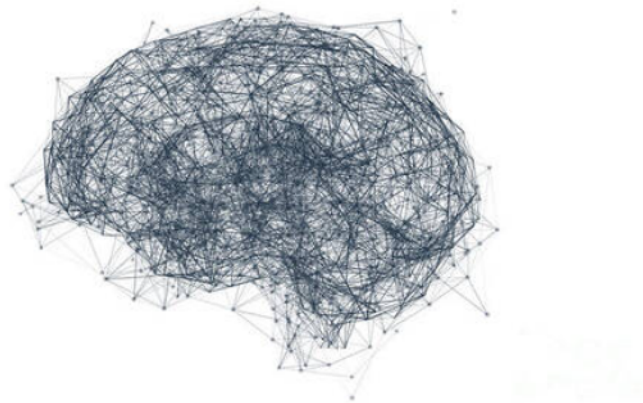


2024/25

Κατζός Γεώργιος:	ics22074
Γηρούση Ευαγγελία:	ics22048
Τίτλος:	Diabetic feet ulcer monitoring
Υπεύθυνος Καθηγητής:	Ευτύχιος Πρωτοπαπαδάκης



Εφαρμοσμένη Πληροφορική

Πανεπιστήμιο Μακεδονίας

Περιεχόμενα

1	Προσδιορισμός και Κατανόηση του Προβλήματος	3
1.1	Περιγραφή του Προβλήματος	3
1.2	Η Σημασία του Προβλήματος	3
1.3	Πρόταση Λύσης	4
1.3.1	Μεθοδολογία	4
1.3.2	Η προσέγγιση της Ανίχνευσης Αντικειμένων (Object Detection) . . .	4
2	Το Σύνολο Δεδομένων (Dataset)	5
2.1	Περιγραφή του Dataset	5
2.2	Περιγραφή Χαρακτηριστικών του Dataset	5
2.3	Λόγοι Επιλογής του Dataset	7
3	Προ-επεξεργασία Δεδομένων	8
3.1	Ανάλυση και Διαχωρισμός Δεδομένων	8
3.2	Μετατροπή Annotations	8
3.3	Κανονικοποίηση Εικόνων και Διαστάσεις	8
3.4	Διαχωρισμός Εισόδων και Ετικετών	8
4	Διερεύνηση Δεδομένων	10
4.1	Η Κατανομή των Bounding Boxes	10
4.2	Έλεγχος της Ποιότητας των Labels	11
4.3	Αποτελέσματα ανά Fold (Cross-Validation)	11
4.4	Μετρικές Αξιολόγησης)	11
4.5	Μέθοδος Αξιολόγησης	11
4.6	Μετρικές Αξιολόγησης	12
4.6.1	Accuracy	12
4.6.2	Precision	12
4.6.3	Recall	12
4.6.4	F1 Score	12
4.6.5	mAP Over Epochs	13
4.6.6	AP (Average Precision)	13
4.6.7	AP50	13
4.6.8	AP75	14
4.6.9	APl (AP large)	14
4.6.10	APm (AP medium)	14
4.6.11	APs (AP small)	14

5	Επιλογή Μοντέλων	15
5.1	YOLOv8n	15
5.2	YOLOv11n	15
5.3	YOLOv11m	15
5.4	Detectron2 R-CNN	16
6	Μεθοδολογία	17
6.1	Ορισμός Υπερπαραμέτρων και Εκπαίδευσης	17
6.2	Εκπαίδευση και Αξιολόγηση	18
6.3	Επιλογή Καλύτερων Μοντέλων	19
6.3.1	Επιλογή Καλύτερου Μοντέλου για το Fold 1	19
6.3.2	Επιλογή Καλύτερου Μοντέλου για το Fold 2	20
6.3.3	Επιλογή Καλύτερου Μοντέλου για το Fold 3	20

Κεφάλαιο 1

Προσδιορισμός και Κατανόηση του Προβλήματος

1.1 Περιγραφή του Προβλήματος

Τα διαβητικά έλκη ποδιών (Diabetic foot ulcer - DFU) είναι μια σοβαρή και ευρέως διαδεδομένη επιπλοκή του διαβήτη, η οποία συντελεί σε μεγάλο βαθμό στη νοσηρότητα των ασθενών. Τα **DFUs** εμφανίζονται εξαιτίας της κακής κυκλοφορίας του αίματος και της νευροπάθειας, με αποτέλεσμα την εμφάνιση χρόνιων τραυμάτων που είναι ιδιαίτερα ευαίσθητα στη μόλυνση. Οι έρευνες δείχνουν ότι περίπου το 15% των διαβητικών ασθενών είναι πιθανό να αναπτύξουν έλκος στο πόδι κατά τη διάρκεια της ζωής τους, ενώ από την άλλη πλευρά το 85% των ακρωτηριασμών των κάτω άκρων που συνδέονται με τον διαβήτη προέρχονται από μη θεραπεύσιμα ή κακώς διαχειριζόμενα έλκη [2].

Ανεξαρτήτως των προόδων στην ιατρική θεραπεία των **DFUs**, η έγκαιρη διάγνωση και η διαρκής παρακολούθηση εξακολουθούν να αποτελούν σημαντικές προκλήσεις. Η παραδοσιακή μέθοδος ανίχνευσης είναι χειροκίνητη και υποκειμενική, με αποτέλεσμα να παρατηρούνται συχνά διαγνωστικές ασυνέπειες. Επιπλέον, η απουσία εξειδικευμένης περίθαλψης σε απομακρυσμένες περιοχές έχει ως αποτέλεσμα την καθυστέρηση της θεραπείας, αυξάνοντας τον κίνδυνο επιπλοκών. Ως αποτέλεσμα, είναι επιτακτική η ανάγκη για αυτοματοποιημένες, ακριβείς και επεκτάσιμες λύσεις για την πρόληψη, τη βελτίωση της ανίχνευσης και της διαχείρισης των **DFU**.

1.2 Η Σημασία του Προβλήματος

Τα **DFUs** αποτελούν πρόβλημα για τα συστήματα υγειονομικής περίθαλψης λόγω των υψηλών ποσοστών υποτροπής που εμφανίζουν, της παρατεταμένης διάρκειας της θεραπείας τους και των επιπλοκών που σχετίζονται άμεσα με αυτά. Οι παραδοσιακές μέθοδοι αξιολόγησης και ταξινόμησης των **DFUs**, όπως το σύστημα ταξινόμησης του Wagner, στηρίζονται στην οπτική εξέταση και την κλινική κατάσταση του ασθενούς, γεγονός που ενδέχεται να έχει ως αποτέλεσμα ασυνέπεια και καθυστερήσεις στη θεραπεία [3].

