|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| \\dsp-disk.eng.tau.ac.il\Documents\eeproj\Forms & Important\פרויקטים\הנחיות\Guidelines\Students\ENG_LOGO-01.png | | | **\\dsp-disk.eng.tau.ac.il\Documents\eeproj\Forms & Important\פרויקטים\הנחיות\Guidelines\Students\TAU_EngineeringENG.png** | |
| זיהוי רגשות מהבעות פנים מתמונה או וידאו | | | |
| פרויקט מס' 17-1-1-1406  דו"ח סיכום | | | |
| מבצעים: | | | |
|  | גל קשי | 204572861 | |
|  | חן אילון | 201617032 | |
| מנחים: | | | |
|  | פרופ' אמיר גלוברזון | אוניברסיטת ת"א | |
| מקום ביצוע הפרויקט: אוניברסיטה | | | |

# 

**תוכן עניינים**

[רשימות איורים וטבלאות 3](#_Toc516679421)

[תקציר 4](#_Toc516679422)

[הקדמה 5](#_Toc516679423)

[רקע תיאורטי 7](#_Toc516679424)

[1.1 רגשות "גלובליים" 7](#_Toc516679425)

[1.2 מאגרי מידע מתוייגים 7](#_Toc516679426)

[1.3 פיצ'רים 8](#_Toc516679427)

[1.4 אלגוריתמים שחקרנו 9](#_Toc516679428)

[1.4.1 Machine Learning - 9](#_Toc516679429)

[1.4.2 Deep Learning- 9](#_Toc516679430)

[סימולציות ובדיקות 10](#_Toc516679431)

[1.5 מציאת פרמטרים לאלגוריתמי Machine Learning שאינם Deep learning 10](#_Toc516679432)

[1.6 מחקר אלגוריתמים וסיווג רגשות - 11](#_Toc516679433)

[1.7 סיווג רגשות Multi Label- 13](#_Toc516679434)

[1.8 נרמול פיצ'רים ביחס להבעה ניטרלית 15](#_Toc516679435)

[מימוש 16](#_Toc516679436)

[1.9 תיאור תוכנה 16](#_Toc516679437)

[1.10 תרחישים נפוצים 19](#_Toc516679438)

[1.11 פירוט נוסף 21](#_Toc516679439)

[ניתוח תוצאות 24](#_Toc516679440)

[סיכום, מסקנות והצעות להמשך 25](#_Toc516679441)

[מקורות 26](#_Toc516679442)

# רשימות איורים וטבלאות

**רשימת איורים**

[איור 1 - בלוקים אבסטרקטיים של המערכת 4](#_Toc516679386)

[איור 2 8](#_Toc516679387)

[איור 3 10](#_Toc516679388)

[איור 4 10](#_Toc516679389)

[איור 5 11](#_Toc516679390)

[איור 6 - תרשים בלוקים של המערכת 16](#_Toc516679391)

[איור 7 - תרחישים באפלקציה (1) 19](#_Toc516679392)

[איור 8 - תרחישים באפלקציה (2) 20](#_Toc516679393)

[איור 9 21](#_Toc516679394)

[איור 10 22](#_Toc516679395)

[איור 11 22](#_Toc516679396)

**רשימת טבלאות**

[טבלה 1 - הרגשות של אקמן 7](#_Toc516679397)

[טבלה 2 - תכולה של מאגר Affectnet 8](#_Toc516679398)

[טבלה 3 -תוצאות מחקר (1) 11](#_Toc516679399)

[טבלה 4- תוצאות מחקר (2) 12](#_Toc516679400)

[טבלה 5 - תוצאות מחקר (3) 12](#_Toc516679401)

[טבלה 6 - תוצאות מחקר (4) 12](#_Toc516679402)

[טבלה 7 - תוצאות מחקר (5) 13](#_Toc516679403)

[טבלה 8 - תוצאות מחקר (6) 13](#_Toc516679404)

[טבלה 9 - תוצאות מחקר (7) 14](#_Toc516679405)

[טבלה 10 - תוצאות מחקר (8) 14](#_Toc516679406)

[טבלה 11 24](#_Toc516679407)

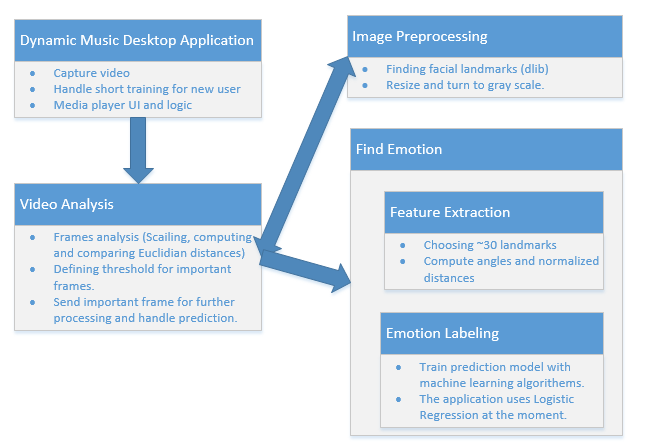
# תקציר

הפרויקט שעוסק בזיהוי רגשות מהבעות פנים מתמונה ומוידאו.

מטרת הפרויקט היא להצליח להבחין בין 6 רגשות מוידאו שמכיל הבעות פנים שונות בReal-Time: שמחה, עצב, הפתעה, פחד, כעס, גועל.

תוצר הפרויקט הוא אפליקציית מוזיקה דינאמית, שמשנה את השירים שהיא משמיעה בהתאם להרגשת המשתמש, מבין הרגשות המפורטים לעיל.

להלן דיאגרמת בלוקים של המערכת -



איור 1 - בלוקים אבסטרקטיים של המערכת

# הקדמה

חיי היום-יום שלנו מבוססים בחלק גדול מהם על תקשורת - בין אם מדובר בהתעדכנות בחדשות, כתיבת "חלב" ברשימת קניות או בדיקה פשוטה אצל רופא.

בני-אדם היו מתקשרים בעיקר עם בני אדם אחרים, או עם בעלי חיים - כאשר היה צורך להעביר מידע כלשהו בינהם.

אז כאשר בן אדם מסויים מדבר אל בן אדם אחר - איזה מידע עובר לבן אדם השני?

"The most important thing in communication is hearing what isn’t said.” – Peter F. Drucker

בנוסף למילים עצמן, האדם מדבר בטון מסויים, ופניו מביעות הבעה מסויימת.

המידע הנוסף הזה מאפשר לבן אדם המאזין לתת הקשר ומשמעות אמיתית למילים עצמן (לדוגמה - הבנת ציניות, כאב, אושר).

