



המרכז האקדמי רופין  
המחלקה למדעי המחשב

## **פיתוח מודל למידת מכונה לחיזוי משתני בית שורשים בחממה**

דוח מסכם חלק א' זה הוגש כחלק מהדרישות לקבלת תואר  
"בוגר במדעים" BSc במדעי המחשב  
במרכז האקדמי רופין  
המחלקה למדעי המחשב

על-ידי

**עדן אליהו**

**אורן צ'אושו**

העבודה הוכנה בהנחייתם של ד"ר אבניר פריאל וד"ר יעלא ג'מאן

## **תוכן עניינים:**

|                                   |  |
|-----------------------------------|--|
| <b>1. תקציר</b>                   | 1.1. מטרת העבודה.....<br>1.2. הערך היישומי.....<br>1.3. שיטת המחקר.....<br>1.4. תוצאות ראשוניות..... |
| <b>4</b>                          |  |
| <b>4</b>                          |  |
| <b>4</b>                          |  |
| <b>5</b>                          |  |
| <b>2. מבוא</b>                    |  |
| 6.....                            | 2.1. רקע כללי: האתגר בבריאות אקלים וبيות שורשים.   |
| 6.....                            | 2.2. הפער בין הרצוי למצוי בפועל.....   |
| 7.....                            | 2.3. הבסיס הפיזיקלי-ביולוגי.....   |
| 8.....                            | 2.4. למה למידת מכונה?.....   |
| 8.....                            | 2.5. הפתרון המוצע- חישוני וירטואלי.....  |
| 8.....                            | 2.6. מטרת הפרויקט.....   |
| <b>3. הגדרת הבעיה</b>             |  |
| 9.....                            | 3.1. ניסוח הבעיה.....  |
| 9.....                            | 3.2. שאלת המחקר.....   |
| 9.....                            | 3.3. השערת המחקר.....  |
| <b>4. מושגים בסיסיים</b>          |  |
| 10.....                           | 4.1. מושגים אגראונומיים.....   |
| 10.....                           | 4.2. מושגים בלימידת מכונה.....   |
| 11.....                           | 5. סקירת ספרות.....  |
| <b>5. גישת הפתרון</b>             |  |
| 12.....                           | 6.1. מasad הנתונים.....  |
| 12.....                           | 6.1.1. נתוני רציפים.....   |
| 12.....                           | 6.1.2. נתוני בדים.....   |
| 13.....                           | 6.2. השלמת נתונים היסטוריים.....   |
| 14.....                           | 6.3. עיבוד מקדים ונסכון זמנים.....   |
| 14.....                           | 6.4. סבירות הפיתוח.....  |
| 14.....                           | 6.5. ארכיטקטורת המערכת.....  |
| 14.....                           | 6.5.1. מודל מיקרו-אקלים.....   |
| 14.....                           | 6.5.2. מודל בית השורשים.....   |
| 15.....                           | 6.6. אסטרטגיית איחוד הנתונים.....  |
| 16.....                           | 6.7. הנדסת פיצרים.....   |
| 16.....                           | 6.8. אסטרטגיית אימון ובדיקה.....   |
| <b>6. תוצאות ראשוניות ונתונים</b> |  |
| 17.....                           | 7.1. תוצאות המודלים.....   |
| 17.....                           | 7.1.1. מודל המיקרו-אקלים.....  |
| 18.....                           | 7.1.2. מודל בית השורשים.....   |
| 19.....                           | 7.2. אתגרים ודרכי פתרון.....   |
| 19.....                           | 7.2.1. היעדר ניתוח רציף.....   |
| 19.....                           | 7.2.2. בעיית הדיליות.....  |
| 19.....                           | 7.2.3. רגישות לערכי EC נמוכים.....   |
| 20.....                           | 7.2.4. אי סדרות זמנית של הדגימות.....  |
| 20.....                           | 7.2.5. תלות במצב התחלתי ואפקט הזיכרון.....   |
| 21.....                           | 7.2.6. אי רציפות נתונים אקלים.....   |
| <b>7. תוכנית עבודה ולוח זמנים</b> |  |
| 22.....                           | 8.1. הנדסת פיצרים ושיפור המודל.....  |
| 22.....                           | 8.2. אופטימיזציה של האלגוריתם.....   |



|         |                                |
|---------|--------------------------------|
| 23..... | 8.3<br>ניתוח מעתפת וסימולציות. |
| 23..... | 8.4<br>סיכום וכתייה.           |
| 24..... | 9. <b>ביבליוגרפיה</b>          |

## **איורים:**

Figure 1: Current Workflow – Manual Sampling with Late Detection and Reactive Correction

Figure 2: The Physical-Biological Basis – Factors influencing rootzone variables (EC & pH) resulting from the balance between Irrigation, Climate, and Plant Activity

Figure 3: Data Sparsity Problem – Continuous vs. Discrete Data

Figure 4: Historical Data Imputation – Results of the Microclimate Data Imputation Model

Figure 5: System Architecture – Two-Stage Prediction Architecture

Figure 6: Microclimate Model Results

Figure 7: Preliminary Results – Rootzone Model

Figure 8: Gantt



## 1. תקציר:

### 1.1. מטרת העבודה:

מטרת פרויקט זה היא לפתח מודל חיזוי דינמי מבוסס למידת מכונה (**Machine Learning**) לניהול אופטימלי של בית החממה במחמה.

הפרויקט מתמקד ביצירת מערכת המסוגלת לחזות את ערכי המolicות החשמלית (EC) והחומריות (H<sub>c</sub>) העתידיים, בהינתן נקודת מוצא ידועה (מדידה הנוכחיית).

המודל משקל שלושה וקטורים עיקריים כדי לבצע את החיזוי:

1. **מצב התחלתי:** נתוני EC וה-H<sub>c</sub> שנמדדו בנקודת הזמן הנוכחיית.
2. **תנאי סבבה:** תחזית מיקרו-אקלים בתוך החממה (קרינה, טמפרטורה, לחות והתאדות).
3. **משק גידול:** פרוטוקול ההשקייה והדישון המתוכנן (כמויות מים, הרכבי דשן ומועד השקיה).

### 1.2. הערך היישומי:

יכולת חיזוי זו נועדה לשמש ככל אופטימיזציה עבור המגדל. היא מאפשרת לעבר מגישה "תגובהית" (תיקון ערכיים חריגים לאחר שכבר התרחשו) לגישה "פרויקטיבית" – דיווי מגמות של המלח או שינוי חומריות מראש, והתאמת משתר הדישון בזמן אמיתי כדי לשמור על תנאים אידיאליים לצמח, לחסוך במשאבים ולמקם את היבול.

### 1.3. שיטת המחקר:

**שיטת המחקר ומקורות הנתונים:** המחקר מתבסס על אינטגרציה של נתונים מחמישה מקורות מייעדים שונים, היוצרים יחד תמונה מוצב מלאה של סביבת הגידול:

1. **נתוני אקלים חיצוני:** נאספו מהשירות המטאורולוגי (תחנת בית דגן) וכוללים קרינה, טמפרטורה, לחות ומהירות רוח.
2. **נתוני חיישנים בחממה:** ניטור רציף של המיקרו-אקלים (אוויר) ונתוני קרקע שנאספו באמצעות Data Logger.
3. **נתוני תכנון גידול (משק):** פרוטוקולים מתוכנים של משתר ההשקייה והרכבי הדישון (כמויות ומועדים).
4. **נתוני התפתחות הצמח:** מעקב אחר CISCO הנוף (Canopy Cover) המציג את גודל הצמח וצריכת המים שלו.
5. **דגימות קרקע ידניות:** מדידות מעבדה תקופתיות ומדדייקות של EC ו-H<sub>c</sub> המשמשות כעוגן לאימון המודל.



על בסיס נתונים אלו פותחה שרשרת מודלים (Pipeline) דו-שלבית:

**מודל מיקרו-אקלים:** חיזוי הטמפרטורה, הלחות וההתאות העתידים בתוך החממה על בסיס נתונים חוץ.

**מודל בית שורשים:** חיזוי EC ו-H<sub>2</sub>O על בסיס המיקרו-אקלים הנוכחי, היסטורית ההשקייה והדישון, במצב הצמח. נעשה שימוש באלגוריתמים מסוג GBM ו-LightGBM ו-XGBoost תוך שימוש בשיטת אימוט Walk-Forward Validation המדמה תרחיש זמן-אמת.

#### 1.4. תוצאות ראשוניות:

**מצאים ראשוניים:** נכון לשלב זה של המחקר, המודל מציג תובנות מעניינות לגבי יכולת החיזוי של המשתנים השונים:

- חיזוי H<sub>2</sub>O:** המודל מפגין יציבות יחסית מרשימה. אף על פי שהשגיאה המוחלטת עומדת על כ-0.48°C, השגיאה היחסית היא נמוכה מאוד ועומדת על כ-5% בלבד. נתון זה מעיד כי המודל מצליח למדוד היטב את הדינמייקה הכתימית הבסיסית של המצע, וכי הסטייה האבסולוטית נובעת ככל הנראה מכיויל (Calibration) ולא מחוסר יכולת חיזוי.
- חיזוי EC:** כאן התמונה מורכבת יותר. אمنם שהשגיאה המוחלטת נראהית זניחה כ-0.10 mS/cm אך בשל ערכי המוליכות הנומוכים שנמדדו בתקופה זו, שגיאיה זו מתרגמת לסטייה יחסית משמעותית של כ-34%. ממצא זה מדגיש את האתגר בחיזוי ערכים נמוכים (Low Range Sensitivity) ומסמן את הצורך בשיפור רגישות המודל לשינויים עדינים ברכיביו החדש בהמשך המחקר.
- יציבות בזמן:** בנייתו Walk-Forward של המודל שומר על עקביות בחיזוי לטווחים קצרים ובינוניים ממציא המחזק את הפוטנציאל לשימוש בו ככלי תכנון עתידי.

