FACIAL FEATURES DETECTION SYSTEM FOR CLASSIFYING CRITICAL SITUATION OF BRAI
N SURGERY POSTOPERATIVE PATIENTS

Mr. Bhatarasiri Slakkham

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science and Information Technology
Department of Mathematics and Computer Science
FACULTY OF SCIENCE
Chulalongkorn University
Academic Year 2019

ระบบการตรวจจับลักษณะของใบหน้าเพื่อการจำแนกสถานะวิกฤติของผู้ป่วยหลังการผ่าตัดสมอง

นายภัทรศิริ สลักคำ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2562

Thesis Title	FACIAL FEATURES DETECTION SYSTEM FOR CLASSIFYING CRITICAL SITUATION OF BRAIN SURGERY POSTOPERATIVE PATIENTS
By	Mr. Bhatarasiri Slakkham
Field of Study	Computer Science and Information Technology
Thesis Advisor	Associate Professor PATTARASINEE BHATTARAKOSOL, Ph.D.
Thesis Co Advisor	Associate Professor Aurauma Chutinet

Accepted by the FACULTY OF SCIENCE, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Science

.... Dean of the FACULTY OF SCIENCE
(Professor Polkit Sangvanich, Ph.D.)

THESIS COMMITTEE

.... Chairman
(Assistant Professor Monnat Pongpanich, Ph.D.)

.... Thesis Advisor
(Associate Professor PATTARASINEE BHATTARAKOSOL, Ph.D.)

.... Thesis Co-Advisor
(Associate Professor Aurauma Chutinet)

.... External Examiner
(Associate Professor Panjai Tantatsanawong, Ph.D.)

ภัทรศิริ สลักคำ : ระบบการตรวจจับลักษณะของใบหน้าเพื่อการจำแนกสถานะวิกฤติของผู้ป่วยหลังการผ่าตัดสมอง. (FACIAL FEATURES DETECTION SYSTEM FOR CLASSIFYING CRITICAL SITUATI ON OF BRAIN SURGERY POSTOPERATIVE PATIENTS) อ.ที่ปรึกษาหลัก : รศ. ดร. ภัทรสินี ภัทรโกศล, อ.ที่ปรึกษาร่วม : รศ. พญ.อรอุมา ชุติเนตร

งานวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อสร้างโมเดลที่ใช้สำหรับตรวจจับสัญญาณขอความช่วยเหลือจากคนไข้โรคหลอดเลือดสมองหลังได้รับการผ่าตัด ผ่านเทคโนโลยีปัญญาประดิษฐ์โดยตรวจจับผ่านรูปภาพใบหน้าของผู้ป่วย เพื่อช่วยลดภาระการทำงานและลดค่าใช้จ่ายของโรงพยาบาลลง การเก็บข้อมูลได้อาสาสมัครเป็นกลุ่มตัวอย่างจาก ศูนย์หลอดเลือดสมองแบบครบวงจร โรงพยาบาลจุฬาลงกรณ์ สภากาชาดไทย ภายในห้องผู้ป่วยวิกฤต จำนวนทั้งสิ้น 8 ราย โดยทั้ง 8 รายนั้นได้รับการคัดกรองจากแพทย์เจ้าของไข้และกรอกเอกสารความยินยอมก่อนได้รับอนุญาตให้เข้ารับการเก็บข้อมูล ซึ่งวิธีในการเก็บข้อมูลนั้นทำได้โดยการติดกล้องขนาดเล็ก 3 ตัว ทั้งด้านซ้าย ขวา และด้านบนของเตียงคนไข้ เพื่อให้ได้ภาพใบหน้าตรงของผู้ป่วยในทุกมุมมอง ระยะเวลาในการเก็บภาพจะอยู่ประมาณ 5-7 วัน หลังจากได้ภาพของคนไข้มาแล้ว จะนำข้อมูลเข้าสู่ขั้นตอนเตรียมข้อมูลโดยการเลือกรูปภาพของผู้ป่วยที่ต้องการความช่วยเหลือจากลักษณะต่าง ๆ บนใบหน้า การสร้างโมเดลนั้นได้เลือก 2 โมเดล เพื่อใช้ในการเปรียบเทียบประสิทธิภาพคือ โมเดลการถดถอยโลจิสติกส์ (Logistic Regression Model) และโมเดลโครงข่ายประสาทเทียม (Neural Network Model) ในส่วนการวัดประสิทธิภาพนั้นจะใช้เมตริกซ์วัดประสิทธิภาพ (Confusion Matrix) เพื่อเปรียบเทียบโมเดลทั้ง 2 และงานวิจัยอื่นที่เป็นที่รู้จักกันอย่างดี ผลการทดลองแสดงให้เห็นว่า โครงข่ายประสาทเทียมที่นำเสนอมีความถูกต้องในการทำนาย 92% ค่าความแม่นยำและค่าความถูกต้องมีค่าเท่ากันคือ 92%

สาขาวิชา	วิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ	ลายมือชื่อนิสิต ....
ปีการศึกษา	2562	ลายมือชื่อ อ.ที่ปรึกษาหลัก ....
		ลายมือชื่อ อ.ที่ปรึกษาร่วม ....

6072620623 : MAJOR COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
KEYWORD: Logistic Regression, Neural Network, Face recognition

Bhatarasiri Slakkham :
FACIAL FEATURES DETECTION SYSTEM FOR CLASSIFYING CRITICAL SITUATIO N OF BRAIN SURGERY POSTOPERATIVE PATIENTS. Advisor: Assoc. Prof. PATTARASINEE BHATTARAKOSOL, Ph.D. Co-advisor: Assoc. Prof. Aurauma Chutinet

The purpose of this research is to create a model for detecting a help signal from stroke patients after surgery. Through artificial intelligence technology by detecting via a patient's face image to help reducing workload and hospital's expenses. The data is collected from the Stroke Unit, King Chulalongkorn Memorial Hospital at the intensive care unit (ICU). The total number of samples is 8 persons, all these 8 were screened by the doctor and completed the consent forms before data being collected. The method of data collection is performed by attaching 3 small cameras on three positions: the left, right and top of the patient bed, to obtain a straight face image of the patient from every angle. The recording time is approximately 5-7 days after that the patient's image will bring into the data preparation step by selecting pictures of the patient who needs help from various facial features. There are two models to be chosen for comparison of efficiency: Logistic Regression Model, and Neural Network Model. The performance measurement, the Confusion Matrix was used to compare both models, including others well-known research. The result shows that the proposed neural network model has accuracy in prediction is 92%, the precision and recall are both equal to 92%.

Field of Study:	Computer Science and Information Technology	Student's Signature ....
Academic Year:	2019	Advisor's Signature ....
		Co-advisor's Signature ....

ACKNOWLEDGEMENTS

The research was supported by The Scholarship from the Graduate School, Chulalongkorn University to commemorate The Celebrations on the Auspicious Occasion of Her Royal Highness Princess Maha Chakri Sirindhorn's 5th Cycle (60th) Birthday. This work can be accomplished well because of the courtesy and assistance from many faculties and parties.

I would like to give my gratitude to Associate Professor PATTARASINEE BHATTARAKOSOL, Ph.D. and Assistant Professor Aurauma Chutinet for the honor of being the advisor and assistant advisor for the thesis. Please kindly transfer your knowledge, advice and suggestions which are very useful for the thesis. They have always been providing encouragement and attention since the beginning of the thesis until the project is completed accomplished well.

I would like to sincerely thanks to my program chair, Assistant Professor Monnat Pongpanich, Ph.D. and external examiner Associate Professor Panjai Tantatsanawong, Ph.D. for providing suggestions and suggestions for creating guidelines for development. This thesis will be much complete.

I would like to show my grateful thanks to all volunteers for their cooperation in collecting data, including all staff members in the stroke unit, both doctors and nurses who all help and advice on data collection.

I would like to express my gratitude to my parents who always take care, support, and encourage me when I had faced with hard problems and obstacles.

Finally, I would like to thank the great kindness from all the people mentioned above, including other people whom are not mentioned here, again, for any assistances in which this thesis is well accomplished.

Bhatarasiri Slakkham

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The components in this chapter will be divided into 5 sections. Section 1.1 will explain the causes and motivation of the research. Section 1.2 Define the objectives. Section 1.3 Define the scope and limitations of the research. Section 1.4 Expected results. The final section, section 1.5, describes the thesis structure.

1.1 Background and Importance

Currently, there are various sicknesses that cause patients to be bedridden and need intensive care from medical staffs. Unfortunately, the number of medical staffs or caregivers in the hospitals are not fit to the number of the existing patients. Moreover, different illnesses also require different treatments and caring. For example, the stroke patients require a close caring from nurses or caregivers for 24 hours, especially after their postoperative. From the last few years, the number of strokes in Thailand is increasing (Hanchaiphiboolkul et al., 2011); this might be the effects from the aging society. Most of the old citizen usually cannot notice their stroke symptoms until the stroke entirely hit them and cause their serious illness. There are 2 types of strokes, ischemic stroke, and hemorrhagic stroke. Thus, the difference types lead to different treatments. Generally, there are two kinds of treatments. Most of the patients are medication; however, some patients have the operation.

For any surgery patients, all of them need to be in the intensive care unit (ICU) because their postoperative condition is not trustable and must be monitored for all times. In such case, many medical staffs, especially nurses and nurse’s assistants, must be on call for 24 hours. Even though the number of caregivers or nurses is needed extremely, there are small amount of these staffs when comparing with the available demands. Furthermore, within the group of caregivers or nurses, there are some inexperience persons. So, the number of qualified persons in the group of caregivers or nurses is small. According to this requirement, the availability of medical staffs in the ICU is not enough and the persons who are on duty will be too tired to perform a complete job. Moreover, in the general situation, when patients feel seriously uncomfortable (Hossain & Ahmed, 2012), they usually ring for a nurse; but not every patient can press the button for assistant.

A serious symptom that patients always cry for help is when they feel pain in some parts of their bodies. This feeling can be immediately detected from patients' faces by the nurse. Then, the nurse will take charge to help the patient as needed. According to the facial detection from nurse's experiences, this detection has been applied to develop a facial action detection technology for detecting patients' pains. The development of this technology can monitor critical patients for all times, and it can inform a nurse whenever a patient has a significant pain. Thus, the nurse can perform a right treatment in the right time. However, there is several patients that always cries for help without necessity. So, implementing a facial detection system for pain interpretation can avoid this annoying case (Chai, Weng, Lin, Chang, & Liu, 2017). Consequently, the face detection technique is applied to the camera capture function. These face detection technology captures patients' faces, and the captured pictures are analyzed with AUs (Action Units) (Zhou & Shi, 2017) (Singh, Majumder, & Behera, 2014), including medical doctor's comments. This method reduces the burden on either the caregivers or nurses.

Since computer technology has been applied for patient monitoring system to support a medical care unit, it is also applied to many medical diagnostic processes to safe human's live. Therefore, implementing medical applications to perform the basic diagnosis for any diseases in the early state can deduct the risk of death. So, inefficient time consuming in the hospital of patients can be eliminated, and patients can receive a good treatment in time as expected. Additionally, the number of doctors, nurses, or nurse assistant in the hospital can be limited for saving hospital's expenses. Furthermore, in the medical profession, technologies have been applied more and more, such as surgery, analysis of the occurrence of heart disease or even death. The results are satisfactory and save a lot of lives.

The improvement in medical applications does not limit to illness diagnosis, but it is also adopted to predict the time to die of people based on their electronic health. This application was implemented by (Avati et al., 2017), using sample sizes of 200,000 persons, both dead and alive. According to the survey of (Bailey & Periyakoi), it found that 80% of American citizen would like to spend the last time of their lives at home. Unluckily, only 20% achieve for what they are wishing. In addition, 60% of deaths happen in an acute care hospital. Although everybody wants to have a peaceful death while they are in a coma and attached to life support machines. Thus, providing the palliative care service for these unconsciousness patients is a solution that patients' families preferred. Nonetheless, most medical doctors do not believe that this method comforts the patients until their last minute. Besides, it is just a deal of families' over-optimism that want to delay the patients' death. Even though insensible persons have good care from their doctors and families, they may obtain either aggressive or non-professional care. In both reasons, they are very costly. Hence, a research expected that patients might please to know their last day of lives. So, it proposed a crucial role to predict the day of death from Electronic Health Record with the Deep learning.

According to the outcomes from medical applications are tremendous valuable to human's live, many researchers have proposed new methods based on the concept of AI, machine learning, and data mining algorithms to create efficient diagnostic tools to indicate patients' sickness. There was a research that proposed the use of the data mining and techniques in AI, the K-NN and the Bayesian algorithm, to diagnose the indication of diabetes. This proposed technique can reduce time for nurses in screening patients for medical doctors (Shetty, Rit, Shaikh, & Patil, 2018).

Medical technology has grown up to improve the abilities of the health care system. The integration between medical knowledge and computer technology has been arose more than decades. Most of these works are focusing in finding methods for identifying causes of illness. In addition, some research works focus in the patient monitoring system for coma persons due to lack of medical staffs.

To perform the medical research, patients' data must be collected, analyzed, and summarized for conclusions. Collected data might be obtained from medical health records, questionnaires, and interviews; all elicitation methods require budgets. In addition, patients must be conscious to answer the questionnaire. Nonetheless, in the case of brain surgery, most patients are always insensible and require at least 24 hours of close supervision. Thus, survey for postoperative from these patients using a questionnaire is impossible; but the study for postoperative requirements of patients is necessary because it will lead medical staffs to provide the right treatment in the right time.

One specific issue that the medical doctors concern is related to the postoperative symptoms of patients; the postoperative may include spine surgery, bone or muscle surgery, and abdominal surgery. Generally, postoperative patients have a tremendous headache or pain around the surgical area with nausea or vomiting. These uncomfortable feelings may be a sign of some abnormalities within the patient's body that needs to have an urgent look up. Thus, the caregivers or nurses must take a close look to the patients under brain surgery to detect an additional unexpected sickness. Therefore, the purpose of the postoperative research is to determine factors of complications, and infections that can occur to the patients so a better treatment can be disclosed and prevent the postoperative mortality. Besides, there was a pilot study in neurosurgery that had studied the postoperative pain of brain surgery patients. This research aimed to disclose factors of patients' pain after they had brain operation. Moreover, the researcher in this pilot study also observed the patients' behavior during pains. According to this study, it suggested that patients with brain surgery should be monitor closely (De Benedittis et al., 1996).

