



**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΑΓΡΟΝΟΜΩΝ ΚΑΙ ΤΟΠΟΓΡΑΦΩΝ ΜΗΧΑΝΙΚΩΝ
ΕΡΓΑΣΤΗΡΙΟ ΤΗΛΕΠΙΣΚΟΠΗΣΗΣ**

**Αξιολόγηση Χρονικών Χαρακτηριστικών Για Την Ετήσια
Χαρτογράφηση Κάλυψης Γης Από Ετήσια Δεδομένα Sentinel-2**

ΓΙΑΝΝΗΣ ΑΝΔΡΕΑΣ ΤΣΙΩΤΑΣ ΝΙΑΧΟΠΕΤΡΟΣ



ΑΘΗΝΑ, ΟΚΤΩΒΡΙΟΣ 2020



NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF RURAL AND SURVEY ENGINEERING
REMOTE SENSING LABORATORY

Evaluating Temporal Features For Land Cover Mapping From Annual Sentinel-2 Data

IOANNIS ANDREAS TSIOTAS NIACHOPETROS



Athens, October 2020



RSLab

Remote Sensing Laboratory
National Technical University of Athens

✓ Sensing ✓ Analytics ✓ Monitoring





**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΑΓΡΟΝΟΜΩΝ ΚΑΙ ΤΟΠΟΓΡΑΦΩΝ ΜΗΧΑΝΙΚΩΝ
ΕΡΓΑΣΤΗΡΙΟ ΤΗΛΕΠΙΣΚΟΠΗΣΗΣ**

**Αξιολόγηση Χρονικών Χαρακτηριστικών Για Την Ετήσια
Χαρτογράφηση Κάλυψης Γης Από Ετήσια Δεδομένα Sentinel-2**

ΓΙΑΝΝΗΣ ΑΝΔΡΕΑΣ ΤΣΙΩΤΑΣ ΝΙΑΧΟΠΕΤΡΟΣ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 14^η Οκτώβρη 2020.

(Υπογραφή)

.....

Κ. Καράντζαλος

Αν. Καθηγητής Ε.Μ.Π.

(επιβλέπων)

(Υπογραφή)

.....

Δ. Αργιαλάς

Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....

Β. Καραθανάση

Καθηγήτρια Ε.Μ.Π.

Ευχαριστίες

Θα ήθελα αρχικά να ευχαριστήσω τον κύριο Καραντζαλο για την εμπιστοσύνη που μου έδειξε ώστε να μου αναθέσει αυτή τη διπλωματική εργασία. Η διαρκής παρουσία του και η άμεση ανταπόκρισή του σε όσα προβλήματα αντιμετώπισα ήταν καθοριστική προκειμένου να εκπονηθεί αυτή η εργασία. Καθοριστική ήταν επίσης και η παροχή των απαιτούμενων πόρων καθώς και το εξαιρετικό κλίμα το οποίο υπήρχε στη μεταξύ μας επικοινωνία.

Στη συνέχεια θα ήθελα να ευχαριστήσω την υποψήφια διδάκτορα Χριστίνα Καρακίζη για την άμεση και αποφασιστική της συμβολή σε όλα τα επίπεδα, την καθοδήγηση σε όλα τα βήματα αυτής της διπλωματικής και την άμεση ανταπόκρισή της σε κάθε μου απορία. Επιπλέον πολύ σημαντική ήταν και η εξαιρετικά φιλική της παρουσία.

Ακολούθως θα ήθελα να ευχαριστήσω τον υποψήφιο διδάκτορα Αριστείδη Βαϊόπουλο για την προεπεξεργασία των δεδομένων χωρίς την οποία η εκπόνηση της εργασίας θα ήταν αδύνατη καθώς και τον υποψήφιο διδάκτορα Ζαχαρία Κανδυλάκη και το μέλος του εργαστηρίου Αλέκο Φαλάγγα για τις πολύτιμες συμβουλές που μου έδωσαν και τη διάθεσή τους να με βοηθήσουν σε κάθε δυσκολία που αντιμετώπισα κατά την διάρκεια της εργασίας.

Τέλος θα ήθελα να ευχαριστήσω τους καθηγητές κύριο Δημήτρη Αργιαλά και κυρία Βασιλεία Καραθανάση για τις χρήσιμες συμβουλές και παρατηρήσεις επί του τεύχους και στην παρουσίαση της διπλωματικής εργασίας.

Γιάννης Τσιώτας

Αθήνα, Οκτώβριος 2020

ΠΕΡΙΛΗΨΗ

Το μεγάλο πλήθος ελεύθερα διαθέσιμων δορυφορικών τηλεπισκοπικών δεδομένων σε συνδυασμό με την ανάπτυξη πληθώρας αξιόπιστων αλγορίθμων μηχανικής μάθησης, καθιστά σήμερα την αυτόματη παραγωγή χαρτών καλύψεων γης πιο εφικτή από ποτέ. Σε λειτουργικό επίπεδο όμως υπάρχουν δυο βασικοί παράγοντες που προσθέτουν επιπλέον δυσκολία στην παραγωγή τους: Η υψηλή υπολογιστική ισχύς που απαιτείται για την υλοποίηση του μεθοδολογικού πλαισίου με χρήση πολυάριθμων διαχρονικών πολυφασματικών δεδομένων, καθώς επίσης και η ανάγκη για μεγάλο πλήθος δεδομένων αναφοράς, τα οποία πρέπει να δοθούν ως είσοδος ώστε να είναι εφικτή η εκπαίδευση του μοντέλου μηχανικής μάθησης.

Στη συγκεκριμένη διπλωματική εργασία διερευνήθηκαν μέθοδοι μείωσης της διαστατικότητας των διαχρονικών πολυφασματικών δεδομένων και η επίπτωσή τους στην ακρίβεια της ταξινόμησης, προκειμένου να μειωθεί η απαιτούμενη υπολογιστική ισχύς για τη διεξαγωγή του μεθοδολογικού πλαισίου. Για τις ανάγκες αυτής της εργασίας χρησιμοποιήθηκαν 10 πολυφασματικές απεικονίσεις υψηλής χωρικής ανάλυσης του Sentinel 2A για τέσσερα διαφορετικά tiles του δορυφόρου, το 34TEK, το 34SEJ, το 35TLF και το 34SEG. Μετά τις απαραίτητες προεπεξεργασίες και διορθώσεις στις δορυφορικές εικόνες, παρήχθησαν δυο σύνολα δεδομένων για κάθε ένα tile. Δημιουργώντας μια στοίβα από διαχρονικά χαρακτηριστικά, πιο συγκεκριμένα έξι φασματικά κανάλια και τρεις φασματικούς δείκτες, σχηματίστηκε αρχικά ένας κύβος φασματικών διαχρονικών χαρακτηριστικών για κάθε ένα tile. Μετέπειτα υπολογίστηκαν επί του φασματικού κύβου κάθε tile στατιστικά μεγέθη (όπως μέγιστη, ελάχιστη, μέση τιμή κ.α.) και έτσι σχηματίστηκαν οι κύβοι των χρονικών (temporal) χαρακτηριστικών. Για κάθε tile και κύβο εφαρμόστηκε μεγάλο πλήθος πειραμάτων με διαφορετικούς συνδυασμούς χαρακτηριστικών, αρχικά για ταξινόμηση και παραγωγή χάρτη ανά tile. Επιπρόσθετα υλοποιήθηκαν και πειράματα γενίκευσης μοντέλων σε άγνωστα δεδομένα, ώστε να βρεθεί το κατά πόσο είναι εφικτή η εκπαίδευση ενός μοντέλου σε μια περιοχή και η εφαρμογή του σε μια διαφορετική περιοχή, καθώς και πειράματα με εκπαίδευση και έλεγχο σε δύο περιοχές (συνδυαστικά πειράματα).

Από το πλήθος των πειραμάτων που πραγματοποιήθηκαν προέκυψαν υψηλές ολικές ακρίβειες άνω του 90% για πειράματα μηχανικής μάθησης στον κύβο των χρονικών χαρακτηριστικών ανά tile. Οι ολικές ακρίβειες για γενίκευση μοντέλων σε άγνωστα δεδομένα κυμάνθηκαν στο εύρος 49%-84%, ενώ παρατηρήθηκε ότι βελτιωμένα αποτελέσματα προέκυψαν με τη χρήση γειτνιαζόντων tiles, ένα για εκπαίδευση και ένα για έλεγχο ή με τη χρήση δύο tiles για εκπαίδευση, περίπτωση κατά την οποία οι ολικές ακρίβειες κυμάνθηκαν στο εύρος 55%-81%. Σε ό,τι αφορά στα συνδυαστικά πειράματα παρατηρείται ότι ανά δύο tiles οι ακρίβειες κυμαίνονται στο εύρος 94%-96%, είναι δηλαδή πολύ υψηλές σε κάθε περίπτωση με τη γειτνίαση να μην παίζει κάποιο ρόλο στις ακρίβειες. Η εφαρμογή της μεθοδολογίας και η ανάλυση των αποτελεσμάτων καταδεικνύει ότι η χρήση χρονικών χαρακτηριστικών είναι αποτελεσματική για την ταξινόμηση κάλυψης γης και καλλιεργειών σε εκτενείς εκτάσεις, με ταυτόχρονη μείωση της διαστατικότητας των τηλεπισκοπικών δεδομένων.

ABSTRACT

The abundance of freely available remote sensing data combined with the development of reliable machine learning algorithms renders the automatic production of land cover maps easier than ever before. However, on a practical level there are two main factors that hinder the process: The high computational power that is required in order to apply the methodology on a large amount of multispectral imagery, and also the need for a large amount of reference data which have to be given as input to the algorithm in order to render the training of a machine learning model feasible.

In this diploma thesis, feature extraction methods for multispectral timeseries and their effect on classification accuracy were studied and implemented, in order to reduce the aforementioned required computational power. 10 Sentinel 2A, high resolution, multispectral images for each of 4 different tiles of Sentinel 2A were used. These tiles are 34TEK, 34SEJ, 35TLF and 34SEG. After the required preprocessing procedures and corrections two datasets were created for every tile. Firstly 6 spectral bands and 3 indices were stacked in order to produce a spectral features cube for every tile. Afterwards statistical derivations of the cube per pixel (temporal features) were calculated and temporal features cubes were created. A large number of experiments was conducted with both training and testing on the same tile, training on one or two tiles and testing on a different tile and training and testing on 2 tiles.

The experiments which were conducted on one tile achieved high overall accuracy rates of over 90%. The overall accuracy range for experiments with training on one tile and testing on another was mediocre (49%-84%), and the method yielded its best results for tiles that shared a common border. The overall accuracy range for experiments with training on 2 tiles and testing on one tile was 55%-81%. The experiments with training and testing on the same tile yielded results with high overall accuracy rates ranging from 94% to 96%. The implementation of the methodology and the analysis of the results highlights that the use of temporal features is an efficient solution for accurate land cover classification over large areas, combined with reduced remote sensing data dimensionality.



ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΚΕΦΑΛΑΙΟ 1 ΕΙΣΑΓΩΓΗ.....	1
1.1 Αντικείμενο - Στόχοι.....	1
1.2 Δομή.....	2
1.3 Κίνητρο και Συνεισφορά.....	3
ΚΕΦΑΛΑΙΟ 2 ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ.....	5
2.1 Ανασκόπηση Βιβλιογραφίας.....	5
2.2 Ορισμοί.....	6
2.3 Ευρωπαϊκό Πρόγραμμα COPERNICUS.....	8
2.4 Η Δορυφορική Αποστολή Sentinel – 2.....	8
2.5 Δείκτες.....	9
2.6 Μέθοδοι Ταξινόμησης.....	10
2.7 Τυχαία Δάση (Random Forests – RF).....	12
2.7.1 Gini impurity.....	13
2.7.2 Πλεονεκτήματα και μειονεκτήματα των τυχαίων δασών.....	14
ΚΕΦΑΛΑΙΟ 3 ΜΕΘΟΔΟΛΟΓΙΑ.....	17
3.1 Εργασίες Βάσης της Μεθοδολογίας.....	17
3.2 Δορυφορικά Δεδομένα - Περιοχή Μελέτης.....	18
3.3 Συλλογή Διαχρονικών Δορυφορικών Δεδομένων.....	23
3.4 Κατηγορίες Ταξινόμησης και Δεδομένα Αναφοράς (Ground Truth).....	26
3.5. Επεξεργασίες Προετοιμασίας Δορυφορικών Δεδομένων για την Ταξινόμηση.....	31
3.5.1 Προεπεξεργασίες-Διορθώσεις δορυφορικών δεδομένων Sentinel-2A.....	31
3.5.2 Επιλογή καναλιών, κανονικοποίηση φασματικών τιμών, υπολογισμός δεικτών και σχηματισμός υπερκύβου φασματικών χαρακτηριστικών.....	31
3.5.3 Προετοιμασία δεδομένων για ταξινόμηση μηχανικής μάθησης.....	32
3.5.4 Προσθήκη DSM.....	33
3.5.5 Προσθήκη Brightness.....	34
3.5.6 Μείωση διαστατικότητας-Εξαγωγή χρονικών χαρακτηριστικών (temporal	

features).....	35
3.5.7 Μείωση διαστατικότητας - Επιλογή σημαντικότερων χαρακτηριστικών.....	36
3.6 Περιγραφή Πειραμάτων.....	37
3.6.1. Αρχικά πειράματα στο tile αναφοράς 34SEJ.....	37
3.6.2 Σχολιασμός μείωσης καναλιών σε σχέση με την ακρίβεια και τη χρονική και υπολογιστική αποδοτικότητα.....	42
3.6.3 Επιλογή πειραμάτων εφαρμογής της μεθοδολογίας ανά tile.....	47
3.6.4 Εφαρμογή της πειραματικής μεθοδολογίας ανά tile, στα tiles 34SEG, 34TEK, 35TLF.....	48
3.6.5 Πειράματα γενίκευσης μοντέλων και συνδυαστικά πειράματα.....	48
ΚΕΦΑΛΑΙΟ 4 ΑΠΟΤΕΛΕΣΜΑΤΑ.....	53
4.1. Ανάλυση Επιλεγμένων Πειραμάτων στο Tile 34SEJ.....	53
4.1.1 Πείραμα 1: Μοντέλο φασματικών χαρακτηριστικών με DSM.....	54
4.1.2 Πείραμα 2: Μοντέλο φασματικών χαρακτηριστικών με DSM και εξαγωγή βάσει σημαντικότητας (35%).....	56
4.1.3 Πείραμα 3: Μοντέλο φασματικών χαρακτηριστικών με DSM και εξαγωγή βάσει σημαντικότητας (65%).....	60
4.1.4 Πείραμα 4: Μοντέλο χρονικών χαρακτηριστικών με DSM.....	63
4.1.5 Πείραμα 5: Μοντέλο χρονικών χαρακτηριστικών με DSM και εξαγωγή βάσει σημαντικότητας (45%).....	66
4.1.6 Πείραμα 6: Μοντέλο χρονικών χαρακτηριστικών με DSM και εξαγωγή βάσει σημαντικότητας (75%).....	69
4.2 Συγκριτική Παρουσίαση των Πειραμάτων για το Tile 34SEJ.....	72
4.3 Συγκριτική Παρουσίαση των Πειραμάτων για το Tile 34TEK.....	76
4.4 Συγκριτική Παρουσίαση των Πειραμάτων για το Tile 35TLF.....	81
4.5 Συγκριτική Παρουσίαση των Πειραμάτων για το Tile 34SEG.....	87
4.6 Πειράματα Γενίκευσης σε Άγνωστα Δεδομένα και Συνδυαστικά Πειράματα.....	92
4.6.1 Πειράματα γενίκευσης σε άγνωστα δεδομένα και συνδυαστικά πειράματα ανά 2 tiles.....	92
4.6.2 Πειράματα γενίκευσης σε άγνωστα δεδομένα ανά 3 tiles (εκπαίδευση στα κοντινά, έλεγχος στα μακρινά).....	97

4.6.3 Πειράματα γενίκευσης σε άγνωστα δεδομένα ανά 3 tiles (εκπαίδευση στα μακρινά, έλεγχος στα κοντινά).....	103
4.6.4 Συγκριτική παρουσίαση πειραμάτων γενίκευσης σε άγνωστα δεδομένα ανά 3 tiles.....	109
4.6.5 Συγκριτική παρουσίαση πειραμάτων γενίκευσης σε άγνωστα δεδομένα και συνδυαστικά πειράματα ανά 2,3,4 tiles.....	112
ΚΕΦΑΛΑΙΟ 5 ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΤΑΣΕΙΣ.....	116
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	123

ΚΕΦΑΛΑΙΟ 1 ΕΙΣΑΓΩΓΗ

Τα τελευταία χρόνια οι χάρτες καλύψεων γης έχουν γίνει απαραίτητο συμπλήρωμα για πλήθος εφαρμογών όπως η ανίχνευση διαχρονικών μεταβολών βλάστησης, η παρακολούθηση της αναγέννησης καμένων εκτάσεων, η βελτιστοποίηση της αγροτικής παραγωγής. Παλιότερα οι χάρτες αυτοί παράγονταν με τη βοήθεια της φωτοερμηνείας και της ψηφιοποίησης. Ωστόσο, με την είσοδο αλγορίθμων μηχανικής μάθησης στις διαδικασίες, οι παραπάνω μέθοδοι αντικαταστάθηκαν σε μεγάλο βαθμό από αυτόματες και ημιαυτόματες μεθόδους ταξινόμησης.

Αυτές σε συνδυασμό με το πλήθος ελεύθερα διαθέσιμων, υψηλής ανάλυσης, πολυφασματικών απεικονίσεων, καθιστούν την παραγωγή χαρτών χρήσεων καλύψεων γης πιο απαραίτητες και εύκολες από ποτέ. Ωστόσο η υψηλή διαστατικότητα των τηλεπισκοπικών δεδομένων σε συνδυασμό με την ανάγκη για μεγάλο πλήθος αναφοράς προκείμενου να επιτευχθεί η ταξινόμηση, αποτελούν σημαντικά εμπόδια στην ταξινόμηση. Σε αυτή την εργασία θα προταθούν μεθοδολογίες υπερπήδησης αυτών των εμποδίων.

1.1 Αντικείμενο - Στόχοι

Το αντικείμενο της παρούσας εργασίας είναι η εφαρμογή και αξιολόγηση τεχνικών επιβλεπόμενης μάθησης για τη χαρτογράφηση κατηγοριών κάλυψης γης και καλλιεργειών σε 4 tiles του δορυφόρου Sentinel 2A, η γενίκευση με εκπαίδευση σε ένα ή περισσότερα tiles και έλεγχο σε κάποιο άλλο και η παραγωγή χαρτών κάλυψης γης για κάθε tile.

Πιο συγκεκριμένα, το αντικείμενο αυτής της εργασίας ήταν η αυτοματοποίηση της παραγωγής χαρτών χρήσεων - καλύψεων γης. Γι' αυτό τον σκοπό συλλέχθηκαν δεδομένα εκπαίδευσης από αρχικά 31 και στην συνέχεια 28 (κατόπιν συγχωνεύσεων) θεματικές κατηγορίες.

Για την εκπαίδευση του μοντέλου και την παραγωγή των χαρτών η εργασία βασίστηκε στον αλγόριθμο random forest. Οι συγκεκριμένος αλγόριθμος χρησιμοποιείται τα τελευταία χρόνια για ταξινόμηση καλύψεων χρήσεων γης και πετυχαίνει υψηλές ακρίβειες.

Οι επιμέρους στόχοι είναι:

- Η μείωση της διαστατικότητας των δεδομένων μέσω της χρήσης κύβου χρονικών χαρακτηριστικών με βάση την δημοσιευμένη εργασία στα ISPRS Archives "Assessing the contribution of spectral and temporal features for annual land cover and crop type mapping" (Karakizi et al., 2020)[1] στην οποία ο συγγραφέας συμμετείχε, και της επιλογής των καναλιών τα οποία συμβάλλουν κατά τον μεγαλύτερο βαθμό στις ακρίβειες (σημαντικά κανάλια). Σε ό,τι αφορά την επιλογή καναλιών αυτή επετεύχθη αλγοριθμικά ενώ η μείωση των καναλιών έγινε με αλληπάλληλες επιλογές των σημαντικότερων κατά περίπτωση καναλιών.
- Η δοκιμή διάφορων πειραμάτων στους κύβους των φασματικών και των χρονικών δεδομένων για την εύρεση εκείνων τα οποία προσφέρουν τις υψηλότερες ακρίβειες ταξινόμησης ή τον βέλτιστο

συνδυασμό μεταξύ ακριβούς ταξινόμησης και χαμηλής διαστατικότητας.

- Η εφαρμογή των παραπάνω πειραμάτων για την εκπαίδευση μοντέλων μηχανικής μάθησης σε διαφορετικές περιοχές του ελλαδικού χώρου.
- Η παραγωγή θεματικών χαρτών από το πείραμα με την καλύτερη ακρίβεια για τα τέσσερα tiles.
- Τα πειράματα γενίκευσης από ένα ή δύο tiles σε ένα άλλο tile για έλεγχο της ποιότητας της ταξινόμησης
- Τα συνδυαστικά πειράματα με δεδομένα αναφοράς από δύο και τρία tiles του Sentinel 2A και η σύγκριση των αποτελεσμάτων.

1.2 Δομή

Μετά από μια σύντομη εισαγωγή στη στοχοθεσία της εργασίας, στο δεύτερο κεφάλαιο γίνεται ανάπτυξη των βασικών θεωρητικών αρχών και μεθοδολογικών εργαλείων που αξιοποιήθηκαν για την παρούσα μελέτη. Αναλύονται λεπτομερώς οι αρχές πάνω στις οποίες βασίστηκαν τα μοντέλα που εκπαιδεύτηκαν στα επόμενα κεφάλαια, καθώς και αυτές βάσει των οποίων έγινε η επιλογή των χαρακτηριστικών και η μείωση της διαστατικότητας.

Στο τρίτο κεφάλαιο γίνεται η ανάπτυξη του μεθοδολογικού πλαισίου. Πιο συγκεκριμένα περιγράφονται:

- Η επιλογή και συλλογή των δεδομένων εκπαίδευσης για τα τέσσερα tiles του Sentinel 2A.
- Η κατηγοριοποίηση των δεδομένων εκπαίδευσης.
- Η προεπεξεργασία που απαιτήθηκε προκειμένου να φτάσουμε σε ατμοσφαιρικά διορθωμένες πολυφασματικές απεικονίσεις ανακλαστικότητας.
- Η προσθήκη δεικτών και DSM.
- Η κανονικοποίηση των τιμών και η δόμηση του υπερκύβου των φασματικών τιμών για κάθε ένα από τα tiles.
- Η δόμηση του υπερκύβου των χρονικών χαρακτηριστικών για κάθε ένα από τα διαφορετικά tiles.
- Το σύνολο των πειραμάτων που έγιναν στα τέσσερα tiles για τους κύβους των φασματικών και χρονικών χαρακτηριστικών.
- Τα πειράματα γενίκευσης σε διαφορετικές περιοχές και τα συνδυαστικά πειράματα με εκπαίδευση και έλεγχο σε πάνω από ένα tile.

Στο τέταρτο κεφάλαιο γίνεται η παράθεση και σχολιασμός των αποτελεσμάτων και των χαρτών όπως αυτοί προέκυψαν από την ταξινόμηση ανά κύβο. Γίνεται επίσης παράθεση και σχολιασμός των αποτελεσμάτων της γενίκευσης μοντέλων σε άγνωστα δεδομένα καθώς και σύγκριση μεταξύ των αποτελεσμάτων. Ακόμα γίνεται παρουσίαση κάποιων συνδυαστικών πειραμάτων που έγιναν με

εκπαίδευση και έλεγχο σε πάνω από ένα tile. Εν τέλει παρουσιάζονται συγκριτικές αξιολογήσεις για κάποια από τα πειράματα γενίκευσης που έχουν ιδιαίτερο ενδιαφέρον για την εξαγωγή συμπερασμάτων.

Στο πέμπτο κεφάλαιο παρατίθενται τα γενικά συμπεράσματα που προέκυψαν από αυτή τη διπλωματική εργασία, καθώς και ειδικότερα συμπεράσματα από τις ταξινομήσεις. Τέλος προσφέρονται προτάσεις που θα μπορούσαν να βελτιώσουν τις επιμέρους αδυναμίες αυτής της εργασίας.

1.3 Κίνητρο και Συνεισφορά

Κίνητρο για τη διερεύνηση της εφαρμογής σύγχρονων τηλεπισκοπικών μεθόδων και τεχνικών για την καταγραφή των καλύψεων γης, αποτέλεσε η ραγδαία ανάπτυξη σύγχρονων τηλεπισκοπικών αισθητήρων που έχει οδηγήσει στην παραγωγή ενός ολοένα και αυξανόμενου όγκου από τηλεπισκοπικά δεδομένα πολύ υψηλής χωρικής ανάλυσης, τα οποία μάλιστα κατά ένα μεγάλο ποσοστό διατίθενται και δωρεάν στο διαδίκτυο. Ακόμα ένα κίνητρο για αυτή την εφαρμογή αποτέλεσε και το πλήθος αλγορίθμων μηχανικής μάθησης που προσφέρονται σήμερα, μέσω των οποίων γίνεται ημιαυτόματη και αυτόματη ταξινόμηση καλύψεων γης, καθώς και ότι οι εν λόγω αλγόριθμοι προσφέρονται ως “πακέτα” από διάφορες βιβλιοθήκες της γλώσσας προγραμματισμού python.

Η συνεισφορά της παρακάτω εργασίας συνοψίζεται στα εξής:

- Υλοποίηση πειραμάτων για τη διερεύνηση των ακριβειών που μπορούν να επιτευχθούν από την ταξινόμηση σε ένα κύβο χρονικών χαρακτηριστικών και σύγκριση με την παραδοσιακή ταξινόμηση σε κύβο φασματικών χαρακτηριστικών.
- Σχεδιασμός αλγορίθμου εξαγωγής χαρακτηριστικών από τα δεδομένα εκπαίδευσης και ελέγχου πολυφασματικού υπερκύβου μέσω του μεγέθους gini impurity του Random Forest.
- Σχεδιασμός και υλοποίηση πειραμάτων γενίκευσης σε άγνωστα δεδομένα ανάμεσα σε 2 και 3 tiles
- Μελέτη και αξιολόγηση απόδοσης των ταξινομήσεων με δεδομένα εκπαίδευσης και ελέγχου από διαφορετικές περιοχές μελέτης.

ΚΕΦΑΛΑΙΟ 2 ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

2.1 Ανασκόπηση Βιβλιογραφίας

Σε αυτό το κεφάλαιο γίνεται η καταγραφή των στοιχείων βιβλιογραφίας που αποτέλεσαν τους πυλώνες πάνω στους οποίους βασίστηκε αυτή η εργασία. Η βιβλιογραφία που παρατίθεται αφορά εφαρμογές χαρτογράφησης καλύψεων γης με πολλές διαφορετικές μεθοδολογίες.

Οι Karakizi et al. (2020) [1] ορμώμενοι από τη συνεισφορά των σύγχρονων τηλεπισκοπικών δεδομένων στην παραγωγή χαρτών καλύψεων γης, κατέδειξαν τη σημασία τόσο των φασματικών όσο και των χρονικών χαρακτηριστικών στη λεπτομερή ταξινόμηση καλύψεων γης. Σε αυτή τη δημοσιευμένη εργασία περιγράφονται καταρχάς κάποια πειράματα για παραγωγή χαρτών καλύψεων χρήσεων γης σε έναν υπερκύβο από φασματικά κανάλια και δείκτες από μια χρονοσειρά δέκα απεικονίσεων του Sentinel 2A για το έτος 2016 με τη βοήθεια του αλγορίθμου Random Forest. Στη συνέχεια παράχθηκαν δύο διαφορετικοί κύβοι μειωμένης διαστατικότητας ο κύβος των φασματοχρονικών και αυτός των χρονικών χαρακτηριστικών και με βάση αυτούς έγινε πλήθος πειραμάτων ταξινόμησης επίσης με Random Forest. Ως φασματοχρονικά και χρονικά χαρακτηριστικά ορίστηκαν στατιστικά παράγωγα των φασματικών χαρακτηριστικών με τη σημαντική διαφορά ότι κάθε φασματοχρονικό χαρακτηριστικό είχε βάθος εννιά κανάλια ενώ κάθε χρονικό μόλις ένα κανάλι. Παρατηρήθηκε ότι η μείωση της διαστατικότητας ήταν 40% και 50% αντίστοιχα. Παρά τη σημαντική μείωση της διαστατικότητας η πτώση στη γενική ακρίβεια ήταν μικρότερη από 3% και για τις δύο περιπτώσεις, κάτι που κατέδειξε την αποτελεσματικότητα αυτής της λύσης. Επιπλέον καθοριστική ήταν και η συνεισφορά της συγκεκριμένης εργασίας ως προς την ανάδειξη της σημασίας που έχει η ανεξαρτησία των δεδομένων εκπαίδευσης και ελέγχου.

Οι Matton et al. (2015) [2] εισήγαγαν μια γενική μέθοδο για τη χαρτογράφηση των καλλιεργειών σε υψηλή ανάλυση με τη χρήση χάρτη υψηλής ανάλυσης και των χρονικών χαρακτηριστικών που αναφέρθηκαν παραπάνω. Αυτή η μέθοδος δοκιμάστηκε για δεδομένα από διάφορες περιοχές (Μαρόκο, Αργεντινή, Γαλλία, Βέλγιο, ΗΠΑ, Νότιος Αφρική, Ουκρανία, Κίνα) με τα πολυφασματικά δεδομένα να προέρχονται από τους δορυφόρους LANDSAT-8 και SPOT-4, χάρτες υψηλής και μεσαίας χωρικής ανάλυσης και in situ δεδομένα από την ομάδα JECAM. Όλα τα παραπάνω αποτελούν εισόδους των μη επιβλεπόμενων αλγορίθμων Kmeans και GMM (Gaussian Mixture Model) με τον πρώτο να χρησιμοποιείται για την ίδια την ταξινόμηση και τον δεύτερο για τη “διαλογή” των χρήσιμων στοιχείων. Οι μέγιστες γενικές ακρίβειες ανά γεωγραφική περιοχή κυμάνθηκαν στο εύρος 81,4%-99% και ως εκ τούτου παρατηρήθηκε ότι οι κύβοι χρονικών χαρακτηριστικών είναι ιδιαίτερα χρήσιμοι και για τις περιπτώσεις της μη επιβλεπόμενης ταξινόμησης.

Οι Valero et al. (2016) [3] χρησιμοποίησαν αντικειμενοστραφείς και pixel based στρατηγικές προκειμένου να δομηθεί ετήσια δυαδική μάσκα με τις κατηγορίες καλλιέργειες - μη καλλιέργειες. Και σε αυτή την περίπτωση προτάθηκαν κύβοι χρονικών χαρακτηριστικών έναντι των κλασικών φασματικών χαρακτηριστικών. Εξαιτίας της φύσης των ανιχνευόμενων εκτάσεων (δυαδική ταξινόμηση σε

καλλιέργειες και μη) χρησιμοποιήθηκαν ως βάση των χαρακτηριστικών μόνο κανάλια σχετικά με τη βλάστηση (NDVI, NDWI, Brightness). Προτάθηκαν και συγκρίθηκαν μία pixel based και δύο αντικειμενοστραφείς τεχνικές για τις ταξινομήσεις. Δώδεκα περιοχές, οι οποίες καταλάμβαναν εκτάσεις και στις πέντε ηπείρους, χρησιμοποιήθηκαν προκειμένου να δομηθούν αυτές οι μάσκες ενώ ιδιαίτερη σημασία δόθηκε στη διαφοροποίηση των περιοχών ως προς το μικροκλίμα, τις αγροικολογικές ζώνες κλπ. Τα δεδομένα που χρησιμοποιήθηκαν προήλθαν από τους δορυφόρους SPOT-4 και LANDSAT-8. Οι γενικές ακρίβειες ήταν σε κάθε περίπτωση άνω του 90%, πράγμα που καταδεικνύει τη χρησιμότητα των χρονικών χαρακτηριστικών σε συνδυασμό με δείκτες σε ό,τι αφορά την ανίχνευση καλλιεργειών.

Η Αντωνίου 2020 [4] στη μεταπτυχιακή διπλωματική της εργασία σχεδίασε και υλοποίησε ένα σύστημα επεξεργασίας με βάση τα δέντρα απόφασης με δεδομένα από τρία διαφορετικά tiles του Sentinel 2A. Οι ταξινομήσεις έγιναν στις εξής τρεις ομάδες πειραμάτων:

- Χρήση του ίδιου tile για εκπαίδευση και έλεγχο του αλγορίθμου ταξινόμησης, με βασικό πεδίο διερεύνησης να αποτελεί το ποσοστό των δεδομένων αναφοράς που χρησιμοποιήθηκαν για εκπαίδευση σε σχέση με αυτά που χρησιμοποιήθηκαν για έλεγχο. Τα αποτελέσματα είχαν πολύ υψηλές ακρίβειες - άνω του 98%.
- Εκπαίδευση του αλγορίθμου ταξινόμησης σε ένα tile του δορυφόρου Sentinel 2A και έλεγχος σε ένα άλλο tile. Εδώ τα αποτελέσματα είναι μέτριας ακρίβειας (64%-70%).
- Εκπαίδευση και έλεγχος το αλγορίθμου ταξινόμησης από όλα τα tiles, περίπτωση για την οποία επιτεύχθηκε ακρίβεια άνω του 99%

2.2 Ορισμοί

Ορισμός Φωτοερμηνείας

Η φωτοερμηνεία ορίζεται ως “η τέχνη ή επιστήμη της λήψης αξιόπιστων μετρήσεων μέσω φωτογραφιών” [5]. Ουσιαστικά μέσω της φωτοερμηνείας είμαστε σε θέση να εξάγουμε ποιοτικές και ποσοτικές πληροφορίες από μια δορυφορική απεικόνιση με τη βοήθεια της παρατήρησης.

Ορισμός Τηλεπισκόπησης

Η τηλεπισκόπηση είναι η επιστήμη και η τέχνη της απόκτησης πληροφοριών για ένα αντικείμενο, μια περιοχή ή ένα φαινόμενο δια μέσου της ανάλυσης δεδομένων, τα οποία έχουν αποκτηθεί από μια συσκευή, η οποία δεν είναι σε επαφή με το υπό έρευνα αντικείμενο, περιοχή ή φαινόμενο [6]. Πιο ειδικά, είναι η καταγραφή πληροφοριών από το υπεριώδες, το ορατό, το εγγύς υπέρυθρο και το μικροκυματικό μέρος του ηλεκτρομαγνητικού φάσματος και η ανάλυση των πληροφοριών που συλλέχθηκαν μέσα από οπτική (φωτοερμηνεία) και ψηφιακή επεξεργασία εικόνων.

Η πολυφασματική τηλεπισκόπηση αναφέρεται σε δέκτες που καταγράφουν λιγότερα από δεκαπέντε φασματικά κανάλια όπως είναι και ο Sentinel-2. Η υπερ-φασματική τηλεπισκόπηση αναφέρεται σε δέκτες που καταγράφουν περισσότερα από δεκαπέντε φασματικά κανάλια.

Ηλεκτρομαγνητικό Φάσμα

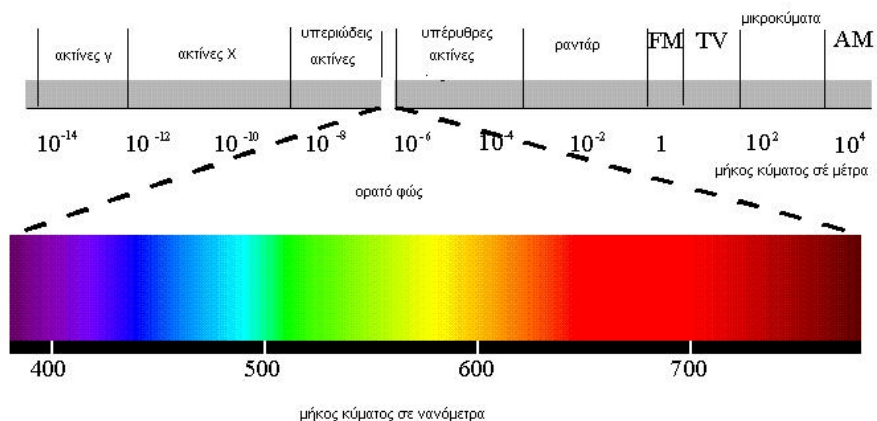
Ηλεκτρομαγνητικό φάσμα ονομάζεται το εύρος της περιοχής συχνοτήτων που καλύπτουν τα ηλεκτρομαγνητικά κύματα. Το μεγαλύτερο τμήμα του ηλεκτρομαγνητικού φάσματος είναι αόρατο για το ανθρώπινο μάτι.

Το φως έχει διττή φύση: Είναι παράλληλα κύμα και σωματίδιο. Η κυματική φύση του φωτός είναι και αυτή που έχει άμεσο ενδιαφέρον για εμάς. Το ηλεκτρομαγνητικό κύμα αποτελείται από δύο ταλαντούμενα πεδία, ένα ηλεκτρικό και ένα μαγνητικό, με τα δύο διανύσματα να σχηματίζουν ορθή γωνία μεταξύ τους και να είναι κάθετα προς τη διεύθυνση διάδοσης [7]. Η συχνότητα ενός κύματος, μετράει τον αριθμό των μηκών κύματος που διέρχονται από ένα σημείο ανά μονάδα χρόνου, ενώ το μήκος κύματος (συμβολίζεται με το ελληνικό γράμμα «λ») ορίζεται ως η μέση απόσταση μεταξύ δύο διαδοχικών μέγιστων ή ελάχιστων στα πλαίσια ενός απλοποιημένου περιοδικού προτύπου. Τα δύο αυτά χαρακτηριστικά είναι αντιστρόφως ανάλογα και το γινόμενό τους δίνει την ταχύτητα του φωτός.

Λόγω του πολύ μεγάλου εύρους του ηλεκτρομαγνητικού φάσματος χωρίζεται σε επιμέρους περιοχές, ανάλογα με το μήκος κύματος, που σε κάθε μία ανήκει ένα μικρό εύρος συχνοτήτων (αύξουσα σειρά συχνότητας):

- ραδιοκύματα
- μικροκύματα
- υπέρυθρη ακτινοβολία
- ορατή ακτινοβολία
- υπεριώδης ακτινοβολία
- ακτίνες X
- ακτίνες γ.

Το ορατό φάσμα κυμαίνεται από 390 με 750 νανόμετρα (μήκος κύματος, δισεκατομμυριοστό του μέτρου) και είναι αυτό που μπορεί να αντιληφθεί ο ανθρώπινος εγκέφαλος.



2.3 Ευρωπαϊκό Πρόγραμμα COPERNICUS

Το Copernicus είναι το πρόγραμμα γεωσκόπησης της Ευρωπαϊκής Ένωσης, το οποίο παρατηρεί το περιβάλλον και τον πλανήτη μας προς όφελος όλων των Ευρωπαίων πολιτών. Παρέχει υπηρεσίες πληροφόρησης με βάση δορυφορικά δεδομένα γεωσκόπησης και επίγεια (μη διαστημικά) δεδομένα.

Τον συντονισμό και τη διαχείριση του προγράμματος έχει αναλάβει η Ευρωπαϊκή Επιτροπή. Το πρόγραμμα υλοποιείται σε συνεργασία με τα κράτη μέλη, τον Ευρωπαϊκό Οργανισμό Διαστήματος (ESA), τον Ευρωπαϊκό Οργανισμό Εκμετάλλευσης Μετεωρολογικών Δορυφόρων (EUMETSAT), το Ευρωπαϊκό Κέντρο Μεσοπρόθεσμων Μετεωρολογικών Προβλέψεων (ECMWF), οργανισμούς της ΕΕ και την εταιρεία Mercator Océan.

Το πρόγραμμα χρησιμοποιεί μεγάλες ποσότητες παγκόσμιων δεδομένων προερχόμενων από δορυφορικά και από επίγεια, αερομεταφερόμενα και θαλάσσια συστήματα μέτρησης, για την παροχή πληροφοριών που βοηθούν τους παρόχους υπηρεσιών, τις δημόσιες αρχές και άλλους διεθνείς οργανισμούς, να βελτιώσουν την ποιότητα ζωής των πολιτών της Ευρώπης. Οι χρήστες του προγράμματος έχουν ελεύθερη και απρόσκοπτη πρόσβαση στις παρεχόμενες υπηρεσίες πληροφόρησης.

Το πρόγραμμα Copernicus περιλαμβάνει ήδη έξι δορυφορικές αποστολές αφιερωμένες στην παρακολούθηση της Γης από το διάστημα, τις Sentinel -1,-2,-3,-4,-5 και -6. Στο πλαίσιο της παρούσας εργασίας θα χρησιμοποιηθούν δεδομένα από τη δορυφορική αποστολή Sentinel-2.

2.4 Η Δορυφορική Αποστολή Sentinel – 2

Η αποστολή Copernicus Sentinel-2 αποτελείται από δύο δορυφόρους πολικής τροχιάς (γυρίζουν σε τροχιά των πόλων) με ίδια ηλιοσύγχρονη τροχιά, διαδοχικά σε γωνία 180° μεταξύ τους. Στόχος του είναι η παρακολούθηση της μεταβολής των επιφανειακών συνθηκών, όπου μαζί με το ευρύ πλάτος σάρωσης (290 χλμ.) και τον μικρό χρόνο επαναπροσπέλασης (δέκα ημέρες στον ισημερινό με ένα δορυφόρο και πέντε ημέρες με δύο δορυφόρους κάτω από συνθήκες χωρίς σύννεφα που έχουν ως αποτέλεσμα να δίνουν αποτελέσματα ανά τρεις ημέρες στα μέσα γεωγραφικά πλάτη) θα επιτρέψει την παρακολούθηση των μεταβολών της γήινης επιφάνειας. Τα όρια κάλυψης κυμαίνονται μεταξύ 56 ° νότιου και 84 ° βόρειου γεωγραφικού πλάτους.

Οι δύο δορυφόροι Sentinel-2A και Sentinel-2B κατασκευάστηκαν κυρίως για τη διαχείριση εκτάκτων αναγκών και περιλαμβάνουν όργανα πολυφασματικής απεικόνισης για την παρακολούθηση και διαχείριση του περιβάλλοντος (μελέτη των υδάτων και υδρογραφικών δικτύων, κάλυψη εδαφών και βλάστηση σε παράκτιες περιοχές, κλπ). Πετάνε σε ύψος 786 km από την επιφάνεια της γης και έχουν γωνιακή απόκλιση 180° μεταξύ τους. Η διάρκεια ζωής τους έχει σχεδιαστεί στα επτά έτη και ζυγίζουν ο

καθένας 1140 kg. Έχουν ένα υψηλής χωρικής διακριτικής ικανότητας πολυφασματικό σαρωτή με δεκατρία φασματικά κανάλια.

Οι δορυφόροι είναι προϊόντα συνεργασίας της Ευρωπαϊκής Επιτροπής Διαστήματος (European Space Agency – ESA), μιας κοινοπραξίας εξήντα εταιρειών με επικεφαλής την Airbus Defense and Space και υποστηρίζονται από το Εθνικό Κέντρο Διαστημικών Μελετών (Centre National d'Études Spatiales – CNES) της Γαλλίας και το Γερμανικό κέντρο Διαστημικής DLR, για τη βελτιστοποίηση των καλύψεων της γήινης επιφάνειας.

2.5 Δείκτες

Τα κανάλια δεικτών είναι ουσιαστικά εικόνες που προκύπτουν μετά από μαθηματικές πράξεις μεταξύ φασματικών καναλιών της ίδιας ή διαφορετικών εικόνων. Αυτές οι πράξεις μπορεί να είναι από απλές αφαιρέσεις ως και λόγοι καναλιών. Οι λόγοι βασίζονται στις φασματικές ιδιότητες των υλικών του εδάφους και στην άμεση σύνδεση που υπάρχει μεταξύ αυτών και των φυσικών ιδιοτήτων. Ως αποτέλεσμα προκύπτει ότι οι λόγοι των φασματικών καναλιών δίνουν πληροφορίες για τη χημική σύσταση κάθε στόχου. Για τις ανάγκες της τηλεπισκόπησης έχει αναπτυχθεί πλήθος τέτοιων δεικτών που κάθε ένας είναι χρήσιμος για την ανάδειξη διαφορετικών χωρικών οντοτήτων.

Οι δείκτες που θα χρησιμοποιηθούν σε αυτή την εργασία είναι οι:

- NDVI (δείκτης βλάστησης)
- NDWI (δείκτης υδάτων)
- NDBI (δείκτης οικιστικής ανάπτυξης)
- BRIGHTNESS (δείκτης

Ο NDVI είναι ένας δείκτης που δομήθηκε για την ανάδειξη της βλάστησης μιας εικόνας. Η μαθηματική του έκφραση διατυπώνεται ως εξής:

$$NDVI = \frac{NIR-RED}{NIR+RED} \text{ και άρα για τον SENTINEL είναι } NDVI = \frac{(B08-B04)}{(B08+B04)}$$

Παίρνει τιμές στο εύρος (-1,1). Χρησιμοποιείται για τον υπολογισμό της βλάστησης και ειδικότερα της ζωτικότητας αυτής, όπως φαίνεται από τις δορυφορικές απεικονίσεις. Η χλωροφύλλη, όπως φαίνεται και στο διάγραμμα στην επόμενη σελίδα, ανακλά σχεδόν το 50% της προσπίπτουσας ακτινοβολίας στο εγγύς υπέρυθρο και ένα ελάχιστο ποσοστό της προσπίπτουσας στο ορατό. Ως εκ τούτου, εάν η τιμή της ανακλαστικότητας στο εγγύς υπέρυθρο είναι μεγαλύτερη από το ορατό, τότε έχουμε ισχυρή ύπαρξη βλάστησης. Γενικά με αυτόν τον δείκτη ελαχιστοποιούνται οι τοπογραφικές επιδράσεις αλλά παρουσιάζονται δείγματα κορεσμού σε πολύ υψηλές συγκεντρώσεις βλάστησης.

Ο NDWI αποτελεί τον καλύτερο δείκτη για τον διαχωρισμό των υδάτινων μαζών από το χερσαίο περιβάλλον [31]. Το νερό έχει μεγάλη απορρόφηση και μικρή ανακλαστικότητα μεταξύ ορατού και

υπερύθρου τμήματος του η/μ φάσματος. Ο δείκτης είναι ευαίσθητος ακόμα στην οικιστική γη και προσφέρεται για ανάδειξη πλημμυρών. Ο NDWI μεγιστοποιεί την ανακλαστικότητα του νερού χρησιμοποιώντας τον εξής τύπο:

$$NDWI = \frac{GREEN-NIR}{GREEN+NIR} \text{ ο οποίος για τον Sentinel-2 είναι: } NDWI = \frac{(B03-B08)}{(B03+B08)} \text{ και έχει εύρος } (-1,1).$$

Ο NDBI [30] χρησιμοποιείται για τον εντοπισμό των αστικών ζωνών που τυπικά έχουν υψηλότερη ανακλαστικότητα στην περιοχή του μικροκυματικού υπέρυθρου (SWIR) σε σχέση με την εγγύς υπέρυθρη ακτινοβολία (NIR). Ο τύπος είναι:

$$NDBI = \frac{NIR-SWIR}{NIR+SWIR} \text{ άρα για τον Sentinel-2 είναι: } NDBI = \frac{(B11-B08)}{(B11+B08)} \text{ και το εύρος πάλι } (-1,1)$$

Ο Brightness θεωρείται ως δείκτης συμπληρωματικός του NDWI, με την έννοια ότι χρησιμοποιείται για την ανίχνευση υγρασίας του εδάφους και της βλάστησης και δίνεται από το άθροισμα των τετραγώνων όλων των καναλιών που υφίστανται στον υπερκύβο. Στη δική μας εργασία ο τύπος είναι:

$$BRIGHTNESS = \sqrt{(BLUE^2 + GREEN^2 + RED^2 + NIR^2 + RE^2 + SWIR^2)} \text{ , δηλαδή ίσο με: } \\ \sqrt{(B01^2 + B02^2 + B03^2 + B04^2 + B05^2 + B06^2)}$$

2.6 Μέθοδοι Ταξινόμησης

Μια πολυφασματική δορυφορική εικόνα αποτελείται από πολλές εικόνες της ίδιας ακριβώς περιοχής στην επιφάνεια της γης, καταγεγραμμένες την ίδια χρονική στιγμή αλλά σε διαφορετικά μήκη κύματος, από αισθητήρα ευαίσθητο σε αυτά. Τα διάφορα αντικείμενα στην επιφάνεια της γης, δηλαδή ουσιαστικά οι διαφορετικές καλύψεις γης, αντιδρούν διαφορετικά στην ηλεκτρομαγνητική ακτινοβολία (του ήλιου για τους παθητικούς αισθητήρες) και εμφανίζονται φωτεινότερα ή σκοτεινότερα, ανάλογα με το μήκος κύματος του ηλεκτρομαγνητικού φάσματος που προσπίπτει σε αυτά. Με αυτόν τον τρόπο προκύπτουν εικόνες ακριβώς της ίδιας περιοχής καταγεγραμμένες την ίδια χρονική στιγμή αλλά με διαφορετική ένταση ακτινοβολίας στις διαφορετικές περιοχές τους. Η διαφορά μεταξύ των καταγραφών στα διαφορετικά μήκη κύματος (όχι μόνον του ορατού φάσματος αλλά και του υπέρυθρου και των θερμικών περιοχών του ηλεκτρομαγνητικού φάσματος) είναι μία σημαντική αρχή πάνω στην οποία βασίστηκε η Τηλεπισκόπηση.

Η διαφοροποίηση καλύψεων γης θα ήταν σχεδόν αδύνατο να εντοπιστεί μόνο σε κάποια εικόνα καταγεγραμμένη σε συγκεκριμένο μήκος κύματος. Μία πολυφασματική δορυφορική εικόνα αποτελείται από τόσες επιμέρους εικόνες, όσα και τα μήκη κύματος που καταγράφονται για την ίδια ακριβώς περιοχή, την ίδια ακριβώς χρονική στιγμή και οι εικόνες αυτές είναι απολύτως υπερθέσιμες μεταξύ τους, δηλαδή το κάθε εικονοστοιχείο απεικονίζει ακριβώς την ίδια περιοχή στα διαφορετικά μήκη κύματος. Κάθε μια από τις πολλές (δισδιάστατες) εικόνες που αποτελείται μια πολυφασματική (τρισδιάστατη) δορυφορική εικόνα ονομάζεται φασματικό κανάλι ή φασματική μπάντα (Περάκης, 2015) [8].

Η ταξινόμηση των πολυφασματικών εικόνων γίνεται με σκοπό να κατηγοριοποιηθεί κάθε εικονοστοιχείο της εικόνας σε κάποια κατηγορία κάλυψης γης. Η κατηγοριοποίηση αυτή βασίζεται στα φασματικά πρότυπα που ενυπάρχουν στα δεδομένα του κάθε εικονοστοιχείου, καθώς χρησιμοποιούνται ως η ψηφιακή βάση της κατηγοριοποίησης [6] και μπορούν να συσχετιστούν με βιοφυσικές ιδιότητες αποκαλύπτοντας στοιχεία για το βιοφυσικό περιεχόμενο του κάθε εικονοστοιχείου. Με βάση αυτή την κατηγοριοποίηση δομούνται και οι θεματικοί χάρτες. Ένας θεματικός χάρτης παρουσιάζει τη χωρική κατανομή των αναγνωρίσιμων χαρακτηριστικών της γήινης επιφάνειας, δηλαδή παρέχει μία πληροφοριακή περιγραφή για κάθε περιοχή που πρόκειται να απεικονιστεί σε ένα τέτοιο προϊόν. Ανάλογα με τον τρόπο προσέγγισης του κάθε προβλήματος η ταξινόμηση μπορεί να κατηγοριοποιηθεί σύμφωνα με τα παρακάτω αλλά και να ανήκει σε μια ή περισσότερες κατηγορίες.

- 1) Μη Επιβλεπόμενη και Επιβλεπόμενη.
- 2) Παραμετρική και Μη Παραμετρική.
- 3) Hard και Soft (Fuzzy).
- 4) Pixel-based και Object-based.

Στη μη επιβλεπόμενη το σύνολο δεδομένων εκπαίδευσης είναι άγνωστο ενώ τα εικονοστοιχεία ομαδοποιούνται σύμφωνα με τα φασματικά τους χαρακτηριστικά. Στην επιβλεπόμενη το σύνολο των δεδομένων εκπαίδευσης είναι γνωστό και η διαδικασία ξεκινά με την εκπαίδευση του ταξινομητή με αυτά τα δεδομένα ή και επιπρόσθετα. Οι παραμετρικοί ταξινομητές βασίζονται στη στατιστική κατανομή πιθανότητας κάθε τάξης, ενώ οι μη-παραμετρικοί χρησιμοποιούνται σε περιπτώσεις όπου η συνάρτηση πυκνότητας είναι άγνωστη, με σκοπό τον υπολογισμό της πιθανότητας της συνάρτησης πυκνότητας [9]. Κατά τη Hard ταξινόμηση, κάθε εικονοστοιχείο ταξινομείται σε μία και μοναδική τάξη, ενώ κατά τη Soft ή Fuzzy προσέγγιση, κάθε εικονοστοιχείο ταξινομείται σε παραπάνω από μία τάξεις με τη χρήση συναρτήσεων συμμετοχής για κάθε μία τάξη. Τέλος, ένας άλλος διαχωρισμός γίνεται με βάση τη στοιχειώδη μονάδα ανάλυσης της εικόνας, η οποία μπορεί να είναι είτε κάθε ένα εικονοστοιχείο (Pixel-based προσέγγιση) ή μια ομάδα εικονοστοιχείων που εκτός από κοινά φασματικά χαρακτηριστικά, συμπεριλαμβάνουν σημαντικές εννοιολογικές πληροφορίες (γεωμετρικές, τοπολογικές).

Στην παρούσα εργασία η ταξινόμηση είναι επιβλεπόμενη, pixel-based. Σε κάθε εικονοστοιχείο αντιστοιχεί ένα διάνυσμα με n τιμές. Οι χαρακτηριστικές τιμές του διανύσματος, είναι οι τιμές που πήρε αυτό το εικονοστοιχείο σε κάθε μια grayscale εικόνα από αυτές που αποτελείται η πολυφασματική απεικόνιση. Το εκπαιδευμένο μοντέλο που προκύπτει από την εφαρμογή ενός αλγορίθμου ταξινόμησης σε ένα σύνολο διανυσμάτων χαρακτηριστικών καλείται και ταξινομητής (classifier).

Συγκεκριμένα για αυτή την εργασία εφαρμόστηκε ο αλγόριθμος Random Forests. Στη συνέχεια

αναλύεται η συγκεκριμένη μεθοδολογία.

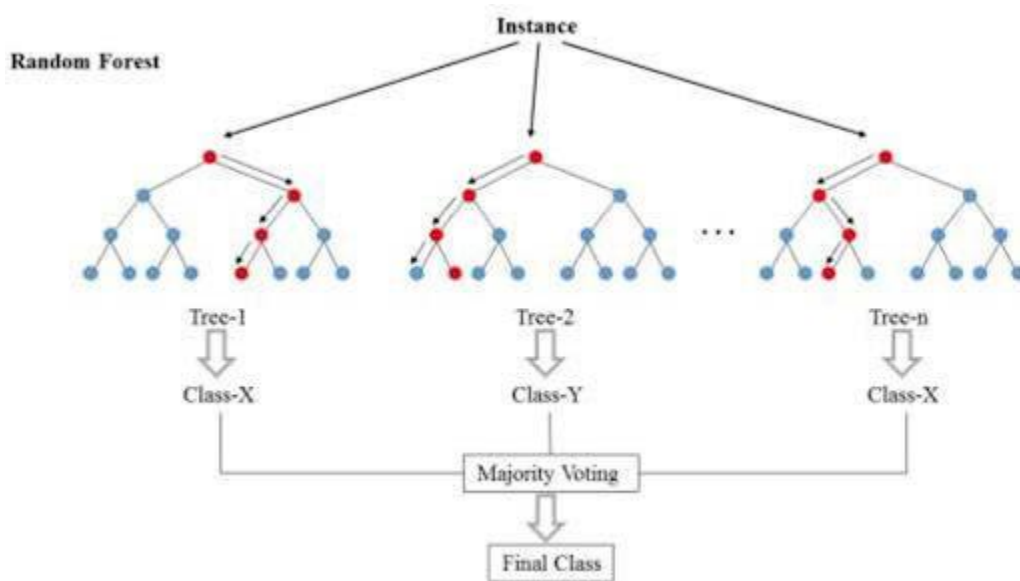
2.7 Τυχαία Δάση (Random Forests – RF)

Περιγραφή

Μία ακόμα μέθοδος ταξινόμησης δεδομένων είναι τα Τυχαία Δάση. Τα Random Forests είναι ουσιαστικά μία συλλογή από δέντρα απόφασης (decision trees). Εμπνευστής της μεθόδου των Random Forests είναι ο Leo Breiman [10]. Τα παρακάτω βήματα δίνουν μια γενική ιδέα για το πώς λειτουργούν τα random forests:

- Αρχικά αναπτύσσονται πολλά classification decision trees σε κάθε ένα από τα οποία ανατίθεται ένα τμήμα των δεδομένων με προσπάθεια κάθε δέντρο να διαχειριστεί δείγματα από κάθε θεματική κατηγορία.
- Κάθε δέντρο ταξινομεί όλα τα εικονοστοιχεία τα οποία του αντιστοιχούν.
- Έτσι κάθε κλάση έχει έναν αριθμό ψήφων (votes).
- Η τελική και οριστική ταξινόμηση γίνεται με το δάσος να διαλέγει την κλάση με τις περισσότερες ψήφους.

Όλα τα παραπάνω φαίνονται καλύτερα στην επόμενη εικόνα:



Σχήμα 2 : Αλγόριθμος random forest

Ο αλγόριθμος random forest είναι στην ουσία ένα σύνολο δέντρων απόφασης. Τα δέντρα εκπαιδεύονται με ένα μέρος των δεδομένων των αρχικών δεδομένων εκπαίδευσης. Κάθε δέντρο απόφασης χρησιμοποιεί ένα σύνολο κόμβων απόφασης από τους οποίους κάθε εικονοστοιχείο του δέντρου διέρχεται (δυαδική απόφαση) και καταλήγει εν τέλει να ταξινομηθεί σε κάποιο κόμβο-φύλλο του οποίου ιδεατά όλα τα εικονοστοιχεία πρέπει να περιλαμβάνονται σε μια κλάση. Ο βέλτιστος διαχωρισμός στον κόμβο θεωρείται εκείνος ο οποίος έχει ως αποτέλεσμα υψηλή “καθαρότητα” (purity)

της κλάσης. Το μέτρο το οποίο μετρά την καθαρότητα είναι το gini impurity το οποίο θα οριστεί στην συνέχεια. [11]

2.7.1 Gini impurity

Το gini impurity ορίζεται ως η πιθανότητα να ταξινομηθεί λάθος ένα τυχαία επιλεγθέν στοιχείο σε ένα σετ δεδομένων, αν το στοιχείο ήταν τυχαία τοποθετημένο σε μια θεματική κατηγορία ανάλογα με την κατανομή των θεματικών κατηγοριών στο σετ δεδομένων. Το μέγεθος αυτό προσεγγίζει το 0 όταν όλα τα στοιχεία του κόμβου ανήκουν σε μια κλάση.

