



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Αυτόματη Περίληψη Φυτοπαθολογιών με Χρήση Πολυτροπικών Μεγάλων Γλωσσικών Μοντέλων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΓΕΩΡΓΙΟΥ Β. ΦΡΑΓΚΟΓΙΑΝΝΗ

Επιτροπή : Καθ. Μιχαήλ Γ. Λαγουδάκης
Καθ. Θρασύβουλος Σπυρόπουλος
Δρ. Βασίλειος Διαχολουκάς

Χανιά, Οκτώβριος 2025



Πολυτεχνείο Κρήτης
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Αυτόματη Περίληψη Φυτοπαθολογιών με Χρήση Πολυτροπικών Μεγάλων Γλωσσικών Μοντέλων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΓΕΩΡΓΙΟΥ Β. ΦΡΑΓΚΟΓΙΑΝΝΗ

Επιβλέπων: Μιχαήλ Γ. Λαγουδάκης

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 08/10/2025.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....

Μιχαήλ Γ. Λαγουδάκης

.....

Θρασύβουλος Σπυρόπουλος

.....

Βασίλειος Διακολουκάς

Χανιά, Οκτώβριος 2025



Πολυτεχνείο Κρήτης
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Copyright ©–All rights reserved Γεώργιος Φραγκογιάννης, 2025.

Με την επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα.

Το περιεχόμενο αυτής της εργασίας δεν απηχεί απαραίτητα τις απόψεις της Σχολής, του Επιβλέποντα, ή της επιτροπής που την ενέκρινε.

Υπεύθυνη Δήλωση

Βεβαιώνω ότι είμαι συγγραφέας της παρούσας διπλωματικής εργασίας, και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην διπλωματική εργασία. Επίσης έχω αναφέρει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επίσης, βεβαιώνω ότι η παρούσα διπλωματική εργασία προετοιμάστηκε από εμένα προσωπικά ειδικά για τις απαιτήσεις του προγράμματος σπουδών της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Πολυτεχνείου Κρήτης.

(Υπογραφή)

.....

Γεώργιος Φραγκογιάννης



TECHNICAL UNIVERSITY OF CRETE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

Automatic Summarization of Phytopathologies Using Multimodal Large Language Models

DIPLOMA THESIS

by

GEORGIOS V. FRAGKOGIANNIS

Committee : Prof. Michail G. Lagoudakis
Prof. Thrasyvoulos Spyropoulos
Dr. Vasileios Diakouloukas

Chania, October 2025

Περίληψη

Η σύγχρονη γεωργία καλείται να ανταποκριθεί στις αυξανόμενες ανάγκες για παραγωγή τροφής, διατηρώντας παράλληλα τη βιωσιμότητα και την αποδοτικότητά της. Η πρόκληση αυτή καθιστά απαραίτητη την αξιοποίηση προηγμένων τεχνολογιών, όπως η υπολογιστική όραση και τα σύγχρονα μεγάλα γλωσσικά μοντέλα, ιδιαίτερα σε πεδία όπου η έγκαιρη και ακριβής διάγνωση φυτοπαθολογιών μπορεί να μειώσει σημαντικά, τόσο τις απώλειες στην παραγωγή, όσο και τη χρήση φυτοφαρμάκων. Στη παρούσα διπλωματική εργασία, επικεντρωνόμαστε στην ανάπτυξη ενός συστήματος ικανού να αναγνωρίζει ασθένειες που προσβάλλουν τα φυτά, με βάση φωτογραφίες των φύλλων τους και να προτείνει τρόπους αντιμετώπισης. Η προτεινόμενη αρχιτεκτονική αποτελείται από τρία βασικά στάδια. Αρχικά, αξιοποιείται μια σειρά από οπτικά μοντέλα για τον εντοπισμό του φύλλου, την ανίχνευση περιοχών με ενδείξεις προσβολής και την ταξινόμηση της πιθανής ασθένειας. Στο επόμενο στάδιο, ένα σύστημα tool calling επιτρέπει σε ένα γλωσσικό μοντέλο να επιλέξει και να ενεργοποιήσει τα κατάλληλα εργαλεία, εμπλουτίζοντας τη διαγνωστική διαδικασία. Τέλος, το σύστημα συνθέτει μία τεκμηριωμένη και ολοκληρωμένη απάντηση, στην οποία παρουσιάζονται, τόσο η αναγνωρισμένη ασθένεια, όσο και οι προτεινόμενοι τρόποι αντιμετώπισής της. Κατά τη διάρκεια της πειραματικής διαδικασίας, χρησιμοποιήθηκαν μοντέλα YOLO για την ανίχνευση και ταξινόμηση εικόνων, τα οποία εκπαιδεύτηκαν σε ειδικά διαμορφωμένα σύνολα δεδομένων, βασισμένα σε αναγνωρισμένα δημόσια datasets του χώρου της έξυπνης γεωργίας, όπως τα PlantVillage και PlantDoc. Παράλληλα, αναπτύχθηκαν συστήματα Retrieval-Augmented Generation (RAG) για την ανάκτηση πληροφοριών σχετικά με μεθόδους αντιμετώπισης των ασθενειών, ενώ πραγματοποιήθηκε και fine-tuning σε vision-language models για την ενίσχυση της ακρίβειας στην αναγνώριση φυτοπαθολογιών.

Λέξεις Κλειδιά

Έξυπνη γεωργία, υπολογιστική όραση, μεγάλα γλωσσικά μοντέλα, ανίχνευση φυτοπαθολογιών, YOLO, vision-language models, fine-tuning, tool calling, retrieval-augmented generation, PlantVillage, PlantDoc, διάγνωση φυτοπαθολογικών ασθενειών, βιώσιμη γεωργία, αυτόματη αναγνώριση εικόνων, αντιμετώπιση ασθενειών φυτών

Abstract

Modern agriculture is called to respond to the increasing needs for food production, while at the same time maintaining its sustainability and efficiency. This challenge makes necessary the exploitation of advanced technologies, such as computer vision and modern large language models, especially in fields where timely and accurate diagnosis of phytopathologies can significantly reduce both production losses and the use of pesticides. In the present diploma thesis, we focus on the development of a system capable of recognizing diseases that affect plants, based on pictures of their leaves. The proposed architecture consists of three basic stages. Initially, a series of visual models is utilized for the detection of the leaf, the identification of areas with signs of infection, and the classification of the possible disease. In the next stage, a tool calling system allows a language model to select and activate the appropriate tools, enriching the diagnostic process. Finally, the system composes a documented and complete answer, in which both the recognized disease and the proposed ways of addressing it are presented. During the experimental process, YOLO models were used for the detection and classification of images, which were trained on specially configured datasets, based on recognized public datasets in the field of smart agriculture, such as PlantVillage and PlantDoc. At the same time, Retrieval-Augmented Generation (RAG) systems were developed for the retrieval of information regarding methods for dealing with the diseases, while fine-tuning was also carried out on vision-language models to enhance accuracy in the recognition of phytopathologies.

Keywords

Smart agriculture, computer vision, large language models, detection of phytopathologies, YOLO, vision-language models, fine-tuning, tool calling, retrieval-augmented generation, PlantVillage, PlantDoc, plant disease diagnosis, sustainable agriculture, automated image recognition, plant disease treatment

Ευχαριστίες

Αυτή η εργασία σηματοδοτεί το τέλος του ακαδημαϊκού μου κύκλου ως προπτυχιακός φοιτητής στο Πολυτεχνείο Κρήτης. Με αυτή την αφορμή θέλω να ευχαριστήσω όλους όσους με βοήθησαν σε αυτήν την πορεία.

Αρχικά, θα ήθελα να ευχαριστήσω τον κ. Λαγουδάκη για τη σημαντική του συνεισφορά στη σύνθεση της παρούσας διπλωματικής εργασίας. Ιδιαίτερη ευγνωμοσύνη οφείλω και στον κ. Διακολουκά, η καθοδήγηση και η στήριξη του οποίου υπήρξαν καθοριστικές για την ολοκλήρωση του έργου αυτού.

Στη συνέχεια θέλω να ευχαριστήσω τους συναδέλφους και πάνω απ' όλα φίλους μου από τα Χανιά. Ήταν εκεί στις ατελείωτες νύχτες και στις απαιτητικές περιόδους των σπουδών μας, και η παρουσία τους έκανε το ταξίδι αυτό πολύτιμο.

Σημαντική αναφορά οφείλω να κάνω και στου φίλους μου στον Πειραιά, που στάθηκαν δίπλα μου όλα αυτά τα χρόνια, αλλά και σε ένα ιδιαίτερο πρόσωπο στην Πάτρα με το οποίο μας χώριζε πάντα η απόσταση, αλλά ποτέ η παρουσία.

Τέλος, θέλω να ευχαριστήσω τους γονείς μου, για όλες τις θυσίες που έκαναν και την ώθηση που μου έδωσαν για να μπορέσω να φτάσω σε αυτό το σημείο.

Για τον Βασίλη και τη Δάμν

Περιεχόμενα

Περίληψη	i
Abstract	iii
Ευχαριστίες	v
Περιεχόμενα	xi
Κατάλογος Σχημάτων	xiv
Κατάλογος Πινάκων	xv
1 Εισαγωγή	1
1.1 Προκλήσεις στην Ταξινόμηση Φυτοπαθολογιών	1
1.2 Αντικείμενο της Διπλωματικής Εργασίας	3
1.3 Σχετική Έρευνα στην Έξυπνη Γεωργία και την Αναγνώριση Φυτοπαθολογιών	3
1.4 Καινοτομία και Συνεισφορά της Διπλωματικής Εργασίας	4
1.5 Δομή της Διπλωματικής Εργασίας	5
2 Θεωρητικό Υπόβαθρο	7
2.1 Βασικοί Ορισμοί	7
2.1.1 Τεχνητή Νοημοσύνη	7
2.1.2 Μηχανική Μάθηση	7
2.1.3 Βαθιά Μάθηση	8
2.2 Convolutional Neural Network (CNN)	8
2.2.1 Convolution Layer	8
2.2.2 Pooling Layer	9
2.2.3 Fully Connected Layer	9
2.2.4 Loss Functions	10
2.2.5 Activation	10
2.2.6 Optimizer	11
2.2.7 Hyperparameters	12
2.3 Μοντέλα Όρασης	13

2.4	YOLO	13
2.4.1	Εισαγωγή στο YOLO	13
2.4.2	Εκπαίδευση του YOLOv11	14
2.5	Μεγάλα Γλωσσικά Μοντέλα	15
2.5.1	Ιστορική Αναδρομή	15
2.5.2	Tokenizer	15
2.5.3	Μηχανισμός Attention	16
2.5.4	Embeddings και Κωδικοποιήσεις Θέσης	17
2.5.5	Fine-tuning με χρήση LoRA Adapters	17
2.5.6	Βελτιστοποιήσεις Υπολογιστικής Απόδοσης	18
2.5.7	Τεχνικές Prompting	19
2.6	Πολυτροπικά Μεγάλα Γλωσσικά Μοντέλα	20
2.7	Retrieval-Augmented Generation (RAG)	20
2.7.1	Πλήθος Ανακτήσεων (top- k)	21
2.7.2	Μοντέλο Embedding	21
2.7.3	Μέγεθος Context Window	21
2.7.4	Μέγεθος Τμημάτων Κειμένου	21
2.7.5	Επικάλυψη Μεταξύ Τμημάτων Κειμένου	21
2.8	Αρχιτεκτονική Πολυτροπικών Μοντέλων	21
2.8.1	Vision Transformer (ViT)	22
2.9	Ροή Επεξεργασίας με LangGraph	22
2.9.1	Η Λογική του State Machine	22
2.9.2	Πλεονεκτήματα για την Παρούσα Εργασία	23
2.10	Αναγνώριση Φυτοπαθολογιών	23
2.11	Ψηφιακές Πλατφόρμες και Εργαλεία Υλοποίησης	24
3	Μοντελοποίηση Συστήματος	25
3.1	Vision-Guided Retrieval System (VGRS)	25
3.1.1	Ανάλυση Εικόνας	25
3.1.2	Ανάκτηση Πληροφοριών από τη Βάση Δεδομένων	26
3.1.3	Ενοποίηση μέσω Μεγάλου Γλωσσικού Μοντέλου	26
3.2	Πολυτροπική Προσέγγιση – Multimodal Expert System (MES)	27
3.2.1	Οπτική Αναγνώριση Χαρακτηριστικών	28
3.2.2	Tool Calling Μηχανισμός	28
3.2.3	Σύνθεση Απάντησης	29
3.3	Τελική Προσέγγιση - Vision Enhanced Analysis Agent (VEAA)	29
4	Πειραματικό Μέρος	31
4.1	Μετρικές Αξιολόγησης	31
4.1.1	Καμπύλη Precision–Recall	31
4.1.2	IoU (Intersection over Union)	32

4.1.3	mAP@0.5	32
4.1.4	mAP@0.5:0.95	33
4.1.5	Ο Πίνακας Confusion Matrix	33
4.1.6	F1 Score	33
4.1.7	Context Precision	34
4.1.8	Context Recall με χρήση LLM	34
4.1.9	Faithfulness	35
4.1.10	Answer Relevance	35
4.2	Πειράματα VGRS	36
4.2.1	Πρώτα Βήματα	36
4.2.2	Εκπαίδευση Ταξινόμησης στο PlantVillage Dataset	37
4.2.3	Αξιολόγηση με χρήση του PlantDoc Dataset	41
4.2.4	Ανίχνευση Ασθενειών με Βάση τις Προσβεβλημένες Περιοχές	45
4.2.5	Εκπαίδευση ανά Τύπο Φυτού	48
4.2.6	Δημιουργία RAG Συστήματος	54
4.2.7	Αξιολόγηση του VGRS Pipeline με RAGAS	57
4.3	Πειράματα MES	65
4.3.1	Επιλογή Μοντέλου	65
4.3.2	Fine Tuning Vision-LLM με χρήση Image-to-Text Dataset	66
4.3.3	Fine Tuning Vision-LLM με χρήση PlantVillage VQA	68
4.3.4	Ενσωμάτωση Web Scraper (Tavily) με tool calling	71
4.3.5	Αξιολόγηση MES	72
4.4	Σύγκριση των Συστημάτων VGRS και MES	77
4.5	Πειράματα VEAA	77
4.5.1	Υλοποίηση του Συστήματος με LangGraph	77
4.5.2	Παράδειγμα Λειτουργίας του Συστήματος	81
4.6	Αξιολόγηση του Ενοποιημένου Συστήματος	83
5	Επίλογος	85
5.1	Συμπεράσματα	85
5.2	Περιορισμοί	86
5.3	Μελλοντικές Επεκτάσεις	86

Κατάλογος Σχημάτων

1.1	Δείγματα από σύνολο δεδομένων με ταξινομημένα άρρωστα φύλλα, των οποίων οι φωτογραφίες έχουν ληφθεί σε ελεγχόμενο περιβάλλον.	2
2.1	Απλοποιημένη μορφή συνελικτικού νευρωνικού δικτύου	9
2.2	Απεικόνιση δύο διαφορετικών διαδικασιών pooling (max και average). (Πηγή: ACM Digital library)	10
2.3	Η απόδοση του Yolov11 σε σχέση με τις προηγούμενες εκδόσεις (Πηγή: Ultralytics YOLO Docs)	14
2.4	Η αρχιτεκτονική του σύγχρονου LLM (Πηγή: GeeksForGeeks)	16
2.5	Η λειτουργία του tokenizer (Πηγή: OpenVINO Docs)	16
2.6	LoRA Adapter	18
2.7	Αρχιτεκτονική του Retrieval-Augmented Generation (RAG).	20
3.1	Αρχιτεκτονική του Vision-Guided Retrieval System (VGRS)	27
3.2	Flowchart για τη λειτουργία του MES συστήματος	28
3.3	Αρχιτεκτονική τελικού συστήματος	29
4.1	Παραδείγματα Εικόνων από το PlantVillage dataset	37
4.2	Ultralytics YOLO Dataset Format για Object Detection	38
4.3	Λίστα διαθέσιμων YOLOv11 μοντέλων	39
4.4	Αποτελέσματα εκπαίδευσης Object Detection πάνω σε ολόκληρο το PlantVillage Dataset	40
4.5	Παραδείγματα αναγνώρισης φύλλου και ασθένειας	40
4.6	Γραφήματα Αξιολόγησης στο PlantVillage Dataset	42
4.7	Δείγματα από διάφορες κατηγορίες του συνόλου δεδομένων PlantDoc δείχνουν το χάσμα μεταξύ εργαστηριακά ελεγχόμενων και πραγματικών εικόνων	43
4.8	Παραδείγματα αναγνώρισης φύλλου και ασθένειας στο PlantDoc Dataset	43
4.9	Γραφήματα Αξιολόγησης στο PlantDoc Dataset	44
4.10	Δείγμα εικόνας από το σετ δεδομένων στο Roboflow	46
4.11	Παραδείγματα αναγνώρισης φύλλου και ασθένειας στο apple_leaf_detection_dataset	46
4.12	Γραφήματα αξιολόγησης με ανίχνευση στις άρρωστες περιοχές	47
4.13	Γραφήματα Αξιολόγησης Ανίχνευσης Τύπου Φυτού	51

4.14 Γραφήματα Αξιολόγησης Ανίχνευσης Τύπου Φυτού μετά απο εκπαίδευση με υβριδικό dataset	52
4.15 Παραδείγματα λανθασμένης αναγνώρισης	53
4.16 Αποτελέσματα Αξιολόγησης Ταξινόμησης ύπαρξης Φυτού	54
4.17 Confusion Matrix του τελικού Yolo Pipeline	55
4.18 Όψιμος περονόσπορος σε φύλλο τομάτας (Tomato late blight)	59
4.19 Οι 8 καλύτερες θέσεις του OpenVLM Leaderboard για μοντέλα όρασης από 4 έως 10 δις. παραμέτρους	66
4.20 Κατανομή των ερωτήσεων του dataset	70
4.21 Παράδειγμα χρήσης εργαλείου κατά τη διάρκεια της εκτέλεσης	72
4.22 Φύλλο φράουλας με καψαλισιά (Strawberry leaf with leaf scorch)	74
4.23 Η VEAA αρχιτεκτονική υλοποιημένη με LangGraph	78
4.24 Ο ορισμός του κανονισμού πάνω στον οποίο υπολογίζεται η μετρική	80
4.25 Φύλλα αχλαδιάς με σκωρίαση (Rust Damage)	81
4.26 Κλήσεις εργαλείων που κάνει το llm για να παράξει την τελική απάντηση	82

Κατάλογος Πινάκων

2.1	Ενδεικτικές συναρτήσεις ενεργοποίησης και οι μαθηματικοί τους ορισμοί. . . .	11
4.1	Ομαδοποίηση αρχικών κλάσεων του PlantVillage ανά τύπο φυτού	49
4.2	RAG retrieval evaluation for different setups	56
4.3	Παράμετροι ρύθμισης του RAG pipeline	57
4.4	Αποτελέσματα αξιολόγησης του VGRS Pipeline με RAGAS	65
4.5	Κατηγορίες ερωτήσεων (ενοποιημένος πίνακας, χωρίς στήλη επιπέδου)	68
4.6	Αποτελέσματα Αξιολόγησης του Fine-tuned Μοντέλου στο PlantVillage-VQA Test Split	71
4.7	Αποτελέσματα Αξιολόγησης Απαντήσεων του MES με RAGAS	73
4.8	Έξοδος του Yolo Pipeline όπου δεν ανιχνεύεται φύλλο στην εικόνα	82
4.9	Αποτελέσματα αξιολόγησης (μέσοι όροι)	84

Κεφάλαιο 1

Εισαγωγή

Η σύγχρονη γεωργία βρίσκεται σε ένα κρίσιμο σταυροδρόμι, όπου οι αυξανόμενες ανάγκες για παραγωγή τροφής συναντούν την επιτακτική ανάγκη για βιωσιμότητα και αποδοτικότητα. Η αύξηση του παγκόσμιου πληθυσμού και οι επιπτώσεις της κλιματικής αλλαγής δημιουργούν έντονες πιέσεις στα γεωργικά συστήματα, απαιτώντας καινοτόμες προσεγγίσεις που συνδυάζουν υψηλή παραγωγικότητα με μειωμένο περιβαλλοντικό αποτύπωμα. Σε αυτό το πλαίσιο, η αξιοποίηση προηγμένων τεχνολογιών, όπως η υπολογιστική όραση και η τεχνητή νοημοσύνη, αποκτά ολοένα και μεγαλύτερη σημασία.

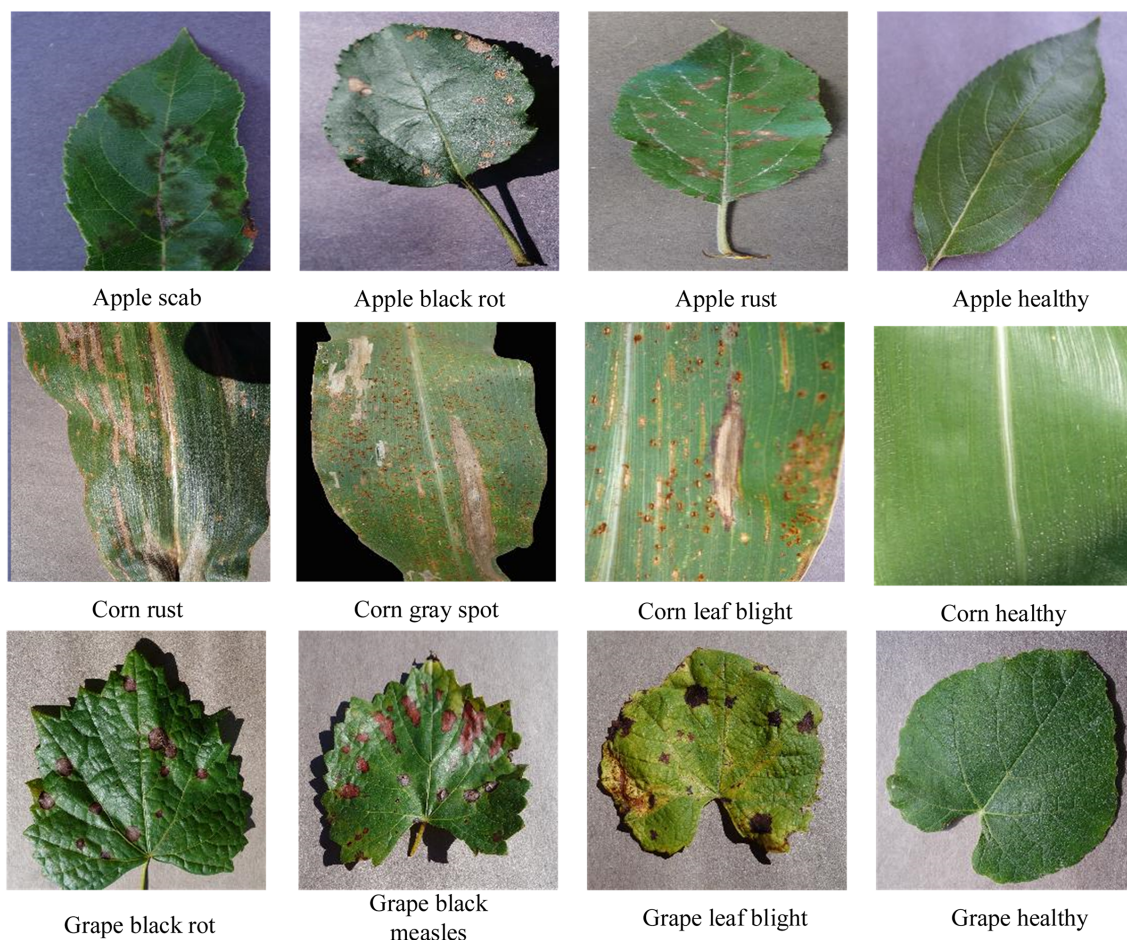
Η υπολογιστική όραση επιτρέπει την αυτοματοποιημένη κατανόηση και ανάλυση του οπτικού περιβάλλοντος των καλλιεργειών, δίνοντας τη δυνατότητα για έγκαιρη διάγνωση φυτοπαθολογιών, παρακολούθηση ανάπτυξης φυτών και εντοπισμό ανωμαλιών με ακρίβεια που συχνά ξεπερνά τις ανθρώπινες δυνατότητες. Η δυνατότητα αυτή ενισχύει τη λήψη έγκαιρων αποφάσεων και συμβάλλει στη μείωση των απωλειών παραγωγής, στην εξοικονόμηση πόρων και στη βελτίωση της ποιότητας των προϊόντων.

Η χρήση προηγμένων μοντέλων οπτικής αναγνώρισης σε συνδυασμό με σύγχρονες μεθόδους ταξινόμησης και ανάκτησης πληροφορίας, προσφέρει πρακτικά εργαλεία για την ενίσχυση της γεωργικής παραγωγής. Μέσα από την αυτόματη επεξεργασία εικόνων, οι αγρότες μπορούν να λαμβάνουν έγκαιρα κρίσιμες πληροφορίες, να μειώνουν τη χρήση φυτοφαρμάκων και να αυξάνουν την αποδοτικότητα των καλλιεργειών τους. Η ενσωμάτωση τέτοιων συστημάτων στην πράξη καθιστά τη γεωργία πιο “έξυπνη”, πιο φιλική προς το περιβάλλον και πιο ανθεκτική στις προκλήσεις της εποχής.

1.1 Προκλήσεις στην Ταξινόμηση Φυτοπαθολογιών

Παρά την πρόοδο στον τομέα της υπολογιστικής όρασης, η ταξινόμηση ασθενειών σε φύλλα φυτών εξακολουθεί να παρουσιάζει σημαντικές προκλήσεις, τόσο σε επίπεδο δεδομένων, όσο και σε επίπεδο γενικευσιμότητας των μοντέλων.

Μία από τις βασικότερες δυσκολίες είναι η περιορισμένη διαθεσιμότητα αξιόπιστων και ποικίλων datasets. Τα περισσότερα δημόσια διαθέσιμα σύνολα δεδομένων περιλαμβάνουν φωτογραφίες που έχουν ληφθεί σε ελεγχόμενα περιβάλλοντα, με σταθερό φωτισμό, ουδέτερο



Σχήμα 1.1: Δείγματα από σύνολο δεδομένων με ταξινομημένα άρρωστα φύλλα, των οποίων οι φωτογραφίες έχουν ληφθεί σε ελεγχόμενο περιβάλλον.

φόντο και καθαρές απεικονίσεις της ασθένειας [Σχήμα 1.1]. Αυτή η συνθήκη, αν και ιδανική για την εκπαίδευση, δεν αντικατοπτρίζει τις πραγματικές συνθήκες πεδίου, όπου ο φωτισμός είναι μεταβαλλόμενος, τα φύλλα μπορεί να είναι μερικώς κρυμμένα ή κατεστραμμένα, και οι εικόνες να λαμβάνονται υπό δύσκολες γωνίες ή με συσκευές χαμηλής ανάλυσης.

Επιπλέον, ακόμη και όταν ένα μοντέλο επιτυγχάνει υψηλή ακρίβεια σε συγκεκριμένο σύνολο δεδομένων, τα αποτελέσματά του συχνά υποβαθμίζονται αισθητά, όταν εφαρμόζεται σε εικόνες από πραγματικές συνθήκες. Το φαινόμενο αυτό αναδεικνύει την έλλειψη γενικευσιμότητας, η οποία αποτελεί κρίσιμο παράγοντα για την πρακτική αξιοποίηση τέτοιων συστημάτων.

Ένα ακόμη ζήτημα είναι ότι η κλασική προσέγγιση της απλής ταξινόμησης εικόνας, αν και χρήσιμη, δεν επαρκεί για να υποστηρίξει ολοκληρωμένες λύσεις. Ο αγρότης δεν χρειάζεται μόνο να γνωρίζει ποια είναι η ασθένεια, αλλά και να λάβει άμεσα προτάσεις αντιμετώπισης προσαρμοσμένες στο είδος της καλλιέργειας, στην ένταση της προσβολής και στις διαθέσιμες πρακτικές.

1.2 Αντικείμενο της Διπλωματικής Εργασίας

Η εμφάνιση ασθενειών στα φύλλα των φυτών αποτελεί βασικό φαινοτυπικό δείκτη για τη διάγνωση φυτοπαθολογικών καταστάσεων και την εκτίμηση της υγείας των φυτών. Οι ασθένειες εκδηλώνονται μέσω μορφολογικών και χρωματικών αλλοιώσεων που οφείλονται σε βιοτικούς παράγοντες, όπως μύκητες, βακτήρια και ιοί, καθώς και σε αβιοτικούς παράγοντες, όπως η έλλειψη θρεπτικών στοιχείων ή οι περιβαλλοντικές καταπονήσεις. Συνηθισμένα συμπτώματα περιλαμβάνουν κηλίδες (νεκρωτικές ή χλωρωτικές), μεταβολές στη χρωματική υφή, παραμορφώσεις του ελάσματος, εξάνθηση ή μάρανση. Οι μορφολογικές μεταβολές μπορεί να σχετίζονται με την αλλοίωση των κυτταρικών τοιχωμάτων ή τη δυσλειτουργία του φωτοσυνθετικού μηχανισμού, ενώ οι χρωματικές αλλαγές αντανακλούν μεταβολές στο περιεχόμενο χλωροφύλλης ή στη συγκέντρωση δευτερογενών μεταβολιτών. Η θεωρητική μελέτη των συμπτωμάτων αυτών, σε συνδυασμό με τεχνικές απεικόνισης και υπολογιστικής ανάλυσης, μπορεί να συμβάλει καθοριστικά στην πρόωμη διάγνωση και στην ανάπτυξη αυτοματοποιημένων συστημάτων εντοπισμού και ταξινόμησης ασθενειών στα φυτά.

Η παρούσα διπλωματική εργασία έχει ως αντικείμενο την ανάπτυξη ενός ολοκληρωμένου συστήματος αναγνώρισης ασθενειών σε φύλλα φυτών και παροχής προτάσεων αντιμετώπισης. Το σύστημα δέχεται ως είσοδο μια εικόνα φύλλου, εντοπίζει την παρουσία και το είδος της ασθένειας χρησιμοποιώντας έναν συνδυασμό μοντέλων, και στη συνέχεια ανακτά από κατάλληλη βάση γνώσης πληροφορίες για τις πιθανές θεραπείες και τα μέτρα πρόληψης.

Ο στόχος της παρούσας διπλωματικής εργασίας είναι η δημιουργία ενός ολοκληρωμένου συστήματος ικανού να αναγνωρίζει την ασθένεια ενός φύλλου και να παρέχει τόσο μια αναλυτική περιγραφή της ασθένειας, όσο και κατάλληλους τρόπους αντιμετώπισής της. Για την επίτευξη του στόχου αυτού, αξιοποιούνται μοντέλα υπολογιστικής όρασης, όπως το YOLO, κλασικά μεγάλα γλωσσικά μοντέλα (Large Language Models – LLMs), πολυτροπικά γλωσσικά μοντέλα όρασης (Vision LLMs), καθώς και συστήματα ανάκτησης γνώσης μέσω επανζημένης παραγωγής (Retrieval-Augmented Generation – RAG).

1.3 Σχετική Έρευνα στην Έξυπνη Γεωργία και την Αναγνώριση Φυτοπαθολογιών

Η αξιοποίηση μεθόδων machine/deep learning στη γεωπονία έχει γνωρίσει ραγδαία ανάπτυξη την τελευταία δεκαετία, με εφαρμογές που εκτείνονται από την παρακολούθηση καλλιεργειών έως την έγκαιρη διάγνωση ασθενειών. Από τις πρώτες ακόμη έρευνες αναδείχθηκε η υπεροχή των Convolutional Neural Networks (CNNs) σε εφαρμογές επεξεργασίας εικόνων, θέτοντας παράλληλα ζητήματα επάρκειας δεδομένων, υπολογιστικής αποδοτικότητας και γενίκευσης σε πραγματικές συνθήκες [56, 47]. Οι πιο πρόσφατες έρευνες εστιάζουν ειδικά στην αναγνώριση ασθενειών και καταγράφουν τη στροφή προς ελαφριές, ταχύτερες και περισσότερο επεξηγηματικές προσεγγίσεις [40, 68].

Στο leaf-based πρόβλημα ταξινόμησης ασθενειών, σημείο αναφοράς αποτελεί το Plant Village, όπου η εκπαίδευση CNNs σε ελεγχόμενα δεδομένα (54.306 εικόνες) πέτυχε εξαιρετικές

αποδόσεις, αποδεικνύοντας τη σκοπιμότητα της προσέγγισης [33]. Ωστόσο, τα αποτελέσματα αυτά περιορίζονται από τη φύση του dataset (ομοιογενής φωτισμός/φόντο), που δεν αντανακλά τον «θόρυβο» του πεδίου. Η κριτική αυτή οδήγησε την κοινότητα σε νέα σύνολα δεδομένων και πειραματικά πρωτόκολλα με μεγαλύτερη ποικιλία συνθηκών.

Για την αντιμετώπιση του domain shift από εργαστηριακές σε in-the-wild εικόνες, προτάθηκε το PlantDoc, το οποίο συγκεντρώνει εικόνες από το διαδίκτυο με ανομοιογενείς συνθήκες λήψης [49]. Παρά το μικρότερο μέγεθός του, χρησιμοποιείται ευρέως ως «γέφυρα» αξιολόγησης σε ρεαλιστικά σενάρια. Πιο πρόσφατα, σύνολα όπως το FieldPlant διευρύνουν τη διασπορά των σκηνών «στο πεδίο» [34], ενώ νέα benchmarks δοκιμάζουν μοντέλα σε συνθήκες πραγματικού κόσμου, τόσο για ταξινόμηση, όσο και για εντοπισμό/τμηματοποίηση (π.χ. PlantWild, PlantSeg) [65, 66].

Η εισαγωγή της υπολογιστικής όρασης στην αγροτική τεχνολογία δεν περιορίστηκε μόνο σε στατικά CNN-based classifiers. Τα τελευταία χρόνια, η έρευνα στράφηκε σε πιο ευέλικτες και ολοκληρωμένες προσεγγίσεις, όπως οι ανιχνευτές αντικειμένων (object detectors) τύπου YOLO και EfficientDet, που επιτρέπουν την εντοπισμό και ταυτόχρονη ταξινόμηση των προσβεβλημένων περιοχών πάνω στο φύλλο. Με αυτόν τον τρόπο, η επεξεργασία δεν περιορίζεται στο σύνολο της επιφάνειας του φύλλου, αλλά επικεντρώνεται στις «πραγματικά ενδιαφέρουσες» περιοχές, αυξάνοντας την ακρίβεια [35].

Παράλληλα, τα τελευταία δύο χρόνια εμφανίζονται Vision-Language Models (VLMs) και Large (Multimodal) Language Models (LMMs/VL-LMs) για διάγνωση και υποστήριξη αποφάσεων στη γεωργία. Ενδεικτικά, το WDLM συνδυάζει ορατές ενδείξεις με περιγραφές/αλυσίδες συλλογισμού για διάγνωση σκωριάσεως στο σιτάρι σε «άγριες» συνθήκες [72]. Σε ευρύτερο, διατομεακό επίπεδο γεωργίας/κτηνοτροφίας, το AgriCLIP προσαρμόζει το CLIP σε αγροτικά image-text ζεύγη μεγάλης κλίμακας, ώστε να βελτιώνεται μέσω μηδενικής/ολιγοδειγματικής (zero-shot/fewshot) μάθησης σε εξειδικευμένα καθήκοντα [36]. Μεθοδολογικά, τα VL-LMs μπορούν να πλαισιωθούν με Retrieval-Augmented Generation (RAG), ώστε να αντλούν επικαιροποιημένες γεωπονικές οδηγίες/πρακτικές και να μειώνουν «παραισθήσεις» [28]. Συνολικά, διαμορφώνεται ένα υβριδικό παράδειγμα: detectors/segmenters για ακριβή εντοπισμό συμπτωμάτων πάνω στο φύλλο και VLMs για ερμηνεία/εξηγησιμότητα και παραγωγή προτάσεων αντιμετώπισης.

Συνολικά, η εξέλιξη της τεχνητής νοημοσύνης στην έξυπνη γεωργία φανερώνει μια πορεία από απλά μοντέλα ταξινόμησης εικόνων προς έξυπνους πολυτροπικούς πράκτορες με δυνατότητες διάγνωσης, επεξήγησης και καθοδήγησης του χρήστη με φυσικό τρόπο. Αυτή η μετάβαση δεν είναι μόνο τεχνική, αλλά και εννοιολογική, καθώς φέρνει τη γεωργική διάγνωση πιο κοντά στο όραμα της πλήρως αυτοματοποιημένης, αλληλεπιδραστικής και βιώσιμης γεωργίας.

1.4 Καινοτομία και Συνεισφορά της Διπλωματικής Εργασίας

Παρόλο που στη βιβλιογραφία υπάρχουν αρκετές εργασίες που έχουν ασχοληθεί με την ταξινόμηση ασθeneιών σε φύλλα, οι περισσότερες περιορίζονται είτε στη χρήση μοντέλων υπο-

λογιστικής όρασης, είτε στη χρήση πολυτροπικών LLMs. Η παρούσα εργασία διαφοροποιείται σημαντικά, καθώς στοχεύει στη διαμόρφωση ενός ενοποιημένου συστήματος υποστήριξης αποφάσεων, το οποίο συνδυάζει τα ισχυρότερα χαρακτηριστικά των μοντέλων YOLO, των LLMs, των Vision LLMs και των τεχνικών επαυξημένης ανάκτησης γνώσης (Retrieval-Augmented Generation – RAG).

Για την επίτευξη αυτής της λειτουργικότητας, αξιοποιείται ένας συνδυασμός πολλαπλών τεχνολογιών, όπως το YOLO για την ανίχνευση οπτικών χαρακτηριστικών, μεγάλα γλωσσικά μοντέλα (LLMs), πολυτροπικά μοντέλα όρασης και γλώσσας (Vision LLMs), καθώς και τεχνικές επαυξημένης ανάκτησης γνώσης (Retrieval-Augmented Generation – RAG). Η σύζευξη αυτών των τεχνολογιών επιτρέπει τη δημιουργία ενός ολοκληρωμένου και πολυτροπικού συστήματος, το οποίο μπορεί όχι μόνο να εντοπίζει με ακρίβεια την ασθένεια, αλλά και να εξηγεί τα συμπτώματα και να προτείνει στοχευμένες ενέργειες αντιμετώπισης. Η προσέγγιση αυτή είναι ιδιαίτερα καινοτόμα και συμβάλλει ουσιαστικά στη σύγχρονη έρευνα στον τομέα της αγροτεχνολογίας.

1.5 Δομή της Διπλωματικής Εργασίας

Η δομή της εργασίας έχει ως εξής:

- **Κεφάλαιο 2** – Ανασκόπηση Βιβλιογραφίας: Παρουσίαση σχετικών μελετών και τεχνολογιών στον τομέα της αναγνώρισης ασθενειών μέσω υπολογιστικής όρασης, ανάλυση των μοντέλων YOLO, καθώς και επισκόπηση μεθόδων Retrieval-Augmented Generation και χρήσης LLMs.
- **Κεφάλαιο 3** – Μεθοδολογία: Περιγραφή της διαδικασίας συλλογής και προεπεξεργασίας δεδομένων, της εκπαίδευσης και αξιολόγησης του μοντέλου YOLO, καθώς και της ανάπτυξης του συστήματος παροχής προτάσεων θεραπείας.
- **Κεφάλαιο 4** – Αποτελέσματα και Συζήτηση: Παρουσίαση των αποτελεσμάτων της αξιολόγησης, ανάλυση της απόδοσης του συστήματος, καθώς και συζήτηση των πλεονεκτημάτων και των περιορισμών του.
- **Κεφάλαιο 5** – Συμπεράσματα και Μελλοντική Εργασία: Σύνοψη των βασικών ευρημάτων, εξαγωγή συμπερασμάτων και προτάσεις για βελτιώσεις και επεκτάσεις στο μέλλον.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

Στο κεφάλαιο αυτό παρουσιάζονται αναλυτικά οι βασικές τεχνολογίες που έχουν σχέση με την εργασία αυτή.

