

Κεφάλαιο 2:

Ταξινόμηση Καλλιεργειών με τη χρήση Πολυχρονικών Εικόνων SENTINEL-2

Παρασκευή Μπουραντά Γεωγράφος, Μεταπτυχιακή φοιτήτρια Πανεπιστήμιο Αιγαίου Τμήμα Γεωγραφίας Email: geo16131@geo.aegean.gr	Χρήστος Βασιλάκος ΕΔΙΠ Πανεπιστήμιο Αιγαίου Τμήμα Γεωγραφίας Email: chvas@aegean.gr	Δημήτρης Καβρουδάκης Επίκουρος Καθηγητής Πανεπιστήμιο Αιγαίου Τμήμα Γεωγραφίας Email: dimitrisk@aegean.gr	Νίκος Σουλακέλλης Καθηγητής Πανεπιστήμιο Αιγαίου Τμήμα Γεωγραφίας Email: nsoul@aegean.gr
---	--	--	--

Περίληψη:

Η χαρτογράφηση χρήσης/κάλυψης γης είναι μια από τις πιο σημαντικές και βασικές εφαρμογές της τηλεπισκόπησης, καθώς αποτελεί σημαντική πηγή πληροφορίας για την υλοποίηση διαφόρων αναπτυξιακών έργων τόσο σε εθνικό αλλά όσο και σε τοπικό επίπεδο. Ιδιαίτερα στον τομέα της γεωργίας μέσω της ταξινόμησης δορυφορικών εικόνων είναι δυνατή καταγραφή με μεγάλη ακρίβεια των διαφόρων τύπων καλλιεργειών. Δεδομένου όμως του κύκλου ανάπτυξης και της φαινολογίας κατά τη διάρκεια αυτού, η ταξινόμηση των καλλιεργειών είναι αρκετές φορές δύσκολη όταν χρησιμοποιούνται πολυφασματικά δεδομένα μιας χρονικής στιγμής. Μεγαλύτερη δυσκολία εμφανίζεται στο διαχωρισμό των καλλιεργειών που έχουν παρόμοια χαρακτηριστικά σε συγκεκριμένες χρονικές στιγμές. Στόχος αυτής της εργασίας είναι η ταξινόμηση των καλλιεργειών με τη χρήση διαχρονικών πολυφασματικών δεδομένων Sentinel-2 στην περιοχή της Καλλονής της Ν. Λέσβου. Χρησιμοποιήθηκαν 12 δορυφορικές εικόνες από τον πολυφασματικό δέκτη Sentinel-2. Πραγματοποιήθηκαν επιβλεπόμενες ταξινομήσεις με τη μέθοδο Random Forest για διάφορα σετ δεδομένων. Σύμφωνα με τα αποτελέσματα η ακρίβεια της ταξινόμησης με τη χρήση διαχρονικών δεδομένων ήταν μεγαλύτερη σε σύγκριση με την ταξινόμηση σε εικόνες μίας χρονικής στιγμής. Επίσης η εφαρμογή συνδυαστικών μοντέλων βασισμένο σε ένα πολλαπλό σχήμα συναίνεσης παρουσίασε υψηλή ακρίβεια. Συμπερασματικά, τα δορυφορικά δεδομένα του Sentinel-2 μπορούν να συμβάλλουν στην απόκτηση αξιόπιστης πληροφορίας για τις γεωργικές περιοχές, λόγω της υψηλής τους διαχρονικής και χωρικής διακριτικής τους ικανότητας.

Λέξεις Κλειδιά: Τηλεπισκόπηση, Ταξινόμηση καλλιεργειών, Random Forest, Πολυχρονικά Δορυφορικά Δεδομένα

1. Εισαγωγή

Η δορυφορική τηλεπισκόπηση αποτελεί μια οικονομική πηγή δεδομένων χρήσεων και κάλυψης γης. Οι τεχνικές της δορυφορικής τηλεπισκόπησης χρησιμοποιούνται ευρέως στη γεωργία (Courault et al, 2016). Τέτοιου είδους τεχνικές είναι ο εντοπισμός ή η ταξινόμηση των καλλιεργειών, η απογραφή της έκτασης των καλλιεργειών, η εκτίμηση της παραγωγής τους, η ανίχνευση των ασθενειών τους, ο σχεδιασμός και η λειτουργία αρδευτικών έργων και η εκτίμηση των ζημιών από πλημμύρες, παγετό, χαλάζι, ξηρασία και από την εκδήλωση των πυρκαγιών (Καρτάλης και Φείδας, 2017). Η γεωργία έχει ένα ιδιαίτερο δυναμικό χαρακτήρα, λόγω της συνεχούς ανάπτυξης των φυτών σε συνδυασμό με τις μεταβολές που αυτά υφίστανται από το φυσικό περιβάλλον. Η γεωργική παραγωγή ακολουθεί ισχυρά εποχιακά πρότυπα που σχετίζονται με τον βιολογικό κύκλο ζωής των καλλιεργειών. Επίσης, η παραγωγή εξαρτάται από τον τύπο εδάφους, τις κλιματικές μεταβολές που αλλάζουν μέσα σε σύντομες χρονικές περιόδους και τις γεωργικές πρακτικές διαχείρισης που χρησιμοποιεί ο κάθε παραγωγός (Atzberger, 2016). Εν κατακλείδι, η διάκριση των τάξεων είναι απαιτητική όσον αφορά την βλάστηση. Πόσο μάλλον η διάκριση των καλλιεργητικών ειδών καθώς απαιτεί προσοχή αφού οι φασματικές αντανάκλασεις τους δεν διαφέρουν σημαντικά (Esch et al, 2014).

Για τον διαχωρισμό των ειδών βλάστησης χρησιμοποιούνται διάφορα φασματικά κριτήρια τα οποία βασίζονται κυρίως στην ανακλαστικότητα στο εγγύς υπέρυθρο κανάλι (NIR) καθώς και στους δείκτες βλάστησης NDVI, EVI κ.α. (Jiang et al, 2008). Σε αντίθεση με τα φασματικά χαρακτηριστικά, η διαφοροποίηση των γεωργικών κατηγοριών όπως οι βοσκότοποι με τις καλλιέργειες και τα δάση με τις φυτείες βασίζονται μόνο σε χαρακτηριστικά υφής και δομής σε εικόνες μιας μόνο χρονικής περιόδου είναι πολύ δύσκολη. Αν όμως, τα χαρακτηριστικά αυτά χρησιμοποιηθούν ως βοηθητικά σε μια ταξινόμηση, μπορεί να δοθεί ένα ικανοποιητικό αποτέλεσμα ταξινόμησης των τάξεων αυτών (Peña-Barragán et al. 2011). Πολλές φορές όμως, οι φασματικές υπογραφές διαφορετικών τύπων καλλιεργειών είναι αρκετά παρόμοιες μεταξύ τους σε ορισμένες περιόδους του έτους, ανάλογα με την κατάσταση ανάπτυξης, η οποία με τη σειρά της καθορίζεται από τη φαινολογία και τον τύπο της καλλιέργειας (Kussul et al, 2017). Επομένως, υπάρχουν ομοιότητες μεταξύ των διαφορετικών τύπων καλλιεργειών ανάλογα με τις φασματικές υπογραφές που παρουσιάζει ο κάθε τύπος γεωργικής καλλιέργειας της περιοχής (Tang Yi and Li Xinrong, 2016). Αυτό έχει ως αποτέλεσμα, η πληροφορία να μην ταξινομηθεί σωστά λόγω της ταύτισης των φασματικών χαρακτηριστικών ορισμένων ειδών καλλιέργειας, όπως για παράδειγμα τα κτηνοτροφικά φυτά, τα λοιπά σιτηρά και το σκληρό σίτος επειδή παρουσιάζουν παρόμοιες φασματικές υπογραφές και εντέλει να ομαδοποιούνται ως μία κατηγορία.

