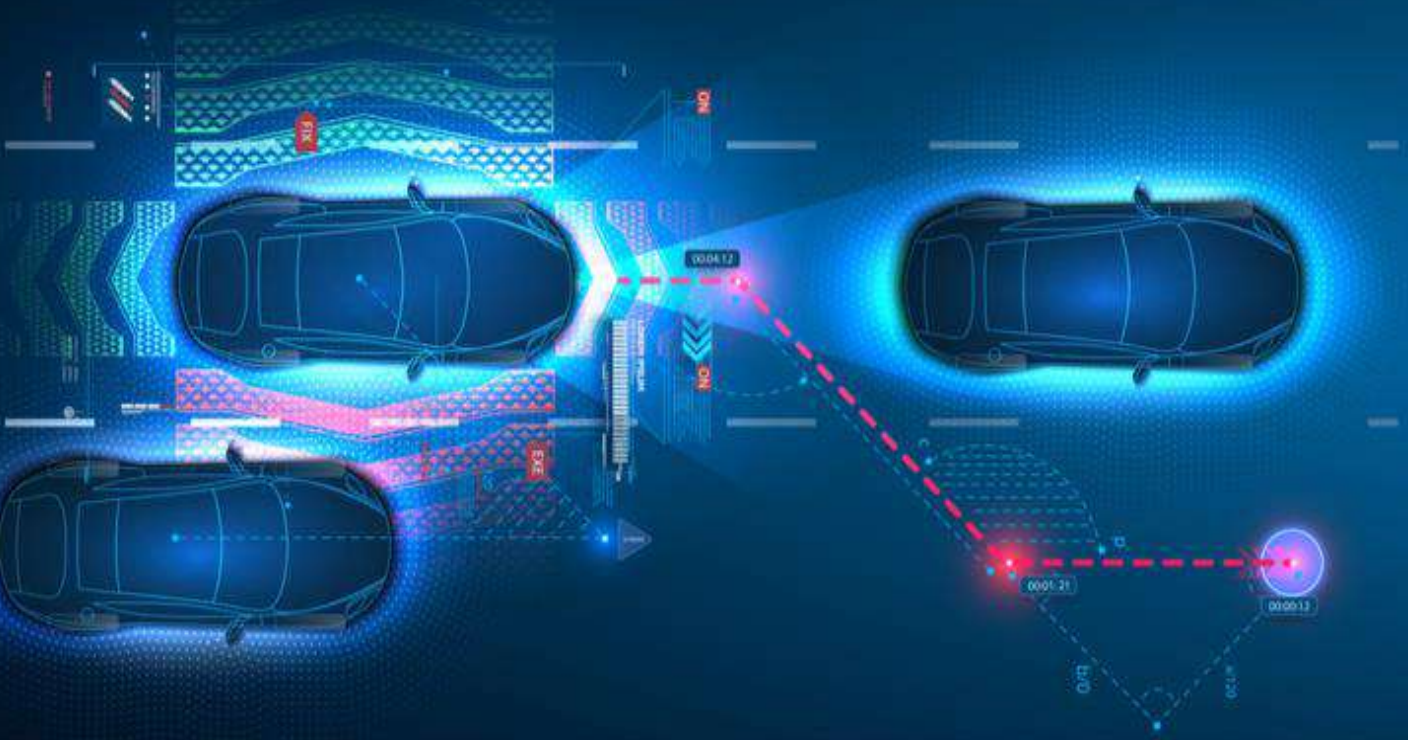




ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ  
ΤΟΜΕΑΣ ΜΕΤΑΦΟΡΩΝ ΚΑΙ ΣΥΓΚΟΙΝΩΝΙΑΚΗΣ ΥΠΟΔΟΜΗΣ

## ΔΙΕΡΕΥΝΗΣΗ ΤΗΣ ΕΠΙΚΙΝΔΥΝΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ ΤΩΝ ΟΔΗΓΩΝ ΜΕΣΩ ΤΗΣ ΑΞΙΟΠΟΙΗΣΗΣ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΟΔΗΓΗΣΗΣ ΣΕ ΠΡΑΓΜΑΤΙΚΕΣ ΣΥΝΘΗΚΕΣ



Ελένη Μαρία Θεοδωράκη  
Επιβλέπων: Γιώργος Γιαννής, Καθηγητής Ε.Μ.Π.  
Αθήνα, Μάρτιος 2024



## Ευχαριστίες

Με την ολοκλήρωση της παρούσας Διπλωματικής Εργασίας ολοκληρώνεται ο κύκλος των προπτυχιακών σπουδών μου στη Σχολή Πολιτικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου.

Θα ήθελα πρωτίστως να ευχαριστήσω θερμά τον κ. Γιαννή, Καθηγητή της Σχολής Πολιτικών Μηχανικών ΕΜΠ, για την εμπιστοσύνη που μου έδειξε με την ανάθεση του θέματος, καθώς επίσης για την υποστήριξη και την καθοδήγηση σε όλα τα στάδια της Διπλωματικής Εργασίας.

Επίσης θα ήθελα να ευχαριστήσω θερμά τον ερευνητή Θοδωρή Γαρεφαλάκη, για την καθοριστική συνεισφορά του στη διεκπεραίωση της παρούσας εργασίας, μέσω των πολύτιμων συμβουλών και υποδείξεων του, καθώς και για το εξαιρετικό κλίμα συνεργασίας και επικοινωνίας που διαμόρφωσε. Τέλος, θα ήθελα να ευχαριστήσω τους φίλους μου και κυρίως του γονείς μου που με στήριξαν καθ' όλη τη διάρκεια των σπουδών μου.

Αθήνα, Μάρτιος 2024

Ελένη Μαρία Θεοδωράκη

# Διερεύνηση της επικίνδυνης συμπεριφοράς των οδηγών μέσω της αξιοποίησης δεδομένων οδήγησης σε πραγματικές συνθήκες

Ελένη Μαρία Θεοδωράκη

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

## Σύνοψη

Ο στόχος της παρούσας Διπλωματικής Εργασίας αποτελεί η ανάπτυξη μοντέλων ταξινόμησης για την αναγνώριση της επικίνδυνης συμπεριφοράς του οδηγού σε τρία επίπεδα ασφαλείας. Για τον σκοπό αυτό συλλέχθηκαν χρήσιμα δεδομένα σχετιζόμενα με την συμπεριφορά του οδηγού μέσω ενός νατουραλιστικού πειράματος οδήγησης σε πραγματικές συνθήκες για το Βέλγιο και την Αγγλία. Με βάση την επεξεργασία και την ανάλυση των δεδομένων καθορίστηκαν τρία επίπεδα κινδύνου. Στο πρώτο μέρος των αναλύσεων αναπτύχθηκαν τρεις συνδυαστικοί αλγόριθμοι μηχανικής και βαθιάς μάθησης με σκοπό την ταξινόμηση της συμπεριφοράς των οδηγών σε ένα από τα τρία επίπεδα ασφαλείας, με το μοντέλο 'Βαθύ Νευρωνικού Δικτύου (DNN)-Τυχαία Δάση (RF)' να σημειώνει την υψηλότερη επίδοση και για τις δύο χώρες. Στο πλαίσιο διερεύνησης της επιρροής των παραγόντων οδήγησης στην αναγνώριση της επικίνδυνης οδήγησης, προέκυψαν ως σημαντικότερες η συνολική απόσταση του ταξιδιού, η ταχύτητα, η εμφάνιση ή μη απότομης επιτάχυνσης και η εμφάνιση ή μη απότομου φρεναρίσματος. Στο δεύτερο μέρος των αναλύσεων εξετάστηκε η επιρροή των οδηγικών χαρακτηριστικών στην πρόβλεψη των συνδυασμένων μοντέλων με την ανάπτυξη του αλγορίθμου Lime (Local Interpretable Model-agnostic explanations). Από τα αποτελέσματα προέκυψε ως σημαντικότερη μεταβλητή η εμφάνιση ή μη απότομης επιτάχυνσης και η εμφάνιση ή μη απότομου φρεναρίσματος στο Βέλγιο και η συνολική απόσταση του ταξιδιού και η εμφάνιση ή μη απότομης επιτάχυνσης στην Αγγλία.

**Λέξεις Κλειδιά:** ανάλυση οδηγικής συμπεριφοράς, ταξινόμηση οδηγικής συμπεριφοράς, πρόβλεψη ατυχημάτων σε πραγματικό χρόνο, μηχανική μάθηση, βαθιά μάθηση, μοντέλα ταξινόμησης, επιλογή χαρακτηριστικών, μη ισορροπημένο σύνολο δεδομένων, μέθοδος επαναδειγματοληψίας, τυχαία δάση, μοντέλο προσαρμοστικής ενδυνάμωσης, βαθιά νευρωνικά δίκτυα, επαναλαμβανόμενο νευρωνικό δίκτυο, συνελεκτικό νευρωνικό δίκτυο, μοντέλο μακράς βραχυπρόθεσμης μνήμης, Local Interpretable Model-agnostic explanations.



# Investigation of dangerous driving behaviour by exploiting naturalistic driving data

Eleni Maria Theodoraki

Supervisor: George Yannis, Professor NTUA

## Abstract

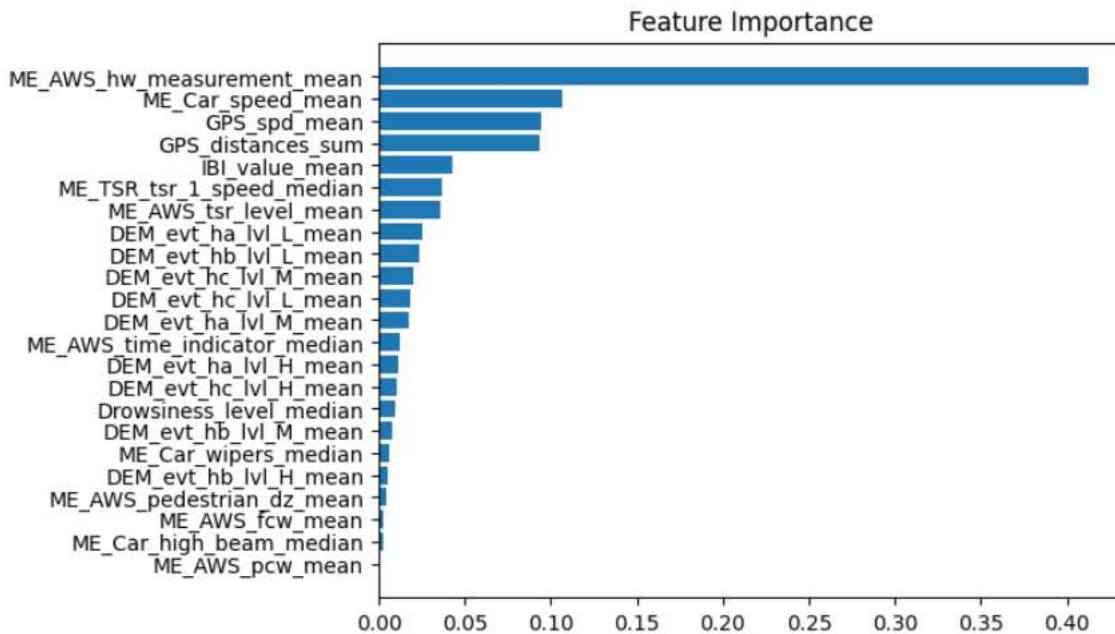
The present research aimed to develop classification models for the identification of dangerous driver behaviour at three safety levels. For this purpose, useful data related to driver behaviour were collected through a naturalistic driving experiment in real-life conditions in Belgium and England. Based on the processing and analysis of the data, three levels of risk were defined. In the first part of the analyses, three combined machine and deep learning algorithms were developed to classify driver behaviour into one of the three safety levels, with the 'Deep Neural Network (DNN)-Reflective Forest (RF) model performing the best for both countries. As part of the investigation of the influence of driving factors on the identification of dangerous driving, the most important factors were found to be total travel distance, speed, occurrence or not of sudden acceleration and occurrence or not of sudden braking. The second part of the analysis examined the influence of driving characteristics on the prediction of the combined models by developing the Lime algorithm (Local Interpretable Model-agnostic explanations). The results showed that the most significant variables were the occurrence or non-occurrence of sudden acceleration and the occurrence or non-occurrence of sudden braking in Belgium and the total travel distance and the occurrence or non-occurrence of sudden acceleration in England.

**Key words:** driving behavior analysis, driving behavior classification, real-time crash prediction, machine learning, classification models, feature selection, imbalanced dataset, resampling methods, random forests, AdaBoost, deep neural networks, recurrent neural networks, convolutional neural networks, long short-term memory networks, Local Interpretable Model-agnostic explanations.

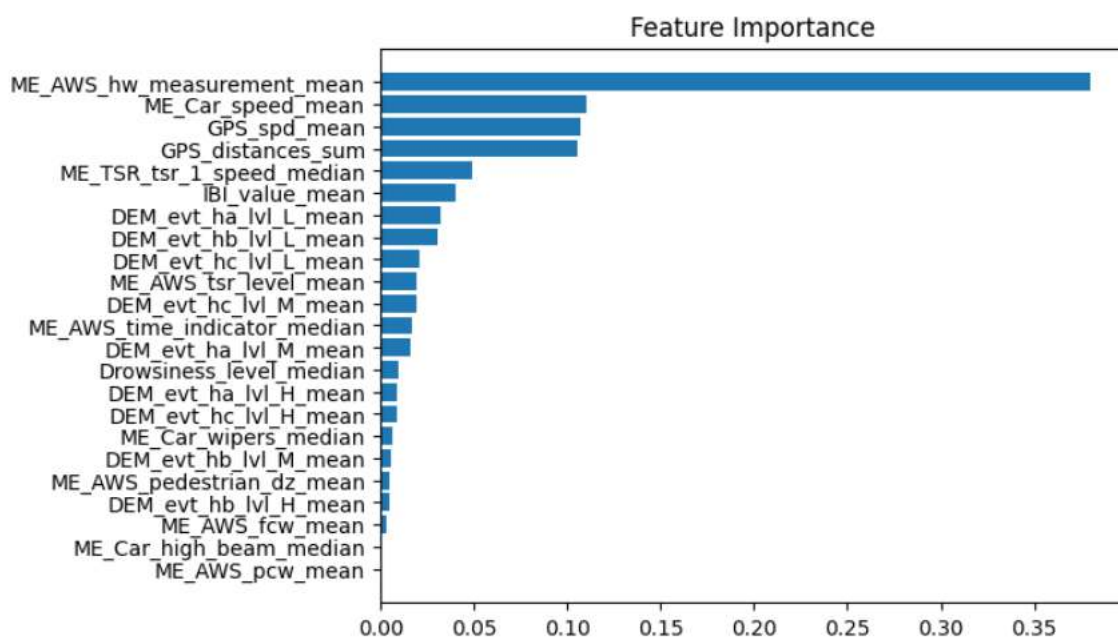
## ΠΕΡΙΛΗΨΗ

Στόχος της παρούσας Διπλωματικής Εργασίας αποτελεί η διερεύνηση της επικίνδυνης συμπεριφοράς του οδηγού (**Ζώνη Ανοχής Ασφαλείας**) με τεχνικές μηχανικής και βαθιάς μάθησης. Τα δεδομένα που αναλύθηκαν συλλέχθηκαν από πείραμα σε πραγματικές συνθήκες οδήγησης στο πλαίσιο του ερευνητικού προγράμματος i-DREAMS, στο οποίο διεξήχθη ένα νατουραλιστικό πείραμα οδήγησης στο οποίο συμμετείχαν οδηγοί από το Βέλγιο και την Αγγλία. Για την ανάλυση της οδηγικής συμπεριφοράς ήταν αναγκαίο να οριστούν διαφορετικά επίπεδα 'Ζώνης Ανοχής Ασφαλείας' βάσει ορισμένων τεχνικών. Τελικά ο καθορισμός των επιπέδων ασφαλείας πραγματοποιήθηκε με την χρήση της μεταβλητής **Headway**, καθώς το επίπεδο 'Κανονικό' να είναι η κύρια κατηγορία με το υψηλότερο ποσοστό δειγμάτων, ενώ τα επίπεδα 'Επικίνδυνο' και 'Αποφυγή Ατυχήματος' είναι η μειοψηφική κατηγορία με το χαμηλότερο ποσοστό δειγμάτων.

Στο πρώτο μέρος των αναλύσεων αναπτύχθηκαν κατάλληλες τεχνικές προσδιορισμού της **σημαντικότητας των μεταβλητών** για την πρόβλεψη του επιπέδου 'Ζώνης Ανοχής Ασφαλείας' που βρίσκεται ο οδηγός. Επισημαίνεται ότι επιλέγοντας να γίνει ο καθορισμός του επιπέδου ασφαλείας με την χρήση της μεταβλητής Headway, τα διάφορα περιγραφικά στατιστικά στοιχεία του παράγοντα Headway δεν θα αποτελέσουν μεταβλητές εισόδου στα μοντέλα. Ο συνυπολογισμός τους στην διαδικασία της ταξινόμησης θα οδηγούσε σε μεροληψία του μοντέλου χωρίς εκείνο να προσφέρει χρήσιμα και σημαντικά αποτελέσματα. Η σημαντικότητα φαίνεται στα γραφήματα που ακολουθούν.



Γράφημα 1: Σημαντικότητα μεταβλητών σύμφωνα με την μέθοδο 'Τυχαίων Δασών' για το Βέλγιο



Γράφημα 2: Σημαντικότητα μεταβλητών σύμφωνα με την μέθοδο 'Τυχαίων Δασών' για την Αγγλία

Στην συνέχεια, αξιοποιώντας τέσσερις πιο σημαντικές μεταβλητές αναπτύχθηκαν τρεις συνδυασμοί αλγορίθμων μηχανικής και βαθιάς μάθησης με σκοπό την ταξινόμηση της οδηγικής συμπεριφοράς σε ένα από τα τρία επίπεδα ασφαλείας. Εφαρμόζοντας την τεχνική **Synthetic Minority Oversampling Technique (SMOTE)** αντιμετωπίζεται το πρόβλημα της ανισορροπίας των δεδομένων εκπαίδευσης ως προς τις διαφορετικές κλάσεις.

Η ονοματολογία και ο συμβολισμός των τριών αλγορίθμων για το Βέλγιο και την Αγγλία παρατίθενται στον πίνακα ενώ οι επιδόσεις τους παρουσιάζονται στον πίνακα και στο γράφημα που ακολουθεί.

Πίνακας 1: Ονοματολογία και συμβολισμός μοντέλων ταξινόμησης

Όνομα Μοντέλου (Ελληνικά)	Όνομα Μοντέλου (Αγγλικά)	Συμβολισμός Μοντέλου
Ταξινομητής Τυχαίων Δασών	Random Forests Classifier	RF
Ταξινομητής Adaboost	Adaboost Classifier	Adaboost
Συνελεκτικά Νευρωνικά Δίκτυα	Convolutional Neural Networks	CNN
Μοντέλο Μακράς-Βραχείας Μνήμης	Long-Short Term Memory Classifier	LSTM
Μοντέλο Επαναλαμβανόμενων Νευρωνικών Δικτύων	Recurrent Neural Network	RNN



---

Μοντέλο Βαθιών Νευρωνικών  
Δικτύων

Deep Neural Network

DNN

Πίνακας 2: Σύγκριση μετρικών αξιολόγησης των μοντέλων ταξινόμησης για το Βέλγιο

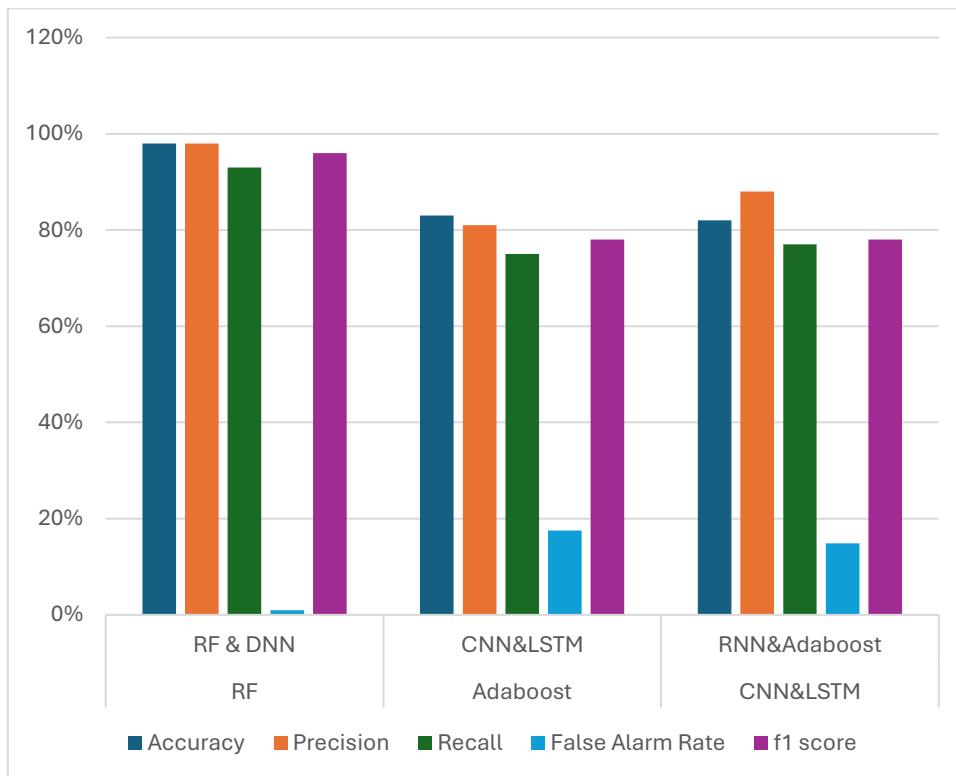
---

	<b>Ορθότητα</b>	<b>Ακρίβεια</b>	<b>Ανάκληση</b>	<b>FPR</b>	<b>f1-score</b>
RF & DNN	98%	98%	93%	0,96%	96%
CNN&LSTM	83%	81%	75%	17,5%	78%
RNN&Adaboost	82%	88%	77%	14,9%	78%

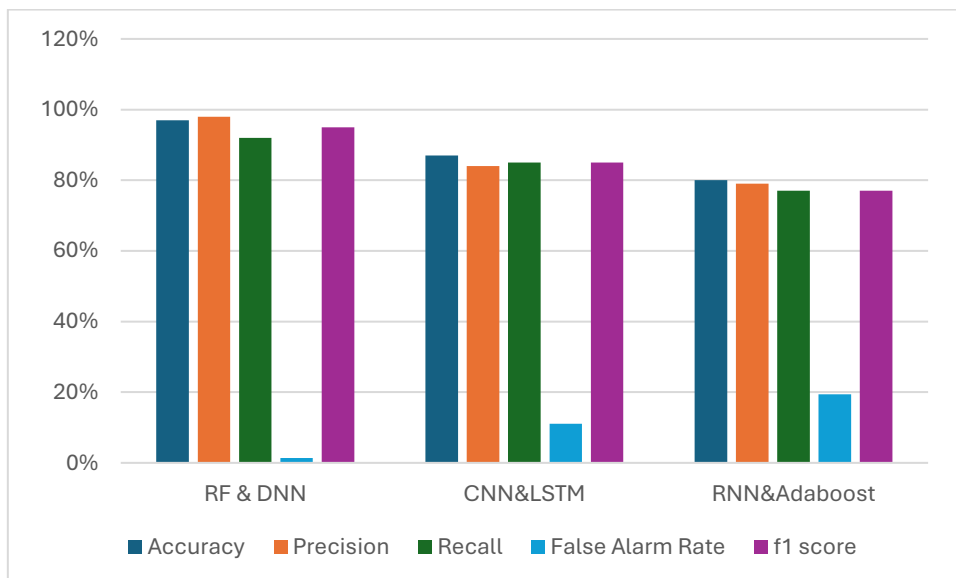
Πίνακας 3: Σύγκριση μετρικών αξιολόγησης των μοντέλων ταξινόμησης για την Αγγλία

---

	<b>Ορθότητα</b>	<b>Ακρίβεια</b>	<b>Ανάκληση</b>	<b>FPR</b>	<b>f1-score</b>
RF&DNN	97%	98%	92%	1,36%	95%
RNN&Adaboost	87%	84%	85%	11,11%	85%
CNN&LSTM	80%	79%	77%	19,4%	77%



Γράφημα 3: Επίδοση των μοντέλων ταξινόμησης σύμφωνα με τις μετρικές αξιολόγησης για το Βέλγιο



Γράφημα 4: Επίδοση των μοντέλων ταξινόμησης σύμφωνα με τις μετρικές αξιολόγησης για την Αγγλία

Στο δεύτερο μέρος των αναλύσεων για να έχουμε μία επισκόπηση των ανεξάρτητων μεταβλητών που είναι πιο σημαντικές και για να κατανοήσουμε καλύτερα τα συνδυασμένα μοντέλα θα αναπτύξουμε τον αλγόριθμο **Lime (Local Interpretable Model-agnostic explanations)**. Σύμφωνα με τα διαγράμματα του αλγόριθμου Lime μπορούμε να συμπεράνουμε τα εξής:

Δεδομένα Βελγίου:

- Τα χαρακτηριστικά που επηρεάζουν περισσότερο την τελική πρόβλεψη είναι οι μεταβλητές DEM\_evt\_ha\_lvl\_L\_mean και η DEM\_evt\_hB\_lvl\_L\_mean.

Δεδομένα Αγγλίας:

- Τα χαρακτηριστικά που επηρεάζουν περισσότερο την τελική πρόβλεψη είναι οι μεταβλητές DEM\_evt\_ha\_lvl\_L\_mean και GPS\_distances\_sum.

Βάσει των αποτελεσμάτων που προέκυψαν κατά την εφαρμογή της μεθοδολογίας, προέκυψαν ορισμένα **συμπεράσματα** άμεσα σχετιζόμενα με τον στόχο της Δπλωματικής Εργασίας.

- Σύμφωνα με τα αποτελέσματα των γραφημάτων 6.1 και 6.2 για το Βέλγιο και την Αγγλία αντίστοιχα, συμπεραίνουμε ότι η χρονική απόσταση του οχήματος από το προπορευόμενο όχημα είναι η σημαντικότερη μεταβλητή για την αναγνώριση της οδηγικής συμπεριφοράς. Καθώς όμως τα επίπεδα 'Ζώνης Ανοχής Ασφαλείας' έχουν προκύψει από την μεταβλητή Headway, η μεταβλητή αυτή δεν θα αποτελέσει μεταβλητή εισόδου για τα μοντέλα ταξινόμησης. Ο συνυπολογισμός τους στην διαδικασία της ταξινόμησης θα οδηγούσε σε μεροληψία του μοντέλου χωρίς εκείνο να προσφέρει χρήσιμα και σημαντικά αποτελέσματα.
- Η ταχύτητα του οχήματος έχει επίσης μεγάλο δείκτη σημαντικότητας. Η ταχύτητα σχετίζεται άμεσα με την πιθανότητα εμφάνισης ατυχήματος καθώς επίσης και με την σοβαρότητα αυτού. Για παράδειγμα, η αυξημένη ταχύτητα του οχήματος ελαχιστοποιεί τον χρόνο αντίδρασης του οδηγού.
- Η συνολική διανυθείσα απόσταση αποτελεί επίσης σημαντική μεταβλητή για την αναγνώριση της οδηγικής συμπεριφοράς. Ανάλογα με την συνολική απόσταση που διανύει ο οδηγός μπορεί να παρατηρηθούν διαφορετικές οδηγικές συμπεριφορές. Για παράδειγμα, οι οδηγοί που διανύουν μεγάλες αποστάσεις είναι πιθανό να εμφανίσουν σημάδια κούρασης και μειωμένης προσοχής, τα οποία οδηγούν σε επικίνδυνη οδηγική συμπεριφορά.
- Η μεταβλητή DEM\_evt\_ha\_lvl\_L\_mean, η οποία αφορά την εμφάνιση ή μη απότομων επιταχύνσεων, επηρεάζει επίσης την διαδικασία ταξινόμησης στα μοντέλα. Για παράδειγμα, οι απότομες εναλλαγές της επιτάχυνσης των οχημάτων ελαχιστοποιούν το χρόνο αντίδρασης του οδηγού και αποσπών την προσοχή του οδηγού από το οδικό περιβάλλον, με αποτέλεσμα να οδηγείται σε επικίνδυνες συμπεριφορές.
- Από τις διαφορετικές τεχνικές αντιμετώπισης του φαινομένου της άνισης κατανομής των δειγμάτων στις διαφορετικές κλάσεις, η τεχνική της συνθετικής μειονοτικής υπερδειγματοληψίας (Synthetic minority oversampling technique, SMOTE) πρόσφερε τα βέλτιστα αποτελέσματα για το σύνολο των ταξινομητών. Όπως αναφέρθηκε στο κεφάλαιο 3.3, η τεχνική SMOTE επικεντρώνεται στον «χώρο των χαρακτηριστικών» και όχι στον «χώρο δεδομένων», με άλλα λόγια, ο αλγόριθμος βασίζεται στις τιμές των

χαρακτηριστικών και στη σχέση τους, αντί να θεωρεί τα σημεία δεδομένων ως ολόκληρος.

- Στην παρούσα εργασία αναπτύχθηκαν τρεις συνδυασμοί μοντέλων μηχανικής και βαθιάς μάθησης οι οποίοι σημείωσαν ικανοποιητικές επιδόσεις. Ο συνδυασμός 'Βαθύ Νευρωνικού Δικτύου (Deep Neural Network) - Τυχαίων Δασών (Random Forests)' σημείωσε τις υψηλότερες επιδόσεις στην πλειοψηφία των μετρικών αξιολόγησης τους.
- Από την αξιολόγηση των μετρικών αξιολόγησης για κάθε χώρα, συμπεραίνεται ότι όλα τα παραπάνω μοντέλα ταξινόμησης μπορούν να αποφέρουν επιθυμητά αποτελέσματα και για τις δύο χώρες.
- Από το σύνολο των μεταβλητών που εξετάστηκαν οι μεταβλητές DEM\_evt\_ha\_lvl\_L\_mean, DEM\_evt\_hB\_lvl\_L\_mean και DEM\_evt\_ha\_lvl\_L\_mean, GPS\_distances\_sum για το Βέλγιο και την Αγγλία αντίστοιχα, είχαν τη μεγαλύτερη επιρροή στην τελική πρόβλεψη, σύμφωνα με την ανάπτυξη του αλγόριθμου Lime.
- Η υψηλή επίδραση των μεταβλητών, DEM\_evt\_ha\_lvl\_L\_mean και DEM\_evt\_hB\_lvl\_L\_mean, που σχετίζονται με τα απότομα περιστατικά οδήγησης υποδεικνύει ότι η οδηγική συμπεριφορά των οδηγών στο Βέλγιο επηρεάζεται κυρίως από την αποφυγή απότομων περιστατικών οδήγησης και είναι ευαίσθητη σε αυτά τα περιστατικά.
- Η υψηλή επίδραση της μεταβλητής DEM\_evt\_ha\_lvl\_L\_mean για την Αγγλία υποδεικνύει ότι η αποφυγή της απότομης οδήγησης είναι σημαντική και στις δύο χώρες. Επιπλέον, η υψηλή επιρροή της GPS\_distances\_sum υποδεικνύει ότι οι αποστάσεις που καλύπτονται έχουν επίσης σημαντική συνεισφορά στον τρόπο οδήγησης. Επομένως, οι οδηγοί στην Αγγλία επηρεάζονται από την απόσταση που διανύουν σε μεγαλύτερο βαθμό.
- Από την εκπόνηση της συγκεκριμένης Διπλωματικής Εργασίας προκύπτει ότι τα δεδομένα που συλλέγονται από τα έξυπνα συστήματα, τα συστήματα αιχμής και περαιτέρω εύρυνες παρέχουν ιδιαίτερα σημαντικές πληροφορίες οι οποίες, με την κατάλληλη επεξεργασία και ανάπτυξη μαθηματικών μοντέλων, μπορούν να ωφελήσουν στην εξαγωγή χρήσιμων συμπερασμάτων για τις κρίσιμες παραμέτρους που επηρεάζουν την συμπεριφορά του οδηγού κατά τη διάρκεια οδήγησης αλλά και τη κυκλοφοριακή συμπεριφορά των οδηγών.
- Συνολικά συμπεραίνεται ότι η ασφάλεια και η αποφυγή απότομης οδήγησης είναι κοινή και για τις δύο χώρες και οι διαφορές στις επιρροές μπορεί να υποδεικνύουν ότι οι κοινωνιο-οικονομικές, πολιτιστικές ή γεωγραφικές διαφορές επηρεάζουν την οδική συμπεριφορά σε κάθε χώρα. Συνοψίζοντας, οι οδηγοί στο Βέλγιο επικεντρώνονται ιδιαίτερα στην αποφυγή απότομης οδήγησης, ενώ στην Αγγλία η ασφάλεια συνδέεται επίσης και με τις αποστάσεις που διανύουν.
- Από τα αποτελέσματα των αναλύσεων, συμπεραίνεται ότι οι δύο χώρες έχουν παρόμοια οδηγικά χαρακτηριστικά. Ωστόσο, παρατηρείται μία μικρή αύξηση στη σημαντικότητα των μεταβλητών ME\_car\_speed\_mean, GPS\_spd\_mean και GPS\_distances\_sum για την Αγγλία, που υποδεικνύει ότι επικρατούν υψηλότερες ταχύτητες και αποστάσεις ταξιδιού.

## ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΕΡΙΛΗΨΗ.....	6
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ .....	<b>Error! Bookmark not defined.</b>
1. ΕΙΣΑΓΩΓΗ.....	1
1.1 Γενική ανασκόπηση.....	1
1.2 Στόχος.....	3
1.3 Μεθοδολογία.....	4
1.4 Δομή Διπλωματική Εργασίας .....	6
2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ.....	8
2.1 Εισαγωγή.....	8
2.2 Συναφείς έρευνες και μεθοδολογίες.....	8
2.2.1 Ανάλυση οδηγικής συμπεριφοράς.....	9
2.2.2 Αλγόριθμοι ταξινόμησης οδηγικής συμπεριφοράς.....	14
2.2.3 Τεχνικές βελτίωσης της εκπαίδευσης και της γενίκευσης μοντέλων .....	16
2.3 Αξιολόγηση επίδρασης μεταβλητών στην επικίνδυνη οδήγηση.....	17
2.4 Σύνοψη.....	17
3. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ .....	18
3.1 Εισαγωγή.....	18
3.2 Επιλογή χαρακτηριστικών (Feature Selection) .....	19
3.3 Μέθοδοι βελτίωσης και γενίκευσης μοντέλων.....	20
3.3.1 Τεχνική Συνθετικής Μειονοτικής Υπερδειγματοληψίας (SMOTE) .....	20
3.3.2 Τεχνική Dropout .....	21
3.4 Αλγόριθμοι Ταξινόμησης.....	22
3.4.1 Μοντέλα Μηχανικής Μάθησης.....	23
3.4.1.1 Τυχαία Δάση (Random Forest).....	24
3.4.1.2 Προσαρμοστική ενδυνάμωση (Adaboost) .....	25
3.4.2 Μοντέλα Βαθιάς Μάθησης.....	25
3.4.2.1 Συνελεκτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks) .....	26
3.4.2.2 Επαναλαμβανόμενο Νευρωνικό Δίκτυο (Recurrent Neural Network – RNN) .....	27
3.5 Τοπικό Ερμηνευτικό Μοντέλο – Lime (Local Interpretable Model-agnostic explanations) .....	29
3.6 Μετρικές αξιολόγησης για ταξινόμηση (Evaluation metrics for classification) .....	30
3.6.1 Μήτρα σύγχυσης (Confusion Matrix).....	30
3.6.2 Ορθότητα (Accuracy) .....	31
3.6.3 Ακρίβεια (Precision).....	32

3.6.4 Ανάκληση (Recall).....	32
3.6.5 Ρυθμός λανθασμένων θετικών προβλέψεων (False positive rate).....	32
3.6.6 f1-score .....	32
3.6.7 Εξειδικευτικότητα (Specificity) .....	32
3.7.8 Δείκτης λάθος συναγερμού (False alarm rate) .....	32
3.6.9 Μακροοικονομικός μέσος όρος (MacroAverage).....	33
3.6.10 Σταθμισμένος μέσος όρος (WeightedAverage) .....	33
4. ΣΥΛΛΟΓΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΣΤΟΙΧΕΙΩΝ .....	33
4.1 Εισαγωγή.....	33
4.2 Πείραμα σε πραγματικές συνθήκες οδήγησης .....	33
4.2.1 Στόχος πειράματος.....	33
4.2.2 ECU & On-Board Diagnostics (OBD II) .....	35
4.2.3 Σύστημα Πρόληψης Ατυχημάτων MobilEye (G4S).....	36
4.2.4 Στοιχεία που συλλέχθηκαν από το πείραμα .....	37
4.3 Επεξεργασία στοιχείων.....	39
4.4 Περιγραφική στατιστική δεδομένων .....	42
4.5 Συσχέτιση μεταβλητών .....	44
4.6 Σύνοψη.....	46
5. ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΟΛΟΓΙΑΣ - ΑΠΟΤΕΛΕΣΜΑΤΑ .....	46
5.1 Εισαγωγή.....	46
5.2 Εντοπισμός του Επιπέδου ‘Ζώνης Ανοχής Ασφαλείας’ .....	47
5.2.1 Καθορισμός Επιπέδων Ασφαλείας.....	47
5.2.2 Επιλογή Χαρακτηριστικών (Feature Selection) .....	49
5.2.3 Προετοιμασία δεδομένων .....	53
5.2.4 Αντιμετώπιση άνισης κατανομής δεδομένων στις κλάσεις .....	54
5.2.5 Ανάπτυξη μοντέλων ταξινόμησης.....	55
5.2.6 Σύγκριση Μετρικών Αξιολόγησης των μοντέλων .....	68
5.3 Εξήγηση λειτουργίας μοντέλων μηχανικής και βαθιάς μάθησης .....	70
5.4 Σύνοψη.....	71
6. ΣΥΜΠΕΡΑΣΜΑΤΑ .....	73
6.1 Σύνοψη αποτελεσμάτων.....	73
6.2 Σύνοψη Συμπερασμάτων .....	77
6.3 Προτάσεις για την αξιοποίηση των αποτελεσμάτων .....	79
6.4 Προτάσεις για περαιτέρω έρευνα.....	80



# 1. ΕΙΣΑΓΩΓΗ

## 1.1 Γενική ανασκόπηση

Στη σύγχρονη εποχή, οι οδικές μεταφορές αποτελούν ένα από τους σημαντικότερους τομείς για την εύρυθμη λειτουργία και ανάπτυξη της κοινωνίας. Η απότομη τεχνολογική ανάπτυξη κατά τον 20ο αιώνα και η οικονομική ευμάρεια των προηγούμενων ετών, σε συνδυασμό με τη ραγδαία εξέλιξη της αυτοκινητοβιομηχανίας τα τελευταία 30 χρόνια έχουν συντελέσει στην κατακόρυφη αύξηση του δείκτη ιδιοκτησίας Ι.Χ. οχημάτων, που δημιούργησε μια πρωτοφανή αύξηση των οδικών μεταφορών. Το αναμφίβολο μεγάλο όφελος αυτής της τεχνολογικής εξέλιξης για τον άνθρωπο με τη χρήση των οχημάτων στη ζωή του επιφέρει και το οδυνηρότερο τίμημα για τον ίδιο μέσω των οδικών ατυχημάτων.

Όπως δείχνουν οι περισσότερες έρευνες, η οδηγική συμπεριφορά αποτελεί τον κύριο παράγοντα για ποσοστό 95% των συνολικών οδικών ατυχημάτων. Γενικά, η παράβαση των κανόνων κυκλοφορίας, το οδηγικό λάθος ή αντίδραση, η μέθη, η κούραση, η επιθετική συμπεριφορά, η απειρία και η μη συμμόρφωση με την σήμανση και την σηματοδότηση αποτελούν βασικούς παράγοντες που καθιστούν την οδηγική στην κορυφή της λίστας των οδικών ατυχημάτων.

C

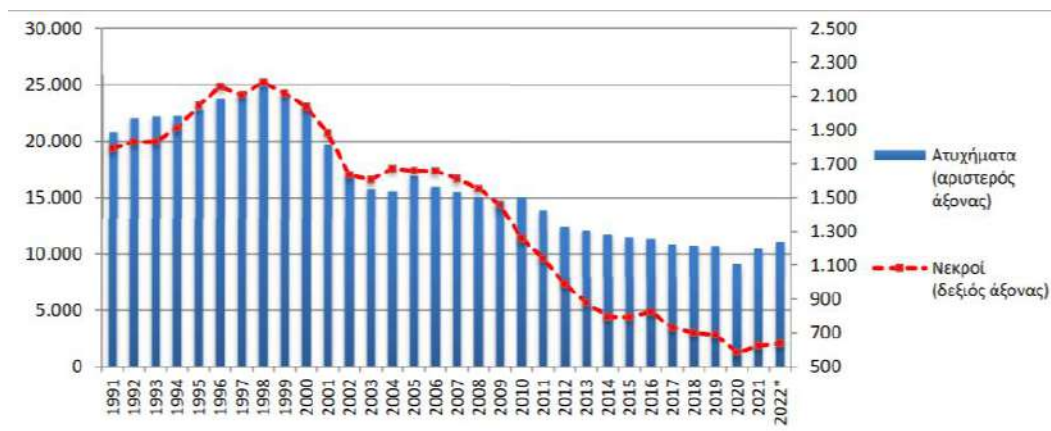


Γράφημα 1.1: Βασικά στοιχεία οδικής ασφάλειας στην Ελλάδα, 2011-2020  
Πηγή: NTUA Road Safety Observatory (2022)

Κατά την διάρκεια των 10 τελευταίων ετών η Ελλάδα παρουσίασε αξιοσημείωτη βελτίωση της οδικής ασφάλειας στην Ευρωπαϊκή Ένωση, με μείωση 28% των θανάτων από τροχαία ατυχήματα από το 2013 και μείωση των σοβαρών τραυματισμών κατά 48%. Το ποσοστό θανάτου ανά εκατομμύριο οχήματα έχει μειωθεί κατά 33% από το 2013.