2.1 Βασικοί Ορισμοί

Αρχικά, είναι σημαντικό να γίνει μια σαφής διάκριση μεταξύ όρων που συχνά συγχέονται, όπως η Τεχνητή Νοημοσύνη (Artificial Intelligence - AI), η Μηχανική Μάθηση (Machine Learning - ML) και η Βαθιά Μάθηση (Deep Learning - DL).

2.1.1 Τεχνητή Νοημοσύνη

Η **Τεχνητή Νοημοσύνη** [69] αναφέρεται στην ικανότητα ενός υπολογιστικού συστήματος ή λογισμικού να εκτελεί εργασίες που υπό φυσιολογικές συνθήκες απαιτούν ανθρώπινη νοημοσύνη, όπως η λήψη αποφάσεων, η επίλυση προβλημάτων, η αναγνώριση προτύπων και η κατανόηση φυσικής γλώσσας. Πρόκειται για έναν διεπιστημονικό τομέα της Πληροφορικής που εστιάζει στον σχεδιασμό και την υλοποίηση «έξυπνων» μηχανών. Η τεχνητή νοημοσύνη χρησιμοποιείται ευρέως σε τομείς, όπως η βιομηχανία, η ιατρική, οι μεταφορές, η κυβερνητική και η επιστημονική έρευνα. Ενδεικτικές εφαρμογές περιλαμβάνουν προηγμένες μηχανές αναζήτησης στο διαδίκτυο, συστήματα προτάσεων, αυτόνομα οχήματα και γενετικά μοντέλα, όπως το ChatGPT και το Gemini. Ο Άλαν Τιούρινγκ υπήρξε από τους πρώτους επιστήμονες που ασχολήθηκαν με το πεδίο, το οποίο αρχικά ονόμασε «μηχανική νοημοσύνη».

2.1.2 Μηχανική Μάθηση

Η **Μηχανική Μάθηση** (Machine Learning - ML) [20] αποτελεί έναν υποκλάδο της Τεχνητής Νοημοσύνης που επικεντρώνεται στην ανάπτυξη αλγορίθμων και μοντέλων που επιτρέπουν σε έναν υπολογιστή να «μαθαίνει» από δεδομένα και να βελτιώνει την απόδοσή του σε συγκεκριμένα καθήκοντα, χωρίς να είναι ρητά προγραμματισμένος. Ουσιαστικά, μέσω της μηχανικής μάθησης, το σύστημα αναγνωρίζει μοτίβα και σχέσεις μέσα σε δεδομένα, δημιουργώντας προγνωστικά ή ταξινομητικά μοντέλα. Οι πιο διαδεδομένες κατηγορίες μηχανικής

μάθησης είναι η εποπτευόμενη μάθηση (supervised learning), η μη εποπτευόμενη μάθηση (unsupervised learning) και η ενισχυτική μάθηση (reinforcement learning). Οι εφαρμογές της ML είναι ευρύτατες και περιλαμβάνουν συστήματα αναγνώρισης φωνής, ανάλυση εικόνας, ανίχνευση απάτης, προτάσεις περιεχομένου και χρηματοοικονομική πρόβλεψη.

2.1.3 Βαθιά Μάθηση

Η **Βαθιά Μάθηση** (Deep Learning - DL) [44] είναι ένας εξειδικευμένος τομέας της Μηχανικής Μάθησης, ο οποίος χρησιμοποιεί τεχνητά νευρωνικά δίκτυα με πολλές ενδιάμεσες στρώσεις (layers) για την επεξεργασία και κατανόηση σύνθετων δεδομένων. Τα βαθιά νευρωνικά δίκτυα έχουν τη δυνατότητα να μαθαίνουν υψηλού επιπέδου αναπαραστάσεις των δεδομένων, γεγονός που τους επιτρέπει να επιτυγχάνουν εντυπωσιακές επιδόσεις σε απαιτητικές εργασίες, όπως η αναγνώριση εικόνας, η κατανόηση φυσικής γλώσσας και η δημιουργία περιεχομένου. Η βαθιά μάθηση έχει καταστεί εφικτή χάρη στην αύξηση της υπολογιστικής ισχύος, στη διαθεσιμότητα μεγάλων συνόλων δεδομένων και στην πρόοδο των αλγορίθμων εκπαίδευσης. Ενδεικτικά μοντέλα βαθιάς μάθησης είναι τα συνελικτικά νευρωνικά δίκτυα (CNNs), τα επαναληπτικά δίκτυα (RNNs) και τα μετασχηματιστικά μοντέλα (Transformers).

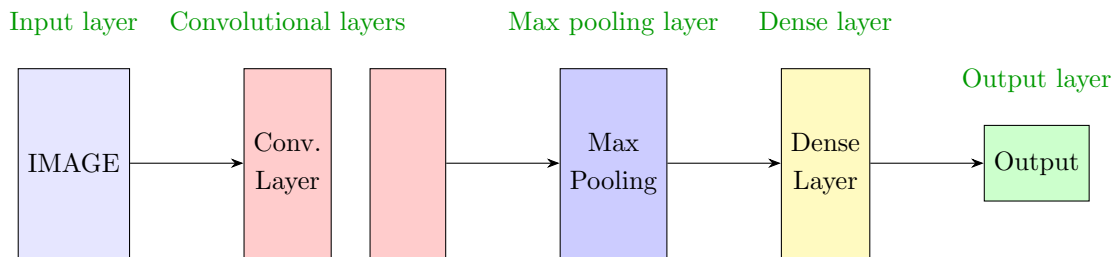
2.2 Convolutional Neural Network (CNN)

Ένα **Convolutional Neural Network (CNN)** [51] είναι ένα feedforward νευρωνικό δίκτυο ικανό να εξαγάγει *χαρακτηριστικά* από δεδομένα μέσω convolutional δομών (Σχήμα 2.1). Σε αντίθεση με τις κλασικές προσεγγίσεις, δεν απαιτεί χειροκίνητη εξαγωγή χαρακτηριστικών. Ο σχεδιασμός των CNN έχει επηρεαστεί από την *οπτική αντίληψη*: ένας βιολογικός νευρώνας αντιστοιχίζεται σε έναν τεχνητό, τα kernels δρουν ως υποδοχείς που ανιχνεύουν ιδιότητες, οι activation functions μιμούνται το κατώφλι διέγερσης, και οι loss functions μαζί με optimizers εκπαιδεύουν το δίκτυο, ώστε να επιτυγχάνει τον στόχο μάθησης. Σε σχέση με πλήρως συνδεδεμένα (Fully Connected - FC) δίκτυα, τα CNN υπερέχουν διότι:

1. **Τοπικές συνδέσεις:** κάθε νευρώνας συνδέεται μόνο με μικρή γειτονιά της προηγούμενης στρώσης, μειώνοντας παραμέτρους και επιταχύνοντας τη σύγκλιση.
2. **Κοινή χρήση βαρών (weight sharing):** πολλαπλές συνδέσεις μοιράζονται τα ίδια βάρη, άρα μειώνονται οι παράμετροι.
3. **Μείωση διάστασης (downsampling):** οι στρώσεις pooling εκμεταλλεύονται τις τοπικές συσχετίσεις για να ελαττώσουν τον όγκο δεδομένων, διατηρώντας βασικές λεπτομέρειες.

2.2.1 Convolution Layer

Για την κατασκευή ενός CNN ρυθμίζουμε συνήθως τέσσερα στοιχεία: **μέγεθος πυρήνα (kernel size)**, **padding**, **αριθμό καναλιών** και **βήμα (stride)**. Οι kernels είναι



Σχήμα 2.1: Απλοποιημένη μορφή συνελικτικού νευρωνικού δικτύου

μικρά φίλτρα που ολισθαίνουν κατά πλάτος και ύψος της εικόνας και υπολογίζουν εσωτερικά γινόμενα με τις τιμές των εικονοστοιχείων, παράγοντας feature maps. Η χρήση πεπερασμένου kernel συνεπάγεται απώλεια πληροφορίας στα άκρα. Για να το αντιμετωπίσουμε, χρησιμοποιούμε **padding**: συνήθως zero padding (συμπλήρωση με μηδενικά) ή replicate padding (αντιγραφή της τιμής του οριακού εικονοστοιχείου).

Ο αριθμός καναλιών είναι 1 για γκριζες εικόνες και 3 για RGB. Το stride ορίζει το βήμα ολίσθησης του kernel και ελέγχει την «πυκνότητα» της συνέλιξης: όσο μεγαλύτερο το stride, τόσο αραιότερη η δειγματοληψία.

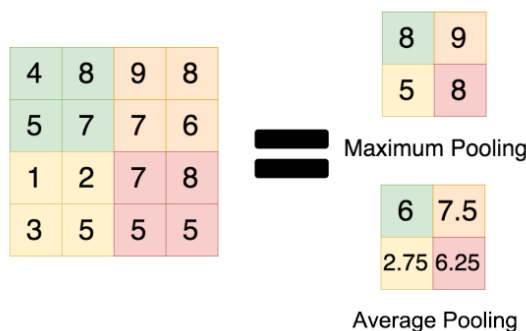
Μετά τη συνέλιξη, οι feature maps μπορεί να είναι πολλοί και να υπερπροσαρμόζουν (overfitting). Για αυτό εισάγουμε **pooling** (*downsampling*) στις αντίστοιχες στρώσεις, για απομάκρυνση πλεονάζουσας πληροφορίας.

2.2.2 Pooling Layer

Οι στρώσεις pooling είναι κρίσιμες στην εκπαίδευση ενός CNN [32]: ακολουθούν τις συνελικτικές στρώσεις και εξάγουν τις πιο ενημερωτικές τιμές από κάθε τοπικό παράθυρο, λειτουργώντας ως γενικευτές χαμηλού επιπέδου. Συνηθέστερες μορφές είναι το max pooling (μέγιστη τιμή) και το average pooling (μέσος όρος) (Σχήμα 2.2). Γενικά, το pooling μειώνει τις διαστάσεις, εκτελεί μη γραμμικό υπολογισμό και περιορίζει τον αριθμό παραμέτρων, μειώνοντας και την ανάλυση των feature maps της προηγούμενης στρώσης. Για πιο σύνθετες εικόνες, αυξάνουμε τον αριθμό των convolutional/pooling στρώσεων, ώστε να συλλάβουμε λεπτομερέστερη πληροφορία, με κόστος υπολογιστικού φόρτου.

2.2.3 Fully Connected Layer

Η **Fully Connected (FC) στρώση** ή **dense** είναι στρώση όπου κάθε νευρώνας συνδέεται με κάθε νευρώνα της προηγούμενης στρώσης. Συνήθως τοποθετείται στο τελικό τμήμα της αρχιτεκτονικής και είναι υπεύθυνη για την παραγωγή των τελικών προβλέψεων. Στα CNN, οι FC στρώσεις ακολουθούν τις convolutional και pooling στρώσεις: κάνουν *flatten* τη διδιάστατη χωρική διάταξη των feature maps σε vectors και εκτελούν την ταξινόμηση. Ο αριθμός νευρώνων της τελευταίας FC στρώσης ισούται με τον αριθμό των κλάσεων εξόδου. Για παράδειγμα, σε ταξινόμηση 8 κλάσεων, η τελική FC έχει 8 νευρώνες, καθένας εκ των οποίων αποδίδει ένα score για μία κλάση.



Σχήμα 2.2: Απεικόνιση δύο διαφορετικών διαδικασιών pooling (max και average). (Πηγή: ACM Digital library)

2.2.4 Loss Functions

Η loss function, γνωστή και ως cost function, χρησιμοποιείται για τον υπολογισμό της απόκλισης μεταξύ της αναμενόμενης και της πραγματικής τιμής. Αποτελεί βασικό κριτήριο μάθησης στο πρόβλημα βελτιστοποίησης και χρησιμοποιείται στα CNNs τόσο για regression όσο και για classification, με στόχο την ελαχιστοποίηση του σφάλματος.

Για regression, συνήθως χρησιμοποιούνται οι Mean Absolute Error (MAE) και Mean Squared Error (MSE). Η MAE υπολογίζει τον μέσο όρο των απόλυτων διαφορών μεταξύ προβλεπόμενης και πραγματικής τιμής, ενώ η MSE βασίζεται στα τετράγωνα των διαφορών. Η MAE είναι προτιμότερη όταν υπάρχουν πολλοί ακραίοι θόρυβοι στο σύνολο δεδομένων.

Για classification, η πιο διαδεδομένη είναι η cross entropy loss, η οποία μετρά τη διαφορά ανάμεσα στην προβλεπόμενη και την πραγματική κατανομή πιθανοτήτων. Η ποινή υπολογίζεται λογαριθμικά: μικρές αποκλίσεις δίνουν μικρές τιμές απώλειας, ενώ μεγαλύτερες αποκλίσεις δίνουν μεγαλύτερες τιμές.

Περισσότερες σύγχρονες loss functions περιλαμβάνουν [12]:

- **Contrastive Loss:** αυξάνει την απόσταση μεταξύ διαφορετικών κλάσεων και μειώνει την απόσταση εντός της ίδιας κλάσης. Είναι πολύ χρήσιμη για face recognition.
- **Center Loss:** προσθέτει περιορισμό για την ενδοκατηγορική συνοχή των χαρακτηριστικών.
- **Large Margin Softmax Loss:** εισάγει περιθώρια μεταξύ των κλάσεων, βελτιώνοντας την διαχωρισιμότητα. Χρησιμοποιείται σε εφαρμογές όπως emotion και speaker recognition.

2.2.5 Activation

Μια συγκεκριμένη συνάρτηση, γνωστή ως **activation function**, χρησιμοποιείται στα τεχνητά νευρωνικά δίκτυα για να μετατρέπει ένα επεξεργασμένο σήμα εισόδου σε σήμα εξόδου. Αυτό το σήμα εξόδου στη συνέχεια μεταβιβάζεται ως είσοδος στην επόμενη στρώση.

Ένα ANN υπολογίζει πρώτα το σταθμισμένο άθροισμα των εισόδων και στη συνέχεια εφαρμόζει τη συνάρτηση ενεργοποίησης στο αποτέλεσμα για να παραγάγει την έξοδο της στρώσης.

Χωρίς activation, ένα δίκτυο συμπεριφέρεται ως linear regression model, περιορισμένο σε εκφραστική ισχύ. Με κατάλληλες συναρτήσεις, το δίκτυο μπορεί να μάθει και να επεξεργαστεί πολύπλοκα δεδομένα όπως εικόνες, ήχο, βίντεο ή κείμενο.

Ενδεικτικές συναρτήσεις ενεργοποίησης φαίνονται στον Πίνακα 2.1

Συνάρτηση	Μαθηματικός Ορισμός	Περιγραφή
Binary Step	$f(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$	Απλή δυαδική έξοδος. Δεν χρησιμοποιείται συχνά λόγω μη διαφορισιμότητας.
Linear	$f(x) = x$	Γραμμική συνάρτηση χωρίς περιορισμό εύρους. Δεν εισάγει μη γραμμικότητα.
Sigmoid	$f(x) = \frac{1}{1+e^{-x}}$	Εξάγει τιμές μεταξύ 0 και 1. Κατάλληλη για πιθανοτικά μοντέλα, αλλά μπορεί να προκαλέσει vanishing gradients.
Tanh	$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	Παράγει τιμές μεταξύ -1 και 1. Έχει μηδενικό μέσο, αλλά μπορεί επίσης να έχει vanishing gradient.
ReLU	$f(x) = \max(0, x)$	Η πιο δημοφιλής ενεργοποίηση. Γρήγορη, αλλά μπορεί να «σκοτώσει» νευρώνες για αρνητικά x .
ELU	$f(x) = \begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$	Ομαλότερη από την ReLU, με τιμές κοντά στο μηδέν για $x < 0$.
Swish	$f(x) = x \cdot \sigma(x)$	Μη γραμμική, ομαλή και αυτορυθμιζόμενη συνάρτηση. Προτείνεται από την Google.
Softmax	$f_i(x) = \frac{e^{x_i}}{\sum_j e^{x_j}}$	Μετατρέπει διανύσματα σε κατανομές πιθανοτήτων. Χρησιμοποιείται στην έξοδο ταξινομητών.

Πίνακας 2.1: Ενδεικτικές συναρτήσεις ενεργοποίησης και οι μαθηματικοί τους ορισμοί.

2.2.6 Optimizer

Οι optimizers είναι αλγόριθμοι που τροποποιούν δυναμικά τις παραμέτρους του νευρωνικού μοντέλου κατά την εκπαίδευση, με στόχο την ελαχιστοποίηση της loss function. Ενημερώνουν τα weights και τα biases με βάση την πληροφορία που προκύπτει από τα δεδομένα.

Κύριοι optimizers που χρησιμοποιούνται στη βαθιά μάθηση [7]:

- **Stochastic Gradient Descent (SGD)**: βασικός αλγόριθμος με ενημερώσεις που βασίζονται στον υπολογισμό ή εκτίμηση της πρώτης παραγώγου (gradient).
- **Adam**: συνδυάζει προσαρμοστικούς ρυθμούς μάθησης και ορμής (momentum).
- **RMSprop**: κατάλληλος για μη σταθερά προβλήματα, ρυθμίζει τον ρυθμό μάθησης για κάθε παράμετρο.

Οι διαφορετικοί optimizers προσφέρουν διαφορετικές ταχύτητες σύγκλισης και σταθερότητα στην εκπαίδευση [64].

2.2.7 Hyperparameters

Κατά τη σχεδίαση ενός CNN, εκτός από τη συνάρτηση ενεργοποίησης, τη loss function και τον optimizer, απαιτείται ρύθμιση επιπλέον hyperparameters που επηρεάζουν σημαντικά την απόδοση του μοντέλου. Δεν υπάρχει ιδανικός συνδυασμός hyperparameters, επομένως η εμπειρία και η συστηματική ρύθμιση είναι κρίσιμες.

Η αναζήτηση του βέλτιστου συνδυασμού μπορεί να γίνει μέσω grid search, αλλά αυξάνεται εκθετικά με τον αριθμό παραμέτρων. Στις σημαντικές hyperparameters περιλαμβάνονται οι παρακάτω:

Learning Rate

Ορίζει το βήμα ενημέρωσης των weights. Ξεκινά υψηλό για ταχύτερη εκπαίδευση και μειώνεται προοδευτικά για σταθερότητα. Ένα κακά επιλεγμένο learning rate μπορεί να οδηγήσει σε απόκλιση ή στασιμότητα.

Number of Epochs

Ο όρος epoch δηλώνει τον αριθμό φορές που το σύνολο εκπαίδευσης περνά από το δίκτυο. Πολύ μικρός αριθμός προκαλεί underfitting, ενώ υπερβολικά μεγάλος προκαλεί overfitting. Στόχος είναι η σύγκλιση του training και validation loss.

Batch Size

Η batch size δηλώνει πόσα δείγματα επεξεργάζεται το μοντέλο κάθε φορά. Μικρή τιμή οδηγεί σε θορυβώδη εκπαίδευση, ενώ πολύ μεγάλη μπορεί να εγκλωβίσει το μοντέλο σε τοπικά ελάχιστα. Συνήθως επιλέγεται τιμή που είναι δύναμη του 2 (π.χ. 16, 32, 64), ανάλογα με τη μνήμη της GPU.

2.3 Μοντέλα Όρασης

Η υπολογιστική όραση (computer vision) αποτελεί έναν από τους σημαντικότερους κλάδους της τεχνητής νοημοσύνης και της μηχανικής μάθησης. Στόχος της είναι η ανάπτυξη συστημάτων με ικανότητα κατανόησης και επεξεργασίας οπτικών ερεθισμάτων, προσεγγίζοντας τις ανθρώπινες ικανότητες αντίληψης. Τα **μοντέλα όρασης** είναι οι αλγοριθμικές και μαθηματικές δομές που υλοποιούν αυτήν τη διαδικασία. Αρχικά, η έρευνα στον τομέα βασιζόταν σε χειροποίητα χαρακτηριστικά (hand-crafted features), όπως οι περιγραφείς ακμών (edges), οι υφές και οι γωνίες. Τεχνικές, όπως οι SIFT (Scale-Invariant Feature Transform) και HOG (Histogram of Oriented Gradients), αποτέλεσαν για χρόνια τη βάση πολλών εφαρμογών αναγνώρισης αντικειμένων [31, 8]. Ωστόσο, οι μέθοδοι αυτές είχαν περιορισμούς, καθώς απαιτούσαν εκτενή προεπεξεργασία και δεν μπορούσαν να γενικευθούν αποτελεσματικά σε σύνθετα δεδομένα.

Η σημαντική τομή ήρθε με την εμφάνιση των **νευρωνικών δικτύων** και ειδικότερα των **Συνελικτικών Νευρωνικών Δικτύων (Convolutional Neural Networks – CNNs)**. Τα CNNs έφεραν επανάσταση στην ανάλυση εικόνας, καθώς μάθαιναν αυτόματα αναπαραστάσεις πολλαπλών επιπέδων (από χαμηλού επιπέδου χαρακτηριστικά, όπως ακμές, έως υψηλού επιπέδου, όπως σχήματα και αντικείμενα). Το 2012, το μοντέλο AlexNet σημείωσε τεράστια βελτίωση στο διαγωνισμό ImageNet, μειώνοντας δραστικά το σφάλμα ταξινόμησης και σηματοδοτώντας την κυριαρχία του deep learning στην υπολογιστική όραση [24].

Έκτοτε, αναπτύχθηκαν πληθώρα αρχιτεκτονικών, όπως οι VGGNet, ResNet, Inception και EfficientNet, οι οποίες βελτίωσαν την ακρίβεια και την αποδοτικότητα των μοντέλων [48, 14, 53, 54]. Πρόσφατα, προτάθηκαν και εναλλακτικές αρχιτεκτονικές όπως το Vision Transformer (ViT) [11], το οποίο υιοθετεί μηχανισμούς self-attention για την καλύτερη κατανόηση μακροπρόθεσμων συσχετίσεων μέσα στην εικόνα. Παρά την επιτυχία των transformers σε διάφορες εφαρμογές, στην πράξη ένα από τα πιο διαδεδομένα και αποτελεσματικά μοντέλα για προβλήματα ανίχνευσης αντικειμένων είναι η οικογένεια **YOLO (You Only Look Once)**.

2.4 YOLO

2.4.1 Εισαγωγή στο YOLO

Ο Joseph Redmon, μαζί με τους Santosh Divvala, Ross Girshick και Ali Farhadi, δημοσίευσαν την εργασία “You Only Look Once: Unified, Real-Time Object Detection” στο συνέδριο CVPR, παρουσιάζοντας ένα επαναστατικό μοντέλο με το όνομα YOLO. Ο κύριος στόχος ήταν η δημιουργία ενός ταχύτερου αλγορίθμου ανίχνευσης αντικειμένων με μία μόνο διέλευση, (single-shot detection), χωρίς να γίνεται συμβιβασμός στην ακρίβεια. [41]

Η προσέγγιση του YOLO διατυπώνει την ανίχνευση ως πρόβλημα παλινδρόμησης, όπου η εικόνα περνά μία φορά από ένα ενιαίο συνελικτικό δίκτυο ώστε να παραχθούν οι συντεταγμένες των οριακών πλαισίων (bounding boxes) και οι αντίστοιχες κατηγορίες για πολλαπλά αντικείμενα. [41].

Η καινοτομία του έγκειται στην προσέγγιση single-shot, κατά την οποία η εικόνα χωρίζεται

γενίκευσης του μοντέλου. Η εκπαίδευση μπορεί να διακοπεί και να συνεχιστεί από το τελευταίο checkpoint, διευκολύνοντας την επανάληψη πειραμάτων ή τη βελτιστοποίηση υπαρχόντων μοντέλων [58, 62, 59].

2.5 Μεγάλα Γλωσσικά Μοντέλα

Τα Μεγάλα Γλωσσικά Μοντέλα (Large Language Models – LLMs) αποτελούν μια υποκατηγορία των νευρωνικών δικτύων, σχεδιασμένα για την επεξεργασία και κατανόηση της φυσικής γλώσσας. Εκπαιδεύονται πάνω σε τεράστιους όγκους δεδομένων κειμένου, με στόχο την πρόβλεψη της επόμενης λέξης σε μια ακολουθία, τη δημιουργία νοηματικά συνεκτικού λόγου, την απάντηση σε ερωτήσεις, τη μετάφραση γλωσσών και την περίληψη κειμένων [5, 22]. Τα μοντέλα αυτά βασίζονται κυρίως στη χρήση μετασχηματιστών (Transformers), οι οποίοι επιτρέπουν την αποδοτική επεξεργασία ακολουθιακών δεδομένων με βάση τους μηχανισμούς προσοχής (attention) [63]. Χάρη στην υψηλή παραμετροποιησιμότητά τους και την εκτενή εκπαίδευση, τα LLMs έχουν έρθει στο επίκεντρο της έρευνας για την επεξεργασία φυσικής γλώσσας. Η χρήση τέτοιων μοντέλων έχει πλέον επεκταθεί πέραν της απλής επεξεργασίας κειμένου, με εφαρμογές που περιλαμβάνουν chatbots συστήματα, υποβοήθηση στη παραγωγή κώδικα, αλλά και ενσωμάτωσή τους σε πολυτροπικά μοντέλα που επεξεργάζονται και εικόνες ή ήχο [39]. Η αρχιτεκτονική του σύγχρονου Transformers-Based LLM φαίνεται στο Σχήμα 2.4.

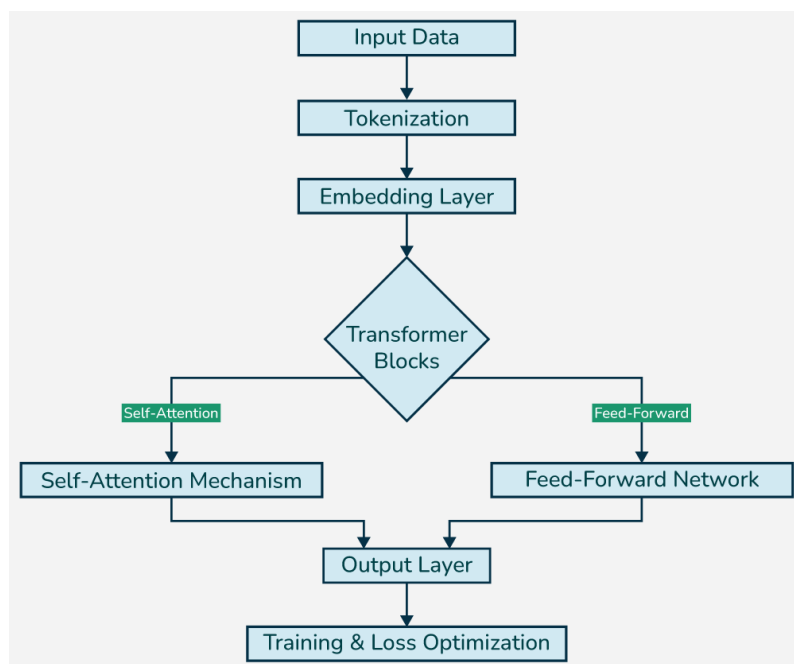
2.5.1 Ιστορική Αναδρομή

Η ιδέα των γλωσσικών μοντέλων ανάγεται στις απαρχές της στατιστικής γλωσσολογίας, με πρώιμα μοντέλα, όπως τα n-gram, να βασίζονται σε απλές πιθανότητες εμφάνισης λέξεων [22]. Με την πρόοδο των υπολογιστικών μεθόδων και την εμφάνιση των νευρωνικών δικτύων, άρχισαν να αναπτύσσονται πιο σύνθετα μοντέλα, όπως τα RNNs (Recurrent Neural Networks) και τα LSTMs (Long Short-Term Memory) [15]. Το 2017, η εισαγωγή του μετασχηματιστή (Transformer) από το άρθρο **“Attention is All You Need”** αποτέλεσε σημείο [63]. Το νέο αυτό αρχιτεκτονικό μοντέλο επέτρεψε την εκπαίδευση πολύ μεγαλύτερων δικτύων, με καλύτερη διαχείριση μακροπρόθεσμων εξαρτήσεων στη γλώσσα.

Η πρώτη σημαντική εφαρμογή του Transformer σε γλωσσικό μοντέλο ήταν το BERT της Google [10], ενώ ακολούθησε το GPT της OpenAI. Η κυκλοφορία του GPT-3 (2020) με 175 δισεκατομμύρια παραμέτρους έδειξε την πραγματική δυναμική των LLMs στην παραγωγή κειμένου και την γενικότερη κατανόηση γλώσσας [5]. Από τότε, έχουν παρουσιαστεί πολλά άλλα μοντέλα, τόσο από εταιρείες, όσο και από την ακαδημαϊκή κοινότητα, με διαρκώς αυξανόμενες δυνατότητες και εφαρμογές.

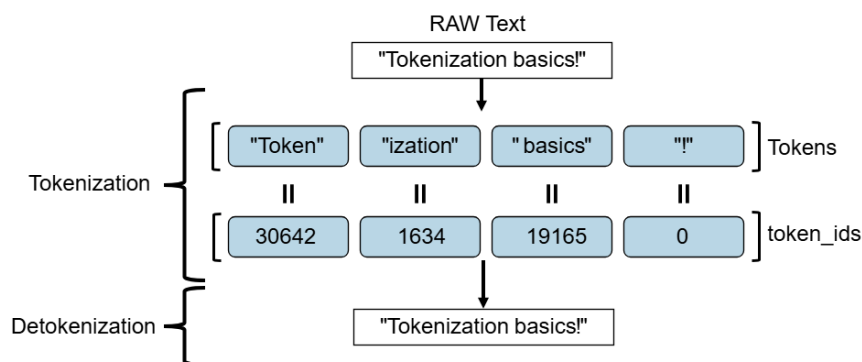
2.5.2 Tokenizer

Ο tokenizer είναι υπεύθυνος για τη μετατροπή του αρχικού κειμένου σε ακολουθία από **tokens** (Σχήμα 2.5, δηλαδή διακριτά σύμβολα που αντιστοιχούν σε αχέραιους δείκτες ενός προκαθορισμένου λεξιλογίου, και το προετοιμάζει για το μοντέλο [17].



Σχήμα 2.4: Η αρχιτεκτονική του σύγχρονου LLM (Πηγή: GeeksForGeeks)

Συνήθως βασίζεται σε υπολέξεις (subword μεθόδους, όπως [45], WordPiece [67] ή Unigram [25]), ώστε να επιτυγχάνεται μια ισορροπία μεταξύ επαρκούς κάλυψης του λεξιλογίου και περιορισμού του συνολικού μήκους των ακολουθιών. Ο tokenizer περιλαμβάνει επίσης κανόνες ομαλοποίησης, όπως η μετατροπή των χαρακτήρων σε πεζά και η διαχείριση σημείων στίξης, καθώς και ειδικά tokens για την ένδειξη θέσεων ή καταστάσεων (όπως [BOS], [EOS], [PAD], [UNK]). Τέλος, η διαδικασία μπορεί να αντιστραφεί μέσω του μηχανισμού detokenization, ο οποίος ανακατασκευάζει το αρχικό κείμενο από τη σειρά των tokens [26, 27, 18].



Σχήμα 2.5: Η λειτουργία του tokenizer (Πηγή: OpenVINO Docs)

2.5.3 Μηχανισμός Attention

Ο μηχανισμός προσοχής (attention) επιτρέπει στο μοντέλο να εστιάζει δυναμικά σε διαφορετικά τμήματα της εισόδου κατά την επεξεργασία κάθε λέξης ή token. Η βασική ιδέα είναι

ότι, αντί να επεξεργάζεται όλες τις λέξεις εξίσου, το μοντέλο μαθαίνει να αποδίδει διαφορετική «βαρύτητα» στην καθεμία, ανάλογα με τη συνάφειά της με την υπό εξέταση λέξη. Η πιο διαδεδομένη μορφή είναι η *κλιμακωτή γινομενική προσοχή* (scaled dot-product attention) [63], η οποία υπολογίζεται ως:

$$\text{Attn}(Q, K, V) = \text{softmax} \left(\frac{QK^\top}{\sqrt{d_k}} \right) V,$$

όπου Q (ερωτήματα), K (κλειδιά) και V (τιμές) είναι προβολές των εισόδων μέσω μαθηματικών μετασχηματισμών. Η προσοχή υπολογίζει τη βαθμολογία ομοιότητας μεταξύ Q και K και τη χρησιμοποιεί για να σταθμίσει τις V . Αυτό επιτρέπει στο μοντέλο να «δανείζεται» πληροφορία από άλλες θέσεις της ακολουθίας, ενισχύοντας την κατανόηση συμφραζομένων, συντακτικών και σημασιολογικών σχέσεων. Η τεχνική αυτή είναι κρίσιμη για την επιτυχία των Transformers και αποτελεί το βασικό συστατικό των σύγχρονων LLMs.

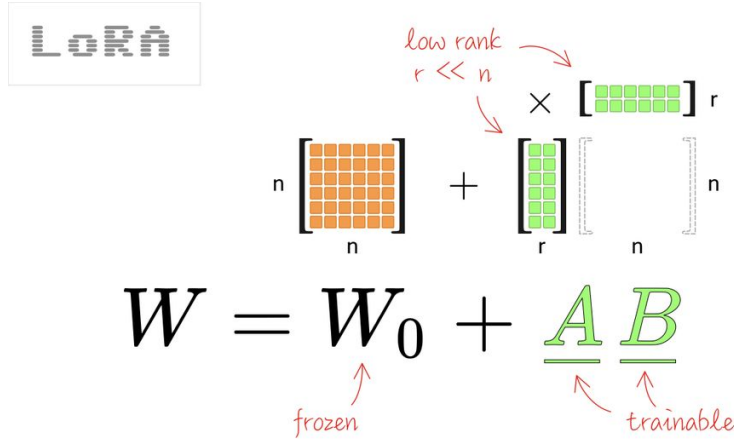
2.5.4 Embeddings και Κωδικοποιήσεις Θέσης

Κάθε token χαρτογραφείται σε ένα διάνυσμα d διαστάσεων μέσω πίνακα ενσωματώσεων $E \in \mathbb{R}^{|\mathcal{V}| \times d}$, όπου $|\mathcal{V}|$: Το πλήθος των tokens στο λεξιλόγιο (vocabulary size) και d : Η διάσταση του διανύσματος ενσωμάτωσης (embedding dimension). Επειδή ο Transformer δεν είναι ακολουθιακός, η πληροφορία θέσης εισάγεται με κωδικοποιήσεις θέσης: απόλυτες (ημιτονοειδείς ή εκπαιδευσιμες) ή σχετικές/περιστροφικές (relative [46], RoPE [52]), οι οποίες προστίθενται ή ενσωματώνονται πολλαπλασιαστικά στα διανύσματα, ώστε το μοντέλο να διακρίνει τη σειρά των tokens.

2.5.5 Fine-tuning με χρήση LoRA Adapters

Το fine-tuning μεγάλων γλωσσικών μοντέλων αποτελεί βασικό βήμα για την προσαρμογή τους σε εξειδικευμένες εργασίες ή θεματικές περιοχές. Παρά τα οφέλη του, η πλήρης αναπροσαρμογή όλων των παραμέτρων ενός προεκπαιδευμένου μοντέλου είναι ιδιαίτερα απαιτητική σε υπολογιστικούς πόρους, καθιστώντας τη διαδικασία μη βιώσιμη σε περιβάλλοντα με περιορισμένες δυνατότητες. Για την αντιμετώπιση αυτού του ζητήματος, έχουν αναπτυχθεί τεχνικές αποδοτικού fine-tuning, με πιο διαδεδομένη τη χρήση των LoRA (Low-Rank Adaptation) adapters.

Οι LoRA adapters βασίζονται στην ιδέα της εισαγωγής μικρών γραμμικών μετασχηματισμών χαμηλής τάξης, οι οποίοι ενσωματώνονται στα ήδη υπάρχοντα βάρη του μοντέλου, όπως για παράδειγμα στους query ή value μετασχηματισμούς ενός μηχανισμού προσοχής. Κατά την εκπαίδευση, οι βασικές παράμετροι του μοντέλου παραμένουν σταθερές και μαθαίνονται μόνο οι πρόσθετες παράμετροι των LoRA, μειώνοντας έτσι σημαντικά τον αριθμό των μαθησιακών παραμέτρων. Η τροποποίηση αυτή εκφράζεται μαθηματικά ως $W = W_0 + BA$, όπου W_0 είναι το αρχικό βάρος, ενώ οι πίνακες B και A μαθαίνονται κατά την προσαρμογή και έχουν πολύ μικρότερες διαστάσεις. Αυτή η προσέγγιση επιτρέπει την αποδοτική εκπαίδευση, ακόμη και σε περιορισμένες υπολογιστικές υποδομές, καθιστώντας εφικτή τη χρήση προηγμένων μοντέλων σε πρακτικές εφαρμογές. Επιπλέον, η ανεξαρτησία των LoRA adapters από το κύριο μοντέλο



Σχήμα 2.6: LoRA Adapter

επιτρέπει την εύκολη εναλλαγή μεταξύ διαφορετικών προσαρμοσμένων εκδόσεων, χωρίς την ανάγκη αναπροσαρμογής του ίδιου του μοντέλου.

Στο πλαίσιο της παρούσας εργασίας, επιλέχθηκε η χρήση LoRA ως η πλέον κατάλληλη μέθοδος φινε-τυνινγ, καθώς προσφέρει την ευελιξία και την αποδοτικότητα που απαιτούνται για την εξειδίκευση σε γεωργικά δεδομένα, χωρίς να επιβαρύνει ιδιαίτερα το σύστημα ή τη διαδικασία ανάπτυξης.

2.5.6 Βελτιστοποιήσεις Υπολογιστικής Απόδοσης

Η ανάπτυξη και εκτέλεση μεγάλων γλωσσικών μοντέλων απαιτεί σημαντικούς υπολογιστικούς πόρους. Για να καταστεί δυνατή η χρήση τους σε πραγματικό χρόνο ή σε συστήματα περιορισμένων πόρων, εφαρμόζονται τεχνικές όπως το quantization και το parallelization.

Quantization Το quantization είναι μία τεχνική μείωσης της αριθμητικής ακρίβειας των παραμέτρων και των ενεργοποιήσεων σε νευρωνικά δίκτυα, με σκοπό την ελάφρυνση του υπολογιστικού φορτίου κατά τη χρήση τους (inference). Στην περίπτωση των Large Language Models (LLMs), τα οποία συχνά περιλαμβάνουν δισεκατομμύρια παραμέτρους, το quantization καθίσταται απαραίτητο για την εκτέλεσή τους σε συστήματα περιορισμένων πόρων. Μέσω της μετατροπής από αριθμητικούς τύπους υψηλής ακρίβειας (π.χ. float32) σε μικρότερης ακρίβειας μορφές, όπως int8 ή float16, επιτυγχάνεται σημαντική μείωση της χρήσης μνήμης και επιτάχυνση των υπολογισμών. Στο πλαίσιο της παρούσας εργασίας, η χρήση τεχνικών quantization ήταν αναγκαία, προκειμένου να καταστεί εφικτό το inference των επιλεγμένων LLMs στους διαθέσιμους servers, οι οποίοι διέθεταν περιορισμένους υπολογιστικούς πόρους. Η μείωση του κόστους εκτέλεσης συνοδεύτηκε από διατήρηση ικανοποιητικής απόδοσης, καθιστώντας την τεχνική αυτή κρίσιμη για την επιτυχία της υλοποίησης.