Για την βελτίωση των αποτελεσμάτων της ταξινόμησης των καλλιεργειών χρησιμοποιούνται διάφορες τεχνικές ταξινόμησης που εφαρμόζουν διάφορους αλγόριθμους μηχανικής μάθησης οι οποίοι είναι τα τεχνικά νευρωνικά δίκτυα (Artificial Neural Networks, ANN), τα δέντρα απόφασης (Decision trees, DT), τα Support Vector Machine (SVM) και τα τυχαία δάση (Random Forest, RF) (Khatami et al, 2016). Προηγούμενες έρευνες, επικεντρώνονται κυρίως στην σύγκριση των αποτελεσμάτων των ταξινομητών. Οι Inglada et al (2015), εξέτασαν πέντε διαφορετικούς αλγόριθμους επιβλεπόμενης ταξινόμησης για την αυτόματη παραγωγή χαρτών καλλιεργειών σε 12 περιοχές μελέτης. Συγκεκριμένα επιλέχθηκαν 4 περιοχές της Ευρώπης, 4 στην Αφρική, 2 στην Αμερική και 2 στην Ασία. Οι ταξινομητές που επιλέχθηκαν απαιτούσαν την συλλογή δεδομένων πεδίου για την εκπαίδευση τους ενώ η παραγωγή χαρτών έγινε αυτόματα. Οι πολυφασματικές εικόνες που χρησιμοποιήθηκαν προέρχονταν από τους δορυφόρους SPOT 4 και Landsat 8 και πραγματοποιήθηκε σύγκριση

μεταξύ των ταξινομητών SVM, Decision trees, GBT (Gradient Boosted Trees) και RF ως προς την ακρίβεια των αποτελεσμάτων τους. Σύμφωνα με τα αποτελέσματα, ο ταξινομητής RF είναι ο καταλληλότερος για ταξινομήσεις με χρονοσειρές, καθώς παρουσίασε συνολική ακρίβεια άνω των 80% για όλες τις περιοχές.

Ο ταξινομητής RF χρησιμοποιήθηκε και από τους Atzberger et al (2016), με δεδομένα Sentinel-2 για την αναγνώριση θερινών και χειμερινών καλλιεργειών σε μια περιοχή μελέτης ενώ σε μια δεύτερη αναγνωρίστηκαν επτά διαφορετικά είδη φυλλοβόλων και κωνοφόρων δέντρων. Αποδείχθηκε και στην συγκεκριμένη εργασία τα αποτελέσματα της ακρίβειας του ταξινομητή ήταν 65% για τα είδη δέντρων και 76% για τα είδη των καλλιεργειών. Παρόλο που είχαν καλά αποτελέσματα οι συγγραφείς διαπίστωσαν ότι με την χρήση πολυχρονικών δεδομένων Sentinel-2 τα αποτελέσματα του ταξινομητή θα ήταν πιο υψηλά σε σχέση με τα μονοχρονικά δεδομένα που χρησιμοποίησαν οι ίδιοι.

Πολυχρονικά δεδομένα Sentinel-2 χρησιμοποιήθηκαν και από τους Akbari et al (2020). Δημιούργησαν διάφορα σύνολα παραμέτρων εισόδου συμπεριλαμβανομένων των φασματικών ζωνών και διαφόρων δεικτών βλάστησης και άλλων δεικτών υφής, τα οποία χρησιμοποιήθηκαν για την ενίσχυση της χαρτογράφησης των διαφορετικών τύπων καλλιεργειών της περιοχής. Για την εξαγωγή χαρτών έγινε η χρήση του ταξινομητή RF. Η απόδοση του RF ήταν αποτελεσματική στην χαρτογράφηση από φασματικά δεδομένα και δεδομένα υφής. Όμως, οι συγγραφείς, εφάρμοσαν και άλλη μια μέθοδο ως προς την ακρίβεια των αποτελεσμάτων, αυτής της πλειοψηφίας (Majority) που βασίστηκε σε δεδομένα της Random Forest και συγκεκριμένα εφαρμόστηκε η διαδικασία RF classifier PSO (Random Forest classifier particle swarm optimization). Η εφαρμογή αυτή οδήγησε σε μια πιο ακριβή ταξινόμηση από ότι ο αλγόριθμος RF, σχεδόν για όλους τους τύπους καλλιεργείας. Οι συγγραφείς προτείνουν την μέθοδο RF classifier PSO για την διάκριση μεταξύ των ετερογενών καλλιεργειών π.χ. των λαχανικών λόγω της μείωσης των λαθών.

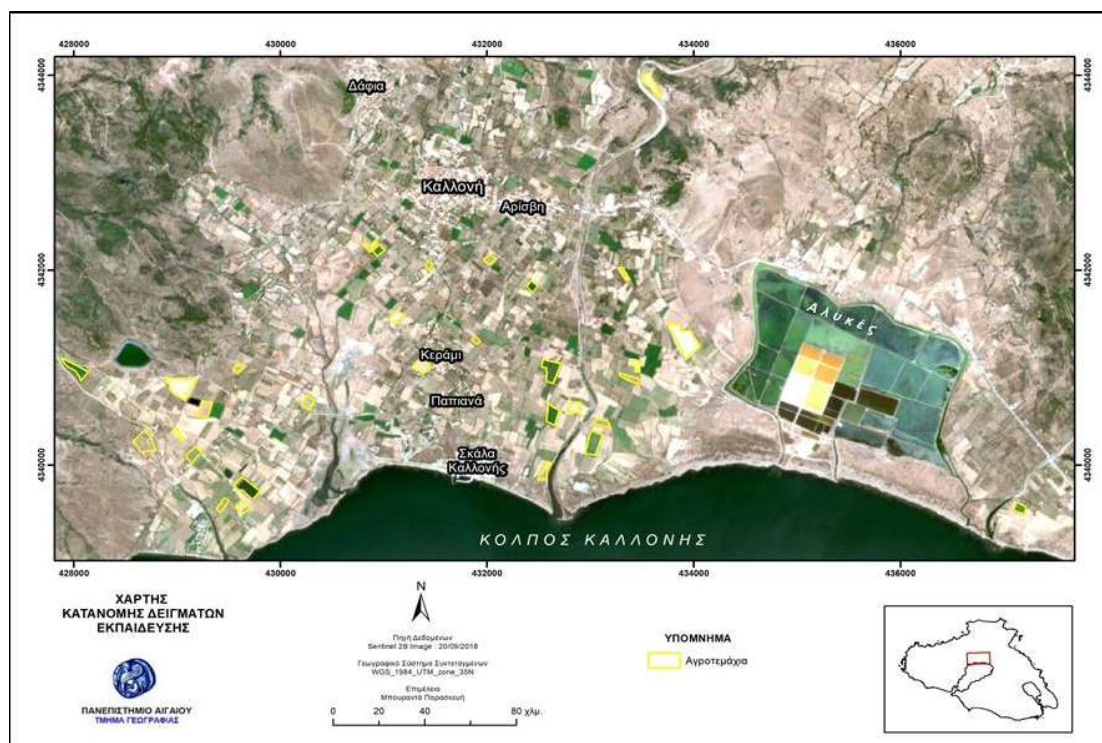
Στόχος της παρούσης εργασίας είναι η αναγνώριση των διαφόρων τύπων καλλιεργειών για μια καλλιεργητική χρονιά στην περιοχή της Καλλονής της Ν. Λέσβου. Σύμφωνα με τα παραπάνω, στην παρούσα εργασία χρησιμοποιήθηκαν πολυχρονικά δεδομένα του Sentinel-2 και συγκεκριμένα 12 μηνιαίες εικόνες για την επιβλεπόμενη ταξινόμηση των καλλιεργειών. Δημιουργήθηκαν διάφορα μοντέλα ταξινόμησης λαμβάνοντας υπόψη διαφορετικό αριθμό παραμέτρων εισόδου ενώ τέλος εφαρμόστηκε ένα συνδυαστικό μοντέλο βασισμένο σε ένα πολλαπλό σχήμα συναίνεσης. Η εκπαίδευση και αξιολόγηση βασίστηκε με δεδομένα πεδίου που συλλέχθηκαν από τοπικούς φορείς και συνεντεύξεις των αγροτών.

