



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 בהנחה יתם של ד"ר אבניר פריאל וד"ר עלaea ג'מאל

תוכן עניינים:

1.	תקציר	1.1
4	מטרת העבודה.....	1.1.1
4	הערך היישומי.....	1.1.2
4	שיטת המחקר.....	1.1.3
5	תוצאות ראשוניות.....	1.1.4
2. מבוא		
6	רקע כללי: האתגר בברית אקלים ובית שורשים.....	2.1
6	הפער בין הרצוי למצוי בניתור.....	2.2
7	הבסיס הפיזיקלי-ביולוגי.....	2.3
8	למה למידת מכונה?.....	2.4
8	הפתרון המוצע- חישון וירטואלי.....	2.5
8	מטרת הפרויקט.....	2.6
3. מושגים בסיסיים		
9	מושגים אגרונומיים.....	3.1
9	מושגים בלמידת מכונה.....	3.2
4. סקירה ספרות		
5. הגדרת הבעיה		
11	ניסוח הבעיה.....	5.1
11	שאלת המחקר.....	5.2
11	השערת המחקר.....	5.3
6. גישת הפתרון		
12	בסיס הנתונים.....	6.1
12	נתונים רציפים.....	6.1.1
12	נתונים בדידים.....	6.1.2
13	השלמת נתונים היסטוריים.....	6.2
14	עיבוד מקדים וונכרון זמנים.....	6.3
14	סביבה הפיתוח.....	6.4
14	ארכיטקטורת המערכת.....	6.5
14	מודל מיקרו-אקלים.....	6.5.1
14	מודל בית השורשים.....	6.5.2
15	אסטרטגיית איחוד הנתונים.....	6.6
16	הנדסת פיצרים.....	6.7
16	הנדסת פיצרים.....	6.8
7. תוצאות ראשוניות ואתגרים		
17	תוצאות המודלים.....	7.1
17	מודל המיקרו-אקלים.....	7.1.1
18	מודל בית השורשים.....	7.1.2
19	אתגרים ודרכי פתרון.....	7.2
19	היעדר ניטור רציף.....	7.2.1
19	בעיית הדילולות.....	7.2.2
19	رجישות לערכי EC במוכים.....	7.2.3

20.....	7.2.4. א' סדרות زمنית של הדגימות
20.....	7.2.5. תלות במצב התחלתי ואפקט הזיכרון.....
21.....	7.2.6. א' רציפות נתונים אקלים.....
	8. תוכנית עבודה ולוח זמנים
22.....	8.1. הנדסת פיצרים ושיפור המודל.....
22.....	8.2. אופטימיזציה של האלגוריתם.....
23.....	8.3. ניתוח מעתפת וסימולציות.....
23.....	8.4. סיכום וכתיבה.....

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



1. תקציר:

1.1. מטרת העבודה:

מטרת פרויקט זה היא לפתח מודל חיזוי דינמי מבוסס למידת מכונה (**Machine Learning**) ומטרת הפרויקט מותאמת ביצירת המערכת להערכת המוליכות החשמלית (EC) והחומריות (H_c) העתידיים, בהינתן נקודת מזאכידואה (מדידה הנוכחיית).

המודל משקל שלושה וקטוריים עיקריים כדי לבצע את החיזוי:

1. **מצב התחלתי:** נתוני EC וה-H_c שנמדדו בנקודת הזמן הנוכחיית.
2. **תנאי סביבה:** תחזית מיקרו-אקלים בתוך החממה (קרינה, טמפרטורה, לחות והתאדות).
3. **משק גידול:** פרוטוקול ההשקייה והדישון המתוכנן (כמויות מים, הרכבי דשן ומועד השקייה).

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" – דיזמי מגמות של המלח או שינוי חומריות מראש, והתאמת משטר 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 משטר 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 ומדדיות של EC ו-H_c 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-H₂O 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 H₂O:** 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 ריגושות המודל לשינויים עדינים ברכיביו החדש בהמשך המחקר.
- יציבות בזמן:** בנייתו Walk-Forward המודל שמר 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 ב芝יה מגמות EC יציבות, בעוד את האתגר המרכזי להמשך הוא דיקון 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. במגוון טכנולוגיות, המבוססות לרבות על מעעים מנוטקים, השיטה על תנאי בית השורשים (Rootzone) היא המפתח להשתת מטרות אלו. בנגוד לגידולי שדה בקרקע טבעי, למצע המנותק יש באפר קיבול (Buffering Capacity) נמוך, משמעות הדבר היא שכל 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 לחסור הדזנה ופגעה בקצב הצימוח, בעוד שערך גבוה מדי ייצור לחץ אוסmoti שיקשה על הצמח לקולט מים, ואף עלול לגרום לצריכות והמתתקת יבול מוקדמת.
- **חומציות (Hc):** ממד הקובע את המסודות והזמינים 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 למוצע בנייטור

על אף החשיבות הקritisית של מדרדים אלו, קיבלת תמונה מצב אמין ורציפה של בית השורשים מהוות כוונון "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 בתדרירות נמוכה (אחת ליום או לפחות ימים). התוצאה היא "נקודות עיוורות" בזמן; יתכן שבין שתי דגימות תקיןות התרחש אירוע קיצוני של המלחה (Spike) שగרם לנזק בלתי הפיך, אך לא תועד. כמו כן, שיטה זו היא ריאקטיבית (תגובהית) – המידע מתתקבל לאחר מעשה.