Parallelization Το parallelization είναι μια τεχνική που επιτρέπει την ταυτόχρονη εκτέλεση υπολογισμών σε περισσότερους από έναν επεξεργαστές, με σκοπό να επιταχυνθεί η εκπαίδευση ή η χρήση ενός μοντέλου. Ειδικά στην περίπτωση των Large Language Models (LLMs), που έχουν πολύ μεγάλο αριθμό παραμέτρων, η παράλληλη επεξεργασία είναι συχνά απαραίτητη, γιατί διαφορετικά η εκπαίδευση ή το inference μπορεί να πάρει υπερβολικά πολύ χρόνο ή να μην είναι καν εφικτό (π.χ. λόγω ανεπαρκούς μνήμης). Τεχνικές, όπως το **data parallelism**, όπου το ίδιο μοντέλο εφαρμόζεται σε διαφορετικά κομμάτια των δεδομένων, ή η **model parallelism**, όπου το μοντέλο «σπάει» και κατανέμεται σε πολλαπλές συσκευές, βοηθούν στην καλύτερη αξιοποίηση των διαθέσιμων πόρων.

2.5.7 Τεχνικές Prompting

Μια ακόμη σημαντική τεχνική αξιοποίησης των Large Language Models (LLMs) είναι οι τεχνικές prompting, δηλαδή η παροχή κατάλληλων εντολών ή παραδειγμάτων προς το μοντέλο με σκοπό τη βελτιστοποίηση της απόδοσής του, χωρίς την ανάγκη περαιτέρω εκπαίδευσης. Οι τεχνικές prompting βασίζονται στη διαμόρφωση του περιβάλλοντος εισόδου, έτσι ώστε το μοντέλο να κατανοήσει με μεγαλύτερη ακρίβεια την πρόθεση του χρήστη και να παράγει πιο σχετικές και στοχευμένες απαντήσεις. Στο πλαίσιο αυτό έχουν αναπτυχθεί διάφορες στρατηγικές:

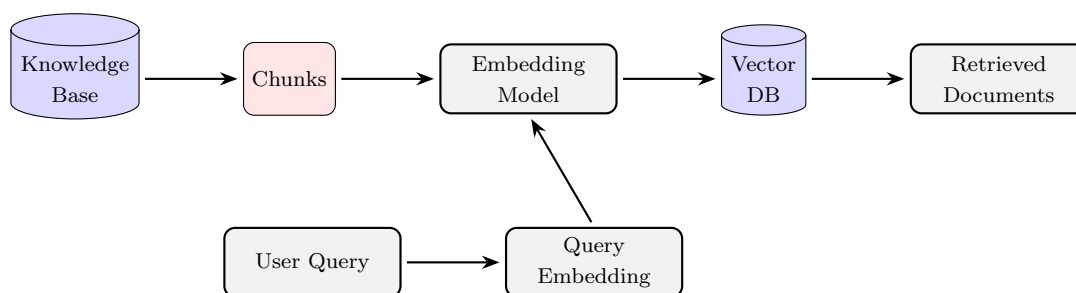
- **Full-Context Prompting:** Η τεχνική full-context prompting περιλαμβάνει την παροχή ολόκληρου του σχετικού περιεχομένου ως είσοδο, ώστε το μοντέλο να έχει πλήρη γνώση των δεδομένων κατά τη φάση παραγωγής. Αυτό επιτρέπει την παραγωγή απαντήσεων που λαμβάνουν υπόψη όλες τις διαθέσιμες πληροφορίες και είναι ιδιαίτερα χρήσιμο σε περιπτώσεις, όπως η σύνοψη εγγράφων ή η σύνθετη ερώτηση-απάντηση.
- **JSON Prompting** Αντίστοιχα, το JSON prompting αξιοποιεί την ικανότητα των μοντέλων να παράγουν δομημένες εξόδους σε μορφή JSON, διευκολύνοντας την ενσωμάτωσή τους σε συστήματα που απαιτούν αυστηρή μορφοποίηση των απαντήσεων. Η μέθοδος αυτή είναι ιδιαίτερα χρήσιμη σε εφαρμογές, όπου η έξοδος πρέπει να μετατραπεί αυτόματα σε αντικείμενο ή να αναλυθεί από άλλες διεργασίες.
- **Chain-Of-Thought Prompting** Η τεχνική Chain-of-Thought (CoT) prompting, επιτρέπει την εκτέλεση πολύπλοκων λογικών συλλογισμών μέσω ενδιάμεσων βημάτων σκέψης. Αντί να ζητείται μια άμεση απάντηση από το μοντέλο, η μέθοδος αυτή το ενθαρρύνει να παράγει πρώτα μια ακολουθία λογικών βημάτων που οδηγούν στην τελική απάντηση. Η CoT prompting τεχνική μπορεί να συνδυαστεί αποτελεσματικά με τεχνικές few-shot prompting (όπου παρέχονται στο μοντέλο λίγα παραδείγματα ερωτήσεων-απαντήσεων ως οδηγός), ώστε να ενισχυθεί η απόδοση του μοντέλου.

2.6 Πολυτροπικά Μεγάλα Γλωσσικά Μοντέλα

Τα πολυτροπικά μεγάλα γλωσσικά μοντέλα (Multimodal Large Language Models – MLLMs) επεκτείνουν τις δυνατότητες των κλασικών LLMs ενσωματώνοντας και άλλες μορφές πληροφορίας, όπως εικόνες, ήχο ή βίντεο [39, 70, 13], πέραν του καθαρού κειμένου. Μέσω της πολυτροπικής επεξεργασίας, τα μοντέλα αυτά είναι σε θέση να κατανοήσουν και να συσχετίσουν πληροφορία από διαφορετικά είδη εισόδων, επιτρέποντας εφαρμογές, όπως η περιγραφή εικόνων, η απάντηση σε ερωτήσεις βασισμένες σε εικόνα (Visual Question Answering) [3].

Στο πλαίσιο της παρούσας εργασίας γίνεται χρήση αποκλειστικά **οπτικών πολυτροπικών μοντέλων**, δηλαδή μοντέλων που επεξεργάζονται εικόνες σε συνδυασμό με φυσική γλώσσα. Τα μοντέλα αυτά βασίζονται σε έναν συνδυασμό οπτικών και γλωσσικών επεξεργαστών, με στόχο την κατανόηση του περιεχομένου μιας εικόνας και την παραγωγή αντίστοιχου κειμένου, όπως διαγνώσεις ή περιγραφές.

2.7 Retrieval-Augmented Generation (RAG)



Σχήμα 2.7: Αρχιτεκτονική του Retrieval-Augmented Generation (RAG).

Το Retrieval-Augmented Generation (RAG) είναι μια προσέγγιση που συνδυάζει την ανάκτηση πληροφορίας (retrieval) με τη δημιουργία φυσικής γλώσσας (generation), με στόχο τη βελτίωση της ακρίβειας και της ενημερότητας των απαντήσεων που παράγονται από τα γλωσσικά μοντέλα. Το βασικό πρόβλημα που επιλύει το RAG είναι το λεγόμενο *knowledge cutoff* και το *hallucination* [16] των LLMs, δηλαδή η τάση τους να απαντούν με τρόπο πειστικό αλλά λανθασμένο, ειδικά όταν δεν έχουν πρόσβαση σε ενημερωμένες ή εξειδικευμένες πληροφορίες. Αντί να βασίζονται αποκλειστικά στη γνώση που έχουν μάθει κατά την προεκπαίδευση, τα μοντέλα τύπου RAG **αναζητούν σχετικά έγγραφα** από μια εξωτερική βάση δεδομένων (π.χ. vector store). Τα αποτελέσματα αυτά χρησιμοποιούνται από το μοντέλο σαν ground truth για να απαντήσουν την ερώτηση του χρήστη [28].

Η αρχιτεκτονική RAG τυπικά αποτελείται από δύο κύρια στάδια: την ανάκτηση πληροφορίας με χρήση ενός retriever, ο οποίος αναζητά σχετικά κείμενα βάσει (embeddings), και τη γλωσσική παραγωγή από έναν generator (συνήθως ένα LLM), ο οποίος χρησιμοποιεί τα ανακτηθέντα κείμενα ως συμφραζόμενα για να παραγάγει την τελική απάντηση. Οι πιο διαδεδομένες μεθοδολογίες για την ανάκτηση περιλαμβάνουν την αναζήτηση με πυκνά διανύσματα

(dense retrieval) με χρήση του FAISS, ή την αποθήκευση και ανάκτηση μέσω εξειδικευμένων βάσεων, όπως το ChromaDB. Παρακάτω, αναλύονται βασικές παράμετροι ενός RAG συστήματος.

2.7.1 Πλήθος Ανακτήσεων (top-k)

Η παράμετρος top-k καθορίζει πόσα έγγραφα θα ανακτηθούν από τη βάση γνώσης για κάθε ερώτημα. Ένα μικρό k μπορεί να οδηγήσει σε απώλεια κρίσιμων πληροφοριών, ενώ ένα πολύ μεγάλο k αυξάνει τον θόρυβο και την πιθανότητα να ξεπεραστεί το context window του μοντέλου. Η τιμή του k πρέπει να επιλέγεται με βάση τη συγκεκριμένη εφαρμογή και τους διαθέσιμους υπολογιστικούς πόρους.

2.7.2 Μοντέλο Embedding

Η ποιότητα των embeddings που χρησιμοποιούνται τόσο για τα έγγραφα όσο και για τα ερωτήματα επηρεάζει άμεσα τη συνάφεια των ανακτηθέντων πληροφοριών. Η χρήση πιο εξελιγμένων μοντέλων embedding (π.χ. Instructor, E5, ή OpenAI text-embedding models) μπορεί να βελτιώσει αισθητά τα αποτελέσματα του retrieval.

2.7.3 Μέγεθος Context Window

Το context window του γλωσσικού μοντέλου ορίζει πόσο μεγάλο κείμενο μπορεί να επεξεργαστεί. Σε μοντέλα με περιορισμένο παράθυρο συμφοραζομένων (π.χ. 2K ή 4K tokens), είναι σημαντικό να γίνεται επιλογή μόνο των πιο σχετικών εγγράφων. Αντίθετα, σε μεγαλύτερα μοντέλα (π.χ. με 32K ή 128K tokens), μπορεί να συμπεριληφθεί περισσότερη πληροφορία, αυξάνοντας την ακρίβεια του generation.

2.7.4 Μέγεθος Τμημάτων Κειμένου

Πριν αποθηκευτούν τα έγγραφα στον διανυσματική βάση δεδομένων, χωρίζονται σε μικρότερα τμήματα (chunks), ώστε να μπορούν να αναζητηθούν πιο εύκολα. Το μέγεθος κάθε chunk καθορίζει πόσες λέξεις ή tokens περιέχει το κάθε τμήμα. Αν είναι πολύ μικρό, μπορεί να χάνεται το νόημα. Αν είναι πολύ μεγάλο, μπορεί να μην «χωράει» στο context window.

2.7.5 Επικάλυψη Μεταξύ Τμημάτων Κειμένου

Η παράμετρος αυτή καθορίζει αν και πόσο επικαλύπτονται τα διαδοχικά chunks. Η επικάλυψη (overlap) βοηθά ώστε να μην κόβεται απότομα η ροή του κειμένου και να διατηρείται η συνοχή, ειδικά όταν σημαντικές πληροφορίες βρίσκονται στο όριο μεταξύ δύο τμημάτων.

2.8 Αρχιτεκτονική Πολυτροπικών Μοντέλων

Η αρχιτεκτονική των πολυτροπικών μοντέλων περιλαμβάνει συνήθως δύο βασικά υποσυστήματα: έναν vision encoder για την επεξεργασία των εικόνων και έναν language model (συ-

νήθως προεκπαιδευμένο LLM) για την παραγωγή ή ερμηνεία γλωσσικών αποτελεσμάτων [4, 71]. Οι έξοδοι του vision encoder μετατρέπονται σε ενδιάμεσες αναπαραστάσεις (π.χ. διανύσματα ή tokens), οι οποίες ενσωματώνονται στο πλαίσιο του γλωσσικού μοντέλου είτε μέσω απευθείας σύντηξης, είτε μέσω ενός ενδιάμεσου adapter [57, 29].

Οι πιο συνήθεις τρόποι σύντηξης των modalities είναι:

- **Early fusion:** Ενοποίηση των εισόδων πριν την είσοδό τους στο μοντέλο.
- **Late fusion:** Επεξεργασία κάθε modality ξεχωριστά και συνδυασμός των εξόδων.
- **Cross-modal attention:** Οι εκπροσωπήσεις κάθε modality αλληλεπιδρούν μέσω προσοχής.

2.8.1 Vision Transformer (ViT)

Για την επεξεργασία της οπτικής πληροφορίας, τα περισσότερα πολυτροπικά μοντέλα χρησιμοποιούν τον Vision Transformer (ViT), έναν τύπο νευρωνικού δικτύου που εφαρμόζει την αρχιτεκτονική των Transformers σε εικόνες. Η είσοδος μιας εικόνας $x \in \mathbb{R}^{H \times W \times C}$ χωρίζεται σε N επιθέματα (patches) διαστάσεων $P \times P$, τα οποία μετατρέπονται σε διανύσματα $x_p^i \in \mathbb{R}^{P^2 \cdot C}$.

Αυτά προβάλλονται γραμμικά σε χώρο ενσωμάτωσης:

$$z_0^i = Ex_p^i + p_i, \quad i = 1, \dots, N,$$

όπου E είναι ο πίνακας ενσωμάτωσης και p_i οι αντίστοιχες κωδικοποιήσεις θέσης. Προστίθεται επίσης ένα ειδικό [CLS] token για την τελική αναπαράσταση της εικόνας. Η ακολουθία εισόδου εισέρχεται σε στοίβα Transformer blocks με μηχανισμούς multi-head self-attention και feed-forward layers, όπως και στα γλωσσικά μοντέλα.

2.9 Ροή Επεξεργασίας με LangGraph

Το LangGraph αποτελεί ένα εργαλείο ανοικτού κώδικα που επεκτείνει τις δυνατότητες των γλωσσικών μοντέλων, επιτρέποντας τη δημιουργία πολύπλοκων ροών διαλόγου και απόφασης μέσω γραφημάτων κατάστασης. Αντί για απλή αλληλουχία εντολών ή προτροπών, με το LangGraph ο χρήστης μπορεί να ορίσει έναν γράφο (graph) στον οποίο κάθε κόμβος (node) αντιπροσωπεύει μία συγκεκριμένη υπολογιστική εργασία, ενώ οι ακμές (edges) καθορίζουν τις μεταβάσεις μεταξύ καταστάσεων.

2.9.1 Η Λογική του State Machine

Κάθε γράφος που δημιουργείται με το LangGraph βασίζεται στη λογική των state machines, όπου το σύστημα βρίσκεται πάντα σε μια συγκεκριμένη κατάσταση (state), εκτελεί μια λειτουργία και στη συνέχεια αποφασίζει – βάσει κάποιου κριτηρίου – ποια θα είναι η επόμενη μετάβαση. Αυτό επιτρέπει τη δημιουργία πολύπλοκων ακολουθιών, όπως:

- ελέγχους ποιότητας των δεδομένων εισόδου,
- λήψη αποφάσεων με βάση εξωτερική γνώση,
- επανατροφοδότηση (feedback) και επανεκτέλεση.

Το LangGraph αξιοποιεί τις δυνατότητες του LangChain για αλυσίδες εργαλείων (chains) και τις ενσωματώνει σε γράφους με παραμετρική, επεκτάσιμη λογική.

2.9.2 Πλεονεκτήματα για την Παρούσα Εργασία

Η χρήση του LangGraph στην παρούσα εργασία κρίθηκε απαραίτητη για τη δομημένη διαχείριση της ροής πληροφορίας. Πιο συγκεκριμένα, το LangGraph ορίζει με σαφήνεια τη ροή από την εισαγωγή της εικόνας έως την τελική απάντηση, επιτρέποντας έτσι τον ακριβή έλεγχο της διαδικασίας. Επιπλέον, παρέχει τη δυνατότητα εύκολου διαχωρισμού των λειτουργιών σε επιμέρους κόμβους, διευκολύνοντας την υλοποίηση μέσω ενός *modular* σχεδιασμού. Η δομή αυτή επιτρέπει την εύκολη προσθήκη νέων δυνατοτήτων, όπως η αποθήκευση ιστορικού ή η ενσωμάτωση επιπλέον πηγών γνώσης, χωρίς να απαιτείται αναδιάρθρωση του συνολικού συστήματος. Τέλος, το LangGraph υποστηρίζει την οπτικοποίηση της διαδικασίας, κάτι που είναι ιδιαίτερα χρήσιμο, τόσο για τον εντοπισμό σφαλμάτων (*debugging*), όσο και για την παρουσίαση της ροής σε εκπαιδευτικό ή ερευνητικό πλαίσιο.

2.10 Αναγνώριση Φυτοπαθολογιών

Οι ασθένειες των φύλλων διαφοροποιούνται ανάλογα με το είδος του δέντρου, τις τοπικές κλιματικές συνθήκες και το μικροπεριβάλλον της καλλιέργειας. Για παράδειγμα, στις μηλιές παρατηρούνται συχνά μυκητολογικές ασθένειες, όπως η *Venturia inaequalis* (χηλίδες μήλου), ενώ στις ελιές κυριαρχεί το κυκλοκόνιο (*Spilocaea oleagina*). Οι ασθένειες αυτές εκδηλώνονται με διαφορετικά μορφολογικά χαρακτηριστικά, όπως αλλαγές στο χρώμα, το σχήμα και την υφή των φύλλων, και συχνά σχετίζονται με την εποχικότητα και την υγρασία.

Η θεραπευτική αντιμετώπιση περιλαμβάνει τη χρήση δραστικών ουσιών με προληπτική ή κατασταλτική δράση. Ενδεικτικά, χρησιμοποιούνται μυκητοκτόνα, όπως το mancozeb, το copper hydroxide και το trifloxystrobin, τα οποία επιδρούν αναστέλλοντας τον κύκλο ζωής του παθογόνου.

Ωστόσο, ολοένα και περισσότερο προτείνεται η ενσωμάτωση εναλλακτικών, μη-χημικών πρακτικών στο πλαίσιο της ολοκληρωμένης φυτοπροστασίας. Τέτοιες πρακτικές περιλαμβάνουν το σωστό και τακτικό κλάδεμα για βελτίωση του αερισμού της κόμης, την απομάκρυνση προσβεβλημένων φύλλων, τη χρήση φυσικών εκχυλισμάτων (π.χ. σκόρδου, τσουκνίδας), καθώς και την ενίσχυση του μικροβιακού οικοσυστήματος του εδάφους με οργανική ουσία ή ωφέλιμους μικροοργανισμούς (π.χ. μυκόρριζες). Ο συνδυασμός φαρμακευτικών και φυσικών μεθόδων μπορεί να προσφέρει αποτελεσματική και οικολογικά βιώσιμη αντιμετώπιση των ασθενειών των φύλλων.

2.11 Ψηφιακές Πλατφόρμες και Εργαλεία Υλοποίησης

Στην παρούσα εργασία αξιοποιήθηκαν διάφορες πλατφόρμες και εργαλεία που διευκολύνουν την πρόσβαση σε δεδομένα (datasets), προεκπαιδευμένα μοντέλα και υπολογιστικούς πόρους.

Το **Kaggle** [23] είναι μια διαδικτυακή πλατφόρμα που παρέχει πρόσβαση σε πλήθος διαγωνισμών μηχανικής μάθησης, καθώς και σε ένα εκτενές αποθετήριο με χιλιάδες σύνολα δεδομένων, τα οποία αξιοποιούνται ευρέως για ερευνητικούς και εκπαιδευτικούς σκοπούς.

Το **Hugging Face** [1] αποτελεί μία από τις πιο σημαντικές πλατφόρμες ανοικτού λογισμικού για μοντέλα μηχανικής μάθησης και επεξεργασίας φυσικής γλώσσας. Προσφέρει βιβλιοθήκες όπως την Transformers, καθώς και ένα online αποθετήριο όπου οι ερευνητές και οι προγραμματιστές μπορούν να δημοσιεύουν και να δοκιμάζουν μοντέλα, να μοιράζονται datasets, και να δημιουργούν pipelines για επεξεργασία κειμένου, εικόνας ή και πολυτροπικών δεδομένων.

Τέλος, το **Ollama** [38] είναι ένα ελαφρύ περιβάλλον που επιτρέπει την εκτέλεση μεγάλων γλωσσικών μοντέλων (LLMs) τοπικά, χωρίς την ανάγκη σύνδεσης σε απομακρυσμένους διακομιστές. Υποστηρίζει πληθώρα μοντέλων, συμπεριλαμβανομένων των LLaMA, Mistral και Gemma, και μπορεί να ενσωματωθεί εύκολα σε εφαρμογές με χρήση API.

Η αξιοποίηση αυτών των εργαλείων συνέβαλε καθοριστικά στην ανάπτυξη και αξιολόγηση του συστήματος που παρουσιάζεται στην παρούσα εργασία.

Κεφάλαιο 3

Μοντελοποίηση Συστήματος

Στο πλαίσιο της παρούσας διπλωματικής εργασίας υλοποιήθηκαν δύο διαφορετικές προσεγγίσεις για την αναγνώριση ασθενειών σε φύλλα, οι οποίες θα συγκριθούν και θα αξιολογηθούν σε επόμενο κεφάλαιο. Η πρώτη προσέγγιση, η οποία θα αναφέρεται ως Vision-Guided Retrieval System (VGRS), συνδυάζει ένα συνδυασμό μοντέλων ανίχνευσης αντικειμένων τύπου YOLO, ένα προεκπαιδευμένο μεγάλο γλωσσικό μοντέλο (LLM) και ένα σύστημα (RAG) για την παραγωγή αναλυτικού κειμένου σχετικά με την ασθένεια και τους τρόπους αντιμετώπισής της. Η δεύτερη προσέγγιση, η οποία θα αναφέρεται ως Specialized Multimodal Expert System (SMES), βασίζεται σε fine-tuning πολυτροπικών μοντέλων τύπου Vision LLM με τη χρήση ενός προσαρμοσμένου συνόλου δεδομένων, ενσωματωμένων σε ένα σύστημα πολλαπλών εξειδικευμένων agents (Mixture of Experts – MoE). Και τα δύο συστήματα σχεδιάστηκαν με στόχο την παραγωγή, όχι μόνο ακριβών προβλέψεων, αλλά και επεξηγηματικού κειμένου που να καθοδηγεί τον χρήστη στην κατανόηση και διαχείριση της ασθένειας.

3.1 Vision-Guided Retrieval System (VGRS)

Το σύστημα αυτό χωρίζεται σε τρία διακριτά στάδια: την αναγνώριση και ανάλυση της εικόνας του φύλλου, την ανάκτηση πληροφοριών από τη βάση δεδομένων και τη συγχώνευση των γνώσεων για τη δημιουργία της τελικής ανάλυσης. Κάθε φάση περιγράφεται αναλυτικά παρακάτω:

3.1.1 Ανάλυση Εικόνας

Σε αυτό το μέρος, στόχος είναι η ανάλυση της εικόνας εισόδου με σκοπό την αναγνώριση της πιο πιθανής ασθένειας, βάσει των οπτικών χαρακτηριστικών του φύλλου. Για την επίτευξη αυτού του στόχου ακολουθείται το εξής pipeline:

1. **Έλεγχος εγκυρότητας εισόδου:** Αρχικά, η εικόνα αναλύεται από έναν **fine-tuned YOLO classifier**. Ο ταξινομητής αποφασίζει εάν η εικόνα περιέχει πράγματι ένα φύλλο και όχι κάποιο άσχετο αντικείμενο. Αυτός ο πρώτος έλεγχος γίνεται ώστε

το σύστημα να μπορεί να χρησιμοποιηθεί σε γενικές συνθήκες, ανεξαρτήτως είδους εισόδου. Αφού διαπιστωθεί ότι η εικόνα περιέχει φύλλο, συνεχίζουμε στο επόμενο βήμα.

2. **Αναγνώριση τύπου φυτού:** Στη συνέχεια η εικόνα αναλύεται από έναν **finetuned YOLO detector**, ο οποίος αναγνωρίζει το είδος του φυτού που απεικονίζεται (π.χ. μηλιά, αμπέλι κ.λπ.). Οι ασθένειες εκδηλώνονται διαφορετικά ανάλογα με το είδος του φυτού, επομένως ο εντοπισμός του τύπου του δέντρου είναι σημαντικός για να πετύχουμε καλή ακρίβεια στο επόμενο κομμάτι της ανάλυσης.
3. **Ανίχνευση ασθένειας:** Τέλος, με βάση τον τύπο του δέντρου που αναγνωρίστηκε στο προηγούμενο βήμα, επιλέγεται ένας δεύτερος **finetuned YOLO detector**. Το μοντέλο αυτό εντοπίζει την πιθανή ασθένεια του φύλλου, εφόσον υπάρχει.

Με τη χρήση του παραπάνω pipeline, καταλήγουμε να γνωρίζουμε τόσο τον τύπο του δέντρου, όσο και την ασθένεια που εμφανίζεται. Σημειώνεται ότι το VGRS έχει σχεδιαστεί έτσι ώστε ο χρήστης να μπορεί, εφόσον το επιθυμεί, να δηλώσει απευθείας την ύπαρξη φύλλου, τον τύπο του δέντρου, ακόμη και την ασθένεια. Σε αυτή την περίπτωση, το σύστημα λειτουργεί περισσότερο ως ψηφιακός γεωπόνος που προτείνει τρόπους αντιμετώπισης.

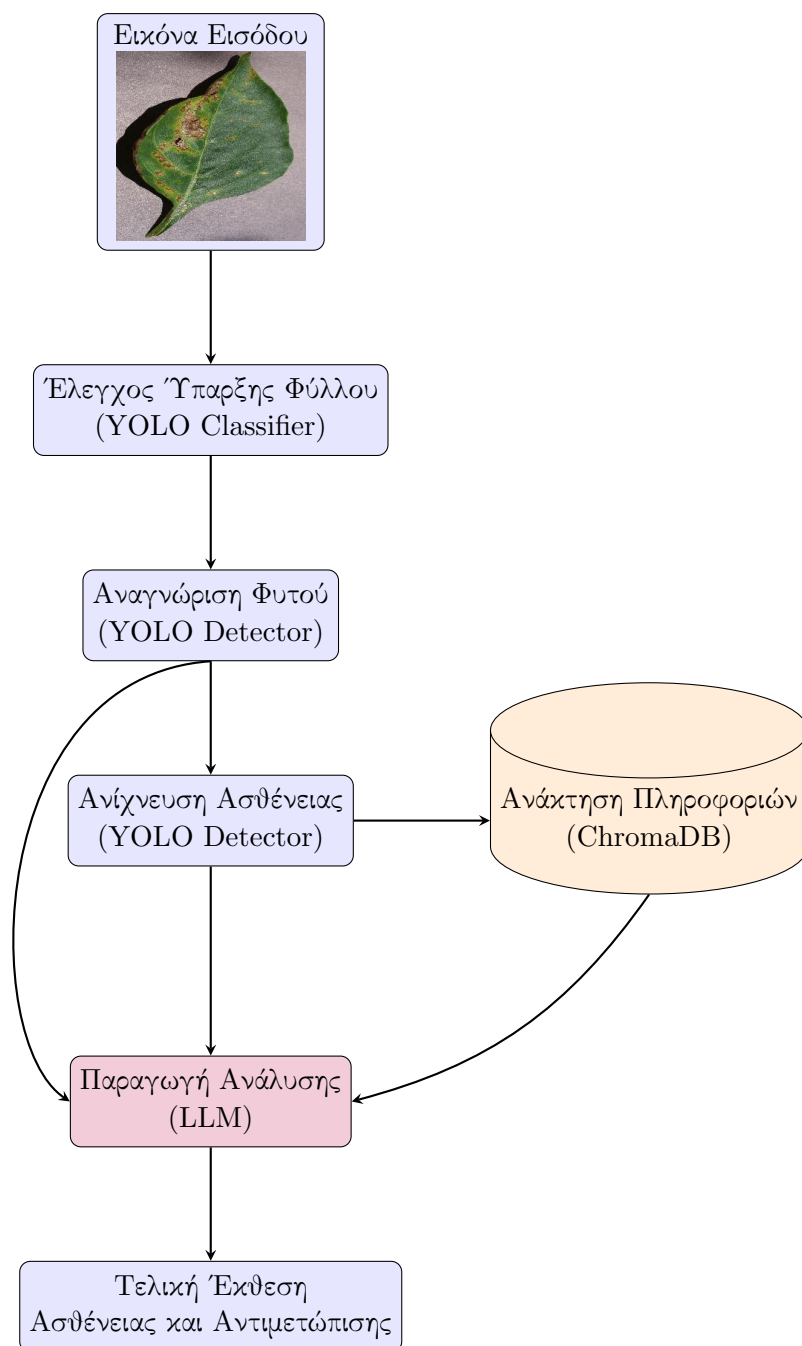
3.1.2 Ανάκτηση Πληροφοριών από τη Βάση Δεδομένων

Αφού εντοπιστεί η ασθένεια, περνάμε στη φάση της ανάκτησης πληροφοριών. Το σύστημα αναζητά, σε μια διανυσματική βάση δεδομένων (ChromaDB), πληροφορίες σχετικά με τα χαρακτηριστικά της ασθένειας (όπως αυτά αποτυπώνονται οπτικά στο φύλλο) και τρόπους αντιμετώπισής της. Πιο συγκεκριμένα, εκτελούνται δύο αναζητήσεις στη βάση: μία για τα χαρακτηριστικά της ασθένειας και μία για τις μεθόδους αντιμετώπισης. Από κάθε αναζήτηση, επιστρέφονται τα τρία πιο σημασιολογικά συναφή αποσπάσματα κειμένου, τα οποία θα χρησιμοποιηθούν στη συνέχεια ως γνώση για τη δημιουργία της τελικής ανάλυσης.

3.1.3 Ενοποίηση μέσω Μεγάλου Γλωσσικού Μοντέλου

Σε αυτό το στάδιο, έχουμε πλέον συλλέξει τις εξής πληροφορίες: τον τύπο του δέντρου, την εντοπισμένη ασθένεια, τα βασικά χαρακτηριστικά της και τους προτεινόμενους τρόπους αντιμετώπισης. Ζητείται λοιπόν από ένα προεκπαιδευμένο γλωσσικό μοντέλο (LLM) να παραγάγει μία συνεκτική ανάλυση, η οποία εξηγεί στον χρήστη την ασθένεια και προτείνει μεθόδους αντιμετώπισης.

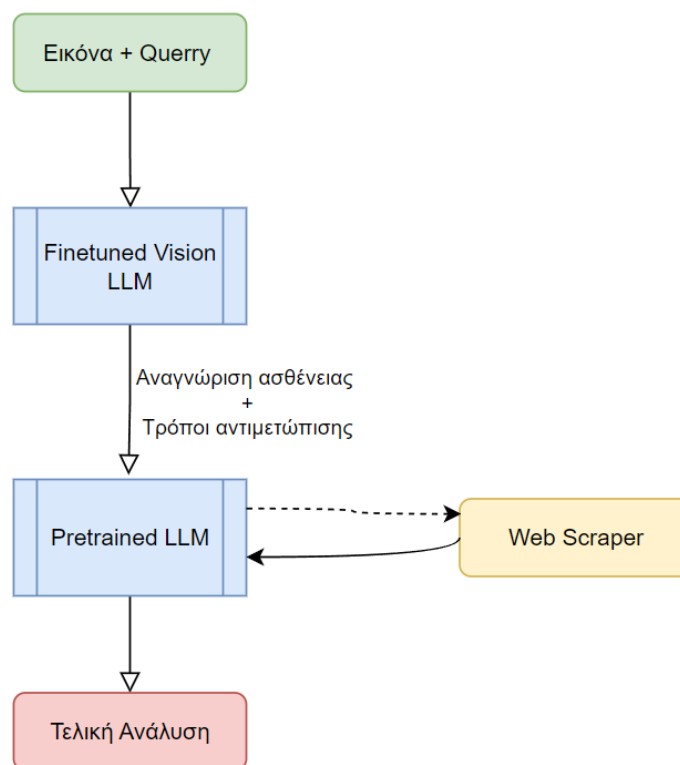
Αξίζει να σημειωθεί ότι δεν ζητείται από το μοντέλο να βασιστεί στις προϋπάρχουσες γνώσεις του, αλλά αποκλειστικά στις πληροφορίες που έχουν ανακτηθεί από τη βάση δεδομένων. Το LLM, συνεπώς, δεν λειτουργεί σαν γεωπόνος που απαντά σε μία ερώτηση, αλλά σαν συγγραφέας/αναλυτής που χρησιμοποιεί συγκεκριμένα δεδομένα για να παραγάγει μία ενοποιημένη και τεκμηριωμένη απάντηση.



Σχήμα 3.1: Αρχιτεκτονική του Vision-Guided Retrieval System (VGRS)

3.2 Πολυτροπική Προσέγγιση – Multimodal Expert System (MES)

Σε αυτό το σύστημα [Σχήμα 3.2], βασικό ρόλο διαδραματίζει ένα finetuned Vision LLM, ενώ γίνεται και εισαγωγική αξιοποίηση μηχανισμών tool calling μέσω γλωσσικού μοντέλου. Η προτεινόμενη αρχιτεκτονική δομείται σε τρία διακριτά στάδια:



Σχήμα 3.2: Flowchart για τη λειτουργία του MES συστήματος

3.2.1 Οπτική Αναγνώριση Χαρακτηριστικών

Η είσοδος του συστήματος αποτελείται από μία εικόνα φυλλώματος (π.χ. φωτογραφία φύλλου) και μία πιθανή ερώτηση χρήστη σχετικά με το περιεχόμενο της εικόνας (query). Η είσοδος αυτή προωθείται αρχικά σε ένα **εκπαιδευμένο Vision LLM**, το οποίο έχει γίνει fine-tuned για να αναγνωρίζει οπτικά χαρακτηριστικά ασθενειών και να επιστρέφει τόσο το όνομα της πιθανής ασθένειας, όσο και σύντομες περιγραφές σχετικά με την εμφάνισή της και πιθανούς τρόπους αντιμετώπισης. Το ερώτημα του χρήστη αξιοποιείται για να καθοδηγήσει το σύστημα, είτε ως προς τον επιθυμητό τρόπο αντιμετώπισης της ασθένειας — για παράδειγμα με φυσικές μεθόδους, χωρίς χρήση φαρμάκων ή με βιολογικά σκευάσματα — είτε ως προς την αναγνώριση του είδους του φυτού και της πιθανής ασθένειας που απεικονίζεται στην εικόνα.

3.2.2 Tool Calling Μηχανισμός

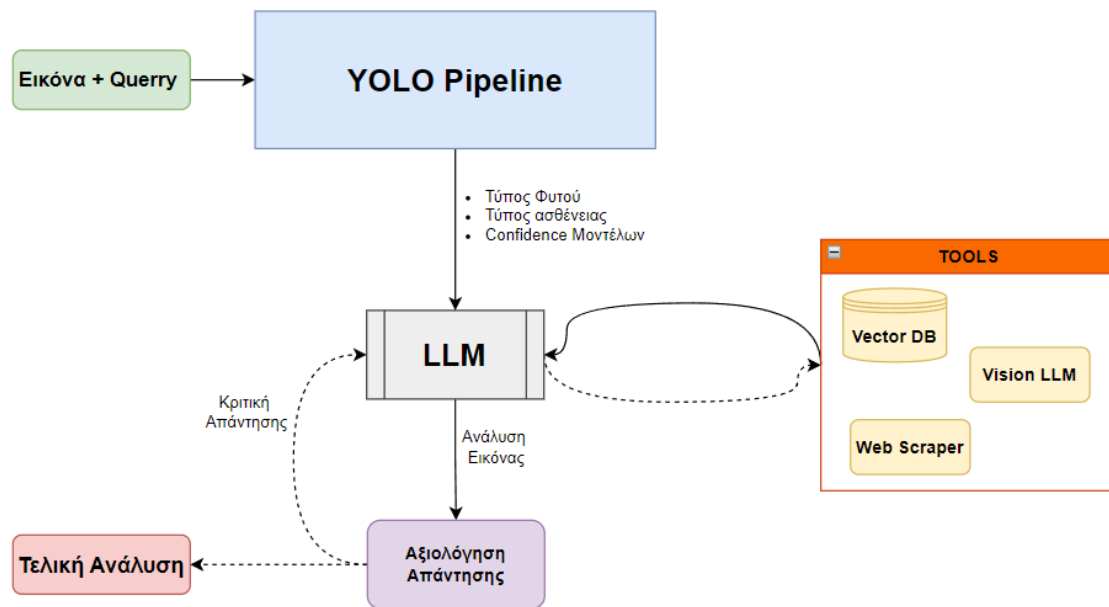
Το αποτέλεσμα από το Vision LLM προωθείται σε ένα **προεκπαιδευμένο γλωσσικό μοντέλο** (Pretrained LLM), το οποίο αναλαμβάνει να εμπλουτίσει την παραπάνω πληροφορία με βάση την ερώτηση του χρήστη και το υπάρχον διαγνωστικό αποτέλεσμα. Το LLM καλείται να χρησιμοποιήσει ένα **μηχανισμός ανάκτησης πληροφορίας μέσω Web Scraper**, ο οποίος συλλέγει επιπρόσθετες πληροφορίες από εξειδικευμένες πηγές διαδικτύου (π.χ. γεωπονικές βάσεις γνώσεων, δημόσια έγγραφα, ιστοσελίδες πανεπιστημίων). Οι πληροφορίες αυτές ενσωματώνονται στο τελικό υπόβαθρο γνώσης που θα αξιοποιήσει το γλωσσικό

μοντέλο για να διαμορφώσει την απάντησή του.

3.2.3 Σύνθεση Απάντησης

Το τελικό στάδιο περιλαμβάνει τη **σύνθεση της απάντησης**, η οποία περιλαμβάνει διάγνωση, περιγραφή της ασθένειας, προτεινόμενα μέτρα αντιμετώπισης και πρόσθετες πληροφορίες που κρίνονται χρήσιμες για τον χρήστη. Το αποτέλεσμα είναι μία κατανοητή και εμπειριστατωμένη αναφορά, βασισμένη τόσο σε οπτική αναγνώριση, όσο και σε γλωσσική επεξεργασία.

3.3 Τελική Προσέγγιση - Vision Enhanced Analysis Agent (VEAA)



Σχήμα 3.3: Αρχιτεκτονική τελικού συστήματος

Τέλος, προτείνεται μια πολυσταδιακή πολυτροπική αρχιτεκτονική, η οποία συνδυάζει τις παραπάνω προσεγγίσεις για την αυτοματοποιημένη αναγνώριση και αντιμετώπιση ασθενειών σε φυτά.

Το σύστημα που σχεδιάστηκε [Σχήμα 3.3] αποτελεί μια πολυσταδιακή αρχιτεκτονική διάγνωσης και αντιμετώπισης ασθενειών σε φυτά, αξιοποιώντας τόσο μοντέλα υπολογιστικής όρασης, όσο και μεγάλες γλωσσικές μονάδες (LLMs). Στόχος του συστήματος είναι να παρέχει μια πλήρως αυτοματοποιημένη ροή από την ανάλυση εικόνας μέχρι την πρόταση τρόπων αντιμετώπισης της εκάστοτε ασθένειας.

