

ระบบสนับสนุนการเลือกยาสำหรับการรักษามะเร็งแบบมุ่งเป้า
ตามเส้นทางการส่งสัญญาณของเซลล์

Drug Selection Support System for Targeted Cancer Therapy Based on Cell
Signal Transduction Pathways

มัศยา สำเนียงงาม¹ และธนันนี เพียรตระกูล²

Massaya Samnienggam¹ and Tanasanee Phienthrakul²

Department of Computer Engineering, Faculty of Engineering, Mahidol University, Thailand

¹f.mas_tsu@hotmail.com, ²tanasanee.phim@mahidol.ac.th

Received: 30 September 2017

Accepted: 20 February 2018

Keywords:

การค้นหาในแนวลึกตาม
ความเป็นจุดศูนย์กลาง,
การรักษามะเร็งแบบมุ่งเป้า,
เส้นทางการส่งสัญญาณ
ของเซลล์

Centrality Depth First
Search, Targeted
Cancer Therapy,
Cell Signal Transduction
Pathways

บทคัดย่อ: งานวิจัยนี้ได้นำเสนอระบบสนับสนุนการเลือกยาสำหรับการรักษามะเร็งแบบมุ่งเป้าตามเส้นทางการส่งสัญญาณของเซลล์ โดยให้นำเทคนิคการสำรวจกราฟ 2 วิธี ได้แก่ การค้นหาในแนวลึก และ การค้นหาในแนวกว้าง มาประยุกต์เข้ากับการวัดความเป็นจุดศูนย์กลาง เพื่อสร้างเป็นเทคนิคที่เรียกว่า การค้นหาในแนวลึกตามความเป็นจุดศูนย์กลาง มาใช้ในการสำรวจเส้นทางการส่งสัญญาณของเซลล์ เพื่อค้นหาข้อยีนที่เกี่ยวข้องกับยีนที่สนใจในเส้นทางการส่งสัญญาณเดียวกัน จากนั้นระบบจะนำข้อยีนเหล่านั้นมาหายาที่สามารถใช้กับยีนเหล่านั้นได้ เพื่อให้ได้ยาที่สามารถใช้รักษามะเร็งนั้นๆ ด้วยวิธีการรักษามะเร็งแบบมุ่งเป้า นอกจากนี้ยังพัฒนาระบบการให้คะแนนยา เพื่อคำนวณน้ำหนักหรือคะแนนให้แก่ยาแต่ละตัว พร้อมทั้งเรียงลำดับตามคะแนนที่ได้ ผลการวิจัยที่ได้แสดงให้เห็นว่า ระบบสนับสนุนการเลือกยาสำหรับการรักษามะเร็งแบบมุ่งเป้าสามารถช่วยลดระยะเวลาในการวิเคราะห์และค้นหาจากเส้นทางการส่งสัญญาณของเซลล์ได้

Abstract: This research proposed drug selection support system for targeted cancer therapy based on cell signal transduction pathways. Two graph traversal algorithms include depth first search (DFS) and breadth first search (BFS) algorithms were brought to apply with centrality measurements. There are called centrality depth first search (CDFS) algorithm. The CDFS was used to traverse in the cell signal transduction pathway for genes that related with the interested gene in the same pathway. Then the system will bring that genes to find drugs that can be used for targeted cancer therapy. Moreover, this research also developed scoring drug system for computing score of that drugs and ranking drugs according to their score. The results shown that drug selection support system for targeted cancer therapy based on signal transduction pathway can decrease time to analyze pathway and find drugs based on cell signal transduction pathway.

1. บทนำ

มะเร็ง คือ กลุ่มของโรคที่เกิดจากความผิดปกติของเซลล์ในร่างกาย ส่งผลให้เซลล์มีการเจริญเติบโตและแบ่งตัวเพิ่มจำนวนเซลล์อย่างรวดเร็วและมากกว่าปกติ (National Cancer Institute, 2017) ซึ่งเป็นอีกโรคหนึ่งที่มีวิวัฒนาการมาพร้อมๆ กับเทคโนโลยี สังเกตได้จากที่ยังเทคโนโลยีเจริญก้าวหน้ามากเท่าไร ประชากรโลกที่เป็นมะเร็งก็ขยายเพิ่มมากขึ้นตามไปด้วย องค์การอนามัยโลก หรือ WHO (2015) ได้รายงานไว้ว่า มะเร็งเป็นสาเหตุการเสียชีวิตของมนุษย์เป็นอันดับสองของโลก ซึ่งบ่งชี้ให้เห็นข้อเท็จจริงที่กล่าวไปข้างต้นว่าความเจริญก้าวหน้าของเทคโนโลยีมีผลกระทบต่อสุขอนามัยของมนุษย์ เนื่องจากอัตราการเป็นมะเร็งของประชากรโลกที่เพิ่มขึ้นนี้เองทำให้เหล่านักวิจัยพยายามคิดค้นและพัฒนาวิธีการรักษามะเร็งให้มีประสิทธิภาพมากขึ้น ซึ่งวิธีการรักษามะเร็งมีหลายวิธี แต่หลักๆ และเป็นที่ยอมรับโดยทั่วไปแล้วมี 3 วิธี ได้แก่ การผ่าตัด (Surgery) การฉายรังสี (Radiation Therapy) และเคมีบำบัด (Chemotherapy) (National Cancer Institute, 2017) โดยการรักษามะเร็งนั้นบางครั้งอาจใช้วิธีการรักษา ร่วมกัน เช่น การใช้วิธีฉายรังสี หรือ เคมีบำบัด เพื่อลดขนาดและหยุดการแพร่กระจายของเซลล์มะเร็งก่อนที่จะทำการผ่าตัดต่อไป ซึ่งเคมีบำบัดเป็นวิธีการรักษาวิธีหนึ่งที่มีประสิทธิภาพและถูกนำมาใช้บ่อยในปัจจุบัน (Bayat Mokhtari R. และคณะ, 2017)

วิธีการรักษาแบบเคมีบำบัด เป็นการให้ยาแก่ผู้ป่วย เพื่อให้ฤทธิ์ของยาส่งผลไปทำลายเซลล์มะเร็งที่มีการแบ่งตัวอย่างรวดเร็ว แต่ยานี้อาจจะไปส่งผลต่อเซลล์ปกติที่มีการแบ่งตัวอย่างรวดเร็วด้วย จึงทำให้เกิดผลข้างเคียงจากการรักษาได้ เช่น ภาวะเม็ดเลือดขาวต่ำ (Neutropenia) โลหิตจาง (Anemia) ภาวะเกร็ดเลือดต่ำ (Thrombocytopenia) ท้องเสีย (Diarrhea) ผอมร่าง และ โรคแผลในปาก เป็นต้น (BNH, 2011) ต่อมาเมื่อเทคโนโลยีเจริญก้าวหน้าจนกระทั่งสามารถศึกษาถึงระดับยีนและโมเลกุลเพื่อตรวจหายีนที่ผิดปกติและการเจริญเติบโตของเซลล์มะเร็งได้แล้ว วิธีการรักษามะเร็งอีกวิธีหนึ่งที่มีชื่อว่า วิธีการรักษามะเร็งแบบมุ่งเป้า (Targeted Cancer Therapy) จึงได้ถือกำเนิดขึ้น วิธีการรักษานี้เป็นการให้ยาแก่ผู้ป่วยเพื่อให้ฤทธิ์ยาส่งผลเฉพาะเจาะจงต่อเซลล์มะเร็งเท่านั้น เพื่อไปยับยั้งกระบวนการส่งสัญญาณ (Signal Transduction Pathway) ของเซลล์มะเร็ง ซึ่งกระบวนการส่งสัญญาณนี้

เป็นสาเหตุหนึ่งที่ทำให้เซลล์มะเร็งมีการเจริญเติบโตและแพร่กระจาย ยาที่ใช้ในการรักษานี้จะไปจับกับเป้าหมายที่อยู่ภายนอกเซลล์หรือบนผิวเซลล์ หรือ เคลื่อนผ่านผนังเซลล์ไปจับกับเป้าหมายภายในเซลล์ได้ และเนื่องจากมะเร็งแต่ละชนิดก็มีเส้นทางการส่งสัญญาณที่แตกต่างกัน ฉะนั้นยาแต่ละตัวก็จะมีกลไกการออกฤทธิ์ที่แตกต่างกันด้วย เช่น ยาที่ยับยั้งกระบวนการส่งสัญญาณที่ไปกระตุ้นการเจริญเติบโตของเซลล์มะเร็ง และ ยาที่ยับยั้งตัวรับ (Receptor) บนผิวเซลล์ เป็นต้น (BNH, 2013)

วิธีการรักษามะเร็งแบบมุ่งเป้ามีความเกี่ยวข้องกับกระบวนการส่งสัญญาณของเซลล์ ซึ่งกระบวนการส่งสัญญาณเกิดขึ้นเมื่อมีการส่งสัญญาณจากภายนอกเซลล์เข้ามาจับกับตัวรับเฉพาะที่อยู่บนผิวเซลล์หรือภายในเซลล์ ตัวรับนี้ก่อให้เกิดการส่งต่อสัญญาณทางชีวเคมีภายในเซลล์เพื่อสร้างการตอบสนองต่างๆ ออกมา ซึ่งการตอบสนองที่เกิดขึ้นนี้จะทำให้เกิดการเปลี่ยนแปลงกระบวนการสร้างและสลายพลังงาน (Metabolism) ของเซลล์ การเปลี่ยนแปลงรูปร่างของเซลล์ การเปลี่ยนแปลงการแสดงออกของยีน (Gene Expression) หรือ การเปลี่ยนแปลงความสามารถในการแบ่งตัวของเซลล์ สัญญาณที่เกิดขึ้นสามารถถูกขยายได้ในทุกๆ ขั้นตอน ดังนั้น การส่งสัญญาณหนึ่งครั้งสามารถทำให้เกิดการตอบสนองได้หลายๆ ครั้ง เช่น การส่งสัญญาณให้เซลล์มีการแบ่งตัว เป็นต้น ซึ่งในการค้นคว้าและพัฒนา ยีนที่ผิดปกติหรือยีนเป้าหมาย (Target Gene) จะถูกใช้เป็นเป้าหมายของยา (Drug Target) หรือ ตัวยับยั้ง (Inhibitor)

วิธีการรักษามะเร็งแบบมุ่งเป้ามักจะใช้ในผู้ป่วยที่เป็นมะเร็งในระยะลุกลาม (Metastatic Cancer) หรือ ใช้เป็นวิธีรักษาเสริมจากวิธีการรักษาหลักอื่นๆ อย่างไรก็ตาม วิธีการรักษานี้มีค่าใช้จ่ายค่อนข้างสูงและไม่สามารถใช้รักษาผู้ป่วยได้ทุกคน เนื่องจากวิธีการรักษานี้จำเป็นต้องทราบชื่อยีนและตำแหน่งของยีนที่ทำให้เกิดมะเร็ง และต้องตรวจสอบยาที่สามารถส่งผลกระทบต่อยีนนั้นได้ ดังนั้น แพทย์จำเป็นต้องนำชิ้นเนื้อส่วนที่เป็นตัวอย่างมาทำการทดสอบหาความผิดปกติ จากนั้นนำผลไปวิเคราะห์หายีนและตำแหน่งของยีนที่ผิดปกติต่อไป เมื่อทราบชื่อยีนที่ผิดปกติแล้วจึงจะสามารถเลือกยามาใช้กับยีนนั้นได้ แต่ยังมีอีกกรณีที่ไม่พบยาที่สามารถใช้กับยีนเป้าหมายได้ หรือ ผู้ป่วยไม่สามารถรับยาที่มีอยู่ได้ แพทย์จำเป็นต้องหามาใช้ทดแทน โดยการวิเคราะห์เส้นทางการส่งสัญญาณของเซลล์เพื่อค้นหาอื่นในเส้น

