

מערכת לזיהוי פגמים ופסילות בקלף מזוזה ותפילין

שם: הודיה מושייב

מנחה: יוסף סולביץ'

תקציר:

קלפי מזוזה ותפילין הם פריטים דתיים יהודיים בעלי חשיבות רבה. פגמים או פסילות בהם עלולים לפגום בכשרותם ולפסול את השימוש בהם. זיהוי פגמים אלו באופן ידני יכול להיות תהליך מורכב, גוזל זמן ומוטה לטעויות אנושיות. לכן המטרה היא לפתח שיטה אוטומטית לזיהוי פגמים ופסילות בקלף מזוזה ותפילין באמצעות שיטות למידה עמוקה, ולהשוות את התוצאות לשיטות מסורתיות שאינן עוסקות בלמידה עמוקה. הניסוי התמקד בשלוש גישות שונות לזיהוי פגמים: השיטה הראשונה היא **Match Template** – השוואת תמונות לתבניות ידועות של אותיות, שיטה נוספת היא **Image Hash** – יצירת טביעת אצבע ייחודית לכל תמונה והשוואתה לטביעות אצבע של תמונות של אותיות ידועות, והשיטה האחרונה היא **CNN** – לימוד רשת נוירונים קונבולוציונית לזהות אותיות בתמונות באופן אוטומטי. רשת ה-CNN הציגה את הביצועים הטובים ביותר, עם דיוק של 99.16%, בהשוואה ל-94.24% עבור **Match Template**, ו-90.04% עבור **Image Hash**. בעזרת שיטות אלו ניתן לבדוק בקלות רבה פגמים כמו: אותיות חסרות, אותיות מיותרות, אותיות מוכפלות וכו' וכך לחסוך משאבי אנוש ותקציבים רבים ולמנוע טעויות אנוש אפשריות.

הקדמה:

בדיקת תפילין ומזוזות היא תהליך חיוני, שמטרתו לוודא שהפריטים עומדים בדרישות ההלכתיות הקפדניות. כיום, בדיקה זו מתבצעת באופן ידני על ידי סופרים מומחים, תהליך הכרוך בזמן רב, עלויות גבוהות וכוח אדם מוגבל. כתוצאה מכך, עלולים להיווצר קשיים בהשגת תפילין ומזוזות כשרים, פגיעה באמון הציבור ובזבז משאבים יקרים. במחקר זה, אנו מציגים שיטה אוטומטית חדשנית לזיהוי פגמים ופסילות בקלף מזוזה ותפילין באמצעות שיטות למידה עמוקה. שיטה זו משווה את ביצועיה לשיטות מסורתיות שאינן עוסקות בלמידה עמוקה, ומדגישה את היתרונות והפוטנציאל שלה לשיפור תהליך הבדיקה והאימות של פריטים אלו. פיתוח מערכת אוטומטית לזיהוי אותיות בתפילין ובמזוזות יכול להוות פתרון יעיל לבעיות אלו. מערכת כזו תוכל לקצר משמעותית את זמן הבדיקה, להפחית את עלויותיה ולשפר את הדיוק. כתוצאה מכך, תפילין ומזוזות יהיו נגישים יותר ליותר אנשים, תוך הבטחת כשרותם וחיסכון משמעותי במשאבים. המטרה העיקרית של מחקר זה היא לפתח שיטה אוטומטית יעילה ומדויקת לזיהוי פגמים ופסילות בקלף מזוזה ותפילין. אנו שואפים להשיג זאת באמצעות שיטות למידה עמוקה, ולשפר משמעותית את תהליך הבדיקה והאימות של פריטים אלו בהשוואה לשיטות מסורתיות. השערות המחקר הן ששיטות למידה עמוקה מסוגלות לזהות פגמים ופסילות בקלף מזוזה ותפילין ברמת דיוק גבוהה יותר משיטות מסורתיות שאינן עוסקות בלמידה עמוקה. בנוסף, שימוש בשיטות אלו יאפשר חיסכון משמעותי בזמן ובמשאבים אנושיים, תוך הפחתת הסיכון לטעויות

אנוש. שיטות אלו יהיו יעילות לאיתור מגוון רחב של פגמים, כגון אותיות חסרות, אותיות מיותרות, אותיות מוכפלות ועוד.