2. Δεδομένα και Μεθοδολογία

2.1 Περιοχή μελέτης

Η Λέσβος είναι το τρίτο νησί της Ελλάδας με έκταση 1672 τετραγωνικών χιλιομέτρων και ποικίλους γεωλογικούς σχηματισμούς, κλιματολογικές συνθήκες και τύπους βλάστησης (π.χ. δάση, θάμνους, λιβάδια και αγροτικές εκτάσεις). Το κλίμα στο νησί είναι μεσογειακό που σημαίνει ζεστά και ξηρά καλοκαίρια και ήπιους μέτρια βροχερούς χειμώνες (Vasilakos et al, 2007). Η Καλλονή βρίσκεται στο κέντρο του νησιού στη βόρεια πλευρά του ομώνυμου κόλπου και αποτελεί την πιο εύφορη πεδιάδα του νησιού (Σχήμα 1).

Σχήμα 1. Απεικόνιση της περιοχής μελέτης.



Οι καλλιέργειες από τις οποίες αποτελείται το νησί της Λέσβου είναι οι παραδοσιακοί ελαιώνες, ενώ στον κάμπο της Καλλονής κυριαρχούν ελιές, σιτάρι, κτηνοτροφικά φυτά, λοιπά σιτηρά, ενώ τα κηπευτικά περιορίζονται κυρίως στους καλοκαιρινούς μήνες.

2.2 Δεδομένα

Για την ταξινόμηση αποκτήθηκαν 12 πολυφασματικές εικόνες από τους δορυφορικούς δέκτες Sentinel-2A και Sentinel-2B για το χρονικό διάστημα Σεπτέμβριος 2018 – Αύγουστος 2019 (Πίνακας 1). Ο πολυφασματικός σαρωτής καλύπτει 13 φασματικές ζώνες (443-2.190nm) με χωρικές αναλύσεις των 10μ. (για 4 φασματικά κανάλια στο ορατό και το κοντινό υπέρυθρο), των 20μ. (για 6 φασματικά κανάλια στις μέσες υπέρυθρες ζώνες) και των 60μ. (3 κανάλια για ατμοσφαιρικές διορθώσεις). Για την παρούσα εργασία χρησιμοποιήθηκαν τα κανάλια B2 (Μπλε), B3 (Πράσινο), B4 (Κόκκινο) και B8 (Κοντινό Υπέρυθρο) χωρικής διακριτικής ικανότητας 10μ., τα οποία προσφέρουν πληροφορίες για το ορατό και το υπέρυθρο τμήμα του φάσματος καθώς και τα κανάλια B5, B6, B7, B8A, B11 και B12 που καλύπτουν το εύρος μεταξύ κοντινού και μικροκυματικού υπέρυθρου ενώ διαθέτουν χαμηλότερη χωρική διακριτική ικανότητα 20μ. Οι δορυφορικές εικόνες που χρησιμοποιήθηκαν ήταν σε επίπεδο επεξεργασίας Level-2 ήτοι γεωμετρικά και ραδιομετρικά διορθωμένες χωρίς ατμοσφαιρικές επιδράσεις συνεπώς οι τιμές των pixel εκφράζουν την ανακλαστικότητα του εδάφους.

Για την επιβλεπόμενη ταξινόμηση ήταν αναγκαία η χρήση δεδομένων εδάφους με γνωστό τον τύπο καλλιέργειας. Για να δημιουργηθούν τα αγροτεμάχια για τα οποία θα αποκτηθεί γνώση για την πραγματική καλλιέργεια την συγκεκριμένη περίοδο πραγματοποιήθηκε μη επιβλεπόμενη ταξινόμηση σε μία νέα εικόνα από την σύνθεση των καναλιών του κοντινού υπέρυθρου (B8) και των 12 εικόνων.

Πίνακας 1. Ημερομηνίες των δορυφορικών εικόνων.

Ημερομηνία	Δορυφόρος
20/09/2018	Sentinel-2B
10/10/2018	Sentinel-2B
09/11/2018	Sentinel-2B
27/12/2018	Sentinel-2A
16/01/2019	Sentinel-2A
22/02/2019	Sentinel-2A
24/03/2019	Sentinel-2A
03/04/2019	Sentinel-2A
28/05/2019	Sentinel-2B
27/06/2019	Sentinel-2B
22/07/2019	Sentinel-2A
31/08/2019	Sentinel-2A

Λαμβάνοντας υπόψη το αποτέλεσμα της ομαδοποίησης των pixel αναγνωρίστηκαν και ψηφιοποιήθηκαν 31 αγροτεμάχια με διαφορετική φασματική υπογραφή σε διάφορα σημεία της περιοχής. Για τα παραπάνω αγροτεμάχια αναζητήθηκε η πραγματική τους καλλιέργεια κατά την καλλιεργητική περίοδο 2018-2019 μέσω του Κέντρου Εξυπηρέτησης Αγροτών Λέσβου (ΚΕΑ Λέσβου), του Ολοκληρωμένου Συστήματος Διαχείρισης και Ελέγχου (ΟΣΔΕ) και συνεντεύξεων των αγροτών. Τα εικονοστοιχεία των παραπάνω αγροτεμαχίων αποτέλεσαν τα δεδομένα εκπαίδευσης και επαλήθευσης των μοντέλων τυχαίων δασών (RF) που αναπτύχθηκαν.