מצאים אלו מדגימים היכולת לפיתוח מודל חיזוי אפקטיבי כאשר חוזקו הנוכחי הוא בזיהוי מגמות EC יציבות, בעוד את האתגר המרכזי להמשך הוא דיקט המודל בערך EC נמוכים.



## 2. מבוא:

### 1.2. רקע כללי: האתגר בברית אקלים ובית שורשים

החקלאות המודרנית ניצבת בפני אתגר כפול: הצורך למקסם את היבול ליחידת שטח, תוך שימוש בשיטות טכנולוגיות. ביחסו של בית השורשים, המבוססות לרובם על מצעים מנוקקים, השיטה על תנאי בית השורשים (Rootzone) היא המפתח להשתתת מטרות אלו. בנגוד לגידולי שדה בקרקע טבעי, למצע המנוקק יש לאחר קיבול (Buffering Capacity) נמוך, משמעות הדבר היא שככל טעות בדישון או בהשקייה משפיעה באופן מיידי וחרייף על הצמח.

בסביבה דינמית זו, שני המדרדים הקritisטיים ביותר לניטור הם:

- **מוליכות חשמלית (EC):** ממד המיצג את סך המלחים המומסים בתמיסה. שמירה על רמת EC אופטימלית היא איזון עדין – ערך נמוך מדי יגרום לחוסר הדנה ופגעה בקצב הצימוח, בעוד שערך גבוה מדי ייצור לחץ אוסmoti שיקשה על הצמח לקלוט מים, ואף עלול לגרום לצרירות והמתתקת יבול מוקדמת.
- **חומציות (Hc):** ממד הקובע את המסימות והэмיניות של יסודות ההזנה (בעיקר מיקרו-אלמנטים כמו ברזל ומangan). סטייה ב-Hc עלולה להוביל למצב חוסר (Deficiency) או רעליות (Toxicity) גם אם הדשן סופק בכמות הנכונה.

## 2. הפער בין הרצוי למצוי בניתוח

על אף חשיבות הקritisטי של מדרדים אלו, קיבלת תמונה מצב אמינה ורציפה של בית השורשים מהווה כוון "צואר בקבוק" טכנולוגי. המגדל נדרש לבחור בין שתי חלופות שאין מושלמות:

### 1. דגימות מעבדה ידניות (Manual Sampling):

זהי שיטת-h Gold Standard מבוחנת דיווק. דגימת נקז או מיצוי מצע נבדקת במכשור מעבדה מכיל. עם זאת, שיטה זו דורשת כוח אדם וזמן, וכן מתבצעת לרוב בתדריות נמוכה (אחת ליום או לפחות ימים). התוצאה היא "נקודות עיוורות" בזמן; יתכן שבין שתי דגימות תקיןות התרחש אירוע קיצוני של המלחה (Spike) שגורם לנזק בלתי הפיך, אך לא תועד. כמו כן, שיטה זו היא ריאקטיבית (תגובהית) – המידע מתקיים לאחר מעשה.

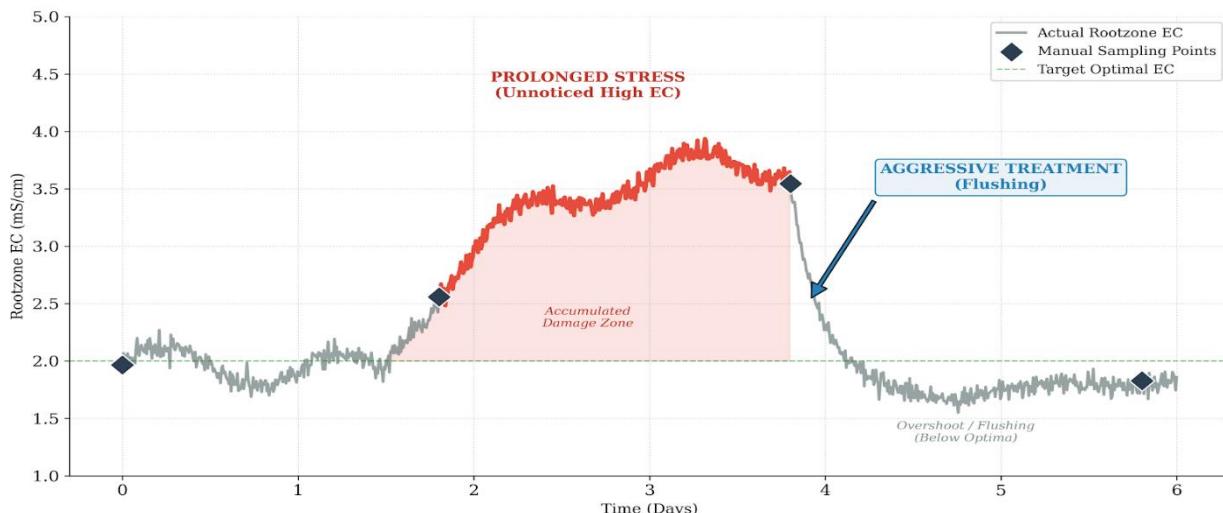


Figure 1: Current Workflow - Manual Sampling with Late Detection and Reactive Correction



## 2. חיישני קרקע רציפים (In-situ Sensors):

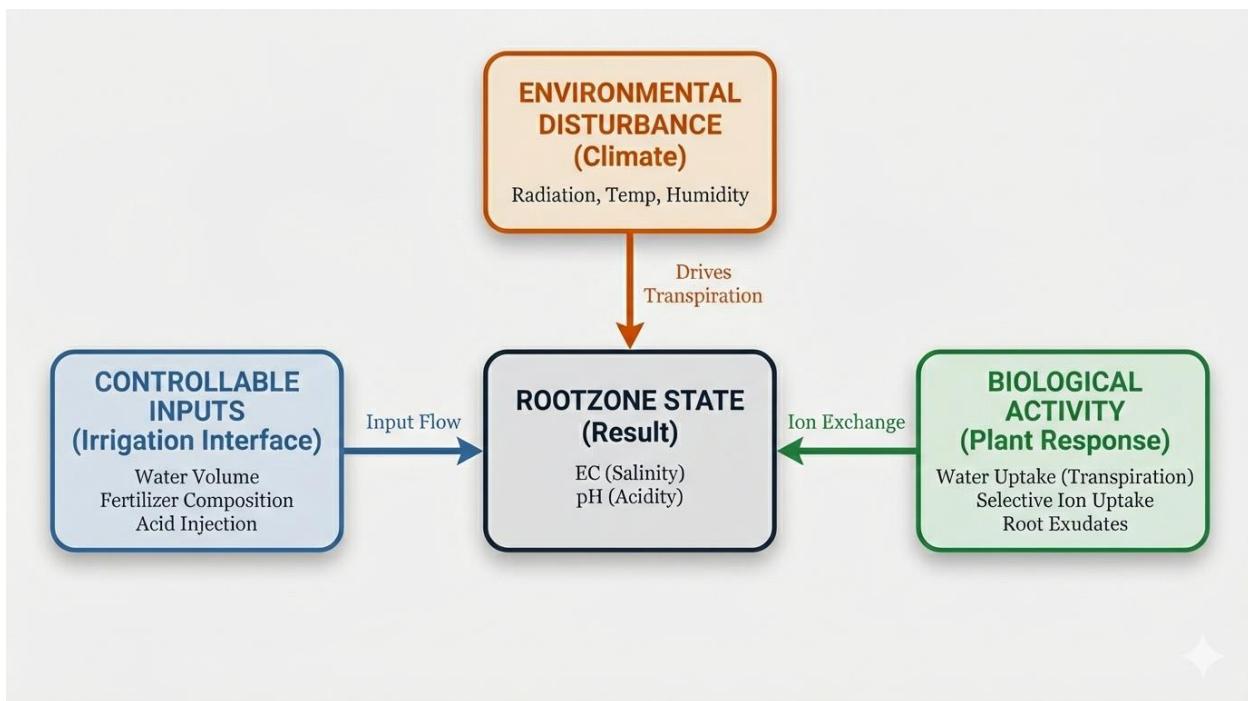
חישנים אלו מספקים רצף נתונים (Time-Series) ומאפשרים זיהוי מגמות בזמן אמת. אולם, השימוש בהם תומן בחובו אתגרים טכניים משמעותיים:

- סחיפת וכיוון (Drift & Calibration): האלקטרודות נוטות לצבור משקעים וביפויים, הגורמים לקריאה "לנדוד" עם הזמן ולספק נתונים שגויים.
- ייצוגיות מרחבית: חישן מודד נקודת בודדת במרחב. ביחס גודלה, או אפילו בתוך עציץ בודד, קיימת שונות מרחבית גבוהה ברטיבות ובמלחיות. חישן הממוקם ליד הטרטיפת יראה ערך שונה לחלוtin מחישן הממוקם בתחום המצע.
- עלות: פרישה רחבה של חיישנים אינטנסיביים דורשת השקעה כספית גבוהה ותחזקה שוטפת.