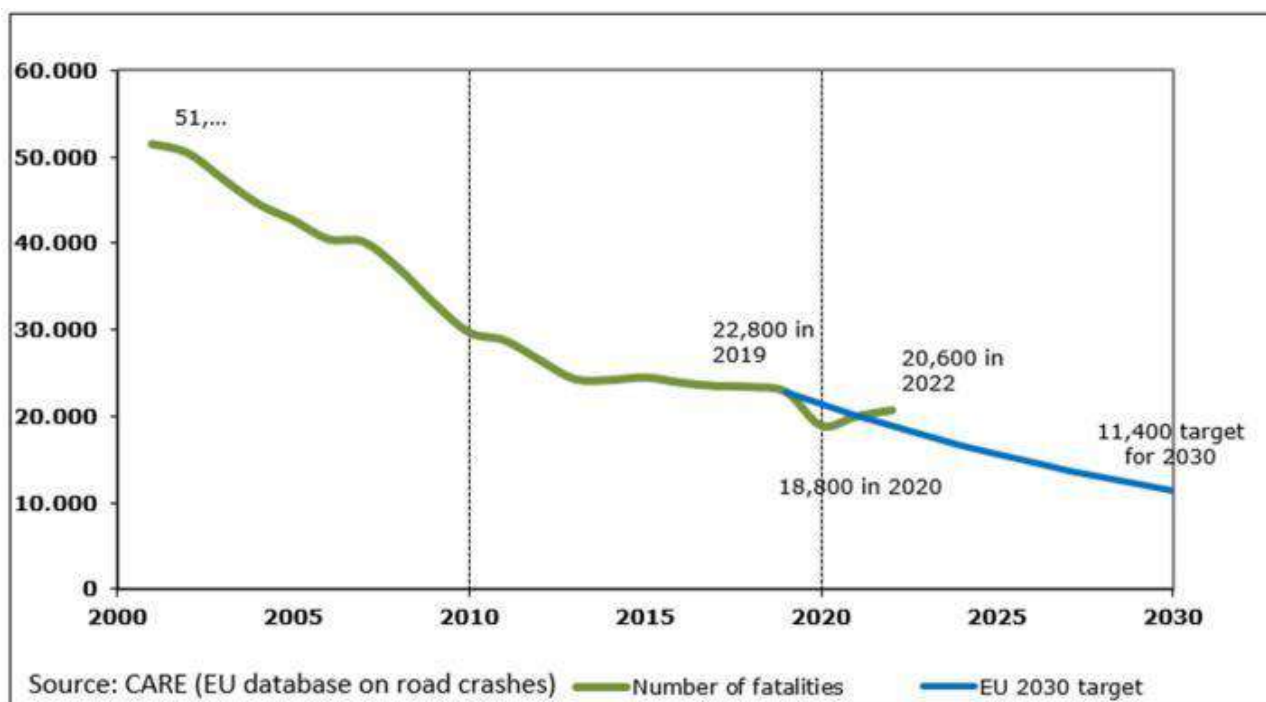


Παρόλα αυτά, έχει παρατηρηθεί μία μικρή αύξηση των θανατηφόρων ατυχημάτων στην Ελλάδα το 2022 σε σχέση με τα στοιχεία το 2021, σύμφωνα με τα προσωρινά στοιχεία της ΕΛΣΤΑΤ. Αυτή η μικρή αύξηση αποδίδεται κυρίως στην αύξηση της κυκλοφορίας μετά την περίοδο της πανδημίας.



Γράφημα 1.2: Σύνολο οδικών ατυχημάτων και νεκρών, 1991-2022  
 Πηγή: ΕΛΣΤΑΤ (2022)

Στο γράφημα 1.3 που ακολουθεί διακρίνεται η σημαντική μείωση των θανατηφόρων οδικών ατυχημάτων στην ΕΕ καθώς και η απόκλιση από τον στόχο του έτους 2020.



Γράφημα 1.3: Στατιστικά Θανατηφόρων ατυχημάτων 2022  
 Πηγή: CARE (EU road accidents database)

Η «Ευρωπαϊκή Ένωση (Ε.Ε)» σε συνεργασία με τον Παγκόσμιο Οργανισμό Υγείας και τα Ηνωμένα Έθνη έχουν θέσει ως στόχο την μείωση των θανατηφόρων ατυχημάτων στην Ε.Ε. κατά 50%. Για την πραγματοποίηση του παραπάνω στόχου δίνεται ιδιαίτερη βάση στην συνεισφορά καινοτόμων τεχνολογιών στον τομέα της αυτοκινητοβιομηχανίας και της αυτοματοποίησης στις μεταφορές, με στόχο τη βελτίωση της οδικής ασφάλειας.

Η κατάσταση των χρηστών της οδού, οι περιβαλλοντικές συνθήκες και οι κυκλοφοριακές συνθήκες μπορούν να αποτελέσουν σημαντικούς παράγοντες πρόκλησης ατυχημάτων (Aljanahi et al., 1999). Ωστόσο, η συμπεριφορά του ανθρώπου ως χρήστη της οδού, αποτελεί το σημαντικότερο παράγοντα για τα οδικά ατυχήματα. Η ραγδαία ανάπτυξη συστημάτων που μας βάζουν στον προθάλαμο της αυτόνομης οδήγησης έχει ως στόχο να περιορίσει την επίδραση του επιρρεπή σε λάθη ανθρώπινου παράγοντα στη συνολική οδική ασφάλεια.

Τα τελευταία χρόνια, παρουσιάζεται μεγάλο ενδιαφέρον στην αξιοποίηση αλγόριθμων μηχανικής εκμάθησης και βαθιάς μάθησης για τη ανάλυση της συμπεριφοράς των οδηγών (Peppes et al., 2021). Επιπρόσθετα, έχει διαπιστωθεί ότι η χρήση έξυπνων συστημάτων παρακολούθησης της συμπεριφοράς του οδηγού με σκοπό τις παρεμβάσεις σε πραγματικό χρόνο, έχουν αποδειχθεί ιδιαίτερα αποτελεσματική στην ελαχιστοποίηση των ατυχημάτων (Michelaraki et al., 2021b). Η αναγκαιότητα ανάπτυξης παρεμφερών συστημάτων με σκοπό την εξασφάλιση της οδικής ασφάλειας καθιστά αναγκαίο τον καθορισμό της επιρροής των διαφορετικών παραγόντων κίνδυνου κατά την οδήγηση.

Συνεπώς, ο εντοπισμός της επικίνδυνης συμπεριφοράς των οδηγών και των παραγόντων που επιδρούν σε αυτήν θα αποτελέσει κύριο αντικείμενο έρευνας στην παρούσα μελέτη.

## 1.2 Στόχος

Στόχος της παρούσας διπλωματικής εργασίας αποτελεί η διερεύνηση της **επικίνδυνης συμπεριφοράς των οδηγών** μέσω της αξιοποίησης δεδομένων οδήγησης σε πραγματικές συνθήκες. Πιο συγκεκριμένα, η διερεύνηση της επικίνδυνης συμπεριφοράς του οδηγού θα πραγματοποιηθεί σε τρία επίπεδα ασφαλείας με έμφαση στη Ζώνη Ανοχής Ασφαλείας καθώς και ο εντοπισμός των χαρακτηριστικών που επιδρούν στη διαδικασία αυτή.

Οι αναλύσεις, βασίζονται σε δύο διαφορετικά σύνολα δεδομένων οδήγησης σε πραγματικές συνθήκες, από δύο διαφορετικές χώρες (Ηνωμένο Βασίλειο και Βέλγιο). Το παραπάνω θα επιτευχθούν μέσω:

- 1) Της ανάπτυξης **μοντέλων ταξινόμησης**, με σκοπό τον προσδιορισμό του επιπέδου της 'Ζώνης Ανοχής Ασφαλείας' που βρίσκεται κάθε οδηγός. Συγκεκριμένα κάθε οδηγός ταξινομείται σε ένα από τα τρία επίπεδα της 'Ζώνης Ανοχής Ασφαλείας' μέσω συνδυασμένων αλγόριθμων μηχανικής εκμάθησης και βαθιάς μάθησης οι οποίοι έχουν αναπτυχθεί, εκπαιδευτεί και αξιολογηθεί καταλλήλως. Αυτό επιτυγχάνεται λαμβάνοντας ως δεδομένα εισόδου τα χαρακτηριστικά οδήγησης του κάθε οδηγού καθώς και του αντίστοιχου περιβάλλοντος οδήγησης. Η ταξινόμηση μέσω της μηχανικής μάθησης αποτελεί σημαντικό εργαλείο για την αναγνώριση της οδηγικής συμπεριφοράς και κατ' επέκταση τη βελτίωση της οδηγικής ασφάλειας (Meiring and Myburgh, 2015; Wu et al., 2016).
- 2) Της ανάπτυξης της **μεθόδου Lime**, η οποία χρησιμοποιείται ως υποστηρικτικό εργαλείο για την ανάλυση της επιρροής των ανεξάρτητων μεταβλητών στην πρόβλεψη των συνδυασμένων μοντέλων, με σκοπό να εξεταστούν και να αξιολογηθούν τα χαρακτηριστικά που επιδρούν στην διάρκεια οδήγησης κάτω από επικίνδυνες συνθήκες.

Χρήσιμα συμπεράσματα θα εξαχθούν για τους παράγοντες που επιδρούν στον προσδιορισμό της επικίνδυνης οδήγησης μέσω του εντοπισμού του επιπέδου της 'Ζώνης Ανοχής Ασφαλείας'. Επιπρόσθετα, εναλλακτικό δείκτη κινδύνου θα αποτελέσει η ανάλυση της διάρκειας οδήγησης σε επικίνδυνες συνθήκες και η ανάλυση του θα ενισχύσει τα συμπεράσματα σχετικά με την επιρροή των παραγόντων οδήγησης.

Η παρούσα Διπλωματική Εργασία επιχειρεί να προσφέρει επιπλέον γνώση και να διευρύνει την υπάρχουσα στον τομέα της ανάλυσης οδηγικής συμπεριφοράς και ανάπτυξης αυτόματων συστημάτων οδήγησης.

### 1.3 Μεθοδολογία

Στην παρούσα ενότητα περιγράφεται συνοπτικά η μεθοδολογία που εφαρμόστηκε για την επίτευξη του στόχου της Διπλωματικής Εργασίας.

Πρώτο βήμα αποτελεί η **οριστικοποίηση του θέματος της μελέτης και ο στόχος** της. Στην συνέχεια, πραγματοποιήθηκε η βιβλιογραφική ανασκόπηση κατά την οποία αναζητήθηκαν στη διεθνή βιβλιογραφία συναφείς με το θέμα της παρούσας Διπλωματικής Εργασίας, έρευνες, μελέτες και μεθοδολογίες ανάλυσης που αξιοποιήθηκαν.

Σε επόμενο στάδιο πραγματοποιήθηκε η **συλλογή και η επεξεργασία των στοιχείων**. Τα στοιχεία που συλλέχθηκαν παράχθηκαν από πείραμα σε προσομοιωτή οδήγησης στο πλαίσιο του ερευνητικού έργου i-DREAMS και αφορούσαν στα χαρακτηριστικά οδήγησης 250 οδηγών από το Βέλγιο και το Ηνωμένο Βασίλειο καθώς και του αντίστοιχου περιβάλλοντος οδήγησης. Με την κατάλληλη επεξεργασία τα δεδομένα προετοιμάστηκαν για την ανάλυση τους.

Το ερευνητικό έργο i-DREAMS με συνολική διάρκεια 4 ετών (2019-2023), χρηματοδοτήθηκε μέσω του προγράμματος έρευνας και καινοτομίας στις μεταφορές της Ευρωπαϊκής Ένωσης, Horizon 2020 (<https://idreamsproject.eu/>). Σκοπός του i-DREAMS ήταν η δημιουργία ενός πλαισίου εργασίας για τον καθορισμό, την ανάπτυξη, τη δοκιμή και την επικύρωση μίας «Ζώνης Ανοχής Ασφαλείας» (Safety Tolerance Zone - STZ), που θα περιλαμβάνει διαφορετικά επίπεδα ασφαλείας. Μέσω ενός έξυπνου συστήματος παρακολούθησης που λαμβάνει υπόψη τους παράγοντες του ιστορικού του οδηγού και τους σχετικούς δείκτες κινδύνου σε πραγματικό χρόνο με τις επιδόσεις οδήγησης καθώς και τους δείκτες πολυπλοκότητας της κατάστασης του οδηγού και των οδηγικών καθηκόντων, είναι δυνατή η αναγνώριση του επιπέδου που βρίσκεται ο κάθε οδηγός και η ανάπτυξη παρεμβάσεων προκειμένου ο οδηγός να μην παρεκκλίνει από την ασφαλή οδήγηση.

Οι παρεμβάσεις σχεδιάστηκαν με σκοπό να πραγματοποιούνται σε δύο φάσεις. Η πρώτη φάση αφορά σε πραγματικό χρόνο δηλαδή κατά την διαδικασία της οδήγησης προκειμένου ο οδηγός να προβεί άμεσα στις απαιτούμενες ενέργειες. Στην συνέχεια, η δεύτερη φάση εφαρμόζεται σε χρόνο μετά την διαδρομή, αποσκοπώντας στην διεύρυνση της γνώσης και κατ' επέκταση την βελτίωση της συμπεριφοράς του οδηγού.

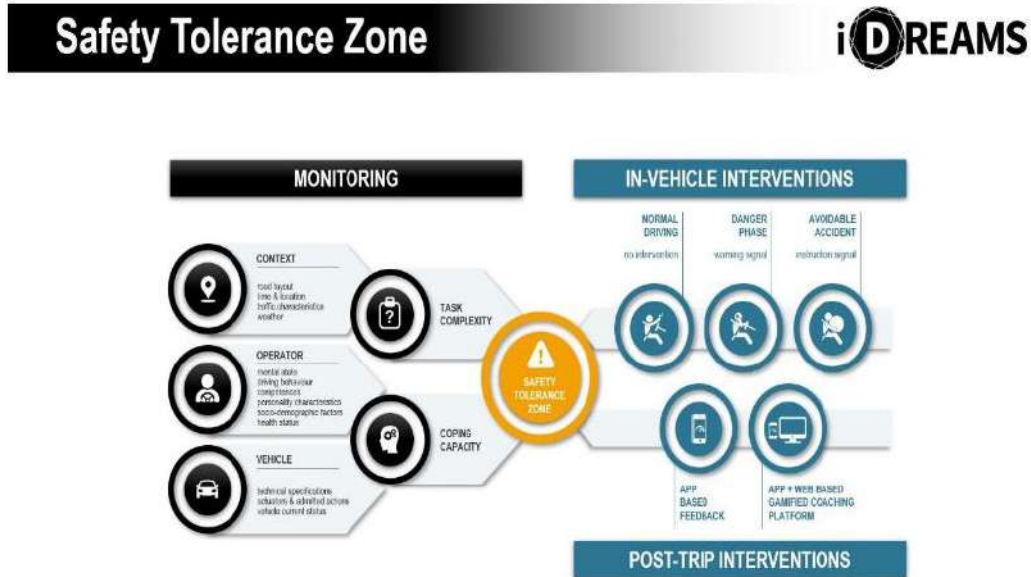
Η «Ζώνη Ανοχής Ασφάλειας» περιλαμβάνει τρία επίπεδα:

1) Κανονικό – Ασφαλές (Normal)

2) Επικίνδυνο (Dangerous)

3) Αποφυγής Ατυχήματος (Avoidable Accident)

Στο πλαίσιο του προγράμματος i-DREAMS, διεξήχθη ένα πείραμα σε πραγματικές συνθήκες οδήγησης με την συμμετοχή 250 οδηγών από το Βέλγιο και το Ηνωμένο Βασίλειο.



Γράφημα 1.4: Μεθοδολογία ερευνητικού έργου i-DREAMS  
Πηγή: i-DREAMS (2023)

Έχοντας πραγματοποιήσει την συλλογή και την επεξεργασία, ακολούθησε η ανάπτυξη των κατάλληλων μοντέλων μηχανικής εκμάθησης και βαθιάς μηχανικής μάθησης με σκοπό την ταξινόμηση. Η επεξεργασία, η ανάπτυξη των μοντέλων και οι αναλύσεις έγιναν με χρήση της γλώσσας προγραμματισμού Python αξιοποιώντας τις βιβλιοθήκες μηχανικής μάθησης TensorFlow, scikit-learn, την βιβλιοθήκη ανάλυσης δεδομένων pandas καθώς και το πακέτο ερμηνείας των μοντέλων Local Interpretable Model-Agnostic Explanations (LIME).

Τέλος, αξιολογήθηκαν τα αποτελέσματα με την εξαγωγή χρήσιμων συμπερασμάτων και προτάσεων για περαιτέρω έρευνα.

Παρακάτω παρουσιάζονται υπό την μορφή διαγράμματος ροής (γράφημα 1.5) τα διαδοχικά στάδια που ακολουθήθηκαν για την εκπόνηση της παρούσας διπλωματικής εργασίας.



Γράφημα 1.5: Διάγραμμα ροής- Μεθοδολογία διπλωματικής εργασίας

#### 1.4 Δομή Διπλωματική Εργασίας

Στην παρούσα ενότητα παρουσιάζεται η δομή της διπλωματικής εργασίας μέσω της συνοπτικής περιγραφής του περιεχομένου κάθε κεφαλαίου.

Το **Κεφάλαιο 1** αποτελεί την εισαγωγή και την ανάδειξη του στόχου της διπλωματικής εργασίας. Αρχικά με την γενική ανασκόπηση παρουσιάζεται το πλαίσιο της διπλωματικής εργασίας που αφορά στην σοβαρή επιρροή των οδικών ατυχημάτων στην σύγχρονη κοινωνία. Παρατίθενται στατιστικά στοιχεία για την οδική ασφάλεια στην Ευρώπη και την Ελλάδα και γίνεται αναφορά στην συνεισφορά των σύγχρονων τεχνολογιών στην μείωση των θανατηφόρων οδικών ατυχημάτων με έμφαση στο ερευνητικό έργο i-DREAMS. Τέλος, περιγράφεται ο στόχος, η μεθοδολογία που ακολουθήθηκε για την επίτευξη του και η δομή της διπλωματικής εργασίας.

Το **Κεφάλαιο 2**, περιλαμβάνει την βιβλιογραφική ανασκόπηση στην οποία παρουσιάζονται συναφείς έρευνες τόσο με το αντικείμενο της διπλωματικής εργασίας όσο και με τις μεθοδολογίες που αξιοποιήθηκαν. Οι έρευνες προέρχονται από την Ελληνική και την Διεθνή Επιστημονική κοινότητα.

Στο **Κεφάλαιο 3**, γίνεται αναφορά στο θεωρητικό υπόβαθρο της έρευνας. Αρχικά αναλύονται οι τεχνικές επεξεργασίας των δεδομένων και δίνεται ιδιαίτερη έμφαση στην αναγκαιότητα αυτού του βήματος για την ανάπτυξη των μοντέλων. Στην συνέχεια παρουσιάζονται, οι διαφορετικοί αλγόριθμοι μηχανικής και βαθιάς μηχανικής μάθησης που αναπτύχθηκαν για την ταξινόμηση και περιγράφονται οι μετρικές αξιολόγησης των μοντέλων.

Στο **Κεφάλαιο 4**, περιγράφονται τα δεδομένα και η διαδικασία συλλογής τους από νατουραλιστικά πειράματα σε πραγματικές συνθήκες οδήγησης (i-DREAMS). Στη συνέχεια, αναλύεται η διαδικασία και τα βήματα της επεξεργασίας των οδηγικών και περιβαλλοντικών χαρακτηριστικών προκειμένου να προετοιμαστούν για την περαιτέρω ανάλυση.

Το **Κεφάλαιο 5**, αποτελεί την κύρια ενότητα της διπλωματικής εργασίας καθώς περιλαμβάνει την αναλυτική παρουσίαση της μεθοδολογίας ανάπτυξης των συνδυασμένων μοντέλων. Η συγκεκριμένη υπο-ενότητα χωρίζεται σε δύο τομείς, την ταξινόμηση και την παλινδρόμηση. Αρχικά επεξηγούνται τα βήματα που ακολουθήθηκαν για την εφαρμογή της μεθοδολογίας, αναλύεται η διαδικασία ανάπτυξης των μοντέλων μηχανικής και βαθιάς μάθησης και περιγράφονται τα δεδομένα εισόδου και εξόδου. Τέλος, παρουσιάζονται τα συνολικά αποτελέσματα της ανάλυσης συγκρίνοντας και περιγράφοντας τα διαφορετικά μοντέλα συνοδευόμενα από τις πολλαπλές μετρικές αξιολόγησης.

Το **Κεφάλαιο 6** περιλαμβάνει τα συμπεράσματα που προέκυψαν από τα τελικά αποτελέσματα του προηγούμενου κεφαλαίου. Στο τέλος παρουσιάζονται προτάσεις που μπορούν να συνδράμουν στην περαιτέρω έρευνα, η οποία αφορά στην αξιοποίηση είτε διαφορετικών μεθόδων, είτε διαφορετικών δεδομένων.

Στο **Κεφάλαιο 7** παρατίθενται οι βιβλιογραφικές αναφορές, οι οποίες αξιοποιήθηκαν για την εκπόνηση της διπλωματικής εργασίας.

## 2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

### 2.1 Εισαγωγή

Στο κεφάλαιο της βιβλιογραφικής ανασκόπησης παρουσιάζονται **συναφείς μεθοδολογίες και έρευνες** που είναι παρεμφερείς με το αντικείμενο της παρούσας διπλωματικής εργασίας. Συγκεκριμένα, διερευνήθηκαν στην διεθνή βιβλιογραφία δημοσιευμένες έρευνες που εστιάζουν στην ανάλυση, αναγνώριση της συμπεριφοράς του οδηγού καθώς και την πρόβλεψη συγκρούσεων σε πραγματικό χρόνο, αξιοποιώντας διάφορες τεχνικές μηχανικής εκμάθησης και βαθιάς μάθησης.

Με την παρουσίαση και σύγκριση των διάφορων ερευνών θα καθοριστεί ο στόχος της παρούσας μελέτης καθώς και η κατάλληλη μεθοδολογία για την επίτευξη του. Τα διάφορα μοντέλα μηχανικής εκμάθησης και βαθιάς μάθησης θα είναι στο επίκεντρο της προσοχής στο παρόν κεφάλαιο. Στην πλειονότητα των ερευνών που διερευνήθηκαν παρουσιάζεται το πρόβλημα της άνισης κατανομής των δεδομένων στις διαφορετικές τάξεις. Για τον λόγο αυτό θα παρουσιαστούν οι διαφορετικές τεχνικές επαναδειγματοληψίας δεδομένων που έχουν εφαρμοστεί σε προγενέστερες έρευνες.

### 2.2 Συναφείς έρευνες και μεθοδολογίες

Υπάρχουν ποικίλοι παράγοντες που επηρεάζουν την επικίνδυνη συμπεριφορά των ατόμων κατά την οδήγηση. Σύμφωνα με τους Wang et al. (2020) η σοβαρότητα της επικίνδυνης οδήγησης σχετίζεται με διάφορους κυκλοφοριακούς παράγοντες συμπεριλαμβανομένων των χαρακτηριστικών συμπεριφοράς του οδηγού, των χαρακτηριστικών του οχήματος και του περιβάλλοντος.

Ο αριθμός των θανατηφόρων τροχαίων ατυχημάτων αυξάνεται ταυτόχρονα με την αύξηση οδικής κυκλοφορίας και την αποδοχή των ιδιωτικών αυτοκινήτων. Για αυτό τα τελευταία χρόνια η ανάλυση της οδηγικής συμπεριφοράς με την χρήση αλγόριθμων μηχανικής εκμάθησης είναι επίκεντρο της επιστημονικής κοινότητας και ταυτόχρονα η ανάπτυξη των Ευφυών Μεταφορικών Συστημάτων (Intelligent Transportation Systems - ITS) έχει δημιουργήσει πρόσφορο έδαφος στον τομέα των αυτόνομων οχημάτων (Peppes et al., 2021).

Εξίσου σημαντικό είναι η ικανότητα προσδιορισμού του επιπέδου ασφάλειας της οδηγικής συμπεριφοράς και η ικανότητα πρόβλεψης σε πραγματικό χρόνο της πιθανότητας ατυχήματος καθώς μπορούν να συνεισφέρουν σημαντικά στην εξέλιξη των Αναπτυσσόμενων Συστημάτων Υποβοήθησης Οδηγού (Advanced Driver Assistance Systems – ADAS) (Yang et al., 2021).

Αρκετές δημοσιοποιημένες έρευνες εμβαθύνουν στην κατανόηση της επίδρασης των διαφορετικών χαρακτηριστικών στην επικίνδυνη οδηγική συμπεριφορά και συμβάλλουν στην επίτευξη όλων των παραπάνω στόχων και στην εξέλιξη στον τομέα της αυτόματης οδήγησης. Η διερεύνηση της επιρροής δύναται να εξελίξει τα κατάλληλα μοντέλα αναγνώρισης επικίνδυνων συμπεριφορών κατά την οδήγηση και κατ' επέκταση να βελτιώσει την αποδοτικότητα των συστημάτων υποβοήθησης του οδηγού.

### 2.2.1 Ανάλυση οδηγικής συμπεριφοράς

Οι βασικές προσεγγίσεις για την ανάλυση της επικίνδυνης οδηγικής συμπεριφοράς αποτελούνται από μελέτες βασισμένες σε δεδομένα από έρευνες, μακροσκοπική ανάλυση δεδομένων ατυχήματος, μελέτες βασισμένες σε πείραμα δεδομένα προσομοιωτή οδήγησης και νατουραλιστικές μελέτες οδήγησης (NDS). Σε προηγούμενες έρευνες έχουν διερευνηθεί ορισμένοι δείκτες ασφαλείας για την αξιολόγηση της πιθανότητας κινδύνου όπως είναι ο χρόνος πρόσκρουσης (TTC) (Shi et al., 2018).

Οι Wang et al. (2010) πρότειναν έναν αλγόριθμο ημι-επιβλεπόμενης μηχανικής εκμάθησης για την χρήση τόσο των επισημασμένων και μη επισημασμένων δεδομένων, καθώς και της αλληλεξάρτησης τους για τη δημιουργία μιας σωστής συνάρτησης στο επίπεδο κινδύνου. Τα αποτελέσματα δείχνουν ότι η προτεινόμενη μέθοδος απαιτεί λιγότερη ώρα εκπαίδευσης και πέτυχε μεγαλύτερη ακρίβεια πρόβλεψης. Ωστόσο, αυτή η πρόβλεψη χρησιμοποιεί δύο περιπτώσεις, δηλαδή την περίπτωση ασφαλούς/επικίνδυνης οδήγησης, η οποία δεν είναι αρκετά πειστική. Οι ολοκληρωμένες πειραματικές αξιολογήσεις μας αποκαλύπτουν ότι, σε σύγκριση με την εκτίμηση επιπέδου επικινδυνότητας οδήγησης χρησιμοποιώντας μεθόδους που βασίζονται σε ταξινόμηση, όπως το κρυφό μοντέλο Markov (HMM) ή ο αλγόριθμος τυχαίου πεδίου υπό όρους, η προτεινόμενη μέθοδος απαιτεί λιγότερο χρόνο εκπαίδευσης και επιτυγχάνει υψηλότερη ακρίβεια πρόβλεψης.

Η έρευνα των P. Ping et al. (2019) εξετάζει τη σχέση μεταξύ της οδηγικής συμπεριφοράς και της κατανάλωσης καυσίμων που μπορεί να συμβάλλει στη μείωση του ενεργειακού κόστους της μεταφοράς και την ανάπτυξη της τεχνολογίας αξιολόγησης συμπεριφοράς για το σύστημα ADAS. Σε αυτή την έρευνα στόχος είναι ο σχεδιασμός μιας μεθόδου που βασίζεται στην μηχανική εκμάθηση, η οποία μπορεί να προβλέψει και να αναλύσει μια λογική σχέση μεταξύ της οδηγικής συμπεριφοράς και της κατανάλωσης καυσίμου και προτείνεται μια προσέγγιση που χρησιμοποιεί 2 μεθόδους μηχανικής εκμάθησης για την αξιολόγηση της απόδοσης καυσίμου της οδηγικής συμπεριφοράς χρησιμοποιώντας τα νατουραλιστικά δεδομένα οδήγησης. Στο πρώτο στάδιο χρησιμοποιείται μια μέθοδος μηχανικής εκμάθησης χωρίς επίβλεψη για να αναλύσει μακροσκοπικά της απόδοση του καυσίμου της οδηγικής συμπεριφοράς. Χρησιμοποιείται ένας αλγόριθμος φασματικής ομαδοποίησης για να ταξινομηθούν τα σύνολα δεδομένων που συλλέγονται από πολλούς οδηγούς. Οι οδηγοί κατηγοριοποιούνται σε 3 κατηγορίες ανάλογα στις ομοιότητες της οδηγικής τους συμπεριφοράς. Στο δεύτερο στάδιο χρησιμοποιούμε ένα επαναλαμβανόμενο τεχνητού νευρωνικού δικτύου (LSTM) για την πρόβλεψη των χαρακτηριστικών της βραχυπρόθεσμης κατανάλωσης καυσίμου και περιγράφει την απόδοση του μοντέλου χρησιμοποιώντας αντιπροσωπευτικό αποτέλεσμα πρόβλεψης χαρακτηριστικών κατανάλωσης καυσίμου.

Η μελέτη των Shi et al. (2019) σχεδιάζει ένα πλαίσιο εξαγωγής και επιλογής χαρακτηριστικών, για την αξιολόγηση της οδήγησης οχήματος και την πρόβλεψη των επιπέδων κινδύνου. Το πλαίσιο ενσωματώνει την επιλογή χαρακτηριστικών που βασίζονται στην μάθηση, την αξιολόγηση κινδύνου χωρίς επίβλεψη και τη μη ισοροπημένη δειγματοληψία δεδομένων. Για την εκτίμηση των δυνατοτήτων κινδύνου των οχημάτων κατά την οδήγηση, προτείνεται η σήμανση δεδομένων χωρίς επίβλεψη. Με βάση τα εξαγόμενα χαρακτηριστικά των δεικτών κινδύνου, τα οχήματα ομαδοποιούνται σε διάφορα επίπεδα επικινδυνότητας με τεχνικές μη επιβλεπόμενης μάθησης. Έπειτα, γίνεται ο καθορισμός των



βασικών χαρακτηριστικών σύμφωνα με την κατάταξη της σημασίας των χαρακτηριστικών και την αναδρομική εξάλειψη. Τα επίπεδα κινδύνου των οχημάτων κατά την οδήγηση προβλέπονται με βάση τα βασικά χαρακτηριστικά που έχουν επιλεγεί. Τα ευρήματα δείχνουν ότι αυτή η προσέγγιση είναι αποτελεσματική και αξιόπιστη για τον εντοπισμό σημαντικών χαρακτηριστικών για την αξιολόγηση οδήγησης και για την επίτευξη ακριβούς πρόβλεψης των επιπέδων κινδύνου.

Οι οδηγοί υψηλού κινδύνου είναι δυνατόν να εμπλακούν είναι πιο πιθανό να εμπλακούν σε τροχαία ατυχήματα και το επίπεδο κινδύνου οδήγησης των οδηγών επηρεάζεται από παράγοντες όπως τα δημογραφικά στοιχεία και χαρακτηριστικά προσωπικότητας. Η μελέτη των Song et al. (2021) με βάση το μοντέλο δοκιμών εξισώσεων (SEM) διερευνά τις σχέσεις μεταξύ των δημογραφικών χαρακτηριστικών των οδηγών (φύλο, ηλικία και αθροιστικά έτη οδήγησης), της αναζήτησης περιπέτειας, της αντίληψης του κινδύνου και της επικίνδυνης οδηγικής συμπεριφοράς του οδηγού. Με την χρήση τεχνικών ομαδοποίησης ανάλογα το φύλο του οδηγού καθορίζονται 3 ομάδες επικινδυνότητας η 'χαμηλή', η 'μεσαία' και η 'υψηλή'. Τέλος, προτείνεται ένα μοντέλο ταξινόμησης του κινδύνου οδήγησης του οδηγού από την ταξινόμηση Τυχαίων Δασών (Random Forests).

Στην έρευνα των Shangguan et al. (2021) προτείνεται μία μεθοδολογία για την αξιολόγηση και την πρόβλεψη της κατάστασης κινδύνου που βρίσκεται ο οδηγός σε πραγματικό χρόνο. Στην μελέτη αναπτύσσονται αλγόριθμοι ομαδοποίησης μέσω των οποίων καθορίζονται 4 στάδια επικινδυνότητας. Έπειτα γίνεται χρήση ορισμένων αλγορίθμων μηχανικής εκμάθησης για την πρόβλεψη της κατάστασης κινδύνου. Τέλος κάνοντας ανάλυση της επιρροής των μεταβλητών προκύπτει ότι η διαφορά ταχύτητας, η απόσταση από το προπορευόμενο όχημα, η ταχύτητα και η επιτάχυνση είναι ιδιαίτερα σημαντικές για την πρόβλεψη της κατάστασης επικινδυνότητας του οδηγού.

Η έρευνα των Yang et al. (2021) που πραγματοποιήθηκε στο πλαίσιο του ερευνητικού έργου i-DREAMS, προτείνει ένα πλαίσιο εργασίας για την ταξινόμηση και την αξιολόγηση διαφορετικών επιπέδων ασφαλείας της οδηγικής συμπεριφοράς. Με την χρήση προσομοιωτή οδήγησης συλλέγονται δεδομένα που σχετίζονται με διάφορα χαρακτηριστικά οδήγησης. Με την χρήση διαφορετικών τεχνικών ομαδοποίησης (clustering) προκύπτουν 4 επίπεδα ασφαλείας. Έπειτα για την αναγνώριση του επιπέδου ασφαλείας που βρίσκεται ο οδηγός με βάση τα οδηγικά χαρακτηριστικά χρησιμοποιούνται και συγκρίνονται διαφορετικοί αλγόριθμοι ταξινόμησης.

Παρά τις προόδους της τεχνολογίας της αυτόνομης οδήγησης τα τροχαία ατυχήματα παραμένουν ένα πρόβλημα που πρέπει να επιλυθεί στο σύστημα μεταφορών καθώς περισσότερα από τα μισά τροχαία ατυχήματα οφείλονται σε μη ασφαλή οδήγηση και η επιθετική οδηγική συμπεριφορά μπορεί να οδηγήσει σε κυκλοφοριακή συμφόρηση. Στην έρευνα των Seong Kyung Kwong et al. (2021) για την μείωση αυτού του φαινομένου προτείνεται μια προσέγγιση τεσσάρων επιπέδων Συνελικτικών Νευρωνικών Δικτύων (Convolutional Neural Networks – CNN) και 2 στοιβάδες Μακράς Βραχυπρόθεσμης Μνήμης (Long-Short Term Memory – LSTM) για ταξινόμηση επιθετικής συμπεριφοράς σε αμυντική, κανονική και επιθετική οδήγηση και ένα σύστημα κοινής χρήσης συμπεριφοράς οδήγησης βασισμένο σε V2X.

Στην έρευνα των Yang et al. (2022) προτείνεται ένα πλαίσιο ταξινόμησης και αξιολόγησης του επιπέδου ασφαλείας της οδηγικής συμπεριφοράς σε πραγματικό χρόνο, το οποίο επικυρώθηκε από μία μελέτη πειραμάτων σε προσομοιωτή οδήγησης.

Η προτεινόμενη μεθοδολογία επικεντρώνεται στον προσδιορισμό του βέλτιστου χρονικού διαστήματος συνάθροισης, την εύρεση του βέλτιστου αριθμού επιπέδων ασφαλείας για την οδηγική συμπεριφορά, την ταξινόμηση των επιπέδων ασφαλείας και αξιολόγηση των επιπέδων ασφαλείας σε πραγματικό χρόνο. Ένα βελτιωμένο cross validation mean square error model βασισμένο σε δεδομένα της οδηγικής συμπεριφοράς για τον προσδιορισμό του βέλτιστου χρονικού διαστήματος συνάθροισης το οποίο βρέθηκε 1 δευτερόλεπτο. Η οδηγική ασφάλεια μπορεί να ομαδοποιηθεί σε 3 επίπεδα τα οποία χαρακτηρίζονται ως «κανονική» οδήγηση, «χαμηλού κινδύνου» οδήγηση και οδήγηση «υψηλού κινδύνου». Εφαρμόστηκαν 3 τεχνικές ομαδοποίησης ο αλγόριθμος ομαδοποίησης k-means, ιεραρχική μέθοδος (hierarchical clustering) και η μέθοδος τύπων/μοντέλων (model-based clustering). Ως μοντέλα ταξινόμησης αναπτύχθηκαν μηχανές διανυσμάτων υποστήριξης (Support Vector Machines SMV), δένδρα αποφάσεων (decision trees) και η τεχνική Naïve Bayes. Η ακρίβεια του συνδυασμού k-means ομαδοποίησης και δένδρων απόφασης αποδείχθηκε η καλύτερη από τις 3 μεθόδους.

Στην έρευνα των Gunapriya Balan et al. (2022) προτείνεται μία μεθοδολογία για την αυτοματοποίηση και εξατομίκευση των προηγμένων συστημάτων υποβοήθησης της οδήγησης (ADAS) και να βοηθήσει σε κρίσιμες εργασίες όπως το σύστημα πέδησης με αποτελεσματικό τρόπο. Στην μελέτη αναπτύσσονται μοντέλα ταξινόμησης βαθιών οδηγών (βαθύ νευρωνικό δίκτυο ως DNN) με τεχνικές μείωσης χαρακτηριστικών (τυχαίο δάσος ως RF και ανάλυση κύριου στοιχείου ως PCA). Χρησιμοποιώντας μοντέλα εργασιών, προσομοιώνεται ένα σύστημα χαμηλού κόστους υποβοηθούμενο από τον οδηγό σε πραγματικό χρόνο, όπου διερευνώνται διάφορα σενάρια και καθορίζεται ο προγραμματισμός των εργασιών πριν από την εφαρμογή τους.

Στον πίνακα 2.1 παρουσιάζονται οι περιορισμοί, οι ελλείψεις και οι προτάσεις για περαιτέρω των παραπάνω μελετών.

Πίνακας 2.1: Ελλείψεις/Προτάσεις για μελλοντική διερεύνηση των ερευνών που παρουσιάστηκαν

Έρευνα	Ελλείψεις	Προτάσεις για μελλοντική διερεύνηση
J. Wang et al., 2010	Τα δείγματα για επικίνδυνες καταστάσεις οδήγησης είναι ελλιπή καθώς είναι δύσκολη η συλλογή τους, δεν υπάρχουν κοινά αποδεκτά κριτήρια για να χαράξουμε ένα σαφές όριο μεταξύ ασφαλών και επικίνδυνων καταστάσεων οδήγησης και ολόκληρος ο χώρος της κατάστασης	Προτείνεται η ενσωμάτωση του μοντέλου σε έναν πιο ρεαλιστικό προσομοιωτή οδήγησης ή σε πραγματικά οχήματα για την εκτέλεση μιας πρόσθετης υποκειμενικής μελέτης χρήστη, η ενσωμάτωση των χαρακτηριστικών που βασίζονται στην οπτική συμπεριφορά του οδηγού για περαιτέρω βελτίωση της απόδοσης,

	<p>οδήγησης είναι πράγματι ένας συνεχής χώρος όπου καταστάσεις που αντιστοιχούν σε ατυχήματα οδήγησης διασκορπίζονται σε πολλά σημεία και οποιοσδήποτε μεταβάσεις μεταξύ ασφαλών και επικίνδυνων καταστάσεων συμβαίνουν μάλλον συνεχώς παρά διακριτά.</p>	<p>η αξιολόγηση πρόσθετων αλγόριθμων μηχανικής εκμάθησης σε μεγαλύτερα σύνολα δεδομένων και η βελτίωση της διέπαφης χρήστη σε πραγματικά οχήματα και να έχει έναν καλό τρόπο να ειδοποιεί τον οδηγό για τον επείγοντα χαρακτήρα των πληροφοριών.</p>
<p>P. Ping et al., 2019</p>	<p>Οι ελλείψεις των συλλεγόμενων δεδομένων θα επηρεάσουν τη μέθοδο που βασίζεται στη βαθιά μάθηση, καθώς τα δεδομένα που συλλέγονται προέρχονται από 2 είδη δρόμων και οι παράγοντες περιβάλλοντος κυκλοφορίας δεν κωδικοποιούνται όλοι στα δεδομένα χρονοσειράς. Οι πληροφορίες εισόδου του μοντέλου έχουν περιορισμένο αριθμό συνθηκών κυκλοφορίας και η μορφή τους είναι σχετικά βασική, έτσι η ακρίβεια της μεθόδου που βασίζεται στο LSTM δεν είναι υψηλή. Γίνεται χρήση μόνο ενός τύπου πειραματικού οχήματος.</p>	<p>Προτείνεται η χρήση φυσιοκρατικών δεδομένων οδήγησης μεγαλύτερης κλίμακας για αποκτήσει το μοντέλο πρόβλεψης μεγαλύτερη στιβαρότητα. Στη συνέχεια, η επίδραση των άλλων παραγόντων, όπως το χαρακτηριστικό της ομαδικής προσωπικότητας ή ο τύπος οχήματος, στην ανάλυση κατανάλωσης καυσίμου θα πρέπει επίσης να μελετηθεί, προκειμένου να γίνει το μοντέλο πρόβλεψης κατανάλωσης καυσίμου πιο γενικό.</p>
<p>Shi et al., 2019</p>	<p>Η έλλειψη περιπτώσεων σύγκρουσης στην διαβάθμιση της επικινδυνότητας καθιστά δύσκολη την επαλήθευση των συνδέσεων μεταξύ του υψηλού επιπέδου κινδύνου και της πραγματικής εμφάνισης ατυχήματος.</p>	<p>Προτείνεται η σε βάθος εξαγωγή χαρακτηριστικών, η οποία θα καλύπτει ένα μεγάλο εύρος οδηγικής συμπεριφοράς και επικίνδυνων συνθηκών, όπως η αλλαγή λωρίδας και η σύγκρουση μεταξύ μοτοσυκλέτας και οχήματος.</p>
<p>Song et al., 2021</p>	<p>Τα δεδομένα της έρευνας ενδέχεται να αντιμετωπίζουν το πρόβλημα της μεροληψίας, καθώς προέρχονται από τα</p>	<p>Σε μελλοντική έρευνα περισσότερα χαρακτηριστικά της προσωπικότητας του οδηγού, όπως η επιθετική συμπεριφορά, και ψυχολογικοί παράγοντες θα μπορούσαν να ληφθούν υπόψη.</p>

	αποτελέσματα υποκειμενικών ερωτηματολογίων.	Επιπρόσθετα προτείνεται σε μελλοντική έρευνα ο συνδυασμός των αποτελεσμάτων των ερωτηματολογίων και αντικειμενικών δεδομένων οδήγησης προκειμένου τα αποτελέσματα να έχουν μεγαλύτερη ακρίβεια.
Shangguan et al., 2021	Η έρευνα περιορίζεται μόνο στην επικίνδυνη επακολουθία των οχημάτων αγνοώντας επικίνδυνες συμπεριφορές κατά την αλλαγή λωρίδας. Επίσης, σημαντικά χαρακτηριστικά του οχήματος και της οδού δεν λαμβάνονται υπόψη στην ανάλυση.	Κρίνεται αναγκαίο από τους μελετητές να συμπεριληφθούν επικίνδυνες ενέργειες αλλαγής λωρίδας σε μελλοντικές έρευνες. Επίσης, μελλοντικά θα μπορούσαν να αναπτυχθούν ορισμένοι αλγόριθμοι βαθιάς εκμάθησης. Τέλος επιπρόσθετες μεταβλητές όπως τα χαρακτηριστικά του οχήματος και τα γεωμετρικά χαρακτηριστικά της οδού θα μπορούσαν να εξεταστούν για την βελτίωση και εξέλιξη του μοντέλου πρόβλεψης.
Yang et al., 2021	Τα δεδομένα που αναλύθηκαν δεν περιλάμβαναν την στάση, την αντίληψη και τα δημογραφικά χαρακτηριστικά των οδηγών.	Προτείνεται η εξέταση διαφορετικών χαρακτηριστικών του οδηγού. Επίσης, οι ερευνητές προτείνουν ότι μελλοντικές μελέτες θα μπορούσαν να εστιάσουν στις πιο σημαντικές μεταβλητές εξετάζοντας την σχέση τους με τα διαφορετικά επίπεδα ασφαλείας καθώς και την μεταξύ τους σχέση.
Seong Kyung Kwong et al., 2021	Ελλείψεις στον αριθμό σεναρίων για την επιθετική οδήγηση στον δρόμο και όχι τόσο λεπτομερής ταξινόμηση για την ασφάλεια της οδήγησης.	Προτείνεται η κατασκευή σεναρίων αλλαγής λωρίδας χωρίς σήμα ή αλλαγής πολλών λωρίδων ταυτόχρονα. Το υπόλοιπο ερευνητικό έργο μπορεί να είναι η κατασκευή ενός μεγάλου αριθμού σεναρίων για επιθετική οδήγηση που μπορεί να συμβεί στο δρόμο και η λεπτομερής ταξινόμηση της επιθετικότητας κατά την οδήγηση.
Yang et al., 2022	Στην έρευνα το σύνολο δεδομένων δεν περιλαμβάνει υπάρχουσες μεταβλητές σχετικά	Προτείνεται σε μελλοντικές έρευνες να ληφθούν υπόψη οι υπάρχουσες μεταβλητές σχετικά με τα

με τα δημογραφικά στοιχεία και τις στάσεις, αντιλήψεις των οδηγών.

