

อัลกอริทึมเรียนรู้สำหรับการวินิจฉัยโรคที่เกี่ยวข้องกับความผิดปกติทางการเคลื่อนไหวโดยใช้ข้อมูล  
การเดิน

นางสาวรัตติยา สุวานะสุนทรฤกษ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2556

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)

เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)  
are the thesis authors' files submitted through the Graduate School.

LEARNING ALGORITHM FOR MEDICAL DIAGNOSIS ON MOVEMENT DISORDER  
USING GAIT DATA

Ms. Rattiya Thanasoontornrek

A Thesis Submitted in Partial Fulfillment of the Requirements  
for the Degree of Master of Engineering Program in Computer Engineering

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2013

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	อัลกอริทึมเรียนรู้สำหรับการวินิจฉัยโรคที่เกี่ยวข้องกับความผิดปกติทางการเคลื่อนไหวโดยใช้ข้อมูลการเดิน
โดย	นางสาวรัตติยา สุานะสุนทรฤกษ์
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.พิชญ์ คนองชัยยศ
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	ดร.พีรพล เวทีกุล

---

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้รับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์  
(ศาสตราจารย์ ดร.บัณฑิต เอื้ออาภรณ์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ  
(ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก  
(ผู้ช่วยศาสตราจารย์ ดร.พิชญ์ คนองชัยยศ)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม  
(ดร.พีรพล เวทีกุล)

..... กรรมการ  
(ศาสตราจารย์ พญ.อารีรัตน์ สุพุทธิธาดา)

..... กรรมการภายนอกมหาวิทยาลัย  
(ผู้ช่วยศาสตราจารย์ ดร.อานนท์ รุ่งสว่าง)

รัตติยา ฐานะสุนทรฤกษ์: อัลกอริทึมเรียนรู้สำหรับการวินิจฉัยโรคที่เกี่ยวข้องกับความผิดปกติทางการเคลื่อนไหวโดยใช้ข้อมูลการเดิน (Learning Algorithm For Medical Diagnosis On Movement Disorder Using Gait Data) อ.ที่ปรึกษาวิทยานิพนธ์หลัก:  
 ผศ. ดร.พิชญ์ คนองชัยยศ, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: ดร.พีรพล เวทีกุล, 74 หน้า

การวิเคราะห์ข้อมูลการเดิน เป็นการศึกษาเกี่ยวกับข้อมูลการเดินของมนุษย์ รวมถึงการเก็บข้อมูลและการวิเคราะห์ข้อมูล ข้อมูลเหล่านี้สามารถนำไปใช้ประโยชน์ในหลาย ๆ ด้าน โดยเฉพาะอย่างยิ่งในทางการแพทย์ได้ใช้ข้อมูลเหล่านี้นำไปวิเคราะห์โรคที่ส่งผลกระทบต่ออาการเดินเพื่อวินิจฉัย ดำเนินรักษาและฟื้นฟูผู้ป่วย กระบวนการการวินิจฉัยนั้น แพทย์ผู้เชี่ยวชาญจะทำการอ่านและวิเคราะห์ผลของข้อมูลที่ได้มา ซึ่งในขั้นตอนนี้ต้องใช้ผู้ที่มีความรู้ที่เฉพาะทางและใช้ผู้ที่มีความเชี่ยวชาญอีกด้วย ดังนั้นวิทยานิพนธ์นี้จึงนำเสนอระบบการวิเคราะห์ข้อมูลการเดินจาก Motion Capture ในการจำแนกผู้ป่วยที่มีโรคเกี่ยวกับการเดิน โดยจะแบ่งออกเป็น 3 คลาส คือ คนปกติ (Normal), ผู้ป่วยข้อเข่าเสื่อม (Sick/Knee OA), ผู้ป่วยพาร์กินสัน (Sick/Parkinson) การทดลองได้ศึกษาสภาพแวดล้อมที่เหมาะสม โดยศึกษาในผู้กลุ่มตัวอย่าง 88 ราย เป็นคนปกติสูงวัย 27 คน ผู้ป่วยข้อเข่าเสื่อม 34 คนและผู้ป่วยพาร์กินสัน 27 คน ซึ่งผลการทดลองได้ให้ค่าความแม่นยำที่ดีที่สุดเป็นพีเจอร์ที่ใช้ตำแหน่งในวงจรถาวรเดินสี่ตำแหน่งโดยใช้ Support vector machine (SVM) โครงสร้างแบบ One VS One ในการจำแนกประเภท ซึ่งได้ค่าความแม่นยำเป็น 70% และนำเสนอโปรแกรมต้นแบบที่มีขอบเขตค่าปกติเพื่อช่วยให้สามารถนำระบบนี้ในการจำแนกผู้ป่วยได้ในเบื้องต้น

ภาควิชา...วิศวกรรมคอมพิวเตอร์... ลายมือชื่อ.....  
 สาขาวิชา...วิศวกรรมคอมพิวเตอร์... ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก.....  
 ปีการศึกษา.....2556..... ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม.....

# # 547 03435 21 : MAJOR COMPUTER ENGINEERING

KEYWORDS : MEDICAL DIAGNOSIS / GAIT DATA / MOVEMENT DISORDER

RATTIYA THANASOONTORNREK : LEARNING ALGORITHM FOR MEDICAL  
DIAGNOSIS ON MOVEMENT DISORDER USING GAIT DATA. ADVISOR:  
ASST.PROF. PIZZANU KANONGCHAIYOS, Ph.D., CO-ADVISOR: PEERAPON  
VATEEKUL, Ph.D., 74 PP.

Gait analysis is the process of collecting and analyzing the data of human walking movement. This kind of information is useful and widely used in many areas, especially in clinical research study. Doctors and physicians use it to identify a type of movement disorder diseases. Different illnesses require different types of doctors and treatments. This thesis presents an automated diagnosis system using gait data to classify patients into three groups: Normal, Sick/Knee Osteoarthritis (OA), and Sick/Parkinson's disease. In the study, there are 88 samples (patients): 27 elder normal, 34 osteoarthritis, and 27 Parkinson. The best classification scheme is based on a feature set of four major positions in the gait cycle and SVM using the One-VS-One strategy. The experimental results show that the proposed system achieved 70% accuracy. To primarily assist the diagnosis, the prototype system was implemented showing a comparison of gait cycle graphs between a patient and normal people.

Department : ...Computer Engineering Student's Signature.....

Field of Study : ...Computer Engineering Advisor's Signature.....

Academic Year : .....2013..... Co-advisor's Signature.....

## กิตติกรรมประกาศ

ในการจัดทำวิทยานิพนธ์นี้ ข้าพเจ้าขอขอบพระคุณอาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก ผู้ช่วยศาสตราจารย์ ดร.พิเชษฐ คนองชัยยศและ อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ดร.พีรพล เวทีกุล ที่ให้คำแนะนำทั้งการดำเนินงานวิจัย ความรู้ต่าง ๆ ที่มีประโยชน์ และเรื่องอื่น ๆ จนสามารถทำให้ข้าพเจ้าสามารถดำเนินการทำวิทยานิพนธ์สำเร็จลุล่วงไปได้ ข้าพเจ้าขอขอบพระคุณศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล ผู้ช่วยศาสตราจารย์ ดร.อานนท์ รุ่งสว่างและ ศาสตราจารย์ พญ.อารีรัตน์ สุพุทธิธาดา ที่กรุณาสละเวลาอันมีค่า มาให้คำแนะนำอันเป็นประโยชน์ต่อการดำเนินการทำวิทยานิพนธ์

ข้าพเจ้าขอขอบพระคุณ ศูนย์ความเป็นเลิศทางการแพทย์ด้านการเดินและเคลื่อนไหว โรงพยาบาลจุฬาลงกรณ์ ที่เอื้อเฟื้อเครื่องมือต่าง ๆ และข้อมูลที่เป็นต่อการศึกษาและทดลองในการทำวิทยานิพนธ์นี้ ขอขอบพระคุณศาสตราจารย์ พญ.อารีรัตน์ สุพุทธิธาดา และ นางสาวฉัตรแก้ว พงษ์มาลา เจ้าหน้าที่ศูนย์สำหรับความรู้และคำแนะนำที่มีประโยชน์

นอกจากนั้น ขอขอบคุณรุ่นพี่ รุ่นน้อง และเพื่อนร่วมงานในห้องปฏิบัติการคอมพิวเตอร์ กราฟิก ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่คอยให้กำลังใจ และคำแนะนำต่าง ๆ ในการทำวิทยานิพนธ์ครั้งนี้

และสุดท้ายข้าพเจ้าขอกราบขอบพระคุณบิดา มารดา และครอบครัว ที่ให้กำลังใจแก่ข้าพเจ้าในยามที่ท้อแท้ และสนับสนุนให้ข้าพเจ้าสามารถศึกษาและจัดทำวิทยานิพนธ์นี้ให้สำเร็จลุล่วงได้

## สารบัญ

หน้า

บทคัดย่อภาษาไทย .....	ง
บทคัดย่อภาษาอังกฤษ .....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ .....	ช
สารบัญตาราง.....	ญ
สารบัญรูป .....	ฐ
บทที่ 1 บทนำ.....	1
ความเป็นมาและความสำคัญของปัญหา .....	1
วัตถุประสงค์ของการวิจัย .....	2
ขอบเขตของการวิจัย .....	2
วิธีการดำเนินการวิจัย .....	2
ประโยชน์ที่คาดว่าจะได้รับ .....	3
ผลงานตีพิมพ์จากวิทยานิพนธ์.....	4
ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์.....	4
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง .....	6
ทฤษฎีที่เกี่ยวข้อง.....	6
1. นิยาม ความหมาย และความรู้เบื้องต้นด้านการแพทย์.....	6
1.1. การเดิน (Gait).....	6
1.2. การวิเคราะห์ท่าทางการเดิน (Gait analysis) .....	9

1.3. โรคข้อเสื่อม (Osteoarthritis) .....	11
1.4. โรคพาร์กินสัน (Parkinson's disease) .....	12
2. ทฤษฎีที่เกี่ยวข้องกับเทคนิคการจำแนกข้อมูลอัตโนมัติ .....	12
2.1. ต้นไม้ตัดสินใจ (Decision Tree) .....	13
2.2. Support Vector Machine .....	14
2.3. Cros Validation .....	16
งานวิจัยที่เกี่ยวข้อง.....	17
1. งานวิจัยทางการวิเคราะห์โรคพาร์กินสันจากข้อมูลการเดิน.....	17
2. งานวิจัยทางการวิเคราะห์โรคข้อเข่าเสื่อมจากข้อมูลข้อมูลโดยใช้ Neural network.....	19
บทที่ 3 วิธีการดำเนินการวิจัย.....	22
ภาพรวมของงานวิจัย .....	22
1. Data Collection.....	22
2. Data Preprocess .....	24
3. Data Analysis .....	27
การวัดผลการทดลอง.....	29
การเตรียมข้อมูล .....	30
บทที่ 4 การทดลองและผลการทดลอง.....	31
1. การทดลองวัดประสิทธิภาพของระบบโดยใช้ข้อมูลโดยใช้ชุดข้อมูล 54 คน .....	31
2. การทดลองวัดประสิทธิภาพของระบบโดยใช้ปริมาณจำนวนตัวอย่างเพิ่มขึ้น .....	37



3.	การทดลองเพื่อเพิ่มประสิทธิภาพโดยใช้วิธีการวิเคราะห์องค์ประกอบ (Principle Component Analysis) และวิธีการปรับช่วงข้อมูล (Data Normalization) .....	42
3.1.	การทดลองเพื่อเพิ่มประสิทธิภาพโดยวิธีการลดตัวแปรโดยใช้วิธีการวิเคราะห์องค์ประกอบ (Principle Component Analysis) .....	42
3.2.	การทดลองเพื่อเพิ่มประสิทธิภาพโดยการปรับช่วงข้อมูล (Data Normalization)....	44
3.3.	การทดลองเพื่อเพิ่มประสิทธิภาพโดยการลดตัวแปรโดยใช้วิธีการวิเคราะห์องค์ประกอบ (Principle Component Analysis) และ การปรับช่วงข้อมูล (Data Normalization) .....	46
4.	การทดลองเพื่อเพิ่มประสิทธิภาพโดยการลดตัวแปรโดยผู้เชี่ยวชาญ.....	49
4.1.	ทดลองวัดประสิทธิภาพโดยใช้ข้อมูลต้นฉบับพีเจอร์ SelectKinematic .....	51
4.2.	ทดลองวัดประสิทธิภาพโดยใช้ข้อมูลต้นฉบับพีเจอร์ SelectKinematic และวิธีการวิเคราะห์องค์ประกอบ (PCA) .....	52
4.3.	ทดลองวัดประสิทธิภาพโดยใช้ข้อมูลต้นฉบับพีเจอร์ SelectKinematic และเพิ่มการปรับช่วงข้อมูล .....	54
4.4.	ทดลองวัดประสิทธิภาพโดยใช้ข้อมูลที่มีการปรับช่วงข้อมูลและวิธีการวิเคราะห์องค์ประกอบ (PCA).....	55
5.	การทดลองเพื่อเปรียบเทียบประสิทธิภาพของโครงสร้าง Classifier.....	58
5.1.	การทดลองเพื่อวัดประสิทธิภาพโดยใช้โครงสร้างแบบ Two-level hierarchical classification .....	58
5.2.	การทดลองเพื่อวัดประสิทธิภาพโดยใช้โครงสร้างแบบ Binary Classifier Approach (One VS One).....	61
6.	ต้นแบบโปรแกรมของระบบวินิจฉัยโรค .....	63

6.1. ส่วนแสดงข้อมูล Kinematic .....	63
6.2. ส่วนแสดงข้อมูล Distance Temporal.....	64
6.3. ส่วนแสดงข้อมูลวิเคราะห์วินิจัยผล .....	65
บทที่ 5 ข้อสรุปและข้อเสนอแนะ .....	66
ข้อสรุป .....	66
ข้อเสนอแนะ .....	68
รายการอ้างอิง.....	70
ประวัติผู้เขียนวิทยานิพนธ์.....	74

## สารบัญตาราง

หน้า

ตารางที่ 3.1 ตารางแสดงรายการฟีเจอร์ Temporal data ที่ใช้ในการทดลอง .....	26
ตารางที่ 3.2 ตาราง macro-averaging ของ precision, recall, $F_1$ .....	30
ตารางที่ 4.1 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้ข้อมูลต้นฉบับ .....	33
ตารางที่ 4.2 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้ข้อมูลที่มีการเติม ค่าที่หายไป .....	35
ตารางที่ 4.3 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้ข้อมูลที่มี จำนวนตัวอย่างเพิ่มขึ้น .....	38
ตารางที่ 4.4 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้ PCA ในการลด ตัวแปร .....	43
ตารางที่ 4.5 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้ข้อมูลที่มีการปรับ ช่วงข้อมูล .....	45
ตารางที่ 4.6 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้ข้อมูลที่มีการปรับ ช่วงข้อมูลและผ่านการวิเคราะห์องค์ประกอบ (PCA) .....	47
ตารางที่ 4.7 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้ข้อมูลต้นฉบับ ฟีเจอร์ SelectKinematic .....	51
ตารางที่ 4.8 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้วิธีวิเคราะห์องค์ ประกอบกับ SelectKinematic .....	53
ตารางที่ 4.9 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้ข้อมูลที่มีการปรับ ช่วงข้อมูลกับ SelectKinematic .....	54
ตารางที่ 4.10 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้ข้อมูลที่มีการปรับ ช่วงข้อมูลและวิธีวิเคราะห์องค์ประกอบ กับฟีเจอร์ SelectKinematic .....	56

ตารางที่ 4.11 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโครงสร้างแบบ Hier โดยใช้ข้อมูลที่มีการปรับช่วงข้อมูล (Data Normalization) .....	59
ตารางที่ 4.12 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโครงสร้างแบบ Hier โดยใช้ข้อมูลที่มีการปรับช่วงข้อมูล (Data Normalization) และใช้วิธีวิเคราะห์ องค์ประกอบ (PCA) .....	60
ตารางที่ 4.13 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโครงสร้างแบบ One vs One .....	62

## สารบัญรูป

หน้า

รูปที่ 2.1	วงจรรการเดิน .....	7
รูปที่ 2.2	ลักษณะและค่าต่าง ๆ ของวงจรรการเดิน ช่วงเวลาที่เกิด Double Support .....	7
รูปที่ 2.3	ภาพแสดง Spatial parameters: stride and step lengths .....	8
รูปที่ 2.4	ภาพแสดงแกนตามการแพทย์ ได้แก่ Sagittal plane, Coronal plane และ Transverse plane.....	10
รูปที่ 2.5	กราฟตัวอย่าง Kinematic gait cycle ของคนปกติ .....	10
รูปที่ 2.6 (ก)	ภาพหัวเข้าของคนปกติ.....	11
รูปที่ 2.6 (ข)	ภาพหัวเข้าของโรคข้อเข่าเสื่อม .....	11
รูปที่ 2.7	ตำแหน่งของสมองส่วนกลางที่สร้างสารโคปามีน.....	12
รูปที่ 2.8	ลักษณะของผู้ป่วยพาร์กินสัน.....	13
รูปที่ 2.9	ตัวอย่างต้นไม้ตัดสินใจ.....	14
รูปที่ 2.10	ตัวอย่างการเลือกเส้นที่ดีที่สุดสำหรับการแบ่งกลุ่มข้อมูลทั้งสองออกจากกัน.....	16
รูปที่ 2.11	3-fold cross validation.....	17
รูปที่ 3.1	ภาพรวมการทำงานของระบบ.....	22
รูปที่ 3.2 (ก)	กล้องสำหรับตรวจจับมาร์คเกอร์ .....	23
รูปที่ 3.2 (ข)	มาร์คเกอร์สะท้อนแสง (reflective marker).....	23
รูปที่ 3.3	ตำแหน่งที่ติดมาร์คเกอร์แบบ Walking marker .....	23
รูปที่ 3.4 (ก)	กราฟแสดงข้อมูล Gait Cycle ที่มีค่าที่หายไป .....	27
รูปที่ 3.4(ข)	กราฟแสดงข้อมูล Gait Cycle ที่มีการเติมค่าที่หายไปแล้ว .....	27
รูปที่ 3.5	โครงสร้างแบบ Single multiclass classification (Flat) .....	28

รูปที่ 3.6 โครงสร้างแบบ Two-level hierarchical classification (Hier) .....	29
รูปที่ 3.7 โครงสร้างแบบ Binary classifier approach (One VS One) .....	29
รูปที่ 4.1 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพของเครื่องมือจำแนกระหว่าง SVM-Linear และต้นไม้ตัดสินใจโดยใช้ชุดข้อมูลต้นฉบับ .....	34
รูปที่ 4.2 กราฟแสดงค่า $F_1$ เปรียบเทียบประสิทธิภาพของเครื่องมือจำแนกระหว่าง SVM-Linear และต้นไม้ตัดสินใจโดยใช้ชุดข้อมูลต้นฉบับ .....	34
รูปที่ 4.3 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพของเครื่องมือจำแนกระหว่าง SVM-Linear และต้นไม้ตัดสินใจโดยใช้ชุดข้อมูลที่มีการเติมค่าที่หายไป .....	35
รูปที่ 4.4 กราฟแสดงค่า $F_1$ เปรียบเทียบประสิทธิภาพของเครื่องมือจำแนกระหว่าง SVM-Linear และต้นไม้ตัดสินใจโดยใช้ชุดข้อมูลที่มีการเติมค่าที่หายไป .....	36
รูปที่ 4.5 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลต้นฉบับกับชุดข้อมูลที่มีการเติมค่าที่หายไป โดยใช้เครื่องมือ SVM-Linear .....	36
รูปที่ 4.6 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพของเครื่องมือจำแนกระหว่าง SVM-Linear และต้นไม้ตัดสินใจโดยใช้ชุดข้อมูลต้นฉบับกลุ่มตัวอย่าง 88 คน .....	38
รูปที่ 4.7 กราฟแสดงค่า $F_1$ เปรียบเทียบประสิทธิภาพของเครื่องมือจำแนกระหว่าง SVM-Linear และต้นไม้ตัดสินใจโดยใช้ชุดข้อมูลต้นฉบับกลุ่มตัวอย่าง 88 คน .....	39
รูปที่ 4.8 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพของเครื่องมือจำแนกระหว่าง SVM-Linear และต้นไม้ตัดสินใจโดยใช้ชุดข้อมูลต้นฉบับกลุ่มตัวอย่าง 88 คนที่มีการเติมค่าที่หายไป .....	39
รูปที่ 4.9 กราฟแสดงค่า $F_1$ เปรียบเทียบประสิทธิภาพของเครื่องมือจำแนกระหว่าง SVM-Linear และต้นไม้ตัดสินใจโดยใช้ชุดข้อมูลต้นฉบับกลุ่มตัวอย่าง 88 คนที่มีการเติมค่าที่หายไป .....	40

รูปที่ 4.10 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลกลุ่ม  
ตัวอย่าง 88 คนต้นฉบับกับชุดข้อมูลกลุ่มตัวอย่าง 88 คนที่มีการเติมค่าที่หายไป  
โดยใช้เครื่องมือ SVM-Linear..... 41

รูปที่ 4.11 กราฟแสดงค่าความแม่นยำเปรียบเทียบระหว่างชุดข้อมูลต้นฉบับของจำนวน  
ตัวอย่าง 54 คน กับชุดข้อมูลต้นฉบับของจำนวนตัวอย่าง 88..... 41

รูปที่ 4.12 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลต้นฉบับ  
กับชุดข้อมูลต้นฉบับและใช้วิธีวิเคราะห์องค์ประกอบ (PCA) โดยใช้เครื่องมือ  
SVM-Linear ..... 44

รูปที่ 4. 13 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลต้นฉบับ  
กับชุดข้อมูลต้นฉบับที่ผ่านการปรับช่วงข้อมูล (Data Normalization) โดยใช้  
เครื่องมือ SVM-Linear ..... 45

รูปที่ 4.14 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลต้นฉบับ  
กับชุดข้อมูลต้นฉบับและใช้วิธีวิเคราะห์องค์ประกอบ (PCA) และชุดข้อมูลที่ผ่าน  
การปรับช่วงข้อมูล (Data Normalization) โดยใช้เครื่องมือ SVM-Linear ..... 48

รูปที่ 4.15 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลต้นฉบับ  
กับชุดข้อมูลต้นฉบับและใช้วิธีวิเคราะห์องค์ประกอบ (PCA) และชุดข้อมูลที่ผ่าน  
การปรับช่วงข้อมูล (Data Normalization) โดยใช้เครื่องมือ SVM-Linear ..... 49

รูปที่ 4.16 รูปวงจรรการเดิน ตำแหน่งที่ขีดเส้นใต้คือจุดที่สำคัญในการวินิจฉัยทางการแพทย์..... 50

รูปที่ 4.17 กราฟค่ามุมของข้อเข้าในแกน Sagittal..... 50

รูปที่ 4.18 ภาพแสดงตำแหน่งในวงจรรการเดินที่เลือกใช้ในการทดลอง..... 50

รูปที่ 4.19 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลต้นฉบับ  
 พีเจอร์ Kinematic กับชุดข้อมูลต้นฉบับพีเจอร์ SelectKinematic โดยใช้เครื่องมือ  
 SVM-Linear ..... 52

รูปที่ 4.20 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลต้นฉบับ  
 พีเจอร์ Kinematic ผ่านวิธีวิเคราะห์องค์ประกอบ (PCA) กับชุดข้อมูลต้นฉบับพีเจอร์  
 SelectKinematic ผ่านวิธีวิเคราะห์องค์ประกอบ (PCA) โดยใช้เครื่องมือ SVM ..... 53

รูปที่ 4.21 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลต้นฉบับ  
 พีเจอร์ Kinematic ผ่านการปรับช่วงข้อมูล (Data Normalization) กับชุดข้อมูล  
 ต้นฉบับพีเจอร์ SelectKinematic ผ่านการปรับช่วงข้อมูล (Data Normalization)  
 โดยใช้เครื่องมือ SVM-Linear ..... 55

รูปที่ 4.22 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพชุดข้อมูลต้นฉบับที่ผ่านการ  
 ปรับช่วงข้อมูล (Data Normalization) และวิธีวิเคราะห์องค์ประกอบ (PCA)  
 ระหว่างพีเจอร์ Kinematic กับพีเจอร์ SelectKinematic โดยใช้เครื่องมือ SVM ..... 56

รูปที่ 4.23 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพสภาพแวดล้อมของชุดข้อมูล  
 ต้นฉบับพีเจอร์ SelectKinematic ..... 57

รูปที่ 4.24 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลต้นฉบับ  
 พีเจอร์ Kinematic ผ่านการปรับช่วงข้อมูล (Data Normalization) กับชุดข้อมูล  
 ต้นฉบับพีเจอร์ SelectKinematic ผ่านการปรับช่วงข้อมูล (Data Normalization)  
 โดยใช้เครื่องมือ SVM-Linear ..... 59

รูปที่ 4.25 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพชุดข้อมูลต้นฉบับพีเจอร์  
 SelectKinematic ผ่านการปรับช่วงข้อมูล (Data Normalization) และใช้วิธี  
 วิเคราะห์องค์ประกอบ (PCA) ระหว่างโครงสร้างแบบ Flat กับโครงสร้างแบบ Hier  
 โดยใช้เครื่องมือ SVM-Linear ..... 61



รูปที่ 4.26 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพชุดข้อมูลต้นฉบับพีเจอร์  
SelectKinematic ผ่านการปรับช่วงข้อมูล (Data Normalization) โดยใช้เครื่องมือ  
SVM-Linear ระหว่างโครงสร้างแบบ Flat, Hier และ One VS One ..... 62

# บทที่ 1

## บทนำ

### ความเป็นมาและความสำคัญของปัญหา

การวิเคราะห์ข้อมูลการเดิน (Gait analysis) เป็นการศึกษาเกี่ยวกับข้อมูลการเดินของมนุษย์ รวมถึงการเก็บข้อมูลและการวิเคราะห์ข้อมูล ซึ่งข้อมูลเหล่านี้มีประโยชน์ในหลายด้าน [1],[2] ในทางการแพทย์ได้ใช้ข้อมูลเหล่านี้ไปวิเคราะห์โรคที่ส่งผลกระทบต่อ การเดิน เพื่อวินิจฉัย ดำเนินการรักษาและฟื้นฟูผู้ป่วย ตัวอย่างเช่น โรคข้อเข่าเสื่อม โรคพาร์กินสัน โรคอัมพาตครึ่งซีก การพิการตั้งแต่กำเนิด เป็นต้น โรคเหล่านี้มีสาเหตุที่ก่อให้เกิดโรคที่ต่างกันย่อมต้องการการวินิจฉัยของแพทย์ที่เฉพาะทางตามด้านนั้น ๆ และมีการรักษาที่แตกต่างกัน ยกตัวอย่างเช่น โรคพาร์กินสัน เป็นโรคทางระบบประสาทที่เกิดจากการเสื่อมของสมองส่วนกลาง แต่โรคข้อเข่าเสื่อมมีสาเหตุมาจากการเสื่อมของเข่าตามอายุและวิธีการใช้งานของข้อเข่าซึ่งไม่ได้เกี่ยวกับระบบประสาทแต่อย่างใด

เทคนิควิธีการเก็บข้อมูลการวิเคราะห์การเดินมีหลายวิธี เช่น การใช้กล้องในการบันทึกการเดิน การใช้ Force plates, ตัววัด goniometer เป็นต้น MoCap (Motion capture) [3] เป็นวิธีหนึ่ง ที่นิยมใช้ โดยระบบจะทำการเก็บข้อมูลโดยใช้กล้องในการจับท่าเดินของมนุษย์ กล้องจะทำการตรวจจับตัวมาร์คเกอร์ (reflective maker) ที่ติดตามข้อต่าง ๆ ของผู้ใช้ และแปลงค่าออกมาเป็นข้อมูลการเปลี่ยนแปลงขององศาตามข้อต่าง ๆ ได้แก่ กระดูกเชิงกราน, ข้อสะโพก, ข้อเข่า และข้อเท้า รวมถึงข้อมูลลักษณะการเดินอื่น ๆ เช่น ความเร็ว ระยะก้าว เป็นต้น

กระบวนการการวินิจฉัยนั้น แพทย์ผู้เชี่ยวชาญจะทำการอ่านและวิเคราะห์ผลของข้อมูลที่ ได้มา ซึ่งในขั้นตอนนี้ต้องใช้ผู้ที่มีความรู้ที่เฉพาะทางและใช้ผู้เชี่ยวชาญ ชำนาญอีกด้วย ในงานวิจัยที่ผ่านมีการทำระบบวิเคราะห์ข้อมูลการเดินแบบอัตโนมัติ โดยใช้ Neural network เพื่อจำแนกผู้ป่วยโรคข้อเข่าเสื่อม [4],[5],[6] หรือใช้ Neural network เพื่อจำแนกผู้ป่วยพาร์กินสัน [7] แต่ในระบบเหล่านี้มีการใช้ลักษณะข้อมูลการเดินที่แตกต่างกันออกไปตามวิธีการเก็บข้อมูล และในงานที่ผ่านมาจะวินิจฉัยโรคออกมาเพียงโรคเดียว ซึ่งโรคที่เกี่ยวข้องกับการเดินที่สามารถวิเคราะห์ได้จากข้อมูลการเดินนั้นมีหลายโรค

งานวิจัยนี้จึงนำเสนอระบบการวิเคราะห์ข้อมูลการเดินจาก Motion Capture โดยใช้เครื่องมือจำแนกประเภทมาตรฐาน ในการจำแนกผู้ป่วยที่มีโรคเกี่ยวกับการเดิน โดยจะแบ่งออกเป็น 3 คลาส คือ คนปกติ (Normal), ผู้ป่วยข้อเข่าเสื่อม (Sick/Knee OA), ผู้ป่วยพาร์กินสัน (Sick/Parkinson) เพื่อช่วยให้สามารถใช้ระบบนี้ในการจำแนกผู้ป่วยได้ในเบื้องต้น และให้ค่าความแม่นยำที่สูง รวมถึงมีการเลือกฟีเจอร์ที่ใช้ในการเรียนเพื่อเพิ่มประสิทธิภาพความแม่นยำของระบบ

### วัตถุประสงค์ของการวิจัย

เพื่อนำเสนอระบบวินิจฉัยโรคที่เกี่ยวข้องกับการเคลื่อนที่ด้วยข้อมูลการเดิน ในการจำแนกผู้ป่วยที่มีโรคเกี่ยวกับการเดินจะแบ่งออกเป็น 3 คลาส คือ คนปกติ (Normal), ผู้ป่วยข้อเข่าเสื่อม (Sick/Knee OA), ผู้ป่วยพาร์กินสัน (Sick/Parkinson) เพื่อช่วยให้สามารถใช้ระบบนี้ในการจำแนกผู้ป่วยได้ในเบื้องต้น รวมถึงศึกษาหาความสัมพันธ์ระหว่างฟีเจอร์และโรคที่เกี่ยวข้องกับการเดิน เพื่อศึกษาและเปรียบเทียบผลลัพธ์ความแม่นยำในแต่ละเครื่องมือมาตรฐานในการจำแนกประเภทข้อมูล

### ขอบเขตของการวิจัย

1. ข้อมูลที่ใช้ศึกษาได้มาจากศูนย์ความเป็นเลิศทางการแพทย์ด้านการเดินและเคลื่อนไหว โรงพยาบาลจุฬาลงกรณ์ เป็นข้อมูลที่ได้มาจากระบบ Motion Capture ชื่อว่า Qualisys
2. ศึกษาข้อมูลเฉพาะ kinematic data และ temporal data
3. ระบบวินิจฉัยแบ่งกลุ่มออกเป็น 3 กลุ่ม คือ กลุ่มปกติ, กลุ่มโรคข้อเข่าเสื่อม และกลุ่มโรคพาร์กินสัน
4. โรคที่ส่งผลต่อการเดินที่เลือกใช้จะเป็นโรคข้อเข่าเสื่อมแบบรุนแรงและโรคพาร์กินสันอย่างอ่อน ไม่รวมโรคอื่น ๆ ที่เกี่ยวข้องกับการเดิน
5. อายุของกลุ่มตัวอย่างเป็นผู้สูงอายุตั้งแต่ 60 ปีขึ้นไป

