**מכללת אפקה להנדסה בתל אביב**

**המחלקה למדעי המחשב**

**פרויקט גמר מדעי המחשב, אפריל 2022**

**שם המנחה: ד"ר דינה גורן-בר**

**Facial Emotion Detection Classifier**

A picture containing different, several

Description automatically generated

**מגישים:**

**שחף מירון 312500457**

**עומר ישי 204308886**

**אדי אברמוב 307947655**

**תודות**

נבקש להודות למכללת אפקה על מתן ההזדמנות בביצוע פרויקט הגמר, ובמיוחד -

למנחת הפרויקט, ד''ר **גורן בר דינה**, תודה על האמון, העצמאות והתמיכה שניתנה לנו במהלך ביצוע עבודת הגמר. עבודת גמר מצריכה יחס האישי, סבלנות ואוזן קשבת. תודה על ההכוונה המקצועית לאורך כל הפרויקט.

בנוסף, נרצה להודות למשפחות שלנו, שתמכו בנו ועזרו לנו לאורך כל התואר.

תודה לכל חברינו שתרמו את עזרתם והשכלתם במהלך הפרויקט.

קריאה מהנה,

שחף, עומר ואדי

**תוכן עניינים**

4. תקציר......................................................................................................4

5. מבוא .......................................................................................................6

5.1 מוטיבציה ...............................................................................................6

5.2 הגדרת הבעיה ..........................................................................................7

5.3 מטרות ויעדים .........................................................................................8

6. סקירת ספרות ...........................................................................................8

7. סקר שוק .................................................................................................12

8. חלופות ....................................................................................................15

9. ארכיטקטורה ...........................................................................................19

10. תכן מפורט .............................................................................................23

11. תיאור התוצר גרסת אלפא .......................................................................25

11.1. אלגוריתמים ........................................................................................25

11.2. קוד ....................................................................................................26

11.3. הדגמה ................................................................................................29

12. הערכה ...................................................................................................30

12.1. סט נתונים ...........................................................................................30

12.2. מדדים וצורת בדיקה .............................................................................31

13. תוצאות ..................................................................................................34

14. סיכום ומסקנות ......................................................................................36

15. מקורות ..................................................................................................37

**4. תקציר**

מקובל להניח שניתן לנתח בקלות את מצבו הרגשי של אדם על פי תנועות הפנים שלו, הנקראות בדרך כלל הבעות רגשיות או הבעות פנים. הנחה זו משפיעה על שיפוטים משפטיים, החלטות מדיניות, פרוטוקולי ביטחון לאומי ושיטות חינוכיות. מנחה את האבחון והטיפול במחלות פסיכיאטריות, כמו גם פיתוח יישומים מסחריים, והוא חודר לאינטראקציות חברתיות יומיומיות כמו גם מחקר בתחומים מדעיים אחרים כגון בינה מלאכותית, מדעי המוח וראייה ממוחשבת. בפרויקט זה, אנו בוחנים את תנועות הפנים והעיניים של האדם באמצעות מודל היברידי, תוך התמקדות בשבע קטגוריות הרגשות הפופולריות ביותר: כעס, גועל, פחד, אושר, עצב, ניטרלי והפתעה. הראיות המדעיות כיום מצביעות על כך שאנשים מחייכים לפעמים כששמחים, מזעיפים את פניהם כשהם עצובים, זועמים פנים כשכועסים, וכן הלאה.

עם זאת, האופן שבו אנשים מתקשרים לכעס, גועל, פחד, אושר, עצב והפתעה משתנה באופן מהותי בין תרבויות, מצבים ואפילו בין אנשים בסיטואציה בודדת. יתר על כן, תצורות דומות של תנועות פנים מבטאות באופן שונה מקרים של יותר מקטגוריית רגשות אחת. למעשה, תצורה נתונה של תנועות פנים, כגון זעף פנים, מתקשרת לעתים קרובות למשהו אחר מלבד מצב רגשי. מדענים מסכימים שתנועות פנים מעבירות מגוון מידע וחשובות לתקשורת חברתית, רגשית או אחרת. אבל הפרוייקט שלנו מעיד על צורך דחוף בפתרון הבוחן כיצד אנשים למעשה מזיזים את פניהם ואת עיניהם כדי להביע רגשות ומידע חברתי אחר במגוון ההקשרים המרכיבים את חיי היומיום, כמו גם מחקר מדוקדק של המנגנונים שבהם אנשים תופסים מקרים של רגש אחד בשני. אנו מציעים פתרון ספציפי שיניב תמונה תקפה יותר של האופן שבו אנשים מזיזים את פניהם ועיניהם כדי להביע רגשות וכיצד הם מסיקים משמעות רגשית מתנועות פנים ותנועות העיניים במצבי חיי היומיום.

**4. Summary**

It is commonly assumed that a person's emotional state can be easily analyzed according to his facial movements, commonly called emotional expressions or facial expressions. This assumption affects judicial decisions, policy decisions, national security protocols, and educational practices. Guides the diagnosis and treatment of psychiatric illnesses as well as the development of commercial applications, and penetrates everyday social interactions as well as research in other scientific fields such as artificial intelligence, neuroscience and computer vision. In this project, we examine a person's facial and eye movements using a hybrid model, focusing on the seven most popular emotion categories: anger, disgust, fear, happiness, sadness, neutral and surprise. As for these days, scientific evidence suggests that people sometimes smile when they are happy, frown when they are sad, making angry face when they are angry, and so on.

However, the way people communicate anger, disgust, fear, happiness, sadness and surprise varies substantially between cultures, situations and even between people in a single situation. Moreover, similar configurations of facial movements differently express cases of more than one category of emotions. In fact, a given configuration of facial movements, such as a frown, often communicates something other than an emotional state. Scientists agree that facial movements convey a variety of information and are important for social, emotional or other communication. But our project testifies about urgent need for a solution that examines how people actually move their faces and eyes to express emotions and other social information in the variety of contexts that make up daily life, as well as a careful study of the mechanisms by which people perceive cases of emotion. We offer a specific solution that will yield a more accurate image of how people move their faces and eyes to express emotions and how they derive emotional meaning from facial movements and eye movements in everyday life situations.

**5. מבוא**

**5.1 מוטיבציה**

התפרצות מגפת הקורונה השפיעה רבות על מוסדות החינוך ברחבי העולם והובילה לסגירה כמעט מוחלטת של בתי ספר, אוניברסיטאות ומכללות בניסיון למנוע את התפשטות המגפה. עם הזמן, הסתמנה מגמה של מעבר ללמידה מקוונת בה התלמידים לומדים מרחוק. המעבר ללמידה מקוונת הציפה נזקים וקשיים רבים בקרב התלמידים, הן ברמה הנפשית והבריאותית והן בהישגים.

בהיותנו סטודנטים, חווינו ועודנו חווים את הקשיים בצל מגפת הקורונה ומתוך הבנה עמוקה של השלכות הלמידה המקוונת ונזקיה, נולד הרעיון אשר מוצג בפרויקט זה, והוא ניסיון לזהות את אותן המצוקות והקשיים שחווים התלמידים באמצעות זיהוי הבעות פנים וניתוח רגשות בזמן אמת.

הבעת פנים היא הביטוי הנראה לעין של המצב הרגשי, פעילות קוגניטיבית, כוונה, אישיות ופסיכופתולוגיה של אדם וממלאת תפקיד תקשורתי ביחסים בין אישיים. הבעת פנים היא אחד האמצעים החזקים והמיידים ביותר עבור בני האדם להביע את הרגשות, הכוונות והדעות שלהם. בשנים האחרונות, חוקרים הובילו להתקדמות ניכרת בפיתוח זיהוי וניתוח רגשות. חלק ממערכות זיהוי הפנים הקיימות היום מסווגות את הפנים למכלול הרגשות הבסיסיים כגון שמחה, עצב, כעס, פחד, גועל ורגש ניטרלי.

זיהוי רגשות יכול להתבצע בעזרת ניתוח מאפיינים שונים כגון הבעות פנים, דיבור, EEG, ואפילו טקסט. מבין תכונות אלו, הבעת הפנים היא הנפוצה ביותר. חלק עיקרי נוסף בזיהוי הרגשות על ידי הפנים הוא העיניים, שכוללות תכונות חיוניות לזיהוי רגשות כמו קוטר האישון, מיקום האישון, ומהירות התנועה של העין.

אנו סבורים כי באמצעות מתן דגש על ניתוח העיניים בשילוב ניתוח הבעות הפנים במערכת למידה מקוונת (דוגמת Zoom), נוכל להעלות את אחוז הדיוק של זיהוי וסיווג רגשות ובכך להתריע למורה או למרצה על זיהוי זה. עצם הסבת תשומת לב המרצה או המורה לרגשות התלמידים תסייע לו לאמוד את מידת הקשב וההבנה שלהם במהלך השיעור ואף לזהות מצוקות נפשיות ובריאותיות שהם חווים.

נציין כי זיהוי אוטומטי של הבעות פנים עשוי להיות מרכיב חשוב במגוון תחומים נוספים כמו ממשקי אדם-מכונה טבעיים, מדעי ההתנהגות, פרקטיקה קלינית ואף שיווק.

בעזרת האלגוריתם שלנו לזיהוי וסיווג הרגש המתאים, נוכל לתת מענה לאנשים שחווים קושי במערכת מקוונת שכזו, או לשפר את השיווק לסרט שאמור לצאת לאקרנים בקרוב.