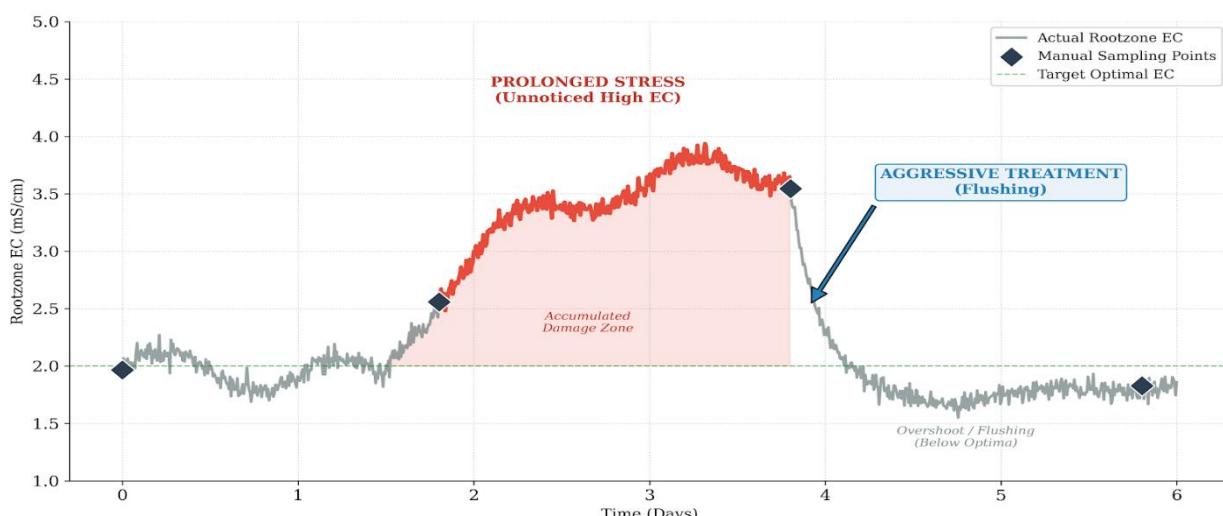


Figure 1: Current Workflow - Manual Sampling with Late Detection and Reactive Correction



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 עצית בודד, קיימת שונות מרחבית גבוהה ברטיבות ובלתי-הומוגניות. חישן הממוקם ליד הטרטיפת יראה ערך שונה של חלוטין מאשר הממוקם בתחום הקרקע.
- עלות: פרישה רחבה של חיישנים אינטגרליים 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:

- ביום של קרינה גבוהה: הצמח מבצע דיות (Transpiration) מוגברת ו"שותה" בעיקר מים נקיים, מה שמותיר את המלחים במצב וגורם לעלייה EC מהירה.
- ביום מעונני: קצב הדיות יורד, וצריכת המים והדשן מתאוזנת. המודל שלנו נדרש ללמוד את הדרושים הללו (הקשר הלא-lienair בין קרינה לצריכת מים) ולהזות את הנסיבות המלחים שנובעת מהם.

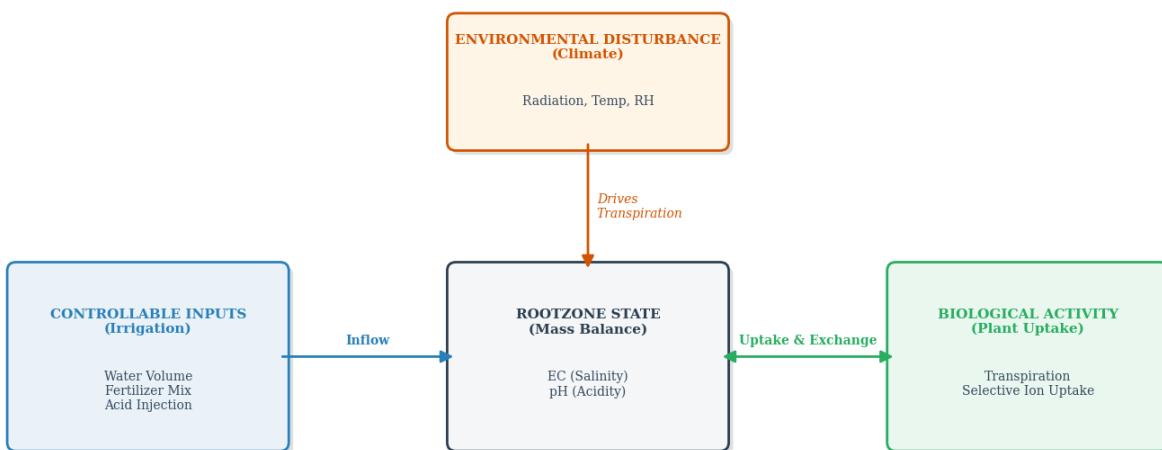


Figure 2: The Physical-Biological Basis - Factors influencing rootzone variables (EC & pH) resulting from the balance between Irrigation, Climate, and Plant Activity.



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 את התחנוגות האפקטיבית של המערכת לשירות מתוך הנתונים ההיסטוריים, ובכך מאפשר דיקוגובה בעלות חשיבות נמוכה.

2.5. הפתרון המוצע: "חישן וירטואלי" מבוסס נתונים

פרויקט זה מציע גישה חדשה לבניית המכונה על תפיסת ה-Soft Sensing ("חישן תוכנה"). ההנחה היא שקיים קשר פיזיקלי וכימי הדוק בין תנאי הסביבה החיצוניים (מזג האוויר), פעולות המגדל (השקייה ודישון) והtagובה 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 בחישני 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 ליישום במערכות תומכות החלטה.