ทางการส่งสัญญาณเดียวกันที่มีการส่งสัญญาณมาหา ยีนเป้าหมาย ยีนที่ค้นพบเหล่านั้นจะถูกนำมาพิจารณาเป็นเป้าหมายของยาต่อไป ในกรณีนี้อาจมีผลข้างเคียงเกิดขึ้นหากยีนที่เลือกที่จะให้ยานั้นไม่ได้ส่งสัญญาณไปให้แค้ยีนเป้าหมายตัวเดียว ซึ่งงานวิจัยนี้ได้มุ่งความสนใจไปที่กรณีหลังนี้ โดยมีแนวคิดที่จะช่วยลดระยะเวลาในการวิเคราะห์เลือกยามาใช้ในการรักษามะเร็งแบบมุ่งเป้า

เส้นทางการส่งสัญญาณของเซลล์ (Cell Signal Transduction Pathways) คือ เส้นทางที่เมื่อเซลล์ได้รับสัญญาณแล้ว อาจมีการแปลผลส่งสัญญาณต่อ หรือกระตุ้นให้เกิดเหตุการณ์อื่น อาจมีขั้นตอนเดียวหรือสลับซับซ้อนแตกต่างกันไป ซึ่งผลของสัญญาณสามารถส่งไปยังเซลล์ที่อยู่ใกล้เคียง และอาจส่งสัญญาณต่อเป็นทอดๆ ไปยังเซลล์อื่นๆ ได้อีก การส่งสัญญาณของเซลล์นี้มีทั้งแบบกระตุ้นให้เกิดการแบ่งตัวของเซลล์และยับยั้ง การแบ่งตัวของเซลล์ในการวิเคราะห์เส้นทางการส่งสัญญาณจึงต้องพิจารณาถึงความสัมพันธ์ของเซลล์ต่างๆ ที่อยู่บนเส้นทางการส่งสัญญาณนั้น และ ชนิดของสัญญาณที่ส่งด้วย

งานวิจัยนี้จึงได้ออกแบบระบบสนับสนุนการเลือกยาสำหรับการรักษามะเร็งแบบมุ่งเป้าตามเส้นทางการส่งสัญญาณของเซลล์ขึ้น โดยนำข้อมูลเส้นทางการส่งสัญญาณของมะเร็งชนิดต่างๆ จำนวน 14 ชนิด มาจาก KEGG (2015) เพื่อนำมาใช้ค้นหายีนอื่นๆ ในเส้นทางการส่งสัญญาณเดียวกันที่เกี่ยวข้องกับยีนเป้าหมาย โดยใช้เทคนิคการค้นหาในแนวลึกตามความเป็นศูนย์กลาง (Centrality Depth First Search: CDFS) ที่ได้พัฒนาขึ้นในงานวิจัยนี้ และนำข้อมูลที่ได้ไปหาญาติที่เกี่ยวข้องต่อไป นอกจากนี้ งานวิจัยนี้ยังได้นำข้อมูลเกี่ยวกับยามาจาก 2 แหล่งด้วยกัน ได้แก่ ClinicalTrial.gov (2015) ซึ่งเป็นฐานข้อมูลเกี่ยวกับการวิจัยยาต่างๆ โดยยาที่อยู่ในฐานข้อมูลนี้จะเป็นยาที่อยู่ในขั้นตอนการทดลอง (Clinical Trial Drug) ยังไม่สามารถนำมาใช้กับผู้ป่วยได้ และ Drugs@FDA (2015) ซึ่งเป็นฐานข้อมูลเกี่ยวกับยาที่ได้รับการอนุมัติให้ใช้กับผู้ป่วยได้โดยองค์การอนามัยโลกแห่งสหรัฐอเมริกา (WHO) ซึ่งข้อมูลยาเหล่านี้จะถูกนำมาใช้เป็นแหล่งข้อมูลสำหรับการค้นหาจากชื่อยีนที่ค้นพบ และยังถูกนำไปใช้ในการให้คำแนะนำเพื่อจัดลำดับความสำคัญของยาในระบบการให้คำแนะนำที่ได้รับการออกแบบจากงานวิจัยนี้อีกด้วย ซึ่งระบบนี้จะสามารถช่วยเหลือผู้ทำการตัดสินใจในการค้นหาที่เกี่ยวข้องกับ

ยีนและมะเร็งที่สนใจ และยังสนับสนุนผู้ทำการตัดสินใจในการเลือกยาออกมาได้อย่างเหมาะสมอีกด้วย

2. งานวิจัยที่เกี่ยวข้อง

เนื่องจากงานวิจัยนี้วางแผนที่จะค้นหาเป้าหมายของยาและยาจากเส้นทางการส่งสัญญาณของเซลล์ ดังนั้นงานวิจัยนี้จึงได้ทำการศึกษางานวิจัยอื่นๆ ที่เกี่ยวข้องกับเทคนิคการสำรวจกราฟ เช่น เทคนิคการสำรวจแบบสุ่ม (Random Walking Algorithm) และ เทคนิคไดจิสตรา (Dijkstra's Algorithm) เป็นต้น อย่างไรก็ตาม งานวิจัยนี้ยังได้ทำการศึกษาเทคนิคการสำรวจกราฟพื้นฐานอย่างเทคนิคการค้นหาในแนวลึก (Depth First Search Algorithm) และเทคนิคการค้นหาในแนวกว้าง (Breadth First Search Algorithm) นอกจากนี้ งานวิจัยนี้ยังได้ศึกษาการวัดความเป็นศูนย์กลางด้วย เพื่อช่วยในการแนะนำเส้นทางที่เหมาะสมในเส้นทางการส่งสัญญาณ ได้แก่ ความเป็นศูนย์กลางโดยวัดจากระดับ (Degree Centrality) ความเป็นศูนย์กลางโดยวัดจากความใกล้ชิด (Closeness Centrality) และความเป็นศูนย์กลางโดยวัดจากการคั่นกลาง (Betweenness Centrality) ยิ่งกว่านั้นยังทำให้ค้นพบแหล่งข้อมูลที่น่าสนใจจากขั้นตอนการศึกษางานวิจัยที่เกี่ยวข้องนี้ด้วย

จากการศึกษางานวิจัยที่เกี่ยวข้องพบว่า มีการค้นคว้าและพัฒนาการค้นหาในแนวทางที่หลากหลายแตกต่างกันออกไป โดยสามารถจำแนกออกเป็น 2 ส่วนหลักๆ ได้แก่ กลุ่มงานวิจัยที่ค้นหาโดยใช้การทำเหมืองข้อมูลเข้ามาช่วย และ กลุ่มงานวิจัยที่มีการพิจารณาเส้นทางการส่งสัญญาณของเซลล์ร่วมด้วย ซึ่งในช่วงแรกๆ นั้น มักมีการใช้การทำเหมืองข้อมูลเข้ามาช่วยในการค้นหา เช่น ในงานของ Pankaj และ David (2008), Yang และคณะ (2009), และ Taysir และคณะ (2012) เป็นต้น งานวิจัยเหล่านี้แสดงให้เห็นว่าวิธีการทำเหมืองข้อมูลเป็นวิธีการแรกๆ ที่ถูกนำมาใช้ในการค้นหาอย่างต่อเนื่องเมื่อเทคโนโลยีก้าวหน้าขึ้น ประกอบกับมีวิธีการรักษามะเร็งแบบใหม่เพิ่มขึ้น จึงทำให้การค้นหาหรือการศึกษาความเกี่ยวข้องของโรคกับยีนไม่ได้จำกัดเฉพาะแค่การทำเหมืองข้อมูลอีกต่อไป การศึกษาความสัมพันธ์ที่เกิดขึ้นในเส้นทางการส่งสัญญาณของเซลล์จึงเริ่มเป็นที่สนใจมากขึ้น มีหลายงานวิจัยที่ศึกษาเกี่ยวกับเส้นทางการส่งสัญญาณของเซลล์ และยังมีการใช้เทคนิค

ในการสำรวจกราฟเข้ามาช่วยวิเคราะห์เส้นทางการส่งสัญญาณ เช่น Meng-Hsiun และคณะ (2010) ได้ประยุกต์ใช้ยีนในการจำแนกเส้นทางการเดินด้วยวิธีการ Dijkstra ร่วมกับทฤษฎีของเบย์ งานวิจัยดังกล่าวสามารถวิเคราะห์เส้นทางการความสัมพันธ์ของเซลล์มะเร็งที่ซับซ้อนได้ แต่ยังคงต้องการการตรวจสอบจากผู้เชี่ยวชาญ เพื่อให้โมเดลการทำนายนี้มีความน่าเชื่อถือมากยิ่งขึ้น

Maya และ Michael (2012) ได้ใช้วิธีการเดินแบบสุ่ม (Random walk) บนกราฟ เพื่อแก้ไขปัญหาความสมบูรณ์ของเส้นทางการเดินกับเครือข่ายปฏิสัมพันธ์ระหว่างโปรตีน พวกเขาใช้เครือข่ายปฏิสัมพันธ์ระหว่างโปรตีนของยีสต์ และเส้นทางการเดินของยีสต์จำนวน 2 เส้นทาง จากนั้นจึงเปรียบเทียบความถูกต้องของวิธีที่นำเสนอกับงานอื่นๆ ที่ใช้วิธีการเดินแบบสุ่มเช่นเดียวกัน ผลที่ได้แสดงให้เห็นว่า วิธีที่พวกเขาเสนอได้ความถูกต้องที่คล้ายกันหรือดีกว่างานอื่นที่นำมาเปรียบเทียบ Deanna และ Neil (2012) ได้ทำการสำรวจความสัมพันธ์ระหว่างเครือข่ายของยีนและความเกี่ยวข้องทางมะเร็งวิทยา จากนั้นประเมินการทำนายของตัววัดความเป็นศูนย์กลางและสัมประสิทธิ์การจัดกลุ่มโดยใช้โมเดลเชิงเส้น และประยุกต์วิธีการเดินแบบสุ่มในการค้นหากลุ่มยีนที่เกี่ยวข้องกับมะเร็ง ผลที่ได้แสดงให้เห็นว่า ยีนมะเร็งในเครือข่ายเมตาบอลิกและเครือข่ายการส่งสัญญาณมีความแตกต่างกันอย่างมีนัยสำคัญ ซึ่งทำให้สามารถจัดลำดับความสำคัญของยีนที่ทำให้เกิดโรค หรือ ทำนายสถานะของโรคได้

3. ขั้นตอนการวิจัย

งานวิจัยนี้ได้แบ่งขั้นตอนการทำวิจัยออกเป็น 6 ขั้นตอน ได้แก่ การศึกษางานวิจัยที่เกี่ยวข้อง การคัดเลือกข้อมูลที่ใช้สำหรับการวิจัย การเตรียมข้อมูลสำหรับการวิจัย การออกแบบระบบและอัลกอริทึม การพัฒนาและทดสอบระบบ และ การสรุปและประเมินผลการวิจัย ดังภาพประกอบ 1



ภาพประกอบที่ 1 ขั้นตอนการวิจัย

3.1 การรวบรวมข้อมูล (Data Collection)

งานวิจัยนี้ได้เลือกข้อมูลจากแหล่งข้อมูลที่อนุญาตให้ดาวน์โหลดข้อมูลมาใช้งานได้ และเป็นข้อมูลที่มีประโยชน์ในการวิเคราะห์เส้นทางการส่งสัญญาณ โดยมีผู้เชี่ยวชาญทางด้านมะเร็งแนะนำ ซึ่งข้อมูลที่จำเป็นสำหรับการวิจัย ได้แก่ ข้อมูลยาที่ได้รับการอนุมัติให้ใช้ในคนได้ ข้อมูลยาที่อยู่ในขั้นตอนการทดลอง และข้อมูลเส้นทางการส่งสัญญาณของเซลล์มะเร็ง ดังนั้น งานวิจัยจึงได้ทำการคัดเลือกแหล่งข้อมูลที่มีความน่าเชื่อถือและเหมาะสมต่อการนำมาใช้ในการวิจัย ได้แก่ Drugs@FDA (2015) สำหรับข้อมูลยาที่ได้รับการอนุมัติให้ใช้ในคนได้จาก FDA จำนวน 32,407 ตัว ซึ่งอยู่ในรูปแบบ Tab Delimiter Text File Format, ClinicalTrials.gov (2015) สำหรับข้อมูลยาที่อยู่ในขั้นตอนการทดลอง จำนวน 22,668 ตัว ซึ่งอยู่ในรูปแบบ Pipe Delimiter Text File Format และ KEGG (2015) สำหรับเส้นทางการส่งสัญญาณของเซลล์มะเร็งจำนวน 14 เส้นทางการส่งสัญญาณ ซึ่งอยู่ในรูป XML File Format ตัวอย่างข้อมูลทั้งสามส่วนแสดงดังภาพประกอบที่ 2, 3 และ 4 ตามลำดับ