### วิธีการดำเนินการวิจัย

1. ขั้นตอนการศึกษาเบื้องต้น
  - 1.1. ศึกษาการทำงานของระบบในโรงพยาบาล

- 1.2. ศึกษาข้อมูลทางการแพทย์ ความหมายและความรู้ที่เกี่ยวกับการเดิน
- 1.3. ศึกษาข้อมูลและงานวิจัยที่เกี่ยวข้องกับการวินิจฉัยโรคจากข้อมูลการเดิน
2. ขั้นตอนการเก็บข้อมูล และเตรียมข้อมูล
  - 2.1. ศึกษาและเก็บข้อมูลดิบในโรงพยาบาล ข้อมูลคนปกติ ข้อมูลผู้ป่วยข้อเข่าเสื่อม ข้อมูลผู้ป่วยพาร์กินสัน
  - 2.2. เตรียมข้อมูล แปลงข้อมูลดิบให้อยู่ตรงตามรูปแบบข้อมูลที่จะใช้ในงานวิจัย
  - 2.3. เพิ่มขั้นตอนในการเติมข้อมูลที่หายไป
3. ขั้นตอนการออกแบบ พัฒนา ทดลองการวินิจฉัยโรคโดยใช้เครื่องมือจำแนกมาตรฐาน
  - 3.1. ออกแบบ พัฒนาเพื่อวัดประสิทธิภาพการวินิจฉัยโรคโดยใช้เครื่องมือจำแนกมาตรฐาน
  - 3.2. ทดลอง ทดสอบและวัดประสิทธิภาพการวินิจฉัยโรคโดยใช้เครื่องมือจำแนกมาตรฐาน
  - 3.3. สรุปและวิเคราะห์ผลของขั้นตอนวิธี
4. ขั้นตอนการพัฒนา ปรับปรุงข้อมูล
  - 4.1. เก็บข้อมูลเพิ่มขึ้น และปรับปรุงข้อมูล
  - 4.2. ขั้นตอนการพัฒนา การปรับช่วงข้อมูล (Normalized data)
5. ขั้นตอนออกแบบ พัฒนา นำวิธีการ Feature Selection มาใช้
  - 5.1. เพิ่มวิธีการ Feature Selection โดยใช้ principal component analysis (PCA)
6. ทดลอง ทดสอบและวัดประสิทธิภาพโดยใช้เครื่องมือจำแนกมาตรฐาน
7. เปรียบเทียบและวิเคราะห์ผลจากการทดลองทั้งหมด
8. วิเคราะห์และสรุปผล

#### **ประโยชน์ที่คาดว่าจะได้รับ**

1. ได้ลักษณะของข้อมูลและเครื่องมือในการจำแนกข้อมูลที่เหมาะสมที่ให้ความแม่นยำมากที่สุด
2. ได้ระบบวินิจฉัยโรคที่เกี่ยวข้องกับการเคลื่อนที่ด้วยข้อมูลการเดิน ในการจำแนกผู้ป่วยที่มีโรคเกี่ยวกับการเดินจะแบ่งออกเป็น 3 คลาส คือ คนปกติ (Normal), ผู้ป่วยข้อเข่าเสื่อม

(Sick/Knee OA), ผู้ป่วยพาร์กินสัน (Sick/Parkinson) เพื่อช่วยให้สามารถใช้ระบบนี้ในการ  
จำแนกผู้ป่วยได้ในเบื้องต้น

### ผลงานตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้รับการตอบรับและตีพิมพ์ในบทความทางวิชาการ 2 หัวข้อ  
เรื่องได้แก่

1. "Tree Induction for Diagnosis on Movement Disorders Using Gait Data" โดย  
"Rattiya Thanasoontornrer" , "Peerapon Vateekul", "Pizzanu Kanongchaiyos",  
"Chatkaew Pongmala" และ "Areerat Suputtitada" ในงานประชุมวิชาการนานาชาติ  
"2013 5th International Conference on Knowledge and Smart Technology (KST)"  
ณ มหาวิทยาลัยบูรพา จังหวัดชลบุรี ประเทศไทย ระหว่างวันที่ 31 มกราคม 2556 ถึงวันที่  
1 กุมภาพันธ์ 2556
2. "Kinect-based Gait Capture and Analysis System" โดย "Rattiya  
Thanasoontornrer", "Saran Sillapaphiromsuk", "Pattaravut Maleehuan", "Deacha  
Moungsri", "Jayakorn Vongkulbhisal" , "Chatkaew Pongmala" , "Peerapon  
Vateekul" , "Pizzanu Kanongchaiyos" และ "Areerat Suputtitada" ในงานประชุม  
วิชาการนานาชาติ "5th AUN/SEED-Net Regional Conference on Information and  
Communications Technology 2012 (RCICT 2012)" ณ ประเทศฟิลิปปินส์ ระหว่างวันที่  
18 ตุลาคม 2555 ถึงวันที่ 19 ตุลาคม 2555

### ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์

ในวิทยานิพนธ์เล่มนี้ได้แบ่งการนำเสนอออกเป็น 5 บทคือ

บทที่ 1 คือ บทนำซึ่งจะกล่าวถึงที่มาและความสำคัญของปัญหาในการทำวิจัย  
วัตถุประสงค์ของการวิจัย ขอบเขตในการวิจัย วิธีการดำเนินการวิจัย ประโยชน์ที่จะได้รับ รวมถึง  
ผลงานที่ได้เผยแพร่ และลำดับในการจัดเรียงเนื้อหาของวิทยานิพนธ์

บทที่ 2 กล่าวถึง 2 ส่วน ได้แก่ แนวคิดและทฤษฎี และงานวิจัยที่เกี่ยวข้อง โดยแนวคิดและทฤษฎีจะประกอบไปด้วย นิยาม ความหมาย และความรู้เบื้องต้นที่เกี่ยวข้อง ทฤษฎีที่เกี่ยวข้องกับเทคนิคการจำแนกข้อมูลอัตโนมัติ และทฤษฎีอื่น ๆ ที่เกี่ยวข้อง

บทที่ 3 ขั้นตอนภาพรวมของงานวิจัยหลัก ๆ การเก็บข้อมูล การเตรียมข้อมูล การวัดผลการทดลอง การเตรียมข้อมูลและรายละเอียดของข้อมูลที่จะนำไปใช้ในการทดลอง

บทที่ 4 กล่าวถึงขั้นตอนการทดลอง วิธีการทดลองและวัดประสิทธิภาพของระบบ และผลการทดลองทั้งหมด เปรียบเทียบเพื่อหาประสิทธิภาพที่ดีที่สุด

บทที่ 5 เป็นการสรุปผลการวิจัย วิเคราะห์ผลลัพธ์ที่ได้จากการทดลอง และข้อเสนอแนะของงานวิจัย

## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีที่เกี่ยวข้อง ซึ่งแบ่งเป็น 3 ส่วน คือ ส่วนแรกจะกล่าวถึง ทฤษฎีทาง นิยาม ความหมาย และความรู้เบื้องต้นด้านการแพทย์ที่เกี่ยวข้องกับวิทยานิพนธ์ ส่วนที่สองจะ กล่าวถึง ทฤษฎีที่เกี่ยวข้องกับเทคนิคการจำแนกข้อมูลอัตโนมัติ และส่วนที่สามจะกล่าวถึงงานวิจัย ที่เกี่ยวข้องกับวิทยานิพนธ์นี้

#### ทฤษฎีที่เกี่ยวข้อง

##### 1. นิยาม ความหมาย และความรู้เบื้องต้นด้านการแพทย์

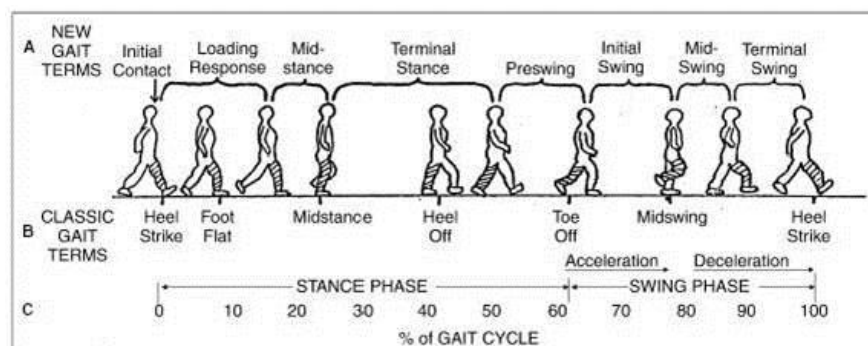
###### 1.1. การเดิน (Gait)

การเดิน เป็นการเคลื่อนไหวโดยใช้ขาทั้งสองข้างสลับกันอย่างเป็นจังหวะ และมีการ เคลื่อนไหวทางลำตัวให้เป็นจุดศูนย์ถ่วง ทำให้เกิดการเคลื่อนไหวไปข้างหน้า ซึ่งการเดินจะเป็นวงจร เรียกว่า วงจรการเดิน (Gait Cycle) ซึ่งจะแบ่งออกเป็นสองช่วงหลัก ได้แก่ ช่วงสแตนซ์ (Stance Phase) และ ช่วงสวิง (Swing Phase) ดังรูปที่ 2.1

ช่วงที่หนึ่งคือช่วง Stance Phase เป็นช่วงการเดินที่เท้าข้างใดข้างหนึ่งสัมผัสกับพื้น ซึ่งใช้ เวลาประมาณ 60 % ของวงจรการเดิน ซึ่งจะแบ่งออกเป็นห้าช่วงย่อย คือช่วงที่หนึ่ง Heel strike เป็นช่วงที่ส้นเท้าเริ่มสัมผัสกับพื้น จนเข้าสู่ช่วงที่สอง Foot flat เป็นช่วงที่เท้าเรียบติดพื้นและร่างกาย เริ่มทิ้งน้ำหนักลงที่เท้า จนถึงช่วงที่สาม Mid stance เป็นช่วงที่ร่างกายทิ้งน้ำหนักทั้งหมดลงกึ่งกลาง มวลร่างกายและขาเป็นแนวตั้งกับเท้า จนเมื่อถึงช่วงที่สี่เท้าเริ่มยกขึ้นจากพื้นจะเป็นช่วงที่สี่คือ Heel off และช่วงสุดท้ายคือ Toe off เป็นช่วงที่นิ้วเท้ายกขึ้น เพื่อจะเข้าสู่ช่วงสวิง (Swing Phase) เป็นห้าช่วงย่อยที่ต่อเนื่องในขาข้างเดียวกัน

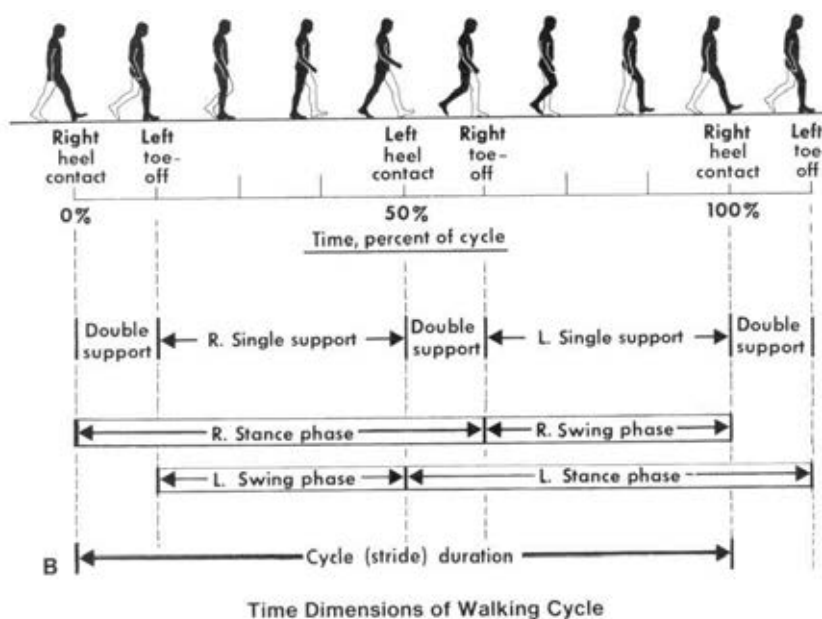
ช่วงที่สองคือช่วง Swing Phase เป็นช่วงการเดินที่เท้าข้างใดข้างนั้นลอยขึ้น ไม่สัมผัสกับพื้น ช่วงนี้ใช้เวลาประมาณ 40 % ของวงจรการเดิน แบ่งออกเป็นสามช่วงย่อย คือ เริ่มต้นจากช่วง Early swing เป็นช่วงเริ่มขึ้นของการที่เท้าลอยขึ้นจากพื้น จนถึงช่วง Mid swing เป็นช่วงกลางของการยก

เท้าซึ่งเท้าจะลอยจากพื้นสูงสุด และช่วงเข้าสู่ช่วง Late swing เป็นช่วงสุดท้ายที่เท้าจะลอยจากพื้น เตรียมเข้าสู่ช่วงสแตนด์ (Stance phase) ต่อไป



รูปที่ 2.1 วงจรการเดิน [8]

นอกจากสองช่วงหลักที่กล่าวไปแล้วข้างต้น ในวงจรการเดินยังมีช่วงดับเบิลซัพพอร์ต (Double Support) เป็นช่วงที่เท้าทั้งสองข้างสัมผัสพื้นพร้อมกันดังรูปที่ 2.2 โดยช่วงนี้จะเริ่มตั้งแต่ที่เท้าข้างหนึ่งกำลังอยู่ในลักษณะ Heel strike จนถึงขณะที่เท้าอีกข้างหนึ่งกำลังอยู่ในลักษณะ Toe off [1],[3],[9],[10]



รูปที่ 2.2 ลักษณะและค่าต่าง ๆ ของวงจรการเดิน ช่วงเวลาที่เกิด Double Support [9]  
ค่าพารามิเตอร์ที่ได้จากการเดินมีดังนี้

1.1.1. Spatial Parameter ได้แก่ (ดังรูปที่ 2.3)

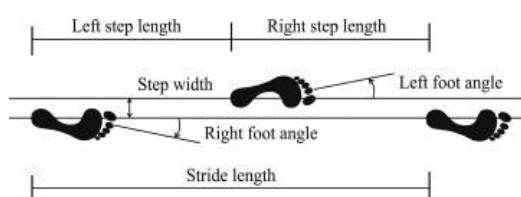


Step Length เป็นระยะทางระหว่าง heel strike ของเท้าข้างหนึ่งถึง heel strike ของเท้าอีกข้างหนึ่ง

Stride Length เป็นระยะทางระหว่าง heel strike ของเท้าข้างหนึ่งถึง heel strike อีกเท้าข้างเดิม

Foot angle (Angle of gait) คือมุมองศาของเท้าที่ออกจากแกนอ้างอิง

Base width (Base of gait) คือระยะห่างระหว่างจุดกึ่งกลางของเท้าข้างหนึ่งและเท้าอีกข้างหนึ่ง



รูปที่ 2.3 ภาพแสดง Spatial parameters: stride and step lengths [10].

#### 1.1.2. Temporal parameters

Step time คือ เวลาที่ใช้ในการเดินระหว่างช่วง heel strike ของเท้าข้างหนึ่งถึง heel strike ของเท้าอีกข้างหนึ่ง

Stride time คือ เวลาที่ใช้ในการเดินระหว่างช่วง heel strike ของเท้าข้างหนึ่งถึง heel strike ของเท้าข้างเดิม

Single support คือ เวลาทั้งหมดที่เท้าข้างใดข้างหนึ่งสัมผัสกับพื้น

Double support คือ เวลาที่เท้าทั้งสองข้างสัมผัสกับพื้น

Swing time คือ เวลาที่เท้าข้างนั้น ๆ กำลังก้าวลอยขึ้น ไม่สัมผัสกับพื้น

Stance time คือ เวลาที่เท้าข้างนั้น ๆ สัมผัสกับพื้น

Total support คือ เวลาทั้งหมดที่ขาข้างนั้น ๆ ใช้ทั้งหมดในหนึ่งวงจรกิจกรรมเดิน นั่นคือเวลาของ single support และเวลาของ double support รวมกัน

Cycle time คือ เวลาทั้งหมดในหนึ่งรอบวงจรการเดิน เวลาที่ใช้ในการเดินระหว่างช่วง heel strike ของเท้าข้างหนึ่งถึง heel strike ของเท้าข้างเดิม

### 1.1.3. พารามิเตอร์อื่น ๆ

Cadence คือ จำนวนก้าวทั้งหมดในหนึ่งนาที

Velocity คือ ความเร็วของการเดิน

$$Velocity = \frac{Step\ length(m) \times cadence\ (steps/min)}{60}$$

Symmetry of step length คือ อัตราส่วนระหว่าง step length ของข้างซ้ายต่อ step length ของข้างขวา

$$Symmetry\ of\ step\ length = \frac{Step\ length\ of\ the\ left}{Step\ length\ of\ the\ right}$$

Symmetry of step time คือ อัตราส่วนระหว่าง step time ของข้างซ้ายต่อ step time ของข้างขวา

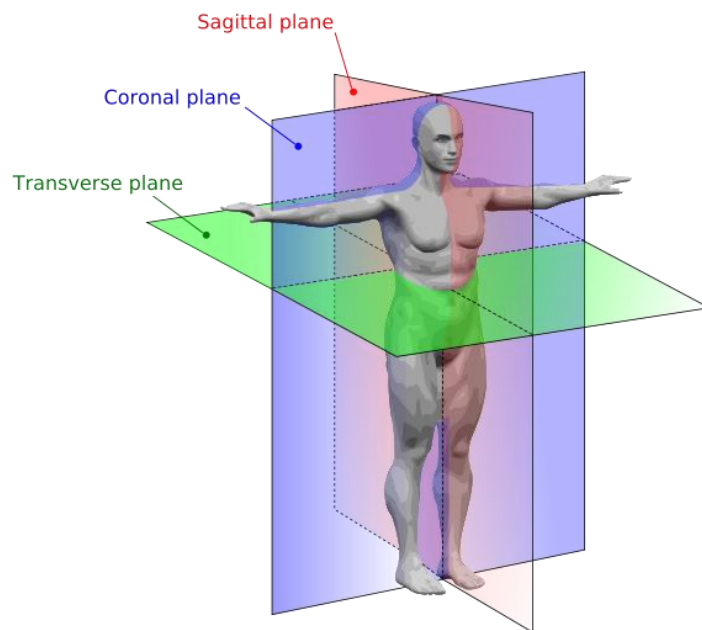
$$Symmetry\ of\ step\ time = \frac{Step\ time\ of\ the\ left}{Step\ time\ of\ the\ right}$$

## 1.2. การวิเคราะห์ท่าทางการเดิน (Gait analysis)

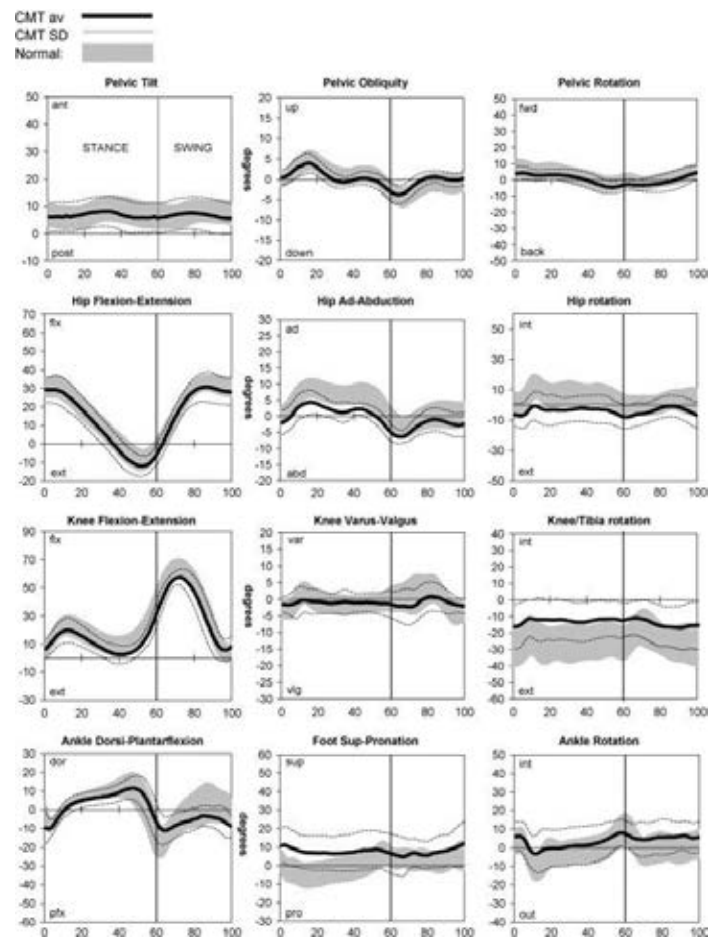
การวิเคราะห์ท่าทางการเดิน (Gait analysis) เป็นขั้นตอนในการเก็บข้อมูลรวมถึงวิเคราะห์แปลงผลรูปแบบท่าเดินของมนุษย์ ซึ่งข้อมูลเหล่านี้มีประโยชน์ในหลายด้าน โดยเฉพาะด้านการแพทย์ ในการเก็บข้อมูลท่าทางการเดินมีหลายวิธี เช่น force plates, video camera, goniometer, motion capture เป็นต้น

ค่ามาตรฐานการวิเคราะห์ท่าทางการเดิน (Measurement of Gait analysis) มีดังนี้

1.2.1. Kinematic data เป็นข้อมูลทางด้านมุมมองศาของตำแหน่งแต่ละข้อ ถม เวลานั้นๆ ซึ่งเป็นค่าองศาในแต่ละข้อตามแกนทางการแพทย์ คือ sagittal plane, coronal plane, traverse plane (รูปที่ 2.4) โดยจะไม่สนใจแรงน้ำหนักที่เกิดขึ้นขณะเคลื่อนที่ ส่วนมากจะแสดงข้อมูลเป็นกราฟ ดังรูปที่ 2.5 ทางกายภาพจะอาศัยการอ่านแนวโน้มของกราฟในการวิเคราะห์โรค รวมถึงการรักษาฟื้นฟู



รูปที่ 2.4 ภาพแสดงแกนตามการแพทย์ ได้แก่ Sagittal plane, Coronal plane และ Transverse plane [11].



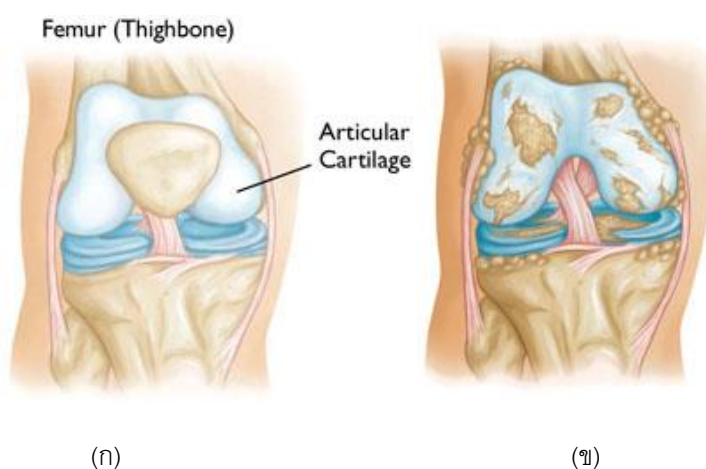
รูปที่ 2.5 กราฟตัวอย่าง Kinematic gait cycle ของคนปกติ [12].

1.2.2. Temporal data เป็นข้อมูลที่บอกระยะเวลาและระยะห่างของการเดิน รวมถึงความเร็ว และ จำนวนก้าว ได้แก่ Stance Time, Swing Time, Step Time, Stride Time, Step Time per minute, Stride Time per minute, Single Support, Double Support, Step Length, Stride Length, Cadence, Velocity เป็นต้น ซึ่งข้อมูลส่วนนี้ ก็ใช้ในการวิเคราะห์ทางการแพทย์เช่นกัน

1.2.3. Kinetics เป็นข้อมูลที่บอกแรงกดที่เกิดขึ้นขณะเคลื่อนไหว ณ เวลานั้น ๆ บริเวณเท้าต่อพื้นทั้งการกดหรือการวางเท้ากับพื้น ซึ่งข้อมูลแบบนี้ได้มาจากแผ่น foot plate รูปแบบพารามิเตอร์ที่นิยมใช้กับข้อมูล kinetics ได้แก่ Ground reaction force (GRF)

### 1.3. โรคข้อเสื่อม (Osteoarthritis)

โรคข้อเสื่อมเป็นโรคที่พบบ่อยในผู้สูงอายุ เกิดจากข้อต่อปกติ (รูปที่ 2.6ก) มีความเสื่อมสภาพจากการใช้งาน ผิวกระดูกอ่อนบริเวณข้อมีการสึกกร่อนเปลี่ยนแปลง (รูปที่ 2.6ข) และบริเวณกระดูกมีการเปลี่ยนแปลงด้วยเช่นมีขอบที่หนาขึ้น ทำให้เกิดอาการปวด ทำให้การเคลื่อนไหวของข้อนั้นลดลง ซึ่งตำแหน่งที่พบโดยส่วนใหญ่จะเป็นตำแหน่งข้อเข่า ลักษณะอาการที่พบได้คือ มีอาการปวดบริเวณข้อเมื่อมีการใช้งาน ข้อฝืดเกิดอาการฝืดขณะทำท่าทางการงอหรือการเหยียด ข้อบวมและผิดรูป สูญเสียการเคลื่อนไหวและการทำงาน เดินไม่สะดวกหรือไม่ถนัด มีเสียงดังกรอบแกรบที่ข้อในขณะเคลื่อนไหว



รูปที่ 2.6 (ก) ภาพหัวเข่าของคนปกติ (ข) ภาพหัวเข่าของโรคข้อเข่าเสื่อม [13].

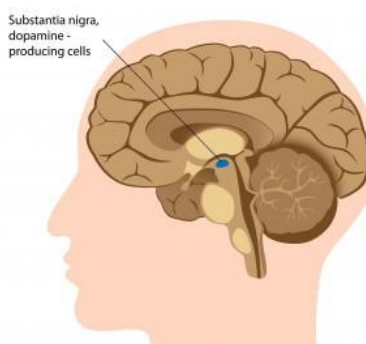
#### 1.4. โรคพาร์กินสัน (Parkinson's disease)

โรคพาร์กินสันเป็นโรคทางระบบประสาทที่เกิดขึ้นในผู้สูงอายุเป็นส่วนใหญ่ เกิดขึ้นจากการเสื่อมของสมองส่วนกลาง (รูปที่ 2.7) ที่มีเซลล์ชื่อว่า นิวรอน (neuron) ได้ตายหรือมีการทำงานที่เสื่อมลง เป็นสาเหตุทำให้การสร้างโดปามีนนั้นไม่เพียงพอ สารโดปามีนเป็นสารสื่อประสาทที่มีผลต่อการแสดงออก ทำหน้าที่ในการแสดงอารมณ์ รวมถึงการประสานงานการทำงานของกล้ามเนื้อ และการเคลื่อนไหวของร่างกายระหว่างสมองส่วนกลาง (substantia nigra) และสมองส่วนอื่น ๆ (corpus striatum) โดยทั่วไปโรคพาร์กินสันจะแบ่งเป็น 2 ประเภท คือ ประเภทที่เกี่ยวข้องกับการเคลื่อนไหว (Motor PD) และประเภทที่ไม่เกี่ยวข้องกับการเคลื่อนไหว (Non-motor PD)

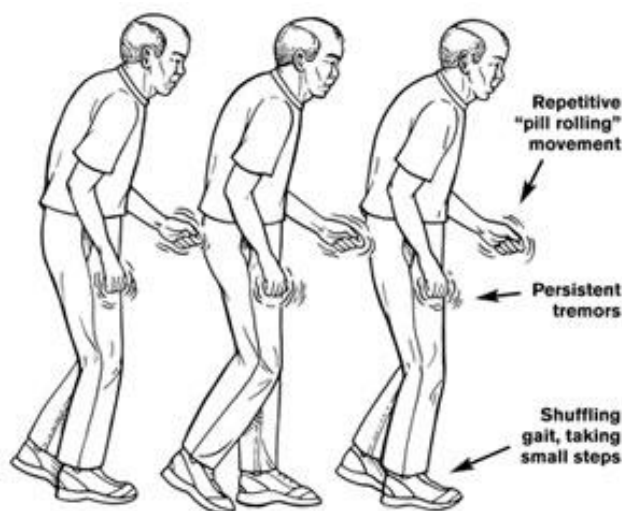
##### 1.4.1. Motor PD โดยปกติจะมีอาการ 4 แบบ

- 1.4.1.1. Handing shaking มีอาการสั่นบริเวณนิ้วหัวแม่มือและนิ้วชี้คล้ายลักษณะการปั้นเม็ดยา (pill-rolling) ลักษณะดังรูปที่ 2.8
- 1.4.1.2. Rigidity มีอาการเกร็งแขนและลำตัว
- 1.4.1.3. Bradykinesia เป็นอาการเคลื่อนไหวช้าและลำบาก สูญเสียการควบคุมการเคลื่อนไหว
- 1.4.1.4. Postural instability เสียการทรงตัว ซึ่งเป็นสาเหตุให้ผู้ป่วยง่ายต่อการล้ม

1.4.2. Non-motor PD เป็นอาการทางจิตประสาท ได้แก่ สะเทือนใจ กังวล เฉยเมย อารมณ์แปรปรวน อาการหลงลืม เคี้ยวอาหารและกลืนอาหารลำบาก ท้องผูก กลั้นปัสสาวะไม่ได้ มีปัญหาในการนอนหลับ เป็นต้น



รูปที่ 2.7 ตำแหน่งของสมองส่วนกลางที่สร้างสารโดปามีน [14].



รูปที่ 2.8 ลักษณะของผู้ป่วยพาร์กินสัน (วาดโดย Kathryn Born) [15].