**5.2 הגדרת הבעיה**

התפרצות הקורונה והוראת הריחוק החברתי הובילו לסגירת הקמפוסים ובתי הספר ואילצו את מוסדות החינוך ברחבי העולם לעבור להוראה מרחוק בחירום על מנת להמשיך את השגרה ככל הניתן ולאפשר לתלמידים וסטודנטים להשלים את לימודיהם.

השגרה היומיומית עברה למערכת היברידית באמצעות למידה ועבודה מרחוק בכל תחום, בתעשייה, אוניברסיטאות, מכללות ובתי ספר.

בשנת 2021 כמות המשתתפים היומית ב-Zoom הגיעה לשיא של כ-300 מיליון (עליה של 3000% משנת 2019) וזמן הפגישות השנתי עמד על 3.3 טריליון דקות שיחה כאשר פגישת זום בממוצע מגיעה ל-60 דקות.

תהליך הלמידה מרחוק טומן בחובו לא מעט אתגרים. תלמידים רבים מתקשים להתרכז למשך שעות רבות מול המסכים, הם חווים תחושת בדידות שלעיתים מתפתחת לתחושות דיכאון וחרדה ומתלוננים על עייפות ובעיות שינה.

מצידם של המורים והמרצים, האתגר הקשה ביותר הוא שמירה על האינטראקציות החברתיות, כיוון שרוב התלמידים אינם מרגישים בנוח לפתוח מצלמה, שיעור ההשתתפות בשיעור הולך ופוחת ולמורים ולמרצים אין שום אינדיקציה למידת הריכוז של התלמידים והאם הבינו את החומר הנלמד.

מסקרים שונים שבוצעו עולה כי 95% מתלמידי בתי הספר בישראל העידו שהם אינם מרוכזים במהלך השיעורים הנלמדים מרחוק. 71% מהם אמרו שהלמידה מרחוק פוגעת בהישגים שלהם בלימודים, 83% מהתלמידים שהשתתפו בסקר העידו שהם מרגישים בודדים בלמידה מקוונת ו-60% אמרו שאינם רוצים יותר ללמוד מרחוק אם הקורונה תסתיים. 56% מהתלמידים העידו כי הם חווים תחושות חרדה ודיכאון בעת השימוש בלמידה מקוונת. 75% מהתלמידים העידו כי הם חשים בכאבים ובעיות פיזיות שלא היו להם בעבר או שהחמירו בעקבות המצב, 72% מהתלמידים אמרו שהם חווים הפרעות שינה. בנוסף, 30% מהנשאלים העידו כי חוו חוסר ביחס ובליווי אישי מצד המורים.

ממצאי הסקר תואמים את עמדות המומחים לחינוך ופסיכולוגים על ההשלכות השליליות של הלמידה מרחוק, בעיקר תוך הסתמכות מסיבית על השימוש ב-Zoom כמערכת העיקרית שבה מתבצעת הלמידה בישראל, על אף שלא פותחה כדי לשמש כפלטפורמה ללמידה.

גם בקרב המבוגרים עלו תחושות דומות כאשר 80% ממשתתפי הסקר טענו כי הם חווים תחושת בדידות ומחסור במפגשים חברתיים ו-57% מהם אמרו שהם חווים חרדה ודיכאון. 78% דיווחו על קשיי ריכוז, 70% על כאבים שהחמירו או לא היו ו-66% אמרו כי פיתחו בעיות שינה.

אנשים רבים חווים התמודדות עם לחצים שונים בחייהם, בין אם מלמידה אינטנסיבית בתואר אקדמי, חששות מהעתיד המקצועי, בעיות שונות בחיים האישיים, עומס שנוצר משילוב של הדברים וכו׳. קשיים אלו, כאשר אינם מטופלים עלולים להוביל להידרדרות במצב הרגשי ואף להוות התחלה של הפרעות נפשיות שונות (כגון הפרעת חרדה ודיכאון). כיום כ- 300 מיליון אנשים בעולם חווים דיכאון, כאשר לכמעט 50% מהם יש גם חרדות. בנוסף, התאבדות היא הגורם השני לתמותה בקרב גילאים 10-34.

כלומר, אלו הן תופעות נפוצות אשר מגבירות את הצורך בזיהוי מוקדם של התסמינים שעשויים לבוא לידי ביטוי בין היתר בהבעות פנים וברגשות.

בפן השיווק והמכירה נמצא כי רווחים שנכנסים מפרסומות עם תוכן רגשי לעומת פרסומות עם תוכן רציונלי גבוהים פי שניים וכי לתכנים מכעיסים יש סיכוי של 40% להפוך לוויראליים ברשת. כמו כן, ממקורות עולה כי 70% מהאנשים שחווים תגובה עזה לפרסומת, רוכשים את המוצר. כלומר גם בתחום השיווק והמכירה יש חשיבות רבה לתכנים רגשיים והבעות פנים עשויות להעיד על הצלחה או אי הצלחה של תוכן שיווקי כזה או אחר.

על אף התקדמות הטכנולוגיה, והרלוונטיות של הבעיה, אחוזי הדיוק בסיווג הרגשות לפי הבעות הפנים והשימוש במודלי למידת מכונה מסוג זה עדיין לא גבוהים מספיק.

**5.3 מטרות ויעדים**

מטרת פרויקט זה הינה לפתח שיטה לזיהוי רגשות באמצעות הבעות פנים ותנועות עיניים, זיהוי מהיר, חזק, פשוט למדי ומדויק ככל הניתן. זאת על ידי:

* שיפור אלגוריתם לזיהוי רגש לפי הבעות פנים
* שיפור אלגוריתם לזיהוי רגש לפי הבעת עיניים
* יצירת אלגוריתם היברידי המשלב את שניהם ומוציא תוצאות מדויקות יותר.

**6. סקירת ספרות**

תיאוריות רגשות קודמות הציעו שיש בין שניים לעשרים רגשות בסיסיים. ארבעת הרגשות הנפוצים ביותר המופיעים ברשימות החוקרים הם פחד, כעס, עצב ושמחה. חוקר ושמו פלוצ'יק הבחין בין שמונה רגשות בסיסיים: פחד, כעס, צער, שמחה, גועל, קבלה, ציפייה והפתעה. חוקר אחר ושמו אייקמן התמקד בקבוצה של שישה עד שמונה רגשות בסיסיים הקשורים להבעות פנים. [3]

חוקרים נוספים מנסים לזהות תנועות שרירים בודדות שהפנים יכולות לייצר על מנת לספק תיאור אובייקטיבי של הפנים. המסגרת הפסיכולוגית הידועה ביותר לתיאור כמעט כל תנועות הפנים הוא מערכת לקידוד פעולת הפנים (FACS) Facial Action Coding System. [1]

הבעות פנים הינן גלויות, הן מכילות תכונות שימושיות רבות לזיהוי רגשות, וקל יותר לאסוף סט נתונים גדול של פרצופים מאשר אמצעים אחרים לזיהוי רגש אנושי. [2]

לפי מאמרים שסקרנו, גילינו כי אלגוריתמי Machine Learning כגון KNN, SVM ו-RF הם הנפוצים ביותר לסיווג רגש לפי הבעות פנים:

* אלגוריתם ה-**SVM (Support Vector Machines)** הוא אלגוריתם למידת מכונה מפוקח פשוט יחסית המשמש לסיווג ו/או רגרסיה. הוא מועדף יותר לשם סיווג אבל לפעמים שימושי מאוד גם עבור רגרסיה. כעיקרון, SVM מוצא מרחב ממימד n-1 בתוך מרחב ממימד n שיוצר גבול בין סוגי הנתונים. במרחב דו מימדי, מישור ההיפר הזה אינו אלא קו ישר.
* אלגוריתם ה-  **KNN(K-Nearest Neighbors)** הוא אחד מאלגוריתמי הסיווג הבסיסיים אך החיוניים ביותר בלמידת מכונה. הוא שייך לתחום למידת מכונה מפוקחת וחזק בזיהוי תבניות, כריית נתונים וזיהוי פריצות.
* אלגוריתם ה- **RF (Random Forest)** מסוגל לבצע הן משימות רגרסיה והן משימות סיווג תוך שימוש בעצי החלטה מרובים ובטכניקה הנקראת Bootstrap and Aggregation, הידועה בכינויה bagging. הרעיון הבסיסי מאחורי זה הוא לשלב עצי החלטה מרובים בקביעת התפוקה הסופית במקום להסתמך על עצי החלטה בודדים.

שלושת אלגוריתמי הסיווג (classification) הללו ידועים כבעלי ביצועים טובים עבור מרחבים בעלי מימדים גבוהים כמו גם למספר רב של דגימות אימון (training).

כל אחד מאלגוריתמים האלו פועל היטב בנסיבות ספציפיות, לֹKNN- עבור נתונים לא-מסודרים, ה- SVM עבור נתונים בלתי נפרדים באופן ליניארי ו-RF עבור מאפיינים קטגוריאליים. [4]

שיטות אלו היו בשימוש נרחב במחקרים ומאמרים שקראנו.

לפי מאמר [4], תהליך ניתוח התמונה מסט הנתונים וניתוח הרגש מתחלק לשתי שכבות.