δημογραφικά στοιχεία και οι στάσεις, αντιλήψεις των οδηγών, για την βελτίωση της ομαδοποίησης και της ταξινόμησης. Επιπλέον, σε μελλοντικές έρευνες να εξεταστούν περισσότερα διαφορετικά σενάρια οδήγησης.

Gunapriya  
Balan et al.,  
2022

Η απόδοση των τεχνικών βαθιάς μάθησης μειώνεται όταν όσο αυξάνονται τα χαρακτηριστικά και οι πρόσθετες λειτουργίες.

Προτείνεται σε μελλοντικές έρευνες για την βελτίωση της απόδοσης του συστήματος αναγνώρισης οδηγού η προσθήκη πρόσθετων λειτουργιών, όπως χαρακτηριστικά Cepstral σε συνδυασμό με λειτουργίες τομέα χρόνου και τομέα συχνότητας. Η κατανάλωση ενέργειας μπορεί να ελαχιστοποιηθεί χρησιμοποιώντας μια ανεπτυγμένη τεχνική ως μελλοντική εργασία.

## 2.2.2 Αλγόριθμοι ταξινόμησης οδηγικής συμπεριφοράς

Στον τομέα της οδικής ασφάλειας υιοθετούνται όλο και περισσότεροι αλγόριθμοι μηχανικής εκμάθησης. Για την ταξινόμηση της οδηγικής συμπεριφοράς και την ανάπτυξη μοντέλων πρόβλεψης συγκρούσεων σε πραγματικό χρόνο πολλές μελέτες επικεντρώνονται στην ανάπτυξη διαφορετικών αλγόριθμων ταξινόμησης. Όπως μπορούμε να δούμε στον πίνακα 2.2 παρουσιάζονται αλγόριθμοι ταξινόμησης με την υψηλότερη απόδοση από τις αντίστοιχες έρευνες.

Πίνακας 2.2: Αποτελεσματικότεροι αλγόριθμοι ταξινόμησης ανά έρευνα ανάλυσης οδηγικής συμπεριφοράς

Έρευνα	Σκοπός αλγόριθμων ταξινόμησης	Αλγόριθμοι ταξινόμησης με το υψηλότερο ποσοστό ορθών προβλέψεων
J. Wang et al., 2010	Χρήση τόσο των επισημασμένων και μη επισημασμένων δεδομένων, καθώς και της αλληλεξάρτησης τους για τη δημιουργία μιας σωστής συνάρτησης στο επίπεδο κινδύνου	Δένδρα απόφασης (DT): ποσοστό ορθών προβλέψεων 98.6%
P. Ping et al., 2019	Αξιολόγηση της απόδοσης καυσίμου της οδηγικής συμπεριφοράς	Επαναλαμβανόμενο τεχνητό νευρωνικό δίκτυο (LSTM): ποσοστό ορθών προβλέψεων 81%

	χρησιμοποιώντας τα νατουραλιστικά δεδομένα οδήγησης	
Shi et al., 2019	Ταξινόμηση της οδηγικής συμπεριφοράς σε 4 επίπεδα	Ακραία Διαβαθμιζόμενη Ενδυνάμωση (XGBoost): ποσοστό ορθών προβλέψεων 89%
Song et al., 2021	Ταξινόμηση επικίνδυνης οδηγικής συμπεριφοράς σε 3 επίπεδα ασφαλείας	Τυχαία δάση (RF): ποσοστό ορθών προβλέψεων 90%
Shangguan et al., 2021	Ταξινόμηση επικίνδυνης οδηγικής κατάστασης σε 4 επίπεδα	Πολυεπίπεδο Perceptron (MLP): ποσοστό ορθών προβλέψεων 85%
Yang et al., 2021	Ταξινόμηση οδηγικής συμπεριφοράς σε 4 επίπεδα ασφαλείας	Μηχανές διανυσμάτων υποστήριξης (SVM): ποσοστό ορθών προβλέψεων 95%
Seong Kyung Kwong et al., 2021	Βελτίωση της οδικής ασφάλειας προβλέποντας την επιθετική οδική συμπεριφορά χρησιμοποιώντας μόνο δεδομένα αισθητήρων χαμηλού κόστους και κοινοποιώντας τα με τα γύρω οχήματα για τη μείωση του κύκλου των δυσμενών επιπτώσεων	Συνελικτικά Νευρωνικά Δίκτυα (CNN) - Μακρά Βραχυπρόθεσμη Μνήμη (LSTM): ποσοστό ορθών προβλέψεων 98%
Yang et al., 2022	Προσδιορισμό του βέλτιστου χρονικού διαστήματος συνάθροισης, την εύρεση του βέλτιστου αριθμού επιπέδων ασφαλείας για την οδηγική συμπεριφορά, την ταξινόμηση των επιπέδων ασφαλείας και αξιολόγηση των επιπέδων ασφαλείας σε πραγματικό χρόνο	Μηχανές διανυσμάτων υποστήριξης (SVM): ποσοστό ορθών προβλέψεων 94,7% Δένδρα απόφασης (DT): ποσοστό ορθών προβλέψεων 98.6% Απλοί ταξινομητές N. Bayes: ποσοστό ορθών προβλέψεων 82.8%
Gunapriya Balan et al., 2022	Για να βοηθήσει στην αυτοματοποίηση και στην εξατομίκευση των συστημάτων ADAS και σε κρίσιμες εργασίες, όπως το σύστημα πέδησης με αποτελεσματικό τρόπο.	Βαθή Νευρωνικό Δίκτυο(DNN)-Τυχαία Δάση(RF): ποσοστό ορθών προβλέψεων 97,05% Βαθή Νευρωνικό Δίκτυο(DNN)-Ανάλυση Κύριου Στοιχείου(PCA): ποσοστό ορθών προβλέψεων 95,55%

Ο αλγόριθμος των 'Τυχαίων Δασών' και τα συνδυαστικά μοντέλα 'Βαθύ Νευρωνικό Δίκτυο (DNN)-Τυχαία Δάση (RF)', 'Συνελικτικά Νευρωνικά Δίκτυα (CNN) - Μακρά Βραχυπρόθεσμη Μνήμη (LSTM)' όπως προκύπτει από τις έρευνες που μελετήθηκαν έχουν τις καλύτερες επιδόσεις. Επιπρόσθετα οι 'Μηχανές διανυσμάτων υποστήριξης', τα 'Δένδρα απόφασης', σημειώνουν εξίσου καλές επιδόσεις.

### 2.2.3 Τεχνικές βελτίωσης της εκπαίδευσης και της γενίκευσης μοντέλων

Πολύ συχνό φαινόμενο είναι το να συναντάμε στα προβλήματα πραγματικού χρόνου το πρόβλημα της ανισοροπίας των δεδομένων όσον αφορά την κατανομή τους σε κάθε τάξη. Ειδικότερα σε σύγκριση με την ασφαλή οδηγική συμπεριφορά και την μη πρόκληση ατυχήματος αντίστοιχα, η επικίνδυνη οδηγική συμπεριφορά και η πιθανότητα ατυχήματος είναι σπάνια γεγονότα στις σχετικές έρευνες. Κύρια κλάση ονομάζεται η κλάση που διαθέτει τα περισσότερα δεδομένα ενώ κλάση μειοψηφίας ονομάζεται η κλάση με τα λιγότερα δεδομένα. Σε θέματα ανάλυσης των ατυχημάτων σε πραγματικό χρόνο η αναλογία των συμβάντων που σχετίζονται με το ατύχημα προς αυτών που δεν σχετίζονται με το ατύχημα κυμαίνεται από 1:5 (Roshandel et al., 2015) και από 1:20 (Xu et al., 2013). Σύμφωνα με τους Katrazakas et al., 2020, συμπεραίνεται ότι η ενσωμάτωση της τεχνικής Edited Nearest Neighbors (ENN) με την τεχνική Synthetic Minority Oversampling Technique (SMOTE) αποδίδει βελτιωμένα και ενδιαφέροντα αποτελέσματα όταν εφαρμόζεται σε πρόβλεψη σύγκρουσης σε πραγματικό χρόνο ή για την αξιολόγηση της ασφάλειας δεδομένων σε προσομοιωτή οδήγησης.

Στην έρευνα των Zhu, Shengxue, et al. (Zhu, Shengxue, et al, 2022) εξετάζονται διάφορες τεχνικές SMOTE με boost διαδικασία, Τυχαία Υπερδειγματοληψία ( Random Oversampling) και SMOTE-Adaboost. Όπως προτείνεται από την βιβλιογραφία για την αντιμετώπιση της ανισοροπίας των δεδομένων είναι αναγκαίο να εφαρμοστούν διάφορες τεχνικές υπερδειγματοληψίας και υποδειγματοληψίας, καθώς ταυτόχρονα με την εφαρμογή τους θα διευρυνθεί περαιτέρω η έρευνα στα προβλήματα ανισοροπίας του τομέα οδικής ασφαλείας.

Η τεχνική Dropout έχει χρησιμοποιηθεί ευρέως για την υπερπροσαρμογή των βαθιών νευρωνικών δικτύων με μεγάλο αριθμό παραμέτρων, οι οποίες συνήθως πανομοιότυπα και ανεξάρτητα σε τυχαία δείγματα συνόλων και νευρώνων να είναι μηδενικές (Zhe Li et al. 2016). Η εγκατάλειψη είναι μια απλή αλλά αποτελεσματική τεχνική για την πρόληψη της υπερπροσαρμογής στην εκπαίδευση των βαθιών νευρωνικών δικτύων. Σύμφωνα με την έρευνα των Zhe Li et al. προτείνεται η χρήση πολυωνυμικής δειγματοληψίας για το dropout, δηλαδή την δειγματοληψία χαρακτηριστικών ή νευρώνων σύμφωνα με μία πολυωνυμική κατανομή με διαφορετικές πιθανότητες για διαφορετικά χαρακτηριστικά- νευρώνες. Συμπεραίνεται ότι η τεχνική dropout επιτυγχάνει μικρότερο κίνδυνο και μεγαλύτερη σύγκλιση και αναπτύχθηκε μια αποτελεσματική εξελικτική εγκατάλειψη για την εκπαίδευση σε βαθιά νευρωνικά δίκτυα που προσαρμόζει τις πιθανότητες δειγματοληψίας στις εξελισσόμενες κατανομές των εξόδων των επιπέδων.

## 2.3 Αξιολόγηση επίδρασης μεταβλητών στην επικίνδυνη οδήγηση

Η βαθιά μάθηση έχει αποτελέσει σημαντική πρόοδο στην τεχνητή νοημοσύνη τα τελευταία χρόνια. Αντίθετα με παραδοσιακές μεθόδους μηχανικής μάθησης, οι μέθοδοι βαθιάς μάθησης έχουν επιτύχει σημαντική βελτίωση σε διάφορες εργασίες πρόβλεψης, όπως στην επεξεργασία εικόνας και στην επεξεργασία φυσικής γλώσσας. Ωστόσο, τα βαθιά νευρωνικά δίκτυα είναι συγκριτικά αδύναμα στην εξήγηση των διαδικασιών εξαγωγής συμπερασμάτων και των τελικών αποτελεσμάτων τους.

Σύμφωνα με την έρευνα Hao Liu et al. (2024) αναπτύσσετε ένα μοντέλο Transformer βαθιάς εκμάθησης για την αναγνώριση της πρόθεσης αλλαγής λωρίδας. Για την αντιμετώπιση του «μαύρου κουτιού» των μοντέλων βαθιάς μάθησης εισάγεται το μοντέλο Lime για να παρέχεται μία πιο ερμηνεύσιμη ανάλυση του μοντέλου αναγνώρισης και αποκαλύπτει ότι το μοντέλο εστιάζει σε χαρακτηριστικά παρακολούθησης ματιών και αλληλεπιδράσεις οχημάτων αλλαγής λωρίδας με την προηγούμενη και την επερχόμενη κυκλοφορία κατά τη διάρκεια συμβάντων.

Σύμφωνα με την έρευνα των Datta et al. (2018) τα δύο βασικά προβλήματα στην μηχανική εκμάθηση και τα βαθιά νευρωνικά μοντέλα είναι η έλλειψη ποσοτικοποίησης της αβεβαιότητας και η αδυναμία τους να επεξηγήσουν ποιες μεταβλητές είναι σημαντικές. Στη μελέτη χρησιμοποιούνται δύο ερμηνευτικά μοντέλα 'Local interpretable model-agnostic explanations' και 'Shapley Values' και συγκρίνεται η ανθεκτικότητα τους στην εξήγηση των προβλέψεων ενός μοντέλου ταξινόμησης κειμένου, με αποτέλεσμα το μοντέλο Lime να είναι πιο ανθεκτικό σε αλλαγές.

Η μελέτη των Michelaraki et al. (2021) έχει εκπονηθεί στο πλαίσιο του ερευνητικού έργου i-DREAMS και έχει στόχο να παρέχει πληροφορίες για την χαρτογράφηση των μεθόδων που θα αξιοποιηθούν για την πραγματοποίηση των στόχων του έργου i-DREAMS. Για κάθε παράγοντα κινδύνου είναι γνωστό το επίπεδο 'Ζώνης Ανοχής Ασφαλείας' και σε κάθε φάση στοχεύεται ένας συγκεκριμένος παράγοντας κινδύνου για την πρόβλεψη του επιπέδου ασφαλείας χωρίς να λαμβάνονται υπόψη άλλες εξίσου σημαντικές μεταβλητές. Αναγκαία είναι η χρήση του συνόλου των παραγόντων στο ίδιο μοντέλο για να γίνει η πρόβλεψη του κινδύνου οδήγησης. Επίσης, για την αξιολόγηση της επίδρασης όλων των μεταβλητών στην επικίνδυνη οδήγηση εξετάζεται η χρήση του αλγόριθμου Lime (Local Interpretable Model-Agnostic Explanation) ως υποστηρικτικό εργαλείο για την ανάλυση της επιρροής των ανεξάρτητων μεταβλητών στην πρόβλεψη των μοντέλων βαθιάς εκμάθησης.

## 2.4 Σύνοψη

Αξιοποιώντας μεθοδολογίες που χρησιμοποιούνται σε παλαιότερες έρευνες που αναφέρονται στην βιβλιογραφική ανασκόπηση, η παρούσα διπλωματική εργασία θα επιχειρήσει να προσφέρει επιπλέον γνώμη στον τομέα των Ευφυών Μεταφορικών Συστημάτων (ITS).

Με βάση τα παραπάνω, στόχος της διπλωματικής εργασίας είναι ο εντοπισμός του επιπέδου και της διάρκειας της επικίνδυνης συμπεριφοράς του οδηγού και ταυτόχρονα να επιλεγεί η κατάλληλη μεθοδολογία αξιολογώντας τα αποτελέσματα των παλαιότερων ερευνών.



Συγκεκριμένα θα εξεταστεί η επίδραση των διάφορων οδηγικών χαρακτηριστικών στην αναγνώριση των διαφορετικών επιπέδων 'οδηγικής συμπεριφοράς' και θα εξεταστεί η μεταξύ τους σχέση.

Ιδιαίτερη προσοχή θα δοθεί στις τεχνικές επαναδειγματοληψίας για την αντιμετώπιση των προβλημάτων ανισορροπίας των δεδομένων σε κάθε τάξη. Στην παρούσα διπλωματική εργασία αξιολογούνται οι μέθοδοι που έχουν χρησιμοποιηθεί σε παλαιότερες έρευνες και αυτές που έχουν προταθεί για μελλοντικές έρευνες.

Με την χρήση έξυπνων συστημάτων, μηχανικής και βαθιάς μάθησης όπως αναφέρθηκε στις παλαιότερες μεθόδους και με βάση την αποτελεσματικότητα των μοντέλων αυτών, επιλέγονται οι εξής συνδυασμοί αλγορίθμων για την ταξινόμηση των οδηγών σε ένα από τα τρία επίπεδα της 'Ζώνης Αντοχής Ασφαλείας' :

1. Ο συνδυασμός '**Βαθύ Νευρωνικού Δικτύου (DNN)**' – '**Τυχαία Δάση (RF)**' ο οποίος παρουσίασε υψηλή επίδοση στην πλειονότητα των ερευνών που μελετήθηκαν.
2. Ο '**Επαναλαμβανόμενου Νευρωνικού Δικτύου (RNN)**' – '**Adaboost**', καθώς το μοντέλο επαναλαμβανόμενου νευρωνικού δικτύου παρουσιάζει ικανοποιητική επίδοση και ο αλγόριθμος Adaboost βελτιώνει σε μεγάλο βαθμό την ακρίβεια των προβλέψεων στα μοντέλα πρόβλεψης ατυχήματος σε πραγματικό χρόνο όταν χρησιμοποιείται για την μείωση των ανομοιογενών δεδομένων.
3. Ο συνδυασμός '**Συνελικτικών Νευρωνικών Δικτύων (CNN)**' – '**Μοντέλων Μακράς Βραχυπρόθεσμης Μνήμης (LSTM)**', ο οποίος παρουσίασε την υψηλότερη επίδοση στο σύνολο των ερευνών.

Με βάση την αξιολόγηση των αλγορίθμων στην βιβλιογραφία θα εξεταστούν ορισμένες μετρικές αξιολόγησης.

Στο δεύτερο κομμάτι της παρούσας εργασίας θα πραγματοποιηθεί η αξιολόγηση της επιρροής των ανεξάρτητων μεταβλητών στην πρόβλεψη των συνδυασμένων μοντέλων, με σκοπό να εξεταστούν και να αξιολογηθούν τα χαρακτηριστικά που επιδρούν στην διάρκεια οδήγησης κάτω από επικίνδυνες συνθήκες.

Η αξιολόγηση της επίδρασης όλων των μεταβλητών στην επικίνδυνη οδήγηση θα πραγματοποιηθεί με τη χρήση του αλγορίθμου **Lime (Local Interpretable Model-Agnostic Explanation)** το οποίο λειτουργεί σαν υποστηρικτικό εργαλείο για την ανάλυση της επιρροής των ανεξάρτητων μεταβλητών στην πρόβλεψη των μοντέλων βαθιάς εκμάθησης.

### 3. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

#### 3.1 Εισαγωγή

Στην παρούσα ενότητα παρουσιάζεται το θεωρητικό υπόβαθρο που πραγματοποιήθηκε η επεξεργασία των δεδομένων και η ανάλυση τους. Αρχικά, πραγματοποιήθηκε η ανάλυση των τεχνικών επεξεργασίας και επαναδειγματοληψίας των δεδομένων καθώς εμφανίζεται ανισορροπία των δειγμάτων σε κάθε τάξη. Έπειτα, παρουσιάζονται τα μοντέλα μηχανικής και βαθιάς εκμάθησης που αναπτύχθηκαν για την ταξινόμηση της οδηγικής συμπεριφοράς στα τρία επίπεδα της Ζώνης Ανοχής Ασφαλείας. Στην συνέχεια, για να αξιολογηθούν τα

χαρακτηριστικά που επιδρούν στην διάρκεια οδήγησης κάτω από επικίνδυνες συνθήκες αναπτύχθηκε η μέθοδος Lime, η οποία χρησιμοποιείται ως υποστηρικτικό εργαλείο για την ανάλυση της επιρροής των ανεξάρτητων μεταβλητών στην πρόβλεψη των συνδυασμένων μοντέλων. Τέλος, δίνεται ιδιαίτερη έμφαση στην σημασία των μετρικών αξιολόγησης, των κριτηρίων αποδοχής των μοντέλων και των στατιστικών ελέγχων των αποτελεσμάτων.

### 3.2 Επιλογή χαρακτηριστικών (Feature Selection)

Η πλειονότητα των προβλημάτων ταξινόμησης του πραγματικού κόσμου απαιτούν εποπτευόμενη μάθηση, όπου οι υποκείμενες πιθανότητες κλάσης και οι πιθανότητες υπό όρους κλάσης είναι άγνωστες και κάθε περίπτωση σχετίζεται με μια ετικέτα κλάσης. Επομένως, για την καλύτερη αναπαράσταση του τομέα, εισάγονται πολλά υποψήφια χαρακτηριστικά, με αποτέλεσμα την ύπαρξη άσχετων/περιττών χαρακτηριστικών με την έννοια του στόχου (Jiliang Tang et al.).

Σε πολλά προβλήματα ταξινόμησης, είναι δύσκολη η εκπαίδευση ικανοποιητικών ταξινομητών πριν αφαιρεθούν τα ανεπιθύμητα χαρακτηριστικά, λόγω του τεράστιου μεγέθους των δεδομένων. Η μείωση του αριθμού των άσχετων/περιττών χαρακτηριστικών μπορεί να μειώσει δραστικά τον χρόνο εκτέλεσης των αλγορίθμων εκμάθησης και να αποφέρει έναν πιο γενικό ταξινομητή. Αυτό βοηθά στην απόκτηση καλύτερης εικόνας σχετικά με την υποκείμενη έννοια ενός προβλήματος ταξινόμησης του πραγματικού κόσμου. Η επιλογή χαρακτηριστικών επηρεάζει κυρίως τη φάση εκπαίδευσης της ταξινόμησης. Μετά τη δημιουργία χαρακτηριστικών, αντί να επιλέγονται δεδομένα με ολόκληρα χαρακτηριστικά, πρώτα θα επιλεγούν χαρακτηριστικά, ώστε να δημιουργηθεί ένα υποσύνολο χαρακτηριστικών και στη συνέχεια θα πραγματοποιηθεί η επεξεργασία των δεδομένων με τα επιλεγμένα χαρακτηριστικά.

Για βέλτιστη επιλογή των χαρακτηριστικών της παρούσας εργασίας συνδυάστηκαν δύο μέθοδοι.

Η πρώτη μέθοδος αφορά στον προσδιορισμό της **συσχέτισης (correlation)** μεταξύ των ανεξάρτητων μεταβλητών. Ο δειγματικός συντελεστής γραμμικής συσχέτισης του Pearson για δύο μεταβλητές συμβολίζεται με  $r$  και ορίζεται από την εξής μαθηματική εξίσωση.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Όπου,

$x_i, y_i$  οι τιμές των δύο μεταβλητών

$\bar{x}, \bar{y}$  ο μέσος όρος των τιμών

Οι τιμές κυμαίνονται από -1 έως 1, όπου  $r=0$  σημαίνει μηδενική συσχέτιση,  $r=1$  πλήρη θετική συσχέτιση και  $r=-1$  πλήρη αρνητική συσχέτιση (Nettleton, 2014). Θετική συσχέτιση μεταξύ δύο μεταβλητών παρατηρείται όταν η τιμή της μίας αυξάνεται και ταυτόχρονα αυξάνεται και η άλλη. Ενώ αρνητική συσχέτιση παρατηρείται όταν η μία αυξάνεται και η άλλη μειώνεται. Το βέλτιστο υποσύνολο αποτελείται από χαρακτηριστικά αρκετά συσχετισμένα με την προβλεπόμενη τάξη αλλά έχοντας ελάχιστη συσχέτιση μεταξύ τους (Hall, 2000). Επομένως,

στο πλαίσιο της μελέτης μεταξύ δύο χαρακτηριστικών με υψηλή συσχέτιση το ένα θα απορριφθεί για την περαιτέρω ανάλυση.

Η δεύτερη μέθοδος αφορά την επιλογή χαρακτηριστικών βασισμένη στον εντοπισμό της **σημαντικότητας των χαρακτηριστικών (feature importance)** του δείγματος. Η παραπάνω διαδικασία μπορεί να πραγματοποιηθεί με την εκπαίδευση αλγόριθμων μηχανικής εκμάθησης όπως είναι ο ταξινομητής τυχαίου δάσους (random forest classifier) ώστε να προσδιοριστεί ο βαθμός επιρροής κάθε μεταβλητής στην τελική πρόβλεψη.

### 3.3 Μέθοδοι βελτίωσης και γενίκευσης μοντέλων

Οι αλγόριθμοι βελτιστοποίησης αποτελούν ένα πολύ σημαντικό κομμάτι της εκπαιδευτικής διαδικασίας ενός νευρωνικού δικτύου. Σκοπός είναι να ελαχιστοποιηθεί η διαφορά μεταξύ των προβλέψεων από το μοντέλο και των πραγματικών ετικετών, προσαρμόζοντας τις παραμέτρους του μοντέλου. Οι αλγόριθμοι αυτοί έχουν σημαντικό ρόλο στην εύρεση του καλύτερου συνόλου των παραμέτρων που οδηγούν σε καλύτερες προβλέψεις, δηλαδή καθοδηγεί το νευρωνικό δίκτυο σε καλύτερες αναπαραστάσεις εισόδου άρα και απόδοση.

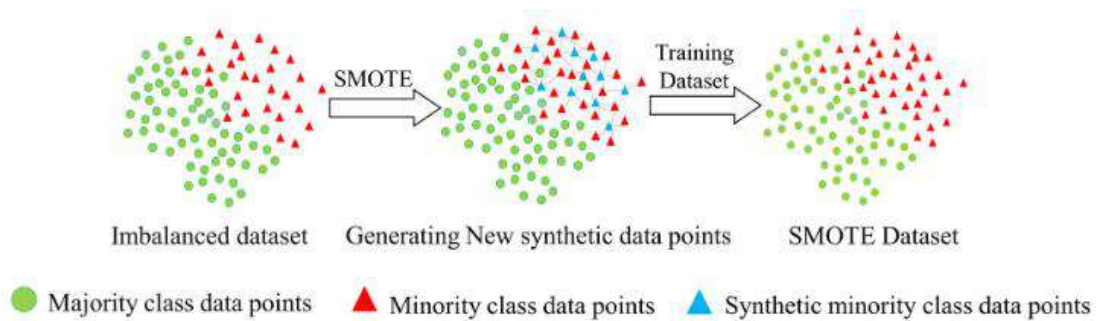
Η εκμάθηση ταξινομητών από μη ισορροπημένα δεδομένα έχει προσελκύσει σημαντικό ενδιαφέρον τα τελευταία χρόνια. Αυτό συμβαίνει επειδή στον πραγματικό κόσμο, υπάρχουν ανισορροπημένα δεδομένα σε πολλές εφαρμογές. Σε αυτές τις εφαρμογές, υπάρχουν συχνά μία ή ορισμένες μειονοτικές τάξεις που διαθέτουν πολύ λίγα δείγματα σε σύγκριση με τις άλλες κλάσεις. Και τις περισσότερες φορές, οι «μικρές» τάξεις είναι πιο σημαντικές από αυτές τις «μεγάλες». Λόγω της ανισορροπίας της κατανομής δεδομένων των προβλημάτων ανισορροπίας μάθησης, είναι συχνά δύσκολο να επιτευχθεί καλή απόδοση για τις περισσότερες περιπτώσεις με τη χρήση παραδοσιακών ταξινομητών όπου θεωρείται ισορροπημένη κατανομή κλάσεων και εκχωρείται ίσο κόστος λανθασμένης ταξινόμησης για κάθε τάξη. Ως αποτέλεσμα, οι παραδοσιακοί ταξινομητές τείνουν να κατακλύζονται από τις πλειοψηφικές τάξεις και να αγνοούν τις μειοψηφικές, κάτι που δεν είναι αποδεκτό σε πολλές πραγματικές εφαρμογές (Yun Qian et al., 2014).

Δεδομένου ότι στην παρούσα μελέτη οι κλάσεις μειοψηφίας είναι τα δύο επίπεδα επικίνδυνης οδήγησης, ενώ η κυρίαρχη κλάση είναι το επίπεδο ασφαλούς οδήγησης καθίσταται σαφής η ανάγκη ανάπτυξης τεχνικών επαναδειγματοληψίας των δεδομένων εκπαίδευσης των αλγόριθμων. Οι επιπτώσεις στην ασφάλεια των οδηγών θα ήταν ιδιαίτερα σοβαρές εάν τα μοντέλα μηχανικής εκμάθησης ταξινομούσαν λανθασμένα επικίνδυνες συμπεριφορές ως ασφαλείς.

#### 3.3.1 Τεχνική Συνθετικής Μειονοτικής Υπερδειγματοληψίας (SMOTE)

Η Τεχνική της συνθετικής μειονοτικής υπερδειγματοληψίας (Synthetic minority oversampling technique, **SMOTE**) αντιμετωπίζει το ενδεχόμενο της υπερπροσαρμογής που μπορεί να προκληθεί από την τυχαία υπερδειγματοληψία. Αυτό επιτυγχάνεται με την προσθήκη νέων συνθετικών μειονοτικών παραδειγμάτων. Αυτά τα νέα δεδομένα δημιουργούνται από παρεμβολή μεταξύ πολλών παρουσιών κλάσεων μειοψηφίας που βρίσκονται εντός μιας καθορισμένης γειτονιάς. Για το λόγο αυτό, η διαδικασία λέγεται ότι επικεντρώνεται στον «χώρο των χαρακτηριστικών» και όχι στον «χώρο δεδομένων», με άλλα λόγια, ο αλγόριθμος βασίζεται στις τιμές των χαρακτηριστικών και στη σχέση τους, αντί να θεωρεί τα σημεία δεδομένων ως ολόκληρος.

Προκειμένου να δημιουργηθεί ένα νέο συνθετικό παράδειγμα η μέθοδος SMOTE βρίσκει τους  $k$ -πλησιέστερους γείτονες για κάθε περίπτωση της κλάσης μειοψηφίας, επιλέγει εκείνον τον πλησιέστερο από τους  $k$ , και στη συνέχεια πολλαπλασιάζει τη διαφορά των διανυσμάτων του κοντινότερου γείτονα και των χαρακτηριστικών με έναν τυχαίο αριθμό  $m$  με  $m \in (0,1)$  και προστίθεται στο διάνυσμα των χαρακτηριστικών του δείγματος. Με αυτόν τον τρόπο, επιτυγχάνεται τόσο η αύξηση της μειονοτικής κλάσης όσο και η αύξηση της ποικιλίας των παραδειγμάτων (Chawla et al., 2002).

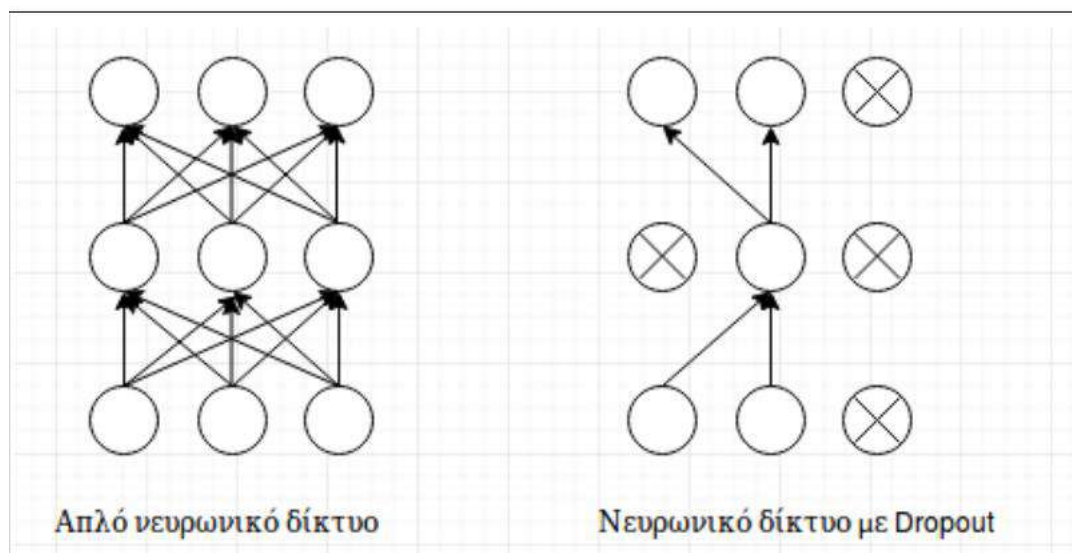


Γράφημα 3.1: SMOTE: Synthetic Minority Oversampling Technique  
 Πηγή: ResearchGate (2020)

### 3.3.2 Τεχνική Dropout

Το **Dropout** είναι μία τεχνική σχεδιασμένη για νευρωνικά δίκτυα που εμποδίζει την υπερπροσαρμογή. Η τεχνική συμβάλλει στη βελτίωση της γενίκευσης του μοντέλου με την τυχαία απομάκρυνση ενός κλάσματος των εξόδων των νευρώνων σε κάθε επανάληψη κατά την διάρκεια της εκπαίδευσης, με αποτέλεσμα να αποτρέπεται το φαινόμενο της υπερπροσαρμογής.

Πιο αναλυτικά, όπου τυχαία επιλεγμένοι νευρώνες απενεργοποιούνται κατά την διάρκεια εκπαίδευσης ενός νευρωνικού δικτύου. Κατά την διάρκεια της εκπαίδευσης ένα τυχαίο υποσύνολο της εξόδου των νευρώνων μηδενίζεται με καθορισμένη πιθανότητα. Καθορίζοντας την έξοδο του νευρώνα στο μηδέν αφαιρείται για την συγκεκριμένη επανάληψη από το δίκτυο, με αποτέλεσμα οι υπόλοιποι νευρώνες να εκπαιδευτούν από τα δυνατά χαρακτηριστικά και έτσι αποφεύγεται η εξάρτηση από έναν μεμονωμένο νευρώνα.

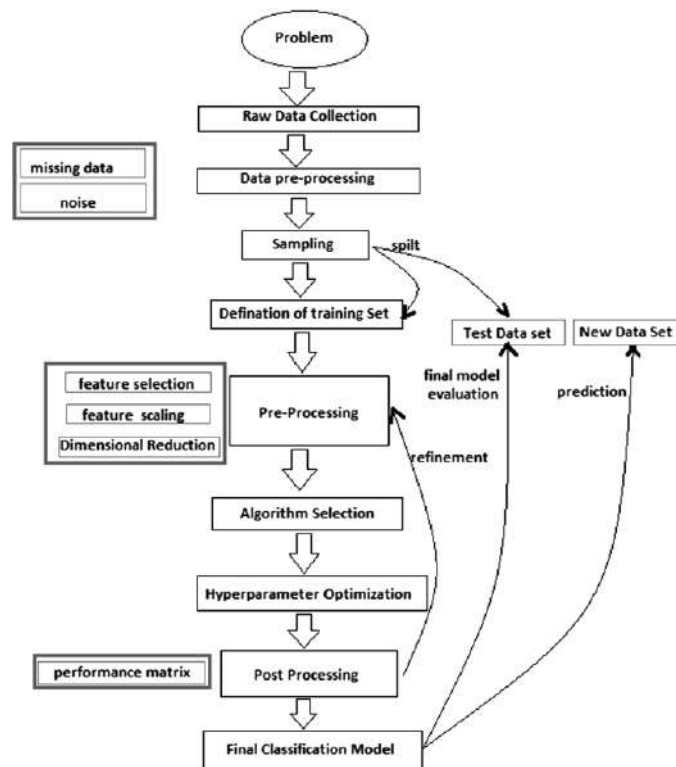


Γράφημα 3.2: Εφαρμογή Dropout σε Νευρωνικό δίκτυο  
 Πηγή: ResearchGate (2020)

### 3.4 Αλγόριθμοι Ταξινόμησης

Στον τομέα της μηχανικής εκμάθησης, οι αλγόριθμοι ταξινόμησης έχουν σημαντικό ρόλο στην επεξεργασία και στην οργάνωση των δεδομένων. Η ταξινόμηση αποτελεί ένα κύριο κομμάτι για την προεπεξεργασία των δεδομένων, για την κατηγοριοποίηση και την πρόβλεψη. Η λειτουργία τους εστιάζει στην ανάπτυξη ενός μοντέλου που μπορεί να εξάγει συμπεράσματα και να προβλέψει την κατηγορία ενός δείγματος. Οι αλγόριθμοι ταξινόμησης εφαρμόζουν τεχνικές ταξινόμησης σε ένα πλήθος χαρακτηριστικών των δεδομένων, με στόχο την αντιστοίχιση των χαρακτηριστικών σε ορισμένες κατηγορίες.

Οι αλγόριθμοι ταξινόμησης στην μηχανική εκμάθηση συνεισφέρουν στην οργάνωση των δεδομένων εκπαίδευσης σε κατηγορίες ή κλάσεις, έτσι ώστε το μοντέλο μηχανικής εκμάθησης να έχει την δυνατότητα να προβλέπει και να μάθει την κατηγορία του εισερχόμενου ή μη εξεταζόμενου δείγματος. Η επιλογή του κατάλληλου αλγόριθμου ταξινόμησης εξαρτάται από τα χαρακτηριστικά των δεδομένων, την έκταση του dataset και τις απαιτήσεις της εφαρμογής. Κατά την διάρκεια της εκπαίδευσης, το μοντέλο προσαρμόζεται στα δεδομένα εκπαίδευσης, ενώ το σύνολο ελέγχου αξιοποιείται για την εκτίμηση της ικανότητας του μοντέλου. Η επιλογή του κατάλληλου μοντέλου ταξινόμησης συμβάλλει στην καλύτερη επίδοση και ακρίβεια των μοντέλων μηχανικής εκμάθησης.

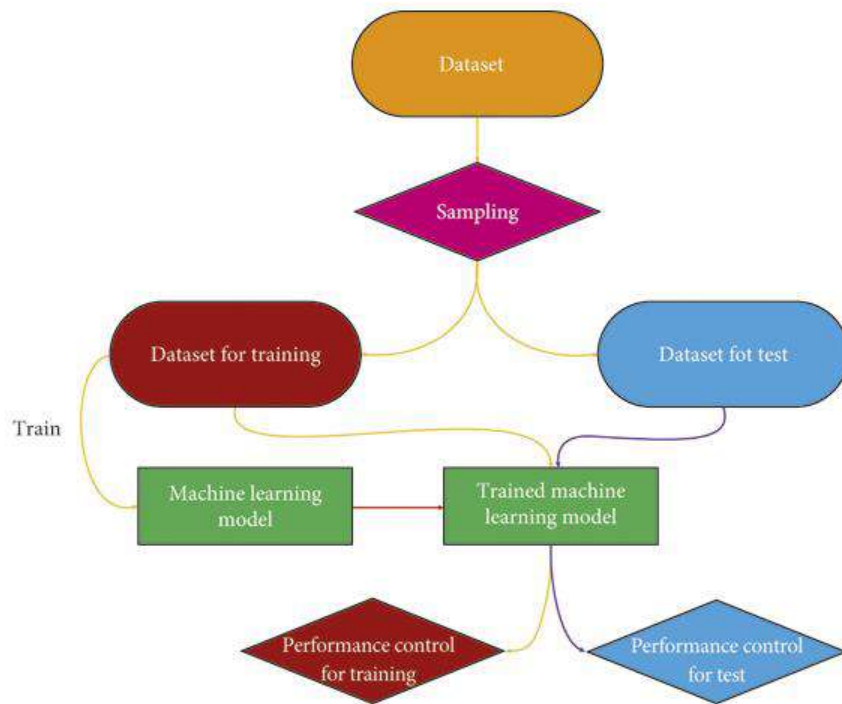


Γράφημα 3.3: Διάγραμμα ροής εργασίας εποπτευόμενου μοντέλου ταξινόμησης  
 Πηγή: Pratap Chandra Sen (2019)

### 3.4.1 Μοντέλα Μηχανικής Μάθησης

Η εποπτευόμενη μηχανική μάθηση (Supervised Machine Learning) είναι η αναζήτηση αλγορίθμων που συλλογίζονται από εξωτερικά παρεχόμενες περιπτώσεις για να παράγουν γενικές υποθέσεις, οι οποίες στη συνέχεια κάνουν προβλέψεις για μελλοντικές περιπτώσεις. Η εποπτευόμενη ταξινόμηση είναι μία από τις εργασίες που εκτελούνται συχνότερα από τα ευφυή συστήματα. Οι αλγόριθμοι μηχανικής μάθησης οργανώνονται σε μια ταξινόμηση με βάση το επιθυμητό αποτέλεσμα του αλγορίθμου. Η εποπτευόμενη μάθηση δημιουργεί μια συνάρτηση που αντιστοιχίζει τις εισόδους στις επιθυμητές εξόδους (Osisanwo et al., 2017).

Η διαδικασία εκμάθησης σε ένα απλό μοντέλο μηχανικής μάθησης χωρίζεται σε δύο βήματα: την εκπαίδευση και την δοκιμή. Αρχικά χωρίζονται τα δεδομένα σε δύο κατηγορίες, στα δεδομένα εκπαίδευσης (training data) και στα δεδομένα εξέτασης (testing data). Στη διαδικασία εκπαίδευσης, δείγματα σε δεδομένα εκπαίδευσης (training dataset) λαμβάνονται ως είσοδος στα οποία τα χαρακτηριστικά μαθαίνονται από τον αλγόριθμο μάθησης ή τον εκπαιδευόμενο και χτίζουν το μοντέλο μάθησης. Στη διαδικασία δοκιμής, το μοντέλο εκμάθησης χρησιμοποιεί τη μηχανή εκτέλεσης για να κάνει την πρόβλεψη για τη δοκιμή ή τα δεδομένα παραγωγής. Δηλαδή εκπαιδεύουμε το μοντέλο σε ένα σύνολο δεδομένων με ετικέτα (δεδομένα εκπαίδευσης) και στη συνέχεια χρησιμοποιούμε το μοντέλο για να προβλέψουμε ετικέτες σε δεδομένα χωρίς ετικέτα (δεδομένα δοκιμής).



Γράφημα 3.4: Διαδικασία Μηχανικής Εκμάθησης  
 Πηγή: Hindawi (2020)

### 3.4.1.1 Τυχαία Δάση (Random Forest)

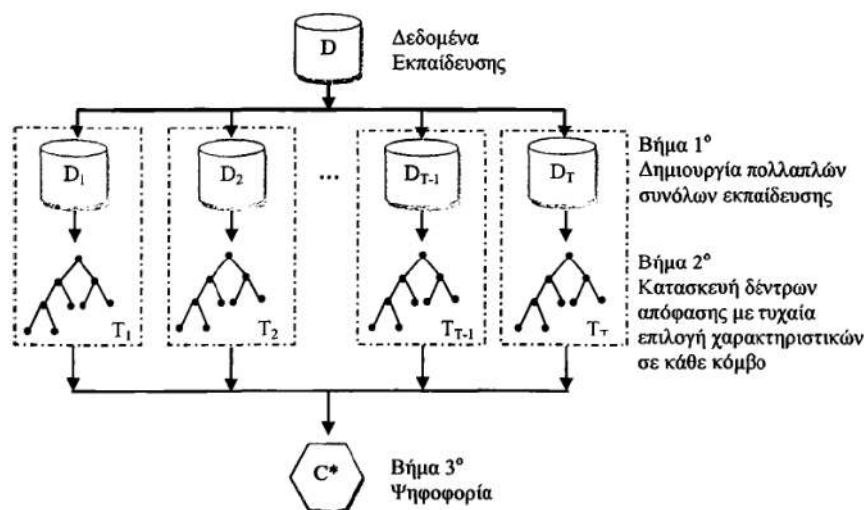
Τα δένδρα απόφασης (**Random Forest**) αποτελούν ευρέως διαδεδομένη τεχνική ταξινόμησης λόγω της απλότητας τους, της εύκολης κατανόησης και λόγω της ακρίβειας των ταξινομήσεων του. Έχουν δενδροειδή μορφή όμοια με τα διαγράμματα ροής και ο κάθε κόμβος χωρίζεται σε δύο μέρη με βάση την αλληλουχία αποφάσεων (Lugo Reyes, 2020). Τα βήματα που ακολουθούνται είναι τα εξής:

Αρχικοποίηση του κόμβου με το σύνολο δεδομένων.

