

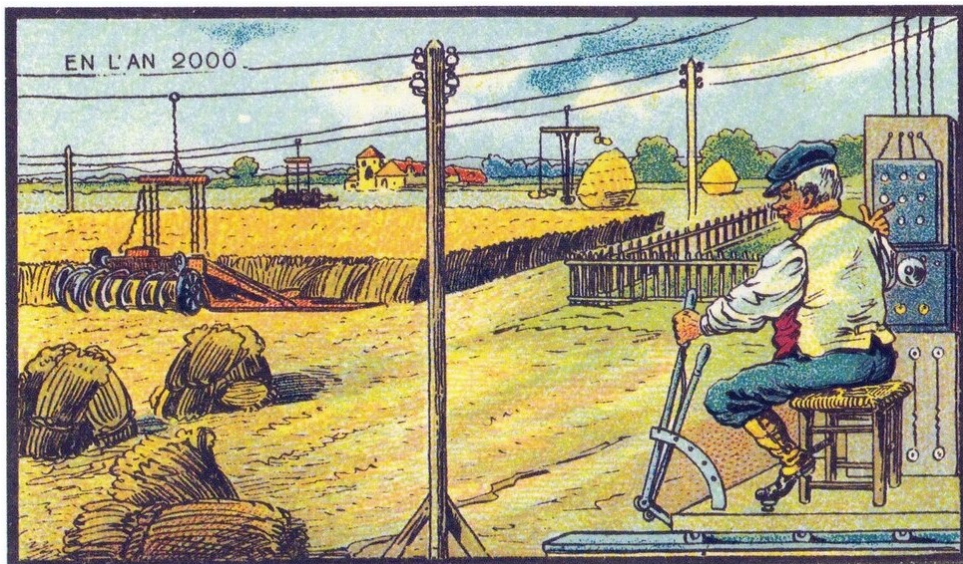


ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΓΕΩΠΟΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΓΕΩΠΟΝΙΑΣ ΦΥΤΙΚΗΣ ΠΑΡΑΓΩΓΗΣ ΚΑΙ ΑΓΡΟΤΙΚΟΥ
ΠΕΡΙΒΑΛΛΟΝΤΟΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**Εκτίμηση παραγωγής σκληρού σίτου μέσω τηλεπισκόπησης
και μηχανικής μάθησης**

Μαρία Μπέμπη



Επιβλέπων Καθηγητής: Κυπαρίσσης Άρης

Βόλος, Σεπτέμβριος 2022

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΓΕΩΠΟΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΓΕΩΠΟΝΙΑΣ ΦΥΤΙΚΗΣ ΠΑΡΑΓΩΓΗΣ ΚΑΙ ΑΓΡΟΤΙΚΟΥ
ΠΕΡΙΒΑΛΛΟΝΤΟΣ

Πτυχιακή Εργασία

**Εκτίμηση παραγωγής σκληρού σίτου μέσω τηλεπισκόπησης
και μηχανικής μάθησης**

**Assessment of durum wheat yield through remote sensing and
machine learning**

Μαρία Μπέμπη

Τριμελής Συμβουλευτική Επιτροπή:

Άρης Κυπαρίσσης (επιβλέπων) Αναπληρωτής Καθηγητής Οικοφυσιολογίας Φυτών, Τμήμα Γεωπονίας Φυτικής Παραγωγής και Αγροτικού Περιβάλλοντος, Πανεπιστήμιο Θεσσαλίας

Χρήστος Καβαλάρης (μέλος) Ε.Δι.Π., Διδάκτωρ Γεωργικής Μηχανολογίας, Τμήμα Γεωπονίας Φυτικής Παραγωγής και Αγροτικού Περιβάλλοντος, Πανεπιστήμιο Θεσσαλίας

Ευθυμία Λεβίζου (μέλος) Αναπληρώτρια Καθηγήτρια Φυσιολογίας Φυτών, Τμήμα Γεωπονίας Φυτικής Παραγωγής και Αγροτικού Περιβάλλοντος, Πανεπιστήμιο Θεσσαλίας

Ευχαριστίες

Θα ήθελα να ευχαριστήσω πρωτίστως τον κύριο Άρη Κυπαρίσση, καθηγητή στο εργαστήριο της οικοφυσιολογίας φυτών και επιβλέποντα της διπλωματικής μου, για την πολύτιμη βοήθεια και καθοδήγηση του καθ' όλη την διάρκεια των πειραμάτων (επιτυχών και μη), και γενικά για την άριστη συνεργασία και επικοινωνία μας.

Επίσης, ευχαριστώ τον κύριο Χρήστο Καβαλάρη, διδάκτορα της γεωργικής μηχανολογίας και την κυρία Έφη Λεβίζου, καθηγήτρια στο εργαστήριο φυσιολογίας φυτών, για τον χρόνο που διέθεσαν για την ανάγνωση αυτής της διπλωματικής και για την συμμετοχή τους στην συμβουλευτική επιτροπή.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου και τους φίλους μου για τις παντός είδους συμβουλές τους και την υποστήριξη τους.

Βεβαιώνω ότι είμαι συγγραφέας αυτής της πτυχιακής εργασίας, η οποία εκπονήθηκε σύμφωνα με τον Κανονισμό Εκπόνησης Πτυχιακής Εργασίας του ΤΓΦΠΑΠ.

Περιεχόμενα

| | |
|---|-----|
| Ευχαριστίες | iii |
| Περιεχόμενα | iv |
| Περίληψη | 1 |
| Abstract | 2 |
| 1. Εισαγωγή | 3 |
| 1.1. Σιτάρι | 3 |
| 1.1.1. Γενικά στοιχεία | 3 |
| 1.1.2. Στάδια ανάπτυξης – κλίμακα Zadoks..... | 5 |
| 1.2. Γεωργία Ακριβείας..... | 7 |
| 1.3. Τηλεπισκόπηση | 8 |
| 1.4. Πρόβλεψη της απόδοσης..... | 10 |
| 1.5. Συστήματα γεωγραφικών πληροφοριών (G.I.S.) | 11 |
| 1.6. Δορυφόροι Sentinel-2 | 12 |
| 1.7. Μηχανική Μάθηση (Machine Learning)..... | 14 |
| 1.7.1. Γενικά στοιχεία | 14 |
| 1.7.2. Βασικές αρχές μηχανικής μάθησης | 15 |
| 1.7.3. ML στον αγροτικό τομέα | 16 |
| 1.8. Αλγόριθμοι ML | 17 |
| 1.8.1. Random Forests (RF)..... | 18 |
| 1.8.2. K-nearest neighbors (KNN) | 19 |
| 1.8.3. Boosting regressions (BR)..... | 20 |
| 1.9. Σκοπός της εργασίας | 20 |
| 2. Υλικά και Μέθοδοι..... | 21 |
| 2.1. Περιοχή μελέτης..... | 21 |
| 2.2. Μέτρηση απόδοσης | 23 |

| | |
|---|----|
| 2.3. Δορυφορικά Δεδομένα | 23 |
| 2.4. Χρονοσειρές NDVI | 24 |
| 2.5. Μοντελοποίηση..... | 25 |
| 2.5.1. Μοντέλο με βάση τους δείκτες βλάστησης (VI-MLR)..... | 25 |
| 2.5.2. Μοντέλα με βάση την μηχανική μάθηση (ML)..... | 27 |
| 2.6. Στατιστική ανάλυση..... | 28 |
| 3. Αποτελέσματα | 29 |
| 4. Συζήτηση..... | 34 |
| 5. Συμπεράσματα | 39 |
| Βιβλιογραφία | 40 |
| Ελληνική | 40 |
| Ξενόγλωσση | 41 |
| Κατάλογος πινάκων..... | 48 |
| Πηγές εικόνων και σχημάτων | 49 |

Περίληψη

Στην παρούσα εργασία γίνεται η σύγκριση δύο προσεγγίσεων μοντελοποίησης για την εκτίμηση της απόδοσης σε καλλιέργεια σκληρού σίτου. Η πρώτη προσέγγιση αφορά την χρήση ενός μοντέλου που έχει αναπτυχθεί προηγουμένως και βασίζεται στους δείκτες βλάστησης με την χρήση της πολλαπλής γραμμικής παλινδρόμησης (VI – MLR). Η δεύτερη προσέγγιση, η οποία αποτελεί το πειραματικό μέρος της παρούσας εργασίας, αξιοποιεί την τεχνική της μηχανικής μάθησης (ML) με την χρήση τριών διαφορετικών αλγορίθμων για την δημιουργία μοντέλου. Στην προσέγγιση ML, αξιοποιούνται ως παράμετροι το σύνολο των φασματικών δεδομένων του δορυφόρου Sentinel-2, στην πρωτογενή μορφή τους, για ένα διάστημα τριών καλλιεργητικών περιόδων (2017 – 2020) που αφορά τους μήνες Οκτώβριο με Μάιο. Οι αλγόριθμοι που χρησιμοποιήθηκαν ήταν οι Random Forest (RF), K-Nearest Neighbors (KNN) και Boosting Regression (BR). Τα αποτελέσματα της μοντελοποίησης εξετάστηκαν με βάση δεδομένα απόδοσης από θεριζοαλωνιστική μηχανή, εξοπλισμένης με σύστημα χαρτογράφησης της απόδοσης.

Η προσέγγιση VI-MLR έδειξε μέτρια απόδοση με $R^2 = 0,532$ και $RMSE = 847 \text{ kg ha}^{-1}$. Όλες οι προσεγγίσεις μηχανικής μάθησης ενίσχυσαν την ακρίβεια του μοντέλου όταν χρησιμοποιήθηκαν όλες οι εικόνες κατά τις περιόδους ανάπτυξης, με καλύτερες τις RF και KNN ($R^2 > 0,91$, $RMSE < 360 \text{ kg ha}^{-1}$). Επιπλέον, η ακρίβεια RF και KNN παρέμεινε υψηλή ($R^2 > 0,87$, $RMSE < 455 \text{ kg ha}^{-1}$) όταν χρησιμοποιήθηκαν εικόνες από την έναρξη της καλλιεργητικής περιόδου μέχρι τον Μάρτιο, δηλαδή τρεις μήνες πριν από τη συγκομιδή. Έτσι υποδεικνύεται πως η μηχανική μάθηση σε συνδυασμό με τα τηλεπισκοπικά δεδομένα δύναται να χρησιμοποιηθεί ως εργαλείο έγκαιρης πρόβλεψης της απόδοσης του σκληρού σίτου και να παρέχει κρίσιμες πληροφορίες για εφαρμογές στην γεωργία ακριβείας.

Abstract

The aim of this thesis is to compare two modeling approaches of durum wheat yield estimation. The first approach consists of the use of a previously developed model based on vegetation indices using multiple linear regression (VI – MLR). The second approach constitutes this thesis' experimental part. It utilizes machine learning (ML) techniques with the use of three different algorithms for modelling. In the ML approach, remote sensing (RS) data, derived from the Sentinel-2 satellites, for a timespan of three growing periods (2017 – 2020) corresponding to the months of October to May, are used as independent parameters. The algorithms used were Random Forest (RF), K-Nearest Neighbors (KNN) and Boosting Regression (BR). Modeling results were examined against yield data recorded by a combine harvester equipped with a yield mapping system.

The VI-MLR approach showed a moderate performance with $R^2 = 0.532$ and $RMSE = 847 \text{ kg ha}^{-1}$. All machine learning approaches improved model accuracy when all images during growing seasons were used, with RF and KNN performing best ($R^2 > 0.91$, $RMSE < 360 \text{ kg ha}^{-1}$). In addition, RF and KNN accuracy remained high ($R^2 > 0.87$, $RMSE < 455 \text{ kg ha}^{-1}$) when images from the beginning of the growing period until March were used, i.e. three months before harvest. Thus, it is indicated that machine learning combined with remote sensing data can be used as a tool for early prediction of durum wheat yield and provide critical information for applications in precision agriculture.

Λέξεις – κλειδιά

Σκληρό σιτάρι, απόδοση, τηλεπισκόπηση, Sentinel-2, μηχανική μάθηση, αλγόριθμος, Random Forest, K-Nearest Neighbors, Boosting Regression.

1. Εισαγωγή

1.1. Σιτάρι

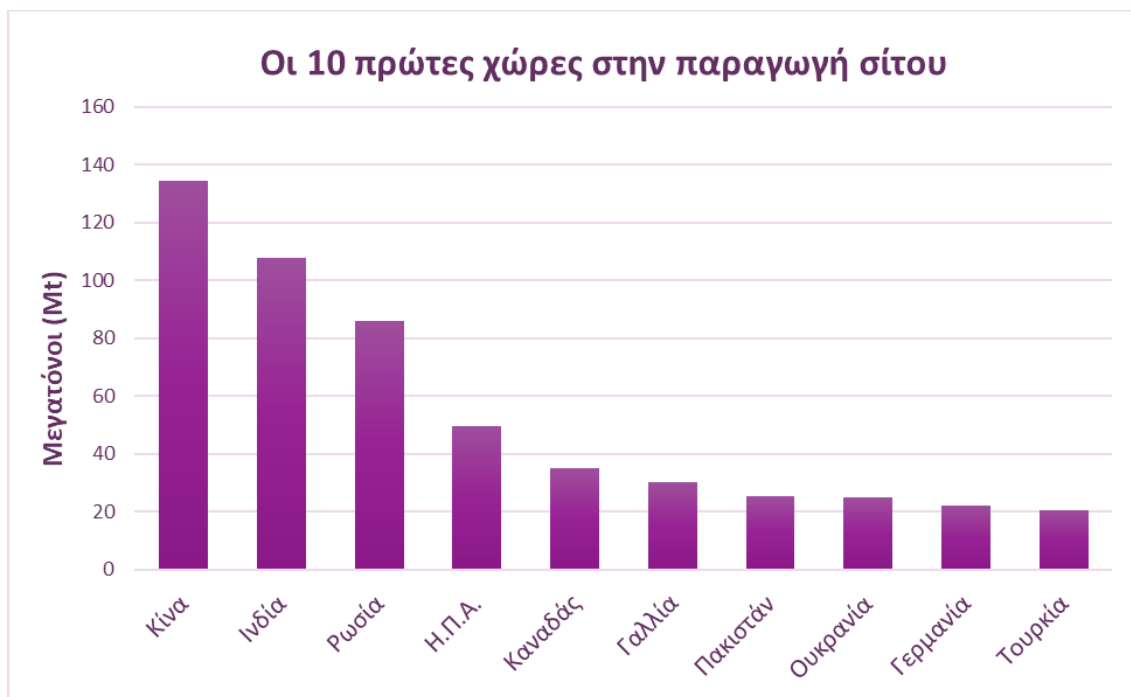
1.1.1. Γενικά στοιχεία

Το σιτάρι αποτελεί την πιο σημαντική και διαδεδομένη καλλιέργεια στον κόσμο. Μετά το κριθάρι, έχει τα ευρύτερα όρια εξάπλωσης σε παγκόσμια κλίμακα, καθώς το κριθάρι πλεονεκτεί στην ανθεκτικότητά του σε ψυχρότερα κλίματα. Πρόκειται για μια καλλιέργεια της Ευκράτου ζώνης, με όρια συνήθως μεταξύ 30 και 60° ΒΠ και 27 και 40° ΝΠ και από παραθαλάσσιες περιοχές, μέχρι και υψόμετρα 3000 μέτρων (Παπακώστα-Τασοπούλου, 2012). Στην εξάπλωση του έχουν συμβάλει η ευρεία προσαρμοστικότητα του, καθώς επίσης και ο μεγάλος αριθμός των βελτιωμένων ποικιλιών που έχουν δημιουργηθεί για να ανταποκρίνονται καλύτερα σε συγκεκριμένες κλιματικές και περιβαλλοντικές απαιτήσεις. Στις εύκρατες περιοχές με ήπιο χειμώνα και σε μεσογειακό κλίμα η καλλιέργεια του σιταριού είναι χειμερινή (δηλαδή η σπορά γίνεται το φθινόπωρο), ενώ στα ηπειρωτικά κλίματα και μεγάλα γεωγραφικά πλάτη είναι εαρινή καλλιέργεια (σπέρνεται την άνοιξη) (Μπιλάλης, 2019).

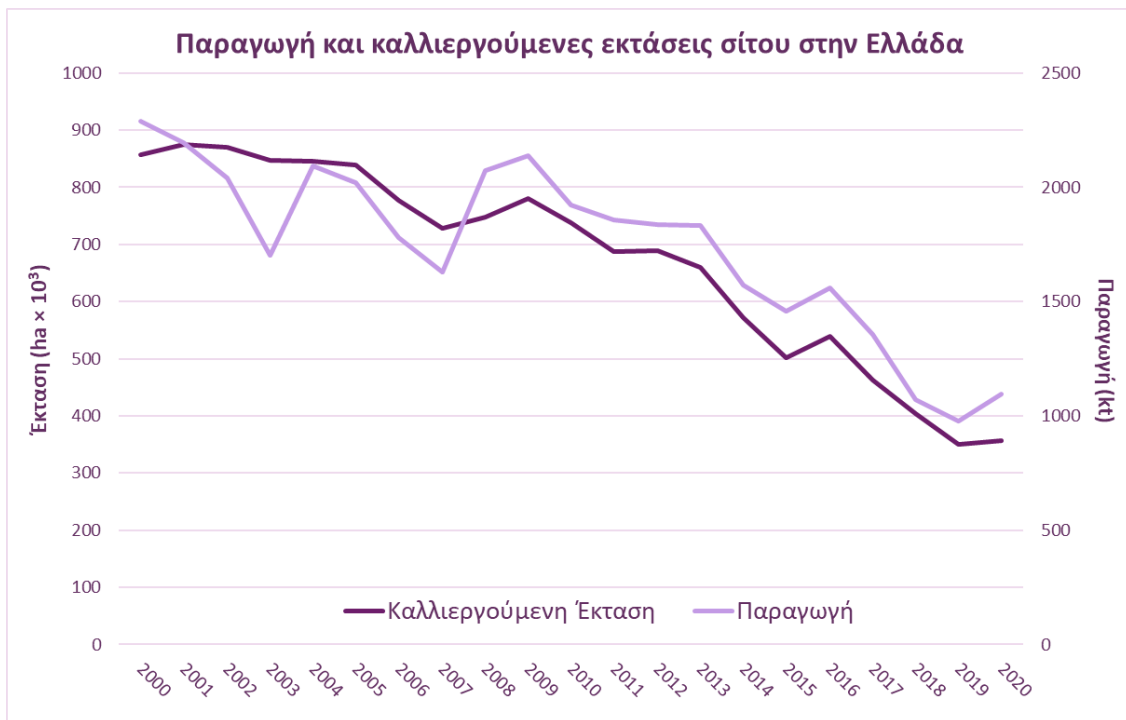
Το σιτάρι καλλιεργείται σε έκταση 2200 – 2300 εκατομμυρίων στρεμμάτων σε παγκόσμιο επίπεδο και η παραγωγή του ανέρχεται σε 550 – 700 εκατομμύρια τόνους (Παπακώστα-Τασοπούλου, 2012). Οι κυριότερες χώρες που παράγουν σιτάρι, σύμφωνα με τα δεδομένα του 2020 ήταν η Κίνα, η Ινδία, η Ρωσία, οι ΗΠΑ, ο Καναδάς, η Γαλλία, το Πακιστάν, η Ουκρανία, η Γερμανία και η Τουρκία, όπως φαίνεται στο Σχήμα 1. Όσον αφορά την Ευρωπαϊκή Ένωση, οι καλλιεργούμενες εκτάσεις σίτου υπολογίζονται στα 250 – 260 εκατομμύρια στρέμματα, ενώ η παραγωγή ανέρχεται 130 – 140 εκατομμύρια τόνους.