השכבה הראשונה נקראת שכבת האימון (training) שהיא בעצם עובדת על סט הנתונים ומאמנת את המודל ולה יש את השלבים הבאים:

1. קלט תמונה ורצף.
2. עיבוד מקדים כגון מיסוך, שינוי קנה מידה, המרה לגווני אפור והפחתת ה-"רעש" בתמונה.
3. שימוש באלגוריתמים של ראייה ממוחשבת לאיתור תכונות ייחודיות (feature extraction)

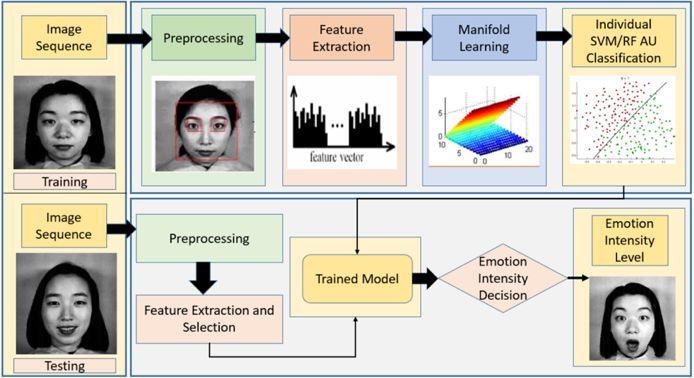
  כגון LBP, Gabor,  ו-HOG  ויוצר וקטור תכונה סופי באמצעות שרשור.

1. שימוש בהפחתת מימד כדי להסיר את התכונות הלא רצויות.
2. אלגוריתמי סיווג כגון SVM, RF ו-KNN המשמשים לסיווג ה-AUs. ( Action Units)

השכבה השנייה, שכבת הבדיקות (testing), כוללת בעיקר שני שלבים: 

1. בדומה לשלבי אימון המודל, מבוצע קלט ורצף של תמונות, עיבוד מקדים וחילוץ ובחירה של תכונות ייחודיות.
2. תכונות אלו מועברות דרך המודל המאומן ולבסוף מתקבלת החלטה על עוצמת הרגש על סמך ה-AUs.

Table

Description automatically generated

תמונה 2. Action Units

תמונה 1. תהליך ניתוח הרגש לפי הבעת הפנים באמצעות המודל.

במאמר [2] , חוקרים מעידים ששימוש באלגוריתמים של הראייה הממוחשבת כגון HOG ו-LBP, בשילוב עם מסווג מאומן, עובדים על סט נתונים של תמונות שצולמו במצב מבוקר בצורה טובה למדי. עם זאת, אלגוריתמים אלו מתקשים לפעול באותה מידה על סט-נתונים מאתגרים יותר עם שונות בתמונות ופנים חלקיות.

מנגד, כשהמשכנו לחקור ולקרוא גילינו שבעזרת למידה עמוקה עם דגש על  - Convolutional neural networks (CNN), ניתן ללמוד ולחלץ תכונות רבות עבור מערכת לזיהוי הבעות פנים טובה. עם זאת, ראוי לציין שבמקרה של הבעות פנים, רבות מהאינדיקציות מגיעות ממספר אזורים בפנים, למשל, הפה והעיניים, בעוד שחלקים אחרים, כגון האוזניים והשיער, משחקים חלקים קטנים בסיווג. משמעות הדבר היא שבאופן אידיאלי, מסגרת למידת המכונה צריכה להתמקד רק בחלקים חשובים של הפנים ועליה לשים פחות דגש על אזורי פנים אחרים.[2]

Diagram

Description automatically generated

*תמונה 3. שלבי תהליך ניתוח הבעת הפנים וסיווג לרגש.*

זיהוי הבעת פנים הינו תחום מחקר פעיל בעשורים האחרונים, והוא עדיין מאתגר בשל השונות הגבוהה בין תתי הקבוצות של הסיווגים השונים.

במסגרת מאמר [5], כאשר סוכמו העדויות על האופן שבו אנשים מזיזים את פניהם בפועל במהלך הבעת רגש ובעקבותיהם אנו בוחנים עדויות על כך שרגשות למעשה מוסקים מהסתכלות על תנועות פנים. אנו מזהים שלושה חסרונות עיקריים במחקר המדעי שתרמו לאי הבנה כללית לגבי האופן שבו רגשות מובעים ונתפסים בתנועות פנים ומגבילים את פירושו של מחקר מדעי זה לשימושים אחרים:

* **מהימנות מוגבלת** - מקרים של אותה קטגוריית רגשות אינם באים לידי ביטוי בצורה מהימנה או נתפסים מתכונות דומות של תנועות פנים. (לדוגמה תנועות פנים דומות לנגעל ומפוחד).
* **חוסר ספציפיות** - מיפוי ייחודי בין תצורה של תנועות פנים לבין מקרים של קטגוריית רגש.
* יכולת הכללה מוגבלת - השפעות ההקשר והתרבויות השונות לא תועדו מספיק.

גילויים ובעיות נוספות שמצאנו:

* ההשפעה של המגדר בהבעות פנים ורגש, הייתה חלשה, למשל כעס עשוי להיתפס יותר כגועל אצל נשים מאשר בפרצופים של גברים, וסביר יותר שגועל נתפס ככעס אצל גברים מאשר אצל נשים.
* ביטויים שמחים וניטרליים זוהו טוב יותר (הן דיוק גדול יותר וזמני תגובה קצרים יותר) מכל הביטויים האחרים, בעוד שביטויים של פחד היו הסבירים והאחרונים ביותר לזיהוי מדויק. דיוק הזיהוי וזמן התגובה לביטויים כועסים, עצובים, נגעלים ומופתעים היו בין שני הקצוות, עם הבדלים קטנים יחסית בין ארבעת הביטויים הללו.
* היו שגיאות שיטתיות של תפיסה שגויה בין ביטויים, במיוחד בין כעס לגועל (ולהיפך) ובין הפתעה לפחד (ולהיפך), כאשר עצב נתפס בצורה שגויה כפחד או גועל.
* סף הזיהוי היה נמוך במיוחד עבור פרצופים מאושרים. [6]

**7. סקר שוק**

**Real Eyes**

מודדת באמצעות ראייה ממוחשבת ולמידת מכונה את תשומת הלב והמעורבות של הקהל המצורף בזמן שהם צופים בתוכן הווידאו אותו אתה מעוניין לבדוק.

1. במהלך הסרטון המצלמה בוחנת את הבעות הפנים של המשתמש.

2. בסוף הסרטון מנפיקה דוח של הרגש המתאים עבור כל רגע נתון במהלך הצפייה.

3. בעזרת ניבוי הרגשות של המשתמשים, החברה המפיקה

Graphical user interface, website

Description automatically generated את הסרט תוכל לדעת האם הצופים יאהבו את הסרט.

*תמונה 4. דף הבית של אתר Real Eyes.*

**Noduls – FaceReader**

היא מערכת אוטומטית לזיהוי של מספר מאפיינים ספציפיים בתמונות פנים, כולל ששת ההבעות הבסיסיות או האוניברסליות: שמח, עצוב, כועס, מופתע, מפחד ונגעל.

בין אם משתתף הבדיקה שלך הוא תינוק, ילד, מבוגר או אדם מבוגר FaceReader ,מתאים את הניתוח למודל המתאים ביותר למחקר שלך.

**שלבים בתהליך ניתוח וסיווג הרגש:**

1. איתור פנים - מוצא פנים באמצעות אלגוריתם איתור פנים מבוסס למידה עמוקה.
2. דוגמנות פנים - יוצר מודל פנים מלאכותי מדויק תוך שימוש בכמעט 500 נקודות מפתח.
3. סיווג פנים - מסווג את ההבעות עם רשתות עצביות מלאכותיות.

**A picture containing text, person, indoor, person

Description automatically generated**

*תמונה 5. הדגמה של ניתוח הבעת הפנים של Noduls.*

**Emteq Labs**

פלטפורמת ניתוח רגשות מהדור הבא למדידה אובייקטיבית של תגובות אנושיות בזמן אמת. אידיאלי למחקר אקדמי וקליני, כמו גם אימות תוכן עבור יוצרים ומפתחים.

ניתוח וסיווג הרגשות מתבצע על ידי לבישת משקפי VR, המשדרות סיטואציות בחיי היום-יום ובזמן אמת מנתחות את הבעת הפנים ומסווגות את הרגש המתאים באותו הרגע.

בנוסף לכך, המוצר מסווג מדדי דופק לב, תנועות עיניים ותנועות שרירים.

**A picture containing diagram

Description automatically generated**

*תמונה 6. דף הבית של אתר Emteq.*

**Visage Technologies – Face Analysis**

FaceAnalysis מזהה פרצופים בתמונות או בווידאו ולאחר מכן משתמשת ביחידות מעקב ופעולה כדי לספק במדויק מין, רגשות וגיל לפנים במיקום חזיתי בערך.

* עוקב ומגיב להתנהגות אנושית בזמן אמת.
* בונה חוויות מרתקות של לקוחות וממקסם את שביעות הרצון שלהם.
* משלב מעקב עיניים לקבלת תובנות מעמיקות יותר לגבי השפעת גירויים שונים על רגשות.

