices, guidance and support in this interesting research work that lead to the completion of this thesis.
Special thanks and gratitude to UMS (University Malaysia Saba).
Special thanks and gratitude to Kota Kinabalu.
Special thanks and gratitude to MALAYSIA.

Ahmed Qusay Sabri
30th September 2019

ABSTRACT

Recognizing the human considered as old and contemporary task. This problem is now solved by using biometrics. Technically biometrics is " the automated technique for measuring an individual's physical or personal characteristic and comparing it to a comprehensive database for identification purposes". This thesis presents a problem with the selection of appropriate human (Volatile Organic Compounds) VOCs emitted from sweat for human odour classification, all gasses emitted by humans through sweat have been collected and detected using the latest technology (High Resolution GCMS / TOF) Gas Chromatograph Mass Spectrometry/Time of Flight. Different people (15 people) with different ages and genders have been tested, some people have been tested several times. There is a total of 198 VOCs detected and methods for selecting features are used to determine which VOCs are suitable for classifying human odour. Two feature selection methods Entropy and Chi Square tests were used to identify and determine the best and most acceptable VOCs. There is a total of 16 stable VOCs extracted from 198 VOCs on the basis of the results obtained. In addition, 10 gasses are detected with zero values for both the entropy and the chi- square test, and these gasses are the strongest candidates to detect and classify odours. The results of this work can be used to classify specific VOCs for the detection of humans by odour. In this thesis, a framework for gender recognition is proposed based on human odour. 20 samples of human odour from male and female are collected, several different activation functions of the neural network (e.g., backpropagation of Levenberg-Marquardt, backpropagation of gradient descent and resilient backpropagation) and several different topologies of the neural network are tested. It is also found that with 2 hidden layers with more neurons in the hidden layers (16 and 16 neurons in which the hidden layer is) Levenberg-Marquardt was able to achieve a higher performance accuracy of 100%. The main investigations conducted in this thesis which is Human Identification from body odour followed by an investigation to prove stability and rigidity of person identification main findings. A framework for human identification is proposed distinctively based on specific human odour features. 15 samples of female and male human odour are collected from different age groups, several diverse functions of neural network activation are tested such as Gradient descent backpropagation, Levenberg-Marquardt backpropagation, and Resilient backpropagation. Besides, numerous neural network topologies are tested by means of a selection of number of neurons and hidden layers. Different activation functions were tested TAN-Sigmoid transfer, Linear transfer, and LOG-Sigmoid transfer. Considering the obtained results, employing two hidden layers with more neurons in the hidden layers- to be specific: 15 neurons in every layer- has yielded better accuracy in performance with an accuracy rate of 100%. The unsurpassed framework for learning algorithm to be used for human identification is Levenberg-Marquardt backpropagation learning algorithm. The best function for activation established in this research is the function of TAN- Sigmoid transfer. Finally, we investigate the effects of missing gases in human odour sample to evaluate the accuracy of classifying individual person. These missing values will be replaced by Random number between 0 and 1 as our research prove, the best accuracy result when missing values are introduced in the odour dataset is the Ensemble Bagged Trees.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

MLP	-	Multi-Layer Perceptron
ANN	-	Artificial Neural Networks
BP	-	Back-propagation network
GC	-	Gas Chromatography Devise
VOC	-	Volatile Organic Compounds
ANN	-	Artificial Neural Network
E-nose	-	Electronic Nose
MHC	-	major histocompatibility complex
GC–MS	-	gas chromatograph–mass spectrometry
SPME	-	solid phase microextraction
LOD	-	limit of detection
ASOB1	-	apocrine secretion odour-binding proteins
GCMS/TOF	-	Gas Chromatograph Mass Spectrometry/Time of Flight
SVM	-	Support Vector Machine
KNN	-	k-nearest neighbor's algorithm
LM	-	Levenberg Marquardt
BPNN	-	Backpropagation Neural Network
PSA	-	Prostate-specific antigen
df	-	Degree Of Freedom

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background

The task of human recognition is old and present at the same time. Biometrics is now solved. From a technical point of view, biometrics is "the automated technique for measuring or comparing an individual's physical or personal characteristics to a comprehensive identification database". (Korotkaya)

Physical or personal biometrics including eye features (iris, retina), facial features, hand geometry, ear shape, fingerprints, wrist / hand veins, chemical composition of DNA and body odour. Personal characteristics include manual signature, keystroke / type patterns and voiceprint. All these physical and personal characteristics are measured and integrated in the human recognition computer system. Therefore, biometrics are used for two key purposes: identification and authentication.

