



## המחלקה להנדסת תוכנה

### פרויקט גמר – התשע"ז

#### למידה וניתוח תמונות רפואיות

#### Machine Learning to Medical Data - Analyzing CT Images

מאת

אלון צליק שמילוביץ

וסתיו ברזני

מנחה אקדמי: מר אסף ב. שפנייר      אישור:      תאריך:

רכז הפרויקטים: דר' ראובן יגל      אישור:      תאריך:

מערכות ניהול הפרויקט:

מיקום	מערכת
<a href="https://github.com/alonshmilo/MedicalData_jce">https://github.com/alonshmilo/MedicalData_jce</a>	מאגר קוד
אתר ניהול פעילות לפי תאריכים <a href="http://alonshmilo.wixsite.com/jce-finalproject">http://alonshmilo.wixsite.com/jce-finalproject</a>	יומן
אתר ויקי <a href="https://github.com/alonshmilo/MedicalData_jce/wiki">https://github.com/alonshmilo/MedicalData_jce/wiki</a>	
<a href="https://huboard.com/alonshmilo/MedicalData_jce">https://huboard.com/alonshmilo/MedicalData_jce</a>	ניהול פרויקט
<a href="https://youtu.be/Zr0yy6b0b8M">https://youtu.be/Zr0yy6b0b8M</a>	סרטון אב-טיפוס



## תוכן העניינים

3.....	<u>מילון מונחים</u>	א.
3.....	<u>מבוא</u>	1.
4.....	<u>תיאור הבעיה</u>	2.
4.....	<u>דרישות ואפיון הבעיה</u>	2.1.
4.....	<u>הבעיה מבחינת הנדסת תוכנה</u>	2.2.
5.....	<u>תיאור הפתרון</u>	3.
5.....	<u>מהי המערכת</u>	3.1.
5.....	<u>תהליכים ונתונים המערכת</u>	3.2.
5.....	<u>תיאור הפתרון המוצע</u>	3.3.
6.....	<u>תיאור הכלים המשמשים לפתרון</u>	3.4.
7.....	<u>תכנית בדיקות</u>	4.
7.....	<u>סקירת עבודות דומות בספרות והשוואה</u>	5.
8.....	<u>סיכום ומסקנות</u>	6.
8.....	<u>נספחים</u>	7.
8.....	<u>רשימת ספרות/ביבליוגרפיה</u>	א.7.
9.....	<u>תרשימים וטבלאות</u>	ב.7.
11.....	<u>תכנון הפרוייקט</u>	ג.7.
11.....	<u>טבלת סיכונים</u>	ד.7.
12.....	<u>רשימת/טבלת דרישות</u>	ה.7.

### מילון מונחים, סימנים וקיצורים

"משתמש" - רופא או איש רפואה שיש לו עניין בניתוח התמונות הרפואיות.

"המערכת" - האפליקציה המוצעת, הנותנת פתרון לבעיית ניתוח התמונות הרפואיות, הלא הן

תמונות הCT.

"תמונות CT" - תמונות אלו מתקבלות לאחר ביצוע בדיקת טומוגרפיה ממוחשבת - Computed

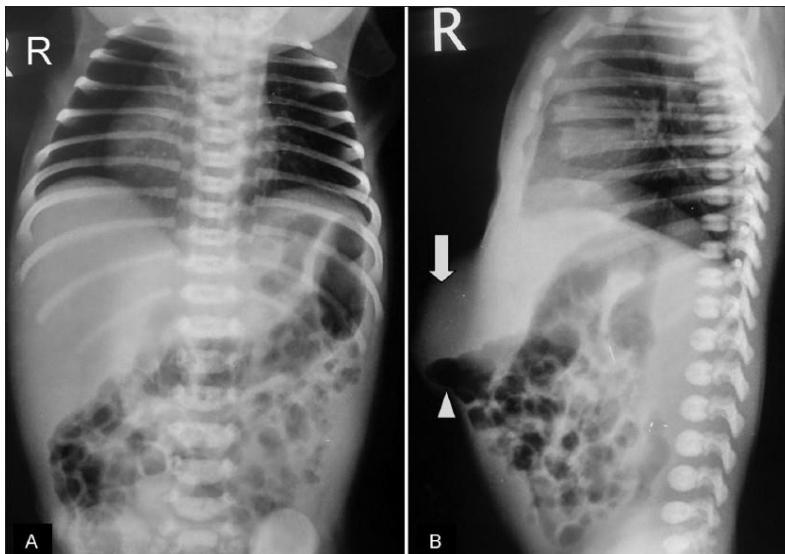
Tomography. בדיקה זו הינה סוג של צילום לא פולשני שנועד להעניק תמונה תלת מימדית