## 2.3. הבסיס הפיזיקלי-ביולוגיי

הקשרי בחיזוי השורשים נובע מכך שבית השורשים הוא מערכת דינמית המושפעת ממאזן מסה (Mass Balance) משתנה. הצמח קולט מים ודשן בקצבים שונים:

- ביוםים של קרינה גבוהה: הצמח מבצע דיות (Transpiration) מוגברת ו"שوتה" בעיקר מים נקיים, מה שמותיר את המלחים במצע וגורם לעלייה EC מהירה.
- ביוםים מעוניינים: קצב הדיות יורך, וצריכת המים והדשן מתאוזנת. המודל שלנו נדרש ללמוד את הדפוסים הללו (הקשר הלא-lienair בין קרינה לצריכת מים) ולהזות את הנסיבות המלחים שנובעת מהם.





## **2.4. למה למידת מכונה?**

באופן מסורתי, ניתן לחזות תהליכיים אלו באמצעות משוואות פיזיקליות מורכבות. אולם, מודלים אלו דורשים פרמטרים רבים שקשה מאוד ללמוד בשטח (כגון מוליכות הידראולית של המצע, נקבוביות משטנה ושטח פנים של השורשים). גישה למידת המכונה (Data-Driven Approach) עוקפת את הצורך במדידת הפרמטרים הללו. במקרה "להבין" את הפיזיקה ברמתה המיקרו, המודל יוכל לומד את הצורך בתנагות האפקטיבית של המערכת. ישירות מתוך הנתונים ההיסטוריים, ובכך מאפשר דיקוגרף בעלות חשיבות נמוכה.

## **2.5. הפתרון המוצע: "חישון תוכנה" מבוסס נתונים**

פרויקט זה מציע גישה חדשה מבוססת על תפיסת ה-Soft Sensing ("חישון תוכנה"). ההנחה היא שקיים קשר פיזיקלי וכימי הדוק בין תנאי הסביבה החיצוניים (מזג האוויר), פועלות המגדל (השקייה ודישון) והtaggova בבית השורשים.

באמצעות אלגוריטמים של למידת מכונה (Machine Learning) ניתן למגדל קשרים מורכבים ובלתי-lienarיים אלו, וליצור תחזית מדויקת של מצב בית השורשים על בסיס נתונים זמינים וזולים, ללא תלות בחישוני קרקע יקרים.

## **2.6. מטרות הפרויקט**

1. אינטגרציית נתונים: ייצירת מסד נתונים מאוחד (Master Dataset) המנरמל ומשלב מקורות מידע הטרוגניים: נתונים אקלים חיצוני, בקרת השקיה, ודוגמאות מעבדה.
2. פיתוח מודל חיזוי (Predictive Modeling): בניית מודל למידת מכונה המסוגל לחזות את ערכי EC וה-Ηק בצעד הזמן הבא בהינתן המצב הנוכחי ותחזית האקלים וההשקה. המטרה היא לספק למגדל יכולת פרואקטיבית לצפות בעיתות לפני שהן מתרחשות.
3. וליזציה בתנאיאמת: בוחינת ביצועי המודל בשיטת Validation Walk-Forward המדמה את התקדמות עונת הגידול, כדי להבטיח שהפתרון יציב, אמין ונitin לשימוש במערכות תומכות החלטה.



## 3. הגדרת הבעיה

### 1.3. ניסוח הבעיה

בגידולים אינטנסיביים על מצע מנוטק, בית השורשים (Rootzone) מהוות "קופסה שחורה" דינמית. הבעיה המרכזית אליה מתמודד המחקר נובעת מהפער הקיים בין הצורך בברכת רציפה ומדיקת לבין אמצעי הניטור הקיימים כיום:

1. מגבלת הניטור המקוטע: בדיקות מעבדה (ניקוז/מצע) הן מדיקות ארוכות תומנות מצע נקודתיות ("Snapshot") ודיליה בזמן. בין דוגמה לדוגמה, המצע עשוי לעבור תנודות חrifיפות ברמת המיליחות (EC) והחומריות (H<sub>c</sub>) כתוצאה משינוי אקלים או תקלות דישון, ללא ידיעת המגדל.
2. אמינותות חיישנים רציפים: חיישנים המוצבים בתוך המצע (in-situ) נתונים לסבול מבעיות כיול, סחיפה (Drift) ורעד אלקטרוני. בנוסף, הם מודדים נקודת בזetta במרחב ומתקשים לייצג את המצע המוצע כולו, הסובל משונות מרחבית גובהה.
3. ניהול תגובה (Reactive Management): בהיעדר יכולת חיזוי, המגדל נאלץ לפעול בגיןה "כיבוי שריפות" תקין ערכי ה-EC וה-H<sub>c</sub> רק לאחר שכבר חרגו מהטוווח הרצוי והנ扎ק הפוטנציאלי לצמח כבר החל.

### 2. שאלת המחקר:

האם ניתן לפתח מודל למידת מכונה (Machine Learning) המסוגל לחזות בדיקות אגרונומיות את ערכי המוליכות החשמלית (EC) והחומריות (H<sub>c</sub>) בבית השורשים לטווח הזמן הקרוב, בהתאם על נתוני אקלים, משטר השקיה ודישון ומדידות עבר?

מתוך שאלת זו נגזרות מספר שאלות משנה:

1. **דיקוק:** מהי רמת הדיקוק המksamילית שניתנת להשיג בחיזוי (نمzd בMAE/RMSE) והאם היא מספקת לקבלת עזר החלטה תפעולי?
2. **חשיבות המשתנים (Feature Importance):** אילו משתנים משפיעים בצורה חזקה ביותר על השינויים בבית השורשים? (למשל: האם קרינה משפיעה יותר מכמויות המים? האם ההיסטוריה של הדישון חשובה יותר מהטמפרטורה הרגעית?).
3. **יציבות:** האם המודל מסוגל לשמר על יציבות בחיזוי לאורך זמן (Walk-Forward) בתנאי אמת משתנים של עונת הגידול?

### 3.3. השערת המחקר:

ההשערה היא שקיים קשר דטרמיניסטי (גם אם לא ליניארי ומורכב) בין עומס האקלים החיצוני הקובע את קצב הדירות (Chpiration), אספקת המים והדישון, ושינוי הרכיבים בקרקע.

אנו משערים כי אלגוריתמים מבוססי עצים כגון XGBoost או Random Forest יצליחו ללמידה את דפוסי "התגובה וההשיה" של המערכת, ולגשר על הפער שבין המדידות הידניות, ובכך לשמש כ"חישון וירטואלי" אמיתי.



## 4. מושגים בסיסיים

לצורך הבנת העבודה, להלן הסבר על המושגים המרכזיים בתחום האגרונומיה ולמידת המכונה בהם נעשה שימוש:

### 4.1. מושגים אגרונומיים

- **מוליכות חשמלית (EC - Electrical Conductivity)**: ממד לכמות המלחים המומסים בתמיסת הקרקע. בחקלאות, זהה הממד המרכזי לרמת הדישון. ערכיהם גבוהים מדי עלולים לגרום להמלחה ופגיעה בצמח, וערכים נמוכים מעידים על חוסר בדשן. נמדד ביחידות של mS/cm.
- **חומציות (H<sub>c</sub>)**: ממד לרכיב יוני המימן בתמיסה. ה-H<sub>c</sub> קובע את זמינות המיקרו-אלמנטים (כגון ברזל ומangan) לצמח.
- **התאדות-דיות (ET0 - Evapotranspiration)**: סך המים המתאדים מהקרקע והמופרשים מהצמח (טרנספירציה). זהה ממד קרייטי לקביעת מנות השקיה, המשפע מקרינה, טמפרטורה, לחות ורוח.
- **בית שורשים (Rootzone)**: נפח הקרקע או המצע שבו מתפתחת מערכת השורשים של הצמח. התנאים באזורי זה משפיעים ישירות על קצב גידילת הצמח ובריאותו.

### 4.2. מושגים בלמידת מכונה

- **מודל רגסיה (Regression Model)**: מודל שנועד לחזות ערך מסוים רציף (בניגוד לשינוי לקטגוריות). בפרויקט זה אנו חוזים בערכים רציפים של משתני מיקרו אקלים, EC ו-H<sub>c</sub>
- **XGBoost**: אלגוריתם מתקדם מבוסס עצי החלטה (Gradient Boosting), הידוע ביעילותו ובמהירותו. שימש בפרויקט להשלמת נתונים אקלים חסרים ובחיזוי משתני בית השורשים.
- **LightGBM (Light Gradient Boosting Machine)**: אלגוריתם למידה מתקדם המבוסס על שיטת Gradient Boosting (שיפור הדרגתית של עצי החלטה). האלגוריתם מתאפיין בשיטת גידילת עץ "יחודית" (Leaf-wise) המאפשרת מהירות אימון גבוהה במיוחד, ניצול יעיל של זיכרון וධיק רב, והוא נחשב לאידיאלי לעיבוד סדרות עיתיות ונתונים בהיקף רחב. שימש בפרויקט לחיזוי ערכי מיקרו אקלים עתידיים בהתאם על בסיס האקלים הנוכחי.
- **Walk-Forward Validation**: שיטת אימوت למודלים של סדרות עיתיות (Time Series) במקומות חלוקה אקראיית ל"אימון" ו" מבחן", אנו מאמנים את המודל על נתונים עבר עד נקודה מסוימת, בודקים על העתיד הקרוב, ואז "מוציאים את החלון" קידמה. שיטה זו מדמה בבדיקה כיצד המערכת תפעל במצבות, יומ אחר יומ.
- **הנדסת פיצ'רים (Feature Engineering)**: תהליך יצירת משתנים חדשים מתוך הנתונים הגולמיים כדי לשפר את יכולת הלמידה של המודל. דוגמאות מפרויקט זה: "זמן שעבר מאז ההשקיה الأخيرة", "סכום הקרן המctrber ב-24 שעות האחרונות".

