**ניסויים ומסקנות:**

את הבדיקות והניסויים אשר ביצענו למערכת ניתן לחלק ל2 סוגים עיקריים: ראשית רצינו לבדוק את טיבו של האלגוריתם שכתבנו ולהשוות את איכות הסיווג ביחס למערכת אחרת כאשר התנאים הם זהים ושנית, רצינו לבדוק את איכותו של ממשק המשתמש : האם השימוש הסטנדרטי במערכת הינו נוח, טבעי ומאפשר לבצע את הפעולות בצורה פשוטה יחסית וכמו כן גם לקרוא את הפלט כראוי.

כאמור, רצינו לבדוק את איכותו של האלגוריתם. לשם כך ביצענו ניסוי אשר בדק את יעילות ודיוק סיווג המערכת שלנו על קבוצת תמונות מסוימת לעומת מערכת אחרת שביצעה אותה מטרה.   
ראשית לשם הניסוי שתי המערכות ביצעו אימונים על אותו Train set (קובץ אשר הכיל אלפי סוגים של תמונות שונות מחמישה סוגים שונים של פרחים) אשר נקבע עוד מבעוד מועד. זאת ועוד , 2 המערכות הופעלו על חומרה זהה (אותו מחשב).  
לאחר אימונים שלקחו מספר שעות, יצרנו Test set (קובץ בדיקה בעל 50 תמונות – 10 מכל קבוצה – שונות) ועליו הרצנו את המערכת שלנו במטרה לקבל סיווג בעבור כל תמונה.

לאחר קבלת התוצאות מ2 המערכות השונות, נעזרנו בפרמטרים כדי לאמוד את רמת הדיוק של כל מערכת:  
**Precision :** דיוק התוצאות שהתקבלו לתוצאות האמת (כמות תוצאות הנכונות שחזרו מתוך כלל התוצאות שחזרו)  
**Recall :** מתייחס ל"רגישות" התוצאות (כמות התוצאות הנכונות שחזרו לעומת כמות התוצאות הנכונות שהיו אמורות להתקבל).  
חישוב פרמטרים אלו מתחיל בסיווג התוצאות ל4 תתי סוגים של תוצאות:  
**חיובי נכון (True positive):** תמונה אשר סווגה מסוג מסוים ואכן כזו.

**חיובי לא נכון (True negative):** תמונה אשר סווגה מסוג מסוים והיא למעשה לא כזו.

**שלילי נכון (False positive):** תמונה אשר סווגה כאחת מסוג אחר ואכן היא כזו.

**שלילי לא נכון (False negative):** תמונה אשר סווגה כאחת מסוג אחר והיא מהסוג הנכון.  
ולאחר הרצת המערכת ביצענו את חישוב הפרמטרים ע"פ הנוסחאות הבאות:

להלן מוצגים הסיווגים השונים שהתקבלו בעבור המערכות השונות:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| System A | | | | | |
|  | Daisy | Dandelion | Tulip | Rose | Sunflower |
| True Positive | 10 | 8 | 9 | 9 | 10 |
| False Positive | 0 | 0 | 1 | 1 | 2 |
| True negative | 40 | 40 | 39 | 39 | 38 |
| False negative | 0 | 2 | 1 | 1 | 0 |
| Precision | 1 | 1 | 0.9 | 0.9 | 0.8334 |
| Recall | 1 | 0.8 | 0.9 | 0.9 | 1 |

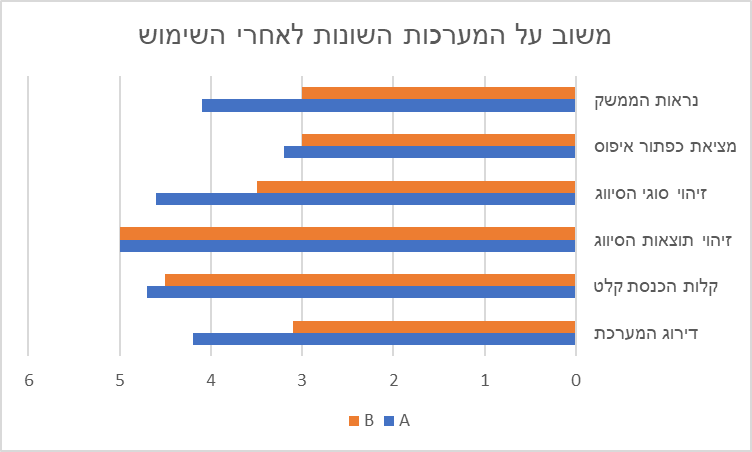
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| System B | | | | | |
|  | Daisy | Dandelion | Tulip | Rose | Sunflower |
| True Positive | 10 | 9 | 10 | 0 | 6 |
| False Positive | 1 | 2 | 10 | 0 | 0 |
| True negative | 39 | 38 | 30 | 40 | 40 |
| False negative | 0 | 1 | 0 | 10 | 4 |
| Precision | 0.8334 | 0.75 | 0.5 | 0 | 1 |
| Recall | 1 | 0.9 | 1 | 0 | 0.6 |

לאחר ריכוז התוצאות הצגנו את הנתונים השונים בגרף של precision כתלות בrecall לצורך בדיקת טיב המערכת. (כאשר דגמנו בעבור כל סוג פרח את הנתונים הנ"ל)

מנתוני הגרף ניתן לראות כי מערכת A הינה יציבה ומחזירה תוצאות מדויקות יותר לעומת המערכת השנייה. היציבות באה לידי ביטוי בכך שבעבור כל סוג של פרח שאובחן במערכת הראשונה ניתן לקבל דיוק יחסית זהה בעוד שהמערכת השנייה מאוד מציגה נתונים לא יציבים וקיצוניים בעבור כל סוג פרח. בכך איששנו את השערת הניסוי ואכן המערכת הראשונה יציבה יותר ומדויקת יותר מהשנייה.

לטובת בדיקת איכות הממשק, יצרנו והעברנו שאלון מקדים באוכלוסיית הנבדקים כדי לאפיין את שגרת שימושם. בשאלון התייחסנו לשימוש במכשיר הפלאפון: מהו השימוש השכיח ביותר, תדירות השימוש במכשיר והכרת המכשיר. בנוסף בדקנו את שגרת שימושו בעולם הפרחים: מהי תדירות שבה מטייל ומהי מידת ההיכרות שלו עם סוגי הפרחים.  
לאחר מכן ביצענו סדרת פעולות אשר קיימת ב2 המערכות :

* 1. בחרו תיקייה ובה תמונות הפרחים אותם תרצו לסווג.
  2. בחרו את המודל הרצוי לסיווג (קובץ מסוג h5).
  3. הריצו את מערכת הסיווג.
  4. לאחר החזרת התוצאות:
     1. בחרו סוג פרח מסוים, כמה תמונות הן מהסוג הנבחר?
     2. כמה סוגי פרחים ישנם?
  5. בצעו reset למערכת.

לאחר ביצוע סדרת הפעולות הנ"ל בדקנו בעזרת שאלון נוסף את שביעות רצונם של המשתמשים השונים. להלן התוצאות:

**סיכום:**