2D08 FORMULATION OF MACHINE LEARNING MODEL TO CLASSIFY HUMAN MOTION BASED ON KINEMATICS DATA OF DAILY ACTIVITIES

ANUAR BIN MOHAMED

BACHELOR OF ELECTRICAL ENGINEERING (ELECTRONICS) WITH HONOURS

UNIVERSITI MALAYSIA PAHANG

UNIVERSITI MALAYSIA PAHANG

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name : Anuar Bin Mohamed

Date of Birth : 5th February 1992

Title : Formulation of Machine Learning Model to Classify Human Motion Based on Kinematics Data of Daily Activities

Academic Session : Semester 1 2021/2022

I declare that this thesis is classified as:

- ☐ CONFIDENTIAL (Contains confidential information under the Official Secret Act 1997)*
- ☐ RESTRICTED (Contains restricted information as specified by the organization where research was done)*
- ☑ OPEN ACCESS I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Universiti Malaysia Pahang reserves the following rights:

1. The Thesis is the Property of Universiti Malaysia Pahang
2. The Library of Universiti Malaysia Pahang has the right to make copies of the thesis for the purpose of research only.
3. The Library has the right to make copies of the thesis for academic exchange.

Certified by:

anuar mohamed

(Student's Signature)

(Supervisor's Signature)

920205-08-5549
New IC/Passport Number
Date: 10th February 2022

Nur Aqilah Binti Othman (Dr.)
Name of Supervisor
Date: 11th February 2022

NOTE : * If the thesis is CONFIDENTIAL or RESTRICTED, please attach a thesis declaration letter.

THESIS DECLARATION LETTER (OPTIONAL)

Librarian,
Perpustakaan Universiti Malaysia Pahang,
Universiti Malaysia Pahang,
Lebuhraya Tun Razak,
26300, Gambang, Kuantan.

Dear Sir,

CLASSIFICATION OF THESIS AS RESTRICTED

Please be informed that the following thesis is classified as RESTRICTED for a period of three (3) years from the date of this letter. The reasons for this classification are as listed below.

Author's Name
Thesis Title

Reasons (i)

(ii)

(iii)

Thank you.

Yours faithfully,

____
(Supervisor's Signature)

Date:

Stamp:

Note: This letter should be written by the supervisor, addressed to the Librarian, *Perpustakaan Universiti Malaysia Pahang* with its copy attached to the thesis.

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the Bachelor of Electrical Engineering (Electronics) with Honours.

____

(Supervisor's Signature)

Full Name : Nur Aqilah Binti Othman (Dr.)

Position : Senior Lecturer

Date : 11th February 2022

____

(Co-supervisor's Signature)

Full Name :

Position :

Date :

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

anuar mohamed

(Student's Signature)

Full Name : Anuar Bin Mohamed

ID Number : EA18122

Date : 10th February 2022

FORMULATION OF MACHINE LEARNING MODEL TO CLASSIFY HIMAN MOTION BASED ON KINEMATICS DATA OF DAILY ACTIVITIES

ANUAR BIN MOHAMED

Thesis submitted in fulfillment of the requirements
for the award of the Bachelor of
Electrical Engineering (Electronics) with Honours

Faculty of Electrical & Electronics Engineering

UNIVERSITI MALAYSIA PAHANG

FEBRUARY 2022

ACKNOWLEDGEMENTS

First and foremost, Alhamdulillah, thanks to Allah S.W.T, the Almighty, for His showers of blessings throughout my research work to complete the research successfully.

Most significantly, I would like to convey my sincere appreciation to Dr. Nur Aqilah Binti Othman, Senior Lecturer at Faculty Technology Electric Engineering Universiti Malaysia Pahang, for providing me with the chance to conduct research and for offering crucial advice during this process. Her energy, vision, genuineness, and determination have left an indelible impression on me. Her knowledge on machine learning has been huge assist for me to complete this project. She taught me the technique for conducting the study and analysing the data as briefly as possible. learning and researching under her supervision was a tremendous honour and respect. I am deeply appreciative for everything she has provided for me. I would also like to thank for her friendship, empathy, and great concern.

I also thank all the coordinator, staff and management of College Engineering, Universiti Malaysia Pahang for their genuine support to complete this thesis successfully.

I am extraordinarily thankful to my parents for their love, prayers, and sacrifices for educating and preparing me for my destiny. Thus, I also appreciate every moment and memories together with my friends that convincing me to continue and completing this degree.

Finally, my thanks go to all the people who have supported me to complete the research work directly or indirectly.

ABSTRAK

Projek ini bertujuan untuk formulasikan kaedah dalam “Machine Learning” untuk mengklasifikasikan aktiviti individual dalam seharian berdasarkan data harian. “Machine Learning” sering kali digunakan dalam teknologi masa kini. Malaysia mempunyai masalah kesihatan yang disebabkan oleh berat badan yang berlebihan. Pelbagai kempen dilakukan untuk mengatasi masalah ini. Antaranya adalah penggunaan jam pintar yang boleh merekod aktiviti seseorang itu dalam bentuk data. Sebagai contoh, jam pintar ini boleh digunakan dalam kehidupan seharian untuk merekod berapa jauh(kilometer) seseorang itu berjalan atau berlari. Objektif dalam projek adalah lokasi yang optimum untuk sensor pada tubuh badan individu. Selain itu, ia bertujuan untuk mengenal pasti kaedah yang paling sesuai dalam “Machine Learning” untuk mengklasifikasikan pergerakan induvidu itu. Seterusnya, ia bertujuan untuk menyiasat ketepatan dalam klasifikasi.

Projek ini menggunakan IMU-MMS sebagai alat untuk merekod data pergerakan individu. Ia akan diletakkan pada lengan, badan dan kaki sebagai lokasi sebelom data direkodkan. Parameter yang direkodkan adalah tiga paksi (x,y,z) “accelerometer” dan tiga paksi (x,y,z) “gyroscope”. Data ini kemudian akan dianalisiskan dalam perincian seperti Orange dan Matlab. Kaedah-kaedah dalam ML akan digunakan seperti SVM,KNN, “neural network” dan “Naïve Bayes” untuk melatih data ini dalam mengklasifikasikan jenis pergerakan individu itu. Selepas itu, penalaan dan analisis dilakukan to mendapatkan parameter yang terbaik bagi gambaran data.

Setelah melakukan Analisa dalam Matlab dan Orange dengan memasukkan 7200 data yang terdiri daripada berjalan kaki, berlari, berdiri, duduk, menaip dan melompat, secara keseluruhannya, pergelangan tangan adalah lokasi yang paling sesuai untuk IMU bagi merekod data pergerakan manusia. “Neural Network” adalah model dalam ML yang memberikan ketepatan paling tinggi berbanding model-model yang lain dalam Orange. Manakala kNN adalah model yang terbaik dalam Matlab. Ketepatan yang dicapai oleh kebanyakan model adalah 90% dan keatas.

ABSTRACT

This project aims to formulate a method in "Machine Learning" to classify individual activities in the day based on daily data. "Machine Learning" is often used in today's technology. Malaysia has health problems caused by being overweight. Various campaigns were conducted to address this problem. Among them is the use of smart watches that can record a person's activities in the form of data. For example, this smart watch can be used in daily life to record how far (kilometers) a person walks or runs. The objective in the project is to find out the optimal location for the sensor on the individual body. In addition, it aims to identify the most appropriate method or model in Machine Learning to classify the activities of the individual. Next, it aims to investigate the accuracy in the classification.

The project uses IMU-MMS as a sensor to record individual movement raw data. It will be placed on the wrist, back of body and shank as a location before the data is recorded. The recorded parameters are three axes (x, y, z) "accelerometer" and three axes (x, y, z) "gyroscope". The raw data will be then downloaded in csv file to the computer and then it will be analyzed in details such as Orange and Matlab. Methods in ML will be used such as Support Vector Machine, k-Nearest Neighbor, Neural Network and Naïve Bayes to train this data in classifying the type of movement of the individual. After that, tuning and analysis is done to get the best parameters for the data picture.

After performing Analysis in Matlab and Orange by inserting 7200 data which consists of walking, running, standing, sitting, typing and jumping, overall, the wrist is the most optimum location for the IMU to record human motion or activites. "Neural Network" is the model in ML that provides the highest accuracy compared to other models in Orange. While kNN is the best model in Matlab. The accuracy achieved by most models is above 90%.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

IMU	Inertial Measurement Unit
PCA	Principal Component Analysis
ML	Machine Learning

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
KJF	Knee Joint Forces
SVM	Support Vector Machine
KNN	k-Nearest Neighbour
ANN	Artificial Neural Network
NB	Naïve Bayes
RBF	Radial Basis Function
ReLu	Rectified Linear unit

CHAPTER 1

INTRODUCTION

1.1 Project Background

Malaysia has the highest rate of obesity and overweight among Asian countries with 64% of male and 65% of female population being either obese or overweight. Wearable devices are widely used nowadays since it can track daily activities and provides information to the users with regards to their daily activities. This, in turn, will create awareness among users on whether they are active or not in that day. Many studies have been conducted to classify the motion during specific movement in sports. Nonetheless, the studies on classification of daily activity motion are still scarce. Therefore, this study aims to formulate a machine learning model to classify of human motion in typical daily activities. This will provide the user more information on their lifestyle, which will create awareness among users and promote healthy lifestyle. Moreover, the sensor placement on the body is another issue. This study will investigate the optimal location of sensor placement by putting the sensors on different locations such as wrist, back and shank. The data collected from these locations will be used as input of machine learning classification model to study which location provides the best classification accuracy. Further, this study seeks to answer the question which machine learning model and which model tuning strategy would provide the best classification accuracy when it comes to daily activity motion classification. The study will be conducted in three phases;

(i) data collection process

(ii) model formulation based on recorded kinematics data, and

(iii) tuning of the model.

The expected outcome of this work is a machine learning model capable of classifying daily activities with high accuracy. This is hoped to promote healthy lifestyle among Malaysians, thus bring down the number of obesity cases in Malaysia.

1.2 Problem Statement

Malaysia has the highest rate of obesity and overweight among Asian countries with 64% of male and 65% of female population being either obese or overweight. This is an alarming situation that needs to be properly addressed by relevant stakeholders such as the government agencies and NGOs. One of the causes of this problem is the unhealthy lifestyle among Malaysians which involves non-healthy food and seldomly exercise. There is a need to promote a healthy lifestyle among Malaysians by creating an awareness of their activities throughout the day, as well as how much calories they have burnt as compared to how much calories they have taken each day. Wearable technology is being widely used nowadays as fitness tracker. More and more smart bands and smart watches have been introduced into the market. In 2021, wearable industry is valued USD116.2 Billion and is anticipated to reach USD265.4 Billion in 2026. More and more people are wearing smart bands or smart watches nowadays. This shows that it has become a trend that people would like to know about how their days went, activity-wise. People now love to analyse how many steps they have taken in a day, and how many calories they have burnt. When it comes to runners, they would like to know how far they have run and also the calories burnt. However, the statistics provided by these wearables, most of the times are limited to only number of steps and calories. It would be better if the wearable could provide the person wearing it a detailed statistics of his/her movements in a day, such as how far this person has walked, how far has he runs, how many stairs did he climbed, how long was he sitting down, how long was he typing on a computer or laptop. This detailed information can be made available through a classification of these different type of activities by means of a machine learning model. To formulate a model to classify the motion, huge amount of data is required. Here comes another research question, where should we put the sensor (inertial measurement unit) on the body? This is another research area that has not been really understood. Some studies experimented with the sensor placed on the chest, while another study put the sensor on the back. A more common sensor placement is on the wrist. This is another focus of this study, that is to find where is the best location to put the sensor, and to understand why this location provides a better data for activity classification. The ability to classify daily activities will allow this information to be easily accessible for the user. This will in turn create an awareness of one's daily locomotion, thus promote an active lifestyle.