Η διαδικασία ξεκινά με την είσοδο μιας εικόνας που περιέχει προσβεβλημένο φύλλο, καθώς

και μια επιπρόσθετη query από τον χρήστη. Η εικόνα αυτή εισέρχεται στον YOLO Pipeline, ο οποίος έχει εκπαιδευτεί για εντοπισμό ασθενειών σε φυτικά φύλλα. Ο YOLO εντοπίζει περιοχές ενδιαφέροντος και εξάγει τις εξής βασικές πληροφορίες: (i) Τύπος φυτού, (ii) Τύπος ασθένειας, (iii) Confidence score του μοντέλου για κάθε αναγνώριση. Οι πληροφορίες αυτές περνούν σε ένα κεντρικό μεγάλο γλωσσικό μοντέλο (LLM), το οποίο έχει πρόσβαση σε ένα σύνολο εργαλείων (tools) μέσω μηχανισμού tool calling. Τα εργαλεία που μπορούν να ενεργοποιηθούν περιλαμβάνουν:

1. Ένα Vector Database που περιέχει επιμελημένη γνώση για ασθένειες και θεραπείες
2. Ένα εξωτερικό Vision LLM για εμπλουτισμένη οπτική ανάλυση, αν αυτό απαιτηθεί
3. Ένα Web Scraper για άντληση επιπρόσθετων πληροφοριών από τον ιστό

Αφού το LLM ανακτήσει και αναλύσει τα απαραίτητα δεδομένα, δημιουργεί μια αρχική απάντηση στην οποία εξηγεί το πρόβλημα και προτείνει πιθανές λύσεις. Η απάντηση αυτή στη συνέχεια περνά από μια φάση **αξιολόγησης**, όπου γίνεται ανάλυση της πληρότητας, της ακρίβειας και της χρήσης κατάλληλων πηγών. Εάν κριθεί απαραίτητο, η απάντηση παράγεται ξανά από το κεντρικό γλωσσικό μοντέλο χρησιμοποιώντας ένα σετ οδηγιών για την σύνθεση μιας πιο ολοκληρωμένης απάντησης. Η ροή αυτή συνδυάζει αλγορίθμους υπολογιστικής όρασης με σύγχρονες τεχνικές επεξεργασίας φυσικής γλώσσας, ενισχύοντας την ακρίβεια, την επεξηγηματικότητα και τη χρηστικότητα του συστήματος.

Κεφάλαιο 4

Πειραματικό Μέρος

Εδώ περιγράφεται η διαδικασία που ακολουθήθηκε για να καταλήξουμε στην υλοποίηση των συστημάτων που περιγράφηκαν στο προηγούμενο κεφάλαιο, καθώς και τις διαδικασίες αξιολόγησης των συστημάτων αυτών. Το σύστημα που χρησιμοποιήθηκε για την εκπαίδευση και την αξιολόγηση των μοντέλων διαθέτει τέσσερις Tesla V100-SXM2-16GB κάρτες γραφικών με συνολική διαθέσιμη μνήμη 64GB.

4.1 Μετρικές Αξιολόγησης

Κατά τη διάρκεια της πειραματικής διαδικασίας, είτε αυτή αφορούσε την εκπαίδευση με διαφορετικά μοντέλα και διαφορετικά σύνολα δεδομένων, είτε την αξιολόγηση των παραγόμενων μοντέλων και των συνολικών συστημάτων, χρησιμοποιήθηκαν συγκεκριμένες μετρικές. Οι μετρικές αυτές συνέβαλαν στην κατανόηση της απόδοσης, στην αξιολόγηση της ορθής λειτουργίας των συστημάτων και στη γενικότερη παρακολούθηση της προόδου. Πριν προχωρήσουμε στην αναλυτική παρουσίαση της πειραματικής διαδικασίας, κρίνεται απαραίτητο να οριστούν οι μετρικές που χρησιμοποιήθηκαν.

4.1.1 Καμπύλη Precision–Recall

Η καμπύλη Precision–Recall αποτελεί ένα βασικό εργαλείο για την αξιολόγηση της απόδοσης ενός μοντέλου ανίχνευσης ή ταξινόμησης. Ας ξεκινήσουμε με το να ορίσουμε τι είναι το Precision και τι το Recall:

Precision (Ακρίβεια)

Η ακρίβεια μετρά το ποσοστό των ανιχνεύσεων που είναι σωστές ως προς το σύνολο των ανιχνεύσεων:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}.$$

όπου:

- TP → True Positives, ο αριθμός των σωστών εντοπισμών που έχει κάνει το μοντέλο.

- FP → False Positives, ο αριθμός των εντοπισμών που έχει κάνει το μοντέλο σε σημείο όπου στην πραγματικότητα δεν υπάρχει κάτι που πρέπει να εντοπιστεί.

Υψηλή τιμή σημαίνει ότι υπάρχουν λίγες ψευδείς ανιχνεύσεις (*false positives*).

Recall (Ανακλησιμότητα)

Η ανακλησιμότητα μετρά το ποσοστό των πραγματικών αντικειμένων που ανιχνεύθηκαν ως προς το σύνολο των πραγματικών αντικειμένων:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

όπου FN → False Negatives, ο αριθμός των αποτυχημένων εντοπισμών αντικειμένων στις εικόνες. Υψηλή τιμή recall σημαίνει λίγα false negatives.

PR–Curve

Η καμπύλη Precision–Recall προκύπτει από τη μεταβολή ενός κατωφλίου εμπιστοσύνης (π.χ. **confidence threshold**) για τις προβλέψεις του μοντέλου και την απεικόνιση της σχέσης μεταξύ Precision και Recall για κάθε τιμή του.

Μια καλή απόδοση του μοντέλου φαίνεται όταν η καμπύλη διατηρεί υψηλές τιμές Precision ακόμη και για μεγάλες τιμές Recall. Η περιοχή κάτω από την καμπύλη αυτή ονομάζεται Average Precision (AP) και χρησιμοποιείται για τον υπολογισμό της mAP.

4.1.2 IoU (Intersection over Union)

Η επικάλυψη μεταξύ προτεινόμενου και πραγματικού *bounding box*:

$$\text{IoU} = \frac{|B_{\text{pred}} \cap B_{\text{gt}}|}{|B_{\text{pred}} \cup B_{\text{gt}}|}.$$

Χρησιμοποιείται ως κατώφλι για να χαρακτηριστεί μια ανίχνευση ως σωστή.

4.1.3 mAP@0.5

Η μετρική mAP@0.5 βασίζεται στην έννοια του Average Precision (AP), η οποία αντιπροσωπεύει το εμβαδό κάτω από την καμπύλη Precision–Recall. Συγκεκριμένα, η AP μετρά πόσο καλά ισορροπεί το μοντέλο μεταξύ ακρίβειας και ανακλησιμότητας για μία συγκεκριμένη κλάση.

Η mAP@0.5 (mean Average Precision στο κατώφλι IoU = 0.5) υπολογίζεται ως ο μέσος όρος των τιμών AP για όλες τις κλάσεις, θεωρώντας μια ανίχνευση σωστή, όταν η επικάλυψη μεταξύ του προβλεπόμενου και του πραγματικού πλαισίου (IoU) είναι τουλάχιστον 0.5:

$$mAP@0.5 = \frac{1}{C} \sum_{c=1}^C AP_c^{IoU=0.5}$$

όπου C είναι ο συνολικός αριθμός των κλάσεων. Όσο μεγαλύτερη είναι η τιμή της mAP@0.5, τόσο καλύτερη είναι η απόδοση του μοντέλου.

4.1.4 mAP@0.5:0.95

Η mAP@0.5:0.95 είναι μια πιο αυστηρή και ολοκληρωμένη μετρική αξιολόγησης, η οποία ακολουθεί το πρότυπο του COCO benchmark. Αντί να χρησιμοποιεί μόνο μία τιμή επικάλυψης (IoU = 0.5), υπολογίζει τη μέση Average Precision για κάθε κλάση σε δέκα διαφορετικά κατώφλια IoU από 0.50 έως 0.95, με βήμα 0.05.

Στη συνέχεια, λαμβάνεται ο μέσος όρος αυτών των τιμών για κάθε κλάση και έπειτα ο μέσος όρος για όλες τις κλάσεις:

$$mAP@0.5 : 0.95 = \frac{1}{C} \sum_{c=1}^C \left(\frac{1}{10} \sum_{t \in \{0.50, 0.55, \dots, 0.95\}} AP_c^{IoU=t} \right)$$

Αυτή η μετρική αξιολογεί πιο αυστηρά τόσο την ακρίβεια της ταξινόμησης όσο και την ποιότητα της εντοπισμένης τοποθεσίας (πλασιού) για κάθε αντικείμενο. Όσο μεγαλύτερη είναι η τιμή της mAP@0.5:0.95, τόσο πιο αξιόπιστο θεωρείται το μοντέλο.

4.1.5 Ο Πίνακας Confusion Matrix

Ο πίνακας Confusion Matrix αποτελεί ένα θεμελιώδες εργαλείο στη μηχανική μάθηση για την αξιολόγηση της απόδοσης ενός αλγορίθμου ταξινόμησης. Σε αντίθεση με έναν απλό δείκτη ακρίβειας (accuracy score) που δείχνει μόνο το ποσοστό των σωστών προβλέψεων, ο πίνακας Confusion Matrix παρέχει μια λεπτομερή αποτύπωση της συμπεριφοράς του μοντέλου για κάθε επιμέρους κλάση. Δεν δείχνει μόνο πότε το μοντέλο κάνει σωστές προβλέψεις, αλλά και πώς κάνει λάθος, αποκαλύπτοντας πού εντοπίζεται η «σύγχυση» του μοντέλου. Αυτός ο βαθμός λεπτομέρειας είναι ιδιαίτερα σημαντικός σε προβλήματα επιβλεπόμενης μάθησης (supervised learning), όπως η ταξινόμηση εικόνων (image classification) και η ανίχνευση αντικειμένων (object detection).

4.1.6 F1 Score

Η μετρική F1 Score αποτελεί έναν αρμονικό μέσο όρο των μετρικών Precision και Recall, και χρησιμοποιείται ευρέως για την αξιολόγηση της συνολικής απόδοσης ενός μοντέλου, ειδικά όταν υπάρχει ανισορροπία μεταξύ των κλάσεων. Εκφράζει μια ισορροπημένη μέτρηση της ακρίβειας και της ανακλησιμότητας, δίνοντας μεγαλύτερη βαρύτητα στα χαμηλά αποτελέσματα — δηλαδή, ένα μοντέλο θα έχει υψηλό F1 Score μόνο όταν και οι δύο αυτές μετρικές είναι υψηλές.

Ο τύπος υπολογισμού του F1 Score είναι:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Η τιμή του F1 Score κυμαίνεται μεταξύ 0 και 1, όπου το 1 υποδηλώνει την ιδανική ισορροπία μεταξύ Precision και Recall.

4.1.7 Context Precision

Η μετρική Context Precision χρησιμοποιείται για την αξιολόγηση της απόδοσης ενός συστήματος ανάκτησης πληροφορίας (retriever) στο πλαίσιο συστημάτων RAG (Retrieval-Augmented Generation). Συγκεκριμένα, μετρά πόσο αποτελεσματικά το σύστημα φέρνει πρώτα τα σχετικά αποσπάσματα (relevant chunks) πληροφορίας μέσα από τα συνολικά ανακτηθέντα, για μια συγκεκριμένη ερώτηση (query).

Η μετρική υπολογίζεται ως ο μέσος όρος της $Precision@k$ για κάθε απόσπασμα που ανακτήθηκε από την βάση γνώσης (retrieved context). Η $Precision@k$ εκφράζει το ποσοστό των σχετικών αποσπασμάτων στη θέση k σε σχέση με το σύνολο των αποσπασμάτων μέχρι εκείνη τη θέση. Ο πλήρης τύπος για την $ContextPrecision@k$ είναι:

$$ContextPrecision@K = \frac{\sum_{k=1}^K (Precision@k \times v_k)}{\text{Συνολικός αριθμός σχετικών αποσπασμάτων στις πρώτες } k \text{ θέσεις}}$$

όπου ο επιμέρους όρος $Precision@k$ δίνεται από:

$$Precision@k = \frac{true\ positives@k}{true\ positives@k + false\ positives@k}$$

με:

- K να είναι το συνολικό πλήθος των αποσπασμάτων στο retrieved context,
- $v_k \in \{0, 1\}$ να είναι ο δείκτης συνάφειας (αν το απόσπασμα στη θέση k είναι σχετικό ή όχι).

Η $ContextPrecision@k$ λαμβάνει υπόψη τόσο την παρουσία, όσο και τη θέση των σχετικών αποσπασμάτων, ενισχύοντας την αξιοπιστία της αξιολόγησης στην ανάκτηση πληροφορίας.

4.1.8 Context Recall με χρήση LLM

Η μετρική Context Recall με χρήση LLM αξιολογεί την πληρότητα των ανακτηθέντων πληροφοριών με βάση το κατά πόσο καλύπτουν το περιεχόμενο της απάντησης που κατά τη διάρκεια της αξιολόγησης θεωρείται το σημείο αναφοράς (reference answer). Η διαδικασία βασίζεται στη διάσπαση της αναφοράς σε επιμέρους δηλώσεις (claims), και στη συνέχεια τον έλεγχο για το αν αυτές οι δηλώσεις υποστηρίζονται από το ανακτηθέν πλαίσιο (retrieved context).

Ο τύπος υπολογισμού είναι:

$$ContextRecall_{LLM} = \frac{\text{Αριθμός claims που καλύπτονται από το ανακτηθέν πλαίσιο}}{\text{Συνολικός αριθμός claims στην αναφορά}}$$

Όσο υψηλότερη είναι η τιμή της Context Recall, τόσο πιο πλήρες θεωρείται το ανακτηθέν περιεχόμενο σε σχέση με την αρχική πληροφορία που πρέπει να καλυφθεί. Είναι μία κρίσιμη μετρική σε εφαρμογές RAG, καθώς διασφαλίζει ότι σημαντικά στοιχεία της γνώσης δεν παραλείπονται από την τελική απάντηση.

4.1.9 Faithfulness

Η μετρική Faithfulness μετρά πόσο οι ισχυρισμοί (claims) που περιέχονται στην απάντηση ενός συστήματος RAG συμφωνούν με το ανακτηθέν πλαίσιο πληροφορίας (retrieved context). Η τιμή της κυμαίνεται από 0 έως 1, όπου υψηλότερες τιμές σημαίνουν καλύτερη συνοχή και λιγότερους ψευδείς ή ανακριβείς ισχυρισμούς.

Για να υπολογιστεί:

1. Εντοπίζονται όλοι οι ισχυρισμοί που κάνει η απάντηση του μοντέλου.
2. Για κάθε ισχυρισμό ελέγχεται αν μπορεί να εξαχθεί ή να υποστηριχτεί από το ανακτηθέν πλαίσιο (retrieved context).
3. Το Faithfulness Score υπολογίζεται ως:

$$\frac{\text{Αριθμός των ισχυρισμών της απάντησης που υποστηρίζονται από το retrieved context}}{\text{Συνολικός αριθμός των ισχυρισμών στην απάντηση}}$$

Μια απάντηση θεωρείται faithful αν **όλοι** οι ισχυρισμοί της απάντησης μπορούν να υποστηριχθούν από τα δεδομένα του πλαισίου.

4.1.10 Answer Relevance

Η μετρική Answer Relevance αξιολογεί το κατά πόσο η παραγόμενη απάντηση είναι σχετική με την αρχική ερώτηση (question). Δεν εστιάζει στην αλήθεια της απάντησης, αλλά στην πληρότητα και στη συνάφεια — τιμωρεί απαντήσεις που είναι ελλιπείς ή περιλαμβάνουν περιττές πληροφορίες.

Το Answer Relevance υπολογίζεται ως ο μέσος όρος της cosine similarity μεταξύ της αρχικής ερώτησης και διάφορων “παραγόμενων” ερωτήσεων που δημιουργούνται με βάση την απάντηση:

$$AnswerRelevance = \frac{1}{N} \sum_{i=1}^N \cos(E_{g_i}, E_o) = \frac{1}{N} \sum_{i=1}^N \frac{E_{g_i} \cdot E_o}{\|E_{g_i}\| \|E_o\|}$$

όπου:

- E_{g_i} είναι το embedding της i -ης ερώτησης που γεννήθηκε από την απάντηση,
- E_o είναι το embedding της αρχικής ερώτησης (original question),
- N είναι ο αριθμός των παραγόμενων ερωτήσεων (προεπιλογή $N = 3$).

Σημειώνεται ότι η μέτρηση της cosine similarity μπορεί θεωρητικά να πάρει τιμές από -1 έως 1, αλλά στην πράξη για αυτή τη χρήση το αποτέλεσμα είναι συνήθως μεταξύ 0 και 1. Μια πιο υψηλή τιμή δείχνει ότι οι παραγόμενες ερωτήσεις που εξάγονται από την απάντηση ευθυγραμμίζονται καλύτερα με την αρχική ερώτηση — άρα η απάντηση είναι πιο σχετική.

4.2 Πειράματα VGRS

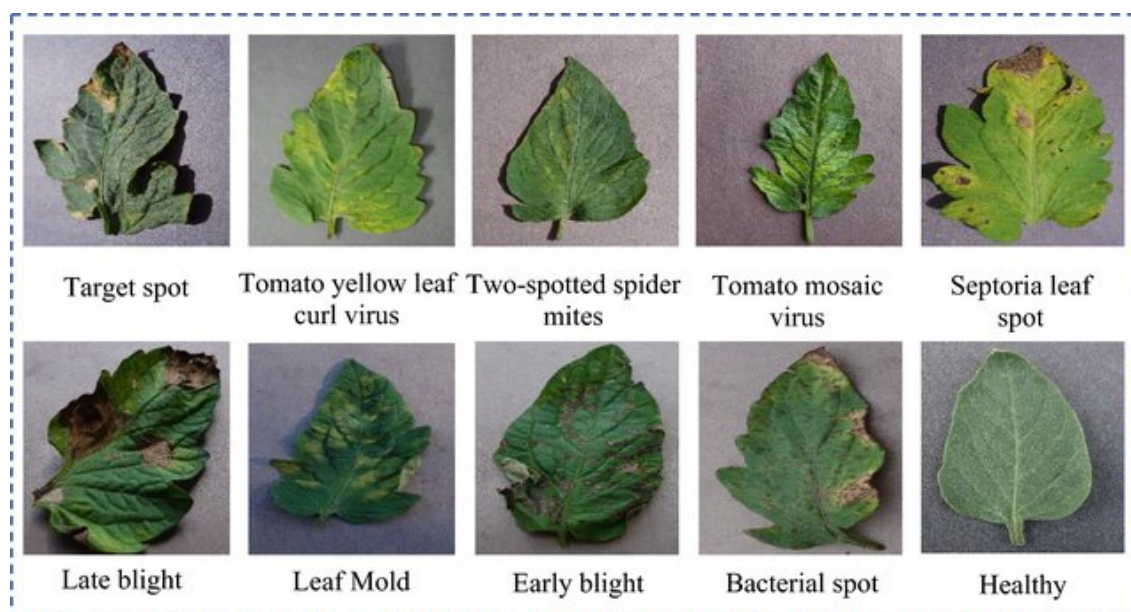
4.2.1 Πρώτα Βήματα

Ένα από τα πιο κρίσιμα στάδια για την ανάπτυξη του συστήματος είναι η επιλογή του κατάλληλου συνόλου δεδομένων για την εκπαίδευση των μοντέλων. Αρχικά, ο στόχος ήταν να εκπαιδευτεί ένα μοντέλο υπολογιστικής όρασης που να μπορεί να αναγνωρίσει ασθένειες οι οποίες επηρεάζουν τα δέντρα, με βάση μία εικόνα του προσβεβλημένου φύλλου. Συνεπώς, το απαιτούμενο σύνολο δεδομένων έπρεπε να περιλαμβάνει **μεγάλο αριθμό εικόνων**, με **σαφή επισήμανση (label)** της ασθένειας που εμφανίζεται σε κάθε μία.

Αφού εξετάστηκαν διάφορες εναλλακτικές, όπως το ImageNet, το Darknet dataset και άλλες μικρότερες συλλογές, καταλήξαμε στη χρήση της αρχιτεκτονικής **YOLO** για το πρόβλημα της αναγνώρισης ασθενειών. Ο λόγος που επιλέχθηκε το YOLO (You Only Look Once) έναντι άλλων μεθόδων είναι η ικανότητά του να εκτελεί ταυτόχρονα εντοπισμό και ταξινόμηση περιοχών ενδιαφέροντος με ένα μόνο πέρασμα από το νευρωνικό δίκτυο, προσφέροντας έτσι σημαντικά πλεονεκτήματα σε επίπεδο ταχύτητας και απόδοσης.

Για την εκπαίδευση του μοντέλου απαιτούνταν ένα πλούσιο, πολυποίκιλο και καλά επισημασμένο σύνολο δεδομένων. Μετά από εκτενή έρευνα, το σύνολο δεδομένων που κρίθηκε καταλληλότερο για τις ανάγκες του συστήματος είναι το PlantVillage dataset [19]. Το συγκεκριμένο σύνολο δημιουργήθηκε από τους David P. Hughes και Marcel Salathé και αποτελεί μία από τις πιο πλήρεις και ευρέως χρησιμοποιούμενες συλλογές για την αναγνώριση ασθενειών σε φύλλα. Περιλαμβάνει 38 κλάσεις και περισσότερες από 50.000 εικόνες από διάφορες καλλιέργειες, όπως η μηλιά, η κερασιά, το καλαμπόκι, η πατάτα, αμπέλια, και η τοματιά, μεταξύ άλλων, και καλύπτει **πληθώρα ασθενειών για κάθε τύπο καλλιέργειας**. Επιπλέον, για κάθε κατηγορία υπάρχουν και **υγιή φύλλα**, γεγονός που επιτρέπει στο μοντέλο να μάθει να διακρίνει μεταξύ υγιών και προσβεβλημένων περιπτώσεων. Το όνομα κάθε κλάσης ακολουθεί τη μορφή *ΤΥΠΟΣ ΦΥΤΟΥ_ΟΝΟΜΑ ΑΣΘΕΝΕΙΑΣ* (π.χ. Apple Black rot, Blueberry healthy, Corn (maize) Cercospora leaf spot Gray leaf spot, κλπ).

Σε σύγκριση με άλλα διαθέσιμα σύνολα δεδομένων, το PlantVillage παρουσίασε σαφή υπεροχή. Πραγματοποιήθηκε έρευνα σε πλατφόρμες, όπως το Hugging Face Datasets Hub, το Kaggle και το Roboflow, οι οποίες φιλοξενούν μεγάλο πλήθος ανοικτών συνόλων δεδομένων για εφαρμογές υπολογιστικής όρασης. Πολλές από τις εναλλακτικές συλλογές που εντοπίστηκαν, είτε περιλάμβαναν πολύ περιορισμένο αριθμό εικόνων, είτε αναφέρονταν αποκλειστικά σε μία καλλιέργεια ή σε μία μόνο ασθένεια, μειώνοντας έτσι τη γενικευσιμότητα των μοντέλων που θα μπορούσαν να εκπαιδευτούν με αυτά. Άλλες αποτελούσαν παραλλαγές ή υποσύνολα του PlantVillage, χωρίς να προσφέρουν επιπλέον πληροφορία ή ουσιαστική διαφοροποίηση. Επιπλέον, σε πλήθος συναφών ερευνητικών εργασιών, το PlantVillage αποτελεί τη συνηθέστερη επιλογή για προβλήματα ταξινόμησης ασθενειών σε φυτά, γεγονός που ενισχύει περαιτέρω την αξιοπιστία και τη χρησιμότητά του. Συνεπώς, το PlantVillage dataset προσέφερε την καλύτερη ισορροπία μεταξύ ποσότητας, ποικιλίας και ποιότητας δεδομένων, καθιστώντας το ιδανική επιλογή για την εκπαίδευση ενός μοντέλου που στοχεύει στην ευρύτερη δυνατή αναγνώριση ασθενειών σε διαφορετικούς τύπους καλλιεργειών.



Σχήμα 4.1: Παραδείγματα Εικόνων από το PlantVillage dataset

4.2.2 Εκπαίδευση Ταξινόμησης στο PlantVillage Dataset

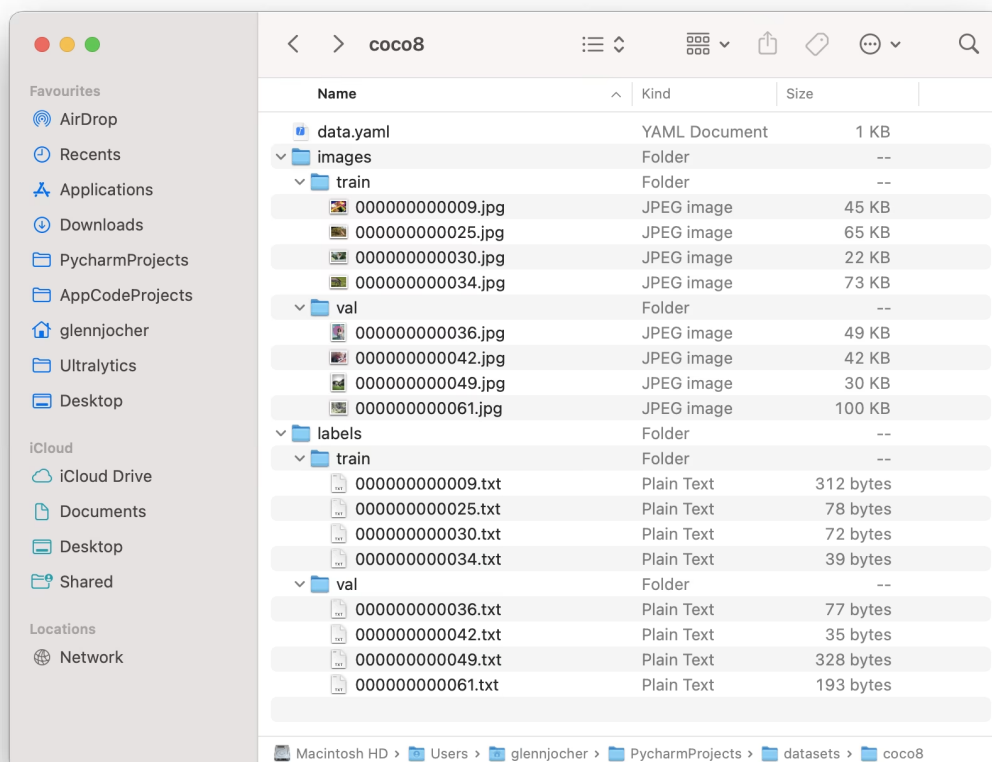
Έχοντας πλέον επιλέξει το κατάλληλο σύνολο δεδομένων, ξεκίνησαν οι πρώτες δοκιμές για την εκπαίδευση του μοντέλου YOLO με στόχο την αναγνώριση ασθενειών μέσω εντοπισμού σε επίπεδο εικόνας. Εφόσον πρόκειται για πρόβλημα object detection, το μοντέλο καλείται να εκτελέσει δύο βασικές λειτουργίες πάνω στην εικόνα εισόδου: πρώτον, να εντοπίσει τη χωρική θέση του φύλλου μέσω παραγωγής ενός κατάλληλου bounding box· και δεύτερον, να ταξινομήσει την προσβεβλημένη περιοχή σε μία από τις διαθέσιμες κατηγορίες ασθενειών ή στην κατηγορία υγιές. Αξίζει να σημειωθεί πως σε αυτή τη φάση η εκπαίδευση πραγματοποιείται πάνω σε ολόκληρο το σύνολο του PlantVillage dataset, το οποίο περιλαμβάνει συνολικά 38 κατηγορίες, που αντιστοιχούν είτε σε συγκεκριμένες ασθένειες είτε σε υγιή φύλλα διαφόρων καλλιεργειών.

Προετοιμασία Dataset

Για να μπορούσαμε να εκπαιδεύσουμε το μοντέλο πρέπει να φέρουμε το σέτ δεδομένων σε μορφή Ultralytics YOLO format. Το πρότυπο αυτό χρησιμοποιείται για την εκπαίδευση αλγορίθμων ανίχνευσης αντικειμένων και οργανώνει τα δεδομένα με συγκεκριμένο τρόπο. Κάθε εικόνα συνοδεύεται από ένα αρχείο .txt, το οποίο περιλαμβάνει μία ή περισσότερες γραμμές, όπου κάθε γραμμή αντιστοιχεί σε ένα αντικείμενο. Η μορφή κάθε γραμμής είναι:

```
class_id x_center y_center width height
```

όπου οι τέσσερις τελευταίες τιμές να είναι κανονικοποιημένες ως προς το πλάτος και το ύψος της εικόνας. Οι εικόνες και οι ετικέτες οργανώνονται σε φακέλους `images/train`, `images/val`, `labels/train`, και `labels/val`, ενώ η περιγραφή του dataset γίνεται μέσω



Σχήμα 4.2: Ultralytics YOLO Dataset Format για Object Detection

ενός αρχείου `.yaml`, το οποίο ορίζει τη διαδρομή των δεδομένων και τα ονόματα των κατηγοριών. Στην παρούσα εργασία, το dataset διαχωρίστηκε, όπως περιγράφηκε ανωτέρω, σε τρία υποσύνολα `train/val/test` με αναλογίες 0.7:0.2:0.1. Κατά τη διαδικασία εκπαίδευσης αξιοποιούνται μόνο τα υποσύνολα `train` και `val`, ενώ το υποσύνολο `test` θα χρησιμοποιηθεί αποκλειστικά για αξιολόγηση.

Εκπαίδευση

Η εκπαίδευση πραγματοποιήθηκε με χρήση της βιβλιοθήκης Ultralytics YOLO, η οποία προσφέρει ένα σύγχρονο και βελτιστοποιημένο περιβάλλον για την ανάπτυξη μοντέλων ανίχνευσης αντικειμένων. Για την εκπαίδευση χρησιμοποιήθηκε το μοντέλο YOLOv11l (large), κατάλληλο για ταχύτερη προπόνηση. Το μέγεθος των εικόνων καθορίστηκε στα 640×640 pixels, η διάρκεια της εκπαίδευσης ήταν 100 εποχές και το μέγεθος της δέσμης δεδομένων (batch size) ορίστηκε σε 16. Η παράμετρος batch size καθορίζει πόσες εικόνες προωθούνται ταυτόχρονα στο νευρωνικό δίκτυο κατά τη διάρκεια κάθε βήματος εκπαίδευσης, επηρεάζοντας τόσο τη μνήμη που απαιτείται, όσο και την ταχύτητα σύγκλισης του μοντέλου. Για την αποτροπή του overfitting, τέθηκε τιμή `patience=6`, που σημαίνει ότι αν δεν παρατηρηθεί βελτίωση στην απόδοση για 6 διαδοχικές εποχές, η εκπαίδευση θα τερματιστεί πρόωρα. Ως βελτιστοποι-

Model	Filenames	Task
YOLO11	yolo11n.pt yolo11s.pt yolo11m.pt yolo11l.pt yolo11x.pt	Detection
YOLO11-seg	yolo11n-seg.pt yolo11s-seg.pt yolo11m-seg.pt yolo11l-seg.pt yolo11x-seg.pt	Instance Segmentation
YOLO11-pose	yolo11n-pose.pt yolo11s-pose.pt yolo11m-pose.pt yolo11l-pose.pt yolo11x-pose.pt	Pose/Keypoints
YOLO11-obb	yolo11n-obb.pt yolo11s-obb.pt yolo11m-obb.pt yolo11l-obb.pt yolo11x-obb.pt	Oriented Detection
YOLO11-cls	yolo11n-cls.pt yolo11s-cls.pt yolo11m-cls.pt yolo11l-cls.pt yolo11x-cls.pt	Classification

Σχήμα 4.3: Λίστα διαθέσιμων YOLOv11 μοντέλων

ητής επιλέχθηκε ο AdamW, μία παραλλαγή του κλασικού Adam optimizer, η οποία εισάγει έναν διαφορετικό τρόπο εφαρμογής της απομείωσης βάρους (decoupled weight decay).

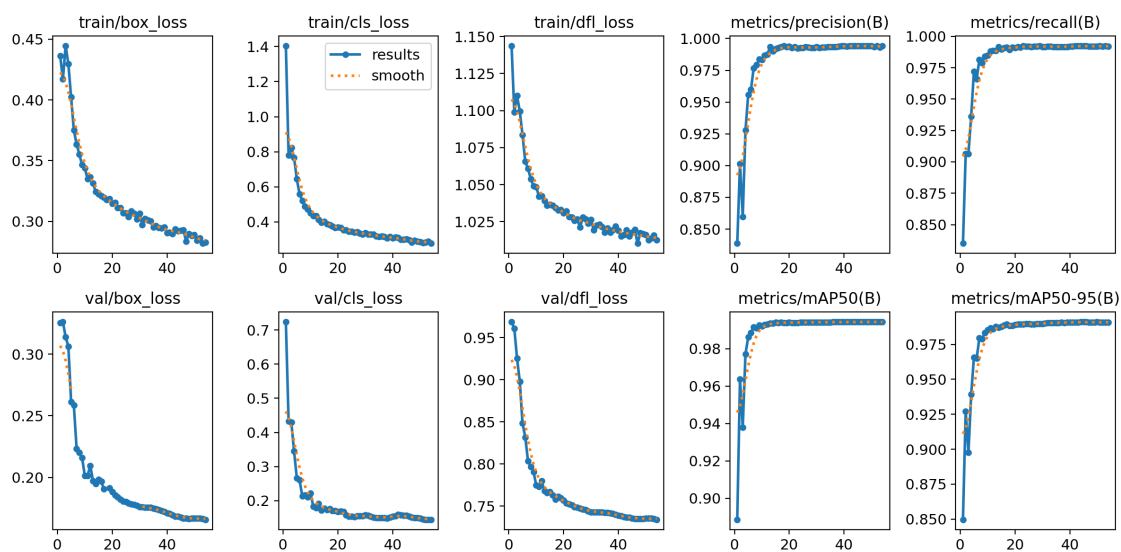
Κατά τη διάρκεια της εκπαίδευσης, το YOLO training pipeline παράγει διάφορες μετρικές και γραφήματα που επιτρέπουν την αξιολόγηση της προόδου και της απόδοσης του μοντέλου. Οι σημαντικότερες μετρικές περιλαμβάνουν την ακρίβεια ανά κατηγορία (per-class precision), την ανάκληση (recall), τη συνολική ακρίβεια (mAP@0.5 και mAP@0.5:0.95), καθώς και τον ρυθμό σύγκλισης του loss function. Παράλληλα, δημιουργούνται γραφήματα που απεικονίζουν την εξέλιξη του training loss και του validation loss, τα οποία βοηθούν στον εντοπισμό φαινομένων όπως το overfitting ή το underfitting.

Όπως φαίνεται στο Σχήμα 4.4 παρόλο που θέσαμε την εκπαίδευση να γίνει για 100 epochs σταμάτησε στις 54, καθώς για 6 εποχές δεν υπήρξε διόρθωση του μοντέλου. Τα train/val box, cls, dfl losses μειώνονται ομαλά και συγχλίνουν σε παρόμοιες τιμές, χωρίς απόκλιση μεταξύ train και val, ένδειξη σταθερής μάθησης και απουσίας έντονου overfitting. Οι precision και recall καμπύλες φτάνουν σε κορεσμό κοντά στο 1.0 μετά τις πρώτες ≈ 10 εποχές, ενώ οι δείκτες mAP@0.5 ≈ 0.99 και mAP@0.5:0.95 ≈ 0.98 υποδηλώνουν πολύ υψηλή απόδοση.

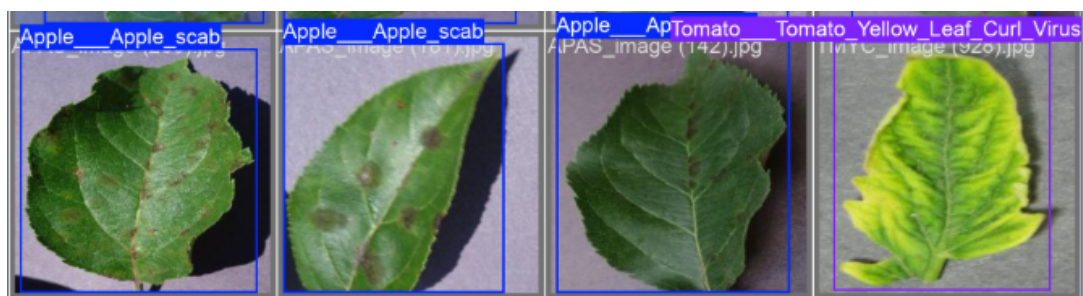
Αξιολόγηση

Προχωρώντας στην αξιολόγηση του μοντέλου με το test split, τα αποτελέσματα παραμένουν ιδιαίτερα ικανοποιητικά.

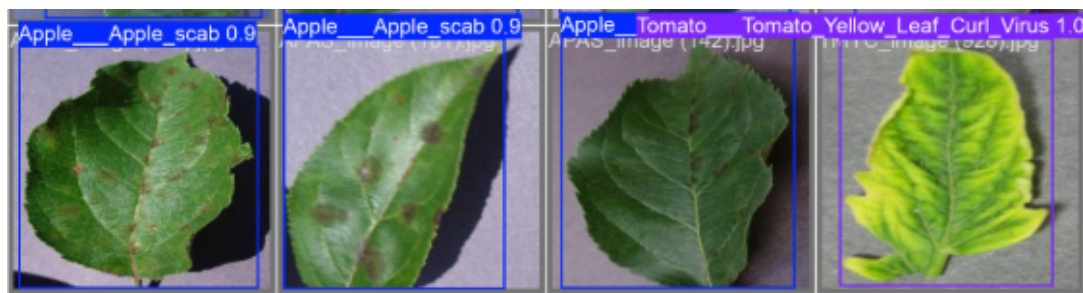
Για την αξιολόγηση του μοντέλου, χρησιμοποιήθηκε η συνάρτηση `val()` της βιβλιοθήκης Ultralytics YOLO, η οποία εφαρμόστηκε πάνω στο test split του συνόλου δεδομένων. Κατά τη διαδικασία, ορίστηκε κατώφλι IoU ίσο με 0.5, ώστε να αυστηροποιήσουμε την αξιολόγηση του μοντέλου



Σχήμα 4.4: Αποτελέσματα εκπαίδευσης Object Detection πάνω σε ολόκληρο το PlantVillage Dataset



(α') Ground Truth για την κλάση και το bounding box



(β') Αποτελέσματα Detection από το εκπαιδευμένο μοντέλο μαζί με το confidence της κάθε πρόβλεψης

Σχήμα 4.5: Παραδείγματα αναγνώρισης φύλλου και ασθένειας

Όπως φαίνεται και στο confusion matrix (Σχήμα 4.6), το μοντέλο παρουσιάζει εξαιρετική απόδοση, γεγονός που αποτυπώνεται στην έντονη διαγώνιο του πίνακα που είναι ένδειξη ότι οι περισσότερες προβλέψεις συμφωνούν με τις πραγματικές κλάσεις. Επιπλέον, έχοντας θέσει το IoU κατώφλι στο 0.5, απαιτείται τουλάχιστον 50% επικάλυψη μεταξύ του παραγόμενου και του πραγματικού bounding box για να θεωρηθεί αποδεκτή η ανίχνευση. Διαφορετικά, η πρόβλεψη απορρίπτεται, και η εικόνα θεωρείται ως μη περιέχουσα φύλλο. Το γεγονός αυτό εξηγεί την παρουσία αχνών γαλάζιων κουκκίδων στη στήλη του background, οι οποίες υποδηλώνουν τέτοιες περιπτώσεις.