כיום, בני אדם מתקשרים יותר ויותר עם מכונות - אך מכונות יודעת לעבד בעיקר data שמכיל מילים, או אותות קבועים מראש.

ההגבלה הזו מונעת ממכונות להיות יכולות לתקשר עם בני אדם באופן שוטף, או, לבצע פעולות רגישות - כגון פיזיותרפיה. נרחיב מעט על דוגמה זו - נניח כי היה רובוט, שהיה יודע לעשות תרגילי פיזיותרפיה מסוימים עם בני אדם. הרובוט היה יכול לעבד פקודות מסוימות - אך הוא לא היה יכול להבחין אם פני האדם מביעות כאב. מכיוון שבני אדם מביעים רגשות באמצעות הבעות פנים - ולא באמצעות פקודות למכונה - הרובוט לא היה יודע מתי צריך להפסיק, והיה עלול לגרום למטופל פגיעה פיזיולוגית.

היינו רוצים שמכונות ידעו לאבד את כל המידע שמועבר בתקשורת עם משתמש, גם את המידע שאינו וורבלי.

כאן נכנס התחום של זיהוי רגשות אוטומטי.

זיהוי רגשות של בני אדם ע'ב הבעת הפנים שלהם זו אינה משימה פשוטה, לעיתים גם לבני אדם.

מטרת הפרוייקט היא לבנות אפלקציה שתדע לפרש את הבעות פניו של אדם ואת השינויים בהם כתגובה לגירויים חיצוניים.

הפרוייקט כולל זיהוי פנים, הבעות ורגשות מתמונה או וידאו.

על מנת לבצע את הפרוייקט אנו צריכים לעבור מספר שלבים:

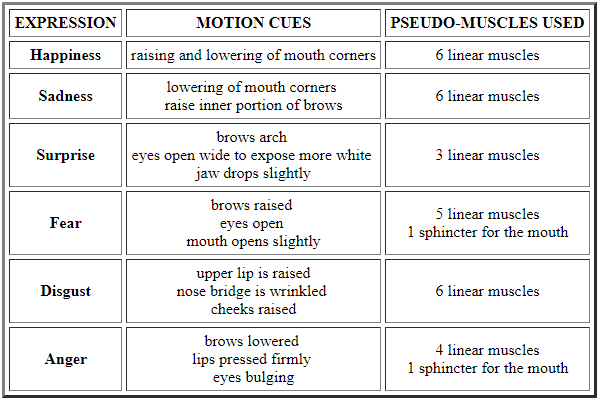
* מציאת Dataset מתוייג מתאים.
* עיבוד תמונה - מציאת פנים בתמונה, חתיכת תמונה והפיכה לשחור לבן.
* מציאת פיצ'רים בפנים, חישוב אוסף פיצ'רים ממנו נרצה ללמד את האלגוריתם וטיובו.
* בחירת אלגוריתם macine learning ואימונו - לימוד אלגוריתמים שונים והשוואת תוצאות חיזוי.
* עיבוד וידאו - הוצאת פריימים בתמונה ועיבוד בזמן אמת.
* אפלקציית מוזיקה דינאמית - אפלקציית desktop.

# 

# רקע תיאורטי

## רגשות "גלובליים"

מה הם רגשות "גלובליים"? אלו הם רגשות שמובעים ע"י הבעות פנים קבועות בתרבויות שונות. עפ"י מחקר שהתבצע ע"י Paul Ekman (1960) קיימים 6 רגשות "גלובליים": Joy, Surprise, Sadness, Anger, Disgust, Fear



טבלה 1 - הרגשות של אקמן

קיים רגש נוסף שהוא שנוי במחלוקת - contempt.

אקמן הרחיב את התיאוריה שלו והתעסק במיקרו הבעות פנים (1990). במחקר הזה הוא הוסיף זיהוי של רגשות נוספים - amusement, contentment, embarrassment, excitement, guilt, pride in achievement, relief, satisfaction, sensory pleasure, and shame..

## מאגרי מידע מתוייגים

מצאנו 4 מאגרי מידע מתוייגים המכילים תמונות ראש עם הבעות פנים, ותיוג של רגש:

* **AffectNet** - מכיל כ-400,000 תמונות שתויגו באופן ידני מהן (מקור [1]):

|  |  |
| --- | --- |
| Label | Exist |
| Neutral | 74,874 |
| Happy | 134,416 |
| Sad | 25,459 |
| Surprise | 14,090 |
| Fear | 6,378 |
| Disgust | 3,803 |
| Anger | 24,882 |
| Contempt | 3,750 |
| Non faces | 112,348 |

טבלה 2 - תכולה של מאגר Affectnet

* **Extended Cohn-Kanade (CK+)**  - מכיל מאות תמונות מתויגות של רגשות. (מקור [2])

המאגר מכיל תמונות של 123 אנשים - כך שלכל מטופל יש בין 2 ל8 סשיינים של רגשות, מהרגשות הבאים - Neutral, Contempt, Happy, Sad, Surprise, Fear, Anger and Disgust.

* **Multimedia Understanding Group (MUG)** - (מקור [3])

המאגר מכיל תמונות של 52 אנשים עבור הרגשות הבאים -Neutral, Happy, Sad, Surprise, Fear, Anger and Disgust.

* **Oulu CASIA NIR&VIS facial expression database** - (מקור [4])

המאגר מכיל תמונות של 80 אנשים עבור הרגשות הבאים -Neutral, Happy, Sad, Surprise, Fear, Anger and Disgust.

## פיצ'רים

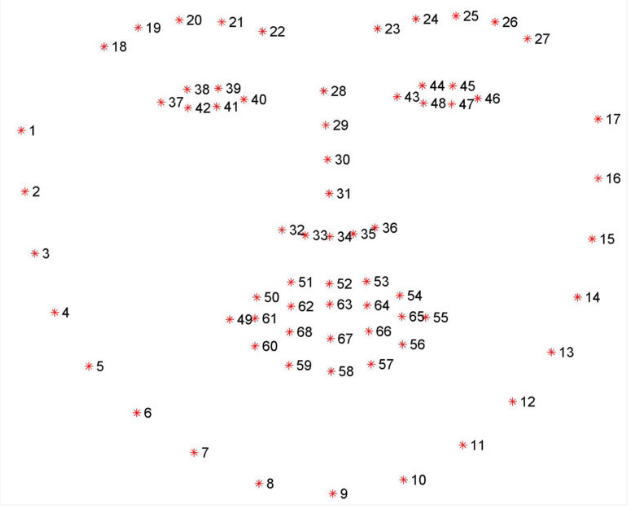
את הבעיה של להצמיד רגש לתמונה המכילה הבעת פנים אפשר לפצל ל2 בעיות קטנות יותר:

* מציאת פנים, וחלקים מסויימים בפנים בתמונה -

על מנת למצוא פנים, וחלקים מסוימים בפנים השתמשנו ברשת מאומנת של סיפריה בשם DLIB.