Συγκεκριμένα στην Ελλάδα, καλλιεργούνται ετησίως 1.5 εκατομμύρια στρέμματα μαλακό σιτάρι και 5.5 εκ. στρέμματα σκληρό σιτάρι (Μπιλάλης, 2019). Όπως φαίνεται στο Σχήμα 2, τόσο η καλλιεργούμενη έκταση, όσο και η παραγωγή σιταριού στην Ελλάδα εμφανίζει μια έντονα πτωτική τάση στο διάστημα της τελευταίας εικοσαετίας.

Πρόκειται για την βασική τροφή της ανθρωπότητας, καθώς παρέχει στους ανθρώπους το 18% της ημερήσιας πρόσληψης θερμίδων και το 20% της πρωτεΐνης τους. Υπάρχουν αρκετά είδη του γένους του σιταριού (*Triticum*). Το πιο διαδεδομένο είναι το κοινό ή μαλακό (*Triticum aestivum* L.), που καταλαμβάνει το 94% της συνολικής έκτασης του καλλιεργούμενου σίτου και χρησιμοποιείται κυρίως στην αρτοποιία. Το σκληρό σιτάρι (*Triticum turgidum* L. var. *durum*) καλλιεργείται σε περίπου 13 εκατομμύρια εκτάρια, εκ των οποίων το 60% περίπου βρίσκεται στην περιοχή της Μεσογείου. Το σκληρό σιτάρι χρησιμοποιείται κυρίως για την παρασκευή ζυμαρικών. Οι γενετικές διαφορές μεταξύ των δύο ειδών έγκεινται στον αριθμό των χρωμοσωμάτων, καθώς το σκληρό σιτάρι είναι τετραπλοειδές είδος (με 28 χρωμοσώματα σε τέσσερα ζεύγη), ενώ το μαλακό σιτάρι είναι εξαπλοειδές είδος (με 42 χρωμοσώματα σε 6 ζεύγη) (Royo et al., 2017).



Σχήμα 1 Παγκόσμια παραγωγή σίτου για το έτος 2020. Παρουσιάζονται οι 10 χώρες με τις μεγαλύτερες παραγωγές σε σιτάρι για το συγκεκριμένο έτος κατά φθίνουσα σειρά. Πηγή: [FAOSTAT](#) (2022).



Σχήμα 2 Η ελληνική παραγωγή (σκούρα γραμμή) και οι καλλιεργούμενες εκτάσεις (ανοιχτή γραμμή) σίτου για το διάστημα 2000 μέχρι 2022. Πηγή: [FAOSTAT](#) (2022).

1.1.2. Στάδια ανάπτυξης – κλίμακα Zadoks

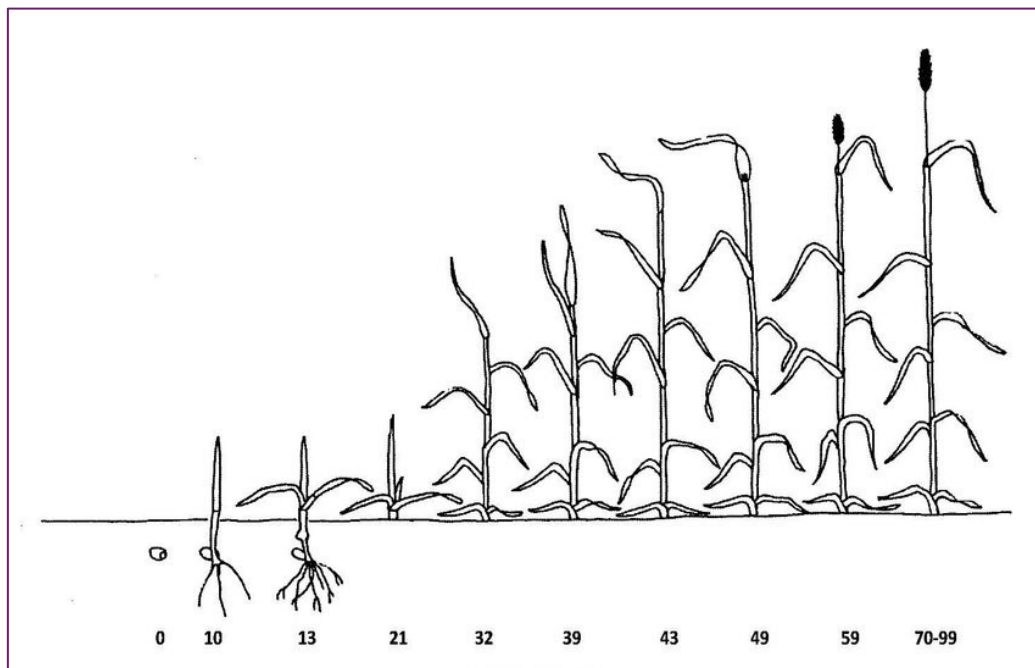
Τα σιτηρά είναι φυτά με καθορισμένο τρόπο ανάπτυξης, όπου η βλαστική ανάπτυξη διαχωρίζεται σαφώς από την αναπαραγωγική, με την εμφάνιση της ταξιανθίας. Για τα σιτηρά έχουν προταθεί εδώ και πολλά χρόνια, διάφορα συστήματα κωδικοποίησης των σταδίων ανάπτυξης, όπως των Large, Haun, Feeks και των Zadoks et al. (1974), το οποίο είναι και το πιο ευρέως χρησιμοποιούμενο. Τα χρόνια που ακολούθησαν την μελέτη του Zadoks, προτάθηκαν περαιτέρω διευκρινήσεις και τροποποιήσεις για την κλίμακα από τους Tottman (1987) και Tottman et al. (1979), καθώς και αναλυτικές οδηγίες για το πως πρέπει να γίνεται η δειγματοληψία για τον ακριβή προσδιορισμό των αναπτυξιακών σταδίων.

Τα δέκα κύρια στάδια ανάπτυξης των σιτηρών είναι τα εξής (Μπιλάλης, 2019):

- Βλάστηση σπόρου (0)
- Αύξηση φυταρίου (1)
- Αδέλφωμα (2)

- Καλάμωμα (3)
- Διόγκωση κολεού (4)
- Ξεστάχιασμα (5)
- Άνθηση (6)
- Γαλακτώδης καρπός (7)
- Στάδιο ζύμης (8)
- Ωρίμανση (9)

Η κλίμακα Zadoks παρέχει την πληρέστερη περιγραφή των σταδίων ανάπτυξης του σιταριού. Χρησιμοποιεί μια κωδικοποίηση που βασίζεται στα δέκα κύρια στάδια που αναφέρθηκαν παραπάνω, τα οποία μπορούν να υποδιαιρεθούν, καθιστώντας τον ένα κατάλληλο εργαλείο για συγκρίσεις με δεδομένα και μοντελοποιήσεις. Η κλίμακα Zadoks αποτελείται από τα 9 πρωτεύοντα στάδια (0-9), όπου το καθένα διαιρείται σε 10 δευτερεύοντα στάδια (έτσι προκύπτουν οι κατηγορίες 00-99) (Εικόνα 1). Για να γίνει ο προσδιορισμός του αναπτυξιακού σταδίου ενός φυτού, είναι απαραίτητο να γίνει πρώτα προσδιορισμός του πρωτεύοντος σταδίου (π.χ. αύξηση φυταρίου – GS1 ή αδέλφωμα – GS2), προτού



Εικόνα 1 Σχηματική αναπαράσταση των αναπτυξιακών σταδίων του σίτου. Από κάτω δίνονται κατά προσέγγιση οι κωδικοί που αντιστοιχούν στην κλίμακα Zadoks. Πηγή: Fowler, 2018.

προχωρήσουμε στον προσδιορισμό των δευτερευόντων σταδίων (π.χ. εμφάνιση 1^{ου} φύλλου – GS11 ή κύριο στέλεχος και 5 αδέρφια – GS25) (Fowler, 2018).

1.2. Γεωργία Ακριβείας

Η γεωργία ακριβείας (precision agriculture) περιλαμβάνει ένα σύνολο τεχνολογιών που συνδυάζει αισθητήρες, συστήματα πληροφοριών, βελτιωμένα μηχανήματα και ενημερωμένη διαχείριση, με κύριο σκοπό τη βελτιστοποίηση της παραγωγής, λαμβάνοντας υπόψη τη μεταβλητότητα και τις αβεβαιότητες που επικρατούν στα γεωργικά συστήματα. Παρέχει ένα καινοτόμο μέσο για την παρακολούθηση της αλυσίδας παραγωγής τροφίμων και διαχείρισης, τόσο της ποσότητας όσο και της ποιότητας των γεωργικών προϊόντων. Η γεωργία ακριβείας, ή αλλιώς η διαχείριση των συστημάτων γεωργικής παραγωγής με βάση τις πληροφορίες, εμφανίστηκε στα μέσα της δεκαετίας του 1980 ως ένας τρόπος για να εφαρμοστεί η σωστή γεωργική παρέμβαση, στο σωστό μέρος τη σωστή στιγμή (Robert, 2002). Η έλευση τεχνολογιών, όπως τα παγκόσμια δορυφορικά συστήματα πλοήγησης (GNSS), τα συστήματα γεωγραφικών πληροφοριών (GIS) και οι μικροϋπολογιστές (microcomputers), τις καθιστά κύρια εργαλεία για την λήψη ακριβέστερων αποφάσεων (Stafford, 2000).

Η γεωργία ακριβείας (ΓΑ) βρίσκεται πλέον σε προχωρημένη φάση και ένας αριθμός αγροτών έχει ήδη αρχίσει να την εφαρμόζει στις καλλιέργειές του, με ελπιδοφόρα αποτελέσματα. Οι δραστηριότητες που σχετίζονται με την ΓΑ περιλαμβάνουν: ταυτοποίηση και οριοθέτηση καλλιεργειών, εντόμων και ζιζανίων, συνεχή παρακολούθηση της απόδοσης, εκμηχάνιση των διεργασιών, δημιουργία μεταβαλλόμενων ζωνών λίπανσης και εφαρμογής ζιζανιοκτόνων, εντομοκτόνων και μυκητοκτόνων, παρακολούθηση και χαρτογράφηση της σποράς κ.α. (Akhter & Sofi, 2021; Buyya & Dastjerdi, 2016).

Λόγω της μεγάλης ποικιλίας των εφαρμογών και των διάφορων παραμέτρων σε κάθε καλλιέργεια, είναι δύσκολο να ποσοτικοποιηθούν τα οφέλη της γεωργίας ακριβείας γενικά. Σύμφωνα με τους Griffin & Lowenberg - DeBoer (2005), σε μια ανασκόπηση 234 μελετών που δημοσιεύθηκαν από το 1988 έως το 2005, η γεωργία ακριβείας βρέθηκε να είναι κερδοφόρα κατά μέσο όρο στο 68% των περιπτώσεων (Gebbers & Adamchuk, 2010).

1.3. Τηλεπισκόπηση

Με τον όρο τηλεπισκόπηση ή τηλεανίχνευση (remote sensing) περιγράφεται η διαδικασία λήψης πληροφοριών για ένα αντικείμενο, περιοχή ή φαινόμενο, με την χρήση ανιχνευτικών συσκευών που δεν βρίσκονται σε επαφή με το αντικείμενο που παρατηρείται. Αυτό ουσιαστικά περιγράφει την συλλογή δεδομένων για ένα μέρος της επιφάνειας της γης, μέσω της μέτρησης και αποτύπωσης της ηλεκτρομαγνητικής ακτινοβολίας που εκπέμπεται και ανακλάται από την επιφάνεια της γης και την ατμόσφαιρα (Καρτάλης & Φείδας, 2013). Σύμφωνα με τον Jensen, (2007) από το 1988 έχει υιοθετηθεί ένας κοινός ορισμός από την ASPRS (American Society for Photogrammetry and Remote Sensing) για την τηλεπισκόπηση και την φωτογραμμετρία, σύμφωνα με τον οποίο πρόκειται για «την τέχνη, επιστήμη και τεχνολογία της συλλογής αξιόπιστων πληροφοριών για φυσικά αντικείμενα και για το περιβάλλον μέσω μιας διαδικασίας καταγραφής, μέτρησης και ερμηνείας εικόνων και ψηφιακών αναπαραστάσεων ενεργειακών προτύπων που λαμβάνονται από απομακρυσμένα συστήματα αισθητήρων» (Colwell, 1997).

Η τηλεπισκόπηση έχει πολλαπλές διαστάσεις, καθώς περιλαμβάνει έννοιες όπως «στόχος» (δηλαδή το προς μελέτη αντικείμενο), την τεχνική συλλογής δεδομένων, καθώς και την ανάλυση και επεξεργασία τους. Τα τελευταία χρόνια οι κλάδοι της τηλεπισκόπησης που παρουσιάζουν το μεγαλύτερο ενδιαφέρον είναι η δορυφορική τηλεπισκόπηση και η τηλεπισκόπηση με την χρήση μη επανδρωμένων αεροσκαφών (drones). Η δορυφορική τηλεπισκόπηση αφορά την άντληση δεδομένων μέσω ανιχνευτών που φέρονται σε δορυφόρους σε τροχιά γύρω από την γη (Καρτάλης & Φείδας, 2013). Η τηλεπισκόπηση ως εργαλείο, έχει κάποια βασικά πλεονεκτήματα, όπως την μη παρεμβατική φύση της, καθώς δεν επηρεάζεται το αντικείμενο ενδιαφέροντος. Επίσης, γίνεται συστηματική συλλογή δεδομένων, περιορίζοντας έτσι τα δειγματοληπτικά σφάλματα, ενώ τα ίδια τα βιοφυσικά δεδομένα είναι πολύτιμες πηγές καινούριας γνώσης σε ευρεία κλίμακα. Αυτά είναι δυνατόν να εφαρμοστούν σε διάφορους κλάδους επιστημών. Ωστόσο, παρουσιάζει κάποιους περιορισμούς. Είναι ένα εργαλείο που προϋποθέτει καλή γνώση της χρήσης του, διότι το περιθώριο για ανθρώπινα σφάλματα είναι μεγάλο. Ακόμη, τα όργανα που χρησιμοποιούνται χάνουν εύκολα την ακρίβεια της αρχικής βαθμονόμησης, επηρεάζοντας έτσι την ποιότητα των

δεδομένων. Τέλος, πρόκειται για μια σχετικά ακριβή μέθοδο, που δεν είναι δυνατό να εφαρμοστεί σε προβλήματα πολύ μικρής κλίμακας (Jensen, 2007).

Τα πρώτα βήματα στην δορυφορική τηλεπισκόπηση έγιναν με την εκτόξευση του δορυφόρου Landsat-1 το 1972 για μη στρατιωτικές εφαρμογές (European Commission. Joint Research Centre. Institute for Environment and Sustainability., 2014). Το 1979, ο Seasat-1 έγινε ο πρώτος δορυφόρος απεικόνισης RADAR και ξεκίνησε έναν νέο τομέα τηλεπισκόπησης. Τα χρόνια που ακολούθησαν, ο τομέας της δορυφορικής τηλεπισκόπησης εξελίχθηκε ταχέως, καθώς εμφανίστηκαν νέα συστήματα οπτικών αισθητήρων και ραντάρ υψηλότερης χωρικής ανάλυσης, υπερφασματικοί αισθητήρες, σημαντικά υποπροϊόντα όπως το ψηφιακό υψομετρικό μοντέλο (Digital Elevation Model – DEM), όπως επίσης και περαιτέρω ανάπτυξη νέων τεχνικών επεξεργασίας χρησιμοποιώντας μηχανική μάθηση (Ali et al., 2015).

Ο συνδυασμός διαφορετικών τύπων δεδομένων τηλεπισκόπησης, μαζί με βοηθητικά δεδομένα από ένα σύνολο διάφορων πηγών, μπορεί να παρέχει νέες πληροφορίες και εργαλεία σε μια πληθώρα κλάδων. Τα δεδομένα τηλεπισκόπησης συμβάλλουν επίσης στην ανάπτυξη ενός πιο προηγμένου συστήματος γεωγραφικών πληροφοριών (GIS) το οποίο με τη σειρά του μπορεί να χρησιμοποιηθεί για την καλύτερη διαχείριση της γης, την διαχείριση των φυσικών πόρων, καθώς και λοιπές περιβαλλοντικές και άλλες εφαρμογές (Roy et al., 2017).

Γενικά, οι δορυφορικοί αισθητήρες κατατάσσονται σε δύο ευρείες κατηγορίες, αναλόγως την πηγή ακτινοβολίας που καταγράφουν. Αυτοί είναι οι παθητικοί (passive sensors) και οι ενεργητικοί αισθητήρες (active sensors). Επίσης, διακρίνονται σε ανεικονιστές (non – imaging sensors) και απεικονιστές (imaging sensors), ανάλογα με την διεργασία ανίχνευσης της περιοχής ενδιαφέροντος. Παραδείγματα ανεικονιστών αποτελούν παθητικά συστήματα, όπως το μικροκυματικό ραδιόμετρο (microwave radiometer) και ενεργητικά, όπως το μικροκυματικό αλτίμετρο (microwave altimeter). Στους απεικονιστές ανήκουν π.χ. ο οπτικός μηχανικός σαρωτής (optical mechanical scanner) και το ραντάρ συνθετικού ανοίγματος (synthetic aperture radar). Μία ακόμη κατηγοριοποίηση των δορυφόρων αφορά τον τρόπο με τον οποίο καταγράφουν την ακτινοβολία, με βάση τον οποίο διαχωρίζονται στα συστήματα πλαισίου (framing systems) και στους σαρωτές (scanning systems) (Καρτάλης & Φείδας , 2013).

