



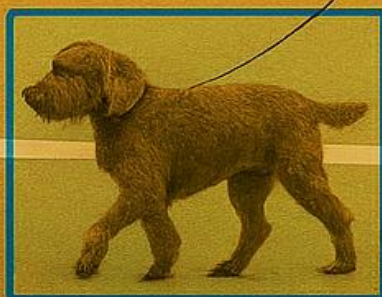
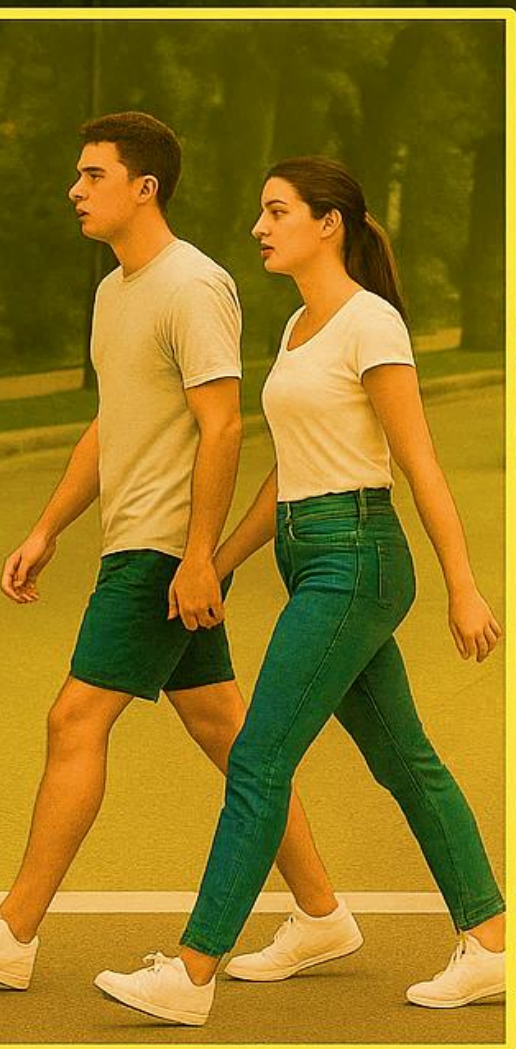
ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

ΤΟΜΕΑΣ ΜΕΤΑΦΟΡΩΝ ΚΑΙ ΣΥΓΚΟΙΝΩΝΙΑΚΗΣ ΥΠΟΔΟΜΗΣ



Διερεύνηση μη συμμορφούμενων διελεύσεων πεζών σε σηματοδοτούμενες
διασταυρώσεις αξιοποιώντας τεχνική υπολογιστικής όρασης



Μυρογιάννη Μελπομένη-Ναταλία
Επιβλέπων: Γεώργιος Γιαννής, καθηγητής Ε.Μ.Π.
Αθήνα, Ιούλιος 2025

ΕΥΧΑΡΙΣΤΙΕΣ

Ολοκληρώνοντας την παρούσα διπλωματική εργασία, θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες προς τον επιβλέποντα καθηγητή, κ. Γεώργιο Γιαννή, για την εμπιστοσύνη που μου έδειξε με την ανάθεση της έρευνας, την ουσιαστική καθοδήγηση σε επιστημονικό και μεθοδολογικό επίπεδο, καθώς και για τις εύστοχες και εποικοδομητικές παρατηρήσεις του καθ' όλη τη διάρκεια της μελέτης.

Ιδιαίτερες ευχαριστίες απευθύνω και στην ερευνήτρια, κα Στυλιανή Ρούσσου, για τη σταθερή υποστήριξη, τις τεκμηριωμένες μεθοδολογικές παρατηρήσεις και την παραγωγική επιστημονική συνεργασία, που ενίσχυσαν σημαντικά τη συγκρότηση και την τελική μορφή του ερευνητικού έργου.

Τέλος, ευχαριστώ την οικογένειά μου και τους στενούς μου φίλους, για την αμέριστη ηθική στήριξη, την υπομονή και την εμπύχωση που μου προσέφεραν καθ' όλη τη διάρκεια αυτής της διαδρομής. Η παρουσία τους υπήρξε πολύτιμη και ουσιαστική.

Αθήνα, Ιούλιος 2025

Μυρογιάννη Μελπομένη - Ναταλία

Διερεύνηση μη συμμορφούμενων διελεύσεων πεζών σε σηματοδοτούμενες διασταυρώσεις αξιοποιώντας τεχνική υπολογιστικής όρασης

Μυρογιάννη Μελπομένη – Ναταλία

Επιβλέπων: Γεώργιος Γιαννής, Καθηγητής Ε.Μ.Π.

Σύνοψη

Η ασφάλεια των πεζών αποτελεί πρόκληση σε πυκνοκατοικημένες περιοχές και, γενικότερα, στα σύγχρονα αστικά περιβάλλοντα, όπως η πόλη της Αθήνας. Η έως τώρα έρευνα επιβεβαιώνει τη σύνδεση μεταξύ συγκρούσεων πεζών–οχημάτων σε σηματοδοτούμενες διασταυρώσεις και των επικίνδυνων ή παράνομων συμπεριφορών που επιδεικνύουν οι πεζοί, αναδεικνύοντας τη θετική συσχέτισή τους με τα ποσοστά ατυχημάτων. Η παρούσα μελέτη εξετάζει μη συμμορφούμενες διασχίσεις πεζών σε μία σηματοδοτούμενη διασταύρωση στο κέντρο της Αθήνας, χρησιμοποιώντας έναν προηγμένο αλγόριθμο ανίχνευσης βασισμένο σε βίντεο. Τα δεδομένα συλλέχθηκαν σε διασταύρωση υψηλής κυκλοφορίας στην πλατεία Ομονοίας, μέσω δύο παράλληλων μεθόδων: επιτόπιας χειροκίνητης παρατήρησης και ενός συστήματος υπολογιστικής όρασης που συνδυάζει τους αλγόριθμους YOLOv8, ResNet-50, ομογραφικό μετασχηματισμό και φιλτράρισμα Kalman. Το σύνολο δεδομένων περιλαμβάνει αναλυτικά στοιχεία για τις συντεταγμένες και τις ταχύτητες πεζών και οχημάτων, τους χρόνους σηματοδότησης και τη μεταβλητή "χρόνος έως σύγκρουση" (time-to-collision). Ο βασικός στόχος της ανάλυσης είναι η σύγκριση της ακρίβειας και της αποτελεσματικότητας των δύο μεθόδων ως προς την ανίχνευση παράνομων διασχίσεων και την κατανόηση του πλαισίου στο οποίο αυτές συμβαίνουν. Η στατιστική ανάλυση πραγματοποιείται με χρήση λογιστικής παλινδρόμησης (logistic regression), τυχαίων δασών (random forest) και διμερούς σημειακής συσχέτισης (point-biserial correlation), με σκοπό την ανάδειξη σημαντικών παραγόντων που προβλέπουν μη συμμορφούμενη συμπεριφορά και την αξιολόγηση της απόδοσης των δύο μεθόδων (χειροκίνητης και αυτοματοποιημένης). Τα ευρήματα της μελέτης συμβάλλουν στην καλύτερη κατανόηση των παραβάσεων πεζών και προσφέρουν χρήσιμες γνώσεις για τη μελλοντική εφαρμογή αυτόματων συστημάτων παρακολούθησης και πολιτικών παρεμβάσεων για ασφαλέστερες διαβάσεις.

Λέξεις-κλειδιά: Παράνομες διελεύσεις πεζών, Υπολογιστική Όραση, YOLO-v8, Λογιστική Παλινδρόμηση, Τυχαίο Δάσος

Investigation of Non-Compliant Pedestrian Crossings at Signalized Intersections Using Computer Vision Techniques

Mirogianni Melpomeni – Natalia

Supervisor: George Yannis, Professor NTUA

Abstract

Pedestrian safety is a challenge in densely populated areas and in general in modern urban environments, like the city of Athens. Research so far confirms the connection between pedestrian-vehicle collisions at signalized intersections to the risky or illegal behaviors pedestrians engage in, and demonstrates the positive correlation with crash rates. This study investigates non-compliant pedestrian crossings at one signalized intersection in the center of Athens, using an advanced video-based detection algorithm. Data are collected from a high-traffic intersection in Omonoia Square, in Athens, using two parallel methods: manual field observation and a computer vision algorithm that integrates YOLOv8, ResNet-50, homography transformation, and Kalman filtering. The dataset includes detailed pedestrian and vehicle coordinates and speed characteristics, signal timing, and time-to-collision metric. Comparing the accuracy and effectiveness of these two methods in detecting illegal crossings and understanding their context is the key purpose of the analysis. Statistical analysis is conducted using logistic regression, random forest classification, and point-biserial correlation to identify significant predictors of non-compliant behaviour and also to compare the effectiveness of the manual field and computer vision algorithm results. The results of this study contribute to the understanding of pedestrian violations and offer valuable insights for future implementation of automated monitoring systems and policy interventions for safer crosswalks.

Keywords: Pedestrian Illegal Crossings, Computer Vision, YOLO-v8, Logistic Regression, Random Forest.

ΠΕΡΙΛΗΨΗ

Η ασφάλεια των πεζών σε πυκνοκατοικημένα αστικά περιβάλλοντα, όπως η Αθήνα, αποτελεί κρίσιμο ζήτημα στον σχεδιασμό των πόλεων. Η μη συμμόρφωση των πεζών προς τη φωτεινή σηματοδότηση στις σηματοδοτούμενες διαβάσεις σχετίζεται άμεσα με αυξημένο κίνδυνο σύγκρουσης. Στην παρούσα εργασία επιχειρείται μια πολυεπίπεδη προσέγγιση της παραβατικής συμπεριφοράς των πεζών, με στόχο τόσο την αξιολόγηση μιας αυτοματοποιημένης μεθόδου παρακολούθησης βασισμένης στην υπολογιστική όραση όσο και την αναζήτηση των παραγόντων που επηρεάζουν τη συμμόρφωση ή μη των πεζών μέσω στατιστικής ανάλυσης.

Τα δεδομένα προήλθαν από βιντεοσκόπηση διάρκειας μίας ώρας σε σηματοδοτούμενη διάβαση στο κέντρο της Αθήνας. Η καταγραφή αναλύθηκε παράλληλα με δύο μεθόδους: (α) **χειροκίνητη καταγραφή** μέσω παρακολούθησης του βιντεοσκοπημένου υλικού, και (β) **αυτοματοποιημένη ανάλυση** μέσω αλγορίθμου σε περιβάλλον MATLAB, ο οποίος ενσωματώνει τεχνικές υπολογιστικής όρασης όπως YOLOv8 για ανίχνευση αντικειμένων, ResNet-50 για εξαγωγή χαρακτηριστικών, δίκτυο Re-Identification (Re-ID) για διατήρηση ταυτοτήτων, Kalman filter για πρόβλεψη κίνησης, αντιστοίχιση με τον Hungarian Algorithm, φιλτράρισμα με Savitzky-Golay Filter, και ανίχνευση της φάσης του φωτεινού σηματοδότη μέσω ROIs (Regions of Interest) στο οπτικό πεδίο.

Στο πρώτο στάδιο της ανάλυσης, αξιολογήθηκε η **ακρίβεια της αυτοματοποιημένης μεθόδου** σε σχέση με τη χειροκίνητη καταγραφή. Η μέση ακρίβεια αναγνώρισης της **φάσης** του φωτεινού σηματοδότη ανήλθε σε **62,77%**, με κάποιες φάσεις να εμφανίζουν πολύ υψηλή απόδοση (έως 90,5%) και άλλες να υπολείπονται λόγω παρεμπόδισης της ορατότητας από μεγάλα οχήματα.

Η ανίχνευση **παράνομων διελεύσεων** προέκυψε ίση με **47.54%** και παρουσίασε σταθερή απόδοση σε αρκετές περιόδους (έως 59,3%), αν και επηρεάστηκε αρνητικά από περιπτώσεις ασαφούς φάσης σηματοδότησης ή διέλευσης σε μεταβατική περίοδο.

Η ανίχνευση των **νόμιμων διελεύσεων** εμφάνισε συνολικά υψηλότερη ακρίβεια, ίση με **64.60%**, με κορυφαία τιμή 92,6% και ποσοστά άνω του 50% στις περισσότερες φάσεις, επιβεβαιώνοντας την αξιοπιστία του συστήματος υπό ομαλές συνθήκες παρατήρησης.

Στο επόμενο στάδιο της εργασίας, εφαρμόστηκαν τεχνικές **στατιστικής ανάλυσης και πρόβλεψης**, με στόχο τον εντοπισμό των παραγόντων που σχετίζονται με τη συμπεριφορά των πεζών ως προς τη χρήση της διάβασης και τη συμμόρφωσή τους με τη φωτεινή σηματοδότηση. Οι δύο εξαρτημένες μεταβλητές που εξετάστηκαν ήταν η παρουσία του πεζού εντός ή εκτός διάβασης και η νομιμότητα ή παρανομία της διέλευσης σε σχέση με το σήμα.

Αρχικά, εφαρμόστηκε **σημειακή διχοτομική συσχέτιση** για τη διερεύνηση γραμμικών σχέσεων μεταξύ των εξαρτημένων μεταβλητών και συνεχών παραμέτρων. Στη συνέχεια, χρησιμοποιήθηκε **λογιστική παλινδρόμηση**, όχι μόνο για την εκτίμηση της πιθανότητας εμφάνισης κάθε μορφής συμπεριφοράς, αλλά και για την κατανόηση της επίδρασης και της αιτιακής σχέσης μεταξύ των υπό εξέταση παραγόντων και των

εξαρτημένων μεταβλητών. Τέλος, η μέθοδος των **τυχαίων δασών** αξιοποιήθηκε για την ενίσχυση της ακρίβειας της πρόβλεψης και την ανάδειξη των σημαντικότερων παραγόντων που σχετίζονται με τη συμπεριφορά των πεζών, ακόμη και όταν οι σχέσεις αυτές είναι μη γραμμικές ή πολύπλοκες.

Η εφαρμογή της point-biserial συσχέτισης ανέδειξε τρεις μεταβλητές με στατιστικά σημαντική διαφοροποιητική ικανότητα ως προς τη θέση του πεζού (εντός ή εκτός διάβασης). Οι παραβάτες εντοπίζονται κυρίως **εντός διαβάσεων**, γεγονός που συνδέεται με τον ορισμό της παραβατικότητας μέσω παραβίασης σηματοδότη. Το εύρημα υποδηλώνει ότι η παραβατική συμπεριφορά δεν είναι αποδιοργανωμένη, αλλά συνειδητή επιλογή μη συμμόρφωσης εντός θεσμικά ρυθμισμένου πλαισίου. Παράλληλα, οι πεζοί εντός διάβασης παρουσιάζουν, κατά μέσο όρο, **υψηλότερη ταχύτητα και μεγαλύτερο χρόνο έως πιθανή σύγκρουση (TTC)**. Η αυξημένη ταχύτητα αντανάκλα είτε στρατηγική ελαχιστοποίησης της έκθεσης των παραβατών είτε σταθερή ροή των νόμιμων πεζών. Το διευρυμένο TTC απορρέει από παύση κυκλοφορίας λόγω σηματοδότη ή αξιοποίησης διαλειμμάτων στην κυκλοφορία, υποδεικνύοντας στοχευμένες στρατηγικές διάσχισης.

Σε δεύτερο επίπεδο, η ανάλυση ανέδειξε διαφοροποιημένες κινητικές συμπεριφορές βάσει νομιμότητας. Οι παραβάτες, μολονότι παραβιάζουν τον σηματοδότη, επιλέγουν τη διάβαση και **κινούνται ταχύτερα**, στοχεύοντας στη μείωση του χρόνου έκθεσης. Αντίθετα, οι νόμιμοι πεζοί κινούνται πιο σταθερά εντός του παρεχόμενου χρονικού πλαισίου. Οι διαφοροποιήσεις αυτές υποδηλώνουν λειτουργικά προσαρμοσμένες στρατηγικές, με έμφαση στη συνειδητή στάθμιση κινδύνου. Ο **αυξημένος TTC** στους παραβάτες αποδίδεται σε επιλεκτική διάσχιση με βάση προσωρινά κενά στην κυκλοφορία, ενώ οι χαμηλότερες τιμές TTC στους νόμιμους πεζούς σχετίζονται με την προσδοκία επιβράδυνσης των οχημάτων. Συνολικά, τα ευρήματα καταδεικνύουν διαφοροποιημένες, αλλά εξίσου στοχευμένες, στρατηγικές αξιολόγησης και αξιοποίησης του χρονικού περιθωρίου ασφαλείας.

Στο πλαίσιο της **λογιστικής παλινδρόμησης**, με εξαρτημένη μεταβλητή την παρουσία του πεζού εντός ή εκτός διάβασης, το μοντέλο παρουσίασε εξαιρετική προβλεπτική ακρίβεια (**AUC = 0,96, ακρίβεια $\approx$ 93%**). Τα αποτελέσματα ανέδειξαν ότι η παρουσία του πεζού στη διάβαση, ακόμη και όταν αυτός παραβιάζει τη σηματοδότηση, αποτελεί συνειδητή και στρατηγική επιλογή. Η θετική επίδραση του πράσινου σηματοδότη επιβεβαίωσε τον ισχυρό ρυθμιστικό ρόλο της φωτεινής σήμανσης, ενώ παράγοντες όπως η αυξημένη ταχύτητα του πεζού και ο διαθέσιμος χρόνος έως πιθανή σύγκρουση υποδήλωσαν προσαρμοστικές συμπεριφορές με στόχο τη μείωση της έκθεσης σε κίνδυνο.

Επιπλέον, η παρουσία **οχήματος** σε κατάσταση προτεραιότητας δεν λειτούργησε αποτρεπτικά, στοιχείο που υποδηλώνει ότι η παραβατικότητα ασκείται με επίγνωση του κινδύνου. Αντίθετα, η αυξημένη ταχύτητα του οχήματος και η φυσική του παρουσία εντός διάβασης συνδέθηκαν με μειωμένη πιθανότητα διέλευσης, επιβεβαιώνοντας την επιρροή των αντιληπτών εμποδίων. Η χωρική απόσταση του οχήματος από τη διάβαση, αντίστροφα, φάνηκε να ενισχύει τη διάθεση των πεζών για διέλευση, υποδηλώνοντας αξιολόγηση ευκαιριών με βάση το οπτικό περιβάλλον.

Στο δεύτερο μοντέλο **λογιστικής παλινδρόμησης**, η ανάλυση επικεντρώθηκε στη νομιμότητα ή παρανομία της διάσχισης, ως προς τη φωτεινή σηματοδότηση. Το

μοντέλο εμφάνισε σχεδόν άριστη προβλεπτική ικανότητα (**AUC = 0,99, ακρίβεια ≈ 99,5%**) και ανέδειξε ένα σύνολο παραγόντων που συνθέτουν ένα προφίλ υπολογισμένου ρίσκου εκ μέρους των πεζών. Η παρουσία πράσινου σήματος για πεζούς, η σταθερότητα της κίνησής τους και η αυξημένη ταχύτητα των οχημάτων συνδέθηκαν ισχυρά με νόμιμη διέλευση, επιβεβαιώνοντας την καθοριστική επίδραση του σηματοδότη και της αντιληπτής κυκλοφορίας.

Αντίθετα, η αυξημένη ταχύτητα του πεζού, η ύπαρξη διαθέσιμου χρόνου έως πιθανή σύγκρουση και η πλαγιομετωπική θέση του οχήματος συνδέθηκαν με αυξημένη πιθανότητα παραβίασης, υποδηλώνοντας ότι οι πεζοί προχωρούν σε παρανομία όταν εκτιμούν πως υφίσταται στιγμιαίο περιθώριο ασφάλειας. Το εύρημα ότι η παρανομία εκδηλώνεται σχεδόν αποκλειστικά εντός της διάβασης ενισχύει περαιτέρω την ερμηνεία της ως συνειδητή πράξη μη συμμόρφωσης, στο πλαίσιο ενός οργανωμένου και προβλέψιμου περιβάλλοντος.

Τα βασικά ευρήματα του μοντέλου συνοψίζονται στον επισυναπτόμενο πίνακα:

1ο Λογιστικό Μοντέλο - Μεταβλητή Στόχος: Παρουσία πεζού εντός/εκτός διάβασης							Pseudo R ²	AIC	BIC	ACCURACY	AUC
Variable	Coef.	Std.Err.	z	P> z	[0.025	0.975]					
const	-2.518	0.069	-36.375	1E-289	-2.654	-2.382	0.643	7084.329	7152.427	0.93	0.96
MinTTC_wins_log	0.961	0.037	26.050	1E-149	0.889	1.033					
vy	0.384	0.024	15.681	2E-55	0.336	0.432					
GroundPlaneCentroid_1_veh	0.077	0.036	2.100	4E-02	0.005	0.148					
magnitude_veh	-0.238	0.028	-8.420	4E-17	-0.293	-0.183					
VehicleStatus	0.371	0.049	7.639	2E-14	0.276	0.466					
PedestrianStatus	-1.969	0.066	-29.778	7E-195	-2.099	-1.840					
ResolvedTrafficLightStatus	1.867	0.046	40.502	0E+00	1.776	1.957					
IsInCrossingArea_veh	-0.117	0.056	-2.098	4E-02	-0.226	-0.008					
2ο Λογιστικό Μοντέλο - Μεταβλητή Στόχος: Νόμιμη/Παρανόμη διέλευση πεζού							0.907	1977.828	2046.51	0.99	0.99
Variable	Coef.	Std.Err.	z	P> z	[0.025	0.975]					
const	5.726	0.145	39.606	0.000	5.443	6.009					
MinTTC_wins_log	-0.213	0.072	-2.965	0.003	-0.353	-0.072					
Confidence	0.842	0.090	9.347	0.000	0.665	1.018					
vy	-0.379	0.044	-8.590	0.000	-0.466	-0.293					
GroundPlaneCentroid_2_veh	-0.445	0.048	-9.214	0.000	-0.540	-0.351					
magnitude_veh	0.798	0.092	8.713	0.000	0.618	0.977					
VehicleStatus	-0.292	0.129	-2.272	0.023	-0.544	-0.040					
ResolvedTrafficLightStatus	4.581	0.160	28.568	0.000	4.267	4.895					
IsInCrossingArea	-3.039	0.091	-33.383	0.000	-3.217	-2.860					

Στο τρίτο στάδιο της ανάλυσης, το μοντέλο **τυχαίου δάσους** αξιοποιήθηκε για την πρόβλεψη της παρουσίας του πεζού εντός ή εκτός διάβασης, αναδεικνύοντας **υψηλή ακρίβεια ταξινόμησης**. Τα αποτελέσματα έδειξαν ότι η συμπεριφορά των πεζών μπορεί να διαφοροποιηθεί με βάση παρατηρήσιμα χαρακτηριστικά, κυρίως ως προς τη χωρική τους θέση, τη δυναμική κίνησης και τη φάση του σηματοδότη. Η θέση του πεζού στον χώρο (σε σχέση με την κάμερα), η ταχύτητα κίνησής του και ο τύπος συμπεριφοράς (π.χ. σταθερή διέλευση ή δισταγμός) αποτέλεσαν τους ισχυρότερους προγνωστικούς δείκτες.

Επιπλέον, η κατάσταση του φωτεινού σηματοδότη συνέβαλε ουσιαστικά στην πρόβλεψη, όπως και μεταβλητές σχετικές με την κυκλοφοριακή ροή, όπως ο διαθέσιμος χρόνος έως πιθανή σύγκρουση. Αντίθετα, χαρακτηριστικά των οχημάτων, όπως η θέση ή η ταχύτητά τους, εμφάνισαν περιορισμένη σημασία, υποδεικνύοντας ότι η τοποθέτηση του πεζού και η δική του στρατηγική κινητικότητα παίζουν καθοριστικό ρόλο στον διαχωρισμό της συμπεριφοράς.

Η ανάλυση ολοκληρώθηκε με την εφαρμογή μοντέλου **τυχαίων δασών** για την πρόβλεψη της παραβατικής συμπεριφοράς πεζών, το οποίο κατέγραψε εξαιρετικά **υψηλή ακρίβεια ταξινόμησης**. Το μοντέλο ανέδειξε την παρουσία του πεζού εντός διάβασης ως τον σημαντικότερο προβλεπτικό παράγοντα για την παρανομία, επιβεβαιώνοντας ότι η παραβατική συμπεριφορά εκδηλώνεται εντός του οριοθετημένου χώρου, ως συνειδητή πράξη μη συμμόρφωσης. Σε συνδυασμό με τη φάση του σηματοδότη και την ταχύτητα κίνησης, το μοντέλο αποτυπώνει τη λήψη απόφασης ως αποτέλεσμα χωρικών και χρονικών εκτιμήσεων.

Παράγοντες όπως ο διαθέσιμος χρόνος έως πιθανή σύγκρουση, η ταχύτητα και η θέση του οχήματος συνέβαλαν επίσης στην πρόβλεψη, ενώ μεταβλητές σχετικές με την αξιοπιστία των ανιχνεύσεων είχαν δευτερεύουσα σημασία. Αντίθετα, η κατάσταση του οχήματος ως νόμιμου ή παράνομου δεν επηρέασε ουσιαστικά την εκδήλωση παραβατικότητας, γεγονός που ενισχύει την υπόθεση ότι η απόφαση του πεζού βασίζεται κυρίως σε αντιληπτικά και δυναμικά στοιχεία της σκηνής, και όχι στην παρατήρηση των κανόνων από τους άλλους χρήστες του οδικού χώρου.

Τα βασικά ευρήματα συνοψίζονται στον επισυναπτόμενο πίνακα:

1ο Μοντέλο Τυχαίου Δάσους - Μεταβλητή Στόχος: Παρουσία πεζού εντός/εκτός διάβασης		ACCURACY	AUC
Variable	Importance		
GroundPlaneCentroid_2	0.234	0.97	0.99
magnitude	0.200		
PedestrianStatus	0.180		
ResolvedTrafficLightStatus	0.106		
MinTTC_wins_log	0.079		
magnitude_veh	0.063		
Confidence	0.045		
GroundPlaneCentroid_1_veh	0.044		
GroundPlaneCentroid_2_veh	0.021		
Confidence_veh	0.017		
IsInCrossingArea_veh	0.010		
VehicleStatus	0.001		
2ο Μοντέλο Τυχαίου Δάσους - Μεταβλητή Στόχος: Νόμιμη / Παράνομη διέλευση πεζού		0.99	0.99
Variable	Importance		
IsInCrossingArea	0.580		
magnitude	0.172		
ResolvedTrafficLightStatus	0.136		
MinTTC_wins_log	0.034		
magnitude_veh	0.028		
Confidence	0.022		
GroundPlaneCentroid_1_veh	0.017		
Confidence_veh	0.011		
VehicleStatus	0.000		

ΠΕΡΙΕΧΟΜΕΝΑ

Κεφάλαιο 1: Εισαγωγή.....	15
1.1 Γενική ανασκόπηση.....	15
1.2 Στόχος διπλωματικής εργασίας.....	17
1.3 Μεθοδολογία διπλωματικής εργασίας	17
1.4 Δομή διπλωματικής εργασίας	19
Κεφάλαιο 2: Βιβλιογραφική Ανασκόπηση	21
2.1 Εισαγωγή.....	21
2.2 Μοτίβα συμπεριφοράς πεζών σε σηματοδοτούμενες διασταυρώσεις	21
2.2.1 Εισαγωγή.....	21
2.2.2 Ταχύτητα πεζών και προσαρμογή της κατά τη διάβαση	22
2.2.3 Αβεβαιότητα και λήψη αποφάσεων κατά τη διάβαση.....	22
2.2.4 Απόσπαση προσοχής και κίνδυνοι κατά τη διάβαση	23
2.3 Παράγοντες επιρροής παραβατικής συμπεριφοράς	23
2.3.1 Διαμόρφωση διαβάσεων και συμμόρφωση πεζών.....	23
2.3.2 Φύλο και συμμόρφωση πεζών	24
2.3.3 Ηλικία και συμμόρφωση πεζών	24
2.3.4 Κοινωνική επιρροή και συμμόρφωση πεζών	25
2.3.5 Κυκλοφοριακή δυναμική και συμμόρφωση πεζών	25
2.4 Αυτόματη διάγνωση θεμάτων ασφαλείας πεζών σε σηματοδοτούμενες διασταυρώσεις	25
2.4.1 Εισαγωγή.....	25
2.4.2 Αυτόματη εξαγωγή και ανάλυση τροχιών πεζών	26
2.4.3 Αυτόματη ανίχνευση πεζών	26
2.4.4 Αυτόματη διάγνωση θεμάτων ασφαλείας πεζών	27
2.4.5 Αυτόματη ανίχνευση-πρόβλεψη παραβατικής συμπεριφοράς πεζών	29
2.4.6 Αυτόματη ανίχνευση-πρόβλεψη συγκρούσεων μεταξύ οχημάτων και πεζών	30
2.4.7 Αυτόματη βελτιστοποίηση της κυκλοφορίας σε σηματοδοτούμενες διασταυρώσεις	31
2.5 Σύνοψη βιβλιογραφίας.....	32
Κεφάλαιο 3: Θεωρητικό υπόβαθρο.....	34
3.1 Εισαγωγή	34
3.2 Βασικές έννοιες στατιστικής	34
3.3 Συσχέτιση και πρόβλεψη: Δύο βασικές προσεγγίσεις στατιστικής ανάλυσης	35
3.4 Συντελεστής συσχέτισης Pearson: Θεωρητική βάση και επεκτάσεις	36

3.5 Μοντέλα παλινδρόμησης: Γραμμική και Λογιστική προσέγγιση στην ανάλυση σχέσεων	37
3.6 Μοντέλο Point-Biserial Correlation.....	37
3.6.1 Εισαγωγή.....	37
3.6.2 Θεωρητικό υπόβαθρο μεθόδου.....	38
3.6.3 Κριτήρια αξιολόγησης και αποδοχής συσχετίσεων	38
3.6.4 Εξήγηση αποτελεσμάτων	41
3.7 Μοντέλο Λογιστικής Παλινδρόμησης	42
3.7.1 Εισαγωγή.....	42
3.7.2 Θεωρητικό υπόβαθρο μεθόδου.....	42
3.7.3 Κριτήρια αξιολόγησης και αποδοχής μοντέλου	44
3.8 Μοντέλο Random Forest	49
3.8.1 Εισαγωγή.....	49
3.8.2 Θεωρητικό υπόβαθρο μεθόδου.....	49
3.8.3 Κριτήρια αξιολόγησης και αποδοχής μοντέλου	51
3.8.4 Εξήγηση αποτελεσμάτων	54
Κεφάλαιο 4: Συλλογή και επεξεργασία στοιχείων.....	55
4.1 Εισαγωγή	55
4.2 Συλλογή δεδομένων	55
4.2.1 Δεδομένα αλγορίθμου	55
4.3 Στατιστική αποτύπωση των δεδομένων της μελέτης.....	59
4.3.1 Εισαγωγή.....	59
4.3.2 Στατιστική αποτύπωση των μεταβλητών της βάσης δεδομένων πεζών	59
4.3.3 Στατιστική αποτύπωση των μεταβλητών της βάσης δεδομένων οχημάτων	60
4.3.4 Στατιστική αποτύπωση των μεταβλητών της βάσης δεδομένων του ελάχιστου χρόνου ως την πιθανή σύγκρουση	62
4.4 Ανάλυση της ακρίβειας των ενδείξεων του φωτεινού σηματοδότη ανά κύκλο εναλλαγής	63
4.4.1 Εισαγωγή.....	63
4.4.2 Ενοποίηση και κανονικοποίηση αρίθμησης καρέ	64
4.4.3 Μετατροπή καρέ σε μονάδες χρόνου	64
4.4.4 Αντιμετώπιση άγνωστων και ελλιπών τιμών	64
4.4.5 Ορισμός κύκλων φωτεινής σηματοδότησης.....	64
4.4.6 Αντιστοίχιση φάσεων και καθαρισμός δεδομένων.....	65
4.4.7 Αξιολόγηση και οπτικοποίηση της ακρίβειας ανίχνευσης φάσεων φωτεινών σηματοδοτών	65
4.5 Ανάλυση της ακρίβειας των παράνομων διαβάσεων πεζών	68

4.5.1 Σύντομη ανασκόπηση και σύνδεση με την τρέχουσα σύγκριση.....	68
4.5.2 Φιλτράρισμα και καταμέτρηση παράνομων διαβάσεων πεζών αυτοματοποιημένης μεθόδου	68
4.5.3 Φιλτράρισμα και καταμέτρηση παράνομων διαβάσεων πεζών χειροκίνητης μεθόδου	68
4.5.4 Αξιολόγηση και οπτικοποίηση της ακρίβειας ανίχνευσης παράνομων διαβάσεων πεζών	69
4.6 Ανάλυση της ακρίβειας των νόμιμων διαβάσεων πεζών	71
4.6.1 Σύντομη ανασκόπηση και σύνδεση με την τρέχουσα σύγκριση.....	71
4.6.2 Φιλτράρισμα και καταμέτρηση νόμιμων διαβάσεων πεζών αυτοματοποιημένης μεθόδου	71
4.6.3 Φιλτράρισμα και καταμέτρηση νόμιμων διαβάσεων πεζών χειροκίνητης μεθόδου	72
4.6.4 Αξιολόγηση και οπτικοποίηση της ακρίβειας ανίχνευσης νόμιμων διαβάσεων πεζών	73
Κεφάλαιο 5: Εφαρμογή μεθοδολογίας και αποτελέσματα	75
5.1 Εισαγωγή	75
5.2 Επεξεργασία και ενοποίηση δεδομένων πεζών	75
5.2.1 Χρονική ενοποίηση και κανονικοποίηση δεδομένων πεζών	75
5.2.2 Εκκαθάριση και διαχείριση τιμών φωτεινού σηματοδότη	76
5.2.3 Ορισμός και μοντελοποίηση κύκλων φωτεινής σηματοδότησης	76
5.2.4 Χρονική αντιστοίχιση πεζών με κύκλους και φάσεις σηματοδότησης.....	77
5.2.5 Ομαδοποίηση συμπεριφοράς πεζών ανά χρονική στιγμή	77
5.3 Επεξεργασία και ενοποίηση δεδομένων οχημάτων	78
5.3.1 Χρονική ενοποίηση και κανονικοποίηση δεδομένων οχημάτων.....	78
5.3.2 Εκκαθάριση και διαχείριση τιμών φωτεινού σηματοδότη	78
5.3.3 Ορισμός και μοντελοποίηση κύκλων φωτεινής σηματοδότησης	78
5.3.4 Χρονική αντιστοίχιση οχημάτων με κύκλους και φάσεις σηματοδότησης	79
5.3.5 Ομαδοποίηση συμπεριφοράς οχημάτων ανά χρονική στιγμή.....	79
5.4 Επεξεργασία και διαχείριση δεδομένων ελάχιστου χρόνου έως τη σύγκρουση (minTTC).....	79
5.4.1 Διαχείριση ακραίων τιμών στη μεταβλητή MinTTC	79
5.4.2 Ενοποίηση χρονικής ροής και μετατροπή Καρέ σε δευτερόλεπτα.....	79
5.4.3 Δημιουργία κύκλων φωτεινής σηματοδότησης.....	80
5.4.4 Χρονική αντιστοίχιση με κύκλους και φάσεις φωτεινής σηματοδότησης.....	80
5.4.5 Ομαδοποίηση δεδομένων ελάχιστου χρόνου έως τη σύγκρουση.....	81
5.5 Ενοποίηση και τελικός καθαρισμός δεδομένων	81
5.6 Διαχείριση ακραίων τιμών και μετασχηματισμός της μεταβλητής MinTTC	82

5.7 Ανάλυση Συσχέτισης με Point-Biserial Correlation.....	83
5.7.1 Ανάλυση Συσχέτισης με τη μεταβλητή IsInCrossingArea	84
5.7.2 Αποτελέσματα συσχέτισης με τη μεταβλητή IsInCrossingArea	84
5.7.3 Ανάλυση Συσχέτισης με τη μεταβλητή PedestrianStatus	86
5.7.4 Αποτελέσματα συσχέτισης με τη μεταβλητή PedestrianStatus	86
5.7.5 Οπτική διερεύνηση της ταχύτητας σε σχέση με τη νομιμότητα της πεζής συμπεριφοράς.....	88
5.7.6 Οπτική διερεύνηση της εγγύτητας σε σύγκρουση σε σχέση με τη νομιμότητα της πεζής συμπεριφοράς	89
5.8 Λογιστική Παλινδρόμηση	91
5.8.1 Εισαγωγή.....	91
5.8.2 Μοντέλο 1: Εξαρτημένη μεταβλητή – Παρουσία πεζού εντός/εκτός διάβασης	91
.....	95
5.8.3 Μοντέλο 2: Εξαρτημένη μεταβλητή – Νόμιμη/Παράνομη διέλευση πεζού	95
.....	97
5.9 Τυχαία Δάση.....	100
5.9.1 Εισαγωγή.....	100
5.9.2 Μοντέλο 1: Εξαρτημένη μεταβλητή – Παρουσία πεζού εντός/εκτός διάβασης	100
.....	102
5.9.3 Μοντέλο 2: Εξαρτημένη μεταβλητή – Νόμιμη/Παράνομη διέλευση πεζού	104
.....	105
Κεφάλαιο 6: Συμπεράσματα	108
6.1 Εισαγωγή.....	108
6.2 Συμπεράσματα σύγκρισης αυτοματοποιημένης – χειροκίνητης μεθόδου	109
6.3 Συμπεράσματα στατιστικών αναλύσεων.....	110
6.3.1 Συσχέτιση σημειακής διχοτομίας – Εξαρτημένη μεταβλητή: Παρουσία πεζού εντός/εκτός διάβασης.....	111
6.3.2 Συσχέτιση σημειακής διχοτομίας – Εξαρτημένη μεταβλητή: Νόμιμη/Παράνομη διέλευση	112
6.3.3 Λογιστική Παλινδρόμηση – Εξαρτημένη μεταβλητή: Παρουσία πεζού εντός/εκτός διάβασης.....	112
6.3.4 Λογιστική Παλινδρόμηση – Εξαρτημένη μεταβλητή: Νόμιμη/Παράνομη διέλευση πεζού	114
6.3.5 Τυχαία δάση – Εξαρτημένη μεταβλητή: Παρουσία πεζού εντός / εκτός διάβασης.....	115
6.3.6 Τυχαία δάση – Εξαρτημένη μεταβλητή: Νόμιμη/Παράνομη διέλευση πεζού	117
6.4 Προτάσεις για περαιτέρω έρευνα	118

Κεφάλαιο 1: Εισαγωγή

1.1 Γενική ανασκόπηση

Η διαχείριση της κυκλοφορίας και η διασφάλιση της οδικής ασφάλειας αποτελούν βασικές προκλήσεις στον σχεδιασμό των σύγχρονων πόλεων. Με την αυξανόμενη πυκνότητα και κινητικότητα στους αστικούς χώρους, καθίσταται αναγκαία η εύρυθμη και ασφαλής συνύπαρξη πεζών και οχημάτων. Σε αυτό το πλαίσιο, οι σηματοδοτούμενες διασταυρώσεις λειτουργούν ως κρίσιμα σημεία ρύθμισης της κυκλοφορίας και αποτροπής συγκρούσεων.

Ωστόσο, η ανθρώπινη συμπεριφορά δεν είναι πάντα προβλέψιμη ή συνεπής με τους κανόνες. Φαινόμενα μη συμμόρφωσης των πεζών με τη φωτεινή σηματοδότηση, όπως η **διέλευση με κόκκινο** ή **εκτός των προβλεπόμενων ορίων** (αυθαίρετη διάσχιση-jaywalking) παραμένουν συχνά, καθιστώντας τη μελέτη τέτοιων συμπεριφορών κρίσιμη για τη βελτίωση της οδικής ασφάλειας. Παραδοσιακά, η καταγραφή τέτοιων φαινομένων βασιζόταν σε χειροκίνητες μεθόδους, όπως **παρατηρήσεις πεδίου** και **ερωτηματολόγια**, που παρουσιάζουν περιορισμούς σε ακρίβεια και κλίμακα. Η εξέλιξη της **υπολογιστικής όρασης** (Computer Vision) προσφέρει πλέον τη δυνατότητα αυτοματοποιημένης ανίχνευσης και ανάλυσης της πεζής συμπεριφοράς, σε πραγματικό χρόνο και με υψηλό βαθμό αξιοπιστίας.

Η **Υπολογιστική Όραση** (Computer Vision) είναι πεδίο της Τεχνητής Νοημοσύνης που επικεντρώνεται στην αυτόματη κατανόηση και εξαγωγή πληροφορίας από ψηφιακά οπτικά δεδομένα, όπως εικόνες και βίντεο. Με την αξιοποίηση **μεθόδων μηχανικής μάθησης** και ιδιαίτερα **αλγορίθμων βαθιάς μάθησης**, όπως τα συνελικτικά νευρωνικά δίκτυα (Convolutional Neural Networks – CNNs), καθίσταται δυνατός ο εντοπισμός, η αναγνώριση και η ταξινόμηση αντικειμένων, σκηνών ή ανθρώπινων συμπεριφορών με υψηλή ακρίβεια. Οι τεχνικές αυτές επιτρέπουν την επεξεργασία μεγάλου όγκου δεδομένων σε πραγματικό χρόνο, καθιστώντας την υπολογιστική όραση κατάλληλη για απαιτητικές εφαρμογές, όπως η αυτόνομη οδήγηση, η επιτήρηση ασφάλειας και η ανάλυση κυκλοφοριακής συμπεριφοράς. Στο πλαίσιο της παρούσας εργασίας, η τεχνολογία εφαρμόζεται για την **παρακολούθηση και ανάλυση της συμπεριφοράς πεζών σε διαβάσεις**, με στόχο την ανίχνευση φαινομένων μη συμμόρφωσης με τη φωτεινή σηματοδότηση και τη συμβολή στη βελτίωση της οδικής ασφάλειας.

Η παραβατική συμπεριφορά των πεζών στις σηματοδοτούμενες διασταυρώσεις αποτελεί αντικείμενο εκτενούς επιστημονικής μελέτης, η οποία έχει πλέον υπερβεί τα όρια της καταγραφής και εστιάζει στη διερεύνηση των παραγόντων που επηρεάζουν τη συμμόρφωση ή μη προς τη σηματοδότηση. Ερευνητικά μοντέλα έχουν δείξει ότι η απόφαση ενός πεζού να τηρήσει ή να παραβεί τους κανόνες επηρεάζεται από ένα σύνολο κοινωνικών, δημογραφικών και περιβαλλοντικών μεταβλητών.

Σε αστικά περιβάλλοντα έχει καταδειχθεί η σημασία της **κοινωνικής δυναμικής** τη στιγμή της απόφασης, επιβεβαιώνοντας ότι η παραβατική συμπεριφορά δεν είναι αποτέλεσμα μόνο ατομικής στάσης, αλλά και της στιγμιαίας κοινωνικής κατάστασης (Rosenbloom, 2009; Brosseau et al., 2013, Herath & Amarasingha, 2022, Jay et al., 2020, Koh et al., 2014). Για παράδειγμα, η **παρουσία άλλων πεζών** που περιμένουν να διασχίσουν τον δρόμο φαίνεται να μειώνει την πιθανότητα παραβίασης του

κόκκινου σηματοδότη, λειτουργώντας αποτρεπτικά μέσω της ενίσχυσης του κοινωνικού ελέγχου (Raoniar & Maurya, 2022).

Η ποιότητα και ο σχεδιασμός της υποδομής επηρεάζουν καθοριστικά τη συμπεριφορά των πεζών στις διαβάσεις. Συγκεκριμένα, η **έλλειψη διαγράμμισης στις διαβάσεις** σχετίζεται με αυξημένες παραβάσεις, καθώς οι πεζοί είναι λιγότερο πιθανό να συμμορφωθούν όταν δεν υπάρχουν οπτικές ενδείξεις που να τους καθοδηγούν (Mukherjee & Mitra, 2020).

Νεότερα ποσοτικά μοντέλα έχουν καταλήξει πως μεταβλητές όπως ο **χρόνος αναμονής**, ο **κυκλοφοριακός φόρτος**, η **ταχύτητα των οχημάτων** και το **μήκος της διάβασης** σχετίζονται, επίσης, σημαντικά με την πιθανότητα παραβίασης (Kumar & Ghosh, 2022; Raoniar & Maurya, 2022; Brosseau et al., 2013, Araya-Porras et al., 2022).

Ως προς αυτή την κατεύθυνση, έχει καταδειχθεί πως **περιοχές χωρίς σήμανση**, με **χαμηλή ορατότητα ή ανεπαρκή φωτισμό** συνδέονται, εξίσου, με αυξημένα ποσοστά παραβίασης, ενώ η παρουσία στοιχείων, όπως **νησίδες και σηματοδότες**, ευνοεί μια πιο προσεκτική συμπεριφορά. (Budzynski et al., 2021, Koh et al., 2014).

Η φθαρμένη ή **ανεπαρκής διαγράμμιση** και η **απουσία** κατακόρυφης και οριζόντιας **σήμανσης** σχετίζονται με αυξημένη μη συμμόρφωση, καθώς καθιστούν το σημείο διάσχισης λιγότερο ορατό και προβλέψιμο (Budzynski et al., 2021) . Επιπλέον, η **παρουσία στάσεων λεωφορείου** πλησίον της διάβασης δύναται να περιορίσει την ορατότητα για οδηγούς και πεζούς, αυξάνοντας την πιθανότητα παραβιάσεων (Budzynski et al., 2021) . Αντιθέτως, η ενσωμάτωση **κουμπιών χειροκίνητης ενεργοποίησης, οπτικών χρονόμετρων αντίστροφης μέτρησης και ηχητικών σημάτων** βελτιώνει τη συμμόρφωση των πεζών, ιδίως για ηλικιωμένους και ευάλωτους χρήστες, καθώς αυξάνει την αντιληπτή προβλεψιμότητα και ασφάλεια του συστήματος (Koh et al., 2014) .

Η **παρουσία χρονόμετρου** φαίνεται να περιορίζει την παραβατικότητα, κυρίως σε γυναίκες και άτομα άνω των 40 ετών, ενώ σε νεότερους πεζούς ενδέχεται να ενισχύει την πρώιμη διέλευση (Lipovac et al., 2013). Αυτό το εύρημα της βιβλιογραφίας είναι πλήρως συμβατό με την γενικότερα παγιωμένη αντίληψη για τη μεταβλητή της **ηλικίας** για την οποία έχει καταδειχθεί ότι όσο εκείνη αυξάνεται τόσο περισσότερο σχετίζεται με πιο προσεκτική συμπεριφορά αλλά και με μειωμένο κίνδυνο παραβίασης (Lipovac et al., 2013, Araya-Porras et al., 2022).

Τέλος, κάνοντας αναφορά και σε κάποια **προσωπικά χαρακτηριστικά** των πεζών, έχει καταγραφεί πως η αυξημένη **ταχύτητα βάδισης** έχει συσχετιστεί με υψηλότερη πιθανότητα παραβίασης της σηματοδότησης (Herath & Amarasingha, 2022). Η συγκεκριμένη διαφορά στην ταχύτητα μεταξύ συμμορφούμενων και παραβατών οφείλεται κυρίως στο μεγαλύτερο μήκος βήματος και όχι στη συχνότητα, γεγονός που υποδηλώνει προσπάθεια επιτάχυνσης της διέλευσης σε φάσεις μη επιτρεπόμενες για πεζούς (Hussein et al., 2015). Η **ταχύτητα**, το είδος της κίνησης (**βάδισμα ή τρέξιμο**), και η **χρήση ή μη της διάβασης** αναδεικνύονται ως σημαντικοί παράγοντες πρόβλεψης μη συμμόρφωσης με τη σηματοδότηση (Herath & Amarasingha, 2022).

Ομολογουμένως, η κατανόηση της παραβατικής συμπεριφοράς των πεζών στις σηματοδοτούμενες διασταυρώσεις αποτελεί κρίσιμο βήμα για τη βελτίωση της οδικής ασφάλειας και του αστικού σχεδιασμού. Ο συνδυασμός τεχνολογικών εργαλείων, όπως η υπολογιστική όραση, με ποσοτικές και κοινωνικές προσεγγίσεις, μπορεί να

προσφέρει πολύτιμες λύσεις για την πρόληψη ατυχημάτων και την ενίσχυση της συμμόρφωσης, λαμβάνοντας υπόψη τα ιδιαίτερα χαρακτηριστικά του περιβάλλοντος και των χρηστών του.

1.2 Στόχος διπλωματικής εργασίας

Η παρούσα εργασία εστιάζει στη μελέτη της συμπεριφοράς των πεζών σε σηματοδοτούμενες διασταυρώσεις, μέσα από τη συλλογή και ανάλυση δεδομένων που προέρχονται τόσο από **χειροκίνητη παρατήρηση και καταγραφή** όσο και από **αυτοματοποιημένες** μεθόδους βασισμένες σε τεχνολογίες υπολογιστικής όρασης. Οι δύο προσεγγίσεις εφαρμόστηκαν ταυτόχρονα στο ίδιο περιβάλλον μελέτης, προκειμένου να διασφαλιστεί η συγκρισιμότητα των δεδομένων και να αξιολογηθεί η αποτελεσματικότητα της τεχνολογικής λύσης σε πραγματικές συνθήκες κυκλοφορίας.

Βασικός στόχος της εργασίας είναι η **συγκριτική αξιολόγηση** των δύο μεθόδων, με έμφαση στον υπολογισμό της **ακρίβειας** της αυτοματοποιημένης διαδικασίας, όπως αυτή προκύπτει από τη σύγκρισή της με τη χειροκίνητη μέθοδο. Η ακρίβεια αφορά κρίσιμα στοιχεία της ανάλυσης, όπως η αναγνώριση της ένδειξης του **φωτεινού σηματοδότη** ανά κύκλο φωτεινού σηματοδότη, καθώς και η καταγραφή **νόμιμων** και **παράνομων** διελεύσεων πεζών στη σηματοδοτούμενη διάβαση.

Επιπλέον, στόχος αποτελεί η επιλογή και εφαρμογή της κατάλληλης μεθόδου **στατιστικής ανάλυσης**, ούτως ώστε να προκύψουν αξιόπιστα και τεκμηριωμένα αποτελέσματα αναφορικά με τις μεταβλητές που σχετίζονται είτε θετικά είτε αρνητικά με την παρουσία ή μη του πεζού εντός της διάβασης. Μέσα από αυτή τη διαδικασία, επιδιώκεται να προσδιοριστεί αν και σε ποιο βαθμό συγκεκριμένοι παράγοντες επηρεάζουν τη **συμμόρφωση των πεζών**, συμβάλλοντας στην κατανόηση των συνθηκών που ευνοούν ή αποθαρρύνουν την **παραβατική συμπεριφορά**. Πέρα από την ανάλυση των συσχετίσεων, αξιοποιήθηκαν επίσης μεθοδολογίες **πρόβλεψης**, ώστε να ενισχυθεί η δυνατότητα κατανόησης και εκτίμησης της πιθανότητας εμφάνισης συγκεκριμένων μορφών συμπεριφοράς.

1.3 Μεθοδολογία διπλωματικής εργασίας

Για την επίτευξη των ερευνητικών στόχων της παρούσας Διπλωματικής Εργασίας, ακολουθήθηκε μια συστηματική μεθοδολογική προσέγγιση, η οποία περιλαμβάνει τη βιβλιογραφική ανασκόπηση, τη συλλογή δεδομένων μέσω χειροκίνητης παρατήρησης και αυτοματοποιημένης ανάλυσης οπτικού υλικού, καθώς και τη στατιστική επεξεργασία των αποτελεσμάτων.

Στο **πρώτο στάδιο** της εργασίας, πραγματοποιήθηκε διαδικτυακή **βιβλιογραφική ανασκόπηση** με σκοπό τον εντοπισμό σχετικών ερευνών και την ανάλυση των μεθοδολογικών προσεγγίσεων που έχουν εφαρμοστεί σε ανάλογες περιπτώσεις. Η αναζήτηση επικεντρώθηκε σε έγκυρες επιστημονικές ιστοσελίδες, διεθνή περιοδικά και ηλεκτρονικές βάσεις δεδομένων, ώστε να εντοπιστούν υπάρχοντα κενά στη γνώση και να προσδιοριστούν οι κατευθύνσεις για περαιτέρω ερευνητική εμβάθυνση.

Στη συνέχεια της εργασίας, πραγματοποιήθηκαν δύο βασικές ενέργειες για την αξιοποίηση του διαθέσιμου βιντεοσκοπικού υλικού. Αρχικά, εφαρμόστηκε **χειροκίνητη ανάλυση του υλικού**, κατά την οποία καταγράφηκαν δεδομένα που

σχετίζονται με τη συμπεριφορά των πεζών. Συγκεκριμένα, σημειώθηκαν η χρονική στιγμή (σε δευτερόλεπτα), η κατάσταση του φωτεινού σηματοδότη ανά κύκλο, ο αριθμός των πεζών που διέσχισαν νόμιμα κατά τη διάρκεια της πράσινης φάσης, καθώς και οι περιπτώσεις παραβατικής διέλευσης σε κόκκινη ή ενδιάμεση φάση. Η διαδικασία αυτή παρείχε ένα αξιόπιστο σύνολο δεδομένων αναφοράς, το οποίο χρησιμοποιήθηκε ως βάση για τη σύγκριση με τα αποτελέσματα της αυτοματοποιημένης μεθόδου.

Παράλληλα, το ίδιο βιντεοσκοπικό υλικό αναλύθηκε με τη **χρήση δομημένου αλγορίθμου σε περιβάλλον Matlab**, ο οποίος σχεδιάστηκε από το Εργαστήριο Κυκλοφοριακής Τεχνικής, με στόχο την εξαγωγή αυτοματοποιημένης βάσης δεδομένων (Ventura et al, 2025). Η βάση δεδομένων περιλάμβανε χαρακτηριστικά που αποτυπώνουν την κινηματική συμπεριφορά πεζών και οχημάτων (π.χ. ταχύτητα στους επιμέρους άξονες και συνολική ταχύτητα), χωρικές και χρονικές πληροφορίες (συντεταγμένες θέσης, χρονική σήμανση καρέ), καθώς και δεδομένα ταυτοποίησης (αναγνωριστικά πεζών και οχημάτων). Επιπλέον, καταγράφονταν η συνθήκη σηματοδότησης και η θέση ως προς τη διάβαση (π.χ. παρουσία εντός ή εκτός διαβάσεων, ένδειξη φωτεινού σηματοδότη), ο χαρακτηρισμός της ενέργειας ως νόμιμης ή μη, καθώς και δείκτες αξιοπιστίας ανίχνευσης και μετρικές επικινδυνότητας, όπως ο ελάχιστος χρόνος έως πιθανή σύγκρουση (Minimum Time To Collision).

Αφού ολοκληρώθηκε η συλλογή δεδομένων, ακολούθησε η ανάλυσή τους με τη χρήση σύγχρονων στατιστικών μεθόδων. Η επεξεργασία των δεδομένων πραγματοποιήθηκε στη γλώσσα προγραμματισμού **Python**, όπου αναπτύχθηκαν κατάλληλα στατιστικά μοντέλα με σκοπό την ερμηνεία και ποσοτικοποίηση της σχέσης μεταξύ διαφορετικών μεταβλητών που σχετίζονται με τη συμπεριφορά των πεζών σε σηματοδοτούμενες διασταυρώσεις.

Το τελικό στάδιο της μεθοδολογικής προσέγγισης επικεντρώνεται στην **ανάλυση των ευρημάτων** και στην εξαγωγή αξιόπιστων συμπερασμάτων. Μέσα από την επεξεργασία των δεδομένων επιδιώκεται η καλύτερη κατανόηση των παραγόντων που επηρεάζουν τη συμπεριφορά των πεζών και η ανάδειξη στοιχείων που μπορούν να αξιοποιηθούν για τη βελτίωση της αστικής κυκλοφορίας και της ασφάλειας. Παράλληλα, διατυπώνονται προτάσεις για **περαιτέρω έρευνα**, με στόχο την ενίσχυση της αποτελεσματικότητας μελλοντικών εφαρμογών και παρεμβάσεων.

Η συνολική μεθοδολογική προσέγγιση της παρούσας Διπλωματικής Εργασίας αποτυπώνεται σχηματικά στο διάγραμμα (Γράφημα 1) που ακολουθεί:



Γράφημα 1 : Μεθοδολογική προσέγγιση

1.4 Δομή διπλωματικής εργασίας

Βάσει της μεθοδολογίας που ακολουθείται, διαμορφώνεται η γενική δομή της παρούσας διπλωματικής εργασίας, με στόχο την ξεκάθαρη και συνεκτική παρουσίαση του θέματος. Ακολουθεί μια σύντομη επισκόπηση των κεφαλαίων που την συνθέτουν.

Το **Κεφάλαιο 1** εισάγει το ευρύτερο **ερευνητικό πλαίσιο της εργασίας**, ξεκινώντας με μια γενική επισκόπηση του προβλήματος της παραβατικής πεζής συμπεριφοράς και της συμβολής της υπολογιστικής όρασης στην παρακολούθηση και ανάλυσή της. Στη συνέχεια, διατυπώνονται οι στόχοι της μελέτης, και η παρουσίαση της μεθοδολογικής προσέγγισης που εφαρμόστηκε. Τέλος, περιγράφεται η δομή της εργασίας και η διάρθρωση των επιμέρους κεφαλαίων.

Στο **Κεφάλαιο 2** παρουσιάζεται η **βιβλιογραφική ανασκόπηση**, όπου παρατίθενται σχετικές έρευνες και οι χρησιμοποιούμενες μεθοδολογίες, προσφέροντας μια εμπειριστατωμένη κατανόηση του ερευνητικού πεδίου.

Το **κεφάλαιο 3** παρουσιάζει το **θεωρητικό πλαίσιο** των στατιστικών μεθόδων που χρησιμοποιήθηκαν στην ανάλυση συμπεριφοράς των πεζών. Περιγράφονται βασικές έννοιες στατιστικής, η διάκριση μεταξύ συσχέτισης και πρόβλεψης, και αναλύονται οι τεχνικές Pearson και Point-Biserial Correlation για τη μελέτη σχέσεων μεταξύ μεταβλητών. Επίσης, εξετάζονται τα μοντέλα γραμμικής και λογιστικής παλινδρόμησης για την πρόβλεψη δυαδικών αποτελεσμάτων, καθώς και η μέθοδος μηχανικής μάθησης Random Forest για ταξινόμηση και εκτίμηση σημαντικότητας χαρακτηριστικών.

Το **Κεφάλαιο 4** περιγράφει τη **διαδικασία συλλογής και επεξεργασίας των δεδομένων** που αξιοποιήθηκαν για την ανάλυση της πεζής και οδικής συμπεριφοράς. Παρουσιάζονται οι πηγές των δεδομένων (χειροκίνητες και αυτοματοποιημένες

καταγραφές), η τεχνολογική υποδομή του αλγορίθμου υπολογιστικής όρασης, καθώς και οι τεχνικές προ-επεξεργασίας και κανονικοποίησης που εφαρμόστηκαν. Ακολουθεί η στατιστική αποτύπωση των μεταβλητών για πεζούς, οχήματα και τον ελάχιστο χρόνο έως τη σύγκρουση, και τέλος πραγματοποιείται αξιολόγηση της ακρίβειας της ανίχνευσης των φάσεων του σηματοδότη και των νόμιμων και παράνομων διαβάσεων πεζών, μέσα από σύγκριση μεταξύ της αυτόματης και της χειροκίνητης μεθόδου.

Το **Κεφάλαιο 5** περιλαμβάνει την παρουσίαση των **μαθηματικών μοντέλων** και των αποτελεσμάτων της στατιστικής ανάλυσης. Αναλύονται οι διαδικασίες ανάπτυξης των μοντέλων και αξιολογείται η αξιοπιστία και η σημασία των συμπερασμάτων.