Διάσπαση του κόμβου με βάση κάποιο κριτήριο διαχωρισμού σε κάποιο από τα γνωρίσματα.

Επανάληψη του βήματος 2 έως ότου ικανοποιηθεί το κριτήριο τερματισμού και τα δεδομένα έχουν ταξινομηθεί με βάση τα γνωρίσματα τους μέσω ενός συστήματος αποφάσεων.

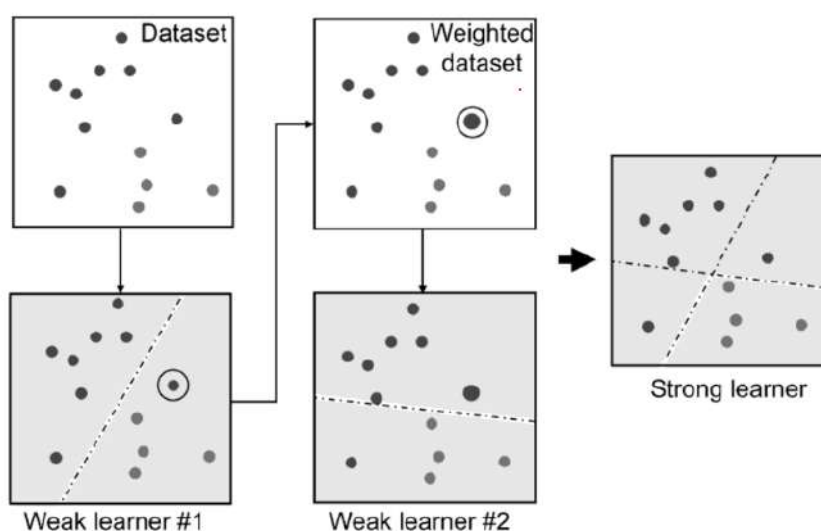
Ο δείκτης gini (gini index) και η εντροπία (entropy) αποτελούν τα κριτήρια υπολογισμού του κέρδους πληροφορίας. Οι αλγόριθμοι δένδρων απόφασης αξιοποιούν το κέρδος πληροφορίας για τον βέλτιστο αριθμό διαχωρισμών του κάθε κόμβου. Ο ταξινομητής Random Forest είναι μία μέθοδος συνόλου που εκπαιδεύει πολλά δένδρα απόφασης παράλληλα με το bootstrapping που ακολουθείτε από το aggregation, που αναφέρεται από κοινού ως τεχνική bagging (Misra and Li, 2020). Το bootstrapping υποδεικνύει ότι πολλά μεμονωμένα δένδρα αποφάσεων εκπαιδεύονται παράλληλα σε διάφορα υποσύνολα του συνόλου δεδομένων εκπαίδευσης χρησιμοποιώντας διαφορετικά υποσύνολα διαθέσιμων χαρακτηριστικών. Οι αποφάσεις των επιμέρους δένδρων απόφασης λαμβάνονται υπόψη για την τελική απόφαση του ταξινομητή.



Γράφημα 3.5: Η Διαδικασία Κατασκευής των Random Forests  
 Πηγή: Τσιάρα (2012)

### 3.4.1.2 Προσαρμοστική ενδυνάμωση (Adaboost)

Η προσαρμοστική ενδυνάμωση (**Adaboost**) είναι ένας αλγόριθμος μηχανικής εκμάθησης που χρησιμοποιείται για την ανάπτυξη ισχυρών ταξινομητών από αδύναμους. Ο αλγόριθμος λειτουργεί επαναληπτικά, ενισχύοντας την ικανότητα του ταξινομητή να επιλέξει τις σωστές απαντήσεις για κάθε παράδειγμα. Κατά την εκπαίδευση, ο Adaboost αναθέτει σημασία στις λανθασμένες ταξινομήσεις, εστιάζοντας έτσι στις περιπτώσεις που ο ταξινομητής αδυνατεί να αντιμετωπίσει. Η προσαρμοστική ενδυνάμωση έχει ευρεία εφαρμογή σε διάφορα προβλήματα μηχανικής εκμάθησης και παρουσιάζει πλεονεκτήματα όπως την αποτελεσματικότητα και τη δυνατότητα αντιμετώπισης συγκεκριμένων προβλημάτων (Ding et al., 2022).



Γράφημα 3.6: Λειτουργία του αλγορίθμου AdaBoost  
 Πηγή: Misra & Li (2020)

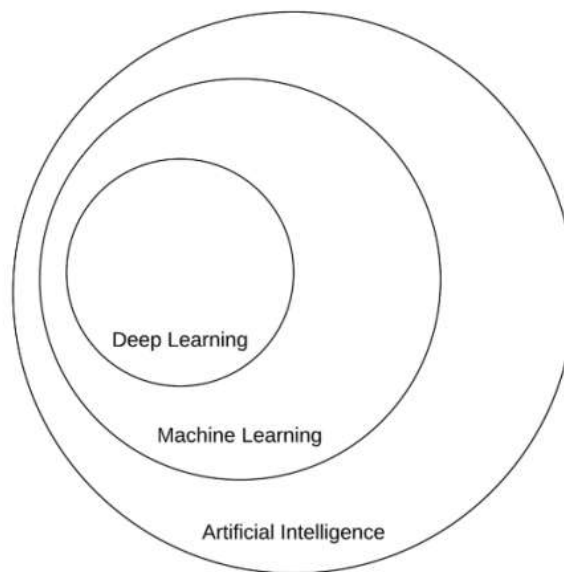
### 3.4.2 Μοντέλα Βαθιάς Μάθησης

Η Βαθιά μάθηση (Deep Learning), γνωστή και ως βαθιά δομημένη μάθηση (deep structured learning) ή ιεραρχική μάθηση (hierarchical learning), είναι ένα μέρος μίας ευρύτερης



κατηγορίας μεθόδων μηχανικής μάθησης βασιζόμενων σε αναπαραστάσεις δεδομένων μάθησης παρά σε αλγόριθμους συγκεκριμένων εργασιών (Goodfellow et al., 2016).

Στην ουσία είναι μία τεχνική που δίνει την ικανότητα στους υπολογιστές να προσεγγίσουν τον τρόπο σκέψης των ανθρώπων. Στην Βαθιά μάθηση το υπολογιστικό μοντέλο μαθαίνει να κατηγοριοποιεί διεργασίες απευθείας από εικόνες, κείμενο και ήχο. Τα μοντέλα βαθιάς εκμάθησης μπορούν να εξασφαλίσουν μεγάλη ακρίβεια και ακόμα μπορούν να ξεπεράσουν την ανθρώπινη επίδοση. Η λέξη *deep* στην φράση *Deep Learning* σχετίζεται με την ιδέα των διαδοχικών στρώσεων (*layers*) αναπαραστάσεων. Ο συνολικός αριθμός των στρώσεων που συμβάλουν σε ένα μοντέλο δεδομένων αποτελεί το βάθος του μοντέλου. Η βαθιά εκμάθηση συχνά μπορεί να αποτελείται από δεκάδες ή εκατοντάδες επίπεδα στρωμάτων τα οποία έχουν την ικανότητα να εκπαιδευτούν από την έκθεση σε δεδομένα εκπαίδευσης.



Γράφημα 3.7: Τεχνική νοημοσύνη, Μηχανική εκμάθηση, Βαθιά μάθηση

Τα νευρωνικά δίκτυα (**Artificial Neural Networks-ANN**) είναι υπολογιστικά συστήματα που βασίζονται σε βιολογικά νευρωνικά δίκτυα. Τα ANN είναι δίκτυα που αποτελούνται από κόμβους που είναι συνδεδεμένοι μεταξύ τους και μέσω αυτής της σύνδεσης στέλνουν σήματα ο ένας στον άλλον. Κάθε νευρώνας δέχεται ένα σύνολο αριθμητικών εισόδων από διαφορετικές πηγές όπως άλλοι νευρώνες ή το περιβάλλον, στην συνέχεια κάνει έναν υπολογισμό με βάση τις εισόδους και παράγει μια έξοδο, όπου με την σειρά της η έξοδος αυτή πηγαίνει στο περιβάλλον ή γίνεται είσοδος σε έναν άλλον νευρώνα. Οι νευρώνες χωρίζονται σε τρεις κατηγορίες:

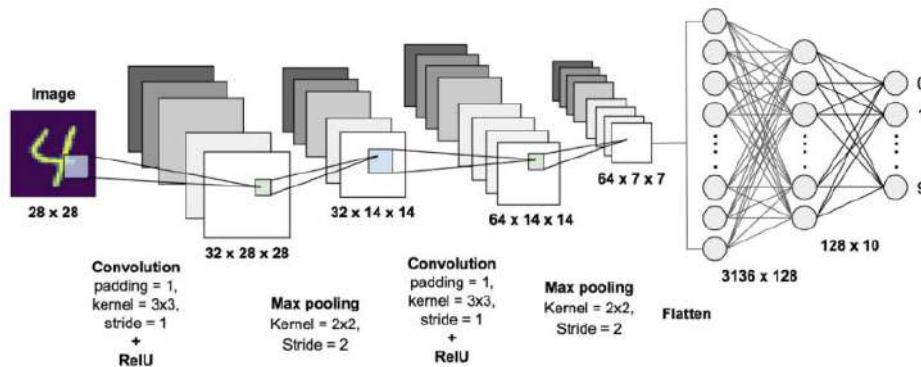
- Νευρώνες εισόδου
- Κρυφοί νευρώνες
- Νευρώνες εξόδου

#### 3.4.2.1 Συνελεκτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks)

Τα Συνελεκτικά Νευρωνικά Δίκτυα (**CNN**) διαχωρίζονται σε δύο μεγάλες κατηγορίες τα Αβαθή Νευρωνικά Δίκτυα (Shallow Neural Networks) και τα Βαθιά Νευρωνικά Δίκτυα (Deep

Neural Networks). Το συνελεκτικό επίπεδο είναι ένας αλγόριθμος βαθιάς μάθησης που λαμβάνει μία εικόνα ως είσοδο και μπορεί να αποδώσει σημασία και χαρακτηριστικά. Συγκεκριμένα εκτελεί ταξινόμηση που ειδικεύεται στην ανάλυση εικόνων αναγνωρίζοντας και αναλύοντας ειδικά μοτίβα κατά τη διάρκεια της εκπαίδευσης του.

Μια τυπική αρχιτεκτονική διάταξη ενός Συνελικτικού Νευρωνικού Δικτύου, για χρήση σε προβλήματα ταξινόμησης εικόνων παρουσιάζεται στην παρακάτω εικόνα.



Γράφημα 3.8: Διάταξη ενός Νευρωνικού Συνελικτικού Δικτύου  
 Πηγή: ResearchGate (2020)

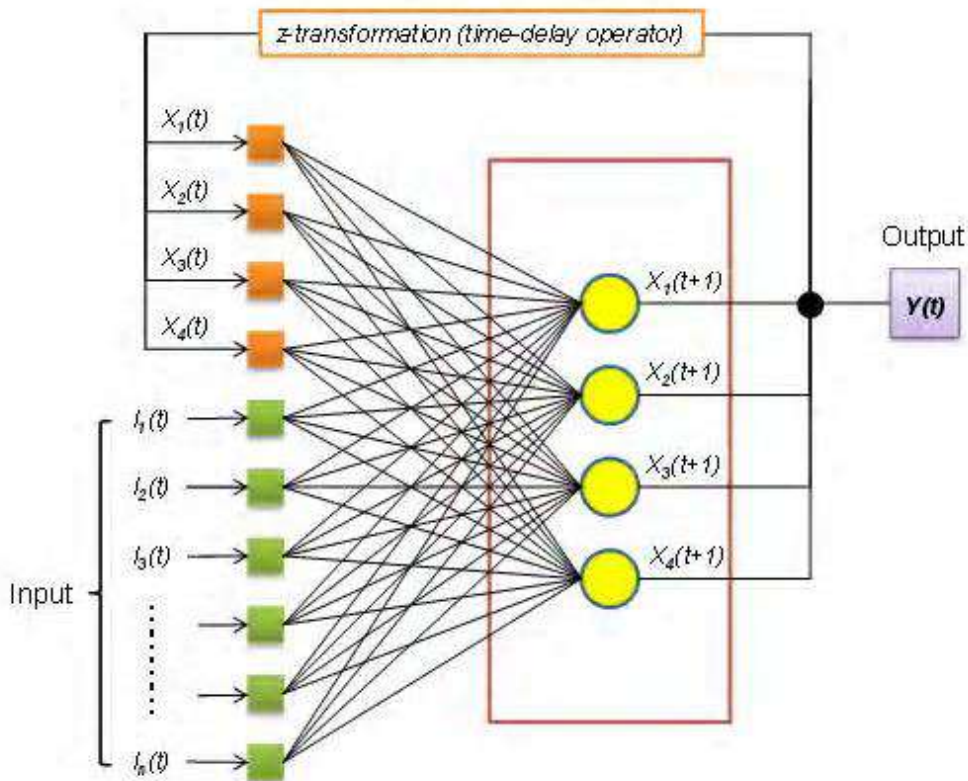
Τα βαθιά νευρωνικά δίκτυα (Deep Neural Networks, DNN) είναι νευρωνικά δίκτυα με πολλαπλά κρυφά επίπεδα. Ένα βαθύ δίκτυο δεν προβλέπει απλώς την έξοδο  $Y$  δεδομένης μίας εισόδου  $X$ , αλλά ταυτόχρονα αντιλαμβάνεται βασικά χαρακτηριστικά της μεταβλητής εισόδου  $X$ . Έχει την ικανότητα να απομνημονεύσει αφηρημένες έννοιες χαρακτηριστικών των παραδειγμάτων εισόδου, να κατανοήσει τα βασικά χαρακτηριστικά των παραδειγμάτων και να κάνει προβλέψεις με βάση αυτά τα χαρακτηριστικά.

### 3.4.2.2 Επαναλαμβανόμενο Νευρωνικό Δίκτυο (Recurrent Neural Network – RNN)

Τα επαναλαμβανόμενα νευρωνικά δίκτυα (**RNN**) αποτελούν μία κλάση του ANN όπου η σύνδεση των νευρώνων σχεδιάζουν ένα κατευθυνόμενο γράφημα κατά μία χρονική ακολουθία, το οποίο τους επιτρέπει να εμφανίζουν χρονική δυναμική συμπεριφορά. Τα RNN έχουν την ικανότητα να χρησιμοποιήσουν την μνήμη τους για να επεξεργαστούν ακολουθίες μεταβλητού μήκους εισόδων. Επομένως, καθίστανται εφαρμόσιμα σε εργασίες όπως η μη καταχωρισμένη, συνδεδεμένη αναγνώριση γραφικού χαρακτήρα ή αναγνώριση ομιλίας. Υπάρχουν δύο είδη Α) Πλήρως Επαναλαμβανόμενο νευρωνικό δίκτυο (Fully Recurrent Neural Networks- FRNN) και Β) Μακράς-Βραχείας Μνήμης (Long –Short Term Memory - LSTM).

#### Α) Πλήρως Επαναλαμβανόμενο νευρωνικό δίκτυο (Fully Recurrent Neural Networks- FRNN)

Τα FRNN συνδέουν τις εξόδους όλων των νευρώνων με τις εισόδους όλων των νευρώνων. Αυτή είναι η πιο γενική τοπολογία νευρικού δικτύου, επειδή όλες οι άλλες τοπολογίες μπορούν να αναπαρασταθούν θέτοντας κάποια βάρη στο μηδέν για προσομοίωση της έλλειψης συνδέσεων μεταξύ αυτών των νευρώνων.

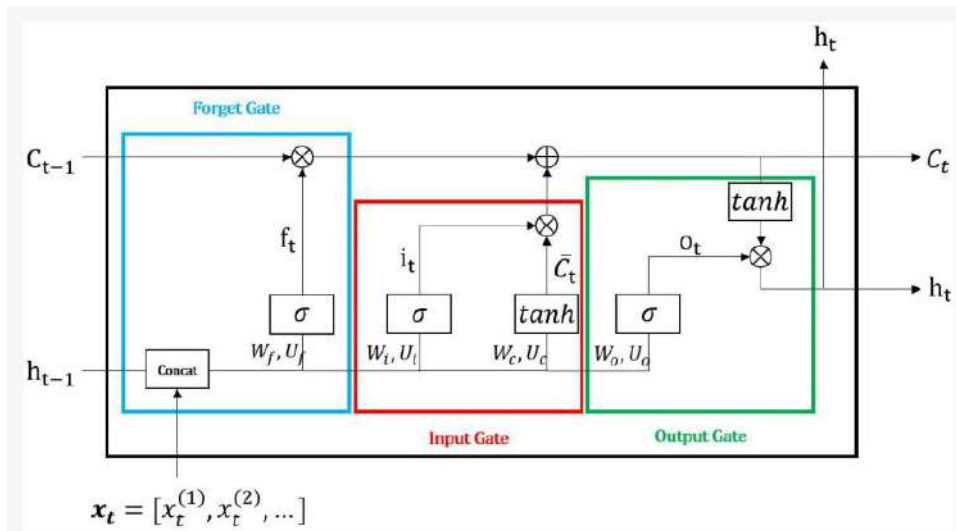


Γράφημα 3.9: Διάταξη ενός Πλήρους Επαναλαμβανόμενου Νευρωνικού Συνελεκτικού Δικτύου  
 Πηγή: ReseachGate (2020)

## B) Νευρωνικό Δίκτυο Μακράς-Βραχυπρόθεσμης Μνήμης (Long –Short Term Memory)

Το μοντέλο μακράς-βραχείας μνήμης (LSTM) αναφέρθηκε για πρώτη φορά το 1997 από τον Sepp Hoch Reiter και Jürgen Schmid Huber. Η προκαταρκτική έκδοση του μοντέλου LSTM ενσωμάτωσε κελιά, πύλες εισόδου και εξόδου. Είναι διάσημο για τον σχεδιασμό του, ο οποίος είναι ένας συγκεκριμένος τύπος επαναλαμβανόμενου νευρωνικού δικτύου (RNN). Το μεγαλύτερο πρόβλημα ενός RNN είναι ότι διατηρεί μόνο τις πληροφορίες της προηγούμενης κατάστασης, προκαλώντας το πρόβλημα της διαβάθμισης. Το LSTM έρχεται να λύσει το πρόβλημα, το οποίο κατασκευάστηκε για να αποφευχθεί το ζήτημα των μακροχρόνιων εξαρτήσεων. Η απομνημόνευση πληροφοριών για μεγάλο χρονικό διάστημα είναι βασικά η προεπιλεγμένη συμπεριφορά του. Το LSTM έχει επίσης μια πανομοιότυπη δομή αλυσίδας αντί να έχει ένα ενιαίο στρώμα νευρωνικού δικτύου (Hochreiter & Schmidhuber, 1997).

Τα νευρωνικά δίκτυα (LSTM) αποτελούν έναν τύπο επαναλαμβανόμενου νευρωνικού δικτύου που έχει σχεδιαστεί για την αντιμετώπιση του προβλήματος της εξαφάνισης της κλίσης που εμφανίζεται στα παραδοσιακά μοντέλα RNN. Το παραπάνω πρόβλημα εμφανίζεται όταν οι διαβαθμίσεις γίνονται πολύ μικρές για να είναι χρήσιμες για την ενημέρωση των βαρών κατά την οπισθοδιάδοση. Τα LSTM λύνουν αυτό το πρόβλημα εισάγοντας μια κυψέλη μνήμης που μπορεί να αποθηκεύσει πληροφορίες για μεγάλες χρονικές περιόδους, επιτρέποντας στο δίκτυο να διατηρεί σημαντικές πληροφορίες από προηγούμενα χρονικά βήματα.



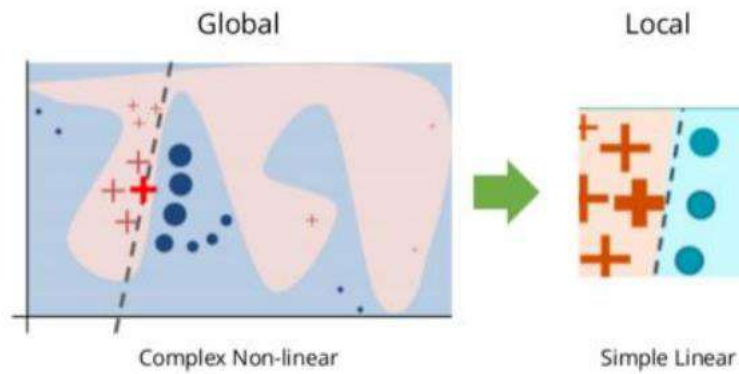
Γράφημα 3.10: Δομή του μοντέλου LSTM  
 Πηγή: Seong Kyung Kwong et al., 2021

### 3.5 Τοπικό Ερμηνευτικό Μοντέλο – Lime (Local Interpretable Model-agnostic explanations)

Το Lime είναι μία τεχνική που προσεγγίζει οποιοδήποτε μοντέλο μηχανικής εκμάθησης μαύρου κουτιού (black box) με ένα τοπικό, ερμηνεύσιμο μοντέλο για την επεξήγηση κάθε μεμονωμένης πρόβλεψης δημιουργώντας γραμμικά μοντέλα γύρω από κάθε μία από αυτές. Το Lime είναι μία μελέτη στην οποία οι συγγραφείς προτείνουν μία συγκεκριμένη εφαρμογή τοπικών υποκατάστατων μοντέλων, τα οποία εκπαιδεύονται ώστε να προσεγγίζουν τις προβλέψεις του υποκείμενου μοντέλου μαύρου κουτιού. Αντί να εκπαιδεύει ένα παγκόσμιο υποκατάστατο μοντέλο, το Lime εστιάζει στην εκπαίδευση των τοπικών υποκατάστατων μοντέλων για να εξηγήσει μεμονωμένες προβλέψεις.

Η μέθοδος Lime διαταράσσει τα αρχικά σημεία δεδομένων, τα τροφοδοτεί στο μοντέλο του «μαύρου κουτιού» και εξαγάγει τις προβλέψεις του μοντέλου μηχανικής εκμάθησης. Έπειτα, δημιουργείται ένα νέο σύνολο δεδομένων που περιλαμβάνει τα διαταραγμένα δείγματα και τις αντίστοιχες προβλέψεις του μαύρου κουτιού. Στο νέο σύνολο δεδομένων το Lime εκπαιδεύει στη συνέχεια ένα ερμηνεύσιμο μοντέλο, το οποίο σταθμίζεται με βάση την εγγύτητα των παρουσιών του δείγματος με το στιγμιότυπο του ενδιαφέροντος. Τελικά, προσαρμόζει ένα υποκατάστατο μοντέλο, όπως η γραμμική παλινδρόμηση (linear regression), στο σύνολο δεδομένων με διακυμάνσεις χρησιμοποιώντας αυτά τα βάρη του δείγματος.

Κάθε αρχικό σημείο δεδομένων μπορεί στη συνέχεια να εξηγηθεί με το πρόσφατα εκπαιδευμένο μοντέλο εξήγησης.



Γράφημα 3.11: Διάγραμμα Lime

Πηγή: <https://c3.ai/glossary/data-science/lime-local-interpretable-model-agnostic-explanations/>

Η βασική ιδέα του Lime είναι να επεξηγήσουμε μία πρόβλεψη ενός σύνθετου μοντέλου βαθιάς μάθησης, με την προσαρμογή ενός τοπικού υποκατάστατου μοντέλου, του οποίου οι προβλέψεις είναι εύκολο να εξηγηθούν. Ως εκ τούτου το Lime αναφέρεται ως τεχνική εξήγησης με βάση το υποκατάστατο. Επομένως, το LIME υφίσταται έναν συμβιβασμό μεταξύ της ακρίβειας του μοντέλου και της πολυπλοκότητας.

### 3.6 Μετρικές αξιολόγησης για ταξινόμηση (Evaluation metrics for classification)

#### 3.6.1 Μήτρα σύγχυσης (Confusion Matrix)

Τα προβλήματα δυαδικής ταξινόμησης και η αξιολόγηση διάκρισης της βέλτιστης λύσης κατά την διαδικασία εκπαίδευσης της παλινδρόμησης μπορεί να οριστεί με βάση την Μήτρα σύγχυσης.

Η γραμμή στον πίνακα αντιπροσωπεύει την προβλεπόμενη κλάση, ενώ η στήλη την πραγματική κλάση. Στην μήτρα σύγχυσης τα  $tp$  και τα  $tn$  δηλώνουν τον αριθμό των θετικών και αρνητικών περιπτώσεων που ταξινομούνται σωστά. Τα  $fp$  και  $fn$  δηλώνουν το αριθμό των λανθασμένων αρνητικών και θετικών περιπτώσεων αντίστοιχα.

	Actual Positive Class	Actual Negative Class
Predicted Positive Class	True positive ( $tp$ )	False negative ( $fn$ )
Predicted Negative Class	False positive ( $fp$ )	True negative ( $tn$ )

Γράφημα 3.12: Πίνακας σύγχυσης για το πρόβλημα δυαδικής ταξινόμησης

Πηγή: Hossin et al. (2015)

Οι διαφορές μεταξύ των προβλεπόμενων και των πραγματικών κλάσεων αναπαρίστανται στην μήτρα σύγχυσης.

Τα τέσσερα πιθανά αποτελέσματα είναι τα εξής:

- ‘Πραγματικά Θετικά’ (True Positive): Το σημείο δεδομένων είναι πραγματικά αληθινό ( $tp$ ) όταν προβλέπεται ένα θετικό αποτέλεσμα και προκύπτει θετικό.
- ‘Ψευδώς Θετικά’ (False Positive): Το σημείο δεδομένων είναι ψευδώς αληθινό ( $fp$ ) όταν προβλέπεται ένα θετικό αποτέλεσμα και προκύπτει ένα αρνητικό αποτέλεσμα. Αυτό το σενάριο είναι γνωστό ως Σφάλμα τύπου 1.



### 3.6.3 Ακρίβεια (Precision)

Η ακρίβεια (precision) χρησιμοποιείται για τη μέτρηση των θετικών μοτίβων που προβλέπονται σωστά από τα συνολικά προβλεπόμενα μοτίβα σε μια θετική τάξη.

$$\text{Ακρίβεια (Precision)} = \frac{TP}{TP+FP} \quad (3.8)$$

### 3.6.4 Ανάκληση (Recall)

Η ανάκληση (recall) χρησιμοποιείται για τη μέτρηση του κλάσματος των θετικών προτύπων που ταξινομούνται σωστά.

$$\text{Ανάκληση (Recall)} = \frac{TP}{TP+TN} \quad (3.9)$$

Στην παρούσα εργασία οι επιπτώσεις της λανθασμένης ταξινόμησης μίας επικίνδυνης κλάσης σαν λιγότερο επικίνδυνη ή ασφαλή θα σημειώνουν σημαντικές επιπτώσεις. Επομένως, ο τύπος σφάλματος 2 είναι κρισιμότερος από τον τύπο σφάλματος 1 και καθιστά την μετρική ανάκλασης ιδιαίτερα σημαντική.

### 3.6.5 Ρυθμός λανθασμένων θετικών προβλέψεων (False positive rate)

Ο ρυθμός λανθασμένων θετικών προβλέψεων (False positive rate, FPR) ή ρυθμός λανθασμένου συναγερμού (False alarm rate, FAR) υπολογίζεται ως ο αριθμός των ψευδώς θετικών προβλέψεων (FP) διαιρεμένος με τον συνολικό αριθμό των αρνητικών (N). Το καλύτερο ποσοστό λανθασμένων θετικών προβλέψεων είναι 0,0 και το χειρότερο είναι 1,0.

$$\text{FPR ή FAR} = \frac{FP}{FP+TN} \quad (3.10)$$

### 3.6.6 f1-score

Το f1-score αποτελεί τον αρμονικό μέσο όρο μεταξύ της ακρίβειας και της ανάκλησης.

$$f1 - score = \frac{2x(\text{Ακρίβεια})x(\text{Ανάκληση})}{(\text{Ακρίβεια})+(\text{Ανάκληση})} \quad (3.11)$$

### 3.6.7 Εξειδικευτικότητα (Specificity)

Υπολογίζει ποιο κλάσμα όλων των αρνητικών δειγμάτων προβλέπονται σωστά ως αρνητικά από τον ταξινομητή. (Khanna, Mohit et al., 2021).

$$\text{specificity} = \frac{TN}{TN + FP}$$

### 3.7.8 Δείκτης λάθος συναγερμού (False alarm rate)

Υπολογίζεται ως ο αριθμός των λανθασμένων θετικών προβλέψεων δια του συνολικού αριθμού των αρνητικών. (Sklearn.Metrics.F1\_score , 2022).

$$\text{False Alarm Rate} = \frac{FP}{TN + FP}$$

### 3.6.9 Μακροοικονομικός μέσος όρος (MacroAverage)

Χρησιμοποιείτε όταν όλες οι κατηγορίες πρέπει να αντιμετωπίζονται ισότιμα για να αξιολογήσετε τη συνολική απόδοση του ταξινομητή σε σχέση με τις πιο συχνές ετικέτες κλάσεων. (Micro-Average & Macro-Average Scoring Metrics – Python, 2020).

$$F1_{class1} + F1_{class2} + \dots + F1_{classN}$$

### 3.6.10 Σταθμισμένος μέσος όρος (WeightedAverage)

Είναι ένας υπολογισμός που λαμβάνει υπόψη τη σχετική τιμή των ακεραίων σε μια συλλογή δεδομένων. Κάθε τιμή στο σύνολο δεδομένων κλιμακώνεται με ένα προκαθορισμένο βάρος πριν ολοκληρωθεί ο τελικός υπολογισμός κατά τον υπολογισμό ενός σταθμισμένου μέσου όρου. (Micro-Average & Macro-Average Scoring Metrics – Python, 2020).

$$F1_{class1} * W1 + F1_{class2} * W2 + \dots + F1_{classN} * WN$$

## 4. ΣΥΛΛΟΓΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΣΤΟΙΧΕΙΩΝ

### 4.1 Εισαγωγή

Σύμφωνα με το ερευνητικό έργο i-Dreams στόχος του είναι ο ορισμός, η ανάπτυξη, η δοκιμή και η επικύρωση μίας 'Ζώνης Ανοχής Ασφαλείας' ώστε να περιορίζεται η επικίνδυνη συμπεριφορά κατά την οδήγηση μέσω παρεμβάσεων σε πραγματικό χρόνο ή μεταγενέστερο χρόνο. Η μελέτη συγκεκριμένων δεδομένων της οδηγικής συμπεριφοράς και του οδικού περιβάλλοντος καίριο βήμα για την επίτευξη των παραπάνω στόχων. Με την μελέτη των δεδομένων της έρευνας επιδιώκεται:

- A. Ο καθορισμός του επιπέδου 'Ζώνης Ανοχής Ασφαλείας' που βρίσκεται ο οδηγός σε πραγματικό χρόνο με στόχο την πρόκληση παρεμβάσεων.
- B. Η αναγνώριση της σχέσης μεταξύ του κινδύνου και των άμεσα σχετιζόμενων με αυτών παραγόντων. Επιδιώκεται η καλύτερη κατανόηση των παραγόντων της οδηγικής συμπεριφοράς και κατ' επέκταση την βελτίωση των παρεμβάσεων.

### 4.2 Πείραμα σε πραγματικές συνθήκες οδήγησης

#### 4.2.1 Στόχος πειράματος

Στο πλαίσιο του ερευνητικού προγράμματος i-DREAMS, διεξήχθη ένα νατουραλιστικό πείραμα οδήγησης στο οποίο συμμετείχαν 43 οδηγοί από το Βέλγιο και δημιουργήθηκε μια μεγάλη βάση δεδομένων με 7163 ταξίδια και 147.337 λεπτά. Για τους οδηγούς του Ηνωμένου Βασιλείου, το σύνολο δεδομένων περιλάμβανε 26 οδηγούς, 8226 ταξίδια και 118.175 λεπτά. Στόχος του πειράματος ήταν η συλλογή δεδομένων σχετιζόμενων με την οδηγική συμπεριφορά και το οδικό περιβάλλον προκειμένου να ακολουθήσει η ανάλυση τους για την επίτευξη των στόχων που έχουν τεθεί.

Όπως φαίνεται στον παρακάτω πίνακα, το δοκιμαστικό πείραμα στο δρόμο πραγματοποιήθηκε σε τέσσερις φάσεις.



Πίνακας 4.1: Φάσεις πειράματος

<b>Φάση 1</b>		
<b>Χώρα</b>	<b>Οδηγοί</b>	<b>Ταξίδια (λεπτά)</b>
Βέλγιο	39	1173 trips (23,725 min)
Αγγλία	25	618 trips (10,803 min)
<b>Φάση 2</b>		
<b>Χώρα</b>	<b>Οδηγοί</b>	<b>Ταξίδια (λεπτά)</b>
Βέλγιο	43	1549 trips (31,414 min)
Αγγλία	26	2243 trips (25,151 min)
<b>Φάση 3</b>		
<b>Χώρα</b>	<b>Οδηγοί</b>	<b>Ταξίδια (λεπτά)</b>
Βέλγιο	51	1973 trips (40,121 min)
Αγγλία	26	2198 trips (24,569 min)
<b>Φάση 4</b>		
<b>Χώρα</b>	<b>Οδηγοί</b>	<b>Ταξίδια (λεπτά)</b>
Βέλγιο	49	2468 trips (52,077 min)
Αγγλία	26	3167 trips (57,652 min)

Το πείραμα διεξήχθη σύμφωνα με καθιερωμένες αρχές, με ιδιαίτερη έμφαση στην αξιολόγηση παρεμβάσεων που έχουν σχεδιαστεί για να ενισχύσουν την τήρηση πρακτικών ασφαλούς οδήγησης. Το πείραμα είχε διάρκεια τεσσάρων μηνών και πραγματοποιήθηκε σε τέσσερις φάσεις. Η φάση 1, που ορίστηκε ως φάση παρακολούθησης, διήρκεσε 4 εβδομάδες και δεν περιλάμβανε παρεμβάσεις. Η φάση 2, η οποία διήρκεσε επίσης 4 εβδομάδες, εισήγαγε παρεμβάσεις εντός του οχήματος με την παροχή προειδοποιήσεων σε πραγματικό χρόνο μέσω προσαρμοστικών προηγμένων συστημάτων υποστήριξης οδηγού (ADAS). Στη φάση 3, που διήρκεσε επίσης 4 εβδομάδες, οι οδηγοί έλαβαν ανατροφοδότηση σχετικά με τις επιδόσεις τους στην οδήγηση μέσω μιας εφαρμογής για κινητά τηλέφωνα. Στη Φάση 4, διάρκειας 6 εβδομάδων, οι οδηγοί συνέχισαν να λαμβάνουν ανατροφοδότηση όπως στη Φάση 3, αλλά με την πρόσθετη ενσωμάτωση στοιχείων παιχνιδιοποίησης. Και οι τέσσερις φάσεις επικεντρώθηκαν στην παρακολούθηση της οδигικής συμπεριφοράς και στην αξιολόγηση του αντίκτυπου των παρεμβάσεων σε πραγματικό χρόνο, συμπεριλαμβανόμενων των προειδοποιήσεων εντός του οχήματος, καθώς και των παρεμβάσεων μετά την οδήγηση, όπως η ανατροφοδότηση και η παιχνιδιοποίηση, στην οδигική συμπεριφορά.

Καθώς το βασικό αποτέλεσμα του έργου i-Dreams είναι ένα ολοκληρωμένο σύνολο εργαλείων παρακολούθησης και επικοινωνίας για παρέμβαση και υποστήριξη, χρησιμοποιήθηκαν τεχνολογίες και συστήματα αιχμής για την παρακολούθηση των δεικτών απόδοσης της οδήγησης. Συγκεκριμένα μια συσκευή OBD-II που είναι εγκατεστημένη σε κάθε όχημα για να καταγράφει τις εκατοντάδες παραμέτρους που μπορούν να αντληθούν από ένα σύγχρονο όχημα. Επιπλέον, με την αξιοποίηση του συστήματος Mobileye που βασίζεται στο δίκτυο κινητής τηλεφωνίας, γίνεται συλλογή δεδομένων χωρίς την συμμετοχή του χρήστη.

#### **4.2.2 ECU & On-Board Diagnostics (OBD II)**

Τα αρχικά ECU προκύπτουν από το Engine Control Unit. Η ECU αποτελεί τον εγκέφαλο του αυτοκινήτου, δηλαδή τη μονάδα εκείνη που επεξεργάζεται τα διάφορα δεδομένα του οχήματος και είναι υπεύθυνη για τις ενδείξεις στο καντράν αλλά και για λειτουργίες, όπως τα συστήματα ESP και ABS καθώς και των αισθητήρων παρκαρίσματος.

Για την επικοινωνία χρησιμοποιείται το σύστημα OBD. Το OBD αποτελεί ένα σύστημα επικοινωνίας των οχημάτων και των βιομηχανικών για επικοινωνία με τον έξω κόσμο. Πιο συγκεκριμένα, το OBD είναι γλώσσα της μονάδας ελέγχου του κινητήρα (ECU) και έχει σχεδιαστεί για να βοηθήσει στην αντιμετώπιση των εκπομπών και των κινητήρων.

Μια συσκευή OBD-II που υποστηρίζει όλα τα πρωτόκολλα OBD-II είναι εγκατεστημένη σε κάθε όχημα. Ένα σύγχρονο όχημα υποστηρίζει εκατοντάδες παραμέτρους, οι οποίες καταγράφονται από τη συσκευή OBD-II που φιλοξενεί το κατάλληλο Εργαλείο ανάπτυξης λογισμικού (SDK) για την εξαγωγή των απαραίτητων δεδομένων καθώς και ένα πλούσιο σύνολο API (Application Programming Interfaces) για επικοινωνία με συστήματα τρίτων. Αυτό το OBD-II ενσωματώνει την τεχνολογία 2G ή 3G GSM/GPRS μέσω της οποίας όλα τα δεδομένα που καταγράφονται από το όχημα μέσω των αισθητήρων του μεταδίδονται σε

απομακρυσμένους διακομιστές (Cloud). Το δίκτυο κινητής τηλεφωνίας χρησιμοποιείται για τη μετάδοση δεδομένων χωρίς καμία συμμετοχή χρήστη.

#### 4.2.3 Σύστημα Πρόληψης Ατυχημάτων MobilEye (G4S)

Τα συστήματα Mobileye είναι συστήματα, τα οποία με την χρήση μίας κάμερας, ειδοποιούν οπτικά και ηχητικά τον οδηγό σε περιπτώσεις αλλαγής λωρίδας χωρίς φλας, ανίχνευσης πεζού ή αντικειμένου στο δρόμο και σε περιπτώσεις πιθανής σύγκρουσης με το προπορευόμενο όχημα, δηλαδή είναι ως ένα δίκτυο αισθητήρων που μετρά παραμέτρους, όπως η απόσταση προόδου.



Εικόνα 4.1: Σύστημα Mobileye  
Πηγή: Γιαννόπουλος (2020)

Πιο συγκεκριμένα, χρησιμοποιήθηκαν δεδομένα από το σύστημα Mobileye (Mobileye, 2022), μια κάμερα ταμπλό και την πύλη Cardio (CardioID Technologies, 2022) το οποίο καταγράφει την οδηγική συμπεριφορά (π.χ. ταχύτητα, επιτάχυνση, επιβράδυνση, τιμόνι) μαζί με σήματα GNSS. Συγκεκριμένα, το σύστημα Mobileye είναι ως αισθητήρας δικτύου και ένα σύστημα που βασίζεται σε κάμερα που είναι τοποθετημένη στο παρμπρίζ, που μετρά παραμέτρους, όπως παρακολούθηση ανόδου, παρακολούθηση θέσης λωρίδας, αναγνώριση σημάτων κυκλοφορίας και αναγνώριση πεζών.

Το σύστημα μπορεί να συνδεθεί με τον δίαυλο CAN και επιτρέπει την ενσωμάτωση με πολλά προϊόντα οικοσυστήματος ADAS. Το Cardio gateway είναι ένα σύστημα που βασίζεται σε αισθητήρες που συνδέεται με τον εξοπλισμό Mobileye μέσω του διαύλου CAN του οχήματος και μπορεί να μεταφέρει δεδομένα μέσω διαφορετικών τεχνολογιών επικοινωνίας (BLE, CAN, I2C, SPI, WiFi). Πληροφορίες σχετικά με το τρέχον στάδιο προειδοποίησης, όπως ορίζεται από το Mobileye, συλλέχθηκαν επίσης για σύγκριση με το προειδοποιητικό στάδιο i-DREAMS (δηλαδή κανονική οδήγηση, φάση κινδύνου, φάση αποφυγής ατυχήματος). Παράλληλα, συλλέχθηκαν πληροφορίες για την τρέχουσα κατάσταση της πλατφόρμας i-DREAMS.

Η θεμελιώδης πρόκληση στο έργο i-DREAMS είναι ο τρόπος με τον οποίο οι επεξηγηματικές μεταβλητές (δηλαδή μετρήσεις απόδοσης διαφόρων μεταβλητών και δείκτες πολυπλοκότητας εργασιών και ικανότητας αντιμετώπισης) συσχετίζονται με την εξαρτημένη μεταβλητή «κίνδυνος» προκειμένου να προβλεφθεί η STZ.

#### 4.2.4 Στοιχεία που συλλέχθηκαν από το πείραμα

Υπάρχουν τρία κύρια στοιχεία της φύσης των μεταβλητών που χρησιμοποιούνται στο i-DREAMS:

- Διακριτές μεταβλητές: μεταβλητές που είναι κατηγορικές (τακτική ή ονομαστική) και μπορούν να λάβουν διακριτές τιμές μόνο από τους πραγματικούς αριθμούς. Μερικά παραδείγματα διακριτών μεταβλητών στο i-DREAMS θα μπορούσαν να είναι η κόπωση (ναι, όχι), η ώρα της ημέρας (ημέρα, νυχτερινή οδήγηση) και STZ (κανονική φάση, φάση κινδύνου, φάση αποφυγής ατυχήματος).
- Συνεχείς μεταβλητές: μεταβλητές που μπορούν να πάρουν οποιοσδήποτε τιμές από τους πραγματικούς αριθμούς. Μερικά παραδείγματα συνεχών μεταβλητών στο i-DREAMS θα μπορούσαν να είναι μεταβλητές ταχύτητας, πρόοδος και σύνθετες μεταβλητές, όπως μεταβλητές σταθμισμένου αθροίσματος ή σταθμισμένου μέσου όρου.
- Λανθάνουσες μεταβλητές: μεταβλητές που δεν είναι μπορούν να παρατηρηθούν εύκολα από τον αναλυτή και επομένως δεν είναι γνωστό αν είναι συνεχείς ή διακριτές. Παραδείγματα λανθανόντων μεταβλητών στο i-DREAMS είναι η πολυπλοκότητα των εργασιών και η ικανότητα αντιμετώπισης που είναι λανθάνουσες επεξηγηματικές μεταβλητές και επομένως απαιτούνται δείκτες που μπορούν να παρατηρηθούν για τη μέτρηση αυτών των λανθάνουσας μεταβλητής. Ο κίνδυνος επίσης αρχικά συλλαμβάνεται στο i-DREAMS ως μια λανθάνουσα μεταβλητή.

Συγκεκριμένα, οι κύριοι παράγοντες κινδύνου που θα διερευνηθούν στο πλαίσιο του έργου i-DREAMS είναι:

- Επιτάχυνση
- Προώθηση
- Προσπέραση
- Κόπωση
- Δύσκολες επιταχύνσεις
- Δυνατό φρενάρισμα
- Συμβάντα ελέγχου οχήματος (συνδυασμός σκληρής επιτάχυνσης, πέδησης και στροφών)

Στον πίνακα 4.1 παρέχεται μία επισκόπηση των μεταβλητών που θα εξεταστούν μαζί με την αντίστοιχη περιγραφή τους.