* הגדרת וקטור מייצג מתמונה של פנים, שיאפשר שיוך לרגש, בעזרת אלגוריתם -

בשביל להגדיר וקטור של פיצ'רים השתמשנו בחישוב זוויות ומרחקים בין נקודות מסויימות באיור הבא.



איור 2

## אלגוריתמים שחקרנו

### Machine Learning -

* SVM
* KNN
* Logistic Regression
* Random Forest

### Deep Learning

* Basic CNN (conv2)
* Dense NN
* sklearn.neural\_network.MLPClassifier

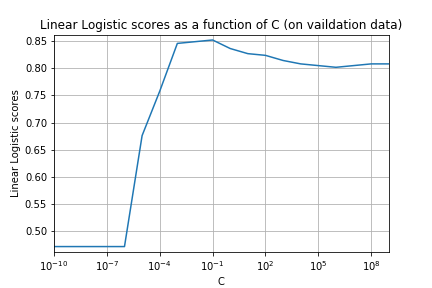
# סימולציות ובדיקות

## מציאת פרמטרים לאלגוריתמי Machine Learning שאינם Deep learning

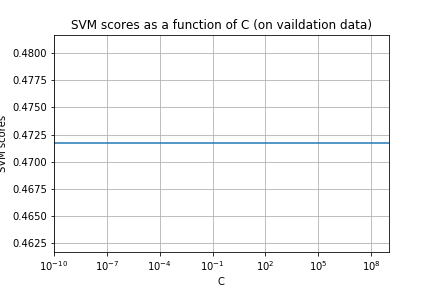
בחרנו לבדוק 3 אלגוריתמים לומדים: SVM, KNN, Logistic-regression.

לכל אחד מהאלגוריתמים האלו יש פרמטר שעלינו לקבוע בצורה אופטימלית ( C עבור SVM ו- Logistic-regression, K עבור KNN).

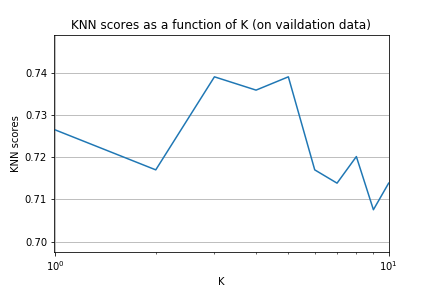
הרצנו את האלגוריתמים על 635 תמונות המכילות 8 רגשות (7 של אקמן והבעה ניטרלית), באופן לא מאוזן. בשביל למצוא את הפרמטרים האופטימלים ביצענו cross validation, וקיבלנו את התוצאות הבאות:



איור 3



איור 4



איור 5

ניתן לראות כי את התוצאה הטובות ביותר קיבלנו עבור:

* Logistic Regression - C=0.1, 85% דיוק על הValidation data.
* KNN - K=3, 74% דיוק על הValidation data.
* SVM - C= Any, 47% דיוק על הValidation data.

## מחקר אלגוריתמים וסיווג רגשות -

מכיוון שלא מצאנו הרבה תמונות של "בוז", בחרנו להסתכל על ההפרדה של האלגוריתמים ל7 קטגוריות, בלעדיו:



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Log Reg** | **KNN** | **SVM** | **Dense NN 1** | **Dense NN 2** | **Random Forest** | **CNN** |
| **all (7)** | 49.5 | 29.2 |  | 49.2 | 14.2 | 44 | 46 |
| **0,1,2,3,5** | 65.9 | 49.1 |  |  |  |  | 59.2 |
| **1,2** | 89.75 | 79.25 | 87.75 | 50 | 88 | 91 |  |
| **1,2,5** | 74.16 | 57 | 67.6 |  |  |  |  |
| **1,2,3,6** |  |  |  |  |  |  | 67.7 |
| **0,1,2** | 72.6 | 59.9 | 67.6 |  |  | 66 |  |
| **2,4,5,6** | 76.125 | 71.75 | 75 |  |  | 52 |  |

טבלה 3 -תוצאות מחקר (1)

הטבלה לעיל מכילה אחוזי זיהוי עבור בעיית למידה ממוחשבת עם מספר מחלקות.

ניתן לראות בראשית כל שורה אילו מספרי מחלקות השתתפו באימון ובבחינה, ובראש כל עמודה את האלגוריתם שנעשה בו שימוש.

המספרים בטבלה הם אחוזי החיזוי על קבוצת המבחן.

בחלק זה הData עליו אומנו האלגוריתמים הינו מאוזן. כל מחלקה הכילה 2000 תמונות - כך ש90% מהתמונות בקבוצת האימון ו-10% בקבוצת המבחן.

* ניתן לראות כי לאחר שהData אוזן, אחוזי הזיהוי צנחו למתחת ל-50%. (בסעיף הקודם כ-70% מהתמונות היו של מחלקת happy).
* כמו כן, ניתן לראות כי ההפרדה בין הבעה שמחה לעצובה היא קלה באופן יחסי - וצולחת בכ-90% מהמקרים.
* ניתן לראות כי הבעות ניטרליות "מבלבלות" את האלגוריתמים.

בדיקה נוספת שעשינו היא חלוקת הבעות פנים ל-2 מחלקות: חיובי ושלילי.

להלן מספר קומביציות שניסינו:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Log Reg** | **KNN** | **Random Forest** |
| **1 vs 2456** | 91 | 88 | 91 |
| **013 vs 2456** | 72.7 | 68 |  |
| **01 vs 2456** | 76.7 | 71.7 |  |
| **13 vs 2456** | 79.5 | 74.4 |  |

טבלה 4- תוצאות מחקר (2)

גם כאן דאגנו לאזן את הData, כן שב-2 המחלקות יהיו סה"כ אותה כמות תמונות.

ניתן לראות כי בעיית ההפרדה בין פרצופים חיוביים לשלילים היא קלה יותר מהבעיה לעיל.