של פנים הגוף. על ידי קרני רנטגן ואותות אלקטרוניים נוצרת הדמיה תלת מימדית של הגוף.

מתוך הדמיה זו ניתן לייצר תמונות של חתכים שונים.

"חתכים" - סדרה של צילומי רנטגן מזוויות שונות שבהמשך מעובדות באמצעות המחשב

לתמונות. דוגמא לשני חתכים שונים :



## 1. מבוא

הפרויקט עוסק בניתוח תמונות רפואיות – תמונות CT, תוך שימוש בעיבוד תמונה ממוחשב ומערכת לומדת.

במרכז הפרויקט עומדת מערכת שתקלוט מהמשתמש סריקת CT.

כיום, מכשיר ה-CT יוצר סדרה של צילומי רנטגן מזוויות שונות שבהמשך מעובדות באמצעות המחשב

לתמונות הנקראות חתכים. ועל ידי תמונות אלו הרופא קובע אבחנות ומסקנות לטיפול.

המערכת תבצע ניתוח של התמונה - תקבע על גביה את מיקום הצלעות ותמספר אותן.



פלט המערכת יהיה אותה סריקת הCT, אך עם סימוני הצלעות והמספור שישמר בחתכים כך שבמעבר בין חתכים, סימון הצלעות ישתנה על פי מיקומם החדש בחתך החדש. המערכת תהיה אפליקציה שיחזיק הרופא בעבודתו השוטפת.

לכן, הפרוייקט יעסוק בזיהוי של מיקום הצלעות ומספורן. נציין כי זוהי פעולה שאורכת כיום זמן רב - נעשית ע"י הרופא בצורה ידנית, כשעה מזמן עבודתו היומית. על כן, המערכת שנבנה היא מערכת חדשנית, שלא קיימת כיום כלל וכלל.

## 2. תיאור הבעיה

### דרישות ואפיון הבעיה

- בעת קבלת החתכים ישנו צורך לזהות בקלות את הצלעות ולמספר אותן, לשם זיהוי האיברים האחרים - ניתן ע"י מספור הצלעות לזהות איברים גדולים/קטנים מהרגיל ולאתר מחלות שונות וע"י כך לסייע למטופלים רבים.
- המציאות כיום היא שרופא מקבל את תוצאות בדיקת הCT, ונאלץ באופן ידני לזהות את הצלעות, לספור אותן, למדוד את המרחקים ביניהם ולהסיק את המסקנות הרפואיות מכך. לדוגמא איבר גדול או קטן מהנורמה ישפיע על גודל הצלעות המוצגות בתמונה. הדרישה היא לאפליקציה שיחזיק הרופא בעבודתו השוטפת, שתעזור לו לדעת את מיקום הצלעות וכך לנתח את תמונת הCT באופן יותר מהיר ולהסיק מסקנות בנוגע למחלות אפשריות. על האפליקציה לתמוך בתמונות רפואיות היוצאות כפלט ממכשירי הבדיקה המוחזקים בבתי חולים, בהם נעשה שימוש כאשר הנבדקים נשלחים על ידי רופאיהם. האפליקציה נועדה:
1. להקל על עבודתו של הרופא.
  2. לחסוך בזמן רופא שהינו זמן יקר מאוד שעלול להגיע למאות שקלים לשעה.
  3. למקד את הרופא בקביעת מסקנות רפואיות ולמנוע עיסוק מיותר בספירה של צלעות וזיהוי מרחקים בעייתיים.
  4. קידום הטכנולוגיה בעבודת הרופא.