Για κάθε δέντρο απόφασης το scikit learn υπολογίζει τη σημαντικότητα ενός κόμβου χρησιμοποιώντας τη σημαντικότητα Gini (Gini impurity), υποθέτοντας την ύπαρξη μόνο δύο θυγατρικών κόμβων.

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)}$$

$N_{sub(j)}$ Η σημαντικότητα του κόμβου j.

$W_{sub(j)}$ Ο βεβαρημένος αριθμός των δειγμάτων που φτάνουν στον κόμβο j.

$C_{sub(j)}$ η τιμή impurity του κόμβου j.

$Left(j)$ ο θυγατρικός κόμβος από το split προς τα αριστερά.

$Right(j)$ ο θυγατρικός κόμβος από το split προς τα δεξιά.

Στη συνέχεια η σημαντικότητα για κάθε θεματική κατηγορία υπολογίζεται ως εξής:

$$fi_i = \frac{\sum_{j: \text{node } j \text{ splits on feature } i} ni_j}{\sum_{k \in \text{all nodes}} ni_k}$$

$F_{sub(j)}$ η σημαντικότητα του χαρακτηριστικού (θεματικής κατηγορίας (j)).

$N_{sub(j)}$ η σημαντικότητα του κόμβου j.

Μετά από αυτό γίνεται η κανονικοποίηση διαιρώντας με το άθροισμα των τιμών όλων των θεματικών κατηγοριών.

$$norm fi_i = \frac{fi_i}{\sum_{j \in \text{all features}} fi_j}$$

Η τελική σημαντικότητα σε επίπεδο random forest υπολογίζεται ως ο μέσος όρος από όλα τα δέντρα. Το άθροισμα των σημαντικοτήτων των θεματικών χαρακτηριστικών σε κάθε δέντρο υπολογίζεται και στην συνέχεια διαιρείται με τον αριθμό των δέντρων.

$RF_{sub(i)}$ η σημαντικότητα του χαρακτηριστικού I υπολογισμένο από όλα τα δέντρα

του RF.

$Norm_{sub}(ij)$ = Η κανονικοποιημένη σημαντικότητα για κάθε θεματική κατηγορία σε κάθε δέντρο j . T = αριθμός δέντρων.

Εν τέλει η μεθοδολογία με την οποία αποφασίζεται κάθε διαχωρισμός περιγράφεται παρακάτω:

- Υπολογίζεται το gini impurity για κάθε κόμβο.
- Αρχικά αναπτύσσονται πολλά classification decision trees σε κάθε ένα από τα οποία ανατίθεται ένα τμήμα των δεδομένων με προσπάθεια κάθε δέντρο να διαχειριστεί δείγματα από κάθε θεματική κατηγορία.
- Κάθε δέντρο ταξινομεί όλα τα εικονοστοιχεία τα οποία του αντιστοιχούν.
- Έτσι κάθε κλάση έχει έναν αριθμό ψήφων (votes).
- Η τελική και οριστική ταξινόμηση γίνεται με το δάσος να διαλέγει την κλάση με τις περισσότερες ψήφους.

Ο αλγόριθμος για εξαγωγή χαρακτηριστικών, που θα αναλυθεί σε επόμενο κεφάλαιο, επιλέγει τα κανάλια που έχουν παραπάνω από μια σημαντικότητα (ορισμένη από τον χρήστη), η οποία υπολογίζεται βάσει του gini impurity και μετά κάνει μετασχηματισμό των δεδομένων για να αποτελούνται μόνο από αυτά.

2.7.2 Πλεονεκτήματα και μειονεκτήματα των τυχαίων δασών

Τα κύρια πλεονεκτήματα των Random Forests είναι:

- Μπορούν να εκπαιδευτούν σε σύνολα δεδομένων υψηλής διάστασης χωρίς να εμφανίσουν σημαντικό βαθμό υπερεκπαίδευσης.
- Εξαιτίας του μεγάλου πλήθους δέντρων στο δάσος, το σφάλμα γενίκευσης είναι περιορισμένο. Αυτό έχει ως αποτέλεσμα τη μη εμφάνιση φαινομένων υπερεκπαίδευσης.
- Μη επαναληπτική διαδικασία εκπαίδευσης, ο αλγόριθμος ολοκληρώνεται σε σταθερό αριθμό βημάτων.
- Η τυχαία επιλογή ενός υποσυνόλου των χαρακτηριστικών για τη διαμέριση των παραδειγμάτων κάθε ενδιάμεσου κόμβου, ελαττώνει τη συσχέτιση ανάμεσα στα δέντρα και διατηρεί την πόλωση (bias) σε χαμηλά επίπεδα καθώς τα δέντρα αναπτύσσονται χωρίς κλάδεμα. Χρησιμοποιώντας ένα σύνολο δέντρων απόφασης μειώνεται και η διακύμανση (variance).
- Η διάσχιση ενός δέντρου από ένα παράδειγμα ξεκινώντας από τη ρίζα και καταλήγοντας σε έναν από τους τερματικούς κόμβους γίνεται λογαριθμικά ως προς το πλήθος των φύλλων του.
- Παρουσιάζουν ανεκτικότητα ως προς το θόρυβο και τα αριθμητικά σφάλματα στα δεδομένα εκπαίδευσης.
- Για την επαγωγή κάθε δέντρου περίπου το $\frac{1}{2}$ των δεδομένων δεν επιλέγεται για εκπαίδευση. Αυτά τα παραδείγματα καλούνται Out-of-Bag και μπορούν να χρησιμοποιηθούν για την

εκτίμηση της πιθανότητας σφάλματος, εξαλείφοντας την ανάγκη ύπαρξης ενός συνόλου ελέγχου ή της τεχνικής cross-validation.

- Παράγει μια εσωτερική αμερόληπτη εκτίμηση του σφάλματος γενίκευσης καθώς εξελίσσεται η διαδικασία κατασκευής του δέντρου.
- Υπάρχει δυνατότητα παράλληλης επαγωγής δέντρων σε αντίθεση με τη μέθοδο Boosting.
- Αναζητά τον καλύτερο διαχωρισμό σε ένα μικρό υποσύνολο των χαρακτηριστικών και δεν κάνει εξαντλητική αναζήτηση όπως ο αλγόριθμος Boosting.
- Μπορεί να χρησιμοποιηθεί για ομαδοποίηση.
- Επιτρέπει τη δημιουργία παραλλαγών της βασικής τεχνικής ως προς την κατασκευή του μοντέλου ταξινόμησης π.χ. χρήση διαφορετικών τεχνικών διαμέρισης των παραδειγμάτων των ενδιάμεσων κόμβων.

Τα τυχαία δάση παρουσιάζουν όμως και κάποια σημαντικά μειονεκτήματα ως προς την εφαρμογή τους:

- Υψηλό υπολογιστικό κόστος.
- Υπάρχει σημαντικό πλήθος ελεύθερων παραμέτρων τις οποίες πρέπει να προσδιορίσει ο χρήστης π.χ. πλήθος δέντρων, βαθμός κόμβων, πλήθος παραδειγμάτων εκπαίδευσης, συνθήκη τερματισμού διαμέρισης των κόμβων.
- Για την επέκταση ενός μοντέλου με στόχο της εισαγωγή μιας ακόμα κατηγορίας απαιτείται η κατασκευή του μοντέλου από την αρχή.
- Κάθε νέο παράδειγμα πρέπει να διασχίσει όλα τα δέντρα του δάσους για την εκτίμηση της κατηγορίας του.

ΚΕΦΑΛΑΙΟ 3 ΜΕΘΟΔΟΛΟΓΙΑ

3.1 Εργασίες Βάσης της Μεθοδολογίας

Η εργασία αυτή αποτελεί συνέχεια του μεθοδολογικού πλαισίου που αναπτύχθηκε στα πλαίσια του μαθήματος θέματος “Εφαρμογές - Φωτοερμηνείας -Τηλεπισκόπησης” της ΣΑΤΜ του συγγραφέως με τίτλο «Χαρτογράφηση Χρήσεων Καλύψεων γης εκτεταμένου μέρους της ελληνικής επικράτειας με ανάπτυξη αλγορίθμων για επιβλεπόμενη μηχανική μάθηση» καθώς επίσης και στη δημοσιευμένη εργασία στα ISPRS Archives “Assessing the contribution of spectral and temporal features for annual land cover and crop type mapping” [1]. Αυτές οι εργασίες αποτέλεσαν βάση για το μεθοδολογικό πλαίσιο το οποίο ακολούθως αναπτύχθηκε και εφαρμόστηκε σε περισσότερες περιοχές της ελληνικής επικράτειας. Πιο συγκεκριμένα στο πλαίσιο του μαθήματος θέματος έγινε η συλλογή των δεδομένων αναφοράς (ground truth) για το tile 34SEJ του δορυφόρου Sentinel-2 και αναπτύχθηκαν πολλαπλοί αλγόριθμοι επεξεργασίας για την προεπεξεργασία και ταξινόμηση μονοχρονικών και διαχρονικών πολυφασματικών δεδομένων. Στο πλαίσιο της προαναφερθείσας επιστημονικής δημοσίευσης, αναπτύχθηκε αλγοριθμική μεθοδολογία για την ταξινόμηση διαχρονικών πολυφασματικών δεδομένων με βάση φασματικά και χρονικά χαρακτηριστικά. Οι εν λόγω εργασίες συμπεριέλαβαν όλες τις κατάλληλες προεπεξεργασίες σε πολυφασματικές απεικονίσεις του tile 34SEJ του δορυφόρου Sentinel-2 ώστε να παραχθούν:

- Ένας υπερκύβος αποτελούμενος από δέκα πολυφασματικές απεικονίσεις του tile 34SEJ του δορυφόρου Sentinel 2A προερχόμενες από δέκα διαφορετικές ημερομηνίες για το έτος 2016 με κάθε μια να αποτελείται από τα κανάλια BLUE, GREEN, RED, NIR, RED EDGE, SWIR και τους δείκτες NDVI, NDWI, NDBI. Οι ακρίβειες που επιτεύχθηκαν από την ταξινόμηση αυτού του υπερκύβου των φασματικών χαρακτηριστικών είναι OA=94,03% και μέσο f1=83,47%.
- Ένας υπερκύβος χρονικών χαρακτηριστικών αποτελούμενος από τα στατιστικά μεγέθη Min, Max, Median, Mean, Standard Deviation για όλα τα φασματικά κανάλια και τους δείκτες του υπερκύβου όπως προαναφέρθηκε (Για περαιτέρω ανάλυση κεφάλαιο 3.3). Οι ακρίβειες που επιτεύχθηκαν για την ταξινόμηση του κύβου των χρονικών χαρακτηριστικών είναι OA=92,11% και η μέση ακρίβεια f1 είναι 81,72%.
- Ο πίνακας του ground truth ο οποίος είναι συμπληρωματικός των παραπάνω υπερκύβων.

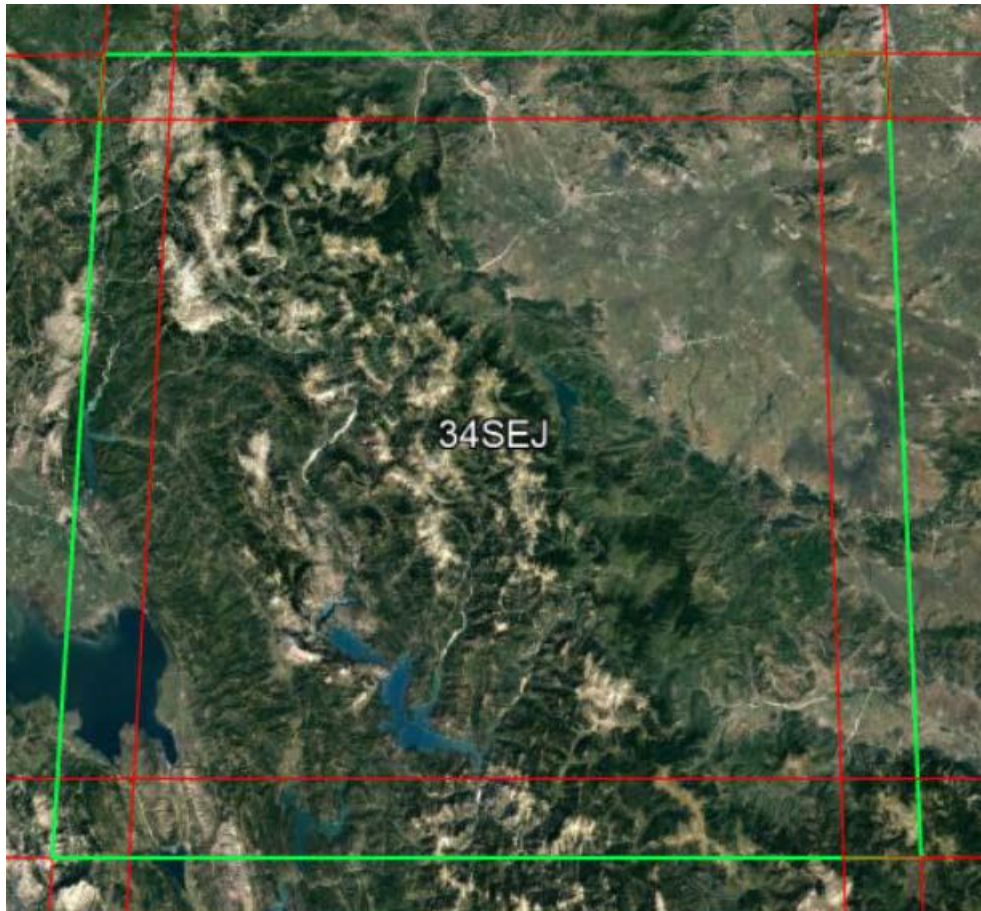
Τα τρία παραπάνω σετ δεδομένων αποτελούν τα βασικά δεδομένα εισόδου για τα αρχικά πειράματα στο tile 34SEJ και στη συνέχεια έγιναν διάφορες προσθήκες ή/και αφαιρέσεις επιπέδων πληροφορίας και εκτέλεση επιπλέον πειραμάτων. Οι επιλογές πειραμάτων για τα υπόλοιπα τρία tiles βασίστηκαν στην ανάλυση αποτελεσμάτων από το tile 34SEJ. Εκτενέστερη περιγραφή ακολουθεί στις επόμενες παραγράφους.

3.2 Δορυφορικά Δεδομένα - Περιοχή Μελέτης

Όπως προαναφέρθηκε, στόχος αυτής της εργασίας είναι η δημιουργία μοντέλων μηχανικής μάθησης με τη βοήθεια του random forest από δεδομένα εκπαίδευσης, μιας περιοχής του ελλαδικού χώρου, προερχόμενα από διαφορετικές προεπεξεργασίες, η αξιολόγηση των αποτελεσμάτων όπως αυτά προκύπτουν και η επιλογή κάποιων πειραμάτων, τα οποία συνδυάζουν με τον βέλτιστο τρόπο, υψηλές ακρίβειες (>90%) με τον χαμηλότερο δυνατό αριθμό καναλιών. Αυτά τα πειράματα στη συνέχεια πραγματοποιούνται σε άλλες περιοχές του Ελλαδικού χώρου και ακολούθως γίνεται γενίκευση των μοντέλων μηχανικής μάθησης από την μια περιοχή στην άλλη.

Ο δορυφόρος από τον οποίο ελήφθησαν τα παραπάνω δεδομένα είναι ο Sentinel 2A με τις περιοχές οι οποίες καλύπτονται να αφορούν τα tiles 34SEJ (tile βάσης στο οποίο θα γίνει και η μεγάλη πλειοψηφία των πειραμάτων, ενώ είναι τα δεδομένα τα οποία χρησιμοποιήθηκαν στις εργασίες που αναφέρθηκαν προηγουμένως), 34TEK, 34SEG και 35TLF τα οποία καλύπτουν πολλές διαφορετικές γεωγραφικές περιοχές με πολύ διαφορετικό μικροκλίμα και ανάγλυφο (από τη δυτική Πελοπόννησο μέχρι και τη Θράκη) οι οποίες είναι κατάλληλες για να ελεγχθούν οι ακρίβειες των μοντέλων μηχανικής μάθησης κατά τη γενίκευση ανάμεσα σε περιοχές ανόμοιες γεωμορφολογικά και κλιματολογικά.

Το tile 34SEJ του Sentinel-2 καλύπτει την κεντρική Ελλάδα. Η περιοχή μελέτης καλύπτει 10000 km² και συμπεριλαμβάνει τμήματα από τέσσερις διαφορετικές περιφερειακές ενότητες της χώρας. Αυτή η περιοχή επιλέχθηκε εξαιτίας της ετερογένειας τοπίων και μικροκλιμάτων τα οποία περιλαμβάνει και των διαφορετικών μορφών αναγλύφου, οι οποίες κυμαίνονται από τον ορεινό όγκο της Πίνδου μέχρι μεγάλες πεδινές εκτάσεις όπως αυτές στη Θεσσαλία. Η Πίνδος διασχίζει την περιοχή από Βόρεια προς Νοτιοανατολικά και παρουσιάζει καλύψεις γης όπως τα δάση κωνοφόρων, τα δάση φυλλοβόλων, τις χορτολιβαδικές εκτάσεις, τη σκληροφυλλική βλάστηση και τις άγονες εκτάσεις. Ανατολικά της Πίνδου η περιοχή μας διατρέχεται κατά ένα μέρος από τη Θεσσαλική πεδιάδα, μια από τις σημαντικότερες εκτάσεις για παραγωγή αγροτικών προϊόντων στην Ελλάδα. Οι αγροτικές εκτάσεις της περιοχής μελέτης αποτελούνται κυρίως από βαμβάκι, καλαμπόκι, δημητριακά, τριφύλλι και άλλες καλλιέργειες. Οι αστικές και ημιαστικές περιοχές αποτελούνται από τα κέντρα της Καρδίτσας και των Τρικάλων και από μικρότερα χωριά, τα οποία βρίσκονται διάσπαρτα ανάμεσα σε αυτά τα δύο κέντρα. Η περιοχή μελέτης συμπεριλαμβάνει και αρκετές υδάτινες εκτάσεις όπως τις τεχνητές λίμνες Πλαστήρα, Κρεμαστών και Πουρναρίου, ενώ στο νοτιοδυτικό άκρο φαίνεται και ένα μέρος του Αμβρακικού κόλπου.



Σχήμα 3: tile 34 SEJ

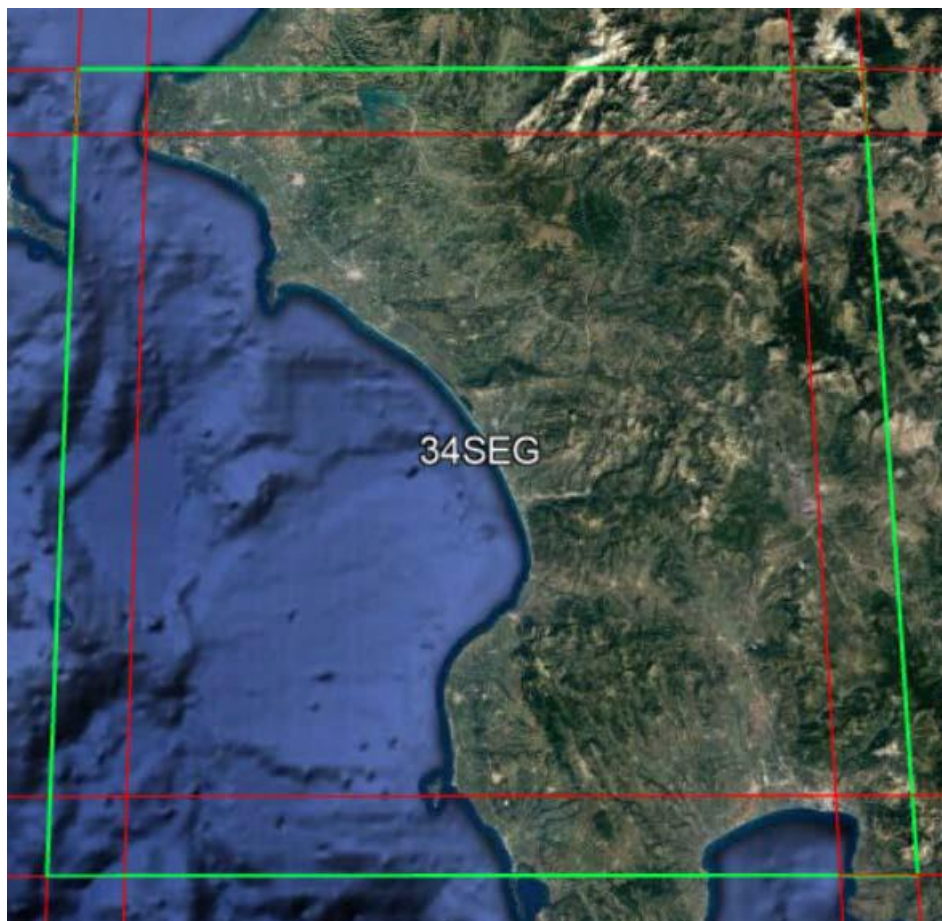
Το tile 34TEK του δορυφόρου καταλαμβάνει τμήμα της δυτικής Μακεδονίας, της βόρειας Θεσσαλίας, της Ηπείρου και συνορεύει προς νότο με το 34SEJ. Η έκταση αυτή είναι 1000 km^2 . Η περιοχή αυτή επιλέχθηκε ως περιοχή καθώς το γεγονός ότι η φορά του δορυφόρου που φιλοξενεί τον πολυφασματικό δέκτη είναι από βορρά προς νότο, συνεπάγεται ότι υπάρχουν για τις ίδιες ημερομηνίες πολυφασματικές απεικονίσεις, πράγμα το οποίο είναι σημαντικό κατά τη μεταφορά των πειραμάτων σχετικών με τον υπερκύβο, ο οποίος αποτελείται από μια στοίβα δέκα πολυφασματικών απεικονίσεων. Ακόμα η γεωγραφική εγγύτητα συνεπάγεται ότι οι πολυφασματικές απεικονίσεις έχουν ληφθεί με πολύ μικρή χρονική απόκλιση και άρα εάν δεν υπάρχουν τοπικές νεφώσεις και κακοκαιρία οι εικόνες που έχουν μικρή νεφοκάλυψη στο 34SEJ αναμένεται να έχουν μικρή νεφοκάλυψη και στο tile 34TEK. Επιπλέον η διαφορά που υπάρχει ανάμεσα στο γεωγραφικό πλάτος μεταξύ των tiles συνεπάγεται κάποιες διαφορές στο μικροκλίμα και άρα στις καλύψεις γης, οι οποίες είναι σημαντικές για να δοκιμαστεί η αποτελεσματικότητα των πειραμάτων σε διαφορετικές συνθήκες. Πιο συγκεκριμένα, την περιοχή κεντρικά διατρέχει η οροσειρά της βόρειας Πίνδου με σημαντικά αστικά κέντρα αυτά της Κοζάνης και της Πτολεμαΐδας και διάσπαρτους πολλούς οικισμούς σε όλη την έκταση. Η περιοχή διατρέχεται από αρκετά ποτάμια, με δεσπόζοντα τον Αλιάκμονα, καθώς και ένα μεγάλο μέρος της Εγνατίας Οδού. Επίσης εντός της περιοχής βρίσκεται και η λίμνη του Πολυφύτου. Επικρατούσες μορφές βλάστησης είναι πρωτίστως τα φυλλοβόλα και τα κωνοφόρα δέντρα, πράγμα

που αναμένεται σε αυτό το γεωγραφικό μήκος. Η περιοχή διατρέχεται από πλήθος μεγάλων ποταμών. Τέλος οι κυριότερες καλλιέργειες της περιοχής είναι τα σιτηρά, ο αραβόσιτος και οι ζωοτροφές.



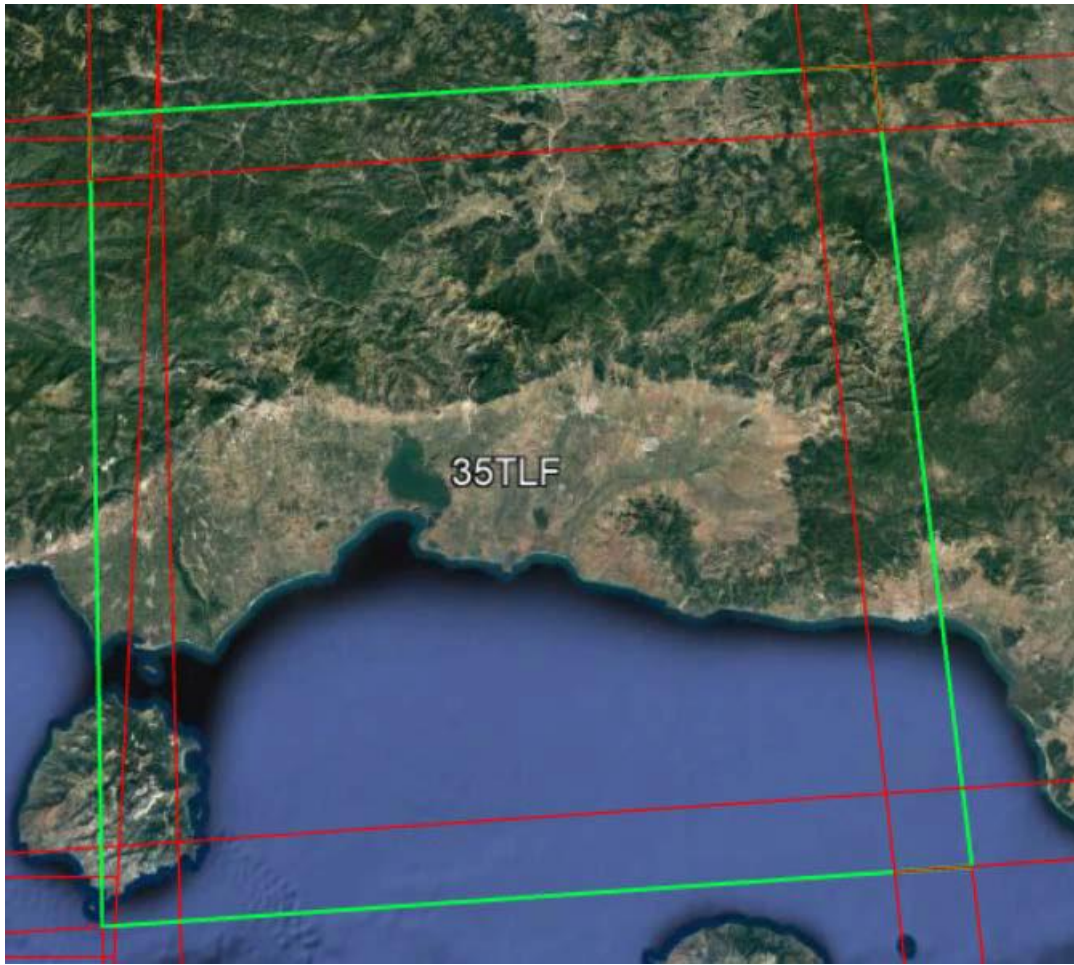
Σχήμα 4: tile 34 TEK

Το tile 34SEG του δορυφόρου καταλαμβάνει κυρίως το δυτικό μέρος της Πελοποννήσου με έκταση 1000 km² ενώ ένα μεγάλο μέρος του tile περιλαμβάνει το Ιόνιο πέλαγος. Η έκταση αυτή καλύπτει ίδια γεωγραφικά μήκη αλλά αρκετά διαφορετικά γεωγραφικά πλάτη σε σχέση με το 34SEJ. Αυτό συνεπάγεται ότι θα υπάρχουν σημαντικές διαφορές ως προς τη χλωρίδα και την πανίδα της περιοχής και άρα, όπως και στην προηγούμενη περίπτωση, η περιοχή προσφέρεται για τον έλεγχο της δυνατότητας μεταφοράς πειραμάτων σε διαφορετικές περιοχές του ελλαδικού χώρου. Πιο συγκεκριμένα οι κυρίαρχες καλλιέργειες στη δυτική Πελοπόννησο είναι τα ελαιόδεντρα και τα καρποφόρα εσπεριδοειδή. Η περιοχή δεν έχει σημαντικό αριθμό άλλων καλλιεργειών (όπως δημητριακά, καπνό ή βαμβάκι) οι οποίες εμφανίζονται σε μεγάλες συγκεντρώσεις στον υπόλοιπο ελλαδικό χώρο. Επιπλέον υπάρχει μεγάλος αριθμός θερμοκηπίων στην περιοχή, ο οποίος εκμεταλλεύεται τις αρκετές ημέρες ηλιοφάνειας του χρόνου. Το μόνο σημαντικό αστικό κέντρο της περιοχής είναι ο Πύργος της Ηλείας και στο κάτω αριστερό άκρο η Καλαμάτα, ενώ υπάρχουν αρκετές κωμοπόλεις και χωριά διάσπαρτα ανάμεσα στις διαφορετικές περιοχές. Το κύριο είδος βλάστησης το οποίο φύεται στην περιοχή είναι αυτό της σκληροφυλλικής βλάστησης (θαμνώδεις εκτάσεις). Τέλος στο tile αυτό φαίνεται και ολόκληρος ο κόλπος της Κυπαρισσίας.



Σχήμα 5: tile 35 SEG

Το tile 35TLF του δορυφόρου καταλαμβάνει την περιοχή της ανατολικής Μακεδονίας και της Θράκης με έκταση και πάλι 1000km^2 . Η έκταση αυτή δεν έχει το ίδιο γεωγραφικό μήκος με τα υπόλοιπα tiles αλλά το γεωγραφικό πλάτος είναι αρκετά όμοιο. Αυτό κατέστησε αδύνατη την εύρεση απεικονίσεων στις ίδιες ημερομηνίες, καθώς όπως προαναφέρθηκε η τροχιά του δορυφόρου είναι στην φορά από βορρά προς νότο και άρα μεσολαβεί ένα διάστημα ανάμεσα στη λήψη των πολυφασματικών απεικονίσεων των διαφορετικών tile. Η περιοχή που καταλαμβάνεται από το tile είναι κυρίως θάλασσα ή εκτός των συνόρων της Ελλάδας, ενώ δεσπόζουσα θέση στην περιοχή έχει και η οροσειρά της Ροδόπης. Τα πιο σημαντικά αστικά κέντρα της περιοχής είναι της Ξάνθης και της Κομοτηνής. Οι κύριες καλλιέργειες είναι ο αραβόσιτος, τα σιτηρά και το βαμβάκι αλλά και τα ελαιόδεντρα. Η περιοχή έχει πολλά είδη βλάστησης που μπορούσαν να θεωρηθούν επικρατούντα, από τα κωνοφόρα και τα φυλλοβόλα στην οροσειρά της Ροδόπης, έως και τα στοιχεία της πυκνής σκληροφυλλικής και αραιής σκληροφυλλικής βλάστησης νότια του ορεινού όγκου της Ροδόπης. Τέλος η περιοχή περιλαμβάνει πολλά μικρά ποτάμια και τη λίμνη Βιστονίδα.



Σχήμα 6:tile TLF

Παρακάτω φαίνονται και τα παραπάνω tiles στη σχετική τους θέση σε σχέση με την υπόλοιπη Ελλάδα



Σχήμα 7: Θέση των tiles στην Ελληνική επικράτεια

3.3 Συλλογή Διαχρονικών Δορυφορικών Δεδομένων









































Για τη δόμηση του υπερκύβου για κάθε ένα από τα tiles αποφασίστηκε να συλλεχθούν δέκα πολυφασματικές απεικονίσεις (συνολικά σαράντα πολυφασματικές απεικονίσεις). Οι απεικονίσεις αυτές αποφασίστηκε να προέρχονται όλες από το έτος 2016. Το χρονικό διάστημα που έχει μεσολαβήσει από το 2016 έως σήμερα, κατέστησε αρκετά δύσκολη την εύρεση των απεικονίσεων, με αποτέλεσμα την αναζήτηση των επιθυμητών απεικονίσεων από τρεις διαφορετικές πηγές.

- Το αρχείο δορυφορικών απεικονίσεων της Αμερικάνικης Γεωλογικής Υπηρεσίας (USGS) <https://earthexplorer.usgs.gov/>
- Πολυφασματικές απεικονίσεις από το αρχείο του εργαστηρίου τηλεπισκόπησης.
- Πολυφασματικές απεικονίσεις από την ιστοσελίδα αναζήτησης του ευρωπαϊκού προγράμματος χαρτογράφησης Copernicus <https://scihub.copernicus.eu/dhus/#/home>.

Τα κριτήρια τα οποία τέθηκαν για τις πολυφασματικές απεικονίσεις ήταν τα εξής:

- Το ποσοστό νεφοκάλυψης να μην ξεπερνά σε καμία περίπτωση το 30%, καθώς σε τέτοια περίπτωση η αναγνώριση των χαρακτηριστικών του εδάφους δυσχεραίνεται σε μεγάλο βαθμό.
- Οι απεικονίσεις να καλύπτουν ολόκληρη την έκταση του tile (μη ύπαρξη στοιχείων με τιμή no data) ή να υπάρχει δυνατότητα για σύνθεση εικόνων εκεί που λείπει η πληροφορία.
- Οι ημερομηνίες συλλογής των απεικονίσεων να διατρέχουν κατά το δυνατόν όλο το έτος με ελάχιστη προϋπόθεση την ύπαρξη κατ' ελάχιστο μιας πολυφασματικής απεικόνισης για κάθε εποχή.

Επιλέχθηκαν τελικά οι λήψεις για τα tile που φαίνονται παρακάτω:

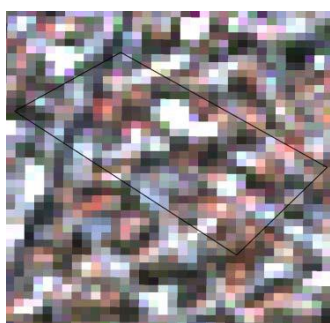
	01	02	03	04	05	06	07	08	09	10
34SEJ	 2016/12/10	 2016/10/01	 2016/08/22	 2016/08/12	 2016/08/02	 2016/07/23	 2016/07/13	 2016/07/03	 2016/06/23	 2016/04/04
34TEK	 2016/12/10	 2016/10/01 ΚΑΙ 2016/10/01	 2016/08/22 και 2016/08/22	 2016/12/10	 2016/08/02	 2016/07/23	 2016/07/13	 2016/07/03	 2016/06/23	 2016/04/04
34SEG	 2016/12/10	 2016/10/01	 2016/08/22	 2016/08/02	 2016/07/23	 2016/07/13	 2016/07/03	 2016/06/23	 2016/04/04	 2016/01/25
35TFL	 2016/12/17	 2016/11/17	 2016/09/28	 2016/09/18	 2016/09/08	 2016/08/19	 2016/07/10	 2016/06/20	 2016/05/01	 2016/04/21

Σχήμα 8: Επιλεγμένες εικόνες

3.4 Κατηγορίες Ταξινόμησης και Δεδομένα Αναφοράς (Ground Truth)

Η κατηγοριοποίηση των τάξεων που θα χρησιμοποιηθούν στην ταξινόμηση έγινε με βάση το πρότυπο του Corine Land Cover Level 3. Αρχικά αφαιρέθηκαν κατηγορίες που δεν απαντώνται στην περιοχή (πχ πάγοι), στη συνέχεια συμπτυχθήκαν κατηγορίες που δεν είχαν αρκετά διακριτές φασματικές υπογραφές ή δεν μπορούσε να ληφθεί αρκετά μεγάλος αριθμός πολυγώνων εκπαίδευσης ώστε να μπορέσουν να ταξινομηθούν ως αυτόνομη κατηγορία. Εν τέλει προστέθηκαν νέες κατηγορίες (φωτοβολταϊκά και θερμοκήπια) και σε κάποιες περιπτώσεις κατηγορίες αναλύθηκαν περαιτέρω. Με αυτόν τον τρόπο προέκυψαν τριάντα μία θεματικές κατηγορίες, με τις οποίες αντιστοιχίστηκαν οι φασματικές κατηγορίες των απεικονίσεων, βάσει των οποίων συλλέχθηκαν τα δεδομένα εκπαίδευσης και ελέγχου.

Η επιλογή επαρκούς αριθμού δειγμάτων εκπαίδευσης και δειγμάτων ελέγχου, καθώς και η αντιπροσωπευτικότητά τους είναι πολύ κρίσιμη για τη σωστή ταξινόμηση [12]. Για την παρούσα εργασία τα δεδομένα αναφοράς (ground truth) είναι ιδιαιτέρως εκτεταμένα αφού αφορούν τέσσερα διαφορετικά tile του Sentinel-2. Τα πολύγωνα για τις κατηγορίες κάλυψης γης και καλλιιεργειών προήλθαν από τις βάσεις δεδομένων του εργαστηρίου Τηλεπισκόπησης και συγκεκριμένα από ψηφιοποιημένα δεδομένα για το 2016 από τις εργασίες της Υποψηφίου Διδάκτορος Χριστίνας Καρακίζη και άλλες διπλωματικές ή μεταπτυχιακές εργασίες ([13], [14], [15]) καθώς και την εργασία του μαθήματος θέματος του συγγραφέα. Το tile αναφοράς ήταν το 34SEJ για το οποίο ψηφιοποιήθηκαν οι τριάντα μία κατηγορίες που παρουσιάζονται ακολούθως, μαζί με αντιπροσωπευτικά πολύγωνα εκπαίδευσης για κάθε κατηγορία:



Εικόνα 1: Πυκνή Αστική δόμηση



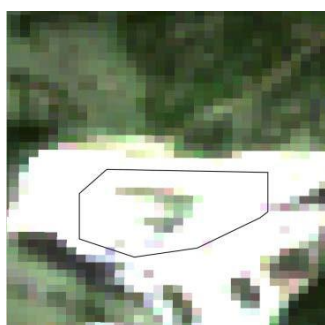
Εικόνα 2: Αραιή αστική δόμηση



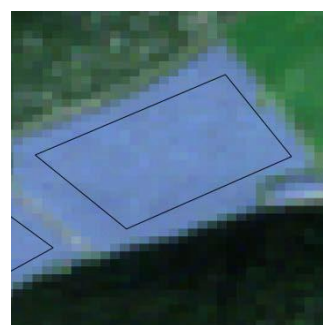
Εικόνα 3: Αραιή αστική δόμηση



Εικόνα 4: Ασφαλτοστρωμένα-Σκυροδετημένα δίκτυα



Εικόνα 5: Μονάδα εξαγωγής Ορυκτών



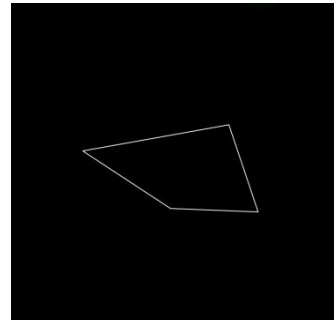
Εικόνα 6: Φωτοβολταϊκά πάνελ



Εικόνα 7: Θερμοκήπια



Εικόνα 8: Φυλλοβόλα δέντρα



Εικόνα 9: Κωνοφόρα δέντρα



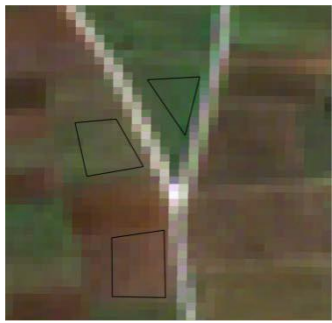
Εικόνα 10: Χορτολιβαδικές εκτάσεις



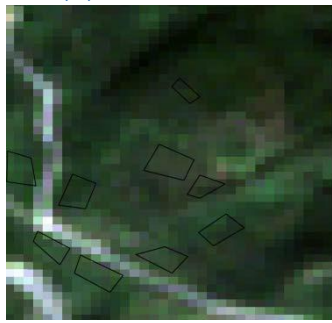
Εικόνα 11: Πυκνή Σκληροφυλλική Βλάστηση



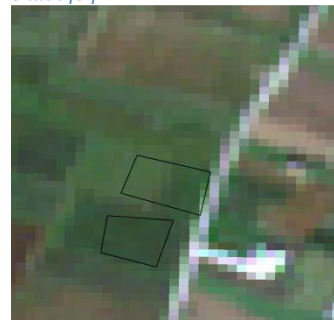
Εικόνα 12: Αραιή Σκληροφυλλική βλάστηση



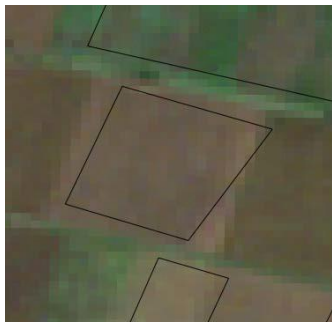
Εικόνα 13: Αμπελώνες



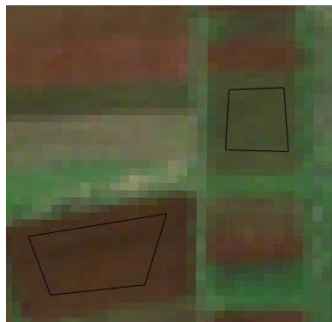
Εικόνα 14: Ελαιώνες



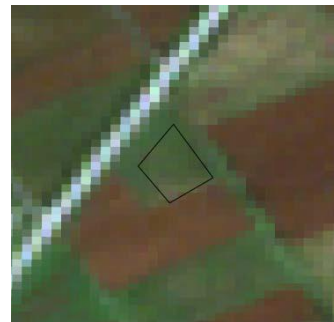
Εικόνα 15: Δέντρα φρούτων



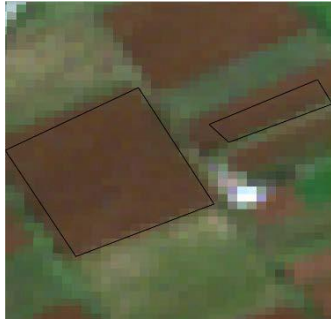
Εικόνα 16: Σιτηρά



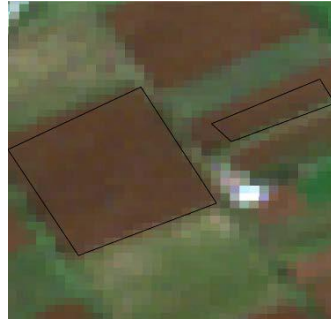
Εικόνα 17: Κριθάρι



Εικόνα 18: Βρώμη



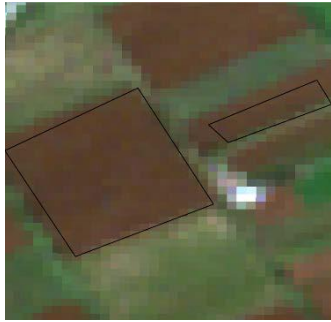
Εικόνα 19: Αραβόσιτος



Εικόνα 20: Βαμβάκι



Εικόνα 21: Καπνός



Εικόνα 22: Τριφύλλι



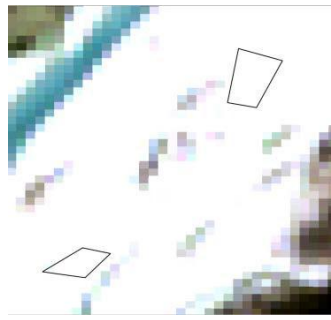
Εικόνα 23: Ζωοτροφές



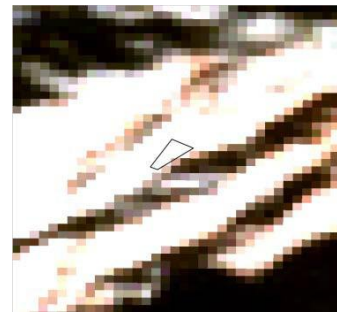
Εικόνα 24: Αγρανάπαυση



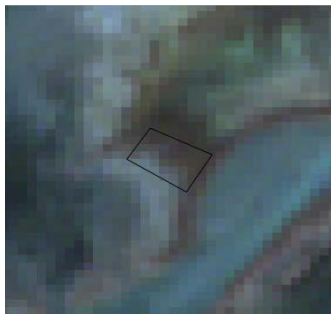
Εικόνα 25: Αραιή βλάστηση



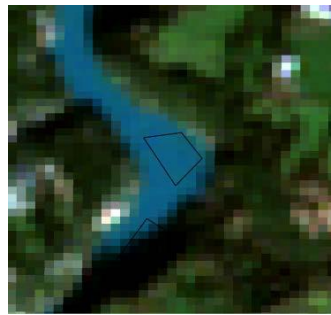
Εικόνα 26: Αμμώδης έκταση



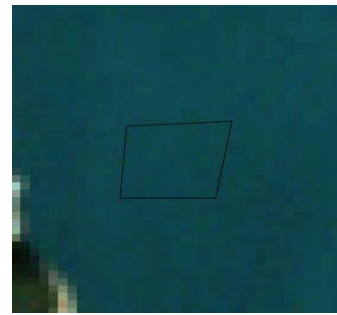
Εικόνα 27: Βραχώδης έκταση



Εικόνα 28: Βαλτώδης έκταση



Εικόνα 29: Ποτάμια



Εικόνα 30: Λίμνη



Εικόνα 31: Παράκτιο νερό

Σχήμα 9: Χαρακτηριστικές περιπτώσεις πολυγώνων εκπαίδευσης όπως φαίνονται από τις πολυφασματικές απεικονίσεις του Sentinel-2A

Οι παραπάνω είναι οι κατηγορίες για τις οποίες συλλέχθηκαν δεδομένα. Ωστόσο αυτές οι κατηγορίες είναι αυτές στις οποίες ταξινομήθηκαν τα δεδομένα μόνο για το tile 34SEJ. Για τα άλλα τρία tiles η συλλογή του ground truth έγινε με κάποιες συγχωνεύσεις μεταξύ κατηγοριών, με τις οποίες οι κατηγορίες έγιναν 28 από 31. Οι συγχωνεύσεις αφορούν τις κατηγορίες του σιταριού, του κριθαριού και της βρώμης, οι οποίες συγχωνεύτηκαν στην κατηγορία των δημητριακών και το τριφύλλι και τις ζωοτροφές που συγχωνεύτηκαν στην κατηγορία των ζωοτροφών.

Τα δεδομένα αναφοράς συλλέχθηκαν με τη βοήθεια του λογισμικού QGIS και στη συνέχεια συγχωνεύθηκαν σε ένα διανυσματικό (vector) αρχείο με τον κωδικό της κάθε κατηγορίας να διατηρείται. Στη συνέχεια τα διανυσματικά αυτά αρχεία μετατράπηκαν σε μορφή raster και έτσι τα δεδομένα μετατράπηκαν σε ένα γεωναφερμένο πίνακα 10980X10980 με κάθε στοιχείο να αντιστοιχεί σε μια κατηγορία ή να είναι 0. Αυτό είναι το ground truth από το οποίο θα εξαχθούν οι θεματικές κατηγορίες για εκπαίδευση και έλεγχο.

Τα παραπάνω φαίνονται και πιο εύληπτα στον παρακάτω πίνακα με τις κατηγορίες και τα πολύγωνα εκπαίδευσης για κάθε κατηγορία:

ΚΑΤΗΓΟΡΙΕΣ ΚΑΙ ΠΟΛΥΓΩΝΑ ΕΚΠΑΙΔΕΥΣΗΣ GROUND TRUTH ANA TILE

CATEGORY	CODE	NAME	34SEG	34TEK	35TFL	34SEJ	
Artificial	1	Continuous urban fabric	101	67	332	41	
	2	Discontinuous urban fabric	665	138	282	171	
	3	Industrial and commercial units	80	72	191	84	
	4	Road/Asphalt networks	53	49	282	77	
	5	Mineral extraction sites	19	33	176	18	
	6	Photovoltaic units	50	44	42	74	
	7	Greenhouses	108	46	9	46	
Forest	8	Broad-leaved forest	103	1069	3505	226	
	9	Coniferous forest	111	489	3441	279	
Grass and Shrub lands	10	Natural grasslands	181	301	939	292	
	11	Dense sclerophyllous	564	287	1061	646	
	12	Sparse sclerophyllous	347	246	600	327	
Crops	13	Vineyards	234	142	64	83	
	14	Olive groves	1001	62	468	187	
	15	Citrus Fruit trees	201	316	52	104	
	16	Cereals	0	427	388	170	Wheat
						31	Barley
						11	Oat
	17	Maize	87	103	413	244	
	18	Cotton	1	0	426	311	
	19	Tobacco	0	62	51	33	
	20	Grass Fodders	45	392	169	22	Clover
		45	392	169	280	Other Grass Fodders	
21	Fallow	0	77	0	23		
Bareland	22	Sparsely vegetated areas	84	317	1000	244	
	23	Beaches, dunes, sands	138	109	157	233	
	24	Bare rocks	93	99	471	189	
Marshes	25	Marshes	19	32	37	43	
Water	26	Water courses	86	124	15	222	
	27	Water bodies	24	51	160	141	
	28	Coastal water	50	0	87	23	

Πίνακας 1: Κατηγορίες Ground Truth για τις πολυφασματικές απεικονίσεις και αριθμός πολυγώνων εκπαίδευσης για κάθε tile

3.5. Επεξεργασίες Προετοιμασίας Δορυφορικών Δεδομένων για την Ταξινόμηση

3.5.1 Προεπεξεργασίες-Διορθώσεις δορυφορικών δεδομένων Sentinel-2A

Η μεθοδολογία η οποία ακολουθήθηκε κατά τη διάρκεια των προεπεξεργασιών είναι αυτή που χρησιμοποιήθηκε για τις αντίστοιχες προεπεξεργασίες στην εργασία “ASSESSING THE CONTRIBUTION OF SPECTRAL AND TEMPORAL FEATURES FOR ANNUAL LAND COVER AND CROP TYPE MAPPING” (Karakizi et al)[1].

Πιο συγκεκριμένα πραγματοποιήθηκαν οι εξής διορθώσεις με βάση το μεθοδολογικό πλαίσιο ανεπτυγμένο από τον υποψήφιο διδάκτορα της ΣΑΤΜ, Αριστείδη Βαϊόπουλο, για τα τέσσερα σετ δεδομένων ανά tile, με βάση :

- Ατμοσφαιρική διόρθωση από τον αλγόριθμο SEN2COR προκειμένου να παραχθούν προϊόντα ανακλαστικότητας επιπέδου 2C.
- Διόρθωση BRDF όπου οι τιμές ανακλαστικότητας των εικονοτοιχείων προσαρμόστηκαν ώστε οι λήψεις από τον δορυφόρο να μετατραπούν σε ναδιριαίες
- Διόρθωση νεφών και σκιών με την βοήθεια του αλγορίθμου Fmask όπου στα εικονοστοιχεία με νέφη/σκιές παρεμβλήθηκαν τιμές από τις εγγύτερες χρονικά εικόνες.
- Διόρθωση των σφαλμάτων μεταβολής γεωεντοπισμού ανάμεσα στις διαφορετικές ημερομηνίες.
- Pansharpning για τα κανάλια με αναλύσεις 20 και 60 m ώστε να έχουν ανάλυση 10 m.

Εδώ είναι σημαντικό να αναφερθεί το γεγονός ότι για το tile 34SEG στην χρονολογία 2016 ο αλγόριθμος SEN2COR δεν είχε επαρκή δεδομένα ώστε να πραγματοποιηθεί ατμοσφαιρική διόρθωση σε όλες τις πολυφασματικές απεικονίσεις του Sentinel 2A και έτσι η χρονοσειρά των εικόνων για το tile 34SEG αποτελείται από εννιά εικόνες, κάτι που θα περιορίσει το εύρος των δυνατών πειραμάτων που θα περιλαμβάνουν το tile 34SEG.

3.5.2 Επιλογή καναλιών, κανονικοποίηση φασματικών τιμών, υπολογισμός δεικτών και σχηματισμός υπερκύβου φασματικών χαρακτηριστικών

Με τη βοήθεια ανεπτυγμένου αλγορίθμου έγινε η διαδικασία της κανονικοποίησης των φασματικών τιμών (ανακλαστικότητας) των επιλεγμένων καναλιών στο διάστημα 0-1 τα οποία είναι το Blue, Green, Red, NIR, Red Edge, SWIR. Αυτό γίνεται με τη διαίρεση των τιμών των ανακλαστικότητας με την τιμή 1000, την μετατροπή όσων τιμών ξεπερνούν το 1000 σε 1 και όσων είναι κάτω του 0 σε 0.

Ακολούθησε ο υπολογισμός, κανονικοποίηση στο διάστημα 0-1 και στη συνέχεια προσθήκη των φασματικών δεικτών NDVI, NDWI, NDBI με τη βοήθεια του παραπάνω αλγορίθμου. Τον υπολογισμό και την κανονικοποίηση των δεικτών ακολουθεί η πρόσθεση των καναλιών μεταξύ τους, ώστε η σειρά των καναλιών να είναι Blue, Green, Red, NIR, Red Edge, SWIR, NDVI, NDWI, NDBI της απεικόνισης της μίας ημερομηνίας μετά την άλλη. Η χρονική σειρά με την οποία προστίθενται τα δεδομένα είναι από

τον 1ο του 2016 προς τον 12ο (ημερολογιακά). Το μέγεθος κάθε ένα από τους τρεις κύβους με τις δέκα απεικονίσεις ανέρχεται σε 43,4 GB, ενώ για τον κύβο του 34SEG ανέρχεται 39,1 GB.

3.5.3 Προετοιμασία δεδομένων για ταξινόμηση μηχανικής μάθησης

Επόμενο στάδιο είναι η προετοιμασία των δεδομένων για την είσοδό τους στις διαδικασίες ταξινόμησης μηχανικής μάθησης. Η ταξινόμηση μέσω της δημιουργίας μοντέλου μηχανικής μάθησης με τη βοήθεια των τυχαίων δέντρων βασίζεται στην εκπαίδευση ενός μαθηματικού μοντέλου με κάποια χειροκίνητα συλλεχθέντα δεδομένα και τον έλεγχο της ακρίβειας του συγκεκριμένου μαθηματικού μοντέλου με τη βοήθεια, επίσης χειροκίνητα συλλεχθέντων, δεδομένων ελέγχου. Για κάθε μία από αυτές τις δύο κατηγορίες δεδομένων απαιτούνται δύο σύνολα. Το πρώτο είναι το σύνολο των διανυσμάτων των χαρακτηριστικών τιμών των εικονοστοιχείων (θα το συμβολίζουμε με X) το οποίο θα πρέπει να περιέχει πληροφορία μόνο για τα εικονοστοιχεία των οποίων την κατηγορία γνωρίζουμε. Η πληροφορία για τη θέση αυτών των στοιχείων βρίσκεται στην εικόνα του ground truth. Η πληροφορία που σχετίζεται με τη φασματική απόκριση των παραπάνω στοιχείων βρίσκεται στον ίδιο τον υπερκύβο. Το δεύτερο είναι ένα διάνυσμα στήλη (θα το συμβολίζουμε y) το οποίο περιέχει την κατηγορία κάλυψης γης που αντιστοιχεί σε κάθε εικονοστοιχείο.

Για την προεπεξεργασία και τη δημιουργία των συνόλων που αναφέρθηκαν παραπάνω (διαχωρισμός δεδομένων εκπαίδευσης και ελέγχου) χρησιμοποιήθηκαν αλγόριθμοι οι οποίοι αναπτύχθηκαν ώστε από τα πολύγωνα από τα οποία θα εξαχθούν τα δεδομένα εκπαίδευσης να μην εξαχθούν και τα δεδομένα ελέγχου, προκειμένου να αποφευχθεί η υπερμοντελοποίηση (overfitting). Με αυτό τον αλγόριθμο εξήχθησαν, ανά θεματική κατηγορία όπως περιγράφονται από το ground truth, περιοχές εκπαίδευσης από συγκεκριμένα πολύγωνα που τους δόθηκε ένας αύξοντας κωδικός και στη συνέχεια κάθε πολύγωνο του οποίου ο αύξοντας κωδικός είχε χρησιμοποιηθεί για την εκπαίδευση διαγράφηκε από τη λίστα των πολυγώνων που χρησιμοποιήθηκαν για τον έλεγχο (parcel split). Το script δομήθηκε κατά τέτοιο τρόπο ώστε αντί ενός ακριβούς ποσοστού διαχωρισμού μεταξύ δεδομένων εκπαίδευσης ελέγχου, να δοθεί ένα εύρος στο οποίο θα πρέπει αυτά να κινούνται. Αυτό το ποσοστό ορίστηκε να είναι 65-70% για τα δεδομένα εκπαίδευσης και 30-35% για τα δεδομένα ελέγχου αντιστοίχως. Οι θέσεις των πολυγώνων εκπαίδευσης κρατήθηκαν σε πίνακες με τις διαστάσεις των εικόνων, προκειμένου να υπάρχει η δυνατότητα εκπαίδευσης μοντέλων για πολλές διαφορετικές πολυφασματικές απεικονίσεις και με οποιαδήποτε μεθοδολογία, με τα ίδια ακριβώς πολύγωνα εκπαίδευσης-ελέγχου. Επιπλέον εντάχθηκαν εντός του αλγορίθμου και κάποιες άλλες απαραίτητες διεργασίες προεπεξεργασίας πριν μπορέσει να γίνει η εκπαίδευση του μοντέλου και οι οποίες βασίστηκαν σε αντίστοιχους αλγορίθμους που συντάχτηκαν από μέλη του εργαστηρίου. Αυτές ήταν οι εξής:

- Μετατροπή των πολυφασματικών απεικονίσεων σε διανύσματα. (Μείωση των διαστάσεων κατά μία.)
- Αφαίρεση όλων εκείνων των εικονοστοιχείων των πολυφασματικών απεικονίσεων για τα

οποία δεν υπάρχει ground truth. (Δεν καλύπτονται από πολύγωνα εκπαίδευσης ή ελέγχου.)

- Τυχαία αναδιάταξη των στοιχείων των δημιουργημένων διανυσμάτων ώστε να έχει νόημα η εκπαίδευση του ταξινομητή.

Στη συνέχεια συντάχθηκε αλγόριθμος με σκοπό οι θέσεις των δεδομένων που προέκυψαν πριν για την εκπαίδευση του μοντέλου μηχανικής μάθησης να χρησιμοποιηθούν στη συνέχεια για να εκπαιδευτούν και όλα τα υπόλοιπα μοντέλα του συγκεκριμένου tile. Οι πίνακες με τις συγκεκριμένες θέσεις εφαρμόστηκαν σαν μάσκες σε κάθε άλλη πολυφασματική απεικόνιση για την οποία θελήσαμε να εκπαιδεύσουμε μοντέλο μηχανικής μάθησης. Με αυτό τον τρόπο εκπαιδεύτηκαν όλα τα μοντέλα μηχανικής μάθησης με το ίδιο ακριβώς ποσοστό δεδομένων και κυρίως με τα ίδια πολύγωνα εκπαίδευσης και ελέγχου. Έτσι επετεύχθη συνέχεια ως προς τα αποτελέσματα, χωρίς η ακρίβεια να επηρεάζεται άμεσα από μεταβολές στα δεδομένα εισόδου. Όλοι οι προαναφερθέντες κώδικες φαίνονται και στο παράρτημα.

Στη συνέχεια ήδη δομημένοι αλγόριθμοι τροποποιήθηκαν προκειμένου τα δεδομένα εκπαίδευσης και ελέγχου να συλλεχθούν για το DSM και τις φωτεινότητες και αυτά να προστεθούν στα ήδη υπάρχοντα δεδομένα εκπαίδευσης και ελέγχου για τον υπερκύβο της χρονοσειράς αλλά και για όλους τους υπερκύβους των χρονικών μεγεθών. Ομοίως δομήθηκε και νέος υπερκύβος ο οποίος περιελάμβανε το DSM μαζί με τα υπάρχοντα κανάλια του υπερκύβου αλλά και τη φωτεινότητα.