3. מושגים בסיסיים

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

1. מושגים אגרונומיים

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

2. מושגים בלמידת מכונה

- **מודל רגסיבי (Regression Model)**: מודל שנועד לחזות ערך מסווני רציף (בניגוד לשיווג לקטגוריות). בפרויקט זה אנו חוזים בערכים רציפים של משתני מיקרו אקלים, EC ו-H_c.
- **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 למיניהם מתקדם המבוסס על שיטת Gradient Boosting (שייפור הדרגתית של עצי ההחלטה). האלגוריתם מתאפיין בשיטת גידלת עצ' ייחודית (Leaf-wise) המאפשרת מהירות אימון גבוהה במיוחד, ניצול יעיל של זיכרון וධוק רב, והוא נחשב לאידיאלי לעיבוד סדרות עיתיות ונתונים בהיקף רחב. שימש בפרויקט לחיזוי ערכי מיקרו אקלים עתידיים בתוך החממה על בסיס האקלים הנוכחי.
- **Walk-Forward Validation**: שיטת אימון למודלים של סדרות עיתיות (Time Series) במקומות חלוקה אקראיית ל"אימון" ו"בדיקה", אנו מאמנים את המודל על נתונים בעבר עד נקודה מסוימת, בודקים על העתיד הקרוב, ואז "chezים את החלון" קדימה. שיטה זו מדמה במדויק כיצד המערכת תפעל במצבאות, יומם אחר יומם.
- **הנדסת פיצ'רים (Feature Engineering)**: תהליכי ייצור משתנים חדשים מתוך הנתונים הגולמיים כדי לשפר את יכולת הלמידה של המודל. דוגמאות מפרויקט זה: "זמן שעבר מאז ההשקה الأخيرة", "סךום הקrinaה המctrber ב-24 שעות אחרונות".



4. סקירת ספרות

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

אחד האתגרים המרכזיים בגידול בחממות הוא השילטה ברמת המליחות (EC) בבית השורשים. במחקרם של Moon et al. (2018) בחנו החוקרים את יכולת לחזות את ה- EC של תמייסת ההזנה במערכות הידרואפוניות סגורות (Closed-loop soilless cultures). החוקרים השתמשו ברשת ניירונים מסווג (Recurrent Neural Network) RNN ובפרט במודל LSTM הידוע בכולתו ללמידה סדרות עיתיות. המחבר הראה כי שילוב של **משתני אקלים** (קרינה, טמפרטורה ולחות יחסית) יחד עם **נתוני פעולה** (כמות השקיה וריכוז דשן במי השקיה), מאפשר לחזות בד"וק רב את ה- EC העתידי במציאות. המסקנה המרכזית ממחקר זה, הרלוונטי לעבודתנו, היא שקיים קשר הדוק ונitin למידול בין עומס האקלים החיצוני לבין הדינמיקה הכימית בבית השורשים, וכי היסטוריית הנתונים המשמשת כ**נתוני פעולה** (Time-lagged features) היא קריטית לביצוע המודל.

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

הספרות הקיימת מצביעה על היכולות גבוהה לשימוש בלמידה מכונה לניטור בית השורשים. בעוד Sh et al. Moon et al. (2018) התמקדו ב- EC ו- Lü et al. (2025) ברטיבות, פרויקט זה שואף לשלב גישות אלו ולפתח מודל אחד החוצה **הן את ה- EC והן את ה- H₂O** תוך שימוש באלגוריתמים של למידה מכונה מבוססי עצים המצביעים איזון בין דיקוק ליכולת פרשנות ומתאים לישום במערכות בקרה תפעוליות בחממות מסחריות.

ביבליוגרפיה

1. 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.
2. Lü, X., Nurmeket, I., Xiao, S., Zhao, J., Yu, X., Aili, Y., & Li, S. (2025). 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.



5. הגדרת הבעיה

5.1. 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 המליחות (EC) והחומריות (H_c) 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 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_c רק לאחר שכבר חרגו מהטוווח הרצוי והנ扎ק הפוטנציאלי לצמח כבר החל.

5.2. שאלת המחקר:

האם ניתן לפתח מודל למידת מכונה (Machine Learning) המסוגל לחזות בדיקות אגרונומיות את ערכי המוליכות החשמלית (EC) והחומריות (H_c) 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 המksamilmית שניתנת להשיג בחיזוי (نمzd בMAE/RMSE) והאם היא מספקת לקבלת עדן החלטה תפעולי?
2. **חשיבות המשתנים (Feature Importance):** אילו משתנים משפיעים בצורה חזקה ביותר על השינויים בביותר על השינויים בבית השורשים? (למשל: האם קירינה 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?

5.3. 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 הריכוזים 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.



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 נימש תמיד) כדי לחזות את המתרחש בתוך המצע.