מחקר זה נושא בחובו חשיבות רבה הן מבחינה דתית והן מבחינה טכנולוגית. מבחינה דתית, הוא תורם לשיפור איכותם וכשרותם של פריטים דתיים בעלי משמעות רבה עבור המאמינים. מבחינה טכנולוגית, הוא מדגים את הפוטנציאל של שיטות למידה עמוקה ליישום בתחומים מגוונים, תוך שיפור יעילות ודיוק בתהליכים מורכבים.

במחקר זה, אנו מציגים שלוש גישות שונות לזיהוי פגמים בקלף:

שיטה ראשונה - Match Template: השיטה משווה תמונות של קלף לתבניות ידועות של אותיות. שיטה שנייה - Image Hash: השיטה יוצרת טביעת אצבע ייחודית לכל תמונה ומשווה אותה לטביעות אצבע של תמונות של אותיות ידועות.

השיטה השלישית היא CNN: השיטה מלמדת רשת נוירונים קונבולוציונית לזהות אותיות בתמונות באופן אוטומטי.

לאחר בדיקת כל אחת מהגישות, אנו משווים את תוצאותיהן ומציגים את הגישה היעילה ביותר. בסופו של דבר, אנו דנים ביתרונות ובמגבלות של שיטות אלו, ומציעים כיוונים עתידיים למחקר ופיתוח.

אנו מאמינים שמחקר זה יהווה תרומה משמעותית הן לתחום הדת והן לתחום טכנולוגיות הלמידה העמוקה.

שיטות:

- קיבלנו מידי סופר סת"ם תמונה של קלף מזוזה, עליה הופעלו פונקציות לחילוץ האותיות ומיון לפי התייגים שלהם לצורך השוואה ע"י כל אחת מהשיטות. עבור הניסוי בקלף זה, שמרנו את המיקומים של כל אות בקלף כדי לשמור את כל האותיות שבקלף לפי הסדר שבו הן נכתבו.
1. **Match Template:** שיטת התאמה לתבנית היא טכניקת עיבוד תמונה שיכולה לשמש לזיהוי אותיות בקלף של מזוזה ותפילין. שלבי התהליך כוללים את חלוקת תמונת קלף המזוזה לאזורים קטנים, תוך התחשבות באפשרויות מיקום האותיות, חילוף תמונות האותיות לכל אזור, לאחר מכן ביצוע התאמה עבור כל תמונת אות ע"י השוואה שלה עם כל תבנית אות ידועה באמצעות פונקציות `cv.matchTemplate` של `OpenCV`, ולאחר מכן בחירת האות המתאימה ע"י התבנית שהציגה את הערך הגבוה ביותר בפונקציה עבור כל תמונת אות.
 - הפונקציה מקבלת 2 תמונות כקלט ומחזירה כתוצאה תמונה שבה כל פיקסל מייצג את דמיון התבנית באותו מיקום בתמונה המקורית.
 - בניסוי זה, אנחנו מבצעים התאמה בין אות מה `data` לאות מהקלף. לכן מה שנותר לבצע הוא לקחת כל אות מהקלף מרשימת האותיות ששמרנו לפי הסדר, ולבדוק לאיזה אות מתוך כל `data` היא הכי קרובה, והאות שתחזיר את הערך המקסימלי תיבחר להיות האות המיוצגת בהסתברות הגבוהה ביותר.
 2. **Image Hash:** בשיטה זו משתמשים בטכניקת חישוב "טביעת אצבע" שיכולה גם כן לשמש לסיווג של אותיות בקלף. התהליך דומה לתהליך שמבוצע בשיטת `OpenCV`, וההבדל הוא שעבור כל תמונת אות מיוצר גיבוב (כמו הצפנה), כך שככל שהאותיות דומות יותר, הגיבובים שלהם יהיו יותר דומים וההפרש בין הגיבובים יהיה יותר קטן.

לאחר שיש לכל תמונה גיבוב משלה נבצע השוואה וניקח את הגיבוב עם ההפרש הקטן ביותר מהתמונה המקורית, כהסתברות הגבוהה ביותר לאות המיוצגת.

3. CNN: רשת נוירונים המתוכנתת במיוחד לזיהוי דפוסים בתמונות. רשתות כאלה משמשות במגוון רחב של משימות ראייה ממוחשבת, כמו זיהוי תמונות, סיווג תמונות, עיבוד תמונה וכו'.

רשתות נוירונים כאלה ואחרות הינן פחות נוקשות ומסוגלות להכליל מערך נתונים ברמה גבוהה מאוד. החיסרון הוא שצריך כמות גדולה של נתונים כדי לאמן רשתות כאלה כדי להגיע לביצועים טובים ולתוצאות דיוק גבוהות.