Στο **Κεφάλαιο 6** παρουσιάζονται τα βασικά **συμπεράσματα** που προκύπτουν από την ανάλυση των δεδομένων και την ερμηνεία των μαθηματικών μοντέλων. Επιπλέον, διατυπώνονται **προτάσεις για μελλοντική έρευνα**, οι οποίες θα μπορούσαν να συμβάλουν ουσιαστικά στη βελτίωση της οδικής ασφάλειας και στην ανάπτυξη νέων στρατηγικών πρόληψης τροχαίων ατυχημάτων.

Κεφάλαιο 2: Βιβλιογραφική Ανασκόπηση

2.1 Εισαγωγή

Το παρόν κεφάλαιο αφορά στη βιβλιογραφική ανασκόπηση που πραγματοποιήθηκε για την εκπόνηση της παρούσας Διπλωματικής Εργασίας. Περιλαμβάνει σχετικές έρευνες, με σκοπό τον προσδιορισμό του αντικειμένου της εργασίας και την επιλογή μεθοδολογιών, ενώ, αναζητήθηκαν μελέτες τόσο σε εγχώριο όσο και σε διεθνές επίπεδο. Μέσω της ανάλυσης και σύνθεσης τους, προκύπτει μια ολοκληρωμένη εικόνα καθώς και χρήσιμα συμπεράσματα. Συγκεκριμένα, οι σηματοδοτημένες διασταυρώσεις θεωρείται διεθνώς πως αποτελούν σημεία με **αυξημένα ποσοστά ατυχημάτων** που αφορούν πεζούς, καθιστώντας την κατανόηση της συμπεριφοράς τους σε αυτά τα περιβάλλοντα ζήτημα κρίσιμης σημασίας (Porter et al., 2020).

Στο πλαίσιο αυτό, η παρούσα βιβλιογραφική ανασκόπηση στοχεύει να διερευνήσει πώς οι πεζοί αντιδρούν στη σηματοδότηση κυκλοφορίας, με έμφαση στη **συχνότητα** και στους **λόγους παραβίασης** της κόκκινης σηματοδότησης. Παράλληλα, εξετάζει τους **παράγοντες** που επηρεάζουν τη λήψη αποφάσεων των πεζών, αναλύοντας τόσο εξωτερικές συνθήκες, όπως η πυκνότητα κυκλοφορίας, όσο και εσωτερικές παραμέτρους, όπως οι κοινωνικές και ψυχολογικές επιρροές. Από τη βιβλιογραφία προέκυψε ότι η συλλογή δεδομένων για την καταγραφή και ανάλυση αυτών των συμπεριφορών πραγματοποιείται με δύο βασικές προσεγγίσεις: αφενός, μέσω **παραδοσιακών μεθόδων**, όπως οι επιτόπιες παρατηρήσεις, που βασίζονται στη φυσική παρουσία ερευνητών, και αφετέρου, μέσω εξελιγμένων **τεχνολογιών υπολογιστικής όρασης** (γνωστό στη διεθνή βιβλιογραφία ως computer vision), που επιτρέπουν την αυτοματοποιημένη και ακριβή ανάλυση του οδικού περιβάλλοντος στο αστικό πλαίσιο σε πραγματικό χρόνο. Η ανασκόπηση αυτή επιδιώκει να συνθέσει τα ευρήματα των ερευνών, προσφέροντας μια ολοκληρωμένη εικόνα για τη συμπεριφορά των πεζών και τις παραμέτρους που συνδέονται με την οδική ασφάλεια στις σηματοδοτούμενες διασταυρώσεις.

2.2 Μοτίβα συμπεριφοράς πεζών σε σηματοδοτούμενες διασταυρώσεις

2.2.1 Εισαγωγή

Οι συγκρούσεις πεζών και οχημάτων στις σηματοδοτούμενες διασταυρώσεις, ιδιαίτερα σε αστικά περιβάλλοντα, αποτελούν σοβαρό ζήτημα ασφάλειας με το ποσοστό παραβίασης των σημάτων από πεζούς να συνδέεται θετικά με τη συχνότητα θανατηφόρων ατυχημάτων (Mukherjee&Mitra, 2020). Διεθνείς μελέτες επιβεβαιώνουν ότι οι συγκρούσεις μεταξύ οχημάτων και πεζών συχνά προκαλούνται από τους ίδιους τους πεζούς, επειδή **συμμετέχουν σε ακατάλληλες ή παράνομες** συμπεριφορές διάβασης (Cookson et al., 2011). Όσον αφορά τη γενική επικράτηση, μετρήσεις πεδίου στις Βρυξέλλες (Populer, 2014), το Παρίσι (Huguenin & Richard, 2010) και το Αμβούργο (Schlabach, 2010) δείχνουν ότι περίπου το 20-25% των πεζών παραβιάζουν τα φανάρια κατά τη διάβαση, ενώ, παρόμοιες εκτιμήσεις έχουν καταγραφεί και στην Αυστραλία (King et al., 2009).

Ωστόσο, εξετάζοντας μεμονωμένες χώρες, αυτό το ποσοστό κυμαίνεται από 38% στη Σλοβενία, 44.6% στην Ινδία, με τη συχνότητα παραβίασης να είναι υψηλότερη στο

δεύτερο μισό της περιόδου του κόκκινου φωτός(Kumar & Ghosh,2022), έως και 82% στην Ισπανία (Torfs et al., 2016). Τέλος, τεκμηριώνεται ότι ο **κίνδυνος σύγκρουσης** αυξάνεται σημαντικά (οκταπλασιάζεται) όταν οι πεζοί διασχίζουν τον δρόμο παράνομα, γεγονός που αναδεικνύει τις επιπτώσεις αυτής της συμπεριφοράς στην οδική ασφάλεια (King et al., 2009). Η συμπεριφορά αυτή φαίνεται να ενισχύεται από την αντίληψη ότι δεν υπάρχει αστυνομική επιτήρηση, καθώς μελέτες δείχνουν πως οι πεζοί είναι πιο πιθανό να παραβούν τους κανόνες όταν θεωρούν απίθανη την επιβολή κυρώσεων (Akin, 2003).

2.2.2 Ταχύτητα πεζών και προσαρμογή της κατά τη διάβαση

Η συμπεριφορά των πεζών κατά τη διάρκεια της διάβασης επηρεάζεται από παράγοντες όπως η αρχική **ταχύτητα βάδισης**, οι ξαφνικές **μεταβολές** της και οι συνθήκες του **αστικού περιβάλλοντος**, περιλαμβανομένων της γεωμετρίας της διάβασης και του φωτεινού σηματοδότη (Alhajyaseen & Iryo-Asano, 2017). Σημαντικές μεταβολές στην ταχύτητα, κυρίως απότομες επιταχύνσεις, παρατηρούνται συχνά κατά την είσοδο του πεζού στις περιοχές πιθανής σύγκρουσης με οχήματα, γεγονός που τις καθιστά κρίσιμες για την κατανόηση της δυναμικής των συγκρούσεων, δεδομένου ότι δεν είναι προβλέψιμες από τους οδηγούς (Alhajyaseen & Iryo-Asano, 2017). Η **ταχύτητα εισόδου**, ο **απαιτούμενος χρόνος** για την ολοκλήρωση της διάβασης πριν την έναρξη της κόκκινης φάσης καθώς και το **μήκος της διάβασης** αποτελούν καθοριστικούς παράγοντες που επηρεάζουν την επιλογή του πεζού να επιταχύνει, να επιβραδύνει ή να διατηρήσει σταθερή ταχύτητα (Alhajyaseen & Iryo-Asano, 2017).

Η μεταβλητότητα στην ταχύτητα βάδισης σχετίζεται επίσης με ατομικά χαρακτηριστικά, όπως η **ηλικία** και το **φύλο**, καθώς και με τη **γεωμετρία** του χώρου (Zhang et al., 2011), ενώ η **αλληλεπίδραση** μεταξύ πεζών και η πυκνότητα του πλήθους διαμορφώνουν σημαντικά τη δυναμική της κίνησης, ιδιαίτερα σε ροές αντίθετων κατευθύνσεων (Feliciani et al., 2018). Επιπλέον, η ετερογένεια στη συμπεριφορά των πεζών, όπως αποτυπώνεται είτε σε φυσικά χαρακτηριστικά - όπως η ταχύτητα βάδισης, ο καταλαμβανόμενος χώρος και η κίνηση σε ζεύγη - είτε σε ψυχολογικούς παράγοντες - όπως η εκτιμώμενη διάρκεια της πράσινης φάσης και η υποκειμενική πίεση χρόνου - επηρεάζει σε μεγάλο βαθμό τις επιλογές ταχύτητας και τις στρατηγικές διάσχισης, ιδιαίτερα σε σηματοδοτημένες διαβάσεις εντός πόλης. (Tian et al., 2023).

2.2.3 Αβεβαιότητα και λήψη αποφάσεων κατά τη διάβαση

Η αντιμετώπιση της ασφαλούς διάσχισης του δρόμου προκύπτει, μεταξύ άλλων , και από την **αβεβαιότητα** που αισθάνονται οι πεζοί σχετικά με την εκτίμηση του κατάλληλου χρόνου διάσχισης. Συχνά, οι πεζοί αντιμετωπίζουν στιγμές αβεβαιότητας λόγω **εσφαλμένης εκτίμησης** της κατάλληλης στιγμής για να διασχίσουν, γεγονός που επηρεάζει τις αποφάσεις τους και μπορεί να οδηγήσει σε καθυστερήσεις ή αλλαγές στην ταχύτητα και την πορεία τους.

Συγκεκριμένα, σε σχετική έρευνα, το ποσοστό αβεβαιότητας βρέθηκε μικρότερο στους Γάλλους πεζούς (5%) σε σχέση με τους Ιάπωνες (10%), ενώ ο χρόνος αβεβαιότητας ήταν μεγαλύτερος στη Ναγκόγια (1,55 δευτερόλεπτα) σε σύγκριση με το Στρασβούργο (1,22 δευτερόλεπτα) (Jay et al., 2020). Οι συνοδευόμενοι πεζοί στην Ιαπωνία παρουσίασαν μικρότερο χρόνο αβεβαιότητας, ενώ η ηλικία φάνηκε να επηρεάζει περισσότερο τους Γάλλους, με τους 20χρονους να εμφανίζουν την υψηλότερη αβεβαιότητα και τους 60χρονους τη χαμηλότερη. Επιπλέον, οι πεζοί ήταν πιο πιθανό

να εγκαταλείψουν τη διάσχιση σε δρόμους με περισσότερες λωρίδες, αν και η πιθανότητα αυτή μειωνόταν όσο περισσότερα άτομα είχαν ήδη ξεκινήσει τη διάσχιση (Jay et al., 2020).

2.2.4 Απόσπαση προσοχής και κίνδυνοι κατά τη διάβαση

Ένα ακόμα παρατηρούμενο μοτίβο συμπεριφοράς των πεζών κατά τη διάβαση του δρόμου είναι η **απόσπαση προσοχής**, η οποία μπορεί να οδηγήσει σε αλλαγές στις κινήσεις τους και να αυξήσει τον κίνδυνο σύγκρουσης. Συγκεκριμένα, η ανάλυση δεδομένων που συλλέχθηκαν μέσω αυτοματοποιημένης επεξεργασίας βίντεο από δύο διασταυρώσεις στο Βανκούβερ έδειξε ότι οι αποσπασμένοι πεζοί διατηρούν **μικρότερες αποστάσεις** από τα οχήματα και κινούνται με μειωμένες ταχύτητες (Alsharif et al., 2024).

Οι πεζοί που είναι αποσπασμένοι από την προσοχή τους, όπως όταν χρησιμοποιούν **κινητό τηλέφωνο**, σπάνια παραχωρούν προτεραιότητα στα οχήματα και σπάνια προσαρμόζουν τη γωνία αλληλεπίδρασης με τα οχήματα, γεγονός που οδηγεί σε μειωμένη αποτελεσματικότητα πλοήγησης (Alsharif et al., 2024). Αντίθετα, οι **μη αποσπασμένοι** πεζοί εκτελούν ασφαλέστερους ελιγμούς, διατηρούν μεγαλύτερες αποστάσεις από τα οχήματα, παραχωρούν προτεραιότητα συχνότερα και ρυθμίζουν κατάλληλα την ταχύτητά τους. Οι αλληλεπιδράσεις τους με τα οχήματα παρουσίασαν μείωση 46,5% στη σοβαρότητα των συγκρούσεων και αύξηση 30,2% στην ελάχιστη απόσταση, σε σύγκριση με τις αλληλεπιδράσεις των αποσπασμένων πεζών (Alsharif et al., 2024).

2.3 Παράγοντες επιρροής παραβατικής συμπεριφοράς

2.3.1 Διαμόρφωση διαβάσεων και συμμόρφωση πεζών

Η **ύπαρξη σηματοδότη για πεζούς** με ή χωρίς ένδειξη αντίστροφης μέτρησης, μειώνει την πιθανότητα παραβιάσεων ενώ η ολική απουσία φωτεινού σηματοδότη για πεζούς την αυξάνει. Έχει παρατηρηθεί πως αν ο **χρόνος εκκαθάρισης** είναι μεγαλύτερος από τον χρόνο που απαιτείται για τη διάβαση του δρόμου, οι χρήστες της οδού είναι πιθανότερο να διαπράξουν παραβίαση ενώ, ειδικά, σε διασταυρώσεις χωρίς σηματοδότες για πεζούς, όπου ο χρόνος εκκαθάρισης είναι πάντα μικρότερος από τον απαιτούμενο χρόνο διάβασης, ακόμα περισσότεροι τείνουν να διασχίσουν επικίνδυνα. Η **στιγμή της άφιξης** στη διασταύρωση φαίνεται να έχει μικρή επίδραση στη συμπεριφορά διέλευσης των πεζών (Brosseau et al., 2013). Αναφορικά, με τη **διάρκεια της κόκκινης φάσης** για πεζούς, έχει διαπιστωθεί πως όσο μικρότερη είναι, τόσο μειώνεται και η πιθανότητα παράνομης διέλευσης, ενώ, αυτή αυξάνεται ραγδαία όταν αυξάνεται ο χρόνος αναμονής (Araya-Porras et al., 2022, Brosseau et al., 2013).

Εκτός από τις μεταβλητές που σχετίζονται με τις κυκλοφοριακές συνθήκες στη διάβαση, ορισμένα τεχνικά χαρακτηριστικά της υποδομής επηρεάζουν ουσιαστικά τη συμπεριφορά των πεζών. Συγκεκριμένα, έχει διαπιστωθεί ότι το **αυξημένο μήκος** της διάβασης συνδέεται με υψηλότερα ποσοστά μη συμμόρφωσης των πεζών, όπως καταγράφηκε στην Κόστα Ρίκα (Araya-Porras et al., 2022). Επιπλέον, σε έρευνα που πραγματοποιήθηκε σε σηματοδοτημένες διαβάσεις στις εννέα μεγαλύτερες πόλεις του Βελγίου (Αμβέρσα, Μπριζ, Βρυξέλλες, Σαρλερουά, Γάνδη, Λουβαίν, Λιέγη, Μονς και Ναμούρ), διαπιστώθηκε ότι τόσο τα **φθαρμένα οριζόντια σήματα** όσο και τα **κουμπιά πίεσης** συνέβαλαν στην αύξηση της συχνότητας παραβάσεων

(Diependaele, 2019). Από την άλλη, τα **βοηθητικά σήματα** στις διαβάσεις, είτε οπτικά είτε ηχητικά, έχουν θετική επίδραση, καθώς συμβάλλουν στη μείωση της συχνότητας παραβάσεων από τους πεζούς (Brousseau et al., 2013).

Επιπλέον, σε ανάλυση δεδομένων από 55 σηματοδοτούμενες διασταυρώσεις στην Καλκούτα της Ινδίας, αναδείχθηκαν και άλλοι παράγοντες που εντείνουν την παραβατικότητα και σχετίζονται με τη διαμόρφωση της διάβασης, όπως η **απουσία διαγράμμισης, η ανεπαρκής ορατότητα και η στάθμευση** κοντά στο σημείο διάβασης (Mukherjee & Mitra, 2020). Το τελευταίο εύρημα έρχεται σε αντίθεση με τα συμπεράσματα των Yannis et al. (2013), οι οποίοι υποστήριξαν ότι οι πεζοί τείνουν να είναι πιο προσεκτικοί όταν υπάρχουν παράνομα σταθμευμένα οχήματα κοντά στη διάβαση, σε περίπτωση μελέτης στην Αθήνα. Τέλος, η παρεκκλίνουσα συμπεριφορά των πεζών φαίνεται, επίσης, να επηρεάζεται από τον **προορισμό** τους μετά τη διάβαση (Mukherjee & Mitra, 2020), καθώς έχει παρατηρηθεί ότι άτομα που μετακινούνται για λόγους εργασίας ή φοίτησης παρουσιάζουν υψηλότερα ποσοστά παραβάσεων των κανόνων κυκλοφορίας (Guo et al., 2011).

2.3.2 Φύλο και συμμόρφωση πεζών

Το φύλο είναι μια ακόμα παράμετρος που φαίνεται να επηρεάζει τη συμπεριφορά των πεζών, με τους **άνδρες** να παραβιάζουν συχνότερα τους κανόνες κυκλοφορίας σε σύγκριση με τις γυναίκες (Zhu et al., 2022, Brousseau et al., 2013). Ωστόσο, η επίδραση του φύλου μπορεί να μεταβάλλεται ανάλογα με το περιβάλλον. Σε **τουριστικά περιβάλλοντα**, όπως στη Viareggio της Ιταλίας, δεν έχουν παρατηρηθεί στατιστικά σημαντικές διαφορές ως προς τη συμμόρφωση των πεζών βάσει φύλου, ηλικίας ή μεγέθους της ομάδας (Pratelli et al., 2017). Αντίθετα, σε **εργασιακά περιβάλλοντα**, δηλαδή αστικά περιβάλλοντα όπου κυριαρχούν οι μετακινήσεις εργαζομένων και φοιτητών, οι άνδρες πεζοί εμφανίζουν υψηλότερα ποσοστά παραβατικότητας σε σύγκριση με τις γυναίκες (Antonio Pratelli et al., 2019). Το ίδιο μοτίβο καταγράφηκε και σε αστικό περιβάλλον του Τελ Αβίβ, όπου η ανάλυση δεδομένων σε σηματοδοτούμενη διάβαση έδειξε ότι οι άνδρες διέσχισαν τον δρόμο με κόκκινο φανάρι συχνότερα από τις γυναίκες, ανεξαρτήτως του αν βρίσκονταν μόνοι ή σε ομάδα (Rosenbloom, 2009). Αναφερόμενοι στο **γυναικείο φύλλο**, σε σχετική μελέτη στο Χονγκ Κονγκ, τα αποτελέσματα ενός μοντέλου λογιστικής παλινδρόμησης, με τυχαίες παραμέτρους, κατέδειξαν ότι η τάση παραβίασης του κόκκινου φωτός από γυναίκες αυξάνεται όταν άλλοι πεζοί παραβιάζουν το κόκκινο φως (Zhu et al. 2021).

2.3.3 Ηλικία και συμμόρφωση πεζών

Η επίδραση της ηλικίας στη συμπεριφορά των πεζών φαίνεται να είναι σημαντική σε ορισμένα περιβάλλοντα, σε αντίθεση με ό,τι καταγράφηκε στην πόλη Viareggio, όπου δεν διαπιστώθηκαν στατιστικά σημαντικές διαφορές βάσει ηλικίας (Pratelli et al., 2017). Στην περιοχή Plateau-Mont-Royal του Μόντρεαλ, για παράδειγμα, καταγράφηκε αυξημένη παραβατικότητα μεταξύ **νεαρών ενηλίκων** (Brousseau et al., 2013), ενώ, έχει επιπλέον διαπιστωθεί ότι οι **μαθητές** είναι πιο πιθανό να ρισκάρουν κατά τη διάβαση, όταν συνομήλικοί τους επιδεικνύουν παραβατική συμπεριφορά, λειτουργώντας ως αρνητικά πρότυπα (Holm et al., 2018).

Αντίθετα, σε έξι σηματοδοτούμενες διαβάσεις του Χονγκ Κονγκ, οι οποίες χαρακτηρίζονται από υψηλό αριθμό τροχαίων ατυχημάτων, παρατηρήθηκε αυξημένη

πιθανότητα παραβίασης από πεζούς μεγαλύτερης ηλικίας, γεγονός που αποδόθηκε εν μέρει στο χαμηλό μορφωτικό επίπεδο των ηλικιωμένων στην περιοχή (Zhu et al., 2021). Συμπληρωματικά, έχει καταγραφεί ότι, παρόλο που οι **ηλικιωμένοι πεζοί** είναι γενικά πιο προσεκτικοί, εμπλέκονται σε μεγαλύτερο αριθμό ατυχημάτων λόγω της φυσικής ευαλωτότητας που τους χαρακτηρίζει (Bernhoft & Carstensen, 2008).

2.3.4 Κοινωνική επιρροή και συμμόρφωση πεζών

Η κοινωνική επιρροή παίζει επίσης καθοριστικό ρόλο στη διαμόρφωση της συμπεριφοράς των πεζών καθώς σε παρατήρηση διασταυρώσεων της Καλκούτας, η ανάπτυξη ενός δυαδικού μοντέλου λογιστικής παλινδρόμησης ανέδειξε ότι **ο αριθμός των πεζών** που διέσχιζαν επιτυχώς αλλά παραβατικά τον δρόμο επηρεάζει τη μη συμμόρφωση των υπολοίπων πεζών (Raoniar & Maurya, 2022). Η πιθανότητα ένας πεζός να περιμένει στη φάση του κόκκινου φωτός αυξάνεται όταν περισσότεροι άνθρωποι περιμένουν στο πεζοδρόμιο (Brosseau et al., 2013, Raoniar & Maurya, 2022), ενώ αντίθετα, η τάση παραβίασης του κόκκινου φωτός ενισχύεται όταν ήδη άλλοι πεζοί έχουν ξεκινήσει να διασχίζουν τον δρόμο παράνομα (Raoniar & Maurya, 2022). Επιπροσθέτως, έχει παρατηρηθεί ότι η **παρουσία άλλων πεζών** που περιμένουν στη διάβαση, κατά την άφιξη ενός πεζού, καθώς και η σταδιακή προσθήκη περισσότερων ατόμων, μειώνουν την πιθανότητα παραβίασης του σηματοδότη, ενισχύοντας τη συλλογική τήρηση του νόμου, ενώ οι πεζοί που είναι μόνοι έχουν μεγαλύτερη πιθανότητα να μη συμμορφωθούν με τον κανόνα (Kumar & Ghosh, 2022, Rosenbloom, 2009).

2.3.5 Κυκλοφοριακή δυναμική και συμμόρφωση πεζών

Οι παράγοντες που σχετίζονται με τη συνολική κυκλοφοριακή δυναμική και την πολυπλοκότητα του οδικού περιβάλλοντος φαίνεται να επηρεάζουν σημαντικά τη συμπεριφορά των πεζών στις διαβάσεις. Σε μελέτη που διεξήχθη σε μεγάλες πόλεις του Βελγίου, η επίδραση στη συχνότητα παραβίασης ερυθρού σηματοδότη από πεζούς **κατά την αύξηση του όγκου κυκλοφορίας** των μηχανοκίνητων οχημάτων εμφανίστηκε να έχει μειωμένη τάση (Araya-Porras et al., 2022; Wang et al., 2011; Yagil, 2000; Yang et al., 2005). Επιπλέον, σε συνθήκες που συνδέονται γενικά με **αυξημένη κυκλοφοριακή πολυπλοκότητα**, όπως οι ώρες αιχμής, ο μεγάλος αριθμός κατευθύνσεων κυκλοφορίας, οι πολλές λωρίδες ανά κατεύθυνση, καθώς και η παρουσία λωρίδων τραμ ή λεωφορείου, βρέθηκαν επίσης να μειώνουν τη συχνότητα των παραβάσεων (Diependaele, 2019).

2.4 Αυτόματη διάγνωση θεμάτων ασφαλείας πεζών σε σηματοδοτούμενες διασταυρώσεις

2.4.1 Εισαγωγή

Η αυτόματη διάγνωση θεμάτων ασφαλείας πεζών σε σηματοδοτούμενες διασταυρώσεις έχει αποτελέσει αντικείμενο εκτεταμένης έρευνας, αξιοποιώντας τόσο παραδοσιακές μεθόδους παρατήρησης όσο και σύγχρονες τεχνολογίες υπολογιστικής όρασης. Οι σύγχρονες προσεγγίσεις, με τη χρήση **αλγορίθμων ανίχνευσης** και ανάλυσης συμπεριφοράς, επιτρέπουν την ακριβή καταγραφή των παραβάσεων και των δυνητικών συγκρούσεων, βελτιώνοντας την κατανόηση των κινδύνων. Η

παρούσα ενότητα εξετάζει τη σχετική βιβλιογραφία, εστιάζοντας στις μεθόδους ανάλυσης, τα κύρια ευρήματα και τις προκλήσεις που σχετίζονται με τη συλλογή και επεξεργασία δεδομένων.

2.4.2 Αυτόματη εξαγωγή και ανάλυση τροχιών πεζών

Η **εξαγωγή τροχιών** πεζών αποτελεί θεμελιώδες βήμα στην ανάλυση δεδομένων μέσω υπολογιστικής όρασης, καθώς επιτρέπει τη μελέτη των προτύπων κίνησης και τη συστηματική παραγωγή στατιστικών ροής (Zhang et al., 2023). Στη βιβλιογραφία έχουν αναπτυχθεί διάφορες προσεγγίσεις που στοχεύουν στην αξιόπιστη εξαγωγή τέτοιων δεδομένων, οδηγώντας στη δημιουργία τριών βασικών τύπων στατιστικών: **χάρτες θερμότητας** διαδρομών, **κατανομές** διαδρομών ανά ζώνη και **σχέσεις ταχύτητας-πυκνότητας** (Zhang et al., 2023). Τα παραγόμενα στατιστικά δεδομένα χρησιμοποιούνται ευρέως στην ανάλυση της κυκλοφορίας πεζών, διευκολύνοντας τον σχεδιασμό αστικών υποδομών και βελτιώνοντας την ποιότητα της κυκλοφοριακής ροής (Zhang et al., 2023).

Για την εξέλιξη της διαδικασίας έχουν προταθεί διάφορες τεχνικές που στοχεύουν στη βελτίωση της ακρίβειας της παρακολούθησης, όπως η **ενσωμάτωση υψηλού επιπέδου χαρακτηριστικών** πεζών ως πρόσθετο στοιχείο, η χρήση ολοκληρωμένων **μέτρων ομοιότητας** που συνδυάζουν πολλαπλές ενδείξεις για την αντιστοίχιση ταυτότητας και η εφαρμογή **μηχανισμών δοκιμαστικής περιόδου** για τη διαχείριση περιπτώσεων απόκρυψης (Wong et al., 2021). Τα πειραματικά αποτελέσματα σε υπάρχουσες μελέτες επιβεβαιώνουν ότι η εφαρμογή αυτών των τεχνικών ενισχύει την ανθεκτικότητα της διαδικασίας παρακολούθησης πεζών, συμβάλλοντας στη βελτίωση της αξιοπιστίας των εξαγόμενων δεδομένων.

2.4.3 Αυτόματη ανίχνευση πεζών

Εκτός από τις τεχνικές που αποσκοπούν στη βελτίωση της παρακολούθησης πεζών, η επιστημονική έρευνα έχει επεκταθεί και στον τομέα της **ανίχνευσης σε απαιτητικά αστικά περιβάλλοντα**, όπου οι κάμερες συχνά διαθέτουν περιορισμένη ανάλυση και οι υπολογιστικοί πόροι είναι ιδιαίτερα περιορισμένοι.

Σε τέτοιες περιπτώσεις εφαρμόζονται αρχιτεκτονικές τύπου master-slave, οι οποίες διαχωρίζουν τη διαδικασία ανίχνευσης σε δύο φάσεις με σκοπό τη βελτιστοποίηση της υπολογιστικής απόδοσης. Συγκεκριμένα, ο **master ανιχνευτής** λειτουργεί σε χαμηλό ρυθμό καρέ και εντοπίζει υποψήφιας περιοχές όπου πιθανώς υπάρχουν πεζοί, ενώ ο **slave ανιχνευτής**, σε υψηλότερη συχνότητα λειτουργίας, επαληθεύει ή απορρίπτει αυτές τις υποθέσεις, προσφέροντας επιπλέον πληροφορίες σε πραγματικό χρόνο για την αναπαράσταση των πεζών που ενδέχεται να βρίσκονται σε επικίνδυνες καταστάσεις (Alahi et al., 2014). Η αρχιτεκτονική αυτή καθιστά το σύστημα κατάλληλο για εγκατάσταση σε οχήματα με περιορισμένες δυνατότητες απόκτησης δεδομένων και επεξεργασίας (Alahi et al., 2014).

Στο ίδιο πλαίσιο, αξιοποιούνται και οι περιγραφείς **FREAK** (Fast Retina Keypoint), οι οποίοι αποτελούν δυαδικούς περιγραφείς σχεδιασμένους για ταχεία εξαγωγή χαρακτηριστικών και ανθεκτικότητα σε παραμορφώσεις εικόνας (Alahi et al., 2014). Οι FREAK βασίζονται σε κατανομή σημείων εμπνευσμένη από τη λειτουργία του αμφιβληστροειδούς και επιτρέπουν υψηλή απόδοση με χαμηλό υπολογιστικό κόστος.

Η εξαγωγή των περιγραφέντων FREAK μπορεί να γίνει με δύο διαφορετικές διαμορφώσεις. Στην **αραιή διαμόρφωση**, οι περιγραφείς εξάγονται μόνο σε επιλεγμένα σημεία-κλειδιά, τα οποία εντοπίζονται με τον ανιχνευτή FAST (Features from Accelerated Segment Test) — έναν ιδιαίτερα γρήγορο αλγόριθμο εντοπισμού γωνιών, κατάλληλο για χρήση σε συστήματα με περιορισμένους υπολογιστικούς πόρους (Alahi et al., 2014). Αντίθετα, στην **πυκνή διαμόρφωση**, οι περιγραφείς υπολογίζονται σε τακτικά κατανεμημένα σημεία, σχηματίζοντας ένα ομοιόμορφο πλέγμα που καλύπτει ολόκληρη την περιοχή του πεζού. Η αραιή προσέγγιση είναι κατάλληλη για χρήση σε ενσωματωμένα (embedded) συστήματα — δηλαδή σε εξειδικευμένες, αυτόνομες υπολογιστικές μονάδες με περιορισμένη υπολογιστική ισχύ, όπως αυτές που χρησιμοποιούνται σε οχήματα (Alahi et al., 2014). Από την άλλη πλευρά, η πυκνή διαμόρφωση προσφέρει μεγαλύτερη ακρίβεια στην ανίχνευση, ξεπερνώντας τις επιδόσεις παλαιότερων μεθόδων (Alahi et al., 2014).

2.4.4 Αυτόματη διάγνωση θεμάτων ασφαλείας πεζών

Η υπολογιστική όραση έχει αξιοποιηθεί στην ανάλυση θεμάτων ασφάλειας πεζών σε σηματοδοτούμενες διασταυρώσεις, επιτρέποντας την αυτόματη καταγραφή και επεξεργασία επικίνδυνων συνθηκών μέσω της παρακολούθησης της αλληλεπίδρασης πεζών και οχημάτων.

Επί παραδείγματι, σε αστικό κόμβο του Βανκούβερ στον Καναδά, χρησιμοποιήθηκαν δεδομένα βιντεοσκόπησης και εφαρμόστηκαν τεχνικές ανάλυσης τροχιών, όπως ο αλγόριθμος **longest common subsequence (LCSS)**, ο οποίος υπολογίζει την ομοιότητα μεταξύ δύο διαδρομών λαμβάνοντας υπόψη επιμέρους αποκλίσεις ή θόρυβο, γεγονός που τον καθιστά κατάλληλο για πραγματικά δεδομένα με ακανόνιστες συμπεριφορές (Zaki et al., 2013). Η χρήση τέτοιων μεθόδων επέτρεψε την εκτίμηση των κινήσεων, την ανίχνευση παραβάσεων και την ανάδειξη πιθανών συγκρούσεων μέσω της χωρο-χρονικής σύγκλισης πορειών. Η αξιοπιστία των εντοπισμένων περιστατικών ενισχύθηκε μέσω χειροκίνητης επικύρωσης, επιτυγχάνοντας ποσοστό επιτυχίας άνω του 85% (Zaki et al., 2013).

Για την αξιολόγηση της σοβαρότητας των συμβάντων, στην προαναφερθείσα εφαρμογή χρησιμοποιήθηκαν δείκτες όπως ο **Time to Collision (TTC)**, ο οποίος εκτιμά τον χρόνο μέχρι πιθανή σύγκρουση μεταξύ πεζού και οχήματος, και ο **Post Encroachment Time (PET)**, που μετρά τον χρόνο ανάμεσα στη διέλευση του πρώτου και του δεύτερου χρήστη από το ίδιο σημείο εντός της περιοχής σύγκρουσης. Η χωρική κατανομή των επικίνδυνων σημείων αποτυπώθηκε μέσω θερμικών χαρτών. Οι περισσότερες επικίνδυνες αλληλεπιδράσεις παρατηρήθηκαν κατά την εκτέλεση δεξιών και αριστερών στροφών, με τις αριστερές να παρουσιάζουν μεγαλύτερη επικινδυνότητα λόγω περιορισμένης ορατότητας και μειωμένης δυνατότητας εκτίμησης της πορείας των πεζών από τους οδηγούς. Αυξημένα ποσοστά παραβάσεων εντοπίστηκαν επίσης σε διασταυρώσεις πλησίον στάσεων λεωφορείων, όπου παρατηρήθηκε επιτάχυνση της κίνησης των πεζών με σκοπό την έγκαιρη επιβίβαση, συχνά χωρίς συμμόρφωση προς τη σηματοδότηση (Zaki et al., 2013).

Αντίστοιχη προσέγγιση με έμφαση στην αυτόματη καταγραφή και αξιολόγηση επικίνδυνων αλληλεπιδράσεων μεταξύ πεζών και οχημάτων υιοθετήθηκε και στο πλαίσιο του **ερευνητικού έργου MOBIS** στην Πολωνία, το οποίο επικεντρώθηκε σε

μη σηματοδοτούμενες διαβάσεις πεζών (Olszewski et al., 2016). Η μεθοδολογία βασίστηκε επίσης στην ανάλυση βιντεοσκοπημένων δεδομένων κυκλοφορίας, με σκοπό την εξαγωγή τροχιών και τη μελέτη των συνθηκών προ της σύγκρουσης, μέσα από παραμέτρους όπως η ταχύτητα οχήματος (VV), η ταχύτητα πεζού (VP), η ελάχιστη απόσταση (S), και οι δείκτες Time to Collision (TTC) και Post Encroachment Time (PET), οι οποίοι χρησιμοποιήθηκαν ως υποκατάστατα δεικτών ασφαλείας. Αν και χρησιμοποιήθηκαν παρόμοιοι δείκτες εκτίμησης επικινδυνότητας με προηγούμενες μεθοδολογίες, η συγκεκριμένη μελέτη προχώρησε ένα βήμα παραπέρα, προτείνοντας μια τυπολογία των **αλληλεπιδράσεων μεταξύ πεζών και οχημάτων βάσει της θέσης και της χρονικής στιγμής εμφάνισης των δύο μερών στο πεδίο της κάμερας**. Οι αλληλεπιδράσεις ταξινομήθηκαν σε τέσσερις τύπους, εκ των οποίων οι πλέον χαρακτηριστικοί ήταν: (A1) διέλευση του οχήματος ακριβώς μπροστά από πεζό που βρίσκεται ήδη στη διάβαση, (B) διέλευση πίσω από πεζό που διασχίζει, και (C) επιβράδυνση ή στάση του οχήματος πριν από τη διάβαση (Olszewski et al., 2016). Οι περιπτώσεις A1 και B θεωρήθηκαν ενδεικτικές παραβίασης της προτεραιότητας των πεζών, ενώ η περίπτωση C συνδέεται με συμμόρφωση προς τον ΚΟΚ. Η κατηγοριοποίηση αυτή παρείχε ένα λειτουργικό πλαίσιο αξιολόγησης της συμπεριφοράς των οδηγών σε σχέση με την ασφάλεια των πεζών, επιτρέποντας ποσοτική αποτύπωση των παραβατικών ή προστατευτικών προσεγγίσεων.

Καινοτομία της παραπάνω προσέγγισης αποτέλεσε η συσχέτιση αυτών των καταστάσεων με την αντίληψη κινδύνου μέσω **ποιοτικής αξιολόγησης από παρατηρητές**. Μέσα από τη συλλογή **ερωτηματολογίων** βασισμένων σε 32 πραγματικά περιστατικά (τύπου A1 και B), κατασκευάστηκε ένας υποκειμενικός δείκτης αντιλαμβανόμενης επικινδυνότητας (danger score), ο οποίος αναλύθηκε στατιστικά ώστε να αποκαλύψει τη σημασία επιμέρους παραμέτρων. Μέσω **πολλαπλής γραμμικής παλινδρόμησης**, εξετάστηκε η επίδραση των μεταβλητών VV, VP, S και του τύπου αλληλεπίδρασης (A1 ή B) στον συγκεκριμένο δείκτη. Διαπιστώθηκε ότι τόσο η ταχύτητα του οχήματος (VV) όσο και η ελάχιστη απόσταση από τον πεζό (S) είχαν στατιστικά σημαντική επίδραση στον δείκτη επικινδυνότητας ($p < 0.05$), ενώ αντίθετα η ταχύτητα του πεζού (VP) και ο τύπος αλληλεπίδρασης δεν εμφάνισαν στατιστική συσχέτιση. Συγκεκριμένα, παρατηρήθηκε ότι όσο μεγαλύτερη είναι η ταχύτητα του οχήματος ή μικρότερη η απόσταση από τον πεζό, τόσο υψηλότερη είναι η εκτιμώμενη επικινδυνότητα, όπως αυτή εκφράστηκε μέσω του δείκτη (Olszewski et al., 2016) Η ένταξη της ανθρώπινης αντίληψης στην αξιολόγηση των συμβάντων προσέφερε μια πιο ρεαλιστική εικόνα του οδικού κινδύνου σε καθημερινές συνθήκες, ενισχύοντας την αξία των δεικτών που εξάγονται αυτόματα από τις τροχιές.

Παράλληλα, η μελέτη παρείχε εμπειρικά δεδομένα για την αποτελεσματικότητα παρεμβάσεων που στόχο έχουν τη βελτίωση της ασφάλειας, όπως **φωτεινά συστήματα ενεργητικής σήμανσης** (SignFlash και Levelite) και **ταχυδρομικά μαξιλάρια**, δηλαδή υπερυψωμένες επιφάνειες στην πορεία των οχημάτων οι οποίες προκαλούν ήπια επιβράδυνση. Η ανάλυση κατέδειξε ότι οι παρεμβάσεις αυτές, πέρα από τη στατιστικά σημαντική μείωση της ταχύτητας προσέγγισης, σχετίζονται και με τη μείωση παραβατικών συμπεριφορών εκ μέρους των οδηγών συμβάλλοντας έτσι στη βελτίωση της ασφάλειας των πεζών (Olszewski et al., 2016).

2.4.5 Αυτόματη ανίχνευση-πρόβλεψη παραβατικής συμπεριφοράς πεζών

Η πρόβλεψη παραβάσεων ερυθρού σηματοδότη από πεζούς σε σηματοδοτούμενες διαβάσεις μπορεί να επιτευχθεί με χρήση νευρωνικών δικτύων που αναλύουν τη χρονική ακολουθία της κίνησης. Ιδιαίτερα αποτελεσματικά κρίνεται ότι είναι τα δίκτυα τύπου **Long Short-Term Memory (LSTM)**, τα οποία διαθέτουν εσωτερικούς μηχανισμούς μνήμης και διατηρούν πληροφορίες από προηγούμενες χρονικές στιγμές, καθιστώντας τα κατάλληλα για την ανάλυση δεδομένων με χρονική εξάρτηση, όπως η σταδιακή κίνηση ενός πεζού στον χώρο (Zhang et al., 2020). Η εκπαίδευση τέτοιων μοντέλων μπορεί να βασιστεί σε χωρικά χαρακτηριστικά (όπως η θέση και η κατεύθυνση κίνησης του πεζού), καθώς και σε συμπληρωματικά γνωρίσματα, όπως το φύλο και η συμμετοχή σε ομάδα πεζών, με στόχο την εκτίμηση της πρόθεσης παραβίασης του σηματοδότη πριν την πραγματοποίησή της.

Η πρόβλεψη μπορεί να πραγματοποιηθεί σε διαδοχικά χρονικά σημεία έως και **1,5 δευτερόλεπτο πριν** από τη διάβαση, επιτρέποντας τη δυναμική εκτίμηση της πιθανής παραβατικής συμπεριφοράς. Η απόδοση του συστήματος αξιολογείται βάσει δεικτών όπως η **ακρίβεια** (accuracy), που μετρά το ποσοστό των συνολικά σωστών προβλέψεων, και η **ευαισθησία** (recall), που αποτυπώνει την ικανότητα του συστήματος να εντοπίζει όλες τις πραγματικές περιπτώσεις παραβίασης.

Σε έρευνα που διεξήχθη σε σηματοδοτούμενη διάβαση πεζών, με έντονη κυκλοφορία, η εφαρμογή του LSTM εμφάνισε **ακρίβεια 91,6%** και **recall 87,6%**, υπερτερώντας έναντι εναλλακτικών αρχιτεκτονικών νευρωνικών δικτύων, όπως το **Multilayer Perceptron (MLP)** – ένα απλό προωθητικό δίκτυο χωρίς μηχανισμούς χρονικής μνήμης – και το **Bidirectional LSTM (bi-LSTM)**, το οποίο επεξεργάζεται την ακολουθία δεδομένων τόσο προς τα εμπρός όσο και προς τα πίσω (Zhang et al., 2020).

Παρά τη συνολική υψηλή απόδοση, η εφαρμογή τέτοιων μοντέλων συνοδεύεται από την εμφάνιση **ψευδώς θετικών προβλέψεων**, δηλαδή περιπτώσεων κατά τις οποίες ο πεζός προβλέπεται εσφαλμένα ως παραβάτης, ενώ στην πραγματικότητα συμμορφώνεται με το σηματοδότη. Αυτές οι αστοχίες παρατηρούνται συνήθως όταν η κίνηση του πεζού προκαλεί αμφισημία — για παράδειγμα, κατά την προσωρινή προσέγγιση προς τη διάβαση ή λόγω της εγγύτητας σε ομάδα που τελικά παραβιάζει τον σηματοδότη. Η ύπαρξη τέτοιων παρανοήσεων υπογραμμίζει τη δυσκολία στην αυτόματη εκτίμηση της ανθρώπινης πρόθεσης και αναδεικνύει την ανάγκη ενσωμάτωσης συμπραζομένων παραμέτρων και πιο εξελιγμένων μεθόδων κατανόησης συμπεριφοράς (Zhang et al., 2020).

Στο πλαίσιο της κατανόησης των παραγόντων που επηρεάζουν την ασφάλεια των πεζών, η αυτοματοποιημένη ανάλυση της συμπεριφοράς τους σε πολυσύχναστες αστικές διασταυρώσεις παρέχει, εξίσου, σημαντικά ευρήματα. Η περίπτωση της διασταύρωσης Park Avenue South & East 28th Street στη Νέα Υόρκη αναλύθηκε με σύγχρονες τεχνικές υπολογιστικής όρασης, φανερώνοντας συστηματικά μοτίβα παραβάσεων και αλληλεπιδράσεων πεζών–οχημάτων (Hussein et al., 2015). Συγκεκριμένα, παρατηρήθηκε ότι περίπου το **30% των πεζών εμπλέκονταν σε παραβάσεις**, είτε διασχίζοντας τον δρόμο εκτός των επιτρεπόμενων χρονικών

φάσεων (**15,3%**, παραβίαση χρόνου) είτε έξω από τις διαγραμμίσεις της διάβασης (**17,9%**, παραβίαση χώρου) (Hussein et al., 2015).

Παράλληλα, μέσω μικροσκοπικής ανάλυσης της βάδισης, διαπιστώθηκε ότι οι παραβάτες πεζοί τείνουν να κινούνται ταχύτερα από τους μη παραβάτες. Η μέση **ταχύτητα βάδισης** των παραβατών ήταν αυξημένη κατά **5,5%**, γεγονός που συνδέθηκε κυρίως με μεγαλύτερο μήκος βήματος (+5,4%), ενώ η συχνότητα των βημάτων παρέμεινε πρακτικά σταθερή (Hussein et al., 2015). Επιπλέον, πεζοί που κινούνταν **μεμονωμένα** είχαν κατά μέσο όρο **9% υψηλότερη ταχύτητα** από όσους κινούνταν σε ομάδες (Hussein et al., 2015), γεγονός που ενδέχεται να αντανάκλα την απουσία κοινωνικής αναστολής ή τη μεγαλύτερη αυτοπεποίθηση κινήσεων. Τέλος, η μελέτη κατέγραψε **144 περιστατικά** πιθανών συγκρούσεων μεταξύ πεζών και οχημάτων, χρησιμοποιώντας το TTC (Time-to-Collision) ως μέτρο επικινδυνότητας. Από αυτά, τα **70 περιστατικά (48.6%) είχαν $TTC \leq 1$ δευτερόλεπτο**, υποδεικνύοντας ιδιαίτερα κρίσιμες καταστάσεις με υψηλή πιθανότητα σύγκρουσης (Hussein et al., 2015).

2.4.6 Αυτόματη ανίχνευση-πρόβλεψη συγκρούσεων μεταξύ οχημάτων και πεζών

Η πρόβλεψη τροχαίων συμβάντων σε σηματοδοτούμενες διασταυρώσεις έχει αναδειχθεί σε βασικό πεδίο έρευνας για την ενίσχυση της προληπτικής οδικής ασφάλειας. Η χρήση συστημάτων υπολογιστικής όρασης σε συνδυασμό με τεχνικές ανάλυσης χωρο-χρονικών αλληλεπιδράσεων επιτρέπει την εξαγωγή δυναμικών χαρακτηριστικών πεζών και οχημάτων από βιντεοσκοπημένα δεδομένα, συμβάλλοντας στη δημιουργία έγκαιρων μηχανισμών προειδοποίησης. Στο πλαίσιο αυτό, έχει προταθεί η εφαρμογή **νευρωνικών δικτύων τύπου Long Short-Term Memory (LSTM)**, τα οποία εξειδικεύονται στην επεξεργασία **χρονικών σειρών** και παρουσιάζουν αυξημένη ικανότητα στη μοντελοποίηση της χρονικής ακολουθίας κινήσεων σε διασταυρώσεις (Zhang et al., 2020).

Η προσέγγιση βασίζεται στον υπολογισμό του δείκτη Post-Encroachment Time (**PET**), δηλαδή του χρόνου που μεσολαβεί μεταξύ της διέλευσης ενός πεζού και ενός οχήματος από το ίδιο σημείο. Οι περιπτώσεις με PET **μικρότερο των 2 δευτερολέπτων** θεωρούνται εν δυνάμει συγκρουσιακές, επιτρέποντας τη χρήση των τιμών αυτών ως βάση ταξινόμησης των αλληλεπιδράσεων (Zhang et al., 2020). Το προτεινόμενο μοντέλο LSTM εκπαιδεύεται σε τέτοια περιστατικά και δύναται να προβλέψει συγκρούσεις **έως και 2 δευτερόλεπτα πριν** την εκτιμώμενη επικάλυψη, παρέχοντας κρίσιμο χρόνο απόκρισης σε υποστηρικτικά συστήματα οχημάτων (Zhang et al., 2020).

Σε εφαρμογή του μοντέλου, που έλαβε χώρα σε διασταύρωση του Πανεπιστημίου της Κεντρικής Φλόριντα, η **ακρίβεια** (accuracy) της πρόβλεψης ανήλθε σε ποσοστό **88,5%**. Όταν, δε, εφαρμόστηκε σε διαφορετική τοποθεσία, με νέα δεδομένα που αντιστοιχούσαν μόλις στο **30%** του συνόλου της εκπαίδευσης, η ακρίβεια παρέμεινε υψηλή, φθάνοντας το **84,9%** (Zhang et al., 2020). Η σταθερότητα αυτή υποδεικνύει ικανοποιητική ικανότητα γενίκευσης και ενισχύει τη δυνατότητα αξιοποίησης του συστήματος σε διαφορετικά αστικά περιβάλλοντα, στο πλαίσιο τεχνολογιών Συνδεδεμένων και Αυτόνομων Οχημάτων (CAVs) και έξυπνων υποδομών προειδοποίησης συγκρούσεων.

Συμπληρωματικά προς τις προσεγγίσεις που βασίζονται σε LSTM, έχει εξεταστεί και η αξιοποίηση συνδυαστικών δεικτών εγγύτητας κινδύνου (Surrogate Measures of Safety – SMOs) για την ενίσχυση της ακρίβειας πρόβλεψης. Πρόκειται για μετρήσιμες παραμέτρους, όπως το **Post-Encroachment Time (PET)** και το **minimum Time to Collision (min TTC)**, οι οποίες επιτρέπουν την **εκτίμηση της επικινδυνότητας μιας αλληλεπίδρασης χωρίς την ανάγκη ύπαρξης πραγματικού ατυχήματος** (Zhang et al., 2020).

Σε σχετική εφαρμογή σε δύο διασταυρώσεις της κομητείας Seminole στη Φλόριντα, εξήχθησαν οι τιμές PET και mTTC από βιντεοσκοπημένα δεδομένα πεζών και οχημάτων, στοχεύοντας στον εντοπισμό πιθανών συγκρούσεων. Για τον καθορισμό των κατωφλίων που διαχωρίζουν τις επικίνδυνες από τις μη επικίνδυνες αλληλεπιδράσεις, εφαρμόστηκε η Θεωρία Ακραίων Τιμών (Extreme Value Theory – EVT), η οποία επιτρέπει την ανάλυση των πιο ακραίων τιμών σε κατανομές δεδομένων. Συγκεκριμένα, η κατανομή GEV εφαρμόζεται στις τιμές PET και η κατανομή Pareto (GP) στις τιμές mTTC, προσδιορίζοντας ως κατώφλι το 5ο εκατοστημόριο κάθε κατανομής, δηλαδή την τιμή κάτω από την οποία εμπίπτει το 5% των μικρότερων παρατηρήσεων. Οι αλληλεπιδράσεις με τιμές κάτω από αυτό το όριο χαρακτηρίζονται ως περιστατικά σχεδόν ατυχήματος (near-miss conflicts) (Zhang et al., 2020).

Τα δεδομένα αυτά τροφοδότησαν ένα νευρωνικό δίκτυο τύπου **Gated Recurrent Unit (GRU)** — παραλλαγή του LSTM με απλούστερη εσωτερική δομή και χαμηλότερες υπολογιστικές απαιτήσεις — το οποίο είναι κατάλληλο για την επεξεργασία χρονικά εξαρτώμενων δεδομένων και την πρόβλεψη κινδύνου (Zhang et al., 2020). Το μοντέλο επέτυχε **ακρίβεια 87,8%** και **τιμή AUC (Area Under the Curve) 0,865**, υπερβαίνοντας την απόδοση του LSTM στο ίδιο σύνολο δεδομένων. Η συνδυαστική αξιοποίηση δεικτών SMOs, EVT και GRU αποδεικνύει τη δυνατότητα υλοποίησης αυτοματοποιημένων μηχανισμών ανίχνευσης κινδύνου σε πραγματικές συνθήκες, με εφαρμογή σε Συνδεδεμένα Οχήματα και Έξυπνες Υποδομές.

Η ανίχνευση πραγματοποιήθηκε μέσω του **YOLOv5**, μοντέλου ταχείας αναγνώρισης αντικειμένων, το οποίο μπορεί να εντοπίζει πεζούς, οχήματα και σχετικές οντότητες στο οπτικό πεδίο. Η υλοποίηση βασίστηκε στις βιβλιοθήκες **OpenCV** και **Torch**, αξιοποιώντας τεχνικές επεξεργασίας εικόνας και μηχανικής μάθησης (Radojicic & Dobrojevic, 2024).

Το σύστημα εμφάνισε **ποσοστό ανίχνευσης 81%** και **ψευδείς θετικές προβλέψεις 8,06%**, ενώ μειωμένη απόδοση παρατηρήθηκε σε περιβάλλοντα χαμηλού φωτισμού και συμφόρησης. Η σημασία τέτοιων εφαρμογών ενισχύεται σε οδικά δίκτυα με παλαιωμένα οχήματα: στη Βοσνία-Ερζεγοβίνη, **38% του στόλου υπερβαίνει τα 23 έτη**, καθιστώντας εξωτερικά συστήματα ανίχνευσης κρίσιμα για την πρόληψη τροχαίων συμβάντων (Radojicic & Dobrojevic, 2024).

2.4.7 Αυτόματη βελτιστοποίηση της κυκλοφορίας σε σηματοδοτούμενες διασταυρώσεις

Η υπολογιστική όραση αξιοποιείται για την αυτόματη προσαρμογή των σηματοδοτών σε διασταυρώσεις, με στόχο τη μείωση του **χρόνου αναμονής** και την **ενίσχυση της ασφάλειας πεζών** και οχημάτων. Σε πιλοτική εφαρμογή στην Τσελιαμπίνσκ της

Ρωσίας, αναπτύχθηκε και δοκιμάστηκε ένα σύστημα **δυναμικής ρύθμισης σηματοδότησης**, βασισμένο σε δύο διαμορφώσεις του νευρωνικού δικτύου YOLOv3 (για ανίχνευση οχημάτων και πεζών), εκπαιδευμένων στο πλαίσιο Darknet, και στον αλγόριθμο παρακολούθησης SORT, που επιτρέπει τη συνεχή ταυτοποίηση αντικειμένων σε πραγματικό χρόνο (Ibrieva et al., 2020).

Το σύστημα **κατένειμε δυναμικά** τον χρόνο του πράσινου σηματοδότη μεταξύ οχημάτων και πεζών, βάσει της παρουσίας πεζών σε καθορισμένες ζώνες ενδιαφέροντος και της εκτιμώμενης διάρκειας διέλευσης κάθε ουράς οχημάτων. Η εκτίμηση αυτή βασίστηκε σε προκαταρκτική ανάλυση που κατέγραψε τον απαιτούμενο χρόνο διέλευσης για ουρές διαφόρων μηκών, σε φάσεις χωρίς πεζούς, δημιουργώντας πίνακες πρόβλεψης ανάλογα με την κατάσταση.

Τα αποτελέσματα έδειξαν ότι το σύστημα πέτυχε **βελτιστοποίηση της κατανομής χρόνου**, επιτρέποντας σε μεγαλύτερο ποσοστό οχημάτων να στρίβουν χωρίς διακοπή, χωρίς να παραβιάζεται η ασφάλεια των πεζών. Σε περιπτώσεις χωρίς πεζούς, ολόκληρος ο διαθέσιμος χρόνος αποδιδόταν στα οχήματα, επιταχύνοντας τη ροή. Όταν εντοπίζονταν πεζοί, εξασφαλιζόταν τουλάχιστον ο ελάχιστος απαιτούμενος χρόνος για την ασφαλή διέλευσή τους (t_{min}), ενώ ο υπόλοιπος χρόνος αποδιδόταν στα οχήματα. Το σύστημα διατήρησε την ομαλή κυκλοφορία ακόμη και σε περιόδους κυκλοφοριακής αιχμής, αποφεύγοντας την ανάγκη για ανθρώπινη παρέμβαση ή σταθερές ρυθμίσεις.

Η μελέτη καταλήγει στο ότι τέτοια προσαρμοστικά συστήματα, με συνεχή επεξεργασία ροής βίντεο και ευέλικτη κατανομή φάσεων, μπορούν να συμβάλουν ουσιαστικά στη **μείωση των καθυστερήσεων και στην αύξηση της ασφάλειας σε σηματοδοτούμενες διασταυρώσεις**, με προϋπόθεση τη διαρκή ενημέρωση των δεδομένων λόγω των έντονων διακυμάνσεων στην κυκλοφορία κατά τη διάρκεια της ημέρας και ανά τοποθεσία (Ibrieva et al., 2020).

2.5 Σύνοψη βιβλιογραφίας

Η βιβλιογραφική επισκόπηση ανέδειξε βασικά **μοτίβα συμπεριφοράς** πεζών σε σηματοδοτούμενες διασταυρώσεις, τους **παράγοντες** που επηρεάζουν την παραβατικότητα και τις **τεχνολογικές μεθόδους** παρακολούθησης και πρόληψης κινδύνων. Μέσα από εμπειρικές μελέτες και σύγχρονες τεχνικές, αποτυπώθηκε η πολυπλοκότητα της συμπεριφοράς των πεζών και προέκυψαν χρήσιμα δεδομένα για τη βελτίωση της οδικής ασφάλειας.

Η βιβλιογραφία καταδεικνύει ότι η συμπεριφορά των πεζών σε σηματοδοτούμενες διασταυρώσεις επηρεάζεται από την **αντίληψη κινδύνου**, τις συνθήκες του **περιβάλλοντος** και επιμέρους ατομικά **χαρακτηριστικά**. Η παραβίαση του κόκκινου φωτός, με ποσοστά που αγγίζουν το 82%, συνδέεται άμεσα με την πιθανότητα σύγκρουσης και ενισχύεται όταν απουσιάζει η επιτήρηση. Η **ταχύτητα βάδισης** και οι ξαφνικές επιταχύνσεις διαφοροποιούνται ανάλογα με τη γεωμετρία της διάβασης, την κυκλοφοριακή πίεση και τη χρονική φάση του σηματοδότη, επηρεάζοντας τη δυναμική της διάβασης. Η **αβεβαιότητα** ως προς τον κατάλληλο χρόνο διέλευσης οδηγεί σε διστακτικότητα ή εγκατάλειψη, ενώ η **απόσπαση προσοχής** — κυρίως λόγω κινητών τηλεφώνων — σχετίζεται με μειωμένες αποστάσεις ασφαλείας και αυξημένη έκθεση στον κίνδυνο.

Η παραβατική συμπεριφορά των πεζών διαμορφώνεται από πλήθος παραγόντων που σχετίζονται με τη διαμόρφωση της **διάβασης**, τα ατομικά **χαρακτηριστικά** και το κοινωνικό και κυκλοφοριακό **περιβάλλον**. Η ύπαρξη σηματοδότη, η διάρκεια της φάσης κόκκινου και η ποιότητα της σήμανσης επιδρούν καθοριστικά στη συμμόρφωση, ενώ ελλείψεις στην υποδομή, όπως φθαρμένα σήματα ή απουσία διαγράμμισης, εντείνουν την παραβατικότητα. Το φύλο και η ηλικία επιδρούν διαφοροποιημένα, με τους άνδρες και τους νεότερους να εμφανίζουν υψηλότερα ποσοστά παραβάσεων, ιδίως σε αστικά περιβάλλοντα. Η κοινωνική επιρροή λειτουργεί καταλυτικά: η παρουσία συμμορφωμένων πεζών ενισχύει τη νομιμότητα, ενώ η θέα παραβατών αυξάνει την πιθανότητα μίμησης. Τέλος, η ένταση και η πολυπλοκότητα της κυκλοφορίας — ειδικά σε ώρες αιχμής ή σε δρόμους με πολλαπλές λωρίδες — λειτουργεί συχνά αποτρεπτικά ως προς τη μη συμμόρφωση, ενισχύοντας στρατηγικές αναμονής και ασφάλειας.

Η αξιοποίηση τεχνολογιών **υπολογιστικής όρασης** και τεχνητής νοημοσύνης έχει προσφέρει νέες δυνατότητες στην **αυτόματη** διάγνωση και πρόληψη κινδύνων σε σηματοδοτούμενες διασταυρώσεις. Μέθοδοι εξαγωγής τροχιών, ανίχνευσης πεζών και πρόβλεψης συγκρούσεων επιτρέπουν την αποδοτική καταγραφή και ανάλυση της ανθρώπινης συμπεριφοράς, παρέχοντας κρίσιμους δείκτες όπως οι TTC και PET για την αξιολόγηση της επικινδυνότητας. Νευρωνικά δίκτυα τύπου LSTM και GRU επιτυγχάνουν υψηλή ακρίβεια στην πρόβλεψη παραβατικής συμπεριφοράς ή εν δυνάμει συγκρούσεων, ενισχύοντας την ανάπτυξη προειδοποιητικών μηχανισμών σε αστικά περιβάλλοντα. Παράλληλα, συστήματα προσαρμοστικής σηματοδότησης, βασισμένα σε πραγματικό χρόνο ανίχνευση, βελτιστοποιούν τη ροή κυκλοφορίας και ενισχύουν την ασφάλεια, επιβεβαιώνοντας τη συμβολή των έξυπνων υποδομών στη διαχείριση της μικτής κυκλοφορίας πεζών και οχημάτων.