Με όμοιο τρόπο εισήχθησαν και τα δεδομένα εκπαίδευσης και ελέγχου στα αντίστοιχα διανύσματα του υπερκύβου. Ακολούθως έγινε δεύτερος κύκλος πειραμάτων κατά τον οποίο χρησιμοποιήθηκαν τόσο η φωτεινότητα όσο και το ψηφιακό μοντέλο εδάφους. Για αυτά τα νέα πειράματα χρησιμοποιήθηκαν ως επί το πλείστον ήδη δημιουργηθέντες αλγόριθμοι, προσαρμοσμένοι στα νέα δεδομένα.

3.5.4 Προσθήκη DSM

Ακολούθησαν πειράματα με την εισαγωγή υψομετρικής πληροφορίας στα δεδομένα με τα οποία εκπαιδεύτηκε το μοντέλο. Γι' αυτό τον σκοπό συλλέχθηκε επιπλέον των ήδη χρησιμοποιούμενων πολυφασματικών απεικονίσεων το DSM "ALOS World 3D 30m" το οποίο είναι ένα παγκόσμιο ψηφιακό μοντέλο επιφανείας με χωρική ανάλυση (οριζοντιογραφικά) 30 m. Η συλλογή των δεδομένων έγινε από τον πανχρωματικό δέκτη για στερεοχαρτογράφηση (PRISM), ο οποίος είναι ένας οπτικός δέκτης στον δορυφόρο ALOS.

Καταρχάς το συγκεκριμένο παγκόσμιο DSM λήφθηκε από την ιστοσελίδα <https://www.eorc.jaxa.jp/ALOS/en/aw3d30/index.htm>. Σε πρώτη φάση εργαστήκαμε στο tile 34SEJ. Επιλέχθηκε αρχικά η περιοχή του ενδιαφέροντός μας, η οποία περιλάμβανε το μεγαλύτερο μέρος της ελληνικής επικράτειας και στη συνέχεια κατέβηκε ένα υποσύνολο αυτής της περιοχής. Στη συνέχεια το τμήμα του DSM που κατεβάσαμε εισήχθη σε πρόγραμμα GIS και εκεί έγιναν οι ακόλουθες

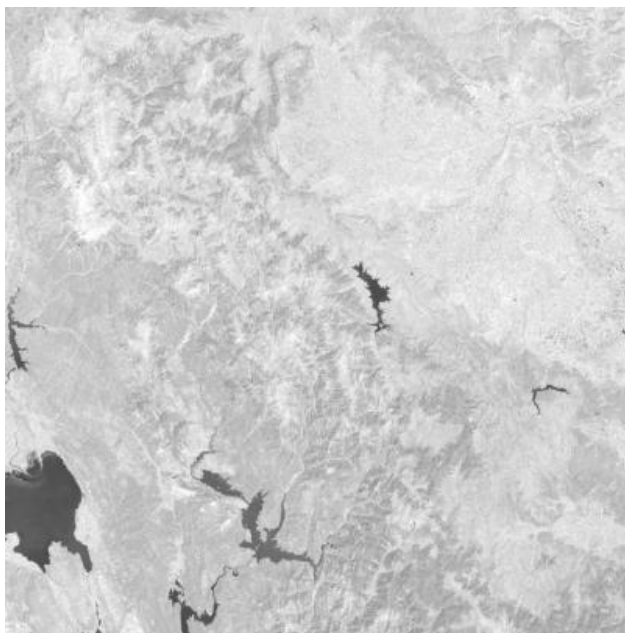
επεξεργασίες, ώστε να μπορέσει να προστεθεί η πληροφορία που περιλαμβάνεται σε αυτό στα δεδομένα εκπαίδευσης, προκειμένου να γίνει η εκπαίδευση μοντέλου μηχανικής μάθησης:

1. Επαναδειγματοληψία προκειμένου να μεταβληθεί η οριζοντιογραφική χωρική ανάλυση του DSM από 30 στα 10 m.
2. Αποκοπή των ορίων της περιοχής που μας ενδιαφέρουν ώστε αυτά να συμπίπτουν ακριβώς με τα γεωγραφικά όρια του tile τα οποία μας ενδιαφέρουν.
3. Μεταβολή του γεωδαιτικού συστήματος αναφοράς ώστε αυτά να ταιριάζουν με τα αντίστοιχα του υπερκύβου.
4. Προσθήκη στον υπερκύβο της χρονοσειράς.

Στην συνέχεια η αντίστοιχη εργασία έγινε και για τα άλλα τρία tiles με αποτέλεσμα μετά το τέλος του βήματος αυτού να υπάρχουν τέσσερις υπερκύβοι χρονοσειράς που προστίθενται μαζί με το ψηφιακό μοντέλο εδάφους.

3.5.5 Προσθήκη Brightness

Στη συνέχεια υπολογίστηκε ο δείκτης της φωτεινότητας για κάθε μια από τις πολυφασματικές απεικονίσεις της χρονοσειράς. Για τον σκοπό αυτό αναπτύχθηκε αλγόριθμος ο οποίος υπολόγιζε για κάθε εικονοστοιχείο κάθε πολυφασματικής απεικόνισης από την οποία αποτελείται ο υπερκύβος τον δείκτη της φωτεινότητας, περισσότερες λεπτομέρειες για τον οποίο δίνονται στο προηγούμενο κεφάλαιο. Ο δείκτης της φωτεινότητας προτείνεται σε αρκετές περιπτώσεις στη διεθνή βιβλιογραφία [3] ως ένας ακόμα δείκτης ο οποίος διευκολύνει ιδιαίτερα τη διάκριση μεταξύ καλλιεργειών. Στην συνέχεια γίνεται κανονικοποίηση των τιμών της φωτεινότητας με τη βοήθεια του εκατοστημορίου 99%. Δηλαδή υπολογίζεται η τιμή του στοιχείου που είναι στο 99% των ταξινομημένων κατά αύξουσα σειρά τιμών της φωτεινότητας. Στη συνέχεια κάθε εικονοστοιχείο του δείκτη διαιρείται με αυτή την τιμή. Ο δείκτης επιλέγεται να κανονικοποιηθεί με αυτό τον τρόπο έναντι της διαίρεσης με το μέγιστο της φωτεινότητας, ώστε να μην επηρεαστεί η κανονικοποίηση από εικονοστοιχεία με τιμές για τις οποίες δεν υπάρχει πληροφορία. Στη συνέχεια κάθε δείκτης της φωτεινότητας ο οποίος δημιουργήθηκε πριν, σώζεται ως ξεχωριστό αρχείο. Ο συγκεκριμένος δείκτης φαίνεται παρακάτω για την απεικόνιση από τις 23/07:



Σχήμα 10: Δείκτης Φωτεινότητας για τις 23/07 όπως φαίνεται στο λογισμικό QGIS

Έτσι δημιουργήθηκαν δέκα διαφορετικοί δείκτες φωτεινότητας (brightness) οι οποίοι στη συνέχεια με τη βοήθεια αλγορίθμου εισήχθησαν συμπληρωματικά των υπολοίπων δεικτών στον υπερκύβο, ακριβώς μετά τον δείκτη του NDBI. Μετά την προσθήκη των επιπλέον δέκα καναλιών και του DSM, ο υπερκύβος απέκτησε διαστάσεις 10980x10980x101 έναντι των αρχικών διαστάσεων. (Όλα τα προαναφερθέντα scripts φαίνονται στο παράρτημα.)

3.5.6 Μείωση διαστατικότητας-Εξαγωγή χρονικών χαρακτηριστικών (temporal features)

Μια εναλλακτική προσέγγιση, η οποία προτείνεται αντί της κλασσικής ταξινόμησης σε υπερκύβο δεδομένων, είναι αυτή της ταξινόμησης σε έναν κύβο αποτελούμενο από χρονικά μεγέθη (temporal features). Η εργασία κατέδειξε την αποδοτικότητα της προταθείσας μεθοδολογίας, επιτυγχάνοντας υψηλή συνολική ακρίβεια άνω του 90% σε όλα τα πειράματα. Η παραπάνω εργασία κατέδειξε ότι οι καλύτερες ακρίβειες επιτυγχάνονται από τη μέτρηση χρονικών μεγεθών και στη συνέχεια απευθείας χρήση των παραπάνω μεγεθών ως δεδομένα εισόδου της ταξινόμησης. Προτείνονται τα εξής χρονικά χαρακτηριστικά: Min, Max, Median, Mean, Standard Deviation για όλα τα φασματικά κανάλια και τους δείκτες του υπερκύβου και επιτυγχάνονται με αυτά υψηλές ακρίβειες (>90%). Σε εξέλιξη των αλγορίθμων της εργασίας [1], κατέστη δυνατή η δημιουργία αρκετών κύβων χρονικών χαρακτηριστικών απαραίτητων για τα πειράματα που θα ακολουθήσουν:

- ◆ min, max, median, mean, stdv for [B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI] (5X9) Ο κύβος ο οποίος προτείνεται από την εργασία.
- ◆ min, max, median, mean, stdv for [B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI]+DEM (5X9)+1
- ◆ min, max, median, mean, stdv for [B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI,BRIGHTNESS] (5X10)
- ◆ min, max, median, mean, stdv for [B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI,BRIGHTNESS]+DEM (5X10)+1

- ◆ min, max, median, mean, stdv for [G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI,BRIGHTNESS]+DEM (5X9)+1
- ◆ min, max, median, mean, stdv, percentile(75%) for [G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI,BRIGHTNESS] +DEM (6X10)+1
- ◆ min, max, median, mean, stdv, percentile(25%) for [G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI,BRIGHTNESS]+DEM (6X10)+1
- ◆ min, median, mean, stdv, percentile(99%) for [B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI,BRIGHTNESS]+DEM (5X10)+1

Οι παραπάνω κύβοι είναι αυτοί που, μαζί με το αρχείο του ground truth, θα αποτελέσουν τη βάση για το σύνολο των μετέπειτα πειραμάτων που είναι απαραίτητα.

3.5.7 Μείωση διαστατικότητας - Επιλογή σημαντικότερων χαρακτηριστικών

Η εξαγωγή χαρακτηριστικών στα πλαίσια της τηλεπισκόπησης μπορεί να οριστεί ως οι τεχνικές που μας επιτρέπουν να αναγνωρίσουμε και να κατηγοριοποιήσουμε αμοιβαίες συσχετίσεις μεταξύ διαφορετικών περιοχών της εικόνας [21]. Οι πολυφασματικοί κύβοι δεδομένων που αποτελούνται από δεκάδες κανάλια τείνουν να έχουν πολύ υψηλό βαθμό συσχέτισης μεταξύ διαφορετικών καναλιών κάτι το οποίο αυξάνει την διαστατικότητα, χωρίς απαραίτητα να αυξάνει και την πληροφορία η οποία δίνεται. Γι' αυτό τον λόγο έγιναν δοκιμές μείωσης των καναλιών τόσο στον κύβο της χρονοσειράς όσο και των χρονικών χαρακτηριστικών ώστε να ελεγχθεί ο συσχετισμός μεταξύ όγκου πληροφορίας και αποτελέσματος. Στις συγκεκριμένες δοκιμές κρατήθηκε ο χρόνος εκπαίδευσης, το ποσοστό της μνήμης το οποίο χρησιμοποιήθηκε για αυτή την εργασία και το ποσοστό ισχύος του επεξεργαστή που αυτή κατανάλωσε. Η εξαγωγή χαρακτηριστικών έγινε με την εξαγωγή των καναλιών βάσει του gini impurity, το οποίο είναι ένα μέγεθος που συσχετίζεται με την ποιότητα διαχωρισμού των δεδομένων από τον εκάστοτε κόμβο και αναλύθηκε εκτενώς στο προηγούμενο κεφάλαιο σε συσχέτιση με τον random forest. Για τον σκοπό αυτό, καθώς και για τις μετρήσεις της απόδοσης που αναφέρθηκαν, αναπτύχθηκε αλγόριθμος βασισμένος στα εκτενή εγχειρίδια της βιβλιοθήκης scikit learn και ο οποίος παρατίθεται στο παράρτημα. Ο αλγόριθμος λειτουργεί ως εξής:

1. Με βάση τα δεδομένα εκπαίδευσης και ελέγχου εκπαιδεύεται ένα μοντέλο μηχανικής μάθησης με τα αρχικά δεδομένα.
2. Δημιουργείται πίνακας με τις σημαντικότητες (ως σημαντικότητα ορίζεται η μείωση στη νόθευση του κόμβου βεβαρημένη από την πιθανότητα η πληροφορία να φτάσει σε αυτόν) για κάθε κανάλι βασισμένος στο gini impurity. Το gini impurity είναι το μέγεθος της πιθανότητας για λανθασμένη ταξινόμηση, αν η νέα περίπτωση ταξινομηθεί με βάση την κατανομή ταξινόμησης στο σετ δεδομένων. Εξηγείται επιπλέον στο κεφάλαιο 2.8.1.
3. Από αυτόν τον πίνακα σημαντικότητας εξήχθησαν τα κανάλια με τη μεγαλύτερη σημαντικότητα βάσει εκατοστημορίων 35-85%. Αυτό επιτυγχάνεται ως εξής: Από τον κατά

αύξουσα σειρά ταξινομημένο πίνακα των σημαντικότητων εξάγονται τα κανάλια των οποίων οι σημαντικότητες υπερβαίνουν το ποσοστό % επί αυτού του πίνακα. (Στιγμή η οποία χρονομετρείται και για την οποία υπολογίζεται χρήση επεξεργαστή και μνήμης.)

4. Με βάση τα εξαχθέντα κανάλια μετασχηματίζουμε το μοντέλο τη μηχανικής μάθησης και ακολούθως τα ίδια τα δεδομένα εκπαίδευσης και ελέγχου.

5. Με βάση αυτά τα νέα δεδομένα εκπαίδευσης και ελέγχου δομήθηκε νέο μοντέλο μηχανικής μάθησης. (Στιγμή η οποία χρονομετρείται και για την οποία υπολογίζεται χρησιμοποιούμενη ισχύς επεξεργαστή και η καταναλωμένη μνήμη).

6. Δημιουργείται υπολογιστικό φύλλο με τους κώδικες των καναλιών τα οποία χρησιμοποιούνται για τη δόμηση του μοντέλου μετά την διαδικασία της μηχανικής μάθησης.

7. Δημιουργείται ο πίνακας σύγχυσης.

3.6 Περιγραφή Πειραμάτων

Στη συνέχεια, και αφού χωρίστηκαν τα δεδομένα μας σε περιοχές εκπαίδευσης και ελέγχου, πραγματοποιήθηκε μια σειρά πειραμάτων για το tile 34SEJ ώστε να βρεθούν εκείνα τα οποία προσέφεραν τις καλύτερες δυνατές ακρίβειες συνολικά αλλά και συναρτήσει της μικρότερης διαστατικότητας και τα οποία θα χρησιμοποιούνταν και στα επόμενα στάδια για τα υπόλοιπα tiles.

Για την εκπαίδευση βασιστήκαμε στη βιβλιοθήκη Scikit Learn της Python. Οι περισσότερες παράμετροι του μοντέλου το οποίο δομείται είναι οι προεπιλεγμένες από τη βιβλιοθήκη, με εξαίρεση τα δέντρα του Random Forest τα οποία αυξήθηκαν από 10 σε 100. Προκειμένου να επιλεγούν τα πειράματα που εν τέλει κρατήθηκαν, έγιναν δοκιμές με συνολικά 37 πειράματα. Παρακάτω ακολουθούν σύντομες περιγραφές για κάθε ένα από αυτά τα πειράματα καθώς και παρουσίαση των ενδιαφέροντων αποτελεσμάτων που προέκυψαν.

3.6.1. Αρχικά πειράματα στο tile αναφοράς 34SEJ

Τα τριάντα επτά πειράματα που έγιναν παρουσιάζονται πολύ συνοπτικά παρακάτω ενώ στην συνέχεια ακολουθεί διάγραμμα με τις γενικές ακρίβειες καθώς και τις μέσες f1 για κάθε πείραμα:

- Πείραμα 10 dates X [B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI] + DEM. Πείραμα με την ταξινόμηση του υπερκύβου, όπως αυτός προέκυψε από την προσθήκη του DEM στον κύβο 10 ημερομηνιών φασματικών χαρακτηριστικών που αναλύθηκε στο κεφάλαιο 3.1. Έδωσε αποτελέσματα υψηλής ακρίβειας τα οποία κινήθηκαν 0,5%-2% ψηλότερα σε σχέση με τον αρχικό κύβο
- Πείραμα 10 dates X [B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI,BRIGHTNESS]. Πείραμα με την ταξινόμηση του υπερκύβου των φασματικών χαρακτηριστικών του 3.1 με την επιπλέον

προσθήκη των δεικτών της φωτεινότητας για κάθε ημερομηνία. Υπάρχουν ελάχιστες διαφορές ως προς την ακρίβεια, αποτελέσματα χαμηλότερα 0,2% κατά μέσο όρο σε σχέση με τον προαναφερθέντα κύβο αλλά αύξηση της διαστατικότητας κατά 10 κανάλια.

- Πείραμα 10 dates X [B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI,BRIGHTNESS]+DEM. Πείραμα με την ταξινόμηση του υπερκύβου των φασματικών χαρακτηριστικών με προσθήκη του δείκτη της φωτεινότητας και του DEM. Αύξηση ακριβειών κατά 0,5-2% σε σχέση με τον κύβο των φασματικών χαρακτηριστικών με το DEM του προηγούμενου πειράματος.
- Πείραμα 10 dates X [B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI,BRIGHTNESS] +DEM with feature extraction using percentile=range(35,85). Σειρά πειραμάτων κατά τα οποία στον κύβο των φασματικών χαρακτηριστικών με την προσθήκη της φωτεινότητας και του DEM εκτελούνται τεχνικές μείωσης της διαστατικότητας όπως εξηγήθηκαν στο κεφάλαιο 2.7 και 3.11. Τα πειράματα μειώνουν την διαστατικότητα μέχρι και στα 15 κανάλια και η ακρίβεια ακόμα και σε αυτή την περίπτωση παραμένει OA=94,14%. Οι ακρίβειες παραμένουν σχεδόν αμετάβλητες μέχρι και την εξαγωγή χαρακτηριστικών με εκατοστημόριο percentile=65 (OA=92,36%) με τις ακρίβειες f1 να είναι επίσης παρεμφερείς με αυτές του πειράματος χωρίς εξαγωγή χαρακτηριστικών.
- Πείραμα min, max, median, mean, stdv for [B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI]+DEM (5X9)+1. Σε αυτό το πείραμα έγινε προσθήκη του DEM στον κύβο των χρονικών χαρακτηριστικών που αναφέρθηκε στο κεφάλαιο 3.1. Αυτή τη φορά η αύξηση των ακριβειών είναι σαφώς υψηλότερη από την αντίστοιχη περίπτωση του υπερκύβου χρονικών χαρακτηριστικών με τις ακρίβειες αυξημένες κατά 1%-3% σε σχέση με τον υπερκύβο αυτό.
- Πείραμα min, max, median, mean, stdv for [B, G, R, NIR, RE, SWIR, NDVI, NDWI, NDBI, BRIGHTNESS] (5X10). Σε αυτό το πείραμα έγινε προσθήκη των δεικτών της φωτεινότητας στον υπερκύβο των χρονικών χαρακτηριστικών του 3.1. Τα αποτελέσματα έχουν τις ίδιες ακρίβειες με αυτά χωρίς την προσθήκη της φωτεινότητας και η διαστατικότητα αυξάνεται κατά πέντε κανάλια.
- Πείραμα min, max, median, mean, stdv for [B, G, R, NIR, RE, SWIR, NDVI, NDWI, NDBI, BRIGHTNESS] +DEM (5X10)+1. Στον κύβο των χρονικών χαρακτηριστικών του 3.1 προστίθενται ο δείκτης της φωτεινότητας για κάθε απεικόνιση και το DSM. Οι ακρίβειες και πάλι αυξάνονται κατά 1-3% όπως φαίνεται και στον πίνακα της επόμενης σελίδας.
- Πείραμα min, max, median, mean, stdv for [B, G, R, NIR, RE, SWIR, NDVI, NDWI, NDBI, BRIGHTNESS]+DEM (5X10)+1 with feature extraction using percentile=range(35,85). Σειρά πειραμάτων κατά τα οποία στον κύβο των χρονικών χαρακτηριστικών του 3.1 με την προσθήκη

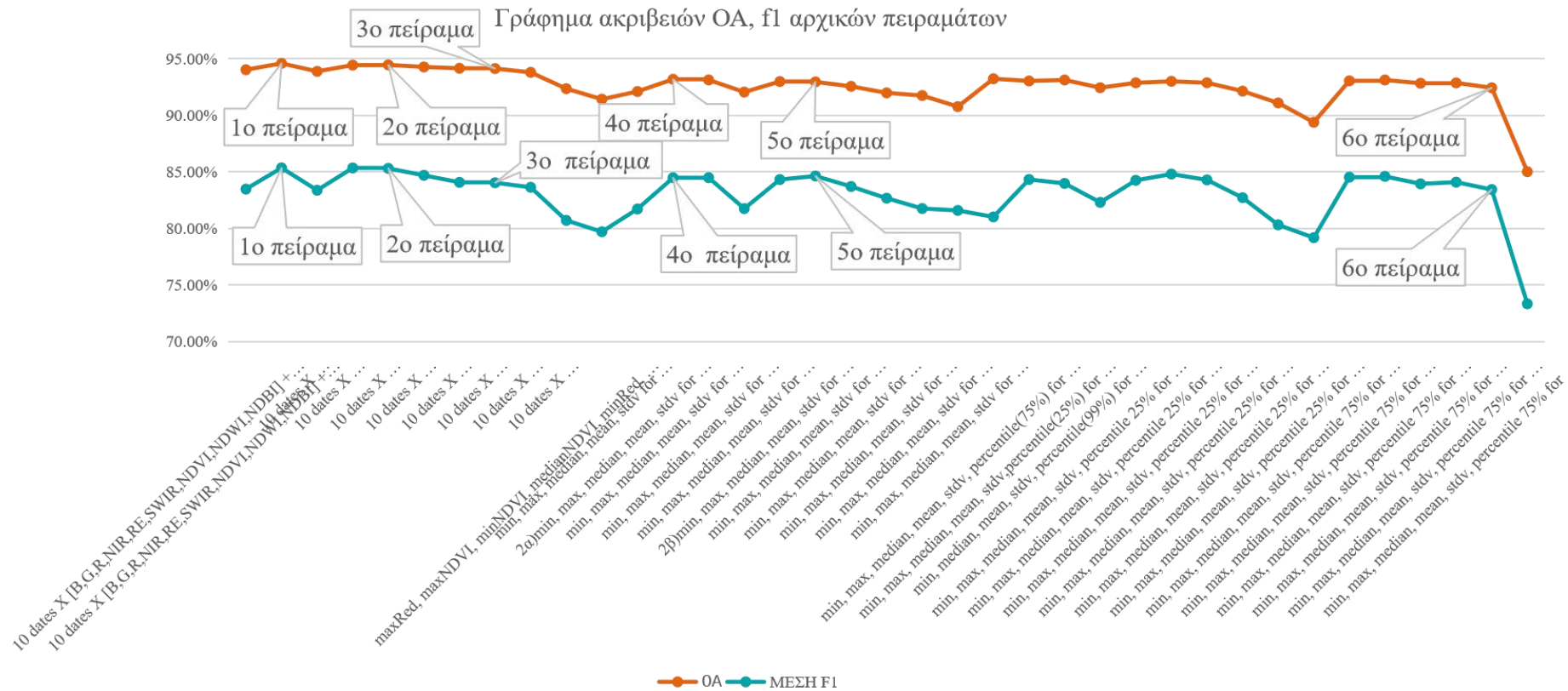
των δεικτών της φωτεινότητας και του DEM, εκτελούνται τεχνικές μείωσης της διαστατικότητας, όπως εξηγήθηκαν στο κεφάλαιο 2.7 και 3.11. Τα πειράματα μειώνουν την διαστατικότητα μέχρι και στα 12 κανάλια και η ακρίβεια ακόμα και σε αυτή την περίπτωση παραμένει OA=90,77%. Οι ακρίβειες παραμένουν σχεδόν αμετάβλητες μέχρι και την εξαγωγή χαρακτηριστικών με εκατοστημόριο percentile=55 (OA=92,56%) με τις ακρίβειες f1 να είναι επίσης παρεμφερείς με αυτές του πειράματος χωρίς εξαγωγή χαρακτηριστικών και τον αριθμό των καναλιών να είναι μειωμένος στα 21. Παρατηρείται ότι ο αριθμός των καναλιών σε αυτή την περίπτωση μειώνεται αρκετά δυσκολότερα σε σχέση με την περίπτωση του υπερκύβου πράγμα που δείχνει την αυξημένη σημαντικότητα.

- Πείραμα min, max, median, mean, stdv for [G, R, NIR, RE, SWIR, NDVI, NDWI, NDBI, BRIGHTNESS]+DEM (5X10)+1. Παρατηρώντας τη λίστα με τα κανάλια που κρατήθηκαν στα προηγούμενα πειράματα, έγινε προφανής η σημασία του καναλιού blue. Δοκιμάστηκε η εκπαίδευση μοντέλου από τον κύβο των χρονικών χαρακτηριστικών του 3.1 χωρίς το κανάλι blue για να βρεθεί η πορεία που ακολουθούν οι ακρίβειες. Η πτώση ήταν, αναμενόμενα, πολύ σημαντική (1-4%) για τους διάφορους δείκτες ακρίβειας f1 σε σχέση με τον αντίστοιχο υπερκύβο με το μπλε κανάλι όμως η γενική ακρίβεια έμεινε σταθερή.
- Πείραμα min, median, mean, stdv, percentile(99%) for [B, G, R, NIR, RE, SWIR, NDVI, NDWI, NDBI, BRIGHTNESS]+DEM (5X10)+1. Αυτό το πείραμα έγινε με αντικατάσταση του χρονικού χαρακτηριστικού του max με αυτό της τιμής που βρίσκεται στο 99% του εύρους τιμών του ταξινομημένου κατά αύξουσα σειρά πίνακα του εύρους των ανακλαστικότητας του κύβου των χρονικών χαρακτηριστικών. Με αυτό τον τρόπο αντιμετωπίζεται πιθανός επηρεασμός που θα υπάρξει στο χρονικό χαρακτηριστικό του max από “καμένα” εικονοστοιχεία. Οι γενικές και μέσες ακρίβειες f1 είχαν πτωτική πορεία σε σχέση με τον κύβο των χρονικών χαρακτηριστικών του 3.1 και κινήθηκαν 0,6% έως 2,5% χαμηλότερα.
- Πείραμα min, max, median, mean, stdv, percentile(75%) for [G, R, NIR, RE, SWIR, NDVI, NDWI, NDBI, BRIGHTNESS]+DEM (6X10)+1. Σε αυτό το πείραμα στον κύβο των πέντε χρονικών χαρακτηριστικών του 3.1 μαζί με τον δείκτη της φωτεινότητας προσθέσαμε το χαρακτηριστικό εκατοστημόριο 75% με το οποίο εξάγεται η τιμή που βρίσκεται στο 75% των τιμών του ταξινομημένου πίνακα. Στη συνέχεια προστέθηκε και το DEM. Οι ακρίβειες για αυτή την περίπτωση κυμαίνονται στα ίδια επίπεδα με τον αντίστοιχο υπερκύβο χωρίς το εκατοστημόριο 75% με την διαφοροποίηση να μην υπερβαίνει σε καμία περίπτωση το $\pm 0,2\%$.
- Πείραμα min, max, median, mean, stdv, percentile 75% for [B, G, R, NIR, RE, SWIR, NDVI, NDWI, NDBI, BRIGHTNESS] (6X10) with feature extraction using percentile=range(35,85). Σειρά πειραμάτων κατά τα οποία στον κύβο χρονικών χαρακτηριστικών του 3.1 μαζί με τον δείκτη της φωτεινότητας εκτελούνται τεχνικές μείωσης της διαστατικότητας, όπως

εξηγήθηκαν στα κεφάλαια 2.7 και 3.11. Τα πειράματα μειώνουν τη διαστατικότητα μέχρι και στα 7 κανάλια και η ακρίβεια ακόμα σε αυτή την περίπτωση πέφτει μόλις κατά 8% στο OA=85,02%. Οι ακρίβειες παραμένουν σχεδόν αμετάβλητες μέχρι και την εξαγωγή χαρακτηριστικών με εκατοστημόριο percentile=65 (OA=92,87%) με τις ακρίβειες f1 να είναι επίσης παρεμφερείς με αυτές του πειράματος χωρίς εξαγωγή χαρακτηριστικών.

- Πείραμα min, max, median, mean, stdv, percentile(25%) for [G, R, NIR, RE, SWIR, NDVI, NDWI, NDBI, BRIGHTNESS]+DEM (6X10)+1. Σε αυτό το πείραμα στον κύβο των πέντε χρονικών χαρακτηριστικών προσθέσαμε το χαρακτηριστικό εκατοστημόριο 25%, με το οποίο εξάγεται η τιμή που βρίσκεται στο 25% των τιμών του ταξινομημένου πίνακα. Στη συνέχεια προστέθηκε και το DEM. Οι ακρίβειες και για αυτή την περίπτωση κυμαίνονται στα ίδια επίπεδα με τον αντίστοιχο υπερκύβο χωρίς το εκατοστημόριο 25% με την μέγιστη διαφορά μεταξύ των δύο πειραμάτων να βρίσκεται στο $\pm 0,2\%$.
- Πείραμα min, max, median, mean, stdv, percentile 25% for [B, G, R, NIR, RE, SWIR, NDVI, NDWI, NDBI, BRIGHTNESS] (6X10) with feature extraction using percentile=range(35,85). Σειρά πειραμάτων κατά τα οποία στον κύβο των χρονικών χαρακτηριστικών με τη φωτεινότητα εκτελούνται τεχνικές μείωσης της διαστατικότητας όπως εξηγήθηκαν στα κεφάλαια 2.7 και 3.11. Τα πειράματα μειώνουν την διαστατικότητα μέχρι και στα 10 κανάλια και η ακρίβεια ακόμα σε αυτή την περίπτωση πέφτει μόλις κατά 3% στο OA=89,39%. Οι ακρίβειες παραμένουν σχεδόν αμετάβλητες μέχρι και την εξαγωγή χαρακτηριστικών με εκατοστημόριο percentile=55 (OA=92,88%) με τις ακρίβειες f1 να είναι επίσης παρεμφερείς με αυτές του πειράματος χωρίς εξαγωγή χαρακτηριστικών.

Στη συνέχεια παρουσιάζονται όλα τα παραπάνω πειράματα με τις ακρίβειες τους κατά αύξουσα σειρά σε ένα διάγραμμα στο οποίο φαίνονται οι γενικές ακρίβειες OA και οι μέσες ακρίβειες f1 κατά περίπτωση. Εδώ θα παρατηρηθούν πιο εποπτικά τα όσα αναφέρθηκαν για τις ακρίβειες παραπάνω:



Σχήμα 11: Διάγραμμα ΟΑ, f1 των αρχικών πειραμάτων

3.6.2 Σχολιασμός μείωσης καναλιών σε σχέση με την ακρίβεια και τη χρονική και υπολογιστική αποδοτικότητα

Στην συνέχεια σχολιάζονται οι ακρίβειες OA και f1 σε σχέση με τον αριθμό καναλιών για τις περιπτώσεις του α) υπερκύβου των φασματικών χαρακτηριστικών β) υπερκύβου των χρονικών χαρακτηριστικών γ) υπερκύβου των χρονικών χαρακτηριστικών με το επιπλέον χρονικό χαρακτηριστικό εκατοστημόριο 75% δ) υπερκύβου των χρονικών χαρακτηριστικών με το επιπλέον χρονικό χαρακτηριστικό εκατοστημόριο 25%.

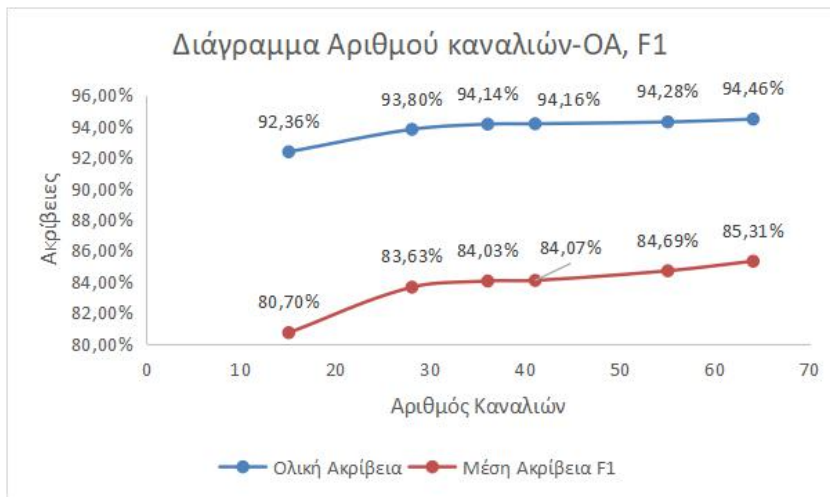
Όπως φαίνεται οι ακρίβειες σε σχέση με τον αριθμό των καναλιών για την περίπτωση α) είναι πολύ υψηλές σε όλα τα επίπεδα εξαγωγής καναλιών αλλά υπάρχει σημαντική πτώση άνω του 2% όταν τα κανάλια είναι κάτω από 20. Παρατηρείται ακόμα μια σαφώς σημαντικότερη πτώση στις ακρίβειες f1 σε σχέση με αυτή στις OA. Η ολική ακρίβεια για εξαγωγή καναλιών σε επίπεδο 35% διαφέρει 2% από την αντίστοιχη για 85%, ενώ η ακρίβεια f1 διαφέρει κατά 5% ανάμεσα στα δύο επίπεδα. Άρα η εξαγωγή χαρακτηριστικών για το πείραμα α) μπορεί να χαρακτηριστεί επιτυχημένη από πλευράς ακριβειών.

Το πείραμα β) ξεκινά από μικρότερες ακρίβειες κατά 2% σε σχέση με το α) αλλά με λιγότερα από 20 κανάλια οι ακρίβειες είναι συγκρίσιμες με τις αντίστοιχες με 30 κανάλια. Ενδιαφέρον είναι το γεγονός ότι οι ακρίβειες παραμένουν σχετικά σταθερές μέχρι ο αλγόριθμος να αφήσει κάτω από 12 κανάλια. Η δυνατότητα εξαγωγής καναλιών με διατήρηση παρεμφερών επιπέδων ακρίβειας είναι σημαντικά εφικτότερη στην περίπτωση του πειράματος β). Από τα 12 κανάλια και κάτω πάντως παρατηρείται απότομη πτώση η οποία αλλάζει την μέχρι εκεί ομαλή κλίση του διαγράμματος. Οι ακρίβειες f1 έχουν μεταβολή 3% από το πρώτο στο τρίτο πείραμα και τα διαγράμματα τους παρουσιάζουν αρκετά ομαλή κλίση.

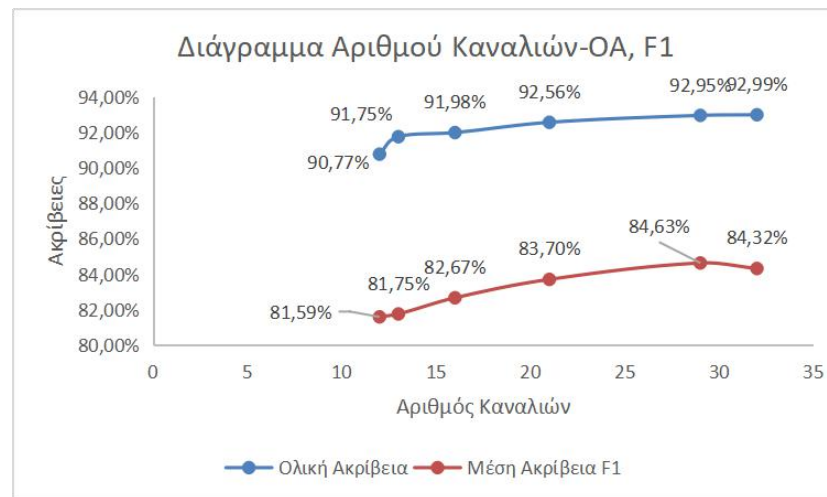
Στο πείραμα γ) παρατηρείται μια σημαντική ομαλοποίηση της κλίσης της OA σε σχέση με το πείραμα β) οπότε και η μείωση των καναλιών έχει μια αρκετά πιο γραμμική σχέση με την μεταβολή της ακρίβειας σε σχέση με πριν. Παρατηρείται πάντως ότι η συγκεκριμένη ομαλότητα της κλίσης σταματά απότομα ανάμεσα στις δύο τελευταίες περιπτώσεις εξαγωγής δεδομένων. Η περίπτωση της f1 για αυτό το πείραμα δείχνει ότι η f1 αυξάνεται στην δεύτερη εξαγωγή αλλά κατά τα άλλα προσεγγίζει αρκετά την κλίση της αντίστοιχης καμπύλης του προηγούμενου πειράματος. Η διαφοροποίηση ανάμεσα στις ακρίβειες είναι περίπου 3% για τις OA και 5% για τις f1.

Εν τέλει για το πείραμα δ) παρατηρείται μια πολύ υψηλή εξαγωγή καναλιών που με τη σειρά της καθιστά την κλίση του διαγράμματος σαφώς πιο απότομη. Δεν υπάρχει αναλογικότητα ως προς τον αριθμό εξαγομένων χαρακτηριστικών και της ακρίβειας, με την κατακόρυφη μείωση για την OA και την f1 όσο μειώνονται τα κανάλια. Η διαφορά των ακριβειών είναι 8% για τις OA και 10% για τις f1

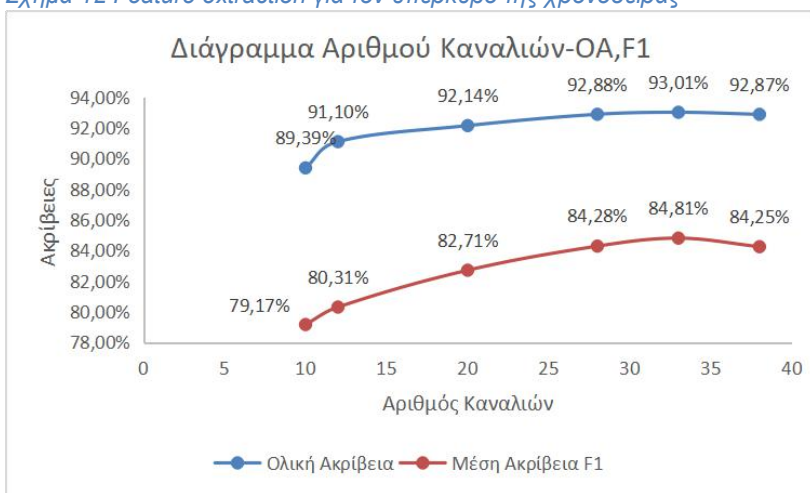
αντίστοιχα.



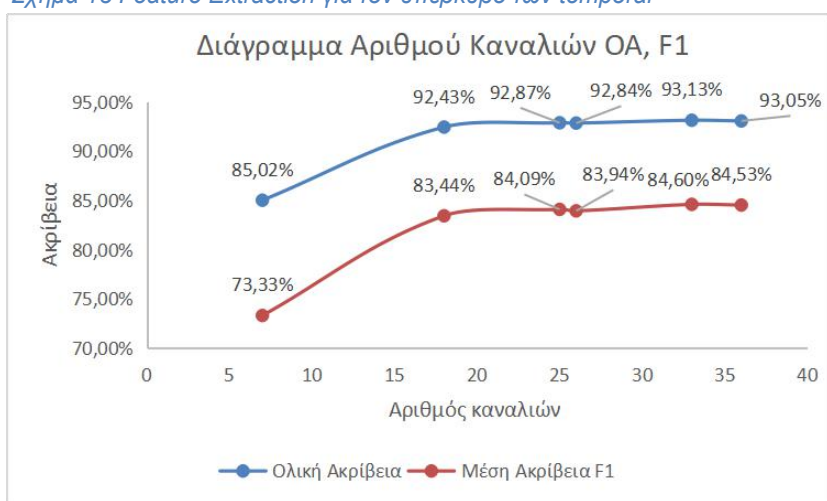
Σχήμα 12 Feature extraction για τον υπερκύβο της χρονοσειράς`



Σχήμα 13 Feature Extraction για τον υπερκύβο των temporal



Σχήμα 14 Feature extraction για τον υπερκύβο των temporal + percentile 25%



Σχήμα 15 Feature extraction για τον υπερκύβο των temporal + percentile 75%

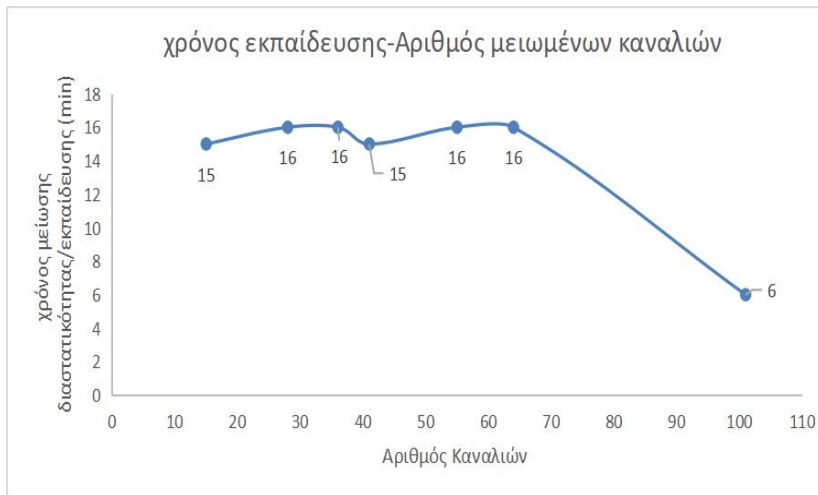
Εν τέλει θα γίνει ένας σχολιασμός της χρονικής και υπολογιστικής αποδοτικότητας της εξαγωγής χαρακτηριστικών, η οποία έγινε με την βοήθεια βιβλιοθηκών της rython αμέσως μετά την εκπαίδευση του μοντέλου μηχανικής μάθησης για δύο περιπτώσεις:

- Τον κύβο των φασματικών χαρακτηριστικών του πειράματος α)
- Τον κύβο των χρονικών χαρακτηριστικών του πειράματος δ)

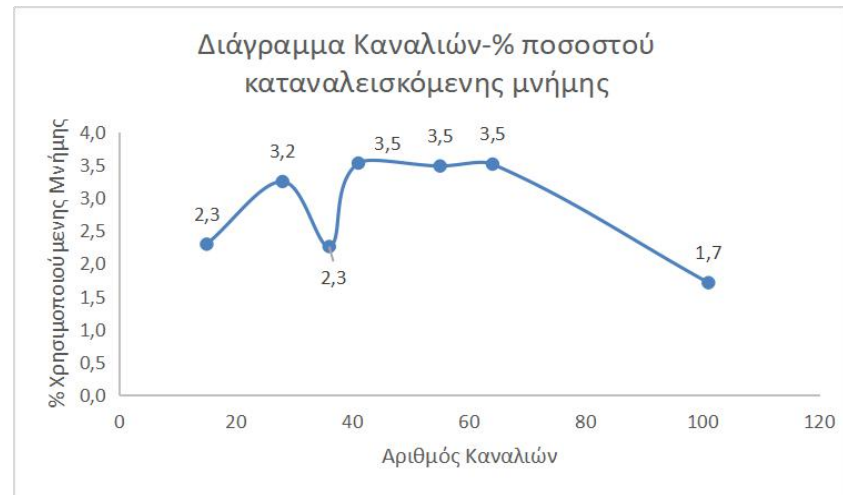
Παρατηρείται ότι για το πρώτο πείραμα η εκπαίδευση και ο έλεγχος του μοντέλου παίρνει περίπου έξι λεπτά χωρίς εξαγωγή χαρακτηριστικών. Για να εφαρμοστεί η εξαγωγή χαρακτηριστικών απαιτείται εκπαίδευση και επανεκπαίδευση του μοντέλου με μετασχηματισμένα χαρακτηριστικά (λιγότερα κανάλια). Έτσι δικαιολογείται ο χρόνος των δεκαέξι λεπτών που απαιτούνται για τη διαδικασία. Σε σχέση με τη μνήμη που καταναλώνεται, παρατηρείται ότι αυτή χωρίς εξαγωγή χαρακτηριστικών είναι 1,7% της μνήμης του υπολογιστή. Για τη διαδικασία της εξαγωγής χαρακτηριστικών παρατηρούμε πως η χρήση της μνήμης διπλασιάζεται. Μοναδικές εξαιρέσεις είναι αυτές της εξαγωγής με εκατοστημόριο 65% και 85% όπου υπάρχει σημαντική μείωση της χρησιμοποιούμενης μνήμης.

Για το δεύτερο πείραμα η εκπαίδευση και ο έλεγχος του μοντέλου απαιτεί πέντε λεπτά χωρίς εξαγωγή χαρακτηριστικών (αναμενόμενη μείωση χρόνου εκπαίδευσης λόγω σημαντικά μειωμένης διαστατικότητας). Για να εφαρμοστεί η εξαγωγή χαρακτηριστικών απαιτείται και εδώ εκπαίδευση και επανεκπαίδευση του μοντέλου με μετασχηματισμένα χαρακτηριστικά (λιγότερα κανάλια). Σε αυτή την περίπτωση απαιτούνται 11-13 λεπτά για κάθε εξαγωγή χαρακτηριστικών ενώ ενδιαφέρον έχει ότι ο χρόνος που απαιτείται μειώνεται λίγο όσο μειώνεται η διαστατικότητα. Σε σχέση με τη μνήμη που καταναλώνεται παρατηρείται ότι αυτή χωρίς εξαγωγή χαρακτηριστικών είναι 1,6% της μνήμης του υπολογιστή. Κατά την διαδικασία εξαγωγής χαρακτηριστικών η χρήση της μνήμης αυξάνεται πρώτα στο 3,6% της συνολικής μνήμης και μετά στο 4,5%.

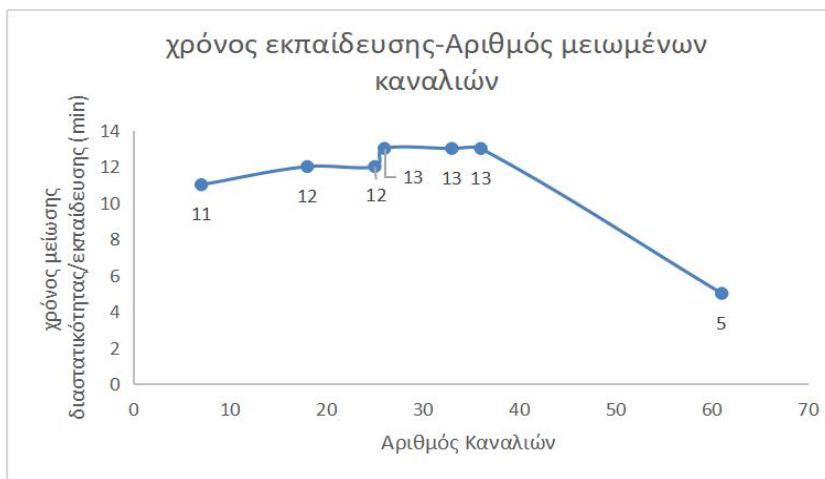
Γενικά παρατηρείται ότι η εξαγωγή χαρακτηριστικών απαιτεί μια σημαντική αύξηση χρησιμοποιούμενης μνήμης και χρόνου σε σχέση με την απλή εκπαίδευση μοντέλου. Σε ό,τι αφορά τον επεξεργαστή η χρήση της ισχύος του μετρήθηκε και βρέθηκε (αναμενόμενα) στο 100%.



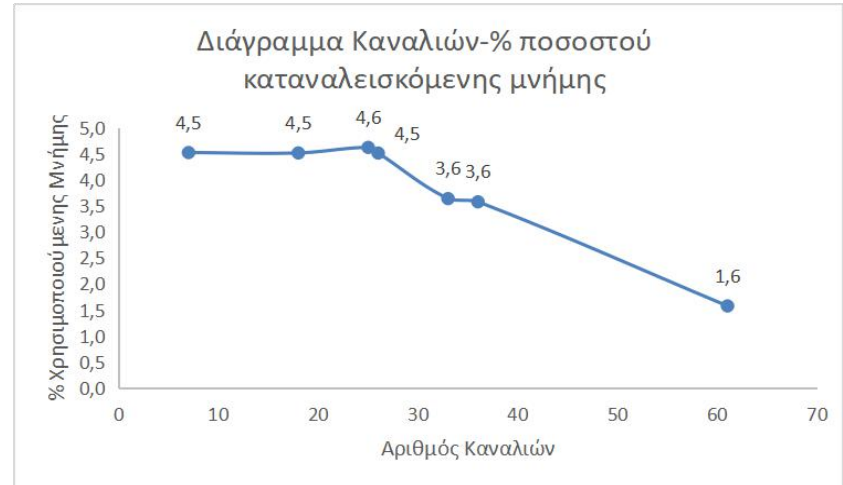
Σχήμα 16 Διάγραμμα Χρονικής αποδοτικότητας για τον φασματικό κύβο των 101 καναλιών



Σχήμα 17 Διάγραμμα Υπολογιστικής αποδοτικότητας για τον φασματικό κύβο των 101 καναλιών



Σχήμα 18 Διάγραμμα Χρονικής αποδοτικότητας για τον κύβο των temporal των 61 καναλιών



Σχήμα 19 Διάγραμμα Υπολογιστικής αποδοτικότητας για τον κύβο των temporal των 61 καναλιών

3.6.3 Επιλογή πειραμάτων εφαρμογής της μεθοδολογίας ανά tile

Από τα πειράματα που έγιναν στο tile 34SEJ επιλέχτηκαν έξι, καθώς συνδύαζαν υψηλές ολικές ακρίβειες άνω του 90% ανεξαρτήτως διαστατικότητας, καθώς επίσης και καλές ακρίβειες f1 για τις επιμέρους κατηγορίες. Για τα τέσσερα από τα έξι αυτά πειράματα επιτεύχθηκαν αρκετά υψηλές ακρίβειες ενώ μάλιστα είχε προηγηθεί η τεχνική μείωσης της διαστατικότητας με επιλογή καναλιών. Τα πειράματα αυτά είναι τα εξής:

1. Πείραμα με τον υπερκύβο 10 εικόνων*[B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI] + DEM. Σε αυτή την περίπτωση επιτυγχάνεται αξιοσημείωτα υψηλή γενική ακρίβεια (δηλαδή 94,59%), ενώ οι επιμέρους ακρίβειες παραγωγού και χρήστη (83,87% και αντιστοίχως 88,48%) είναι επίσης πάρα πολύ υψηλές, με εξαίρεση την ακρίβεια χρήστη για τις καλλιέργειες, το οποίο είναι όμως κάτι που επαναλαμβάνεται σε όλα τα πειράματα. Πάντως όλα τα παραπάνω συνδυάζονται με υψηλή διαστατικότητα (91 κανάλια).
2. Πείραμα με εξαγωγή χαρακτηριστικών από τον υπερκύβο με percentile 35% στις 10 εικόνες* [B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI, Brightness] + DEM και σύνολο 64 κανάλια. Οι ακρίβειες σε αυτή την περίπτωση είναι όμοιες με αυτές του αρχικού υπερκύβου (OA=94,46% AVPA=83,77% και AVUA=88,45%) αλλά επιτυγχάνονται με μείωση των χαρακτηριστικών από 101 στα 64. Σε κάποιες περιπτώσεις μάλιστα όπως του f1 στις καλλιέργειες έχουμε καλύτερες ακρίβειες σε σχέση με όλο τον υπερκύβο. Αυτό δείχνει τη μικρή σημαντικότητα που έχει ένας μεγάλος αριθμός καναλιών .
3. Πείραμα με εξαγωγή χαρακτηριστικών από τον υπερκύβο με percentile 65% 10 εικόνες*[B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI, Brightness] + DEM. Οδηγούμαστε σε αμελητέα μείωση συνολικής ακρίβειας (δηλαδή OA=94,14% AVPA=82,48% και AVUA=87,26%) σε σχέση με τη δραστηκή μείωση των καναλιών στα 36 από 101. Οι επιμέρους ακρίβειες παραγωγού και χρήστη βρίσκονται 1,5-2% κάτω από αυτές του κύβου.
4. Πείραμα με τον υπερκύβο των temporal min, max, median, mean, stdv for [B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI]+DEM (5X9)+1. Σε αυτή την περίπτωση έχουμε την καλύτερη ακρίβεια που επιτεύχθηκε για τον κύβο των temporal (OA=93,19%, AVPA=82,74, AVUA=84,48%) με καλές ακρίβειες σε όλους τους επιμέρους δείκτες, εκτός από την ακρίβεια παραγωγού για τις καλλιέργειες που είναι αρκετά μέτρια.
5. Πείραμα με τον υπερκύβο των temporal min, max, median, mean, stdv for [B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI,BRIGHTNESS]+DEM (5X10)+1 με percentile 45%. Εδώ η ολική ακρίβεια είναι ελάχιστα χειρότερη (92,95%) από αυτή ολόκληρου του κύβου. Οι υπόλοιπες επιμέρους ακρίβειες (82,77%, 88,26%) είναι στα ίδια επίπεδα ή και υψηλότερες από αυτές του

κύβου των temporal. Έχουμε μείωση των καναλιών στα 29 από 46.

6. Πείραμα με τον υπερκύβο των temporal min, max, median, mean, stdv, perc75 for [B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI,BRIGHTNESS]+DEM (6X10)+1 με percentile 75%. Εδώ η ολική ακρίβεια (OA=94,43%) είναι σημαντικά χειρότερη από αυτή ολόκληρου του κύβου. Ωστόσο έχουμε μικρή διάσταση (18) σε συνδυασμό με αποδεκτές ακρίβειες σε όλα τα στατιστικά μεγέθη εκτός των ακριβειών για τις καλλιέργειες.

3.6.4 Εφαρμογή της πειραματικής μεθοδολογίας ανά tile, στα tiles 34SEG, 34TEK, 35TLF

Στη συνέχεια τα πειράματα που επιλέχθηκαν θα πρέπει, όπως προαναφέρθηκε, να εφαρμοστούν σε κάποια tiles διαφορετικά από το 34SEJ και ακολούθως σε κάποια από αυτά να γίνουν πειράματα γενίκευσης με άγνωστα δεδομένα ανάμεσα σε tiles (εκπαίδευση στο ένα έλεγχος στο άλλο) ή αθροιστική μάθηση (εκπαίδευση και έλεγχος σε όλα τα tiles μαζί). Οι διαδικασίες που αναφέρθηκαν περιγράφονται παρακάτω:

Τα έξι πειράματα τα οποία επιλέχθηκαν πριν ξαναέγιναν με σύμπτυξη κάποιων από τις κατηγορίες σε νέες - και άρα μεταβολή ως προς τις κατηγορίες του ground truth - με σκοπό να υπάρχει το ίδιο σύστημα κατηγοριοποίησης για το tile 34SEJ και για τα άλλα τρία tiles και να είναι εφικτά τα πειράματα γενίκευσης με άγνωστα δεδομένα από το ένα στο άλλο. Αυτά τα πειράματα στη συνέχεια εκτελέστηκαν και στα άλλα τρία tiles τοποθετημένα σε πολλές διαφορετικές περιοχές του ελλαδικού χώρου. Για τα παρακάτω πειράματα δομήθηκαν οι εξής επιπλέον αλγόριθμοι:

1. Αλγόριθμος ο οποίος μεταβάλλει τους κωδικούς των κατηγοριών από το εύρος 1-31 σε 1-28 λαμβάνοντας υπόψιν τις συμπτώξεις που υπάρχουν ανάμεσα στις κατηγορίες.
2. Αλγόριθμος ο οποίος διαβάζει τα υπολογιστικά φύλλα με τα τελικώς εξαχθέντα χαρακτηριστικά και στη συνέχεια δημιουργεί δεδομένα εκπαίδευσης και ελέγχου μόνο με τα παραπάνω χαρακτηριστικά.

Στην συνέχεια τα έξι αυτά πειράματα έγιναν για τα τρία νέα tiles με τις συμπτυχθείσες κατηγορίες. Οι σειριακές εργασίες που απαιτήθηκαν προκειμένου να γίνει αυτό, έγιναν με τους αλγόριθμους που χρησιμοποιήθηκαν και στις προηγούμενες περιπτώσεις.

3.6.5 Πειράματα γενίκευσης μοντέλων και συνδυαστικά πειράματα

Η τελική σειρά πειραμάτων ήταν αυτή για την οποία έγιναν πειράματα γενίκευσης σε άγνωστα δεδομένα καθώς και συνδυαστικά πειράματα μεταξύ 2, 3 και 4 tiles.

- Ως πειράματα γενίκευσης σε άγνωστα δεδομένα ορίζονται πειράματα για τα οποία τα δεδομένα εκπαίδευσης προέρχονται από ένα tile και τα δεδομένα ελέγχου από ένα άλλο. Ουσιαστικά δηλαδή

ένα μοντέλο εκπαιδευμένο σε ένα tile ελέγχεται σε ένα άλλο.

- Ως συνδυαστικά πειράματα ορίζονται τα πειράματα κατά τα οποία γίνεται εκπαίδευση και έλεγχος του μοντέλου με δεδομένα αναφοράς που προέρχονται από περισσότερα του ενός tiles.

Σε κάθε περίπτωση τα πειράματα θα γίνουν για τους κύβους των χρονικών και φασματικών χαρακτηριστικών στις περιπτώσεις που δεν συμμετέχει το tile 34SEG και μόνο για τους κύβους των χρονικών χαρακτηριστικών για τα πειράματα που συμμετέχει. Συγκεκριμένα ο κύβος των φασματικών χαρακτηριστικών αποτελείται από τα χαρακτηριστικά [B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI] + DEM για κάθε μία από τις 9 ή 10 ημερομηνίες ενώ ο κύβος των χρονικών χαρακτηριστικών από τα χαρακτηριστικά min, max, median, mean, stdv for [B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI]+DEM. Παρατηρείται λοιπόν ότι η διαστατικότητα του κύβου των φασματικών χαρακτηριστικών μεταβάλλεται ανάλογα με τον αριθμό των ημερομηνιών που δομούν τον κύβο ενώ η διαστατικότητα του κύβου των χρονικών χαρακτηριστικών παραμένει σταθερή και αμετάβλητη ανεξαρτήτως του αριθμού των χαρακτηριστικών. Από τα παραπάνω εξηγείται το γεγονός πως το tile 34SEG δεν μπορεί να χρησιμοποιηθεί σε πειράματα γενίκευσης ή συνδυαστικά τα οποία έχουν ως είσοδο τον κύβο των φασματικών χαρακτηριστικών. Αντίθετα μπορεί να χρησιμοποιηθεί για πειράματα τα οποία δέχονται ως είσοδο τον κύβο των χρονικών χαρακτηριστικών η διαστατικότητα του οποίου παραμένει σταθερή ανεξάρτητα με τον αριθμό των ημερομηνιών από τις οποίες αποτελείται. Επιπλέον όλα τα πειράματα αναγωγής και τα συνδυαστικά έγιναν με είσοδο κύβους στους οποίους δεν προηγήθηκε εξαγωγή χαρακτηριστικών. Ο λόγος είναι ότι η τεχνική με την οποία γίνεται η εξαγωγή χαρακτηριστικών δεν επιτρέπει να επαναληφθεί η εξαγωγή των ίδιων καναλιών για κύβους από διαφορετικές περιοχές.

