



Ταξινόμηση σκηνών σε εναέριες εικόνες με χρήση τεχνικών μηχανικής μάθησης

1. Εισαγωγή

Η παρούσα συνοπτική έκθεση συντάχθηκε στα πλαίσια Διπλωματικής Εργασίας της Σχολής Αγρονόμων και Τοπογράφων Μηχανικών – Μηχανικών Γεωπληροφορικής του ΕΜΠ με πιθανό θέμα «Ταξινόμηση σκηνών σε εναέριες εικόνες με χρήση τεχνικών μηχανικής μάθησης».

Σκοπός της παρούσας μελέτης ήταν η εφαρμογή τεχνικών μηχανικής μάθησης για την ταξινόμηση σκηνών που σχετίζονται με υποδομές και κρίσιμες εγκαταστάσεις, όπως αεροδρόμια, γέφυρες, λιμάνια και άλλα, καθώς και η επιλογή του βέλτιστου μοντέλου, το οποίο επρόκειτο να αξιολογηθεί ως προς τη δυνατότητα γενίκευσης σε άλλο σετ δεδομένων.

2. Μεθοδολογία

2.1 Δεδομένα εκπαίδευσης – επικύρωσης

Για την εκπαίδευση (train) και επικύρωση (validation) αξιοποιήθηκε μέρος του σετ δεδομένων *Aerial Image Dataset* (AID – [Λήψη](#)) και συγκεκριμένα 8 κλάσεις που αφορούσαν σε υποδομές και κρίσιμες εγκαταστάσεις, ενώ για κάθε κλάση το σύνολο των εικόνων μοιράστηκε σε 80% για την εκπαίδευση και σε 20% για την επικύρωση.

AID Classes – Train Images:

Airport	Bridge	Center	Industrial	Port	RailwayStation	StorageTanks	Viaduct	Total
288	288	208	312	304	208	288	336	2232

AID Classes – Validation Images:

Airport	Bridge	Center	Industrial	Port	RailwayStation	StorageTanks	Viaduct	Total
72	72	52	78	76	52	72	84	558

