**NATIONAL TECHNICAL UNIVERSITY OF ATHENS**
**School of Civil Engineering**
**Dept. of Transportation Planning and Engineering**

**Doctoral Dissertation**

Panagiotis Fafoutellis

# Traffic Theory Inspired Spatiotemporal and Multimodal Deep Learning Models for Trustworthy Short-Term Urban Traffic Forecasting

Supervisor: Eleni I. Vlahogianni, Professor NTUA

Co-Supervisor: Ioannis Golias, Professor Emeritus NTUA

Co-Supervisor: George Yannis, Professor NTUA

Athens, February 2024

**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ**
**Σχολή Πολιτικών Μηχανικών**
**Τομέας Μεταφορών και Συγκοινωνιακής Υποδομής**

**Διδακτορική Διατριβή**

Παναγιώτης Φαφουτέλλης, Πολιτικός Μηχανικός ΕΜΠ &
Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών ΕΜΠ

# Αξιόπιστη Βραχυπρόθεσμη Πρόβλεψη Κυκλοφορίας σε Αστικά Δίκτυα με Χωροχρονικά και Πολυτροπικά Μοντέλα Βαθιάς Μάθησης Εμπνευσμένα από τη Θεωρία της Κυκλοφοριακής Ροής

Επιβλέπουσα: Ελένη Ι. Βλαχογιάννη, Καθηγήτρια ΕΜΠ

Συνεπιβλέπων: Ιωάννης Γκόλιας, Ομότιμος Καθηγητής ΕΜΠ

Συνεπιβλέπων: Γιώργος Γιαννής, Καθηγητής ΕΜΠ

Αθήνα, Φεβρουάριος 2024

## Contact Details

For any inquiries about the contents of this Ph.D. dissertation, please contact the author using any of the following contact details:


## Panagiotis Fafoutellis

National Technical University of Athens (NTUA), Department of Transportation Planning and Engineering, 5 Heroon Polytechniou Str., Zografou Campus, GR-15773, Athens, Greece

Email: panfaf@mail.ntua.gr

Personal Email: pfafoutellis@gmail.com

Phone: 210 7721289

*Αφιερώνεται με πολλή αγάπη στην Έλενα,*

*στην κόρη μας, την Γεωργία, για να το δει όταν μεγαλώσει*

*και στο σκυλάκι μας, την Lucy, που συνέβαλε με τον δικό της τρόπο*

# ΣΥΝΟΨΗ

Στην παρούσα διδακτορική διατριβή προτείνεται ένα πλήρες μεθοδολογικό πλαίσιο και μια συνοδευτική εργαλειοθήκη μεθόδων και εννοιών για την βραχυπρόθεσμη πρόβλεψη της κυκλοφορίας σε αστικά οδικά δίκτυα, που στόχο έχουν να αυξήσουν την ακρίβεια, την αξιοποιστία και τη εφαρμοσιμότητα των μοντέλων Βαθιάς Μάθησης στην διαχείριση κυκλοφορίας βασιζόμενα σε χαρακτηριστικά όπως η πολυτροπικότητα και η αιτιότητα και σε στοιχεία της θεωρίας της Κυκλοφοριακής Ροής.

Συγκεκριμένα αναπτύσσεται ένα πολυτροπικό πλαίσιο πρόβλεψης βασισμένο στην έννοια των Πολυεπίπεδων Δικτύων (Multiplex Networks) για την πρόβλεψη σε αστικά οδικά δικτύα λαμβάνοντας υπόψη χωροχρονικές σχέσεις μεταξύ των δύο εξεταζόμενων μέσων (οδική κυκλοφορία και ζήτηση του μετρό). Προς την κατεύθυνση της διεύρυνσης της πρόβλεψης σε επίπεδο δικτύου, αναπτύσσονται νευρωνικά δίκτυα πολλαπλών εξόδων, βασισμένα στις αρχές της Μάθησης Πολλαπλών Διεργασιών, τα οποία είναι σε θέση να παρέχουν προβλέψεις χρόνων διαδρομής για πολλαπλές διαδρομές σε όλο το δίκτυο, χρησιμοποιώντας ένα ενιαίο μοντέλο και δεδομένα από πολλαπές πηγές κυκλοφοριακών δεδομένων. Ακόμα, για να ενισχυθεί η αξιοπιστία της διαδικασίας πρόβλεψης, εφαρμόζεται η προσαρμογή Βαθιάς Μάθησης της αιτιότητας Granger, το μοντέλο Νευρωνικού Granger, για την ανακάλυψη αιτιωδών σχέσεων μεταξύ των θέσεων του οδικού δικτύου. Επιπλέον, ενσωματώνονται πτυχές της θεωρίας της κυκλοφοριακής ροής σε ένα αιτιώδες και πολλαπλών διεργασιών μοντέλο πρόβλεψης με βάση τη Φυσική (Physics-Informed Neural Network), χρησιμοποιώντας μια καινοτόμο συνάρτηση απωλειών εμπνευσμένη από τη θεωρία της Κυκλοφοριακής Ροής, η οποία, εκτός από το σφάλμα πρόβλεψης, λαμβάνει υπόψη την απόσταση των προβλέψεων από το αντίστοιχο θεμελιώδες διάγραμμα. Το παραπάνω πλάισιο εφαρμόζεται σε δεδομένα τροχιών τμημάτων από την πόλη Xi'an της Κίνας και σε κυκλοφοριακά δεδομένα και δεδομένα ζήτησης μέσων μαζικής μεταφοράς από την ευρύτερη περιοχή των Αθηνών.

Τα αποτελέσματα αποκαλύπτουν σημαντικά κυκλοφοριακά μοτίβα που περιγράφουν τη μηχανική του οδικού δικτύου και έχουν τη δυνατότητα να αυξήσουν την προβλεψιμότητα των κυκλοφοριακών συνθηκών, όπως υποδεικνύεται από τα αντίστοιχα πειράματα. Ακόμα, η ανίσχευση σχέσεων αιτιότητας και η ενσωμάτωση βασικών γνώσεων της Κυκλοφορίακής Ροής στην συνάρτηση απωλειών αυξάνει σημαντικά τόσο την ακρίβεια και την αξιοπιστία των βραχυπρόθεσμων προβλέψεων, όσο και την ανθεκτικότητα του μοντέλου σε θορυβώδη δεδομένα.

Η μελλοντική έρευνα θα επικεντρωθεί στην αξιολόγηση της δυνατότητας μεταφοράς των παραπάνω δομών σε άλλες τοποθεσίες του ίδιου οδικού δικτύου και σε άλλα οδικά δίκτυα, καθώς και στη χρήση πιο σύνθετων δομών και τεχνικών Βαθιάς Μάθησης, προκειμένου να βελτιωθεί περαιτέρω η απόδοση πρόβλεψης. Τέλος, θα επιχειρηθεί ο συνδυασμός όλων των παραπάνω δομών σε ένα θεωρητικά ενημερωμένο, αιτιώδες, πολυτροπικό και σε επίπεδο δικτύου πλαίσιο πρόβλεψης.

**Λέξεις-κλειδιά:** πρόβλεψη κυκλοφορίας, εφαρμοσιμότητα μοντέλου, πολυτροπική πρόβλεψη, μάθηση πολλαπλών διεργασιών, αιτιότητα κατά Granger, νευρωνικά δίκτυα με βάση τη φυσική

# ABSTRACT

In this doctoral dissertation, a complete methodological framework and an associative toolkit of methods and concepts are proposed for the problem of short term urban traffic forecasting that aim to increase the accuracy, trustworthiness and actionability of deep learning models for traffic management based on features such as multimodality and causality and on aspects of Traffic Flow Theory.

Specifically, a multimodal forecasting framework based on the concept of Multiplex Networks is developed for urban road networks, taking into account spatio-temporal relationships between the two considered modes (road traffic and metro demand). Towards the extension to network-level forecasting, multiplex neural networks based on the principles of Multitask Learning are being developed, which are able to provide travel time forecasts for multiple routes across the network, using a single model and data from multiple traffic data sources. Moreover, to enhance the trustworthiness of the prediction process, the Granger Causality Deep Learning adaptation, Neural Granger, is applied to detect causal relationships between road network locations. In addition, aspects of traffic flow theory are incorporated into a causal and multitask Physics-Informed Neural Network forecasting model, using an innovative loss function derived from Traffic Flow Theory, which, in addition to the prediction error, takes into account the distance of the forecasts from the corresponding fundamental diagram. The above framework is applied to segment trajectory data from the city of Xi'an, China, and to traffic and public transport demand data from the greater area of Athens.

The results reveal important traffic patterns that describe the mechanics of the road network and have the potential to increase the predictability of traffic conditions, as indicated by the corresponding experiments. Furthermore, the detection of causal relations and the incorporation of basic knowledge of Traffic Flow into the loss function significantly increases both the accuracy and trustworthiness of the short-term predictions and the robustness of the model to noisy data.

Future research will focus on evaluating the feasibility of transferring these structures to other locations on the same road network and other road networks, as well as using more complex structures and Deep Learning techniques to further improve the prediction performance. Finally, an attempt will be made to combine all the above structures into a theoretically informed, causal, multimodal and network-level prediction framework.

**Keywords:** traffic forecasting, model actionability, multimodal forecasting, multitask learning, Granger causality, physics-informed neural networks

# TABLE OF CONTENTS

# TABLE OF FIGURES AND TABLES

# ΕΚΤΕΝΗΣ ΠΕΡΙΛΗΨΗ

## Εισαγωγή

Η άνοδος των δεικτών ιδιοκτησίας οχημάτων και, γενικότερα, της ζήτησης για μεταφορές τις τελευταίες δεκαετίες, ως αποτέλεσμα της αύξησης του πληθυσμού και της αστικοποίησης, έχει εντείνει σε σημαντικό βαθμό την πίεση που ασκείται στα αστικά συστήματα μεταφορών. Ως αποτέλεσμα των παραπάνω, η συχνότητα εμφάνισης, καθώς και η ένταση, φαινομένων κυκλοφοριακής συμφόρησης έχει αυξηθεί, με σημαντικές άμεσες και έμμεσες συνέπειες στην υγεία των πολιτών, την οικονομία, την οδική ασφάλεια και το περιβάλλον. Σε αυτό το πλαίσιο, η άμβλυνση των φαινομένων συμφόρησης έχει καταστεί μία απαιτητική και πολύπλοκη αποστολή. Τα τελευταία 20 χρόνια, η ανάπτυξη έξυπνων συστημάτων μεταφορών (Intelligent Transportation Systems – ITS) έχει αναδειχθεί ως το πιο αποτελεσματικό εργαλείο προς αυτή την κατεύθυνση. Η δε πρόοδος της Τεχνητής Νοημοσύνης, των υπολογιστικών συστημάτων και των τηλεπικοινωνιών προσφέρει νέες προοπτικές στην αποδοτική διαχείριση της κυκλοφορίας.

Η ακριβής βραχυπρόθεσμη πρόβλεψη της κυκλοφορίας θεωρείται καθοριστική για την αποτελεσματική εφαρμογή των ITS, δεδομένου ότι επιτρέπει την έγκαιρη αντίδραση στις αναμενόμενες συνθήκες κυκλοφορίας, όπως η εφαρμογή κατάλληλων στρατηγικών αποτροπής της εμφάνισης κυκλοφοριακής συμφόρησης. Η αυξημένη ζήτηση για ακριβείς προβλέψεις σε συστήματα διαχείρισης κυκλοφορίας έχει αυξήσει το ενδιαφέρον στην ερευνητική περιοχή της πρόβλεψης των κυκλοφοριακών συνθηκών. Μάλιστα, η άνευ προηγουμένου διαθεσιμότητα κυκλοφοριακών δεδομένων, ως αποτέλεσμα της ραγδαίας ανάπτυξης της τεχνολογίας και των τηλεπικοινωνιών στις μέρες μας, έστρεψε την προσοχή των ερευνητών προς μεθόδους οδηγούμενες από τα δεδομένα και, κυρίως, την Βαθιά Μηχανική Μάθηση (Deep Learning).Τα μοντέλα πρόβλεψης που ανήκουν σε αυτή την κατηγορία θεωρούνται ως αυτά με την μεγαλύτερη ακρίβεια καθώς μπορούν να προσομοιώσουν με ακρίβεια οποιαδήποτε μαθηματική σχέση, ανεξαρτήτως της πολυπλοκότητάς της, όπως οι χωροχρονικές συσχετίσεις μεταξύ των κυκλοφοριακών συνθηκών των θέσεων ενός οδικού δικτύου και μάλιστα χωρίς να απαιτείται ιδιαίτερη προσπάθεια από τον χρήστη.

Από την άλλη μεριά, τα συγκεκριμένα μοντέλα απαιτούν μεγάλο όγκο δεδομένων και αυξημένη υπολογιστική ισχύ για να εκπαιδευτούν αποτελεσματικά, ενώ είναι δυσχερής η κατανόηση του μηχανισμού των προβλέψεων και η ερμηνεία των αποτελεσμάτων τους, γεγονός που μειώνει την αξιοπιστία και τη διαφάνειά τους. Οι παραπάνω λόγοι περιορίζουν σε μεγάλο βαθμό τη χρησιμοποίηση αυτών των μοντέλων για τους σκοπούς της διαχείρισης κυκλοφορίας σε πραγματικές συνθήκες, η οποία παραμένει αντιστρόφως ανάλογη του μεγάλου ερευνητικού ενδιαφέροντος και του όγκου των δημοσιεύσεων στην συγκεκριμένη ερευνητική περιοχή. Προκύπτει συνεπώς η ανάγκη για επαναπροσδιορμό των κριτηρίων αξιολόγησης των μοντέλων πρόβλεψης, βάζοντας σε κεντρικό ρόλο όχι μόνο την ακρίβεια των προβλέψων αλλά και την αξιοπιστία, την αιτιότητα, την επεξηγησιμότητα και την αποδοτικότητα, ο συνδυασμός των οποίων καθορίζει την εφαρμοσιμότητα ενός μοντέλου.

*Στόχοι*

Ο κύριος στόχος της παρούσας διατριβής είναι η ανάπτυξη ενός εφαρμόσιμου πλαισίου πρόβλεψης της κυκλοφορίας πολλαπλών χρονικών και χωρικών κλιμάκων για την προγνωστική διαχείριση της κυκλοφορίας, το οποίο εκπαιδεύεται σε πολυτροπικά περιβάλλοντα (λαμβάνοντας υπόψη περισσότερα του ενός μέσα μετακίνησης), αξιοποιώντας αιτιώδεις χωροχρονικές σχέσεις και πτυχές της θεωρίας της Κυκλοφοριακής Ροής.

Για τον σκοπό αυτό, προτείνεται μια εργαλειοθήκη λειτουργικών μονάδων, καθεμία από τις οποίες μπορεί να αξιοποιηθεί από τις αρχές διαχείρισης της κυκλοφορίας υπό διαφορετικές συνθήκες για την πραγματοποίηση πολλαπλών εργασιών πρόβλεψης από διαφορετικές οπτικές γωνίες, π.χ. πρόβλεψη πολλαπλών μεταβλητών εξόδου, πολυτροπικών ή μεμονωμένων εργασιών, πρόβλεψη σε όλο το δίκτυο ή σημειακά και βραχυπρόθεσμες ή μακροπρόθεσμες προβλέψεις (με χρήση δεδομένων υψηλής ή χαμηλής ανάλυσης, αντίστοιχα).

Προς την κατεύθυνση αυτή, η διατριβη θέτει τους παρακάτω επιμέρους στόχους:

1. Προσδιορισμός χωρικών συσχετίσεων στα οδικά δίκτυα, με τη χρήση σύγχρονων μεθόδων από την εφαρμοσμένη στατιστική και τη θεωρία της πληροφορίας, και αξιολόγηση της επίδρασής τους στην ακρίβεια της πρόβλεψης.
2. Διερεύνηση πιθανών σημαντικών αλληλεπιδράσεων μεταξύ των συνθηκών οδικής κυκλοφορίας και της ζήτησης για άλλους τρόπους μετακίνησης και αξιποίησή τους για την παραγωγή πολυτροπικών προβλέψεων.
3. Αξιοποίηση μιας εύστοχης και ουσιαστικής αναπαράστασης του οδικού δικτύου για την αποτελεσματική μοντελοποίηση, τόσο των χωρικών, όσο και των χρονικών σχέσεων του συστήματος μεταφορών (ζήτηση για διαφορετικά μέσα).
4. Ανίχνευση σημαντικών αιτιωδών σχέσεων μεταξύ των κυκλοφοριακών συνθηκών σε διάφορες θέσεις του οδικού δικτύου και των κυκλοφοριακών μοτίβων που αναδεικνύουν σε επίπεδο πόλης, τόσο βραχυπρόθεσμα όσο και μακροπρόθεσμα.
5. Διερεύνηση της επίδρασης των αιτιωδών σχέσεων στην ενίσχυση της προβλεψιμότητας των κυκλοφοριακών συνθηκών σε επίπεδο δικτύου, καθώς και την αξιοπιστία των προβλέψεων.
6. Εισαγωγή στοιχείων από τη θεωρία της Κυκλοφοριακής Ροής στο στάδιο εκπαίδευσης του μοντέλου με την ανάπτυξη μιας συνάρτησης απωλειών συμβατής με τη φύση της κυκλοφορίακής ροής και της εξέλιξής της σε χώρο και χρόνο. Διερεύνηση της επίδρασής της στην απόδοση και την αξιοπιστία του πλαισίου πρόβλεψης.
7. Αξιολόγηση της επίδρασης της προσέγγισης που καθοδηγείται από τη θεωρία όσον αφορά στη δυνατότητα αξιοποίησης του πλαισίου σε πραγματικές συνθήκες (απόδοση, αποδοτικότητα και αξιοπιστία).

## Προκλήσεις για την ανάπτυξη εφαρμόσιμων μοντέλων πρόβλεψης της κυκλοφορίας

Σε αυτή την ενότητα, παρουσιάζονται οι σημαντικότερες προκλήσεις που σχετίζονται με τη δυνατότητα αξιοποίησης μοντέλων Βαθιάς Μάθησης στην προγνωστική διαχείριση της κυκλοφορίας, σύμφωνα με την πρόσφατη βιβλιογραφία, καθώς και τα ερευνητικά ερωτήματα στα οποία επιχειρεί να απαντήσει η παρούσα διατριβή.

*Πρόκληση 1: Περιορισμοί της Βαθιάς Μάθησης*

Παρά την αδιαμφισβήτητη αποτελεσματικότητα της Βαθιάς Μάθησης στην ακριβή πρόβλεψη των μελλοντικών συνθηκών, υπάρχουν ορισμένα εμπόδια για την εφαρμογή των σχετικών μοντέλων σε πραγματικές συνθήκες:

- Απαίτηση μεγάλου όγκου δεδομένων, με επαρκή γεωγραφική κάλυψη του οδικού δικτύου, καθώς και αντιπροσωπευτικότητας των δεδομένων αυτών.
- Χρονοβόρα διαδικασία εκπαίδευσης καθώς και βαθμονόμησης της δομής και των υπερπαραμέτρων των μοντέλων.
- Απαιτείται υψηλή υπολογιστική ισχύς για την εκπαίδευση και συντήρηση του μοντέλου, καθώς και για την αποθήκευση και χρήση των αντίστοιχων βάσεων δεδομένων.
- Χαμηλή δυνατότητα ερμηνείας των αποτελεσμάτων, η οποία είναι απαραίτητη προκειμένου να δικαιολογηθεί η διαδικασία λήψης αποφάσεων και να αυξηθεί η αξιοπιστία του συγκεκριμένου μοντέλου.
- Περιορισμένη δυνατότητα μεταφοράς και γενίκευσης των μοντέλων σε άλλα δίκτυα ή σε άλλα σημεία του ίδιου δικτύου.

*Πρόκληση 2: Αναπαράσταση του οδικού δικτύου και των δεδομένων εισόδου*

Η εύστοχη αναπαράσταση του οδικού δικτύου και των χωροχρονικών συσχετίσεων μεταξύ των περιοχών του δύναται να ενισχύσει την απόδοση του μοντέλου, καθώς και την ευχέρεια ερμηνείας του. Επίσης, σε αντίθεση με την πιο συνηθισμένη πρακτική της βιβλιογραφίας, όπου το μοντέλο τροφοδοτείται με όλες τις διαθέσιμες ακατέργαστες πληροφορίες χωρίς καμία προηγούμενη ανάλυση ή επιλογή των πιο σχετικών χαρακτηριστικών, η ακριβής αναπαράσταση των συσχετίσεων μεταξύ των δεδομένων εισόδου μειώνει την πολυπλοκότητα και τη διαστατικότητα του χώρου εισόδου του μοντέλου. Οι πιο διαδεδομένες αναπαραστάσεις είναι οι εξής:

- Στοίβα διανυσμάτων (Stacked Vector): Οι χρονοσειρές των κυκλοφοριακών μετρήσεων από τις διάφορες περιοχές του δικτύου (που μπορούν να θεωρηθούν ως διανύσματα) στοιβάζονται σε ένα διάνυσμα διανυσμάτων. Αυτή είναι η πιο απλή μέθοδος αναπαράστασης, ωστόσο είναι κατάλληλη, κατά κύριο λόγο, για απλά οδικά δίκτυα με λίγες θέσεις μέτρησης.
- Εικόνα ή πλέγμα (Image or Grid): Ορίζεται ένα τετραγωνικό πλέγμα με το μέγεθος του οδικού δικτύου και σε κάθε εικονοστοιχείο (πίξελ) του πλέγματος αποδίδεται μια τιμή που αντιπροσωπεύει τις συνθήκες κυκλοφορίας στο εσωτερικό του, το οποίο είναι ακριβώς ανάλογη με μια εικόνα σε κλίμακα του γκρι. Το μειονέκτημα αυτής της μεθόδου είναι ότι, συνήθως, το οδικό δίκτυο καλύπτει ένα μικρό μέρος του πλέγματος, αφήνοντας κενά τα περισσότερα εικονοστοιχεία. Συνεπώς, το μοντέλο τροφοδοτείται με έναν σχετικά μεγάλο χώρο εισόδου, ο οποίος όμως περιέχει περιορισμένη πληροφορία.
- Γράφος: Οι γράφοι μπορούν να χρησιμοποιηθούν για να εκφράσουν πιο σύνθετες σχέσεις μεταξύ των δεδομένων εισόδου από διαφορετικές θέσεις του οδικού δικτύου, οι οποίες απορρέουν από τη συνδεσιμότητα των τμημάτων του οδικού δικτύου, την επίδραση των διασταυρώσεων και των φωτεινών σηματοδοτών και τα μοτίβα κυκλοφορίας/συμφόρησης μεταξύ μη-γειτονικών θέσεων. Αυτού του είδους οι συσχετίσεις εκφράζονται μέσω της συνδεσιμότητας των κόμβων του γράφου και ενός συντελεστή βαρύτητας που αντιστοιχεί σε κάθε σύνδεση. Υπάρχουν τρεις βασικοί τύποι συντελεστών βαρύτητας: αυτοί που εκφράζουν την πραγματική (γεωγραφική) συνδεσιμότητα των κόμβων, αυτοί που είναι

ανάλογοι της απόστασης των κόμβων και αυτοί που εκφράζουν την ύπαρξη στατιστικής συσχέτισης μεταξύ των κυκλοφοριακών συνθηκών των αντίστοιχων κόμβων.

*Πρόκληση 3: Συσχέτιση και Αιτιότητα*

Στον επιστημονικό κλάδο της πρόβλεψης της κυκλοφορίας, παρά το γεγονός ότι οι περισσότεροι ερευνητές συμφωνούν σχετικά με την ύπαρξη σημαντικών αιτιωδών σχέσεων μεταξύ διαφορετικών θέσεων ενός οδικού δικτύου, συνήθως βασίζονται σε μοντέλα βαθιάς μάθησης για εντοπισμό συσχετίσεων μεταξύ του μεγάλου όγκου των δεδομένων εισόδου. Ωστόσο, οι συγκεκριμένες συσχετίσεις δεν είναι συνήθως διαφανείς και κατανοητές για τους χρήστες, ενώ δεν είναι εγγυημένο ότι είναι αιτιώδεις. Κατά τη μοντελοποίηση οποιουδήποτε φαινομένου, οι αιτιώδεις σχέσεις μεταξύ των μεταβλητών εισόδου και εξόδου παίζουν πολύ σημαντικό ρόλο. Σε αντίθεση με τη στατιστική συσχέτιση που μπορεί να παρατηρηθεί σε ένα συγκεκριμένο σύνολο δεδομένων, η έννοια της αιτιότητας αναφέρεται σε πιο σημαντικές σχέσεις μεταξύ των μεταβλητών, με την έννοια ότι, π.χ., η τιμή μιας ή περισσότερων μεταβλητών εισόδου θα "προκαλέσει" την έξοδο να λάβει μια συγκεκριμένη τιμή. Πιο συγκεκριμένα, η μελέτη των αιτιωδών σχέσεων είναι πολύ σημαντική για τρεις βασικούς λόγους:

- Χρήση μεταβλητών εισόδου που σχετίζονται αιτιωδώς με το αναμενόμενο αποτέλεσμα και των οποίων ο αντίκτυπος σε αυτό παραμένει σταθερός και είναι ανεξάρτητος από τις μεταβολές των τιμών των μεταβλητών.
- Συνήθως, ο χώρος εισόδου περιλαμβάνει τόσο αιτιώδεις όσο και μη αιτιώδεις μεταβλητές και, κατά τη χρήση τεχνικών επεξήγησης των αποτελεσμάτων, δεν μπορούν να διαχωριστούν, γεγονός που οδηγεί σε αναξιόπιστες επεξηγήσεις.
- Όταν χρησιμοποιούνται τόσο αιτιώδη όσο και μη αιτιώδη χαρακτηριστικά στην πρόβλεψη, η διαστατικότητα του χώρου εισόδου αυξάνεται σημαντικά, υπονομεύοντας την απόδοση του μοντέλου, ενώ η χρήση μόνο αιτιωδών χαρακτηριστικών θα ήταν επαρκής.

Όταν πρόκειται για δεδομένα χρονοσειρών, η πιο κατάλληλη και δημοφιλής μέθοδος για την ανίχνευση αιτιωδών σχέσεων είναι ο έλεγχος αιτιότητας κατά Granger. Με απλά λόγια, λέμε ότι μια χρονοσειρά "x" προκαλεί κατά Granger τη χρονοσειρά "y", όταν η "y" προβλέπεται με μεγαλύτερη ακρίβεια από ένα μοντέλο, εάν προηγούμενες τιμές της "x" περιλαμβάνονται στο χώρο εισόδου. Στην κλασική αιτιότητα κατά Granger, το μοντέλο που χρησιμοποιείται για την αξιολόγηση της ύπαρξης αιτιωδών σχέσεων είναι το Διανυσματικό Αυτοπαλίνδρομο μοντέλο (Vector Autoregressive - VAR). Γίνεται δηλαδή η παραδοχή ότι υπάρχουν μόνο γραμμικές σχέσεις μεταξύ των δεδομένων, κάτι που μπορεί να είναι υπεραπλουστευμένο για συστήματα του πραγματικού κόσμου, π.χ. για δεδομένα κυκλοφορίας

Το σημαντικότερο πλεονέκτημα της αιτιότητας Granger, ωστόσο, είναι ότι πρόκειται για μια πολυμεταβλητή μέθοδο, δηλαδή αξιολογεί την ύπαρξη αιτιώδους σχέσης μεταξύ δύο μεταβλητών λαμβάνοντας υπόψη και την επίδραση όλων των άλλων μεταβλητών εισόδου, σε αντίθεση με τις περισσότερες άλλες μεθόδους που εξετάζουν μόνο τις σχέσεις ανά ζεύγη. Σύμφωνα με το πρότυπο VAR, η μεταβλητή εξόδου $y_t$ είναι ένας γραμμικός συνδυασμός των προηγούμενων χρονικών βημάτων των μεταβλητών εισόδου $x_{i,t-l}$:

$$y_t = \sum_{l=1}^{T} a_{1,l} x_{1,t-l} + \sum_{l=1}^{T} a_{2,l} x_{2,t-l} + \dots + \sum_{l=1}^{T} a_{n,l} x_{n,t-l} \qquad (I)$$

όπου T είναι ο μέγιστος αριθμός χρονικών υστερήσεων που λαμβάνονται υπόψη, n είναι ο συνολικός αριθμός των χρονοσειρών εισόδου και $a_{i,l}$ είναι ο συντελεστής της i-οστής χρονοσειράς που σχετίζεται με το χρονικό βήμα t-l. Οι τιμές των συντελεστών προσδιορίζονται έτσι ώστε να ελαχιστοποιείται η ακόλουθη ποσότητα:

$$\frac{1}{N}(\sum^{N} y_t - \sum_{i=1}^{n}\sum_{l=1}^{T} a_{i,l}x_{i,t-l}) + \lambda \sum_{i=1}^{n}\sum_{l=1}^{T}|a_{i,l}| \qquad (II)$$

όπου N είναι ο συνολικός αριθμός των παρατηρήσεων.

*Πρόκληση 4: Επεξηγησιμότητα και χωροχρονική ανάλυση*

Η ερμηνευσιμότητα ή επεξηγησιμότητα αναφέρονται στη διαφάνεια ενός μοντέλου, η οποία συνεπάγεται ότι τα δεδομένα ή ο αλγόριθμος και ο μηχανισμός που παρέχει τις προβλέψεις είναι προσβάσιμα σε κάποιο βαθμό από τον χρήστη, χωρίς να απαιτείται πλήρης κατανόηση της δομής του μοντέλου και του μαθηματικού του υποβάθρου από την πλευρά του.

Η επεξηγησιμότητα και ο εντοπισμός των χωρικών και χρονικών σχέσεων μεταξύ των θέσεων του οδικού δικτύου, με σαφήνεια και διαφάνεια, είναι πολύ σημαντικές για τους σκοπούς της διαχείρισης της κυκλοφορίας, προκειμένου:

- Να δικαιολογείται η διαδικασία λήψης αποφάσεων και να αυξάνεται η εμπιστοσύνη στο συγκεκριμένο μοντέλο, πράγμα απαραίτητο για τη συμμόρφωση των χρηστών του δικτύου.
- Να εξαχθεί νέα επιστημονική γνώση σχετικά με τη μηχανική του δικτύου.
- Να βελτιωθεί η απόδοση του μοντέλου πρόβλεψης και η μεταφερσιμότητά του.

Η εξαγωγή των χωροχρονικών σχέσεων και η ερμηνεία των αποτελεσμάτων των μοντέλων Βαθιάς Μάθησης γίνεται με μεθόδους οι οποίες δεν έχουν ισχυρό μαθηματικό υπόβαθρο και εξαρτώνται εξ ολοκλήρου από τα διαθέσιμα δεδομένα - κατά συνέπεια, είναι πολύ ευάλωτες σε θορυβώδη σύνολα δεδομένων και ενδέχεται να παρέχουν αναξιόπιστα αποτελέσματα. Επιπλέον, υποδηλώνουν μόνο στατιστικές και όχι αιτιώδεις σχέσεις. Για την αντιμετώπιση αυτού του ζητήματος, οι ερευνητές θα πρέπει να λάβουν υπόψη τους αφενός τη γνώση που προέρχεται από τη θεωρία της Κυκλοφοριακής Ροής σχετικά με τη χωροχρονική διάδοση και των δυναμικών της κυκλοφορίας και της συμφόρησης και αφετέρου μεθόδους για την εξέταση και την ποσοτικοποίηση των αιτιωδών σχέσεων στα διαθέσιμα δεδομένα.

*Πρόκληση 5: Μάθηση πολλαπλών διεργασιών: Η ανάγκη για πολυμεταβλητές προβλέψεις*

Η έννοια των μοντέλων πολλαπλών διεργασιών ή πολυμεταβλητών προβλέψεων αναφέρεται στην ανάπτυξη ενός μοντέλου που μπορεί να προβλέψει περισσότερα του ενός μεγέθη (διεργασίες) ταυτόχρονα, δηλαδή έχει πολλές μεταβλητές εξόδου. Στον τομέα της πρόβλεψης της κυκλοφορίας, ένα μοντέλο πολλαπλών διεργασιών μπορεί να χρησιμοποιηθεί για την πρόβλεψη περισσότερων του ενός κυκλοφοριακών μεγεθών (φόρτος, ταχύτητα, πυκνότητα ή κατάληψη). Το βασικό πλεονέκτημα αυτής της μεθόδου σε σχέση με τη χρήση μοντέλων για την πρόβλεψη μίας μεταβλητής είναι ότι, σύμφωνα με τις θεμελιώδεις σχέσεις της Κυκλοφοριακής Τεχνικής, χρειάζονται τουλάχιστον δύο μεταβλητές για τον χαρακτηρισμό των κυκλοφοριακών συνθηκών ως συμφορημένες ή μη. Για παράδειγμα, μία τιμή του φόρτου αντιστοιχεί σε δύο πιθανές καταστάσεις. Παρέχοντας στο μοντέλο την τιμή μιας επιπλέον μεταβλητής, αυτό δύναται να διαχωρίζει τις δύο καταστάσεις και τις διαφορετικές δυναμικές που αντιστοιχούν στην καθεμιά.

Μία διαφορετική προσέγγιση πολυμεταβλητών προβλέψεων στην πρόβλεψη της κυκλοφορίας είναι οι πολυτροπικές προβλέψεις, δηλαδή η πρόβλεψη της κυκλοφορίας στο οδικό δίκτυο σε συνδυασμό με τη ζήτηση για άλλα μέσα μεταφοράς, π.χ. Μέσα Μαζικής Μεταφοράς. Ένα σχετικό παράδειγμα παρουσιάζεται στο Κεφάλαιο 5 της παρούσας διατριβής. Πρέπει να τονιστεί ότι, ιδιαίτερα σε αυτή την περίπτωση, εκτός από τον εκτεταμένο χώρο εξόδου, αυξάνεται σημαντικά και η πολυπλοκότητα του χώρου εισόδου. Συνεπώς, αυτή η προσέγγιση πρέπει να επιλέγεται όταν υπάρχει σημαντική συσχέτιση των μεταβλητών εξόδου και αναμένεται να αυξηθεί η συνολική ακρίβεια πρόβλεψης για όλες.

*Πρόκληση 6: Ενισχυμένη αξιοπιστία με Νευρωνικά Δίκτυα Βασισμένα Στη Φυσική*

Η πρόσφατη βιβλιογραφία στην πρόβλεψη της κυκλοφορίας βασίζεται κυρίως στην Βαθιά Μάθηση, παραβλέποντας την επιστημονική γνώση που προέρχεται από τη θεωρία της Κυκλοφοριακής Τεχνικής. Το γεγονός αυτό μειώνει περεταίρω την αξιοπιστία των μοντέλων και, σε συνδυασμό με τους περιορισμούς που αναφέρθηκαν νωρίτερα, μειώνει και τις προοπτικές αξιοποίησής τους σε πραγματικές συνθήκες. Από την άλλη μεριά, τα αναλυτικά (ή στατιστικά) κυκλοφοριακά μοντέλα οδηγούν σε ανακριβείς εκτιμήσεις για τέσσερις κύριους λόγους: μπορούν να αποτυπώσουν μόνο ένα περιορισμένο υποσύνολο των πιθανών συνθηκών κυκλοφορίας και της δυναμικής τους, αναφέρονται σε ιδανικές συνθήκες, περιλαμβάνουν αρκετές παραμέτρους που πρέπει να καθοριστούν από τον χρήστη και είναι ευάλωτα σε θορυβώδη δεδομένα. Ωστόσο, αυτού του είδους τα μοντέλα είναι ερμηνεύσιμα και δεν απαιτούν μεγάλες ποσότητες δεδομένων.

Ο συνδυασμός των παραπάνω προσεγγίσεων υιοθετείται από τα Νευρωνικά Δίκτυα Βασισμένα στη Φυσική (Physics-Informed Neural Networks) που είναι πολύ αποδοτικά όσον αφορά τις απαιτήσεις δεδομένων, επιτυγχάνοντας παρόμοιες ή και καλύτερες επιδόσεις από τα κλασικά μοντέλα Βαθιά Μάθησης. Μάλιστα η εισαγωγή στοιχείων από τη θεωρία αναμένεται να μειώσει την πολυπλοκότητα του μοντέλου, να βελτιώσει την απόδοση και την αξιοπιστία και, κυρίως, να οδηγήσει σε έγκυρες προβλέψεις και ορθολογική λήψη αποφάσεων. Η έννοια της Ενημερωμένης από τη Θεωρία Βαθιάς Μάθησης στοχεύει στο να συνδυάσει τα πλεονεκτήματα των δύο παραπάνω προσεγγίσεων μοντελοποίησης. Συγκεκριμένα, εφαρμόζεται όταν ένα μοντέλο οδηγούμενο από τα δεδομένα περιλαμβάνει, ως είσοδο ή έξοδο, μία ή περισσότερες μεταβλητές για τις οποίες είναι γνωστή μια αναλυτική/μαθηματική σχέση από το αντίστοιχο επιστημονικό πεδίο. Η σχέση αυτή συνήθως ενσωματώνεται στη συνάρτηση απωλειών του μοντέλου ή χρησιμοποιείται σε μια ανεξάρτητη οντότητα και αποσκοπεί στην προσαρμογή ιδιαίτερα λανθασμένων, παράλογων προβλέψεων του μοντέλου, προς τις "θεωρητικά" αναμενόμενες.

*Πρόκληση 7: Προβλέψεις σε επίπεδο δικτύου*

Η ανάπτυξη ενός πλαισίου μοντελοποίησης που θα παρέχει προβλέψεις για ένα ολόκληρο οδικό δίκτυο ταυτόχρονα, παραμένει ανεπαρκώς διερευνημένη, αν και μπορεί να αυξήσει σημαντικά τη δυνατότητα αξιοποίησης οποιουδήποτε μοντέλου. Στην πρόσφατη βιβλιογραφία, τα περισσότερα μοντέλα παρέχουν έξοδο για μία μόνο θέση και όχι για όλες ταυτόχρονα, που συνήθως αναφέρεται ως θέση-στόχος, και εκμεταλλεύονται μόνο ένα σχετικά μικρό τμήμα του οδικού δικτύου ως είσοδο. Οι λόγοι που συμβαίνει αυτό περιλαμβάνουν την ανεπαρκή διαθεσιμότητα και ποιότητα δεδομένων, τις υψηλές υπολογιστικές απαιτήσεις και την πολυπλοκότητα ενός τέτοιου μοντέλου.

Για την αντιμετώπιση των παραπάνω περιορισμών, οι ερευνητές στρέφονται κυρίως σε αλγορίθμους και μεθόδους της Υπολογιστικής Επιστήμης που εμφανίστηκαν σχετικά πρόσφατα. Η πιο σημαντική από αυτές είναι η Υπολογιστική Παρυφών (Edge Computing), στην οποία οι υπολογιστικές εργασίες εκτελούνται στις παρυφές του δικτύου, δηλαδή εκεί που συλλέγονται τα δεδομένα, και όχι σε ένα κεντρικό υπολογιστικό νέφος ή κέντρο δεδομένων. Για παράδειγμα, η Υπολογιστική Παρυφών μπορεί να χρησιμοποιηθεί για την επεξεργασία μεγάλων ποσοτήτων δεδομένων κυκλοφορίας που παράγονται από αισθητήρες, κάμερες, συσκευές GPS και άλλες πηγές τοπικά, προκειμένου να παρέχονται προβλέψεις των συνθηκών κυκλοφορίας σχεδόν σε πραγματικό χρόνο.

Βασισμένο στις αρχές της Υπολογιστικής Παρυφών, η Ομοσπονδιακή Μάθηση (Federated Learning), είναι ένα πρότυπο μηχανικής μάθησης που επιτρέπει σε πολλαπλούς συμμετέχοντες, όπως συσκευές ή κόμβους ακμών, να μαθαίνουν συνεργατικά ένα μοντέλο χωρίς να μοιράζονται τα ακατέργαστα δεδομένα τους με μια κεντρική οντότητα. Αντ' αυτού, ανταλλάσσουν μεταξύ τους ενημερώσεις του μοντέλου με αποκεντρωμένο τρόπο, επιτρέποντας τη βελτίωση του συνολικού μοντέλου μέσω των συλλογικών συνεισφορών όλων των συμμετεχόντων. Αυτή η προσέγγιση έχει τα πλεονεκτήματα της δυνατότητας κλιμάκωσης, δηλαδή το φορτίο του υπολογιστικού έργου κατανέμεται στις συμμετέχουσες συσκευές, επιτρέποντας την κλιμάκωση σε προβλήματα που θα ήταν ανέφικτο να επιλυθούν με μια κεντρική προσέγγιση, και της ανθεκτικότητας, καθώς το σύστημα θα είναι σε θέση να συνεχίσει να λειτουργεί, ακόμη και αν κάποιοι συμμετέχοντες παρουσιάσουν οποιουδήποτε είδους αστοχία.

Προς το παρόν, η αξιοποίηση των παραπάνω μεθόδων στην πρόβλεψη της κυκλοφορίας, καθώς και γενικότερα της πρόβλεψης σε επίπεδο δικτύου, βρίσκεται σε πολύ πρώιμο στάδιο.

*Πρόκληση 8: Αποδοτικότητα και επεκτασιμότητα σε πολυτροπικά περιβάλλοντα*

Η πρόβλεψη της ζήτησης των μέσων μαζικής μεταφοράς, καθώς και άλλων τρόπων μεταφοράς (π.χ. ταξί, υπηρεσίες μετακίνησης, ποδήλατα κ.λπ.), αποτελούν παραλλαγές του προβλήματος της πρόβλεψης της κυκλοφορίας, που επίσης μπορούν να αντιμετωπιστούν με παρόμοιο τρόπο. Μια ολοκληρωμένη προσέγγιση, που θα λαμβάνει υπόψη όλα τα μέσα μεταφοράς, θα έφερνε επανάσταση στη διαχείριση της κυκλοφορίας και στη λήψη αποφάσεων σε επίπεδο πόλης και θα παρείχε στις αρχές ένα εργαλείο που θα επέτρεπε τη βελτιστοποίηση των συνθηκών κυκλοφορίας σε ολόκληρο το οδικό δίκτυο. Τα παραπάνω είναι δυνατά με την επέκταση της έννοιας της πρόβλεψης πολλαπλών διεργασιών στην πρόβλεψη πολυτροπικής κυκλοφορίας.

Εάν αυξηθεί η διαθεσιμότητα δεδομένων πολλαπλών πηγών, αναμένεται να αυξηθεί και η δημοτικότητα της πολυτροπικής πρόβλεψης κυκλοφορίας. Προκειμένου να αναπτυχθεί ένα μοντέλο ικανοποιητικά χαμηλής πολυπλοκότητας που θα είναι αποδοτικό, παρά το γεγονός ότι έχει έναν τόσο μεγάλο χώρο εισόδου, είναι απαραίτητη μια ακριβής αναπαράσταση των χώρων εισόδου και των μεταξύ τους σχέσεων.

*Ερευνητικά Ερωτήματα*

Σύμφωνα με τα ευρήματα της βιβλιογραφικής ανασκόπησης και τις προκλήσεις που εντοπίστηκαν, και όσον αφορά στους στόχους της παρούσας διατριβής, μπορούν να διατυπωθούν τα ακόλουθα ερευνητικά ερωτήματα:

Ε1: Μπορούν να εντοπιστούν σημαντικά χωροχρονικά μοτίβα κυκλοφορίας σε ένα οδικό δίκτυο; Πόσο ευαίσθητα είναι αυτά στη διαδικασία εξαγωγής; Ποιος είναι ο αντίκτυπος αυτών των μοτίβων στη βραχυπρόθεσμη δυνατότητα πρόβλεψης της κυκλοφορίας;

Ε2: Πώς μπορούν να επεκταθούν οι τυπικές αναπαραστάσεις του οδικού δικτύου και των χωροχρονικών συσχετίσεων ώστε να συμπεριλάβουν πληροφορίες από πολλαπλούς τρόπους μετακίνησης και ποιος θα είναι ο αντίκτυπος στην προβλεψιμότητα της κυκλοφορίας;

Ε3: Ποια θα ήταν η κατάλληλη διατύπωση του προβλήματος και η δομή μοντελοποίησης πολλαπλών εργασιών ώστε να καταστεί δυνατή η αποτελεσματική και ακριβής πρόβλεψη της κυκλοφορίας σε επίπεδο δικτύου;

Ε4: Μπορεί να καταρτιστεί ένα πλαίσιο αιτιώδους πρόβλεψης που θα οδηγήσει στην ανίχνευση σημαντικών προτύπων κυκλοφορίας, ενισχύοντας παράλληλα την αξιοπιστία και την ακρίβεια των προβλέψεων;

Ε5: Πώς θα μπορούσαν να ενσωματωθούν πτυχές της μοντελοποίησης με επίγνωση της θεωρίας στη διαδικασία πρόβλεψης για να βελτιώσουν την αξιοπιστία και τη εφαρμοσιμότητα των μοντέλων πρόβλεψης;

## Εξόρυξη Γραμμικών και Μη Γραμμικών Χωροχρονικών Χαρακτηριστικών σε Αστικά Δίκτυα

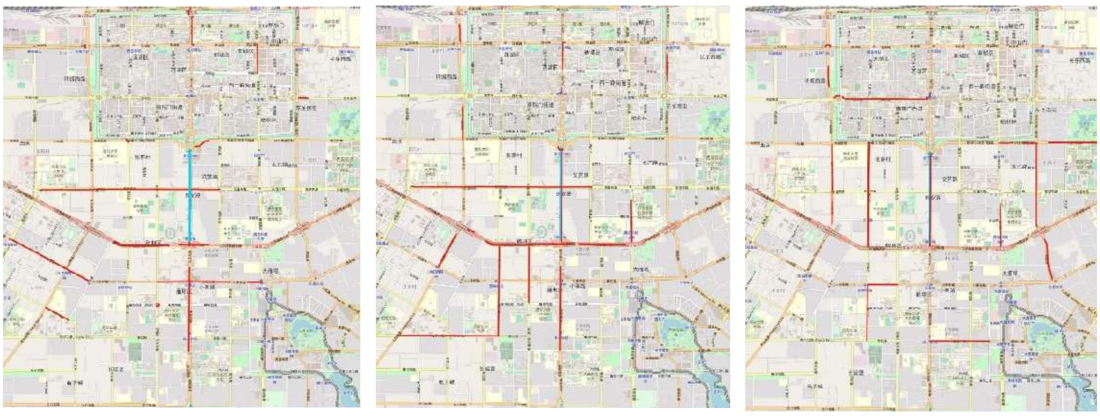
Σύμφωνα με όσα έχουν ήδη συζητηθεί, ένα προφανές πρώτο βήμα προς την ανάπτυξη ενός εφαρμόσιμου μοντέλου πρόβλεψης είναι να εξεταστεί κατά πόσον υπάρχουν χωρικές εξαρτήσεις μεταξύ των κυκλοφοριακών συνθηκών στις διάφορες θέσεις ενός οδικού δικτύου και ποιες είναι οι επιπτώσεις τους στη βραχυπρόθεσμη πρόβλεψη της κυκλοφορίας. Για το σκοπό αυτό, εφαρμόζονται τρεις μέθοδοι για τον εντοπισμό τους: Συσχέτιση κατά Pearson, Αμοιβαία Πληροφορία (Mutual Information) και Δυναμική Χρονική Παραμόρφωση (Dynamic Time Warping). Η προτεινόμενη μεθοδολογική προσέγγιση εφαρμόζεται στο οδικό δίκτυο της πόλης Xi'an της Κίνας, χρησιμοποιώντας δεδομένα τροχιάς που παρέχονται από την Didi Chuxing Technology Co, μια κινεζική εταιρεία παροχής υπηρεσιών ταξί κινητικότητας, τα οποία αξιοποιήθηκαν για την εκτίμηση της χρονοσειράς της μέσης ταχύτητας για κάθε τμήμα του οδικού δικτύου. Πιο συγκεκριμένα, πραγματοποιείται εντοπισμός των 20 πιο συσχετισμένων θέσεων με τη θέση-στόχο σύμφωνα με κάθε μέθοδο συσχέτισης και συγκρίνονται οι ακρίβειες πρόβλεψης όταν χρησιμοποιούνται αποκλειστικά οι προαναφερθείσες θέσεις ως είσοδος σε ένα μοντέλο κατηγοριοποίησης Μπεϋζιανού δικτύου.

*Προεπεξεργασία δεδομένων*

Τα δεδομένα που χρησιμοποιούνται σε αυτή την ενότητα αποτελούνται από 3,2 εκατομμύρια διαδρομές GPS των οχημάτων της Didi στο οδικό δίκτυο της Xi'an, που πραγματοποιήθηκαν μεταξύ 2 και 30 Νοεμβρίου 2016. Κάθε τροχιά αντιστοιχεί στην ακριβή θέση του οχήματος κάθε 2 έως 4 δευτερόλεπτα, καθώς και την ταχύτητά του, που είναι και η μεταβλητή που χρησιμοποιήθηκε για την πρόβλεψη. Στη συνέχεια, εφαρμόστηκε ένας αλγόριθμος αντιστοίχισης των σημείων στα οδικά τμήματα του δικτύου και, ομαδοποιώντας τα δεδομένα κάθε τμήματος ανά μία ώρα, προέκυψαν οι ωριαίες χρονοσειρές μέσης ταχύτητας των τμημάτων.