מבנה רשת ה-CNN מורכב משכבות קונבולוציה (Conv2D) – שכבות אלו מריצות חלון קטן (הנקרא "פילטר") על פני התמונה, מחשבות את המכפלה הפנימית בין הפילטר לאזור התמונה וסוכמות את התוצאות. תהליך זה מאפשר להן ללמוד דפוסים מקומיים בתמונה. שכבות נוספות נקראות Pooling – אלה מצמצמות את מימד המאפיין של הפלט משכבת הקונבולוציה הקודמת. הן יכולות לבצע פעולות כמו מקסימום (MaxPooling2D) או ממוצע (AveragePooling2D), אשר מפחיתות את רגישות הרשת לשינויים קטנים במיקום המאפיין בתמונה. סוג נוסף של שכבה הוא שכבת עיבוד מלא – Fully connected (Dense) שמטרתה לחבר את המאפיינים שהופקו מהשכבות הקודמות לווקטור יחיד, ולחשב את הפלט הסופי באמצעות פונקציית אקטיבציה.

נוכל לחלק את הדאטא ל-datan, test, validation- כדי לאמן את הרשת על קבוצת נתונים אחת של אותיות, ולבצע אימות בדיקה על קלפים מסוים של מזוזה או תפילין.

את התוצאות האלה נוכל להשוות אל מול התוצאות שנקבל בשיטות הקודמות, ונוכל לראות איך הרשת תתנהג עם האותיות שלא זוהו נכון ברוב הפעמים בשיטות הראשונות.

תוצאות:

שיטת ה-Match Template הוכחה כיעילה בזיהוי אותיות בקלף במצבים מסוימים, אך הציגה גם מגבלות משמעותיות. היתרונות הם שהשיטה יחסית פשוטה להבנה וליישום, מה שהופך אותה לבחירה טובה עבור פרויקטים הדורשים פתרון מהיר. השיטה מצטיינת בזיהוי אותיות ברורות וחדות, כאשר התאמת התבנית לאות המקורית היא גבוהה. לעומת זאת, השיטה רגישה מאוד לרעש בתמונה, שינויים בתאורה, עיוותים קלים באותיות או שינויים בגודל. כל שינוי קטן בתמונה עלול להשפיע משמעותית על התוצאות. בנוסף, הכנת תבניות איכותיות לכל אות יכולה להיות תהליך ארוך ומייגע, במיוחד עבור כתב יד או אותיות עם וריאציות רבות. כמו כן, השיטה מתקשה להתמודד עם אותיות שמופיעות בגדלים שונים או בזוויות שונות, ולכן בשל המגבלות הנ"ל, הדיוק של השיטה הוא בדרך כלל נמוך יותר בהשוואה לשיטות מבוססות למידה עמוקה.

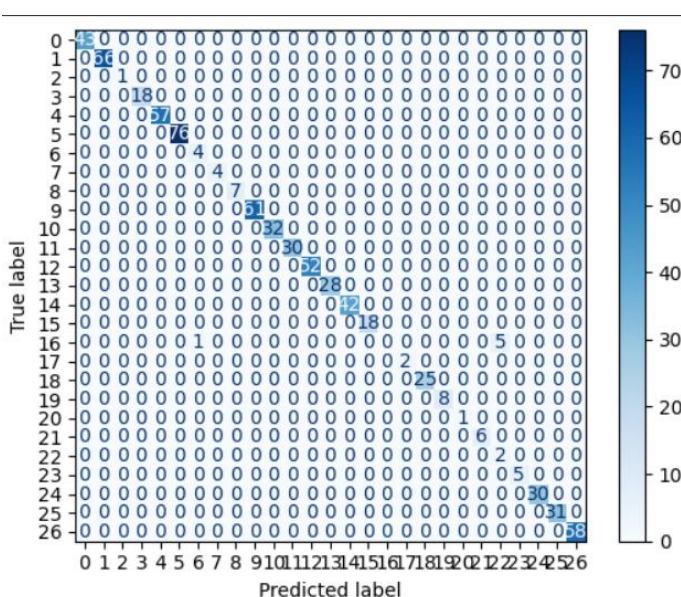
כשביצענו השוואה בין כל אחת מאותיות הקלף לפי הסדר, לבין כל אחת מאותיות הא"ב, קיבלנו דיוק של 90.88%, וכשביצענו התאמה בין כל אות מהקלף עם כל האותיות שנמצאות ב-datan, אחוז הדיוק גדל ל-94.24%, כאשר רוב הטעויות היו בין האותיות ס-ם, ו-ן, וכן בין אותיות דומות על תגים, כשההבדלים נראים כך:

כשהגענו לשיטת ה-CNN, התחלנו לפתח רשת בעלת ארכיטקטורה פשוטה אך יעילה. הרשת מורכבת משלוש שכבות קונבולוציה, כאשר כל שכבה מבצעת חילוץ תכונות ברמות מורכבות שונות. לאחר כל שכבת קונבולוציה, הוספנו שכבת Dropout כדי למנוע התאמת יתר. שכבות Max pooling צמצמו את הממד של התמונה ושמרו על התכונות החשובות. לבסוף, שתי שכבות Fully connected ביצעו את סיווג האותיות. לאימון הרשת, השתמשנו באופטימיזטור Adam ופונקציית Categorical Crossentropy - Loss. תהליך האימון כלל 30 epochs, כאשר בכל

Confusion matrix for the CIFAR-10 dataset. The x-axis is 'Predicted label' and the y-axis is 'True label', both ranging from 0 to 9. The matrix shows counts for each combination of true and predicted labels. A color bar on the right indicates the count scale from 0 to 60. The diagonal elements (where true label equals predicted label) are the highest, indicating good classification performance.

(Recall). ה-F1-Score, המשלב את ה-Precision, וה-Recall מספק מדד כולל לביצועי המודל ועומד על 0.8922, מה שמעיד על איזון טוב בין היכולת של המודל לזהות נכון מקרים חיוביים והיכולת שלו לא להחמיץ מקרים חיוביים. ניתוח ה-Confusion Matrix חשף כי רוב הטעויות שנעשו על ידי המודל התרכזו בבלבול בין אותיות דומות בצורתן, כגון ו-ר, ע-ש, צ-ע. תוצאה זו מצביעה על כך שהמודל רגיש להבדלים עדינים בין האותיות, ודורש שיפור נוסף באימון על מערך נתונים גדול ומגוון יותר.

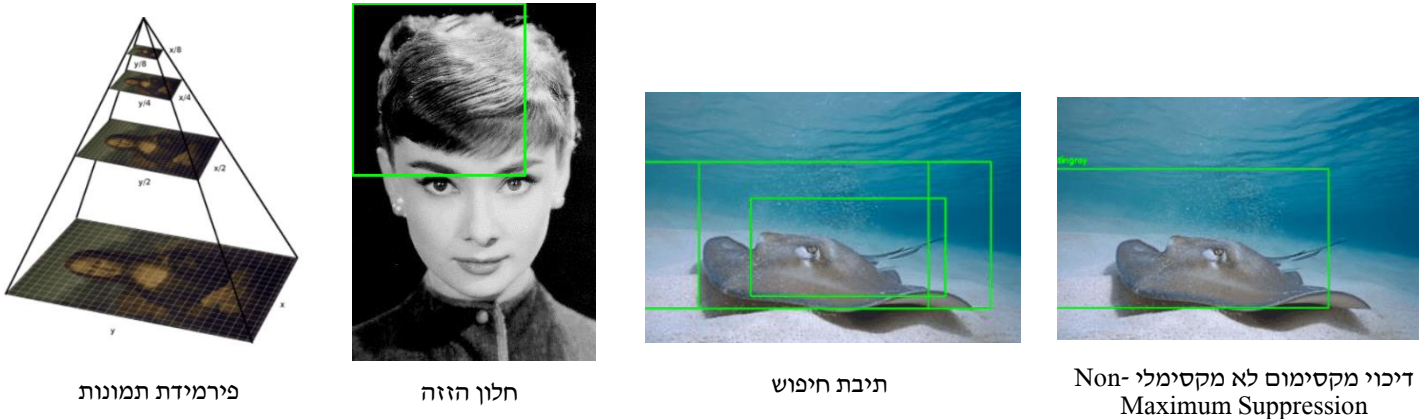
כדי לשפר את יכולת הכללה של המודל (generalizability) ולהפחית את הסיכון להתאמת יתר (overfitting), השתמשנו בטכניקת הגדלת נתונים (Data Augmentation). טכניקה זו יוצרת תמונות חדשות על ידי ביצוע טרנספורמציות קלות בתמונות המקוריות. השתמשנו ב-ImageDataGenerator של Keras, שהגדיר את הפרמטרים הבאים: **Rotation_range=10** - סיבוב התמונה עד 10 מעלות, **width_shift_range=0.1, height_shift_range=0.1** - הזזת התמונה אופקית ואנכית עד 10% מגודלה, **shear_range=0.1** - גזירת התמונה באופן אלכסוני עד 10%, **zoom_range=0.1** - הגדלה או הקטנה של התמונה עד 10%, **horizontal_flip=False** - בחרנו שלא לבצע היפוך אופקי של התמונות, כי במציאות באמת לא נקבל תמונות הפוכות, **fill_mode='nearest'** - קובע את אופן מילוי הפיקסלים החסרים לאחר טרנספורמציות. הערכים בפרמטרים הם מאוד קטנים ומזעריים, כי רוב התמונות לא מגיעות מסובבות או חתוכות, אלא בצורה ישרה, ואולי קצת מוטות בזווית מאוד קטנה. על ידי הגדלת מלאכותיות של מערך הנתונים, המודל נחשף לווריאציות שונות של אותיות, כגון סיבוב קל, הזזה או שינויים קלים בצורה. כתוצאה מכך, המודל לומד תכונות כלליות יותר של האותיות, ופחות רגיש לרעש בתמונות המקוריות. יישום ה-Data Augmentation הוביל לשיפור משמעותי בביצועי המודל. הדיוק עלה מ 90.04% ל-97.9%, וניתן לראות שיפור גם במדדי ה-Precision, Recall ו-F1 Score. תוצאות אלו מחזקות את יעילות הטכניקה במניעת התאמת יתר ושיפור יכולת הכללה של המודל. ניתן לראות לפי Confusion Matrix את הטעויות הבודדות בין האותיות ב-ג, כ-ץ ששתיהן אותיות עם תגים דומים.