Θα επιχειρηθεί η σύγκριση μεταξύ πειραμάτων που έγιναν στους κύβους των φασματικών και χρονικών χαρακτηριστικών καθώς και η εφαρμογή πειραμάτων γενίκευσης και συνδυαστικών πειραμάτων με διαφορετικό αριθμό tiles. Επιπλέον τα αποτελέσματα των καινούργιων πειραμάτων θα συγκριθούν και με τα αποτελέσματα των προηγούμενων πειραμάτων που εφαρμόστηκαν σε ένα tile. Σε κάθε περίπτωση η επιτυχία ή μη των πειραμάτων αξιολογείται από τις ακρίβειες που επιτυγχάνουν. Τα πειράματα που έγιναν φαίνονται παρακάτω:

- Ανά 2 tiles πειράματα γενίκευσης με άγνωστα δεδομένα, δηλαδή εκπαίδευση μοντέλου με όλα τα δεδομένα από ένα tile και έλεγχος με όλα τα δεδομένα από άλλο tile.
- Ανά 2 tiles συνδυαστικά πειράματα, δηλαδή εκπαίδευση ενός μοντέλου με όλα τα δεδομένα εκπαίδευσης και έλεγχος με όλα τα δεδομένα ελέγχου και των δύο tiles.
- Ανά 3 tiles πειράματα γενίκευσης σε άγνωστα δεδομένα, δηλαδή εκπαίδευση μοντέλου με όλα τα δεδομένα από δύο tiles και έλεγχος σε ένα tile.
- Συγκριτική ανάλυση αποτελεσμάτων γενίκευσης σε άγνωστα δεδομένα με εκπαίδευση σε 1, 2, 3 tiles.
- Συγκριτική ανάλυση αποτελεσμάτων συνδυαστικών πειραμάτων με εκπαίδευση και έλεγχο σε 2, 3, 4 tiles.

Η πρώτη σειρά πειραμάτων αποτελείται από ένα πολύ μεγάλο αριθμό από πειράματα με εκπαίδευση σε κάποιο από τα tiles 34TEK, 34SEG, 35TLF και έλεγχο σε κάποιο από τα άλλα. Αξίζει να αναφερθεί ότι δοκιμάστηκαν όλοι οι συνδυασμοί για εκπαίδευση και έλεγχο που ήταν εφικτοί και δεν έγινε επιλογή συνδυασμών δεδομένων εκπαίδευσης/ελέγχου. Έτσι προέκυψαν τα εξής πειράματα:

- Εκπαίδευση στο tile 35TLF και έλεγχος στο tile 34SEJ (φασματικά χαρακτηριστικά).
- Εκπαίδευση στο tile 34SEJ και έλεγχος στο tile 35TLF (φασματικά χαρακτηριστικά).
- Εκπαίδευση στο tile 35TLF και έλεγχος στο tile 34SEJ (χρονικά χαρακτηριστικά).
- Εκπαίδευση στο tile 34SEJ και έλεγχος στο tile 35TLF (χρονικά χαρακτηριστικά).
- Εκπαίδευση στο tile 35TLF και έλεγχος στο tile 34TEK (φασματικά χαρακτηριστικά).
- Εκπαίδευση στο tile 34TEK και έλεγχος στο tile 35TLF (φασματικά χαρακτηριστικά).
- Εκπαίδευση στο tile 35TLF και έλεγχος στο tile 34TEK (χρονικά χαρακτηριστικά).
- Εκπαίδευση στο tile 34TEK και έλεγχος στο tile 35TLF (χρονικά χαρακτηριστικά).
- Εκπαίδευση στο tile 34TEK και έλεγχος στο tile 34SEJ (φασματικά χαρακτηριστικά).
- Εκπαίδευση στο tile 34SEJ και έλεγχος στο tile 34TEK (φασματικά χαρακτηριστικά).
- Εκπαίδευση στο tile 34TEK και έλεγχος στο tile 34SEJ (χρονικά χαρακτηριστικά).
- Εκπαίδευση στο tile 34SEJ και έλεγχος στο tile 34TEK (χρονικά χαρακτηριστικά).

Η δεύτερη σειρά πειραμάτων αποτελείται από συνδυαστικά πειράματα μεταξύ 2 tiles. Όπως και πριν έγιναν όλοι οι συνδυασμοί ανά 2 tiles ανάμεσα στα 34SEJ, 34TEK, 35TLF. Παρακάτω φαίνονται αυτά τα πειράματα:

- Εκπαίδευση και έλεγχος στα tiles 35TLF και 34SEJ (φασματικά χαρακτηριστικά).
- Εκπαίδευση και έλεγχος στα tiles 35TLF και 34SEJ (χρονικά χαρακτηριστικά).
- Εκπαίδευση και έλεγχος στα tiles 34TEK και 34SEJ (φασματικά χαρακτηριστικά).
- Εκπαίδευση και έλεγχος στα tiles 34TEK και 34SEJ (χρονικά χαρακτηριστικά).
- Εκπαίδευση και έλεγχος στα tiles 35TLF και 34TEK (φασματικά χαρακτηριστικά).
- Εκπαίδευση και έλεγχος στα tiles 35TLF και 34TEK (χρονικά χαρακτηριστικά).

Για την τρίτη σειρά πειραμάτων ενδιέφερε ιδιαίτερα η ακρίβεια του μοντέλου μηχανικής μάθησης όταν τα tiles που χρησιμοποιούνται για πειράματα γενίκευσης σε άγνωστα δεδομένα γειτνιάζουν, όταν δεν γειτνιάζουν και η σύγκριση των ακριβειών μεταξύ των δύο περιπτώσεων. Η πρώτη περίπτωση (γειτνίαση δεδομένων εκπαίδευσης) έχει ιδιαίτερο ενδιαφέρον καθώς είναι σημαντικό να εξεταστεί ποια είναι η ακρίβεια μοντέλων που εκπαιδεύονται σε δεδομένα εκπαίδευσης πολλαπλάσια σε σχέση με τα δεδομένα ελέγχου για την περίπτωση που αυτά γειτνιάζουν. Η δεύτερη περίπτωση έχει ιδιαίτερο ενδιαφέρον ώστε να εξεταστεί τι συμβαίνει όταν τα tiles που χρησιμοποιούνται για την εκπαίδευση ενός μοντέλου είναι διάσπαρτα στον χώρο.

Εν τέλει προκύπτουν τα εξής πειράματα από γενίκευση σε άγνωστα δεδομένα στους κύβους των

χρονικών χαρακτηριστικών:

- Εκπαίδευση στα tiles 34SEJ και 34TEK και έλεγχος στο tile 35TLF (πρώτη περίπτωση).
- Εκπαίδευση στα tiles 34SEJ και 34TEK και έλεγχος στο tile 34SEG (πρώτη περίπτωση).
- Εκπαίδευση στα tiles 34TLF και 34SEG και έλεγχος στο tile 34SEJ (δεύτερη περίπτωση).
- Εκπαίδευση στα tiles 34TLF και 34SEG και έλεγχος στο tile 34TEK (δεύτερη περίπτωση).

Στην συνέχεια έγινε σειρά από πειράματα γενίκευσης σε άγνωστα δεδομένα για τον κύβο των χρονικών χαρακτηριστικών με εκπαίδευση σε 1,2,3 tiles και συγκριτική παρουσίαση αυτών προκειμένου να προσδιοριστούν οι διαφοροποιήσεις στην ακρίβεια . Παρακάτω φαίνονται τα πειράματα που θα γίνουν:

- Εκπαίδευση στο tile 35TLF και έλεγχος στο tile 34SEJ.
- Εκπαίδευση στα tiles 35TLF και 34SEG και έλεγχος στο 34SEJ.
- Εκπαίδευση στα tiles 35TLF 34SEG και 34TEK και έλεγχος στο 34SEJ.

Τέλος έγινε σειρά από συνδυαστικά πειράματα για τον κύβο των χρονικών χαρακτηριστικών με εκπαίδευση και έλεγχο σε 1, 2, 3 και 4 tiles και συγκριτική παρουσίαση αυτών προκειμένου να προσδιοριστούν οι διαφοροποιήσεις στην ακρίβεια και παρακάτω φαίνονται αυτά τα πειράματα:

- Εκπαίδευση και έλεγχος στα tiles 34TEK και 34SEJ.
- Εκπαίδευση και έλεγχος στα tiles 35TLF, 34SEJ και 34TEK.
- Εκπαίδευση στα tiles 35TLF, 34SEJ, 34SEG και 34TEK.

ΚΕΦΑΛΑΙΟ 4 ΑΠΟΤΕΛΕΣΜΑΤΑ

4.1. Ανάλυση Επιλεγμένων Πειραμάτων στο Tile 34SEJ

Η αξιολόγηση των αποτελεσμάτων και της ακρίβειάς τους γίνεται μέσω του πίνακα σύγχυσης. Οι πίνακες σύγχυσης είναι ένα ευρέως διαδεδομένο εργαλείο για εφαρμογές μηχανικής μάθησης και ειδικά αν αυτή σχετίζεται με εφαρμογές ταξινόμησης. Στους πίνακες σύγχυσης στο κεφάλαιο αυτό οι σειρές αντιστοιχούν στις κλάσεις του ground truth και οι στήλες στις κλάσεις του αποτελέσματος ταξινόμησης. Τα μεγέθη τα οποία θα χρησιμοποιηθούν σε αυτή την εργασία είναι τα εξής:

1. Ακρίβεια Παραγωγού - Producer Accuracy: Το μέγεθος που αντιπροσωπεύει την ακρίβεια της ταξινόμησης για την κάθε κλάση. Προκύπτει από τον λόγο των σωστά ταξινομημένων εικονοστοιχείων σε σχέση με όλα τα εικονοστοιχεία της κλάσης.
2. Ακρίβεια Χρήστη - User Accuracy: Το μέγεθος που αντιπροσωπεύει την αξιοπιστία της κάθε κλάσης στην ταξινομημένη εικόνα. Πρόκειται για το κλάσμα των σωστά ταξινομημένων pixel σε σχέση με όλα τα εικονοστοιχεία που ταξινομούνται σε αυτή την κατηγορία στην ταξινομημένη εικόνα.
3. Συνολική Ακρίβεια - Overall Accuracy: Υπολογίζεται ως ο συνολικός αριθμός των σωστά ταξινομημένων εικονοστοιχείων (διαγώνιος) διαιρούμενος με τον συνολικό αριθμό των εικονοστοιχείων ελέγχου.

4.1.1 Πείραμα 1: Μοντέλο φασματικών χαρακτηριστικών με DSM

Αρχικά σχολιάζεται το πείραμα με τον υπερκύβο αποτελούμενο από τα έξι κανάλια και τους δείκτες και το DSM που έχει συνολική διάσταση καναλιών 91. Σε αυτό το πείραμα παρατηρείται με βάση την αναφορά της ταξινόμησης συνολική ακρίβεια 94,59% και μέσος δείκτης f_1 85,35%.

Παρατηρείται ότι οι κατηγορίες του ανθρωπογενούς περιβάλλοντος (DUF, SUF, ICU, MES, RAN, GRH, PHV) έχουν ακρίβεια χρήστη η οποία κυμαίνεται στο εύρος 54%-99% και χρήστη στο εύρος 66-99%. Η μικρότερη ακρίβεια χρήστη βρίσκεται στην κατηγορία της αραιής αστικής δόμησης, η μικρότερη ακρίβεια παραγωγού στην κατηγορία του οδικού δικτύου (54%) και η μικρότερη ακρίβεια χρήστη στην κατηγορία του οδικού δικτύου (81%).

Για τις κατηγορίες των κωνοφόρων και των φυλλοβόλων (CNF, BLF) παρατηρείται ακρίβεια παραγωγού 99%-100% και χρήστη 97-99% οι οποίες είναι αξιοσημείωτα υψηλές. Για τις κατηγορίες των χορτολιβαδικών εκτάσεων και σκληροφυλλικής βλάστησης (NGR, DSV,SSV) παρατηρείται ότι το εύρος της ακρίβειας παραγωγού είναι 93%-96% με τα περισσότερα σφάλματα να υπάρχουν για την κατηγορία των εκτάσεων με γρασίδι (NGR), οι εκτάσεις του οποίου κατά ένα σημαντικό ποσοστό ταξινομούνται ως περιοχές με αραιή βλάστηση. Οι ακρίβειες χρήστη κινούνται και αυτές σε πολύ υψηλά επίπεδα στο εύρος 89%-97%, και προφανώς δεν υπάρχουν σημαντικά σφάλματα συμπερίληψης για καμία από αυτές τις κατηγορίες.

Οι κατηγορίες των καλλιεργειών (WHT, BRL, OAT,MAI, CLV, GRF, TBC, VNY, OLG, FRT) κινούνται σε ένα ιδιαίτερα εκτεταμένο εύρος ακριβειών παραγωγού 34%-97% και εύρος ακριβειών χρήστη 35%-99%. Και εδώ οι καλύτερες ακρίβειες παραγωγού παρατηρούνται για το σιτάρι και το βαμβάκι ενώ οι χειρότερες για τη βρώμη και τις ζωτροφές (ακρίβεια χρήστη 34% και 48% αντίστοιχα). Οι μικρότερες ακρίβειες χρήστη υπάρχουν στις κατηγορίες της βρώμης και του κριθαριού (35% και 60% αντίστοιχως).

Στις κατηγορίες που είναι σχετικές με το γυμνό έδαφος (RCK, SVA, BDS) παρατηρούνται και πάλι υψηλές ακρίβειες παραγωγού, που κινούνται στο εύρος 90%-99%. Οι χειρότερες ακρίβειες παραγωγού βρίσκονται στην κατηγορία των περιοχών με αραιή βλάστηση (90%) και οι μεγαλύτερες στην κατηγορία των βραχωδών εκτάσεων (100%). Οι ακρίβειες χρήστη αντίστοιχα κινούνται στο εύρος 71%-98% .

Οι τελευταίες κατηγορίες με τις οποίες θα ασχοληθούμε είναι αυτές των υδάτινων σωμάτων (WBD, WCR, MRS, CWT). Και εδώ έχουμε ιδιαίτερα υψηλές ακρίβειες, παρεμφερείς με το προηγούμενο πείραμα, με ακρίβεια παραγωγού στο εύρος 93%-100% και ακρίβεια χρήστη 86%-100%. Οι χειρότερες ακρίβειες παραγωγού είναι στην κατηγορία των βαλτωδών εκτάσεων με ακρίβεια 93% ενώ οι χειρότερες χρήστη είναι σε αυτή των ποταμών με ακρίβεια 86%.

ΠΙΝΑΚΑΣ ΣΥΓΧΥΣΗΣ ΥΠΕΡΚΥΒΟΥ ΦΑΣΜΑΤΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΜΑΖΙ ΜΕ DEM																																	
	DUF	SUF	ICU	RAN	MES	PHV	GRH	BLF	CNF	NGR	DSV	SSV	VNY	OLG	FRT	WHT	BRL	OAT	MAI	CTN	TBC	CLV	GRF	FLW	SVA	BDS	RCK	MRS	WCR	WBD	CWT	sum	PA
DUF	903	111	16	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	1053	86%	
SUF	147	1243	0	4	4	0	0	0	0	0	0	13	3	3	1	3	0	1	6	5	0	0	0	0	9	1	1	0	0	0	1444	86%	
ICU	36	7	436	0	21	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	30	0	0	0	544	80%		
RAN	77	70	0	264	1	17	0	15	0	0	3	11	0	1	0	0	0	0	8	2	0	0	0	0	16	4	0	0	1	0	490	54%	
MES	2	6	80	4	1717	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	1017	1	0	2	0	2846	60%	
PHV	0	14	0	0	0	466	0	5	0	10	0	23	0	19	1	9	1	0	6	2	0	0	0	0	6	0	0	0	0	0	562	83%	
GRH	0	6	0	0	0	0	605	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	612	99%	
BLF	0	4	0	0	0	0	0	12149	5	1	66	0	5	0	6	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	12248	99%	
CNF	0	0	0	0	0	0	0	37	34402	0	7	4	0	2	0	0	0	0	0	0	0	0	0	0	8	0	1	0	0	0	34461	100%	
NGR	0	0	0	0	0	0	0	7	0	4344	1	21	0	9	0	0	0	0	1	0	0	14	0	0	108	0	0	0	0	0	4505	96%	
DSV	0	0	0	0	0	0	0	292	11	0	9225	71	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	9618	96%	
SSV	0	34	0	0	0	0	0	27	0	3	144	5174	0	54	0	0	0	0	0	0	0	0	0	0	147	0	0	0	0	0	5583	93%	
VNY	0	49	0	0	0	0	0	0	0	1	5	24	623	13	9	0	0	0	1	42	0	2	0	0	0	0	0	0	0	0	769	81%	
OLG	0	7	0	0	0	0	0	0	0	101	9	260	0	1563	15	0	0	0	0	0	0	16	0	0	0	0	0	0	0	0	1971	79%	
FRT	0	22	0	0	0	0	0	4	0	0	13	9	11	144	688	0	0	0	28	14	0	24	0	0	0	0	0	0	0	0	957	72%	
WHT	11	19	0	0	0	0	0	0	0	0	0	0	0	0	0	10106	109	1	8	21	0	81	51	3	0	0	0	0	0	10410	97%		
BRL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	207	288	25	3	4	0	1	0	0	0	0	0	0	0	0	528	55%	
OAT	0	1	0	0	0	0	0	0	0	29	0	16	0	1	0	64	2	70	22	0	0	1	0	0	0	0	0	0	0	0	206	34%	
MAI	0	31	0	0	0	0	0	0	0	0	28	0	1	2	0	36	28	0	11155	218	0	405	0	0	0	0	0	0	0	11904	94%		
CTN	5	127	0	4	31	0	0	4	0	1	0	15	4	20	7	71	50	82	264	20747	0	279	19	3	0	63	0	0	0	21796	95%		
TBC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	76	0	0	33	3	1410	14	0	0	0	0	0	0	0	1536	92%		
CLV	0	29	0	0	0	0	0	0	0	4	0	43	0	20	0	130	0	18	40	768	17	11830	0	25	0	0	0	0	0	12924	92%		
GRF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	537	0	0	0	10	0	0	541	19	0	0	0	0	0	1107	49%		
FLW	0	71	0	0	0	0	0	0	0	39	0	12	1	1	3	111	0	3	2	48	0	5	0	348	0	0	0	4	0	648	54%		
SVA	3	16	0	1	0	0	0	0	1	163	0	146	0	11	0	0	0	0	0	0	0	0	0	0	3342	0	21	0	0	3704	90%		
BDS	0	6	3	0	66	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2681	8	0	106	0	2871	93%	
RCK	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	6	2748	0	0	0	2762	99%	
MRS	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	566	41	0	0	611	93%
WCR	0	0	1	0	0	0	0	0	3	0	5	0	0	0	0	0	0	0	1	2	0	0	0	0	0	3	0	0	1081	0	0	1096	99%
WBD	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29	11174	0	11214	100%
CWT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7202	7202	100%	
sum	1187	1873	536	277	1856	485	610	12540	34422	4696	9517	5846	648	1881	730	11350	478	200	11578	21887	1427	12672	611	398	3668	3792	2810	570	1261	11174	7202	kappa: 94%	
UA	76%	66%	81%	95%	93%	96%	99%	97%	100%	93%	97%	89%	96%	83%	94%	89%	60%	35%	96%	95%	99%	93%	89%	87%	91%	71%	98%	99%	86%	100%	100%	OA: 95%	

Πίνακας 2 Πίνακας Σύγκρισης του πρώτου πειράματος (με κόκκινο φαίνονται οι κατηγορίες στις οποίες κατευθύνονται σφάλματα παράλειψης>5%), με ακρίβεια OA=94,59% και MO f1=85,5%

4.1.2 Πείραμα 2: Μοντέλο φασματικών χαρακτηριστικών με DSM και εξαγωγή βάσει σημαντικότητας (35%)

Στην συνέχεια σχολιάζεται το πείραμα το οποίο έγινε για τον υπερκύβο των φασματικών χαρακτηριστικών με εξαγωγή χαρακτηριστικών στο εκατοστημόριο 35% με αριθμό καναλιών=64. Η συνολική ακρίβεια για το συγκεκριμένο πείραμα βρίσκεται στο 94,46% ενώ ο συντελεστής f_1 έχει τιμή 85,31%.

Παρατηρείται ότι τα γενικά αποτελέσματα μεταξύ αυτού και του προηγούμενου πειράματος δε διαφέρουν σχεδόν καθόλου από την άποψη των ακριβειών παραγωγού και χρήστη. Συγκεκριμένα παρατηρείται ότι οι κατηγορίες του ανθρωπογενούς περιβάλλοντος (DUF, SUF, ICU, MES, RAN, GRH, PHV) έχουν ακρίβεια χρήστη η οποία κυμαίνεται στο εύρος 55%-99% και χρήστη 66-99%. Η μικρότερη ακρίβεια χρήστη αλλά και παραγωγού παρατηρείται στην κατηγορία των οδικών δικτύων.

Για την κατηγορία των δασικών εκτάσεων (CNF, BLF) παρατηρείται ακρίβεια παραγωγού 99%-100% και χρήστη 97-99% και είναι ίδιες με τις ακρίβειες του προηγούμενου πειράματος. Για τις κατηγορίες των χορτολιβαδικών εκτάσεων και σκληροφυλλικής βλάστησης (NGR, DSV, SSV) παρατηρείται ότι το εύρος της ακρίβειας παραγωγού είναι 93%-97% με τα περισσότερα σφάλματα να υπάρχουν για την κατηγορία των εκτάσεων με γρασίδι (NGR) η οποία, όπως αναφέρθηκε και στο προηγούμενο πείραμα, έχει αρκετά σφάλματα παράλειψης που ταξινομούνται ως αραιή βλάστηση. Οι ακρίβειες χρήστη κινούνται και αυτές σε πολύ υψηλά επίπεδα στο εύρος 90%-97%.

Οι κατηγορίες των καλλιεργειών (WHT, BRL, OAT, MAI, CLV, GRF, TBC, VNY, OLG, FRT) έχουν και πάλι ένα υψηλό εύρος ακριβειών παραγωγού 37%-97% και εύρος χρήστη 38%-98%. Και εδώ οι καλύτερες ακρίβειες παραγωγού και χρήστη παρατηρούνται στο σιτάρι και το βαμβάκι ενώ οι χειρότερες ακρίβειες παραγωγού στη βρώμη και στις εκτάσεις με ζωτροφές (ακρίβεια παραγωγού 37% και 48% αντίστοιχα). Οι χειρότερες ακρίβειες χρήστη είναι στις κατηγορίες της βρώμης και του κριθαριού (38% και 59% αντίστοιχως). Το πολύ μεγάλο εύρος ακριβειών που υπάρχει σε αυτή την περίπτωση μπορεί να γίνει σημαντικά μικρότερο αν η βρώμη συγχωνευτεί με το σιτάρι σε μια γενικότερη κατηγορία τα δημητριακά.

Για τις κατηγορίες που είναι σχετικές με το γυμνό έδαφος (RCK, SVA, BDS) παρατηρούνται και πάλι υψηλές ακρίβειες παραγωγού που κινούνται στο εύρος 90%-100%. Και πάλι οι χαμηλότερες ακρίβειες παραγωγού βρίσκονται στην κατηγορία των περιοχών με αραιή βλάστηση (SVA) (90%) και οι υψηλότερες στην κατηγορία των βραχωδών εκτάσεων (BDS) (100%). Οι ακρίβειες χρήστη αντίστοιχα κινούνται στο εύρος 69%-99%, λίγο κατώτερες σε σχέση με το προηγούμενο πείραμα.

Οι τελευταίες κατηγορίες που θα εξεταστούν είναι αυτές των υδάτινων σωμάτων (WBD, WCR, MRS,

CWT). Και εδώ βρίσκονται πολύ υψηλές ακρίβειες, παρεμφερείς με το προηγούμενο πείραμα, με ακρίβεια παραγωγού στο εύρος 94%-100% και ακρίβεια χρήστη 87%-100%. Οι χαμηλότερες ακρίβειες παραγωγού είναι στην κατηγορία των βαλτωδών εκτάσεων (MRS) ενώ οι χαμηλότερες χρήστη είναι σε αυτή των ποταμών με ακρίβεια 87%.

Παρατηρείται ότι παρά τη σημαντική μείωση που υπήρξε στον αριθμό των χαρακτηριστικών μέσω των πειραμάτων (64 από 101), η ολική ακρίβεια, ο δείκτης f1 αλλά και οι ακρίβειες χρήστη και παραγωγού ανά κατηγορία, μένουν σχετικά σταθερές. Μοναδική εξαίρεση αποτελούν οι ακρίβειες για τις καλλιέργειες που βρίσκονται 0,2% χαμηλότερα σε σχέση με το προηγούμενο πείραμα. Κάτι τέτοιο εξηγείται από το γεγονός ότι 40 από τα 101 κανάλια είναι δείκτες παράγωγοι των καναλιών του Sentinel 2A. Αυτό συνεπάγεται ότι αυτά τα κανάλια δεν εισάγουν νέα πληροφορία, απλά βελτιστοποιούν την απεικόνιση της ήδη υπάρχουσας. Ως εκ τούτου, ακόμα και αν παραβλεφθεί η συσχέτιση μεταξύ διαφορετικών καναλιών, θεωρητικά μπορεί να επιτευχθούν παρεμφερείς ακρίβειες από ταξινόμηση υπερκύβου με όλα τα 90 κανάλια, με ταξινόμηση 60 καναλιών.

Σε ό,τι αφορά στα κανάλια τα οποία επιλέχθηκαν μπορούν να γίνουν οι εξής παρατηρήσεις. Το πρώτο κανάλι που επιλέγεται προκειμένου να κρατηθεί είναι αυτό του DSM πράγμα το οποίο αναμενόταν, αφού αυτό προσφέρει σημαντική υψομετρική πληροφορία, η οποία παίζει σημαντικό ρόλο σε μια περιοχή με τόσο μεγάλες υψομετρικές διαφορές. Εξάλλου, όπως έχει αναφερθεί, κάτι τέτοιο φαίνεται και από τη διαφορά της ολικής ακρίβειας ανάμεσα στα πειράματα με και χωρίς DSM που συνήθως είναι 0,5%. Στις πρώτες θέσεις των καναλιών που κρατούνται βρίσκονται και τα κανάλια του μπλε για τους ανοιξιάτικους και θερινούς μήνες του 2016 (03/16-08/16), ενώ ιδιαίτερο ενδιαφέρον έχει και το γεγονός πως τα κανάλια του μπλε κρατούνται από όλες τις απεικονίσεις. Κάτι τέτοιο μπορεί να εξηγηθεί από το γεγονός ότι το μπλε κανάλι είναι ιδιαίτερα χρήσιμο στην διάκριση του εδάφους από τις καλλιέργειες, στη διάκριση μεταξύ κωνοφόρων και φυλλοβόλων δέντρων και στη διάκριση μεταξύ νεφών και χιονιού. Όπως προαναφέρθηκε η συγκεκριμένη περιοχή καλύπτεται σε σημαντικό βαθμό από βλάστηση με τα κωνοφόρα και τα φυλλοβόλα να είναι επίσης πολυάριθμα. Εν τέλει, ειδικά για τους χειμερινούς μήνες, η δυνατότητα διάκρισης μεταξύ νεφών και χιονιού παίζει κομβικό ρόλο, ιδιαίτερα για ορεινές περιοχές.

Το επόμενο κανάλι που παίζει πολύ μεγάλο ρόλο είναι αυτό του πράσινου. Και εδώ διατηρούνται όλα τα κανάλια του πράσινου με αυτά των ανοιξιάτικων και θερινών μηνών να είναι και πάλι τα πιο σημαντικά (κρατιούνται κατά κύριο λόγο κανάλια από απεικονίσεις όπως αυτή από τις 23/07). Όπως είναι γνωστό το πράσινο κανάλι συμβάλλει κατά κόρον στη διάκριση μεταξύ διαφορετικών κατηγοριών βλάστησης καθώς και στη διάκριση του αστικού πρασίνου από τον υπόλοιπο αστικό ιστό. Με βάση την έκταση που καλύπτει η βλάστηση επί του συνόλου της περιοχής είναι λογική η επιλογή του. Τα επόμενα πολύ σημαντικά κανάλια που κρατιούνται είναι οι δείκτες NDWI και NDVI οι οποίοι έτσι και αλλιώς θεωρούνται συμπληρωματικοί, με την έννοια ότι ο NDVI χρησιμοποιείται για να ανιχνευθεί η βλάστηση, με έμφαση να δίνεται στην υγιή βλάστηση, ενώ ο δείκτης NDWI

χρησιμοποιείται για την ανίχνευση της υγρασίας και του νερού στα φυτά αλλά και το έδαφος. Και οι δύο αυτοί δείκτες έχουν όπως αναμένεται μεγάλη σημασία για την περιοχή μας, η οποία καλύπτεται σε μεγάλο ποσοστό από βλάστηση και τα ποσοστά υγρασίας είναι υψηλά, ειδικότερα στη δυτική πλευρά της Πίνδου. Παρατηρείται ότι τα κανάλια που κρατιούνται για αυτούς τους δείκτες προέρχονται πρώτα από τη φθινοπωρινή απεικόνιση (απεικόνιση από τις 01/10/2016) και μετά από τις θερινές όπως και πριν. Τα τελευταία κανάλια που κρατιούνται σε σημαντικό ποσοστό είναι το RED και το NIR. Και τα δύο αυτά κανάλια θεωρούνται πολύ χρήσιμα σε ό,τι αφορά τη βλάστηση. Πιο συγκεκριμένα το RED, με τη μικρή απορροφητικότητα που έχει στη βλάστηση, είναι ιδιαίτερος σημαντικό για τη διάκριση μεταξύ εδάφους και βλάστησης, ενώ το NIR είναι ιδιαίτερος σημαντικό για τη διάκριση μεταξύ διαφορετικών ειδών βλάστησης. Εν τέλει αρκετά συχνά κρατιέται και ο δείκτης της φωτεινότητας(Brightness). Γενικά κρατιούνται κανάλια από όλες τις απεικονίσεις, ενώ δεν υπάρχει κάποιο μήκος κύματος καναλιού που να μη διατηρείται ως σημαντικό.

ΠΙΝΑΚΑΣ ΣΥΓΧΥΣΗΣ ΥΠΕΡΚΥΒΟΥ ΦΑΣΜΑΤΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΜΑΖΙ ΜΕ DEM ΜΕ ΕΞΑΓΩΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΜΕ ΕΚΑΤΟΣΤΗΜΟΡΙΟ 35%																																	
	DUF	SUF	ICU	RAN	MES	PHV	GRH	BLF	CNF	NGR	DSV	SSV	VNY	OLG	FRT	WHT	BRL	OAT	MAI	CTN	TBC	CLV	GRF	FLW	SVA	BDS	RCK	MRS	WCR	WBD	CWT	sum	PA
DUF	903	113	16	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	1053	86%	
SUF	129	1250	1	4	7	0	0	0	0	0	19	3	2	1	3	0	1	5	7	0	2	0	0	7	1	2	0	0	0	1444	87%		
ICU	56	7	424	0	25	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	10	0	0	0	544	78%			
RAN	86	64	0	269	1	5	0	14	0	0	3	13	0	0	1	1	0	0	9	2	0	0	0	13	7	0	1	1	0	490	55%		
MES	3	8	54	5	1665	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	1105	0	0	1	0	2846	59%		
PHV	0	20	0	1	0	450	0	5	0	15	0	17	0	24	0	13	0	2	6	6	0	0	0	3	0	0	0	0	562	80%			
GRH	0	7	1	0	0	0	603	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	612	99%			
BLF	0	3	0	0	0	0	0	12124	5	1	96	1	1	0	6	0	0	0	0	0	0	0	0	0	11	0	0	0	0	12248	99%		
CNF	0	1	0	0	0	0	0	38	34405	0	6	4	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	1	0	34461	100%		
NGR	0	0	0	0	0	0	0	11	0	4366	1	16	0	4	0	0	0	0	0	0	0	7	0	0	100	0	0	0	0	4505	97%		
DSV	0	0	0	0	0	0	0	289	5	0	9233	68	0	22	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	9618	96%		
SSV	0	20	0	0	0	0	0	27	0	6	146	5202	0	61	0	0	0	0	0	0	0	0	0	0	121	0	0	0	0	5583	93%		
VNY	0	44	0	0	0	0	0	0	0	1	10	15	636	13	10	0	0	0	1	35	0	4	0	0	0	0	0	0	0	769	83%		
OLG	0	9	0	0	0	0	0	0	0	110	7	225	0	1557	20	0	0	0	0	0	0	43	0	0	0	0	0	0	0	1971	79%		
FRT	0	25	0	0	0	0	0	4	0	0	11	11	11	133	680	0	0	0	38	24	0	20	0	0	0	0	0	0	0	957	71%		
WHT	10	16	0	2	0	0	0	0	0	0	0	0	0	0	0	10115	94	0	11	14	0	75	64	9	0	0	0	0	10410	97%			
BRL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	210	285	16	9	4	0	4	0	0	0	0	0	0	528	54%			
OAT	0	11	0	0	0	2	0	0	0	34	0	6	0	1	0	38	4	76	34	0	0	0	0	0	0	0	0	0	206	37%			
MAI	0	34	0	0	0	0	0	0	0	0	11	0	0	2	0	29	29	2	11118	220	0	459	0	0	0	0	0	0	11904	93%			
CTN	1	140	0	10	34	0	0	3	0	0	0	15	1	23	13	65	70	86	239	20715	0	294	20	9	0	58	0	0	21796	95%			
TBC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	76	0	0	59	0	1385	16	0	0	0	0	0	0	1536	90%			
CLV	0	43	0	0	0	0	0	0	0	3	0	43	0	25	0	188	0	16	91	769	6	11716	0	24	0	0	0	0	12924	91%			
GRF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	554	0	0	6	0	0	529	18	0	0	0	0	0	1107	48%			
FLW	0	70	0	0	0	0	0	0	0	39	0	6	0	4	0	120	0	0	6	54	0	6	0	341	0	0	0	0	648	53%			
SVA	0	14	0	2	0	0	0	0	1	163	0	147	0	17	0	0	0	0	0	0	0	0	0	3341	0	19	0	0	3704	90%			
BDS	4	4	3	0	76	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2688	3	0	2871	94%			
RCK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	3	2753	0	0	2762	100%			
MRS	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	572	36	0	611	94%	
WCR	0	0	1	1	0	0	0	0	3	0	5	0	0	0	0	0	0	0	2	0	0	0	0	0	0	2	0	0	1096	99%			
WBD	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25	11178	0	11214	100%		
CWT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7202	7202	100%		
sum	1192	1903	500	294	1823	457	607	12515	34419	4738	9540	5811	652	1888	731	11412	482	199	11626	21858	1391	12647	613	401	3613	3889	2787	575	1239	11178	7202	kappa: 94%	
UA	76%	66%	85%	91%	91%	98%	99%	97%	100%	92%	97%	90%	98%	82%	93%	89%	59%	38%	96%	95%	100%	93%	86%	85%	92%	69%	99%	87%	100%	100%	OA: 94%		

Πίνακας 3 Πίνακας Σύγκρισης του δεύτερου πειράματος (με κόκκινο φαίνονται οι κατηγορίες στις οποίες κατευθύνονται σφάλματα παράλειψης>5%) με ακρίβεια OA=94,46% και MO f1=85,31%

4.1.3 Πείραμα 3: Μοντέλο φασματικών χαρακτηριστικών με DSM και εξαγωγή βάσει σημαντικότητας (65%)

Σε αυτή την περίπτωση το πείραμα γίνεται με εξαγωγή χαρακτηριστικών με εκατοστημόριο 65%, η οποία έχει ως αποτέλεσμα τα αρχικά δεδομένα εκπαίδευσης και ελέγχου με διαστατικότητα των 101 καναλιών να μετασχηματιστούν σε παράγωγα με διαστατικότητα 36 καναλιών. Σε αυτό το πείραμα η συνολική ακρίβεια είναι 94,14% και ο μέσος δείκτης $f1$ είναι 84,03%.

Για τις κατηγορίες που σχετίζονται με το ανθρωπογενές περιβάλλον, δηλαδή τις κατηγορίες DUF, SUF, ICU, RAN, MES, PHV και GRH, οι ακρίβειες παραγωγού κινούνται στο εύρος 51%-98%, κάτι που συνεπάγεται ότι σε σχέση με το πρώτο πείραμα το κάτω όριο βρίσκεται 3% κάτω. Κάτι αντίστοιχο ισχύει και με τις υπόλοιπες κατηγορίες. Η κατηγορία με τη μικρότερη ακρίβεια παραγωγού είναι η κατηγορία των οδικών δικτύων RAN με ακρίβεια 51% και αυτή με τη μεγαλύτερη, η κατηγορία των θερμοκηπίων (GRH). Οι ακρίβειες χρήστη κινούνται στο εύρος 61%-100% με τη μικρότερη ακρίβεια χρήστη να είναι στην κατηγορία της αραιής αστικής δόμησης.

Οι κατηγορίες των κωνοφόρων και φυλλοβόλων δέντρων (CNF,BLF) έχουν ακρίβειες χρήστη 97% και 100% και παραγωγού 100% - παρεμφερείς με αυτές των κατηγοριών στα προηγούμενα πειράματα.

Για τις κατηγορίες που σχετίζονται με τις χορτολιβαδικές εκτάσεις (NGR, DSV, SSV) και τη σκληροφυλλική βλάστηση, παρατηρείται ότι το εύρος για τις ακρίβειες παραγωγού είναι 92%-96% και το εύρος ακριβειών χρήστη είναι 91%-97%. Η κατηγορία με τη μικρότερη ακρίβεια παραγωγού και χρήστη είναι η αραιή σκληροφυλλική βλάστηση, με ακρίβειες 92% και 91% αντιστοίχως. Παρατηρείται λοιπόν ότι δεν υπάρχει διαφοροποίηση για αυτές τις κατηγορίες σε σχέση με τα προηγούμενα πειράματα.

Στις κατηγορίες των καλλιεργειών (VNY, OLG, FRT, WHT, BRL, OAT, MAI, CTN, TBC, FLW, GRF) παρατηρούμε ακρίβειες παραγωγού οι οποίες βρίσκονται στο εύρος 27%-97%, δηλαδή αρκετά χαμηλότερα ως προς το κάτω όριο σε σχέση με αυτές των προηγούμενων πειραμάτων (-7%). Τάση για σημαντική μείωση των ακριβειών παραγωγού σε σχέση με τα προηγούμενα πειράματα παρατηρείται στις κατηγορίες των οποίων οι ακρίβειες παραγωγού ήταν κοντά στο κάτω όριο του εύρους τιμών ενώ στις υψηλότερες ακρίβειες (>80%) παρατηρείται ελάχιστη μεταβολή. Οι ακρίβειες χρήστη βρίσκονται στο εύρος 26%-96% και για αυτές εξάγονται τα ίδια συμπεράσματα που εξήχθησαν και στην περίπτωση των ακριβειών παραγωγού. Τη μικρότερη ακρίβεια παραγωγού και χρήστη την έχουμε για την κατηγορία της βρώμης.

Οι κατηγορίες που είναι σχετικές με το γυμνό έδαφος (BDS,RCK,MRS), έχουν ακρίβειες παραγωγού 90%-100% και ακρίβειες χρήστη 69%-98%. Και στις δύο περιπτώσεις είναι παρεμφερείς με αυτές των

αρχικών πειραμάτων, ενώ η κατηγορία με τη μικρότερη ακρίβεια χρήστη είναι η BDS και αυτή με την μικρότερη ακρίβεια παραγωγού είναι η SVA.

Εντέλει για τα υδάτινα σώματα (MRS, WCR, WBD, CWT), παρατηρούνται σημαντικά υψηλές ακρίβειες παραγωγού και χρήστη, οι οποίες κινούνται στο εύρος 89%-100% και 91%-100% αντιστοίχως. Οι ακρίβειες δε διαφοροποιούνται σχεδόν καθόλου σε σχέση με τις προηγούμενες περιπτώσεις.

Τα χαρακτηριστικά τα οποία επιλέγονται σε αυτή την περίπτωση είναι ως επί το πλείστον τα ίδια τα οποία επιλέχθηκαν στο προηγούμενο πείραμα, με εξαγωγή χαρακτηριστικών για τα πρώτα 36 κανάλια. Και εδώ κρατιούνται κυρίως τα κανάλια Blue και Green κυρίως από τους θερινούς μήνες (δηλαδή τον 7ο και τον 8ο του 2016) καθώς και κάποιοι δείκτες NDVI και NDWI με την πρώτη απεικόνιση που διατηρείται να είναι αυτή από 01/10/2016. Για δύο ημερομηνίες δηλαδή τις 03/07/2016 και 02/08/2016 έχουμε μόνο ένα κανάλι να διατηρείται, το μπλε. Πάντως διατηρούνται κανάλια από όλες τις απεικονίσεις.

ΠΙΝΑΚΑΣ ΣΥΓΧΥΣΗΣ ΥΠΕΡΚΥΒΟΥ ΦΑΣΜΑΤΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΜΑΖΙ ΜΕ DEM ΜΕ ΕΞΑΓΩΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΜΕ ΕΚΑΤΟΣΤΗΜΟΡΙΟ 65%																																	
	DUF	SUF	ICU	RAN	MES	PHV	GRH	BLF	CNF	NGR	DSV	SSV	VNY	OLG	FRT	WHT	BRL	OAT	MAI	CTN	TBC	CLV	GRF	FLW	SVA	BDS	RCK	MRS	WCR	WBD	CWT	sum	PA
DUF	899	114	18	1	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	5	0	0	0	0	1053	85%	
SUF	133	1212	0	5	8	0	0	1	0	1	0	19	1	2	2	3	0	0	8	30	0	2	0	0	13	1	3	0	0	0	1444	84%	
ICU	30	24	409	0	38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22	21	0	0	0	544	75%	
RAN	56	67	0	251	3	27	0	11	0	0	2	15	0	1	0	1	0	0	13	3	0	0	0	0	27	10	3	0	0	0	490	51%	
MES	1	11	45	9	1631	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	4	1141	1	0	0	0	2846	57%	
PHV	0	21	0	0	0	447	0	0	0	10	0	33	0	17	0	10	1	3	7	1	0	1	0	0	9	0	0	2	0	0	562	80%	
GRH	4	4	4	0	0	0	599	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	612	98%	
BLF	0	14	0	0	0	0	0	12203	5	1	15	3	1	0	1	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	12248	100%	
CNF	0	0	0	0	0	0	0	35	34402	2	4	8	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	34461	100%	
NGR	0	0	0	0	0	0	0	13	0	4347	0	27	0	1	0	0	0	0	0	0	0	0	0	0	117	0	0	0	0	0	4505	96%	
DSV	0	0	0	0	0	0	0	289	4	0	9254	49	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	9618	96%	
SSV	0	22	0	0	0	0	0	28	0	8	153	5139	0	87	0	0	0	0	0	0	0	0	0	0	146	0	0	0	0	5583	92%		
VNY	0	49	0	0	0	0	0	0	0	4	15	8	620	11	27	0	0	0	15	16	0	4	0	0	0	0	0	0	0	0	769	81%	
OLG	0	3	0	0	0	0	0	0	0	56	12	157	1	1701	16	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0	1971	86%	
FRT	0	24	0	0	0	0	0	3	0	1	10	16	15	114	683	0	0	0	38	24	0	29	0	0	0	0	0	0	0	0	957	71%	
WHT	0	8	0	12	0	0	0	0	0	0	0	3	0	0	0	10097	137	0	14	16	0	43	71	7	2	0	0	0	0	10410	97%		
BRL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	198	269	51	7	3	0	0	0	0	0	0	0	0	0	0	528	51%	
OAT	0	1	0	0	0	0	0	0	0	27	0	22	0	0	0	20	4	55	67	0	0	7	0	3	0	0	0	0	0	0	206	27%	
MAI	0	55	0	0	0	0	0	2	0	0	21	1	1	2	0	27	23	7	11116	243	1	405	0	0	0	0	0	0	0	0	11904	93%	
CTN	1	209	0	10	45	0	0	2	0	1	0	12	10	13	19	104	55	77	206	20596	0	342	20	14	9	51	0	0	0	0	21796	94%	
TBC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	76	5	0	16	98	1289	52	0	0	0	0	0	0	0	0	1536	84%	
CLV	0	56	0	0	0	0	0	0	0	1	0	16	1	47	8	236	1	18	106	1000	2	11409	0	23	0	0	0	0	0	0	12924	88%	
GRF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	550	0	0	0	1	0	1	534	21	0	0	0	0	0	0	1107	48%	
FLW	0	86	0	0	0	0	0	0	0	33	0	6	0	0	0	149	0	0	8	20	0	24	2	318	1	0	0	1	0	648	49%		
SVA	3	7	0	0	0	0	0	0	3	183	0	131	0	10	0	0	0	0	0	0	0	0	0	0	3349	0	18	0	0	3704	90%		
BDS	32	7	2	0	58	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2711	2	0	59	0	2871	94%	
RCK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	1	2754	0	0	0	0	2762	100%	
MRS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	556	55	0	0	611	91%	
WCR	3	0	0	0	0	0	0	0	0	0	4	1	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	1086	0	1096	99%	
WBD	0	0	0	0	0	1	0	1	0	0	7	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	11181	0	11214	100%
CWT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7202	7202	100%	
sum	1162	1994	478	288	1798	477	599	12588	34414	4675	9497	5670	650	2027	756	11471	495	211	11621	22052	1292	12345	627	386	3699	3944	2802	559	1222	11181	7202	kappa: 94%	
UA	77%	61%	86%	87%	91%	94%	100%	97%	100%	93%	97%	91%	95%	84%	90%	88%	54%	26%	96%	93%	100%	92%	85%	82%	91%	69%	98%	99%	89%	100%	100%	OA: 94%	

Πίνακας 4 Πίνακας Σύγκρισης του τρίτου πειράματος (με κόκκινο φαίνονται οι κατηγορίες στις οποίες κατευθύνονται σφάλματα παράλειψης >5%) με OA=94,14% και MO f1=84,03%

4.1.4 Πείραμα 4: Μοντέλο χρονικών χαρακτηριστικών με DSM

Σε αυτή την περίπτωση το πείραμα γίνεται με την αντικατάσταση του υπερκύβου των φασματικών χαρακτηριστικών με έναν υπερκύβο με τα μεγέθη min, max, median, mean και stdv, εξαχθέντα από όλα τα ομόλογα κανάλια που βρίσκονται στον υπερκύβο. Σε αυτό το πείραμα η συνολική ακρίβεια είναι 93,19%, ο μέσος δείκτης f1 είναι 84,18% και ο αριθμός των καναλιών είναι 46.

Για τις κατηγορίες που σχετίζονται με το ανθρωπογενές περιβάλλον δηλαδή τις κατηγορίες DUF, SUF, ICU, RAN, MES, PHV, GRH, οι ακρίβειες παραγωγού κινούνται στο εύρος 58%-99%, δηλαδή είναι αρκετά υψηλές σε σχέση με τα προηγούμενα πειράματα παρά τις μικρότερες διαστάσεις (29 κανάλια), ειδικά σε κατηγορίες με ακρίβεια κοντά στο 50%. Η κατηγορία με τη μικρότερη ακρίβεια είναι αυτή των μεταλλείων (MES) (58%), ενώ αυτή με τη μεγαλύτερη τα θερμοκήπια (99%). Οι ακρίβειες χρήστη κινούνται στο εύρος 75%-99%, με την μικρότερη ακρίβεια χρήστη στην κατηγορία της αραιής αστικής δόμησης (75%) και τη μεγαλύτερη στα φωτοβολταϊκά.

Οι κατηγορίες των κωνοφόρων και φυλλοβόλων δέντρων (CNF,BLF) έχουν ακρίβειες χρήστη που είναι 99% και 100% και παραγωγού 96% και 100% αντίστοιχα, οι οποίες παραμένουν αμετάβλητες σε σχέση με προηγούμενα πειράματα.

Για τις κατηγορίες που σχετίζονται με τις χορτολιβαδικές εκτάσεις (NGR, DSV, SSV) και την σκληροφυλλική βλάστηση, παρατηρείται ότι το εύρος ακριβειών παραγωγού είναι 91%-96% και το εύρος ακριβειών χρήστη είναι 88%-98%. Η αραιή σκληροφυλλική βλάστηση είναι η κατηγορία με τη μικρότερη ακρίβεια παραγωγού, 91% και με τη μικρότερη ακρίβεια χρήστη, 88% .

Στις κατηγορίες των καλλιεργειών (VNY, OLG, FRT, WHT, BRL, OAT, MAI, CTN, TBC,FLW,GRF) αφενός παρατηρούνται ακρίβειες παραγωγού οι οποίες βρίσκονται στο εύρος 26%-97%, αφετέρου παρατηρείται ότι το κάτω όριο είναι χαμηλότερα σε σχέση με τα προηγούμενα πειράματα. Και πάλι παρατηρείται μείωση των ακριβειών στις κατηγορίες με χαμηλές ακρίβειες επί του εύρους τιμών, ενώ στις υψηλότερες ακρίβειες (>80%) παρατηρείται ελάχιστη μεταβολή. Οι ακρίβειες χρήστη βρίσκονται στο εύρος 36%-99% και για αυτές εξάγονται τα ίδια συμπεράσματα που εξήχθησαν και για την περίπτωση των ακριβειών παραγωγού. Τη μικρότερη ακρίβεια παραγωγού και χρήστη την έχουμε για την κατηγορία της βρώμης.

Οι κατηγορίες που είναι σχετικές με το γυμνό έδαφος (BDS,RCK,MRS) έχουν ακρίβειες παραγωγού 91%-100% και ακρίβειες χρήστη 69%-98%. Και σε αυτή την περίπτωση έχουμε παρεμφερείς ακρίβειες με όλα τα προηγούμενα πειράματα, ενώ η κατηγορία με τη μικρότερη ακρίβεια χρήστη είναι η BDS και αυτή με τη μικρότερη ακρίβεια παραγωγού είναι η SVA.

Εν τέλει για τα υδάτινα σώματα (MRS, WCR, WBD, CWT), παρατηρούμε πάρα πολύ υψηλές ακρίβειες παραγωγού και χρήστη, οι οποίες κινούνται στο εύρος 95%-100% και 85%-100% αντίστοιχως. Οι

ακρίβειες δεν διαφοροποιούνται σχεδόν καθόλου σε σχέση με τις προηγούμενες περιπτώσεις.

ΠΙΝΑΚΑΣ ΣΥΓΧΥΣΗΣ ΥΠΕΡΚΥΒΟΥ ΦΑΣΜΑΤΟΧΡΟΝΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΜΑΖΙ ΜΕ DEM																																	
	DUF	SUF	ICU	RAN	MES	PHV	GRH	BLF	CNF	NGR	DSV	SSV	VNY	OLG	FRT	WHT	BRL	OAT	MAI	CTN	TBC	CLV	GRF	FLW	SVA	BDS	RCK	MRS	WCR	WBD	CWT	sum	PA
DUF	893	113	20	3	16	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	6	0	0	0	0	1053	85%	
SUF	134	1218	0	4	6	0	0	0	0	0	0	21	4	5	2	1	0	0	5	15	0	1	0	0	24	2	2	0	0	0	1444	84%	
ICU	27	7	427	0	42	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	3	0	0	0	544	78%		
RAN	53	44	0	310	1	1	0	5	0	1	3	7	0	3	0	3	0	0	8	6	0	0	0	0	23	6	14	0	2	0	490	63%	
MES	5	4	63	5	1640	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	1117	1	0	2	0	2846	58%	
PHV	0	9	0	2	1	441	0	3	0	21	0	27	0	10	0	13	0	1	16	3	0	0	0	0	14	1	0	0	0	562	78%		
GRH	0	1	2	0	0	0	608	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	612	99%		
BLF	0	0	0	0	0	0	0	12178	6	21	6	2	7	0	10	0	0	0	3	0	0	0	0	0	15	0	0	0	0	12248	99%		
CNF	0	0	0	0	0	0	0	43	34404	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	1	0	34461	100%	
NGR	0	0	0	0	0	0	0	74	0	4240	1	15	0	3	0	0	0	3	0	53	0	25	0	0	91	0	0	0	0	0	4505	94%	
DSV	0	0	0	0	0	0	0	288	21	0	9212	75	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	9618	96%	
SSV	0	84	0	0	0	1	0	26	0	4	145	5065	0	77	0	0	0	0	0	0	0	0	0	0	181	0	0	0	0	5583	91%		
VNY	0	38	0	0	0	0	0	0	0	2	0	16	607	17	27	0	0	19	43	0	0	0	0	0	0	0	0	0	0	769	79%		
OLG	0	0	0	0	0	0	0	0	0	72	5	287	0	1525	27	8	2	0	0	3	0	41	0	0	1	0	0	0	0	1971	77%		
FRT	0	22	0	0	0	0	0	1	0	0	18	19	5	107	651	0	0	0	89	7	0	38	0	0	0	0	0	0	0	957	68%		
WHT	9	3	0	2	0	0	0	0	0	0	0	0	0	0	0	10093	66	0	23	104	0	52	44	14	0	0	0	0	0	10410	97%		
BRL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	205	272	0	21	12	0	18	0	0	0	0	0	0	0	528	52%		
OAT	0	2	0	0	0	0	0	0	0	18	0	1	0	0	1	42	0	58	81	0	0	0	0	0	3	0	0	0	0	206	28%		
MAI	0	10	0	0	0	0	0	9	0	0	18	2	14	17	21	37	33	0	10588	594	2	554	0	2	3	0	0	0	0	11904	89%		
CTN	2	36	0	6	53	0	1	3	0	3	0	19	14	37	14	130	95	88	275	20617	10	326	19	0	0	48	0	0	0	21796	95%		
TBC	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	73	2	0	4	61	1384	8	0	1	0	0	0	0	0	1536	90%		
CLV	0	11	0	0	0	0	0	0	0	260	0	19	12	50	21	515	0	8	74	1401	0	10551	0	1	1	0	0	0	0	12924	82%		
GRF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	456	0	1	0	37	0	5	597	11	0	0	0	0	0	1107	54%		
FLW	0	11	0	0	0	0	0	0	0	29	0	1	0	0	6	160	1	0	3	78	0	21	2	335	0	0	0	1	0	648	52%		
SVA	3	11	0	1	0	0	0	0	0	164	0	117	0	8	0	0	0	0	0	1	0	0	0	0	3371	0	28	0	0	3704	91%		
BDS	1	2	4	1	76	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2652	14	0	119	0	2871	92%	
RCK	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	2756	0	0	2762	100%		
MRS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	581	30	0	611	95%	
WCR	0	0	0	0	0	0	0	6	0	4	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2	0	0	1081	1	1096	99%	
WBD	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	36	11169	0	11214	100%
CWT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	7200	7202	100%
sum	1127	1626	516	335	1835	444	625	12630	34437	4838	9429	5693	663	1879	780	11736	471	159	11209	23039	1396	11640	662	364	3743	3859	2818	582	1277	11170	7200	kappa: 0,92	
UA	79%	75%	83%	93%	89%	99%	97%	96%	100%	88%	98%	89%	92%	81%	83%	86%	58%	36%	94%	89%	99%	91%	90%	92%	90%	69%	98%	100%	85%	100%	100%	OA: 0,93	

Πίνακας 5 Πίνακας Σύγκρισης του τέταρτου πειράματος (με κόκκινο φαίνονται οι κατηγορίες στις οποίες κατευθύνονται σφάλματα παράλειψης>5%) με OA=93,19% και MO f1=84,48%

4.1.5 Πείραμα 5: Μοντέλο χρονικών χαρακτηριστικών με DSM και εξαγωγή βάσει σημαντικότητας (45%)

Σε αυτή την περίπτωση στον κύβο του προηγούμενου πειράματος έγινε εξαγωγή χαρακτηριστικών με εκατοστημόριο 45, με το οποίο γίνεται μείωση χαρακτηριστικών από 46 σε 29 κανάλια. Σε αυτό το πείραμα η συνολική ακρίβεια είναι 93,19% και ο μέσος δείκτης f_1 είναι 84,18%, δηλαδή παρουσιάζονται παρεμφερείς ακρίβειες με τα προηγούμενα πειράματα.

Για τις κατηγορίες του αστικού ιστού (DUF, SUF, ICU, RAN, MES, PHV, GRH) οι ακρίβειες παραγωγού κυμαίνονται στο εύρος 55%-99% και είναι πολύ κοντά στις αντίστοιχες ακρίβειες του προηγούμενου πειράματος που είχε 17 περισσότερα κανάλια. Η κατηγορία με τη μικρότερη ακρίβεια είναι αυτή των μεταλλείων (55%), ενώ αυτή με τη μεγαλύτερη τα θερμοκήπια (99%). Οι ακρίβειες χρήστη κινούνται στο εύρος 76%-99% με τη χαμηλότερη ακρίβεια χρήστη να είναι στην κατηγορία της αραιής αστικής δόμησης και την υψηλότερη στα φωτοβολταϊκά. Παρατηρείται λοιπόν ότι οι ακρίβειες παραγωγού και χρήστη δε μεταβάλλονται σε σχέση με τα προηγούμενα πειράματα.

Για τις κατηγορίες των κωνοφόρων και φυλλοβόλων δέντρων (CNF,BLF) παρατηρούνται ακρίβειες χρήστη 99% και 100% και παραγωγού 97% και 100% αντίστοιχα, οι οποίες παραμένουν αμετάβλητες σε σχέση με προηγούμενα πειράματα.

Για τις κατηγορίες που σχετίζονται με τις χορτολιβαδικές εκτάσεις (NGR, DSV, SSV) και την σκληροφυλλική βλάστηση παρατηρείται ότι το εύρος ακριβειών παραγωγού είναι 91%-96% και το εύρος ακριβειών χρήστη είναι 87%-97%. Η κατηγορία με τη μικρότερη ακρίβεια παραγωγού είναι η αραιή σκληροφυλλική βλάστηση, με ακρίβεια παραγωγού 91%, ενώ αυτή με τη μικρότερη ακρίβεια χρήστη 87% είναι εκτάσεις που καλύπτονται από γρασίδι NGR.

Στις κατηγορίες των καλλιεργειών (VNY, OLG, FRT, WHT, BRL, OAT, MAI, CTN, TBC,FLW,GRF) παρατηρούνται ακρίβειες παραγωγού οι οποίες βρίσκονται στο εύρος 45%-97%, με το κάτω όριο σαφώς υψηλότερα σε σχέση με το προηγούμενο πείραμα. Γενικότερα παρατηρείται το φαινόμενο οι κατηγορίες οι οποίες είχαν ακρίβεια παραγωγού κοντά ή κάτω από 50% να έχουν μια σημαντική αύξηση, ενώ κατηγορίες οι οποίες είχαν ήδη υψηλή ακρίβεια παραμένουν στάσιμες. Η μικρότερη ακρίβεια παραγωγού είναι και πάλι στην κατηγορία της βρώμης 45%. Σε σχέση με τις ακρίβειες χρήστη, αυτές βρίσκονται στο εύρος 47%-97% και για αυτές εξάγονται τα ίδια συμπεράσματα που εξήχθησαν και για την περίπτωση των ακριβειών παραγωγού, δηλαδή ότι η χαμηλότερη ακρίβεια χρήστη παρατηρείται για την κατηγορία της βρώμης και είναι 47%.

Οι κατηγορίες που είναι σχετικές με το γυμνό έδαφος (BDS,RCK,MRS) έχουν ακρίβειες παραγωγού 91%-100% και ακρίβειες χρήστη 67%-99%. Και σε αυτή την περίπτωση έχουμε παρεμφερείς ακρίβειες $\pm 1\%$ με όλα τα προηγούμενα πειράματα, ενώ η κατηγορία με τη μικρότερη ακρίβεια χρήστη, όπως και πριν, είναι η BDS και αυτή με τη μικρότερη ακρίβεια παραγωγού είναι η SVA.