*Εντοπισμός χωρικών συσχετίσεων*

Χρησιμοποιώντας τις τρεις προαναφερθείσες μεθόδους, υπολογίζεται η συσχέτιση όλων των οδικών τμημάτων με το τμήμα-στόχος. Ως τμήμα-στόχος επιλέχθηκε ένα από τα πιο πολυσύχναστα τμήματα που βρίσκεται στο κέντρο της Xi’an. Παρατηρώντας τα αποτελέσματα, φαίνεται ότι οι τρεις προσεγγίσεις αποτυπώνουν διαφορετικά χωρικά μοτίβα στο ίδιο σύνολο δεδομένων. Πιο συγκεκριμένα, σύμφωνα με τους υπολογισμούς της αμοιβαίας πληροφορίας και της συσχέτισης Pearson, τα κοντινά κάθετα οδικά τμήματα συσχετίζονται περισσότερο με το επιλεγμένο, καθώς και ορισμένα ανάντη και κατάντη οδικά τμήματα. Επιπλέον, για την αμοιβαία πληροφορία, εντοπίζονται ορισμένα παράλληλα τμήματα. Αντίθετα, με την DTW, δεν ανιχνεύεται κανένα κάθετο οδικό τμήμα, αλλά αντίθετα κάποια γειτονικά παράλληλα (τα οποία δεν ανιχνεύονται από τις άλλες μεθόδους). Ο αντίκτυπος αυτών των διαφορών θα πρέπει να διερευνηθεί περαιτέρω όσον αφορά την ακρίβεια πρόβλεψης. Το τμήμα-στόχος καθώς και τα 20 πιο συσχετισμένα με αυτό τμήματα, σύμφωνα με κάθε μέθοδο, παρουσιάζονται στην Εικόνα 1.



(a) (b) (c)

**Εικόνα 1. Τα 20 πιο συσχετισμένα οδικά τμήματα (κόκκινο) με το επιλεγμένο τμήμα (μπλε), από την άποψη (α) της Συσχέτισης Pearson, (β) της Αμοιβαίας Πληροφορίας και (γ) της Δυναμικής Χρονικής Παραμόρφωσης.**

*Αποτελέσματα πρόβλεψης και σύγκριση μεθόδων*

Προκειμένου να συγκριθούν τα αποτελέσματα των μεθόδων, αναπτύσσονται μοντέλα πρόβλεψης της ταχύτητας του οδικού τμήματος-στόχου. Τα μοντέλα είναι Μπεϋζιανοί ταξινομητές που προβλέπουν την ταχύτητα του τμήματος σε τρεις ισορροπημένες κλάσεις: <20, 20-26, >26 km/h. Συνολικά αναπτύχθηκαν τρία μοντέλα: τα δύο πρώτα δέχονται σαν είσοδο τις τιμές της ταχύτητας στα 20 πιο συσχετισμένα τμήματα σύμφωνα με τις μεθόδους της Αμοιβαίας Πληροφορίας και της Δυναμικής Χρονικής Παραμόρφωσης, αντίστοιχα, ενώ το τρίτο από όλα τα τμήματα. Η ακρίβεια των τριών μοντέλων παρουσιάζεται στον Πίνακας 1.

**Πίνακας 1. Μετρικές ταξινόμησης των τριών μοντέλων**

| **Metrics** | **Model 1 (Mutual Info)** | **Model 2 (DTW)** | **Model 3 (all locations)** |
|---|---|---|---|
| Accuracy | 89% | 86% | 84% |
| Recall (Sensitivity) | 89% | 86% | 84% |
| Precision | 89% | 86% | 85% |
| F1 - score | 89% | 86% | 85% |

Από τον παραπάνω πίνακα μπορεί κανείς να συμπεράνει ότι η χρήση της αμοιβαίας πληροφορίας είναι η πιο κατάλληλη επιλογή για την περιγραφή των χωρικών συσχετίσεων, τουλάχιστον για την συγκεκριμένη εφαρμογή. Αφετέρου, τα δύο πρώτα μοντέλα έχουν σαφώς καλύτερη απόδοση από το τρίτο. Το αποτέλεσμα αυτό αναδεικνύει τη χρησιμότητα της εκτέλεσης ακόμη και μιας απλής χωροχρονικής ανάλυσης για την επιλογή των πιο κατάλληλων δεδομένων εισόδου. Επιπλέον, η παραπάνω διαδικασία μειώνει τη διαστατικότητα του προβλήματος, η οποία είναι ένα πολύ συνηθισμένο ζήτημα κατά τη χρήση αλγορίθμων μηχανικής μάθησης. Στην παρούσα περίπτωση, ήταν διαθέσιμες 283 μεταβλητές (οδικά τμήματα), που είναι μια υψηλή τιμή, αλλά εν τέλει ήταν απαραίτητα μόνο 20, μειώνοντας και τους απαιτούμενους υπολογιστικούς πόρους, κάτι που είναι εξίσου σημαντικό.

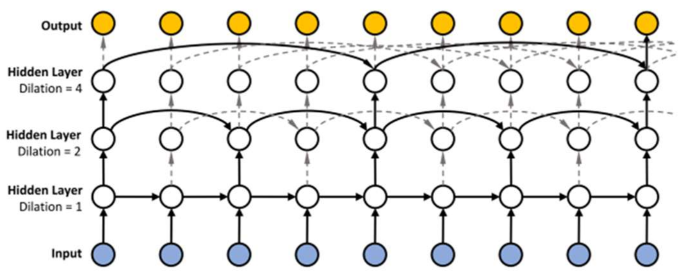
## Πρόβλεψη Κυκλοφορίας με Χρήση Πληροφοριών Πολλαπλής Χρονικής Ανάλυσης με Δεδομένων Τροχιάς Οχημάτων Σε Επίπεδο Δικτύου

Η ανάλυση που παρουσιάστηκε στο προηγούμενο κεφάλαιο κατέδειξε ότι, πράγματι, υπάρχουν σημαντικές χωρικές σχέσεις μεταξύ των κυκλοφοριακών συνθηκών ενός οδικού δικτύου, οι οποίες μπορούν να εντοπιστούν με τη χρήση κατάλληλων μεθόδων από τη στατιστική και τη θεωρία της πληροφορίας. Επιπλέον, προέκυψε ότι η επίδρασή τους στην ακρίβεια της πρόβλεψης είναι αξιοσημείωτη. Στην παρούσα ενότητα, δίνεται έμφαση στις χρονικές σχέσεις, χρησιμοποιώντας μια πιο κατάλληλη δομή μοντελοποίησης, δηλαδή ένα μοντέλο Βαθιάς Μάθησης, του οποίου ο χώρος εισόδου αποτελείται από χρονοσειρές. Επιπλέον, τα δεδομένα συγκεντρώνονται σε υψηλότερη χρονική ανάλυση και, σε αντίθεση με την προηγούμενη ενότητα, παρέχονται βραχυπρόθεσμες προβλέψεις, κάτι που αποτελεί πιο δύσκολο εγχείρημα.

*Μοντελοποίηση σύνθετων χρονικών μοτίβων*

Η πολυπλοκότητα του οδικού περιβάλλοντος, των μηχανισμών ελέγχου, καθώς και το πλήθος των οδηγικών συμπεριφορών και των μη επαναλαμβανόμενων συνθηκών, οδηγούν σε συνεχώς μεταβαλλόμενες χρονικές δυναμικές που καθιστούν εξαιρετικά δύσκολη την μοντελοποίησή τους από μια απλή δομή. Για παράδειγμα, στα Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks), που θεωρούνται τα πιο κατάλληλα και ακριβή μοντέλα στην ανάλυση χρονοσειρών, δεν προβλέπεται η χρήση δεδομένων εισόδου με μεταβλητή χρονική συχνότητα.

Σε αυτή την ενότητα, επιχειρείται η αποτύπωση των βραχυπρόθεσμων και μακροπρόθεσμων επιδράσεων που παρατηρούνται στην κυκλοφορία, η εσωτερίκευση του ζητήματος των χρονικά μεταβαλλόμενων χρονικών εξαρτήσεων που παρατηρούνται στη ροή κυκλοφορίας στη διαδικασία μάθησης και η βελτίωση της ικανότητα πρόβλεψης σε σχέση με τα κλασικά Αναδρομικά Νευρωνικά Δίκτυα, με την εφαρμογή προηγμένων χρονικά μεταβαλλόμενων επαναλαμβανόμενων νευρωνικών δομών. Συγκεκριμένα, η αρχιτεκτονική των Διαστελλόμενων Αναδρομικών Δικτύων (Dilated Neural Networks) είναι μια πολυεπίπεδη αρχιτεκτονική νευρωνικού δικτύου που είναι πολύ απλή στην κατανόησή της, αλλά αποτελεσματική και αποδοτική και βασίζεται στα απλά Αναδρομικά δίκτυα. Χαρακτηρίζεται επίσης από τις διασταλμένες επαναλαμβανόμενες συνδέσεις παράλειψης και τη χρήση εκθετικά αυξανόμενης διαστολής, η οποία απεικονίζεται στην Εικόνα 2.




**Εικόνα 2. Η αρχιτεκτονική των Διαστελλόμενων Αναδρομικών Δικτύων (Dilated Neural Networks)**

*Προετοιμασία δεδομένων*

Τα δεδομένα που χρησιμοποιούνται σε αυτή την ενότητα είναι αυτά των μέσων ταχυτήτων του οδικού δικτύου της Xi'an που χρησιμοποιήθηκε και στην προηγούμενη ενότητα, με τη διαφορά ότι οι χρονοσειρές ομαδοποιούνται ανά πεντάλεπτο. Ο χώρος εισόδου περιλαμβάνει τις χρονοσειρές όλων των οδικών τμημάτων (497 συνολικά), για 12 χρονικά βήματα (μία ώρα) ανά παρατήρηση, οι οποίες χρησιμοποιούνται για την πρόβλεψη της ταχύτητας του τμήματος-στόχου 5 λεπτά μετά.

Αποφασίστηκε ότι οι ταχύτητες του τμήματος-στόχου θα πρέπει να ταξινομηθούν σε δύο κατηγορίες, οπότε, και πάλι, έχουμε ένα πρόβλημα ταξινόμησης. Οι δύο κατηγορίες είναι λιγότερο από 33 km/h και περισσότερο από 33 km/h, που είναι η διάμεσος των ταχυτήτων του τμήματος για τις 29 ημέρες των δεδομένων. Με αυτόν τον τρόπο, υποθέτουμε κακές έως μέτριες συνθήκες κυκλοφορίας όταν η κατηγορία ταχύτητας είναι η πρώτη και μέτριες έως καλές συνθήκες όταν είναι η δεύτερη.

*Αποτελέσματα*

Για την πρόβλεψη των ταχυτήτων χρησιμοποιείται ένα Διαστελλόμενο δίκτυο που βασίζεται συγκεκριμένα στην αρχιτεκτονική του Δικτύου Μακράς Βραχυπρόθεσμης Μνήμης (Long Short-Term Memory – LSTM). Στον Πίνακας 2 παρουσιάζονται οι μετρικές κατηγοριοποίησης για το προαναφερθέν διαστελλόμενο δίκτυο σε σύγκριση με ένα απλό δίκτυο ίδιου τύπου.

Τα αποτελέσματα της σύγκρισης μεταξύ ενός Διαστελλόμενου δικτύου LSTM και ενός κλασικού στην πρόβλεψη μελλοντικών συνθηκών κυκλοφορίας με χρήση δεδομένων από όλο το δίκτυο έδειξαν ότι το διαστελλόμενο μπορεί να βελτιώσει την ακρίβεια πρόβλεψης, χωρίς να εισάγει υπερβολική πολυπλοκότητα στη διαδικασία. Η αρχιτεκτονική του προτεινόμενου δικτύου είναι αρκετά απλή και επίσης εύκολα αντιληπτή. Το κύριο πλεονέκτημα της προτεινόμενης αρχιτεκτονικής σε σύγκριση με το απλό LSTM είναι ότι αποτυπώνει και εκμεταλλεύεται τις μακροπρόθεσμες συσχετίσεις που μπορεί να επηρεάσουν τις μελλοντικές συνθήκες κυκλοφορίας πιο καθοριστικά σε σύγκριση με τις βραχυπρόθεσμες

**Πίνακας 2. Σύγκριση αποτελεσμάτων Διαστελλόμενου και κλασσικού Αναδρομικού δικτύου**

| | **Dilated LSTM** | **Simple LSTM** |
|---|---|---|
| **Precision** | 0.85 | 0.83 |
| **Recall** | 0.85 | 0.82 |
| **F1 score** | 0.85 | 0.82 |
| **Accuracy** | 0.85 | 0.819 |

## Πολυτροπική Πρόβλεψη σε Επίπεδο Δικτύου

Προχωρώντας ένα βήμα παρακάτω από την απλή επιλογή μεταβλητών εισόδου και την μοντελοποίηση των χωροχρονικών συσχετίσεων με βάση κάποια μετρική, σε αυτή την ενότητα προτείνεται η υιοθέτηση του Πολυεπίπεδου Δικτύου (Multiplex Network) από την ερευνητική περιοχή της Ανάλυσης Κοινωνικών Δικτύων για την αναπαράσταση του μεταφορικού δικτύου της Αθήνας και των αντίστοιχων συσχετίσεων. Επίσης, χρησιμοποιείται ένας αλγόριθμος Εντοπισμού Κοινοτήτων στο Πολυεπίπεδο Δίκτυο για τον προσδιορισμό των πιο σχετικών στοιχείων του χώρου εισόδου για τα μοντέλα πρόβλεψης που αναπτύσσονται στη συνέχεια.

Μάλιστα, σε αυτή την εφαρμογή, ακολουθείται μια προσέγγιση πολυτροπικών και πολλαπλών πηγών δεδομένων, καθώς, όχι μόνο προβλέπεται ο μελλοντικός φόρτος κυκλοφορίας και η μελλοντική ζήτηση του μετρό, αλλά αξιοποιούνται επίσης ως χαρακτηριστικά εισόδου για την πρόβλεψη του ενός από το άλλο. Αυτή η προσέγγιση ενός πολυτροπικού πλαισίου, το οποίο εντοπίζει σημαντικές συσχετίσεις μεταξύ των διαφόρων μέσων μεταφοράς (και μπορεί επίσης να επεκταθεί σε περισσότερα από δύο μέσα), θα έχει πολύ σημαντικές προεκτάσεις σε ένα ολιστικό σύστημα διαχείρισης της κυκλοφορίας.

*Πολυεπίπεδα δίκτυα*

Τα Πολυεπίπεδα Δίκτυα είναι πολύπλοκες δομές γράφων που διαθέτουν στρώματα εκτός από κόμβους και ακμές, τα οποία διαθέτουν οι απλοί γράφοι. Ένα πολυεπίπεδο δίκτυο αποτελείται από δύο ή περισσότερους διασυνδεδεμένους απλούς γράφους που βρίσκονται σε διαφορετικά στρώματα. Κάθε στρώμα έχει διαφορετική δομή συνδεσιμότητας εντός του. Αυτό το είδος δικτύου επιτρέπει μια πιο ρεαλιστική προσέγγιση στη μελέτη της αλληλεπίδρασης ατόμων/πρακτόρων, τα οποία μπορούν να επικοινωνούν μέσω διαφορετικών τύπων καναλιών. Η παρούσα εργασία υιοθετεί τον ορισμό ενός πολυεπίπεδου δικτύου ως ένα δίκτυο που αποτελείται από πολλά στρώματα, όπου ένας κόμβος δεν συνδέεται με κόμβους που βρίσκονται σε διαφορετικό στρώμα, αλλά μόνο εντός του ίδιου στρώματος.

*Εντοπισμός κοινοτήτων*

Τα πραγματικά δίκτυα παρουσιάζουν μεγάλες ανομοιογένειες, δηλαδή κορυφές με χαμηλό βαθμό συνδεσιμότητας (λίγες συνδέσεις) συνυπάρχουν με κάποιες κορυφές με υψηλό βαθμό. Έτσι, η κατανομή των ακμών είναι συνολικά και τοπικά ανομοιογενής. Δομικές διαφορές μπορούν να παρατηρηθούν εντός συγκεκριμένων ομάδων κορυφών, όπως υψηλές ή χαμηλές συγκεντρώσεις ακμών μεταξύ αυτών των ομάδων. Αυτό το χαρακτηριστικό των πραγματικών δικτύων ονομάζεται δομή κοινότητας ή ομαδοποίηση. Έτσι, οι κοινότητες, οι οποίες αποκαλούνται επίσης συστάδες ή ενότητες, είναι ομάδες κορυφών που μοιράζονται κοινές ιδιότητες και/ή παίζουν παρόμοιο ρόλο μέσα στον γράφο.

Για τους σκοπούς της παρούσας εργασίας, χρησιμοποιείται ο αλγόριθμος Louvain. Ο αλγόριθμος ανίχνευσης κοινοτήτων Louvain βασίζεται στη βελτιστοποίηση της αρθρωσιμότητας (modularity) κατά την πρόοδο του αλγορίθμου. Η αρθρωσιμότητα είναι μια τιμή κλίμακας μεταξύ -0,5 (μη αρθρωτή ομαδοποίηση) και 1 (πλήρως αρθρωτή ομαδοποίηση) που μετρά τη σχετική πυκνότητα των ακμών εντός κοινοτήτων σε σχέση με τις ακμές εκτός κοινοτήτων και η βελτιστοποίηση αυτής της τιμής οδηγεί στην καλύτερη δυνατή ομαδοποίηση των κόμβων ενός συγκεκριμένου δικτύου.

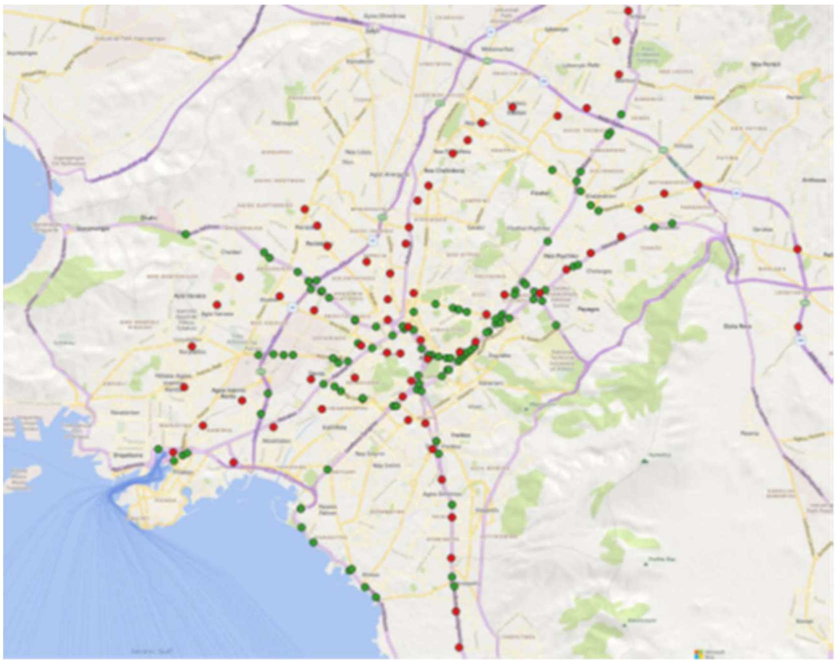
Για Πολυεπίπεδα Δίκτυα, η αρθρωσιμότητα δίνεται από την παρακάτω σχέση:

$$Q_m = \frac{1}{2\mu} \sum_{i,j\,sr} \left[ \left( a_{ijs} - \frac{k_{is} k_{js}}{2m_s} \right) \delta(s,r) + \omega \delta(i,j) \right] \delta(\gamma_{is}, \gamma_{jr}) \qquad (III)$$

όπου i, j είναι οι πράκτορες, s, r είναι τα στρώματα, $a_{ijs}$ είναι 1 αν i, j είναι γειτονικά στο στρώμα s, $k_{js}$ είναι ο βαθμός του πράκτορα i στο στρώμα s, μ είναι ο αριθμός των ζευγών κορυφών που είτε είναι γειτονικές στο στρώμα α είτε αντιστοιχούν στον ίδιο πράκτορα, $m_s$ είναι ο αριθμός των ακμών στο στρώμα s, $\gamma_{is}$ είναι η κοινότητα στην οποία έχει ανατεθεί ο πράκτορας i στο στρώμα s, δ είναι το δέλτα Kronecker και ω είναι ένα βάρος: όταν ο ίδιος πράκτορας ανήκει στην ίδια κοινότητα σε δύο διαφορετικά στρώματα, τότε το $Q_m$ αυξάνεται κατά ω. Το ωμέγα είναι μια παράμετρος που παίρνει τιμές από 0 έως 1. Ο καθορισμός υψηλότερων τιμών του ωμέγα θα έχει ως αποτέλεσμα κοινότητες που θα εκτείνονται σε πολλαπλά στρώματα και θα αποτελούνται από τους ίδιους πράκτορες, αφού με αυτόν τον τρόπο αυξάνεται η τιμή της αρθρωσιμότητας. Από την άλλη πλευρά, αν το ωμέγα οριστεί ίσο με 0, η ύπαρξη των ίδιων πρακτόρων σε διαφορετικά στρώματα στην ίδια κοινότητα δεν συμβάλλει στην αρθρωσιμότητα.

*Διαθέσιμα δεδομένα*

Για την ανάπτυξη του μοντέλου πρόβλεψης της κυκλοφορίας χρησιμοποιήθηκαν δεδομένα κυκλοφοριακής ροής και ζήτησης του μετρό, τα οποία αντιστοιχούν στον αριθμό των διερχόμενων οχημάτων και των επιβατών, αντίστοιχα, από διάφορα σημεία του οδικού δικτύου για περίοδο μιας ώρας. Συγκεκριμένα, χρησιμοποιήθηκαν δεδομένα 10 μηνών (Ιανουάριος έως Οκτώβριος του 2021) από 113 από ανιχνευτές βρόχου που βρίσκονται γύρω από το κέντρο της πόλης της Αθήνας, καθώς και η ζήτηση όλων των 63 σταθμών του μετρό. Στην Εικόνα 3 φαίνεται η ακριβής θέση των ανιχνευτών βρόχου με πράσινο χρώμα και των σταθμών του μετρό με κόκκινο χρώμα.



**Εικόνα 3. Γεωγραφική κατανομή ανιχνευτών (πράσινο) και σταθμών του μετρό (κόκκινο)**

*Το δίκτυο της Αθήνας ως Πολυεπίπεδος Γράφος και εντοπισμός κοινοτήτων*

Το συγκοινωνιακό δίκτυο της Αθήνας (ανιχνευτές και σταθμοί μετρό) μπορεί να αναπαρασταθεί ως ένα πολυεπίπεδο γράφημα των χωροχρονικών σχέσεων των κόμβων, προκειμένου να ανιχνευθούν κοινότητες και να αντληθούν πολύτιμες πληροφορίες για τις χωρικές και χρονικές σχέσεις που το διέπουν. Η κατασκευή του γραφήματος βασίστηκε στην ιδέα ότι κάθε επίπεδο θα αντιστοιχεί σε κάθε ώρα της ημέρας, οπότε το γράφημα αποτελείται από 24 επίπεδα. Κάθε στρώμα αποτελείται από 176 κόμβους, καθένας από τους οποίους αντιπροσωπεύει έναν ανιχνευτή βρόχου ή έναν σταθμό του μετρό. Στη συνέχεια, για να καθοριστούν οι ακμές κάθε στρώματος, σχηματίστηκε η χρονοσειρά της ζήτησης κάθε κόμβου την αντίστοιχη ώρα κάθε ημέρας (π.χ. κάθε μέρα στις 9 π.μ. για το 9ο στρώμα) και στη συνέχεια υπολογίστηκε η αμοιβαία πληροφορία μεταξύ των χρονοσειρών όλων των κόμβων για την ίδια ώρα. Εάν η αμοιβαία πληροφορία μεταξύ δύο κόμβων είναι μεγαλύτερη από 0,5, γεγονός που υποδηλώνει σημαντική συσχέτιση, δημιουργείται μια ακμή μεταξύ τους. Με αυτόν τον τρόπο μοντελοποιήθηκε η χωροχρονική συσχέτιση μεταξύ των κόμβων. Χρησιμοποιήθηκε η παραπάνω τεχνική για να δημιουργηθούν οι ακμές των γραφημάτων, αντί για τις πραγματικές συνδέσεις, επειδή, σε επίπεδο δικτύου, δεν είναι ιδιαίτερα ευχερής ο προσδιορισμός της συνδεσιμότητας των κόμβων, ιδίως μεταξύ ετερογενών κόμβων (σταθμοί του μετρό και ανιχνευτές βρόχων) και, κυρίως, η μέθοδος αυτή οδηγεί σε ξεχωριστή δομή γραφημάτων σε κάθε επίπεδο, η οποία αντικατοπτρίζει τη δυναμική φύση των χωρικών σχέσεων.

Μετά την κατασκευή του πολυεπίπεδου γραφήματος ακολουθεί η ανίχνευση κοινοτήτων χωροχρονικά συσχετισμένων κόμβων. Εφαρμόζοντας τον αλγόριθμο ανίχνευσης κοινοτήτων του Louvain, εντοπίζονται 7 κοινότητες, όπου οι κόμβοι διαφορετικών στρωμάτων αντιστοιχίζονται στην ίδια κοινότητα εάν διαδραματίζουν τον ίδιο ρόλο, ο καθένας στο αντίστοιχο στρώμα και, δεδομένου ότι οι κόμβοι αντιπροσωπεύουν χρονοσειρές και οι συνδέσεις τους αντιστοιχούν σε στατιστικές σχέσεις, οι κόμβοι που ανήκουν στην ίδια κοινότητα έχουν παρόμοια πρότυπα κυκλοφορίας και ζήτησης. Επιπλέον, λόγω της ύπαρξης κόμβων από διαφορετικά στρώματα στις ίδιες κοινότητες, μπορεί κανείς να υποθέσει ότι αναδεικνύονται όχι μόνο χωρικές, αλλά και χρονικές σχέσεις και πιο συγκεκριμένα, υποδεικνύεται η ύπαρξη σημαντικής στατιστικής συσχέτισης μεταξύ των κόμβων σε διαφορετικές ώρες της ημέρας, η οποία μπορεί να αποδοθεί σε παρόμοια πρότυπα κυκλοφορίας που εμφανίζονται στους συγκεκριμένους κόμβους κατά τις διαφορετικές ώρες της ημέρας, όπως η εμφάνιση συνθηκών συμφόρησης.

*Ανάπτυξη μοντέλων πρόβλεψης και αποτελέσματα*

Για την πραγματοποίηση προβλέψεων, αναπτύσσεται ένα μοντέλο για κάθε κόμβο και για κάθε ώρα της ημέρας για την πρόβλεψη της ζήτησης στον αντίστοιχο ανιχνευτή ή σταθμό, χρησιμοποιώντας ως χαρακτηριστικά εισόδου τις τελευταίες τιμές της ζήτησης των υπόλοιπων κόμβων που ανήκουν στην ίδια κοινότητα. Έτσι, ο αριθμός των χαρακτηριστικών εισόδου κάθε μοντέλου είναι ίσος με το μέγεθος της κοινότητας στην οποία ανήκει το ζητούμενο ζεύγος (κόμβος, ώρα) και η τιμή κάθε χαρακτηριστικού αντιστοιχεί στις μετρήσεις των προηγούμενων 24 ωρών. Τα παραπάνω δεδομένα αξιοποιούνται για την πρόβλεψη των κυκλοφοριακών συνθηκών ή της ζήτησης (για τους κόμβους των σταθμών του μετρό) με ορίζοντα πρόβλεψης 1 ώρα. Σε αυτό το κεφάλαιο, οι τιμές των μεταβλητών-στόχων δεν κατηγοριοποιούνται, αλλά προβλέπεται η ακριβής τιμή τους.

Στη συνέχεια, για κάθε ζεύγος κόμβου-χρόνου, ένα μοντέλο παλινδρόμησης Gradient Boosting αναπτύσσεται, το οποίο είναι ένα μοντέλο Μηχανικής Μάθησης με σημαντικά μικρότερη πολυπλοκότητα σε σχέση με τα Βαθιά Νευρωνικά Δίκτυα. Τα αποτελέσματα συγκρίνονται με ένα μοντέλο LSTM για κάθε ανιχνευτή και σταθμό μετρό, το οποίο εφαρμόζεται στα αρχικά δεδομένα, χωρίς να λαμβάνεται υπόψη η δομή του πολυεπίπεδου γράφου και ο εντοπισμός των κοινοτήτων.

Ο μέσος όρος των σφαλμάτων των μοντέλων που αναπτύχθηκαν με τις δύο παραπάνω στρατηγικές παρουσιάζεται στον Πίνακας 3. Παρατηρούμε ότι για το προτεινόμενο μοντέλο, επιτυγχάνεται μία ικανοποιητικά μικρή τιμή του Μέσου Απόλυτου Ποσοστιαίου Σφάλματος (MAPE) 9,5%, ενώ οι κόμβοι ανιχνευτών βρόχου προβλέπονται με μεγαλύτερη ακρίβεια σε σχέση με τους σταθμούς του μετρό. Επιπλέον, η προτεινόμενη αναπαράσταση και επιλογή των πιο σχετικών μεταβλητών εισόδου σε συνδυασμό με ένα απλούστερο μοντέλο οδηγούν σε σημαντικά καλύτερη απόδοση ως προς το σφάλμα πρόβλεψης σε σχέση με ένα μοντέλο υψηλότερων δυνατοτήτων και, θεωρητικά, πιο κατάλληλο για τη συγκεκριμένη εργασία.

**Πίνακας 3. Σύγκριση μετρικών σφαλμάτων των μοντέλων**

| **Metrics** | **Multiplex-Gradient Boosting** | | **LSTM** | |
|---|---|---|---|---|
| | Traffic | Transit | Traffic | Transit |
| **MAE** | 91.28 | 30.81 | 133.14 | 52.02 |
| **MAPE** | 9.0% | 11.3% | 16.8% | 23.0% |
| **Overall MAPE** | 9.5% | | 18.8% | |

## Πρόβλεψη σε Επίπεδο Δικτύου

Τα μοντέλα που έχουν παρουσιαστεί μέχρι στιγμής, χρησιμοποιούν μεταβλητές εισόδου από θέσεις σε όλο το δίκτυο. Εκμεταλλευόμενα τις χωροχρονικές σχέσεις τους, παρέχουν προβλέψεις των μελλοντικών συνθηκών κυκλοφορίας σε συγκεκριμένες θέσεις-στόχους. Για κάθε θέση-στόχο απαιτείται ένα ξεχωριστό μοντέλο, με τις αντίστοιχες μεταβλητές εισόδου και τιμές υπερπαραμέτρων. Ένα μοντέλο πρόβλεψης που θα είναι σε θέση να παρέχει προβλέψεις για όλο το δίκτυο σε περιβάλλον πολλαπλών διεργασιών, δηλαδή να προβλέπει τις συνθήκες κυκλοφορίας σε όλες τις εξεταζόμενες θέσεις ταυτόχρονα, χρησιμοποιώντας μία ενιαία δομή και αξιοποιώντας περισσότερες από μία πηγές δεδομένων, που αντιστοιχούν σε κάθε θέση θα είχε πολύ σημαντικές επιπτώσεις στη διαχείριση της κυκλοφορίας, καθώς θα παρείχε όλες τις απαραίτητες πληροφορίες. Επιπλέον, για να είναι εφαρμόσιμο, το μοντέλο αυτό θα πρέπει να είναι ακριβές και ταυτόχρονα αποδοτικό, όσον αφορά τον χρόνο εκπαίδευσης και τις απαιτήσεις σε υπολογιστική ισχύ και δεδομένα, κάτι που δεν μπορεί να εξασφαλιστεί ακόμη ούτε για μοντέλα μίας θέσης.

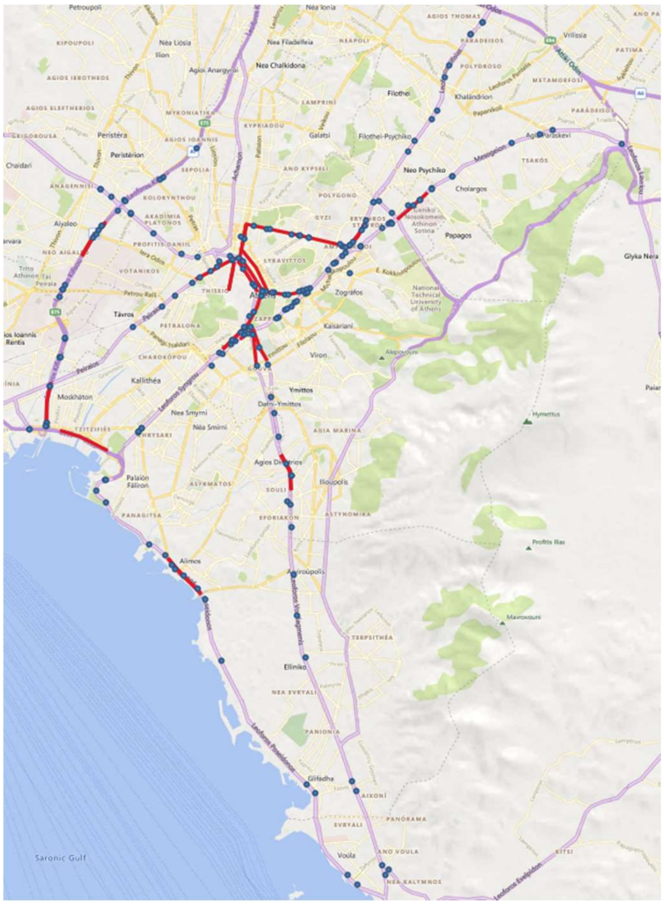
Σε αυτήν την ενότητα, παρουσιάζεται μια από τις πρώτες προσπάθειες στη βιβλιογραφία για την ταυτόχρονη πρόβλεψη των χρόνων διαδρομής για τη διέλευση 30 σημαντικών οδικών τμημάτων γύρω από το κέντρο της Αθήνας, χρησιμοποιώντας ως δεδομένα εισόδου δεδομένα κυκλοφοριακού φόρτου από ανιχνευτές βρόχου. Για την αξιολόγηση των αποτελεσμάτων της πρόβλεψης, εκτός από τις κλασικές μετρικές σφάλματος, παρέχονται επεξηγήσεις των αποτελεσμάτων, με βάση την εκτίμηση των τιμών Shapley (ή SHAP).

*Διαθέσιμα δεδομένα*

Για την ανάπτυξη του μοντέλου πρόβλεψης, συνδυάζονται τα δεδομένα κυκλοφοριακού φόρτου των ανιχνευτών βρόχου από την πόλη της Αθήνας με δεδομένα χρόνου ταξιδιού που ανακτήθηκαν από τη δημοφιλή υπηρεσία χαρτών και πλοήγησης Google Maps. Οι χρόνοι ταξιδιού αναφέρονται στο χρόνο (σε δευτερόλεπτα) για τη διέλευση 30 από τα πιο πολυσύχναστα και κρίσιμα οδικά τμήματα κατά τις πρωινές και απογευματινές ώρες αιχμής, δηλαδή από τις 8 π.μ. έως τις 10 π.μ. και από τις 12 μ.μ. έως τις 7 μ.μ., για κάθε ημέρα μεταξύ Ιουνίου και Δεκεμβρίου του 2021. Ο συνολικός αριθμός των καταγραφών είναι 1110. Οι φωρατές και οι διαδρομές που λήφθηκαν υπόψη φαίνονται στην Εικόνα 4.

*Ανάπτυξη μοντέλου και αποτελέσματα πρόβλεψης*

Το μοντέλο πρόβλεψης που αξιοποιείται είναι ένα σχετικά ρηχό Νευρωνικό Δίκτυο με 3 κρυφά στρώματα. Η είσοδος του μοντέλου είναι οι ωριαίες μετρήσεις των επιλεγμένων ανιχνευτών βρόχων (μία μέτρηση ανά ανιχνευτή), ενώ η έξοδος περιλαμβάνει και τους 30 χρόνους διαδρομής που αντιστοιχούν στην επόμενη ώρα από αυτή των δεδομένων εισόδου. Παρόλο που το μοντέλο που χρησιμοποιείται είναι απλό σε σύγκριση με την πολυπλοκότητα του χώρου εξόδου, εκπαιδεύεται αποτελεσματικά και αναμένεται να έχει αρκετά καλές επιδόσεις. Εξάλλου, επιλέχθηκε να χρησιμοποιείται ως είσοδος μόνο η τελευταία μέτρηση κάθε ανιχνευτή βρόχου και όχι χρονοσειρές μετρήσεων, οι οποίες θα αύξαναν, χωρίς να είναι απαραίτητο, την πολυπλοκότητα.



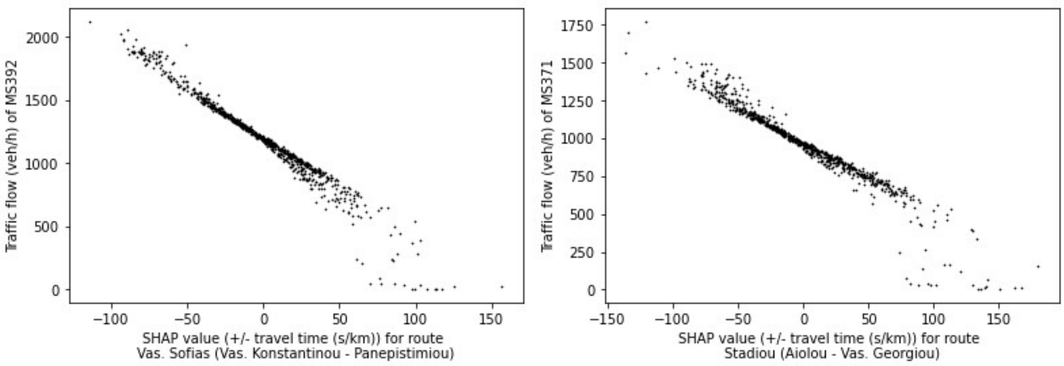
**Εικόνα 4. Τοποθεσίες φωρατών και διαδρομών**

Το μοντέλο αποδίδει σε γενικές γραμμές ικανοποιητικά, επιτυγχάνοντας ένα μέσο Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (MAPE) 12,3%, ενώ υπάρχει ένας σημαντικός αριθμός διαδρομών των οποίων οι χρόνοι ταξιδιού προβλέπονται με MAPE μικρότερο από 9% (τεταρτημόριο 25%). Επιπλέον, οι υψηλότερες παρατηρούμενες τιμές MAPE είναι 20,4% και 19,4%, ενώ το τεταρτημόριο 75% του MAPE είναι 15,8%, το οποίο θεωρείται αποδεκτό. Η μέση τιμή του μέσου απόλυτου σφάλματος είναι 23,1 s/km. Τα σφάλματα πρόβλεψης για όλες τις διαδρομές παρουσιάζονται αναλυτικά στον Πίνακας 4.

Επιπλέον, από τον υπολογισμό των τιμών SHAP, παρατηρείται ένα σταθερό μοτίβο για τη σχέση μεταξύ των χρόνων διαδρομής και των φόρτων που καταγράφηκαν στους φωρατές: όταν παρατηρείται χαμηλή τιμή του κυκλοφοριακού φόρτου, η οποία υποδηλώνει συνθήκες συμφόρησης, η τιμή SHAP του χρόνου διαδρομής αυξάνεται σημαντικά, δηλαδή αναμένονται υψηλότεροι χρόνοι διαδρομής. Ειδικότερα για τις διαδρομές που βρίσκονται πιο κοντά στο κέντρο της πόλης (π.χ. Σταδίου και Βασ. Σοφίας) παρατηρούνται επίσης υψηλότερες τιμές SHAP γενικά, οι οποίες είναι επίσης ενδεικτικές της μεγαλύτερης διακύμανσης των χρόνων διαδρομής τους και της σημαντικότητας της επίδρασης της εμφάνισης συμφόρησης. Δύο ενδεικτικά παραδείγματα παρουσιάζονται στην Εικόνα 5.

**Πίνακας 4. Αποτελέσματα πρόβλεψης χρόνων διαδρομών**

| **Διαδρομή** | **MAPE** | **MAE** |
|---|---|---|
| Panepistimiou (Vas. Sofias - Patision) | 9.8% | 19.8 |
| Akadimias (Patision - Vas. Sofias) | 11.3% | 26.4 |
| Stadiou (Aiolou - Vas. Georgiou) | 15.9% | 53.5 |
| Athinas (Ermou - Stadiou) | 9.4% | 25.2 |
| Athinas (Stadiou - Ermou) | 8.6% | 21.8 |
| Vas. Sofias (Vas. Konstantinou - Panepistimiou) | 19.4% | 56.4 |
| Vas. Amalias (Ath. Diakou - Panepistimiou) | 17.0% | 44.7 |
| Patision (Alexandras - Stadiou) | 8.3% | 16.6 |
| Pireos (Kolokinthous - Omonoia Sq.) | 15.8% | 51.3 |
| Syngrou Av. (Vas. Amalias - Frantzi) | 6.3% | 9.1 |
| Pireos (Kolokinthous - Iera Odo) | 17.3% | 41.5 |
| Syngrou Av. (Frantzi - Vas. Amalias) | 20.4% | 44.8 |
| Alexandras (Kifisias - Patision) | 11.1% | 21.7 |
| Kallirois (Petmeza - Ardittou) | 10.3% | 32.0 |
| Patision (Ioulianou - Chalkokondili) | 9.2% | 13.0 |
| Patision (Stournari - Ioulianou) | 10.4% | 22.9 |
| Kifisias (Alexandras - Panormou) | 16.0% | 24.5 |
| Kifisias (Panormou - Alexandras) | 13.4% | 21.7 |
| Mesogion (Katechaki - Kiprou) | 13.9% | 16.1 |
| Vouliagmenis (Arditou - Ilia Iliou) | 12.9% | 15.1 |
| Vouliagmenis (Ag. Konstantinou - Pirronos) | 4.9% | 1.9 |
| Vouliagmenis (Pirronos - Ag. Konstantinou) | 5.8% | 1.7 |
| Ilioupoleos (Ilia Iliou - Arditou) | 15.2% | 32.2 |
| Cephissus (Posidonos - Pireos) | 15.0% | 11.7 |
| Cephissus (Pireos - Posidonos) | 13.7% | 6.3 |
| Cephissus (Athinon - Moudrou) | 8.0% | 1.3 |
| Cephissus (Moudrou - Athinon) | 17.3% | 24.7 |
| Posidonos (Niarchos - Cephissus) | 5.7% | 0.4 |
| Posidonos (Amfitheas - Alimou) | 13.8% | 14.4 |
| Posidonos (Alimou - Amfitheas) | 14.0% | 21.0 |
| **Μέσος όρος** | **12.3** | **23.1** |




**Εικόνα 5. Τιμές SHAP ενδεικτικών διαδρομών.**

Λαμβάνοντας υπόψη την αυξημένη αποδοτικότητα ενός μοντέλου που μπορεί να παρέχει έγκυρες προβλέψεις σε όλο το δίκτυο, η σημασία του στην προγνωστική διαχείριση της κυκλοφορίας και στη λήψη αποφάσεων είναι μεγάλη και η έλλειψη παρόμοιων προσεγγίσεων είχε προηγουμένως επισημανθεί ως μία από τις κύριες ανοικτές προκλήσεις για την ανάπτυξη μοντέλων που μπορούν να χρησιμοποιηθούν σε πραγματικές συνθήκες. Τα αποτελέσματα δείχνουν επίσης ότι υπάρχουν σημαντικές βραχυπρόθεσμες και μακροπρόθεσμες σχέσεις μεταξύ των κυκλοφοριακών συνθηκών και των χρόνων διαδρομής των επιλεγμένων διαδρομών- έτσι, μια προσέγγιση μάθησης πολλαπλών διεργασιών, όπως αυτή που παρουσιάζεται, μπορεί να τις εκμεταλλευτεί αποτελεσματικά και αποδοτικά.

## Αιτιώδης Βαθιά Μάθηση για Βραχυπρόθεσμη Πρόβλεψη της Κυκλοφορίας

Η μέθοδος της θεωρίας της πληροφορίας (αμοιβαία πληροφορία), καθώς και οι άλλες στατιστικές μέθοδοι που χρησιμοποιήθηκαν σε προηγούμενες ενότητες για την ανίχνευση χωροχρονικών σχέσεων μεταξύ των θέσεων του οδικού δικτύου, παρουσιάζουν ένα σοβαρό μειονέκτημα: υπονοούν την ύπαρξη μιας στατιστικής σχέσης, η οποία μπορεί να είναι σημαντική, αλλά δεν εγγυάται την αιτιότητα. Κατά συνέπεια, οι σχέσεις που ανιχνεύονται μπορεί να αφορούν αποκλειστικά το συγκεκριμένο σύνολο δεδομένων, δηλαδή δεν είναι γενικεύσιμες και πιθανόν να μην ισχύουν για ακραίες συνθήκες ή/και μελλοντικά δεδομένα. Ως εκ τούτου, περιορίζουν την αξιοπιστία και την επεξηγησιμότητα του μοντέλου.

Στην παρούσα ενότητα, παρουσιάζεται ένα πλαίσιο για τον προσδιορισμό της αιτιώδους δομής ενός συνόλου κυκλοφοριακών δεδομένων που αποτελείται από μετρήσεις σε διαφορετικές θέσεις του οδικού δικτύου, το οποίο μπορεί επίσης να εφαρμοστεί σε άλλα προβλήματα χρονοσειρών, με βάση το νευρωνικό μοντέλο Granger (Neural Granger). Αξιοποιούνται δεδομένα κυκλοφοριακού φόρτου από περισσότερους από 300 φωρατές που είναι εγκατεστημένοι στο οδικό δίκτυο της Αθήνας και το προτεινόμενο πλαίσιο εφαρμόζεται τόσο για την εξήγηση των αποτελεσμάτων του μοντέλου όσο και για την επιλογή χαρακτηριστικών. Στη συνέχεια, αναπτύσσεται ένα μάλλον απλό μοντέλο Νευρωνικού Δικτύου Μακράς Βραχυπρόθεσμης Μνήμης (LSTM) για την πρόβλεψη των κυκλοφοριακών συνθηκών σε πολλές σημαντικές τοποθεσίες και, αξιοποιώντας τα αποτελέσματα του πλαισίου, επιτυγχάνει πολύ ικανοποιητικές επιδόσεις.

*Νευρωνικό Μοντέλο Αιτιότητας Granger*

Προκειμένου να ξεπεραστούν οι αδυναμίες της κλασικής προσέγγισης και κυρίως η παραδοχή της γραμμικής σχέσης, αναπτύσσουμε ένα νευρωνικό δίκτυο και συγκεκριμένα ένα LSTM για την προσομοίωση της σχέσης μεταξύ των μεταβλητών εισόδου και εξόδου, με βάση τις αρχές του Νευρωνικού Μοντέλου Granger.

Το μοντέλο εκπαιδεύεται με την ακόλουθη συνάρτηση απωλειών:

$$\frac{1}{N}\sum^{N}(x_t - h(x_{i,<t}))^2 + \lambda\sum_{i=1}^{n}\|W_i^1\|) \qquad (IV)$$

Ο πρώτος όρος είναι το μέσο τετραγωνικό σφάλμα της εκτίμησης του Granger LSTM, ενώ ο δεύτερος είναι το άθροισμα των βαρών εισόδου (βάρος του στρώματος εισόδου) που συνδέονται με κάθε χρονοσειρά του χώρου εισόδου. Η παράμετρος λ παίζει τον ίδιο ρόλο όπως και στην κλασική, γραμμική προσέγγιση, δηλαδή ελέγχει τη σημαντικότητα του δεύτερου όρου. Αντίστοιχα, όταν μια χρονοσειρά δεν έχει σημαντική συνεισφορά στον πρώτο όρο (δηλαδή δεν μειώνει σημαντικά το σφάλμα εκτίμησης), το βάρος εισόδου της συρρικνώνεται στο μηδέν (δεύτερος όρος) και δεν εισέρχεται στο μοντέλο. Στο τέλος της διαδικασίας εκπαίδευσης, οι χρονοσειρές με μη μηδενικά βάρη εισόδου θεωρείται ότι προκαλούν κατά Granger την έξοδο.

*Υλοποίηση*

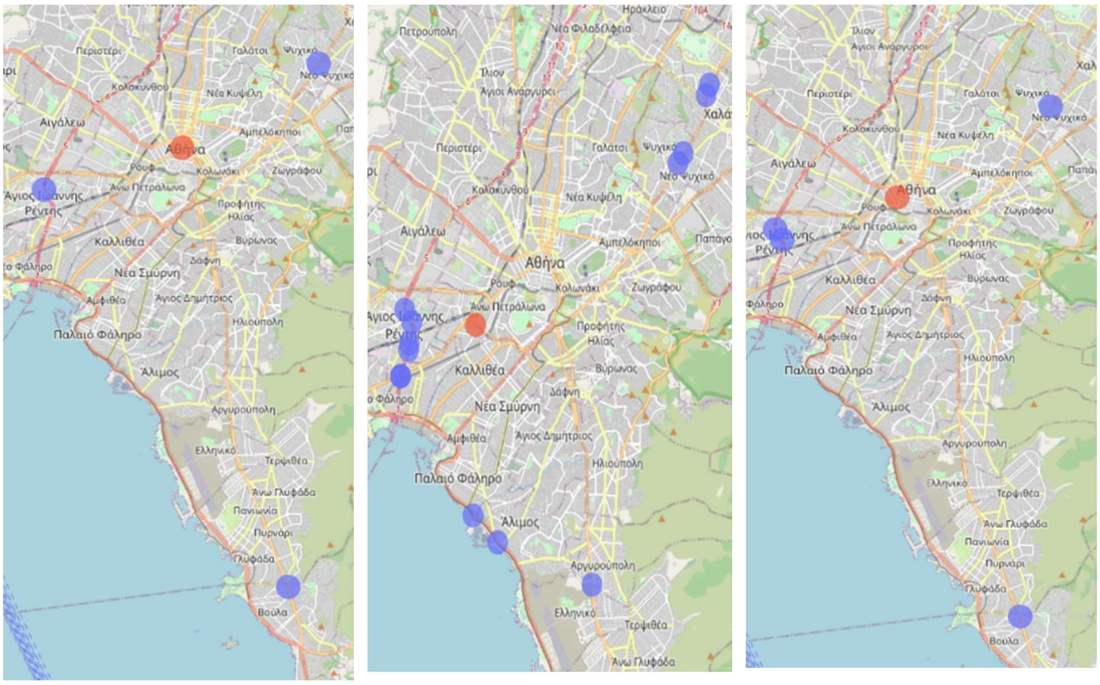
Στην προτεινόμενη προσέγγισή, αρχικά αξιοποιείται το νευρωνικό μοντέλο αιτιότητας Granger για να κατασκευαστεί ο γράφος αιτιότητας του οδικού δικτύου της Αθήνας, ο οποίος αποτελείται από περισσότερες από 330 θέσεις/φωρατές που είναι εγκατεστημένοι σε ολόκληρο το οδικό δίκτυο. Το Neural Granger, υλοποιημένο ως δίκτυο LSTM, επιλέχθηκε ως η καταλληλότερη μέθοδος για να χειριστεί ένα τόσο μεγάλο δίκτυο με πολύπλοκες εξαρτήσεις μεταξύ των θέσεων. Κατά συνέπεια, αναπτύσσεται ένα μικρότερο και απλούστερο δίκτυο LSTM για την εκτέλεση του έργου πρόβλεψης, το οποίο αναφέρεται επίσης ως "Αραιό LSTM", καθώς ο χώρος εισόδου του περιλαμβάνει μόνο τις χρονοσειρές που βρέθηκαν να προκαλούν κατά Granger τη χρονοσειρά-στόχο κατά το πρώτο βήμα του προτεινόμενου πλαισίου. Θα πρέπει να καταστεί σαφές ότι για κάθε θέση-στόχο θα πρέπει να αναπτυχθούν ένα Granger LSTM και ένα Sparse LSTM. Η βάση δεδομένων που χρησιμοποιήθηκε είναι αυτή με χρονική ανάλυση μίας ώρας που παρουσιάστηκε και σε προηγούμενα κεφάλαια.