3.2 การเตรียมข้อมูล (Data Preparation)

เนื่องจากข้อมูลที่งานวิจัยนี้คัดเลือกมานั้นมีรูปแบบข้อมูลที่แตกต่างกัน ได้แก่ Drugs@FDA มีรูปแบบข้อมูลแบบคั่นด้วยแท็บ (tab-delimiter) ClinicalTrials.gov มีรูปแบบข้อมูลแบบคั่นด้วยไปป์ (pipe-delimiter) และ KEGG มีรูปแบบข้อมูลแบบ XML (Extensible Markup Language) จึงต้องนำข้อมูลเหล่านั้นมาแปลงให้อยู่ในรูปแบบที่สามารถนำมาวิเคราะห์ได้สะดวก โดยการแปลงให้อยู่ในรูปตารางความสัมพันธ์บนฐานข้อมูลเชิงสัมพันธ์ (Relational Database) และยังได้ทำความสะอาดข้อมูล (cleansing data) จัดการกับข้อมูลที่ไม่ถูกต้อง ไม่สมบูรณ์ และไม่สอดคล้องกัน หากข้อมูลแถวใดไม่ครบ (missing data) จะทำการตัดออกเนื่องจากข้อมูลประเภทนี้ค่อนข้างละเอียดอ่อน

3.3 การออกแบบระบบและเทคนิคการสำรวจเส้นทางการส่งสัญญาณ (System and Algorithm Design)

ในขั้นตอนนี้ได้แบ่งการออกแบบออกเป็น 5 ส่วน ได้แก่ การออกแบบฐานข้อมูล การออกแบบระบบสนับสนุนการเลือกยา และการออกแบบระบบการให้คำแนะนำ ในส่วนของการออกแบบฐานข้อมูล ได้ทำการออกแบบฐาน

1	000004	004	SOLUTION/DROPS;OPHTHALMIC	1%	3	0	PAREDINE	HYDROXYAMPHETAMINE	HYDROBROMIDE
2	000159	001	TABLET;ORAL	500MG	3	0	SULFAPYRIDINE	SULFAPYRIDINE	
3	000552	001	INJECTABLE;INJECTION	20,000 UNITS/ML	3	0		LIQUAEMIN SODIUM	HEPARIN SODIUM
4	000552	002	INJECTABLE;INJECTION	40,000 UNITS/ML	3	0		LIQUAEMIN SODIUM	HEPARIN SODIUM
5	000552	003	INJECTABLE;INJECTION	5,000 UNITS/ML	3	0		LIQUAEMIN SODIUM	HEPARIN SODIUM

ภาพประกอบที่ 2 ตัวอย่างข้อมูลยาที่ได้รับการอนุมัติให้ใช้ในคนได้จาก FDA

```

1 INTERVENTION_ID|NCT_ID|INTERVENTION_TYPE|INTERVENTION_NAME|DESCRIPTION
2 1|NCT00000102|Drug|Nifedipine|
3 2|NCT00000110|Drug|dietary fat|
4 3|NCT00000176|Drug|Estrogen|
5 4|NCT00000681|Drug|Sargramostim|
6 5|NCT00005780|Drug|autologous tumor cell vaccine|N/A

```

ภาพประกอบที่ 3 ตัวอย่างข้อมูลยาที่อยู่ในขั้นตอนการทดลองจาก ClinicalTrials.gov

```

1 <?xml version="1.0"?>
2 <!DOCTYPE pathway SYSTEM "http://www.kegg.jp/kegg/xml/KGML_v0.7.1_.dtd">
3 <!-- Creation date: Oct 2, 2012 14:04:48 +0900 (GMT+09:00) -->
4 <pathway name="path:hsa05210" org="hsa" number="05210"
5     title="Colorectal cancer"
6     image="http://www.kegg.jp/kegg/pathway/hsa/hsa05210.png"
7     link="http://www.kegg.jp/kegg-bin/show_pathway?hsa05210">
8   <entry id="1" name="hsa:1630" type="gene"
9     link="http://www.kegg.jp/dbget-bin/www_bget?hsa:1630">
10   <graphics name="DCC, CRC18, CRCR1, IGDC1, MRMV1, NTN1R1" fgcolor="#000000" bgcolor="#BFFFBF"
11     type="rectangle" x="276" y="520" width="46" height="17"/>
12 </entry>

```

ภาพประกอบที่ 4 ตัวอย่างข้อมูลเส้นทางการส่งสัญญาณของเซลล์มะเร็งจาก KEGG

ข้อมูลเพื่อให้สามารถจัดการข้อมูลได้สะดวก เช่น การเพิ่ม การลบ หรือ การแก้ไขข้อมูล เป็นต้น ซึ่งตารางหลักๆ ที่มีในฐานข้อมูล ได้แก่ ตารางข้อมูลจาก Drugs@FDA, ClinicalTrials.gov และ KEGG ฐานข้อมูลส่วนหลักแสดง ดังภาพประกอบที่ 5

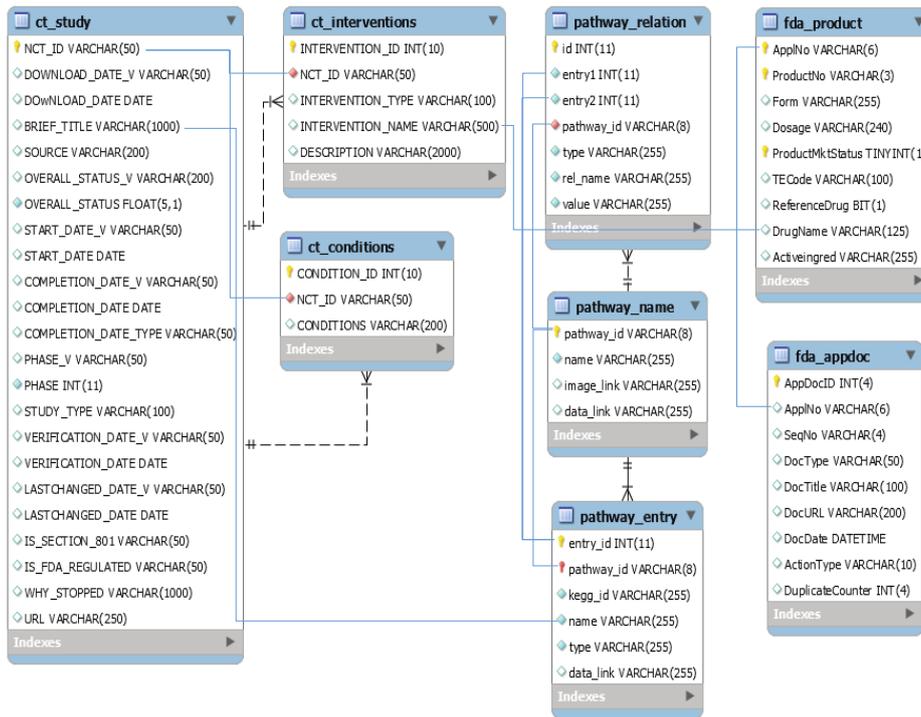
สำหรับในส่วนของระบบสนับสนุนการเลือกยา ประกอบด้วย การสำรวจเส้นทางการส่งสัญญาณของ เซลล์ และ ระบบการให้คำแนะนำ ซึ่งระบบการสนับสนุน ยาถือเป็นหัวใจหลักของการวิจัยนี้ โดยมีรายละเอียดการ ออกแบบในหัวข้อที่ 4

3.4 การพัฒนาและทดสอบระบบ (Implementation and Testing)

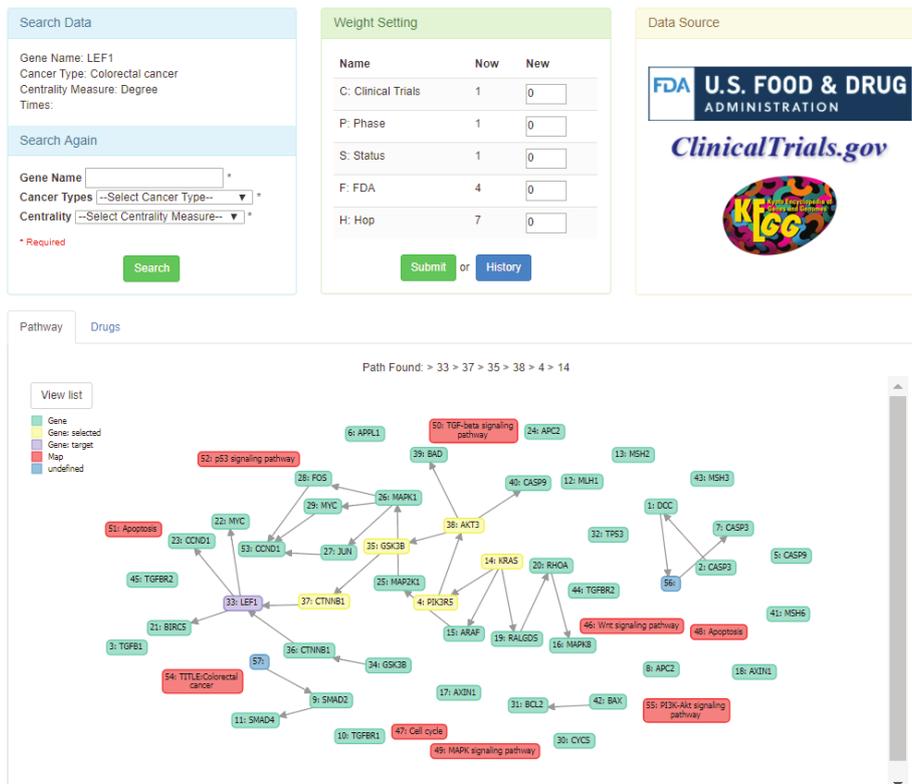
ในขั้นตอนนี้จะเป็นขั้นตอนการพัฒนา ระบบ โดยเริ่มจากการสร้างฐานข้อมูลตามที่ได้ออกแบบไว้ จากนั้นทำการแปลงข้อมูลที่ได้เลือกไว้เข้าฐานข้อมูล ส่วน ค่าความเป็นศูนย์กลางนั้น งานวิจัยนี้ได้ นำ MATLAB มาช่วยในการคิดค่าความเป็นศูนย์กลางออกมาแล้วใส่

ผลที่ได้ลงในฐานข้อมูลที่เตรียมไว้ ส่วนเทคนิคในการ สำรองเส้นทางการส่งสัญญาณและระบบให้คำแนะนำนั้น งานวิจัยนี้ได้พัฒนาโดยใช้ภาษา PHP และการทดสอบ จะแบ่งออกเป็น 2 ส่วน คือ การเปรียบเทียบผลคะแนน ยาตามการวัดความเป็นศูนย์กลาง และการเปรียบเทียบ ผลคะแนนยาตามเทคนิคการสำรวจกราฟ

การเปรียบเทียบผลคะแนนยาตามการวัดความเป็น ศูนย์กลาง ได้ทำการทดสอบ 3 แบบ ได้แก่ ความเป็น ศูนย์กลางโดยวัดจากระดับ (Degree Centrality) ความเป็น ศูนย์กลางโดยวัดจากความใกล้ชิด (Closeness Centrality) และความเป็นศูนย์กลางโดยวัดจากการ คั่นกลาง (Betweenness Centrality) เพื่อหาการวัดค่า ความเป็นศูนย์กลางที่เหมาะสมกับเทคนิค CDFS มาก ที่สุด และในส่วนของการเปรียบเทียบผลคะแนนยาตาม เทคนิคการสำรวจกราฟ มี 3 เทคนิค ได้แก่ เทคนิค การค้นหาในแนวลึก (Depth First Search Algorithm) เทคนิคการค้นหาในแนวกว้าง (Breadth First Search Algorithm) และ เทคนิคการค้นหาในแนวลึกตามความ