## 2. ทฤษฎีที่เกี่ยวข้องกับเทคนิคการจำแนกข้อมูลอัตโนมัติ

### 2.1. ต้นไม้ตัดสินใจ (Decision Tree)

ต้นไม้ตัดสินใจ (Decision Tree) เป็นการจำแนกข้อมูลออกเป็นกลุ่ม ๆ โดยพิจารณาจากคุณลักษณะของข้อมูล ซึ่งจะแสดงผลในรูปแบบของต้นไม้ (รูปที่ 2.9) ประกอบไปด้วยจุดเริ่มต้นเรียกว่า โหนดราก และแตกกิ่งไปโหนดถัดไป ซึ่งแต่ละโหนดคือคุณลักษณะของข้อมูล โดยมีกิ่งเป็นค่าของคุณลักษณะของโหนดนั้น จะแตกกิ่งไปเรื่อย ๆ จนถึงโหนดใบจะเป็นผลลัพธ์ของการจำแนกประเภทข้อมูล เทคนิคการเรียนรู้ของต้นไม้ตัดสินใจแบบ C4.5 เป็นอัลกอริทึมที่นิยมทั่วไป พัฒนาขึ้นโดย J.Ross Quinlan [16] จะใช้ค่ามาตรฐานอัตราส่วนเกน (Gain Ratio) เพื่อเลือกคุณลักษณะที่จะใช้เป็นโหนด โดยการเลือกคุณลักษณะนั้นดูจากค่าเกนความรู้ (Information gain) ซึ่งจะต้องมีการหาค่าเอนโทรปี (Entropy) มีนิยามดังนี้

Entropy:

$$E(S) = - \sum_{j=1}^n p_S(j) \log_2 p_S(j)$$

โดยที่  $S$  คือ แอตทริบิวต์ที่เราสนใจ

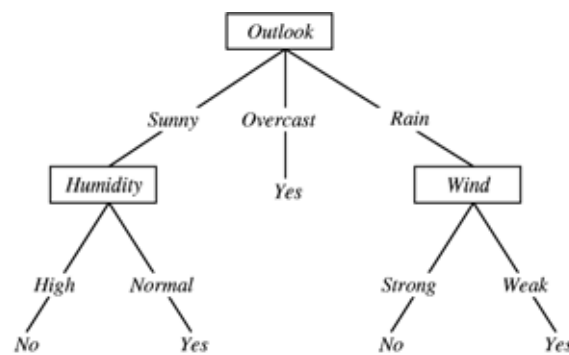
$p_S(j)$  คือ สัดส่วนของจำนวนสมาชิกของกลุ่ม  $j$  กับจำนวนสมาชิกทั้งหมดของกลุ่ม

ตัวอย่าง

Information Gain:

$$Gain(S, A) = E(S) - \sum_{v=value(A)} \frac{|S_v|}{|S|} E(S_v)$$

- โดยที่  $A$  คือ แอตทริบิวต์หรือตัวแปรที่เราสนใจ  
 $E$  คือ ค่าเอนโทรปี  
 $|S_v|$  คือ สมาชิกของแอตทริบิวต์  $A$  ที่มีค่า  $v$   
 $|S|$  คือ จำนวนสมาชิกของกลุ่มตัวอย่าง



รูปที่ 2.9 ตัวอย่างต้นไม้ตัดสินใจ

ลักษณะของต้นไม้ตัดสินใจ จะมีความทนทานต่อข้อมูลรบกวน (noisy data) มีความรวดเร็วในการเรียนรู้ ในแต่ละเส้นทางตั้งแต่ไหนตรงไปจนถึงไหนใดสามารถเขียนในรูปแบบกฎ IF-THEN ได้ โมเดลที่ได้จากการเรียนรู้แสดงผลในรูปแบบที่สามารถเข้าใจได้ง่าย ทำให้สามารถวิเคราะห์คุณลักษณะหรือเหตุผลในการจำแนกข้อมูลของกลุ่มต่าง ๆ

## 2.2. Support Vector Machine

Support Vector Machine (SVM) [17] เป็นอัลกอริทึมที่ใช้หลักการของการหาสัมประสิทธิ์ โดยสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ดีที่สุด สำหรับใช้ในการจำแนกหรือวิเคราะห์ข้อมูล ก่อนป้อนเข้าสู่กระบวนการสอนให้ระบบเรียนรู้ต่อไป การทำงานของ SVM เริ่มจากนำค่าข้อมูลมาวางลงใน Feature Space แล้วจึงหาเส้นที่ดีที่สุดที่ใช้แบ่งข้อมูลทั้งสองกลุ่มออกจากกัน เรียกว่า Hyperplane แต่เดิมแล้วแนวคิดของ SVM ถูกออกแบบมาสำหรับใช้กับกลุ่มข้อมูลที่เป็นเชิงเส้น แต่โดยส่วนมากแล้วข้อมูลที่น่ามาใช้ในการสอนให้ระบบเรียนรู้มักจะเป็นข้อมูลที่ไม่เป็นเชิง

เส้น ซึ่งสามารถแก้ไขปัญหาดังกล่าวได้โดยการใช้ Kernel Function สำหรับการจำแนกข้อมูลบนระนาบหลายมิติ จะใช้การแบ่งส่วนกลุ่มข้อมูลที่มีความเหมาะสมที่สุด เรียกว่า Feature Selection ซึ่งเลือกมาจากข้อมูลที่ได้สอนให้ระบบเรียนรู้ และเซตของข้อมูลที่ใช้อธิบายสำหรับแต่ละกรณี เรียกว่า Vector จากที่กล่าวมานี้ SVM มีจุดมุ่งหมายคือการแบ่งแยกกลุ่มของ Vector ในแต่ละระนาบออกจากกัน ซึ่งเวกเตอร์ที่อยู่ข้างระนาบหลายมิติทั้งหมดเรียกว่า Support Vectors ดังสมการที่ (1) เป็นสมการแสดงเวกเตอร์

$$\phi(w, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \quad (1)$$

โดยที่

$$y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0 ; \\ i = 1, \dots, N$$

เมื่อ	$C$	คือ ค่าคงที่ของความจุ (capacity constant)
	$w$	คือ ค่าเวกเตอร์ coefficient
	$i$	คือ ค่าดัชนี
	$N$	คือ จำนวนข้อมูลที่ใช้สำหรับ training
	$\xi_i$	คือ ค่า slack variable ซึ่งเป็นพารามิเตอร์ของข้อมูลผิดพลาด
	$x_i$	คือ ค่าตัวแปรอิสระ
	$b$	คือ ค่าคงที่
	$y$	คือ ค่าของ class; $y \in \pm 1$
	$\phi$	คือ ค่า kernel ซึ่งควรมีค่ามากกว่า $C$

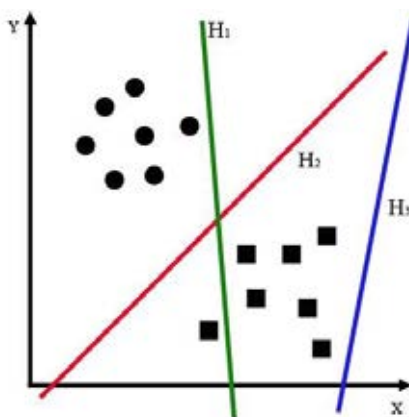
Kernel Functions:

$$K(X_i, X_j) = \begin{cases} X_i \cdot X_j, & \text{Linear} \\ (\gamma X_i \cdot X_j + C)^d, & \text{Polynomial} \\ \exp(-\gamma |X_i - X_j|^2), & \text{RBF} \\ \tanh(\gamma X_i \cdot X_j + C), & \text{Sigmoid} \end{cases}$$

เมื่อ  $K(X_i, X_j) = \phi(X_i) \cdot \phi(X_j)$

$i, j$  คือ ค่าดัชนี



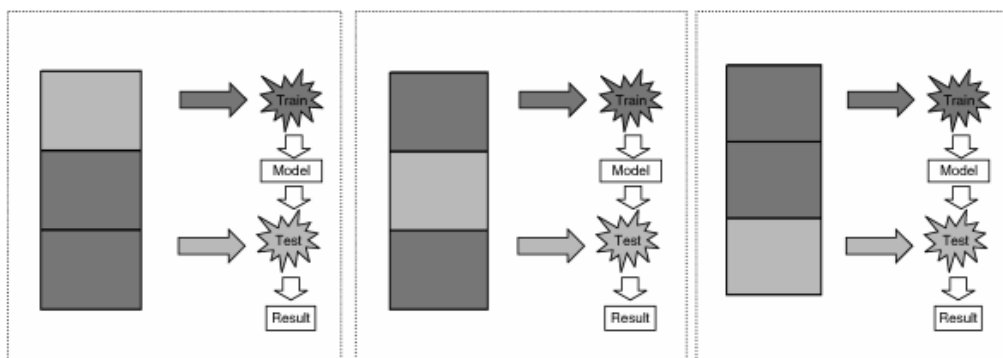


รูปที่ 2.10 ตัวอย่างการเลือกเส้นที่ดีที่สุดสำหรับการแบ่งกลุ่มข้อมูลทั้งสองออกจากกัน

จากรูปที่ 2.10 สมมติว่าต้องการคัดแยกข้อมูลออกเป็น 2 กลุ่ม ด้วยการใช้เส้นแบ่งที่เป็นเส้นตรง จะเห็นว่าสามารถเลือกเส้นตรงได้หลายเส้น เช่น เส้นตรง  $H_1$ ,  $H_2$  หรือ  $H_3$  แต่สำหรับการเลือกเส้นที่ดีที่สุด จะสามารถทำได้โดยการใช้ค่า Margin ซึ่งเป็นค่าที่แสดงถึงผลรวมระยะห่างระหว่างเส้นแบ่ง กับเส้นตรงที่ขนานกันที่ใกล้จุดกึ่งกลางของกลุ่มข้อมูลมากที่สุด จากรูปที่ 2.10 จะเห็นได้ว่า เส้นตรง  $H_2$  เป็นเส้นตรงที่สามารถแบ่งกลุ่มข้อมูลทั้งสองกลุ่มออกจากกันได้ดีที่สุด นั่นคือมีค่า margin สูงที่สุด ซึ่งเราเรียกข้อมูลที่อยู่บน margin นี้ว่า Support Vector

### 2.3. Cross Validation

การตรวจสอบไขว้กัน (Cross Validation) [18] เป็นวิธีทางสถิติสำหรับการประเมินและเปรียบเทียบอัลกอริทึมในการเรียนรู้โดยจะทำการแบ่งข้อมูลออกเป็นสองส่วน ส่วนแรกเป็นข้อมูลใช้สำหรับการเรียนรู้เพื่อสร้างโมเดล และส่วนที่สองเป็นข้อมูลที่ใช้สำหรับตรวจสอบโมเดล ข้อมูลชุดเรียนรู้และชุดตรวจสอบจะทำการสลับสับเปลี่ยนวนเป็นรอบ ทำให้ทุกชุดข้อมูลจะถูกนำมาตรวจสอบโมเดลเสมอ ซึ่ง k-fold cross validation เป็นรูปแบบทั่วไปที่นิยมใช้ โดยจะแบ่งข้อมูลออกเป็น k ชุดเท่า ๆ กัน จะทำการประเมินค่าจำนวน k รอบ ในแต่ละรอบจากข้อมูลทั้งหมด k ชุด จะถูกเลือกหนึ่งชุดเพื่อเป็นชุดตรวจสอบ และ ข้อมูลจำนวน k-1 ชุดจะใช้เป็นชุดข้อมูลสำหรับการเรียนรู้ อย่างในรูปที่ 2.11 เป็นการทำให้ 3-fold validation



รูปที่ 2.11 การทำ 3-fold cross validation [17]

### งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้อง จะแบ่งเป็นงานวิจัยทางการวิเคราะห์โรคพาร์กินสันจากข้อมูลการเดิน และงานวิจัยทางการวิเคราะห์โรคข้อเข่าเสื่อมจากข้อมูลโดยใช้ Neural network

#### 1. งานวิจัยทางการวิเคราะห์โรคพาร์กินสันจากข้อมูลการเดิน

H. Hany et al. [7] กล่าวถึงงานวิจัยต่าง ๆ ที่เกี่ยวข้องกับจากวินิจฉัยโรคพาร์กินสันจากการเดิน ซึ่งจะสังเกตได้ว่าค่าเฉลี่ยของความเร็วในการเดินและค่าระยะก้าวเท้า (stride Length) ของคนปกติจะมีค่าสูงกว่าเมื่อเทียบกับผู้ป่วยพาร์กินสันและผู้ป่วยจะใช้เวลาในการเดินนานกว่าต่อหนึ่งรอบการเดิน (gait cycle) เมื่อเทียบจากค่า Kinematic มาตรฐานแล้วมุมมองศาของขาช่วงล่าง จะมี แอมพลิจูดที่ลดลงจากคนปกติ จากงานวิจัยของ Morris et al. ที่ศึกษาผู้หญิงอายุ 71 ปีที่เป็นโรคพาร์กินสันสรุปได้ว่าจะมีความเร็ว, จำนวนก้าวต่อนาที, ระยะก้าวเท้า (stride Length), มุมองศาของสะโพก ข้อเข่า ข้อเท้าจะลดลงจากคนทั่วไป รวมถึงการเคลื่อนไหวของข้อในระนาบ Sagittal และแรงที่สะโพก ข้อเข่า ข้อเท้าจะลดลงอีกด้วย

Delval et al. [8] แสดงให้เห็นว่าความเร็วในการเดิน (walking speed) จำนวนเท้าในหนึ่งนาที (cadence) และระยะก้าวเท้า (stride Length) ในผู้ป่วยพาร์กินสัน 32 คน มีค่าน้อยกว่าเมื่อเทียบกับผู้สูงอายุที่ปกติ 30 คน ในขณะที่ผู้ป่วยจะใช้เวลาในการก้าวเท้า (stride Time) เพิ่มมากขึ้น และค่า Hip extension, Knee flexion จะต่ำกว่าคนปกติ Azulay et al. และ Wu&Krishnan ได้รายงานไว้ว่าเวลาที่ใช้ในการก้าวเท้า (Stride Time) ของผู้ที่เป็นพาร์กินสันจะสูงกว่าปกติ

H. Hany et al. [8] ได้ทำการทดลองเพื่อยืนยันงานวิจัยข้างต้น ซึ่งทดสอบโดยใช้ 3 พารามิเตอร์ คือ ข้อมูลพื้นฐาน, Kinematic และ Kinetic เพื่อตรวจสอบลักษณะฟีเจอรของการเดิน สำหรับตรวจจับโรคพาร์กินสัน ซึ่งข้อมูลพื้นฐานที่ใช้ได้แก่ stride time, step length และ walking speed ส่วนข้อมูล Kinematic parameter ได้แก่ มุมองศาของข้อตามแต่ละเวลา ซึ่งจะได้แก่ สะโพก ข้อเข่า ข้อเท้า และข้อมูล Kinetic parameter เป็นค่าที่ได้จากแรงบริเวณเท้าต่อพื้น ซึ่งได้ใช้ค่า Ground reacting force (GRF) เป็นมาตรฐาน งานวิจัยนี้ได้ทำการทดลองโดยใช้คนปกติ (Control Group) 20 คน อายุ 19-30 ปี และผู้ป่วยพาร์กินสัน 12 คน เป็น Test Group โดยข้อมูลการเดินมาจากซอฟต์แวร์ Vicon® โดยการทดลองแรกทำการนำค่าข้อมูลพื้นฐาน (basic gait parameter) มาหาค่า Mean, Standard deviation และ p-value ของทั้งกลุ่ม คนปกติ (Normal Subject) และกลุ่มพาร์กินสัน (PD Subject) ผลปรากฏว่าค่าของ จำนวนก้าวในหนึ่งนาที (cadence), ระยะก้าวเท้า (Stride Length) และความเร็วในการเดิน (walking speed) ของกลุ่มพาร์กินสันจะมีค่าต่ำกว่ากลุ่มปกติ ในขณะที่เวลาที่ใช้ในการก้าวเท้า (Stride Time) สูงกว่า และผลที่ได้จากการทดลองค่า p-value แสดงให้เห็นว่าค่า Step length และค่าความเร็วในการเดิน (walking speed) มีนัยยะสำคัญในการทดลองที่สองเป็นการนำค่ามุมองศาของข้อต่าง ๆ ได้แก่ Hip angle, Knee angle และ Ankle angle ณ ตำแหน่ง Heel Strike, Toe off, Max flexion, Max extension มาหาค่า Mean, Standard deviation และ p-value ของทั้งกลุ่ม คนปกติ (Normal Subject) และ กลุ่มพาร์กินสัน (PD Subject) ได้ผลคือค่ามุมของแต่ละข้อของคนปกติจะมีค่าสูงกว่ากลุ่มพาร์กินสัน ยกเว้นค่า heel strike ของ ankle การทดลองที่สามได้นำค่า Kinetic มาทำการทดลองเช่นเดียวกับการทดลองที่ผ่านมา ได้ผลดังนี้ค่า mean ทั้ง vertical และ horizontal GRF สูงกว่าในกลุ่มคนปกติ จากผลการทดลองที่ผ่านมาจะมี 4 ฟีเจอรของการเดินที่มีผลต่อกลุ่มพาร์กินสันคือ step length, walking speed, knee angle และ ground reaction force (โดยเฉพาะค่า Fz2) สำหรับการทดลองสุดท้าย จะทำการวัดค่า performance โดยใช้ ANN Classifier (โครงข่ายประสาทเทียม : Artificial neural network) และใช้การประเมินค่าแบบ 4-fold cross-validation ซึ่งได้ผลการทดลองจะได้ว่าค่าฟีเจอรที่มีลักษณะเด่นทั้งสี่ตัวจากจากการทดลอง 1-3 ซึ่งมีชื่อฟีเจอรคือ Four significant features selection จะได้ค่าความแม่นยำสูงที่สุด เป็น 95.63%

## 2. งานวิจัยทางการการวิเคราะห์โรคข้อเข่าเสื่อมจากข้อมูลข้อมูลโดยใช้ Neural network

J. McBride [5] ได้ทำทดลองเพื่อการวิเคราะห์ข้อมูลการเดินโดยใช้โมเดล Neural network ในการทำนายระดับความเสื่อมของข้อระดับความเจ็บปวดของโรคข้อเข่าเสื่อม โดยใช้ผู้ร่วมทดสอบ 30 คน อายุ 60-85 ปี มีอาการข้อเข่าเสื่อม และได้แบ่งออกเป็น 3 กลุ่ม คือ กลุ่มที่ไม่ได้มีการฝึกร่างกาย (Group1:Control Group) กลุ่มที่มีการฝึกกล้ามเนื้อ (Group2) กลุ่มที่มีการเล่นไทเก๊ก (Group3) ข้อมูลการเดินของผู้เข้าร่วมได้มาจาก Ground reaction forces และ Three-dimensional kinematic data ซึ่งจะได้เวกเตอร์ที่ใช้ในการวัดทั้ง kinematic และ kinetic มีทั้งหมด 21 เวกเตอร์ ได้มาจาก มุมของข้อเท้า (ankle angle), อัตราเร็วเชิงมุมของข้อเท้า (ankle angular velocity), มุมของข้อเข่า (knee angle), อัตราเร็วเชิงมุมของข้อเข่า (knee angular velocity), มุมของสะโพก (hip angle), อัตราเร็วเชิงมุมของสะโพก (hip angular velocity) และ Ground reaction force จากการเดินของผู้เข้าร่วม ทำให้มีข้อมูลทั้งหมด 300 ระเบียบ ในการทดลองแรกจะทำการแบ่งประเภท (classify) โดยใช้วิธี Kellgren/Lawrence scores (K/L scores) range 1-5 ใช้ข้อมูล 200 ระเบียบจาก 20 คน เป็น train set และ 100 ระเบียบเป็น test set โดยขั้นแรกได้นำทั้ง 21 เวกเตอร์มาใช้ ซึ่งได้ค่าความแม่นยำเฉลี่ย 67% ต่อมาได้ทำการ Different combination เวกเตอร์ ทำให้ได้ผลลัพธ์เวกเตอร์ที่เลือกใช้เป็นข้อมูลมุม (angle) และอัตราเร็วเชิงมุม (angular velocities) ใน sagittal plane และ ground reaction force ซึ่งมีเวกเตอร์ที่ใช้ 12 เวกเตอร์ และนำจากการหาค่าความแม่นยำของ K/L scores ซึ่งได้ค่าความแม่นยำเพิ่มขึ้น 2% เป็น 67% ความแม่นยำในบาง range มีค่าที่ต่ำ อาจเนื่องมาจากมีการจัดข้อมูลที่ไม่ครอบคลุมและมีความเกี่ยวข้องต่อกัน ดังนั้นการทดลองต่อมาจึงทำการจัดคลาสให้กับ K/L scores โดย K/L score 1, 2,3 เป็น low class (Class 1) และ K/L score 4,5 เป็น high class (Class2) ระเบียบที่ใช้เป็น 150 ระเบียบในการ train และ 150 ระเบียบในการ test แล้วนำไปทดสอบได้ผลดังนี้ เมื่อใช้ 21 เวกเตอร์เป็น input ได้ความแม่นยำเป็น 62% เมื่อใช้ตัวแปรใน sagittal plane และ ground reaction force เป็น input ได้ความแม่นยำเป็น 76% และการทดลองต่อมาเป็นการหาระดับของอาการเจ็บป่วย ซึ่ง pain score มีช่วงตั้งแต่ 0-403.5 จัดการแบ่งเป็น 2 คลาสได้ คือ ค่าคะแนนที่สูงกว่า 100 จะอยู่ใน low class (Class 1) และค่าคะแนนต่ำกว่า 100 จัดอยู่ใน high class (Class 2) ระเบียบที่ใช้เป็น 150

ระเบียนในการ train และ 150 ระเบียนในการ test แล้วนำไปทดสอบได้ผลดังนี้ เมื่อใช้เวกเตอร์ทั้งหมดเป็น input ได้ความแม่นยำเป็น 74.7% เมื่อใช้ตัวแปรใน sagittal plane และ ground reaction force เป็น input ได้ความแม่นยำเป็น 74.4%

R. Lafuente [4] ได้อธิบายถึงวิธีในการออกแบบและทดสอบระบบการวินิจฉัยการวิเคราะห์การเคลื่อนไหวและการตรวจสอบการใช้ประโยชน์ทางการแพทย์ของเครือข่ายประสาทและประสิทธิภาพในการวิเคราะห์ของกรณีศึกษาโดยใช้ข้อมูลการเดินที่มีผู้ป่วยเป็นโรคข้อเสื่อม 148 คน (arthritis) คนปกติ 88 คน (Control subjects) และได้ข้อมูลมาจาก force plate ได้พีเจอร์ทั้งหมด 78 พีเจอร์ มีการทำ Common transformed และ Diffential transformed กับพีเจอร์ทั้งหมด หลังจากนั้นได้ทำการเลือกจัดอันดับและเลือกพีเจอร์โดยให้ FDA algorithm และเลือกเพียงสิบอันดับแรกจากพีเจอร์ที่เลือกมา ต่อมาได้ทำการนำข้อมูลเหล่านี้มาเรียนรู้อยู่ใช้ multilayer perception (MLP) ด้วย 6 hidden neurons และ Bayes quadratics ซึ่งผลรับที่ได้มีประสิทธิภาพของเครือข่ายประสาทคิดเป็น 80%

แต่ในงานวิจัยนี้มีการเปลี่ยนแปลง (transform) พีเจอร์ที่ใช้และแบ่งประเภทโดยใช้โครงสร้างเครือข่ายประสาทซึ่งเป็นโมเดลทางคณิตศาสตร์ ทำให้ไม่สามารถให้เหตุผลที่สอดคล้องกับทฤษฎีทางการแพทย์ได้

N.S. Kostas [6] กล่าวถึงวิธีการแบ่งประเภทข้อมูลเดินโดยใช้การศึกษาเบื้องต้นเพื่อการออกแบบระบบกึ่งอัตโนมัติในการการวินิจฉัยโรค โดยมีผู้เข้าร่วมเป็นโรคข้อเสื่อม 110 คน (sick subjects : Knee OA) และคนปกติ 91 คน (normal subject) ในงานวิจัยนี้จะแบ่งวินิจฉัยออกเป็นสองกลุ่มคือ คนปกติ (healthy) และผู้ป่วย (patient) โดยใช้วิธี stereo metric ในการเก็บข้อมูลการเดิน ซึ่งจะใช้ charge coupled device (CCD) และ frame-grabber electronics ซึ่งจะส่งภาพดิจิทัลที่ถูกกล้องเก็บภาพขณะเคลื่อนที่ ซึ่งในงานวิจัยนี้จะเลือกใช้พีเจอร์ที่เกี่ยวกับข้อเข่ามาทำการทดลอง ได้แก่ knee flexion, knee flexion moment, knee valgus moment และ total knee power รวมถึงค่าความเร็วในการเดิน (walking velocity), single support, step length ได้ใช้ MLP ทำการเรียนรู้อพีเจอร์เหล่านั้น และแบ่งประเภทข้อมูลออกมา ซึ่งจากการทดลองได้ผลว่าสามารถทำการวินิจฉัยโรคแบ่งข้อมูลโดยใช้โครงสร้างเครือข่ายประสาทได้ แต่ระบบกึ่งอัตโนมัติในการ

วินิจฉัยโรคเป็นเพียงการวินิจฉัยโรคเพียงขั้นต้นเท่านั้น ยังไม่สามารถมาแทนแพทย์หรือผู้เชี่ยวชาญโดยสิ้นเชิง

ในแต่ละงานวิจัยข้างต้นใช้พีเจอร์ที่ต่างกันตามวิธีการเก็บข้อมูลที่ต่างกัน และจำแนกโรคที่เกี่ยวข้องของการเดินเพียงโรคเดียว จากงานวิจัยข้างต้นแสดงให้เห็นว่าเมื่อทำการเลือกพีเจอร์ที่เหมาะสมมาใช้ในการจำแนกข้อมูลแต่ละคลาสจะทำให้ค่าความแม่นยำเพิ่มขึ้น ดังนั้นในวิทยานิพนธ์นี้ จะนำการเลือกพีเจอร์ที่เหมาะสมมาใช้ในการจำแนกข้อมูล และหาสภาพแวดล้อมที่เหมาะสมกับข้อมูลที่มีอยู่ เพื่อให้มีความแม่นยำมากที่สุด และทำการจำแนกโรคที่เกี่ยวข้องของการเดินคือ โรคข้อเข่าเสื่อมและโรคพาร์กินสัน

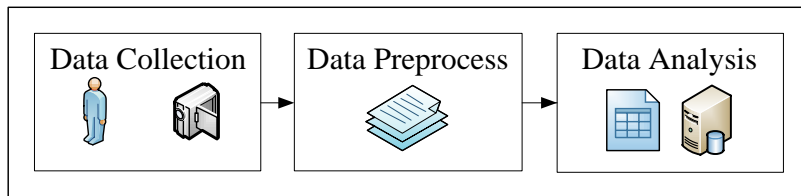
## บทที่ 3

### วิธีการดำเนินการวิจัย

ในบทนี้ จะกล่าวถึงขั้นตอนภาพรวมของงานวิจัยหลัก ๆ การวัดผลการทดลอง การเตรียมข้อมูลและรายละเอียดของข้อมูลที่จะนำไปใช้ในการทดลอง

#### ภาพรวมของงานวิจัย

ระบบจำลองการวิเคราะห์ข้อมูลการเดินประกอบไปด้วย 3 ส่วน ดังนี้ Data collection, Data preprocess, Data analysis ดังรูปที่ 3.1 ซึ่งมีขั้นตอนการทำงานคร่าว ๆ ดังนี้



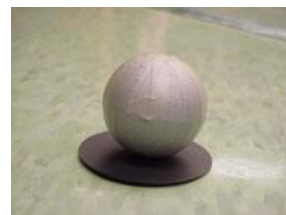
รูปที่ 3.1 ภาพรวมการทำงานของระบบ

#### 1. Data Collection

ส่วนแรก การเก็บข้อมูล (Data Collection) ข้อมูลการเดินในการวิจัยนี้ได้มาจากศูนย์ความเป็นเลิศทางการแพทย์ด้านการเดินและเคลื่อนไหว โรงพยาบาลจุฬาลงกรณ์ (Excellence center for Gait and Motion, King Chulalongkorn Memorial Hospital, Thailand) ซึ่งได้ข้อมูลจากชุดโปรแกรม Qualisys [19] เป็นระบบ Motion capture ในระบบนี้จะใช้กล้อง (รูปที่ 3.2ก) ทั้งหมด 6 ตัวในการบันทึกข้อมูล โดยผู้ใช้งานจะต้องติดมาร์คเกอร์สะท้อนแสง (reflective marker) ลักษณะตามรูปที่ 3.2ข ตามข้อต่าง ๆ ในตำแหน่งตามรูปที่ 3.3 ซึ่งจะมีวิธีติด 2 แบบ คือแบบครึ่งตัวและแบบเต็มตัว แบบครึ่งตัวจะใช้มาร์คเกอร์ทั้งหมด 26 ตำแหน่ง แบบเต็มตัวจะใช้มาร์คเกอร์ทั้งหมด 45 ตำแหน่ง สำหรับงานวิจัยนี้จะใช้การติดแบบครึ่งตัวที่ตำแหน่งช่วงล่างของร่างกาย ขณะที่ผู้ใช้งานกำลังเคลื่อนไหวกล้องทั้ง 6 ตัวจะตรวจจับที่ตำแหน่งมาร์คเกอร์เหล่านั้นและจัดเก็บข้อมูลตำแหน่งที่เปลี่ยนแปลงไป



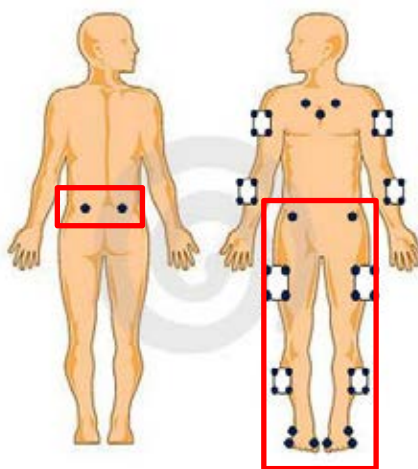
(ก)



(ข)

รูปที่ 3.2 (ก) กล้องสำหรับตรวจจับมาร์คเกอร์ (ข) มาร์คเกอร์สะท้อนแสง(reflective marker)

ค่าที่ได้จากโปรแกรมจะเป็นค่ามุมมองสามในแต่ละแกนทางการแพทย์ตามช่วงเวลาของการเดินทั้งหมด ซึ่งมีสามแกนคือ sagittal plane, coronal plane และ transverse plane (รูปที่ 2.4) ของแต่ละข้อต่อซึ่งมี 4 ตำแหน่งหลัก ๆ คือ ตำแหน่งกระดูกเชิงกราน, ข้อสะโพก, ข้อเข่า และข้อเท้า เมื่อผู้ใช้งานเดินครบรอบและโปรแกรมประมวลผลเบื้องต้นเสร็จแล้ว หลังจากนั้นเจ้าหน้าที่จะทำการระบุตำแหน่ง 4 จุดในวงจรการเดิน คือ Left Heel Strike (LHS), Left Toe Off (LTO), Right Heel Strike (Right Heel Strike), Right Toe Off (RTO) ในข้อมูลการเดิน โปรแกรมจะนำค่าเหล่านั้นไปประมวลผลลัพธ์ค่าอื่น ๆ จนสมบูรณ์ ซึ่งข้อมูลดิบที่นำไปใช้ในการทดลองที่ได้จากโปรแกรมคือ ค่ามุมมองสามตามช่วงเวลาในทั้งสามแกนทางการแพทย์ของทั้งสี่ข้อ และข้อมูลตำแหน่งสี่จุดในวงจรการเดิน



รูปที่ 3.3 ตำแหน่งที่ติดมาร์คเกอร์แบบ Walking marker แต่ในงานวิจัยนี้จะติดทั้งหมด 26

ตำแหน่งในกรอบสีแดง [20]



## 2. Data Preprocess

หลังจากที่ได้เก็บข้อมูลเรียบร้อยแล้ว ต่อไปเป็นส่วนที่สอง Data Preprocess เป็นขั้นตอนสำหรับการเตรียมข้อมูลจากข้อมูลดิบให้เป็นข้อมูลที่พร้อมสำหรับการนำไปวิเคราะห์ ประกอบไปด้วย 3 ขั้นตอนหลัก ๆ คือ

1. การแปลงข้อมูลการเดิน ข้อมูลดิบที่ได้มาจากการเก็บข้อมูลจะอยู่ในรูปค่ามุมมองศาในแต่ละเฟรมวิดีโอซึ่งเราจะต้องทำการแปลงข้อมูลให้อยู่ในรูปค่ามุมมองศาในแต่ละเปอร์เซ็นต์ของวงจรรจากเดิน โดยจะเลือกช่วงของข้อมูลเริ่มที่ตำแหน่งแรกของตำแหน่ง heel strike ไปจนถึงตำแหน่ง heel strike ตำแหน่งที่ 2 แล้วทำการปรับข้อมูลในช่วงที่เลือกมาโดยใช้วิธีการ interpolation ให้อยู่ในรูปเปอร์เซ็นต์ของวงจรรการเดิน หากข้อมูลดิบที่ได้มีตำแหน่ง heel strike มากกว่า 2 ขึ้นไปจะทำให้มีช่วงวงจรรการเดินมากกว่าหนึ่งช่วง ให้นำแต่ละช่วงผ่านการ interpolation ให้อยู่ในรูปของเปอร์เซ็นต์ของวงจรรการเดินแล้วนำช่วงทั้งหมดมาเฉลี่ยกัน จึงจะได้ข้อมูลที่สามารถนำไปวิเคราะห์ต่อได้

ข้อมูลของคนหนึ่งคนนั้นจะแปลงข้อมูลทั้งหมดในทุก ๆ ระยะเวลาของทุกข้อ ๆ ของขาทั้งสองข้าง ซึ่งจะแสดงผลข้อมูลออกมาในรูปของกราฟ ซึ่งจะมีทั้งหมด 24 กราฟ ดังรูปที่ 2.5 ในแต่ละกราฟแกนตั้งจะแสดงค่าของมุมมองศา และแกนนอนเป็นค่าเปอร์เซ็นต์ของวงจรรการเดิน ในกราฟหนึ่ง ๆ จะมีทั้งหมด 100 ข้อมูล ซึ่งทำให้ได้คนหนึ่งคนจะมีข้อมูลทั้งหมด 2,400 พีเจอร์ มาจากข้อต่อสี่ข้อ ขาทั้งสองข้าง ในสามระยะนาบ และกราฟละ 100 จุด ( $4 * 2 * 3 * 100 = 2,400$  พีเจอร์) ซึ่งเซตของข้อมูลเหล่านี้เรียกว่า Kinematic data

2. การคำนวณหาค่า temporal data เป็นข้อมูลเกี่ยวกับระยะทางและความเร็ว ซึ่งไม่ได้มาจาก Motion capture ในขั้นตอนการเก็บข้อมูลโดยตรง แต่สามารถคำนวณได้จากข้อมูลดิบและตำแหน่งสี่จุดในวงจรรการเดิน (LHS, LTO, RHS, RTO) ซึ่งยกตัวอย่างการหา temporal data ของขาข้างซ้ายดังนี้

Left Stance Time: ช่วงเวลาที่ส้นเท้าซ้ายสัมผัสพื้นจนถึงนิ้วเท้าซ้ายพ้นจากพื้น

$$\text{Left Stance Time} = (\text{LTO}_{\text{time}} - \text{LHS}_{\text{time}})$$

Left Swing Time: ช่วงเวลาที่เท้าซ้ายไม่ได้สัมผัสกับพื้น

$$\text{Left Swing Time} = (\text{LHS}_{\text{time}} - \text{LTO}_{\text{time}})$$

Left Cycle Time: ช่วงเวลาตั้งแต่เท้าซ้ายสัมผัสกับพื้นจนถึงเวลาที่ส้นเท้าซ้ายสัมผัสกับพื้นอีกครั้ง

$$\text{Left Cycle Time} = (\text{LHS}_{\text{position2}} - \text{LHS}_{\text{position1}})$$

Left Stride Length: ระยะทางตั้งแต่ที่ตำแหน่งของเท้าซ้ายสัมผัสกับพื้นจนถึงตำแหน่งที่ส้นเท้าซ้ายสัมผัสกับพื้นอีกครั้ง

$$\text{Left Stride Length} = (\text{LHS}_{\text{position2}} - \text{LHS}_{\text{position1}})$$

Left Step Time: ช่วงเวลาที่ส้นเท้าขวาสัมผัสกับพื้นจนถึงเวลาที่ส้นเท้าซ้ายสัมผัสกับพื้น

$$\text{Left Step Time} = (\text{LHS}_{\text{time}} - \text{RHS}_{\text{time}})$$

Left Step Length: ระยะทางของตำแหน่งที่ส้นเท้าขวาสัมผัสกับพื้นจนถึงตำแหน่งที่ส้นเท้าซ้ายสัมผัสกับพื้น

$$\text{Left Step Length} = (\text{LHS}_{\text{position}} - \text{RHS}_{\text{position}})$$

Left Steps Time Per Minute: จำนวนก้าวของขาซ้ายในหนึ่งนาที

$$\text{Left Steps Time Per Minute} = (60 / \text{Left Steps Time})$$

Left Strides Time Per Minute: จำนวน Left Cycle ต่อหนึ่งนาที

$$\text{Left Strides Time Per Minute} = (60 / \text{Left Cycle Time})$$

Speed: ความเร็วในการเดิน

$$\text{Speed} = \text{Distance} / \text{Time}$$

นอกจากนี้ ยังมี Base width คือความกว้างของช่วงก้าวระหว่างเท้าขวาและเท้าซ้าย, Double Support Time คือระยะเวลาที่เท้าทั้งสองสัมผัสกับพื้นพร้อมกัน Cycle time คือเวลาในหนึ่งรอบวงจรกิจกรรมเดิน รวมถึงยังนำค่าฟิเจอร์ระหว่างเท้าซ้ายและเท้าขวามาหาค่าในรูปแบบเปอร์เซ็นต์ เช่น เปอร์เซ็นต์ของ Left Stance Time =  $(\text{Left Stance Time} \times 100) / (\text{Left Stance Time} + \text{Right Stance Time})$  และค่าอัตราส่วนระหว่างขาทั้งสองข้าง A ratio of Left Stance Time =  $(\text{Left Stance Time}) / (\text{Right Stance Time})$  รวมมีฟิเจอร์ทั้งหมด 50 ฟิเจอร์ ตามตารางที่ 3.1 ซึ่งข้อมูลเหล่านี้เรียกว่า Temporal Data