*Εντοπισμός αιτιωδών σχέσεων*

Η εκπαίδευση του Neural Granger LSTM καταλήγει στον αιτιώδη γράφο του οδικού δικτύου, δηλαδή σε ένα σύνολο εντοπισμένων αιτιωδών σχέσεων. Κατά μέσο όρο, ένα μικρό τμήμα του οδικού δικτύου, περίπου το 7,3% των χρονοσειρών (περίπου 25 θέσεις), βρέθηκε να προκαλεί κατά Granger κάθε χρονοσειρά-στόχο. Το παραπάνω εύρημα δείχνει ότι σε ένα τόσο μεγάλο οδικό δίκτυο, θα πρέπει να καταβληθεί σοβαρή προσπάθεια στην επιλογή χαρακτηριστικών, καθώς οι πραγματικές αιτιώδεις σχέσεις είναι πολύ λίγες και η διαστατικότητα του χώρου εισόδου μπορεί να μειωθεί δραματικά, χωρίς να θυσιαστεί η απόδοση πρόβλεψης του μοντέλου.

Στην Εικόνα 6 παρουσιάζονται ενδεικτικά οι περιοχές του δικτύου που βρέθηκε ότι προκαλούν κατά Granger κάποια σημεία στο κέντρο της πόλης.

Ενδιαφέρον παρουσιάζει το γεγονός ότι παρατηρούμε ένα πολύ ισχυρό, κοινό μοτίβο για όλες τις θέσεις-στόχους: οι ανιχνευτές βρόχων που βρίσκονται στο κέντρο της πόλης προκαλούνται κατά Granger από ανιχνευτές που βρίσκονται σε σημαντικά τμήματα του οδικού δικτύου στην περίμετρο της πόλης της Αθήνας, που λειτουργούν ως είσοδοι στο κέντρο της πόλης και, ως επί το πλείστον, όχι πολύ κοντά στο κέντρο της πόλης. Επιπλέον, ένα άλλο σημαντικό εύρημα είναι ότι υπάρχουν λίγοι (3 ή 4) συγκεκριμένοι ανιχνευτές βρόχων που προκαλούν κατά Granger σχεδόν όλους τους ανιχνευτές βρόχων του κέντρου της πόλης- αποδεικνύεται ότι οι τελευταίοι ανιχνευτές μπορούν να παρέχουν ζωτικής σημασίας πληροφορίες σχετικά με τις μελλοντικές κυκλοφοριακές συνθήκες στο κέντρο της πόλης, οπότε οι αρχές διαχείρισης της κυκλοφορίας και άλλοι επαγγελματίες θα πρέπει πάντα να λαμβάνουν υπόψη τις μετρήσεις τους για σκοπούς πρόβλεψης και λήψης αποφάσεων. Τέλος, μπορεί κανείς να παρατηρήσει ότι οι αιτιώδεις σχέσεις που εντοπίστηκαν είναι μεταξύ των ανιχνευτών βρόχων-στόχων και άλλων ανιχνευτών που δεν βρίσκονται πολύ κοντά σε αυτούς.




**Εικόνα 6. Περιοχές που προκαλούν κατά Granger τις περιοχές-στόχους**

*Αποτελέσματα πρόβλεψης*

Για την αξιολόγηση της ακρίβειας του προτεινόμενου Αραιού LSTM για κάθε τοποθεσία, πραγματοποιούνται συγκρίσεις με βάση ένα δίκτυο LSTM που λαμβάνει ως είσοδο τη χρονοσειρά μόνο από την τοποθεσία-στόχο (Single-point LSTM) και ένα δίκτυο LSTM που λαμβάνει είσοδο από όλες τις τοποθεσίες (Inclusive LSTM), όσον αφορά την ακρίβεια πρόβλεψης και την υπολογιστική αποδοτικότητα. Πιο συγκεκριμένα, για κάθε ένα από τα 334 μοντέλα μετρήθηκαν οι τιμές MAPE και ο χρόνος εκπαίδευσης.

Όπως φαίνεται ξεκάθαρα στον Πίνακας 5, τα προτεινόμενα αραιά μοντέλα, εκμεταλλευόμενα τη στρατηγική επιλογής χαρακτηριστικών που εφαρμόζεται με βάση την ανίχνευση αιτιώδους δομής, επιτυγχάνουν καλύτερη επίδοση πρόβλεψης από τα δύο άλλα, ενώ είναι σημαντικά πιο αποδοτικά από το Inclusive μοντέλο, καθώς η αρχιτεκτονική τους είναι πιο απλή, λόγω της μειωμένης διαστατικότητας του χώρου εισόδου.

**Πίνακας 5. Αξιολόγηση μοντέλων πρόβλεψης (μέση τιμή όλων των τοποθεσιών)**

| Model | Average MAPE (and deviation) | Efficiency (time to train per model) |
|---|---|---|
| Inclusive LSTM | 12.6% ± 3.1% | 104s |
| Single-point LSTM | 13.3% ± 3.7% | 27s |
| Sparse LSTM | 9.1% ± 1.8% | 46s |

## Ένα θεωρητικά ενημερωμένο, πολυμεταβλητό, αιτιώδες πλαίσιο για βραχυπρόθεσμες προβλέψεις κυκλοφορίας

Σε αυτή την ενότητα, ενσωματώνονται πρόσθετες πτυχές που σχετίζονται με τη θεωρία της ροής της κυκλοφορίας και προτείνεται μια ολιστική προσέγγιση (από την προεπεξεργασία των δεδομένων έως την εκπαίδευση και την αξιολόγηση των μοντέλων) για ένα πλαίσιο πρόβλεψης της κυκλοφορίας που βασίζεται στη θεωρία, είναι αιτιώδες και πολυμεταβλητό, με στόχο να είναι επίσης εφαρμόσιμο. Στον πυρήνα της προτεινόμενης μεθοδολογίας, υπάρχει ένα καινοτόμο νευρωνικό δίκτυο πολλαπλών διεργασιών που βασίζεται στη θεωρία της κυκλοφοριακής ροής, το οποίο χρησιμοποιείται για την από κοινού βραχυπρόθεσμη πρόβλεψη δύο μεταβλητών της κυκλοφορίας ( κυκλοφοριακός φόρτος και ταχύτητα), οι οποίες συνιστούν μια κατάσταση κυκλοφορίας. Για την εκπαίδευση του μοντέλου, προτείνεται μια προσαρμοσμένη συνάρτηση απωλειών, η οποία ενσωματώνει την απόσταση της προκύπτουσας πολυμεταβλητής πρόβλεψης (ζεύγη κυκλοφοριακών μεταβλητών) από το πραγματικό θεμελιώδες διάγραμμα της αντίστοιχης θέσης. Για να ενισχυθεί η απόδοση και η ερμηνευσιμότητα του μοντέλου, επιλέγονται πληροφορίες εισόδου κυκλοφορίας σε επίπεδο δικτύου από τις πιο σχετικές θέσεις, χρησιμοποιώντας το νευρωνικό μοντέλο Granger. Για τα πειράματα που παρουσιάζονται στην παρούσα εργασία, αναπτύσσεται ένα δίκτυο Μακράς Βραχυπρόθεσμης Μνήμης (LSTM), ωστόσο, ολόκληρη η μεθοδολογία (συμπεριλαμβανομένης της συνάρτησης απωλειών) είναι συμβατή με οποιαδήποτε δομή βαθιάς μάθησης. Ολόκληρη η μεθοδολογία εφαρμόζεται σε δεδομένα υψηλής ανάλυσης που προέρχονται από το οδικό δίκτυο της Αθήνας.

*Συνάρτηση απωλειών με βάση τη θεωρία ροής κυκλοφορίας*

Η εκπαίδευση ενός μοντέλου βαθιάς μάθησης είναι ένα πρόβλημα βελτιστοποίησης που αποσκοπεί στον προσδιορισμό των βέλτιστων τιμών των βαρών του για την ελαχιστοποίηση μιας συνάρτησης απωλειών, όπως το μέσο τετραγωνικό σφάλμα (MSE) ή το μέσο τετραγωνικό σφάλμα (RMSE). Λαμβάνοντας υπόψη την περίπτωση της κοινής πρόβλεψης κυκλοφοριακού φόρτου και ταχύτητας, σε μια διάταξη μοντελοποίησης πολλαπλών διεργασιών, η πιο δημοφιλής προσέγγιση θα ήταν η εκτίμηση μιας τιμής απώλειας (μετρική σφάλματος) για κάθε μία από τις προβλέψεις φόρτου και ταχύτητας σε κάθε εποχή εκπαίδευσης. Όσο χαμηλότερη είναι η απώλεια τόσο καλύτερες είναι οι επιμέρους προβλέψεις. Ωστόσο, εκτός από μια χαμηλή τιμή σφάλματος, θα πρέπει να περιμένουμε ότι όλα τα προβλεπόμενα ζεύγη τιμών θα πρέπει να βρίσκονται κοντά στην αντίστοιχη θεμελιώδη καμπύλη ταχύτητας-φόρτου. Το τελευταίο δεν εξασφαλίζεται από μια χαμηλή τιμή σφάλματος- στην πραγματικότητα, μια αξιοπρεπής ατομική μέση τιμή σφάλματος για κάθε μεταβλητή μπορεί να καλύπτει ζητήματα όπως τα προαναφερθέντα.

Για να ληφθούν υπόψη οι παραπάνω προκλήσεις, προτείνεται η συνάρτηση απώλειας με βάση τη θεωρία ροής κυκλοφορίας (Traffic flow theory-informed loss function - TFTI loss), η οποία συνδυάζει το MSE των δύο επιμέρους μεταβλητών με την απόσταση της πρόβλεψης από το πλησιέστερο σημείο στο εκτιμώμενο θεμελιώδες διάγραμμα. Το τελευταίο ορίζεται για κάθε θέση/τμήμα του οδικού δικτύου που παρακολουθείται και μπορεί να έχει συγκεκριμένη συναρτησιακή μορφή. Έστω $\hat{y}_i = (\hat{v}_i, \hat{s}_i)$ η πρόβλεψη για ένα πραγματικό ζεύγος $y_i = (v_i, s_i)$ και $g_j = \left(v_j^e, s_j^e\right)$ το πλησιέστερο σημείο στο εκτιμώμενο θεμελιώδες διάγραμμα που χαρακτηρίζει τη θέση ενδιαφέροντος, η απώλεια TFTI ορίζεται ως εξής:

$$TFTI\,loss = a * \frac{1}{2}\sqrt{[MSE_v + MSE_s]} + (1 - a) * d(\hat{y}_i, g_j) \qquad (V)$$

όπου $MSE_v$, $MSE_s$ είναι το μέσο τετραγωνικό σφάλμα των προβλέψεων του όγκου και της ταχύτητας αντίστοιχα, $d(\hat{y}_i, g_j)$ είναι η ευκλείδεια απόσταση των προβλεπόμενων ζευγών και του πλησιέστερου σημείου του θεμελιώδους διαγράμματος στην αντίστοιχη πραγματική τιμή. Ο παράγοντας $a$ ελέγχει τη σημασία του δεύτερου παράγοντα έναντι του πρώτου, $a \in [0,1]$.

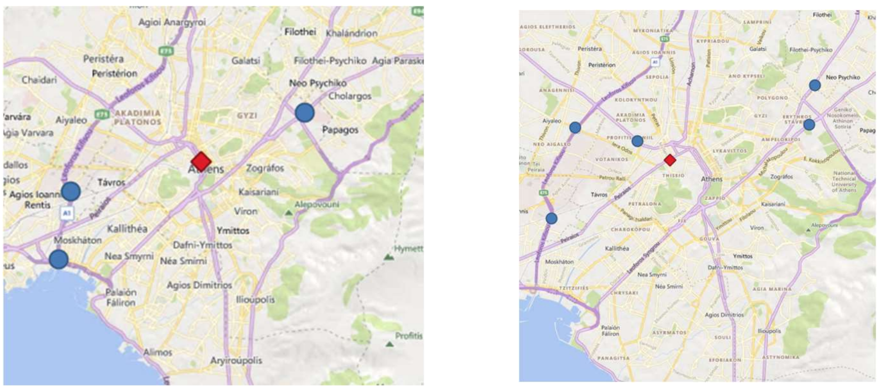
*Διαθέσιμα δεδομένα και διάταξη μοντελοποίησης*

Για τους σκοπούς της παρούσας εργασίας, τέθηκαν στη διάθεσή μας μετρήσεις από 420 φωρατές του δικτύου της Αθήνας, οι οποίες αποτελούνται από τον κυκλοφοριακό φόρτο (οχήματα/ώρα), τη μέση ταχύτητα (km/h) και την κατάληψη (%). Η χρονική ανάλυση των δεδομένων είναι 6 λεπτά, δηλαδή τα δεδομένα που ανιχνεύονται συναθροίζονται σε διαστήματα των 6 λεπτών, συνεπώς είναι κατάλληλα για βραχυπρόθεσμες προβλέψεις. Τα δεδομένα αφορούν 40 ημέρες μετρήσεων, μεταξύ 20 Μαρτίου και 30 Απριλίου 2023.

Είναι σημαντικό να τονιστεί ότι η συνάρτηση απωλειών που βασίζεται στη θεωρία της κυκλοφοριακής ροής, καθώς και ολόκληρο το πλαίσιο που παρουσιάζεται σε αυτή την ενότητα, είναι συμβατά με οποιοδήποτε μοντέλο βαθιάς μάθησης. Στην παρούσα εργασία, προτιμήθηκε να χρησιμοποιηθεί μια μάλλον απλή αρχιτεκτονική, δηλαδή ένα δίκτυο LSTM, για να τονιστεί ότι το προτεινόμενο πλαίσιο μπορεί να επιτύχει κορυφαίες επιδόσεις ακόμη και όταν χρησιμοποιείται μια απλούστερη δομή. Οι δύο προσεγγίσεις που συγκρίνονται είναι η εκπαίδευση του μοντέλου με (i) απλό σφάλμα MSE και (ii) την προτεινόμενη συνάρτηση απωλειών TFTI, με παράμετρο a=0,7. Εκτός από τη συνάρτηση απωλειών, τα δύο μοντέλα είναι απολύτως όμοια μεταξύ τους.

*Εντοπισμός αιτιωδών συσχετίσεων*

Μετά την εφαρμογή της μεθόδου νευρωνικού Granger στο διαθέσιμο σύνολο δεδομένων, προέκυψαν οι τοποθεσίες που προκαλούν κατά Granger τις τοποθεσίες-στόχους. Ο μέσος αριθμός των εντοπισμένων θέσεων ήταν 6,6 από τις 420 θέσεις, γεγονός που δείχνει ότι η διαστατικότητα του χώρου εισόδου για κάθε μοντέλο πρόβλεψης μπορεί να μειωθεί κατά πολύ, χωρίς ωστόσο να χαθούν σημαντικές πληροφορίες. Οι τοποθεσίες που βρέθηκαν να προκαλούν κατά Granger κάποιες ενδεικτικές τοποθεσίες-στόχους παρουσιάζονται στην Εικόνα 7.




**Εικόνα 7. Θέσεις (μπλε) που προκαλούν κατά Granger τις θέσεις-στόχους (κόκκινο)**

*Αποτελέσματα πρόβλεψης*

Στον Πίνακας 6 παρουσιάζονται το μέσο απόλυτο σφάλμα (MAE) και το μέσο απόλυτο ποσοστιαίο σφάλμα (MAPE) των προβλέψεων για τις δύο συναρτήσεις απωλειών συνολικά και για κάθε μεταβλητή ξεχωριστά για ορισμένες ενδεικτικές θέσεις στόχων.

**Πίνακας 6. Σύγκριση αποτελεσμάτων πρόβλεψης**

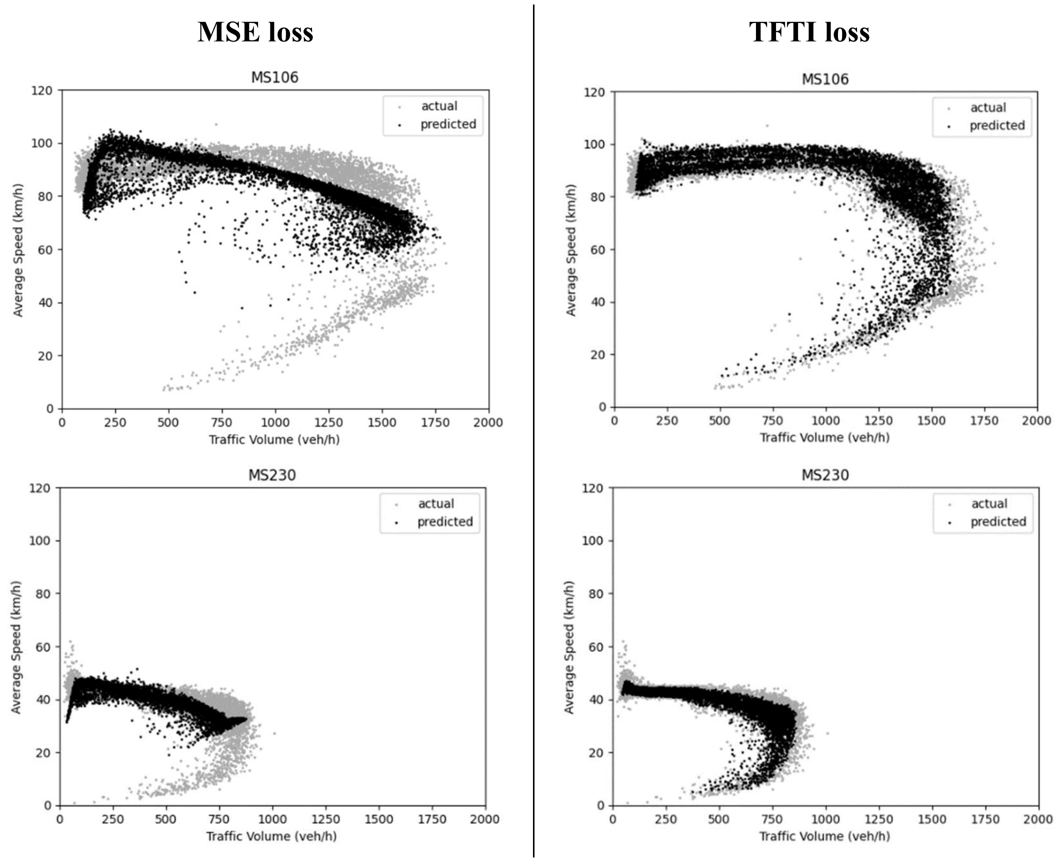
| | *MSE loss* | ***TFTI loss*** | *MSE loss* | ***TFTI loss*** |
|---|---|---|---|---|
| *Location* | **MS106** | | **MS230** | |
| *Volume MAE* | 40.6 | 62.3 | 35.4 | 39.5 |
| *Volume MAPE* | 6.8% | 9.1% | 9.2% | 10.5% |
| *Speed MAE* | 8.0 | 3.3 | 3.9 | 2.1 |
| *Speed MAPE* | 17.3% | 6.4% | 23.0% | 9.4% |
| ***Overall MAPE*** | 12.0% | **7.8%** | 15.6% | **9.9%** |
| *Location* | **MS443** | | **MS634** | |
| *Volume MAE* | 40.4 | 49.2 | 29.5 | 31.0 |
| *Volume MAPE* | 9.6% | 13.0% | 8.6% | 9.7% |
| *Speed MAE* | 4.3 | 2.8 | 3.2 | 1.8 |
| *Speed MAPE* | 17.0% | 9.3% | 11.1% | 6.7% |
| ***Overall MAPE*** | 13.3% | **11.1%** | 9.8% | **8.2%** |

Όπως μπορεί κανείς να παρατηρήσει, η TFTI συνάρτηση απωλειών βοηθά το μοντέλο να επιτύχει χαμηλότερο συνολικό σφάλμα για τις εξεταζόμενες τοποθεσίες. Το μέσο MAPE για την απώλεια TFTI είναι 10,9%, έναντι 13,0% για την απώλεια MSE για τις 12 θέσεις-στόχους. Το τελευταίο οφείλεται κυρίως στο γεγονός ότι το μοντέλο με την απώλεια MSE αποτυγχάνει σχεδόν σε όλες τις περιπτώσεις να προβλέψει με ακρίβεια τις τιμές της ταχύτητας, αν και αποδίδει ικανοποιητικά στην πρόβλεψη του φόρτου κυκλοφορίας. Μια πιθανή εξήγηση γι' αυτό είναι ότι το σύνολο δεδομένων είναι ιδιαίτερα ανισοβαρές, δηλαδή τα περισσότερα σημεία ανήκουν στον κλάδο χωρίς κυκλοφοριακή συμφόρηση, γεγονός που προσθέτει μεροληψία στο μοντέλο ως προς την πρόβλεψη υψηλότερων τιμών ταχύτητας. Από την άλλη πλευρά, η TFTI loss ενσωματώνει την πληροφορία σχετικά με τον κλάδο στον οποίο θα έπρεπε να ανήκει το σημείο (δηλαδή την απόσταση από το θεμελιώδες διάγραμμα) και διατηρεί μια αξιοπρεπή απόδοση και για τον συμφορημένο κλάδο.

*Αξιολόγηση αξιοπιστίας*

Ένας απλός, αλλά ακριβής ορισμός ενός αξιόπιστου μοντέλου είναι ότι πρόκειται για ένα μοντέλο που διατηρεί μια σωστή συμπεριφορά που συνάδει με τα δεδομένα εκπαίδευσης και, γενικά, με την κατανόηση των ειδικών του αντικειμένου για τους παράγοντες που επηρεάζουν το αποτέλεσμα των προβλέψεων. Ως εκ τούτου, καθίσταται σαφές ότι ένα ακριβές μοντέλο δεν είναι πάντα αξιόπιστο.

Για να εξετάσουμε την αξιοπιστία του μοντέλου, παραθέτουμε τα θεμελιώδη διαγράμματα που προκύπτουν από τις προβλέψεις σε σύγκριση με το πραγματικό διάγραμμα ενδεικτικών θέσεων στην Εικόνα 8. Τα γραφήματα δείχνουν ότι οι προβλέψεις που γίνονται από τα μοντέλα που εκπαιδεύονται με την TFTI loss είναι πιο αξιόπιστες, καθώς η κατανομή των σημείων πρόβλεψης είναι πιο κοντά στην πραγματική και στο αντίστοιχο θεμελιώδες διάγραμμα, ενώ τα μοντέλα με την MSE αποτυγχάνουν συστηματικά να προβλέψουν τον κλάδο συμφόρησης. Επιπλέον, η TFTI είναι λιγότερο ευάλωτη σε θορυβώδη δεδομένα και ακραίες τιμές, π.χ. σημεία που βρίσκονται μεταξύ των δύο κλάδων του αναμενόμενου θεμελιώδους διαγράμματος.




**Εικόνα 8. Προβλεπόμενα θεμελιώδη διαγράμματα σε σχέση με τα πραγματικά για την MSE (αριστερά) και την TFTI (δεξιά)**

Τέλος, προχωράμε στην εκτίμηση της ακρίβειας των μοντέλων στην πρόβλεψη της σωστής κατηγορίας συνθηκών, ανεξάρτητα από τις ακριβείς τιμές των μεταβλητών. Δεδομένου ότι οι συνθήκες συμφόρησης αντιπροσωπεύουν περίπου το 11% του συνόλου δεδομένων, ο αντίκτυπος των λανθασμένα ταξινομημένων συνθηκών συμφόρησης δεν είναι τόσο σημαντικός στη συνολική ακρίβεια. Έτσι, εκτιμούμε επιπλέον το συνολικό F1-score και το F1-score της κλάσης με συμφόρηση. Οι αντίστοιχες μέσες τιμές για όλες τις θέσεις-στόχους είναι για την MSE: Ακρίβεια = 0,92, F1-score (συμφορημένη κλάση) = 0,43, Overall F1-score = 0,69, ενώ για την TFTI: Ακρίβεια = 0,96, F1-score (κλάση συμφόρησης) = 0,74, Overall F1-score = 0,86. Εμφανώς, τα μοντέλα που εκπαιδεύονται με την TFTI loss παρέχουν βελτιωμένη απόδοση σε σύγκριση με εκείνα που εκπαιδεύονται με την MSE. Συγκεκριμένα, όσον αφορά τα αποτελέσματα F1, υπερέχουν σημαντικά, ειδικά για την κλάση συμφόρησης. Λαμβάνοντας υπόψη όλα τα παραπάνω, τα μοντέλα που εκπαιδεύτηκαν με την συνάρτηση απωλειών TFTI φαίνεται να είναι πιο αξιόπιστα, εκτός από πιο ακριβή.

## Συμπεράσματα

*Γενικά*

Στην παρούσα διατριβή, παρουσιάστηκε μια εργαλειοθήκη εφαρμόσιμων μεθόδων και τεχνικών για την εκτέλεση διαφορετικών εργασιών πρόβλεψης, και συγκεκριμένα πολυτροπικές προβλέψεις, προβλέψεις πολλαπλών εργασιών σε όλο το δίκτυο, προβλέψεις με βάση τη θεωρία κυκλοφορίας, ανίχνευση στατιστικών και αιτιωδών σχέσεων, καθώς και ένα πλαίσιο για την αξιολόγηση της αξιοπιστίας των προβλέψεων. Εκτός από τις διαφορετικές χωρικές κλίμακες, οι προτεινόμενες δομές μπορούν να αξιοποιηθούν για διαφορετικούς χρονικούς ορίζοντες πρόβλεψης (βραχυπρόθεσμες και μακροπρόθεσμες προβλέψεις), χρησιμοποιώντας δεδομένα υψηλότερης ή χαμηλότερης ανάλυσης. Επιπλέον, διαπιστώθηκε ότι, χρησιμοποιώντας τις παραπάνω μεθόδους, μπορεί κανείς να επιτύχει κορυφαίες επιδόσεις πρόβλεψης, ακόμη και με σχετικά απλές αρχιτεκτονικές μοντελοποίησης, αν και το όλο πλαίσιο είναι συμβατό με οποιοδήποτε μοντέλο Βαθιάς Μάθησης. Έτσι, μια αρμόδια για τη διαχείριση της κυκλοφορίας αρχή, εξοπλισμένη με την παραπάνω εργαλειοθήκη, θα είναι σε θέση να εφαρμόσει προγνωστική διαχείριση της κυκλοφορίας, προγραμματισμό και λήψη αποφάσεων για τον χειρισμό σχεδόν οποιασδήποτε κατάστασης σε επίπεδο πόλης και υπό οποιεσδήποτε συνθήκες.

*Συνεισφορά και καινοτομία*

Η κύρια συνεισφορά της παρούσας εργασίας είναι ένα νέο πλαίσιο πολλαπλών διεργασιών που βασίζεται στη θεωρία της κυκλοφοριακής ροής, το οποίο χρησιμοποιείται για την από κοινού βραχυπρόθεσμη πρόβλεψη δύο μεταβλητών κυκλοφορίας. Επιπλέον, προτείνεται μια νέα προσέγγιση για τον συνδυασμό πολυτροπικών δεδομένων (οδικής κυκλοφορίας και ζήτησης δημόσιων μεταφορών), ώστε να αυξηθεί η προβλεψιμότητα των μελλοντικών τιμών και των δύο μέσων και συγκεκριμένα, μια καινοτόμος αναπαράσταση του οδικού δικτύου και των δεδομένων εισόδου για πολυτροπικές διατάξεις, βασισμένη στην έννοια των πολυεπίπεδων δικτύων (Multiplex Networks) από την ερευνητική περιοχή της ανάλυσης κοινωνικών δικτύων, η οποία αξιοποιείται για πρώτη φορά στην πρόβλεψη της κυκλοφορίας.

Ακόμη, παρουσιάστηκε μια από τις πρώτες προσπάθειες για πρόβλεψη σε όλο το δίκτυο, με την έννοια ότι, όχι μόνο η είσοδος του μοντέλου περιλαμβάνει πληροφορίες από πολλές ή όλες τις θέσεις του οδικού δικτύου, αλλά επίσης, και το πιο σημαντικό, η έξοδός του αναφέρεται επίσης σε όλες τις θέσεις, οι οποίες προβλέπονται ταυτόχρονα με τη χρήση ενός ενιαίου μοντέλου, με βάση την έννοια της πολυδιεργασιακής μάθησης (Multitask Learning).

Μια άλλη σημαντική συνεισφορά της παρούσας εργασίας είναι ότι προτείνεται ένα νέο πλαίσιο αξιολόγησης που αξιολογεί την αξιοπιστία και την ανθεκτικότητα του μοντέλου σε θορυβώδη και ανισοβαρή δεδομένα. Τέλος, το πλαίσιο που αναπτύχθηκε είναι σχεδιασμένο ώστε να μην βασίζεται υπερβολικά σε πολύπλοκες δομές Βαθιάς Μάθησης, ενώ παράλληλα επιτυγχάνει ικανοποιητική επίδοση.

*Περιορισμοί και μελλοντική έρευνα*

Οι σημαντικότεροι περιορισμοί της παρούσας εργασίας σχετίζονται με την σχετικά περιορισμένη διαθεσιμότητα κυκλοφοριακών δεδομένων υψηλής χρονικής συχνότητας, καθώς και δεδομένων χρόνων διαδρομής και ζήτησης Μέσων Μαζικής Μεταφοράς που δεν επέτρεψαν την ανάπτυξη αντίστοιχων μοντέλων βραχυπρόθεσμων προβλέψεων.

Η μελλοντική έρευνα θα περιλαμβάνει τη χρήση πιο σύνθετων αρχιτεκτονικών Βαθιάς Μάθησης, ώστε να εξεταστεί η περεταίρω βελτίωση της ακρίβειας των μοντέλων. Επίσης θα αξιολογηθεί η δυνατότητα μεταφοράς των εκπαιδευμένων μοντέλων σε άλλα δίκτυα ή σε άλλα σημεία του ίδιου δικτύου. Τέλος, θα επιχειρηθεί ο συνδυασμός των σημαντικότερων εργαλείων που αναπτύχθηκαν για τη δημιουργία ενός πολυτροπικού και βασισμένου στη θεωρία πλαισίου που θα παρέχει προβλέψεις για όλο το δίκτυο.

# 1 INTRODUCTION

## 1.1 Background and Motivation

The increase of private vehicle ownership rates and the growing demand for public and private transport services during the last decades, as a consequence of population growth and urbanization, has put significant pressure on urban transportation systems (X. Yin et al., 2021a). As a result, congested conditions can be more frequently observed and last for longer, having significant direct or indirect effects on public health, society, economy, road safety and the environment (Lana et al., 2018). Within this framework, mitigating the effects of traffic congestion has become a very complex and challenging task for traffic managers. The most prominent tool for alleviating traffic congesting has been the implementation of Intelligent Transportation Systems (ITS) for more than 20 years. The recent surge of Artificial Intelligence algorithms and cutting-edge information and communication technologies has revealed new opportunities but also severe challenges for the efficient management of flows (Boukerche & Wang, 2020; Ye et al., 2022).

Accurate short-term traffic forecasting is considered key for efficient ITS implementation, since – at least conceptually – it enables the early response to the anticipated traffic conditions, such as the initiation of proper mitigation strategies to prevent congestion from occurring (Jiang & Luo, 2021; Kumar & Raubal, 2021). The most important implications of traffic forecasting, as a vital part of ITS, include traffic management and control of traffic flow, user information and travel time estimation and traffic light control, aiming at optimizing the network's conditions and level of service (Vlahogianni et al., 2014; Yao et al., 2019; Boukerche & Wang, 2020).The growing need for traffic forecasts embedded in real-time Connected and Cooperative Intelligent Transportation Systems (C-ITS) has also increased the interest in the research area of traffic forecasting, which has been blooming for the last three decades (Vlahogianni et al., 2004, 2014; Lana et al., 2018; X. Yin et al., 2021a).

Typically, short term traffic forecasting research has been governed by two schools of thought; the statistical thinking and the connectionists, both sharing similarities, but also a lot of differences (Karlaftis & Vlahogianni, 2011). Recently, the unprecedented data availability, which emerged as the result of vast advancements in telecommunications and sensing technology (e.g., smartphones, seamless connectivity, 5G networks, connected vehicles, etc.), combined with the reignited interest in Machine Learning and Deep Learning, shifted researchers' attention, from statistical, parametric and analytical traffic flow models, towards data-driven ones (K. Lee et al., 2021; Mantouka et al., 2021).

Deep Learning methods are commonly acknowledged as having the best performance in terms of forecasting errors, compared to previous approaches, which is also mentioned as their main advantage (Y. Wang et al., 2019). The main reason behind that is their potential, at least theoretically, to approximate almost any function, regardless of its degree of non-linearity and model underlying, complex temporal and spatial relations, such as those in network-wide traffic data (Ye et al., 2022). Additionally, Deep Learning models can extract features from large-scale raw data automatically and, thus, the demanding tasks of feature engineering and selection, which require much effort and, especially, domain knowledge, are not always necessary (Z. Liu et al., 2018; Ye et al., 2022).

However, due to the complexity of their structures and the number of hyperparameters involved, deep learning models require increased resources for training, in terms of data availability and computational power. Also, they tend to be very difficult to interpret and understand the reasoning behind their outcomes, which is the main factor that limits the exploitation of these models in decision- and policy-making, for example for traffic management (Lipton, 2018; Pavlyuk, 2019; Y. Wang et al., 2019; X. Yin et al., 2021a; Fafoutellis & Vlahogianni, 2023a). According to recent literature, the highest-performing methods, namely complex deep learning structures, can be the least explainable, while less accurate models illustrate increased explainability (Gunning et al., 2019). Model explainability refers to models whose outcome is understandable to a human, without requiring the complete understanding of its structure and the algorithm used to train it, but also includes the proof of developing an interpretable model, e.g., by selecting an appropriate structure and input variables (Ribera & Lapedriza, 2019). Furthermore, it relates the model's transparency, trustworthiness and fairness; explainable outcome mechanisms can be used for the evaluation of a system, as well as for improving it and extracting knowledge from it (Ribera & Lapedriza, 2019; Manibardo et al., 2021; Laña et al., 2021).

Interestingly, despite the very high research interest in traffic forecasting, indicated by the big volume of research works getting published, the exploitation of Deep Learning forecasting modules as part of a traffic management scheme remains, disproportionally, low. This fact has raised reasonable considerations about the usability and actionability of Deep Learning in traffic forecasting, as well as the direction towards which researchers are moving and the approach they follow (Laña et al., 2021; Manibardo et al., 2021). Therefore, without underestimating the effectiveness of the more complex deep learning structures on the forecasting task, lately, several researchers propose focusing on other properties of the forecasting models, such as explainability, efficiency, trustworthiness and transferability, and not exclusively on their performance in terms of error metrics. The above are expected to increase the models' actionability in real-world conditions and adoption by traffic management authorities (Manibardo et al., 2021).

In this dissertation, it is attempted to develop a fully actionable traffic forecasting framework, that would be trustworthy and robust, as the developed modules will be enhanced with aspects from traffic flow theory and Granger causality and the results will be evaluated on this basis as well. Thus, the framework would not be deployed as a black box, but would be transparent and suitable for traffic management, decision-making and planning. Moreover, the complexity of the input space and the modeling structure will be kept as low as possible, in order for the framework to be efficient in terms of training time and computational resources requirements. Besides, we argue that by incorporating theory aspects and causal features it is possible to use a less complex modeling architecture without sacrificing the model's performance.

## 1.2 Problem Overview

Traffic forecasting at a network level is the process of estimating the traffic conditions at a future time, in one or more locations of the network, given their past traffic conditions (i.e., historical data). More formally, let's assume that there are $n$ locations whose traffic conditions are observed in a specific road network (e.g. road sections, sensors or intersections) during $T$ time periods; so the observed data can be represented by an $n \times T$ matrix, let $X$. Correspondingly, let $Y$ be an $m \times t$ matrix containing the traffic conditions of the $m$ target locations for $t$ timesteps ahead of $T$ ($T$+1, $T$+2, …, $T$+$t$). The task of traffic forecasting is the determination of a function/model $f$, such that $Y = f(X)$.

The most common indicators of traffic conditions are traffic volume and mean vehicle speed, which are the easiest and most straightforward to collect. Traffic volume refers to the number of passing vehicles from a certain point or road section aggregated in a certain period of time (usually one hour). The mean speed can be expressed both as time mean vehicle speed, which refers to the average speed of all vehicles when passing from a certain point or a road section (where a sensor is installed) over a period of time, or as space mean speed, which is the average speed of all vehicles over a certain distance or road section at a specific moment.

The traffic conditions at a road section can as well be expressed as the travel time for passing the specific segment. Also, the travel times of predetermined routes within the road network can serve as traffic conditions indicators. Finally, the occupancy of specific points of the road network can be used, which is defined as the percentage of time that the point is occupied by a vehicle.

Traffic data are an example of spatial time series data, i.e. they have emerged in a successive temporal order and exhibit both temporal and spatial dependencies. Traffic time series are usually non-stationary, especially when examined in high temporal resolutions (Boukerche & Wang, 2020). Moreover, the influence of traffic patterns in different locations of the road network on the target locations is complex, highly non-linear and varying over time (X. Yin et al., 2021a). The existence of dependencies between the traffic conditions at different locations is also supported by traffic flow theory (Pavlyuk, 2019).

The spatiotemporal dependencies are not limited by the connectivity and the proximity of the locations in space and time; the traffic states of road sections that are close to the target section are not necessarily the most correlated to its traffic state, while the same applies with the traffic states of far time steps, which sometimes are more correlated with the predicted time step than more recent ones (Do, Taherifar, et al., 2019; Jiang & Luo, 2021; X. Yin et al., 2021a). Therefore, the selection of the most relevant historical observations for the prediction remains a challenging, yet vital, task. The identification of the spatiotemporal dependencies is considered an instance of the feature selection problem, whose aim is to identify a subset of the input space that would lead to predictions of decent accuracy, while simplifying the model's structure and its fitting procedure (Pavlyuk, 2019).

Except for the above, traffic forecasting is more challenging than other time series prediction problems, because it incorporates various external natural and human factors that should be taken into account, such as the weather conditions, traffic accidents and special events, and whose effect on the traffic conditions is difficult to be modeled, due to their unpredictable nature (Boukerche & Wang, 2020; Jiang & Luo, 2021; Ye et al., 2022).

## 1.3 Objectives

The main objective of the present dissertation is to develop an actionable multi-scale traffic forecasting framework for predictive traffic management, also applicable to multimodal settings, by exploiting causal spatiotemporal relations and traffic flow theory aspects.

For this purpose, a toolkit of modules for various forecasting tasks is proposed, each of which can be exploited by traffic management authorities under different conditions to accomplish different forecasting tasks from different perspectives, e.g., multitask, multimodal or single task prediction, network-wide or point prediction and short- or longer-term predictions (using high or low-resolution data).

Some more specific objectives towards this direction are the following:

1. Identify spatial relations in road networks, using state-of-the-art methods from applied statistical modeling and information theory, and evaluate their effect on forecasting accuracy.
2. Investigate potential significant interrelations between road traffic conditions and the demand for other modes of transport and how they can be utilized for multimodal forecasting.
3. Use a meaningful representation of the road network to efficiently model both the spatial and temporal relations of the transportation system (demand for different modes).
4. Investigate the efficiency and effectiveness of a multitask (multi-output) model to deliver network-wide predictions.
5. Detect significant causal relations between the traffic conditions at different locations of the road network and the traffic patterns they reveal at a city level, both in the short and long term.
6. Examine the effect of causal relations on enhancing the predictability of the traffic conditions in network-level forecasting, as well as the trustworthiness of the predictions.
7. Introduce theory aspects in the model's training stage by deploying a loss function that evaluates the outcomes of the model and steers them towards a theory-compatible direction. Examine its effect on the performance and trustworthiness of the modeling framework.
8. Evaluate the effect of the theory-guided approach in terms of actionability (performance, efficiency, and trustworthiness) using a dedicated evaluation framework.

## 1.4 Structure

The remainder of this dissertation is structured as follows: In the next section, several challenges related to the development of actionable traffic forecasting models are presented according to the findings of a thorough literature review. Moreover, relevant research questions are raised. In Section 3, three statistical and information-theoretic metrics are used to detect spatial relations in a road network and are compared to each other, while in Section 4 temporal relations are considered as well, by using a multi-resolution version of the LSTM network. In Section 5, we present a novel representation of the spatiotemporal relations between multimodal data based on the concept of Multiplex Networks. Section 6 includes an explainable, multitask (multioutput) framework for network-wide predictions and in Section 7 we present an adaptation of the Granger causality test for extracting causal relations between different locations of the road network. In Section 8, we develop a theory-informed, causal framework for short-term traffic forecasting with increased trustworthiness. Finally, Section 9 includes the most important conclusions of this work, as well as identified limitations and proposed directions for future research. In Figure 1, the structure of the dissertation is presented, along with the relation of each section with one or more objectives, as well as the motivation for developing each corresponding module.

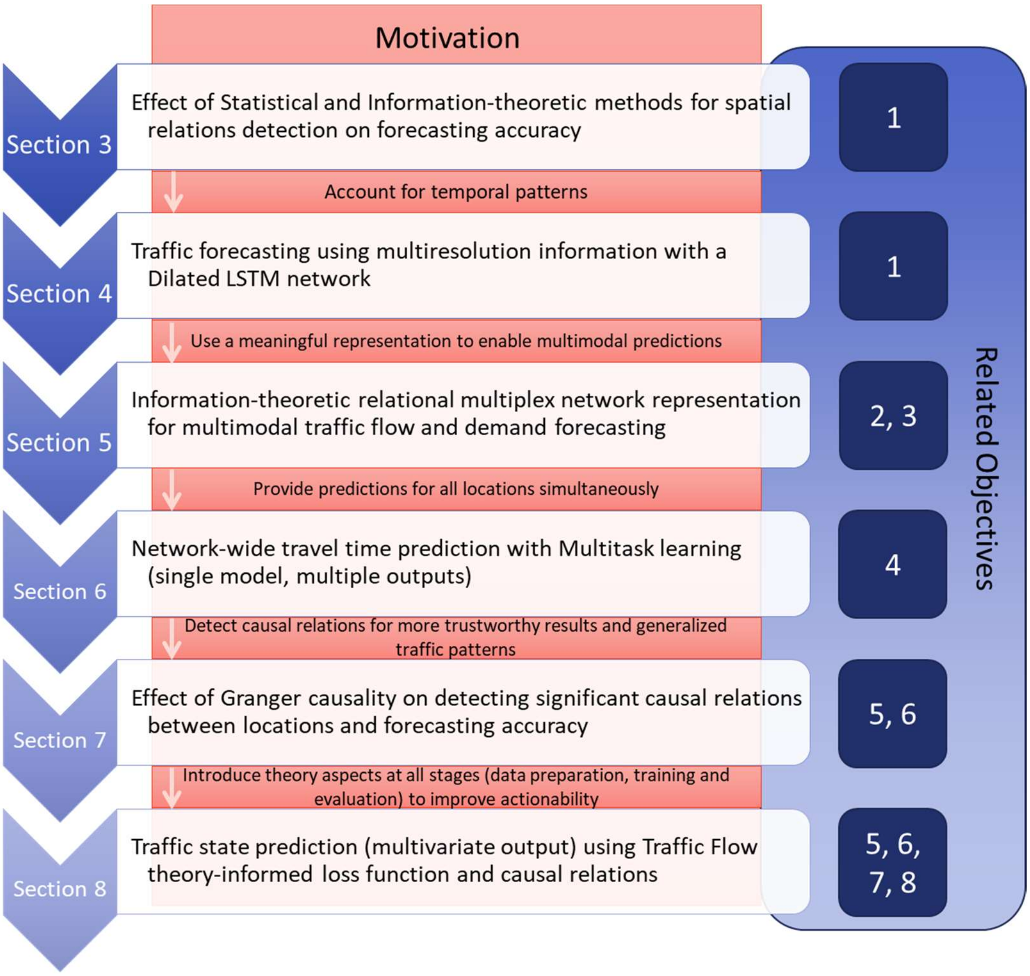



Figure 1. Structure overview and relation with objectives

# 2 TOWARDS ACTIONABLE TRAFFIC FORECASTING: CURRENT PRACTICE AND FUTURE CHALLENGES

In this section, several challenges that are related to a Deep Learning model's actionability in predictive traffic management are discussed, according to recent literature. Moreover, the research questions that the present dissertation attempts to answer are presented.

## 2.1 Models' Taxonomy

Traffic forecasting has been a very popular research area for, at least, the last four decades. Departing from the first naïve forecasting models, such as Historical Average that had very low complexity, but poor forecasting performance in most cases, a large variety of methodologies have been exploited to model traffic time series, spanning from simpler statistical and parametric models to Deep Learning, which is state-of-the-art right now. The different methodologies have been analyzed in detail in the numerous review papers that have been published during the last few years and, thus, presenting the specifications of each method is out of the scope of this work. Instead, in this section, the most prevailing taxonomies of the methods are presented and, as far as it concerns Deep Learning, a critical view of the main characteristics of the most popular architectures is provided, as well as some comments on their applicability.

The first separation that is found in recent literature is between parametric and non-parametric models (Vlahogianni et al., 2004; Polson & Sokolov, 2017; Do, Taherifar, et al., 2019). The main difference between the two approaches is that in the non-parametric, there is typically a very large number of parameters, with no physical interpretation, while in parametric methods most of the parameters have some physical interpretation.

More specifically, the parametric methods rely on assumptions about the population (data) distribution and their mathematical formulation includes parameters that should be a priori determined. The determination of the parameters requires good domain knowledge, as well as intuition about the mathematical foundations of the method and modeling experience. For traffic forecasting problems, such a model would exploit traffic flow theory and the inputs/parameters may include the road network's demand, route choice, road capacity, etc. Some examples of this class of models are the family of Linear Models (linear and logistic regression), Naïve Bayes, Kalman filtering methods, exponential smoothing, Autoregressive Integrated Moving Average (ARIMA) and others.

Non-parametric methods rely on a flexible mathematical formulation and structure and have adjustable parameters. The exact structure and parameters' values are learned from the data, during the fitting of the model and have the potential to simulate more complex and non-linear relations. Neural Networks and some Machine Learning models (e.g., Support Vector Machines) belong to this category.

The methods that have been applied to traffic forecasting can also be divided into Statistical, Machine Learning and Deep Learning. Correspondingly to the order they are mentioned, there has been a trend towards higher accuracy and lower interpretability between the three categories.

Statistical models are also parametric models. They have widely accepted and strong mathematical foundations, that allow getting insights into the mechanisms creating the estimations (Karlaftis & Vlahogianni, 2010). On the other hand, their input data are subjected to various assumptions, such as linearity and stationarity, which are over-simplifications for traffic data (Ryu et al., 2021). Although they offer a straightforward interpretation of the results and require less data for their fitting, their performance is relatively low because of the above assumptions, which are prerequisites for their implementation. Thus, such models can be exploited to understand the relations between the input features and the predictions (Karlaftis & Vlahogianni, 2010; Tedjopurnomo et al., 2020). In other words, statistical models can be used to explain a phenomenon and not to precisely predict the future.

Statistical models tend to disappear and are considered deprecated in recent literature, despite the advantages of the strong mathematical foundation and high interpretability, because they cannot handle the large amounts of data, that incorporate high complexity and dimensionality (especially for network-level prediction), and the complex and strongly non-linear spatiotemporal relations effectively and provide predictions of satisfying accuracy. However, models of this category are those that are more often adopted in real-world conditions and, especially, the most naïve ones, mainly due to issues that are discussed in the next section (J. J. Q. Yu et al., 2021).

Unlike statistical methods, Machine Learning models are focused on making accurate predictions. They can deal with complex and non-linear relations more effectively, compared with statistical models, and that is the main reason they reach better performance, in general (Ermagun & Levinson, 2018). The user interference is also reduced, as the models of this category adjust (learn) the values of their hyperparameters based on the input data. On the other hand, due to their shallow architecture and the manual feature selection, their prediction performance remains limited, especially in cases of large input spaces (Ma et al., 2017). Most Machine Learning models are non-parametric.