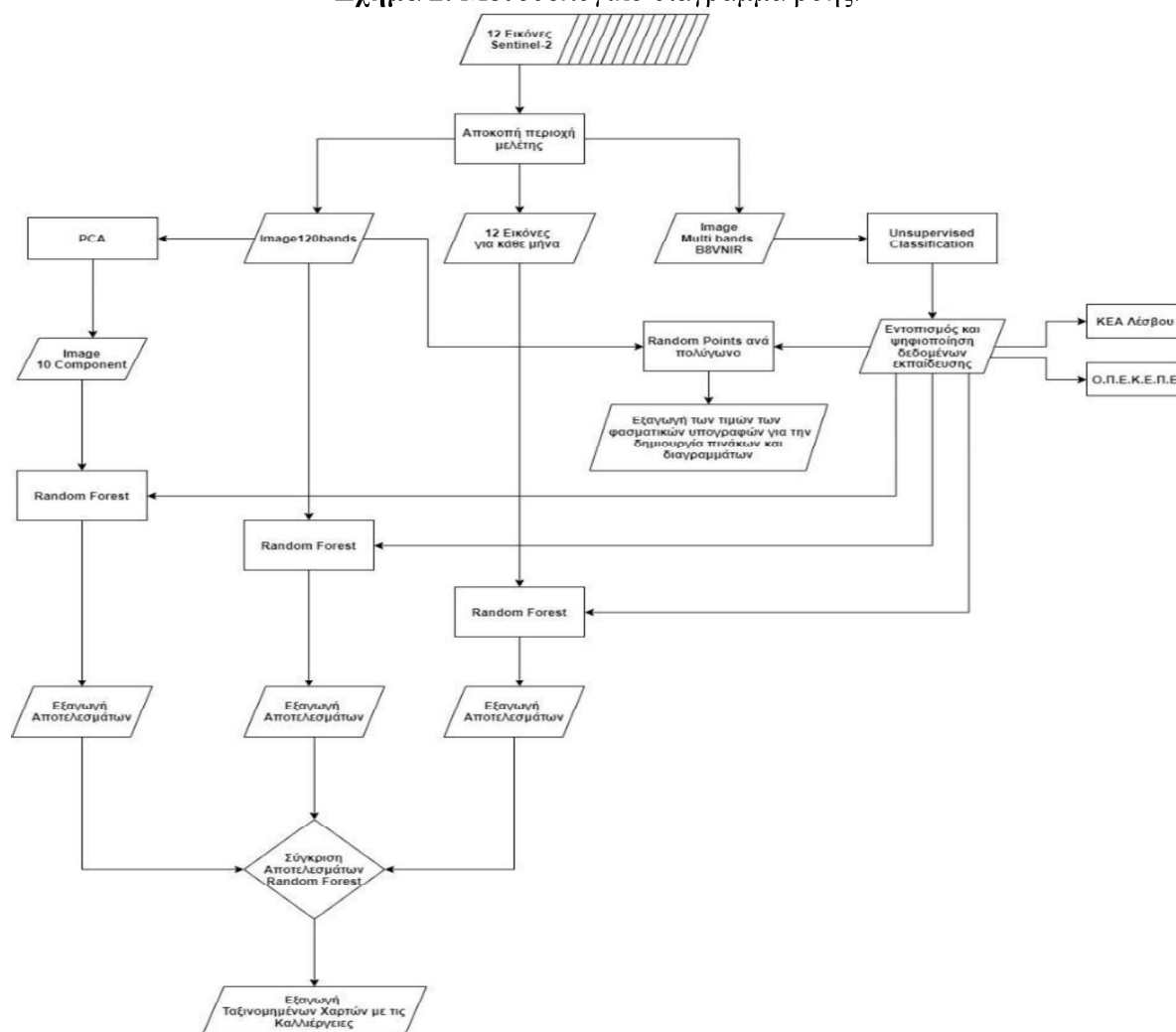
2.3 Μεθοδολογία

Η συνολική ροή της εργασίας απεικονίζεται στο Σχήμα 2. Αναπτύχθηκαν 15 μοντέλα ταξινόμησης Random Forest για διάφορα set δεδομένων. Καταρχήν, για κάθε μήνα ξεχωριστά αναπτύχθηκε ένα μοντέλο ώστε να διαπιστωθεί αν η ημερομηνία λήψης των δεδομένων επηρεάζει την αποτελεσματικότητα της ταξινόμησης αγροτικών καλλιεργειών. Έπειτα, όλα τα κανάλια όλων των ημερομηνιών ενώθηκαν σε μια εικόνα η οποία αποτέλεσε τα δεδομένα εισόδου σε επόμενο μοντέλο ταξινόμησης. Λόγω όμως του μεγάλου αριθμού καναλιών ως δεδομένων εισόδου, στη συνέχεια εφαρμόστηκε η ανάλυση κυρίων συνιστωσών (Principal Component Analysis-PCA).

Με την εφαρμογή της PCA μειώνονται οι διαστάσεις των δεδομένων (μεταβλητές) και διατηρείται η πληροφορία της εικόνας από 120 κανάλια σε 10 συνιστώσες (component). Αυτή η μέθοδος χρησιμοποιείται συχνά για την συμπίεση δεδομένων μετασχηματίζοντας ένα σύνολο από συσχετισμένες μεταβλητές σε ένα νέο σύνολο ασυσχέτιστων μεταβλητών. Τα νέα κανάλια (οι κύριες συνιστώσες της μεθόδου PCA) που δημιουργούνται είναι ασυσχέτιστα και ανεξάρτητα μεταξύ τους (Καρτάλης και Φείδας, 2017). Τα 120 κανάλια της πολυφασματικής εικόνας συσχετίζονται μεταξύ τους ιδιαίτερα τα ίδια κανάλια διαφορετικών χρονικών στιγμών καθώς η ανακλαστικότητα των στόχων διαφέρει ελάχιστα από ημερομηνία σε ημερομηνία για το ίδιο κανάλι. Μέσω της PCA ο αρχικός διανυσματικός χώρος συμπίεστηκε σε 10 κανάλια τα οποία είναι στατιστικά ασυσχέτιστα μεταξύ τους. Τα συγκεκριμένα κανάλια (συνιστώσες) ήταν τα δεδομένα εισόδου σε νέο μοντέλο ταξινόμησης RF. Τέλος, δεδομένου του μεγάλου αριθμού ταξινομητών και των πιθανών διαφορετικών αποτελεσμάτων εφαρμόστηκε ένα συνδυαστικό μοντέλο βασισμένο σε ένα πολλαπλό σχήμα συναίνεσης ώστε να δημιουργηθεί ένας τελικός

χάρτης με τα είδη των καλλιεργειών της Καλλονής. Η τελική απόφαση για κάθε εικονοστοιχείο ελήφθη μέσω πλειοψηφικής σύνθεσης των προηγούμενων 14 αποτελεσμάτων ταξινόμησης και προέκυψε ένας νέος χάρτης, όπου η τάξη του εικονοστοιχείου της εικόνας θα καθορίζεται από το ποια τάξη εμφανίζει τη μεγαλύτερη συχνότητα για το ίδιο εικονοστοιχείο στους 14 χάρτες των προηγούμενων μεθόδων (base classifiers). Τέλος, για κάθε χάρτη ταξινόμησης δημιουργήθηκαν και οι αντίστοιχοι χάρτες εμπιστοσύνης με το ποσοστό της ψήφου εμπιστοσύνης της επικρατούσας τάξης σε κάθε εικονοστοιχείο της εικόνας. Οι χάρτες εμπιστοσύνης είναι ένα εργαλείο για την αξιολόγηση των δεδομένων στην χωρική ανάλυση. Τα αποτελέσματα επιτρέπουν την σύγκριση και την ανάλυση της εμπιστοσύνης των δεδομένων και συνεπώς την αξιολόγησή τους.

Σχήμα 2. Μεθοδολογικό διάγραμμα ροής.



Όσον αφορά τη διαδικασία εκπαίδευσης, για τα δεδομένα εκπαίδευσης πρέπει να οριστούν κάποιοι παράμετροι οι οποίοι είναι το μέγιστο μέγεθος δείγματος εκπαίδευσης ανά τάξη, το μέγιστο μέγεθος δείγματος επικύρωσης ανά τάξη και η αναλογία του δείγματος εκπαίδευσης και επικύρωσης. Έπειτα, και για την επιλογή του ταξινομητή έπρεπε να καθοριστούν κάποιοι παράμετροι, όπως το μέγιστο βάθος του δέντρου, ο ελάχιστος αριθμός δειγμάτων σε κάθε κόμβο, το κριτήριο τερματισμού για το δέντρο παλινδρόμησης, το μέγεθος