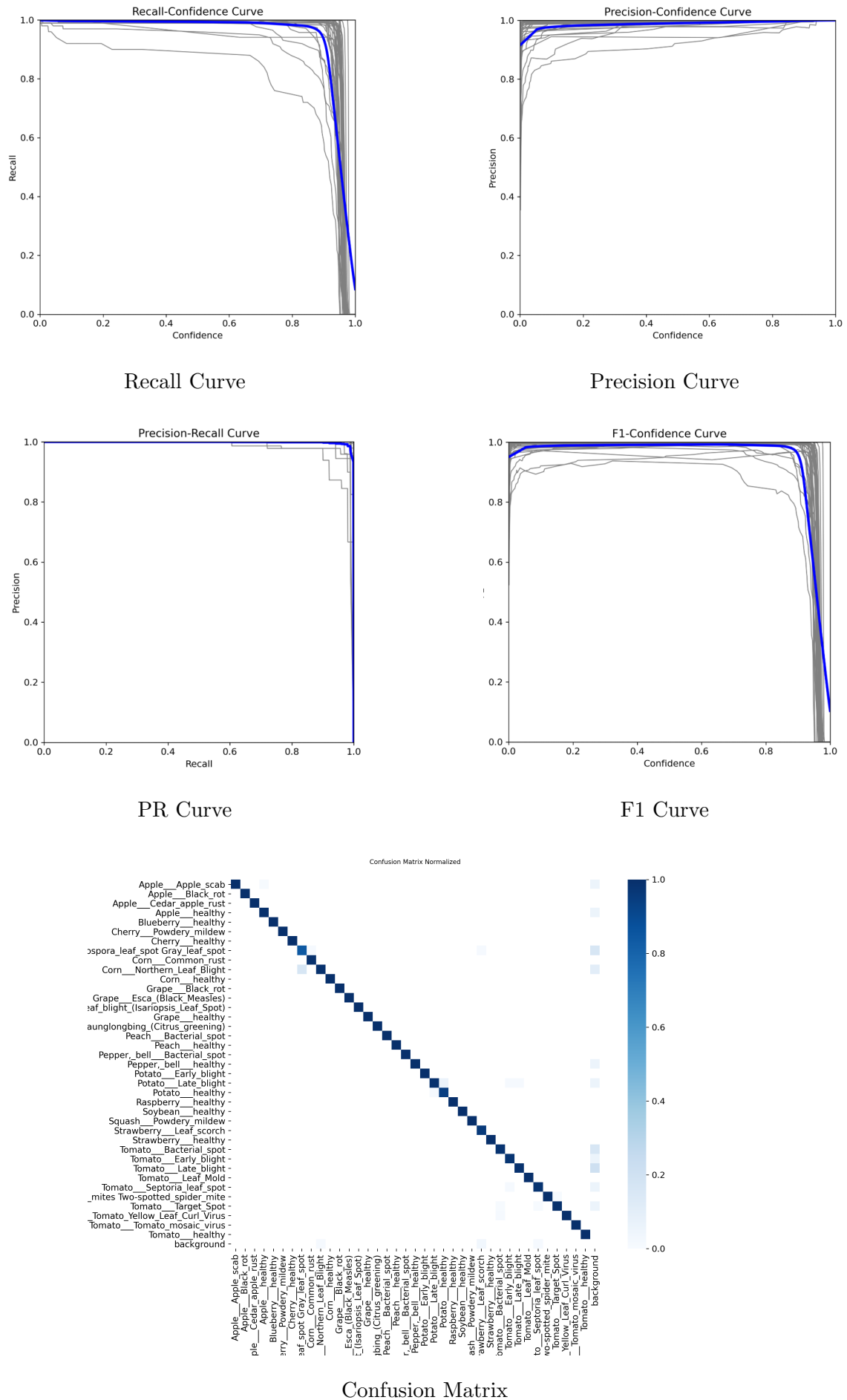
Αναλύοντας περαιτέρω τις γραφικές παραστάσεις αξιολόγησης στο Σχήμα 4.6, παρατηρείται ότι τόσο η Precision-Confidence Curve όσο και η Recall-Confidence Curve εμφανίζουν ιδιαίτερα υψηλές τιμές (κοντά στο 1.0) για μεγάλο εύρος τιμών εμπιστοσύνης (confidence), κάτι που αποδεικνύει τη βεβαιότητα του μοντέλου στις προβλέψεις του. Ειδικά στην καμπύλη Precision-Confidence, η ακρίβεια παραμένει σταθερά υψηλή ακόμα και σε χαμηλές τιμές εμπιστοσύνης, γεγονός που υποδεικνύει χαμηλό ποσοστό ψευδώς θετικών. Παράλληλα, η καμπύλη F1-Confidence, που συνδυάζει αρμονικά την ακρίβεια και την ανάκληση, παρουσιάζει κορυφή στο 0.99 για confidence 0.648, υποδεικνύοντας ότι σε αυτό το κατώφλι το μοντέλο επιτυγχάνει ιδανική ισορροπία μεταξύ σωστών ανιχνεύσεων και ελαχιστοποίησης λαθών. Συνολικά, οι καμπύλες δείχνουν ότι το μοντέλο δεν είναι μόνο ακριβές, αλλά και συνεπές, με υψηλή αξιοπιστία στις προβλέψεις του.

4.2.3 Αξιολόγηση με χρήση του PlantDoc Dataset

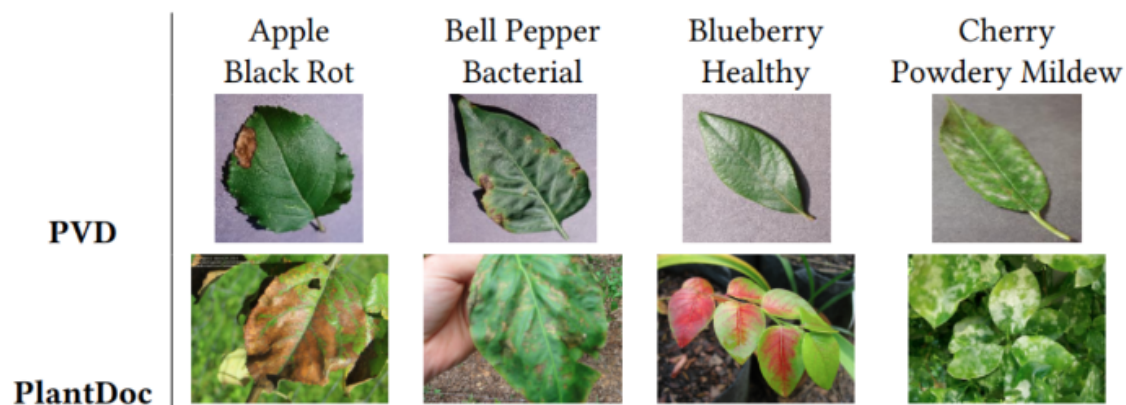
Για να διασφαλιστεί η γενικευσιμότητα του μοντέλου, έγινε χρήση ενός διαφορετικού συνόλου δεδομένων, παρόμοιας μορφής με το PlantVillage, του PlantDoc dataset [50]. Η κύρια διαφοροποίησή του είναι ότι περιλαμβάνει εικόνες που έχουν ληφθεί σε πραγματικές συνθήκες πεδίου, αντί για συνθετικά ή εργαστηριακά περιβάλλοντα [4.7]. Παρότι οι εικόνες αυτές είναι γενικά καθαρές και υψηλής ανάλυσης, η χρήση δεδομένων πεδίου συμβάλλει στη βελτίωση της γενικευσιμότητας και επιτρέπει μια πιο ρεαλιστική και αντικειμενική αξιολόγηση της απόδοσης του μοντέλου, κάτι ιδιαίτερα σημαντικό καθώς ο τελικός στόχος είναι η εφαρμογή του συστήματος σε πραγματικά αγροτικά περιβάλλοντα.

Στο Σχήμα 4.9 φαίνονται τα αποτελέσματα της αξιολόγησης με τη χρήση του PlantDoc Dataset.

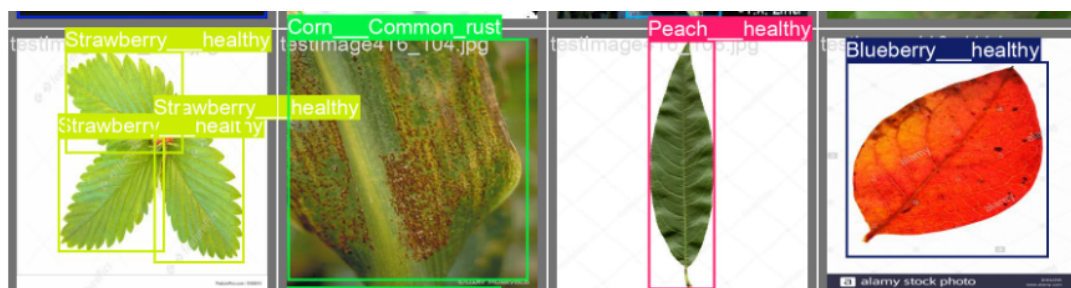
Όπως φαίνεται από τις γραφικές παραστάσεις αξιολόγησης του μοντέλου 4.9, η απόδοση του στο νέο dataset είναι σημαντικά υποδεέστερη. Ξεκινώντας από τη confusion matrix, παρατηρείται έντονη λανθασμένη επισήμανση (mislabeling) των εικόνων, χωρίς να υπάρχει κάποια κλάση την οποία το μοντέλο να μπορεί να αναγνωρίσει σταθέρα. Επιπλέον, διαπιστώνεται δυσκολία στον εντοπισμό, καθώς οι περισσότερες εικόνες ταξινομούνται ως background, γεγονός που αποτυπώνεται και στην αδυναμία σωστής σχεδίασης των bounding boxes. Το χαμηλό recall σε όλο το εύρος των τιμών confidence επιβεβαιώνει αυτή την παρατήρηση. Παράλληλα η χαμηλή καμπύλη του precision καταδεικνύει μεγάλο αριθμό εσφαλμένων ανιχνεύσεων, με την ακρίβεια να βελτιώνεται όταν η σιγουριά του μοντέλου είναι κοντά στο 1



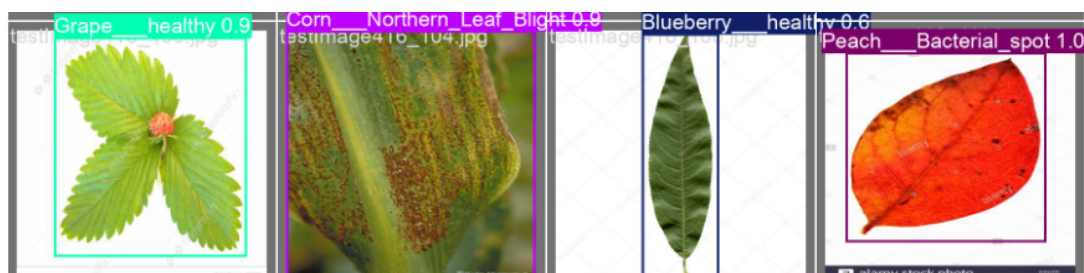
Σχήμα 4.6: Γραφήματα Αξιολόγησης στο PlantVillage Dataset



Σχήμα 4.7: Δείγματα από διάφορες κατηγορίες του συνόλου δεδομένων PlantDoc δείχνουν το χάσμα μεταξύ εργαστηριακά ελεγχόμενων και πραγματικών εικόνων

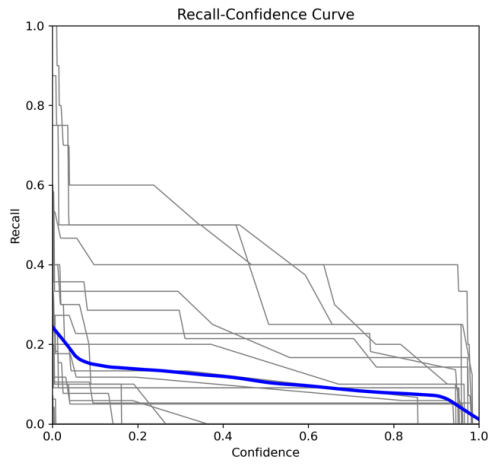


(α') Ground Truth για την κλάση και το bounding box

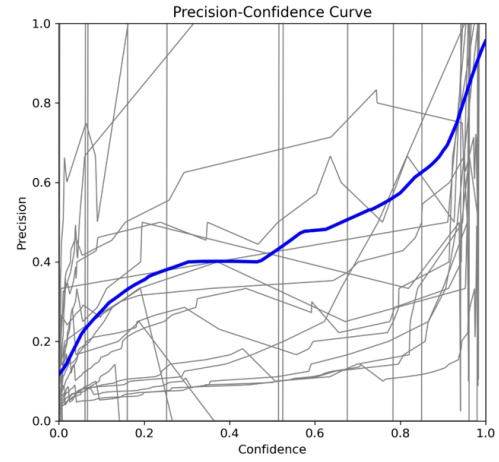


(β') Αποτελέσματα Detection από το εκπαιδευμένο μοντέλο μαζί με το confidence της κάθε πρόβλεψης

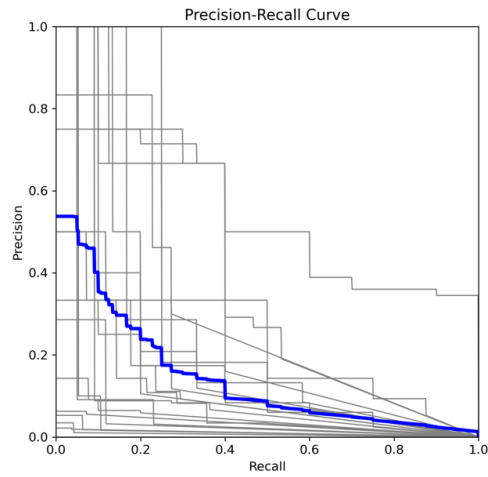
Σχήμα 4.8: Παραδείγματα αναγνώρισης φύλλου και ασθένειας στο PlantDoc Dataset



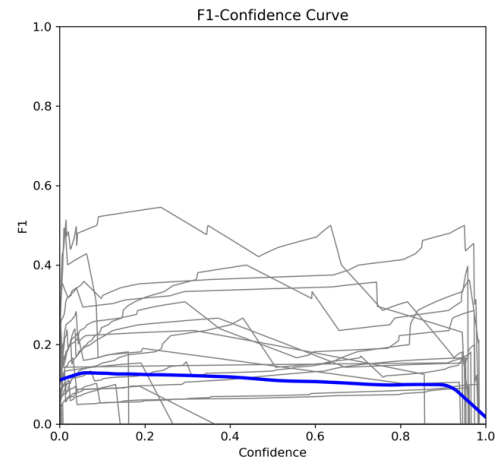
Recall Curve



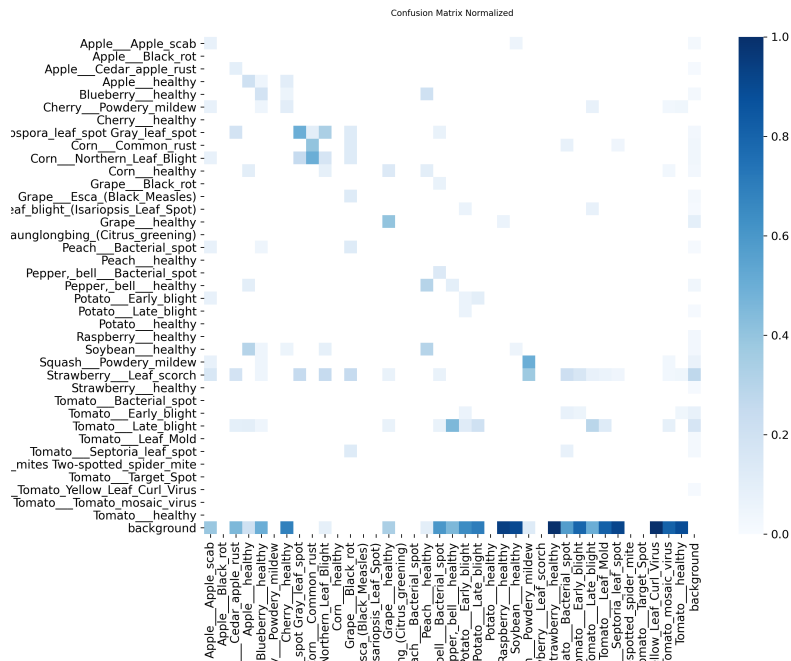
Precision Curve



PR Curve



F1 Curve



Confusion Matrix

Σχήμα 4.9: Γραφήματα Αξιολόγησης στο PlantDoc Dataset

(σημείο που στην καμπύλη του recall είδαμε πως χάνονται οι περισσότερες ανιχνεύσεις). Η καμπύλη Precision-Recall καταδεικνύει φτωχή ισορροπία μεταξύ ευαισθησίας και ακρίβειας, ενώ οι τιμές του F1-score παραμένουν χαμηλές σε όλο το εύρος των τιμών εμπιστοσύνης, χωρίς να εντοπίζεται κάποιο σαφές βέλτιστο σημείο. Συνολικά, το μοντέλο εμφανίζει ασθενή γενική απόδοση στο συγκεκριμένο σύνολο δεδομένων.

Όπως φαίνεται και στο Σχήμα 4.11, παρά τους υψηλούς δείκτες confidence, το μοντέλο αδυνατεί να εντοπίσει όλα τα φύλλα που απεικονίζονται, ενώ στις περισσότερες περιπτώσεις προβαίνει σε λανθασμένη ταξινόμηση. Οι παραπάνω παρατηρήσεις καταδεικνύουν την ανάγκη για μια πιο αποτελεσματική και στοχευμένη προσέγγιση.

4.2.4 Ανίχνευση Ασθενειών με Βάση τις Προσβεβλημένες Περιοχές

Αυτή η προσέγγιση βασίζεται στη μελέτη του άρθρου *A novel deep learning method for detection and classification of plant diseases* [35]. Σε αντίθεση με τις παραδοσιακές μεθόδους που βασίζονται στην ανάλυση ολόκληρου του φύλλου, εδώ προτείνεται η ανίχνευση της ασθένειας να πραγματοποιείται αποκλειστικά πάνω στις προσβεβλημένες περιοχές. Με τον τρόπο αυτό, απομονώνονται τα σημεία ενδιαφέροντος από το φόντο και τα υγιή τμήματα, ενισχύοντας έτσι τη σχετικότητα της πληροφορίας που παρέχεται στο μοντέλο.

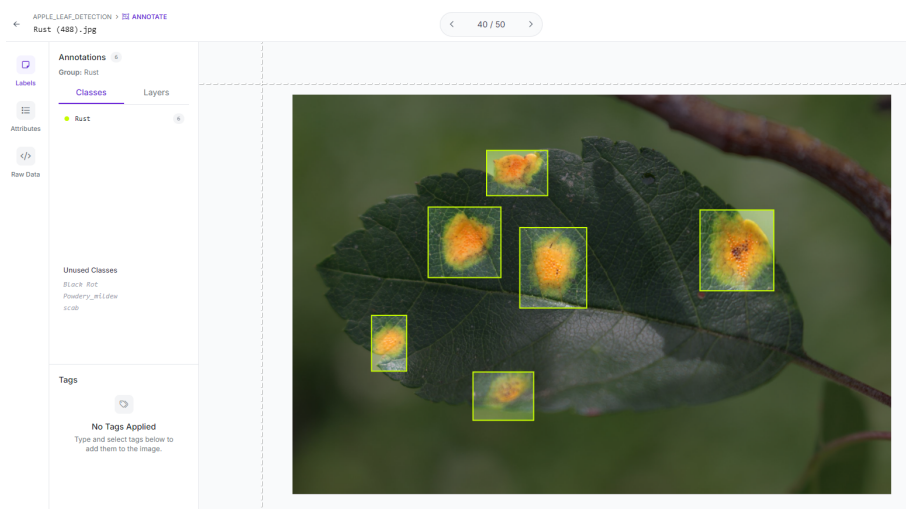
Η συγκεκριμένη στρατηγική καθιστά μη κατάλληλα τα σύνολα δεδομένων που χρησιμοποιήθηκαν μέχρι τώρα, καθώς τα bounding boxes σε αυτά περιλάμβαναν ολόκληρο το φύλλο και όχι μόνο τις περιοχές με εμφανή συμπτώματα.

Έπειτα από εκτενή αναζήτηση στο Roboflow, εντοπίστηκε το `apple_leaf_detection_dataset` [6], το οποίο αποτελεί ένα εξειδικευμένο σύνολο δεδομένων αφιερωμένο αποκλειστικά σε φύλλα μηλιάς. Το συγκεκριμένο dataset επικεντρώνεται στη δημιουργία bounding boxes γύρω από όλες τις προσβεβλημένες περιοχές των φύλλων, όπως απεικονίζεται στο Σχήμα [4.10]. Περιλαμβάνει τέσσερις κατηγορίες ασθενειών ('Black Rot', 'Powdery Mildew', 'Rust', 'Scab'), όλες εκ των οποίων περιλαμβάνονται και στο PlantVillage dataset.

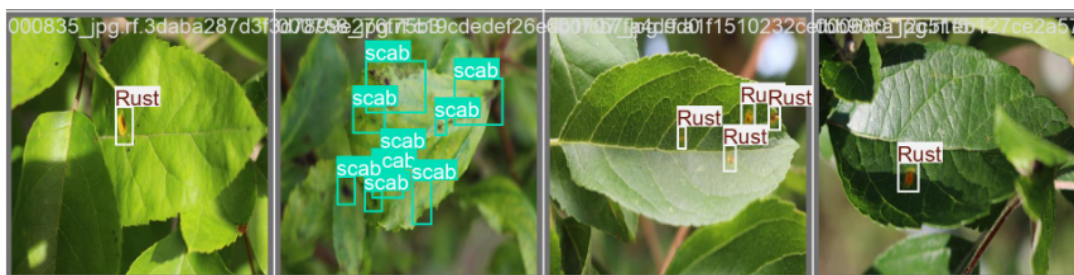
Το dataset χωρίστηκε εκ νέου σε train, validation και test splits. Εκπαιδεύτηκε μοντέλο YOLOv11 για object detection πάνω σε αυτό το dataset και στη συνέχεια αξιολογήθηκε στο test split. Καθώς οι εικόνες προέρχονται από πραγματικές συνθήκες πεδίου, η γενικευσιμότητα των αποτελεσμάτων θεωρείται πιο αξιόπιστη. Τα αποτελέσματα της αξιολόγησης στο test split παρουσιάζονται στο Σχήμα 4.12.

Σε επίπεδο αξιολόγησης, αυτή η προσέγγιση εμφανίζει σαφώς βελτιωμένα αποτελέσματα. Υπάρχει ικανοποιητικό trade-off ανάμεσα σε precision και recall, καθώς από την καμπύλη Precision-Recall η συνολική επίδοση (mAP@0.5) είναι ≈ 0.797 , ενώ ακόμη και η πιο απαιτητική κλάση πετυχαίνει απόδοση ≈ 0.713 —τιμή υψηλότερη από κάθε απόδοση που είχε επιτευχθεί με την προηγούμενη προσέγγιση.

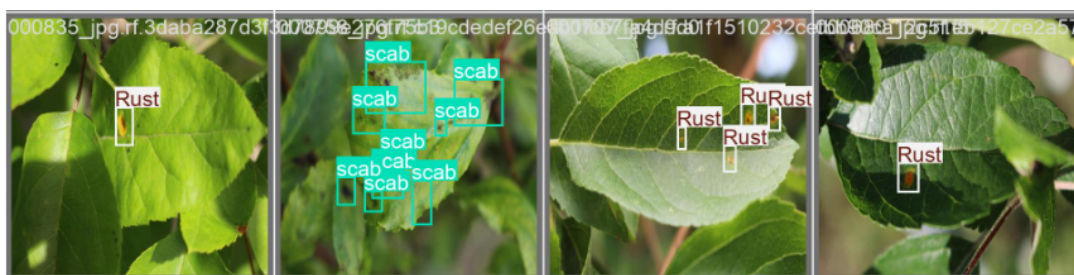
Από τις καμπύλες Precision-Confidence και Recall-Confidence προκύπτει ότι, παρότι η αύξηση του confidence οδηγεί σε ολοένα και πιο σωστές προβλέψεις, για confidence score μεγαλύτερο από ≈ 0.4 το recall μειώνεται απότομα. Η εικόνα αυτή συμφωνεί με το confusion



Σχήμα 4.10: Δείγμα εικόνας από το σετ δεδομένων στο Roboflow

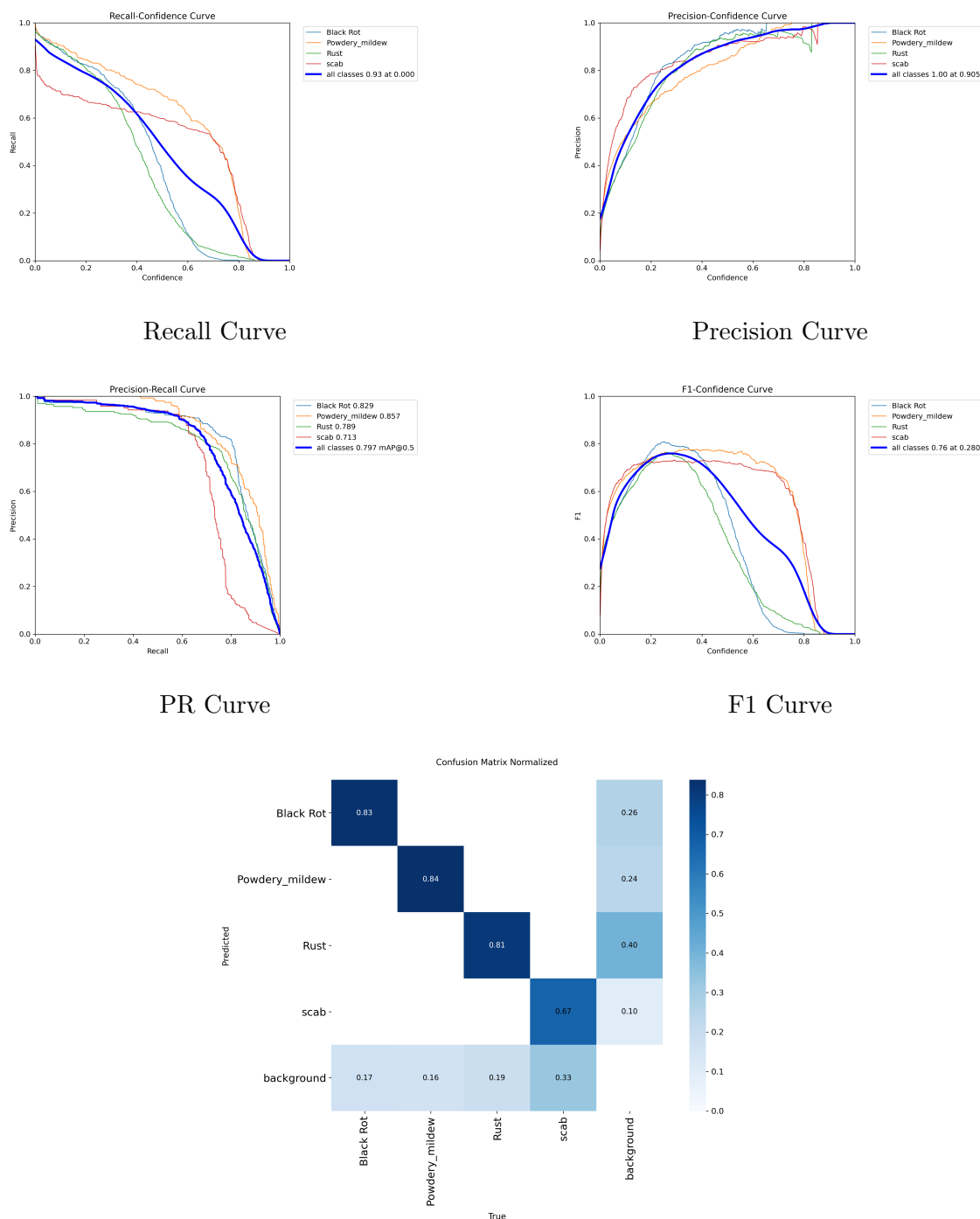


(α') Ground Truth για την κλάση και το bounding box



(β') Αποτελέσματα Detection από το εκπαιδευμένο μοντέλο μαζί με το confidence της κάθε πρόβλεψης

Σχήμα 4.11: Παραδείγματα αναγνώρισης φύλλου και ασθένειας στο apple-leaf-detection-dataset



Confusion Matrix

Σχήμα 4.12: Γραφήματα αξιολόγησης με ανίχνευση στις άρρωστες περιοχές

matrix, το οποίο δείχνει ότι οι ασθένειες δεν είναι mislabelled ως άλλες κλάσεις, αλλά κυρίως ως background — δηλαδή, είτε το μοντέλο δεν εντοπίζει τα προσβεβλημένα σημεία, είτε τα ευρήματα δεν ταυτίζονται τουλάχιστον κατά $\approx 50\%$ με ground truth, ώστε να θεωρηθούν valid. Το βέλτιστο επιχειρησιακό σημείο προκύπτει από την καμπύλη F1–Confidence, όπου το μέγιστο F1 ≈ 0.76 επιτυγχάνεται σε confidence ≈ 0.28 . Συνολικά, το μοντέλο αποδίδει καλύτερα στις κλάσεις Powdery mildew και Black Rot. Αυτή η τεχνική λοιπόν ήταν ωφέλιμη για την αναγνώριση ασθενειών, το πρόβλημα που υπάρχει όμως είναι πως δεν υπάρχουν ανάλογα δεδομένα για άλλα είδη δέντρων. Παρ’ όλα αυτά υπάρχουν τεχνικές με τις οποίες μπορούμε να πετύχουμε καλύτερη ανίχνευση με τα δεδομένα τα οποία είναι διαθέσιμα.

Η τεχνική αυτή αποδείχθηκε ωφέλιμη για την αναγνώριση ασθενειών. Το βασικό, όμως, πρόβλημα είναι ότι δεν υπάρχουν αντίστοιχα σύνολα δεδομένων για άλλα είδη δέντρων. Παρ’ όλα αυτά, μπορούμε να αξιοποιήσουμε τα διαθέσιμα δεδομένα και να βελτιώσουμε την ανίχνευση με κατάλληλες μεθόδους, ώστε να επιτύχουμε αξιόπιστα αποτελέσματα και σε διαφορετικά είδη.

4.2.5 Εκπαίδευση ανά Τύπο Φυτού

Από τα πειράματα που παρουσιάστηκαν στην Ενότητα 4.2.2, παρατηρήθηκε ότι, παρά το γεγονός πως η ταξινόμηση των εικόνων ήταν κατά κανόνα λανθασμένη, η ανίχνευση των προσβεβλημένων περιοχών ήταν σχετικά ικανοποιητική — γεγονός που αποτυπώνεται χαρακτηριστικά στο Σχήμα 4.11. Λαμβάνοντας υπόψη αυτό το συμπέρασμα, καθώς και τα ενισχυμένα αποτελέσματα που επιτεύχθηκαν στην Ενότητα 4.2.4, όπου το μοντέλο καλούνταν να επιλέξει μόνο μεταξύ τεσσάρων κατηγοριών ασθενειών μήλου, αντί για τις 38 του πλήρους PlantVillage dataset, προτείνεται μια στρατηγική περιορισμού του πλήθους των πιθανών κλάσεων κατά την ανίχνευση.

Η νέα προσέγγιση βασίζεται στον διαχωρισμό της διαδικασίας σε δύο διακριτά στάδια: αρχικά, ένα κατάλληλα fine-tuned μοντέλο YOLO θα χρησιμοποιείται για την αναγνώριση του τύπου του φυτού που απεικονίζεται στην εικόνα. Στη συνέχεια, με βάση το αναγνωρισμένο είδος φυτού, θα εφαρμόζεται ένας εξειδικευμένος ανιχνευτής ασθενειών, επίσης fine-tuned, εκπαιδευμένος αποκλειστικά για την ανίχνευση ασθενειών που σχετίζονται με το συγκεκριμένο φυτό.

Αναδιαμόρφωση του Συνόλου Δεδομένων

Για την υλοποίηση της παραπάνω προσέγγισης, το PlantVillage αναδιαμορφώθηκε, ώστε να υποστηρίζει δύο τύπους ανιχνευτών:

- **Ανίχνευση Τύπου Φυτού:** Δημιουργήθηκε Python script για την παραγωγή νέου dataset. Το script ομαδοποιεί όλες τις κλάσεις που εμφανίζονται στα αρχικά label .txt αρχεία του συνόλου δεδομένων, έτσι ώστε όλες οι ασθένειες που ανήκουν στον ίδιο τύπο φυτού να αντιστοιχίζονται στην ίδια κλάση (δηλαδή επανεγγράφονται τα `class-ids` μόνο ως `plant-type`). Η ομαδοποίηση έγινε σύμφωνα με τον Πίνακα 4.1.

Πίνακας 4.1: Ομαδοποίηση αρχικών κλάσεων του PlantVillage ανά τύπο φυτού

Τύπος Φυτού	Αρχικές Κλάσεις που Ομαδοποιούνται
Apple	Apple__Apple_scab, Apple__Black_rot, Apple__Cedar_apple_rust, Apple__healthy
Blueberry	Blueberry__healthy
Cherry	Cherry__Powdery_mildew, Cherry__healthy
Corn	Corn__Cercospora_leaf_spot_Gray_leaf_spot, Corn__Common_rust, Corn__Northern_Leaf_Blight, Corn__healthy
Grape	Grape__Black_rot, Grape__Esca_(Black_Measles), Grape__Leaf_blight_(Isariopsis_Leaf_Spot), Grape__healthy
Orange	Orange__Haunglongbing_(Citrus_greening)
Peach	Peach__Bacterial_spot, Peach__healthy
Pepper, bell	Pepper, bell__Bacterial_spot, Pepper, bell__healthy
Potato	Potato__Early_blight, Potato__Late_blight, Potato__healthy
Raspberry	Raspberry__healthy
Soybean	Soybean__healthy
Squash	Squash__Powdery_mildew
Strawberry	Strawberry__Leaf_scorch, Strawberry__healthy
Tomato	Tomato__Bacterial_spot, Tomato__Early_blight, Tomato__Late_blight, Tomato__Leaf_Mold, Tomato__Septoria_leaf_spot, Tomato__Spider_mites_Two-spotted_spider_mite, Tomato__Target_Spot, Tomato__Tomato_Yellow_Leaf_Curl_Virus, Tomato__Tomato_mosaic_virus, Tomato__healthy

- **Ανίχνευση Ασθένειας ανά Τύπο Φυτού:** Για κάθε τύπο φυτού δημιουργήθηκε εξειδικευμένο υποσύνολο dataset με τις αντίστοιχες κλάσεις ασθενειών του (π.χ. για Apple: Apple_scab, Black_rot, Cedar_apple_rust, healthy). Η διαδικασία περιλαμβάνει: (i) φιλτράρισμα των εικόνων/ετικετών που αντιστοιχούν στον συγκεκριμένο τύπο φυτού (βάσει του προθέματος πριν από το “___”), (ii) επανεγγραφή των labels ώστε τα `class_ids` να αντιστοιχούν στις επιμέρους ασθένειες του φυτού (χωρίς το πρόθεμα του φυτού), (iii) παραγωγή ξεχωριστού αρχείου YAML ανά φυτό με τα νέα class names και `nc` (number of classes), και (iv) εκπαίδευση ενός YOLO ανιχνευτή ανά φυτό πάνω στο υποσύνολο (τα train/val/test splits διατηρούνται).

Εκπαίδευση Ανιχνευτή Τύπου Φυτού

Ο ανιχνευτής τύπου φυτού (YOLO) εκπαιδεύτηκε στο αναδιαμορφωμένο PlantVillage που περιγράφηκε παραπάνω (14 κλάσεις τύπων φυτών). Η αξιολόγηση και πάλι γίνεται με τη χρήση του PlantDoc dataset.

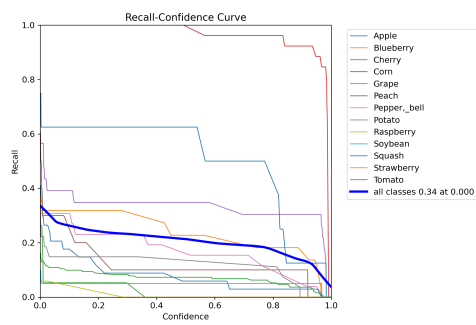
Από τα αποτελέσματα της αξιολόγησης, βλέπουμε πως οι καμπύλες Recall–Confidence και Precision–Confidence δείχνουν χαμηλή ανάκληση συνολικά ($\text{recall} \approx 0.34$ στο $\text{confidence} = 0$, φθίνουσα όσο αυξάνει το confidence) και αύξηση της ακρίβειας (precision) έως ≈ 0.89 , πολύ κοντά στο $\text{confidence} = 1$, κάτι που υποδηλώνει ότι με αυστηρά κατώφλια εντοπίζονται λίγα, αλλά καθαρά, δείγματα. Η καμπύλη F1–Confidence κορυφώνεται περίπου στο σε τιμή $\text{F1} \approx 0.22$ γύρω από $\text{confidence} \approx 0.12$. Η Precision–Recall δίνει συνολικό $\text{mAP}@0.5 = 0.237$, με έντονες διαφοροποιήσεις ανά κλάση: πολύ καλή επίδοση στο Corn (≈ 0.929) και μέτρια σε Squash, Grape, Blueberry ($\approx 0.45/0.37/0.31$), αλλά χαμηλή σε Strawberry, Raspberry, Cherry και Soybean ($\approx 0.00/0.04/0.05/0.07$). Ο κανονικοποιημένος πίνακας σύγχυσης εμφανίζει ισχυρή πόλωση προς background προβλέψεις, εξηγώντας τη χαμηλή ανάκληση (πολλές αστοχίες ανίχνευσης). Συνολικά, η απόδοση του μοντέλου δεν είναι επαρκής για χρήση ως προ-ταξινομητής τύπου φυτού σε πραγματικές περιπτώσεις.

Για να λυθεί το πρόβλημα της αδυναμίας της γενικευσιμότητας του μοντέλου δημιουργήθηκε ένα νέο, υβριδικό σύνολο δεδομένων που συνδυάζει εργαστηριακές λήψεις (όπου τα φύλλα εμφανίζονται καθαρά και απομονωμένα) με φωτογραφίες πεδίου (όπου τα φύλλα παραμένουν πάνω στο δέντρο και μπορεί να συνυπάρχουν πολλαπλά φύλλα ανά εικόνα). Για τη σύνθεσή του αξιοποιήθηκε ολόκληρο το PlantVillage, καθώς και υποσύνολο του PlantDoc, με διαχωρισμό σε σύνολο εκπαίδευσης 70% και σύνολο αξιολόγησης 30%, ώστε να διασφαλιστεί αξιόπιστη εκτίμηση της απόδοσης του μοντέλου.

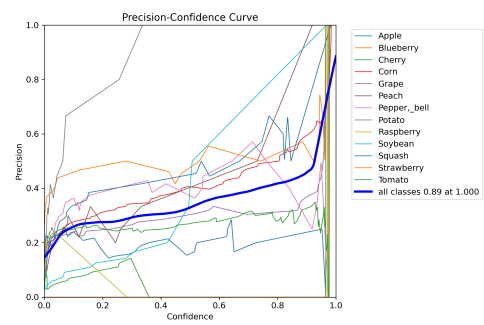
Μετά την επανεκπαίδευση, το μοντέλο αξιολογήθηκε και πάλι [Σχήμα 4.14]. Οι καμπύλες Precision–Confidence, Recall–Confidence και F1–Confidence δείχνουν εξαιρετικά σταθερή συμπεριφορά σε όλο το εύρος του confidence : η precision προσεγγίζει το 1.00, η recall παραμένει πολύ κοντά στο 1 έως και υψηλά κατώφλια, ενώ η συνολική επίδοση κορυφώνεται σε $\text{F1} \approx 0.97$ γύρω από $\text{confidence} \approx 0.525$. Η Precision–Recall καμπύλη είναι σχεδόν «κολλημένη» στο άνω αριστερό όριο, με $\text{mAP}@0.5 = 0.989$ και ομοιόμορφα υψηλές τιμές ανά κλάση (ελαφρώς χαμηλότερες—αλλά και πάλι άριστες—για Blueberry και Raspberry, τα οποία είχαν δυσκολέψει πολύ τις προηγούμενες προσεγγίσεις). Η κανονικοποιημένη confusion matrix είναι σχεδόν διαγώνια, με πολύ μικρή διαρροή μεταξύ κλάσεων, στοιχείο που επιβεβαιώνει την υψηλή διακριτική ικανότητα του ανιχνευτή τύπου φυτού. Συνολικά, τα ευρήματα υποδεικνύουν ένα ώριμο μοντέλο με ασφαλές λειτουργικό κατώφλι γύρω στο 0.5, κατάλληλο για χρήση ως πρώτο στάδιο στο σύστημά μας.