Deep Learning is, actually, a sub-category of Machine Learning models, which are based on the concepts of Neural Networks and multi-layer architecture. However, due to their popularity and different characteristics compared to the rest of Machine Learning techniques, they are usually mentioned as a separate category. They are, as well, non-parametric and with Machine Learning models, they are also often referred to as data-driven models. As already mentioned in a previous section, Deep Learning models reach high prediction accuracy, thanks to their capability of handling vast amounts of data and simulating complex, non-linear relations, but they have the disadvantages of high computation power and data requirements, as well as limited interpretability (Ma et al., 2017; Manibardo et al., 2021).

Deep Learning models belong to the broader family of Machine learning and all related techniques are based on the Artificial Neural Network (ANN) paradigm, which was first proposed by (McCulloch & Pitts, 1943). An ANN is an abstract mathematical model that can be defined as a complex linear system based on a collection of connected units or nodes called artificial neurons, which loosely resemble the neurons in a biological brain. ANNs have a multi-layer architecture, which is used to progressively extract higher-level features from the input data, with each layer consisting of Neurons (or Perceptrons) (LeCun et al., 2015). An ANN does not require any empirical formulas to obtain the inherent relationship of the data from the dataset. By observing many input and output samples, the ANN model automatically adjusts and establishes the input and output map model during its training phase.

Although quite an old concept, ANNs were not given a lot of attention when first presented, at least in the field of traffic forecasting, due to shortcomings that could not be overcome 20 or 30 years ago, such as the high computational power and data requirements. The most vital step towards the rise of Deep Learning has been the evolution of sensing technology and the exploitation of novel data-collection tools, such as smartphones, which made the collection of an unprecedented amount of traffic and mobility data possible (Z. Liu et al., 2018). Moreover, other aspects that supported the development and implementation of Deep Learning models are (Geron, 2017a; Zhao et al., 2017):

- The unprecedented development of telecommunication and computing systems, which as well contributed to the collection of big amounts of data, their manipulation and storage. Also, the tremendous increase in computing power and the faster hardware (e.g., GPU) make the efficient (in a reasonable amount of time) training of Deep Learning models possible.
- The training algorithms have been, at least slightly, improved, making the convergence of the algorithms easier. Also, some theoretical limitations of ANNs have not been proven to be so serious in practice, such as the training algorithms stucking in local optima.
- The development of user-friendly software and libraries (e.g. Tensorflow, Keras), that do not require high levels of expertise to implement very complex models.
- Research on Deep Learning techniques and applications has attracted significant funding.

The numerous variations of Deep Learning algorithms have been applied to various scientific fields, with significantly successful outcomes. Deep Learning has been an obvious direction for researchers and practitioners in traffic forecasting, too, during the last decades. The intense work and high research interest have constituted traffic forecasting as a domain for testing, benchmarking, and comparing new modeling structures and techniques. Deep Learning methods are commonly acknowledged as the methods with the most accurate performance in terms of prediction errors, compared to previous approaches, which is also mentioned as their main advantage (Y. Wang et al., 2019). The main reason behind that is their potential to approximate almost any function, regardless of its degree of non-linearity and model underlying, complex temporal and spatial relations, such as those in network-wide traffic data (Ye et al., 2022).

Additionally, Deep Learning models can extract features from large-scale raw data automatically and, thus, the difficult tasks of feature engineering and selection, which require much effort and, especially, domain knowledge, are not always necessary (Z. Liu et al., 2018; Ye et al., 2022). However, the aforementioned processes can improve the model's performance, facilitate and speed up the learning process and, most importantly, reduce the model complexity and the computational resources required, so it is recommended not to disregard its importance.

All variations of the ANN model have been widely and successfully used in traffic forecasting. Among them, the most relevant categories are the Recurrent Neural Networks (RNN), the Convolutional Neural Networks (CNN) and the Graph Convolutional Neural Networks

(GCNN). Each category includes several alternative architectures, while there are also approaches that include aspects (layers) from more than one category. The most important specifications of each category are described below.

RNN are developed to model time series and other sequence data, such as text. In contrast to other structures that consider each input as independent, they can perceive the existence of a temporal relationship between the input data, exploiting the memory (or internal state) unit that is included in their architecture (Geron, 2017a). Long Short-Term Memory (LSTM) networks are the most popular RNN architecture. Their structure is more complex than simple RNN's, as it allows the interaction between the Internal state of the current and all previous timesteps, while in RNN only with the exact previous. LSTM's memory unit can be anticipated as a pipe, connecting all inputs, emphasizing those of the previous that relate the most with the current and weakening those that don't (Gulli & Pal, 2017). LSTM are considered one of the most powerful methodologies for time series prediction and are used a lot in traffic flow forecasting (Zhao et al., 2017; Bogaerts et al., 2020; Fafoutellis et al., 2020).

One drawback of RNN (and, especially, LSTM) is that when processing long sequences, the effect of the first (oldest) timesteps may decline as the training phase proceeds to later ones, also known as the "vanishing gradient" issue (X. Yin et al., 2021a). Another equally important issue when dealing with traffic data is that RNN, although they are suitable for time series data, they cannot capture the spatial relations among them (Boukerche & Wang, 2020).

On the other hand, Convolution Neural Networks (CNN), have been applied to traffic forecasting problems in order to exploit spatial relations (X. Yin et al., 2021a). CNN's main application areas are image recognition and computer vision. The input data should be modeled as a 2- (or 3-) dimensional grid before being fed to the CNN or, equivalently, as an image. The most significant drawback of this method is that the CNN operation is performed in neighboring grids, i.e. road sections that are close to each other are processed together, and, thus, relations with further sections are ignored, which is not desirable for traffic forecasting (Jiang & Luo, 2021). CNN are often combined with RNN, so that both spatial and temporal correlations are deployed (Dai et al., 2019; Ma et al., 2017).

Moreover, CNN are limited to data and spatial relations on the Euclidean domain. However, road networks can be intuitively represented as a graph, where, e.g., the road sections can be thought of as the edges and intersections as the nodes (Jiang & Luo, 2021). Consequently, Graph Convolutional Neural Networks (GCNN) are more suitable for traffic forecasting, as they can capture spatial dependencies in non-Euclidean structures, namely graphs. The input of this model is the adjacency matrix of the graph, which reflects the connectivity or other relations of the nodes, e.g. statistical correlation of traffic measurements, and, additionally, a set of features for each node (Cui, Ke, Pu, Ma, et al., 2020; Leiser & Yildirimoglu, 2021; J. J. Q. Yu et al., 2021). As traffic forecasting is a time series prediction problem, a variation of GCNN, the Spatial-Temporal Graph Neural Networks (STGNN), is also very often employed. In STGNN, the features of each node change dynamically over time. However, correlations between different nodes at different time steps are not explicitly modeled (Zheng et al., 2021).

There are many alternatives to how to represent a road network and the traffic data in a graph. For example, as both nodes and edges can have their own attributes, a road section or a detector (point of the road network where there are traffic measurements) can be correspondingly represented as a node or as an edge. These attributes should include one or more traffic state indicators, such as traffic volume, speed, etc. The alternative representations are described in detail in a following section (Section 2.3).

## 2.2 Challenge 1: Limitations of Deep Learning

There is currently a disconnection between Deep Learning and traffic management mostly due to the low interpretability of the models and, also, the lack of a proper and straightforward way of identifying the spatial and temporal relations (Lana et al., 2018; Manibardo et al., 2021). In addition, the intensive data and computational requirements significantly limit the probability of using the models for network-wide predictions; although predicting the traffic conditions of a single road section is technically feasible, developing models for large road networks has not been adequately explored. Moreover, most of the recent studies focus on one-step prediction, while each step usually corresponds to 5 to 15 minutes. The latter period is usually relatively short for the purposes of traffic management, planning, and decision-making. However, the research on multistep forecasting is considerably lower, while the prediction accuracy deteriorates rapidly with the increase in the number of steps.

So, despite the unquestionable effectiveness of Deep Learning in predicting very accurately future traffic conditions (at least for a limited number of road sections), compared to previous approaches, several obstacles have been identified in recent literature, concerning the applicability of Deep Learning in network-wide traffic management. The most important of them are presented below.

**Data requirements**: As has been already mentioned, Deep Learning models require vast amounts of data, that cover all possible traffic states, in order to train and converge (X. Yin et al., 2021a). The amount of data needed increases with the model's complexity, as well as the input space's size (dimensionality), i.e., the number of road sections there are traffic measurements from. Such data are rarely available for all the road sections of a network and, especially, for an adequately long time in the past.

**Long training time**: Even with today's technology and powerful hardware, training a complex Deep Learning model for network-wide predictions, using the data volume described above, is a very time-consuming task that may last for hours or even days (Do, Taherifar, et al., 2019). Furthermore, the model should be re-calibrated and its parameters need to be updated regularly and as soon as new data become available, which is also a time-consuming and demanding process. On the other hand, Statistical and simpler Machine Learning models have a simpler and shallow architecture that does not include so many learnable parameters and, thus, their training time is extremely lower (up to some hundreds of times lower), facilitating their exploitation in traffic management (Boukerche & Wang, 2020).

**Low interpretability**: Due to their "black-box" nature, as it is often referred to, Deep Learning models lack the simple and straightforward interpretability of other models, such as Multinomial Regression (Do, Taherifar, et al., 2019; Karlaftis & Vlahogianni, 2010). With Deep Learning, it is difficult to understand the reasoning behind the predictions provided by the model, i.e. the type and magnitude of the effect of each input feature, which limits the exploitation of the model in policy-making (Y. Wang et al., 2019). Understanding the model's mechanisms of producing its output, would also reveal information about the spatiotemporal dynamics of the road network, which is vital for traffic management. Moreover, interpretability is essential in order to justify the decision-making process and increase the trustworthiness of the specific model, extract new scientific knowledge regarding the network's mechanics, but also to find ways to improve its performance and its transferability (Barredo Arrieta et al., 2020).

**Hyperparameter selection**: As there is no recognized methodology for determining the optimal architecture, i.e., number of hidden layers, values of hyperparameters to use, etc., of a Deep Learning model, usually it is determined by experience or by following a trial-and-error approach. Alternatively, one may apply a grid search procedure, the required time for which increases exponentially with the complexity of the model (K. Lee et al., 2021). Moreover, given that a relatively simple structure does not ensure the good performance of the model, especially when dealing with a big input space, using a shallower structure is not a possible direction.

**Computational power requirements**: Along with all the above, one should not neglect the high computational power required to train and maintain the model and to store and use the corresponding datasets (Boukerche & Wang, 2020; Manibardo et al., 2021). On many occasions, traffic management authorities do not own modern and so powerful computers and servers and, thus, must turn to other models.

**Transferability**: A trained model is fitted and optimized for predicting the traffic conditions at specific points or road sections. Therefore, it cannot provide reliable outcomes for any other point, not even of the same network (J. Li et al., 2022). Given the time and effort required to develop a complex Deep Learning model, the limited transferability and generalizability of the methods is a major drawback (Barredo Arrieta et al., 2020). Model transferability in traffic forecasting still remains an under-researched topic (X. Yin et al., 2021a).

Except for the above issues, one of the fundamental principles of Machine Learning is the assumption that the data are independent and identically distributed (i.i.d.) to achieve statistically significant results and good forecasting accuracy. The latter implies that the test set, as well as other future unseen input, should follow the same distribution as the training set, or else the i.i.d. assumption will be violated, and the model would perform poorly (Kaddour et al., 2022). The above is almost impossible for a large urban road network with hundreds of sections (J. Li et al., 2022). In practice, when abnormal conditions occur, e.g., due to a road accident, and new congestion patterns appear, the i.i.d. assumption gets violated, at least for some of the road sections. If these models are exclusively relying on data and not on relations from traffic flow theory, they are restricted to specific regions and periods and is most likely that they will get outdated and their performance will drop significantly, as soon as the i.i.d. assumption no longer holds.

### *2.2.1 Challenges Using Deep Learning for Traffic Forecasting*

From the discussion provided in the previous chapter, it becomes clear that several open issues should be mitigated to enable the deployment of Deep Learning in traffic forecasting, at a larger scale. Several review papers have been working on extracting future research paths from the abundant existing literature on traffic forecasting. These are summarized in Table 1 and critically discussed in the following sections.

**Table 1. Identified challenges in traffic forecasting literature.**

| *Research Work* | *Identified Challenges* |
|---|---|
| **Fafoutellis & Vlahogianni (2023b)** | Need for network-wide predictions<br>Consider multiple modes (multimodal predictions)<br>Causality, explainability and spatiotemporal analysis |
| **Manibardo et al. (2021)** | Focus on other properties of the model than performance (faster training time, requirement of less computational resources and easier interpretability)<br>Consider explainability and actionability when developing forecasting models<br>Generation of "standard" data testbeds for model testing |
| **Boukerche & Wang (2020)** | Addressing issues of data availability and scarcity<br>Computational power requirements<br>Efficient training and retraining with the rate of arrival of new data |
| **X. Yin et al. (2021a)** | Lack of sufficient available data at a city level<br>Focus on efficient and lightweight models for real-time prediction applications.<br>Design of interpretable and transparent Deep Learning models. |
| **Tedjopurnomo et al. (2020)** | High requirements of data and computing resources<br>Develop responsive prediction schemes (e.g., to unexpected changes in traffic) |
| **Y. Wang et al. (2019)** | High data and computing resources requirements<br>Develop methods to address the limited capability of interpreting the outcomes<br>Exploitation of different data sources |
| **Jiang & Luo (2021)** | Data of insufficient quality and quantity<br>Include external factors (e.g., weather conditions)<br>Multitask prediction models |
| **Do, Taherifar, et al. (2019)** | Addressing the high complexity of road network topologies in relation to the deployment of Deep Learning structures<br>Focus on theory-driven training (non-recurrent events, traffic flow fundamental diagrams) |
| **Pavlyuk (2019)** | Extract spatiotemporal features to increase the models' performance, interpretability and efficiency |
| **K. Lee et al. (2021)** | Focus on selecting the proper representations of the network and spatiotemporal dependencies |
| **Lana et al. (2018)** | Consider the context of forecasting, from point based to network-based predictions.<br>Extend the prediction horizon, towards long-term estimation approaches that boost the actionability of predictive models<br>Include different data sources to prediction, applying data fusion techniques.<br>Big Data implementations and distributed computing. |

Looking closer at the recent literature, researchers take, almost exclusively, the models' accuracy into account when comparing different approaches and deciding on the most effective, disregarding issues such as computational efficiency and ignoring shortcomings such as limited interpretability and explanatory power. The latter leads to implementing Deep Learning models as "black boxes" (Karlaftis & Vlahogianni, 2010). Therefore, a balance should be found between the complexity of the model that is developed and the time and resources it requires, if they are going to be applied in real-world conditions (Boukerche & Wang, 2020). In this chapter, the future challenges of traffic forecasting are presented, as discussed in the most significant recent review papers, concerning the implementation of Deep Learning techniques in traffic forecasting.

In a recent study, Manibardo et al. (2021) showed that the performance of complex Deep Learning structures is only slightly better (and sometimes the same or worse) than that of shallow methods and ensembles, after following a well-established benchmarking process, using different traffic datasets on several predicting horizons. They state that such slight performance improvements do not translate into practical advantages in real-world applications. Thus, the authors suggest that simpler methods should be as well considered by researchers and practitioners, as they offer several practical advantages, which are translated into increased actionability of the model. As a future direction, the authors recommend that model actionability should be the goal for works in the field, which has not exclusively to do with the precision of the forecasts. In traffic management, the interpretation of the outcome, to make informed decisions, is considered more important than slightly better performance. The authors also highlight the need for a benchmarking datasets repository.

Boukerche & Wang (2020) provide a comparison of the performance of different Machine Learning and Deep Learning models in traffic prediction tasks, as observed in recent literature. As far as it concerns the applicability of this kind of models in traffic management, the authors focus on the prompt availability of data of high temporal resolution and computers of high computational power, in order to deal with the requirements of a complex, network-wide prediction model. X. Yin et al., (2021a) also provide a benchmarking of different prediction models, spanning from classical statistical methods to state-of-the-art Deep Learning, using several public datasets. Furthermore, they proceed to propose the following directions for future research: As there are not sufficiently big datasets available in most cities, researchers should focus on transfer learning, but also on designing more interpretable and transparent models, which are important for management purposes. Moreover, an efficient and lightweight model is vital for real-time applications, as well as the use of multi-source data.

Tedjopurnomo et al. (2020) have also highlighted the issues of the large data requirements, long and computationally expensive training and difficulty to interpret, due to the large number of input features and the trainable parameter values. The authors proposed the development of prediction schemes that are responsive (e.g., to unexpected changes in traffic) by incorporating external data (weather, accidents), highlighting existing spatiotemporal correlations, and can be efficiently updated and operate in real-time. Y. Wang et al. (2019) mention the high accuracy of the predictions as the main advantage of Deep Learning methods. However, they as well highlight the high requirements of data and computing resources, as well as the limited capability of interpreting the outcomes. Furthermore, the authors recommend the exploitation of different data sources to improve efficiency and applicability.

In (Jiang & Luo, 2021), the authors review Graph Neural Networks application in traffic forecasting, along with the most widespread traffic data sources. Moreover, they mention the most important challenges of traffic forecasting, in terms of data quality and availability. Firstly, crowdsourced and GPS data are sometimes of doubtful quality and suffer from sparsity, high missing data ratios and noise. Also, external factors should be taken into account, such as weather conditions and special events, when modeling network-wide traffic conditions. Finally, the authors recommend focusing on multi-task prediction (e.g., predicting traffic conditions at multiple road sections with the same model.

In a recent work, Do, Taherifar, et al. (2019) conclude that the high complexity of road network topologies does not favor the deployment of Deep Learning structures. They propose that the use of effective representations of the network would increase the chance of the models being used in real-time applications. Moreover, training the models at non-recurrent events and incorporating basic theoretical knowledge, such as the fundamental relationships, is also expected to increase performance and interpretability. Pavlyuk (2019) reviews the methods of spatiotemporal features extraction and their exploitation in Machine Learning and Deep Learning models. The author highlights the importance of the above features in increasing the models' performance, interpretability and efficiency, as well. K. Lee et al. (2021) reviewed the different representations of road networks and spatial-temporal dependencies, as well as Deep Learning methods. The authors conclude that accurate representation is equally important with the selection of the appropriate modeling technique.

Finally, Lana et al. (2018) discuss some considerations regarding urban and network-wide predictions, as well as hybrid models and evaluation metrics and comparison. As the most important future challenges, they mention going to a larger, network scale and longer predicting horizons and using techniques for efficient handling of Big Data, such as parallel computing. Moreover, they highlight the importance of an efficient input representation and data fusion techniques for incorporating heterogeneous data.

## 2.3 Challenge 2: Road Network and Input Space Representation

In theory, Deep Neural Networks can simulate any relation between the input data, regardless of its complexity or the size of the input space. In practice, however, the performance of the model heavily depends on the representation of the input data, and the amount of supplementary information they provide about the spatial relations (Manibardo et al., 2021). This valuable information enhances the performance of the model, as well as its interpretability. For example, by using an image representation, the proximity of the locations is implied, as well as the geometry of the road network. In traffic forecasting, an accurate and meaningful representation can also reduce uncertainty and is considered equally important with the modeling technique that is used (Barredo Arrieta et al., 2020; K. Lee et al., 2021).

However, in numerous recent studies, a data-intensive approach is followed by feeding the model with all the available raw information, without any prior analysis or feature selection, which adds complexity to the model and increases the dimensionality of the input space, undermining the model's performance. The complex Deep Learning structure required to successfully handle such input spaces will eventually lack the properties of actionability and interpretability and would be very demanding in terms of training time and computational resources. In these cases, the complex Deep Learning structure may achieve good prediction accuracy, but only by just implying the existence of correlation, and disregarding any causal features that could be useful in traffic management.

In recent years, a large variety of road network and input data representations have been proposed in the research area of traffic forecasting, depending on the task of the prediction (single point or network level) and the kind of input that is compatible with the corresponding prediction model. In general, they can be classified into three categories: stacked vectors, grid- (or image-) based and graph.

*Stacked Vector*

The first class of representations is the stacked vector, where the road network data are organized into a single vector. More precisely, the time series of the measurements of each location (e.g., loop detector, road section, intersection, region, etc.), which can already be considered as vectors, are simply stacked in a vector of vectors, which can also be thought of as a two-dimensional matrix of dimensions (number of locations) x (number of timesteps). This representation still remains the most popular one and was already proposed by the initial network-wide traffic forecasting works (K. Lee et al., 2021).

There is no predefined way of stacking the vectors of the input data into a single vector, but it depends on the researcher's or practitioner's intuition, domain knowledge or personal preference. The order in which each location's vector appears plays an important role, especially in cases where a model that takes account of locality and/or proximity is exploited, such as a Convolutional Neural Network (Modi et al., 2022). Thus, although this kind of representation is simple, the user should not disregard organizing the input data in a suitable way. When the entire road network and the corresponding data locations have a simple geometry, the order in which the input data are organized is rather straightforward. For example, when a circular road network or a corridor is represented (Figure 2), the vectors of each location can be stacked in a clockwise or connection/proximity order, respectively. According to Figure 2, the input data would be represented as follows:

$$X = \{V(L_1), V(\,L_2), \ldots, V(L_6)\} \qquad (1)$$

where $L_i$ is the location and $V(L_i)$ is the value of the traffic variable considered, e.g., mean speed or flow.


**Figure 2. Examples of road networks that can be represented as a stacked vector.**

The main disadvantage of this method is that there is no effective way to represent even slightly more complex road network geometries. For example, if an additional location that does not follow the same pattern is added to the above examples, as shown in Figure 3, the stacked vector representation is not easily applied. Moreover, when a more complex network with a large number of locations should be modeled, it is unclear which locations are close or adjacent and in which order they should appear. In such cases, the location vectors can also be stacked randomly and, thus, the spatial relations between the locations are not at all provided to the prediction model. In order to mitigate the effect of this issue, it is recommended to follow a feature selection strategy to reduce the dimensionality of the input space and utilize only the locations that are most correlated with the target one. A very naïve and straightforward approach for feature selection is by calculating a correlation metric (e.g. Pearson's correlation) between the target location and all other locations and including in the input data only a subset of the most correlated locations with the target location (Ermagun & Levinson, 2018). Of course, more complex approaches, taking into account proximity or other properties can also be followed (Cai et al., 2015; Ryu et al., 2018).

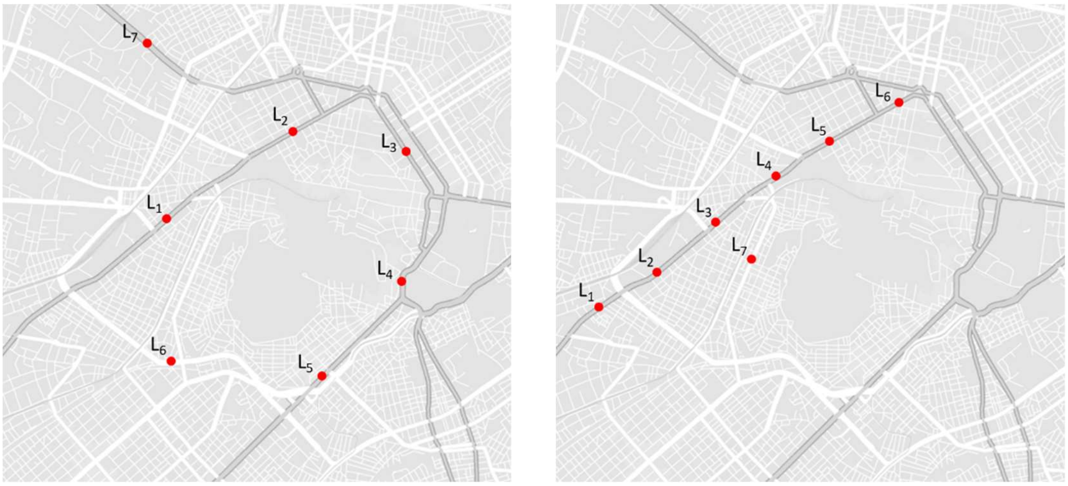


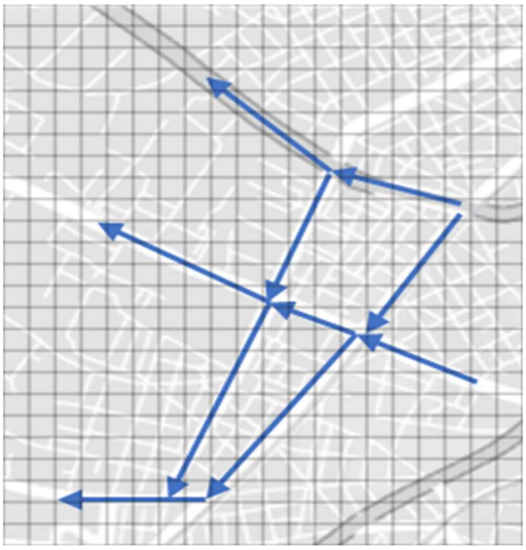

**Figure 3. Examples of low efficiency of the stacked vector representation.**

Stacked vector representation has been very popular, due to its simplicity and flexibility (Ye et al., 2022). However, it can only pass a limited amount of information about the spatial dependencies and the road network's geometry to the prediction model. Finally, it can be used for any road network and is compatible with all prediction models, although it is not very suitable for road networks with complex geometry and relationships between the locations.

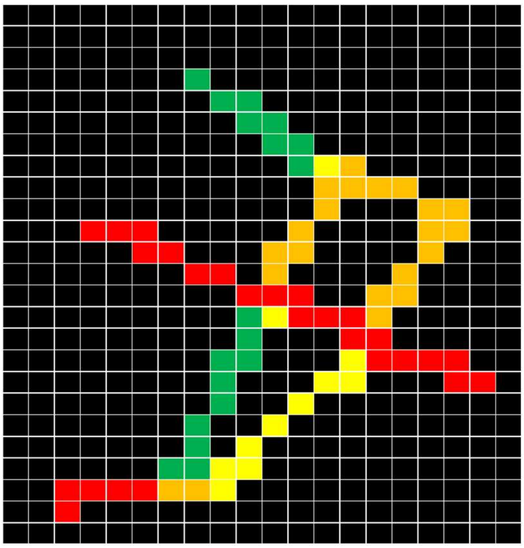
*Grid or Image*

The second representation method is the grid or image-like representation. The data are organized into a two-dimensional grid, which is a very intuitive choice for 2-dimensional, Euclidean, spatial data from locations with latitude and longitude. More specifically, a square grid with the size of the road network is defined and a value that represents the traffic conditions inside it is assigned to each square of the grid, which is exactly proportional to a greyscale image. Thus, the input data can directly be passed to the prediction model, without any modification. As Convolutional Neural Networks (CNN) is the most proper technique to handle image data, any hybrid Deep Learning model is compatible with the grid representation, as long as it includes a CNN layer. Other techniques, e.g. Statistical and Machine Learning, cannot be used in this case (K. Lee et al., 2021).

An example of grid representation is given in Figure 4. The value of each pixel (square) of the grid is equal to the value of the respective traffic variable measured at the road section that lies inside it, and is sometimes represented by a color scale. Moreover, one may also witness in Figure 4 the first drawback of this kind of representation, which is its inefficiency. First, the majority of the pixels do not usually match any road section, so their values are zero (black color). Consequently, the model is fed with a larger input space, which increases the complexity of the computations needed and the time they require, while it only contains a relatively low amount of useful information. The latter issue is even more noteworthy when the input data consist of measurements from single points (e.g. loop detectors) that occupy only one pixel and not road sections as in the example of Figure 4.



*(a) Example road network*



*(b) Grid representation*

**Figure 4. Grid representation of indicative road network**

Second, as is also obvious in Figure 4, each road section of the network is contained in more than one squares, but the corresponding traffic variable value is the same along its entire length. As a result, the same value is passed multiple times to the model (all pixels have the same value), which increases the input's size, without increasing the amount of useful information correspondingly.

Another issue that should be considered is the size of the pixels or, equivalently, the resolution of the grid. A higher resolution (more pixels with smaller dimensions) may provide more detailed information about the traffic conditions, but it also intensifies the two drawbacks described earlier; the smaller the pixels, the more possible it is that a road section occupies more pixels and the more pixels will remain empty. On the other hand, when a lower resolution is selected, a significant number of pixels may contain two or more different locations, e.g., road sections, as shown in Figure 5. In this case, as each pixel can have only one representative value, the average of all road sections may be calculated. The latter is usually undesirable, as the two or more sections may not have a very significant correlation (e.g., when heading in opposite directions) and, by this aggregation, a significant amount of information may be lost.

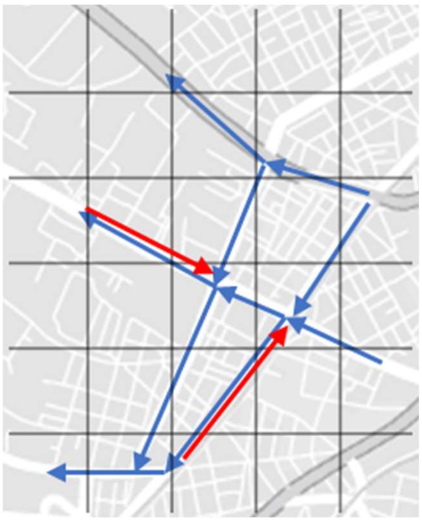


**Figure 5. Road network represented as low-resolution grid.**

Despite the aforementioned issues, the image or grid representation has been very popular so far because it very accurately depicts the road network's geometry and the relations and proximity of the locations. However, its dependence on the method of CNN brings with it all its drawbacks, the most important of which being that they only take into account relationships between pixels that are close to each other in the Euclidean domain (local dependencies) which may not be sufficient for road networks, where dependencies between distant locations are often stronger (Ermagun & Levinson, 2018). Finally, the grid representation, although it can be used to extract spatial information, does not express all the properties of a road network, which is physically organized as a graph (Ye et al., 2022).

*Graph*

The third main class of representations is the graph. Compared to images, graphs can be used to express more complex relations between the input data from different locations of the road network, which cannot be explained only by (Euclidean) proximity information, stemming from the connectivity of the sections of the road network, the impact of intersections and traffic lights and traffic/congestion patterns of distant locations.

In general, a graph is a mathematical structure that is used to model pairwise relationships between different objects and can be represented as $G = (V, E)$, where $V = \{v_1, v_2, \dots, v_n\}$ is the set of vertices or nodes and $E$ is the set of edges, consisting of pairs of nodes that are connected to each other, $(v_i, v_j)$, and $1 \leqslant i, j \leqslant n$. An edge $(v_i, v_j)$ may be directed, i.e. connecting the nodes asymmetrically with direction from $v_i$ to $v_j$, or undirected, i.e. connecting the two nodes symmetrically in both directions. A graph consisting exclusively of undirected edges is also called undirected; otherwise, it is called directed.

The most efficient way of representing a graph is with an adjacency matrix $A \in R^{|V| \times |V|}$. The simplest definition of the adjacency matrix is the following: $A = (a_{ij})$, where $a_{ij} = 1$, if $(v_i, v_j)$

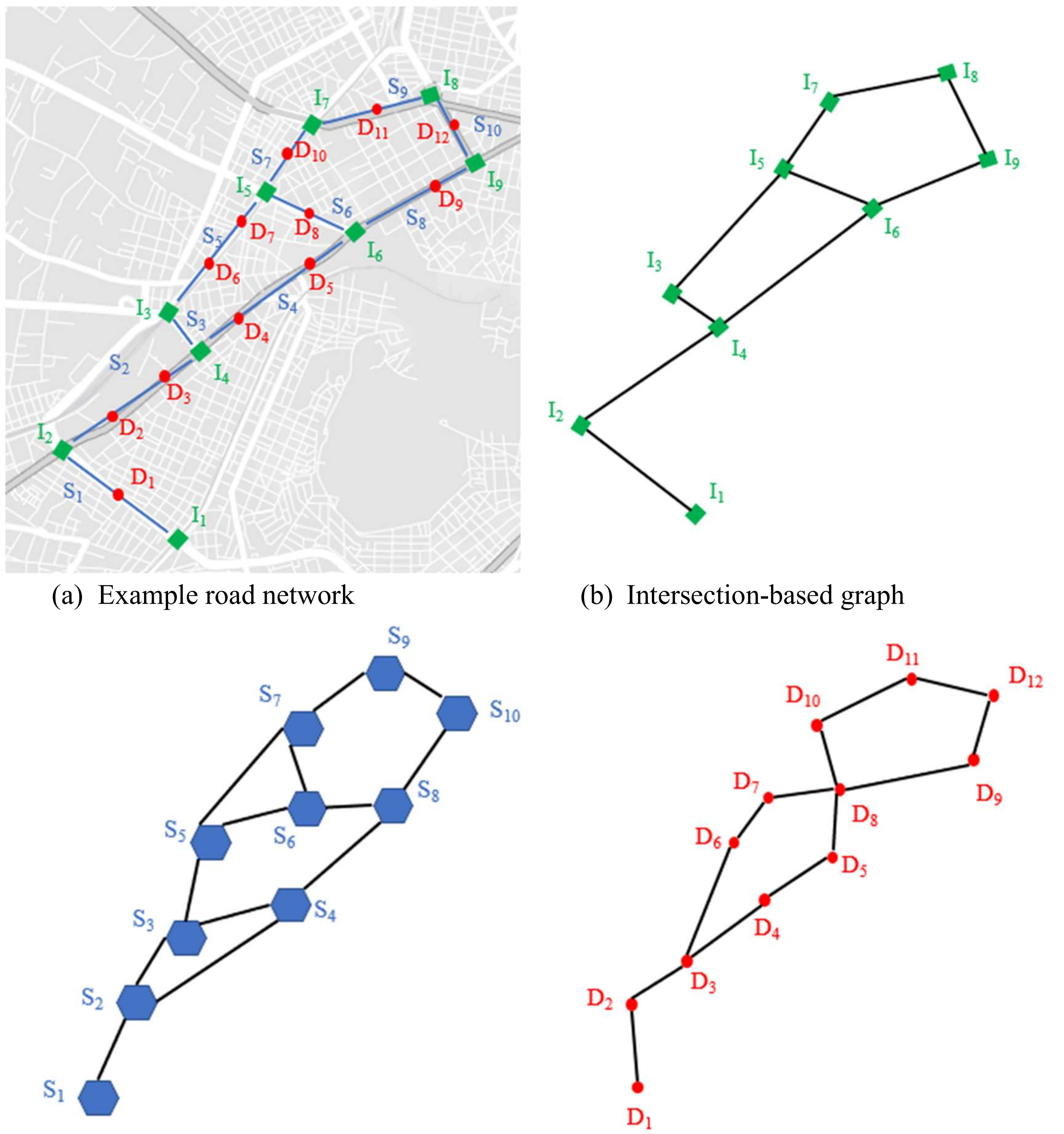
is an edge of *G* and *0* otherwise. Moreover, a weight can be assigned to each edge, representing usually the strength of the connections or a cost, depending on the specific occasion. In this case, the values of the elements of the adjacency matrix are equal to the value of the corresponding weight.

Graph Convolutional Neural Networks (GCNN), as well as their variation Spatio-Temporal GCNN, are the only modeling technique that can handle input data organized in a graph. Except for the adjacency matrix, the input also includes vectors of features for each node, e.g., the time series of the traffic variables measurements.

In traffic forecasting literature, several approaches to how exactly to represent the road network as a graph have been proposed. Firstly, depending on the type of locations that are exploited and data availability, the nodes of the graph can be defined as the intersections of the road network (which are the "physical" nodes as well), loop detectors that are installed at the network or its road sections (Jiang & Luo, 2021). Furthermore, the connections between them may be their physical ones or connections that express some kind of similarity or statistical relationship between the nodes, which can also be weighted (Ye et al., 2022).

Figure 6 displays various ways to depict the same road network. Among them, the representation where intersections are utilized as nodes (Figure 6(b)), closely resembles the actual network visually. In this representation, the adjacency matrix includes the information about the traffic variable measurements at the road sections (weights of the adjacency matrix), which are the edges of the graph. However, this representation is not ideal as the most important input, which is also the output of the model, i.e., traffic conditions, should typically correspond to the nodes and not the edges of the graph, according to the architecture of the Graph Convolutional Networks.

The two other representations are quite similar to each other. In the first one, each node of the graph corresponds to a road section of the network (Figure 6(c)), while in the second (Figure 6(d)) to the exact point the measurements refer to (where a loop detector or other sensor is installed). Often the two approaches may result in the same graph, except for the case that two or more detectors are installed at the same road section, which is very possible for long road sections. In this case, for the road section graph, an average value of all the corresponding detectors' measurements should be calculated as representative for the section, which may not be desirable, as it decreases the level of detail that the input data have.




(a) Example road network

(b) Intersection-based graph

(c) Road section-based graph

(d) Detector (sensor)-based graph

**Figure 6. Examples of graph representations considering different node types.**

The most important aspect for defining the spatial relations between the locations of the road network, which are key to enhancing the prediction model's performance, is the adjacency matrix, which contains their pairwise relationships (connectivity or traffic conditions pattern similarity) (Ye et al., 2022). In recent literature, a variety of approaches have been proposed to the definition of the adjacency matrix, which are presented below:

<u>Physical connectivity matrix</u>

This type of adjacency matrix reflects the actual connectivity of the road network, e.g., consecutive road sections. This approach is quite intuitive and the values of the elements of the matrix are 1 if the corresponding nodes are connected and 0 otherwise. In the examples of Figure 6, the connections are determined based on the connectivity of the nodes. Although this representation is simple and suitable for smaller networks, in complex networks it is not always clear which nodes are directly connected, and, especially, when only a relatively small part of the network is covered by sensors. An example of the latter is given in Figure 7; the road sections with loop detectors are far apart and cannot be considered adjacent. In this case, one of the following approaches should be considered.

Distance-based matrix

In this approach, two nodes the distance between which is below a threshold that is decided by the model developer are considered as connected (adjacency matrix value equals 1). Alternatively, the order of neighboring can be used and compared to a threshold: if a node $v_i$ is reachable from node $v_j$ with $m$ steps, based on the natural connectivity of the nodes, or, equivalently, $v_i$ is an $m$-order neighbor of $v_j$, the two nodes are adjacent. In order to provide the model with more detailed information, weights that are equal to the distance or the neighbor order, respectively, can be assigned to the corresponding edges.

Similarity/Correlation-based matrix

The two above representations suffer from two main disadvantages: first they are static, i.e. remain the same over time and during different periods, and, secondly, they take into account only local dependencies, e.g. they consider that the traffic conditions at a location are only related with and affected exclusively by nearby locations, which is not accurate (Jiang & Luo, 2021; Zheng et al., 2021). On the other hand, when using a correlation-based matrix, a statistical metric of the similarity between the time series (e.g. Pearson or Spearman correlation) of the traffic conditions of each pair of nodes or a similar metric from Information theory (e.g. Mutual Information) is estimated; if their correlation is significant (higher than a threshold), which implies that they have similar behavior in terms of the emergence of certain traffic patterns at the same time during the day, the two nodes are considered adjacent (Ermagun & Levinson, 2019; Ryu et al., 2018). Of course, the value of the correlation metric can be utilized as the edge's weight. In general, this approach can theoretically represent more complex spatiotemporal relations, as it captures the dependencies between pairs of distant and nearby nodes the same way and, in addition, it is dynamic, as the connectivity of the nodes can change over time, depending on the similarity of traffic conditions, and the corresponding prediction model would be fed with an adjacency matrix that is not fixed.

Combined methods

In this case, a function that includes the distance of two nodes, the existence of a physical connection and a correlation metric is used to estimate the weights of the adjacency matrix. Examples of this approach from recent literature are presented in the next section.

The graph representation is currently the state-of-the-art and the most popular in traffic forecasting, because it is a simple and intuitive, yet efficient way to represent any network. In addition, by incorporating novel correlation concepts, the spatial and temporal relations, which are vital for the interpretability and actionability of the model, are extracted. The graph representation is compatible with GCNNs and their variations, as well as hybrid Deep Learning models. An example of how the same road network would be represented according to each one is provided in Figure 8. $V_i$ denotes the traffic variable value at section $i$.

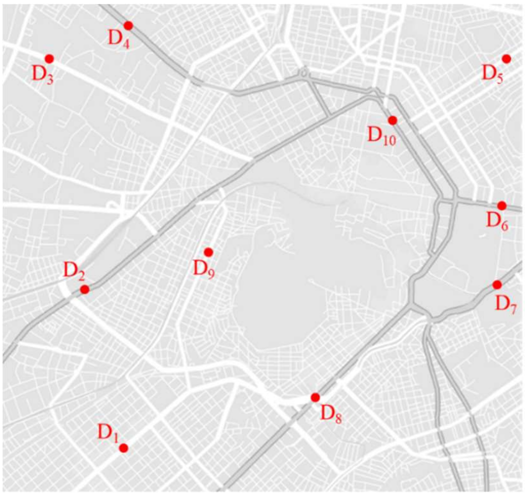



**Figure 7. Network that cannot be effectively represented with physical connectivity matrix.**

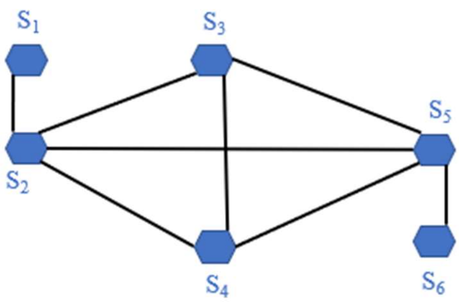

(a) Example road network

| | | | | | |
|---|---|---|---|---|---|
| V1 | V2 | V3 | V4 | V5 | V6 |

(b) Stacked vector representation

| | | | | | | |
|---|---|---|---|---|---|---|
| V1 | | | V3 | | | |
| V1 | | | V3 | | | |
| $\frac{V1 + V2}{2}$ | V2 | V2 | $\frac{V2 + V3 + V4 + V5}{4}$ | V5 | V5 | $\frac{V5 + V6}{2}$ |
| | | | V4 | | | V6 |
| | | | V4 | | | V6 |

(c) Grid/image representation




(d) Graph representation

**Figure 8. Comparison between the three representation methods.**

### *2.3.1 Spatiotemporal Representations Modeling in Recent Literature*

In this section, a selection of the most significant and interesting papers of recent literature, in terms of representation approach, are presented and discussed. Emphasis is given to the details of the method of modeling the spatial and temporal relations between the road network's locations. An overview of the aforementioned papers is given in Table 2, Table 3, Table 4 and Table 5, for stacked vectors, grid, and graph representations respectively. The modeling technique that was exploited in each paper is mentioned in the second column (if it is based on a well-known model, the latter is mentioned inside a parenthesis). In the next column, the technique used to determine the spatial and temporal correlations or other strategies of feature selection are described. Finally, some implementation details are presented, namely the traffic variable that is predicted, the input data resolution (separated with a comma if more than one datasets of different resolutions are exploited) and the performance of the model, in terms of the Mean Absolute Percentage Error (MAPE) or Mean Absolute Error (MAE), in case MAPE is not estimated. As in most papers more than one experiments are conducted or more than one datasets may be exploited, various error values are presented separated with a comma, which refer to different datasets, or a range of values, which refer to one- and multi-step forecasting (first value to one-step and last to longest-step).

**Table 2. Overview of significant recent literature works using stacked vector representation**

| Research Work | Base Model* | Spatial relation/feature selection method | Traffic variable** | Resolution (min) | Performance (MAPE) |
|---|---|---|---|---|---|
| **Zhu et al. (2023)** | Bayesian clustering ensemble Gaussian process (BCEGP) | Clustering of detectors based on traffic measurements and position | Volume | 5 | (MAE 5.3-6.4) |
| **Sun et al. (2022)** | kNN-GRU | Cluster traffic patterns with Kmeans, detect most relevant input with kNN | Volume | 5 | 10.2% |
| **Afrin & Yodo (2022)** | LSTM | Spatial trend, temporal trend | Volume, Speed(t) | 60, 10 | (MAE 80.9, 4.7) |
| **Lin et al. (2022)** | Support Vector Regression-kNN | Maximum information coefficient (MIC) between lagged versions of time series for feature selection | Volume | 5 | 8.6% |
| **Modi et al. (2022)** | Autoencoder | Distance, Pearson correlation between sensors | Speed | 5 | 4.4%-8.4% |
| **Fang et al. (2022)** | Attention LSTM | Temporal Attention | Volume | 10 | 10%-12% |
| **X. Shi et al. (2021)** | Attention Neural Network | Short- and long-term spatial-temporal attention | Volume | 5 | 8%-11% |
| **Hu et al. (2021)** | LSTM | Spatial attention, temporal attention | Volume | 5 | (MAE 2-3.5) |
| **Cheng et al. (2021)** | kNN | traffic clustering by partitioning time periods | Speed(t) | 5 | 13.5% |
| **S. Guo et al. (2019)** | CNN | Sort vectors based on coordinates | Volume | 6 | 4%-11% |
| **Z. Li et al. (2019)** | Gradient Boosting | Partial correlation | Volume | 60, 10, 3 | 6%, 9%, 16% |
| **L. Li et al. (2019)** | Deep Belief Network | Corridor, vectors stacked based on connectivity | Volume | 10 | 10% |
| **Y. Gu et al. (2019)** | LSTM-GRU | Feature selection with entropy-based grey relation analysis | Speed | 2 | 6% |
| **Ermagun & Levinson (2019)** | Linear | Network weight matrix (based on proximity, correlation and network structure, after temporal detrending) | Volume | 1 | 17%-39% |
| **Ryu et al. (2018)** | kNN | Mutual information | Volume | 5 | 7%-10% |

**abbreviations: LSTM-Long Short Term Memory, kNN-k Nearest Neighbors, CNN-Convolutional Neural Network, GRU-Gated Recurrent Unit, GCNN-Graph Convolutional Neural Network, STGCNN-SpatioTemporal Graph Convolutional Neural Network*

***A "t" next to the traffic variable indicates that its values are extracted from a trajectory dataset; else, they were measured with road sensors or loop detectors*

**Table 3. Overview of significant recent literature works using grid/image representation**

| Research Work | Base Model* | Spatial relation/feature selection method | Traffic variable** | Resolution (min) | Performance (MAPE) |
|---|---|---|---|---|---|
| **J. Guo et al. (2021)** | CNN-RNN | 3-dimensional grid representation (x,y are coordinates, z is time and the colour of each pixel is speed) | Speed(t) | 5 | Not estimated (RMSE 0.08) |
| **Ranjan et al. (2020)** | CNN-LSTM-Transpose CNN | Image with colorscale, retrieved from website service | Speed | 5 | Accuracy (0.87-0.83) |
| **Dai et al. (2019)** | CNN | Rearrange grid so that sensors with high correlation (Spearman) are placed closer | Volume | 5 | 10%-12% |
| **W. Zhang et al. (2019)** | CNN | Spatio-temporal feature selection algorithm (STFSA) based on Pearson correlation between locations, spatial-temporal grid (x-axis is time) | Volume | 5 | 5%-8.7% |
| **Ma et al. (2017)** | CNN | Images depicting the temporal evolution of traffic conditions (x-axis is time) | Speed(t) | 2 | Not estimated (MSE 22-39) |
| **H. Yu et al. (2017)** | CNN-LSTM | Color-scale for traffic conditions | Speed(t) | 2 | 20%, 35% |

**abbreviations: LSTM-Long Short Term Memory, kNN-k Nearest Neighbors, CNN-Convolutional Neural Network, GRU-Gated Recurrent Unit, GCNN-Graph Convolutional Neural Network, STGCNN-SpatioTemporal Graph Convolutional Neural Network*

***A "t" next to the traffic variable indicates that its values are extracted from a trajectory dataset; else , they were measured with road sensors or loop detectors*

**Table 4. Overview of significant recent literature works using graph representation (from 2022)**

| Research Work | Base Model* | Node type | Spatial relation/feature selection method | Traffic variable** | Resolution (min) | Performance (MAPE) |
|---|---|---|---|---|---|---|
| **Fafoutellis & Vlahogianni (2023a)** | Gradient Boosting | detectors | Multiplex network (each layer corresponds to an hour of the day), adjacency matrix based on mutual information, community detection | Volume | 60 | 9.5% |
| **J. Gu et al. (2023)** | GCNN-Temporal Convolution | detectors | Adjacency matrix based on distance and Pearson correlation | Volume | 5 | (MAE 4.7-5.1, 15.3-22.1) |
| **R. Huang et al. (2023)** | STGCNN | detectors | Cross-approximation entropy matrix and physical connectivity matrix | Volume | 5 | 12.9%-16.2%, 8.7%-12.9%, 6.9%-9.2%, 6.7%-10.2% |
| **H. Liu et al. (2023)** | GCNN-GRU-Attention | detectors | Proximity adjacency matrix, attention mechanism | Volume | 5 | 10.4% |
| **Huo et al. (2023)** | STGCNN-Transformer | detectors | Two Deep Learning modules to extract long- and short- term spatiotemporal relations | Volume | 5 | 2.7%-4.5%, 5%-8.6% |
| **Shin & Yoon (2022)** | Progressive GCNN | detectors and road sections | Progressive graph: connections between nodes changing over time and learnable | Volume, Speed | 5 | 2.7%-4.5%, 7%-9.9%, 9.3%-11.2%, 7.6%-11.9% |
| **Rahman et al. (2022)** | GCNN-LSTM | intersections | Travel distance adjacency (day, time, trip attraction and production, land use) | Volume | 60 | (MAE 23.07) |
| **W. Zhang et al. (2022)** | GCNN, attention | detectors | Traffic state similarity matrix | Volume | 5 | 9%-15% |
| **J. J. Q. Yu (2022)** | GCNN | detectors | Relational graph with learnable adjacency matrix based on attention mechanism outcomes | Speed | 5 | 9.3%-12.4%, 9.7%-12.4%, 2.2%-4.9% |
| **S. Wang et al. (2022)** | GCNN | detectors | Node aggregation with community detection, clustering of nodes based on traffic flow | Volume | 5 | (MAE 10.4-11.3, 7.4-9.2) |

**abbreviations: LSTM-Long Short Term Memory, kNN-k Nearest Neighbors, CNN-Convolutional Neural Network, GRU-Gated Recurrent Unit, GCNN-Graph Convolutional Neural Network, STGCNN-SpatioTemporal Graph Convolutional Neural Network*

***A "t" next to the traffic variable indicates that its values are extracted from a trajectory dataset; else, they were measured with road sensors or loop detectors*

**Table 5. Overview of significant recent literature works using graph representation (until 2021)**

| Research Work | Base Model* | Node type | Spatial relation/feature selection method | Traffic variable** | Resolution (min) | Performance (MAPE) |
|---|---|---|---|---|---|---|
| **Bai et al. (2021)** | GCNN-GRU | road sections | Connectivity adjacency matrix | Speed(t) | 15 | (MAE 2.7-4.2) |
| **Leiser & Yildirimoglu (2021)** | GCNN-LSTM | intersections | Congestion pattern-based clustering, data of the same cluster are used as features | Speed | 10, 60 | (MAE 2-3.65, 2.63-3.65) |
| **X. Yin et al. (2021b)** | STGCNN | detectors | Neighboring order-based attention mechanism | Volume | 5 | 7.6%-10.3% |
| **Z. Zhang et al. (2021)** | Temporal GCNN | road sections | Fusion of spatial proximity, cosine similarity and graph betweenness adjacency matrices | Speed(t) | 5 | (MAE 10.6-12.7) |
| **K. Zhang et al. (2021)** | GCNN | road sections | Physical connectivity adjacency, spatial attention | Speed(t) | 5 | 10%-18%, 6%-10% |
| **Ye et al. (2021)** | STGCNN | detectors | Include external factors, temporal attention | Speed | 5 | 6%-10% |
| **J. J. Q. Yu et al. (2021)** | STGCNN | road sections | Physical (geographical) connections matrix | Speed(t) | 5 | 7.3%-9.5% |
| **Bogaerts et al. (2020)** | GCNN-LSTM | road sections | Connectivity adjacency matrix, sorted based on latitude and longitude of nodes | Speed(t) | 5 | 12%-14% |
| **Chen et al. (2020)** | GCNN | road sections | Edge-wise adjacency matrix (stream connectivity or competitive) | Speed | 5 | 3%-10% |
| **Q. Zhang et al. (2020)** | Structure Learning Conv. | detectors | Graphs structure (adjacency matrix) is learnable | Volume | 15, 20 | 5.2%-8.2%, 3%-4.8%, 6.7%-9.7% |
| **Zheng et al. (2020)** | GCNN | detectors | Distance-based adjacency, spatial and temporal attention | Volume, Speed | 5 | 15%, 4% |
| **Y. Zhang et al. (2020)** | GCNN-LSTM | detectors and road sections | Neighboring order and similarity-based adjacency matrix | Volume, Speed | 5, 10 | 9%, 9.2% |
| **Do, Vu, et al. (2019)** | GCNN | road sections | Neigboring order-based adjacency | Volume | 5 | 14%-19% |
| **M. Wang et al. (2018)** | GCNN | road sections | Distance (affinity) adjacency matrix, based on travel times between nodes | Volume | 5, 30 | 7%-10%, 12%-14% |

**abbreviations: LSTM-Long Short Term Memory, kNN-k Nearest Neighbors, CNN-Convolutional Neural Network, GRU-Gated Recurrent Unit, GCNN-Graph Convolutional Neural Network, STGCNN-SpatioTemporal Graph Convolutional Neural Network*

***A "t" next to the traffic variable indicates that its values are extracted from a trajectory dataset; else, they were measured with road sensors*

From Tables 2-5, one may observe the popularity of Deep Learning models in recent literature. It is also clear that, apart from the hybrid Deep Learning structures, such as those including GCNN, RNN or CNN layers, a significant number of works adopt the exploitation of an Attention mechanism, which assigns learnable weights to the input data based on their influence on the expected output, to detect the most important spatial and temporal features (Fang et al., 2022). Besides, evidence from recent literature shows that an Attention mechanism can significantly improve a model's performance, especially for long-term (multistep) forecasting (X. Yin et al., 2021a).