1.4. Πρόβλεψη της απόδοσης

Τα πρώτα μοντέλα για την πρόβλεψη της απόδοσης που δημιουργήθηκαν, ήταν εμπειρικές στατιστικές και προσανατολισμένες στην διαδικασία προσεγγίσεις που στόχευαν στην προσομοίωση της ανάπτυξης μιας καλλιέργειας. Οι παράμετροι που λαμβανόταν υπόψιν συμπεριλάμβαναν βιοφυσικές, περιβαλλοντικές και διαχειριστικές παραμέτρους. Τα δεδομένα αυτά ήταν δύσκολο να ληφθούν, ειδικά όσο αυξανόταν η κλίμακα της υπό μελέτη περιοχής ή η απαιτούμενη ανάλυση (Basso et al., 2013; Han et al., 2020). Βάσει αυτών των δυσκολιών, έγινε διαχωρισμός των μοντέλων σε τρεις κατηγορίες από τους Claverie et al. (2012). Η πρώτη κατηγορία περιλάμβανε ορισμένα σύνθετα μοντέλα, τα οποία λάμβαναν υπόψιν μια πληθώρα φαινολογικών και φυσιολογικών παραμέτρων. Ορισμένα παραδείγματα αποτελούν τα CERES (Richie, 1985), AFRCWHEAT2 (Porter, 1993), SWHEAT (Porter et al., 1993). Η δεύτερη κατηγορία περιλαμβάνει απλούστερα μοντέλα που χρησιμοποιούν την προσέγγιση LUE (Light Use Efficiency) και αξιοποιούν τους δείκτες βλάστησης. Στην τρίτη κατηγορία ανήκουν μοντέλα που χρησιμοποιούν ημι-εμπειρικές προσεγγίσεις και αξιοποιούνται σε μεγαλύτερες εκτάσεις. Χαρακτηριστικό παράδειγμα αποτελεί το SAFY (Duchemin et al., 2008; Μεγγούδη, 2020).

Το CERES (Crop Environment Resource Synthesis), είναι ένα μοντέλο, το οποίο χρησιμοποιείται ευρέως σε καλλιέργειες και αξιοποιεί παραμέτρους όπως το κλίμα, τον γονότυπο, το έδαφος και τις μεθόδους καλλιέργειας σε παγκόσμια κλίμακα (Basso et al. 2016). Το μοντέλο CropSyst (Cropping Systems Simulation Model) (Van Evert & Campbell, 1994) εφαρμόζεται σε πολλαπλές καλλιέργειες για μεγάλα χρονικά διαστήματα και διερευνά τις κλιματικές και εδαφολογικές επιπτώσεις σε σχέση με το περιβάλλον και την απόδοση της καλλιέργειας (Stöckle et al., 2003).

Με την ταχεία ανάπτυξη της τηλεπισκόπησης (RS), δεδομένα ευρείας κλίμακας με υψηλή χωρική και χρονική ανάλυση έχουν γίνει ευρέως διαθέσιμα, από την δεκαετία του 1980. Έκτοτε, πολυφασματικά όργανα σε διάφορους δορυφόρους, όπως οι LANDSAT (Pollock & Kanemasu, 1979), MODIS (Trombetta et al., 2016), SPOT (Boissard et al., 1989) και πρόσφατα οι δορυφόροι Sentinel παρέχουν δωρεάν, ανεξάρτητα, και συνεχώς τηλεπισκοπικά δεδομένα σε παγκόσμιο

επίπεδο για τη μοντελοποίηση της ανάπτυξης των καλλιεργειών (Hunt et al., 2019).

1.5. Συστήματα γεωγραφικών πληροφοριών (G.I.S.)

Τα GIS ορίζονται ως «μια οργανωμένη συλλογή εξοπλισμού, λογισμικού, γεωγραφικών δεδομένων και εξειδικευμένου προσωπικού, σχεδιασμένη κατά τέτοιο τρόπο ώστε να συλλέγει και επιβεβαιώνει, να επεξεργάζεται, αποθηκεύει, ενημερώνει και να διαχειρίζεται όλους τους τύπους γεωγραφικών πληροφοριών» (Καρτάλης & Φείδας , 2013).

Η πρώτη προσπάθεια για συστηματική χρήση των χαρτογραφικών δεδομένων έγινε κατά τη διάρκεια των δεκαετιών του '60 και του '70. Η πρώτη οργανωμένη απόπειρα χρήσης των χαρτογραφικών δεδομένων σε Η/Υ έγινε από τον Howard T. Fisher, το 1963. Το πρόγραμμα του Fisher πήρε το όνομα SYMAP (SYnagraphic MAPping system) και δημιουργούσε απλούς χάρτες τυπώνοντας στατιστικές τιμές πάνω σε έναν κάρναβο. Το πρόγραμμα SYMAP ακολουθήθηκε από μία σειρά άλλων προγραμμάτων χαρτογράφησης, όπως το GRID και το IMGRID. Από τότε μία σειρά εξελίξεων, όχι μόνο στα λογισμικά αυτά αλλά και στη τεχνολογία των υπολογιστών, είχαν ως αποτέλεσμα τη δημιουργία νέων συστημάτων ολοένα πιο ισχυρών που χειρίζονται, αναλύουν και παρουσιάζουν, γεωγραφικής φύσεως πληροφορίες (Ευελπίδου & Αντωνίου , 2015).

Σήμερα, υπάρχουν στην αγορά G.I.S. που έχουν αναπτυχθεί από διαφορετικές εταιρείες, όμως όλα λειτουργούν με βάση την ίδια φιλοσοφία και τις ίδιες αρχές, έχουν παρόμοιες δυνατότητες και τηρούν συγκεκριμένα πρότυπα που υπαγορεύονται από τη διεθνή οργάνωση “Open G.I.S.”

Τα δεδομένα σε ένα G.I.S. αναπαρίστανται μέσω επιπέδων πληροφορίας (information layers), τα οποία περιέχουν το καθένα διαφορετικό είδος πληροφορίας. Τα δεδομένα ανήκουν σε δύο μεγάλες κατηγορίες, ανάλογα με τον τρόπο καταχώρισης και επεξεργασίας των χαρτογραφικών στοιχείων. Αυτές είναι τα raster (ψηφιδωτά) και vector (διανυσματικά). Παρουσιάζουν μεγάλες διαφορές, καθώς εξυπηρετούν διαφορετικές λειτουργίες (Mesev, 2007).

Η σειρά με την οποία τοποθετούνται αυτά τα επίπεδα έχει σημασία, έτσι ώστε να είναι εύκολα διαχειρίσιμα. Συνήθως, πρώτα τοποθετούνται τα σημειακά επίπεδα πληροφορίας, από κάτω τα γραμμικά, πιο κάτω τα επιφανειακά και τελευταία τα ψηφιδωτά (Βαϊόπουλος et al., 2008).

1.6. Δορυφόροι Sentinel-2

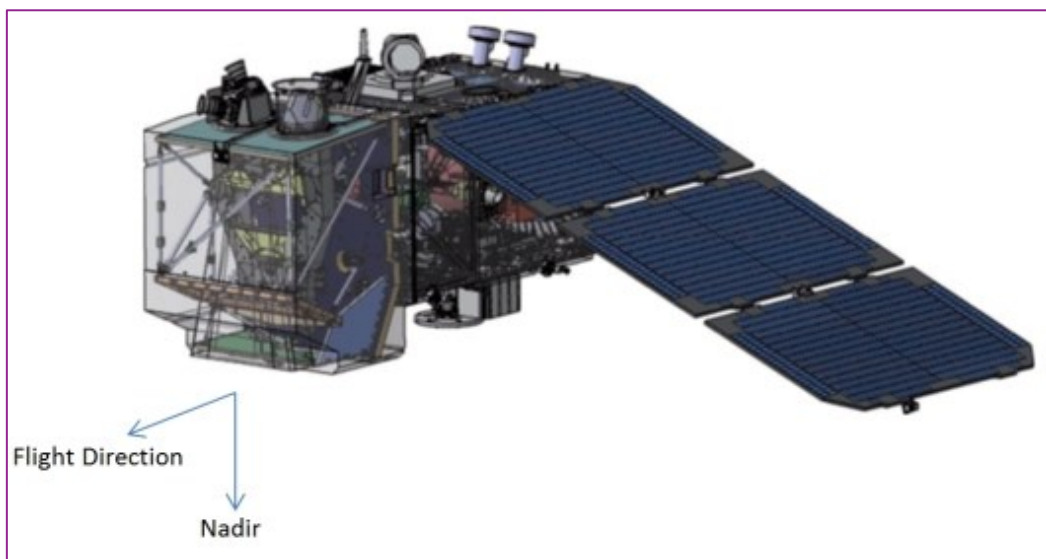
Το Ευρωπαϊκό επιστημονικό πρόγραμμα Copernicus δημιουργήθηκε με πρωτοβουλία της Ευρωπαϊκής Επιτροπής (European Commission) σε συνεργασία με τον Ευρωπαϊκό Οργανισμό Διαστήματος (European Space Agency – ESA). Το πρόγραμμα αυτό αποτελεί την συνέχεια του προγράμματος GMES (Global Monitoring for Environment and Security), δηλαδή του Ευρωπαϊκού Προγράμματος για την «Παγκόσμια Παρακολούθηση του Περιβάλλοντος και της Ασφάλειας» (Περάκης et al., 2015). Το 2014, το Copernicus έθεσε σε τροχιά τον πρώτο δορυφόρο Sentinel, τον Sentinel-1A. Μέχρι και σήμερα, το ερευνητικό πρόγραμμα Copernicus έχει θέσει σε τροχιά μια σειρά αποστολών με δορυφόρους Sentinel (1, 2, 3, 4, 5, 5P και 6).

Οι δορυφόροι Sentinel-2 είναι προϊόντα της συνεργασίας μεταξύ ESA και μιας κοινοπραξίας εξήντα εταιριών με επικεφαλής την Airbus Defense and Space και υποστηρίζονται από το γαλλικό CNES (Centre national d'études spatiales, Εθνικό Κέντρο Διαστημικών Μελετών) και το DLR (Deutsches Zentrum für Luft- und Raumfahrt, Γερμανικό Αεροδιαστημικό Κέντρο) (The European Space Agency, 2022).

Πρόκειται για 2 παρόμοιους δορυφόρους, τον Sentinel-2A και τον 2B (Εικόνα 2) (The European Space Agency, 2022). Ο δορυφόρος Sentinel-2A τέθηκε σε τροχιά στις 23/06/2015, μέσω ενός πυραύλου Vega από τον Ευρωπαϊκό Διαστημικό Σταθμό, κοντά στο Κούρου, στην Γαλλική Γουινέα (Περάκης et al., 2015). Οι πρώτες εικόνες ελήφθησαν λίγες μέρες αργότερα μετά την εκτόξευσή του (Immitzer et al., 2016). Ο δορυφόρος Sentinel-2B εκτοξεύτηκε αρκετά αργότερα, στις 7 Μαρτίου 2017. Οι δορυφόροι Sentinel-2 φέρουν ενσωματωμένα όργανα πολυφασματικής απεικόνισης (Multispectral Imaging Instruments – MSI), υψηλής διακριτικής ικανότητας, με τη δυνατότητα καταγραφής δεδομένων σε 13 ευρείες φασματικές ζώνες (The European Space Agency, 2022).

Ο πρωταρχικός στόχος της αποστολής Sentinel-2 είναι να παρέχει δορυφορικά δεδομένα υψηλής χωρικής, αλλά και χρονικής ανάλυσης για την παρακολούθηση της κάλυψης και της χρήσης της γης, την παρακολούθηση της κλιματικής αλλαγής και διάφορων φυσικών καταστροφών (Malenovsky et al., 2012). Ένας ακόμη βασικός στόχος των αποστολών Sentinel-2 είναι να παρέχει συμπληρωματικές πληροφορίες με αυτές άλλων παγκόσμιων δορυφορικών προγραμμάτων, όπως το Landsat και το SPOT (Satellite Pour l'Observation de la Terre), διασφαλίζοντας έτσι την συνεχή παρακολούθηση της επιφάνειας της γης (Pesaresi et al., 2016; Phiri et al., 2020).

Οι δορυφόροι Sentinel-2 έχουν εύρος σάρωσης 290 χλμ. και συχνή επαναδιέλευση και καταγραφή της επιφάνειας της γης (οι δύο δορυφόροι παρέχουν χρονική ανάλυση 5 ημερών στον Ισημερινό, καθώς διανύουν την ίδια τροχιά). Από τα 13 κανάλια, τα 4 έχουν χωρική ανάλυση στα 10 m, τα 6 στα 20 m και τα 3 στα 60 m. Αναλυτικές πληροφορίες για τα φασματικά κανάλια παρατίθενται



Εικόνα 2 Σχηματική αναπαράσταση των δορυφόρων Sentinel-2. Πηγή: [Sentinel-2 – ESA](#) (2022)

στον Πίνακα 1. Κάθε ένας από τους δορυφόρους ζυγίζει 1140 kg. Η διάρκεια ζωής τους έχει σχεδιαστεί στα 7 έτη, με προοπτική να επιμηκυνθεί (Περάκης et al., 2015). Οι μπαταρίες και τα προωθητικά με τα οποία είναι εξοπλισμένοι επαρκούν για 12 χρόνια λειτουργίας (The European Space Agency, 2022). Βρίσκονται σε ύψος 786 χλμ. πάνω από την επιφάνεια της γης και σε γωνιακή απόκλιση 180° μεταξύ τους (Περάκης et al., 2015). Η ακριβής θέση κάθε δορυφόρου Sentinel-2

στην τροχιά του μετράται από έναν δέκτη διπλής συχνότητας Global Navigation Satellite System (GNSS). Η τροχιακή ακρίβεια διατηρείται από ένα αποκλειστικό σύστημα πρόωσης.

Πίνακας 1 Φασματικές ζώνες του δορυφόρου Sentinel-2, Πηγή: [Sentinel-2 – ESA](#) (2022)

| Αριθμός καναλιού | S2A | | S2B | | Χωρική Ανάλυση (m) |
|------------------|-----------------------------|---------------------|-----------------------------|---------------------|--------------------|
| | Κεντρικό μήκος κύματος (nm) | Εύρος καναλιού (nm) | Κεντρικό μήκος κύματος (nm) | Εύρος καναλιού (nm) | |
| 1 | 442.7 | 20 | 442.3 | 20 | 60 |
| 2 | 492.7 | 65 | 492.3 | 65 | 10 |
| 3 | 559.8 | 35 | 558.9 | 35 | 10 |
| 4 | 664.6 | 30 | 664.9 | 31 | 10 |
| 5 | 704.1 | 14 | 703.8 | 15 | 20 |
| 6 | 740.5 | 14 | 739.1 | 13 | 20 |
| 7 | 782.8 | 19 | 779.7 | 19 | 20 |
| 8 | 832.8 | 105 | 832.9 | 104 | 10 |
| 8a | 864.7 | 21 | 864 | 21 | 20 |
| 9 | 945.1 | 19 | 943.2 | 20 | 60 |
| 10 | 1373.5 | 29 | 1376.9 | 29 | 60 |
| 11 | 1613.7 | 90 | 1610.4 | 94 | 20 |
| 12 | 2202.4 | 174 | 2185.7 | 184 | 20 |

1.7. Μηχανική Μάθηση (Machine Learning)

1.7.1. Γενικά στοιχεία

Ως μηχανική μάθηση (ML), ορίζεται «το φαινόμενο κατά το οποίο ένα σύστημα βελτιώνει την απόδοσή του κατά την εκτέλεση μιας συγκεκριμένης εργασίας, χωρίς να υπάρχει ανάγκη να προγραμματιστεί εκ νέου (Γεωργούλη, 2015)». Η ML έχει ως σκοπό την δημιουργία μηχανών που έχουν την ικανότητα να μαθαίνουν, δηλαδή να βελτιώνουν την απόδοσή τους σε διάφορους τομείς, μέσω της αξιοποίησης προηγούμενων γνώσεων και εμπειριών (Γεωργούλη, 2015).

Στην ML, χρησιμοποιούνται συνήθως τέσσερις κατηγορίες μεθόδων μάθησης, η καθεμία χρήσιμη για την επίλυση διαφορετικών εργασιών (Choi et al., 2020):

- 1) Επιβλεπόμενη μάθηση (Supervised Learning)
- 2) Μη επιβλεπόμενη μάθηση (Unsupervised Learning)
- 3) Ημιεπιβλεπόμενη μάθηση (Semisupervised Learning)
- 4) Ενισχυτική μάθηση (Reinforcement Learning)

Η ML είναι ένας κλάδος που εστιάζει στην μαθησιακή πτυχή της τεχνητής νοημοσύνης (Artificial Intelligence – AI). Χρησιμοποιεί υποσύνολα δεδομένων, αναπτύσσοντας έτσι αλγορίθμους που χρησιμοποιούν νέους ή διαφορετικούς συνδυασμούς χαρακτηριστικών από αυτά που μπορούν να προκύψουν από τις βασικές αρχές του προγραμματισμού (James et al., 2013).

Από όλες τις κατηγορίες που προαναφέρθηκαν, η επιβλεπόμενη μάθηση θεωρείται η πιο κοινώς εφαρμοζόμενη. Ένας υποκλάδος της επιβλεπόμενης μάθησης που αποκτά ιδιαίτερο ενδιαφέρον τα τελευταία χρόνια, περιλαμβάνει τα βαθιά νευρωνικά δίκτυα (deep neural networks), τα οποία είναι δίκτυα πολλαπλών επιπέδων μονάδων κατωφλιού (threshold units), καθένα από τα οποία υπολογίζει μια παραμετροποιημένη συνάρτηση των δεδομένων που εισάγονται (Schmidhuber, 2015). Τα συστήματα βαθιάς μάθησης (Deep Learning) χρησιμοποιούν αλγορίθμους βελτιστοποίησης, βασισμένους σε βαθμίδες, έτσι ώστε να προσαρμόσουν τις παραμέτρους σε ένα τέτοιο πολυσύνθετο δίκτυο, με βάση τα σφάλματα που προκύπτουν στα τελικά αποτελέσματα (output) (Jordan & Mitchell, 2015).

1.7.2. Βασικές αρχές μηχανικής μάθησης

Γενικότερα, στόχος των αλγορίθμων μηχανικής μάθησης είναι η βελτιστοποίηση της απόδοσης μιας εργασίας, μέσω της εκμετάλλευσης προηγούμενης εμπειρίας. Πιο συγκεκριμένα, η ML μπορεί να δημιουργήσει αποτελεσματικές σχέσεις μεταξύ των εισροών δεδομένων (data inputs) και να ανακατασκευάσει ένα γνωστικό σχήμα. Στην μεθοδολογία που βασίζεται σε δεδομένα, ο μεγαλύτερος όγκος τους συνεπάγεται και καλύτερη λειτουργία της μηχανικής μάθησης (Vieira et al., 2020). Σε γενικές γραμμές, τα συστήματα ML λειτουργούν με δύο

βασικές διαδικασίες, αυτήν της μάθησης (learning), η οποία χρησιμοποιείται για την εκπαίδευση του συστήματος (training) και την δοκιμή (testing) (Lopez-Arevalo et al., 2020).

Στην απλή μορφή της, η ML βασιζόμενη σε δεδομένα εκπαίδευσης, στη φάση της εκμάθησης μαθαίνει να εκτελεί την εργασία από την εμπειρία. Μόλις η μαθησιακή επίδοση φτάσει σε ένα ικανοποιητικό σημείο (το οποίο εκφράζεται μέσω μαθηματικών και στατιστικών σχέσεων), τελειώνει. Στη συνέχεια, το μοντέλο που αναπτύχθηκε μέσω της εκπαιδευτικής αυτής διαδικασίας χρησιμοποιείται περαιτέρω για ταξινόμηση (classification), ομαδοποίηση (clustering) ή πρόβλεψη (prediction) (Benos et al., 2021).