בדקנו גם דיוק לפי 2 התיוגים הטובים ביותר - מסווג ה-CNN מדרג עבור כל תמונה את הרגשות האפשריים לפי ההסתברות שזהו הרגש המובע. הגדרנו סיווג כנכון אם הרגש נמצא בשני המקומות הראשונים בדירוג, וקיבלנו את אחוזי הדיוק הבאים:

|  |  |
| --- | --- |
|  | **CNN** |
| **0,1,2,3,6** | 78.3 |
| **1,2,3,6** | 88.2 |

טבלה 5 - תוצאות מחקר (3)

התפלגות הזוגות שסווגו כ-2 הטובים ביותר, כלומר זוגות הרגשות שנפוצים שנחזו יחד במקומות הראשונים, בהרצה על **0,1,2,3,6**, באחוזים:

|  |  |
| --- | --- |
| **Labels** | **Percentage** |
| 0, 1 | 5.48 |
| 0, 2 | 11.48 |
| 0, 3 | 11.24 |
| 0, 6 | 14.1 |
| 1, 2 | 6.42 |
| 1, 3 | 12.52 |
| 1, 6 | 4.36 |
| 2, 3 | 11.32 |
| 2, 6 | 17.54 |
| 3, 6 | 5.54 |

טבלה 6 - תוצאות מחקר (4)

## סיווג רגשות Multi Label-

על מנת לחקור את ההבדל בין המחלקות, הסתכלנו על סיווג Multi Label.

המוטיבציה לעשות זאת היא בדיקה האם סקאלת הרגשות הם כמו פלטה של צבעים, ולא מחלקות זרות ובדידות. כלומר, האם פנים ניטרליות הן למעשה פנים שהם בסיכוי שווה שמחות ועצובות.

בשביל לענות על השאלה הזו, השתמשנו ב2 אלגוריתמים:

1. Random Forest עם 100 עצים.

ראשית הסתכלנו על אימון עבור כל 7 המחלקות:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Neutral** | **Happy** | **Sad** | **Surprise** | **Fear** | **Disgust** | **Anger** |
| **Neutral** | 0.149  (std 0.0195) | 0.1163  (std 0.0272) | 0.1715  (std 0.0212) | 0.0623  (std 0.0162) | 0.1126  (std 0.0218) | 0.1747  (std 0.0195) | 0.2133  (std 0.0272) |
| **Happy** | 0.1330  (std 0.0160) | 0.1857  (std 0.0198) | 0.1753  (std 0.0205) | 0.0847  (std 0.0154) | 0.1012  (std 0.0214) | 0.1539  (std 0.0242) | 0.1659  (std 0.0252) |
| **Sad** | 0.1578  (std 0.0249) | 0.1078  (std 0.0310) | 0.1654  (std 0.0253) | 0.0682  (std 0.0190) | 0.1227  (std 0.0267) | 0.1710  (std 0.0256) | 0.2067  (std 0.0294) |
| **Surprise** | 0.1889  (std 0.0330) | 0.0646  (std 0.0218) | 0.1393  (std 0.0218) | 0.1205  (std 0.0151) | 0.1526  (std 0.0246) | 0.1624  (std 0.0235) | 0.1713  (std 0.0277) |
| **Fear** | 0.1326  (std 0.0265) | 0.1644  (std 0.0278) | 0.1646  (std 0.0242) | 0.0864  (std 0.0197) | 0.1437  (std 0.0342) | 0.1476  (std 0.0230) | 0.1606  (std 0.0272) |
| **Disgust** | 0.1403  (std 0.0282) | 0.1279  (std 0.0313) | 0.1769  (std 0.0238) | 0.0441  (std 0.0164) | 0.0987  (std 0.019) | 0.1790  (std 0.028) | 0.2328  (std 0.0272) |
| **Anger** | 0.1622  (std 0.0295) | 0.1231  (std 0.0348) | 0.1851  (std 0.0215 ) | 0.056  (std 0.0175) | 0.0972  (std 0.0214) | 0.1698  (std 0.0176) | 0.2063  (std 0.0333) |

טבלה 7 - תוצאות מחקר (5)

מסקנות מהטבלה לעיל:

* ניתן לראות כי כל הרגשות הם בסיכוי יחסית גבוה כועסים, נגעלים או עצובים.
* ההתפלגויות של הבעות עצובות וניטרליות מאוד דומות.

שנית, הסתכלנו על הפרדה בין 3 מחלקות: ניטרלי, שמח ועצוב:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Neutral** | **Happy** | **Sad** |
| **Neutral** | 0.2864  (std 0.0302) | 0.2571  (std 0.0243) | 0.4564  (std 0.0263) |
| **Happy** | 0.2448  (std 0.0208) | 0.3906  (std 0.0302) | 0.3646  (std 0.0287) |
| **Sad** | 0.2932  (std 0.0356) | 0.2429  (std 0.0287) | 0.4638  (std 0.0261) |

טבלה 8 - תוצאות מחקר (6)

גם באימון הזה ניתן לראות שההתפלגות של פנים עצובות וניטרליות היא כמעט זהה.

כלומר, פנים ניטרליות מאוד דומות לפנים עצובות.

1. אלוגריתם Deep Learning: MLPClassifie.

ראשית מצאנו כי הalpha האופטימלית היא 0.02 (בעזרת Cross validation).

הסתכלנו על אימון עבור כל 7 המחלקות:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Neutral** | **Happy** | **Sad** | **Surprise** | **Fear** | **Disgust** | **Anger** |
| **anger** | 0.3206 (std 0.1853) | 0.0153 (std 0.0311) | 0.1053 (std 0.0638) | 0.0108 (std 0.0199) | 0.0375 (std 0.0463) | 0.2097 (std 0.1364) | 0.3007 (std 0.1481) |
| **disgust** | 0.0889 (std 0.0972) | 0.0131 (std 0.0310) | 0.0494 (std 0.0409) | 0.0022 (std 0.0025) | 0.0081 (std 0.0088) | 0.6736 (std 0.1602) | 0.1646 (std 0.1070) |
| **fear** | 0.0885 (std 0.0523) | 0.0895 (std 0.1505) | 0.1742 (std 0.1269) | 0.0828 (std 0.1152) | 0.4297 (std 0.2093) | 0.1146 (std 0.1425) | 0.0208 (std 0.0274) |
| **happy** | 0.0112 (std 0.0247) | 0.8854 (std 0.1013) | 0.0296 (std 0.0317) | 0.0129 (std 0.0244) | 0.0226 (std 0.0235) | 0.0336 (std 0.0331) | 0.0046 (std 0.0066) |
| **neutral** | 0.5737 (std 0.1710) | 0.0077 (std 0.0143) | 0.1004 (std 0.0664) | 0.0296 (std 0.0300) | 0.0868 (std 0.0651) | 0.1126 (std 0.1231) | 0.0892 (std 0.0721) |
| **sad** | 0.3788 (std 0.1510) | 0.0013 (std 0.0013) | 0.2631 (std 0.1221) | 0.0211 (std 0.0260) | 0.1020 (std 0.0783) | 0.1271 (std 0.0945) | 0.1067 (std 0.1123) |
| **surprise** | 0.0486 (std 0.0940) | 0.0006 (std 0.0032) | 0.0163 (std 0.0151) | 0.3566 (std 0.1598) | 0.5218 (std 0.1589) | 0.0410 (std 0.0386) | 0.0151 (std 0.0162) |

טבלה 9 - תוצאות מחקר (7)

התוצאות בהתפלגות הזו מפרידות את הData יותר מאלו שהתקבלו בRandom Forest.