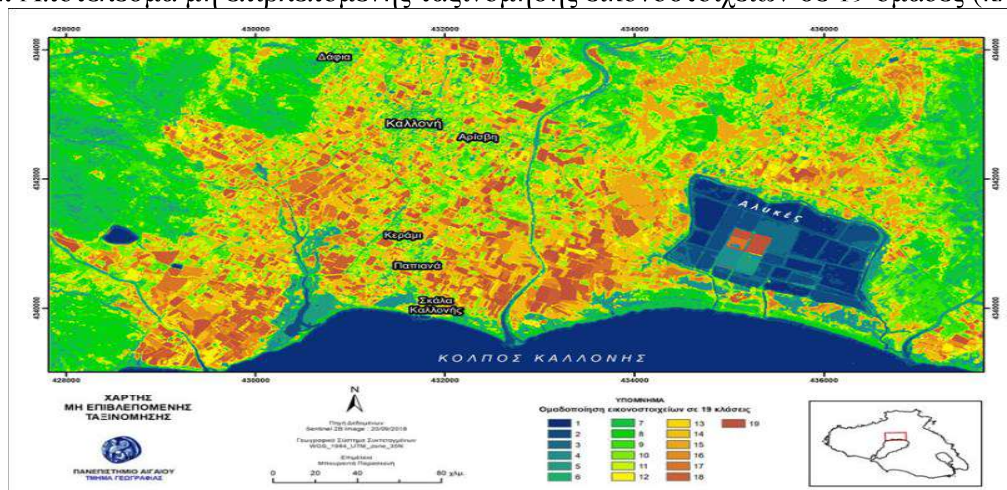
του τυχαία επιλεγμένου υποσυνόλου δυνατοτήτων σε κάθε κόμβο δέντρου και ο μέγιστος αριθμός των δέντρων του δάσους. Για την εφαρμογή της ταξινόμησης Random Forest έχουν αναπτυχθεί διάφορα εργαλεία ανοικτού και κλειστού κώδικα τα οποία μπορούν να χρησιμοποιηθούν για διάφορες μελέτες. Στην εργασία αυτή, χρησιμοποιήθηκε το λογισμικό QGIS και συγκεκριμένα η βιβλιοθήκη Orfeo Toolbox (OTB).

3. Αποτελέσματα και συζήτηση

3.1 Συλλογή δεδομένων εκπαίδευσης

Με την αρχική μη επιβλεπόμενη ταξινόμηση της περιοχής μελέτης σε διακριτές ομάδες εικονοστοιχείων κατηγοριοποιήθηκε η περιοχή σε 19 τάξεις. Όπως παρατηρούμε και στο Σχήμα 3 θα μπορούσαμε να πούμε ότι χωρίζεται σε τρεις γενικές κατηγορίες από το 1 έως το 6 απεικονίζεται το νερό που υπάρχει στην περιοχή μελέτης, όπως για παράδειγμα οι αλυκές και ο Κόλπος Καλλονής, οι κατηγορίες 7 έως 12 απεικονίζουν το γυμνό έδαφος της περιοχής κυρίως το πράσινο χρώμα και λίγο το κίτρινο σε κάποια σημεία της περιοχής, ενώ οι κατηγορίες 13 έως 19 απεικονίζουν τα διαφορετικά είδη καλλιεργειών, που σύμφωνα με τις κατηγορίες τα διαφορετικά είδη είναι 7.

Σχήμα 2. Αποτέλεσμα μη επιβλεπόμενης ταξινόμησης εικονοστοιχείων σε 19 ομάδες (κλάσεις).



Πίνακας 2. Κατανομή των δειγμάτων εκπαίδευσης ανά τάξη.

Τύπος Καλλιέργειας	Αριθμός αγροτεμαχίων	Αριθμός εικονοστοιχείων
Κτηνοτροφικά φυτά	17	2082
Λοιπά σιτηρά	2	424
Δεν δηλώνεται κάποια έκταση	2	143
Βοσκοτοπική έκταση	1	88
Σκληρό σίτος	2	481
Γεώμηλα	2	157
Αγρανάπαυση	2	562
Ελαιώνες	3	450
Σύνολο	31	4387

Σύμφωνα με τα δεδομένα εκπαίδευσης τα αγροτεμάχια κατηγοριοποιήθηκαν στις τάξεις σύμφωνα με τον Πίνακα 2. Οι κατηγορίες είναι τα κτηνοτροφικά φυτά (π.χ. τριφύλλι, σανός, μηδική), τα λοιπά σιτηρά (π.χ. μαλακό σιτάρι, σίκαλη, κριθάρι, βρώμη), σκληρό σίτος, γεώμηλα (πατάτες εαρινές), ελαιώνες, βοσκοτοπικές εκτάσεις καθώς επίσης ψηφιοποιήθηκαν και αγροτεμάχια που την δεδομένη χρονική περίοδο βρίσκονταν σε αγρανάπαυση. Τέλος, συμπεριλήφθηκε και μια κατηγορία όπου στα αγροτεμάχια δεν δηλώνεται κάποια έκταση. Αυτά τα αγροτεμάχια είναι εγκαταλελειμμένα αλλά παρουσιάζουν φασματική ομοιογένεια με άλλους τύπους κάλυψης γης όπως τις βοσκοτοπικές εκτάσεις και τα αγροτεμάχια που βρίσκονται σε αγρανάπαυση.

3.2 Ταξινόμηση Random Forest

Για τη διαδικασία της εκπαίδευσης το μέγιστο μέγεθος δείγματος εκπαίδευσης και επαλήθευσης ανά τάξη ορίστηκε σε 1000 και η αναλογία του δείγματος εκπαίδευσης και επαλήθευσης ορίστηκε σε 0,6. Αυτό σημαίνει ότι από το αρχικό σύνολο δεδομένων, το δεδομένα εκπαίδευσης ήταν το 60% των εικονοστοιχείων και το υπόλοιπο 40% τα δεδομένα επαλήθευσης. Τα δεδομένα εκπαίδευσης χρησιμοποιούνται ώστε να διερευνηθεί ποια είναι η κατάλληλη τιμή για τον αριθμό των δέντρων που θα χρησιμοποιηθούν για την εκπαίδευση του ταξινομητή, ενώ τα δεδομένα επικύρωσης για την αξιολόγηση της Random Forest μέσω της ακρίβειας που προκύπτει από τον πίνακα ταξινόμησης.

Για την επιλογή του ταξινομητή RF έπρεπε να καθοριστούν κάποιοι παράμετροι. Το μέγιστο βάθος του δέντρου ορίστηκε σε 5, ο ελάχιστος αριθμός δειγμάτων σε κάθε κόμβο ορίστηκε σε 10 ενώ δεν επιλέχτηκε κάποιο κριτήριο τερματισμού για το δέντρο παλινδρόμησης, οπότε η τιμή αυτή ήταν μηδέν.

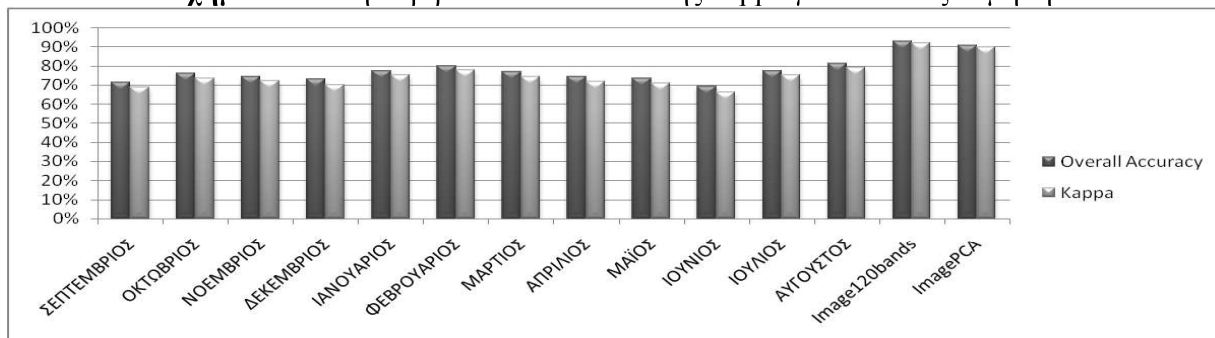
Το μέγεθος του τυχαία επιλεγμένου υποσυνόλου δυνατοτήτων σε κάθε κόμβο δέντρου ήταν μηδέν και ο μέγιστος αριθμός των δέντρων του δάσους ορίστηκε σε 100. Από διάφορες εργασίες έχει αποδειχθεί ότι αν είναι πάνω από 50, 100, ανάλογα κάθε φορά τι εξετάζεται, έχει πολύ καλά αποτελέσματα (Akbari et al, 2020). Οι παραπάνω επιλογές εφαρμόστηκαν σε όλα τα μοντέλα της Random Forest της συγκεκριμένης εργασίας.