A person's face on a poster

Description automatically generated with low confidence

*תמונה 7. דף הבית של אתר Visage.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Our Solution | Emteq | Realeyes | Visage Technologies  Face Analysis | Noldus  FaceReader | קריטריונים / מוצר |
| 2021 | 2015 | 2007 | 2002 | 1989 | נוסדה |
| שמחה, עצב, כעס, הפתעה, פחד, הגעלות, ניטרליות, זלזול. | לא צוין. | שמחה, הפתעה, הגעלות בלבול, אמפתיה וזלזול | שמחה, עצב, כעס, הפתעה, פחד הגעלות. | שמחה, עצב, כעס, הפתעה, פחד, הגעלות, ניטרליות, זלזול. | רגשות |
| קיים (אופציונלי) | לא קיים | לא קיים | קיים | קיים | ניתוח גיל |
| קיים (אופציונלי) | לא קיים | לא קיים | קיים | קיים | ניתוח מגדר |
| מוסדות לימוד, חוקרים, עסקים, מפתחי תוכנה ומאמנים. | חוקרים, מפתחי תוכנה ומאמנים | חברות פרסום | עסקים וחוקרים | חוקרים | קהל יעד |
| קיים | קיים | קיים | קיים | קיים | אתר אינטרנט |
| מצלמה ומחשב | מערכת מציאות מדומה | מצלמה ומחשב | מצלמה ומחשב | מצלמה ומחשב | ציוד נדרש |
| קיים | לא קיים | לא קיים | לא קיים | לא קיים | ניתוח באמצעות מודל היברידי המתמקד בעיניים |

**טבלת ניתוח המתחרים**

אנו מאמינים כי המוצר שלנו ייחודי משאר המוצרים הקיימים, ונותן מענה הכרחי לסיווג רגש לפי הבעת פנים בעזרת תמיכה נוספת באמצעות סיווג רגש לפי תנועות עיניים. בנוסף, נוכל גם לנתח את הגיל ואת המגדר של בן אדם על מנת להוסיף עוד נתונים למחקר ולסטטיסטיקה, למשל, אם רוצים לבצע מחקר מדגמי על הרגשות של בני האדם לפי גילאים ומגדרים.

**8. חלופות**בחלק זה אנו בחנו את החלופות האפשריות עבור כל שלב בתהליך סיווג הרגשות, חילוץ תכונות מהתמונה וסיווג התמונה לרגש המתאים.

**Feature Extraction1.**

עבור שלב הFeature Extraction גילנו שנעשים שימושים באלגוריתמים נפוצים שלComputer Vision ו-Neural Networks.

**HOG -** הטכניקה סופרת מופעים של כיוון שיפוע בחלקים מקומיים של תמונה שיטה זו דומה לזו של היסטוגרמות של כיוון קצה מתארי טרנספורמציה של מאפיינים בקנה מידה והקשרי צורות אך שונה בכך שהיא מחושבת על רשת צפופה של תאים ברווח אחיד ומשתמשת בנורמליזציה של ניגודיות מקומית חופפת לשיפור הדיוק

**LBP -** הוא אופרטור טקסטורה פשוט אך יעיל מאוד אשר מסמן את הפיקסלים של תמונה על ידי קביעת הסף של הסביבה של כל פיקסל ומציג את התוצאה כמספר בינארי באמצעות LBP בשילוב עם היסטוגרמות נוכל לייצג את תמונות הפנים עם וקטור נתונים פשוט.

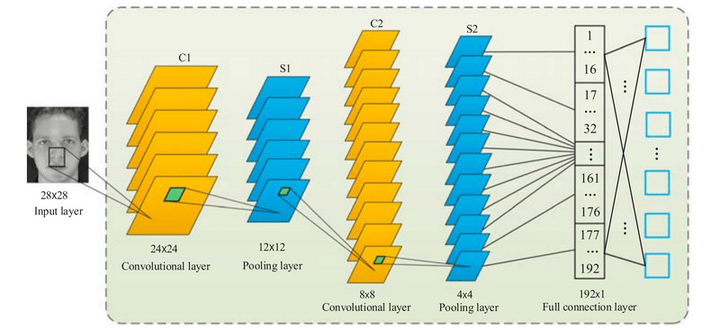
**CNN -** בDeep Learning - ה - CNN היא סוג של רשת עצבית מלאכותית עם שכבות קונבולציוניות המיושמת לרוב לניתוח דימויים חזותיים.

באופן כללי, CNN מכיל שני סוגים של שכבות נסתרות: Convolutional Layer ו – Pooling Layer שהן בדרך כלל מסודרות לסירוגין ברשת העצבית.

לפיכך זמן האימון של ה CNN יכול להיות קצר להפליא ברוב המקרים.

בפרט מתי שתמונה היא הקלט של CNN, ניתן להכניס את התמונה ישירות לרשת העצבית כדי למנוע שלבים מסובכים כגון חילוץ התכונות (Feature Extraction) ושחזור נתונים (Date Reconstruction).

*תמונה 8. הדגמה של חילוץ תכונות התמונה ברשת הקונבנציונלית*



**שילוב של HOG & LBP ב - Feature Extraction**

* שניהם מנסים להשתמש באותו סוג של מידע - שיפועים סביב פיקסל. ההבדל העיקרי בין HOG ל-LBP הוא האופן שבו כל שיטה משתמשת במידע ההדרגתי. הכוח של LBP נובע מהעובדה שהוא משתמש בכל 8 הכיוונים עבור כל פיקסל, לעומתו HOG משתמש רק בכיוון אחד לכל פיקסל. עם זאת, הגסות שבה האלגוריתם המופעל על ידי LBP, גורמת לו לאבד מידע בהשוואה ל-HOG.
* ידוע ש - HOG מעולה בלכידת קצוות ופינות בתמונות. מצד שני, LBP לוכד את הדפוסים המקומיים. בסופו של דבר HOG ו-LBP לוכדים סוגים שונים של מידע, מה שהופך אותם למשלימים אחר לאחר. ואכן, ניתן לשלב את HOG ו-LBP יחד ביישומי עיבוד תמונה כגון זיהוי פנים ולכן נשתמש באלגוריתמים אלו עבור שלב ה-Feature Extraction.
* עם זאת, החיסרון הוא שאלגוריתמים אלו מתקשים לפעול באותה מידה על סט-נתונים מאתגרים יותר עם שונות בתמונות ופנים חלקיות.

**Classification2.**

**SVM (Support Vector Machines) -** ידוע כבעל יכולת הכללה על ידי מיפוי קלט באופן לא ליניארי למרחבי תכונה ממדיים גבוהים יותר באמצעות היכולת שלו להפריד נתוני אימון עם מישור היפר (היפר-מישור הוא תת-מרחב שהמימד שלו קטן מזה של המרחב הסביבתי שלו(.

**KNN (K- Nearest Neighbors) -**הוא סוג של למידה מבוססת מופעים, כולל את השכנים שמחליטים לאיזו מחלקה (בין k מחלקות) שייכת נקודת נתונים ספציפית. השכנים הקרובים ביותר מוקצים בשיטות פופולריות כגון מרחק אוקלידי או האמינג.

**RF (Random Forest)**

Diagram

Description automatically generated with medium confidenceהוא אוסף של מספר עצי החלטה שאינם זקוקים לתכונות ליניאריות או אפילו לתכונות המקיימות אינטראקציה ליניארית.

**CNN (Convolutional Neural Network)**

ניתן להשתמש ברשתות קונבנציונליות גם בתור מסווג.

*תמונה 9. המחשה של שכבות הרשת הקונבנציונלית*

**יתרונות וחסרונות**

**Feature Extraction**

**HOG**

יתרונות:

* מתמקד במבנה או בצורתו של אובייקט, טוב יותר מכל ״מתאר קצה״ מכיוון שהוא משתמש בגודל וגם בזווית של השיפוע כדי לחשב את התכונות.
* עבור אזורי התמונה הוא מייצר היסטוגרמות תוך שימוש בגודל והכיוון של הגרדיאנט.

חסרונות:

* רגיש מאוד לסיבוב תמונה. לכן, HOG אינו בחירה טובה לסיווג של טקסטורות או אובייקטים שלעיתים קרובות ניתן לזהות כתמונה מסובבת.

**LBP**

יתרונות:

* חוסן לשינויים מונוטוניים בגווני אפור הנגרמים, למשל, על ידי שינויים בתאורה.
* פשטות חישובית, המאפשרת לנתח תמונות בהגדרות מאתגרות בזמן אמת.

חסרונות:

* יצירת היסטוגרמות ארוכות למדי, שמאטות את מהירות הזיהוי במיוחד במסד נתונים בקנה מידה גדול של פנים.
* לעיתים מפספסים את המבנה המקומי מכיוון שהם לא מתחשבים בהשפעה של הפיקסל המרכזי.
* הנתונים הבינאריים המופקים על ידם רגישים לרעש לכן, עלול להיות בעייתי לפתרון שלנו שמשתמש במצלמות אינטרנט והתמונות ככל הנראה יכילו רעש רב.

**CNN**

יתרונות:

* רשתות CNN מספקות חילוץ אוטומטי של תכונות, וזה היתרון העיקרי.

חסרונות:

* כבד מבחינת זיכרון ולא מקודד את המיקום והכיוון של האובייקט לתחזיות שלהם.

**Classification**

**SVM**

יתרונות:

* עובד טוב יחסית כאשר יש מרווח ברור של הפרדה בין המחלקות.
* יעיל יותר במרחבים בעלי ממדים גבוהים.
* יעיל במקרים שבהם מספר הממדים גדול ממספר הדגימות
* יעיל יחסית בזיכרון.

חסרונות:

* אלגוריתם SVM אינו מתאים למערכות נתונים גדולות.
* אינו מתפקד טוב במיוחד כאשר מערך הנתונים מכיל יותר רעש, כלומר מחלקות היעד חופפות.
* במקרים בהם מספר התכונות עבור כל נקודת נתונים עולה על מספר דגימות נתוני האימון, ה-SVM יבצע פחות.

**KNN**

יתרונות:

* זמן חישוב מהיר
* אלגוריתם פשוט לניתוח
* רב תכליתי - שימושי עבור רגרסיה וסיווג
* דיוק גבוה - אין צורך להשוות עם מודלים של למידה בפיקוח טוב יותר.