6.1. מסד הנתונים

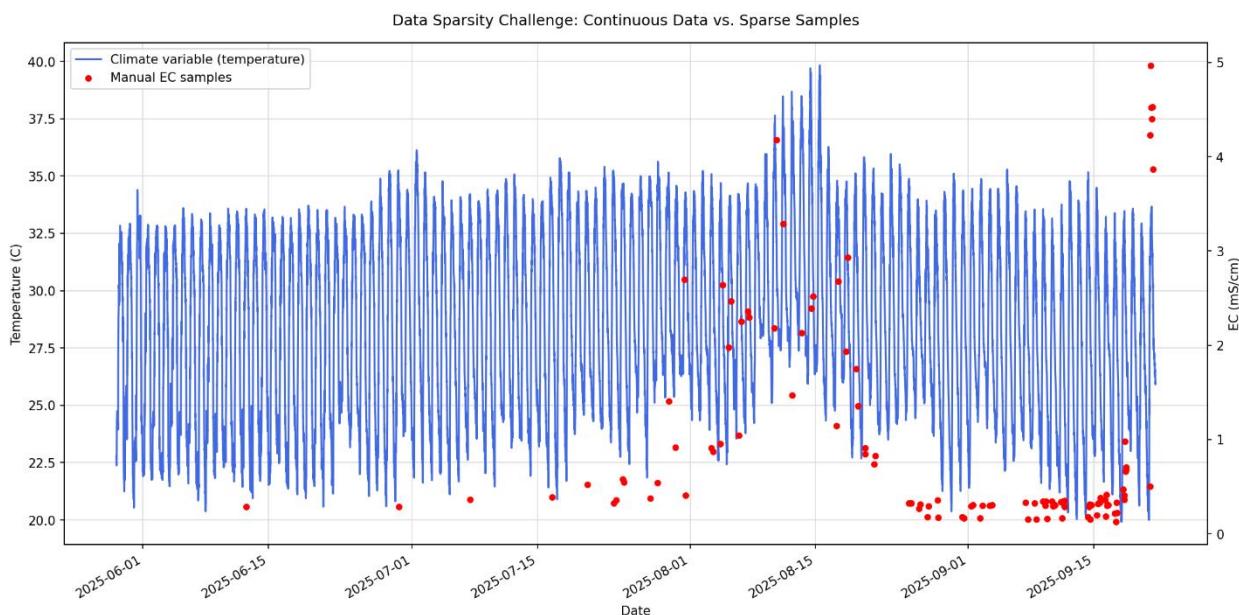
המחקר מתבסס על נתונים אגרו-מטאורולוגיים שנאספו בחמתת המחקר במהלך עונת הגידול 2025 (בחודשים מאי-ספטמבר). לצורך אימון ובדיקה המודלים, נבנה מסד נתונים מאוחד המשלב מקורות מידע הטרוגניות, שעברה תהליכי ניקוי וסנכרון. הנתונים נחלקים לשתי קטגוריות עיקריות, הנבדלות ברחולוציה הזמן שלהן:

1. נתונים רציפים (High-Frequency Data) :

- מקור: תחנה מטאורולוגית (בית דגן) ולוגר נתונים פנימי בחמתה.
- היקף: כ-16,000 רשומות זמן.
- רוחולוציה: דגימה כל 10 דקות באופן רציף.
- משתנים: קרינה גלובלית, טמפרטורה, לחות יחסית, מהירות רוח וחישוב 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) למשתנה המטרה (מעבדה) עומד על 1:200.



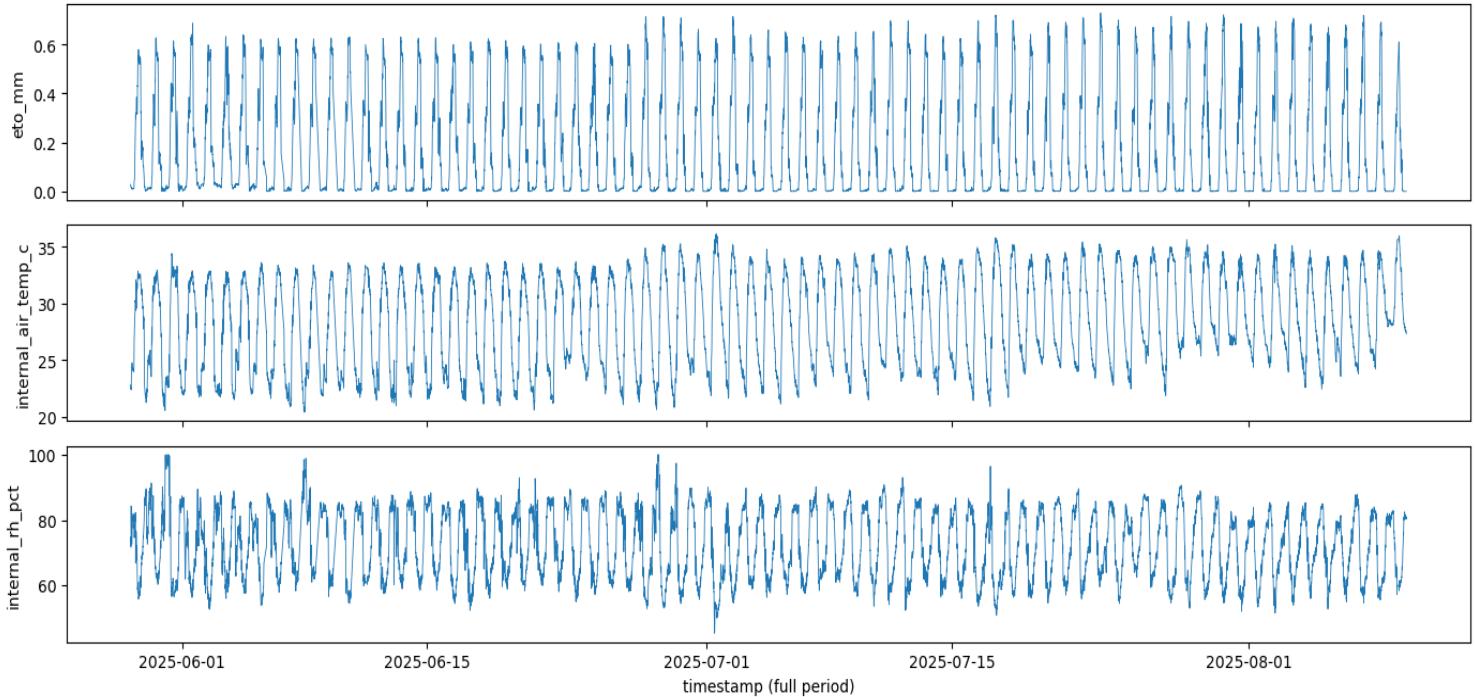
6.2. השלמת נתונים היסטוריים (Data Imputation using XGBoost):

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 (H₂O/EC) 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 "לאחור" כדי לשחזר בדיקוק רב את הטמפרטורה והלחות הפנימית בעבר התקופה שקדמה להתקנת החישון. פועלה זו אפשרה לנו להגדיל את סט האימון האפקטיבי בכ-30%.