Generally, families of the coma patients usually hire the palliative care service because these patients need an intensive care. Nevertheless, most of postoperative patients also need this treatment because 20%-30% of them may have PONV (postoperative nausea and vomiting); PONV is defined as nausea, retching or vomiting. From the medical record, 30% to 60% of patients experienced PONV. The factors that cause PONV can be classified into 3 groups: individual reasons, anesthesia, and surgery. The first group, the individual reason group, denotes factors such as sex, taking medicine, migraine. The second group represents a group of patients who sleeps more than 1 hour after medical treatment. The last group refers to a group of patients that have surgery. Consequently, to solve the patient uncomfortable problem, patients' monitoring technology is applied to detect posture of patients and predict what would be happened. Accordingly, the nurse can provide an intensive care in time. Another feature is the face detection mechanism that is used to determine either patients are sleeping or waking up. This monitoring technology is also support caregivers to perform their jobs completely even when they are at homes (Chai et al., 2017).

In this research, technology aims to help patients after their brain surgery. Since some patients cannot tell their needs, the intensive care is required. This intensive care is a responsibility of caregivers or nurses who must have knowledge, understanding, and patience. Unfortunately, the number of required staffs is small when comparing with the number of patients. To solve this problem, integrating the computer technology with medical technology is a solution that has been implemented. In this research, the face detection system using video camera integrates with some programs is proposed to monitor patients who are in the postoperative brain surgery. This interpretation will be validated by comparing with the observation outcome from the medical doctors or nurses.

1.2 Objectives

To implement a system that can indicate the patients' emotion: pain, or normal, during the postoperative brain surgery using a face detection technique.

1.3 Scope of thesis and constraints

This research aims to determine the change of emotions from neutral to help mode of unconscious postoperative brain surgery patients. The scope of this thesis is listed below.

1. All samples must be postoperative brain surgery patients in the ICU.
2. All samples are patients from the ICU stroke unit of King Chulalongkorn Memorial Hospital.

3. All samples are selected by the physicians and received informed consent from relatives before their facial images are captured.
4. All samples must be Thai citizen.

1.4 Expected Outcome

The expected outcomes are system and software that can auto-detect normal and pain feelings of the postoperative brain surgery patients.

1.5 Thesis structure

This thesis consists of 4 main chapters, which Chapter 1 provides the introduction. Chapter 2 contains the principle knowledge background and literature reviews. Chapter 3 explains the methodology and analysis of data. Chapter 4 discusses the experimental results and some final thoughts in the conclusion.

CHAPTER 2
LITERATURE REVIEW / RELATED WORKS

2.1 Stroke, a critical illness of death

2.1.1 What is the Stroke

Stroke is a disease, it happened when the arteries are blocked from clot or burst then the blood cannot transfer oxygen and nutrients to the brain (Rink & Khanna, 2011); so, it starts to die. The stroke has 2 main types: Ischemic stroke (clot), and Hemorrhagic stroke (bleeds).

The effect of stroke depends on the damaged brain area. For example, if a part of either the Broca's area or the Wernicke's area are damaged, then the patients will have speech problems. The speech problem caused by the injured Broca is differ from the speech problem caused by the injured Wernicke. Nevertheless, the brain area of patients can be damaged in both areas. Another common problem that mostly occurs when the brain was attacked is the vision's problem. This problem arises when the primary visual cortex located nearby the occipital was strongly hit by a very hard item. In addition, the right-body control mechanism is paralyzed when the left-brain area is wounded. Like the left-brain's effect, the control of the left-body is down because of the right-brain's injury. So, the physical's problems arise when the brain is injured.

Since there are various types of strokes, there are various types of treatments. Therefore, patients must have the correct diagnostic of the stroke disease; so, the right treatments can be delivered. In the case of the ischemic stroke where a clot blocks a vessel, the patient was treated by either medication a clot removed in the acute phase. Different from ischemic stroke, hemorrhagic stroke caused by the bleeding in the brain needs surgery if the bleeding is significantly serious and in the proper location; otherwise, the patients can be treated by medication.

2.1.2 Effects of the Stroke

From the above, the pain assessment is used for self-evaluation in the pain's levels by the patients who are conscious. The advantage of being able to tell the doctor about the pain that help the doctor to identify unusual the pain locations, determine the dosage of the drug used for the pain level, check the reaction. Nevertheless, some patients cannot communicate with others because they are unconsciousness or have language, speech problems, such as aphasia, and dysarthria. However, if the brain of the patient is damaged, the chemistry in the brain is usually altered; and, this change affects to the emotions and behavior of the patient. The mood disorders like a depression, anxiety, and pseudo-bulbar are usually happened for the post stroke patients. In cases of the pseudo-bulbar, the patients will react oppositely between feeling and expression: laughing when sad, crying when hear joke. Fortunately, there are some therapies, but it must take times to restore the patient's state.

2.1.3 Symptoms after the operation

As some case of the stroke needs to be treated by operation to remove the clot or cranium blood that blocks a vessel, the patient and relatives must be informed for the surgery's situation, risk, and ways of caring the patients after the operation. After finishing the surgical procedure, the patient is transferred to the intensive care unit (ICU) to monitor for any emergencies, such as complication, fever, vomit, nausea, and severe pain. In the ICU, there are various types of patients' managements. For example, airway management, blood pressure management, management of Cerebral Edema, management, and monitoring of Elevated ICP etc. Moreover, there are assessments, such as eyes response, motor response, brainstem reflexes, respiration, and pain, for estimating the patient's condition so that the treatment step can be determined.

2.1.4 Stroke caring unit

In the stroke caring unit, if patients are not in the critical situation, like coma or surgery, the patients will be sent to the recovery room. Generally, the recovery room always has proper equipment for taking care of the patients, for example, pulse meter, intravenous fluid set, oxygen tank, medical treatment. Besides, the patients are safe because they are located nearby nurses and doctors who have skills for curing them. Moreover, nurses and their assistants always come to check blood pressure, wipe the patient, give the medicine, or flip the patient's body for bedridden patients. However, these caring cannot be compared to the intensive care unit (ICU) where the most patients in the ICU are in the coma state or the postoperative patients who requires very close attention. In the ICU, there are respirator, pulse oximeter that may attach to the patients for close monitoring (Rink & Khanna, 2011). In addition, in the situation that there is an infected patient, a save zone can be managed for this specific patient.

As mentioned above, the in-patients have many types: conscious, unconscious, and conscious because of paralyzing or damaged brain. Therefore, when any problems occur or some situations that patients need some treatments, the norm conscious patients can give voice to medical staffs. In contrast, the other groups of patients must wait for turn around period of the medical staffs; so, some bad consequences can arise. For examples, in some situations, patients might dissatisfy with the patient care system in the hospital (Gottschalk et al., 2007), or patients need intermediate care; otherwise, they can lose their lives. In such situations, every hospital must face the same issues, but their management is usually dissimilar. The differentiation in management policies can be clearly seen when comparing between the government hospitals and private hospitals since the number of staffs in the government hospitals is much smaller than the private hospitals (Eggleston et al., 2010). Therefore, the detection healthcare system for the emotion of patients in the government hospitals is seriously needs to be improved to increase the quality of patients.

From the above reasons, the researcher aims to apply knowledge in computer science, such as machine learning, to create automate tools that can indicate the emotion of patients who really feel uncomfortable without limited to the conscious patients. Consequently, the problems of lacking medical staffs and high expenses medical caring system can be solved. Moreover, the quality of lives of patients will be much better.

2.2 Detection Tools' Development

2.2.1 Machine Learning Mechanism

According to the abilities of machine learning (ML), the machines or computer's equipment are able to perform self-learning from the data to create new elements which is based on mathematics, statistics and probability that may equivalent to or greater than the abilities of humans. Therefore, many researchers tend to apply this technology to support their research in various areas, such as medical image processing, robotics, etc.

Currently, ML has been applied to many applications, such as advertising, automotive, Internet of Things (IoT), security, trading stocks, and diagnosing diseases; so, in medical area, the use of ML is significant useful because of fast with low cost of diagnostics. For example, in the case of immediate chest pain visit at Emergency room, ML can firstly be deployed to analyze the symptoms for a possibility of being heart disease (Liu et al., 2014) from age, gender, chest pain, fasting blood sugar, etc.; then, first diagnosis result is delivered, following with a proper treatment. Similar to heart disease identification, ML can also determine the patients' situations as diabetes (Kavakiotis et al., 2017) using features of the glucose in blood, insulin, and meal digestion. It helps the patient to treat themselves in early stage, such as food control, exercise, and relaxation.

As the fact that ML can improve its own abilities by self-learning from big-data sets, applying ML in the image processing can provide tremendous and trustable analysis outcomes. For example, after learning various types of x-ray films (Kuang et al., 2019), ML can diagnosis abnormality tissues and its location as same as manual analysis. Besides, ML also can foretell whether the patients will survive or not (Avati et al., 2017). Based on the statistical record of ML's prediction, the accuracy is much better than humans' works.

ML does not only analyze the image for diagnosis diseases, but it also analyzes the patients' emotions (Hosseini & Krechowec, 2004). The work of Hosseini and Krechowec is through image analysis via remote patient monitoring systems (RPMS). The images used in the analysis are black and white images. The result is the ability to distinguish emotional changes that occur with the patient. Unfortunately, it still cannot specify the mood of the patient. After that, people were interested to work in this area, and to develop a monitoring system for detecting changes of patients' mood who lying on the bed in real-time (Chiang, Chen, Chou, & Chao, 2012).

One of the emotions that is important and should be observed is sadness because sadness affects the physical conditions of the patient very much. So, people are studying the detection of sadness from moving their heads (Alghowinem, Goecke, Wagner, Parkerx, & Breakspear, 2013). Moreover, researchers have used wave detection electroencephalogram (EEG) to detect sadness in patients, which normally uses EEG waves to see abnormalities in the brain (Katyal, Alur, Dwivedi, & Menaka, 2015). Then, other types of waves used to detect the transformation of emotions are such as electrocardiogram (ECG) waves. ECG is usually used to diagnose heart related diseases; but the research (Tivatansakul & Ohkura, 2015) uses ECG to improve patient care by detecting stress conditions in patients.

2.2.2 Pain Assessment Test

Presently, the most used the assessment forms for the surgery patients is the pain assessment form. For example, the assessment form using facial images as reflection of pain's levels; each painful face represents the level of pain as shown in Figure 1. Another assessment form uses simple numbers or visual analogue scale that explains the pain level, running from 1 to 10. In both forms the patients must be aware and conscious enough to understand their conditions and the meaning of assessments that are explained by the medical staffs.

Figure 1 Wong-Baker Faces Pain Scale

In evaluating emotions, they usually are evaluated by means of asking the patient's symptoms. In the surgery patient, the most commonly used assessment form is the pain assessment form. There are many pain assessments forms such as using facial images as a reflection of pain that show the painful face image with the number (more number that mean more pain and very hurt face image), using numbers or visual analogue scale that explain the more number with more pain from 0 to 10, both of which the patients themselves must be aware and conscious enough to understand his conditions and the meaning of assessment that is explained by the medical staff.

Another way to judge the emotion of patients who are unconscious or very childish is to observe their behaviors, so called as the external symptoms' observation. The outcome of this inspection is the pain's level based on patient's behaviors. In such case, face expression is one of the most important part of emotion expression since human always changes faces when their feelings changes. Therefore, it is not surprised that many articles refer to the human mind from the face expression. Additionally, the body movement, such as the upper body movement like a hand, or arms, usually leads to the interpretation of patients' uncomfortably which might be caused by pain in the upper body part. Moreover, when a patient has a severe pain, it can be observed from the flexion of the finger, and the tight of the muscle. Furthermore, the statuses when using a respirator, such as coughing, fighting ventilator, and uncontrollable ventilator, can be used to evaluate the pain score by the medical staff.

The pain assessment is significantly vital because the physical and mental of patients can be altered by their pain. Once the pain starts, the patients can feel anxious, uncomfortable, or stress. As a result, the patients become the sleepless or insomnia; this stress leads to some irregular breathing, poor immunity and so on. Consequently, the patients might have slow recovery, and, more than that, it may cause risk to the treatment because of sleepless. In such situation, the medical staff or the doctor must, firstly, pick up the suitable method of pain assessment for determining the level of the patient's pain. Then, the suitable treat can be delivered afterwards. Nevertheless, to determine the level of pain by medical staffs is always dissimilar according to unique experiences. Thus, patients may receive different treatments from different medical staffs under the same sickness and the same level of real pain. For example, young medical staffs may have distinct diagnostic results from high experience medical staffs for the same patients

2.2.3 Pain Assessment Test by Machine Learning Technique

Since the pain assessment test using personal experiences does not a properly work as mentioned above, the concept to deploy computer technology with machine learning technique has been arisen. The example of this research is to identify the level of pain through facial expressions. The data used is from the UNBC McMaster Shoulder Pain Archive database. The pain level is classified according to the standard PSPI scale, which is caused by the sum of action units. (Neshov & Manolova, 2015). Research to analyze the pain also occurs in patients with appendicitis who have to lie in the bed early. (Sikka et al., 2015). In fact, it would not have been better if the camera had to be installed around the patient's bed. So, someone proposed using a mobile phone which is easy to be found currently, is a device to analyze the pain level of the patient (Adibuzzaman et al., 2015)

Though, there are 3 types of behavior pain scales, the face expression is counted as the main criterion among all. Therefore, combining the face expression pain scale with the machine learning technique, which can process images as needed, will classify the pain or the signal from patients who are trying to communicate to the nurses. This combining method is suitable for the post-operative stroke patients, who are unconscious, speechless, but can express their feelings via their faces.

As the fact that ML has abilities to learn and analyze some meaningful context from data, this finding will be derived as a model to classify or cluster items. The main process of ML combines with 7 steps.

Step 1: data collection

This is the most importance procedure because the model of ML can be obtained by learning from the training dataset. This is as teach a baby; if the baby learns from good things, it will have a chance to become a good person. Similarly, in ML, if the model is taught by the qualified data and has no conflict with needs or objectives of the project, the accomplished model from the qualified input will be able to answer the significant issues of the interest. For example, if the aim of the model is to classify numbers from an image but an alphabet-image is used as the input data, even though, the ML process is still successfully achieved without any numbers obtained. In contrast, this failure will not occur if a number-image is the input data because the model will be trained by the right data so numbers can be classified easily.