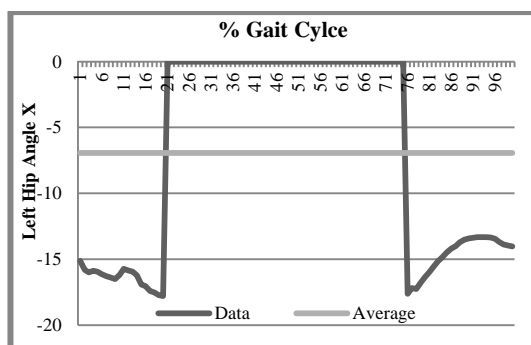
ตารางที่ 3.1 ตารางแสดงรายการฟิเจอร์ Temporal data ที่ใช้ในการทดลอง

Temporal Data Features	
Speed	Cycle Time
Base Width	Double Support Time
Left Step Length	Right Step Length
Left Step Time	Right Step Time
Left Stride Length	Right Stride Length
Left Cycle Time	Right Cycle Time
Left Stance Time	Right Stance Time
Left Swing Time	Right Swing Time
Left Steps Per Minute	Right Steps Per Minute
Left Strides Per Minute	Right Strides Per Minute
Percent of Left Stance Time	Percent of Right Stance Time
Percent of Left Swing Time	Percent of Right Swing Time
Percent of Left Stance Time in Cycle	Percent of Right Stance Time in Cycle
Percent of Left Swing Time in Cycle	Percent of Right Swing Time in Cycle
Left Stance:Swing Time	Right Stance:Swing Time
Left:Right Stance Time	Left:Right Swing Time
Percent of Left Step Time	Percent of Right Step Time
Percent of Left Step Length	Percent of Right Step Length
Left:Right Step Time	Left:Right Step Length

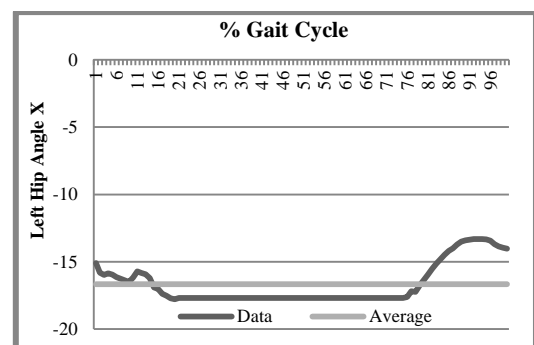
ตารางที่ 3.1 (ต่อ) ตารางแสดงรายการฟีเจอร์ Temporal data ที่ใช้ในการทดลอง

Temporal Data Features	
Percent of Left Stride Length	Percent of Right Stride Length
Percent of Left Steps Time Per Minute	Percent of Right Steps Time Per Minute
Left:Right Steps Time Per Minute	Left:Right Strides Time Per Minute
Percent of Left Strides Time Per Minute	Percent of Right Strides Time Per Minute

3. การแก้ไขข้อมูลที่ผิดพลาด ในขั้นตอนการทำงานของ Motion Capture อาจเกิดความผิดพลาดจากในขณะที่กำลังบันทึกข้อมูลมีช่วงเวลาหนึ่งที่กล้องไม่สามารถตรวจจับมาร์คเกอร์บางจุดได้ ทำให้ไม่สามารถคำนวณตำแหน่งและทำให้ข้อมูลในช่วงเวลานั้นมีค่าเป็นศูนย์ ในงานวิจัยนี้ได้มีการแก้ไขข้อมูลที่หายไปโดยใช้วิธีการเฉลี่ยค่า โดยนำค่าจากจุดสุดท้ายก่อนที่ข้อมูลจะหายไปและค่าแรกที่ปรากฏขึ้นหลังช่วงที่ข้อมูลหายไปนำมาเฉลี่ยกันและแทนผลเฉลี่ยในช่วงที่ข้อมูลหายไป อย่างเช่นรูปที่ 3.4ก ข้อมูลตั้งแต่ตำแหน่งที่ 20 ถึงตำแหน่งที่ 75 ขาดไป จึงทำการเฉลี่ยค่าที่ตำแหน่งที่ 19 และตำแหน่งที่ 76 และแทนผลเฉลี่ยลงในช่วงนั้นดังรูปที่ 3.4ข



(ก)



(ข)

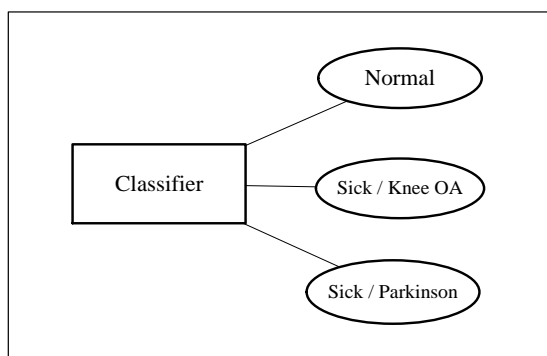
รูปที่ 3.4 (ก) กราฟแสดงข้อมูล Gait Cycle ที่มีค่าที่หายไป (ข) กราฟแสดงข้อมูล Gait Cycle ที่มีเติมค่าที่หายไปแล้ว

### 3. Data Analysis

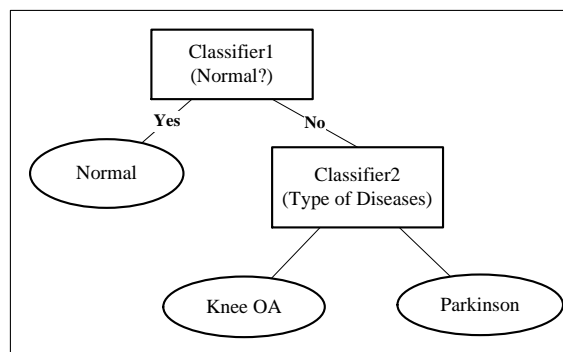
ส่วนที่สาม Data analysis สำหรับส่วนการวิเคราะห์ข้อมูล จะจำแนกข้อมูลออกเป็น 3 คลาส คือ Normal (คนปกติ), Sick/KneeOA (ผู้ป่วยข้อเข่าเสื่อม), Sick/Parkinson (ผู้ป่วยพาร์กินสัน)

สั้น) ในงานวิจัยนี้ได้ใช้เครื่องมือจำแนกประเภทสองตัว ได้แก่ อัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree) แบบ C4.5 ซึ่งเป็นอัลกอริทึมของต้นไม้ตัดสินใจพัฒนาขึ้นโดย Quinlan [16] โดยใช้ข้อมูล gain ratio ในการจำแนกประเภท และเครื่องมือที่สองคือ อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนส์ (Support Vector Machine – SVM) เป็นอัลกอริทึมที่สามารถช่วยแก้ปัญหาในการจำแนกข้อมูล โดยใช้หลักการของการหาสมประสิทธิผลของสมการ ในการสร้างสมการเชิงเส้นเพื่อใช้ในการแบ่งแยกกลุ่มข้อมูล ในงานวิจัยนี้ได้เลือกใช้ SVM แบบ Linear [21]

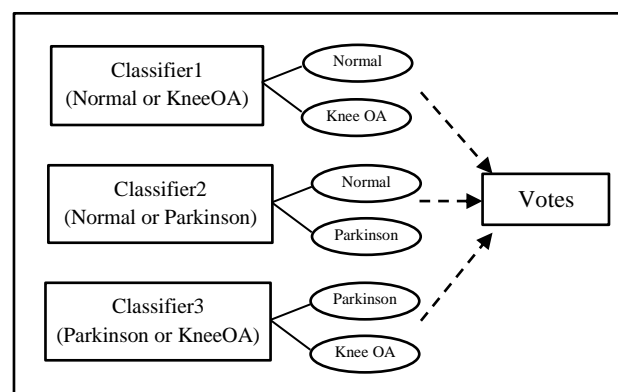
ในงานวิจัยนี้จะมีวิธีการแบ่งข้อมูลออกเป็น 3 โครงสร้างคือ Flat (รูปที่ 3.5) เป็นโครงสร้างแบบ single multiclass classifier คือจะจำแนกออกเป็น 3 คลาส ได้แก่ Normal , Sick/KneeOA , Sick/Parkinson โครงสร้างที่สองแบบ Hier (Hieratical) เป็นโครงสร้างแบบจำแนกข้อมูลออกเป็นสองระดับ (รูปที่ 3.6) โดยระดับแรกจะจำแนกผู้ป่วยออกเป็น Normal หรือ Sick ถ้าเป็นผู้ป่วยจะไปจำแนกในระดับที่สองว่าเป็น Parkinson หรือ KneeOA และโครงสร้างที่สามแบบ One VS One (Binary classifier approach) ดังรูปที่ 3.7 เป็นโครงสร้างที่มีแบบจำแนกข้อมูล 3 classifier โดยข้อมูลจะต้องผ่าน classifier ทั้ง 3 ตัว และนำผลลัพธ์ที่ได้มาทำการโหวต โดยจะยึดคำตอบที่มีค่าโหวตสูงสุด (ในกรณีที่ผลโหวตเท่ากัน จะพิจารณาค่าความแม่นยำของ Training data)



รูปที่ 3.5 โครงสร้างแบบ Single multiclass classification (Flat)



รูปที่ 3.6 โครงสร้างแบบ Two-level hierarchical classification (Hier)



รูปที่ 3.7 โครงสร้างแบบ Binary classifier approach (One VS One)

### การวัดผลการทดลอง

ในการวัดประสิทธิภาพของระบบจะพิจารณาจากค่า Recall, ค่า Precision และค่า F-measure

Precision (Pr) คือ ค่าเปอร์เซ็นต์ของ true positive ของค่าทั้งหมด

Recall (Re) คือ ค่าเปอร์เซ็นต์ของ true positive ของค่าที่สนใจ

$$Precision = \frac{TP}{TP + FP} \quad Recall = \frac{TP}{TP + FN}$$

เมื่อ TP = True positives,

FN = False Negatives,

FP = False Positives,

TN = True Negatives

F-measure คือ ค่าเฉลี่ยของ precision และ recall

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

สำหรับค่าการวัดประสิทธิภาพของ multi-class domain จะใช้ค่า *macro-averaging* ซึ่งเป็นค่าเฉลี่ยของแต่ละคลาส ตามตารางที่ 3.2

ตารางที่ 3.2 ตาราง *macro-averaging* ของ precision, recall ,  $F_1$  (สำหรับ multi-class domain)

	Precision	Recall	$F_1$
Macro	$Pr = \frac{\sum_j^k Pr_k}{k}$	$Re = \frac{\sum_j^k Re_k}{k}$	$F_1 = \frac{\sum_j^k F_{1k}}{k}$

โดยที่

$k$  คือ จำนวนข้อมูลทั้งหมด

$j$  คือ ค่าดัชนี มีค่าตั้งแต่ 1, ...,  $k$

### การเตรียมข้อมูลและเครื่องมือที่ใช้ในการทดลอง

สำหรับงานวิจัยนี้เราใช้ข้อมูลที่ได้มาจากศูนย์ความเป็นเลิศทางการแพทย์ด้านการเดินและเคลื่อนไหว โรงพยาบาลจุฬาลงกรณ์ โดยเป็นข้อมูลผู้สูงอายุที่มีอายุตั้งแต่ 60 ปีขึ้นไป คณะแพศ ซึ่งมีข้อมูลทั้งหมด 102 คน โดยจะแบ่งออกเป็น 3 กลุ่ม : ผู้สูงอายุปกติ (Normal) 40 คน, ผู้ป่วยข้อเข่าเสื่อม (ชนิดขั้นรุนแรง) 34 คน, ผู้ป่วยพาร์กินสัน (ชนิด mild) 28 คน ซึ่งได้ผ่านการวินิจฉัยและแบ่งกลุ่มโดยแพทย์ผู้เชี่ยวชาญ เนื่องจากในการหาเก็บข้อมูลและการคำนวณหา Kinematic data และ Temporal Data จะต้องใช้ตำแหน่งที่สำคัญทั้งสี่ตำแหน่ง ได้แก่ LHS, LTO, RHS และ RTO ให้ครบอย่างน้อยตำแหน่งละสองจุด เพื่อให้สามารถคำนวณหาข้อมูลได้ครบถ้วน ทำให้ในการเก็บข้อมูลต้องมีการคัดเลือกข้อมูลกลุ่มตัวอย่างที่มีข้อมูลครบตามเงื่อนไข ข้อมูลที่ใช้ในการทดลองคือข้อมูล Kinematic เป็นข้อมูลบอกมุมองศาของข้อต่อสี่ข้อตามแต่ละแกนระนาบทั้งสามระนาบของทั้งขาข้างซ้ายและขาข้างขวา ซึ่งต้องมีการแปลงให้อยู่ในรูปค่าเปอร์เซ็นต์ของวงจรรการเดิน และ

ข้อมูล Temporal เป็นข้อมูลทางด้านความเร็วและเวลา ได้มาจากการคำนวณของค่า kinematic และ สี่ตำแหน่งที่สำคัญของวงจรถ่ายการเดิน

ข้อมูลที่ใช้ในการทดลองมีสองแบบคือ แบบ Original เป็นข้อมูลต้นฉบับและแบบ ImputeMissing เป็นข้อมูลที่มีการเติมค่าที่หายไปแล้ว ฟีเจอร์ที่ใช้ในการทดลองจะมีทั้งหมด 3 เซต คือ Kinematic, KinematicWithTemporal (รวมฟีเจอร์ Kinematic และฟีเจอร์ Temporal ไว้ด้วยกัน) และ Temporal เครื่องมือจำแนกประเภทที่ใช้ในการทดลอง คือ SVM-linear และ ต้นไม้ตัดสินใจ (Decision Tree) และมีโครงสร้างในการจำแนกข้อมูลทั้งหมด 3 โครงสร้างคือแบบ Flat, แบบ Hier และแบบ One VS One ในแต่ละการทดลองจะวัดประสิทธิภาพโดยวัดจากค่าความแม่นยำ (accuracy) และค่า  $F_1$

ในการทดลองแรกได้เลือกกลุ่มตัวอย่างตามเงื่อนไขมา 54 คนเป็นผู้สูงอายุปกติ 24 คน ผู้ป่วยข้อเข่าเสื่อม 15 คนและผู้ป่วยพาร์กินสัน 15 คน จำนวนฟีเจอร์ที่ใช้มีฟีเจอร์ Kinematic ทั้งหมด 2,400 ฟีเจอร์, Temporal ทั้งหมด 30 ฟีเจอร์ และ KinematicWithTemporal มีทั้งหมด 2,430 ฟีเจอร์

ในการทดลองต่อมา มีการปรับข้อมูลจากคำแนะนำจากผู้เชี่ยวชาญ ทำให้การคัดเลือกกลุ่มตัวอย่างมีจำนวนเพิ่มขึ้นเป็นผู้สูงอายุปกติ 27 คน ผู้ป่วยข้อเข่าเสื่อม 34 คนและผู้ป่วยพาร์กินสัน 27 คน รวมทั้งหมด 88 คน ฟีเจอร์ที่ใช้มีฟีเจอร์ Kinematic ทั้งหมด 2,400 ฟีเจอร์, Temporal มีการคำนวณเพิ่มจึงมีจำนวนเพิ่มขึ้นเป็นทั้งหมด 50 ฟีเจอร์ และ KinematicWithTemporal มีทั้งหมด 2,450 ฟีเจอร์



## บทที่ 4

### การทดลองและผลการทดลอง

ในบทนี้จะกล่าวถึงขั้นตอนการทดลอง และผลการทดลองทั้งหมด ในการทดลองต้องการที่จะหาสภาพแวดล้อมที่ให้ค่าความแม่นยำมากที่สุด ซึ่งข้อมูลที่จะนำมาทดลองจะผ่านกระบวนการเตรียมข้อมูลมาก่อนแล้ว ในการทดลองแรกจะนำข้อมูลเบื้องต้นที่เก็บมาใช้ทดลอง หลังจากนั้นทำการแก้ไขและเก็บข้อมูลเพิ่มเติม มีการเปรียบเทียบข้อมูลต้นฉบับกับข้อมูลที่มีการเติมค่าที่หายไป รวมถึงนำวิธีการปรับช่วงข้อมูล การลดลดปริมาณฟีเจอร์โดยใช้วิธีวิเคราะห์องค์ประกอบและวิธีเลือกเฉพาะจุดที่สำคัญ มีการหาโครงสร้างของวิธีการจำแนกที่เหมาะสมอีกด้วย ซึ่งมีรายละเอียดตามขั้นตอนการทดลอง ดังต่อไปนี้

#### 1. การทดลองวัดประสิทธิภาพของระบบโดยใช้ข้อมูลโดยใช้ชุดข้อมูล 54 คน

อันดับแรก ข้อมูลที่จะนำมาใช้ในการทดลองนี้มีกลุ่มตัวอย่างที่มีอายุตั้งแต่ 60 ปีขึ้นไป คณะแพทย์ โดยคัดเลือกมาตามเงื่อนไข จำนวนทั้งหมด 54 คน ประกอบไปด้วย ผู้สูงวัยปกติ 24 คน ผู้ป่วยข้อเข่าเสื่อม 15 คน และ ผู้ป่วยพาร์กินสัน 15 คน ในการทดลองจะมีชุดข้อมูล 2 ชุด โดยข้อมูลชุดแรกจะแทนที่ค่าข้อมูลที่หายไป (Missing Value) ด้วยค่าศูนย์ และข้อมูลชุดที่สองจะทำการเติมค่าข้อมูลที่หายไป (Missing Value) ด้วยค่าเฉลี่ยจุดสองจากจุดต้นและจุดปลายของช่วงที่จะมีการเกิดค่าข้อมูลที่หายไป (Missing Values) ซึ่งการเติมค่าที่หายไปจะมีผลต่อฟีเจอร์ Kinematic ไม่มีผลต่อฟีเจอร์ Temporal

ในการทดลอง จะทำการทดลองโดยใช้เครื่องมือมาตรฐานคือ SVM-Linear [18],[22] และ ต้นไม้ตัดสินใจ (Decision Tree) [16],[24] ในการเปรียบเทียบประสิทธิภาพจะใช้วิธีการแบ่งข้อมูลด้วยโครงสร้างแบบ Flat (รูปที่ 3.4) เป็นโครงสร้างแบบ single multiclass classifier คือจะจำแนกออกเป็น 3 คลาส ได้แก่ Normal , Sick/KneeOA , Sick/Parkinson บน 3-fold cross validation และมีฟีเจอร์ที่ใช้ในการทดสอบทั้งหมดสามชุดคือ Kinematic, Kinematic with temporal และ Temporal ซึ่งเมื่อทดลองโดยใช้ข้อมูลชุดแรก (แทนค่าข้อมูลที่หายไปด้วยศูนย์) มีผลการทดลองดังตารางที่ 4.1 และทดลองด้วยข้อมูลชุดที่สอง (แทนค่าข้อมูลที่หายไปด้วยค่าเฉลี่ยสองจุด) มีผลการทดลองดังตารางที่ 4.2

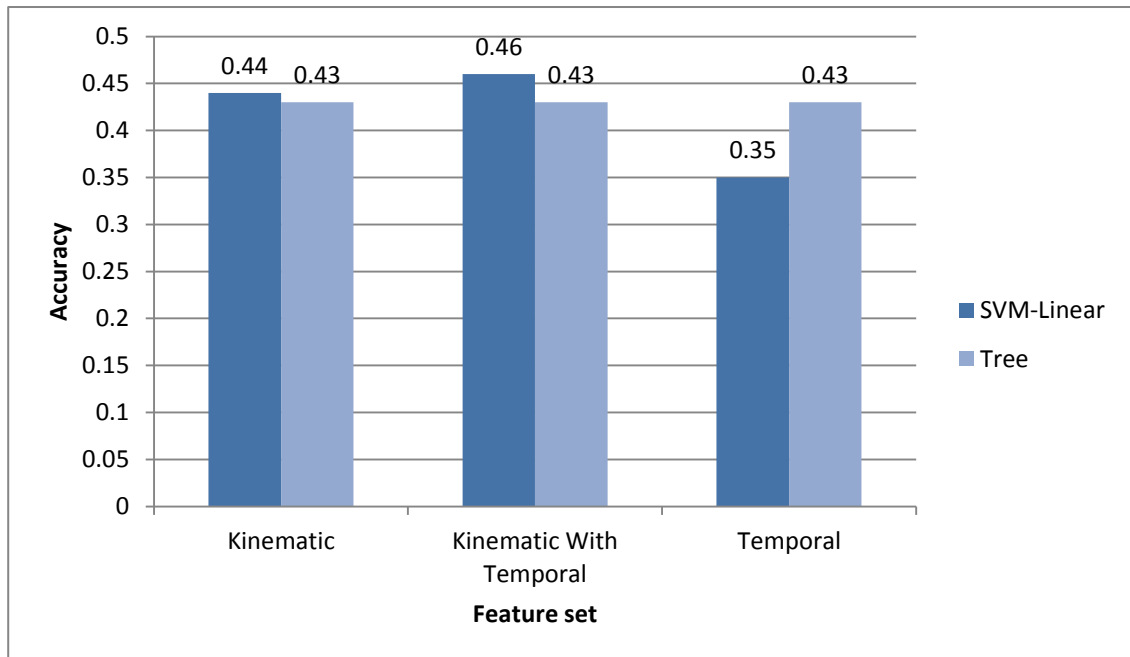
จากรูปที่ 4.1 และ 4.2 เป็นกราฟแสดงผลลัพธ์จากตารางที่ 4.1 จะเห็นได้ว่าพีเจอร์ Kinematic และ Kinematic with temporal เครื่องมือ SVM-Linear จะให้ค่าประสิทธิภาพที่สูงกว่า ต้นไม้ตัดสินใจ สำหรับพีเจอร์ Temporal เครื่องมือต้นไม้ตัดสินใจจะให้ประสิทธิภาพที่สูงกว่า SVM-Linear เช่นเดียวกับรูปที่ 4.3 และ 4.4 เป็นกราฟแสดงผลลัพธ์ตารางที่ 4.2 ซึ่ง Kinematic และ Kinematic with temporal เครื่องมือ SVM-Linear จะให้ค่าประสิทธิภาพที่สูงกว่าต้นไม้ตัดสินใจ แต่ค่าประสิทธิภาพยังแตกต่างกันไม่มากนัก ซึ่งในการทดลองนี้ทำให้ยังไม่สามารถสรุปได้ว่าเครื่องมือใดระหว่าง SVM และต้นไม้ตัดสินใจให้ประสิทธิภาพที่เหมาะสมกว่ากัน ในรูปที่ 4.5 เป็นกราฟการเปรียบเทียบระหว่างข้อมูลต้นฉบับและข้อมูลที่มีการเติมส่วนที่หายไป ซึ่งมีค่าความแม่นยำต่างกันเพียงเล็กน้อย ผลการทดลองที่ได้ยังไม่เห็นชัดเจนมากนักจึงยังไม่สามารถสรุปว่าการเติมค่าที่หายไปนั้นส่งผลต่อระบบหรือไม่

จากผลลัพธ์ตารางที่ 4.1 และ 4.2 จะเห็นได้ว่าสภาพแวดล้อมที่ให้ค่าประสิทธิภาพด้านความแม่นยำมากที่สุดคือ พีเจอร์ Kinematic with temporal เป็น 46% โดยใช้ SVM-Linear เป็นเครื่องมือในการจำแนกข้อมูล และใช้โครงสร้างแบบ Flat

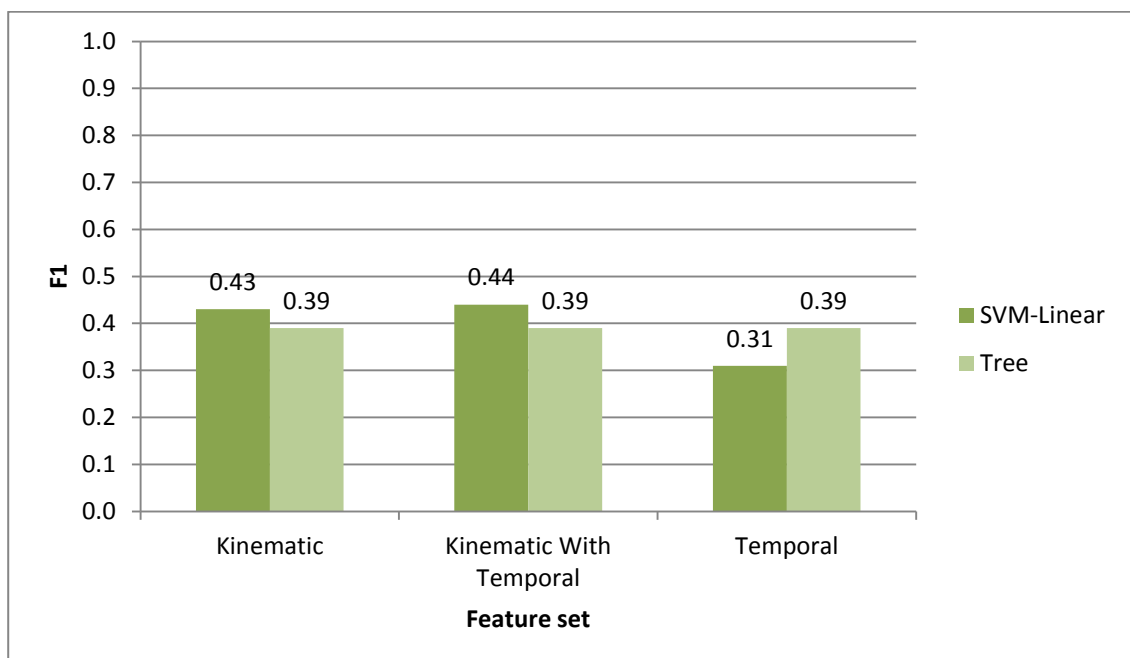
ตารางที่ 4.1 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้ข้อมูลต้นฉบับ (มีข้อมูลที่ขาดหาย)

Data	Feature set*	Classifier	Accuracy	F1	Precision	Recall
Original	Kinematic (2,400)	svm-linear	0.44±0.24	0.43±0.21	0.45±0.26	0.44±0.24
		Tree	0.43±0.32	0.39±0.29	0.37±0.27	0.43±0.32
	Kinematic With Temporal (2,430)	svm-linear	<b>0.46±0.25</b>	<b>0.44±0.21</b>	0.47±0.25	0.46±0.25
		Tree	0.43±0.32	0.39±0.29	0.37±0.27	0.43±0.32
	Temporal (30)	svm-linear	0.35±0.31	0.31±0.23	0.31±0.15	0.35±0.31
		Tree	0.43±0.20	0.39±0.17	0.38±0.14	0.43±0.20

\* ค่าในวงเล็บคือจำนวนพีเจอร์ในแต่ละเซตพีเจอร์



รูปที่ 4.1 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพของเครื่องมือจำแนกระหว่าง SVM-Linear และต้นไม้ตัดสินใจโดยใช้ชุดข้อมูลต้นฉบับ

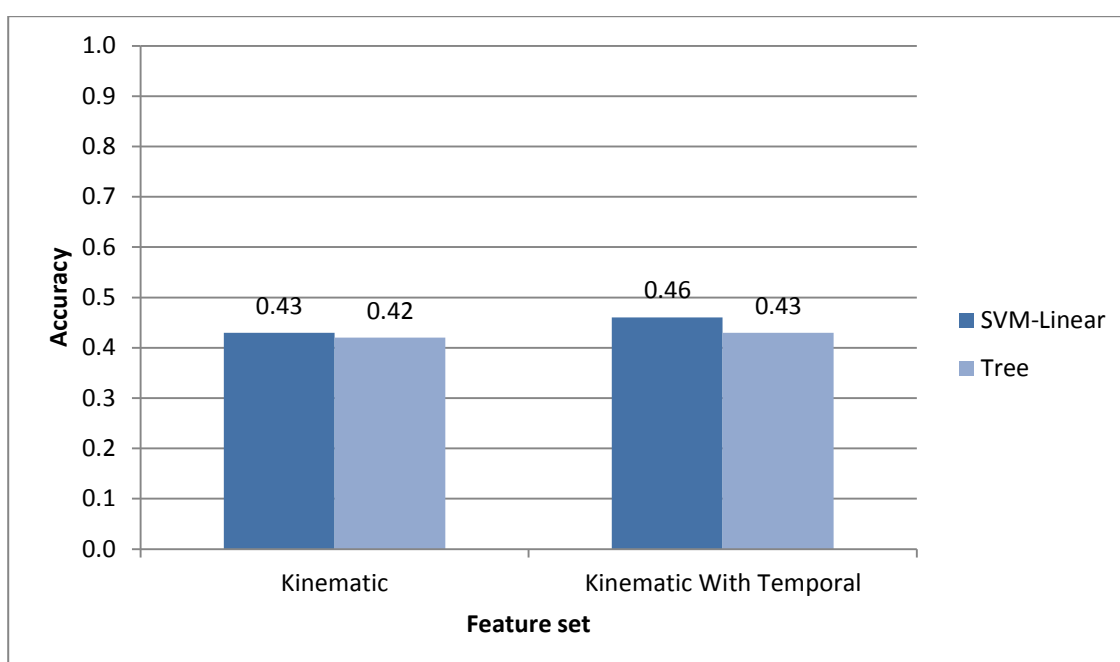


รูปที่ 4.2 กราฟแสดงค่า  $F_1$  เปรียบเทียบประสิทธิภาพของเครื่องมือจำแนกระหว่าง SVM-Linear และต้นไม้ตัดสินใจโดยใช้ชุดข้อมูลต้นฉบับ

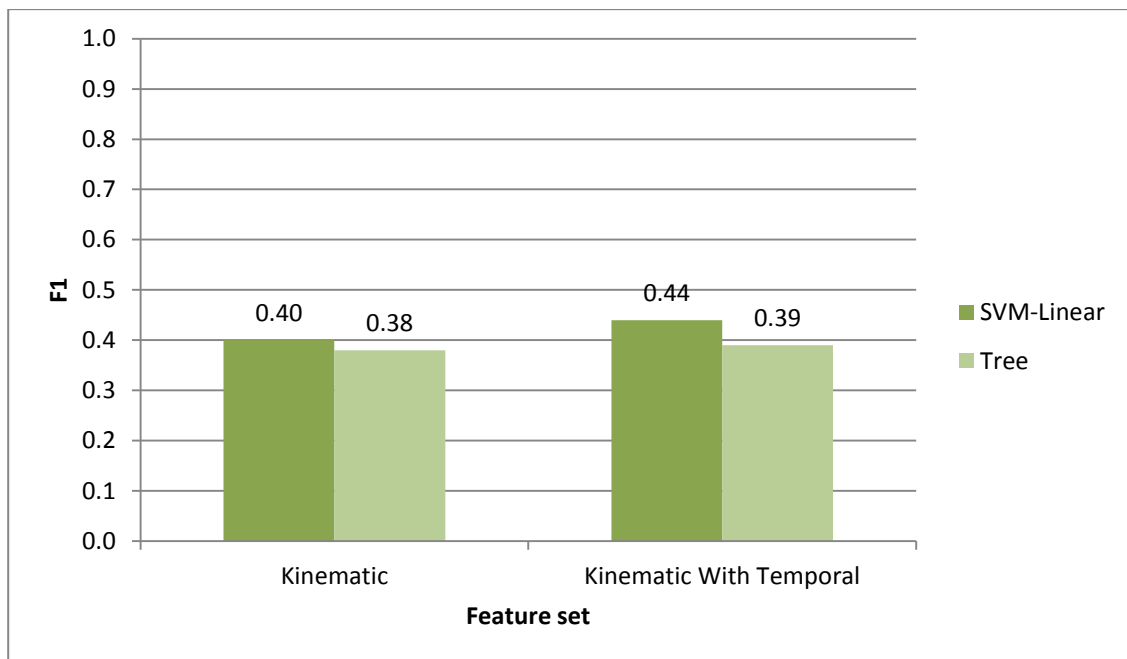
ตารางที่ 4.2 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้ข้อมูลที่มีการเติมค่าที่หายไป

Data	Feature set*	Classifier	Accuracy	F1	Precision	Recall
Original imputeMissing	Kinematic (2,400)	svm-linear	0.43±0.30	0.40±0.24	0.41±0.21	0.43±0.30
		Tree	0.42±0.30	0.38±0.27	0.37±0.26	0.42±0.30
	Kinematic WithTemporal (2,430)	svm-linear	<b>0.46±0.27</b>	<b>0.44±0.23</b>	0.44±0.21	0.46±0.27
		Tree	0.43±0.32	0.39±0.29	0.37±0.27	0.43±0.32