Επιπλέον, με την παρατεταμένη παραμονή στο νοσοκομείο λόγω των **DFUs** και τις δαπανηρές ιατρικές πράξεις η οικονομική επιβάρυνση των παρόχων υγειονομικής περίθαλψης και των ασθενών αυξάνεται σημαντικά. Έρευνες επισημαίνουν ότι η έγκαιρη ανίχνευση και η έγκαιρη

παρέμβαση δύναται να μειώσει αισθητά τα ποσοστά ακρωτηριασμών και το κόστος θεραπείας [5]. Ωστόσο, εξακολουθεί να υπάρχει έλλειψη τυποποιημένων, αυτοματοποιημένων και κλιμακούμενων μεθόδων για την ανίχνευση και την παρακολούθηση των ελκών.

Δεδομένων αυτών των προκλήσεων, η ενσωμάτωση της τεχνητής νοημοσύνης (AI) και της βαθιάς μάθησης στην ανίχνευση των DFUs αποτελεί μια ευαίσιμη λύση. Τα διαγνωστικά εργαλεία με βάση την Τεχνητή Νοημοσύνη μπορούν να παρέχουν συνεπή και αποτελεσματική ανάλυση, μειώνοντας την εξάρτηση από τον ανθρώπινο παράγοντα και επιτρέπουν την παρακολούθηση εξ αποστάσεως, ιδίως για ασθενείς σε υποβαθμισμένες περιοχές.

1.3 Πρόταση Λύσης

Για να αντιμετωπιστούν οι περιορισμοί των παραδοσιακών μεθόδων ανίχνευσης των **DFUs**, η παρούσα μελέτη προτείνει μια μέθοδο αυτοματοποιημένης ανίχνευσης και παρακολούθησης τους με τη χρήση της τεχνητής νοημοσύνης και πιο συγκεκριμένα μοντέλων βαθιάς μάθησης (Deep Learning). Οι τεχνικές ανίχνευσης αντικειμένων, επιτρέπουν τον ακριβή εντοπισμό των ελκών εντός των εικόνων, επιτρέποντας την έγκαιρη διάγνωση και την σταδιακή παρακολούθηση με την πάροδο του χρόνου.

1.3.1 Μεθοδολογία

Το προτεινόμενο σύστημα χρησιμοποιεί ένα σύνολο δεδομένων από εικόνες που εντοπίζονται **DFUs**. Το σύνολο δεδομένων θα περιέχει εικόνες με επισημασμένες περιοχές έλκους, επιτρέποντας στο μοντέλο να μάθει πρότυπα για μελλοντική αποτελεσματική ανίχνευση. Το εκπαιδευμένο μοντέλο θα εφαρμοστεί για τον εντοπισμό περιοχών έλκους και την παρακολούθηση της εξέλιξης με την πάροδο του χρόνου. Αυτό θα επιτρέψει τη συνεχή παρακολούθηση της επούλωσης των τραυμάτων και θα βοηθήσει τους κλινικούς γιατρούς στη λήψη αποφάσεων θεραπείας βάσει δεδομένων.

1.3.2 Η προσέγγιση της Ανίχνευσης Αντικειμένων (Object Detection)

Σε αντίθεση με τα παραδοσιακά μοντέλα ταξινόμησης εικόνας που δύναται να προσδιορίσουν μόνο αν υπάρχει έλκος σε μια εικόνα, το συγκεκριμένο προσφέρει μια πιο ολοκληρωμένη λύση, καθώς όχι μόνο ανιχνεύει τα έλκη αλλά και τα εντοπίζει εντός των εικόνων. Με την χρήση των οριοθετημένων πλαισίων, τα μοντέλα ανίχνευσης αντικειμένων παρέχουν ακριβείς πληροφορίες αναφορικά με τις θέσεις των **DFUs**, γεγονός που τα καθιστά ιδιαίτερα πολύτιμα για την παρακολούθηση της εξέλιξης της νόσου. Επιπλέον, ο εντοπισμός αντικειμένων επιτρέπει την αυτοματοποιημένη μέτρηση του μεγέθους και του σχήματος του έλκους, η οποία είναι απαραίτητη για την χρονική εκτίμηση της επούλωσης του τραύματος.

Κεφάλαιο 2

Το Σύνολο Δεδομένων (Dataset)

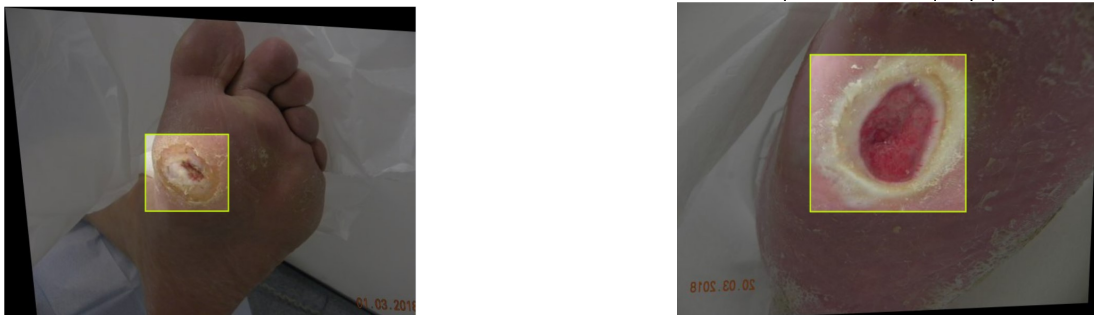
2.1 Περιγραφή του Dataset

Το **dataset** που χρησιμοποιείται σε αυτό το έργο επικεντρώνεται στην ανίχνευση και τον εντοπισμό **DFUs**, μέσω της χρήσης του **object detection**. Αποκτήθηκε από το Roboflow Universe [12], μια πλατφόρμα που χρησιμοποιείται ευρέως για τη δημιουργία και τη διαχείριση dataset υπολογιστικής όρασης (computer vision). Το σύνολο δεδομένων, με τίτλο **Diabetic Foot Ulcer - v6**, περιέχει 10.163 εικόνες, με **DFUs** και έχει υποστεί **data augmentation** για την ενίσχυση της γενικότητας και της αξιοπιστίας του μοντέλου ανίχνευσης, μέσω της χρήσης διαφορετικών σεναρίων.

Κάθε εικόνα απεικονίζει ένα ανθρώπινο πόδι και συνοδεύεται από ένα **label** σχετικό με την παρουσία του διαβητικού έλκους, η οποία γίνεται αντιληπτή μέσω ενός οριοθετημένου πλαισίου (**bounding box**). Οι σχολιασμοί παρέχονται σε πολλαπλές μορφές (YOLO, COCO, Pascal VOC), επιτρέποντας την υλοποίηση διαφορετικών μοντέλων μηχανικής μάθησης.