Εν τέλει για τα υδάτινα σώματα (MRS, WCR, WBD, CWT) παρατηρούνται οι ίδιες ακρίβειες παραγωγού και χρήστη με το προηγούμενο πείραμα, οι οποίες κινούνται στο εύρος 95%-100% και 85%-100% αντιστοίχως. Οι ακρίβειες δε διαφοροποιούνται σχεδόν καθόλου σε σχέση με τις προηγούμενες περιπτώσεις.

Σε ό,τι αφορά τα κανάλια τα οποία επιλέχθηκαν, παρατηρείται ότι τα μετασηματισμένα κανάλια περιλαμβάνουν και τους πέντε δείκτες. Κάτι επιπλέον που παρατηρείται είναι το γεγονός ότι το μόνο κανάλι το οποίο δεν εμφανίζεται με κανένα δείκτη του είναι το SWIR. Τα πρώτα σε σημαντικότητα κανάλια που κρατιούνται είναι αυτά του blue και πιο συγκεκριμένα τα mean blue και median blue, πράγμα το οποίο εξηγείται τόσο από την σημασία που έχει το blue για τη σωστή ταξινόμηση, και έχει εξηγηθεί από την περίπτωση του υπερκύβου, αλλά και από τη σημασία που έχει ο μέσος όρος των τιμών και η διάμεσος, ώστε να υπάρχει η καλύτερη δυνατή εικόνα για το εύρος των τιμών της ανακλαστικότητας και αυτή να μην επηρεάζεται από ακραίες τιμές. Στην περίπτωση του κύβου των χρονικών χαρακτηριστικών παρατηρείται ότι ιδιαίτερη σημασία έχουν οι δείκτες, με όλους τους δείκτες να διατηρούνται ως σημαντικοί. Πιο συγκεκριμένα, καταρχάς κρατιέται ο δείκτης NDWI με τα χρονικά χαρακτηριστικά mean, min και max. Παρατηρείται λοιπόν ότι εδώ διατηρείται τόσο η ενδιάμεση τιμή του εύρους, όσο και οι ακραίες τιμές. Η σημασία του δείκτη NDWI έχει τονιστεί σε προηγούμενο πείραμα. Οι δείκτες κατά σειρά σημαντικότητας μετά από τον NDWI είναι ο NDVI, ο NDBI και ο brightness και παρατηρείται ότι και για τους τρεις αυτούς δείκτες διατηρούνται το min και το max όπως και στην περίπτωση του NDWI αλλά το mean στις τρεις αυτές περιπτώσεις κρίνεται σαφώς σημαντικότερο από το min. Γενική τάση που παρατηρείται είναι να προτιμώνται τα χρονικά χαρακτηριστικά για τις ενδιάμεσες τιμές, έναντι αυτών για τις ακραίες.

ΠΙΝΑΚΑΣ ΣΥΓΧΥΣΗΣ ΥΠΕΡΚΥΒΟΥ ΦΑΣΜΑΤΟΧΡΟΝΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΜΑΖΙ ΜΕ DEM ΚΑΙ ΕΞΑΓΩΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΜΕ ΕΚΑΤΟΣΤΗΜΟΡΙΟ 45																																	
	DUF	SUF	ICU	RAN	MES	PHV	GRH	BLF	CNF	NGR	DSV	SSV	VNY	OLG	FRT	WHT	BRL	OAT	MAI	CTN	TBC	CLV	GRF	FLW	SVA	BDS	RCK	MRS	WCR	WBD	CWT	sum	PA
DUF	908	101	20	1	15	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	6	0	0	0	0	0	1053	86%
SUF	119	1229	0	8	5	0	0	0	0	0	0	28	8	5	3	1	0	0	2	11	0	1	0	0	21	1	2	0	0	0	1444	85%	
ICU	37	6	420	0	50	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	2	0	0	0	0	544	77%	
RAN	70	49	0	298	2	1	0	5	0	3	5	6	0	1	3	3	0	0	6	0	0	0	0	0	21	8	7	0	2	0	490	61%	
MES	6	5	32	4	1570	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	1223	0	0	0	0	2846	55%	
PHV	2	12	0	2	0	437	0	4	0	23	0	23	0	10	0	12	0	0	12	4	0	0	0	1	20	0	0	0	0	0	562	78%	
GRH	0	2	3	0	0	0	606	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	612	99%	
BLF	0	0	0	0	0	0	0	12128	10	57	17	3	3	1	12	0	0	0	3	0	0	0	0	0	14	0	0	0	0	0	12248	99%	
CNF	0	0	0	0	0	0	0	53	34391	0	6	0	0	1	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	34461	100%	
NGR	0	0	0	0	0	0	0	13	0	4341	0	17	0	4	0	0	0	3	2	27	0	26	0	72	0	0	0	0	0	0	4505	96%	
DSV	0	0	0	0	0	0	0	292	19	0	9215	67	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	9618	96%	
SSV	0	46	0	0	0	0	0	26	0	3	146	5089	0	79	0	0	0	0	0	0	0	0	0	0	194	0	0	0	0	0	5583	91%	
VNY	0	10	0	0	0	0	0	0	0	1	0	43	602	17	16	0	0	0	24	43	0	13	0	0	0	0	0	0	0	0	769	78%	
OLG	0	0	0	1	0	0	0	0	0	70	11	268	1	1551	21	3	1	1	2	1	0	39	0	0	1	0	0	0	0	0	1971	79%	
FRT	0	22	0	0	0	0	0	4	0	1	13	18	4	115	652	0	1	0	86	10	0	31	0	0	0	0	0	0	0	0	957	68%	
WHT	10	4	0	0	0	0	0	0	0	0	0	0	1	0	0	10047	56	2	49	124	0	53	54	10	0	0	0	0	0	10410	97%		
BRL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	200	256	13	40	14	0	5	0	0	0	0	0	0	0	0	528	48%	
OAT	0	3	0	0	0	0	0	0	0	19	0	3	0	0	0	11	2	93	72	0	0	3	0	0	0	0	0	0	0	0	206	45%	
MAI	0	13	0	0	0	0	0	1	0	0	34	2	15	17	18	25	15	0	10523	751	2	485	0	0	3	0	0	0	0	0	11904	88%	
CTN	3	70	0	7	54	0	0	2	0	7	0	16	12	42	12	127	79	82	285	20568	2	359	22	0	0	47	0	0	0	0	21796	94%	
TBC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	76	0	0	100	1337	22	0	0	0	0	0	0	0	0	1536	87%		
CLV	0	14	0	0	0	0	0	0	0	280	0	19	21	43	21	487	0	4	171	1529	0	10335	0	0	0	0	0	0	0	0	12924	80%	
GRF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	480	0	1	0	43	0	3	566	14	0	0	0	0	0	0	1107	51%	
FLW	0	11	0	0	0	0	0	0	0	31	0	7	1	2	7	163	0	0	1	67	0	32	0	326	0	0	0	0	0	0	648	50%	
SVA	3	8	0	0	0	0	0	0	0	169	0	107	0	9	0	0	0	0	0	1	0	0	0	0	3385	0	22	0	0	3704	91%		
BDS	2	5	2	1	64	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2673	5	0	119	0	0	2871	93%
RCK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2	2755	0	2	0	0	2762	100%
MRS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	611	95%	
WCR	0	0	0	0	0	0	0	0	6	0	4	0	0	0	0	0	0	0	0	0	2	0	0	0	0	2	1	0	1081	0	0	1096	99%
WBD	0	0	0	0	0	0	0	0	0	0	8	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	35	11169	0	11214	100%
CWT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	7200	7202	100%	
sum	1160	1610	477	322	1760	438	622	12528	34426	5005	9459	5718	668	1920	766	11635	410	199	11278	23297	1341	11407	642	351	3751	3975	2794	582	1270	11171	7200	kappa:	92%
UA	78%	76%	88%	93%	89%	100%	97%	97%	100%	87%	97%	89%	90%	81%	85%	86%	62%	47%	93%	88%	100%	91%	88%	93%	90%	67%	99%	100%	85%	100%	100%	OA:	93%

Πίνακας 6 Πίνακας Σύγκρισης του πέμπτου πειράματος (με κόκκινο φαίνονται οι κατηγορίες στις οποίες κατευθύνονται σφάλματα παράλειψης>5%) με OA=92,95% και MO f1=84,63%

4.1.6 Πείραμα 6: Μοντέλο χρονικών χαρακτηριστικών με DSM και εξαγωγή βάσει σημαντικότητας (75%)

Σε αυτή την περίπτωση στον κύβο ο οποίος περιγράφηκε παραπάνω με τα χρονικά χαρακτηριστικά του προηγούμενου πειράματος προστέθηκε ένα νέο χρονικό χαρακτηριστικό, το εκατοστημόριο 75%. Η εξαγωγή καναλιών οδηγεί σε έναν κύβο με 18 κανάλια αντί των αρχικών 60. Σε αυτό το πείραμα η συνολική ακρίβεια είναι 92,43% και ο μέσος δείκτης f_1 είναι 83,44%, δηλαδή παρατηρούνται παρεμφερείς ακρίβειες με τα προηγούμενα πειράματα.

Για τις κατηγορίες του αστικού ιστού (DUF, SUF, ICU, RAN, MES, PHV, GRH) οι ακρίβειες παραγωγού κυμαίνονται στο εύρος 57%-97% και είναι ελάχιστα χαμηλότερα από το προηγούμενο πείραμα το οποίο είχε 32 περισσότερα κανάλια, για το άνω όριο του εύρους και ελάχιστα ψηλότερα για το κάτω όριο του εύρους. Η κατηγορία με τη μικρότερη ακρίβεια είναι αυτή των μεταλλείων, (57%), ενώ αυτή με τη μεγαλύτερη τα θερμοκήπια (97%). Οι ακρίβειες χρήστη κινούνται στο εύρος 77%-98%, με τη μικρότερη ακρίβεια χρήστη να είναι στην κατηγορία της αραιής αστικής δόμησης και την υψηλότερη στα φωτοβολταϊκά. Παρατηρείται λοιπόν ότι οι ακρίβειες παραγωγού και χρήστη μεταβάλλονται μεν ανάμεσα στα πειράματα, ωστόσο αυτό που βασικά παρατηρείται είναι, όπως και στο πείραμα 3, ότι οι χαμηλότερες ακρίβειες χρήστη και παραγωγού αυξάνονται όσο μειώνεται ο αριθμός των καναλιών ενώ οι μεγαλύτερες ακρίβειες χρήστη και παραγωγού μειώνονται ανάλογα με τον αριθμό των καναλιών.

Οι κατηγορίες των κωνοφόρων και φυλλοβόλων δέντρων (CNF, BLF) έχουν ακρίβειες χρήστη 97% και 100% και παραγωγού 97% και 99% αντίστοιχα, οι οποίες παραμένουν ουσιαστικά αμετάβλητες σε σχέση με προηγούμενα πειράματα.

Για τις κατηγορίες που σχετίζονται με τις χορτολιβαδικές εκτάσεις (NGR, DSV, SSV) και την σκληροφυλλική βλάστηση παρατηρείται ότι το εύρος για τις ακρίβειες παραγωγού είναι 96%-100% και το εύρος ακριβειών χρήστη είναι 88%-100%. Η κατηγορία με τη μικρότερη ακρίβεια παραγωγού είναι η αραιή σκληροφυλλική βλάστηση με ακρίβεια παραγωγού 96% και αυτή με τη μικρότερη ακρίβεια χρήστη (UA=88%) είναι εκτάσεις που καλύπτονται από γρασίδι NGR και παρατηρείται ότι υπάρχει άνοδος σε αυτές τις ακρίβειες σε σχέση με προηγούμενα πειράματα.

Στις κατηγορίες των καλλιεργειών (VNY, OLG, FRT, WHT, BRL, OAT, MAI, CTN, TBC, FLW, GRF) παρατηρούμε ακρίβειες παραγωγού οι οποίες βρίσκονται στο εύρος 33%-95%. Εδώ βλέπουμε ότι ενώ το άνω όριο του εύρους παραμένει σχετικά σταθερό, το κάτω όριο μειώνεται σημαντικά (33% από 45%). Παρατηρείται δηλαδή ότι οι ακρίβειες παραγωγού οι οποίες ήταν χαμηλότερες ή λίγο υψηλότερα από 50% μειώνονται σημαντικά (δηλαδή παραπάνω από 3%) σε αυτή την περίπτωση, ενώ αυτές που βρίσκονται κοντά στο άνω όριο των ακριβειών δεν έχουν σημαντικές αυξομειώσεις. Η μικρότερη

ακρίβεια παραγωγού είναι και πάλι στην κατηγορία της βρώμης (PA=33%). Σε σχέση με τις ακρίβειες χρήστη αυτές βρίσκονται στο εύρος 33%-98% και για αυτές εξάγονται τα ίδια συμπεράσματα που εξήχθησαν και για την περίπτωση των ακριβειών παραγωγού. Τη μικρότερη ακρίβεια χρήστη την έχουμε για την κατηγορία της βρώμης και πάλι, η οποία είναι UA=33%.

Οι κατηγορίες που είναι σχετικές με το γυμνό έδαφος (BDS,RCK,MRS) έχουν ακρίβειες παραγωγού 92%-100% και ακρίβειες χρήστη 64%-99%. Σε αυτή την περίπτωση παρατηρείται ότι σε σχέση με τα προηγούμενα πειράματα οι ακρίβειες παραγωγού και χρήστη παραμένουν πρακτικά αμετάβλητες για τις κατηγορίες εκείνες που είχαν ήδη υψηλές ακρίβειες και μειώνονται σημαντικά για τις κατηγορίες που δεν ισχύει κάτι τέτοιο.

Εν τέλει για τα υδάτινα σώματα (MRS, WCR, WBD, CWT) παρατηρούνται οι ίδιες ακρίβειες παραγωγού και χρήστη με το προηγούμενο πείραμα οι οποίες κινούνται στο εύρος 98%-100% και 86%-100% αντιστοίχως. Οι ακρίβειες δε διαφοροποιούνται σχεδόν καθόλου σε σχέση με τις προηγούμενες περιπτώσεις.

Σε ό,τι αφορά τα κανάλια τα οποία επιλέχθηκαν παρατηρείται ότι τα μετασχηματισμένα κανάλια περιλαμβάνουν και τους έξι δείκτες. Αυτό το οποίο παρατηρείται είναι ότι, σε συνέχεια του προηγούμενου πειράματος, διατηρούνται όλα τα χαρακτηριστικά του υπερκύβου που ήταν σημαντικά και ειδικότερα παρατηρούμε ότι διατηρούνται τα χρονικά χαρακτηριστικά median, mean, max κυρίως και δευτερευόντως τα υπόλοιπα. Πολύ ενδιαφέρον είναι το γεγονός ότι υπάρχουν λίγα κανάλια με τον δείκτη του εκατοστημρίου (percentile), αλλά με σχετικά μεγάλη σημαντικότητα (είναι στις πρώτες 10 θέσεις).

ΠΙΝΑΚΑΣ ΣΥΓΧΥΣΗΣ ΥΠΕΡΚΥΒΟΥ ΦΑΣΜΑΤΟΧΡΟΝΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΜΑΖΙ ΜΕ DEM ΚΑΙ ΕΞΑΓΩΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΜΕ ΕΚΑΤΟΣΤΗΜΟΡΙΟ 75																																	
	DUF	SUF	ICU	RAN	MES	PHV	GRH	BLF	CNF	NGR	DSV	SSV	VNY	OLG	FRT	WHT	BRL	OAT	MAI	CTN	TBC	CLV	GRF	FLW	SVA	BDS	RCK	MRS	WCR	WBD	CWT	sum	PA
DUF	912	105	18	1	13	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	1053	87%	
SUF	111	1217	0	4	6	1	0	0	0	1	0	38	2	10	3	0	0	2	4	15	0	1	0	1	21	3	2	0	2	0	1444	84%	
ICU	20	15	378	1	37	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	81	0	0	0	544	69%		
RAN	66	41	0	278	5	5	0	8	0	6	2	16	0	1	2	1	0	0	2	3	0	0	0	0	38	11	3	0	2	490	57%		
MES	12	2	7	3	1470	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	1345	0	0	2	2846	52%		
PHV	0	8	0	0	0	437	0	2	0	29	0	33	0	2	0	7	1	0	14	6	0	0	0	0	19	0	0	4	0	562	78%		
GRH	3	7	2	0	4	0	595	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	612	97%		
BLF	0	0	0	0	0	0	0	12160	5	32	23	3	0	1	9	0	0	0	0	0	0	0	0	0	15	0	0	0	0	12248	99%		
CNF	0	0	0	0	0	0	0	63	34374	0	10	2	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	1	34461	100%		
NGR	0	0	0	0	0	1	0	20	0	4309	0	15	0	2	0	0	0	2	3	24	0	16	0	113	0	0	0	0	0	4505	96%		
DSV	1	0	0	0	0	0	0	293	4	0	9241	51	0	27	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9618	96%		
SSV	0	44	0	0	0	0	0	25	0	2	160	5035	2	103	7	2	0	1	0	0	0	0	0	202	0	0	0	0	0	5583	90%		
VNY	0	4	0	0	0	0	0	2	0	3	1	41	493	19	33	0	0	0	116	54	0	1	0	2	0	0	0	0	0	769	64%		
OLG	0	1	0	1	0	0	0	0	0	53	10	226	7	1616	19	3	0	0	1	0	0	33	0	0	1	0	0	0	0	1971	82%		
FRT	0	22	0	0	0	0	0	3	0	0	8	18	8	93	660	0	1	0	113	13	0	18	0	0	0	0	0	0	0	957	69%		
WHT	0	3	0	13	0	0	0	0	0	0	0	0	0	0	0	9838	104	11	92	274	0	52	18	3	2	0	0	0	0	10410	95%		
BRL	0	0	0	0	0	0	0	0	0	1	0	1	2	3	0	206	245	20	29	16	0	5	0	0	0	0	0	0	0	528	46%		
OAT	0	1	0	0	0	0	0	0	0	8	0	3	1	0	0	13	3	67	106	0	0	4	0	0	0	0	0	0	0	206	33%		
MAI	0	19	0	0	0	0	0	0	0	0	0	35	4	17	7	21	57	6	5	10467	727	12	526	0	1	0	0	0	11904	88%			
CTN	3	39	0	12	64	0	0	2	0	5	0	17	6	58	21	131	77	78	500	20267	10	426	19	0	24	37	0	0	21796	93%			
TBC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	76	0	0	0	59	1380	21	0	0	0	0	0	0	1536	90%			
CLV	0	11	0	0	0	0	0	0	0	251	0	24	12	51	21	384	0	12	172	1669	0	10315	1	1	0	0	0	0	12924	80%			
GRF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	526	2	2	0	21	0	7	532	17	0	0	0	0	1107	48%			
FLW	2	18	0	0	0	0	0	0	0	28	0	3	12	3	9	141	0	0	9	43	0	38	1	338	0	0	0	3	648	52%			
SVA	0	23	0	0	0	0	0	0	0	174	1	91	0	8	0	0	0	0	0	0	0	0	0	0	3390	0	17	0	3704	92%			
BDS	2	6	1	1	91	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2638	6	0	126	0	2871	92%	
RCK	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2757	0	1	0	2762	100%	
MRS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	611	0	0	0	611	100%
WCR	1	0	0	0	1	0	0	0	6	0	5	1	0	0	0	0	0	0	0	0	0	0	0	0	1	5	0	0	1075	1	0	1096	98%
WBD	0	1	0	0	0	0	0	0	0	0	7	3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	44	11158	0	11214	100%
CWT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7202	100%	
sum	1135	1588	406	314	1691	444	608	12578	34389	4902	9503	5625	562	2004	806	11385	439	200	11628	23192	1402	11463	571	362	3844	4124	2785	618	1253	11159	7202	καρρα:	92%
UA	80%	77%	93%	89%	87%	98%	98%	97%	100%	88%	97%	90%	88%	81%	82%	86%	56%	34%	90%	87%	98%	90%	93%	93%	88%	64%	99%	99%	86%	100%	100%	OA:	92%

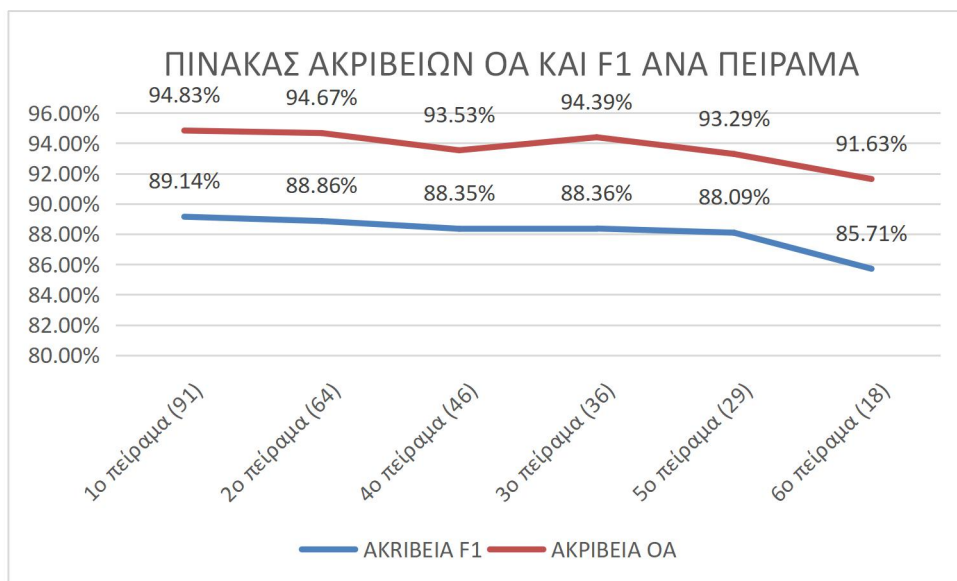
Πίνακας 7 Πίνακας Σύγκρισης του έκτου πειράματος (με κόκκινο φαίνονται οι κατηγορίες στις οποίες κατευθύνονται σφάλματα παράλειψης>5%) και εκατοστημόριο 75 με OA=92,43% και MO f1=83,44%

4.2 Συγκριτική Παρουσίαση των Πειραμάτων για το Tile 34SEJ

Παρακάτω γίνεται μια συγκριτική παρουσίαση για τις ακρίβειες των έξι πειραμάτων όπως αυτά εφαρμόστηκαν για το tile 34SEJ του δορυφόρου και τις συγχωνευόμενες 28 κατηγορίες. Παρατηρείται ότι οι ακρίβειες για το tile 34SEJ μετά τη συγχώνευση, είναι σαφώς υψηλότερες σε σχέση με τις αντίστοιχες πριν τη συγχώνευση, οπότε αυτή κρίνεται επιτυχημένη εκτός από απαραίτητη για τη συνέχεια των πειραμάτων. Η γενική ακρίβεια κινείται στο εύρος 91,63%-94,83%, ενώ η μέση ακρίβεια f1 στο εύρος 85,71%-89,14%.

ΑΚΡΙΒΕΙΕΣ F1,ΟΑ ΑΝΑ ΚΑΤΗΓΟΡΙΑ ΓΙΑ ΤΟ TILE 34SEJ ΜΕ ΛΙΓΟΤΕΡΕΣ ΚΑΤΗΓΟΡΙΕΣ						
CODE	ΥΠΕΡΚΥΒΟΣ ΗΜΕΡΟΜΗΝΙΩΝ +DSM (91 κανάλια)	ΥΠΕΡΚΥΒΟΣ+BRI GHTNESS +DSM ΜΑΖΙ ΜΕ ΕΞΑΓΩΓΗ ΧΑ/ΚΩΝ PERC 35 (64 κανάλια)	ΥΠΕΡΚΥΒΟΣ+BRI GHTNESS +DSM ΜΑΖΙ ΜΕ ΕΞΑΓΩΓΗ ΧΑ/ΚΩΝ PERC 65 (36 κανάλια)	ΥΠΕΡΚΥΒΟΣ TEMPORAL+ DSM (46 κανάλια)	ΥΠΕΡΚΥΒΟΣ TEMPORAL+BRI GHTNESS +DSM ΜΑΖΙ ΜΕ ΕΞΑΓΩΓΗ ΧΑ/ΚΩΝ PERC 45 (29 κανάλια)	ΥΠΕΡΚΥΒΟΣ TEMPORAL+BRI GHTNESS+PERC 75+DSM ΜΑΖΙ ΜΕ ΕΞΑΓΩΓΗ ΧΑ/ΚΩΝ PERC 46 (18 κανάλια)
DUF	81,25%	81,80%	81,25%	82,95%	82,89%	66,87%
SUF	75,95%	76,34%	73,02%	80,25%	80,25%	74,47%
ICU	79,95%	83,35%	83,64%	80,89%	84,58%	75,95%
RAN	72,03%	70,68%	66,16%	77,55%	76,86%	57,87%
MES	71,59%	70,85%	71,28%	70,23%	70,54%	67,78%
PHV	90,30%	87,86%	86,86%	88,27%	87,88%	85,99%
GRH	99,00%	99,50%	98,99%	98,48%	98,48%	98,50%
BLF	97,99%	97,99%	97,99%	97,48%	97,48%	97,99%
CNF	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
NGR	94,96%	94,48%	95,48%	92,48%	90,28%	89,96%
DSV	96,50%	96,50%	96,50%	96,99%	96,99%	95,99%
SSV	90,96%	90,96%	90,99%	89,99%	88,99%	89,50%
VNY	89,46%	88,87%	86,86%	85,43%	82,99%	84,65%
OLG	80,42%	80,95%	84,50%	79,47%	79,95%	78,11%
FRT	82,56%	80,90%	78,99%	74,54%	73,74%	71,59%
CER	93,28%	93,28%	92,73%	90,60%	91,17%	92,83%
MAI	95,48%	94,99%	94,99%	91,90%	91,37%	90,73%
CTN	95,00%	95,00%	93,99%	92,43%	91,96%	93,99%
TBC	94,83%	94,83%	93,05%	94,83%	94,37%	93,62%
GRF	90,43%	89,90%	88,36%	86,15%	83,57%	84,84%
FLW	67,22%	62,96%	61,35%	66,12%	65,28%	65,58%
SVA	90,48%	89,99%	89,47%	90,50%	90,50%	86,71%
BDS	79,88%	79,22%	79,22%	78,86%	78,56%	78,75%
RCK	97,99%	98,99%	98,99%	98,48%	98,99%	94,91%
MRS	96,96%	95,83%	95,83%	97,44%	97,44%	92,04%
WCR	91,61%	92,04%	93,73%	91,47%	91,47%	91,04%
WBD	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
CWT	100,00%	100,00%	100,00%	100,00%	100,00%	99,50%
F1	89,14%	88,86%	88,36%	88,35%	88,09%	85,71%
O.A	94,83%	94,67%	94,39%	93,53%	93,29%	91,63%

Πίνακας 8 Ακρίβειες f1, ΟΑ ανά κατηγορία για το tile 34SEJ με λιγότερες κατηγορίες



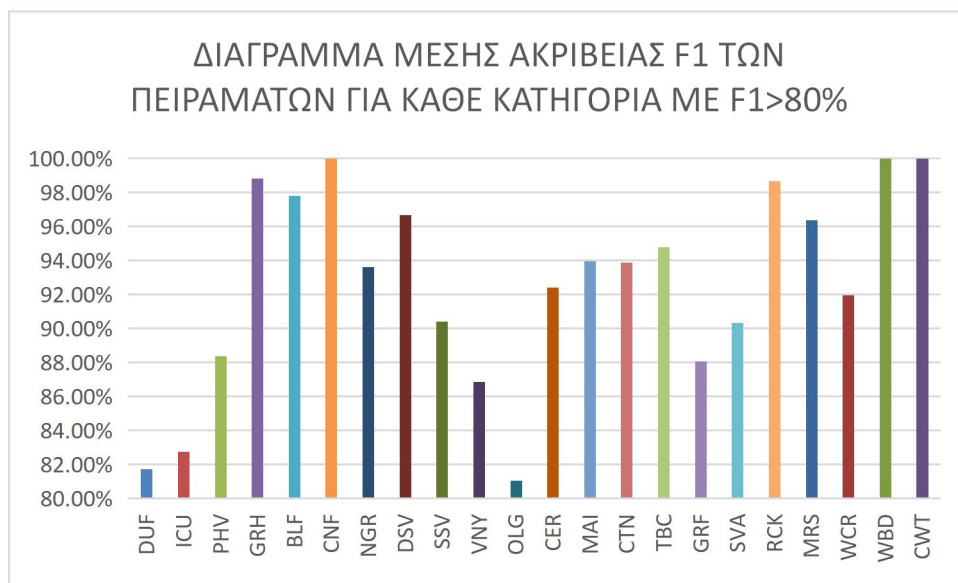
Σχήμα 20 Διάγραμμα ακριβειών ΟΑ και f1 ανά πείραμα (σε παρένθεση ο αριθμός των καναλιών)

Σε αυτό το tile οι παρατηρήσεις για τα πειράματα είναι οι εξής:

- Το πρώτο πείραμα πετυχαίνει υψηλές ακρίβειες f1 με εύρος 71,59%-100% (σαφώς ανώτερες από τις αντίστοιχες του αντίστοιχου πειράματος πριν τη συγχώνευση. Όλες οι ακρίβειες f1 καθώς και η γενική ακρίβεια καταδεικνύουν ότι το πείραμα είναι πολύ επιτυχημένο παρά την αυξημένη διαστατικότητα σε σχέση με τα υπόλοιπα.
- Το δεύτερο πείραμα έχει ακρίβεια παρεμφερή με αυτή του προηγούμενου πειράματος $\pm 1\%$ με εξαίρεση την κατηγορία της αγρανάπαυσης όπου η ακρίβεια f1 μειώνεται κατά 5%. Η μέση f1 είναι 88,86%. Μάλιστα τα κανάλια είναι 64 αντί των 91 αρχικών.
- Το τρίτο πείραμα έχει υψηλές ακρίβειες συγκρίσιμες με αυτές των δύο προηγούμενων για τις περισσότερες περιπτώσεις ($\pm 1\%$), με εξαίρεση τις κατηγορίες του αστικού ιστού για τις οποίες παρατηρούνται σημαντικές διαφοροποιήσεις σε σχέση με το πρώτο πείραμα ($\text{anf1}=88,36\%$). Η διαστατικότητα σε αυτή την περίπτωση είναι 36 κανάλια.
- Το τέταρτο πείραμα πετυχαίνει ακρίβειες οι οποίες στην πλειοψηφία των κατηγοριών βρίσκονται 1-3% κάτω από τα πειράματα του υπερκύβου φασματικών χαρακτηριστικών ($\text{anf1}=88,35\%$). Οι καλλιέργειες ειδικά έχουν ακρίβειες 1-8% χαμηλότερες σε σχέση με το πρώτο πείραμα. Παρ' όλα αυτά για τρεις κατηγορίες της αστικής δόμηση (RAN, SUF, DUF) επιτυγχάνονται με αυτό το πείραμα ακρίβειες έως και 5% ανώτερες σε σχέση με τα προηγούμενα πειράματα.
- Το πέμπτο πείραμα έχει πολύ μειωμένη διαστατικότητα (29 κανάλια) αλλά και τις μικρότερες ακρίβειες σε σχέση με τα προηγούμενα πειράματα ($\text{anf1}=88,09\%$).
- Το έκτο πείραμα έχει, αναμενόμενα, τη χαμηλότερη γενική ακρίβεια και μέση f1 σε σχέση με τα

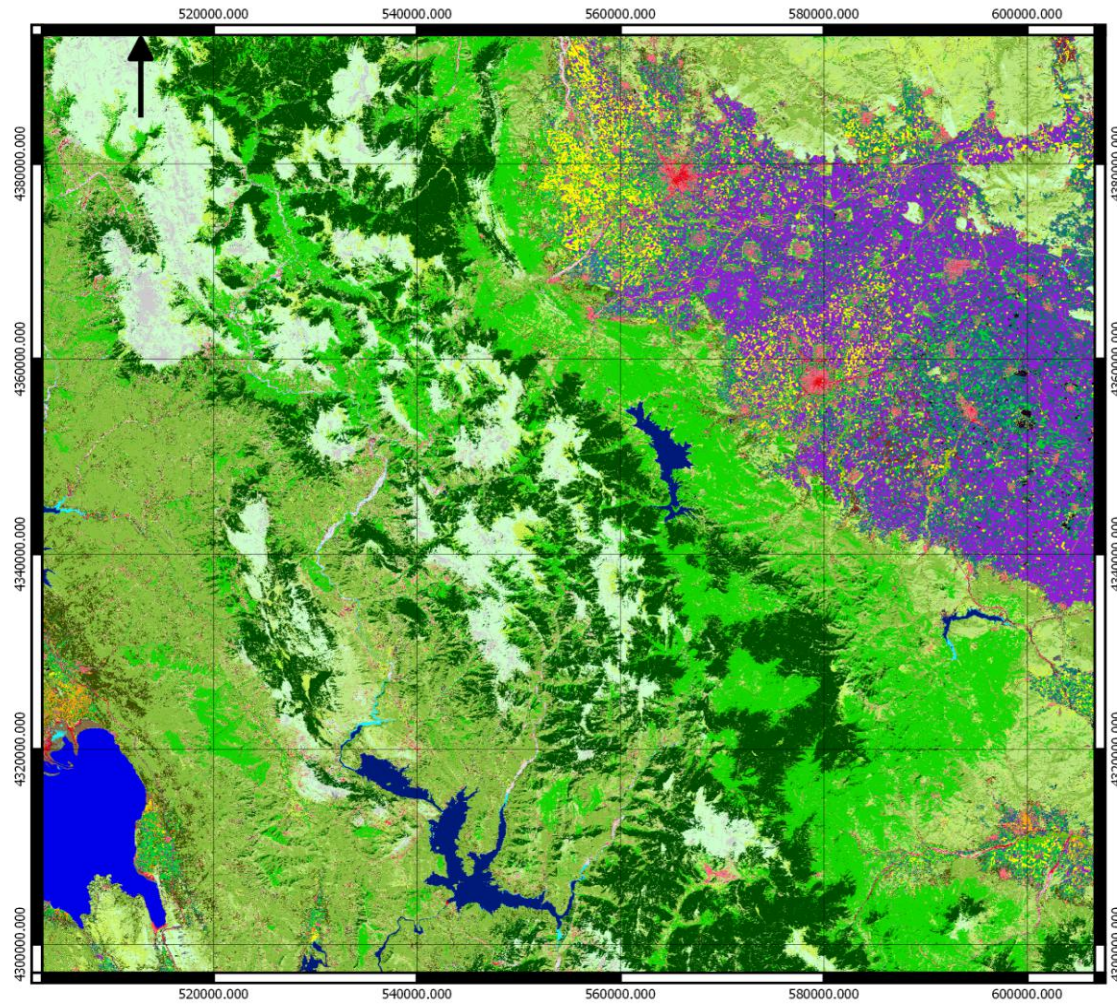
προηγούμενα πειράματα (OA=91,63% και $anf1=85,71\%$) αλλά η διαστατικότητα για αυτή την περίπτωση είναι πολύ σημαντικά μειωμένη με μόλις 18 κανάλια στον υπερκύβο.

Στην συνέχεια θα αναλυθούν λίγο παραπάνω οι κατηγορίες με υψηλή ακρίβεια όπως αυτές φαίνονται στο παρακάτω διάγραμμα, όπου ως κατηγορίες υψηλής ακρίβειας ορίζονται αυτές για τις οποίες το μέσο $f1 > 80\%$. Το μέσο $f1$ ορίζεται ως ο μέσος όρος των έξι πειραμάτων στο tile:



Σχήμα 21 Διάγραμμα μπάρα για κάθε κατηγορία με $anf1 > 80\%$

Παρατηρείται ότι η ακρίβεια των περισσότερων καλλιεργειών είναι πολύ υψηλή, γεγονός που οφείλεται στην αφθονία πολυγώνων εκπαίδευσης που υπήρχαν εξαιτίας του μεγάλου ποσοστού κάλυψης της περιοχής από καλλιέργειες, καθώς επίσης και της μη ύπαρξης φασματικά όμοιων κατηγοριοποιήσεων στην περιοχή όπως συνέβη για το 345EJ με τις 31 κατηγορίες. Επιπλέον παρατηρείται το φαινόμενο η μεγάλη πλειοψηφία των κατηγοριών να έχουν ακρίβειες $> 80\%$. Αυτό εξηγείται από τον πολύ μεγάλο αριθμό πολυγώνων εκπαίδευσης που συλλέχθηκαν αλλά και την αξιοσημείωτη ακρίβειά τους, η οποία επετεύχθη χάρη στους εκτενείς ελέγχους που έγιναν από την υποψήφια διδάκτορα Χριστίνα Καρακίζη. Επίσης παρατηρείται ότι με τη συγχώνευση των κατηγοριών της βρώμης, του σιταριού και του κριθαριού επιτυγχάνεται πολύ μεγάλη ακρίβεια για την κατηγορία που προήλθε από την συγχώνευση (δημητριακά), παρότι η βρώμη είχε μικρή ακρίβεια στην προηγούμενη περίπτωση. Επομένως οι κατηγορίες αυτές μπορούν να θεωρηθούν ως μη φασματικά διαχωρίσιμες. Συνολικά 22 από τις 28 κατηγορίες σε αυτή την περίπτωση έχουν ακρίβεια άνω του 80% και τα αποτελέσματα κρίνονται ιδιαίτερα ικανοποιητικά. Ενδιαφέρον έχει και το γεγονός πως δεν υπάρχουν κατηγορίες με ακρίβειες κάτω από 60% και άρα δε θα φτιαχτεί αντίστοιχο διάγραμμα για τις κατηγορίες με χαμηλές ακρίβειες $f1$.



**ΧΑΡΤΗΣ ΚΑΛΥΨΕΩΝ ΓΗΣ ΜΕ
ΤΑΞΙΝΟΜΗΣΗ ΥΠΕΡΚΥΒΟΥ
ΧΡΟΝΟΣΕΙΡΑΣ 2016 ΚΑΙ DSM
ΑΝΑΛΥΣΗΣ 10 Μ ΣΤΟ TILE
34SEJ ΤΟΥ SENTINEL**

ΓΕΩΔΑΙΤΙΚΟ ΣΥΣΤΗΜΑ ΑΝΑΦΟΡΑΣ: WGS 84
ΧΑΡΤΟΓΡΑΦΙΚΗ ΠΡΟΒΟΛΗ: ΠΑΓΚΟΣΜΙΑ
ΕΓΚΑΡΣΙΑ ΜΕΡΚΑΤΟΡΙΚΗ ΠΡΟΒΟΛΗ

ΥΠΟΜΝΗΜΑ

■ DUF	■ FRT
■ SUF	■ CER
■ ICU	■ MAI
■ RAN	■ CTN
■ MES	■ TBC
■ PHV	■ GRF
■ GRH	■ FLW
■ BLF	■ SVA
■ CNF	■ BDS
■ NGR	■ RCK
■ DSV	■ MRS
■ SSV	■ WCR
■ VNY	■ WBD
■ OLG	■ CWT



Σχήμα 22 Χάρτης καλύψεων γης tile 34SEJ

4.3 Συγκριτική Παρουσίαση των Πειραμάτων για το Tile 34TEK

Παρακάτω γίνεται μια συγκριτική παρουσίαση για τις ακρίβειες των έξι πειραμάτων όπως αυτά εφαρμόστηκαν για το tile 34TEK του δορυφόρου. Καταρχάς, κάποιες παρατηρήσεις σε σχέση με την γενική ακρίβεια (ΟΑ) είναι ότι αυτή κινείται σε πολύ υψηλά επίπεδα για όλα τα πειράματα (>94%). Μάλιστα παρατηρείται ότι η γενική ακρίβεια για αυτή τη σειρά είναι σε όλες τις περιπτώσεις 1-2,5% υψηλότερη σε σχέση με την αντίστοιχη για τις ταξινομήσεις του προηγούμενου tile.

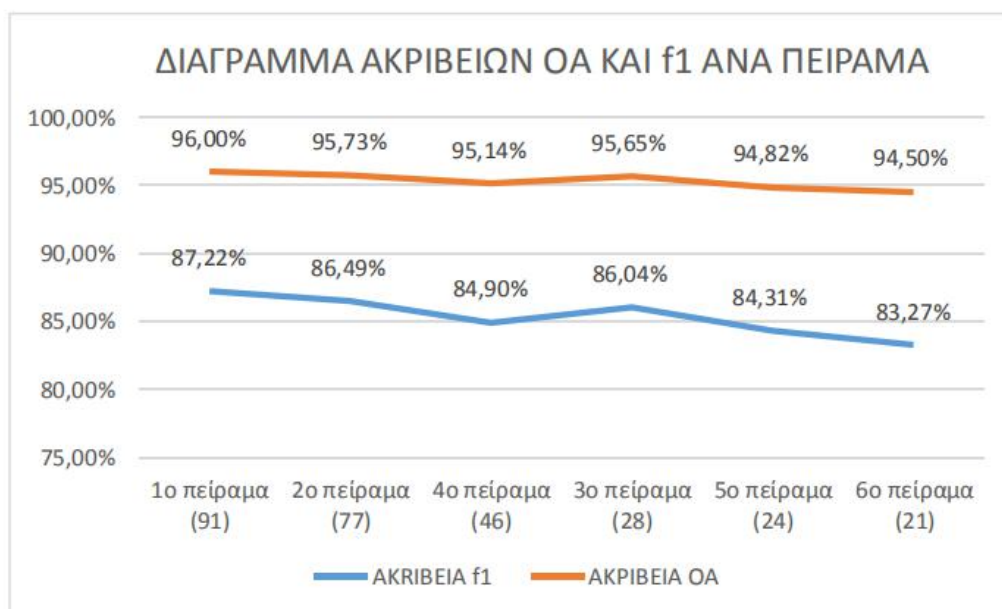
Αξίζει ωστόσο να σημειωθεί ότι οι μέσες ακρίβειες f1 είναι σημαντικά μειωμένες σε σχέση με την προηγούμενη σειρά πειραμάτων, με τη μείωση αυτή να κινείται μεταξύ του 2% και 3,5% ανάμεσα στα αντίστοιχα πειράματα.

ΑΚΡΙΒΕΙΕΣ F1,ΟΑ ΑΝΑ ΚΑΤΗΓΟΡΙΑ ΓΙΑ ΤΟ TILE 34TEK ΜΕ ΛΙΓΟΤΕΡΕΣ ΚΑΤΗΓΟΡΙΕΣ						
CODE	ΥΠΕΡΚΥΒΟΣ Σ ΗΜΕΡΟΜΗ ΝΙΩΝ+DSM (91 κανάλια)	ΥΠΕΡΚΥΒΟΣ+BRI GHTNESS +DSM ΜΑΖΙ ΜΕ ΕΞΑΓΩΓΗ ΧΑ/ΚΩΝ PERC 35 (77 κανάλια)	ΥΠΕΡΚΥΒΟΣ+BRI GHTNESS +DSM ΜΑΖΙ ΜΕ ΕΞΑΓΩΓΗ ΧΑ/ΚΩΝ PERC 65 (24 κανάλια)	ΥΠΕΡΚΥΒΟΣ Σ TEMPORAL +DSM (46 κανάλια)	ΥΠΕΡΚΥΒΟΣ TEMPORAL+BRI GHTNESS +DSM ΜΑΖΙ ΜΕ ΕΞΑΓΩΓΗ ΧΑ/ΚΩΝ PERC 45 (28 κανάλια)	ΥΠΕΡΚΥΒΟΣ TEMPORAL+BRIGHTN ESS+PERC75+DSM ΜΑΖΙ ΜΕ ΕΞΑΓΩΓΗ ΧΑ/ΚΩΝ PERC 46 (21 κανάλια)
DUF	55,73%	52,71%	50,75%	51,18%	48,07%	48,88%
SUF	66,81%	66,81%	65,03%	62,08%	62,44%	63,75%
ICU	60,81%	57,92%	55,58%	45,25%	42,44%	39,91%
RAN	97,48%	96,96%	96,96%	95,99%	95,50%	96,48%
MES	90,46%	89,29%	88,87%	87,08%	85,23%	84,84%
PHV	93,05%	92,47%	93,05%	93,05%	91,89%	87,48%
GRH	99,50%	98,99%	100,00%	97,99%	99,50%	99,50%
BLF	100,00%	100,00%	100,00%	100,00%	100,00%	99,50%
CNF	99,50%	99,50%	99,50%	99,00%	99,00%	98,50%
NGR	92,50%	92,50%	90,50%	88,50%	87,50%	84,95%
DSV	94,91%	94,43%	93,90%	94,43%	94,43%	94,43%
SSV	88,36%	88,36%	88,36%	87,90%	87,90%	86,90%
VNY	81,80%	80,35%	78,89%	74,67%	72,47%	72,00%
OLG	89,60%	89,16%	89,16%	86,15%	85,58%	84,14%
FRT	84,81%	85,36%	84,14%	83,26%	84,14%	83,57%
CER	93,18%	92,73%	92,17%	91,17%	90,03%	90,73%
MAI	97,00%	96,50%	97,50%	96,48%	96,48%	95,99%
TBC	78,95%	75,67%	73,74%	78,50%	73,50%	65,03%
GRF	84,71%	85,26%	84,81%	81,80%	79,89%	77,95%
FLW	54,35%	51,83%	53,17%	56,46%	57,93%	61,17%
SVA	95,50%	94,99%	95,50%	93,99%	93,99%	93,48%
BDS	91,50%	90,96%	90,96%	84,42%	86,27%	79,13%
RCK	96,99%	96,50%	95,48%	96,50%	95,50%	93,48%
MRS	85,99%	85,71%	85,71%	86,63%	84,97%	84,33%
WCR	94,29%	93,73%	93,18%	94,83%	97,48%	99,00%
WBD	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
AVERAGE F1	87,22%	86,49%	86,04%	84,90%	84,31%	83,27%
O.A	96,00%	95,73%	95,65%	95,14%	94,82%	94,50%

Πίνακας 9 Ακρίβειες f1, ΟΑ ανά κατηγορία για το tile 34TEK με λιγότερες κατηγορίες

Σε ειδικότερη ανάλυση παρατηρείται ότι για άλλη μια φορά στην πλειοψηφία των κατηγοριών το πρώτο πείραμα δίνει και τα καλύτερα αποτελέσματα. Εξάιρεση είναι η κατηγορία των θερμοκηπίων

όπου όμως η ακρίβεια έχει εύρος 98%-100%, οπότε πρακτικά η κατηγορία αυτή διακρίνεται σωστά σε απόλυτο βαθμό. Στην περίπτωση των καλλιεργειών πάλι παρατηρούμε ότι τα δέντρα με καρπούς ταξινομούνται με 1% μεγαλύτερη ακρίβεια στο δεύτερο πείραμα (εξαγωγή χαρακτηριστικών με εκατοστημόριο 35%). Η κατηγορία των ζωοτροφών παρατηρείται ότι ταξινομείται με την καλύτερη ακρίβεια σε ποσοστό 85,26% σωστά στο δεύτερο πείραμα. Αξιοσημείωτη περίπτωση αποτελεί η αγρανάπαυση η οποία συναντιέται ελάχιστα ως κατηγορία στην περιοχή και έχει ως εκ τούτου πολύ χαμηλή ακρίβεια f1, αλλά πετυχαίνει την καλύτερη της επίδοση στο έκτο πείραμα (61,17%) το οποίο έχει τα λιγότερα κανάλια. Τις υψηλότερες ακρίβειες πετυχαίνει στο συγκεκριμένο πείραμα και η κατηγορία των βάλτων (99%). Παρακάτω φαίνεται το διάγραμμα των πειραμάτων συναρτήσει της ακρίβειας η οποία επιτυγχάνεται, με τα πειράματα να τοποθετούνται στον άξονα x ανάλογα με τον αριθμό των καναλιών τους. Το πείραμα αυτό χρησιμοποιείται για να υπάρχει καλύτερη γενική εποπτεία των ακριβειών:



Σχήμα 23 Διάγραμμα ακριβειών ΟΑ και f1 ανά πείραμα (σε παρένθεση ο αριθμός των καναλιών)

Σε αυτό το tile οι παρατηρήσεις για τα πειράματα είναι οι εξής:

- Το πρώτο πείραμα δίνει τις καλύτερες ακρίβειες f1 ανά κατηγορία, με όλες να κινούνται σε ποσοστά άνω του 50%. Ωστόσο η διαστατικότητα του είναι σαφώς μεγαλύτερη από αυτές των υπολοίπων πειραμάτων. Η μέση ακρίβεια f1 είναι 87,22% και τα κανάλια 91
- Το δεύτερο πείραμα δίνει παρεμφερή, στην πλειοψηφία των κατηγοριών, αποτελέσματα με το πρώτο (μέγιστη διαφοροποίηση 3%) παρά τη μείωση της διαστατικότητας κατά 14 κανάλια. Η μέση ακρίβεια f1 είναι 86,49% και ο αριθμός των καναλιών 77.
- Το τρίτο πείραμα δίνει σημαντικά μειωμένες ακρίβειες σε σχέση με τα προηγούμενα, αλλά ως προς την πλειονότητα των περιπτώσεων υψηλότερες σε σχέση με τον κύβο των χρονικών

χαρακτηριστικών, παρότι έχει τα μισά κανάλια (έως και 10% υψηλότερες ακρίβειες για την κατηγορία των ICU σε σχέση με το προηγούμενο πείραμα). Η μέση ακρίβεια f1 είναι 86,04% και ο αριθμός των καναλιών 46.

- Το τέταρτο πείραμα σε σχέση με τα δύο πρώτα δίνει 2,5%-3% χαμηλότερες ακρίβειες f1 ανά κατηγορία σε σχέση με όλο τον υπερκύβο φασματικών χαρακτηριστικών και μέση ακρίβεια f1=84,90% ωστόσο παρατηρείται ότι στις κατηγορίες με λίγα δεδομένα εκπαίδευσης (πχ αγρανάπαυση) επιτυγχάνονται πολύ υψηλές ακρίβειες.
- Το πέμπτο πείραμα έχει και σε αυτή την περίπτωση μειωμένη διαστατικότητα με τις ακρίβειες ωστόσο να απέχουν σημαντικά από τις ακρίβειες των προηγούμενων πειραμάτων (έως και 5,5% χαμηλότερες ακρίβειες σε σχέση με το τέταρτο πείραμα). Η μέση ακρίβεια f1 είναι 84,31% και ο αριθμός των καναλιών 28.
- Το έκτο πείραμα δίνει ακρίβειες συγκρίσιμες και πολλές φορές ανώτερες σε σχέση με το προηγούμενο πείραμα με 7 λιγότερα κανάλια (έως +1,5%). Οι μόνες εξαιρέσεις όπου οι ακρίβειες είναι πολύ κατώτερες σε σχέση με το πέμπτο είναι στις περιπτώσεις του καπνού, των βραχιδών εκτάσεων του γρασιδιού και των φωτοβολταϊκών, περιπτώσεις κατά τις οποίες η ακρίβεια κρίνεται αποδεκτή. Η μέση f1 είναι 83,27% και ο αριθμός των καναλιών 21.

Στην συνέχεια θα αναλυθούν λεπτομερέστερα οι κατηγορίες με υψηλή ακρίβεια όπως αυτές φαίνονται στο παρακάτω διάγραμμα - μπάρα. Το μέσο f1 ορίζεται ως ο μέσος όρος των έξι πειραμάτων

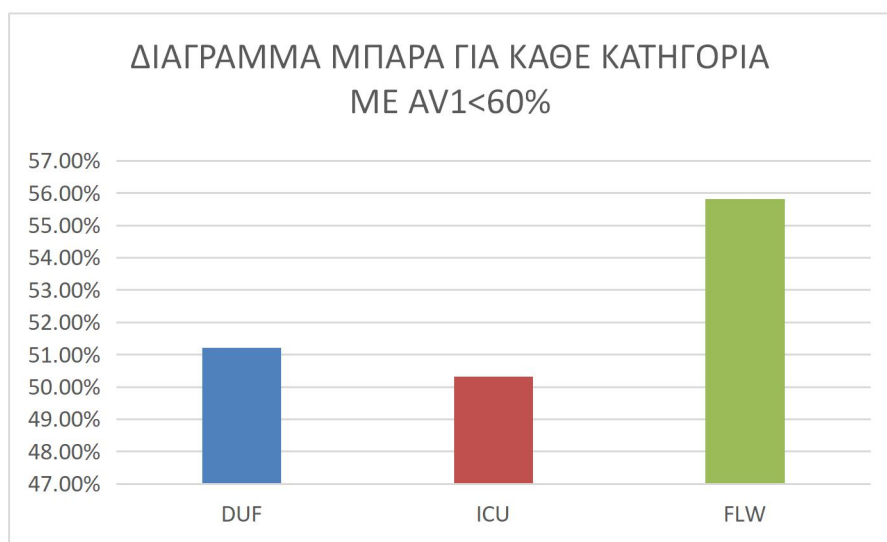
το tile:



Σχήμα 24 Διάγραμμα μπάρα για κάθε κατηγορία με $anf1 > 80\%$

Παρατηρείται ότι η ακρίβεια των περισσότερων καλλιεργειών είναι υψηλή, πράγμα που οφείλεται στην αφθονία πολυγώνων εκπαίδευσης που υπήρχαν για αυτές, εξαιτίας του μεγάλου ποσοστού κάλυψης της περιοχής από καλλιέργειες, καθώς επίσης και της μη ύπαρξης φασματικά όμοιων κατηγοριοποιήσεων στην περιοχή όπως συνέβη για το 34SEJ με τις 31 κατηγορίες. Επιπλέον, κατηγορίες οι οποίες στο προηγούμενο tile παρουσίαζαν υψηλές ακρίβειες, παρουσιάζουν και εδώ υψηλές ακρίβειες, όπως αυτές που σχετίζονται με το γυμνό έδαφος, οι οποίες είναι εύκολα φασματικά διαχωρίσιμες ή αυτές των κωνοφόρων και των φυλλοβόλων δέντρων, που έχουν ακρίβεια σχεδόν 100%. Συνολικά 20 από τις 26 κατηγορίες σε αυτή την περίπτωση έχουν ακρίβεια άνω του 80% και τα αποτελέσματα κρίνονται πολύ ικανοποιητικά.

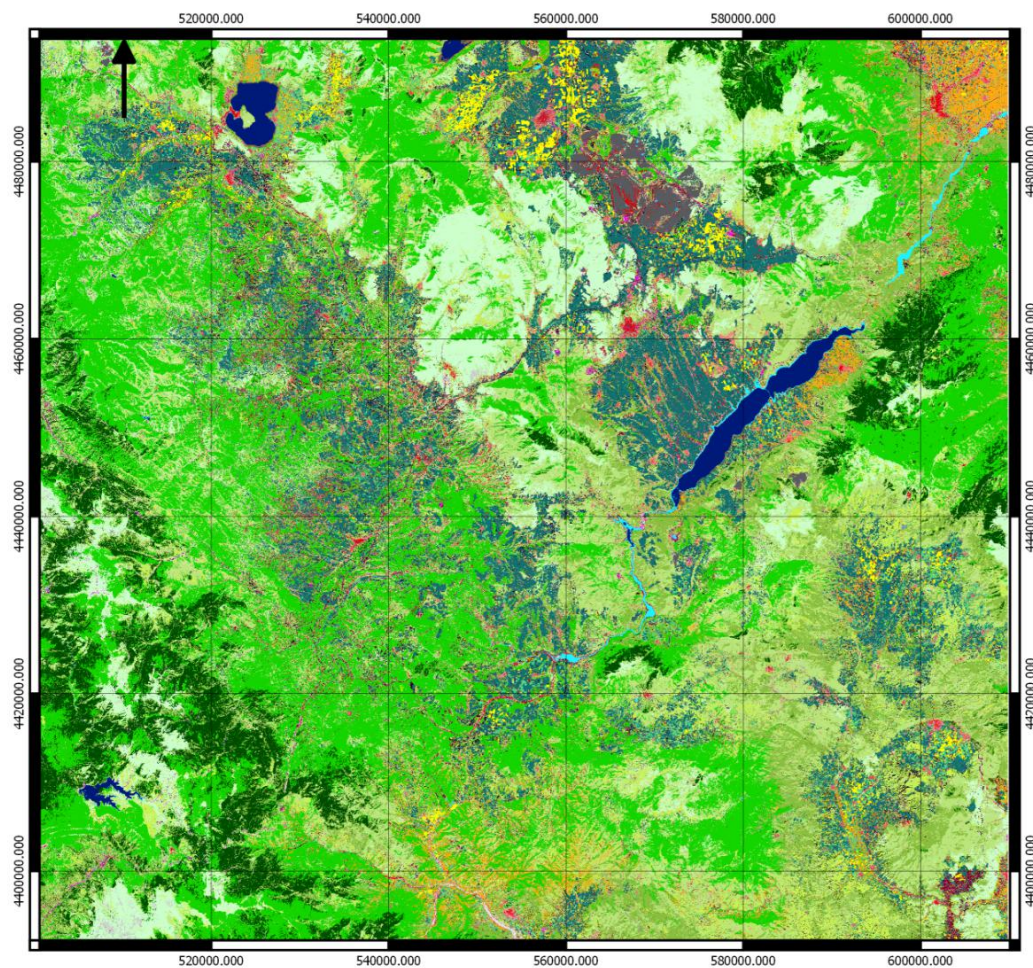
Εν τέλει γίνεται μια σύντομη ανάλυση των κατηγοριών με τις χαμηλές ακρίβειες, οι οποίες, για αυτή τη σειρά πειραμάτων, ορίζονται με $avf1 < 60\%$ και το διάγραμμα τους φαίνεται παρακάτω:



Σχήμα 25 Διάγραμμα αράχνη με $avf1 < 60\%$

Παρατηρείται ότι για αυτό το tile χαμηλές ακρίβειες έχουν οι κατηγορίες της πυκνής αστικής δόμησης (50%), της αγρανάπαυσης (55%) και των εμπορικών/βιομηχανικών μονάδων. Με έλεγχο των πινάκων σύγκυσης παρατηρείται ότι σφάλματα παράλειψης για την κατηγορία της πυκνής αστικής δόμησης κατευθύνονται προς αυτά της αραιής αστικής δόμησης, καθώς οι φασματικές υπογραφές μοιάζουν και η αραιή αστική δόμηση είναι σαφώς πιο πολυπληθής κατηγορία. Όσον αφορά στην κατηγορία των εμπορικών βιομηχανικών χρήσεων αυτές είναι ολιγοπληθείς και ως εκ τούτου συγχέονται με τα μεταλλεία που είναι πιο πολυπληθής κατηγορία και έχει όμοια φασματική υπογραφή. Η κατηγορία της αγρανάπαυσης έχει ελάχιστα πολύγωνα εκπαίδευσης.

Στην επόμενη σελίδα φαίνεται ο χάρτης καλύψεων γης, όπως έχει προκύψει για το συγκεκριμένο tile από το πρώτο πείραμα, το οποίο έδωσε και τις μεγαλύτερες ακρίβειες.