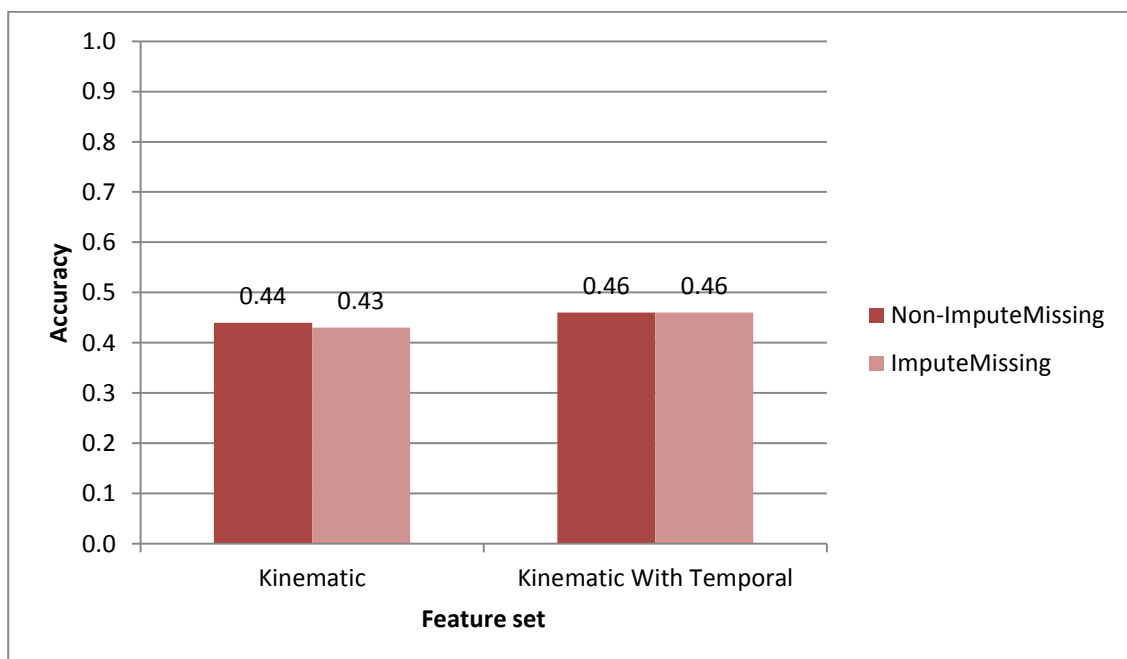
\* ค่าในวงเล็บคือจำนวนฟีเจอร์ในแต่ละเซตฟีเจอร์



รูปที่ 4.3 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพของเครื่องมือจำแนกระหว่าง SVM-Linear และต้นไม้ตัดสินใจโดยใช้ชุดข้อมูลที่มีการเติมค่าที่หายไป



รูปที่ 4.4 กราฟแสดงค่า  $F_1$  เปรียบเทียบประสิทธิภาพของเครื่องมือจำแนกระหว่าง SVM-Linear และต้นไม้ตัดสินใจโดยใช้ชุดข้อมูลที่มีการเติมค่าที่หายไป



รูปที่ 4.5 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลต้นฉบับกับชุดข้อมูลที่มีการเติมค่าที่หายไป โดยใช้เครื่องมือ SVM-Linear

ค่าความแม่นยำในการทดลองนี้จะเห็นได้ว่ายังต่ำอยู่ซึ่งมีเพียง 46% อาจเนื่องมาจากจำนวนตัวอย่างที่นำมาใช้นั้นยังมีปริมาณน้อย สำหรับการทดลองต่อไปจะทำการเก็บจำนวนตัวอย่างเพิ่มเติมและปรับปรุงข้อมูลของกลุ่มตัวอย่าง หาสภาพแวดล้อมที่เหมาะสมเพื่อทดสอบว่า จะทำให้มีประสิทธิภาพด้านความแม่นยำมากยิ่งขึ้นหรือไม่

## 2. การทดลองวัดประสิทธิภาพของระบบโดยใช้ปริมาณจำนวนตัวอย่างเพิ่มขึ้น

การทดลองนี้ จะทำการเก็บข้อมูลเพื่อเพิ่มจำนวนตัวอย่างในการทดลองมากขึ้น โดยขอคำแนะนำจากเจ้าหน้าที่เรื่องบางข้อมูลที่มีปัญหาและทำการปรับแก้ไข ซึ่งทำให้ในการทดลองนี้ได้มีการใช้จำนวนในกลุ่มตัวอย่างมากขึ้น โดยจะใช้จำนวนตัวอย่างทั้งหมด 88 คน เป็นผู้สูงอายุปกติ จำนวน 27 คน ผู้ป่วยข้อเข่าเสื่อม 34 คน ผู้ป่วยพาร์กินสันจำนวน 27 คน ซึ่งเงื่อนไขในการเลือกตัวอย่างยังคงเหมือนกับการทดลองแรก และมีเงื่อนไขเพิ่มเติมคือ ถ้าตัวอย่างใดมีข้อมูลที่ขาดหายไปเกิน 50 % จะไม่นำมาใช้

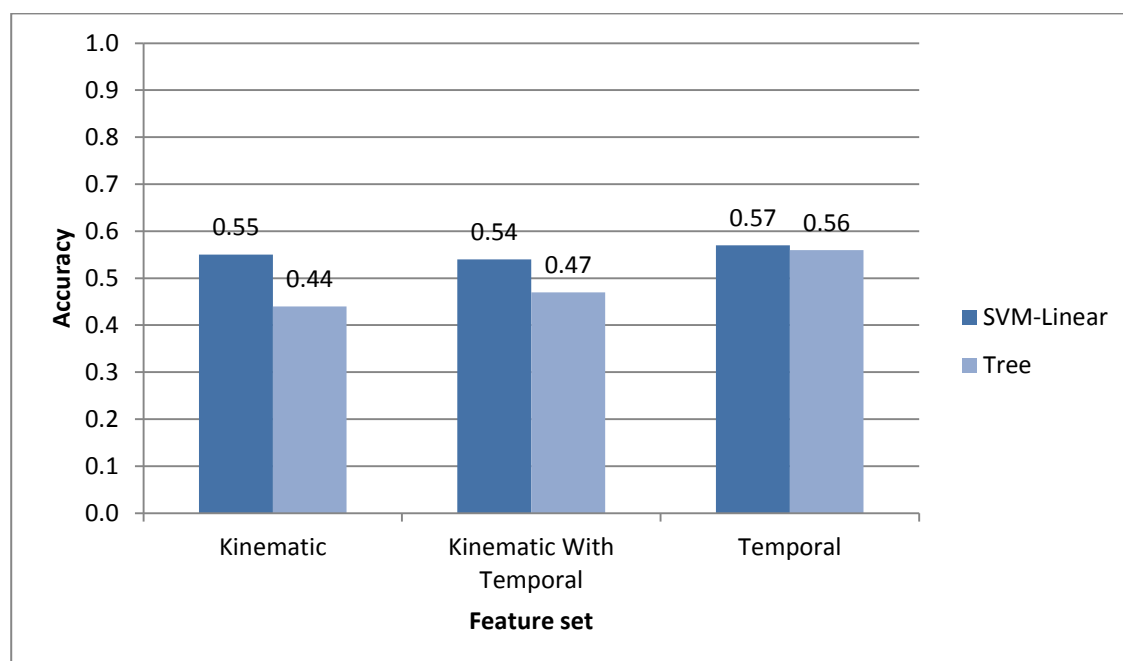
ในการทดลองจะทำการทดลองโดยใช้เครื่องมือมาตรฐานคือ SVM-Linear และต้นไม้ตัดสินใจ ในการเปรียบเทียบประสิทธิภาพใช้วิธีการแบ่งข้อมูลด้วยโครงสร้างแบบ Flat บน 3-fold cross validation โดยจะมีฟีเจอร์ทั้งหมดสามชุดคือ Kinematic, Kinematic with temporal และ Temporal ข้อมูลที่นำมาทดลองจะมีสองชุดคือชุดต้นฉบับและชุดข้อมูลที่ผ่านการเติมส่วนที่หายไป ซึ่งมีผลการทดลองเป็นไปตามตารางที่ 4.3 และจากผลลัพธ์ตารางที่ 4.3 จะเห็นได้ว่าสภาพแวดล้อมที่ให้ค่าประสิทธิภาพด้านความแม่นยำมากที่สุดคือ ฟีเจอร์ Temporal เป็น 57% โดยใช้ SVM-Linear เป็นเครื่องมือในการจำแนกข้อมูล และใช้โครงสร้างแบบ Flat

จากรูปที่ 4.6, 4.7, 4.8 และ 4.9 เป็นกราฟเปรียบเทียบผลจากตารางที่ 4.3 จะเห็นได้ว่าในชุดข้อมูลกลุ่มตัวอย่าง 88 คน ทั้งฟีเจอร์ Kinematic, Kinematic with temporal และ Temporal เครื่องมือ SVM-Linear จะให้ค่าประสิทธิภาพทั้งค่าความแม่นยำและค่า  $F_1$  ที่สูงกว่าต้นไม้ตัดสินใจในทุก ๆ เซตฟีเจอร์

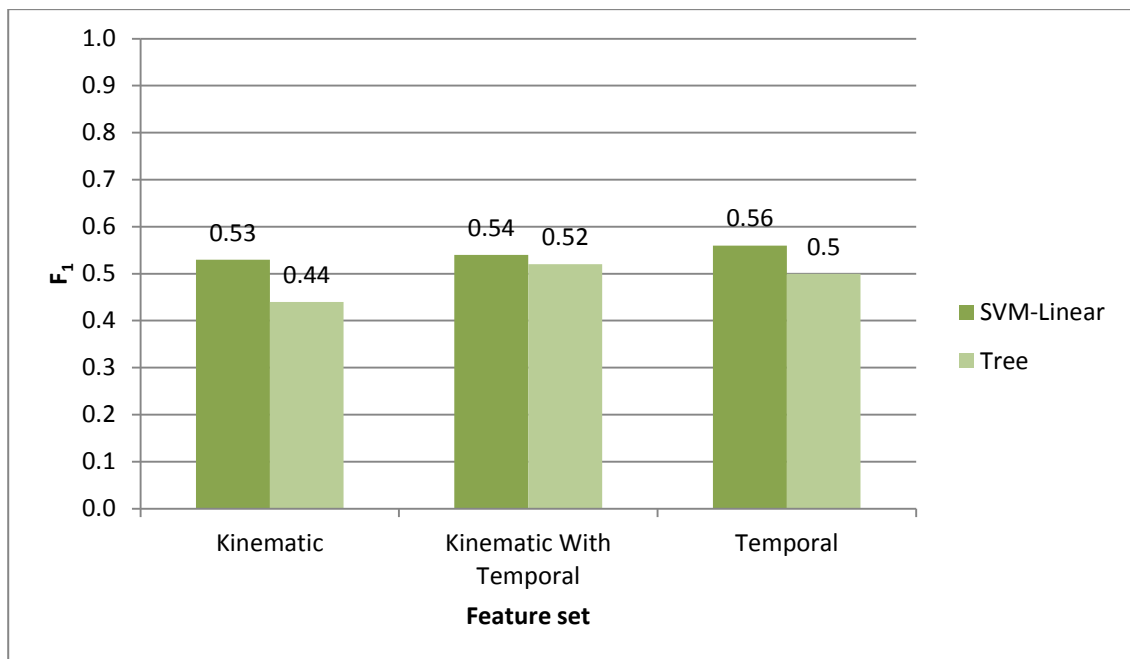
ตารางที่ 4.3 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้ข้อมูลที่มีจำนวนตัวอย่างเพิ่มขึ้น

Data	Feature set*	Classifier	Accuracy	F1	Precision	Recall
Original	Kinematic (2,400)	svm-linear	0.55±0.13	0.53±0.13	0.55±0.12	0.55±0.13
		Tree	0.44±0.12	0.44±0.11	0.45±0.10	0.44±0.12
	Kinematic WithTemporal (2,450)	svm-linear	0.54±0.13	0.52±0.14	0.54±0.12	0.54±0.13
		Tree	0.47±0.13	0.46±0.13	0.46±0.13	0.47±0.13
	Temporal (50)	svm-linear	<b>0.57±0.19</b>	<b>0.56±0.23</b>	0.59±0.07	0.57±0.19
		Tree	0.51±0.12	0.50±0.11	0.51±0.10	0.51±0.12
Original imputeMissing	Kinematic (2,400)	svm-linear	<b>0.54±0.07</b>	<b>0.53±0.10</b>	0.55±0.11	0.54±0.07
		Tree	0.49±0.12	0.48±0.12	0.48±0.12	0.49±0.12
	Kinematic WithTemporal (2,450)	svm-linear	0.52±0.10	0.51±0.10	0.54±0.08	0.52±0.10
		Tree	0.48±0.10	0.47±0.10	0.47±0.10	0.48±0.10

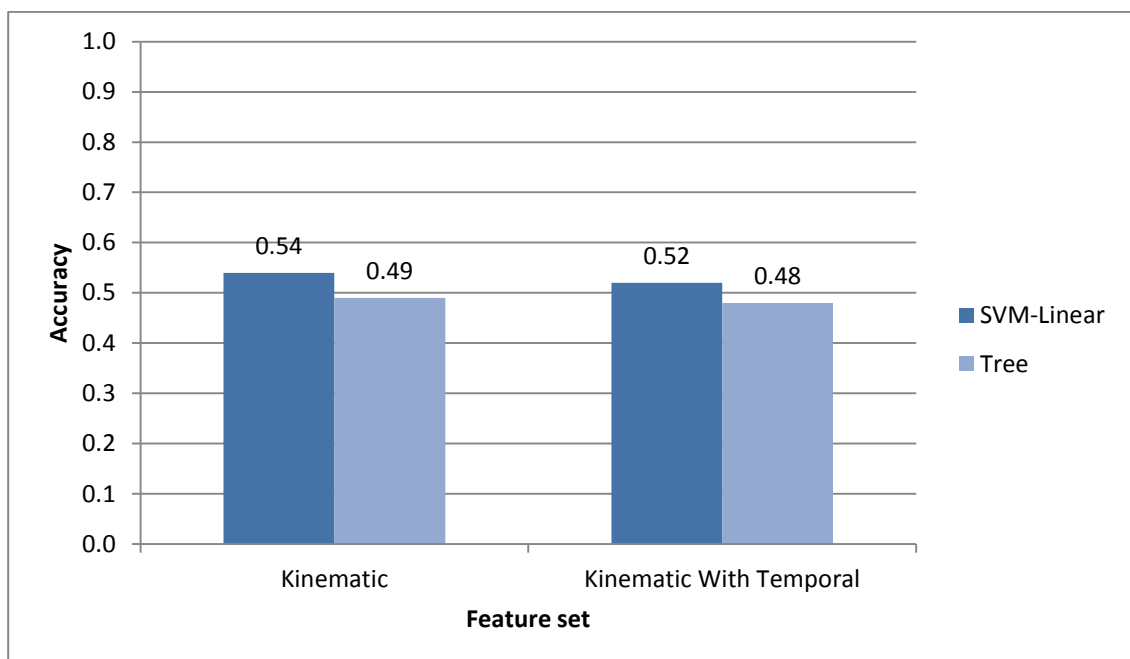
\* ค่าในวงเล็บคือจำนวนฟีเจอร์ในแต่ละเซตฟีเจอร์



รูปที่ 4.6 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพของเครื่องมือจำแนกระหว่าง SVM-Linear และต้นไม้ตัดสินใจโดยใช้ชุดข้อมูลต้นฉบับกลุ่มตัวอย่าง 88 คน

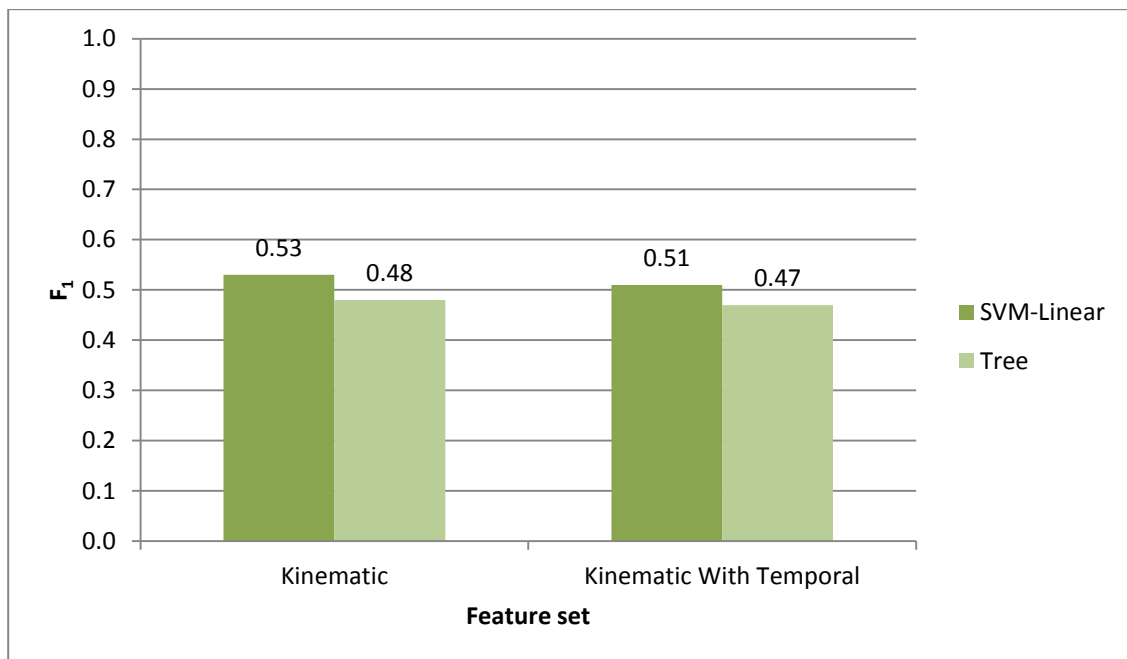


รูปที่ 4.7 กราฟแสดงค่า  $F_1$  เปรียบเทียบประสิทธิภาพของเครื่องมือจำแนกระหว่าง SVM-Linear และต้นไม้ตัดสินใจโดยใช้ชุดข้อมูลต้นฉบับกลุ่มตัวอย่าง 88 คน



รูปที่ 4.8 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพของเครื่องมือจำแนกระหว่าง SVM-Linear และต้นไม้ตัดสินใจโดยใช้ชุดข้อมูลกลุ่มตัวอย่าง 88 คนมีการเติมค่าที่หายไป

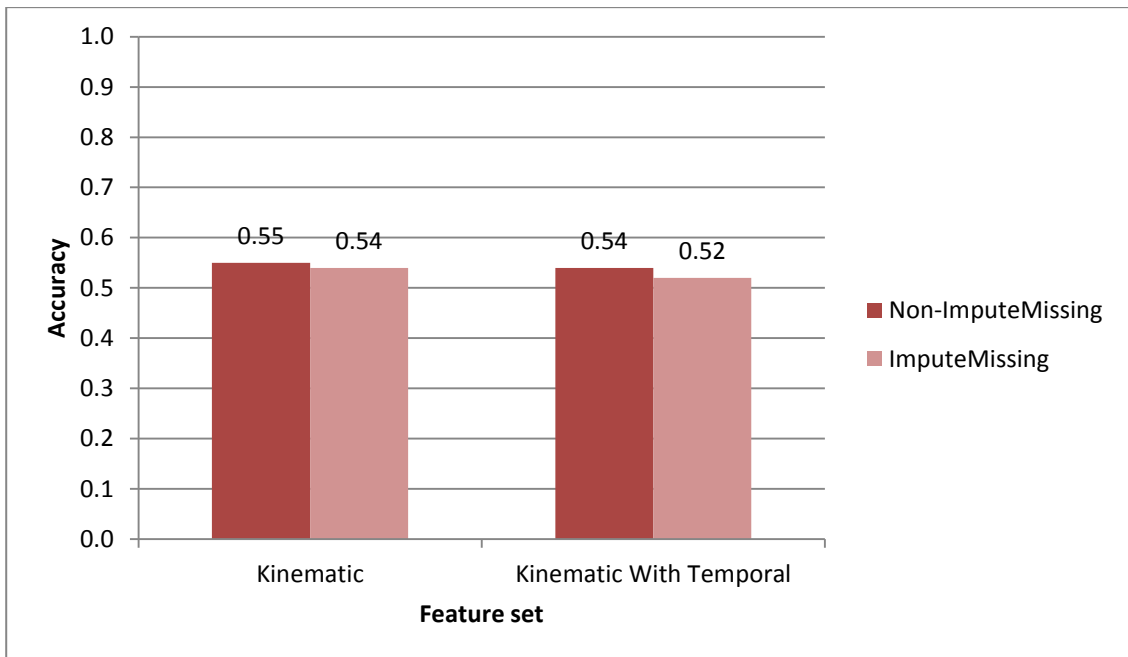




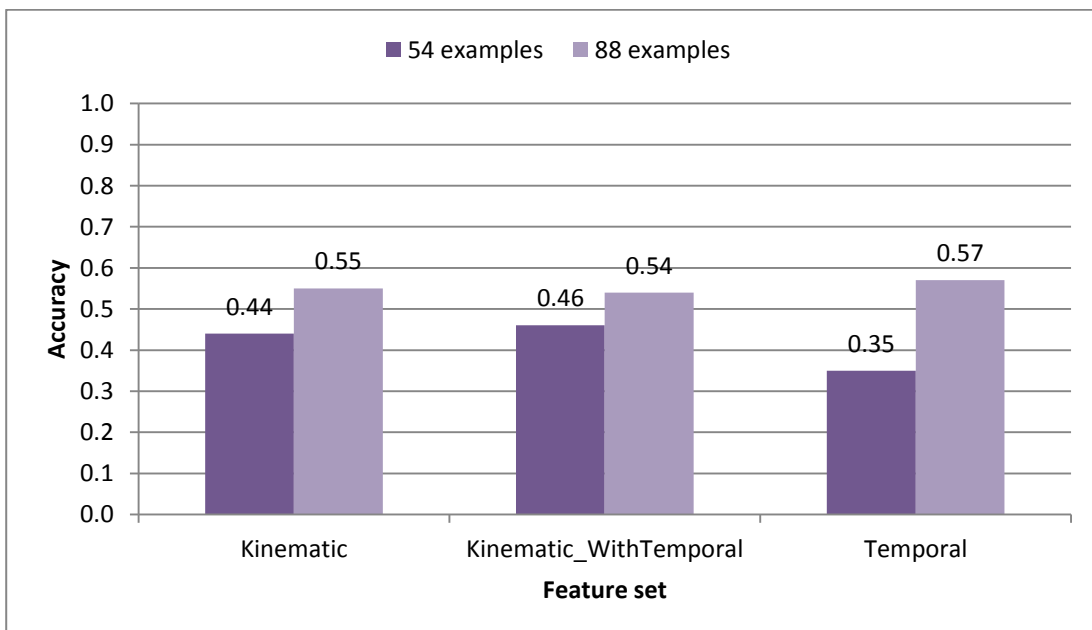
รูปที่ 4.9 กราฟแสดงค่า  $F_1$  เปรียบเทียบประสิทธิภาพของเครื่องมือจำแนกระหว่าง SVM-Linear และต้นไม้ตัดสินใจโดยใช้ชุดข้อมูลกลุ่มตัวอย่าง 88 คนมีการเติมค่าที่หายไป

รูปที่ 4.10 เป็นผลเปรียบเทียบจากตารางที่ 4.3 เป็นกราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลกลุ่มต้นฉบับกับชุดข้อมูลมีการเติมค่าที่หายไป จะเห็นได้ว่าชุดข้อมูลต้นฉบับนั้นมีค่าประสิทธิภาพที่สูงกว่าชุดข้อมูลมีการเติมค่าที่หายไป อาจจะเป็นเนื่องจากอัลกอริทึมที่ใช้ยังไม่สอดคล้องกับข้อมูลจริงและค่าที่เดานั้นเป็นค่าเฉลี่ยและค่าจะอยู่ในช่วงที่ไม่แตกต่างกับค่าศูนย์มากนัก ทำให้การเติมข้อมูลด้วยอัลกอริทึมนี้จึงไม่ส่งผลต่อค่าความแม่นยำเท่าที่ควร

จากรูปที่ 4.11 เป็นกราฟแสดงค่าความแม่นยำเปรียบเทียบระหว่างชุดข้อมูลต้นฉบับของจำนวนตัวอย่าง 54 คน กับชุดข้อมูลต้นฉบับของจำนวนตัวอย่าง 88 คน ซึ่งมาจากตารางที่ 4.1 และ 4.3 จะเห็นได้ว่าเมื่อจำนวนตัวอย่างเพิ่มขึ้นค่าความแม่นยำมีค่าสูงขึ้น เพราะฉะนั้นจึงจะใช้จำนวนตัวอย่าง 88 คนเป็นข้อมูลในการทดลองลำดับถัดไป



รูปที่ 4.10 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลกลุ่มตัวอย่าง 88 คนต้นฉบับกับชุดข้อมูลกลุ่มตัวอย่าง 88 คนที่มีการเติมค่าที่หายไป โดยใช้เครื่องมือ SVM-Linear



รูปที่ 4.11 กราฟแสดงค่าความแม่นยำเปรียบเทียบระหว่างชุดข้อมูลต้นฉบับของจำนวนตัวอย่าง 54 คน กับชุดข้อมูลต้นฉบับของจำนวนตัวอย่าง 88 คน

จากการทดลองนี้ สภาพแวดล้อมที่ให้ค่าประสิทธิภาพด้านความแม่นยำมากที่สุดคือ ฟีเจอร์ Temporal เป็น 57% โดยใช้ SVM-Linear เป็นเครื่องมือในการจำแนกข้อมูล และใช้ โครงสร้างแบบ Flat ซึ่งมีค่าความแม่นยำเพิ่มขึ้นจากการทดลองที่ผ่านมา ทำให้สรุปได้ว่าเมื่อ จำนวนตัวอย่างเพิ่มขึ้นส่งผลให้ระบบมีประสิทธิภาพมากยิ่งขึ้น และในผลการทดลองจะเห็นได้ว่า ในทุก ๆ เซตฟีเจอร์เครื่องมือ SVM-Linear จะให้ค่าประสิทธิภาพทั้งค่าความแม่นยำและค่า  $F_1$  ที่สูงกว่าต้นไม่ตัดสินใจ ในการทดลองถัดไปจึงควรเลือกใช้ SVM-Linear ในการทดลอง และจะเห็นได้ว่า ชุดข้อมูลต้นฉบับนั้นให้ค่าประสิทธิภาพที่สูงกว่าชุดข้อมูลที่มีการเติมค่าที่หายไป การเติมค่าข้อมูลที่หายไปไม่ส่งผลต่อค่าความแม่นยำไม่มากนัก เพราะฉะนั้นในการทดลองลำดับถัดไปจะใช้ชุดข้อมูล จำนวนตัวอย่าง 88 คนที่เป็นข้อมูลต้นฉบับที่ไม่ผ่านการเติมข้อมูลที่หายไป จะเลือกใช้เครื่องมือ SVM-Linear ในการจำแนกข้อมูล

ในการทดลองนี้ค่าความแม่นยำนั้นยังต่ำอยู่เพียง 57% ซึ่งมีข้อสันนิษฐานสามกรณีคือ กรณีแรกเกิดจากปริมาณจำนวนตัวอย่างกับปริมาณของจำนวนฟีเจอร์มีปริมาณที่แตกต่างกันมากเกินไป กรณีที่สองในลักษณะหรือหน่วยของข้อมูลในบางฟีเจอร์มีช่วงข้อมูลที่ต่างกัน และกรณีที่สามอาจจะเป็นเพราะทั้งกรณีที่หนึ่งและกรณีที่สองร่วมกัน จึงจะนำวิธีการลดตัวแปรโดยใช้วิธีการวิเคราะห์องค์ประกอบ (Principle Component Analysis) เพื่อลดปริมาณของจำนวนฟีเจอร์ และวิธีการปรับช่วงข้อมูล (Data Normalization) เพื่อลดความแตกต่างของช่วงข้อมูล ในการทดลองถัดไป

### 3. การทดลองเพื่อเพิ่มประสิทธิภาพโดยใช้วิธีการวิเคราะห์องค์ประกอบ (Principle Component Analysis) และวิธีการปรับช่วงข้อมูล (Data Normalization)

ในการทดลองจะทดลองเพื่อหาวิธีเพิ่มประสิทธิภาพ โดยจะมีสามการทดลองย่อย การทดลองแรกจะนำวิธีการวิเคราะห์องค์ประกอบ (PCA) มาใช้ การทดลองที่สองจะนำวิธีการปรับช่วงข้อมูล (Data Normalization) และการทดลองสุดท้ายจะนำทั้งวิธีการวิเคราะห์องค์ประกอบ (PCA) และวิธีการปรับช่วงข้อมูล (Data Normalization) มาใช้ร่วมกัน

#### 3.1. การทดลองเพื่อเพิ่มประสิทธิภาพโดยวิธีการลดตัวแปรโดยใช้วิธีการวิเคราะห์องค์ประกอบ (Principle Component Analysis)

การทดลองนี้จะใช้จำนวนตัวอย่างทั้งหมด 88 คน ซึ่งข้อมูลก็นำมาใช้เป็นชุดข้อมูลต้นฉบับที่จะผ่านการปรับลดตัวแปรโดยใช้วิธีวิเคราะห์องค์ประกอบ (PCA) โดยมีค่า variance อยู่ที่ 95% ในการทดลองจะทำการทดลองโดยใช้เครื่องมือมาตรฐานคือ SVM-Linear ในการเปรียบเทียบประสิทธิภาพ ใช้วิธีการแบ่งข้อมูลด้วยโครงสร้างแบบ Flat โดยจะมีฟีเจอร์ทั้งหมดสามชุดคือ Kinematic, Kinematic with temporal และ Temporal โดยมีผลลัพธ์ดังตารางที่ 4.4

ตารางที่ 4.4 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้ PCA ในการลดตัวแปร

Data	Feature set*	Classifier	Accuracy	F1	Precision	Recall
Original + PCA	Kinematic (35-33-30)	svm-linear	0.51±0.08	0.50±0.10	0.52±0.10	0.51±0.08
	<b>Kinematic With Temporal (36-34-31)</b>	svm-linear	<b>0.56±0.12</b>	<b>0.54±0.12</b>	0.56±0.09	0.56±0.12
	Temporal (10-9-9)	svm-linear	0.54±0.23	0.51±0.17	0.53±0.09	0.54±0.23

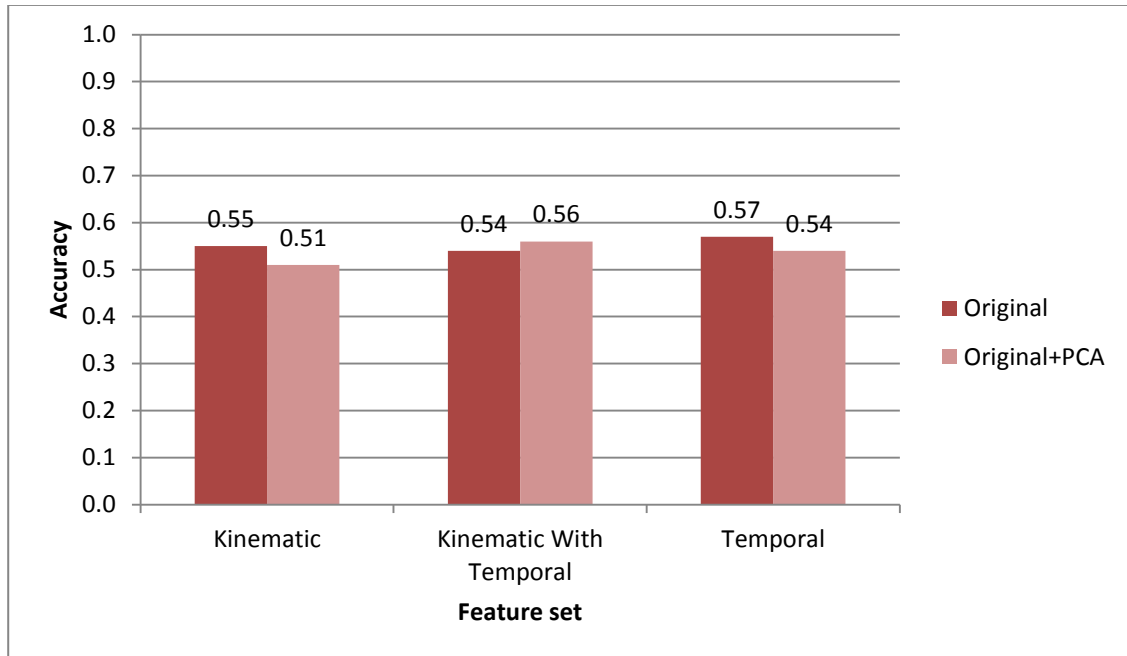
\* ค่าในวงเล็บคือจำนวนฟีเจอร์ในแต่ละเซตฟีเจอร์ (fold1-fold2-fold3)

การปรับลดตัวแปรโดยใช้วิธีวิเคราะห์องค์ประกอบ (PCA) จะทำการหาลักษณะเด่นและสร้างความสัมพันธ์ของแต่ละฟีเจอร์ ทำให้ในแต่ละเซตฟีเจอร์มีจำนวนฟีเจอร์ลดลง จากตารางที่ 4.4 จำนวนฟีเจอร์ของเซตฟีเจอร์ Kinematic จาก 2,400 ฟีเจอร์ เป็น 30-35 ฟีเจอร์ เซตฟีเจอร์ Kinematic with temporal จาก 2,450 ฟีเจอร์ เป็น 31-36 ฟีเจอร์ และเซตฟีเจอร์ Temporal จาก 50 ฟีเจอร์เป็น 9-10 ฟีเจอร์