In the present day, there are many methods to collect data from its source. Some projects need multiple methods to elicit the data for training the model. Therefore, understanding the goal of the project and types of the data are very vital to pick up the right data collection method for the project. Some interesting examples of the data collection method are described as follow.

The easiest way is to collect from the secondary data sources or can be called as the secondary databases. For examples, the websites like Kaggle dataset, Amazons dataset, UCI Machine Learning Repository provide the dataset for practicing skills in ML started from basic to expert. Some companies find their employees by deploying challenge datasets for people who are interested to join. More than that, google provides Dataset Search browser that accesses only the datasets in the world. The result of the Dataset Search represents data description, file type, last update date of each file; anyone who are interested in a specific dataset can access the dataset's website to download.

Presently, the big data is obtained from the daily used websites, but the important issue is techniques to retrieve data from those sources. To achieve the goal, using the web scraping technique is a suitable solution for the new trend of every data science currently, where the understanding only logic, maths, and statistics is not enough. Since the web scraping can collect the data from the websites, such as IMDB, YouTube, Wikipedia, most collected data are mainly displayed in the public and non-owner websites. On the other hand, tracking the usage data using JavaScript is interesting issue since the behavior of the web users can be identified, such as the number of click over a web page, the most hover position on your website, etc. Thus, the website usage record can be improved that is a way to develop an interactive website with value added for the business.

Though, big data is obtained from the Internet or Cyberspace, another traditional data elicitation is the data survey which is commonly use in a large sample size using questionnaire. The concepts of using a questionnaire for data capturing are that all questions must be stated clearly and contain some possible choices to be selected. Nonetheless, to distribute the questionnaire is not simple since the sample size for this data collection is usually big. Thus, there are two methods to hand out the questionnaire: online survey and person-to-person survey. The online survey comes in action to solve the time usages problem by spreading the form to groups in social media that are the target group. Even if the online survey can reduce the time, the target can only read the question without any suggestions. So, an error can occur.

Step 2: data preparation

Data preparation, or wrangle data, is the process to make the suitable dataset for training by the ML algorithm. As the fact that the collected data from the method in step 1 may be imperfect because of missing value, duplicated record, wrong format, etc.; these sorts of problems can affect to the model. For example, the model might not be able to be trained, low accuracy, and lack of incredibility from unsupported reasons. So, the data preparation must be error-awared.

Before jumping into the problem, the data should be explored beforehand to check the suitable method to handle the error mentioned in the previous paragraph. Each error is manipulated differently. For example, one missing value can be substituted by mean, median of feature while multiple missing values might lead to the record deletion instead. Similarly, to the multiple missing values, the duplicate record is solved by the deletion the duplicated one. Besides, the same feature must be in the same format. Thus, in case of having unstable formats, the reformatting must be performed. Moreover, the range of the data is needed to be concerned because the large difference in ranges of the data can affect to the algorithm of the model where the bias occurs. To solve such problem, data normalization is needed. In addition, the data preparation processes may apply statistical methods to solve the problem like unnormal distribution for creating unbiased data before the training process.

In the ML process, data is called as a dataset because of the ML process applies groups of datasets to create a model. The data that is used in the model derivation will be divided into 3 types: training dataset, validation dataset, and testing dataset; nonetheless, only two types are sufficient for building an efficient model.

Step 3: choosing a model

Before choosing a model for the dataset, the available models of machine learning must be clearly understood. The ML model is divided into 3 types: supervised learning, unsupervised learning, and reinforcement learning. The supervised learning is a trained model by the labeled data. The labeled data is the dataset that every record has only 1 column with a tag of the correct answer. The human teaches the machine with the labeled data that contains the true answers. For example, if the model is used to classify the weather of the day, in the dataset must have features like hot, or rain, or cloud, or windy. It is called a label. In contrast, the unsupervised learning means the dataset does not have any labels. So, the model can only search for the structure of the data. The human just put the dataset, and the algorithm will automatically find the group of the records that are the answer. The last one is the reinforcement learning that has some famous example like Alpha GO, OpenAI. The main idea is trial and error with the reward, if collect, in another way, give the punishment, if incorrect. The reinforcement learning is the newest learning method that allows the machine learns from its failures. Thus, every obstructed process will be solved in the next learning step.

Based on the above contents, it is about all the input data. Next, the output data will be described. It the output data, it has been separated to 3 main categories: regression, classification, and cluster. First, the regression is the numerical data that is the qualitative dataset; the linear regression is one of the example models. Second, the classification is class of the records in a dataset which can represent in numeric or word to show groups of the data. The examples of the classification model are such as the logistic regression, a binary classification model, can classify the dataset into 2 groups. Moreover, there are some ML models that perform both regression and classification. Such as, neural network model simulates the working principles of the nervous system. For the cluster model, the model is used to classify data into groups using distances between them. The example is k-mean clustering that create the centroid by distance between the points which represent the group of data.

After understanding the input and output types of ML model. The next step is to specify the dataset's type. Before matching the group of ML models, the suitable dataset must be declared. Though, there are varieties of ML models with 5-10 lines of codes available in the ML module, every ML algorithm has limitation, specific input, pros and cons of the model. So, a suitable ML model can be identified. However, the obtained model must be refined to support specific objective of the project. For example, if the object is used with the patient by classifying the disease from x-ray films, so, the accuracy should be high, then the result in the complex model must come from high cost. Another example is to use the model to identify the trend of stock market; the stock prediction is the hardest problem. Moreover, some of the great traders need a short period of a regression prediction trend combining with the stock knowledges is enough to predict the next trend of the stock price. In such case, the accuracy is not always necessary.

Step 4: training the model

From the above, this process is like someone wants to ride a car. At first, they do not know how to start the car, use the pandal, or break, but after a long-time training, they know how to drive, pass the driving license test, or even turn themselves to be racing drivers. In machine, it is similar in teaching the machine with the training dataset, sequentially, step-by-step, and repeats as many as possible. The training dataset is used for training a model, as named. The training process starts when a model is defined, then the training dataset is input to the model to solve all constants and parameters of the model.

In fact, the model is the mathematic equation or algorithm. Therefore, the linear equation is considered as a model. The fundamental model is a linear regression model, yi = mxi + b where xi is the input, m is the slop, b is yi-intercept, and yi is the value in the line from input xi. According to the sampling dataset, each dataset has individual classifier xi and yi values because every time the input xi has been changed, then it affects to the values of m and b, including yi. In such case, it can be said that m and b have direct impact to the model of yi.

As the fact that the value of the variable m does not only depend on the sample dataset, but also depends on the number of features to be interested. So, from many features or independent variables, there are a group of m where they can be derived from a matrix, and, called as w or weight. Similarly, for b, that is called bias. The initial of training process is started with random w and b for feature xi, then compare the result of prediction with the answer, this is called loss of the model. Then, the model needs to adapt the values of w and b to reduce the loss for every training dataset. It is like driving, it cannot be good at the first time, the repeat training process is always needed.

Moreover, the outcome of this training process is the full defined model with constants, variables, and defined conditions. During the training process, the validation process can be performed by dividing data in the training dataset for an unbiased of model effectiveness. So, the outcome of the training process is the unbiased model. This step aims for improving the ability to predict, cluster, answer the gold of the project.

Step 5: evaluation the model

After obtained the model, this model must be evaluated in the testing process using a testing dataset instead of the training dataset. The purpose of this process is to confirm the fitness of the achieved model in the use of real-life situation. Thus, using the training dataset in the testing process is not suitable because the model is derived by the training dataset which cannot be assumed as the real-life situation as needed. In such case, another dataset, testing dataset, must be applied in the testing process to simulate variety situations for the model to test the model-fitting, and the model accuracy.

The method of model's evaluation has 2 main types: holdout, and cross-validation. For the holdout method, where the use of the training and the testing datasets to evaluate the model's performance, the testing dataset (used features of the dataset) is used to predict the result from the completed model. Then, comparing the result of prediction with the true result of the testing dataset with an evaluation metric is performed. Considering the ratio of the dataset in each procedure; there is no certain number to be defined but most researches put the size of the training dataset bigger than the size of the testing dataset, such as 80:20, or 60:40, etc. Second method, the cross-validation or K-folds cross-validation, this method will separate the dataset into K-folds with equal size. Then, K-1 folds are used to train the model and used the rest fold to test the model with evaluation metric like the holdout method. This counted as 1 iteration, the process needs to repeat by changing the training dataset, and the testing dataset with another fold until every fold is used to test the model. The last result of evaluation comes from the average of the result in every round. The benefit of K-folds cross-validation reduces the bias and variance, but the holdout method has speeded up and flexibility.

The evaluation metric of the obtained model is calculated to indicate its efficiency; and the proper metric is chosen based on the model. Examples of the evaluation metric for classification model are such as accuracy, confusion matrix, logarithmic loss, area under curve, and F-measure. On the other hand, the mean absolute error, and root mean square error are examples of evaluation of a regression model.

Step 6: parameter tuning

According to the performance metric obtained in Step 5, it is possible to improve performance of the model by changing some parameters in the training process, such as learning rate, iteration, hyperparameters; then, execute the training and evaluation procedures again. According to the tuning process, the parameters to be adjusted in this process are categorized into 2 types: parameters, and hyperparameters. Details of these types are described in the following paragraph.

The parameters refer to values obtained during the training process. Thus, the parameters in the model will be altered according to the training data and algorithms. Examples of the commonly used algorithms are such as Gradient Descent, Stochastic Gradient Descent (SGD), Mini-batch Gradient Descent. Thus, no matter what algorithms are applied, parameters will be adjusted until least error is obtained. In such case, the learning rate is like the speed of learning. So, if the learning rate is high, the parameter adjustment will be large too. In contrast, it has the opposite effect. In addition, another adjustable variable is the number of times that each record is entered to the training process; this may be referred to as steps, number of estimators, and so on. With this variable, the model will clearly understand the information. However, the values of learning rate and the number of times do not imply that the higher is the better for the model’s performance because the model’s performance is only relied on the smallest error of prediction. Additionally, if there are too many steps, the model may be overfitting.

Consider the hyperparameters. The hyperparameter is dissimilar from any common parameters because the value of this hyperparameter will not be changed during the training process. In such case, the value of the hyperparameter is determined prior starting the training process. Moreover, the hyperparameter’s value will be varied on the chosen model. For example, there are various hyperparameters for a Tree model, like, the maximum dept of the tree, the height of the tree, and the number of children per a node; these hyperparameters reflex to the resolution of the model and so on.

Although the classification model is pre-defined and a training dataset is known, there is no doubt that the hyperparameters can be pre-defined, but not parameters. Moreover, during the training process, the parameter tuning in the tuning process does not guarantee that the new values of parameters are the precise values for what have to be used.

The tuning parameters and hyperparameter process cannot specify more precisely what the value must be used. It all depends on the model, the data, and the training process.

Step 7: prediction step

For the final steps to check the reliability of the model that can be used in the real life is to test the gained model with testing dataset; then, checking whether the model's performance is acceptable.

2.2.4 Neural Network

Neural Network is a model created to solve data problems like other ML models in general, but there is a misconception that their working principles are the same as the work of the human brain. In fact, it is still difficult to understand the true working principles of the human brain.

The general working principle of the neural network is to transfer data through multiple nodes or perceptron, which can be designed freely by users to extract important information and analyze the results. For example, the calculation of a single-node neural network with multiple inputs by X is input data, W is weight for each connection, while input via perceptron is calculated as equation. (1)

$$Z_1 = \sum_{i=1}^{n} x_i w_{1i} + bias \tag{1}$$

Until getting Z_1. After that, Z_1 will enter into various activation function, such as relu, tanh, sigmoid, etc. As show in Figure 2.

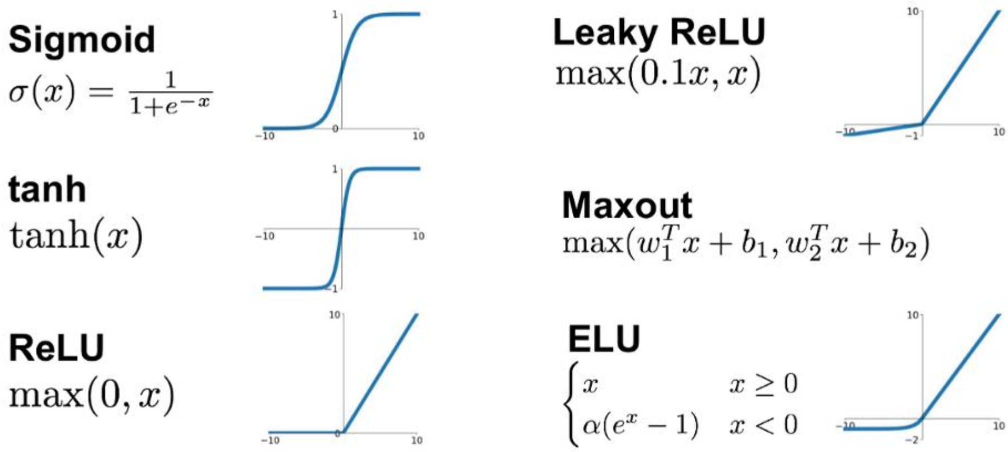

Figure 2 Activation Function.

(Available from: https://mc.ai/complete-guide-of-activation-functions/ Accessed: July 15, 2020)

The activation function is for converting the equation to a non-linear function. For the neural network, the activation function can be any forms, both linear function, and non-linear function. According to Figure 3, the output O1 is the result after the perceptron. The weight and bias are adjusted according to the error occurred during the training process using backpropagation method, such as stochastic gradient descent, to calculate from the gradient to find the minimum value.