2.2 Περιγραφή Χαρακτηριστικών του Dataset

Στην παρούσα ενότητα παρουσιάζονται τα χαρακτηριστικά του **dataset**. Τα δεδομένα αφορούν εικόνες με σχολιασμούς (**annotations**) (Σχήμα 2.1) για την ανίχνευση των **DFUs**, και είναι δυνατό να χρησιμοποιηθούν σε προβλήματα υπολογιστικής όρασης (**computer vision**). Παρακάτω στον Πίνακα 2.1 αναλύονται τα δεδομένα μορφοποιημένα σύμφωνα με το πρότυπο (COCO format). Ο αριθμός των εικόνων ισούται με 10.163, οι διαστάσεις της κάθε εικόνας είναι 640×640 pixels και τα χαρακτηριστικά (**features**) της κάθε εγγραφής είναι 8.



Σχήμα 2.1: Εικόνες από το σύνολο δεδομένων Diabetic Foot Ulcer με σημειωμένα έλκη για χρήση σε εφαρμογές ιατρικής ανάλυσης εικόνας.

Name	Type	Description
Ulcer	Categorical (0 or 1)	Υποδεικνύει αν υπάρχει παρουσία διαβητικού έλκους. Στην COCO μορφή, η ύπαρξη του έλκους αντιστοιχεί στην κατηγορία με id: 0 .
File Name	Categorical (String)	Το όνομα του αρχείου εικόνας (π.χ. 100048.jpg), αποτελεί μοναδικό αναγνωριστικό κάθε εγγραφής.
Image Width	Integer	Το οριζόντιο μέγεθος της εικόνας σε pixels, σταθερά ορισμένο στα 640 pixels.
Image Height	Integer	Το κατακόρυφο μέγεθος της εικόνας σε pixels, σταθερά ορισμένο στα 480 pixels.
Bounding Box X	Float	Η οριζόντια θέση του κέντρου του πλαισίου (bounding box) που περιλαμβάνει το έλκος.
Bounding Box Y	Float	Η κατακόρυφη θέση του κέντρου του πλαισίου (bounding box) που περιλαμβάνει το έλκος.
Bounding Box Width	Float	Το οριζόντιο μήκος του πλαισίου (bounding box) που περιλαμβάνει το έλκος.
Bounding Box Height	Float	Το κατακόρυφο μήκος του πλαισίου (bounding box) που περιλαμβάνει το έλκος.
Split	Categorical (String)	Ο διαχωρισμός των εικόνων σε σύνολο εκπαίδευσης (train), επικύρωσης (val), ή ελέγχου (test).

Πίνακας 2.1: Πίνακας Μεταβλητών και Χαρακτηριστικών

2.3 Λόγοι Επιλογής του Dataset

Το συγκεκριμένο **dataset** επιλέχθηκε για την παρούσα εργασία για μια σειρά από λόγους που σχετίζονται τόσο με την ποιότητα όσο και με τη συνάφειά του με το πρόβλημα της ανίχνευσης ελκών σε ασθενείς με διαβήτη. Πιο συγκεκριμένα, κατέχει:

1. **Ιατρική Σημασία:** Τα **DFUs** αποτελούν σοβαρή επιπλοκή για ασθενείς με σακχαρώδη διαβήτη, καθώς μπορούν να οδηγήσουν σε ακρωτηριασμό αν δεν εντοπιστούν έγκαιρα. Το συγκεκριμένο **dataset** επιτρέπει την εκπαίδευση μοντέλων, τα οποία θα δύναται να συμβάλλουν στην πρόωπη διάγνωση.
2. **Ποιότητα Σχολιασμών:** Οι σχολιασμοί (**annotations**) έχουν πραγματοποιηθεί με ακρίβεια, με τη χρήση αναγνωρισμένων προτύπων (YOLO και COCO).
3. **Augmentation:** Το ενσωματωμένο **augmentation** ενισχύει τη γενίκευση των μοντέλων που θα υλοποιηθούν.
4. **Ανοιχτή Πρόσβαση:** Το **dataset** διατίθεται δημόσια, γεγονός που επιτρέπει την ελεύθερη χρήση, τροποποίηση και δημοσίευση αποτελεσμάτων.
5. **Δομή:** Το **dataset** έχει σαφώς καθορισμένη δομή, με ελάχιστα έως καθόλου κενά δεδομένα (**null**), καθιστώντας το εύκολο να αναλυθεί και να επεκταθεί μελλοντικά.

Κεφάλαιο 3

Προ-επεξεργασία Δεδομένων

Η διαδικασία προ-επεξεργασίας των δεδομένων αποτελεί απαραίτητο στάδιο στην ανάπτυξη κάθε αλγορίθμου μηχανικής μάθησης και ειδικότερα σε εφαρμογές **object detection**, όπως για παράδειγμα το YOLOv8 [11] που χρησιμοποιήθηκε στην παρούσα εργασία. Είναι σημαντικό να αναφερθεί ότι η ακρίβεια και η συνέπεια των δεδομένων διαδραματίζει πρωταγωνιστικό ρόλο στην τελική απόδοση του μοντέλου.

3.1 Ανάλυση και Διαχωρισμός Δεδομένων

Αρχικά, το **dataset** ήταν διαχωρισμένο σε **training** και **validation** σύνολα. Καθώς απαιτούνταν και **test set**, ενοποιήθηκαν όλα τα δεδομένα και πραγματοποιήθηκε νέα διαίρεση. Η διαδικασία επαναλαμβάνεται για κάθε **fold** στο πλαίσιο της διασταυρούμενης επικύρωσης (**cross-validation**).

3.2 Μετατροπή Annotations

Τα δεδομένα λήφθηκαν από την ιστοσελίδα του **Roboflow** στο κατάλληλο **format** τύπου **COCO**. Το **dataset** υπό κανονικές συνθήκες θα περιείχε μία μόνο κατηγορία αντικειμένου, τα έλκη (**ulcers**). Κατά τη λήψη, υπήρχε επιπλέον μια κατηγορία '0' που δεν αντιστοιχούσε σε καμία εικόνα, οπότε και αφαιρέθηκε.

3.3 Κανονικοποίηση Εικόνων και Διαστάσεις

Όλες οι εικόνες αναπροσαρμόστηκαν στο σταθερό μέγεθος των 640×640 pixels με χρήση **padding** και **resize**, διαδικασία που υποστηρίζεται εσωτερικά από το **YOLOv8** κατά το **training**. Η τυποποίηση αυτή διευκολύνει τη σύγκριση αποτελεσμάτων μεταξύ διαφορετικών μοντέλων και διασφαλίζει ότι η πληροφορία αξιοποιείται ισότιμα.

3.4 Διαχωρισμός Εισόδων και Ετικετών

Καθώς πρόκειται για πρόβλημα εντοπισμού αντικειμένων (**object detection**), κάθε δείγμα (εικόνα) συνδέεται με μία ή περισσότερες ετικέτες (**bounding boxes** με κλάσεις). Η είσοδος προς το μοντέλο είναι μία εικόνα, ενώ η έξοδος είναι οι συντεταγμένες και τα **bounding boxes** των αντικειμένων. Τα δεδομένα οργανώνονται αυτόματα μέσω του **YAML**, το οποίο

είναι αρχείο ρύθμισης που καθορίζει τη διαδρομή είτε για **train**, είτε για **validation** είτε για **test**.