จากผลลัพธ์ตารางที่ 4.4 จะเห็นได้ว่า เมื่อนำวิธีวิเคราะห์องค์ประกอบ(PCA) โดยมีค่า variance อยู่ที่ 95% สำหรับข้อมูลต้นฉบับ ฟีเจอร์ Kinematic with temporal ใช้เครื่องมือจำแนกแบบ SVM-Linear จะให้ค่าความแม่นยำสูงสุดซึ่งมีค่าความแม่นยำเป็น 56%

รูปที่ 4.12 เป็นกราฟแสดงการเปรียบเทียบค่าความแม่นยำจากตารางที่ 4.3 และตารางที่ 4.4 เพื่อเปรียบเทียบระหว่างชุดข้อมูลต้นฉบับกับชุดข้อมูลต้นฉบับที่ใช้วิธีวิเคราะห์องค์ประกอบ (PCA) โดยใช้เครื่องมือ SVM-Linear จะเห็นได้ว่าเซตฟีเจอร์ Kinematic With Temporal ข้อมูลต้นฉบับที่ใช้วิธี PCA จะให้ผลที่ดีกว่า แต่เซตฟีเจอร์ Kinematic และ Temporal นั้นชุดข้อมูลต้นฉบับที่ใช้วิธี PCA จะให้ผลค่าความแม่นยำที่ต่ำลง เพราะฉะนั้นยังสรุปชัดเจนไม่ได้ว่าการนำวิธี

วิเคราะห์ห้องค์ประกอบจะช่วยในการเพิ่มประสิทธิภาพของการทดลอง ในการทดลองต่อไปจะทำการปรับช่วงข้อมูลก่อนนำข้อมูลไปทดลอง



รูปที่ 4.12 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลต้นฉบับกับชุดข้อมูลต้นฉบับและใช้วิธีวิเคราะห์ห้องค์ประกอบ (PCA) โดยใช้เครื่องมือ SVM-Linear

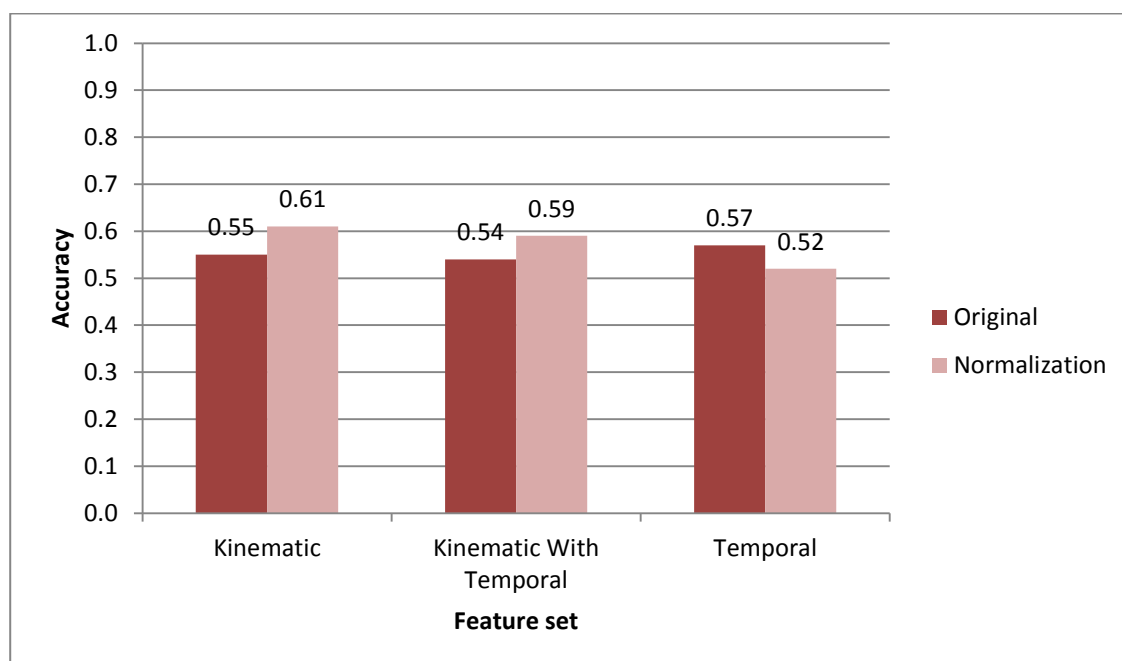
### 3.2. การทดลองเพื่อเพิ่มประสิทธิภาพโดยการปรับช่วงข้อมูล (Data Normalization)

การทดลองนี้จะใช้จำนวนตัวอย่างทั้งหมด 88 คนซึ่งข้อมูลที่น่ามาใช้เป็นชุดข้อมูลต้นฉบับที่จะผ่านการปรับช่วงข้อมูล (Data Normalization) โดยข้อมูลจะอยู่ในช่วง 0-1 ในการทดลองจะทำการทดลองโดยใช้เครื่องมือมาตรฐานคือ SVM-Linear ในการเปรียบเทียบประสิทธิภาพ ใช้วิธีการแบ่งข้อมูลด้วยโครงสร้างแบบ Flat โดยจะมีฟีเจอร์ทั้งหมดสามชุดคือ Kinematic, Kinematic with temporal และ Temporal ซึ่งมีผลการทดลองดังตารางที่ 4.5 และจากตารางที่ 4.5 สภาพแวดล้อมที่ให้ค่าความแม่นยำสูงสุดคือฟีเจอร์ Kinematic ที่ใช้เครื่องมือ SVM-Linear ซึ่งมีค่าความแม่นยำเป็น 61%

ตารางที่ 4.5 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้ข้อมูลที่มีการปรับช่วงข้อมูล

Data	Feature set*	Classifier	Accuracy	F1	Precision	Recall
Original	<b>Kinematic</b> (2,400)	svm-linear	<b>0.61±0.10</b>	<b>0.60±0.11</b>	0.62±0.13	0.61±0.10
	Kinematic WithTemporal (2,450)	svm-linear	0.59±0.12	0.58±0.11	0.60±0.12	0.59±0.12
	Temporal (50)	svm-linear	0.52±0.34	0.46±0.25	0.49±0.13	0.52±0.34

\* ค่าในวงเล็บคือจำนวนฟีเจอร์ในแต่ละเซตฟีเจอร์



รูปที่ 4.13 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลต้นฉบับกับชุดข้อมูลต้นฉบับที่ผ่านการปรับช่วงข้อมูล (Data Normalization) โดยใช้เครื่องมือ SVM-Linear

รูปที่ 4.13 เป็นกราฟแสดงการเปรียบเทียบค่าความแม่นยำจากตารางที่ 4.3 และตารางที่ 4.5 เพื่อเปรียบเทียบระหว่างชุดข้อมูลต้นฉบับกับชุดข้อมูลต้นฉบับที่ผ่านการปรับช่วงข้อมูล (Data Normalization) โดยใช้เครื่องมือ SVM-Linear จะเห็นได้ว่าเซตฟีเจอร์ Kinematic และเซตฟีเจอร์ Kinematic With Temporal ชุดข้อมูลต้นฉบับที่ผ่านการปรับช่วงข้อมูล จะให้ผลค่าความแม่นยำที่ดีกว่าชุดข้อมูลต้นฉบับ แต่เซตฟีเจอร์ Temporal นั้นชุดข้อมูลต้นฉบับที่ผ่านการปรับช่วงข้อมูลจะให้ผลค่าความแม่นยำที่ต่ำลง จะเห็นได้ว่าเมื่อข้อมูลมีการปรับช่วงข้อมูล (Data Normalization)

ส่งผลให้ผลลัพธ์มีประสิทธิภาพที่ค่อนข้างดียิ่งขึ้น ในการทดลองต่อไปจะทำการนำข้อมูลไปทำการปรับช่วงข้อมูลและวิธีวิเคราะห์องค์ประกอบแล้วนำไปทดลอง

### 3.3. การทดลองเพื่อเพิ่มประสิทธิภาพโดยการลดตัวแปรโดยใช้วิธีการวิเคราะห์องค์ประกอบ (Principle Component Analysis) และ การปรับช่วงข้อมูล (Data Normalization)

การทดลองนี้จะใช้จำนวนตัวอย่างทั้งหมด 88 คนซึ่งข้อมูลที่นำมาใช้เป็นชุดข้อมูลต้นฉบับที่นำมาใช้จะผ่านการปรับช่วงข้อมูล (Data Normalization) โดยข้อมูลจะถูกปรับให้มีค่าอยู่ในช่วง 0-1 และต่อมาจะผ่านการปรับลดตัวแปรโดยใช้วิธีวิเคราะห์องค์ประกอบ (PCA) จากการทดลองที่ผ่านมาจะเห็นได้ว่าการใช้วิธีวิเคราะห์องค์ประกอบส่งผลต่อความแม่นยำของระบบให้ดีขึ้น ในการทดลองนี้จึงปรับค่า variance อยู่ที่ 99%

ในการทดลองจะทำการทดลองโดยใช้เครื่องมือคือ SVM-Linear ในการจำแนกประเภทโดยใช้โครงสร้างแบบ Flat จะมีฟีเจอร์ในการทดลองทั้งหมดสามชุดคือ Kinematic, Kinematic with temporal และ Temporal ซึ่งมีผลการทดลองดังตารางที่ 4.6

จากตารางที่ 4.6 การปรับลดตัวแปรโดยใช้วิธีวิเคราะห์องค์ประกอบ (PCA) ค่า variance อยู่ที่ 99% จำนวนฟีเจอร์ของเซตฟีเจอร์ Kinematic จาก 2,400 ฟีเจอร์ เป็น 45-52 ฟีเจอร์ เซตฟีเจอร์ Kinematic with temporal จาก 2,450 ฟีเจอร์ เป็น 45-53 ฟีเจอร์ และเซตฟีเจอร์ Temporal จาก 50 ฟีเจอร์ เป็น 13-14 ฟีเจอร์ จากการทดลองสภาพแวดล้อมที่ให้ค่าความแม่นยำสูงสุดคือเซตฟีเจอร์ Kinematic ที่ใช้เครื่องมือ SVM-Linear ซึ่งมีค่าความแม่นยำเป็น 60%

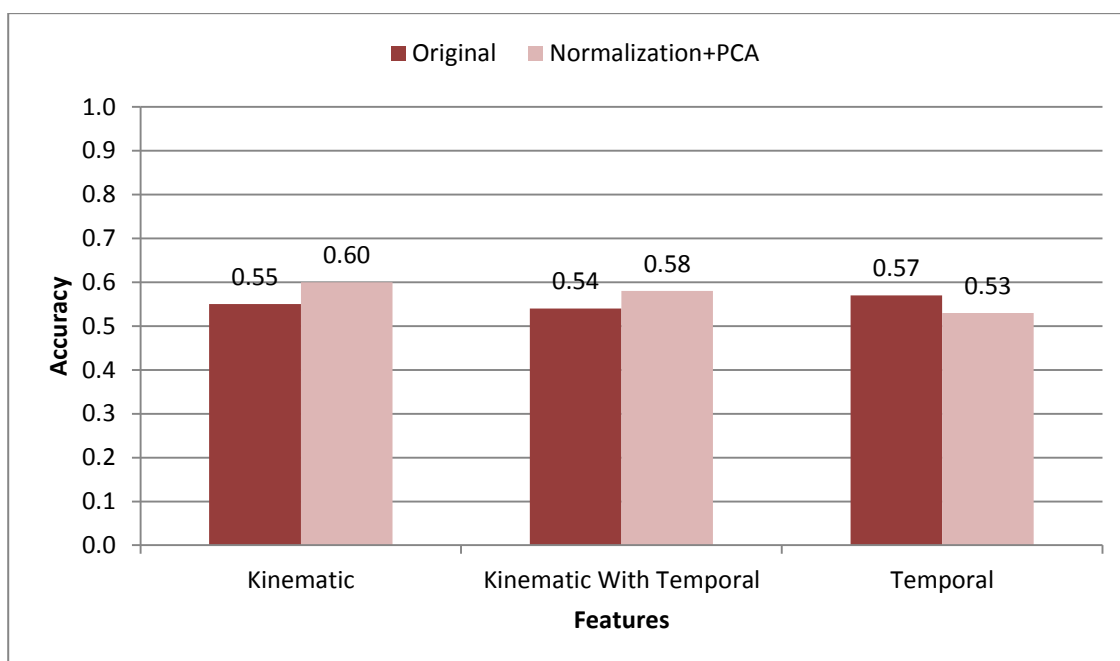
ตารางที่ 4.6 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้ข้อมูลที่มีการปรับช่วงข้อมูลและผ่านการวิเคราะห์องค์ประกอบ (PCA)

Data	Feature set*	Classifier	Accuracy	F1	Precision	Recall
Original Normalize PCA	Kinematic (52-49-45)	svm-linear	<b>0.60±0.10</b>	<b>0.58±0.12</b>	0.60±0.11	0.60±0.10
	Kinematic WithTemporal (53-49-46)	svm-linear	0.58±0.12	0.57±0.12	0.58±0.13	0.58±0.12
	Temporal (14-13-13)	svm-linear	0.53±0.17	0.53±0.12	0.53±0.08	0.53±0.17

\* ค่าในวงเล็บคือจำนวนฟีเจอร์ในแต่ละเซตฟีเจอร์ของแต่ละ fold (fold1-fold2-fold3)

รูปที่ 4.14 เป็นกราฟแสดงการเปรียบเทียบค่าความแม่นยำจากตารางที่ 4.3 และตารางที่ 4.6 เพื่อเปรียบเทียบระหว่างชุดข้อมูลต้นฉบับกับชุดข้อมูลต้นฉบับที่ผ่านการปรับช่วงข้อมูลและผ่านการวิเคราะห์องค์ประกอบ (PCA) โดยใช้เครื่องมือ SVM-Linear จะเห็นได้ว่าเซตฟีเจอร์ Kinematic และเซตฟีเจอร์ Kinematic With Temporal ชุดข้อมูลต้นฉบับที่ผ่านการปรับช่วงข้อมูล จะให้ผลค่าความแม่นยำที่ดีกว่าชุดข้อมูลต้นฉบับ แต่เซตฟีเจอร์ Temporal นั้นชุดข้อมูลต้นฉบับที่ผ่านการปรับช่วงข้อมูลจะให้ผลค่าความแม่นยำที่ต่ำลง



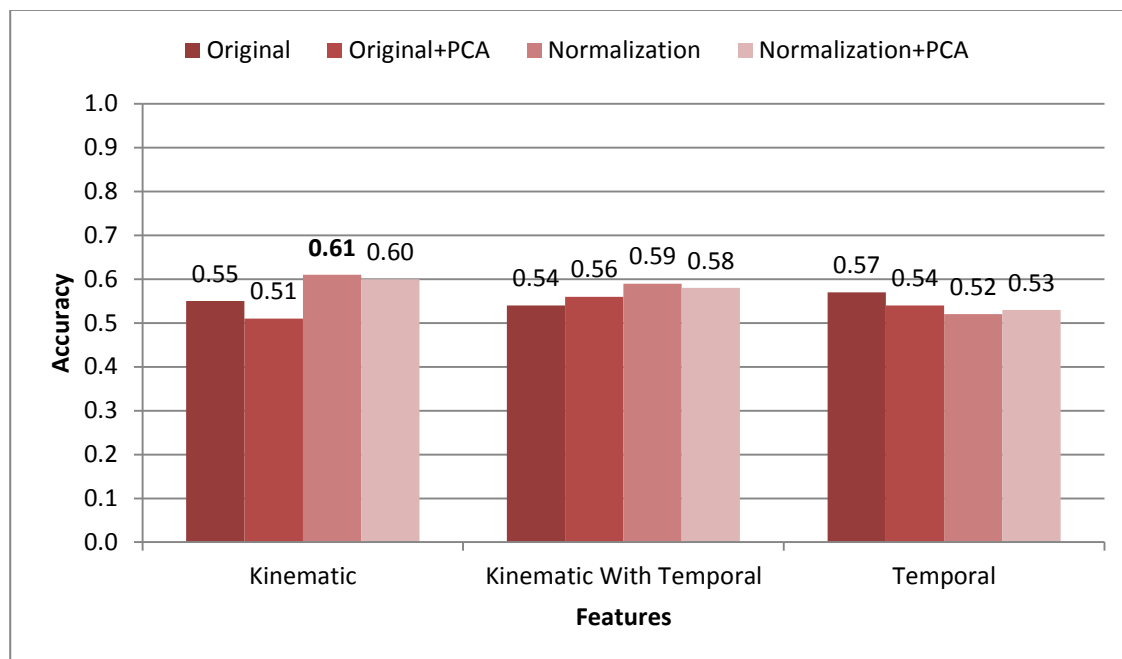


รูปที่ 4.14 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลต้นฉบับกับชุดข้อมูลต้นฉบับและใช้วิธีวิเคราะห์องค์ประกอบ (PCA) และชุดข้อมูลที่ผ่านการปรับช่วงข้อมูล (Data Normalization) โดยใช้เครื่องมือ SVM-Linear

จากรูปที่ 4.15 เป็นข้อมูลจากตารางที่ 4.3, 4.4, 4.5, 4.6 แสดงกราฟเปรียบเทียบค่าความแม่นยำระหว่างชุดข้อมูลต้นฉบับกับชุดข้อมูลที่ใช้วิธีวิเคราะห์องค์ประกอบกับชุดข้อมูลที่ผ่านการปรับช่วงข้อมูลกับชุดข้อมูลที่ผ่านการปรับช่วงข้อมูลและวิธีวิเคราะห์องค์ประกอบ โดยใช้เครื่องมือ SVM-Linear จากกราฟจะเห็นได้ว่าเซตฟีเจอร์ส่วนใหญ่เมื่อทำการปรับช่วงข้อมูล (Data Normalization) จะส่งผลให้ผลลัพธ์มีประสิทธิภาพที่ค่อนข้างดียิ่งขึ้นกว่าข้อมูลต้นฉบับ แต่จะเห็นได้ว่าเมื่อทำการลดจำนวนฟีเจอร์โดยใช้วิธีวิเคราะห์องค์ประกอบทำให้ผลความแม่นยำลดลง อาจเนื่องมาจากวิธีนี้ไปทำการลดฟีเจอร์ที่มีนัยยะสำคัญต่อการจำแนกประเภท ทำให้ผลลัพธ์มีค่าต่ำลง

จากการทดลองที่ผ่านมาทั้งหมดสภาพแวดล้อมที่ให้ค่าประสิทธิภาพค่าความแม่นยำสูงสุดคือชุดข้อมูลต้นฉบับมีกลุ่มตัวอย่าง 88 คน ที่มีการปรับช่วงข้อมูล (Data Normalization) โดยใช้เครื่องมือ SVM-Linear ในการจำแนกประเภท และมีฟีเจอร์เป็น Kinematic ซึ่งมีค่าประสิทธิภาพเป็น 61% ในการทดลองขั้นต่อไปจะทำการหาวิธีลดตัวแปรเพื่อให้มีประสิทธิภาพที่ดียิ่งขึ้น โดยจะ

ทำการขอคำแนะนำจากผู้เชี่ยวชาญเกี่ยวกับว่ามีฟีเจอร์ใดบ้างที่สำคัญและจำเป็นในการวิเคราะห์ทางด้านการแพทย์



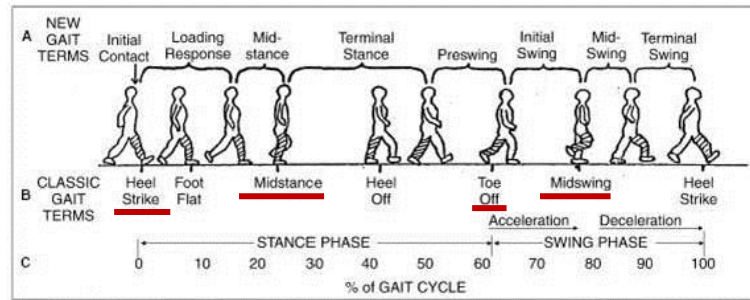
รูปที่ 4.15 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลต้นฉบับกับชุดข้อมูลต้นฉบับและใช้วิธีวิเคราะห์องค์ประกอบ (PCA) และชุดข้อมูลที่ผ่านการปรับช่วงข้อมูล (Data Normalization) โดยใช้เครื่องมือ SVM-Linear

#### 4. การทดลองเพื่อเพิ่มประสิทธิภาพโดยการลดตัวแปรโดยผู้เชี่ยวชาญ

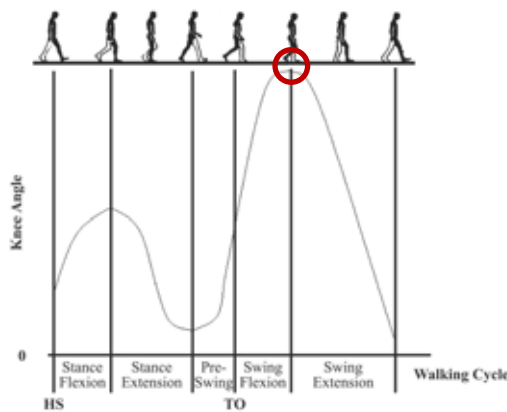
ในการทดลองนี้มีวัตถุประสงค์เพื่อเพิ่มประสิทธิภาพค่าความแม่นยำมากยิ่งขึ้น จึงทำการปรึกษากับเจ้าหน้าที่และได้รับคำแนะนำให้นำค่าเฉพาะจุดที่สำคัญเพื่อตัดข้อมูลที่มีสัญญาณรบกวนออก ซึ่งได้แก่จุด Heel Strike, Toe Off, Mid Stance, Mid Swing (รูปที่ 4.16) ซึ่งเป็นจุดที่สำคัญที่ใช้ในการวินิจฉัยทางการแพทย์

การหาค่าตำแหน่งจุด Heel Strike (ตำแหน่งที่ส้นเท้าสัมผัสกับพื้น) และ Toe Off (ตำแหน่งที่ปลายเท้ากำลังจะลอยจากพื้น) ทราบได้จากการกำหนดจากเจ้าหน้าที่ในขั้นตอนเบื้องต้น ส่วนตำแหน่ง Mid Stance เป็นตำแหน่งที่เท้าข้างนั้นรับน้ำหนักจุดศูนย์กลางมวลของร่างกาย เนื่องจากไม่สามารถหาค่าตำแหน่งได้โดยตรง จึงจะใช้ตำแหน่ง Toe off ของขาอีกข้างแทน เพราะการเดินใน

ขณะที่เท้าอีกข้างกำลังจะลอยจากพื้นนั้นเท้าอีกข้างจะรับน้ำหนักของร่างกายทั้งหมด และตำแหน่ง Mid Swing เป็นจุดที่เท้าลอยสูงสุดจากพื้น หาได้จากตำแหน่งที่ค่ามุมของข้อเข่าในแกน Sagittal มีค่าสูงสุด ดังรูปที่ 4.17

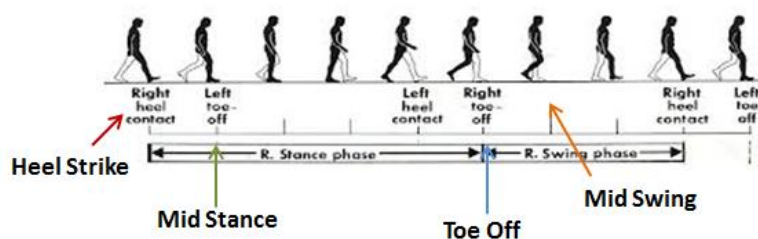


รูปที่ 4.16 รูปวงจรกิจกรรมเดิน ตำแหน่งที่ขีดเส้นใต้คือจุดที่สำคัญในการวินิจฉัยทางการแพทย์ [25]



รูปที่ 4.17 กราฟค่ามุมของข้อเข่าในแกน Sagittal [26]

การทดลองนี้จะใช้จำนวนตัวอย่างทั้งหมด 88 คน จะลดตัวแปรในพีเจอร์ Kinematic ลง โดยเลือกเฉพาะจุด Heel Strike, Toe Off, Mid Stance, Mid Swing (รูปที่ 4.18) ทำให้เซตพีเจอร์ Kinematic จาก 2,400 พีเจอร์ เป็น 96 พีเจอร์ โดยใช้ชื่อกลุ่มพีเจอร์นี้ว่า SelectKinematic



รูปที่ 4.18 ภาพแสดงตำแหน่งในวงจรกิจกรรมเดินที่เลือกใช้ในการทดลอง [27]

#### 4.1. ทดลองวัดประสิทธิภาพโดยใช้ข้อมูลต้นฉบับฟีเจอร์ SelectKinematic

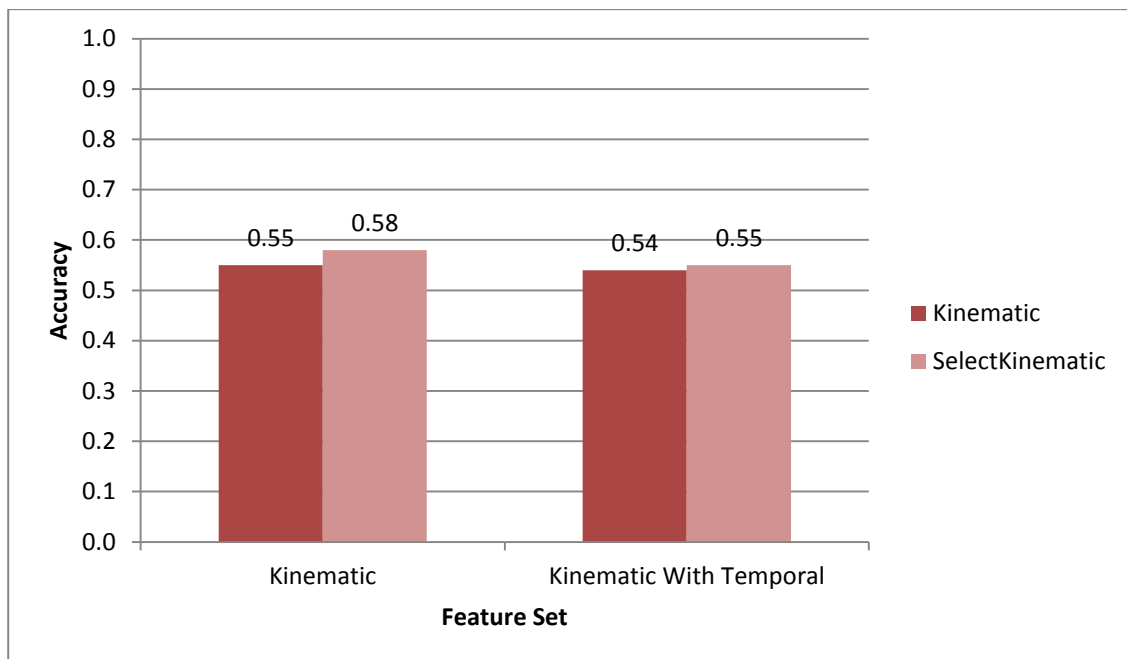
ในการทดลองจะทำการทดลองโดยใช้เครื่องมือมาตรฐานคือ SVM-Linear ในการเปรียบเทียบประสิทธิภาพ ใช้วิธีการแบ่งข้อมูลด้วยโครงสร้างแบบ Flat โดยจะมีฟีเจอร์ทั้งหมดสามชุดคือ SelectKinematic, SelectKinematic with temporal และ Temporal ข้อมูลที่นำมาทดลองเป็นชุดข้อมูลต้นฉบับ ซึ่งมีผลการทดลองตามตารางที่ 4.7 และจากตารางที่ 4.7 สภาพแวดล้อมที่ทำให้ค่าความแม่นยำสูงสุดคือฟีเจอร์ SelectKinematic ที่ใช้เครื่องมือ SVM-Linear ซึ่งมีค่าความแม่นยำเป็น 58%

ตารางที่ 4.7 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้ข้อมูลต้นฉบับฟีเจอร์ SelectKinematic

Data	Feature set*	Classifier	Accuracy	F1	Precision	Recall
Original	SelectKinematic (96)	svm-linear	<b>0.58±0.16</b>	<b>0.56±0.12</b>	0.58±0.08	0.58±0.16
	SelectKinematic WithTemporal (146)	svm-linear	0.55±0.18	0.54±0.12	0.55±0.07	0.55±0.18
	Temporal (50)	svm-linear	0.57±0.19	0.56±0.13	0.59±0.07	0.57±0.19

\* ค่าในวงเล็บคือจำนวนฟีเจอร์ในแต่ละเซตฟีเจอร์

จากรูปที่ 4.19 เป็นกราฟแสดงผลค่าความแม่นยำจากตารางที่ 4.3 และ 4.7 เพื่อเปรียบเทียบระหว่างชุดข้อมูลต้นฉบับโดยใช้ฟีเจอร์เดิม (ตามการทดลองที่ผ่านมา) และชุดข้อมูลต้นฉบับที่ใช้ฟีเจอร์ชุดใหม่ (ฟีเจอร์ SelectKinematic) โดยใช้กลุ่มตัวอย่างจำนวน 88 คน และใช้ SVM-Linear เป็นเครื่องมือในการจำแนก จะเห็นได้ว่า ข้อมูลต้นฉบับฟีเจอร์ SelectKinematic จะให้ค่าความแม่นยำสูงกว่าฟีเจอร์เดิม ในการทดลองลำดับถัดไปจะนำวิธีวิเคราะห์องค์ประกอบ (PCA) ใช้กับเซตฟีเจอร์ SelectKinematic เพื่อเปรียบเทียบกับเซตฟีเจอร์ Kinematic



รูปที่ 4.19 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลต้นฉบับฟีเจอร์ Kinematic กับชุดข้อมูลต้นฉบับฟีเจอร์ SelectKinematic โดยใช้เครื่องมือ SVM-Linear

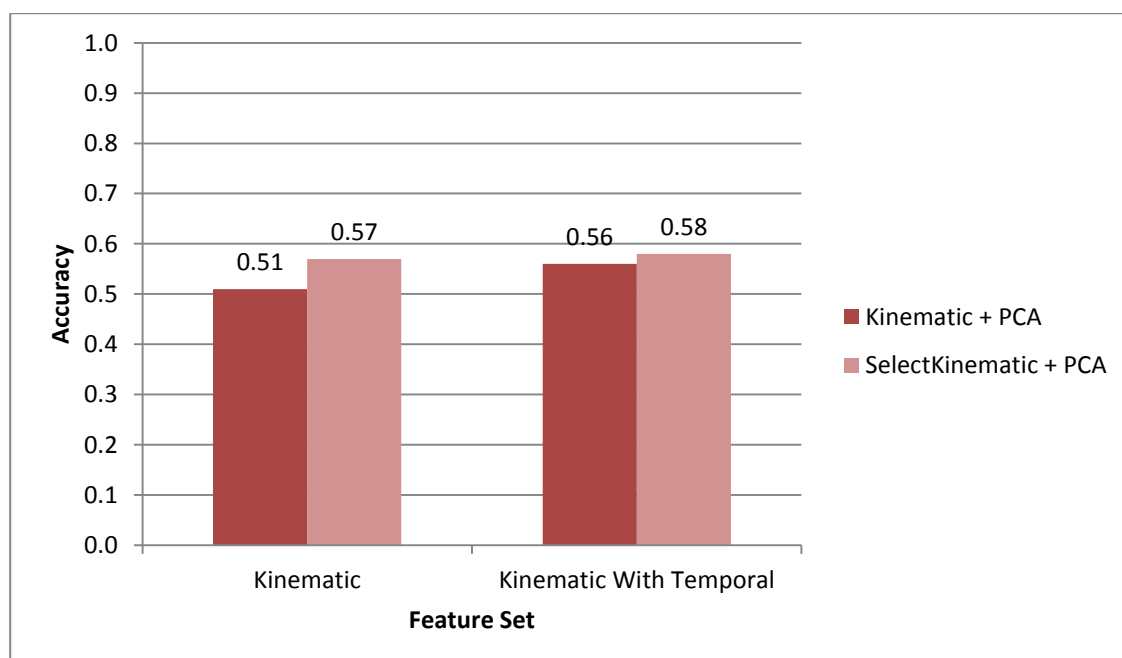
#### 4.2. ทดลองวัดประสิทธิภาพโดยใช้ข้อมูลต้นฉบับฟีเจอร์ SelectKinematic และวิธีวิเคราะห์องค์ประกอบ (PCA)

การทดลองนี้จะใช้จำนวนตัวอย่างทั้งหมด 88 คน ซึ่งข้อมูลที่น่ามาใช้จะเป็นข้อมูลต้นฉบับที่ผ่านการปรับลดตัวแปรโดยใช้วิธีวิเคราะห์องค์ประกอบ (PCA) โดยมีค่า variance อยู่ที่ 99% ในการทดลองจะทำการทดลองโดยใช้เครื่องมือมาตรฐานคือ SVM-Linear ในการเปรียบเทียบประสิทธิภาพจะใช้วิธีการแบ่งข้อมูลด้วยโครงสร้างแบบ Flat โดยจะมีฟีเจอร์ทั้งหมดสามชุดคือ SelectKinematic, SelectKinematic with temporal และ Temporal ซึ่งผลการทดลองเป็นไปตามตารางที่ 4.8 และจากตารางที่ 4.8 สภาพแวดล้อมที่ให้ค่าความแม่นยำสูงสุดคือฟีเจอร์ SelectKinematic with temporal ที่ใช้เครื่องมือ SVM-Linear ซึ่งมีค่าความแม่นยำเป็น 58%

ตารางที่ 4.8 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้วิธีวิเคราะห์องค์ประกอบ  
กับ SelectKinematic