1.3 Objectives

The goal of this project is to find sensor optimal location. It is still unsure where is the best location for the sensor on the body. As the body has lots of moving parts such as wrist, shank and foot, these three locations can be best location for the sensor. Since this project to classify human’s activities such as walking, running, climbing, sitting down and lying down, every location on the body has the best data classification for every of these activities. If the sensor is put on shank, it must be decided to place it on the right or left shank. By placing the sensor on the location mention above, data classification can be obtained by getting the classification accuracy. This result determines where the best location on the body for raw data collection by using Inertial Measurement Unit (IMU) sensor.

In this project, one of the most important aspects is formulation model in machine learning (ML). Support vector machine, k-Nearest Neighbour, Neural Network and Naïve Bayes are the common techniques used in ML. This study aims to determine the best ML model or algorithm to classify human’s activities.

Lastly, the objective in this study is to investigate the classification accuracy. This includes the process of tuning in variables and data analysis. Data split will be in data training and testing data. This proportion plays crucial role before training the data. The best variable in dataset will be determined to perform the best accuracy for four ML models mentioned above.

1.4 Scopes

Overweight and obesity are major problems in Malaysia nowadays. Because of non-healthy lifestyle, people in Malaysia suffer from non-communicable disease (NCD) like cancer. Device such as smart watch can be used to supervised human’s activities throughout the day. They can use this device to measure how long (in kilometre) walk, jog and run. This device can also measure the data for climbing, sitting down, lying down and other activities.

This project aims to use sensor which is IMU-MMS as hardware and software for data analysis and classification to predict and train the data so human's activities can be visualised into a pattern. The sensor is placed on specific body location to record and obtain the optimum raw data which gives the best result. This location will be switched after completing all data measurements for all activities. Targeted data volume in human's motions is approximately 7200 data.

After the raw data is collected, the data will then be saved in CSV file and transferred via Bluetooth of the IMU-MMS to the computer. Orange software does predict and train the raw data collected into model development which are SVM, KNN, neural network and NB. The other alternative to train the data is Matlab as it contains Classification Learner Application to train and test the data.

Further analysis and tuning are carried out to obtain the best classification for human's activities. This means the best 2-axis parameter will be determined to plot or visualise the data pattern. Thus, the classification accuracy of SVM, KNN, neural network and NB can be achieved at their best. By executing this, human's motions can be classified accurately every time he/she switches type of activities such as from walking to jumping or from standing up to sitting down.

CHAPTER 2

LITERATURE REVIEW

Most of the cases use machine learning to supervised individual body movements especially in sports. Some of them are used to supervised post-injury knee movement of a patient. Activity in a gym such as weightlifting is also supervised by using ML. Observing, classifying and assessing human movements is important in many applied fields, including human-computer interface, clinical assessment, activity monitoring and sports performance (Preatoni et al., 2020). Participants performed a variety of movements, including linear motions, changes of direction, and jumps(Stetter et al., 2019).

2.1 Activity/Gait Movement

Machine learning methods have been widely used for gait assessment through the estimation of spatial-temporal parameters (Mannini et al., 2016). As a further step, the objective of this work is to propose and validate a general probabilistic modelling approach for the classification of different pathological gaits. Specifically, the presented methodology was tested on gait data recorded on two pathological populations (Huntington's disease and post-stroke subjects) and healthy elderly controls using data from inertial measurement units placed at shank and waist (Mannini et al., 2016). ML is used to predict ground reaction forces (GRFs), which could serve as quantitative indicators of sports performance or rehabilitation (Lim et al., 2020). Most of the literature did ML for walking activity. But the difference is the parameters that are measured for particular purposes. For particular study, they are some researchers who determine the estimation of Knee-Joint Forces (KJF).

Table 2.1 Description of the five walking conditions

Walking condition	Description
Level ground walking	The subjects walked along a 10-m straight path inside the gait laboratory
Upslope treadmill walking	The subjects walked on a +15° inclination on the treadmill
Downslope treadmill walking	The subjects walked on a -15° inclination on the treadmill
Stair ascent	The subjects climbed up a staircase of six steps. Each step had a height of 15 cm and a breadth of 20 cm
Stair descent	The subjects descended the same staircase as for stair ascent

Source: (Lau et al., 2008)

2.2 Equipment

The raw data is collected by using wearable sensor. This is then applied to a number of participants. Inertial Measurement Units (IMU) seems to be the most relevant device to record or measure the raw data. Most of the researcher used this device on participants for gait movements. As a consequence, alternative technologies, such as wearable IMU, have experienced tremendous advances within the last two decades (Stetter et al., 2019). Wearable technologies for motion analysis are predominantly IMU which, thanks to their low cost and minimal obtrusiveness, represent an optimal solution for tracking and assessing sports movement on-field (Preatoni et al., 2020). IMU includes combination of accelerometers and gyroscopes have been successfully used for assessing gait characteristics (Mannini et al., 2016). In particular cases, foot-ground reaction forces in the vertical and anterior–posterior directions were recorded using two forces sensing platforms (Begg & Kamruzzaman, 2005). Movement of the lower limb was recorded using a 3D PEAK Motion analysis system via reflective markers attached to lower limb joints and segment (Begg & Kamruzzaman, 2005).

2.3 Location of Sensor

The sensor units were attached at two locations on the lower limb. One sensor unit was attached at the tibial tuberosity of the shank. The second sensor unit was attached at the back of the heel on the shoe (Lau et al., 2008). This sensor locations are used for walking activity which all conditions. In KJF case, thirteen participants were equipped with two IMUs located on the right leg (Stetter et al., 2019). With these sensors attached on the leg, participant can perform activity like linear motions, change of direction and jumps. For activities for functional fitness exercise like workout drill, there a lot of locations on the body for sensor placement. IMUs were located onto specific anatomical landmarks, which included the left ankle, thigh, upper arm and wrist, and trunk (Preatoni et al., 2020). These positions were chosen to reproduce the locations where commercially available devices with embedded motion monitors. This also aims to capture whole body information and drill dynamics whilst allowing the natural execution of movements, avoiding obstruction or discomfort for the subject. (Preatoni et al., 2020).

2.4 Machine Learning Model

Most of cases select SVM and Artificial Neural Network (ANN) as their model because these activity classifications are more reliable in term of algorithm and also result more classification accuracy. However, there are also algorithm used in their study such as KNN. All these model developments are processed by using Python Software like Keras, TensorFlow, Scikit-learn, NumPy and Matlab. After the extraction of the features and the labelling of associated windows, different type of automatic classifiers was trained by using dataset. K-Nearest Neighbours (KNN), with different types of metrics and number of neighbours and Support Vector Machine (SVM) with several types of kernel functions were selected as the classifying algorithms to be tested (Preatoni et al., 2020). Accelerations and angular velocities were acquired continuously from the units and used to train and test different supervised learning models, including k-Nearest Neighbours (KNN) and support-vector machine (SVM) algorithms. The use of different kernel functions, as well as different strategies to segment continuous inertial data were explored (Preatoni et al., 2020). The ANN could predict unmeasured force information from the IMUs, find the global optima of the loss function during backpropagation, generalize without overfitting, and select appropriate input types and size (Lim et al., 2020)

2.5 Parameter/Variable

Accelerations and angular velocities were acquired continuously from the units and used to train and test different supervised learning models, including k-Nearest Neighbours (KNN) and support-vector machine (SVM) algorithms. The use of different kernel functions, as well as different strategies to segment continuous inertial data were explored (Preatoni et al., 2020). Each accelerometer measured both the linear acceleration of the limb it was attached to and an acceleration component due to gravity. The gyroscope measured the angular velocity through the Coriolis force. The rate of increase of the tangential speed which caused by the radial velocity, is the Coriolis acceleration (Lau et al., 2008).

2.6 Result

It appeared that SVM always performed with the highest accuracy for all of the classification tasks, and it achieved 100% classification for two-class and three-class problems. Meanwhile, the classification accuracy of SVM was found to be monotonously increasing as the number of classes was reduced (Lau et al., 2008). ANN is quite reliable in determining human gait as it gives good accuracy. The estimation accuracy of the ANN varied between movements, but that accuracy was good for most movements. With respect to the three different force components, vertical KJF showed the highest agreement between the ANN-predicted outcomes and the inverse dynamics-calculated data, followed by the anterior–posterior KJF, and finally the medial–lateral (Stetter et al., 2019). In gym workout case, human gait movements were classified with high accuracy. When data input included all the five available sensors, both SVM- and KNN-type classifiers achieved good level of overall accuracy (Preatoni et al., 2020).Accuracy ranged from 82.5% (SVM classifier with fine gaussian kernel, and 300 ms−10% overlap windows) to 97.8% (Preatoni et al., 2020). SVM would great approach to classify motions as it can seek the optimal parameter. In the former step, a projection function is designed to mapping the high-dimensional data into a low dimension subspace (Zhu, Zhang, Zho, 2017)

Table 2.2 Literature Findings for Human Motions and Machine Learning Models

Activity	Paper	Model/Technique	Equipment	Position of sensor	Output/Parameter of study	Result/ Finding
walking	Prediction of Lower Limb Kinetics and Kinematics during Walking by a Single IMU on the Lower Back Using Machine Learning, SENSOR, 2020	ANN (artificial neural network) Keras, tensorflow	IMU (inertial measurement units) (sensor)	HipCoM Near sacrum	Segment angles of 1. Foot 2. Shank 3. Thigh of stance, swing, torque, GRF (x,y)	7 subjects walking on treadmill, GRF & joint motion as weighted sum of the COM, IMU measurement used in pre-processed to identify each step to obtain position and velocity
walking	Support vector machine for classification of walking conditions using miniature e kinematic sensors,	SVM Software : Scikit-learn, numpy	Gyrometer & Accelerometer	Shank , foot	Level ground, upslope, downslope, stair ascent, stair descent .Signal in form of Shank angular velocity, shank acc,	SVM more accurate compared to other technique 2-class comp., 3-class comp., 5-class comp.