**ΧΑΡΤΗΣ ΚΑΛΥΨΕΩΝ ΓΗΣ ΜΕ
ΤΑΞΙΝΟΜΗΣΗ ΥΠΕΡΚΥΒΟΥ
ΧΡΟΝΟΣΕΙΡΑΣ 2016 ΚΑΙ DSM
ΑΝΑΛΥΣΗΣ 10 Μ ΣΤΟ TILE
34TEK ΤΟΥ SENTINEL**

ΓΕΩΔΑΙΤΙΚΟ ΣΥΣΤΗΜΑ ΑΝΑΦΟΡΑΣ: WGS 84
ΧΑΡΤΟΓΡΑΦΙΚΗ ΠΡΟΒΟΛΗ: ΠΑΓΚΟΣΜΙΑ
ΕΓΚΑΡΣΙΑ ΜΕΡΚΑΤΟΡΙΚΗ ΠΡΟΒΟΛΗ

ΥΠΟΜΝΗΜΑ

■ DUF	■ FRT
■ SUF	■ CER
■ ICU	■ ΜΑΙ
■ RAN	■ CTN
■ MES	■ TBC
■ PHV	■ GRF
■ GRH	■ FLW
■ BLF	■ SVA
■ CNF	■ BDS
■ NGR	■ RCK
■ DSV	■ MRS
■ SSV	■ WCR
■ VNY	■ WBD
■ OLG	■ CWT



Σχήμα 26 Χάρτης καλύψεων γης tile 34TEK

4.4 Συγκριτική Παρουσίαση των Πειραμάτων για το Tile 35TLF

Στη συνέχεια γίνεται μια συγκριτική παρουσίαση για τις ακρίβειες των έξι πειραμάτων, όπως αυτά εφαρμόστηκαν για το tile 35TLF του δορυφόρου. Η γενική ακρίβεια (ΟΑ) κινείται σε υψηλά επίπεδα για όλα τα πειράματα (>94,5%). Μάλιστα παρατηρείται ότι η γενική ακρίβεια για αυτή τη σειρά είναι σε όλες τις περιπτώσεις 1-2,5% καλύτερη σε σχέση με την αντίστοιχη για τις ταξινομήσεις των προηγούμενων tiles και το εύρος κινείται μεταξύ των τιμών 94,87-96,98%.

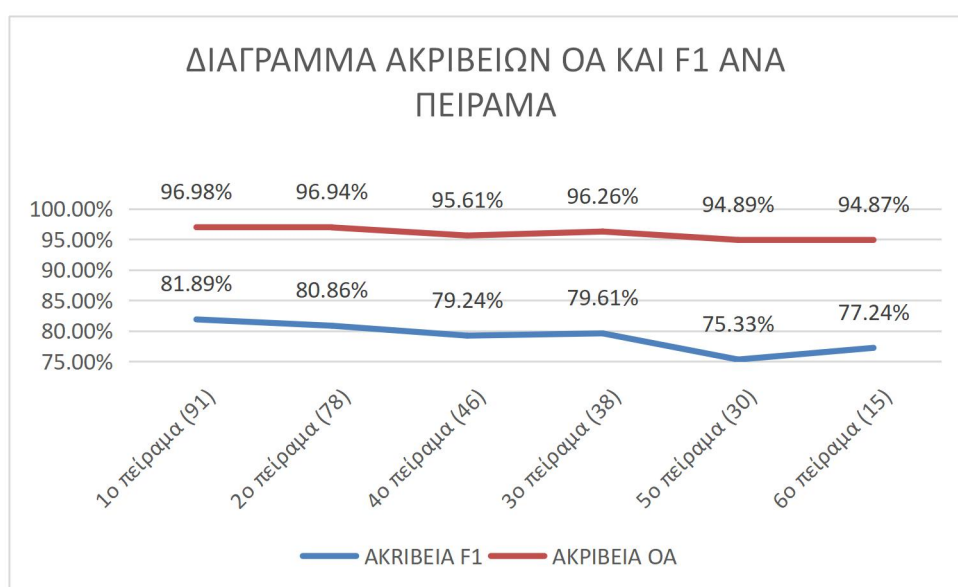
Οι μέσες ακρίβειες f1 είναι πολύ μειωμένες σε σχέση με την προηγούμενη σειρά πειραμάτων, με τη μείωση αυτή να κινείται μεταξύ του 10% και 14% στις αντίστοιχες κατηγορίες.

ΑΚΡΙΒΕΙΕΣ F1, ΟΑ ΑΝΑ ΚΑΤΗΓΟΡΙΑ ΓΙΑ ΤΟ TILE 35TLF ΜΕ ΛΙΓΟΤΕΡΕΣ ΚΑΤΗΓΟΡΙΕΣ						
CODE	ΥΠΕΡΚΥΒΟΣ ΗΜΕΡΟΜΗΝΙΩ N+DSM (91 κανάλια)	ΥΠΕΡΚΥΒΟΣ+BRIGHTNESS +DSM ΜΑΖΙ ΜΕ ΕΞΑΓΩΓΗ ΧΑ/ΚΩΝ PERC 35 (78 κανάλια)	ΥΠΕΡΚΥΒΟΣ+BRI GHTNESS +DSM ΜΑΖΙ ΜΕ ΕΞΑΓΩΓΗ ΧΑ/ΚΩΝ PERC 65 (38 κανάλια)	ΥΠΕΡΚΥΒΟΣ TEMPORAL +DSM (46 κανάλια)	ΥΠΕΡΚΥΒΟΣ TEMPORAL+BRI GHTNESS +DSM ΜΑΖΙ ΜΕ ΕΞΑΓΩΓΗ ΧΑ/ΚΩΝ PERC 45 (30 κανάλια)	ΥΠΕΡΚΥΒΟΣ TEMPORAL+BRI GHTNESS+PERC 75+DSM ΜΑΖΙ ΜΕ ΕΞΑΓΩΓΗ ΧΑ/ΚΩΝ PERC 46 (15 κανάλια)
DUF	83,26%	84,36%	78,79%	83,89%	73,66%	79,42%
SUF	80,56%	79,98%	75,36%	78,80%	62,75%	71,22%
ICU	87,90%	87,95%	87,36%	87,82%	83,35%	85,81%
RAN	96,00%	96,00%	95,00%	93,99%	89,50%	91,48%
MES	99,00%	99,00%	98,50%	99,00%	96,99%	98,50%
PHV	71,81%	71,01%	68,79%	66,96%	53,55%	62,61%
GRH	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
BLF	99,50%	99,50%	99,50%	99,50%	98,99%	98,99%
CNF	96,50%	96,50%	95,99%	94,99%	93,99%	93,50%
NGR	93,43%	92,96%	92,96%	90,90%	90,90%	87,99%
DSV	88,36%	87,36%	84,71%	78,79%	77,18%	75,36%
SSV	92,48%	92,48%	91,96%	90,43%	86,99%	88,43%
VNY	75,70%	76,96%	75,76%	77,89%	60,71%	76,15%
OLG	94,99%	94,99%	93,99%	92,99%	86,15%	92,99%
FRT	34,12%	31,98%	31,24%	33,84%	20,69%	30,55%
CER	90,99%	91,48%	87,50%	87,43%	87,95%	85,99%
MAI	98,50%	98,50%	98,50%	98,00%	98,00%	97,00%
CTN	82,99%	83,50%	77,47%	75,50%	75,42%	72,47%
TBC	48,25%	50,62%	50,35%	44,68%	44,07%	32,05%
GRF	90,43%	89,90%	89,36%	84,14%	86,59%	82,70%
SVA	92,96%	92,48%	92,96%	91,96%	83,81%	89,96%
BDS	93,37%	92,83%	95,48%	71,68%	92,96%	83,57%
RCK	97,00%	97,50%	97,50%	97,00%	94,99%	96,50%
MRS	91,73%	90,03%	84,59%	91,43%	74,08%	89,29%
WCR	31,72%	5,80%	5,81%	28,80%	21,55%	23,92%
WBD	99,50%	99,50%	100,00%	99,50%	99,50%	99,50%
CWT	100,00%	100,00%	100,00%	99,50%	99,50%	99,50%
AVERAGE F1	81,89%	80,86%	79,61%	79,24%	75,33%	77,24%
O.A	96,98%	96,94%	96,26%	95,61%	94,89%	94,87%

Πίνακας 10 Ακρίβειες f1, ΟΑ ανά κατηγορία για το tile 35TLF με λιγότερες κατηγορίες

Σε ειδικότερη ανάλυση παρατηρείται ότι για άλλη μια φορά ως προς την πλειοψηφία των κατηγοριών το πρώτο πείραμα δίνει και τα καλύτερα αποτελέσματα. Εξάιρεση είναι η κατηγορία του πυκνού αστικού ιστού όπου η μέγιστη ακρίβεια f1 84,36% εμφανίζεται στο δεύτερο πείραμα (εξαγωγή χαρακτηριστικών με εκατοστημόριο 35%). Στην περίπτωση των καλλιιεργειών πάλι

παρατηρείται ότι οι αμπελώνες ταξινομούνται με 2% μεγαλύτερη ακρίβεια στο τέταρτο πείραμα (κύβος χρονικών χαρακτηριστικών). Η κατηγορία του καπνού παρατηρείται ότι ταξινομείται με την υψηλότερη ακρίβεια σε ποσοστό 50,62% σωστά στο δεύτερο πείραμα. Οι αμμώδεις εκτάσεις, οι οποίες συναντώνται πολύ εκτεταμένα ως κατηγορία στην περιοχή, έχουν ως εκ τούτου πολύ υψηλή ακρίβεια f1 95,48%. Αυτή η ακρίβεια επιτυγχάνεται στο τρίτο πείραμα το οποίο έχει τα λιγότερα κανάλια από όλα τα πειράματα του υπερκύβου των φασματικών χαρακτηριστικών. Αξιοσημείωτη είναι η περίπτωση των θερμοκηπίων, τα οποία αποτελούνται από μόλις 21 εικονοστοιχεία, οπότε η ορθή ταξινόμηση είναι πολύ δύσκολη. Παρακάτω φαίνεται το διάγραμμα των πειραμάτων, συναρτήσει της ακρίβειας η οποία επιτυγχάνεται με τα πειράματα να τοποθετούνται στον άξονα x ανάλογα με τον αριθμό των καναλιών τους. Το διάγραμμα αυτό χρησιμοποιείται για να υπάρχει καλύτερη γενική εποπτεία των ακριβειών:



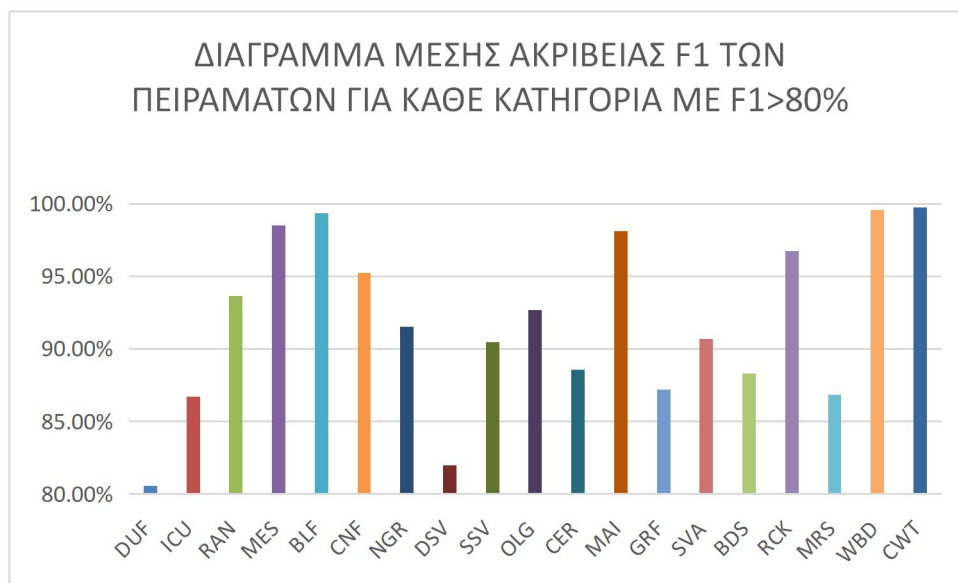
Σχήμα 28 Διάγραμμα ακριβειών ΟΑ και f1 ανά πείραμα (σε παρένθεση ο αριθμός των καναλιών)

- Το πρώτο πείραμα δίνει τις υψηλότερες ακρίβειες f1 ανά κατηγορία, με όλες τις κατηγορίες που έχουν επαρκή αριθμό πολυγώνων εκπαίδευσης να έχουν υψηλές ακρίβειες (>85% στις περισσότερες περιπτώσεις). Οι χαμηλές ακρίβειες για κάποιες περιπτώσεις (GRH, WCR) οφείλονται στον ανεπαρκή αριθμό ή την κακή ποιότητα δεδομένων. Η διαστατικότητα του πειράματος είναι σαφώς μεγαλύτερη από αυτές των υπολοίπων πειραμάτων (91 κανάλια και δείκτες) και η $anf1=81,89\%$.
- Το δεύτερο πείραμα δίνει παρεμφερή αποτελέσματα στην πλειοψηφία των κατηγοριών $anf1=80,86\%$, παρά την μείωση της διαστατικότητας κατά 13 κανάλια.
- Το τρίτο πείραμα δίνει πολύ μειωμένες ακρίβειες σε σχέση με τα προηγούμενα δύο (79,61%) και δεν είναι εξίσου βιώσιμη επιλογή σε σχέση με το αντίστοιχο για το προηγούμενο tile, ειδικά

αν ληφθεί υπόψιν ότι, λόγω της διαδικασίας εξαγωγής χαρακτηριστικών, η μείωση στη διαστατικότητα δεν είναι εξίσου μεγάλη (38 κανάλια) με αυτή του αντίστοιχου πειράματος για το προηγούμενο tile.

- Το τέταρτο πείραμα δίνει ακρίβειες f1 οι οποίες σε κάποιες περιπτώσεις είναι όμοιες, σε άλλες είναι υψηλότερες και σε άλλες σημαντικά κατώτερες σε σχέση με τα προηγούμενα πειράματα (anf1=79,24%).
- Το πέμπτο πείραμα έχει χαμηλές ακρίβειες (μέση ακρίβεια anf1=75,33%) στις περισσότερες κατηγορίες και άρα δεν μπορεί να θεωρηθεί επιτυχημένο.
- Το έκτο πείραμα είναι αρκετά πιο επιτυχημένο από το πέμπτο (anf1=77,24%), με πολύ καλύτερες ακρίβειες στις περισσότερες κατηγορίες και μειωμένη διαστατικότητα (έως και 10% υψηλότερη ακρίβεια από το πείραμα 5 με μόλις 15 κανάλια). Και σε αυτή την περίπτωση μπορούμε να πούμε ότι είναι προτιμότερη η χρήση του σε σχέση με το πέμπτο πείραμα.

Στην συνέχεια θα αναλυθούν λεπτομερέστερα οι κατηγορίες με υψηλή ακρίβεια όπως αυτές φαίνονται στο παρακάτω διάγραμμα με μπάρες για κάθε κατηγορία για την οποία ο μέσος όρος των πειραμάτων f1 υπερβαίνει το 80%:

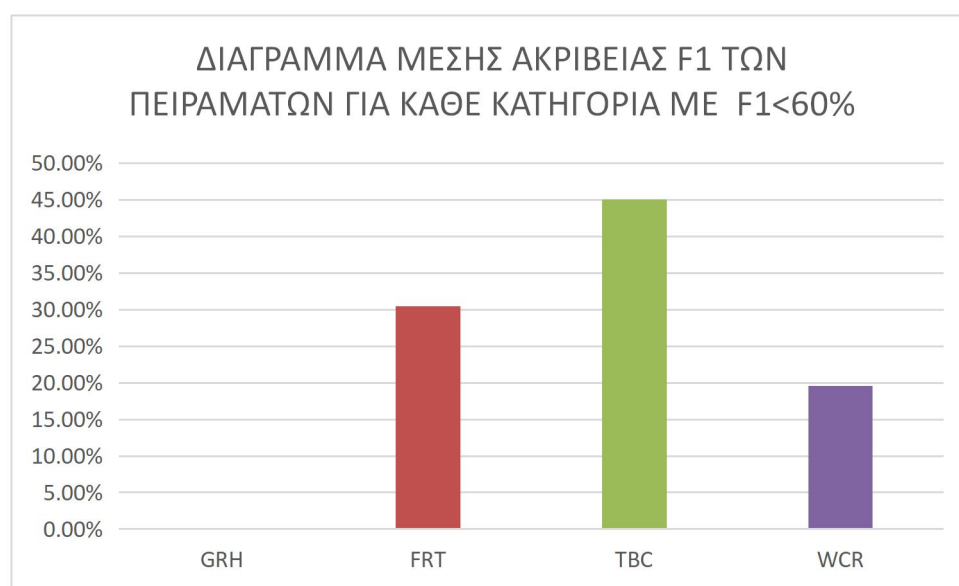


Σχήμα 29 Διάγραμμα μπάρα με anf1>80%

Παρατηρείται ότι για αυτό το tile οι περισσότερες κατηγορίες οι οποίες έχουν ακρίβεια f1 άνω του 80%, κινούνται σε πολύ υψηλές ακρίβειες οι οποίες είναι συνήθως >90%. Χαρακτηριστικά αναφέρονται παραδείγματα όπως αυτά των βραχυδών εκτάσεων (97%), των φυλλοβόλων και των κωνοφόρων δέντρων με 95% και 99% μέσες ακρίβειες αντιστοίχως. Ιδιαίτερα υψηλές με μέσο f1 κοντά στο 80% από όλα τα πειράματα της σειράς είναι οι ακρίβειες για τα στοιχεία του δομημένου ανθρωπογενούς περιβάλλοντος, με την εξαίρεση των φωτοβολταϊκών, η ακρίβεια των οποίων είναι

66%, κάτι το οποίο εξηγείται αν αναλογιστούμε ότι υπάρχουν μόλις 42 εικονοστοιχεία που καλύπτονται από φωτοβολταϊκά σε όλη την έκταση, σε σχέση με τα 74 του 34SEJ. Σε ό,τι αφορά τις καλλιέργειες αυτές, ως προς την πλειονότητα τους έχουν ακρίβειες άνω του 90%, με αυτές του αραβόσιτου (MAI) με 98% και των ελαιοδέντρων (OLG) 93% να ταξινομούνται με πολύ μεγαλύτερη ακρίβεια σε σχέση με τα προηγούμενα tiles. Οι κατηγορίες που έχουν ακρίβεια ταξινόμησης άνω του 80% είναι 19 από τις 27 και τα αποτελέσματα μπορούν να κριθούν ως χειρότερα αυτών του προηγούμενου tile παρά την υψηλότερη γενική ακρίβεια.

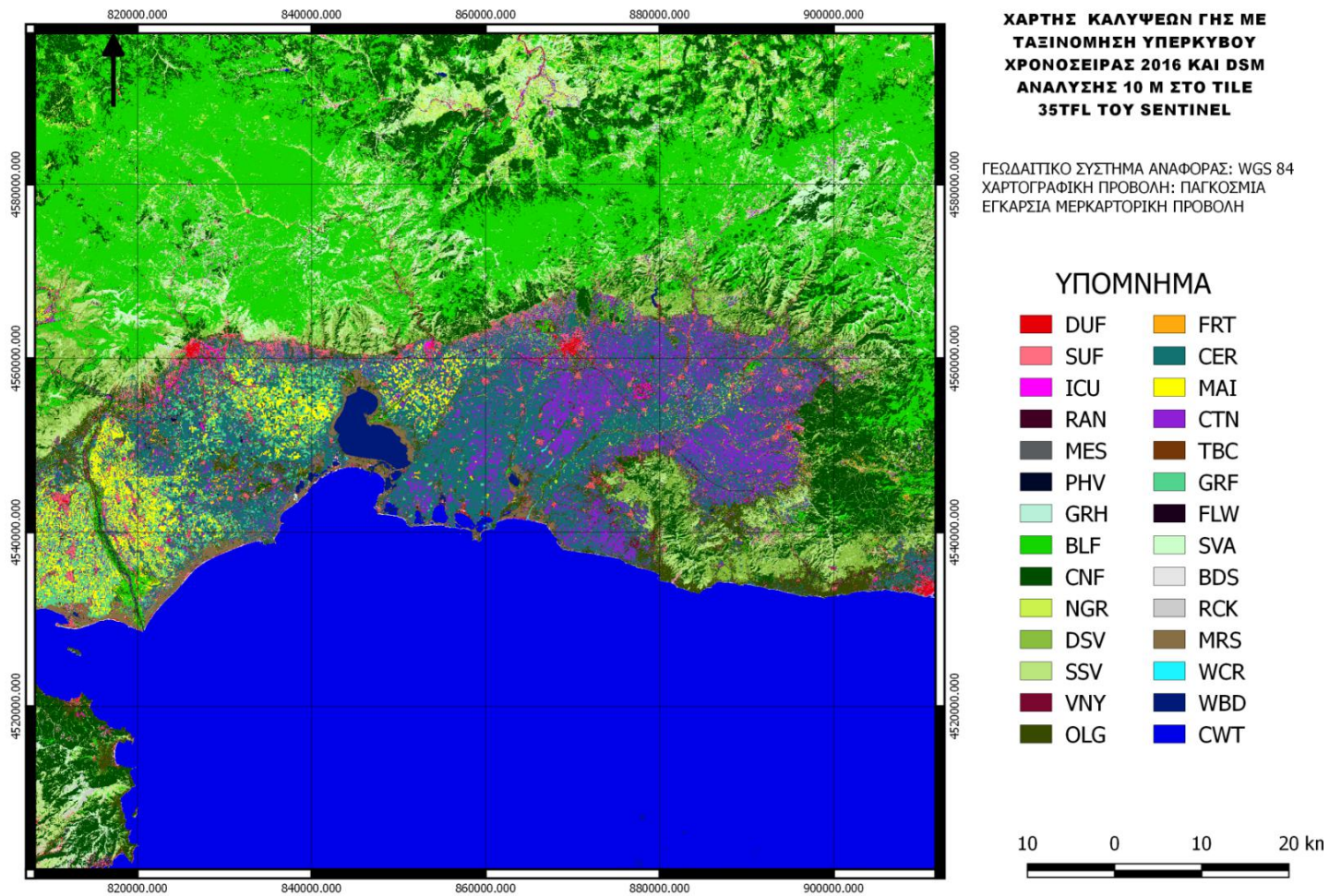
Τελικά γίνεται μια σύντομη ανάλυση των κατηγοριών με τις κακές ακρίβειες, οι οποίες για αυτή την σειρά πειραμάτων ορίζονται με $avf1 < 60\%$.



Σχήμα 30 Διάγραμμα μπάρα με $avf1 < 60\%$

Παρατηρείται μια πολύ ιδιαίτερη κατάσταση, όπου οι κατηγορίες για τις οποίες η ακρίβεια ταξινόμησης είναι κατώτερη από 60% έχουν πολύ χαμηλή ακρίβεια (0-45%). Για όλες τις περιπτώσεις ο βασικός λόγος που συμβαίνει αυτό οφείλεται στο γεγονός ότι ο αριθμός πολυγώνων εκπαίδευσης είναι περιορισμένος. Αναλυτικότερα για την κατηγορία των θερμοκηπίων μόνο εννιά πολύγωνα εκπαίδευσης καλύπτονται από αυτή, με αποτέλεσμα να είναι ανέφικτη η ορθή ταξινόμηση έστω και ενός πολυγώνου αυτής της κατηγορίας, παρά την αισθητά διακριτή φασματική υπογραφή. Για την κατηγορία των δέντρων με καρπούς (FRT) παρατηρούμε ακρίβειες 30%. Κάτι τέτοιο επίσης αναμένεται από το γεγονός ότι σε αυτή την περίπτωση τα πολύγωνα εκπαίδευσης είναι κατά 88% λιγότερα σε σχέση με αυτά του 34TEK. Για την κατηγορία των καπνοκαλλιεργειών οι λόγοι είναι πιο σύνθετοι καθώς τα πολύγωνα εκπαίδευσης είναι παρόμοιου αριθμού με αυτά των άλλων tiles. Ο λόγος για την κακή ακρίβεια βρίσκεται στα στοιχεία αυτής της κατηγορίας που ταξινομούνται ως βαμβάκι, όπως παρατηρούμε από τους πίνακες σύγχυσης. Τα όμοια μοτίβα που ακολουθούν οι παραπάνω καλλιέργειες και η ομοιότητα μεταξύ της κώμης των δύο φυτών οδηγεί σε αυτά τα

σφάλματα, ενώ ο πολύ μεγαλύτερος αριθμός πολυγώνων εκπαίδευσης για το βαμβάκι συνηγορεί περαιτέρω προς αυτές τις ακρίβειες. Εν τέλει τα ποτάμια έχουν πολύ χαμηλότερο αριθμό πολυγώνων εκπαίδευσης από κατηγορίες με παρεμφερείς φασματικές υπογραφές όπως οι λίμνες.



Σχήμα 31 Χάρτης καλύψεων γης tile 34TLF

4.5 Συγκριτική Παρουσίαση των Πειραμάτων για το Tile 34SEG

Στη συνέχεια γίνεται μια συγκριτική παρουσίαση για τις ακρίβειες των έξι πειραμάτων, όπως αυτά εφαρμόστηκαν για το tile 34SEG του δορυφόρου. Η γενική ακρίβεια (ΟΑ) κινείται σε υψηλά επίπεδα για όλα τα πειράματα (>88,81%). Παρατηρείται ότι η γενική ακρίβεια για αυτή την σειρά είναι σε όλες τις περιπτώσεις 4%-6% κατώτερη σε σχέση με την αντίστοιχη για τις ταξινομήσεις των προηγούμενων tiles και το εύρος κινείται μεταξύ των τιμών 88,81%-90,49%.

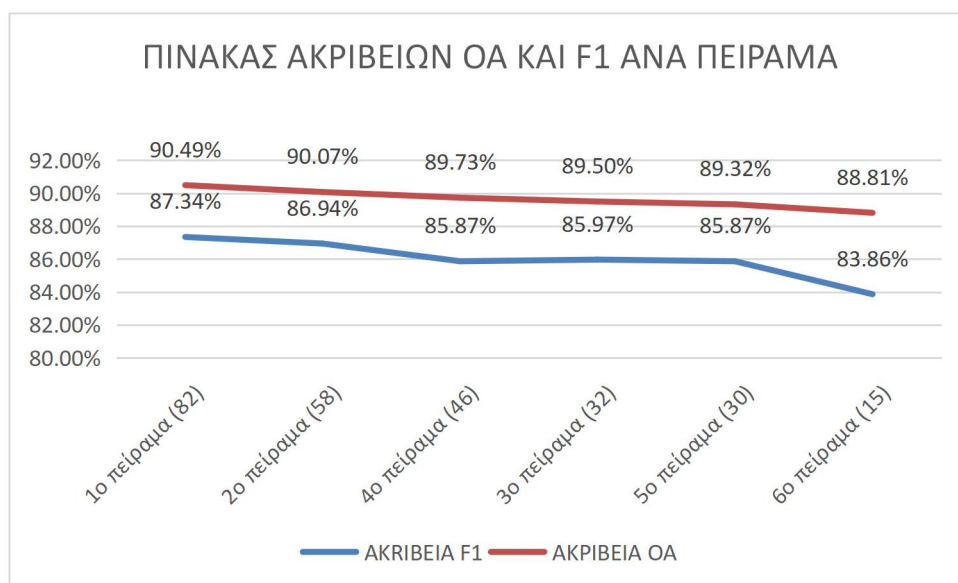
Οι μέσες ακρίβειες f1 αντιθέτως είναι πολύ υψηλές σε σχέση με τις προηγούμενες 2 σειρές πειραμάτων και συγκρίνονται μόνο με τις αντίστοιχες του πρώτου πειράματος, με το εύρος να είναι 83,86%-87,34%

ΑΚΡΙΒΕΙΕΣ F1,ΟΑ ΑΝΑ ΚΑΤΗΓΟΡΙΑ ΓΙΑ ΤΟ TILE 34SEG ΜΕ ΛΙΓΟΤΕΡΕΣ ΚΑΤΗΓΟΡΙΕΣ						
CODE	ΥΠΕΡΚΥΒΟΣ ΗΜΕΡΟΜΗΝΙΩΝ+ DSM (82 κανάλια)	ΥΠΕΡΚΥΒΟΣ+BRIG HTNESS +DSM ΜΑΖΙ ΜΕ ΕΞΑΓΩΓΗ ΧΑ/ΚΩΝ PERC 35 (58 κανάλια)	ΥΠΕΡΚΥΒΟΣ+BRI GHTNESS +DSM ΜΑΖΙ ΜΕ ΕΞΑΓΩΓΗ ΧΑ/ΚΩΝ PERC 65 (32 κανάλια)	ΥΠΕΡΚΥΒΟΣ TEMPORAL+ DSM (46 κανάλια)	ΥΠΕΡΚΥΒΟΣ TEMPORAL+BRIG HTNESS +DSM ΜΑΖΙ ΜΕ ΕΞΑΓΩΓΗ ΧΑ/ΚΩΝ PERC 45 (30 κανάλια)	ΥΠΕΡΚΥΒΟΣ TEMPORAL+BRIG HTNESS+PERC75 +DSM ΜΑΖΙ ΜΕ ΕΞΑΓΩΓΗ ΧΑ/ΚΩΝ PERC 46 (15 κανάλια)
DUF	41,77%	37,14%	34,11%	36,46%	35,29%	39,52%
SUF	88,02%	87,02%	84,84%	88,60%	84,84%	81,01%
ICU	25,83%	28,55%	28,19%	25,76%	26,78%	39,91%
RAN	75,19%	73,66%	71,01%	79,95%	74,90%	58,43%
MES	77,30%	76,54%	75,00%	75,49%	75,78%	77,00%
PHV	100,00%	100,00%	99,50%	99,50%	98,50%	99,50%
GRH	89,82%	89,72%	85,26%	88,96%	81,22%	75,83%
BLF	99,50%	99,50%	99,50%	98,50%	98,50%	98,48%
CNF	98,00%	98,00%	97,50%	97,50%	97,50%	96,99%
NGR	91,48%	88,96%	89,99%	88,90%	88,82%	89,36%
DSV	85,50%	84,50%	83,99%	84,50%	83,50%	83,00%
SSV	85,00%	84,00%	83,50%	84,00%	83,00%	82,00%
VNY	88,47%	88,96%	88,47%	81,80%	81,80%	70,68%
OLG	97,96%	97,96%	96,96%	97,96%	97,48%	96,96%
FRT	79,92%	79,95%	77,84%	74,14%	73,85%	71,15%
MAI	100,00%	100,00%	99,50%	100,00%	99,50%	99,50%
GRF	89,70%	88,87%	81,80%	82,81%	82,13%	77,58%
SVA	93,28%	92,28%	92,83%	89,46%	89,03%	88,45%
BDS	95,00%	95,50%	94,50%	94,43%	93,90%	93,37%
RCK	94,43%	95,48%	96,50%	95,48%	95,99%	97,00%
MRS	100,00%	100,00%	100,00%	99,50%	98,99%	98,99%
WCR	100,00%	100,00%	100,00%	99,50%	99,50%	97,99%
WBD	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
CWT	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
AVERAGE F1	87,34%	86,94%	85,87%	85,97%	85,03%	83,86%
O.A	90,49%	90,07%	89,50%	89,73%	89,32%	88,81%

Πίνακας 12 Ακρίβειες f1, ΟΑ ανά κατηγορία για το tile 34SEG με λιγότερες κατηγορίες

Ειδικότερα παρατηρείται ότι για άλλη μια φορά ως προς την πλειοψηφία των κατηγοριών το πρώτο πείραμα δίνει και τα καλύτερα αποτελέσματα. Εξάιρεση είναι η κατηγορία των εμπορικών χρήσεων γης, όπου η μέγιστη ακρίβεια 39,91% εμφανίζεται στο έκτο πείραμα (εξαγωγή χαρ/κών με perc75 από τον κύβο των temporal). Στην περίπτωση του οδικού δικτύου πάλι, παρατηρείται ότι μέγιστη ακρίβεια (79,95%) επιτυγχάνεται στο τέταρτο πείραμα (κύβος χρονικών χαρακτηριστικών) και ακολουθείται για άλλη μια φορά ένα μοτίβο το οποίο παρατηρείται αρκετά συχνά, δηλαδή για το

οδικό δίκτυο ο κύβος των χρονικών χαρακτηριστικών να προσφέρει τις βέλτιστες ακρίβειες. Εν τέλει η κατηγορία των γυμνών βράχων παρατηρείται ότι έχει πολύ καλά αποτελέσματα για το πείραμα 6 (97%) σαφώς ανώτερα αυτών της ταξινόμησης στον υπερκύβο. Αυτό ακολουθεί το ίδιο μοτίβο με άλλα κατηγορίες με μικρό αριθμό δεδομένων εκπαίδευσης (μόλις 93 πολύγωνα εκπαίδευσης έναντι των 471 του προηγούμενου πειράματος) να ταξινομούνται με σαφώς υψηλότερες ακρίβειες για τον κύβο των χρονικών χαρακτηριστικών σε σχέση με αυτές του υπερκύβου των φασματικών χαρακτηριστικών. Παρακάτω φαίνεται το διάγραμμα των πειραμάτων συναρτήσει της ακρίβειας η οποία επιτυγχάνεται με τα πειράματα να τοποθετούνται στον άξονα x ανάλογα με τον αριθμό των καναλιών τους. Το πείραμα αυτό χρησιμοποιείται για να υπάρχει καλύτερη γενική εποπτεία των ακριβειών:



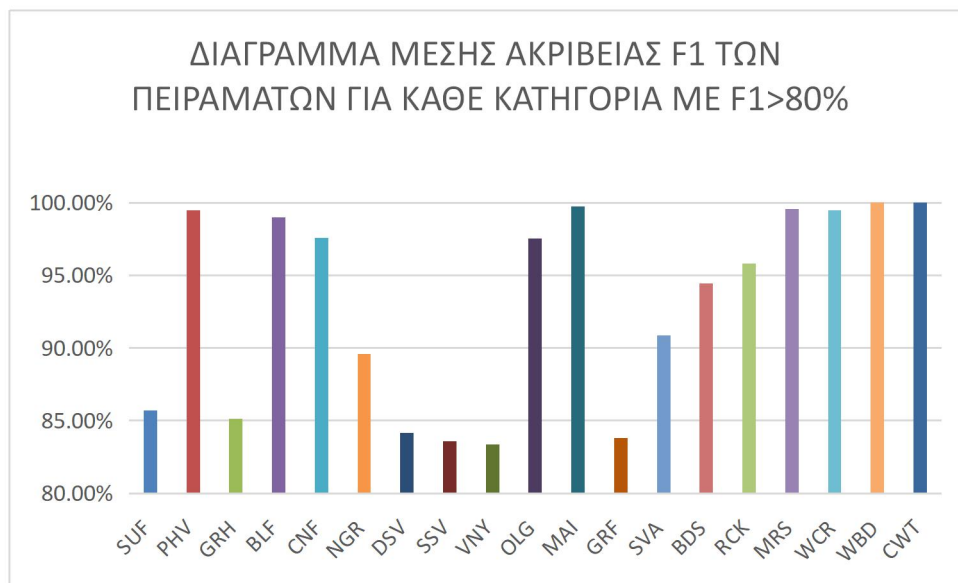
Σχήμα 32 Διάγραμμα ακριβειών ΟΑ και f1 ανά πείραμα (σε παρένθεση ο αριθμός των καναλιών)

- Το πρώτο πείραμα δίνει τις υψηλότερες ακρίβειες f1 ανά κατηγορία (87,34%) με όλες σχεδόν τις κατηγορίες να έχουν πολύ υψηλές ακρίβειες. Οι μικρές ακρίβειες για κάποιες περιπτώσεις οφείλονται ακριβώς στον ανεπαρκή αριθμό σε απόλυτο επίπεδο ή τον μικρό αριθμό δεδομένων εκπαίδευσης σε σχέση με κατηγορίες με τις οποίες υπάρχουν όμοιες φασματικές υπογραφές. Η διαστατικότητα του πειράματος (82 κανάλια) είναι σαφώς μεγαλύτερη από αυτές των υπολοίπων πειραμάτων.
- Το δεύτερο πείραμα δίνει παρεμφερή αποτελέσματα (anf1=86,94%) στην πλειοψηφία των κατηγοριών, όμως αυτά είναι σημαντικά μειωμένα σε περιπτώσεις με έτσι κι αλλιώς χαμηλές ακρίβειες όπως η πυκνή αστική δόμηση. Εδώ η μείωση της διαστατικότητας είναι κατά 13 κανάλια.
- Το τρίτο πείραμα δίνει πολύ μειωμένες ακρίβειες σε σχέση με τα προηγούμενα δύο και δεν

είναι εξίσου βιώσιμη επιλογή σε σχέση με το πείραμα 3 στα προηγούμενα 2 tiles, ειδικά αν ληφθεί υπόψιν ότι το επόμενο πείραμα που γίνεται στον κύβο των χρονικών χαρακτηριστικών έχει σημαντικά μεγαλύτερη ακρίβεια από αυτό. OA=89,50% και AVf1=85,97%

- Το τέταρτο πείραμα δίνει ακρίβειες f1 οι οποίες είναι υψηλότερες για την περίπτωση αυτού του πειράματος (anf1=85,97%) σε σχέση με το προηγούμενο πείραμα με διαστατικότητα 46 κανάλια.
- Το πέμπτο πείραμα έχει αρκετά καλύτερη ακρίβεια (anf1=85,87%) με αριθμό καναλιών 30 σε σχέση με το έκτο και σημαντικά χειρότερη από όλα τα υπόλοιπα.
- Το έκτο πείραμα είναι αυτό με τις μικρότερες ακρίβειες (anf1=83,86%) και 15 κανάλια, πράγμα αναμενόμενο αν αναλογιστούμε την πολύ μειωμένη διαστατικότητα.

Στην συνέχεια θα αναλυθούν λίγο παραπάνω οι κατηγορίες με υψηλή μέση ακρίβεια από όλα τα πειράματα όπως αυτές φαίνονται στο παρακάτω διάγραμμα- μπάρα για κάθε κατηγορία με f1>80%. Ως μέση ακρίβεια ορίζεται ο μέσος όρος των ακριβειών f1 για κάθε κατηγορία:

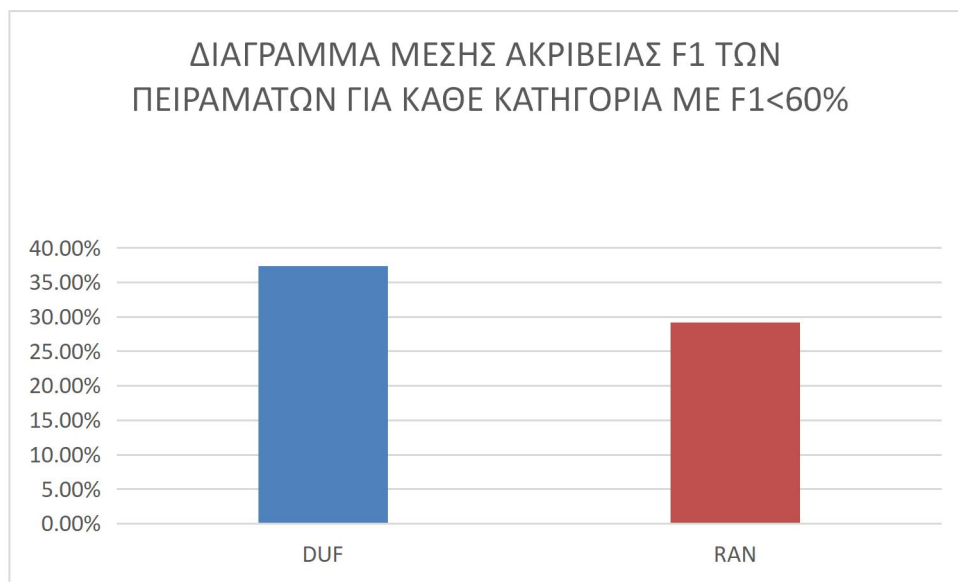


Σχήμα 33 Διάγραμμα μπάρα με anf1>80%

Παρατηρείται ότι και για αυτό το tile οι περισσότερες κατηγορίες οι οποίες έχουν ακρίβεια f1 άνω του 80%, κινούνται σε υψηλές ακρίβειες οι οποίες είναι ως επί το πλείστον >95%. Ο σχολιασμός για κατηγορίες με ακρίβεια άνω του 95% περιττεύει, καθώς μεγάλο πλήθος από τις ακρίβειες είναι σχεδόν 100% ακόμα και σε κατηγορίες όπως τα ελαιόδεντρα που σε προηγούμενα tiles δύσκολα ξεπερνούσαν το 80% ή τα φωτοβολταϊκά που στην περίπτωση αυτή έχουν 99% ακρίβεια. Σε ό,τι αφορά τις καλλιέργειες αυτές, είναι λίγων ειδών στην περιοχή, με τα ελαιόδεντρα και τον αραβόσιτο να κυριαρχούν. Αυτές οι 2 έχουν και ακρίβειες >95%, ενώ οι υπόλοιπες αρκετά χαμηλότερη. Οι κατηγορίες που έχουν ακρίβεια ταξινόμησης άνω του 80% είναι 19 από τις 25 και τα αποτελέσματα

μπορούν να κριθούν για άλλη μια φορά πολύ καλά.

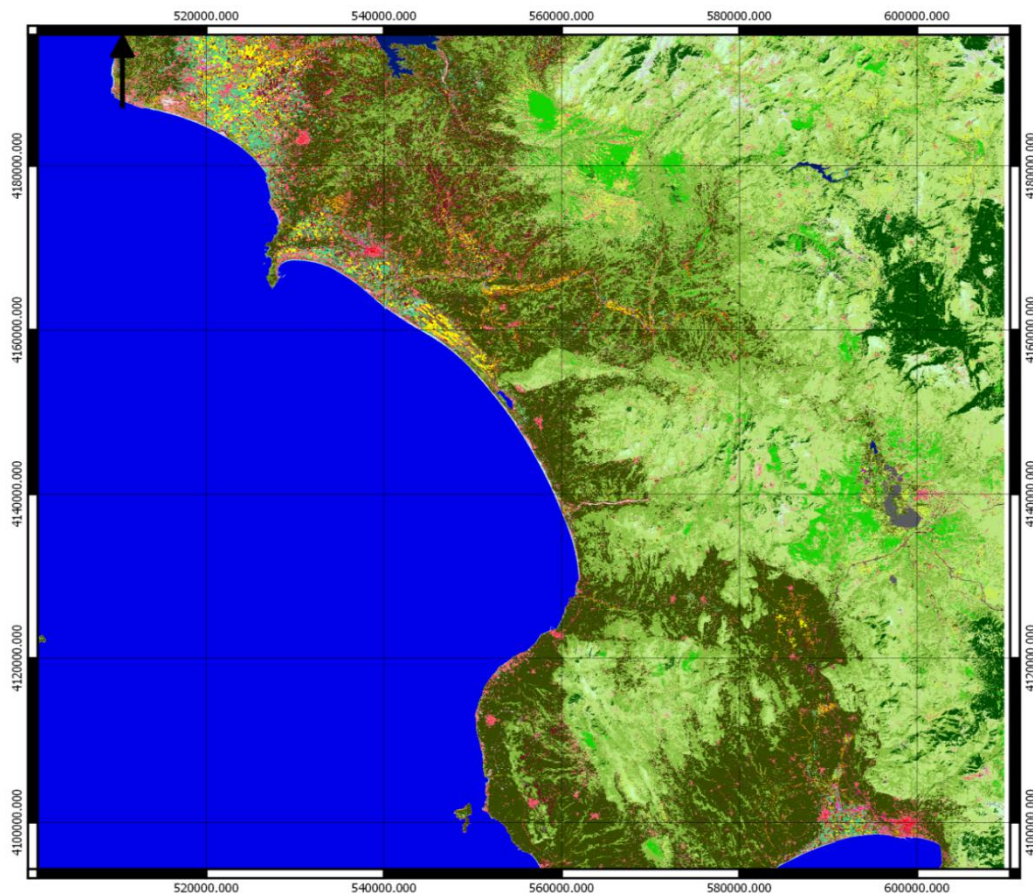
Τελικά γίνεται μια σύντομη ανάλυση των κατηγοριών με τις κακές ακρίβειες, οι οποίες για αυτή την σειρά πειραμάτων ορίζονται εκείνα με $f1 < 60\%$. Οι ακρίβειες φαίνονται πιο εποπτικά στο επόμενο διάγραμμα- μπάρα:



Σχήμα 34 Διάγραμμα μπάρα με $anf1 < 60\%$

Αυτές οι δυο κατηγορίες είναι η πυκνή αστική δόμηση και οι εμπορικές-βιοτεχνικές χρήσεις γης (DUF και ICU)

- Ο λόγος για την μειωμένη ακρίβεια στην κατηγορία της πυκνής αστικής δόμησης σχετίζεται με το γεγονός ότι παρά τον μεγάλο αριθμό πολυγώνων εκπαίδευσης για αυτή την κατηγορία, ο ακόμα μεγαλύτερος αριθμός πολυγώνων αραιής αστικής δόμησης δεν επιτρέπει την σωστή ταξινόμηση.
- Ο λόγος για την μειωμένη ακρίβεια των οδικών δικτύων σχετίζεται με τον μικρό τους αριθμό.



**ΧΑΡΤΗΣ ΚΑΛΥΨΕΩΝ ΓΗΣ ΜΕ
ΤΑΞΙΝΟΜΗΣΗ ΥΠΕΡΚΥΒΟΥ
ΧΡΟΝΟΣΕΙΡΑΣ 2016 ΚΑΙ DSM
ΑΝΑΛΥΣΗΣ 10 Μ ΣΤΟ TILE
34SEG ΤΟΥ SENTINEL**

ΓΕΩΔΑΙΤΙΚΟ ΣΥΣΤΗΜΑ ΑΝΑΦΟΡΑΣ: WGS 84
ΧΑΡΤΟΓΡΑΦΙΚΗ ΠΡΟΒΟΛΗ: ΠΑΓΚΟΣΜΙΑ
ΕΓΚΑΣΙΑ ΜΕΡΚΑΤΟΡΙΚΗ ΠΡΟΒΟΛΗ

ΥΠΟΜΝΗΜΑ

- | | |
|---------------------------------------------------|-----------------------------------------------|
| ■ DUF | ■ FRT |
| ■ SUF | ■ CER |
| ■ ICU | ■ ΜΑΙ |
| ■ RAN | ■ CTN |
| ■ MES | ■ TBC |
| ■ PHV | ■ GRF |
| ■ GRH | ■ FLW |
| ■ BLF | ■ SVA |
| ■ CNF | ■ BDS |
| ■ NGR | ■ RCK |
| ■ DSV | ■ MRS |
| ■ SSV | ■ WCR |
| ■ VNY | ■ WBD |
| ■ OLG | ■ CWT |



Σχήμα 35 Χάρτης Καλύψεων Γης για το tile 34SEG

4.6 Πειράματα Γενίκευσης σε Άγνωστα Δεδομένα και Συνδυαστικά Πειράματα

Στις επόμενες παραγράφους φαίνονται τα αποτελέσματα των πειραμάτων που προέκυψαν από γενίκευση σε άγνωστα δεδομένα και των συνδυαστικών πειραμάτων. Θα εξεταστούν πειράματα που πραγματοποιήθηκαν ανά 2, 3, 4 tiles με εκπαίδευση και έλεγχο σε διαφορετικά tiles, αλλά και πειράματα που πραγματοποιήθηκαν με εκπαίδευση και έλεγχο στο ίδιο tile.

4.6.1 Πειράματα γενίκευσης σε άγνωστα δεδομένα και συνδυαστικά πειράματα ανά 2 tiles

Στη επόμενη σελίδα δίνονται τα αποτελέσματα γενίκευσης σε άγνωστα δεδομένα για τα 3 πρώτα tiles που έχουν τον ίδιο αριθμό εικόνων. Επομένως για αυτά τα tiles χρησιμοποιήθηκαν οι μεθοδολογίες των πειραμάτων 1 και 4 για αναγωγή σε άγνωστα δεδομένα καθώς και τα συνδυαστικά πειράματα με όλους τους δυνατούς συνδυασμούς. Παρατηρήθηκε καταρχάς ότι τόσο οι μέσες ακρίβειες f1, όσο και οι γενικές ακρίβειες (OA) είχαν μεγάλη πτώση σε σχέση με οποιοδήποτε πείραμα έγινε σε ένα tile. Το εύρος των γενικών ακριβειών για τα πειράματα γενίκευσης σε άγνωστα δεδομένα είναι 48,90%-83,99% και παρότι είναι αρκετά εκτεταμένο παρατηρούμε ότι ακόμα και στην καλύτερη περίπτωση από άποψης ακριβειών (η οποία παρουσιάζεται στα πειράματα γενίκευσης με άγνωστα δεδομένα από το 34SEJ στο 34TEK), οι γενικές ακρίβειες παραμένουν 10% χαμηλότερες σε σχέση με το αντίστοιχο πείραμα για το tile 34SEJ. Οι μέσες ακρίβειες f1 κινούνται κατά τα αναμενόμενα σε ακόμα μικρότερο εύρος 26,14%-63,34%, πράγμα το οποίο αναμένεται εξαιτίας των εγγενών δυσκολιών που υπάρχουν στον έλεγχο μοντέλων που εκπαιδεύτηκαν από ένα σετ δεδομένων σε αρκετά “μακρινή” γεωγραφική θέση και με διαφοροποιημένη σε ένα βαθμό γεωμορφολογία.

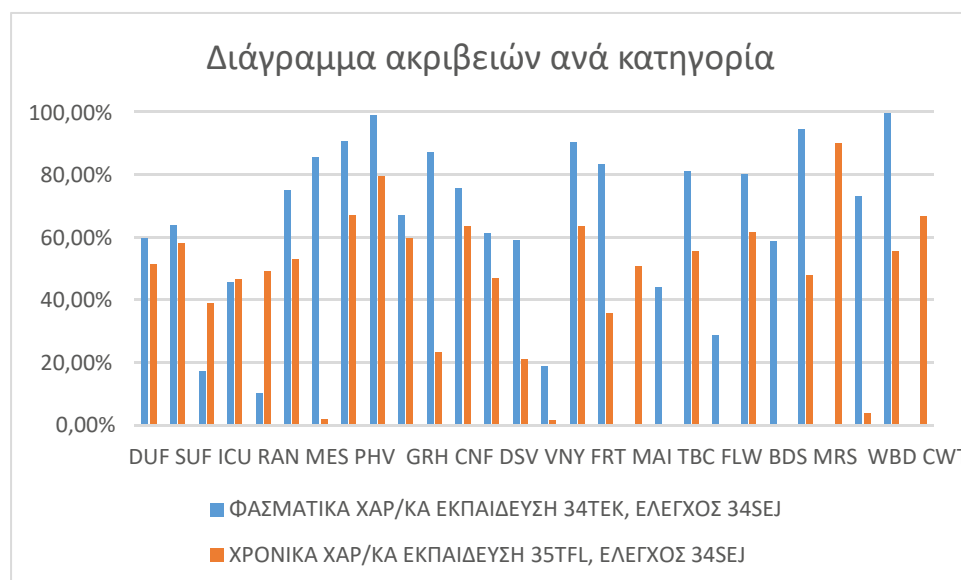
Υποστηρικτικά στην παραπάνω θέση, ότι δηλαδή η γεωγραφική εγγύτητα αποτελεί κομβικό παράγοντα για τις γενικές και επιμέρους ακρίβειες της γενίκευσης σε άγνωστα δεδομένα, λειτουργεί και το γεγονός ότι στην περίπτωση της γενίκευσης μοντέλων ανάμεσα στα γειτονικά tiles 34TEK και 34SEJ έχουμε αποτελέσματα των οποίων το εύρος των γενικών ακριβειών για τα 2 πειράματα είναι μεταξύ των τιμών 75,73% και 83,99%. Εδώ μάλιστα καταδεικνύεται και η σημασία της όμοιας γεωμορφολογίας, αφού και τα 2 tiles έχουν ως δεσπόζον στοιχείο της περιοχής τον ορεινό όγκο της Πίνδου. Συγκριτικά τα αντίστοιχα αποτελέσματα για πειράματα τα οποία γίνονται σε tiles που δεν γειτνιάζουν παρατηρείται ότι δίνουν αποτελέσματα με εύρος γενικών ακριβειών (58,90-74,68%) δηλαδή γενικές ακρίβειες 1,5%-25% κατώτερες σε σχέση με τις αντίστοιχες των γειτονικών tiles. Σε παρεμφερείς διαφορές ανάμεσα στις τιμές τους κινούνται και οι ακρίβειες f1 με την περίπτωση των tiles που γειτνιάζουν να έχουν εύρος τιμών (46,09%-63,44%) και το αντίστοιχο εύρος για tiles που δεν γειτνιάζουν να είναι 28,50%-44,15%.

Τα μόνα tiles για τα οποία τα πειράματα γενίκευσης με άγνωστα δεδομένα φαίνεται να παράγουν καλά αποτελέσματα παρά την γεωγραφική απόσταση, είναι κατά την περίπτωση υλοποίησης τους από το tile 35TLF προς το 34TEK με γενική ακρίβεια 74,68%. Ωστόσο η μέση ακρίβεια f1 (32,05%) αλλά και οι ακρίβειες f1 ανά κατηγορία είναι χαμηλές, ακόμα και σε σύγκριση με αντίστοιχα αποτελέσματα μεταξύ “μακρινών” tiles, πράγμα το οποίο καθιστά το πείραμα σχετικά αποτυχημένο.

Γενικά παρατηρείται ότι η ακρίβειες f1 και OA είναι στην πλειονότητα των περιπτώσεων σημαντικά καλύτερες για την περίπτωση του μοντέλου από υπερκύβο φασματικών χαρακτηριστικών σε σχέση με την περίπτωση του μοντέλου από κύβο των χρονικών χαρακτηριστικών. Μοναδική εξαίρεση αποτελούν οι περιπτώσεις αναγωγής από το TLF στο SEJ και το αντίθετο. Εκεί παρατηρείται ότι οι κύβοι των χρονικών χαρακτηριστικών έχουν 3-4% υψηλότερη ακρίβεια σε σχέση με τους αντίστοιχους των φασματικών χαρακτηριστικών.

Τα αποτελέσματα από τα συνδυαστικά πειράματα κινούνται σε υψηλές ακρίβειες, συγκρίσιμες ή ακόμα και ανώτερες σε πολλές περιπτώσεις από ταξινομήσεις σε ένα tile, πράγμα σημαντικό ώστε να μπορεί να δικαιολογηθεί το αυξημένο υπολογιστικό κόστος που απαιτείται για την εκπαίδευση τους. Πιο συγκεκριμένα παρατηρείται ότι το εύρος των γενικών ακριβειών είναι 94,24%-96,42% και το εύρος των ακριβειών f1 είναι 84,03%-88,16%. Και πάλι οι ακρίβειες των κύβων των φασματικών χαρακτηριστικών κινούνται 1,5% ψηλότερα σε σχέση με τις αντίστοιχες των χρονικών χαρακτηριστικών όπως και πριν.

Στην συνέχεια φαίνεται ένα διάγραμμα ακριβειών f1 ανά κατηγορία για τα πειράματα γενίκευσης του υπερκύβου των φασματικών χαρακτηριστικών 35TLF προς το 34TEK (1), και του υπερκύβου χρονικών χαρακτηριστικών 35TLF προς το 34SEJ(2). Τα πειράματα αυτά επιλέχθηκαν γιατί πετύχαιναν τον καλύτερο συνδυασμό ανάμεσα στα υπόλοιπα πειράματα μηχανικής μάθησης μεταξύ των ακριβειών f1 και γενικών ακριβειών OA.



Σχήμα 36 Διάγραμμα Ακριβειών $anf1$ Ανά Κατηγορία

Οι ακρίβειες των κατηγοριών αστικού ιστού έχουν εύρος για το πείραμα (1) 0-69,96%, και για το πείραμα (2) 0-99,50%.

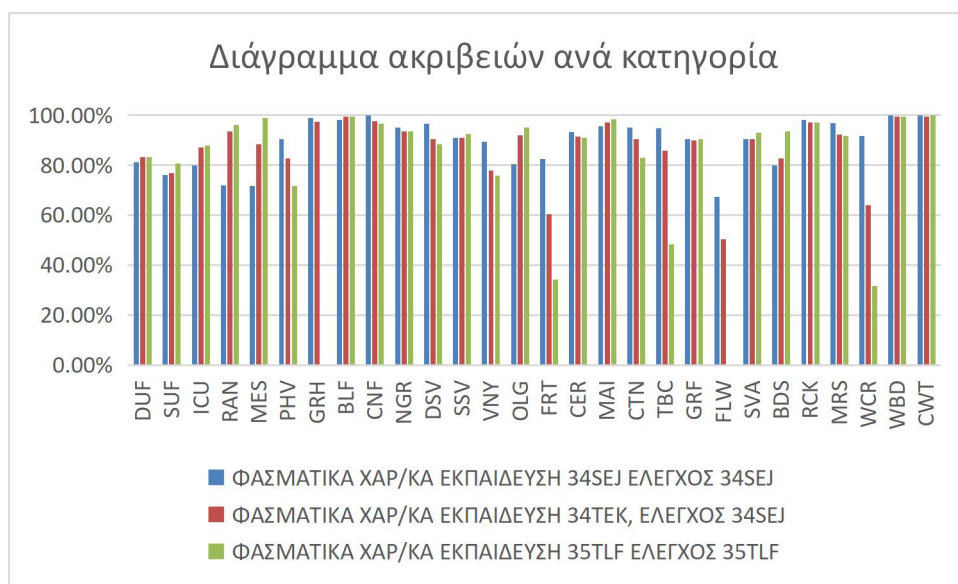
Για την κατηγορία του αστικού ιστού παρατηρούνται ακρίβειες οι οποίες διαφοροποιούνται αρκετά τόσο μεταξύ διαφορετικών κατηγοριών, όσο και στις αντίστοιχες κατηγορίες μεταξύ των πειραμάτων. Οι ακρίβειες πάντως όπως φαίνεται και από το διάγραμμα είναι πιο σταθερές για τον κύβο των χρονικών χαρακτηριστικών σε σχέση με τον αντίστοιχο των φασματικών χαρακτηριστικών. Πιο συγκεκριμένα για το πείραμα (1) το εύρος στον αστικό ιστό είναι 1,98-57,98% και για το πείραμα (2) 10,20%-85,61%. Στην πρώτη περίπτωση οι περισσότερες κατηγορίες έχουν ακρίβεια >40%. Εξαιρέση σε αυτό αποτελεί η κατηγορία των θερμοκηπίων, η οποία έχει 0% ακρίβεια f1, εξαιτίας των πολύ λίγων δεδομένων εκπαίδευσης. Για το πείραμα (2) οι ακρίβειες f1 έχουν σημαντικές διακυμάνσεις και οι τιμές τους είναι από πολύ χαμηλές για την περίπτωση των RAN και των PHV με ακρίβειες f1 17% και 10%, μέχρι υψηλές για την κατηγορία GRH (85%).

Για τις δασικές και σκληροφυλλικές εκτάσεις παρατηρούνται έντονες διακυμάνσεις και στα 2 πειράματα, με τις μεγαλύτερες ακρίβειες να υπάρχουν και πάλι για το δεύτερο πείραμα για το οποίο κυμαίνονται στο εύρος 67,06%-98,99%. Υψηλές ακρίβειες επιτυγχάνονται για τις κατηγορίες BLF και CNF (99% και 90%) ενώ χαμηλές για την κατηγορία NGR. Το αντίστοιχο εύρος για το πρώτο πείραμα είναι 23,16%-79,55%. Οι μικρότερες ακρίβειες παρουσιάζονται στην κατηγορία DSV (23%) και οι μεγαλύτερες την κατηγορία CNF (80%), και παρατηρούμε και πάλι ότι άμεση σχέση με την ακρίβεια f1 έχει το πλήθος των συλλεγμένων δεδομένων και η ποιότητα τους.