ניתן לראות כי פה יותר הבעות עצובות נתפסו כניטרליות מאשר כעצובות.

הסתכלנו על הפרדה בין 3 מחלקות: ניטרלי, שמח ועצוב:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Neutral** | **Happy** | **Sad** |
| **happy** | 0.0215 (std 0.0474) | 0.8928 (std 0.1192) | 0.0858 (std 0.0892) |
| **neutral** | 0.6661 (std 0.1603) | 0.0141 (std 0.0209) | 0.3198 (std 0.1650) |
| **sad** | 0.30 (std 0.1636) | 0.0018 (std 0.0026) | 0.6982 (std 0.1649) |

טבלה 10 - תוצאות מחקר (8)

התוצאות האלו ניראות טובות יותר.

אך עדיין ניתן לראות שפחות מ70% מהפרצופים הניטרלים והעצובים מפורשים כמו שצריך.

## נרמול פיצ'רים ביחס להבעה ניטרלית

השערה שהעלנו במהלך הבדיקות לעיל היא שתווי פנים של אנשים שונים מספיק בשביל לבלבל חזאי רגשות.

כלומר, נניח שקיבלנו תמונה של אדם לא מוכר, בהבעה לא ידועה, ואנו יודעים שהמרחק בין העין לגבה שלו הוא x - החזאי לא ידע בהכרח אם האיש מרים גבה, או שזהו המצב הניטרלי שלו.

כדי לבדוק זאת יש לנרמל וקטור של פיצ'רים שחושבו מהבעת פנים לא ידועה, עם וקטור פיצ'רים של ההבעה הניטרלית של אותו אדם.

עבור כ-200 נבחנים, הרצנו נרמול פיצ'רים ואימון של 2 אלגוריתמים:

* Logistic Regression - הC האידיאלי עפ"י cross validation הוא 0.1.

הפרדה של 6 רגשות: Happy, Sad, Surprise, Fear, Anger and Disgust הצליחה ב-**85%.**

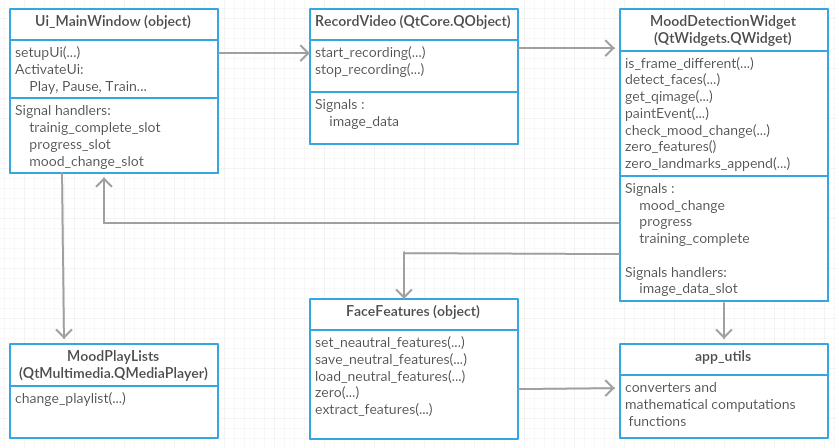
* MLPClassifier - ה האידיאלי עפ"י cross validation הוא 1.

הפרדה של 6 רגשות: Happy, Sad, Surprise, Fear, Anger and Disgust הצליחה ב-**85%.**

זוהי שיטת החיזוי המוצלחת ביותר שמצאנו, ובה השתמשנו באפלקציית המוזיקה הדינאמית.

# מימוש

להלן דיאגראמת בלוקים של הפרויקט:



איור - תרשים בלוקים של המערכת

## תיאור תוכנה

התוצר הסופי של הפרוייקט הוא אפלקציית Desktop.

האפלקציה היא נגן מוזיקה דינאמי, שמשנה את הPlaylist שהוא מנגן בהתאם לרגש של המשתמש. אנו מפרידים בין 6 סוגי רגשות: שמח, עצוב, מופתע, מפחד, כועס ונגעל.

שפת מימוש - פייתון.

בלוקים בתרשים לעיל:

* **Ui\_MainWindow -**

יורשת מ - object

ספרית UI - PyQt5

תיאור - המחלקה אחראית על ממשק האפלקציה.

האפלקציה מאפשרת למשתמש -

* + - אימון ראשוני למשתמש, כולל הצגת Progress Bar.
    - ניגון (play, pause) מוזיקה בהתאם להבעת הפנים של המשתמש.
    - ניגון שיר לפי מיקום על פס תזוזה, והצגת התקדמות שיר.
    - תצוגה של תמונת הפנים שהמערכת עורכת (סימון ריבוע סביב הפנים).

עיצוב בעזרת הסיפריה qdarkstyle.

* **MoodPlaylists -**

יורשת מ - Qmultimedia.Qmediaplayer

תיאור - המחלקה מבוססת על אובייקט Media Player של PyQt. המחלקה מייצרת ושומרת 6 סוגים של playlistים, עפ"י הרגשות המסווגים.

בעת שינוי רגש, המחלקה מאפשרת שינוי פלייליסט לפי הרגש הרלוונטי.

* **app\_utils -**

תיאור - המחלקה מכילה פונקציות עזר לחישובים הדרושים בMoodDetectionWidget ובFaceFeatures. פונקציות העזר מתמקדמות בהמרות של Typeים, ובחישוב מרחקים וזוויות בין וקטורים.

בנוסף המחלקה מכילה פונ' עיבוד תמונה בסיסי הנעזר בסיפריה OpenCV - הפונקציה משנה את גודל התמונה לגודל אחיד, ומעבירה אותה לשחור-לבן.