חסרונות:

* הדיוק תלוי באיכות הנתונים.
* עם נתונים גדולים, שלב החיזוי עשוי להיות איטי.
* רגיש לקנה המידה של הנתונים ותכונות לא רלוונטיות.
* דורש זיכרון גבוה - צריך לאחסן את כל נתוני האימון.
* בהתחשב בכך שהוא מאחסן את כל ההכשרה, זה יכול להיות יקר מבחינה חישובית.

**RF**

יתרונות:

* יכול לפתור את שני סוגי הבעיות שהוא סיווג ורגרסיה ועושה הערכה הגונה בשתי החזיתות.
* כוח גדול בלטפל בסט נתונים גדולים עם מימדיות גבוהה יותר. הוא יכול להתמודד עם אלפי משתני קלט ולזהות את המשתנים המשמעותיים ביותר ולכן הוא נחשב לאחת משיטות הפחתת הממדיות.
* המודל מציג חשיבות של משתנה, מה שיכול להיות תכונה שימושית מאוד.

חסרונות:

* עושה עבודה טובה בסיווג אבל לא לגבי רגרסיה מכיוון שהוא לא נותן חיזוי רציף מדויק. הוא לא מנבא מעבר לטווח בנתוני האימון, ושהם עשויים להתאים יותר לקבוצות נתונים רועשות במיוחד.
* יכול להרגיש כמו גישת קופסה שחורה, יש לנו מעט מאוד שליטה על מה שהמודל עושה. אתה יכול במקרה הטוב לנסות פרמטרים שונים וSeeds- אקראיים.

**9. ארכיטקטורה**

לחילוץ תכונות התמונה (Feature extraction) בחרנו להשתמש במודל מאומן של ספריית Deep Face שאומן באמצעות DNN. עבור האלגוריתם לזיהוי מהעיניים, בחרנו להשתמש במודל הCNN מכיוון שהוא בעל אחוזי הדיוק הגבוהים ביותר כשמדובר בחילוץ התכונות של התמונה לפי הבעות פנים.

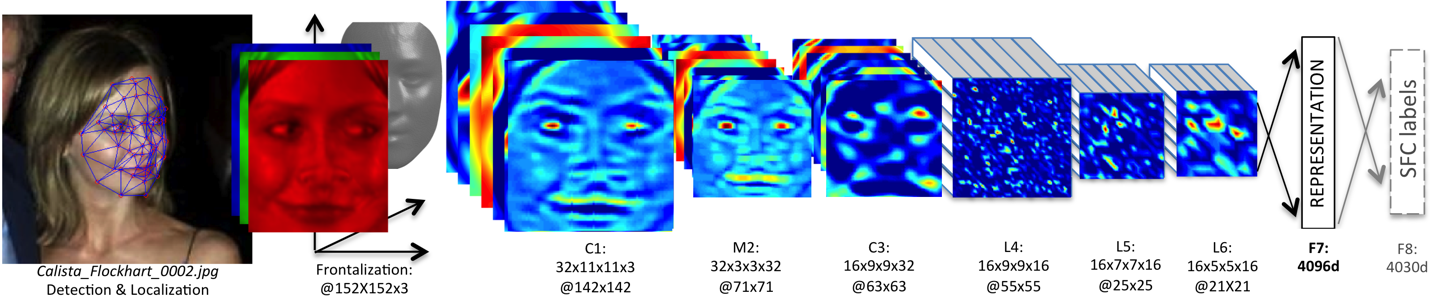
נשתמש באלגוריתם ה-CNN גם כדי לבצע את הסיווג לרגש (Classification). לכל השכבות של CNN יש מסננים קונבולוציונים מרובים הפועלים וסורקים את מטריצת התכונות ומבצעים את הפחתת מימדים, מה שמאפשר ל-CNN להיות רשת מתאימה מאוד לסיווג ועיבוד תמונות.

נממש זאת באמצעות אלגוריתם היברידי, הכולל שני מודלים לזיהוי רגשות מתכונות בפנים:

* מודל מאומן באמצעות DNN לזיהוי רגשות מהבעות פנים Deep Face.
* מודל לזיהוי רגשות מתנועות עיניים באמצעות CNN.

**מודל מאומן באמצעות DNN לזיהוי רגשות מהבעות פנים Deep Face**

מתוך ספריית deepFace אנו משתמשים במודל DNN מאומן לסיווג הרגשות ל7 הקטגוריות.



*תמונה 10. המחשה של שכבות ארכיטקטורת Deep Faceבאמצעות DNN.*

convolution-pooling-convolution filteringיחיד על הקלט המיושר (שכבות C1,M2,C3), ואחריו שלוש שכבות המחוברות מקומית (L4,L5,L6) ושתי שכבות מחוברות במלואן (F7,F8). הצבעים ממחישים מיפוי של התכונות המיוצרות בכל שכבה. הרשת כוללת יותר מ-120 מיליון פרמטרים, כאשר יותר מ-95% מגיעים מהשכבות המקומיות המחוברות במלואן.

**Diagram

Description automatically generated**

*איור 1. דיאגרמת בלוקים של מודל Deep Face.*

המודל מאומן באמצעות DNN על מנת לנתח ולזהות תמונת פנים. כפי שהארכיטקטורה מוצגת באיור 1. תחילתה, תמונת פנים תלת מימדית RGB בגודל 152 על 152 פיקסלים מועברת לשכבה הקונבנציונלית C1 עם 32 פילטרים בגודל 11x11x3. לאחר מכן 32 מטריצות מיפויי התכונות מוזנות לשכבת הMax Pooling אשר לוקחת את מקסימום עבור 3x3 קבוצות של שכנים במרווחים של 2 (Stride).

לאחר מכן מועברים לשכבה הקונבנציונלית השלישית C3 עם 16 מטריצות מיפויי התכונות בגודל 9x9x16.

מטרת שלוש השכבות הללו היא לחלץ תכונות ברמה נמוכה, כמו קצוות ומרקמים פשוטים. שכבות pooling הופכות את הפלט של רשתות הקונבנציונליות להיות חזק יותר עבור ניתוחים מקומיים של התמונה.

השכבות הבאות הן L4,L5,L6 והן מחוברות באופן מקומי, כמו שכבה קונבולוציונית רגילה הן מיישמות סינון, אבל כל מיקום במטריצת מיפויי התכונות לומדת קבוצה אחרת של מסננים. זה הכרחי מכיוון שלאזורים שונים של תמונה מיושרת יש נתונים סטטיסטיים מקומיים שונים. לדוגמה, אזורים בין העיניים והגבות מציגים מראה שונה מאוד ובעלי יכולת הבחנה גבוהה בהרבה בהשוואה לאזורים בין האף לפה.

לבסוף, שתי השכבות העליונות (F7 ו-F8) מחוברות במלואן: כל יחידת פלט מחוברת לכל הכניסות. שכבות אלו מסוגלות ללכוד מתאמים בין תכונות שנתפסו בחלקים מרוחקים של תמונות הפנים, למשל, מיקום וצורת העיניים ומיקום וצורת הפה. הפלט של השכבה הראשונה המחוברת במלואה (F7) ברשת ישמש כווקטור של תכונת ייצוג הפנים הגולמי.

**מודל לזיהוי רגשות מתנועות עיניים באמצעות CNN**

אימון המודל מתחלק לשני חלקים, שלב האימון ושלב הבדיקות.

**Trainin**Graphical user interface, application

Description automatically generated**g**

*איור 2. דיאגרמת בלוקים של אימון מודל העיניים באמצעות CNN.*

1. Image Input From Dataset

קלט תמונה מהסט-נתונים ועיבוד מקדים כגון מיסוך, שינוי קנה מידה, המרה לגווני אפור והפחתת "רעשים" בתמונה.

1. Feature Extraction

שימוש במודל ה-CNN על מנת ליצור מאפיינים אוטומטית ושילובם עם המסווג בשלב הבא.

תחילתה, תמונת פנים תלת מימדית בשחור-לבן בגודל 48 על 24 פיקסלים מועברת לשכבה הקונבנציונלית C1 עם 32 פילטרים בגודל 3. לאחר מכן 32 מטריצות מיפויי התכונות מוזנות לשכבה הקונבנציונלית השנייה עם 64 פילטרים בגודל 3 ומועברת לשכבת הMax-Pooling

אשר לוקחת את מקסימום עבור 2x2 קבוצות של שכנים במרווחים של 2 (Stride).

לאחר מכן, מתבצע בשכבה הבאה Dropout. שכבה זו מונעת התאמת יתר של נתוני האימון, אם הם מוסתרים (נוירונים), המקבץ של דוגמאות האימון ישפיע על הלמידה בצורה לא פרופורציונלית. לכן נוכל למנוע למידה של תכונות מיותרות המופיעות בשלבים מאוחרים יותר באימון.

נעבור לשכבות הבאות, C5 עם 128 פילטרים בגודל 3 ולאחריה שכבה M6 בדומה לM3.

נעבור באופן דומה לשכבות C7 ו-M8 עם אותם נתונים ולבסוף נבצע Dropout.

לבסוף, נבצע שיטוח (Flatten) וDense על מנת ליצור fully connected layers.

1. Feature Classification

כל הנוירונים מהשכבות הקודמות מחוברים. ה-CNN יסווג את התווית לפי המאפיינים מהשכבות הקונבולוציוניות ויצטמצם עם שכבת ה-pooling.

1. Trained Model

שמירת המבנה והמשקולות של המודל המאומן.