### הבעיה מבחינת הנדסת תוכנה - הבעיה האלגוריתמית

הפרוייקט מספק אתגרים בעיבוד תמונה רפואית ובלמידה של תמונות רפואיות, למציאת הצלעות וסימוןן. זהו פרויקט מחקרי-מדעי. מבחינה אלגוריתמית - לזהות את הצלעות מבלי לפספס, אלגוריתמי עיבוד תמונה, אלגוריתמי מערכות לומדות - כדי ללמוד היכן נמצאות הצלעות השונות בחתכים השונים.



### 3. תיאור הפתרון

#### מהי המערכת

אל מול המשתמש המערכת תאפשר :

1. קליטת תמונה.
2. ניתוח התמונה על פי הלימוד.
3. הצגת נתונים.
4. עקיבה.

Use cases

1. הזנת תצלום CT.
2. ניתוח התמונה.
3. התמקדות בחתך מסוים.
4. החלפה בין החתכים השונים.
5. מספור וזיהוי הצלעות ע"י שימוש במכונה ה"חכמה" שלנו.

#### תהליכים ונתוני המערכת

1. מצב הצגת תמונה מוגדלת.
2. מצב עבודה רגיל - מעבר בין חתכים.

#### תיאור הפתרון המוצע

הפתרון הינו מערכת שתקבל את תמונת הCT מהמשתמש.  
המערכת תקבל את התמונה, ותמצא על גביה את הצלעות.  
המערכת תמספר את הצלעות על פי סדר מוסכם מראש וידוע.  
במעבר בין חתכים, סימון הצלעות ישתנה על פי מיקומם החדש בחתך החדש.  
מספר שלבים :

1. קליטת התמונה וקביעת מיקום הצלעות עפ"י מערכת לומדת.
2. סריקת התמונה וזיהוי הצלעות.



### 3. מספור הצלעות.

בנוסף יש לבחור באלגוריתם מערכת לומדת machine learning ורשתות נוירונים. בעזרת אלגוריתם זה נלמד את המערכת כיצד נראות הצלעות בחתכים השונים. לאחר מכן תיקלט תמונה חדשה ועל פי הלמידה שבוצעה, לקבוע היכן הצלעות על התמונה החדשה. בעצם נשתמש בשילוב של שלוש מערכות:

1. קביעת קטגוריית התמונה, על ידי רשת נוירונים קונבולוציוניות CNN.
  2. הבלטה וחיזוד של התמונה התלת מימדית המתקבלת של הצלעות בעזרת CRF. שיטה זו מתייחסת לכל פיקסל במונה בהקשר של הסביבה שלו ולא כיחיד.
  3. מסווג מהיר ומדויק יותר, תוך שימוש ברשת RNN שנחשבת למסווג טוב ומהיר, בעזרת סימון מסגרות לאובייקטים.
- שילוב של שלושת מערכות אלו יספקו לנו רשתות תלת מימדיות ואלו יהיו רשתות מתקדמות ביותר שיציגו פתרון טוב לבעיה המתוארת לעיל.**

## תיאור הכלים המשמשים לפתרון

בפרויקט נעשה שימוש באלגוריתמים מתאימים ללמידה ועיבוד תמונות. הפרוייקט יבוצע על המחשבים האישיים, כאשר יהיה שימוש במערכת הפעלה macOS Sierra. שפת התכנות תהיה Python על ידי העורך Pycharm. מניפולציות על תמונות יבוצעו בעזרת תוכנת MATLAB. צפייה בתמונות בחתכים השונים וסגמנטציות בעזרת תוכנת ITKSNAP. הקוד יימצא במאגר Github. הפרוייקט ינוהל על ידי אתר שיהווה בלוג/יומן לתיעוד והתקדמות בפרויקט. אתר זה יכיל:

1. לוח מטלות.
2. יומן אירועים.
3. מאגר קוד במערכת Github.



## 4. תכנית בדיקות

במערכת מסוג זה, הבדיקות מבוצעות על תמונות וסריקות CT. הבדיקות שיבוצעו יהיו על סריקות שונות, עם בעיות שונות על מנת לבחון את עמידות האלגוריתמים וחוזק הרשתות. ישנה גם חשיבות לזמן ביצוע הסיווג, היות ונרצה סיווג מהיר לקלאסים רלוונטיים.