* **RecordVideo -**

יורשת מ - QCore.Qobject

תיאור - המחלקה אחראית על הקלטת וידאו ושליחת סיגנאל התמונה לעיבוד. בעת אימון המערכת, או לחיצה על Play, האפלקציה מתחילה סשן הקלטה של מצלמת המשתמש, בעזרת timer של PyQt.

(סשן הצילום נפסק כאשר מפסיקים את הtimer).

כל Frame שנקלט במצלמה משודר כסיגנאל למחלקה MoodDetectionWidget, שם הוא עובד עיבוד מתקדם יותר.

* **FaceFeatures -**

יורשת מ - object

תיאור - המחלקה מאפשרת הוצאה של וקטור פיצ'רים מאוסף נקודות על הפנים. השימוש במחלקה זו נעשה בשלב הtraining ובשלב בו רוצים להעריך רגש של frame.

המרחקים בין הlandmarks מחושבים כמרחק אוקלידי, ומנורמלים לפי מרחק קבוע בפנים של אדם.

הזוויות מחושבת בעזרת נוסחה לcos של זווית.

וקטור הפיצ'רים הסופי של frame יהיה וקטור הפיצ'רים המחושב במחלקה, פחות וקטור הפיצ'רים שחושב עבור הבעת הפנים הניטרלית של המשתמש. באופן זה המערכת יכולה לעכוב אחר התזוזות בפניו של אדם ספיציפי.

* **MoodDetectionWidget -**

יורשת מ - Qwidgets.Qwidget

תיאור ואלגוריתמים - זו היא המחלקה המרכזית בפרוייקט.

המחלקה מקבלת סיגנאל של frame ומכילה את החלקים הבאים -

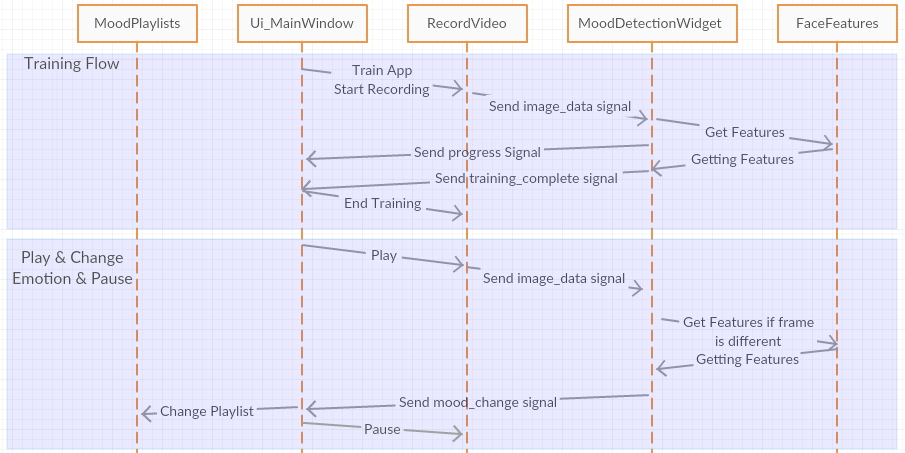
* זיהוי פנים ותחימתם בריבוע - עיבוד תמונה בעזרת OpenCV, מציאת פנים ו-68 נקודות (landmarks) בפנים בעזרת רשת מאומנת מסיפריה בשם dlib.
* מציאת frameים חשודים בשינוי רגש באופן יעיל - על מנת להגדיר frame כ"חשוד", נרצה שהמרחק שלו מהframe החשוב האחרון יעבור delta מסויימת. מרחק בין frameים יחושב באופן הבא -

לכל frame נחשב וקטור של כל המרחקים האוקילידים בין הlandmarks וננרמל אותם לפי אורך קבוע בפנים. המרחק בין הframeים יהיה המרחק האוקילידי בין הוקטורים האלו.

* חישוב פיצ'רים מלא לframeים החשודים בשינוי רגש - חישוב פיצ'רים נעשה ע"י המחלקה FaceFeatures.
* חיזוי רגש - חיזוי הרגש נעשה באמצעות אלגוריתם logistic regression שאימנו, כפי שתואר בפרקים הקודים. האלגוריתם מבדיל בין 6 מחלקות של רגשות, וצודק ב85% מהמקרים.
* מנגנון המחליט אם נעשה שינוי ברגש ואם צריך להעשות שינוי ברשימת ההשמעה - מכיוון שאיננו רוצים שהאפלקציה תשנה שיר בשל שינוי של מיקרו-הבעת פנים, אנו דורשים 2 פריימים שחשודים באותו רגש על מנת להחליף רגש.
* ניהול Training ושליחת אותות עדכון התקדמות שלו.

## תרחישים נפוצים

להלן דיאגרמת רצף של 3 תרחישים נפוצים באפלקציה:



איור 7 - תרחישים באפלקציה (1)

תרחיש ראשון - Training -

לפני תחילת השימוש באפלקציה, המשתמש צריך לעבור תהליך אימון קצר, בו המערכת שומרת את הפיצ'רים של ההבעה הניטרלית של אותו אדם.

השלבים של התהליך הם:

1. נשלחת בקשת Training מממשק המשתמש למחלקה RecordVideo.
2. RecordVideo מתחילה להקליט ולשלוח סיגנאלי פריימים לMoodDetectionWidget
3. MoodDetectionWidget מנהל את תהליך האימון. הוא אוסף 20 פריימים שמכילים פנים ומוציא מהם landmarks. תוך כדי הוא מעדכן עם סיגנאל process את הui של המשתמש. את מערך הlandmarks הוא מעביר לFaceFeatures.
4. FaceFeatures בוחר 3 פריימים מהמערך, מחשב עבורם פיצ'רים, ממצע אותם ושומר.
5. בסיום התהליך נשלח סיגנאל training\_complete מ MoodDetectionWidget לui, והui סוגר את ההקלטה.

תרחיש שני - Play & Pause, with mood change -

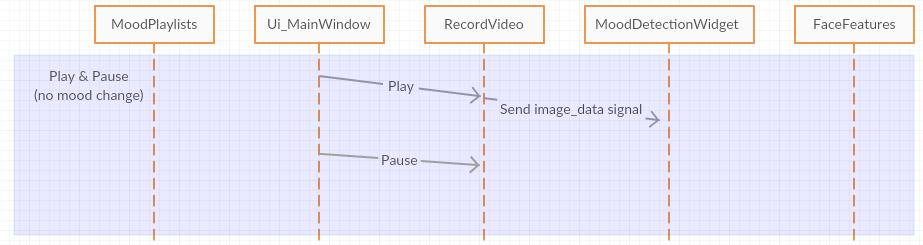
לאחר האימון של המערכת, המשתמש רשאי להתחיל להשתמש בנגן.