**Testing**

לאחר שאימנו את המודל בשלב האימון, ביצענו את שלב הבדיקות על מנת להעריך ולנבא את אחוזי הדיוק

**Diagram

Description automatically generated**של המודל המאומן עם סט נתונים שגודלו 25% מגודל הסט נתונים של שלב האימון.

*איור 3. דיאגרמת בלוקים של מודל Deep Face.*

**Diagram

Description automatically generatedמודל היברידי משולב הבעות פנים ותנועות עיניים**

*איור 3. דיאגרמת בלוקים של המודל ההיברידי.*

בעזרת שני המודלים, יצרנו מודל היברידי המשלב סיווג רגשות באמצעות הבעות פנים ובאמצעות תנועות עיניים. תמונת המשתמש תעבור דרך זיהוי הפנים, במידה והיא לא תזהה היא תעצור. אחרת, היא תמשיך ותעבור למודל DeepFace ותבצע סיווג רגשות לפי הבעות פנים, במידה ואחוז הדיוק לסיווג הרגש הנכון יהיה מעל 75% אנו נציג למשתמש את הרגש לפי הבעת הפנים. אחרת, אנו נעבור למודל העיניים ונבצע סיווג רגש לפי תנועות העיניים. במידה ואחוז הדיוק בסיווג הרגש יהיה נמוך יותר אנו נשתמש במודל של הבעות הפנים, אחרת אנו נציג למשתמש את הרגש לפי תנועות העיניים.

**10. תכן מפורט**

**Main.py**

קובץ ההרצה הראשי, משתמש במודל לזיהוי רגשות מהבעות פנים ובמודל זיהוי רגשות בעזרת העיניים. בעת הריצה, מתבצעת לולאה אינסופית, בה הוא לוקח פריים (Frame) מהמצלמה של המשתמש ומנתח אותה:

1. תחילה האלגוריתם מזהה את הפרצופים בפריים.
2. על כל פנים שנמצא, צייר מלבן על מנת לתת אינדיקציה שזוהה פנים.
3. לאחר מכן, החלק בפריים שמכיל את הפנים, נשלח למודל לזיהוי רגשות באמצעות הבעות פנים של DeepFace.
4. המודל מחזיר מערך המייצג את 7 הרגשות כאשר בכל תא יש ערך בין 0-100 המייצג את אחוז ההסתברות לאותו רגש. את המערך שומרים במשתנה results.
5. כעת, מתבצע זיהוי של העיניים, וסימון במלבן.
6. גזור את המלבן המייצג את העיניים מתוך הפריים ושלח למודל זיהוי רגשות על ידי העיניים.
7. מתוך מערך הרגשות במשתנה results, נמצא את הערך המקסימלי ונשמור אותו ואת האינדקס שלו.
8. בדוק האם ההסתברות לאותו רגש קטנה מ75%.
   1. אם כן, בדוק האם מודל העיניים חזה את אותו רגש.
      1. אם כן, נבע את הרגש והצג על המסך.
      2. אם לא, אל תציג תוצאה ועבור לפריים הבא.
   2. אם לא, נבא את הרגש והצג על המסך.

**בעיות וקשיים**

במהלך הפרויקט נתקלנו בבעיות במציאת datasets המייצגים רגשות באמצעות העיניים, לבסוף כשמצאנו סט נתונים הוא לא היה עקבי, למשל חתוך בצורה לא טובה, או שהתמונות צולמו מזוויות שונות. בנוסף, בכל קטגוריה היה מספר שונה של תמונות (כלומר סט הנתונים לא היה מאוזן) ולכן המודל היה ״משוחד״ לקטגוריות בעלות מספר רב יותר של תמונות לאימון.

כמו כן, המידע שמצאנו במהלך סקירת הספרות שביצענו היה מועט בנוגע לנושא של סיווג רגשות באמצעות העיניים ובהרבה מהסקירות שמצאנו היה צורך בכלי מדידה יעודי לעיניים ולא רק מצלמת רשת. יתר על כן, חווינו קשיים במהלך הבדיקות עם מצלמת הרשת היות ורוב מצלמות הרשת שהיו ברשותינו היו באיכות נמוכה, ונפגעו מתנאי תאורה שנים.

**11. תיאור התוצר גרסת ALPHA**

**11.1 אלגוריתמים וספריות**

**OpenCV** – ספריה המשמשת לזיהוי אובייקטים, באלגוריתם שלנו נעשה שימוש בספריה זאת לזיהוי הפנים וזיהוי העיניים.

ספריה המשמשת למחשוב מדעי , באלגוריתם שלנו נעשה שימוש בספריה זאת לטיפול - **NumPy**

במערכים ומטריצות.

- ספריה המשמשת לבניית מודלים של למידת מכונה , באלגוריתם שלנו נעשה שימוש **Keras**

בספריה זאת להגדרת רשת עצבית שתבנה לאחר מכן על ידי TensorFlow.

**TensorFlow** – ספריה המשמשת לבניית רשת עצבית , באלגוריתם שלנו נעשה שימוש בספריה לבנית רשת עצבית.

**Deepface** – ספריה הכוללת מודל מאומן לזיהוי רגשות בעזרת הפנים במודל זה עשינו שימוש בפרויקט.

באמצעות **Keras** מימשנו אלגוריתמים להוספת רשתות קונבנציונליות, פונקציות שיטוח וכו':

Conv2D, MaxPooling2D, Dense, Dropout, Flatten.

כמו כן, השתמשנו בפונקציית Sequential על מנת לאתחל את המודל לפני הוספת השכבות המתאימות.

עבור כל הוספה של שכבה קונבנציונלית, הוספנו את פונקציית ReLU, מוציאה את הקלט ישירות אם הוא חיובי, אחרת היא מוציאה אפס. היא הפכה לפונקציית ההפעלה המוגדרת כברירת מחדל עבור סוגים רבים של רשתות עצביות מכיוון שמודל שמשתמש בה קל יותר לאימון ולעתים קרובות משיג ביצועים טובים יותר.

השתמשנו באלגוריתם Adam, הוא אלגוריתם אופטימיזציה שניתן להשתמש בו במקום הליך הירידה האקראי הקלאסי של ירידה בשיפוע לעדכון משקלי רשת הנוירונים בהתבסס על נתוני האימון.

השתמשנו בפונקציית ImageDataGenerator של Keras, המספקת דרך מהירה וקלה להגדיל את התמונות. היא מספקת שורה של טכניקות הגדלה שונות כמו סטנדרטיזציה, סיבוב, תזוזות, סיבובים, שינוי בהירות ועוד רבים אחרים.

**חלוקת העבודה**

חלוקת העובדה על הפרויקט בוצעה שווה בשווה, נציין כי את רוב תהליך כתיבת הקוד, אימון המודלים ואיחודם למודל אחד נעשו בשיתוף פעולה מלא, במהלך העבודה על הפרויקט נתקלנו במכשולים ואתגרים רבים מהם למדנו והפקנו ידע רב בתחום שהיה חדש לנו ונלמד תוך כדי עבודה על הפרויקט.

את העבודה חילקנו לשלושה חלקים:

מודל לזיהוי רגשות בעזרת הפנים – שחף מרון.

מודל לזיהוי רגשות בעזרת העיניים – אדי אברמוב.

אופטימיזציה ואיחוד המודלים למודל אחד – עומר ישי.

**11.2 קוד**

* main.py
* train\_eyes\_model.py

את הקובץ train\_eyes\_model.py אין צורך להריץ כאשר רוצים להשתמש בתוכנית, בקבצים אלו עשינו שימוש מוקדם לאימון המודלים של זיהוי רגשות לפי העיניים.

כעת נפרט את קוד אימון מודל העיניים (train\_eyes\_model.py ):

Text

Description automatically generated

עיבוד התמונות לביצוע האימון.

Text

Description automatically generated

בניית המודל.

Text

Description automatically generated

אימון המודל בעזרת מסווג CNN.

Text

Description automatically generated

שמירת מבנה המודל המאומן בקובץ בו נשתמש בכל פעם שנרצה להריץ את המודל בעזרת קוד ה-.main.py

Text

Description automatically generated

שמירת המשקולות אותם יצר המודל בזמן האימון בקובץ שבו נשתמש בכל פעם שנרצה להריץ את קובץ ה-main.py.

לאחר שאימנו את מודל העיניים ושמרנו את המודל המאומן וייבאנו את הספרייה ובה מודל הפנים המאומן, כל שנותר לעשות הוא להריץ את קובץ ה main.py.

פירוט קוד ה: main

Text

Description automatically generated with medium confidence

טעינה ויצירת מודל זיהוי הרגשות על פי העיניים שאומן על ידינו טרם הרצת קוד הmain.



טעינת המשקולות למודלים של זיהוי העיניים.



הפעלת מצלמת הרשת לצורך הפעלת המודלים של זיהוי הרגשות בזמן אמת.

Text

Description automatically generated

סימון הפנים והעיניים (בריבוע) אשר מזוהים במצלמה בזמן אמת.

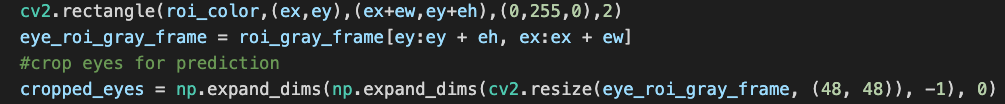
Text

Description automatically generated

זיהוי הפנים המופיעות במצלמה וקבלת ניבוי לרגש המזוהה מהבעת הפנים.



זיהוי כל העיניים המופיעות במצלמה.



