



אורי טביבי
Orita4@ac.sce.ac.il



שון אסולין
Shonas@ac.sce.ac.il

זיהוי מגדר על בסיס תמונות כתבי יד

BS_SE-21-104

מנחות: ד"ר אירינה רבייב
ד"ר מרינה ליטבק

תקציר

זיהוי מגדר לפי כתב יד נחשבת לאחת הבעיות הנחקרות והמעניינות המעוררת עניין רב בשל מגוון התחומים הרחב שנעשה בהם שימוש, החל מרפואה לצרכי מחקר על התנהגות ובריאות האדם ועד לשימושים נרחבים כמו אימות נתונים, ניתוח מסמכים היסטוריים וחקירות קרימינולוגיות.

פרויקט זה עוסק במחקר ופיתוח כלי אוטומטי המבוסס על רשתות נוירונים לזיהוי מגדר על בסיס תמונות כתבי יד, חוקרים רבים מראים שבעיה זו לא פשוטה לפתרון על ידי בני-אדם שלא מומחים בנושא זה.

במסגרת המחקר בחנו שימוש במספר מודלים של רשתות נוירונים עמוקות לזיהוי תמונה, המבצעים חילוף מאפיינים מסריקות כתבי היד ולאחר מכן מזהות את מין הכותב. השימוש במודלים אלו מראה יכולות טובות לזיהוי המאפיינים המבדילים בין השוני של כתבי היד לפי מגדר.

הניסויים התבססו על מאגר HHD_gender המכיל תמונות כתבי יד בשפה העברית, כמו כן, על מנת להרחיב את המחקר ולהעריך את טיב המודלים, ביצענו ניסויים נוספים על מאגר QUWI המכיל תמונות כתבי יד בשפות האנגלית והערבית.

בנוסף, במהלך המחקר ערכנו סקר שהופץ לקהל הרחב כדי להשוות ביצועים בין זיהוי המתבצע על ידי בני אדם שלא מומחים בניתוח כתבי יד לבין שימוש ברשתות נוירונים.

בשלב השני לפרויקט, בנינו מערכת (אתר) המאפשרת למשתמש להעלות סריקה של כתב יד ולקבל ניתוח מפורט לגבי המידע שהתקבל בתהליך זיהוי המין של הכותב.

במסגרת המחקר ראינו כי קיים שוני בין כתבי היד של שני המינים וכי שימוש בכלים אוטומטיים מביא לתוצאות טובות יותר בזיהוי לעומת זיהוי הנעשה על ידי בני אדם שלא בקיאים בנושא.

במהלך המחקר נכתב מאמר אקדמי על ידי צוות הפרויקט הכולל את ד"ר מרינה ליטבק וד"ר אירינה רבייב, המתאר את המחקר והעבודה שבוצעה. מאמר זה נמצא בשלבי שיפוט.

מילות מפתח: זיהוי מגדר, כתבי יד, למידה עמוקה, רשתות נוירונים עמוקות.