Regarding the accuracy achieved by different models, the MAPE metric ranges around 10% and can be even lower in some cases. The values presented in the above Tables cannot be directly compared to each other, as they refer to different experimental setups and also different target variables (e.g., speed, traffic volume). Moreover, according to recent literature, forecasting accuracy is site-specific, meaning that it depends on the specific dataset that is used, the size and geometry of the corresponding road networks and the length of the prediction horizon; thus models evaluated in different datasets and forecasting tasks cannot be directly compared to each other (Manibardo et al., 2021). When using the same benchmarking datasets, it is evident that, in general, the graph representation and the GCNN-based structures are more effective in modeling bigger and more complex road networks, compared to simpler representations and structures. The latter is also reflected in the comparisons between the aforementioned models and simpler baselines in most research papers listed above, indicatively (Chen et al., 2020; Y. Zhang et al., 2020; Ye et al., 2021). X. Yin et al. (2021a), who provide a benchmarking of different forecasting models, spanning from classical statistical methods to state-of-the-art Deep Learning, using several public datasets, come also to the same conclusion. However, simpler representation and modeling methods perform equally well on simpler tasks (e.g. less complex network, fewer locations) and their deployment should be as well considered, taking into account other advantages of these models, such as interpretability, actionability, and simplicity (Nair & Dekusar, 2020; Manibardo et al., 2021).

Finally, regarding the prediction horizon, it is evident that multi-step forecasting is associated with significantly higher error values when compared to one-step ahead forecasting, which is mostly ought to the way the models are trained, i.e., for one-step forecasting. More specifically, for performing multi-step (or long-term) forecasting, the one-step model is used recursively to provide the expected values for each future step, taking as input the forecasted values of the previous ones. That way, each additional step's forecasting (after the first step) incorporates the forecasting errors of all the previous steps and, thus, naturally, the performance metrics gradually drop after each step.

Furthermore, for stacked vector representation, most studies also include a feature selection process, which is vital for reducing the input space's dimensionality and the model's complexity, as well as for introducing the spatial-temporal relations to the model. The simplest approaches include the estimation of a statistical metric to determine the most significant input, such as Pearson and Partial correlation (Z. Li et al., 2019). However, such metrics are subject to several assumptions, such as normality of the distributions of the data, linearity of their relations and independence that are not usually met for traffic data. Thus, the emerging dependencies may not be accurate. Another equally straightforward approach is to select features based on the distance or the neighboring order. But as long as distant locations may be equally or even more correlated than closer ones, the suitability of this method is also questionable. Approaches combining both methods are also popular in recent literature. For example, in (Modi et al., 2022), the authors use a combination of distance and statistical correlation measures to determine the most relevant sensors, while Cai et al. (2015) also include the neighboring order.

In order to overcome the limitations of the aforementioned approaches, metrics from the area of Information theory, such as Mutual Information, have also been exploited (Ryu et al., 2018), while Lin et al. (2022) estimate the maximum information coefficient between the target location's time series and lagged versions of the time series of the rest locations. On the other hand, Cheng et al. (2021) perform K-Means clustering to detect locations with similar traffic patterns. In contrast to the Attention mechanism, the latter correlation metrics and feature selection approaches are not learnable (not learned during the training phase of the model) but are predetermined. For this reason, selecting the most appropriate method is a very important task.

A typical example of image representation is presented by H. Yu et al. (2017), where a color scale is used to represent traffic conditions. The grid values are assigned as the mean value of the corresponding road sections that lie inside them. Ranjan et al. (2020) use a similar representation and image data that are retrieved from a map service's website. In order to increase the efficiency of the classic image representation, Dai et al. (2019) estimate the Spearman correlation between the traffic conditions of all pairs of locations and developed an algorithm to rearrange the grid, so that the most correlated locations are placed closer, while Ma et al. (2017) model a single road section as an image, with the x-axis corresponding to the evolution of traffic conditions over time and y-axis to the traffic conditions across the road section. W. Zhang et al. (2019) use also a grid where the x-axis corresponds to time, but the y-axis contains the traffic measurements of different locations. Additionally, the authors propose a feature selection algorithm, based on the Pearson correlation between the road network's locations. In (J. Guo et al., 2021), the road network and the evolution of traffic conditions are represented as a 3-dimensional grid, where the x- and y-axes correspond to the coordinates of the locations and the z-axis to time. Moreover, a value is assigned to each pixel of the grid, corresponding to the location's speed at the specific timestamp.

For the graph representations, it should be noted, additionally, that there is a variation of node types and hybrid Deep Learning models that are utilized with graph road network representations. Considering the papers that are examined, which can be considered indicative of the works of recent literature, intersection nodes are the less common, mainly due to their lower compatibility with GCNN, which is also mentioned before. Road section and detector nodes, which are very similar in most cases, seem to be equally popular and the choice between the two depends mainly on the available dataset. The most usual modeling approach is a combination of GCNN and a model from the Recurrent Neural Networks family (LSTM or GRU), to take account of both the spatial and temporal relations.

The spatial relations, which are expressed through the adjacency matrix, are most of the times based on the physical connectivity, the distance (or the neighboring order or travel time) between the nodes or a correlation metric. Y. Zhang et al. (2020) use a combined similarity and neighboring order-based adjacency matrix, while Bai et al. (2021) a connectivity-based one, and S. Wang et al. (2022) construct an affinity matrix based on travel times. The more sophisticated approaches, which include a function that combines the latter aspects, lead to more accurate detection of the dependencies. For example, Z. Zhang et al. (2021) use a matrix that includes physical proximity information fused with cosine similarity and graph betweenness metrics. Furthermore, in (Q. Zhang et al., 2020) a Structure Learning Convolution (SLC) framework is proposed, which is able to learn the adjacency matrix during the model's training phase, given the nodes of the graph. J. J. Q. Yu (2022) follows a similar approach, using a learnable adjacency matrix, based on the outcomes of an attention mechanism. In (Leiser & Yildirimoglu, 2021) and in (S. Wang et al., 2022), the authors utilize two different clustering methods in order to enrich the spatiotemporal dependencies information.

Lastly, the typical temporal resolution for traffic data is 5 minutes, but higher resolutions are becoming increasingly popular. Forecasting with a shorter-term horizon of 1-3 minutes poses a significant scientific challenge and is also deemed important for near real-time traffic management incorporated into Intelligent Transportation Systems. In general, data with higher resolution contain higher variability as well, even between consecutive timesteps, which makes the forecasting task harder and more complicated. In this case, an effective representation and an accurate spatiotemporal analysis are crucial to achieving decent forecasting accuracy. On the other hand, data with a lower temporal resolution are smoother due to aggregating traffic measurements of a relatively longer horizon; as a result, simpler and less sophisticated models may perform very satisfactorily.

## 2.4 Challenge 3: Correlation Versus Causation

### *2.4.1 A Historical Perspective*

In contrast to most time series data, traffic data evolve with both space and time as highlighted in research for the past 30 years. The first spatiotemporal representations either in autoregressive models or neural networks were linear in nature and introduced the influence of upstream data, which varies during different periods of the day, in order to increase the accuracy of the model (Tebaldi et al., 2002; Stathopoulos & Karlaftis, 2003; Vlahogianni et al., 2005). Since then, a series of papers showed evidence that the influence of traffic patterns in different locations of the road network on the target locations is complex, highly non-linear and varying over time and, thus, difficult to model (Vlahogianni et al., 2004, 2014; Ermagun & Levinson, 2018; X. Yin et al., 2021a; K. Lee et al., 2021). The existence of dependencies between the traffic conditions (traffic flow or volume, speed, etc.) at different locations is also supported by traffic flow theory for both signalized and unsignalized road networks (Vlahogianni et al., 2008, 2014; Pavlyuk, 2019).

Interestingly, a clearcut literature finding is that the spatiotemporal dependencies are not limited by the connectivity and the proximity of the locations in space and time; the traffic states of road sections that are close to the target section are not necessarily the most correlated to its traffic state, while the same applies to the traffic states of far time steps, which sometimes are more correlated with the predicted time step than more recent onesΑιτι (Ma et al., 2015; Do, Taherifar, et al., 2019; X. Yin et al., 2021a; Jiang & Luo, 2021). The selection of the most relevant historical observations for the forecasting task remains a challenging, yet vital, task for modeling, understanding and decision-making.

The effect of the spatiotemporal analysis on traffic forecasting was mainly limited to the feature selection process, until recently. Except from the simplest approach of using the direct upstream and downstream links or higher-order neighboring links, researchers have also considered more sophisticated strategies. The first one is based on the distance of the locations: it was assumed that closer locations would have similar traffic conditions patterns, especially in the short term. The distance could also be defined as the travel time between the locations, instead of its actual (Euclidean) value (Pavlyuk, 2019). Alternatively, a correlation coefficient was estimated between the time series of traffic conditions of different locations, in order to detect the most correlated ones. A variety of metrics and methods can be used for this purpose, with the most popular being Pearson's correlation, cross-correlation, mutual information and custom metrics proposed by the corresponding authors (Ermagun & Levinson, 2018; Fafoutellis et al., 2020). The latter methods allowed researchers to come to the conclusion that not only near, but also distant locations have high (sometimes higher) correlation coefficients. From a traffic theory perspective, this, intuitively, depends on the size and geometry of the road network, as well as the time resolution of the examined data.

These techniques were very popular before the emergence of Machine Learning and, especially, Deep Learning and were very suitable and efficient for models that cannot cope with a very large and complex input space and, although more sophisticated inputs and methods have been proposed, they are still utilized.

### *2.4.2 Causality for Time Series Data: Granger Causality Test*

In the field of traffic forecasting, despite most researchers agree about the existence of significant causal relationships between different locations of a road network, they usually depend on deep learning models to cope with large datasets and find correlations between the data. The specific correlations are not usually transparent and understandable for human users, while it is not guaranteed that they are causal. Thus, to account for the spatiotemporal evolution of traffic in a consistent manner that is interpretable and not spurious, causal traffic patterns should be detected.

When modeling any phenomenon, causal relationships between input and output variables play a very important role. In contrast to statistical correlation which may be observed at a specific dataset, the notion of causality refers to more significant relations between the variables, in a sense that, e.g., the value of one or more input variables will "cause" the output to take a specific value (Miller, 2019). However, detecting the causal relations between variables is not a straightforward task, especially when the input space is extensive, and may rely completely on expert knowledge on the specific field.

Taking the above into account, studying the causal relations of a dataset is very important for three main reasons:

- To use input features that are causally related to the expected outcome and whose impact on it remains constant and is independent of the variations of the features' values.
- Usually, the input space includes both causal and non-causal variables and, using explainability techniques, they cannot be separated, which leads to unreliable explanations.
- When both causal and non-causal features are used for the prediction task, the dimensionality of the input space is significantly increased, undermining the model's performance; using only causal features should be adequate.

Causal relations are usually represented as a directed acyclic graph (causal DAG), whose nodes are the variables that are considered and there exist directed connections between nodes that have causal relations (from the variable that has a causal effect on the other to the second). The causal DAG is created based on probabilistic relations, namely estimation of the conditional probability as proposed by Bayes and the relevant graphical model, called a Bayesian Network (Kaddour et al., 2022).

When dealing with time series data, the most appropriate and popular method for detecting causal relations is the Granger causality test (L. Li et al., 2015). Simply put, we say that a time series "x" Granger-causes time series "y", when "y" is more accurately predicted by a model if previous values of "x" are included in the input space (Schwab & Karlen, 2019). In classic Granger causality, the model that is used to evaluate the existence of causal relations is the Vector Autoregressive model (VAR), i.e., only linear relations are assumed between the data, which may be oversimplified for real-world systems, e.g., for traffic data (Tank et al., 2021).

The most important advantage of Granger causality, though, is that it is a multivariate method, i.e., evaluates the existence of causal relations between two variables taking into account the effect of all other input variables as well, in contrast to most other methods that consider only pairwise relations (Tank et al., 2021). More formally, according to the VAR model, the output variable $y_t$ is a linear combination of the previous time steps of the input variables $x_{i,t-l}$:

$$y_t = \sum_{l=1}^{T} a_{1,l} x_{1,t-l} + \sum_{l=1}^{T} a_{2,l} x_{2,t-l} + \cdots + \sum_{l=1}^{T} a_{n,l} x_{n,t-l} \qquad (2)$$

where T is the maximum number of time lags considered, n is the total number of input time series and $a_{i,l}$ is the coefficient of the ith time series related with t-l time step. The coefficients values are determined to minimize the following quantity:

$$\frac{1}{N}(\sum^{N} y_t - \sum_{i=1}^{n}\sum_{l=1}^{T} a_{i,l} x_{i,t-l}) + \lambda \sum_{i=1}^{n}\sum_{l=1}^{T} |a_{i,l}| \qquad (3)$$

where N is the total number of observations. The above quantity has two terms: the first one expresses the estimation error of the VAR model, while the second is the sum of the coefficients multiplied by a factor λ. For the above quantity to be minimized, when including a specific time step of a time series in the input space does not contribute enough to reducing the estimation error (first term), its corresponding coefficient's value becomes zero and, thus, it is excluded from the model. The parameter λ is set by the user and controls the significance of the second term over the first. At the end of the process, the time series with at least one non-zero coefficient are considered to Granger-cause y.

Due to the difficulties in detecting causal relations in complex datasets, causality concepts remain under-utilized in traffic forecasting literature, despite their significance in developing more stable and interpretable models. Most state-of-the-art approaches rely on fixed graph representations of the input space, representing the spatial dependencies between different locations of the road network based on their Euclidean distance, their neighboring order or the estimation of a statistical correlation metric, most of which are not a very suitable choice for time series data (Ye et al., 2022). However, as long as causal relations of traffic conditions are not sufficiently considered, the above methods are vulnerable to spurious correlation, as a moderate correlation between locations may be the result of confounding variables and biased data (L. Zhang et al., 2022).

Regarding the use of the concept of Granger causality in recent traffic forecasting literature, the authors of (L. Li et al., 2015) have adopted the Lasso method-based Granger causality model to perform feature selection, which quickly filters out irrelevant data. The method was found to enhance the model's performance but does not cope with the linearity assumption of Granger causality. In (Y. Wu & Tan, 2016) Granger causality is used to interpret the outcomes of a developed model, while the authors of (He et al., 2022) use simple linear Granger causality as a component of a deep learning framework, in order to incorporate causal spatiotemporal relations. Finally, Zhang et al. developed a deep learning framework which first learns the causal relations and then performs graph convolution on the causal graph (L. Zhang et al., 2022).

## 2.5 Challenge 4: Explainability and Spatiotemporal Analysis

In general, interpretability and explainability refer to a model's transparency, which implies that the data or algorithm and the mechanism that provides the outcomes are accessible to some extent (Miller, 2019). Models such as Linear Regression, Decision Trees, and rule-based models are considered easy to interpret; linear models offer explanations for predictions generated by the signs and magnitude of the coefficients, while Decision Trees and rule-based models, on the other hand, have a certain degree of interpretability due to their reliance on decision rules. Tree-like models, in particular, can provide immediate information on the most relevant attributes of a specific rule because of their hierarchical structure.

Explainability aims to make complex Machine Learning and Deep Learning models explainable using dedicated tools and methods after their development. It is a broader concept that encompasses the development of AI systems that are understandable, fair, accountable, trustworthy, and transparent, allowing the end-user to comprehend the "what," "why," and "how" of the models (Gunning et al., 2019). Explainability aims to provide customized and relevant information to different stakeholders, taking into account their goals, privacy, and adaptability to human understandability (Barredo Arrieta et al., 2020).

Explainability and identification of the spatial and temporal relations between the locations of the road network, clearly and transparently, are very important for traffic management purposes (Vlahogianni et al., 2014). Specifically, it is essential in order to (Barredo Arrieta et al., 2020):

- Justify the decision-making process and increase the trust in the specific model, which is necessary for the compliance of the network users.
- Extract new scientific knowledge regarding the network's mechanics
- Find ways to improve the prediction model's performance and its transferability

Inducing knowledge on the spatial and temporal relations of the forecasting process, via the road network and input data representation, as well as the detection of causal relations, enhances the model's performance, but also increases its actionability, as it allows the usage of less complex and more interpretable models.

Moreover, understanding the spatial and temporal relations is important for coping with non-recurrent conditions: when extreme conditions (heavy congestion) emerge, e.g., due to an accident, a Deep Learning prediction model would not be able to predict the evolution of the phenomenon and the locations that would be affected, because these models are dependent on the input training data, which most probably would not include a sufficient number of non-recurrent events emerging at all the locations of the road network. Thus, they would not be able to predict future conditions that are not observed in the input data. As a result, a network management authority would not be able to timely implement corrective measures to prevent the spread of congestion. On the other hand, if the spatiotemporal dependencies have been identified, either before the development of the model (for the representation of the input data or the detection of causal relations) or after (for interpreting the results), it would be clear which locations are going to be directly affected and the corresponding measures would be enforced.

Extracting the spatiotemporal relations and interpreting the outcomes of Deep Learning models is not a straightforward task and is usually performed post hoc, using model agnostic methods, such as LIME (Local Interpretable Model-Agnostic Explanations), SHAP (SHapley Additive exPlanations) and Partial Dependence Plots (Molnar, 2019). However, these methods do not have a strong mathematical foundation and depend entirely on the available data; consequently, they are very vulnerable to noisy datasets and may provide unreliable outcomes. Moreover, they only imply statistical and not causal relations. To address this issue, researchers should take into consideration the knowledge coming from traffic flow theory concerning traffic spatiotemporal propagation and congestion dynamics, so that the noisy information can be translated into important and potentially causal features that can reduce the dimensionality of the forecasting problem and improve its reliability. Along the same line, the emerging field of Causal Machine Learning proposes a variety of methods to examine and quantify the causal relationships in the available data (Y. Zhao & Liu, 2023). The exploitation of such methods for short-term traffic forecasting remains to be researched.

Finally, in recent literature, it is very common to observe researchers developing a very sophisticated model, evaluating its performance, and indicating its superiority compared to baselines, but, at the same time, not elaborating on analyzing, understanding and presenting the spatial and temporal dependencies (Manibardo et al., 2021; X. Yin et al., 2021a). However, for a forecasting process to be actionable in real-world traffic management scenarios, it should be assessed not only based on the values of error metrics, but also on the statistical properties of the error and the error bias that affect its trustworthiness as well (Karlaftis & Vlahogianni, 2011; Vlahogianni & Karlaftis, 2013b). Understanding the effect of bias on the model's trustworthiness in the presence of extensive network-level spatiotemporal information is at an early stage, especially in relation to real-world applications.

## 2.6 Challenge 5: Multitask Learning: The Need for Multivariate Predictions

### *2.6.1 Fundamental Relations of Traffic Variables*

The three fundamental variables are the most popular traffic conditions indicators as well, are the following: traffic flow rate or traffic volume (vehicles/hour), velocity/speed (km/h) and density (vehicles/km), which can also be expressed as occupancy (% of the time that a car is on a detector). Measurements of the above variables at different locations of a road network (i.e. network-level data) are an instance of spatiotemporal correlated time series data, as there exist strong and complex dependencies between them, that are also varying over time (X. Yin et al., 2021a).

For the case of a single road section, the three most common indicator variables of traffic congestion mentioned above are related to each other according to the fundamental equation (Bramich et al., 2022):

$$q = kv \qquad (4)$$

where q is the traffic flow, k is the vehicle density and v is the mean speed of the vehicles. The graphical presentation of the above relationship for each pair of variables constitute the fundamental diagrams of traffic flow theory (Koch et al., 2022). The fundamental diagrams have several properties, that are very important in various traffic engineering tasks, e.g., traffic management, calculations using theoretical concepts such as shockwave theory, etc. (Knoop & Daamen, 2017). More specifically, the q versus k diagram has a triangle-like shape. An indicative example is illustrated in Figure 9. The point that is at the top of the triangle represents the capacity (or maximum flow) of the specific road section, while the corresponding density is called critical density. For densities higher than the critical density, the vehicles' flow becomes congested and starts decreasing. On the other hand, for lower density values the flow is uncongested. Most traffic management methods are implemented with the objective to keep flow at the uncongested area (Bramich et al., 2022).

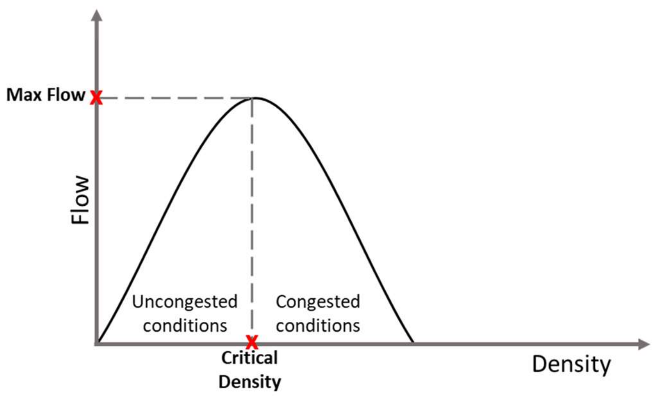
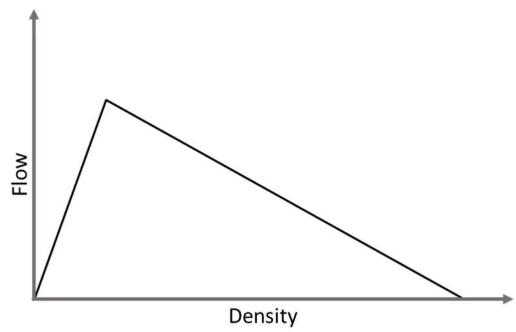



**Figure 9. Indicative traffic flow – density fundamental diagram**

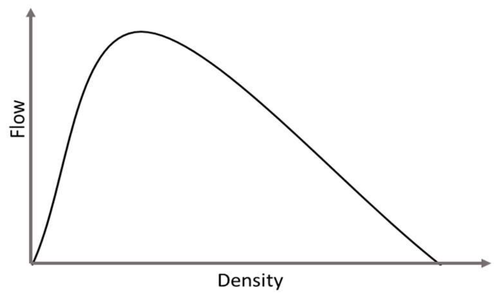
Several approaches to the exact way of fitting the fundamental diagram to available measurements and the shape and number of the curves to use have been proposed in the literature. Starting from the 1930s, Grienshields et al. proposed the use of one curve per fundamental diagram (Greenshields et al., 1935). The relation between speed and flow is parabolic and the one between speed and density is linear. Another univariate model (use of one curve) with similar characteristics is Drake's, where, however, the relation between speed and density is exponential (Drake et al., 1967).

Alternative approaches propose the use of two different curves to describe the congested and uncongested branches (two-variate models). These models allow a more detailed illustration of the fundamental relation and are considered more accurate (Knoop & Daamen, 2017). The most naïve of them is the triangular, where the two branches of the flow versus density diagram are straight lines (linear relation) and, more specifically, the slope of the congested branch is equal to the shockwave speed and the one of the uncongested to the free flow speed (Knoop & Daamen, 2017). Daganzo introduced the truncated triangular flow-density diagram, where, for a range of densities close to the critical value, the flow retains its maximum value (is constant) (Daganzo, 1997). According to the above models, the speed when traffic is at the uncongested branch remains constant.

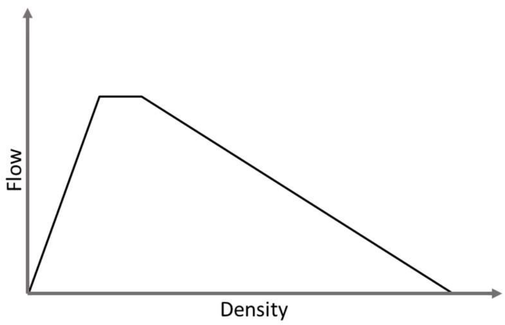
More recently, researchers argue that a discontinuous fundamental diagram, taking capacity drop into account, might be more accurate. Here, the uncongested branch does not start at the maximum flow (capacity) but at a lower level (Knoop & Daamen, 2017). Moreover, it was also suggested that it should not be assumed that the relation between flow and speed is linear and that speed during uncongested conditions is constant (N. Wu, 2002). The most significant of the above-described models are presented in Figure 10.



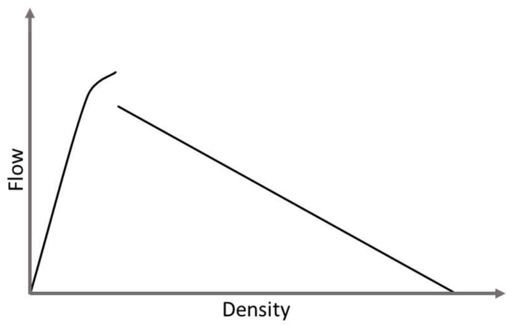

(a) Triangular fundamental diagram




(b) Parabolic fundamental diagram (Grienshields)




(c) Truncated triangular fundamental diagram




(d) Inverse lambda fundamental diagram (Wu) – capacity drop

**Figure 10. Most significant fundamental diagram models.**

It should be noted that the aforementioned approaches require the knowledge of various parameters, such as the free flow speed, the wave speed, the free flow capacity and the corresponding capacity drop, as well as the functional form of the curve or curves used to describe the fundamental diagram, i.e., they are deterministic.

### *2.6.2 Multivariate Forecasting Schemes*

Producing multivariate predictions, meaning using multiple variables in the input but also in the output space, using classical time series approaches or neural networks has been a common approach in traffic forecasting since the early 2000s (Vlahogianni et al., 2004; Vlahogianni, 2009a; Dunne & Ghosh, 2012; Vlahogianni et al., 2014; Y. Yin & Shang, 2016; K. Zhang et al., 2020).

In traffic forecasting, the concept of multitask prediction can be implemented by exploiting more than one traffic variables (volume, speed, occupancy/density), whose dependence on each other can be described with the fundamental relations (and fundamental diagrams), as discussed in the previous section. The main advantage of this approach can be explained with the following example: although the same traffic volume measurement may correspond to extremely different traffic conditions, namely congested and uncongested traffic respectively, a single-task model would interpret them the same way. On the other hand, the knowledge of the value of an additional variable would allow the model to distinguish and learn the different dynamics of the two states and, probably, estimate their evolvement more accurately. An example of the above is depicted in Figure 11.

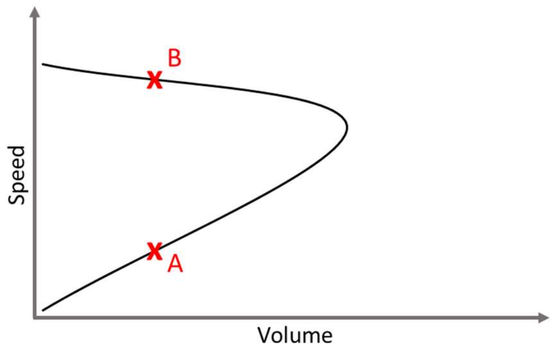



**Figure 11. Example of the advantage of a multitask model: in a single-task model with traffic volume data, points A and B would be considered identical.**

Two representative examples of this approach are the work of (Du et al., 2019), where the authors use all three fundamental variables and also include travel times and of (Liao et al., 2022), which, in addition, takes into account weather and special events data to predict taxi demand. It is to be noted that the output can include multiple stream data coming from the same generation mechanism or coming from correlated generation mechanisms. To illustrate the above, (Liang et al., 2022) identified correlations between ride-hailing and subway demand, while (Z. Liu & Chen, 2022) developed a multitask model that takes as input taxi and metro demand data. (Fafoutellis & Vlahogianni, 2023a) used a combination of data from different travel modes, e.g., public transport demand and traffic data in a multi-task, deep learning framework enhanced with a multiplex network representation to forecast traffic volume and metro (subway) demand.

Regardless of whether the outputs are produced by the same mechanisms or by multiple ones, the main characteristic of a multitask model is that, in addition to the extended output space, its input space includes the input spaces of all tasks (e.g., in the case of time series, the past values of all variables), allowing for data exchange between the tasks. The output variables should be (significantly) correlated; in this case, it is expected that feeding all input spaces to the model would increase the overall prediction accuracy, while, otherwise, it would unnecessarily increase the dimensionality, decreasing the model's efficiency and adding noise to the input.

## 2.7 Challenge 6: Enhanced Trustworthiness with Physics-Informed Neural Networks

The state of the art in short-term traffic forecasting relies almost exclusively on advancements in deep learning and performance-driven modeling, disregarding the knowledge stemming from empirical and analytical investigations of traffic flow. This leads to forecasting constructs that are usually extremely complex, difficult to understand and hard to generalize on unseen events, and, eventually, of limited trustworthiness (Laña et al., 2021). Several researchers argue that deep learning models can hardly claim applicability in large-scale scenarios (city-level traffic management), due to significant computational resources requirements and the inability to generalize well since there is a need for large amounts of data that should be also representative of the phenomenon they intend to model (Di et al., 2023). On the other hand, physics-based traffic models, when applied in real-world conditions, lead to inaccurate estimations for four main reasons (R. Shi et al., 2021; Usama et al., 2022): they can only capture a limited subset of the possible traffic conditions and related dynamics, they refer to ideal conditions, they include several parameters that should be defined by the user and they are vulnerable to noisy data. However, this kind of models are interpretable and do not require massive amounts of data (Di et al., 2023).

Combining the above approaches leads to models that are very efficient in terms of data demands, achieving similar or even better performance than classic Deep Learning models (Usama et al., 2022). (Laña et al., 2021) argue that considering theoretical concepts in data-driven modeling at all phases, including preprocessing, modeling, and evaluating, can reduce model complexity, improve performance and trustworthiness and, most importantly, lead to feasible predictions and rational decision-making.

The concept of Physics- (or Theory-) Informed Deep Learning combines the advantages of the two main modeling approaches, namely analytical or physics-based and data-driven, in order to achieve better generalizability and higher accuracy (Di et al., 2023). Specifically, it can be deployed when a data-driven model includes, as input or output, one or more variables for which an analytical/mathematical relation is known from the corresponding scientific field. This relation is usually incorporated into the model's loss function or is used in an independent module and aims to adjust highly erroneous, unreasonable predictions of the model, towards what is "in theory" expected. In this case, the model's outcomes are assessed based on their distance from the observed (actual) outcome and the one expected by the relevant physics law (Usama et al., 2022).

Combining analytical and data-driven models has been previously followed by many researchers. For example, combinations of analytical traffic models with unsupervised learning methods have been used for queue profile estimation at signalized intersections (Ramezani & Geroliminis, 2015). (X. Wu et al., 2018) introduced a data-driven computational graph that implements the traditional 4-step process on a road network and uses the back-propagation algorithm to refine the estimated travel demand. Other approaches include combinations of analytical traffic models with neural networks for queue estimation (Lu et al., 2023), and for the development of new car-following models (Mo et al., 2021).

Regarding using theory-driven knowledge for training, (R. Shi et al., 2021) proposed a framework where a physics-informed neural network, which encodes analytical equations of either macroscopic or microscopic traffic flow models, is used for regularizing the predictions of a simple one. (Usama et al., 2022) incorporate the residuals of traffic flow continuity and flow conservation equations into the loss function. For their experiments, they use a small circular simulated road network. Recently, (A. J. Huang & Agarwal, 2023) also used analytical equations relating speed and density and incorporated them in the cost function. The above approaches documented improved accuracy and decreased computational time.

It is without doubt that both empirical and analytical investigations can provide valuable insights into the mechanics of network traffic and, when combined with powerful modeling techniques, they can enhance their actionability, resilience and efficiency, as well as reduce prediction error. In the case of short-term traffic forecasting, inducing knowledge from theory to the forecasting process is not a straightforward process, but should be done holistically, ranging from issues of model structure, learning, as well as possible insights extraction from the trained models.

## 2.8 Challenge 7: Network-Wide Forecasting

Developing a modeling framework that would provide predictions for an entire road network, at once, remains under-researched, although it can increase significantly the actionability of any model. In recent literature, there are very few network-wide approaches (Cui, Ke, Pu, & Wang, 2020); most models provide output for a single location and not all of them simultaneously, usually referred to as target location, and/or only exploit a relatively small part of the road network as input. The reason for this is that there are several limitations to network-wide traffic prediction, including the following:

- Data availability and quality: Network-wide traffic prediction requires access to large amounts of traffic data, whose quality and reliability can vary greatly, limiting the accuracy of traffic predictions.
- Computational requirements: Deep Learning traffic prediction models can be computationally expensive, especially when they are applied at a so large scale, which limits the ability to deploy these algorithms in real time.
- Model complexity: Traffic prediction models can be highly complex, especially when they incorporate additional factors, such as road geometry and traffic flow dynamics. These models can be difficult to interpret and validate, which prevents them from being applied effectively in practice.

Researchers are developing algorithms and methods that can address these challenges and enable the deployment of network-wide traffic prediction, including new concepts of Computational Science that emerged relatively recently. The most important of them is Edge Computing, in which the computational tasks are performed at the edge of the network, where the data are generated, rather than in a centralized cloud or data center. The main advantages of Edge Computing over traditional cloud computing are (W. Shi et al., 2016; Cao et al., 2020):

- Latency and bandwidth reduction: By processing the data closer to where they are generated, reduced amounts of data need to be transmitted to a centralized location and, thus, the related latency and bandwidth requirements are reduced, while the efficiency of the network is improved.
- Privacy and security: The local processing of data, as well as the reduced amount of them that has to be transmitted, reduces the risk of leaking of sensitive data.

By processing traffic data at the edge of the road network, Edge Computing can enable real-time, low-latency traffic forecasting that can provide valuable information to traffic management authorities. The combination of local data processing, real-time analytics and deployment of machine learning models on edge devices with optimized decentralized coordination (edge nodes exchange necessary information and data without the need for a central entity), there is no need for transmitting large amounts of data to process them in a centralized way, which empowers transportation systems to respond quickly to changing traffic conditions and enables timely decision-making for traffic management, route optimization, and congestion mitigation. For example, Edge Computing can be used to process large amounts of traffic data generated by sensors, cameras, GPS devices and other sources, in order to provide near real-time predictions of traffic conditions.

Based on the principles of Edge Computing, Federated Learning, is a machine learning paradigm that allows multiple participants, such as devices or edge nodes, to collaboratively learn a model without sharing their raw data with a central entity. Instead, the participants train their local models based on their own data and then exchange model updates with each other in a decentralized manner, allowing the overall model to be improved through the collective contributions of all participants (Bonawitz et al., 2019). This approach has the advantages of scalability, i.e. the load of the computational task is distributed across the participating devices, enabling scaling up to problems that would be infeasible to solve with a centralized approach, and robustness, as the system would be able to continue to operate, even if some participants experience any kind of failure (T. Li et al., 2020).

Additionally, Federated Learning can enable the deployment of traffic forecasting algorithms that would be impractical or impossible to run in a centralized cloud or data center due to their computational requirements or the large amounts of data they generate. For example, by distributing the data analysis and model training tasks to various devices close to where the data are collected, the deployment of sophisticated Deep Learning algorithms, that would be able to learn from network-wide traffic data in real-time and provide highly accurate predictions of future traffic conditions, and in a scalable and efficient manner would be made possible. These devices constitute the nodes of a real-world graph. So, representing the relationships between them becomes even more vital for both the training of the local models and, most importantly, for defining the contribution of each of them to the overall model.

For the time being, the exploitation of Federated Learning in traffic forecasting is at a very early stage, with only a few works taking advantage of its potential (Xia et al., 2022). (Y. Liu et al., 2020) propose the “Federated GRU”, which uses a safe data aggregation mechanism, based on a federated averaging algorithm, which prevents sharing private data among, e.g., different organizations. On the other hand, (Zeng et al., 2021) developed a framework for traffic forecasting using data from different traffic stations, without the need for sharing the data between them. The method depends on partitioning the data into different clusters, based on a hierarchical clustering approach. Moreover, (Zhou et al., 2022) have exploited federated learning for vehicle trajectory prediction in order to preserve the drivers’ privacy. In the near future, with the penetration of connected and automated vehicles in traffic and the wider development of C-ITS, it is expected that massive amounts of network-wide data would be made available, making technologies such as Edge computing and Federated learning an essential part of real-time traffic forecasting and management.

## 2.9 Challenge 8: Efficiency and Scalability to Multimodal Environments

Demand prediction of public transport, as well as other modes (e.g. taxi, ride-hailing services, bicycles, etc.), are all instances of variations of the traffic forecasting problem, that also can be addressed similarly. An inclusive approach, considering all modes, would revolutionize traffic management and decision-making at a city level, and provide authorities with a tool that would enable the optimization of traffic conditions across the entire road network. The above is possible with the extension of the concept of multi-task prediction in traffic forecasting. As discussed earlier, multi-task prediction is a novel machine learning technique where multiple related tasks are learned and predicted simultaneously (by the same model), in contrast to single-task prediction, where each task is learned and predicted separately (Jiang & Luo, 2021). In multi-task prediction, the model gets as input shared representations for the tasks, allowing for knowledge transfer between the tasks and improving the overall performance. This is particularly useful when there is limited data available for each task and the tasks are related, as the shared representations can help the model generalize better to new data. Multi-task prediction has already been applied in various domains with great success (Kendall et al., 2018; X. Liu et al., 2019).

In traffic forecasting, multi-task prediction can be associated with multimodal prediction, i.e. predicting the traffic demand of different modes, e.g. volume of private cars, passengers of buses, subway, trains, etc. As it is widely assumed, but also indicated by recent works, there have to be very significant correlations between road traffic and the demand for public transport, which can be utilized to increase the predictability of each individual variable (Fafoutellis & Vlahogianni, 2023a). Moreover, this type of prediction would have a very significant impact and implications on traffic management, and is vital for model actionability, as it constitutes an integrated, multimodal tool for city-level traffic management. If the availability of such multi-source data increases, the popularity of multi-task traffic prediction is expected to increase as well. Such a model would require more than one input dataset and network representations (one for each mode/variable), or, equivalently, a multi-layer graph structure. In order to develop a model of satisfyingly low complexity that would be efficient, despite having a so large input space, an accurate representation of the input spaces and the interrelations between them is necessary.

## 2.10 Research Questions

According to the literature review findings and the identified challenges, and concerning the objectives of the present dissertation, the following research questions can be formulated:

**Q1: Can significant spatiotemporal traffic patterns be identified in a road network? How sensitive are they to the extraction process? What is the impact of those patterns on traffic's short-term predictability?**

A variety of correlation concepts, spanning from statistical to information-theoretic, are utilized to detect significant relations between locations of a road network. We showcase that different methodologies may lead to different results in terms of spatiotemporal data evolution, affecting the predictability of traffic conditions differently.

**Q2: How can the typical graph-based representations of the road network and the spatiotemporal relations be extended to include information from multiple transport modes and what will the impact be on traffic's predictability?**

We investigate the use of the novel multiplex network representation, adapted from the field of Social Networks Analysis, to model the relations between road traffic and public transport demand. More specifically, the daily multimodal traffic patterns are represented as a 24-layer graph (one per hour) and a community detection algorithm is used to identify significant spatiotemporal relations between the two modes which leads to increased predictability for both modes.

**Q3: What will the proper problem setup and multitask modeling structure be to enable efficient and accurate network-level traffic forecasting?**

We use traffic volume data from loop detectors to forecast travel times of multiple routes simultaneously, using a single multi-output model. Within this multitask context, the model is capable of learning significant relations between the travel times of routes and traffic volumes at neighboring and distant locations, despite using a simple yet efficient model structure, achieving decent forecasting accuracy and leading to a resilient and sustainable network-level prediction schema.

**Q4: Can we establish a causal forecasting framework that would lead to the detection of meaningful traffic patterns enhancing the trustworthiness and the accuracy of the forecasts?**

In order to detect causal relationships between different locations of the road network, we utilize the innovative Neural Granger concept, the Deep Learning adaptation of the classic Granger causality test, which reveals generalizable (not dataset-specific) traffic patterns at a city level. Moreover, using the aforementioned concept for feature selection purposes, we were able to increase the forecasting accuracy, while reducing the dimensionality of the input space and the forecasting model's complexity.

**Q5: How can theory-aware modeling aspects be incorporated into the forecasting process and improve the trustworthiness and actionability of the prediction models?**

We introduce traffic flow theory aspects at all modeling stages, from the data preparation and fitting process to the evaluation of the outcomes, that lead, as results indicate, to increased performance and trustworthiness. More specifically, we use a multivariate input and output space (instead of a single variable), which correspond to certain traffic states and train the model with a novel loss function based on the fundamental relations of traffic flow, according to the concept of Physics-Informed Neural Networks. Moreover, we propose a dedicated evaluation framework that assesses the model's trustworthiness by comparing the distribution of the predictions with the actual and accounting for the accurate classification of the traffic state as congested or uncongested, instead of just the values of common error metrics.

# 3 LINEAR AND NON-LINEAR SPATIOTEMPORAL PATTERNS OF NETWORK TRAFFIC

According to what is already discussed, an obvious first step towards the development of an actionable forecasting model is to examine whether spatial dependencies exist between the traffic conditions at different locations of a road network and what their impact on short-term traffic forecasting is. The present section attempts to introduce a sober analysis of spatiotemporal dependencies disengaged from Deep Learning, aiming to increase the understanding of the above topics.

To this end, we implement concepts spanning from classical correlation analysis to Information Theory, Time Series Analysis and Bayesian Networks. The proposed methodological approach is implemented on the road network of the city of Xi'an, China using trajectory data provided by Didi Chuxing Technology Co, a Chinese taxi and private car-hailing company, which were exploited to estimate the time series of mean speed for each section of the road network. More specifically, we proceed to detect the 20 most correlated locations to the target location according to each correlation concept and compare the forecasting accuracies when using exclusively the aforementioned locations as input to a low-complexity model.

## 3.1 Linear Correlation and Mutual Information

To investigate whether there are road sections that are significantly correlated in terms of travel speed, we apply the concept of Mutual Information (MI) and compare the results with the classical correlation analysis (Pearson's correlation). Based on information theory, MI of two random variables or stochastic processes (e.g., time series) is a metric that quantifies the amount of information obtained for one random variable when observing the other. Unlike the classical correlation analysis, the mutual information takes into account nonlinear correlations as well, because the computed measure is not connected to the linear or non-linear evolution rules of the quantities involved, but to Shannon Entropy (Abarbanel, 2012; Kantz & Schreiber, 2004). Let $x_n$ and $y_n$ be two equally spaced sets of random variables with joint probability density $p(x_n, y_n)$ and individual probability densities $p(x_n)$ and $p(y_n)$. The MI $I(x_n, y_n)$, which quantifies the expected information gained about $x_n$ when observing $y_n$ is given by:

$$I(x_n, y_n) = -\sum_{x_n, y_n} p(x_n, y_n) log_2 \frac{p(x_n, y_n)}{p(x_n)p(y_n)} \qquad (5)$$

For the case of traffic time series, this approach exhibits two main limitations: first, it is static in the sense that time is not introduced in the analysis of travel speed interrelations between different network locations. Second, interrelations are assessed in a pairwise manner without providing an understanding of how information from multiple locations may interact with each other and affect predictability.

## 3.2 Distance-Based Time Series Similarity

To address the first limitation mentioned above, the present work implements the fast dynamic time warping (Fast DTW) algorithm. Dynamic time warping (DTW) is a dynamic programming technique to find an optimal alignment between two given time series with the objective of minimizing a specific distance measure (Berndt & Clifford, 1994). For the time series X = $x_1,x_2,\ldots,x_n$ and Y = $y_1,y_2,\ldots,y_n$, DTW distance is given by the following recurrent equation to the matrix $\gamma(i\ldots n, j\ldots n)$ using dynamic programming (M. Lee et al., 2017):

$$\gamma(i,j) = dist(x_i, y_j) + \min\,[\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1) \qquad (6)$$

The path that provides the optimum, namely minimum, distance is the warping path. The DTW distance $DTW(X,Y) = \gamma(n,n)$ is the Euclidean distance along the warping path. DTW has a quadratic time and space complexity that limits its use to only short time series data sets.

To alleviate this limitation, an extension on classical DTW may be used, which first transforms high high-dimensional time series to low dimensional time series and then obtains DTW distances on the low-dimensional time series. This extension known as Fast DTW operates in three steps (Salvador & Chan, 2007): coarsening to reduce the dimensionality, projection to calculate DTW distance in the lowest time series resolution, and refinement to project the warping path to an incrementally higher resolution. The last two steps repeat until the path is projected to the original time series' resolution.

## 3.3 Bayesian Classifier

Finally, to address the limitation of pairwise time series comparison, we develop a Bayesian Network, which presents the relations between all the road sections and is based on the calculation of conditional probability between their speeds' distributions. A Bayesian Network (BN) is a directed acyclic graph whose nodes represent variables. The weights of the connections of the nodes are proportional to the relationship between the variables of the corresponding nodes. With the above model, it is possible to calculate the conditional probability of a variable getting a certain value when knowing the values of all the variables that are connected to it (child nodes) (Pearl, 2000).