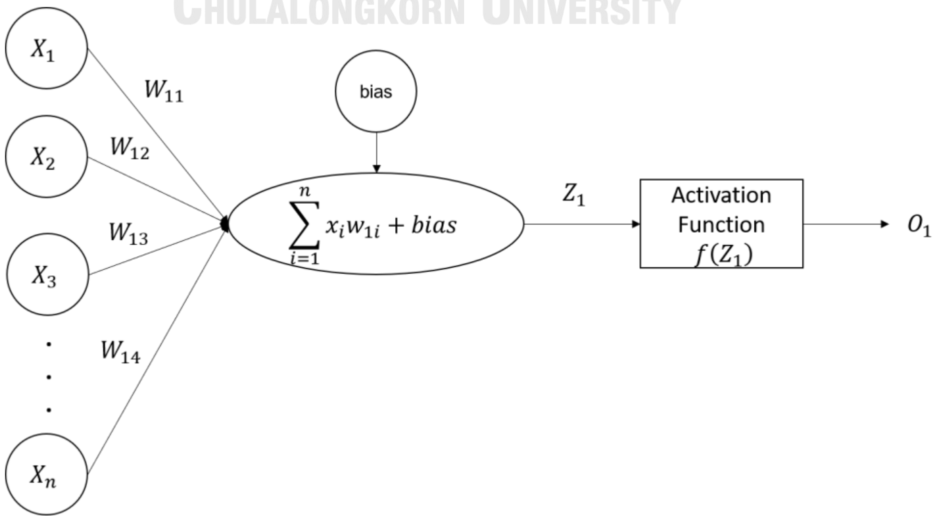

Figure 3 Multiple inputs with a single neural network.

2.2.5 Convolutional Neural Network

Convolutional layer

This also known as the kernel layer, it is a layer that contains a kernel or a filter that screens low level features out from images, such as lines, edges, and corners. The process of this method aims to reduce the size of an image; it starts from setting dimension of the image, such as 5*5, including the number of channels e.g. RGB. Therefore, the image can be input to the conventional layer mechanisms as 5*5*1 where 5*5 is 5 rows and 5 columns with the number of channels as 1. After setting the image's dimension, the dimension of the filter or the kernel must be identified without the number of channels, such as 3*3. Later, the number of shifting bits (strides) of the filter within the image must be defined before starting the filtering process. To initiate the filtering process, the defined filter is put to the top and left-most of the image. Then, the multiplications of bits that are overlapping to each other, when the filter is put over the image, are executed and summarized to the new image with the smaller size before the filter striding to the next position. The multiplication and summarization processes are repeated until the filter completely moves around the image blocks. At this end, the similar image likes the original one but smaller size as 3 * 3 * 1 can be obtained. This filtering processes are shown in Figure 4.

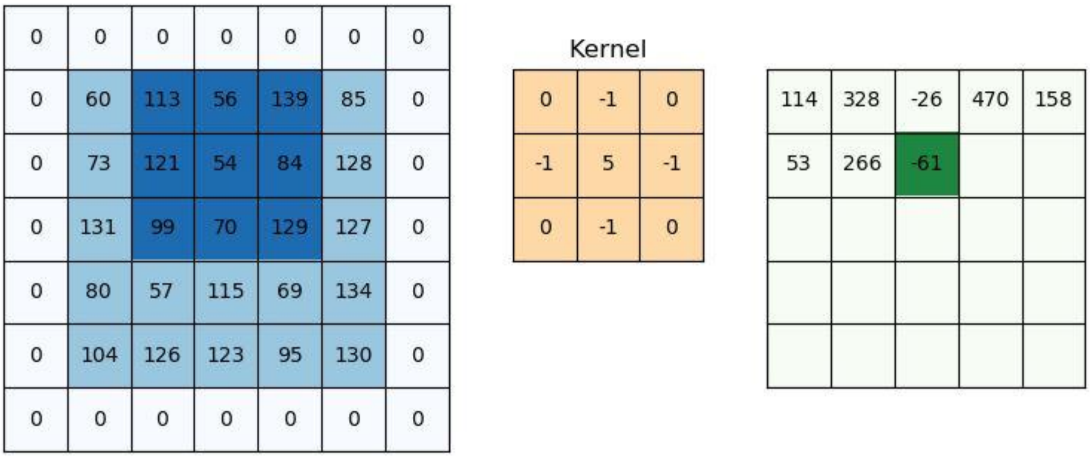

Figure 4 Convolutional layer.

(Available from: https://www.pyimagesearch.com/2018/12/31/keras-conv2d-and-convolutional-layers/ Accessed: July 11, 2020)

Pooling layer

This is a procedure that reduces the image size to save the computing resources and to reduce the noise of the image. There are two main pooling layers: max pooling, and average pooling. The pooling layer technique likes the convolutional layer, except it selects the maximum or average values instead of multiplying the matrix to represent all the pixels of the filter.

For example, if you have a 5 * 5 * 1 image, a 3 * 3 * 1 filter, and a stride length equal 1, if you choose the max pooling method, each filtering time, the maximum value within the filter matrix is selected. As the consequent, the result of the max pooling is shown in Figure 5.

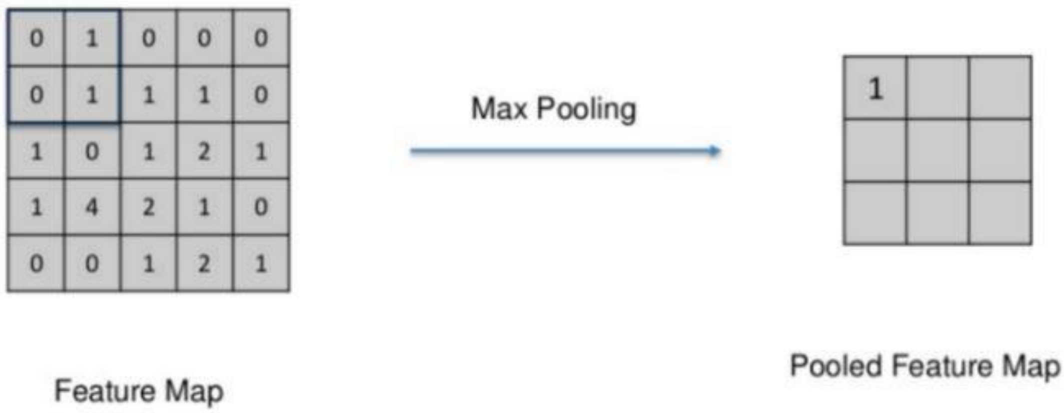

Figure 5 Max-pooling layer.

(Available from: https://medium.com/@pradyasin/what-is-convolution-neural-network-bf2e525089f5 Accessed: July 11, 2020)

Both the convolutional layer and the pooling layer don't need just one layer. It can overlap those layers to extract many features out of one image. As in the case of studying the patient's face image in the first layer, the convolutional layer extracts a line or a border in various directions until those lines construct lines of an organ like the eyes, nose, mouth before bringing these features to other machine learning models. The integration between the convolutional layer and machine learning model had produced a new method, namely CNN (Convolutional Neural Network). This CNN method has been derived to various techniques, such as R-CNN, Fast R-CNN, YOLO, and FaceNet. Nevertheless, this research will apply only the FaceNet method, the reason is explained in the following section.

2.2.6 MTCNN

The Multi-task Cascaded Convolutional Network (MTCNN) is a model used in the face detection process before passing the facial images to face recognition or face verification processes. The MTCNN consists of 3 sub-networks as P-Net, R-Net, and O-Net; its structure is shown in Figure 6 (Zhang, Zhang, Li, & Qiao, 2016).

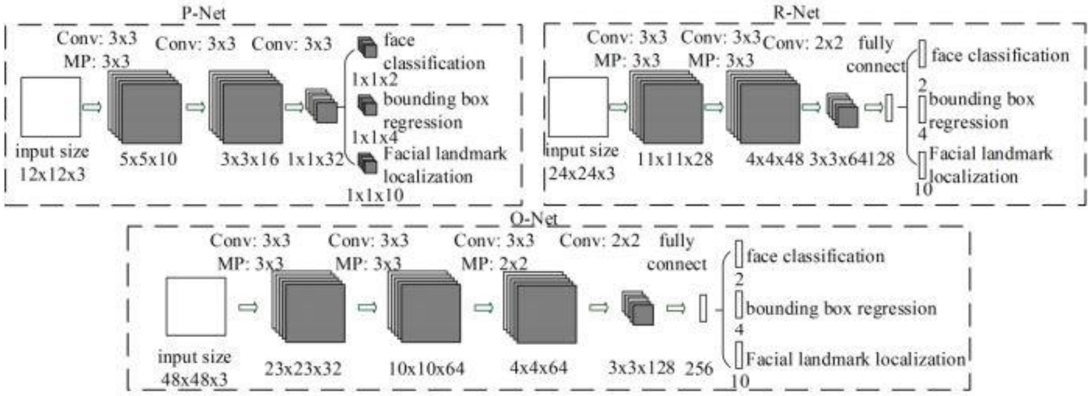

Figure 6 Structure of the MTCNN.

When receiving an input image, the output image will be created as Image pyramid, copy of the image into various sizes. After that, the image pyramid will pass through P-Net, R-Net, and O-Net, respectively. There is a similar procedure: pass to convolution neural networks to find the boundary box of the face in the image. Remove the boundary box with the least confidence in prediction and use the Non-Maximum Suppression method to reduce the number of duplicate boundary boxes. The difference in each part is the output in the P-Net and R-Net layers, which have prediction value and boundary box position of face image. Finally, at the O-Net stage, facial landmarks will be added. The process pipeline is shown in Figure 7 (Zhang et al., 2016).

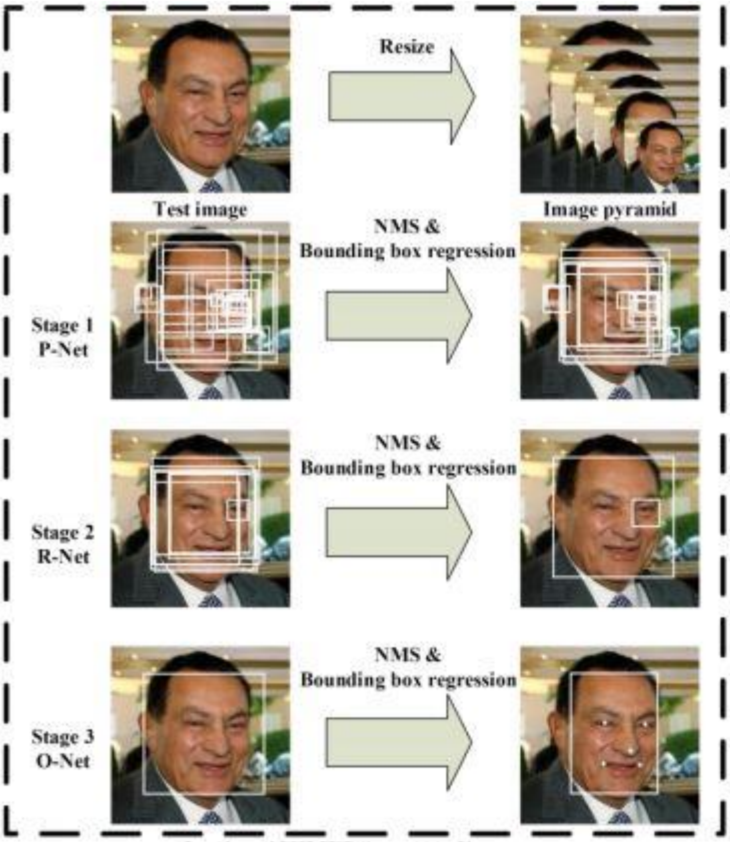

Figure 7 The process pipeline of MTCNN.

2.2.7 FaceNet

The FaceNet (Schroff, Kalenichenko, & Philbin, 2015) is a deep CNN model that can be applied to face verification and face classification; the process pipeline of the FaceNet method is shown in Figure 8. Though, FaceNet deploys the CNN mechanism, it also applies L2-norm to convert images to 128-dimension vector, including the Triplet Loss to compare images in Euclidean space. The accuracy of using FaceNet for classifying a person is 99.63%.

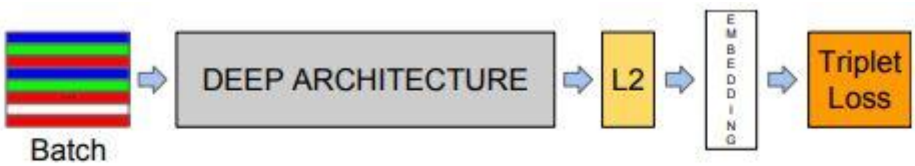

Figure 8 The process pipeline of FaceNet.

In such case, for similar images, the Euclidean spaces must be closed to each other. On the other hand, dissimilar images usually have large distances, see Figure 9. Although the FaceNet deploys the Triplet Loss within its mechanism, the researcher can replace the Triplet Loss module with a machine learning model, such as SVM, KNN, Neural Network etc for process suitability.

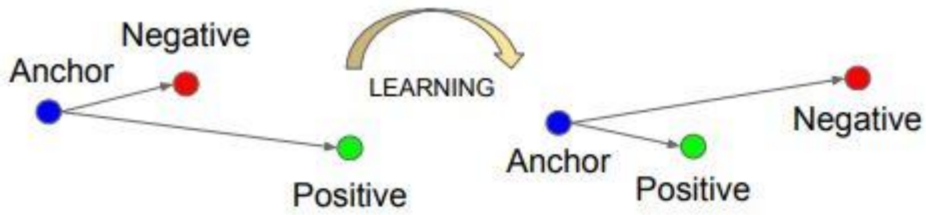

Figure 9 The Triplet Loss's mechanism.

Since FaceNet is used for face classification, the general process to be performed is the face training where varieties of faces are input to the learning process to create a classification model.

CHAPTER 3
RESEARCH METHODOLOGY

In this chapter, the experimental research is considered under the assumption that the samples are patients who are the stroke patients and had undergone surgery not longer than 72 hours. With this assumption, the model to classify the feelings of the patients via their face after the surgeries is created. The entire processes in this research are elaborated below.

3.1 Population and Samples

This experiment focuses in the patients who are postoperative stroke patients; so, the sample group is obtained from patients in King Chulalongkorn Memorial Hospital during July 2019 – March 2020. However, to collect data from these samples, it cannot perform without medical approval before hand because the patients who are samples must be evaluated by the medical staffs to confirm their appropriation to participate in the experiment. Once the medical staffs approved the condition of the patient, the patient’s relatives will receive the details of the experiment and the to be signed for permission. After receiving the consent form with signature of the patient's relatives, the device will be attached to the patient's bed and start the data collection process.