Εκπαίδευση Ανιχνευτή Ύπαρξης Φύλλου

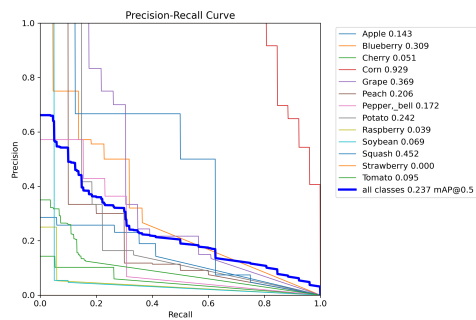
Επειδή ο στόχος είναι η αξιοποίηση του συστήματος σε πραγματικές συνθήκες, οφείλουμε να λάβουμε υπόψη την πιθανότητα η εικόνα που παρέχει ο χρήστης να μην περιέχει φύλλο. Σε έναν προκαταρκτικό έλεγχο με τη χρήση του COCO [30] Dataset διαπιστώθηκε ότι ο ανιχνευ-



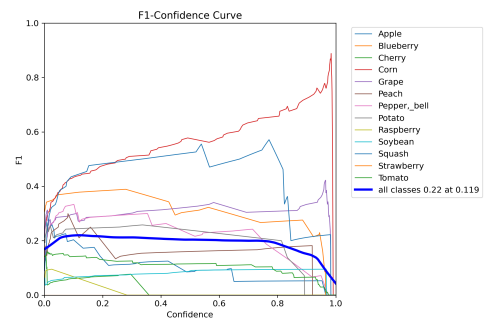
Recall Curve



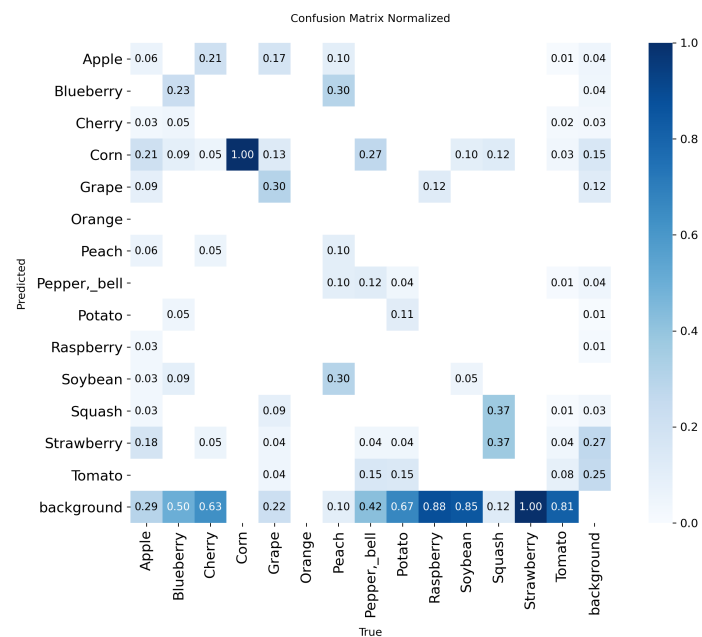
Precision Curve



PR Curve

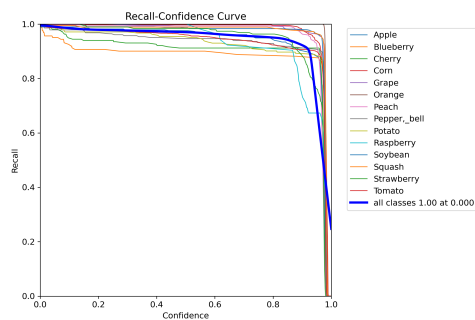


F1 Curve

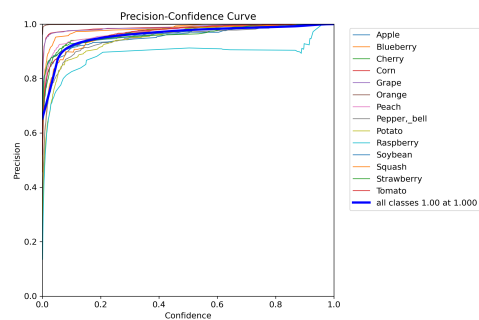


Confusion Matrix

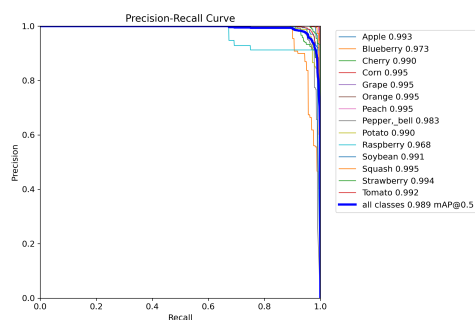
Σχήμα 4.13: Γραφήματα Αξιολόγησης Ανίχνευσης Τύπου Φυτού



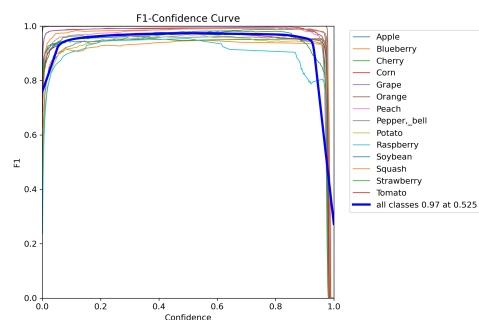
Recall Curve



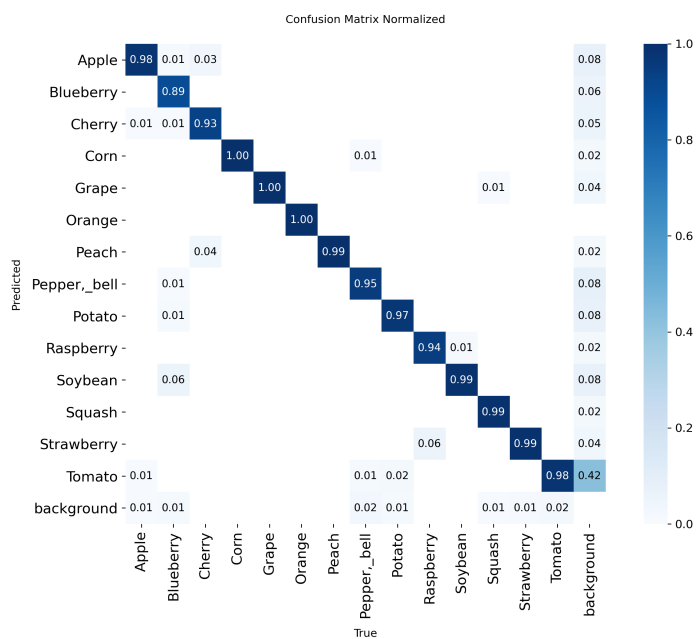
Precision Curve



PR Curve



F1 Curve



Confusion Matrix

Σχήμα 4.14: Γραφήματα Αξιολόγησης Ανίχνευσης Τύπου Φυτού μετά απο εκπαίδευση με υβριδικό dataset



Σχήμα 4.15: Παραδείγματα λανθασμένης αναγνώρισης

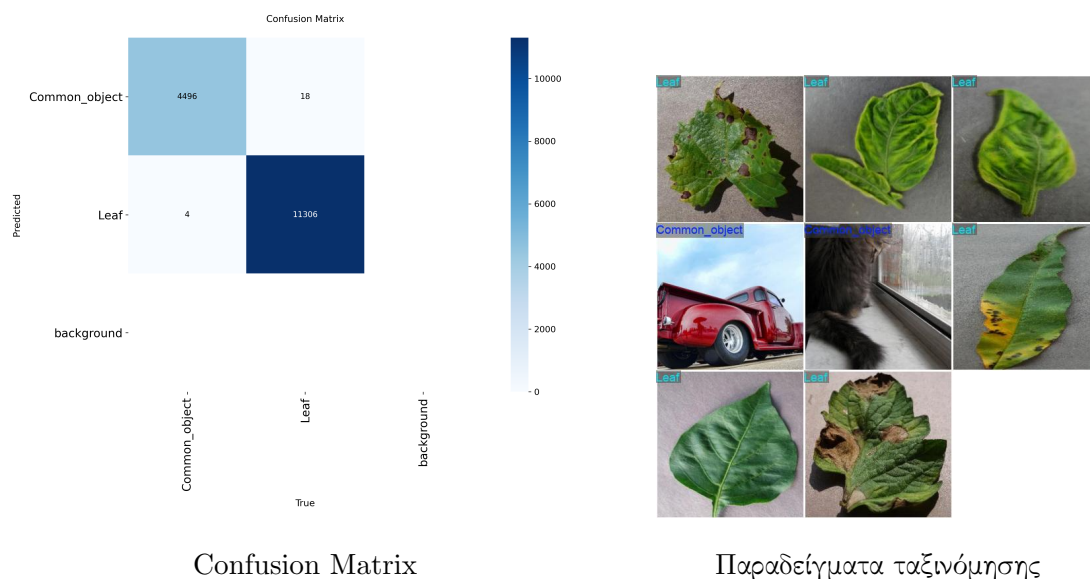
τής τύπου φύλλου παρήγαγε συχνά false positives, εντοπίζοντας «φύλλα» σε σημεία όπου δεν υπάρχουν, όπως φαίνεται στο Σχήμα 4.15. Το COCO είναι ένα μεγάλο μέγεθος σύνολο δεδομένων με καθημερινές σκηνές και ποικιλία αντικειμένων, ευρέως χρησιμοποιούμενο για εργασίες object detection, segmentation και classification. Η αξιοποίησή του εδώ λειτουργεί ως negative control για τη μέτρηση του ρυθμού λανθασμένων ανιχνεύσεων σε εικόνες χωρίς στοχευμένο αντικείμενο (φύλλο).

Για να αποφευχθεί η διάδοση σφαλμάτων στα επόμενα στάδια, εκπαιδεύτηκε ένα μοντέλο YOLO σε λειτουργία classification που λειτουργεί ως πύλη (gate): αποφασίζει αν στην είσοδο υπάρχει τουλάχιστον ένα φύλλο. Μόνο εφόσον ανιχνευθεί «παρουσία φύλλου» προχωρά η ροή σε αναγνώριση τύπου φυτού και κατόπιν ασθένειας. Ο ταξινομητής εκπαιδεύτηκε σε **δυαδική** διατύπωση (Leaf / Common Object) με θετικά παραδείγματα από το υβριδικό μας σύνολο (εργαστηριακές και πεδίου λήψεις, όπου τα φύλλα είναι σαφώς ορατά) και αρνητικά παραδείγματα από επιλεγμένο υποσύνολο του COCO με σκηνές που τυπικά δεν περιέχουν κοντινά φύλλα. Κατά την προεπεξεργασία εφαρμόστηκαν ήπιες αυξήσεις δεδομένων (data augmentation), όπως τυχαίες περικοπές, μικρές χρωματικές μεταβολές και ελαφρά θόλωση, ώστε το μοντέλο να γενικεύει σε διαφορετικές συνθήκες φωτισμού και background. Η εκπαίδευση πραγματοποιήθηκε με τυπικό σχήμα cross-entropy loss και early stopping βάσει της επίδοσης στο validation σύνολο.

Όπως φαίνεται από τα αποτελέσματα της αξιολόγησης του ταξινομητή (Σχήμα 4.16), το μοντέλο κάνει ελάχιστα λάθη (22 λάθος ταξινομήσεις στις 15822 φωτογραφίες), άρα μπορεί να χρησιμοποιηθεί στο σύστημά μας.

Εκπαίδευση Ανιχνευτή Ασθένειας ανά Τύπου Φυτού

Έχοντας πλέον ένα σταθερό μοντέλο που μπορεί με ακρίβεια να αναγνωρίζει τον τύπο του φυτού που εμφανίζεται στη φωτογραφία, προχωρήσαμε στην εκπαίδευση 14 εξειδικευμένων ταξινομητών, καθένας εκ των οποίων αναγνωρίζει ασθένειες για έναν συγκεκριμένο τύπο φυτού, με βάση το αρχικό PlantVillage dataset. Για τον σκοπό αυτό δημιουργήσαμε ένα υβριδικό σύνολο δεδομένων, όπως περιγράφηκε προηγουμένως, χρησιμοποιώντας εικόνες



Σχήμα 4.16: Αποτελέσματα Αξιολόγησης Ταξινόμησης ύπαρξης Φυτού

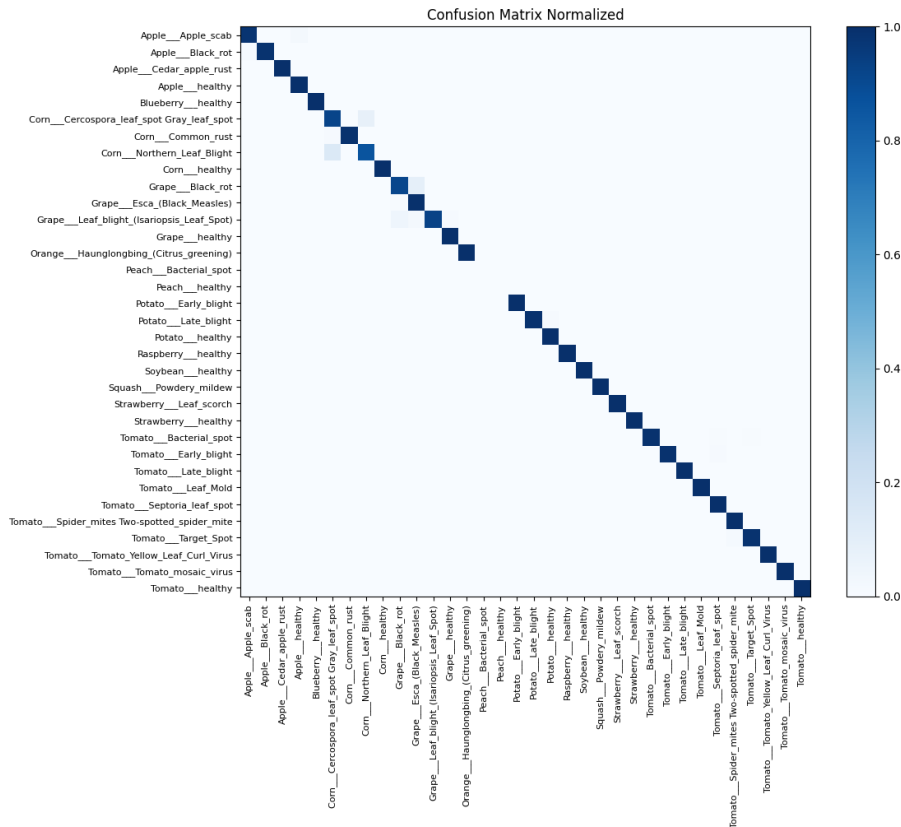
από τα PlantVillage και PlantDoc. Αρχικά, ενοποιήσαμε τα ονόματα των κλάσεων των δύο datasets ώστε να είναι συνεπή. Από το συνολικό αυτό σύνολο δεδομένων προέκυψαν υποσύνολα που χρησιμοποιήθηκαν για την εκπαίδευση των εξειδικευμένων ταξινομητών. Τα μοντέλα εκπαιδεύτηκαν και αξιολογήθηκαν χρησιμοποιώντας το 30% του PlantDoc (το οποίο δεν χρησιμοποιήθηκε για εκπαίδευση).

Όπως προκύπτει από τα αποτελέσματα της αξιολόγησης στην Εικόνα 4.17, το υλοποιημένο YOLO pipeline παρουσιάζει σαφώς ανώτερη απόδοση σε σύγκριση με όλες τις υπόλοιπες προσεγγίσεις. Το σύστημα καταφέρνει σταθερά να αναγνωρίζει σωστά την κατηγορία της ασθένειας σε συνάρτηση με τον σωστό τύπο φύλλου, ενώ ταυτόχρονα επιδεικνύει την ικανότητα να ανιχνεύει αξιόπιστα πότε δεν υπάρχει φύλλο στην εικόνα εισόδου.

4.2.6 Δημιουργία RAG Συστήματος

Έχοντας πλέον ολοκληρώσει την υλοποίηση του συστήματος ανίχνευσης της ασθένειας, προχωρούμε στην ανάπτυξη ενός συστήματος Retrieval-Augmented Generation (RAG), το οποίο έχει ως στόχο την παροχή έγκυρης και εξειδικευμένης γνώσης σε ένα γλωσσικό μοντέλο. Με τον τρόπο αυτό, το μοντέλο θα μπορεί να παράγει μια τεκμηριωμένη και ακριβή ανάλυση της ασθένειας που εντοπίστηκε στο φύλλο.

Για τη δημιουργία της βάσης γνώσης επιλέχθηκε η χρήση μιας Chroma database, πάνω στην οποία δοκιμάστηκαν διαφορετικά μοντέλα ενσωμάτωσης (embedding models) — τα οποία θα περιγραφούν παρακάτω. Δεδομένου ότι το σύστημα καλείται να διαχειριστεί ένα συγκεκριμένο σύνολο ασθενειών και τύπων φυτών, δημιουργήθηκε ένα ειδικό Python script, το οποίο ανέλαβε τη συλλογή (scraping) κειμένων από διαδικτυακές σελίδες αξιόπιστων πανεπιστημιακών ιδρυμάτων, που αναλύουν μεθόδους διαχείρισης των ασθενειών που περιλαμβάνονται στο PlantVillage dataset.



Σχήμα 4.17: Confusion Matrix του τελικού Yolo Pipeline

Για κάθε ασθένεια επιλέχθηκαν χειροκίνητα (manually) δύο πηγές από έγκριτους οργανισμούς, όπως το Penn State University και το Cornell University. Από τις σελίδες αυτές εξήχθη περιγραφικό κείμενο και προτάσεις αντιμετώπισης· τα κείμενα καθαρίστηκαν και μετατράπηκαν σε documents με κατάλληλα metadata (π.χ. ασθένεια, είδος φυτού, προέλευση), τα οποία ενσωματώθηκαν στη Chroma DB.

Η παραμετροποίηση του συστήματος ανάκτησης πραγματοποιήθηκε με συνδυαστική χρήση RAGAS και Optuna Study. Πιο συγκεκριμένα, αξιοποιήθηκαν οι μετρικές context_precision και context_recall της βιβλιοθήκης RAGAS για να εκτιμηθεί σε ποιον συνδυασμό hyper-parameters και embedding model το σύστημα αποδίδει καλύτερα. Κατά τη διάρκεια της διερεύνησης, το Optuna μπορούσε:

- να αλλάζει το chunk size από 80 έως 1000 με βήμα 20,
- να ορίσει επικάλυψη ως $\text{max_overlap} = \max(0, \text{chunk_size} - 20)$,
- να ρυθμίσει το chunk overlap στο $[0, \text{max_overlap}]$ με βήμα 10,
- να επιλέξει batch size από το σύνολο 16, 32, 64, 128, 256,
- να δοκιμάσει διαφορετικά embedding models από τα: sentence-transformers/all-MiniLM-L6-v2, sentence-transformers/all-mpnet-base-v2, google/embeddinggemma-300m.

Για τον υπολογισμό των `context_precision` και `context_recall` χρησιμοποιήθηκε το `gemini-2.0-flash` (μέσω API) ως **LLM-as-judge**. Η χρήση LLM είναι απαραίτητη, διότι οι συγκεκριμένες μετρικές **δεν βασίζονται σε απλή αντιστοίχιση λέξη-προς-λέξη, αλλά σε σημασιολογική αξιολόγηση**: το LLM εντοπίζει τις βασικές μονάδες πληροφορίας της απάντησης και κρίνει (α) ποια από τα ανακτηθέντα αποσπάσματα τις τεκμηριώνουν (`context_precision`: αναλογία σχετικών αποσπασμάτων στα ανακτηθέντα) και (β) σε ποιο βαθμό το σύνολο των ανακτηθέντων αποσπασμάτων καλύπτει τις απαιτούμενες πληροφορίες της απάντησης (`context_recall`: κάλυψη απαιτούμενων πληροφοριών από τα ανακτηθέντα).

Η μελέτη με Optuna διήρκεσε 50 trials και η αξιολόγηση γινόταν στα top-3 documents που ανέκτησε το σύστημα της Chroma DB ανά ερώτημα, χρησιμοποιώντας τις προαναφερθείσες μετρικές της RAGAS. Τα αποτελέσματα της μελέτης φαίνονται στον Πίνακα 4.2:

Table 4.2: RAG retrieval evaluation for different setups

number	value	embedding model	chunk overlap/chunk size/batch size
0	0.474692	all-mpnet-base-v2	750/ 980/ 64
1	0.398183	embeddinggemma-300m	430/ 660/ 64
2	0.306818	embeddinggemma-300m	750/ 840/ 16
3	0.380652	embeddinggemma-300m	550/ 820/ 16
4	0.517702	all-MiniLM-L6-v2	50/ 360/ 64
5	0.444263	all-mpnet-base-v2	180/ 320/ 16
6	0.556422	all-MiniLM-L6-v2	310/ 500/ 128
7	0.151059	embeddinggemma-300m	80/ 320/ 16
8	0.588714	all-MiniLM-L6-v2	70/ 540/ 16
9	0.529848	all-MiniLM-L6-v2	290/ 380/ 128
10	0.612743	all-MiniLM-L6-v2	150/ 600/ 256
11	0.600478	all-MiniLM-L6-v2	160/ 620/ 256
12	0.590079	all-MiniLM-L6-v2	180/ 680/ 256
13	0.346697	all-MiniLM-L6-v2	0/ 80/ 256
14	0.584313	all-MiniLM-L6-v2	200/ 640/ 32
15	0.536158	all-MiniLM-L6-v2	130/ 520/ 256
16	0.539811	all-mpnet-base-v2	400/ 780/ 256
17	0.422936	all-MiniLM-L6-v2	0/ 140/ 256
18	0.584408	all-MiniLM-L6-v2	510/ 980/ 32
19	0.431373	all-mpnet-base-v2	230/ 440/ 256
20	0.507983	all-MiniLM-L6-v2	120/ 240/ 256
21	0.516360	all-MiniLM-L6-v2	240/ 640/ 256
22	0.559065	all-MiniLM-L6-v2	170/ 720/ 256
23	0.531405	all-MiniLM-L6-v2	280/ 600/ 256
24	0.568095	all-MiniLM-L6-v2	350/ 740/ 256
25	0.610059	all-MiniLM-L6-v2	520/ 880/ 256
26	0.572737	all-MiniLM-L6-v2	660/ 900/ 32
27	0.577295	all-MiniLM-L6-v2	530/ 880/ 256
28	0.304169	embeddinggemma-300m	130/ 460/ 128
29	0.531493	all-mpnet-base-v2	600/ 1000/ 64

Continued on next page

number	value	embedding model	chunk overlap/chunk size/batch size
30	0.434614	all-mpnet-base-v2	840/ 920/ 256
31	0.549158	all-MiniLM-L6-v2	450/ 700/ 256
32	0.582136	all-MiniLM-L6-v2	240/ 580/ 256
33	0.548044	all-MiniLM-L6-v2	360/ 680/ 256
34	0.351364	embeddinggemma-300m	440/ 780/ 64
35	0.582130	all-MiniLM-L6-v2	170/ 580/ 256
36	0.604371	all-MiniLM-L6-v2	590/ 820/ 256
37	0.351853	embeddinggemma-300m	670/ 840/ 64
38	0.606895	all-MiniLM-L6-v2	490/ 780/ 128
39	0.584131	all-MiniLM-L6-v2	610/ 940/ 128
40	0.327174	embeddinggemma-300m	480/ 780/ 128
41	0.603993	all-MiniLM-L6-v2	610/ 860/ 128
42	0.587434	all-MiniLM-L6-v2	570/ 880/ 128
43	0.583257	all-MiniLM-L6-v2	690/ 840/ 128
44	0.598997	all-MiniLM-L6-v2	770/ 940/ 128
45	0.610018	all-MiniLM-L6-v2	500/ 760/ 16
46	0.595666	all-MiniLM-L6-v2	490/ 800/ 16
47	0.602703	all-MiniLM-L6-v2	410/ 740/ 16
48	0.514994	all-mpnet-base-v2	550/ 820/ 16
49	0.562022	all-MiniLM-L6-v2	560/ 740/ 16

Άρα, το setup που θα χρησιμοποιείται για το RAG pipeline είναι αυτές που φαίνονται στον Πίνακα 4.3, καθώς με αυτό το setup επιτεύχθηκαν τα καλύτερα αποτελέσματα κατά τη διάρκεια της έρευνας.

Πίνακας 4.3: Παράμετροι ρύθμισης του RAG pipeline

Παράμετρος	Τιμή
Embedding model	all-MiniLM-L6-v2
Chunk overlap	150
Chunk size	600
Batch size	256

4.2.7 Αξιολόγηση του VGRS Pipeline με RAGAS

Πλέον έχει οριστεί τόσο το σύστημα αναγνώρισης της ασθένειας του φυτού, όσο και το σύστημα ανάκτησης έγκυρων πληροφοριών — τόσο για την περιγραφή τρόπων αντιμετώπισης, όσο και για την περιγραφή της ίδιας της ασθένειας. Τα υποσυστήματα που υλοποιήθηκαν, συνδέθηκαν στο ολοκληρωμένο **VGRS Pipeline 3.1**.

Ρόλος του Γλωσσικού Μοντέλου στο VGRS Pipeline

Το LLM στο VGRS Pipeline λειτουργεί ως συνδετικός κρίκος ανάμεσα στα δύο υποσυστήματα που αναλύθηκαν προηγουμένως. Ο ρόλος του είναι να συνθέσει μια τελική απάντηση προς τον χρήστη, η οποία εξηγεί γιατί η φωτογραφία που παρείχε στο σύστημα αντιστοιχεί σε ένα φύλλο προσβεβλημένο από συγκεκριμένη ασθένεια — περιγράφοντας τα οπτικά χαρακτηριστικά της — και να προτείνει πρακτικούς τρόπους αντιμετώπισης. Το κρίσιμο σημείο εδώ είναι ότι δεν ζητείται από το μοντέλο να βασιστεί στη γνώση που έχει αποκτήσει κατά την προεκπαίδευσή του, αλλά να συνδυάσει τη διάγνωση που παρήγαγε το YOLO pipeline με την πληροφορία που ανακτήθηκε από το vector database. Πιο συγκεκριμένα, η έξοδος του YOLO pipeline (δηλαδή τα `disease_type` και `plant_type`) χρησιμοποιείται για τη διατύπωση δύο ερωτημάτων προς τη βάση γνώσης:

- What are the characteristics of {disease_type} in {plant_type} trees?
- How to manage {disease_type} in {plant_type} trees?

Για κάθε ερώτημα ανακτώνται τα τρία πιο σχετικά κείμενα, τα οποία ενσωματώνονται στο prompt του μοντέλου. Ο σχεδιασμός του system prompt προσδιορίζει τον ρόλο του LLM ως εικονικού γεωπόνου-παθολόγου φυτών. Το μοντέλο καλείται να εξηγήσει με σαφήνεια τι είναι η ασθένεια, να περιγράψει τα εμφανή συμπτώματα που την χαρακτηρίζουν, και να προτείνει τρόπους αντιμετώπισης, οργανικούς ή χημικούς. Εφόσον η παρεχόμενη πληροφορία δεν επαρκεί, το μοντέλο οφείλει να το δηλώσει ρητά. Η προσέγγιση αυτή διασφαλίζει την αξιοπιστία της εξόδου και την πρακτική χρησιμότητα της πληροφορίας για τον τελικό χρήστη.

Αξιολόγηση με RAGAS

Η ακρίβεια της αναγνώρισης ασθενειών από το σύστημα αντιστοιχεί στη συνολική ακρίβεια ταξινόμησης που προέκυψε κατά την αξιολόγηση του YOLO pipeline (Σχήμα 4.17). Η ποιότητα της διαδικασίας ανάκτησης γνώσης από τη Chroma database έχει επίσης αξιολογηθεί ξεχωριστά. Αυτό που απομένει, λοιπόν, είναι η συνολική αξιολόγηση του RAG pipeline που έχει δημιουργηθεί. Η αξιολόγηση αυτή θα πραγματοποιηθεί με χρήση της βιβλιοθήκης RAGAS, εστιάζοντας στις μετρικές **faithfulness** 4.1.9 και **answer relevancy** 4.1.10.

Δημιουργήθηκε ένα σύνολο δεδομένων πάνω στο οποίο θα πραγματοποιηθεί η αξιολόγηση με RAGAS. Στο σύστημα δόθηκαν τρεις εικόνες από κάθε κλάση του PlantVillage dataset και, κατά την ανάλυση των εικόνων, αποθηκεύτηκαν: οι ερωτήσεις που υποβλήθηκαν στο LLM, το context που ανακτήθηκε για κάθε ερώτημα, η ανάλυση που συνέθεσε το μοντέλο μας (δηλαδή το συνολικό αποτέλεσμα του συστήματος), καθώς και η απάντηση του Gemini 2.0 Flash όταν του ζητήθηκε να αναλύσει την ασθένεια και να προτείνει τρόπους αντιμετώπισης. Οι απαντήσεις του Gemini θα χρησιμοποιηθούν ως σημείο αναφοράς για σύγκριση με τις απαντήσεις του συστήματός μας. Το μεγάλο γλωσσικό μοντέλο που επιλέχθηκε για το VGRS είναι το Gemma 3 4B, καθώς εμφανίζει ισχυρές επιδόσεις για το μέγεθός του σε καθιερωμένα benchmarks [9] και είναι κατάλληλο για ανάπτυξη σε περιβάλλοντα με περιορισμένους πόρους. Το μοντέλο ανακτήθηκε από το Hugging Face.

Παράδειγμα από το dataset αξιολόγησης



Σχήμα 4.18: Όψιμος περονόσπορος σε φύλλο τομάτας (Tomato late blight)

Question

Based on the image and analysis, a leaf from a **Tomato** tree is most likely affected by **Lateblight**. Explain what is wrong with the leaf, what disease is affecting it, define its visual symptoms, and how the grower should manage the disease.

Contexts

Τα παρακάτω σημεία αποτελούν τα πιο σχετικά αποσπάσματα που εξήχθησαν από διάφορα έγγραφα και χρησιμοποιήθηκαν για την υποστήριξη της παραγόμενης στρατηγικής διαχείρισης:

- “Integrated Management: With no commercially available host resistance, the primary method to control target spot on tomato in Florida is with the routine application of protective fungicides.”
- “Other cultural management strategies include rotating tomato fields with nonsolanaceous crops, starting with clean and healthy transplants, and eliminating weeds (especially Solanaceae family), volunteers, and plant debris that may harbor inoculum.”

- “Everything that improves circulation to reduce humidity deep in the plant canopy can reduce Leaf Mold infection.”
- “Remove all leaf tissue below the lowest fruiting cluster... these lower leaves are often where the infection starts, so de-leafing can remove inoculum.”
- “Rotate fungicide modes of action and use the few known materials that reduce Leaf Mold.”
- “Scout tomatoes regularly and remove diseased leaves.”
- “Plant tomatoes where no tomatoes, potatoes, peppers or eggplants have been for the past 3–4 years.”
- “Keep leaves as dry as possible. Water with drip irrigation or a soaker hose. Water in the morning so leaves dry quickly.”
- “Stake or cage plants. Space plants so that air flows between them.”
- “When pruning lower tomato leaves, clean the cutting tool with 70% rubbing alcohol to limit the spread of disease.”
- “Control volunteer potato plants, as infected plants can grow from infected tubers.”
- “Tomato plants can tolerate high levels of leaf loss from leaf spot diseases without affecting the number of tomatoes produced by the plant.”

LLM Answer

Diagnosis: Based on the information provided, the leaf from the **Tomato tree** is most likely affected by **Lateblight**. **Visual Symptoms:** The leaf exhibits large, dark brown blotches with a green-gray edge, not confined by major leaf veins. These symptoms are characteristic of Lateblight, a fungal disease caused by *Phytophthora infestans*.

Disease Characteristics: Lateblight is a water mold that favors cool (60°F to 70°F), damp conditions. It is a highly mutagenic disease, meaning it can produce new strains rapidly, making management challenging. The disease can infect leaves, stems, tomato fruit, and potato tubers.

Management Tactics: To manage Lateblight, the grower should:

1. **Remove and destroy infected plant material:** Remove and dispose of any infected leaves, stems, and fruit to prevent the spread of the disease.
2. **Improve air circulation:** Increase air circulation around the plants by staking or caging them, pulling weeds, and spacing plants far apart. This will help reduce humidity and prevent the spread of the disease.
3. **Use resistant varieties:** Plant resistant tomato varieties, such as Mountain Magic, Juliet, Defiant PhR, Verona, or Jasper, which have been bred to resist Lateblight.
4. **Maintain good crop rotation:** Rotate crops to non-susceptible crops like small grains or brassica crops for three years to break the disease cycle.
5. **Monitor for disease progression:** Regularly inspect the plants for signs of disease progression and take action promptly to prevent the disease from spreading.
6. **Avoid overwatering:** Avoid overwatering, which can create an ideal environment for the disease to develop.
7. **Use fungicides as a last resort:** If the disease is severe, use fungicides as a last resort, but always follow the label instructions and take necessary precautions to avoid harming the environment and non-target organisms.

It is essential to note that Lateblight is a complex disease, and a combination of these management tactics may be necessary to effectively control its spread.

Ground Truth

Okay, let's break down the likely diagnosis of Late Blight on a tomato leaf based on the image and analysis, and then discuss management strategies.

What's Wrong with the Leaf?

Based on the description, the leaf is showing signs of a disease, specifically Late Blight. The key indicators are likely:

- **Irregular, water-soaked lesions:** These are the initial spots that appear on the leaves. They often start small and quickly expand.

- **Dark brown or black discoloration:** As the lesions develop, they turn dark brown or black.
- **White, fuzzy growth (sporulation) on the underside of the leaf:** This is a crucial diagnostic feature. The white fuzz is the *sporangia* (structures containing spores) of the *Phytophthora infestans* pathogen. This is most visible in humid conditions.
- **Rapid spread:** Late blight is known for its aggressive and rapid spread, especially in favorable (cool and wet) weather.

The Disease: Late Blight

- **Causal Agent:** *Phytophthora infestans* (an oomycete, often called a water mold, not a true fungus).
- **Host Range:** While most known for affecting tomatoes and potatoes, it can also affect other members of the Solanaceae family (e.g., peppers, eggplants).
- **Conditions Favoring Disease:** Cool, wet weather (temperatures between 60–80°F or 15–27°C) and high humidity are ideal for Late Blight development and spread. Free moisture on the leaves is essential for infection.

Visual Symptoms (Detailed)

- **Leaf Lesions:**
 - Start as small, water-soaked spots, often near the leaf margins or tips.
 - Expand rapidly, becoming irregular in shape.
 - Turn dark brown to black, often with a slightly greasy appearance.
 - A characteristic feature is the presence of a light green to yellow halo around the lesion's edge in the early stages.
 - On the underside of the leaf, especially in humid conditions, you'll see the white, fuzzy growth (sporulation) at the edges of the lesions. This is the pathogen producing spores.
- **Stem Lesions:** Dark brown to black lesions can also appear on the stems of the tomato plant. These lesions can girdle the stem, leading to plant death.
- **Fruit Symptoms:**
 - Firm, brown, greasy-looking blotches on the fruit.
 - These blotches can quickly rot the fruit, making it inedible.
 - Fruit infection often occurs later in the disease cycle.

How the Grower Should Manage the Disease

Late Blight is a serious disease that requires a multi-pronged approach to manage:

1. Prevention is Key:

- **Start with Disease-Free Seedlings:** Purchase transplants from reputable nurseries that are certified disease-free.
- **Choose Resistant Varieties:** Select tomato varieties that have some level of resistance to Late Blight. While resistance isn't immunity, it can significantly slow down the disease's progress.
- **Good Air Circulation:** Space plants adequately to promote good air circulation. This helps leaves dry quickly and reduces humidity.
- **Avoid Overhead Watering:** Water at the base of the plants to keep the foliage dry. If overhead watering is unavoidable, do it early in the day so the leaves have time to dry before nightfall.
- **Monitor Regularly:** Inspect plants frequently for early signs of the disease. Early detection is crucial.

2. Early Detection and Action:

- **Remove Infected Leaves Immediately:** If you spot any leaves with suspicious lesions, remove them promptly and dispose of them properly (burn them, bury them deeply, or bag them and discard them in the trash – *do not compost*).
- **Prune Lower Leaves:** Prune the lower leaves of the plant to improve air circulation and reduce the chance of soil splash carrying the pathogen up to the foliage.

3. Fungicides (If Necessary):

- **Protectant Fungicides:** Apply protectant fungicides *before* the disease appears, especially if Late Blight has been reported in your area or if weather conditions are favorable. These fungicides create a barrier on the leaf surface that prevents the pathogen from infecting. Examples include chlorothalonil, mancozeb, and copper-based fungicides. *Follow label instructions carefully.*
- **Systemic Fungicides:** If Late Blight is already present, systemic fungicides may be necessary. These fungicides are absorbed by the plant and can kill the pathogen from within. Examples include those containing active ingredients like mefenoxam, dimethomorph, or protectant fungicides mixed with systemic fungicides. *Follow label instructions carefully.*

- **Rotate Fungicides:** To prevent the pathogen from developing resistance to fungicides, rotate between different classes of fungicides with different modes of action.
- **Organic Options:** Copper-based fungicides are often used in organic gardening, but they are protectants and must be applied preventatively.

4. Sanitation:

- **Clean Up Plant Debris:** After the growing season, remove and destroy all tomato and potato plant debris. Do not compost it.
- **Control Volunteer Plants:** Eliminate any volunteer tomato or potato plants that may sprout up in the spring, as they can serve as a source of inoculum for the disease.

5. Community Awareness:

- **Report Outbreaks:** If you suspect Late Blight, report it to your local agricultural extension office. This helps track the disease and alert other growers in the area.
- **Communicate with Neighbors:** If you grow tomatoes or potatoes, talk to your neighbors about Late Blight and encourage them to monitor their plants as well. Coordinated efforts are essential to control the disease.

Important Considerations:

- **Accurate Diagnosis:** It's crucial to confirm the diagnosis of Late Blight. If you're unsure, contact your local agricultural extension office for assistance.
- **Fungicide Resistance:** *Phytophthora infestans* can develop resistance to fungicides. Therefore, it's essential to use fungicides judiciously and rotate between different classes of fungicides.
- **Weather Monitoring:** Pay close attention to weather forecasts. If cool, wet weather is predicted, be prepared to take preventative measures.
- **Organic vs. Conventional:** The management strategies will differ depending on whether you are gardening organically or conventionally. Organic growers will rely more on preventative measures, resistant varieties, and copper-based fungicides.

By implementing these management strategies, growers can significantly reduce the risk of Late Blight and protect their tomato crops. Remember that early detection and prompt action are critical for successful control.

Αποτελέσματα αξιολόγησης του VGRS Pipeline

Μετρική	Μέση τιμή (σε 100 δείγματα)
Faithfulness	0.817
Answer Relevance	0.642

Πίνακας 4.4: Αποτελέσματα αξιολόγησης του VGRS Pipeline με RAGAS

Η αξιολόγηση με RAGAS έδειξε μέσο faithfulness 0.817 και μέσο answer relevance 0.642 σε 100 δείγματα (Πίνακας 4.4), γεγονός που υποδηλώνει ότι οι απαντήσεις του συστήματος είναι γενικά καλά τεκμηριωμένες από το ανακτημένο context (χαμηλή τάση για hallucination), αλλά η συνάφειά τους με τα ερωτήματα παραμένει μέτρια. Πρακτικά, το pipeline μας «πατάει» σωστά πάνω στις πηγές, όμως χρειάζεται βελτίωση στον μηχανισμό ανάκτησης και διατύπωσης απαντήσεων ώστε να στοχεύει πιο άμεσα το ζητούμενο της ερώτησης. Αυτό δείχνει ότι επόμενα βήματα όπως καλύτερη ρύθμιση chunking και top-k, χρήση query expansion/multi-query, ενσωμάτωση cross-encoder reranker και αυστηρότερες οδηγίες σύνθεσης απάντησης (π.χ. ζητούμενα bullets/αιτιολόγηση/παραπομπές) μπορούν να αυξήσουν τη συνάφεια χωρίς να θυσιάσουν τη γειωμένη τεκμηρίωση.