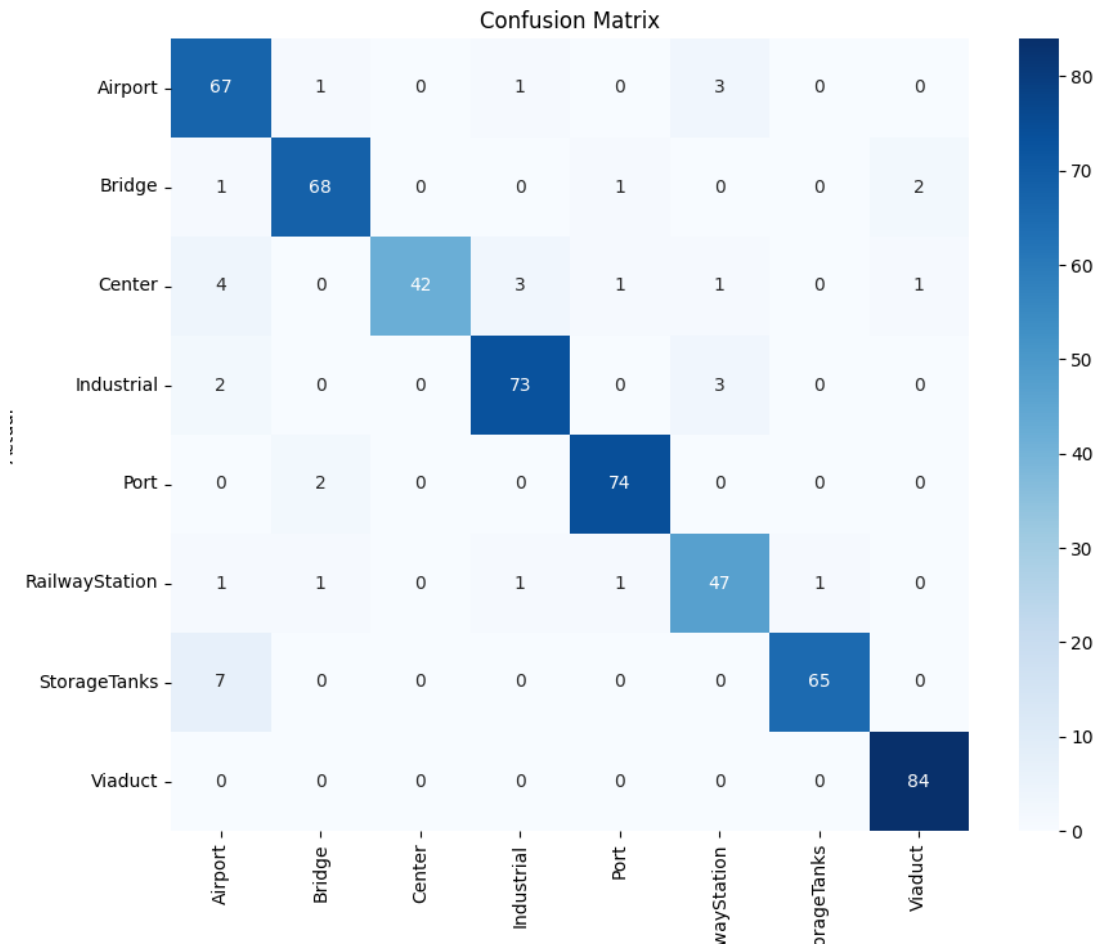
2.2 Μοντέλα μηχανικής μάθησης

Για τις ανάγκες της μελέτης αναπτύχθηκε κώδικας σε γλώσσα Python αξιοποιώντας τις βιβλιοθήκες torch και torchvision του Google Colab, ο οποίος βρίσκεται στο ακόλουθο αποθετήριο <https://github.com/lewisdoukas/scene-classification>.

Η επιλογή του βέλτιστου μοντέλου πραγματοποιήθηκε μετά από δοκιμές συνελκτικών νευρωνικών δικτύων, visual transformers και υβριδικών μοντέλων ως συνδυασμός των ανωτέρω. Συγκεκριμένα χρησιμοποιήθηκε το μοντέλο VGG-16 για 10 και 20 εποχές χωρίς αξιόλογο αποτέλεσμα (περίπου 17% accuracy – δεν αποθηκεύτηκε), το μοντέλο ResNet-50 για 10 και 20 εποχές, ο visual transformer ViT για 10 εποχές και υβριδικό μοντέλο ResNet-50 + ViT για 10 εποχές. Στους Πίνακες 1 – 3 παρουσιάζονται οι μετρικές των παραπάνω μοντέλων και στις Εικόνες 1 – 3 οι αντίστοιχοι πίνακες σύγχυσης:

Πίνακας 1. ResNet-50 Metrics

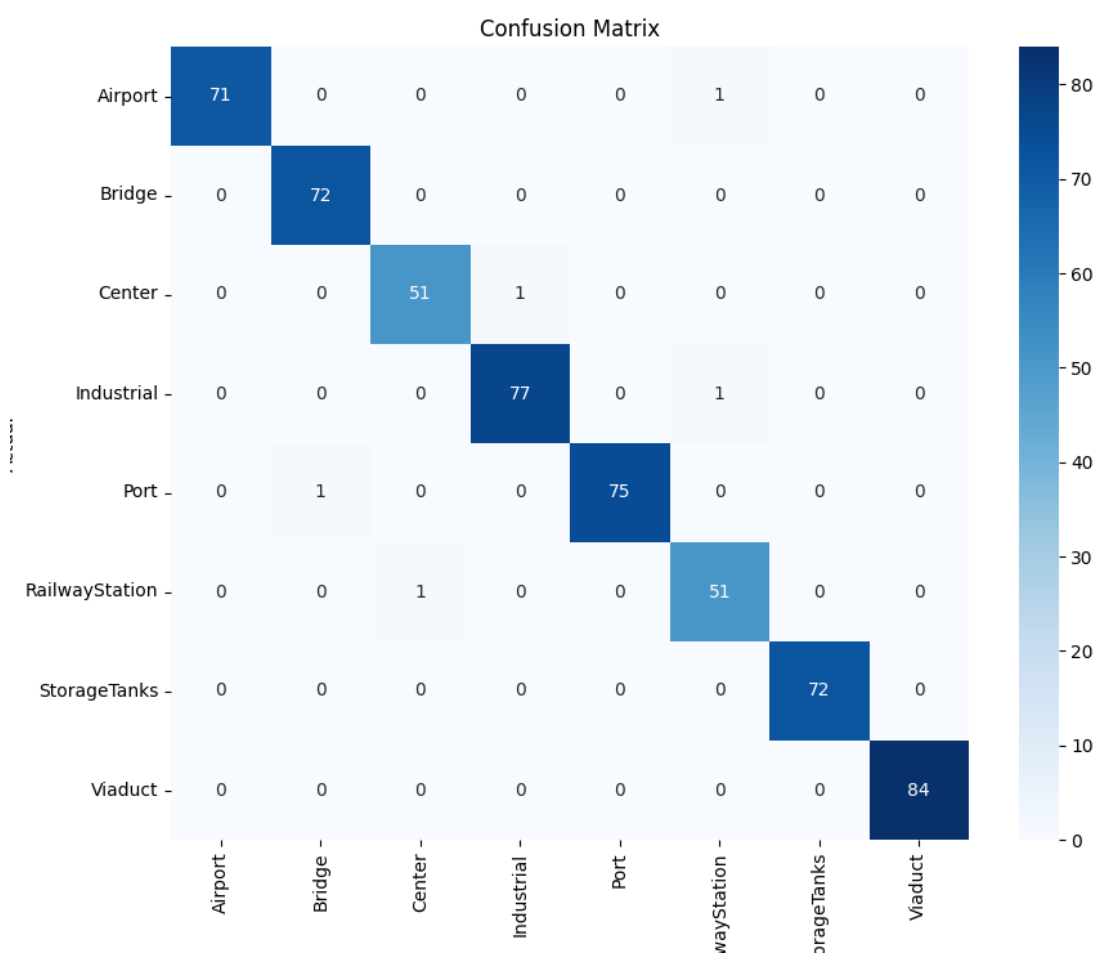
Epoch	Train		Validation			
	Loss	Accuracy	Accuracy	Precision	Recall	F1-score
1	0.9324	0.7003	0.7401	0.7729	0.7237	0.7186
2	0.5447	0.8257	0.8459	0.8543	0.8485	0.8407
3	0.5030	0.8401	0.7599	0.7933	0.7487	0.7499
4	0.4230	0.8683	0.8728	0.8785	0.8776	0.8645
5	0.3583	0.8875	0.8674	0.8674	0.8579	0.8588
6	0.2910	0.8961	0.9176	0.9216	0.9095	0.9124
7	0.2272	0.9292	0.9050	0.9092	0.8992	0.8982
8	0.2716	0.9091	0.7133	0.7971	0.7174	0.7130
9	0.2833	0.9176	0.9319	0.9349	0.9249	0.9278
10	0.2003	0.9306	0.9194	0.9184	0.9140	0.9137