Κεφάλαιο 4

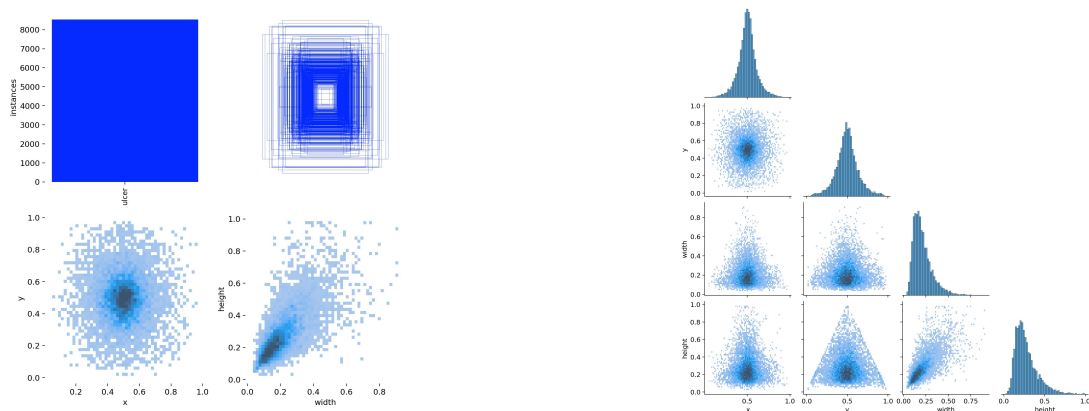
Διερεύνηση Δεδομένων

Η διερεύνηση δεδομένων αποτελεί το επόμενο στάδιο αυτής της μελέτης και έχει ως σκοπό την αποτύπωση της ποιότητας, της κατανομής και των συσχετίσεων στα δεδομένα. Στο πλαίσιο της παρούσας εργασίας, όπου χρησιμοποιούνται εικόνες και σημειώσεις **bounding boxes** για ανίχνευση διαβητικών ελκών, η διερεύνηση περιλαμβάνει:

- Ανάλυση της πληρότητας και της ορθότητας των annotations.
- Κατανομή της ύπαρξης DFU ανά εικόνα.
- Σύγκριση του **ground truth** με τα **predicted bounding boxes**.
- Στατιστικά ανά **fold cross-validation**.

4.1 Η Κατανομή των Bounding Boxes

Κατά τη διερεύνηση του **dataset** εντοπίστηκαν συνολικά 10.163 εικόνες, στις οποίες τα DFUs έχουν επισημανθεί με **bounding boxes**. Η κλάση ulcer είναι categorical (class_id: 0). Δεν εντοπίστηκαν εικόνες χωρίς αντίστοιχο **label**, γεγονός που επιβεβαιώνει τη συνέπεια του **dataset**. Η κατανομή των ετικετών ήταν ισορροπημένη μεταξύ **training** και **validation set** (Σχήμα 4.1).



Σχήμα 4.1: Τα δεδομένα για τα bounding boxes για όλο το dataset.

Metric	Description
Precision (Ακρίβεια)	Εντοπισμοί που ήταν σωστοί (true positives) σε σχέση με όλους τους εντοπισμούς.
Recall (Ανακλητικότητα)	Σωστοί εντοπισμοί σε σχέση με όλα τα πραγματικά έλκη (ground truth objects).
mAP@0.5	Μέση ακρίβεια για IoU > 0.5.
mAP@0.5:0.95	Μέση ακρίβεια για IoU μεταξύ 0.5 και 0.95, με βήμα 0.05.
F1 Score	Ο μέσος όρος μεταξύ precision και recall, δηλαδή ο αρμονικός μέσος τους.

Πίνακας 4.1: Πίνακας Μετρικών Αξιολόγησης

4.2 Έλεγχος της Ποιότητας των Labels

Με τη βοήθεια των συναρτήσεων `load_yolo_labels()` και `yolo_to_bbox()`, πραγματοποιήθηκε έλεγχος των συντεταγμένων των **bounding boxes**, επιβεβαιώνοντας ότι όλα τα **label** ήταν εντός των ορίων των εικόνων (640 × 480). Κάτα την αρχικοποίηση του **dataset**, δεν εντοπίστηκαν ελλιπή ή κατεστραμμένα **annotations**.

4.3 Αποτελέσματα ανά Fold (Cross-Validation)

Κατά τη συγγραφή του κώδικα χρησιμοποιήθηκε **5-Fold Cross-Validation** ώστε να εξεταστεί η απόδοση του μοντέλου σε πολλαπλές υποδιαίρεσεις του **dataset**. Για το κάθε fold του κάθε μοντέλου που μελετήθηκε στην παρούσα εργασία, καταγράφηκαν οι μετρικές που παρουσιάζονται στον Πίνακα 4.1

4.4 Μετρικές Αξιολόγησης

Για την αξιολόγηση της απόδοσης των μοντέλων, χρησιμοποιήθηκαν διάφορες μετρικές που καλύπτουν διαφορετικές πτυχές της απόδοσης, παρέχοντας μια ολοκληρωμένη εικόνα των αποτελεσμάτων. Στο κεφάλαιο αυτό θα περιγραφούν οι μετρικές που χρησιμοποιήθηκαν στην παρούσα εργασία.

4.5 Μέθοδος Αξιολόγησης

Για την αξιολόγηση, χρησιμοποιήθηκε **cross-validation** στη μορφή της τεχνικής **k-Fold**, και πιο συγκεκριμένα **5-Fold**, όπως αναφέρθηκε και προηγουμένως. Σύμφωνα με το **5-Fold**,

τα δεδομένα χωρίζονται σε 5 τμήματα, και τα μοντέλα εκπαιδεύονται και αξιολογούνται σε κάθε τμήμα ξεχωριστά. Οι τελικές μετρήσεις προκύπτουν από τον μέσο όρο των αποτελεσμάτων όλων των folds. Η χρήση του **5-Fold cross-validation** επιτρέπει την πλήρη αξιοποίηση των δεδομένων και διασφαλίζει την αξιοπιστία των αποτελεσμάτων, μειώνοντας τον κίνδυνο για overfit και παρέχοντας μια σταθερή εκτίμηση της απόδοσης των μοντέλων.

4.6 Μετρικές Αξιολόγησης

4.6.1 Accuracy

Η ακρίβεια (**accuracy**) υπολογίζει το ποσοστό των ορθά ταξινομημένων δειγμάτων (τόσο θετικών όσο και αρνητικών) στο σύνολο των παρατηρήσεων. Υπολογίζεται ως εξής:

$$Accuracy = \frac{TruePositives + TrueNegatives}{TruePositives + TrueNegatives + FalsePositive + FalseNegatives}$$

Η ακρίβεια είναι μια διαισθητική μετρική, ωστόσο μπορεί να είναι λιγότερο χρήσιμη σε περιπτώσεις μη ισορροπημένων δεδομένων, όπου η μία κλάση υπερισχύει. Για το λόγο αυτό, δεν πρέπει να λαμβάνεται υπόψη από μόνη της. Η ακρίβεια υπολογίζεται μέσω του παρακάτω τμήματος κώδικα [14] [15].