## 5. סקירה ספרות

השימוש במודלים מבוססי נתונים (Data-Driven Models) בחקלאות מדיקת צבר תאוצה משמעותית בשנים האחרונות. מחקרים שונים הראו כיצד ניתן להחליף או לתגבר מדידות פיזיקליות יקרות באמצעות אלגוריתמים המבוססים נתונים סביבה ותפעול.

אחד האתגרים המרכזיים בגידול בחממות הוא השליטה ברמת המלחות (EC) בבית השורשים. במחקר של **Moon et al. (2018)** בחנו החוקרים את היכולת לחזות את ה- EC של תמייסת ההזנה במערכות הידרопוניות סגורות (Closed-loop soilless cultures). החוקרים השתמשו ברשת ניירונים מסווג (Recurrent Neural Network) ופרט במודל LSTM הידוע ביכולתו ללמידה סדרות עיתיות. המחקר הראה כי שילוב של **משתני אקלים** (קרינה, טמפרטורה וחותם יחסית EC) יחד עם **נתוני פעולה** (כמויות השקיה ורכיב דשן במיל השקיה), מאפשר לחזות בדוק רב את ה- EC העתידי במציאות. המשקנה המרכזית ממחקירות זה, הרלוונטיות לעובdotנו, היא שקיים קשר הדוק ונitin למידול בין עומס האקלים החיצוני לבין הדינמיקה הכימית בבית השורשים, וכי היסטוריית הנתונים (Time-lagged features) היא קריטית לביצוע המודל.

בעוד ש-Moon התמקד בכימיה של המים (**Lü et al. (2025)** הציגו גישה מתקדמת לחיזוי רטיבות קרקע בבית השורשים (Root Zone Soil Moisture – RZSM). המחקר התמודד עם הקושי למדוד רטיבות בעומק הקרקע והציג מודל היברידי המשלב מודל פיזיקלי הידרולוגי (Hydrus-1D) עם מודל למידה עמוק CNN-LSTM-Attention . המחקר הדגים כיצד מודלים של למידה עמוקה (Deep Learning) יכולים למודד את הקשרים הלא-ליניאריים המורכבים בין רטיבות פני השטח (הניתנת למידה קלה או חישמה מרוחק) לבין המתרחש בעומק בית השורשים. השימוש במנגנון קשב (Attention Mechanism) אפשר למודל להתמקד בטוויח הזמן המשפיעים ביותר על המצב הנוכחי. אף על פי שמחקר זה בוצע בתנאי שדה (Oasis field) ולא בחממה, הוא מחזק את ההנחה כי ניתן לייצר "חיישן וירטואלי" לתנאי עומק הקרקע על בסיס נתונים סביבתיים, וכי שילוב של אלגוריתמות המבאות בחשבון את המימד הרציף של הזמן כגון LSTM או נגזרות משפר משמעותית את הדיקון בהשוואה למודלים סטטיים.

הספרות הקיימת מצביעה על היתכנות גבוהה לשימוש במכשיר מכונה לניטור בית השורשים. בעוד Moon et al. התמקד ב- EC-הן, **Lü et al. (2025)** תור שימוש באlgorigthms של למידת מכונה מודל אחד החוצה **הן את ה- EC** והן את ה- **pH** תוך יכולת פרשנות ומתאים לשימוש במערכות בקרה תפעוליות בחממות מסחריות.

Literature Review: Comparative Analysis

| Research Source    | Technology / Method    | Key Application               | Target Variable |
|--------------------|------------------------|-------------------------------|-----------------|
| Moon et al. (2018) | RNN                    | EC Prediction - Hydroponics   | EC              |
| Lü et al. (2025)   | Hybrid (Attn-LSTM-CNN) | Soil Moisture - Field (RZSM)  | RZSM            |
| This Project       | Gradient Boosting      | Integrated Greenhouse Control | EC & pH         |



## 6. גישת הפתרון

כדי להתמודד עם האתגר של חיזוי תנאים ביו-כימיים מורכבים בבית השורשים, פיתחנו גישה מערכית המבוססת על **ארכיטקטורה דו-שלבית (Two-Stage Pipeline)**. גישה זו מפרקת את הבעיה לשני תתי-מודלים, כאשר הפלט של הראשון משמש כקלט לשני. מבנה זה מאפשר לנו להשתמש בנתוני אקלים חיצוניים (הזמן נימודים תמיד) כדי לחזות את המתරחש בתוך המצע.

### 6.1. מסד הנתונים

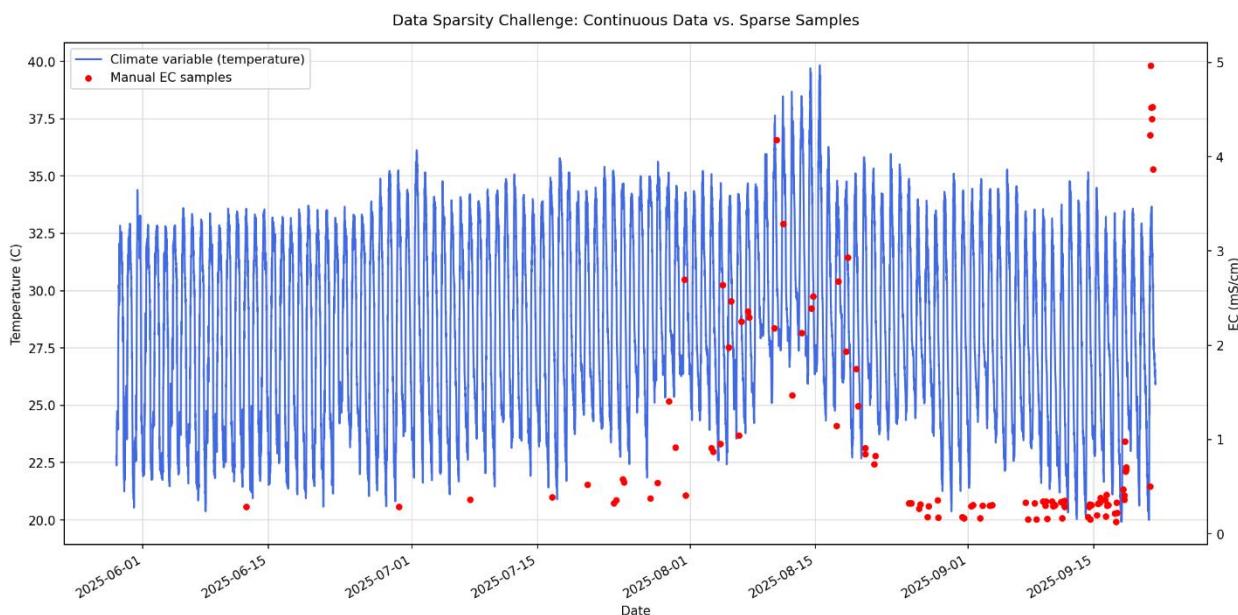
המחקר מתבסס על נתונים אגרו-מטאורולוגיים שנאספו בחמתת המחקר במהלך עונת הגידול 2025 (בחודשים מאי-ספטמבר). לצורך אימון ובדיקה המודלים, נבנה מסד נתונים מאוחד המשלב מקורות מידע הטרוגניים, שעבר תהליכי ניקוי וסנכרון. הנתונים נחלקים לשתי קטגוריות עיקריות, הנבדלות ברצולוציית הזמן שלהן:

#### 1. נתונים רציפים (High-Frequency Data) :

- מקור: תחנה מטאורולוגית (בית דגן) ולוגר נתונים פנימי בחמתה.
- היקף: כ-16,000 רשומות זמן.
- רזולוציה: דגימה כל 10 דקות באופן רציף.
- משתנים: קרינה גלובלית, טמפרטורה, לחות יחסית, מהירות רוח וчисוב התאדות פוטנציאלית( $ET_0$ ).

