

Μάθημα: Διαχείριση και Επεξεργασία Μεγάλων Δεδομένων Παρατήρησης Γης

ΔΠΜΣ: ΕΔΕΜΜ

Άσκηση: 2η - Ταξινόμηση χρονοσειρών πολυφασματικών δορυφορικών δεδομένων

Ονοματεπώνυμο: Στέφανος Πέτκοβιτς

Αριθμός Μητρώου: 03400281

Προθεσμία Παράδοσης: 8/6/2025

Περίληψη Άσκησης

Στην παρούσα άσκηση, θα υλοποιηθεί μια μεθοδολογία ταξινόμησης σε κατηγορίες κάλυψης αγροτικής γης από χρονοσειρές Sentinel-2, με χρήση αρχιτεκτονικής τεχνητού νευρωνικού δικτύου τύπου transformers, παρόμοια με αυτή που αναλύεται στην δημοσίευση [1]. Στα παρακάτω κεφάλαια θα περιγραφεί η αρχιτεκτονική του μοντέλου, τα απαιτούμενα βήματα για την προ-επεξεργασία του συνόλου δεδομένων καθώς και η στρατηγική εκπαίδευσης.

Δεδομένα & Προεπεξεργασία

Το σύνολο δεδομένων της παρούσας άσκησης αποτελείται από 5.000 χρονοσειρές Sentinel-2 με 10 φασματικά κανάλια οι οποίες αποτυπώνουν αγροτεμάχια στην Δανία. Τα διαθέσιμα μεταδεδομένα που συνοδεύουν τις χρονοσειρές είναι οι ετικέτες αληθείας για κάθε αγροτεμάχιο καθώς και οι ημερομηνίες απόκτησης κάθε χρονοσειράς. Για να καταστεί δυνατό να δοθούν όλες οι χρονοσειρές με την μορφή ενός PyTorch ταυστή σαν είσοδος στο μοντέλο ταξινόμησης τύπου “transformer”, έγιναν οι παρακάτω επεξεργασίες μέσω των μεθόδων της κλάσης TimeMatchPreprocessor (όλες οι κλάσεις περιέχονται στο Jupyter notebook που συνοδεύει την τεχνική έκθεση):

1. Ανάγνωση του αρχείου με τις ετικέτες αληθείας (ground truth labels) και αφαίρεση ετικετών οι οποίες αντιστοιχούν σε λιγότερες από 200 χρονοσειρές.
2. Ανάθεση νέων labels στο επεξεργασμένο αρχείο ετικετών.
3. Μετασχηματισμός ημερομηνιών από YYYYMMDD σε DOY (Day of the Year) format.
4. Κανονικοποίηση χρονοσειρών σε επίπεδο εικονοστοιχείου (διαίρεση με την τιμή 65535 για “συμπίεση” του εύρους τιμών από 16bit σε [0,1]).

Επίσης, σε αυτό το στάδιο αναπτύχθηκε και η κλάση PixelSamplerDataset για την επιλογή 32 τυχαίων εικονοστοιχείων για κάθε χρονοσειρά σε κάθε εποχή εκπαίδευσης του μοντέλου τύπου “transformer” για την ταξινόμηση σε κατηγορίες κάλυψης αγροτικής γης. Τέλος, επισημαίνεται ότι κατόπιν της προ-επεξεργασίας οι ταυστές που περιέχουν θα έχουν διαστάσεις (B,T,S,C), όπου:

- I. B → Μέγεθος batch (υπερπαραμέτρος που ορίζεται από τον χρήστη).
- II. T → Μήκος ακολουθίας ημερομηνιών απόκτησης απόκτησης απεικονίσεων Sentinel-2 (σταθερός αριθμός ίσος με 52).
- III. S → Αριθμός φασματικών καναλιών Sentinel-2 (σταθερός αριθμός ίσος με 10).
- IV. C → Πλήθος τυχαία επιλεγμένων εικονοστοιχείων από κάθε χρονοσειρά.