Οι κατηγορίες των καλλιεργειών έχουν πολύ χαμηλές ακρίβειες για το πρώτο πείραμα (0-63,50%) και αρκετά μεγαλύτερες ακρίβειες (18,76%-90,28%) για το δεύτερο πείραμα. Για το πρώτο πείραμα η ακρίβεια 0% εμφανίζεται για την κατηγορία του καπνού ενώ η μέγιστη (63,50%) για τα δημητριακά. Η μέγιστη ακρίβεια για το δεύτερο πείραμα είναι 90% για την κατηγορία ΜΑΙ και η ελάχιστη 18% για τα δημητριακά. Οι ακρίβειες για τις καλλιέργειες παρατηρείται ότι είναι χαμηλές <50% για όλα τα πειράματα, ενώ παρατηρείται ότι σε πολλές περιπτώσεις καλλιέργειες με υψηλές ακρίβειες στα πειράματα για ένα tile, έχουν χαμηλές ακρίβειες στην περίπτωση γενίκευσης ανάμεσα σε tiles.

Για τις τελευταίες κατηγορίες παρατηρούνται όμοια φαινόμενα με αυτά που παρατηρήθηκαν προηγουμένως, με τα εύρη να κυμαίνονται από 0% έως 99,5% για το δεύτερο πείραμα και 0%-89,82% για το πρώτο. Οι χειρότερες ακρίβειες παρουσιάζονται στην περίπτωση των αμμωδών εκτάσεων στο πείραμα (1) πράγμα που σχετίζεται άμεσα με την σύγχυση που υπάρχει μεταξύ των ποταμών και των αμμωδών εκτάσεων. Οι καλύτερες ακρίβειες δίνονται στην περίπτωση των βαλτώδων εκτάσεων. Για το πείραμα (2) παρατηρείται ότι τα χειρότερα αποτελέσματα βρίσκονται στην περίπτωση των βάλτων (δεν υπήρχαν καθόλου πολύγωνα εκπαίδευσης) και τα καλύτερα στην περίπτωση των ποταμών (99,5% και 0% αντίστοιχα).

Σε ότι αφορά τα συνδυαστικά πειράματα τα αποτελέσματα είναι εξαιρετικά, οπότε θα συγκρίνουμε τις ακρίβειες για κάθε κατηγορία σε σχέση με τις αντίστοιχες των μοντέλων από ένα tile.



Σχήμα 37 Συγκριτικό Διάγραμμα Ακριβειών ανφ1 Ανά Κατηγορία

Παρατηρείται ότι τα συνδυαστικά πειράματα έχουν παρόμοιες ακρίβειες σε σχέση με τα 2 μοντέλα μηχανικής μάθησης όταν δομούνται ανεξάρτητα το ένα από το άλλο. Οι μόνες κατηγορίες με χαμηλή ακρίβεια για τα συνδυαστικά πειράματα είναι τα WBD (58,22%) και οι SVA (50,40%)

Συμπερασματικά, παρατηρείται ότι οι γενικές ακρίβειες της γενίκευσης μοντέλων σε άγνωστα δεδομένα καθώς και οι f1 δεν είναι καλές και άρα δεν μπορούν να προταθούν. Ωστόσο παρατηρείται ότι οι ακρίβειες από τα συνδυαστικά πειράματα είναι συνολικά πολύ υψηλές. Επιπλέον οι κύβοι των χρονικών χαρακτηριστικών έχουν χειρότερες ακρίβειες σε σχέση με τους αντίστοιχους των φασματικών χαρακτηριστικών. Εν τέλει λοιπόν, εξάγεται το συμπέρασμα:

- Τα πειράματα γενίκευσης μοντέλων σε άγνωστα δεδομένα έχουν εξαιρετικά χαμηλές ακρίβειες f1 ανφ1<50%, σχεδόν σε όλες τις περιπτώσεις οι οποίες δεν μπορούν να θεωρηθούν ως βιώσιμη επιλογή για εκπαίδευση μοντέλων, είτε με τον παραδοσιακό κύβο των φασματικών, είτε με αυτόν των χρονικών χαρακτηριστικών
- Μοναδική εξαίρεση αποτελούν τα μοντέλα τα οποία εκπαιδεύονται από tiles που γειτνιάζουν όπου έχουμε ικανοποιητικά υψηλές ακρίβειες.

ΠΙΝΑΚΑΣ ΑΚΡΙΒΕΙΩΝ F1 ΓΙΑ ΓΕΝΙΚΕΥΣΗ ΜΟΝΤΕΛΟΥ ΣΕ ΑΓΝΩΣΤΑ ΔΕΔΟΜΕΝΑ ΑΝΑ 2 TILES																		
CODE	ΦΑΣΜΑΤΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ Η 35ΤΦΛ, ΕΛΕΓΧΟΣ 34ΣΕΙ	ΦΑΣΜΑΤΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ Η 34ΣΕΙ, ΕΛΕΓΧΟΣ 35ΤΦΛ	ΦΑΣΜΑΤΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ Η 34ΣΕΙ, ΕΛΕΓΧΟΣ 35ΤΦΛ	ΧΡΟΝΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ Η 35ΤΦΛ, ΕΛΕΓΧΟΣ 34ΣΕΙ	ΧΡΟΝΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ Η 34ΣΕΙ, ΕΛΕΓΧΟΣ 35ΤΦΛ	ΧΡΟΝΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ Η 34ΣΕΙ, ΕΛΕΓΧΟΣ (34ΣΕΙ, 35ΤΦΛ)	ΦΑΣΜΑΤΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ Η 35ΤΦΛ, ΕΛΕΓΧΟΣ 34ΤΕΚ	ΦΑΣΜΑΤΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ Η 34ΤΕΚ, ΕΛΕΓΧΟΣ 35ΤΦΛ	ΦΑΣΜΑΤΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ Η 35ΤΦΛ, ΕΛΕΓΧΟΣ (34ΤΕΚ, 35ΤΦΛ)	ΧΡΟΝΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ Η 35ΤΦΛ, ΕΛΕΓΧΟΣ Σ 34ΤΕΚ	ΧΡΟΝΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ Η 34ΤΕΚ, ΕΛΕΓΧΟΣ Σ 35ΤΦΛ	ΧΡΟΝΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ Η 34ΤΕΚ, ΕΛΕΓΧΟΣ (34ΤΕΚ, 35ΤΦΛ)	ΦΑΣΜΑΤΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ Η 34ΤΕΚ, ΕΛΕΓΧΟΣ 34ΣΕΙ	ΦΑΣΜΑΤΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ Η 34ΣΕΙ, ΕΛΕΓΧΟΣ 34ΤΕΚ	ΦΑΣΜΑΤΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ Η 34ΣΕΙ, ΕΛΕΓΧΟΣ (34ΤΕΚ, 34ΣΕΙ)	ΧΡΟΝΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ Η 34ΤΕΚ, ΕΛΕΓΧΟΣ 34ΣΕΙ	ΧΡΟΝΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ Η 34ΣΕΙ, ΕΛΕΓΧΟΣ 34ΤΕΚ	ΧΡΟΝΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ Η 34ΣΕΙ, ΕΛΕΓΧΟΣ (34ΤΕΚ, 34ΣΕΙ)
DUF	42,43%	52,22%	83,35%	51,32%	60,40%	82,89%	21,49%	28,50%	64,81%	25,37%	44,24%	62,69%	59,72%	28,22%	59,45%	58,41%	19,95%	55,93%
SUF	50,26%	17,87%	76,77%	57,98%	25,50%	78,24%	28,49%	27,69%	72,52%	24,62%	39,10%	69,70%	63,86%	25,54%	68,46%	65,19%	20,94%	67,45%
ICU	50,78%	68,69%	86,95%	39,12%	64,84%	85,90%	6,86%	41,82%	73,88%	6,00%	38,65%	68,06%	17,30%	16,76%	64,98%	17,14%	15,35%	53,53%
RAN	43,45%	68,89%	93,43%	46,66%	71,88%	91,48%	43,40%	70,64%	95,99%	7,36%	66,70%	92,96%	45,64%	29,84%	92,96%	38,77%	34,12%	90,99%
MES	49,98%	71,89%	88,45%	49,33%	67,52%	88,02%	39,98%	18,87%	90,60%	48,88%	11,48%	88,28%	10,20%	67,78%	88,28%	11,20%	54,23%	84,84%
PHV	49,50%	20,82%	82,69%	53,02%	22,64%	78,55%	1,92%	13,63%	86,36%	1,96%	10,20%	84,69%	75,19%	48,00%	90,71%	74,43%	54,86%	90,89%
GRH	0,00%	3,79%	97,44%	1,98%	7,24%	96,50%	0,00%	5,47%	95,29%	18,18%	8,86%	91,83%	85,61%	98,50%	99,50%	91,04%	96,91%	98,00%
BLF	59,98%	74,29%	99,50%	66,96%	89,60%	99,50%	66,37%	86,86%	99,50%	95,50%	94,50%	99,50%	90,73%	96,99%	99,50%	88,60%	91,28%	99,50%
CNF	80,29%	35,06%	97,50%	79,55%	33,60%	95,99%	33,22%	58,39%	97,50%	68,56%	43,70%	94,99%	98,99%	96,96%	100,00%	99,50%	90,30%	99,50%
NGR	70,23%	5,81%	93,50%	59,67%	16,40%	92,50%	5,45%	49,39%	91,00%	23,27%	17,06%	87,47%	67,06%	49,20%	92,48%	55,20%	28,33%	87,99%
DSV	13,02%	33,00%	90,43%	23,16%	33,44%	84,71%	10,84%	38,49%	91,37%	23,39%	34,02%	85,71%	86,95%	86,59%	94,91%	89,96%	86,44%	94,91%
SSV	56,95%	54,84%	91,00%	63,40%	57,50%	89,50%	54,16%	57,46%	88,96%	63,50%	58,73%	87,47%	75,47%	67,94%	90,28%	77,11%	65,51%	88,72%
VNY	19,29%	14,27%	77,73%	46,98%	31,11%	78,86%	29,92%	25,49%	76,57%	42,57%	23,84%	75,13%	61,11%	69,99%	84,14%	45,98%	61,34%	79,69%
OLG	8,53%	32,61%	91,99%	20,87%	50,09%	89,99%	0,00%	0,00%	94,50%	0,00%	0,00%	91,99%	58,99%	33,75%	81,25%	57,38%	26,49%	79,12%
FRT	2,40%	2,86%	60,44%	1,50%	2,67%	57,51%	3,87%	8,91%	78,36%	26,83%	7,20%	72,36%	18,76%	48,85%	85,15%	21,24%	46,87%	79,69%
CER	74,75%	42,93%	91,43%	63,50%	46,49%	88,82%	81,25%	48,98%	91,43%	74,79%	42,45%	88,36%	90,28%	81,41%	92,73%	88,02%	50,75%	91,04%
MAI	58,62%	84,97%	96,99%	35,83%	1,85%	95,48%	31,58%	85,50%	97,50%	25,83%	1,92%	97,50%	83,43%	63,50%	95,48%	82,83%	43,40%	92,37%
CTN	66,16%	48,50%	90,50%	50,72%	48,75%	86,47%			83,00%			75,50%			95,00%			92,48%
TBC	0,00%	0,00%	85,71%	0,00%	0,00%	79,61%	18,02%	1,89%	66,04%	9,75%	1,93%	62,45%	44,24%	14,73%	89,90%	59,36%	14,73%	84,99%
GRF	29,71%	28,00%	89,90%	55,55%	40,88%	84,71%	18,16%	8,40%	86,71%	7,14%	22,67%	82,13%	80,98%	47,54%	88,36%	76,96%	1,75%	83,81%
FLW									53,17%			52,53%	28,89%	0,00%	57,31%	10,88%	0,00%	58,37%
SVA	57,78%	46,00%	50,40%	61,46%	40,98%	31,50%	70,49%	34,50%	95,00%	65,51%	36,28%	93,50%	79,98%	77,11%	94,99%	78,19%	70,94%	93,48%
BDS	1,97%	13,87%	90,50%	0,00%	5,14%	88,99%	1,94%	42,44%	89,00%	0,00%	0,00%	76,96%	58,79%	35,20%	82,81%	56,77%	19,40%	77,47%
RCK	50,22%	62,89%	82,83%	47,73%	46,75%	82,57%	35,20%	60,34%	97,50%	26,19%	44,21%	96,00%	94,48%	83,58%	97,48%	93,90%	79,26%	97,48%
MRS	90,17%	19,20%	96,99%	89,82%	3,83%	96,48%	0,00%	8,78%	88,96%	1,71%	1,71%	88,27%	0,00%	0,00%	89,10%	0,00%	0,00%	87,88%
WCR	1,98%	10,74%	92,28%	3,92%	3,89%	92,96%	25,85%	1,98%	89,88%	23,08%	1,98%	87,86%	73,25%	31,06%	93,73%	73,25%	34,22%	92,83%
WBD	32,16%	3,69%	63,98%	55,38%	1,78%	58,39%	83,99%	5,83%	99,50%	91,17%	1,98%	99,50%	99,50%	92,47%	100,00%	99,50%	90,90%	100,00%
CWT	61,11%	58,22%	99,50%	66,67%	62,94%	99,50%			100,00%			99,50%			1			1
AVERAGE F1	41,54%	36,14%	86,75%	44,15%	34,73%	84,28%	28,50%	33,21%	86,96%	32,05%	26,14%	84,03%	63,44%	53,52%	88,18%	61,95%	46,09%	85,82%
OA	57,71%	48,90%	96,27%	58,96%	51,22%	95,00%	51,66%	52,84%	96,42%	74,68%	51,74%	94,83%	83,73%	83,99%	95,54%	82,73%	75,73%	94,24%

Πίνακας 13 Συγκριτικός Πίνακας ακριβειών f1 ανά κατηγορία για κάθε πείραμα γενίκευσης μοντέλων σε άγνωστα δεδομένα και συνδυαστικά πειράματα

4.6.2 Πειράματα γενίκευσης σε άγνωστα δεδομένα ανά 3 tiles (εκπαίδευση στα κοντινά, έλεγχος στα μακρινά)

Στην συνέχεια αναλύονται τα αποτελέσματα της γενίκευσης σε άγνωστα δεδομένα ανά 3 tiles με εκπαίδευση σε 2 tiles που γειτνιάζουν και έλεγχο σε ένα τρίτο tile που δεν γειτνιάζει με κανένα από τα παραπάνω. Ο λόγος που από όλα τα πιθανά πειράματα επιλέχθηκαν αυτά, ήταν ότι ένας από τους σημαντικότερους στόχους αυτής της εργασίας είναι η εξέταση της ακρίβειας μοντέλων τα οποία εκπαιδεύονται σε μια εκτεταμένη περιοχή και μετά ελέγχονται σε μια δεύτερη, λιγότερο εκτεταμένη περιοχή. Αυτό επιτυγχάνεται με τα πειράματα γενίκευσης με άγνωστα δεδομένα με αυτόν τον τρόπο, αφού τα δύο γειτονικά tiles μπορούν να θεωρηθούν ως ενιαία εκτεταμένη περιοχή. Ο υπερκύβος ο οποίος δομείται είναι και για τις τέσσερις περιπτώσεις ο κύβος των χρονικών χαρακτηριστικών τα χαρακτηριστικά του οποίου είναι: min, max, median, mean, stdv for [B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI]+DEM (5X9)+1 και η διαστατικότητα του είναι σε κάθε περίπτωση 46. Τα πειράματα που θα αναλυθούν παρακάτω είναι τα εξής:

- Πείραμα γενίκευσης μοντέλου από υπερκύβο χρονικών χαρακτηριστικών με εκπαίδευση μοντέλου στα tiles 34SEJ και 34TEK και έλεγχο στο 34SEG (πείραμα 1).
- Πείραμα γενίκευσης μοντέλου από υπερκύβο χρονικών χαρακτηριστικών με εκπαίδευση μοντέλου στα tiles 34SEJ και 35TEK και έλεγχο στο 35TLF (πείραμα 2).

Πείραμα Γενίκευσης Μοντέλου με Εκπαίδευση στα Tiles 34SEJ και 34TEK και Έλεγχο στο 34SEG

Στο πείραμα αυτό η γενική ακρίβεια είναι 67,40% και ο μέσος δείκτης f1 είναι 53,53%. Εδώ τα γενικά στατιστικά στοιχεία και οι ακρίβειες είναι για την περίπτωση γενίκευσης σε άγνωστα δεδομένα σημαντικά ακριβέστερα από τα αποτελέσματα για δύο tiles. Η γενική ακρίβεια για αυτό το πείραμα είναι σημαντικά καλύτερη σε σχέση με την προηγούμενη σειρά πειραμάτων, +10% στα περισσότερα συγκρίσιμα πειράματα, δηλαδή τα πειράματα με εκπαίδευση και έλεγχο σε tiles που δεν γειτνιάζουν. Οι ακρίβειες f1 διαφοροποιούνται κατά 5%-15%.

Για τις κατηγορίες του αστικού περιβάλλοντος, δηλαδή τις κατηγορίες DUF, SUF, ICU, RAN, MES, PHV, GRH, οι ακρίβειες παραγωγού κυμαίνονται στο εύρος 2%-91%, δηλαδή κινούνται από την πλήρη αδυναμία του μοντέλου να ταξινομήσει τις κατηγορίες, μέχρι πολύ υψηλές ακρίβειες. Οι κατηγορίες οι οποίες είναι πρακτικά αδύνατο να ταξινομηθούν είναι αυτές των βιοτεχνικών και εμπορικών μονάδων με ακρίβειες f1 2% και 8% αντίστοιχα. Παρατηρείται ότι η πλειοψηφία των πολυγώνων εκπαίδευσης και των δύο κατηγοριών ταξινομούνται ως μεταλλεία. Ο λόγος που συμβαίνει αυτό είναι ότι τα μεταλλεία έχουν πολλαπλάσια πολύγωνα εκπαίδευσης σε σχέση με τα άλλες δύο κατηγορίες. Τα καλύτερα αποτελέσματα παρατηρούνται στην περίπτωση των θερμοκηπίων (GRH 91%). Παρεμφερή προβλήματα παρατηρούνται και με την ακρίβεια χρήστη όπου τα αποτελέσματα κινούνται στο εύρος

93%-100%

Οι κατηγορίες των κωνοφόρων και φυλλοβόλων δέντρων (CNF,BLF) έχουν ακρίβειες χρήστη που είναι 96% και 87% και παραγωγού 96% και 83% αντίστοιχα, οι οποίες παραμένουν αμετάβλητες σε σχέση με προηγούμενα πειράματα.

Για τις κατηγορίες που σχετίζονται με τις χορτολιβαδικές εκτάσεις (NGR, DSV, SSV) και την σκληροφυλλική βλάστηση, παρατηρείται ότι το εύρος για τις ακρίβειες παραγωγού είναι 65%-79% και το εύρος ακριβειών χρήστη είναι 61%-74% και μπορούν να θεωρηθούν ικανοποιητικές για αυτό το πείραμα. Αναμενόμενα η μικρότερη ακρίβεια παραγωγού εμφανίζεται για την κατηγορία της αραιής σκληροφυλλικής βλάστησης 61% και οι λάθος ταξινομήσεις της κατευθύνονται προς την κατηγορία της πυκνής σκληροφυλλικής βλάστησης. Αυτό εξηγείται από το γεγονός ότι η πυκνή, σε σχέση με την αραιή σκληροφυλλική βλάστηση ορίζεται σε σχέση με την αντιληπτική ικανότητα του φωτοερμηνευτή και όχι κάποιο ποσοτικό όριο. Η κατηγορία με τη μικρότερη ακρίβεια παραγωγού είναι το γρασίδι με ακρίβεια παραγωγού 27%. Το εύρος τιμών των ακριβειών παραγωγού είναι 27%-54%.

Στις κατηγορίες των καλλιεργειών (VNY, OLG, FRT, WHT, BRL, OAT, MAI, CTN, TBC,FLW,GRF) παρατηρούνται ακρίβειες παραγωγού οι οποίες βρίσκονται στο εύρος 0%-66% με το κάτω όριο να απαρτίζεται από τα στοιχεία του καπνού, των ζωοτροφών και της αγρανάπαυσης. Για τις περιπτώσεις του καπνού και της αγρανάπαυσης υπήρχε αδυναμία ταξινόμησης και στις προηγούμενες περιπτώσεις της γενίκευσης. Οι λόγοι για αυτή τη δυσκολία ταξινόμησης σχετίζονται με παρεμφερείς παράγοντες με την προηγούμενη περίπτωση. Οι ακρίβειες χρήστη βρίσκονται στο εύρος 0%-95% και κυμαίνονται και αυτές από πολύ κακές έως καλές.

Οι κατηγορίες που είναι σχετικές με το γυμνό έδαφος (BDS,RCK,MRS) έχουν ακρίβειες παραγωγού 7%-70% και ακρίβειες χρήστη 66%-91%. Σε αυτή την περίπτωση παρατηρούμε ότι υπάρχει δυσκολία ταξινόμησης για τις αμμώδεις εκτάσεις όπως και σε προηγούμενη περίπτωση.

Τέλος για τα υδάτινα σώματα (MRS, WCR, WBD, CWT), παρατηρούμε εξαιρετικές ακρίβειες παραγωγού και χρήστη οι οποίες κινούνται στο εύρος 33%-100% και 9%-100% αντιστοίχως. Στις λίμνες υπάρχουν οι μικρότερες ακρίβειες παραγωγού με ένα μεγάλο μέρος των λιμνών να ταξινομούνται ως ποτάμια και οι μικρότερες χρήστη στα ποτάμια, όπου μόλις το 9% είναι όντως ποτάμια με τα υπόλοιπα στοιχεία να είναι λίμνες ή θαλάσσια νερά.

ΠΙΝΑΚΑΣ ΣΥΓΧΥΣΗΣ ΥΠΕΡΚΥΒΟΥ ΦΑΣΜΑΤΟΧΡΟΝΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ (ΕΚΠΑΙΔΕΥΣΗ 34ΤΕΚ, 34ΣΕJ- ΕΛΕΓΧΟΣ 34ΣΕG)																										
	DUF	SUF	ICU	RAN	MES	PHV	GRH	BLF	CNF	NGR	DSV	SSV	VNY	OLG	FRT	MAI	GRF	SVA	BDS	RCK	MRS	WCR	WBD	CWT	sum	PA
DUF	1372	84	388	56	399	0	10	0	0	0	0	1	0	0	0	0	0	0	82	0	0	3	0	0	2395	57%
SUF	1391	8167	13	198	216	2	4	26	0	18	2	486	47	33	44	13	33	279	16	40	4	0	0	0	11032	74%
ICU	404	16	1260	196	363	1	22	0	1	0	0	0	0	0	0	0	1	0	344	25	2	16	0	0	2651	48%
RAN	70	16	20	3402	77	27	0	0	0	6	0	20	0	0	1	1	0	16	11	0	1	140	0	0	3808	89%
MES	20758	272	2184	1190	12849	265	22	0	0	57	0	751	6	1	98	27	68	67	406	19	56	791	1	0	39888	32%
PHV	385	4	15	237	120	1853	0	0	0	0	0	0	0	0	44	52	13	10	0	0	0	0	0	0	2733	68%
GRH	28	332	254	50	425	0	860	0	0	3	0	5	0	0	0	9	45	0	118	0	0	0	0	0	2129	40%
BLF	0	0	0	0	0	0	0	29279	91	25	547	1	1	0	19	2	1	0	0	0	0	0	0	0	29966	98%
CNF	0	0	0	0	0	0	0	6	55701	0	1037	20	0	0	0	0	0	5	0	0	2	0	0	0	56771	98%
NGR	0	0	0	0	0	0	0	85	1	3162	3	125	26	5	16	4	162	158	0	0	0	0	0	0	3747	84%
DSV	0	3	0	0	0	0	0	256	11824	5	104912	2701	2	174	99	12	7	0	0	0	8	0	0	0	120003	87%
SSV	0	8	0	0	0	0	0	439	3093	237	56578	54482	6	1059	435	19	458	583	0	0	2	0	0	0	117399	46%
VNY	0	371	0	0	1	1	0	28	0	20	1	130	833	524	1505	196	2676	0	0	0	0	0	0	0	6286	13%
OLG	0	431	0	5	1	0	0	96	0	39	15693	18728	144	112801	2506	167	4757	0	0	0	0	0	0	0	155368	73%
FRT	0	114	0	0	0	0	0	0	0	0	1914	680	6	3241	2320	31	202	0	0	0	14	2	0	0	8524	27%
MAI	0	0	0	0	0	3	0	0	0	0	0	33	0	40	3170	10000	424	0	0	0	0	0	0	0	13670	73%
GRF	0	0	0	0	0	21	0	0	0	460	0	52	5	102	247	21	1364	15	0	0	0	0	0	0	2287	60%
SVA	3	28	0	0	6	0	0	0	0	773	0	1638	0	0	1	0	227	7111	0	115	0	0	0	0	9902	72%
BDS	8779	50	659	64	4127	0	46	0	0	0	0	0	0	0	0	0	0	4	611	26	1	43	0	0	14410	4%
RCK	0	0	28	0	7	0	0	0	0	0	0	0	0	0	0	0	0	859	0	12633	0	0	0	0	13527	93%
MRS	18	45	3	16	55	3	0	70	0	0	111	215	16	169	868	5925	109	0	0	0	2320	661	1	0	10605	22%
WCR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	105	1423	0	0	1528	93%
WBD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3789	1841	0	5630	33%
CWT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9341	59	10117	19517	52%
sum	33208	9941	4824	5414	18646	2176	964	30285	70711	4805	180798	80068	1092	118149	11373	16479	10547	9107	1588	12858	2515	16209	1902	10117	kappa:	62%
UA	4%	82%	26%	63%	69%	85%	89%	97%	79%	66%	58%	68%	76%	95%	20%	61%	13%	78%	38%	98%	92%	9%	97%	100%	OA:	67%

Πίνακας 14 Πίνακας Σύγκρισης του πρώτου πειράματος (με κόκκινο φαίνονται οι κατηγορίες στις οποίες κατευθύνονται σφάλματα παράλειψης>5%) με OA=67,40% και MO f1=53,53%

Πείραμα Γενίκευσης Μοντέλου με Εκπαίδευση στα Tiles 34SEJ και 34TEK και Έλεγχο στο 35TLF

Στο πείραμα αυτό η γενική ακρίβεια είναι 54,42% και ο μέσος δείκτης f1 είναι 37,40%. Εδώ τα γενικά στατιστικά στοιχεία είναι για την περίπτωση γενίκευσης σε άγνωστα δεδομένα σημαντικά ακριβέστερα από τα αποτελέσματα για δύο tiles. Οι ακρίβειες για αυτό το πείραμα είναι σαφώς χειρότερες σε σχέση με το προηγούμενο πείραμα, τόσο σε επίπεδο γενικών ακριβειών στο οποίο η ταξινόμηση είναι κατά 13% χαμηλότερα σε σχέση με την προηγούμενη, όσο και στο επίπεδο των ακριβειών f1 οι οποίες είναι χαμηλότερες έως και 15% σε σχέση με το προηγούμενο.

Για τις κατηγορίες του αστικού περιβάλλοντος, δηλαδή τις κατηγορίες DUF, SUF, ICU, RAN, MES, PHV, GRH, οι ακρίβειες παραγωγού κυμαίνονται στο εύρος 26%-77% και επομένως κινούνται από χαμηλές έως αρκετά υψηλές. Ενδιαφέρον παρουσιάζει το γεγονός ότι οι ακρίβειες για τις περισσότερες κατηγορίες του ανθρωπογενούς περιβάλλοντος είναι σταθερά σε ποσοστά >70%. Η χειρότερη ακρίβεια παραγωγού παρουσιάζεται για την περίπτωση των φωτοβολταϊκών και βρίσκεται μόλις στο 26%. Όλες οι κατηγορίες του αστικού περιβάλλοντος έχουν χαμηλές ακρίβειες χρήστη με εξαίρεση το οδικό δίκτυο. Τα πιο πολλά ποσοστά λάθος ταξινομημένων εικονοστοιχείων τα έχουν οι κατηγορίες της αραιής αστικής δόμησης με 21% και αυτή των θερμοκηπίων όπου μόλις το 5% των ταξινομημένων εικονοστοιχείων είναι όντως θερμοκήπια. Πολύ μεγάλο μέρος των εικονοστοιχείων που ταξινομούνται ως θερμοκήπια είναι βιοτεχνικές μονάδες και οδικό δίκτυο.

Οι κατηγορίες των κωνοφόρων και φυλλοβόλων δέντρων (CNF,BLF) έχουν ακρίβειες χρήστη που είναι 94% και 26% και παραγωγού 96% και 70% αντίστοιχα. Παρατηρούμε ότι για πρώτη φορά τα φυλλοβόλα δέντρα έχουν εξαιρετικά χαμηλή ακρίβεια παραγωγού και επίσης χαμηλή ακρίβεια χρήστη.

Για τις κατηγορίες που σχετίζονται με τις χορτολιβαδικές εκτάσεις (NGR, DSV, SSV) και την σκληροφυλλική βλάστηση, παρατηρείται ότι το εύρος για τις ακρίβειες παραγωγού είναι 62%-83% και το εύρος ακριβειών χρήστη είναι 22%-58% και μπορούν να θεωρηθούν ικανοποιητικές για αυτό το πείραμα. Η μικρότερη ακρίβεια παραγωγού είναι για την κατηγορία της αραιής σκληροφυλλικής βλάστησης 62% και οι λάθος ταξινομήσεις της κατευθύνονται προς την κατηγορία της πυκνής σκληροφυλλικής βλάστησης. Κατά ένα μέρος για αυτό ευθύνεται η ομοιότητα στις φασματικές υπογραφές των δύο κατηγοριών. Η κατηγορία με τη μικρότερη ακρίβεια χρήστη είναι η πυκνή σκληροφυλλική βλάστηση, με ακρίβεια παραγωγού 22% και το εύρος τιμών 22%-58%.

Στις κατηγορίες των καλλιεργειών (VNY, OLG, FRT, WHT, BRL, OAT, MAI, CTN, TBC,FLW,GRF) παρατηρούνται ακρίβειες παραγωγού οι οποίες βρίσκονται στο εύρος 2%-55% με το κάτω όριο να απαρτίζεται από τα στοιχεία του αραβοσίτου (2%), και του καπνού. Γενικά οι ακρίβειες σε αυτή την περίπτωση είναι οι χειρότερες σε σχέση με όλα τα προηγούμενα και επόμενα πειράματα. Οι κατηγορίες που είναι σχετικές με το γυμνό έδαφος (BDS,RCK,MRS) έχουν ακρίβειες παραγωγού 13%-33% και ακρίβειες χρήστη 34%-95%. Και εδώ παρατηρούνται οι χειρότερες ακρίβειες παραγωγού

και χρήση να είναι στην περίπτωση των αμμωδών εκτάσεων.

Τέλος για τα υδάτινα σώματα (MRS, WCR, WBD, CWT), παρατηρούνται εξαιρετικές ακρίβειες παραγωγού και χρήστη, οι οποίες κινούνται στο εύρος 1%-94% και 1%-74% αντιστοίχως. Οι ακρίβειες στις λίμνες είναι οι μικρότερες ακρίβειες παραγωγού, με μόλις 1% των λιμνών να ταξινομούνται ως λίμνη και οι μικρότερες χρήση στα ποτάμια όπου μόλις το 1% είναι όντως ποτάμι με τα υπόλοιπα στοιχεία να είναι λίμνες ή θαλάσσια νερά.

ΠΙΝΑΚΑΣ ΣΥΓΧΥΣΗΣ ΥΠΕΡΚΥΒΟΥ ΦΑΣΜΑΤΟΧΡΟΝΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ (ΕΚΠΑΙΔΕΥΣΗ 34ΤΕΚ, 34ΣΕΙ- ΕΛΕΓΧΟΣ 35ΤΦΙ)																													
	DUF	SUF	ICU	RAN	MES	PHV	GRH	BLF	CNF	NGR	DSV	SSV	VNY	OLG	FRT	CER	MAI	CTN	TBC	GRF	SVA	BDS	RCK	MRS	WCR	WBD	CWT	sum	PA
DUF	4429	414	724	273	530	1	25	0	0	0	0	1	0	0	0	10	0	5	0	0	1	351	0	0	92	0	0	6856	65%
SUF	823	4134	8	30	139	1	8	0	0	7	0	25	27	5	11	44	11	81	2	11	6	6	1	0	0	0	0	5380	77%
ICU	457	31	4777	74	544	1	145	0	0	0	0	0	0	0	0	2	0	0	0	0	0	185	1	1	36	0	0	6254	76%
RAN	317	89	833	6864	49	159	198	0	0	9	0	8	0	0	0	57	0	26	4	9	21	128	0	134	198	0	0	9103	75%
MES	0	0	2667	0	8521	0	53	0	0	0	0	0	0	0	0	0	0	0	0	0	3	198	5	0	0	0	0	11447	74%
PHV	98	25	0	116	8	322	0	0	0	57	0	10	1	0	0	450	0	69	0	34	30	2	0	0	1	0	0	1223	26%
GRH	0	4	7	20	0	0	30	0	0	0	0	0	0	0	1	0	0	7	1	2	0	0	0	0	0	0	0	72	42%
BLF	0	3	0	0	0	0	0	487837	22119	2637	159	4	11	1	183	0	460	5	4	5478	6	0	0	0	0	0	0	518907	94%
CNF	2	15	0	20	0	12	0	14878	54575	5	137637	2286	2	75	410	4	85	6	1	15	79	0	0	1062	197	0	0	211366	26%
NGR	0	5	0	0	0	0	0	817	4	3040	4	14	11	0	15	33	46	3	7	560	1	0	0	12	0	0	0	4572	66%
DSV	0	13	0	0	0	0	0	3240	881	7	41158	929	80	411	2678	6	87	1	3	37	7	0	0	2	1	0	0	49541	83%
SSV	3	1375	0	279	21	26	8	220	1	129	2146	15602	145	3038	1453	27	138	82	19	308	187	0	0	2	12	0	0	25221	62%
VNY	321	2867	0	0	8	0	0	0	0	38	0	241	1355	11	589	200	48	252	1	203	58	0	0	0	0	0	0	6192	22%
OLG	0	2255	0	364	10	34	11	96	6	49	1909	5386	357	14324	3683	2694	1722	1142	1	1539	6	0	0	5	8	0	0	35601	40%
FRT	0	270	0	12	4	6	25	155	0	38	19	206	328	40	496	129	267	198	0	408	2	0	0	1	0	0	0	2604	19%
CER	450	3312	0	27	2618	46	4	0	0	1792	0	67	134	10	17	42061	226	27625	4608	5977	532	0	0	0	0	0	0	89506	47%
MAI	0	41	0	1	17	11	0	0	0	50	1	2	2	2	159	39065	1012	1721	145	19090	0	0	0	0	0	0	0	61319	2%
CTN	1913	3667	0	64	1408	355	7	0	0	1285	0	28	182	14	61	6351	243	25803	783	1212	204	0	0	0	0	0	0	43580	59%
TBC	2	344	0	1	196	0	0	0	0	3	0	7	5	0	0	735	18	247	45	29	5	0	0	0	0	0	0	1637	3%
GRF	1	70	0	1	39	5	0	0	0	1995	0	3	163	2	302	3524	225	2692	623	11724	1	0	0	2	0	0	0	21372	55%
SVA	139	506	5	337	206	213	9	22	14	403	1	1062	19	0	139	47	19	2	57	190	1728	4	117	48	21	0	0	5308	33%
BDS	538	148	1356	119	779	0	89	3	0	3	0	2	0	1	1	9	0	5	0	1	10	449	3	1	62	0	0	3579	13%
RCK	41	79	49	34	7634	0	16	0	0	2	0	74	0	0	2	8	0	1	23	39	126	7	2357	0	2	0	0	10494	22%
MRS	67	154	0	44	9	34	1	0	0	2	0	1105	353	26	1286	266	123	360	2	11	15	0	2	2570	209	28	0	6667	39%
WCR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	150	2501	0	14	2667	94%
WBD	3	21	0	40	9	91	0	0	0	0	0	3	12	0	26	1	36	29	0	1	2	0	0	266	200821	2428	44702	248491	1%
CWT	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	64505	5796	128993	199298	65%
sum	9604	19842	10426	8721	22749	1317	629	507268	77600	11551	183034	27065	3187	17960	11512	95724	4766	60362	6329	46878	3030	1331	2486	4259	268666	8252	173709	kappa:	48%
UA	46%	21%	46%	79%	37%	24%	5%	96%	70%	26%	22%	58%	43%	80%	4%	44%	21%	43%	1%	25%	57%	34%	95%	60%	1%	29%	74%	OA:	55%

Πίνακας 15 Πίνακας Σύγκρισης του δεύτερου πειράματος (με κόκκινο φαίνονται οι κατηγορίες στις οποίες κατευθύνονται σφάλματα παράλειψης>5%) με OA=54,72% και MO f1=37,4%

4.6.3 Πειράματα γενίκευσης σε άγνωστα δεδομένα ανά 3 tiles (εκπαίδευση στα μακρινά, έλεγχος στα κοντινά)

Στην συνέχεια αναλύονται τα αποτελέσματα της γενίκευσης σε άγνωστα δεδομένα ανά 3 tiles με εκπαίδευση σε 2 tiles που δε γειτνιάζουν και έλεγχο σε ένα τρίτο tile που δε γειτνιάζει με κανένα από τα παραπάνω. Επιλέχθηκαν αυτά τα πειράματα, από όλα τα άλλα πιθανά, προκειμένου να είμαστε σε θέση να συγκρίνουμε στη συνέχεια με τα προηγούμενα πειράματα και να καταλήξουμε έτσι στην βέλτιστη μεθοδολογία για τα πειράματα γενίκευσης με άγνωστα δεδομένα μεταξύ διαφορετικών περιοχών. Ο υπερκύβος ο οποίος δομείται είναι και για τις τέσσερις περιπτώσεις ο κύβος των χρονικών χαρακτηριστικών, τα χαρακτηριστικά του οποίου είναι: min, max, median, mean, stdv for [B,G,R,NIR,RE,SWIR,NDVI,NDWI,NDBI]+DEM (5X9)+1 και η διαστατικότητα του είναι σε κάθε περίπτωση 46. Τα πειράματα που θα αναλυθούν παρακάτω είναι τα εξής:

- Πείραμα γενίκευσης σε άγνωστα δεδομένα από υπερκύβο χρονικών χαρακτηριστικών με εκπαίδευση μοντέλου στα tiles 34SEG και 35TLF και έλεγχο στο 34SEJ (πείραμα 1).
- Πείραμα γενίκευσης σε άγνωστα δεδομένα από υπερκύβο χρονικών χαρακτηριστικών με εκπαίδευση μοντέλου στα tiles 34SEG και 35TLF και έλεγχο στο 34TEK (πείραμα 2).

Πείραμα Γενίκευσης Μοντέλου με Εκπαίδευση στα Tiles 34SEG και 35TLF και Έλεγχο στο 34SEJ

Στο πείραμα αυτό παρατηρείται ότι η γενική ακρίβεια είναι 73,18% και ο μέσος δείκτης $f1$ είναι 53,67%. Παρατηρείται από τα γενικά στατιστικά ότι οι ακρίβειες για αυτή την περίπτωση γενίκευσης σε άγνωστα δεδομένα είναι σημαντικά υψηλότερες από τις αντίστοιχες ανά 2 tiles με τις ακρίβειες να είναι 10%-15% υψηλότερες για αυτή την περίπτωση.

Για τις κατηγορίες που σχετίζονται με το ανθρωπογενές περιβάλλον δηλαδή τις κατηγορίες DUF, SUF, ICU, RAN, MES, PHV, GRH οι ακρίβειες παραγωγού κυμαίνονται από χαμηλές έως υψηλές και κινούνται στο εύρος 33%-88%, δηλαδή κινούνται από πολύ χαμηλές ακρίβειες μέχρι μέσες ακρίβειες. Η κατηγορία με τη μικρότερη ακρίβεια είναι αυτή των θερμοκηπίων (GRH) (33%) και RAN (42%), ενώ αυτή με τη μεγαλύτερη η αραιή αστική δόμηση SUF (99%). Στην περίπτωση των θερμοκηπίων ένα μεγάλο ποσοστό των λάθος ταξινομήσεων κινείται προς τις εμπορικές και βιοτεχνικές μονάδες σε ποσοστό 48% πράγμα που εξηγείται από την ομοιότητα των φασματικών υπογραφών. Για την περίπτωση των οδικών δικτύων RAN οι χαμηλές ακρίβειες εξηγούνται βάσει της μικρής έκτασης που καλύπτουν και του μακρόστενου τους σχήματος που τις καθιστά πολύ δύσκολα ταξινομήσιμες καλύψεις γης. Τα σφάλματα ταξινόμησης της παραπάνω κατηγορίας αναμενόμενα κατευθύνονται προς τις κατηγορίες της πυκνής και αραιής αστικής δόμησης και των μεταλλείων σε ποσοστό 14% για το πρώτο και 17% αντιστοίχως για τα άλλα δύο. Οι μικρότερες ακρίβειες χρήση υπάρχουν στην

περίπτωση των εμπορικών μονάδων (ICU) με 47% και με 40% στην περίπτωση των μεταλλείων.

Οι κατηγορίες των κωνοφόρων και φυλλοβόλων δέντρων (CNF,BLF) έχουν ακρίβειες χρήστη που είναι 96% και 97% και παραγωγού 75% και 91% αντίστοιχα, οι οποίες παραμένουν αμετάβλητες σε σχέση με προηγούμενα πειράματα.

Για τις κατηγορίες που σχετίζονται με τις χορτολιβαδικές εκτάσεις (NGR, DSV, SSV) και την σκληροφυλλική βλάστηση, παρατηρείται ότι το εύρος για τις ακρίβειες παραγωγού είναι 65%-79% και το εύρος ακριβειών χρήστη είναι 71%-91%. Η κατηγορία με τη μικρότερη ακρίβεια παραγωγού και χρήστη είναι το γρασίδι, με ακρίβεια παραγωγού 65% και χρήστη 71%.

Στις κατηγορίες των καλλιεργειών (VNY, OLG, FRT, WHT, BRL, OAT, MAI, CTN, TBC,FLW,GRF) παρατηρούνται ακρίβειες παραγωγού οι οποίες βρίσκονται στο εύρος 0%-86% και διαπιστώνεται ότι το κάτω όριο είναι για τα στοιχεία του καπνού και της αγρανάπαυσης. Για αυτές τις περιπτώσεις υπήρχε αδυναμία ταξινόμησης και στις προηγούμενες περιπτώσεις της γενίκευσης. Σε σχέση με τις ακρίβειες χρήστη αυτές βρίσκονται στο εύρος 0%-76% και για αυτές εξάγονται τα ίδια συμπεράσματα που εξήχθησαν και για την περίπτωση των ακριβειών παραγωγού με τις μικρότερες ακρίβειες να παρουσιάζονται πάλι για καπνό και αγρανάπαυση. Παρατηρείται δυσκολία ταξινόμησης για τις καλλιέργειες, οι οποίες είναι σημαντικά μειωμένες σε αριθμό σε σχέση με αυτές των προηγούμενων κατηγοριών.Οι δυσκολίες που παρατηρήθηκαν στην αναγνώριση των καλλιεργειών της προηγούμενης σειράς πειραμάτων επιμένουν και σε αυτή.

Οι κατηγορίες που είναι σχετικές με το γυμνό έδαφος (BDS,RCK) και τα υδάτινα σώματα (WCR, BDS) έχουν ακρίβειες παραγωγού 81%,19% και 13%,98% αντίστοιχως. Αυτό σχετίζεται με το γεγονός ότι υπάρχει μεγάλη σύγχυση του μοντέλου μηχανικής μάθησης σε σχέση με τις δύο αυτές κατηγορίες λόγω εγγύτητας και επικάλυψης κατά τους χειμερινούς μήνες.

ΠΙΝΑΚΑΣ ΣΥΓΧΥΣΗΣ ΥΠΕΡΚΥΒΟΥ ΦΑΣΜΑΤΟΧΡΟΝΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ (ΕΚΠΑΙΔΕΥΣΗ 34ΣΕΓ, 35ΤΛΦ- ΕΛΕΓΧΟΣ 34ΣΕ)																														
	DUF	SUF	ICU	RAN	MES	PHV	GRH	BLF	CNF	NGR	DSV	SSV	VNY	OLG	FRT	CER	MAI	CTN	TBC	GRF	FLW	SVA	BDS	RCK	MRS	WCR	WBD	CWT	sum	PA
DUF	1642	1286	84	58	166	48	0	0	0	0	0	1	1	1	0	1	0	20	0	0	0	0	45	7	0	0	0	0	3360	49%
SUF	154	3498	1	10	45	5	0	0	4	0	4	157	64	83	1	48	0	17	0	2	0	35	2	0	0	0	0	0	4130	85%
ICU	187	58	1011	12	366	13	8	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	19	21	0	0	0	0	1696	60%
RAN	198	157	71	585	163	27	2	15	18	4	1	65	7	26	0	3	3	6	0	1	0	38	5	11	0	0	0	0	1406	42%
MES	297	121	851	16	7341	0	1	0	0	0	0	0	0	0	0	5	0	6	0	0	0	20	118	284	0	0	0	0	9060	81%
PHV	0	5	0	14	203	1120	0	3	1	2	0	25	0	193	0	15	0	25	2	2	0	146	0	0	1	0	0	0	1757	64%
GRH	2	10	866	3	295	0	590	0	0	0	0	0	1	2	0	0	0	3	0	1	0	0	2	0	0	0	0	0	1775	33%
BLF	0	7	0	0	0	0	0	37049	838	92	228	491	4	1	0	0	0	0	0	0	0	12	0	0	0	0	0	0	38722	96%
CNF	0	0	0	0	0	0	0	2275	97587	0	848	64	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1E+05	97%
NGR	0	0	0	0	62	0	0	3271	33	8427	18	112	36	246	2	83	1	6	0	264	0	372	0	0	1	0	0	0	12934	65%
DSV	0	0	0	2	1	0	0	778	4278	0	21134	1303	0	150	17	0	0	0	0	0	0	2	0	0	0	0	0	0	27665	76%
SSV	0	134	0	0	213	13	0	0	82	40	371	12840	0	1576	0	7	0	0	0	1	0	940	0	0	0	0	0	0	16217	79%
VNY	0	54	0	0	16	0	0	3	11	4	1	20	1563	396	62	17	0	88	0	22	0	0	0	0	4	0	0	0	2261	69%
OLG	0	0	0	0	0	0	0	12	14	17	328	482	42	4215	546	0	20	0	0	0	0	0	0	0	1	0	0	0	5677	74%
FRT	0	21	0	0	1	0	0	207	62	0	35	116	163	848	925	20	81	14	0	138	0	4	0	0	110	0	0	0	2745	34%
CER	5	85	0	6	5	32	0	22	0	47	0	1	93	2534	6	28456	6	1808	1	22	0	0	0	0	3	0	0	0	33132	86%
MAI	0	20	0	0	0	55	0	965	53	47	108	32	1050	2276	849	709	24976	2123	0	413	0	0	0	0	718	0	0	0	34394	73%
CTN	17	310	12	0	36	16	4	346	0	1427	12	45	2146	2646	121	7439	2067	21313	1	7813	0	4	30	0	16547	0	0	0	62352	34%
TBC	0	0	0	0	0	0	0	1	0	10	0	0	426	7	0	108	1	1511	0	2554	0	0	0	0	47	0	0	0	4665	0%
GRF	0	9	0	0	0	86	0	4434	1	1087	7	47	2636	3761	466	4477	5534	850	0	16838	0	0	0	0	207	0	0	0	40440	42%
FLW	0	15	0	0	0	8	0	0	0	3	0	0	75	238	28	539	0	630	0	84	0	1	0	0	232	0	0	0	1853	0%
SVA	0	319	0	37	241	172	8	37	167	744	0	489	87	114	0	305	0	385	0	23	0	7554	1	121	0	0	0	0	10804	70%
BDS	616	59	783	2	6164	0	2	0	0	0	11	4	0	0	0	0	0	2	0	0	0	0	574	51	0	0	2	3	8273	7%
RCK	1	11	76	0	2892	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	11	4982	0	0	0	0	7997	62%
MRS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1611	33	154	0	1798	90%
WCR	5	3	22	30	145	0	0	0	22	0	3	5	0	0	0	0	0	1	0	0	0	0	68	0	24	1109	986	782	3205	35%
WBD	0	0	0	4	0	0	0	0	8	0	2	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	32086	100%
CWT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21638	100%
sum	3124	6182	3777	779	18355	1595	615	49418	1E+05	11951	23111	16299	8394	19313	3023	42233	32689	28808	4	28178	0	9154	875	5477	19507	1142	33204	22431	kappa: 71%	
UA	53%	57%	27%	75%	40%	70%	96%	75%	95%	71%	91%	79%	19%	22%	31%	67%	76%	74%	0%	60%	0%	83%	66%	91%	8%	97%	97%	96%	OA: 73%	

Πίνακας 16 Πίνακας Σύγκρισης του πρώτου πειράματος (με κόκκινο φαίνονται οι κατηγορίες στις οποίες κατευθύνονται σφάλματα παράλειψης>5%) με OA=73,18% και MO f1=53,67%

Πείραμα Γενίκευσης Μοντέλου με Εκπαίδευση στα Tiles 34SEG και 35TLF και Έλεγχο στο 34TEK

Στο πείραμα αυτό διαπιστώνεται ότι η γενική ακρίβεια είναι 80,88% και ο μέσος δείκτης f1 είναι 43,45%. Εδώ τα γενικά στατιστικά στοιχεία και οι ακρίβειες είναι για την περίπτωση γενίκευσης μοντέλου σε άγνωστα δεδομένα αρκετά ακριβέστερα από τα αντίστοιχα ανά 2 tiles. Οι ακρίβειες για αυτό το πείραμα είναι σημαντικά καλύτερες στην περίπτωση της γενικής ακρίβειας σε σχέση με το προηγούμενο πείραμα (+8%) αλλά και σε ό,τι αφορά τις ακρίβειες f1 τόσο τις γενικές όσο και ανά κατηγορία παρατηρείται μείωση η οποία κυμαίνεται στο εύρος 10-15%.

Για τις κατηγορίες του αστικού περιβάλλοντος δηλαδή τις κατηγορίες DUF, SUF, ICU, RAN, MES, PHV, και GRH οι ακρίβειες παραγωγού κυμαίνονται στο εύρος 2%-91%, δηλαδή κινούνται από την πλήρη αδυναμία του μοντέλου να ταξινομήσει τις κατηγορίες μέχρι πολύ υψηλές ακρίβειες. Οι κατηγορίες οι οποίες είναι πρακτικά αδύνατο να ταξινομηθούν είναι αυτές των βιοτεχνικών και εμπορικών μονάδων με ακρίβειες f1 2% και 8% αντίστοιχα. Παρατηρείται ότι η πλειοψηφία των πολυγώνων εκπαίδευσης και των δύο κατηγοριών ταξινομούνται ως μεταλλεία. Ο λόγος που συμβαίνει αυτό είναι ότι τα μεταλλεία έχουν πολλαπλάσια πολύγωνα εκπαίδευσης σε σχέση με τις άλλες δύο κατηγορίες. Τα καλύτερα αποτελέσματα παρατηρούνται στην περίπτωση των θερμοκηπίων (GRH 91%). Παρεμφερή προβλήματα διαπιστώνονται και με την ακρίβεια χρήστη, όπου τα αποτελέσματα κινούνται στο εύρος 93%-100%.

Οι κατηγορίες των κωνοφόρων και φυλλοβόλων δέντρων (CNF,BLF) έχουν ακρίβειες χρήστη που είναι 96% και 87% και παραγωγού 96% και 83% αντίστοιχα, οι οποίες παραμένουν αμετάβλητες σε σχέση με προηγούμενα πειράματα.

Για τις κατηγορίες που σχετίζονται με τις χορτολιβαδικές εκτάσεις (NGR, DSV, SSV) και την σκληροφυλλική βλάστηση, παρατηρείται ότι το εύρος για τις ακρίβειες παραγωγού είναι 65%-79% και το εύρος ακριβειών χρήστη είναι 61%-74% και μπορούν να θεωρηθούν ικανοποιητικές για αυτό το πείραμα. Αναμενόμενα η μικρότερη ακρίβεια παραγωγού είναι για την κατηγορία της αραιής σκληροφυλλικής βλάστησης 61% και οι λάθος ταξινομήσεις της κατευθύνονται προς την κατηγορία της πυκνής σκληροφυλλικής βλάστησης. Η κατηγορία με τη μικρότερη ακρίβεια χρήστη είναι το γρασίδι, με ακρίβεια παραγωγού 27% και το εύρος τιμών 27%-54%.

Στις κατηγορίες των καλλιεργειών (VNY, OLG, FRT, WHT, BRL, OAT, MAI, CTN, TBC,FLW,GRF) παρατηρούνται ακρίβειες παραγωγού οι οποίες βρίσκονται στο εύρος 0%-66% με το κάτω όριο να απαρτίζεται από τα στοιχεία του καπνού, των ζωοτροφών και της αγρανάπαυσης. Για τις περιπτώσεις του καπνού και της αγρανάπαυσης υπήρχε αδυναμία ταξινόμησης και στις προηγούμενες περιπτώσεις της μεταφορικής μάθησης. Οι λόγοι για αυτή την δυσκολία ταξινόμησης σχετίζονται με παρεμφερείς παράγοντες με την προηγούμενη περίπτωση. Οι ακρίβειες χρήστη βρίσκονται στο εύρος 0%-95% και κυμαίνονται και αυτές από πολύ κακές έως καλές.

Οι κατηγορίες που είναι σχετικές με το γυμνό έδαφος (BDS,RCK,MRS) έχουν ακρίβειες παραγωγού 7%-70% και ακρίβειες χρήστη 66%-91%. Σε αυτή την περίπτωση παρατηρούμε ότι υπάρχει δυσκολία ταξινόμησης για τις αμμώδεις εκτάσεις όπως και σε προηγούμενη περίπτωση.

Τέλος για τα υδάτινα σώματα (MRS, WCR, WBD, CWT), παρατηρείται σημαντικό εύρος στις ακρίβειες παραγωγού και χρήστη, οι οποίες κινούνται στο εύρος 35%-100% και 8%-100% αντιστοίχως. Οι ακρίβειες παραγωγού για τα ποτάμια είναι πολύ χαμηλές (δεν υπάρχει ικανοποιητικός συσχετισμός μεταξύ μιας κατηγορίας στο έδαφος με την αντίστοιχη στον χάρτη) ενώ οι ακρίβειες χρήστη των βάλτων είναι επίσης πολύ χαμηλές αφού δεν υπάρχει ακριβής συσχετισμός μεταξύ των αντικειμένων του χάρτη και των πραγματικών κατηγοριών.

ΠΙΝΑΚΑΣ ΣΥΓΧΥΣΗΣ ΥΠΕΡΚΥΒΟΥ ΦΑΣΜΑΤΟΧΡΟΝΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ (ΕΚΠΑΙΔΕΥΣΗ 34ΣΕΓ, 35ΤΛΓ- ΕΛΕΓΧΟΣ 34ΤΕΚ)																												
	DUF	SUF	ICU	RAN	MES	PHV	GRH	BLF	CNF	NGR	DSV	SSV	VNY	OLG	FRT	CER	MAI	TBC	GRF	FLW	SVA	BDS	RCK	MRS	WCR	WBD	sum	PA
DUF	1010	4226	10	7	2425	36	0	2	0	0	0	9	10	0	0	0	0	0	0	0	99	3	97	0	0	0	7934	13%
SUF	16	3330	0	1	145	0	0	1	1	3	8	338	7	26	0	5	0	0	0	0	222	0	18	0	0	0	4121	81%
ICU	16	86	281	6	3149	0	0	0	0	4	0	1	0	0	0	1	0	0	0	0	16	2	63	0	0	0	3625	8%
RAN	3	15	314	78	4753	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	36	0	1	0	0	0	5200	2%
MES	1997	9597	784	0	56722	0	0	0	0	2	0	8	1	0	0	17	0	24	0	0	364	408	4586	0	0	0	74510	76%
PHV	0	0	5	18	1176	706	0	0	9	0	0	0	0	0	0	0	0	0	0	0	143	0	59	0	0	0	2116	33%
GRH	2	4	56	17	52	0	1360	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	6	0	0	0	1500	91%
BLF	0	0	0	0	0	0	0	457029	3197	13924	642	240	1	26	0	0	0	0	0	0	6	0	0	0	0	0	475065	96%
CNF	0	0	0	0	0	0	0	3212	62574	3	5868	136	0	5	0	0	0	0	0	0	1	0	0	0	0	0	71799	87%
NGR	0	7	0	0	8	0	0	1065	15	12532	17	1585	35	1098	0	146	0	0	406	0	1469	0	0	0	0	0	18383	68%
DSV	0	2	0	0	0	0	0	721	4354	0	27501	12575	0	42	0	0	0	0	0	0	3	0	0	0	0	0	45198	61%
SSV	0	1079	0	0	69	0	0	61	286	111	584	30586	21	30	0	0	0	0	0	0	8542	0	135	0	0	0	41504	74%
VNY	13	271	0	1	15	0	0	133	5	176	12	940	1969	404	102	48	0	0	15	0	194	0	5	36	0	0	4339	45%
OLG	0	6	0	0	0	0	0	0	0	0	38	492	0	1036	0	0	0	0	0	0	1	0	6	0	0	0	1579	66%
FRT	0	138	0	0	12	0	0	2132	1046	375	752	3422	542	1210	2465	54	879	0	126	0	425	0	3	6	0	0	13587	18%
CER	0	874	0	0	123	0	0	0	0	7546	0	19	23	397	13	11226	4	85	0	0	1829	0	94	0	0	0	22233	50%
MAI	0	14	0	0	0	0	0	7210	14	459	4	83	14	1	2	26	1878	0	4	0	178	0	70	0	0	0	9957	19%
TBC	0	389	0	0	2	0	0	0	0	630	0	98	46	10	0	620	52	71	107	0	815	0	48	0	0	0	2888	2%
GRF	0	261	0	0	1	0	0	3467	1	10142	9	334	98	153	0	564	15	1	514	0	1184	0	60	0	0	0	16804	3%
FLW	0	164	0	0	143	0	0	1	0	465	0	188	20	60	0	79	2	0	10	0	1235	0	28	0	0	0	2395	0%
SVA	0	678	111	49	1510	16	2	15	200	320	5	5184	64	1	0	18	0	0	2	0	50667	2	5118	106	0	0	64068	79%
BDS	394	60	337	21	2820	2	0	1	0	0	0	1	0	0	0	2	0	0	0	0	24	66	32	0	0	0	3760	2%
RCK	0	14	0	0	1722	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	16	9	7668	0	0	0	9431	81%
MRS	0	0	0	0	24	0	0	1639	3184	16	39	38	0	0	0	0	14	0	0	0	73	0	0	1150	1	4	6182	19%
WCR	0	0	0	4	503	0	0	10	80	0	11	20	0	0	0	0	0	0	0	0	7	0	0	246	1370	7960	10211	13%
WBD	0	0	0	0	9	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	931	51173	52134	98%
sum	3451	21215	1898	202	75383	760	1362	476699	74967	46708	35490	56299	2851	4500	2582	12807	2844	181	1184	0	67549	491	18097	1564	2302	59137	kappa:	74%
UA	29%	16%	15%	39%	75%	93%	100%	96%	83%	27%	77%	54%	69%	23%	95%	88%	66%	39%	43%	0%	75%	13%	42%	74%	60%	87%	OA:	81%

Πίνακας 17 Πίνακας Σύγκυσης του πρώτου πειράματος (με κόκκινο φαίνονται οι κατηγορίες στις οποίες κατευθύνονται σφάλματα παράλειψης>5%) με OA=80,88% και MO f1=43,45%

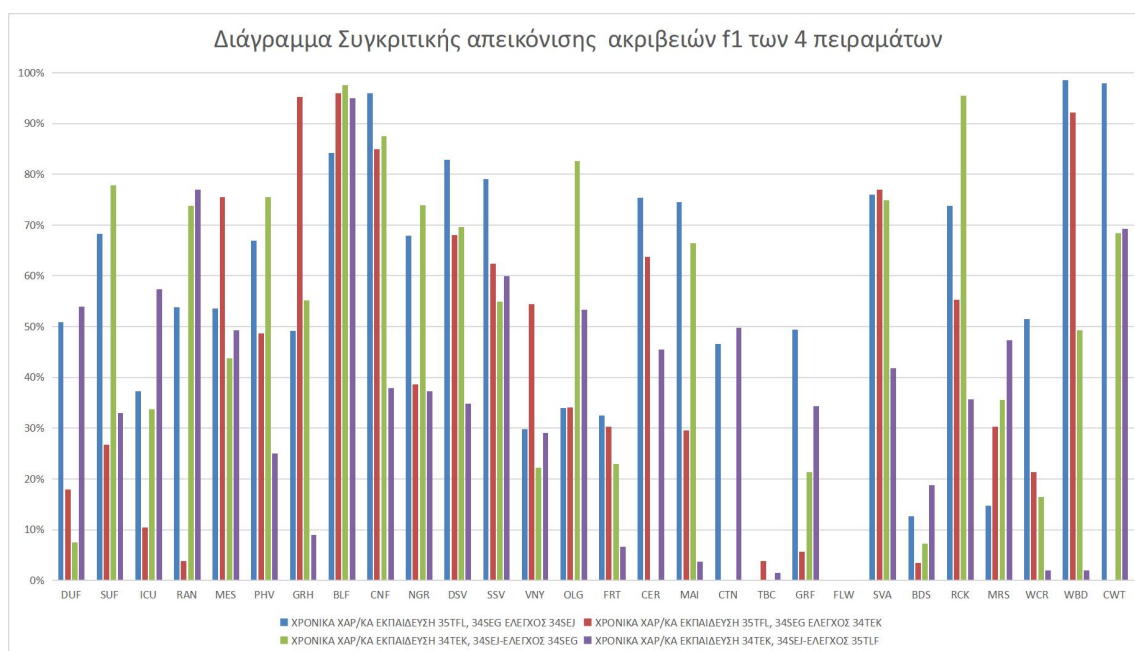
4.6.4 Συγκριτική παρουσίαση πειραμάτων γενίκευσης σε άγνωστα δεδομένα ανά 3 tiles

Στην συνέχεια θα γίνει μια σύγκριση μεταξύ των ακριβειών f1 ανά κατηγορία για κάθε ένα από τα τέσσερα πειράματα που αναλύθηκαν πιο πάνω. Αυτή θα γίνει με τη βοήθεια του παρακάτω πίνακα, στον οποίο υπάρχει χρωματική διαβάθμιση με τις μικρότερες τιμές να είναι με κόκκινο και τις μεγαλύτερες με πράσινο χρώμα. Επιπλέον οι 13 υψηλότερες τιμές ακριβειών f1 σημειώνονται με bold και οι 13 χαμηλότερες με διπλή υπογράμμιση.