#### 2. נתונים בדידים (Low-Frequency Data) :

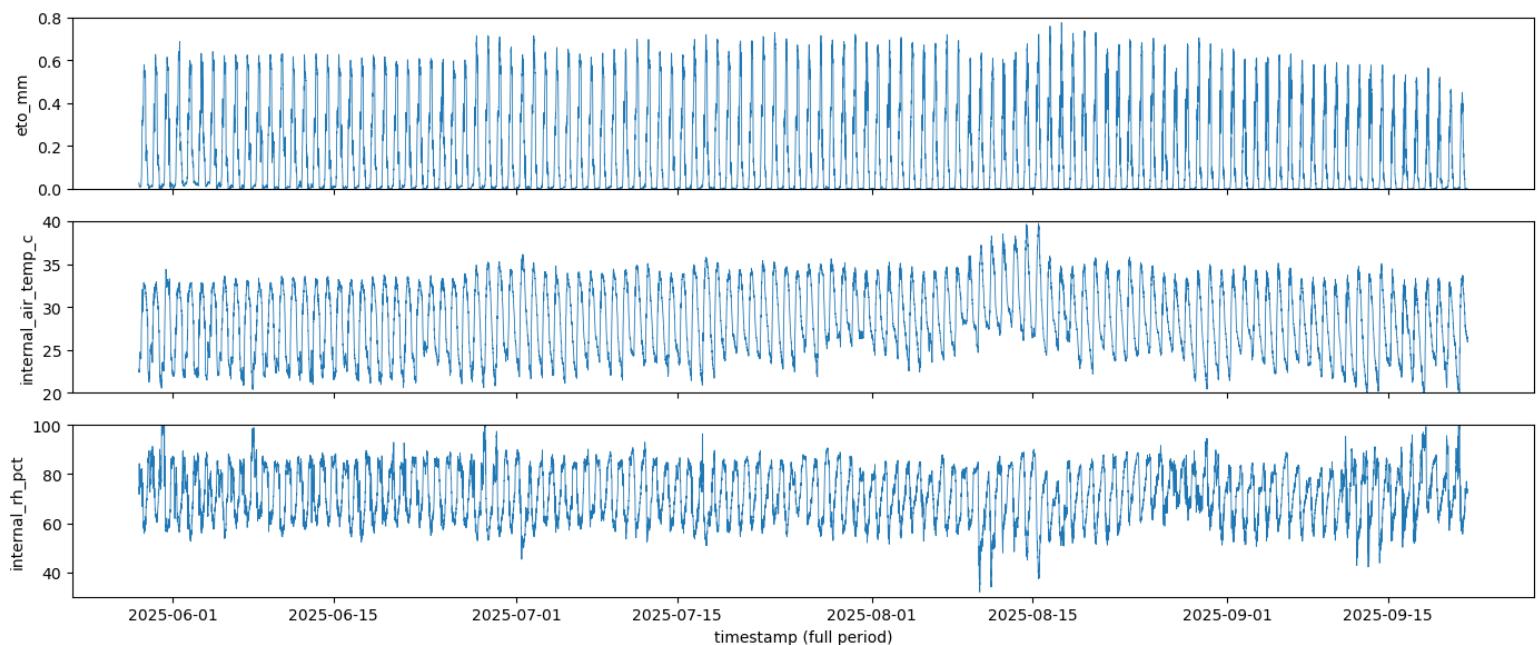
- מקור: דגימות מעבדה ידניות של מי נקי ומצע.
- היקף: כ-100 דגימות תקינות של EC ו- $H^+$ .
- האתגר: נתונים אלו מאופיינים בדלילות (Sparsity) גבוהה ובמרווחי זמן לא קבועים (בין יומם ל-5 ימים בין דגימות). עובדה זו היזגה את האתגר האלגוריתמי המרכזי בפרויקט, שכן היחס בין שני המשתנים הקלט (אקלים) למשתני המטרה (מעבדה) עומד על 1:200.



## 6.2. השלמת נתונים היסטוריים (Data Imputation using XGBoost):

אחד האתגרים המשמעותיים בפרויקט היה היעדר רציפות נתונים האקלים הפנימי. חישון הנתונים (Data Logger) הותקן בחממה רק במהלך העונה, כך שבעבר התקופה הראשונה של הניסוי היו בידינו נתונים מעבדה (H/C) ונתוני חוץ (בית דגון), אך ללא נתונים אקלים פנימי תואמים.

כדי לא לשב את דגימות המעבדה היקרות מתקופה זו, פיתחנו מודל רגסיה "יעודי" מבוסס XGBoost להשלמת הנתונים החסרים. המודל אומן על התקופה המאוחרת (בה היו נתונים מלאים) כדי למדוד את הקורלציה בין האקלים החיצוני לפנימי ( $R^2 > 0.95$ ), ולאחר מכן הופעל "לאחר" כדי לשחזר בדיקוק רב את הטמפרטורה והלחות הפנימית עבור התקופה שקדמה להתקנת החישון. פעולה זו אפשרה לנו להגדיל את סט האימון האפקטיבי בכ-30%. נתונים האקלים לאחר השלמה:



Microclimate Model Performance Results

| Target Variable        | MAE    | R <sup>2</sup> Score (%) |
|------------------------|--------|--------------------------|
| EtO (mm)               | 0.0211 | 94.56%                   |
| Internal Air Temp (°C) | 0.3832 | 98.02%                   |
| Internal RH (%)        | 2.3773 | 90.08%                   |



### 6.3. עיבוד מקדים:

בוצע תהליך ניקוי רעשים (Noise Reduction) לסינון נתונים לא פיזיולוגיים מהחומרيين. כמו כן, בוצע סנכרון זמנים (Time Synchronization) כדי להציגם לכל דגימות מעבדה בוודת היסטורית האקלים וההשקייה המדוקנית שהובילה אליה ("חולון השפעה").

### 6.4. סביבת הפיתוח

מערכת החיזוי פותחה בסביבת Python 3.13. תוכן שימוש בספריות המרכזיות הבאות:

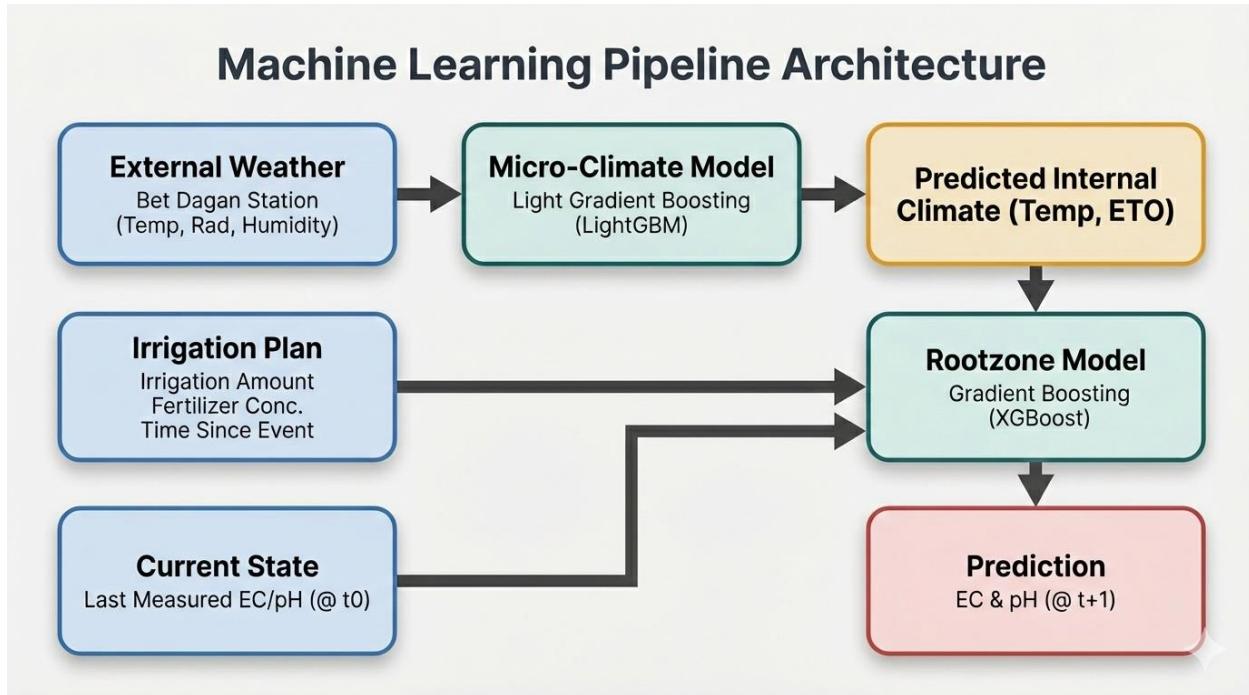
- **Pandas** לטיפול ועיבוד סדרות עיתיות.
- **Scikit-Learn** לימוש אלגוריתמי ML קלואסיים (Random Forest) חלוקת DATA-Train/Test Split ומטריקות הערכה.
- **Gradient Boosting** עילה ששימשה למידול המיקרו-אקלים ובית השורשים.
- **Matplotlib/Seaborn** לוויזואליזציה וניתוח נתונים. הפיתוח בוצע על גבי מחברות Jupyter Notebooks המאפשרות מחקר אינטראקטיבי ותיעוד התהילר.

### 6.5. ארכיטקטורת המערכת

המערכת בנויה משני מודלים עוקבים המזינים זה את זה:

- שלב 1: מודל "תרגום" אקלים (Micro-Climate Model):  
מודל זה משמש כ"מתווך". הוא מקבל כקלט את תחזית מג' האוויר החיצונית (תחנת בית דגן) וחוצה את תנאי המיקרו-אקלים בתוך החממה (טמפרטורה פנימית, לחות יחסית, קירינה ו-ET0).
  - **הרצינול:** הצמח וbeit השורשיםמושפעים מהתנאים בתוך המבנה, השונים משמעויות מתנאים בחוץ בשל אפקט החממה. שלב זה מאפשר לנו להשתמש בתחזיות מטאורולוגיות סטנדרטיות כבסיס לחיזוי.
  - **אלגוריתם LightGBM (Light Gradient Boosting Machine):** הבחירה ב-LightGBM נעשתה בשל יתרונותיו המובהקים בעיבוד סדרות עיתיות ובਮירות האימון הגבוהה שלו בהשוואה לאלגוריתמים מסורתיים. המודל אומן למדוד את פונקציית המעבר הלא-ליניארית שבין הקירינה, הלחות והטמפרטורה בחוץ, לבין הטמפרטורה, הלחות וההתאות (ET0) בתוך המבנה המוגן.
- שלב 2: מודל בית השורשים (Rootzone Dynamics Model):  
זהו מודל הליבה. הוא מקבל את המיקרו-אקלים החזוי (שלב 1), את תוכנית ההשקייה והדישון, ואת המצב הנוכחי וחוצה את ערכי ה-EC וה-H<sub>2</sub>O בצעד הזמן הבא.

- אלגוריתם XGBoost Regressor זה נבחר בשל עמידותו לרעש (Noise Robustness) ויכולתו להתמודד היטב עם מערבי נתוניים קטנים יחסית (כ-100 דוגמאות ידניות) מבל' להיכנס לבעיית התאמת-יתר(Overfitting).