Identification: The Biometrics Glossary states that identification, the primary purpose of biometrics, is "the one to many comparisons of a single human biometric sample with the entire database of biometric templates. It enables us to determine whether the sample corresponds to any of the templates and, if so, to identify the enrollee whose template has been matched."

Personal Authentication: The second objective of the biometric technique is authentication. It is the verification of information like identity, property or authorization. The answer to the authentication

question is, "Are you the one you claim to be?" in order to achieve this purpose. For this purpose, one - to - one comparison is used (Oyeleye et al., 2012).

Biometric characteristics: This new area of science and technology is born by the growing need for privacy and security in our everyday lives. Biometrics covers a set of approaches for the unique recognition of people based on one or more intrinsic physical or behavioral characteristics. Biometrics is used in computer science as a way to manage and control identity access. It is also used to identify persons in groups under supervision.

Physiological characteristics related to the body's shape include but not limited to fingerprint, face detection, DNA, palm print, hand geometry, iris detection and odour. The behavioral characteristics associated with a person's behavior include but are not limited to rhythm, gait and voice typing. Biometrics works by unobtrusively matching live person patterns in real time against registered records. Biometric technology leading examples can recognize and authenticate faces, hands, fingers, signatures, irises, voices and fingerprints.

Research on hand-shape biometric system based on a novel feature extraction methodology using the morphological pattern spectrum or pecstrum. Identification experiments were carried out using the obtained feature vectors as an input to some recognition systems using neural networks and support vector machine (SVM) techniques, obtaining in average an identification of 98.5%. The verification case was analyzed through an Euclidean distance classifier, obtaining the acceptance rate. (Ramirez-Cortes et al. 2011)

There are a wide range of odourants to be identified in many sectors, such as food, environmental monitoring and medicine. For substances as automatic detection systems artificial nose systems are therefore currently being developed. Such devices are extremely important in applications such as inspection of food quality, process control, detection of environmental gas

leaks, medical diagnosis and much more (Omato, 2013). The sensor system and the pattern recognition system are two main components of an artificial nose. Each odour creates a characteristic odour pattern in the sensor system. This pattern is prepared and given to the system of pattern recognition that classifies the material (Ludermir and Yamazaki, 2003). Figure 1.1 below show variety of biometric characteristics.

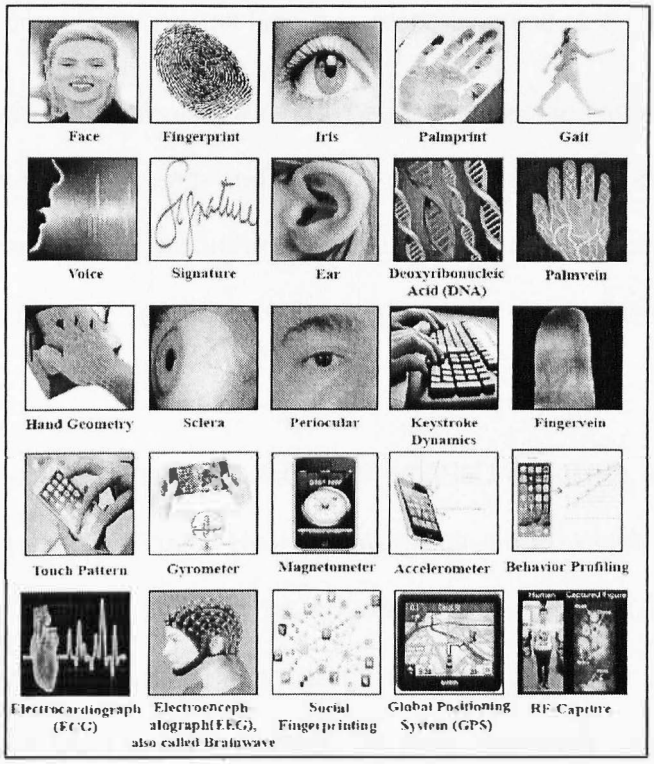

Figure 1. 1 Various Biometric Characteristics

A wide range of high-sensitivity odourants can be detected in the mammalian nose and substances recognized by combination and relative proportions of compounds. This can be done by combining a set of sensing elements with overlapping large selectivity profiles. In artificial nose systems, therefore, the sensor systems consist of elements with large and overlapping sensitivities. For this purpose, polymer sensors are widely used. This sensor type is based on the

conductivity variation of polymers exposed to volatile gases (Ludermir and Yamazaki, 2003) (He et al. 2008).