Κεφάλαιο 3: Θεωρητικό υπόβαθρο

3.1 Εισαγωγή

Το παρόν κεφάλαιο αποτελεί το θεωρητικό θεμέλιο της παρούσας διπλωματικής εργασίας, παρέχοντας το πλαίσιο κατανόησης των στατιστικών μεθόδων και μοντέλων που εφαρμόστηκαν. Η μελέτη της ανθρώπινης συμπεριφοράς στο αστικό περιβάλλον, και συγκεκριμένα της θέσης των πεζών ως προς τις διαβάσεις, απαιτεί την αξιοποίηση εργαλείων ποσοτικής ανάλυσης, ικανά να αποκαλύψουν πρότυπα, συσχετίσεις και προβλέψεις βάσει εμπειρικών δεδομένων.

Αρχικά, παρουσιάζονται **βασικές έννοιες** στατιστικής που διευκολύνουν την κατανόηση των τεχνικών που εφαρμόστηκαν, όπως οι τύποι μεταβλητών και τα βασικά μέτρα αξιοπιστίας και σημαντικότητας. Στη συνέχεια, αναλύονται οι δύο βασικές προσεγγίσεις της στατιστικής ανάλυσης – η **συσχέτιση** και η **πρόβλεψη** – καθώς και τα πιο αντιπροσωπευτικά εργαλεία κάθε προσέγγισης: ο συντελεστής συσχέτισης **Pearson** και το μοντέλο **παλινδρόμησης**.

Ειδική έμφαση δίνεται στη μέθοδο **Point-Biserial Correlation**, κατάλληλη για τη μελέτη σχέσεων μεταξύ δυαδικών και αριθμητικών μεταβλητών, καθώς και στο μοντέλο **Logistic Regression**, που χρησιμοποιείται για την πρόβλεψη πιθανοτήτων εμφάνισης ενός γεγονότος. Τέλος, περιγράφεται η μέθοδος **Random Forest**, ένα ισχυρό εργαλείο μηχανικής μάθησης που επιτρέπει την ταξινόμηση και αξιολόγηση πολύπλοκων σχέσεων μεταξύ μεταβλητών.

Η θεωρητική επισκόπηση του παρόντος κεφαλαίου αποτελεί αναγκαίο υπόβαθρο για την ερμηνεία των αποτελεσμάτων και την υποστήριξη των συμπερασμάτων της ανάλυσης που ακολουθεί.

3.2 Βασικές έννοιες στατιστικής

Στο πλαίσιο μιας στατιστικής μελέτης, τα δεδομένα που συλλέγονται προέρχονται είτε από το σύνολο του πληθυσμού είτε από ένα δείγμα που τον αντιπροσωπεύει. Τα δεδομένα αυτά οργανώνονται με τέτοιο τρόπο ώστε να είναι δυνατή η συστηματική ανάλυσή τους και η εξαγωγή τεκμηριωμένων συμπερασμάτων (Δαμιανού et al., 2003). Ένα από τα θεμελιώδη στοιχεία σε κάθε στατιστική προσέγγιση είναι η έννοια της μεταβλητής.

Μεταβλητές (variables) ονομάζονται τα χαρακτηριστικά που παρατηρούνται και μελετώνται στο πλαίσιο μιας έρευνας, είτε αυτά αφορούν άτομα, ομάδες, αντικείμενα ή γεγονότα. Ανάλογα με τη φύση των τιμών που λαμβάνουν, οι μεταβλητές διακρίνονται σε ποσοτικές και ποιοτικές. **Ποσοτικές** χαρακτηρίζονται οι μεταβλητές που λαμβάνουν αριθμητικές τιμές και επιτρέπουν μαθηματικούς υπολογισμούς. Ανάλογα με την κατανομή των τιμών τους, οι ποσοτικές μεταβλητές χωρίζονται περαιτέρω σε **συνεχείς και διακριτές**. Μία μεταβλητή είναι συνεχής όταν μπορεί, θεωρητικά, να λάβει οποιαδήποτε τιμή εντός ενός καθορισμένου διαστήματος, όπως για παράδειγμα το ύψος ή η ταχύτητα. Αντίθετα, διακριτές είναι οι μεταβλητές που λαμβάνουν συγκεκριμένες, συνήθως ακέραιες, τιμές, όπως ο αριθμός τροχαίων ατυχημάτων σε μια περίοδο (Moore, McCabe&Craig, 2017). Οι **ποιοτικές** ή

κατηγορικές μεταβλητές, από την άλλη, χρησιμοποιούνται για να εκφράσουν χαρακτηριστικά που δεν έχουν αριθμητική διάσταση αλλά ταξινομούνται σε κατηγορίες, όπως για παράδειγμα το φύλο ή ο τύπος οχήματος. Ειδική περίπτωση κατηγορικών μεταβλητών αποτελούν οι **δυσδικές** (ή διχοτομικές) μεταβλητές, οι οποίες λαμβάνουν μόνο δύο τιμές, τυπικά 0 και 1, και χρησιμοποιούνται συχνά σε αναλύσεις κατηγοριοποίησης (Agresti & Finlay, 2009).

Η αξιολόγηση της αξιοπιστίας των αποτελεσμάτων που προκύπτουν από τη στατιστική ανάλυση είναι κρίσιμη για την εγκυρότητα των συμπερασμάτων. Στο πλαίσιο αυτό, ιδιαίτερη σημασία έχουν δύο αλληλένδετες αλλά αντίστροφα σχετιζόμενες έννοιες: το επίπεδο εμπιστοσύνης και το επίπεδο στατιστικής σημαντικότητας. Το **επίπεδο εμπιστοσύνης (confidence level)** εκφράζει την πιθανότητα μια εκτίμηση να αντανakλά την πραγματική τιμή στον πληθυσμό. Συνήθως χρησιμοποιούνται επίπεδα εμπιστοσύνης της τάξης του 90%, 95% ή 99%, με το 95% να αποτελεί το πιο συχνά χρησιμοποιούμενο στη στατιστική ανάλυση. Για παράδειγμα, ένα επίπεδο εμπιστοσύνης 95% σημαίνει ότι, αν η ίδια μελέτη επαναλαμβανόταν πολλές φορές, σε 95 από τις 100 περιπτώσεις η εκτίμηση που προκύπτει από το δείγμα θα ήταν κοντά στην πραγματική τιμή του πληθυσμού (Triola, 2018). Συμπληρωματικά, το **επίπεδο σημαντικότητας (significance level)** αφορά την πιθανότητα να απορριφθεί εσφαλμένα μια σωστή μηδενική υπόθεση. Το πιο συχνό όριο είναι το 5% (ή 0,05), που υποδηλώνει αποδοχή ενός ποσοστού σφάλματος 5% στην κρίση μας ότι υπάρχει στατιστικά σημαντική σχέση ή διαφορά. Επομένως, όταν η τιμή p (p -value) που προκύπτει από την ανάλυση είναι μικρότερη από το επιλεγμένο επίπεδο σημαντικότητας, απορρίπτεται η μηδενική υπόθεση και το αποτέλεσμα θεωρείται στατιστικά σημαντικό (Field, 2018).

Ένα ακόμη βασικό στοιχείο στη στατιστική ανάλυση αποτελεί ο **συντελεστής συσχέτισης** (correlation coefficient), ο οποίος αποτυπώνει τον βαθμό και την κατεύθυνση της σχέσης μεταξύ δύο μεταβλητών. Η τιμή του συντελεστή συσχέτισης κυμαίνεται θεωρητικά από -1 έως +1. Θετικές τιμές υποδηλώνουν ότι οι δύο μεταβλητές αυξάνονται ταυτόχρονα, ενώ αρνητικές τιμές υποδηλώνουν ότι καθώς η μία αυξάνεται, η άλλη μειώνεται. Τιμές κοντά στο μηδέν υποδηλώνουν ασθενή ή ανύπαρκτη γραμμική συσχέτιση, ενώ τιμές κοντά στο ± 1 δηλώνουν ισχυρή σχέση. Είναι σημαντικό να τονιστεί ότι η ερμηνεία του μεγέθους του συντελεστή εξαρτάται και από το πεδίο της έρευνας, καθώς διαφορετικά αντικείμενα μπορεί να θεωρούν διαφορετικά όρια για "ισχυρή" ή "ασθενή" συσχέτιση (Hinkle, Wiersma & Jurs, 2003). Η κατεύθυνση της σχέσης (θετική ή αρνητική) προσδιορίζεται από το πρόσημο του συντελεστή, ενώ η έντασή της από την απόλυτη τιμή του.

Η κατανόηση όλων των παραπάνω εννοιών είναι απαραίτητη για την ερμηνεία των αποτελεσμάτων που παρουσιάζονται στα επόμενα υποκεφάλαια, καθώς αποτελούν τη βάση πάνω στην οποία στηρίζονται οι στατιστικές μέθοδοι που εφαρμόζονται στο πλαίσιο της παρούσας εργασίας.

3.3 Συσχέτιση και πρόβλεψη: Δύο βασικές προσεγγίσεις στατιστικής ανάλυσης

Οι στατιστικές μέθοδοι που χρησιμοποιούνται στις ερευνητικές εφαρμογές μπορούν να ταξινομηθούν σε δύο κύριες κατηγορίες, ανάλογα με τον σκοπό της ανάλυσης: αφενός σε μεθόδους που αποσκοπούν στη διερεύνηση της ύπαρξης και της φύσης

της σχέσης μεταξύ μεταβλητών (**συσχετιστικές μέθοδοι**), και αφετέρου σε μεθόδους που στοχεύουν στην πρόβλεψη της τιμής ή της κατηγορίας μιας εξαρτημένης μεταβλητής βάσει ενός συνόλου ανεξάρτητων μεταβλητών (**προβλεπτικά ή επεξηγηματικά μοντέλα**) (Field, 2018; Agresti & Finlay, 2009).

Οι **συσχετιστικές μέθοδοι** επιτρέπουν την ποσοτική αποτύπωση της στατιστικής σχέσης μεταξύ δύο μεταβλητών, χωρίς απαραίτητα να τεκμηριώνουν αιτιακή σύνδεση. Τέτοιου είδους μέθοδοι χρησιμοποιούνται συχνά σε πρώιμα στάδια μιας ανάλυσης ή όταν ο στόχος είναι η εξερεύνηση συσχετίσεων και όχι η πρόβλεψη. Η ερμηνεία τους εστιάζει στη δύναμη και την κατεύθυνση της σχέσης μεταξύ των μεταβλητών (Moore, McCabe & Craig, 2017).

Αντίθετα, τα **μοντέλα πρόβλεψης** ή ταξινόμησης έχουν ως στόχο την ανάπτυξη μαθηματικής σχέσης ικανής να προβλέψει την τιμή μιας εξαρτημένης μεταβλητής με βάση συγκεκριμένα χαρακτηριστικά (ανεξάρτητες μεταβλητές). Σε αυτά τα πλαίσια, ο ερευνητής ενδιαφέρεται όχι μόνο για την ύπαρξη σχέσης, αλλά και για τη δυνατότητα πρόβλεψης ή γενίκευσης σε νέες περιπτώσεις. Τέτοιες προσεγγίσεις εντάσσονται στην επόπτευση μοντέλων μάθησης (supervised learning) και χρησιμοποιούνται ευρέως στην ανάλυση συμπεριφοράς, την κατηγοριοποίηση, τη διάγνωση ή τη λήψη αποφάσεων (James, Witten, Hastie & Tibshirani, 2021).

Η επιλογή της κατάλληλης μεθόδου εξαρτάται από τον ερευνητικό στόχο, τη φύση της εξαρτημένης μεταβλητής (συνεχής, δυαδική, κατηγορική), καθώς και τις ιδιότητες των ανεξάρτητων μεταβλητών. Η διάκριση μεταξύ συσχετιστικών και προβλεπτικών προσεγγίσεων είναι καθοριστική για την ορθή εφαρμογή των στατιστικών τεχνικών, καθώς και για την ερμηνεία και την εγκυρότητα των αποτελεσμάτων που προκύπτουν από αυτές.

3.4 Συντελεστής συσχέτισης Pearson: Θεωρητική βάση και επεκτάσεις

Η συσχέτιση αποτελεί βασική έννοια της στατιστικής ανάλυσης και αναφέρεται στην εκτίμηση της γραμμικής σχέσης μεταξύ δύο μεταβλητών (Moore, McCabe & Craig, 2017). Η πιο ευρέως χρησιμοποιούμενη μορφή συσχέτισης είναι ο συντελεστής του **Pearson**, ο οποίος εφαρμόζεται όταν και οι δύο μεταβλητές είναι ποσοτικές και **συνεχείς** (Field, 2018). Η ερμηνεία του βασίζεται τόσο στην κατεύθυνση της σχέσης (θετική ή αρνητική) όσο και στην έντασή της. Η μέτρηση αυτή δεν προϋποθέτει αιτιακή σχέση, αλλά αποτυπώνει τον βαθμό στον οποίο οι μεταβλητές μεταβάλλονται ταυτόχρονα (Agresti & Finlay, 2009).

Στις περιπτώσεις όπου μία από τις δύο μεταβλητές είναι δυαδική, έχει αναπτυχθεί η μέθοδος **Point-Biserial Correlation**, η οποία αποτελεί ειδική εφαρμογή της λογικής του Pearson σε διαφορετικού τύπου δεδομένα (Hinkle, Wiersma & Jurs, 2003). Η συγκεκριμένη προσέγγιση είναι ιδιαίτερα κατάλληλη όταν ο ερευνητικός στόχος είναι η μελέτη της σχέσης ανάμεσα σε μια κατηγορική (δικοτομική) εξαρτημένη μεταβλητή και συνεχείς ανεξάρτητες μεταβλητές, όπως συμβαίνει και στην παρούσα εργασία, όπου η μέθοδος αυτή αξιοποιήθηκε στο πλαίσιο της στατιστικής διερεύνησης των δεδομένων.

3.5 Μοντέλα παλινδρόμησης: Γραμμική και Λογιστική προσέγγιση στην ανάλυση σχέσεων

Η **παλινδρόμηση** αποτελεί μία από τις βασικότερες μεθόδους στατιστικής ανάλυσης, καθώς επιτρέπει τη μοντελοποίηση της σχέσης μεταξύ μιας εξαρτημένης μεταβλητής και μίας ή περισσότερων ανεξάρτητων μεταβλητών (Agresti & Finlay, 2009). Η θεμελιώδης ιδέα στηρίζεται στην προσπάθεια πρόβλεψης ή ερμηνείας της εξαρτημένης μεταβλητής, λαμβάνοντας υπόψη τις μεταβολές που παρατηρούνται στις μεταβλητές εισόδου (Field, 2018). Ανάλογα με τη φύση των μεταβλητών και τη γραμμικότητα της σχέσης, υπάρχουν διαφορετικοί τύποι παλινδρόμησης που καλύπτουν διαφορετικές ερευνητικές ανάγκες (James et al., 2021).

Η πιο βασική μορφή είναι η **Γραμμική Παλινδρόμηση**, η οποία εφαρμόζεται όταν η εξαρτημένη μεταβλητή είναι **συνεχής** και η σχέση που υποτίθεται μεταξύ των μεταβλητών είναι γραμμική. Σε αυτή την περίπτωση, το μοντέλο επιδιώκει να περιγράψει τη μεταβολή της εξαρτημένης μεταβλητής ως συνάρτηση των ανεξάρτητων μεταβλητών μέσω μιας ευθείας γραμμής (Moore, McCabe & Craig, 2017).

Ωστόσο, όταν η εξαρτημένη μεταβλητή είναι **δυναδική**, δηλαδή περιγράφει δύο διακριτές κατηγορίες, η γραμμική παλινδρόμηση δεν είναι κατάλληλη. Στην περίπτωση αυτή χρησιμοποιείται η **Λογιστική Παλινδρόμηση**, η οποία επιτρέπει τη μοντελοποίηση της πιθανότητας μιας έκβασης με βάση τις τιμές των ανεξάρτητων μεταβλητών, μεταφράζοντας το πρόβλημα πρόβλεψης σε πλαίσιο κατηγοριοποίησης (James et al., 2021· Field, 2018).

Η διάκριση μεταξύ των δύο αυτών μορφών παλινδρόμησης είναι ουσιώδης, καθώς κάθε μία βασίζεται σε διαφορετικές στατιστικές υποθέσεις και εξυπηρετεί διαφορετικά είδη εξαρτημένων μεταβλητών. Στην παρούσα εργασία, εφαρμόστηκε η Λογιστική Παλινδρόμηση, καθώς η εξαρτημένη μεταβλητή έχει δυναδική μορφή. Παράλληλα, έγινε υποτυπώδης αναφορά και στη Γραμμική Παλινδρόμηση, η οποία αν και δεν χρησιμοποιείται στην ανάλυση, παρουσιάζεται για την πληρέστερη κατανόηση του εννοιολογικού πλαισίου της παλινδρόμησης και της διαφοροποίησης μεταξύ συνεχών και κατηγορικών αποτελεσμάτων.

3.6 Μοντέλο Point-Biserial Correlation

3.6.1 Εισαγωγή

Στο πλαίσιο της παρούσας διπλωματικής εργασίας, για τη διερεύνηση της σχέσης μεταξύ της θέσης ενός πεζού ως προς τη διάβαση και επιμέρους χαρακτηριστικών της κίνησης, εφαρμόστηκε η στατιστική μέθοδος Point-Biserial Correlation. Η συγκεκριμένη τεχνική επιλέχθηκε λόγω της φύσης των μεταβλητών: η **εξαρτημένη** μεταβλητή είναι **δυναδική**, ενώ οι **ανεξάρτητες** είναι είτε **αριθμητικές**, είτε **κατηγορικές**, οι οποίες μετατράπηκαν σε αριθμητικές, και αφορούν τόσο τους πεζούς όσο και τα οχήματα. Τα δεδομένα συλλέχθηκαν από βιντεοσκοπικό υλικό, το οποίο αναλύθηκε με τη χρήση MATLAB και Python, και στη συνέχεια υποβλήθηκαν σε στατιστική επεξεργασία για την εξαγωγή συσχετίσεων. Η παρούσα ενότητα

περιγράφει αναλυτικά τη θεωρητική βάση της μεθόδου και την εφαρμογή της στα δεδομένα του συστήματος.

3.6.2 Θεωρητικό υπόβαθρο μεθόδου

Το μοντέλο **Point-Biserial Correlation** αποτελεί ειδική περίπτωση του συντελεστή συσχέτισης **Pearson** και χρησιμοποιείται για την αξιολόγηση της γραμμικής σχέσης μεταξύ μιας **δихοτομικής εξαρτημένης μεταβλητής** και μίας **συνεχούς ανεξάρτητης μεταβλητής** (Cohen et al., 2003; Bonett, 2020). Ο συγκεκριμένος συντελεστής είναι κατάλληλος όταν η εξαρτημένη μεταβλητή διαχωρίζει τις παρατηρήσεις σε δύο κατηγορίες, ενώ η ανεξάρτητη είναι αριθμητική (Kornbrot, 2007).

Στην παρούσα εργασία, ως εξαρτημένη μεταβλητή θεωρείται το αν ένας πεζός βρίσκεται εντός ή εκτός διάβασης καθώς και η κατάσταση της διέλευσης του, ενώ ως ανεξάρτητες μεταβλητές χρησιμοποιούνται διάφορες αριθμητικές και κατηγορικές μεταβλητές που αντιπροσωπεύουν τόσο τους πεζούς όσο και τα οχήματα. Η μεθοδολογία εφαρμόζεται με στόχο τον εντοπισμό στατιστικά σημαντικών συσχέτισεων μεταξύ της θέσης του πεζού και των υπό εξέταση μεταβλητών.

Η εξίσωση του συντελεστή Point-Biserial είναι η εξής (Glass, 1966):

$$r_{pb} = \frac{\bar{X}_1 - \bar{X}_0}{s} \cdot \sqrt{\frac{n_1 n_0}{n(n-1)}} \quad (\text{Εξ. 1})$$

όπου:

- $\bar{X}_1, \bar{X}_0$: οι μέσοι όροι της αριθμητικής μεταβλητής για τις δύο κατηγορίες της δυαδικής μεταβλητής,
- s : η συνολική τυπική απόκλιση της αριθμητικής μεταβλητής,
- n_1, n_0 : τα πλήθη των παρατηρήσεων σε κάθε κατηγορία,
- n : το συνολικό πλήθος παρατηρήσεων.

3.6.3 Κριτήρια αξιολόγησης και αποδοχής συσχέτισεων

Η αξιολόγηση των αποτελεσμάτων που προκύπτουν από την εφαρμογή της Point-Biserial Correlation είναι απαραίτητη για την εξαγωγή αξιόπιστων και επιστημονικά έγκυρων συμπερασμάτων. Αν και πρόκειται για ένα σχετικά απλό μέτρο συσχέτισης, υπάρχουν βασικά κριτήρια που πρέπει να πληρούνται για να θεωρηθεί ότι μια σχέση μεταξύ δύο μεταβλητών είναι σημαντική και χρήσιμη για την ερμηνεία των φαινομένων που μελετώνται.

Ερμηνεία του συντελεστή συσχέτισης

Η Point-Biserial Correlation χρησιμοποιείται για να μελετηθεί η σχέση μεταξύ μιας **δυαδικής μεταβλητής** (που παίρνει τιμές 0 και 1) και μιας **συνεχούς αριθμητικής μεταβλητής**. Ο υπολογισμός της δίνει έναν συντελεστή συσχέτισης, συμβολιζόμενο ως r_{pb} , ο οποίος κυμαίνεται από -1 έως +1 (Cohen et al., 2003; Glass, 1966).

Το **πρόσημο** του r_{pb} δείχνει σε ποια κατηγορία της δυαδικής μεταβλητής αντιστοιχούν υψηλότερες τιμές της αριθμητικής μεταβλητής:

- Αν $r_{pb} > 0$, τότε η μέση τιμή της αριθμητικής μεταβλητής είναι μεγαλύτερη όταν η δυαδική μεταβλητή έχει τιμή 1,
- Αν $r_{pb} < 0$, η μέση τιμή είναι μεγαλύτερη όταν η δυαδική μεταβλητή έχει τιμή 0 (Kornbrot, 2007).

Η ερμηνεία της κατεύθυνσης της σχέσης πρέπει να λαμβάνει υπόψη το θεωρητικό πλαίσιο της έρευνας, καθώς και το πώς έχουν κωδικοποιηθεί οι τιμές της δυαδικής μεταβλητής. Τυχόν αντίστροφα αποτελέσματα σε σχέση με τις αρχικές υποθέσεις δεν σημαίνουν απαραίτητα ότι η σχέση είναι λανθασμένη, αλλά πρέπει να διερευνηθούν προσεκτικά.

Στατιστική σημαντικότητα συσχέτισης

Η **στατιστική σημαντικότητα** είναι το βασικό κριτήριο που χρησιμοποιούμε για να κρίνουμε αν η συσχέτιση που παρατηρούμε μεταξύ δύο μεταβλητών είναι πραγματική ή απλώς οφείλεται σε τύχη (Cohen et al., 2003; Field, 2013). Στην περίπτωση της **Point-Biserial Correlation**, η οποία μελετά τη γραμμική σχέση ανάμεσα σε μία **δυαδική μεταβλητή** (με δύο τιμές, π.χ. 0 και 1) και μία **αριθμητική μεταβλητή**, η έννοια της σημαντικότητας μας επιτρέπει να αξιολογήσουμε κατά πόσο ο υπολογιζόμενος συντελεστής συσχέτισης (r_{pb}) είναι αξιόπιστος (Kornbrot, 2007).

Αρχικά, υπολογίζεται ο συντελεστής r_{pb} , ο οποίος εκφράζει τον βαθμό συσχέτισης μεταξύ των δύο μεταβλητών. Ωστόσο, από μόνος του δεν αρκεί για να πούμε αν η συσχέτιση είναι στατιστικά σημαντική. Για να το κρίνουμε αυτό, προχωρούμε στον υπολογισμό της στατιστικής **τιμής t**, χρησιμοποιώντας τον εξής τύπο (Glass, 1966):

$$t = r_{pb} \cdot \sqrt{\left(\frac{(n-2)}{(1-r_{pb}^2)} \right)} \quad (\text{Εξ. 2})$$

όπου:

- r_{pb} : ο συντελεστής Point-Biserial,
- n : το πλήθος των παρατηρήσεων του δείγματος.

Η τιμή t που προκύπτει ακολουθεί την **t-κατανομή του Student**, η οποία χρησιμοποιείται για την ανάλυση δειγμάτων μικρού ή μεσαίου μεγέθους (Bonett, 2020). Συγκεκριμένα, η τιμή t συγκρίνεται με την t -κατανομή με **($n - 2$)** βαθμούς ελευθερίας.

Από αυτή τη σύγκριση, υπολογίζεται το **p-value**, το οποίο αποτελεί τη πιθανότητα να παρατηρηθεί μία τιμή t ίση ή πιο ακραία από την υπολογισμένη, αν η μηδενική υπόθεση (H_0) είναι αληθής. Η μηδενική υπόθεση στην περίπτωση της Point-Biserial Correlation δηλώνει ότι **δεν υπάρχει συσχέτιση** μεταξύ των δύο μεταβλητών, δηλαδή $r_{pb} = 0$ (Field, 2013).

Ο υπολογισμός του p-value βασίζεται αποκλειστικά στη στατιστική τιμή t και στους βαθμούς ελευθερίας. Μπορεί να πραγματοποιηθεί με δύο τρόπους:

- Μέσω στατιστικού λογισμικού (όπως SPSS, R, Python), το οποίο υπολογίζει το p-value αυτόματα με βάση την t -κατανομή,

- Χειροκίνητα μέσω πινάκων t-κατανομής, όπου εντοπίζεται η τιμή t στη σειρά που αντιστοιχεί στους βαθμούς ελευθερίας ($n-2$) και εκτιμάται η αντίστοιχη πιθανότητα.

Αν το **p-value** που προκύπτει είναι μικρότερο από το προκαθορισμένο επίπεδο **σημαντικότητας** (συνήθως $\alpha = 0.05$), τότε απορρίπτεται η μηδενική υπόθεση και η συσχέτιση θεωρείται **στατιστικά σημαντική** (Cohen et al., 2003). Αυτό σημαίνει ότι υπάρχει ισχυρή ένδειξη ότι η συσχέτιση που παρατηρείται στο δείγμα αντανάκλα μια πραγματική σχέση στον πληθυσμό και είναι πολύ απίθανο να έχει προκύψει λόγω τύχης.

Αντίθετα, αν το p-value είναι μεγαλύτερο από 0.05, τότε **δεν απορρίπτεται η μηδενική υπόθεση**, πράγμα που σημαίνει ότι δεν υπάρχουν επαρκή στατιστικά στοιχεία για να υποστηρίξουμε την ύπαρξη συσχέτισης μεταξύ των μεταβλητών (Bonett, 2020).

Ποιότητα μοντέλου

Η Point-Biserial Correlation δεν αποτελεί μαθηματικό μοντέλο πρόβλεψης, αλλά στατιστικό δείκτη γραμμικής συσχέτισης (Cohen et al., 2003). Συνεπώς, δεν χρησιμοποιούνται μέτρα προσαρμογής όπως ο **συντελεστής προσδιορισμού R^2** , που συναντάται σε γραμμικά ή λογιστικά μοντέλα (Field, 2013).

Ωστόσο, η ένταση της συσχέτισης μπορεί να εκτιμηθεί από την απόλυτη τιμή του r_{pb} . Ενδεικτικά (Cohen, 1988):

- $|r| < 0.10 \rightarrow$ αμελητέα συσχέτιση
- $0.10 \leq |r| < 0.30 \rightarrow$ ασθενής
- $0.30 \leq |r| < 0.50 \rightarrow$ μέτρια
- $|r| \geq 0.50 \rightarrow$ ισχυρή

Επιπλέον, το r^2_{pb} μπορεί να χρησιμοποιηθεί ερμηνευτικά ως ποσοστό της διακύμανσης της αριθμητικής μεταβλητής που εξηγείται από τη δυαδική (McGrath & Meyer, 2006). Παρά το ότι δεν έχει την ίδια χρήση όπως το R^2 σε μοντέλα παλινδρόμησης, μπορεί να προσφέρει επιπλέον κατανόηση της δύναμης της σχέσης (Bonett, 2020).

Σφάλμα

Η Point-Biserial Correlation δεν βασίζεται σε εξισώσεις πρόβλεψης και δεν συνοδεύεται από υπόλοιπα σφάλματος (residuals), όπως συμβαίνει σε μοντέλα παλινδρόμησης (Field, 2013; Cohen et al., 2003). Συνεπώς, **δεν εφαρμόζονται υποθέσεις** που σχετίζονται με την κατανομή του σφάλματος, τη σταθερότητα της διασποράς ή την ανεξαρτησία των σφαλμάτων (McGrath & Meyer, 2006).

Η εγκυρότητα της ανάλυσης βασίζεται κυρίως:

- στην κατάλληλη χρήση του μοντέλου για το είδος των μεταβλητών (Kornbrot, 2007),
- στον έλεγχο της στατιστικής σημαντικότητας (p-value) (Bonett, 2020),

- και στο επαρκές μέγεθος δείγματος (Cohen, 1988).

Η ερμηνεία των αποτελεσμάτων πρέπει να γίνεται στο πλαίσιο της συνολικής έρευνας και όχι μεμονωμένα ώστε να αποφεύγονται υπεραπλουστεύσεις ή λανθασμένα συμπεράσματα (Field, 2013).

3.6.4 Εξήγηση αποτελεσμάτων

Μετά την αξιολόγηση της στατιστικής σημαντικότητας των συσχετίσεων, ακολουθεί η ερμηνεία των αποτελεσμάτων που προέκυψαν από την ανάλυση Point-Biserial Correlation. Για να επιτευχθεί αυτό, εξετάζονται οι εξής παράμετροι:

Εξήγηση σχέσεων εξαρτημένης και ανεξάρτητων μεταβλητών

Η σχέση μεταξύ της δυαδικής εξαρτημένης μεταβλητής (π.χ. αν ο πεζός βρίσκεται εντός ή εκτός διάβασης) και κάθε αριθμητικής ανεξάρτητης μεταβλητής (όπως ταχύτητα, απόσταση, κατεύθυνση κίνησης κ.λ.π) εκφράζεται με τον συντελεστή συσχέτισης r_{pb} (Kornbrot, 2007; Glass, 1966).

Το **πρόσημο** του συντελεστή καθορίζει την κατεύθυνση της σχέσης (Field, 2013):

- **Θετικό πρόσημο** σημαίνει ότι οι τιμές της αριθμητικής μεταβλητής είναι υψηλότερες στις περιπτώσεις όπου η εξαρτημένη μεταβλητή έχει τιμή 1 (π.χ. εντός διάβασης).
- **Αρνητικό πρόσημο** σημαίνει ότι οι τιμές είναι υψηλότερες στις περιπτώσεις με τιμή 0 (π.χ. εκτός διάβασης).

Η **ένταση της συσχέτισης** κρίνεται από την απόλυτη τιμή του r_{pb} , σύμφωνα με τις εξής κατηγορίες (Cohen, 1988):

- $|r| < 0.10 \rightarrow$ αμελητέα συσχέτιση
- $0.10 \leq |r| < 0.30 \rightarrow$ ασθενής
- $0.30 \leq |r| < 0.50 \rightarrow$ μέτρια
- $|r| \geq 0.50 \rightarrow$ ισχυρή

Επιβεβαίωση ή απόρριψη της διεθνούς βιβλιογραφίας

Τα αποτελέσματα της ανάλυσης συσχετίσεων συγκρίνονται με ευρήματα που παρουσιάστηκαν στη βιβλιογραφική ανασκόπηση, ώστε να διαπιστωθεί αν οι σχέσεις που εντοπίστηκαν συμφωνούν ή αποκλίνουν από προηγούμενες μελέτες. Η σύγκριση αυτή συμβάλλει στην αξιολόγηση της αξιοπιστίας των ευρημάτων, ενισχύει ή αμφισβητεί υπάρχουσες θεωρίες και αναδεικνύει πιθανά νέα μοτίβα συμπεριφοράς των πεζών στο οδικό περιβάλλον.

Τυχόν αποκλίσεις δεν σημαίνουν απαραίτητα ότι τα αποτελέσματα είναι λανθασμένα, αλλά ενδέχεται να οφείλονται σε διαφοροποιήσεις στο δείγμα, στο πλαίσιο της μελέτης ή στη μέθοδο συλλογής των δεδομένων (π.χ. παρατήρηση μέσω βίντεο και επεξεργασία με MATLAB και στη συνέχεια στατιστική επεξεργασία με χρήση Python, όπως εφαρμόστηκε στην παρούσα εργασία).

3.7 Μοντέλο Λογιστικής Παλινδρόμησης

3.7.1 Εισαγωγή

Στο πλαίσιο της παρούσας διπλωματικής εργασίας, για τη διερεύνηση της επίδρασης επιμέρους χαρακτηριστικών της κίνησης στην πιθανότητα παρουσίας ενός πεζού εντός ή εκτός διάβασης αλλά και για τον προσδιορισμό της κατάστασης διέλευσης του, εφαρμόστηκε η στατιστική μέθοδος της **Λογιστικής Παλινδρόμησης**. Η συγκεκριμένη τεχνική επιλέχθηκε λόγω της καταλληλότητάς της για την πρόβλεψη **δυναμικών εκβάσεων**, με βάση **συνεχείς ή/και κατηγορικές** ανεξάρτητες μεταβλητές. Τα δεδομένα προήλθαν από ανάλυση βιντεοσκοπημένου υλικού, μέσω εργαλείων MATLAB, και υποβλήθηκαν σε επεξεργασία με χρήση Python προκειμένου να αξιολογηθεί η πιθανότητα εμφάνισης του φαινομένου υπό συγκεκριμένες συνθήκες. Η παρούσα ενότητα παρουσιάζει τη θεωρητική θεμελίωση της μεθόδου και την υλοποίησή της επί των συλλεχθέντων δεδομένων.

3.7.2 Θεωρητικό υπόβαθρο μεθόδου

Το μοντέλο **Λογιστικής Παλινδρόμησης** χρησιμοποιείται για τη **μοντελοποίηση της πιθανότητας εμφάνισης ενός δυαδικού αποτελέσματος** ως συνάρτηση μίας ή περισσότερων ανεξάρτητων μεταβλητών (Hosmer et al., 2013). Δηλαδή, εφαρμόζεται σε περιπτώσεις όπου το αποτέλεσμα μπορεί να πάρει μόνο δύο τιμές, όπως «εντός» ή «εκτός» διάβασης, «ναι» ή «όχι», «0» ή «1».

Η Λογιστική Παλινδρόμηση ανήκει στην κατηγορία της **γραμμικής παλινδρόμησης**, ωστόσο είναι ειδικά σχεδιασμένη για καταστάσεις όπου η εξαρτημένη μεταβλητή δεν είναι συνεχής, αλλά **δυναμική** (Menard, 2002). Αντί να προβλέπει έναν αριθμό, προβλέπει **την πιθανότητα** να συμβεί ένα γεγονός, με τιμές μεταξύ 0 και 1.

Αρχικά, υπολογίζεται ένας **γραμμικός συνδυασμός** των ανεξάρτητων μεταβλητών, δηλαδή ένα σταθμισμένο άθροισμα της μορφής:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (\text{Εξ. 3})$$

όπου:

- x_i : οι ανεξάρτητες μεταβλητές,
- β_i : οι συντελεστές παλινδρόμησης που εκτιμώνται από το μοντέλο και εκφράζουν τη βαρύτητα κάθε μεταβλητής στο τελικό αποτέλεσμα,
- β_0 : ο σταθερός όρος (intercept).

Η τιμή **z** μπορεί να πάρει οποιαδήποτε πραγματική τιμή, θετική ή αρνητική. Όμως, επειδή ο στόχος είναι να υπολογιστεί **μία πιθανότητα**, η τιμή αυτή μετατρέπεται μέσω της **συνάρτησης λογιστικής (sigmoid)** σε τιμή εντός του διαστήματος [0, 1] (Agresti, 2018).

Η τελική πιθανότητα ότι το αποτέλεσμα είναι ίσο με 1, δηλαδή ότι το γεγονός συμβαίνει, δίνεται από την εξίσωση:

$$P(Y = 1|X) = \frac{1}{1+e^{-(\beta_0+\beta_1X_1+\beta_2X_2+\dots+\beta_nX_n)}} \quad (\text{Εξ. 4})$$

όπου:

- $P(Y=1|X)$: η πιθανότητα εμφάνισης του γεγονότος (π.χ. πεζός εντός διάβασης),
- β_0 : σταθερός όρος (intercept),
- β_i : συντελεστής παλινδρόμησης για την ανεξάρτητη μεταβλητή X_i ,
- X_i : ανεξάρτητες μεταβλητές

Η συνάρτηση αυτή εξασφαλίζει ότι όσο αυξάνεται το z , τόσο η πιθανότητα προσεγγίζει το 1, ενώ όσο μειώνεται το z , η πιθανότητα προσεγγίζει το 0. Το μοντέλο επομένως δεν προβλέπει απευθείας κατηγορία, αλλά εκφράζει την **πιθανότητα υπαγωγής σε μια κατηγορία**, επιτρέποντας πιο ευέλικτη και ερμηνεύσιμη πρόβλεψη (Long, 1997).

Οι συντελεστές β_i του μοντέλου εκτιμώνται με τη **μέθοδο της μέγιστης πιθανοφάνειας** (Maximum Likelihood Estimation). Η μέθοδος αυτή επιλέγει τις παραμετρικές τιμές που **μεγιστοποιούν την πιθανότητα εμφάνισης των παρατηρούμενων δεδομένων** σύμφωνα με το προτεινόμενο μοντέλο (Menard, 2002; Hosmer et al., 2013). Σε πρακτικό επίπεδο, αυτό επιτυγχάνεται με αριθμητικές μεθόδους βελτιστοποίησης, οι οποίες εντοπίζουν τις τιμές των συντελεστών που καθιστούν τα δεδομένα του δείγματος στατιστικά πιο αναμενόμενα υπό τις υποθέσεις του μοντέλου.

Στην παρούσα εργασία, η εξαρτημένη μεταβλητή είναι η θέση του πεζού (εντός ή εκτός διάβασης) και η κατάσταση της διέλευσης του, ενώ οι ανεξάρτητες μεταβλητές περιλαμβάνουν διάφορα αντιπροσωπευτικά χαρακτηριστικά τόσο των πεζών όσο και των οχημάτων και κατατάσσονται είτε στην κατηγορία των συνεχών μεταβλητών είτε στην κατηγορία των κατηγορικών όπου και χρήζουν τροποποίησης. Μέσω της Λογιστικής Παλινδρόμησης επιχειρείται η **ποσοτική ανάλυση της επίδρασης αυτών των παραγόντων στην πιθανότητα παρουσίας του πεζού εντός διάβασης**, με στόχο την ερμηνεία της συσχέτισης και της στατιστικής σημασίας κάθε μεταβλητής.

Τέλος, κατά την **εκπαίδευση** του μοντέλου, η **αυτόματη εξισορρόπηση βαρών κατηγοριών** μέσω της παραμέτρου **class_weight='balanced'** μπορεί να ενισχύσει τη μάθηση από λιγότερο συχνές κατηγορίες, μειώνοντας την προκατάληψη υπέρ των συχνότερων (King & Zeng, 2001). Εκτός από την αυτόματη εξισορρόπηση βαρών, εφαρμόζεται και η μέθοδος **SMOTE (Synthetic Minority Over-sampling Technique)** για την αντιμετώπιση του προβλήματος της άνισης κατανομής τάξεων στο σύνολο εκπαίδευσης. Η μέθοδος SMOTE δημιουργεί τεχνητά δείγματα της μειοψηφούσας τάξης (π.χ. "πεζός εντός διάβασης") με χρήση γραμμικών συνδυασμών υπαρχόντων παραδειγμάτων, αυξάνοντας έτσι την αναπαράσταση της μειοψηφίας χωρίς να εισάγει απλή επανάληψη (duplication) (Chawla et al., 2002). Η τεχνική αυτή συμβάλλει σημαντικά στη βελτίωση της γενίκευσης του μοντέλου, αποτρέποντας την τάση να ευνοούνται μόνο οι περιπτώσεις της πλειοψηφούσας κατηγορίας. Ο συνδυασμός **SMOTE** με **class_weight='balanced'** προσφέρει μια πιο ισόρροπη εκπαίδευση, ιδιαίτερα χρήσιμη σε εφαρμογές όπου η αποτυχία ανίχνευσης της μειοψηφικής κατηγορίας έχει κρίσιμη σημασία.

3.7.3 Κριτήρια αξιολόγησης και αποδοχής μοντέλου

Η αξιολόγηση των αποτελεσμάτων της Λογιστικής Παλινδρόμησης είναι κρίσιμη για την εξαγωγή επιστημονικά έγκυρων συμπερασμάτων. Αν και η Λογιστική Παλινδρόμηση προσφέρει τη δυνατότητα πρόβλεψης της πιθανότητας εμφάνισης ενός δυαδικού γεγονότος, απαιτείται προσεκτική ανάλυση των συντελεστών, της στατιστικής σημαντικότητας και των συνολικών μετρικών ποιότητας του μοντέλου για να διασφαλιστεί η αξιοπιστία των ευρημάτων.

Ερμηνεία συντελεστών παλινδρόμησης

Οι συντελεστές β_i που προκύπτουν από τη λογιστική παλινδρόμηση **δεν εκφράζουν άμεσα μεταβολή στην πιθανότητα**, αλλά στη **λογαριθμική τιμή του λόγου των πιθανοτήτων**, δηλαδή στα log-odds (Hosmer et al., 2013; Menard, 2002). Συγκεκριμένα, κάθε συντελεστής β_i δείχνει πόσο αυξάνεται ή μειώνεται ο λογάριθμος των odds του εξαρτώμενου δυαδικού αποτελέσματος όταν η αντίστοιχη ανεξάρτητη μεταβλητή X_i αυξάνεται κατά μία μονάδα, κρατώντας τις υπόλοιπες σταθερές (Agresti, 2018).

Επειδή οι log-odds είναι δύσκολο να ερμηνευτούν άμεσα, χρησιμοποιείται η **εκθετική μορφή των συντελεστών**, δηλαδή:

$$OR_i = e^{\beta_i} \quad (\text{Εξ. 5})$$

Αυτό το μέγεθος OR_i ονομάζεται **odds ratio** (λόγος πιθανοτήτων) και εκφράζει **πόσες φορές αυξάνεται (ή μειώνεται)** η πιθανότητα να συμβεί το γεγονός, **όταν η μεταβλητή X_i αυξάνεται κατά μία μονάδα** (Long, 1997).

- Αν $OR > 1$: η μεταβλητή έχει **θετική επίδραση** στο ενδεχόμενο εμφάνισης του γεγονότος (αυξάνει την πιθανότητα).
- Αν $OR < 1$: η μεταβλητή έχει **αρνητική επίδραση** (μειώνει την πιθανότητα).
- Αν $OR = 1$: η μεταβλητή **δεν επηρεάζει** την πιθανότητα.

Η ερμηνεία των συντελεστών και των odds ratios εξαρτάται από το πώς έχει κωδικοποιηθεί η εξαρτημένη μεταβλητή (π.χ. αν η τιμή 1 αντιστοιχεί σε «εντός διάβασης»), καθώς και από το θεωρητικό και πρακτικό πλαίσιο της έρευνας (Hosmer et al., 2013).

Στατιστική σημαντικότητα συντελεστών

Κατά την εφαρμογή λογιστικής παλινδρόμησης, είναι κρίσιμο να εξεταστεί αν κάθε ανεξάρτητη μεταβλητή συμβάλλει ουσιαστικά στην πρόβλεψη του εξαρτημένου δυαδικού αποτελέσματος. Αυτό γίνεται μέσω της αξιολόγησης της **στατιστικής σημαντικότητας** των αντίστοιχων συντελεστών β_i (Hosmer et al., 2013; Agresti, 2018).

Η στατιστική σημαντικότητα ελέγχεται με βάση το **p-value** κάθε συντελεστή, το οποίο προκύπτει από τον **έλεγχο Wald**. Ο έλεγχος αυτός εξετάζει τη μηδενική υπόθεση ότι $\beta_i = 0$, δηλαδή ότι η αντίστοιχη ανεξάρτητη μεταβλητή δεν έχει επίδραση στην εξαρτημένη (Menard, 2002). Για να υπολογιστεί το p-value, χρησιμοποιείται η τιμή του

Wald στατιστικού, το οποίο δίνεται από τον λόγο του εκτιμημένου συντελεστή προς την τυπική του απόκλιση:

$$W = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \quad (\text{Εξ. 6})$$

Αυτή η τιμή συγκρίνεται με την κανονική ή χ^2 κατανομή (ανάλογα την υλοποίηση), και από εκεί προκύπτει το p-value, δηλαδή η **πιθανότητα να παρατηρηθεί τέτοια ή εντονότερη επίδραση, αν στην πραγματικότητα $\beta_i = 0$** (Field, 2013).

Αν το p-value είναι **μικρότερο από το προκαθορισμένο επίπεδο σημαντικότητας** (π.χ. $\alpha=0.05$), τότε απορρίπτεται η μηδενική υπόθεση και θεωρείται ότι η μεταβλητή έχει **στατιστικά σημαντική επίδραση** στο εξαρτημένο αποτέλεσμα.

Εκτός από τον έλεγχο κάθε μεταβλητής ξεχωριστά, είναι συχνά χρήσιμο να ελεγχθεί αν ένα σύνολο μεταβλητών βελτιώνει συνολικά το μοντέλο. Αυτό γίνεται με τον έλεγχο λόγου πιθανοφάνειας (**Likelihood Ratio Test – LRT**), ο οποίος συγκρίνει δύο ένθετα μοντέλα: ένα μικρό μοντέλο που περιέχει μόνο τις βασικές μεταβλητές και ένα πλήρες μοντέλο που περιλαμβάνει τις βασικές και επιπλέον μεταβλητές. Ο έλεγχος βασίζεται στις τιμές log-likelihood των δύο μοντέλων και υπολογίζεται ως:

$$\text{LRT} = -2 \cdot (\text{LogLik}_{\text{μικρό}} - \text{LogLik}_{\text{πλήρες}}) \quad (\text{Εξ. 7})$$

Η τιμή αυτή συγκρίνεται με τη χ^2 κατανομή με βαθμούς ελευθερίας ίσους με τον αριθμό των προστιθέμενων παραμέτρων (Long, 1997). Αν το προκύπτον p-value είναι μικρότερο από 0.05, τότε το πλήρες μοντέλο θεωρείται ότι έχει σημαντικά καλύτερη προσαρμογή και συνεπώς οι επιπλέον μεταβλητές βελτιώνουν την προβλεπτική ικανότητα του μοντέλου. Αν το p-value είναι μεγαλύτερο ή ίσο με 0.05, οι επιπλέον μεταβλητές δεν προσφέρουν στατιστικά σημαντική βελτίωση.

Ποιότητα μοντέλου (Goodness-of-Fit)

Σε αντίθεση με την Point-Biserial Correlation, η οποία απλώς μετρά τον βαθμό γραμμικής σχέσης μεταξύ μίας δυαδικής και μίας συνεχούς μεταβλητής (Cohen, 1988), η **Λογιστική Παλινδρόμηση** είναι **προβλεπτικό μοντέλο** (Hosmer et al., 2013; Agresti, 2018). Επειδή έχει ως στόχο την πρόβλεψη της πιθανότητας εμφάνισης ενός γεγονότος, η αξιολόγησή του απαιτεί τη χρήση **ειδικών μετρικών καταλληλότητας** που λαμβάνουν υπόψη τόσο τη στατιστική προσαρμογή όσο και την προβλεπτική ικανότητα του μοντέλου (Long, 1997).

Τα κυριότερα κριτήρια αξιολόγησης περιλαμβάνουν:

- **Pseudo-R²** (όπως το McFadden's R²): Το McFadden's Pseudo-R² είναι ένα μέτρο που δείχνει **πόσο καλύτερο είναι το μοντέλο μας σε σχέση με ένα «άδειο» μοντέλο**, δηλαδή ένα μοντέλο που **δεν περιλαμβάνει καμία ανεξάρτητη μεταβλητή** και απλώς προβλέπει το ίδιο αποτέλεσμα για όλες τις παρατηρήσεις. Συγκεκριμένα, συγκρίνει τη λογαριθμική πιθανοφάνεια (log-likelihood) του τελικού μοντέλου με αυτήν του null model. Ο τύπος υπολογισμού είναι (McFadden, 1974):

$$R^2_{\text{McFadden}} = 1 - \frac{\log L_{\text{πλήρες}}}{\log L_{\text{κενό}}} \quad (\text{Εξ. 8})$$

όπου:

- ο $\log L_{\text{πλήρες}}$ είναι το log-likelihood του μοντέλου με τις μεταβλητές,
- ο $\log L_{\text{κενό}}$ είναι το log-likelihood του μοντέλου χωρίς μεταβλητές.

Αν και δεν ερμηνεύεται όπως το R^2 στη γραμμική παλινδρόμηση (δηλαδή δεν δείχνει πόσο ποσοστό της μεταβλητότητας εξηγείται), **μεγαλύτερες τιμές δηλώνουν καλύτερη προσαρμογή**. Ενδεικτικά, τιμές:

- ο κοντά στο 0 υποδηλώνουν φτωχή προσαρμογή,
 - ο πάνω από 0.2 θεωρούνται ικανοποιητικές,
 - ο πάνω από 0.4 πολύ καλές, ειδικά σε κοινωνικές και βιολογικές επιστήμες.
- **AIC (Akaike Information Criterion):** Το AIC είναι ένα μέτρο που συνδυάζει δύο πράγματα, πρώτον, πόσο καλά "ταιριάζει" το μοντέλο στα δεδομένα (με βάση τη λογαριθμική πιθανοφάνεια), και δεύτερον, πόσο απλό ή πολύπλοκο είναι το μοντέλο, δηλαδή πόσες παραμέτρους έχει (Burnham & Anderson, 2002). Ο τύπος υπολογισμού είναι:

$$AIC = 2k - 2 \log L \quad (\text{Εξ. 9})$$

όπου:

- ο k : ο αριθμός των παραμέτρων του μοντέλου,
- ο $\log L$: το log-likelihood του μοντέλου.

Όσο **μικρότερη είναι η τιμή του AIC**, τόσο **καλύτερο θεωρείται το μοντέλο**, επειδή καταφέρνει να προβλέπει καλά χωρίς να είναι περιττά πολύπλοκο. Το AIC χρησιμοποιείται κυρίως για **σύγκριση μοντέλων μεταξύ τους**: δεν λέει αν ένα μοντέλο είναι «καλό» ή «κακό» από μόνο του, αλλά ποιο μοντέλο είναι καλύτερο μέσα σε ένα σύνολο υποψήφιων μοντέλων. Η καλύτερη επιλογή είναι αυτή με το χαμηλότερο AIC.

- **BIC (Bayesian Information Criterion):** Το BIC είναι ένα ακόμη μέτρο καταλληλότητας μοντέλου, παρόμοιο με το AIC, το οποίο επίσης συνδυάζει την ποιότητα προσαρμογής και την πολυπλοκότητα του μοντέλου. Η διαφορά του από το AIC είναι ότι το BIC τιμωρεί περισσότερο την προσθήκη παραμέτρων, ιδίως όταν το μέγεθος δείγματος είναι μεγάλο. Ο υπολογισμός του βασίζεται στον εξής τύπο (Schwarz, 1978):

$$BIC = \ln(n) \cdot k - 2 \log L \quad (\text{Εξ. 10})$$

όπου:

- ο n : το μέγεθος του δείγματος,
- ο k : ο αριθμός των παραμέτρων του μοντέλου,
- ο $\log L$: το log-likelihood του μοντέλου.

Το BIC προτιμά πιο απλά μοντέλα όταν η διαφορά στην προσαρμογή δεν είναι μεγάλη. Όπως και το AIC, χρησιμοποιείται για **σύγκριση μοντέλων**: το μοντέλο με το **χαμηλότερο BIC** θεωρείται καλύτερο, εφόσον επιτυγχάνει καλή ισορροπία ανάμεσα σε ακρίβεια και απλότητα.

- **Ποσοστά ορθής ταξινόμησης, ROC καμπύλη και Area Under the Curve (AUC)**: Αυτά τα μέτρα εστιάζουν στην **προβλεπτική ικανότητα** του μοντέλου. Το ποσοστό ορθής ταξινόμησης δείχνει σε τι βαθμό το μοντέλο προβλέπει σωστά τις κατηγορίες. Η **ROC καμπύλη** απεικονίζει τη σχέση μεταξύ ευαισθησίας και ειδικότητας για διαφορετικά κατώφλια ταξινόμησης, ενώ το **AUC** (εμβαδόν κάτω από την καμπύλη) συνοψίζει την **ικανότητα του μοντέλου να διαχωρίζει σωστά τις κατηγορίες**: όσο πλησιάζει το 1, τόσο καλύτερη η διακριτική ικανότητα (Fawcett, 2006). Ενδεικτικά, **AUC $\geq$ 0.70** θεωρείται αποδεκτό, ενώ **AUC $\geq$ 0.90** εξαιρετικό.

Επιπρόσθετα, για την αναλυτικότερη αξιολόγηση της επίδοσης του μοντέλου Λογιστικής Παλινδρόμησης χρησιμοποιείται ο **πίνακας σύγχυσης (confusion matrix)**. Ο πίνακας αυτός καταγράφει τον αριθμό των σωστών και λανθασμένων προβλέψεων για κάθε κατηγορία της εξαρτημένης μεταβλητής, διακρίνοντας μεταξύ:

- **True Positives (TP)**: σωστά προβλεφθείσες θετικές περιπτώσεις,
- **False Positives (FP)**: περιπτώσεις που προβλέφθηκαν ως θετικές αλλά ήταν αρνητικές,
- **False Negatives (FN)**: περιπτώσεις που προβλέφθηκαν ως αρνητικές ενώ ήταν θετικές,
- **True Negatives (TN)**: σωστά προβλεφθείσες αρνητικές περιπτώσεις (Sokolova & Lapalme, 2009).

Από τον πίνακα σύγχυσης προκύπτουν βασικές μετρικές αξιολόγησης ταξινομητικών μοντέλων:

- **Accuracy (Ακρίβεια)**: το ποσοστό των συνολικών σωστών προβλέψεων επί του συνόλου των παρατηρήσεων (Chicco & Jurman, 2020):

$$Accuracy = (TP + TN) / (TP + FP + FN + TN) \quad (\text{Εξ. 11})$$

- **Precision (Εκλεκτικότητα)**: το ποσοστό των σωστών θετικών προβλέψεων επί όλων των θετικών προβλέψεων (Chicco & Jurman, 2020):

$$Precision = TP / (TP + FP) \quad (\text{Εξ. 12})$$

- **Recall (Ευαισθησία)**: η ικανότητα του μοντέλου να εντοπίζει τις πραγματικά θετικές περιπτώσεις (Chicco & Jurman, 2020):

$$Recall = TP / (TP + FN) \quad (\text{Εξ. 13})$$

- **F1-score**: ο αρμονικός μέσος των precision και recall, ο οποίος χρησιμοποιείται ιδιαίτερα σε περιπτώσεις μη ισορροπημένων τάξεων (Chicco & Jurman, 2020):

$$F1 - score = 2 \cdot (Precision \cdot Recall) / (Precision + Recall) \quad (\text{Εξ. 14})$$

Οι μετρικές αυτές συμβάλλουν στην πληρέστερη κατανόηση της συμπεριφοράς του μοντέλου, καθώς επιτρέπουν την αποτίμηση όχι μόνο της συνολικής ακρίβειας, αλλά και της ικανότητάς του να εντοπίζει σωστά συγκεκριμένες κατηγορίες. Ιδίως όταν το πλήθος των περιπτώσεων ανά κατηγορία δεν είναι ισορροπημένο, η ερμηνεία των αποτελεσμάτων με βάση τον πίνακα σύγχυσης καθίσταται απαραίτητη (Fernández et al., 2018).

Επιπλέον, η συμπεριφορά του μοντέλου επηρεάζεται από την τιμή του **κατωφλίου ταξινόμησης** (threshold), δηλαδή του ορίου πιθανότητας πάνω από το οποίο μία πρόβλεψη θεωρείται θετική. Η προεπιλεγμένη τιμή (συνήθως 0.5) μπορεί να τροποποιηθεί ανάλογα με τις απαιτήσεις της εφαρμογής, επηρεάζοντας τις τιμές των TP, FP, FN και TN και κατ' επέκταση τις σχετικές μετρικές (π.χ. precision ή recall) (Hossin & Sulaiman, 2015). Η ρύθμιση του threshold επιτρέπει την προσαρμογή του μοντέλου στην επιθυμητή ισορροπία μεταξύ ψευδών θετικών και ψευδών αρνητικών.

Για περιπτώσεις **ανισόρροπων τάξεων**, μία χρήσιμη μετρική είναι η **balanced accuracy**, η οποία προσφέρει μια πιο δίκαιη εκτίμηση της απόδοσης. Υπολογίζεται ως ο μέσος όρος της ευαισθησίας (recall) και της ειδικότητας (specificity), λαμβάνοντας εξίσου υπόψη και τις δύο κατηγορίες (Brodersen et al., 2010):

$$\text{Balanced Accuracy} = \frac{1}{2} \left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right) \quad (\text{Εξ. 15})$$

Η χρήση της είναι ιδιαίτερα σημαντική όταν το μοντέλο ενδέχεται να ευνοεί την πολυπληθέστερη τάξη, παραβλέποντας τις σπάνιες περιπτώσεις.

Υπόλοιπα σφάλματος και έλεγχος υποθέσεων

Αν και η **Λογιστική Παλινδρόμηση** δεν στηρίζεται στις ίδιες στατιστικές υποθέσεις με τη γραμμική παλινδρόμηση — όπως η κανονικότητα των σφαλμάτων ή η ομοσκεδαστικότητα — εξακολουθεί να απαιτεί τον έλεγχο διαφόρων παραγόντων που μπορούν να επηρεάσουν την εγκυρότητα του μοντέλου (Hosmer et al., 2013). Μεταξύ των βασικών διαγνωστικών ελέγχων περιλαμβάνονται:

- **Κίνδυνος υπερπροσαρμογής (overfitting):** Το φαινόμενο αυτό εμφανίζεται όταν το μοντέλο είναι υπερβολικά πολύπλοκο σε σχέση με το μέγεθος του δείγματος, «μαθαίνει» τις ιδιαιτερότητες των συγκεκριμένων δεδομένων και χάνει την ικανότητα γενίκευσης σε νέα. Η πιθανότητα εμφάνισης overfitting αυξάνεται όταν το μοντέλο περιλαμβάνει μεγάλο αριθμό μεταβλητών ή αλληλεπιδράσεων, ειδικά σε μικρά δείγματα (Babyak, 2004).
- **Πολυσυγγραμμικότητα (multicollinearity):** Αναφέρεται στην ύπαρξη ισχυρής συσχέτισης μεταξύ δύο ή περισσότερων ανεξάρτητων μεταβλητών. Η πολυσυγγραμμικότητα μπορεί να προκαλέσει αστάθεια στους συντελεστές παλινδρόμησης, οδηγώντας σε φουσκωμένα σφάλματα και παραπλανητικές στατιστικές σημαντικότητες. Ένας συνηθισμένος τρόπος εντοπισμού της είναι ο **συντελεστής διόγκωσης της διακύμανσης (VIF – Variance Inflation Factor):** υψηλές τιμές VIF (συνήθως > 5 ή 10) υποδηλώνουν προβληματική συσχέτιση (Kutner et al., 2005).