בשביל לעשות זאת על המשתמש ללחוץ על "Play", שפותח הקלטה, או על "Pause" שסוגר הקלטה.

ראשית נציג תרחיב בו המערכת תופסת שינוי של רגש:

השלבים של התהליך הם:

1. נשלחת בקשת Play מממשק המשתמש למחלקה RecordVideo.
2. RecordVideo מתחילה להקליט ולשלוח סיגנאלי פריימים לMoodDetectionWidget
3. MoodDetectionWidget מנהל את תהליך הניגון. על מנת לאפשר פעילות בזמן אמת, לא נוכל להוציא 8555 פיצ'רים לכל פריים. עבור כל פריים מתבצע חישוב של landmarks ומרחקים אוקלידים בין הנקודות. לאחר מכן מחושב מרחק בין וקטור המרחקים הנוכחי, לבין הוקטור של הפריים החשוד האחרון שנשמר. אם המרחק בינהם גדול מגבול שהגדרנו מראש, אנו מעבירים את הפריים לבדיקה קפדנית יותר.
4. הנקודות של הפריים נשלחות ל-FaceFeatures שמחשב עבורם וקטור פיצ'רים בהתאם לוקטור הפיצ'רים השמור משלב האימון. וקטור הפיצ'רים מוחזר ל MoodDetectionWidget
5. MoodDetectionWidget בודק בעזרת החזאי שאימנו אם הרגש בפריים שונה, והאם יש צורך לשנות את הPlaylist.
6. אם יש צורך לשנות רשימת השמעה, נשלח סיגנאל mood\_change לui, והui קורא לMoodPlaylists ומשנה את רשימת ההשמעה המושמעת.
7. כעת המשתמש לוחץ על Pause בממשק, והקלטת הוידאו נסגרת.



איור - תרחישים באפלקציה (2)

תרחיש שלישי - Play & Pause, without mood change -

תרחיש מאוד דומה לתרחיש השני - כאן לא התבצע שינוי ברגש.

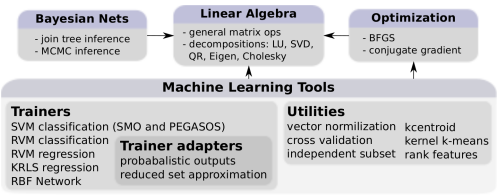
השלבים של התהליך הם:

1. נשלחת בקשת Play מממשק המשתמש למחלקה RecordVideo.
2. RecordVideo מתחילה להקליט ולשלוח סיגנאלי פריימים לMoodDetectionWidget
3. MoodDetectionWidget מנהל את תהליך הניגון. על מנת לאפשר פעילות בזמן אמת, לא נוכל להוציא 8555 פיצ'רים לכל פריים. עבור כל פריים מתבצע חישוב של landmarks ומרחקים אוקלידים בין הנקודות. לאחר מכן מחושב מרחק בין וקטור המרחקים הנוכחי, לבין הוקטור של הפריים החשוד האחרון שנשמר. אם המרחק בינהם גדול מגבול שהגדרנו מראש, אנו מעבירים את הפריים לבדיקה קפדנית יותר.
4. מכיוון שלא היה שינוי בהבעת הפנים, לא עברנו את הגבול של הבדיקה, ולא הוצאנו פיצ'רים לפריימים.
5. כעת המשתמש לוחץ על Pause בממשק, והקלטת הוידאו נסגרת.

## פירוט נוסף

1. זיהוי פנים בעזרת DLIB, המבוססת על openCV-

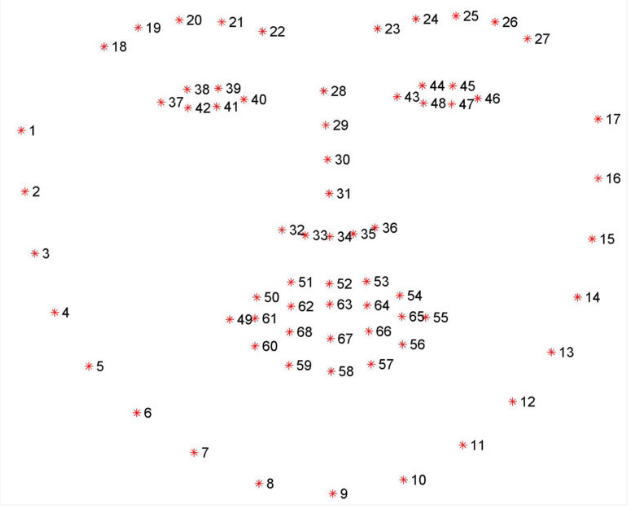
הסיפריה מאפשרת שימוש ברשתות מאומנות למציאת פנים, ונקודות מסוימות בפנים. האלגוריתמים בסיפריה:



איור

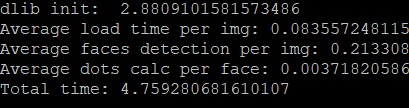
כמו כן, הסיפריה מאפשרת למצוא 68 נקודות בתמונת פנים, בעזרת רשת מאומנת.

להלן תמונת הנק' שהרשת מוצאת:



איור

זמני הרצה בשניות עבור 6 תמונות:



איור

1. גזירת פיצ'רים:

ניסינו לחשוב מה מאפיין שינוי בהבעות פנים.

בחרנו לחקור את הפיצ'רים הבאים:

* + מרחקים בין כל 2 נקודות.
  + זוויות בין כל 3 נקודות.

הבעיה - חישוב של הפיצ'רים לעיל עבור 68 נקודות בפנים, לוקח המון זמן.

כמות הפיצ'רים:

דרכים לצמצום פיצ'רים:

* + עבור כל 3 נקודות, קיימות 3 זוויות. מכיוון ששלושת הזוויות מגדירות משולש - ניתן להשתמש רק ב2 מהן (השלישית תלויה בהן).
  + להתייחס לפחות נקודות בפנים - ככה"נ קיימים הרבה פיצ'רים שהקורלציה בינהם גבוהה, ואין צורך להתחשב בשניהם. את חלקם אפשר להסיק באופן לוגי ואת האחרים בעזרת מטריצת קורלציה.