Σύμφωνα με τα αποτελέσματα, όλες οι ταξινομήσεις είχαν υψηλά ποσοστά ακρίβειας η οποία κυμάνθηκε άνω του 60%, με την μεγαλύτερη ακρίβεια να σημειώνεται για την εικόνα του Αυγούστου σε ποσοστό 81% (Σχήμα 3).

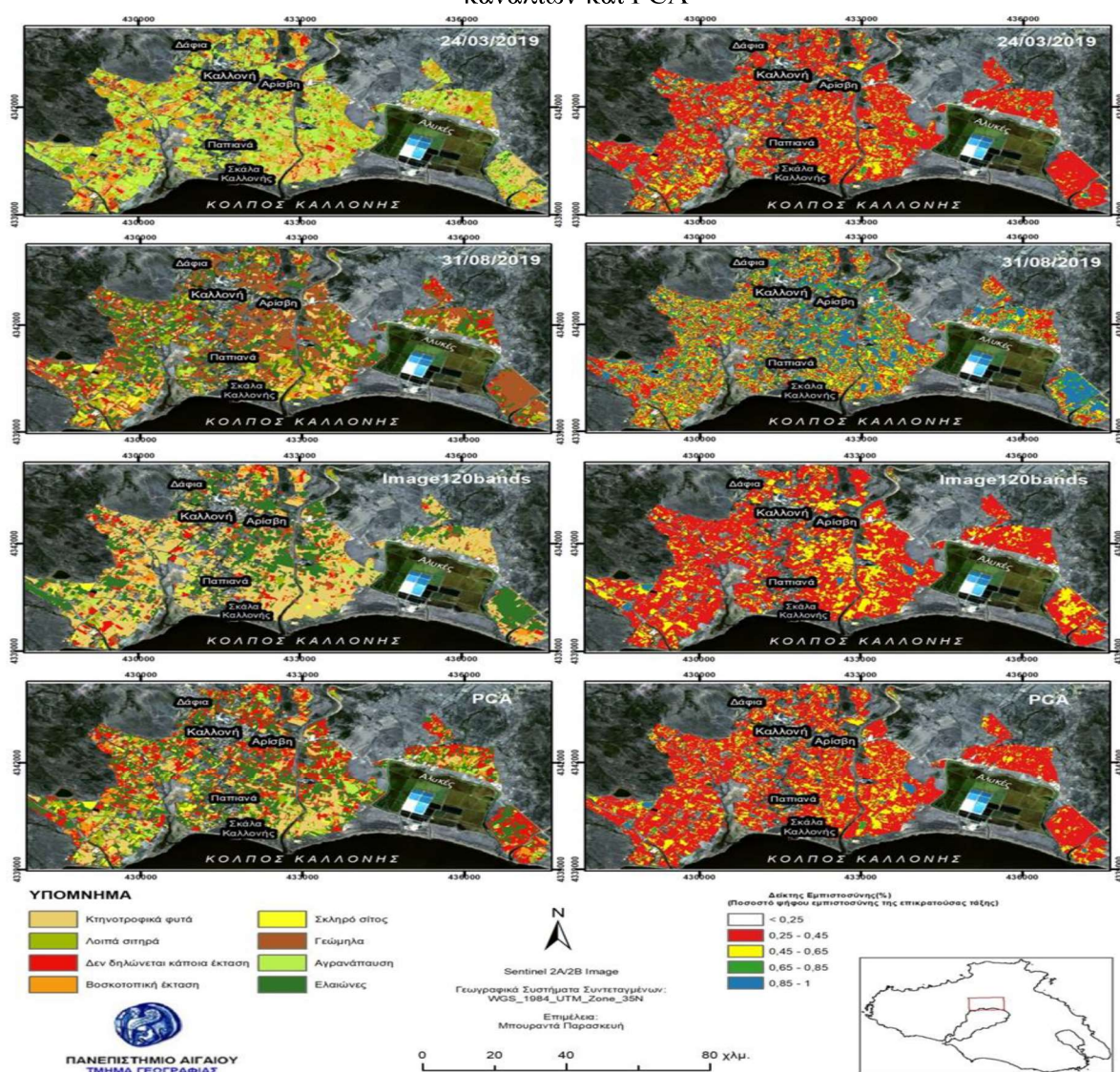
Παρόλα αυτά, από τη σύγκριση των χαρτών προκύπτει η διαφορετικότητα των αποτελεσμάτων. Τα σφάλματα οφείλονταν κυρίως λόγω της ταύτισης των φασματικών χαρακτηριστικών ορισμένων ειδών καλλιέργειας, όπως για παράδειγμα τα κτηνοτροφικά φυτά, τα λοιπά σιτηρά και το σκληρό σίτος επειδή παρουσιάζουν παρόμοιες φασματικές υπογραφές, ο ταξινομητής στο αποτέλεσμα του χάρτη τις εμφάνιζε ως μία κατηγορία (Σχήμα 4).

Στην εργασία οι χάρτες αυτοί αναπαριστούν το ποσοστό της ψήφου εμπιστοσύνης της επικρατούσας τάξης για κάθε εικονοστοιχείο, δηλαδή τα εικονοστοιχεία που ταξινομήθηκαν στην συγκεκριμένη κατηγορία πόσο σωστά είναι. Η ποσοτική ανάλυση των χαρτών είναι ότι οι τιμές των εικονοστοιχείων που κυμαίνονται από < 25% έχουν χαμηλή αξιοπιστία, από 0,25% - 0,45% χαμηλή προς μέτρια, 0,45% - 0,65% μέτρια, 0,65% - 0,85% μέτρια προς υψηλή και τέλος από 0,85% - 100% υψηλό ποσοστό αξιοπιστίας.

Σχήμα 3. Ολική ακρίβεια και συντελεστής kappa για κάθε ταξινόμηση.



Σχήμα 4. Χάρτες ταξινόμησης και δείκτη εμπιστοσύνης για 24/3/2019, 31/8/2019, σύνθεση 120 καναλιών και PCA



Αναλύοντας τα αποτελέσματα της ταξινόμησης για κάθε μήνα διαπιστώθηκε ότι σε όλες τις εικόνες στα κτηνοτροφικά φυτά ταξινομούνται εσφαλμένα οι υπόλοιπες κατηγορίες των

καλλιεργειών, τις περισσότερες φορές, σύμφωνα με την ακρίβεια τους. Αυτό συμβαίνει διότι η φασματική συμπεριφορά και ο ετήσιος κύκλος ανάπτυξης των συγκεκριμένων ειδών καλλιεργειών παρουσιάζουν αρκετές ομοιότητες. Επίσης, είναι είδη που έχουν ίδια χαρακτηριστικά ως προς το χρονικό διάστημα της φύτευσης και της συγκομιδής τους. Επομένως, για την ταξινόμηση των καλλιεργειών δεν μπορούμε να βασιστούμε σε μονοχρονικά δεδομένα, δηλαδή να χρησιμοποιήσουμε μια εικόνα ενός μηνός και να εξάγουμε συμπεράσματα για ολόκληρη τη περιοχή.

Παρατηρώντας τους χάρτες εμπιστοσύνης φαίνονται οι διαφορές που υπάρχουν στις εικόνες για κάθε μήνα και στην κατανομή των δεδομένων πάνω σε αυτές. Συγκεκριμένα στην εικόνα του Σεπτεμβρίου, τα δεδομένα που χρησιμοποιούνται για την ταξινόμηση, με βάση την χωρική τους ανάλυση είναι 85% αξιόπιστα και έμπιστα με την πραγματική εικόνα της περιοχής μελέτης. Για την εικόνα του Μαρτίου δεν συμβαίνει κάτι τέτοιο καθώς το ποσοστό αξιοπιστίας των δεδομένων είναι πάρα πολύ χαμηλό. Τα δεδομένα είναι 0,25% - 0,45% αξιόπιστα σε σχέση με την πραγματικότητα, και αυτό εξακολουθεί να συμβαίνει και στους επόμενους μήνες Απρίλιο - Μάιο καθώς τα ποσοστά αξιοπιστίας των δεδομένων εξακολουθούν να είναι χαμηλά, συγκεκριμένα στο 0,45%. Ενώ, το αποτέλεσμα για την εικόνα του Αύγουστο είναι το καλύτερο.