חיתוך העיניים שזוהו לתמונות נפרדות.

Text

Description automatically generated

קבלת חיזוי רגשות עבור העיניים.

Text

Description automatically generated

בדיקת הניבוי ויצירת חיווי למשתמש.

אם מודל זיהוי הפנים ניבא רגש מסוים בסבירות גדולה מ-75%, המודל יתן חיווי בזמן אמת על הרגש שהתקבל. אחרת, (כאשר אין רגש עם סבירות גדולה מ-75%) אם הרגש בעל הסבירות הגבוהה ביותר שהתקבל מזיהוי הפנים שווה לחיזוי אותו קיבל מודל העיניים, רגש זה ישודר בזמן אמת. אחרת, המערכת תמשיך לבדוק עד אשר תתקבל תשובה חד משמעית.

על מנת להריץ את הקוד יש לוודא כי למחשב יש מצלמת רשת מובנת או מחוברת וכן סביבת עבודה לפיתוח בשפת פייתון כגון pycharm, vs code וכדומה.

לבסוף, יש לוודא כי כל הספריות בהן נעשה שימוש בקוד מותקנות ולאחר מכן יש להריץ את הקובץ main.py.

**11.3 הדגמה**



ניתן לראות כי מודל חיזוי הרגשות ניבא כי הרגש של הפנים אותם זיהה הוא ב90.7% ניטרלי, לכן החיזוי של מודל העיניים איננו רלוונטי.

A person wearing headphones

Description automatically generated with medium confidence

*תמונה 11. המחשה של סיווג הרגש בזמן אמת*

כפי שניתן לראות בתמונה, הרגש אותו זיהה המודל (ניטרלי) מוצג על גבי המסך.

**12. הערכה**

**12.1.1 Face Recognition Data Sets**

**Deep Face משתמשת בשלושה סטים של נתונים:**

* **SFC dataset**

זהו סט הנתונים שנוצר על ידי פייסבוק. הוא מכיל כמעט 4.4 מיליון תמונות של 4030 אנשים שלכל אחד מהם 800 עד 1200 תמונות פנים. לצורך הבדיקה, הם לוקחים 5% מהתמונות העדכניות ביותר מכל קטגוריה.

* **LFW dataset**

זהו אחד ממאגרי הנתונים הפופולריים ביותר בתחום זיהוי הפנים. הוא מכיל יותר מ-13,000 תמונות אינטרנט של יותר מ-5,700 סלבריטאים.

* **YTF dataset**

הוא מכיל 3425 סרטונים של 1595 סלבריטאים (קבוצת משנה של ידוענים מ-LFW). סרטונים אלו מחולקים ל-5,000 זוגות וידאו של 10 פיצולים ומשמשים להערכת ביצועים באימות פנים באמצעות וידאו.

**12.1.2 Eye Data Set**

* **Eyes Data Set**

סט נתונים מבוסס על סט הנתונים הידוע FER-2013 שנועד לזיהוי רגשות לפי הבעות פנים, הוא מכיל 28,717 תמונות חתוכות של העיניים בלבד. התמונות בגודל 24x48 המסווגות ל 7 סוגי רגשות:

* + כעס – 3996 סה"כ תמונות
  + נגעל – 495 סה"כ תמונות
  + פחד- 4098 סה"כ תמונות
  + שמח – 7216 סה"כ תמונות
  + נטרלי – 4966 - סה"כ תמונות
  + עצוב – 4831 - סה"כ תמונות
  + מופתע – 3172 סה"כ תמונות

כמו במאגר הנתונים של זיהוי הפנים ניתן לראות כי זהו סט נתונים של תמונות שהרגש נגעל איננו כולל מספר גדול של תמונות ולכן אחוזי הדיוק נמוכים משמעותית לעומת שאר הרגשות.

**12.2 מדדים וצורת הבדיקה**

תחילה ביצענו מדדים על סט הנתונים לזיהוי וניתוח רגש לפי הבעת פנים הידוע FER-2013 המכיל

7,178 תמונות פנים (Validation) לפי 7 סוגי הרגשות.

כדי להעריך את אחוזי הדיוק ואת אחוזי השגיאה השתמשנו בשיטות הערכת ביצועים הידועות בלמידת מכונה, Classification Report ו-Confusion Matrix.

Table, calendar

Description automatically generated

A screenshot of a computer

Description automatically generated with low confidence

*תמונה 13. Classification Report*

*תמונה 12. Confusion Matrix*

על פי ה – Classification Report, קיבלנו אחוז דיוק נמוך מאוד ועומד על 18%.

לאחר שדגמנו את מדדים אלו ראינו כי יש צורך לשפר את המודל ולכן עבור מודל זיהוי הפנים השתמשנו במודל Deep Face והמדדים והבדיקות שבוצעו עבורו:

* 1. **SFC dataset**

המודל אומן באמצעות רשת זרימה קדימה (FeedForward neural network) רשתות זרימה קדימה היו הסוג הראשון והפשוט ביותר של רשתות נוירונים מלאכותיות. ברשת כזו, המידע זורם רק בכיוון אחד, קדימה, מצמתי הקלט, דרך הצמתי החבויים (באם קיימים) אל צמתי הפלט, ללא מעגלים או לולאות עצמיות.

האימון סווג באמצעות SGD עם ,Momentum=0.9 ו-Batch Size=128 וקצב הלמידה זהה עבור כל שכבת הנוירונים, שהוא 0.01 ועם 15 epochs.

Table

Description automatically generatedהמודל מאומן על שלוש קבוצות תתי קבוצות (Validation) של סט נתונים של 1,500 אנשים (1.5 מיליון תמונות), 3,000 אנשים (3.3 מיליון תמונות) ו-4,000 אנשים (4.4 מיליון תמונות).

אחוז שיעור טעויות הסיווג בתתי קבוצות אלו הוא 7%, 7.2% ו-8.7% בהתאמה. טבלה [1] – השוואה בין שגיאות הסיווג

*טבלה 1. השוואה בין שגיאות הסיווג בסט הנתונים SFC*

* 1. **LFW dataset**

הביצועים נמדדו בשלוש שיטות:

**Restricted Method -** שיטה מוגבלת שבה זוג התמונות שסופק לדגמן והמטרה היא לזהות אם התמונה זהה או לא זהה.

**- Unrestricted Method**שיטה בלתי מוגבלת שבה יותר תמונות מזוג בודד נגישות לאימון.

Table

Description automatically generatedChart

Description automatically generated**Unsupervised method -** שיטה לא מפוקחת שבה המודל לא עבר הכשרה על מערך נתונים של LFW.

*גרף1 . השוואה בין אחוזי הדיוק בין אלגוריתמים אחרים לסיווג הרגשות*

*טבלה 2 . השוואה בין אחוזי הדיוק בסט הנתונים LFW*

, DeepFace-ensambleמייצג שילוב של מודל DeepFace-Single שונה שמשתמש במדדי אימות שונים. כפי שאנו יכולים להסיק DeepFace-ensamble מגיע לדיוק מירבי של 97.35% דיוק שהוא קרוב מאוד לרמה האנושית 97.53%. ניתן לראות שהוא מתעלה על שאר האלגוריתמים כמו Baysian ו- LBP.

* 1. **Table

     Description automatically generatedYTF dataset**

*טבלה 3 . השוואה בין אחוזי הדיוק בסט הנתונים YTF עם אלגוריתמים שונים*

נשים לב, שבגלל טשטוש תנועה, "לכלוכים" וגורמים אחרים, איכות התמונה של מערכי נתונים של וידאו היא בדרך כלל גרועה יותר ממערך הנתונים של תמונות. עם זאת, 91.4% הוא עדיין הדיוק המתקדם באותה תקופה ומפחית את שיעור השגיאות ביותר מ-50%.

מבחינת זמן הבדיקה,ל- DeepFace לוקח 0.33 שניות כאשר נבדק על מעבד אינטל ליבה אחת במהירות 2.2GHz. זה כולל 0.05 שניות של יישור ו-0.18 שניות בחילוץ תכונה.

לכן, זהו אחד ממודלי זיהוי הפנים הטובים ביותר כעת, כמובן שישנם מודלים נוספים כגון Google-FaceNet ומודלים אחרים המספקים אחוזי דיוק של עד 99.6% על סט הנתונים של LFW. הבעיה העיקרית שה-DeepFace הצליחה לפתור היא לבנות מודל שאינו מושפע מחשיפה לאור, תנוחה, הבעת הפנים וכו' ובגלל זה הוא משמש ברוב משימות זיהוי הפנים של פייסבוק. כמו כן, הגישה החדשנית לשימוש ביישור תלת מימד תרמה גם היא לעלייה ברמת הדיוק של המודל.

הסט השני בו השתמשנו לבדיקת תוצאות של מודל חיזוי הרגשות בעזרת העיניים הוא – datasetEyes שבו 8,136 תמונות והתוצאות שהתקבלו הן:

Graphical user interface, application, table

Description automatically generated

A screen shot of a computer

Description automatically generated with low confidence

*תמונה 15. Classification Report*

*תמונה 14. Confusion Matrix*

על פי ה – Classification Report, קיבלנו אחוז דיוק נמוך מאוד ועומד על 17%.

חישוב מדדי מודל הפנים הם 7 סוגי הרגשות אשר מהם כל רגש קיבל אחוז מסוים שאותו המודל חזה לעומת מודל העיניים שבו הרגש שהמודל זיהה קיבל את הספרה 1 לעומת שאר הרגשות בהם סומן 0.