The BN for a set of variables $X_i = \{X_1, \ldots, X_n\}$ also consists of a set $P_i = \{P_1, \ldots, P_n\}$ of local conditional probability distributions associated with each node and its parents. BN's causal interpretation is as follows: a directed edge from one variable to another Y, represents the claim that *X* is a direct cause of *Y* with respect to other variables in the DAG (Friedman et al., 1997). The joint distribution *p* can be factorized as a product of conditional probabilities, by specifying the distribution of each node conditional on its parents. For a given structure *B* of a BN, the joint probability distribution $P(X)$ for *X* can be written as:

$$P(X) = \prod_{i=1}^{n} P_i(X_i|pa_i) \qquad (7)$$

where $pa_i$ denotes the set of parents for $X_i$. The BN can be used as a classifier of $X_i$ inputs to a set of classes, in our case, the travel speed classes (*C*), by the rule (Friedman et al., 1997):

$$classify(x_1, \ldots, x_n) = argmax_n\, p\,(R) \prod_{i=1}^{n} p(X_i = x_i|C) \qquad (8)$$

By the BN classification task, the influence of each variable can be determined with respect to the prevailing speed class *C*. The selection of influential spatio-temporal relations of travel speeds will be based on the mutual information criterion. Mutual information quantifies the amount of information flow between a node $X_i$ and the knowledge of traffic speed levels *C*. The mutual information I(X,C) between a variable *X* and a class C measures the expected information gained about *C*, after observing the value of the variable *X*:

$$I(X, C) = \sum_{X,C} P(X|C)P(C)\, log \frac{P(X|C)}{\sum_{c \in C} P(X|C)P(c)} \qquad (9)$$

## 3.4 Implementation and Findings

### *3.4.1 Data Preprocessing*

Data preprocessing is an essential procedure when conducting statistical analysis or applying machine learning techniques. Well-prepared input data lead to better performing and easier trained and tuned prediction models. The data used in this section consist of 3.2 million GPS trajectories of Didi's vehicles in the road network of Xi'an, conducted between the 2nd and 30th of November 2016. Each trajectory consists of the exact position of the vehicle every 2 to 4 seconds, which corresponds to about 250 points per trajectory or 750 million points in total. More specifically, the attributes of each point are: longitude and latitude, a unique ID specifying the ride, a second unique ID specifying the driver and the timestamp that corresponds to the exact time and date that the vehicle was at the particular position. The data were granted by the company to be used for research purposes.

Xi'an is the 12th biggest city of China with population of about 6.5 million residents and is located in central China. The road network of the city center of Xi'an consists of 498 primary and secondary road sections(Figure 12*).*

The present work employs a data preprocessing strategy as follows: First, the coordinates of the points had to be transformed from the Chinese State Bureau of Surveying and Mapping coordinate system – GCJ-02 to the World Geodetic System – WGS-84. In this way it was possible to depict them accurately on the most popular map web services, OpenStreetMap for example, as well as further processing.

Next, each one of the points was snapped to the road section where it is most likely that it belongs, using a Nearest Neighbor based algorithm (Tveite, 2014). The road network of Xi'an was retrieved from OpenStreetMap and the above procedure was implemented using QGIS, which is an open-source software for editing geographical data. The algorithm's complexity is O(nxm), where n is the number of points and m is the number of sections. At the end of this process, our database was enriched with another ID number, which refers uniquely to the road section each point belongs to.

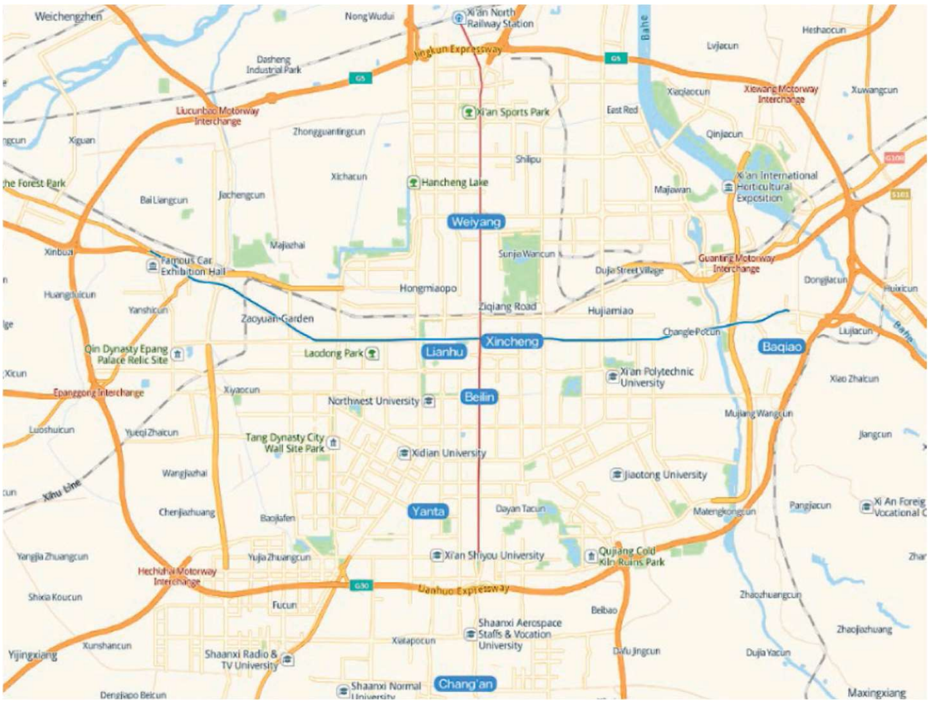



**Figure 12. Xi'an's road network.**

For the purposes of this work, an aggregation time period of 1 hour was used; as already mentioned, a rather simple modeling structure will be used for the forecasting task which would not be able to handle a higher time resolution and, consequently, a shorter term prediction, which is a more difficult task. Besides, the focus of this section is on the methods used for detecting the relationship between the locations, and secondarily on the method and accuracy of the forecasts.

It is also worth mentioning that road sections that did not have any record on any of the twenty-nine chosen days were excluded from further analysis, as it is clear that they do not play an important role in Xi'an's transportation system. The same applies to road sections that do not have any record for more than an hour on any of the twenty-nine days. Figure 13 and Figure 14 show the available full-length time series and one day time series of a specific road section respectively; although a daily seasonality is evident, there seem to be also some short-term features during the day that may significantly affect the magnitude and evolution of speed and, consequently, the prediction accuracy.

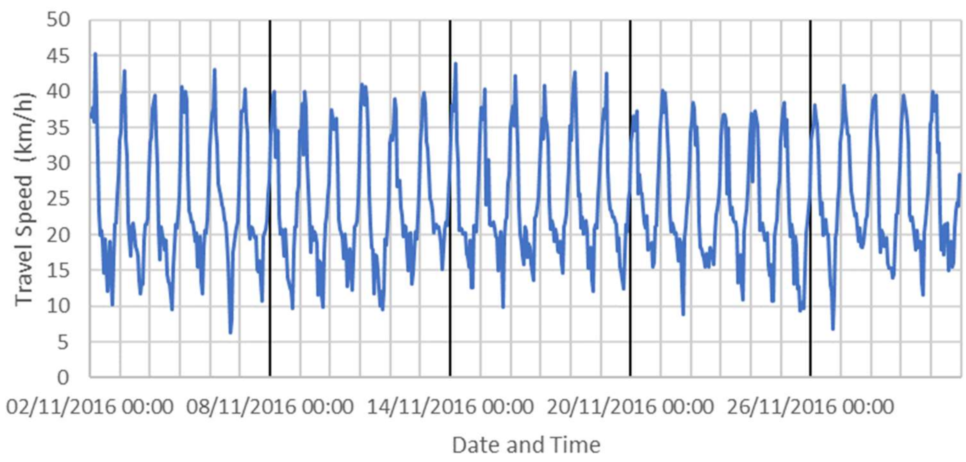



**Figure 13. Sample 30-days time series**

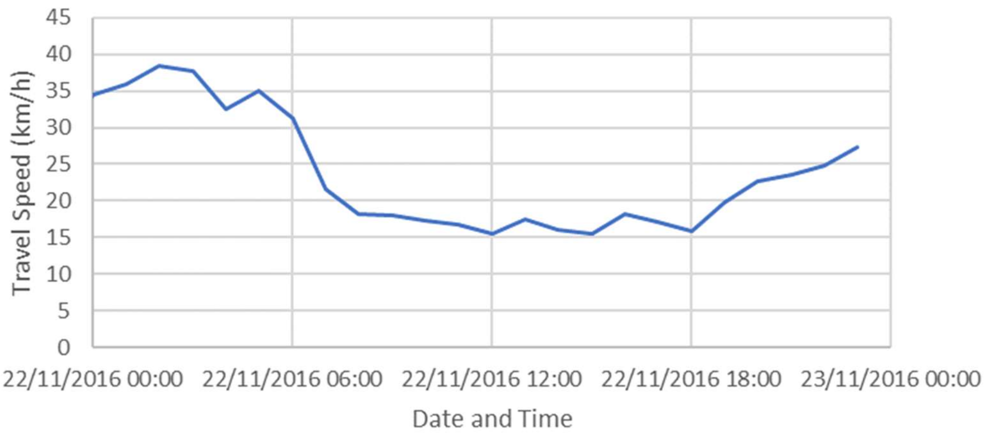



**Figure 14. Sample 1-day time series**

### *3.4.2 Detecting Spatial Correlations*

To detect significant spatial correlations, Pearson's correlation coefficient between each pair of time series was calculated and is presented as a heatmap in Figure 15. This heatmap gives a clear indication of which sections are related to each other. In addition, the mutual information of each pair of the series was calculated. The results are also presented below as a heatmap in Figure 16. In both heatmaps, the lighter the color bands the stronger the relationship between road sections. In the two heatmaps, some common patterns are noticeable in identifying similar correlations between the time series of the same locations. The MI criterion seems to produce lower values in terms of the strength of the identified spatial patterns in general (fewer areas with light color). However, both methods detect locations with almost perfect correlation (close to 1), which indicates the existence of very significant spatial patterns in the road network. By examining each column (or row) separately, one can observe which sections are mostly related to that of the column (or row, respectively).

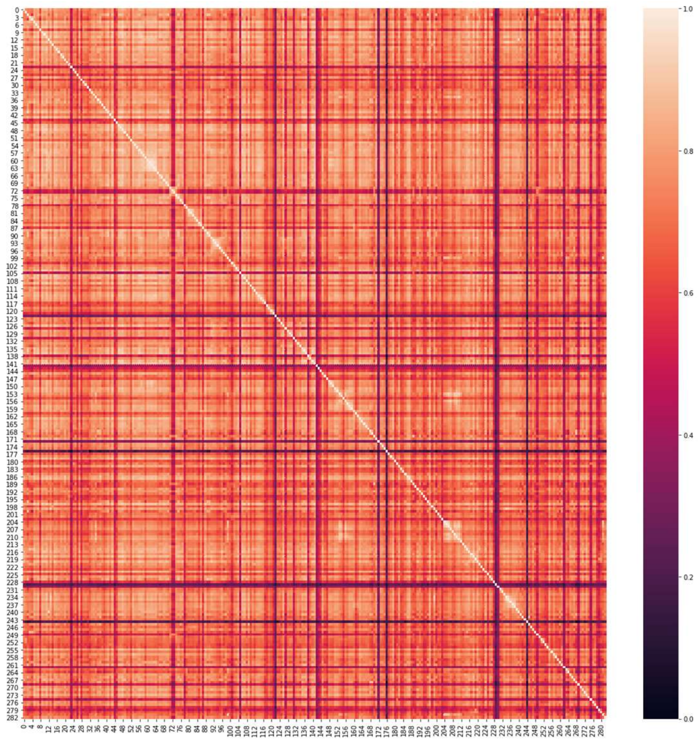


**Figure 15. Pearson correlation heatmap of time series**

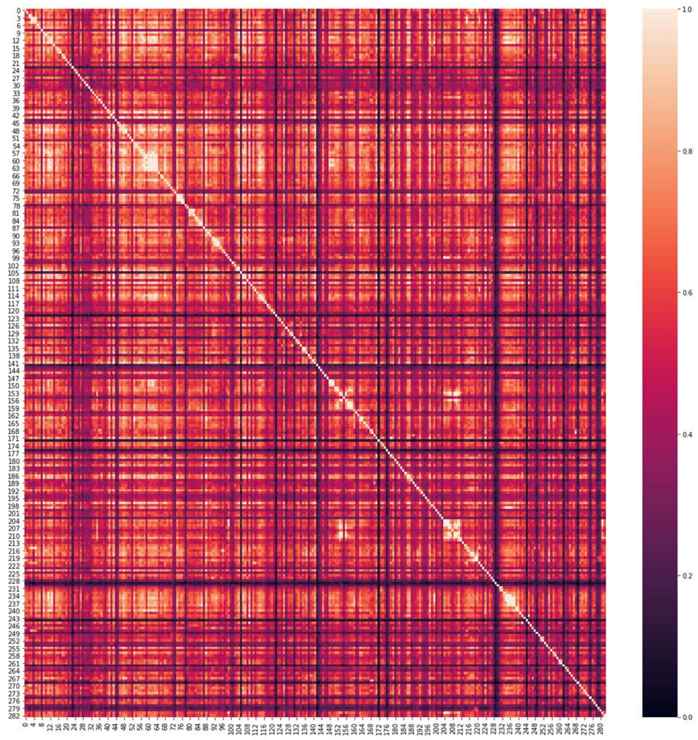


**Figure 16. Mutual Information heatmap of time series**

Further, the implementation results of the Fast DTW algorithm on the available 1-h times series, which gives an indication of which road sections' speeds are related to each other in terms of temporal evolution is seen in Figure 17. A smaller value (darker color) means a higher correlation in this case. Again, we may observe that a significant number of highly correlated pairs can be detected.

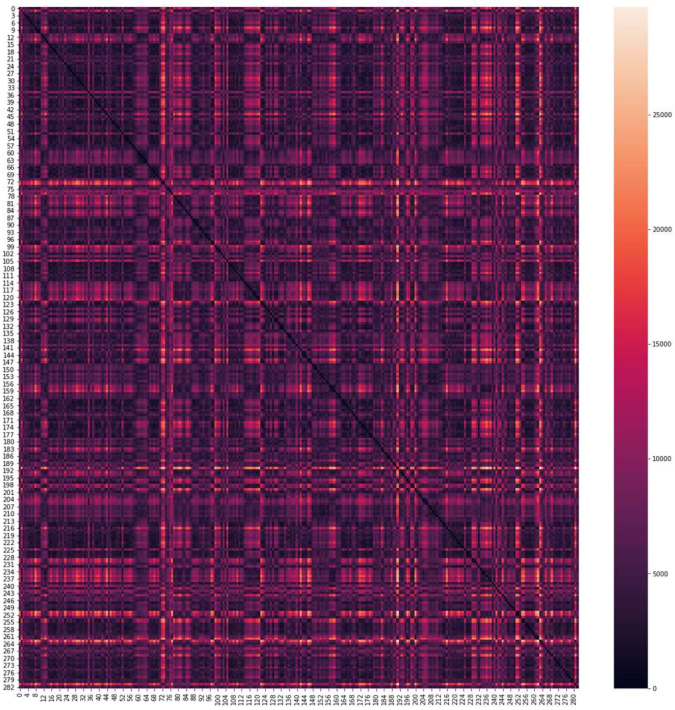



**Figure 17. Dynamic Time Warping heatmap of time series**

In order to compare the three metrics to each other in terms of predictability enhancement, the road section with ID number 28258922 on Open Street Map was selected as the target road section. The above road section is one of the most crowded road sections in Xi'an at the center of the city. The exact position of the road section is shown in Figure 18.

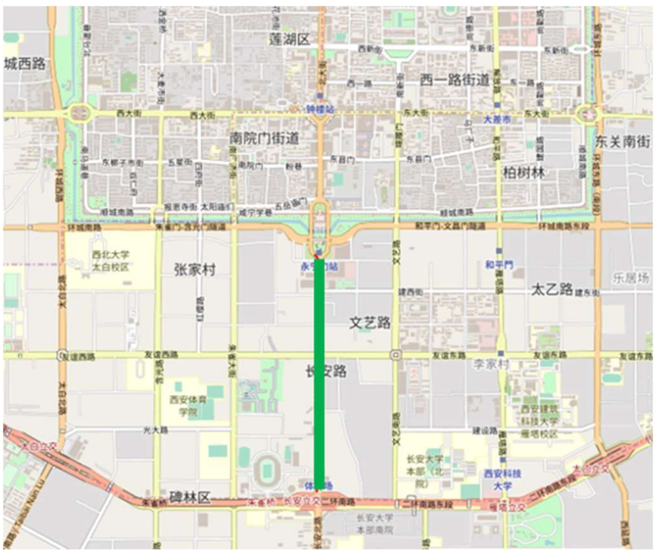
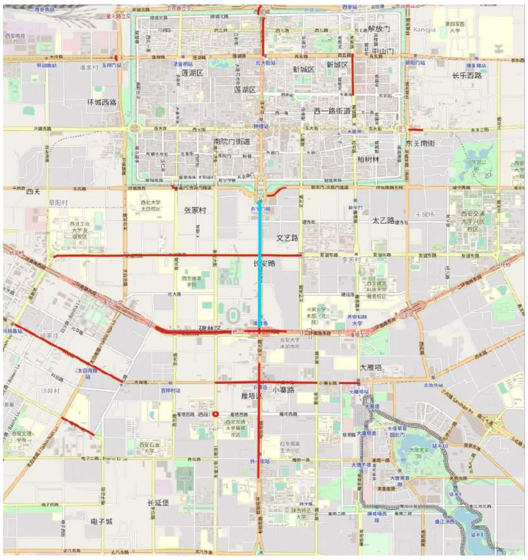
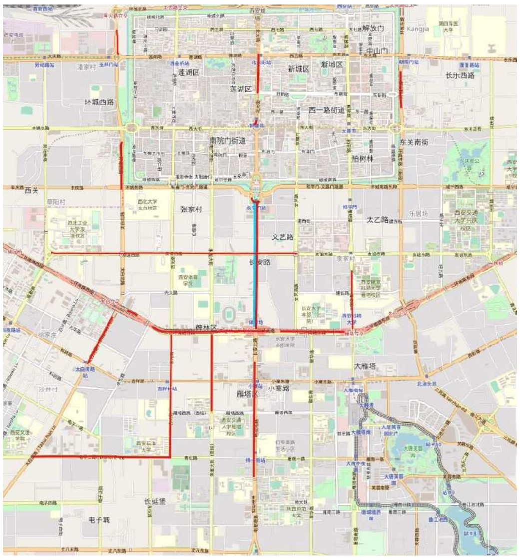



**Figure 18. Selected road section (green) on Xi'an City map.**

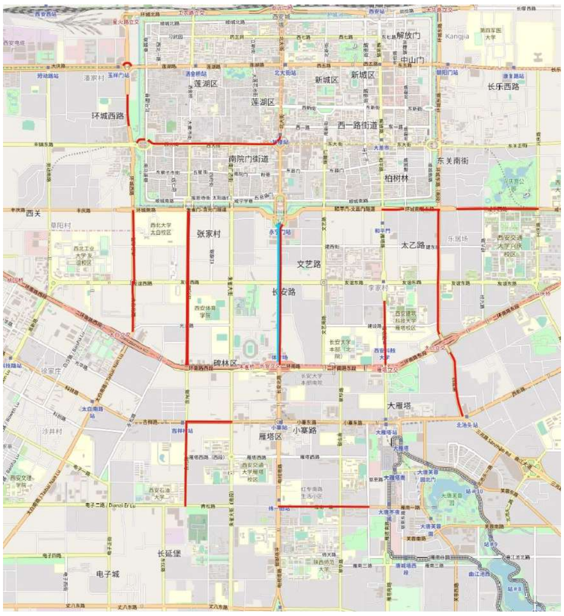
Figure 19 depicts the 20 most related road sections (red) to the selected section, in terms of Pearson's Correlation Mutual Information and DTW. It seems that the three approaches capture different spatial patterns on the same dataset. More specifically, according to the mutual information and Pearson correlation calculations, close perpendicular road sections are the most correlated to the selected one, as well as some upstream and downstream road sections. In addition, for mutual information, some further parallel sections are also detected. In contrast, with DTW, no perpendicular road section is detected, but instead some neighboring parallel ones (which are not detected by the other methods). The impact of these differences should be further investigated in terms of prediction accuracy.



*(a) Pearson's correlation*



*(b) Mutual information*



*(c) Dynamic Time Warping*

**Figure 19. The 20 most related road sections (red) to the selected section (blue), in terms of (a) Pearson's Correlation, (b) Mutual Information and (c) Dynamic Time Warping**

### 3.4.3 Comparison Between Mutual Information and Dynamic Time Warping

In order to compare each method's results, we proceed to develop two prediction models of the speed of the target road section. Both models are Bayesian Network Classifiers that assign the section's speed to three balanced classes: <20, 20-26, >26 km/h, which is a reasonable choice for signalized road sections, especially when we refer to travel speeds that include possible stops (e.g. upstream of a signalized intersection), as in our case. The choice to set up this problem as a classification one (and not regression) was made because the Bayesian model is only compatible with categorical data. The first model uses only input data from the twenty road sections with the highest Mutual Information value, while the second uses only the twenty road sections with the smaller values of DTW distance. The exact input space of each model consists of the speeds of the corresponding road sections the time step (one hour) before the one it is predicted.

As can be clearly seen in Table 6, the model using Mutual Information as the metric to select features outperforms the second one. The accuracy of the two models is 89.1% and 85.6% respectively. Hence, one can assume that using Mutual Information is a more accurate choice for the present application. This can be explained by considering again the definitions of the two metrics. Dynamic Time Warping is a measure of similarity between two time series, which is highly affected by the absolute size of the section's speed and only slightly by the time series' pattern. On the other hand, the estimation of Mutual Information considers the trend of the time series and the proportional variation of speeds rather than their absolute values.

**Table 6. Classification metrics of the two models developed**

| Metrics | Model 1 (Mutual Info) | Model 2 (DTW) |
|---|---|---|
| Accuracy | 89% | 86% |
| Recall (Sensitivity) | 89% | 86% |
| Precision | 89% | 86% |
| F1 - score | 89% | 86% |

### 3.4.4 The Impact of Spatiotemporal Dependencies in Short-Term Traffic Forecasting

In order to identify if the analysis conducted in the previous chapters is relevant and improves the prediction task, we evaluate the findings by comparing them to the performance of a Bayesian classifier that is "free" to make predictions using data from any road section, i.e., data from all road sections are provided as input to the model. In this case, each road section's contribution to forecasting is proportionate to its probabilistic relationship with the selected one. The classification results are summarized in Table 7. The accuracy of this model is 84.4%.

**Table 7. Classification metrics of Naïve Bayes model**

| Metrics | Model 3 (Naïve Bayes) |
|---|---|
| Accuracy | 84% |
| Recall (Sensitivity) | 84% |
| Precision | 85% |
| F1 - score | 85% |

Results indicate that the models presented in the previous chapter produce obviously better predictions. The first one, which uses the sections with the highest Mutual Information to the selected one, is performing noticeably better, while the second only slightly, but still better. This result highlights the usefulness of performing even a simple spatiotemporal analysis in order to select the most relevant input.

Moreover, the above procedure decreases the dimensionality of the problem, which is a very common issue when using Machine Learning algorithms. In the current case, we originally had 283 features (road sections), which is a significant high value. After performing spatiotemporal analysis feature selection, we used only the 20 most related. Furthermore, the above procedure reduces the computational resources needed, which is equally important.

## 3.5 Summary of Findings

According to the results presented earlier, analyzing large-scale spatiotemporal characteristics of a road network before proceeding to the development of prediction models is essential to produce accurate predictions. In addition, they provide better interpretation of the model's outcomes.

General use metrics, such as Pearson's Correlation, as well as more specific time series metrics such as Dynamic Time Warping distance and Mutual Information, provide a clear insight into the spatiotemporal relationships of the road sections of an urban road network. These relationships occur from the relative position of the road sections and the traffic flow each one serves; therefore, they provide explainable results.

In order to underline the importance of identifying spatiotemporal traffic patterns, a Bayesian Classifier was developed to classify speeds taking advantage of the identified spatiotemporal relations. Although the model that was used is not the best-suited choice for time series data and is relatively simple-structured, the improved performance by introducing the network-level spatiotemporal data is noticeable and important from the standpoint of accuracy and computational efficiency.

# 4 EFFICIENT TRAFFIC FORECASTING USING MULTI-RESOLUTION INFORMATION AND NETWORK-WIDE VEHICLE TRAJECTORY DATA

The analysis presented in the previous chapter revealed that, indeed, there are significant spatial relations between the traffic conditions of a road network, which can be detected by using suitable concepts from statistics and information theory. Moreover, it was shown that their impact on the forecasting accuracy is noteworthy. In this section, we focus on the temporal relations, by using a more suitable modeling structure, i.e. a Deep Learning model, whose input space consists of time series. Moreover, the data are aggregated in a higher temporal resolution and, in contrast with the previous section, we provide shorter-term predictions, which is a more challenging task.

## 4.1 Deep Learning and Deeper Understanding of Traffic Dynamics

Traffic temporal patterns are not easily recognized, especially when studying traffic's short-term behavior. Moreover, the relatively frequent shifts to boundary conditions, such as the onset of congestion, pose a difficult modeling problem to solve. To some extent, the key issue in traffic forecasting has always been the decoding of this complex behavior in order to improve its predictability. The complexity of the road environment, the control mechanisms, as well as the plurality of driving behaviors and non-recurrent conditions, lead to continuously changing temporal dynamics that make it extremely difficult for a single simple modeling structure to answer the critical question: How far back in time should the model look in order to accurately predict the future traffic conditions?

Ideally, to answer this question one should produce a time-variant environment to incorporate the evolving temporal dynamics of traffic flow and produce accurate short-term traffic forecasts. Time variance, if not induced externally to the process of prediction, can be internally represented in modern Deep Learning recurrent networks (Laña et al., 2019; Vlahogianni, 2009b). Time variant multi-scale designs of recurrent neural networks can exhibit increased efficiency in learning various scales of dynamics across multiple steps in time.

By applying advanced time-variant recurrent neural structures to short-term traffic flow we aim to capture the short and long-memory effects observed in traffic (Karlaftis & Vlahogianni, 2009; Yuan & Lin, 2017), internalize the issue of time varying temporal dependencies observed in traffic flow to the learning process and improve the prediction power over the classical recurrent neural networks.

The variation of Deep Neural Networks that is most often used in traffic forecasting is the Recurrent Neural Networks (RNN). Recurrent Neural Networks' main characteristic is the use of an internal memory, which holds an output from the previously processed input, that is used alongside the current input in order to produce a new output (Geron, 2017b). Consequently, unlike other Deep Neural Networks, RNN do not consider each input as independent of any other. This makes them suitable for input data that are temporally related, e.g. time series. From that point of view, RNNs are the most appropriate methodology to use in traffic flow prediction problems.

One of the most famous RNN architectures is the Long Short-Term Memory Networks (LSTM). Their structure is more complex than RNN's, allowing interaction between current and all previous inputs, and not only the last one like RNNs do (Olah, 2015). More details on the structure of the above type of models can be found in a previous section. LSTMs encompass certain limitations, namely not capturing long-term dependencies in favor of short-term and vanishing or exploding gradients during the training phase, that prevent the network from converging, especially when processing long time series.

The aim of this section is to evaluate the implementation of Dilated Recurrent Neural Networks introduced by Chang et al. (Chang et al., 2017) to the problem of short-term travel speed forecasting in order to overcome the quite common issues of LSTM, referred to earlier and investigate the impact of temporal relations in the forecasting accuracy.

## 4.2 Dilated Recurrent Neural Networks

Dilated RNN is a neural network architecture that is very simple, yet effective and efficient. It was developed in order to alleviate three major issues of classic RNN and LSTM networks (Chang et al., 2017):

- Vanishing and exploding gradients, which occur during the training phase of the model, not allowing the algorithm to converge and produce accurate results, especially when time series are long
- Capture of long-term dependencies
- Time-consuming training phase

Dilated RNN's architecture is independent of the type of the cells (RNN, LSTM and GRU cells can be used) and multi-layer. It is also characterized by its dilated recurrent skip connections and its use of exponentially increasing dilation, which are proven to perform better (Oord et al., 2016). The described architecture can be better understood by observing Figure 20, which illustrates the architecture with dilation or skip length equal to 4.

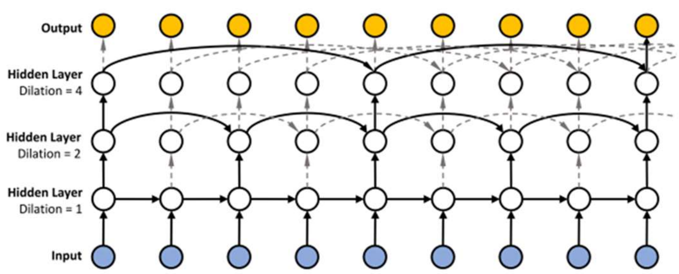



**Figure 20. Dilated RNN architecture**

More formally, let $c_t^{(l)}$ be the cell in layer l at time t. The dilated skip connection can be represented as:

$$c_t^{(l)} = f\left(x_t^{(l)}, c_{t-s^l}^{l}\right) \qquad (10)$$

where $s^{(l)}$ is the skip length or dilation of layer l, $x_t^{(l)}$ is the input to layer l at time t and $f(\cdot)$ denotes any RNN cell and output operation, e.g. simple RNN, LSTM, GRU.

## 4.3 Data Preparation

In this paper, we perform traffic conditions forecasting of a single road section using a Dilated LSTM Network, exploiting the speed's timeseries of the road sections of Xi'an's road network. The temporal resolution of the time series used for this application is 5 minutes.

In order for the data to be fed to the Dilated LSTM Network model, they were organized as a three-dimensional array, in which the first dimension is the total number of the available instances (5568 in our case), the second is the number of timesteps per time series (12 in our case, which multiplied by 5 minutes equals to 1 hour) and the third dimension is the number of attributes of each timestep (number of road sections, 497). More specifically we use a rolling window of one hour in order to match the speed of each timestep of the target section to the speeds of the rest sections from 5 minutes to 65 minutes before (the length of the rolling window is equal to 1 hour). This is the reason why each time series has 12 timesteps with a 5-minute duration each. The above practice results in 192 one-hour time series per section and per day (or 5568 in total per section), each of them assigned to a target value.

A network's discrete input is the time series of the speeds of the 497 road sections for one window. The output is the predicted category of the target section's speed.

Due to the existence of many missing values between 00:00 and 07:00, only data from 07:00 to 23:59 of each day were used in the following steps. In addition, the small number of missing values of the above period were imputed as the mean of the previous and the next timestep.

The road section with ID number 152428518 was chosen in order to perform predictions of its traffic conditions for a horizon of 5 minutes (one timestep). It belongs to the second ring of Xi'an city's center and serves hundreds of thousands of vehicles every day. The exact position of section 152428518 is shown in Figure 21.

It was decided that the target section's speeds should be classified into two categories, so, again, we have a classification problem. The two categories are less than 33 km/h and more than 33 km/h, which is the median of the section's speeds for the 29 days of the data. That way, we assume bad to moderate traffic conditions when the speed category is the first and moderate to good conditions when it is the second.

Regarding the Network's architecture, each LSTM cell has 256 neurons, the batch size we used is 160 and we trained the model for just 20 training epochs to avoid overfitting. The Adam optimizer was used and the Softmax cross-entropy with logits as the cost function. The learning rate was set to 0.0001. Finally, we selected three different dilations, 1, 2 and 4 timesteps, following the advice of (Oord et al., 2016), who have shown that the length of the dilations should be a power of 2. We could also use a dilation of 8 timesteps, as our time series have 12 timesteps, but the above scheme was proven to perform better.


**Figure 21. Position of the target section in Xi'an's road network**

## 4.4 Results

The first 22 days of our data were used for training the model, while the rest 7 (about 25%) constitute the test set. The evaluation of the model on the test set indicated that it achieves about 85.0% accuracy on unseen data, which is a decent result for data of that temporal resolution. The first category (<33 km/h) is in general predicted more accurately. Evaluation results are summarized below. Table 8 is the confusion matrix of the predictions and in Table 9 some metrics are presented.

**Table 8. Confusion matrix of Dilated LSTM network**

| | | Predicted | |
|---|---|---|---|
| | | <33 km/h | >33 km/h |
| **True value** | <33 km/h | 630 | 89 |
| | >33 km/h | 113 | 512 |

**Table 9. Classification report of Dilated LSTM network**

| | Category | | |
|---|---|---|---|
| | <33 km/h | >33 km/h | **Average** |
| **Precision** | 0.85 | 0.85 | **0.85** |
| **Recall** | 0.88 | 0.82 | **0.85** |
| **F1 score** | 0.86 | 0.84 | **0.85** |
| **Accuracy** | | | **0.85** |

Next, we compared the efficiency of the proposed Dilated LSTM Network in predicting the future traffic conditions of the selected road section to the performance of a simple LSTM network. A Dilated LSTM with a single dilation is practically, as already described, the classic implementation of an LSTM Network.

The classic LSTM Network hyperparameters were selected after performing an extensive Grid search. The LSTM cell consists of 120 neurons, the Adam optimizer was once again used and the Binary cross-entropy as the loss function. The batch size was 128 samples and the model was trained for 20 epochs. The accuracy of the second model was 81.9%, significantly lower than the Dilated LSTM's. The rest of the metrics also show that the use of Dilated LSTM Network instead of the classic architecture indeed produces more accurate results. In Table 10 below the metrics of the LSTM Network are presented.

**Table 10. Classification report of vanilla LSTM network**

| | Category | | |
|---|---|---|---|
| | <33 km/h | >33 km/h | **Average** |
| **Precision** | 0.88 | 0.77 | **0.83** |
| **Recall** | 0.75 | 0.89 | **0.82** |
| **F1 score** | 0.81 | 0.83 | **0.82** |
| **Accuracy** | | | **0.819** |

## 4.5 Summary of Findings

The results of the comparison between a Dilated LSTM Network and a classic LSTM Network in predicting future traffic conditions using network-wide data have shown that the Dilated one can enhance the predicting accuracy, without introducing extreme complexity to the process. The architecture of the proposed network is simple enough and easily anticipated as well. The main advantage of the proposed architecture compared to the simple LSTM is that it captures and exploits long-term dependencies that may influence future traffic conditions more significantly compared to shorter-term ones. In addition, it prevents the appearance of exploding or vanishing gradients, which is often when training Recurrent Neural Networks.

For the application presented in this section, it was attempted to exploit just the temporal relationship between the locations of the road network using a suitable Deep Learning structure, so no feature selection was conducted by determining the spatial relationship between the locations of the road network, as presented in the previous section. The predicting accuracy of 85.0% is decent, taking into account the complexity of the problem and the high dimensionality of the input space (497 road sections). However, it is assumed that the model has the potential to perform better on this problem if only relevant features were introduced in the input space. Therefore, in the next section, a novel spatiotemporal representation is proposed in order to enhance the predictability of the phenomenon.

# 5 MULTIMODAL NETWORK-LEVEL FORECASTING

Departing from simple feature selection, a meaningful and accurate representation of the road network, which refers to detecting and modeling the spatial and temporal relations between different locations, as well as to how they will be incorporated into the model's input space and learning process, is an efficient way to pass vital information to the forecasting model and improve its interpretability, while also decreasing the required complexity. Although most early studies assume that it is sufficient to deal with the traffic variables at different locations of the road network as independent features, it has now become widely understood that there do exist significant spatial and temporal relations, which, in addition, are dynamic, complex and nonlinear (K. Lee et al., 2021). While various approaches have been proposed for modeling the temporal dimension of the problem (i.e. statistical and data-driven time series analysis methods), capturing the spatiotemporal relations between different locations is a more complex and under-researched task.

Taking into account the above, in this section, we adopt the concept of Multiplex Networks (Magnani, Rossi, et al., 2021) from the research area of Social Network analysis to model the spatiotemporal relationships of the transportation network of the city of Athens, Greece and develop an innovative framework of feature selection based on Community Detection, a clustering concept from graph theory, on multilayer graphs. The transportation network is defined by loop detectors and metro stations, where measurements of traffic volume and ticket validations, respectively, are available.

The Multiplex Network proposed consists of 24 layers, each representing an hour of the day. Each layer is a graph, whose nodes are the loop detectors and the metro stations. A connection between two nodes of the same layer is established if they are statistically significantly correlated. So, similarly to social network analysis, two nodes are connected only if they are correlated, and not when they are geographically close or adjacent to each other. By representing the road network in this way, spatio-temporal correlations can emerge after performing community detection.

Thus, in this application, a multimodal and multisource data approach is followed, as, not only future traffic volume and subway (metro) demand are both predicted but are also exploited as features to predict one another. This approach of a multimodal framework, which identifies significant relations between the different modes and can also be extended to more than two modes, would have very important implications in a holistic traffic management scheme. Moreover, since the complexity of the models that are widely proposed and used in recent literature is high, this kind of feature selection allows the use of a simpler model, namely a Gradient Boosting Regressor with a simple input space, which is more suitable for a real-time application, as it can be efficiently retrained and calibrated even in real-time, if necessary, and does not require high computational power, making it suitable to operate on devices with limited resources, such as smartphones.

## 5.1 Theoretical Background

In this section, we aim to develop a forecasting framework that can operate in real-time, so a lightweight predictive model is necessary, supported by an accurate and meaningful representation of the spatial and temporal relations and an effective feature selection strategy. Besides, the representation of the transportation network and the input data is considered equally important to the modeling technique (K. Lee et al., 2021).

In order to extract spatio-temporal correlations, traffic and transit demand data can be modeled as a Multiplex Network. Specifically, in the Multiplex Graph, each layer corresponds to each hour of the day and has as many nodes as the total number of the detectors and the metro stations. The nodes' connection for each layer was based on the mutual information of the corresponding time series, which emerged as the most suitable metric according to the analysis presented in a previous section.

Afterward, by performing community detection, spatio-temporal correlations can emerge, since the algorithm searches for groups of vertices that share common properties. In this work, community detection is used for feature selection, by using as input features of the prediction model the measurements of the nodes which belong to the same community as the target node.

For the prediction task, a Gradient Boosting Regressor is used due to its relatively simple and shallow architecture, compared with Deep Learning models. Gradient Boosting is an ensemble algorithm, i.e. fits a predefined number of simpler predictors to the training set. In contrast to other ensembles, which use a simple or weighted average of the predictors' outputs, it integrates a greedy criterion, according to which each predictor is fitted on the residuals of its predecessor rather than the training set and the predictors are then added sequentially (Fafoutellis et al., 2022; Geron, 2017b). Although the predictors usually exploited are Decision Trees, other Machine Learning algorithms can be used as well. A prediction is, therefore, given according to the following equation:

$$\hat{y} = \sum_{i=1}^{N} a_i * d_i(\bar{x}) \qquad (11)$$

where N is the number of predictors, $d_i(\bar{x})$ is the output of each predictor and $a_i$ is a weight that is assigned to each predictor during the training phase of the model (Bonaccorso, 2017). Gradient Boosting is considered one of the most powerful Machine Learning models and remains very popular despite the rise of more complex deep learning structures, because it offers the advantages of more straightforward interpretability, efficiency and does not require as big amounts of data to get trained. Thus, a real-time traffic forecasting model that is based on spatiotemporal correlation can be built.

In this section, we will present some basic definitions of graph theory, which is a branch of mathematics useful for describing systems that are organized in graphs, such as transportation networks. Then, Multiplex Networks will be described, which consist of two or more graphs organized in distinct layers, used to represent different types of relations of the same agents. In network analysis, community detection is one of the fundamental problems, where the goal is to find groups of nodes that are, in some sense, more similar to each other than to the other nodes (H. Zhang et al., 2022). This concept and the corresponding algorithms are extended for Multiplex Graphs.

### 5.1.1 Graph Definitions

The central object of study in graph theory is the graph, which is a mathematical structure used to model relationships between objects and its definition is presented below.

**Definition 1 (Graph):** A graph $G = (V, E)$ is given by its vertices $i \in V$, where $V = \{v_1, ..., v_n\}$, and its edges $i, j \in E$, where $E \subseteq V \times V$. The edges $e \in E$ are pairs of 2 vertices $v_i$ and $v_j$ where $1 \leqslant i, j \leqslant n$.

A distinction is made between undirected edges, the ends of which connect two vertices symmetrically, and directed edges, the ends of which connect two vertices asymmetrically with an arrow. When the graph consists exclusively of undirected edges it is called undirected, otherwise, it is called directed (Hartmann & Weigt, 2005).

An efficient way to mathematically represent a graph $G = (V, E)$ is the adjacency matrix. The latter is a square matrix $A \in R^{|V| \times |V|}$ which is defined as: $A = (a_{i,j})$; $a_{i,j} = 1$, if $(i, j)$ is an edge of $G$ and 0 otherwise (Harary, 1962).

There are occasions where a cost should be assigned to the connections between nodes. In this case, a weighted graph is used, where a numerical weight $w_{i,j}$ is assigned to the corresponding edge and usually represents a cost. In the adjacency matrix of a weighted graph, the values in position $(i, j)$ are equal to the weight of the edge $(i, j)$.

### 5.1.2 Multiplex Networks

Multiplex Networks are complex graph structures that have layers in addition to nodes and edges, which simple graphs have. A Multiplex Network consists of two or more interconnected graphs lying in distinct layers. Each layer has a different connectivity structure within it. This kind of network allows for a more realistic approach to the study of the interaction of individuals, who can communicate through different types of channels (Amato et al., 2017). This kind of structure is widely used in public transportation systems, social network analysis and research on infectious diseases (Kivelä et al., 2014). Graph theory is suitable for studying these topics, as individual agents (e.g. locations of the city) can be represented as nodes and the relationships between them as edges.

This work adopts the definition of a Multiplex Network as a network consisting of many layers where a node does not connect to another node across layers, but only within layers and its definition is below (Magnani, Rossi, et al., 2021):

**Definition 2 (Multiplex Network):** A Multiplex Network is a tuple (A, L, V, E) where A is a set of actors, L is a set of layers, $V \subseteq A \times L$ and $E \subseteq V \times V$ where $\forall(\alpha_1, l_1, \alpha_2, l_2) \in E : l1 = l2$.

In this kind of network, $E$ is restricted to intra-layer edges, that is, an edge is allowed only if its nodes are at the same layer. An example of a Multiplex Network is shown in Figure 22, where $L = l_1, l_2$, $A = \alpha_1, ..., \alpha 5$ and $(\alpha_1, l_2, \alpha_2, l_2)$ is an example of an edge in $E$. The definition of the Multiplex Network allows some of the nodes not to be present in some layers. In some cases, when the term multiplex is used, it is assumed that all nodes are present in all layers. To avoid confusion, in this work, we explicitly refer to a kind of graph that is defined as the node-aligned

Multiplex Network. In this most strict form of a Multiplex Graph, all nodes will be present at all layers. More formally (Bródka et al., 2018):

**Definition 3 (Node-aligned Multiplex Network):** A node-aligned Multiplex Network is a Multiplex Network (A, L, V, E) where $\forall n \in A, l \in L : (n, l) \in V$.

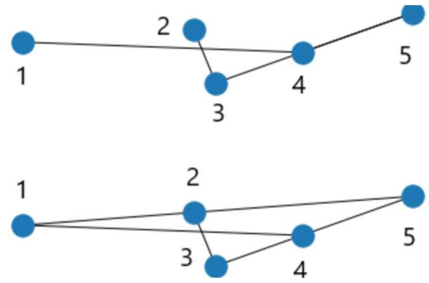



**Figure 22. Node-aligned Multiplex Network**

### *5.1.3 Community Detection*

Real networks display big inhomogeneities, i.e. vertices having low degrees (few connections) coexist with some vertices with high degrees. So, the distribution of edges is globally and locally inhomogeneous. Structural differences may be observed within special groups of vertices such as high or low concentrations of edges between these groups. This feature of real networks is called community structure, or clustering. So, communities, which are as well called clusters or modules, are groups of vertices that share common properties and/or play similar roles within the graph (Magnani, Hanteer, et al., 2021).

In the case of Multiplex Graphs, we distinguish the community detection algorithms into actor-overlapping or actor-disjoint. That is, if there is at least one actor belonging to more than one clusters, then we call the clustering actor-overlapping, otherwise, it is called actor-disjoint. Also, we distinguish the methods into semi-pillar and pillar. A multiplex community is called semi-pillar on layers $L' \subseteq L$ if for each actor $\alpha \in A$ in the community all nodes in $(\alpha, l) \in V : l \in L'$ are included in the community. When $L' = L$, we have a pillar community.

Community detection algorithms for Multiplex Networks can be grouped into three typical main classes, described below:

- Flattening: The first approach consists of simplifying the Multiplex Network into a graph by merging its layers, using a so-called flattening algorithm, and then applying a community detection algorithm for simple graphs. The algorithms in this class are only able to identify pillar communities, and some communities may emerge because of edges spread on different layers that would not constitute a community on an individual layer, because of the flattening process.
- Layer by Layer: This kind of methods first process each layer and merge the results of the processing. As a consequence of the layer-by-layer community detection step, these methods include actors in the same community only when they are part of the same community in at least one layer. Also, due to the merging of layer-specific communities, these methods can in principle only identify pillar communities.
- Multilayer: The third class of algorithms operates directly on the Multiplex Network. These methods can detect non-pillar and actor-overlapping communities, unlike the

previous 2 classes. That is, they will span on multiple layers and will not have the same actors on different layers necessarily in the same community (Magnani, Hanteer, et al., 2021). Instances of these methods are the Multi-Layer Clique Percolation Method (ML-CPM), Infomap, Locally Adaptive Random Transitions (LART), Louvain and Multi-Dimensional Label Propagation (MDLPA) (Bródka et al., 2018; Magnani, Hanteer, et al., 2021).

For the purposes of this work, the Louvain algorithm is exploited. The Louvain community detection algorithm is based on the optimization of modularity as the algorithm progresses. Modularity is a scale value between −0.5 (non-modular clustering) and 1 (fully modular clustering) that measures the relative density of edges inside communities to edges outside communities and optimizing this value results in the best possible grouping of the nodes of a given network (Nguyen et al., 2008).

In this community detection algorithm, communities are found by optimizing modularity locally on all nodes, then each small community is grouped into one node and the process is repeated. For a weighted graph, modularity is defined as:

$$Q = \frac{1}{2m}\sum_{i,j}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j) \qquad (12)$$

where $A_{ij}$ represents the edge weight between nodes i and j, $k_i$ and $k_j$ are the sums of the weights of the edges attached to nodes i and j respectively, m is the sum of all of the edge weights in the graph, $c_i$ and $c_j$ are the communities of the nodes and δ is the Kronecker delta function where $\delta(x, y) = 1$ if $x = y$, 0 otherwise (Newman, 2006).

The goal is to maximize this value efficiently, using the algorithm's two phases that are repeated iteratively. In the first phase, each node in the network is assigned to its own community. Then, for each node i, the change in modularity is calculated for removing it from its own community and moving it into the community of each one of its neighbors j. This value is calculated for all communities to which node i is connected, and it is placed into the community that resulted in the greatest modularity increase. If no increase is possible, i remains in its original community. This process is applied repeatedly and sequentially to all nodes until no modularity increase can occur.

In the second phase of the Louvain algorithm, the nodes of each community are grouped and a new network where nodes of the previous phase are the communities is built. The links between nodes of the same community are now represented by self-loops on the new community node and links from multiple nodes in the same community to a node in a different community are represented by weighted edges between communities. Once the new network is created, the first phase can be re-applied to the new network (Nguyen et al., 2008).

The above description of the Louvain algorithm applies to the case of simple graphs and it is useful to generalize this process to Multiplex Networks that are studied in the present work. Multiplex Network nodes are in the same community when they have similarities and tend to share common properties. Therefore, revealing the community structure in a Multiplex Graph can provide a better understanding of the overall functioning of the network (Magnani, Hanteer, et al., 2021).

Using R programming language and the multinet package (Magnani, Hanteer, et al., 2021), community detection can be implemented. The goal is to find community structures across layers, where vertices in different layers can belong to the same or a different community despite corresponding to the same actor. This community detection algorithm is also based on modularity optimization. Multiplex modularity is a quality metric function that takes higher values if most of the edges are between vertices in the same community and if vertices corresponding to the same actors are also often in the same community. Modularity for Multiplex Networks is defined as:

$$Q_m = \frac{1}{2\mu} \sum_{i,j\, sr} \left[ \left( a_{ijs} - \frac{k_{is} k_{js}}{2m_s} \right) \delta(s,r) + \omega \delta(i,j) \right] \delta(\gamma_{is}, \gamma_{jr}) \qquad (13)$$

where i, j are actors, s, r are layers, $\alpha_{ijs}$ is 1 if i, j are adjacent on layer s, $k_{js}$ is the degree of actor i on layer s, μ is the number of pairs of vertices either adjacent on α layer or corresponding to the same actor, $m_s$ is the number of edges in layer s, $\gamma_{is}$ is the community to which actor i on layer s is assigned to, δ is the Kronecker delta, and ω is a weight; when the same actor belongs to the same community on two different layers, then $Q_m$ is increased by ω.

Omega is a parameter that takes values from 0 to 1. Setting higher values of omega will result in communities that will span on multiple layers and will consist of the same actors, since this way the value of modularity increases. On the other hand, with omega set to 0, having the same actors on different layers in the same community does not contribute to modularity (Bródka et al., 2018).

## 5.2 Implementation and Results

### *5.2.1 The Dataset*

To develop the traffic forecasting model, traffic flow and transit demand data were used, which stand for the number of passing vehicles and passengers, respectively, from different points of the road network over a period of one hour. The data were collected from an open-source database that was developed by the Greek Government and the Region of Attica and includes the hourly traffic volume passing from more than 400 loop detectors and the number of passengers embarking on the subway (metro) train at each of the 63 stations. The data are available for download for academic purposes. For this paper, we used data from 10 months (January to October of 2021) from 113 of the most significant loop detectors that are located around the center of the city of Athens, as well as the demand of all the metro stations. Figure 23 shows the exact location of the loop detectors in green and the metro stations in red.

In addition to traffic flow data, in this paper, it was decided to include public transport data, and especially metro stations demand, which is a means of public transport that serves a large part of the population daily. It is widely understood that the two datasets have significant correlations with each other, which can be exploited to provide more accurate predictions. Moreover, the latter has not been extensively studied in recent literature and only a few studies exploit such heterogeneous data.