4.6.2 Precision

Το **precision** εστιάζει στη θετική κλάση και μετρά το πραγματικό ποσοστό των ορθών θετικών προβλέψεων σε σχέση με το συνολικό αριθμό θετικών προβλέψεων που είναι είτε **true** είτε **false**:

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives}$$

Το **precision** είναι σημαντικό όταν το κόστος των ψευδώς θετικών είναι υψηλό. Και σε αυτή την περίπτωση, δεν πρέπει να λαμβάνεται υπόψη από μόνη της, αυτή η μετρική [14] [15].

4.6.3 Recall

Η μετρική **recall** υπολογίζει το ποσοστό των πραγματικών θετικών δειγμάτων που ανιχνεύθηκαν ορθά από το μοντέλο σε σχέση με το άθροισμα των πραγματικών θετικών και των ψευδών αρνητικών:

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives}$$

Η **recall** είναι ιδιαίτερα σημαντική όταν το κόστος των ψευδώς αρνητικών είναι υψηλό [14] [15].

4.6.4 F1 Score

Στην περίπτωση εντοπισμού αντικειμένων, το **F1 Score** υπολογίζεται ως ο αρμονικός μέσος της ακρίβειας (**precision**) και της ανάκλησης (**recall**), με βάση τις προβλέψεις πλαίσιων (**bounding boxes**). Μία πρόβλεψη θεωρείται ορθή (**true positive**) όταν το **IoU** (Intersection over Union) μεταξύ προβλεπόμενου και πραγματικού πλαισίου είναι μεγαλύτερο από ένα κατώφλι, συνήθως 0.5.

$$F_1Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Το **F1 Score** αποτελεί έναν συνδυασμό των δύο και είναι κατάλληλη μετρική για μη ισορροπημένα δεδομένα, καθώς εξισορροπεί τις επιδόσεις μεταξύ **precision** και **recall** [14] [15] [16] [17].

4.6.5 mAP Over Epochs

Η μέση ακρίβεια εντοπισμού (mean Average Precision, **mAP**) είναι μια βασική μετρική αξιολόγησης μοντέλων ανίχνευσης αντικειμένων. Υπολογίζει, πόσο βέλτιστα το μοντέλο προβλέπει την παρουσία και θέση των αντικειμένων σε σχέση με τα πραγματικά δεδομένα. Ο υπολογισμός της **mAP** βασίζεται στον υπολογισμό της ακρίβειας (**precision**) και της ανάκλησης (**recall**) για διαφορετικά κατώφλια απόφασης. Η ακρίβεια εκφράζει το ποσοστό των σωστών θετικών προβλέψεων επί του συνόλου των θετικών προβλέψεων, ενώ η ανάκληση μετράει το ποσοστό των σωστών θετικών προβλέψεων επί του συνόλου των πραγματικών θετικών.

Η Average Precision (**AP**) υπολογίζεται ως η επιφάνεια κάτω από την καμπύλη ακρίβειας-ανάκλησης (**precision-recall curve**) για ένα συγκεκριμένο όριο Intersection over Union (**IoU**), δηλαδή την ελάχιστη απαιτούμενη επικάλυψη μεταξύ προβλεπόμενου και πραγματικού πλαισίου εντοπισμού.

Το **mAP** προκύπτει ως ο μέσος όρος των **AP** για όλα τα κλάσεις αντικειμένων και, συχνά, για πολλαπλά επίπεδα **IoU**. Για παράδειγμα:

- Το $mAP@0.5$ υπολογίζει την ακρίβεια για **IoU** τουλάχιστον 0.5, επιτρέποντας σχετικά χαλαρή ευθυγράμμιση.
- Το $mAP@0.5 : 0.95$ είναι ο μέσος όρος των **AP** που υπολογίζονται για **IoU** από 0.5 έως 0.95 σε βήματα 0.05, παρέχοντας πιο αυστηρή και ολοκληρωμένη αξιολόγηση.

4.6.6 AP (Average Precision)

Η **AP** είναι η βασική μετρική αξιολόγησης που χρησιμοποιείται στο **Detectron2** και αντιπροσωπεύει τη μέση ακρίβεια του μοντέλου για διαφορετικά κατώφλια **IoU**, από 0.5 έως 0.95 με βήμα 0.05. Αποτελεί έναν συνολικό δείκτη απόδοσης, ενσωματώνοντας τόσο το **precision** όσο και το **recall** σε διάφορα επίπεδα απαιτούμενης επικάλυψης μεταξύ προβλεπόμενων και πραγματικών πλαισίων.

4.6.7 AP50

Η **AP50** μετρά την ακρίβεια των προβλέψεων όταν το ελάχιστο αποδεκτό ποσοστό επικάλυψης μεταξύ πραγματικού και προβλεπόμενου πλαισίου είναι 50%. Πρόκειται για μια πιο επιεική εκδοχή της **AP** και δείχνει πόσο καλά εντοπίζονται τα αντικείμενα, χωρίς απαραίτητα να απαιτείται υψηλή ακρίβεια στο περιγράμματά τους.

4.6.8 AP75

Η **AP75** είναι μια αυστηρότερη παραλλαγή, όπου η ακρίβεια υπολογίζεται μόνο αν το **IoU** ξεπερνά το 75%. Εστιάζει στον ακριβή εντοπισμό και χρησιμοποιείται συχνά για να αξιολογηθεί η λεπτομέρεια στην πρόβλεψη των **bounding boxes**.

4.6.9 AP1 (AP large)

Η **AP1** αξιολογεί την απόδοση του μοντέλου σε μεγάλα αντικείμενα. Εδώ εξετάζεται αν το μοντέλο μπορεί να εντοπίζει σωστά και με ακρίβεια αντικείμενα μεγαλύτερων διαστάσεων, όπου συνήθως υπάρχουν περισσότερα οπτικά στοιχεία για να αξιοποιηθούν.

4.6.10 APm (AP medium)

Η **APm** μετρά την ακρίβεια για αντικείμενα μεσαίου μεγέθους. Είναι κρίσιμη σε πολλές ιατρικές εφαρμογές, όπου οι βλάβες δεν είναι ούτε πολύ μικρές ούτε μεγάλες αλλά απαιτούν ακρίβεια εντοπισμού.

4.6.11 APs (AP small)

Η **APs** εστιάζει στα μικρού μεγέθους αντικείμενα. Πρόκειται για την πιο απαιτητική υποκατηγορία, καθώς η εντόπιση μικρών αντικειμένων είναι συνήθως δύσκολη λόγω περιορισμένων χαρακτηριστικών και χαμηλής ανάλυσης.