הבדיקות יבוצעו בכל שלב, בהתאמה לרשת בה אנו עובדים. הבדיקות יכללו:

1. הדפסת פרמטרים ותיעודם, וזאת על מנת לטייב את הפרמטרים בקונפיגורציות שונות.
2. בדיקות לא פונקציונליות – נתיבים לא נכונים, קריאת קבצים לא נכונה, שמירת קבצים לא נכונה.
3. בדיקות פונקציונליות – אימות פעילות המערכת, ביצוע הלמידה כראוי.

## 5. סקירת עבודות דומות בספרות והשוואה

1. במהלך תקופת העבודה על האב טיפוס נקראו מאמרים בנושאים שונים המוצגים לעיל. את עבודת הסקירה ניתן למצוא במסמכי העבודה שלהלן: [מסמך עבודה 1](#) [מסמך עבודה 2](#)
2. מאמרם של Tobias Klinder, Cristian Lorenz, Jens Von Berg, Sebastian P.M. Dries, Thomas Bulow, and Jorn Osterman משנת 2007 עוסק בזיהוי ומספור צלעות מתמונות CT. בעבודה זו יצרו מודל של צלעות על ידי 29 data-sets של תמונות CT. לאחר שיצרו את המודל ומיקמו את הצלעות במקום הנכון, בדקו 18 data-sets חדשים שלא נראו (unseen), מתוכם הצליחו לזהות 16 סטים. עבודה זו דומה להצעה שלנו ואנו מקווים להשיג דיוק גבוה יותר. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.295.3328&rep=rep1&type=pdf>
3. מאמרם של Jaesung Lee and Anthony P. Reeves משנת 2010. בעבודה זו גם מבקשים למצוא את מיקום הצלעות בסריקות CT, אך כאן מציעים לחלק את האלגוריתם ל-4 חלקים: ראשית, לסמן את כל העצמות שבסריקה. שנית, סימון עמוד השדרה. לאחר מכן, סימון שורש הצלע המחובר לעמוד השדרה ולבסוף זיהוי הצלעות. כאן בעצם מדברים על ביזור העבודה, מה שאמור לחסוך בזמן עבודה וריצה. כאן לא מדובר על מספור הצלעות. <http://wonko.via.cornell.edu/publications/JLAR10apub.pdf>



## 6. סיכום ומסקנות

1. עד כה בוצעה עבודה בנושא CNN על מערכת DeepMedic. העבודה כללה:
  - א. התקנה והרצה של הרשתות על פי דוגמאות נתונות.
  - ב. הכנת קונפיגורציות לסטי אימון חדשים.
  - ג. הכנת סטי אימון עבור הבעיה שלנו.
  - ד. שינויים רלוונטיים להתאמה לסטים החדשים.
  - ה. הרצה וקבלת תוצאות.
2. הבנה והכנה של מערכות Cascaded-FCN שמתוכה נבצע CRF להבלטה של האיבר הדרוש, ומערכת Faster RNNLM שמתוכה נממש מסווג מהיר, יעיל וטוב.
3. ביצוע מניפולציות על תמונות בעזרת MATLAB.