	Springer, 2008				foot acceleration	
Knee Joint Forces	Estimation of Knee Joint Forces in Sport Movements Using Wearable Sensors and Machine Learning, Sensors, 2019	ANN	2 IMUs	Right leg via knee sleeve	Fv(vertival), Fap(anterior-posterior), Fml(medial-lateral), -> GRF data Matlab,	Of 16 movements, 13 had %Diff for summed Fv smaller than 6.8%. +ve correlation, F∗v for moderate running showed the highest correlation with Fv (0.94 ± 0.33).
Workout drills(clean & jerk, american swing, Box Jump, Burpee,	Supervised Machine Learning Applied to Wearable Sensor Data can Accurately Classify Functional Fitness Exercises Within a Continuous Workout, Frontiers, 2020	kNN & SVM	IMU, tri-axial accelerometer, gyroscope and magnetometers – base station via Bluetooth Low Energy	Upper arm, thigh, left ankle, wrist & trunk	Acceleration & angular velocity 1. Clean & jerk 2. American swing 3. Box jump 4. Burpee	Both knn and svm achieved overall accuracy. (82.5% - 97.8%). Only in the AS, the classifier seemed to perform relatively better when using data from the

						sensor placed on the lumbar spine (single sensor configuration.). The worst overall performance was obtained when considering the data acquired by the only sensor placed on the ankle.
Walking disabilities (drop foot)	The reliability of using accelerometer and gyroscope for gait event identification on persons with dropped foot, Elsevier, 2007	threshold detection in which ROC analysis employed. Receiver Operating Characteristic curve	Accelerometers and gyroscopes	Thigh, shank, foot	Linear acceleration & angular velocity, FSR Ap = anterior posterior SI = superior inferior	Fsr obtained 20 gait cycles. For subjects with dropped foot, the average duration of the stance and swing phases were 0.757 s (S.D. =

					Walking speed, swing time, stance time for : 1. Thigh 2. Shank 3. Foot	0.116 s) and 0.603 s (S.D. = 0.137 s), respectively. For non-impaired subjects, the average values were 0.632s (S.D. = 0.002 s) and 0.399 s (S.D. = 0.048 s), respectively (Table 2). Patients had a slightly prolonged swing phase (44.04%) compared with controls (38.61%). The duration for a complete gait cycle was longer for patients (

Walking back and forth for about one minute along a 12-m walkway	A Machine Learning Framework for Gait Classification Using Inertial Sensors: Application to Elderly, Post-Stroke and Huntington's Disease Patients, Sensors, 2016	ANN and SVM Training of class-specific HMM (Hidden Markov Models)	3 IMUs, 1 tri-axial accelerometers and 1 tri-axial gyroscope	Both shanks	Post stroke, huntingtons's disease, healthy elderlyML and VT accelerations .(Medio-lateral,vertical) 1. HMM based features 2. Time domain features 3. Frequency Domain Features HMM, SVM, MV	both group-specific HMMs and SVM classifiers were evaluated using an LOSO cross-validation approach. To improve the final prediction and summarize the results, a majority voting (MV) strategy was finally applied. The MV step summarized results by smoothing the effects of intra-subject variability

						, providing a high overall accuracy in terms of subjects classification up to 90.5%.
sitting, sitting down, standing up, standing, and walking	Recognition of Human Activities using Machine Learning Methods with Wearable Sensors, IEEE,	SVM, HMM , ANNSoftware: matlab	wearable accelerometers, 12 features	left thigh, right arm, right ankle and abdomen	overall 12 values were read for each subject, which means 12 features were generated for each subject.	Future work on SVM approach would be to find a more efficient way to find the optimal parameters rather than using cross validation since it is time consuming.All models score above 90%.
standing, sitting, lying,	Human Daily Activity Recognitio	CNN – deep learning	MPU6050	waist	X,y,z-axis for acceleromet	the accuracy reached 93.77%,

walking upstairs, walking downstairs, and walking	n Performed Using Wearable Inertial Sensors Combined With Deep Learning Algorithms, IEEE Access, 2020				er & gyroscope	the precision was 93.82%, the recall was 93.82%, and the F1-Score was 93.82,
Walk, sitting, lying, stair climb, standing, running, cycling	Machine Learning Methods for Classifying Human Physical Activity from On-Body Accelerometers, Sensors, 2010	Hidden Markov Models, NB, SVM, kNN, ANN	IMU	hip, wrist, arm, ankle, and thigh	Triaxial Accelerometer Magnetometer	Classification accuracy (%): NB - 97.4 SVM - 97.8 kNN-98.3 ANN-96.1 HMM-99.1
Walk, jump, run, sprint	Human Motion Recognition by Textile Sensors Based on Machine	ANN, MANN, SVM, Random Forest	single-walled carbon nanotubes, spandex fabric	thigh	Average amplitude, standard deviation, averave cycle	Accuracy is achieved (%): RD – 90

	Learning Algorithms , Sensors, 2018					SVM-84 ANN-85 MANN - 88

CHAPTER 3

METHODOLOGY

3.1 Fundamental/Theories

Artificial Intelligence (AI), the appearance of intelligence in machines, is a huge topic today's technology. It is getting bigger day by day. The way people interact with mobile phone through speaking is an example of AI. Machine learning, a subset of artificial intelligence, refers to systems that can learn by themselves. It involves teaching a computer to recognize patterns, rather than programming it with specific rules (Mueller & Massaron, 2016) . The training process involves feeding large amounts of data to the algorithm and allowing it to learn from that data and identify patterns. Machine Learning (ML) is part of technology that is used to use algorithm to determine the pattern of the data. Machine Learning does rely on algorithms to perform data analysis. By doing this, the people can actually see the pattern on how the walks, runs or climbing stairs. These algorithms perform predictive analysis faster than human being. As a result, machine can definitely help human to work more efficiently.

Many people confuse that ML and Statistics are the same things. The fact is ML works with huge data in forms of networks and graph. That means the raw data is obtained by the sensors and then split into both training and test data. Meanwhile statistic use models to create predictive power on small samples. In Machine Learning, the data obtained is sampled, randomized and transformed to maximize the accuracy score. The basis for machine learning is math which means the algorithms are used to determine how to interpret big data in specific ways (Mueller & Massaron, 2016). The reason machine learning is used because it can work out the data so that the pattern can be visualised to see the classification. Machine learning does depend on Python and R to decipher the data.

3.1.1 Support Vector Machine

People use support vector machine for linear and nonlinear classification, regression and unsupervised detection of outliers (Mueller & Massaron, 2016).

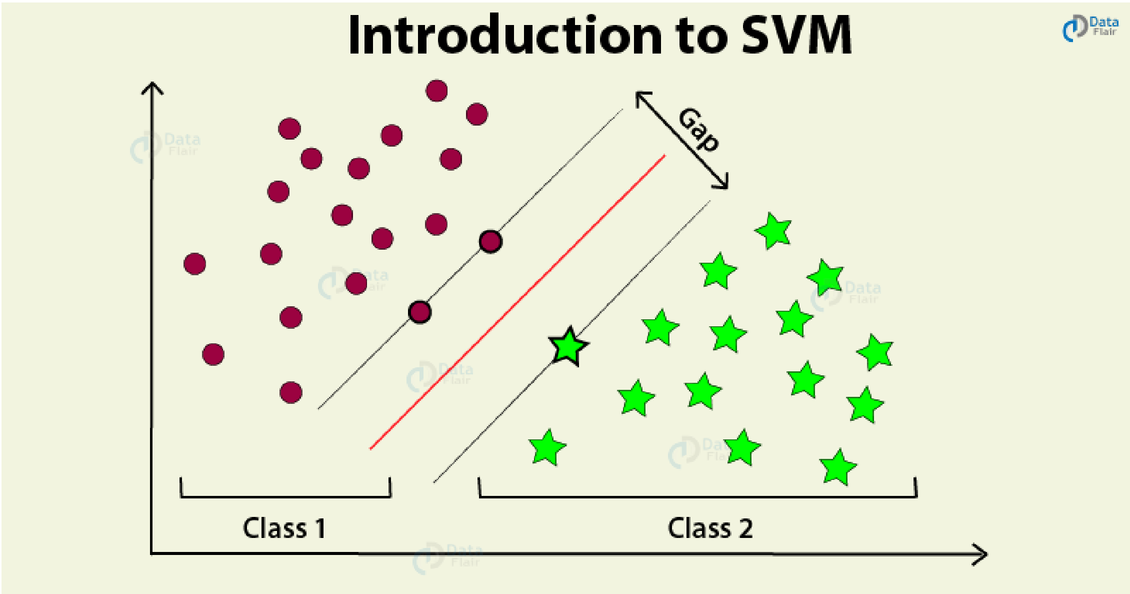


Figure 3.1: SVM classification

Source: https://data-flair.training/blogs/svm-support-vector-machine-tutorial/

By observing the line difference in Figure 3.1, the gap is the margin which separates the gutter of each class. Using SVM terminology, the separating line is the one with the largest margin which places the separating line in the middle of the margin (Mueller & Massaron, 2016). SVM optimization process can be expressed into:

$$y(X^T w + b) \geq M(1 + \epsilon_i$$

y: the response values which can be valued 1 or -1

X^T: the transposed features (the dimensions of the matrix are switched)

w: the weighting values (a vector of coefficients)

b: the bias (a constant)

M: the margin (the optimal value for classification)

ϵ_i : the slack variable, a correction value

In the optimization process, each example's prediction is adjusted by a specific epsilon value (a value equal or greater than zero), in such a way that as many wrong classifications as possible are corrected (Mueller & Massaron, 2016). If an example is already correctly classified, then the assigned epsilon is 0. Values of epsilon between 0 and 1 indicate that the example is still correct yet it is inside the margin. Finally, when the assigned epsilon is above 1, the case is misclassified and is located on the wrong side of the optimal separating hyperplane. Epsilon allows the algorithm to correct mismatched cases so that the optimization process will be more successful (Mueller & Massaron, 2016).

3.1.2 K-Nearest Neighbours

The k value, an integer number, is the number of neighbours that the algorithm has to consider in order to figure out an answer. The smaller the k parameter, the more the algorithm will adapt to the data you are presenting, risking overfitting but nicely fitting complex separating boundaries between classes. The larger the k parameter, the more it abstracts from the ups and downs of real data, which derives nicely smoothed curves between classes in data, but does so at the expense of accounting for irrelevant examples (Mueller & Massaron, 2016).

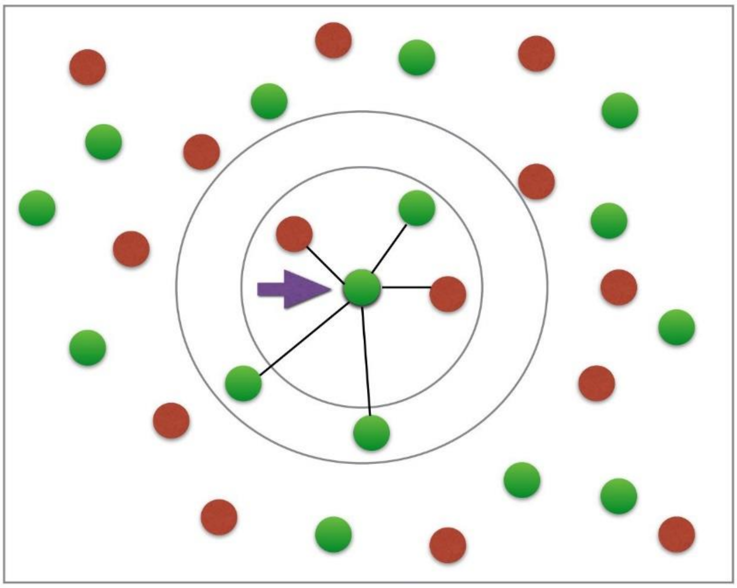


Figure 3.2: KNN of two classes of data

Source: https://towardsdatascience.com/building-a-k-nearest-neighbors-k-nn-model-with-scikit-learn-51209555453a

The learning strategy in a KNN is more like memorization. It is just like remembering what the answer should be when the question has certain characteristics rather than really knowing the answer, because you understand the question by means of specific classification rules (Mueller & Massaron, 2016). In a sense, KNN is often defined as a lazy algorithm because no real learning is done at the training time, just data recording. It is fast at training data using a lazy algorithm. Nonetheless, the prediction of data is slow. In conclusion, KNN can be used in machine learning for working on classification especially when one has many labels to encounter. Compared to other models, they need to specify a different model for each label (Mueller & Massaron, 2016).

3.1.3 Neural Network

The core neural network algorithm is the neuron. Many neurons arranged in an interconnected structure make up a neural network, with each neuron linking to the inputs and outputs of other neurons. Thus, a neuron can input features from examples or the results of other neurons, depending on its location in the neural network. Something similar to the neuron, the perceptron, appears earlier in this book, although it uses a simpler structure and function(Mueller & Massaron, 2016). Neurons in a neural network are a further evolution of the perceptron: They take many weighted values as inputs, sum them, and provide the summation as the result, just as a perceptron does. However, they also provide a more sophisticated transformation of the summation, something that the perceptron cannot do(Mueller & Massaron, 2016).