צמצמנו ל-30 נקודות מעניינות יותר -

זמני הרצה בשניות עבור 635 תמונות:

* Extract landmarks - 82.74
* Extract features - 49.65
* Calculate Correlation - 78.28
* Apply PCA dimension reduction - 0.25

# ניתוח תוצאות

להלן טבלה המסכמת את הדיוק של מסווגים שאומנו לפי גישות שונות במחקרי עבר (מקור [1], טבלה 7), בהשוואה לדיוק המסווג שלנו. ההשוואה הינה לפי רגש:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Emotion | Imbalanced | Down-Sampling | Up-Sampling | Weighted-Loss | Our Classifier |
| Happy | 88% | 85% | 85% | 82% | 98% |
| Sad | 63% | 64% | 60% | 63% | 85% |
| Surprise | 61% | 53% | 57% | 51% | 93% |
| Fear | 52% | 54% | 56% | 56% | 83% |
| Disgust | 52% | 53% | 53% | 48% | 91% |
| Anger | 65% | 62% | 63% | 60% | 89% |

טבלה 11

ניתן לראות כי נירמול הפיצ'רים ביחס להבעה הניטרלית מביא לשיפור משמעותי באחוזי הזיהוי, בהשוואה לאלגוריתמים חלופיים.

מדדנו זמני הרצה ממוצעים של מספר פונקציות במערכת:

|  |  |  |
| --- | --- | --- |
| Function | Run Time [s] | Comments |
| Train | 2.28 | Applied once on initial run |
| Image difference check | 0.01 | Applied on every frame |
| Feature extraction | 0.15 | Applied during training and new emotion calculation |
| New emotion calculation | 0.16 | Applied if and only if a difference was found |

* ניתן לראות כי זמני ההרצה טובים ומאפשרים עבודה בזמן אמת.
* עבור פונקצית האימון קיבלנו זמן מעט ארוך, אך זניח בהתחשב בכך שזוהי פונקציה שמריצים פעם אחת בשימוש אופייני.
* ניתן לראות כי רוב זמן הריצה של חישוב הרגשות הוא למעשה חישוב הפיצ'רים. כלומר, קיים חיסכון משמעותי בזמן בלא להריץ מציאת פיצ'רים על כל פריים.

# סיכום, מסקנות והצעות להמשך

מטרת הפרויקט שלנו היא לממש תוכנה לזיהוי רגשות לפי הבעות פנים. לאחר מאמצים רבים ובחינה של מספר גישות, ניתן לומר כי התוכנה פועלת בהתאם למטרה עם אחוזי דיוק גבוהים. כלומר, אנו מצליחים להפריד בין 6 רגשות המובעים בהבעות פנים בזמן אמת, באחוז דיוק של 85%.

מרבית המחקר שלנו נעשה באזור של חיזוי רגשות מתמונות של הבעות פנים.

במהלך המחקר נתקלנו בהרבה קשיים, למשל -

* אנשים מביעים יותר רגשות בדיבור עם אנשים אחרים, בניגוד לתקשורת מול מחשב. כלומר, ההבעות שאנשים ישנו מול מחשב הן מרומזות יותר באופן משמעותי, ויותר קשות לזיהוי.
* קשה לחזאי, וגם לאנשים, להבדיל לפעמים בין הבעה של כעס להבעה של גועל, כיוון שיש חפיפה כלשהי בין השניים. אבל – קל להגיד שהאדם מרגיש רגש שלילי כלפי משהו (בין אם זה גועל או כעס).
* ההגדרה של פנים עצובות או פחד היא מאוד כללית ומוגזמת לעיתים בעולם האמיתי. הן אולי נכונות למקרים קיצוניים של עצב/ פניקה, אבל כנראה שלא למקרים יותר מתונים.
* קל להגיד אם אדם מחייך, אבל לא קל להגיד אם זה חיוך מזויף.

הצעות לשיפור המערכת ומטרות עתידיות:

* שיפור אחוזי הדיוק:
  + אימון על קבוצת נבדקים גדולה יותר.
  + בחינת מודלים (אלגוריתמי חיזוי) נוספים.
* זיהוי רגשות נוספים.
* שיפור ביצועי המערכת:
  + חישוב מהיר יותר של פיצ'רים, או מציאת מודל שאינו דורש שימוש בפיצ'רים.
  + דגימה תכופה יותר של פריימים לשיפור חוויית המשתמש.
* הוספת פיצ'רים נוספים של נגן מוזיקה שלא הספקנו לממש.
* ללמוד את המשתמש יותר טוב:
  + להוסיף מערכת המלצות פר-משתמש.
  + ללמוד אונליין איזו מוזיקה משתמש מעדיף לשמוע כאשר הוא מרגיש רגש כלשהו.
* להשתמש בחזאי למטרות נוספות, לדוגמה רובוט שמבצע פיזיותרפיה.

מקורות

**פרסומים באנגלית:**

**מאגרי מידע**

1. Mollahosseini; B. Hasani; M. H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," in IEEE Transactions on Affective Computing, 2017.

https://arxiv.org/pdf/1708.03985.pdf

1. P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression", in the Proceedings of IEEE workshop on CVPR for Human Communicative Behavior Analysis, San Francisco, USA, 2010.
2. N. Aifanti, C. Papachristou and A. Delopoulos, ”The MUG Facial Expression Database,” in Proc. 11th Int. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), Desenzano, Italy, April 12-14 2010.
3. G. Zhao, X. Huang, M. Taini, S.Z. Li & M. Pietikäinen (2011): Facial expression recognition from near-infrared videos. Image and Vision Computing, 29(9):607-619.

**מאמר:**

1. Davis E. King, " Dlib-ml: A Machine Learning Toolkit", Journal of machine learning research 10, 2009

http://jmlr.csail.mit.edu/papers/volume10/king09a/king09a.pdf

1. Neeta Sarode et. al. ,"Facial Expression Recognition" (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 05, 2010, 1552-1557
2. Rituparna Halder, Sushmit Sengupta, Arnab Pal, Sudipta Ghosh and Debashish Kundu - "Real Time Facial Emotion Recognition based on Image Processing and Machine Learning", International Journal of Computer Applications (0975 – 8887) Volume 139 – No.11, April 2016

**קישורים למקורות באינטרנט:**

1. OpenCV tutorials - "Face Detection using Haar Cascades"

https://docs.opencv.org/3.3.0/d7/d8b/tutorial\_py\_face\_detection.html

1. "The Universally Recognized Facial Expressions of Emotion"

https://www.kairos.com/blog/the-universally-recognized-facial-expressions-of-emotion

1. "Hard Feelings: Science’s Struggle to Define Emotions"

https://www.theatlantic.com/health/archive/2015/02/hard-feelings-sciences-struggle-to-define-emotions/385711/