Based on the statistical data during the year 2016-2018 in the database of the Government Health Service under Ministry of Public Health, the number and the rate of inpatients with stroke per 100,000 population (including all diagnoses) in the year 2018, there are 331,086 cases by the population of 65,406,320; or can be the case rate of 506.20 per 100,000 persons. On the other hand, this rate is equivalent to the probability of 0.01. Therefore, from the Cochran’s formula (2)

$$n = \frac{z_{\frac{\alpha}{2}}^{2} pq}{e^{2}} \tag{2}$$

According to the need for a 95% confidence percentage, the result is $Z^2_{\frac{0.05}{2}}$ =1.96, p=0.01 = q=1-p and the precision (e)=0.1. So, the number of samples is 0.38 or 4 people.

3.2 Data collection

Since the data used in the experiment is very sensitive, there is no available free-of-charged images of the desired targets existed on the Internet. Thus, this research has to occupied random samples from King Chulalongkorn Memorial Hospital. Nonetheless, the first step before the data collection is to obtain the Institutional Review Board (IRB) which contains Research Ethic Committees (REC) before the research starts and corrects data at the ICU Stoke Hospital.

In this research, the profile data of patients will not be used as input dataset because there is no difference between young-unconscious and old-unconscious patients, or unconscious male and unconscious female. Thus, the only concerned in this experiment is the patients' faces. To collect facial information of patients, both hardware and software design are required as tools for this process as described below.

Hardware Design

Since the stroke patients who must undergo for surgery usually be bedridden and must be turned left and right in order to prevent the pressure ulcers. Thus, it is necessary to capture the images of the patient while lying in the bed; the designed instrument used to perform data collection is shown in Figure 10, 11.

Figure 10 The Camera that attaches to the left and the right of the patient's bed.

Figure 11 The Camera that attaches to the top of the patient's bed.

The main structure of the designed instrument is divided into 2 parts, namely a client and a server as shown in Figure 12. For the client side, it is used to save images before sending to the server site. The client part is installed on top of the patient's bed using a tripod device; this part made of aluminum profile building by the maintenance technician of the Faculty of Science Chulalongkorn University. Furthermore, the camera that records images is the raspberry pi (rpi) board.

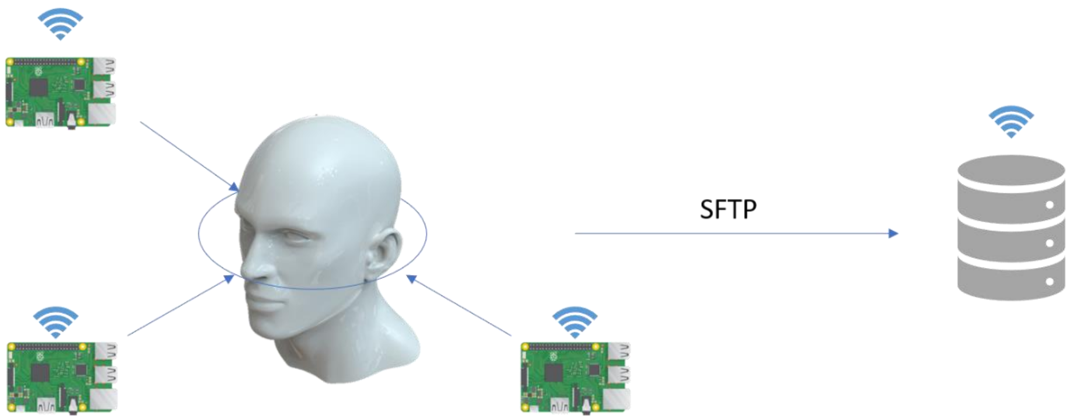

Figure 12 The Facial Collection System.

Since the rpi is like a small computer, it is possible to attach some additional equipment such as a small camera, and infrared device, and can embedded with additional codes to save images conveniently. In the area of the patient's bed, there are 3 cameras attached on the top and asides of bed, both left and right. Consequently, the frontal face image can be captured in every angle although the patient turns the head to the right or the left sides.

Software Design

To collect data from the bedridden patients, commands must be embedded to the camera's board to manage the image capturing process and quality. In such case, using rpi in the capturing process is the right solution since the rpi board easily supports the scripts' embedding that are responsible for defining functions and parameters; those are the recording frame rate, the image size, the resolution of the image, the type of file, and the record duration. In this research, the file type of the recorded file is .h264 with the resolution of 640 * 480 px; this resolution is large enough for the size of a face after cropping. Moreover, the recorded file is set to 25 frame rate, or 25 frame per second (FPS), so it can be easily chopping and sent to the classification model for being classified easily. The average data collection time is approximately 5 to 7 days. However, if it is necessary for patients to relocate or cancel the experiment, the equipment will be uninstalled immediately. The Table 1 shows the defined values of parameters that are used to collect data via rpi.

Table 1 Defined values of all parameters related to data collection via rpi.

parameter	Value
File type	.h264
Image resolution	640 * 480 px
Frame rate	25 frames per second
Record duration	5 – 7 days

According to these defined values, each file usually has the size approximately 35-40 Giga Bytes for 5 days, or average 300 MB per hour. The file is so large because of its coloring.

After the video recording is completed, the files from 3 cameras will be sent to the server via the hospital's internet using sftp protocol. Even though, there are options like cloud servers available over the Internet, but for the security reason of the sensitive data, the local storage is selected and the sftp protocol is used to prevent the sniffing of data.

The information required from each patient is the set of facial images at the time of pain, uncomfortable, including the neutral faces. The frontal face image is the desired image because it is the most noticeable aspect of facial expressions. Therefore, the frontal face provides a lot of information that can be useful for emotional classification.

Since the brain surgery patients are rare in King Chulalongkorn University Hospital, thus, within the data collection period, there are only 8 patients to be collected and two patients out of 8 cannot be used in this research according to their sickness condition. Therefore, there are only 6 patients who are satisfied with the collection criteria. Nevertheless, all pictures in this dissertation are not the pictures of real patients who are samples of our study because of their privacy and IRB.

3.3 Preprocessing data

After obtaining the video datasets, scenes in the video will be classified into 2 groups: a group of neutral faces, and a group of wagging faces. According to the wagging faces, the researcher interprets as the help-sign of the patients by observing from the video contents. Based on the contents in the video, the abnormal movements of other organs such as arms, shoulders may be seen by nurses before the nurse comes for examining. Another criterion is the use of action unit code sheet for pain as sorting (Neshov & Manolova, 2015). The separation of the video's groups is manually process that had been judged and accepted by the researcher.

Therefore, the frames indicate the neutral face and the frames indicate the help-sign face must be separated. Thus, the software named FFmpeg is deployed. This process is the manual process that uses 25 FPS to determine the number of frames to be cut off. Nonetheless, the selected frames are only frames that clearly show the frontal faces without any interferences. Since one frame is one image, this experiment will use 30 images per emotion. Thus, each sample contains 60 images: 30 neutral-images, and 30 help-images. The face classification procedure after obtaining all images mentioned previously are shown in Figure 13.

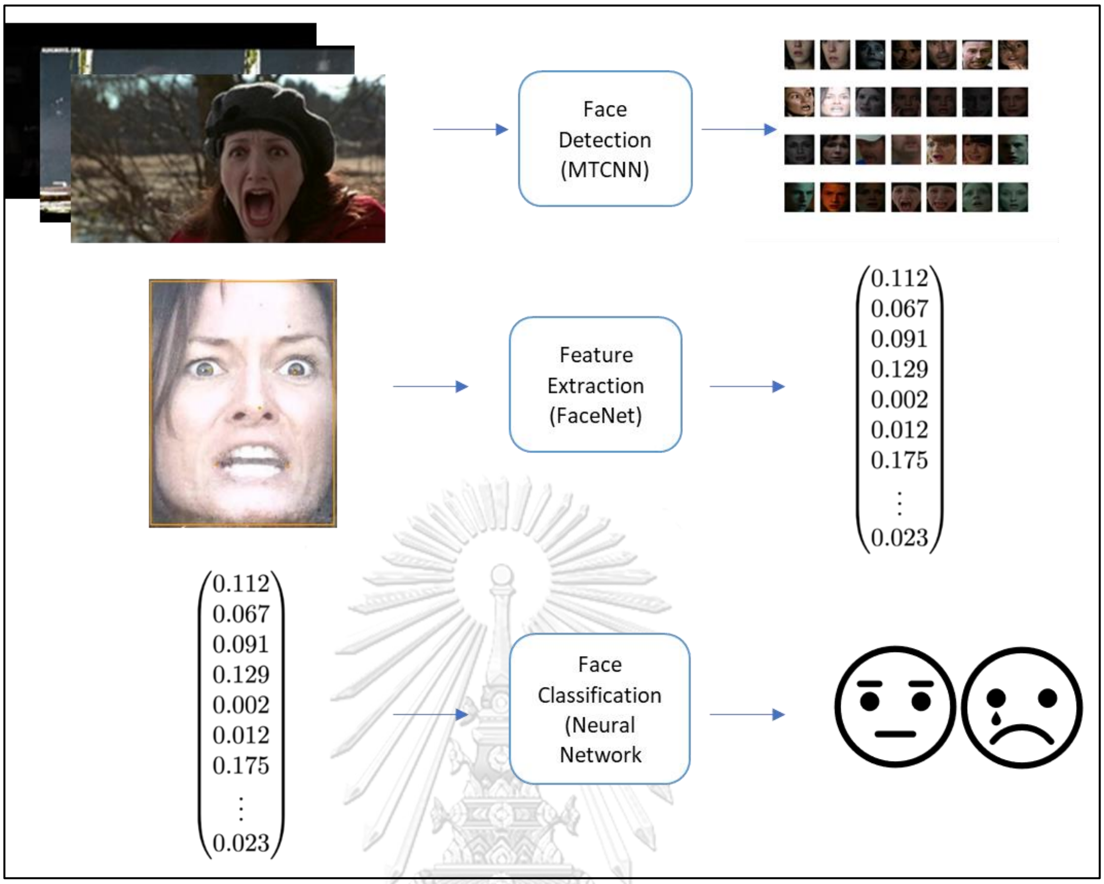

Figure 13 Face classification pipeline.

From Figure 13, the output of the face classification procedure is the status of the face either neutral or help. The input of this process are images obtained from the FFmpeg. Out of the FFmpeg, the function that is deployed to detect the face from the pile of the images is MTCNN; therefore, the outcomes of the MTCNN model is only the faces within the images. The next step is to extract all features from the face which is usually appeared as one-dimensional vector with the size of 128 features, as shown in Figure 13; the neural network work is responsible for interpreting the emotional of a face to either neutral or help requesting. Details of each process will be described below.

3.4 Face Detection

The objective of this process is to detect only faces that appeared in the image. The input image resolution using MTCNN is unquestionable and the output from this model is the pictures of the face where its resolution is equal to 160x160 pixels. Moreover, all images must be standardized before passing through the next step so they can be compared when needed. Formula (3) is the standardization of the value x_i

$$Z = \frac{x_i - \mu}{\sigma} \tag{3}$$

3.5 Feature Extraction

In this process, all faces are put through feature extraction process so features on the face can be figured out individually. Once the elements on the face are separately identified, the emotion testing can be executed easily. The main model that is deployed in this process is the FaceNet; its input is the facial image, 160x160 px, obtained from the first process while the output is the one-dimension vector with 128 numbers that represent the most important features of the face. The generated vector obtained from the FaceNet is named as the feature vector. The benefit of this feature vector is that it supports the comparison between two facial images, including image grouping.

Though all images were put to the FaceNet for features extraction, the researcher also performs manual observation to find some similarity of patients under the situation that they might need some caring from medical staffs. The results of this finding are shown in Table 2.

Table 2 Patient's movement when the medical caring is requested.

	Eyes	Eyebrows	Nose	Mouth
Case 1		x		x
Case 2			x	
Case 3				x
Case 4	x		x	x
Case 5		x		x
Case 6		x	x	x

From Table 2, it is obvious that these 6 patients either move their nose or mouth whenever they need some help from the medical staffs.

3.6 Face Classification

Consider numbers within each feature vector. These numbers might have large variance because of high differences among them. Thus, these differences must be resolved by normalizing the feature vector with the L2Norm function before transmitting the feature vectors to the classification process. The reason of running L2Norm is to adjust numbers' differences of feature vector so the variance of each vector is small and suitable for being classified.

After adjusting the feature vector to be in the norm scale, these vectors were input to the classification process. Nonetheless, this research applied two classification processes to gain the best classification model that has the highest accuracy in the classification. The first classification method in this research is the Machine Learning Mechanism while the second classification method is the Neural Network.

Before execution both methods, the data of 6 patients are divided to 2 groups: training and testing. Since there are only 6 patients, data from 4 patients were used for training and data from other 2 patients were used for testing. Thus, there are 120 neutral images and 120 help images in the training dataset. On the other hand, there are 60 neutral images and 60 help images in the testing dataset. Details of each method are elaborated as follows.

3.7 Machine Learning (ML)

Although the ML mechanism has various classification model, each model had been implemented in dissimilar objectives. After considering objectives of various classification models of the ML, the most suitable model for this research is Logistic Regression Model because this model is suitable for binary classification problem: neutral, or help. In addition, the library scikit-learn was also implemented in this command list to support the classification process. Figure 14 shows the list of commands that execute the Logistic Regression Model using the feature vectors as input data.

```
from numpy import load
import matplotlib.pyplot as plt
from sklearn.metrics import classification_report,confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import roc_curve, auc
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import Normalizer
from sklearn.linear_model import LogisticRegression

data = load('thesis30_2-embeddings.npz')
trainX, trainy, testX, testy = data['arr_0'], data['arr_1'], data['arr_2'], data['arr_3']

print('Dataset: train=%d, test=%d' % (trainX.shape[0], testX.shape[0]))
# normalize input vectors
in_encoder = Normalizer(norm='l2')
trainX = in_encoder.transform(trainX)
testX = in_encoder.transform(testX)
# one-hot encoder
out_encoder = LabelEncoder()
out_encoder.fit(trainy)
trainy = out_encoder.transform(trainy)
testy = out_encoder.transform(testy)

log = LogisticRegression()
# train model
log.fit(trainX, trainy)

# predict
yhat_train_log = log.predict(trainX)
yhat_test_log = log.predict(testX)
```

Figure 14 The commands for Logistic Regression Model using the feature vectors as input data.