## 7. נספחים

### א. רשימת ספרות \ ביבליוגרפיה

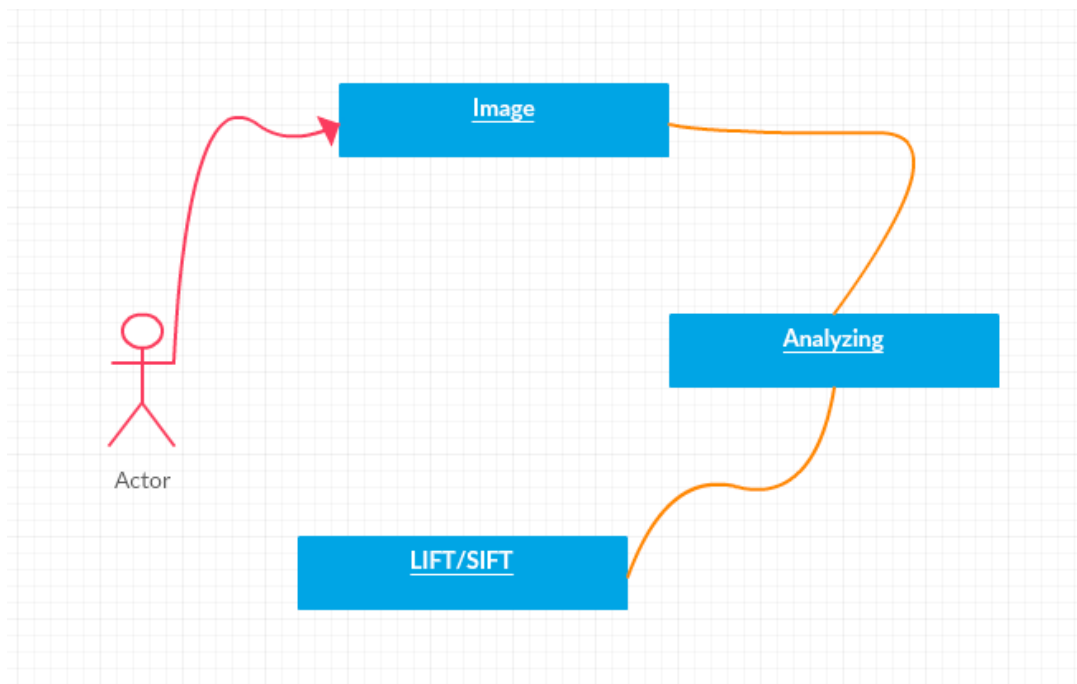
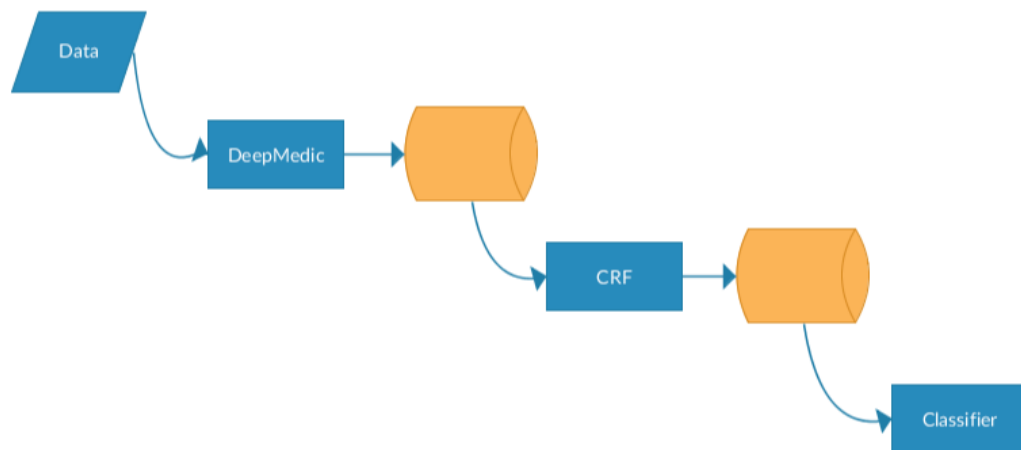
1. הערכת מיקומים בצילומי CT של החזה, מאת Jaesung Lee, Alberto M. Biancardi, Anthony P. Reeves, David F. Yankelevitz, and Claudia I. Henschke שנת 2009  
<http://www.via.cornell.edu/publications/JLAB09pub.pdf>
2. הבנה ומספור של מבנה האנטומי של הצלעות בצילומי CT, מאת X. Zhou, T. Hayashi, M. Han, H. Chen, T. Hara, H. Fujita, R. Yokoyama, M. Kanematsu and H. Hoshi משנת 2009  
<http://www.fjt.info.gifu-u.ac.jp/publication/602.pdf>
3. <http://cs231n.stanford.edu/syllabus.html> קורס מקוון של אוניברסיטת סטנפורד, מכיל חומר ומידע על רשתות נוירונים, רגרסיה לינארית ורשתות קונבולוציה CNN, בהם יש כוונה להשתמש במהלך העבודה כעיבוד תמונה.