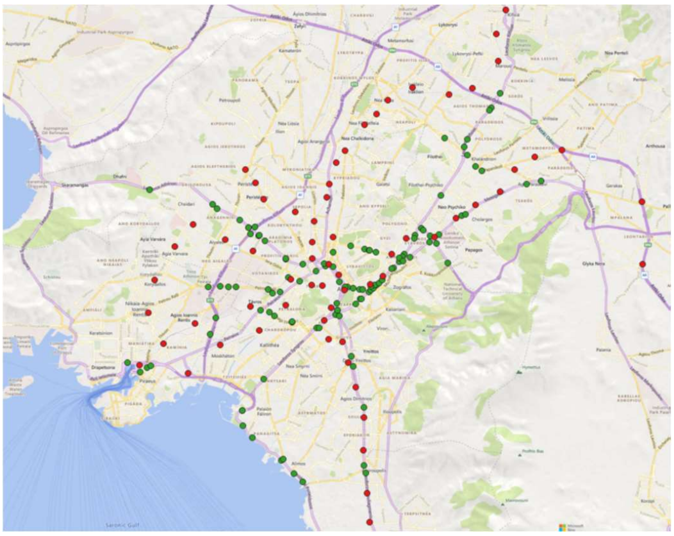
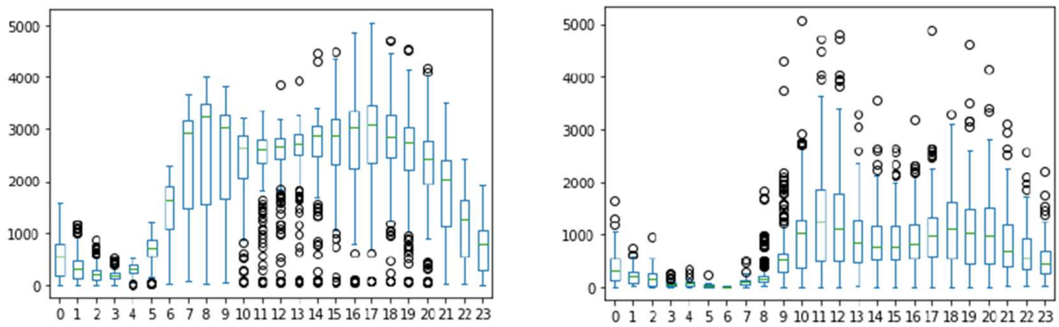


**Figure 23. Geographic distribution of loop detectors (green) and metro stations (red)**

To gain a better understanding of the data, indicative boxplots of the measured flow and demand per hour are presented for the loop detector on Kyfisos Avenue, and Panepistimio metro station, which are two of the most crowded, in Figure 24. Considering the values from all detectors, high values arise in the morning until the afternoon, and from 7 PM onwards there is a gradual decrease in the number of passing vehicles. In most locations the maximum values are observed at 8-9 o'clock, then there is a slight reduction of passing vehicles early in the afternoon but at 5 PM we observe higher values again. A similar trend is observed in the metro data where most passengers use public transportation between 10-11 AM and 5-6 PM.




(a) Hourly traffic volume at Kyfisos Ave.

(b) Number of ticket validations in Panepistimio metro station

**Figure 24. Hourly Boxplots for passing vehicles and number of ticket validations at indicative locations**

*5.2.2 Multiplex Network of Athens Transportation System*

The transportation network of Athens (detectors and metro stations) can be represented as a Multiplex Graph of the spatiotemporal relations of the nodes, in order to detect communities and obtain valuable insights into the spatial and temporal relations that define it. The construction of the graph was based on the idea that each layer would correspond to each hour of the day, so the graph consists of 24 layers. Each layer consists of 176 nodes, each of which represents a loop detector or a metro station. Next, to define the edges of each of the layers, we got the time series of the demand of each node at the corresponding time of each day (e.g. every day at 9 a.m. for the 9$^{th}$ layer) and calculated the mutual information between the time series of all nodes for the same hour. If the mutual information between two nodes is higher than 0.5, which indicates a significant correlation, an edge is created between them. Of course, we scaled the 2 types of data in order to place them in the same structure, as they refer to different measurements. That way we model the spatio-temporal correlation between the nodes. We use the above technique to create the edges of the graphs, instead of the actual connections, because, at a network level, it is not very straightforward to determine the connectivity of the nodes, especially between heterogeneous nodes (metro stations and loop detectors) and, most importantly, this method leads to distinct graph structure at each layer, that reflects the dynamic nature of spatial relations.

Next, a comparative analysis of the 24 layers is conducted, where each layer corresponds to an hour of the day, to better understand the data and the structure of the multiplex graph that emerged from them. We can compare these layers using several different approaches and concepts from graph theory. Firstly, the dissimilarity between degree distributions is computed using the Jeffrey dissimilarity function, where the higher the values, the more dissimilar the two layers are (Bródka et al., 2018). Figure 25 shows the heatmap of Jeffrey dissimilarity of the layers (darker color indicates higher dissimilarity). One may observe that the graphs corresponding to the early morning hours are quite different from those corresponding to the evening and late evening hours. Furthermore, we observe that the degree distributions of 7 AM until 3 PM are similar, pointing at similar network structures and, thus, similar traffic conditions.

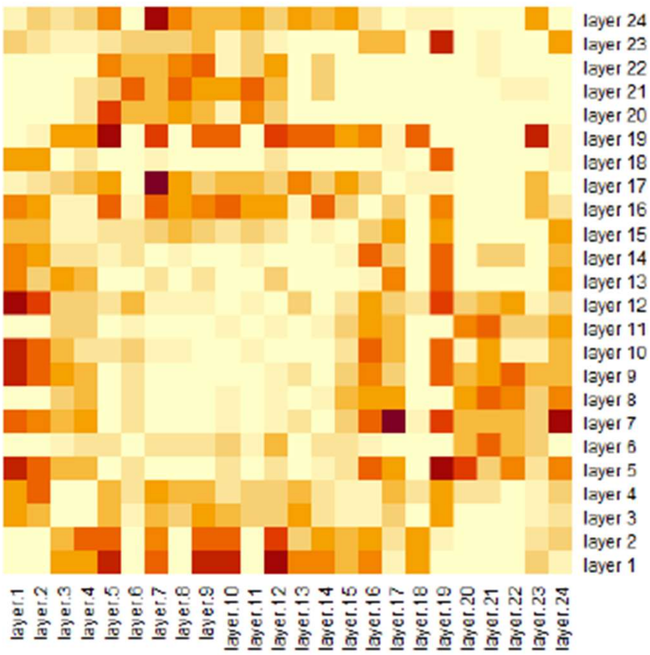



**Figure 25. Heatmap with Jeffrey dissimilarity of the layers**

Moreover, to assess layer similarity, the correlation between the degrees of the nodes is computed. For this task, the Pearson (or linear) correlation between the degrees of actors in each couple of layers is estimated. This value ranges between 1 and −1, where values close to -1 and 1 indicate a high negative and positive correlation, respectively, and values close to 0 low correlation. It is important to note that the correlation only depends on the number of existing edges for each node at each layer, and not on which actors are adjacent; they can be the same or different actors (Bródka et al., 2018; Magnani, Rossi, et al., 2021). Figure 26 shows with dark color the layers with high correlation. Indicatively, some of the layers which correspond to 10:00, 11:00, and 12:00 have a high correlation as well as those which correspond to 4:00 and 5:00.

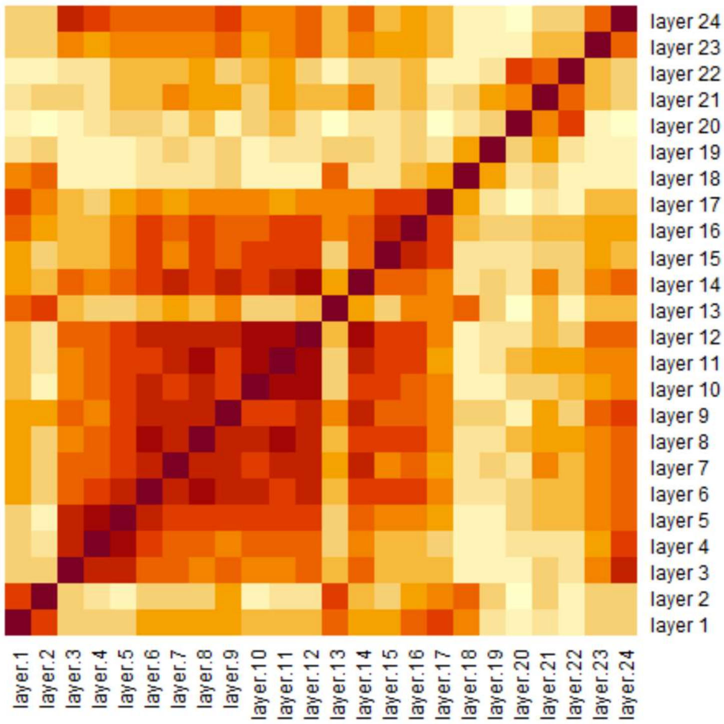



**Figure 26. Heatmap with Pearson Correlation of the layers**

Finally, it is interesting to evaluate to what extent actors are adjacent to the same other actors in different layers, by checking the amount of overlapping between edges in the two layers, which will be 0 if no actors that are adjacent in one layer are also adjacent in the other and 1 if all pairs of actors are either adjacent in both layers or in none, which is expressed by the Jaccard metric and the corresponding heatmap is presented in Figure 27. Practically, by looking at Jaccard overlapping on the actors we see which couples of layers share many edges. The latter answers the question of whether the spatial relations between nodes change over time. By observing the heatmap, and comparing it to Figure 26, we can see that, in most cases, layers with a similar number of connections actually have the exact same connections, since the two figures are quite similar. This indicates that the specific couples of layers that have several edges in common are also organized according to similar dynamics (Magnani, Rossi, et al., 2021).

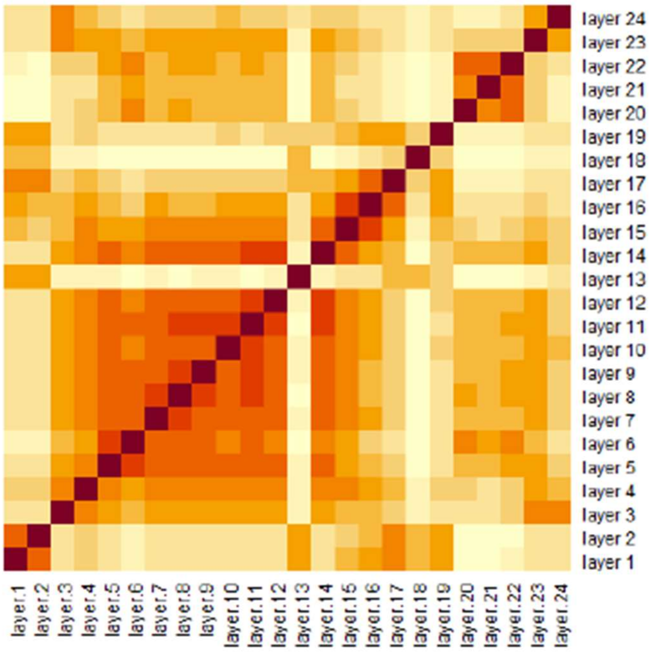



**Figure 27. Heatmap with Jaccard overlapping of the edges**

### *5.2.3 Detecting Communities*

After the construction of the multiplex graph, and the exploration of some spatial and temporal properties, we proceed to detect communities of spatially and temporally related nodes. It should be reminded that the existence of an edge between two nodes is very important when estimating the modularity metric and detecting communities of nodes. It is desirable that each community contains different nodes at each time of the day (i.e. layer) and each instance of a node does not necessarily belong to the same community, in order to capture the temporal relations as well.

The community detection algorithm that was selected is Louvain. The parameter omega's value was set to 0.001. It is reminded that this parameter takes values from 0 to 1 and setting higher values of omega will result in communities that will span on multiple layers and all instances of each actor would belong to the same community, since this way the value of modularity increases (Bródka et al., 2018). In contrast, if the parameter is selected as 0 the communities will contain nodes from different layers. We have chosen a rather low value of omega, so that different actors from different layers are included in the same community, in order to extract more accurately the spatiotemporal relations.

From the different groups that were created, since the nodes that belong to the same community have many structural similarities and common characteristics, it is deduced that the number of passing vehicles or validated tickets in one area of Athens at a specific hour may be affected or significantly related with other areas at the same or different hour. This is a useful feature and is used later to build a forecasting model.

By applying the Louvain community detection algorithm, we end up with 7 communities which are presented in Figure 28.

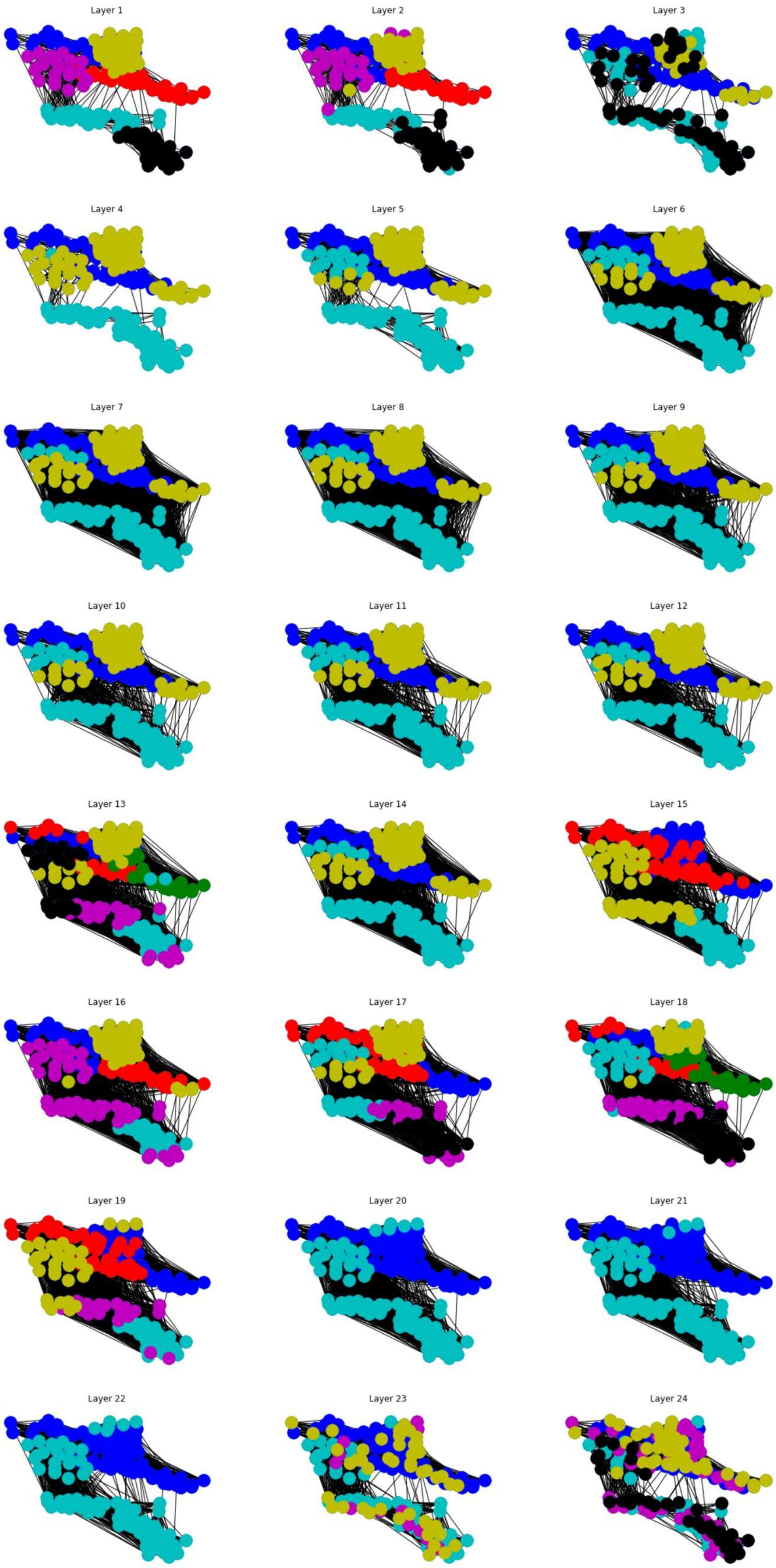



**Figure 28. Detected communities**

Different colors correspond to different communities. As long as nodes of different layers belong to the same community, it can be assumed that they have similar structural properties, i.e. they are connected to instances of the same nodes. Therefore, nodes of different layers are assigned to the same community if they play the same role, each at the corresponding layer and, since the nodes represent time series and their connections correspond to statistical relations, nodes belonging to the same community illustrate similar traffic and demand patterns, so it is reasonable to use them as features in a prediction model. Moreover, due to the existence of nodes from different layers in the same communities, we may assume that not only spatial, but also temporal relations are highlighted. Actually, it indicates the existence of a significant statistical correlation between nodes at different times of the day, which can be attributed to similar traffic patterns emerging at the specific nodes during different times of the day, such as the occurrence of congested conditions. In the case the graphs and the connections between the nodes were generated based on the geographical proximity of the nodes, we would end up with clusters of nodes that do not have common spatial and temporal demand characteristics, which is not desired in the present study, where the community detection is performed to detect the most relevant features to use as the prediction model's input.

### *5.2.4 Model Development*

After detecting the 7 communities presented earlier, a model for each node and for each time of the day can be developed to predict the demand at the corresponding detector or station, using as features the latest values of demand of the rest of the nodes that belong to the same community. For example, the demand at node i at 9:00 can be predicted from the demands of nodes j at 7:00 and k at 22:00 (of the previous day) if these three nodes belong to the same community. So, the number of features of each model is equal to the size of the community to which the requested pair (device, hour) belongs, and the value of each feature corresponds to the measurements of the previous 24 hours. The above data are exploited to predict the traffic conditions or the demand (for metro station nodes) with a predicting horizon of 1 hour.

Then, for each device-time pair, a Gradient Boosting Regressor was developed with the following values of hyperparameters: the number of estimators (decision trees) for each ensemble is 100, the learning rate 0.1 and the maximum depth of each decision tree is 3, with a minimum number of 2 samples required to split an internal node and 1 for a leaf node. The train-test data split was random in which the test set consisted of 33% of the records. When adjusting the hyperparameters of the model, the validation set consisted of 20% of the training set. Finally, a total number of 4224 models were trained, which is the number of the node-hour pairs.

### *5.2.5 Baseline*

Since the available dataset consists of long (10 months) and heterogeneous (traffic and transit) time series, a reasonable and suitable direction is the development of LSTM models that are able to handle and exploit such a large amount of data. LSTM networks are the state-of-the-art in modeling time series data, especially heterogeneous ones, and, thus, it is preferred compared to other Deep Learning models. Therefore, as a baseline, an LSTM network is developed for each target point (loop detectors and metro stations respectively). The exact architecture of the networks includes an LSTM layer followed by two fully connected (dense) layers. The hyperparameters of each model were defined after an extensive grid search process, where each model architecture was evaluated on the cross-validation set and are as follows: The LSTM layer consists of 32 to 64 neurons and the first dense layer of 4 to 16 neurons (different for each point), while the second dense layer, which gives the output of the model, of 1 neuron. The ReLU activation was selected for the two dense layers and the Adam optimizer with a learning rate of 0.0001. The Mean Squared Error was the loss function, the models were trained for 100 epochs and the batch size was 32. The optimal "look back period" (previous timesteps exploited) was 12 steps, which corresponds to 12 hours as well. In contrast with the previous sections, the output variables have continuous values, so we have a regression problem.

## 5.3 Results

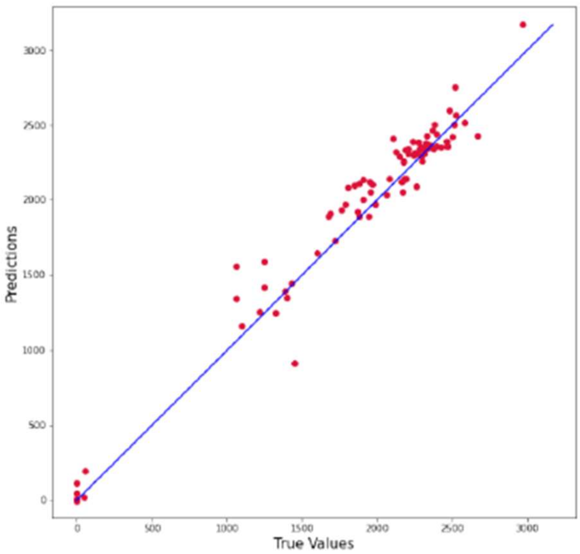
Following the above strategy, predictions on the test sets for all pairs (device, hour) were produced. For the sake of brevity, we selected to present in more detail an evaluation of the Gradient Boosting Regressors corresponding to 9:00 and 17:00, as these are the peak hours for the road network of Athens. So, from now on, the results of the models for the entire transport network (all detectors and stations) corresponding to these hours are presented. The models were evaluated by the calculation of the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE) of the test set. As it is presented in Table 11, the models achieve a satisfying overall accuracy, which justifies the use of the proposed methodology. The MAE and MAPE values refer to the average error of the predictions concerning each node and are presented separately for traffic flow and transit demand prediction, as well as for the two different time periods. The overall MAPE value, which is about 0.095 (9.5%), refers to all predictions.

It is evident from the table that traffic demand can be predicted more accurately from the models, which may be because the majority of the data refer to loop detectors. We can also observe that both variables are better predicted for the afternoon peak (17:00), achieving a significantly improved MAPE. The reason for this may be that there is a higher deviation of the data values during the morning hours.

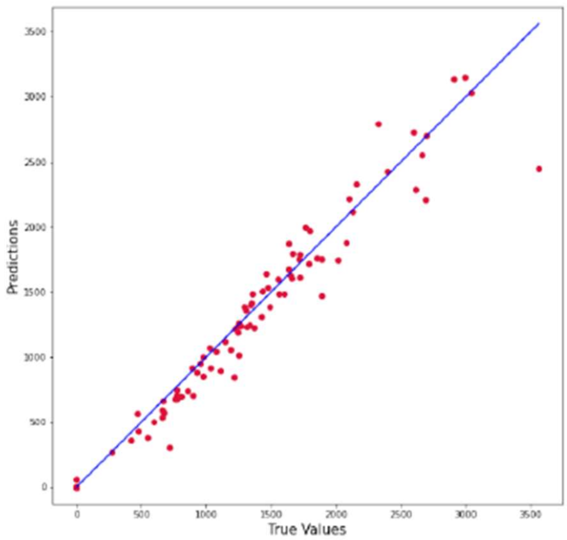
**Table 11. Accuracy metrics of trained models**

| Hour of the day | Metric | Traffic | Transit |
|---|---|---|---|
| **9:00** | MAE | 83.11 | 30.10 |
| | MAPE | 10.4% | 13.1% |
| **17:00** | MAE | 99.45 | 31.51 |
| | MAPE | 7.5% | 9.5% |
| **Overall MAPE** | | **9.5%** | |

Moreover, in Figure 29 the scatter plots of the predicted and actual values of two indicative nodes (one loop detector and a metro station) are presented. One may observe that the points are close to the $x = y$ line and no systematic error can be detected.




(a) Detector of Vasilissis Sofias street




(b) Evangelismos metro station

**Figure 29. Predicted versus Actual values scatter plots of indicative nodes**

Concerning the baseline model, it also achieves a good performance, especially for predicting future traffic conditions. However, the overall MAPE value is about 0.188 (18.8%), which is significantly higher than the proposed model's. In Table 12 the detailed error metrics of the LSTM network are presented. Again, the MAPE and MAE values refer to the average value for all loop detectors and metro stations, respectively, as a different model for each of them was developed.

**Table 12. Baseline LSTM model evaluation metrics**

| Metric | Traffic | Transit |
|---|---|---|
| MAE | 133.14 | 52.02 |
| MAPE | 16.8% | 23.0% |
| Overall MAPE | **18.8%** | |

Furthermore, the overall MAPE distribution for both approaches and for all nodes is presented in Figure 30 below. It is evident that the values of the MAPE have a relatively high variation, which is reasonable, due to the high number of different models developed for the entire network (176 detectors and metro stations). The median value of the errors for the proposed Multiplex-Gradient Boosting model is 0.08, meaning that half of the models have even lower error, and 25% of the models have a MAPE value lower than 0.06, with the minimum error being 0.03. Interestingly, the maximum error is 0.18 for the specific approach, which is lower than the average error of the state-of-the-art LSTM approach. Also, the superiority of the proposed approach is established by the fact that the minimum error of an LSTM model (0.12) is also higher than the average error of the first one. As is usually the case with Deep Learning models, if higher amounts of data become available, the LSTM networks may achieve improved performance, however, the latter fact can be also thought of as an additional drawback of the specific method.

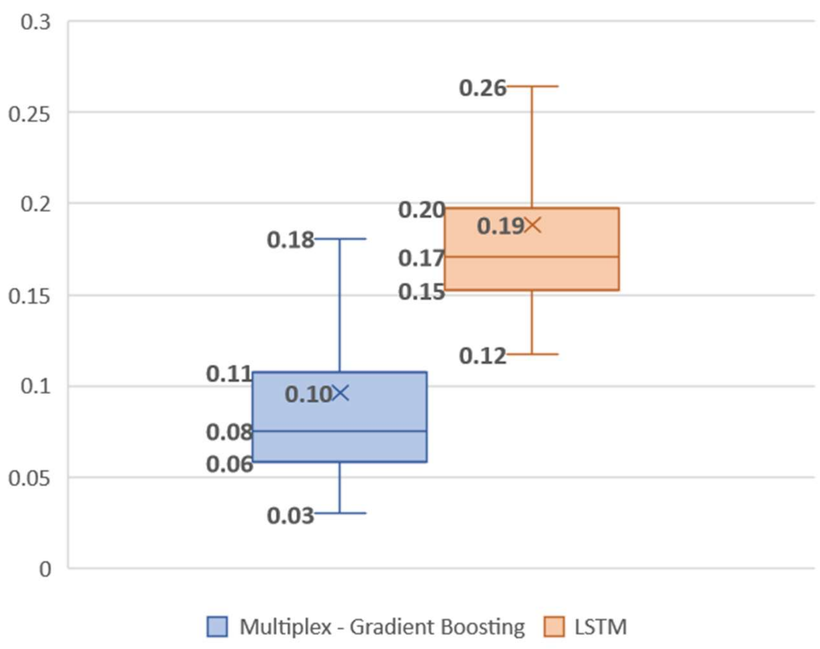



**Figure 30. Box plot of overall MAPE for traffic and transit demand for both modeling applications and for all nodes**

Finally, it is very important to mention that the training time and the computational resources needed for the development of the LSTM networks are incomparably higher than those of the Multiplex-Gradient Boosting. Indicatively, the grid search and training process for the LSTM models of each of the 176 detectors and stations lasted for about half an hour (80 hours in total), while for all the Gradient Boosting Regressors for all detectors-hours and stations-hours pairs, respectively, the same processes took 1 hour in total.

## 5.4 Summary of findings

In this section, a Multiplex Network representation of a road network is proposed along with a novel feature selection framework, that is adopted from the research area of Social Network analysis. More specifically, the road network is represented as a multilayer graph, by connecting nodes that are statistically correlated, instead of geographically close. By modeling the data in this way, the spatial and temporal relations of its nodes can be highlighted. Also, a community detection algorithm provided an efficient feature selection framework that enhanced the models' performance, by significantly decreasing the input space's dimensionality. Moreover, a multimodal and multisource dataset was exploited, including traffic flow and public transport demand data. Significant correlations have been discovered between the two datasets and the data were fused to provide accurate predictions of all kinds of nodes. The implications of such a multimodal forecasting framework in an inclusive and unified urban traffic management scheme are significant. Also, the findings indicated that the resulting shallow Machine Learning model is more suitable for providing accurate predictions compared to a Deep Learning model.

The entire methodology is suitable for real-time applications, as the feature selection process can be executed offline and the model development for specific nodes is extremely lightweight, as well as efficient in terms of time and computational resources required and can even be trained in real-time in cases that the latter is necessary, e.g. extreme traffic conditions, and, moreover, on devices with limited computational and memory resources, such as smartphones. It should be noted that the real-time operation of the model would be necessary for a lower predicting horizon (e.g. a few minutes). Due to the unavailability of such data, we use a 1-hour horizon, but the methodology can be applied, as it is, for shorter horizons. Therefore, an effective representation and feature selection strategy, such as the proposed, can improve the prediction accuracy, while reducing complexity and computational and time resources requirements.

In this work we attempted to showcase the importance of the goals of the analysis instead of the tools used, especially when driven by domain-specific theory. We showed that there are always (implicit) assumptions in all modeling approaches, and that complex, nonlinear Deep Learning structures have both advantages and limitations and frequently simpler models facilitated by domain-specific representations and feature selection strategies give as good results as complex ones. However, many unanswered questions still remain. The first relates to the data characteristics used and, especially, the temporal granularity and how it may affect the spatial dependence and, further, the accuracy of the model. Research on unimodal and univariate analysis thus far indicates that aggregation eliminates long memory characteristics and variance heterogeneity; which leads to smoothing traffic variation and creating a time series structure that has reduced sensitivity to changes in traffic (Vlahogianni & Karlaftis, 2013). It is of interest to understand how data granularity may affect multimodal correlations and their evolution in space and time.

Next, we believe that testing complex or less complex modeling structures has been systematically neglected, leading to inconclusive modeling procedures and biased predictions. Interestingly, in recent years, there has been an obvious trend in traffic modeling of researchers departing from simpler, linear, low-dimensional and turning to complex, nonlinear, high dimensional systems, largely because of significant increases in computing power, but still not necessarily justified by logic or fundamental research needs. The synergy with statistics is deemed necessary to achieve actionable forecasting models that retain the properties of transferability and interpretability.

# 6 EXPLAINABLE NETWORK-WIDE TRAVEL TIMES PREDICTION

The modules presented so far, use network-wide input. Taking advantage of the spatiotemporal relations between the locations of the road network, they provide predictions of the future traffic conditions at specific target locations. For each target location, a separate model is required, with its corresponding input and hyperparameters' values.

One of the identified issues of recent literature in traffic forecasting that remains open is a forecasting model that would be able to provide network-wide forecasts in a multitask environment, i.e., predict the traffic conditions at all examined locations simultaneously, using a single model and leveraging more than one data sources, corresponding to each location. Such a model would have very significant implications in traffic management, as it would provide all the necessary information at once. Moreover, in order to be actionable, this model should be both accurate and efficient, in terms of training time and computational and data requirements, which cannot be guaranteed even for single-location models.

In this chapter, we present one of the first efforts in the literature to predict simultaneously the travel times for crossing 30 significant road sections around the city center of Athens, using as input traffic volume data from loop detectors. As discussed in Section 2.8, most attempts from recent literature propose the solution of using innovative computing schemes (e.g., Edge Computing and Federated Learning), which focus on reducing the complexity and data and computational requirements of the forecasting models, to the task of network-wide forecasting. However, in this case, one model is needed for each location of interest, which, regardless of their low complexity, require different input data and frequent fine-tuning and calibration. In contrast, the approach presented in this section exploits one forecasting model, which is also relatively simple, to forecast 30 outputs at once, being significantly more efficient than using 30 separate models.

## 6.1 Methodological Approach

We develop a prediction framework of the travel times, using loop detector data that correspond to the average flow of the hour before each travel time measurement. For example, hourly traffic volumes recorded between 8am and 9am will be used to predict the average travel times recorded between 9am and 10am. The above problem is treated as an instance of multitask prediction, i.e., instead of using the corresponding (e.g., nearest) loop detectors for predicting each travel time separately, we provide the input data of all loop detectors to the model and allow it to learn more complex relations between each target variable (route travel time) and neighboring or distant detectors and expect the future travel times of all routes as output. In order to evaluate the forecasting results, in addition to classic error metrics, we provide explanations of the output, based on the estimation of the Shapley (or SHAP) values.

### 6.1.1 Model Explainability with SHAP Values

For the better understanding of the effect that each input variable has on the output variables, the estimation of the SHapley Additive exPlanations (SHAP) values takes place. The SHAP value is computed by carefully perturbing input features and seeing how changes to the input features correspond to the final model prediction. The main advantage of SHAP values is that the difference between the prediction and the average prediction is fairly distributed among the feature values of the instance. More specifically, the SHAP value is the average expected marginal contribution of one input feature after all possible combinations have been considered. SHAP value closer to zero means the feature contributes little to the prediction whereas SHAP value away from zero indicates the feature contributes more. Concerning the mathematical definition, the Shapley value for the $i^{th}$ input variable is:

$$\phi_i(v) = \frac{1}{|N|!}\sum_R \left[v\left(P_i^R \cup \{i\}\right) - v\left(P_i^R\right)\right] \qquad (14)$$

where the sum ranges over all |N|! orders R of the input variables, and $Pi^R$ is the set of input variables that precede input variable i in the order R. Here, we can imagine that there are |N|! different ways of adding input variables to the coalition one at a time; the SHAP value is just the mean of the additional payout input variable i brings across all these different ways (Aas et al., 2021).

For example, for each given instance of the test set, a positive or negative SHAP value is assigned to each input variable, reflecting the effect of the specific variable on the output. By adding the SHAP values of all input variables for a specific instance to the average value of the output, the exact value of the output for the specific instance emerges.

### 6.1.2 Data

We combine the traffic volume data of the loop detectors from the city of Athens with travel time data retrieved from the popular maps and navigation service Google Maps. The travel times refer to the time (in seconds) for crossing 30 of the most crowded and critical road sections during the morning and afternoon peak hours, namely from 8 am to 10 am and 12pm to 7pm, for each day between June and December of 2021. The main specifications of the routes and descriptive statistics of the travel times are presented in Table 13.

As the routes do not have the same length, their travel times were divided with their length, so that they are comparable and reflect the prevailing traffic conditions (congested or uncongested). The total number of records is 1110. As one may observe, most routes have an average crossing time around 4-5 minutes per kilometer for those close to the city center and 2 minutes per kilometer for sections at highways of the perimeter that serve as entrances or exits to the city center. Moreover, the first category also have higher deviation of their travel times and maximum values, spanning around 1000 s/km in some cases, indicating their varying nature and the extremely high intensity of congestion at the specific locations. In Figure 31, we present the considered routes.

**Table 13. Main routes' specifications and travel time descriptive statistics**

| a/a | Street name and direction (from-to) | Length (m) | Mean travel time (s/km) | Standard deviation | Maximum (s/km) |
|---|---|---|---|---|---|
| **1** | Panepistimiou (Vas. Sofias - Patision) | 1001 | 181.2 | 63.7 | 753.9 |
| **2** | Akadimias (Patision - Vas. Sofias) | 1339 | 235.0 | 91.7 | 622.1 |
| **3** | Stadiou (Aiolou - Vas. Georgiou) | 862 | 343.8 | 175.7 | 829.9 |
| **4** | Athinas (Ermou - Stadiou) | 667 | 273.2 | 66.0 | 595.7 |
| **5** | Athinas (Stadiou - Ermou) | 666 | 260.8 | 47.2 | 469.0 |
| **6** | Vas. Sofias (Vas. Konstantinou - Panepistimiou) | 1214 | 350.8 | 166.3 | 887.3 |
| **7** | Vas. Amalias (Ath. Diakou - Panepistimiou) | 969 | 229.2 | 113.7 | 688.3 |
| **8** | Patision (Alexandras - Stadiou) | 980 | 205.4 | 49.7 | 448.0 |
| **9** | Pireos (Kolokinthous - Omonoia Sq.) | 600 | 290.1 | 137.7 | 962.6 |
| **10** | Syngrou Av. (Vas. Amalias - Frantzi) | 850 | 149.8 | 28.7 | 422.1 |
| **11** | Pireos (Kolokinthous - Iera Odo) | 500 | 219.7 | 139.5 | 1067.2 |
| **12** | Syngrou Av. (Frantzi - Vas. Amalias) | 817 | 222.0 | 155.5 | 1108.9 |
| **13** | Alexandras (Kifisias - Patision) | 2667 | 214.3 | 74.5 | 473.2 |
| **14** | Kallirois (Petmeza - Ardittou) | 350 | 280.3 | 101.8 | 819.2 |
| **15** | Patision (Ioulianou - Chalkokondili) | 600 | 149.9 | 37.9 | 525.3 |
| **16** | Patision (Stournari - Ioulianou) | 400 | 217.7 | 54.7 | 545.6 |
| **17** | Kifisias (Alexandras - Panormou) | 850 | 167.6 | 81.6 | 502.9 |
| **18** | Kifisias (Panormou - Alexandras) | 850 | 173.3 | 67.6 | 663.9 |
| **19** | Mesogion (Katechaki - Kiprou) | 950 | 124.5 | 60.2 | 544.7 |
| **20** | Vouliagmenis (Arditou - Ilia Iliou) | 900 | 112.4 | 68.9 | 910.7 |
| **21** | Vouliagmenis (Ag. Konstantinou - Pirronos) | 900 | 87.4 | 14.1 | 193.7 |
| **22** | Vouliagmenis (Pirronos - Ag. Konstantinou) | 900 | 77.3 | 11.0 | 189.3 |
| **23** | Ilioupoleos (Ilia Iliou - Arditou) | 850 | 185.3 | 123.0 | 923.8 |
| **24** | Cephissus (Posidonos - Pireos) | 1000 | 58.8 | 75.3 | 918.7 |
| **25** | Cephissus (Pireos - Posidonos) | 1000 | 51.8 | 26.9 | 406.7 |
| **26** | Cephissus (Athinon - Moudrou) | 1000 | 50.9 | 10.6 | 164.4 |
| **27** | Cephissus (Moudrou - Athinon) | 950 | 100.3 | 95.5 | 868.4 |
| **28** | Posidonos (Niarchos - Cephissus) | 1300 | 52.0 | 9.9 | 196.7 |
| **29** | Posidonos (Amfitheas - Alimou) | 1200 | 116.0 | 36.8 | 545.0 |
| **30** | Posidonos (Alimou - Amfitheas) | 1200 | 179.8 | 63.8 | 508.3 |

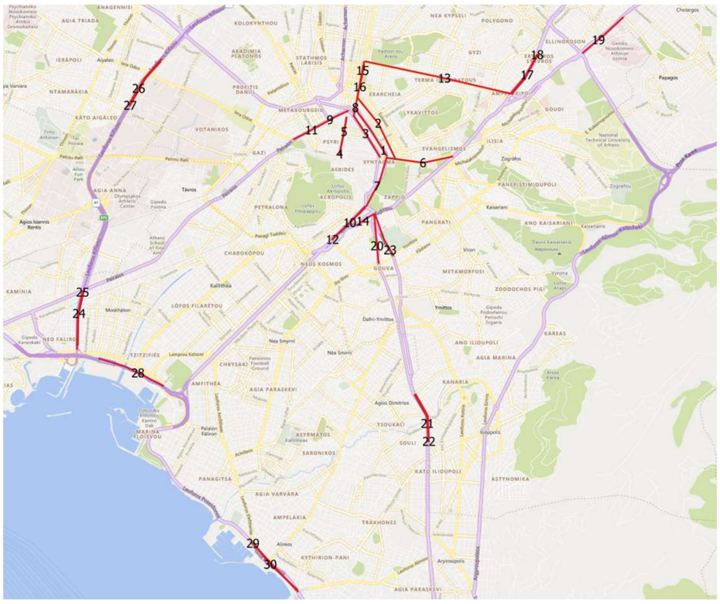


**Figure 31. Geographical distribution of routes**

The traffic dataset used in this application is the one with the hourly Athens loop detectors measurements presented in the previous section as well, from which 190 locations that geographically cover the broader area of the city of Athens were selected. Similarly to travel time data, the traffic volume data were also scaled between their minimum and maximum values. The measurements of all 190 loop detectors are passed to the model as input in order to implement a network input – network output scheme, without the need for any data preprocessing step. The 190 considered loop detectors are presented in Figure 32.

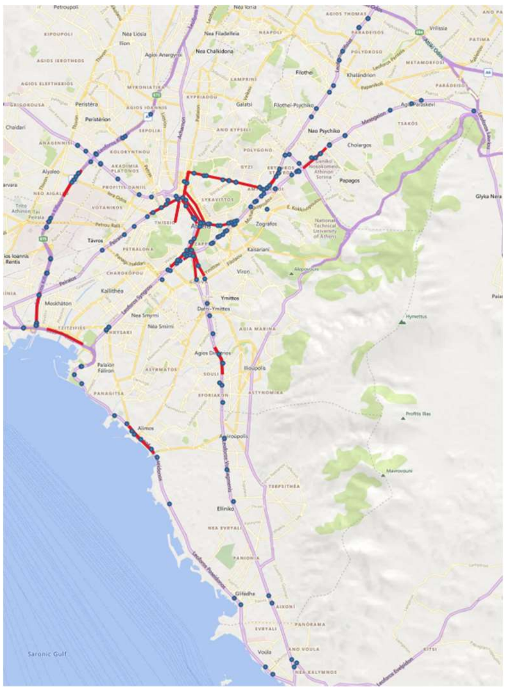


**Figure 32. Actual locations of loop detectors**

## 6.2 Implementation and Results

### 6.2.1 Problem Setup and Modeling

A multivariate input and output framework is utilized in this section for the prediction of travel times for crossing 30 road sections at the broader area of the Athens city center. The forecasting model that is exploited is a relatively shallow Feed Forward Neural Network with 3 hidden layers, consisting of 128,128 and 64 neurons respectively. The activation function that was used is the Relu (Rectified linear unit) and Mean Squared Error (MSE) as the loss function, while the learning rate was set to 0.001 and the batch size 8. The above architecture and the values of the hyperparameters have emerged after an extensive grid search process.

The input of the model is the hourly measurements of the selected loop detectors (one measurement per detector), while the output includes all 30 travel times that correspond to the next hour of that of the input. Although the model that is used is a rather simple one compared to the complexity of the output space, it is trained efficiently and it is expected to perform sufficiently well. Besides, it was selected to use as input only the latest measurement of each loop detector and not time series of measurements, which would, unnecessarily, increase the complexity and the data and computational resources requirements.

The available dataset was separated into training and test sets, with 25% of the available data corresponding to the test set. The model was trained for 300 epochs.

### 6.2.2 Forecasting Results and Explanations

In this section, the model's accuracy is assessed and the corresponding error metrics are presented. The Mean Absolute Error's mean value is 23.1 s/km. The prediction errors for all routes are presented in detail in Table 14.

The model performs quite decent, in general, achieving an average Mean Absolute Percentage Error (MAPE) of 12.3%, while there is a significant number of routes whose travel times are predicted with a MAPE lower than 9% (25% quartile). Moreover, the highest observed MAPE values are 20.4% and 19.4%, while the 75% quartile of the MAPE is 15.8%, which is considered acceptable.

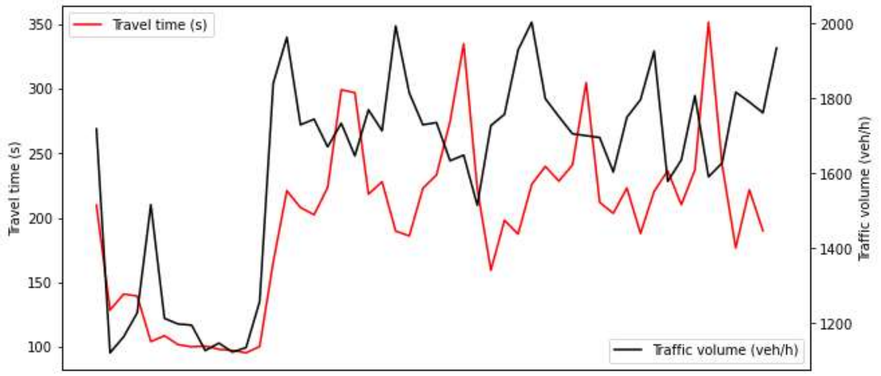
Regarding the model's training efficiency, it should be noted that the fitting process lasted for less than 6 minutes, while relatively simple, typical single output RNN models (like those presented in the following sections) may require more than 15 minutes (450 minutes for all 30 locations) and more complex Deep Learning models from recent literature may be trained for hours. Besides, the proposed model does not require very large amounts of data, e.g., long sequences of traffic measurements, which also accelerates the training process. In addition, the proposed model exhibits increased efficiency during operation as well, as it provides all required output at once, while the alternative approach requires several models, with different input variables and calibration to run.

**Table 14. Forecasting accuracy for each route (test set).**

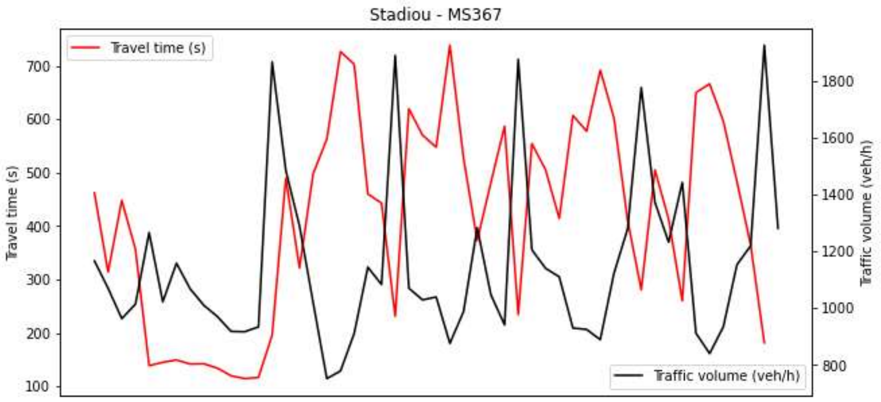
| Route | MAPE |
|---|---|
| Panepistimiou (Vas. Sofias - Patision) | 9.8% |
| Akadimias (Patision - Vas. Sofias) | 11.3% |
| Stadiou (Aiolou - Vas. Georgiou) | 15.9% |
| Athinas (Ermou - Stadiou)/ (Stadiou - Ermou) | 9.4%/ 8.6% |
| Vas. Sofias (Vas. Konstantinou - Panepistimiou) | 19.4% |
| Vas. Amalias (Ath. Diakou - Panepistimiou) | 17.0% |
| Patision (Alexandras - Stadiou) | 8.3% |
| Pireos (Kolokinthous - Omonoia Sq.) | 15.8% |
| Syngrou Av. (Vas. Amalias - Frantzi) | 6.3% |
| Pireos (Kolokinthous - Iera Odo) | 17.3% |
| Syngrou Av. (Frantzi - Vas. Amalias) | 20.4% |
| Alexandras (Kifisias - Patision) | 11.1% |
| Kallirois (Petmeza - Ardittou) | 10.3% |
| Patision (Ioulianou - Chalkokondili) | 9.2% |
| Patision (Stournari - Ioulianou) | 10.4% |
| Kifisias (Alexandras - Panormou) / (Panormou - Alexandras) | 16.0%/ 13.4% |
| Mesogion (Katechaki - Kiprou) | 13.9% |
| Vouliagmenis (Arditou - Ilia Iliou) | 12.9% |
| Vouliagmenis (Ag. Konstantinou - Pirronos) / (Pirronos - Ag. Konstantinou) | 4.9% / 5.8% |
| Ilioupoleos (Ilia Iliou - Arditou) | 15.2% |
| Cephissus (Posidonos - Pireos) / (Pireos - Posidonos) | 15.0%/ 13.7% |
| Cephissus (Athinon - Moudrou) / (Moudrou - Athinon) | 8.0%/ 17.3% |
| Posidonos (Niarchos - Cephissus) | 5.7% |
| Posidonos (Amfitheas - Alimou) / (Alimou - Amfitheas) | 13.8% / 14.0% |
| **OVERALL AVERAGE** | **12.3** |

In order to gain some insights into the predictions, we first plotted the time series of the predicted travel times and the measurements of the loop detectors that are installed at the same road section (closest to the middle of the route) for 4 indicative routes (Figure 33). As one may see, higher values of travel time are expected right after significant and abrupt drops in the traffic flow, which correspond to the emergence of congested conditions. In contrast, the lower values of travel time coincide, most of the time, with higher values of traffic volume. This outcome seems reasonable and indicates the suitability of the model for the specific task and its good fit.

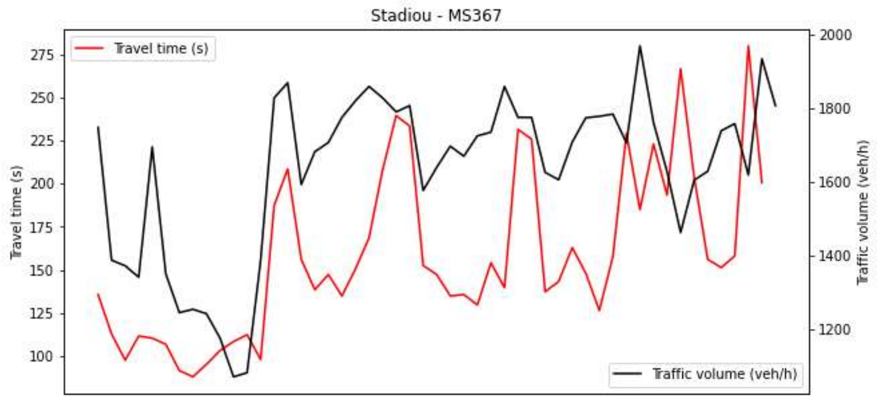
SHAP values are a model-agnostic interpretation method that can be effectively used for Deep Learning model. Its outcomes provide useful insights into the model and the prediction mechanisms, but they can also be used to assess the trustworthiness of the model: explanations that are consistent with what is in theory expected are a sign of a well-fitted model. In addition, this kind of explanation can be utilized to provide, at least, a rough estimation of future travel times in cases where no data are available or extreme and non-recurrent conditions occur.