4.3 Πειράματα MES

4.3.1 Επιλογή Μοντέλου

Ο στόχος της παρούσας αρχιτεκτονικής είναι η υλοποίηση ενός fine-tuned vision-language model (VLM) το οποίο θα λειτουργεί ως φυτοπαθολόγος, προσφέροντας ανάλυση της ασθένειας με βάση την εικόνα ενός φύλλου. Για τον σκοπό αυτό χρησιμοποιείται το Qwen2.5-VL-7B, ένα open-source μοντέλο χαμηλών παραμέτρων. Πρόκειται για ένα από τα ισχυρότερα VLMs της κατηγορίας του, καθώς κατατάσσεται στην 8η θέση ανάμεσα στα καλύτερα open-source models με παραμέτρους μεταξύ 4 και 10 δισεκατομμυρίων στο OpenVLM Leaderboard του OpenCompass [55] [2]. Το γεγονός αυτό επιβεβαιώνει την ποιότητα και την αξιοπιστία του μοντέλου.

Η επιλογή του Qwen2.5-VL αντί κάποιου άλλου μοντέλου υψηλότερης κατάταξης στο leaderboard έγινε για δύο βασικούς λόγους: Πρώτον, επιθυμούμε να ξεκινήσουμε την εκπαίδευση από μια ουδέτερη (clean-slate) κατάσταση, χωρίς να χρησιμοποιήσουμε ήδη fine-tuned εκδόσεις του ίδιου μοντέλου, όπως συμβαίνει με άλλες καταχωρήσεις υψηλής βαθμολογίας. Δεύτερον, η διαθεσιμότητά του στο Unsloth επιτρέπει την αποδοτική εκπαίδευση σε υπολογιστικά περιβάλλοντα περιορισμένων πόρων, κάτι που είναι κρίσιμο για τις ανάγκες της εργασίας αυτής.

Rank ▲	Method ▲	Param (B) ▲	Language Model ▲
1	SAIL-VL-1.6-8B	8.33	Qwen2.5-7B
2	InternVL 3-8B	7.94	Qwen2.5-7B
3	InternVL 3-9B	9.14	InternLM3-8B
4	Ola-7b	8.88	Qwen2.5-7B
5	WeThink-Qwen2.5V L-7B	8.29	Qwen2.5-7B
6	SAIL-VL-1.5-8B	7.95	Qwen2.5-7B
7	Ovis2-8B	8.94	Qwen2.5-7B
8	Qwen2.5-VL-7B	8.29	Qwen2.5-7B

Σχήμα 4.19: Οι 8 καλύτερες θέσεις του OpenVLM Leaderboard για μοντέλα όρασης από 4 έως 10 δισ. παραμέτρους

4.3.2 Fine Tuning Vision-LLM με χρήση Image-to-Text Dataset

Δημιουργία Synthetic Dataset

Για να καταστεί εφικτή η εκπαίδευση του μοντέλου όρασης, ήταν απαραίτητο να χρησιμοποιηθεί ένα σύνολο δεδομένων, το οποίο να συνδέει εικόνες ασθενών φύλλων με κείμενο που περιγράφει τόσο τα χαρακτηριστικά της ασθένειας, όσο και τους τρόπους αντιμετώπισής της. Στις αρχές ακόμα της εκπόνησης της συγκεκριμένης εργασίας, δεν υπήρχε ακόμα διαθέσιμο κάποιο τέτοιο σετ δεδομένων, γεγονός που οδήγησε στη δημιουργία ενός προσαρμοσμένου synthetic dataset, ειδικά σχεδιασμένου για τις ανάγκες της παρούσας μελέτης.

Πιο συγκεκριμένα, χρησιμοποιήθηκε το Gemini 2.0 Flash μέσω API call για την αυτόματη δημιουργία συνθετικών δειγμάτων. Αρχικά, η διαδικασία περιορίστηκε αποκλειστικά στις ασθένειες που σχετίζονται με τα φύλλα της μηλιάς, ώστε να αξιολογηθεί η απόδοση της εκπαίδευσης του μοντέλου σε περιορισμένο εύρος δεδομένων. Κάνοντας χρήση των κατηγοριών ασθενειών μήλου που περιλαμβάνονται στο PlantVillage dataset, ζητήθηκε από το Gemini να παραγάγει για κάθε κατηγορία μια παράγραφο που περιγράφει τα οπτικά χαρακτηριστικά της ασθένειας και προτείνει μεθόδους αντιμετώπισης — τόσο φυσικές, όσο και με χρήση φυτοφαρμάκων. Στην περίπτωση υγιούς φύλλου, το μοντέλο κλήθηκε να παραγάγει συμβουλές που βοηθούν στη διατήρηση της υγείας του δέντρου.

Εκπαίδευση Qwen2.5-VL-7B με τη βοήθεια του Unsloth

Το Unsloth είναι βιβλιοθήκη για αποδοτικό fine-tuning πολυτροπικών (και μη) μοντέλων. Μειώνει τις απαιτήσεις μνήμης και τον χρόνο εκπαίδευσης μέσω κβαντοποίησης 4-bit και βελτιστοποιημένου gradient checkpointing, ενώ ενσωματώνεται ομαλά με PEFT/LoRA και το οικοσύστημα TRL. Στη παρούσα εργασία, το FastVisionModel φορτώνει το Qwen2.5-VL-7B-Instruct-bnb-4bit με load_in_4bit και use_gradient_checkpointing="unsloth", εφαρμόζοντας LoRA σε επίπεδα vision, language, attention και MLP. Η εκπαίδευση γίνεται με SFTTrainer και τον απαραίτητο UnslothVisionDataCollator, αφού ενεργοποιηθεί η λειτουργία εκπαίδευσης με FastVisionModel.for_training. Η παρακάτω διαμόρφωση υπερπαραμέτρων επιλέχθηκε ως για τον περιορισμό πόρων και δεδομένων:

- **load_in_4bit = true:** Κβαντοποίηση 4-bit για δραστική μείωση VRAM χωρίς σημαντική απώλεια ποιότητας σε LoRA fine-tuning. Επιτρέπει εκπαίδευση σε κάρτες 8-12GB.
- **use_gradient_checkpointing = "unsloth":** Μειώνει τη μνήμη ενεργοποιώντας αποδοτικό checkpointing με ελάχιστο compute κόστος, επιτρέποντας μεγαλύτερο context και σταθερότητα.
- **finetune_vision/language/attention/mlp = true:** Επιτρέπει προσαρμογή, τόσο στα οπτικά, όσο και στα γλωσσικά υποσυστήματα, κρίσιμο για το domain shift της φυτοπαθολογίας (υφές, κηλίδες, όροι).
- **LoRA: r = 16, lora_alpha = 16, lora_dropout = 0, bias = "none":** r και alpha ισορροπούν ικανότητα μάθησης και σταθερότητα για μοντέλο 7B. Το Dropout = 0 συχνά αποδίδει καλύτερα σε σύντομες προπονήσεις και η ρύθμιση bias="none" μειώνει τις παραμέτρους που ενημερώνονται κατά τη διάρκεια του fine-tuning.
- **optim = "adamw_8bit":** Ο AdamW σε 8-bit συμπιέζει τις καταστάσεις του βελτιστοποιητή, εξοικονομώντας μνήμη με παρόμοια σύγκλιση.
- **learning_rate = 5e-5:** Τυπικό εύρος για LoRA σε 7B· αρκετά υψηλό για γρήγορη προσαρμογή, όχι τόσο ώστε να αποσταθεροποιεί την εκπαίδευση.
- **lr_scheduler_type = "cosine", warmup_ratio = 0.05:** Ο «cosine» χρονοπρογραμματιστής με 5% warmup προσφέρει ομαλή αύξηση και σταδιακή μείωση του ρυθμού μάθησης, βελτιώνοντας τη γενίκευση.
- **per_device_train_batch_size = 2, gradient_accumulation_steps = 8:** Αποτελεσματικό batch size ≈ 16 ανά συσκευή με χαμηλή VRAM. Καλή ισορροπία signal-to-noise στον υπολογισμό βαθμίδων.
- **num_train_epochs = 2:** Λίγες εποχές περιορίζουν την υπερπροσαρμογή στο συνθετικό dataset, ενώ επαρκούν για αισθητή προσαρμογή με LoRA.

- **weight_decay = 0.01**: Ήπια κανονικοποίηση που μειώνει το overfitting, χωρίς να εμποδίζει την προσαρμογή.
- **max_length = 2048**: Επαρκές μήκος για image tokens και κείμενο προτροπών/απαντήσεων χωρίς αποκοπές.
- **remove_unused_columns = false, dataset_text_field = ""**, **dataset_kwargs = {"skip_prepare_dataset": true}**: Αποφεύγεται η αυτόματη απάλειψη απαραίτητων columns από την TRL, και παρακάμπτεται ο προσανατολισμένος-στο-κείμενο προεπεξεργαστής, όπως απαιτείται για vision fine-tuning.
- **save_strategy = "epoch", output_dir = "outputs"**: Απλή και αξιόπιστη αποθήκευση ανά εποχή για έλεγχο/επιστροφή σε καλύτερο checkpoint.
- **logging_steps = 1, report_to = "none"**: Πυκνή τηλεμετρία για έγκαιρη ανίχνευση αστάθειας, χωρίς εξωτερική αναφορά σε συστήματα όπως το weights and biases.
- **seed = 3407**: Αναπαραγωγιμότητα πειραμάτων και συγκρισιμότητα αποτελεσμάτων.
- **Λοιπά LoRA γνωρίσματα: use_rslora = false, loftq_config = None, random_state = 3407**: Δεν απαιτείται RS-LoRA/LoftQ στο παρόν budget/διάστημα εκπαίδευσης.

4.3.3 Fine Tuning Vision-LLM με χρήση PlantVillage VQA

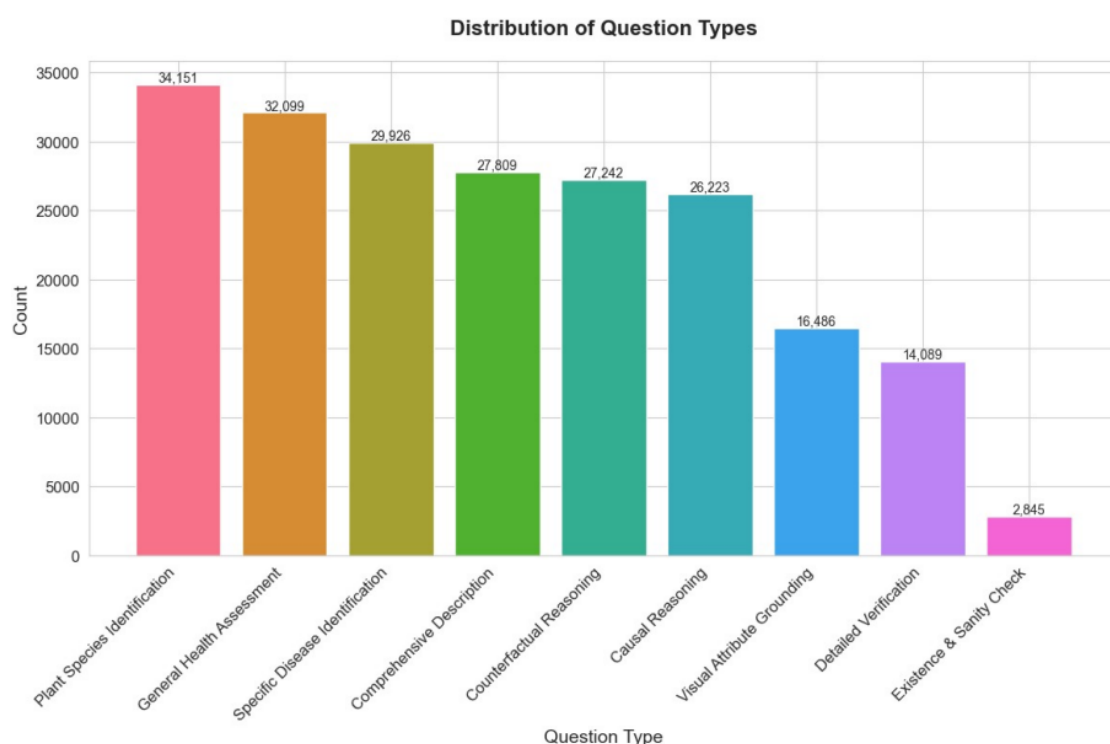
Κατά τη διάρκεια της πειραματικής περιόδου δημοσιεύθηκε ένα άρθρο [43] που παρουσίασε ένα dataset που ανταποκρινόταν ακριβώς στις ανάγκες μας! Το άρθρο προτείνει ένα multiturn visual question answering dataset, βασισμένο στο αρχικό PlantVillage Dataset. Για κάθε φωτογραφία δημιουργείται μια ακολουθία ερωτήσεων και απαντήσεων που αναφέρονται στη συγκεκριμένη εικόνα. Πιο συγκεκριμένα, οι ερωτήσεις εμπίπτουν στις παρακάτω κατηγορίες:

Πίνακας 4.5: Κατηγορίες ερωτήσεων (ενοποιημένος πίνακας, χωρίς στήλη επιπέδου)

Κατηγορία Ερώτησης	Σκοπός	Πρότυπο ερώτησης (παράδειγμα)
Υπαρξη & Έλεγχος Ορθότητας	Σχεδιάστηκε για φιλτράρισμα άσχετων ή εικόνων φόντου.	«What is the primary subject of this image?»
συνέχεια στην επόμενη σελίδα		

Κατηγορία Ερώτησης	Σκοπός	Πρότυπο ερώτησης (παράδειγμα)
Αναγνώριση Είδους Φυτού	Για ασθένειες που αφορούν συγκεκριμένα είδη, είναι απαραίτητος ο προσδιορισμός του φυτού-ξενιστή. Αυτή η κατηγορία ελέγχει τις δυνατότητες οπτικής κατηγοριοποίησης του μοντέλου με λεπτομέρεια. Τα πρότυπα σχεδιάστηκαν εξάγοντας το όνομα της καλλιέργειας από το όνομα του φακέλου. Δημιουργήθηκαν τόσο θετικά όσο και αρνητικά παραδείγματα για τη δημιουργία μιας ισχυρής εργασίας ταξινόμησης.	«Is this a [Crop] leaf?»
Γενική Εκτίμηση Υγείας	Ελέγχει την ικανότητα του μοντέλου να κάνει υψηλού επιπέδου, δυαδική ταξινόμηση για τη συνολική υγεία του φυτού.	«Is the plant in this image healthy?»
Σύνδεση Οπτικών Γνωρισμάτων	Δημιουργήθηκαν πρότυπα που περιγράφουν τα κανονικά οπτικά συμπτώματα για καθεμία από τις 38 ασθένειες, βάσει καθιερωμένης φυτοπαθολογικής βιβλιογραφίας.	«Does the leaf exhibit dark, concentric 'bullseye' rings?»
Λεπτομερής Επαλήθευση	Συνδυάζει αναγνώριση είδους και ασθένειας σε ένα σύνθετο ερώτημα. Τα πρότυπα αναφέρουν ρητά την καλλιέργεια και την ασθένεια, δημιουργώντας εξαιρετικά συγκεκριμένες δυαδικές ερωτήσεις.	«Is this [Crop] leaf infected with [Disease Name]?»
Αναγνώριση Συγκεκριμένης Ασθένειας	Άμεση, ανοικτού τύπου διαγνωστική εργασία: δεν δίνεται υποψήφια ασθένεια, το μοντέλο ανακαλεί και παρέχει τη σωστή από τη γνώση του.	«Please provide a diagnosis for the condition shown.»
Ολοκληρωμένη Περιγραφή	Ελέγχει την ικανότητα του μοντέλου να δημιουργεί ολιστικές, συνοπτικές περιγραφές. Τα πρότυπα ήταν ανοιχτές προτροπές για περιγραφικό κείμενο.	«Provide a full description of the plant and its condition.»
συνέχεια στην επόμενη σελίδα		

Κατηγορία Ερώτησης	Σκοπός	Πρότυπο ερώτησης (παράδειγμα)
Αιτιολογικός Συλλογισμός	Αυτό διερευνά εάν το μοντέλο μπορεί να συμπεράνει τον αιτιολογικό παράγοντα που ευθύνεται για τα ορατά συμπτώματα. Οι ερωτήσεις σχεδιάστηκαν για να ζητήσουν την αιτία, τον παθογόνο παράγοντα ή τον λόγο.	«What is the cause of the unhealthy appearance of this leaf?»
Αντιπραγματικός Συλλογισμός	Αυτή η εργασία ελέγχει την ικανότητα του μοντέλου να συλλογίζεται για μια υποθετική κατάσταση που είναι αντίθετη με τα οπτικά στοιχεία. Τα πρότυπα έθεσαν υποθετικά σενάρια.	«If this plant were healthy, what visual features would be different?»



Σχήμα 4.20: Κατανομή των ερωτήσεων του dataset

Αποτελέσματα Εκπαίδευσης

Η εκπαίδευση πραγματοποιήθηκε με τις ίδιες παραμέτρους που αναφέρθηκαν προηγουμένως. Το fine-tuned μοντέλο αξιολογήθηκε χρησιμοποιώντας, τόσο μετρικές εξειδικευμένες για την εκάστοτε απάντηση (όπως Exact Match, Token-based F1 Score, Accuracy), όσο και μετρικές σημασιολογικής ομοιότητας (όπως Cosine Similarity, BERT F1 Score).

Αξίζει να σημειωθεί ότι, όταν γίνεται αναφορά στη μετρική Accuracy, αυτή υπολογίζεται μόνο για τις ερωτήσεις που απαιτούν από το μοντέλο να αναγνωρίσει τον τύπο της ασθένειας του φυτού. Σε αυτή την περίπτωση, ελέγχεται αν το μοντέλο επιστρέφει την ορθή κλάση. Επιπλέον, γίνεται διάκριση ανάμεσα στο Token-based F1 Score, το οποίο αξιολογεί την αλληλοκάλυψη λέξεων-κλειδιών σε σχέση με την αναμενόμενη απάντηση, και στο BERT F1 Score, το οποίο αξιολογεί σημασιολογικά την απόδοση του μοντέλου.

Η αξιολόγηση έγινε στο test split του PlantVillage-VQA dataset και τα αποτελέσματα παρουσιάζονται στον Πίνακα 4.6.

Πίνακας 4.6: Αποτελέσματα Αξιολόγησης του Fine-tuned Μοντέλου στο PlantVillage-VQA Test Split

Μετρική	Τιμή
Exact Match	0.612
Token-based F1 Score	0.745
BERT F1 Score	0.801
Cosine Similarity	0.86
Accuracy (Διάγνωση Ασθένειας)	0.783

Τα αποτελέσματα της αξιολόγησης δείχνουν πως το μοντέλο αποδίδει ιδιαίτερα καλά στο καθήκον της αναγνώρισης ασθενειών σε εικόνες φύλλων. Η υψηλή τιμή στο BERT F1 Score (80.1%) και η σημαντική Accuracy (78.3%) όσον αφορά τη σωστή διάγνωση της ασθένειας, επιβεβαιώνουν πως το μοντέλο έχει πλέον εδραιωμένες γνώσεις στον συγκεκριμένο τομέα και μπορεί να θεωρηθεί ένας ικανός «ψηφιακός φυτοπαθολόγος». Ωστόσο, η εξειδικευμένη αυτή εκπαίδευση έχει οδηγήσει σε απώλεια της γενικής ικανότητάς του να απαντά ερωτήσεις που δεν σχετίζονται άμεσα με τη διάγνωση φυτικών ασθενειών. Το μοντέλο εμφανίζει πλέον έντονη εξάρτηση από τη δομή και το περιεχόμενο του PlantVillage-VQA dataset. Αξίζει επίσης να σημειωθεί πως το εν λόγω dataset, αν και ιδιαίτερα χρήσιμο για σκοπούς αναγνώρισης, δεν περιλαμβάνει πληροφορίες σχετικά με τρόπους αντιμετώπισης των ασθενειών.

4.3.4 Ενσωμάτωση Web Scraper (Tavily) με tool calling

Εφόσον ο στόχος του συστήματος δεν περιορίζεται μόνο στην αναγνώριση ασθενειών, αλλά συμπεριλαμβάνει και την παροχή προτάσεων αντιμετώπισης προς τον χρήστη, το fine-tuned μοντέλο από μόνο του δεν είναι επαρκές. Λαμβάνοντας υπόψη ότι το μοντέλο, μετά την εξειδίκευσή του, παρουσιάζει μειωμένη ικανότητα στη γενική σύνθεση λόγου, κρίθηκε απαραίτητο να προστεθεί ένας επιπλέον κόμβος στο σύστημα για την παραγωγή της τελικής απάντησης.

Πιο συγκεκριμένα, το σύστημα που δημιουργήθηκε είναι το εξής: δίνεται στο finetuned Qwen μοντέλο η εικόνα και του τίθενται οι εξής τρεις ερωτήσεις:

1. What is the primary subject of this image? (Έλεγχος ύπαρξης και ορθότητας)

```
User input: What is the disease? Provide only natural ways to manage the disease

> Entering new AgentExecutor chain...

Invoking: `get_knowledge_from_the_web` with `{'query': 'natural management of Strawberry Leaf scorch'}`
```

Σχήμα 4.21: Παράδειγμα χρήσης εργαλείου κατά τη διάρκεια της εκτέλεσης

2. What is the crop? (Αναγνώριση του είδους του φυτού)
3. Provide a full description of the plant and its condition. (Ολοκληρωμένη περιγραφή)

Οι απαντήσεις στις παραπάνω ερωτήσεις διοχετεύονται σε ένα δεύτερο, τοπικό γλωσσικό μοντέλο, το GPT:OSS (μέσω Ollama). Το μοντέλο αυτό λειτουργεί σαν agent ο ρόλος του οποίου είναι η σύνθεση της τελικής ανάλυσης, η οποία οφείλει να περιλαμβάνει τις προτεινόμενες μεθόδους αντιμετώπισης της ασθένειας (με βάση αυτό που έχει ζητήσει ο χρήστης) και την ασθένεια που επηρεάζει το δέντρο.

Για να καλυφθεί η ανάγκη ανάκτησης γνώσης για τους τρόπους αντιμετώπισης, υλοποιήθηκε εργαλείο αναζήτησης διαδικτυακών πηγών με χρήση του Tavily. Το Tavily είναι ένα εργαλείο web search, το οποίο, λαμβάνοντας ως είσοδο μια ερώτηση (query), επιστρέφει αποσπάσματα από τα τρία πιο σχετικά άρθρα, συνοδευόμενα από τους αντίστοιχους συνδέσμους (URLs). Αν και το Tavily παρέχει τη δυνατότητα επιστροφής και συνοπτικών περιλήψεων, στη συγκεκριμένη υλοποίηση χρησιμοποιείται αποκλειστικά για την ανάκτηση των αρχικών αποσπασμάτων. Η τελική περίληψη και σύνθεση της πληροφορίας πραγματοποιείται εξολοκλήρου από το γλωσσικό μας μοντέλο, το οποίο αξιοποιεί τα επιστραφέντα αποσπάσματα ως εξωτερική γνώση για την παραγωγή της απάντησης.

Η διαδικασία της tool calling επιτρέπει στο LLM να καλεί εξωτερικές συναρτήσεις, όταν απαιτείται πρόσβαση σε πληροφορία εκτός του ίδιου του μοντέλου. Στη δική μας περίπτωση, ορίστηκε ένα μόνο εργαλείο, το `get_knowledge_from_the_web`, το οποίο δέχεται ως είσοδο ένα ερώτημα (`query: str`) και επιστρέφει μια απάντηση τύπου `str`, η οποία περιλαμβάνει συγχωνευμένα αποτελέσματα από τις τρεις πιο σχετικές διαδικτυακές πηγές. Εφόσον το μοντέλο κρίνει απαραίτητη τη χρήση του εργαλείου, θα το καλέσει αυτόματα, δημιουργώντας επιτόπου ένα προσαρμοσμένο query για να αντλήσει πληροφορίες από το διαδίκτυο σχετικές με την ερώτηση του χρήστη, όπως φαίνεται στο Σχήμα 4.21.

Η προσέγγιση αυτή ακολουθεί τη φιλοσοφία του **agentic Retrieval-Augmented Generation (RAG)**, όπου το γλωσσικό μοντέλο λειτουργεί ως ενεργός πράκτορας που καθορίζει πότε και πώς θα αναζητήσει εξωτερικές πληροφορίες. Η αξιολόγηση της αποτελεσματικότητας αυτού του συστήματος παρουσιάζεται στην επόμενη ενότητα.

4.3.5 Αξιολόγηση MES

Η ικανότητα αναγνώρισης ασθενειών και η ορθότητα της διάγνωσης έχουν ήδη αξιολογηθεί στον Πίνακα 4.6. Συνεπώς, για την ολοκληρωμένη αξιολόγηση του συστήματος απομένει να εξεταστεί κατά πόσο το κεντρικό γλωσσικό μοντέλο (LLM):

Πίνακας 4.7: Αποτελέσματα Αξιολόγησης Απαντήσεων του MES με RAGAS

Μετρική	Μέση Τιμή
Faithfulness	0.7367
LLM Context Precision (No Reference)	0.6320
Guideline Adherence	0.8026
Answer Relevancy	0.6277

- χρησιμοποιεί ορθά τα εργαλεία που του παρέχονται,
- αξιοποιεί επαρκώς τη γνώση που αντλεί από αυτά,
- απαντά με επάρκεια στην ερώτηση του χρήστη,
- συμμορφώνεται με τις οδηγίες ή τις προτιμήσεις που εκφράζει ο χρήστης (π.χ. αποκλειστική χρήση φυσικών μεθόδων αντιμετώπισης).

Για τον σκοπό αυτό, χρησιμοποιήθηκαν οι μετρικές `faithfulness`, `context_precision`, `answer_relevancy` και `guideline_aspect`. Όλες οι μετρικές, εκτός από την `guideline_aspect`, έχουν αναλυθεί εκτενώς στην Ενότητα 4.1.

Η μετρική `guideline_aspect` υλοποιείται στο RAGAS μέσω του μηχανισμού `Aspect Critic`. Πρόκειται για ένα πλαίσιο αξιολόγησης βασισμένο σε γλωσσικά μοντέλα (LLMs), το οποίο επιτρέπει την ευέλικτη ποιοτική αποτίμηση μιας απάντησης με βάση συγκεκριμένες πτυχές (`aspects`) που καθορίζονται. Ο τρόπος λειτουργίας του στηρίζεται στην εξής αρχή: δίνεται στο LLM ένα πρότυπο `prompt`, το οποίο περιλαμβάνει την ερώτηση του χρήστη, την απάντηση του μοντέλου, και μία ή περισσότερες οδηγίες ή κριτήρια που πρέπει να πληροί η απάντηση. Το LLM καλείται να αξιολογήσει εάν η απάντηση συμμορφώνεται με τις συγκεκριμένες οδηγίες και επιστρέφει μία ποσοτική εκτίμηση σε μορφή βαθμολογίας, μεταξύ 0 και 1. Στη δική μας περίπτωση ο κανονισμός πάνω στον οποίο γίνεται ο έλεγχος είναι κατά πόσο το παραγόμενο κείμενο είναι συνεπές με οδηγίες όπως «δώσε μόνο φυσικούς τρόπους αντιμετώπισης», «απόφυγε χημικά σκευάσματα» ή «γράψε απλά και κατανοητά».

Η αξιολόγηση πραγματοποιήθηκε σε 1000 εικόνες από το `test split` του `PlantVillage VQA dataset`. Για κάθε εικόνα, το σύστημα έλαβε ως είσοδο, μαζί με την εικόνα, μια ερώτηση τύπου VQA και αξιολογήθηκε η παραγόμενη απάντηση βάσει των προαναφερθεισών μετρικών. Τα αποτελέσματα συνοψίζονται στον Πίνακα 4.7.

Από τα αποτελέσματα του ελέγχου φαίνεται ότι το σύστημα παρουσιάζει υψηλή συμμόρφωση προς τις οδηγίες του χρήστη, όπως καταδεικνύεται από τη μετρική `Guideline Adherence` (0.8026). Αυτό σημαίνει ότι το μοντέλο προσαρμόζει επιτυχώς τις απαντήσεις του σε απαιτήσεις, όπως η αποκλειστική χρήση φυσικών μεθόδων αντιμετώπισης ή η αποφυγή χημικών σκευασμάτων.

Από τη μετρική `Faithfulness` (0.7367) φαίνεται επίσης ότι το μοντέλο βασίζεται σε μεγάλο βαθμό στις πληροφορίες που του παρέχονται από το `web-search` και αποφεύγει να εφευρίσκει

μη τεκμηριωμένο περιεχόμενο. Αυτή η μετρική μπορεί να διορθωθεί περαιτέρω με prompt engineering. Αντιθέτως, οι χαμηλότερες τιμές στις μετρικές LLM Context Precision (No Reference) (0.6320) και Answer Relevancy (0.6277) υποδεικνύουν ότι οι πληροφορίες που αξιοποιεί το μοντέλο δεν είναι επαρκώς σχετικές με το αρχικό ερώτημα του χρήστη, με αποτέλεσμα η παραγόμενη απάντηση να μην καλύπτει πλήρως το ζητούμενο. Στην παραπάνω απόδοση φαίνεται να συμβάλλουν τρεις βασικοί παράγοντες: (α) η αδυναμία του γλωσσικού μοντέλου να συνθέσει ένα αποδοτικό query προς το Tavily, (β) η συχνή παρουσία «θορύβου» στα κείμενα που επιστρέφει το Tavily, όπου με τον όρο «θόρυβος» αναφερόμαστε σε αποσπάσματα με κακή χωρική ή λογική δομή ή ελλιπή συμφραζόμενα, τα οποία δυσχεραίνουν την επεξεργασία και κατανόηση του περιεχομένου από το γλωσσικό μοντέλο και (γ) το περιορισμένο context window του LLM, που καθιστά προβληματική την ταυτόχρονη εισαγωγή μεγάλου όγκου πληροφορίας, οδηγώντας σε απώλεια εστίασης κατά τη σύνθεση της τελικής απάντησης.

Παράδειγμα λειτουργίας MES pipeline



Σχήμα 4.22: Φύλλο φράουλας με καψαλισιά (Strawberry leaf with leaf scorch)

User Query

What is the disease? Provide only natural ways to manage the disease

Usage of fine-tuned Vision LLM

Q1:What is the primary subject of this image

A1:This is a Strawberry leaf suffering from Leaf scorch.

Q2:What is the crop?

A2:Strawberry

Q3:Provide a full description of the plant and its condition.

A3:The leaf has a rough texture with several circular dark purple to black spots scattered across its surface, mainly near the edges and along the veins

Web Search Tool

Query: 'natural management of Strawberry leaf scorch'

Retrieved Context: can reduce the abundance of leaf scorch. Where the fungus has been a problem, remove all strawberry plant debris from the site at the end of the growing season and dispose of it away from where strawberries will be planted the next growing season. To keep foliage drier, use drip or furrow irrigation instead of overhead sprinkling. Irrigate in the morning instead of late during the day or at night so that foliage dries more quickly. Where leaf scorch has been a problem, these recommended [...] even though it is still edible. Solutions Seek resistant varieties for planting where leaf scorch has been a problem or conditions are commonly wet. Grow strawberries as an annual so the fungus cannot persist in the same plant for multiple growing seasons and become more abundant over time. Plant strawberries in a sunny, open location and on a berm or mound of raised soil. Avoid crowding plants and control nearby weeds. These cultural practices will increase air circulation around plants and [...] strawberries—*Diplocarpon earlium* = *Marssonina fragariae* Leaf scorch is a fungal disease that spots and kills strawberry leaves and reduces fruit abundance and quality. Identification Leaf scorch causes brown to purple spots on the upper surface of strawberry leaves. When infections are abundant, the leaf tissue between lesions commonly takes on a bright red to purplish color, but coloration varies with the cultivar, temperature, and other factors. Spotted leaves eventually turn brown, dry up,

Brown edges with red field and purple spots

This is advanced leaf scorch. Remove affected leaves if possible to increase airflow. Clean up the strawberry bed at the end of the growing season since the fungus can overwinter on fallen leaves. Start new plants to replace severely infected older plants next year.

Gray to tan blotches [...] If your strawberry leaves are turning bright red-orange

and have small solid purple spots, they have leaf scorch. Eventually, the edges will turn brown and dry, appearing scorched. Remove the infected leaves and weeds to increase airflow.

Yellow

Overwatering [...] Dark purple to reddish spots with white, tan, or gray centers on young tender leaves are a sign of strawberry leaf spot.

Solid dark purple spots

Leaves that have small, solid dark purple spots without light-colored centers are in the early stages of a leaf scorch infection. If left to spread, the leaves will turn red-orange and eventually develop brown scorched edges.

This is a fungal disease. Remove the affected leaves and increase airflow between plants.

Like most of the fungal pathogens mentioned above, leaf scorch (*Diplocarpon earlianum*) takes hold in extra wet weather. It is one of the most common foliar diseases in strawberries, but it doesn't typically cause any problems with the fruit itself.

How to Identify

Leaf scorch makes strawberry leaves look like they've been burnt by something.

Beginning signs of leaf scorch include: [...] Large sunken lesions where spots have merged together Purplish to brown colored areas that can girdle leaves from the base Brown, curled up leaves Scorched dying leaves Weakened vigor Dead leaves, flowers, or fruit (in severe infections)

How to Prevent

The preventative measures for leaf scorch are very similar to other fungal strawberry diseases. It all comes down to reducing plant stress and preventing excess moisture: [...] When the leaf scorch fungus takes hold of your leaves, there's not much you can do. It may look unsightly but it probably won't harm your fruit yields. The only treatment options are removing infected leaves and practicing the preventative measures outlined above.

Alternaria Fruit SpotThe strawberry leaf in the image is suffering from Leaf Scorch, a fungal disease. To manage it naturally, remove all strawberry plant debris at the end of the growing season. Ensure good air circulation by planting in a sunny, open location on raised soil, avoiding crowding, and controlling weeds. Use drip or furrow irrigation instead of overhead sprinkling, and irrigate in the morning so foliage dries quickly. Remove affected leaves to increase airflow.

Final Analysis (from central LLM)

The strawberry leaf in the image is suffering from Leaf Scorch, a fungal disease. To manage it naturally, remove all strawberry plant debris at the end of the growing season. Ensure good air circulation by planting in a sunny, open location on raised soil, avoiding crowding, and controlling weeds. Use drip or furrow irrigation instead of overhead sprinkling, and irrigate in the morning so foliage dries quickly. Remove affected leaves to increase airflow.

4.4 Σύγκριση των Συστημάτων VGRS και MES

Το VGRS αποδεικνύεται ισχυρό σύστημα ως καθαρός οπτικός ταξινομητής: συνδυάζει ανίχνευση και ταξινόμηση με το YOLO pipeline και πετυχαίνει πολύ υψηλή ακρίβεια στο σύνολο δεδομένων της Εικόνας 4.13. Το μειονέκτημα αυτής της προσέγγισης είναι ότι κατά την παραγωγή της ανάλυσης, επειδή το μοντέλο δεν «βλέπει» πραγματικά την εικόνα, βασίζεται στην υπόθεση ότι το YOLO έχει ταξινομήσει σωστά την ασθένεια και, με αυτή την πρόβλεψη, επιχειρεί να συντάξει το τελικό κείμενο. Αν η πρόβλεψη είναι εσφαλμένη — π.χ. επειδή οι ανιχνευτές δεν έχουν εκπαιδευτεί σε αυτόν τον τύπο ασθένειας/φυτού ή επειδή η εικόνα που έχει δοθεί για ανάλυση είναι κακής ποιότητας — τότε το σύστημα θα παραγάγει κείμενο που, παρόλο που θα περιγράφει σωστά την αντιμετώπιση της ασθένειας, δεν θα αντιστοιχεί στη φωτογραφία που δόθηκε ως είσοδος στο σύστημα.

Η αρχιτεκτονική MES αντιμετωπίζει το παραπάνω ζήτημα, καθώς αποτελεί ένα multimodal σύστημα κατανόησης και παραγωγής απαντήσεων. Παρότι υστερεί έναντι του VGRS στην καθαρή επίδοση ταξινόμησης εικόνων (Σχήμα 4.6), κατά τη διατύπωση απαντήσεων αξιοποιεί άμεσα τα ορατά συμπτώματα και μπορεί να περιγράψει με αξιοπιστία την κατάσταση του φύλλου. Η περιγραφή αυτή μπορεί να συνδυαστεί με διαδικτυακή έρευνα, ώστε να προτείνονται τελικά τρόποι διαχείρισης της ασθένειας, ακόμη κι όταν ο τύπος του φύλλου ή η ασθένεια δεν ανήκουν με σαφήνεια στο σύνολο πάνω στο οποίο εκπαιδεύτηκε το μοντέλο.

Ενοποιώντας τα δύο συστήματα σε ένα, οι αδυναμίες τους αλληλοαναιρούνται: το YOLO pipeline του VGRS μπορεί να ταξινομεί με μεγάλη ακρίβεια τις ασθένειες στα φύλλα. Όταν όμως δεν μπορεί να πραγματοποιήσει ταξινόμηση με υψηλή βεβαιότητα, αξιοποιούμε το fine-tuned VLM για μια πιο open-ended διάγνωση, καλύπτοντας έτσι ένα ευρύτερο φάσμα ασθενειών και διατηρώντας ταυτόχρονα υψηλή ακρίβεια. Η υλοποίηση αυτή θα αναλυθεί παρακάτω.

4.5 Πειράματα VEAA

4.5.1 Υλοποίηση του Συστήματος με LangGraph

Η αρχιτεκτονική της Εικόνας 3.3 υλοποιήθηκε ως γράφος καταστάσεων (state machine) με το LangGraph. Κάθε κόμβος εκτελεί ένα σαφώς ορισμένο βήμα, ενώ οι μεταβάσεις μεταξύ κόμβων ελέγχονται από συγκεκριμένους κανόνες και ελέγχους εμπιστοσύνης. Αφού ο γράφος

ορίστηκε με τη χρήση του LangGraph, εκτυπώθηκε η ορισμένη αρχιτεκτονική για να ελεγχθεί η ορθότητά της όπως φαίνεται στο Σχήμα 4.23.



Σχήμα 4.23: Η VEEA αρχιτεκτονική υλοποιημένη με LangGraph

Η είσοδος του συστήματος είναι μία εικόνα ενός άρρωστου φυλλώματος και μια ερώτηση. Στην ερώτηση μπορεί να δηλώνονται πρόσθετα στοιχεία, όπως το πιθανό είδος φυτού (π.χ. μηλιά) ή προτιμήσεις για τρόπους αντιμετώπισης (π.χ. μόνο φυσικές πρακτικές), στοιχεία που πρέπει να ληφθούν υπόψιν κατά τη σύνθεση της τελικής απάντησης. Παρακάτω θα αναλυθεί η λειτουργία έκαστου κόμβου μαζί με τους ελέγχους που γίνονται για την μετάβαση από τον ένα κόμβο στον άλλο.

0) Ορισμός State Η μεταβλητή state αποτελεί βασικό στοιχείο της αρχιτεκτονικής του LangGraph, καθώς μέσω αυτής μεταφέρονται κρίσιμες πληροφορίες από κόμβο σε κόμβο. Με άλλα λόγια, το state λειτουργεί ως κοινός χώρος μνήμης που επιτρέπει την αλληλεπίδραση και τον συγχρονισμό μεταξύ των επιμέρους βημάτων της ροής. Οι πληροφορίες που κρίνονται απαραίτητες για τον συγκεκριμένο γράφο περιλαμβάνουν:

- **messages:** Τα μηνύματα της εσωτερικής σκέψης που γίνονται στο γράφο.
- **user_in:** Η αρχική είσοδος του χρήστη.
- **crop_type:** Ο τύπος καλλιέργειας που έχει επιλεγεί ή ανιχνευθεί.
- **toolsUsed:** Η λίστα εργαλείων που χρησιμοποιήθηκαν για την παραγωγή της τελικής ανάλυσης.
- **yolo_existence και yolo_existence_conf:** Πληροφορίες για το αν ανιχνεύθηκε φυτό στην εικόνα και με τι σιγουριά.