ΣΥΓΚΡΗΤΙΚΗ ΠΑΡΟΥΣΙΑΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΑΝΑΓΩΓΗΣ ΜΟΝΤΕΛΩΝ ΣΕ ΑΓΝΩΣΤΑ ΔΕΔΟΜΕΝΑ				
	ΧΡΟΝΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥ ΣΗ 35ΤΛΦ, 34SEG ΕΛΕΓΧΟΣ 34SEJ	ΧΡΟΝΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣ Η 35ΤΛΦ, 34SEG ΕΛΕΓΧΟΣ 34ΤΕΚ	ΧΡΟΝΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣ Η 34ΤΕΚ, 34SEJ- ΕΛΕΓΧΟΣ 34SEG	ΧΡΟΝΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ 34ΤΕΚ, 34SEJ- ΕΛΕΓΧΟΣ 35ΤΛΦ
DUF	50,92%	17,95%	7,48%	53,87%
SUF	68,24%	26,72%	77,79%	33,00%
ICU	37,24%	10,43%	33,73%	57,31%
RAN	53,85%	<u>3,80%</u>	73,78%	76,95%
MES	53,55%	75,50%	43,72%	49,33%
PHV	66,87%	48,71%	75,56%	24,96%
GRH	49,12%	95,29%	55,19%	8,94%
BLF	84,21%	96,00%	97,50%	94,99%
CNF	95,99%	84,95%	87,48%	37,92%
NGR	67,87%	38,65%	73,92%	37,30%
DSV	82,83%	68,07%	69,60%	34,78%
SSV	79,00%	62,44%	54,88%	59,93%
VNY	29,80%	54,47%	22,20%	29,11%
OLG	33,92%	34,11%	82,56%	53,33%
FRT	32,43%	30,27%	22,98%	<u>6,61%</u>
CER	75,32%	63,77%		45,45%
MAI	74,47%	29,51%	66,46%	<u>3,65%</u>
CTN	46,59%			49,75%
TBC	<u>0,00%</u>	<u>3,80%</u>		<u>1,50%</u>
GRF	49,41%	<u>5,61%</u>	21,37%	34,38%
FLW	<u>0,00%</u>	<u>0,00%</u>		
SVA	75,95%	76,95%	74,88%	41,80%
BDS	12,66%	<u>3,47%</u>	<u>7,24%</u>	18,81%
RCK	73,75%	55,32%	95,43%	35,73%
MRS	14,69%	30,24%	35,51%	47,27%
WCR	51,44%	21,37%	16,41%	<u>1,98%</u>
WBD	98,48%	92,17%	49,25%	<u>1,93%</u>
CWT	97,96%		68,42%	69,21%
AVG F1	53,67%	43,45%	52,53%	37,40%
OA	73,17%	80,88%	67,40%	54,72%

Πίνακας 18 Συγκρητικός πίνακας ακριβειών μεταξύ των 4 πειραμάτων της γενίκευσης σε άγνωστα δεδομένα ανά 3 tiles

Οι ακρίβειες f1 του πίνακα ανά κατηγορία φαίνονται πιο εύληπτα και στο παρακάτω διάγραμμα μπάρα:



Σχήμα 38 Διάγραμμα Συγκριτικής Απεικόνισης ανf1 των 4 πειραμάτων

Από τον παραπάνω πίνακα παρατηρείται ότι οι υψηλότερες γενικές ακρίβειες παρουσιάζονται για τη δεύτερη σειρά πειραμάτων, δηλαδή για τα πειράματα τα οποία εκπαιδεύτηκαν με 2 tiles τα οποία δεν γειτνιάζαν και ελέγχθηκαν σε 1 tile με το οποίο κανένα από τα παραπάνω δεν γειτνιάζε. Συγκεκριμένα αυτές οι ακρίβειες είναι 73% για το πείραμα με εκπαίδευση μοντέλου στα tiles 34SEG και 35TLF και έλεγχο στο 34SEJ (πείραμα 1 στο εξής), 81% για το πείραμα με εκπαίδευση μοντέλου στα tiles 34SEG και 35TLF και έλεγχο στο 34TEK (πείραμα 2 στο εξής), 67% για το πείραμα με εκπαίδευση μοντέλου στα tiles 34SEJ και 34TEK και έλεγχο στο 34SEG (πείραμα 3 στο εξής) και 55% για το πείραμα με εκπαίδευση μοντέλου στα tiles 34SEJ και 35TEK και έλεγχο στο 34TFL (πείραμα 4 στο εξής).

Παρατηρείται ότι η μέση ακρίβεια f1 είναι η καλύτερη για το πείραμα 1 (54%) και με εξαίρεση τον καπνό και την αγρανάπαυση (κατηγορίες που από όλα τα μέχρι πειράματα φαίνεται ότι είναι αδύνατο να ταξινομηθούν σωστά στα πλαίσια της μεταφερόμενης μάθησης) και τις βαλτώδεις και αμμώδεις εκτάσεις (οι οποίες ταξινομούνται με υψηλότερη ακρίβεια σε άλλες περιπτώσεις) όλες οι υπόλοιπες κατηγορίες ταξινομούνται με ικανοποιητικές ακρίβειες για αυτή την σειρά πειραμάτων (>30%). Μάλιστα η αύξηση της ακρίβειας σε σχέση με την ταξινόμηση που γίνεται με διαφορετική περιοχή ελέγχου είναι 11%. Αυτό εξηγείται από το γεγονός ότι παρότι τα tiles στα οποία έγινε η εκπαίδευση έχουν αρκετά διαφορετικές χρήσεις γης, και άρα διαφορετικό ground truth, το ένα από το άλλο έχουν πλήθος ποιοτικά συλλεγμένων δεδομένων και στις κατηγορίες που υπάρχει κενό στα δεδομένα από το ένα tile συμπληρώνεται από το άλλο. Επιπλέον το αρκετά διαφορετικό μήκος και πλάτος βοηθά τον ταξινομητή να γίνει αρκετά ευέλικτος ως προς τις ταξινομήσεις του.

Σε ότι αφορά το πείραμα 3, το οποίο επίσης έχει υψηλή γενική ακρίβεια $OA=67\%$ και υψηλή μέση ακρίβεια $f1=53\%$, παρατηρείται ότι οι ακρίβειες αυτές συνοδεύονται από πολύ χαμηλές ακρίβειες $f1=7\%$ για τις βραχώδεις εκτάσεις και την πυκνή αστική δόμηση καθώς και για τα ποτάμια. (16% ακρίβεια.)

Εν τέλει παρατηρείται ότι οι καλύτερες γενικές ακρίβειες OA για αυτή την σειρά πειραμάτων εμφανίζονται για την εκπαίδευση σε απομακρυσμένα tiles του δορυφόρου. Ωστόσο οι ακρίβειες $f1$ δεν φαίνεται να επηρεάζονται από αυτό. Φαίνεται λοιπόν ότι οι μεγαλύτερες ακρίβειες OA επιτυγχάνονται για απομακρυσμένα tiles τα οποία έχουν και την μεγαλύτερη διαφοροποίηση ως προς την γεωμορφολογία και άρα διαφοροποίηση ως προς της κατηγορίες που απαντώνται. Αντίθετα σε ότι αφορά τις ακρίβειες $f1$ παρατηρείται ότι αυτές τείνουν να μην επηρεάζονται από την εγγύτητα ή μη των δεδομένων εκπαίδευσης αλλά από την δυνατότητα κατηγοριών με ελλιπή δεδομένα εκπαίδευσης στην μια ημερομηνία να συμπληρώνονται από δεδομένα εκπαίδευσης της άλλης.

4.6.5 Συγκριτική παρουσίαση πειραμάτων γενίκευσης σε άγνωστα δεδομένα και συνδυαστικά πειράματα ανά 2,3,4 tiles

Στην συνέχεια θα αναλυθούν συγκριτικά αποτελέσματα για την αναγωγή σε άγνωστα δεδομένα

ΣΥΓΚΡΗΤΙΚΗ ΠΑΡΟΥΣΙΑΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΑΝΑΓΩΓΗΣ ΜΟΝΤΕΛΩΝ ΣΕ ΑΓΝΩΣΤΑ ΔΕΔΟΜΕΝΑ ΜΕ ΕΚΠΑΙΔΕΥΣΗ ΣΕ 1, 2 ΚΑΙ 3 TILES			
	ΧΡΟΝΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ 35TLF, 34SEG-ΕΛΕΓΧΟΣ 34SEJ	ΧΡΟΝΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ 35TLF, 34SEG-ΕΛΕΓΧΟΣ 34SEJ	ΧΡΟΝΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ 35TLF, 34SEG, 34TEK-ΕΛΕΓΧΟΣ 34SEJ
DUF	51,32%	50,92%	67,63%
SUF	57,98%	68,24%	70,94%
ICU	39,12%	37,24%	44,50%
RAN	46,66%	53,85%	49,98%
MES	49,33%	53,55%	67,05%
PHV	53,02%	66,87%	82,93%
GRH	1,98%	49,12%	72,87%
BLF	66,96%	84,21%	89,88%
CNF	79,55%	95,99%	98,50%
NGR	59,67%	67,87%	57,87%
DSV	23,16%	82,83%	91,90%
SSV	63,40%	79,00%	83,50%
VNY	46,98%	29,80%	42,48%
OLG	20,87%	33,92%	45,26%
FRT	1,50%	32,43%	33,75%
CER	63,50%	75,32%	82,02%
MAI	35,83%	74,47%	70,32%
CTN	50,72%	46,59%	14,38%
TBC	0,00%	0,00%	35,49%
GRF	55,55%	49,41%	61,74%
FLW		0,00%	3,68%
SVA	61,46%	75,95%	80,21%
BDS	0,00%	12,66%	57,92%
RCK	47,73%	73,75%	94,48%
MRS	89,82%	14,69%	16,40%
WCR	3,92%	51,44%	77,12%
WBD	55,38%	98,48%	98,99%
CWT	66,67%	97,96%	99,50%
AVG F1	44,15%	53,67%	64,93%
OA	58,96%	73,18%	76,19%

Πίνακας 19 Πίνακας Συγκριτικής Παρουσίασης Γενίκευσης Μοντέλων σε Άγνωστα Δεδομένα

ΣΥΓΚΡΗΤΙΚΗ ΠΑΡΟΥΣΙΑΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΣΥΝΔΥΑΣΤΙΚΩΝ ΠΕΙΡΑΜΑΤΩΝ ΜΕ ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΕΛΕΓΧΟ ΣΕ 1, 2, 3 ΚΑΙ 4 TILES			
	ΧΡΟΝΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ 34SEJ-34TEK, ΕΛΕΓΧΟΣ 34SEJ-34TEK	ΧΡΟΝΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ 35TLF, 34SEJ, 34TEK-ΕΛΕΓΧΟΣ 35TLF, 34SEJ, 34TEK	ΧΡΟΝΙΚΑ ΧΑΡ/ΚΑ ΕΚΠΑΙΔΕΥΣΗ 35TLF, 34SEJ, 34TEK, 34SEG-ΕΛΕΓΧΟΣ 35TLF, 34SEJ, 34TEK, 34SEG
DUF	59,45%	64,69%	62,78%
SUF	68,46%	70,31%	75,76%
ICU	64,98%	68,87%	54,47%
RAN	92,96%	91,50%	88,36%
MES	88,28%	86,26%	85,46%
PHV	90,71%	85,61%	90,30%
GRH	99,50%	95,50%	90,36%
BLF	99,50%	99,50%	99,50%
CNF	100,00%	96,50%	95,50%
NGR	92,48%	86,47%	86,99%
DSV	94,91%	87,72%	84,95%
SSV	90,28%	87,90%	84,47%
VNY	84,14%	77,73%	75,16%
OLG	81,25%	89,50%	95,48%
FRT	85,15%	71,03%	69,80%
CER	92,73%	89,27%	89,27%
MAI	95,48%	95,48%	95,48%
CTN	95,00%	85,99%	86,50%
TBC	89,90%	74,41%	79,13%
GRF	88,36%	83,26%	82,70%
FLW	57,31%	46,98%	46,55%
SVA	94,99%	92,99%	91,43%
BDS	82,81%	79,39%	84,42%
RCK	97,48%	96,48%	96,99%
MRS	89,10%	89,82%	94,96%
WCR	93,73%	81,42%	84,95%
WBD	100,00%	99,50%	99,50%
CWT	100,00%	99,50%	99,50%
AVG F1	88,18%	84,77%	84,67%
OA	96,21%	94,75%	94,36%

Πίνακας 20 Πίνακας Συγκριτικής Παρουσίασης Συνδυαστικών Πειραμάτων

Εδώ θα αναλυθούν συγκριτικά αποτελέσματα όπως αυτά προέκυψαν από τη συγκριτική παρουσίαση των πειραμάτων, καταρχάς για την περίπτωση των πειραμάτων γενίκευσης σε άγνωστα δεδομένα. Το πείραμα από την ανάλυση των δύο προηγούμενων κεφαλαίων που εν τέλει επιλέχθηκε ήταν αυτό της μεταφοράς μάθησης από 35TFL και το 34SEG προς το 34TEK. Το συγκεκριμένο πείραμα επιλέχθηκε καθώς όπως φάνηκε από την προηγούμενη σύγκριση είναι ένα πείραμα το οποίο έδωσε πολύ ικανοποιητικές ακρίβειες στα προηγούμενα πειράματα γενίκευσης σε άγνωστα δεδομένα. Στα πλαίσια των πειραμάτων γενίκευσης σε άγνωστα δεδομένα και των μειωμένων ακριβειών αυτής οποιαδήποτε ακρίβεια >70% θεωρείται ικανοποιητική.

Η σύγκριση θα γίνει μεταξύ των ακριβειών f1 και OA για τις εξής περιπτώσεις:

- Εκπαίδευση στο tile 35TLF και έλεγχος στο tile 34SEJ.
- Εκπαίδευση στα tiles 35TLF και 34SEG και έλεγχος στο tile 34SEJ.
- Εκπαίδευση στα tiles 35TLF , 34TEK και 34SEG και έλεγχος στο tile 34SEJ.

Η σύγκριση φαίνεται στον παραπάνω πίνακα όπου επισημαίνεται η μεταβολή ανά κατηγορία με τη βοήθεια της χρωματικής παλέτας - οι 13 υψηλότερες τιμές f1 με bold γράμματα και οι 13 χαμηλότερες τιμές f1 με την βοήθεια της διπλής υπογράμμισης.

Παρατηρείται σε πρώτη φάση, μέσα από την εποπτική ικανότητα που δίνεται με την βοήθεια της διαφοροποίησης στην χρωματική παλέτα, ότι οι ακρίβειες f1 είναι στην μεγάλη τους πλειοψηφία καλύτερες για την περίπτωση της εκπαίδευσης σε 3 tiles. Πιο συγκεκριμένα παρατηρείται ότι από τις 28 συνολικά κατηγορίες που υπάρχουν στα δύο από τα τρία πειράματα και τις 27 στο τρίτο οι 22 έχουν τις υψηλότερες τους ακρίβειες f1 για το τρίτο πείραμα. Αντιστοίχως η πλειοψηφία των χαμηλότερων ακριβειών ανά κατηγορία παρατηρείται στην περίπτωση του πρώτου πειράματος με τις 21 από τις 27 για αυτή την περίπτωση κατηγορίες να παρουσιάζουν τις χαμηλότερες τους ακρίβειες σε αυτό το πείραμα.

Πιο συγκεκριμένα η σημαντικότερη μεταβολή στην ακρίβεια είναι για την κατηγορία των θερμοκηπίων των οποίων η ακρίβεια f1 στην πρώτη περίπτωση είναι μόλις 2% στη δεύτερη περίπτωση είναι 49% και στην τρίτη περίπτωση φτάνει στο 73% δηλαδή γίνεται ικανοποιητική. Οι περισσότερες κατηγορίες έχουν όμοια τάση χωρίς φυσικά να προσεγγίζουν παρόμοιες διαφοροποιήσεις ως προς τις ακρίβειες. Εν τέλει το μοναδικό αντιπαράδειγμα που υπάρχει σε αυτή την τάση είναι αυτό του βαμβακιού για το οποίο η τάση αντιστρέφεται με την ακρίβεια ανάμεσα στα τρία πειράματα να είναι 51% για το πρώτο, 47% για το δεύτερο και 14% για το τρίτο.

Στην συνέχεια αναλύονται τα πειράματα που φαίνονται στον πίνακα 20. Σε αυτή την περίπτωση πραγματοποιήθηκαν τρία συνδυαστικά πειράματα με πρόσθεση ενός ακόμα tile στα ήδη υπάρχοντα για κάθε νέο πείραμα. Επιλέχθηκε ως πείραμα βάσης το συνδυαστικό με εκπαίδευση και έλεγχο στα tiles 34SEJ και 34TEK με το οποίο θα συγκρίνουμε όλα τα υπόλοιπα. Το πείραμα αυτό επιλέχθηκε γιατί έδινε πολύ υψηλές ακρίβειες τόσο f1=88% όσο και OA=96%. Στη συνέχεια

παρουσιάζονται τα αποτελέσματα για τα συνδυαστικά πειράματα όπως φαίνονται στον παραπάνω πίνακα. Τα τρία πειράματα για τα οποία θα γίνει η σύγκριση είναι τα εξής:

- Εκπαίδευση και έλεγχος στα tile 34SEJ και 34TEK.
- Εκπαίδευση και έλεγχος στα tiles 35TLF, 34TEK και 34SEJ.
- Εκπαίδευση και έλεγχος στα tiles 35TLF, 34TEK, 34SEG και 34SEJ.

Όμοια με πριν η χρωματική παλέτα επιτρέπει την καλύτερη διάκριση ανάμεσα στα πειράματα ανάλογα με τις ακρίβειες τους, ενώ οι 13 κατηγορίες με τις υψηλότερες ακρίβειες είναι με bold γράμματα και οι 13 με τις χαμηλότερες είναι με την διπλή υπογράμμιση.

Σε αυτή την περίπτωση παρατηρείται ότι η τάση αντιστρέφεται σε σχέση με το προηγούμενο πείραμα και ότι εδώ εμφανίζονται στο πρώτο πείραμα οι υψηλότερες ακρίβειες για τις 22 από τις 28 κατηγορίες. Τα άλλα δύο πειράματα, παρά την αύξηση των δεδομένων εκπαίδευσης και ελέγχου, έχουν σημαντικά μικρότερες ακρίβειες για τις περισσότερες κατηγορίες. Εδώ έχει σημασία να παρατηρηθούν και οι μέγιστες και ελάχιστες διαφορές ανάμεσα στις ακρίβειες των τριών πειραμάτων. Η μέγιστη διαφορά του δεύτερου πειράματος από το πρώτο στις f1 ανά θεματική κατηγορία είναι 10% και του τρίτου από το πρώτο 11%. Η μόνη περίπτωση στην οποία η ακρίβεια του πρώτου πειράματος είναι σημαντικά κατώτερη από αυτή του δεύτερου είναι αυτή της πυκνής αστικής δόμησης, για την οποία το δεύτερο πείραμα έχει 6% υψηλότερη ακρίβεια από το πρώτο. Με τη βοήθεια της επισήμανσης των υψηλότερων και χαμηλότερων ακριβειών f1, παρατηρείται ότι το εύρος ακριβειών εντός κάθε πειράματος παραμένει υψηλό αλλά η ψαλίδα έχει κλείσει σημαντικά σε σχέση με τα πειράματα γενίκευσης σε άγνωστα δεδομένα.

Η ανάλυση των γενικών ακριβειών και της μέσης f1 επιβεβαιώνει ότι υπάρχει μικρή μείωση των ακριβειών με αύξηση των δεδομένων. Κάτι τέτοιο αναμενόταν καθώς η γειννίαση των tile στο πρώτο πείραμα είναι λογικό να επιτρέπει εξαιρετικά υψηλές ακρίβειες. Η πρόσθεση δεδομένων εκπαίδευσης από περιοχές με διαφορετικό ανάγλυφο αναμενόμενα δίνει και μικρότερες ακρίβειες.

Εν τέλει παρατηρείται ότι ενώ για την περίπτωση των πειραμάτων γενίκευσης σε άγνωστα δεδομένα οι ακρίβειες αυξάνονται πολύ σημαντικά με την προσθήκη επιπλέον δεδομένων εκπαίδευσης από άλλα tiles, με την αύξηση αυτή να παρατηρείται σε όλες τις κατηγορίες, στην περίπτωση των συνδυαστικών πειραμάτων όπου οι ακρίβειες είναι ήδη εξαιρετικά υψηλές η σημαντική αύξηση της ακρίβειας λόγω προσθήκης και άλλων δεδομένων εκπαίδευσης δεν συμβαίνει. Μάλιστα στην περίπτωση που τα tiles συνορεύουν, προσθήκη επιπλέον δεδομένων μειώνει τις ακρίβειες.

ΚΕΦΑΛΑΙΟ 5 ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΤΑΣΕΙΣ

Στην παρούσα διπλωματική εργασία αξιοποιήθηκαν τα ελεύθερα διαθέσιμα πολυφασματικά διαχρονικά δεδομένα του δορυφόρου Sentinel-2 του έτους 2016 από 4 tiles για την παραγωγή μοντέλων μηχανικής μάθησης με τη χρήση υπερκύβων φασματικών και χρονικών χαρακτηριστικών και τη σύγκριση της αποδοτικότητας των μοντέλων μηχανικής μάθησης με αυτούς τους δύο τύπους των δεδομένων. Για κάθε tile και κύβο εφαρμόστηκε μεγάλο πλήθος πειραμάτων με διαφορετικούς συνδυασμούς χαρακτηριστικών, αρχικά για ταξινόμηση ανά tile και στη συνέχεια διεξήχθησαν πειράματα με συνδυασμό των tiles. Επιπρόσθετα υλοποιήθηκαν και πειράματα γενίκευσης μοντέλων σε άγνωστα δεδομένα ώστε να βρεθεί το κατά πόσο είναι εφικτή η εκπαίδευση ενός μοντέλου σε μια περιοχή και η εφαρμογή του σε μια διαφορετική περιοχή. Επιπλέον παράχθηκαν χάρτες καλύψεων γης και καλλιιεργειών για όλα τα tiles.

Η προτεινόμενη μεθοδολογία εφαρμόστηκε αρχικά στο tile 34SEJ με 34 πειράματά με διαφορετικό συνδυασμό χαρακτηριστικών ταξινόμησης, ώστε να επιλεγούν τα πειράματα τα οποία εν τέλει θα εφαρμόζονταν και στα υπόλοιπα tiles. Διαπιστώθηκε ότι επιτυγχάνονται πολύ υψηλές γενικές ακρίβειες OA άνω του 90% στην μεγάλη πλειοψηφία των περιπτώσεων. Σε ότι αφορά τον κύβο των φασματικών χαρακτηριστικών παρατηρήθηκαν γενικές ακρίβειες άνω του 92,36% οι οποίες ήταν υψηλότερες σε σχέση με την γενική ακρίβεια της πλειονότητας του κύβου των φασματικών χαρακτηριστικών ακόμα και στην περίπτωση εξαγωγής 15 καναλιών. Σχετικά με τον κύβο των χρονικών χαρακτηριστικών οι ακρίβειες παρότι μικρότερες σχεδόν κατά 2% στα περισσότερα αντίστοιχα πειράματα (όπου αυτή η αντιστοιχία ορίζεται) σε περιπτώσεις όπως αυτές των κατηγοριών του αστικού ιστού πέτυχαν υψηλότερες ακρίβειες σε σχέση με τον κύβο των φασματικών χαρακτηριστικών. Η χαμηλότερη γενική ακρίβεια (OA) για τον κύβο των φασματικών χαρακτηριστικών ήταν 85,02% για ταξινόμηση σε μόλις 7 κανάλια. Η προσθήκη του DSM οδήγησε σε σημαντική αύξηση των γενικών ακριβειών άνω του 0,6% σε όλες τις περιπτώσεις και μπορεί να θεωρηθεί επιτυχημένη. Μάλιστα παρατηρήθηκε ότι οι ακρίβειες αυξήθηκαν σχεδόν κατά 1% για την περίπτωση του κύβου των χρονικών χαρακτηριστικών και επομένως φαίνεται να επηρεάζονται περισσότερο οι ακρίβειες των χρονικών σε σχέση με τα φασματικά χαρακτηριστικά. Αντιθέτως η προσθήκη του δείκτη της φωτεινότητας οδηγεί σε σχεδόν μηδενικές διαφορές στην ακρίβεια τόσο την γενική όσο και την μέση ακρίβεια f1. Εξάγεται λοιπόν το συμπέρασμα ότι εν αντιθέσει με το DSM ο δείκτης της φωτεινότητας δεν προσφέρει επιπρόσθετη πληροφορία προς αξιοποίηση στην ταξινόμηση.

Για τα 6 πειράματα τα οποία εφαρμόστηκαν σε όλα τα tiles παρατηρήθηκε ότι το tile με τις υψηλότερες γενικές ακρίβειες ήταν το 35TLF με το εύρος των γενικών ακριβειών (OA) από τα πειράματα να κινείται στο (94,67%-96,98%) και το εύρος f1 να είναι 75,33%-81,89%. Και σε αυτή την περίπτωση τα χρονικά χαρακτηριστικά έχουν χειρότερες ακρίβειες σε σχέση με τον κύβο των φασματικών χαρακτηριστικών. Ακόμα παρατηρήθηκε ότι παρότι οι γενικές ακρίβειες είναι πολύ υψηλές οι ακρίβειες f1 είναι οι χαμηλότερες ανάμεσα στα πειράματα. Οι υψηλές γενικές ακρίβειες OA σχετίζονται με το μεγάλο πλήθος δεδομένων αναφοράς, ενώ οι χαμηλές ακρίβειες f1 σχετίζονται με το γεγονός ότι αυτά τα

δεδομένα εκπαίδευσης για αυτό το tile ήταν τα μόνα που δεν είχαν ελεγχθεί και διορθωθεί από την υποψήφια διδάκτορα Χριστίνα Καρακίζη και άρα η ποιότητα τους ήταν αναμενόμενα λίγο κατώτερη. Το εύρος στο οποίο κινούνται οι γενικές ακρίβειες και οι ακρίβειες f1 για το tile 34TEK είναι 94,50%-96,00% και 83,27%-87,22%. Και εδώ παρατηρούνται πολύ μικρές διαφορές ανάμεσα στον κύβο των φασματικών και των χρονικών χαρακτηριστικών. Για το 34SEJ το εύρος ακριβειών είναι 93,29%-94,83% και οι ακρίβειες f1 είναι οι υψηλότερες 88,09%-89,18% από όλα τα πειράματα. Εν τέλει οι γενικές ακρίβειες του tile 34SEG είναι 88,81%-90,49% και οι ακρίβειες f1 έχουν εύρος όμοιο με αυτό του 34SEJ 83,86%-87,34%.

Σε ότι αφορά τα χαρακτηριστικά τα οποία διατηρήθηκαν κατά την εφαρμογή τεχνικών επιλογής καναλιών παρατηρήθηκε ότι για την εξαγωγή καναλιών από τον κύβο των φασματικών και των χρονικών χαρακτηριστικών σε μεγαλύτερο ποσοστό διατηρήθηκε το κανάλι μπλε από τους ανοιξιάτικους και θερινούς μήνες του 2016. Κάτι τέτοιο θεωρείται αναμενόμενο αφού το κανάλι μπλε είναι εξαιρετικά χρήσιμο για την διάκριση μεταξύ κωνοφόρων και φυλλοβόλων δέντρων τα οποία εμφανίζονται σε πολύ μεγάλες συγκεντρώσεις σε όλα τα tiles. Επιπλέον εξαιρετικά χρήσιμο είναι το μπλε και στην διάκριση μεταξύ νεφών και χιονιού ή καλλιεργειών και γυμνού εδάφους. Ακόμα διατηρούνται και τα περισσότερα κανάλια του πράσινου επίσης από τους ανοιξιάτικους και θερινούς μήνες. Και αυτό αναμένεται καθώς το πράσινο κανάλι συμβάλει σημαντικά στην διάκριση ανάμεσα στα είδη βλάστησης καθώς και την διάκριση του αστικού πρασίνου με τον υπόλοιπο αστικό ιστό. Παρατηρήθηκε ακόμα ότι οι δείκτες NDVI και NDWI διατηρούνται στις περισσότερες περιπτώσεις. Αυτό εξηγείται από το γεγονός ότι οι δείκτες βοηθούν στην διάκριση μεταξύ των διαφορετικών ειδών βλάστησης ανάλογα και με την υγρασία σε αυτά. Τέλος σε σχέση με τα χρονικά χαρακτηριστικά που διατηρήθηκαν στον μεγαλύτερο βαθμό παρατηρήθηκε πως αυτά ήταν ο μέσος όρος και η διάμεσος.

Σε ότι αφορά τα πειράματα γενίκευσης με άγνωστα δεδομένα (εκπαίδευση στο ένα έλεγχος στο άλλο) παρατηρήθηκε ότι για τα πειράματα γενίκευσης με άγνωστα δεδομένα τα αποτελέσματα έχουν από χαμηλές μέχρι υψηλές γενικές ακρίβειες με το εύρος της OA να κυμαίνεται στα 48,90%-83,99%. Οι τιμές γενικών ακριβειών δεν επιτρέπουν να μπορεί να θεωρηθούν τα πειράματα γενίκευσης σε άγνωστα δεδομένα μεταξύ 2 tiles ως μια αξιόπιστη εναλλακτική σε κάθε περίπτωση. Μάλιστα οι τιμές των μέσων ακριβειών f1 είναι σε πολλές περιπτώσεις ακόμα και κάτω από 40%. Μοναδική εξαίρεση κατά την οποία οι γενικές ακρίβειες και οι μέσες ακρίβειες f1 μπορούν να θεωρηθούν αποδεκτές είναι οι περιπτώσεις κατά τις οποίες το μοντέλο μεταφέρεται ανάμεσα σε 2 γειτνιάζοντα tiles, δηλαδή τα 34TEK και 34SEJ. Σε αυτή την περίπτωση οι γενικές ακρίβειες βρίσκονται στο εύρος 75,73%-83,99% και οι f1 46,09%-63,44%. Αυτό καταδεικνύει την σημασία που παίζει η γειτνίαση καθώς και οι ομοιότητες σε ανάγλυφο και μικρόκλιμα που υπάρχουν μεταξύ των 2 περιοχών. Σε σχέση με τα συνδυαστικά πειράματα ανάμεσα σε 2 tiles (εκπαίδευση και έλεγχος και στα 2 tiles) παρατηρήθηκαν γενικές ακρίβειες οι οποίες σε κάποιες περιπτώσεις (μεμονωμένων κατηγοριών) υπερβαίνουν την ακρίβεια των συγκεκριμένων κατηγοριών για την ταξινόμηση σε ένα tile (εύρος 94,27%-96,42%). Σημαντικό ρόλο για την επιτυχία της ταξινόμησης με αυτή την μέθοδο παίζει το να είναι δυνατή η συμπλήρωση πιθανών κενών που υπάρχουν στα δεδομένα αναφοράς του ενός tile από το άλλο.

Ακολούθως σχολιάζονται τα αποτελέσματα για τα πειράματα γενίκευσης με άγνωστα δεδομένα μεταξύ 3 tiles τα οποία έγιναν ώστε να συγκριθούν τα αποτελέσματα:

- Από την εκπαίδευση σε 2 tiles που γειτνιάζουν προς ένα tile το οποίο δεν γειτνιάζει με κανένα από τα προηγούμενα
- Από την εκπαίδευση σε 2 tiles που δεν γειτνιάζουν προς ένα με το οποίο επίσης δεν γειτνιάζουν.

Σε αυτή την σειρά πειραμάτων παρατηρήθηκε ότι τα αποτελέσματα βελτιώθηκαν σημαντικά σε σχέση με την πειράματα γενίκευσης με άγνωστα δεδομένα ανά 2 tiles. Συγκεκριμένα παρατηρήθηκε ότι το εύρος των γενικών ακριβειών OA βρίσκεται ανάμεσα στις τιμές 54,72%-80,88% το οποίο κατέδειξε ότι οι γενικές ακρίβειες είναι σημαντικά αυξημένες σε σχέση με τις ακρίβειες για γενίκευση σε άγνωστα δεδομένα από εκπαίδευση σε ένα tile (48,90%-83,99%). Οι ακρίβειες f1 επίσης εμφανίστηκαν αρκετά αυξημένες. Παρατηρήθηκε ότι οι γενικές ακρίβειες επηρεάστηκαν από την γειτνίαση αφού τα καλύτερα αποτελέσματα επιτεύχθηκαν για τις περιπτώσεις όπου τα tiles ήταν διάσπαρτα στον χώρο και δεν γειτνιάζουν. Οι γενικές ακρίβειες OA για tiles που γειτνιάζουν είχαν εύρος (73,17%-80,88%) ενώ για αυτά που δεν γειτνιάζουν εύρος 54,72%-67,40%). Αντίθετα οι ακρίβειες f1 δεν φαίνεται να επηρεάστηκαν άμεσα την γειτνίαση ή μη αλλά κυρίως φάνηκε να επηρεάζονται από την ποιότητα των δεδομένων και την δυνατότητα των δεδομένων αναφοράς από το ένα tile να συμπληρώνονται από το άλλο. Τέλος στα πειράματα που έγιναν με αύξηση των tiles που χρησιμοποιήθηκαν για εκπαίδευση παρατηρήθηκε σημαντική αύξηση των ακριβειών.

Σε ότι αφορά την σύγκριση μεταξύ πειραμάτων γενίκευσης με εκπαίδευση σε διαφορετικό αριθμό tiles που έγινε στην παράγραφο 4.6.5. παρατηρήθηκε ότι η προσθήκη παραπάνω tiles για την εκπαίδευση συνεπάγεται σημαντική αύξηση στις ακρίβειες f1. Το εύρος των ακριβειών f1 βρίσκεται ανάμεσα στις τιμές 44,15% και 64,93%. Παρατηρήθηκε ότι η ακρίβεια f1 αυξάνεται κατά $\approx 10\%$ σε κάθε περίπτωση που προστέθηκε ένα νέο tile. Αξίζει επιπλέον να αναφερθεί ότι οι διαφοροποιήσεις στις ακρίβειες f1 ανά κατηγορία είναι ενίοτε ακόμα υψηλότερες σε κάποιες περιπτώσεις. Υπάρχουν περιπτώσεις όπου κατηγορίες με ακρίβεια f1 $\approx 0\%$ για πείραμα με εκπαίδευση σε ένα tile έχουν ακρίβειες f1 $> 70\%$ για εκπαίδευση μοντέλου σε 3 tiles. Σε ότι αφορά τις γενικές ακρίβειες OA το εύρος τους βρίσκεται ανάμεσα στις τιμές 58,96% και 76,19%. Παρατηρείται οι ακρίβειες OA διαφοροποιούνται σημαντικά ($> 3\%$) με την προσθήκη επιπλέον tiles για την εκπαίδευση. αλλά όχι στον ίδιο βαθμό με τις ακρίβειες f1. Επομένως η προσθήκη tiles είναι πολύ χρήσιμη κατά την γενίκευση μοντέλων. Αντίθετα τα συνδυαστικά πειράματα παρουσιάζουν μια εντελώς διαφορετική εικόνα με την προσθήκη νέων tiles για εκπαίδευση και έλεγχο. Οι ακρίβειες f1 κινούνται στο εύρος 84,67%-86,18% και οι OA στο εύρος 94,36%-96,21%. Σε αυτή την περίπτωση παρατηρήθηκε ότι οι ακρίβειες είναι υψηλότερες όταν χρησιμοποιούνται 2 tiles (ειδικά όταν αυτά γειτνιάζουν) για εκπαίδευση και έλεγχο από την χρήση 3 ή παραπάνω tiles. Αυτό αναμένεται από το γεγονός ότι τα tiles που γειτνιάζουν τείνουν να έχουν πολύ υψηλές ακρίβειες σε συνδυαστικά πειράματα.

Συμπερασματικά αποδείχθηκε ότι ο κύβος των χρονικών χαρακτηριστικών είναι μια αποτελεσματική

εναλλακτική λύση για τη δόμηση μοντέλου μηχανικής μάθησης ικανή να δώσει πολύ υψηλές ολικές ακρίβειες ταξινόμησης, υψηλότερες του 90%, στην περίπτωση δόμησης μοντέλου μηχανικής μάθησης για έναν κύβο. Επιπρόσθετα η χρήση των χρονικών χαρακτηριστικών στην περίπτωση των πειραμάτων γενίκευσης με άγνωστα δεδομένα αλλά και των συνδυαστικών πειραμάτων επιτρέπει τα πειράματα ανάμεσα σε δεδομένα (κύβους) που δεν έχουν προβλεφθεί από το ίδιο πλήθος πολυφασματικών απεικονίσεων ή τις ίδιες ημερομηνίες, δίνοντας έτσι τη δυνατότητα να ξεπεραστεί ένας σημαντικός περιορισμός της δόμησης μοντέλων μηχανικής μάθησης σε διαφορετικές περιοχές/tiles.

Στις επόμενες παραγράφους αναλύονται τα ειδικότερα συμπεράσματα που προέκυψαν από την ανάλυση της ποιότητας των ταξινομήσεων και κυρίως τη μελέτη της σύγχυσης στην ταξινόμηση μεταξύ των διαφορετικών κατηγοριών. Πιο συγκεκριμένα παρατηρήθηκε ότι στις κατηγορίες από τις οποίες απαρτίζεται το ανθρωπογενές περιβάλλον οι ακρίβειες της ταξινόμησης κινούνται συνήθως από χαμηλές έως και πολύ υψηλές ενώ υπάρχουν κατηγορίες με πολύ υψηλά ποσοστά σύγχυσης σε κάθε πείραμα. Οι κατηγορίες πυκνή και αραιή αστική δόμηση και εμπορικές/βιομηχανικές χρήσεις είναι αισθητά δύσκολο να διαχωριστούν από το μοντέλο μηχανικής μάθησης σχεδόν σε όλα τα πειράματα. Κάτι τέτοιο μπορεί να θεωρηθεί αναμενόμενο αφού το σκυρόδεμα το οποίο είναι το κύριο δομικό υλικό για τις τρεις αυτές χρήσεις γης έχει κοινή φασματική υπογραφή και στις τρεις περιπτώσεις. Ένας επιπλέον λόγος της σύγχυσης, όχι μόνο μεταξύ της πυκνής και αραιής αστικής δόμησης αλλά και όλων των κατηγοριών οι οποίες διαχωρίζονται ανάλογα με την πυκνότητα των δομικών τους στοιχείων, σχετίζεται με το γεγονός ότι αυτές οι κατηγορίες δε διαχωρίζονται ποσοτικά αλλά ποιοτικά από τον φωτοερμηνευτή που δύσκολα μεταφράζεται από τον υπολογιστή. Η κατηγορία των οδικών δικτύων έχει γενικώς κακές ακρίβειες ταξινόμησης κυρίως λόγω του γραμμικού μοτίβου της, το οποίο σε συνδυασμό με το μικρό πλάτος και την ανάλυση των 10 m δεν επέτρεψε καλή ποιότητα ταξινόμησης. Εν τέλει σε ό,τι αφορά τις ανθρωπογενείς κατηγορίες παρατηρήθηκε ότι η κατηγορία των λατομείων/μεταλλείων, ειδικά στις περιπτώσεις της μεταφοράς μοντέλων μηχανικής μάθησης, τείνει να έχει πολύ χαμηλές ακρίβειες χρήστη και έτσι ανήκει στις κατηγορίες που μειώνουν την ακρίβεια χρήστη. Στην κατηγορία των λατομείων παρατηρείται ότι τα κύρια σφάλματα παράλειψης κατευθύνονται προς την κατηγορία των εμπορικών και βιομηχανικών εγκαταστάσεων πράγμα που δείχνει μεγάλη δυσκολία του ταξινομητή να ταξινομήσει τα δεδομένα αυτής της κατηγορίας σωστά ενώ μάλιστα δεν υπάρχει ομοιότητα μεταξύ των φασματικών υπογραφών των 2 κατηγοριών.

Τα κωνοφόρα και τα φυλλοβόλα δέντρα παρατηρήθηκε ότι έχουν πολύ υψηλές ακρίβειες f1, οι οποίες στις περισσότερες περιπτώσεις ξεπέρασαν το 90%. Τα κωνοφόρα δέντρα έχουν πολύ χαρακτηριστική κώμη, η οποία είναι εύκολα διαχωρίσιμη από την κώμη όλων των άλλων δέντρων. Επιπλέον τα δέντρα αυτά φύονται σε υψηλά έως και πολύ υψηλά υψόμετρα και είναι η επικρατούσα βλάστηση σε τέτοια υψόμετρα. Ο μεγάλος αριθμός κωνοφόρων και η αδιάκοπη συγκέντρωσή τους σε συγκεκριμένες περιοχές του χάρτη, σε συνδυασμό με την χαρακτηριστική τους φασματική υπογραφή, καθιστά αναμενόμενες τις πολύ υψηλές ακρίβειες ταξινόμησης που επιτυγχάνονται για αυτή την κατηγορία. Τα φυλλοβόλα δέντρα έχουν επίσης χαρακτηριστική κώμη ενώ φύονται σε μικρότερα υψόμετρα σε σχέση με τα κωνοφόρα. Τα φυλλοβόλα δέντρα το φθινόπωρο ρίχνουν τα φύλλα τους, και έτσι η φασματική

τους συμπεριφορά μεταβάλλεται, γεγονός που τονίζει τη σημασία της χρήσης διαχρονικών (multi-temporal) δεδομένων για την ορθή ταξινόμηση τους.

Σε ό,τι αφορά τη σκληροφυλλική βλάστηση, παρατηρήθηκε σε όλες τις περιπτώσεις σύγχυση μεταξύ των υποκατηγοριών της (πυκνή και αραιή σκληροφυλλική βλάστηση) για τους ίδιους λόγους που υπήρξε και μεταξύ της πυκνής και αραιής αστικής δόμησης (συναφής φασματική υπογραφή, έλλειψη σαφώς καθορισμένου ορίου μεταξύ των κατηγοριών). Επίσης παρατηρήθηκε σύγχυση μεταξύ της σκληροφυλλικής βλάστησης και των πλατύφυλλων δέντρων εξαιτίας της ομοιότητας στην κώμη. Εν τέλει η τρίτη κατηγορία για την οποία παρατηρήθηκε μεγάλο ποσοστό σύγχυσης ήταν αυτή των χορτολιβαδικών εκτάσεων με τις περιοχές με αραιή βλάστηση.

Στην περίπτωση των καλλιεργειών υπήρξε ιδιαίτερη σύγχυση στα πρώτα πειράματα μεταξύ κατηγοριών παρεμφερών φασματικών υπογραφών που ανήκουν στην ίδια οικογένεια καλλιέργειας (π.χ. σιτάρι, βρώμη, κριθάρι). Γενικά για όλα τα πειράματα το εύρος των ακριβειών ήταν εκτεταμένο, με τις τιμές να κυμαίνονται από πολύ χαμηλές έως και αρκετά υψηλές. Παρατηρήθηκε ότι οι κατηγορίες του καπνού και της αγρανάπαυσης ήταν αδύνατο να ταξινομηθούν σχεδόν σε όλες τις περιπτώσεις μηχανικής μάθησης με τις ακρίβειες $f1$ να είναι $\approx 0\%$. Για την αγρανάπαυση αυτό αναμένεται καθώς είναι μια μη σταθερή κάλυψη γης μέσα στο έτος και επιπλέον είναι μια χρήση γης που μπορεί να εμφανιστεί σε περιοχές όπου στην προηγούμενη ή την επόμενη εικόνα υπήρχε οποιαδήποτε καλλιέργεια. Παρατηρήθηκε επίσης μεγάλο ποσοστό σύγχυσης μεταξύ ελαιόδεντρων και αραιής σκληροφυλλικής βλάστησης καθώς η κώμη των δυο φυτών σε πολλές περιπτώσεις μοιάζει. Σε πολλές περιπτώσεις κατηγορίες οι οποίες είναι "ομπρέλα" για πολλές μικρής έκτασης καλλιέργειες (όπως οι ζωτροφές και τα δημητριακά) έχουν πολύ υψηλή ακρίβεια για μηχανική μάθηση σε ένα tile αφού η επικρατούσα καλλιέργεια συμβάλλει κυρίως στον σχηματισμό της φασματικής υπογραφής. Αντιθέτως σε περιπτώσεις πειραμάτων γενίκευσης με άγνωστα δεδομένα κάτι τέτοιο πολλές φορές δε συμβαίνει με αποτέλεσμα για την ίδια θεματική κατηγορία να υπάρχουν εντελώς διαφορετικές φασματικές υπογραφές

Οι τελευταίες κατηγορίες που αναλύθηκαν αφορούν το γυμνό έδαφος και τα υδάτινα σώματα. Για το γυμνό έδαφος παρατηρήθηκαν υψηλές ακρίβειες στις περιπτώσεις ταξινόμησης για ένα tile σε όλες τις κατηγορίες. Η κατηγορία των αμμωδών εκτάσεων είναι αρκετά δυσκολότερο να ταξινομηθεί σωστά στα πειράματα γενίκευσης με άγνωστα δεδομένα για τους λόγους που θα αναλυθούν παρακάτω. Για την περίπτωση των υδάτινων σωμάτων οι ακρίβειες ταξινόμησης είναι πολύ υψηλές για την ταξινόμηση σε ένα tile αλλά οι υψηλές ακρίβειες για αυτή την περίπτωση δεν μεταφράστηκαν σε υψηλές ακρίβειες στην περίπτωση της γενίκευσης με άγνωστα δεδομένα. Πιο συγκεκριμένα παρατηρήθηκε ότι σε πολλές περιπτώσεις υπήρξε σύγχυση μεταξύ των λιμνών και των ποταμών. Αυτό εξηγείται καθώς η διάκριση μεταξύ των δύο δυσχεραίνεται από την εγγύτητα καθώς και την ομοιότητα στις φασματικές υπογραφές. Εν τέλει παρατηρήθηκε για τα πειράματα γενίκευσης σε άγνωστα δεδομένα υψηλή σύγχυση μεταξύ των ποταμών και των αμμωδών εκτάσεων. Αυτό εξηγείται από το γεγονός ότι τα δεδομένα αναφοράς για τις αμμώδεις εκτάσεις προέρχονται συχνά από κοίτες ποταμών,

που είναι σε γειτνίαση με το νερό.

Προτάσεις

Από την αναπτυχθείσα μεθοδολογία προκύπτει πλήθος προτάσεων για περαιτέρω επέκταση και εμπλουτισμό. Μια πρώτη πρόταση θα ήταν η επέκταση της μεθοδολογίας σε άλλες περιοχές. Πιο συγκεκριμένα η γενίκευση των μοντέλων σε άγνωστα δεδομένα από εν μέρει διαφορετικές κλιματολογικές και αγροοικολογικές ζώνες θα προσέφερε μια καλή ένδειξη για τα όρια του μοντέλου. Λόγου χάρη ιδιαίτερα χρήσιμα θα ήταν τα αποτελέσματα των ακριβειών για γενίκευση του μοντέλου προερχόμενου από δεδομένα της Πελοποννήσου σε άλλα μεσογειακά εδάφη με παρόμοιο γεωγραφικό πλάτος όπως είναι η Σικελία ή η νότια Ισπανία. Με βάση το γεγονός ότι η εκπαίδευση σε 2 tiles και ο έλεγχος σε κάποιο άλλο αποδείχθηκε ήδη πως είναι εφικτό να δώσει υψηλές ακρίβειες θα ήταν ιδιαίτερα χρήσιμη η δυνατότητα γενίκευσης μοντέλων στον ευρύτερο μεσογειακό χώρο. Εδώ κομβικό ρόλο μπορεί να παίξει η ποικιλία κλιματολογικών συνθηκών και οι μεγάλες διαφοροποιήσεις στο ανάγλυφο που υπάρχουν εντός του Ελλαδικού χώρου διότι πιθανόν οι δυο παραπάνω παράγοντες να συμβάλλουν σε σημαντικό βαθμό στην δυνατότητα γενίκευσης του μοντέλου.

Ακόμα και προκειμένου να αυξηθούν ακόμα περισσότερο οι γενικές ακρίβειες και οι μέσες ακρίβειες $f1$ να είναι υψηλότερες του 90% προτείνεται να δομηθούν κύβοι από ακόμα περισσότερες ημερομηνίες. Εν τέλει θα μπορούσε να προταθεί ο πειραματισμός με προσθήκη πολλών νέων χρονικών χαρακτηριστικών τα οποία θα μπορούσαν να αντικαταστήσουν κάποια από τα υπάρχοντα ή να προστεθούν σε αυτά. Ενδεικτικά αναφέρονται τα παρακάτω:

- Η κλίση (slope) είναι ένα χαρακτηριστικό χρησιμοποιούμενο σε κάποιες περιπτώσεις στην διεθνή βιβλιογραφία το οποίο μάλιστα προσδίδει και αρκετά υψηλές ακρίβειες.
- Ο πειραματισμός με άλλα εκατοστημόρια πλην του 25% και 75% θα μπορούσε να αποτελέσει ενδιαφέρουσα εναλλακτική καθώς τα παραπάνω εκατοστημόρια φάνηκε να βελτιώνουν την ακρίβεια.
- Η αντικατάσταση του μέγιστου με το εκατοστημόριο 99% το οποίο θα μπορούσε να αντιμετωπίσει πιθανά “καμένα” εικονοστοιχεία που υπάρχουν στις απεικονίσεις.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- 1: Karakizi *, I. A. Tsiotas, Z. Kandyllakis, A. Vaiopoulos, K. Karantzas : (2020) 'Assessing the Contribution of Spectral and Temporal Features for Annual Land Cover and Crop Type Mapping'
- 2: Nicolas Matton et al. (2015) An Automated Method for Annual Cropland Mapping along the Season for Various Globally-Distributed Agrosystems Using High Spatial and Temporal Resolution Time Series
- 3: Silvia Valero, David Morin, Jordi Inglada, Guadalupe Sepulcre, Marcela Arias, Olivier Hagolle, Gérard Dedieu, Sophie Bontemps, Pierre Defourny and Benjamin Koetz (2016) Production of a Dynamic Cropland Mask by Processing Remote Sensing Image Series at High Temporal and Spatial Resolutions
- 4: Γεωργία Αντωνίου 2020 Ετήσια Χαρτογράφηση Κάλυψης Γης Και Καλλιεργειών Με Χρήση Δεδομένων Sentinel-2 Και Τεχνικές Μηχανικής Μάθησης
- 5: Robert N Conwell, The photo interpretation picture in 1960, 1960
- 6: Δημήτρης Αργιαλάς, «Φωτοερμηνεία- Τηλεπισκόπηση», Αθήνα: Εθνικό Μετσόβιο Πολυτεχνείο. , 1999
- 7: John R Jensen, Remote Sensing of the Environment: An Earth Resource Perspective, 2007
- 8: Περάκης, Η τηλεπισκόπηση σε 13 ενότητες, 2015
- 9: Yugal & Sahoo (2012) Analysis of Parametric & Non Parametric Classifiers for Classification Technique using WEKA
- 10: Quinlan, J. R. (1986) Induction of decision trees
- 11: Leo Breimman (2001) Random Forest
- 12: Lu Weng (2007) A Survey of Image Classification Methods and Techniques for Improving Classification Performance
- 13: Κωνσταντίνα Τσιώκου, (2019) Χαρτογράφηση Καλλιεργειών και Κάλυψης Γης από Διαχρονικά Δεδομένα Sentinel-2 για τα έτη 2016 και 2017,
- 14: Γεωργία Αντωνίου, (2018) Χαρτογράφηση κάλυψης γης με χρήση δεδομένων sentinel-2 από πολλαπλές ημερομηνίες μέσω μηχανών διανυσματικής υποστήριξης,
- 15: Ουζούνoglou Ελευθέριος (2018) Μέθοδοι Βιοπληροφορικής, Υπολογιστικής Συστημικής Βιολογίας και Μηχανικής Μάθησης, Εφαρμοσμένες στην In Silico Ογκολογία
- 16: Ολυμπία Γούναρη, (2018) Χαρτογράφηση Κάλυψης Γης και Καλλιεργειών με Τεχνικές Μηχανικής Μάθησης από Διαχρονικά Δεδομένα Sentinel-2,
- 17: Karakizi, C.; Karantzas, K.; Vakalopoulou, M.; Antoniou, G., (2018) Detailed Land Cover Mapping from Multitemporal Landsat-8 Data of Different Cloud Cover,
- 18: Γεωργία Αντωνίου, (2018) Χαρτογράφηση κάλυψης γης με χρήση δεδομένων sentinel-2 από πολλαπλές ημερομηνίες μέσω μηχανών διανυσματικής υποστήριξης,
- 19: Zhu and Woodcock, (2012) Improvement and Expansion of the Fmask Algorithm: Cloud, Cloud Shadow, and Snow Detection for Landsats 4-7, 8, and Sentinel 2 Images,
- 20: Sherrie Wang, George Azzari, David B. Lobell 2019, Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques, 2019
- 21: Baatz M and Schäpe A (2000) Multiresolution Segmentation – an optimization approach for high quality multi-scale image segmentation.

- 22: Ιωάννης Γκίνης, (2019) Διερεύνηση Τεχνικών Μεταφοράς Μάθησης για Χαρτογράφηση Κάλυψης Γης με Δεδομένα Αναφοράς Προηγούμενων Ετών
- 23: Khan et al (2018), Evaluating Landsat and RapidEye Data for Winter Wheat Mapping and Area Estimation in Punjab, Pakistan
- 24: Shuai Xie 1,2, Liangyun Liu 1,* , Xiao Zhang etal (2019) Automatic Land-Cover Mapping using Landsat Time-Series Data based on Google Earth Engine
- 25: Shuai Xie, Liangyun Liu , Xiao Zhang, Jiangning Yang , Xidong Chen, and Yuan Gao (2016) Cropland Mapping over Sahelian and Sudanian Agrosystems: A Knowledge-Based Approach Using PROBA-V Time Series at 100-m
- 26: Yongguang Zhai 1,2 ID , Zhongyi Qu 1,* and Lei Hao 3 (2018) Land Cover Classification Using Integrated Spectral, Temporal, and Spatial Features Derived from Remotely Sensed Images
- 27: Alexey V. Egorov 1 ID , David P. Roy 1,* ID , Hankui K. Zhang 1 ID , Matthew C. Hansen 2 and Anil Kommareddy 2 Demonstration of Percent Tree Cover Mapping Using Landsat Analysis Ready Data (ARD) and Sensitivity with Respect to Landsat ARD Processing Level
- 28: Antonio D’Ambrosio Valerio A. Tutore (2011) Conditional Classification Trees by Weighting the Gini Impurity Measure
- 29: Qi J., Chehbouni A., Huete A.R., Kerr Y.H., 1994. Modified Soil Adjusted Vegetation Index (MSAVI). *Remote Sens Environ* 48:119-126.
- 30: Zha, Y., Gao, J. and Ni, S. (2003) Use of Normalized Difference Built-Up Index in Automatically Mapping Urban Areas from TM Imagery. *International Journal of Remote Sensing*, 24, 583-594.
- 31: McFeeters, S.K. (1996) The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features. *International Journal of Remote Sensing*, 17, 1425-1432.
- 32: Rouse, J.W., R.H. Haas, J.A. Schell, and D.W. Deering, 1974. Monitoring vegetation systems in the Great Plains with ERTS, In: S.C. Freden, E.P. Mercanti, and M. Becker (eds) *Third Earth Resources Technology Satellite—1 Symposium. Volume I: Technical Presentations*, NASA SP-351, NASA, Washington, D.C., pp. 309-317.
- 33: Bannari, A., Morin, D., Bonn, F. and Huete, A. R. (1995) 'A review of vegetation indices', *Remote Sensing Reviews*, 13: 1, 95 — 120.
- 34: Cortes, Corinna and Vapnik Vladimir N., (1995) "Support-Vector Networks", *Journal of Machine Learning*, 20 (3), p.p. 273-297.
- 35: Inglada, J.; Arias, M.; Tardy, B.; Hagolle, O.; Valero, S.; Morin, D.; Dedieu, G.; Sepulcre, G.; Bontemps, S.; Defourny, P.; Koetz, B. Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Satellite Optical Imagery. *Remote Sens.* 2015, 7, 12356-12379.