#### 6.6. אסטרטגיית איחוד הנתונים

אחד האתגרים המרכזיים בפרויקט היה השוני בrzולוציות הזמן של הנתונים:

- **נתונים "צפויים"**: נתונים אקליםולוגיים נתונים (כל 10 דקות).
- **נתונים "دليلים"**: דוגמאות מעבדה ידניות (אחת ליום או למספר ימים רוב הזמן).

הפתרון: ייצור Master Dataset בrzולוציה אינידית של 10 דקות.

הנתונים הדليلים "הוזמדו" לנקודות הזמן הקרובות ביותר (Nearest Neighbor Interpolation), אך האימון של מודל בית השורשים בוצע רק על גבי "חלונות זמן" שיש בהם דוגמת אמת בעוד החזוי יכול להתבצע באופן רציף.



## (Feature Engineering 6.7)

כדי לאפשר למודל "להבין" את התהליכים הביולוגיים והפיזיקליים, הנתונים הגלומיים עובדים למשתנים חדשים בעלי משמעות(Domain Knowledge) :

- **משתני צבירה (Cumulative Features):** במקומם להסתכל רק על הקרןנה הרגעית, המודל מחשב את "סך הקרןנה ב-24 שעות האחרונות" או "סך ההשקייה ב-8 שעות האחרונות".  
משתנים אלו מייצגים את ה"זיכרון" של המערכת.
- **זמן מאז אירוע (Time-Since Events):** משתנים המודדים כמה זמן עבר מאז ההשקייה או הדישון האחרון. משתנים אלו קריטיים לחיוי תכניות של עלייה וריכוז מלחים.
- **מצב הצמח (Canopy Cover):** שילוב נתוןCisco הנוף אינדיקציה לגודל הצמח וליכולת שלו לקלוט מים ודשן (Transpiration rate).

## (Walk-Forward Validation 6.8)

מכיון שמדובר בסדרה עיתית (Time Series), לא ניתן להשתמש בחלוקת אקראית רגילה (Random Split) לסת אימון וסת מבחן, מחשש לצלגת מידע (Data Leakage) מהעתיד לעבר.

במקום זאת, יושמה שיטת Walk-Forward Validation המדמה את התקדמות הזמן.

בפרויקט זה הגדרנו שני פרוטוקולים שונים של אופק חיזוי (Forecast Horizon) בהתאם לסוג המודל וזמןנות הנתונים:

1. עברו מודל המיקרו-אקלים (נתונים רציפים):
    - המודל אומן על נתונים היסטוריים וביצע חיזוי של 24 שעות קידמה בכל איטרציה.
    - הרצינל: לאפשר למגדל לתוכנן את משטר ההשקייה ליום המחרת במלואו על בסיס התחזית.
  2. עברו מודל בית השורשים (נתונים בדידים/دلילים):
    - המודל אומן על כל הדגימות הידניות עד נקודה  $t$ , וביצע חיזוי של דגימה אחת קידמה  $1 + t$ .
    - הרצינל: מכיוון שמדידות האמת (Ground Truth) של EC ו-H<sub>2</sub>O מגיעות מהמעבדה באופן לא סדיר, המודל נבחן ביכולתו לחזות את תוצאת המעבדה הבאה בתור, ללא תלות במשך הזמן ש עבר (שעות או ימים).
- לאחר כל חיזוי, חלון האימון הורחב לכלול את הנתונים החדשניים ("Real-Time Update") והתהlijר חזר חלילה עד סוף תקופת הניסוי.



## 7. תוצאות ראשוניות ותגירים

### 7.1. תוצאות המודלים

נכון לשלב זה של המחבר, המערכת נבחנה על נתונים שנאספו במשך מספר חודשים (מיוני עד ספטמבר). להלן הביצועים של שני המודלים המרכזיים:

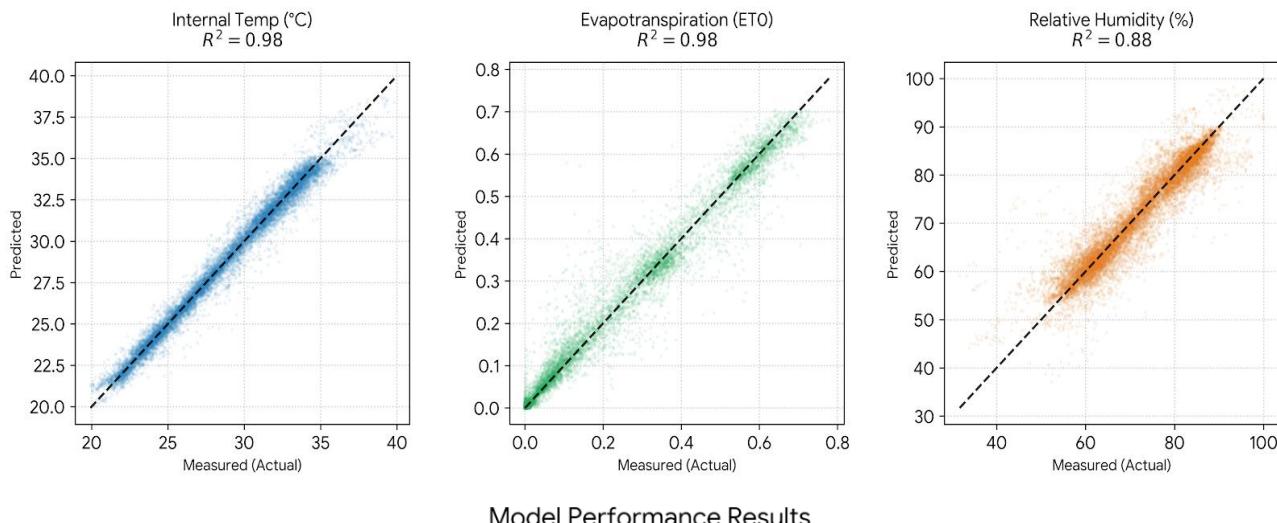
#### A. מודל המיקרו-אקלים:

רמת הדיוק המוגבהת בחיזוי האקלים הפנימי מבוססת שלב זה כעוגן פיזיקלי אמין בתוך הפיפליין, המבטיח למודל הליבת קלטים סביבתיים יציבים לחיזוי ערכי EC וה-PH.

- טמפרטורה פנימית: התקבל מקדם מתאם  $R^2$  של 0.97 עם שגיאה ממוצעת של פחות מ- $0.8^\circ$  צלזיוס.

- לחות יחסית: הדיק עמד על  $R^2$  של 0.87 עם שגיאה ממוצעת של 2.5% בלבד וטעות יחסית של 3.5%, נתון קרייטי לחיזוי התוצאות.

- התוצאות (ET0): התקבל דיק גבוה במיוחד במיחוד עם מקדם מתאם  $R^2$  של 0.97, ושגיאה מוחלטת ממוצעת (MAE) של 0.016 מ"מ בלבד. דיק זה הוא קרייטי שכן ET0 מהווה את הבסיס לחישוב מנתת השקיה.



| Target Variable                          | MAE     | RMSE    | $R^2$ Score |
|--|---------|---------|-------------|
| Internal Air Temp ( $^{\circ}\text{C}$ ) | 0.3863  | 0.5062  | 97.86%      |
| Internal Radiation ( $\text{W/m}^2$ )    | 20.5592 | 39.1329 | 96.89%      |
| Evapotranspiration (mm)                  | 0.0158  | 0.0296  | 97.18%      |
| Internal Relative Humidity (%)           | 2.5209  | 3.2580  | 86.77%      |

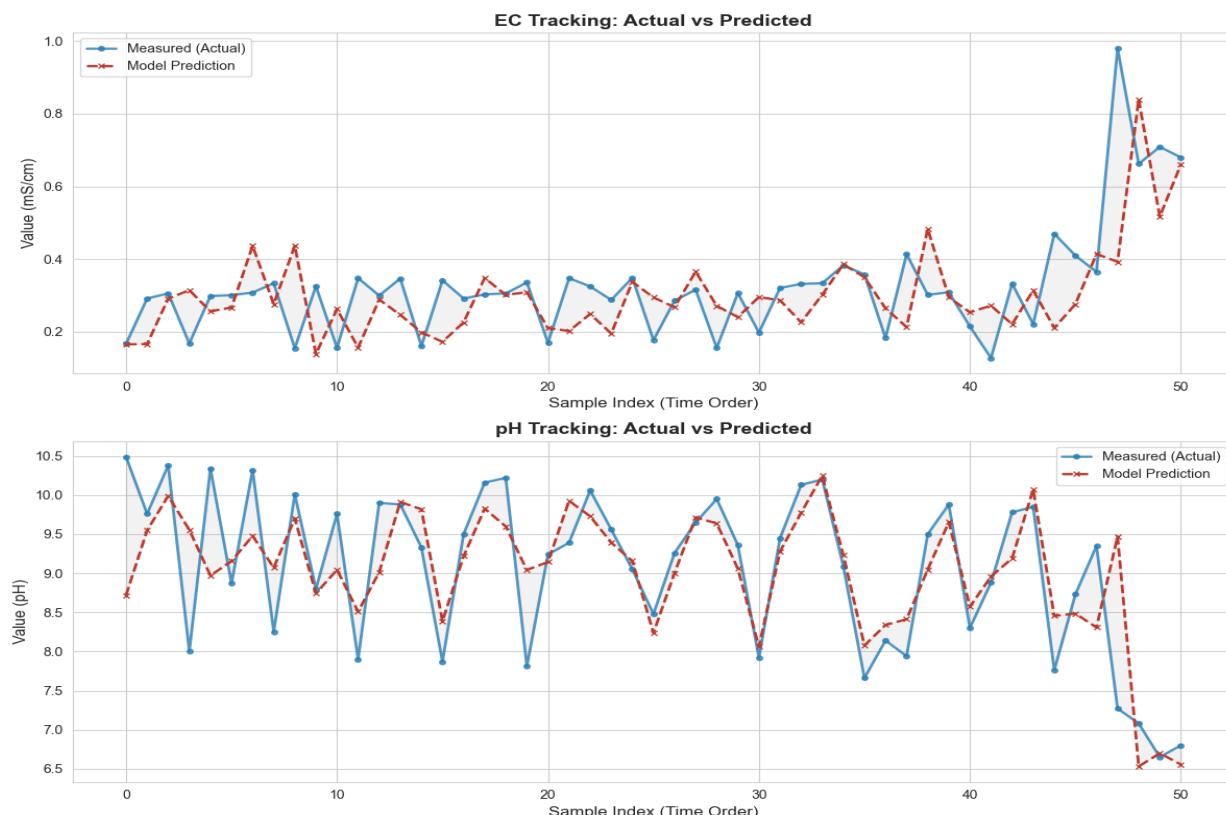
- משמעות: המודל משמש כ"תחנה מטאורולוגית וירטואלית" אמינה מאוד עבור המודל הבא בשירות, יוכל לחזות בצורה אמינה את האקלים הפנימי בחמאמה לטווח של 24 שעות קדימה.