Data	Feature set*	Classifier	Accuracy	F1	Precision	Recall
Original PCA	SelectKinematic (44-41-38)	svm-linear	0.57±0.12	0.56±0.10	0.56±0.10	0.57±0.12
	<b>SelectKinematic WithTemporal</b> (45-43-40)	svm-linear	<b>0.58±0.13</b>	<b>0.56±0.12</b>	0.55±0.12	0.58±0.13
	Temporal (13-12-12)	svm-linear	0.53±0.20	0.51±0.14	0.52±0.09	0.53±0.20

\* ค่าในวงเล็บคือจำนวนฟีเจอร์ในแต่ละเซตที่เจอรันของแต่ละ fold (fold1-fold2-fold3)



รูปที่ 4.20 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลต้นฉบับฟีเจอร์ Kinematic ผ่านวิธีวิเคราะห์องค์ประกอบ (PCA) กับชุดข้อมูลต้นฉบับฟีเจอร์ SelectKinematic ผ่านวิธีวิเคราะห์องค์ประกอบ (PCA) โดยใช้เครื่องมือ SVM-Linear

จากรูปที่ 4.20 เป็นกราฟแสดงผลค่าความแม่นยำจากตารางที่ 4.4 และ 4.8 เพื่อเปรียบเทียบระหว่างชุดข้อมูลต้นฉบับที่ใช้วิธีโดยใช้ฟีเจอร์เดิม (ตามการทดลองที่ผ่านมา) ผ่านวิธีวิเคราะห์องค์ประกอบ (PCA) และชุดข้อมูลต้นฉบับที่ใช้ฟีเจอร์ชุดใหม่ (ฟีเจอร์ SelectKinematic) ผ่านวิธีวิเคราะห์องค์ประกอบ (PCA) โดยใช้กลุ่มตัวอย่างจำนวน 88 คน และใช้ SVM-Linear เป็น

เครื่องมือในการจำแนก จะเห็นได้ว่า ข้อมูลต้นฉบับพีเจอร์ SelectKinematic จะให้ค่าความแม่นยำสูงกว่าพีเจอร์เดิม ในการทดลองลำดับถัดไปจะนำวิธีการปรับช่วงข้อมูล (Data Normalization) ใช้กับเซตพีเจอร์ SelectKinematic เพื่อเปรียบเทียบกับเซตพีเจอร์ Kinematic

#### 4.3. ทดลองวัดประสิทธิภาพโดยใช้ข้อมูลต้นฉบับพีเจอร์ SelectKinematic และเพิ่มการปรับช่วงข้อมูล

การทดลองนี้จะใช้จำนวนตัวอย่างทั้งหมด 88 คนซึ่งข้อมูลก็นำมาใช้เป็นข้อมูลต้นฉบับที่ผ่านการปรับช่วงข้อมูล (Data Normalization) โดยข้อมูลจะอยู่ในช่วง 0-1 ในการทดลองจะทำการทดลองโดยใช้เครื่องมือมาตรฐานคือ SVM-Linear ในการเปรียบเทียบประสิทธิภาพจะใช้วิธีการแบ่งข้อมูลด้วยโครงสร้างแบบ Flat โดยจะมีพีเจอร์ทั้งหมดสามชุดคือ SelectKinematic, SelectKinematic with temporal และ Temporal ซึ่งผลการทดลองเป็นไปตามตารางที่ 4.9 และจากตารางที่ 4.9 สภาพแวดล้อมที่ให้ค่าความแม่นยำสูงสุดคือพีเจอร์ SelectKinematic ที่ใช้เครื่องมือ SVM-Linear ซึ่งมีค่าความแม่นยำเป็น 0.66%

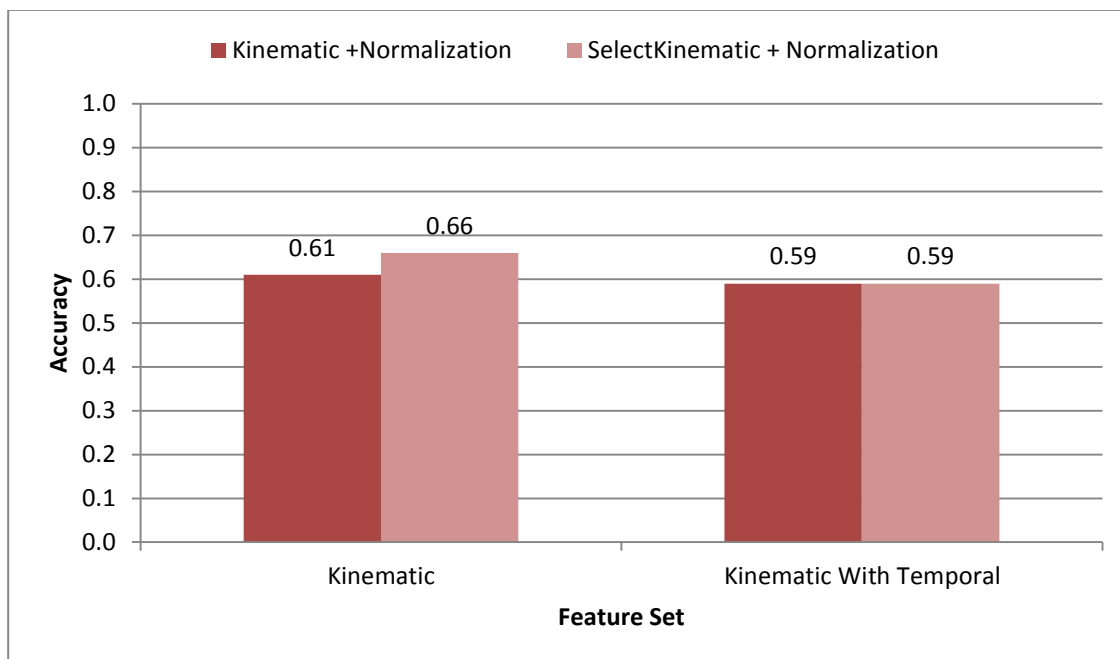
ตารางที่ 4.9 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้ข้อมูลที่มีการปรับช่วงข้อมูลกับ SelectKinematic

Data	Feature set	Classifier	Accuracy	F1	Precision	Recall
Normalize	SelectKinematic (96)	svm-linear	<b>0.66±0.18</b>	<b>0.65±0.11</b>	0.68±0.11	0.66±0.18
	SelectKinematic WithTemporal (146)	svm-linear	0.59±0.11	0.57±0.09	0.59±0.08	0.59±0.11
	Temporal (50)	svm-linear	0.52±0.34	0.46±0.25	0.49±0.13	0.52±0.34

\* ค่าในวงเล็บคือจำนวนพีเจอร์ในแต่ละเซตพีเจอร์

จากรูปที่ 4.21 เป็นการเปรียบเทียบระหว่างชุดข้อมูลต้นฉบับโดยใช้พีเจอร์เดิม (ตามการทดลองที่ผ่านมา) ผ่านการปรับช่วงข้อมูล (Data Normalization) และชุดข้อมูลต้นฉบับที่ใช้พีเจอร์ชุดใหม่ (พีเจอร์ SelectKinematic) ผ่านการปรับช่วงข้อมูล (Data Normalization) ใช้กลุ่มตัวอย่างจำนวน 88 คน และใช้ SVM-Linear เป็นเครื่องมือในการจำแนก จะเห็นได้ว่า ข้อมูลต้นฉบับพีเจอร์ SelectKinematic จะให้ค่าความแม่นยำสูงกว่าพีเจอร์เดิม ในการทดลองลำดับถัดไปจะนำวิธี

วิเคราะห์องค์ประกอบ (PCA) และการปรับช่วงข้อมูล (Data Normalization) ใช้กับเซตฟีเจอร์ SelectKinematic เพื่อเปรียบเทียบกับเซตฟีเจอร์ Kinematic



รูปที่ 4.21 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลต้นฉบับฟีเจอร์ Kinematic ผ่านการปรับช่วงข้อมูล (Data Normalization) กับชุดข้อมูลต้นฉบับฟีเจอร์ SelectKinematic ผ่านการปรับช่วงข้อมูล (Data Normalization) โดยใช้เครื่องมือ SVM-Linear

#### 4.4. ทดลองวัดประสิทธิภาพโดยใช้ข้อมูลที่มีการปรับช่วงข้อมูลและวิธีวิเคราะห์องค์ประกอบ (PCA)

การทดลองนี้จะใช้จำนวนตัวอย่างทั้งหมด 88 คนซึ่งข้อมูลที่น่ามาใช้จะเป็นข้อมูลต้นฉบับที่ผ่านการปรับช่วงข้อมูล (Data Normalization) โดยข้อมูลจะถูกปรับให้มีค่าอยู่ในช่วง 0-1 และต่อมากจะผ่านการปรับลดตัวแปรโดยใช้วิธีวิเคราะห์องค์ประกอบ (PCA) มีค่า variance อยู่ที่ 99% ในการทดลองจะทำการทดลองโดยใช้เครื่องมือมาตรฐานคือ SVM-Linear ในการเปรียบเทียบประสิทธิภาพจะใช้วิธีการแบ่งข้อมูลด้วยโครงสร้างแบบ Flat โดยจะมีฟีเจอร์ทั้งหมดสามชุดคือ SelectKinematic, SelectKinematic with temporal และ Temporal ซึ่งผลการทดลองเป็นไปตามตารางที่ 4.10 และจากตารางที่ 4.10 สภาพแวดล้อมที่ให้ค่าความแม่นยำสูงสุดคือฟีเจอร์ SelectKinematic with twmporal ที่ใช้เครื่องมือ SVM-Linear ซึ่งมีค่าความแม่นยำเป็น 0.57%

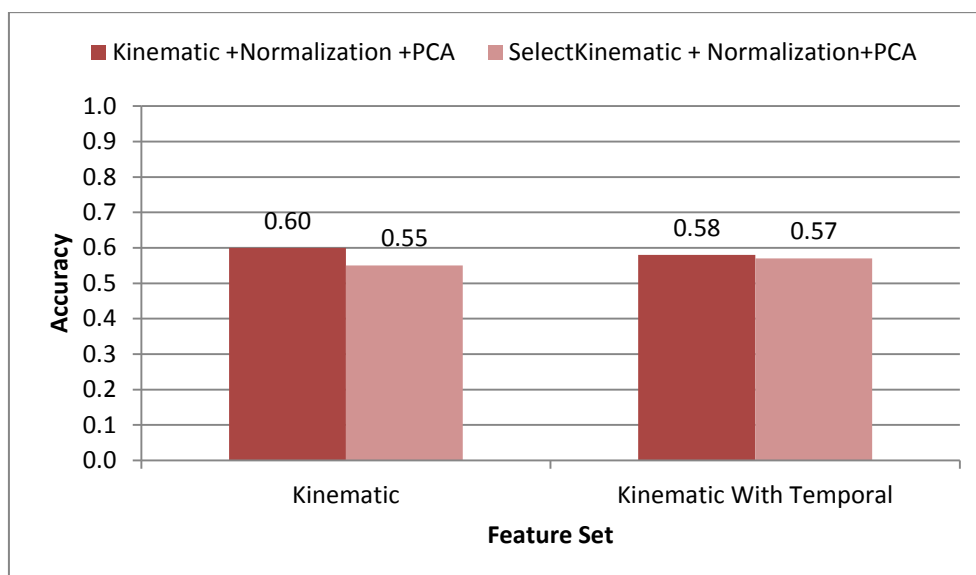


ตารางที่ 4.10 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโดยใช้ข้อมูลที่มีการปรับช่วงข้อมูลและวิธีวิเคราะห์องค์ประกอบ กับพีเจอร์ SelectKinematic

Data	Feature set*	Classifier	Accuracy	F1	Precision	Recall
Normalize PCA	SelectKinematic (44-41-38)	svm-linear	0.55±0.12	0.54±0.11	0.53±0.11	0.55±0.12
	<b>SelectKinematic WithTemporal</b> (45-43-40)	svm-linear	<b>0.57±0.12</b>	<b>0.55±0.13</b>	0.54±0.13	0.57±0.12
	Temporal (13-12-12)	svm-linear	0.53±0.17	0.53±0.08	0.53±0.12	0.53±0.17

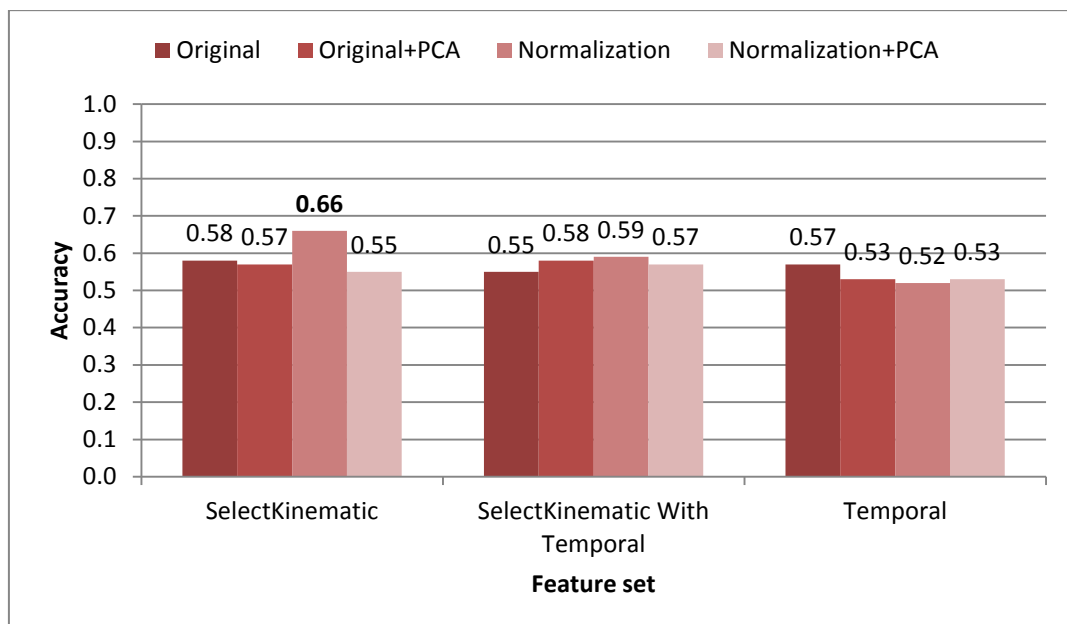
\* ค่าในวงเล็บคือจำนวนพีเจอร์ในแต่ละชุดพีเจอร์ของแต่ละ fold (fold1-fold2-fold3)

จากรูปที่ 4.22 เป็นการเปรียบเทียบระหว่างชุดข้อมูลต้นฉบับโดยใช้พีเจอร์เดิม (ตามการทดลองที่ผ่านมา) ผ่านการปรับช่วงข้อมูล (Data Normalization) และวิธีวิเคราะห์องค์ประกอบ (PCA) และชุดข้อมูลต้นฉบับที่ใช้พีเจอร์ชุดใหม่ (พีเจอร์ SelectKinematic) ผ่านการปรับช่วงข้อมูล (Data Normalization) และวิธีวิเคราะห์องค์ประกอบ (PCA) ใช้กลุ่มตัวอย่างจำนวน 88 คน และใช้ SVM-Linear เป็นเครื่องมือในการจำแนก จะเห็นได้ว่า ข้อมูลต้นฉบับพีเจอร์ SelectKinematic with temporal จะให้ค่าความแม่นยำต่ำกว่าพีเจอร์เดิม



รูปที่ 4.22 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพชุดข้อมูลต้นฉบับที่ผ่านการปรับช่วงข้อมูล (Data Normalization) และวิธีวิเคราะห์องค์ประกอบ (PCA) ระหว่างพีเจอร์ Kinematic กับพีเจอร์ SelectKinematic โดยใช้เครื่องมือ SVM-Linear

รูปที่ 4.23 เป็นการเปรียบเทียบประสิทธิภาพค่าความแม่นยำของข้อมูลต้นฉบับ, ข้อมูลต้นฉบับที่ผ่านวิธีวิเคราะห์องค์ประกอบ (PCA), ข้อมูลต้นฉบับที่ผ่านการปรับช่วงข้อมูล (Data Normalization) และข้อมูลต้นฉบับที่ผ่านการปรับช่วงข้อมูล (Data Normalization) กับวิธีวิเคราะห์องค์ประกอบ (PCA) ในพีเจอร์ SelectKinematic จากการทดลองที่ผ่านมาทั้งหมดสภาพแวดล้อมที่ให้ค่าประสิทธิภาพค่าความแม่นยำสูงสุดคือชุดข้อมูลต้นฉบับที่มีกลุ่มตัวอย่าง 88 คนที่มีการปรับช่วงข้อมูล (Data Normalization) โดยใช้เครื่องมือ SVM-Linear ในการจำแนกประเภท และมีพีเจอร์เป็น SelectKinematic ซึ่งมีค่าประสิทธิภาพเป็น 66%



รูปที่ 4.23 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพสภาพแวดล้อมของชุดข้อมูลต้นฉบับพีเจอร์ SelectKinematic

จากการทดลองจะเห็นได้ว่าการลดปริมาณจำนวนพีเจอร์ซึ่งได้เป็นเซตพีเจอร์ SelectKinematic จากคำแนะนำของเจ้าหน้าที่และข้อมูลที่มีการปรับช่วงข้อมูลจะส่งผลให้ผลการทดลองมีประสิทธิภาพที่ดียิ่งขึ้น เพราะฉะนั้นในการทดลองถัดไปจะใช้ข้อมูลเซต SelectKinematic ที่ผ่านการปรับช่วงข้อมูล (Data Normalization)

จากการทดลองที่ผ่านมา จากการทดลองแรกมีค่าความแม่นยำเป็น 57% จนถึงการทดลองนี้มีค่าความแม่นยำเพิ่มขึ้นเป็น 66% โดยมีสภาพแวดล้อมคือชุดข้อมูลต้นฉบับที่มีกลุ่มตัวอย่าง 88

คนที่มีการปรับช่วงข้อมูล (Data Normalization) โดยใช้เครื่องมือ SVM-Linear ในการจำแนกประเภท มีฟีเจอร์เป็น SelectKinematic และในการทดลองที่ผ่าน ๆ มาใช้โครงสร้างแบบ Flat ในการทดลองต่อไปจะทำการทดลองเพื่อเพิ่มประสิทธิภาพโดยใช้โครงสร้าง Classifier แบบ Hier และแบบ One VS One บนชุดข้อมูลที่ผ่านการปรับช่วงข้อมูล

## 5. การทดลองเพื่อเปรียบเทียบประสิทธิภาพของโครงสร้าง Classifier

ในการทดลองที่ผ่านมาได้ใช้วิธีการแบ่งข้อมูลด้วยโครงสร้างแบบ Flat ในการทดลองถัดไปจะทดสอบและเปรียบเทียบโดยใช้วิธีการแบ่งข้อมูลด้วยโครงสร้างรูปแบบอื่น ๆ เพื่อหารูปแบบที่ส่งผลให้มีประสิทธิภาพที่มากขึ้น โดยจะใช้ข้อมูลจำนวนตัวอย่างทั้งหมด 88 คน ผ่านวิธีปรับช่วงข้อมูล (Data Normalization) และฟีเจอร์ SelectKinematic ซึ่งจากการทดลองที่ผ่านมามีค่าความแม่นยำที่มีค่าสูงสุดเป็น 66%

### 5.1. การทดลองเพื่อวัดประสิทธิภาพโดยใช้โครงสร้างแบบ Two-level hierarchical classification

ในการทดลองนี้จะใช้โครงสร้างแบบ Hier เป็นแบบ Two-level hierarchical classification โดยระดับแรกจะจำแนกผู้ป่วยออกเป็น Normal หรือ Sick ถ้าเป็นผู้ป่วยจะไปจำแนกในระดับที่สองว่าเป็น Parkinson หรือ KneeOA

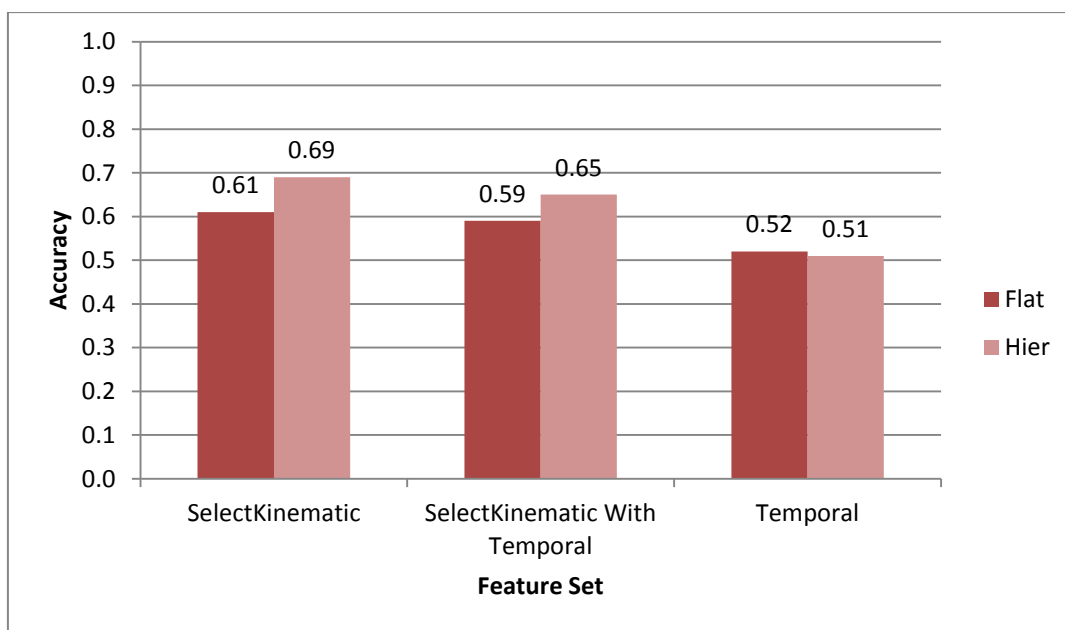
ข้อมูลที่ใช้มีจำนวนตัวอย่างทั้งหมด 88 คน เป็นข้อมูลต้นฉบับที่ผ่านวิธีปรับช่วงข้อมูล (Data Normalization) จะทำการทดลองโดยใช้เครื่องมือมาตรฐานคือ SVM-Linear ในการเปรียบเทียบประสิทธิภาพจะใช้วิธีการแบ่งข้อมูลด้วยโครงสร้างแบบ Hier โดยจะมีฟีเจอร์ทั้งหมดสามชุดคือ SelectKinematic, SelectKinematic with temporal และ Temporal ซึ่งมีผลการทดลองดังตารางที่ 4.11 และ จากตารางที่ 4.11 สภาพแวดล้อมที่ให้ค่าความแม่นยำสูงสุดคือฟีเจอร์ SelectKinematic ที่ใช้เครื่องมือ SVM-Linear ซึ่งมีค่าความแม่นยำเป็น 0.69% ซึ่งมีค่าความแม่นยำมากขึ้นจากการทดลองที่ผ่านมา

ตารางที่ 4.11 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโครงสร้างแบบ Hier โดยใช้ข้อมูลที่มีการปรับช่วงข้อมูล (Data Normalization)

Data	Feature set*	Classifier	Accuracy	F1	Precision	Recall
Normalize	SelectKinematic (96)	svm-linear	<b>0.69±0.08</b>	<b>0.68±0.08</b>	0.68±0.06	0.69±0.08
	SelectKinematic WithTemporal (146)	svm-linear	0.65±0.06	0.64±0.05	0.64±0.05	0.65±0.06
	Temporal (50)	svm-linear	0.51±0.26	0.49±0.18	0.51±0.11	0.51±0.26

\* ค่าในวงเล็บคือจำนวนฟีเจอร์ในแต่ละเซตฟีเจอร์

จากรูปที่ 4.24 เป็นการเปรียบเทียบระหว่างโครงสร้างแบบ Flat และ Hier ของชุดข้อมูลต้นฉบับฟีเจอร์ SelectKinematic ผ่านการปรับช่วงข้อมูล (Data Normalization) ใช้กลุ่มตัวอย่างจำนวน 88 คน และใช้ SVM-Linear เป็นเครื่องมือในการจำแนก จะเห็นได้ว่าโครงสร้างแบบ Hier จะให้ค่าความแม่นยำสูงกว่าโครงสร้างแบบ Flat ในฟีเจอร์ SelectKinematic และ SelectKinematic with temporal



รูปที่ 4.24 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพระหว่างชุดข้อมูลต้นฉบับฟีเจอร์ Kinematic ผ่านการปรับช่วงข้อมูล (Data Normalization) กับชุดข้อมูลต้นฉบับฟีเจอร์ SelectKinematic ผ่านการปรับช่วงข้อมูล (Data Normalization) โดยใช้เครื่องมือ SVM-Linear

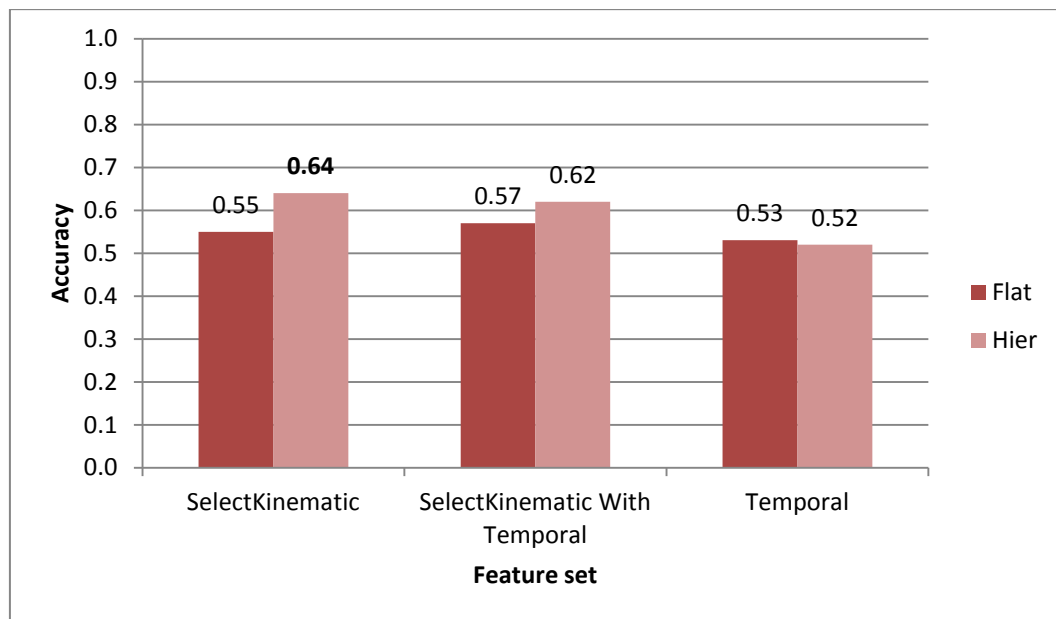
ถัดไปจะเพิ่มการทดลองโดยใช้ข้อมูลต้นฉบับและเพิ่มวิธีวิเคราะห์หองค์ประกอบ จะทำการทดลองโดยใช้เครื่องมือมาตรฐานคือ SVM-Linear ในการเปรียบเทียบประสิทธิภาพ ใช้วิธีการแบ่งข้อมูลด้วยโครงสร้างแบบ Hier โดยจะมีฟีเจอร์ทั้งหมดสามชุดคือ SelectKinematic, SelectKinematic with temporal และ Temporal ซึ่งมีผลการทดลองเป็นไปตามตารางที่ 4.12 และจากตารางที่ 4.12 สภาพแวดล้อมที่ให้ค่าความแม่นยำสูงสุดคือฟีเจอร์ SelectKinematic ที่ใช้เครื่องมือ SVM-Linear ซึ่งมีค่าความแม่นยำเป็น 0.64%

ตารางที่ 4.12 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโครงสร้างแบบ Hier โดยข้อมูลที่มีการปรับช่วงข้อมูล (Data Normalization) และใช้วิธีวิเคราะห์หองค์ประกอบ (PCA)

Data	Feature set*	Classifier	Accuracy	F1	Precision	Recall
Normalize PCA	SelectKinematic (LV1 : 44-41-38) (LV2 : 35-31-31)	svm-linear	<b>0.64±0.05</b>	<b>0.63±0.08</b>	0.62±0.07	0.64±0.05
	SelectKinematic WithTemporal (LV1 : 45-43-40) (LV2 : 36-32-33)	svm-linear	0.62±0.12	0.61±0.10	0.60±0.12	0.62±0.12
	Temporal (LV1 : 13-12-12) (LV2 : 12-12-12)	svm-linear	0.52±0.20	0.52±0.16	0.52±0.18	0.52±0.20

\* ค่าในวงเล็บคือจำนวนฟีเจอร์ในแต่ละเซตฟีเจอร์ของแต่ละ fold แต่ละ Level (Level : fold1-fold2-fold3)

จากรูปที่ 4.25 เป็นการเปรียบเทียบระหว่างโครงสร้างแบบ Flat และ Hier ของชุดข้อมูลต้นฉบับฟีเจอร์ SelectKinematic ผ่านการปรับช่วงข้อมูล (Data Normalization) และใช้วิธีวิเคราะห์หองค์ประกอบ (PCA) ใช้กลุ่มตัวอย่างจำนวน 88 คน และใช้ SVM-Linear เป็นเครื่องมือในการจำแนก จะเห็นได้ว่าโครงสร้างแบบ Hier จะให้ค่าความแม่นยำสูงกว่าโครงสร้างแบบ Flat ในฟีเจอร์ SelectKinematic และ SelectKinematic with temporal แต่ค่าความแม่นยำสูงสุดยังต่ำกว่าการทดลองที่ผ่านมา ในการทดลองนี้จะเห็นได้ว่าฟีเจอร์ SelectKinematic นั้นให้ค่าความแม่นยำสูงที่สุด ในการทดลองถัดไปจึงเลือกใช้เฉพาะฟีเจอร์ SelectKinematic ในการทดลองโครงสร้างแบบ One VS One



รูปที่ 4.25 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพชุดข้อมูลต้นฉบับพีเจอร์ SelectKinematic ผ่านการปรับช่วงข้อมูล (Data Normalization) และใช้วิธีวิเคราะห์องค์ประกอบ (PCA) ระหว่างโครงสร้างแบบ Flat กับโครงสร้างแบบ Hier โดยใช้เครื่องมือ SVM-Linear

## 5.2. การทดลองเพื่อวัดประสิทธิภาพโดยใช้โครงสร้างแบบ Binary Classifier Approach (One VS One)

ในการทดลองนี้จะใช้โครงสร้างแบบ Binary Classifier Approach (One vs One) ซึ่งข้อมูลจะผ่านตัวจำแนกประเภทสามตัว ตัวแรกจะเป็นการจำแนกระหว่างคนปกติกับผู้ป่วยข้อเข่าเสื่อม ตัวที่สองจะจำแนกระหว่างคนปกติกับผู้ป่วยพาร์กินสัน ตัวที่สามจะจำแนกระหว่างผู้ป่วยข้อเข่าเสื่อมกับผู้ป่วยพาร์กินสัน หลังจากนั้นจะนำคำตอบทั้งสามมาทำการโหวต คำตอบใดมีค่ามากที่สุดจะเลือกคำตอบนั้น หากเกิดกรณีที่คำตอบมีปริมาณเท่ากันจะยึดคำตอบที่ได้จากตัวจำแนกที่มีความแม่นยำใน Training data ที่มากที่สุด

ข้อมูลที่ใช้มีจำนวนตัวอย่างทั้งหมด 88 คน เป็นข้อมูลต้นฉบับที่ผ่านวิธีปรับช่วงข้อมูล (Data Normalization) จะทำการทดลองโดยใช้เครื่องมือมาตรฐานคือ SVM-Linear ในการเปรียบเทียบประสิทธิภาพจะใช้วิธีการแบ่งข้อมูลด้วยโครงสร้างแบบ Binary Classifier Approach โดยจะมีพีเจอร์ คือ SelectKinematic ซึ่งมีผลการทดลองตามตารางที่ 4.13 และ จากตารางที่ 4.13

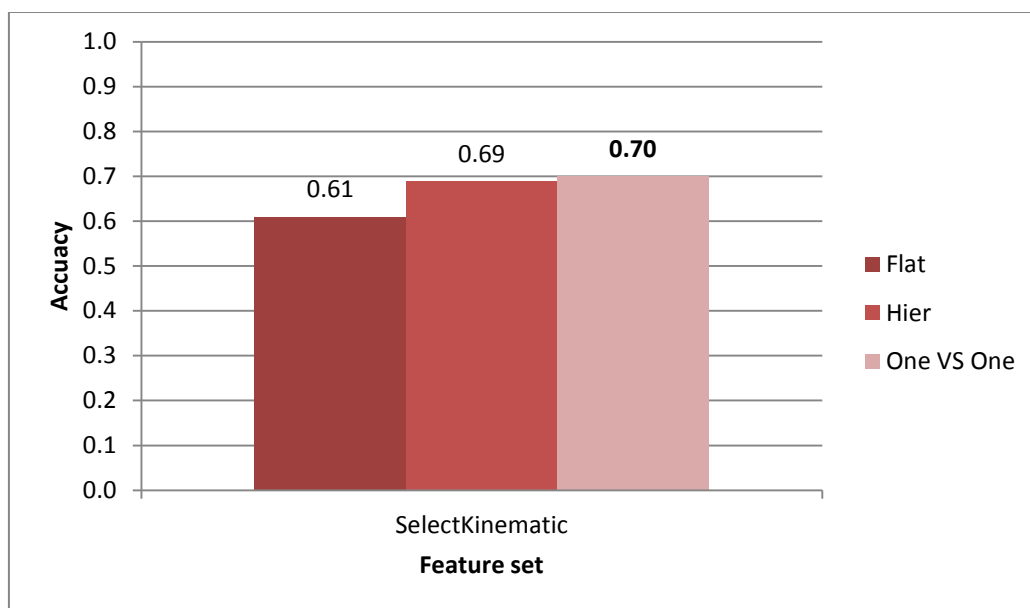
สภาพแวดล้อมที่ให้ค่าความแม่นยำสูงสุดสำหรับข้อมูลต้นฉบับที่ผ่านการปรับช่วงข้อมูล (Data Normalization) ในพีเจอร์ SelectKinematic ที่ใช้เครื่องมือ SVM-Linear โครงสร้างแบบ One VS One ให้ค่าความแม่นยำเป็น 0.70%