Εικόνα 1. Πίνακας σύγχυσης ResNet-50

Πίνακας 2. ViT Metrics

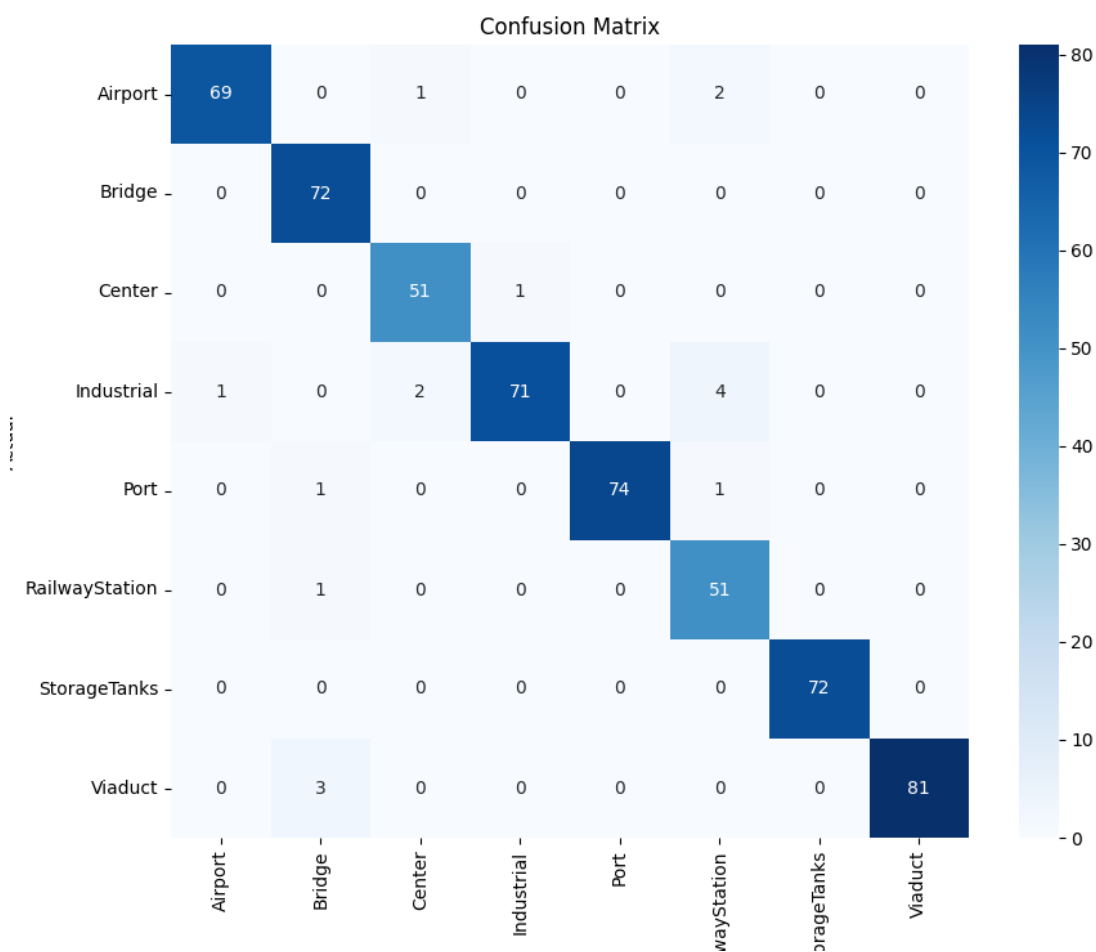
Epoch	Train		Validation			
	Loss	Accuracy	Accuracy	Precision	Recall	F1-score
1	0.3561	0.8840	0.9749	0.9736	0.9738	0.9733
2	0.0296	0.9915	0.9767	0.9750	0.9765	0.9752
3	0.0060	0.9987	0.9624	0.9585	0.9654	0.9596
4	0.0129	0.9964	0.9785	0.9751	0.9773	0.9757
5	0.0493	0.9843	0.9570	0.9558	0.9584	0.9552
6	0.0400	0.9888	0.9677	0.9712	0.9640	0.9666
7	0.0221	0.9933	0.9875	0.9882	0.9859	0.9869
8	0.0078	0.9964	0.9839	0.9836	0.9839	0.9835
9	0.0008	1.0000	0.9857	0.9834	0.9860	0.9845
10	0.0002	1.0000	0.9910	0.9896	0.9902	0.9899