Πίνακας 4.2: Επισκόπηση μεταβλητών που συλλέχθηκαν

---

Μεταβλητή	Περιγραφή	Μονάδες μέτρησης	Τύπος
-----------	-----------	------------------	-------

---

ME_Car_speed	Ταχύτητα οχήματος	χιλιόμετρα ανά ώρα	αριθμητική
ME_AWS_hw_measurement	Μέτρηση της μεταβλητής headway	Δευτερόλεπτα	αριθμητική
ME_AWS_fcw	Προειδοποίηση μπροστινής σύγκρουσης		διακριτή
ME_AWS_pcw	Προειδοποίηση σύγκρουσης πεζών		διακριτή
ME_AWS_ldw	Προειδοποίηση αναχώρησης από τη λωρίδα		διακριτή
ME_AWS_pedestrian_dz	Πεζός σε επικίνδυνη ζώνη		διακριτή
GPS_distances	Συνολική απόσταση του ταξιδιού	Χιλιόμετρα	αριθμητική
GPS_spd	Ταχύτητα	Χιλιόμετρα ανά ώρα	αριθμητική
ME_TSR_tsr_1_speed	Εμφάνιση κωδικού πινακίδας 1 ταχύτητας		διακριτή
ME_AWS_time_indicator	Υποδεικνύει συνθήκες φωτισμού		διακριτή
DEM_evt_ha_lvl_M	Δύσκολα γεγονότα επιτάχυνσης μεσαίου επιπέδου		διακριτή
DEM_evt_hb_lvl_M	Δυνατό φρενάρισμα μεσαίου επιπέδου		διακριτή
DEM_evt_hc_lvl_M	Δύσκολες στροφές μεσαίου επιπέδου		διακριτή
DEM_evt_ha_lvl_L	Δύσκολα γεγονότα επιτάχυνσης χαμηλού επιπέδου		διακριτή
DEM_evt_hb_lvl_L	Δυνατό φρενάρισμα χαμηλού επιπέδου		διακριτή

DEM_evt_hc_lvl_L	Δύσκολες στροφές χαμηλού επιπέδου		διακριτή
DEM_evt_ha_lvl_H	Δύσκολα γεγονότα επιτάχυνσης υψηλού επιπέδου		διακριτή
DEM_evt_hb_lvl_H	Δυνατό φρενάρισμα υψηλού επιπέδου		διακριτή
DEM_evt_hc_lvl_H	Δύσκολες στροφές υψηλού επιπέδου		διακριτή
ME_AWS_time_indicator	Υποδεικνύει συνθήκες φωτισμού (ημέρα, σούρουπο, νύχτα)		διακριτή
ME_Car_wipers	Υποδεικνύει καιρικές συνθήκες (υαλοκαθαριστήρες on/off)		διακριτή
ME_Car_high_beam	Μεγάλη σκάλα		διακριτή
IBI_value	Χρονικό διάστημα μεταξύ διαδοχικών καρδιακών παλμών	χιλιοστά του δευτερολέπτου	αριθμητική
ME_LDW_Map_type_L	Προειδοποίηση αναχώρησης από την αριστερή λωρίδα		διακριτή
ME_LDW_Map_type_R	Προειδοποίηση αναχώρησης από την δεξιά λωρίδα		διακριτή

### 4.3 Επεξεργασία στοιχείων

Τα δεδομένα που συλλέχθηκαν από τους οδηγούς συγκεντρώθηκαν σε δύο .csv αρχεία και η πληροφορία σχετικά με την περιοχή που διεξήχθη το πείραμα (Βέλγιο ή Αγγλία) αναφέρεται στον τίτλο του αρχείου.

Αξιολογήθηκαν οι επεξηγηματικές μεταβλητές κινδύνου και οι πιο αξιόπιστοι δείκτες της πολυπλοκότητας της εργασίας και της ικανότητας αντιμετώπισης, όπως ο χρόνος προόδου, η απόσταση που διανύθηκε, η πρόσθια σύγκρουση ή οι καιρικές συνθήκες. Ιδιαίτερη έμφαση δόθηκε στη μέση ταχύτητα και στη χρονική απόσταση από το προπορευόμενο όχημα. Έπειτα δημιουργήθηκε μια νέα μεταβλητή, η STZ\_speed και STZ\_headway, η οποία λαμβάνει υπόψη τα διαφορετικά επίπεδα STZ για την ταχύτητα και τη χρονική απόσταση

από το προπορευόμενο όχημα. Έτσι, οι εξαρτημένες μεταβλητές ήταν το επίπεδο του STZ για την ταχύτητα (δηλαδή, STZ\_speed) και το επίπεδο του STZ για τη χρονική απόσταση από το προπορευόμενο όχημα (δηλαδή, STZ\_headway), χωρισμένες σε τρία επίπεδα (δηλαδή, Φάση Κανονικής Οδήγησης: 0, Επικίνδυνη φάση: 1, Φάση Αποφυγής Ατυχήματος: 2)

Προκειμένου να απλοποιηθεί η διαδικασία τα δεδομένα μορφοποιήθηκαν σε διαστήματα των 30 δευτερολέπτων. Συγκεκριμένα, για κάθε 30 δευτερόλεπτα υπολογίστηκαν τα περιγραφικά στατιστικά κάθε μεταβλητής όπως η μέση τιμή, η τυπική απόκλιση, η ελάχιστη τιμή, η μέγιστη τιμή και η διάμεσος

Στον πίνακα 4.2 παρατίθενται οι συγκεντρωμένες μεταβλητές των 30 δευτερολέπτων που προέκυψαν από το παραπάνω βήμα.

Πίνακας 4.3: Περιγραφή μεταβλητών μετά την επεξεργασία που αφορούν σε διαστήματα των 30 δλ.

Μεταβλητή	Περιγραφή	Μονάδες μέτρησης	Τύπος
Headway	Χρονική απόσταση από το προπορευόμενο όχημα	δευτερόλεπτα	αριθμητική
ME_Car_speed_mean	Ταχύτητα οχήματος	χιλιόμετρα ανά ώρα	αριθμητική
ME_AWS_hw_measurement_mean	Μέτρηση της μεταβλητής headway	δευτερόλεπτα	αριθμητική
ME_AWS_fcw_mean	Προειδοποίηση μπροστινής σύγκρουσης		διακριτή
ME_AWS_pcw_mean	Προειδοποίηση σύγκρουσης πεζών		διακριτή
ME_AWS_pedestrian_dz_mean	Πεζός σε επικίνδυνη ζώνη		διακριτή
GPS_distances_sum	Συνολική απόσταση του ταξιδιού	χιλιόμετρα	αριθμητική
GPS_spd_mean	Ταχύτητα	Χιλιόμετρα ανά ώρα	αριθμητική

ME_TSR_tsr_1_speed_median	Εμφάνιση κωδικού πινακίδας 1 ταχύτητας	διακριτή
ME_AWS_time_indicator_median	Υποδεικνύει συνθήκες φωτισμού	διακριτή
DEM_evt_ha_lvl_M_mean	Δύσκολα γεγονότα επιτάχυνσης μεσαίου επιπέδου	διακριτή
DEM_evt_hb_lvl_M_mean	Δυνατό φρενάρισμα μεσαίου επιπέδου	διακριτή
DEM_evt_hc_lvl_M_mean	Δύσκολες στροφές μεσαίου επιπέδου	διακριτή
DEM_evt_ha_lvl_L_mean	Δύσκολα γεγονότα επιτάχυνσης χαμηλού επιπέδου	διακριτή
DEM_evt_hb_lvl_L_mean	Δυνατό φρενάρισμα χαμηλού επιπέδου	διακριτή
DEM_evt_hc_lvl_L_mean	Δύσκολες στροφές χαμηλού επιπέδου	διακριτή
DEM_evt_ha_lvl_H_mean	Δύσκολα γεγονότα επιτάχυνσης υψηλού επιπέδου	διακριτή
DEM_evt_hb_lvl_H_mean	Δυνατό φρενάρισμα υψηλού επιπέδου	διακριτή
DEM_evt_hc_lvl_H_mean	Δύσκολες στροφές υψηλού επιπέδου	διακριτή
DEM_evt_ha_lvl_M_sum	Δύσκολα γεγονότα επιτάχυνσης μεσαίου επιπέδου	διακριτή
DEM_evt_hb_lvl_M_sum	Δυνατό φρενάρισμα μεσαίου επιπέδου	διακριτή
DEM_evt_hc_lvl_M_sum	Δύσκολες στροφές μεσαίου επιπέδου	διακριτή



DEM_evt_ha_lvl_L_sum	Δύσκολα γεγονότα επιτάχυνσης χαμηλού επιπέδου		διακριτή
DEM_evt_hb_lvl_L_sum	Δυνατό φρενάρισμα χαμηλού επιπέδου		διακριτή
DEM_evt_hc_lvl_L_sum	Δύσκολες στροφές χαμηλού επιπέδου		διακριτή
DEM_evt_ha_lvl_H_sum	Δύσκολα γεγονότα επιτάχυνσης υψηλού επιπέδου		διακριτή
DEM_evt_hb_lvl_H_sum	Δυνατό φρενάρισμα υψηλού επιπέδου		διακριτή
DEM_evt_hc_lvl_H_sum	Δύσκολες στροφές υψηλού επιπέδου		διακριτή
ME_AWS_time_indicator_median	Υποδεικνύει συνθήκες φωτισμού (ημέρα, σούρουπο, νύχτα)		διακριτή
ME_Car_wipers_median	Υποδεικνύει καιρικές συνθήκες (υαλοκαθαριστήρες on/off)		διακριτή
ME_Car_high_beam_median	Μεγάλη σκάλα		διακριτή
IBI_value_mean	Χρονικό διάστημα μεταξύ διαδοχικών καρδιακών παλμών	χιλιοστά του δευτερολέπτου	αριθμητική
ME_LDW_Map_type_L_median	Προειδοποίηση αναχώρησης από την αριστερή λωρίδα		διακριτή
ME_LDW_Map_type_R_median	Προειδοποίηση αναχώρησης από την δεξιά λωρίδα		διακριτή

#### 4.4 Περιγραφική στατιστική δεδομένων

Αξιοποιώντας τη βιβλιοθήκη ανάλυσης δεδομένων pandas στο προγραμματιστικό περιβάλλον rython, πραγματοποιήθηκε περιγραφική στατιστική των δεδομένων μετά την

επεξεργασία τους. Στους πίνακες 4.3 και 4.4 παρατίθενται ορισμένα περιγραφικά στατιστικά στοιχεία των μεταβλητών που συλλέχθηκαν όπως η μέση τιμή, η τυπική απόκλιση, η ελάχιστη και η μέγιστη τιμή για το Βέλγιο και την Αγγλία αντίστοιχα.

Πίνακας 4.4: Περιγραφική στατιστική αριθμητικών δεδομένων από τον προσομοιωτή οδήγησης για το Βέλγιο.

Μεταβλητή	Μέση τιμή	Τυπική απόκλιση	Ελάχιστη τιμή	Μέγιστη τιμή
Drowsiness_level_median	35,00	0,12	35,00	39,00
IBI_value_mean	816,70	97,32	375,98	2191,41
ME_Car_speed_mean	53,92	34,72	0,00	172,43
ME_Car_wipers_median	0,05	0,22	0,00	1,00
ME_Car_high_beam_median	0,02	0,15	0,00	1,00
ME_AWS_hw_measurement_mean	76225,80	35943,32	0,27	99999,00
ME_AWS_tsr_level_mean	0,51	1,25	0,00	7,00
ME_AWS_fcw_mean	0,00	0,00	0,00	0,13
ME_AWS_pcw_mean	0,00	0,00	0,00	0,13
ME_AWS_pedestrian_dz_mean	0,00	0,03	0,00	1,00
ME_AWS_time_indicator_median	1,31	0,71	1,00	3,00
ME_TSR_tsr_1_speed_median	125,89	119,75	0,00	254,00
GPS_spd_mean	50,92	33,81	0,00	158,74
GPS_distances_sum	431,75	393,11	0,00	42712,52
DEM_evt_ha_lvl_L_mean	0,34	0,43	0,00	1,00
DEM_evt_ha_lvl_M_mean	0,11	0,29	0,00	1,00
DEM_evt_ha_lvl_H_mean	0,06	0,22	0,00	1,00
DEM_evt_hb_lvl_L_mean	0,21	0,36	0,00	1,00
DEM_evt_hb_lvl_M_mean	0,02	0,13	0,00	1,00
DEM_evt_hb_lvl_H_mean	0,00	0,05	0,00	1,00
DEM_evt_hc_lvl_L_mean	0,14	0,31	0,00	1,00
DEM_evt_hc_lvl_M_mean	0,16	0,34	0,00	1,00
DEM_evt_hc_lvl_H_mean	0,02	0,14	0,00	1,00

Πίνακας 4.5: Περιγραφική στατιστική αριθμητικών δεδομένων από τον προσομοιωτή οδήγησης για την Αγγλία.

Μεταβλητή	Μέση τιμή	Τυπική απόκλιση	Ελάχιστη τιμή	Μέγιστη τιμή
Drowsiness_level_median	35,00	0,00	35,00	35,00
IBI_value_mean	799,42	98,04	306,64	1580,08
ME_Car_speed_mean	44,28	34,08	0,00	173,07
ME_Car_wipers_median	0,07	0,26	0,00	1,00
ME_Car_high_beam_median	0,01	0,08	0,00	1,00
ME_AWS_hw_measurement_mean	81104,20	31793,46	0,38	99999,00
ME_AWS_tsr_level_mean	0,10	0,37	0,00	7,00
ME_AWS_fcw_mean	0,00	0,00	0,00	0,13
ME_AWS_pcw_mean	0,00	0,00	0,00	0,07
ME_AWS_pedestrian_dz_mean	0,00	0,03	0,00	1,00

ME_AWS_time_indicator_median	1,80	0,96	1,00	3,00
ME_TSR_tsr_1_speed_median	116,28	115,97	0,00	254,00
GPS_spd_mean	40,74	32,99	0,00	165,49
GPS_distances_sum	341,00	274,92	0,00	3017,56
DEM_evt_ha_lvl_L_mean	0,37	0,44	0,00	1,00
DEM_evt_ha_lvl_M_mean	0,10	0,26	0,00	1,00
DEM_evt_ha_lvl_H_mean	0,03	0,16	0,00	1,00
DEM_evt_hb_lvl_L_mean	0,19	0,33	0,00	1,00
DEM_evt_hb_lvl_M_mean	0,01	0,10	0,00	1,00
DEM_evt_hb_lvl_H_mean	0,00	0,04	0,00	1,00
DEM_evt_hc_lvl_L_mean	0,16	0,34	0,00	1,00
DEM_evt_hc_lvl_M_mean	0,17	0,35	0,00	1,00
DEM_evt_hc_lvl_H_mean	0,02	0,13	0,00	1,00

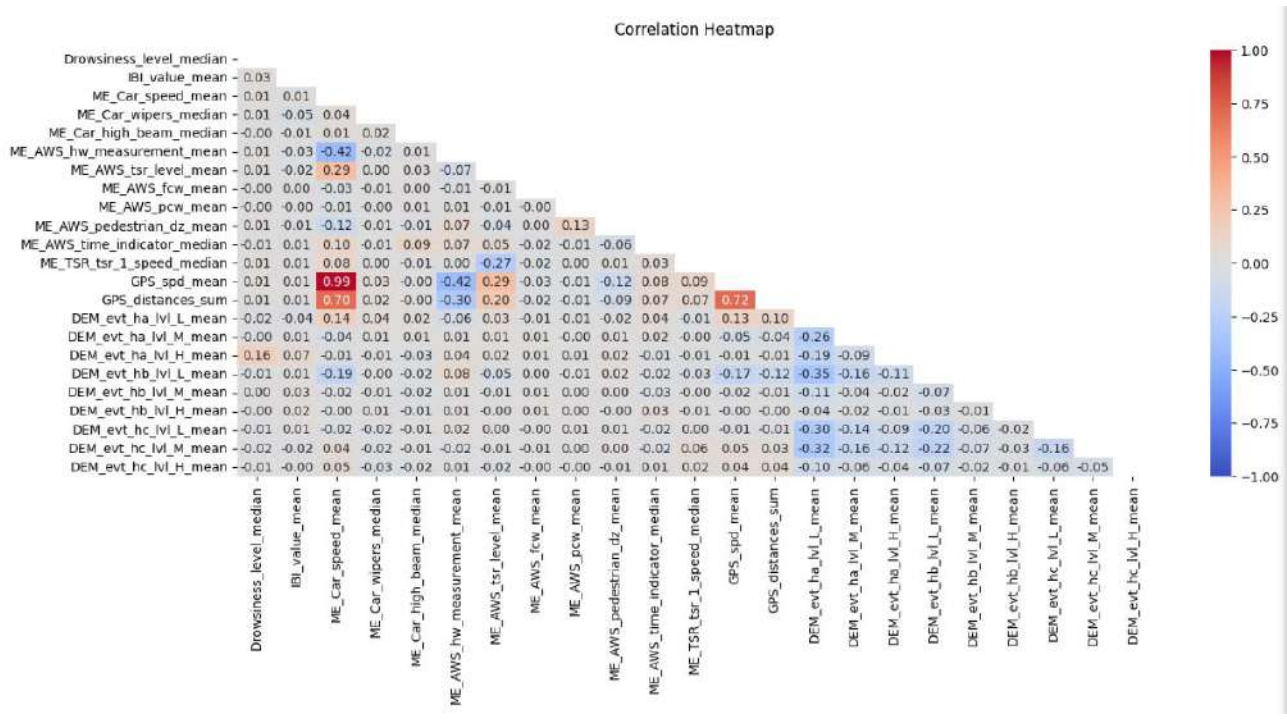
#### 4.5 Συσχέτιση μεταβλητών

Για την ανάπτυξη των μοντέλων ταξινόμησης και παλινδρόμησης είναι απαραίτητο να διερευνηθεί η συσχέτιση μεταξύ των ανεξάρτητων μεταβλητών. Οι τιμές του συντελεστή συσχέτισης Pearson κυμαίνονται στο διάστημα  $[-1,1]$  και η σχέση των ανεξάρτητων μεταβλητών χαρακτηρίζεται ως εξής:

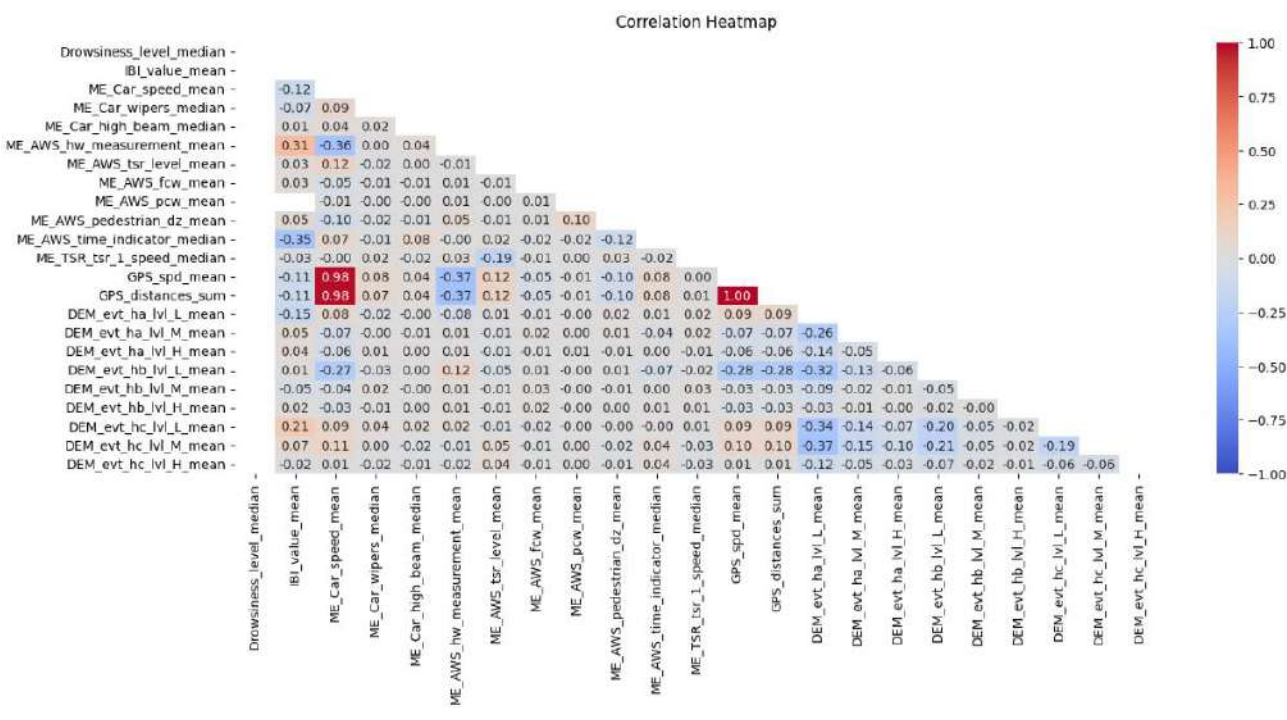
- Ελάχιστη συσχέτιση για  $0.00 \leq |r| \leq 0.30$
- Μέτρια συσχέτιση για  $0.31 \leq |r| \leq 0.70$
- Υψηλή συσχέτιση για  $0.71 \leq |r| \leq 1.00$

Για τον λόγο αυτό, χρησιμοποιώντας την ίδια βιβλιοθήκη ανάλυσης στο προγραμματιστικό περιβάλλον R, αναπτύχθηκαν κατάλληλες τεχνικές υπολογισμού και απεικόνισης της συσχέτισης των μεταβλητών.

Στους παρακάτω τριγωνικούς χάρτες θερμότητας παρουσιάζεται η συσχέτιση μεταξύ των διαφορετικών μεταβλητών. Η θετική συσχέτιση συμβολίζεται με θερμό χρώμα, ενώ η αρνητική με ψυχρό.



Γράφημα 4.2: Τριγωνικός χάρτης συσχέτισης μεταβλητών για το Βέλγιο



Γράφημα 4.3: Τριγωνικός χάρτης συσχέτισης μεταβλητών για την Αγγλία

Από τα γραφήματα 4.1 και 4.2 προκύπτουν τα εξής συμπεράσματα

- Ανάμεσα στα διαφορετικά περιγραφικά στοιχεία της ίδιας μεταβλητής παρουσιάζεται υψηλή συσχέτιση. Η υψηλή αυτή συσχέτιση είναι λογική καθώς αφορά τη σχέση μεταξύ διαφορετικών εκφάνσεων του ίδιου στοιχείου.
- Η μεταβλητή της ταχύτητας ME\_Car\_speed\_mean και των ορίων ταχύτητας GPS\_spd\_mean παρουσιάζει σημαντική συσχέτιση. Η αύξηση του ορίου ταχύτητας προκαλεί αύξηση της ταχύτητας που αναπτύσσει ο οδηγός.

- Επίσης η μεταβλητή `GPS_distances_sum` παρουσιάζει σημαντική συσχέτιση με τις `ME_Car_speed_mean` και `GPS_spd_mean`.

## 4.6 Σύνοψη

Συνοψίζοντας στο πλαίσιο του ερευνητικού προγράμματος i-DREAMS, διεξήχθη ένα νατουραλιστικό πείραμα οδήγησης στο οποίο συμμετείχαν 250 οδηγοί από το Βέλγιο και το Ηνωμένο Βασίλειο, τη Γερμανία, την Ελλάδα και την Πορτογαλία και αντλήθηκαν τα σχετιζόμενα με την οδηγική συμπεριφορά χαρακτηριστικά. Στην συνέχεια, ακολούθησε η κατάλληλη επεξεργασία των δεδομένων και ο υπολογισμός των περιγραφικών στατιστικών στοιχείων για την καλύτερη κατανόηση τους. Τέλος, εξετάστηκε η συσχέτιση μεταξύ των μεταβλητών το οποίο αποτελεί απαραίτητο βήμα για τις προκαταρκτικές διαδικασίες των αναλύσεων που θα ακολουθήσουν.

## 5. ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΟΛΟΓΙΑΣ - ΑΠΟΤΕΛΕΣΜΑΤΑ

### 5.1 Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζεται αναλυτικά η μεθοδολογία που εφαρμόστηκε καθώς και τα αποτελέσματα που προέκυψαν στο πλαίσιο της μελέτης. Ο στόχος και η κατάλληλη μεθοδολογία για την επίτευξη του προσδιορίστηκαν στη βιβλιογραφική ανασκόπηση.

Σύμφωνα με τις μεθοδολογίες παλαιότερων ερευνών, που παρουσιάστηκαν στην βιβλιογραφική ανασκόπηση, για την διερεύνηση της επιρροής των διαφορετικών παραγόντων της οδηγικής συμπεριφοράς θα αναπτυχθούν **κατάλληλοι αλγόριθμοι μηχανικής και βαθιάς εκμάθησης για την ταξινόμηση**. Αναλυτικότερα θα αξιολογηθεί η σημαντικότητα των μεταβλητών στην ταξινόμηση και η ερμηνευτική τους ικανότητα στην παλινδρόμηση. Η επικίνδυνη οδήγηση θα εξεταστεί βάσει δύο προσεγγίσεων επομένως, η ανάλυση θα χωριστεί σε δύο μέρη.

Στο πρώτο μέρος των αναλύσεων θα αναπτυχθούν τα μοντέλα ταξινόμησης για τον καθορισμό του επιπέδου της 'Ζώνης Ανοχής Ασφαλείας' που βρίσκεται οδηγός για κάθε χρονικό πλαίσιο των 30 δευτερολέπτων. Τα δεδομένα που συλλέχθηκαν από την συμμετοχή 250 οδηγών στο Βέλγιο και την Αγγλία, αποτελούν τις ενδογενείς μεταβλητές, ενώ το επίπεδο της 'Ζώνης Ανοχής Ασφαλείας' αποτελεί την εξωγενή μεταβλητή.

Στο δεύτερο μέρος των αναλύσεων θα αναπτυχθεί ο **αλγόριθμος Lime** με σκοπό να εξεταστούν και αξιολογηθούν τα χαρακτηριστικά που επιδρούν στην διάρκεια οδήγηση σε κάθε ένα από τα τρία επίπεδα ασφαλείας για κάθε οδηγό.

Με την χρήση μετρικών αξιολόγησης θα πραγματοποιηθεί η αξιολόγηση της προγνωστικής ικανότητας των μοντέλων. Με την χρήση της προγραμματιστικής γλώσσας **Python** θα πραγματοποιηθεί η ανάλυση. Θα αξιοποιηθούν οι παρακάτω ειδικές βιβλιοθήκες και εργαλεία:

- Υπολογισμοί: NumPy
- Ανάλυση και χειρισμός δεδομένων: Pandas
- Χειρισμός ανομοιογένειας δεδομένων: Imbalanced Learn
- Γραφική απεικόνιση: Matplotlib, Seaborn

- Μηχανική εκμάθηση: Scikit-Learn
- Βαθιά εκμάθηση: Tensorflow
- Ερμηνεία Μοντέλων: Local Interpretable Model-Agnostic Explanations (LIME)

## 5.2 Εντοπισμός του Επιπέδου ‘Ζώνης Ανοχής Ασφαλείας’

Για τον εντοπισμό της ‘Ζώνης Ανοχής Ασφαλείας’ εστιάζουμε στον υπολογισμό της επιρροής του κάθε παράγοντα κινδύνου στην αναγνώριση της επικίνδυνης οδηγικής συμπεριφοράς του οδηγού. Η μεθοδολογία που θα ακολουθηθεί περιλαμβάνει την ανάλυση των διαφορετικών παραγόντων κινδύνου με βάση την ανάπτυξη τριών συνδυασμένων αλγόριθμων ταξινόμησης. Τέλος, με βάση την συνολική επίδοση των μοντέλων ταξινόμησης θα αξιολογήσουμε τους κρίσιμους παράγοντες.

### 5.2.1 Καθορισμός Επιπέδων Ασφαλείας

Αρχικά, πριν την ανάπτυξη των αλγόριθμων ταξινόμησης και την διερεύνηση της επιρροής των μεταβλητών στην επικίνδυνη οδήγηση, ήταν αναγκαία η **κατηγοριοποίηση** των δεδομένων οδήγησης σε ένα από τα τρία επίπεδα της ‘Ζώνης Ανοχής Ασφαλείας’. Σύμφωνα με την εφαρμογή i-DREAMS, καταγράφηκαν ορισμένα επίπεδα παρέμβασης σε πραγματικό χρόνο, τα οποία θεσπίστηκαν στα πλαίσια του έργου με βάση ορισμένες αρχές και όρια από την διεθνή βιβλιογραφία.

Η αντιστοίχιση των επιπέδων βασίστηκε στην ταχύτητα (Speed) και την χρονική απόσταση από το προπορευόμενο όχημα (Headway) για τις δύο χώρες. Προκειμένου τα αποτελέσματα να εναρμονίζονται με τα αποτελέσματα της διεθνούς βιβλιογραφίας είναι θεμιτό τα δείγματα της επικίνδυνης οδήγησης να αποτελούν την κλάση μειοψηφίας. Συγκεκριμένα, το επίπεδο ‘Κανονικό’ πρέπει να είναι η κύρια κατηγορία με το υψηλότερο ποσοστό δειγμάτων, ενώ τα επίπεδα ‘Επικίνδυνο’ και ‘Αποφυγή Ατυχήματος’ πρέπει να είναι η μειοψηφική κατηγορία με το χαμηλότερο ποσοστό δειγμάτων.

Αναλυτικότερα, από το πείραμα συλλέξαμε τις μεταβλητές `iDreams_Headway_Map_level_i` και `iDreams_Speeding_Map_level_i`, όπου  $i = -1, 0, 1, 2, 3$ . Οι παραπάνω μεταβλητές αντιστοιχούν στα επίπεδα παρέμβασης της χρονικής απόστασης από το προπορευόμενο όχημα και παρέμβασης της ταχύτητας σε πραγματικό χρόνο αντίστοιχα. Κάθε επίπεδο παρέμβασης αντιστοιχεί σε ένα από τα τρία επίπεδα της ‘Ζώνης Ανοχής Ασφαλείας’. Οι πιθανές τιμές αυτών των μεταβλητών είναι 0 (όταν το επίπεδο παρέμβασης διάφορο του  $i$ ) και 1 (όταν το επίπεδο παρέμβασης ίσο με  $i$ ).

Πίνακας 5.1: Επεξήγηση μεταβλητών iDreams

Μεταβλητές iDreams	Επεξήγηση Μεταβλητών
<code>iDreams_Headway_Map_level_-1</code>	Για επίπεδο παρέμβασης σε πραγματικό επίπεδο -1 δεν εντοπίζεται κανένα όχημα και το επίπεδο Ζώνης Ανοχής Ασφαλείας είναι Normal.
<code>iDreams_Headway_Map_level_0</code>	Για επίπεδο παρέμβασης σε πραγματικό χρόνο 0 εντοπίζεται όχημα αλλά με $Headway \geq 2,5$ και το επίπεδο Ζώνης Ανοχής Ασφαλείας είναι Normal.

iDreams_Headway_Map_level_1	Για επίπεδο παρέμβασης σε πραγματικό χρόνο 1 εντοπίζεται όχημα με Headway <2,5 που υπερβαίνει τις οριακές τιμές και το επίπεδο Ζώνης Ανοχής Ασφαλείας είναι Normal.
iDreams_Headway_Map_level_2	Για επίπεδο παρέμβασης σε πραγματικό χρόνο 2 έχουμε το πρώτο προειδοποιητικό στάδιο και επίπεδο Ζώνης Ανοχής Ασφαλείας Dangerous.
iDreams_Headway_Map_level_3	Για επίπεδο παρέμβασης σε πραγματικό χρόνο 3 έχουμε το δεύτερο προειδοποιητικό στάδιο και το επίπεδο Ζώνης Ανοχής Ασφαλείας Avoidable Accident.
iDreams_Speeding_Map_level_0	Για επίπεδο παρέμβασης σε πραγματικό χρόνο 0 δεν υπάρχει προειδοποίηση και το επίπεδο Ζώνης Ανοχής Ασφαλείας είναι Normal.
iDreams_Speeding_Map_level_1	Για επίπεδο παρέμβασης σε πραγματικό χρόνο 1 έχουμε οπτική ένδειξη και το επίπεδο Ζώνης Ανοχής Ασφαλείας είναι Normal.
iDreams_Speeding_Map_level_2	Για επίπεδο παρέμβασης σε πραγματικό χρόνο 2 έχουμε οπτική προειδοποίηση υπερβολικής ταχύτητας και το επίπεδο Ζώνης Ανοχής Ασφαλείας είναι Dangerous.
iDreams_Speeding_Map_level_3	Για επίπεδο παρέμβασης σε πραγματικό χρόνο 3 έχουμε οπτική, ακουστική προειδοποίηση υπερβολικής ταχύτητας και το επίπεδο Ζώνης Ανοχής Ασφαλείας είναι Avoidable Accident.

Για τον καθορισμό των επιπέδων της Ζώνης Ανοχής Ασφαλείας που καταγράφονται, ελέγχθηκαν αρχικά για κάθε 30 δευτερόλεπτα οι παραπάνω μεταβλητές σε κάθε επίπεδο παρέμβασης. Στην συνέχεια, επιλέχθηκε το δυσμενέστερο επίπεδο από άποψη ασφαλείας, δηλαδή επιλέχθηκε η μεταβλητή στην οποία εντοπίζετε μία τιμή και αντιστοιχεί στο δυσμενέστερο επίπεδο ασφαλείας.

Επομένως, για κάθε 30 δευτερόλεπτα προκύπτει το επίπεδο της Ζώνης Ανοχής Ασφαλείας το οποίο χαρακτηρίζεται ως:

- **Normal** όταν η μεταβλητή που επιλέγεται είναι επιπέδου παρέμβασης -1,0,1
- **Dangerous** όταν η μεταβλητή που επιλέγεται είναι επιπέδου παρέμβασης 2
- **Avoidable Accident** όταν η μεταβλητή που επιλέγεται είναι επιπέδου παρέμβασης 3

Στον πίνακα 5.2 παρατίθενται τα αποτελέσματα της παραπάνω ανάλυσης και για τα δεδομένα που συλλέχθηκαν στην Αγγλία και το Βέλγιο.

‘Επίπεδο ‘Ζώνης Ανοχής Ασφαλείας’			
	Normal	Dangerous	Avoidable Accident
Βέλγιο (Headway)	68472	17165	4787
Αγγλία (Headway)	53942	13960	3365
Βέλγιο (Speeding)	67411	4354	17555
Αγγλία ( Speeding)	54092	2291	8118

Συμπεραίνοντας ο καθορισμός του επιπέδου ασφαλείας θα γίνει με την χρήση της μεταβλητής Headway, καθώς καταγράφονται περισσότερα επίπεδα Normal και λιγότερα επίπεδα Dangerous και Avoidable Accident σε σύγκριση με την μεταβλητή Speeding. Οι μεταβλητές που αφορούν τον παράγοντα Headway δεν θα αποτελέσουν μεταβλητές εισόδου στα μοντέλα. Ο συνυπολογισμός τους στην διαδικασία της ταξινόμησης θα οδηγούσε σε μεροληψία του μοντέλου χωρίς εκείνο να προσφέρει χρήσιμα και σημαντικά αποτελέσματα.

### 5.2.2 Επιλογή Χαρακτηριστικών (Feature Selection)

Βασικό κομμάτι της μεθοδολογίας αποτελεί η διαδικασία **επιλογής χαρακτηριστικών**. Στόχος της διαδικασίας είναι η μείωση του αριθμού των μεταβλητών εισόδου με ταυτόχρονη μείωση του υπολογιστικού κόστους του μοντέλου και βελτίωση της προγνωστικής του απόδοσης.

Η επιλογή των χαρακτηριστικών γίνεται με γνώμονα την συσχέτιση των μεταβλητών και την επιρροή της μεταβλητής στην διαδικασία της ταξινόμησης. Η μέθοδος αυτή αποτελεί μία αρχική προσέγγιση για την μείωση των μεταβλητών εισόδου και τη βελτίωση των μοντέλων.

Διερευνήθηκαν διάφορα σύνολα συνδυάζοντας διαφορετικές μεταβλητές βάσει της συσχέτισης τους και της επιρροής τους στις προβλέψεις.

Για τον εντοπισμό της σημαντικότητας των μεταβλητών στην ταξινόμηση χρησιμοποιήθηκε η τεχνική σημαντικότητας χαρακτηριστικών βάσει την μετάθεση χαρακτηριστικών (feature importance based on feature permutation).

Αρχικά τα δεδομένα χωρίστηκαν σε δύο υποσύνολα:

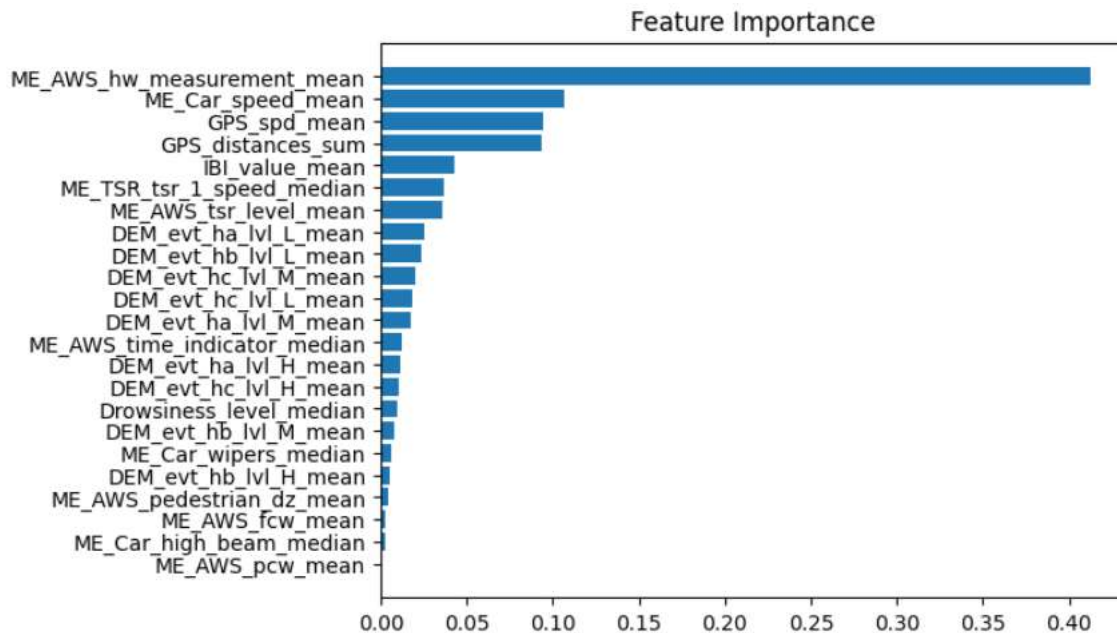
- A. Το πρώτο υποσύνολο περιλαμβάνει όλες τις μεταβλητές εισόδου της μεθόδου δηλαδή τα δεδομένα που συλλέχθηκαν από το πείραμα που διεξήχθη σε πραγματικές συνθήκες οδήγησης.
- B. Το δεύτερο υποσύνολο περιλαμβάνει την μεταβλητή εξόδου που αποτελείται από το επίπεδο της ‘Ζώνης Ανοχής Ασφαλείας’.

Έπειτα χρησιμοποιώντας τα ειδικά εργαλεία της βιβλιοθήκης scikit-learn, αναπτύχθηκε ο ταξινομητής των ‘Τυχαίων Δασών’ (Random Forests Classifier) και υπολογίστηκε η

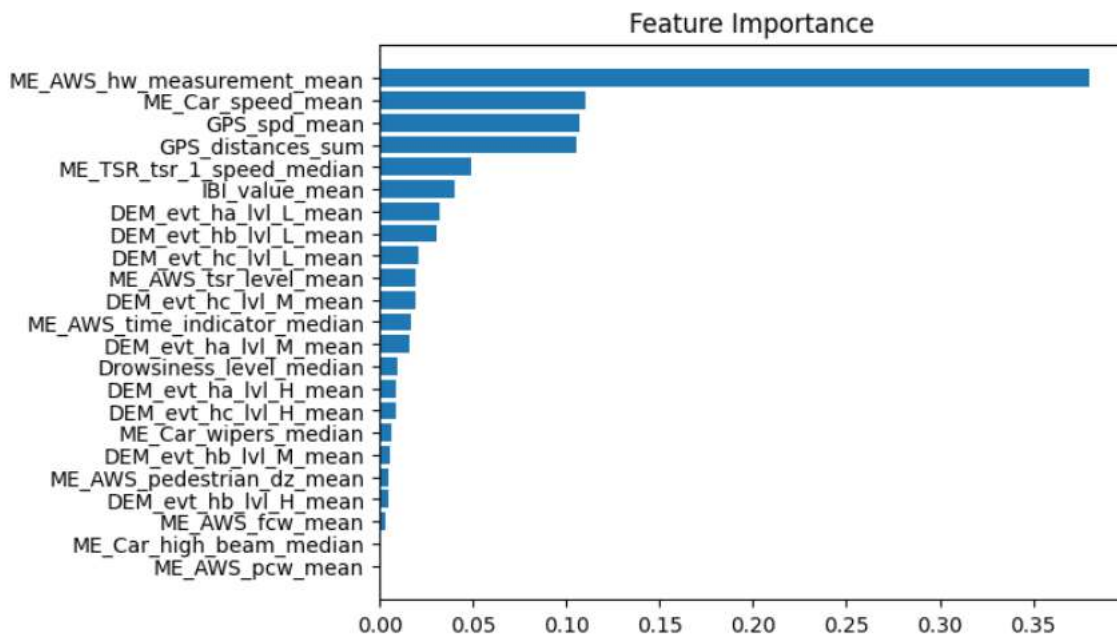


επιρροή της κάθε μεταβλητής (feature importance) με βάση την μετάθεση των χαρακτηριστικών (feature permutation) στη διαδικασία ταξινόμησης.

Στο γράφημα 5.1 και 5.2 καθώς και στον πίνακα 5.3 και 5.4 απεικονίζεται η επιρροή της κάθε μεταβλητής στην ταξινόμηση σε κλίμακα τιμών [0,1], για το Βέλγιο και την Αγγλία αντίστοιχα.



Γράφημα 5.1: Σημαντικότητα μεταβλητών σύμφωνα με την μέθοδο ‘Τυχαίων Δασών’ για το Βέλγιο



Γράφημα 5.2: Σημαντικότητα μεταβλητών σύμφωνα με την μέθοδο ‘Τυχαίων Δασών’ για την Αγγλία

Όπως προκύπτει από τα γραφήματα 5.3 και 5.4, η χρονική απόσταση από το προπορευόμενο όχημα, η μέση ταχύτητα του οχήματος, η ταχύτητα, το χρονικό διάστημα μεταξύ διαδοχικών καρδιακών παλμών και η συνολική απόσταση που διένυσε το όχημα

έχουν την μεγαλύτερη επιρροή στην διαδικασία του καθορισμού του επιπέδου της 'Ζώνης Ανοχής Ασφαλείας' που βρίσκεται ο οδηγός.