1.7.3. ML στον αγροτικό τομέα

Οι αλγόριθμοι μηχανικής μάθησης έχουν χρησιμοποιηθεί τα τελευταία χρόνια για την εκμετάλλευση των πολύπλοκων σχέσεων μεταξύ των τηλεπισκοπικών δεδομένων και των διάφορων παραμέτρων μιας καλλιέργειας. Η επιτυχία τους οφείλεται εν μέρει στην ικανότητά τους να χειρίζονται με αποτελεσματικό τρόπο γραμμικές και μη – γραμμικές σχέσεις. Έτσι μπορούν να διερευνηθούν σε βάθος νέες σχέσεις μεταξύ τηλεπισκοπικών δεδομένων και απόδοσης των καλλιεργειών (Gonzalez-Sanchez et al., 2014).

Ακόμη, παρουσιάζουν ευελιξία στο να ορίσουν τα χρησιμότερα φασματικά κανάλια για την ακριβή πρόβλεψη της απόδοσης σε μια καλλιέργεια (Hunt et al., 2019). Δεδομένου ότι οι αλγόριθμοι ML απαιτούν μεγάλους όγκους δεδομένων εκπαίδευσης και επικύρωσης για να εξάγουν ακριβή συμπεράσματα, είναι σημαντικό να υπάρχουν διαθέσιμα επαρκή δεδομένα απόδοσης, τα οποία θα έχουν και παρόμοια ανάλυση με αυτή των τηλεπισκοπικών δεδομένων. Αυτό είναι δυνατό με τη χρήση μηχανημάτων συγκομιδής που φέρουν συσκευές χαρτογράφησης και καταγραφής της απόδοσης.

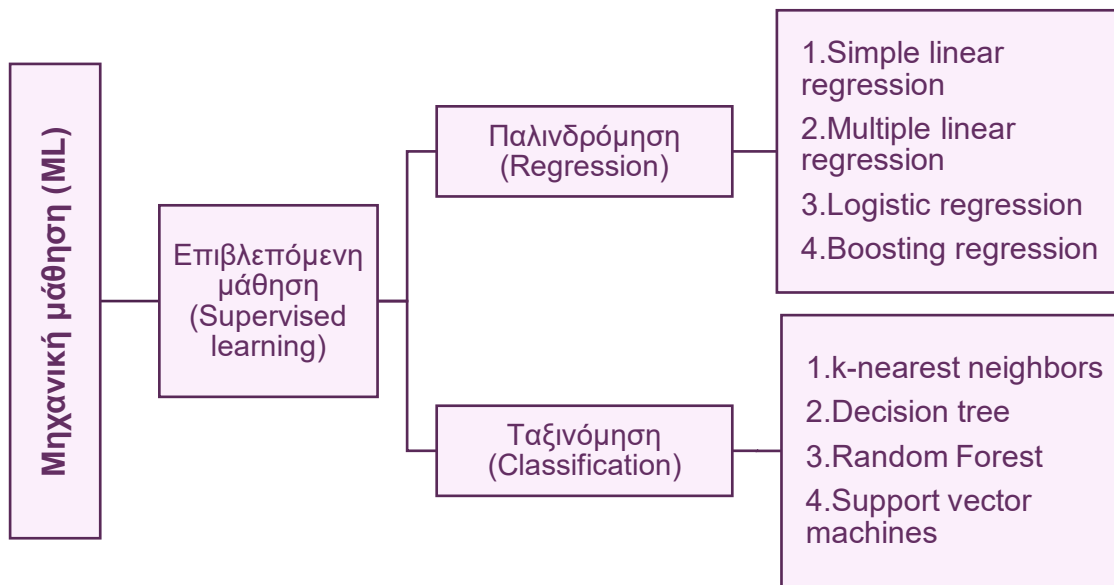
Υπάρχει ένα πλήθος τεχνικών μηχανικής μάθησης που χρησιμοποιούνται για την αναπαράσταση διαφόρων βιοφυσικών διεργασιών. Η επιλογή της κάθε τεχνικής γίνεται με βάση τα μεμονωμένα πλεονεκτήματα και μειονεκτήματα που παρουσιάζει η εκάστοτε τεχνική σε διαφορετικά είδη προβλημάτων. Ορισμένες από αυτές είναι το Artificial Neural Networks (ANN) (Chlingaryan et al., 2018), το

Support Vector Machine (SVM) (Zou et al., 2022), το Random Forests (RF) (Ali et al., 2022), το K-nearest neighbors (KNN) (Zou et al., 2022), το Boosting Regression (BR) (Ali et al., 2022) κ.λπ.

1.8. Αλγόριθμοι ML

Σήμερα, οι αλγόριθμοι ML χρησιμοποιούνται ευρέως για την επίλυση ποικίλων προβλημάτων. Αυτό οφείλεται στην ολοένα αυξανόμενη διαθεσιμότητα καινοτόμων αλγορίθμων, καθώς και στην ύπαρξη μεγάλου όγκου δεδομένων που μπορούν να αξιοποιηθούν ελεύθερα από το διαδίκτυο. Όπως έχει ήδη αναφερθεί, οι αλγόριθμοι ML κατηγοριοποιούνται στην επιβλεπόμενη, την μη-επιβλεπόμενη και την ενισχυτική μάθηση. Οι αλγόριθμοι επιβλεπόμενης μάθησης που χρησιμοποιούνται για την πρόβλεψη κατηγορικών (ποιοτικών) μεταβλητών είναι γνωστοί ως αλγόριθμοι ταξινόμησης (classification). Οι αλγόριθμοι που χρησιμοποιούνται για την πρόβλεψη αριθμητικών (ποσοτικών) μεταβλητών είναι γνωστοί ως αλγόριθμοι παλινδρόμησης (regression).

Οι μη-επιβλεπόμενοι αλγόριθμοι ML λειτουργούν με δεδομένα χωρίς κατηγορία και εξάγουν αποτελέσματα μέσω της ομαδοποίησης παρόμοιων αντικειμένων (Sharma et al., 2021). Παρακάτω δίνεται ένα διάγραμμα (Σχήμα 3) που κατηγοριοποιεί ορισμένους κοινώς χρησιμοποιούμενους αλγορίθμους ML, στους οποίους περιλαμβάνονται και οι τρεις αλγόριθμοι που χρησιμοποιήθηκαν στην παρούσα εργασία.



Σχήμα 3 Διαγραμματική επεξήγηση των κατηγοριών ορισμένων αλγορίθμων μηχανικής μάθησης. Πηγή: Sharma et al., (2021)

1.8.1. Random Forests (RF)

Ο αλγόριθμος RF, ανήκει στους μη-παραμετρικούς αλγορίθμους επιβλεπόμενης μάθησης και η λειτουργία του βασίζεται στην δημιουργία δέντρων αποφάσεων (decision trees) σε διαφορετικά υποσύνολα ενός συνόλου δεδομένων. Κάθε δέντρο αποτελεί έναν προγνωστικό παράγοντα που δημιουργείται με την επιλογή ενός τυχαίου δείγματος του αρχικού συνόλου δεδομένων, αλλά όλα τα «δέντρα» στο «δάσος» έχουν τα ίδια χαρακτηριστικά κατανομής. Μετά τη δημιουργία ενός μεγάλου αριθμού μεμονωμένων δέντρων, ο αλγόριθμος θα επιλέξει τις πιο δημοφιλείς κλάσεις με βάση τον συνδυασμό των προβλέψεων με ένα πλειοψηφικό σύστημα ψηφοφορίας (majority voting system) (Han et al., 2020). Επομένως, ο RF παρέχει υψηλή αποτελεσματικότητα στο χειρισμό μεγάλων συνόλων δεδομένων και μειώνει τον κίνδυνο της υπερπροσαρμογής (overfitting) (Rhee & Im, 2017; Vincenzi et al., 2011), μια ιδιότητα που κρίθηκε ιδιαίτερα χρήσιμη στην παρούσα εργασία. Ακόμη, ο RF μπορεί να ποσοτικοποιήσει την σχετική σημαντικότητα των μετρούμενων μεταβλητών (δηλ. δημιουργεί πίνακες σημαντικότητας) και φαίνεται ότι είναι λιγότερο επιρρεπής σε σφάλματα λόγω ακραίων τιμών (Strobl et al., 2007).

Ορισμένα από τα βασικά πλεονεκτήματα των RF συμπεριλαμβάνουν (Breiman, 2001, Τσιάρα, 2012):

- το περιορισμένο σφάλμα γενίκευσης (που ως αποτέλεσμα έχει την αποφυγή υπερπροσαρμογής),
- την ολοκλήρωση του αλγορίθμου με σταθερό αριθμό βημάτων,
- τη διατήρηση χαμηλής μεροληψίας (bias) και διακύμανσης (variance) στη συσχέτιση,
- την ανεκτικότητα προς το θόρυβο και
- την δυνατότητα χρήσης για ομαδοποίηση.

Αντίθετα, οι αλγόριθμοι RF μειονεκτούν κυρίως ως προς:

- το ψηλό υπολογιστικό κόστος.

1.8.2. K-nearest neighbors (KNN)

Ο KNN (Zhang, 2020), είναι μια μη-παραμετρική μέθοδος ταξινόμησης και ανήκει στις επιβλεπόμενες τεχνικές ML. Ταξινομεί τα δεδομένα δοκιμής (test data) με βάση τα k πιο κοντινά παραδείγματα εκπαίδευσης σε ένα σύνολο δεδομένων. Ο αλγόριθμος αυτός χρησιμοποιείται για κατηγοριοποίηση δεδομένων, όπως επίσης και για μάθηση συναρτήσεων (Pallathadka et al., 2021). Θεωρείται ανεκτικός στο θόρυβο και κατάλληλος να επικεντρωθεί σε μικρότερες, πιο συγκεκριμένες χρονικές περιόδους (Aha et al., 1991; Zou et al., 2022). Η λειτουργία του βασίζεται στη μάθηση μέσω παραδειγμάτων, καθώς υποθέτει την ομοιότητα μεταξύ μιας καινούριας προγνωστικής μεταβλητής (predictor variable) και μια ομάδας εκπαίδευσης (training group). Έπειτα ταξινομεί τον νέο παράγοντα στην πιο όμοια κατηγορία. Πρόκειται για ταξινόμηση με βάση τις αποστάσεις των μεταβλητών από την πλησιέστερη ομάδα εκπαίδευσης (Han et al., 2020). Ο KNN φαίνεται να μην αποδίδει επαρκώς όταν χρησιμοποιηθεί σε πολύ μικρές ομάδες εκπαίδευσης (Aha et al., 1991).

1.8.3. Boosting regressions (BR)

Ο αλγόριθμος BR είναι ένας γενικός αλγόριθμος ML με βελτιωμένη ακρίβεια πρόβλεψης (Ali et al., 2015; Jia et al., 2022) που βασίζεται σε μια οικογένεια μεμονωμένων τεχνικών μηχανικής μάθησης. Η λειτουργία του βασίζεται στην υπόθεση πως κάθε μεμονωμένος αλγόριθμος ML είναι αδύναμος. Έτσι, εκτελεί παράλληλους υπολογισμούς μέσω μιας επαναληπτικής διαδικασίας και χρησιμοποιεί εκτιμήσεις αριθμητικού και σταθμισμένου (weighted mean) μέσου όρου για να ψηφίσει την κυρίαρχη πρόβλεψη. Μετά από πολλές επαναλήψεις, ο αλγόριθμος BR συνδυάζει αυτές τις αδύναμες προβλέψεις σε έναν μοναδικό, ισχυρότερο κανόνα πρόβλεψης. Κατ' επέκταση, είναι ικανό να παρέχει πρόσθετη ακρίβεια πρόβλεψης σε σύγκριση με άλλους αλγορίθμους, όμως πρόκειται για ένα μοντέλο ιδιαίτερα υψηλής πολυπλοκότητας (Ali et al., 2022).

1.9. Σκοπός της εργασίας

Σκοπός της εργασίας αυτής είναι η ανάπτυξη μοντέλων πρόβλεψης της απόδοσης σε χωράφια σκληρού σίτου, μέσω της χρήσης τεχνικών της μηχανικής μάθησης και η σύγκρισή τους με ένα ήδη υπάρχον μοντέλο πρόβλεψης που χρησιμοποιεί τους δείκτες βλάστησης (VI-MLR).

Για την δημιουργία των μοντέλων χρησιμοποιήθηκαν τηλεπισκοπικά δεδομένα από τους δορυφόρους Sentinel-2, των οποίων η επεξεργασία έγινε με την χρήση των λογισμικών ελεύθερης πρόσβασης. Τα αποτελέσματα των μοντέλων συσχετίστηκαν με δεδομένα απόδοσης των χωραφιών, τα οποία καταγράφηκαν από θεριζοαλωνιστική μηχανή εξοπλισμένη με σύστημα χαρτογράφησης της απόδοσης.

Για την κατασκευή των μοντέλων χρησιμοποιήθηκαν τρεις διαφορετικοί αλγόριθμοι, οι RF, KNN και BR, και έγιναν πολλαπλές δοκιμές με διαφορετικά διαστήματα ανά καλλιεργητική περίοδο και ορισμένες ρυθμίσεις στις παραμέτρους των αλγορίθμων. Τέλος, έγινε στατιστική ανάλυση των αποτελεσμάτων και σύγκρισή τους με την πρώτη προσέγγιση (VI-MLR).

2. Υλικά και Μέθοδοι

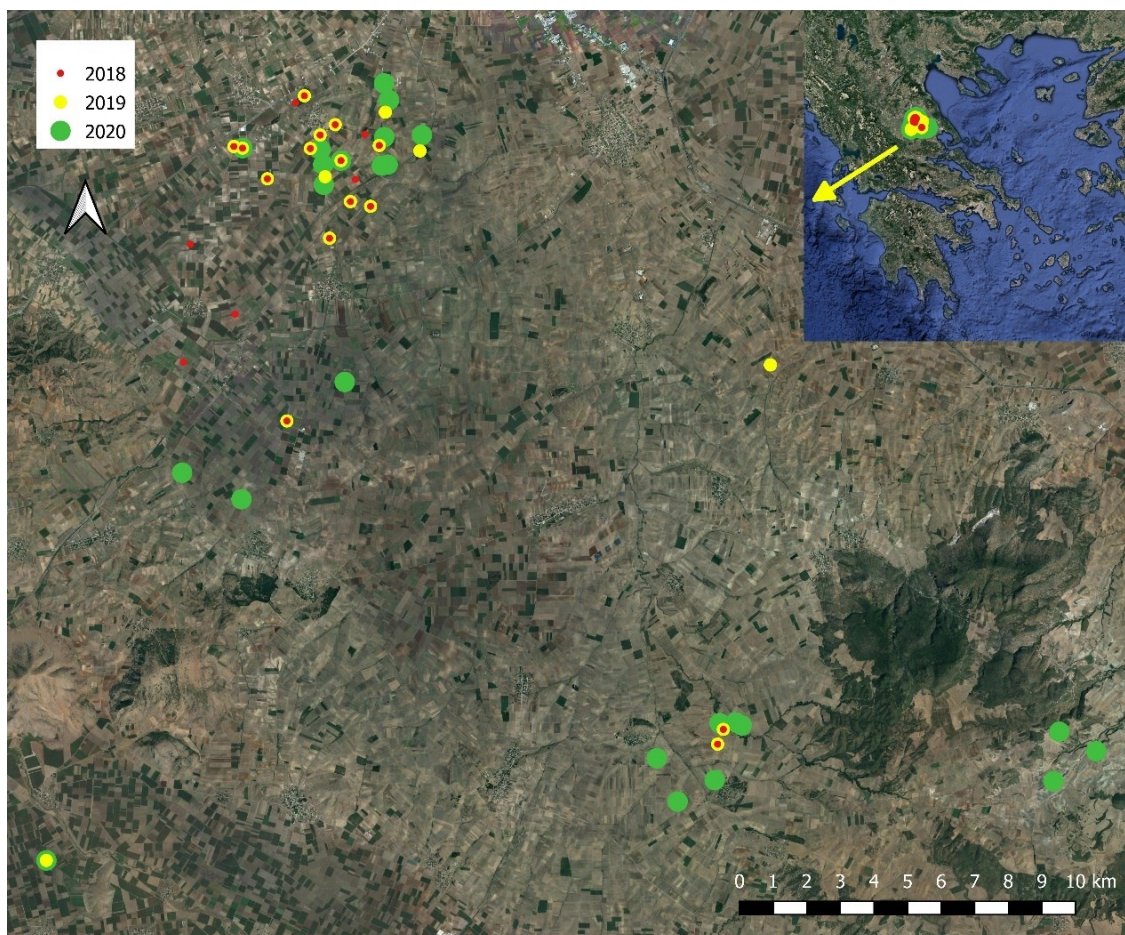
2.1. Περιοχή μελέτης

Όλοι οι αγροί που χρησιμοποιήθηκαν στην μελέτη ανήκουν στην περιοχή της Θεσσαλίας, στην Κεντρική Ελλάδα (Εικόνα 3). Η παρακολούθηση της περιοχής και η συλλογή των δεδομένων αφορά τις καλλιεργητικές περιόδους 2017 – 2018, 2018 – 2019 και 2019 – 2020. Ο συνολικός αριθμός των χωραφιών που χρησιμοποιήθηκαν στην μελέτη ήταν 66, εκ των οποίων 21 χωράφια την περίοδο 2017 – 2018, 21 την περίοδο 2018 – 2019 και 24 την περίοδο 2019 – 2020 (Πίνακας 2).

Τα κριτήρια επιλεξιμότητας των χωραφιών για την μελέτη ήταν:

- 1) Η διαθεσιμότητα δεδομένων από τους χάρτες απόδοσης. Αυτό ήταν δυνατό μόνο στα χωράφια που χρησιμοποιήθηκε θεριζοαλωνιστική μηχανή ακριβείας με αισθητήρα μέτρησης σπόρου.
- 2) Το κατάλληλο μέγεθος, σχήμα και προσανατολισμός του κάθε χωραφιού. Συγκεκριμένα, το κάθε χωράφι έπρεπε να έχει επαρκή έκταση, έτσι ώστε να εξάγεται μεγάλος αριθμός pixels στην ανάλυση που παρέχει ο δορυφόρος Sentinel – 2 (10m × 10m). Όσον αφορά το σχήμα των χωραφιών, προτιμήθηκαν τα πιο τετραγωνισμένα χωράφια που έχουν συνεχόμενες εκτάσεις με δεδομένα απόδοσης σπόρου. Αποφεύχθηκε η χρήση χωραφιών με φυσικά εμπόδια όπως δέντρα, κατασκευές, ρυάκια που διακόπτουν την βλάστηση και αλλοιώνουν τις τιμές των pixels.