Microclimate Model Performance Results

Target Variable	MAE	R ² Score (%)
Et0 (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 ערכים לא פיזיולוגיים מהחומר המקורי. כמו כן, בוצע סנכרון זמנים (Time Synchronization) כדי להציגם לכל דגימות מעבדה בוודת את היסטורית האקלים וההשקייה המדוקנית שהובילה אליה ("חולון השפעה").

6.4. סביבת הפיתוח

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

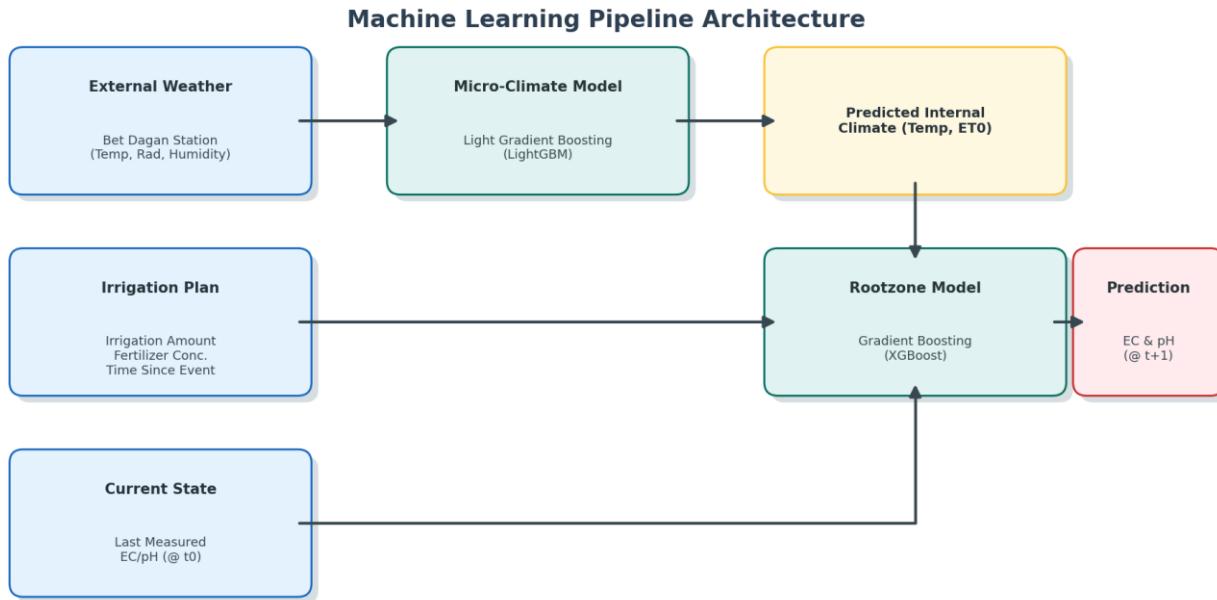
- **Pandas** לטיפול ועיבוד סדרות עיתיות.
- **Scikit-Learn** למימוש אלגוריתמי ML קלואסיים (Random Forest) חלוקת DATA-Train/Test Split ומטריקות הערכה.
- **Gradient Boosting** יעה לשימושה למידול המיקרו-אקלים ובית השורשים.
- **Matplotlib/Seaborn** לוויזואליזציה וניתוח נתונים. הפיתוח בוצע על גבי מחברות Jupyter 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 -ET0).
 - **הרצינול:** הצמח ובית השורשים מושפעים מהתנאים בתוך המבנה, השונים ממשמעות מתנים 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 ובਮירותו האימון הגבוהה שלו בהשוואה לאלגוריתמים מסורתיים. המודל אומן למדוד את פונקציית המעבר הלא-ליニアרית 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-H₂O בצעד הזמן הבא.

- אלגוריתם XGBoost Regressor זה נבחר בשל עמידתו לרעש (Noise Robustness) ויכולתו להתמודד היטב עם מערבי נתונים קטנים יחסית (-100 דגימות ידניות) מבלי להיכנס לבעיית התאמת-יתר(Overfitting).



6.6. אסטרטגיית איחוד הנתונים

אחד האתגרים המרכזיים בפרויקט היה השוני בrzולוציות הזמן של הנתונים:

- **נתונים "צפויים":** נתונים אקלים ולוגר נתונים (כל 10 דקות).
- **נתונים "دلילים":** דגימות מעבדה ידניות (אחת ליום או למספר ימים רוב הזמן).

הפתרון: יצירת Master Dataset בrzולוציה אחידה של 10 דקות.

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



(Feature Engineering) 6.7.

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

- **משתני צבירה (Cumulative Features)**: במקום להסתכל רק על הקרינה הרגעית, המודל מחשב את "סך הקרינה ב-24 השעות האחרונות" או "סך ההשקייה ב-8 השעות האחרונות".
משתנים אלו מייצגים את ה"זיכרון" של המערכת.
- **זמן מאז אירוע (Time-Since Events)**: משתנים המודדים כמה זמן עבר מזמן ההשקייה או הזמן האחרון. משתנים אלו קרייטיים לחיוה תכניות של עלייה וריכוז מלחים.
- **מצב הצמח (Canopy Cover)**: שילוב נתון כיסוי הנוף Caindy קציה לגודל הצמח וליכולת שלו לקלוט מים ודשן (Transpiration rate).