Πίνακας 5.3: Αριθμητικές τιμές της σημαντικότητας των μεταβλητών για το Βέλγιο

Drowsiness_level_median	0,0093
IBI_value_mean	0,0424
ME_Car_speed_mean	0,107
ME_Car_wipers_median	0,0063
ME_Car_high_beam_median	0,0027
ME_AWS_hw_measurement_mean	0,4129
ME_AWS_tsr_level_mean	0,0356
ME_AWS_fcw_mean	0,0031
ME_AWS_pcw_mean	0,0001
ME_AWS_pedestrian_dz_mean	0,004
ME_AWS_time_indicator_median	0,0122
ME_TSR_tsr_1_speed_median	0,0363
GPS_spd_mean	0,0941
GPS_distances_sum	0,0936
DEM_evt_ha_lvl_L_mean	0,0256
DEM_evt_ha_lvl_M_mean	0,0178
DEM_evt_ha_lvl_H_mean	0,0112
DEM_evt_hb_lvl_L_mean	0,0232
DEM_evt_hb_lvl_M_mean	0,0076
DEM_evt_hb_lvl_H_mean	0,0057
DEM_evt_hc_lvl_L_mean	0,0187
DEM_evt_hc_lvl_M_mean	0,0202
DEM_evt_hc_lvl_H_mean	0,0105

Πίνακας 5.4: Αριθμητικές τιμές της σημαντικότητας των μεταβλητών για την Αγγλία

Drowsiness_level_median	0,009
IBI_value_mean	0,040
ME_Car_speed_mean	0,110
ME_Car_wipers_median	0,007
ME_Car_high_beam_median	0,001
ME_AWS_hw_measurement_mean	0,380
ME_AWS_tsr_level_mean	0,019
ME_AWS_fcw_mean	0,003
ME_AWS_pcw_mean	0,000
ME_AWS_pedestrian_dz_mean	0,005
ME_AWS_time_indicator_median	0,017
ME_TSR_tsr_1_speed_median	0,049
GPS_spd_mean	0,107

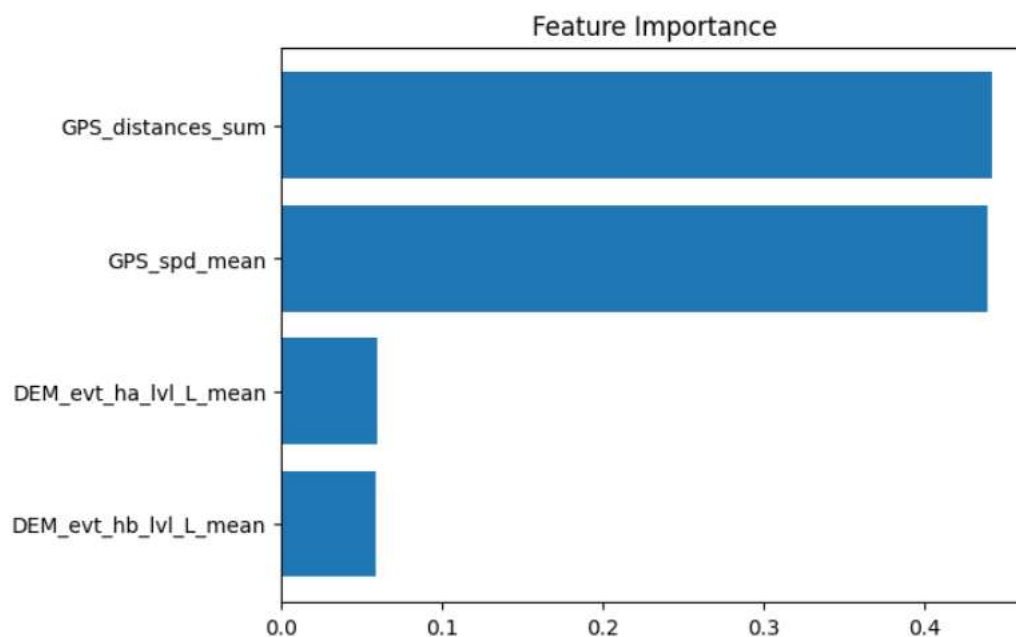
GPS_distances_sum	0,106
DEM_evt_ha_lvl_L_mean	0,032
DEM_evt_ha_lvl_M_mean	0,016
DEM_evt_ha_lvl_H_mean	0,009
DEM_evt_hb_lvl_L_mean	0,031
DEM_evt_hb_lvl_M_mean	0,006
DEM_evt_hb_lvl_H_mean	0,004
DEM_evt_hc_lvl_L_mean	0,021
DEM_evt_hc_lvl_M_mean	0,019
DEM_evt_hc_lvl_H_mean	0,008

Με βάση την συσχέτιση και την σημαντικότητα που προέκυψε γίνεται η επιλογή των μεταβλητών εισόδου στα μοντέλα ταξινόμησης. Επιλέγοντας να γίνει ο καθορισμός του επιπέδου ασφαλείας με την χρήση της μεταβλητής Headway, τα διάφορα περιγραφικά στατιστικά στοιχεία του παράγοντα Headway δεν θα αποτελέσουν μεταβλητές εισόδου στα μοντέλα. Ο συνυπολογισμός τους στην διαδικασία της ταξινόμησης θα οδηγούσε σε μεροληψία του μοντέλου χωρίς εκείνο να προσφέρει χρήσιμα και σημαντικά αποτελέσματα. Επομένως, η μεταβλητή ME\_AWS\_hw\_measurement\_mean δεν θα αποτελέσει μία από τις μεταβλητές εισόδου. Με βάση την σημαντικότητα οι επόμενες μεταβλητές εισόδου που επιλέγονται είναι οι GPS\_distances\_sum και GPS\_spd\_mean. Οι μεταβλητές ME\_Car\_speed\_mean και ME\_TSR\_tsr\_1\_speed\_median δεν θα συμπεριληφθούν στις μεταβλητές εισόδου καθώς είναι ιδιαίτερα συσχετισμένες με την GPS\_spd\_mean.

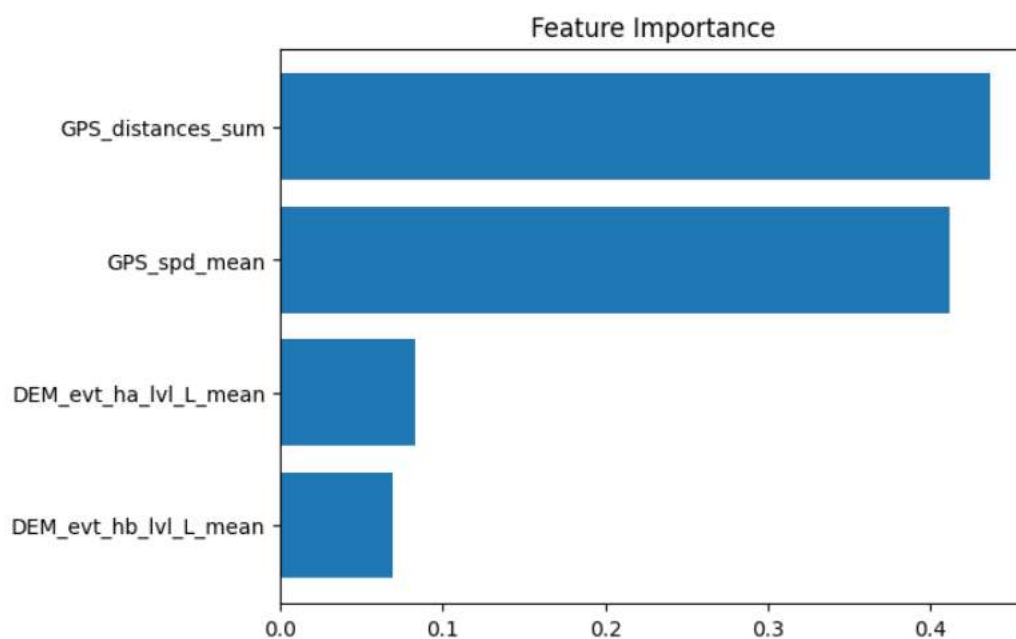
Το σύνολο μεταβλητών εισόδου που επιλέχθηκε είναι:

1. GPS\_distances\_sum
2. GPS\_spd\_mean
3. DEM\_evt\_ha\_lvl\_L\_mean
4. DEM\_evt\_hb\_lvl\_L\_mean

Στα γραφήματα 5.7 και 5.8, προκύπτει η σημαντικότητα των τελικών μεταβλητών εισόδου στην διαδικασία της ταξινόμησης.



Γράφημα 5.3: Σημαντικότητα τελικών μεταβλητών εισόδου ταξινόμησης για το Βέλγιο



Γράφημα 5.4: Σημαντικότητα τελικών μεταβλητών εισόδου ταξινόμησης για την Αγγλία

Όπως προκύπτει η συνολική απόσταση του ταξιδιού αποτελεί την σημαντικότερη μεταβλητή για την πρόβλεψη των διαφορετικών επιπέδων ασφαλείας, ενώ ακολουθεί η ταχύτητα του οχήματος, δύσκολα γεγονότα επιτάχυνσης χαμηλού επιπέδου και Δυνατό φρενάρισμα χαμηλού επιπέδου.

### 5.2.3 Προετοιμασία δεδομένων

Όπως αναλύθηκε στην επιλογή χαρακτηριστικών, τα δεδομένα που αντλήθηκαν από το πείραμα διαχωρίστηκαν σε δύο υποσύνολα: (A) στα δεδομένα εισόδου που αποτελούνται από τις 4 μεταβλητές που αναλύθηκαν προηγουμένως και (B) στις μεταβλητές εξόδου που αποτελούνται από τα τρία επίπεδα της 'Ζώνης Ανοχής Ασφαλείας'. Τα δύο υποσύνολα διαχωρίστηκαν στα **δεδομένα εκπαίδευσης (training dataset)** και τα **δεδομένα εξέτασης**

**(testing dataset)** με το 90% των δεδομένων να αποτελούν το σύνολο δεδομένων εκπαίδευσης και το υπόλοιπο 10% του συνόλου δεδομένων εξέτασης, αντίστοιχα. Σύμφωνα με την περιγραφή της λειτουργίας των μοντέλων μηχανικής και βαθιάς εκμάθησης στο κεφάλαιο 3, τα δεδομένα εκπαίδευσης χρησιμοποιούνται για την εκπαίδευση του μοντέλου στην αναγνώριση του επιπέδου ασφαλείας δημιουργώντας μοτίβα αναγνώρισης βάσει συγκεκριμένων χαρακτηριστικών. Το μοντέλο στην διαδικασία αξιολόγησης επεξεργάζεται τα εισαγόμενα δεδομένα εξέτασης και τα κατηγοριοποιεί σε ένα από τα τρία επίπεδα ώστε να γίνει η σύγκριση τους με τα πραγματικά επίπεδα ασφαλείας.

#### 5.2.4 Αντιμετώπιση άνιση κατανομής δεδομένων στις κλάσεις

Σύμφωνα με την βιβλιογραφική ανασκόπηση, η πλειονότητα των μελετών αντιμετωπίζει **το πρόβλημα της ανισορροπίας δεδομένων** ως προς τις διαφορετικές κλάσεις, με τα δείγματα των επικίνδυνων συνθηκών να είναι σημαντικά μικρότερα σε σχέση με τα δείγματα των συνθηκών ασφαλούς οδήγησης. Επιπρόσθετα, όπως προαναφέρθηκε στο κεφάλαιο 3, τα μοντέλα ταξινόμησης υποθέτουν την ομοιόμορφη κατανομή των δεδομένων στις κλάσεις καθιστώντας τα ιδιαίτερα ευάλωτα σε σφάλματα ταξινόμησης για δεδομένα με άνιση κατανομή.

Καθορίζοντας τα επίπεδα ασφαλείας και κατηγοριοποιώντας τα δεδομένα σε διαφορετικά επίπεδα προκύπτει η άνιση κατανομή που παρουσιάζεται στο παρακάτω πίνακα 5.4 και 5.5, για το Βέλγιο και την Αγγλία αντίστοιχα.

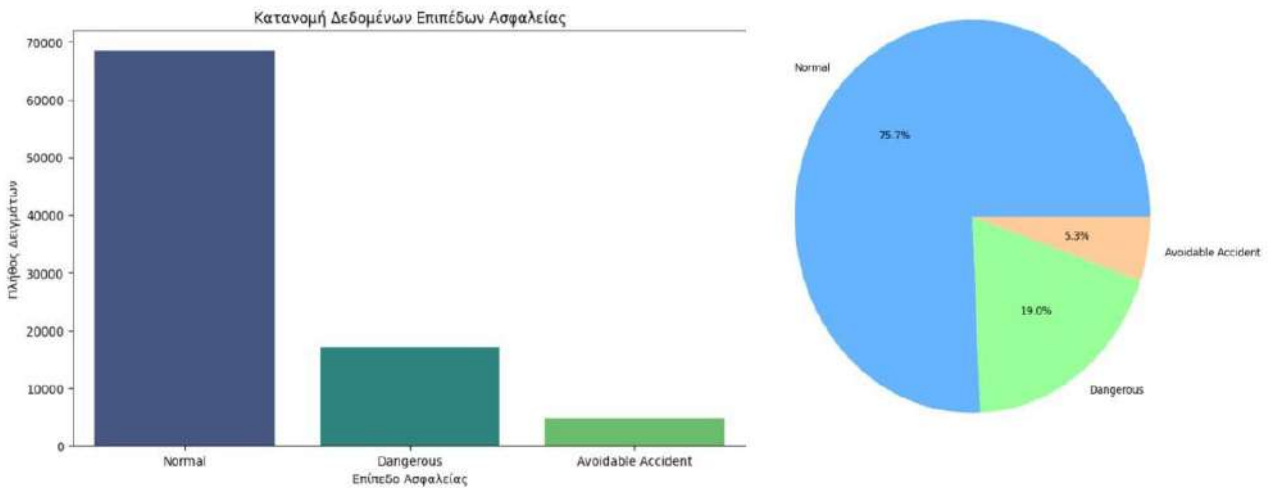
Πίνακας 5.5: Κατανομή δειγμάτων στα διαφορετικά επίπεδα ασφαλείας για το Βέλγιο

Επίπεδο 'Ζώνης Ανοχής Ασφαλείας'	Αριθμός δειγμάτων	Ποσοστό δειγμάτων
Επίπεδο 0 (Nomal)	68472	75,72%
Επίπεδο 1 (Dangerous)	17165	18,98%
Επίπεδο 2 (Avoidable Accident)	4787	5,29%

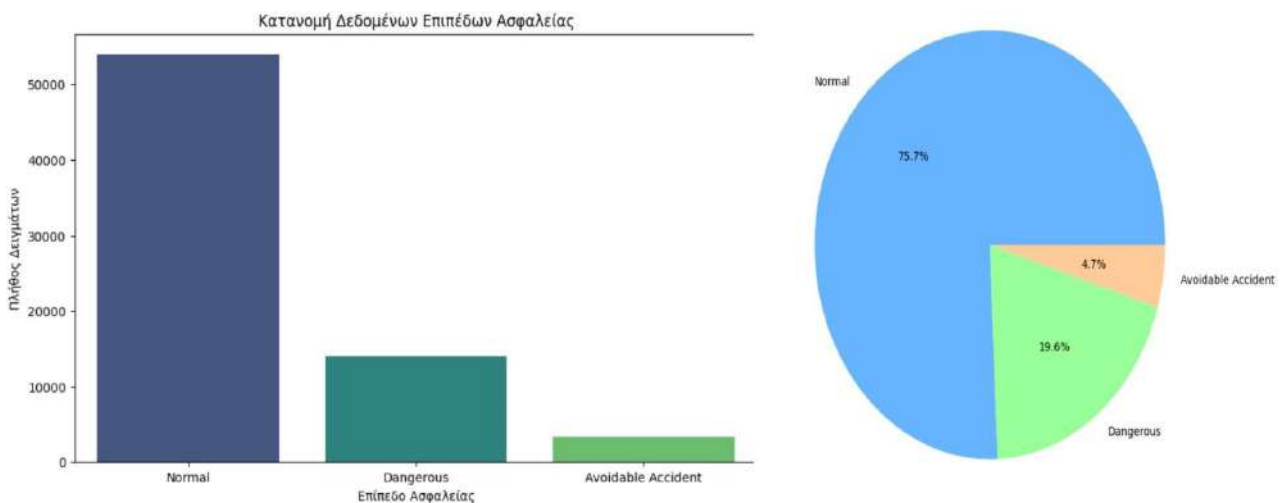
Πίνακας 5.6: Κατανομή δειγμάτων στα διαφορετικά επίπεδα ασφαλείας για την Αγγλία

Επίπεδο 'Ζώνης Ανοχής Ασφαλείας'	Αριθμός δειγμάτων	Ποσοστό δειγμάτων
Επίπεδο 0 (Nomal)	53942	75,69%
Επίπεδο 1 (Dangerous)	13960	19,59%
Επίπεδο 2 (Avoidable Accident)	3365	4,72%

Αντίστοιχη κατανομή θα έχουν και τα δεδομένα εκπαίδευσης που θα εκπαιδεύσουν τον αλγόριθμο ταξινόμησης. Η ανισορροπία στην κατανομή των δεδομένων εκπαίδευσης (training dataset) με βάση τα οποία θα εκπαιδευτούν τα μοντέλα ταξινόμησης παρουσιάζεται στα γραφήματα 5.5 και 5.6, για το Βέλγιο και την Αγγλία αντίστοιχα.



Γράφημα 5.5: Κατανομή δεδομένων εκπαίδευσης στα διαφορετικά επίπεδα ασφαλείας πριν την διαδικασία επαναδειγματοληψίας για το Βέλγιο



Γράφημα 5.6: Κατανομή δεδομένων εκπαίδευσης στα διαφορετικά επίπεδα ασφαλείας πριν την διαδικασία επαναδειγματοληψίας για την Αγγλία

Στην συνέχεια, θα αναλυθούν τα μοντέλα ταξινόμησης που αναπτύχθηκαν καθώς και οι επιδόσεις που σημείωσαν, με βάση τις μεταβλητές εισόδου της υποενότητας 5.2.2 και την τεχνική Synthetic Minority Oversampling Technique (SMOTE).

### 5.2.5 Ανάπτυξη μοντέλων ταξινόμησης

Όπως έχει αναφερθεί στις προηγούμενες ενότητες, ορισμένοι αλγόριθμοι ταξινόμησης αναπτύχθηκαν με στόχο την αναγνώριση του επιπέδου 'Ζώνης Ανοχής Ασφαλείας' που βρίσκεται ο οδηγός σε κάθε χρονικό πλαίσιο των 30 δευτερολέπτων. Η επιλογή των τριών

μοντέλων πραγματοποιήθηκε με βάση την βιβλιογραφική ανασκόπηση. Στον πίνακα 5.4 επεξηγούνται η ονοματολογία και ο συμβολισμός των μοντέλων.

Πίνακας 5.7: Ονοματολογία και συμβολισμός μοντέλων ταξινόμησης

Όνομα Μοντέλου (Ελληνικά)	Όνομα Μοντέλου (Αγγλικά)	Συμβολισμός Μοντέλου
Ταξινομητής Τυχαίων Δασών	Random Forests Classifier	RF
Ταξινομητής Adaboost	Adaboost Classifier	Adaboost
Συνελεκτικά Νευρωνικά Δίκτυα	Convolutional Neural Networks	CNN
Μοντέλο Μακράς-Βραχείας Μνήμης	Long-Short Term Memory Classifier	LSTM
Μοντέλο Επαναλαμβανόμενων Νευρωνικών Δικτύων	Recurrent Neural Network	RNN
Μοντέλο Βαθιών Νευρωνικών Δικτύων	Deep Neural Network	DNN

Λόγω της ιδιαιτερότητας ορισμένων αλγορίθμων, κρίθηκε αναγκαία η εφαρμογή ορισμένων πρόσθετων τεχνικών προ-επεξεργασίας των δεδομένων τους πριν την εκπαίδευση και την αξιολόγηση τους. Ο μετασχηματισμός των δεδομένων θεωρείται αναγκαίος καθώς λόγω της διαφορετικής κλίμακας των τιμών τους δεν συνεισφέρουν ίσα στην εκπαίδευση του μοντέλου με αποτέλεσμα να κινδυνεύει το μοντέλο από μεροληψία.

Στην περίπτωση των συνδυαστικών αλγορίθμων και των δένδρων αποφάσεων δεν είναι απαραίτητος ο μετασχηματισμός των μεταβλητών καθώς η απόσταση μεταξύ των τιμών των δεδομένων δεν θεωρείται σημαντική. Επομένως, επειδή ο αλγόριθμος RF και ο αλγόριθμος AdaBoost είναι μοντέλα που βασίζονται σε δένδρα αποφάσεων, τα δεδομένα εισόδου τους δεν μετασχηματίζονται.

Η εκπαίδευση όλων των μοντέλων έγινε με την τεχνική SMOTE για να μπορέσουν με καλύτερο τρόπο να προβλέψουν τις 3 κατηγορίες που θέσαμε. Επιπλέον, εφαρμόστηκε η τεχνική βελτιστοποίησης των παραμέτρων τους GridSearchCV μέσω της βιβλιοθήκης scikit-learn της python. Με συνδυασμό των παραπάνω βελτιώθηκε η επίδοση των αλγορίθμων.

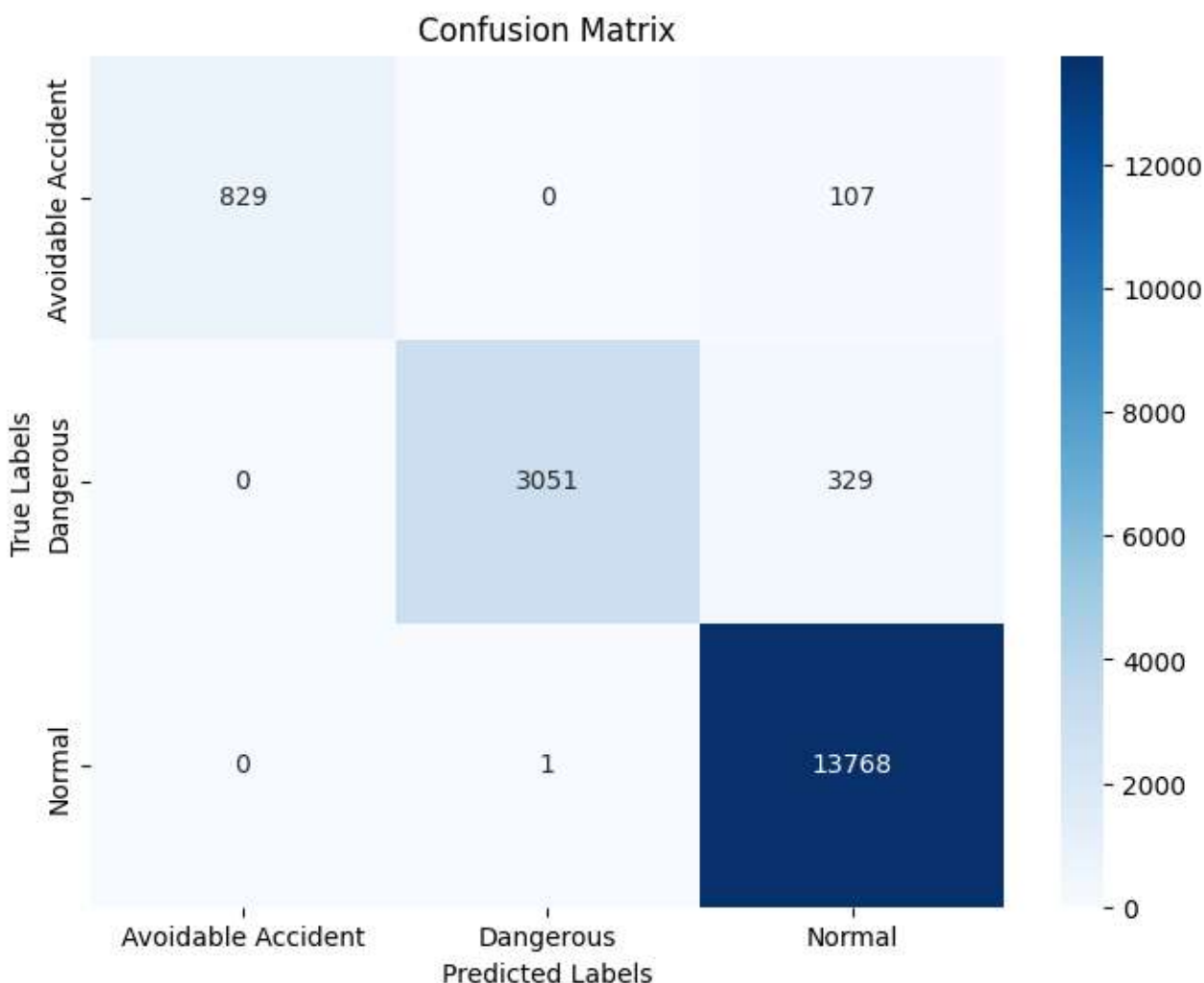
Για την εκπαίδευση των μοντέλων Βαθιάς μάθησης χρησιμοποιήθηκε η βιβλιοθήκη Tensorflow της python για την βελτιστοποίηση των μοντέλων.

Παρακάτω παρουσιάζονται για κάθε συνδυαστικό μοντέλο ταξινόμησης και για τις δύο χώρες οι μήτρες σύγκρισης για την γραφική αναπαράσταση της επίδοσης τους. Επιπλέον, παρουσιάζονται οι μετρικές αξιολόγησης που προέκυψαν μετά την εξέταση του κάθε μοντέλου.

## A) Δεδομένα Βελγίου

### 1. Συνδυασμός Βαθιού Νευρωνικού Δικτύου (DNN)-Τυχαία Δάση (RF)

Όπως φαίνεται στο γράφημα 5.11 ο αλγόριθμός σημείωσε υψηλά ποσοστά αναγνώρισης και των τριών επιπέδων της 'Ζώνης Ανοχής Ασφαλείας'. Το ποσοστό λανθασμένων προβλέψεων του επιπέδου 'Normal' είναι ιδιαίτερα χαμηλό με τιμή 3,07%. Ωστόσο σχετικά με το επίπεδο ασφαλείας 'Avoidable Accident' το αντίστοιχο ποσοστό έχει τιμή 0%. Επομένως, συνολικά θεωρείται ένα ικανοποιητικό μοντέλο που έχει υψηλή ικανότητα αναγνώρισης επικίνδυνων συμπεριφορών.



Γράφημα 5.7: Μήτρα σύγκρισης αλγόριθμου RF-DNN

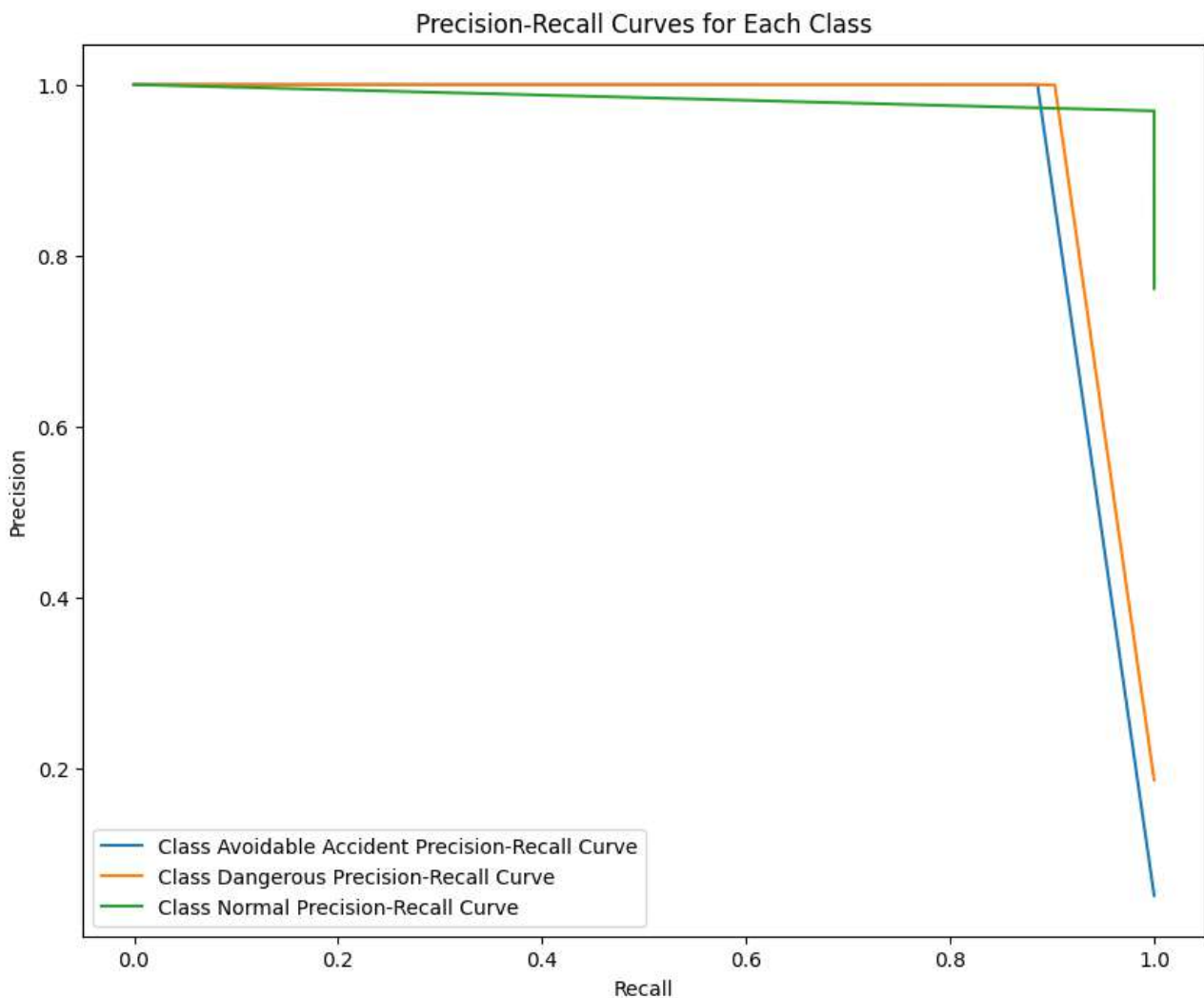
Για τον συνδυασμό RF-DNN συγκρίνοντας τις μετρικές αξιολόγησης του πίνακα 5.8 με τις αντίστοιχες των ερευνών που αναλύθηκαν, προκύπτει ότι η απόδοση του μοντέλου κινήθηκε σε όμοια ποσοστά.



Πίνακας 5.8: Σύνοψη μοντέλου RF-DNN

Επίδοση μοντέλου ταξινόμησης RF-DNN (ορθότητα: 98%)

Επίπεδο Ανοχής Ασφαλείας	‘Ζώνης Ακρίβεια	Ανάκληση	F1-score
Normal	98%	100%	98%
Dangerous	98%	90%	95%
Avoidable Accident	100%	89%	94%
Μέση Τιμή	98%	93%	96%



Γράφημα 5.8: Καμπύλη Ακρίβειας - Ανάκλησης του μοντέλου RF-DNN

## 2. Συνδυασμός Επαναλαμβανόμενου Νευρωνικού Δικτύου (RNN) – Αλγόριθμος Προσαρμοστικής Ενδυνάμωσης (AdaBoost)

Ομοίως με προηγουμένως, ο συνδυασμός σημείωσε ικανοποιητικά αποτελέσματα για το σύνολο των επιπέδων ασφαλείας. Τα ποσοστά λανθασμένων ταξινομήσεων για τα επίπεδα

‘Dangerous’ και ‘Avoidable Accident’ ήταν σχετικά χαμηλά με τιμές 45,4% και 4,01% αντίστοιχα.

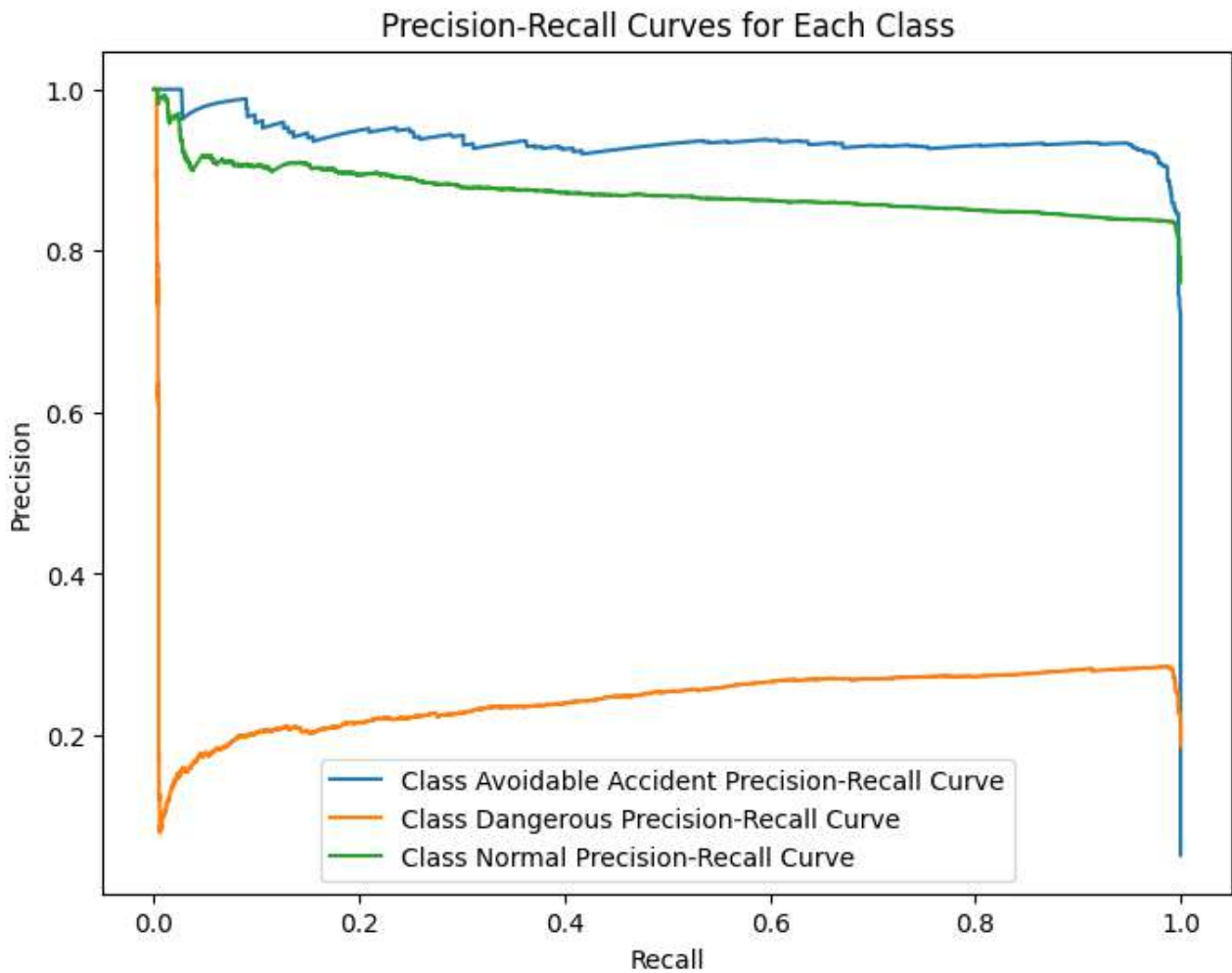


Γράφημα 5.9: Μήτρα σύγχυσης αλγόριθμου RNN- Adaboost

Στην βιβλιογραφική ανασκόπηση δεν εντοπίστηκε η ανάπτυξη του αλγορίθμου RNN-AdaBoost. Παρόλα αυτά, όπως φαίνεται από τις μετρικές αξιολόγησης του πίνακα 5.9 το μοντέλο σημείωσε υψηλές επιδόσεις.

Πίνακας 5.9: Σύνοψη μοντέλου RNN-Adaboost

Επίδοση μοντέλου ταξινόμησης Adaboost (ορθότητα: 82 %)				
Επίπεδο Ανοχής Ασφαλείας	‘Ζώνης	Ακρίβεια	Ανάκληση	F1-score
Normal		82%	90%	86%
Dangerous		83%	49%	52%
Avoidable Accident		99%	92%	94%



Γράφημα 5.10: Καμπύλη Ακρίβειας - Ανάκλησης του μοντέλου RNN -Adaboost

### 3. Συνδυασμός των Συνελικτικών Νευρωνικών Δικτύων (CNN) και Μακράς Βραχυπρόθεσμης Μνήμης (LSTM)

Ο συνδυασμός CNN&LSTM κατάφερε να ταξινομήσει ορθά μεγάλο ποσοστό των δειγμάτων για κάθε επίπεδο της 'Ζώνης Ανοχής Ασφαλείας'. Τα ποσοστά λανθασμένων προβλέψεων για τα δύο επίπεδα επικινδυνότητας για το επίπεδο ασφαλείας 'Normal' έχει ποσοστό λανθασμένων ταξινομήσεων τιμή 8,32% και το επίπεδο 'Avoidable Accident' έχει 15,33%.



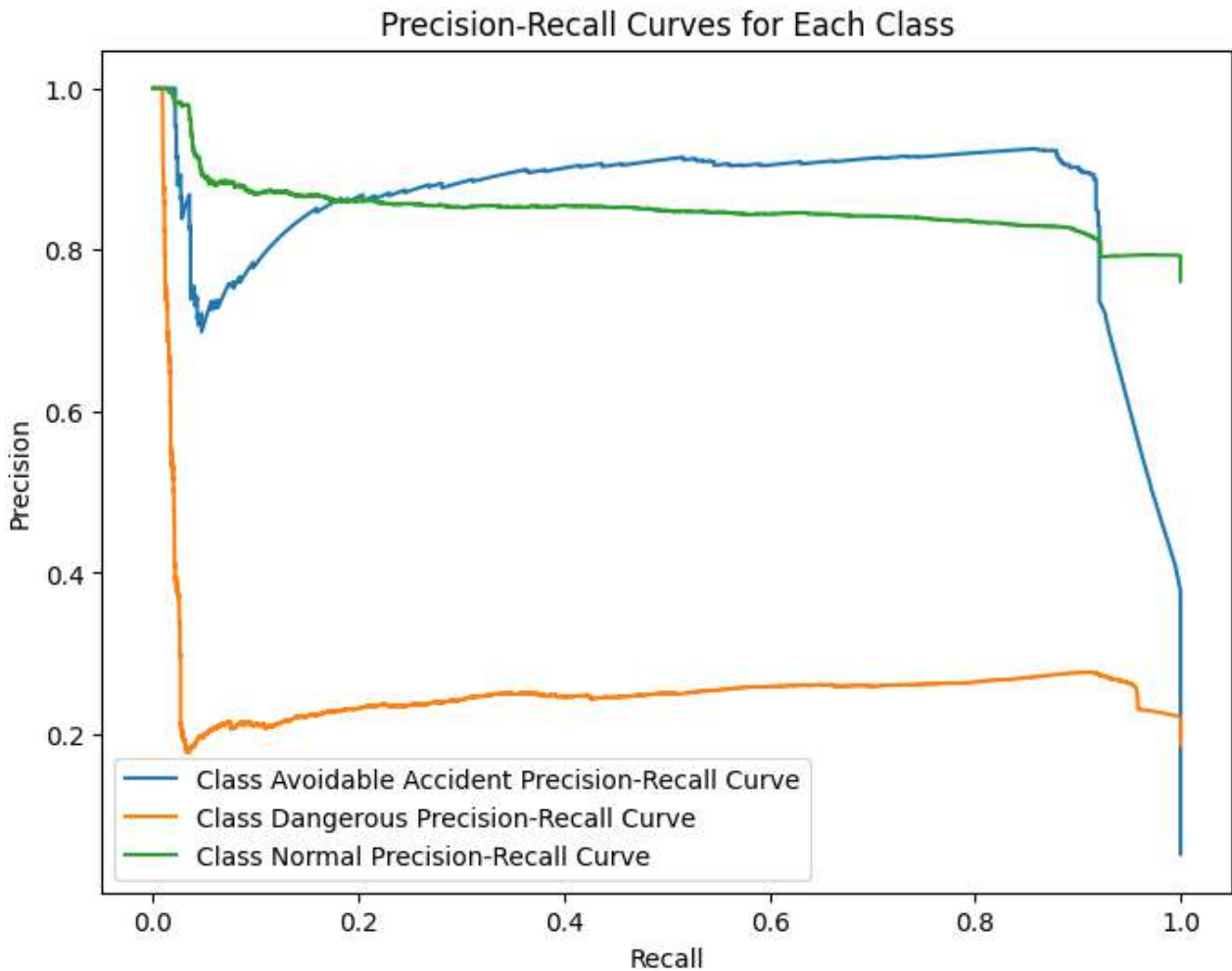
Γράφημα 5.11: Μήτρα σύγχυσης αλγόριθμου CNN-LSTM

Σε σχέση με την έρευνα των Seong Kyung Kwong et al. (2021) που η ορθότητα του μοντέλου ισούταν με 98%, η παρούσα μελέτη κατέγραψε μικρή διαφορά του ποσοστού ορθών προβλέψεων.

Πίνακας 5.10: Σύνοψη μοντέλου CNN-LSTM

Επίδοση μοντέλου ταξινόμησης CNN-LSTM (ορθότητα: 83%)

Επίπεδο Ανοχής Ασφαλείας	‘Ζώνης Ακρίβεια	Ανάκληση	F1-score
Normal	87%	91%	89%
Dangerous	57%	47%	52%
Avoidable Accident	99%	87%	92%
Μέση Τιμή	81%	75%	78%

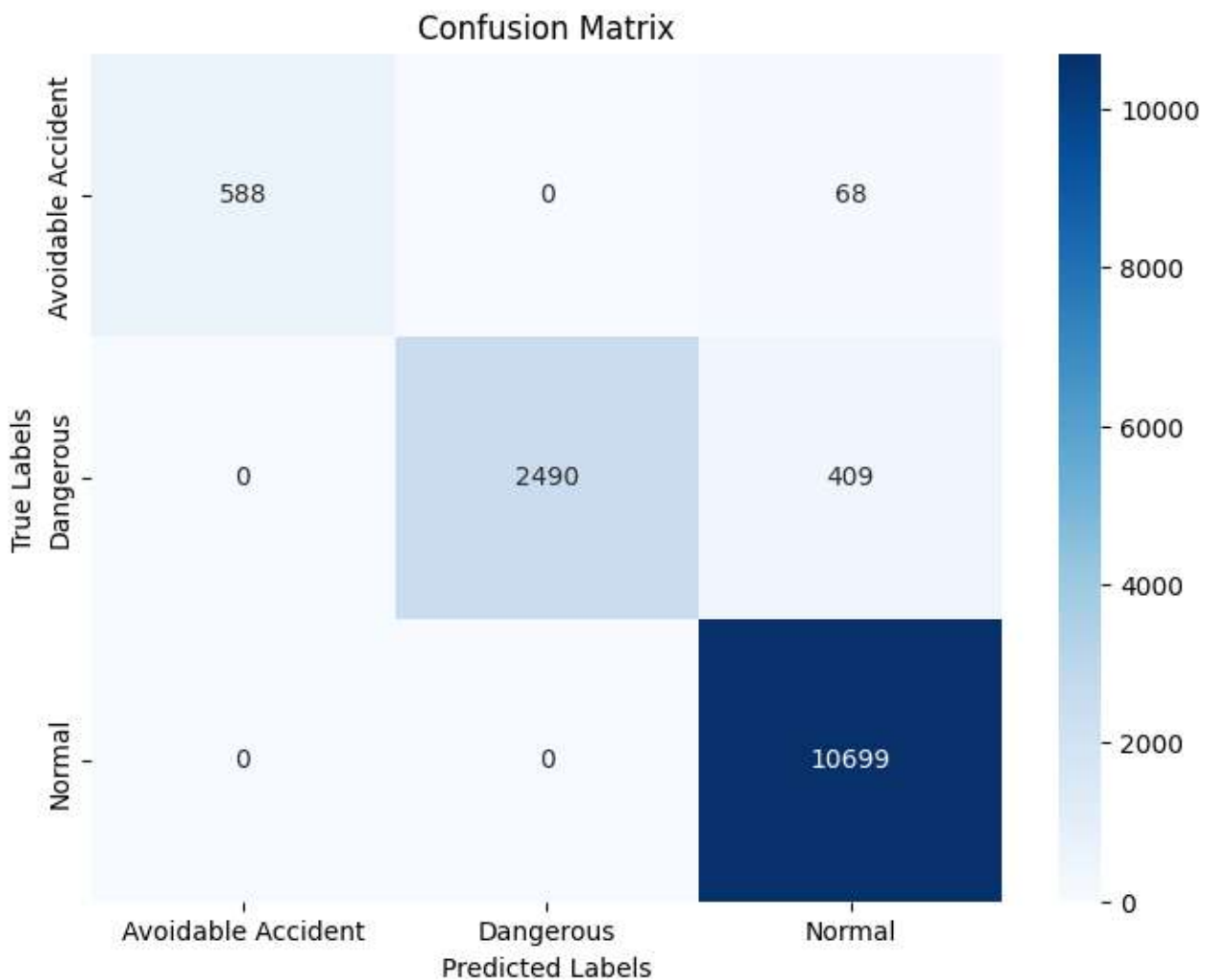


Γράφημα 5.12: Καμπύλη Ακρίβειας - Ανάκλησης του μοντέλου CNN-LSTM

## B) Δεδομένα Αγγλίας

### 1. Συνδυασμός Βαθιού Νευρωνικού Δικτύου (DNN) - Τυχαία Δάση (RF)

Όπως φαίνεται στο γράφημα 5.11 ο αλγόριθμός σημείωσε υψηλά ποσοστά αναγνώρισης και των τριών επιπέδων της 'Ζώνης Ανοχής Ασφαλείας'. Το ποσοστό λανθασμένων προβλέψεων του επιπέδου 'Normal' είναι ιδιαίτερα χαμηλό με τιμή 4,27%. Ωστόσο σχετικά με το επίπεδο ασφαλείας 'Avoidable Accident' το αντίστοιχο ποσοστό έχει τιμή 0%. Επομένως, συνολικά θεωρείται ένα ικανοποιητικό μοντέλο που έχει υψηλή ικανότητα αναγνώρισης επικίνδυνων συμπεριφορών.