**13. תוצאות**

לצורך בחינת תוצאות המודלים ביצענו סימולציה ובה הכנסנו למודל שבעה סרטונים (מתוך האתר <https://mixkit.co>) המייצגים אנשים אמיתיים.

כל סרטון מורכב ממאות תמונות (המכונות Frames) של אדם המביע רגש ספציפי לאורך הסרטון.

נדגיש כי המודל לא נתקל בסרטונים או בפרצופים אלו לפני. כפלט, האלגוריתם מייצא לטבלת אקסל את הרגש המזוהה והזמן בו זוהה.

Table

Description automatically generated

*טבלה 4. הצגה של הרגש בכל פריים.*

מצורפות בזאת 3 תמונות מתוך פלט התוכנית בהן ניתן לראות כי האלגוריתם חוזה נכונה את הרגש. כמו כן, ניתן לראות כי האלגוריתם הצליח ללכוד את הפנים ואת העיניים על ידי הצגת המלבן הירוק על התמונה.

A person with the hands on the face

Description automatically generated with low confidenceA picture containing person, outdoor, green, accessory

Description automatically generated

*תמונה 17. פלט התוכנית בעבור אדם שמח באמצעות מודל עיניים.*

*תמונה 16. פלט התוכנית בעבור אדם עצוב.*

A picture containing text, person, green

Description automatically generated

*תמונה 17. פלט התוכנית בעבור אדם מפוחד.*

חשוב לציין, כי רגשות כמו גועל וכעס הראו תוצאות פחות מדויקות ואף היו סרטונים שלא זהותה התוצאה הנכונה כלל. בעוד שלשמחה, עצב, ניטרליות ופחד היו אחוזי הצלחה גבוהים ומעל 80% מזמן הסרטון נובאה תוצאה, לגועל וכעס היו סרטונים שלא נובה רגש נכון כלל, או לחלק קטן מזמן הסרטון (קטן מ20%).

|  |  |
| --- | --- |
| emotion | count |
| sad | 0 |
| angry | 1 |
| surprise | 0 |
| fear | 0 |
| happy | 574 |
| disgust | 0 |
| neutral | 1 |

בגרףHappy video מטה, ניתן לראות כי מתוך 576 פריימים בסרטון של אישה שמחה ומחייכת קיבלנו חיזוי רגש happy על 574 פריימים. מדובר בדיוק של מעל ל- 99% על 576 פריימים שונים. בכל תמונה, פריים האובייקט (המצולם) משנה מעט את הבעת פניו ואת מיקומו.

*גרף 2. הצגה של הרגשות לאורך סרטון שלם.*

על אף התוצאות הגבוהות שקיבלנו בסרטון השמחה, היו רגשות שעדיין היה קשה לחזות כפי שניתן לראות בגרף Disgust Video. ניתן לראות שמתוך סרטון בעל 275 פריימים המודל לא הצליח לחזות את הרגש הנכון, ״גועל״ וניתן לראות כי הרגשות ״עצב״ ו״עצבנות״ נובו בכמות הגבוהה ביותר. הדבר עולה בקנה אחד עם מחקרים שקראנו הטוענים כי יש קושי רב להבחין בין רגשות אלו.

*גרף 3. הצגה של הרגשות לאורך סרטון שלם המתאר אדם נגעל.*

**14. סיכום ומסקנות**

לסיכום, מתוצאות הפרויקט ניתן לראות כי ניתן לחזות רגשות מתוך הבעות פנים ברמה מדויקת.

בעזרת למידת מכונה ולמידה עמוקה מודלים מאומנים יכולים להגיע לאחוזי דיוק גבוהים מאוד.

בעזרת שילוב עם מודל המסווג רגשות באמצעות העיניים בלבד, הצלחנו להוריד את מדד הFalse positive.

בתחילת הבדיקות שהרצנו, גילינו שמודל החיזוי הרגשות בעזרת הבעות הפנים שלנו עושה עבודה טובה פחות ממה שיצא לנו בשלב הvalidation ומדד הדיוק נפגע משמעותית.

לאחר כיוונון הפרמטרים וניסיונות נוספים הבנו שהמודל לא מספיק טוב וחווינו קושי בלנסות לאמן מודל חדש באמצעות החומרה שהייתה ברשותנו. (דרש זמן רב ).

בשלב זה החלטנו לעבור למודל מאומן מתוך ספריית DeepFace בעל אחוזי דיוק של ~95%.

כמו כן, גילינו שיש רגשות שקשה יותר לסווג כגון: הפתעה, הגעלות וכעס.

עבור מודל החיזוי בעזרת העיניים אותו אימנו בעזרת data set הכולל המון תמונות אשר פחות טובות לביצוע האימון קיבלנו תוצאות שאינן מדויקת עם אחוז דיוק של 64% אך מכיוון שבתחום זיהוי רגשו בעזרת העיניים אין מודלים מאומנים ולמרות מאמצים רבים לא מצאנו data set אחר לבצע עליו את האימון נשארנו עם מודל זה.

מכיוון שמודל זיהוי הרגשות היה בעל אחוז דיוק גבוהה מאוד בשילוב עם מודל לחיזוי העיניים קיבלנו אחוז דיוק גבוהה.

המלצותינו להמשך המחקר הם:

* מציאת או הרכבת data set לאימון מודל העיניים הכולל מספר גדול של תמונות באיכות גבוהה.
* אימון מודל הפנים בו עשינו את הבדיקות הראשוניות עםdata sets והשוואה עם תוצאות המודל deep face בו השתמשנו.
* הגדלת הdata sets של מודל הפנים והעיניים ברגשות בהם אחוז הדיוק היו הנמוכים ביותר (הפתעה, הגעלות וכעס).
* שילוב האלגוריתם בפלטפורמות העובדות בלייב כגון zoom ,teams וכו׳ על מנת לראות את יעילות האלגוריתם ומוכנותו.
* החלפת מודל העיניים במודל אשר מחזיר את התוצאות עבור הרגשות באחוזים ולא ערך בוליאני על מנת לקבל אינדיקציה יותר מדויקת ולשלבה עם מודל הפנים בצורה מדויקת וטובה יותר.
* מומלץ להשתמש בחומרה חזקה על מנת לבצע את הבדיקות והשינויים שהמלצנו.

**15. מקורות**

**[1]**  **Automatic facial expression recognition for intelligent tutoring systems**, · June 2008

Jacob Whitehill, Marian Bartlett, and Javier Movellan Machine Perception Laboratory University of California, San Diego

<https://www.researchgate.net/profile/Javier-Movellan/publication/228907849_Automatic_facial_expression_recognition_for_intelligent_tutoring_systems/links/53fc997c0cf22f21c2f3ee01/Automatic-facial-expression-recognition-for-intelligent-tutoring-systems.pdf>

**[2]**  **Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network**

 Shervin Minaee , Mehdi Minaei  and Amirali Abdolrashidi.

 Snapchat Inc., Santa Monica, CA 90405, USA

 CS Department, Sama Technical College, Azad University, Tonekabon 46817,  Iran.

 CS Department, University of California, Riverside, CA 92521, USA

 27 April 2021.

<https://www.mdpi.com/1424-8220/21/9/3046/htm#B14-sensors-21-03046>

**[3]** **An Affective Model of Interplay Between Emotions and Learning: Reengineering Educational Pedagogy—Building a Learning Companion**

 Barry Kort, Rob Reilly, Rosalind W. Picard

 M.I.T. Media, Laboratory

 2001

<https://www.academia.edu/568626>

**[4]**  **Recognition of Emotion Intensities Using Machine Learning Algorithms: A Comparative Study,** 21 April 2019

Dhwani Mehta, Mohammad Faridul Haque Siddiqui and Ahmad Y. Javaid

Electrical Engineering and Computer Science Department, The University of Toledo, 2801 W Bancroft St, MS 308, Toledo, OH 43606, USA;

<https://www.mdpi.com/1424-8220/19/8/1897/html>

**[5]**  **Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements,** July 17, 2019

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D.

1Department of Psychology, Northeastern University

2Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts

3Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, Massachusetts

4Division of Humanities and Social Sciences, California Institute of Technology

5College of Computer and Information Science, Northeastern University

6Institute of Neuroscience & Psychology, University of Glasgow

7Department of Electrical and Computer Engineering and Center for Cognitive and Brain Sciences, The Ohio State University

8Department of Psychology, University of Wisconsin–Madison

https://journals.sagepub.com/eprint/SAUES8UM69EN8TSMUGF9/full

**[6]**  **Facial expressions of emotion (KDEF): Identification under different display-duration conditions,** 2008

MANUEL G. CALVO, DANIEL LuNDQvIST

University of La Laguna, Tenerife, Spain

Karolinska Institutet, Stockholm, Sweden

<https://link.springer.com/content/pdf/10.3758/BRM.40.1.109.pdf>

\_

**[7]**  **Comparing the Accuracy of Deep Neural Networks (DNN) and Convolutional Neural Network (CNN) in Music Genre Recognition (MGR): Experiments on Kurdish Music,** 2021

Aza Zuhair and Hossein Hassani University of Kurdistan Hewlˆer Kurdistan Region - Iraq

<https://arxiv.org/pdf/2111.11063.pdf>

**[8]**  **DeepFace: Closing the Gap to Human-Level Performance in Face Verification,** 2020

Yaniv Taigman, Ming Yang , Marc’Aurelio Ranzato, Facebook AI Researches from Menlo Park, CA,

Lior Wolf Tel Aviv University Tel Aviv, Israel

<https://www.cs.toronto.edu/~ranzato/publications/taigman_cvpr14.pdf>