ภาพประกอบที่ 5 ตัวอย่างโครงสร้างฐานข้อมูลแสดงความสัมพันธ์ระหว่างตาราง



ภาพประกอบที่ 6 ตัวอย่างหน้าระบบสนับสนุนการเลือกยาสำหรับการรักษามะเร็งแบบมุ่งเป้าตามเส้นทางการส่งสัญญาณของเซลล์

เป็นจุดศูนย์กลาง (Centrality Depth First Search Algorithm) เพื่อหาเทคนิคที่เหมาะสมกับระบบการเลือกยาสำหรับการรักษามะเร็งแบบมุ่งเป้าตามเส้นทางการส่งสัญญาณของเซลล์มากที่สุด ภาพประกอบที่ 6 แสดงตัวอย่างระบบที่ได้ทำการพัฒนาขึ้น

3.5 การสรุปและประเมินผลการวิจัย (Conclusion and Evaluation)

ผลที่ได้จากการเปรียบเทียบทั้งสองส่วนที่กล่าวมานั้น จะถูกนำมาประเมินผลว่าการวัดค่าความเป็นศูนย์กลางแบบใดเหมาะสมกับเทคนิค CDFS มากที่สุด และ เทคนิคการสำรวจเส้นทางการส่งสัญญาณแบบใดที่เหมาะสมกับระบบสนับสนุนการเลือกยามากที่สุด นอกจากนี้ยังต้องพิจารณาถึงปัญหาต่างๆ ที่เกิดขึ้นพร้อมให้แนวทางการแก้ปัญหา และแนวทางการพัฒนาในอนาคต

4. ระบบสนับสนุนการเลือกยา (Drug Selection Support System)

ระบบสนับสนุนการเลือกยา เป็นระบบที่ประกอบด้วยอัลกอริทึมการค้นหาในแนวลึกตามความเป็นศูนย์กลาง (CDFS) และ ระบบการให้คำแนะนำ ระบบนี้ถูกออกแบบเพื่อสนับสนุนผู้ทำการตัดสินใจที่ต้องการค้นหาสำหรับการรักษามะเร็งแบบมุ่งเป้าตามเส้นทางการส่งสัญญาณของเซลล์ กระบวนการทำงานของระบบสามารถแสดงได้ดังภาพประกอบที่ 7

ระบบเริ่มต้นด้วยการรับคำค้นเข้ามา 3 ตัว ได้แก่ ชื่อยีน ชนิดของมะเร็ง และ ตัววัดความเป็นศูนย์กลาง จากนั้นคำค้นเหล่านั้นจะถูกนำไปหาอินและยาที่เกี่ยวข้องด้วยอัลกอริทึม CDFS แล้วยาที่ได้นี้จะถูกนำไปคิดคะแนนเพื่อจัดลำดับด้วยระบบการให้คะแนนยาต่อไป สุดท้ายแล้วจะได้ยาที่เกี่ยวข้องพร้อมคะแนนของยาออกมา ระบบการให้คำแนะนำมีการกำหนดน้ำหนักของคะแนนในแต่ละส่วน ซึ่งหากผู้ใช้ไม่พอใจผลที่ได้จากการกำหนดน้ำหนักพื้นฐาน ผู้ใช้สามารถกำหนดได้

เองในเว็บแอปพลิเคชัน โดยรายละเอียดของอัลกอริทึม CDFS และ ระบบการให้คำแนะนำสามารถอธิบายได้ดังต่อไปนี้

4.1 การค้นหาในแนวลึกตามความเป็นจุดศูนย์กลาง (Centrality Depth First Search: CDFS)

การค้นหาในแนวลึกตามความเป็นจุดศูนย์กลาง หรือ CDFS เป็นเทคนิคที่ถูกพัฒนาขึ้นในงานวิจัยนี้ เพื่อใช้ในการสำรวจเส้นทางการส่งสัญญาณของเซลล์มะเร็ง อัลกอริทึมนี้เกิดจากการประยุกต์เทคนิคการค้นหาในแนวลึก (Depth First Search) และ การค้นหาในแนวกว้าง (Breadth First Search) เข้ากับการวัดความเป็นศูนย์กลาง (Centrality Measurements) งานวิจัยนี้จะมองเส้นทางการส่งสัญญาณของเซลล์เป็นกราฟแบบมีทิศทาง (Directed Graph) ยีนในเส้นทางการส่งสัญญาณของเซลล์เป็นโหนด (Node) และ ลักษณะความสัมพันธ์ระหว่างยีนเป็นขอบ (Edge) ค่าที่ได้จากการวัดความเป็นศูนย์กลางจะช่วยนำทางให้กับการสำรวจกราฟหรือเส้นทางการส่งสัญญาณของเซลล์จากยีนเป้าหมาย หรือ โหนดเริ่มต้น (Source Node) ไปหาโหนดพ่อแม่ (Parent Node) ที่มีค่าความเป็นศูนย์กลางมากที่สุด หรือ มีความสำคัญมากที่สุดในระดับเดียวกัน

โหนดที่มีค่าความเป็นศูนย์กลางมากที่สุดนั้น ผู้วิจัยให้สมมติฐานว่าน่าจะมียาที่สามารถใช้รักษามะเร็งได้ หรือ สามารถนำทางไปหาโหนดอื่นๆ ที่เป็นโหนดที่มีความสำคัญได้ โดยตัววัดความเป็นศูนย์กลางที่นำมาพิจารณาในงานวิจัยนี้ได้แก่ ความเป็นศูนย์กลางโดยวัดจากระดับ (Degree Centrality) ความเป็นศูนย์กลางโดยวัดจากความใกล้ชิด (Closeness Centrality) และ ความเป็นศูนย์กลางโดยวัดจากการคั่นกลาง (Betweenness Centrality) ซึ่งในขั้นตอนการวัดผล การวัดความเป็นศูนย์กลางทั้งสามตัวนี้จะถูกนำมาเปรียบเทียบเพื่อหาตัววัดความเป็นศูนย์กลางที่เหมาะสมที่สุดสำหรับการพัฒนาเทคนิคการค้นหาในแนวลึกตามความเป็นจุดศูนย์กลาง (CDFS) จากนั้น เทคนิค CDFS กับตัววัดความเป็นศูนย์กลางที่เหมาะสมที่สุดจะถูกนำไปเปรียบเทียบกับ



ภาพประกอบที่ 7 กระบวนการทำงานของระบบสนับสนุนการเลือกยา

เทคนิคการสำรวจกราฟอีกสองตัว ได้แก่ การค้นหาในแนวลึก และ การค้นหาในแนวกว้าง ต่อไป เพื่อหาเทคนิคที่เหมาะสมกับระบบการเลือกยานี้มากที่สุด

4.1.1 การคำนวณและเตรียมค่าความเป็นจุดศูนย์กลาง

งานวิจัยนี้ได้นำฟังก์ชันใน MATLAB มาใช้ในการคำนวณหาค่าความเป็นศูนย์กลางของยีนแต่ละตัวตามเส้นทางการส่งสัญญาณของเซลล์ ซึ่งตัววัดความเป็นศูนย์กลางที่นำมาใช้ในงานวิจัยนี้ ได้แก่ ความเป็นศูนย์กลางโดยวัดจากระดับ (Degree Centrality) ความเป็นศูนย์กลางโดยวัดจากความใกล้ชิด (Closeness Centrality) และ ความเป็นศูนย์กลางโดยวัดจากการคั่นกลาง (Betweenness Centrality) โดยมีขั้นตอนในการเตรียมข้อมูลดังภาพประกอบ 8

ขั้นตอนแรก ข้อมูลความสัมพันธ์ของยีนจาก KEGG จะถูกส่งออกจากฐานข้อมูล เพื่อสร้างรายการ (Edge List) เป็นไฟล์รูปแบบ CSV จากนั้นสคริปต์ที่ไม่ได้นำมาใช้ประโยชน์จะถูกลบทิ้งไป แล้วขอบทุกขอบจะถูกกำหนดค่าให้เป็น 1 โดยโปรแกรม MS Excel

ในขั้นตอนที่ 2 นั้นหมายความว่า น้ำหนักของขอบจะไม่มีผลต่อการวิเคราะห์เส้นทางการส่งสัญญาณของเซลล์ ในตอนนี้ เนื่องจากผู้เชี่ยวชาญทางมะเร็งไม่สามารถให้นำหนักหรือจัดลำดับของความสัมพันธ์แต่ละแบบระหว่างยีนในเส้นทางการส่งสัญญาณเดียวกันได้ เพราะฉะนั้น ไฟล์ของรายการขอบจะประกอบด้วย โหนดเริ่มต้น (entry1) โหนดเป้าหมาย (entry2) และ น้ำหนักของขอบ จากนั้น ข้อมูลเหล่านี้จะถูกแยกออกเป็น 14 ไฟล์ ตามชนิดของมะเร็ง ไฟล์รายการขอบจะถูกนำมาใช้หาค่าความเป็นศูนย์กลางต่อไปด้วย MATLAB

ขั้นตอนที่สาม คือ ขั้นตอนการเขียนโปรแกรมเพื่อคำนวณหาค่าความเป็นศูนย์กลาง ซึ่ง MATLAB มีฟังก์ชันในการคำนวณหาค่าความเป็นศูนย์กลางที่ชื่อว่า "centrality" ผลที่ได้จากการคำนวณโดยฟังก์ชันดังกล่าวจะประกอบด้วย ค่าความเป็นศูนย์กลางทั้งสามตัวดังกล่าวมาแล้วข้างต้นตามโหนดหรือไอดีของยีนและไอดีของเส้นทางการส่งสัญญาณที่อยู่ในรูปแบบไฟล์ CSV จากนั้นขั้นตอนสุดท้าย ข้อมูลของค่าความเป็นศูนย์กลางเหล่านี้จะถูกนำเข้าสู่ฐานข้อมูลเพื่อใช้สำหรับการสำรวจเส้นทางการส่งสัญญาณของเซลล์ต่อไป

ตัววัดความเป็นศูนย์กลางที่นำมาใช้ในงานวิจัยนี้ ประกอบด้วย ความเป็นศูนย์กลางโดยวัดจากระดับ (Degree Centrality) ความเป็นศูนย์กลางโดยวัดจากความใกล้ชิด (Closeness Centrality) และ ความเป็นศูนย์กลางโดยวัดจากการคั่นกลาง (Betweenness Centrality) โดยค่าความเป็นศูนย์กลางแต่ละแบบมีความหมายและวิธีการคำนวณแตกต่างกันออกไป

ความเป็นศูนย์กลางโดยวัดจากระดับ (Degree centrality) คือ จำนวนของการเชื่อมโดยตรงที่เกิดขึ้นกับจุด หรือจำนวนของจุดที่ใกล้เคียงที่สุด (Keen & Scott, 1978) ซึ่งสามารถคำนวณได้ดังสมการที่ 1

$$C_D(i) = k_i = \sum_j A_{ij} \quad (1)$$

โดยที่ $C_D(i)$ คือ ค่าความเป็นจุดศูนย์กลางของจุด i , k_i คือ จำนวนของเส้นที่เชื่อมถึงจุด i และ A_{ij} จะเท่ากับ 1 ถ้า i และ j เชื่อมโยงกันและเท่ากับ 0 ถ้าเป็นกรณีอื่นๆ

ความเป็นศูนย์กลางโดยวัดจากความใกล้ชิด (Closeness centrality) คือ ผลรวมของระยะทางระหว่างจุดที่พิจารณากับจุดอื่นทุกจุดภายในกราฟ (Bidgoli, 1989) ซึ่งสามารถคำนวณได้ดังสมการที่ 2



- 1) Export pathway relation as CSV file format from database.
- 2) Data Preparation
- 3) Compute centrality by MATLAB then save it as CSV file.
- 4) Insert the centrality values to database by PHP coding.