- **Ανεξαρτησία των παρατηρήσεων:** Προϋποθέτει ότι κάθε παρατήρηση στο δείγμα είναι στατιστικά ανεξάρτητη από τις υπόλοιπες. Η παραβίαση αυτής της υπόθεσης (π.χ. σε επαναλαμβανόμενες μετρήσεις ή ομαδοποιημένα δεδομένα) μπορεί να οδηγήσει σε λανθασμένες εκτιμήσεις και στατιστικά συμπεράσματα (Field, 2013).
- **Ορθή κωδικοποίηση μεταβλητών:** Ιδιαίτερα σημαντική είναι η σωστή αναπαράσταση των κατηγορικών μεταβλητών μέσω κατάλληλων dummy variables, ώστε να διασφαλίζεται η ερμηνευσιμότητα και η τεχνική σταθερότητα του μοντέλου. Η μη ορθή κωδικοποίηση μπορεί να οδηγήσει σε παρανοήσεις σχετικά με την επίδραση συγκεκριμένων κατηγοριών (Menard, 2002).

3.8 Μοντέλο Random Forest

3.8.1 Εισαγωγή

Στο πλαίσιο της παρούσας διπλωματικής εργασίας, για τη διερεύνηση και ταξινόμηση της θέσης ενός πεζού ως προς τη διάβαση και ως προς την κατάσταση διέλευσης του, βάσει επιμέρους χαρακτηριστικών της κίνησης, εφαρμόστηκε η μέθοδος μηχανικής μάθησης Random Forest. Η συγκεκριμένη τεχνική επιλέχθηκε λόγω της ικανότητάς της να χειρίζεται **μη γραμμικές σχέσεις**, να ενσωματώνει αλληλεπιδράσεις μεταξύ μεταβλητών και να παρέχει υψηλή ακρίβεια στην κατηγοριοποίηση. Η εξαρτημένες μεταβλητές της μελέτης είναι δυαδικές (εντός ή εκτός διάβασης και νόμιμη ή παράνομη διέλευση), ενώ οι ανεξάρτητες μεταβλητές είναι συνεχείς αλλά και κατηγορικές, οι οποίες χρήζουν τροποποίησης, και αφορούν τόσο τους πεζούς όσο και τα οχήματα. Τα δεδομένα εξήχθησαν από ανάλυση βιντεοσκοπημένου υλικού με χρήση MATLAB και Python και στη συνέχεια εισήχθησαν στο μοντέλο Random Forest για την εκπαίδευση και αξιολόγηση της ταξινομητικής του ικανότητας. Η παρούσα ενότητα αναλύει τη θεωρητική βάση της μεθόδου και την εφαρμογή της στο σύνολο των δεδομένων.

3.8.2 Θεωρητικό υπόβαθρο μεθόδου

Η μέθοδος Random Forest αποτελεί έναν αλγόριθμο εποπτευόμενης **μηχανικής μάθησης** (supervised learning), ο οποίος ανήκει στην κατηγορία των τεχνικών συνόλου (ensemble methods). Συνδυάζει την προβλεπτική ισχύ πολλών **δέντρων απόφασης** (Decision Trees), καθένα από τα οποία εκπαιδεύεται σε διαφορετικά τυχαία υποσύνολα του αρχικού συνόλου δεδομένων. Στόχος είναι η δημιουργία ενός μοντέλου με **βελτιωμένη ακρίβεια** και ικανότητα γενίκευσης (Breiman, 2001). Η μέθοδος χαρακτηρίζεται από ευελιξία, δυνατότητα χειρισμού **μη γραμμικών** και περίπλοκων σχέσεων και **ανθεκτικότητα στο overfitting**, ιδίως σε σχέση με μεμονωμένα δέντρα (Cutler et al., 2007).

Στην παρούσα εργασία, η Random Forest εφαρμόστηκε σε πρόβλημα **δυαδικής ταξινόμησης**, με εξαρτημένη μεταβλητή τη θέση του πεζού ως προς τη διάβαση (εντός ή εκτός) και την κατάσταση της διέλευσης του, και με μεταβλητές εισόδου διάφορες αριθμητικές αλλά και κατηγορικές μεταβλητές που αφορούσαν τόσο τους πεζούς όσο και τα οχήματα.

Η βασική λειτουργία της Random Forest περιλαμβάνει τρία βασικά στάδια (Breiman, 2001):

1. **Δημιουργία υποσυνόλων εκπαίδευσης** (bootstrap sampling): Από το αρχικό σύνολο δεδομένων D δημιουργούνται B δείγματα με επαναληπτική δειγματοληψία. Κάθε δείγμα χρησιμοποιείται για την εκπαίδευση ενός ξεχωριστού δέντρου απόφασης T_b , με $b = 1, 2, \dots, B$.
2. **Τυχαία επιλογή χαρακτηριστικών σε κάθε κόμβο**: Σε κάθε διαχωριστικό κόμβο του δέντρου, επιλέγεται τυχαία ένα υποσύνολο χαρακτηριστικών, με μέγεθος $m < p$, όπου p το πλήθος των διαθέσιμων μεταβλητών, εντός του οποίου γίνεται η επιλογή του βέλτιστου διαχωρισμού.
3. **Συνδυασμός προβλέψεων**: Η τελική πρόβλεψη προκύπτει μέσω **πλειοψηφικής ψήφου** (majority vote) των επιμέρους δέντρων στην περίπτωση ταξινόμησης, ή μέσω μέσου όρου στην περίπτωση παλινδρόμησης (Cutler et al., 2007).

Η μαθηματική έκφραση της τελικής πρόβλεψης σε πρόβλημα ταξινόμησης είναι (Breiman, 2001):

$$\hat{y} = \text{mode}\{T_b(x)\}_{b=1}^B \quad (\text{Εξ. 16})$$

όπου:

- $\hat{y}$: η τελική προβλεπόμενη κατηγορία
- $T_b(x)$: η πρόβλεψη του b -οστού δέντρου για την παρατήρηση x ,
- B : το πλήθος των δέντρων

Κατά την κατασκευή των κόμβων των δέντρων, χρησιμοποιούνται **δείκτες καθαρότητας** για την επιλογή διαχωρισμών, όπως:

- **Gini Impurity**: $G = \sum_{i=1}^C p_i(1 - p_i)$ (Εξ. 17)
- **Entropy**: $H = - \sum_{i=1}^C p_i \log_2(p_i)$ (Εξ. 18)

όπου p_i είναι η σχετική συχνότητα της κατηγορίας i στον κόμβο, και C ο αριθμός των κατηγοριών (στην παρούσα περίπτωση $C=2$) (Breiman, 2001; Quinlan, 1993).

Ένα σημαντικό πλεονέκτημα της Random Forest είναι η δυνατότητα εκτίμησης της **σημαντικότητας μεταβλητών** (feature importance). Η συνεισφορά κάθε μεταβλητής στην απόδοση του μοντέλου αποτιμάται βάσει της συνολικής **μείωσης καθαρότητας** (Gini ή Entropy) που επιφέρει στα δέντρα όπου συμμετέχει. Αυτό προσδίδει ερμηνευτική ισχύ στη μέθοδο, καθιστώντας την πολύτιμη και για ανάλυση επιρροών (Louppe et al., 2013).

Η **αξιολόγηση της απόδοσης** του μοντέλου γίνεται εσωτερικά μέσω της τεχνικής **Out-of-Bag (OOB) error estimation**. Κάθε παρατήρηση που δεν χρησιμοποιείται για την εκπαίδευση ενός δέντρου μπορεί να αξιοποιηθεί ως «δοκιμαστικό» παράδειγμα για την εκτίμηση της ακρίβειας του δάσους συνολικά, χωρίς την ανάγκη ξεχωριστού συνόλου επικύρωσης (Breiman, 2001).

Συνολικά, η Random Forest συνδυάζει **υπολογιστική αποδοτικότητα**, **υψηλή ακρίβεια** και **διαφάνεια στην ερμηνεία**, γεγονός που την καθιστά ιδιαίτερα κατάλληλη για την παρούσα εφαρμογή πρόβλεψης της συμπεριφοράς πεζών με βάση παρατηρήσιμα χαρακτηριστικά κίνησης.

3.8.3 Κριτήρια αξιολόγησης και αποδοχής μοντέλου

Η αξιολόγηση της απόδοσης του μοντέλου Random Forest είναι ουσιαστική για την εξαγωγή έγκυρων και αξιόπιστων συμπερασμάτων. Αν και η μέθοδος βασίζεται σε πολύπλοκες εσωτερικές διαδικασίες (όπως η τυχαία δημιουργία δέντρων και η ενσωμάτωσή τους), παρέχει πληθώρα δεικτών που επιτρέπουν την πλήρη ποσοτική και ποιοτική εκτίμηση της ταξινομητικής του ικανότητας.

Ερμηνεία προβλέψεων και σημασία χαρακτηριστικών (feature importance)

Η Random Forest, ως μοντέλο κατηγοριοποίησης, μπορεί να δώσει για κάθε παρατήρηση είτε την τελική **προβλεπόμενη κατηγορία** (π.χ. «εντός διάβασης»), είτε την εκτιμώμενη **πιθανότητα** αυτής της κατηγορίας. Η πιθανότητα αυτή υπολογίζεται με βάση το **ποσοστό των δέντρων** του μοντέλου που «ψήφισαν» υπέρ της συγκεκριμένης κατηγορίας (Breiman, 2001).

Η **σπουδαιότητα κάθε μεταβλητής** (feature importance) για την τελική πρόβλεψη μπορεί να εκτιμηθεί με δύο βασικούς τρόπους:

- Μέσω της **συνολικής μείωσης της καθαρότητας** (όπως του **Gini Impurity** ή της εντροπίας) που προκαλεί η μεταβλητή στους κόμβους όλων των δέντρων όπου χρησιμοποιείται. Όσο μεγαλύτερη η μείωση, τόσο πιο σημαντική θεωρείται η μεταβλητή (Louppe et al., 2013).
- Εναλλακτικά, μέσω της τεχνικής **permutation importance**, όπου η μεταβλητή αναδιατάσσεται τυχαία (δηλαδή διαταράσσεται η αρχική της πληροφορία) και παρατηρείται πόσο **μειώνεται η ακρίβεια του μοντέλου** στις παρατηρήσεις που δεν χρησιμοποιήθηκαν στην εκπαίδευση (Out-of-Bag δείγματα). Αν η ακρίβεια πέσει σημαντικά, αυτό σημαίνει ότι η μεταβλητή είχε σημαντικό ρόλο (Altmann et al., 2010).

Με αυτούς τους τρόπους, η Random Forest αποκτά ένα βαθμό **ερμηνευσιμότητας**, καθώς μπορούμε να εκτιμήσουμε **ποια χαρακτηριστικά επηρεάζουν περισσότερο τις προβλέψεις του μοντέλου**.

Στατιστική σημαντικότητα και αξιοπιστία

Αν και η Random Forest δεν βασίζεται σε παραμετρικά μοντέλα και δεν εφαρμόζει άμεσα **κλασικούς ελέγχους υποθέσεων** (όπως t-tests για τους συντελεστές), παρέχει εναλλακτικούς, μη παραμετρικούς τρόπους για την αξιολόγηση της σημασίας και της αξιοπιστίας των μεταβλητών και του ίδιου του μοντέλου (Strobl et al., 2008). Οι βασικές τεχνικές είναι:

- **Σφάλμα Out-of-Bag (OOB):** Κατά την εκπαίδευση κάθε δέντρου, χρησιμοποιείται μόνο ένα τυχαίο δείγμα από τα αρχικά δεδομένα (bootstrap sample). Οι παρατηρήσεις που δεν περιλαμβάνονται σε αυτό το δείγμα (λέγονται OOB παρατηρήσεις) χρησιμοποιούνται ως ένα είδος «εσωτερικού» σετ ελέγχου. Η **OOB ακρίβεια** είναι η μέση απόδοση του μοντέλου σε αυτές τις παρατηρήσεις και παρέχει μια **αντικειμενική εκτίμηση της γενίκευσης** του μοντέλου χωρίς να απαιτείται ξεχωριστό σετ επικύρωσης (Breiman, 2001).

- **Permutation tests** (δοκιμές με αναδιάταξη): Σε αυτές, επιλέγεται μία μεταβλητή και οι τιμές της ανακατεύονται τυχαία, σπάζοντας τη σχέση της με την εξαρτημένη μεταβλητή. Στη συνέχεια επανυπολογίζεται η απόδοση του μοντέλου. Αν η **ακρίβεια μειωθεί σημαντικά**, τότε η μεταβλητή θεωρείται σημαντική. Αν δεν αλλάξει, η μεταβλητή πιθανώς δεν έχει προβλεπτική αξία. Πρόκειται ουσιαστικά για έναν **έμμεσο έλεγχο σημαντικότητας**, βασισμένο στη συμβολή κάθε μεταβλητής στην απόδοση (Altmann et al., 2010).

Αυτές οι τεχνικές δεν βασίζονται σε παραδοχές για την κατανομή των δεδομένων και είναι κατάλληλες για πολύπλοκα μοντέλα όπως η Random Forest. Παρέχουν επομένως **ισχυρές ενδείξεις αξιοπιστίας και σημασίας**, ειδικά σε περιπτώσεις όπου η κλασική στατιστική δεν μπορεί να εφαρμοστεί άμεσα.

Ποιότητα μοντέλου (μετρικές ταξινόμησης)

Η ποιότητα της Random Forest ως μοντέλου **δυναμικής ταξινόμησης** αξιολογείται μέσω **τυπικών μετρικών ακρίβειας**, όπως η **Accuracy**, η **Recall** (Ευαισθησία), η **Specificity** (Ειδικότητα), το **F1-score** και άλλες. Οι ορισμοί και οι μαθηματικές εκφράσεις αυτών των δεικτών έχουν ήδη παρουσιαστεί σε προηγούμενη ενότητα.

Επιπλέον, για πιο πλήρη αξιολόγηση της ταξινομητικής ικανότητας του μοντέλου, χρησιμοποιείται η **καμπύλη ROC** (*Receiver Operating Characteristic*), η οποία απεικονίζει τη σχέση μεταξύ **True Positive Rate** (ευαισθησία) και **False Positive Rate** (1 – ειδικότητα), για διαφορετικά κατώφλια πρόβλεψης.

Η **επιφάνεια κάτω από την καμπύλη ROC (AUC – Area Under Curve)** παρέχει μία **συνολική μέτρηση της διακριτικής ικανότητας** του μοντέλου:

- $AUC \geq 0.70 \rightarrow$ **Αποδεκτή απόδοση**,
- $AUC \geq 0.90 \rightarrow$ **Εξαιρετική διακριτική ικανότητα** (Fawcett, 2006).

Η απόδοση της Random Forest σε προβλήματα δυναμικής ταξινόμησης επηρεάζεται και από την επιλογή **κατωφλίου** (threshold), δηλαδή της ελάχιστης απαιτούμενης πιθανότητας ώστε μία παρατήρηση να ταξινομηθεί στην «θετική» κατηγορία. Θεωρητικά, η πιθανότητα αυτή προκύπτει ως το ποσοστό των δέντρων του δάσους που προβλέπουν την θετική κατηγορία. Αν το ποσοστό αυτό ξεπερνά το threshold (τυπικά 0.5), η παρατήρηση ταξινομείται ως θετική (Ferri et al., 2009). Ωστόσο, σε περιπτώσεις ανισόρροπων τάξεων, μπορεί να είναι σκόπιμη η τροποποίηση του κατωφλίου ώστε να βελτιστοποιηθούν άλλες μετρικές απόδοσης όπως το recall ή το F1-score (Jeni et al., 2013).

Για την αντιμετώπιση της ανισορροπίας τάξεων, όπως αναφέρθηκε και στο κομμάτι της Λογιστικής Παλινδρόμησης, χρησιμοποιείται η παράμετρος `class_weight='balanced'`, η οποία αυξάνει το βάρος των παρατηρήσεων της μειοψηφούσας τάξης κατά τον υπολογισμό των διαχωρισμών, αποτρέποντας την υπερεκπαίδευση στην πλειοψηφούσα κατηγορία (King & Zeng, 2001). Επιπλέον, εφαρμόζεται η μέθοδος SMOTE (Chawla et al., 2002), η οποία δημιουργεί συνθετικά δείγματα της μειοψηφούσας κατηγορίας πριν τη διαδικασία εκπαίδευσης, ενισχύοντας την ικανότητα του μοντέλου να αναγνωρίζει σπάνιες περιπτώσεις.

Σφάλματα και αξιοπιστία

Η **Random Forest** δεν βασίζεται σε γραμμικά μοντέλα ούτε προϋποθέτει **παραμετρικές υποθέσεις** για την κατανομή του σφάλματος, όπως συμβαίνει στη Γραμμική ή Λογιστική Παλινδρόμηση. Αντί για την εκτίμηση σφαλμάτων μέσω υπολοίπων (residuals) ή κατανομών σφάλματος, χρησιμοποιεί **εναλλακτικές μεθόδους εκτίμησης σφάλματος** και αξιοπιστίας που προσαρμόζονται στη μη παραμετρική της φύση (Breiman, 2001).

Οι βασικοί τρόποι εκτίμησης του σφάλματος είναι:

- **Υπολογισμός του σφάλματος Out-of-Bag (OOB Error):** Κατά την εκπαίδευση κάθε δέντρου, χρησιμοποιείται μόνο ένα υποσύνολο των δεδομένων. Τα δεδομένα που δεν περιλαμβάνονται (OOB δείγματα) χρησιμοποιούνται σαν ένα εσωτερικό σετ επικύρωσης για να ελεγχθεί η απόδοση του δέντρου. Ο μέσος όρος του σφάλματος στα OOB δείγματα δίνει μια **αντικειμενική εκτίμηση της γενικής ακρίβειας** του μοντέλου, ισοδύναμη με σφάλμα επικύρωσης (validation error) (Cutler et al., 2007).
- **Διασταυρούμενη επικύρωση (cross-validation):** Αν εφαρμοστεί επιπλέον, παρέχει μια πιο σταθερή και γενικεύσιμη εκτίμηση της απόδοσης του μοντέλου σε μη παρατηρημένα δεδομένα. Είναι ιδιαίτερα χρήσιμη για την επιλογή υπερπαραμέτρων, όπως ο αριθμός των δέντρων ή το μέγεθος των υποσυνόλων χαρακτηριστικών (Hastie et al., 2009).

Το γεγονός ότι **κάθε δέντρο εκπαιδεύεται σε διαφορετικό δείγμα** συμβάλλει στην **μείωση της διακύμανσης (variance)** του τελικού μοντέλου, καθώς το μοντέλο γίνεται λιγότερο ευαίσθητο σε αλλαγές στα δεδομένα εκπαίδευσης. Ταυτόχρονα, **η μεροληψία (bias)** παραμένει περιορισμένη, με αποτέλεσμα ένα **σταθερό και αξιόπιστο μοντέλο**, ακόμη και σε σύνθετες ή θορυβώδεις συνθήκες (James et al., 2013).

Ερμηνεία και εγκυρότητα

Η αξιολόγηση της Random Forest δεν περιορίζεται μόνο σε αριθμητικές μετρικές. Είναι απαραίτητο να:

- Ελέγχεται η **ερμηνευσιμότητα** μέσω της σημαντικότητας μεταβλητών (Louppe et al., 2013),
- Επιβεβαιώνεται η **γενίκευση** μέσω OOB σφάλματος ή εξωτερικής επικύρωσης (Hastie et al., 2009),
- Εξετάζεται η **συνοχή των αποτελεσμάτων** με το θεωρητικό πλαίσιο της έρευνας.

Η ερμηνεία των προβλέψεων και η αποδοχή του μοντέλου πρέπει να βασίζεται συνδυαστικά σε ποσοτικά και ποιοτικά κριτήρια, ώστε να διασφαλιστεί ότι το μοντέλο δεν είναι απλώς «στατιστικά αποδοτικό», αλλά και επιστημονικά έγκυρο (Lipton, 2018).

3.8.4 Εξήγηση αποτελεσμάτων

Μετά την αξιολόγηση της απόδοσης του μοντέλου Random Forest και των σχετικών μετρικών ταξινόμησης, ακολουθεί η ερμηνεία των αποτελεσμάτων με βάση τις προβλέψεις του μοντέλου και τη συνεισφορά των επιμέρους χαρακτηριστικών. Στόχος είναι η κατανόηση του τρόπου με τον οποίο οι ανεξάρτητες μεταβλητές επηρεάζουν την πιθανότητα ένας πεζός να βρίσκεται εντός ή εκτός διάβασης.

Ερμηνεία σχέσεων εξαρτημένης και ανεξάρτητων μεταβλητών

Η Random Forest, ως **μη γραμμικό και μη παραμετρικό μοντέλο**, δεν παρέχει συντελεστές παλινδρόμησης με άμεση μαθηματική ερμηνεία, όπως συμβαίνει στη Λογιστική Παλινδρόμηση. Αυτό σημαίνει ότι δεν μπορούμε να προσδιορίσουμε απευθείας το μέγεθος ή την κατεύθυνση της επίδρασης κάθε ανεξάρτητης μεταβλητής στην εξαρτημένη (Breiman, 2001).

Ωστόσο, είναι δυνατή η **ποσοτική εκτίμηση της σημαντικότητας (feature importance)** των μεταβλητών, δηλαδή πόσο συνεισφέρουν στην τελική πρόβλεψη του μοντέλου. Οι σημαντικότερες μεταβλητές είναι εκείνες που προκαλούν τη μεγαλύτερη **μείωση της αβεβαιότητας** στα δέντρα, είτε μέσω του **Gini Impurity** είτε μέσω της **εντροπίας (Entropy)**, σε όλους τους κόμβους όπου χρησιμοποιούνται (Louppe et al., 2013).

Μια **υψηλή τιμή σημασίας** υποδηλώνει ότι η μεταβλητή χρησιμοποιήθηκε συχνά και κρίσιμα στη διαδικασία λήψης αποφάσεων του μοντέλου, άρα συνέβαλε ουσιαστικά στην πρόβλεψη.

Ωστόσο, η **κατεύθυνση της σχέσης** (δηλαδή αν η μεταβλητή αυξάνει ή μειώνει την πιθανότητα του αποτελέσματος) **δεν προκύπτει άμεσα** από το μοντέλο. Για να μελετηθεί αυτό, χρησιμοποιούνται εξειδικευμένες τεχνικές όπως:

- **Partial Dependence Plots (PDPs)**: Γραφήματα που δείχνουν πώς αλλάζει η προβλεπόμενη πιθανότητα καθώς μεταβάλλεται μία συγκεκριμένη μεταβλητή, ενώ οι υπόλοιπες διατηρούνται σταθερές (Friedman, 2001). Έτσι, απομονώνεται η μέση επίδραση της μεταβλητής.
- **SHAP values (SHapley Additive exPlanations)**: Παρέχουν τοπικές και συνολικές εξηγήσεις, δείχνοντας **πώς και προς ποια κατεύθυνση** κάθε μεταβλητή επηρέασε την πρόβλεψη σε κάθε παρατήρηση (Lundberg & Lee, 2017). Προσφέρουν πιο λεπτομερή και αξιόπιστη ερμηνεία σε σχέση με άλλα εργαλεία.

Κεφάλαιο 4: Συλλογή και επεξεργασία στοιχείων

4.1 Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζονται τα δεδομένα που αξιοποιήθηκαν για να διερευνηθούν οι παράγοντες παράνομης διέλευσης των πεζών και συγκεκριμένα οι μεταβλητές που σχετίζονται με την αλληλεπίδραση του πεζού και του οχήματος. Συγκεκριμένα, περιγράφεται η **διαδικασία συλλογής και επεξεργασίας των δεδομένων**, ώστε να δοθεί πλήρης εικόνα της αξιοπιστίας και της ποιότητάς τους. Αρχικά, παρουσιάζεται η **μεθοδολογία καταγραφής των στοιχείων** που χρησιμοποιήθηκαν, ενώ στη συνέχεια δίνεται έμφαση στον **προσδιορισμό της ακρίβειας** των ενδείξεων του φωτεινού σηματοδότη ανά κύκλο εναλλαγής, καθώς και της ακρίβειας στον εντοπισμό νόμιμων και παράνομων διαβάσεων πεζών.

Τα δεδομένα αφορούν τη **διάβαση στην οδό Πανεπιστημίου**, στο ύψος της Πλατείας Ομονοίας, στην Αθήνα, και περιλαμβάνουν:

1. Δεδομένα αλγορίθμου για την αυτόματη καταγραφή παράνομων διαβάσεων πεζών και οχημάτων,
2. Χειροκίνητες καταγραφές από μετρήσεις πεδίου,
3. Την ταχύτητα και τον χρόνο αντίδρασης πεζών και οχημάτων.

Επιπλέον, παρουσιάζονται σε σχετικούς πίνακες τα **περιγραφικά στατιστικά στοιχεία** της βάσης δεδομένων, με σκοπό την κατανόηση της κατανομής, της μεταβλητότητας και των βασικών χαρακτηριστικών των υπό ανάλυση μεταβλητών.

4.2 Συλλογή δεδομένων

4.2.1 Δεδομένα αλγορίθμου

Τα δεδομένα που χρησιμοποιήθηκαν στα πλαίσια της συγκεκριμένης Διπλωματικής Εργασίας συλλέχθηκαν από **καταγραφή βίντεο με χρήση κινητού τηλεφώνου** τοποθετημένο σε τρίποδο σε ύψος διάρκειας **μιας ώρας αιχμής**. Η τοποθεσία που επιλέχθηκε είναι στο **κέντρο της Αθήνας**, στο τέλος της οδού Πανεπιστημίου, πριν από την πλατεία Ομονοίας.

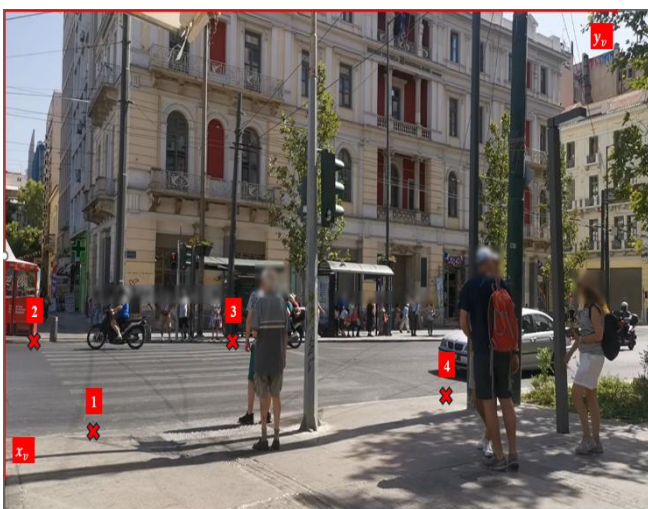
Το βίντεο που καταγράφηκε εισήχθη στον αλγόριθμο που έχει υλοποιηθεί από την ερευνητική ομάδα του εργαστηρίου Κυκλοφοριακής Τεχνικής (Ventura et al., 2025), βασισμένος στην **υπολογιστική όραση** (computer vision), και χρησιμοποιείται για την παρακολούθηση πεζών και οχημάτων, την ανάλυση της συμπεριφοράς τους και την εκτίμηση του Χρόνου έως τη Σύγκρουση (Time-to-Collision - TTC). Ο αλγόριθμος αποτελείται από τα παρακάτω:

1. Χρήση των μοντέλων **YOLOv8** για ανίχνευση αντικειμένων και **ResNet-50** για εξαγωγή χαρακτηριστικών.
2. Εφαρμογή **δικτύου Re-Identification (Re-ID)** για αναγνώριση και διατήρηση των ταυτοτήτων των αντικειμένων.

3. Ορισμός **ομογραφίας** (homography) για μετατροπή των συντεταγμένων των αντικειμένων από το επίπεδο του βίντεο στο πραγματικό επίπεδο του δρόμου.
4. Πρόβλεψη κίνησης με **Kalman Filter** για να μειώσει το θόρυβο και να διατηρήσει σταθερότητα στην παρακολούθηση.
5. Αντιστοίχιση αντικειμένων μεταξύ διαδοχικών καρέ με χρήση του **Hungarian Algorithm**.
6. Φιλτράρισμα θέσεων σε χαμένα αντικείμενα με **Savitzky-Golay Filter**.
7. Ανίχνευση της κατάστασης του φωτεινού σηματοδότη μέσω ανάλυσης χρωμάτων σε συγκεκριμένες **περιοχές ενδιαφέροντος** (Region of Interest).
8. **Ανάλυση συμπεριφοράς πεζών και οχημάτων**, εντοπίζοντας παραβάσεις σηματοδότησης (π.χ. παράνομες διαβάσεις).
9. **Καταμέτρηση πεζών και οχημάτων** σε διάφορα σημεία διέλευσης.
10. Υπολογισμός ταχύτητας αντικειμένων και εκτίμηση **Χρόνου έως Σύγκρουση** (TTC).

Η ανάλυση του βίντεο πραγματοποιήθηκε έχοντας ορίσει ως ανώτατο όριο **30 καρέ (Frames per Second - fps)**, ώστε να αποφευχθεί η υπερφόρτωση της μνήμης.

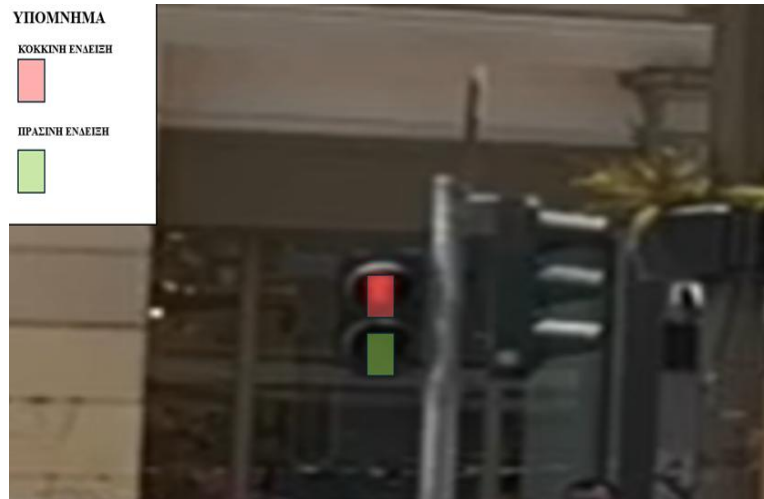
Η εικόνα (Εικόνα 1) δείχνει τα **σημεία αναφοράς** στην προοπτική του βίντεο, ενώ η δεύτερη εικόνα (Εικόνα 2) δείχνει τα αντίστοιχα σημεία στο **επίπεδο του χάρτη**. Αυτά τα σημεία είναι απαραίτητα για την προσαρμογή των συντεταγμένων μεταξύ των δύο επιπέδων, επιτρέποντας την ακριβή αντιστοίχιση των δεδομένων μεταξύ βίντεο και πραγματικού κόσμου. Ουσιαστικά, χρησιμοποιούνται για την ακριβή μετατροπή και χαρτογράφηση της θέσης αντικειμένων στο βίντεο σε πραγματικές συντεταγμένες.



Εικόνα 1: Σημεία αναφοράς βίντεο 1



Εικόνα 2: Σημεία αναφοράς στο επίπεδο χάρτη

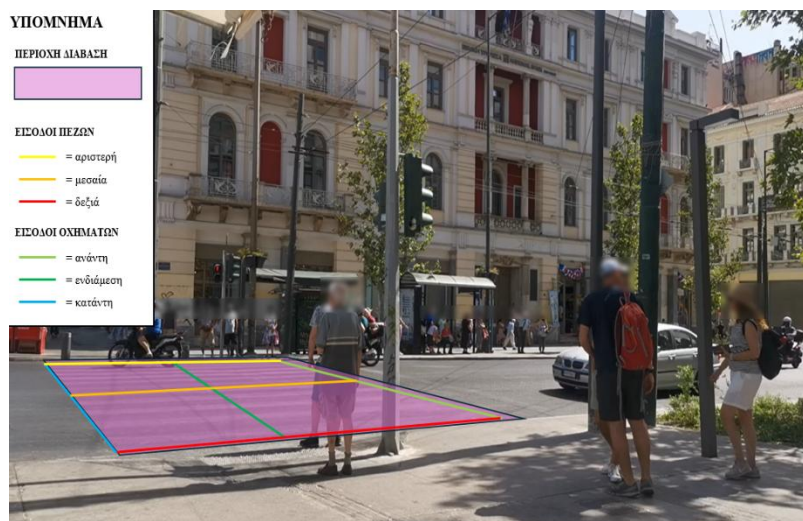


Εικόνα 3: Φωτεινοί Σηματοδότες Πεζών

Στην ανίχνευση της κατάστασης φωτεινών σηματοδοτών πεζών, καθορίστηκαν **περιοχές ενδιαφέροντος** (Region Of Interest -ROIs) για τα **κόκκινα και πράσινα φώτα** του σηματοδότη (Εικόνα 3). Παρόμοια, στην ανάλυση της συμπεριφοράς πεζών και οχημάτων, καθορίστηκε η περιοχή ενδιαφέροντος της **διάβασης πεζών** (Εικόνα 4).

Για την καταμέτρηση πεζών και οχημάτων, ορίστηκαν **έξι «είσοδοι» ανίχνευσης** για κάθε σημείο: τρεις για πεζούς και τρεις για οχήματα. Οι είσοδοι τοποθετήθηκαν στη μέση της διάβασης και στις άκρες της, ενώ οι είσοδοι των οχημάτων τοποθετήθηκαν στη μέση, ανάντη και κατάντη της διάβασης (Εικόνα 4). Ο στόχος ήταν η διασταύρωση των μετρήσεων για αξιόπιστα αποτελέσματα και η αποφυγή υπερ- ή υπο-καταμέτρησης.

Οι **μέγιστες επιτρεπτές ταχύτητες** ορίστηκαν σε 10 m/s για πεζούς (ταχύτητα τρεξίματος) και 40 m/s για οχήματα, ώστε να ληφθούν υπόψη σενάρια υπερβολικής ταχύτητας αλλά να αποφευχθούν ακραίες, μη ρεαλιστικές τιμές. Αυτές οι ρυθμίσεις διασφαλίζουν ότι το σύστημα είναι σταθερό, αποδοτικό και προσαρμόζεται στις διαφορετικές συνθήκες του περιβάλλοντος.



Εικόνα 4: Περιοχή Ενδιαφέροντος Διάβασης Πεζών

Η συλλογή υλικού μέσω βιντεοσκόπησης της εξεταζόμενης διάβασης αποτέλεσε τη βάση για την ανάπτυξη μιας ολοκληρωμένης μεθοδολογίας ανάλυσης δεδομένων. Μέσω προηγμένων τεχνικών επεξεργασίας στο **λογισμικό MATLAB**, πραγματοποιήθηκε ανάλυση και επεξεργασία των καταγραφών, οδηγώντας στη δημιουργία της **βάσης δεδομένων της αυτοματοποιημένης μεθόδου**. Παράλληλα, το ίδιο υλικό παρακολουθήθηκε λεπτομερώς, καταγράφοντας τα δεδομένα **χειροκίνητα**. Με αυτόν τον τρόπο, διαμορφώθηκε τη **βάση δεδομένων της χειροκίνητης μεθόδου**.

Από τον αυτοματοποιημένο αλγόριθμο χρησιμοποιήθηκαν 3 βάσεις δεδομένων, Pedestrian Data, Vehicle Data και MinTTC Data οι οποίες ενοποιήθηκαν σε μια βάση ώστε να μπορέσει να γίνει η ανάλυση πιο γρήγορα και σωστά, όπως θα περιγραφεί στη συνέχεια. Οι κύριες μεταβλητές από την κάθε βάση δεδομένων της αυτοματοποιημένης μεθόδου που θα χρησιμοποιηθούν στις παρακάτω αναλύσεις είναι:

1. το **Video Part** όπου αντιπροσωπεύει το τμήμα του βίντεο από το οποίο προέρχονται τα δεδομένα,
2. το **Frame** όπου αναπαριστά το στιγμιότυπο του βίντεο από το οποίο καταγράφηκε η συγκεκριμένη εγγραφή,
3. το **Pedestrian ID** όπου αποτελεί το μοναδικό αναγνωριστικό για κάθε πεζό που ανιχνεύεται στο βίντεο,
4. το **Vehicle ID** όπου αποτελεί το μοναδικό αναγνωριστικό για κάθε όχημα που ανιχνεύεται στο βίντεο,
5. τη στήλη **Is In Crossing Area** που δείχνει αν ο πεζός ή το όχημα βρίσκεται εντός ή εκτός της περιοχής διάβασης,
6. την στήλη **Traffic Light Status** που αντιπροσωπεύει την κατάσταση του φωτεινού σηματοδότη πεζών τη δεδομένη χρονική στιγμή,
7. τη στήλη **Confidence** που εκφράζει το βαθμό βεβαιότητας του συστήματος ότι η ανίχνευση αντιστοιχεί πράγματι σε πεζό ή όχημα,
8. τις στήλες **GroundPlaneCentroid_1** και **GroundPlaneCentroid_2** οι οποίες αντιπροσωπεύουν τις συντεταγμένες (x, y) του κεντροειδούς της ανίχνευσης πεζού ή οχήματος στο επίπεδο του εδάφους,
9. τη στήλη **Pedestrian Status** η οποία δείχνει αν η διέλευση του πεζού είναι νόμιμη, παράνομη ή άγνωστη,
10. τη στήλη **Vehicle Status** η οποία δείχνει αν η διέλευση του οχήματος είναι νόμιμη, παράνομη ή άγνωστη,
11. τις στήλες **vx** και **vy** που αντιστοιχούν στις επιμέρους συνιστώσες της ταχύτητας του πεζού ή του οχήματος στους άξονες x και y,
12. τη στήλη **magnitude** η οποία εκφράζει το συνολικό μέτρο της ταχύτητας του πεζού ή του οχήματος, υπολογιζόμενο βάσει των δύο παραπάνω συνιστωσών (vx και vy),
13. τη στήλη **minTTC**, η οποία δείχνει το ελάχιστο εκτιμώμενο χρονικό διάστημα μέχρι πιθανή σύγκρουση με άλλο πεζό ή όχημα.

Το excel το οποίο προέκυψε **χειροκίνητα**, και του οποίου τα δεδομένα θα αξιοποιηθούν παρακάτω αποτελείται από τις στήλες:

1. **Traffic Light**, όπου αναφέρεται στην κατάσταση του φωτεινού σηματοδότη τη δεδομένη χρονική στιγμή,

2. **Time (second)**, που καταγράφει τον χρόνο σε δευτερόλεπτα από την αρχή της καταγραφής,
3. **Illegal Crossings**, που καταγράφει τις παράνομες διελεύσεις πεζών, δηλαδή περιπτώσεις όπου οι πεζοί διέσχισαν τη διάβαση χωρίς να έχουν δικαίωμα,
4. **Illegal on the Intergreen (in the middle of the zebra section)**, η οποία αναφέρεται στους πεζούς που βρίσκονται στη μέση της διάβασης κατά τη φάση του intergreen, δηλαδή κατά τη μεταβατική περίοδο μεταξύ κόκκινου και πράσινου φωτός, και
5. **Legal Crossings** όπου αφορά στον αριθμό των νόμιμων διελεύσεων πεζών, δηλαδή εκείνων που πέρασαν τη διάβαση όταν το φανάρι τούς το επέτρεπε.

Όλα όσα προαναφέρθηκαν αξιοποιήθηκαν κατά τρόπο που θα αναλυθεί παρακάτω με τη χρήση της **γλώσσας προγραμματισμού Python**, εφαρμόζοντας κατάλληλες βιβλιοθήκες και εργαλεία για την επεξεργασία, ανάλυση και οπτικοποίηση των δεδομένων.

4.3 Στατιστική αποτύπωση των δεδομένων της μελέτης

4.3.1 Εισαγωγή

Στο πλαίσιο της παρούσας εργασίας πραγματοποιήθηκε **στατιστική αποτύπωση** των μεταβλητών που περιλαμβάνονται στις τρεις επιμέρους βάσεις δεδομένων: πεζών, οχημάτων και δεδομένων ελάχιστου χρονικού περιθωρίου πριν από πιθανή σύγκρουση. Για τις **αριθμητικές μεταβλητές** υπολογίστηκαν βασικά μεγέθη, όπως η **μέση τιμή** (ενδεικτικό του κεντρικού επιπέδου των τιμών), η **τυπική απόκλιση** (ένδειξη της διασποράς τους), τα **ποσοστημόρια** (για την κατανόηση της κατανομής) και οι **ακραίες τιμές** (ελάχιστο και μέγιστο). Αντίστοιχα, για τις **κατηγορικές και δυαδικές μεταβλητές** υπολογίστηκαν οι **απόλυτες και σχετικές συχνότητες εμφάνισης** των κατηγοριών τους, περιλαμβανομένων και των περιπτώσεων αβεβαιότητας (π.χ. τιμές unknown), ώστε να αξιολογηθεί η πληρότητα των δεδομένων. Η παρουσίαση των αποτελεσμάτων οργανώθηκε σε πίνακες, με στόχο την παροχή μιας καθαρής εικόνας της δομής, της ποιότητας και της ποικιλίας των διαθέσιμων δεδομένων πριν την εφαρμογή περαιτέρω αναλυτικών μεθόδων.

4.3.2 Στατιστική αποτύπωση των μεταβλητών της βάσης δεδομένων πεζών

Η ανάλυση των αριθμητικών μεταβλητών που αφορούν τους πεζούς αναδεικνύει ικανοποιητική ποιότητα δεδομένων ως προς την εμπιστοσύνη ανίχνευσης και τα κινητικά χαρακτηριστικά. Η μεταβλητή Confidence εμφανίζει υψηλή συνοχή και περιορισμένη διασπορά, υποδεικνύοντας αξιόπιστη αναγνώριση. Οι ταχύτητες (vx, vy) και η συνολική ταχύτητα (magnitude) φανερώνουν χαμηλά επίπεδα κινητικότητας, με σημαντικό ποσοστό παρατηρήσεων να αντιστοιχεί σε ακίνητους πεζούς. Αντιθέτως, οι χωρικές μεταβλητές (GroundPlaneCentroid_1 και _2) παρουσιάζουν **έντονη παρουσία απομακρυσμένων τιμών**. Συγκεκριμένα, στο GroundPlaneCentroid_1 παρατηρείται μια μέγιστη τιμή της τάξης των 7900, η οποία υπερβαίνει κατά πολύ την κεντρική τάση του δείγματος, ενώ στο GroundPlaneCentroid_2 καταγράφεται μια εξαιρετικά ακραία αρνητική τιμή (-184.949). Και οι δύο περιπτώσεις συνιστούν στατιστικά **outliers**, τα οποία **χρηζουν προσοχής** κατά τη στατιστική ερμηνεία και

ενδεχομένως επηρεάζουν τη συνολική εικόνα των δεδομένων. Η παρουσία τους δεν αναιρεί την αξία του συνόλου, αλλά επισημαίνει την ανάγκη για διερεύνηση και πιθανό φιλτράρισμα στις επόμενες φάσεις της ανάλυσης. Τα αποτελέσματα παρουσιάζονται στον επισυναπτόμενο Πίνακα 4.1:

Μεταβλητή	Πλήθος (count)	Μέση Τιμή (mean)	Τυπ. Απόκλιση (std)	Ελάχιστο (min)	25%	Διάμεσος (50%)	75%	Μέγιστο (max)
Confidence	622,700	0.729	0.113	0.500	0.637	0.744	0.827	0.966
V _x	653,660	-0.167	1.141	-9.985	-0.164	0	0.078	9.663
V _y	653,660	0.078	1.043	-9.818	-0.081	0	0.146	9.977
magnitude	653,660	1.273	1.810	0	0	0.656	1.624	9.999
GroundPlaneCentroid_1	653,660	15.413	124.229	-86034.09	10.483	13.071	16.071	7900.479
GroundPlaneCentroid_2	653,660	12.914	260.038	-184948.9	3.965	10.211	20.197	17322.57

Πίνακας 4.1: Στατιστικά αριθμητικών μεταβλητών πεζών

Η κατανομή των κατηγορικών μεταβλητών αναδεικνύει κρίσιμα πρότυπα συμπεριφοράς πεζών στον αστικό χώρο. Ποσοστό μεγαλύτερο του **84%** καταγράφηκε **εκτός διαβάσεων** (IsInCrossingArea), γεγονός που πιθανόν αντανακλά ελλιπή χρήση ή έλλειψη διαθεσιμότητας ασφαλών σημείων διέλευσης. Παράλληλα, σχεδόν **60%** των παρατηρήσεων αφορούν χρονικές στιγμές με **κόκκινο φωτεινό σήμα** (TrafficLightStatus), ενώ μόνο το **26%** αντιστοιχεί σε **πράσινο**, στοιχείο που μπορεί να σχετίζεται με περιορισμένη χρονική διάρκεια ασφαλούς διέλευσης ή αναμονής. Παρά την υπεροχή των αντίξων συνθηκών, η πλειονότητα των πεζών (PedestrianStatus) εμφανίζεται να **δρα εντός του νομικού πλαισίου**, με **92%** των περιπτώσεων να χαρακτηρίζονται ως νόμιμες. Η παρουσία μικρού αλλά υπαρκτού ποσοστού παραβατικής συμπεριφοράς υποδεικνύει την ανάγκη για στοχευμένες παρεμβάσεις ευαισθητοποίησης ή υποδομής, χωρίς ωστόσο να αμαυρώνει τη γενική εικόνα συμμόρφωσης. Τα αποτελέσματα εμφανίζονται στον παρακάτω Πίνακα 4.2.

Μεταβλητή	Παρουσία πεζών		Κατάσταση φωτεινού σηματοδότη			Νομιμότητα διέλευσης πεζών		
Τιμή	0	1	red	green	unknown	legal	illegal	unknown
Πλήθος	550,085	103,575	391,906	169,245	92,509	600,376	44,568	8,716
Ποσοστό	84.15%	15.85%	59.96%	25.89%	14.15%	91.85%	6.82%	1.33%

Πίνακας 4.2: Στατιστικά κατηγορικών μεταβλητών πεζών

4.3.3 Στατιστική αποτύπωση των μεταβλητών της βάσης δεδομένων οχημάτων

Η ανάλυση των αριθμητικών μεταβλητών που σχετίζονται με τα οχήματα αποκαλύπτει σημαντικές διαφοροποιήσεις σε σχέση με εκείνες των πεζών, αντανακλώντας τις διαφορετικές δυναμικές της κίνησης. Μέσες τιμές των συνιστωσών **ταχύτητας** είναι αρνητική στον άξονα x (-2.74) και θετική στον άξονα y (0.78), στοιχείο που υποδηλώνει ότι τα οχήματα στο σύνολο της καταγραφής **κινούνται προσανατολισμένα** προς μια συγκεκριμένη κατεύθυνση στο χώρο. Η κατεύθυνση αυτή είναι συνεπής και δεν αποτελεί σφάλμα, αλλά **χαρακτηριστικό της γεωμετρίας** ή της ροής του οδικού περιβάλλοντος όπου έγινε η καταγραφή.

Αντιθέτως, οι μεταβλητές **θέσης στο επίπεδο** του εδάφους (GroundPlaneCentroid_1 και _2) εμφανίζουν σαφείς **ενδείξεις ακραίων τιμών**. Συγκεκριμένα, στο πρώτο επίπεδο καταγράφεται μέγιστο άνω των 410.000, όταν η διάμεσος είναι μόλις 22.14 και η τυπική απόκλιση 693. Αντίστοιχα, στο δεύτερο επίπεδο, το μέγιστο ξεπερνά τις

137.000, με ελάχιστο στις -5.466. Οι αποκλίσεις αυτές υπερβαίνουν κατά χιλιάδες φορές την κεντρική τάση και συνιστούν **ισχυρές ενδείξεις outliers ή σφαλμάτων καταγραφής**. Οι ακραίες αυτές τιμές απαιτούν ιδιαίτερη προσοχή, καθώς ενδέχεται να επηρεάσουν δυσανάλογα τα αποτελέσματα σε στατιστικά υποδείγματα, αν δεν αντιμετωπιστούν κατάλληλα (μέσω καθαρισμού ή κανονικοποίησης). Τα αποτελέσματα βρίσκονται στον Πίνακα 4.3:

Μεταβλητή	Πλήθος (count)	Μέση Τιμή (mean)	Τυπ. Απόκλιση (std)	Ελάχιστο (min)	25%	Διάμεσος 50%	75%	Μέγιστο (max)
Confidence	335,795	0.805	0.116	0.5	0.735	0.8470	0.897	0.971
GroundPlane Centroid_1	351,777	26.350	693.036	-16,459.08	16.077	22.137	30.648	410,073.30
GroundPlane Centroid_2	351,777	10.854	233.811	-5,466.88	6.550	9.456	12.860	137,895.90
V _x	351,777	-2.741	4.536	-39.420	-4.583	-2.196	-0.038	38.941
V _y	351,777	0.777	2.549	-38.145	0	0.595	1.631	38.952
magnitude	351,777	6.288	6.019	0	1.653	5.144	8.833	39.968

Πίνακας 4.3: Στατιστικά αριθμητικών μεταβλητών οχημάτων

Η ανάλυση των κατηγορικών μεταβλητών που σχετίζονται με τα οχήματα αποκαλύπτει χαρακτηριστικά πρότυπα κυκλοφοριακής συμπεριφοράς και συνθηκών καταγραφής. Η μεταβλητή *IsInCrossingArea* δείχνει ότι η πλειονότητα των οχημάτων **(80,4%)** καταγράφηκε **εκτός περιοχών διάβασης πεζών**, γεγονός απολύτως αναμενόμενο, καθώς οι διαβάσεις αποτελούν τμήματα του οδικού χώρου με ειδική χρήση, κατά κανόνα όχι για συνεχή διέλευση οχημάτων. Η παρουσία ποσοστού **19,6% εντός των ορίων διάβασης** αποδίδεται στις περιπτώσεις διάσχισης της διάβασης από τους πεζούς.

Η μεταβλητή *TrafficLightStatus* φανερώνει ότι περίπου **60%** των οχημάτων καταγράφηκαν κατά τη φάση **κόκκινου σηματοδότη**, ενώ μόλις το **22,9%** κατά τη φάση του **πράσινου**. Ένα επιπλέον **17%** χαρακτηρίζεται ως **άγνωστο**. Η επικράτηση της κόκκινης φάσης οφείλεται στην πραγματική χρονική κατανομή των σηματοδοτικών κύκλων, δηλαδή στη μεγαλύτερη διάρκεια της κόκκινης φάσης σε σχέση με την πράσινη στο υπό μελέτη περιβάλλον. Το γεγονός αυτό είναι ουσιώδες, καθώς επηρεάζει τη στατιστική αναλογία των φάσεων χωρίς να υποδηλώνει απαραίτητα κυκλοφοριακή παραβατικότητα.

Τέλος, η μεταβλητή *VehicleStatus* δείχνει ότι το **93,5%** των οχημάτων καταγράφηκαν να κινούνται **εντός του νομικού πλαισίου**, με τα ποσοστά παραβατικότητας να παραμένουν χαμηλά (**3,2% “illegal” και 3,3% “unknown”**). Η γενική εικόνα είναι αυτή της υψηλής συμμόρφωσης με τους κανόνες κυκλοφορίας, στοιχείο που μπορεί να αντανάκλα είτε τις συνθήκες οδικής εποπτείας στο σημείο καταγραφής είτε την αποδοτικότητα του συστήματος ανίχνευσης στην αναγνώριση νόμιμων συμπεριφορών. Τα αποτελέσματα είναι στον παρακάτω Πίνακα 4.4.

Μεταβλητή	Παρουσία οχημάτων		Κατάσταση φωτεινού σηματοδότη			Νομιμότητα διέλευσης οχημάτων		
Τιμή	0	1	red	green	unknown	legal	Illegal	unknown
Πλήθος	282,753	69,024	211,414	80,472	59,891	329,055	11,177	11,545
Ποσοστό	80.38%	19.62%	60.10%	22.88%	17.03%	93.54%	3.18%	3.28%

Πίνακας 4.4: Στατιστικά κατηγορικών μεταβλητών οχημάτων

4.3.4 Στατιστική αποτύπωση των μεταβλητών της βάσης δεδομένων του ελάχιστου χρόνου ως την πιθανή σύγκρουση

Η ανάλυση των αριθμητικών μεταβλητών του συνόλου δεδομένων **MinTTC** αποκαλύπτει κρίσιμες πληροφορίες για τη δυναμική αλληλεπίδρασης πεζών και οχημάτων υπό συνθήκες πιθανής σύγκρουσης. Η μεταβλητή **MinTTC** εμφανίζει **ακραίες τιμές**, με την ύπαρξη **απειρών (inf)** να οδηγεί σε μη υπολογίσιμη μέση τιμή και τυπική απόκλιση. Η παρουσία τέτοιων τιμών είναι αναμενόμενη σε περιπτώσεις όπου η τροχιά του πεζού και του οχήματος δεν τέμνονται ή δεν συγκλίνουν χρονικά, υποδηλώνοντας ότι η σύγκρουση δεν θεωρείται πιθανή. Η κατανομή της μεταβλητής είναι έντονα ασύμμετρη, με το **25% των παρατηρήσεων** να συγκεντρώνεται **κάτω από τα 17 δευτερόλεπτα** και τη **διάμεσο** να ανέρχεται στα **40**, υποδεικνύοντας **σημαντικό εύρος χρόνου αντίδρασης** σε αρκετές περιπτώσεις.

Όσον αφορά τις **ταχύτητες**, οι πεζοί παρουσιάζουν **χαμηλή κινητικότητα**, με τουλάχιστον το **25%** να είναι **ακίνητοι**, και **μέση ταχύτητα** μόλις **1.43**, γεγονός αναμενόμενο σε περιβάλλοντα όπου συνήθως παρατηρείται παύση κίνησης (π.χ. αναμονή για διάβαση). Τα **οχήματα**, αντίθετα, κινούνται με **μέση ταχύτητα 8.31** και **μέγιστη περίπου 40**, τιμές απολύτως ρεαλιστικές για αστικό περιβάλλον. Γενικά, οι μεταβλητές ταχύτητας χαρακτηρίζονται από φυσιολογική κατανομή, με συνέπεια και ερμηνευσιμότητα.

Συνολικά, τα δεδομένα εμφανίζουν ποιοτική κινηματική πληροφόρηση, με μοναδική έντονη τεχνική πρόκληση την παρουσία απείρων ή μη έγκυρων τιμών στο **MinTTC**, οι οποίες χρήζουν προσεκτικής διαχείρισης σε επόμενο στάδιο της ανάλυσης. Τα αποτελέσματα οπτικοποιούνται στον επισυναπτόμενο Πίνακα 4.5:

Μεταβλητές	Πλήθος (count)	Μέση τιμή(mean)	Τυπική Απόκλιση(std)	Ελάχιστο(min)	25%	Διάμεσος 50%	75%	Μέγιστο(max)
MinTTC	100408	inf	-	0.06	16.864	40.699	134.65	inf
Confidence_Pedestrian	82263	0.694	0.198	0	0.614	0.741	0.828	0.956
Confidence_Vehicle	82263	0.785	0.157	0	0.712	0.838	0.894	0.964
PedestrianSpeedMagnitude	82263	1.428	1.924	0	0	0.839	1.855	9.999
VehicleSpeedMagnitude	82263	8.314	6.638	0	3.825	7.221	10.973	39.960

Πίνακας 4.5: Στατιστικά αριθμητικών μεταβλητών ελάχιστου χρόνου ως τη σύγκρουση

Η κατανομή των κατηγορικών μεταβλητών στο σύνολο δεδομένων **MinTTC** αποτυπώνει σαφώς τον τρόπο με τον οποίο κατεγράφησαν τα επεισόδια πιθανής σύγκρουσης μεταξύ πεζών και οχημάτων. Σε όλες τις μεταβλητές, εμφανίζεται σταθερά ένα ποσοστό **18,07% με ελλιπή τιμή (NaN)**, το οποίο αφορά περιπτώσεις όπου δεν υπήρξε πλήρες ζεύγος πεζού και οχήματος, και άρα δεν καταγράφηκε επαρκώς η σχετική κατάσταση. Αυτό το ποσοστό πρέπει να ληφθεί υπόψη ως εν δυνάμει περιορισμός της ανάλυσης.

Η μεταβλητή **IsInCrossingArea_Pedestrian** δείχνει ότι μόλις **7,26%** των πεζών βρίσκονταν **εντός διάβασης** τη στιγμή της εκτίμησης του **MinTTC**, ενώ το **74,67%** **εκτός**. Η εικόνα είναι παρόμοια και για τα οχήματα (**IsInCrossingArea_Vehicle**), με **16,62% εντός διάβασης** και **65,30% εκτός**. Η ασυμμετρία αυτή υποδεικνύει ότι η

πλειονότητα των δυνητικά επικίνδυνων αλληλεπιδράσεων δεν έλαβε χώρα σε τυπικά προστατευμένους χώρους διέλευσης, αλλά σε περιβάλλοντα χωρίς προτεραιότητα ή σαφή οριοθέτηση κυκλοφορίας.

Σε σχέση με τη **συμμόρφωση με το νομικό πλαίσιο**, οι περισσότερες καταγραφές αφορούν νόμιμη συμπεριφορά: **78,17% των πεζών** και **76,65% των οχημάτων** χαρακτηρίζονται ως **"legal"**. Η παραβατικότητα είναι σχετικά χαμηλή, με **3,05% των πεζών** και **2,85% των οχημάτων** να καταγράφονται ως **"illegal"**, ενώ υπάρχει και ένα μικρό ποσοστό **"unknown"** (0,72% και 2,43% αντίστοιχα). Οι τιμές αυτές υποδεικνύουν ότι οι καταγεγραμμένες επικίνδυνες καταστάσεις δεν οφείλονται απαραίτητα σε παραβατική συμπεριφορά, αλλά ενδέχεται να αντικατοπτρίζουν δομικούς κινδύνους της κυκλοφοριακής αλληλεπίδρασης, ακόμη και όταν όλα τα μέρη τηρούν τους κανόνες.

Συνολικά, τα ευρήματα υποδεικνύουν ένα πλαίσιο όπου η **μεγάλη πλειονότητα των δυνητικών συγκρούσεων εξελίσσεται σε μη προστατευμένους χώρους** και παρά τη **σχεδόν καθολική συμμόρφωση με τους κανόνες κυκλοφορίας**, αναδεικνύεται η σημασία της χωρικής και λειτουργικής αναδιοργάνωσης του περιβάλλοντος για την πρόληψη των συγκρούσεων. Τα αποτελέσματα οπτικοποιούνται στον επισυναπτόμενο Πίνακα 4.6:

Μεταβλητή	Παρουσία πεζών/οχημάτων						Νομιμότητα διέλευσης οχημάτων / πεζών							
Πεζός / Όχημα	Όχημα			Πεζός			Όχημα				Πεζός			
Τιμή	0	1	NaN	0	1	NaN	legal	illegal	NaN	unknown	legal	illegal	NaN	unknown
Πλήθος	65,571	16,692	18,145	74,976	7,287	18,145	76,964	2,859	18,145	2,440	78,485	3,058	18,145	720
Ποσοστό	65.30%	16.62%	18.07%	74.67%	7.26%	18.07%	76.65%	2.85%	18.07%	2.43%	78.17%	3.05%	18.07%	0.72%

Πίνακας 4.6: Στατιστικά κατηγορικών μεταβλητών ελάχιστου χρόνου ως τη σύγκρουση

4.4 Ανάλυση της ακρίβειας των ενδείξεων του φωτεινού σηματοδότη ανά κύκλο εναλλαγής

4.4.1 Εισαγωγή

Η ανάλυση της ακρίβειας των ενδείξεων του φωτεινού σηματοδότη βασίστηκε στη **σύγκριση** μεταξύ της **αυτοματοποιημένης** και της **χειροκίνητης** μεθόδου. Για την εξασφάλιση της εγκυρότητας των αποτελεσμάτων, κρίθηκε απαραίτητη η **προεπεξεργασία των δεδομένων** της αυτοματοποιημένης μεθόδου, ώστε να εναρμονιστούν οι διαφορετικές μορφές καταγραφής και να διευκολυνθεί η σύγκριση.

