

การสร้างแบบจำลองการทำนายสถานะการรอดชีวิตของผู้ป่วยมะเร็งเต้านมโดยใช้ปัจจัยทางคลินิก PREDICTIVE MODELLING OF BREAST CANCER SURVIVAL STATUS USING CLINICAL FACTORS

ภูริ อุณจันนา¹, กัสกร สืบดี¹, ภูมิ ศรีรัตน์¹, ชินกริต คงทวีบุญ¹ และ เฉลิมรัช นนทะภา^{2*}

Phuri Ounjanum¹, Katsakon Suepde¹, Bhumi Sriratna¹, Chinnagrit Kongthaweeboon¹ and Chalermrat Nontapa^{2*}

¹ศูนย์วิจัยวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

²ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

* ผู้นิพนธ์ประสานงาน: เฉลิมรัช นนทะภา อีเมล: chalermrat.n@cmu.ac.th

บทคัดย่อ:

การศึกษานี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองการเรียนรู้ของเครื่องสำหรับการทำนายสถานะการรอดชีวิตของผู้ป่วยมะเร็งเต้านมโดยใช้ปัจจัยทางคลินิกและพยาธิวิทยา ข้อมูลที่ใช้ในการศึกษาได้มาจากฐานข้อมูลผู้ป่วยจำนวน 4,024 ราย ซึ่งประกอบด้วยตัวแปรสำคัญ ได้แก่ ขนาดของเนื้องอก (Tumor Size), ระยะของโรค (6th Stage), ระดับการจำแนกเซลล์ (Cell Differentiation), สถานะฮอร์โมนโปรเจสเตอโรน (Progesterone Status), อายุ (Age), สถานภาพสมรส (Marital Status), จำนวนต่อมน้ำเหลืองที่มีการแพร่กระจายของมะเร็ง (Regional Node Positive) และ จำนวนต่อมน้ำเหลืองที่ได้รับการตรวจ (Regional Node Examined) โดยมีตัวแปรตาม คือ สถานะการรอดชีวิตของผู้ป่วย ซึ่งแบ่งออกเป็น 2 กลุ่ม ได้แก่ Alive และ Dead ในการศึกษานี้ได้ประยุกต์ใช้ตัวแบบการเรียนรู้ของเครื่อง จำนวน 3 ตัวแบบ ได้แก่ Random Forest, XGBoost และ CatBoost เพื่อใช้ในการจำแนกสถานะการรอดชีวิตของผู้ป่วย และทำการประเมินประสิทธิภาพของแบบจำลองโดยใช้ตัวชี้วัด ได้แก่ Accuracy, Precision, Recall และ F1-Score ผลการทดลองพบว่า ตัวแบบ XGBoost ให้ค่า Accuracy เท่ากับ 0.91, Precision เท่ากับ 0.90, Recall เท่ากับ 0.91 และ F1-Score เท่ากับ 0.90 เป็นตัวแบบที่มีประสิทธิภาพดีที่สุดในการทำนายสถานะการรอดชีวิตของผู้ป่วยมะเร็งเต้านมจากปัจจัยทางคลินิกและสามารถนำไปประยุกต์ใช้เพื่อสนับสนุนการตัดสินใจทางการแพทย์และการวางแผนการรักษาได้อย่างมีประสิทธิภาพ

คำสำคัญ : มะเร็งเต้านม, การเรียนรู้ของเครื่อง, การจำแนกข้อมูล

Abstract:

This study aims to develop a machine learning model to predict the survival status of breast cancer patients using clinical and pathological factors. The dataset used in this study consisted of 4,024 patient records, including important variables such as Tumor Size, 6th Stage, Cell Differentiation, Progesterone Status, Age, Marital Status, Regional Node Positive, and Regional Node Examined. The target variable was the patients' survival status, which was classified into two groups: Alive and Dead. In this study, three machine learning models were applied, including Random Forest, XGBoost, and CatBoost, to classify patient survival status. The performance of the models was evaluated using several metrics, including Accuracy, Precision, Recall, and F1-Score. The experimental results showed that the XGBoost model achieved the best performance, with an Accuracy of 0.91, Precision of 0.90, Recall of 0.91, and F1-Score of 0.90. The results indicate that the XGBoost model provides the highest performance in predicting the survival status of breast cancer patients based on clinical factors and can be applied to support medical decision-making and treatment planning effectively.

Keywords: Breast Cancer, Machine Learning, Classification