- **yolo_plant_type** και **yolo_plant_type_conf**: Ο τύπος φυτού που αναγνωρίστηκε και ο βαθμός εμπιστοσύνης της πρόβλεψης.
- **yolo_disease_type** και **yolo_disease_type_conf**: Η ασθένεια που εντοπίστηκε στο φυτό και η σχετική μέτρηση σιγουριάς του YOLO μοντέλου.
- **guideline_aspect**: Μετρική αξιολόγησης της απάντησης που θα δώσει το μοντέλο για το κατά πόσο ακολουθεί τους κανονισμούς που θέτει ο χρήστης στην ερώτηση του (π.χ. μόνο φυσικοί τρόποι αντιμετώπισης). Ο υπολογισμός της μετρικής και η αξιοποίηση της θα εξηγηθεί παρακάτω.

Η δομή αυτή διασφαλίζει ότι κάθε κόμβος έχει πρόσβαση στα απαραίτητα δεδομένα για να εκτελέσει τον ρόλο του σωστά, επιτρέποντας έτσι την εύρυθμη και ευέλικτη λειτουργία του γράφου.

1) Κόμβος Yolo_Preprocessing και αρχική ανάλυση. Αρχικά η εικόνα διέρχεται από τον κόμβο Yolo_Preprocessing. Εδώ χρησιμοποιείται το pipeline ταξινόμησης και ανιχνευτών που υλοποιήθηκε για το σύστημα VGRS. Το υποσύστημα αυτό, όπως είδαμε και παραπάνω, δίνει σταθερά καλύτερα αποτελέσματα από το fine-tuned πολυτροπικό μεγάλο γλωσσικό μοντέλο όρασης που εκπαιδεύτηκε, οπότε είναι καλό εργαλείο για να κάνουμε τους πρώτους ελέγχους, ώστε να σιγουρευτούμε για την ύπαρξη φυλλου στην εικόνα που δόθηκε, και να πάρουμε μια καλή εκτίμηση για το ποιος είναι ο τύπος του φύτου και ποιά είναι η ασθένεια που το επηρεάζει. Αυτός ο κόμβος ενημερώνει τις τιμές **yolo_existence/yolo_existence_conf**, **yolo_plant_type/yolo_plant_type_conf** και **yolo_disease_type** και **yolo_disease_type_conf** του LangGraph state. Αυτή η «πρώτη γνώμη» χρησιμεύει ως οδηγός για τα επόμενα βήματα, αλλά δεν είναι δεσμευτική.

2) Ο κόμβος Tools Ένα από τα βασικά χαρακτηριστικά του LangGraph είναι η δυνατότητα δημιουργίας Tool Nodes, δηλαδή κόμβων που είναι υπεύθυνοι για την εκτέλεση εξωτερικών εργαλείων κατόπιν εντολής του γλωσσικού μοντέλου. Η βιβλιοθήκη **langgraph.prebuilt** παρέχει έτοιμες υλοποιήσεις τέτοιων κόμβων, με πιο χαρακτηριστική την κλάση ToolNode.

3) Κεντρικός κόμβος Assistant. Ο κεντρικός κόμβος Assistant είναι ένα LLM στο οποίο έχουμε συνδέσει τρία εργαλεία. Το LLM είναι υπεύθυνο για την παραγωγή της τελικής ανάλυσης που θα δει ο χρήστης. Για να το πετύχει αυτό, δεν βασίζεται στις εσωτερικές του γνώσεις, αλλά χρησιμοποιεί τα διαθέσιμα εργαλεία για να συλλέξει πληροφορίες και να συνθέσει την απάντησή του βάσει αυτών.

Τα εργαλεία στα οποία έχει πρόσβαση μέσω του ToolNode είναι τα εξής:

1. **consult_vector_database()** (το οποίο υλοποιήθηκε στο VGRS και περιέχει συμπυκνωμένες περιγραφές οπτικών χαρακτηριστικών και τακτικές αντιμετώπισης)
2. **consult_web_search()** για αναζήτηση στο διαδίκτυο (που υλοποιήθηκε στο σύστημα MES)

```
# 2) custom aspect for user guidelines
guideline_aspect = AspectCritic(
    name="guideline_adherence",
    definition=(
        "The answer must strictly follow the user's request regarding the type of management tactics (e.g., natural, chemical, or both). "
        "If no specific preference is stated, the response must include both natural and chemical management tactics. "
        "If the plant is healthy, the response must clearly state this. "
        "Return PASS only if the answer fully complies with these guidelines, with no deviations or omissions."
    )
)
```

Σχήμα 4.24: Ο ορισμός του κανονισμού πάνω στον οποίο υπολογίζεται η μετρική

3. `analyze_image()`, το οποίο χρησιμοποιεί το εκπαιδευμένο πολυτροπικό μοντέλο, που υλοποιήθηκε για το σύστημα MES, για την εξαγωγή εναλλακτικής διάγνωσης της ασθένειας του φύλλου, σε περίπτωση που η σιγουριά του YOLO pipeline είναι χαμηλή.

Το LLM επιλέγει ποιο εργαλείο να καλέσει με βάση: (α) την περιγραφή του κάθε εργαλείου σε φυσική γλώσσα, (β) την ερώτηση του χρήστη και (γ) την προϋπάρχουσα ανάλυση από το YOLO pipeline. Έχει τη δυνατότητα να καλέσει πολλαπλά εργαλεία, ακόμα και περισσότερες από μία φορές το καθένα, μέχρι να καταλήξει στην τελική του απάντηση.

Η μετάβαση από τον κόμβο Assistant στον κόμβο ToolNode εξαρτάται από την έξοδο του LLM. Αν η απάντηση είναι tool call, τότε μεταβαίνουμε στον ToolNode, όπου το αποτέλεσμα της αντίστοιχης συνάρτησης επιστρέφεται στο μοντέλο μαζί με τα προηγούμενα μηνύματα ως είσοδος. Αντίθετα, αν η απάντηση του LLM είναι κείμενο, τότε έχει παραχθεί η τελική απάντηση και η ροή συνεχίζεται στον επόμενο κόμβο, `answer_critique`.

4) Ο κόμβος αξιολόγησης `answer_critique`. Ο κόμβος αυτός ενσωματώθηκε στο σύστημα ως κρίσιμος μηχανισμός ελέγχου ποιότητας, με σκοπό να διασφαλιστεί ότι η τελική απάντηση είναι πλήρης, συνεπής με τις προτιμήσεις του χρήστη και ουσιαστικά χρήσιμη. Η ανάγκη ύπαρξής του προκύπτει από την πολυπλοκότητα του προβλήματος: το κεντρικό LLM καλείται να διαχειριστεί μεγάλο όγκο πληροφορίας από διαφορετικές πηγές —οπτικά μοντέλα, εξωτερικά εργαλεία, βάσεις γνώσης, οδηγίες του χρήστη— κάτι που συχνά οδηγεί σε συγχύσεις, παρερμηνείες ή παραλείψεις. Έτσι, η τελική απάντηση μπορεί να αποκλίνει από το ζητούμενο, ακόμα και όταν τα επιμέρους βήματα έχουν εκτελεστεί σωστά.

Για να ελεγχθεί η ποιότητα της απάντησης, χρησιμοποιείται η βιβλιοθήκη RAGAS και πιο συγκεκριμένα η μετρική `aspect_critique`, η οποία επιτρέπει την αξιολόγηση της συμμόρφωσης με συγκεκριμένες απαιτήσεις. Η βασική οδηγία επί της οποίας βασίζεται ο έλεγχος ορίζεται ρητά παρακάτω:

Η τιμή της μετρικής υπολογίζεται από ένα εξωτερικό LLM, το οποίο αποδίδει βαθμολογία 1, αν η απάντηση συμμορφώνεται με τον κανονισμό, διαφορετικά 0. Αν το σκορ δεν είναι 1, τότε το runtime επιστρέφει στον κόμβο του assistant, ο οποίος καλείται να αναπαράγει την απάντηση με αυξημένη συμμόρφωση προς τις οδηγίες του χρήστη. Η διαδικασία αυτή επαναλαμβάνεται, έως ότου η απάντηση συμμορφώνεται με τις ανάγκες της ερώτησης του χρήστη.

Κατά τη διάρκεια αυτών των επαναλήψεων, το μοντέλο **δεν χρειάζεται** να καλέσει ξανά όλα τα εργαλεία από την αρχή. Αντιθέτως, επαναξιοποιεί το υπάρχον context με πιο στοχευμένο τρόπο, ώστε να παραχθεί μια απάντηση που να είναι ταυτόχρονα ακριβής, συνεπής και προσαρμοσμένη στις ανάγκες του χρήστη.

4.5.2 Παράδειγμα Λειτουργίας του Συστήματος



Σχήμα 4.25: Φύλλα αχλαδιάς με σκωρίαση (Rust Damage)

Παρακάτω παρουσιάζεται ένα παράδειγμα λειτουργίας του συστήματος, το οποίο αναδεικνύει τη συνεργασία όλων των υποσυστημάτων.

Για το συγκεκριμένο παράδειγμα, χρησιμοποιήθηκε ως είσοδος στο Σχήμα 4.25, συνοδευόμενη από την ερώτηση Q: **What fungicides can be used to manage this?**.

Επιλέχθηκε η παραπάνω εικόνα, καθώς προέρχεται από πραγματικές συνθήκες πεδίου και παρατηρήθηκε πως, σε αυτήν, το YOLO pipeline αποτυγχάνει να διαγνώσει σωστά την ασθένεια. Πιο συγκεκριμένα, ο YOLO classifier φαίνεται να μπερδεύεται, αναγνωρίζοντας με πολύ υψηλή βεβαιότητα ότι δεν υπάρχουν φύλλα στην εικόνα. Τα αποτελέσματα της διάγνωσης φαίνονται στον Πίνακα 4.8.

Η περίπτωση αυτή αποτελεί ένα χρήσιμο παράδειγμα για να παρατηρηθεί η αντίδραση του κεντρικού LLM, όταν του παρέχεται μία τέτοια μη έγκυρη διάγνωση. Όπως φαίνεται στο Σχήμα 4.26, το πρώτο βήμα του μοντέλου είναι να συμβουλευτεί το αποτέλεσμα του εξειδικευμένου VLM, μέσω της χρήσης του εργαλείου `analyze_image()`.

Key	Value
leaf_exists	False
plant_type	None
disease_type	None
leaf_existence_conf	0.9999999
plant_type_conf	None
diseases_type_conf	None
context_results	None
disease_analysis	No leaf in image

Πίνακας 4.8: Έξοδος του Yolo Pipeline όπου δεν ανιχνεύεται φύλλο στην εικόνα

```

**GRAPH HISTORY:**
**Human:** What fungicides to use to manage this?
  image path: /home/gfrag/GFRAG_WORKSPACE/FINAL_GRAPH_WORKSPACE/leaf_image.png
**AI calls tool ->** analyze_image
**Tool _analyze_image_ responds ->** Cedar apple rust...
**AI calls tool ->** consult_web_search
**Tool _consult_web_search_ responds ->** In North Carolina, fungicide applications targeting cedar apple rust should begin around tight cluster and continue on 7 to 14 day intervals through first cover (approximately 10 to 14 days after petal...

```

Σχήμα 4.26: Κλήσεις εργαλείων που κάνει το llm για να παράξει την τελική απάντηση

Το VLM καταφέρνει να εντοπίσει σωστά την ασθένεια ως σκωρίαση, αν και αναγνωρίζει λανθασμένα το φυτό ως μηλιά αντί για αχλαδιά. Το σφάλμα αυτό είναι αναμενόμενο, καθώς το VLM δεν έχει εκπαιδευτεί ειδικά σε δεδομένα αχλαδιάς με σκωρίαση. Επιπλέον, η ομοιότητα στα μορφολογικά χαρακτηριστικά μεταξύ μηλιάς και αχλαδιάς δικαιολογεί τη σύγχυση.

Αφού ολοκληρωθεί η αναγνώριση, το LLM προχωρά με την χρήση του εργαλείου αναζήτησης σε διαδικτυακές πηγές, `consult_websearch()`. Συνδυάζοντας τις πληροφορίες που αντλήθηκαν από τα δύο εργαλεία, το σύστημα συνθέτει την τελική απάντηση.

Diagnosis Result

Based on the image analysis, the leaf is affected by **Rust**. To manage this fungal disease, consider using fungicides containing active ingredients like myclobutanil, triflumizole, or mancozeb. Sterol Inhibitor (S.I.) fungicides (FRAC 3) are generally effective. Apply fungicides starting around the tight cluster stage and continue every 7 to 14 days through the first cover stage. Always check the fungicide label for specific instructions and pre-harvest intervals. In addition to chemical control, prune infected branches and maintain good cultural practices like proper watering and fertilization to keep your apple trees healthy. Biofungicides containing *Bacillus subtilis* or plant extracts can also be used as a preventative measure.

Τέλος, η απάντηση αξιολογείται βάσει της μετρικής `guideline_aspect`, όπως έχει οριστεί μέσω της βιβλιοθήκης RAGAS. Η βαθμολογία που λαμβάνει είναι 1.0, καθώς ανταποκρίνεται πλήρως στην ερώτηση του χρήστη, προτείνοντας αποκλειστικά μεθόδους αντιμετώπισης που εκείνος ζήτησε (*fungicides*). Συνεπώς, αυτή θεωρείται η τελική απάντηση του συστήματος.

4.6 Αξιολόγηση του Ενοποιημένου Συστήματος

Για την αξιολόγηση του τελικού συστήματος χρησιμοποιήθηκαν οι μετρικές **context-precision**, **guideline_adherence**, **answer_relevance** και **accuracy** (όλες εκ των οποίων έχουν χρησιμοποιηθεί και σε προηγούμενες αξιολογήσεις).

Στο συγκεκριμένο σύστημα, εφόσον για τη σύνθεση της τελικής απάντησης δεν χρησιμοποιείται μία βάση γνώσης, αλλά το περιεχόμενο τριών εργαλείων, θεωρούμε ως retrieved context τις απαντήσεις που λαμβάνουμε από τα εργαλεία της διανυσματικής βάσης δεδομένων και του web search, πάνω στο οποίο υπολογίζεται η μετρική `context_precision`. Τα αποτελέσματα και των τριών (δηλαδή συμπεριλαμβανομένης της διάγνωσης του πολυτροπικού μεγάλου γλωσσικού μοντέλου) εργαλείων χρησιμοποιούνται για τον υπολογισμό της `guideline_adherence`.

Για τον υπολογισμό της μετρικής `accuracy` χρησιμοποιήθηκε ένα γλωσσικό μοντέλο πολύ χαμηλών παραμέτρων, το Llama 3.2 1B. Γνωρίζοντας την πραγματική κλάση της εικόνας εισόδου, το μοντέλο αποφασίζει αν το κείμενο που παράγεται από το σύστημα αναφέρεται στην πραγματική ασθένεια και αν έχει ανιχνευθεί ο σωστός τύπος φυτού. Η απόφαση που παράγει το LLM είναι 1 για σωστή ανίχνευση, διαφορετικά 0. Οι μετρικές ανίχνευσης ασθένειας και ανίχνευσης τύπου φυτού παρουσιάζονται ξεχωριστά (`plant_accuracy`, `disease_accuracy`). Η αξιολόγηση πραγματοποιήθηκε στο test split του υβριδικού dataset που δημιουργήθηκε για την εκπαίδευση των ταξινομητών (PlantDoc + PlantVillage). Τα αποτελέσματα φαίνονται στον Πίνακα 4.9.

Συνοψίζοντας, το ενοποιημένο σύστημα χάρη στον συνδυασμό ανάκτησης από το vector database και την διαδικτυακή αναζήτηση με στοχευμένα queries, παραγμένα από τον κεντρικό agent, αναχτά πιο σχετικό περιεχόμενο από ότι στα προηγούμενα δύο συστήματα, όπως απο-

Πίνακας 4.9: Αποτελέσματα αξιολόγησης (μέσοι όροι)

Μέσοι όροι μετρικών	Τιμές
plant accuracy	0.94
disease accuracy	0.89
context precision	0.77
guideline adherence	1.00
tools used	1.9

τυπώνεται στη context precision 0.77. Οι απαντήσεις τηρούν πλήρως τους περιορισμούς του χρήστη (guideline adherence 1.00), ενώ η οπτική συνιστώσα διατηρεί πολύ υψηλή απόδοση τόσο στην αναγνώριση του φυτού (94%), όσο και της ασθένειας (89%). Επιπλέον, ο μέσος αριθμός εργαλείων που ενεργοποιούνται ανά ερώτημα παραμένει κάτω από δύο (tools used 1.9), ένδειξη αποδοτικής χρήσης πόρων.

Κεφάλαιο 5

Επίλογος

5.1 Συμπεράσματα

Η παρούσα εργασία είχε ως κεντρικό στόχο την αυτόματη διάγνωση ασθενειών σε φύλλα φυτών από εικόνες και την παροχή τεκμηριωμένων προτάσεων αντιμετώπισης. Με βάση τα αποτελέσματα που παρουσιάστηκαν, ο στόχος αυτός επιτεύχθηκε: το σύστημα δέχεται μια εικόνα, αναγνωρίζει αν υπάρχει φύλλο και σε ποιο φυτό ανήκει, εντοπίζει την ασθένεια, και στη συνέχεια παράγει απάντηση που εστιάζει στις ζητούμενες πρακτικές διαχείρισης, περιορίζοντας τις συστάσεις στο είδος παρέμβασης που ζητά ο χρήστης, όπως μετρήθηκε και από την άριστη επίδοση στη μετρική `guideline_adherence` (1.0). Η συνολική συμπεριφορά του συστήματος τεκμηριώνεται τόσο από τις μετρικές ταξινόμησης (`plant accuracy` 0.94, `disease accuracy` 0.89), όσο και από μετρικές ποιοτικής αξιολόγησης παραγόμενης απάντησης.

Καθοριστικό βήμα για αυτή την επίδοση ήταν η ενοποίηση δύο αρχιτεκτονικών που είχαν υλοποιηθεί αρχικά, του (VGRS) και του συστήματος (MES), σε ένα ενιαίο *agent*ic σύστημα, το Vision-Enhanced Analysis Agent. Η ενοποίηση επέτρεψε οι αδυναμίες της κάθε προσέγγισης να αλληλοαντισταθμιστούν. Το πολυτροπικό αυτό σύστημα αξιοποιεί την ακρίβεια των YOLO μοντέλων μαζί με την ικανότητα περιγραφής των ασθενειών του Vision LLM για να πετυχαίνει πάντα τη καλύτερη δυνατή ανίχνευση.

Εξίσου κρίσιμη ήταν η ένταξη κόμβων ελέγχου ποιότητας στην αλυσίδα συλλογισμού. Η χρήση των μετρικών RAGAS—και ειδικότερα του `guideline_aspect`—λειτουργήσε ως δι-κλείδα ασφαλείας, ώστε η τελική απάντηση να συμμορφώνεται με τους ρητούς ή άρρητους περιορισμούς του χρήστη. Επίσης, ο κόμβος κριτικής της απάντησης βοήθησε να μετριαστεί η σύγχυση του μοντέλου, όταν έπρεπε να διαχειριστεί μεγάλο όγκο ανακτημένης πληροφορίας, βελτιώνοντας τη συνέπεια και την πληρότητα της τελικής ανάλυσης.

Τα ευρήματα των αξιολογήσεων δείχνουν μια σαφή εικόνα: το σύστημα είναι ισχυρό στη βασική του αποστολή (αναγνώριση φυτού και ασθένειας) και αξιόπιστο στην εφαρμογή κανόνων απάντησης, ενώ παραμένουν περιθώρια βελτίωσης στην αυστηρή τεκμηρίωση κάθε ισχυρισμού και στη συνάφεια όλων των ανακτημένων στοιχείων προς την αρχική ερώτηση. Οι τιμές `faithfulness` και `answer relevancy`—αν και ικανοποιητικές—υποδεικνύουν ότι η ποιότητα και η χωρική/λογική δομή των πηγών που αναχτώνται από το διαδίκτυο μπορεί να επηρεάσουν

τη διαύγεια της τελικής απάντησης.

Η στόχευση στην ακρίβεια της οπτικής αναγνώρισης, η πειθαρχημένη ανάκτηση γνώσης και ο έλεγχος ποιότητας της παραγόμενης απάντησης συνθέτουν μια ρεαλιστική πορεία μεταφοράς της τεχνολογίας από το εργαστήριο στην πράξη, με σαφή περιθώρια εξέλιξης, αλλά και με αποδεδειγμένη χρησιμότητα στο σήμερα.

5.2 Περιορισμοί

Κατά την εκπόνηση της παρούσας διπλωματικής εργασίας, εντοπίστηκαν αρκετοί πρακτικοί και τεχνικοί περιορισμοί, οι οποίοι επηρέασαν την ποιότητα και την εμβέλεια των αποτελεσμάτων. Η μελέτη επηρεάστηκε από συγκεκριμένους περιορισμούς που διαμόρφωσαν τόσο τον σχεδιασμό, όσο και την ερμηνεία των αποτελεσμάτων. Πρώτον, τα σύνολα δεδομένων που χρησιμοποιήθηκαν (PlantVillage, PlantDoc, PlantVillage VQA) καλύπτουν περιορισμένο εύρος φυτικών ειδών και ασθενειών, ενώ μεγάλο μέρος των εικόνων προέρχεται από ελεγχόμενα περιβάλλοντα και όχι από το πεδίο. Η έλλειψη ικανού όγκου εικόνων πεδίου (με φυσικό φωτισμό, ποικιλία υποβάθρων, διαφορετικά στάδια ανάπτυξης) ενίσχυσε το domain shift και περιόρισε τη γενικευσιμότητα. Επιπλέον, απουσίαζαν επαρκώς σχολιασμένα δεδομένα εντοπισμού των προσβεβλημένων περιοχών (π.χ. bounding boxes) στις κλάσεις των σετ δεδομένων, γεγονός που εμπόδισε την εκπαίδευση τέτοιων ανιχνευτών, ωστόσο αποδείξαμε πως έχουν την καλύτερη δυνατή ακρίβεια.

Η υπολογιστική υποδομή αποτέλεσε πρόσθετο περιορισμό. Τα πειράματα εκτελέστηκαν σε GPU τύπου NVIDIA GeForce RTX 3090 με VRAM 24 GB, καθώς και σε σύστημα με τέσσερις GPU 16 GB έκαστη όπου η παλαιότερη έκδοση CUDA δεν επέτρεπε τη χρήση σύγχρονων βιβλιοθηκών για parallelization (όπως Accelerate). Ως αποτέλεσμα επιλέχθηκαν LLM χαμηλότερων παραμέτρων και εφαρμόστηκε έντονη quantization, περιορίζοντας τόσο την ακρίβεια όσο και το διαθέσιμο context window. Η απουσία συστηματικής συνεργασίας με γεωπόνους/φυτοπαθολόγους στέρησε πολύτιμη επιμέλεια ετικετών και καθοδήγηση για οριοθέτηση δύσκολων περιπτώσεων. Μια τέτοια συνεργασία θα μπορούσε να μειώσει το label noise και να βελτιώσει την αξιοπιστία της ταξινόμησης. Τέλος, δεν πραγματοποιήθηκε in-the-wild testing σε πραγματικές καλλιέργειες. Ως εκ τούτου, πιθανές μεροληψίες (biases) στο σύστημα παραμένουν μη ανιχνευμένες—όπως ευαισθησία σε συγκεκριμένες συνθήκες φωτισμού/κάμερας, σε ποικιλίες ή στάδια ανάπτυξης, σε επίπεδα σοβαρότητας της ασθένειας ή σε ανισοροπίες κλάσεων.

5.3 Μελλοντικές Επεκτάσεις

Η παρούσα εργασία αποτελεί ένα πρώτο βήμα προς την αυτοματοποιημένη αναγνώριση ασθενειών σε φυλλώματα με τη χρήση πολυτροπικών μεγάλων γλωσσικών μοντέλων. Ωστόσο, υπάρχουν αρκετές κατευθύνσεις στις οποίες το σύστημα μπορεί να επεκταθεί στο μέλλον, τόσο για τη βελτίωση της ακρίβειας, όσο και για την ενίσχυση της πρακτικής του χρησιμότητας.

Μια από τις βασικές μελλοντικές στοχεύσεις είναι η δημιουργία νέου συνόλου δεδομένων

από εικόνες που θα ληφθούν σε πραγματικές συνθήκες πεδίου. Αυτό θα επιτρέψει την καλύτερη γενίκευση του μοντέλου σε περιβάλλοντα που διαφέρουν από τις ιδανικές εργαστηριακές συνθήκες. Παράλληλα, θα επιδιωχθεί η κάλυψη περισσότερων τύπων φυτών, εστιάζοντας σε καλλιέργειες που είναι πιο διαδεδομένες στην Ελλάδα, ώστε το σύστημα να ανταποκρίνεται καλύτερα στις ανάγκες της τοπικής αγροτικής παραγωγής.

Επιπλέον, σχεδιάζεται η επέκταση του συστήματος για την αναγνώριση προσβολών από έντομα και παράσιτα, καθώς αυτά συνιστούν εξίσου σημαντικό παράγοντα υποβάθμισης της υγείας των φυτών. Η προσθήκη αυτής της δυνατότητας θα ενισχύσει τη διαγνωστική επάρκεια του συστήματος και θα το καταστήσει πιο ολοκληρωμένο εργαλείο φυτοπροστασίας.

Ένα ακόμη σημαντικό βήμα προς την πρακτική αξιοποίηση είναι η ενσωμάτωση πληροφορίας από την επίσημη ελληνική βάση δεδομένων φυτοπροστατευτικών προϊόντων του Υπουργείου Αγροτικής Ανάπτυξης και Τροφίμων, ώστε οι προτεινόμενες θεραπευτικές αγωγές να αντιστοιχούν σε εγκεκριμένα και διαθέσιμα σκευάσματα στην εγχώρια αγορά [37].

Σε ερευνητικό επίπεδο, θα εξεταστεί η απόδοση του συστήματος με τη χρήση μεγαλύτερων μοντέλων, τα οποία ενδεχομένως να προσφέρουν βελτιωμένη κατανόηση και παραγωγή απαντήσεων. Επιπλέον, η χρήση μεγαλύτερου context window μπορεί να επιτρέψει τη διαχείριση μεγαλύτερου όγκου πληροφορίας και άρα να οδηγήσει σε πιο σωστές και καλά τεκμηριωμένες αναλύσεις.

Τέλος, για την ενίσχυση της επιστημονικής εγκυρότητας των αποτελεσμάτων, προβλέπεται η συνεργασία με επαγγελματίες γεωπόνους. Μέσω της συνεργασίας αυτής θα πραγματοποιείται συστηματική αξιολόγηση της ακρίβειας των διαγνώσεων και των προτεινόμενων πρακτικών, ώστε το σύστημα να διατηρεί υψηλά πρότυπα αξιοπιστίας και χρηστικότητας.

Bibliography

- [1] Transformers: State-of-the-art natural language processing, 2020. <https://huggingface.co>.
- [2] Vlmevalkit: An open-source toolkit for evaluating large multi-modality models. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 11198–11201, 2024.
- [3] Jean-Baptiste Alayrac et al. Flamingo: a visual language model for few-shot learning. In *Advances in Neural Information Processing Systems*, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/960a172bc7fbf0177ccccbb411a7d800-Paper-Conference.pdf.
- [4] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. Multimodal machine learning: A survey and taxonomy. *arXiv preprint*, 2017. URL <https://arxiv.org/abs/1705.09406>.
- [5] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al. Language models are few-shot learners. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. URL <https://arxiv.org/abs/2005.14165>.
- [6] Cljy. Apple leaf detection dataset, Feb 2025. URL https://universe.roboflow.com/cljy/apple_leaf_detection.
- [7] DigitalOcean Community. Intro to optimization in deep learning: Momentum, rmsprop, adam. <https://www.digitalocean.com/community/tutorials/intro-to-optimization-momentum-rmsprop-adam>, 2024. Intro to SGD, Adam, RMSprop and how they are used.
- [8] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 886–893, 2005. URL <https://dl.acm.org/doi/10.1109/CVPR.2005.177>.
- [9] Google DeepMind. Gemma 3 4b — model card, 2025. URL <https://huggingface.co/google/gemma-3-4b-it>. Includes benchmark results for the 4B variant.

- [10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, 2019. URL <https://arxiv.org/abs/1810.04805>.
- [11] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2021. URL <https://arxiv.org/abs/2010.11929>.
- [12] Omar Elharrouss, Yasir Mahmood, Yassine Bechqito, Mohamed Serhani, Elarbi Baidi, Jamal Riffi, and H. Tairi. Loss functions in deep learning: A comprehensive review, 04 2025. URL https://www.researchgate.net/publication/390569686_Loss_Functions_in_Deep_Learning_A_Comprehensive_Review.
- [13] Gemini Team. Gemini 1.5: Unlocking multimodal understanding across longer contexts. *arXiv preprint*, 2024. URL <https://arxiv.org/abs/2403.05530>.
- [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. doi: 10.1109/CVPR.2016.90. URL https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf.
- [15] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735. URL <https://direct.mit.edu/neco/article/9/8/1735/6109/Long-Short-Term-Memory>.
- [16] Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, and Ting Liu. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *arXiv preprint*, 2023. URL <https://arxiv.org/abs/2311.05232>.
- [17] Hugging Face. Transformers — tokenizer (main classes). https://huggingface.co/docs/transformers/main/en/main_classes/tokenizer, 2025.
- [18] Hugging Face. Tokenizers — the tokenization pipeline (encoding/decoding). <https://huggingface.co/docs/tokenizers/en/pipeline>, 2025.
- [19] David. P. Hughes and Marcel Salathe. An open access repository of images on plant health to enable the development of mobile disease diagnostics, 2016. URL <https://arxiv.org/abs/1511.08060>.
- [20] IBM. What is machine learning. <https://www.ibm.com/think/topics/machine-learning>, 2024.

-
- [21] Glenn Jocher and Jing Qiu. Ultralytics yolo11, 2024. URL <https://github.com/ultralytics/ultralytics>.
- [22] Daniel Jurafsky and James H. Martin. *Speech and Language Processing*. 3 edition, 2025. URL <https://web.stanford.edu/~jurafsky/slp3/>.
- [23] Kaggle. Kaggle: Your machine learning and data science community, 2023. <https://www.kaggle.com>.
- [24] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25 (NeurIPS 2012)*, 2012. URL <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks>.
- [25] Taku Kudo. Subword regularization: Improving neural network translation models with multiple subword candidates. In *Proceedings of ACL 2018*, pages 66–75, 2018. URL <https://aclanthology.org/P18-1007/>.
- [26] Taku Kudo and John Richardson. Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In *Proceedings of EMNLP 2018: System Demonstrations*, pages 66–71, 2018. URL <https://aclanthology.org/D18-2012/>.
- [27] Taku Kudo and John Richardson. Sentencepiece: Unsupervised text tokenizer and detokenizer. <https://github.com/google/sentencepiece>, 2025.
- [28] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. URL <https://proceedings.neurips.cc/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf>.
- [29] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. *arXiv preprint arXiv:2301.12597*, 2023. URL <https://arxiv.org/abs/2301.12597>.
- [30] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft coco: Common objects in context, 2015. URL <https://arxiv.org/abs/1405.0312>.
- [31] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. URL <https://link.springer.com/article/10.1023/B:VISI.0000029664.99615.94>.

- [32] Machine Learning Mastery. A gentle introduction to pooling layers for convolutional neural networks, 2019. URL <https://machinelearningmastery.com/pooling-layers-for-convolutional-neural-networks/>.
- [33] Sharada P. Mohanty, David P. Hughes, and Marcel Salathé. Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7:1419, 2016. URL <https://www.frontiersin.org/articles/10.3389/fpls.2016.01419>.
- [34] E. Moupojou, A. Tagne, F. Retraint, A. Tadonkemwa, Wilfried D., and H. Tapamo. Fieldplant: A dataset of field plant images for plant disease detection and classification with deep learning. *IEEE Access*, 11:35398–35410, 2023.
- [35] Marriam Nawaz, Ali Javed, Momina Masood, and Saleh Albahli. A novel deep learning method for detection and classification of plant diseases. *Complex Intelligent Systems*, 8, 09 2021.
- [36] Umair Nawaz, Muhammad Awais, Hanan Gani, Muzammal Naseer, Fahad Khan, Salman Khan, and Rao Muhammad Anwer. Agriclip: Adapting clip for agriculture and livestock via domain-specialized cross-model alignment. arXiv preprint, 2024. URL <https://arxiv.org/abs/2410.01407>.
- [37] Greek Ministry of Rural Development and Food. Database of plant protection products. <https://1click.minagric.gr/oneClickUI/frmFytoPro.zul>, 2025.
- [38] Ollama. Ollama: Run llms locally, 2023. <https://ollama.com>.
- [39] OpenAI. GPT-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023. doi: 10.48550/arXiv.2303.08774. URL <https://arxiv.org/abs/2303.08774>.
- [40] İlker Paçal, Pınar Poyraz, and Mehmet Ceylan. A systematic review of deep learning techniques for plant diseases. *Artificial Intelligence Review*, 2024. doi: 10.1007/s10462-024-10944-7. URL <https://link.springer.com/article/10.1007/s10462-024-10944-7>.
- [41] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, 2016. doi: 10.1109/CVPR.2016.91. URL https://openaccess.thecvf.com/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html.
- [42] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems 28 (NeurIPS 2015)*, 2015. URL <https://proceedings.neurips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf>.

- [43] Syed Nazmus Sakib, Nafiul Haque, Mohammad Zabed Hossain, and Shifat E. Arman. Plantvillagevqa: A visual question answering dataset for benchmarking vision-language models in plant science, 2025. URL <https://arxiv.org/abs/2508.17117>.
- [44] Imran H. Sarker. Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2(6):420, 2021. doi: 10.1007/s42979-021-00815-1. URL <https://doi.org/10.1007/s42979-021-00815-1>.
- [45] Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1715–1725, 2016. URL <https://aclanthology.org/P16-1162/>.
- [46] Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. Self-attention with relative position representations. In *Proceedings of NAACL-HLT*, pages 464–468, 2018. doi: 10.18653/v1/N18-2074. URL <https://aclanthology.org/N18-2074/>.
- [47] Muhammad Shoaib, Babar Shah, Shaker El-Sappagh, Akhtar Ali, Asad Ullah, Fayadh Alenezi, Tsanko Gechev, Tariq Hussain, and Farman Ali. An advanced deep learning models-based plant disease detection: A review of recent research. *Frontiers in Plant Science*, 14:1158933, 2023. doi: 10.3389/fpls.2023.1158933.
- [48] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*, 2015. URL <https://arxiv.org/abs/1409.1556>.
- [49] Davinder Singh, Naman Jain, Pranjali Jain, Pratik Kayal, Sudhakar Kumawat, and Nipun Batra. Plantdoc: A dataset for visual plant disease detection. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, pages 249–253, 2020.
- [50] Davinder Singh, Naman Jain, Pranjali Jain, Pratik Kayal, Sudhakar Kumawat, and Nipun Batra. Plantdoc: A dataset for visual plant disease detection. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, CoDS COMAD 2020, page 249–253. ACM, January 2020. URL <http://dx.doi.org/10.1145/3371158.3371196>.
- [51] CS231n Stanford University. Convolutional neural networks: Course notes. <https://cs231n.github.io/convolutional-networks/>. Describes convolutional, pooling, fully-connected layers in CNN architectures.
- [52] Jianlin Su, Yu Lu, Shengfeng Pan, Ahmed Murtadha, Bo Wen, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. *arXiv preprint arXiv:2104.09864*, 2021. doi: 10.48550/arXiv.2104.09864. URL <https://arxiv.org/abs/2104.09864>.
- [53] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going

- deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015. doi: 10.1109/CVPR.2015.7298594. URL https://openaccess.thecvf.com/content_cvpr_2015/html/Szegedy_Going_Deeper_With_2015_CVPR_paper.html.
- [54] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, volume 97 of *Proceedings of Machine Learning Research*, pages 6105–6114. PMLR, 2019. URL <https://proceedings.mlr.press/v97/tan19a.html>.
- [55] OpenCompass Team. Openvlm leaderboard. https://huggingface.co/spaces/opencompass/open_vlm_leaderboard, 2024.
- [56] Nandini Thakur, Meenakshi Kaur, Munish Kumar, Mamta Juneja, Jatinder Saini, Tansir Masood, et al. Trends in vision-based machine learning techniques for plant disease identification: A systematic review. *Expert Systems with Applications*, 206: 117734, 2022. doi: 10.1016/j.eswa.2022.117734.
- [57] Maria Tsimpoukelli, Jacob Menick, Serkan Cabi, S. M. Ali Eslami, Oriol Vinyals, and Felix Hill. Multimodal few-shot learning with frozen language models. In *Advances in Neural Information Processing Systems (NeurIPS 2021)*, 2021. URL <https://arxiv.org/abs/2106.13884>.
- [58] Ultralytics. Tensorboard integration. <https://docs.ultralytics.com/integrations/tensorboard/>, 2024.
- [59] Ultralytics. Trainer reference (BaseTrainer). <https://docs.ultralytics.com/reference/engine/trainer/>, 2024.
- [60] Ultralytics. Ultralytics yolo11 — models, tasks and performance. <https://docs.ultralytics.com/models/yolo11/>, 2024.
- [61] Ultralytics. Configuration guide (project, name, etc.). <https://docs.ultralytics.com/usage/cfg/>, 2025.
- [62] Ultralytics. Model training with ultralytics yolo. <https://docs.ultralytics.com/modes/train/>, 2025. Accessed: 2025-09-15.
- [63] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. doi: 10.48550/arXiv.1706.03762. URL <https://arxiv.org/abs/1706.03762>.
- [64] Analytics Vidhya. Optimizers in deep learning: A detailed guide. <https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-on-deep-learning-optimizers/>, 2021.

- [65] Tianqi Wei, Zhi Chen, and Xin Yu. Benchmarking in-the-wild multimodal plant disease recognition and a versatile baseline. In *Proceedings of the 32nd ACM International Conference on Multimedia (MM '24)*, 2024.
- [66] Tianqi Wei, Zhi Chen, Xin Yu, Scott Chapman, Paul Melloy, and Zi Huang. Plantseg: A large-scale in-the-wild dataset for plant disease segmentation. *arXiv preprint*, 2024. URL <https://arxiv.org/abs/2409.04038>.
- [67] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, et al. Google’s neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*, 2016. URL <https://arxiv.org/abs/1609.08144>.
- [68] Mingle Xu and Ji-Eun Park. Plant disease recognition datasets in the age of deep learning: Challenges and opportunities. *Frontiers in Plant Science*, 15:1452551, 2024. doi: 10.3389/fpls.2024.1452551.
- [69] Y. Xu, X. Liu, X. Cao, C. Huang, E. Liu, S. Qian, X. Liu, Y. Wu, F. Dong, C. W. Qiu, J. Qiu, K. Hua, W. Su, J. Wu, H. Xu, Y. Han, C. Fu, Z. Yin, M. Liu, R. Roepman, and J. Zhang. Artificial intelligence: A powerful paradigm for scientific research. *Innovation (Cambridge (Mass.))*, 2(4):100179, 2021. URL <https://doi.org/10.1016/j.xinn.2021.100179>.
- [70] Shukang Yin, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, and Enhong Chen. A survey on multimodal large language models. *arXiv preprint arXiv:2306.13549*, 2023. URL <https://arxiv.org/abs/2306.13549>.
- [71] Shukang Yin, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, and Enhong Chen. A survey on multimodal large language models. *arXiv preprint arXiv:2306.13549*, 2023. URL <https://arxiv.org/abs/2306.13549>.
- [72] Kunpeng Zhang, Li Ma, Beibei Cui, Xin Li, Boqiang Zhang, and Na Xie. Visual large language model for wheat disease diagnosis in the wild. *Computers and Electronics in Agriculture*, 227:109587, 2024. doi: 10.1016/j.compag.2024.109587.