4.4.2 Ενοποίηση και κανονικοποίηση αρίθμησης καρέ

Δεδομένου ότι το βίντεο που επεξεργάστηκε το λογισμικό ήταν χωρισμένο σε **διαφορετικά τμήματα** (Video Parts), κρίθηκε απαραίτητο να δημιουργηθεί ένα **ενιαίο σύστημα αρίθμησης των καρέ** (frames) για τη διατήρηση της χρονικής συνέχειας. Για τον σκοπό αυτό, κάθε τμήμα αντιστοιχίστηκε με έναν κατάλληλο αντισταθμιστικό δείκτη (offset), ώστε να εξασφαλιστεί η ομαλή μετάβαση μεταξύ των διαφορετικών τμημάτων και η αποφυγή επικαλύψεων ή κενών στην αρίθμηση των καρέ. Το αποτέλεσμα αυτής της διαδικασίας ήταν η δημιουργία της μεταβλητής **Continuous Frame**, η οποία αποδίδει έναν μοναδικό και συνεχόμενο αριθμό σε κάθε καρέ του βίντεο, ανεξάρτητα από το τμήμα στο οποίο ανήκει.

4.4.3 Μετατροπή καρέ σε μονάδες χρόνου

Για τη διευκόλυνση της **χρονικής σύγκρισης** και ανάλυσης, τα καρέ μετατράπηκαν σε μονάδες χρόνου. Ειδικότερα, δημιουργήθηκε η στήλη **Seconds** μέσω της διαίρεσης του αριθμού των καρέ με τον ρυθμό καταγραφής του βίντεο (29.74 frames/sec, περίπου 30 fps), παρέχοντας μια ακριβή αναπαράσταση του χρόνου σε δευτερόλεπτα. Επιπλέον, δημιουργήθηκε η στήλη **Minutes**, η οποία προέκυψε από τη μετατροπή των δευτερολέπτων σε λεπτά, προσφέροντας μια ευρύτερη χρονική κλίμακα για ανάλυση. Για λόγους ακρίβειας, τα δευτερόλεπτα στρογγυλοποιήθηκαν στην πλησιέστερη ακέραια τιμή, ενώ στη συνέχεια αναπροσαρμόστηκαν, ώστε η αρίθμηση τους να ξεκινά από το 1. Με αυτόν τον τρόπο, διασφαλίστηκε η συνέπεια στη χρονική αναφορά, διατηρώντας μια λογική και σταθερή βάση μέτρησης του χρόνου.

4.4.4 Αντιμετώπιση άγνωστων και ελλιπών τιμών

Για τη διασφάλιση της αξιοπιστίας και της συνέπειας των δεδομένων, εφαρμόστηκε **διαδικασία καθαρισμού** και εξομάλυνσης των τιμών στη στήλη **Traffic Light Status**. Αρχικά, όλα τα δεδομένα μετατράπηκαν σε πεζούς χαρακτήρες, ώστε να αποφευχθούν διαφοροποιήσεις λόγω μορφοποίησης, και ταξινομήθηκαν με βάση τη μεταβλητή Continuous Frame, διατηρώντας τη σωστή χρονική σειρά κατά την επεξεργασία. Προκειμένου να αντιμετωπιστούν **άγνωστες ή μη καταγεγραμμένες τιμές** (unknown ή NaN), ελέγχθηκε αν αυτές εμφανίζονταν μεταξύ δύο διαδοχικών πλαισίων με την ίδια ένδειξη φωτεινού σηματοδότη (π.χ. πράσινο ή κόκκινο). Στις περιπτώσεις αυτές, οι εν λόγω τιμές αντικαταστάθηκαν από την **κυρίαρχη ένδειξη**, εξασφαλίζοντας τη συνέπεια στη χρονοσειρά. Επιπλέον, εάν οι άγνωστες ή ασαφείς τιμές διαρκούσαν για **παρατεταμένη χρονική περίοδο** (έως 750 καρέ) και περιβάλλονταν από την ίδια τιμή φωτεινού σηματοδότη πριν και μετά, τότε αντικαθίσταντο και πάλι από την επικρατούσα τιμή. Η διαδικασία αυτή συνέβαλε στη βελτίωση της ποιότητας των δεδομένων, διασφαλίζοντας μεγαλύτερη ακρίβεια στην ανάλυση της αυτοματοποιημένης μεθόδου και επιτρέποντας μια πιο αξιόπιστη σύγκριση των αποτελεσμάτων.

4.4.5 Ορισμός κύκλων φωτεινής σηματοδότησης

Για την ανάλυση των δεδομένων, ήταν απαραίτητο να καθοριστούν οι **κύκλοι λειτουργίας του φωτεινού σηματοδότη** και να αντιστοιχιστεί η χρονική στιγμή κάθε εγγραφής με τον σωστό κύκλο και τη φάση στην οποία ανήκει. Αρχικά, προσδιορίστηκε ο πρώτος κύκλος σηματοδότησης, ο οποίος διέφερε από τους

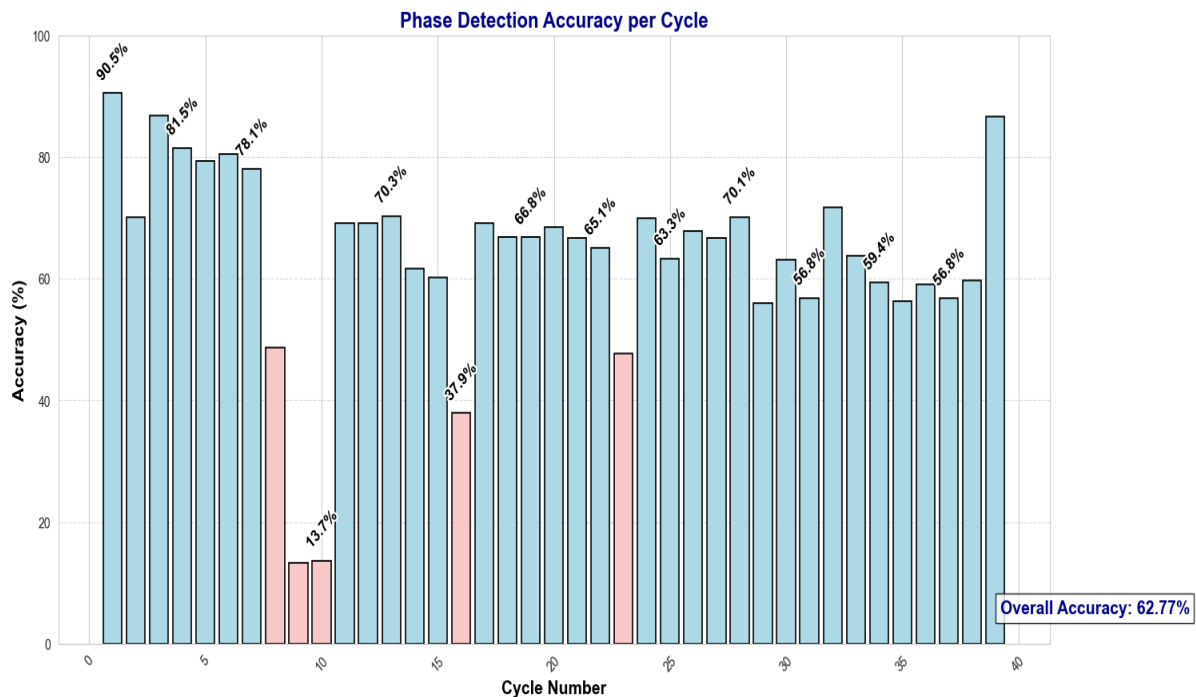
υπόλοιπους, καθώς η αρχική διάρκεια της **κόκκινης φάσης** ήταν μόλις **18 δευτερόλεπτα**. Αυτό οφείλεται στο γεγονός πως η έναρξη καταγραφής βίντεο δεν ήταν συγχρονισμένη με τις φάσεις διάρκειας του φωτεινού σηματοδότη. Στη συνέχεια, ορίστηκαν οι σταθερές διάρκειες για τις επόμενες φάσεις του σηματοδότη: **10 δευτερόλεπτα** για το **πράσινο φως**, **13 δευτερόλεπτα** για τη φάση **intergreen** ή αλλιώς **interval time** που αφορά το χρόνο μετάβασης από το κόκκινο στο φανάρι του πεζού στο πράσινο των οχημάτων και **67 δευτερόλεπτα** για το **κόκκινο φως**. Με βάση αυτά τα δεδομένα, δημιουργήθηκαν διαδοχικοί κύκλοι από τον δεύτερο έως τον 42ο, προσδιορίζοντας το χρονικό τους εύρος και τη συνολική διάρκεια κάθε κύκλου. Αφού ορίστηκαν οι κύκλοι, κάθε χρονική στιγμή του συνόλου δεδομένων αντιστοιχήθηκε στον σωστό κύκλο σηματοδότησης, ελέγχοντας αν ανήκει στο χρονικό διάστημα ενός συγκεκριμένου κύκλου και καταγράφοντας τον αριθμό του στη στήλη "Cycle Number" όταν εντοπιζόταν αντιστοιχία.

4.4.6 Αντιστοίχιση φάσεων και καθαρισμός δεδομένων

Στη συνέχεια, για κάθε χρονική στιγμή προσδιορίστηκε η **αντίστοιχη φάση σηματοδότησης**, με την ένδειξη **"Red"** να καταγράφεται αν ο χρόνος βρισκόταν μεταξύ της έναρξης του κύκλου και του τέλους της κόκκινης φάσης, την ένδειξη **"Green"** αν αντιστοιχούσε στη διάρκεια της πράσινης φάσης, και την ένδειξη **"Intergreen"** αν ανήκε στη μεταβατική περίοδο. Για την εξασφάλιση της συνοχής στα δεδομένα, εφαρμόστηκαν τεχνικές συμπλήρωσης τιμών (**forward** και **backward filling**), ώστε να καλυφθούν τυχόν κενά στην αντιστοίχιση κύκλων και φάσεων. Επιπλέον, η στήλη "Cycle Number" μετατράπηκε σε ακέραια τιμή για μεγαλύτερη ακρίβεια. Τέλος, οι ενδείξεις "Intergreen" αντικαταστάθηκαν με την τιμή "Red", καθώς θεωρήθηκε πως η φάση αυτή αποτελεί μέρος της κόκκινης ένδειξης.

4.4.7 Αξιολόγηση και οπτικοποίηση της ακρίβειας ανίχνευσης φάσεων φωτεινών σηματοδοτών

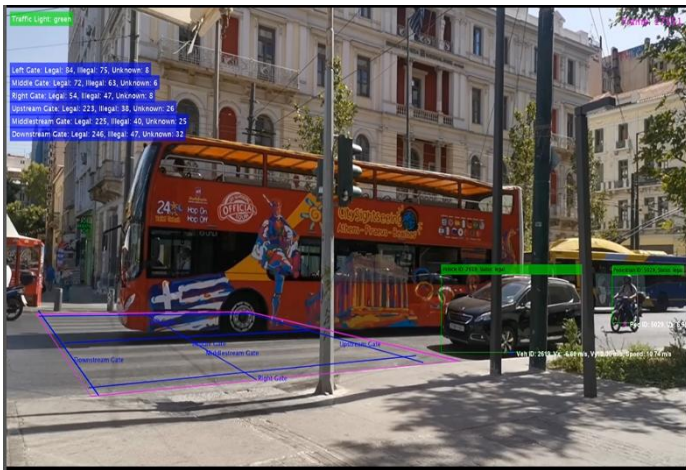
Για την εκτίμηση της αξιοπιστίας της ανίχνευσης των φάσεων του φωτεινού σηματοδότη, πραγματοποιήθηκε σύγκριση μεταξύ των δεδομένων της χειροκίνητης μεθόδου (**"Phase"**) και των διορθωμένων τιμών της αυτοματοποιημένης (**"Resolved Traffic Light Status"**). Αρχικά, διασφαλίστηκε ότι οι τιμές των φάσεων μετατράπηκαν σε πεζούς χαρακτήρες ώστε να αποφευχθούν προβλήματα σύγκρισης λόγω διαφορετικής μορφοποίησης. Στη συνέχεια, υπολογίστηκε η **ακρίβεια ανίχνευσης σε επίπεδο κύκλου σηματοδότησης**, μετρώντας το ποσοστό των περιπτώσεων όπου οι δύο τιμές συμφωνούσαν. Για τον συνολικό υπολογισμό της ακρίβειας, συγκρίθηκαν όλες οι εγγραφές και προσδιορίστηκε το συνολικό ποσοστό επιτυχίας της ανίχνευσης φάσεων, το οποίο εμφανίστηκε ως συνολική ακρίβεια του συστήματος. Η ανάλυση οπτικοποιήθηκε μέσω **γραφήματος στηλών**, όπου κάθε κύκλος σηματοδότησης απεικονίστηκε με το αντίστοιχο ποσοστό ακρίβειας. Για την καλύτερη κατανόηση των αποτελεσμάτων, χρησιμοποιήθηκε χρωματική διαφοροποίηση, με **pale blue** για ακρίβεια άνω του 50% και **pale red** για χαμηλότερη ακρίβεια.



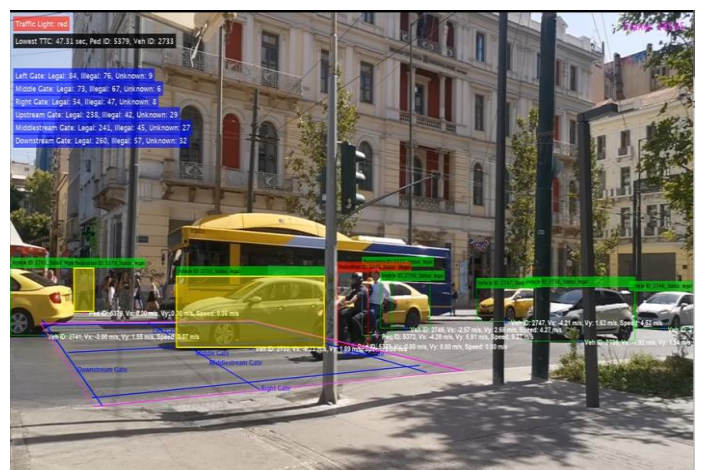
Εικόνα 5: Ακρίβεια ανίχνευσης φάσεων του φωτεινού σηματοδότη ανά κύκλο σηματοδότησης

Η ακρίβεια της ανίχνευσης παρουσιάζει **ιδιαίτερα ικανοποιητικά αποτελέσματα**, με την πλειονότητα των κύκλων να καταγράφει τιμές μεταξύ 50% και 80%, γεγονός που υποδηλώνει ότι το σύστημα αποδίδει αξιόπιστα στη μεγαλύτερη διάρκεια της λειτουργίας του. Σε πολλές περιπτώσεις, η απόδοσή του είναι εξαιρετικά υψηλή, με χαρακτηριστικά παραδείγματα κύκλους όπου η ακρίβεια φτάνει το **90.5%, το 81.5% και το 78.1%**, αποδεικνύοντας ότι ο αλγόριθμος μπορεί να εντοπίσει με μεγάλη ακρίβεια τις φάσεις του φωτεινού σηματοδότη. Το συνολικό ποσοστό ακρίβειας, όπως επισημαίνεται στο γράφημα, ανέρχεται σε **62.77%**, επιβεβαιώνοντας ότι, κατά μέσο όρο, η ανίχνευση της φάσης είναι σωστή στο **63%** των περιπτώσεων, ένα αποτέλεσμα που καταδεικνύει τη σταθερή και αποτελεσματική λειτουργία του συστήματος.

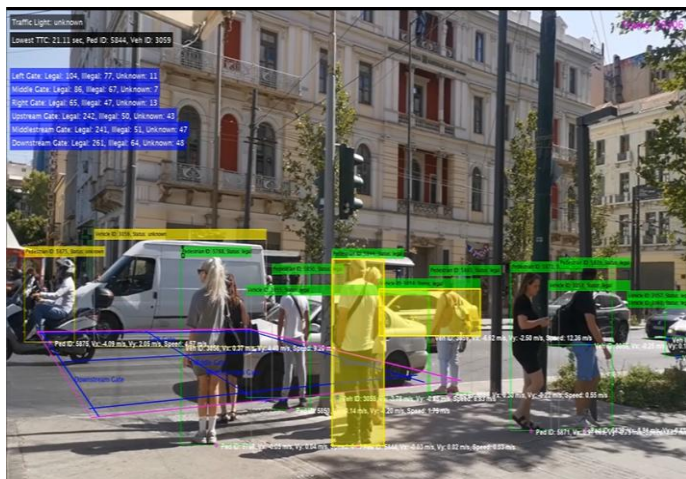
Ωστόσο, υπάρχουν και κύκλοι με σημαντικά **χαμηλότερη ακρίβεια**, όπως εκείνοι με **13.7% και 37.9%**, όπου παρατηρούνται αποκλίσεις μεταξύ της ανιχνευθείσας και της πραγματικής φάσης. Αυτά τα χαμηλά ποσοστά μπορούν να αποδοθούν στην **απόκρυψη του φωτεινού σηματοδότη** από μεγάλα οχήματα, όπως λεωφορεία και φορτηγά, όπως φαίνεται στις εικόνες των κύκλων 8 (Εικόνα 6,7,8), 10 (Εικόνα 9) και 16 (Εικόνα 10). Σε αυτές τις περιπτώσεις ο αλγόριθμος στηριζόμενος στο σηματοδότη και στη χρωματική διαφοροποίηση, δεν μπορούσε να αναγνωρίσει τη σωστή φάση και οδήγησε είτε σε μη ταξινομημένες περιπτώσεις (unknown status) είτε σε λανθασμένες ανιχνεύσεις.



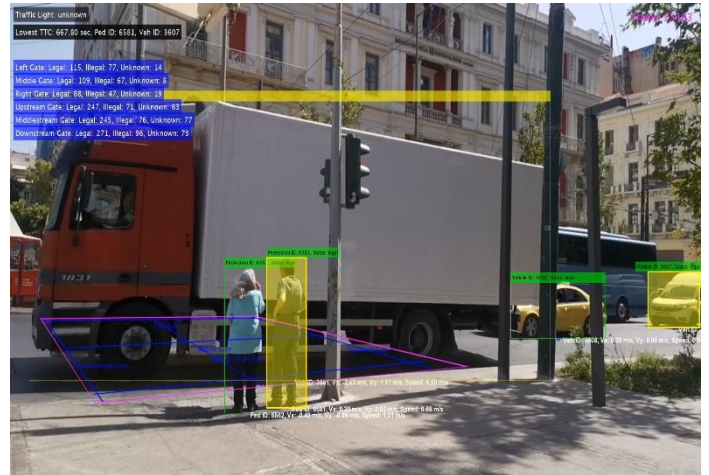
Εικόνα 6: Κύκλος 8 – Απόκρυψη φωτεινού σηματοδότη



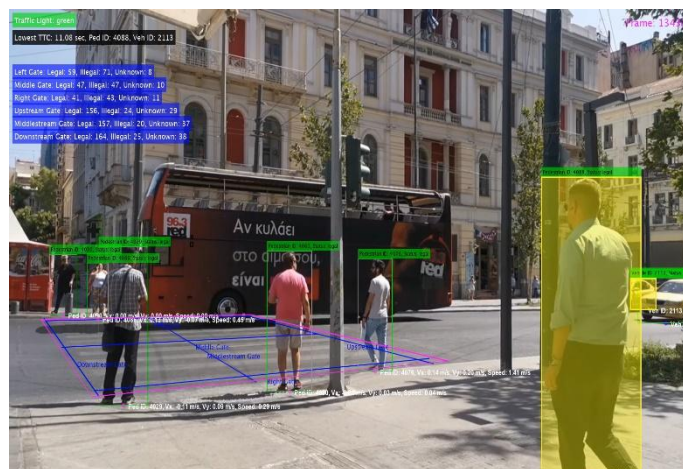
Εικόνα 7: Κύκλος 8 – Απόκρυψη φωτεινού σηματοδότη



Εικόνα 8: : Κύκλος 8 – Απόκρυψη φωτεινού σηματοδότη



Εικόνα 9: Κύκλος 10 – Απόκρυψη φωτεινού σηματοδότη



Εικόνα 10: Κύκλος 16 – Απόκρυψη φωτεινού σηματοδότη

4.5 Ανάλυση της ακρίβειας των παράνομων διαβάσεων πεζών

4.5.1 Σύντομη ανασκόπηση και σύνδεση με την τρέχουσα σύγκριση

Στην προηγούμενη ανάλυση, πραγματοποιήθηκε μια ολοκληρωμένη διαδικασία **προ-επεξεργασίας** των δεδομένων, με στόχο τη διατήρηση της **χρονικής συνέχειας** και την ενίσχυση της ακρίβειας στην ανάλυση. Ειδικότερα, ενοποιήθηκε η αρίθμηση των καρτέ μερών της μεταβλητής Continuous Frame, διευκολύνοντας την αντιστοίχιση των δεδομένων στον πραγματικό χρόνο. Παράλληλα, εφαρμόστηκαν **τεχνικές καθαρισμού και συμπλήρωσης τιμών** για την εξομάλυνση της μεταβλητής Traffic Light Status, διατηρώντας τη συνέπεια στη χρονοσειρά. Επιπλέον, προσδιορίστηκαν οι **κύκλοι λειτουργίας** του φωτεινού σηματοδότη και αντιστοιχήθηκαν με κάθε χρονική εγγραφή, επιτρέποντας μια ακριβή κατηγοριοποίηση των δεδομένων σε σχέση με τις **φάσεις σηματοδότησης**. Τα αποτελέσματα και οι μέθοδοι που αναπτύχθηκαν σε αυτήν τη φάση θα αξιοποιηθούν και στην τρέχουσα σύγκριση, διασφαλίζοντας τη συνοχή στη μεθοδολογία και την εγκυρότητα των ευρημάτων.

4.5.2 Φιλτράρισμα και καταμέτρηση παράνομων διαβάσεων πεζών αυτοματοποιημένης μεθόδου

Αρχικά, φιλτραρίστηκαν οι περιπτώσεις όπου οι πεζοί βρέθηκαν στην περιοχή διάβασης ενώ το φανάρι δεν ήταν πράσινο (**Resolved Traffic Light Status $\neq$ "green"**) και για να διασφαλιστεί ότι εξετάζονται μόνο οι πραγματικές παραβάσεις, αφαιρέθηκαν περιπτώσεις όπου η μεταβλητή **Phase** είχε αποκλειστικά την τιμή **"Green"**. Τα δεδομένα ομαδοποιήθηκαν ανά πεζό και κύκλο σηματοδότησης, και υπολογίστηκαν βασικά χαρακτηριστικά για κάθε παράνομη διάβαση, όπως, το συνολικό χρονικό διάστημα παραβίασης (**TotalSeconds**), το αρχικό και τελικό καρέ (**StartFrame**, **EndFrame**), η διαφορά των καρτέ μερών (**Frame Difference**), καθώς και οι επικρατέστερες τιμές στις μεταβλητές Resolved Traffic Light Status και Phase. Πιο συγκεκριμένα, διατηρήθηκαν μόνο περιπτώσεις με **TotalSeconds ≥ 3.5** , απορρίφθηκαν εγγραφές με ταυτόσημα **StartFrame** και **EndFrame**, ενώ διατηρήθηκαν μόνο παραβάσεις με διαφορά καρτέ **τουλάχιστον ίση με 100**. Στη συνέχεια, αφαιρέθηκαν διπλότυπα πεζών, κρατώντας μόνο την πρώτη καταγραφή για να εξασφαλιστεί η αποφυγή διπλομέτρησης. Τέλος, οι παράνομες διαβάσεις συνοψίστηκαν ανά κύκλο σηματοδότησης, μετρώντας τον συνολικό αριθμό παραβάσεων (**Illegal Crossings Total Auto**) και καταγράφοντας τις επικρατούσες τιμές στις μεταβλητές Traffic Light Status και Phase.

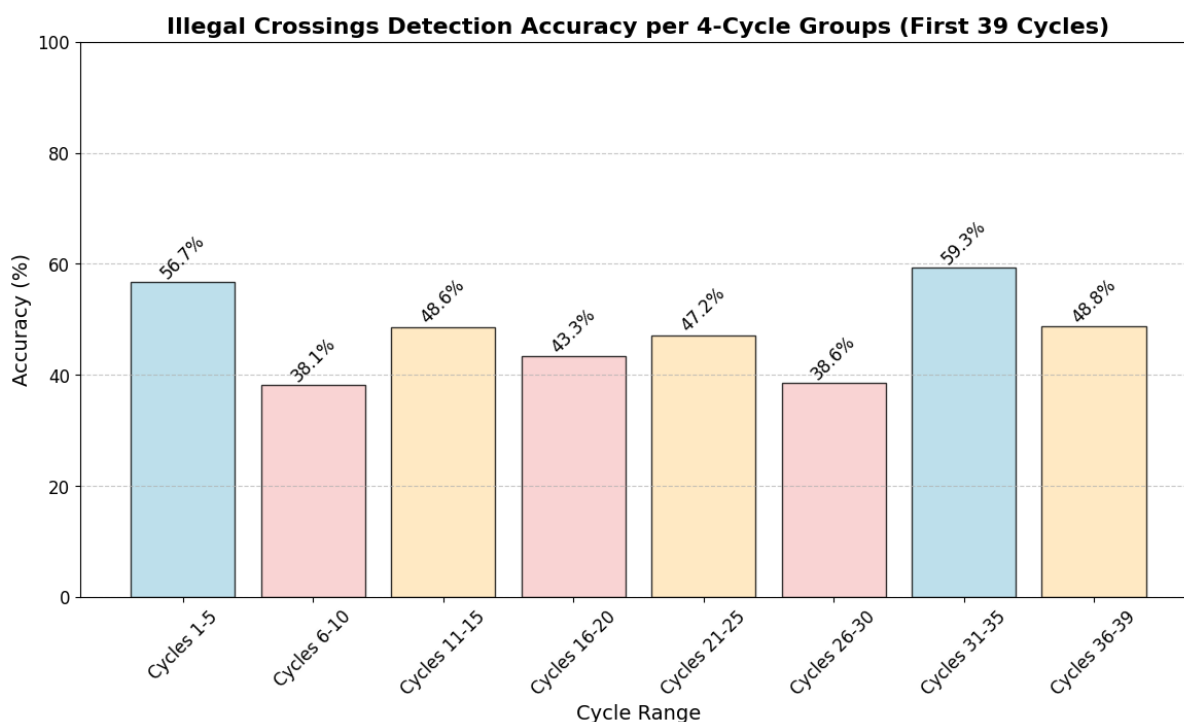
4.5.3 Φιλτράρισμα και καταμέτρηση παράνομων διαβάσεων πεζών χειροκίνητης μεθόδου

Αφού φορτώθηκε το dataset από το αρχείο της χειροκίνητης βάσης δεδομένων, η στήλη **Time (second)** μετατράπηκε σε αριθμητική μορφή, διασφαλίζοντας ότι τυχόν κόμματα (,) αντικαταστάθηκαν με τελείες (.) για σωστή αριθμητική επεξεργασία. Στη συνέχεια, καθορίστηκαν οι **χρονικές διάρκειες των κύκλων σηματοδότησης**: ο πρώτος κύκλος αποτελούνταν από 18 δευτερόλεπτα για το κόκκινο, 10 δευτερόλεπτα για το πράσινο και 13 δευτερόλεπτα για τη μεταβατική φάση (intergreen), ενώ οι υπόλοιποι κύκλοι είχαν 67 δευτερόλεπτα για το κόκκινο, 10 για το πράσινο και 13 για

τη μεταβατική φάση. Έπειτα, πραγματοποιήθηκε επαναληπτική ανάθεση του σωστού κύκλου και της κατάστασης του φωτεινού σηματοδότη σε κάθε χρονική στιγμή στο dataset και η επεξεργασία διακόπηκε μετά τον 39ο κύκλο, καθώς πέρα από αυτό το σημείο δεν υπήρχαν έγκυρα δεδομένα. Τα δεδομένα ομαδοποιήθηκαν ανά **Cycle Number** και **Traffic Light Status**, και υπολογίστηκε το άθροισμα των παράνομων διαβάσεων (**Illegal Crossings**), τόσο για την κόκκινη φάση όσο και για τη μεταβατική φάση (**intergreen**). Στη συνέχεια, τα δεδομένα συγχωνεύτηκαν, ώστε κάθε κύκλος και κάθε κατάσταση σηματοδότησης να περιείχε τις συνολικές παράνομες διαβάσεις, εξασφαλίζοντας έτσι την πλήρη και σωστή καταγραφή των παραβάσεων για κάθε φάση του φωτεινού σηματοδότη.

4.5.4 Αξιολόγηση και οπτικοποίηση της ακρίβειας ανίχνευσης παράνομων διαβάσεων πεζών

Σε αυτό το στάδιο της ανάλυσης, πραγματοποιήθηκε η **σύγκριση** μεταξύ της αυτόματης και της χειροκίνητης καταγραφής παράνομων διαβάσεων πεζών. Αρχικά, τα δύο σύνολα δεδομένων, χειροκίνητες καταμετρήσεις και αυτόματες καταγραφές μέσω αλγορίθμου, συγχωνεύτηκαν βάσει του αριθμού κύκλου σηματοδότησης (**Cycle Number**), διατηρώντας όλους τους κύκλους, ακόμη και εκείνους στους οποίους δεν είχαν καταγραφεί παραβάσεις, εξασφαλίζοντας, έτσι, συνολική πληρότητα δεδομένων αλλά και μια ολοκληρωμένη αξιολόγηση της ακρίβειας του αυτόματου συστήματος. Η ακρίβεια της αυτόματης ανίχνευσης υπολογίστηκε συγκρίνοντας τις δύο μεθόδους για κάθε κύκλο σηματοδότησης με τον δείκτη ακρίβειας να ορίζεται ως το ποσοστό της ελάχιστης τιμής των δύο μεθόδων σε σχέση με τη μέγιστη, πολλαπλασιασμένο με 100. Μετά τον υπολογισμό της ακρίβειας για κάθε κύκλο, υπολογίστηκε, επίσης, και η **συνολική ακρίβεια** ως ο μέσος όρος των επιμέρους ακριβειών και προέκυψε ίση με **47.54%**, ποσοστό το οποίο χαρακτηρίζεται οριακά αποδεκτό. Τα αποτελέσματα της παραπάνω σύγκρισης οπτικοποιούνται μέσω ραβδογράμματος (Εικόνα 11), το οποίο παρατίθεται παρακάτω, και στο οποίο οι κύκλοι ομαδοποιήθηκαν σε **ομάδες των πέντε** για την καλύτερη ανάλυση της απόδοσης του αλγορίθμου σε μεγαλύτερα διαστήματα. Για κάθε ομάδα, υπολογίστηκε ο **μέσος όρος της ακρίβειας των κύκλων** που την απαρτίζουν, παρέχοντας μια πιο ολοκληρωμένη εικόνα της διακύμανσης της ακρίβειας κατά τη διάρκεια του πειράματος. Για να γίνει η διαφοροποίηση πιο εμφανής, εφαρμόστηκε ένα σύστημα χρωματικής κωδικοποίησης: **pale blue** για ακρίβεια $\geq 50\%$, υποδεικνύοντας καλή συμφωνία μεταξύ των δύο μεθόδων, **pale orange** για ακρίβεια μεταξύ 45% και 50%, δείχνοντας οριακή ακρίβεια και **pale red** για ακρίβεια $< 45\%$, που σηματοδοτεί σημαντικές διαφορές μεταξύ της αυτόματης και της χειροκίνητης μέτρησης.



Εικόνα 11: Ακρίβεια ανίχνευσης παράνομων διελεύσεων ανά ομάδα 5 κύκλων σηματοδότησης

Παρατηρούμε ότι σε ορισμένες περιόδους η ανίχνευση ήταν **αποτελεσματική**, με το υψηλότερο ποσοστό να καταγράφεται στην ομάδα κύκλων **31-35 (59.3%)** και 1-5 (56.7%), δείχνοντας ισχυρή απόδοση του συστήματος. Επιπλέον, η ακρίβεια διατήρησε ένα ικανοποιητικό επίπεδο στις περιόδους 11-15 (48.6%), 21-25 (47.2%) και 36-39 (48.8%), υποδηλώνοντας σταθερή ανίχνευση σε πολλές φάσεις. Ωστόσο, υπήρχαν και περίοδοι όπου η απόδοση ήταν **χαμηλότερη**, όπως οι **6-10 (38.1%)**, 16-20 (43.3%) και 26-30 (38.6%), γεγονός που δείχνει ότι το σύστημα αντιμετώπισε προκλήσεις στην ακρίβεια ανίχνευσης κατά αυτές τις φάσεις. Η μειωμένη αυτή απόδοση μπορεί να αποδοθεί σε συγκεκριμένα αίτια που αντιμετώπισε το σύστημα κατά την ανίχνευση παράνομων διελεύσεων. Ένας βασικός παράγοντας ήταν η **αδυναμία του αλγορίθμου να διακρίνει με σαφήνεια την ένδειξη του φωτεινού σηματοδότη** σε συγκεκριμένους κύκλους, όπως 8, 10 και 16, όπου μεγάλα οχήματα (λεωφορεία, φορτηγά) παρεμπόδιζαν την ορατότητα. Ως αποτέλεσμα, το σύστημα αδυνατούσε να προσδιορίσει με ακρίβεια ποιοι πεζοί διέσχισαν νόμιμα και ποιοι παραβίαζαν τον σηματοδότη, επηρεάζοντας την ακρίβεια της ταξινόμησης. Ένας ακόμα λόγος που συνέβαλε στη μείωση της ανίχνευσης ήταν **ο τρόπος με τον οποίο το λογισμικό κατέγραφε τις διελεύσεις**. Συγκεκριμένα, σε περιπτώσεις όπου ένας πεζός ξεκινούσε τη διάσχιση της διάβασης κατά τη διάρκεια της πράσινης ένδειξης, αλλά ολοκλήρωνε τη διάβαση κατά τη φάση εκκαθάρισης, το σύστημα ενδέχεται να τον κατέγραφε εσφαλμένα ως παράνομο, μια τακτική που οφείλεται στη δυσκολία αναγνώρισης της δυναμικής της κίνησης των πεζών σε σχέση με τη χρονική διάρκεια των φωτεινών σηματοδοτών (Εικόνα 12).



Εικόνα 12: Διέλευση πεζών κατά την νόμιμη φάση εκκαθάρισης και λανθασμένη ταξινόμηση αυτών

4.6 Ανάλυση της ακρίβειας των νόμιμων διαβάσεων πεζών

4.6.1 Σύντομη ανασκόπηση και σύνδεση με την τρέχουσα σύγκριση

Όπως και στην προηγούμενη ανάλυση ακρίβειας, έτσι και στην παρούσα, θα αξιοποιήθηκαν οι ίδιες μεθοδολογικές αρχές και τεχνικές **προεπεξεργασίας δεδομένων**. Συγκεκριμένα, η χρήση της μεταβλητής **Continuous Frame** για την ενοποίηση της αρίθμησης των καρέ συνέχισε να διαδραματίζει κεντρικό ρόλο, εξασφαλίζοντας την αντιστοίχιση των δεδομένων με ακρίβεια στον πραγματικό χρόνο. Παράλληλα, οι τεχνικές καθαρισμού και συμπλήρωσης τιμών που εφαρμόστηκαν στη μεταβλητή **Traffic Light Status** χρησιμοποιήθηκαν εκ νέου, διατηρώντας τη συνέπεια στη χρονοσειρά και αποτρέποντας πιθανές στρεβλώσεις στα δεδομένα. Επιπλέον, η κατηγοριοποίηση των δεδομένων με βάση τους **κύκλους** λειτουργίας του φωτεινού σηματοδότη διατηρήθηκε, διασφαλίζοντας την ορθότητα των συσχετίσεων μεταξύ των φάσεων σηματοδότησης και των καταγεγραμμένων παρατηρήσεων.

4.6.2 Φιλτράρισμα και καταμέτρηση νόμιμων διαβάσεων πεζών αυτοματοποιημένης μεθόδου

Στην παρούσα διαδικασία προεπεξεργασίας, τα δεδομένα φιλτραρίστηκαν, ομαδοποιήθηκαν και καθαρίστηκαν, προκειμένου να αναλυθούν μόνο οι **νόμιμες διαβάσεις πεζών**, δηλαδή αυτές που πραγματοποιούνται υπό συνθήκες πράσινου φωτεινού σηματοδότη. Αρχικά, επιλέχθηκαν οι καταγραφές όπου ο πεζός βρισκόταν εντός της περιοχής διάβασης και η κατάσταση του φωτεινού σηματοδότη αντιστοιχούσε στην ένδειξη **"green"**, απορρίπτοντας, έτσι, τα δεδομένα που σχετίζονταν με κόκκινο φανάρι ή μεσοδιάστημα σηματοδότησης. Στη συνέχεια, τα

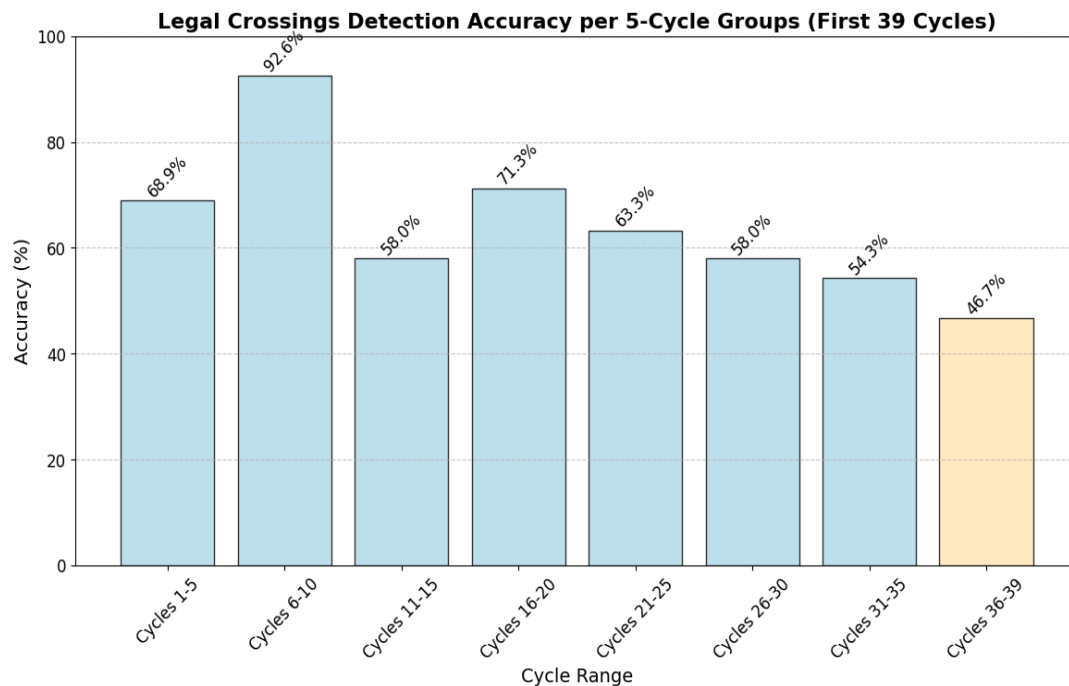
δεδομένα ομαδοποιήθηκαν ανά πεζό και κύκλο σηματοδότησης, επιτρέποντας την εξαγωγή βασικών στατιστικών μεταβλητών. Υπολογίστηκε η **συνολική διάρκεια της διάβασης** μέσω του αριθμού μοναδικών χρονικών στιγμών, ενώ προσδιορίστηκαν το **αρχικό και το τελικό καρέ**, καθώς και η διαφορά τους, η οποία εκφράζει τη συνολική χρονική διάρκεια της διάβασης σε επίπεδο καρέ. Επιπλέον, αναλύθηκε η **επικρατούσα κατάσταση του φωτεινού σηματοδότη** κατά τη διάρκεια της διάβασης, καθώς και η κυρίαρχη φάση του σηματοδότη. Έπειτα, τα δεδομένα ομαδοποιήθηκαν σε επίπεδο κύκλου σηματοδότησης και, έτσι, εξήχθησαν τα συνολικά στατιστικά για κάθε κύκλο. Τέλος, υπολογίστηκε ο συνολικός αριθμός νόμιμων διαβάσεων, καθώς και οι κυρίαρχες καταστάσεις του φωτεινού σηματοδότη και της φάσης σηματοδότησης και προκειμένου να διατηρηθεί η πληρότητα των δεδομένων, δημιουργήθηκε ένας πίνακας που περιλάμβανε όλους τους κύκλους σηματοδότησης, ακόμα και εκείνους χωρίς καταγεγραμμένες διαβάσεις, ούτως ώστε να διασφαλιστεί ότι δεν παραλείφθηκε καμία χρονική περίοδος.

4.6.3 Φιλτράρισμα και καταμέτρηση νόμιμων διαβάσεων πεζών χειροκίνητης μεθόδου

Αφού φορτώθηκε το dataset της χειροκίνητης βάσης δεδομένων, η στήλη **Time (second)** μετατράπηκε σε αριθμητική μορφή, διασφαλίζοντας ότι τυχόν κόμματα (,) αντικαταστάθηκαν με τελείες (.) για σωστή αριθμητική επεξεργασία. Στη συνέχεια, καθορίστηκαν οι **χρονικές διάρκειες** των κύκλων σηματοδότησης: ο πρώτος κύκλος αποτελούνταν από 18 δευτερόλεπτα για το κόκκινο, 10 δευτερόλεπτα για το πράσινο και 13 δευτερόλεπτα για τη μεταβατική φάση (intergreen), ενώ οι υπόλοιποι κύκλοι είχαν 67 δευτερόλεπτα κόκκινο, 10 δευτερόλεπτα πράσινο και 13 δευτερόλεπτα μεσοδιάστημα. Για κάθε χρονική στιγμή στο dataset, πραγματοποιήθηκε επαναληπτική διαδικασία αντιστοίχισης του σωστού κύκλου σηματοδότησης και της αντίστοιχης κατάστασης του φωτεινού σηματοδότη. Αν η τιμή του χρόνου βρισκόταν εντός του χρονικού διαστήματος του κόκκινου φωτός, η κατάσταση καταγράφηκε ως **"Red"**, αν ανήκε στο πράσινο φως ως **"Green"**, ενώ αν βρισκόταν στη μεταβατική φάση, καταγράφηκε ως **"Intergreen"**. Η διαδικασία επαναλήφθηκε για όλες τις χρονικές στιγμές έως τον 39ο κύκλο, όπου και διακόπηκε η ανάλυση, καθώς πέρα από αυτό το σημείο δεν υπήρχαν έγκυρα δεδομένα. Τα δεδομένα εμπλουτίστηκαν με νέες στήλες που αποθήκευαν τον αριθμό κύκλου και την κατάσταση του σηματοδότη για κάθε χρονική στιγμή. Έπειτα, πραγματοποιήθηκε ομαδοποίηση των δεδομένων βάσει του Cycle Number και του Traffic Light Status, ώστε να υπολογιστεί το άθροισμα των νόμιμων διαβάσεων (Legal Crossings) για κάθε κύκλο και κάθε φάση σηματοδότησης. Οι συνολικές διαβάσεις συγχωνεύτηκαν στο dataset, με την τιμή των διαβάσεων να καταγράφεται μόνο στην πρώτη εγγραφή κάθε φάσης, αποτρέποντας την εμφάνιση διπλοεγγραφών. Στη συνέχεια, εφαρμόστηκε φιλτράρισμα, ώστε να διατηρηθούν μόνο οι χρονικές στιγμές που αντιστοιχούσαν στην πράσινη φάση του σηματοδότη, καθώς η ανάλυση επικεντρώθηκε αποκλειστικά στις νόμιμες διαβάσεις. Τα δεδομένα ομαδοποιήθηκαν εκ νέου ανά **Cycle Number**, ώστε να υπολογιστεί το συνολικό πλήθος των νόμιμων διαβάσεων για κάθε κύκλο. Τέλος, η στήλη που περιείχε το σύνολο των διαβάσεων μετονομάστηκε σε **"Legal Crossings Total Manual"**, εξασφαλίζοντας την πλήρη και σωστή καταγραφή της ροής των πεζών κατά τη διάρκεια του πράσινου σηματοδότη.

4.6.4 Αξιολόγηση και οπτικοποίηση της ακρίβειας ανίχνευσης νόμιμων διαβάσεων πεζών

Σε αυτό το στάδιο της ανάλυσης, πραγματοποιήθηκε η **σύγκριση** μεταξύ της αυτόματης και της χειροκίνητης καταγραφής νόμιμων διαβάσεων πεζών. Αρχικά, τα δύο σύνολα δεδομένων, χειροκίνητες καταμετρήσεις και αυτόματες καταγραφές μέσω αλγορίθμου, συγχωνεύτηκαν βάσει του αριθμού **κύκλου σηματοδότησης** (Cycle Number), διατηρώντας όλους τους κύκλους, ακόμη και εκείνους στους οποίους δεν είχαν καταγραφεί παραβάσεις, εξασφαλίζοντας, έτσι, συνολική πληρότητα δεδομένων αλλά και μια ολοκληρωμένη αξιολόγηση της ακρίβειας του αυτόματου συστήματος. Η ακρίβεια της αυτόματης ανίχνευσης υπολογίστηκε συγκρίνοντας τις δύο μεθόδους για κάθε κύκλο σηματοδότησης με τον δείκτη ακρίβειας να ορίζεται ως το ποσοστό της ελάχιστης τιμής των δύο μεθόδων σε σχέση με τη μέγιστη, πολλαπλασιασμένο με 100. Μετά τον υπολογισμό της ακρίβειας για κάθε κύκλο, υπολογίστηκε, επίσης, και η **συνολική ακρίβεια** ως ο μέσος όρος των επιμέρους ακριβειών και προέκυψε ίση με **64.60%**, ποσοστό το οποίο χαρακτηρίζεται αρκετά ικανοποιητικό. Τα αποτελέσματα της παραπάνω σύγκρισης οπτικοποιούνται μέσω **ραβδογράμματος** (Εικόνα 13), το οποίο παρατίθεται παρακάτω, και στο οποίο οι κύκλοι ομαδοποιήθηκαν σε **ομάδες των πέντε** για την καλύτερη ανάλυση της απόδοσης του αλγορίθμου σε μεγαλύτερα διαστήματα. Για κάθε ομάδα, υπολογίστηκε ο μέσος όρος της ακρίβειας των κύκλων που την απαρτίζουν, παρέχοντας μια πιο ολοκληρωμένη εικόνα της διακύμανσης της ακρίβειας κατά τη διάρκεια του πειράματος. Για να γίνει η διαφοροποίηση πιο εμφανής, εφαρμόστηκε ένα σύστημα χρωματικής κωδικοποίησης: **pale blue** για ακρίβεια $\geq 50\%$, υποδεικνύοντας καλή συμφωνία μεταξύ των δύο μεθόδων, **pale orange** για ακρίβεια μεταξύ 45% και 50%, δείχνοντας οριακή ακρίβεια και **pale red** για ακρίβεια $< 45\%$, που σηματοδοτεί σημαντικές διαφορές μεταξύ της αυτόματης και της χειροκίνητης μέτρησης.



Εικόνα 12: Ακρίβεια ανίχνευσης νόμιμων διελεύσεων ανά ομάδα 5 κύκλων σηματοδότησης

Παρατηρούμε ότι η ακρίβεια της ανίχνευσης των νόμιμων διαβάσεων διατηρείται σε **υψηλά επίπεδα**, καθώς οι περισσότερες ομάδες κύκλων καταγράφουν ποσοστά άνω του **50%**, επιβεβαιώνοντας τη συνολική αξιοπιστία της μεθόδου. Ιδιαίτερα αξιοσημείωτη είναι η απόδοση της ομάδας κύκλων **6-10**, η οποία επιτυγχάνει ακρίβεια **92,6%**, αποδεικνύοντας εξαιρετική συμφωνία μεταξύ των δύο μεθόδων καταγραφής. Αντίστοιχα, οι ομάδες 1-5 (68,9%), 16-20 (71,3%), 21-25 (63,3%), 11-15 (58,0%) και 26-30 (58,0%) διατηρούν την ακρίβεια σε υψηλά επίπεδα, εξασφαλίζοντας σταθερή και αξιόπιστη απόδοση.

Ωστόσο, παρατηρείται μείωση της ακρίβειας μόνο στην τελευταία **ομάδα κύκλων 36-39**, όπου το ποσοστό ανέρχεται στο **46,7%**, καταγράφοντας τη χαμηλότερη επίδοση της ανάλυσης. Η μείωση αυτή αποδίδεται σε εξωτερικούς παράγοντες που επηρέασαν την αξιοπιστία της καταγραφής, και συγκεκριμένα στην **παρεμπόδιση της ορατότητας του φωτεινού σηματοδότη από μεγάλα οχήματα** (όπως φορτηγά) που βρίσκονταν σταθμευμένα ή διερχόμενα μπροστά από την κάμερα. Το γεγονός αυτό είχε ως αποτέλεσμα, για μεγάλο αριθμό διαδοχικών καρέ (frames per second), να μην είναι εφικτή η σαφής αναγνώριση της κατάστασης του σηματοδότη. Δεδομένου ότι ο αλγόριθμος βασίζεται αποκλειστικά στη **χρωματική πληροφορία** και όχι στη **χρονική ακολουθία**, η αδυναμία ανίχνευσης σταθερού χρώματος (κόκκινο ή πράσινο) οδηγεί σε χαρακτηρισμούς "unknown". Συνεπώς, η μερική ή πλήρης απόκρυψη του σηματοδότη επιδρά έμμεσα στην ακρίβεια καταγραφής των νόμιμων και παράνομων διελεύσεων, μειώνοντας την τελική επίδοση της ομάδας αυτής.

Κεφάλαιο 5: Εφαρμογή μεθοδολογίας και αποτελέσματα

5.1 Εισαγωγή

Το παρόν κεφάλαιο παρουσιάζει τα αποτελέσματα της εφαρμογής τεσσάρων μοντέλων πρόβλεψης που αναπτύχθηκαν για την ανάλυση της συμπεριφοράς πεζών σε σηματοδοτούμενες διαβάσεις. Συγκεκριμένα, κατασκευάστηκαν δύο μοντέλα **Λογιστικής Παλινδρόμησης** και δύο μοντέλα τύπου **Τυχαίων Δασών**, με στόχο την πρόβλεψη δύο εξαρτημένων μεταβλητών: (α) της παρουσίας πεζού εντός ή εκτός διάβασης και (β) της νομιμότητας ή παραβατικότητας της διέλευσής του. Πριν από την εφαρμογή των μοντέλων, πραγματοποιήθηκε προκαταρκτική ανάλυση συσχετίσεων με τη χρήση **σημειακής διχοτομικής συσχέτισης**.

Τα μοντέλα εκπαιδεύτηκαν με βάση ένα τελικό ενιαίο αρχείο δεδομένων, το οποίο προέκυψε από τη συγχώνευση και χρονική ευθυγράμμιση πληροφορίας που αφορούσε πεζούς, οχήματα και εκτιμήσεις του ελάχιστου χρόνου έως πιθανή σύγκρουση (MinTTC). Τα αρχικά δεδομένα αντλήθηκαν από βιντεοσκοπημένες σκηνές αστικής κυκλοφορίας και επεξεργάστηκαν μέσω αλγορίθμου υπολογιστικής όρασης στο περιβάλλον MATLAB, ενώ η ενοποίηση και ανάλυση πραγματοποιήθηκαν στην Python.

Η αξιολόγηση των μοντέλων βασίστηκε σε πλήθος στατιστικών δεικτών. Για τα μοντέλα λογιστικής παλινδρόμησης εξετάστηκαν ο πίνακας σύγχυσης, η ταξινομική αναφορά (classification report), οι καμπύλες ROC με τις αντίστοιχες τιμές AUC, ο δείκτης Pseudo R^2 , τα κριτήρια AIC και BIC, καθώς και η στατιστική σημαντικότητα και οι συντελεστές των ανεξάρτητων μεταβλητών. Ιδιαίτερη έμφαση δόθηκε στην ερμηνευτική ικανότητα της λογιστικής παλινδρόμησης, καθώς παρέχει πληροφορίες για το πώς και σε ποιο βαθμό κάθε μεταβλητή επηρεάζει την πιθανότητα εμφάνισης του φαινομένου. Αντίστοιχα, για τα μοντέλα Random Forest αξιολογήθηκαν η απόδοση ταξινόμησης, ο πίνακας σύγχυσης, οι τιμές AUC και η σχετική σημασία (feature importance) των χαρακτηριστικών.

5.2 Επεξεργασία και ενοποίηση δεδομένων πεζών

5.2.1 Χρονική ενοποίηση και κανονικοποίηση δεδομένων πεζών

Αρχικά, πραγματοποιήθηκε ταξινόμηση των δεδομένων πεζών με βάση το πλάνο του βίντεο (**VideoPart**) και τον αριθμό του καρέ (**Frame**), ώστε να διασφαλιστεί η σωστή **χρονική σειρά** εμφάνισης των γεγονότων. Στόχος ήταν να συνενωθούν τα επιμέρους βίντεο σε μια ενιαία και συνεχόμενη χρονική ροή, ανεξαρτήτως των τεχνικών διαχωρισμών σε πλάνα.

Για να το πετύχουμε αυτό, δημιουργήθηκε μια νέα στήλη με το όνομα **ContinuousFrame**, η οποία αναπροσαρμόζει τα καρέ κάθε πλάνου έτσι ώστε να συνεχίζουν χρονικά από εκεί που σταμάτησε το προηγούμενο. Αυτό εξασφαλίζει ότι δεν υπάρχει χρονικό κενό ή επικάλυψη μεταξύ των πλάνων.

Στη συνέχεια, τα συνεχόμενα καρέ μετατράπηκαν σε **δευτερόλεπτα** και **λεπτά**, χρησιμοποιώντας το frame rate των 29.74 καρέ ανά δευτερόλεπτο, ενώ εφαρμόστηκε και μια μετατόπιση ώστε ο χρονισμός να ξεκινά από το 1ο δευτερόλεπτο αντί για το 0. Οι τιμές των δευτερολέπτων στρογγυλοποιήθηκαν στον πλησιέστερο ακέραιο, διευκολύνοντας έτσι τη σύγκριση και την ομαδοποίηση χρονικών στιγμών.

Το τελικό αποτέλεσμα είναι ένα χρονολογικά ευθυγραμμισμένο dataset, όπου κάθε γεγονός τοποθετείται σε ένα κοινό χρονικό πλαίσιο, ανεξάρτητα από το αρχικό του πλάνο. Αυτή η ενιαία χρονική δομή αποτελεί βασική προϋπόθεση για συνδυαστική ανάλυση με άλλα δεδομένα (όπως οχήματα ή δείκτες εγγύτητας) και για τη δημιουργία οπτικοποιήσεων που βασίζονται στον χρόνο.

5.2.2 Εκκαθάριση και διαχείριση τιμών φωτεινού σηματοδότη

Σε αυτό το στάδιο, στόχος ήταν να αντιμετωπίσουμε **ελλιπίεις ή ασαφείς ενδείξεις** σχετικά με την κατάσταση του φωτεινού σηματοδότη όπως καταγράφηκε στα δεδομένα πεζών. Οι αρχικές τιμές περιλάμβαναν ασαφείς καταγραφές όπως **"unknown"** ή **κενά πεδία (NaN)**, οι οποίες έπρεπε να ερμηνευτούν ή να αντικατασταθούν με πιο ουσιαστική πληροφορία για να επιτραπεί μια σταθερή και αξιόπιστη ανάλυση.

Αρχικά, όλες οι τιμές ομογενοποιήθηκαν ως προς τη μορφή τους (κεφαλαία/πεζά), και κάθε καταγραφή καρέ ταξινομήθηκε χρονικά, ώστε να διατηρηθεί η σωστή χρονική ακολουθία. Στη συνέχεια, ασαφείς τιμές όπως **"unknown"** ή **κενά πεδία** αντικαταστάθηκαν προσωρινά με την ετικέτα **"Intergreen"** – έναν ενδιάμεσο όρο που χρησιμοποιήθηκε για να υποδηλώσει καταστάσεις που δεν μπορούν άμεσα να κατηγοριοποιηθούν.

Για να αποσαφηνιστούν αυτές οι μεταβατικές καταστάσεις, ελέγχθηκε το περιβάλλον κάθε καρέ: αν ένα καρέ ήταν **"Intergreen"** και τόσο το προηγούμενο όσο και το επόμενο είχαν την ίδια σαφή κατάσταση (π.χ. "green"), τότε το καρέ υιοθετούσε αυτήν την κατάσταση. Επιπλέον, για ομάδες συνεχόμενων **"Intergreen"** καρέ (**μέχρι 750 καρέ**), εφαρμόστηκε η ίδια λογική: αν τα καρέ που τα περιβάλλουν έχουν την ίδια τιμή, όλη η ομάδα υιοθετεί αυτήν την κατάσταση.

Τέλος, όσα **"Intergreen"** δεν μπορούσαν να αποσαφηνιστούν με βάση την προαναφερθείσα λογική, μετατράπηκαν ξανά σε **"unknown"** για να διατηρηθεί η διαφάνεια ως προς την αβεβαιότητα αυτών των σημείων.

Με τη διαδικασία αυτή, η στήλη που περιγράφει την κατάσταση των φαναριών (**Traffic Light Status**) καθαρίστηκε και έγινε πιο συνεπής, επιτρέποντας στο εξής τη χρήση της ως αξιόπιστου δείκτη για τη συμπεριφορά των πεζών ή τη χρονική συσχέτισή της με άλλα γεγονότα.

5.2.3 Ορισμός και μοντελοποίηση κύκλων φωτεινής σηματοδότησης

Σε αυτό το στάδιο δημιουργήθηκε ένα **ενιαίο χρονοδιάγραμμα** φωτεινών σηματοδοτών που βασίζεται σε κυκλική επανάληψη των φάσεων φωτεινής σηματοδότησης (κόκκινο, πράσινο, intergreen). Ο πρώτος κύκλος καταγράφηκε

χειροκίνητα, περιλαμβάνοντας τη χρονική έναρξη και λήξη του, καθώς και τη διάρκεια κάθε επιμέρους φάσης. Η διαχείριση του πρώτου κύκλου διαφοροποιήθηκε από τους υπόλοιπους καθώς η εκκίνηση της καταγραφής ξεκίνησε σε σημείο τέτοιο όπου η διάρκεια της κόκκινης φάσης ήταν μόλις 18 δευτερόλεπτα.

Στη συνέχεια, με βάση σταθερές διάρκειες για κάθε φάση (**10 δευτερόλεπτα για το πράσινο, 13 για το intergreen και 67 για το κόκκινο**), κατασκευάστηκαν αυτόματα οι επόμενοι 40 κύκλοι. Κάθε κύκλος ξεκινά από το σημείο που τελείωσε ο προηγούμενος, ενώ για κάθε κύκλο υπολογίζονται συνοπτικά οι χρονικές στιγμές έναρξης, λήξης και η συνολική διάρκειά του.

Το αποτέλεσμα είναι ένας δομημένος πίνακας (DataFrame) που περιέχει 41 κυκλικές ακολουθίες σηματοδότησης με σαφή χρονική πληροφορία, έτοιμος να χρησιμοποιηθεί για χρονική αντιστοίχιση με τα υπόλοιπα δεδομένα.

5.2.4 Χρονική αντιστοίχιση πεζών με κύκλους και φάσεις σηματοδότησης

Σε αυτή τη φάση της επεξεργασίας, κάθε παρατήρηση πεζού αντιστοιχίστηκε χρονικά με τον κύκλο φωτεινής σηματοδότησης στον οποίο ανήκει. Με βάση τη χρονική στιγμή (σε δευτερόλεπτα) κάθε καταγραφής, εντοπίστηκε ο αντίστοιχος κύκλος μέσα από το προκαθορισμένο χρονοδιάγραμμα σηματοδότησης.

Αφού εντοπίστηκε ο κύκλος, ταξινομήθηκε η χρονική στιγμή σε μία από τις τρεις φάσεις: **κόκκινο**, **πράσινο** ή **intergreen** (μεταβατική φάση μεταξύ αλλαγών). Λόγω της δομής του κύκλου, η φάση υπολογίστηκε συγκρίνοντας το χρονικό στίγμα με τα χρονικά όρια κάθε φάσης.