Fingerprint classification is an effective technique for reducing the candidate numbers of fingerprints in the stage of matching in automatic fingerprint identification system (AFIS). In recent years, deep learning is an emerging technology which has achieved great success in many fields, such as image processing, natural language processing and so on. (Wang et al. 2014)

Fingerprint classification adopting a neural network as decision stage. The neural network is ready to perform matching process and is successfully developed to identify and classify the fingerprint using backpropagation algorithm. The experimental results show the method proposed could improve fingerprint image quality classification accuracy more effectively than others. (Pati and Suralkar, 2012)

Pattern recognition is an important subject of artificial intelligence, also a primary field for the application of ANN (LI et al., 2006). Neural networks have been widely used in artificial noses as pattern recognition systems. Some of the advantages of this approach are: (1) the ability to handle non-linear sensor signals; (2) adaptability; (3) fault and noise tolerance; and (4) the inherent parallelism that leads to fast operation. The most commonly used artificial noses odour classification network was the Multi-Layer Perceptron(MLP) and the back-propagation learning algorithm Neural Networks(BPNN), which analyzes complex data and recognizes patterns. ANN is an important method for the recognition of electronic nose patterns.

The ANN 's most useful pattern is the backpropagation network (BP) or its modified form. Although the measurement speed and accuracy of the electronic nose system and some improved algorithms in the BP neural network have significantly improved, it is still difficult to sufficiently simplify the network architecture (He et al., 2008).

Investigation possibilities of incorporating artificial neural networks into fingerprint recognition process, implemented and documented our own software solution for fingerprint identification based on neural networks whose impact on feature extraction accuracy and overall recognition rate was evaluated. The result of this research is a fully functional software system for fingerprint recognition that consists of fingerprint sensing module using high resolution sensor, image enhancement module responsible for image quality restoration. (Marak and Hambalik, 2016)

Human scent evidence can be of a crucial use in many cases where other types of evidence such as DNA, fingerprints, or fibers are not available. The use of canines within the area of law enforcement scent detection and forensic science often involves locating general human scent but this is increasingly related to their ability to correctly match a human scent sample with its originating subject. Various challenges that have surfaced in courts of law across the United States relate to the validity and efficiency of this technique that employ biological detectors to alert to a specific human scent. Additional concerns correspond to the actual human scent composition and its ability to transfer and be detected by the canine at different locations. There is little information as to the specific odour signatures the canine is alerting to when it makes a positive scent match with a subject. Until 2008, studies of human odour have been focused mainly on the composition of human sweat or body malodours rather than the total volatile odour chemicals that emanate from humans. The use of human scent discriminating canines for the purposes of criminal investigations lies in the idea that human odour is a unique physical characteristic of every individual and that this odour is left at every location, object, or path which the subject has come in contact with. The definition of human odour, however, cannot be limited to one factor and thus can be attributed to skin oils, sweat and volatile compounds emanating from the skin surface. In turn, some of these natural body processes can be affected by heredity, environment, and

daily lifestyle activities which allow an individual to produce a characteristic odour Figure 1.2. (Prada and Furton, 2008).

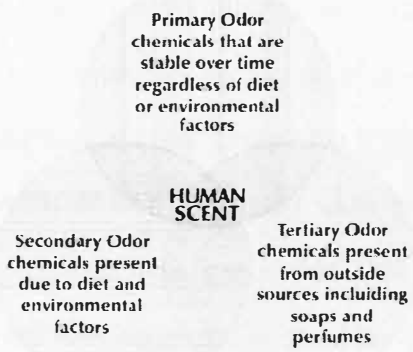

Figure 1.2 Human Odour Definition

The biological function of body odour production relies on the three types of secretory glands in the human skin. Two of these are normally called the "sweat glands" which are the eccrine and apocrine glands, with the third being the sebaceous glands. The aqueous portion of skin secretions originates mostly from the eccrine sweat glands which entirely consist of water along with dilute salts. The sebaceous glands are closely connected to hair follicles and continuously secrete oils, or sebum. Sebaceous glands are found throughout the body, but have a higher concentration in the face and scalp. The apocrine gland is located primarily in the axillary and genital regions. Much of the work performed for the elucidation of human odour composition has been focused on the axillary region of the human body which has helped to identify the chemical composition of sweat and has provided some insight into the understanding of this biological fluid. A number of studies have demonstrated that the characteristic human axillary odours branched and unsaturated aliphatic acids, alcohol, carbonyls and some steroids as major contributors to underarm malodour, with a major contribution originating from (E)-3- methyl-2-

