

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

ภาษาล้านนาเป็นภาษาที่ใช้กันในอาณาจักรล้านนาในศตวรรษที่ 13-18 แต่หลังจากที่อาณาจักรล้านนาถูกยึดครองโดยอาณาจักรสยามในปี ค.ศ. 1774 ตัวอักษรล้านนาได้ถูกละเลยและถูกแทนที่ด้วยตัวอักษรไทยจนในปัจจุบันมีเพียงไม่กี่คนเท่านั้นที่รู้วิธีอ่านและเขียนภาษานี้ จากการสืบค้นข้อมูลทางประวัติศาสตร์พบว่าคนล้านนาจะใช้ใบลาน พับสาหรือศิลาจารึกในการจารึกอักษรล้านนาเพื่อบันทึกเกี่ยวกับวิถีชีวิต ความเชื่อ กฎหมาย เรื่องเล่า สุตฺรยาสมุนไพโร ประวัติศาสตร์ โหราศาสตร์ และความรู้ทั่วไปอื่นๆ เมื่อกาลเวลาผ่านไป เอกสารโบราณเหล่านี้ได้สูญหาย ถูกทำลาย หรือสูญหายไป ดังนั้นเพื่อที่จะอนุรักษ์ข้อมูลทางประวัติศาสตร์อันล้ำค่าที่เอกสารเหล่านี้ได้บันทึกไว้ จึงควรมีการนำเอาระบบคอมพิวเตอร์มาใช้เพื่อแปลตัวอักษรเหล่านี้ให้เป็นตัวอักษรไทยที่ใช้กันในปัจจุบัน

จากการศึกษาวิจัยต่างๆ ที่ผ่านมามีการพัฒนาอัลกอริทึมในการรู้จำตัวอักษร (OCR) จะประกอบไปด้วยขั้นตอน 3 ขั้นตอนได้แก่ การประมวลผลภาพเบื้องต้น การดึงคุณสมบัติของตัวอักษร และการเรียนรู้จำตัวอักษร ในขั้นตอนของการประมวลผลภาพเบื้องต้นจะเป็นกระบวนการที่ทำการปรับปรุงคุณภาพของรูปภาพตัวอักษรให้ดีขึ้นและทำการตัดตัวอักษรให้เป็นตัวอักษรเดี่ยวเพื่อเข้าสู่ขั้นตอนการดึงคุณสมบัติและขั้นตอนการเรียนรู้จำตัวอักษรต่อไป ในการตัดตัวอักษรนั้นปัญหาแรกๆ มักจะพบคือปัญหาตัวอักษรสัมผัสกัน โดยเฉพาะตัวเขียน ตัวอักษรล้านนาก็พบปัญหานี้เช่นกัน ดังนั้นการที่จะทำให้การรู้จำตัวอักษรมีประสิทธิภาพที่ดีจึงจำเป็นต้องแก้ไขปัญหาคือตัวอักษรติดกันเสียก่อน ในการศึกษาและการวิจัยในครั้งนี้ได้มุ่งเน้นการแยกตัวอักษรสัมผัสกันโดยได้มีการนำเอาศาสตร์ทางด้านปัญญาประดิษฐ์มาทำการพัฒนาในการแยกตัวอักษรที่สัมผัสกันและเป็นที่ยอมรับว่ามีประสิทธิภาพได้แก่อัลกอริทึมการเรียนรู้แบบมีข้อมูลในการสอน อัลกอริทึมการเรียนรู้แบบไม่มีข้อมูลในการสอน และอัลกอริทึมการใช้โครงสร้างของตัวอักษร มาช่วยในการแยกตัวอักษรที่ติดกัน

การศึกษาในครั้งนี้มุ่งเน้นการประยุกต์ใช้อัลกอริทึมการใช้โครงสร้างของตัวอักษร ในการแยกตัวอักษรที่ติดกันให้เป็นตัวอักษรเดี่ยว

## 1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

เพื่อที่จะทำการพัฒนาวิธีที่เหมาะสมในการตัดแยกตัวอักษรที่สัมผัสกันของตัวอักษรล้านนาให้ออกเป็นตัวอักษรเดี่ยวได้อย่างถูกต้อง

## 1.3 ขอบเขตของการวิจัย

นำเสนอโมเดลสำหรับการแยกตัวอักษรที่สัมผัสกันของตัวอักษรล้านนา โดยทำการทดลองกับชุดข้อมูลภาพถ่ายตัวอักษรล้านนาที่บันทึกในใบลานของภาคเหนือตอนบน

## 1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. มีความรู้ความเข้าใจอัลกอริทึมในการแยกตัวอักษรที่สัมผัสกัน
2. ได้โมเดลใหม่สำหรับการแยกตัวอักษรที่สัมผัสกันของตัวอักษรล้านนา ซึ่งจะนำไปสู่งานวิจัยด้านการรู้จำตัวอักษรล้านนาต่อไป

## 1.5 ขั้นตอนของการดำเนินงานวิจัย

1. ศึกษางานวิจัยทางด้านการแยกตัวอักษรที่สัมผัสกัน
2. ศึกษาทฤษฎีที่เกี่ยวข้องกับการพัฒนาโมเดลการแยกตัวอักษรสัมผัสกันที่มีประสิทธิภาพ และทดลองนำโมเดลเหล่านั้นมาตัดแยกตัวอักษรล้านนา
3. วิเคราะห์และออกแบบโมเดลใหม่
4. พัฒนาโปรแกรมตามรูปแบบโมเดลที่ได้ออกแบบไว้ โดยใช้โปรแกรม MATLAB
5. จัดเตรียมข้อมูลสำหรับการทดลองวิจัยแล้วดำเนินการทดลองเพื่อวัดประสิทธิภาพของโมเดลที่นำเสนอ

6. สรุปผลการทดลอง

7. จัดทำรายงานการวิจัย