Για τις εγγραφές που δεν ήταν δυνατό να εντοπιστεί άμεσα κύκλος ή φάση, εφαρμόστηκε εμπρόσθια και οπίσθια συμπλήρωση τιμών (forward/backward fill) ώστε να διατηρηθεί η χρονική συνέχεια στα δεδομένα. Επιπλέον, για λόγους απλοποίησης της ανάλυσης αλλά και βάση των παγιωμένων ενδείξεων των φωτεινών σηματοδοτών των πεζών (κόκκινη και πράσινη ένδειξη), οι καταγραφές που ανήκαν στη φάση "Intergreen" επανακαταστηρίστηκαν ως "Red", αφού, ως γνωστόν, η ενδιάμεση φάση αποτελεί μέρος της κόκκινης φάσης.

Το αποτέλεσμα είναι ένα εμπλουτισμένο dataset όπου κάθε καταγραφή πεζού φέρει πληροφορία για τον **κύκλο φωτεινής σηματοδότησης** στον οποίο ανήκει, καθώς και τη **φάση** στην οποία βρισκόταν η κυκλοφορία τη δεδομένη χρονική στιγμή. Αυτό διευκολύνει την ανάλυση της συμπεριφοράς πεζών σε συνάρτηση με τη ροή της σηματοδότησης.

5.2.5 Ομαδοποίηση συμπεριφοράς πεζών ανά χρονική στιγμή

Σε αυτό το στάδιο πραγματοποιήθηκε ομαδοποίηση των δεδομένων σε επίπεδο **δευτερολέπτου και μοναδικού πεζού**, με στόχο την παραγωγή ενός πιο συνοπτικού και διαχειρίσιμου dataset.

Για τις **αριθμητικές μεταβλητές** (όπως ταχύτητα, θέση, εμπιστοσύνη ανίχνευσης), υπολογίστηκε ο **μέσος όρος** ανά ομάδα, ώστε να αποτυπώνεται η συνολική συμπεριφορά του πεζού μέσα στο συγκεκριμένο χρονικό διάστημα.

Για τις **κατηγορικές μεταβλητές** (όπως αν ο πεζός ήταν εντός διασταύρωσης, το στάδιο του φαναριού, η φάση σηματοδότησης κ.ά.), επιλέχθηκε η πιο συχνή τιμή (**mode**) της ομάδας. Αυτό επιτρέπει τη διατήρηση της επικρατούσας κατάστασης σε περιπτώσεις πολλαπλών καταγραφών για τον ίδιο πεζό στο ίδιο δευτερόλεπτο.

Το αποτέλεσμα είναι ένα καθαρό και συνοπτικό αρχείο δεδομένων όπου κάθε γραμμή αντιστοιχεί σε έναν πεζό ανά δευτερόλεπτο, με αριθμητικούς δείκτες συμπεριφοράς και περιγραφικά χαρακτηριστικά.

5.3 Επεξεργασία και ενοποίηση δεδομένων οχημάτων

Όπως και με τα δεδομένα πεζών, εφαρμόστηκε μια σειρά βημάτων επεξεργασίας και εναρμόνισης και στα δεδομένα των οχημάτων. Σε περιπτώσεις όπου η μεθοδολογία ήταν πανομοιότυπη, γίνεται απλή αναφορά στην αντιστοιχία με τα αντίστοιχα βήματα των πεζών, ώστε να αποφεύγεται η περιττή επανάληψη.

5.3.1 Χρονική ενοποίηση και κανονικοποίηση δεδομένων οχημάτων

Κατά αντιστοιχία με την επεξεργασία των δεδομένων πεζών, εφαρμόστηκε η ίδια διαδικασία και στα δεδομένα οχημάτων. Τα επιμέρους πλάνα βίντεο συγχωνεύτηκαν σε ένα ενιαίο και συνεχές χρονικό πλαίσιο μέσω υπολογισμού ενός συνολικού δείκτη καρέ (**ContinuousFrame**), ο οποίος στη συνέχεια μετατράπηκε σε **δευτερόλεπτα και λεπτά**. Με αυτόν τον τρόπο, κάθε καταγραφή οχήματος τοποθετείται σε έναν κοινό χρονικό άξονα, γεγονός που επιτρέπει τη συνεπή σύγκριση και συσχέτιση με άλλα χρονικά δεδομένα.

5.3.2 Εκκαθάριση και διαχείριση τιμών φωτεινού σηματοδότη

Αντίστοιχα με τη διαδικασία που ακολουθήθηκε στα δεδομένα πεζών, πραγματοποιήθηκε καθαρισμός και διόρθωση της στήλης **TrafficLightStatus** και για τα δεδομένα οχημάτων. Ελλιπείς ή ασαφείς τιμές ("**unknown**" ή **κενά**) αντικαταστάθηκαν προσωρινά με την ένδειξη "**Intergreen**", ώστε να ερμηνευθούν βάσει συμφραζομένων. Όπου ήταν δυνατόν, οι τιμές αυτές αποσαφηνίστηκαν με βάση τη χρονική γειτνίαση με προηγούμενα και επόμενα καρέ. Τέλος, όσες τιμές δεν μπορούσαν να αποδοθούν με βεβαιότητα, επανήλθαν στην αρχική ένδειξη "**unknown**".

5.3.3 Ορισμός και μοντελοποίηση κύκλων φωτεινής σηματοδότησης

Ακολουθήθηκε η ίδια διαδικασία με αυτή των πεζών για τη δημιουργία κυκλικού χρονοδιαγράμματος σηματοδότησης. Ο **πρώτος κύκλος** ορίστηκε **χειροκίνητα**, ενώ οι επόμενοι προέκυψαν **αυτόματα με βάση σταθερές διάρκειες** για τις φάσεις κόκκινου, πράσινου και intergreen. Το τελικό αποτέλεσμα είναι ένας πίνακας κύκλων,

όπου κάθε καταγραφή οχήματος μπορεί να αντιστοιχηθεί χρονικά σε συγκεκριμένη φάση σηματοδότησης.

5.3.4 Χρονική αντιστοίχιση οχημάτων με κύκλους και φάσεις σηματοδότησης

Όπως και στα δεδομένα πεζών, κάθε καταγραφή οχήματος αντιστοιχίστηκε χρονικά στον αντίστοιχο **κύκλο φωτεινής σηματοδότησης**, με βάση τη χρονική στιγμή σε δευτερόλεπτα. Στη συνέχεια, καθορίστηκε η **φάση του σηματοδότη** (κόκκινο, πράσινο ή intergreen), και τα όποια κενά ή ασαφή σημεία συμπληρώθηκαν με εμπρόσθια και οπίσθια πλήρωση. Για λόγους ενοποίησης, η φάση "Intergreen" αναταξινομήθηκε ως "Red". Έτσι, κάθε εγγραφή οχήματος φέρει πλέον πληροφορία για τη φάση και τον κύκλο στον οποίο συνέβη.

5.3.5 Ομαδοποίηση συμπεριφοράς οχημάτων ανά χρονική στιγμή

Ακολουθώντας την ίδια προσέγγιση που εφαρμόστηκε για τα δεδομένα πεζών, τα δεδομένα οχημάτων ομαδοποιήθηκαν με βάση τον **μοναδικό αναγνωριστικό αριθμό οχήματος** και το **δευτερόλεπτο καταγραφής**. Για τις **αριθμητικές μεταβλητές** (όπως ταχύτητα και συντεταγμένες θέσης) υπολογίστηκε ο **μέσος όρος**, ενώ για τις **κατηγορικές** (όπως κατάσταση οχήματος ή φάση σηματοδότησης) χρησιμοποιήθηκε η **επικρατούσα τιμή**. Το αποτέλεσμα ήταν ένα συνοπτικό και χρονικά ευθυγραμμισμένο dataset, κατάλληλο για ανάλυση ανά όχημα και χρονική στιγμή.

5.4 Επεξεργασία και διαχείριση δεδομένων ελάχιστου χρόνου έως τη σύγκρουση (minTTC)

5.4.1 Διαχείριση ακραίων τιμών στη μεταβλητή MinTTC

Στο πλαίσιο της επεξεργασίας του συνόλου δεδομένων που περιλαμβάνει τον ελάχιστο χρόνο έως πιθανή σύγκρουση (**MinTTC**), πραγματοποιήθηκε στοχευμένος καθαρισμός ακραίων τιμών. Συγκεκριμένα, οι **τιμές απείρου** (θετικού ή αρνητικού), οι οποίες συνήθως υποδηλώνουν απουσία επικείμενης σύγκρουσης, αντικαταστάθηκαν με μια **υψηλή αριθμητική τιμή** (=9999) που λειτουργεί ως ένδειξη για καταστάσεις μη επικίνδυνες εντός του εξεταζόμενου χρονικού πλαισίου. Με τον τρόπο αυτό, διασφαλίζεται η αριθμητική συμβατότητα του πεδίου και διευκολύνεται η περαιτέρω ανάλυση και συσχέτιση των δεδομένων.

5.4.2 Ενοποίηση χρονικής ροής και μετατροπή Καρέ σε δευτερόλεπτα

Για την εξασφάλιση χρονικής συνέχειας στα δεδομένα minTTC, πραγματοποιήθηκε ενοποίηση των επιμέρους πλάνων βίντεο (**VideoPart**) σε ένα κοινό χρονικό πλαίσιο. Πιο συγκεκριμένα, κάθε καρέ (**FrameNumber**) αναπροσαρμόστηκε ώστε να δημιουργηθεί ένα νέο πεδίο με την ονομασία **ContinuousFrame**, το οποίο απεικονίζει τη θέση κάθε καταγραφής σε μια ενιαία, συνεχόμενη χρονική ακολουθία. Για τον

υπολογισμό του, κάθε επιμέρους πλάνο ευθυγραμμίστηκε με το προηγούμενο, με βάση τον μέγιστο αριθμό καρέ που είχε ήδη καταγραφεί.

Στη συνέχεια, οι τιμές του ContinuousFrame μετατράπηκαν σε **δευτερόλεπτα** και **λεπτά**, χρησιμοποιώντας το frame rate της καταγραφής (**29.74 καρέ ανά δευτερόλεπτο**). Για να υπάρχει σαφές χρονικό σημείο έναρξης, η στήλη Seconds αναπροσαρμόστηκε ώστε η μέτρηση να ξεκινά από το 1. Επιπλέον, οι τιμές των δευτερολέπτων στρογγυλοποιήθηκαν στον πλησιέστερο ακέραιο για να διευκολυνθεί η μελλοντική ομαδοποίηση και συσχέτιση με άλλα δεδομένα.

5.4.3 Δημιουργία κύκλων φωτεινής σηματοδότησης

Για τις ανάγκες χρονικής ανάλυσης των δεδομένων minTTC σε σχέση με τη λειτουργία του φωτεινού σηματοδότη, κατασκευάστηκε ένα **πρότυπο κυκλικό χρονοδιάγραμμα**. Ο πρώτος κύκλος ορίστηκε χειροκίνητα, με καθορισμένη διάρκεια για κάθε φάση (κόκκινο, πράσινο, intergreen) και συνολική διάρκεια 41 δευτερολέπτων. Οι επόμενοι κύκλοι δημιουργήθηκαν με βάση σταθερές διάρκειες για τις φάσεις: **67 δευτερόλεπτα για το κόκκινο, 10 για το πράσινο και 13 για το intergreen**.

Η δημιουργία των κύκλων έγινε με χρονική συνέχεια, έτσι ώστε κάθε κύκλος να ξεκινά ακριβώς από το σημείο όπου τελείωσε ο προηγούμενος. Για κάθε κύκλο καταγράφηκαν τα χρονικά σημεία έναρξης και λήξης, η διάρκεια κάθε φάσης, καθώς και η συνολική διάρκεια του κύκλου. Το αποτέλεσμα είναι ένας δομημένος πίνακας με 41 διαδοχικούς κύκλους σηματοδότησης, έτοιμος για χρονική αντιστοίχιση με τις εγγραφές του dataset minTTC.

5.4.4 Χρονική αντιστοίχιση με κύκλους και φάσεις φωτεινής σηματοδότησης

Σε αυτό το στάδιο, κάθε εγγραφή του dataset minTTC αντιστοιχίστηκε με τον αντίστοιχο **κύκλο φωτεινής σηματοδότησης**, βάσει της χρονικής στιγμής σε δευτερόλεπτα. Για κάθε καταγραφή προσδιορίστηκε ποιος κύκλος περιλαμβάνει τη συγκεκριμένη χρονική στιγμή, και στη συνέχεια αποδόθηκε η κατάλληλη **φάση σηματοδότησης**.

Η φάση καθορίστηκε με βάση τα χρονικά όρια εντός του κύκλου: αν η τιμή χρόνου βρισκόταν εντός της διάρκειας της φάσης κόκκινου ή intergreen, χαρακτηρίστηκε ως **Red**, ενώ αν βρισκόταν εντός της διάρκειας της πράσινης φάσης, χαρακτηρίστηκε ως **Green**. Η ενδιάμεση φάση **Intergreen** δεν διατηρήθηκε διακριτά, αλλά ενσωματώθηκε στην κόκκινη για σκοπούς ανάλυσης.

Για τις εγγραφές που δεν εντάσσονταν απευθείας σε κάποιο κύκλο, εφαρμόστηκε συμπλήρωση τιμών από γειτονικές εγγραφές (μέθοδος forward και backward fill), προκειμένου να διασφαλιστεί η πληρότητα των πεδίων **Cycle Number** και **Phase**. Το τελικό αποτέλεσμα είναι ένα εμπλουτισμένο dataset minTTC, χρονικά ευθυγραμμισμένο με τους κύκλους σηματοδότησης και έτοιμο για συνδυαστική ανάλυση με τα υπόλοιπα δεδομένα.

5.4.5 Ομαδοποίηση δεδομένων ελάχιστου χρόνου έως τη σύγκρουση

Το σύνολο δεδομένων minTTC ομαδοποιήθηκε σε **επίπεδο δευτερολέπτου** και **ζεύγους οχήματος-πεζού**, με στόχο τη δημιουργία ενός συνοπτικού, διαχειρίσιμου πίνακα. Για κάθε μοναδικό συνδυασμό Seconds, VehicleID και PedestrianID, υπολογίστηκε ο μέσος όρος του πεδίου MinTTC, που αποτελεί το βασικό αριθμητικό μέγεθος του συνόλου. Παράλληλα, για τις **κατηγορικές μεταβλητές** Phase και Cycle Number, επιλέχθηκε η επικρατούσα τιμή (**mode**), ώστε να διατηρείται η πληροφορία της φάσης σηματοδότησης στην οποία παρατηρήθηκε η μικρότερη χρονική εγγύτητα σύγκρουσης. Με τον τρόπο αυτό, παράγεται ένα συμπυκνωμένο dataset που διατηρεί την κρίσιμη πληροφορία ανά χρονική στιγμή και ανά ζεύγος οντοτήτων, επιτρέποντας πιο αποτελεσματική ανάλυση και συνδυασμό με τα υπόλοιπα δεδομένα.

5.5 Ενοποίηση και τελικός καθαρισμός δεδομένων

Αφού ολοκληρώθηκε η επεξεργασία των τριών επιμέρους συνόλων δεδομένων (MinTTC, πεζών και οχημάτων), ακολούθησε η **ενοποίησή** τους σε έναν κοινό πίνακα. Η συγχώνευση πραγματοποιήθηκε σταδιακά: αρχικά συνδυάστηκαν τα δεδομένα του MinTTC με εκείνα των πεζών, με βάση τη χρονική στιγμή (**Seconds**) και το **PedestrianID**. Στη συνέχεια, στο ενδιάμεσο αποτέλεσμα προστέθηκαν και τα δεδομένα των οχημάτων, με αντιστοίχιση ως προς τα πεδία **Seconds** και **VehicleID**. Και στις δύο περιπτώσεις χρησιμοποιήθηκε αριστερή συγχώνευση (left join), ώστε να διατηρηθούν όλες οι εγγραφές του αρχικού πίνακα MinTTC, ακόμη και όταν δεν υπήρχε αντίστοιχη πληροφορία για πεζό ή όχημα.

Το αποτέλεσμα ήταν ένας ενιαίος πίνακας που, για κάθε χρονική στιγμή και ζεύγος οχήματος-πεζού, περιλάμβανε τις τιμές MinTTC, τα χαρακτηριστικά κίνησης και θέσης, καθώς και τις αντίστοιχες φάσεις σηματοδότησης – ξεχωριστά για πεζούς και οχήματα.

Στη συνέχεια, δημιουργήθηκε μία καθαρή εκδοχή του ενοποιημένου πίνακα, από την οποία **αφαιρέθηκαν** όλες οι γραμμές που περιείχαν **ελλιπή δεδομένα (NaN)**. Το αρχείο αυτό περιλάμβανε μόνο πλήρεις εγγραφές και ήταν κατάλληλο για εφαρμογή σε αναλυτικές μεθόδους που απαιτούν ένα πλήρες σύνολο χαρακτηριστικών.

Ωστόσο, για τη διασφάλιση της ποιοτικής συνέπειας και αξιοπιστίας των κατηγορικών μεταβλητών, εφαρμόστηκε ένα επιπλέον στάδιο φιλτραρίσματος. Συγκεκριμένα, διατηρήθηκαν μόνο οι εγγραφές στις οποίες οι τιμές των πεδίων **Phase**, **ResolvedTrafficLightStatus**, **PedestrianStatus** και **VehicleStatus** περιείχαν αποδεκτές κατηγορίες. Ειδικότερα, έγιναν αποδεκτές μόνο οι τιμές **"Green"** και **"Red"** για τις φάσεις σηματοδότησης, και **"legal"** ή **"illegal"** για τις καταστάσεις πεζών και οχημάτων, εξαιρώντας περιπτώσεις με ασαφείς ή άγνωστες τιμές όπως **"unknown"**.

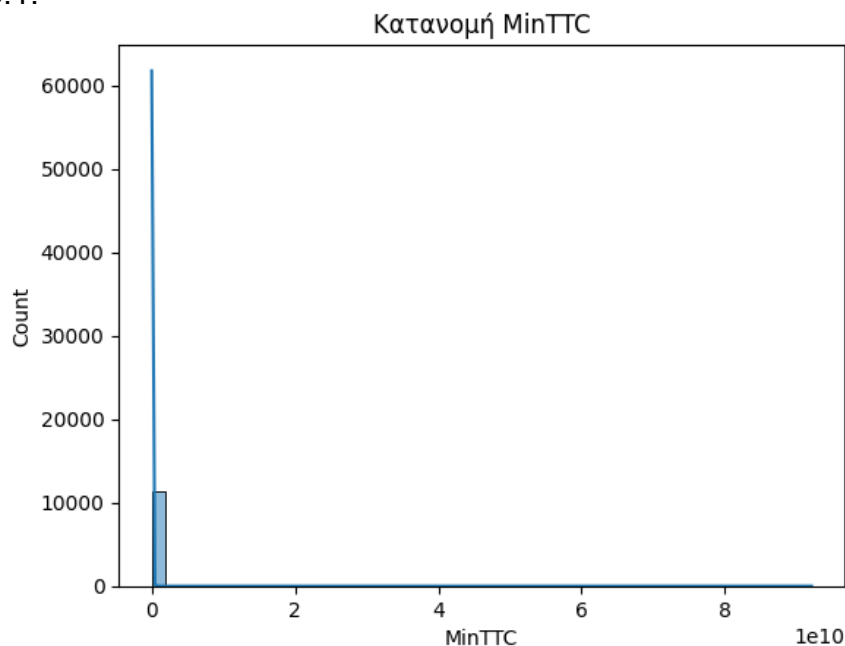
Αυτό το τελικό φιλτράρισμα οδήγησε στην περαιτέρω μείωση του πλήθους των εγγραφών, εξασφαλίζοντας ότι η ανάλυση θα βασίζεται αποκλειστικά σε δεδομένα πλήρως καθορισμένα και αξιόπιστα. Το αποτέλεσμα αυτής της διαδικασίας αποθηκεύτηκε σε νέο αρχείο Excel με την ονομασία filtered_final_combined_info.xlsx, το οποίο περιλαμβάνει αποκλειστικά τις εγγραφές που πληρούν τα προκαθορισμένα

ποιοτικά κριτήρια και είναι έτοιμο προς χρήση σε αναλυτικές και μοντελοποιητικές διεργασίες.

5.6 Διαχείριση ακραίων τιμών και μετασχηματισμός της μεταβλητής MinTTC

Η μεταβλητή **MinTTC** (Minimum Time to Collision) εκφράζει τον ελάχιστο υπολογιζόμενο χρόνο μέχρι την πιθανή σύγκρουση μεταξύ πεζού και οχήματος. Κατά την αρχική διερεύνηση της κατανομής των τιμών της, παρατηρήθηκε ένα ιδιαίτερα προβληματικό στατιστικό χαρακτηριστικό: οι περισσότερες τιμές ήταν συγκεντρωμένες σε **χαμηλά επίπεδα**, κοντά στο μηδέν, ενώ ταυτόχρονα υπήρχαν λίγες τιμές εξαιρετικά **μεγάλου μεγέθους**.

Αυτή η ανισορροπία δημιουργεί αυτό που ονομάζεται **ασύμμετρη κατανομή με δεξιά ουρά** (right-skewed distribution). Με απλά λόγια, αυτό σημαίνει ότι η κατανομή «γέρνει» προς τη μία πλευρά: η πλειοψηφία των παρατηρήσεων βρίσκεται στα αριστερά (δηλαδή σε μικρές τιμές), ενώ μια μικρή ομάδα πολύ μεγάλων τιμών εκτείνεται στα δεξιά — δημιουργώντας μια “ουρά”. Το φαινόμενο απεικονίζεται καθαρά στο Σχήμα 5.1.



Σχήμα 5.1: Ιστόγραμμα της μεταβλητής MinTTC πριν από τον μετασχηματισμό

Η ύπαρξη τέτοιων ακραίων τιμών μπορεί να επηρεάσει σημαντικά τα στατιστικά μοντέλα και να οδηγήσει σε στρεβλά ή ασταθή αποτελέσματα. Ιδίως σε τεχνικές όπως η Λογιστική Παλινδρόμηση ή η point-biserial συσχέτιση, που βασίζονται σε γραμμικές σχέσεις και εύλογη κατανομή των τιμών, τέτοια απόκλιση από την κανονικότητα μπορεί να αλλοιώσει τη συμπεριφορά των μοντέλων. Ακόμη και πιο ευέλικτοι αλγόριθμοι, όπως τα Τυχαία Δάση, είναι ευαίσθητοι στην υπεροχή των outliers στη διαδικασία επιλογής χαρακτηριστικών.

Για την αντιμετώπιση του προβλήματος, εφαρμόστηκε μία συνδυασμένη διαδικασία κανονικοποίησης της μεταβλητής, αποτελούμενη από τα εξής βήματα:

1. **Winsorization:** Αρχικά, **περιορίστηκε η επιρροή των πιο ακραίων τιμών** με τη χρήση του Winsorization. Συγκεκριμένα, ορίστηκαν ως κατώτατο και ανώτατο όριο το 1ο και το 99ο εκατοστημόριο της κατανομής αντίστοιχα. Το 1ο εκατοστημόριο είναι η τιμή κάτω από την οποία πέφτει το 1% των μικρότερων τιμών του δείγματος, ενώ το 99ο εκατοστημόριο είναι η τιμή πάνω από την οποία βρίσκεται το 1% των μεγαλύτερων τιμών. Όλες οι τιμές εκτός αυτού του διαστήματος αντικαταστάθηκαν με τα αντίστοιχα όρια. Έτσι, **διατηρήθηκε το μέγεθος του δείγματος**, αλλά **περιορίστηκε η επιρροή των ακραίων παρατηρήσεων**. Η νέα εκδοχή της μεταβλητής αποθηκεύτηκε ως **MinTTC_wins**.
2. **Λογαριθμικός μετασχηματισμός:** Παρά τον περιορισμό των εξαιρέσεων, η **MinTTC_wins** εξακολουθούσε να εμφανίζει μεγάλες διαφορές μεταξύ τιμών, ακόμη και κατά τάξεις μεγέθους. Για την **περαιτέρω εξομάλυνση της κατανομής** εφαρμόστηκε λογαριθμικός μετασχηματισμός με τη συνάρτηση $\log_{10}(x)$ (δηλαδή $\log(1 + x)$), η οποία είναι κατάλληλη ακόμη και για τιμές κοντά στο μηδέν. Η εφαρμογή λογαρίθμου βασίζεται στο γεγονός ότι η συνάρτηση αυξάνει πιο αργά όσο μεγαλώνει η τιμή της εισόδου. Έτσι, μειώνεται η αριθμητική απόκλιση μεταξύ τιμών και περιορίζεται η επίδραση των μεγαλύτερων παρατηρήσεων. Για παράδειγμα, ενώ η διαφορά μεταξύ των αριθμών 1 και 100 είναι 99 μονάδες, η αντίστοιχη διαφορά μεταξύ $\log(1 + 1)$ και $\log(1 + 100)$ είναι σημαντικά μικρότερη. Ο μετασχηματισμός αυτός **συμπίεζει τις μεγάλες τιμές**, εξισορροπεί την κατανομή και διευκολύνει την εφαρμογή αναλυτικών μοντέλων.

Η κανονικοποιημένη αυτή εκδοχή ενσωματώθηκε στις τρεις βασικές τεχνικές που εφαρμόστηκαν: **Λογιστική Παλινδρόμηση**, **σημειακή διχοτομική συσχέτιση** και **Τυχαία Δάση**, συμβάλλοντας στη βελτίωση της σταθερότητας, της αξιοπιστίας και της ερμηνευσιμότητας των αποτελεσμάτων.

5.7 Ανάλυση Συσχέτισης με Point-Biserial Correlation

Η ανάλυση συσχέτισης αποτελεί βασικό εργαλείο για την προκαταρκτική διερεύνηση σχέσεων μεταξύ μεταβλητών, ιδιαίτερα όταν σκοπός είναι η επιλογή ή η απόρριψη χαρακτηριστικών στο πλαίσιο της ανάπτυξης προβλεπτικών μοντέλων. Στην παρούσα μελέτη χρησιμοποιήθηκε η τεχνική της **point-biserial συσχέτισης**, η οποία είναι κατάλληλη για τη μέτρηση της γραμμικής σχέσης μεταξύ μιας **δυναμικής** (binary) και μιας **συνεχούς** μεταβλητής.

Σε αντίθεση με τον παραδοσιακό συντελεστή Pearson, που προϋποθέτει τη συμμετοχή δύο συνεχών μεταβλητών, η point-biserial επιτρέπει την ανάλυση περιπτώσεων όπου η μία μεταβλητή είναι δυναμικής μορφής (π.χ. 0 και 1). Ο δείκτης συσχέτισης r που προκύπτει κυμαίνεται από -1 έως $+1$ και συνοδεύεται από μία τιμή p (p -value), η οποία δηλώνει το επίπεδο στατιστικής σημαντικότητας της συσχέτισης.

Πριν την εφαρμογή της μεθόδου, διαμορφώθηκε ένα σύνολο ανεξάρτητων μεταβλητών που προέρχονταν από τα τρία βασικά υποσυστήματα της αλληλεπίδρασης: τον πεζό, το όχημα και το περιβάλλον. Περιλήφθηκαν χαρακτηριστικά που αφορούσαν τη θέση, την κατεύθυνση και την ταχύτητα κίνησης (όπως οι συνιστώσες v_x , v_y , η ταχύτητα magnitude, οι χωρικές συντεταγμένες του κέντρου μάζας κ.ά.), καθώς και δείκτες αξιοπιστίας εντοπισμού (confidence),

μεταβλητές σχετικές με τη φάση του σηματοδότη και την κυκλοφοριακή κατάσταση. Οι κατηγορικές μεταβλητές μετατράπηκαν σε δυαδικές μορφές ώστε να καταστούν συμβατές με τη στατιστική τεχνική.

Όσον αφορά τη μεταβλητή MinTTC, εφαρμόστηκε, όπως έχει ήδη αναφερθεί, διαδικασία κανονικοποίησης με **Winsorization** και **λογαριθμικό μετασχηματισμό**, προκειμένου να περιοριστεί η επίδραση ακραίων τιμών και να εξομαλυνθεί η ασυμμετρία της κατανομής. Η διαδικασία αυτή θεωρήθηκε κρίσιμη, ώστε η εν λόγω μεταβλητή να συμμετέχει με σταθερό και ερμηνεύσιμο τρόπο στη συσχετιστική ανάλυση.

Η μέθοδος εφαρμόστηκε σε δύο διαφορετικές περιπτώσεις, με στόχο να εξεταστούν διαφορετικές όψεις της συμπεριφοράς του πεζού. Στην πρώτη περίπτωση, η εξαρτημένη (δυαδική) μεταβλητή ήταν η **IsInCrossingArea**, που δηλώνει αν ο πεζός βρισκόταν εντός της περιοχής διάσχισης. Στη δεύτερη περίπτωση, χρησιμοποιήθηκε ως μεταβλητή στόχου η **PedestrianStatus**, η οποία καταγράφει αν η συμπεριφορά του πεζού χαρακτηριζόταν ως «νόμιμη» ή «παράνομη». Οι δύο αναλύσεις παρουσιάζονται στις υποενότητες που ακολουθούν.

5.7.1 Ανάλυση Συσχέτισης με τη μεταβλητή IsInCrossingArea

Η πρώτη φάση της ανάλυσης point-biserial συσχέτισης επικεντρώθηκε στη μεταβλητή **IsInCrossingArea**, η οποία δηλώνει αν ο πεζός βρισκόταν **εντός** (τιμή 1) ή **εκτός** (τιμή 0) της περιοχής διάσχισης κατά τη στιγμή καταγραφής. Στόχος ήταν η διερεύνηση της σχέσης μεταξύ αυτής της χωρικής μεταβλητής και αριθμητικών χαρακτηριστικών που αφορούν την κινητική συμπεριφορά πεζών και οχημάτων.

Η στατιστική ανάλυση εφαρμόστηκε σε κάθε αριθμητική μεταβλητή ξεχωριστά, με υπολογισμό του συντελεστή συσχέτισης (r) και της τιμής στατιστικής σημαντικότητας (p -value). Οι μεταβλητές ταξινομήθηκαν βάσει του p -value, προκειμένου να εντοπιστούν εκείνες που παρουσιάζουν στατιστικά σημαντική σχέση με τη μεταβλητή θέσης του πεζού.

Στη συνέχεια παρουσιάζονται τα αποτελέσματα και οι σημαντικότερες παρατηρήσεις που προκύπτουν από την ανάλυση.

5.7.2 Αποτελέσματα συσχέτισης με τη μεταβλητή IsInCrossingArea

Η στατιστική ανάλυση point-biserial αποκάλυψε πώς διάφορες ανεξάρτητες μεταβλητές σχετίζονται με τη μεταβλητή IsInCrossingArea, η οποία δηλώνει αν ένας πεζός βρίσκεται εντός της περιοχής διάσχισης. Ο βαθμός και η κατεύθυνση κάθε συσχέτισης επιτρέπουν τη στατιστική ερμηνεία της συμβολής κινηματικών, σηματοδοτικών, χωρικών και εντοπιστικών παραμέτρων. Πιο συγκεκριμένα, τα αποτελέσματα παρουσιάζονται στον παρακάτω πίνακα:

Variable	Correlation	P-value
MinTTC_wins_log	0.288	0
GroundPlaneCentroid_1	-0.058	0
vy	0.168	0
vx	-0.071	0
magnitude	0.422	0
vy_veh	-0.175	0
vx_veh	0.275	0
GroundPlaneCentroid_1_veh	0.099	0
VehicleStatus	0.048	0
PedestrianStatus	-0.625	0
magnitude_veh	-0.208	0
ResolvedTrafficLightStatus	0.194	0
IsInCrossingArea_veh	-0.151	0
GroundPlaneCentroid_2_veh	-0.047	0
Phase	-0.038	0
Confidence_veh	-0.030	0.01
GroundPlaneCentroid_2	-0.03	0.001
Confidence	0.018	0.055

Πίνακας 5.1: Αποτελέσματα συσχέτισης

Από την ανάλυση συσχέτισης point biserial μεταξύ της παρουσίας πεζού εντός ή εκτός διάβασης και επιλεγμένων μεταβλητών, προέκυψαν τρεις παράγοντες που εμφάνισαν στατιστικά σημαντική και τουλάχιστον μέτρια συσχέτιση:

- PedestrianStatus** ($r = -0.625$, $p < 0.001$): Η ισχυρή αρνητική συσχέτιση δείχνει ότι οι παραβάτες πεζοί εντοπίζονται συχνά **εντός διάβασης**. Το εύρημα αυτό συνδέεται με τη **μεθοδολογική δομή** της μελέτης, στην οποία η παρανομία ορίζεται με βάση την **παραβίαση** φωτεινού σηματοδότη και καταγράφεται μόνο **εντός** προκαθορισμένων σηματοδοτούμενων διαβάσεων. Συνεπώς, ένας πεζός χαρακτηρίζεται ως παραβάτης μόνο όταν εισέρχεται ενεργά στη διάβαση με κόκκινο, γεγονός που καθιστά την παραβατικότητα οργανωμένη και **ενσυνείδητη**, όχι χαοτική ή άτυπη. Αντίθετα, η παρουσία πεζών **εκτός διάβασης** δεν μπορεί να αξιολογηθεί βάσει του σηματοδότη και συνεπώς **δεν καταγράφεται ως παραβατική**, αν και πιθανόν να υποδηλώνει **άτυπη** ή ακανόνιστη μορφή διάσχισης, εκτός του τυπικού πλαισίου. Το στοιχείο αυτό ενδέχεται να σχετίζεται είτε με μη συμμόρφωση που δεν καταγράφεται από το μοντέλο, είτε με αδυναμία του αστικού περιβάλλοντος να καθοδηγήσει τη ροή των πεζών.
- magnitude** ($r = 0.422$, $p < 0.001$): Η συνολική ταχύτητα του πεζού παρουσιάζει μέτρια θετική συσχέτιση με την παρουσία εντός διάβασης, γεγονός που υποδηλώνει ότι όσο **αυξάνεται η ταχύτητα**, αυξάνεται και η πιθανότητα ο πεζός να βρίσκεται **εντός διάβασης**. Η αυξημένη ταχύτητα φαίνεται να συνδέεται με δύο διαφορετικά προφίλ πεζών: παραβάτες που επιταχύνουν καθώς διασχίζουν με κόκκινο, επιδιώκοντας να μειώσουν τον χρόνο έκθεσης σε πιθανή σύγκρουση, και νόμιμους πεζούς που κινούνται ταχύτερα πιθανώς λόγω αίσθησης ασφάλειας, σαφούς προτεραιότητας ή ομαλής ροής. Σε κάθε

περίπτωση, η ταχύτητα εντός διάβασης συνδέεται με πιο αποφασιστική και συνεχή κινητική συμπεριφορά.

- **MinTTC_log** ($r = 0.288$, $p < 0.001$): Η μεταβλητή που εκφράζει τον ελάχιστο αναμενόμενο χρόνο έως πιθανή σύγκρουση παρουσιάζει θετική συσχέτιση με την παρουσία εντός διάβασης. Αυτό δείχνει ότι, είτε πρόκειται για νόμιμους είτε για παραβάτες πεζούς, όταν η διάσχιση γίνεται **εντός διάβασης**, παρατηρείται κατά μέσο όρο **μεγαλύτερο περιθώριο ασφαλείας**. Στην περίπτωση των παραβατών, ο αυξημένος χρόνος πιθανόν οφείλεται στο ότι η διάσχιση πραγματοποιείται όταν εντοπίζεται προσωρινό κενό στην κυκλοφορία, παρά την ύπαρξη κόκκινου σηματοδότη. Στην περίπτωση των νόμιμων, ο αυξημένος χρόνος σχετίζεται με τη διακοπή της κυκλοφορίας οχημάτων λόγω πράσινου σηματοδότη για πεζούς.

Οι υπόλοιπες μεταβλητές παρουσίασαν ασθενείς συσχετίσεις ($|r| < 0.28$) και για λόγους πληρότητας παρατίθενται συνοπτικά στον πίνακα, χωρίς περαιτέρω σχολιασμό.

5.7.3 Ανάλυση Συσχέτισης με τη μεταβλητή PedestrianStatus

Στο δεύτερο σκέλος της ανάλυσης, εξετάστηκε η σχέση μεταξύ της μεταβλητής PedestrianStatus, η οποία αποτυπώνει τη νομιμότητα της πεζής συμπεριφοράς (1 = νόμιμη, 0 = παράνομη), και μιας σειράς ανεξάρτητων αριθμητικών μεταβλητών που σχετίζονται με την κινηματική κατάσταση, τα χωρικά δεδομένα και τις συνθήκες σηματοδότησης.

Η εφαρμογή της point-biserial συσχέτισης επέτρεψε την αξιολόγηση κάθε μεταβλητής ως προς τη στατιστική της συσχέτιση με το αν ένας πεζός κινήθηκε εντός ή εκτός των ορίων του Κ.Ο.Κ. Η ταξινόμηση των μεταβλητών έγινε βάσει του p-value, με στόχο την ανάδειξη εκείνων που σχετίζονται περισσότερο με την εκδήλωση **παραβατικής ή συμμορφωμένης** συμπεριφοράς.

Τα αποτελέσματα παρουσιάζονται στη συνέχεια, μαζί με τα σημαντικότερα ευρήματα που προέκυψαν από την ανάλυση.

5.7.4 Αποτελέσματα συσχέτισης με τη μεταβλητή PedestrianStatus

Η στατιστική ανάλυση point-biserial αποκάλυψε πώς διάφορες ανεξάρτητες μεταβλητές σχετίζονται με τη μεταβλητή IsInCrossingArea, η οποία δηλώνει αν ένας πεζός κινείται νόμιμα ή παράνομα κατά την διάσχιση της διάβασης. Ο βαθμός και η κατεύθυνση κάθε συσχέτισης επιτρέπουν τη στατιστική ερμηνεία της συμβολής κινηματικών, σηματοδοτικών, χωρικών και εντοπιστικών παραμέτρων. Πιο συγκεκριμένα, τα αποτελέσματα παρουσιάζονται στον παρακάτω πίνακα:

Variable	Correlation	P-value
MinTTC_wins_log	-0.173	0
magnitude	-0.283	0
vy	-0.104	0
vx	0.123	0
magnitude_veh	0.113	0
ResolvedTrafficLightStatus	0.14	0
vx_veh	-0.144	0
vy_veh	0.087	0
IsInCrossingArea_veh	0.065	0
IsInCrossingArea	-0.625	0
Phase	0.0453	0
Confidence_veh	0.043	0
GroundPlaneCentroid_1_veh	-0.042	0
GroundPlaneCentroid_1	0.036	0
Confidence	-0.021	0.03
VehicleStatus	-0.020	0.04
GroundPlaneCentroid_2	0.016	0.089
GroundPlaneCentroid_2_veh	0.003	0.694

Πίνακας 5.2: Αποτελέσματα συσχέτισης

Από την ανάλυση συσχέτισης point biserial με εξαρτημένη μεταβλητή την PedestrianStatus (0 = παράνομος, 1 = νόμιμος), προέκυψαν τρεις μεταβλητές με στατιστικά σημαντική και τουλάχιστον μέτρια συσχέτιση:

- **IsInCrossingArea** ($r = -0.625$, $p < 0.001$): Η ισχυρή αρνητική συσχέτιση δείχνει ότι οι παραβάτες πεζοί εντοπίζονται συχνά εντός της περιοχής διάβασης. Το εύρημα αυτό συνδέεται άμεσα με τον τρόπο ορισμού της παραβατικότητας στη μελέτη, ο οποίος βασίζεται στην παραβίαση φωτεινού σηματοδότη και όχι στην τοποθεσία του πεζού ως προς τη διάβαση. Η **παραβατική συμπεριφορά**, επομένως, δεν σχετίζεται με άτυπη ή αποδιοργανωμένη διέλευση (π.χ. εκτός διάβασης), αλλά εκδηλώνεται **εντός οργανωμένου οδικού πλαισίου** και ενδέχεται να αποτελεί συνειδητή επιλογή μη συμμόρφωσης.
- **magnitude** ($r = -0.283$, $p < 0.001$): Η συνολική ταχύτητα του πεζού εμφανίζει μέτρια αρνητική συσχέτιση με το αν είναι νόμιμος. Αυτό δείχνει ότι οι **παράνομοι** πεζοί τείνουν να κινούνται **ταχύτερα** κατά τη διάσχιση, πιθανότατα επιδιώκοντας να περιορίσουν τον χρόνο έκθεσής τους στον κίνδυνο. Αντίθετα, οι νόμιμοι πεζοί κινούνται με πιο σταθερό και ομαλό ρυθμό, εντός της προβλεπόμενης χρονικής φάσης για διάσχιση.
- **MinTTC_log** ($r = -0.173$, $p < 0.001$): Η αρνητική συσχέτιση δείχνει ότι όσο **μεγαλύτερος είναι ο διαθέσιμος χρόνος έως πιθανή σύγκρουση**, τόσο αυξάνεται η πιθανότητα ο πεζός να είναι **παραβάτης**. Το εύρημα αυτό ερμηνεύεται από τη στρατηγική συμπεριφορά των παραβατών, οι οποίοι συνήθως περιμένουν να δημιουργηθεί ένα προσωρινό κενό στην κυκλοφορία ώστε να διασχίσουν με κόκκινο — άρα η καταγραφή γίνεται σε συνθήκες όπου τα οχήματα είναι μακριά και το περιθώριο ασφαλείας είναι μεγάλο (υψηλό TTC). Αντίθετα, οι νόμιμοι πεζοί διασχίζουν όταν έχουν πράσινο, ακόμη κι αν κάποιο όχημα πλησιάζει για να σταματήσει. Επειδή το TTC υπολογίζεται βάσει της στιγμιαίας ταχύτητας του οχήματος, καταγράφεται ως μικρότερο, ακόμα και

όταν δεν υφίσταται πραγματικός κίνδυνος. Αυτό οδηγεί σε χαμηλότερες τιμές TTC για τους νόμιμους πεζούς, χωρίς να αντανakλά απαραίτητα αυξημένο ρίσκο.

Οι υπόλοιπες μεταβλητές παρουσίασαν χαμηλότερη συσχετιστική ισχύ ($|r| < 0.17$) και δεν αξιολογούνται περαιτέρω στο παρόν σημείο, αλλά παρατίθενται στον συνοδευτικό πίνακα για λόγους πληρότητας.

5.7.5 Οπτική διερεύνηση της ταχύτητας σε σχέση με τη νομιμότητα της πεζής συμπεριφοράς

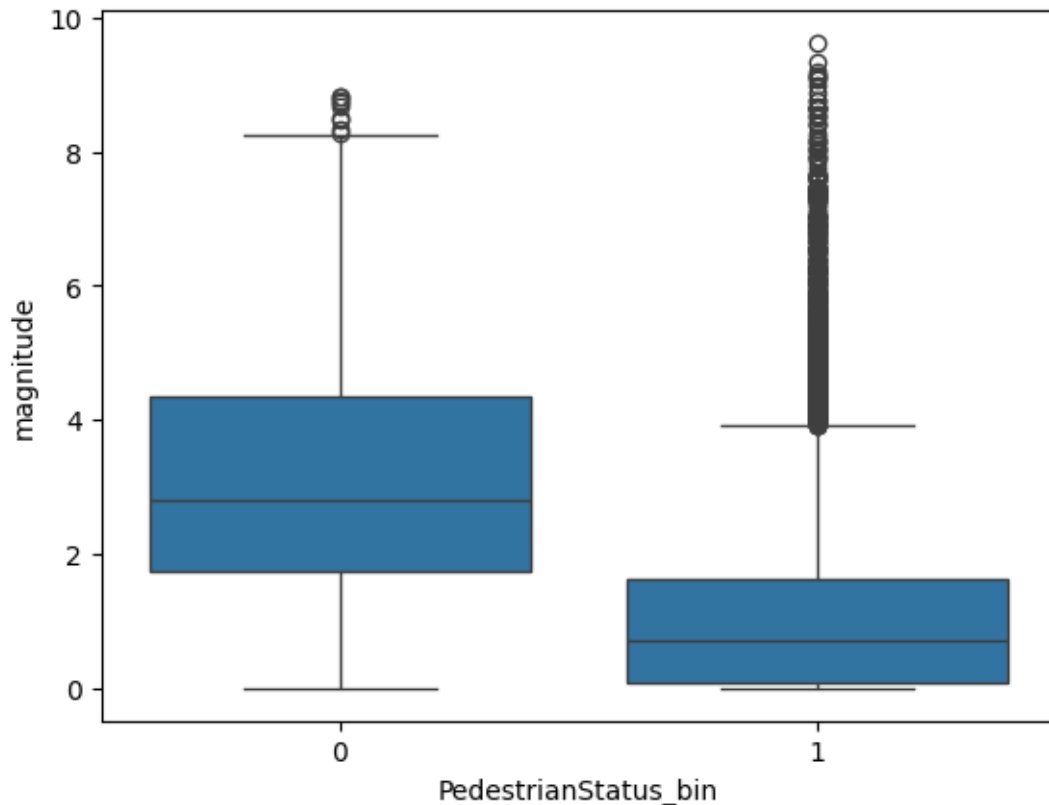
Προκειμένου να ενισχυθούν τα ευρήματα της στατιστικής ανάλυσης που παρουσιάστηκαν στην ενότητα 5.7.4, δημιουργήθηκε οπτική απεικόνιση της μεταβλητής **magnitude**, δηλαδή της συνολικής ταχύτητας του πεζού, σε σχέση με τη μεταβλητή **PedestrianStatus**, που υποδηλώνει αν η συμπεριφορά ήταν νόμιμη (1) ή παράνομη (0). Το Σχήμα 5.2 παρουσιάζει ένα boxplot, μέσω του οποίου είναι δυνατή η απευθείας σύγκριση της κατανομής ταχύτητας μεταξύ των δύο κατηγοριών συμπεριφοράς.

Παρατηρείται ότι οι **παράνομοι πεζοί** εμφανίζουν κατά μέσο όρο **υψηλότερες τιμές ταχύτητας**, όπως δείχνει η μεγαλύτερη διάμεσος της κατηγορίας 0. Το εύρος των τιμών είναι επίσης μεγαλύτερο, γεγονός που υποδηλώνει **μεγαλύτερη ποικιλία στη συμπεριφορά**, με πεζούς που κινούνται είτε πολύ γρήγορα είτε πιο αργά — πιθανόν λόγω βιασύνης, αβεβαιότητας ή απόπειρας αποφυγής κινδύνου. Αντίθετα, οι **νόμιμοι πεζοί** (τιμή 1) παρουσιάζουν **χαμηλότερες και πιο συμπυκνωμένες τιμές ταχύτητας**, κάτι που υποδηλώνει πιο **σταθερή** και ελεγχόμενη **συμπεριφορά**, πιθανότατα επειδή διασχίζουν υπό συνθήκες προτεραιότητας (π.χ. πράσινο φανάρι).

Ιδιαίτερη σημασία έχουν και οι **outliers** που εμφανίζονται στην κατηγορία των νόμιμων πεζών: δείχνουν ότι, σε κάποιες περιπτώσεις, ακόμη και όταν ο πεζός κινείται νομίμως, μπορεί να αναπτύξει **υψηλή ταχύτητα**, είτε λόγω ατομικών διαφορών είτε λόγω τοπικών συγκυριών.

Η εικόνα αυτή συμφωνεί με τα αριθμητικά αποτελέσματα που παρουσιάστηκαν στην ενότητα 5.7.4, όπου βρέθηκε **αρνητική συσχέτιση** μεταξύ της **ταχύτητας** (magnitude) και της **νομιμότητας** (PedestrianStatus) ($r = -0.283$, $p < 0.001$). Το boxplot ενισχύει αυτό το συμπέρασμα και προσφέρει διαισθητική κατανόηση του φαινομένου: οι πεζοί που παραβιάζουν τους κανόνες συχνά κινούνται πιο γρήγορα και πιο απρόβλεπτα, ενώ εκείνοι που τηρούν τον Κ.Ο.Κ. παρουσιάζουν πιο σταθερή κινητική συμπεριφορά.

Το εύρημα αυτό, σε συνδυασμό με τα υπόλοιπα αποτελέσματα της συσχέτισης (π.χ. MinTTC, φάση σηματοδότη, χωρικά χαρακτηριστικά), συμβάλλει στη διαμόρφωση ενός πιο πλήρους προφίλ ρίσκου, βάσει του οποίου η **ταχύτητα του πεζού** μπορεί να αξιοποιηθεί ως **προγνωστικός δείκτης παραβατικής συμπεριφοράς** στο πλαίσιο μοντελοποίησης ή λήψης αποφάσεων.



Σχήμα 5.2: Boxplot ταχύτητας (magnitude) σε σχέση με τη νομιμότητα της πεζής συμπεριφοράς (PedestrianStatus)

5.7.6 Οπτική διερεύνηση της εγγύτητας σε σύγκρουση σε σχέση με τη νομιμότητα της πεζής συμπεριφοράς

Για να κατανοηθεί καλύτερα η σχέση μεταξύ της **νομιμότητας της πεζής συμπεριφοράς** και του **χρόνου που απομένει έως μια πιθανή σύγκρουση** με όχημα, δημιουργήθηκε το Σχήμα 5.3, το οποίο παρουσιάζει ένα boxplot της μεταβλητής MinTTC_wins_log (λογαριθμισμένος ελάχιστος χρόνος έως σύγκρουση), ανάλογα με τη μεταβλητή PedestrianStatus (0 = παράνομος, 1 = νόμιμος).

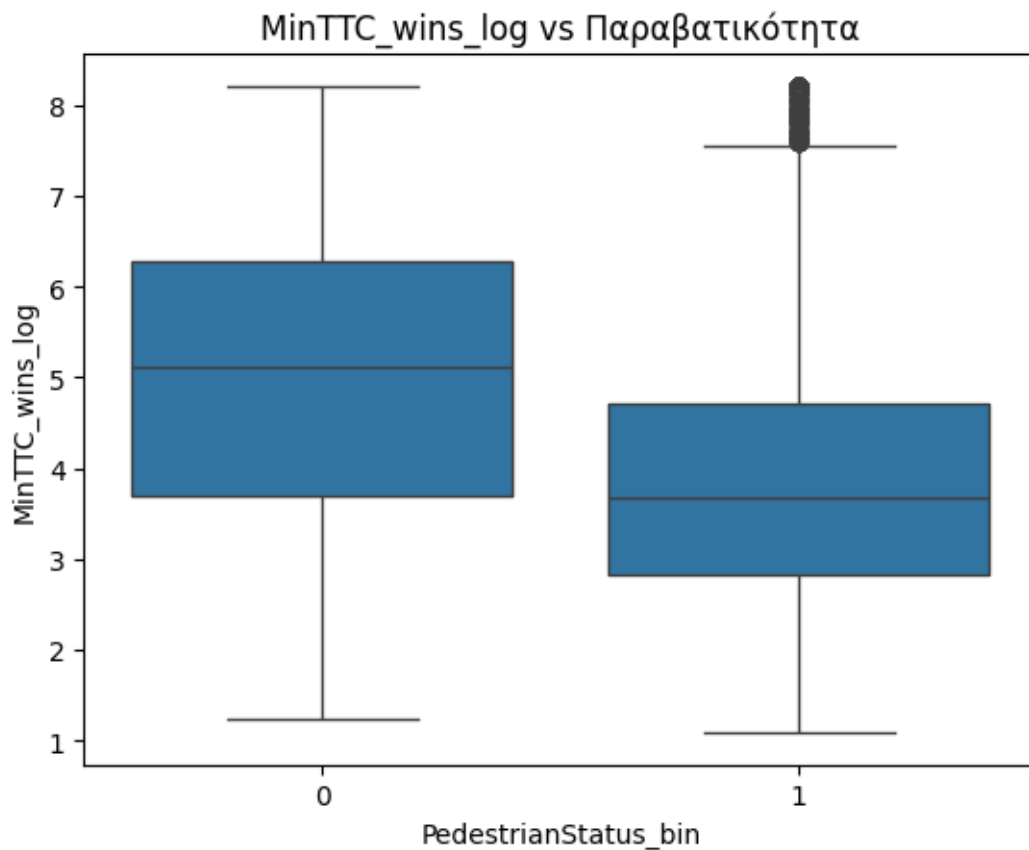
Το boxplot είναι ένας τρόπος να δούμε πώς κατανέμονται οι τιμές μιας μεταβλητής. Η **γραμμή μέσα στο κουτί** δείχνει τη **διάμεσο**, δηλαδή τη μέση τιμή. Το ίδιο το **κουτί περιλαμβάνει το 50% των τιμών** (από το 1ο έως το 3ο τεταρτημόριο), και τα "μουστάκια" που εκτείνονται δεξιά και αριστερά δείχνουν την ελάχιστη και μέγιστη "κανονική" τιμή. Τέλος, οι **τελείες έξω από τα μουστάκια** δείχνουν **ακραίες τιμές** (outliers).

Στο συγκεκριμένο γράφημα, παρατηρούμε ότι οι **παράνομοι πεζοί** (τιμή 0) έχουν **υψηλότερη διάμεσο** σε σχέση με τους νόμιμους. Αυτό σημαίνει πως, κατά μέσο όρο, **έχουν περισσότερο χρόνο μέχρι την πιθανή σύγκρουση** (δηλαδή η απόσταση από τα οχήματα είναι μεγαλύτερη). Το κουτί είναι επίσης πιο "ψηλό", κάτι που δείχνει **μεγαλύτερη ποικιλία στις τιμές** – άλλοι πεζοί διασχίζουν πολύ νωρίς, άλλοι πιο οριακά.

Αντίθετα, οι **νόμιμοι πεζοί** (τιμή 1) έχουν **χαμηλότερη διάμεσο**, άρα **καταγράφονται πιο κοντά σε επερχόμενα οχήματα**, πιθανώς επειδή διασχίζουν τη στιγμή που τα οχήματα έχουν ήδη επιβραδύνει ή σταματήσει λόγω φαναριού. Εμφανίζονται όμως και αρκετοί **outliers** με πολύ υψηλές τιμές: πρόκειται πιθανότατα για περιπτώσεις όπου η διάσχιση ξεκίνησε νωρίς, και τα οχήματα βρίσκονταν μακριά.

Αυτό το μοτίβο **συμφωνεί** απόλυτα με τη στατιστική ανάλυση της ενότητας 5.7.4, όπου διαπιστώθηκε αρνητική συσχέτιση ανάμεσα στις δύο μεταβλητές ($r = -0.173$, $p < 0.001$). Με απλά λόγια, όσο μεγαλύτερος είναι ο χρόνος έως πιθανή σύγκρουση (MinTTC), τόσο πιο πιθανό είναι ο πεζός να παρανομεί.

Συνολικά, το γράφημα δείχνει ότι οι **παράνομοι πεζοί** τείνουν να **διασχίζουν** όταν ο **δρόμος είναι άδειος**, δηλαδή όταν δεν υπάρχουν κοντινά οχήματα, ενώ οι **νόμιμοι πεζοί διασχίζουν** όταν έχουν **προτεραιότητα**, αλλά **ο χρόνος μέχρι την επαφή με το όχημα είναι μικρότερος** επειδή το όχημα πλησιάζει ή έχει μόλις σταματήσει. Η οπτικοποίηση αυτή επιβεβαιώνει και συμπληρώνει τα αριθμητικά ευρήματα του μοντέλου.



Σχήμα 5.3: Boxplot του λογαριθμισμένου ελάχιστου χρόνου έως πιθανή σύγκρουση (MinTTC_wins_log) σε σχέση με τη νομιμότητα της πεζής συμπεριφοράς (PedestrianStatus)

5.8 Λογιστική Παλινδρόμηση

5.8.1 Εισαγωγή

Η λογιστική παλινδρόμηση εφαρμόστηκε με στόχο την πρόβλεψη των μεταβλητών **IsInCrossingArea** και **PedestrianStatus**, οι οποίες δηλώνουν αν ένας πεζός εντοπίζεται εντός (τιμή 1) ή εκτός (τιμή 0) της περιοχής διάβασης και αν ένας πεζός κινείται νόμιμα (τιμή 1) ή παράνομα (τιμή 0), αντίστοιχα. Πρόκειται για δυαδικές μεταβλητές, δηλαδή παίρνουν μόνο δύο τιμές, και για αυτόν τον λόγο η λογιστική παλινδρόμηση αποτελεί κατάλληλη μεθοδολογική επιλογή. Δεν λειτουργεί απλώς ως εργαλείο πρόβλεψης, αλλά και ως μέσο ερμηνείας, καθώς επιτρέπει την ποσοτική εκτίμηση της επίδρασης κάθε ανεξάρτητης μεταβλητής στην πιθανότητα εμφάνισης του υπό εξέταση φαινομένου. Επιπλέον, μέσω στατιστικών δεικτών, καθίσταται δυνατός ο έλεγχος της στατιστικής σημαντικότητας κάθε παράγοντα, προσφέροντας τόσο προβλεπτική ακρίβεια όσο και ερμηνευσιμότητα.

Για την επιλογή των καταλληλότερων λογιστικών μοντέλων, εξετάστηκαν ποικίλοι **συνδυασμοί ανεξάρτητων μεταβλητών**, με σκοπό τη βελτιστοποίηση της προβλεπτικής απόδοσης και της στατιστικής εγκυρότητας. Η αξιολόγηση των μοντέλων βασίστηκε σε πλήθος δεικτών: υπολογίστηκε ο **πίνακας συσχετίσεων** (correlation matrix) και ο **VIF** (Variance Inflation Factor) για την ανίχνευση και αντιμετώπιση πιθανής πολυσυγγραμμικότητας μεταξύ των μεταβλητών· παράλληλα, η αποδοτικότητα κάθε μοντέλου εκτιμήθηκε μέσω του **πίνακα σύγχυσης** (confusion matrix), της αναφοράς ταξινόμησης (classification report) και της **καμπύλης ROC** με την αντίστοιχη τιμή **AUC**, ενώ τα κριτήρια **AIC** και **BIC** χρησιμοποιήθηκαν για τη συγκριτική αξιολόγηση της ποιότητας προσαρμογής. Ο **pseudo-R²** συνέβαλε στην εκτίμηση του βαθμού εξήγησης της διακύμανσης, ενώ η στατιστική σημαντικότητα των παραγόντων εξετάστηκε μέσω των τιμών **p-value**. Η συνδυαστική ανάλυση όλων αυτών των παραμέτρων επέτρεψε την επιλογή των πλέον αξιόπιστων, προβλεπτικά ισχυρών και ερμηνεύσιμων μοντέλων για την πρόβλεψη των μεταβλητών **IsInCrossingArea** και **PedestrianStatus**.

5.8.2 Μοντέλο 1: Εξαρτημένη μεταβλητή – Παρουσία πεζού εντός/εκτός διάβασης

Ποιότητα προσαρμογής μοντέλου

Η συνολική προσαρμογή του μοντέλου αξιολογήθηκε μέσω του log-likelihood και του likelihood ratio test (LLR). Το μοντέλο εμφάνισε τιμή log-likelihood ίση με **-3533.2**, η οποία είναι σημαντικά βελτιωμένη σε σχέση με την τιμή του null model (**-9896.8**), που περιλαμβάνει μόνο τη σταθερά. Η τιμή του **LLR p-value** ήταν **< 0.001**, γεγονός που υποδηλώνει ότι το μοντέλο με τις ανεξάρτητες μεταβλητές παρουσιάζει στατιστικά σημαντικά καλύτερη προσαρμογή από το μοντέλο χωρίς εξηγητικούς παράγοντες. Επιπλέον, το **Pseudo R²** (McFadden) του μοντέλου ήταν **0.643**, τιμή που θεωρείται υψηλή για λογιστική παλινδρόμηση, υποδεικνύοντας ότι το μοντέλο εξηγεί σε πολύ μεγάλο βαθμό τη διακύμανση της εξαρτημένης μεταβλητής.