Once the classification model, Logistic Regression Model, is determined, the training dataset were input to gain the proper coefficient values of the Logistic Regression Model. However, the obtained model must be validated by the testing dataset.

3.8 Neural Network

Beside the machine learning, neural network is another method that has been utilized in the classification process. Furthermore, the face classification process using neural network method is the same as the ML. Thus, the training dataset was input to the neural network process as follow.

In the beginning, to execute each feature vector or 128 values in the network, 5 layers of neural are employed: 1 input layer, 3 hidden layers, and 1 output layer. For each layer, there are nodes where each node represents a feature. So, at the first layer or the input layer, the number of nodes is equal to the number in the feature vector; it is 128 nodes. Similarly, the output layer has only one node which indicates the status of the patient: neutral or help. For other 3 hidden layers, there are 64 nodes, 32 nodes, and 8 nodes in the second, third, and forth layers, respectively.

The results from executing the neural network with 5 layers are weights of the connections between nodes and bias of each node. Sample outcome of the neural network execution is shown in Figure 15.

```
Weight layer 2 node 1 :  [-0.23313304  0.09790544  0.15135265  0.19221887  0.03531433  0.08971281
 -0.04268375  0.15529453  0.04851326  0.00145657 -0.11033373  0.23331783
  0.18128507 -0.10315155  0.20630371 -0.07511627  0.19315593  0.2076725
 -0.13083054 -0.03854848 -0.06658459 -0.09750367  0.07671287 -0.05661486
 -0.04618166 -0.07725259  0.23369389  0.20922907 -0.1746355   0.02368591
 -0.10019657  0.23589383]
Bias layer 2 :  [-2.06705804e-02  1.88562499e-01 -2.68214687e-01  7.62496010e-02
  2.61447063e-02  9.39247547e-02 -3.58288212e-02 -1.14170690e-01
 -5.71566782e-02 -9.63395483e-02 -3.56623735e-02  2.04040199e-01
  1.48387258e-01  1.16129265e-01 -1.55094569e-01 -3.96471177e-03
 -1.25522024e-01  2.19120981e-01  2.33744360e-01 -9.17642869e-02
  2.61857139e-01  1.40517175e-04  1.16673343e-02  6.98951056e-02
  1.07108531e-02 -2.37448324e-01 -2.60425192e-02 -2.08376898e-02
  2.77616225e-02 -1.60619811e-01  3.73344666e-02 -9.41063348e-02]
```

Figure 15 Sample outcomes from the execution of neural network with 5 layers.

After obtaining all weights of all nodes in the neural network, the accuracy of this neural network was validated using the testing dataset as same as the testing procedure in the ML process.

CHAPTER 4
RESULTS

This chapter displays all outputs obtained from the processes in Chapter 3. So, the results from the machine learning and the neural network processes, both training and testing, are elaborated in this chapter.

4.1 Feature Extraction

In the feature extraction, the FaceNet model has derived feature vectors for the face classification procedure. So, all 360 images have been transformed to 360 feature vectors. Examples of feature vectors are written in Figure 16. In Figure 16, the feature vector of a patient in the neural mode is dissimilar from the feature vector in the help mode of the same person. Moreover, the elements in each feature vector are 128 elements that represents significant features on the patient's face under a certain feeling. According to this differentiation, the comparison to find the alter of patients' emotion can be performed easily.

```
array([ 0.47233108,  2.0590835 , -1.9747272 ,  0.65745544, -0.7293263 ,
        0.13442393,  0.51246583,  0.16898192, -0.09018771, -0.4225717 ,
        0.9048385 , -0.12885891,  0.4739393 ,  0.6117623 ,  0.90136015,
        0.497006  ,  1.0777632 , -0.88247806, -0.7886207 ,  0.72773045,
       -0.4582012 ,  1.3489904 , -1.7832385 , -1.0022447 ,  0.63692176,
       -0.8040101 , -0.10897799, -0.5606362 , -0.12550221,  0.87313473,
        0.2435964 ,  0.80276513,  0.7642509 , -0.6101491 ,  0.66032267,
       -1.2045912 ,  1.2042263 ,  0.14062169, -0.20401385,  1.0614078 ,
        0.39793852,  1.209531  , -1.6325686 ,  0.14068943, -1.3178754 ,
       -0.94952583, -0.8559455 ,  0.52710664,  2.064635  ,  1.7167673 ,
        0.3261717 , -0.25029057,  0.66864175,  0.30491835,  1.0323321 ,
       -1.0972782 ,  0.74921376, -0.18359876,  0.16923462, -1.3725495 ,
       -0.74877286, -0.3572278 ,  0.539706  ,  0.42567828,  1.243135  ,
       -0.82000613, -0.47486353,  0.11879253, -0.91858226, -0.5717472 ,
       -0.6483919 ,  0.3183433 ,  0.46674573, -1.6336414 , -0.6680045 ,
       -1.2580148 ,  1.109404  ,  0.6625618 , -0.22199148,  0.52269584,
       -0.9433569 ,  0.02984457, -0.36161476, -0.20788135, -0.23882297,
        1.1730917 , -1.6900126 , -0.7401642 , -1.0880473 , -0.28025272,
        1.742019  ,  0.28632295, -0.6092259 , -1.1062785 ,  0.3608358 ,
       -0.02815157,  2.096127  ,  0.14795059, -0.99499166, -0.75452244,
       -0.40601707,  1.176596  , -0.70796216,  1.100652  , -0.08966868,
        0.03227214,  0.56316394, -0.17139004,  0.5160634 ,  0.79515827,
       -0.98774993, -0.83459425,  0.9187367 ,  0.7443392 , -1.0679026 ,
       -1.7215062 ,  0.05126145, -1.1231704 ,  1.6090177 ,  0.45982257,
       -0.38299727, -0.15469788,  0.3674025 , -2.2264502 , -0.25624692,
        0.38123697, -0.8611271 ,  0.97593486], dtype=float32)
```

(a) Feature vector of neutral emotion.

```
array([ 1.0726047 ,  2.0765452 , -1.2102355 ,  0.9298247 , -0.39932832,
       -0.0621765 ,  0.9391937 , -0.3909644 , -0.48584294, -0.39334473,
        0.25404388,  0.14166418,  1.6073494 ,  1.0019342 , -0.6300583 ,
        0.56408215,  0.9595349 , -1.3719865 , -0.5083765 ,  0.9105434 ,
       -0.5635817 ,  1.788767  , -1.2075716 , -2.2713118 ,  0.83336955,
       -0.6132851 ,  0.1673167 ,  1.1902683 , -1.3726363 ,  0.8577611 ,
       -0.45166358,  1.2178104 , -0.2897997 ,  0.02222073, -0.7031704 ,
       -0.9723001 ,  1.4285302 , -0.8973721 , -0.3461248 ,  0.6523386 ,
       -0.1217722 ,  1.4067686 , -0.7701404 , -0.02828353, -0.46949223,
       -0.91228235, -0.77604854,  1.0538529 ,  1.7177553 ,  2.7253191 ,
       -0.31236458, -0.47423446,  0.2724916 ,  0.6816371 ,  0.00752961,
       -1.8977752 ,  0.32059723,  0.15269436,  0.38035408, -1.1784856 ,
       -1.8317356 , -0.57486576,  0.88530725,  0.77867603,  0.76476747,
       -1.2387152 , -0.95252144, -0.40065426, -0.8675538 , -0.4208253 ,
       -0.620014  ,  0.87592673,  0.55212235, -0.5346438 , -1.1096131 ,
       -0.7980685 ,  1.6102626 ,  0.554409  ,  1.3810703 ,  0.46673143,
       -0.86199814,  0.46175542,  0.43205836, -0.19329019,  0.08156294,
        1.0343429 , -0.5331849 , -1.0448205 , -0.80823517,  0.4553213 ,
        1.3447698 ,  0.4261738 , -0.5237043 , -1.15949   ,  1.0062628 ,
        0.14108755,  1.0109067 ,  0.05116225, -1.3271422 , -1.0263956 ,
        0.9595174 ,  1.1975217 , -0.44956374, -0.37868613, -1.6132694 ,
        0.6666271 , -0.0179999 , -0.0744139 , -0.32557997, -0.250233  ,
       -0.17856063, -0.5039797 ,  0.49954644,  2.0019133 ,  0.69831055,
       -0.6319939 ,  0.32595542, -1.1039591 ,  0.8795818 ,  0.7095121 ,
        0.13396177, -0.23236245,  0.29136238, -1.5250018 , -0.4025234 ,
       -0.27308026,  0.51185846,  0.7411115 ], dtype=float32)
```

(b) Feature vector of help emotion.

Figure 16 Examples of feature vectors of a patient, both neutral and help.

4.2 Face Classification

There are two classification methods are applied to obtain the classification model. The first method is the use of machine learning mechanism where the logistic regression model was applied as the facial classification model. The second method is the use of neural network technique where feature vectors were input to 5 layers of neural to receive the prediction. Details for each method are elaborated as follow.

4.3 Machine Learning Process

To obtain the suitable logistic regression model, the dataset is divided to one training dataset and one testing dataset. For the training dataset, there are 120 neutral images and 120 help images from the same 4 patients. Thus, the remaining data, 60 neutral images and 60 help images from other 2 patients were saved to be the testing dataset.

Training process

After the feature vectors are derived, the 240 vectors in the training dataset are input to Logistic Regression Model so that the coefficients of the model can be identified. As the fact that the general logistic regression model can be written in the form of equation (4)., the result from the training process using dataset from 4 patients can demonstrated in Figure 17.

$$p = \frac{1}{1 + e^{-(a+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_{128} x_{128})}} \tag{4}$$

Where p is the probability to be success.

a is the intercept.

β_i is the regression coefficient.

```
Coef :  [[-0.05486602 -0.80919789 -0.95426703 -0.1996033  -0.00930306 -0.63314539
  -0.69373091 -0.78967475 -0.76535523  0.16293304  0.34164001 -0.16736286
  -0.55763142 -0.03876639  1.16907463 -0.62717736  0.67090354  0.43466674
  -0.22611172 -1.20372752 -0.46707255 -1.2975367  -0.50551679  0.30594658
   0.46099347  0.02917883 -0.03905272  0.1424076   0.27577074 -0.19754838
   0.53223774  0.68996208  0.62210757 -0.16831982  0.52560633  0.71445767
  -0.20297485  0.53863835 -0.36335544  0.45031044  0.94416688  0.23228123
  -0.27216372  0.42467696 -0.18123836 -0.92124593 -0.17881229  0.60057854
   0.70074535 -0.92127634  0.25008318 -0.61142827 -0.2257006   0.26188679
   1.19983113  1.37100812  0.05676472 -0.4028263   0.25319818  0.34525659
   0.3025241  -0.33591115 -0.20652002  0.34446612  0.56800976  0.07262619
  -0.59896138  0.91369152  0.0449839   0.09953913 -0.70326335  0.5554717
  -0.37444207 -0.84026262  0.51389087 -0.51641411 -0.13771729 -0.3499527
  -1.20821499 -0.20067263 -0.44435259  0.05524772 -0.82963835 -0.10916181
  -0.16664034 -0.19822642 -1.29798276 -0.01145817  0.31992416 -0.71670183
   0.4529907  -0.14656638  0.0397678  -0.40966491  0.27907019 -0.24847376
   0.51941986 -0.19369321  0.56997611  0.33530456 -0.1630171   0.433872
  -0.30134338 -0.31600169  0.87748097 -0.34121561  0.55077842 -0.9520326
   0.72379162 -0.53810933 -0.99904534  0.51491509 -0.64902781 -0.65407109
  -0.47650557 -0.3451892   0.28447287  0.44820609  0.41241379  1.01588122
  -0.0938749   0.30156761  1.70060916 -0.95958939  0.09354453  0.57919956
   0.11192854  0.69988715]]
Intercept :  [-0.30038423]
```

Figure 17 Values of intercept and all regression coefficients of the logistic regression model in the training process.

From this logistic regression model, the performance of the model has been evaluated. Figure.18 and Table 3 display the confusion matrix of the derived Logistic Regression Model in the training period.

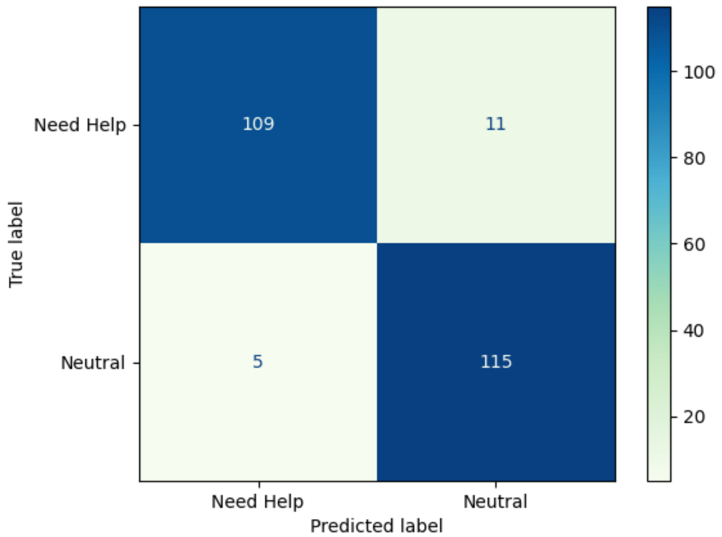

Figure 18 Confusion matrix of the Logistic Regression Model in the training period.

Table 3 Confusion matrix of Logistic Regression Model in the training period.

Predict / Actual	Help	Neutral
Help	109(TN)	11(FP)
Neutral	5(FN)	115(TP)

Based on the confusion matrix in Table 3, the true positive (TP) is 115, false positive (FP) is 11, true negative is 109, and false negative (FN) is 5. Consequently, other performance index such as precision, recall, accuracy, and F-1 score can be computed to indicate the efficiency of the obtained classification model.

First, the precision, the value that is used to measure the proportion of positive identifications was actually correct can be computed using equation (5).

$$Precision = \frac{TP}{(TP + FP)} \tag{5}$$

So, the precision for the Logistic Regression Model in the training state is $Precision = \frac{115}{(115+11)} = 0.9127$.