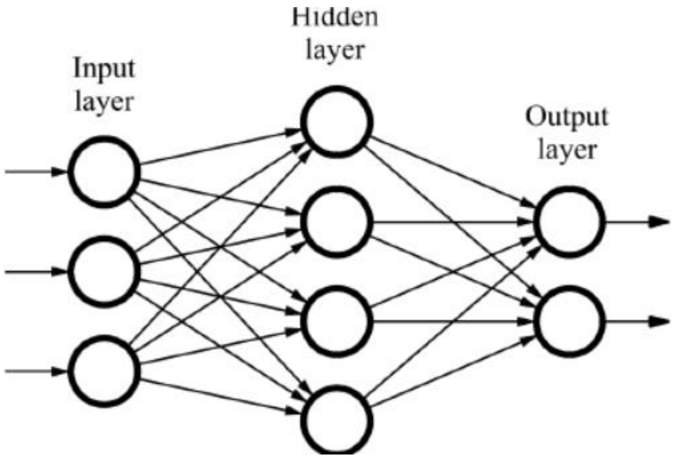


Figure 3.3: In-depth look of neural network structure

Source: https://databricks.com/glossary/neural-network

3.1.4 Naïve Bayes

Naïve Bayes (NB) is a method of using probability. In statistics, naive Bayes classifiers are a type of "probabilistic classifier" based on Bayes' theorem and strong (nave) independence assumptions between features (see Bayes classifier). They are one of the most basic Bayesian network models, but when combined with kernel density estimation, they can achieve higher levels of accuracy. Naive Bayes is a straightforward method for building classifiers, which are models that give classification to problem cases represented as vectors of feature values, with the class labels selected from a limited set. For training such classifiers, there is no one algorithm, but rather a variety of algorithms based on the same principle: all naive Bayes classifiers assume that the value of one feature is independent of the value of any other feature, given the class variable. For example, if a fruit is red, round, and around 10 cm in diameter, it is termed an apple. A naive Bayes classifier examines each of these characteristics to contribute independently to the likelihood that this fruit is an apple, regardless of any possible confounding variables.

The number of parameters required by Nave Bayes classifiers is linear in the number of variables (features/predictors) in a learning problem. Maximum-likelihood training can be done in linear time by evaluating a closed-form expression, rather than the time-consuming iterative approximation required by many other forms of classifiers.

Simple Bayes and Independent Bayes are two terms used in the statistics field to describe naive Bayes models.

3.2 Method of Solution

MMS-IMU sensor will be used as the solely hardware to gain the kinematics data. It is equipped with 3-axis accelerometer, 3-axis gyroscope and 3-axis magnetometer. Since its weight is easy to carry, it can be placed anywhere on the body. With the memory of 512MB of NAND Flash, it is sufficient to collect raw data from human motions. MMS-IMU sensor record the raw sensor data via Bluetooth at up to 400Hz. The data is then accessed and downloaded in Microsoft Excel file on smartphone or computer. Accelerometer, gyroscope and magnetometer are all combined to form as sensor fusion so that sturdy absolute orientation can be achieved.

Figure 3.4: MMS-IMU sensor

Source: https://mbientlab.com/store/metamotions/

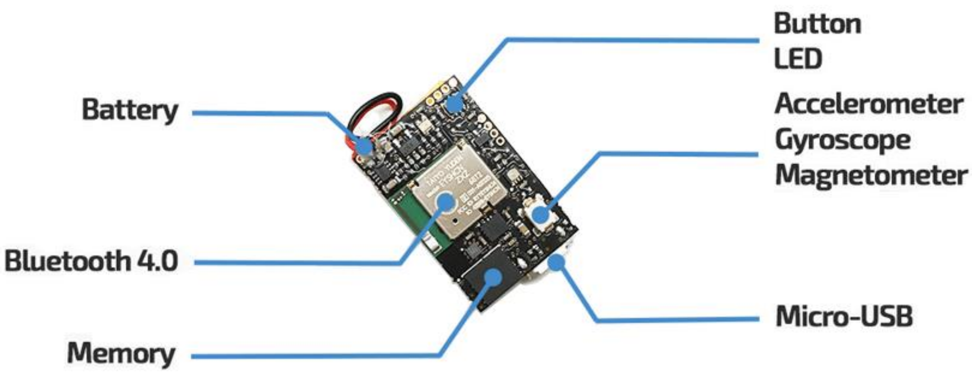


Figure 3.5: In-depth look of MMS-IMU

Source: https://mbientlab.com/store/metamotions/

The first step in this project is placing the sensor on the body. As suggested before, the sensor will be first placed on the wrist by using wrist band kit. It will look like placing watch on wrist. After this body location is completed, the sensor is then placed on the back of the body and finally the shank. It is the best to place the sensor where the part of body is moving, particularly in angular direction or speed (accelerometer). The data can be easily numerously presented to highlight human motions whether he/she is walking, running, or climbing.

The human motions for this project have been decided. Data from the human will be recorded based on walking, running, standing up, sitting down, typing on computer and jumping. After placing the sensor on particular body location, the raw data is then recorded methodically. A flat (180°) treadmill is used to record walking and running raw data to get pattern consistency. A chair will be used to record sitting and standing raw data. A flat ground will a platform for a subject to collect jumping data. Lastly, a keyboard (any keyboard) is the last tool to record typing data. This method is crucial as the data accuracy depends the location where participant can start doing these activities. The subject will warm-up for five minutes to get his pace in moving activities such as walking and running. A 1-min-gap between activities or the same activity is set to ensure the participant gets enough time to recover or settle down. Then, the participant will proceed to the next activity or the same activity which is repeated until 1200 data is obtained for each activity.

After all raw data is collected, the data will be accessed on CSV file through Bluetooth. Software like Orange and Matlab will be used for model development. For the first step, Orange will be used to visualise all the data so the data can be classified according to its pattern. Machine Learning models that will be used are Support Vector Machine (SVM), k-Nearest Neighbour (KNN), Neural Network (NN) and Naïve Bayes (NB). The data is then split in 70% training data and 30% testing data. Since there are a lots parameter in this data, Principal Component analysis (PCA) is used to transform linearly the input data as few as possible. This step is taken to reduce dimension of large datasets. That means it will reduce the number of variables in the data. It does still contain most of information in the large dataset. The target in this project is to classify human motion. Thus, the human motion is the output of the system to differentiate every motion. The data is then predictively analysed and finally trained to find out score in term of classification accuracy (percentage).

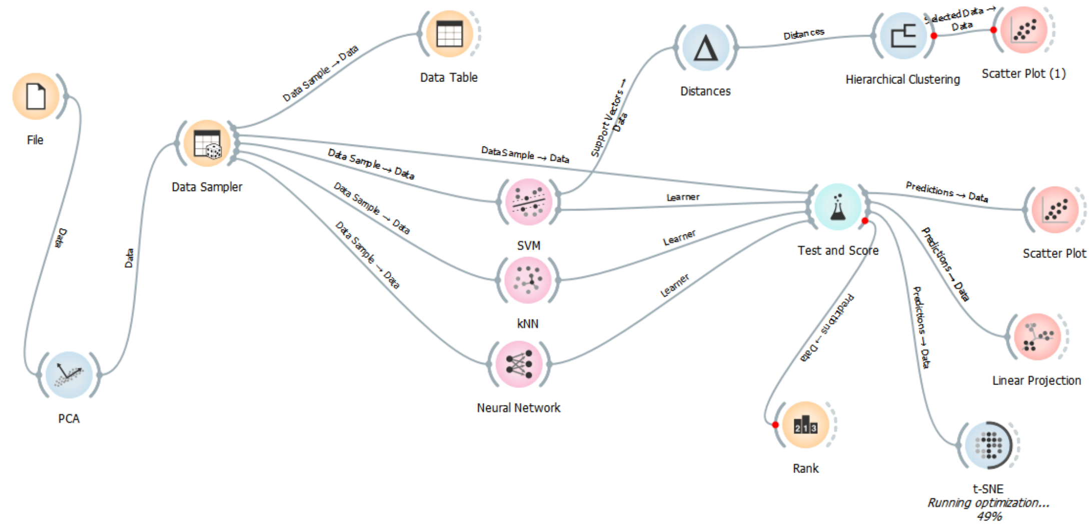


Figure 3.6: Project widgets in Orange Data Mining Software (Example)

Finally, analysis and model tuning will be carried out to determine the best accuracy for every human motion classification. The best two-parameter for every motion will be selected to visualise the data to get the best margin between the motions. With this method, high accuracy will be acquired and thus every motion can be differentiated accordingly and accurately.

Figure 3.7: Block Diagram of Project for Human Motion Activities

3.3 Design Development

Three locations on human body are selected to carried this experiment which are wrist, back and shank. Data collection of next location of IMU sensor is collected after one location is completed. Six human activities which are walking, running, standing up, sitting down, typing and jumping are selected to perform this experiment.

Figure 3.3.8: IMU Sensor on back of body

Figure 3.3.9: IMU Sensor on left shank of subject

Figure 3.3.10: IMU sensor on left wrist of subject

Table 3.1 Description of Human Motion Activities.

Walking condition	**Description**
Walking	The subject walks on treadmill with consistent speed
Running	The subject runs on treadmill with consistent speed
Standing up	The subject stands up from a chair from siting position with 90° body posture to the chair
Sitting down	The subject sits down from standing position
Jumping	The subject jumps by leaping to the air and land with bending legs.
Typing	The subject types on the keyboard

Before the raw data is collected, the IMU is connected to smartphone via Bluetooth by using Metabase Application. Gyroscope and accelerometer are set to ±2000°/sec and ±16g respectively on the application Metabase on Smartphone. Both sensors are set to the same frequency of 50Hz.

The experiment is carried out indoor to collect the raw data. As a treadmill is used to collect data of running and walking, the raw data can be obtained steadily and not in dispersed pattern corresponds to the activities. A consistent speed from treadmill can result in consistency in running and walking raw data.

Figure 3.3.3.11 Metabase Accelerometer and Gyroscope Parameter Settings

Figure 3.3.12: The participant is collecting running and walking raw data on a treadmill.

Figure 3.3.3.13: (Left) Subject initial position for standing raw data,(Right)Subject initial position for sitting raw data

Figure 3.3.3.14 Typing position with IMU on wrist

Figure 3.3.3.15 Subjects leaps to the air during jumping data collection

Before the experiment is performed, the participant will warm up for about five minutes to find his own pace during walking and running. Total of data collected is 7200 which means 1200 raw data is for each activity. 1-minute-gap is applied for each activity so that the data will not lose its pattern consistency. Sitting down, standing up and jumping require more repetition as its duration is lower than others. Thus, one-rep of these activities will only produce about 180 raw data.

After raw data is collected, it will be then downloaded to computer via csv file to perform analysis on workspaces (Matlab and Orange). Three parameters of x-y-z axis from each gyroscope and accelerometer will be used as manipulative data and six activities will be marked as targets of experiment. The proportion of 70/30 is used in this experiment which means 70% of the trained data and the rest remains as test data. Two components of PCA and variance of 97% are used to reduce the dimensionality of x-y-z parameters. First analysis will determine the best sensor location on the body. After finding out the best location out of three, machine learning models will be tested in term of accuracy. SVM, kNN, NN and NB are the models used in both Orange and Matlab to train the raw data and then will be tested with 30% of the raw data. The best machine learning model will be chosen based on their precision. The precision can also be viewed by using the confusion matrix based on 30% (2160) test data. In this matrix, every data precision can be analysed for every 360 test-data which will sum up to 2160 for all activities. Any misclassified data will be calculated in percentage.