Η καλλιέργεια των χωραφιών έγινε με διάφορες ποικιλίες σκληρού σίτου (Iridae, Meridiano, Normano, Simeto, Svevo) και ανήκαν σε διαφορετικούς αγρότες, άρα κατ' επέκταση ακολουθήθηκαν ξεχωριστές καλλιεργητικές πρακτικές. Με αυτόν τον τρόπο, τα αποτελέσματα της μελέτης αφορούσαν την ευρεία εφαρμογή τους, ανεξαρτήτως των καλλιεργητικών πρακτικών. Για όλα τα χωράφια, η σπορά γινόταν τον Νοέμβριο και η συγκομιδή τον Ιούνιο.



Εικόνα 3 Χάρτης της περιοχής της Θεσσαλίας που περιλαμβάνει όλα τα χωράφια της μελέτης. Στην πάνω-δεξιά γωνία φαίνεται η θέση της περιοχής στην Ελλάδα. Τα χωράφια είναι σημειωμένα με κόκκινο (2018), κίτρινο (2019) και πράσινο (2020) χρώμα, ανάλογα την καλλιεργητική περίοδο. Η εικόνα δημιουργήθηκε με το λογισμικό [QGIS](#).

Πίνακας 2 Πληροφορίες των χωραφιών που μελετήθηκαν.

| Καλλιεργητική περίοδος | Αριθμός χωραφιών | Έκταση (ha) | Αριθμός εικονοστοιχείων (pixels) | Αριθμός εικόνων |
|------------------------|------------------|---------------|----------------------------------|-----------------|
| 2017-2018 | 21 | 53.04 | 5304 | 35 |
| 2018-2019 | 21 | 50.31 | 5031 | 26 |
| 2019-2020 | 24 | 85.91 | 8591 | 32 |
| Σύνολο | 66 | 189.26 | 18926 | 93 |

2.2. Μέτρηση απόδοσης

Στο τέλος της κάθε καλλιεργητικής περιόδου (Ιούνιος) έγινε η συγκομιδή των χωραφιών με τη θεριζοαλωνιστική μηχανή John Deere S660i, εξοπλισμένη με σύστημα χαρτογράφησης της απόδοσης που παρέχει, μέσω του σχετικού λογισμικού MyJD (My John Deere, 2022), χάρτες απόδοσης σε διανυσματικά δεδομένα σε μορφή σημείου (point vector data) με χωρική ανάλυση $1.75\text{m} \times 2.5\text{m}$ περίπου (ανάλογα με την ταχύτητα συγκομιδής).

Ακολούθησε περαιτέρω επεξεργασία των χαρτών, χειροκίνητα, στο λογισμικό QGIS (QGIS Geographic Information System, 2022), για να αφαιρεθούν οι τιμές στα περιθώρια των χωραφιών. Οι τιμές αυτές αποτελούν «θόρυβο» στα δεδομένα, καθώς δεν ανταποκρίνονται στην πραγματική εικόνα της απόδοσης. Αυτό συμβαίνει διότι η ροή του σπόρου στην μηχανή καθυστερεί στα άκρα του χωραφιού, λόγω της αλλαγής ταχύτητας και της περιστροφής της θεριζοαλωνιστικής.

Επίσης, χρειάστηκε να γίνει μετατροπή των χαρτών απόδοσης σε δεδομένα raster με τη χρήση της διαδικασίας inverse distance weighting (IDW) του QGIS και τελικά εφαρμόστηκε η διαδικασία της επαναδειγματοληψίας (resampling), με σκοπό να μετατραπεί το μέγεθος των εικονοστοιχείων σε $10\text{m} \times 10\text{m}$. Αυτό είναι απαραίτητο, διότι τα pixels των εικόνων του Sentinel-2 έχουν μέγεθος $10\text{m} \times 10\text{m}$, κι έτσι τα δύο επίπεδα δεδομένων (layers) συμπίπτουν ακριβώς και είναι δυνατή η σύγκριση τους.

2.3. Δορυφορικά Δεδομένα

Στην μελέτη χρησιμοποιήθηκαν συνολικά 93 εικόνες από τους δορυφόρους Sentinel-2 (A και B), για το διάστημα Οκτώβριο έως Ιούνιο καθ' όλη την περίοδο μελέτης (2017–2020). Οι εικόνες αυτές ελήφθησαν από την ιστοσελίδα Copernicus Open Access Hub της ESA (Open Access Hub, 2022). Για να αξιοποιηθεί μια εικόνα στην μελέτη ήταν απαραίτητο να μην υπάρχουν σύννεφα που να σκιάζουν τα προς ανάλυση χωράφια. Κατά τη διάρκεια της 3ετούς μελέτης, η μεγαλύτερη περίοδος χωρίς διαθέσιμες εικόνες (λόγω νεφοκάλυψης) δεν ξεπέρασε τις 20 ημέρες.

Τα πολυφασματικά όργανα (MultiSpectral Instruments, MSI) που φέρουν οι δορυφόροι Sentinel-2, παρέχουν πληροφορίες σε 13 φασματικές ζώνες (443–2190 nm), σε χωρική ανάλυση μεγέθους pixel 10, 20 ή 60 m, ανάλογα το μήκος κύματος, και με χρονική ανάλυση 5 ημερών (οι δορυφόροι A και B λαμβάνουν ο καθένας εικόνες ανά 10 μέρες, με 5 μέρες διαφορά μεταξύ τους). Στην παρούσα μελέτη, οι δορυφορικές εικόνες που χρησιμοποιήθηκαν αποτελούν προϊόντα ανακλαστικότητας της επιφάνειας της γης (Bottom of Atmosphere, BOA), επεξεργασμένες σε επίπεδο Level2A, που σημαίνει ότι είναι γεωμετρικά, ατμοσφαιρικά και ραδιομετρικά διορθωμένες. Πρόκειται για προϊόντα ελεύθερης πρόσβασης που παρέχονται από τον Ευρωπαϊκό Οργανισμό Διαστήματος (European Space Agency, ESA).

Περαιτέρω επεξεργασία των δορυφορικών εικόνων πραγματοποιήθηκε με την χρήση του ελεύθερου λογισμικού ανοιχτού κώδικα SNAP – ESA Sentinels Application Platform έκδοση 7.0 (STEP—Science Toolbox Exploitation Platform, 2022). Σκοπός του βήματος αυτού ήταν να ληφθούν όλα τα φασματικά δεδομένα για όλα τα χωράφια στο διάστημα της 3ετίας. Για να μπορέσουν να είναι πιο εύκολα διαχειρίσιμες οι εικόνες, λόγω μεγέθους, έπρεπε να γίνει περικοπή της μικρότερης δυνατής έκτασης που να περιλαμβάνει όλα τα χωράφια που επιλέχθηκαν. Αυτό έγινε με την διαδικασία του subset. Επίσης, έγινε επαναδειγματοληψία (resampling) των εικόνων σε μέγεθος pixel 10m ×10m. Αυτό ήταν απαραίτητο, καθώς για διαφορετικά μήκη κύματος τα pixels έχουν διαφορετική ανάλυση, πράγμα που καθιστά την σύγκριση δύσκολη. Ο συνολικός αριθμός pixel για τα υπό μελέτη 66 χωράφια που περιλαμβάνονταν στις εικόνες ήταν 18.926.

Εφόσον οι εικόνες ήταν έτοιμες, έγινε εξαγωγή των φασματικών δεδομένων των 13 καναλιών για κάθε pixel. Δημιουργήθηκε ένα αρχείο στο excel που περιλάμβανε όλες τις πληροφορίες σε μορφή πίνακα.

2.4. Χρονοσειρές NDVI

Προτού γίνει η μοντελοποίηση της απόδοσης, δημιουργήθηκαν οι χρονοσειρές του δείκτη βλάστησης NDVI (Normalized Difference Vegetation Index) για το σύνολο των χωραφιών σε κάθε καλλιεργητική περίοδο. Σκοπός αυτού ήταν να υπάρχει μια γενική εικόνα για την πορεία των καλλιεργειών σε κάθε έτος, έτσι

ώστε να αποκλειστεί το ενδεχόμενο να υπήρξαν μεγάλες διαφορές στην ανάπτυξη των καλλιεργειών στο διάστημα των τριών ετών που μας ενδιέφερε. Κάτι τέτοιο θα μείωνε την αποδοτικότητα των μοντέλων. Οι χρονοσειρές αυτές περιλαμβάνουν τον μέσο όρο των τιμών του δείκτη βλάστησης NDVI για όλα τα pixels των 66 χωραφιών που χρησιμοποιήθηκαν ανά περίοδο (21, 21 και 24 χωράφια για τις περιόδους 2018, 2019 και 2020 αντίστοιχα). Χρησιμοποιήθηκαν συνολικά 93 εικόνες χωρίς σύννεφα για την περίοδο (2017-2020). Για την εξαγωγή του δείκτη χρησιμοποιήθηκαν τα φασματικά κανάλια B4 ($R_{664.6}$) και B8 ($R_{832.8}$) του Sentinel-2 από το σύνολο των 18.926 εικονοστοιχείων. Η εξίσωσή του δίνεται από την σχέση:

$$\text{Normalized Difference Vegetation Index, NDVI} = \frac{(R_{832.8} - R_{664.6})}{(R_{832.8} + R_{664.6})}$$

όπου R_x , είναι η ανακλαστικότητα στο μήκος κύματος x (το x αντιστοιχεί στο κεντρικό μήκος κύματος του εκάστοτε φασματικού καναλιού του Sentinel-2).

2.5. Μοντελοποίηση

Εφαρμόστηκαν δύο προσεγγίσεις μοντελοποίησης για την εκτίμηση της απόδοσης σίτου, μια προηγουμένως αναπτυγμένη προσέγγιση πολλαπλής γραμμικής παλινδρόμησης (multiple linear regression), με την χρήση δεικτών βλάστησης ως παράγοντες πρόβλεψης της απόδοσης (VI-MLR) και μια προσέγγιση μηχανικής μάθησης (Machine Learning, ML) που διερευνά την απόδοση τριών διαφορετικών αλγορίθμων ML στο σύνολο των φασματικών δεδομένων του Sentinel-2, επίσης ως παράγοντες πρόβλεψης της απόδοσης.

2.5.1. Μοντέλο με βάση τους δείκτες βλάστησης (VI-MLR)

Η πρώτη προσέγγιση δεν αποτελεί μέρος της παρούσας μελέτης, αλλά αντικείμενο της μελέτης των Cavalaris et al. (2021). Τα χωράφια για τις δύο πρώτες καλλιεργητικές περιόδους που ερευνώνται στην παρούσα εργασία (2017 – 2018, 2018 – 2019) είναι το αντικείμενο μελέτης της παραπάνω έρευνας. Για λόγους

σύγκρισης με το μοντέλο μηχανικής εκμάθησης επιλέχθηκαν τα αποτελέσματα με τις καλύτερες αποδόσεις του προηγούμενου μοντέλου.

Αυτό το μοντέλο περιλαμβάνει το ολοκλήρωμα του δείκτη βλάστησης EVI (Enhanced Vegetation Index) από τις 20 Απριλίου έως τις 31 Μαΐου ως το σήμα που εκπέμπουν τα φυτά, τον δείκτη NMDI (Normalized Multiband Drought Index) πριν από την ημερομηνία σποράς ως το σήμα που εκπέμπεται από το έδαφος και τον δείκτη NMDI στα τέλη Απριλίου, ως το σήμα που αφορά το νερό της βλάστησης (ανεξάρτητες μεταβλητές).

Όσον αφορά το εύρος των μηκών κύματος, αξιοποιήθηκαν τα φασματικά κανάλια B2 (R₄₉₀), B4 (R₆₆₅), B8 (R₈₄₂), B11 (R₁₆₁₀) και B12 (R₂₁₉₀) του Sentinel-2 από το σύνολο των 18.926 εικονοστοιχείων. Με αυτά δημιουργήθηκαν οι αντίστοιχες χρονοσειρές των δεικτών EVI και NMDI, των οποίων οι εξισώσεις δίνονται παρακάτω:

$$\text{Enhanced Vegetation Index, EVI} = 2.5 \frac{R_{842} - R_{665}}{R_{842} + 6R_{665} - 7.5R_{490} + 1}$$

$$\text{Normalized Multiband Drought Index, NMDI} = \frac{R_{842} - (R_{1610} - R_{2190})}{R_{842} + (R_{1610} - R_{2190})}$$

όπου R_x, είναι η ανακλαστικότητα στο μήκος κύματος x (το x αντιστοιχεί στο κεντρικό μήκος κύματος του εκάστοτε φασματικού καναλιού του Sentinel-2). Έπειτα, εφαρμόστηκε η διαδικασία της γραμμικής παρεμβολής (linear interpolation). Η μέθοδος αυτή επιτρέπει την παρεμβολή νέων σημείων δεδομένων μεταξύ δύο γνωστών τιμών σε μια καμπύλη (Farin, 2002). Αυτό γίνεται στην προκειμένη περίπτωση, διότι υπάρχει μια τιμή ανά 5 μέρες, αλλά είναι απαραίτητη η εξαγωγή τιμών για κάθε μέρα ξεχωριστά. Έτσι εξήχθησαν τα δεδομένα για κάθε μέρα για την περίοδο που εξετάστηκε (20 Απριλίου έως 31 Μαΐου, συν μία εικόνα από τον Οκτώβριο, πριν γίνει η σπορά).

2.5.2. Μοντέλα με βάση την μηχανική μάθηση (ML)

Στην δεύτερη προσέγγιση, που αποτελεί το αντικείμενο της παρούσας εργασίας, χρησιμοποιήθηκαν όλα τα φασματικά δεδομένα και των 13 καναλιών του δορυφόρου Sentinel-2, ως ανεξάρτητες μεταβλητές και εφαρμόστηκαν σε 3 διαφορετικούς αλγόριθμους μηχανικής μάθησης. Οι αλγόριθμοι που επιλέχθηκαν είναι ευρέως χρησιμοποιούμενοι. Αυτοί είναι οι random forest, k-nearest neighbors και boosting regression. Η αναλυτική επεξήγηση της λειτουργίας και των ιδιοτήτων του κάθε αλγόριθμου γίνεται στο [Κεφάλαιο 1.8](#).

Από τις 93 εικόνες που υπάρχουν για την 3ετή περίοδο που μελετήθηκε, επιλέχθηκαν οι 33, οι οποίες αφορούν τους μήνες από τέλη Οκτωβρίου έως τέλη Μαΐου. Αυτό αντιστοιχεί σε 11 εικόνες για κάθε καλλιεργητική περίοδο, με μία έως τρεις εικόνες ανά μήνα. Διευκρινίζεται ότι παρ' όλο που οι περισσότερες ημερομηνίες μεταξύ καλλιεργητικών περιόδων δεν συμπίπτουν, η μεγαλύτερη απόκλιση τους δεν ξεπερνά τις 5 ημέρες. Αυτό είναι σημαντικό, καθώς ένα διάστημα μιας εβδομάδας είναι αρκετό ώστε να υπάρχουν διαφορές στην ανάπτυξη μιας καλλιέργειας σίτου (αυτό εξαρτάται και από το αναπτυξιακό στάδιο στο οποίο βρίσκεται το φυτό).

Για κάθε ημερομηνία που υπήρχαν δεδομένα, χρησιμοποιήθηκαν και οι 13 φασματικές ζώνες του δορυφόρου Sentinel-2, ως ανεξάρτητες μεταβλητές στους 3 αλγόριθμους ML. Έτσι δημιουργήθηκαν συνολικά 143 μεταβλητές (13 φασματικά κανάλια × 11 ημερομηνίες).

Η διαδικασία που ακολουθήθηκε στο ML έχει ως εξής: Στην πρώτη δοκιμή των αλγορίθμων χρησιμοποιήθηκαν όλα τα δεδομένα από όλες τις ημερομηνίες. Έπειτα, άρχισε να αφαιρείται μια ημερομηνία κάθε φορά, ξεκινώντας από το τέλος της περιόδου (δηλαδή Μάιο, έπειτα Απρίλιο, Μάρτιο, κ.ο.κ.). Αυτό έγινε για να βρεθεί το μικρότερο δυνατό διάστημα από την αρχή της καλλιεργητικής περιόδου (δηλαδή να χρησιμοποιηθούν όσο λιγότερες εικόνες γίνεται) που να παρέχει πρόβλεψη με επαρκή ακρίβεια.

Σε όλες τις προσεγγίσεις μοντελοποίησης, τα δεδομένα χωρίστηκαν τυχαία σε ένα σύνολο εκπαίδευσης – training set (50% των δεδομένων) και σε ένα σύνολο επικύρωσης – validation set (50% των δεδομένων) για την αξιολόγηση της

απόδοσης της μοντελοποίησης. Ο ίδιος διαχωρισμός δεδομένων εφαρμόστηκε σε όλα τα μοντέλα.

2.6. Στατιστική ανάλυση

Η διαδικασία της στατιστικής ανάλυσης πραγματοποιήθηκε με την χρήση του ελεύθερου λογισμικού JASP, έκδοση 0.16 (JASP—A Fresh Way to Do Statistics Inc., 2022). Στην πρώτη προσέγγιση μοντελοποίησης εφαρμόστηκε η μέθοδος της πολλαπλής γραμμικής παλινδρόμησης μεταξύ των ανεξάρτητων μεταβλητών και της τελικής απόδοσης των χωραφιών, ενώ στην δεύτερη εφαρμόστηκαν οι 3 αλγόριθμοι της μηχανικής μάθησης.

Η απόδοση των μοντέλων αξιολογήθηκε με τον συντελεστή προσδιορισμού R^2 (coefficient of determination), την τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος RMSE (Root mean square error) και την κλίση της πιο κατάλληλης γραμμής που προκύπτει μεταξύ των μετρημένων και των μοντελοποιημένων αποδόσεων.