Εικόνα 2. Πίνακας σύγχυσης ViT

Πίνακας 3. Hybrid (ResNet-50 + ViT) Metrics

Epoch	Train		Validation			
	Loss	Accuracy	Accuracy	Precision	Recall	F1-score
1	0.3280	0.9005	0.9695	0.9663	0.9710	0.9679
2	0.0228	0.9946	0.9659	0.9635	0.9680	0.9641
3	0.0211	0.9942	0.9803	0.9797	0.9775	0.9785
4	0.0080	0.9982	0.9785	0.9790	0.9763	0.9770
5	0.0041	0.9996	0.9892	0.9884	0.9871	0.9877
6	0.0150	0.9960	0.9910	0.9908	0.9903	0.9905
7	0.0034	0.9996	0.9839	0.9827	0.9820	0.9822
8	0.0013	0.9996	0.9982	0.9984	0.9976	0.9980
9	0.0002	1.0000	0.9982	0.9984	0.9976	0.9980
10	0.0002	1.0000	0.9982	0.9976	0.9983	0.9979



Εικόνα 3. Πίνακας σύγχυσης Hybrid (ResNet-50 + ViT)

Από τα παραπάνω, ως βέλτιστο μοντέλο θεωρήθηκε το υβριδικό μοντέλο το οποίο και αξιολογήθηκε ως προς τη δυνατότητα γενίκευσης σε άλλο σετ δεδομένων.

2.3 Αξιολόγηση γενίκευσης μοντέλου

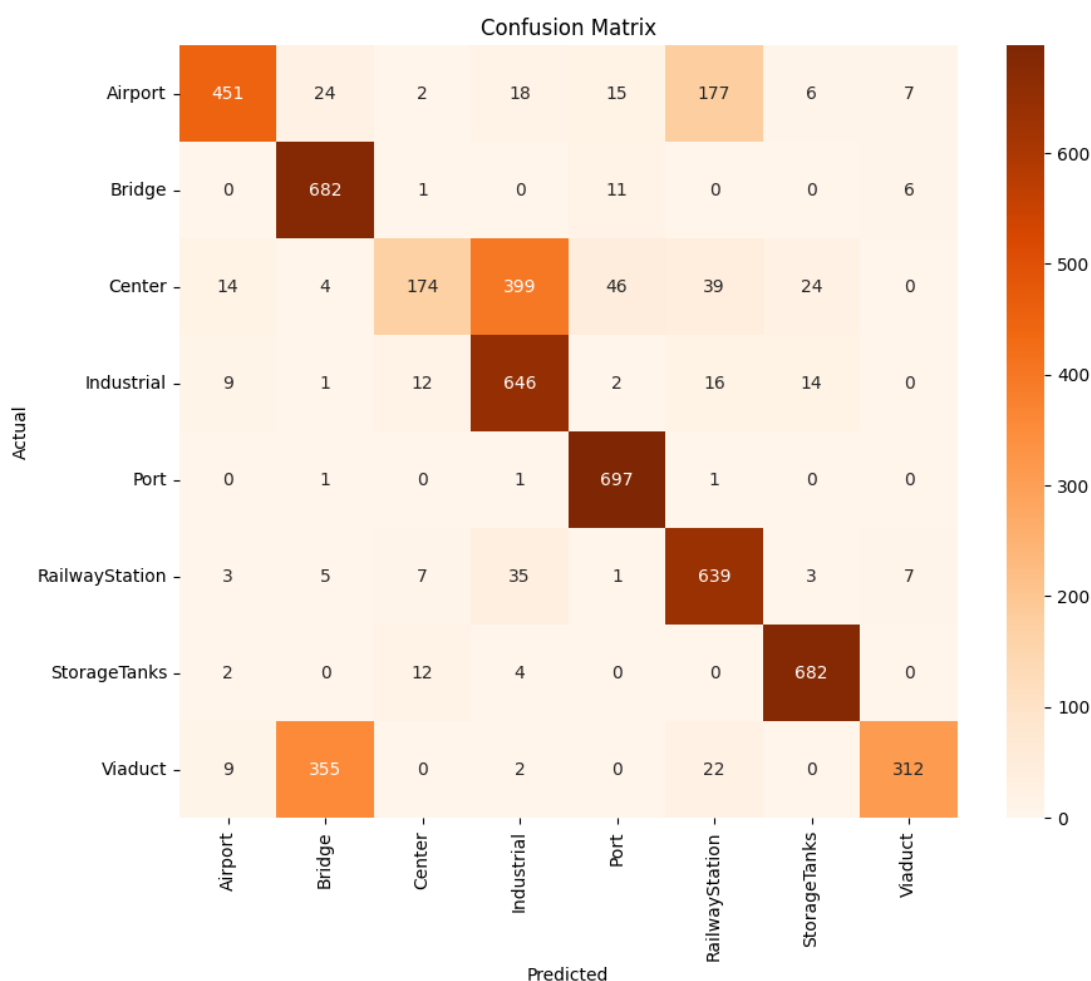
Η αξιολόγηση της δυνατότητας γενίκευσης του υβριδικού μοντέλου πραγματοποιήθηκε με εφαρμογή στο σετ *NWPU-RESISC45* ([Λήψη](#)) και συγκεκριμένα σε 8 κοινές ή παρόμοιες κλάσεις (σύνολο $8 \times 700 = 5600$ εικόνες):