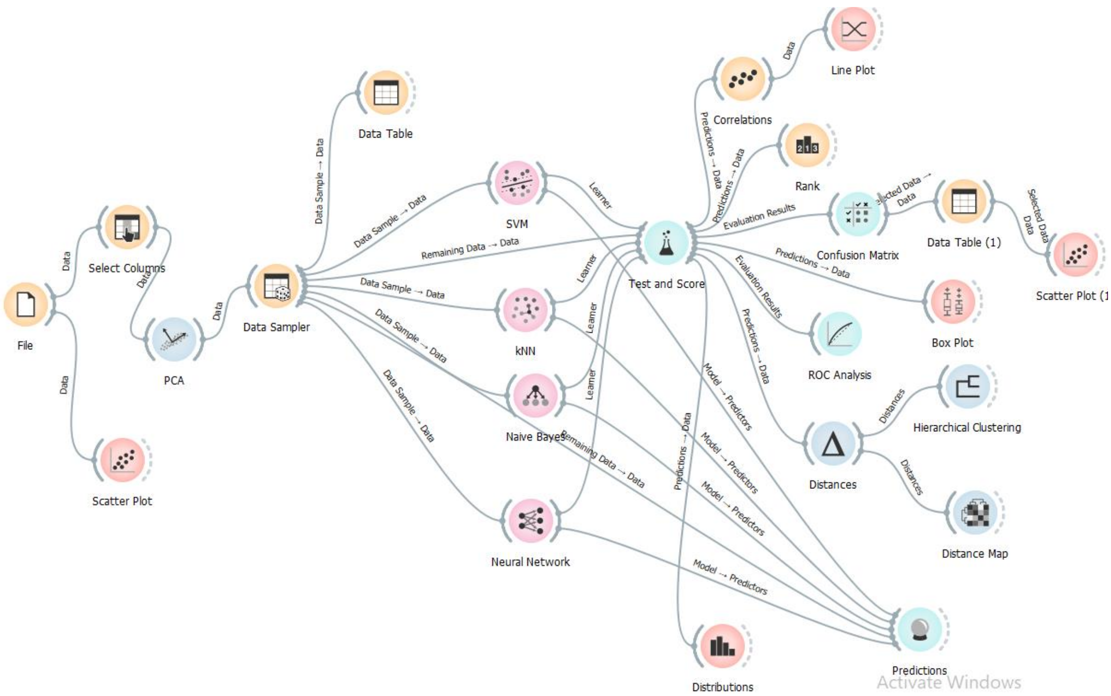


Figure 3.3.16: Orange Visual Programming Design

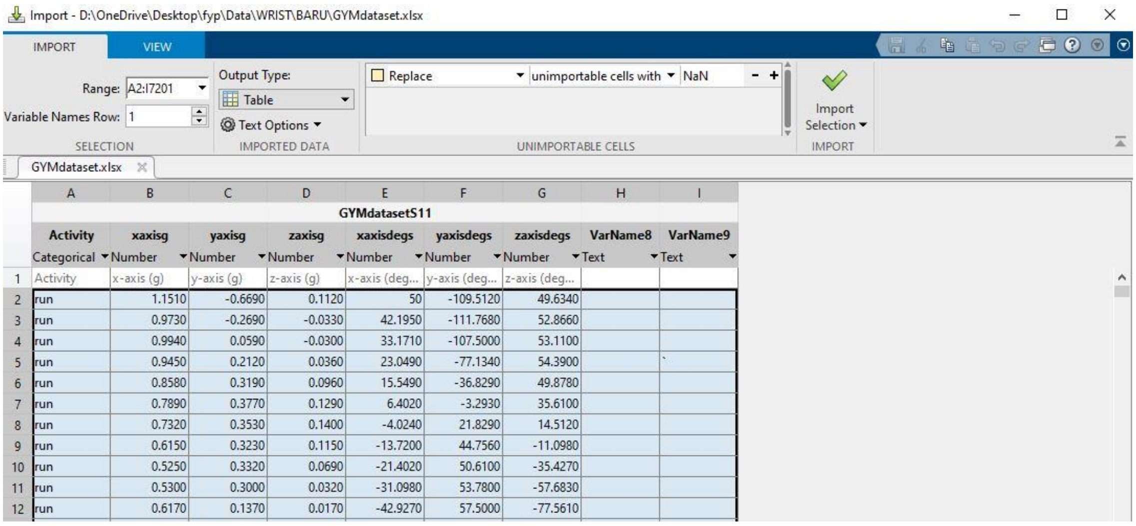


Figure 3.3.17: Importing Data in Matlab

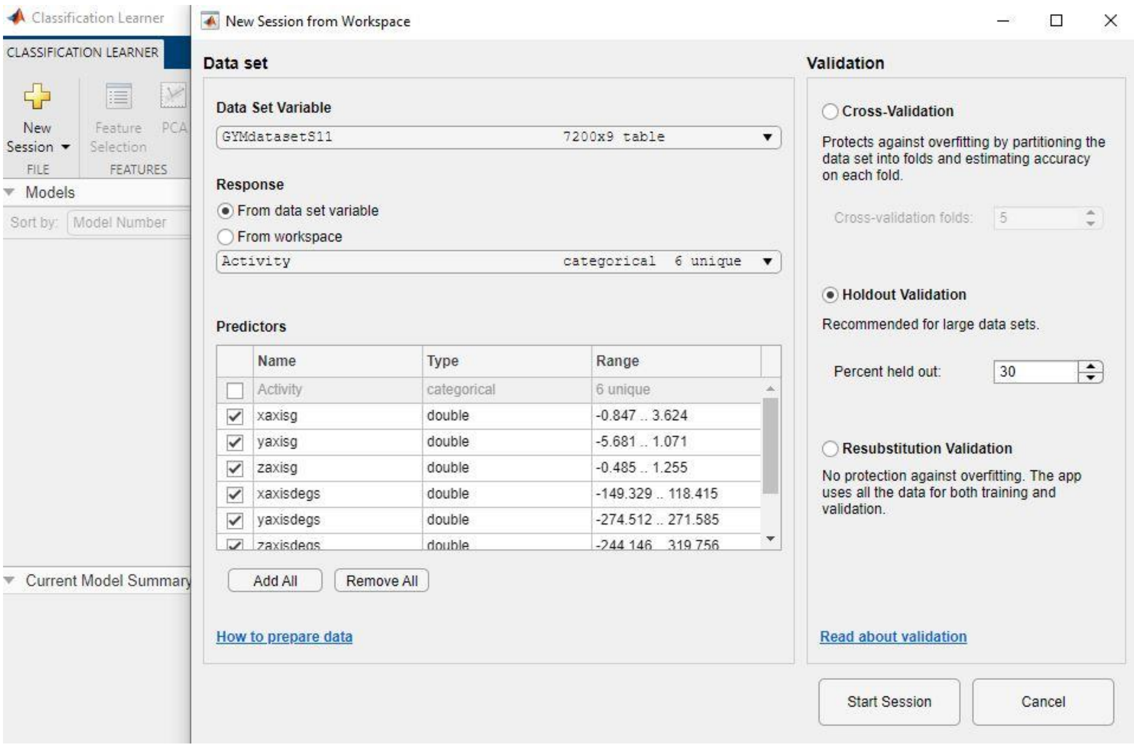


Figure 3.3.18: Classification Learner Application as workspace to train and test data

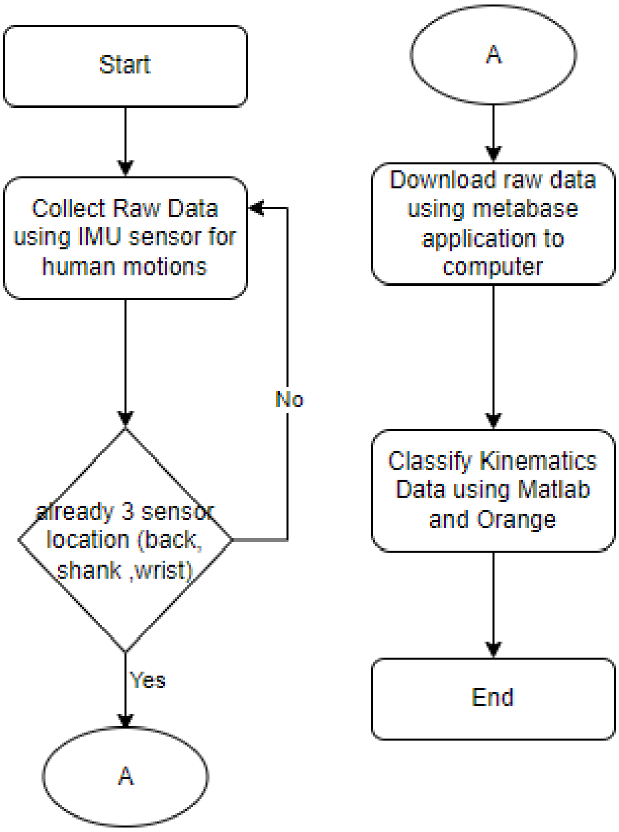


Figure 3.3.19: Flow chart of the Experiment

CHAPTER 4

RESULT AND DISCUSSION

The experiment is performed in a personal computer with Microsoft Windows 10, an Intel Core Processor i5-8400, 8 GB RAM and GPU of Intel Graphic Technology. After 7200 raw data is imported to Orange and Matlab to be analysed. Orange and Matlab use visual programming that does not need for writing any program or code. This step is used to get the user familiar with data analysis and model development in machine learning. Table below shows that the number of data corresponds to the activities. Using all learning models, this experiment is evaluated by the resulting models in terms of average 0–100 accuracy. (Yoshikawa, Losing & Demircan)

Table 4.1 Number of Datased based on Human Activities

Activity	Number of Data
Walk	1200
Run	1200
Stand Up	1200
Sit Down	1200
Jump	1200
Type	1200
	7200

4.1 Sensor Location Determination

By using both Matlab and Orange, wrist is the best among three sensor locations on human body based on the score (precision). SVM scores 92% on wrist meanwhile it only scores 76.7% and 68.3% on back and shank respectively. Meanwhile, Naïve Bayes results in 80% on wrist and then gets 63% and 63.5% on back and shank respectively. By using kNN method, 73.1% precision is gained on wrist and this method scores 41% and 62.1% on back and shank respectively. The last method which is Neural Network scores 94.1% on wrist and it also scores 75% and 56.9% on back and shank respectively.

Matlab results higher precision than Orange and wrist still scores the highest in term of precision. SVM scores 96.6% on wrist meanwhile it results 85.3% and 89.3% on back and shank respectively. Meanwhile, Naïve Bayes results in 88.7% on wrist and then gets 75% and 76.1% on back and shank respectively. By using kNN method, 97.7% precision is gained on wrist and this method scores 92% and 92.3% on back and shank respectively. The last method which is Neural Network scores 97% on wrist and it also scores 90% both back and shank. From tables above, it can be concluded that wrist the optimal sensor location on the body as both Matlab and Orange score the highest precision on wrist based on the trained data. Wrist gives higher precision as it moves in angular position and acceleration of moving hand. Compared to the back of the body, the precision is lower because less consistency in angular speed and acceleration during activities. That is why only part of the data on particular activity gets true positive value and the rest results in false negative. The performance is calculated through algorithm which involve equation for precision.

$$Precision_i(\%) = \frac{TP_i}{TP_i + FP_i} \ x \ 100\% \qquad 4.1$$

where *i* is the type of six activity, TP is True Positive and FP is False Positive.