3. Αποτελέσματα

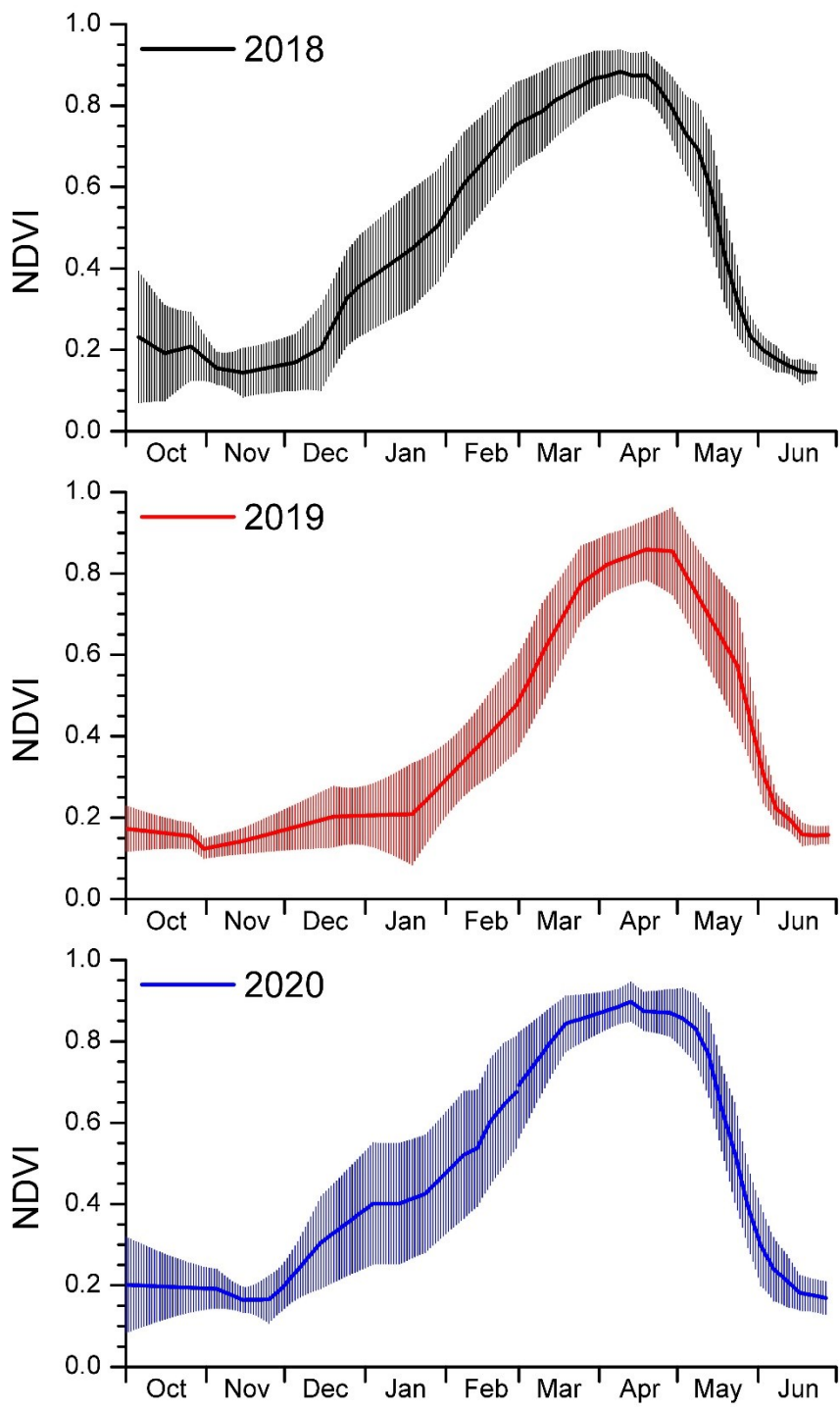
Στο Σχήμα 4 παρουσιάζονται οι χρονοσειρές του δείκτη βλάστησης NDVI για όλες τις καλλιεργητικές περιόδους που μελετήθηκαν (2017-2020). Όπως φαίνεται, οι τιμές του NDVI είναι αντιπροσωπευτικές για το είδος και την περιοχή της καλλιέργειας. Η τιμή του και στις τρεις περιόδους μειώνεται λίγο πριν την σπορά (που έγινε τον Νοέμβριο), πράγμα που συμβαδίζει με την πραγματική εικόνα του αγρού, καθώς αντιστοιχεί σε γυμνό, οργωμένο έδαφος. Έπειτα, ο δείκτης αυξάνεται, όσο αναπτύσσεται η καλλιέργεια. Εδώ γίνεται αντιληπτή μια διαφορά στην ανάπτυξη του σιταριού το 2019 σε αντίθεση με τα έτη 2018 και 2020, καθώς ο δείκτης αυξάνεται πιο αργά τους μήνες Ιανουάριο – Φεβρουάριο. Αυτό πιθανώς να οφείλεται σε δυσμενείς καιρικές συνθήκες (χαμηλές θερμοκρασίες, χιονόπτωση) που να προκάλεσαν επιβράδυνση της ανάπτυξης του σιταριού. Τα άλλα δύο έτη φαίνεται να έχουν παρόμοια πορεία. Παρόλα αυτά, οι υπόλοιποι μήνες της αναπτυξιακής περιόδου δεν έχουν σημαντικές διαφορές ούτε χρονικά, αλλά ούτε και ποσοτικά όσον αφορά τον NDVI. Σε όλες τις περιπτώσεις, ο δείκτης παίρνει τις μέγιστες τιμές κατά τα τέλη Μαρτίου μέχρι τέλη Απριλίου, που αντιστοιχούν σε υγιή ανεπτυγμένη βλάστηση ($NDVI = 0.8 - 0.9$). Η μόνη διαφορά εντοπίζεται στο έτος 2020, όπου ο δείκτης φαίνεται να διατηρεί τις μέγιστες τιμές για λίγο μεγαλύτερο χρονικό διάστημα σε σύγκριση με τα άλλα δύο έτη (αρχές Μαρτίου μέχρι αρχές Μαΐου). Τέλος, όλα τα έτη παρουσιάζουν παρόμοια πτώση του δείκτη κατά του καλοκαιρινούς μήνες, γεγονός που συμβαδίζει με την εικόνα της καλλιέργειας, καθώς βρίσκεται στο στάδιο του γεμίσματος του κόκκου (μετά το ξεστάχασμα και την ανθοφορία), όπου τα φυτά χάνουν το πράσινο χρώμα τους και όλη η νεοσυντιθέμενη ξηρά ουσία μεταφέρεται στους κόκκους.

Στο Σχήμα 5 παρουσιάζονται μερικές από τις δοκιμές μοντελοποίησης που έγιναν, με βάση τον αριθμό εικόνων που χρησιμοποιήθηκαν και τον αλγόριθμο που επιλέχθηκε. Αρχικά, η απόδοση της πρώτης προσέγγισης μοντελοποίησης μέσω δεικτών βλάστησης (VI-MLR) φαίνεται να είναι μέτρια, με $R^2 = 0,532$ και $RMSE = 847 \text{ kg ha}^{-1}$ (Σχήμα 5α). Όπως φαίνεται στο σχήμα α, υπάρχουν ορισμένα σημεία, ειδικά κατά την περίοδο ανάπτυξης 2018 – 2019, όπου δεν προβλέπονται επαρκώς από το μοντέλο. Αυτά είναι τα «κόκκινα νέφη» που παρατηρούνται επάνω αριστερά και επάνω δεξιά, που αντιστοιχούν σε δεδομένα του

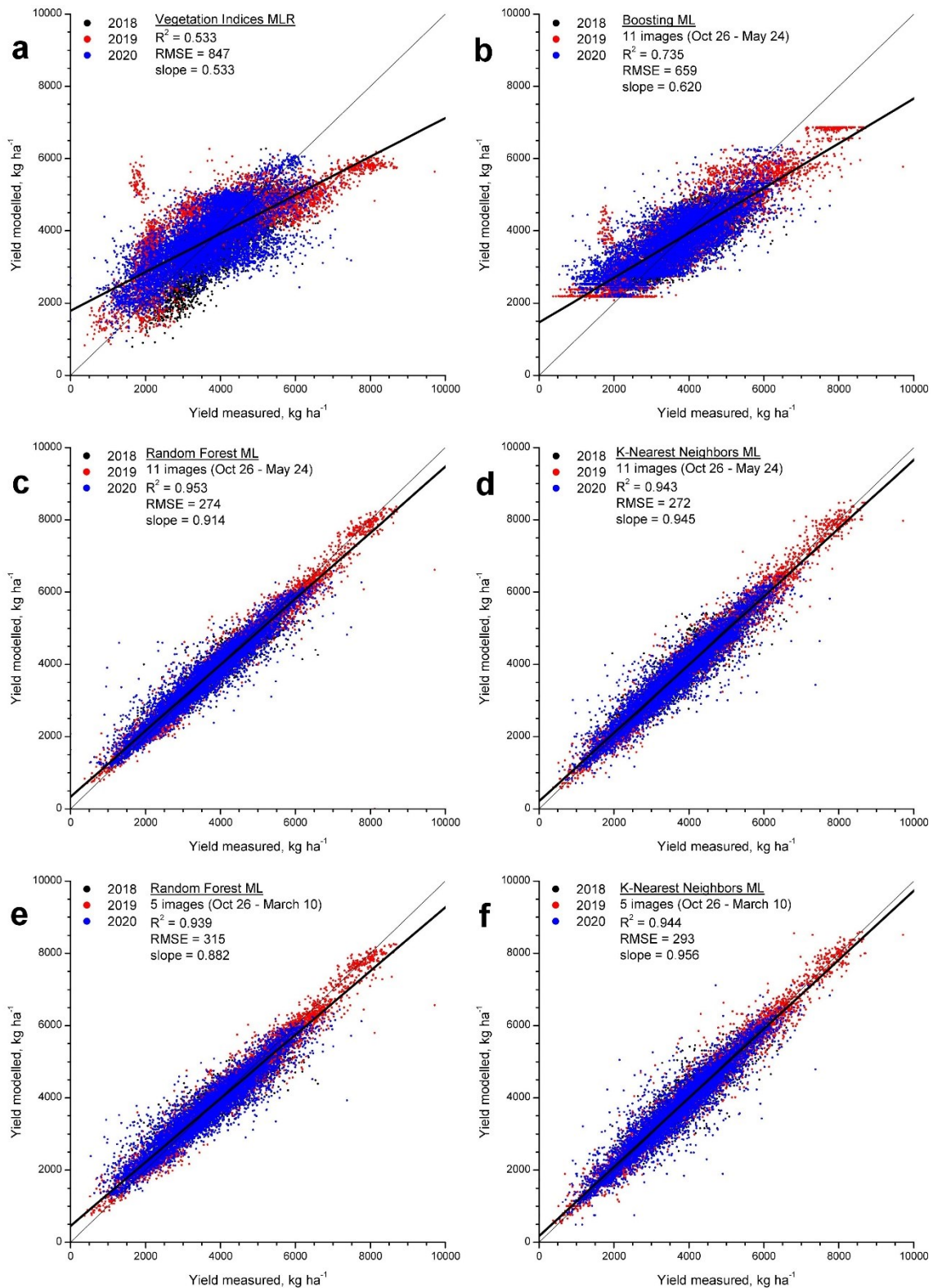
2019 και αποκλίνουν από τα υπόλοιπα δεδομένα. Αυτό μειώνει την συνολική απόδοση του μοντέλου σε μεγάλο βαθμό. Έτσι επηρεάζεται η κλίση της γραμμής, η οποία ισούται με 0,536, δηλαδή απέχει αρκετά από την μονάδα (απόλυτη ταύτιση).

Η δεύτερη προσέγγιση μοντελοποίησης φαίνεται να αποδίδει σαφώς καλύτερα συγκριτικά με την πρώτη, καθώς και οι τρεις αλγόριθμοι ML δίνουν καλύτερα αποτελέσματα από το μοντέλο VI-MLR. Αυτό ισχύει κυρίως όταν χρησιμοποιηθεί το σύνολο των 33 εικόνων (11 ανά έτος, όπως παρατίθεται στα σχήματα), δηλαδή από την αρχή της καλλιεργητικής περιόδου έως τον Μάιο. Ωστόσο, παρόλο που ο αλγόριθμος BR αποδίδει καλύτερα από το μοντέλο VI-MLR ($R^2 = 0,723$), διατηρεί ένα σχετικά υψηλό RMSE (668 kg ha^{-1}) και χαμηλή κλίση (0,622) (Πίνακας 3 και Σχήμα 5b). Εδώ αξίζει να σημειωθεί ότι το μοντέλο VI-MLR και το μοντέλο BR εμφανίζουν τα ίδια αδύναμα σημεία, δηλαδή φαίνεται πως και στις δύο περιπτώσεις υπάρχουν κοινά χωράφια για τα οποία δεν υπάρχει ακριβής πρόβλεψη.

Σε αντίθεση με το BR, οι αλγόριθμοι RF και KNN παρουσιάζουν πολύ καλές και παρόμοιες επιδόσεις, με $R^2 > 0,91$, $\text{RMSE} < 360 \text{ kg ha}^{-1}$ και κλίσεις κοντά στο 1 (Σχήμα 5c-d). Και στις δύο περιπτώσεις, όλα τα χωράφια εμφανίζονται πολύ κοντά στη γραμμή 1:1 χωρίς να υπάρχουν ακραίες τιμές. Ακόμη, παρατηρείται ότι ο αλγόριθμος BR χάνει την υπεροχή του σε σχέση με το μοντέλο VI-MLR όταν μειώνεται ο αριθμός των εικόνων που χρησιμοποιούνται (πριν το τέλος Απριλίου). Αυτό δεν συμβαίνει με τον RF και τον KNN, καθώς και οι δύο διατηρούν την υψηλή απόδοσή τους, ακόμη και όταν οι εικόνες που χρησιμοποιούνται μειωθούν σε 5 ανά καλλιεργητική περίοδο (δηλαδή μέχρι τα μέσα Απριλίου) ή και σε 3 (δηλαδή μέχρι τα μέσα Ιανουαρίου) (Σχήμα 5e-f και Πίνακας 3).



Σχήμα 4 Χρονοσειρές δείκτη βλάστησης NDVI για τις τρεις καλλιεργητικές περιόδους 2017-2018, 2018-2019 και 2019-2020 αντίστοιχα (από τον Οκτώβριο μέχρι τέλη Ιουνίου). Οι τιμές του δείκτη αντιπροσωπεύουν τον μέσο όρο των rixels όλων των χωραφιών που μελετήθηκαν την κάθε καλλιεργητική περίοδο. Στο διάγραμμα φαίνεται και η τυπική απόκλιση (standard deviation) για κάθε τιμή.



Σχήμα 5 Συσχέτιση μεταξύ υπολογισμένης και μοντελοποιημένης απόδοσης. Στο (a) παρουσιάζεται το μοντέλο με δείκτες βλάστησης, μέσω της πολλαπλής γραμμικής παλινδρόμησης (VI-MLR). Στα (b-f) παρουσιάζονται τα μοντέλα ML. Σε κάθε υπόμνημα αναγράφεται ο αλγόριθμος, ο αριθμός των εικόνων ανά καλλιεργητική περίοδο (με το αντίστοιχο διάστημα ημερομηνιών), το R^2 , το RMSE και η κλίση (slope) της γραμμής. Τα δεδομένα αφορούν 66 χωράφια σε τρεις καλλιεργητικές περιόδους (που υποδεικνύονται με διαφορετικά χρώματα). Η λεπτή, μαύρη γραμμή αντιστοιχεί στη γραμμή 1:1 και η έντονη μαύρη γραμμή στην γραμμή τάσης.

Πίνακας 3. Σύγκριση αποδοτικότητας των τριών αλγορίθμων ML για την εκτίμηση της απόδοσης. Σε κάθε σειρά χρησιμοποιείται διαφορετικός αριθμός εικόνων ανά καλλιεργητική περίοδο. Παρουσιάζεται το R² και το RMSE (kg ha⁻¹) μεταξύ της μετρημένης και της μοντελοποιημένης απόδοσης. Τα δεδομένα αφορούν συνολικά 66 χωράφια σε 3 περιόδους ανάπτυξης, τα οποία αντιστοιχούν σε 9463 pixels.

| Ημερομηνίες | Random Forest | | K – Nearest Neighbors | | Boosting Regression | |
|--------------------------------------|----------------|------|-----------------------|------|---------------------|------|
| | R ² | RMSE | R ² | RMSE | R ² | RMSE |
| 26/10 – 24/05 (11 εικόνες) | 0.953 | 274 | 0.943 | 272 | 0.735 | 659 |
| 26/10 – 29/04 (8 εικόνες) | 0.948 | 287 | 0.940 | 303 | 0.687 | 711 |
| 26/10 – 10/03 (5 εικόνες) | 0.939 | 315 | 0.944 | 293 | 0.460 | 941 |
| 26/10 – 28/02 (4 εικόνες) | 0.932 | 331 | 0.932 | 324 | 0.408 | 982 |
| 26/10 – 19/01 (3 εικόνες) | 0.924 | 353 | 0.923 | 345 | 0.359 | 1010 |

4. Συζήτηση

Η χρήση τηλεπισκοπικών δεδομένων από δορυφόρους για τη δημιουργία μοντέλων πρόβλεψης απόδοσης έχει γίνει ολοένα και πιο διαδεδομένη, καθώς πλέον υπάρχει ελεύθερη πρόσβαση στα περισσότερα δεδομένα. Η υψηλή χωρική ανάλυση και η μεγάλη συχνότητα επανεπίσκεψης των δορυφόρων Sentinel-2 της ESA έχουν προσδώσει επιπλέον ενδιαφέρον στην παραπάνω προσέγγιση και έχει γίνει μεγάλος αριθμός προσπαθειών να εκτιμηθεί ή να προβλεφθεί η απόδοση σίτου σε μικρότερα χωράφια μέσω τηλεπισκοπικών δεδομένων. Παρ' όλα αυτά, ο όγκος της βιβλιογραφίας παραμένει χαμηλός προς το παρόν. Αυτό οφείλεται κυρίως στην χαμηλή διαθεσιμότητα χωρικών δεδομένων για την απόδοση των χωραφιών, τα οποία παράγονται από τις θεριζοαλωνιστικές μηχανές που φέρουν συστήματα χαρτογράφησης της απόδοσης (Hunt et al., 2019, Segarra et al., 2022). Για να είναι εφικτή η κατασκευή πιο εξελιγμένων και αποδοτικών μοντέλων αλλά και η επαλήθευσή τους, απαιτούνται μεγάλοι όγκοι δεδομένων, που είναι δύσκολο να συλλεχθούν για μια περιοχή περιορισμένης έκτασης.

Στην παρούσα εργασία έγινε παρουσίαση και συγκριτική αξιολόγηση δύο διαφορετικών προσεγγίσεων μοντελοποίησης για την εκτίμηση της απόδοσης σε χωράφια σκληρού σίτου. Τα δεδομένα που χρησιμοποιήθηκαν προέρχονται από τον δορυφόρο Sentinel-2. Στην πρώτη προσέγγιση του μοντέλου VI-MLR, οι δείκτες βλάστησης χρησιμοποιήθηκαν σε πολλαπλές γραμμικές παλινδρομήσεις με την απόδοση, ενώ στη δεύτερη προσέγγιση αξιοποιήθηκαν εργαλεία της μηχανικής μάθησης, μέσω της χρήσης τριών διαφορετικών αλγορίθμων, των random forest (RF), k-nearest neighbors (KNN) και boosting regressions (BR).