6.8. אסטרטגיית אימון ובדיקה(Walk-Forward Validation)

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

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 וזמןות הנתונים:

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



7. תוצאות ראשוניות ותגירים

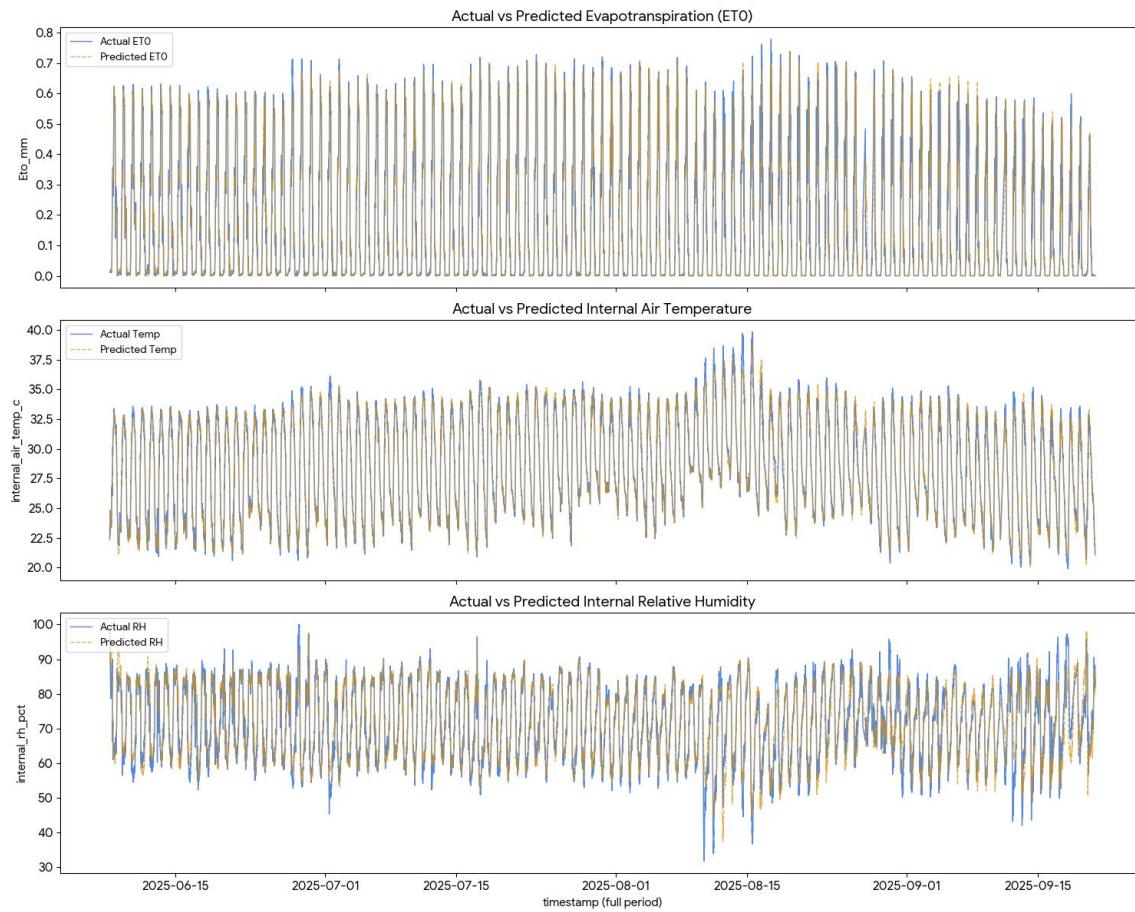
7.1. תוצאות המודלים

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:

מודל זה הציג ביצועים יוצאי דופן והוכיח כי ניתן לחזות את האקלים הפנימי בדיק רכיב על בסיס נתונים חיצוניים בלבד.

- טמפרטורה פנימית: התקבל מקדם מתאם R^2 של 0.97 עם שגיאה ממוצעת של פחות מ- 0.8° צלזיוס.
- לחות יחסית: הדיק עמד על R^2 של 0.87 עם שגיאה ממוצעת של 2.5% בלבד וטעות יחסית של 3.5%, נתון קритי לחיזוי התאדות.
- התאדות (ET0): התקבל דיק גבוה במיוחד במיחוד עם מקדם מתאם R^2 של 0.97, ושגיאה מוחלטת ממוצעת (MAE) של 0.016 מ"מ בלבד לכל מרוח זמן של 10 דקות. דיק זה הוא קритי שכן ET0 מראה את הבסיס לחישוב מנוגת ההשקייה.