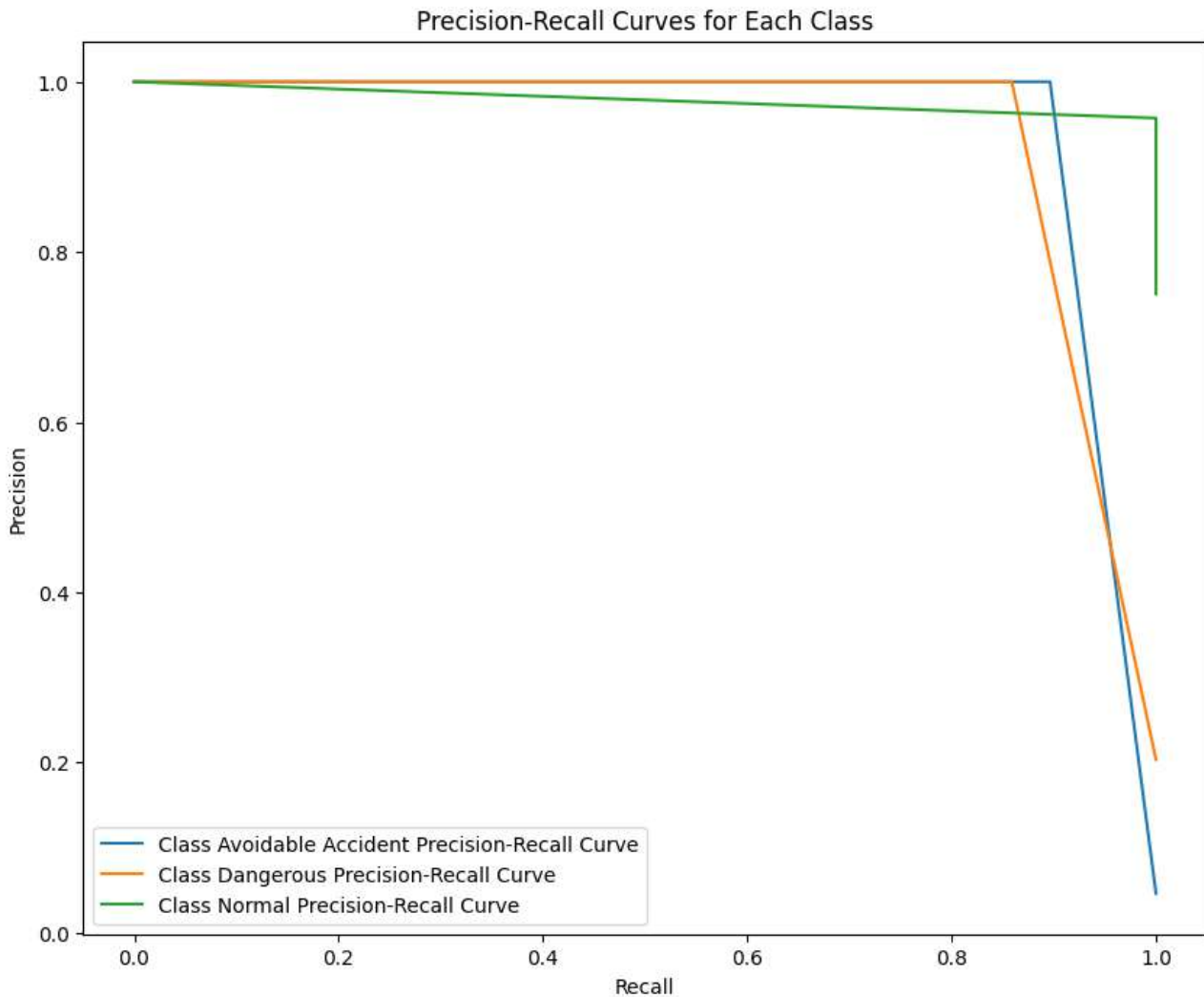
Γράφημα 5.13: Μήτρα σύγχυσης αλγόριθμου RF-DNN

Για τον συνδυασμό RF-DNN συγκρίνοντας τις μετρικές αξιολόγησης του πίνακα 5.11 με τις αντίστοιχες των ερευνών που αναλύθηκαν, προκύπτει ότι η απόδοση του μοντέλου κινήθηκε σε όμοια ποσοστά.

Πίνακας 5.11: Σύνοψη μοντέλου RF-DNN

Επίδοση μοντέλου ταξινόμησης RF (ορθότητα: 97%)

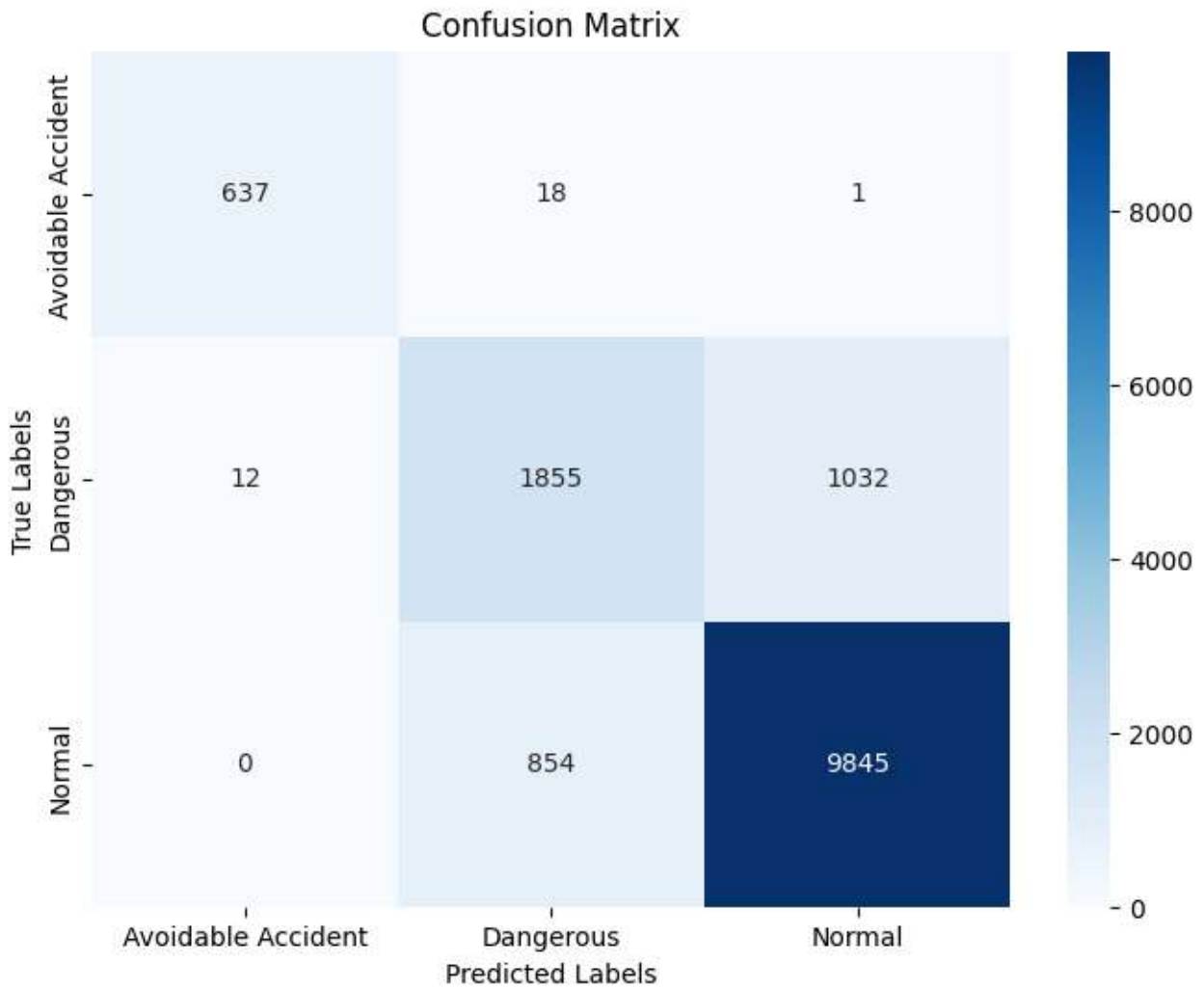
Επίπεδο Ανοχής Ασφαλείας	‘Ζώνης Ακρίβεια	Ανάκληση	F1-score
Normal	97%	100%	98%
Dangerous	97%	86%	92%
Avoidable Accident	100%	90%	95%
Μέση Τιμή	98%	92%	95%



Γράφημα 5.14: Καμπύλη Ακρίβειας - Ανάκλησης του μοντέλου RF-DNN

## 2. Συνδυασμός Επαναλαμβανόμενου Νευρωνικού Δικτύου (RNN) – Αλγόριθμου Προσαρμοστικής Ενδυνάμωσης (AdaBoost)

Ομοίως με προηγουμένως, ο συνδυασμός σημείωσε ικανοποιητικά αποτελέσματα για το σύνολο των επιπέδων ασφαλείας. Τα ποσοστά λανθασμένων ταξινομήσεων για τα επίπεδα 'Dangerous' και 'Avoidable Accident' ήταν σχετικά χαμηλά με τιμές 31,98% και 1,85% αντίστοιχα.



Γράφημα 5.15: Μήτρα σύγχυσης αλγόριθμου RNN- Adaboost

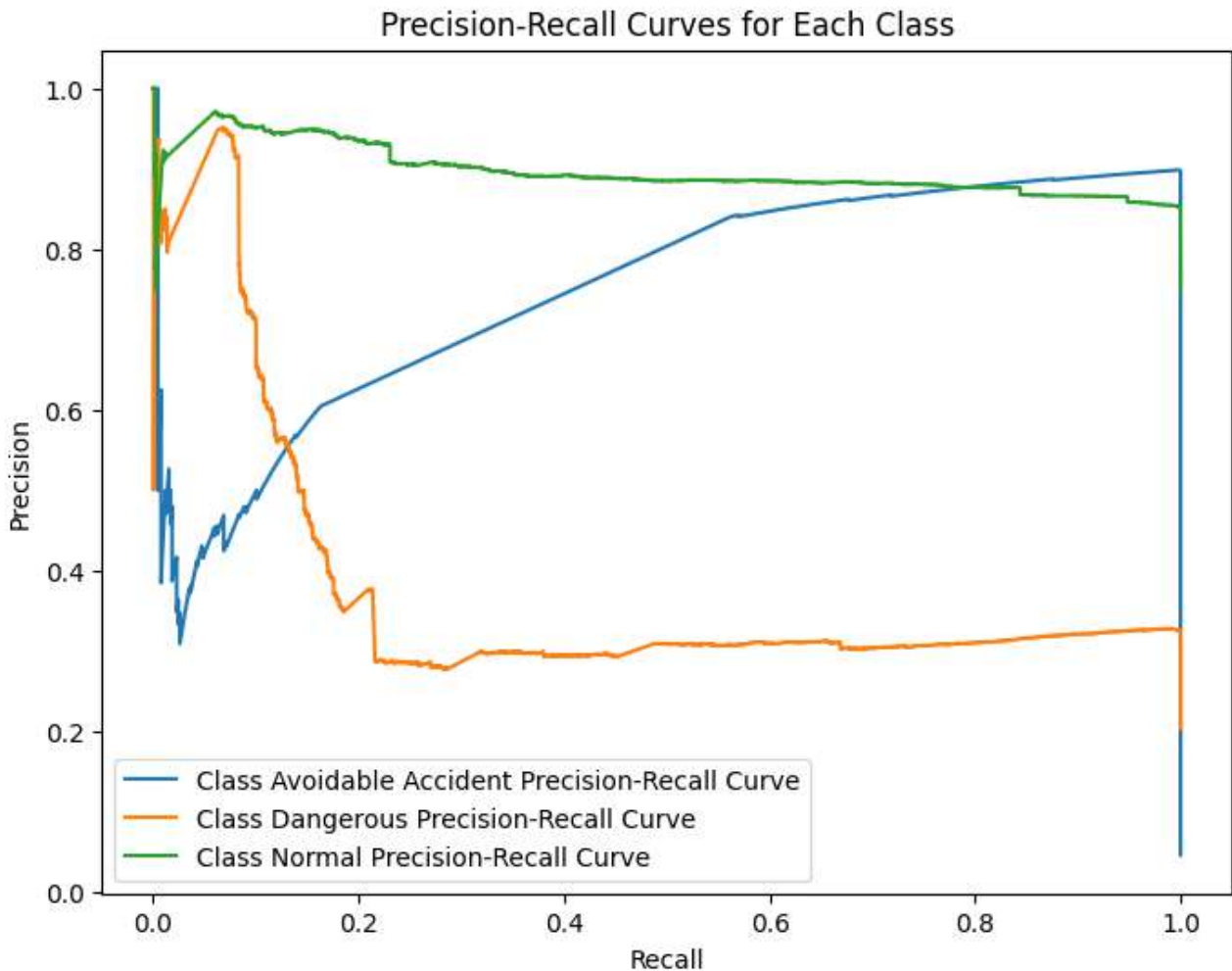
Στην βιβλιογραφική ανασκόπηση δεν εντοπίστηκε η ανάπτυξη του αλγορίθμου RNN-AdaBoost. Παρόλα αυτά, όπως φαίνεται από τις μετρικές αξιολόγησης του πίνακα 5.12 το μοντέλο σημείωσε υψηλές επιδόσεις.

Πίνακας 5.12: Σύνοψη μοντέλου RNN- Adaboost

Επίδοση μοντέλου ταξινόμησης Adaboost (ορθότητα: 87 %)

Επίπεδο Ανοχής Ασφαλείας	‘Ζώνης Ακρίβεια	Ανάκληση	F1-score
Normal	87%	92%	91%
Dangerous	87%	64%	66%
Avoidable Accident	100%	97%	98%
Μέση Τιμή	91%	84%	85%

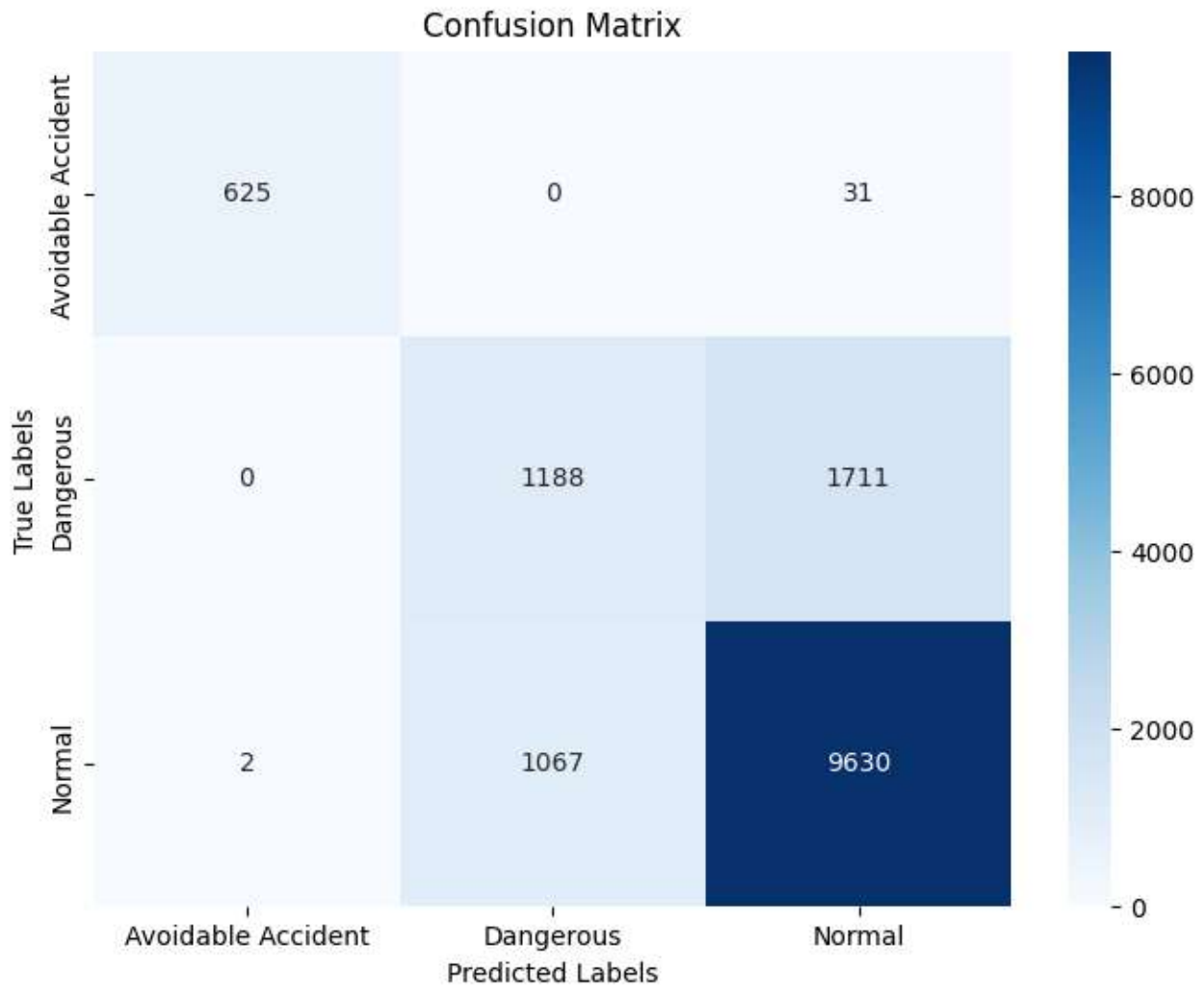




Γράφημα 5.16: Καμπύλη Ακρίβειας - Ανάκλησης του μοντέλου RNN-Adaboost

### 3. Συνδυασμός των Συνελικτικών Νευρωνικών Δικτύων (CNN) και Μακράς Βραχυπρόθεσμης Μνήμης (LSTM)

Ο συνδυασμός CNN&LSTM κατάφερε να ταξινομήσει ορθά μεγάλο ποσοστό των δειγμάτων για κάθε επίπεδο της 'Ζώνης Ανοχής Ασφαλείας'. Τα ποσοστά λανθασμένων προβλέψεων για τα δύο επίπεδα επικινδυνότητας για το επίπεδο ασφαλείας 'Normal' έχει ποσοστό λανθασμένων ταξινομήσεων τιμή 9,4% και το επίπεδο 'Avoidable Accident' έχει 4,94%.



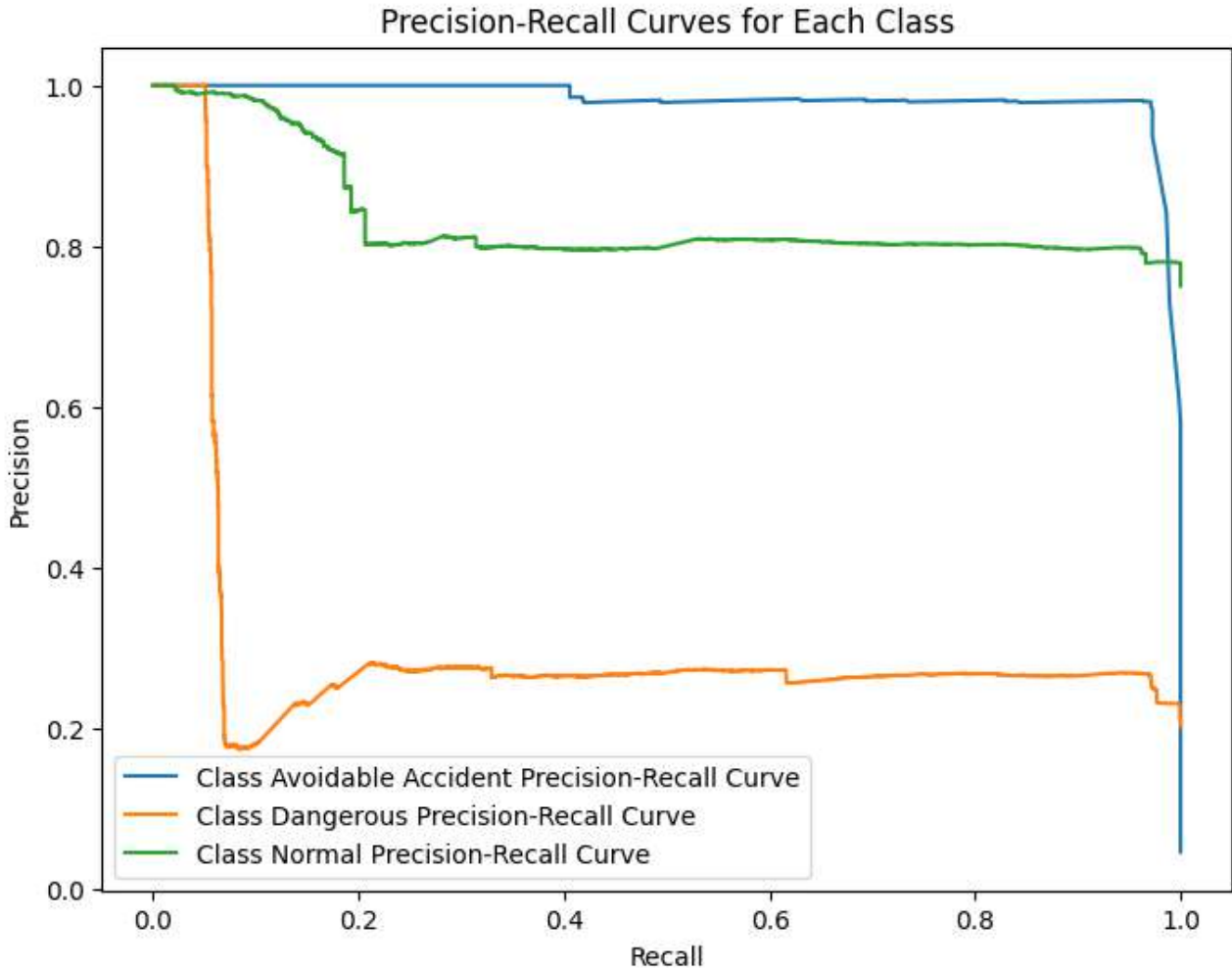
Γράφημα 5.17: Μήτρα σύγχυσης αλγόριθμου CNN-LSTM

Σε σχέση με την έρευνα των Seong Kyung Kwong et al. (2021) που η ορθότητα του μοντέλου ισούταν με 98%, η παρούσα μελέτη κατέγραψε μικρή διαφορά του ποσοστού ορθών προβλέψεων.

Πίνακας 5.13: Σύνοψη μοντέλου CNN-LSTM

Επίδοση μοντέλου ταξινόμησης CNN-LSTM (ορθότητα: 80%)

Επίπεδο Ανοχής Ασφαλείας	‘Ζώνης Ακρίβεια	Ανάκληση	F1-score
Normal	85%	90%	87%
Dangerous	53%	41%	47%
Avoidable Accident	100%	90%	97%
Μέση Τιμή	79%	75%	77%



Γράφημα 5.18: Καμπύλη Ακρίβειας - Ανάκλησης του μοντέλου CNN-LSTM

### 5.2.6 Σύγκριση Μετρικών Αξιολόγησης των μοντέλων

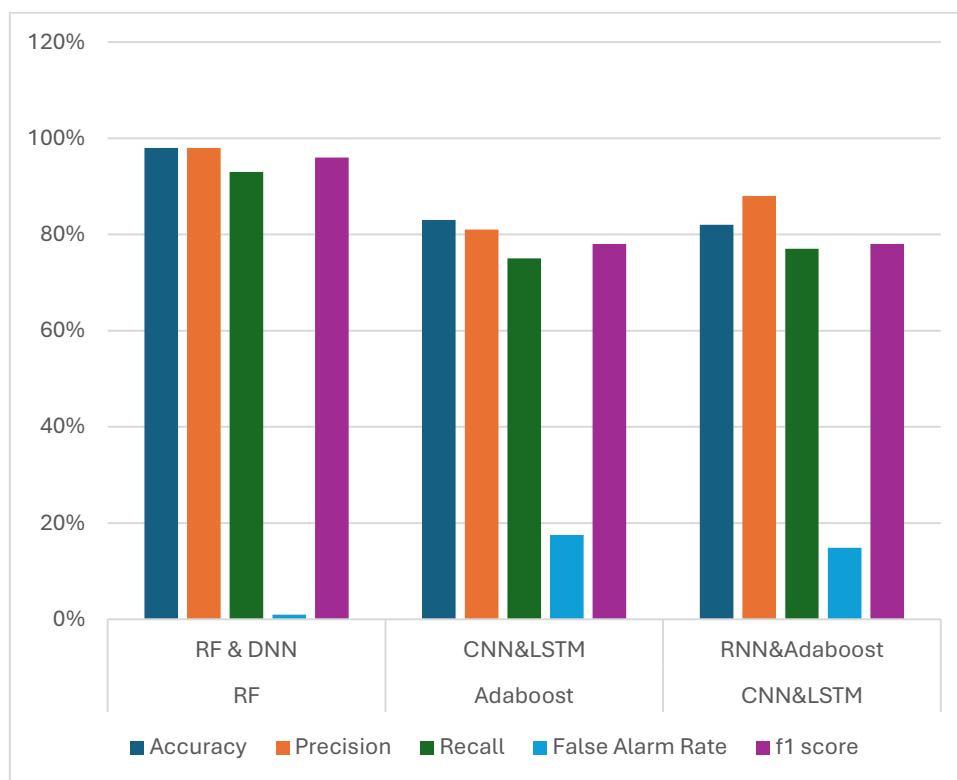
Οι διαφορετικές τεχνικές επεξεργασίας των δεδομένων καθώς και η βελτιστοποίηση των παραμέτρων των αλγορίθμων είχαν ως στόχο την βελτίωση της προγνωστικής ικανότητας των μοντέλων. Στους πίνακες 5.14 και 5.15 και στα γραφήματα 5.19 και 5.20 παρουσιάζονται μερικές σημαντικές μετρικές αξιολόγησης των τεσσάρων μοντέλων προς σύγκριση.

Πίνακας 5.14: Σύγκριση μετρικών αξιολόγησης των μοντέλων ταξινόμησης για το Βέλγιο

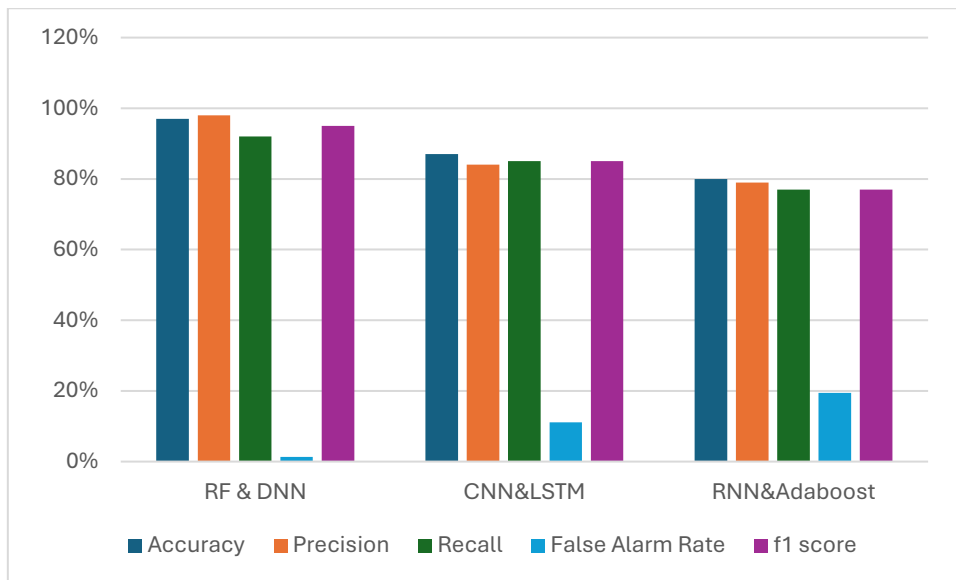
	Ορθότητα	Ακρίβεια	Ανάκληση	FPR	f1-score
RF & DNN	98%	98%	93%	0,96%	96%
CNN&LSTM	83%	81%	75%	17,5%	78%
RNN&Adaboost	82%	88%	77%	14,9%	78%

Πίνακας 5.15: Σύγκριση μετρικών αξιολόγησης των μοντέλων ταξινόμησης για την Αγγλία

	<b>Ορθότητα</b>	<b>Ακρίβεια</b>	<b>Ανάκληση</b>	<b>FPR</b>	<b>f1-score</b>
RF&DNN	97%	98%	92%	1,36%	95%
RNN&Adaboost	87%	84%	85%	11,11%	85%
CNN&LSTM	80%	79%	77%	19,4%	77%



Γράφημα 5.19: Επίδοση των μοντέλων ταξινόμησης σύμφωνα με τις μετρικές αξιολόγησης για το Βέλγιο



Γράφημα 5.20: Επίδοση των μοντέλων ταξινόμησης σύμφωνα με τις μετρικές αξιολόγησης για την Αγγλία

Σύμφωνα με τους πίνακες 5.14 και 5.15, οι αλγόριθμοι αποδίδουν υψηλά ποσοστά ορθότητας (accuracy), ανάκλησης (recall), ακρίβειας (precision) και f1-score τα οποία δεν έχουν μεγάλη απόκλιση μεταξύ τους. Όπως αναφέρθηκε στην ενότητα 3.6, η λανθασμένη ταξινόμηση δεδομένων επικίνδυνου επιπέδου σε λιγότερο επικίνδυνο θα είχε σοβαρές επιπτώσεις στην οδική ασφάλεια. Αυτό καθιστά την **ανάκληση** σημαντική μετρική αξιολόγησης. Για την κλάση ‘Avoidable Accident’ η εμφάνιση υψηλότερης ανάκλασης παράλληλα με χαμηλότερο ποσοστό ακρίβειας συνεπάγεται υψηλή ικανότητα αναγνώρισης του πραγματικού επικίνδυνου επιπέδου αλλά μεγαλύτερο ποσοστό λανθασμένων ταξινομήσεων των κλάσεων ‘Normal’ και ‘Dangerous’ ως ‘Avoidable Accident’. Η παραπάνω περίπτωση θεωρείται ανεκτή στα πλαίσια της παρούσας διπλωματικής εργασίας. Όμως σε περίπτωση αντίθετων αποτελεσμάτων θα παρουσιάζονταν σοβαρά προβλήματα.

Ελέγχοντας την ορθότητα (Accuracy), την ανάκληση (Recall) και τον ρυθμό λανθασμένων θετικών προβλέψεων (FPR) των τεσσάρων μοντέλων για κάθε χώρα προκύπτει ότι τα πιο επιθυμητά αποτελέσματα προσφέρει ο συνδυαστικός αλγόριθμος ‘Βαθύ Νευρωνικού Δικτύου και αλγόριθμου Προσαρμοστικής Ενδυνάμωσης’ (Deep Neural Network-DNN & AdaBoost) για το Βέλγιο και για την Αγγλία.

### 5.3 Εξήγηση λειτουργίας μοντέλων μηχανικής και βαθιάς μάθησης

Τέλος, θέλοντας να κατανοήσουμε καλύτερα τα συνδυασμένα μοντέλα θα αναπτύξουμε τον αλγόριθμο Lime (Local Interpretable Model-agnostic explanations). Όπου ο αλγόριθμος Lime χρησιμοποιείται ως υποστηρικτικό εργαλείο για την ανάλυση της επιρροής των ανεξάρτητων μεταβλητών στην πρόβλεψη των συνδυασμένων μοντέλων. Το συνδυασμένο μοντέλο που απέδωσε καλύτερα είναι το ‘Βαθύ Νευρωνικό Δίκτυο και αλγόριθμο Προσαρμοστικής Ενδυνάμωσης’ για το οποίο προέκυψε το παρακάτω διάγραμμα για κάθε χώρα. Τα διαγράμματα που προκύπτουν από την εξήγηση LIME παρέχουν πληροφορίες σχετικά με την επίδραση των χαρακτηριστικών στις προβλέψεις του συνδυασμένου μοντέλου (Random Forest και Deep Neural Network). Η ερμηνεία παρέχει πληροφορίες

σχετικά με τον τρόπο με τον οποίο κάθε χαρακτηριστικό επηρεάζει την τελική πρόβλεψη του συνδυασμένου μοντέλου.

Πίνακας 5.16: Αποτελέσματα Lime για το Βέλγιο

Μεταβλητή	Σημαντικότητα Μεταβλητής
GPS_spd_mean	0,14
GPS_distances_sum	0,31
DEM_evt_ha_lvl_L_mean	2,00
DEM_evt_hb_lvl_L_mean	0,56

Πίνακας 5.16: Αποτελέσματα Lime για την Αγγλία

Μεταβλητή	Σημαντικότητα Μεταβλητής
GPS_spd_mean	0,29
GPS_distances_sum	0,58
DEM_evt_ha_lvl_L_mean	2,00
DEM_evt_hb_lvl_L_mean	0,13

Από τα διαγράμματα μπορούμε να συμπεράνουμε τα εξής:

1. **Σημαντικότητα Χαρακτηριστικών:** Καθένα από τα χαρακτηριστικά που παρουσιάζονται στο πίνακα έχει μια συγκεκριμένη επιρροή στις προβλέψεις.
2. **Κατεύθυνση Επίδρασης:** Προκύπτει η κατεύθυνση που προσφέρει κάθε χαρακτηριστικό στην πρόβλεψη.
3. **Αξιοπιστία Προβλέψεων:** Με τις εξηγήσεις LIME, μπορεί να αξιολογηθεί η αξιοπιστία των προβλέψεων σε συγκεκριμένες περιπτώσεις.

Δεδομένα Βελγίου:

- Τα χαρακτηριστικά που επηρεάζουν περισσότερο την τελική πρόβλεψη είναι οι μεταβλητές DEM\_evt\_ha\_lvl\_L\_mean και η DEM\_evt\_hB\_lvl\_L\_mean.

Δεδομένα Αγγλίας:

- Τα χαρακτηριστικά που επηρεάζουν περισσότερο την τελική πρόβλεψη είναι οι μεταβλητές DEM\_evt\_ha\_lvl\_L\_mean και GPS\_distances\_sum.

## 5.4 Σύνοψη

Στο κεφάλαιο αυτό εξετάστηκαν διάφοροι οδηγικοί παράγοντες που επηρεάζουν την επικίνδυνη οδήγηση. Η διερεύνηση της επιρροής των παραγόντων αυτών έγινε βάσει ενός

συστήματος αναγνώρισης του επιπέδου τη 'Ζώνης Ανοχής Ασφαλείας'. Συγκεκριμένα αναπτύχθηκαν τρία συνδυασμένα μοντέλα βαθιάς και μηχανικής εκμάθησης και εξετάστηκε η επίδραση των διαφορετικών μεταβλητών στο σύνολο αυτών μέσω ορισμένων τεχνικών επιλογής και επεξεργασίας στοιχείων. Οι μεταβλητές που επιδρούσαν άμεσα στην αναγνώριση και κατ' επέκταση στην ίδια την επικίνδυνη οδηγική συμπεριφορά ήταν για το Βέλγιο η **DEM\_evt\_ha\_lvl\_L\_mean** και η **DEM\_evt\_hB\_lvl\_L\_mean** και για την Αγγλία οι μεταβλητές **DEM\_evt\_ha\_lvl\_L\_mean** και **GPS\_distances\_sum**.

Τα μοντέλα ταξινόμησης στο σύνολο τους είχαν ικανοποιητικά αποτελέσματα για την αναγνώριση του επιπέδου της 'Ζώνης Ανοχής Ασφαλείας' που βρίσκεται ο οδηγός σε κάθε χρονικό πλαίσιο των 30 δευτερολέπτων. Παρόλα αυτά συγκρίνοντας τις μετρικές αξιολόγησης προέκυψε ότι το μοντέλο **'Βαθιών Νευρωνικών Δικτύων (Deep Neural Network) - Τυχαίων Δασμών (Random Forests)'** είχε τα πιο ικανοποιητικά αποτελέσματα για το σύνολο των επιπέδων ασφάλειας.

Επίσης, με βάση την δεύτερη προσέγγιση της παρούσας μελέτης στην οποία αναπτύχθηκε ο αλγόριθμος Lime, προκύπτει ότι οι ανεξάρτητες μεταβλητές που επηρεάζουν περισσότερο την τελική πρόβλεψη και συμβάλουν θετικά στην ταξινόμηση του επιπέδου οδήγησης ως Normal είναι οι **DEM\_evt\_ha\_lvl\_L\_mean**, **DEM\_evt\_hB\_lvl\_L\_mean** για το Βέλγιο. Οι ανεξάρτητες μεταβλητές που επηρεάζουν περισσότερο την τελική πρόβλεψη για την Αγγλία είναι οι **GPS\_distances\_sum**, **DEM\_evt\_ha\_lvl\_L\_mean**, οι οποίες συμβάλουν θετικά στην ταξινόμηση του επιπέδου οδήγησης ως Normal και ως Not Normal αντίστοιχα.

Αξίζει να αναλυθούν σε βάθος τα συμπεράσματα και τα αποτελέσματα που προέκυψαν, ώστε να συμβάλλουν στην καλύτερη κατανόηση των παραγόντων που επιδρούν στην αναγνώριση της επικίνδυνης οδήγησης. Τέλος, είναι αναγκαίο να επισημανθούν η συνεισφορά και οι σημαντικές ελλείψεις της έρευνας, ώστε να αποτελέσουν τη βάση για περαιτέρω διερεύνηση.

## 6. ΣΥΜΠΕΡΑΣΜΑΤΑ

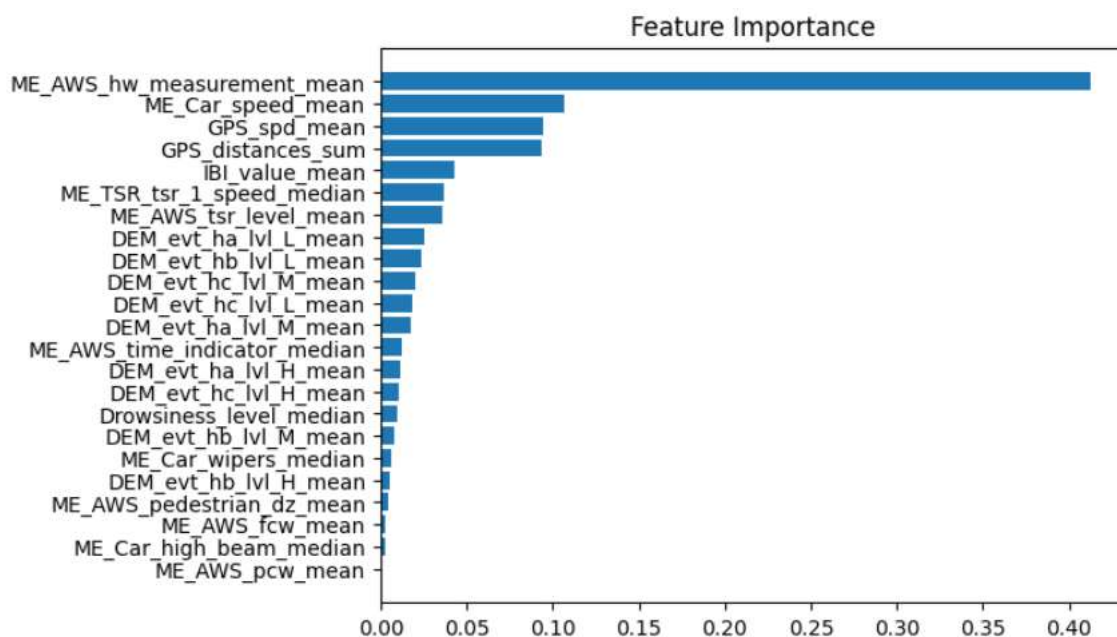
### 6.1 Σύνοψη αποτελεσμάτων

Στόχος της παρούσας διπλωματικής εργασίας αποτελεί η διερεύνηση της επικίνδυνης συμπεριφοράς του οδηγού (**Ζώνη Ανοχής Ασφαλείας**) μέσω της αξιοποίησης δεδομένων οδήγησης σε πραγματικές συνθήκες. Τα δεδομένα που συλλέχθηκαν παράχθηκαν από πείραμα σε πραγματικές συνθήκες οδήγησης στο πλαίσιο του ερευνητικού προγράμματος i-DREAMS, στο οποίο διεξήχθη ένα νατουραλιστικό πείραμα οδήγησης στο οποίο συμμετείχαν 43 οδηγοί από το Βέλγιο και δημιουργήθηκε μια μεγάλη βάση δεδομένων με 7163 ταξίδια και 147.337 λεπτά. Για τους οδηγούς του Ηνωμένου Βασιλείου, το σύνολο δεδομένων περιλάμβανε 26 οδηγούς, 8226 ταξίδια και 118.175 λεπτά. Συγκεκριμένα η διερεύνηση της επικίνδυνης συμπεριφοράς του οδηγού θα πραγματοποιηθεί σε τρία επίπεδα ασφαλείας με έμφαση στη Ζώνη Ανοχής Ασφαλείας, καθώς και ο εντοπισμός των χαρακτηριστικών που επιδρούν στη διαδικασία αυτή.