AID Class	NWPU-RESISC45 Class
Airport	Airport
Bridge	Bridge
Center	Commercial Area
Industrial	Industrial Area
Port	Harbor
RailwayStation	Railway Station
StorageTanks	Storage Tank
Viaduct	Overpass

Η κλάση Center δεν ανταποκρινόταν σημασιολογικά ακριβώς με την κλάση Commercial Area, όπως αντίστοιχα η κλάση Viaduct με την Overpass, όμως αξιοποιήθηκαν ως κοντινότερες, γεγονός που επιβεβαιώθηκε και από τον πίνακα σύγχυσης (Εικόνα 4):

Hybrid – NWPU-RESISC45 Metrics

Accuracy	0.7648
Precision	0.8093
Recall	0.7648
F1-score	0.7420



Εικόνα 4. Πίνακας σύγχυσης Hybrid (ResNet-50 + ViT) – NWPU-RESISC45

Συμπεράσματα – προεκτάσεις

Από την παραπάνω μελέτη συμπεραίνεται ότι και τα τρία μοντέλα που εφαρμόστηκαν ανταποκρίνονται σε πολύ ικανοποιητικό επίπεδο για τον σκοπό της ταξινόμησης σκηνών υποδομών και κρίσιμων εγκαταστάσεων, με το υβριδικό μοντέλο να προηγείται.

Κατά την αξιολόγηση της γενίκευσης του υβριδικού μοντέλου στο σετ *NWPU-RESISC45* διαπιστώθηκε εσφαλμένη ταξινόμηση των κλάσεων Viaduct, Center και Airport, γεγονός που οφείλεται στη διαφοροποίηση τόσο της σημασιολογίας όσο και στις ίδιες τις εικόνες αναφοράς.

Ως πιθανή προέκταση θα μπορούσε να αντιστραφεί το πρόβλημα, δηλαδή να εκπαιδευτούν τα αντίστοιχα μοντέλα πάνω στο σετ *NWPU-RESISC45* το οποίο είχε μεγαλύτερο αριθμό εικόνων και να δοκιμαστεί στο *AID*, ή και να αφαιρεθούν οι κλάσεις που δεν έχουν ακριβώς την ίδια σημασιολογία.