Τα αποτελέσματα έδειξαν την σαφή υπεροχή της απόδοσης των αλγορίθμων RF και KNN σε σύγκριση με τα μοντέλο VI-MLR και τον αλγόριθμο BR. Αυτές οι συγκρίσεις φαίνεται να συμφωνούν με τα ευρήματα των Hunt et al. (2019), Jeong et al. (2016) και Segarra et al. (2022), όπου το μοντέλο RF για την πρόβλεψη απόδοσης σε σίτο είχε υψηλότερες επιδόσεις έναντι των κλασικών μοντέλων πολλαπλής γραμμικής παλινδρόμησης (multiple linear regression). Τα μοντέλα RF παρουσιάζουν ορισμένα βασικά πλεονεκτήματα έναντι των

παραδοσιακών μοντέλων παλινδρόμησης για την εκτίμηση της απόδοσης, καθώς διερευνούν τις σχέσεις μεταξύ επεξηγηματικών μεταβλητών (δηλαδή ανεξάρτητων) για τον έλεγχο συγχυτικών παραγόντων (Hunt et al., 2019). Διαχωρίζουν ένα τυχαίο υποσύνολο από την βαθμονόμηση για τη δοκιμή της απόδοσης και χρησιμοποιούν μόνο το υπόλοιπο σύνολο δεδομένων για την εκπαίδευση του μοντέλου (model training), διατηρώντας έτσι ορισμένες πληροφορίες για να αξιολογηθεί η ακρίβεια του μοντέλου (Jeong et al., 2016). Ο αλγόριθμος RF φαίνεται ότι σε πολλές περιπτώσεις υπερσχύει έναντι άλλων τεχνικών μηχανικής μάθησης, όπως στην μελέτη των Wang et al., (2016), όπου η ακρίβεια εκτίμησης της βιομάζας σε σιτάρι του RF ξεπέρασε αυτήν των μοντέλων SVR (support vector regression) και ANN (artificial neural networks).

Η συγκριτική υπεροχή του αλγορίθμου RF επιβεβαιώθηκε και στην μελέτη των Han et al., (2020). Σύμφωνα με αυτή, έγινε σύγκριση της απόδοσης οκτώ διαφορετικών αλγορίθμων ML για να προβλεφθεί η απόδοση σίτου με τηλεπισκοπικά δεδομένα. Αποδείχτηκε πως μεταξύ όλων των αλγορίθμων το RF έδωσε τις υψηλότερες αποδόσεις. Ωστόσο, φαίνεται πως το μοντέλο KNN είχε την χειρότερη επίδοση. Το KNN παρουσίασε ένα R^2 περίπου 0,65 μεταξύ της προβλεπόμενης και της παρατηρούμενης απόδοσης σίτου και RMSE πάνω από 1000 kg ha⁻¹. Αυτά τα ευρήματα έρχονται σε πλήρη αντιδιαστολή με το παρούσα εργασία, καθώς οι επιδόσεις του KNN είναι παρόμοιες με αυτές του RF, με $R^2 = 0,917$ για ολόκληρη την περίοδο ανάπτυξης (11 εικόνες ανά περίοδο) και RMSE = 357 kg ha⁻¹. Η διαφορά αυτή μπορεί να οφείλεται στο ότι οι συγγραφείς χρησιμοποίησαν τα τηλεπισκοπικά δεδομένα του δορυφόρου MODIS, ο οποίος έχει αρκετά χαμηλότερη χωρική ανάλυση (250m). Όπως έχει εξηγηθεί στην βιβλιογραφική ανασκόπηση (Βλ. [Κεφ.1.8.2](#)), ο αλγόριθμος KNN βασίζεται στην απόσταση των προγνωστικών μεταβλητών από την πλησιέστερη ομάδα εκπαίδευσης (Appelhans et al., 2015). Αυτό σημαίνει ότι πιθανώς η χαμηλή χωρική ανάλυση παρείχε κάποιες λανθασμένες ή μεροληπτικές τιμές γειτόνων. Γενικά, ο αλγόριθμος KNN είναι πολύ ευαίσθητος στην επιλογή της τιμής k: η αύξηση της k μειώνει τη διακύμανση, αλλά μπορεί να αυξήσει την μεροληψία (Gonzalez-Sanchez et al., 2014). Επίσης, αξίζει να αναφερθεί ότι το μοντέλο BR παρουσίασε τη χειρότερη απόδοση, παρόλο που ο αλγόριθμος θεωρείται ότι αντλεί επιπλέον ισχύ σε σύγκριση με άλλες τεχνικές ML (Chen et al., 2022).

Σε αντίθεση με πολλές από τις σχετικές έρευνες που έχουν γίνει, η προσέγγιση της εργασίας αυτής ακολουθεί μία απλή ροή εργασιών, καθώς αξιοποιεί τα πρωτογενή, δηλαδή μη επεξεργασμένα δεδομένα ανακλαστικότητας Level2A, χωρίς να μεσολαβεί το δευτερογενές επίπεδο των δεικτών βλάστησης. Επίσης, δεν γίνεται χρήση πρόσθετων βιοφυσικών ή μετεωρολογικών παραμέτρων. Προκειμένου να προσδιοριστεί η καλύτερη περίοδος για την ακριβή πρόβλεψη της απόδοσης και να διερευνηθεί η δυνατότητα πρώιμων εκτιμήσεών της, οι 11 εικόνες της συνολικής περιόδου χωρίστηκαν σε μικρότερες υποπεριόδους με βάση την ανάπτυξη των καλλιεργειών, μέχρι μεμονωμένες ημερομηνίες. Ένα πλεονέκτημα των δεδομένων που λήφθηκαν είναι η διαθεσιμότητα πολλών εικόνων του δορυφόρου χωρίς σύννεφα στην ευρύτερη περιοχή της Νότιας Ευρώπης. Έτσι, κατέστη δυνατή η βελτιστοποίηση των μοντέλων σε μικρότερα χρονικά πλαίσια, ακόμη και στις 5 ημέρες που είναι η συχνότητα επανεπίσκεψης του δορυφόρου.

Τα αποτελέσματά υποδεικνύουν ότι ακόμη και με την χρήση ελάχιστων εικόνων, από την αρχή έως τα μέσα της περιόδου ανάπτυξης, η ακρίβεια της πρόβλεψης παραμένει πολύ υψηλή. Το χαμηλότερο RMSE για το μοντέλο RF ήταν 347 kg ha^{-1} όταν αξιοποιήθηκε ολόκληρη η περίοδος ανάπτυξης και αυξήθηκε σε 419 kg ha^{-1} για την περίοδο σποράς έως τα τέλη Φεβρουαρίου, ένα σφάλμα που εξακολουθεί να είναι χαμηλό, και κατ' επέκταση αποδεκτό.

Ο προσδιορισμός της βέλτιστης περιόδου για την πρόβλεψη της απόδοσης έχει ερευνηθεί και από άλλες μελέτες. Για παράδειγμα, οι Hunt et al. (2019) διαπίστωσαν ότι η ακρίβεια της εκτίμησης αυξάνεται σημαντικά όταν παρέχονται πρόσθετα τηλεπισκοπικά δεδομένα από τον Δεκέμβριο έως τον Ιούνιο, όμως τα ευρήματά τους βασίζονται μόνο σε τρεις διαθέσιμες εικόνες χωρίς σύννεφα στη μελέτη τους που διεξήχθη στο Ηνωμένο Βασίλειο. Στην μελέτη των Han et al. (2020) φαίνεται ότι τα μοντέλα ML που βασίζονται σε τηλεπισκοπικά δεδομένα του δορυφόρου MODIS μπορούν να προβλέψουν με ακρίβεια την απόδοση 1 – 2 μήνες πριν από την συγκομιδή στην Κίνα.

Τα αποτελέσματα της παρούσας εργασίας είναι άκρως ενθαρρυντικά, καθώς διερευνούν λεπτομερώς ολόκληρη την περίοδο καλλιέργειας και καταδεικνύουν σαφώς ότι είναι εφικτό να γίνει ακριβής πρόβλεψη της απόδοσης του σκληρού σίτου σε περιοχές με παρόμοια χαρακτηριστικά με αυτά της μελετηθείσας. Τέτοιου είδους πληροφορίες μπορούν να συμβάλλουν στην λήψη σημαντικών

διαχειριστικών αποφάσεων, καθώς επίσης και σε εφαρμογές της γεωργίας ακριβείας. Ακόμη, η χρήση των δορυφορικών δεδομένων υψηλής ανάλυσης θα βοηθήσει στην εξέλιξη ήδη εφαρμόσιμων γεωργικών πρακτικών, όπως αυτήν της δημιουργίας χαρτών αζώτου (με τον σχεδιασμό ζωνών διαχείρισης), όπου η εφαρμογή της λίπανσης γίνεται λαμβάνοντας υπόψιν τα τηλεπισκοπικά δεδομένα (Guerrero & Mouazen, 2021).

Η βέλτιστη περίοδος πρόβλεψης της απόδοσης έχει ερευνηθεί και συσχετιστεί με τα αναπτυξιακά στάδια των φυτών σιταριού από διάφορες μελέτες. Σύμφωνα με την μελέτη των Uribeetxebarria et al. (2022), οι εικόνες του δορυφόρου Sentinel-2 ήταν επαρκείς για να γίνει οριοθέτηση των ζωνών διαχείρισης μετά το αναπτυξιακό στάδιο 30 της κλίμακας Zadoks, επομένως ήταν χρήσιμες για την παραγωγή χαρτών λίπανσης της επερχόμενης περιόδου. Στην μελέτη των Stettmer et al. (2022) παρατηρήθηκε ότι τα δορυφορικά δεδομένα είναι περισσότερο αντιπροσωπευτικά της πρόσληψης αζώτου στα αναπτυξιακά στάδια 39 και 55 της κλίμακας BBCH (η κλίμακα BBCH είναι η ίδια με την κλίμακα Zadoks). Όπως έχει ήδη επισημανθεί στην βιβλιογραφική ανασκόπηση, στην κλίμακα Zadoks, τα αναπτυξιακά στάδια από το 30 έως το 39 περιγράφουν το στάδιο του καλαμώματος του φυτού. Αυτό ξεκινάει με την αύξηση του ψευδοβλαστού και ολοκληρώνεται με την πλήρη ανάπτυξη του τελευταίου φύλλου του φυτού (φύλλο σημαία) (Μπιλάλης, 2019). Κατά τη διάρκεια αυτών των σταδίων γίνεται συνήθως η εφαρμογή του αζώτου μέσω της επιφανειακής λίπανσης στην Νότια Ευρώπη, τα οποία αντιστοιχούν με τα τέλη Φεβρουαρίου έως τις αρχές Απριλίου. Όπως είναι κατανοητό, μια εκτίμηση της απόδοσης με την χρήση αυτών των πρωτογενών δεδομένων, που παρέχεται εγκαίρως, έχει την δυνατότητα να αξιοποιηθεί με πολλαπλούς τρόπους στην γεωργία ακριβείας.

Η παρούσα εργασία παρείχε σημαντικές πληροφορίες για τις μεθόδους μηχανικής μάθησης, που φαίνονται ελπιδοφόρες για την εκτίμηση ή την πρόβλεψη της απόδοσης. Ωστόσο, ένα κοινό μειονέκτημα που παρουσιάζει με προηγούμενες μελέτες, είναι το μικρό εύρος εφαρμοσιμότητας της. Τα μοντέλα που κατασκευάζονται με αυτόν τον τρόπο χρήζουν τόσο χωρικής, όσο και χρονικής επέκτασης, έτσι ώστε να μπορέσουν να εφαρμοστούν σε διαφορετικές περιπτώσεις καλλιεργειών και κλιματικών συνθηκών. Όλα τα χωράφια που χρησιμοποιήθηκαν στην συγκεκριμένη μελέτη βρίσκονται στην ίδια περιοχή, σε κοντινή απόσταση,

με μέγιστη απόσταση μεταξύ τους περίπου τα 30 χλμ., επομένως δεν υφίστανται μετεωρολογικές διακυμάνσεις μεταξύ των χωραφιών.

Επομένως, είναι σημαντικό να εξεταστεί η απόδοση προσεγγίσεων που λαμβάνουν υπόψιν περισσότερες παραμέτρους. Πέρα από τα δορυφορικά, φασματικά δεδομένα, μπορεί να γίνει αξιοποίηση δεδομένων από drones εξοπλισμένα με φασματικές κάμερες, τα οποία είναι συνήθως υψηλότερης χωρικής ανάλυσης και μπορούν να παρέχουν περισσότερες λεπτομέρειες για μια καλλιέργεια σε επίπεδο αγρού (Bian et al., 2022; Li et al., 2021; Panday et al., 2020; Xu et al., 2021). Ακόμη, όπως έχει γίνει ήδη σε προηγούμενες μελέτες, μπορούν να ενσωματωθούν κλιματικά δεδομένα (π.χ. θερμοκρασία, βροχοπτώσεις, ένταση ακτινοβολίας), δεδομένα για την κατάσταση του εδάφους (αλατότητα, υδραυλική αγωγιμότητα, περιεκτικότητα σε θρεπτικά συστατικά), καθώς και διάφορες άλλοι παράμετροι, όπως ο δείκτης φυλλικής επιφάνειας (Leaf Area Index, LAI) κ.α. Τέλος, για να δημιουργηθούν μοντέλα που είναι εφαρμόσιμα σε μεγαλύτερο εύρος περιπτώσεων, πρέπει να αξιοποιηθεί πληθώρα δεδομένων (μεγαλύτερος αριθμός χωραφιών, περισσότερες καλλιεργητικές περίοδοι). Έτσι θα διερευνηθούν καλύτερα οι δυνατότητες και τα όρια των αλγορίθμων μηχανικής μάθησης.

5. Συμπεράσματα

Στην παρούσα εργασία, έγινε χρήση της μηχανικής μάθησης ως μια καινοτόμος προσέγγιση για την εκτίμηση της απόδοσης σε χωράφια σκληρού σίτου. Τα δεδομένα που χρησιμοποιήθηκαν αποτελούν τηλεπισκοπικά δεδομένα, και συγκεκριμένα τα δορυφορικά δεδομένα του δορυφόρου Sentinel-2. Με την χρήση τριών διαφορετικών αλγορίθμων μηχανικής μάθησης, δημιουργήθηκαν μοντέλα πρόβλεψης. Η εργασία αυτή παραθέτει μια σύγκριση σε σχέση με προηγούμενη μελέτη των Cavalaris et al. (2021), όπου το μοντέλο βασίζεται σε δείκτες βλάστησης και χρησιμοποιεί την πολλαπλή γραμμική παλινδρόμηση για να εκτιμήσει την απόδοση χωραφιών σκληρού σίτου (VI-MLR). Τα αποτελέσματα της μοντελοποίησης εξετάστηκαν με βάση τα δεδομένα απόδοσης που συλλέχθηκαν από μια θεριζοαλωνιστική μηχανή εξοπλισμένη με ένα σύστημα χαρτογράφησης της απόδοσης.

Κατέστη σαφές πως όλες οι προσεγγίσεις της μηχανικής μάθησης έδειξαν βελτιωμένη ακρίβεια στην εκτίμηση της απόδοσης, σε σύγκριση με την προσέγγιση VI-MLR. Η πρώτη προσέγγιση με την χρήση δεικτών βλάστησης είχε μέτρια ακρίβεια, με $R^2 = 0,532$ και $RMSE = 847 \text{ kg ha}^{-1}$. Κατά την προσέγγιση με τη χρήση μηχανικής μάθησης, ο αλγόριθμος BR, παρουσίασε συγκριτικά την μικρότερη ακρίβεια, αν και η απόδοσή του ήταν καλύτερη από το μοντέλο VI-MLR, με $R^2 = 0,723$, $RMSE 668 \text{ kg ha}^{-1}$ και κλίση 0,622 όταν χρησιμοποιήθηκαν όλες οι εικόνες. Το μέγιστο της ακρίβειας παρατηρήθηκε όταν χρησιμοποιήθηκαν οι αλγόριθμοι RF και KNN με το σύνολο των εικόνων των καλλιεργητικών περιόδων ($R^2 > 0,91$, $RMSE < 360 \text{ kg ha}^{-1}$). Ακόμη, είναι σημαντικό το γεγονός πως η ακρίβεια αυτών των δύο αλγορίθμων παρέμεινε υψηλή, ακόμα και όταν χρησιμοποιήθηκαν δεδομένα μέχρι τον Ιανουάριο για κάθε καλλιεργητική περίοδο, δηλαδή πέντε μήνες πριν τη συγκομιδή της καλλιέργειας (ο RF είχε $R^2 = 0.924$ και $RMSE = 353$, ο KNN είχε $R^2 = 0.923$ και $RMSE = 345$).

Αυτό υποδεικνύει πως η τεχνική της ML μπορεί να συνδράμει στην έγκαιρη και ακριβή πρόβλεψη της απόδοσης, καθιστώντας την ένα πολύτιμο εργαλείο στην γεωργία ακριβείας.