Κεφάλαιο 5

Επιλογή Μοντέλων

Για την παρούσα εργασία ανίχνευσης των διαβητικών ελκών ποδιού, επιλέχθηκε η χρήση μοντέλων εντοπισμού αντικειμένων (**object detection**) βασισμένων σε τεχνολογίες **deep learning**. Συγκεκριμένα, δοκιμάστηκαν και συγκρίθηκαν τέσσερα μοντέλα, τα οποία περιγράφονται στις παρακάτω ενότητες.

5.1 YOLOv8n

Το **YOLOv8n** (nano) αποτελεί την πιο απλή παραλλαγή της οικογένειας YOLOv8, παρέχει υψηλή ταχύτητα και έχει χαμηλές απαιτήσεις υπολογιστικών πόρων. Παρότι είναι εξαιρετικά αποδοτικό σε περιβάλλοντα με περιορισμένους πόρους (π.χ. edge devices), διατήρησε ικανοποιητική απόδοση για την ανίχνευση των ελκών. Στην εργασία αυτή αποτέλεσε αφετηρία για πειραματισμό και benchmarking. [11]

5.2 YOLOv11n

Το **YOLOv11n** αποτελεί μία ερευνητική παραλλαγή του YOLO, με βελτιωμένη αρχιτεκτονική. Η εκπαίδευσή του παραγματοποιείται πιο γρήγορα σε σχέση με το **YOLOv8n**. Παρείχε ταχύτερη σύγκλιση και σημείωσε ικανοποιητική απόδοση στις βασικές περιπτώσεις, με ελαφρώς μειωμένο **recall** σε πιο πολύπλοκα σενάρια, για παράδειγμα σε φωτογραφίες με χαμηλή αντίθεση μεταξύ του έλκους και της υπόλοιπης εικόνας. [11]

5.3 YOLOv11m

Το **YOLOv11m** (medium) είναι μεγαλύτερο μοντέλο σε σχέση με τις nano παραλλαγές, που αξιοποιήθηκαν στην παρούσα εργασία, και παρέχει σημαντικά βελτιωμένη ακρίβεια και ανακλητικότητα (**recall**). Η αυξημένη χωρητικότητα του δικτύου του επιτρέπει να εντοπίζει ακόμα και μικρά ή οριακά ορατά έλκη, με υψηλή αξιοπιστία. Αποτελεί το πιο ισχυρό μοντέλο που χρησιμοποιήθηκε στη μελέτη, αν και απαιτεί μεγαλύτερους χρόνους εκπαίδευσης και μεγαλύτερη χρήση της GPU. [11]

5.4 Detectron2 R-CNN

Το **Detectron2 R-CNN** είναι μια σύγχρονη και ευέλικτη υλοποίηση της αρχιτεκτονικής R-CNN, αναπτυγμένη από το Facebook AI Research ως η επόμενη γενιά της δημοφιλούς βιβλιοθήκης Detectron. Υποστηρίζει πολλές παραλλαγές, όπως τις Faster R-CNN, Mask R-CNN, Cascade R-CNN και χρησιμοποιεί Feature Pyramid Networks (FPN) για βελτιωμένη αναγνώριση αντικειμένων σε διαφορετικές κλίμακες. Το **Detectron2** βελτιώνει σημαντικά την αποδοτικότητα μέσω μοντέρνων βελτιώσεων στην αρχιτεκτονική, όπως batch normalization, multi-scale training και εκμεταλλεύεται πλήρως τις δυνατότητες των σύγχρονων GPU για ταχύτατη εκπαίδευση και inference. Παράλληλα, διατηρεί υψηλή ακρίβεια στον εντοπισμό και την ταξινόμηση αντικειμένων, καθιστώντας το ιδανικό για ερευνητικές αλλά και βιομηχανικές εφαρμογές όπου απαιτείται αξιοπιστία και ταχύτητα. Χρησιμοποιείται συχνά ως σημείο αναφοράς για σύγκριση με άλλες σύγχρονες αρχιτεκτονικές, όπως το YOLO, λόγω της ισορροπίας του μεταξύ απόδοσης και υπολογιστικής πολυπλοκότητας. [20]

Κεφάλαιο 6

Μεθοδολογία

Η διαδικασία βελτιστοποίησης υπερ-παραμέτρων (**hyperparameter tuning**) πραγματοποιήθηκε για να εξασφαλιστεί η βέλτιστη απόδοση του κάθε μοντέλου. Για το σκοπό αυτό, εφαρμόστηκε η τεχνική **K-Fold Cross-Validation** με 5 υποσύνολα. Σε κάθε επανάληψη (**fold**), το μοντέλο εκπαιδεύεται χρησιμοποιώντας τα $\frac{4}{5}$ του συνόλου των δεδομένων, ενώ το υπόλοιπο $\frac{1}{5}$ χρησιμοποιείται για την επικύρωση (**validation**). Η διαδικασία αυτή επαναλαμβάνεται πέντε φορές, ώστε κάθε υποσύνολο του **dataset** να χρησιμοποιηθεί ακριβώς μία φορά ως **validation set**. Μετά την ολοκλήρωση της εκπαίδευσης, επιλέγεται το καλύτερο μοντέλο βάσει των επιδόσεών του και αξιολογείται τελικά στο **test set**.

6.1 Ορισμός Υπερπαραμέτρων και Εκπαίδευσης

Οι υπερ-παραμέτροι που επιλέχθηκαν για κάθε μοντέλο καθορίστηκαν ρητά κατά την εκπαίδευση, όπως φαίνεται στο παρακάτω απόσπασμα 6.1 του κώδικα. Οι τιμές αυτές επιλέχθηκαν βάσει βέλτιστων πρακτικών και πειραματισμού για το συγκεκριμένο πρόβλημα.

```
1  model.train(  
2      data=yaml_path,  
3      epochs=60,  
4      imgsz=416,  
5      batch=16,  
6      project=project_dir,  
7      name=f'yolov11_fold{fold_num}',  
8      exist_ok=True,  
9      optimizer='SGD',  
10     lr0=0.01,  
11     patience=15,  
12     amp=True,  
13     hsv_h=0.015, hsv_s=0.7, hsv_v=0.4,  
14     degrees=5,  
15     scale=0.5,  
16     shear=0.1,  
17     perspective=0.0,  
18     flipud=0.0,  
19     fliplr=0.5,  
20     mosaic=0.8,  
21     mixup=0.0,  
22     copy_paste=0.0,  
23     auto_augment=True,
```

Listing 6.1: Training of YOLOv11 model per fold

6.2 Εκπαίδευση και Αξιολόγηση

Σε κάθε επανάληψη του **cross-validation** και για κάθε fold το μοντέλο εκπαιδεύεται και αξιολογείται μετρώντας τις μετρικές που περιγράφηκαν στο Κεφάλαιο 4, όπως αναγράφονται στο παρακάτω τμήμα κώδικα. Τα αποτελέσματα του κάθε fold αποθηκεύονται σε μία λίστα αρχικά και τελικά σε ένα αρχείο **csv**.

```

1  metrics = {
2      'precision': float(results.box.p.mean()),
3      'recall': float(results.box.r.mean()),
4      'map50': float(results.box.map50.mean()),
5      'map': float(results.box.map.mean()),
6      'f1': float(results.box.f1.mean())
7  }
8
9
10 fold_results.append({
11     'Fold': fold_num,
12     'Precision': round(metrics['precision'], 4),
13     'Recall': round(metrics['recall'], 4),
14     'mAP@0.5': round(metrics['map50'], 4),
15     'mAP@0.5:0.95': round(metrics['map'], 4),
16     'F1': round(metrics['f1'], 4)
17 })

```

Έτσι διασφαλίζεται η αναλυτική καταγραφή των αποτελεσμάτων και η δυνατότητα σύγκρισης τους σε μορφή τέτοια, όπως για παράδειγμα ο Πίνακας 6.1.