ตารางที่ 4.13 ผลลัพธ์ของการทดลองประสิทธิภาพตัวจำแนกประเภทโครงสร้างแบบ One vs One

Data	Feature set*	Classifier	Accuracy	F1	Precision	Recall
Normalize	SelectKinematic (96)	svm-linear	0.70±0.13	0.68±0.09	0.70±0.05	0.70±0.13

\* ค่าในวงเล็บคือจำนวนพีเจอร์ในแต่ละเซตพีเจอร์

จากรูปที่ 4.26 จะเห็นได้ว่าในชุดข้อมูลต้นฉบับพีเจอร์ SelectKinematic ผ่านการปรับช่วงข้อมูล โดยใช้เครื่องมือ SVM-Linear โครงสร้างแบบ One vs One จะให้ค่าความแม่นยำสูงสุด ซึ่งสูงกว่าโครงสร้างแบบ Flat และ Hier มีค่าความแม่นยำเป็น 70%



รูปที่ 4.26 กราฟแสดงค่าความแม่นยำเปรียบเทียบประสิทธิภาพชุดข้อมูลต้นฉบับพีเจอร์ SelectKinematic ผ่านการปรับช่วงข้อมูล (Data Normalization) โดยใช้เครื่องมือ SVM-Linear ระหว่างโครงสร้างแบบ Flat, Hier และ One VS One

จากทดลองที่ผ่านมาทั้งหมดจะเห็นได้ว่ารูปแบบที่ให้ค่าความแม่นยำที่สูงที่สุด ซึ่งมีความแม่นยำเท่ากับ 70% คือการนำข้อมูลต้นฉบับผ่านการปรับช่วงข้อมูล โดยมีพีเจอร์เป็น

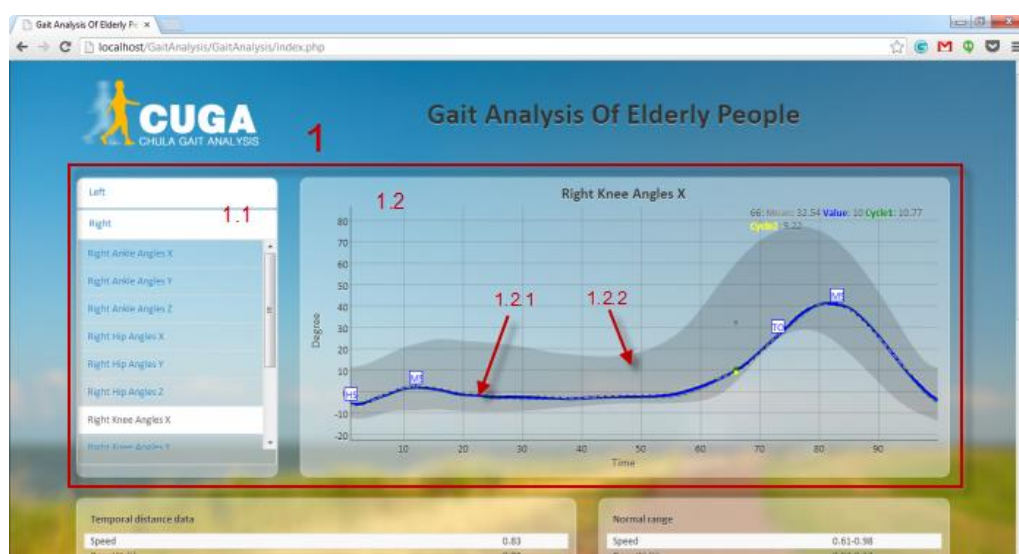
SelectKinematic ใช้ SVM-Linear เครื่องมือในการจำแนกประเภทของข้อมูลที่มีโครงสร้างแบบ One VS One

## 6. ต้นแบบโปรแกรมของระบบวินิจฉัยโรค

ในโปรแกรมจะมีส่วนที่แสดงค่าข้อมูลที่สำคัญของการเดิน ทั้งข้อมูล Kinematic Data และ Distance Temporal Data โดยจะแบ่งเป็น สาม ส่วนหลัก ๆ คือ ส่วนแสดงข้อมูล Kinematic , ส่วนแสดงข้อมูล Temporal และส่วนวิเคราะห์ทำนายผลวินิจฉัย มีรายละเอียดดังนี้

### 6.1. ส่วนแสดงข้อมูล Kinematic

เมื่อทำการรับข้อมูลการเดินเรียบร้อยแล้ว โปรแกรมจะทำการแสดงผลค่าข้อมูล Kinematic ซึ่งแสดงผลเป็นกราฟระหว่างมุมมองศำวัดตามแกนทางการแพทย์ของข้อต่าง ๆ กับเวลาในรูปแบบเปอร์เซ็นต์วงจรการเดิน (% Gait Cycle) ตามรูปที่ 4.27 โดย หมายเลข 1.1 จะแสดงรายการของกราฟทั้งหมดแบ่งเป็น 2 หมวดคือขาข้างซ้ายและขาข้างขวา แต่ละหมวดมี 12 กราฟ หมายเลข 1.2 เป็นกราฟที่จะแสดงข้อมูลค่าระหว่างมุมมองศำวัดตามแกนทางการแพทย์ของข้อต่าง ๆ กับเวลาในรูปแบบเปอร์เซ็นต์วงจรการเดิน (% Gait Cycle) โดยเส้นสีน้ำเงิน (หมายเลข 1.2.1) หมายถึงค่าเฉลี่ยของข้อมูล หากมีมากกว่าหนึ่งวงจรการเดินเส้นปะจะหมายถึงค่าข้อมูลในแต่ละวงจรการเดิน พื้นที่สีเทา (หมายเลข 1.2.2) หมายถึงขอบเขตค่าเกณฑ์ของผู้สูงปกติ



รูปที่ 4.27 ตัวอย่างต้นแบบโปรแกรมของระบบวินิจฉัยโรคส่วนแรก



พื้นที่สีเทาในรูปที่ 4.27 จะเป็นตัวช่วยให้กับผู้ใช้งานในวิเคราะห์ข้อมูล ถ้าเส้นสีน้ำเงินอยู่ในขอบเขตพื้นที่สีเทาแสดงว่าคนนั้นอยู่ในเกณฑ์ปกติ แต่ถ้าเส้นสีน้ำเงินออกนอกพื้นที่สีเทามากเกินไป ควรสามารถวิเคราะห์เบื้องต้นได้ว่าคนนั้นมีความผิดปกติที่เกี่ยวกับการเดิน

## 6.2. ส่วนแสดงข้อมูล Distance Temporal

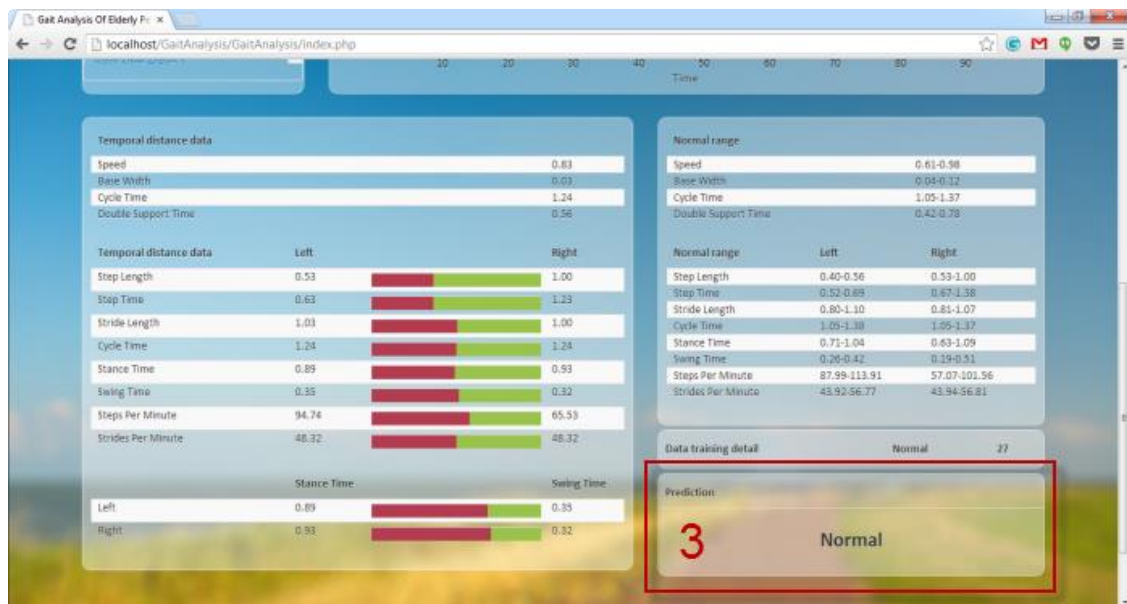
รูปที่ 4.28 เป็นการแสดงข้อมูล Distance Temporal เป็นการแสดงข้อมูลในรูปแบบของเวลาและระยะทาง เมื่อรับข้อมูลการเดินสมบูรณ์ โปรแกรมจะทำการคำนวณและแสดงผลในส่วนต่อที่หมายเลข 2.1 จะแสดงข้อมูล Distance Temporal ทั้งหมดทั้งข้างซ้ายและข้างขวา แต่ละค่ามีการแสดงอัตราส่วนการเปรียบเทียบระหว่างทั้งสองค่า ในหมายเลข 2.2 เป็นส่วนที่แสดงข้อมูลขอบเขตของแต่ละค่า Distance Temporal ในคนสูงอายุปกติ ส่วนหมายเลข 2.3 บอกจำนวนของข้อมูลคนปกติที่ใช้ในการบอกขอบเขตของคนปกติ ซึ่งสามารถนำค่าที่แสดงผลในหมายเลข 2.1 มาเปรียบเทียบในส่วน 2.2 นี้ได้ หากค่านั้นเกินขอบเขตมากเกินไป แสดงว่าคนนั้นอาจมีความผิดปกติเกิดขึ้น อัตราส่วนระหว่างพื้นที่สีแดงและสีเขียวจะเป็นตัวช่วยสำหรับผู้ใช้งานในการวิเคราะห์ข้อมูลเบื้องต้น ทำให้สามารถเปรียบเทียบค่าระหว่างข้างซ้ายและข้างขวาได้ ส่วนมากค่าเฉลี่ยแต่ละฟิเจอร์ของขาแต่ละข้างจะมีค่าแตกต่างกันไม่มากนัก หากน้ำหนักไปข้างใดข้างหนึ่งมากเกินไปจะสังเกตได้ว่าคนนั้นมีความผิดปกติเกิดขึ้น หรืออย่างบางฟิเจอร์เช่น ค่า Stance time ตามทฤษฎีจะมีค่าประมาณ 70% ของวงจรรการเดิน และค่า Swing time จะมีค่าประมาณ 30% ของวงจรรการเดิน



รูปที่ 4.28 ตัวอย่างต้นแบบโปรแกรมของระบบวินิจฉัยโรคส่วนที่สอง

### 6.3. ส่วนแสดงข้อมูลวิเคราะห์วินิจฉัยผล

รูปที่ 4.29 เป็นการแสดงผลส่วนสุดท้ายที่ผ่านการคำนวณตามสภาพแวดล้อมที่ได้จากการทดลองที่มีค่าความแม่นยำมากที่สุด และส่งผลคำตอบค่าทำนายกลับมา ซึ่งคำตอบได้แก่ คนปกติ (Normal), ผู้ป่วยพาร์กินสัน (Parkinson), ผู้ป่วยข้อเข่าเสื่อม (Knee OA)



รูปที่ 4.29 ตัวอย่างต้นแบบโปรแกรมของระบบวินิจฉัยโรคส่วนที่สาม

ตัวต้นแบบของโปรแกรมวินิจฉัยโรค ประกอบไปด้วย ส่วนแสดงผลข้อมูลและส่วนแสดงผลวินิจฉัย ในส่วนแสดงผลข้อมูล จะมีข้อมูลสองส่วนคือ ข้อมูล Kinematic และข้อมูล Distance temporal ในส่วนแสดงผลวินิจฉัยจะสามารถวินิจฉัยออกมาได้เป็นสามแบบคือ คนปกติ (Normal), ผู้ป่วยพาร์กินสัน (Parkinson), ผู้ป่วยข้อเข่าเสื่อม (Knee OA) ในส่วนแสดงผลข้อมูลนอกจากจะแสดงข้อมูลที่ได้รับมาแล้ว มีการแสดงข้อมูลขอบเขตค่าของผู้สูงอายุที่ปกติ เพื่อใช้ในวินิจฉัยโรค ประกอบกับค่าทำนายผลที่ได้จากโปรแกรม

## บทที่ 5

### ข้อสรุปและข้อเสนอแนะ

#### ข้อสรุป

งานวิจัยชิ้นนี้ได้นำเสนอระบบวิเคราะห์ข้อมูลการเดินโดยใช้ข้อมูลจากระบบ Motion Capture เพื่อวินิจฉัยโรคโดยใช้เครื่องมือจำแนกประเภทมาตรฐานในการจำแนกผู้ป่วยที่มีโรคเกี่ยวกับการเดิน โดยจะแบ่งออกเป็น 3 คลาส คือ คนปกติ (Normal), ผู้ป่วยข้อเข่าเสื่อม (Sick/Knee OA), ผู้ป่วยพาร์กินสัน (Sick/Parkinson) เพื่อช่วยให้สามารถใช้ระบบนี้ในการจำแนกผู้ป่วยได้ในเบื้องต้น

ภาพรวมของงานวิจัยมีสามส่วนหลัก ๆ คือ ส่วนการเก็บข้อมูล (Data Collection) การเตรียมข้อมูล (Data Preprocess) และการวิเคราะห์ข้อมูล (Data Analysis) ขั้นตอนในการเก็บข้อมูล ข้อมูลที่ใช้ได้นั้นได้มาจากศูนย์ความเป็นเลิศทางการแพทย์ด้านการเดินและเคลื่อนไหว โรงพยาบาลจุฬาลงกรณ์ ซึ่งเป็นระบบ Motion Capture โดยข้อมูลนั้นแบ่งออกเป็นสามกลุ่ม ได้แก่ ผู้สูงอายุปกติ ผู้ป่วยพาร์กินสัน ผู้ป่วยข้อเข่าเสื่อม ซึ่งมีอายุ 60 ปีขึ้นไป คณะแพทย์ และได้รับการวินิจฉัยจากแพทย์เฉพาะทาง เนื่องจากว่าข้อมูลที่ได้รับมานั้นเป็นข้อมูลดิบไม่สามารถนำไปวิเคราะห์โรคได้ จึงต้องมีการเตรียมข้อมูลโดยขั้นแรกทำการแปลงข้อมูลการเดินให้อยู่ในรูปของเปอร์เซ็นต์ของวงจรถ่วงการเดินซึ่งเป็นค่าของมุมในแต่ละข้อของแต่ละแกนทางการแพทย์ของขาทั้งสองข้าง ซึ่งข้อมูลชุดนี้มีชื่อว่า Kinematic มีทั้งหมด 2,400 ฟิเจอร์ ต่อจากนั้นคำนวณหาค่าตัวแปรในมุมมองของระยะทางและเวลา เช่น ความเร็วในการเดิน (Speed), ระยะก้าว (Step Length), ระยะเวลาที่เท้าสัมผัสกับพื้น (Stance Time), ระยะเวลาที่เท้าไม่สัมผัสกับพื้น (Swing Time) ฯลฯ ซึ่งเรียกตัวแปรชุดนี้ว่า Distance Temporal มีทั้งหมด 30-50 ฟิเจอร์ เนื่องจากข้อมูลที่ได้มาอาจมีข้อผิดพลาดจากของระบบ Motion Capture ที่ไม่สามารถบันทึกข้อมูลบางช่วงได้ จึงเกิดการขาดหายของข้อมูลบางช่วง ในขั้นตอนนี้จึงมีการเติมข้อมูลที่หายไปโดยใช้ค่าเฉลี่ยของค่าก่อนและหลังจากช่วงที่หายไปเติมแทน เมื่อเตรียมข้อมูลพร้อมแล้ว ส่วนสุดท้ายจะนำข้อมูลไปทำการวิเคราะห์ โดยใช้เครื่องมือมาตรฐานในการจำแนกประเภท ในการทดลองจะเลือกใช้อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine - SVM) และอัลกอริทึมต้นไม้ตัดสินใจ

(Decision Tree) มีโครงสร้างในการจำแนกสามแบบคือ แบบ Flat เป็นโครงสร้างแบบ Single multiclass classifier แบบที่สองแบบ Hier (Hierarchical) เป็นแบบจำแนกออกเป็นสองระดับ และแบบที่สามแบบ Binary classifier approach (One VS One) โดยทั้งสามจะจำแนกข้อมูลออกเป็นสามกลุ่มคือ ปกติ (Normal), ผู้ป่วยพาร์กินสัน (Parkinson), ผู้ป่วยข้อเข่าเสื่อม (Knee OA) และใช้ค่าความแม่นยำ (Accuracy) และ  $F_1$  ในการหาประสิทธิภาพของเครื่องมือและสภาพแวดล้อมที่เหมาะสมกับระบบ

สำหรับการทดลองจะใช้ข้อมูลที่เป็นต้นฉบับและข้อมูลที่เติมค่าที่หายไป มีฟีเจอร์สามชุดคือ Kinematic, Kinematic with temporal และ Temporal โดยใช้ SVM และต้นไม้ตัดสินใจเป็นเครื่องในการจำแนกประเภทในโครงสร้างแบบ Flat ในการทดลองแรกใช้กลุ่มตัวอย่างที่คัดเลือกตามเงื่อนไขทั้งหมด 54 คน ได้ผลความแม่นยำเป็น 46% ซึ่งผลยังไม่ค่อยดีนัก อาจเนื่องมาจากใช้จำนวนกลุ่มตัวอย่างที่น้อยเกินไป จึงมีการปรับปรุงแก้ไขข้อมูลและคัดเลือกกลุ่มตัวอย่างเพิ่มเป็น 88 คน ซึ่งการเพิ่มจำนวนคนในกลุ่มตัวอย่างทำให้ได้ผลความแม่นยำมากยิ่งขึ้นเป็น 57% ในการทดลองนี้ทำให้สรุปได้ว่าจำนวนกลุ่มตัวอย่างที่มากขึ้นทำให้ประสิทธิภาพดียิ่งขึ้น เครื่องมือจำแนกประเภท แบบ SVM ให้ผลการทดลองที่ดีกว่าต้นไม้ตัดสินใจ และการเติมข้อมูลที่ขาดหายไปนั้นไม่เพิ่มประสิทธิภาพในการทดลองอาจเนื่องมาจากอัลกอริทึมยังไม่ถูกตามแนวโน้มของข้อมูลจริง ซึ่งในการทดลองถัดไปจึงเลือกใช้ชุดข้อมูลต้นฉบับ 88 คน และใช้เครื่องมือ SVM ในการจำแนกข้อมูล

เนื่องจากในการทดลองที่ผ่านมาในแต่ละค่ามีนัยยะที่ต่างกันจึงมีการเพิ่มการปรับช่วงข้อมูล (Data Normalization) และอาจเนื่องมาจากฟีเจอร์มีปริมาณที่มากจึงเพิ่มวิธีการวิเคราะห์องค์ประกอบ (PCA) มาช่วยในการลดจำนวนฟีเจอร์ โดยผลลัพธ์ที่ได้จากการใช้การปรับช่วงข้อมูล และใช้วิธีวิเคราะห์องค์ประกอบส่งผลช่วยเพิ่มประสิทธิภาพในฟีเจอร์ Kinematic และ Kinematic with temporal มากยิ่งขึ้น ซึ่งสภาพแวดล้อมที่ให้ความแม่นยำมากที่สุดในการทดลองนี้คือข้อมูลต้นฉบับฟีเจอร์ Kinematic มีการปรับช่วงข้อมูล ใช้เครื่องมือ SVM และโครงสร้างแบบ Flat ซึ่งมีค่าเป็น 61%

การทดลองต่อมาจะลดจำนวนตัวแปรในฟีเจอร์ Kinematic โดยเลือกตำแหน่ง 4 จุดที่สำคัญคือจุด Heel Strike, Toe Off, Mid Stance และ Mid Swing ซึ่งฟีเจอร์กลุ่มนี้ชื่อว่า

SelectKinematic ซึ่งผลการทดลองได้ค่าความแม่นยำมากขึ้นเป็น 66% ในสภาพแวดล้อมข้อมูล ต้นฉบับพีเจอร์ SelectKinematic ที่มีการปรับช่วงข้อมูล (Data Normalization) โดย SVM โครงสร้างแบบ Flat ซึ่งต่อมาได้นำข้อมูลชุดนี้ไปทดสอบในโครงสร้างแบบอื่น โดยโครงสร้างแบบ Hier ได้ความแม่นยำเป็น 69% และแบบ One-VS-One ค่าความแม่นยำเป็น 70% ซึ่งเป็นค่าความแม่นยำที่สูงสุดโดยมีสภาพแวดล้อมเป็นข้อมูลต้นฉบับที่ผ่านการปรับช่วงข้อมูลและใช้เฉพาะพีเจอร์ SelectKinematic ผ่านเครื่องมือจำแนกคือ SVM-Linear และใช้โครงสร้างแบบ One VS One

สำหรับการทดลองเรื่องการเติมข้อมูลที่ขาดหายไปนั้นให้ผลไม่แตกต่างกับข้อมูลต้นฉบับ อาจเนื่องมาจากอัลกอริทึมที่ใช้ยังไม่สอดคล้องกับข้อมูลจริงและค่าที่เติมนั้นอยู่ในช่วงที่ไม่แตกต่างกับค่าศูนย์มากนัก ทำให้การเติมข้อมูลด้วยอัลกอริทึมนี้จึงไม่ส่งผลต่อค่าความแม่นยำเท่าที่ควร

การเลือกใช้เฉพาะจุดที่สำคัญทั้งสิ้นช่วยให้อรรถภาพดีขึ้นอาจเนื่องมาจากว่าเป็นการลดจำนวนพีเจอร์ที่ไม่จำเป็นและเป็นการตัดช่วงข้อมูลที่อาจเกิดสัญญาณรบกวน (noise) หรือมีการแกว่งของข้อมูลมาก ตำแหน่งทั้งสี่เป็นตำแหน่งที่ทางแพทย์ได้ใช้วิเคราะห์ข้อมูล เป็นการให้ระบบเลียนแบบการวิเคราะห์ของมนุษย์อีกด้วย

ส่วนโปรแกรมต้นแบบของระบบวินิจฉัยมีสองส่วนหลัก ๆ คือส่วนแสดงผลข้อมูลและส่วนวิเคราะห์ทำนายผล ในส่วนแรกจะแสดงผลข้อมูล Kinematic ซึ่งเป็นแสดงผลเป็นกราฟ และแสดงข้อมูล Distance temporal มีการแสดงอัตราส่วนของขาซ้ายและขวา รวมถึงในส่วนนี้จะมีการแสดงขอบเขตค่าคนปกติของผู้สูงอายุ เพื่อให้ผู้ใช้งานสามารถอ่านค่าจริงและนำมาเปรียบเทียบเพื่อใช้ในการประเมินผลเบื้องต้นได้ ในส่วนวิเคราะห์ทำนายผลโปรแกรมจะทำการคำนวณโดยใช้สภาพแวดล้อมตามการทดลองที่ผ่านมาและส่งคำตอบการทำนายกลับมาว่ามีแนวโน้มเป็น ปกติ (Normal), พาร์กินสัน (Parkinson) หรือ ข้อเข่าเสื่อม (Knee OA) ซึ่งส่วนนี้เมื่อนำไปประกอบกับการแสดงผลข้อมูลในส่วนแรกจะเป็นตัวช่วยในการวินิจฉัยเบื้องต้นสำหรับผู้ใช้งาน

### ข้อเสนอแนะ

ทางด้านจำนวนข้อมูล จำนวนและความสมบูรณ์ข้อมูลมีผลสำคัญต่อความแม่นยำของระบบ หากข้อมูลมีความผิดพลาดน้อย และมีความสมบูรณ์มากจะส่งผลให้เครื่องมือในการจำแนก

ประเภทมีความแม่นยำมากยิ่งขึ้น ในการวินิจฉัยโรคตามทางการแพทย์นั้นใช้ปัจจัยหลายอย่างในการวิเคราะห์ด้วย เช่น อายุ เพศ น้ำหนัก ฯลฯ หากสามารถนำปัจจัยเหล่านั้นมาใช้ในระบบ น่าจะสามารถช่วยให้ระบบมีประสิทธิภาพดียิ่งขึ้น

ทางด้านเครื่องมือในการจำแนกประเภท จากการทดลองจะเห็นได้ว่า SVM-Linear ให้ประสิทธิภาพดีกว่า Decision Tree ซึ่งในงานวิจัยนี้ได้ตั้งค่าตัวแปรต่าง ๆ ของ SVM โดยใช้ค่าตั้งต้น ซึ่งหากมีการปรับจนค่าให้เหมาะสม คาดว่าจะช่วยให้มีประสิทธิภาพมากยิ่งขึ้น หรือการเลือกใช้เครื่องตัวอื่น เช่น SVM-Polynomial ที่มีการปรับจนตัวแปรให้เหมาะสมจะสามารถช่วยในประสิทธิภาพต่อระบบมากยิ่งขึ้น

ทางด้านตัวแปรและฟีเจอร์ ในงานวิจัยนี้ได้เลือกใช้ข้อมูลทุกข้อทั้งสี่ข้อที่เกี่ยวกับการเดิน และใช้ข้อมูลมาจากทั้งสามแกน ซึ่งอาจทำให้มีข้อมูลที่ไม่จำเป็นรวมอยู่ด้วย ในการพัฒนาด้านประสิทธิภาพต่อไป ควรจะเลือกใช้เฉพาะบางข้อในบางแกนที่มีนัยยะสำคัญต่อโรคที่เกี่ยวกับเดิน เช่น ใช้ค่าข้อเท้าในแกน Sagittal , ค่าของข้อเท้าในแกน Sagittal และ Coronal มาใช้เป็นฟีเจอร์ในการทดลองต่อไป

## รายการอ้างอิง

- [1] Richards, J. Biomechanics in Clinic and Research. London, UK : Churchill Livingstone Elsever, 2008.
- [2] Benedetti, M.G., Catani, F., Leardini, A., Pignotti, E., Giannini, S.. Data management in gait analysis for clinical applications. Clinical Biomechanics. (1998) : 204-215.
- [3] Kirtley, C. Clinical Gait Analysis: Theory And Practice. London, UK : Churchill Livingstone Elsever, 2006.
- [4] Lafuente, R., Belda, J.M., Sánchez-Lacuesta, J., Soler, C., Prat, J. Design and test of neural networks and statistical classifiers in computer-aided movement analysis: a case study on gait analysis. Clinical Biomechanics. (1998) : 216-229.
- [5] McBride, J., Zhang, S., Wortley, M., Paquette, M., Klipple, G., Byrd, E., Baumgartner, L., Zhao, X. Neural network analysis of gait biomechanical data for classification of knee osteoarthritis. Biomedical Sciences and Engineering Conference (BSEC), 1-4. 2011.
- [6] Koktas, N.S., Yalabik, N., Yavuzer, G. Ensemble Classifiers for Medical Diagnosis of Knee Osteoarthritis Using Gait Data. 5th International Conference on Machine Learning and Applications (ICMLA '06), pp. 225-230. 2006.
- [7] Manap, H.H., Tahir, N.M., Yassin, A.I.M. Statistical analysis of parkinson disease gait classification using Artificial Neural Network. 2011 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pp.060-065. 2011.
- [8] Paranya, A. and Phayung, M.. A comparative Efficiency of Neural Network Classification for the Diagnosis of Heatstork. The 6TH National Conference on Computing and Information Technology (NCCIT2010). 2010.

- [9] Inman, V.T., Ralston, H.J., and Todd, F., Human Walking, p.26. Baltimore: Williams & Wilkins, 1981.
- [10] Racic, V., Pavic, A., Brownjohn, J.M.W. Experimental identification and analytical modelling of human walking forces: Literature review. Journal of Sound and Vibration, Vol. 326. 2009 : 1-49.
- [11] LoStrangolatore. Human anatomy planes [Online]. 2012. Available from: [http://en.wikipedia.org/wiki/File:Human\\_anatomy\\_planes.svg](http://en.wikipedia.org/wiki/File:Human_anatomy_planes.svg) [2013, Jan 05]
- [12] Newman, C.J., Walsh, M., O'Sullivan, R., Jenkinson, A., Bennett, D., Lynch, B., O'Brien, T.. The characteristics of gait in Charcot-Marie-Tooth disease types I and II. Gait & Posture, Vol. 26. 2007 : 120-127.
- [13] American Academy of Orthopaedic Surgeons. OrthoInfo [Online]. 2012. Available from: <http://orthoinfo.aaos.org/topic.cfm?topic=a00212> [2013, Jan 15]
- [14] mdCurrent-India staff. Protection against Parkinson disease could be an added benefit of statins [Online]. 2012. Available from: <http://mdcurrent.in/pharmacology/protection-against-parkinson-disease-could-be-an-added-benefit-of-statins/> [2013, Jan 15]
- [15] Mahlon, R.D. Parkinson's Disease — The Dana Guide [Online]. 2009. Available from: <https://www.dana.org/news/brainhealth/detail.aspx?id=9860> [2013, April 18]
- [16] Quinlan, J.R. C4.5: Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning). San Francisco, USA : Morgan Kaufmann Publishers Inc., 1993.
- [17] Payam, R., Lei, T. and Huan, L. Cross Validation. In Tamer, A.M. and Ling, L. (ed.), Encyclopedia of Database Systems. Springer, 2009.



- [18] Hill, T. and Lewicki, P. Support Vector Machine (SVM). In STATISTICS: Methods and Applications. Tulsa, OK : StatSoft. 2007.
- [19] Qualisys AB. Qualisys Motion Capture System [Online]. 2012. Available from: <http://www.qualisys.com> [2012, Dec 20]
- [20] ศูนย์ความเป็นเลิศทางการแพทย์ผู้สูงวัยสุขภาพดี โรงพยาบาลจุฬาลงกรณ์. คู่มือการใช้เครื่องวิเคราะห์การเคลื่อนไหว (Motion Analysis Instruction Manual). 2012.
- [21] Cristianini, N. and Shawe-Taylor, J. An Introduction to Support Vector Machines. Cambridge University Press, Cambridge, 2000.
- [22] Chang, C. and Lin, C. LIBSVM -- A Library for Support Vector Machines [Online]. 2012. Available from: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/> [2012, Dec 03]
- [23] Hsu, C., Chang, C. and Lin, C. A Practical Guide to Support Vector Classification. 2010.
- [24] Rokach, L. and Maimon, O. Decision Trees in Data Mining. In The Data Mining and Knowledge Discovery Handbook, pp. 165-192. 2005.
- [25] Heikki, U. and Edgardo, B. Gait Analysis. In Physical Medicine and Rehabilitation Board Review. New York : Demos Medical Publishing, 2004.
- [26] Hugh, H. and Ari, W. User-adaptive control of a magnetorheological prosthetic knee. Industrial Robot: An International Journal. (2003) : 42-42-55.
- [27] Warwick. Engineering » Non GPS Based Personal Navigation System – Gait Lab [Online]. 2013. Available from : <http://www2.warwick.ac.uk/fac/sci/eng/meng/nongps/rnd/gait/> [2013, March 10]
- [28] Breiman, L. Random Forests. In Machine Learning, Vol.45. MA, USA : Kluwer Academic Publishers Hingham, (October 2001) : 5-32.

- [29] Rijsbergen, C.J.V. Information Retrieval. London, UK : Butterworths ,1979.
- [30] Deluzio, K.J., Astephan, J.L. Biomechanical features of gait waveform data associated with knee osteoarthritis. An application of principal component analysis. Gait & Posture, Vol. 25. 2007 : 86-93.
- [31] Oana, G., Ioan, U., Valentin, P., Cornel, O.T., and Nicoleta, C. Gait in Parkinson's disease – signal processing and modeling. 11th International Conference on DEVELOPMENT AND APPLICATION SYSTEMS. pp.166-170. 2012.
- [32] Moshe, G., Erin R., Assaf S., and Ran G. Full Body Gait Analysis with Kinect. In Proceedings of EMBC. Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2012.

## ประวัติผู้เขียนวิทยานิพนธ์

นางสาวรัตติยา สุวานะสุนทรฤกษ์ เกิดเมื่อวันที่ 31 ธันวาคม พ.ศ.2531 ที่จังหวัด นครราชสีมา สำเร็จการศึกษาระดับมัธยมศึกษาจาก โรงเรียนมารีย์วิทยา นครราชสีมา สำเร็จ การศึกษาระดับปริญญาบัณฑิต จากคณะวิศวกรรมศาสตร์ มหาวิทยาลัยขอนแก่น จังหวัดขอนแก่น ในปีการศึกษา 2553