Ανάπτυξη Μοντέλου Ταξινόμησης Χρονοσειρών

Το μοντέλο ταξινόμησης χρονοσειρών τύπου “transformer” αποτελείται από 2 μέρη:

- A. Μια pixel set encoder αρχιτεκτονική η οποία δέχεται σαν είσοδο τανυστές (με την διαστατικότητα που περιγράφηκε στην παραπάνω ενότητα), δίνει σαν έξοδο έναν τανυστή ο οποίος περιέχει την χωρική πληροφορία του μοντέλου και αποτελείται από τις παρακάτω αρχιτεκτονικές:
- Ένα Multi Layer Perceptron (MLP) με μια κρυφή στρώση μεγέθους 16 νευρώνων και batch normalization σε κάθε στρώση, το οποίο δίνει σαν έξοδο έναν τανυστή χαρακτηριστικών με διαστάσεις (B,T,32,32).
 - 2 pooling operations στην χρονική διάσταση μέσω average & standard deviation pooling. Κατόπιν, τα pooled χαρακτηριστικά συνενώνονται και έτσι παράγεται ένας τανυστής διαστάσεων (B,32,64).
 - Ένα MLP χωρίς κρυφές στρώσεις το οποίο εξάγει έναν τανυστή διαστάσεων (B,128).
- B. Εφαρμογή μεθοδολογίας positional encoding στον τανυστή εξόδου του pixel set encoder έτσι ώστε όλα τα δείγματα της χρονοσειράς να καταταχθούν χρονικά βάσει των διαθέσιμων ημερομηνιών απόκτησης τους. Ο σχετικός πίνακας του positional encoding υπολογίζεται μέσω της ημιτονικής χρονικής κωδικοποίησης η οποία φαίνεται στην παρακάτω σχέση (όπου DOY είναι η ακολουθία χρόνων απόκτησης των δειγμάτων της χρονοσειράς, d_{model} το μήκος της ακολουθίας που περιέχει τα χαρακτηριστικά που εξάγονται από τον pixel set encoder ενώ αναφέρεται και ότι ο όρος του παρονομαστή εντός του ημιτόνου τέθηκε ίσος με 1.000 και όχι 10.000 όπως συνηθίζεται σε NLP εργασίες καθώς το μήκος της χρονικής ακολουθίας είναι σχετικά μικρό):

$$PE = \sin(DOY/1000^{2i/d_{model}})$$

Επισημαίνεται ότι ο positionally aware τανυστής χαρακτηριστικών (πρόσθεση τανυστή με PE) συνενώνεται με ένα classification token.

- C. Μια temporal attention encoder αρχιτεκτονική η οποία βασίζεται σε μια μικρή προσαρμογή των attention based νευρωνικών δικτύων και αποτελείται από τις παρακάτω συνιστώσες:
- Πλήρως συνδεδεμένη στρώση από την οποία βάσει του αριθμού των attention heads (υπερπαραμέτρος) υπολογίζονται οι προβολές των key, query και values. Επισημαίνεται ότι σαν είσοδος στην πλήρως συνδεδεμένη στρώση δίνεται το positionally encoded (βάσει των ημερομηνιών απόκτησης των χρονοσειρών οι οποίες μοντελοποιούνται ορίζοντας ένα classification token) διάνυσμα χαρακτηριστικών.
 - Από τον μέσο όρο της query προβολής υπολογίζεται το master query από μια δεύτερη πλήρως συνδεδεμένη στρώση η οποία δεν επιφέρει αλλαγή της διαστατικότητας.
 - Υπολογισμός attention ανά head και συνένωση (concatenation μέσω πολλαπλασιασμού όλων των πινάκων attention που δημιουργούνται ανά head) των αποτελεσμάτων από όλα τα heads. Ο κάθε πίνακας attention υπολογίζεται βάσει της παρακάτω σχέσης (όπου d_k η κοινή διάσταση των queries και keys):

$$Attention = softmax(\frac{qk}{\sqrt{d_k}})$$

Τελικά, από την έξοδο του temporal attention encoder (classification token) εφαρμόζεται το τελικό MLP για την ταξινόμηση των χρονοσειρών στις διαθέσιμες κλάσεις. Όλες οι παραπάνω διαδικασίες υλοποιούνται προγραμματιστικά στο συνημμένο notebook μέσω των μεθόδων των κλάσεων MLP_1, MLP_2, PositionalEncoding, TemporalAttentionEncoder & Decoder.