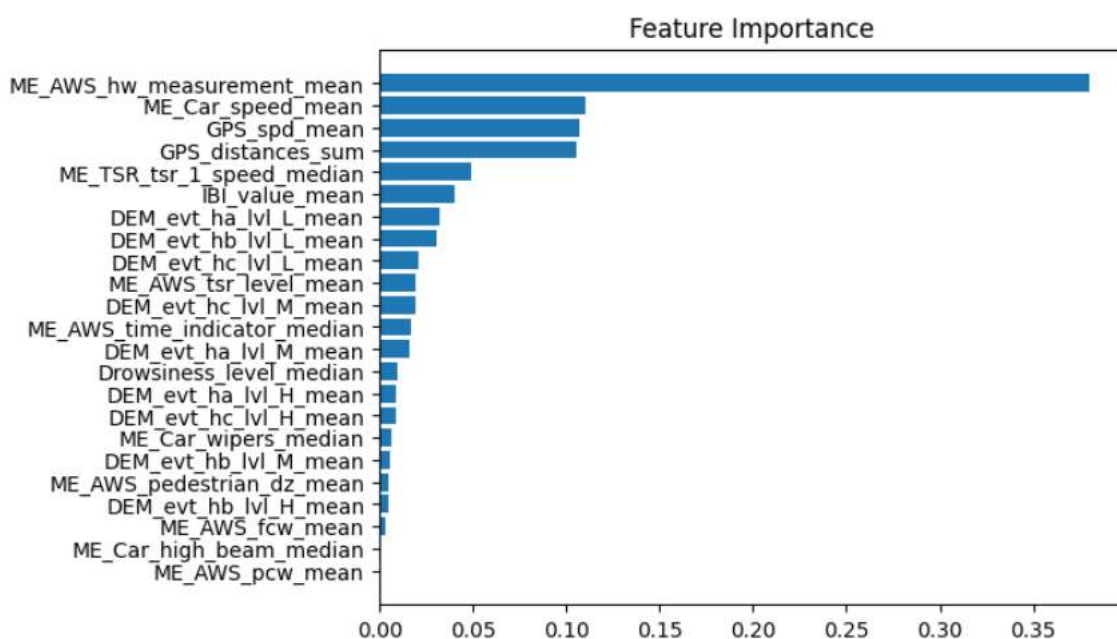
Τελικά ο καθορισμός των επιπέδων ασφαλείας πραγματοποιήθηκε με την χρήση της μεταβλητής **Headway**, καθώς το επίπεδο 'Κανονικό' να είναι η κύρια κατηγορία με το υψηλότερο ποσοστό δειγμάτων, ενώ τα επίπεδα 'Επικίνδυνο' και 'Αποφυγή Ατυχήματος' είναι η μειοψηφική κατηγορία με το χαμηλότερο ποσοστό δειγμάτων.

Στο πρώτο μέρος των αναλύσεων αναπτύχθηκαν κατάλληλες τεχνικές προσδιορισμού της **σημαντικότητας των μεταβλητών** στην πρόβλεψη του επιπέδου 'Ζώνης Ανοχής Ασφαλείας' που βρίσκεται ο οδηγός. Επισημαίνεται ότι επιλέγοντας να γίνει ο καθορισμός του επιπέδου ασφαλείας με την χρήση της μεταβλητής Headway, τα διάφορα περιγραφικά στατιστικά στοιχεία του παράγοντα Headway δεν θα αποτελέσουν μεταβλητές εισόδου στα μοντέλα. Ο συνυπολογισμός τους στην διαδικασία της ταξινόμησης θα οδηγούσε σε μεροληψία του μοντέλου χωρίς εκείνο να προσφέρει χρήσιμα και σημαντικά αποτελέσματα. Επομένως, η μεταβλητή ME\_AWS\_hw\_measurement\_mean δεν θα αποτελέσει μία από τις μεταβλητές εισόδου. Η σημαντικότητα φαίνεται στα γραφήματα που ακολουθούν.





Γράφημα 6.1: Σημαντικότητα μεταβλητών σύμφωνα με την μέθοδο ‘Τυχαίων Δασών’ για το Βέλγιο



Γράφημα 6.2: Σημαντικότητα μεταβλητών σύμφωνα με την μέθοδο ‘Τυχαίων Δασών’ για την Αγγλία

Έπειτα, αξιοποιώντας τις τέσσερις πιο σημαντικές μεταβλητές αναπτύχθηκαν τρεις συνδυασμοί αλγορίθμων μηχανικής και βαθιάς μάθησης με σκοπό την ταξινόμηση της οδηγικής συμπεριφοράς σε ένα από τα τρία επίπεδα ασφαλείας. Με την εφαρμογή της τεχνικής **Synthetic Minority Oversampling Technique (SMOTE)** αντιμετωπίζεται το πρόβλημα της ανισορροπίας των δεδομένων εκπαίδευσης ως προς τις διαφορετικές κλάσεις.

Η ονοματολογία και ο συμβολισμός των τριών αλγορίθμων για το Βέλγιο και την Αγγλία παρατίθενται στον πίνακα, ενώ οι επιδόσεις τους παρουσιάζονται στον πίνακα και στο γράφημα που ακολουθεί.

Πίνακας 6.1: Ονοματολογία και συμβολισμός μοντέλων ταξινόμησης

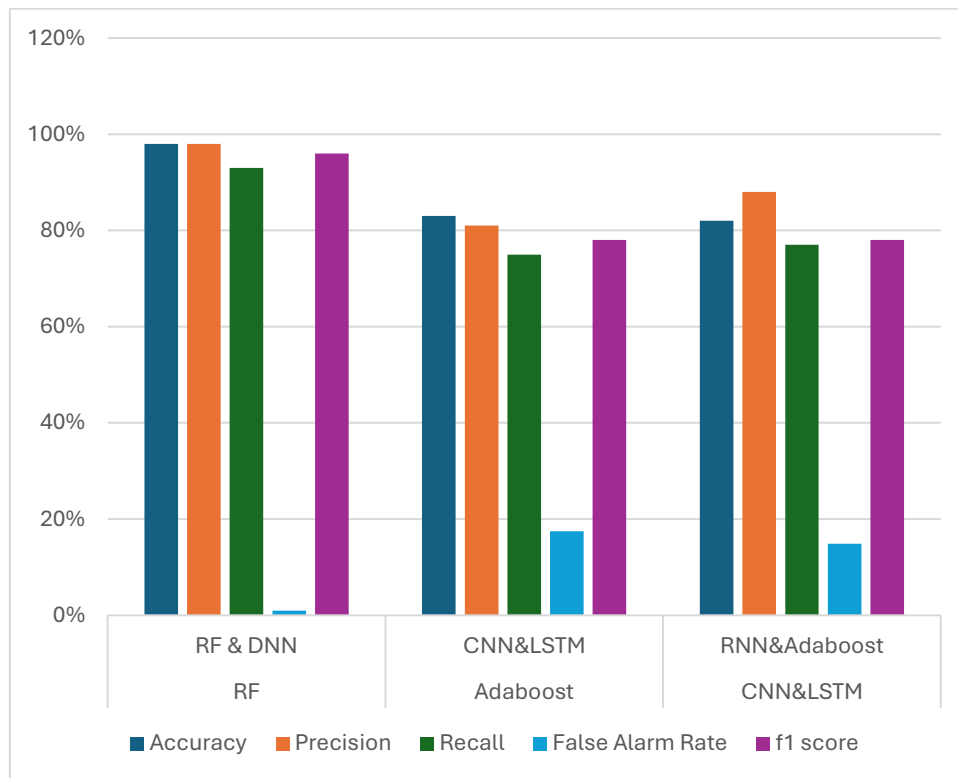
Όνομα Μοντέλου (Ελληνικά)	Όνομα Μοντέλου (Αγγλικά)	Συμβολισμός Μοντέλου
Ταξινομητής Τυχαίων Δασών	Random Forests Classifier	RF
Ταξινομητής Adaboost	Adaboost Classifier	Adaboost
Συνελεκτικά Νευρωνικά Δίκτυα	Convolutional Neural Networks	CNN
Μοντέλο Μακράς-Βραχείας Μνήμης	Long-Short Term Memory Classifier	LSTM
Μοντέλο Επαναλαμβανόμενων Νευρωνικών Δικτύων	Recurrent Neural Network	RNN
Μοντέλο Βαθιών Νευρωνικών Δικτύων	Deep Neural Network	DNN

Πίνακας 6.2: Σύγκριση μετρικών αξιολόγησης των μοντέλων ταξινόμησης για το Βέλγιο

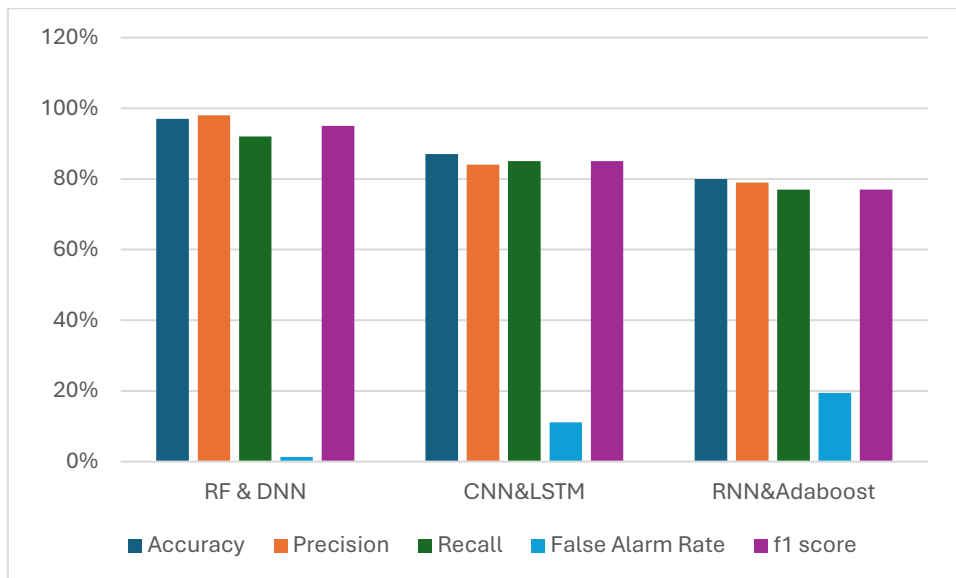
	Ορθότητα	Ακρίβεια	Ανάκληση	FPR	f1-score
RF & DNN	98%	98%	93%	0,96%	96%
CNN&LSTM	83%	81%	75%	17,5%	78%
RNN&Adaboost	82%	88%	77%	14,9%	78%

Πίνακας 6.3: Σύγκριση μετρικών αξιολόγησης των μοντέλων ταξινόμησης για την Αγγλία

	<b>Ορθότητα</b>	<b>Ακρίβεια</b>	<b>Ανάκληση</b>	<b>FPR</b>	<b>f1-score</b>
RF&DNN	97%	98%	92%	1,36%	95%
RNN&Adaboost	87%	84%	85%	11,11%	85%
CNN&LSTM	80%	79%	77%	19,4%	77%



Γράφημα 6.3: Επίδοση των μοντέλων ταξινόμησης σύμφωνα με τις μετρικές αξιολόγησης για το Βέλγιο



Γράφημα 6.4: Επίδοση των μοντέλων ταξινόμησης σύμφωνα με τις μετρικές αξιολόγησης για την Αγγλία

Στο δεύτερο μέρος των αναλύσεων για να έχουμε μία επισκόπηση των ανεξάρτητων μεταβλητών που είναι πιο σημαντικές και για να κατανοήσουμε καλύτερα τα συνδυασμένα μοντέλα θα αναπτύξουμε τον αλγόριθμο **Lime (Local Interpretable Model-agnostic explanations)**. Σύμφωνα με τα διαγράμματα του αλγόριθμου Lime μπορούμε να συμπεράνουμε τα εξής:

Δεδομένα Βελγίου:

- Τα χαρακτηριστικά που επηρεάζουν περισσότερο την τελική πρόβλεψη είναι οι μεταβλητές DEM\_evt\_ha\_lvl\_L\_mean και η DEM\_evt\_hB\_lvl\_L\_mean.

Δεδομένα Αγγλίας:

- Τα χαρακτηριστικά που επηρεάζουν περισσότερο την τελική πρόβλεψη είναι οι μεταβλητές DEM\_evt\_ha\_lvl\_L\_mean και GPS\_distances\_sum.

## 6.2 Σύνοψη Συμπερασμάτων

Βάσει των αποτελεσμάτων που προέκυψαν κατά την εφαρμογή της μεθοδολογίας, προέκυψαν ορισμένα **συμπεράσματα** άμεσα σχετιζόμενα με τον στόχο της διπλωματικής εργασίας.

- Σύμφωνα με τα αποτελέσματα των γραφημάτων 6.1 και 6.2 για το Βέλγιο και την Αγγλία αντίστοιχα, συμπεραίνουμε ότι η χρονική απόσταση του οχήματος από το προπορευόμενο όχημα είναι η σημαντικότερη μεταβλητή για την αναγνώριση της οδηγικής συμπεριφοράς. Καθώς όμως τα επίπεδα 'Ζώνης Ανοχής Ασφαλείας' έχουν προκύψει από την μεταβλητή Headway, η μεταβλητή αυτή δεν θα αποτελέσει μεταβλητή εισόδου για τα μοντέλα ταξινόμησης. Ο συνυπολογισμός τους στην διαδικασία της ταξινόμησης θα οδηγούσε σε μεροληψία του μοντέλου χωρίς εκείνο να προσφέρει χρήσιμα και σημαντικά αποτελέσματα.

- Η ταχύτητα του οχήματος έχει επίσης μεγάλο δείκτη σημαντικότητας. Η ταχύτητα σχετίζεται άμεσα με την πιθανότητα εμφάνισης ατυχήματος καθώς επίσης και με την σοβαρότητα αυτού. Για παράδειγμα, η αυξημένη ταχύτητα του οχήματος ελαχιστοποιεί τον χρόνο αντίδρασης του οδηγού.
- Η συνολική διανυθείσα απόσταση αποτελεί επίσης σημαντική μεταβλητή για την αναγνώριση της οδηγικής συμπεριφοράς. Ανάλογα με την συνολική απόσταση που διανύει ο οδηγός μπορεί να παρατηρηθούν διαφορετικές οδηγικές συμπεριφορές. Για παράδειγμα, οι οδηγοί που διανύουν μεγάλες αποστάσεις είναι πιθανό να εμφανίσουν σημάδια κούρασης και μειωμένης προσοχής, τα οποία οδηγούν σε επικίνδυνη οδηγική συμπεριφορά.
- Η μεταβλητή DEM\_evt\_ha\_lvl\_L\_mean, η οποία αφορά την εμφάνιση ή μη απότομων επιταχύνσεων, επηρεάζει επίσης την διαδικασία ταξινόμησης στα μοντέλα. Για παράδειγμα, οι απότομες εναλλαγές της επιτάχυνσης των οχημάτων ελαχιστοποιούν το χρόνο αντίδρασης του οδηγού και αποσπούν την προσοχή του οδηγού από το οδικό περιβάλλον, με αποτέλεσμα να οδηγείται σε επικίνδυνες συμπεριφορές.
- Από τις διαφορετικές τεχνικές αντιμετώπισης του φαινομένου της άνισης κατανομής των δειγμάτων στις διαφορετικές κλάσεις, η τεχνική της συνθετικής μειονοτικής υπερδειγματοληψίας (Synthetic minority oversampling technique, SMOTE) πρόσφερε τα βέλτιστα αποτελέσματα για το σύνολο των ταξινομητών. Όπως αναφέρθηκε στο κεφάλαιο 3.3, η τεχνική SMOTE επικεντρώνεται στον «χώρο των χαρακτηριστικών» και όχι στον «χώρο δεδομένων», με άλλα λόγια, ο αλγόριθμος βασίζεται στις τιμές των χαρακτηριστικών και στη σχέση τους, αντί να θεωρεί τα σημεία δεδομένων ως ολόκληρος.
- Στην παρούσα εργασία αναπτύχθηκαν τρεις συνδυασμοί μοντέλων μηχανικής και βαθιάς μάθησης, οι οποίοι σημείωσαν ικανοποιητικές επιδόσεις. Ο συνδυασμός 'Βαθύ Νευρωνικού Δικτύου (Deep Neural Network) - Τυχαίων Δασών (Random Forests)' σημείωσε τις υψηλότερες επιδόσεις στην πλειοψηφία των μετρικών αξιολόγησης τους.
- Από την αξιολόγηση των μετρικών αξιολόγησης για κάθε χώρα, συμπεραίνεται ότι όλα τα παραπάνω μοντέλα ταξινόμησης μπορούν να αποφέρουν επιθυμητά αποτελέσματα και για τις δύο χώρες.
- Από το σύνολο των μεταβλητών που εξετάστηκαν οι μεταβλητές DEM\_evt\_ha\_lvl\_L\_mean, DEM\_evt\_hB\_lvl\_L\_mean και DEM\_evt\_ha\_lvl\_L\_mean, GPS\_distances\_sum για το Βέλγιο και την Αγγλία αντίστοιχα, είχαν την μεγαλύτερη επιρροή στην τελική πρόβλεψη σύμφωνα με την ανάπτυξη του αλγόριθμου Lime.
- Η υψηλή επίδραση των μεταβλητών, DEM\_evt\_ha\_lvl\_L\_mean και DEM\_evt\_hB\_lvl\_L\_mean, που σχετίζονται με τα απότομα περιστατικά οδήγησης υποδεικνύει ότι η οδηγική συμπεριφορά των οδηγών στο Βέλγιο επηρεάζεται κυρίως από την αποφυγή απότομων περιστατικών οδήγησης και είναι ευαίσθητη σε αυτά τα περιστατικά.
- Η υψηλή επίδραση της μεταβλητής DEM\_evt\_ha\_lvl\_L\_mean για την Αγγλία υποδεικνύει ότι η αποφυγή της απότομης οδήγησης είναι σημαντική και στις δύο χώρες. Επιπλέον, η υψηλή επιρροή της GPS\_distances\_sum υποδεικνύει ότι οι αποστάσεις που καλύπτονται έχουν επίσης σημαντική συνεισφορά στον τρόπο οδήγησης. Επομένως, οι οδηγοί στην Αγγλία επηρεάζονται από την απόσταση που διανύουν σε μεγαλύτερο βαθμό.
- Από την εκπόνηση της συγκεκριμένης Διπλωματικής Εργασίας προκύπτει ότι τα δεδομένα που συλλέγονται από τα έξυπνα συστήματα, τα συστήματα αιχμής και περαιτέρω εύρυνες παρέχουν ιδιαίτερα σημαντικές πληροφορίες οι οποίες, με την

κατάλληλη επεξεργασία και ανάπτυξη μαθηματικών μοντέλων, μπορούν να ωφελήσουν στην εξαγωγή χρήσιμων συμπερασμάτων για τις κρίσιμες παραμέτρους που επηρεάζουν την συμπεριφορά του οδηγού κατά τη διάρκεια οδήγησης, αλλά και την κυκλοφοριακή συμπεριφορά των οδηγών.

- Συνολικά συμπεραίνεται ότι η ασφάλεια και η αποφυγή απότομης οδήγησης είναι κοινή και για τις δύο χώρες και οι διαφορές στις επιρροές μπορεί να υποδεικνύουν ότι οι κοινωνιο-οικονομικές, πολιτιστικές ή γεωγραφικές διαφορές επηρεάζουν την οδική συμπεριφορά σε κάθε χώρα. Συνοψίζοντας, οι οδηγοί στο Βέλγιο επικεντρώνονται ιδιαίτερα στην αποφυγή απότομης οδήγησης, ενώ στην Αγγλία η ασφάλεια συνδέεται επίσης και με τις αποστάσεις που διανύουν.
- Από τα αποτελέσματα των αναλύσεων, συμπεραίνεται ότι οι δύο χώρες έχουν παρόμοια οδηγικά χαρακτηριστικά. Ωστόσο παρατηρείται μία μικρή αύξηση στην σημαντικότητα των μεταβλητών `ME_car_speed_mean`, `GPS_spd_mean` και `GPS_distances_sum` για την Αγγλία, που υποδεικνύει ότι επικρατούν υψηλότερες ταχύτητες και αποστάσεις ταξιδιού.

### 6.3 Προτάσεις για την αξιοποίηση των αποτελεσμάτων

Με γνώμονα τα αποτελέσματα και τα συμπεράσματα που εξήχθησαν κατά την εκπόνηση της παρούσας διπλωματικής εργασίας, παρακάτω παραθέτετε ένα σύνολο **προτάσεων αξιοποίησης** των αποτελεσμάτων, οι οποίες θα μπορούσαν να συμβάλλουν στην καλύτερη κατανόηση της επιρροής των διάφορων παραγόντων στην οδική ασφάλεια, αλλά και στην εξέλιξη της έρευνας των Ευφυών Μεταφορικών Συστημάτων (ITS).

- **Αξιοποίηση των συνδυαστικών μοντέλων** ταξινόμησης για τον καθορισμό του επιπέδου ασφαλείας των οδηγών σε πραγματικές συνθήκες οδήγησης. Σύμφωνα με τις επιδόσεις των τεσσάρων συνδυαστικών αλγόριθμων ταξινόμησης, συμπεραίνεται ότι μπορούν να προσφέρουν ικανοποιητικά αποτελέσματα και έτσι θα μπορούσαν να χρησιμεύσουν για την περαιτέρω διερεύνηση της οδηγικής συμπεριφοράς.
- Μία **περαιτέρω μελέτη των σημαντικότερων μεταβλητών** που επηρεάζουν την αναγνώριση της επικίνδυνης οδηγικής συμπεριφοράς. Με αυτόν τον τρόπο θα ενισχυθεί η προσπάθεια της επιστημονικής κοινότητας και της αυτοκινητοβιομηχανίας για την βελτίωση των προηγμένων συστημάτων υποστήριξης οδηγού.
- Η ανάπτυξη ενός κατάλληλου **συστήματος αναγνώρισης** του επιπέδου 'Ζώνης Ανοχής Ασφαλείας' που βρίσκεται ο οδηγός σε πραγματικό χρόνο εντός του οχήματος. Επομένως, ο οδηγός θα έχει την ικανότητα κατά την διάρκεια της διαδρομής να παρακολουθεί την οδηγική του συμπεριφορά και να αντιδρά σε περίπτωση μη ασφαλούς οδήγησης.
- Την ανάπτυξη μίας **εφαρμογής για έξυπνα κινητά τηλέφωνα**, η οποία θα καταγράφει τα δεδομένα οδήγησης και θα προβλέπει την συνολική διάρκεια στην οποία ο οδηγός θα βρίσκεται σε κάθε επίπεδο της 'Ζώνης Ανοχής Ασφαλείας'. Με αυτό τον τρόπο ο οδηγός θα μπορεί να έχει μια πλήρη εικόνα της οδηγικής του συμπεριφοράς και να προβεί στις απαραίτητες παρεμβάσεις για την βελτίωση της.
- Εξοπλισμός των οχημάτων με συστήματα που αξιοποιώντας τον αλγόριθμο θα μπορέσουν να προβλέπουν την ύπαρξη ατυχήματος και θα προειδοποιούν τον οδηγό μέσω προειδοποιητικών μηνυμάτων, έτσι ώστε να εκτελέσει κατάλληλες ενέργειες για την αποφυγή του.

## 6.4 Προτάσεις για περαιτέρω έρευνα

Οι σύγχρονες μέθοδοι επεξεργασίας και ανάλυσης αξιοποιούνται όλο και περισσότερο στον τομέα της οδικής ασφάλειας. Η ανάλυση της οδηγικής συμπεριφοράς με την αξιοποίηση μεθόδων μηχανικής και βαθιάς εκμάθησης, αποτελεί αντικείμενο υψηλού ενδιαφέροντος για τους ερευνητές. Στην μελέτη που αναλύθηκε προέκυψαν ορισμένα ζητήματα. Επομένως, οι ερευνητές συνέστησαν την εξέταση επιπλέον παραγόντων και μεθόδων.

Η παρούσα Διπλωματική Εργασία είχε ως στόχο να καλύψει το κενό που προέκυψε από την βιβλιογραφική ανασκόπηση εξετάζοντας διάφορες τεχνικές μηχανικής και βαθιάς μάθησης και να αποτελέσει βάση για ακόμα περισσότερες συγκριτικές αναλύσεις. Ωστόσο κατά την ανάπτυξη της μεθοδολογίας και την αξιολόγηση των αποτελεσμάτων εντοπίστηκαν ορισμένες ελλείψεις, οι οποίες θα μπορούσαν να ληφθούν υπόψη σε μελλοντικές μελέτες. Παρακάτω παρατίθενται **προτάσεις για περαιτέρω έρευνα**, οι οποίες θα μπορούσαν συμβάλουν στην εμπάθυνση, στην καλύτερη κατανόηση και κατ' επέκταση στην αντιμετώπιση των ελλείψεων που εντοπίστηκαν.

- Αξιοποίηση του **μεγαλύτερου όγκου δεδομένων** με στόχο την βελτίωση της προγνωστικής ικανότητας των μοντέλων ταξινόμησης. Όσο αυξάνεται ο αριθμός των δεδομένων, παράλληλα μειώνεται η πιθανότητα σφάλματος του μοντέλου.
- Ανάπτυξη **εναλλακτικών τεχνικών εξέτασης σημαντικότητας** των μεταβλητών (feature importance). Η περαιτέρω διερεύνηση της σημαντικότητας των μεταβλητών μπορεί να προσδιορίσει με μεγαλύτερη ακρίβεια την σχέση των μεταβλητών με την ικανότητα αναγνώρισης του επιπέδου ασφαλείας που βρίσκεται κάθε οδηγός.
- Διερεύνηση της **επιρροής πρόσθετων παραγόντων**. Με βάση την παρούσα μελέτη αλλά και τις έρευνες που αναζητήθηκαν κατά την βιβλιογραφική ανασκόπηση, οι παράγοντες που θα μπορούσαν μελλοντικά να εξεταστούν αφορούν τις καιρικές συνθήκες, τα στοιχεία της οδού, τα χαρακτηριστικά και τις αντιλήψεις (σχετικά με την επικινδυνότητα κατά την οδήγηση) των οδηγών, τη γεωμετρία των οδών, τους κανόνες οδικής κυκλοφορίας, και τους πολιτισμικούς παράγοντες που διαφέρουν μεταξύ των χωρών και επηρεάζουν εξίσου την οδηγική συμπεριφορά.
- Ανάπτυξη κατάλληλων μοντέλων για ταξινόμηση ακολουθίας (**sequence classification**) του επιπέδου 'Ζώνης Ανοχής Ασφαλείας' που θα βρίσκεται ο οδηγός στο επόμενο χρονικό πλαίσιο των 30 δευτερολέπτων.

## BIBΛΙΟΓΡΑΦΙΑ

1. Nettleton, D., 2014. Chapter 6 - Selection of Variables and Factor Derivation, in: Nettleton, D. « Selection of Variables and Factor Derivation, in: Nettleton, D. (Ed.) » , D. (Ed.), *Commercial Data Mining. Morgan Kaufmann, Boston, pp. 79–104.* <https://doi.org/10.1016/B978-0-12-416602-8.00006-6>.
2. Shi, X., Wong, Y.D., Li, M.Z.F., Chai, C., 2018. «Key risk indicators for accident assessment conditioned on pre-crash vehicle trajectory. *Accident Analysis & Prevention* 117, 346–356» <https://doi.org/10.1016/j.aap.2018.05.007>.
3. Alex Goldstein, Adam Kapelner, Justin Bleich & Emil Pitkin, 2015. «Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation.» *p.p. 44-65.* <https://doi.org/10.1080/10618600.2014.907095>.
4. Aljanahi, A.A.M., Rhodes, A.H., Metcalfe, A.V. « Speed, speed limits and road traffic accidents under free flow conditions.» *Accident Analysis & Prevention* 31. 161–168. 1999. [https://doi.org/10.1016/S0001-4575\(98\)00058-X](https://doi.org/10.1016/S0001-4575(98)00058-X).
5. Chawla, N. v., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. «SMOTE: Synthetic Minority Over-sampling Technique.» <https://doi.org/10.1613/jair.953>.
6. Christopher K. Wikle, Abhirup Datta, Bhava Vyasa Hari, Edward L. Boone, Indranil Sahoo, Indulekha Kavila, Stefano Castruccio, Susan J. Simmons, Wesley S. Burr, Won Chang, 2022. «An illustration of model agnostic explainability methods applied to environmental data.» <https://doi.org/10.1002/env.2772>.
7. Dimitrios Rimpas, Andreas Papadakis, Maria Samarakou, 2019. «OBD-II sensor diagnostics for monitoring vehicle operation and consumption.» *Science Direct.* <https://doi.org/10.1016/j.egy.2019.10.018>.
8. Directorate-General for Mobility and Transport, 2023. *Road safety in the EU: fatalities below pre-pandemic levels but progress remains too slow.* *European Commission (europa.eu).* URL. [https://transport.ec.europa.eu/news-events/news/road-safety-eu-fatalities-below-pre-pandemic-levels-progress-remains-too-slow-2023-02-21\\_en](https://transport.ec.europa.eu/news-events/news/road-safety-eu-fatalities-below-pre-pandemic-levels-progress-remains-too-slow-2023-02-21_en).
9. Eva Michelaraki, Christos Katrakazas, Tom Brijs, George Yannis, 2021. «Modelling the Safety Tolerance Zone: Recommendations from the i-DREAMS project.» 10th International Congress on Transportation Research. <https://www.nrso.ntua.gr/geyannis/wp-content/uploads/geyannis-pc398.pdf>.
10. Eva Michelaraki, Christos Katrakazas, Tom Brijs, George Yannis. «Modelling the Safety Tolerance Zone: Recommendations from.» 2021. <https://www.nrso.ntua.gr/geyannis/wp-content/uploads/geyannis-pc398.pdf>.
11. Eva Michelaraki, Marianthi Kallidoni, Christos Katrakazas, Tom Brijs, George Yannis. 2023. «How to define a safety tolerance zone for speed? Insights from the i-DREAMS project.» *Science Direct. Transportation Research Procedia*, p.p. 415-422. <https://doi.org/10.1016/j.trpro.2023.11.422>.



12. Garefalakis, T, Christos Katrazakas, Yannis, G. «Data-Driven Estimation of a Driving Safety Tolerance Zone Using Imbalanced Machine Learning.» *MDPI*. 2022. <https://www.mdpi.com/1424-8220/22/14/5309>.
13. Guibin Dong, Xiucheng Li , Jingxiao Zhao , Shuai Su , R.D.K. Misra , Ruoxiu Xiao , Chengjia Shang, 2020. «Machine learning guided methods in building chemical composition-hardenability model for wear-resistant steel.» *Science Direct*. <https://doi.org/10.1016/j.mtcomm.2020.101332>.
14. Gunapriya Balan, Singaravelan Arumugam, Suresh Muthusamy, Hitesh Panchal, Hossam Kotb, Mohit Bajaj, Sherif S. M. Ghoneim, Kitmo, 2022. «An Improved Deep Learning-Based Technique for Driver Detection and Driver Assistance in Electric Vehicles with Better Performance.» *Hindawi*. <https://www.hindawi.com/journals/itees/2022/8548172/>.
15. Hall, M.A., 2000. «Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning.» , in: ICML.
16. Hao Liu, Tao Wang, Wenyong Li, Xiaofei Ye, Quan Yuan, 2024. «Lane-change intention recognition considering oncoming traffic: Novel insights revealed by advances in deep learning.» *Science Direct* . <https://doi.org/10.1016/j.aap.2024.107476>.
17. Hochreiter, Sepp, και 1997. Jürgen Schmidhuber. «Long Short-Term Memory.» *IEEE*. p.p 1735 - 1780. <https://ieeexplore.ieee.org/abstract/document/6795963>.
18. Hossin, M., Sulaiman, M.N. «A review on evaluation metrics for data classification evaluations .» *International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.2, March 2015* . [https://d1wqtxts1xzle7.cloudfront.net/37219940/5215ijdkp01-libre.pdf?1428316763=&response-content-disposition=inline%3B+filename%3DA\\_REVIEW\\_ON\\_EVALUATION\\_METRICS\\_FOR\\_DATA.pdf&Expires=1709467450&Signature=AsIBS~STR3YP~qWDhMhH856b18R4axW5-N8B5VRv5Q21Y4d89TQ](https://d1wqtxts1xzle7.cloudfront.net/37219940/5215ijdkp01-libre.pdf?1428316763=&response-content-disposition=inline%3B+filename%3DA_REVIEW_ON_EVALUATION_METRICS_FOR_DATA.pdf&Expires=1709467450&Signature=AsIBS~STR3YP~qWDhMhH856b18R4axW5-N8B5VRv5Q21Y4d89TQ).
19. Ian Goodfellow, Yoshua Bengio, Aaron Courville. *Deep Learning*. London, England: The MIT press, Cambridge, 2016.
20. Jair Ferreira Júnior, Eduardo Carvalho, Bruno V. Ferreira, Cleidson de Souza, Yoshihiko Suhara, Alex Pentland, Gustavo Pessin. «Driver behavior profiling: An investigation with different smartphone sensors and machine learning.» 2017. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0174959>.
21. Jason Brownlee, 2020. «How to Use the Keras Functional API for Deep Learning.» <https://machinelearningmastery.com/keras-functional-api-deep-learning/>.
22. Jaswanth K.S., D. Stalin David, 2020. «A Novel Based 3D Facial Expression Detection Using Recurrent Neural Network.» *IEEE*. <https://ieeexplore.ieee.org/abstract/document/9262287>.

23. Junshi Xu, Marc Saleh, Marianne Hatzopoulou, 2020. «A machine learning approach capturing the effects of driving behaviour and driver characteristics on trip-level emissions.» *Science Direct*. <https://doi.org/10.1016/j.atmosenv.2020.117311>.
24. Katrakazas, C., Antoniou, C., Yannis, G., 2020. «Identification of driving simulator sessions of depressed drivers: A comparison between aggregated and time-series classification.» *Transportation Research Part F: Traffic Psychology and Behaviour* 75, 16–25. <https://doi.org/10.1016/j.trf.2020.09.015>.
25. Kui Yang, Christelle Al Haddad, George Yannis, Constantinos Antoniou, 2022. «Classification and Evaluation of Driving Behavior Safety Levels: A Driving Simulation Study.» *IEEE*. 111-125. <https://ieeexplore.ieee.org/abstract/document/9705507>.
26. Llundunfeld H., Alexander G J. «User's guide to positive guidance. 3rd edition.» *UC Berkeley Transportation Library*. 1990. <https://trid.trb.org/view/489067>.
27. MathWorks:. «Long Short-Term Memory Neural Networks, 2023. URL. .» χ.χ. [https://au.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html#responsive\\_offcanvas](https://au.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html#responsive_offcanvas).
28. Md. Al Mehedi Hasan, Mohammed Nasser, Shamim Ahmad, Khademul Islam Molla. *Feature Selection for Intrusion Detection Using Random Forest*. 2016. <https://www.scirp.org/journal/paperinformation?paperid=65359>.
29. Michelaraki, E., Katrakazas, C., Brijs, T., Yannis, G., 2021a. «Modelling the Safety Tolerance Zone: Recommendations from the i-DREAMS project, in: 10th International Congress on .» *Transportation Research. Rhodes Island, Greece*.
30. Michelaraki, E., Katrakazas, C., Yannis, G., Konstantina Frantzola, E., Kalokathi, F., Kaiser, S., Brijs, K., Brijs, T. «A Review of Real-Time Safety Intervention Technologies.» *in; 7th Humanist Conference. Rhodes Island, Greece*. 2021b.
31. NTUA Road Safety Observatory [WWW Document], 2022. URL. <https://www.nrso.ntua.gr/> (accessed 2.1.22).
32. Osisanwo F.Y., Akinsola J.E.T., Awodele O., Hinmikaiye J. O., Olakanmi O., Akinjobi J., 2017. «Supervised Machine Learning Algorithms: Classification and Comparison.» *International Journal of Computer Trends and Technology (IJCTT) – Volume 48 Number 3 June 2017*. [https://www.researchgate.net/profile/J-E-T-Akinsola/publication/318338750\\_Supervised\\_Machine\\_Learning\\_Algorithms\\_Classification\\_and\\_Comparison/links/596481dd0f7e9b819497e265/Supervised-Machine-Learning-Algorithms-Classification-and-Comparison.pdf](https://www.researchgate.net/profile/J-E-T-Akinsola/publication/318338750_Supervised_Machine_Learning_Algorithms_Classification_and_Comparison/links/596481dd0f7e9b819497e265/Supervised-Machine-Learning-Algorithms-Classification-and-Comparison.pdf).
33. Peng Ping, Wenhui Qin, Yang Xu, Chiyomi Miyajima, Kazuya Takeda, 2019. «Impact of Driver Behavior on Fuel Consumption: Classification, Evaluation and Prediction Using Machine Learning.» *IEEE*. <https://ieeexplore.ieee.org/abstract/document/8727915>.

34. Peppes, N., Alexakis, T., Adamopoulou, E., Demestichas, K. «Driving Behaviour Analysis Using Machine and Deep Learning Methods for Continuous Streams of Vehicular Data.» *Sensors* 21. . 2021. <https://doi.org/10.3390/s21144704>.
35. Peppes, N., Alexakis, T., Adamopoulou, E., Demestichas, K., 2021. « Driving Behaviour Analysis Using Machine and Deep Learning Methods for Continuous Streams of Vehicular Data. *Sensors* 21.» <https://doi.org/10.3390/s21144704>.
36. PRASHANT, BANERJEE. «Explain your model predictions with LIME.» *Kaggle*. <https://www.kaggle.com/code/prashant111/explain-your-model-predictions-with-lime>.
37. Qingyan Yang, Ju Fan, Jianyong Wang, Lizhu Zhou, 2010. «Personalizing Web Page Recommendation via Collaborative Filtering and Topic-Aware Markov Model.» *IEEE*. <https://ieeexplore.ieee.org/abstract/document/5694099>.
38. Roshandel, S., Zheng, Z., Washington, S., 2015. «Impact of real-time traffic characteristics on freeway crash occurrence: Systematic review and meta-analysis. *Accident Analysis & Prevention* 79, 198–211.» <https://doi.org/10.1016/j.aap.2015.03.013>.
39. Roussou, S., T. Garefalakis, E. Michelaraki, T. Brijs, και G, 2024. Yannis. «Machine Learning Insights on Driving Behaviour Dynamics among Germany, Belgium, and UK Drivers.» *MDPI*. <https://www.mdpi.com/2071-1050/16/2/518>.
40. Seong Kyung Kwon, Ji Hwan Seo, Jun Young Yun, Kyoung-Dae Kim , 2021. «Driving Behavior Classification and Sharing System Using CNN-LSTM Approaches and V2X Communication.» *MDPI*. <https://doi.org/10.3390/app112110420>.
41. Shangguan, Q., Fu, T., Wang, J., Luo, T., Fang, S., 2021. «An integrated methodology for real-time driving risk status prediction using naturalistic driving data. *Accident Analysis & Prevention* 156, 106122.» <https://doi.org/10.1016/j.aap.2021.106122>.
42. Shi, X., Wong, Y.D., Li, M.Z.-F., Palanisamy, C., Chai, C., 2019. «A feature learning approach based on XGBoost for driving assessment and risk prediction. *Accident Analysis & Prevention* 129, 170–179.» <https://doi.org/10.1016/j.aap.2019.05.005>.
43. Shinde, Pramila P., και 2019. Seema Shah. «A Review of Machine Learning and Deep Learning Applications.» *IEEE*. <https://ieeexplore.ieee.org/abstract/document/8697857>.
44. Song, X., Yin, Y., Cao, H., Zhao, S., Li, M., Yi, B., 2021. «The mediating effect of driver characteristics on risky driving behaviors moderated by gender, and the classification model of driver's driving risk. *Accident Analysis & Prevention* 153, 106038. .» <https://doi.org/10.1016/j.aap.2021.106038>.
45. Spyridon Alexakos, 2021. «Assessment of operator competence using machine learning.» *Academic press*. p.p. 23-30. <https://polynoe.lib.uniwa.gr/xmlui/handle/11400/1557>.

46. Visual Studio Code, 2023. «Jupyter Notebooks in VS Code.URL.»  
<https://code.visualstudio.com/docs/datascience/jupyter-notebooks>.
47. Vujovic, Zeljko. «Classification Model Evaluation Metrics.International Journal of Advanced Computer Science and Applications Volume 12(Issue 6):599-606.»  
*Research Gate*. 2021.  
[https://www.researchgate.net/publication/352902406\\_Classification\\_Model\\_Evaluation\\_Metrics](https://www.researchgate.net/publication/352902406_Classification_Model_Evaluation_Metrics).
48. Wang, Jiaqi, Ma, Y., Yang, X., Li, T., Wei, H., 2021. «Short-Term Traffic Prediction considering Spatial-Temporal Characteristics of Freeway Flow. Journal of Advanced Transportation 2021.» <https://doi.org/10.1155/2021/5815280>.
49. World Health Organization, 2018. «Global Status Report On Road Safety 2018 [WWW Document].» *World Health Organization*.  
 URL.<https://www.who.int/publications/i/item/9789241565684> (accessed 2.2.22).
50. World Health Organization, 2021. «Global Plan: Decade of Action for Road Safety 2021- 2030 [WWW Document].» *World Health Organization*. URL.  
<https://www.who.int/teams/social-determinants-of-health/safety-and-mobility/decade-ofaction-for-road-safety-2021-2030> (accessed 2.2.22).
51. Xu, C., Tarko, A.P., Wang, W., Liu, P., 2013. «Predicting crash likelihood and severity on freeways with real-time loop detector data. Accident Analysis & Prevention 57, 30–39. .» <https://doi.org/10.1016/j.aap.2013.03.035>.
52. Shengxue Zhu, Chongyi Li, Kexin Fang, Yichuan Peng, Yuming Jiang, Yajie Zou, 2022. «An Optimized Algorithm for Dangerous Driving Behavior Identification Based on Unbalanced Data.» *MDPI*. <https://doi.org/10.3390/electronics11101557>.
53. Yi Ding, Hongyang Zhu, ,Ruyun Chen ,Ronghui Li 1. «An Efficient AdaBoost Algorithm with the Multiple Thresholds Classification.» *MDPI*. 2022 .  
<https://www.mdpi.com/2076-3417/12/12/5872>.
54. Yun Qian, Yanchun Liang , Mu Li , Guoxiang Feng , Xiaohu Shi, 2014. «A resampling ensemble algorithm for classification of imbalance problems.» *Science Direct* .  
<https://doi.org/10.1016/j.neucom.2014.06.021>.
55. Yuru Zhang, Chun Su , Jiajun Wu , Hao Liu , Mingjiang Xie, 2024. «Trend-augmented and temporal-featured Transformer network with multi-sensor signals for remaining useful life prediction.» *Science Direct*. <https://doi.org/10.1016/j.ress.2023.109662>.
56. Zhe Li, Boqing Gong, Tianbao Yang, 2016. «Improved Dropout for Shallow and Deep Learning.»  
[https://proceedings.neurips.cc/paper\\_files/paper/2016/file/7bb060764a818184ebb1cc0d43d382aa-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2016/file/7bb060764a818184ebb1cc0d43d382aa-Paper.pdf).
57. Zoumana Keita, 2022. «Classification in Machine Learning: An Introduction.» *Datacamp*. <https://www.datacamp.com/blog/classification-machine-learning>.

58. Βύτανος Λέων, 2020. «Gesture Recognition System using Machine Learning.» *Academic press*, p.p 14-15. <https://ikee.lib.auth.gr/record/319053>.
59. Γιαννουλόπουλος Ιωάννης, 2020. «Διερεύνηση της επίδρασης προηγμένων συστημάτων υποστήριξης οδηγού στη βελτίωση της οδηγικής του συμπεριφοράς.» *Academic press*. p.p. 39-40. <https://apothesis.eap.gr/archive/item/158122>.
60. Νικόλαος, Γαλάνης. «Επεξηγηματική βαθιά μάθηση σε εφαρμογές ταξινόμησης κειμένου.» 2023. <http://hdl.handle.net/11610/25928>.
61. Σκοτίδας Αναστάσιος, 2013. «Παρακολούθηση δεδομένων οχήματος με χρήση Arduino ADK και Android smartphome.» *Academic press*. <https://dspace.uowm.gr/xmlui/handle/123456789/2631>.