Η αξιοπιστία των δεδομένων αγγίζει το 0,85% - 100%, που σημαίνει ότι τα δεδομένα που χρησιμοποιήθηκαν είναι σωστά και ανταποκρίνονται στην πραγματικότητα.

Ακόμη, ένα πρόβλημα ως προς την ταξινόμηση των καλλιεργειών για τις εικόνες για κάθε μήνα αποτελεί και το μικρό δείγμα που χρησιμοποιείται ως δείγμα εκπαίδευσης για την εικόνα, καθώς έχουμε στην διάθεση μας λίγα δεδομένα εκπαίδευσης. Με περισσότερα δεδομένα εκπαίδευσης θα είχαμε καλύτερη εικόνα των καλλιεργειών που υπάρχουν στην περιοχή, με αποτέλεσμα καλύτερη ακρίβεια που συνεπάγεται με ένα αξιόπιστο αποτέλεσμα ταξινόμησης με την πραγματικότητα. Παρόλα αυτά τα στατιστικά αποτελέσματα της Random Forest για κάθε μήνα κυμάνθηκαν σε επίπεδα άνω των 65%, με την μεγαλύτερη ακρίβεια να σημειώνεται για την εικόνα του Αυγούστου σε ποσοστό 81% αλλά σύμφωνα με τα χαρτογραφικά αποτελέσματα υπάρχει ασυμφωνία των τάξεων ανά εικονοστοιχείων παρουσιάστηκαν και παραπάνω στους χάρτες εμπιστοσύνης.

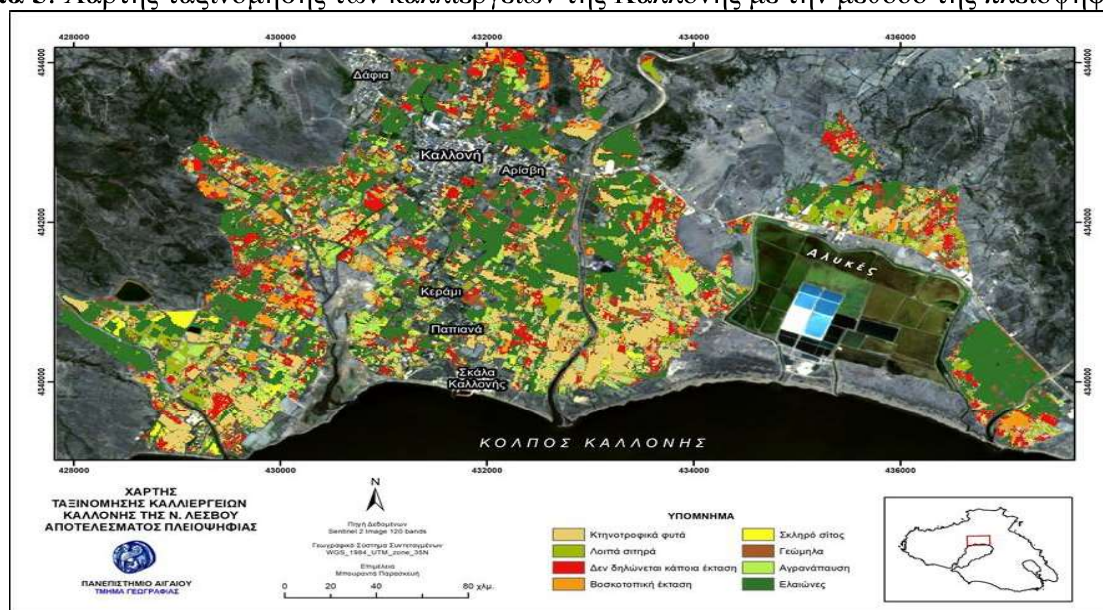
Η απόδοση της ταξινόμησης στην εικόνα σύνθεσης όλων των καναλιών ήταν υψηλότερη καθώς και σύμφωνα με τη βιβλιογραφία η απόδοση της ταξινόμησης RF βελτιώνεται με τη χρήση πολλαπλών ημερομηνιών Atzberger et al (2016). Συγκεκριμένα, ο ταξινομητής είχε την καλύτερη απόδοση με ακρίβεια 93%, παρόλα αυτά παρατηρώντας το αποτέλεσμα του χάρτη εμπιστοσύνης για την πολυφασματική εικόνα των 120 καναλιών, το ποσοστό αξιοπιστίας των δεδομένων που χρησιμοποιήθηκαν για την ταξινόμηση της εικόνας είναι πάρα πολύ χαμηλό. Πολύ πιθανόν αυτό οφείλεται στο μεγάλο αριθμό των δεδομένων εισόδου σε συνδυασμό με τον αριθμό των δεδομένων εκπαίδευσης. Έπειτα, μειώθηκε ο διανυσματικός χώρος των 120 καναλιών σε 10 συνιστώσες εφαρμόζοντας την ανάλυση κυρίων συνιστωσών όπου σύμφωνα με τον πίνακα ιδιοτιμών, οι 10 συνιστώσες εμπεριέχουν τουλάχιστον το 80% της αρχικής διασποράς.

3.3 Συνδυαστικό μοντέλο ταξινόμησης

Τέλος, μέσω της πλειοψηφίας και των 14 αποτελεσμάτων που δημιουργήθηκαν προέκυψε ένας νέος χάρτης (Σχήμα 5). Κάθε εικονοστοιχείο της εικόνας ταξινομήθηκε σύμφωνα συχνότητα εμφάνισης κάθε τάξης στους υπόλοιπους 14 χάρτες. Έπειτα για να αξιολογηθεί ο τελικός χάρτης χρησιμοποιήθηκαν τα αρχικά αγροτεμάχια. Για κάθε αγροτεμάχιο εκτιμήθηκε η τάξη του σύμφωνα με το συνδυαστικό μοντέλο ταξινόμησης με βάση την πλειοψηφία των

εικονοστοιχείων εντός του αγροτεμαχίου. Η εκτιμηθείσα τάξη συγκρίθηκε με την γνωστή τάξη του Πίνακα 2 και δημιουργήθηκε ο πίνακας σύγκυσης (Πίνακας 3). Από τα αποτελέσματα προκύπτει ότι η ολική ακρίβεια του χάρτη είναι 84% και ο συντελεστής kappa είναι 78%. Πιο αναλυτικά, όλα τα αγροτεμάχια ταξινομήθηκαν σωστά πλην 4 από τα 17 των κτηνοτροφικών φυτών και 1 από τα 2 σκληρού σίτου τα οποία ταξινομήθηκαν σε άλλες.

Σχήμα 5. Χάρτης ταξινόμησης των καλλιεργειών της Καλλονής με την μέθοδο της πλειοψηφίας.



Πίνακας 3. Πίνακας σύγκυσης συνδυαστικού μοντέλου ταξινόμησης.