Εκπαίδευση Μοντέλου και Αξιολόγηση Αποτελεσμάτων

Η στρατηγική εκπαίδευσης θα είναι η διασταυρούμενη επικύρωση (5 fold) με το βέλτιστο fold να επιλέγεται από τον καλύτερο συνδυασμό μέσου όρου και τυπικής απόκλισης των μετρικών ταξινόμησης. Στους παρακάτω πίνακες παρουσιάζονται τα αποτελέσματα μέσου όρου και τυπικής απόκλισης αντίστοιχα όλων των διαθέσιμων μετρικών, για όλες τις εποχές εκπαίδευσης ανά fold στο σύνολο επικύρωσης.

Fold	Mean Accuracy (%)	Mean Weighted Prec. (%)	Mean Micro Prec. (%)	Mean Weighted Rec. (%)	Mean Micro Rec. (%)	Mean Weighted F1 (%)	Mean Micro F1 (%)
1	22.94	5.29	22.94	22.94	22.94	8.59	22.93
2	41.11	35.56	41.11	41.11	41.11	36.17	41.11
3	23.45	0.55	23.45	23.45	23.45	8.92	23.45
4	23.20	5.40	23.20	23.20	23.20	8.76	23.20
5	62.93	60.37	62.93	62.93	62.93	60.03	62.93

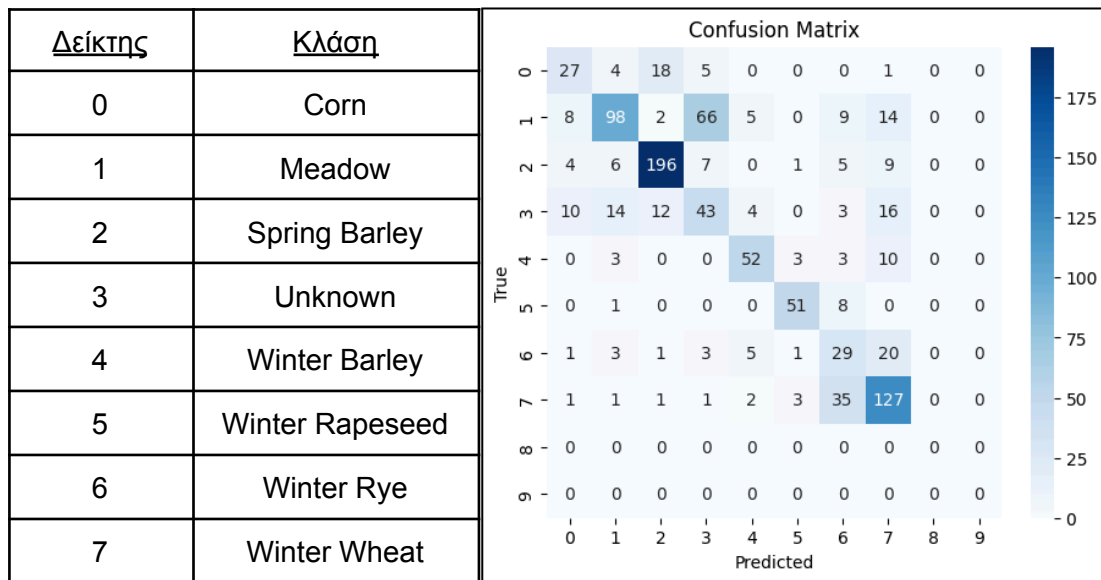
Σχήμα 1. Πίνακα μέσων όρων μετρικών ανά fold διασταυρούμενης επικύρωσης

Fold	Std Accuracy (%)	Std Weighted Prec. (%)	Std Micro Prec. (%)	Std Weighted Rec. (%)	Std Micro Rec. (%)	Std Weighted F1 (%)	Std Micro F1 (%)
1	1.75	0.75	1.75	1.75	1.75	1.12	1.75
2	8.61	14.28	8.61	0.86	0.86	12.92	8.61
3	1.05	0.48	1.05	1.05	1.05	0.70	1.05
4	1.22	0.55	1.22	1.22	1.22	0.81	1.22
5	7.52	11.59	7.52	7.52	7.52	10.62	7.52