בשלב הבא, הוספנו את מנגנון ה-Early Stopping. מנגנון זה מפסיק את תהליך האימון באופן אוטומטי כאשר הביצועים על מערך הוולידציה מתחילים להידרדר, לפני שהמודל מתחיל ללמוד את הרעש בנתונים ולא את התכונות האמיתיות. ובכך מונע התאמת יתר למערך האימון. שילוב של שתי הטכניקות הללו הביא לשיפור משמעותי נוסף בדיוק המודל, שהגיע ל-99.16%. תוצאה זו מצביעה על יעילותה של הגישה המשולבת בהגדלת נתונים ובמניעת התאמת יתר.

דיון:

בנוסף לזיהוי אותיות בודדות, נבחנה האפשרות לזהות את כל האותיות בקלף במקביל באמצעות גישה מבוססת זיהוי עצמים. גישה זו נועדה להפחית את הצורך בהכנה מוקדמת של תמונות האותיות ולספק תמונה כוללת של הקלף. שיטה זו מבוססת על השלבים הבאים: **פירמידת תמונות**: התמונה של הקלף כולו חולקה למספר רמות של רזולוציה. כל רמה מייצגת "פירמידה" של התמונה, כאשר הרמות הנמוכות מכילות פרטים ברזולוציה גבוהה והרמות הגבוהות מכילות פרטים ברזולוציה נמוכה. **חלון הזזה**: חלון בגודל קבוע "חולף" על פני כל רמה בפירמידה. בכל מיקום, החלון מכיל תת-תמונה קטנה. **הזנת תת-תמונה לרשת CNN**: כל תת-תמונה מוזנת לרשת CNN שהוכשרה לזיהוי אותיות. הרשת מחזירה את ההסתברות לכך שתת-תמונה מכילה אות מסוימת. **תיבת חיפוש**: אם ההסתברות עולה על סף מסוים, המערכת יוצרת תיבת חיפוש סביב האות הזוהתה. **זיכוי מקסימום לא מקסימלי (Non-Maximum Suppression)**: תהליך זה משמש כדי למזג תיבות חיפוש חופפות ולהשאיר רק את התיבות עם ההסתברות הגבוהה ביותר.



עם זאת, תוצאות הניסוי הראו כי גישה זו לא הביאה לשיפור משמעותי בביצועים. הסיבות לכך עשויות להיות מגוונות, ובהן: **מורכבות התמונה** - קלפי מזוזה ותפילין מציגים מגוון רחב של וריאציות חזותיות, כגון שינויים בגודל האותיות, זווית הכתיבה, סוג הנייר והדיו. **חפיפות בין אותיות** - במקרים רבים, אותיות חופפות זו עם זו, מה שמקשה על מערכת זיהוי העצמים להפריד ביניהן. **איכות תמונה נמוכה** - תמונות של קלפים עתיקים או פגומים עשויות להכיל רעש חזותי, מה שמקשה על זיהוי האותיות.