#### ב. מודל בית השורשים:

כאן התוצאות משקפות את המרכיבות הביולוגית של המערכת:

- חיזוי H<sub>c</sub>: המודל הציג יציבות מרשים בזיהוי המגמה, עם שגיאה יחסית נמוכה של כ-5% (MAE 0.48) זהו הישג שימושי בהינתן שה- H<sub>c</sub> מושפע מתהליכים כימיים צעירים בקרקע.

- חיזוי C<sub>E</sub>: המודל הצליח לחזות את ערכי המוליכות בשגיאה אבסולוטית נמוכה של 0.10 mS/cm. עם זאת, בשל ערכי המלחות הנמוכים שנמדדו בתקופה זו, השגיאה היחסית לנורמה גבוהה (כ-34%), אתגר המפורט בהמשך.



Overall Next-Observation Performance (with Relative Error)

| Target Variable | MAE    | RMSE   | Total Rel. Error |
|-----------------|--------|--------|------------------|
| pH              | 0.4820 | 0.6635 | 5.46%            |
| EC (mS/cm)      | 0.0996 | 0.1393 | 33.95%           |

## 7.2. אתגרים ודרכי פתרון

במהלך הפיתוח נתקלנו בשלושה אתגרים מרכזים הנובעים מאופי הנתונים והמערכת. להלן הניתוח והפתרונות:

### 7.2.1. אתגר 1: "הкопסה השחורה" – העדר ניתוח רציף

- **הבעיה:** בבחינה הנחקרה לא קיים כלל **חישון לניטור רציף של EC** |-EC המשמעות היא שהמערכת נמצאת ב"עיוורון" מוחלט רוב הזמן. ישן נקודות אמת בודדות (דוגמאות מעבדה) המרחקות זו מזו בזמן (לעתים ימים שלמים), ואין שום דרך לדעת מה קרה לבית השורשים בתווך שביניהן – האם הייתה קפיצה מלאכות? האם ה-EC צנחה לאחר הדישון?
- **הפתרון:** פיתוח **"חישון וירטואלי"** טהור, המודל תוכנן כך שהוא אינו מסתמן על משוב (Feedback) רצוף מהblkיע, אלא בונה מחדש את "סיפור" בית השורשים אך ורק על סמך **הקלט** (כਮויות המים והדשן שנכנסו) וה**סיבבה** (מיקרו אקלים). זהה משימה מורכבת יותר מכיוון **חישון קיים**, שכן המודל נדרש לסייע את התהילה הכימי שלו.

### 7.2.2. אתגר 2: בעיית הדليلות

- **הבעיה:** לאחר אין חישנים, ה-Ground Truth שלנו מוגבל לכ-100 דגימות ידניות בלבד שנאספו לאורך מספר חודשים. אימון מודל למידת מכונה הדורש בדרך כלל Big Data על סט נתונים כה דליל יוצר סיכון גבוה ל-Overfitting (שינון הנתונים).
- **הפתרון:**

1. **הנדסת פיצרים אגרסיבית:** במקומם להזין למודל נתונים רגילים, יצרנו משתנים "היסטריים" (כגון: "ס'ך הדשן שהצטבר ב-48 השעות האחרונות" או "זמן שלף מאז ההשקייה الأخيرة"). משתמשים אלו מפיצים על חוסר הנתונים בכך שהם נתונים למודל הקשר (Context) רחב לכל דוגמה.
2. **שימוש באלגוריתם XGBoost:** נבחר אלגוריתם המבוסס על אנסמבל של עצים ומוגבל בעומקו, אשר הוכיח עמידות גבוהה בעבודה עם דגימות מעטות.

### 7.2.3. אתגר 3: ריגישות המודל בערכי קיצון נמכרים

- **הבעיה:** במהלך תקופת הניסוי נמדד בביית השורשים הרבה ערכי EC נמכרים במיוחד, בטוחה של  $\text{mS/cm}$  0.2-0.5 (ערכים האופייניים למצע שטוף כמעט לחלוין). בטוחה דינמי זה, אנו נתקלים בבעיה מתמטית ופיזיקלית כפולה:

  1. **הטיה סטטיסטית:** מבחינה מתמטית, שגיאת חיזוי אבסולוטית של  $0.1 \text{ mS/cm}$  נחשבת מצוינת. כאשר ה-EC הוא 2.5, מדובר בעיות של 4% בלבד. אולם כאשר ה-EC הוא 0.2, אותה טיטה בדיקת מתורגם לטעות יחסית של 50%. הדבר יוצר עייפות במידה העריכת המודל.
  2. **יחסאות לרעיש:** בערכים כה נמכרים, ה"אות" (השינוי האמתי ברכיב המלחים כתוצאה מדישון) הוא חלש מאד, וקשה להפרדה מ"רעש הרקע" (טעויות מדידה קטנות, השפעת טמפרטורה על המolicות, או שאריות מלחים במים). המודל מתקשה להזות דפוסים ברורים כאשר השינויים הם בסדר גודל מיקרוסקופיים.

המשמעות האגראונומית: חשוב לציין כי מבחינה מעשית ההבדל בין EC של 0.2 ל-0.3 הוא זניח עבור הצמיחה – שניהם מצבים על חומר דשן ודורשים אותה פעולה מתקנת. לכן, הטעות הגבואה באחזois אינה מעידה בהכרח על כישלון המודל ככל תומך החלטה.

- **הפתרון:** טרנספורמציה לוגריתמית של המטרה(Log-Target Transformation) – משתנים כימיים כמו ריכוזים (EC) נוטים להתפלג בצורה שאינה נורמלית, והשונות (Variance) של השגיאה גדלה ככל שהערך גדול. במקום לאמן את המודל לחזות את הערך הגלומי ( $y$ ) נמן אותו לחזות את הלוגריתם של הערך ( $\log(y)$ ). טרנספורמציה זו "מודחת" את הטווחים הנמכרים ומקווצת את הגבויים. כאשר ממירם את התחזית צרצה על ידי שימוש ב- $\text{Exponent}$  המודל נוטה באופן טבעי להימנע מחיצוי ערכים שליליים או אפסיים לא היגיוניים, ומשפר את הדיקוק היחסיבי בערכים הנמכרים.

#### 7.2.4. אתגר 4: אי-סידירות זמנית של הדגימות

- **הבעיה:** בגיןו נתונים אקלים המגיעים בקצב קבוע (כל 10 דקות), דגימות המעבדה נלקחות באופן לא סדר – לעיתים הפרש הוא כמה שעות, לעיתים יומ אחד, ולעיתים 4-5 ימים (סופי שבוע/חגים). מצב זה מקשה על שימוש באלגוריתמים קלאסיים של סדרות עיתיות כגון ARIMA או LSTM המניחים צעד זמן קבועים ( $t, t+1, t+2, \dots$ ). המודל נדרש ללמוד לחזות את המצב הבא, בלי לדעת מראש כמה זמן יעבור עד אליו.
- **הפתרון:** שימוש בפיזר של "זמן יחס" הינו למודל משטנה מפורש המציג את מספר השעות שעברו מאז הדגימה האחרונה, כך שהוא למד לשקלל את "דעיכת הזמן" – ככל שעבר יותר זמן, עליו להסתמך פחות על המצב ההתחלתי ויתר על האינטגרציה של ההשקיה והאקלים בתווך.

#### 7.2.5. אפקט הזיכרון ותלות במצב ההתחלתי

- **הבעיה:** תגובת המצע לדישון אינה ליניארית ואנייה אחדית, אלא תלואה בהיסטוריה ובמצבי הנוכחי שלו.
- 1. **אפקט הזיכרון:** השקיה בודדת במקרים מותקים לא מורידה את EC-מיד, אלא מבצעת שטיפה הדרגתית.
- 2. **תלות במצב (State Dependency):** אותה מנת דשן בדיקון תוביל לתוצאה כימית שונה לחולוטין אם היא ניתנת למצוע שכביר רווי במחלכים וחומצוי, לעומת מצוע שטופ ובסיסי. התעלמות מה המצב ההתחלתי גורמת לשגיאות חיזוי גסות.
- 3. **אפקט המהילה:** השפעת הדשן נגזרת מהרכיב הסופי בתמיisha, ולא רק מהכמויות המוחלטות (גרמיים). השקיה בנוף מים גדול "mdlalt" את הדשן וממתנת את השפעתו על EC.
- **הפתרון:** הנדסת פיצרים מרובדת(Multi-Layer Feature Engineering):



1. **חלונות צבירה (Accumulation Windows)**: ייצור משתנים מסכמים לחלונות זמן שונים (8, 16, 24 שעות) כדי לאפשר למודל לזהות דפוסים של הצלבות מליחים או מחזורי שטיפה.