Model Performance Results

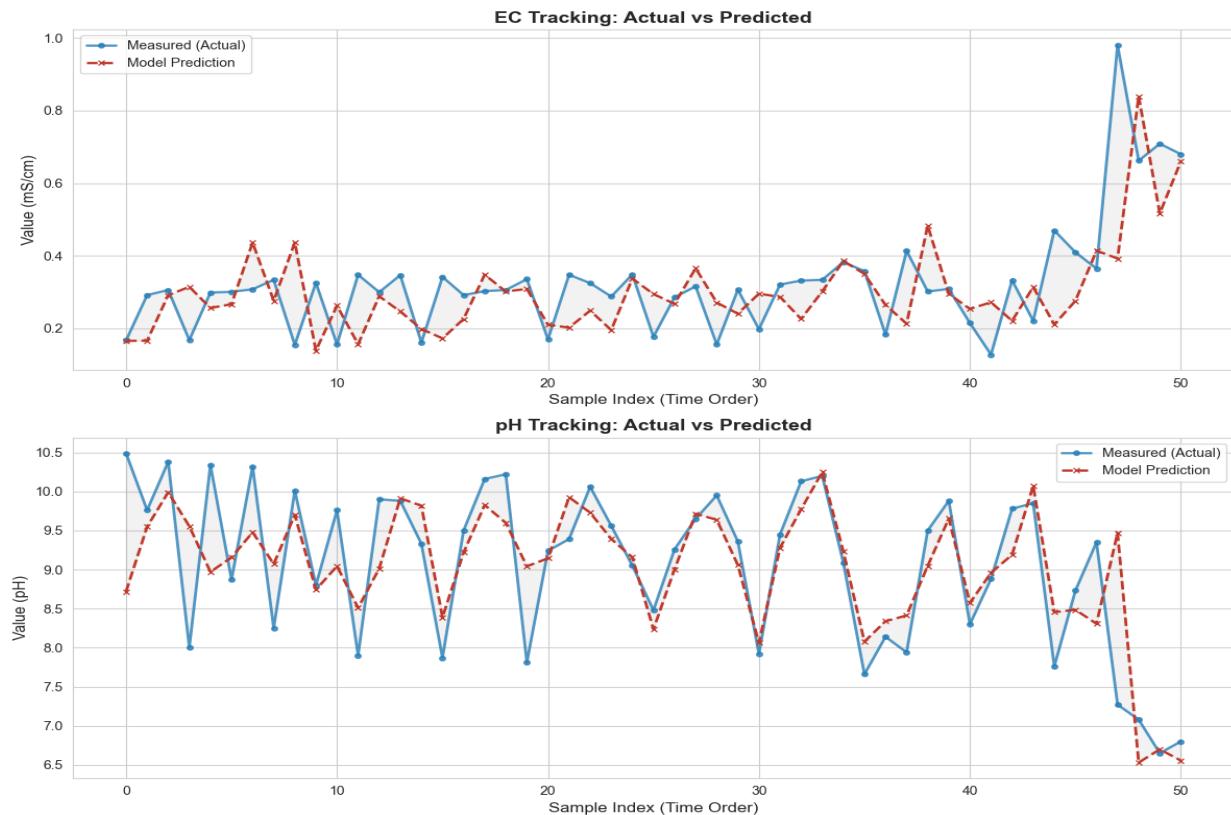
Target Variable	MAE	RMSE	R ² Score
Internal Air Temp (°C)	0.3863	0.5062	97.86%
Internal Radiation (W/m ²)	20.5592	39.1329	96.89%
Evapotranspiration (mm)	0.0158	0.0296	97.18%
Internal Relative Humidity (%)	2.5209	3.2580	86.77%

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

ב. מודל בית השורשים:

כאן התוצאות משלב את המרכיבות הביוולוגיות של המערכת:

- חיזוי EC: המודל הצליח לחשות את ערכי המוליכות בשגיאה אבסולוטית נמוכה של כ-5% (MAE 0.48) זהו הישג שימושי בהינתן שה-EC מושפע מתהליכים כימיים צעירים בפרקע.
- חיזוי pH: המודל הצליח לחשות את ערכי ה-pH הנמוכים שנמדדו בתקופה זו, השגיאה היחסית נותרה 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%

2. אתגרים ודרך פתרו

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

7.2.1. אתגר 1: "הkopsea השחורה" – העדר ניטור רציף

- **הבעיה:** במחמתה הנחקרה לא קיים כלל **חייבן לניטור רציף של EC** או **-Hק** המשמעות היא שהמערכת נמצאת ב"עיוורון" מוחלט רוב הזמן. ישן נקודות אמת בודדות (dagimot meubda) המרוחקות זו מזו בזמן (לעתים ימים שלמים), ואין שום דרך לדעת מה קרה לבית השורשים בתוך שביניהן – האם הייתה קפיצת מליחות? האם ה-**Hק** צנחה לאחר הדישון?
- **הפתרון:** פיתוח "חייבן וירטואלי" טהור, המודל תוכנן כך שהוא אינו מסתמך על מושב (Feedback) רציף מהפרקע, אלא בונה מחדש את "סיפור" בית השורשים אך ורק על סמרק הקלט (כਮויות המים והדשן שנכנסו) והסבירה (מיקרו אקלים). זהה משימה מורכבת יותר מכיוון **חייבן קיים**, שכן המודל נדרש לספק את התהילה הכימי שלו.

7.2.2. אתגר 2: בעיית הדיליות

- **הבעיה:** לאחר שאין חיישנים, 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 (שינוי הנתונים).

- **הפתרון:**

1. **הנדסת פיצ'רים אגרטיבית:** במקומם להזין למודל נתונים רגילים, יצרנו משתנים "היסטרויים" (כגון: "סך הדשן שהצטבר ב-48 שעות האחרונות" או "זמן שחלף מאז ההשקייה האחרונות"). משתנים אלו מפצים על חוסר הנתונים בכך שהם נתונים למודל הקשור (Context) רחב לכל דגימה.
2. **שימוש באלגוריתם XGBoost:** נבחר אלגוריתם המבוסס על אנסמבל של עצים ומוגבל בעומקו, אשר הוכח עמידות גבוהה בעבודה עם דגימות מעטות.

7.2.3. אתגר 3: ריגשות המודל בערכי **קייזן נמכים**

- **הבעיה:** במהלך תקופת הניסוי נמדדנו בבית השורשים הרבה EC נמכים במינוח, בטוווח של cm/mS/cm 0.2-0.5 (ערכים האופייניים למצו שטוף כמעט לחלוין). בטוווח DIN מי זה, אנו נתקלים בבעיה מתמטית ופייזיקלית כפולה:



1. הטיה סטטיסטית : מבחינה מתמטית, שגיאת חיזוי אבסולוטית של 0.1 mS/cm נחשבת מצינית. כאשר ה- EC הוא 2.5 , מדובר בטעות של 4% בלבד. לעומת זאת, כאשר ה- EC הוא 0.2 , אותה סטייה בדיקת מתורגמת לטעות יחסית של 50% . הדבר יוצר עיוות במדד הערצת המודל.