Table 4.2 Orange Precision and Test Score for Wrist

Wrist	Precision (%)
SVM	92.00
Naïve Bayes	80.00
kNN	73.10
Neural Network	94.10

Table 4.3 Matlab Precision and Test Score for Wrist

Wrist	Precision (%)
SVM	96.60
Naïve Bayes	88.70
kNN	97.70
Neural Network	97.00

Table 4.4 Orange Precision and Test Score for Back of body

Back	Precision (%)
SVM	76.70
Naïve Bayes	63.00
kNN	41.00
Neural Network	75.00

Table 4.5 Matlab Precision and Test Score for Back of body

Back	Precision (%)
SVM	85.30
Naïve Bayes	75.00
kNN	92.00
Neural Network	90.00

Table 4.6 Orange Precision and Test Score for shank

Shank	Precision (%)
SVM	68.30
Naïve Bayes	63.50
kNN	62.10
Neural Network	56.90

Table 4.7 Matlab Precision and Test Score for shank

Shank	Precision (%)
SVM	89.30
Naïve Bayes	76.10
kNN	92.30
Neural Network	90.00

4.2 Determination of Machine Learning Model

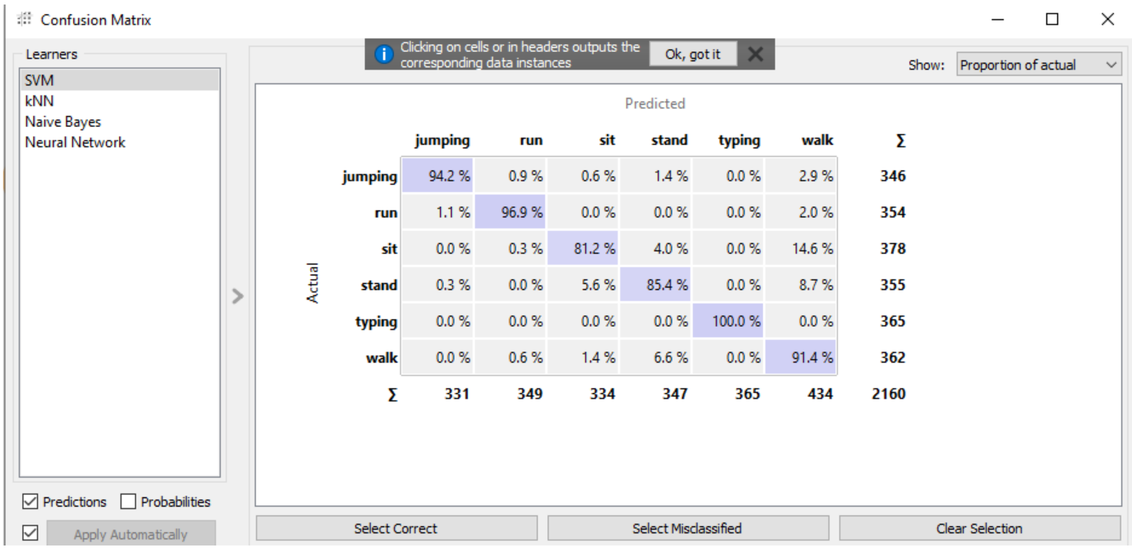
The analysis is then continued to determine the best machine learning model in both Matlab and Orange. In Orange workspace, SVM is set to guassian Radial Basis Function with kernel of exp $(-g|x-y|^2)$ and iteration limit of 100. In SVM, Cost (C) is set to 100 which means penalty term for loss and applies for classification and regression. Regression Loss Epsilon is set to 0.10 which means distance from true values within which no penalty is associated with predicted values. KNN is set with uniform weight, metric of Euclidean and five number of neigbours. There is no setup required for naïve bayes and it can be used directly from the 70% train data. In Neural Network, 200 neurons in hidden layers and 100 iteration is used for setup in machine learning learner. As all six activities involve most of acceleration, x-axis(g) and z-axis(g) are the best parameter to view the data pattern.

In Matlab, SVM is set to Cubic kernel function as the best algorithm among all SVMs available in Classification Learner Application. This model is set to automatic kernel scale. Fine kNN is used as it is the best model among all kNNs with one number of neighbors, Euclidean distance metric and equal distance weight. Kernel Naïve Bayes is selected in the application with kernel type of Gaussian and unbounded support.

In Wide Neural Network, 100 first layer is fixed value for this NN model with Rectified Linear Unit activation function. The iteration limit is fixed-set to 1000. All 70% data will be trained in these models and then be tested with rest of the data.

To summarize the performance, confusion matrix will be used to for classification algorithm. In confusion matrix, any misclassification can be spotted in percentage of this 30% test data.

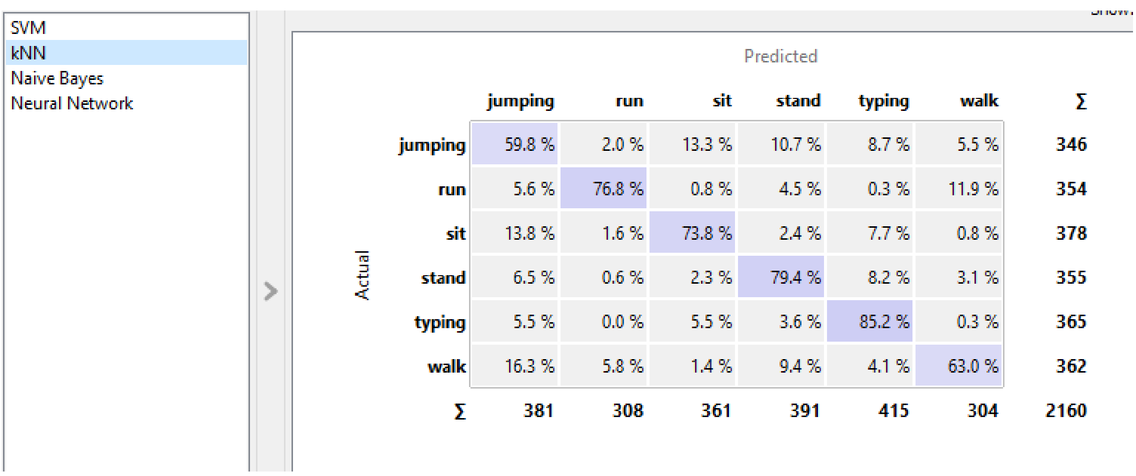
Below are the results of test data in confusion matrix in Matlab and Orange. Tables below show the summary of the confusion matrix result based on the activities in Orange and Matlab.



Predicted

Actual	jumping	run	sit	stand	typing	walk	Σ
jumping	94.2 %	0.9 %	0.6 %	1.4 %	0.0 %	2.9 %	346
run	1.1 %	96.9 %	0.0 %	0.0 %	0.0 %	2.0 %	354
sit	0.0 %	0.3 %	81.2 %	4.0 %	0.0 %	14.6 %	378
stand	0.3 %	0.0 %	5.6 %	85.4 %	0.0 %	8.7 %	355
typing	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %	0.0 %	365
walk	0.0 %	0.6 %	1.4 %	6.6 %	0.0 %	91.4 %	362
Σ	331	349	334	347	365	434	2160

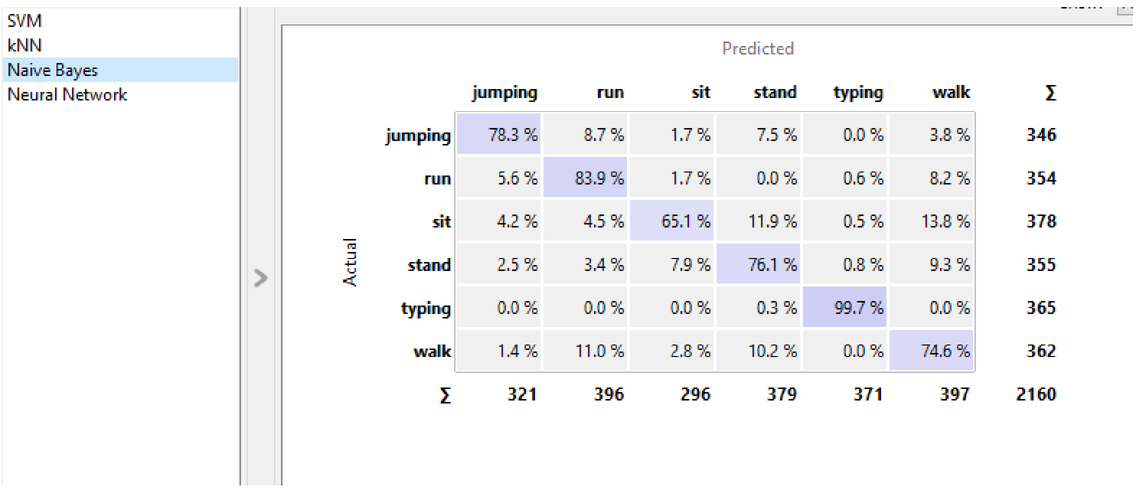
Figure 4.1 SVM Confusion Matrix in Orange



Predicted

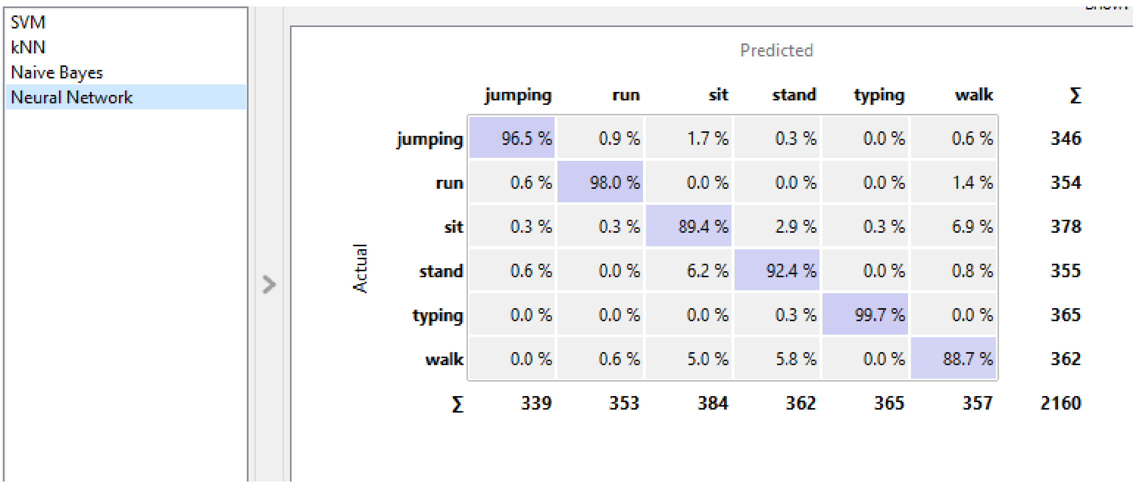
Actual	jumping	run	sit	stand	typing	walk	Σ
jumping	59.8 %	2.0 %	13.3 %	10.7 %	8.7 %	5.5 %	346
run	5.6 %	76.8 %	0.8 %	4.5 %	0.3 %	11.9 %	354
sit	13.8 %	1.6 %	73.8 %	2.4 %	7.7 %	0.8 %	378
stand	6.5 %	0.6 %	2.3 %	79.4 %	8.2 %	3.1 %	355
typing	5.5 %	0.0 %	5.5 %	3.6 %	85.2 %	0.3 %	365
walk	16.3 %	5.8 %	1.4 %	9.4 %	4.1 %	63.0 %	362
Σ	381	308	361	391	415	304	2160