	ΚΦ	ΛΣ	ΔΔ	ΒΕ	ΣΣ	ΓΜ	ΑΑ	ΕΛ	Σύνολο	CE	UA
ΚΦ	13	0	0	0	0	0	0	0	13	0	1
ΛΣ	1	2	0	0	0	0	0	0	3	0,33	0,67
ΔΔ	0	0	2	0	0	0	0	0	2	0	1
ΒΕ	0	0	0	1	0	0	0	0	1	0	1
ΣΣ	0	0	0	0	1	0	0	0	1	0	1
ΓΜ	1	0	0	0	0	2	0	0	3	0,33	0,67
ΑΑ	1	0	0	0	0	0	2	0	3	0,33	0,67
ΕΛ	1	0	0	0	1	0	0	3	5	0,4	0,6
Σύνολο	17	2	2	1	2	2	2	3	31		
ΟΕ	0,235	0	0	0	0,5	0	0	0			
ΡΑ	0,765	1	1	1	0,5	1	1	1			

* ΚΦ: Κτηνοτροφικά φυτά, ΛΣ: Λοιπά σιτηρά, ΔΔ: Δεν δηλώνεται κάποια έκταση, ΒΕ: Βοσκοτοπική έκταση, ΣΣ: Σκληρό σίτος, ΓΜ: Γεώμηλα, ΑΑ: Αγροανάπαυση, ΕΛ: Ελαιώνες, CE: Σφάλμα παρερμηνείας, ΟΕ: Σφάλμα παράβλεψης, UA: Ακρίβεια χρήση, ΡΑ: Ακρίβεια ταξινομητή

4. Συμπεράσματα και περαιτέρω έρευνα

Στην παρούσα εργασία έγινε εφαρμογή και αξιολόγηση του αλγορίθμου των τυχαίων δασών (RF) για την χαρτογράφηση των καλλιεργειών της περιοχής της Καλλονής της Ν. Λέσβου, με τη χρήση πολυχρονικών πολυφασματικών δεδομένων του δορυφόρου Sentinel-2. Από την εφαρμογή της μεθοδολογίας στις εικόνες για κάθε μήνα προέκυψαν χάρτες ταξινόμησης με ακρίβειες μεγαλύτερες από 65% με τη μεγαλύτερη ακρίβεια να σημειώνεται στην εικόνα του Αυγούστου σε ποσοστό 81%.

Παρόλα αυτά, κάθε χάρτης πρέπει να ερμηνευτεί όχι μόνο ως προς τα ποσοστά ακρίβειας για τα δεδομένα ακρίβειας αλλά και το δείκτη εμπιστοσύνης του. Από τον τελευταίο προέκυψαν ότι τα αποτελέσματα πρέπει να ερμηνευτούν με μεγάλη προσοχή δεδομένου ότι σημειώθηκαν χαμηλοί δείκτες εμπιστοσύνης για αρκετές τάξεις σε ορισμένες μονοχρονικές εικόνες. Εικάζεται ότι αυτό είναι αποτέλεσμα του μικρού αριθμού δεδομένων εκπαίδευσης. Αυτό έγινε περισσότερο εμφανές στην ταξινόμηση των συνθετικής εικόνας των 120 καναλιών όπου ο ταξινομητής είχε υψηλότερη απόδοση με ακρίβεια 93% αλλά επίσης ο δείκτης εμπιστοσύνης ήταν χαμηλός.

Παρόμοια αποτελέσματα είχε και η εικόνας με τις 10 συνιστώσες μετά την εφαρμογή της PCA με ακρίβεια 91%. Από το σύνολο των αποτελεσμάτων προκύπτει ότι αν και οι 14 ταξινομητές έχουν υψηλή ακρίβειας τόσο ο δείκτης εμπιστοσύνης όσο και η οπτική σύγκριση των αποτελεσμάτων μας οδηγεί στην ερμηνεία και αποδοχή των αποτελεσμάτων με μεγάλη προσοχή. Σε αυτή την περίπτωση προτείνεται η εφαρμογή συνδυαστικών μοντέλων ταξινόμησης ώστε να επιλεγθεί η βέλτιστη κλάση για κάθε εικονοστοιχείο είτε σύμφωνα με την πλειοψηφίας των αποφάσεων των επιμέρους ταξινομητών ή άλλων κριτηρίων (Vasilakos et al, 2020).

Λαμβάνοντας υπόψη τα παραπάνω, προτείνεται να διερευνηθούν και άλλοι αλγόριθμοι μηχανικής μάθησης όπως για παράδειγμα τα Τεχνητά Νευρωνικά Δίκτυα. Επίσης, μπορούν να ληφθούν υπόψη επιπλέον χαρακτηριστικά της βλάστησης όπως διάφοροι δείκτες βλάστησης NDVI, EVI κ.α., καθώς επίσης και διάφορα μορφολογικά χαρακτηριστικά των αγροτεμαχίων. Σε κάθε περίπτωση προτείνεται η χρήση περισσότερων δεδομένων εκπαίδευσης δεδομένου ότι από τον μικρό αριθμό δεδομένων μπορεί να προκύπτουν υψηλά ποσοστά ακριβείας για το σετ των δεδομένων επαλήθευσης εντούτοις μια πιο προσεκτική διερεύνηση όλων των αποτελεσμάτων να αποδείξει ότι η υψηλή ακρίβεια σε ένα μόνο ταξινομητή δεν σημαίνει ότι τα αποτελέσματα ανταποκρίνονται και στην πραγματικότητα.

5. Βιβλιογραφία

- Akbari E., Boloorani D.A., Samary N., Hamzeh S., Soufizadeh S. and Pignoutti S. (2020) Crop Mapping Using Random Forest and Particle Swarm Optimization based on Multi-Temporal Sentinel-2. *Remote Sensing* 12(9): 1449.
- Atzberger C., Vuolo F. and Immitzer M. (2016) First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sensing*, 8(3): 166.
- Courault D., Demarez V., Guerif M., Le Page M., Simonneaux V., Ferrant S. and Veloso A. (2016) Contribution of Remote Sensing for crop and water Monitoring. *Land Surface Remote Sensing in Agriculture and Forest*, 113-177.
- Esch T., Metz A., Marconcini M. and Keil M. (2014) Combined use of multi-seasonal high and medium resolution satellite imagery for parcel-related mapping of cropland and grassland. *International Journal of Applied Earth Observation and Geoinformation*, 28: 230-237.

- Inglada J., Arias M., Tardy B., Hagolle O., Valero S., Morin D., Dedieu G., Sepulcre G., Bontemps S., Defourny P., Koetz B. (2015) Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Satellite Optical Imagery. *Remote Sensing*, 7(9):12356-12379.
- Καρτάλης Κ. και Φείδας Χ. (2017) Αρχές Δορυφορικής Τηλεπισκόπησης. ΤΖΙΟΛΑ, Θεσσαλονίκη.
- Khatami R., Mountrakis G., Stehman S.V. (2016) A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research, *Remote Sensing of Environment*, 177: 89-100.
- Kussul N., Lavreniuk M., Skakun S., Shelestov S. (2017) Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data, *IEEE Geoscience and Remote Sensing Letters*, 14(5): 778-782.
- Peña-Barragán J.M., Ngugi M.K., Plant R.E., Six J. (2011) Object-based crop identification using multiple vegetation indices, textural features and crop phenology, *Remote Sensing of Environment*, 115: 1301-1316.
- Tang Y. and Li X. (2016) Set-based Similarity Learning in subspace for agricultural remote Sensing classification. *Neurocomputing*, 173: 332-338.
- Vasilakos C., Kalabokidis K., Hatzopoulos J., Kallos G., Matsinos I. (2007). Integrating New Methods and Tools in Fire Danger Rating, *International Journal of Wildland Fire*, 16(3): 306-316.
- Vasilakos C., Kavroudakis D., Georganta A. (2020) Machine Learning Classification Ensemble of Multitemporal Sentinel-2 Images: The Case of a Mixed Mediterranean Ecosystem. *Remote Sensing*. 12(12): 2005.
- Jiang Z., Huete A., Didan K., Miura T. (2008) Development of a two-band enhanced vegetation index without a blue band, *Remote Sensing of Environment*, 112(10): 3833-3845.