ภาพประกอบที่ 8 ขั้นตอนการเตรียมค่าความเป็นศูนย์กลาง

$$C_C(i) = \frac{1}{\sum_j d(i, j)} \quad (2)$$

โดยที่ $C_C(i)$ คือ ค่าความเป็นจุดศูนย์กลางของจุด i และ $d(i, j)$ คือ ระยะทางระหว่างจุด i และจุด j
ความเป็นศูนย์กลางโดยวัดจากการคั่นกลาง (Betweenness centrality) คือ จำนวนครั้งที่จุดที่พิจารณาทำตัวเป็นสะพานเชื่อมระหว่างเส้นทางที่สั้นที่สุด (Turban และคณะ, 2005) ซึ่งสามารถคำนวณได้ดังสมการที่ 3

$$C_B(i) = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}} \quad (3)$$

โดยที่ $C_B(i)$ คือ ค่าความเป็นจุดศูนย์กลางของจุด i , $\sigma_{st}(i)$ คือ จำนวนของเส้นทางที่สั้นที่สุดระหว่างจุด s และจุด t ที่ผ่านจุด i , σ_{st} คือ จำนวนทั้งหมดของเส้นทางที่สั้นที่สุดระหว่างจุด s และจุด t

4.1.2 วิธีการสำรวจกราฟของ CDFS

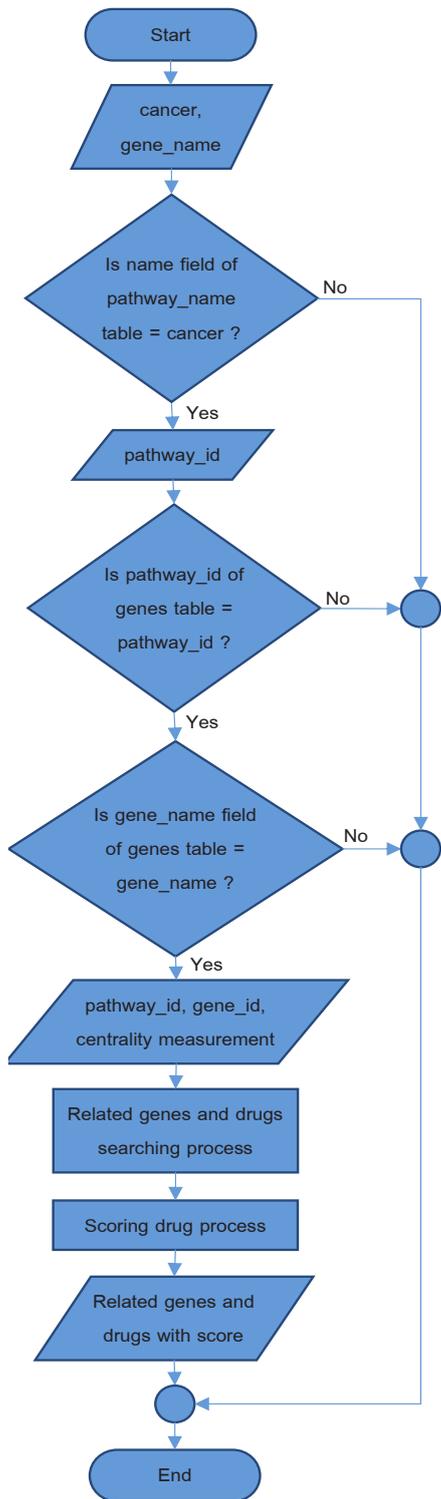
จากการศึกษางานวิจัยที่เกี่ยวข้อง พบว่ามีเทคนิคการสำรวจกราฟหลายเทคนิคที่น่าสนใจ แต่เทคนิคเหล่านั้นต้องการตัวแปรนำเข้าที่มากเกินไปกว่าที่งานวิจัยนี้มี และ เทคนิคเหล่านั้นยังไม่เหมาะสมสำหรับการสำรวจเส้นทางการส่งสัญญาณของเซลล์ เพราะฉะนั้นงานวิจัยนี้จึงได้พยายามนำเทคนิคการสำรวจกราฟพื้นฐาน เช่น การค้นหาในแนวลึก และ การค้นหาในแนวกว้าง มาประยุกต์รวมกับการวัดความเป็นศูนย์กลาง ค่าความเป็นศูนย์กลางแต่ละตัวจะช่วยให้หาเส้นทางที่เหมาะสมในการสำรวจเส้นทางการส่งสัญญาณ ซึ่งการประยุกต์นี้เกิดเป็นเทคนิคที่เรียกว่า “การค้นหาในแนวลึกตามความเป็นจุดศูนย์กลาง”

คำค้นหรือตัวแปรนำเข้าของเทคนิคนี้ในเว็บแอปพลิเคชันและการทดสอบจะมีความแตกต่างกัน แต่ผลลัพธ์ที่ได้นั้นเหมือนกัน เนื่องจากการทดสอบผู้วิจัยไม่ต้องการเสียเวลาในการป้อนคำค้นเข้าระบบ จึงได้เขียนโปรแกรมให้หน้าเข้าอัตโนมัติ ซึ่งคำค้นของเว็บแอปพลิเคชันจะประกอบด้วย ชื่อยีน ชนิดของมะเร็ง และ ตัววัดความเป็นศูนย์กลาง โดยกระบวนการค้นหาไอดีของยีนและไอดีของเส้นทางการส่งสัญญาณของเว็บแอปพลิเคชันแสดงดังภาพประกอบที่ 9 ก.

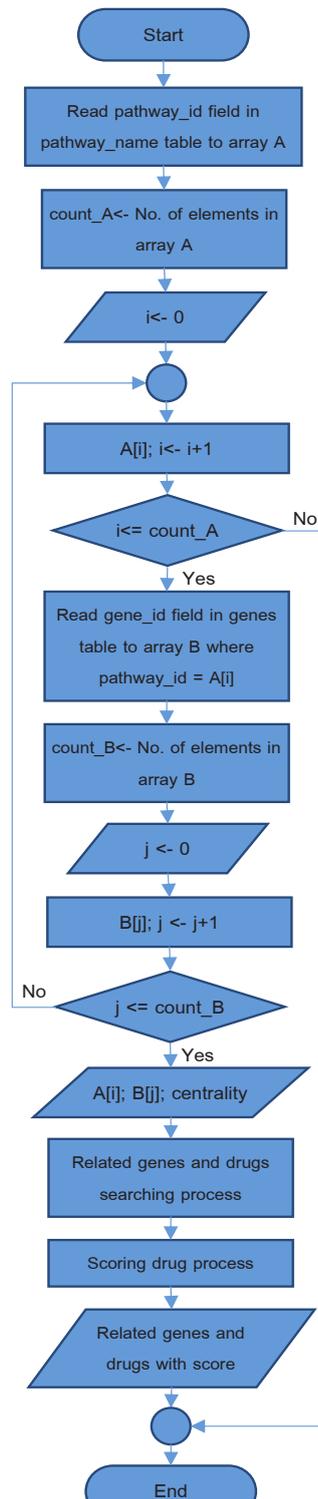
การทำงานเริ่มต้นด้วยการจับคู่คำค้นกับข้อมูลที่มีอยู่ในฐานข้อมูล ถ้าชนิดของมะเร็งตรงกับคุณลักษณะที่ชื่อว่า name ในตาราง pathway_name แล้ว ไอดีของเส้นทางการส่งสัญญาณจะถูกเลือกออกมา จากนั้นระบบจะนำไอดีของเส้นทางการส่งสัญญาณและชื่อยีนไปจับคู่กับข้อมูลในคุณลักษณะที่ชื่อว่า pathway_id และ gene_name จากตาราง genes ตามลำดับ สุดท้ายจะได้รับไอดีของยีนและไอดีของเส้นทางการส่งสัญญาณออกมา จากนั้นจึงนำไอดีของยีน ไอดีของเส้นทางการส่งสัญญาณ และ ตัววัดความเป็นศูนย์กลาง เข้าสู่กระบวนการค้นหา ยีนและยาที่เกี่ยวข้องต่อไป ส่วนกระบวนการเตรียมไอดีของยีนและไอดีของเส้นทางการส่งสัญญาณของการทดสอบแสดงดังภาพประกอบที่ 9 ข.

จากภาพประกอบที่ 9 ข. อินพุตของการทดสอบประกอบด้วย ไอดีของเส้นทางการส่งสัญญาณและไอดีของยีนเริ่มต้น โดยไอดีของเส้นทางการส่งสัญญาณจะถูกคิวรี (Query) จากตาราง pathway_name (จำนวน 14 เส้นทางการส่งสัญญาณ) เข้ามาเก็บเป็นอาเรย์ไว้ จากนั้นระบบจะเลือกไอดีของเส้นทางการส่งสัญญาณตัวแรกและไอดีของยีนที่ไม่ซ้ำกันจากคุณลักษณะที่ชื่อว่า entry_id ออกมา โดยที่คุณลักษณะ pathway_id จะเหมือนกับไอดีของเส้นทางการส่งสัญญาณ ไอดีของยีนทุกตัวจะถูกเพิ่มเข้าอาเรย์ภายใต้ไอดีของเส้นทางการส่งสัญญาณ และทำซ้ำไปเรื่อยๆ จนกระทั่งได้ไอดีของเส้นทางการส่งสัญญาณและไอดีของยีนครบหมดทุกตัวที่มีในฐานข้อมูล ซึ่งไอดีของยีนและไอดีของเส้นทางการส่งสัญญาณที่เตรียมเป็นชุดๆ ใว้จะถูกนำไปใช้ในกระบวนการค้นหา ยีนและยาที่เกี่ยวข้องต่อไป และกระบวนการค้นหา ยีนและยาที่เกี่ยวข้องแสดงดังภาพประกอบที่ 10

จากภาพประกอบที่ 10 จะเห็นว่าขั้นตอนที่ 9 และ 10 จะเป็นการทำงานของ CDFS ซึ่งจะกำหนด ไอดีของยีนที่ส่งเข้ามาตัวแรกเป็นโหนดเริ่มต้น และเลือกเดินไปหนึ่งในโหนดพ่อแม่ในระดับเดียวกันที่มีค่าความเป็นจุดศูนย์กลางสูงที่สุด แล้วพิจารณาต่อไปเช่นนี้ในแนวลึกลงไปจนกระทั่งเจอโหนดสุดท้าย ซึ่งทิศทางการสำรวจกราฟของอัลกอริทึมนี้สามารถอธิบายได้ดังภาพประกอบที่ 11 และแสดงทิศทางการสำรวจของการค้นหาในแนวลึก และ การค้นหาในแนวกว้าง ในภาพประกอบที่ 12 และ 13 ตามลำดับ เพื่อให้ในการเปรียบเทียบทิศทางการสำรวจกราฟ



ก. สำหรับเว็บแอปพลิเคชัน



ข. สำหรับการทดสอบ

ภาพประกอบที่ 9 กระบวนการค้นหาไอดีของยีนและไอดีของเส้นทางการส่งสัญญาณ



ภาพประกอบที่ 10 กระบวนการค้นหาเอนไซม์และยาที่เกี่ยวข้อง

จากภาพประกอบที่ 10 โหนด S คือ โหนดเริ่มต้น และ ตัวเลขในโหนด คือ ค่าความเป็นจุดศูนย์กลางโดย สมมติ ซึ่งจากภาพประกอบที่ 11, 12 และ 13 แสดงให้เห็นถึงความแตกต่างของทิศทางสำรวจกราฟ ผลลัพธ์ที่ได้จาก CDFS คือ S -> A -> E -> F ด้วยจำนวน 3 ก้าว ผลลัพธ์ที่ได้จาก DFS คือ S -> A -> E -> F -> D -> B -> C ด้วยจำนวน 6 ก้าว และ ผลลัพธ์ที่ได้จาก BFS คือ S -> A -> B -> C -> D -> E -> F ด้วยจำนวน 6 ก้าว จากผลลัพธ์ของทั้งสามอัลกอริทึมจะเห็นว่า CDFS ใช้จำนวนก้าวน้อยที่สุด เนื่องจากไม่ได้สำรวจทุกโหนด ขณะที่ DFS และ BFS ในภาพประกอบที่ 12 และ 13 ใช้จำนวนก้าวเท่ากัน แต่ลำดับของโหนดที่พบนั้นแตกต่างกัน เนื่องจากมีการสำรวจทุกโหนดเหมือนกัน แต่มีรูปแบบการเดินทางที่แตกต่างกันจึงทำให้พบโหนดที่ลำดับต่างกัน ซึ่งแต่ละอัลกอริทึมมีรูปแบบการเดินทางที่แตกต่างกันทำให้ลำดับการพบโหนดนั้นแตกต่างกันด้วย และยังใช้เวลาในการสำรวจแตกต่างกันตามจำนวนของโหนดที่ค้นพบและโหนดนั้นๆ มียาที่เกี่ยวข้องเท่าไร แม้ว่า CDFS จะไม่ได้สำรวจทุกโหนดเหมือน DFS และ BFS แต่วิธีการของ CDFS จะสามารถลดระยะเวลาในการสำรวจและวิเคราะห์เส้นทางการส่งสัญญาณของเซลล์ลงในเส้นทางการส่งสัญญาณที่มีขนาดใหญ่ได้