Figure 4.2 KNN Confusion Matrix in Orange



	jumping	run	sit	stand	typing	walk	Σ
jumping	78.3 %	8.7 %	1.7 %	7.5 %	0.0 %	3.8 %	346
run	5.6 %	83.9 %	1.7 %	0.0 %	0.6 %	8.2 %	354
sit	4.2 %	4.5 %	65.1 %	11.9 %	0.5 %	13.8 %	378
stand	2.5 %	3.4 %	7.9 %	76.1 %	0.8 %	9.3 %	355
typing	0.0 %	0.0 %	0.0 %	0.3 %	99.7 %	0.0 %	365
walk	1.4 %	11.0 %	2.8 %	10.2 %	0.0 %	74.6 %	362
Σ	321	396	296	379	371	397	2160

Figure 4.3 Naïve Bayes Confusion Matrix in Orange



	jumping	run	sit	stand	typing	walk	Σ
jumping	96.5 %	0.9 %	1.7 %	0.3 %	0.0 %	0.6 %	346
run	0.6 %	98.0 %	0.0 %	0.0 %	0.0 %	1.4 %	354
sit	0.3 %	0.3 %	89.4 %	2.9 %	0.3 %	6.9 %	378
stand	0.6 %	0.0 %	6.2 %	92.4 %	0.0 %	0.8 %	355
typing	0.0 %	0.0 %	0.0 %	0.3 %	99.7 %	0.0 %	365
walk	0.0 %	0.6 %	5.0 %	5.8 %	0.0 %	88.7 %	362
Σ	339	353	384	362	365	357	2160

Figure 4.4 Neural Network Confusion Matrix in Orange

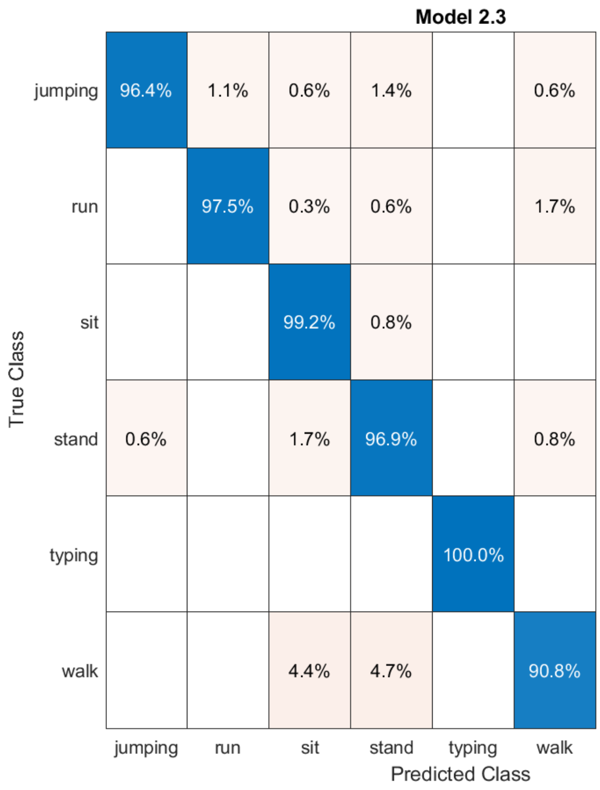


Figure 4.5 SVM Confusion Matrix in Matlab

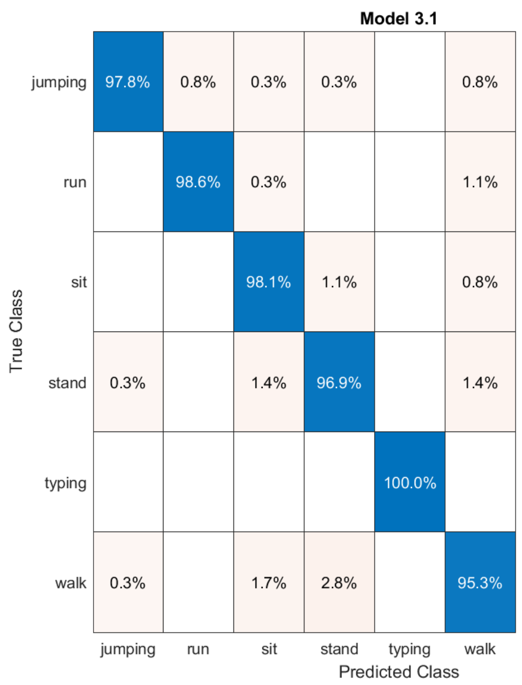


Figure 4.6 kNN Confusion Matrix in Matlab

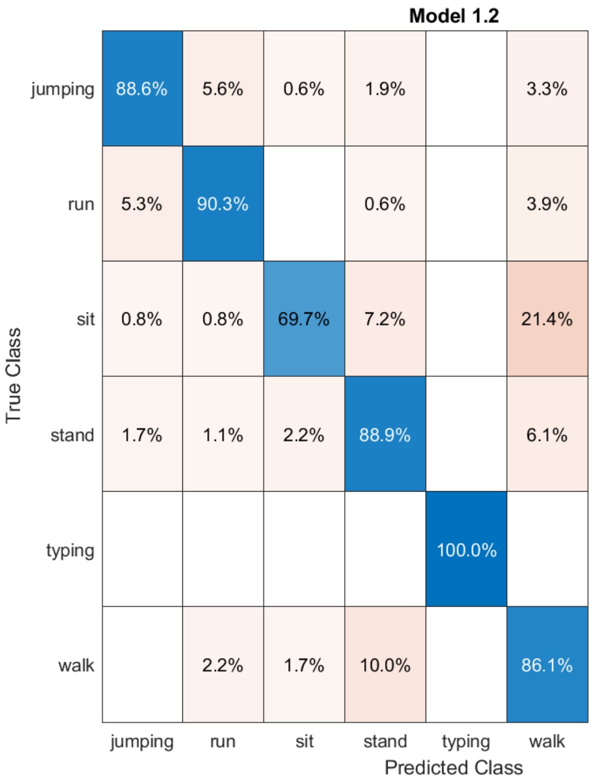


Figure 4.7 Naïve Bayes Confusion Matrix in Matlab

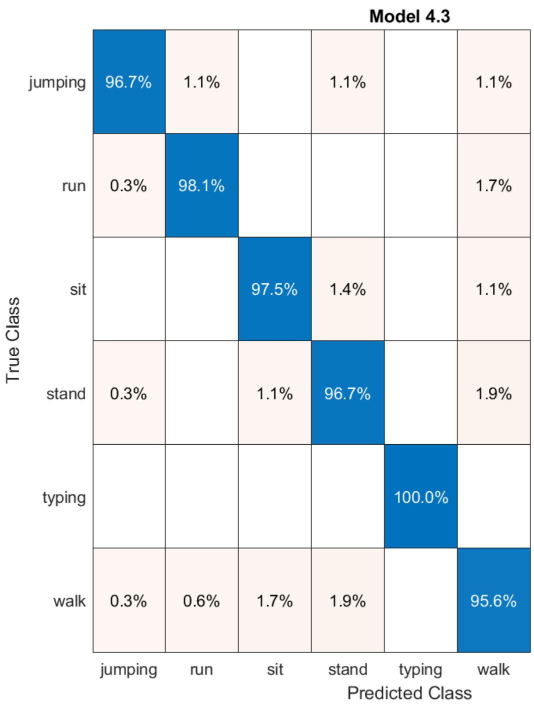


Figure 4.8 Neural Network Confusion Matrix in Matlab

Table 4.8 Orange Accuracy in Confusion Matrix of Test Data for Walking

Model	**Accuracy (%)**
SVM	91.4
Naïve Bayes	74.6
kNN	63.0
Neural Network	88.7

Table 4.9 Orange Accuracy in Confusion Matrix of Test Data for Jumping

Model	**Accuracy (%)**
SVM	94.20
Naïve Bayes	78.30
kNN	59.80
Neural Network	96.50

Table 4.10 Orange Accuracy in Confusion Matrix of Test Data for Running

Model	**Accuracy (%)**
SVM	96.90
Naïve Bayes	83.90
kNN	76.80
Neural Network	98.00

Table 4.11 Orange Accuracy in Confusion Matrix of Test Data for Sitting

Model	**Accuracy (%)**
SVM	81.20
Naïve Bayes	65.10
kNN	73.10
Neural Network	89.40

Table 4.12 Orange Accuracy in Confusion Matrix of Test Data for Standing

Model	**Accuracy (%)**
SVM	85.4
Naïve Bayes	76.10
kNN	79.40
Neural Network	92.4

Table 4.13 Orange Accuracy in Confusion Matrix of Test Data for Typing

Model	**Accuracy (%)**
SVM	100.00
Naïve Bayes	99.70
kNN	85.20
Neural Network	99.70

Table 4.14 Matlab Accuracy in Confusion Matrix of Test Data for Walking

Model	**Accuracy (%)**
SVM	90.80
Naïve Bayes	86.10
kNN	95.30
Neural Network	95.60

Table 4.15 Matlab Accuracy in Confusion Matrix of Test Data for Jumping

Model	**Accuracy (%)**
SVM	96.40
Naïve Bayes	88.60
kNN	97.80
Neural Network	96.70

Table 4.16 Matlab Accuracy in Confusion Matrix of Test Data for Running

Model	**Accuracy (%)**
SVM	97.50
Naïve Bayes	90.30
kNN	98.60
Neural Network	98.10

Table 4.17 Matlab Accuracy in Confusion Matrix of Test Data for Sitting

Model	**Accuracy (%)**
SVM	99.20
Naïve Bayes	69.70
kNN	98.10
Neural Network	97.50

Table 4.18 Matlab Accuracy in Confusion Matrix of Test Data for Standing

Model	Accuracy (%)
SVM	96.90
Naïve Bayes	88.90
kNN	96.90
Neural Network	96.70

Table 4.19 Matlab Accuracy in Confusion Matrix of Test Data for Typing

Model	Accuracy (%)
SVM	100.00
Naïve Bayes	100.00
kNN	100.00
Neural Network	100.00