2. יחסאות לרשע: בערכים כה נמוכים, ה"אות" (השינוי האמייתי ברכיב המלחים כתוצאה מדישון) הוא חלש מאוד, וקשה להפרדה מ"רשע הרקע" (טעויות מדידה קטנות, השפעת טמפרטורה על המolicות, או שרירות מלחים במים). המודל מתבקש לזרות דפוסים ברורים כאשר השינויים הם בסדר גודל מיקרוסקופיים.
- המשמעות האגרונומית: חשוב לציין כי מבחינה מעשית הבדל בין EC של 0.2 ל- 0.3 הוא זניח לעומת הצמח – שניהם מצביעים על חוסר דשן ודורשים אותה פעולה מתבקשת. לכן, הטעות הגבוהה באחוזים אינה מעידה בהכרח על כישלון המודל ככל תומך החלטה.
- **פתרונות:** טרנספורמציה לוגריתמית של המטרה(Log-Target Transformation)
- משתנים כימיים כמו ריכוזים (EC) נוטים להתפלג בצורה שאינה נורמלית, והשונות (Variance) של השגיאה אדלה ככל שהערך גדול.

במוקם לאמן את המודל לחזות את הערך הגולמי (y) נאמן אותו לחזות את הלוגריתם של הערך ($\log(y)$). טרנספורמציה זו "זוויה" את התווחים הנמוכים ומכובצת את הגבוהים. כאשר ממירם את התחזית צרצה על ידי שימוש ב- Exponent- ARIMA המודל נוטה באופן טבעי להימנע מחיזוי ערכים שליליים או אפסיים לא הגיוניים, ומשפר את הדיק היחסיבי בערכים הנמוכים.

7.2.4. אתגר 4: אי-סידרות זמניות של הדגימות

- **הבעיה:** בגיןו אקלים המגיעים בקצב קבוע (כל 10 דקות), דגימות המעבדה נלקחות באופן לא סדיר - לעיתים הפרש הרוא כמה שעות, לעיתים יומ אחד, ולעיתים 5-4 ימים (סופ' שבוע/חגים). מצב זה מקשה על שימוש באלגוריתמים קלאסיים של סדרות עיתיות כגון ARIMA או LSTM המניחים צעד-זמן קבועים ($t, t+1, t+2$). המודל נדרש ללמוד לחזות את המצב הבא, בלי לדעת מראש כמה זמן יעבור עד אליו.
- **פתרונות:** שימוש בפיצ'ר של "זמן יחסי" הזרים למודל משטנה מפורש המציג את מסוף השעות שעברו מאז הדגימה האחרונה, כך שהוא למד לשקלל את "דעתת הזמן" – ככל שעבר יותר זמן, עליו להסתמך פחות על המצב ההתחלתי ויוטר על האינטגרציה של ההשכלה והאקלים בתווך.

7.2.5. אתגר 5: אפקט הזיכרון ותלות במצב ההתחלתי

- **הבעיה:** תגובה המצע לדישון אינה ליניארית ואנייה אחדשה, אלא תלולה בהיסטוריה ובמצב הנוכחי שלו.
- 1. **אפקט הזיכרון:** השקיה בודדת במים מתוקים לא מורידה את ה-EC-מיד, אלא מבצעת שטיפה הדרגתית.

2. **תלות במצב:** (State Dependency) אוטה מנת דשן בדיק תוביל לתוצאה כימית שונה לחילוtin אם היא ניתנת למצע שכבר רווי במלחים וחומצى, לעומת מצע שטוף ובסיסי. התוצאות מה המצב ההתחלתי גורמת לשגיאות חיזוי גסות.
 3. **אפקט המהילה:** השפעת הדשן נגרת מהרכיב הסופי בתמיסה, ולא רק מהכמאות המוחלטת (גרמיים). השקיה בנוף מים גדול "mdlalt" את הדשן וממתנת את השפעתו על ה-EC.
- **הפתרון:** הנדסת פיצ'רים 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):** במקומות להשתמש רק בכמויות מוחלטות של דשן (mg), יצרנו פיצ'רים המחשבים את היחס בין כמות הדשן לנוף המים באותה השקיה. משתנים אלו מייצגים את הריכוז בפועל (wkcm) ואת אפקט המהילה, ומשיעים למודל לחוק את הפיזיקה האמיתית של תמיית הקרחן.

7.2.6. אתגר 6: אי-ריציפות בנתוני האקלים

- **הבעיה:** מסד הנתונים של דגימות המעבדה (H/c/EC) החל להיבנות עוד לפני שהו התקנה בחממה מערכת הניטור הרציפה (Data Logger) נוצר מצב שבו עבר חלק נכבד מדגימות הקרחן המוקדמות, לא היה קיימם תיעוד מקביל של האקלים הפנימי (טמפרטורה/לחות), מה שמנע את האפשרות להשתמש בהן לאימון המודל.
- **הפתרון:** שחזור נתונים לאחר. במקומות לוותר על הדגימות היקרות, פיתחנו מודל רגרסיה "יעודי" מבוסס XGBoost שאומן על התקופה המאוחרת יותר, בה היו נתונים מלאים. המודל למד את המתאם הגבוי שבין נתונים התחנה המטאורולוגית החיצונית (בית דגן) לבין האקלים בתוך החממה. לאחר שהמודל הגיע לדיק גובה, ($R^2 > 0.95$) השתמשנו בו כדי לשחזר ("לחזרות לאחר") את תנאי האקלים הפנימיים עבור התקופה החסורה. 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_c בדגימות האחרונות.
- בדיקה האם הוסף מייד 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 שטיינו את הנתונים.

- **משימה 2.1 - סלקטיבית 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 15-10 משתנים משפיעים.

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

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



3.8. שלב 3: ניתוח מעטפת וסימולציות (העשרה הדוח)

המטרה: יצירת נפח עבודה נוספת נוספת המדגים את יכולות המערכת מעבר למספר היבש של הדיקט.

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

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

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

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

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

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

4.8. שלב 4: סיכום וכתיבה

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