อย่างไรก็ตาม CDFS ยังคงต้องมีการปรับปรุงเพิ่มเติมในอนาคต เพราะงานวิจัยนี้ยังไม่ได้เพิ่มน้ำหนักของความสัมพันธ์ระหว่างยีนหรือขอบเข้ามาพิจารณาในการวิเคราะห์เส้นทางการส่งสัญญาณนี้ เนื่องจากผู้เชี่ยวชาญทางมะเร็งไม่สามารถจัดลำดับหรือให้น้ำหนักแก่ความสัมพันธ์แต่ละรูปแบบได้ เพราะมีรายละเอียดต่างๆ มากมายที่ต้องนำมาวิเคราะห์ร่วมด้วย แต่ในอนาคตผู้วิจัยวางแผนจะเพิ่มน้ำหนักของความสัมพันธ์ระหว่างยีนเข้ามาวิเคราะห์ด้วย เนื่องจากความสัมพันธ์ระหว่างยีนเป็นข้อมูลที่มีความสำคัญที่ไม่ควรมองข้าม

4.2 ระบบให้คะแนนยา (Scoring Drug System)

การให้คะแนนยา เป็นวิธีประเมินยาและจัดลำดับยาตามคะแนนเพื่อแนะนำยาที่เหมาะสมแก่การรักษา มะเร็งแบบมุ่งเป้า ข้อมูลที่นำมาใช้คำนวณคะแนนของยาประกอบด้วย 2 ส่วน ได้แก่ ข้อมูลยา (clinicaltrials.gov และ Drugs@FDA) และ ข้อมูลเส้นทางการส่งสัญญาณของเซลล์จาก KEGG ซึ่งคะแนนแบ่งออกเป็น 5 ส่วน

ได้แก่ Clinical Trials Score, Study Phase Score, Study Status Score, FDA Score, และ Hop Score โดยมีรายละเอียดดังต่อไปนี้

4.2.1 Clinical Trials Score มีค่าเป็น 1 ถ้าชื่อยีนถูกค้นพบในข้อมูลของ clinicaltrials.gov และ ถ้าไม่พบจะมีค่าเป็น 0

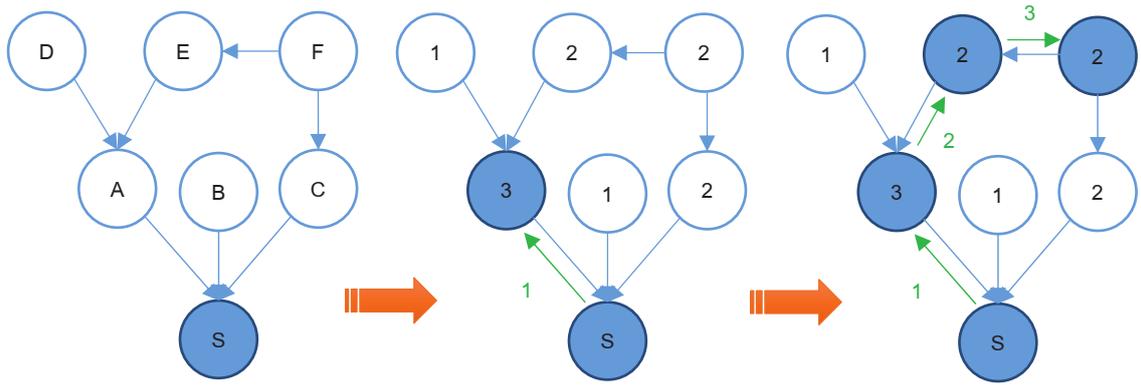
4.2.2 Study Phase Score มีค่าตามระยะการทดลองในงานศึกษาตามข้อมูลจาก clinicaltrials.gov ในกรณีที่ระบบพบงานศึกษาที่ศึกษายีนตัวเดียวกัน ระบบจะเลือกงานศึกษาที่อยู่ในระยะทดลองมากที่สุด ระยะในการทดลองประกอบด้วย 4 ระยะ ได้แก่ (1) ระยะตรวคัดกรอง (2) กำหนดขั้นตอนการทดสอบ (3) การทดสอบขั้นสุดท้าย และ (4) การอนุมัติงานศึกษา

4.2.3 Study Status Score เป็นการให้คะแนนตามสถานะของงานศึกษาตามข้อมูลจาก clinicaltrials.gov การให้คะแนนจะกำหนดตามความหมายและความสำคัญของสถานะของงานศึกษานั้นๆ ในกรณีที่มียานศึกษาที่ทำการศึกษายีนตัวเดียวกัน ระบบจะเลือกสถานะจากงานศึกษาที่มีระยะการทดลองมากที่สุดและวันที่เสร็จสมบูรณ์ล่าสุด ซึ่งสถานะที่เป็นไปได้ประกอบด้วย “ยังไม่ได้รับการคัดเลือก” หรือ “ยกเลิก” หรือ “ถอน” คะแนนจะเท่ากับ 0 ถ้าสถานะของงานศึกษา คือ “กำลังคัดเลือก” หรือ “ลงทะเบียน” คะแนนจะเท่ากับ 1 ถ้าสถานะคือ “ใช้งานอยู่ หยุดการคัดเลือก” คะแนนจะเท่ากับ 2 ถ้าสถานะคือ “เสร็จสมบูรณ์” คะแนนจะเท่ากับ 3 และถ้าสถานะของงานศึกษา คือ “ถูกระงับ” คะแนนจะเท่ากับ 0.5 ซึ่งสถานะนี้สามารถถูกเปลี่ยนให้เป็น “ถอน” หรือ “กำลังคัดเลือก” ได้ในภายหลัง

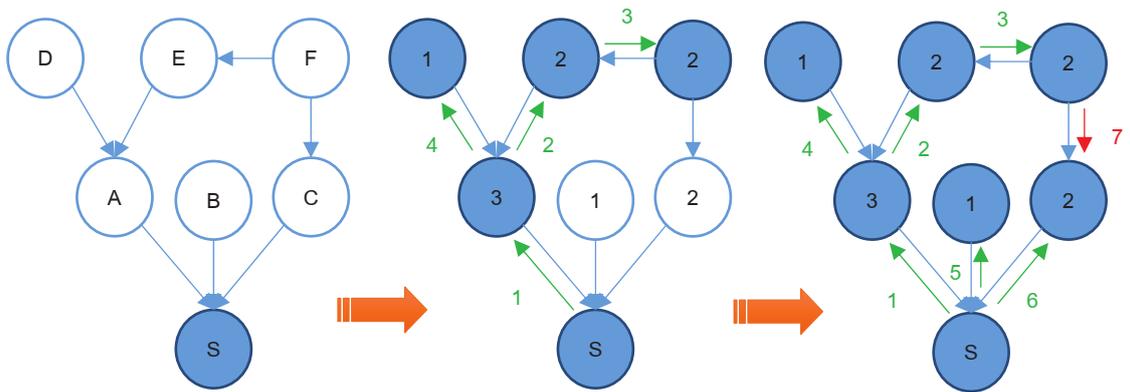
4.2.4 FDA Score จะเท่ากับ 1 ถ้า ชื่อยาถูกค้นพบในฐานข้อมูลของ FDA และ ถ้าไม่พบจะเท่ากับ 0

4.2.5 Hop Score กำหนดให้เป็นสัดส่วนผกผันของระยะทางจากโหนดต้นทางไปสู่โหนดที่พิจารณาบวกหนึ่ง ถ้าชื่อยาถูกพบจากโหนดมากกว่าหนึ่งโหนดในเส้นทางการส่งสัญญาณเดียวกัน ระบบจะเลือกคะแนนในส่วนนี้จากโหนดที่มีจำนวนก้าว (hop) มากที่สุด

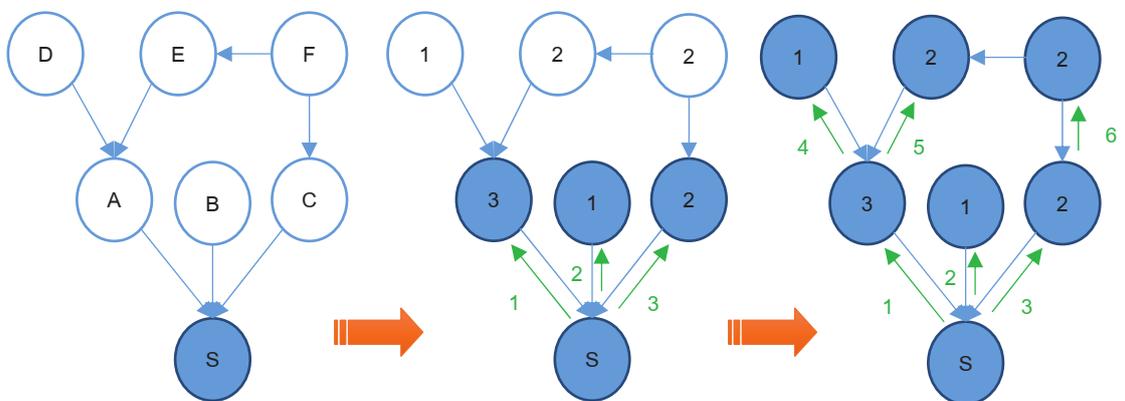
คะแนนแต่ละส่วนจะถูกทำให้ค่าอยู่ระหว่าง 0 ถึง 1 โดย Study Phase Score และ Study Status Score จะถูกหารด้วยคะแนนที่มากที่สุดของแต่ละส่วนคะแนน จากนั้นคะแนนแต่ละส่วนจะถูกคูณด้วยค่าน้ำหนักที่งานวิจัยนี้ได้กำหนดเป็นค่าเริ่มต้นตามความสำคัญของคะแนนแต่ละส่วนไว้ดังสมการที่ 4



ภาพประกอบที่ 11 ตัวอย่างของทิศทางการสำรวจกราฟของ CDFA



ภาพประกอบที่ 12 ตัวอย่างของทิศทางการสำรวจกราฟของ DFS



ภาพประกอบที่ 13 ตัวอย่างของทิศทางการสำรวจกราฟของ BFS

$$\frac{(C \times w_C + P \times w_P + S \times w_S + F \times w_F + H \times w_H)}{\text{Full score}} \times 100 \quad (4)$$

โดยที่ C คือ Clinical Trials score, P คือ Study Phase score, S คือ Study Status score, F คือ FDA score, H คือ Hop score และ w_C, w_P, w_S, w_F และ w_H คือ ค่าน้ำหนัก ซึ่งในงานวิจัยนี้ได้ทำการทดสอบจนได้ค่าที่เหมาะสมเป็น 1, 1, 1, 4, และ 7 ตามลำดับ ทำให้ Full score มีค่าเท่ากับ 14

5. ผลการวิจัยและการอภิปรายผล

ผลที่ได้จากการเปรียบเทียบคะแนนยาตามการวัดความเป็นจุดศูนย์กลาง และการเปรียบเทียบคะแนนยาตามเทคนิคการสำรวจเส้นทางการส่งสัญญาณของเซลล์มะเร็งสามารถแสดงได้ดังต่อไปนี้