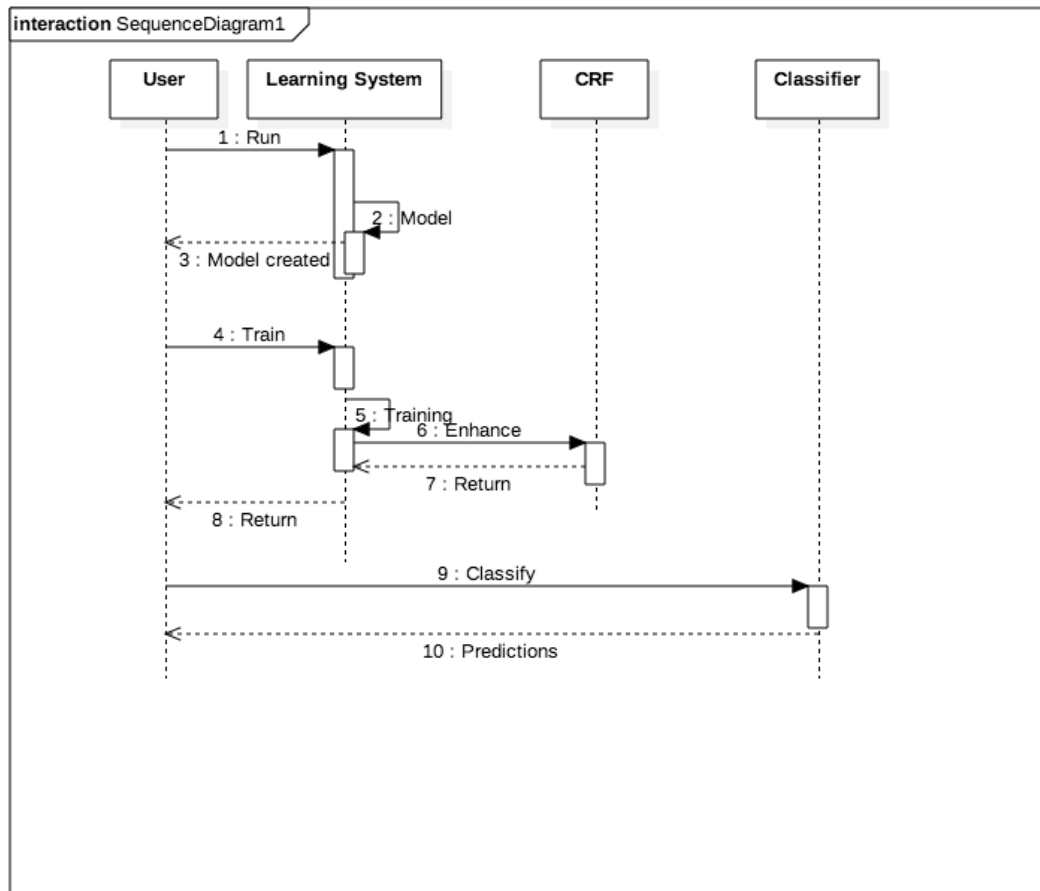
## ב. תרשימים וטבלאות

ארכיטקטורת המערכת : UML – Modeling

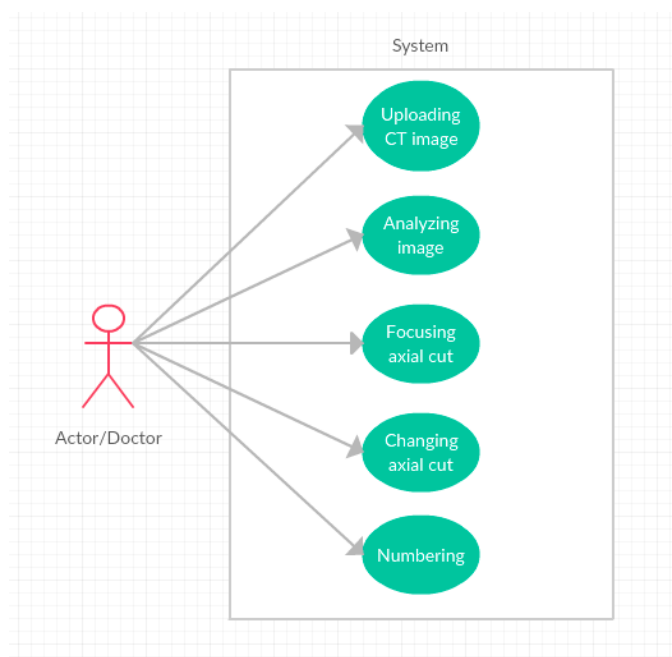




תרשים Sequence Diagram :