Σχήμα 2. Πίνακα τυπικών αποκλίσεων μετρικών ανά fold διασταυρούμενης επικύρωσης

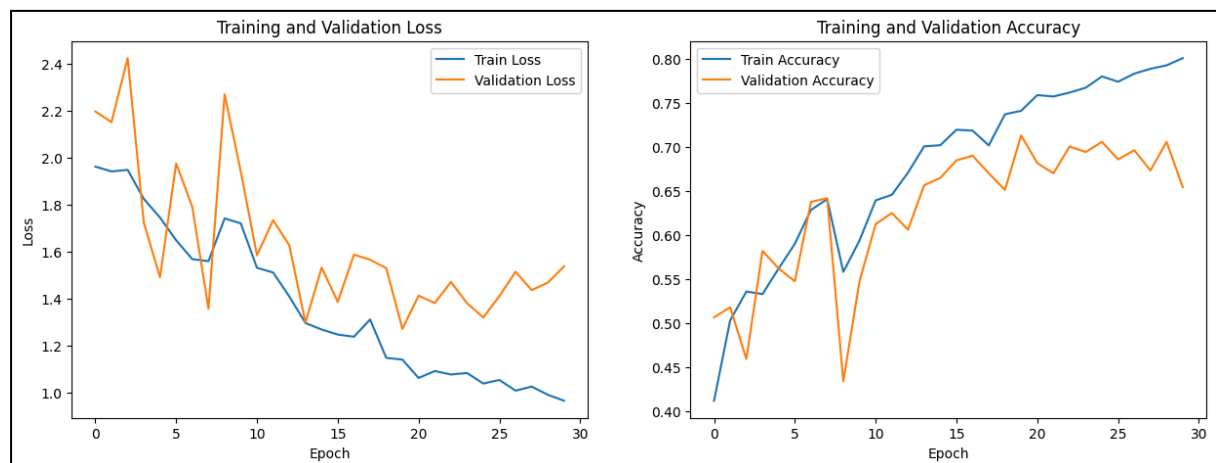
Βάσει των παραπάνω αποτελεσμάτων ως βέλτιστο fold επιλέγεται το 5ο. Επισημαίνεται ότι στα folds με αρκετά χαμηλότερο μέσο όρο ακριβειών ανά εποχή (1ο και 3ο και 5ο), όλα τα δείγματα ταξινομούνται λανθασμένα σε 1 ή 2 κλάσεις, κάτι το οποίο μπορεί να οφείλεται σε

αρκετούς λόγους όπως η άνιση κατανομή των ετικετών κατά μήκος του συνόλου δεδομένων. Από τον πίνακα σύγχυσης για το βέλτιστο fold παρατηρείται ότι παρόλη την άνιση κατανομή των κλάσεων στο αρχικό σύνολο δεδομένων, το μοντέλο προβλέπει με ακρίβεια άνω του ~70% όλες τις κλάσεις εκτός από τις ‘unknown’ (όπου η χαμηλή ακρίβεια ταξινόμησης είναι λογική), ‘Meadow’ (ο ταξινομητής αποτυγχάνει να αναγνωρίσει ορισμένα από τα χαρακτηριστικά της συγκεκριμένης κλάσης και επομένως κάνει αρκετές λανθασμένες προβλέψεις στην κλάση ‘unknown’) και ‘winter rye’ (ενδεχομένως λόγω όμοιων χαρακτηριστικών με την κλάση ‘winter wheat’ να δημιουργείται “σύγχυση” στον ταξινομητή).