5.1 การเปรียบเทียบคะแนนยาตามการวัดความเป็นจุดศูนย์กลาง

ในการทดสอบเพื่อเลือกวิธีการวัดความเป็นจุดศูนย์กลางของ CDFS ได้ทำการทดลองเปรียบเทียบวิธีวัดความเป็นจุดศูนย์กลาง 3 แบบ คือ ความเป็นจุดศูนย์กลางโดยวัดจากระดับความเป็นจุดศูนย์กลางโดยวัดจากความใกล้ชิด และ ความเป็นจุดศูนย์กลางโดยวัดจากการค้นกลาง ซึ่งงานวิจัยนี้ได้นำยื่นจำนวน 3,871 ตัว จาก 14 เส้นทางการส่งสัญญาณมาใช้เป็นตัวทดสอบสำหรับการหาที่เกี่ยวข้อง โดยทำการคำนวณคะแนนยาด้วยระบบการให้คะแนนยา และคำนวณคะแนนเฉลี่ยของยาสิบอันดับแรกในเส้นทางการส่งสัญญาณแต่ด้วยกันผลที่ได้แสดงดังตารางที่ 1 ซึ่งคะแนนยาเฉลี่ยสิบอันดับแรกนี้จะทำให้ทราบว่าวิธีการวัดความเป็นจุดศูนย์กลางวิธีใดที่สามารถค้นหายีนที่มีความสำคัญและพบยาที่มีความน่าเชื่อถือมากที่สุดเมื่อพิจารณาจากตัวแปรในระบบการให้คะแนนยา

จากตารางที่ 1 แสดงให้เห็นว่า เส้นทางการส่งสัญญาณหลายเส้นทางตามการวัดความเป็นจุดศูนย์กลางมีคะแนนยาเฉลี่ยเท่ากัน ซึ่งบ่งบอกได้ว่าลำดับยาของเส้นทางการส่งสัญญาณเหล่านั้นเหมือนกัน เนื่องจากเส้นทางการส่งสัญญาณที่นำมาใช้ทดสอบมีขนาดใหญ่ไม่เพียงพอที่จะทำให้เห็นความแตกต่างของเส้นทางการเดิน และคะแนนเฉลี่ยของยาตามการวัดความเป็นจุดศูนย์กลางโดยวัดจากระดับ และวัดจากความใกล้ชิด มีค่าน้อยกว่าคะแนนเฉลี่ยของยาตามการวัดความเป็นจุดศูนย์กลางโดยวัดจากการค้นกลาง ฉะนั้น

จึงสามารถสรุปได้ว่า การวัดความเป็นจุดศูนย์กลางโดยวัดจากการค้นกลางเหมาะสมสำหรับ CDFS มากที่สุดตามคะแนนเฉลี่ยของยาสิบอันดับแรกและเวลาที่เหมาะสม

อย่างไรก็ตาม การวัดความเป็นจุดศูนย์กลางโดยวัดจากการค้นกลางใช้เวลาในการวิเคราะห์มากที่สุด แต่ก็ไม่ต่างจากการวัดความเป็นจุดศูนย์กลางแบบอื่นมากนัก งานวิจัยนี้จึงเลือกใช้การวัดความเป็นจุดศูนย์กลางโดยวัดจากการค้นกลางบน CDFS ในการเปรียบเทียบผลคะแนนยาตามวิธีการสำรวจกราฟต่อไป ซึ่งคะแนนและเวลาเฉลี่ยของยาตามการวัดความเป็นจุดศูนย์กลางในรูปแบบของกราฟแสดงดังภาพประกอบที่ 14 และ 15 ตามลำดับ

5.2 การเปรียบเทียบคะแนนยาบนเทคนิคการสำรวจเส้นทางการส่งสัญญาณของเซลล์

การเปรียบเทียบผลคะแนนยาตามเทคนิคการสำรวจเส้นทางการส่งสัญญาณของเซลล์นั้น วิธีการสำรวจกราฟที่ถูกนำมาเปรียบเทียบ ประกอบด้วย การค้นหาในแนวลึก (DFS) การค้นหาในแนวกว้าง (BFS) และ การค้นหาในแนวลึกตามความเป็นจุดศูนย์กลาง (CDFS) งานวิจัยนี้ได้เลือกยาสิบอันดับแรกจากระบบการให้คะแนนยา และคำนวณเช่นเดียวกันกับการเปรียบเทียบแบบแรก ผลลัพธ์ที่ได้แสดงดังตารางที่ 2

จากตารางที่ 2 จะเห็นว่า CDFS มีคะแนนเฉลี่ยของยาสูงที่สุด และ ยังใช้เวลาในการวิเคราะห์น้อยกว่าเทคนิคอื่นๆ ด้วย ซึ่งสามารถสรุปได้ว่า CDFS เหมาะสมสำหรับสนับสนุนการเลือกยามากที่สุดด้วยคะแนนเฉลี่ยของยาสิบอันดับแรกมากที่สุดและเวลาที่เหมาะสม ซึ่งคะแนนและเวลาเฉลี่ยของยาตามเทคนิคการสำรวจเส้นทางการส่งสัญญาณของเซลล์ในรูปแบบของกราฟแสดงดังภาพประกอบที่ 16 และ 17 ตามลำดับ

6. สรุปผลการวิจัย

งานวิจัยนี้ได้นำเสนอวิธีการค้นหาในแนวลึกตามความเป็นจุดศูนย์กลาง หรือ CDFS เพื่อสำรวจเส้นทางการส่งสัญญาณของเซลล์ และแนะนำยาที่เหมาะสมสำหรับการรักษามะเร็งแบบมุ่งเป้าด้วยการวิเคราะห์เส้นทางการส่งสัญญาณของเซลล์ นอกจากนั้นยังได้นำเสนอวิธีการให้คะแนนยาเพื่อจัดอันดับตามความน่าเชื่อถือของยา ซึ่งสามารถนำไปใช้เรียงลำดับยาที่เป็นผลลัพธ์จากการสืบค้นเพื่อแนะนำยาที่เกี่ยวข้องได้ ส่วนการประเมินผลนั้น แบ่ง

ตารางที่ 1 คะแนนเฉลี่ยของยาตามการวัดความเป็นจุดศูนย์กลาง

รหัสเส้นทางการส่งสัญญาณ	Degree		Closeness		Betweenness	
	คะแนน (%)	เวลา (sec.)	คะแนน (%)	เวลา (sec.)	คะแนน (%)	เวลา (sec.)
hsa05210	33.728	35.039	33.728	37.879	33.728	35.759
hsa05211	43.083	13.118	43.053	12.227	43.053	12.695
hsa05212	57.219	33.847	57.024	33.549	57.150	33.469
hsa05213	45.461	65.150	45.461	65.370	45.461	66.168
hsa05214	52.129	58.562	52.129	56.019	52.129	55.663
hsa05215	50.200	32.196	50.200	31.821	50.200	33.794
hsa05216	42.255	12.504	42.255	12.700	42.255	12.610
hsa05217	4.737	0.638	4.737	0.601	4.737	0.601
hsa05218	56.630	20.753	56.630	20.692	56.630	20.872
hsa05219	37.769	9.581	37.769	9.740	37.769	9.686
hsa05220	38.607	5.652	38.607	5.609	38.607	5.469
hsa05221	42.748	7.016	42.747	6.768	47.904	13.291
hsa05222	35.242	4.061	35.242	4.034	37.706	4.152
hsa05224	52.675	15.086	52.675	14.779	52.675	15.037
Total (%)	44.761	23.723	44.729	23.560	45.202	24.014

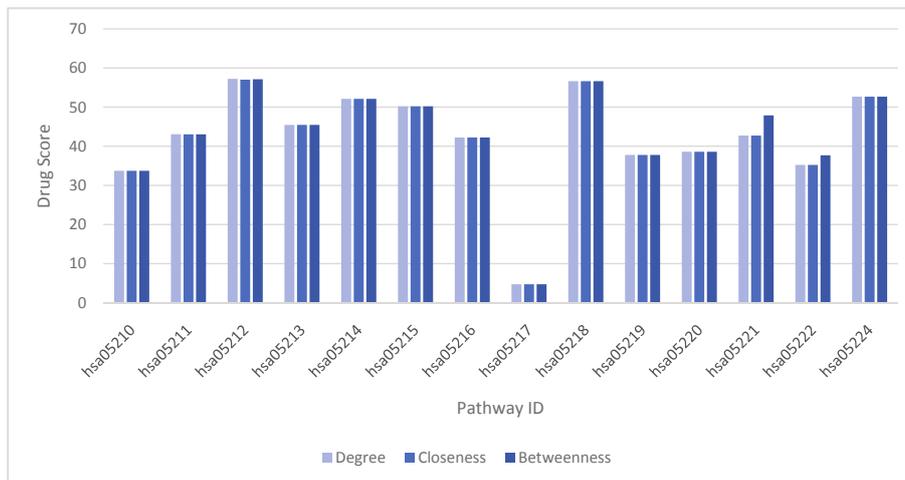
ตารางที่ 2 คะแนนเฉลี่ยของยาตามเทคนิคการสำรวจเส้นทางการส่งสัญญาณของเซลล์

รหัสเส้นทางการส่งสัญญาณ	DFS		BFS		CDFS	
	คะแนน (%)	เวลา (sec.)	คะแนน (%)	เวลา (sec.)	คะแนน (%)	เวลา (sec.)
hsa05210	32.771	29.565	33.487	19.477	33.728	35.759
hsa05211	35.426	12.870	37.050	13.220	43.053	12.695
hsa05212	56.286	30.131	55.904	33.390	57.150	33.469
hsa05213	45.461	62.926	45.444	69.799	45.461	66.168
hsa05214	47.456	55.608	47.775	58.159	52.129	55.663
hsa05215	48.891	68.967	49.784	70.563	50.200	33.794
hsa05216	39.047	47.236	41.345	52.838	42.255	12.610
hsa05217	4.737	2.072	4.737	1.620	4.737	0.601
hsa05218	56.759	84.551	56.708	87.913	56.630	20.872
hsa05219	33.981	32.412	35.867	49.345	37.769	9.686
hsa05220	38.198	20.371	38.414	19.123	38.607	5.469
hsa05221	46.335	69.386	46.502	78.636	47.904	13.291
hsa05222	36.625	14.0549	36.525	13.477	37.706	4.152
hsa05224	47.829	98.599	47.898	106.774	52.675	15.037
Total (%)	42.538	50.076	43.011	53.353	45.202	24.014

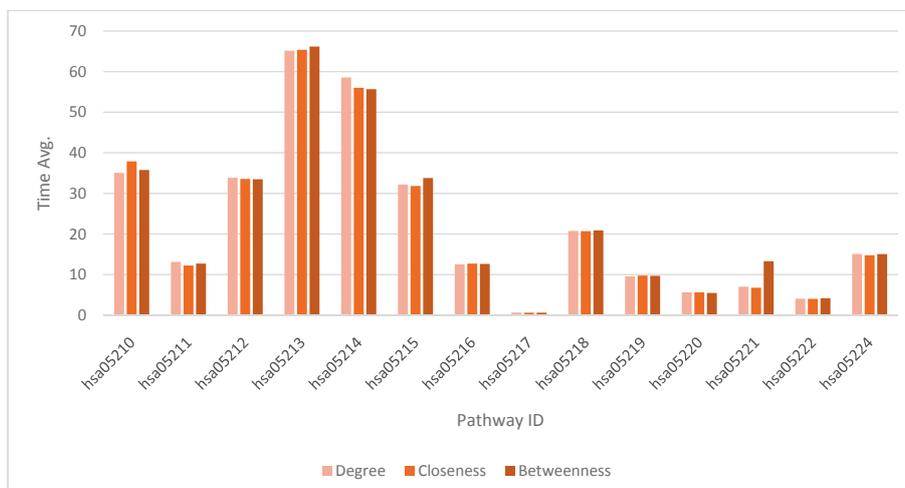
ออกเป็น 2 ส่วน ได้แก่ การเปรียบเทียบผลคะแนนของยาตามการวัดความเป็นจุดศูนย์กลาง ได้แก่ ความเป็นศูนย์กลางโดยวัดจากระดับ ความเป็นศูนย์กลางโดยวัดจากความใกล้ชิด และความเป็นศูนย์กลางโดยวัดจากการคั่นกลาง เพื่อหาวิธีวัดความเป็นศูนย์กลางที่เหมาะสมสำหรับ CDFS ซึ่งผลลัพธ์แสดงให้เห็นว่า การวัดความเป็นศูนย์กลางจากการคั่นกลางเหมาะสมกับ CDFS มากที่สุดด้วยคะแนนยาเฉลี่ยอันดับแรกมากที่สุดและเวลาที่เหมาะสม และการเปรียบเทียบผลคะแนนของยาตามเทคนิคสำรวจเส้นทาง การส่งสัญญาณของเซลล์ ได้แก่ การค้นหาในแนวลึก (DFS) การค้นหาในแนวกว้าง (BFS) และ การค้นหาในแนวลึกตาม