Τα αποτελέσματα παρουσιάζονται στον επισυναπτόμενο πίνακα:

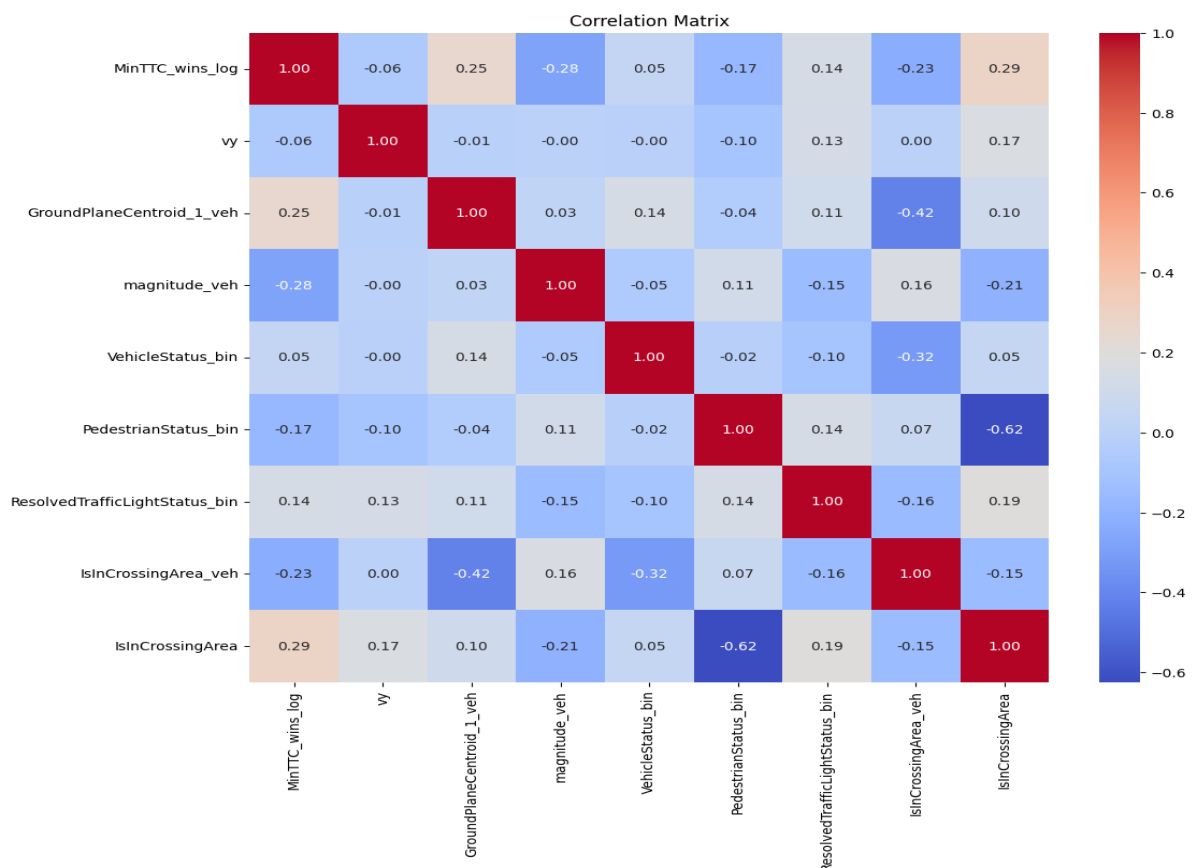
0	1	2	3
Model:	Logit	Method:	MLE
Dependent Variable:	IsInCrossingArea	Pseudo R-squared:	0.643
Date:	2025-07-10 07:54	AIC:	7084.3290
No. Observations:	14278	BIC:	7152.4273
Df Model:	8	Log-Likelihood:	-3533.2
Df Residuals:	14269	LL-Null:	-9896.8
Converged:	1.0000	LLR p-value:	0.0000
No. Iterations:	10.0000	Scale:	1.0000

Πίνακας 5.3: Στατιστικά χαρακτηριστικά λογιστικού μοντέλου

Έλεγχος πολυσυγγραμμικότητας

Πραγματοποιήθηκε έλεγχος πολυσυγγραμμικότητας μεταξύ των ανεξάρτητων μεταβλητών μέσω του πίνακα συσχετίσεων και των δεικτών VIF. Καμία μεταβλητή δεν παρουσίασε υψηλή συσχέτιση με άλλη ($r > 0.7$), ενώ όλες οι τιμές του δείκτη **VIF** ήταν **κάτω από 1.5**, με τη μεγαλύτερη να ανέρχεται σε 1.41 (για τη μεταβλητή IsInCrossingArea_veh). Οι τιμές αυτές υποδεικνύουν **απουσία πολυσυγγραμμικότητας**, ενισχύοντας τη σταθερότητα των συντελεστών και τη γενική αξιοπιστία του μοντέλου.

Τα σχετικά αποτελέσματα παρουσιάζονται στον επισυναπτόμενο πίνακα:



Πίνακας 5.4: Πίνακας συσχέτισης λογιστικού μοντέλου

Variable	VIF
const	1
MinTTC_wins_log	1.23
vy	1.05
GroundPlaneCentroid_1_veh	1.28
magnitude_veh	1.14
VehicleStatus	1.14
PedestrianStatus	1.09
ResolvedTrafficLightStatus	1.15
IsInCrossingArea_veh	1.41

Πίνακας 5.5: Τιμές VIF λογιστικού μοντέλου

Συντελεστές και στατιστική σημαντικότητα

Όλες οι ανεξάρτητες μεταβλητές στο τελικό μοντέλο βρέθηκαν στατιστικά σημαντικές σε επίπεδο $p < 0.05$, γεγονός που υποδηλώνει ότι συμβάλλουν ουσιαστικά στην πρόβλεψη της παρουσίας πεζού εντός ή εκτός διάβασης. Συγκεκριμένα, οι μεταβλητές **MinTTC_wins_log** ($\beta = 0.96$, $p < 0.001$), **vy** ($\beta = 0.38$, $p < 0.001$), **GroundPlaneCentroid_1_veh** ($\beta = 0.08$, $p = 0.036$) και **VehicleStatus** ($\beta = 0.37$, $p < 0.001$) εμφάνισαν θετικούς συντελεστές, γεγονός που υποδηλώνει ότι η αύξησή τους σχετίζεται με αυξημένη πιθανότητα παρουσίας πεζού **εντός** διάβασης.

Αντίθετα, οι μεταβλητές **magnitude_veh** ($\beta = -0.24$, $p < 0.001$), **PedestrianStatus** ($\beta = -1.97$, $p < 0.001$) και **IsInCrossingArea_veh** ($\beta = -0.12$, $p = 0.036$) είχαν αρνητικούς συντελεστές, υποδεικνύοντας ότι η αύξησή τους μειώνει την πιθανότητα παρουσίας πεζού εντός διάβασης.

Εντυπωσιακά, η μεταβλητή **ResolvedTrafficLightStatus** παρουσίασε τον μεγαλύτερο θετικό συντελεστή ($\beta = 1.87$, $p < 0.001$), γεγονός που υποδηλώνει ισχυρή συσχέτιση με την πιθανότητα παρουσίας πεζού εντός διάβασης, πιθανώς λόγω της σχέσης μεταξύ φωτεινής σηματοδότησης και οργανωμένης διάβασης.

Τα αποτελέσματα παρουσιάζονται στον επισυναπτόμενο πίνακα:

Variable	Coef.	Std.Err.	z	P> z	[0.025	0.975]
const	-2.518	0.069	-36.374	1.1E-289	-2.653	-2.382
MinTTC_wins_log	0.961	0.037	26.050	1.4E-149	0.889	1.033
vy	0.384	0.024	15.681	2.05E-55	0.336	0.432
GroundPlaneCentroid_1_veh	0.076	0.036	2.099	0.035763	0.005	0.148
magnitude_veh	-0.238	0.028	-8.420	3.76E-17	-0.293	-0.183
VehicleStatus	0.371	0.048	7.639	2.19E-14	0.276	0.466
PedestrianStatus	-1.969	0.066	-29.778	7.5E-195	-2.099	-1.840
ResolvedTrafficLightStatus	1.867	0.046	40.502	0	1.776	1.957
IsInCrossingArea_veh	-0.117	0.056	-2.098	0.035934	-0.225	-0.008

Πίνακας 5.6: Εκτιμώμενοι συντελεστές λογιστικού μοντέλου

Απόδοση του μοντέλου

Για την αξιολόγηση της απόδοσης του μοντέλου, χρησιμοποιήθηκαν η μήτρα σύγχυσης, οι βασικές μετρικές ταξινόμησης (ακρίβεια, ανάκληση, F1-score), καθώς και η καμπύλη ROC. Πριν από την εκπαίδευση του μοντέλου, εφαρμόστηκε η μέθοδος SMOTE (Synthetic Minority Oversampling Technique), προκειμένου να αντιμετωπιστεί η αρχική ανισορροπία μεταξύ των δύο κατηγοριών της εξαρτημένης μεταβλητής.

Η μήτρα σύγχυσης δείχνει συνολική ακρίβεια **92,96%**, με **2892** σωστές ταξινομήσεις για την κατηγορία “εκτός διάβασης” (0) και **289 σωστές ταξινομήσεις** για την κατηγορία “εντός διάβασης” (1). Οι ψευδώς αρνητικές περιπτώσεις (δηλαδή οι περιπτώσεις εντός διάβασης που ταξινομήθηκαν λανθασμένα ως εκτός) ήταν **78**, ενώ οι ψευδώς θετικές ήταν **163**.

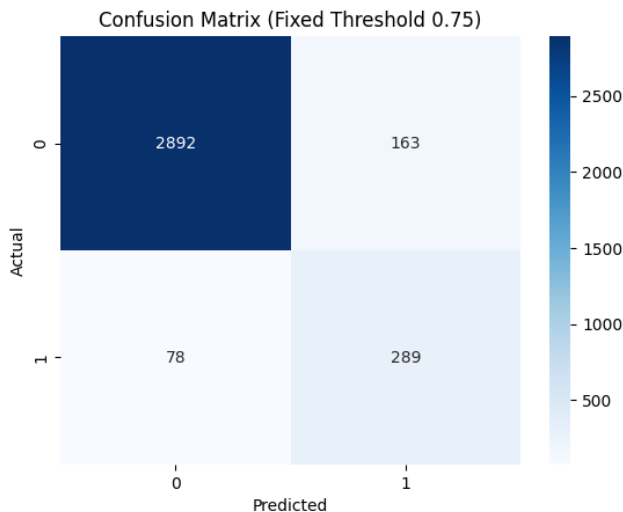
Αναλύοντας τις επιμέρους μετρικές, η κατηγορία **0** εμφάνισε υψηλές επιδόσεις (ακρίβεια: **0,97**, ανάκληση: **0,95**, F1-score: **0,96**), ενώ η κατηγορία **1** παρουσίασε πιο μέτριες τιμές (ακρίβεια: **0,64**, ανάκληση: **0,79**, F1-score: **0,71**), εν μέρει λόγω της αρχικής ανισορροπίας στο πλήθος των παρατηρήσεων, παρότι εφαρμόστηκε oversampling. Ο μέσος F1-score (macro average) ήταν **0,83**, ενώ ο σταθμισμένος μέσος όρος (weighted average) ανήλθε σε **0,93**, ενισχύοντας τη γενική εικόνα καλής απόδοσης του μοντέλου.

Η καμπύλη **ROC** επιβεβαιώνει την υψηλή διακριτική ικανότητα του μοντέλου, με **AUC = 0,96**, ένδειξη ότι το μοντέλο διαχωρίζει αποτελεσματικά τις δύο κατηγορίες.

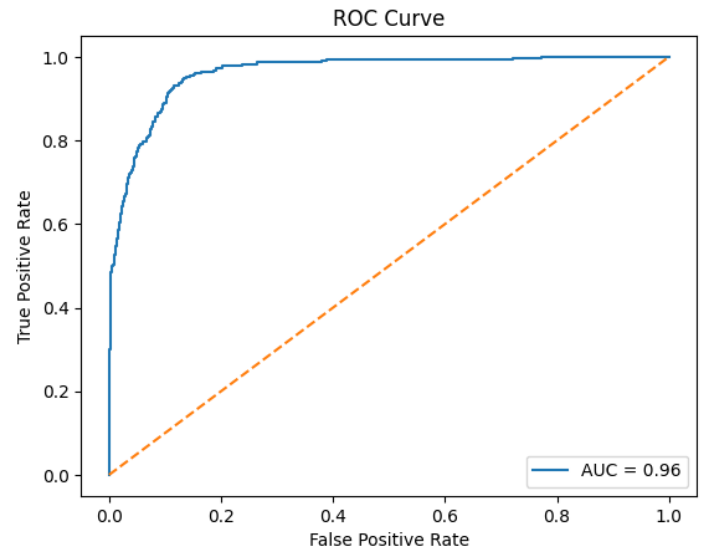
Παρακάτω παρουσιάζονται τα σχετικά αποτελέσματα:

	precision	recall	f1-score	support
0	0.973	0.946	0.960	3055
1	0.639	0.787	0.706	367
accuracy	0.929	0.929	0.929	0.929
macro avg	0.806	0.867	0.833	3422
weighted avg	0.938	0.929	0.933	3422

Πίνακας 5.7: Αναφορά ταξινόμησης λογιστικού μοντέλου



Πίνακας 5.8: Πίνακας σύγχυσης λογιστικού μοντέλου



Σχήμα 5.4: Καμπύλη ROC λογιστικού μοντέλου

Τελική αξιολόγηση του μοντέλου και δείκτες AIC&BIC

Συνολικά, το λογιστικό μοντέλο που κατασκευάστηκε εμφανίζει **υψηλή στατιστική ποιότητα και αξιοπιστία**. Όλες οι ανεξάρτητες μεταβλητές είναι στατιστικά σημαντικές, χωρίς παρουσία πολυσυγγραμμικότητας, και παρουσιάζουν ερμηνεύσιμη σχέση με την εξαρτημένη μεταβλητή. Το μοντέλο επιτυγχάνει πολύ καλή απόδοση, τόσο από πλευράς ακρίβειας όσο και διακριτικής ικανότητας (**AUC = 0,96**), γεγονός που ενισχύει την εμπιστοσύνη στην πρακτική του χρησιμότητα.

Αν και τα στατιστικά κριτήρια **AIC (7084,33)** και **BIC (7152,43)** δεν αποτέλεσαν τον βασικό άξονα επιλογής του μοντέλου, παρέχουν μια επιπλέον ένδειξη καλής προσαρμογής, με σχετικά χαμηλές τιμές σε σχέση με την πολυπλοκότητα του μοντέλου και τον αριθμό των παρατηρήσεων. Συνεπώς, το μοντέλο κρίνεται κατάλληλο για την ερμηνεία και πρόβλεψη της παρουσίας πεζού εντός ή εκτός διάβασης, βάσει των υπό εξέταση χαρακτηριστικών.

5.8.3 Μοντέλο 2: Εξαρτημένη μεταβλητή – Νόμιμη/Παράνομη διέλευση πεζού

Ποιότητα προσαρμογής μοντέλου

Η συνολική προσαρμογή του μοντέλου αξιολογήθηκε μέσω του log-likelihood και του likelihood ratio test (LLR). Το μοντέλο εμφάνισε τιμή log-likelihood ίση με **-979.91**, η οποία είναι σημαντικά βελτιωμένη σε σχέση με την τιμή του null model (**-10561.00**), που περιλαμβάνει μόνο τη σταθερά. Η τιμή του **LLR p-value** ήταν **< 0.001**, γεγονός που υποδηλώνει ότι το μοντέλο με τις ανεξάρτητες μεταβλητές παρουσιάζει στατιστικά σημαντικά καλύτερη προσαρμογή από το μοντέλο χωρίς εξηγητικούς παράγοντες. Επιπλέον, το **Pseudo R² (McFadden)** του μοντέλου ήταν **0.907**, τιμή ιδιαίτερα υψηλή για λογιστική παλινδρόμηση, υποδεικνύοντας ότι το μοντέλο εξηγεί σε εξαιρετικά μεγάλο βαθμό τη διακύμανση της εξαρτημένης μεταβλητής.

Τα σχετικά αποτελέσματα παρουσιάζονται στον επισυναπτόμενο πίνακα:

0	1	2	3
Model:	Logit	Method:	MLE
Dependent Variable:	PedestrianStatus_bin	Pseudo R-squared:	0.907
Date:	2025-07-04 10:56	AIC:	1977.8277
No. Observations:	15236	BIC:	2046.5105
Df Model:	8	Log-Likelihood:	-979.91
Df Residuals:	15227	LL-Null:	-10561.
Converged:	1.0000	LLR p-value:	0.0000
No. Iterations:	11.0000	Scale:	1.0000

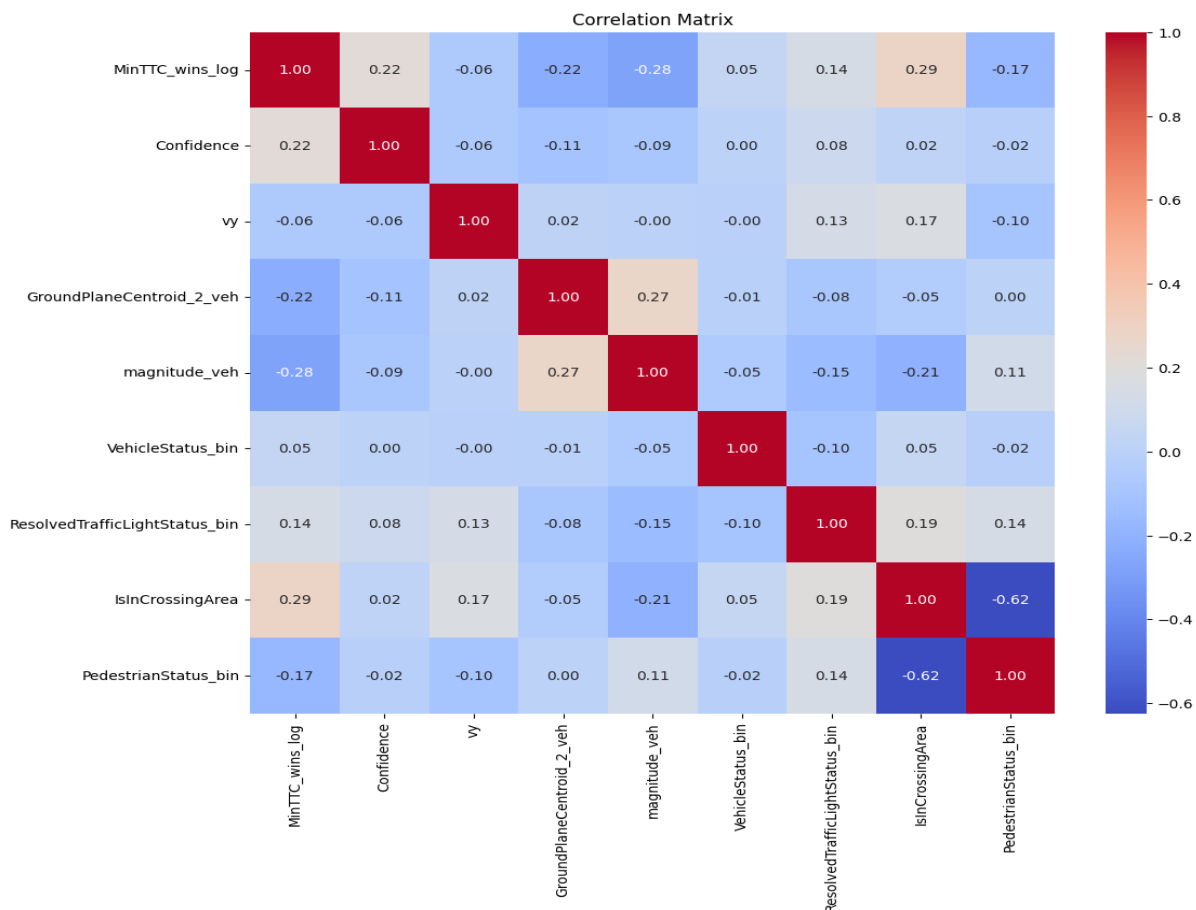
Πίνακας 5.9: Στατιστικά χαρακτηριστικά λογιστικού μοντέλου

Έλεγχος πολυσυγγραμμικότητας

Ο έλεγχος πολυσυγγραμμικότητας μεταξύ των ανεξάρτητων μεταβλητών πραγματοποιήθηκε μέσω του **πίνακα συσχετίσεων** (Pearson) και των δεικτών **VIF** (Variance Inflation Factor). Καμία μεταβλητή δεν παρουσίασε ισχυρή συσχέτιση με άλλη (όλες οι τιμές $r < 0.3$), γεγονός που υποδεικνύει ότι δεν υπάρχει σημαντική αλληλεξάρτηση μεταξύ των μεταβλητών του μοντέλου.

Επιπλέον, οι τιμές του δείκτη VIF για όλες τις ανεξάρτητες μεταβλητές κυμάνθηκαν μεταξύ 1.02 και 1.25, με τη μεγαλύτερη να αφορά τη μεταβλητή MinTTC_wins_log (VIF = 1.25). Οι τιμές αυτές βρίσκονται πολύ κάτω από το συνήθως αποδεκτό όριο των 5, γεγονός που ενισχύει την ένδειξη απουσίας πολυσυγγραμμικότητας. Η **χαμηλή πολυσυγγραμμικότητα** συμβάλλει στη σταθερότητα των συντελεστών και στη συνολική αξιοπιστία των εκτιμήσεων του μοντέλου.

Τα σχετικά αποτελέσματα παρουσιάζονται παρακάτω:



Πίνακας 5.10: Πίνακας συσχέτισης λογιστικού μοντέλου

Variable	VIF
const	1
MinTTC_wins_log	1.25
Confidence	1.06
vy	1.06
GroundPlaneCentroid_2_veh	1.12
magnitude_veh	1.18
VehicleStatus	1.02
ResolvedTrafficLightStatus	1.09
IsInCrossingArea	1.18

Πίνακας 5.11: Τιμές VIF λογιστικού μοντέλου

Συντελεστές και στατιστική σημαντικότητα

Όλες οι ανεξάρτητες μεταβλητές βρέθηκαν στατιστικά σημαντικές ($p < 0.05$), υποδεικνύοντας ουσιαστική συμβολή στην πρόβλεψη της νόμιμης διέλευσης πεζού. Θετικούς συντελεστές παρουσίασαν οι μεταβλητές **Confidence** ($\beta = 0.84$), **magnitude_veh** ($\beta = 0.80$) και ιδιαίτερα η **ResolvedTrafficLightStatus** ($\beta = 4.58$), υποδεικνύοντας ότι υψηλότερες τιμές σχετίζονται με αυξημένη πιθανότητα νόμιμης διέλευσης.

Αρνητικοί συντελεστές παρατηρήθηκαν για τις **MinTTC_wins_log** ($\beta = -0.21$), **vy** ($\beta = -0.38$), **GroundPlaneCentroid_2_veh** ($\beta = -0.45$), **VehicleStatus** ($\beta = -0.29$) και **IsInCrossingArea** ($\beta = -3.04$). Ειδικά για την τελευταία, η αρνητική σχέση ερμηνεύεται με βάση τον ορισμό της παραβατικότητας για το πλαίσιο της παρούσας διπλωματικής εργασίας: παραβιάσεις καταγράφονται όταν ο πεζός βρίσκεται εντός διάβασης και η κατάσταση του φωτεινού σηματοδότη είναι ερυθρή. Επομένως, ο υψηλά αρνητικός συντελεστής της μεταβλητής καταδεικνύει ότι οι παράνομοι πεζοί εισέρχονται ενεργά και συνειδητά στη διάβαση.

Τα αποτελέσματα συνοψίζονται στον πίνακα που ακολουθεί:

Variable	Coef.	Std.Err.	z	P> z	[0.025	0.975]
const	5.726	0.144	39.606	0	5.442	6.009
MinTTC_wins_log	-0.213	0.072	-2.965	0.003	-0.353	-0.072
Confidence	0.841	0.09	9.347	9.04E-21	0.665	1.018
vy	-0.379	0.044	-8.590	8.69E-18	-0.466	-0.293
GroundPlaneCentroid_2_veh	-0.445	0.048	-9.214	3.15E-20	-0.540	-0.351
magnitude_veh	0.798	0.091	8.713	2.95E-18	0.618	0.977
VehicleStatus	-0.292	0.128	-2.272	0.023071	-0.544	-0.040
ResolvedTrafficLightStatus	4.581	0.160	28.568	1.7E-179	4.266	4.895
IsInCrossingArea	-3.039	0.091	-33.383	2.4E-244	-3.217	-2.860

Πίνακας 5.12: Εκτιμώμενοι συντελεστές λογιστικού μοντέλου

Απόδοση του μοντέλου

Για την αξιολόγηση της απόδοσης του μοντέλου, χρησιμοποιήθηκαν η μήτρα σύγχυσης, οι βασικές μετρικές ταξινόμησης (ακρίβεια, ανάκληση, F1-score), καθώς και η καμπύλη ROC. Πριν από την εκπαίδευση του μοντέλου, εφαρμόστηκε η μέθοδος SMOTE (Synthetic Minority Oversampling Technique), προκειμένου να αντιμετωπιστεί η αρχική ανισορροπία μεταξύ των δύο κατηγοριών της εξαρτημένης μεταβλητής.

Η μήτρα σύγχυσης δείχνει συνολική ακρίβεια **99,53%**, με **171** σωστές ταξινομήσεις για την κατηγορία “παράνομη διέλευση” (0) και **3235** σωστές ταξινομήσεις για την κατηγορία “νόμιμη διέλευση” (1). Οι ψευδώς αρνητικές περιπτώσεις (δηλαδή οι περιπτώσεις νόμιμης διέλευσης που ταξινομήθηκαν λανθασμένα ως παράνομες) ήταν **7**, ενώ οι ψευδώς θετικές ήταν **9**.

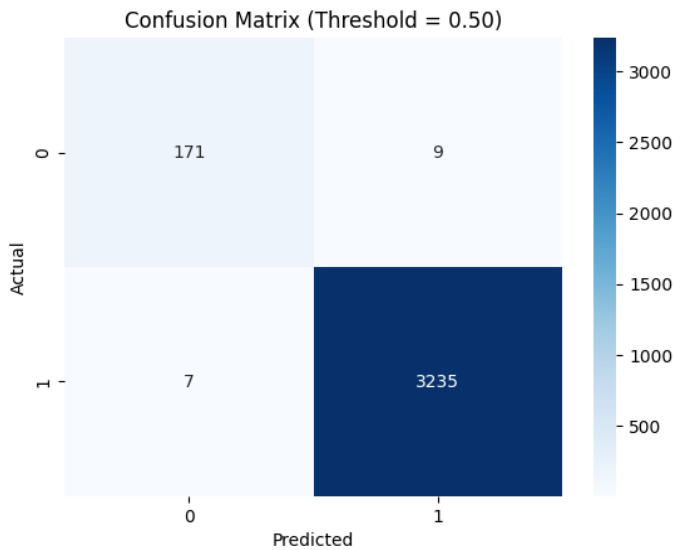
Αναλύοντας τις επιμέρους μετρικές, η κατηγορία 0 εμφάνισε πολύ καλές επιδόσεις (ακρίβεια: **0,96**, ανάκληση: **0,95**, F1-score: **0,96**), ενώ η κατηγορία 1 παρουσίασε εξαιρετικά υψηλές τιμές (ακρίβεια: **0,997**, ανάκληση: **0,998**, F1-score: **0,998**). Ο μέσος F1-score (macro average) ήταν 0,98, ενώ ο σταθμισμένος μέσος όρος (weighted average) ανήλθε σε 0,995, ενισχύοντας την εικόνα εξαιρετικής απόδοσης του μοντέλου.

Η καμπύλη ROC επιβεβαιώνει την πολύ υψηλή διακριτική ικανότητα του μοντέλου, με **AUC = 0,99**, ένδειξη ότι το μοντέλο διαχωρίζει αποτελεσματικά τις δύο κατηγορίες.

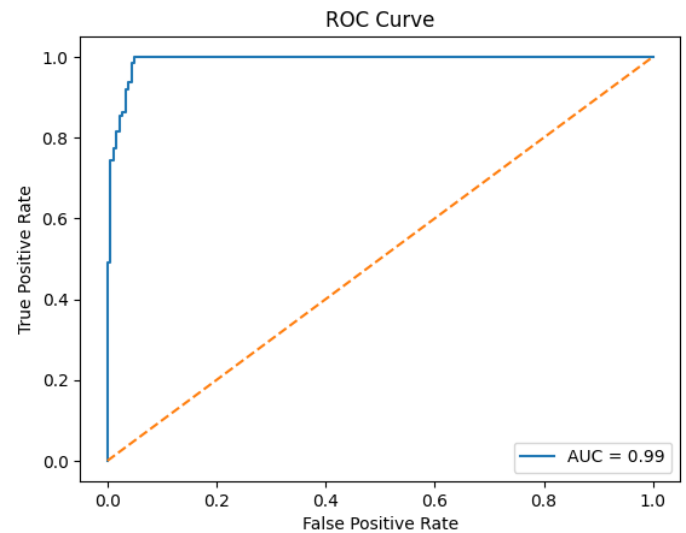
Παρακάτω παρουσιάζονται και τα σχετικά αποτελέσματα:

	precision	recall	f1-score	support
0	0.960	0.95	0.955	180
1	0.997	0.997	0.997	3242
accuracy	0.995	0.995	0.995	0.995
macro avg	0.978	0.973	0.976	3422
weighted avg	0.995	0.995	0.995	3422

Πίνακας 5.13: Αναφορά ταξινόμησης λογιστικού μοντέλου



Πίνακας 5.14: Πίνακας σύγκρισης λογιστικού μοντέλου



Σχήμα 5.5: Καμπύλη ROC λογιστικού μοντέλου

Τελική αξιολόγηση του μοντέλου και δείκτες AIC & BIC

Συνολικά, το λογιστικό μοντέλο που κατασκευάστηκε εμφανίζει υψηλή στατιστική ποιότητα και αξιοπιστία. Όλες οι ανεξάρτητες μεταβλητές είναι στατιστικά σημαντικές, απουσιάζει πολυσυγγραμμικότητα και οι συντελεστές παρουσιάζουν ερμηνεύσιμη σχέση με την πιθανότητα νόμιμης διέλευσης πεζού. Το μοντέλο επιτυγχάνει πολύ υψηλή απόδοση, τόσο ως προς την ακρίβεια όσο και ως προς τη διακριτική του ικανότητα (**AUC = 0,99**), γεγονός που ενισχύει την εμπιστοσύνη στη χρηστικότητα του στην πράξη.

Παρότι οι δείκτες **AIC (1977,83)** και **BIC (2046,51)** δεν αποτέλεσαν βασικό κριτήριο επιλογής, προσφέρουν πρόσθετη ένδειξη καλής προσαρμογής, διατηρώντας χαμηλές τιμές σε σχέση με την πολυπλοκότητα του μοντέλου και τον συνολικό αριθμό παρατηρήσεων. Συνεπώς, το μοντέλο κρίνεται κατάλληλο για την πρόβλεψη και ερμηνεία της νομιμότητας ή παρανομίας της διέλευσης πεζού, με βάση τα διαθέσιμα χαρακτηριστικά.

5.9 Τυχαία Δάση

5.9.1 Εισαγωγή

Η μέθοδος **Random Forest** εφαρμόστηκε με σκοπό την πρόβλεψη των μεταβλητών **IsInCrossingArea** και **PedestrianStatus**, οι οποίες δηλώνουν, αντίστοιχα, αν ένας πεζός βρίσκεται εντός της περιοχής διάβασης και αν κινείται νόμιμα ή παράνομα. Η Random Forest αποτελεί αλγόριθμο εποπτευόμενης μάθησης που βασίζεται σε πλήθος δέντρων απόφασης, τα οποία συνδυάζονται ώστε να παράγουν ένα πιο σταθερό και ισχυρό μοντέλο.

Για την αξιολόγηση της απόδοσης, χρησιμοποιήθηκαν βασικοί δείκτες ταξινόμησης, όπως το **confusion matrix**, η **αναφορά ταξινόμησης** (classification report), η οποία περιλαμβάνει τις μετρικές **ακρίβεια** (precision), **ανάκληση** (recall) και **F1-score**, καθώς και η **καμπύλη ROC** με την αντίστοιχη τιμή **AUC**.

Επιπλέον, αξιοποιήθηκε η **σημαντικότητα των μεταβλητών** (feature importances), μέσω της οποίας προσδιορίστηκε η σχετική επιρροή κάθε μεταβλητής στη διαδικασία λήψης απόφασης του μοντέλου. Μέσα από αυτή τη διαδικασία αναδείχθηκαν τα πιο αξιόπιστα και προβλεπτικά ισχυρά μοντέλα για την εκτίμηση της πιθανότητας εμφάνισης των φαινομένων που περιγράφουν οι μεταβλητές **IsInCrossingArea** και **PedestrianStatus**.

5.9.2 Μοντέλο 1: Εξαρτημένη μεταβλητή – Παρουσία πεζού εντός/εκτός διάβασης

Ανάλυση σημαντικότητας μεταβλητών

Η ανάλυση σημαντικότητας μεταβλητών του μοντέλου Random Forest, το οποίο στόχευε στην πρόβλεψη της παρουσίας πεζού εντός ή εκτός διάβασης, ανέδειξε ως σημαντικότερη μεταβλητή τη **θέση του πεζού στον άξονα βάθους του οπτικού πεδίου** (GroundPlaneCentroid_2), δηλαδή την απόστασή του από την κάμερα κατά τη διάρκεια της κίνησης. Η ισχυρή σημασία αυτής της παραμέτρου υποδηλώνει ότι ο χωρικός εντοπισμός του πεζού σε σχέση με το σύστημα αντίληψης παίζει καθοριστικό ρόλο για την κατηγοριοποίηση της σκηνής.

Ακολούθησε σε σημασία η συνολική **ταχύτητα του πεζού** (magnitude), η οποία διαφοροποιεί με σαφήνεια τις περιπτώσεις, ενδεχομένως αντανakλώντας διαφορές στον ρυθμό κίνησης ανάλογα με τη χωρική τοποθέτηση ή τη συμπεριφορά.

Η μεταβλητή **PedestrianStatus**, η οποία αποτυπώνει τη **νομιμότητα ή παραβατικότητα της διάσχισης**, εμφάνισε επίσης υψηλή σημασία. Ωστόσο, η ερμηνεία της σχετίζεται στενά με τη μεθοδολογία της μελέτης: η παραβατικότητα ορίζεται αποκλειστικά ως παραβίαση σηματοδότη και καταγράφεται μόνο εντός σηματοδοτούμενων διαβάσεων. Συνεπώς, ένας πεζός χαρακτηρίζεται ως παραβάτης μόνον όταν εισέρχεται σε διάβαση με απαγορευτικό σήμα, κάτι που αντανakλά μια ενσυνείδητη και δομημένη επιλογή. Αντίθετα, η παρουσία εκτός διάβασης δεν αξιολογείται ως παραβατική, ακόμη κι αν μπορεί να υποδηλώνει άτυπη ή μη κανονιστική συμπεριφορά. Η υψηλή σημασία της μεταβλητής ενδέχεται επομένως να

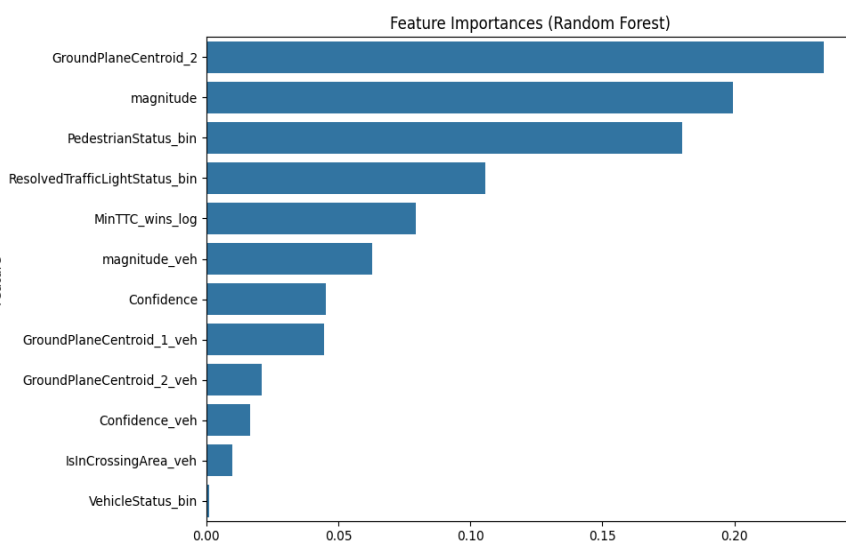
αντανακλά τόσο τη συμπεριφορική διαφοροποίηση, όσο και τη μεθοδολογική φύση της σήμανσης.

Η **κατάσταση του φωτεινού σηματοδότη** για πεζούς (ResolvedTrafficLightStatus) κατετάγη επίσης μεταξύ των πλέον κρίσιμων παραγόντων, επιβεβαιώνοντας τον καθοριστικό ρόλο της κυκλοφοριακής ρύθμισης στην κατανόηση της σκηνής.

Μεταβλητές όπως ο **ελάχιστος χρόνος έως πιθανή σύγκρουση** (MinTTC_wins_log), η **ταχύτητα του οχήματος** (magnitude_veh) και η **βεβαιότητα του αλγορίθμου για τον εντοπισμό πεζού** (Confidence) παρουσίασαν μέτρια σημασία, αναδεικνύοντας τη συμβολή τόσο της σκηνικής δυναμικής όσο και της ποιότητας ανίχνευσης στην τελική πρόβλεψη.

Αντίθετα, μεταβλητές που σχετίζονται με το **όχημα**, όπως η **χωρική του θέση** (GroundPlaneCentroid_1,2_veh), η **βεβαιότητα εντοπισμού** (Confidence_veh), η **παρουσία του εντός διάβασης** (IsInCrossingArea_veh) και η **προτεραιότητά του** (VehicleStatus), εμφάνισαν χαμηλή ή αμελητέα σημασία. Το εύρημα αυτό υποδηλώνει ότι, για το συγκεκριμένο μοντέλο, η πληροφορία που αφορά το όχημα είναι δευτερεύουσας σημασίας σε σχέση με τις παραμέτρους που περιγράφουν άμεσα τη συμπεριφορά και τη θέση του πεζού.

Τα παραπάνω συνοψίζονται στο διάγραμμα σημαντικότητας που παρατίθεται και στον αντίστοιχο πίνακα:



Σχήμα 5.6: Ραβδόγραμμα σημασίας μεταβλητών

Feature	Importance
GroundPlaneCentroid_2	0.234
magnitude	0.2
PedestrianStatus	0.181
ResolvedTrafficLightStatus	0.106
MinTTC_wins_log	0.08
magnitude_veh	0.063
Confidence	0.05
GroundPlaneCentroid_1_veh	0.04
GroundPlaneCentroid_2_veh	0.02
Confidence_veh	0.017
IsInCrossingArea_veh	0.01
VehicleStatus	0.01

Πίνακας 5.15: Τιμές σημασίας μεταβλητών

Αξιολόγηση επίδοσης μοντέλου

Το μοντέλο τυχαίου δάσους παρουσίασε υψηλή συνολική ακρίβεια της τάξης του **97,2%**, ενώ η τιμή του **AUC** ανήλθε σε **0,99**, αποτυπώνοντας εξαιρετική διακριτική ικανότητα μεταξύ των δύο κατηγοριών της εξαρτημένης μεταβλητής (παρουσία πεζού εντός ή εκτός διάβασης).

Η μήτρα σύγχυσης δείχνει χαμηλό αριθμό σφαλμάτων ταξινόμησης, με μόλις **96** λανθασμένες προβλέψεις σε σύνολο **3.422** παρατηρήσεων. Από αυτές, **53** περιπτώσεις εντός διάβασης ταξινομήθηκαν λανθασμένα ως εκτός, ενώ **43** περιπτώσεις εκτός διάβασης ταξινομήθηκαν ως εντός.

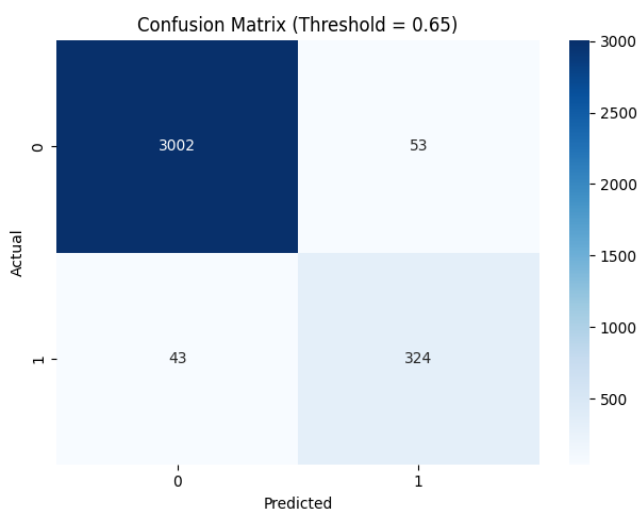
Η κατηγορία “εκτός διάβασης” (0) παρουσίασε πολύ υψηλές επιδόσεις (precision: **0.99**, recall: **0.98**, F1-score: **0.98**), ενώ η κατηγορία “εντός διάβασης” (1) εμφάνισε ελαφρώς χαμηλότερα αλλά ικανοποιητικά αποτελέσματα (precision: **0.86**, recall: **0.88**, F1-score: **0.87**). Οι μέσοι δείκτες macro και weighted average ξεπέρασαν το **0.92**, δείχνοντας ισχυρή γενίκευση του μοντέλου και στις δύο κατηγορίες.

Η εφαρμογή της μεθόδου **SMOTE** συνέβαλε στην εξισορρόπηση του συνόλου δεδομένων και βελτίωσε τις επιδόσεις στην προβλεπτικά ασθενέστερη κατηγορία (εντός διάβασης). Επιπλέον, η υψηλή διακριτική δύναμη συγκεκριμένων μεταβλητών (όπως η χωρική θέση του πεζού, η συνολική ταχύτητα και ο φωτεινός σηματοδότης) φαίνεται να ενίσχυσε σημαντικά την αποτελεσματικότητα του μοντέλου.

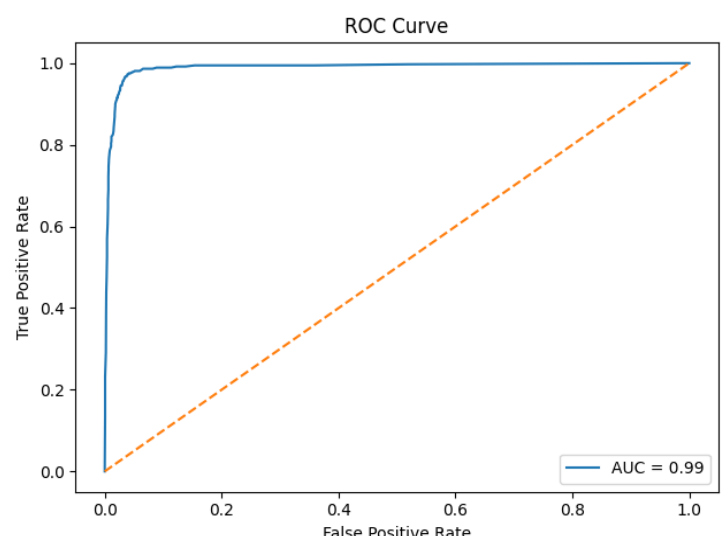
Τα αποτελέσματα συνοψίζονται παρακάτω:

	precision	recall	f1-score	support
0	0.986	0.983	0.984	3055
1	0.859	0.883	0.871	367
accuracy	0.972	0.972	0.972	0.972
macro avg	0.923	0.933	0.928	3422
weighted avg	0.972	0.972	0.972	3422

Πίνακας 5.16: Αναφορά ταξινόμησης μοντέλου μηχανικής μάθησης



Πίνακας 5.17: Πίνακας σύγχυσης μοντέλου μηχανικής μάθησης



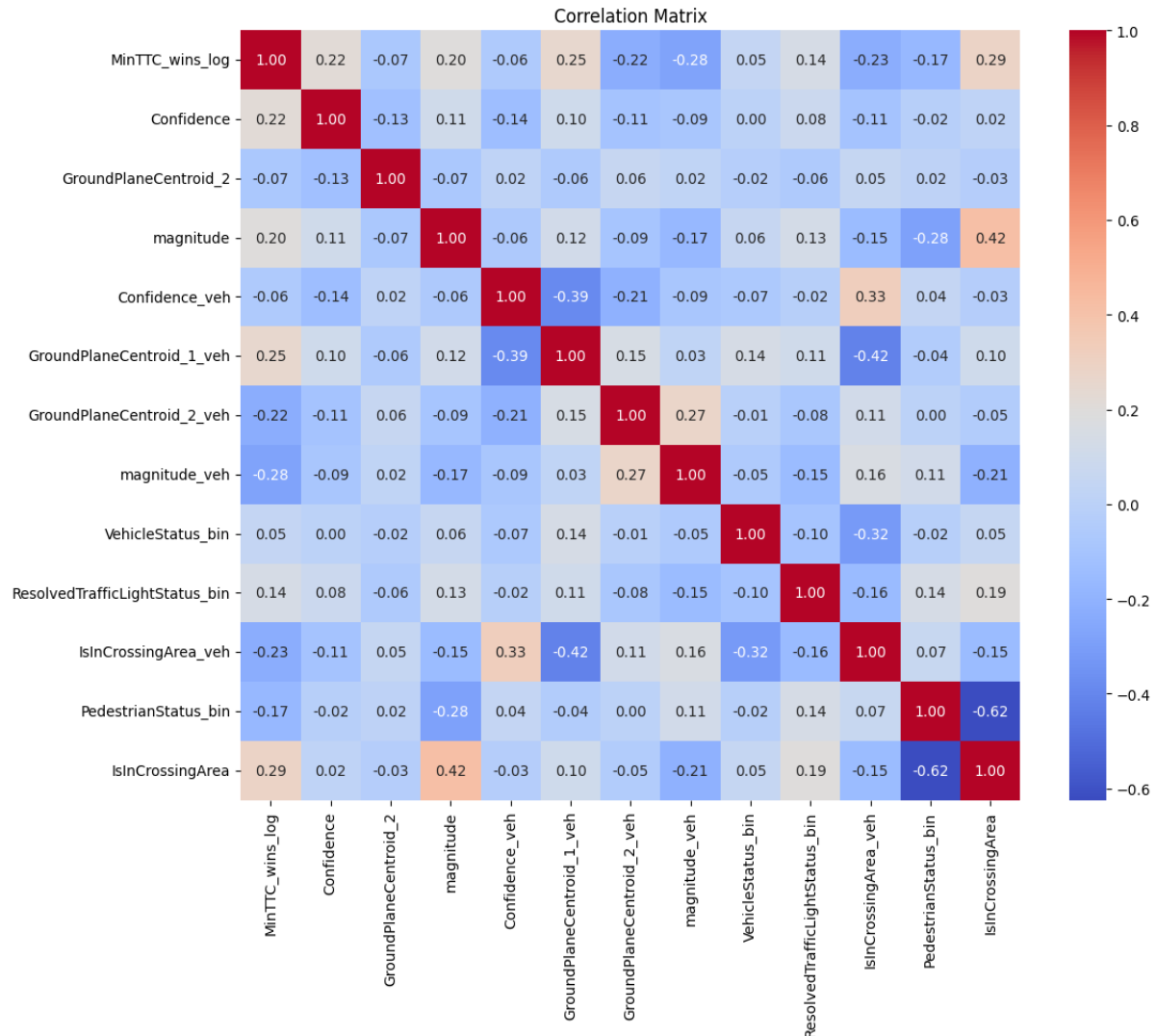
Σχήμα 5.7: Καμπύλη ROC μοντέλου μηχανικής μάθησης

Έλεγχος πολυσυγγραμμικότητας και συσχετίσεων

Ο έλεγχος πολυσυγγραμμικότητας επιβεβαίωσε την απουσία προβληματικών αλληλεξαρτήσεων μεταξύ των ανεξάρτητων μεταβλητών. Όλες οι τιμές VIF ήταν κάτω από το όριο του 1.6, με τη μέγιστη να ανέρχεται σε 1.53 (IsInCrossingArea_veh), γεγονός που υποδεικνύει περιορισμένη επικάλυψη πληροφορίας και ενισχύει τη

σταθερότητα του μοντέλου. Παράλληλα, καμία συσχέτιση μεταξύ μεταβλητών δεν υπερέβη το όριο του $r = 0.7$, στοιχείο που επιβεβαιώνει την ανεξαρτησία τους σε αποδεκτό βαθμό και την εγκυρότητα της συμβολής τους στην ταξινόμηση.

Τα σχετικά αποτελέσματα συνοψίζονται παρακάτω:



Πίνακας 5.18 :Πίνακας συσχέτισης μοντέλου μηχανικής μάθησης

Feature	VIF
IsInCrossingArea_veh	1.53
GroundPlaneCentroid_1_veh	1.45
Confidence_veh	1.32
MinTTC_wins_log	1.30
GroundPlaneCentroid_2_veh	1.24
magnitude_veh	1.20
magnitude	1.17
PedestrianStatus	1.16
VehicleStatus	1.15
ResolvedTrafficLightStatus	1.14
Confidence	1.09
GroundPlaneCentroid_2	1.02

Πίνακας 5.19: Τιμές VIF μοντέλου μηχανικής μάθησης

5.9.3 Μοντέλο 2: Εξαρτημένη μεταβλητή – Νόμιμη/Παράνομη διέλευση πεζού

Ανάλυση σημαντικότητας μεταβλητών

Η ανάλυση σημαντικότητας μεταβλητών με το μοντέλο Random Forest, το οποίο στόχευε στην πρόβλεψη της νομιμότητας ή παραβατικότητας της διάσχισης πεζού, ανέδειξε ως κυρίαρχη τη μεταβλητή που αποτυπώνει την **παρουσία του πεζού εντός ή εκτός της διάβασης**. Η εξαιρετικά υψηλή σημασία αυτής της μεταβλητής υποδηλώνει ότι το ίδιο το χωρικό πλαίσιο στο οποίο πραγματοποιείται η διάσχιση αποτελεί το βασικότερο στοιχείο διαφοροποίησης των κατηγοριών. Το εύρημα συνδέεται άμεσα με τη μεθοδολογική οριοθέτηση της παραβατικότητας, η οποία ορίζεται αποκλειστικά εντός σηματοδοτούμενων διαβάσεων.

Σημαντική ήταν και η συμβολή της **ταχύτητας** του πεζού, γεγονός που δείχνει ότι ο τρόπος κίνησης του πεζού διαφοροποιεί την πρόβλεψη ως προς τη νομιμότητα ή μη της διάσχισης.

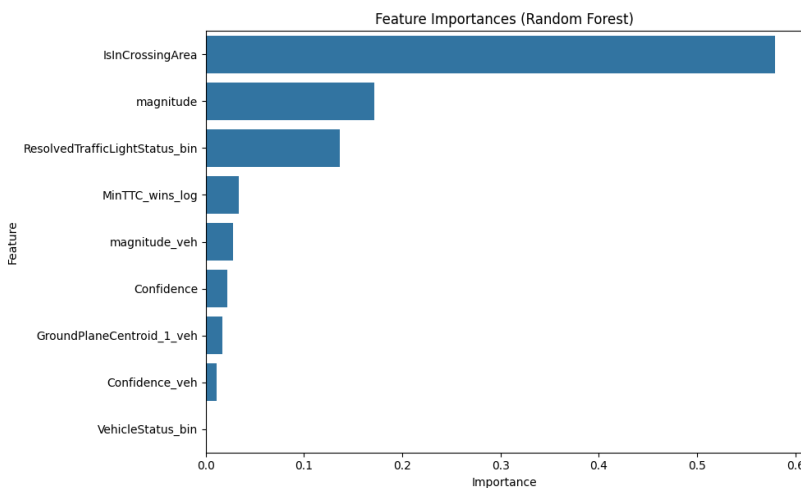
Αντίστοιχα, η κατάσταση του **φωτεινού σηματοδότη** για πεζούς καταγράφηκε ως μεταβλητή υψηλής σημασίας, γεγονός που επιβεβαιώνει τη λειτουργία του σηματοδότη ως κρίσιμου εξωτερικού παράγοντα για την πρόβλεψη της κατηγορίας. Ακολουθούν μεταβλητές όπως ο **ελάχιστος χρόνος έως πιθανή σύγκρουση (MinTTC)** και η **ταχύτητα του οχήματος**, οι οποίες παρότι μικρότερης σχετικής σημασίας, εξακολουθούν να συνεισφέρουν στο διαχωρισμό των περιπτώσεων βάσει του συνδυασμού των χαρακτηριστικών.

Αντίθετα, μεταβλητές όπως η **εμπιστοσύνη** του συστήματος για την ανίχνευση του πεζού ή του οχήματος (confidence) και η **σχετική θέση του οχήματος** ως προς τη διάβαση είχαν περιορισμένη συνεισφορά, γεγονός που υποδηλώνει ότι πληροφορίες

σχετικές με τη σταθερότητα εντοπισμού αντικειμένων ή την εκτίμηση της κίνησης ενδέχεται να είναι λιγότερο διακριτικές για τη συγκεκριμένη ταξινόμηση.

Τέλος, η **κατάσταση του οχήματος** ως προς την προτεραιότητα (VehicleStatus) εμφανίστηκε σχεδόν μηδενικής σημασίας. Αυτό πιθανώς οφείλεται στο ότι η πληροφορία αυτή εμπεριέχεται ήδη έμμεσα μέσω της μεταβλητής του φωτεινού σηματοδότη, καθιστώντας τη πλεονάζουσα στο πλαίσιο του μοντέλου.

Τα προαναφερθέντα αποτελέσματα επισυνάπτονται παρακάτω:



Σχήμα 5.8: Ραβδόγραμμα σημασίας μεταβλητών

Feature	Importance
IsInCrossingArea	0.579
magnitude	0.171
ResolvedTrafficLightStatus	0.136
MinTTC_wins_log	0.034
magnitude_veh	0.028
Confidence	0.021
GroundPlaneCentroid_1_veh	0.017
Confidence_veh	0.011
VehicleStatus	0.0002

Πίνακας 5.20: Τιμές σημασίας μεταβλητών

Αξιολόγηση επίδοσης μοντέλου

Το μοντέλο τυχαίου δάσους παρουσίασε εξαιρετικά υψηλή διακριτική ικανότητα, γεγονός που αποτυπώνεται στη συνολική **ακρίβεια** της τάξης του **99,5%**. Η ισχυρή αυτή επίδοση επιβεβαιώνεται και από την υψηλή τιμή του **AUC (0.99)**, που δείχνει εξαιρετική ικανότητα διάκρισης μεταξύ των δύο κατηγοριών — δηλαδή, των παραβατικών και των νόμιμων πεζών. Η μήτρα σύγχυσης αποτυπώνει την πολύ χαμηλή συχνότητα σφαλμάτων, καθώς το μοντέλο αναγνωρίζει σωστά την πλειοψηφία των περιπτώσεων, με μόλις **15 συνολικά λανθασμένες ταξινομήσεις** (false positives και false negatives).

Η τάξη που αντιστοιχεί στη νόμιμη διάσχιση παρουσιάζει πολύ υψηλές τιμές precision (**0.997**) και recall (**0.998**), καταδεικνύοντας την αξιοπιστία του μοντέλου στην αναγνώριση αυτών των περιστατικών. Αντίστοιχα, ακόμα και η τάξη της παραβατικής διάσχισης, παρότι υποεκπροσωπείται στο δείγμα, παρουσιάζει ικανοποιητικές επιδόσεις με precision **0.966** και recall **0.95**. Οι δείκτες macro και weighted averages κινούνται επίσης σε πολύ υψηλά επίπεδα (**άνω του 0.97**), υποδηλώνοντας συνολικά ισχυρή γενικευσιμότητα του μοντέλου σε διαφορετικές κατανομές.

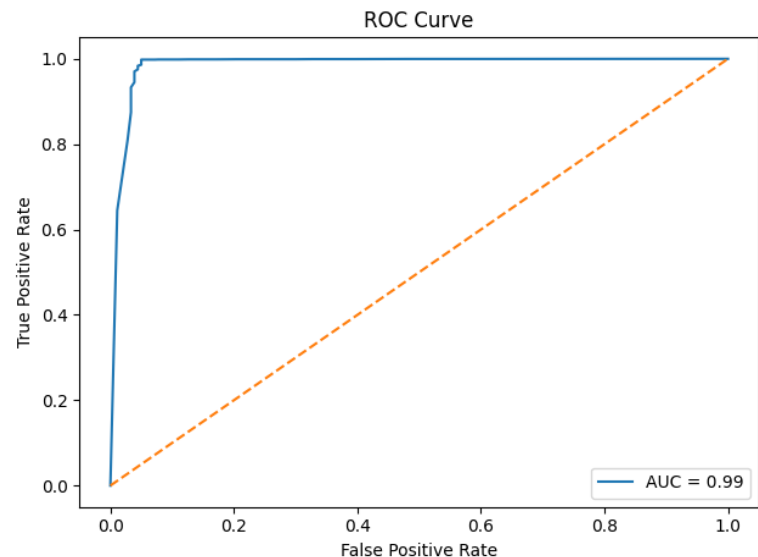
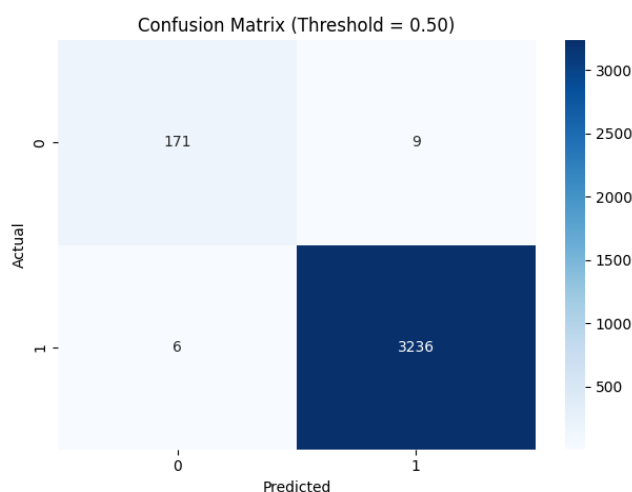
Η εξαιρετική απόδοση αυτή μπορεί εν μέρει να αποδοθεί στον πολυδιάστατο χαρακτήρα των μεταβλητών εισόδου, αλλά και στην ενίσχυση του δείγματος μέσω της τεχνικής **SMOTE**, η οποία συνέβαλε στην καλύτερη αναγνώριση της μειοψηφικής

τάξης. Επιπλέον, η ύπαρξη μεταβλητών με έντονα διακριτικό χαρακτήρα, όπως η παρουσία εντός ή εκτός διάβασης και η κατάσταση του σηματοδότη, φαίνεται να βελτιστοποιεί την απόδοση του αλγορίθμου, παρέχοντας κρίσιμες πληροφορίες για τη συμπεριφορά του πεζού στο αστικό περιβάλλον.