Σχήμα 1. Πίνακας αντιστοίχισης δείκτη σε κλάση και πίνακα σύγχυσης για το fold το οποίο δίνει τα βέλτιστα αποτελέσματα

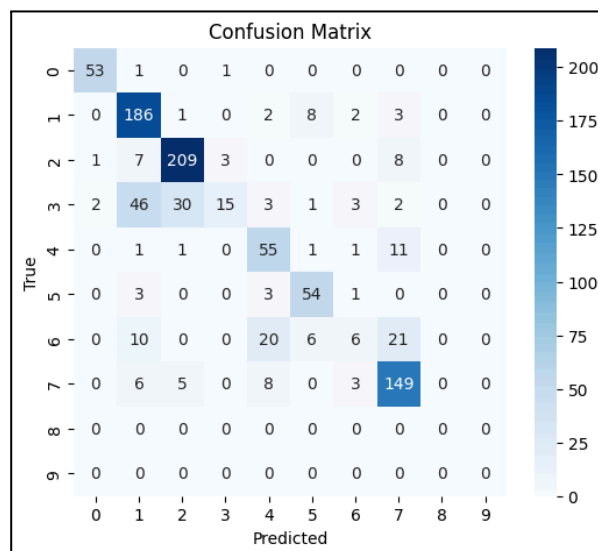
Στο παρακάτω σχήμα (πάλι για το fold το οποίο δίνει τα βέλτιστα αποτελέσματα) παρατίθενται τα διαγράμματα στα οποία αποτυπώνεται η απώλεια και η ακρίβεια ταξινόμησης ανά εποχή για τα σύνολα εκπαίδευσης και επικύρωσης. Σε γενικές γραμμές, παρατηρείται “ελαφρώς” το φαινόμενο της υπερ-προσαρμογής (overfitting) λόγω της σταθερής απόκλισης μεταξύ απώλειας εκπαίδευσης και επικύρωσης όσο αυξάνονται οι εποχές, κάτι το οποίο ενδέχεται να έχει σχέση με τον διαφορετικό τρόπο κανονικοποίησης στις συνιστώσες της αρχιτεκτονικής (π.χ. σε κάποια επίπεδα εφαρμόζεται batch normalization ενώ σε άλλα dropout).



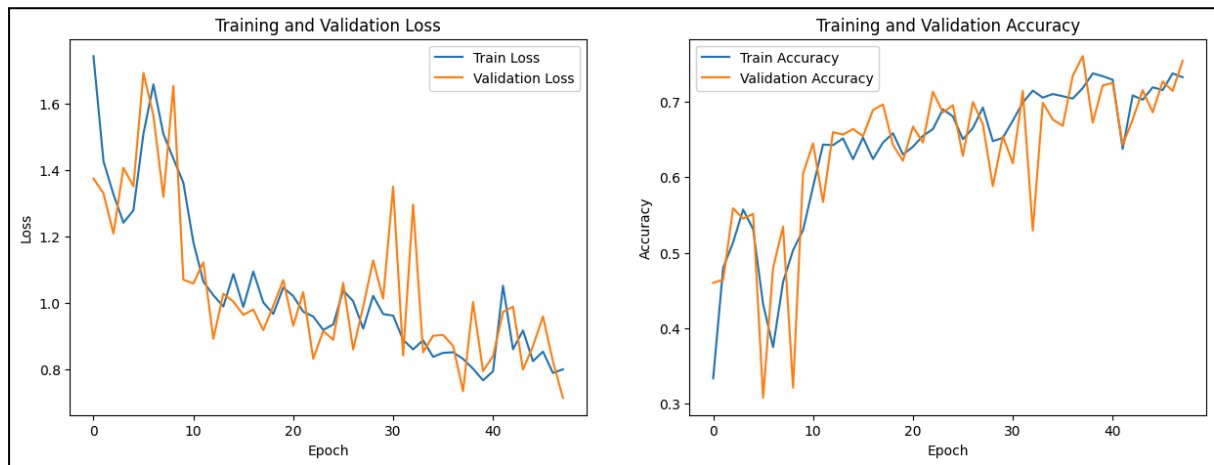
Τέλος, ακολουθούν τα διαγράμματα αποτύπωσης των μετρικών precision (αναλογία true positives προς όλα τα δείγματα που ο ταξινομητής θεώρησε ότι πρόβλεψε σωστά), recall (αναλογία true positives προς true positives και false negatives) και F1 score (αρμονικός μέσος των precision και recall) με μέθοδο averaging micro (εκτίμηση συνολικής απόδοσης χωρίς να λαμβάνονται υπόψη τυχόν ιδιαιτερότητες στην κατανομή των κλάσεων) και weighted (βεβαρημένος μέσος όρος μετρικών ανά κλάση).