ความเป็นจุดศูนย์กลาง (CDFS) เพื่อหาวิธีการที่เหมาะสมสำหรับการแนะนำยาในระบบสนับสนุนการเลือกยาสำหรับการรักษาแบบมุ่งเป้าตามเส้นทางการส่งสัญญาณของเซลล์

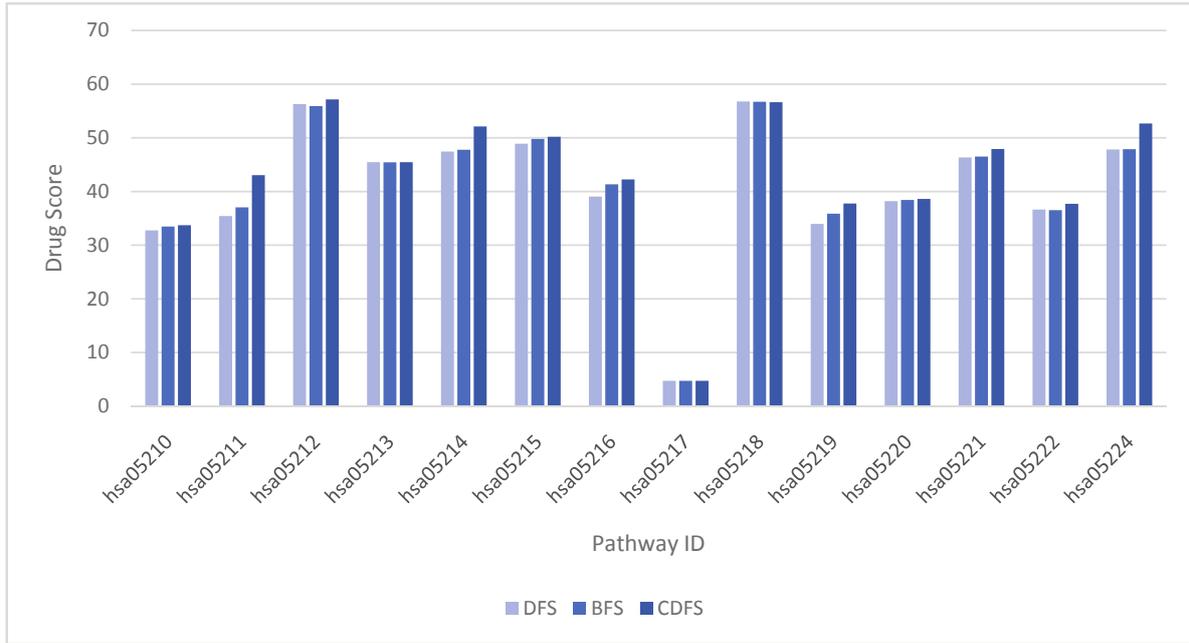
ผลการทดสอบแสดงให้เห็นว่า CDFS เหมาะสำหรับระบบสนับสนุนการเลือกยามากที่สุดด้วยคะแนนเฉลี่ยของยาอันดับแรกมากที่สุดและเวลาเฉลี่ยที่ใช้ประมวลผลน้อยที่สุด แต่การตัดสินใจสุดท้ายว่าจะเลือกยาตัวใดมาใช้ในการรักษานั้นยังขึ้นอยู่กับดุลยพินิจของแพทย์ เทคนิคนี้เป็นเพียงส่วนช่วยสนับสนุนการตัดสินใจ โดยเสนอตัวเลือกยาที่เพิ่มมากขึ้นให้ผู้ทำการตัดสินใจได้พิจารณา ปัญหาที่



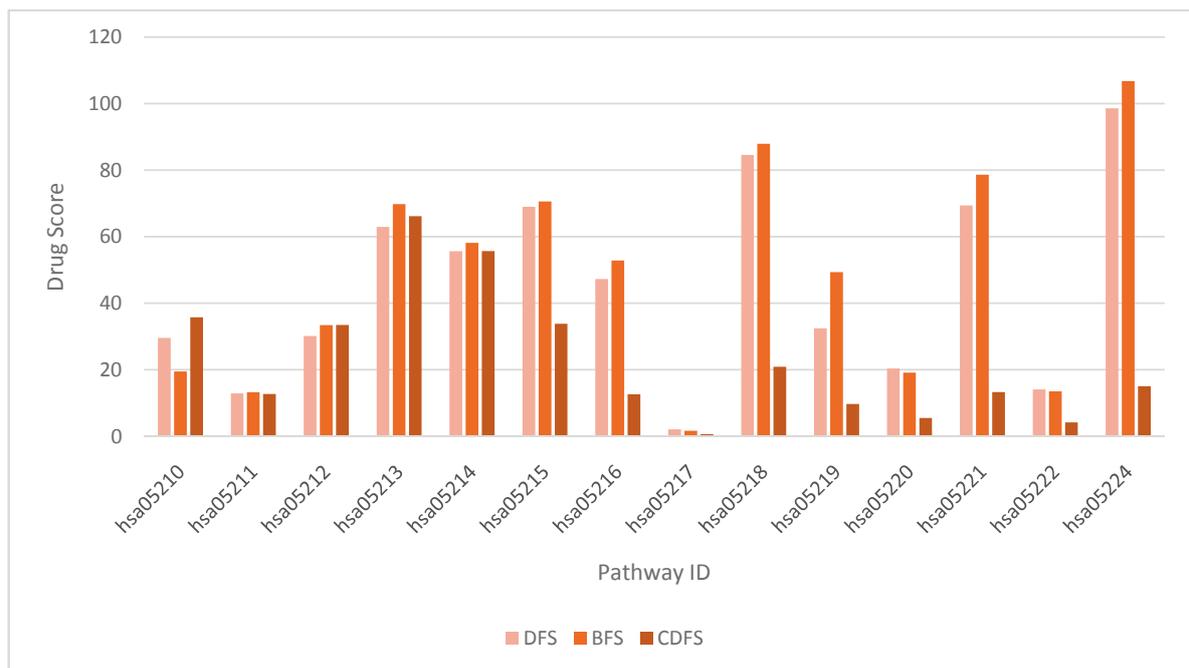
ภาพประกอบที่ 14 คะแนนเฉลี่ยของยาตามการวัดความเป็นจุดศูนย์กลางที่แสดงในรูปแบบกราฟแท่ง



ภาพประกอบที่ 15 เวลาเฉลี่ยของยาตามการวัดความเป็นจุดศูนย์กลางที่แสดงในรูปแบบกราฟแท่ง



ภาพประกอบที่ 16 คะแนนเฉลี่ยของยาตามการวัดความเป็นจุดศูนย์กลางที่แสดงในรูปแบบกราฟแท่ง



ภาพประกอบที่ 17 เวลาเฉลี่ยของยาตามการวัดความเป็นจุดศูนย์กลางที่แสดงในรูปแบบกราฟแท่ง

เกิดกับงานวิจัยนี้ ได้แก่ ชุดเส้นทางการส่งสัญญาณของเซลล์ที่นำมาทดสอบนี้มีขนาดใหญ่ไม่เพียงพอที่จะมองเห็นความแตกต่างของการเลือกเส้นทางการสำรวจในการสำรวจเส้นทางการส่งสัญญาณของเซลล์ ซึ่งในอนาคตอาจมีการเพิ่มชุดข้อมูลเข้ามาในการทดสอบ อีกทั้งการเตรียมข้อมูลยาและเส้นทางการส่งสัญญาณค่อนข้างยากที่จะ

ทำให้สมบูรณ์ เนื่องจากข้อมูลเหล่านั้นมีรูปแบบข้อมูลที่แตกต่างกัน และงานวิจัยนี้ถึงแม้จะได้รับคำแนะนำในเรื่องของการให้คะแนนยาโดยผู้เชี่ยวชาญ แต่รายการของยาที่ได้ ซึ่งเป็นผลลัพธ์สุดท้ายของงานวิจัยนี้ ยังคงต้องการการตรวจสอบจากผู้เชี่ยวชาญอยู่เพื่อเพิ่มความเชื่อมั่นในผลการวิจัยที่ได้

เอกสารอ้างอิง

- Bayat Mokhtari R. et al. (2017). Combination therapy in combating cancer. *Oncotarget*. 8(23). 38022-38043.
- Bidgoli, H. (1989). *Decision Support Systems: Principles and Practices*. St. Paul: West Publishing Company. 1989.
- BNH. (2011). Cancer Treatment Trends. *Better Health Magazine Issue 2*. Available from : <https://www.bumrungrad.com/en/betterhealth/2011/cancer-treatments/cancer-treatment-trends>. [August 18, 2011].
- BNH. (2013). Targeted Therapy. *Health Spot*. Available from: <https://www.bumrungrad.com/healthspot/November-2013>. [August 36, 2013].
- ClinicalTrials.gov. (2015). Available from: <https://clinicaltrials.gov>. [February 26, 2015].
- Deanna Petrochilos and Neil Abernethy. (2012). Assessing Network Characteristics of Cancer Associated Genes in Metabolic and Signaling Networks. *Proceeding of the 2012 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, San Diego, CA, USA, May 9-12, 2012, 290-297.
- FDA@Drugs (2015). Available from: <https://www.accessdata.fda.gov>. [February 26, 2015].
- Keen, P.G.W. and Scott Morton, M.S. (1978). *Decision Support Systems: An Organization Perspective*. Reading, MA : Addison-Wesley. 1978.
- KEGG. (2015). KEGG CANCER. Available from: <http://www.genome.jp/kegg>. [February 26, 2015].
- Maya El Dayeh and Michael Hahsler. (2012). Biological Pathway Completion using Network Motifs and Random Walks on Graphs. *Proceeding of the 2012 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, San Diego, CA, USA, May 9-12, 2012, 229-236.
- Meng-Hsiun Tsai et al. (2010). Genetic Regulatory Networks Established by Shortest Path Algorithm and Conditional Probabilistic for Ovarian Carcinoma Microarray Data. *Proceeding of the 9th International Conference on Machine Learning and Cybernetics*, Qingdao, July 11-14, 2010, 32-36.
- National Cancer Institute (2017). NCI Dictionary of Cancer Terms. Available from: <https://www.cancer.gov/publications/dictionaries/cancer-terms>. [January 10, 2018].
- National Cancer Institute (2017). Types of Cancer Treatment. Available from: <https://www.cancer.gov/about-cancer/treatment/types>. [January 10, 2018].
- Pankaj Agarwal and David B. Searls. (2008). Literature Mining in Support of Drug Discovery. *Briefings in Bioinformatics*. 9(6), 479-492.
- Taysir Hassan A. Soliman et al. (2012). Mining Disease Integrated Ontology. *Proceeding of the 2012 IEEE 12th International Conference on Bioinformatics and Bioengineering (BIBE)*, Larnaca, Cyprus, November 11-13, 2012, 40-45.
- Turban, E et al. (2005). *Introduction to Information Technology*. Hoboken : John Wiley & Sons. 2005.
- WHO. (2015). World health report. Available from: <http://www.who.int>. [January 12, 2015].
- Yongliang Yang et al. (2009). Target Discovery from Data Mining Approaches. *Drug Discovery Today*. 14, 147-154.