Fold	Precision	Recall	mAP@0.5	mAP@0.5:0.95	F1
1	0.7756	0.6999	0.7562	0.3363	0.7358
2	0.7711	0.6803	0.7551	0.3364	0.7228
3	0.8004	0.6895	0.7674	0.3352	0.7408
4	0.7920	0.6717	0.7499	0.3346	0.7269
5	0.7829	0.7048	0.7712	0.3403	0.7418
Avg (Fold 3×2)	0.7844	0.6892	0.7600	0.3366	0.7336

Πίνακας 6.1: Αρχείο csv για το μοντέλο YOLOv8n

Στην περίπτωση του **Detectron2 R-CNN**, η διαδικασία υπολογισμού των μετρικών διαφέρει από άλλες υλοποιήσεις. Στο παρακάτω πίνακα (Πίνακα 6.2), παρουσιάζονται αναλυτικά ο υπολογισμός των μετρικών απόδοσης για το συγκεκριμένο μοντέλο, για το Fold 1.

Iteration	AP	AP50	AP75	AP1	APm	APs
624	10.81	36.69	2.60	13.70	5.99	0.00
1249	16.53	51.23	5.75	19.57	10.58	0.00
1874	19.84	58.01	7.76	23.12	12.09	0.00
2499	20.52	59.23	8.29	23.50	13.39	2.57
3749	20.96	60.23	8.53	23.30	14.85	1.29
4374	22.45	62.20	10.55	25.42	15.34	7.72
4999	22.96	63.56	10.47	26.01	15.76	7.72
5624	22.82	63.63	10.34	25.77	15.61	7.72
6250	23.03	63.60	10.73	26.00	15.99	6.44

Πίνακας 6.2: Εξέλιξη των μετρικών αξιολόγησης για το μοντέλο Detectron2 R-CNN ανά iteration

6.3 Επιλογή Καλύτερων Μοντέλων

6.3.1 Επιλογή Καλύτερου Μοντέλου για το Fold 1

Αναλύοντας τα αποτελέσματα για όλα τα μοντέλα (Yolov8n, Yolov11n, Yolov11m, Detectron2) στο Fold 1, προκύπτει ο παρακάτω Πίνακας 6.3 .

Model	mAP50	mAP50-95	Val Box Loss	Val Cls Loss	Precision	Recall
Yolov8n	0.75616	0.3363	1.90664	1.20159	0.77565	0.69992
Yolov11n	0.75954	0.3346	1.92994	1.21736	0.77941	0.69992
Yolov11m	0.76746	0.3402	1.93944	1.22461	0.78162	0.69992
Detectron2	-	-	-	-	Accuracy: 0.9453	False Neg: 0.09526

Πίνακας 6.3: Συνοπτική Σύγκριση των Μοντέλων

Παρατηρείται ότι το **Yolov11m** εμφανίζει το υψηλότερο **mAP50** (0.76746) και **mAP50-95** (0.3402), που αποτελούν τις πιο αξιόπιστες μετρικές για **object detection**, υπερβαίνοντας τόσο το **Yolov8n** όσο και το **Yolov11n**. Στη συνέχεια, το **Yolov11n** έχει ελαφρώς καλύτερο **mAP50** από το **Yolov8n**, αλλά υστερεί έναντι του **Yolov11m**. Συνολικά όλα τα **YOLO** μοντέλα έχουν παρόμοιο **recall** (0.69992), αλλά το **Yolov11m** έχει και το υψηλότερο **precision** (0.78162). Οι απώλειες (**losses**) είναι πολύ κοντά μεταξύ των **YOLO** μοντέλων, χωρίς σημαντική διαφοροποίηση που να αναιρεί την υπεροχή του **Yolov11m** στη βασική μετρική **mAP**. Το **Detectron2** δεν παρέχει **mAP50/mAP50-95**, που αποτελεί το industry standard για **object detection**, οπότε δεν μπορεί να συγκριθεί άμεσα με τα **YOLO** μοντέλα ως προς αυτή τη βασική μετρική. Αν και έχει υψηλό **classification accuracy** (0.9453), αυτό δεν αρκεί για να κριθεί καλύτερο σε **object detection**, όπου η χωρική ακρίβεια (**mAP**) είναι κρίσιμη.

Συμπερασματικά, το **Yolov11m** είναι το καλύτερο μοντέλο για το **Fold 1**, καθώς υπερέχει στις βασικές μετρικές **object detection** (**mAP50**, **mAP50-95**, **precision**).

6.3.2 Επιλογή Καλύτερου Μοντέλου για το Fold 2

Αναλύοντας τα αποτελέσματα για όλα τα μοντέλα (**Yolov8n**, **Yolov11n**, **Yolov11m**, **Detectron2**) στο **Fold 2**, προκύπτει ο παρακάτω Πίνακας 6.4.

Model	mAP50	mAP50-95	Precision	Recall	F1	Accuracy	False Neg	False Pos
Yolov8n	0.75507	0.33642	0.77109	0.68028	–	–	–	–
Yolov11n	0.73618	0.32413	0.75859	0.67134	–	–	–	–
Yolov11m	0.75385	0.33542	0.77558	0.68282	–	–	–	–
Detectron2	–	–	0.773	0.662	0.713	0.946	0.091	0.072

Πίνακας 6.4: Συνοπτική Σύγκριση των Μοντέλων στο Fold 2

Παρατηρείται ότι το **Yolov8n** εμφανίζει το υψηλότερο **mAP50** (0.75507) και **mAP50-95** (0.33642), που αποτελούν τις πιο αξιόπιστες μετρικές για **object detection**, όπως αναφέρθηκε και προηγουμένως, με το **Yolov11m** να ακολουθεί πολύ κοντά (**mAP50** 0.75385, **mAP50-95** 0.33542) και να υπερτερεί ελαφρώς σε **precision** και **recall**. Το **Detectron2** έχει **precision** (0.773) και **recall** (0.662), τιμές παρόμοιες με τα **YOLO** μοντέλα, άρα βασικό καλύτερο υποψήφιο μοντέλο. Η συνολική του ακρίβεια (**accuracy** 0.946) είναι επίσης υψηλή.