Βιβλιογραφία

Ελληνική

1. Jensen, J. R. (2007). *Τηλεπισκόπηση Περιβάλλοντος, Μια προοπτική για τα φυσικά διαθέσιμα*. Αθήνα: Πανεπιστημιακές Εκδόσεις Ε.Μ.Π.
2. Βαϊόπουλος , Δ., Βασιλόπουλος , Α., & Ευελπίδου , Ν. (2008). *Γεωγραφικά Συστήματα Πληροφοριών από την θεωρία στην πράξη*. Αθήνα: Συμμετρία.
3. Γεωργούλη, Κ. (2015). *Τεχνητή Νοημοσύνη, Μια Εισαγωγική Προσέγγιση*. Αθήνα: ΣΥΝΔΕΣΜΟΣ ΕΛΛΗΝΙΚΩΝ ΑΚΑΔΗΜΑΪΚΩΝ ΒΙΒΛΙΟΘΗΚΩΝ.
4. Ευελπίδου, Ν., & Αντωνίου , Β. (2015). *Γεωγραφικά Συστήματα Πληροφοριών*. Αθήνα: Κάλλιπος.
5. Καρτάλης, Κ., & Φεΐδας , Χ. (2013). *Αρχές και Εφαρμογές Δορυφορικής Τηλεπισκόπησης*. Θεσσαλονίκη: Εκδόσεις Τζίολα.
6. Μεγγούδη, Σ. (2020, Μάιος). *Μετρήσεις αναπτυξιακών παραμέτρων και υδατικής κατάστασης σε καλλιέργεια σίτου για την ανάπτυξη μοντέλου πρόβλεψης της παραγωγής με τη χρήση δορυφορικών εικόνων Sentinel-2*. Βόλος: Πανεπιστήμιο Θεσσαλίας.
7. Μπιλάλης, Δ. (2019). *Γεωργία, Φυτά μεγάλης καλλιέργειας*. Αθήνα: Εκδόσεις Πεδίο.
8. Παπακώστα-Τασοπούλου, Δ. (2012). *Ειδική Γεωργία, Σιτηρά και Ψυχανθή*. Θεσσαλονίκη: Εκδόσεις Σύγχρονη Παιδεία.
9. Περάκης , Κ., Φαρασλής , Ι., & Μωυσιάδης, Α. (2015). *Η Τηλεπισκόπηση σε 13 Ενότητες*. Αθήνα: ΣΥΝΔΕΣΜΟΣ ΕΛΛΗΝΙΚΩΝ ΑΚΑΔΗΜΑΪΚΩΝ ΒΙΒΛΙΟΘΗΚΩΝ.
10. Τσιάρα, Α. (2012). *Ταξινόμηση Εικόνων με Τυχαία Δάση*. <https://olympias.lib.uoi.gr/jspuibitstream/123456789/6766/1/M.E.Αγγελική%20Τσιάρα.pdf>

Ξενογλώσση

1. Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning*, 6(1), 37–66. <https://doi.org/10.1007/BF00153759>
2. Akhter, R., & Sofi, S. A. (2021). Precision agriculture using IoT data analytics and machine learning. *Journal of King Saud University - Computer and Information Sciences*, S1319157821001282. <https://doi.org/10.1016/j.jksuci.2021.05.013>
3. Ali, I., Greifeneder, F., Stamenkovic, J., Neumann, M., & Notarnicola, C. (2015). Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data. *Remote Sensing*, 7(12), 16398–16421. <https://doi.org/10.3390/rs71215841>
4. Ali, U., Esau, T. J., Farooque, A. A., Zaman, Q. U., Abbas, F., & Bilodeau, M. F. (2022). Limiting the Collection of Ground Truth Data for Land Use and Land Cover Maps with Machine Learning Algorithms. *ISPRS International Journal of Geo-Information*, 11(6), 333. <https://doi.org/10.3390/ijgi11060333>
5. Appelhans, T., Mwangomo, E., Hardy, D. R., Hemp, A., & Nauss, T. (2015). Evaluating machine learning approaches for the interpolation of monthly air temperature at Mt. Kilimanjaro, Tanzania. *Spatial Statistics*, 14, 91–113. <https://doi.org/10.1016/j.spasta.2015.05.008>
6. Basso, B., Cammarano, D., & Carfagna, E. (2013). Review of Crop Yield Forecasting Methods and Early Warning Systems. FAO Headquarters, Rome, Italy, Vol. 18, p. 19(Proceedings of the first meeting of the scientific advisory committee of the global strategy to improve agricultural and rural statistics).
7. Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors*, 21(11), 3758. <https://doi.org/10.3390/s21113758>
8. Bian, C., Shi, H., Wu, S., Zhang, K., Wei, M., Zhao, Y., Sun, Y., Zhuang, H., Zhang, X., & Chen, S. (2022). Prediction of Field-Scale Wheat Yield Using Machine Learning Method and Multi-Spectral UAV Data. *Remote Sensing*, 14(6), 1474. <https://doi.org/10.3390/rs14061474>

9. Boissard, P., Guérif, M., Pointel, J.-G., & Guinot, J.-P. (1989). Application of SPOT data to wheat yield estimation. *Advances in Space Research*, 9(1), 143–154. [https://doi.org/10.1016/0273-1177\(89\)90479-1](https://doi.org/10.1016/0273-1177(89)90479-1)
10. Breiman, L. (2001). Random Forests. *Machine Learning*. <https://doi.org/doi:10.1023/a:1010933404324>
11. Buyya, R., & Dastjerdi, A. V. (Eds.). (2016). *Internet of Things: Principles and paradigms*. Morgan Kaufmann.
12. Cavalaris, C., Megoudi, S., Maxouri, M., Anatolitis, K., Sifakis, M., Levizou, E., & Kyparissis, A. (2021). Modeling of Durum Wheat Yield Based on Sentinel-2 Imagery. *Agronomy*, 11(8), 1486. <https://doi.org/10.3390/agronomy11081486>
13. Chen, S., Liu, W., Feng, P., Ye, T., Ma, Y., & Zhang, Z. (2022). Improving Spatial Disaggregation of Crop Yield by Incorporating Machine Learning with Multisource Data: A Case Study of Chinese Maize Yield. *Remote Sensing*, 14(10), 2340. <https://doi.org/10.3390/rs14102340>
14. Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, 151, 61–69. <https://doi.org/10.1016/j.compag.2018.05.012>
15. Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., & Campbell, J. P. (2020). Introduction to Machine Learning, Neural Networks, and Deep Learning. *Neural Networks, Trans Vis Sci Tech*, 12. <https://doi.org/10.1167/tvst.9.2.14>
16. Claverie, M., Demarez, V., Duchemin, B., Hagolle, O., Ducrot, D., Marais-Sicre, C., Dejoux, J.-F., Huc, M., Keravec, P., Béziat, P., Fieuzal, R., Ceschia, E., & Dedieu, G. (2012). Maize and sunflower biomass estimation in southwest France using high spatial and temporal resolution remote sensing data. *Remote Sensing of Environment*, 124, 844–857. <https://doi.org/10.1016/j.rse.2012.04.005>
17. Colwell, R. N. (1997). History and place of photographic interpretation (W.R. Phillipson εκδ., Τόμ. Manual of photographic interpretation).
18. Duchemin, B., Maisongrande, P., Boulet, G., & Benhadj, I. (2008). A simple algorithm for yield estimates: Evaluation for semi-arid irrigated winter wheat

- monitored with green leaf area index. *Environmental Modelling & Software*, 23(7), 876–892. <https://doi.org/10.1016/j.envsoft.2007.10.003>
19. European Commission. Joint Research Centre. Institute for Environment and Sustainability. (2014). Phenology-based land cover classification using Landsat 8 time series. Publications Office. <https://data.europa.eu/doi/10.2788/15561>
 20. Farin, G. (2002). Linear Interpolation. 18. <https://doi.org/10.1016/B978-155860737-8/50003-X>
 21. Fowler, D. B. (2018). Growth Stages of Wheat. Ducks Unlimited Canada and Conservation Production Systems Limited.
 22. Gebbers, R., & Adamchuk, V. I. (2010). Precision Agriculture and Food Security. *Science*, 327(5967), 828–831. <https://doi.org/10.1126/science.1183899>
 23. Gonzalez-Sanchez, A., Frausto-Solis, J., & Ojeda-Bustamante, W. (2014). Predictive ability of machine learning methods for massive crop yield prediction. *Spanish Journal of Agricultural Research*, 12(2), 313. <https://doi.org/10.5424/sjar/2014122-4439>
 24. Griffin, T. W., & Lowenberg - DeBoer, J. (2005). Worldwide adoption and profitability of precision agriculture Implications for Brazil. *Reviste de Politica Agricola*, 20–37.
 25. Guerrero, A., & Mouazen, A. M. (2021). Evaluation of variable rate nitrogen fertilization scenarios in cereal crops from economic, environmental and technical perspective. *Soil and Tillage Research*, 213, 105110. <https://doi.org/10.1016/j.still.2021.105110>
 26. Han, J., Zhang, Z., Cao, J., Luo, Y., Zhang, L., Li, Z., & Zhang, J. (2020). Prediction of Winter Wheat Yield Based on Multi-Source Data and Machine Learning in China. *Remote Sensing*, 12(2), 236. <https://doi.org/10.3390/rs12020236>
 27. Hunt, M. L., Blackburn, G. A., Carrasco, L., Redhead, J. W., & Rowland, C. S. (2019). High resolution wheat yield mapping using Sentinel-2. *Remote Sensing of Environment*, 233, 111410. <https://doi.org/10.1016/j.rse.2019.111410>
 28. Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. 27.

29. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 103). Springer New York. <https://doi.org/10.1007/978-1-4614-7138-7>
30. JASP—A Fresh Way to Do Statistics Ie. (2022, Μάιος 22). Ανάκτηση από <https://jasp-stats.org>
31. Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K.-M., Gerber, J. S., Reddy, V. R., & Kim, S.-H. (2016). Random Forests for Global and Regional Crop Yield Predictions. *PLOS ONE*, 11(6), e0156571. <https://doi.org/10.1371/journal.pone.0156571>
32. Jia, P., Zhang, J., He, W., Hu, Y., Zeng, R., Zamanian, K., Jia, K., & Zhao, X. (2022). Combination of Hyperspectral and Machine Learning to Invert Soil Electrical Conductivity. *Remote Sensing*, 14(11), 2602. <https://doi.org/10.3390/rs14112602>
33. Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260. <https://doi.org/10.1126/science.aaa8415>
34. Li, K.-Y., Burnside, N. G., Sampaio de Lima, R., Villoslada Peciña, M., Sepp, K., Yang, M.-D., Raet, J., Vain, A., Selge, A., & Sepp, K. (2021). The Application of an Unmanned Aerial System and Machine Learning Techniques for Red Clover-Grass Mixture Yield Estimation under Variety Performance Trials. *Remote Sensing*, 13(10), 1994. <https://doi.org/10.3390/rs13101994>
35. Lopez-Arevalo, I., Aldana-Bobadilla, E., Molina-Villegas, A., Galeana-Zapién, H., Muñoz-Sanchez, V., & Gausin-Valle, S. (2020). A Memory-Efficient Encoding Method for Processing Mixed-Type Data on Machine Learning. *Entropy*, 22(12), 1391. <https://doi.org/10.3390/e22121391>
36. Malenovský, Z., Rott, H., Cihlar, J., Schaepman, M. E., García-Santos, G., Fernandes, R., & Berger, M. (2012). Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sensing of Environment*, 120, 91–101. <https://doi.org/10.1016/j.rse.2011.09.026>
37. Mesev, V. (2007). *Integration of GIS and Remote Sensing*. John Wiley & Sons, Ltd.
38. My John Deere. (2022, Μάιος 22). Ανάκτηση από <https://myjohndeere.deere.com/>

39. Open Access Hub. (2022, Μαΐου 22). Ανάκτηση από <https://scihub.copernicus.eu/>
40. Pallathadka, H., Mustafa, M., Sanchez, D. T., Sekhar Sajja, G., Gour, S., & Naved, M. (2021). Impact of Machine learning on Management, healthcare and agriculture. *Materials Today: Proceedings*, S221478532104894X. <https://doi.org/10.1016/j.matpr.2021.07.042>
41. Panday, U. S., Pratihast, A. K., Aryal, J., & Kayastha, R. B. (2020). A Review on Drone-Based Data Solutions for Cereal Crops. *Drones*, 4(3), 41. <https://doi.org/10.3390/drones4030041>
42. Pesaresi, M., Corbane, C., Julea, A., Florczyk, A., Syrris, V., & Soille, P. (2016). Assessment of the Added-Value of Sentinel-2 for Detecting Built-up Areas. *Remote Sensing*, 8(4), 299. <https://doi.org/10.3390/rs8040299>
43. Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V. R., Murayama, Y., & Ranagalage, M. (2020). Sentinel-2 Data for Land Cover/Use Mapping: A Review. 35.
44. Pollock, R. B., & Kanemasu, E. T. (1979). Estimating leaf-Area Index of Wheat with LANDSAT Data. *REMOTE SENSING OF ENVIRONMENT*, 6.
45. Porter, J. R. (1993). AFRCWHEAT2: A model of the growth and development of wheat incorporating responses to water and nitrogen. *European Journal of Agronomy*, 2(2), 69–82. [https://doi.org/10.1016/S1161-0301\(14\)80136-6](https://doi.org/10.1016/S1161-0301(14)80136-6)
46. Porter, J. R., Jamieson, P. D., & Wilson, D. R. (1993). Comparison of the wheat simulation models Afrcwheat2, Ceres-wheat and Swheat for non-limiting conditions of crop growth. *Field Crops Research*, 33(1–2), 131–157. [https://doi.org/10.1016/0378-4290\(93\)90098-8](https://doi.org/10.1016/0378-4290(93)90098-8)
47. QGIS Geographic Information System. (2022, Μάιος 21). Ανάκτηση από QGIS.org, 2022: <http://www.qgis.org>
48. Rhee, J., & Im, J. (2017). Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data. *Agricultural and Forest Meteorology*, 237–238, 105–122. <https://doi.org/10.1016/j.agrformet.2017.02.011>
49. Richie, J. (1985). Description and performance of CERES wheat: A user-oriented wheat yield model. *ARS wheat yield project*, σσ. 159-175.

50. Robert, P. C. (2002). Precision agriculture: A challenge for crop nutrition management. Kluwer Academic Publishers, *Plant and Soil*, 143–149. https://doi.org/doi:10.1007/978-94-017-2789-1_11
51. Roy, P. S., Behera, M. D., & Srivastav, S. K. (2017). Satellite Remote Sensing: Sensors, Applications and Techniques. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, 87(4), 465–472. <https://doi.org/10.1007/s40010-017-0428-8>
52. Royo, C., Soriano, J. M., & Alvaro, F. (2017). Wheat: A Crop in the Bottom of the Mediterranean Diet Pyramid. In B. Fuerst-Bjelic (Ed.), *Mediterranean Identities—Environment, Society, Culture*. InTech. <https://doi.org/10.5772/intechopen.69184>
53. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
54. Segarra, J., Araus, J. L., & Kefauver, S. C. (2022). Farming and Earth Observation: Sentinel-2 data to estimate within-field wheat grain yield. *International Journal of Applied Earth Observation and Geoinformation*, 107, 102697. <https://doi.org/10.1016/j.jag.2022.102697>
55. Sharma, A., Jain, A., Gupta, P., & Chowdary, V. (2021). Machine Learning Applications for Precision Agriculture: A Comprehensive Review. *IEEE Access*, 9, 4843–4873. <https://doi.org/10.1109/ACCESS.2020.3048415>
56. Stafford, J. V. (2000). Implementing Precision Agriculture in the 21st Century. *Journal of Agricultural Engineering Research*, 76(3), 267–275. <https://doi.org/10.1006/jaer.2000.0577>
57. STEP—Science Toolbox Exploitation Platform. (2022, Μάιος 22). Ανάκτηση από <http://step.esa.int>
58. Stettmer, M., Maidl, F.-X., Schwarzensteiner, J., Hülsbergen, K.-J., & Bernhardt, H. (2022). Analysis of Nitrogen Uptake in Winter Wheat Using Sensor and Satellite Data for Site-Specific Fertilization. *Agronomy*, 12(6), 1455. <https://doi.org/10.3390/agronomy12061455>
59. Stöckle, C. O., Donatelli, M., & Nelson, R. (2003). CropSyst, a cropping systems simulation model. *European Journal of Agronomy*, 18(3–4), 289–307. [https://doi.org/10.1016/S1161-0301\(02\)00109-0](https://doi.org/10.1016/S1161-0301(02)00109-0)

60. Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1), 25. <https://doi.org/10.1186/1471-2105-8-25>
61. Tottman, D. R. (1987). The decimal code for the growth stages of cereals, with illustrations. *Annals of Applied Biology*, 110(2), 441–454. <https://doi.org/10.1111/j.1744-7348.1987.tb03275.x>
62. Tottman, D. R., Makepeace, R. J., & Broad, H. (1979). An explanation of the decimal code for the growth stages of cereals, with illustrations. *Annals of Applied Biology*, 93(2), 221–234. <https://doi.org/10.1111/j.1744-7348.1979.tb06534.x>
63. The European Space Agency. (2022, 09 14). Sentinel 2. Ανάκτηση από <https://sentinel.esa.int/web/sentinel/missions/sentinel-2>
64. The European Space Agency. (2022, 09 14). Sentinel Overview. Ανάκτηση από Sentinel Missions: <https://sentinel.esa.int/web/sentinel/missions>
65. Trombetta, A., Iacobellis, V., Tarantino, E., & Gentile, F. (2016). Calibration of the AquaCrop model for winter wheat using MODIS LAI images. *Agricultural Water Management*, 164, 304–316. <https://doi.org/10.1016/j.agwat.2015.10.013>
66. Uribeetxebarria, A., Castellón, A., & Aizpurua, A. (2022). A First Approach to Determine If It Is Possible to Delineate In-Season N Fertilization Maps for Wheat Using NDVI Derived from Sentinel-2. *Remote Sensing*, 14(12), 2872. <https://doi.org/10.3390/rs14122872>
67. Van Evert, F. K., & Campbell, G. S. (1994). CropSyst: A Collection of Object-Oriented Simulation Models of Agricultural Systems. *Agronomy Journal*, 86(2), 325–331. <https://doi.org/10.2134/agronj1994.00021962008600020022x>
68. Vieira, S., Lopez Pinaya, W. H., & Mechelli, A. (2020). Introduction to machine learning. In *Machine Learning* (pp. 1–20). Elsevier. <https://doi.org/10.1016/B978-0-12-815739-8.00001-8>
69. Vincenzi, S., Zucchetto, M., Franzoi, P., Pellizzato, M., Pranovi, F., De Leo, G. A., & Torricelli, P. (2011). Application of a Random Forest algorithm to predict spatial distribution of the potential yield of *Ruditapes philippinarum* in the Venice lagoon, Italy. *Ecological Modelling*, 222(8), 1471–1478. <https://doi.org/10.1016/j.ecolmodel.2011.02.007>

70. Wang, L., Zhou, X., Zhu, X., Dong, Z., & Guo, W. (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *The Crop Journal*, 4(3), 212–219. <https://doi.org/10.1016/j.cj.2016.01.008>
71. Xu, W., Chen, P., Zhan, Y., Chen, S., Zhang, L., & Lan, Y. (2021). Cotton yield estimation model based on machine learning using time series UAV remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 104, 102511. <https://doi.org/10.1016/j.jag.2021.102511>
72. Zadoks, J. C., Chang, T. T., & Konzak, C. F. (1974). A decimal code for the growth stages of cereals. *Weed Research*, 14(6), 415–421. <https://doi.org/10.1111/j.1365-3180.1974.tb01084.x>
73. Zhang, S. (2020). Cost-sensitive KNN classification. *Neurocomputing*, 391, 234–242. <https://doi.org/10.1016/j.neucom.2018.11.101>
74. Zou, X., Zhu, S., & Möttus, M. (2022). Estimation of Canopy Structure of Field Crops Using Sentinel-2 Bands with Vegetation Indices and Machine Learning Algorithms. *Remote Sensing*, 14(12), 2849. <https://doi.org/10.3390/rs14122849>

Κατάλογος πινάκων

| | |
|--|----|
| Πίνακας 1 Φασματικές ζώνες του δορυφόρου Sentinel-2..... | 14 |
| Πίνακας 2 Πληροφορίες των χωραφιών που μελετήθηκαν..... | 22 |
| Πίνακας 3. Σύγκριση αποδοτικότητας των τριών αλγορίθμων ML..... | 33 |

Πηγές εικόνων και σχημάτων

Εικόνα Εξωφύλλου: <https://publicdomainreview.org/collection/a-19th-century-vision-of-the-year-2000>

Εικόνα 1: Fowler, D. B. (2018). Growth Stages of Wheat. Ducks Unlimited Canada and Conservation Production Systems Limited.

Εικόνα 2: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2/satellite-description>

Εικόνα 3: Δημιουργήθηκε μέσω του λογισμικού: <https://www.qgis.org/en/site/>

Σχήμα 1: <https://www.fao.org/faostat/en/#data/QCL/visualize>

Σχήμα 2: <https://www.fao.org/faostat/en/#data/QCL/visualize>

Σχήμα 3: Sharma, A., Jain, A., Gupta, P., & Chowdary, V. (2021). Machine Learning Applications for Precision Agriculture: A Comprehensive Review. IEEE Access, 9, 4843–4873. <https://doi.org/10.1109/ACCESS.2020.3048415>