תרשים Use cases :





## ג. תכנון הפרויקט

פגישת קבלת נושא לפרויקט ממנחה	11.8.2016
פגישת זוג לסגירת שלב ההתנעה ופתיחת שלב ההצעה	26.10.2016
סיום שלב ההצעה - הגשת הדו"ח	4.12.2016
הצגת אב-טיפוס - אלגוריתמים ראשוניים לניתוח תמונת CT	12.2.2017
בדיקת מקרי קצה ותמונות מיוחדות	30.4.2017
סיום שלב בניית המערכת - דו"ח בנייה	2.4.2017
מסירת הפרויקט	19.6.2017

## ד. טבלת סיכונים

#	הסיכון	חומרה	מענה אפשרי
1	הערכת חוסר לגודל המערכת - המערכת יותר גדולה ויותר מורכבת ממה שחשבנו בהתחלה.	1	מסמך דרישות, תכנון מפורט, ראשית לבצע את החלקים החשובים וה"כבדים" יחסית של המערכת.
2	שינוי מפרטים - הלקוח- המנחה משנה את הדרישות שלו באופן שמסכן את סיום הפרויקט בזמן.	1	הגברת שיתוף הפעולה, עבודת צוות, קשר שותף עם המנחה. בנוסף צוות הפיתוח יעבוד בצורה מודולרית, מסודרת וגמישה תוך מתן מקום לשינויים.
3	כלי הפיתוח לא עומדים בציפיות - מערכות שאמורות לסייע לנו בפיתוח, לא ממלאות את כל הצפיות שיש לנו מהן	2	ביצוע בדיקות מעמיקות אודות כלי הפיתוח, כגון השפה בה נשתמש, השוואות בין כלי הפיתוח הקיימים השונים.
4	חוסר הכרות של הצוות עם כלי הפיתוח - פייתון, מערכות לומדות, עיבוד תמונה	2	ראשית, למידה בנפרד - כל חבר צוות יתמחה בנושא מסוים. לאחר מכן באמצעות למידה משותפת- העברת הידע של כל חבר צוות לחבר הצוות האחר.



## ה. רשימת טבלת דרישות

פורמט טבלת הדרישות יהיה לפי המקובל בארגון. להלן דוגמא:

### דרישות (User Requirement Document)

#### דרישות חומרה

1. המערכת תתמוך בגרפיקה מתאימה.
2. המערכת תרוץ על מכונה בה מותקנים כל ה Tools הדרושים.

#### דרישות תוכנה

1. המערכת תהיה קלה ונוחה לתפעול.
2. למערכת ילוו הוראות תפעול ברורות.
3. המערכת תספק תוצאות בדיוק גבוה ובזמן נמוך.
4. המערכת תאפשר למשתמש להעלות סט תמונות CT דו ותלת מימדיות.
5. המערכת תאפשר למשתמש להגדיל ולהקטין את התמונה המוצגת בפניו.
6. המערכת תאפשר למשתמש לתת פקודה לניתוח התמונה ולסימון הצלעות.
7. המערכת תאפשר למשתמש לבקש מרחק בין הצלעות.
8. המערכת תאפשר למשתמש לעבור בין תמונה לתמונה עם הסימונים הנדרשים.
9. המערכת תאפשר למשתמש להוסיף הערה לצד התמונה כ-meta-data.
10. המערכת תאפשר למשתמש להוסיף הערה על גבי התמונה במקום מסוים.
11. המערכת תאפשר למשתמש לתעד פעולות להמשך טיפול.
12. המערכת תאפשר דיוק מרבי בזיהוי הצלעות.
13. המערכת תבצע ניתוח התמונה והצגת תוצאה בזמן קצר שלא יעלה על מספר שניות.

#### דרישות נוספות

1. המערכת תיכתב בשפת פייתון.
2. התוכנה תיערך על ידי העורך Pycharm.