Συμπερασματικά, το **Yolov8n** παραμένει το καλύτερο μοντέλο για το **Fold 2**, καθώς υπερέρχει οριακά στις βασικές μετρικές του **object detection** (**mAP50**, **mAP50-95**), με το **Yolov11m** να ακολουθεί πολύ κοντά κυρίως σε **precision** και **recall**.

6.3.3 Επιλογή Καλύτερου Μοντέλου για το Fold 3

Αναλύοντας τα αποτελέσματα για όλα τα μοντέλα (**Yolov8n**, **Yolov11n**, **Yolov11m**, **Detectron2**) στο **Fold 3**, προκύπτει ο παρακάτω Πίνακας 6.5.

Model	mAP50	mAP50-95	Precision	Recall	F1	Accuracy	False Neg	False Pos
Yolov8n	0.76740	0.33519	0.80039	0.68953	–	–	–	–
Yolov11n	0.76503	0.33784	0.77764	0.69915	–	–	–	–
Yolov11m	0.76395	0.33601	0.78212	0.69812	–	–	–	–
Detectron2	–	–	0.818	0.701	0.755	0.955	0.072	0.068

Πίνακας 6.5: Συνοπτική Σύγκριση των Μοντέλων στο Fold 3

Παρατηρείται ότι το **Detectron2** εμφανίζει το υψηλότερο **precision** (0.818), **recall** (0.701) και **f1-score** (0.755), υπερβαίνοντας τα μοντέλα **YOLO** σε όλες τις βασικές μετρικές ταξινόμησης. Παρόλο που δεν παρέχει **mAP** μετρικές, η συνολική του απόδοση είναι ανώτερη σε αυτό το **fold**. Επομένως, το **Detectron2** αποτελεί την πιο αξιόπιστη επιλογή για το **Fold 3**.

Συμπερασματικά, το **Detectron2** είναι το καλύτερο μοντέλο για το **Fold 3**, καθώς υπερέρχει στις βασικές μετρικές **precision**, **recall** και **f1-score**, παρότι δεν υπολογίζει **μΑΠ**.

Bibliography

- [1] G. V. Georgoulas, A. T. Tzallas, D. I. Fotiadis, and M. Tsipouras, *A review of non-invasive sensors and artificial intelligence models for diabetic foot monitoring*, *Frontiers in Physiology*, vol. 13, 2022. DOI: <https://doi.org/10.3389/fphys.2022.924546>.
- [2] S. B. Reddy, A. K. Jadhav, and S. S. Reddy, *Wagner’s Classification as a Tool for Treating Diabetic Foot Ulcers: Our Observations at a Suburban Teaching Hospital*, *Cureus*, vol. 14, no. 2, 2022. DOI: <https://doi.org/10.7759/cureus.22492>.
- [3] S. M. Abolarinwa, *Diabetic foot ulcer classifications: A critical review*, *International Journal of Diabetes and Clinical Research*, vol. 5, no. 2, pp. 1-6, 2018. DOI: <https://doi.org/10.23937/2377-3634/1410085>.
- [4] A. A. Al-Rubeaan, *Diabetic foot ulcers: Classification, risk factors and management*, *Saudi Journal of Biological Sciences*, vol. 26, no. 7, pp. 1403-1411, 2019. DOI: <https://doi.org/10.1016/j.sjbs.2019.06.002>.
- [5] M. K. Wurdemann et al., *Hospital stays and costs of telemedical monitoring versus standard follow-up for diabetic foot ulcer: an open-label randomised controlled study*, *The Lancet Regional Health – Europe*, vol. 3, 2021. DOI: <https://doi.org/10.1016/j.lanepe.2021.100053>.
- [6] M. H. Yap et al., *Artificial intelligence for automated detection of diabetic foot ulcers: A real-world proof-of-concept clinical evaluation*, *Diabetes Research and Clinical Practice*, vol. 165, 2020. DOI: <https://doi.org/10.1016/j.diabres.2020.108252>.
- [7] S. K. Mishra et al., *The role of machine learning in advancing diabetic foot: a review*, *Frontiers in Endocrinology*, vol. 12, 2021. DOI: <https://doi.org/10.3389/fendo.2021.708433>.
- [8] J. W. Su et al., *Updating methods for artificial intelligence-based clinical prediction models: a scoping review*, *BMJ Open*, vol. 11, no. 5, 2021. DOI: <https://doi.org/10.1136/bmjopen-2020-045202>.
- [9] R. Sarmun et al., *Diabetic Foot Ulcer Detection: Combining Deep Learning Models for Improved Localization*, *Cognitive Computation*, vol. 16, no. 6, pp. 1413-1431, 2024. DOI: <https://doi.org/10.1007/s12559-024-10267-3>.
- [10] Roboflow, *Diabetic Foot Ulcer Dataset*, Available at: <https://universe.roboflow.com/my-datasets-g2gqc/diabetic-foot-ulcer/dataset/6>.
- [11] Ultralytics. (2024). YOLOv8 Documentation. Retrieved from <https://docs.ultralytics.com>

- [12] The dataset used for this assignment was retrieved from: <https://universe.roboflow.com/my-datasets-g2gqc/diabetic-foot-ulcer/dataset/6>
- [13] Roboflow. (2023). What is R-CNN? Retrieved from <https://blog.roboflow.com/what-is-r-cnn/>
- [14] Google Developers, *Classification: Accuracy, Precision, Recall*, Machine Learning Crash Course, 2023. Available at: <https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall>
- [15] Balaji G., *Mastering Classification Metrics: A Deep Dive from F1 Score to AUC-ROC*, Medium, 2023. Available at: <https://medium.com/@balaji92/mastering-classification-metrics-a-deep-dive-from-f1-score-to-auc-roc-87e82b0ebfae>
- [16] Deepchecks, *Understanding F1 Score, Accuracy, ROC-AUC & PR-AUC Metrics*, 2023. Available at: <https://www.deepchecks.com/f1-score-accuracy-roc-auc-and-pr-auc-metrics-for-models/>
- [17] Keylabs, *Understanding the F1 Score and AUC-ROC Curve*, October 25, 2024. Available at: <https://keylabs.ai/blog/understanding-the-f1-score-and-auc-roc-curve/>
- [18] Neptune.ai, *F1 Score vs ROC AUC vs Accuracy vs PR AUC: Which Evaluation Metric Should You Choose?*, 2022. Available at: <https://neptune.ai/blog/f1-score-accuracy-roc-auc-pr-auc>
- [19] Kurtis Pykes, *AUC and the ROC Curve in Machine Learning*, DataCamp, September 10, 2024. Available at: <https://www.datacamp.com/tutorial/auc>
- [20] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick, *Detectron2*, Facebook AI Research, October 2019. Available at: <https://github.com/facebookresearch/detectron2>