hexenoic acid. Human axillary compounds may not accurately reflect what a subject is leaving behind at a crime scene. Therefore, the true composition of body odour may not be entirely covered simply by studying human sweat. Other research in the field has included analysis of human skin emanations by solid phase microextraction/ gas chromatography-mass spectrometry (SPME/ GC-MS). These studies have revealed key volatile organic compounds which may play a key role in determining the characteristic odour originating from an individual. A wide variety of compounds were found, the main of which included short chain aldehydes and long chain hydrocarbons. Another result obtained was the notion that the levels of these common compounds varied among different individuals. Yet another study performed also highlighted the efficacy of SPME-GC/MS to study the chemical components of human odour. The results of this technique yielded the identification of functional groups such as aldehydes, alcohols, alkanes, and esters in individual armpit samples. The common compounds were shown to be present in a differing ratio pattern between the males and females in the study thus portraying qualitative similarities but with quantitative differences. The skin in general has been described as a continuous source of "rafts" that are dead skin cells and are deposited in the external environment as a person touches objects in the environment. A common description of this raft theory essentially adds to the composition of the human odour by the contribution of these rafts to the air currents after an individual meets a certain place or object of interest. (Prada and Furton, 2008).

A method for improving the capability of odor classification in dynamical change of concentration often encountered in the ambient air. Our method employs a short-time Fourier transform (STFT) algorithm and a stepwise discriminant analysis for feature extraction and dimensional reduction. Finally, using learning vector quantization (LVQ) method to evaluate the classification performance, we successfully achieved high classification rate even if the odor

concentration changes irregularly at different humidity levels whereas the classification rate was insufficient in the case of using only magnitudes of sensor responses. (Nimsuk and Nakamoto, 2007)

A research present invention directed to methods of identifying and/or comparing humans. More particularly, the present invention is directed to methods of collecting human scent compounds from a subject, extracting the compounds, analyzing the compounds, and correlating the compounds to a unique compound profile for the subject. These unique compound profile can be used to distinguish one subject from another, or to identify a specific subject based upon a sample. (Furton and Curran, 2006)

The human body produces VOCs that can be utilized to develop relevant information about the behavior, diseases, health status, and emotional state of a person. The odour of body is one of the human physical features that can be employed to identify different people. There are certain researches piloted to distinguish and categorize humans according to body odour that is extracted from persons' armpits. An electronic nose, for instance, was invented with a sensing array and a system for data analysis. Body odour recognition is viewed as a method of a contactless physical biometric which attempts to confirm the identity of a person. The body odour is extracted from several parts of the body existing in many forms such as armpits, exhalation, stools, farts, urine, or feet. Armpits can be one of the top sources for sampling some volatile chemicals that are released from human body. Armpit is indeed a body part where a huge number of bacteria and glands cooperate to release a strong scent which may produce a unique pattern that can assist in identifying different persons.

Biometrics permits an individual to be authenticated and identified by computer systems following a set of verifiable and identifiable data that are precise and unique in nature. This

mechanism constitutes a cutting-edge method of identifying an individual since it precisely establishes more explicit and direct connection with humans than mere passwords since biometrics use measurable behavioral and physiological features of people.

Main assumption of or research is to prove valid human detection from odour, but the restriction and constraint of this assumption is to find approximately constant list of gases to be the seed of human detection.

The irrelevant features can obviously often lead to poor modeling, as they are not well connected to the class label. Several feature extraction methods have been proposed to generate patterns from time series data for classification purposes (Bahrampour et al., 2013). In fact, these features typically increase the accuracy of classification due to over-fitting when the training data set is small and these features can be included in the model.

Automatic face identification and verification from facial images attain good accuracy with large sets of training data while face attribute recognition from facial images still remain challengeable. Hence introducing an efficient and accurate facial image classification based on facial attributes is an important task. A methodology for automatic age and gender classification based on feature extraction from facial images. In contrast to the other mechanisms proposed in the literature, this methodology main concern on the biometric feature variation of male and females for the classification. (Kalansuriya and Dharmaratne, 2014). Gender classification using artificial neural networks is detected and applied. (Roy and Bandyopadhyay, 2014)

1.2 Problem Statement

Based on the literature review, there are limited researches that have been conducted to investigate the effectiveness of detecting and recognizing human subjects based on odour. There