\

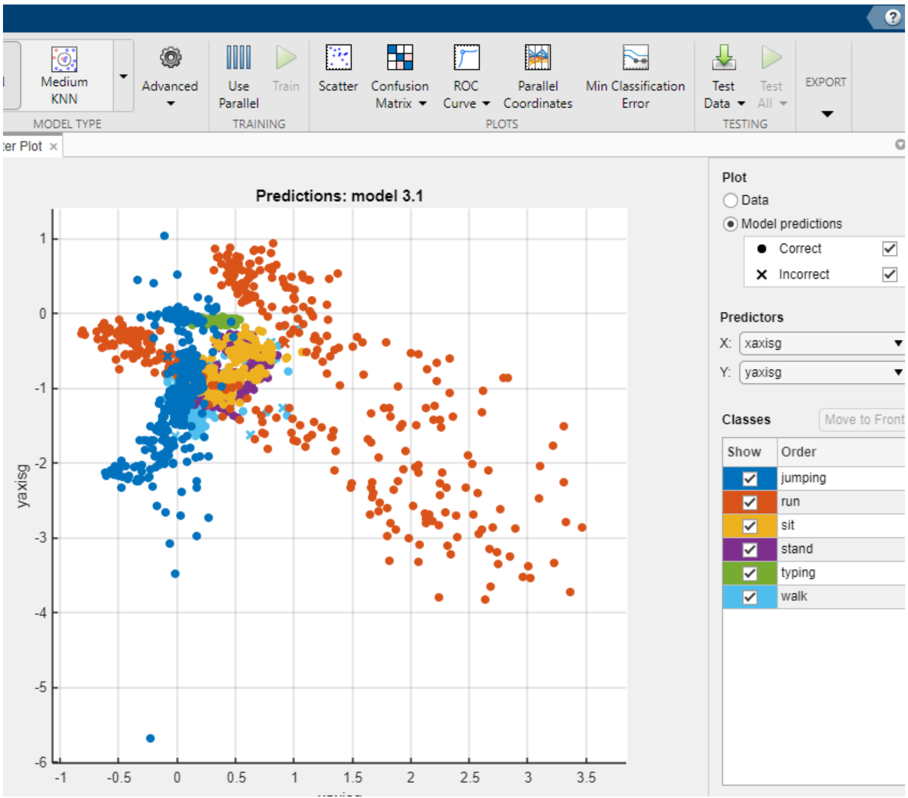


Figure 4.9 Data Scatter of Wrist in Matlab

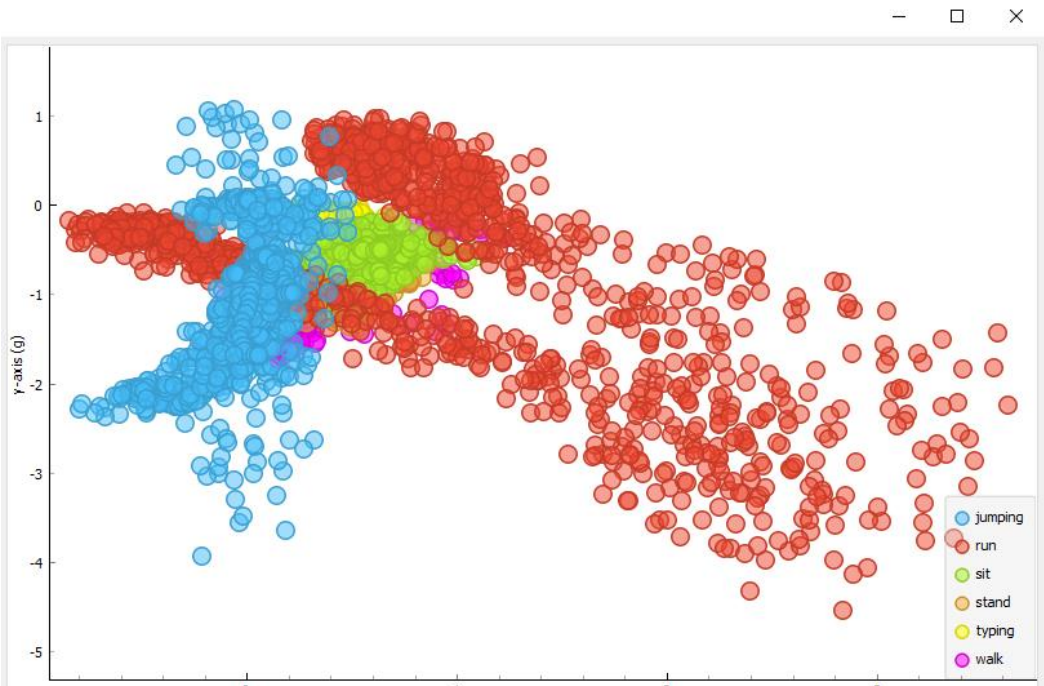


Figure 4.10 Data Scatter of Wrist in Orange

As wrist is the best sensor location, this section only focuses on data on this location. Based on Test and Score Result in Orange in table 4.1.1, Neural Network scores the highest with 94.1% of accuracy meanwhile in table 4.1.2 Matlab's kNN scores the highest accuracy of 97.7%. There are differences in precision on Matlab and Orange as difference algorithm used in every machine learning models.

Different type of SVM used in both Matlab and Orange which are Cubic SVM and RBF SVM. That means different algorithm used are Cubic Polynomial and Gaussian. These are what Cubic SVM model is trained with polynomial kernel meanwhile the other one is trained with Gaussian kernel. Cubic SVM can be expressed into:

$$k(x_1, x_2) = (x_1^T . x_2)^3 \qquad 4.2.1$$

Where: x is class

X^T is transposed feature

Gaussian kernel approaches different pattern than Polynomial's. It can be expressed into:

$$k(x_1, x_2) = exp\ (-(\frac{\|x_1 - x_2\|^2}{2\gamma^2} \qquad 4.2.2$$

Where: x is class

γ^2 is inverse of the standard deviation of the RBF kernel

Wide neural network is used in Matlab with fixed setting of neuron layers. First layer size is set to 100 with 1000 iterations limit and it uses rectified linear activation function. This differs from what neural network on Orange which set to 200 hidden layers, 100 iterations, Adam solver and regularization of 0.0001. Adam solver is an optimization solver which suits with large dataset that uses memory. Regularization is a model to create a small change to algorithm and this makes the model generalizes better.

Fine kNN is chosen on Matlab as it is the best kNN on wrist among all kNNs available. Fine kNN uses only one number of neighbour compared to five number of neighbours of kNN set on Orange. Both kNN uses Euclidean distance metric which is the shortest distance between to points of coordinates.

$$d = ((a_1 - b_1)^2 + (a_2 - b_2)^2)^{1/2}$$

$$\text{Where: } a \text{ and } b \text{ is two different point of coordinates} \qquad 4.2.3$$

$$d \text{ is the distance between two points}$$

Kernel Naïve Bayes is the best NB model on Matlab what trained on wrist data. This is applied on Golf-Testset by using Apply model moderator. To load Golf-Testset, a retrieve operator is needed to run the algorithm with default value. Bayes theorem on Orange applies default preprocessing method by removing empty column. To execute Bayes theorem, it ought to discretise the numeric values to four bins with same frequency.

CHAPTER 5

CONCLUSION

Machine learning has been popularised in modern technology to supervise or unsupervised learning. As with creating machine learning tasks, people who create future environments will be experts in their particular craft, rather than be computer scientists or data scientists. Solving the science of machine learning will eventually turn into an engineering exercise that will give anyone with a good idea the required access. (Mueller, Massaron, 2021)

Based on the literature review, most of cases place the sensor on wrist, foot, shank, knee, thigh and hip. Algorithms that are used in development to classify human's activities are SVM, KNN and neural network. These three are popular methods used by researcher to train the data by implementing mathematical modelling in algorithms.

As the primary dataset of from all sensor location of wrist, back and shank, it resulted that wrist is the optimal location among all three locations as the accuracy on wrist data is more accurate where most of machine learning models scored above 90%. Placing IMU on wrist causes more accuracy in angular speed (°/sec) and acceleration (gals) in most of the activities. Typing on computer gives accuracy above 90% on all sensor locations. When it involves the movement of human body, wrist can produce consistency pattern of data to result in high precision and accuracy.

After the dataset is trained with 70% of data, Neural Network scored the highest accuracy of 94.1% in Orange. Neural Network is Meanwhile, k-Nearest Neighbour obtained the highest accuracy of 97.7% on Matlab. This is because different algorithms are used on both Matlab and Orange. Different algorithms can manipulate the parameter in functions and equations to produce different result. Neural network and k-Nearest Neighbour works relatively well when there is a clear margin of separation between classes on Orange and Matlab. They are more effective in high dimensional spaces. It is effective in cases where the number of dimensions is greater than the number of samples.

SVM can be considered effective as really close percentage in accuracy to Neural Network and kNN.

In overall classification accuracy of three Machine Learning models, it can achieve approximately 90% and above which is highly accurate. Matlab is currently the best tool for data analysis purpose for its convenience in term of visual programming according to accuracy result. Based on three results above, all data can be visualised into a pattern clearly.

Larger dataset would absolutely give higher precision that experiment’s result. By doubling volume of the dataset (14400 data), the precision can be increase to classify human motion or activities. Fewer activities classification would give higher accuracy as well as machine learning algorithm can easily classify the human activities. For example, sitting down and standing up classification would give higher accuracy in Test and Score result. It can achieve from 95% accuracy and above.

REFERENCES

Begg, R., & Kamruzzaman, J. (2005). A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data. *Journal of Biomechanics*, *38*(3). https://doi.org/10.1016/j.jbiomech.2004.05.002

Lau, H. Y., Tong, K. Y., & Zhu, H. (2008). Support vector machine for classification of walking conditions using miniature kinematic sensors. *Medical and Biological Engineering and Computing*, *46*(6). https://doi.org/10.1007/s11517-008-0327-x

Lim, H., Kim, B., & Park, S. (2020). Prediction of lower limb kinetics and kinematics during walking by a single IMU on the lower back using machine learning. *Sensors (Switzerland)*, *20*(1). https://doi.org/10.3390/s20010130

Mannini, A., Trojaniello, D., Cereatti, A., & Sabatini, A. M. (2016). A machine learning framework for gait classification using inertial sensors: Application to elderly, post-stroke and huntington's disease patients. *Sensors (Switzerland)*, *16*(1). https://doi.org/10.3390/s16010134

Mueller, J., & Massaron, L. (2016). Machine learning for dummies. In *For dummies*.

Preatoni, E., Nodari, S., & Lopomo, N. F. (2020). Supervised Machine Learning Applied to Wearable Sensor Data Can Accurately Classify Functional Fitness Exercises Within a Continuous Workout. *Frontiers in Bioengineering and Biotechnology*, *8*. https://doi.org/10.3389/fbioe.2020.00664

Stetter, B. J., Ringhof, S., Krafft, F. C., Sell, S., & Stein, T. (2019). Estimation of knee joint forces in sport movements using wearable sensors and machine learning. *Sensors (Switzerland)*, *19*(17). https://doi.org/10.3390/s19173690

Zhang, X., Xu, Z., & Liao, H. (2022). Human motion tracking and 3D motion track detection technology based on visual information features and machine learning. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-021-06703-2

Vu, C., & Kim, J. (2018). Human Motion Recognition by Textile Sensors Based on Machine Learning Algorithms. *Sensors*, *18*(9), 3109. https://doi.org/10.3390/s18093109

Mannini, A., & Sabatini, A. M. (2010). Machine Learning Methods for Classifying Human Physical Activity from On-Body Accelerometers. *Sensors*, *10*(2), 1154–1175. https://doi.org/10.3390/s100201154

Yen, C. T., Liao, J. X., & Huang, Y. K. (2020). Human Daily Activity Recognition Performed Using Wearable Inertial Sensors Combined With Deep Learning Algorithms. *IEEE Access*, *8*, 174105–174114. https://doi.org/10.1109/access.2020.3025938

Randhawa, P., Shanthagiri, V., & Kumar, A. (2019). Recognition of Violent Activity Response Using Machine Learning Methods with Wearable Sensors. *Journal of Advanced Research in Dynamical and Control Systems*, *11*(11-SPECIAL ISSUE), 592–601. https://doi.org/10.5373/jardcs/v11sp11/20193071

Yoshikawa, T., Losing, V., & Demircan, E. (2020). Machine learning for human movement understanding. *Advanced Robotics, 34*(13), 828–844. https://doi.org/10.1080/01691864.2020.1786724

Wang, J., Fleet, D., & Hertzmann, A. (2008). Gaussian Process Dynamical Models for Human Motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(2), 283–298. https://doi.org/10.1109/tpami.2007.1167

Zhu, Z., Zhang, J., Zou, J., & Deng, C. (2019). Multi-kernel Gaussian process latent variable regression model for high-dimensional sequential data modeling. *Neurocomputing*, *348*, 3–15. https://doi.org/10.1016/j.neucom.2018.07.082

APPENDIX A
SAMPLE APPENDIX 1

https://drive.google.com/drive/folders/1MbWmOaS6hoLhgvibuHLNI4_X_nxaApwH?usp=sharing

(This link contains raw data and works during Experiment)

APPENDIX B
SAMPLE APPENDIX 2

Figure 5.1 Equipment used to record running and walking data

Figure 5.2 Subject is setting appropriate speed for run and walking mode