Second, the recall, the value that is used to identify the correctness of the proportion of actual positives. This value can be computed using equation (6).

$$Recall = \frac{TP}{(TP + FN)} \tag{6}$$

Therefore, the recall for the Logistic Regression Model in the training state is $Recall = \frac{115}{(115+5)} = 0.9583$

Third, the accuracy, the ratio of right predictions that the classification model can provide. The accuracy can be computed using equation (7).

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{7}$$

Therefore, the accuracy for the Logistic Regression Model in the training state is $Accuracy = \frac{115+109}{(115+109+11+5)} = \frac{224}{240} = 0.9333$

The last performance metric is the F-1 score. This F-1 score is the harmonic mean of precision and recall. This score is useful when distributions of precision and recall are unbalance. The F-1 score can be calculated from equation (8).

$$F - 1\ score = \frac{2 * Recall * Precision}{(Recall + Precision)} \tag{8}$$

Thus, the F-1 score for the Logistic Regression Model in the training state is $F - 1\ score = \frac{2*0.9583*0.9127}{(0.9583+0.9127)} = \frac{1.749}{(1.871)} = 0.935$

Testing process

To confirm that the logistic regression model obtained from the training process is efficient, the model must be validated using testing dataset in the testing process. In the testing process, the testing dataset has only 120 feature vectors: 60 neutral images and 60 help images. The result of this execution is the confusion as same as obtained in the training process. Figure 19 and Table 4 display the confusion matrix of the Logistic Regression Model in the testing period.

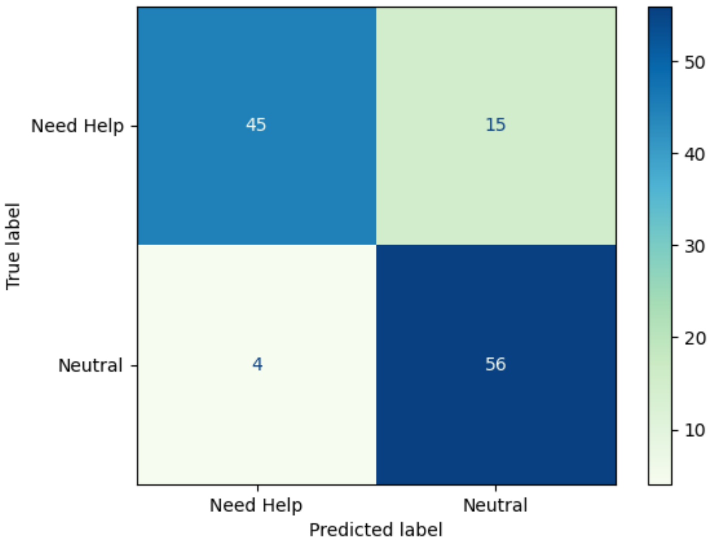

Figure 19 Confusion matrix of the Logistic Regression Model in the testing period.

Table 4 Confusion matrix of the Logistic Regression Model in the testing period.

Predict / Actual	Help	Neutral
Help	45(TN)	15(FP)
Neutral	4(FN)	56(TP)

Based on the confusion matrix in Table 4, the true positive (TP) is 56, false positive (FP) is 15, true negative is 45, and false negative (FN) is 4. Consequently, other performance index such as precision, recall, accuracy, and F-1 score can be computed to confirm the model's efficiency.

First, the precision, the value that is used to measure the proportion of positive identifications was actually correct can be computed using equation (5). So, the precision for the Logistic Regression Model in the training state is $Precision = \frac{56}{(56+15)} = 0.7887$

Second, the recall, the value that is used to identify the correctness of the proportion of actual positives. This value can be computed using equation (6). Therefore, the recall for the Logistic Regression Model in the training state is $Recall = \frac{56}{(56+4)} = 0.9333$

Third, the accuracy, the ratio of right predictions that the classification model can provide. The accuracy can be computed using equation (7). Therefore, the accuracy for the Logistic Regression Model in the training state is $Accuracy = \frac{56+45}{(56+45+15+4)} = \frac{101}{120} = 0.8417$

The last performance metric is the F-1 score. This F-1 score is the harmonic mean of precision and recall. This score is useful when distributions of precision and recall are unbalance. The F-1 score can be calculated from equation (8). Thus, the F-1 score for the Logistic Regression Model in the training state is $F-1\,score = \frac{2*0.9333*0.7887}{(0.9333+0.7887)} = \frac{1.472}{(1.722)} = 0.8548$

4.6 Neural Network Process

Another classification method is the neural network method. This method is like the machine learning method in the point that the dataset is separated to two different datasets: training dataset and testing dataset. For the training dataset, there are 120 neutral images and 120 help images from the same 4 patients. Thus, the remaining data, 60 neutral images and 60 help images from other 2 patients were saved to be the testing dataset. The expected outputs from the training process are weights of nodes in every layer of the network.

Training process

In the training process, all 240 feature vectors derived from 240 images of 4 patients were input to the first layer where 128 nodes are available. After the feature vectors were executed by activation function and solver function, the suitable weights of all nodes in every layer are determined with bias value of each layer. Since there are too many nodes to be implemented in this neural network, a fraction solution has been demonstrated in Figure 20.

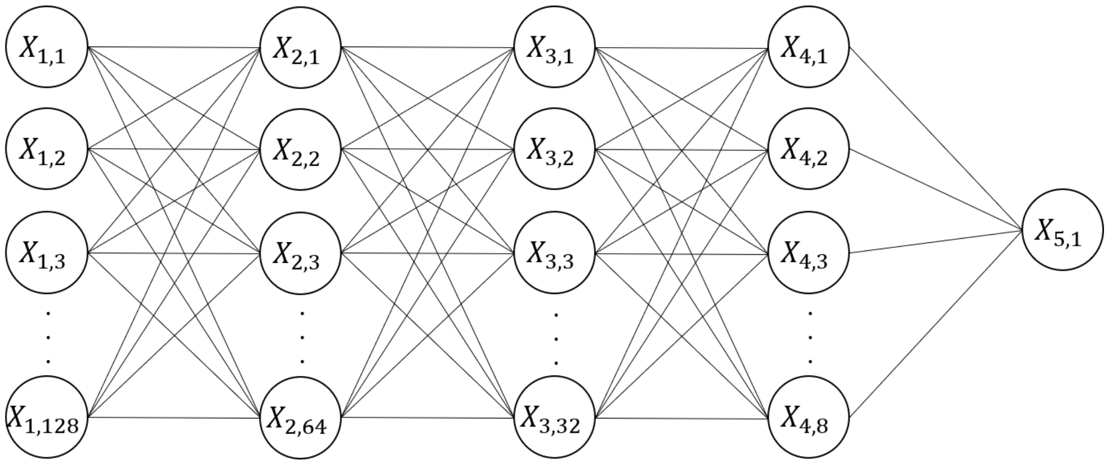

Figure 20 A fraction solution of the derived neural network.

Though all weights are defined with bias of every layer, the solution of this 5-layer neural network must be evaluated as same as the machine learning techniques. Therefore, the confusion matrix of this training process is drawn in Figure 21 and Table 5.

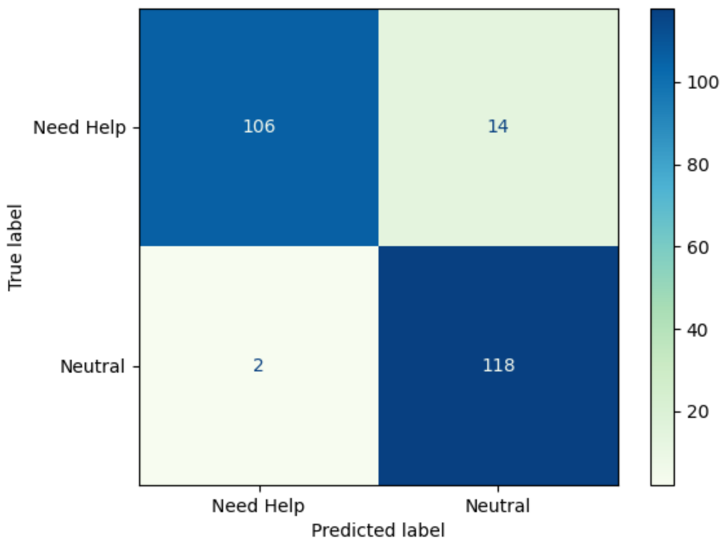

Figure 21 Confusion matrix of the 5-layer neural network in the training period.

Table 5 Confusion matrix of the 5-layer neural network in the training period.

Predict	Help	Neutral
Actual		
Help	106(TN)	14(FP)
Neutral	2(FN)	118(TP)

Based on the confusion matrix in Table 5, the true positive (TP) is 116, false positive (FP) is 14, true negative is 106, and false negative (FN) is 2. Consequently, other performance index such as precision, recall, accuracy, and F-1 score can be computed to confirm the model's efficiency.

First, the precision, the value that is used to measure the proportion of positive identifications was actually correct can be computed using equation (5). So, the precision for the Neural Network Model in the training state is $Precision = \frac{118}{(118+14)} = 0.8939$

Second, the recall, the value that is used to identify the correctness of the proportion of actual positives. This value can be computed using equation (6). Therefore, the recall for the Neural Network Model in the training state is $Recall = \frac{118}{(118+2)} = 0.9833$

Third, the accuracy, the ratio of right predictions that the classification model can provide. The accuracy can be computed using equation (7). Therefore, the accuracy for the Neural Network Model in the training state is $Accuracy = \frac{118+106}{(118+106+14+2)} = \frac{224}{240} = 0.9333$

The last performance metric is the F-1 score. This F-1 score is the harmonic mean of precision and recall. This score is useful when distributions of precision and recall are unbalance. The F-1 score can be calculated from equation (8). Thus, the F-1 score for the Neural Network Model in the training state is $F-1\,score = \frac{2*0.9833*0.8939}{(0.9833+0.8939)} = \frac{1.7579}{(1.8772)} = 0.9364$

Testing process

In the testing process, all 120 feature vectors derived from 120 images of 2 patients were input to the first layer where 128 nodes are available. The weighted neural network was tested using all 120 feature vectors and the confusion matrix of this testing process is drawn in Figure 22 and Table 6.

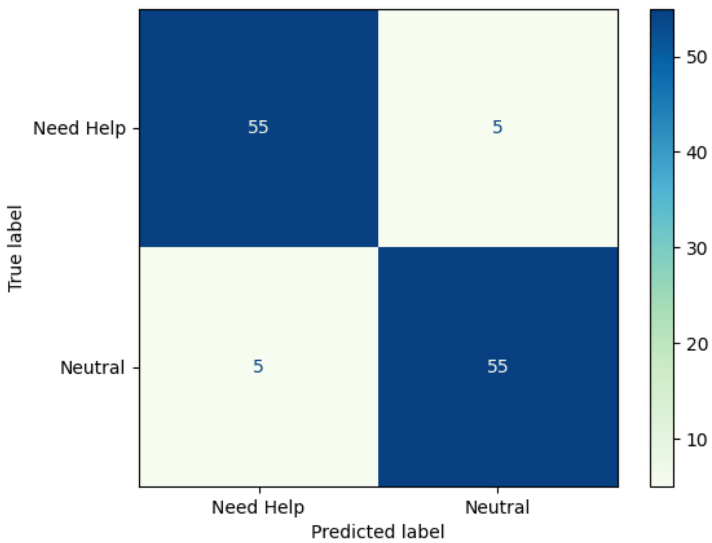

Figure 22 The confusion matrix of the testing process.

Table 6 The confusion matrix of the testing process.

Predict / Actual	Help	Neutral
Help	55(TN)	5(FP)
Neutral	5(FN)	55(TP)

Based on the confusion matrix in Table 6, the true positive (TP) is 55, false positive (FP) is 5, true negative is 55, and false negative (FN) is 5. Consequently, other performance index such as precision, recall, accuracy, and F-1 score can be computed to confirm the model's efficiency.

First, the precision, the value that is used to measure the proportion of positive identifications was actually correct can be computed using equation (5). So, the precision for the Neural Network Model in the training state is $Precision = \frac{55}{(55+5)} = 0.9167$

Second, the recall, the value that is used to identify the correctness of the proportion of actual positives. This value can be computed using equation (6). Therefore, the recall for the Neural Network Model in the training state is $Recall = \frac{55}{(55+5)} = 0.9167$

Third, the accuracy, the ratio of right predictions that the classification model can provide. The accuracy can be computed using equation (7). Therefore, the accuracy for the Neural Network Model in the training state is $Accuracy = \frac{55+55}{(55+55+5+5)} = \frac{110}{120} = 0.9167$

The last performance metric is the F-1 score. This F-1 score is the harmonic mean of precision and recall. This score is useful when distributions of precision and recall are unbalance. The F-1 score can be calculated from equation (8). Thus, the F-1 score for the Neural Network Model in the training state is $F - 1\, score = \frac{2*0.9167*0.9167}{(0.9167+0.9167)} = \frac{1.6807}{(1.8334)} = 0.9167$

4.7 Comparison between Machine Learning Model and Neural Network Model

Based on all computations in both sections, the performance evaluation between two different models can be compared in the following metric, Table 7.

Table 7 Comparison metric between the Logistic Regression Model and Neural Network Model.

Training period	Precision	Recall	Accuracy	F-1 Score
Logistic Regression Model	0.9127	0.9583	0.9333	0.935
Neural Network Model	0.8938	0.9833	0.9333	0.9364
Testing period	Precision	Recall	Accuracy	F-1 Score
Logistic Regression Model	0.7887	0.9333	0.8417	0.8548
Neural Network Model	0.9167	0.9167	0.9167	0.9167

From Table 7 in the testing period, it is obvious that the prediction performance of the logistic regression model is lower than the neural network model. To confirm this conclusion, the Area Under Curve of both models were drawn and compared.

The Area Under Curve analysis as known in the name of the AUC analysis is applied to classify binary classes operations: neutral and help. The AUC is the area under the Receiver Operating Characteristic curve (ROC curve), which is used to inform the prediction accuracy. Therefore, if the value of AUC is high then the accuracy of the prediction model is also high. The AUC curves of both models are drawn in Figure 23.