2. **יחס ריכוז (Concentration Ratios)**: במקומ להשתמש רק בכמות מוחלטת של דשן (mg), יצרנו פיצרים המחשבים את היחס בין כמות הדשן לנפח המים באותה השקיה. משתנים אלו מייצגים את הריכוז בפועל (wkcm) ואת אפקט המהילה, ומשיעים למודל לזהות את הפיזיקה האמיתית של תמיית הקרקע.

### **7.2.6. אתגר 6: אי-ריציפות בנתוני האקלים**

• **הבעיה:** מסד הנתונים של דגימות המעבדה (H/C) החל להיבנות עוד לפני שהותקנה בחומרה מערכת הניטור הרציפה (Data Logger) נוצר מצב שבו עבר חלק נכבד מדגימות הקרקע המוקדמות, לא היה קיים תיעוד מקביל של האקלים הפנימי (טמפרטורה/לחות), מה שמנע את האפשרות להשתמש בהן לאימון המודל.

• **הפתרון:** שחזור נתונים לאחר. במקומות לווטר על הדגימות היקרות, פיתחנו מודל רגסיה "יעודי מבוסס XGBoost" שאומן על התקופה המאוחרת יותר, בה היו נתונים מלאים. המודל למד את המתאם הגבוי שבין נתוני התchanנה המטאורולוגית החיצונית (בית דגן) לבין האקלים בתוך החממה. לאחר שהמודל הגיע לדיקון גובה, ( $R^2 > 0.95$ ) השתמשנו בו כדי לשחזר ("לחזרות לאחר") את תנאי האקלים הפנימיים עבור התקופה החסרה. פועלה זו אפשרה לנו להציג عشرות דגימות מעבדה ולהגדיל את סט האימון האפקטיבי בכ-30%.

## 8. תוכנית עבודה ולוח זמנים

תוכנית העבודה להמשך הממחקר מתמקדת במאזע העיקרי: **מבחן הדיק של מודל ה- EC והח'ק.** תהליך זה יבוצע לצורך איטרטיבית (ניסיונית וטעה מובנה) של הוסף משתנים, סינון רעים וכיוון האלגוריתם. במקביל, יוצאו ניתוחים משלימים ("מעטפת") כדי להדגים את יציבות המערכת וליצור תוצרים ויזואליים לדוח הסופי.

### **8.1. שלב 1: הנדסת פיצרים ושיפור המודל (המאזע העיקרי)**

המטרה: הורדת שגיאת החיזוי (MAE) באופן עקבי על ידי טיב הנתונים הנוכחיים. זהו לב העבודה בחודשים הקרובים.

- **משימה 1.1 - בחינת משתני מגמה (Trend Features)**

- יצירת משתנים חדשים המציגים את קצב השינוי (נגזרת) של ה- EC וה- H<sub>c</sub>.
- בדגימות האחרונות.
- בדיקה האם הוסף מיידע על "כיוון המגמה" (עליה/ירידה) משפרת את החיזוי יותר מאשר הערך האבסולוטי בלבד.

- **משימה 1.2 - ניסוי חלונות זמן (Time Windows Optimization)**

- ביצוע סדרת ניסויים כדי למצוא את "חלון הזיכרון" האופטימלי של המוצע: האם המודל מדייך יותר כשהוא מסתכל על צבירת דשן של 12 שעות? 24 שעות? 48 שעות?
- אופטימיזציה של משתני השהיה – (Lag Features) כמה זמן לוקח להשקייה להשפיע על המדייה?

- **משימה 1.3 - יחס גומלי (Interaction Features)**

- בדיקת פיצרים מורכבים המשלבים מספר גורמים, למשל: יחס דשן-למים (רכיב בפועל), או מכפלת קרינה בכיסוי נוף (צריכה פוטנציאלית).

### **8.2. שלב 2: אופטימיזציה של האלגוריתם**

המטרה: סחיטת הביצועים המקוריים מהמודל הקיים לאחר שטייבנו את הנתונים.

- **משימה 2.1 - סלקציית פיצרים (Feature Selection)**

- שימוש בשיטות כמו (RFE) Recursive Feature Elimination כדי לזרוק משתנים "רעים" שלא תורמים לדיק ורך מבלבים את המודל. המטרה: להישאר עם סט מזוקק של 10-15 משתנים משפיעים.

- **משימה 2.2 - יכול היפר-פרמטרים (Grid Search)**

- הרצת סריקה ממוחשבת למציאת ההגדרות הטובות ביותר למודל: עומק העץ המקסימלי, מספר העצים ביער, וכמוות הדגימות המינימלית לעלה. פועלה זו נועדה למנוע Overfitting ולשפר את הכללה.



### 3.8. שלב 3: ניתוח מעטפת וסימולציות (הכללה)

המטרה: יצירת נפח עבודה נוספת נוספת המאפשרת מעבר למספר היבש של הדיק.

- **משימה 3.1 - ניתוח רגישות(Sensitivity Analysis)**

- מיפוי הגורמים המשפיעים: יצירת גרפים המראים אילו פרמטרים היכי משפיעים על EC (האם זה המים? הטמפרטורה? הדשן?).

- **משימה 3.2 - בדיקת תרחישי קיצון(Stress Testing)**

- הרצת המודול על נתונים סינטטיים קיצוניים (למשל: "מה קורה לחיזוי אם יש שרב קיצוני וההשקייה מפסיקת?"). זה יאפשר לנו להראות שהמודול מתנהג בצורה "הגיונית" גם במצבים שלא קרו בפועל.

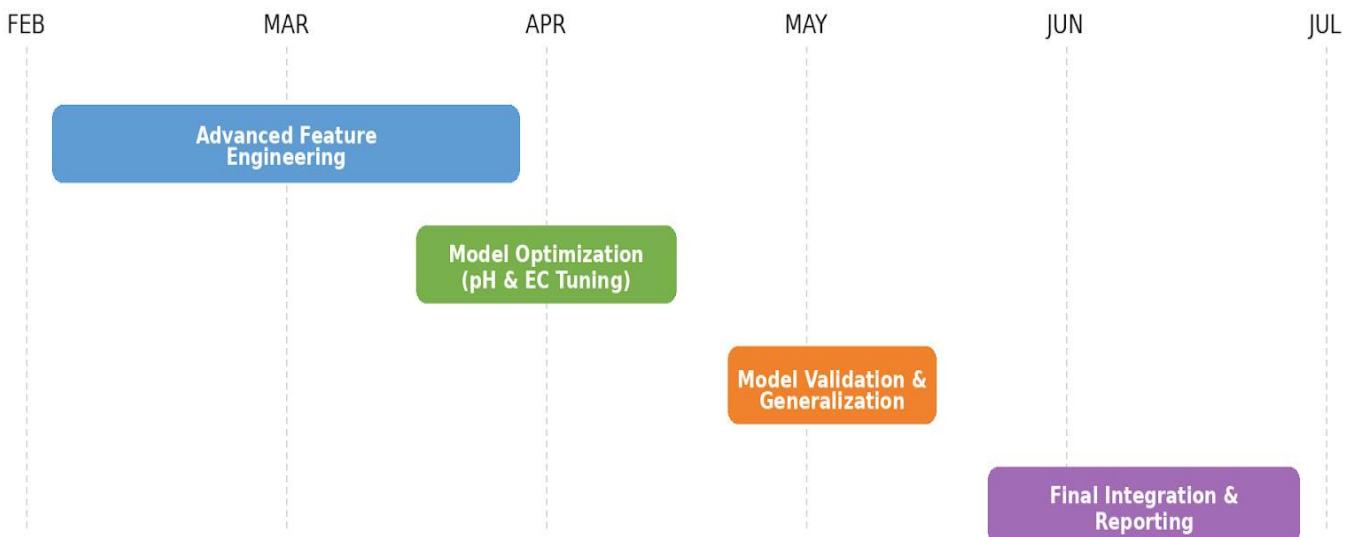
- **משימה 3.3 - השוואת מודלים נאיביים:**

- הוכחת עליונות המודול על ידי השוואת לחיזוי טיפש (כמו "הערך מחר יהיה זהה לערך היום"). זה יסייע נפח סטטיסטי לעבודה.

### 4. שלב 4: סיכון ואינטגרציה

- **משימה 4.1:** כתיבת ספר הפרויקט, הפקת גרפים מסכימים וממצאות.

- **משימה 2.4:** פיתוח Dashboard בסיסי להציג חיזוי המודול.





**9.ביבליוגרפיה**

- Moon, T., Ahn, T. I., & Son, J. E. (2018).** Forecasting Root-Zone Electrical Conductivity of Nutrient Solutions in Closed-Loop Soilless Cultures via a Recurrent Neural Network Using Environmental and Cultivation Information. *Frontiers in Plant Science*, 9, 859.
- Lü, X., Nurmemet, I., Xiao, S., Zhao, J., Yu, X., Aili, Y., & Li, S. (2025).** .2 Spatial-temporal simulation and prediction of root zone soil moisture based on Hydrus-1D and CNN-LSTM-attention models in Yutian Oasis, southern Xinjiang, China. *Pedosphere*, 35(5), 846-857.