Τα αποτελέσματα συνοψίζονται παρακάτω:

	precision	recall	f1-score	support
0	0.966	0.95	0.958	180
1	0.997	0.998	0.998	3242
accuracy	0.996	0.996	0.996	0.996
macro avg	0.982	0.974	0.978	3422
weighted avg	0.995	0.996	0.995	3422

Πίνακας 5.21: Αναφορά ταξινόμησης μοντέλου μηχανικής μάθησης



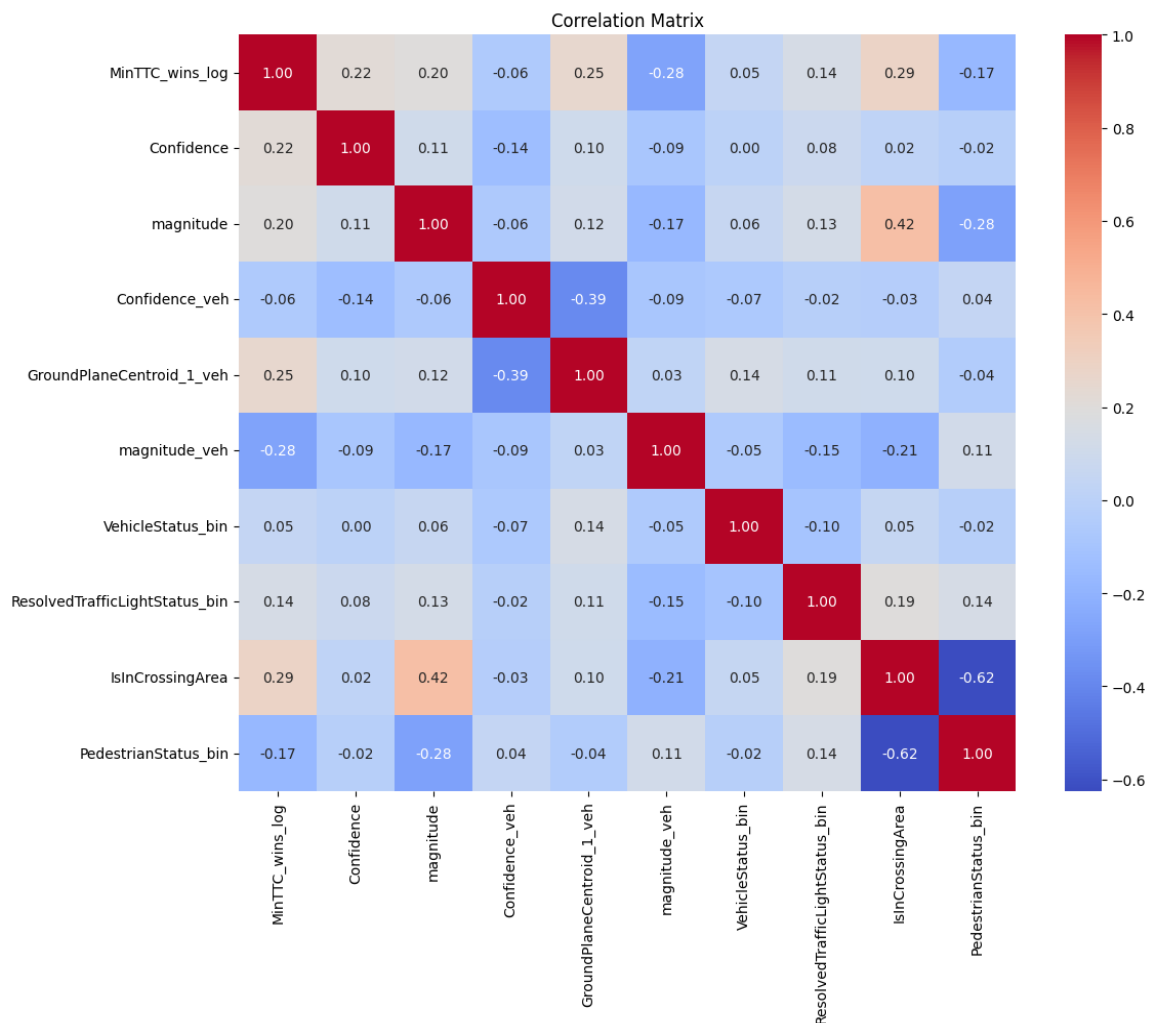
Πίνακας 5.22: Πίνακας σύγχυσης μοντέλου μηχανικής μάθησης

Σχήμα 5.9: Καμπύλη ROC μοντέλου μηχανικής μάθησης

Έλεγχος πολυσυγγραμμικότητας και συσχετίσεων

Ο έλεγχος πολυσυγγραμμικότητας έδειξε ότι όλες οι ανεξάρτητες μεταβλητές παρουσιάζουν χαμηλές τιμές **VIF (κάτω από 1.4)**, ενώ και οι μεταξύ τους συσχετίσεις παραμένουν σε αποδεκτά επίπεδα (**$r < 0.7$**), γεγονός που διασφαλίζει την αξιοπιστία των αποτελεσμάτων και την απουσία επικάλυψης πληροφορίας.

Τα σχετικά αποτελέσματα συνοψίζονται παρακάτω:



Πίνακας 5.23: Πίνακας συσχέτισης μοντέλου μηχανικής μάθησης

Feature	VIF
MinTTC_wins_log	1.281
Confidence	1.083
magnitude	1.253
Confidence_veh	1.207
GroundPlaneCentroid_1_veh	1.297
magnitude_veh	1.153
VehicleStatus	1.044
ResolvedTrafficLightStatus	1.089
IsInCrossingArea	1.329

Πίνακας 5.24: Τιμές VIF μοντέλου μηχανικής μάθησης

Κεφάλαιο 6: Συμπεράσματα

6.1 Εισαγωγή

Ο βασικός στόχος της παρούσας Διπλωματικής Εργασίας είναι η **συγκριτική αξιολόγηση** της χειροκίνητης και της αυτοματοποιημένης μεθόδου καταγραφής της συμπεριφοράς πεζών σε σηματοδοτούμενες διασταυρώσεις, με έμφαση στην ακρίβεια της **αναγνώρισης** της κατάστασης του **φωτεινού σηματοδότη** και της **ταξινόμησης διελεύσεων** ως νόμιμες ή παράνομες. Παράλληλα, στόχος αποτέλεσε η εφαρμογή κατάλληλων στατιστικών τεχνικών για τη διερεύνηση, ερμηνεία και πρόβλεψη παραβατικής συμπεριφοράς, βάσει παραμέτρων που προέκυψαν από την αυτοματοποιημένη ανάλυση βίντεο.

Τα δεδομένα που χρησιμοποιήθηκαν στην παρούσα μελέτη προήλθαν από **βιντεοσκόπηση** σηματοδοτούμενης διασταύρωσης στο κέντρο της Αθήνας (τέλος οδού Πανεπιστημίου, πλατεία Ομονοίας), κατά τη διάρκεια ώρας αιχμής. Το υλικό καταγράφηκε με σταθερό κινητό τηλέφωνο τοποθετημένο σε τρίποδο και αποτέλεσε τη βάση για δύο παράλληλες διαδικασίες ανάλυσης: μία **αυτοματοποιημένη** και μία **χειροκίνητη**.

Η αυτοματοποιημένη ανάλυση πραγματοποιήθηκε μέσω **αλγορίθμου υπολογιστικής όρασης** ανεπτυγμένου στο περιβάλλον **MATLAB**. Ο αλγόριθμος ενσωματώνει τεχνικές ανίχνευσης και παρακολούθησης αντικειμένων, καθώς και επεξεργασία της κατάστασης φωτεινών σηματοδοτών. Η έξοδος του συστήματος οργανώθηκε σε τρεις βάσεις δεδομένων (πεζών, οχημάτων και minTTC), οι οποίες στη συνέχεια ενοποιήθηκαν σε μία ενιαία βάση για τις ανάγκες της στατιστικής ανάλυσης και περιλάμβαναν πληροφορίες ταυτοποίησης αντικειμένων, χωρικών συντεταγμένων, ταχύτητας, κατάστασης φωτεινού σηματοδότη, παρουσία αντικειμένου ως προς τη διάβαση και χαρακτηρισμό της διέλευσης (νόμιμη/παράνομη).

Παράλληλα, το ίδιο υλικό αναλύθηκε με **χειροκίνητη παρατήρηση** από μέλη του Εργαστηρίου Κυκλοφοριακής Τεχνικής. Οι παρατηρήσεις καταγράφηκαν σε δομημένο φύλλο Excel, περιλαμβάνοντας πληροφορίες για την κατάσταση του φωτεινού σηματοδότη, τον χρόνο, τις νόμιμες και παράνομες διελεύσεις πεζών, καθώς και διελεύσεις κατά την ενδιάμεση φάση του σηματοδότη (Intergreen).

Η στατιστική ανάλυση των δεδομένων πραγματοποιήθηκε στο περιβάλλον **Python**, αξιοποιώντας κατάλληλες βιβλιοθήκες για επεξεργασία, συσχέτιση και μοντελοποίηση. Η ανάλυση επικεντρώθηκε σε δύο εξαρτημένες μεταβλητές: **την παρουσία του πεζού εντός ή εκτός της διάβασης και τη νομιμότητα ή παρανομία της διέλευσης**.

Στο πρώτο στάδιο εφαρμόστηκε η μέθοδος **point-biserial correlation**, με σκοπό τον εντοπισμό μεταβλητών που παρουσιάζουν στατιστικά σημαντική συσχέτιση με την κάθε εξαρτημένη μεταβλητή. Στη συνέχεια, εφαρμόστηκε **λογιστική παλινδρόμηση** και η μέθοδος των **τυχαίων δασών (Random Forests)**, με στόχο τόσο την **πρόβλεψη** της εξαρτημένης μεταβλητής σε κάθε περίπτωση (π.χ. αν ο πεζός θα βρίσκεται εντός ή εκτός διάβασης, ή αν η διέλευση θα είναι νόμιμη ή όχι), όσο και την ερμηνεία της επίδρασης των ανεξάρτητων μεταβλητών. Η λογιστική παλινδρόμηση

επέτρεψε τον υπολογισμό της κατεύθυνσης και της ισχύος της σχέσης κάθε μεταβλητής με την παρατηρούμενη συμπεριφορά, ενώ τα τυχαία δάση παρείχαν πληροφορίες για τη σχετική σημασία κάθε παράγοντα στη διαδικασία πρόβλεψης.

6.2 Συμπεράσματα σύγκρισης αυτοματοποιημένης – χειροκίνητης μεθόδου

Αναγνώριση κατάστασης φωτεινού σηματοδότη πεζών ανά κύκλο εναλλαγής

Η αξιολόγηση της ακρίβειας στην αναγνώριση της κατάστασης του φωτεινού σηματοδότη ανέδειξε θετικά αποτελέσματα. Το μέσο ποσοστό **ακρίβειας** διαμορφώθηκε στο **62,77%** με την πλειονότητα των κύκλων να παρουσιάζει τιμές μεταξύ **50% και 80%** και ορισμένες φάσεις να φτάνουν σε επίπεδα άνω του 90%, γεγονός που υποδεικνύει την σταθερή και σε πολλές περιπτώσεις **υψηλή απόδοση του συστήματος**. Ωστόσο, καταγράφηκαν και φάσεις με σημαντικά χαμηλότερη ακρίβεια, κυρίως σε περιπτώσεις όπου η **ορατότητα του σηματοδότη επηρεάστηκε από μεγάλα οχήματα** (π.χ. λεωφορεία ή φορτηγά), οδηγώντας είτε σε **λανθασμένη ανίχνευση**, είτε σε **αδυναμία ταξινόμησης** της φάσης. Παρά τις επιμέρους αποκλίσεις, τα αποτελέσματα καταδεικνύουν ότι ο αλγόριθμος δύναται να εντοπίσει αποτελεσματικά τις φάσεις φωτεινού σηματοδότη σε πραγματικές κυκλοφοριακές συνθήκες.

Καταγραφή παράνομων διελεύσεων πεζών

Η αξιολόγηση της αυτοματοποιημένης μεθόδου ως προς την καταγραφή παράνομων διελεύσεων πεζών ανέδειξε **μέτριο επίπεδο ακρίβειας (47,54%)**, με ποσοστά που κυμαίνονται μεταξύ 38% και 59% ανά ομάδα κύκλων. Σε αρκετές φάσεις (π.χ. κύκλοι 1–5 και 31–35) καταγράφηκε ικανοποιητική απόδοση, ωστόσο σε άλλες περιόδους εμφανίστηκαν αποκλίσεις, οφειλόμενες κυρίως σε δύο παράγοντες: (α) **την απόκρυψη του φωτεινού σηματοδότη από μεγάλα οχήματα**, η οποία περιόρισε την ικανότητα σωστής αναγνώρισης της ένδειξης, και (β) **την ερμηνεία της χρονικής στιγμής της διέλευσης**, καθώς το σύστημα ταξινομούσε ως παράνομες τις περιπτώσεις πεζών που ξεκίνησαν νόμιμα με πράσινο, αλλά ολοκλήρωσαν τη διάβαση ενώ το φανάρι είχε αλλάξει σε κόκκινο. Ελλείψει χρονικής ανάλυσης της αλληλουχίας, η αυτοματοποιημένη μέθοδος κατέγραψε **λανθασμένα** αυτές τις περιπτώσεις ως παραβάσεις. Παρά τις παραπάνω προκλήσεις, τα αποτελέσματα δείχνουν ότι το σύστημα μπορεί να υποστηρίξει μαζική και αυτοματοποιημένη παρακολούθηση συμμόρφωσης πεζών σε αστικές περιοχές, με περιθώρια περαιτέρω βελτίωσης στην ερμηνεία δυναμικής κίνησης.

Καταγραφή νόμιμων διελεύσεων πεζών

Η αυτοματοποιημένη μέθοδος παρουσίασε την **υψηλότερη συνολική απόδοση** στην κατηγορία των **νόμιμων διελεύσεων**, με μέση ακρίβεια **64,60%**, ξεπερνώντας την επίδοση που καταγράφηκε τόσο για τις παράνομες διελεύσεις όσο και για την ανίχνευση της φάσης του φωτεινού σηματοδότη. Σε πολλές ομάδες κύκλων (π.χ. 6–10 με 92,6%, 16–20 με 71,3%) παρατηρήθηκε εξαιρετικά υψηλή συμφωνία με τη χειροκίνητη μέθοδο, γεγονός που αναδεικνύει τη συστηματική ικανότητα του συστήματος να αναγνωρίζει τη συμμόρφωση των πεζών με τη σήμανση. Παρότι σε

κάποιας περιπτώσεις (π.χ. κύκλοι 36–39 με 46,7%) η ακρίβεια μειώθηκε, αυτό αποδόθηκε σε **οπτικά εμπόδια**, όπως διερχόμενα ή σταθμευμένα μεγάλα οχήματα, τα οποία περιόρισαν την ορατότητα του φωτεινού σηματοδότη. Δεδομένου ότι ο αλγόριθμος βασίζεται σε χρωματική πληροφόρηση, η **αδυναμία σαφούς ανίχνευσης του φωτός** σε συνεχόμενα καρέ οδηγεί σε ταξινομήσεις «άγνωστης κατάστασης» ή εσφαλμένες αναγνώσεις. Παρά τις επιμέρους αδυναμίες, η ανίχνευση των νόμιμων διελεύσεων αποτέλεσε το πιο αξιόπιστο και σταθερό σκέλος της μεθόδου, ενισχύοντας τη δυνατότητα της τεχνολογίας υπολογιστικής όρασης να αξιοποιηθεί σε εφαρμογές αυτοματοποιημένης επιτήρησης συμμορφωμένης συμπεριφοράς πεζών.

6.3 Συμπεράσματα στατιστικών αναλύσεων

Στον πίνακα που ακολουθεί παρουσιάζονται οι σημαντικότεροι δείκτες των μοντέλων που υλοποιήθηκαν, ενώ στα επόμενα εδάφια ακολουθεί αναλυτική παρουσίαση των κυριότερων συμπερασμάτων που αυτά προσέφεραν στην μελέτη.

1ο Λογιστικό Μοντέλο - Μεταβλητή Στόχος: Παρουσία πεζού εντός/εκτός διάβασης							Pseudo R^2	AIC	BIC	ACCURACY	AUC
Variable	Coef.	Std.Err.	z	P> z	[0.025	0.975]					
const	-2.518	0.069	-36.375	1E-289	-2.654	-2.382					
MinTTC_wins_log	0.961	0.037	26.050	1E-149	0.889	1.033					
vy	0.384	0.024	15.681	2E-55	0.336	0.432					
GroundPlaneCentroid_1_veh	0.077	0.036	2.100	4E-02	0.005	0.148					
magnitude_veh	-0.238	0.028	-8.420	4E-17	-0.293	-0.183					
VehicleStatus	0.371	0.049	7.639	2E-14	0.276	0.466					
PedestrianStatus	-1.969	0.066	-29.778	7E-195	-2.099	-1.840					
ResolvedTrafficLightStatus	1.867	0.046	40.502	0E+00	1.776	1.957					
IsInCrossingArea_veh	-0.117	0.056	-2.098	4E-02	-0.226	-0.008					
2ο Λογιστικό Μοντέλο - Μεταβλητή Στόχος: Νόμιμη/Παράνομη διέλευση πεζού							0.907	1977.828	2046.51	0.99	0.99
Variable	Coef.	Std.Err.	z	P> z	[0.025	0.975]					
const	5.726	0.145	39.606	0.000	5.443	6.009					
MinTTC_wins_log	-0.213	0.072	-2.965	0.003	-0.353	-0.072					
Confidence	0.842	0.090	9.347	0.000	0.665	1.018					
vy	-0.379	0.044	-8.590	0.000	-0.466	-0.293					
GroundPlaneCentroid_2_veh	-0.445	0.048	-9.214	0.000	-0.540	-0.351					
magnitude_veh	0.798	0.092	8.713	0.000	0.618	0.977					
VehicleStatus	-0.292	0.129	-2.272	0.023	-0.544	-0.040					
ResolvedTrafficLightStatus	4.581	0.160	28.568	0.000	4.267	4.895					
IsInCrossingArea	-3.039	0.091	-33.383	0.000	-3.217	-2.860					
1ο Μοντέλο Τυχαίου Δάσους - Μεταβλητή Στόχος: Παρουσία πεζού εντός/εκτός διάβασης										0.97	0.99
Variable	Importance										
GroundPlaneCentroid_2	0.234										
magnitude	0.200										
PedestrianStatus	0.180										
ResolvedTrafficLightStatus	0.106										
MinTTC_wins_log	0.079										
magnitude_veh	0.063										
Confidence	0.045										

GroundPlaneCentroid_1_veh	0.044					
GroundPlaneCentroid_2_veh	0.021					
Confidence_veh	0.017					
IsInCrossingArea_veh	0.010					
VehicleStatus	0.001					
2ο Μοντέλο Τυχαίου Δάσους - Μεταβλητή Στόχος: Νόμιμη/Παράνομη διέλευση πεζού						
Variable	Importance					
IsInCrossingArea	0.580					
magnitude	0.172					
ResolvedTrafficLightStatus	0.136					
MinTTC_wins_log	0.034					
magnitude_veh	0.028				0.99	0.99
Confidence	0.022					
GroundPlaneCentroid_1_veh	0.017					
Confidence_veh	0.011					
VehicleStatus	0.000					

6.3.1 Συσχέτιση σημειακής διχοτομίας – Εξαρτημένη μεταβλητή: Παρουσία πεζού εντός/εκτός διάβασης

Η ανάλυση ανέδειξε ότι οι παραβάτες πεζοί εμφανίζονται συχνά εντός διαβάσεων, γεγονός που οφείλεται αφενός στον μεθοδολογικό ορισμό της παραβατικότητας —ο οποίος προϋποθέτει την ύπαρξη φωτεινής ρύθμισης— και αφετέρου στη φύση της συμπεριφοράς αυτής, η οποία χαρακτηρίζεται από πρόθεση και δομή. Η **παραβατική διάσχιση** δεν πραγματοποιείται τυχαία ή άναρχα, αλλά **εντός πλαισίου** που επιτρέπει τον **σχεδιασμό της ενέργειας**, ακόμα κι αν αυτή παραβιάζει τους κανόνες. Αντίθετα, η κίνηση εκτός διάβασης, παρότι ενδέχεται να εμπεριέχει στοιχεία μη συμμόρφωσης, δεν καταγράφεται ως παραβατική, γεγονός που επισημαίνει τα όρια του εργαλείου παρατήρησης και αναδεικνύει την ανάγκη περαιτέρω μελέτης των άτυπων μορφών διάσχισης.

Η **ταχύτητα** κίνησης βρέθηκε αυξημένη στους πεζούς που κινούνται εντός διαβάσεων, κάτι που υποδηλώνει αποφασιστικότητα και συνέχεια στην κινητική συμπεριφορά. Η παρατήρηση αυτή αντανακλά δύο πιθανά προφίλ: από τη μία, **παραβάτες** που **επιταχύνουν** για να περιορίσουν την έκθεσή τους στον κίνδυνο· από την άλλη, **νόμιμους** πεζούς που κινούνται με ρυθμό λόγω αίσθησης **προτεραιότητας** ή ομαλής κυκλοφορίας. Σε κάθε περίπτωση, η ταχύτερη κίνηση εντός διαβάσεων υποδεικνύει ότι ο χώρος αυτός προσφέρει ένα οργανωμένο πλαίσιο εντός του οποίου η πεζή μετακίνηση εξελίσσεται με μεγαλύτερη συνοχή.

Ο **χρόνος έως πιθανή σύγκρουση** εμφανίζεται μεγαλύτερος όταν η διάσχιση πραγματοποιείται εντός διάβασης, στοιχείο που υποδηλώνει βελτιωμένες συνθήκες ασφάλειας. Αυτό ισχύει τόσο για τους **νόμιμους** πεζούς —οι οποίοι επωφελοούνται από την **παύση της κυκλοφορίας**— όσο και για τους **παραβάτες**, οι οποίοι επιλέγουν να διασχίσουν όταν εκτιμούν πως υπάρχει **πρόσκαιρο κενό** στη ροή των οχημάτων. Το εύρημα αυτό επιβεβαιώνει ότι η διάβαση, ως θεσμικά κατοχυρωμένος

χώρος, εξακολουθεί να λειτουργεί προστατευτικά, ακόμη και όταν δεν τηρούνται πλήρως οι κανόνες κυκλοφορίας.

6.3.2 Συσχέτιση σημειακής διχοτομίας – Εξαρτημένη μεταβλητή: Νόμιμη/Παράνομη διέλευση

Η ανάλυση ανέδειξε ότι οι παραβάτες πεζοί εντοπίζονται συχνά εντός της οριοθετημένης περιοχής διάβασης. Το εύρημα αυτό σχετίζεται με τον τρόπο ορισμού της παραβατικότητας στη μελέτη, ο οποίος βασίζεται στην παραβίαση σηματοδότη και όχι στην τοποθεσία του πεζού. Ως εκ τούτου, η **παράνομη συμπεριφορά** δεν ταυτίζεται με τυχαία ή χαοτική διέλευση, αλλά **εκδηλώνεται εντός οργανωμένου πλαισίου κυκλοφορίας**. Η επιλογή παραβίασης προκύπτει, επομένως, ως συνειδητή πράξη μη συμμόρφωσης και όχι ως ένδειξη λειτουργικής αδυναμίας ή άγνοιας του συστήματος.

Η **ταχύτητα** κίνησης των πεζών βρέθηκε **αυξημένη** μεταξύ όσων **παραβιάζουν** τον σηματοδότη, γεγονός που υποδηλώνει την ύπαρξη στρατηγικής μείωσης του χρόνου έκθεσης στον κίνδυνο. Αντίθετα, οι νόμιμοι πεζοί κινούνται με σταθερότερο ρυθμό, αξιοποιώντας τον διαθέσιμο χρόνο που προσφέρει η φάση πράσινου. Η διαφοροποίηση αυτή υποδηλώνει ότι η παραβατική διάσχιση δεν είναι απλώς ζήτημα παραβίασης κανόνων, αλλά συνοδεύεται από προσαρμοσμένες κινητικές τακτικές που εξυπηρετούν την αντιληπτή ανάγκη αυτοπροστασίας.

Ο διαθέσιμος χρόνος έως πιθανή σύγκρουση (**TTC**) βρέθηκε αυξημένος σε περιπτώσεις **παραβατικής συμπεριφοράς**, στοιχείο που ερμηνεύεται από τη στρατηγική επιλογή των πεζών να διασχίσουν όταν διαπιστώνουν κενό στην κυκλοφορία, ακόμη και αν το φανάρι είναι κόκκινο. Αντίθετα, οι νόμιμοι πεζοί ενδέχεται να ξεκινούν τη διάσχιση με οχήματα κοντά που αναμένεται να επιβραδύνουν, οδηγώντας σε χαμηλότερες τιμές TTC, χωρίς όμως αυτό να συνεπάγεται μεγαλύτερο πραγματικό ρίσκο. Συνεπώς, οι υψηλές τιμές TTC στους παραβάτες δεν υποδηλώνουν μεγαλύτερη ασφάλεια, αλλά διαφορετικό τρόπο αξιολόγησης των χρονικών παραθύρων κινδύνου.

6.3.3 Λογιστική Παλινδρόμηση – Εξαρτημένη μεταβλητή: Παρουσία πεζού εντός/εκτός διάβασης

Συνοπτική απόδοση μοντέλου

Το τελικό μοντέλο παρουσίασε υψηλή προβλεπτική ικανότητα (**AUC = 0,96, συνολική ακρίβεια $\approx$ 93%**) και δεν εμφάνισε προβλήματα πολυσυγγραμμικότητας ή στατιστικής αστάθειας. Όλες οι μεταβλητές ήταν στατιστικά σημαντικές.

Ερμηνεία συντελεστών

- **PedestrianStatus ($\beta = -1.969, p < 0.001$)**

Η αρνητική τιμή του συντελεστή δείχνει ότι οι **παραβάτες πεζοί** (τιμή 0) εμφανίζονται **συχνά εντός διάβασης**. Το εύρημα αυτό συνδέεται άμεσα με τη μεθοδολογική λογική της μελέτης, καθώς η παραβατικότητα ορίζεται μόνο όταν ο πεζός διασχίζει εντός

διάβασης με κόκκινο. Συνεπώς, η παραβίαση δεν σχετίζεται με άτυπη ή ακανόνιστη διάσχιση, αλλά με **ενσυνείδητη επιλογή** παράκαμψης του σηματοδότη εντός του θεσμικά προβλεπόμενου χώρου. Ο παραβάτης **επιλέγει τη διάβαση** ως σκηνικό παραβίασης, γεγονός που αναδεικνύει μια δομημένη, προσχεδιασμένη στρατηγική παραβατικότητας.

• **VehicleStatus ($\beta = +0.371$, $p < 0.001$)**

Η παρουσία **νόμιμου οχήματος** (τιμή 1) – δηλαδή οχήματος που έχει προτεραιότητα λόγω πράσινου σηματοδότη – συσχετίζεται θετικά με την παρουσία του πεζού εντός διάβασης. Αυτό δείχνει ότι **ο πεζός επιλέγει να διασχίσει, ακόμα και όταν δεν του επιτρέπεται**, γεγονός που αποκαλύπτει παραβίαση του σηματοδότη σε κατάσταση πλήρους επίγνωσης της προτεραιότητας του οχήματος. Η συμπεριφορά αυτή δηλώνει ότι η πεζή παραβατικότητα δεν προκύπτει από άγνοια ή σύγχυση, αλλά από **συνειδητή απόφαση** να διακινδυνεύσει τη διάσχιση, παρά την αντίθετη ροή.

• **ResolvedTrafficLightStatus ($\beta = +1.867$, $p < 0.001$)**

Ο **πράσινος σηματοδότης** για πεζούς αποτελεί τον ισχυρότερο θετικό προγνωστικό παράγοντα για την παρουσία εντός διάβασης. Το εύρημα αποτυπώνει την ισχυρή **ρυθμιστική δύναμη** του σηματοδότη στη συμπεριφορά των πεζών, καθώς υπό συνθήκες σαφούς προτεραιότητας η χρήση της διάβασης είναι αυξημένη και προβλέψιμη.

• **MinTTC_wins_log ($\beta = +0.961$, $p < 0.001$)**

Ο αυξημένος ελάχιστος χρόνος έως πιθανή σύγκρουση σχετίζεται θετικά με την παρουσία εντός διάβασης, καταδεικνύοντας ότι η διάσχιση στον τυπικό χώρο συνοδεύεται από **μεγαλύτερα περιθώρια ασφαλείας**. Η παραβατική διάσχιση φαίνεται να πραγματοποιείται σε στιγμές όπου ο πεζός εντοπίζει **πρόσκαιρο κενό** στην κυκλοφορία, υπολογίζοντας το διαθέσιμο περιθώριο χρόνου. Η επιλογή αυτή δεν αντανακλά αυθορμητισμό αλλά στιγμιαία στρατηγική **προσαρμογή στο δυναμικό περιβάλλον**. Από την άλλη μεριά, οι νόμιμοι επωφελούνται από τη διακοπή της κυκλοφορίας. Σε κάθε περίπτωση, η επιλογή της διάβασης συνδέεται με συνθήκες σχετικής ασφάλειας.

• **vy ($\beta = +0.384$, $p < 0.001$)**

Η συνιστώσα ταχύτητας του πεζού κατά τον άξονα διάσχισης παρουσιάζει θετική συσχέτιση με την παρουσία εντός διάβασης. Η κίνηση σε αυτό το πλαίσιο φαίνεται να είναι πιο **σταθερή**, στοχευμένη και αποφασιστική, κάτι που ενδέχεται να υποδηλώνει είτε αίσθηση νομιμότητας, είτε **επιτάχυνση** για μείωση χρόνου έκθεσης σε περίπτωση παραβίασης.

• **GroundPlaneCentroid_1_veh ($\beta = +0.077$, $p = 0.036$)**

Η σχετική θέση του οχήματος στο οπτικό πεδίο συσχετίζεται θετικά με την παρουσία πεζού εντός διάβασης. Όταν το όχημα βρίσκεται **χωρικά πιο μακριά από τη διάβαση**, ενδέχεται ο πεζός να εκτιμά ότι υπάρχει περιθώριο για ασφαλή διάσχιση, ακόμη και αν δεν έχει προτεραιότητα.

- **magnitude_veh** ($\beta = -0.238$, $p < 0.001$)

Η συνολική ταχύτητα του οχήματος σχετίζεται αρνητικά με την παρουσία του πεζού εντός διάβασης. Η αυξημένη κινητικότητα λειτουργεί **αποτρεπτικά** για την πεζή διάσχιση, πιθανόν λόγω αυξημένης αντίληψης κινδύνου ή απουσίας χρονικού παραθύρου για δράση. Το εύρημα καταδεικνύει τη σημασία της δυναμικής των οχημάτων στη λήψη απόφασης του πεζού.

- **IsInCrossingArea_veh** ($\beta = -0.117$, $p = 0.036$)

Η παρουσία οχήματος εντός της διάβασης μειώνει την πιθανότητα παρουσίας πεζού στο ίδιο σημείο. Η **φυσική παρεμπόδιση** λειτουργεί ανασταλτικά, ενισχύοντας τη χωρική αποτροπή. Το εύρημα είναι ευθυγραμμισμένο με τη λογική αλληλοαποκλεισμού των δύο ροών, ανθρώπινης και μηχανοκίνητης.

6.3.4 Λογιστική Παλινδρόμηση – Εξαρτημένη μεταβλητή: Νόμιμη/Παράνομη διέλευση πεζού

Συνοπτική απόδοση μοντέλου

Το τελικό μοντέλο εμφάνισε εξαιρετική προβλεπτική ικανότητα (**AUC = 0,99**, **συνολική ακρίβεια $\approx 99,5\%$**) και δεν παρουσίασε προβλήματα πολυσυγγραμμικότητας ή στατιστικής αστάθειας. Όλες οι ανεξάρτητες μεταβλητές ήταν στατιστικά σημαντικές.

Ερμηνεία συντελεστών

- **ResolvedTrafficLightStatus** ($\beta = +4.581$, $p < 0.001$)

Ο **πράσινος σηματοδότης** για πεζούς αποτελεί τον ισχυρότερο θετικό δείκτη για νόμιμη διέλευση, επιβεβαιώνοντας τον ρυθμιστικό του ρόλο στη συμμόρφωση των πεζών.

- **Confidence** ($\beta = +0.842$, $p < 0.001$)

Η αυξημένη βεβαιότητα του αλγορίθμου για την παρουσία πεζού σχετίζεται με νόμιμη διάσχιση. Το εύρημα υποδηλώνει ότι οι **νόμιμοι πεζοί** κινούνται με πιο σταθερό, ομαλό και **προβλέψιμο τρόπο**, γεγονός που ενισχύει την ακρίβεια εντοπισμού τους από το σύστημα υπολογιστικής όρασης.

- **magnitude_veh** ($\beta = +0.798$, $p < 0.001$)

Όταν τα οχήματα κινούνται με μεγαλύτερη **ταχύτητα**, αυξάνεται η πιθανότητα οι πεζοί να περιμένουν και να διασχίζουν **νόμιμα**. Η έντονη κυκλοφορία λειτουργεί αποτρεπτικά ως προς την παραβίαση.

- **MinTTC_wins_log** ($\beta = -0.213$, $p = 0.003$)

Η **αύξηση** του διαθέσιμου χρόνου έως πιθανή σύγκρουση συνδέεται με μεγαλύτερη πιθανότητα παραβίασης. Όταν ο πεζός αντιλαμβάνεται επαρκές χρονικό περιθώριο,

φαίνεται πιο πρόθυμος να αγνοήσει τον φωτεινό σηματοδότη και να διασχίσει. Το εύρημα δείχνει ότι η παρανομία βασίζεται, συχνά, σε υπολογισμένο ρίσκο.

• **vy ($\beta = -0.379$, $p < 0.001$)**

Η αυξημένη **ταχύτητα** του πεζού κατά τη διάσχιση σχετίζεται με **παρανομία**. Οι πεζοί κινούνται ταχύτερα πιθανόν για να μειώσουν τον χρόνο έκθεσης ή να προλάβουν να περάσουν πριν την κίνηση των οχημάτων.

• **GroundPlaneCentroid_2_veh ($\beta = -0.445$, $p < 0.001$)**

Η **πλευρική μετακίνηση του οχήματος** από τον κεντρικό άξονα του δρόμου συνδέεται με αυξημένη πιθανότητα παράνομης διέλευσης πεζού. Όταν το όχημα βρίσκεται σε πλάγια θέση (δεξιά ή αριστερά), ενδέχεται ο πεζός να θεωρεί ότι η πορεία του οχήματος δεν τον επηρεάζει άμεσα, οδηγώντας σε λανθασμένη εκτίμηση κινδύνου και παραβίαση των κανόνων διάσχισης. Αντίθετα, όταν το όχημα βρίσκεται ευθυγραμμισμένο με τον άξονα του δρόμου, η παρουσία του λειτουργεί πιο «πειθαρχητικά» και η παράνομη διέλευση περιορίζεται.

• **VehicleStatus ($\beta = -0.292$, $p = 0.023$)**

Η παρουσία οχήματος με προτεραιότητα (**πράσινο σήμα** για όχημα) σχετίζεται με παράνομη διέλευση πεζού, αφού προκύπτει πως επιλέγει να διασχίσει ακόμη και όταν δεν του επιτρέπεται.

• **IsInCrossingArea ($\beta = -3.039$, $p < 0.001$)**

Η ισχυρή αρνητική συσχέτιση δείχνει ότι οι **παραβάσεις** εντοπίζονται **εντός της διάβασης**. Η παρανομία στη μελέτη ορίζεται βάσει παραβίασης του σηματοδότη, όχι βάσει θέσης. Οι πεζοί παρανομούν εντός του οργανωμένου πλαισίου της διάβασης, συχνά με πλήρη επίγνωση και συνειδητή επιλογή μη συμμόρφωσης.

6.3.5 Τυχαία δάση – Εξαρτημένη μεταβλητή: Παρουσία πεζού εντός / εκτός διάβασης

Προβλεπτική ικανότητα μοντέλου

Το μοντέλο τυχαίου δάσους παρουσίασε πολύ υψηλή προβλεπτική ικανότητα, με συνολική **ακρίβεια** περίπου **97%** και **AUC = 0.99**. Η απόδοση αυτή επιβεβαιώνει τη δυνατότητα του αλγορίθμου να διαχωρίζει με αξιοπιστία τις περιπτώσεις παρουσίας πεζού εντός ή εκτός διάβασης, με βάση ένα ευρύ φάσμα χαρακτηριστικών που αποτυπώνουν τόσο τη χωρική θέση όσο και τη δυναμική της σκηνής. Παρότι η κατηγορία “εντός διάβασης” υποεκπροσωπείται, το μοντέλο διατήρησε ικανοποιητικά επίπεδα ακρίβειας και ευαισθησίας, γεγονός που καταδεικνύει σταθερή γενικευσιμότητα.

Σημαντικές μεταβλητές πρόβλεψης

Η σημαντικότερη μεταβλητή ήταν η `GroundPlaneCentroid_2`, που εκφράζει τη **θέση του πεζού στον άξονα βάθους** (δηλαδή η τοποθεσία που εντοπίζεται σε σχέση με την εγκατεστημένη κάμερα). Το εύρημα δείχνει ότι το σημείο καταγραφής του πεζού στον χώρο αποτελεί κρίσιμο παράγοντα για την πρόβλεψη.

Η συνολική του **ταχύτητα** (`magnitude`) αναδείχθηκε ως δεύτερος σημαντικός παράγοντας, γεγονός που υποδηλώνει ότι η δυναμική του πεζού φέρει πληροφορία σχετικά με τη θέση του ως προς τη διάβαση. Παρόμοια, η μεταβλητή `PedestrianStatus`, που κωδικοποιεί το **είδος της πεζής συμπεριφοράς**, ενίσχυσε την πρόβλεψη, επιβεβαιώνοντας ότι ο τρόπος κίνησης είναι διακριτός.

Η **κατάσταση του φωτεινού σηματοδότη** (`ResolvedTrafficLightStatus`) βρέθηκε επίσης μεταξύ των μεταβλητών με υψηλή σημασία, γεγονός που καταδεικνύει τη σημασία του ρυθμιστικού πλαισίου. Ο **ελάχιστος αναμενόμενος χρόνος σύγκρουσης** (`MinTTC_wins_log`) παρείχε πληροφορία για τη χρονική συγκυρία της διάσχισης, πιθανώς αποτυπώνοντας τη σχέση της παρουσίας εντός διάβασης με ασφαλέστερες συνθήκες κυκλοφορίας.

Οι **ταχύτητες των οχημάτων** (`magnitude_veh`) και η **χωρική τους θέση** (`GroundPlaneCentroid_1_veh`, `GroundPlaneCentroid_2_veh`) συνεισέφεραν μεσαίας έντασης στην πρόβλεψη, χωρίς να είναι απόλυτα καθοριστικές. Αυτό ενδέχεται να οφείλεται στο ότι, σε πολλές περιπτώσεις, η θέση και η ταχύτητα του οχήματος δεν διαφοροποιούν σαφώς το αν ο πεζός είναι εντός ή εκτός διάβασης.

Η μεταβλητή `Confidence`, δηλαδή η **εμπιστοσύνη** του συστήματος στην ανίχνευση του πεζού, εμφάνισε σχετικά μικρό ρόλο στο τελικό αποτέλεσμα. Το ίδιο ισχύει και για την `Confidence_veh` για τα οχήματα, ένδειξη ότι η ποιότητα της ανίχνευσης δεν αποτελεί βασικό διακριτικό στοιχείο.

Η `IsInCrossingArea_veh`, που αφορά την **παρουσία οχήματος εντός διάβασης**, είχε πολύ χαμηλή προβλεπτική σημασία. Το γεγονός αυτό πιθανόν σχετίζεται με το ότι, στη συγκεκριμένη εργασία, τόσο οι πεζοί εντός όσο και εκτός διάβασης μπορούν να συναντήσουν οχήματα στον ίδιο χώρο, οπότε η μεταβλητή αυτή δεν διαχωρίζει τις δύο κατηγορίες. Τέλος, η μεταβλητή `VehicleStatus_bin`, που αποτυπώνει την **προτεραιότητα του οχήματος**, εμφάνισε σχεδόν μηδενική συνεισφορά, γεγονός που ενισχύει την υπόθεση περί πλεονάζουσας πληροφορίας μέσω του σηματοδότη.

Συνολικά, οι ισχυρότεροι προβλεπτικοί παράγοντες σχετίζονται κυρίως με τον ίδιο τον πεζό — ως προς τη θέση του, την ταχύτητά του και τις συνθήκες κυκλοφορίας τη στιγμή της διάσχισης — αναδεικνύοντας ότι η ταξινόμηση βασίζεται σε παρατηρήσιμα μοτίβα συμπεριφοράς και χωρικής τοποθέτησης.

6.3.6 Τυχαία δάση – Εξαρτημένη μεταβλητή: Νόμιμη/Παράνομη διέλευση πεζού

Προβλεπτική ικανότητα μοντέλου

Το μοντέλο τυχαίου δάσους εμφάνισε εξαιρετικά υψηλή ακρίβεια στην πρόβλεψη της παραβατικής συμπεριφοράς πεζών, με ποσοστά σωστής ταξινόμησης που αγγίζουν το **99%**. Η επίδοση αυτή υποδηλώνει ότι το μοντέλο μπορεί να αποτυπώσει με σαφήνεια τα μοτίβα που διαφοροποιούν τις νόμιμες από τις παραβατικές διελεύσεις, γεγονός που αναδεικνύει την αξιοπιστία και την εφαρμοστική του αξία.

Σημαντικές μεταβλητές πρόβλεψης

Η **παρουσία του πεζού εντός της διάβασης** (IsInCrossingArea) αναδείχθηκε ως ο ισχυρότερος παράγοντας πρόβλεψης της παραβατικής συμπεριφοράς, επιβεβαιώνοντας τη θεμελιώδη σχέση μεταξύ χωρικής θέσης και επιλογής δράσης. Η **ταχύτητα** κίνησης του πεζού καθώς και η **κατάσταση του φωτεινού σηματοδότη** (ResolvedTrafficLightStatus) συνέβαλαν σημαντικά στην πρόβλεψη, αποτυπώνοντας τον συνδυασμό χωρικών και χρονικών ενδείξεων που επηρεάζουν τη λήψη απόφασης.

Επιπλέον, ο **ελάχιστος αναμενόμενος χρόνος έως πιθανή σύγκρουση** (MinTTC_wins_log) παρουσίασε θετική επίδραση, υποδεικνύοντας ότι ο πεζός ενδέχεται να εκτιμά ενεργά το διαθέσιμο περιθώριο ασφαλείας πριν παρανομήσει. Σημαντικές, αν και λιγότερο έντονες, ήταν και μεταβλητές που σχετίζονται με τα **οχήματα**, όπως η **ταχύτητά** τους (magnitude_veh) και η **διαμήκης τους θέση** στον δρόμο (GroundPlaneCentroid_1_veh).

Η **αξιοπιστία των εντοπισμών**, τόσο του πεζού (Confidence) όσο και του οχήματος (Confidence_veh), φαίνεται να έχει μέτρια επίδραση, ενδεχομένως ως ένδειξη της σαφήνειας της σκηνής ή της σταθερότητας του οπτικού σήματος. Αντίθετα, η δυαδική μεταβλητή VehicleStatus, η οποία εκφράζει αν το όχημα βρίσκεται σε **νόμιμη ή παράνομη κατάσταση** βάσει της οδικής του συμπεριφοράς, εμφάνισε εξαιρετικά χαμηλή σημασία. Το εύρημα αυτό υποδηλώνει ότι η παραβατικότητα του πεζού δεν επηρεάζεται άμεσα από το αν το όχημα τηρεί ή παραβιάζει τους κανόνες, αλλά μάλλον διαμορφώνεται ανεξάρτητα από τέτοιου είδους εξωτερικές παραβάσεις.

Συνολικά, τα αποτελέσματα αποτυπώνουν ένα σύνθετο προφίλ παραγόντων που επηρεάζουν την πεζή παραβατικότητα, συνδυάζοντας τη θέση, την κίνηση και την αντίληψη κινδύνου με τον βαθμό κατανόησης του περιβάλλοντος, όπως αυτό αποτυπώνεται από χωρικά, χρονικά και ποιοτικά χαρακτηριστικά.

6.4 Προτάσεις για περαιτέρω έρευνα

1. **Αξιοποίηση χρονικής πληροφορίας για την εκτίμηση της κατάστασης του σηματοδότη:** Η παρούσα μεθοδολογία βασίζεται στην ανάλυση του χρώματος του φωτεινού σηματοδότη σε μεμονωμένα καρέ. Σε περιπτώσεις όπου η ορατότητα παρεμποδίζεται προσωρινά (π.χ. από διερχόμενα ή σταθμευμένα οχήματα), η κατάσταση του σηματοδότη δεν μπορεί να εκτιμηθεί με σαφήνεια. Μελλοντικά, η ενσωμάτωση χρονικής πληροφορίας (π.χ. μέσω πρόβλεψης ή διατήρησης προηγούμενης τιμής) θα μπορούσε να περιορίσει την εμφάνιση τιμών "unknown" και να ενισχύσει τη συνέχεια στην ανάλυση.
2. **Ανάλυση της χρονικής στιγμής εισόδου στο οδόστρωμα για ακριβέστερη εκτίμηση παραβάσεων:** Το σύστημα καταγράφει παραβάσεις όταν ένας πεζός εντοπίζεται στον δρόμο ενώ ο σηματοδότης είναι κόκκινος. Ωστόσο, δεν λαμβάνεται υπόψη η χρονική στιγμή κατά την οποία ο πεζός ξεκίνησε τη διέλευση. Η μελλοντική ενσωμάτωση της πληροφορίας αυτής μπορεί να επιτρέψει την ακριβέστερη διάκριση μεταξύ νόμιμων και παράνομων διελεύσεων, αποφεύγοντας εσφαλμένους χαρακτηρισμούς.
3. **Επέκταση της μεθόδου σε διαφορετικές τοποθεσίες και συνθήκες:** Η παρούσα εργασία περιορίστηκε σε ένα συγκεκριμένο σημείο στο κέντρο της Αθήνας και σε προκαθορισμένο χρονικό πλαίσιο. Μελλοντική εφαρμογή της μεθοδολογίας σε διαφορετικά σημεία του αστικού ιστού, υπό μεταβαλλόμενες συνθήκες φωτισμού, καιρικές καταστάσεις και επίπεδα κυκλοφορίας, θα επέτρεπε την αξιολόγηση της γενικευσιμότητας και της προσαρμοστικότητας του συστήματος.
4. **Συσχέτιση παραβάσεων με περιβαλλοντικές και κυκλοφοριακές συνθήκες:** Διεθνείς μελέτες έχουν δείξει ότι παράγοντες όπως η διάρκεια αναμονής, η πυκνότητα πεζών και η ένταση της κυκλοφορίας μπορούν να επηρεάσουν τη συμπεριφορά των πεζών. Μελλοντική έρευνα θα μπορούσε να καταγράψει τέτοιες μεταβλητές και να διερευνήσει τη στατιστική τους συσχέτιση με την εμφάνιση παραβάσεων, προσφέροντας βαθύτερη κατανόηση των αιτιών πίσω από αυτές.
5. **Εφαρμογή spatio-temporal μοντέλων πρόβλεψης πρόθεσης:** Η αξιοποίηση νευρωνικών δικτύων με χρονική μνήμη, όπως τα GRU, LSTM ή Transformer, μπορεί να ενισχύσει τη δυνατότητα πρόβλεψης της πρόθεσης διέλευσης πεζών. Τέτοια μοντέλα μπορούν να επεξεργάζονται διαδοχικά καρέ και να αντλούν πληροφορία από την πορεία και το μοτίβο κίνησης του κάθε πεζού, επιτρέποντας προληπτική αναγνώριση παραβατικής ή επικίνδυνης συμπεριφοράς.
6. **Ανάλυση τροχιών πεζών για ενίσχυση της κατανόησης και πρόβλεψης της συμπεριφοράς:** Η παρακολούθηση των πεζών σε διαδοχικά καρέ επιτρέπει την εξαγωγή των τροχιών τους, οι οποίες αποτυπώνουν τη χρονική εξέλιξη της θέσης τους στον χώρο. Παρότι οι τροχιές αυτές παράγονται, δεν αξιοποιούνται για περαιτέρω ανάλυση. Μελλοντική έρευνα θα μπορούσε να

εστιάσει στην επεξεργασία και κατηγοριοποίησή τους, εντοπίζοντας επαναλαμβανόμενα μοτίβα κίνησης, απότομες αλλαγές πορείας ή ξαφνικές στάσεις. Αυτά τα στοιχεία μπορούν να εμπλουτίσουν τη συνολική κατανόηση της συμπεριφοράς των πεζών και να υποστηρίξουν προηγμένα μοντέλα πρόβλεψης ή ανίχνευσης παραβάσεων.

BIBΛΙΟΓΡΑΦΙΑ

1. Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Prentice Hall.
2. Alahi, A., Bierlaire, M., & Vandergheynst, P. (2014). Robust real-time pedestrians detection in urban environments with low-resolution cameras. *Transportation research part C: emerging technologies*, 39, 113-128.
3. Alhajyaseen, W. K., & Iryo-Asano, M. (2017). Studying critical pedestrian behavioral changes for the safety assessment at signalized crosswalks. *Safety science*, 91, 351-360.
4. Alsharif, T., Lanzaro, G., & Sayed, T. (2024). Distracted Walking: Does it impact pedestrian-vehicle interaction behavior?. *Accident Analysis & Prevention*, 208, 107789.
5. Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). *Permutation importance: A corrected feature importance measure*. *Bioinformatics*, 26(10), 1340–1347.
6. Araya-Porras, E., Mora-Calderón, A., & Aguero-Valverde, J. (2022). Pedestrian crossing light violation in Costa Rica: exploring factors affecting mid-block crossing behavior. *Ingeniería*, 32(2), 115-134.
7. Babyak, M. A. (2004). *What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models*. *Psychosomatic Medicine*, 66(3), 411–421.
8. Bonett, D. G. (2020). *Point-biserial correlation: Interval estimation, hypothesis testing, meta-analysis, and sample size determination*. *British Journal of Mathematical and Statistical Psychology*, 73(S1), 113–144.
9. Breiman, L. (2001). *Random forests*. *Machine Learning*, 45(1), 5–32.
10. Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). *The balanced accuracy and its posterior distribution*. In *Proceedings of the 20th International Conference on Pattern Recognition (ICPR)*
11. Brosseau, M., Zangenehpour, S., Saunier, N., & Miranda-Moreno, L. (2013). The impact of waiting time and other factors on dangerous pedestrian crossings and violations at signalized intersections: A case study in Montreal. *Transportation research part F: traffic psychology and behaviour*, 21, 159-172.
12. Budzynski, M., Gobis, A., Guminska, L., Jelinski, L., Kiec, M., & Tomczuk, P. (2021). Assessment of the influence of road infrastructure parameters on the behaviour of drivers and pedestrians in pedestrian crossing areas. *Energies*, 14(12), 3559.
13. Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer.
14. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). *SMOTE: Synthetic minority over-sampling technique*. *Journal of Artificial Intelligence Research*, 16, 321–357.
15. Chicco, D., & Jurman, G. (2020). *The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation*. *BMC Genomics*, 21, 6.
16. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

17. Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Routledge.
18. Cookson, R. E., Richards, D. C., & Cuerden, R. W. (2011). *The characteristics of pedestrian road traffic accidents and the resulting injuries*. IHS.
19. Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). *Random forests for classification in ecology*. *Ecology*, 88(11), 2783–2792.
20. Diependaele, K. (2019). Non-compliance with pedestrian traffic lights in Belgian cities. *Transportation research part F: traffic psychology and behaviour*, 67, 230-241.
21. Dupriez, B., & Houdmont, A. (2008). Accidents de piétons sur un passage pour piétons non réglé par feux en Région de Bruxelles-Capitale: le point de vue de l'infrastructure. *Le piéton et son environnement: Quelles interactions? Quelles adaptations?*
22. Fawcett, T. (2006). *An introduction to ROC analysis*. *Pattern Recognition Letters*, 27(8), 861–874.
23. Feliciani, C., & Nishinari, K. (2018). Empirical analysis of the lane formation process in bidirectional pedestrian flow. *Physical Review E*, 97(1), 012304
24. Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). *Learning from imbalanced data sets*. Springer.
25. Ferri, C., Hernández-Orallo, J., & Modroi, R. (2009). *An experimental comparison of performance measures for classification*. *Pattern Recognition Letters*, 30(1), 27–38.
26. Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). SAGE Publications Ltd
27. Field, A. P. (2018). *Discovering Statistics Using IBM SPSS Statistics* (5η έκδ.). SAGE Publications Ltd.
28. Friedman, J. H. (2001). *Greedy function approximation: A gradient boosting machine*. *Annals of Statistics*, 29(5), 1189–1232.
29. Glass, G. V. (1966). *Statistical methods in education and psychology*. Allyn & Bacon.
30. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
31. Herath, C., & Amarasingha, N. (2022). Factors Affecting Red-Light Running of Pedestrians at Signalized Intersections.
32. Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences* (5th ed.). Houghton Mifflin.
33. Holm, A., Jaani, J., Eensoo, D., & Piksööt, J. (2018). Pedestrian behaviour of 6th grade Estonian students: Implications of social factors and accident-prevention education at school. *Transportation research part F: traffic psychology and behaviour*, 52, 112-119
34. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley.
35. Hossin, M., & Sulaiman, M. N. (2015). *A review on evaluation metrics for data classification evaluations*. *International Journal of Data Mining & Knowledge Management Process*, 5(2), 1–11.
36. Hussein, M., Sayed, T., Reyad, P., & Kim, L. (2015). Automated pedestrian safety analysis at a signalized intersection in New York City: Automated data

- extraction for safety diagnosis and behavioral study. *Transportation Research Record*, 2519(1), 17-27.
37. Ibriaeva, O., Shepelev, V., Zhulev, A., Chizhova, M., Yakupova, G., & Fatikhova, L. (2020, November). The Use of the YOLO neural network in the task of separating vehicles and pedestrians at a signal-controlled intersection. In *2020 Global Smart Industry Conference (GloSIC)* (pp. 303-308). IEEE.
 38. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer.
 39. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning with applications in R* (2nd ed.). Springer
 40. Jay, M., Régnier, A., Dasnon, A., Brunet, K., & Pelé, M. (2020). The light is red: Uncertainty behaviours displayed by pedestrians during illegal road crossing. *Accident Analysis & Prevention*, 135, 105369.
 41. Jeni, L. A., Cohn, J. F., & De La Torre, F. (2013). *Facing imbalanced data—Recommendations for the use of performance metrics*. In *Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (pp. 245–251). IEEE.
 42. King, G., & Zeng, L. (2001). *Logistic regression in rare events data*. *Political Analysis*, 9(2), 137–163
 43. King, M. J., Soole, D., & Ghafourian, A. (2009). Illegal pedestrian crossing at signalised intersections: incidence and relative risk. *Accident Analysis & Prevention*, 41(3), 485-490.
 44. Koh, P. P., Wong, Y. D., & Chandrasekar, P. (2014). Safety evaluation of pedestrian behaviour and violations at signalised pedestrian crossings. *Safety science*, 70, 143-152
 45. Kornbrot, D. (2005). *Point-Biserial Correlation*. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science*
 46. Kornbrot, D. (2005). *Point-Biserial Correlation*. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science*
 47. Kumar, A., & Ghosh, I. (2022). Non-compliance behaviour of pedestrians and the associated conflicts at signalized intersections in India. *Safety science*, 147, 105604.
 48. Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5th ed.). McGraw-Hill/Irwin.
 49. Lipovac, K., Vujanic, M., Maric, B., & Nesic, M. (2013). Pedestrian behavior at signalized pedestrian crossings. *Journal of transportation engineering*, 139(2), 165-172.
 50. Lipton, Z. C. (2018). *The mythos of model interpretability*. *Communications of the ACM*, 61(10), 36–43.
 51. Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. SAGE Publications.
 52. Louppe, G., Wehenkel, L., Sutura, A., & Geurts, P. (2013). *Understanding variable importances in forests of randomized trees*. In *Advances in Neural Information Processing Systems*, 26, 431–439.
 53. Lundberg, S. M., & Lee, S.-I. (2017). *A unified approach to interpreting model predictions*. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS 2017)* (pp. 4765–4774).
 54. McFadden, D. (1974). *Conditional logit analysis of qualitative choice behavior*. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105–142). New York: Academic Press.

55. McGrath, R. E., & Meyer, G. J. (2006). *When effect sizes disagree: The case of r and d* . *Psychological Methods*, 11(4), 386–401.
56. Menard, S. (2002). *Applied logistic regression analysis* (2nd ed.). SAGE Publications.
57. Moore, D. S., McCabe, G. P., & Craig, B. A. (2017). *Introduction to the practice of statistics* (9th ed.). W. H. Freeman.
58. Mukherjee, D., & Mitra, S. (2020). A comprehensive study on factors influencing pedestrian signal violation behaviour: Experience from Kolkata City, India. *Safety science*, 124, 104610.
59. Olszewski, P., Buttler, I., Czajewski, W., Dąbkowski, P., Kraśkiewicz, C., Szagała, P., & Zielińska, A. (2016). Pedestrian safety assessment with video analysis. *Transportation research procedia*, 14, 2044-2053.
60. PRATELLI, A., LUPI, M., & Daniele, R. (2017). Illegal pedestrian crossing at a traffic light: a study on tourist behaviour. *International Journal of Transport Development and Integration*, 1(4), 633-639.
61. Pratelli, A., Lupi, M., Farina, A., & Iannelli, C. (2019). Pedestrian rule compliance at signalized intersections in different urban contexts: an observational study. *Transport Problems*, 14(2), 29-41.
62. Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann.
63. Radojcic, V., & Dobrojevic, M. ANALYSIS OF VEHICLE-PEDESTRIAN CONTACT USING COMPUTER VISION AND OPENCV.
64. Raoniar, R., & Maurya, A. K. (2022). Pedestrian red-light violation at signalised intersection crosswalks: Influence of social and non-social factors. *Safety science*, 147, 105583.
65. Rosenbloom, T. (2009). Crossing at a red light: Behaviour of individuals and groups. *Transportation research part F: traffic psychology and behaviour*, 12(5), 389-394.
66. Schwarz, G. (1978). *Estimating the dimension of a model*. *The Annals of Statistics*, 6(2), 461–464.
67. Sokolova, M., & Lapalme, G. (2009). *A systematic analysis of performance measures for classification tasks*. *Information Processing & Management*, 45(4), 427–437.
68. Tian, Z., Zhang, Y., Li, M., & Zhou, X. (2023). Modeling pedestrian crossing behavior at signalized intersections considering heterogeneity and compliance. *Transportation Research Part C: Emerging Technologies*, 149, 104004.
69. Triola, M. F. (2018). *Elementary statistics* (13th ed.). Pearson.
70. Ventura, R., Roussou, S., Ziakopoulos, A., Barabino, B., & Yannis, G. (2025). Using computer vision and street-level videos for pedestrian-vehicle tracking and behaviour analysis. *Transportation Research Interdisciplinary Perspectives*, 30, 101366.
71. Wang, W., Guo, H., Gao, Z., & Bubb, H. (2011). Individual differences of pedestrian behaviour in midblock crosswalk and intersection. *International Journal of Crashworthiness*, 16(1), 1-9.
72. Wong, P. K. Y., Luo, H., Wang, M., Leung, P. H., & Cheng, J. C. (2021). Recognition of pedestrian trajectories and attributes with computer vision and deep learning techniques. *Advanced Engineering Informatics*, 49, 101356.
73. Yang, J., Deng, W., Wang, J., Li, Q., & Wang, Z. (2006). Modeling pedestrians' road crossing behavior in traffic system micro-simulation in China. *Transportation Research Part A: Policy and Practice*, 40(3), 280-290.

74. Zaki, M. H., Sayed, T., Tageldin, A., & Hussein, M. (2013). Application of computer vision to diagnosis of pedestrian safety issues. *Transportation research record*, 2393(1), 75-84.
75. Zhang, J., Klingsch, W., Schadschneider, A., & Seyfried, A. (2011). Transitions in pedestrian fundamental diagrams of straight corridors and T-junctions. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(06), P06004
76. Zhang, S., Abdel-Aty, M., Cai, Q., Li, P., & Ugan, J. (2020). Prediction of pedestrian-vehicle conflicts at signalized intersections based on long short-term memory neural network. *Accident Analysis & Prevention*, 148, 105799.
77. Zhang, S., Abdel-Aty, M., Wu, Y., & Zheng, O. (2020). Modeling pedestrians' near-accident events at signalized intersections using gated recurrent unit (GRU). *Accident Analysis & Prevention*, 148, 105844.
78. Zhu, D., Sze, N. N., & Bai, L. (2021). Roles of personal and environmental factors in the red light running propensity of pedestrian: Case study at the urban crosswalks. *Transportation research part F: traffic psychology and behaviour*, 76, 47-58.
79. Zhu, D., Sze, N. N., Feng, Z., & Yang, Z. (2022). A two-stage safety evaluation model for the red light running behaviour of pedestrians using the game theory. *Safety science*, 147, 105600.
80. Δαμιανού, Χ. Χ., Παπαδάτου, Ν. Δ., & Χαραλαμπίδης, Χ. Α. (2003). *Εισαγωγή στις πιθανότητες και τη στατιστική*. Αθήνα: Σύμμετρία.