למרות האתגרים, גישה זו מציעה פוטנציאל רב לשיפור עתידי. על ידי שיפור איכות נתוני האימון, פיתוח אלגוריתמים יעילים יותר לזיהוי עצמים בתמונות מורכבות, והתאמה של הרשת לזיהוי טקסט עתיק, ניתן יהיה להשיג תוצאות טובות יותר.

בנוסף לגישות הקודמות, ניסינו לשפר את ביצועי המערכת באמצעות יישום מודל YOLO. YOLO - (You Only Look Once) הוא אלגוריתם זיהוי עצמים מהיר ומדויק. בניגוד לגישות מסורתיות שמבצעות סריקה של התמונה במספר שלבים, YOLO מבצעת את זיהוי העצמים בשלב אחד בלבד, מה שהופך אותה ליעילה מאוד. YOLO מחלק את התמונה לרשת של תאים. כל תא אחראי על חיזוי קופסאות חיפוש (bounding boxes) והסתברויות לכך שעצם מסוים נמצא בתוך התא. המודל מנבא גם את המחלקה של העצם.

מודל זה הותאם לזיהוי אותיות בקלפים על ידי יצירת מערך נתונים מותאם אישית, הכולל תמונות של קלפים עם תיוגים מדויקים של מיקום האותיות בקובץ YAML. למרות הפוטנציאל הגלום בגישה זו, התוצאות לא היו משביעות רצון. אנו מייחסים זאת לכמות המוגבלת של נתוני האימון.

מסקנות:

מחקר זה השווה שלוש שיטות שונות לזיהוי אותיות בקלפי מזוזה ותפילין: התאמת תבנית (Match Template), חישוב הפרשים בין גיבובים של תמונות (Image Hash) ורשתות נוירוניות קונבולוציוניות (CNN). תוצאות הניסוי הראו כי רשתות ה-CNN הציגו ביצועים משמעותיים טובים יותר משתי השיטות האחרות. רשת ה-CNN שהוכשרה במחקר זה השיגה דיוק של 99.16%, לעומת 94.24% עבור התאמת תבנית ו-90.04% עבור חישוב הפרשים בין גיבובים של תמונות. ממצא זה מחזק את ההנחה כי רשתות נוירוניות עמוקות, וביניהן רשתות CNN, מצטיינות במשימות זיהוי תבניות מורכבות, כגון זיהוי אותיות בקלפי מזוזה ותפילין עתיקים. יכולתן ללמוד מאפיינים מופשטים בתמונות מאפשרת להן להתמודד עם מגוון רחב של וריאציות חזותיות, כגון שינויים בגודל, זווית וכתיבה ידנית. הדיוק הגבוה שהושג במחקר פותח פתח ליישומים מעשיים בתחום זה, כגון: **אוטומציה של תהליך הבדיקה** - מערכת מבוססת CNN יכולה להחליף את הבדיקה הידנית של קלפים, להפחית טעויות אנוש ולהאיץ את התהליך. **שימור המורשת** - המערכת יכולה לסייע בשימור המורשת היהודית על ידי זיהוי וקטלוג של קלפים עתיקים. **פיתוח כלים חינוכיים** - ניתן להשתמש במערכת כדי לפתח כלים חינוכיים ללמידת הכתב העברי. תוצאות המחקר מעודדות המשך פיתוח של מערכות מבוססות למידת עומק לזיהוי טקסטים עתיקים ומסמכים היסטוריים. למרות ההצלחה שהושגה במחקר זה, ישנם מספר אתגרים שנוותרו פתוחים. למשל, זיהוי אותיות בקלפים פגומים או מטושטשים עדיין מהווה אתגר. בנוסף, פיתוח מערכת שתוכל לזהות קלפים בשפות עתיקות נוספות דורש מחקר נוסף. מחקר עתידי יוכל להתמקד בשיפור דיוק המערכת עבור קלפים באיכות נמוכה, פיתוח אלגוריתמים יעילים יותר לזיהוי טקסטים עקומים או מקוטעים, והרחבת היכולות של המערכת לזיהוי שפות עתיקות נוספות.