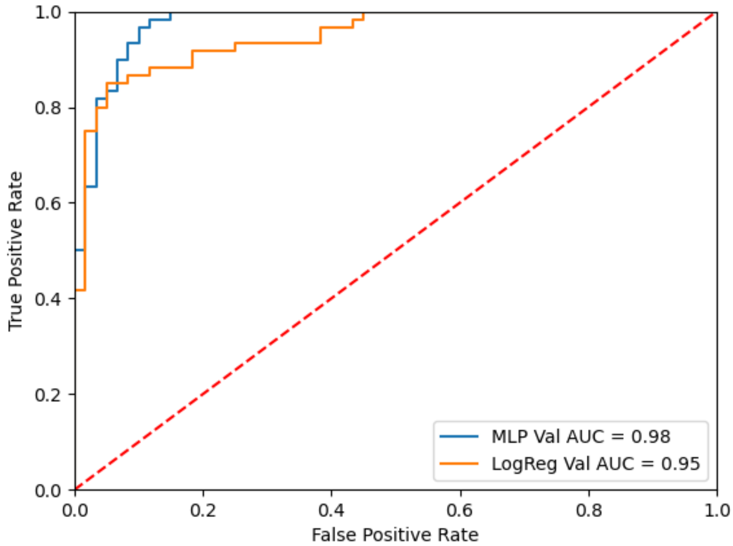

Figure 23 the Area Under Curve (AUC) of the performance prediction rate of logistic regression model and neural network model.

According to Figure 23, the Area Under Curve (AUC) of the performance prediction rate of the logistic regression model, the brown line, is 0.95 while the AUC of the performance prediction rate of the neural network model, the blue line, is 0.98. Thus, the neural network model provides a better prediction than the logistic regression model obtained from ML.

CHAPTER 5
DISCUSSION AND CONCLUSION

This chapter contains a discussion that demonstrates the distinguishing between the proposed solution in this research with previous research in Chapter 2. Moreover, the conclusion that summarizes the outcome of this problem finding is drawn. However, the limitations of this research are described, following with the future work.

5.1 DISCUSSION

Patients in the postoperative stroke are usually in the unconsciousness and need highly intensive care for few days; otherwise, some critical symptoms can occur. Therefore, medical staffs with high skills are assigned to responsible for this group. Unfortunately, the number of the required staffs is not enough although some patients may hire personal caregiver by their own. Thus, lacking the experience medical staffs or caregivers is a serious problem in the healthcare unit of the postoperative stroke patients.

Since there were various researchers worked in the facial classification, none of them can differentiate the patient's face between neutral and help situation. Presently, many researchers relied their works on either ML or neural network to perform their emotional classifications because of the complexity of the input data, such as changes of emotions (Chiang et al., 2012), sadness detection based head movement (Alghowinem et al., 2013), and sadness detection based EEG (Katyal et al., 2015). However, these works were also relied on expensive tools to collect data. Moreover, none of them run both neural network and ML to obtain the optimized classification model.

In contrast, when comparing with this proposed method that used to detect help-signature from the patient's face, only three rpi were installed, one personal computer, and WLAN. This proposed solution used common tools that can be found in the market with low cost of investment. The final emotion classification model is obtained from running the neural network after proving that the performance metric of the classification model from neural network is better than the classification model obtained from ML. In such case, the solution obtained in this research is much reliable than other research mentioned previously.

Quite similar to this research, the works related to pain detection have been performed by many researchers using ML for pain assessment test. As the work of (Neshov & Manolova, 2015) using support vector machine (SVM) and cross validation, the performance of the model was higher than 95%. Nevertheless, the training and testing data are obtained from the UNBC McMaster Shoulder Pain Archive database that contains only the shoulder pain situation and the expression of pain is easy to be seen from the face than unconscious postoperative patients. Thus, the difficulty to detect pain using the classification model from the shoulder pain should not fit for detecting the pain or help-sign of the postoperative brain surgery patients.

Another research work that was performed as the post-operation testing is the patients' pain with appendicitis who are bedridden (Sikka et al., 2015). The objective of this work was to study the change of patients' face when the lesion was hit. The facial pain-images were captured by the camera that put in front of the patients' face. Although this work considered the classification model from the facial pain, the dissimilarity of this work and the author's work is all facial pain-images from the appendicitis patients are obtained from the conscious persons while the patients of the postoperative brain surgery patients are the unconscious persons. Thus, the expression from the conscious persons is able to detect easier than the unconscious persons.

Table 8 shows the comparisons among various research and the proposed solution in this problem domain.

Table 8 Comparisons among various Emotion classification research.

Research	Classify emotion	Machine Learning	Neural network	Unconscious patient
(Chiang et al., 2012)	X	X	/	X
(Alghowinem et al., 2013)	/	/	X	X
(Katyal et al., 2015)	/	X	/	X
(Neshov & Manolova, 2015)	/	/	X	X
(Sikka et al., 2015)	/	/	X	X
Proposed solution	/	/	/	/

5.2 Conclusion

The situation of postoperative brain surgery patients is very critical as the medical staffs must monitor them all the time according to their unconsciousness. Therefore, it is difficult for these medical staffs to leave patients alone because some unexpected situation can occur; and patients cannot cry for help. So, to protect patients from the risk condition, many medical staffs or caregivers must be supported. Unfortunately, the number of required staffs is always not enough, and the risk condition becomes unavoidable.

Although several solutions have been proposed, none of them can fit the problem. Thus, this research proposed a solution that help medical staffs in the intensive care unit (ICU) of the postoperative brain surgery patients who cannot response to any stimuli but they may express their pain and their needs through their faces without showing any consciousness. The research used 6 patients as the samples in this experiment; 4 patients' dataset are used as the training and the rest is the testing dataset. Prior the computation, the manual classification of the item's movement under the patients' pain were drawn; and it can point out that these 6 patients moved their mouths and noses when they want to send a request-signal to the medical staff in the ICU.

Besides, this research applied both logistic regression model, an ML model, and the 5-layer neural network model to derive help-emotion classification model for patients. The performance of both models is measured using the confusion matrix where precision, recall, accuracy, and F-1 score are calculated. Since the objective of the logistic regression model supports the requirement in the emotional classification, the computations of precision, recall, accuracy, and F-1 score in the testing state are closed to 1.0, or 0.7887, 0.9333, 0.8417, and 0.8548, respectively. However, when calculate the precision, recall, accuracy, and F-1 score in the testing state of the 5-layer neural network, the obtained values are 0.9167, 0.9167, 0.9167, and 0.9167 which are higher than values of the logistic regression model. Therefore, the most suitable classification model for help-signature should be the 5-layer neural network that provide 91.67% accuracy.

5.3 Limitations

Since this research works with real patients at King Chulalongkorn Memorial Hospital, there are numerous problems to be faced.

1. Time for requesting the Institutional Review Board (IRB) which contains Research Ethic Committees (REC). It is the condition that before the research starts and corrects data at the ICU Stoke Hospital, all researchers must request the IRB from the REC of King Chulalongkorn Memorial Hospital. According to this condition, this research spend one year for requesting the IRB. Therefore, the research procedure started very slow because of this process.
2. Nurses at the ICU did not feel comfort with the camera and their clips. Thus, to protect the clips removal from the nurses, the researcher must fix the screws as best as possible.
3. Even though the IRB has been granted to this research, only small amount of patients' family agrees to participate in this project. After 8 months past, there were only 6 patients joined in this research, this number might be too small when comparing with other machine learning or neural network

problems. So, the data might not cover all cases of the postoperative brain surgery patients. Consequently, the obtained model may be overfitting.

4. The last limitation is about the Internet condition. As the fact that the system must be installed at the hospital for data collection, thus, the wireless network of the system belongs to the hospital. Regrettably, the reliability of the hospital network is low.

5.4 Future work

It is the fact that Thailand is moving to the aging society; many elderlies need responsible caregivers. Moreover, from the rate of being stroke in Thailand is also arising; some of them need brain surgery and being unconscious after the operation. According to the proposed solution, a system for automatic classification of help-signature from the patients with warning system can be created in the next step; so, the medical staffs can come and take proper action in time. As the result, the quality of the unconscious patients will be increase.

REFERENCES

Adibuzzaman, M., Ostberg, C., Ahamed, S., Povinelli, R., Sindhu, B., Love, R., . . . Ahsan, G. M. T. (2015). Assessment of Pain Using Facial Pictures Taken with a Smartphone. *Proceedings - International Computer Software and Applications Conference, 2*, 726-731. doi:10.1109/COMPSAC.2015.150

Alghowinem, S., Goecke, R., Wagner, M., Parkerx, G., & Breakspear, M. (2013). Head pose and movement analysis as an indicator of depression. *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, 283-288. doi:10.1109/ACII.2013.53

Avati, A., Jung, K., Harman, S., Downing, L., Ng, A., & Shah, N. H. (2017). Improving Palliative Care with Deep Learning.

Bailey, A., & Periyakoi, V. Where do americans die? Retrieved from https://palliative.stanford.edu/home-hospice-home-care-of-the-dying-patient/where-do-americans-die/

Chai, S., Weng, M. C., Lin, C. J., Chang, S. T., & Liu, Y. T. (2017). Establish a dynamic monitoring system for post-operative recovery room. *Proceedings of 2017 International Conference on Machine Learning and Cybernetics, ICMLC 2017, 2*, 521-527. doi:10.1109/ICMLC.2017.8108970

Chiang, H.-h., Chen, W.-m., Chou, C.-s., & Chao, H.-c. (2012). Facial Feature Matching. (Ispacs), 569-573.

De Benedittis, G., Lorenzetti, A., Migliore, M., Spagnoli, D., Tiberio, F., & Villani, R. M. (1996). Postoperative pain in neurosurgery: A pilot study in brain surgery. *Neurosurgery, 38*(3), 466-470. doi:10.1097/00006123-199603000-00008

Eggleston, K., Lu, M., Li, C., Wang, J., Yang, Z., Zhang, J., & Quan, H. (2010). Comparing public and private hospitals in China: Evidence from Guangdong. *BMC Health Services Research, 10*. doi:10.1186/1472-6963-10-76

Gottschalk, A., Berkow, L. C., Stevens, R. D., Mirski, M., Thompson, R. E., White, E. D., . . . Yaster, M. (2007). Prospective evaluation of pain and analgesic use following major elective intracranial surgery. *Journal of Neurosurgery, 106*(2), 210-216.

doi:10.3171/jns.2007.106.2.210

Hanchaiphiboolkul, S., Poungvarin, N., Nidhinandana, S., Suwanwela, N. C., Puthkhao, P., Towanabut, S., . . . Samsen, M. (2011). Prevalence of stroke and stroke risk factors in thailand: Thai epidemiologic stroke (TES) study. *Journal of the Medical Association of Thailand, 94*(4), 427-436.

Hossain, M. A., & Ahmed, D. T. (2012). Virtual caregiver: An ambient-aware elderly monitoring system. *IEEE Transactions on Information Technology in Biomedicine, 16*(6), 1024-1031. doi:10.1109/TITB.2012.2203313

Hosseini, H. G., & Krechowec, Z. (2004). Facial expression analysis for estimating patient's emotional states in RPMS. *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings, 26 II*, 1517-1520. doi:10.1109/iembs.2004.1403465

Katyal, Y., Alur, S. V., Dwivedi, S., & Menaka, R. (2015). EEG signal and video analysis based depression indication. *Proceedings of 2014 IEEE International Conference on Advanced Communication, Control and Computing Technologies, ICACCCT 2014*(978), 1353-1360. doi:10.1109/ICACCCT.2014.7019320

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine Learning and Data Mining Methods in Diabetes Research. *Computational and Structural Biotechnology Journal, 15*, 104-116. doi:10.1016/j.csbj.2016.12.005

Kuang, H., Najm, M., Chakraborty, D., Maraj, N., Sohn, S. I., Goyal, M., . . . Qiu, W. (2019). Automated aspects on noncontrast CT scans in patients with acute ischemic stroke using machine learning. *American Journal of Neuroradiology, 40*(1), 33-38. doi:10.3174/ajnr.A5889

Liu, N., Koh, Z. X., Goh, J., Lin, Z., Haaland, B., & Ting, B. P. (2014). Prediction of adverse cardiac events in emergency department patients with chest pain using machine learning for variable selection. *BMC medical informatics and decision making, 14*, 75-75.

Neshov, N., & Manolova, A. (2015). Pain detection from facial characteristics using supervised descent method. *Proceedings of the 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems:*

Technology and Applications, IDAACS 2015, 1(September), 251-256. doi:10.1109/IDAACS.2015.7340738

Rink, C., & Khanna, S. (2011). MicroRNA in ischemic stroke etiology and pathology. *Physiological Genomics, 43*(10), 521-528. doi:10.1152/physiolgenomics.00158.2010

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 07-12-June*, 815-823. doi:10.1109/CVPR.2015.7298682

Shetty, D., Rit, K., Shaikh, S., & Patil, N. (2018). Diabetes disease prediction using data mining. *Proceedings of 2017 International Conference on Innovations in Information, Embedded and Communication Systems, ICIIECS 2017, 2018-Janua*, 1-5. doi:10.1109/ICIIECS.2017.8276012

Sikka, K., Ahmed, A. A., Diaz, D., Goodwin, M. S., Craig, K. D., Bartlett, M. S., & Huang, J. S. (2015). Automated assessment of children's postoperative pain using computer vision. *Pediatrics, 136*(1), e124-e131. doi:10.1542/peds.2015-0029

Singh, M., Majumder, A., & Behera, L. (2014). Facial expressions recognition system using Bayesian inference. *Proceedings of the International Joint Conference on Neural Networks*, 1502-1509. doi:10.1109/IJCNN.2014.6889754

Tivatansakul, S., & Ohkura, M. (2015). Improvement of emotional healthcare system with stress detection from ECG signal. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2015-Novem*, 6792-6795. doi:10.1109/EMBC.2015.7319953

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Processing Letters, 23*(10), 1499-1503. doi:10.1109/LSP.2016.2603342

Zhou, Y., & Shi, B. E. (2017). Action unit selective feature maps in deep networks for facial expression recognition. *Proceedings of the International Joint Conference on Neural Networks, 2017-May*, 2031-2038. doi:10.1109/IJCNN.2017.7966100

VITA

NAME	Bhatarasiri Slakkham
PLACE OF BIRTH	Bangkok
INSTITUTIONS ATTENDED	He received a Bachelor's degree in Computer Science from Chulalongkorn University. Now he is a Master's degree student in Computer Science and Information Technology, Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University
PUBLICATION	A Monitoring and Analyzing Pain System for Postoperative Brain Surgery Patients Based Facial Detection