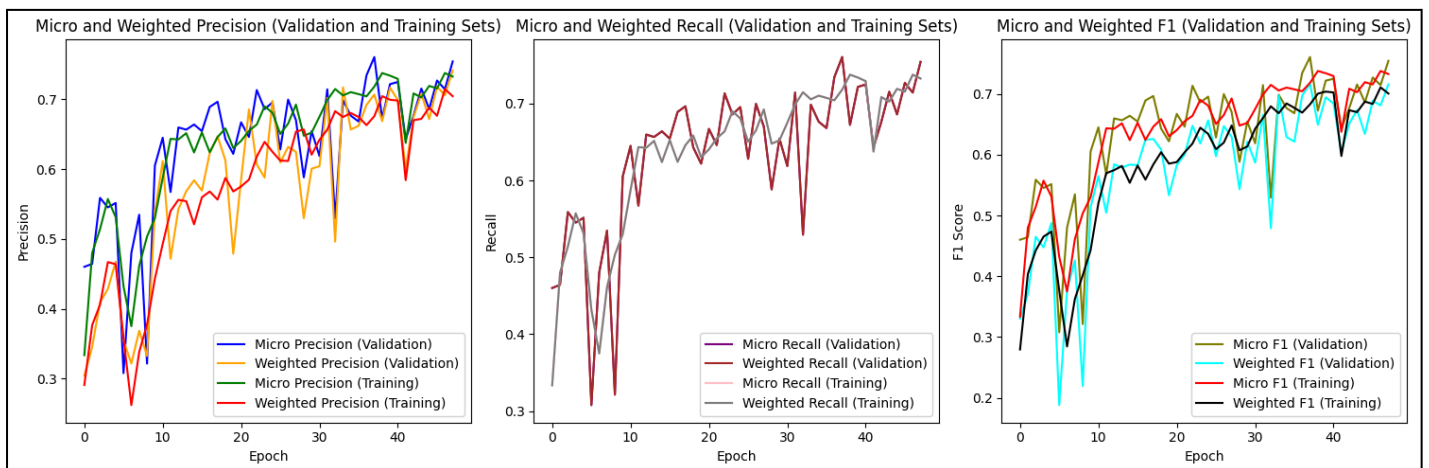
Τέλος, επισημαίνεται ότι κατά την επίλυση της άσκησης έγινε και ένα “τρέξιμο” της transformer αρχιτεκτονικής χωρίς τον μετασχηματισμό mean pooling στην προβολή των queries και εφαρμογή του 2ου πλήρους συνδεδεμένου επιπέδου εντός του temporal attention encoder, με τα αποτελέσματα στις ίδιες μετρικές να είναι αρκετά καλύτερα (μεγαλύτερη ακρίβεια ταξινόμησης και απουσία overfitting).



Σχήμα 4. Πίνακας σύγκρισης για το fold το οποίο δίνει τα βέλτιστα αποτελέσματα και την υλοποίηση χωρίς τον προσδιορισμό του “master query” εντός του temporal attention encoder



Σχήμα 5. Διαγράμματα απώλειας και ακρίβειας ταξινόμησης ανά εποχή στα σύνολα εκπαίδευσης και επικύρωσης για την υλοποίηση χωρίς τον προσδιορισμό του “master query” εντός του temporal attention encoder



Σχήμα 6. Διαγράμματα αποτύπωσης μετρικών precision, recall και F1 score ανά εποχή εκπαίδευσης για την υλοποίηση χωρίς τον προσδιορισμό του “master query” εντός του temporal attention encoder

Παραπομπές

- [1]: Vivien Sainte Fare Garnot and Loic Landrieu and Sebastien Giordano and Nesrine Chehata, Satellite Image Time Series Classification with Pixel-Set Encoders and Temporal Self-Attention (2019). arXiv:1911.07757
- [2]: Ashish Vaswani and Noam Shazeer and Niki Parmar and Jakob Uszkoreit and Llion Jones and Aidan N. Gomez and Lukasz Kaiser and Illia Polosukhin, Attention Is All You Need (2017). arXiv:1706.03762