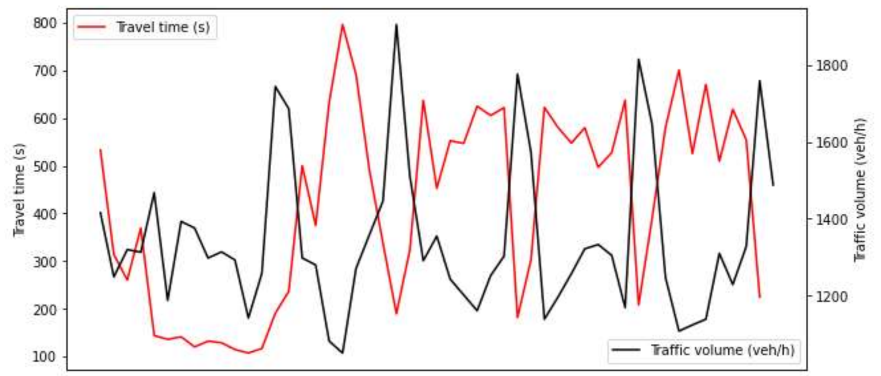

(a) Alexandras (Kifisias - Patision)




(b) Vas. Sofias (Vas. Konstantinou - Panepistimiou)




(c) Kifisias (Panormou - Alexandras)




(d) Stadiou (Aiolou - Vas. Georgiou)

**Figure 33. Time series of traffic volumes and travel times for indicative locations**

For the purposes of this work, we estimate the effect of the traffic conditions at neighboring and distant locations on the travel times. More specifically, the SHAP values of all locations for each route are calculated and, for the sake of brevity, we present the results for the four routes and their two corresponding loop detectors with the highest average SHAP values that are depicted in Figure 34.

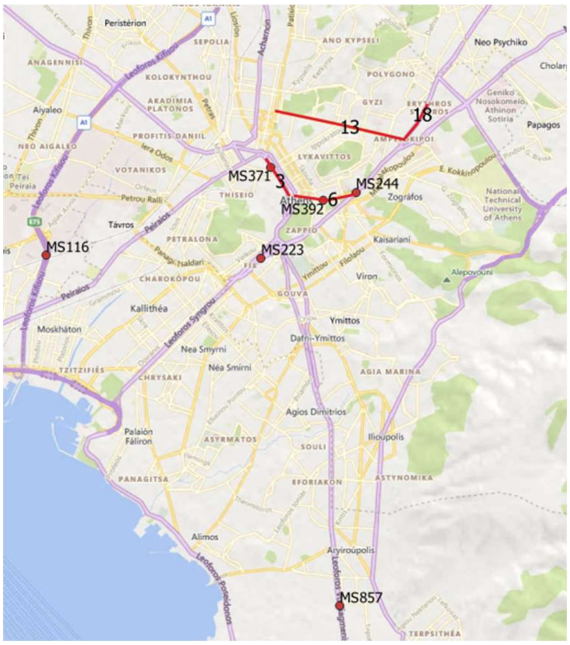



**Figure 34. Selected indicative routes and most influential loop detectors**

In Figure 35 we present the relation between the traffic flow at specific locations and the estimated travel time, i.e., the scatter plot of the SHAP values associated with the estimated travel time of the indicative routes and the traffic flow of the two locations that emerged as the most influential. As one may observe, for most pairs of routes-locations, there is a common pattern: when a low value of traffic volume is observed, which indicates congested conditions, the SHAP value of travel time increases significantly, i.e., higher travel times are expected. Especially for routes that are closer to the city center (e.g., Stadiou and Vas. Sofias), there are also higher SHAP values observed in general, which are also indicative of the higher deviation in their travel times and the significance of the effect of congestion's emergence. For example, depending on the traffic volume passing from specific locations, the travel time for the specific road sections may be increased or decreased by 100 s/km, which is a significant contribution. Moreover, the relation between the SHAP values and the traffic volume is almost linear in most cases. Finally, a very interesting finding is that in some cases, the SHAP values associated with more distant locations are higher than those of locations inside the corresponding route, which implies a significant relation between them, the knowledge of which is also essential for traffic management purposes. Therefore, it is indicated by the results that the utilization of a multitask setup is beneficial for the model's accuracy, as such traffic patterns can be detected. More specifically, it seems that, despite the multivariate output, which would raise concerns about the model's fitting and accuracy, there are significant correlations between the travel times and several loop detectors, and not only those that are close to each route, indicating important short- and long-term traffic patterns, which allow the model to achieve an acceptable performance.

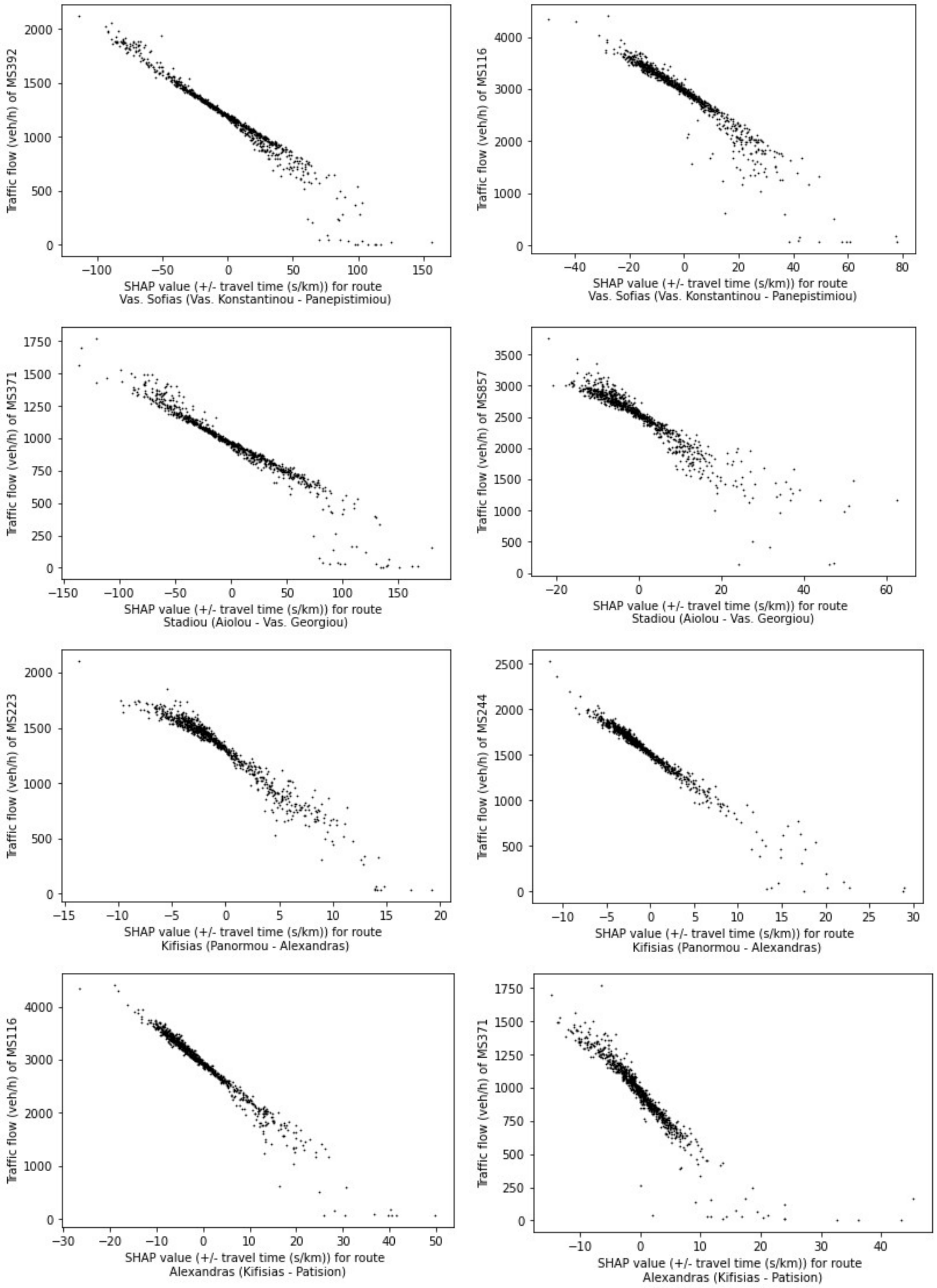



**Figure 35. SHAP values vs traffic flow at the most influential locations for indicative routes**

## 6.3 Summary of Findings

In this section a first attempt is presented to forecast multiple travel times at a network level, using a single multitask model. The neural network that was employed achieves a very good performance, while being efficient in terms of data requirements and training time. The explanations of the results that were extracted using SHAP analysis give meaningful insights that are very useful for traffic management purposes and indicate the trustworthiness of the predictions. The results also indicate that there are significant short- and longer-term relations between the traffic conditions and the travel times of the selected routes; thus, a multitask learning approach such as the one presented can take advantage of them effectively and efficiently.

The model that was used, as well as the input space, are deliberately selected to be relatively simple compared to the output space, which consists of the travel time of all the selected routes, to keep the approach efficient, explainable and actionable. However, more complex models and representations of the input space may lead to more accurate predictions, and their exploitation is left to future research.

Considering the increased efficiency of a model that can provide valid network-wide predictions, its importance in predictive traffic management and decision-making is high and the lack of similar approaches was previously highlighted as one of the main open challenges for developing actionable models. The present approach is a first step towards this direction. Other implications of the proposed framework include user information systems and data generation, e.g., when travel time data from third-party services are not available. Finally, the model can be incorporated into a traffic simulation platform and be used to efficiently and effectively estimate the travel times of the network's road sections, which are essential for the path assignment process.

# 7 CAUSAL DEEP LEARNING FOR SHORT-TERM TRAFFIC FORECASTING

## 7.1 The Need for Causal Relations

The information-theoretic method (mutual information), as well as the other statistical methods that were utilized in previous sections to detect spatiotemporal relations between the locations of the road network, suffer from a serious drawback: they imply the existence of a statistical relation, which may be significant, but do not guarantee causality. Consequently, the relations detected may just exist in the specific dataset, i.e., they are not generalizable and may not apply to extreme conditions or/and future data. Therefore, they limit the model's trustworthiness and explainability.

Model explainability refers to models whose outcome is understandable to a human, but also includes the proof of developing an interpretable model, e.g., by selecting an appropriate structure and input variables (Ribera & Lapedriza, 2019). Furthermore, it is related to the model's transparency, trustworthiness, and fairness. Due to the complexity of their structures and the number of hyperparameters involved, Deep Learning models are very difficult to interpret and understand the reasoning behind their outcomes, which is the main factor that limits the exploitation of the model in decision and policy making (Lipton, 2018; Y. Wang et al., 2019; Fafoutellis & Vlahogianni, 2023a).

In order to enhance Deep Learning models' explainability, the concept of Explainable AI (xAI) has emerged recently. More specifically, xAI proposes a set of methods that enable the representation of the outcomes of a model and the impact of the input features on the outcome, in a human-understandable way, without requiring a complete understanding of its structure or the algorithm for training and processing the data (Barredo Arrieta et al., 2020). Based on the above concept, the most widespread explainability methods are model-agnostic, i.e. they do not depend on the model's structure, namely permutation feature importance, Shapley values and partial dependence plots (Frye et al., 2020; Molnar, 2019; Parr et al., 2020).

Despite being very useful in providing meaningful explanations of the outcomes of very complex models, xAI methods have been accused of being entirely dependent on the specific dataset that the model was trained with and, thus, the outcomes are not generalizable, in a sense that the explanations provided may not be representative of the phenomenon that is studied but apply only to the specific dataset. Also, they are vulnerable to noisy and distorted datasets and may provide unreliable outcomes. For example, it was observed that even slight permutations of the input data may result in significant differences in the explanations, while the outcome of the model remains unchanged (Schwab & Karlen, 2019).

The most important issue of xAI techniques, however, is that they tend to ignore the causal structure of the input and output data, i.e. causal relations between them (Heskes et al., 2020). Causality is not just a statistical relation that is observed in a specific dataset; it is proof or indication that an "event", e.g. the value of a variable, has a significant impact on the outcome and "causes" it to take a specific value (Miller, 2019). As described earlier, xAI techniques depend on mathematical methods to simulate a model's output and explain its relationship with the input and are dataset-specific.

For example, Shapley values, which attribute the deviation of the model's output from an average (baseline) value to each input variable, assume that all input variables are independent, ignoring all causal interrelations between the input variables and between input and output variables. The above often leads to counterintuitive and misleading explanations of a model's operation. Lately, variations of the original method have been proposed, such as Causal Shapley and Asymmetric Shapley values, which provide explanations that also take account of the causal relations between the data (Frye et al., 2020; Heskes et al., 2020). The drawback of these techniques is that they require a priori knowledge of the causal structure, which should be provided by the user. However, detecting the causal relations in complex datasets, such as traffic-related ones, is not a straightforward task and requires expert knowledge, especially for high-dimensional datasets.

In this section, we present a framework for determining the causal structure of a traffic dataset consisting of measurements at different locations of the road network, which may also be applied in other time series problems, based on Neural Granger proposed by (Tank et al., 2021). Traffic volume data of more than 300 loop detectors installed in the road network of Athens, Greece, are exploited and the proposed framework is applied for both explaining the model's outcomes and for feature selection, which is an equally important task, if not a more important one. Then, a rather simple Long Short Term Memory (LSTM) Neural Network model is developed to forecast the traffic conditions in multiple significant locations and, taking advantage of the outcomes of the framework, achieves a very satisfying performance.

## 7.2 Neural Granger

In order to overcome the shortcomings of the classic approach and mainly the linear relation assumption, we develop a neural network and, specifically, an LSTM to simulate the relation between the input and output variables, based on the Neural Granger concept proposed by (Tank et al., 2021).

The Granger LSTM is trained with the following objective (loss function):

$$\frac{1}{N}\sum^{N}(x_t - h(x_{i,<t}))^2 + \lambda\sum_{i=1}^{n}\|W_i^1\|) \qquad (15)$$

The loss function is very similar to the corresponding of the classic Granger causality test: The first term is the Mean Squared Error of the Granger LSTM's estimation, while the second is the sum of the input weights (weight of the input layer) connected to each time series of the input space. The parameter λ plays the same role as in the classic, linear approach, i.e., controls the significance of the second term. Correspondingly, when a time series does not have a significant contribution to the first term (i.e., reduce the estimation error), its input weight shrinks to zero (second term) and is not passed to the model. At the end of the training process, time series with non-zero input weights are considered to Granger-cause the output.

It should be noted that the above neural network is only deployed to detect the time series that Granger-cause the target, by observing the values of the corresponding input layer weights. Due to its architecture and loss function, it is not recommended to use it for forecasting tasks. It can be therefore thought of as a feature selection mechanism that, additionally, has the advantage that is learnable and does not rely on the calculation of a metric.

## 7.3 Implementation

By utilizing sophisticated Deep Learning models, one can accurately forecast the future traffic conditions using network-wide data. The major drawback of this type of models is that they are not explainable, at least at a satisfactory level, and when too many time series are used as input, there are too many hyperparameters involved, which make the training process of the model difficult and data demanding. Granger causality is a concept that may respond to both challenges; it can be used as a feature selection method, in order to reduce the dimensionality of the input space by keeping input that is causally related to the output, which at the same time enhances the explainability, stability and trustworthiness of the model.

In our proposed approach, we firstly exploit the Neural Granger causality model in order to construct the causal graph of the road network of Athens, Greece, which consists of more than 330 locations/loop detectors that are installed in the entire road network. Neural Granger implemented as an LSTM network, was selected as the most appropriate method to handle a so large network with complex dependencies between the locations. Consequently, a smaller and simpler LSTM network is developed as the model to perform the forecasting task, which is also referred to as “Sparse LSTM”, as its input space includes only the time series that were found to Granger cause the target time series during the first step of the proposed framework. It should be made clear that one Granger LSTM and one Sparse LSTM should be developed for each target location.

For the purposes of this work, a dataset of low temporal resolution (1 hour) was preferred, compared to data of higher resolution, that are more popular in recent literature, e.g. 5 minutes. The reason is that the aim of this paper is to identify significant city-level traffic patterns and spatiotemporal causal relations between the locations of the road network; by using time series with a few minutes resolution the Granger LSTM network would most probably detect relations that are local in space, due to high correlation between consecutive locations and temporally close, due to the high similarity between consecutive timesteps and the inability of most models to handle long time series and their tendency to underestimate the effect of distant time steps.

After an extensive ablation study, the time series length (look back period) that was selected as the most appropriate is 8 time steps (8 hours) which was found to lead to the Granger LSTM with the lower loss value during training, but also is suitable and long enough in order to detect generalized traffic patterns. Moreover, the value of parameter $\lambda$ was set to 100 and the learning rate 0.01 for the Granger LSTM. The other hyperparameters vary for each loop detector (a different model is trained for each detector). The input data were arranged in 8-step time series using a rolling window process and each LSTM takes as input the data from all locations. The output of this model is the set of the other time series that Granger cause each one.

The sparse LSTMs (also one model corresponds to each location) have the same look-back period of 8 hours and their learning rate is 0.01 as well. Their loss function is the Root Mean Squared Error of the predictions. Their architecture also includes a small number of Dense layers (2 or 3), which were necessary in order to achieve more stable performance. The number of neurons of each layer varied between 8 and 32, depending on the detector. The sparse LSTMs take as input the time series that were found to Granger cause the corresponding target time series and are used to forecast their future hourly traffic volume.

## 7.4 Results

### *7.4.1 Detected Causal Relations*

The training of the Neural Granger LSTM results to the causal graph of the road network, i.e. to a set of detected causal relationships. On average, a small part of the road network, approximately 7.3% of the time series (about 25 locations) are found to Granger cause each target time series. The latter value may be a little misleading, as there are only a few time series that are affected by such a high number of others; the corresponding median value is 3.8%, which is more representative of the actual case. The above outcome indicates that in such a large road network, a serious effort should be put in feature selection, as the actual causal relations are very few and the input space's dimensionality can be dramatically decreased, without sacrificing the model's forecasting performance.

The average final value of the loss function of the 334 trained models is 423.8, which is a fairly good value that implies that the models are reliable to extract the causal relations. The loss function, as described in the previous subsection, consists of two terms that account for the accuracy in estimating the target time series' value and the number of time series considered, respectively. Consequently, the values of the loss function are not indicative of the forecasting errors, and, as already discussed, these models should not be used for forecasting, but only for extracting the causal structure; however, their Mean Absolute Percentage Error (MAPE) for forecasting the target values has been estimated, as they are also indicative of the validity of the models. The average MAPE is 21.1%, which indicates a relatively good fit of the causal LSTM networks.

Below, for the sake of brevity, the locations that Granger cause 6 loop detectors that are installed in some of the most significant and crowded locations of the road network are presented. The 6 indicative locations were selected to be around the historical city center of Athens, which is the most critical part of the road network and faces severe congestion problems, especially during the morning and afternoon peak hours. Also, it attracts most of the business and commercial activity, as well as government authorities and organizations buildings. The detected causal parents of the selected loop detectors are illustrated in Figure 36.

Interestingly, we observe a very strong, common pattern for all target locations: loop detectors that are in the city center are Granger caused by detectors that are at significant parts of the road network in the perimeter of the city of Athens, acting as entrances to the city center, and, mostly, not very close to the city center. Furthermore, another important finding is that there are a few (3 or 4) specific loop detectors that Granger cause almost all loop detectors of the city center; it turns out that the latter detectors can provide vital information about the future traffic conditions at the city center, so traffic management authorities and other practitioners should always consider their measurements for forecasting and decision-making purposes. Finally, one can notice that the detected causal relations are between the target loop detectors and other detectors that are not very close to them. This was one of the aims of this work and the reason why low-resolution data and a large look-back period were used. So now, we may assume that the detected relations are more generalizable (not data- or time of the day- specific) and are ought to Origin-Destination relations and significant traffic patterns.

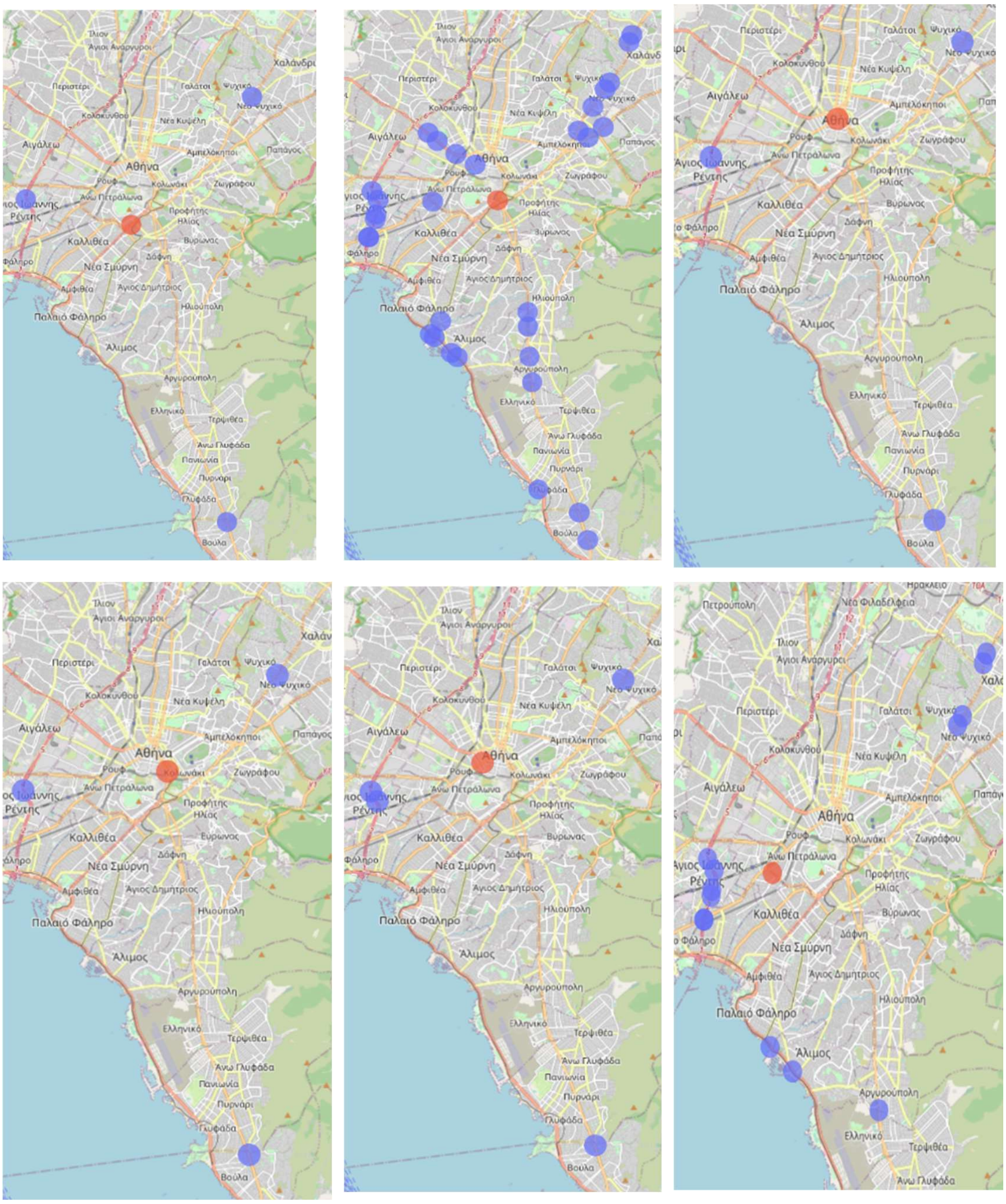


**Figure 36. Locations (blue) that Granger-cause target location (red).**

### *7.4.2 Forecasting Results*

To evaluate the accuracy of the proposed Sparse LSTM forecasting model for each location, comparisons are established based on an LSTM network that takes as input the time series from the target location only (Single-point LSTM) and an LSTM network that takes input from all locations (Inclusive LSTM), in terms of forecasting accuracy and computational efficiency.

More specifically, for each of the 334 models, the MAPE values and the time to train were measured.

As can be clearly seen in Table 15, the proposed Sparse models, taking advantage of the causal structure detection feature selection strategy implemented, achieve a better forecasting performance than both baseline models, while being significantly more efficient than the inclusive model, as their architecture is more compact, due to the reduced dimensionality of the input space. The Single-points LSTM's error is slightly higher than Inclusive LSTM's, while it is the most efficient in terms of training time.

**Table 15. Forecasting models evaluation (average from of all detectors)**

| Model | Average MAPE (and deviation) | Efficiency (time to train per model) |
|---|---|---|
| Inclusive LSTM | 12.6% ± 3.1% | 104s |
| Single-point LSTM | 13.3% ± 3.7% | 27s |
| **Sparse LSTM** | **9.1% ± 1.8%** | **46s** |

## 7.5 Summary of Findings

In this section, the innovative Neural Granger approach for detecting causal relations between time series was applied to the large road network of Athens. Findings show that using the concept of Neural Granger for detecting causal features, the dimensionality of the input space can be significantly decreased, while the emerging relations are more meaningful and interpretable, compared to other feature selection strategies, such as proximity- and statistical correlation-based, increasing the model's transparency as well. Simultaneously, it is indicated that the model's performance is also enhanced.

Using the above methodology, except for the significant relations between its locations, significant traffic patterns were identified. The most important of them was that the traffic conditions at locations of the perimeter Granger cause the traffic conditions of locations at the city center, while also specific locations at the perimeter emerged as the most significant, as they were found to Granger cause most of the loop detectors located at the city center. Moreover, by using the Neural Granger for feature selection, it was possible to achieve an improved forecasting performance, while increasing the efficiency of the model as well.

# 8 A THEORY-INFORMED MULTIVARIATE CAUSAL FRAMEWORK FOR SHORT-TERM URBAN TRAFFIC FORECASTING

In the previous section, the importance of detecting causal relations between the locations of a road network has been highlighted both for increasing the forecasting accuracy and trustworthiness, but also for identifying significant traffic patterns. In this section, we incorporate additional traffic flow theory-related aspects and propose a holistic approach (from data engineering to model training and evaluation) for a traffic forecasting framework that is theory-driven, causal, and multivariate, aiming to be actionable as well.

At the core of our methodology, there is a novel traffic flow theory-informed multitask neural network, which is used for the joint short-term forecasting of two traffic variables (traffic volume and speed), that constitute a traffic state, as mentioned earlier.

For training the model, we propose a custom-made loss function, which incorporates the distance of the emerging multivariate forecast (pairs of traffic variables) from the actual fundamental diagram of the corresponding location. To enhance the model's performance and interpretability, network-level traffic information is selected from the most relevant locations, specific to each target location, using the Neural Granger adaptation of classic Granger causality.

For the experiments presented in this work, we deploy a Long Short-Term Memory (LSTM) Network; nevertheless, the entire methodology (including the loss function) is compatible with any Deep Learning structure. Furthermore, we present a reliable and exclusively data-driven unsupervised learning method for fitting the fundamental diagram in loop detector measurements, which is a prerequisite for using the loss function. The entire methodology is trained and tested on data coming from the road network of Athens (Greece).

An overview of the methodological approach is given in Figure 37.

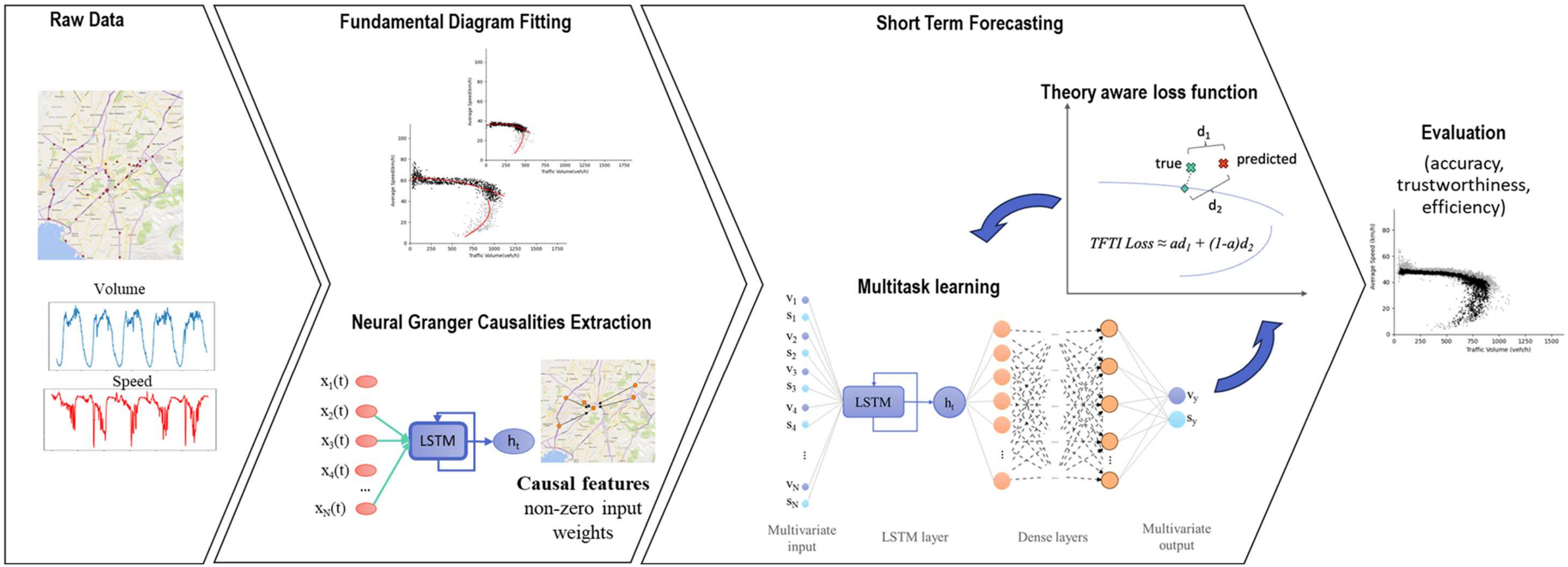



**Figure 37. Methodological approach overview**

## 8.1 Traffic Flow Theory-Informed Loss Function

Training a deep learning model is an optimization problem aiming to determine the optimal values of its weights to minimize a loss function, such as the Mean Squared Error (MSE) or the Root Mean Squared Error (RMSE). Considering the case of the joint forecasting of volume and speed data, in a multitask modeling setup, the most popular approach would be to estimate a loss value (error metric) for each of volume and speed predictions at every training epoch. The lower the loss the better the individual predictions. However, except for a low error value, we shall expect that all predicted pairs of values should lie close to the corresponding fundamental speed-volume curve emerging by fitting a functional form to the real traffic data and constitute valid pairs of traffic states. The latter is not guaranteed by a low error value; actually, a decent individual mean error value for each variable may cover issues such as the aforementioned.

To account for the above challenges, we propose the Traffic Flow Theory-Informed loss function (TFTI loss), which combines the MSE of both individual variables with the distance of the prediction from the closest point on the estimated fundamental diagram. The latter is defined for each location/section of the road network that is monitored and may have a specific functional form. Let $\hat{y}_i = (\hat{v}_i, \hat{s}_i)$ the prediction for a real pair $y_i = (v_i, s_i)$ and $g_j = (v_j^e, s_j^e)$ the closest point on the estimated fundamental graph which characterizes the location of interest, TFTI loss is defined as follows:

$$TFTI\ loss = a * \frac{1}{2}\sqrt{[MSE_v + MSE_s]} + (1 - a) * d(\hat{y}_i, g_j) \qquad (16)$$

where $MSE_v$, $MSE_s$ are the mean square error of the predictions of volume and speed respectively, $d(\hat{y}_i, g_j)$ is the Euclidean distance of the predicted pairs and the closest point of the fundamental diagram to the corresponding actual value. The factor $a$ controls the significance of the second factor over the first, $a \in [0,1]$.

In addition, by observing an individual error metric for each predicted variable, three major concerns may arise that require a special treatment. First, in multitask forecasting, one or more of the output variables may be predicted with a significantly higher error than the other and, thus, a loss function taking both into account in a fairer way is required. Second, when evaluating two predictions for the same point that have the same distance from the actual point, the distances from an estimated speed-volume graph, which reflects the validity of the one against the other should also be examined and the two predictions should be treated differently (Figure 38 (a)). Third, depending on the shape of the fundamental diagram, it is possible that two points may be close enough (low anticipated error) but belong in different branches of the diagram (congested-uncongested), i.e., although the forecasting error may be acceptable, the prediction may suggest, for example, congested conditions, while this is not the case (Figure 38 (b)). In both cases, a mechanism that penalizes and steers the prediction closer to the fundamental diagram, such as the proposed TFTI loss, has the potential to eliminate the above phenomena of infeasible predictions and wrongly interpreted traffic states.

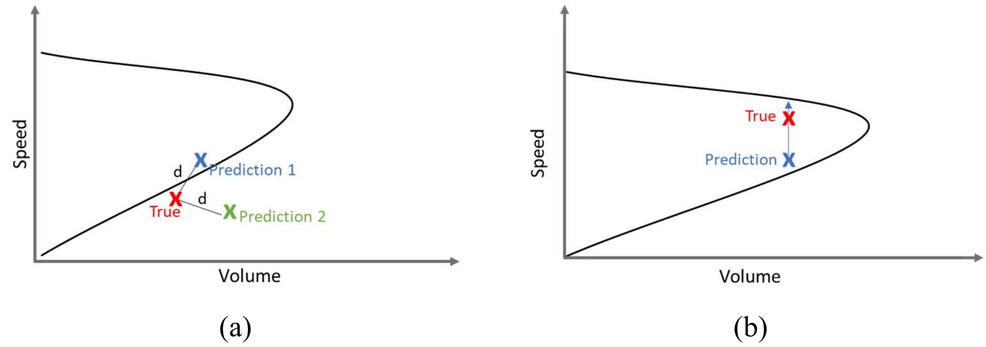



**Figure 38. Examples of challenging forecasting tasks requiring special treatment: (a) two predictions with same error but only one suggests a valid volume-speed, (b) one prediction close to the wrong branch of the speed-volume diagram that should be steered to the correct branch.**

In simple words, the rationale of the proposed loss function is that predictions with high error and/or predictions that constitute a non-feasible speed-volume pair should be "steered" closer to the fundamental diagram to improve the forecasting accuracy. Moreover, in relation to the two examples in, the loss value would be significantly higher for predictions that lie further from the fundamental diagram and propose a traffic state that is not valid for the specific location compared to those that are closer. At the same time, by penalizing and steering the prediction closer to the point of the fundamental diagram that is closer to the actual point, it is expected that the phenomenon of incorrectly anticipated traffic states will also be eliminated.

## 8.2 Implementation

### *8.2.1 Data*

The data used in this section were derived from the database that is maintained by the Traffic Management Center of the Region of Attica. For the purposes of this work, measurements from 420 detectors were made available to us, consisting of the traffic volume (veh/hour), mean speed (km/h) and occupancy (%). The time resolution of the data is 6 minutes, i.e., the sensed data are aggregated in 6-minute intervals. The locations of the loop detectors that are used in this section are presented in Figure 39. As one may see, the detectors cover a large area around the city center and their density is relatively high, especially at some very significant road sections. The data refer to 40 days of measurements, between March 20$^{th}$ and April 30$^{th}$ of 2023.

For the sake of brevity, 12 locations that lie close to the city center and other significant economic/business areas were selected as the target locations, i.e., models were developed to forecast the future traffic conditions at the specific locations. It was attempted to include all types of locations, namely loop detectors at highways and secondary roads (urban and interurban), signalized and non-signalized, as well as uniformly distributed from a geographical perspective. The target locations are presented in Figure 40 below, along with their unique ID number.

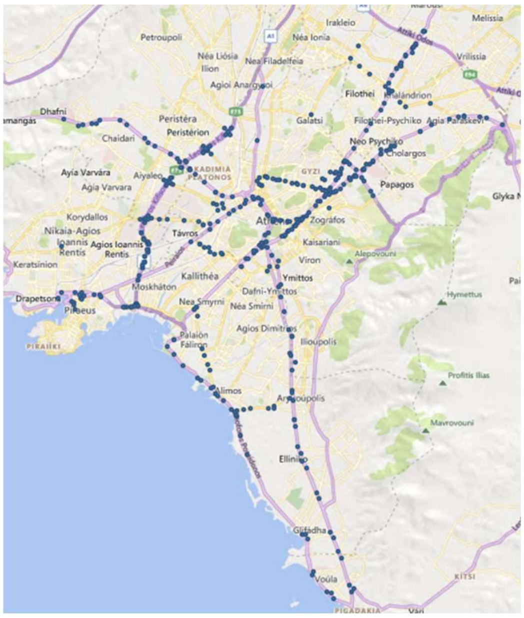


**Figure 39. Locations of loop detectors in the Athens Metropolitan Area.**

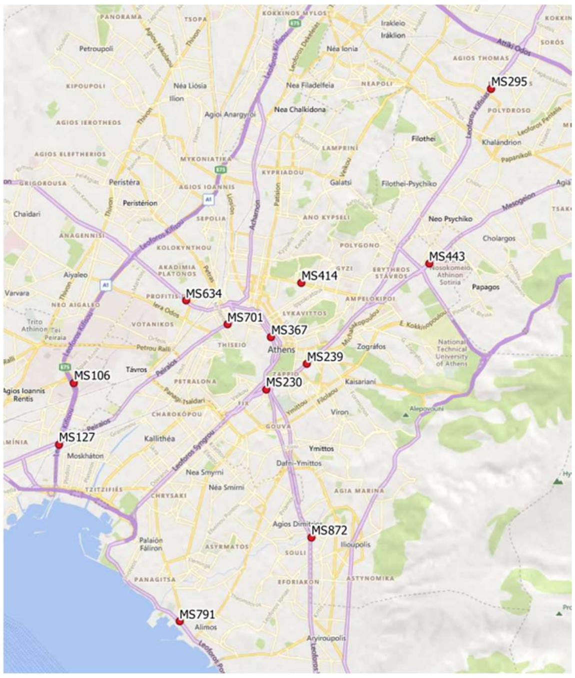



**Figure 40. Target loop detectors' locations and unique ID number**

### *8.2.2 Fundamental Diagrams Fitting with Loop Detector Data*

Loop detectors are the most popular data source in traffic forecasting literature, due to their effectiveness and efficiency in collecting large amounts of network-wide data. This type of sensors provides temporal measurements, i.e., traffic conditions at fixed locations over a specific time interval. Thus, they cannot provide traffic density measurements, which are spatial; instead, they can measure the occupancy, which can be considered approximately proportional. However, occupancy measurements are highly dependent on the actual location of the sensor. For example, a sensor at the vicinity of an intersection, where vehicles move slower, would overestimate the actual occupancy/density of the specific road section. The effect of the sensor's location is not so important on the two other variables (volume and speed) (Bramich et al., 2022). For this reason, in this work, traffic volume and speed data were preferred both for the forecasting task, but also for the fundamental diagrams.

Moreover, as far as it concerns the estimation of the fundamental diagram, loop detector data are not considered the most suitable, as neither speed nor occupancy measurements are spatial. Indeed, all traffic flow analytical models, including the fundamental relation, require section- or lane-level data (Chiabaut et al., 2009). For example, mean speed should be the average value of measurements over the section's length (space mean speed), instead of the average over a time period at a specific point (time mean speed) (Knoop & Daamen, 2017). Taking the above into account, a data-driven method for fitting the fundamental diagram is, thus, more suitable than analytical models.

To use the TFTI loss, one should first produce a functional form of the fundamental diagram by fitting a curve to the target location's measurements and develop a multitask prediction model, as both traffic volume and speed variables are considered in the loss function. To account for the discontinuities observed in the fundamental relationships of traffic variables, i.e., two-variate model (two curves) we use an unsupervised learning approach and curve fitting (Knoop & Daamen, 2017). First, k-means clustering is undertaken to identify groups in data that should be considered separately in the fitting process (congested and uncongested branches), and then polynomial curve fitting is conducted in each group.

More specifically, for the fundamental diagram of all locations (not only target locations), first, the (traffic volume, speed) pair measurements of each location are separated into congested and uncongested using the k-Means clustering algorithm. The emerging mean silhouette index value is 0.57, indicating that the traffic measurements' patterns are very strong and the points are relatively easily distinguished by the unsupervised learning algorithm. Then, a second-order polynomial curve was fitted at each cluster of each location, so, at each location, we have a fundamental diagram that consists of two curves (two-variate, discontinuous diagram). The fit of the curves was assessed by estimating the $R^2$ metric for each one. Its average value is 0.64, which is considered a very satisfying value.

It should be noted that the above results in fitting the diagrams are partially ought to the validity and good quality of the data that were used, as well as their adequate amount: the two branches are clearly visible (and distinguishable) in the vast majority of the cases and there do not seem to be erroneous and noisy data, at least to an extend that would prevent the good fitting of the curves. Otherwise, it would have been necessary to put a significant amount of effort into data cleaning and imputation.

In Figure 41, Figure 42 and Figure 43 twelve indicative fundamental diagrams from target locations are presented. Points that were classified as uncongested are black in color, while the corresponding congested points are grey. Except for the good fit that the reader can observe ($R^2$ metric is also provided next to the location's ID), the capacity drop, which was also detected in a purely data-driven way, is also clearly visible and can also be attributed to the good quality of the data.

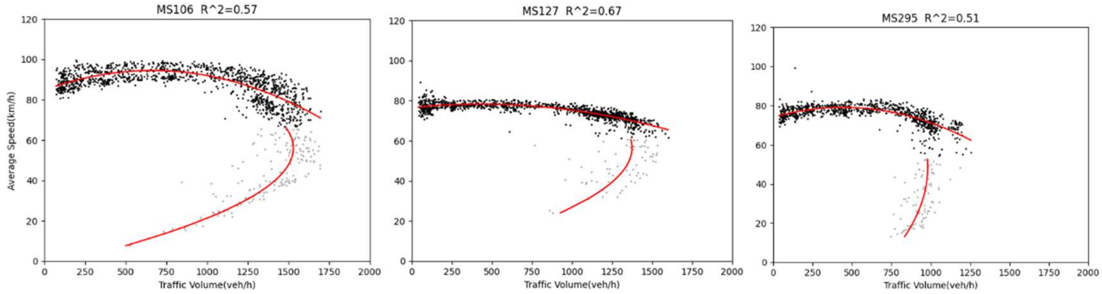



**Figure 41. Indicative fitted fundamental diagrams for high free flow speed locations (above 60 km/h)**

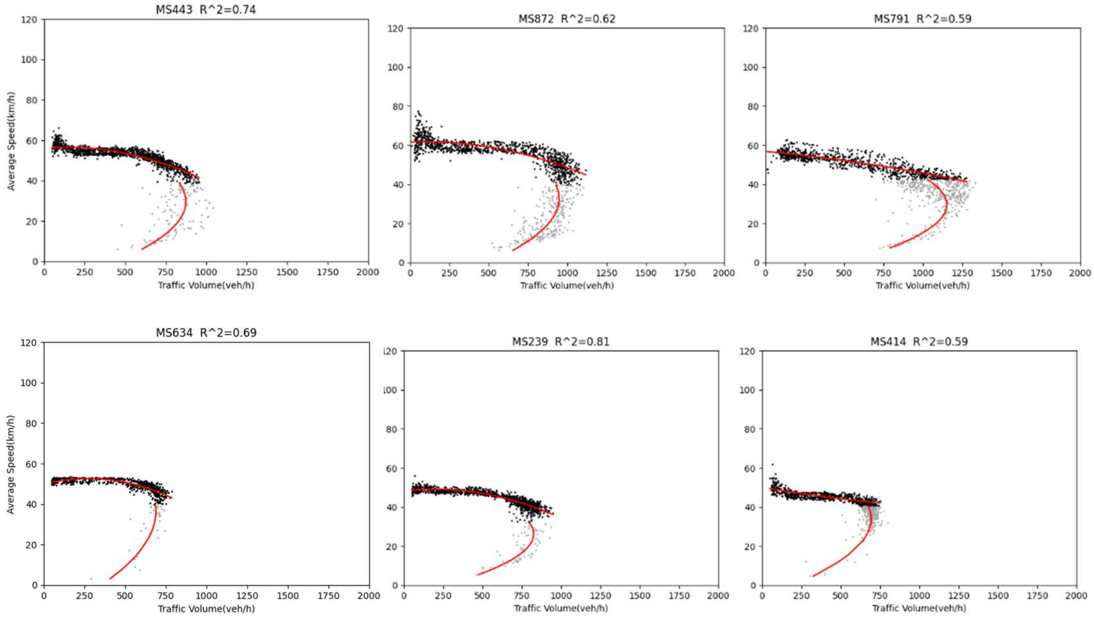



**Figure 42. Indicative fitted fundamental diagrams for medium free flow speed locations (between 40 and 60 km/h)**

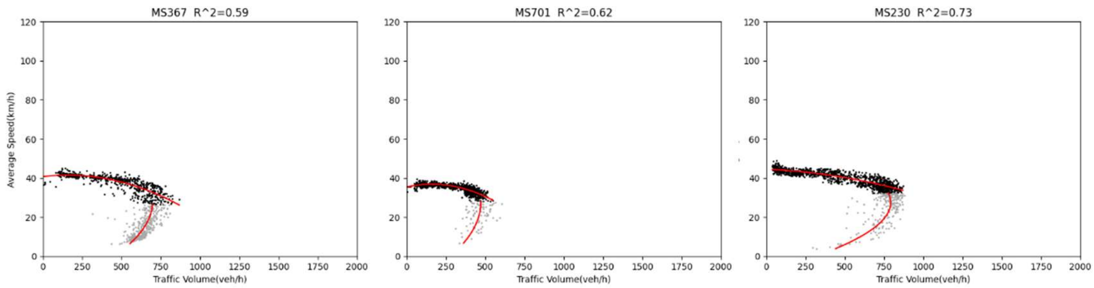



**Figure 43. Indicative fitted fundamental diagrams for low free flow speed locations (about 40 km/h)**

### *8.2.3 Modeling Setup*

It is important to highlight that the traffic flow theory-informed loss function, as well as the entire framework presented so far, is compatible with any Deep Learning model. In this work, it was preferred to use a rather simple architecture, i.e., an LSTM network followed by a small number of fully connected (dense) layers and focus on investigating the effect of using the proposed theory-informed framework. Evidently, by using a more complex structure (e.g., Graph Neural Networks), or other techniques, such as an attention mechanism, we could potentially achieve a better performance. But this falls beyond the scope of this section. Moreover, we aim to show that the proposed framework can achieve state-of-the-art performance even when using a simpler model structure. The forecasting horizon is 6 minutes, which corresponds to one timestep. The experiment aims to evaluate the effectiveness of the TFTI loss function in reducing the overall forecasting error (for both variables) but also increasing the trustworthiness of the model.

The two approaches compared are training the model with (i)simple MSE error and (ii)proposed TFTI loss function, with parameter $a = 0.7$, which corresponds to 70% weight for the first term (MSE) and 30% for the second (distance from the fundamental diagram). The specific weights emerged as those providing the most accurate forecasting results after an extensive ablation study. Moreover, to have a valid and fair comparison, the two models that are developed for each target location (MSE and TFTI loss) are identical, except for the loss function: they have exactly the same architecture, i.e., same number and type of layers and neurons, same values of all hyperparameters, as well as the same input space (previous measurements of locations causally related to the target location), were trained using the same training set and evaluated on the same test set. Moreover, the optimal architecture and hyperparameters' values for each target location were defined after an extensive grid search with a 5-fold cross-validation scheme using the network with the MSE loss, which may favor it over the TFTI loss.

The main specifications of the models for the 12 target locations, as emerged from the above-described process are the following: An LSTM layer is used with 128 or 256 units (neurons), depending on the location, and a look-back period of 10 timesteps, which equals to 1 hour. The models also have up to 5 additional hidden dense layers, each consisting of 8 to 64 neurons. The Adam optimizer was used for fitting the models and the Rectified linear unit (Relu) as the activation function of all layers. The learning rate was varying from 0.0001 to 0.0010. The maximum number of training epochs is 200, but an early-stopping strategy was also adopted, so most networks were trained for a lower number of epochs.

The input of each model is the time series of traffic volume and speed of the loop detectors that were found to be causally related with the corresponding target location, while the output is the values of both variables after 6 minutes (one-step forecasting). The data are separated randomly into train and test sets; 67% for training and 33% for testing. Also, a 5-fold cross-validation scheme is deployed. Finally, due to the two variables (traffic volume and speed) having very different value ranges, they were scaled between 0 and 1.

## 8.3 Results

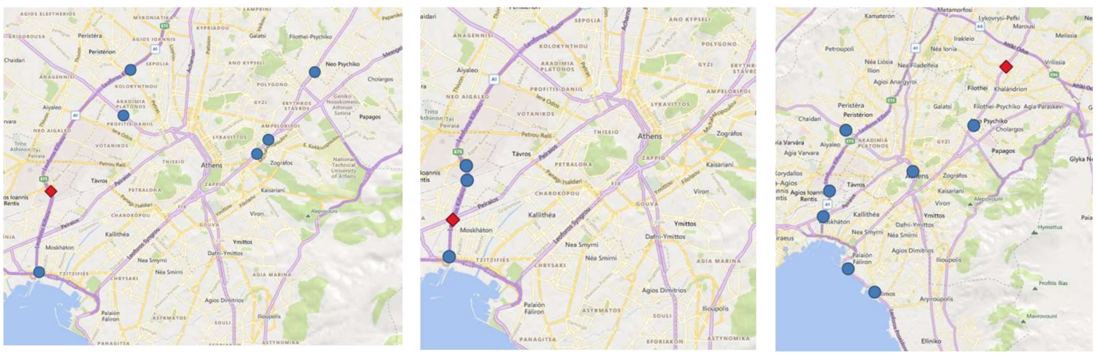
### *8.3.1 Causal Relations*

As described earlier, both in classic and neural Granger causality, the parameter λ plays an important role in defining the number of time series that will be considered to Granger-cause the target time series. In our work, we use a relatively high value of λ, λ=500, which results in only a few time series being causally related to the respective target time series. Therefore, we achieve to significantly reduce the dimensionality of the input space, without sacrificing the forecasting accuracy, as the results indicate.

After applying the Neural Granger method on the available dataset, the locations that Granger-cause the target locations emerged. The average number of detected locations was 6.6 out of the 420 locations, which indicates that the dimensionality of the input space for each forecasting model can be vastly decreased, without, however, losing significant information. The locations that were found to Granger-cause the target locations are presented in Figure 44, Figure 45 and Figure 46.

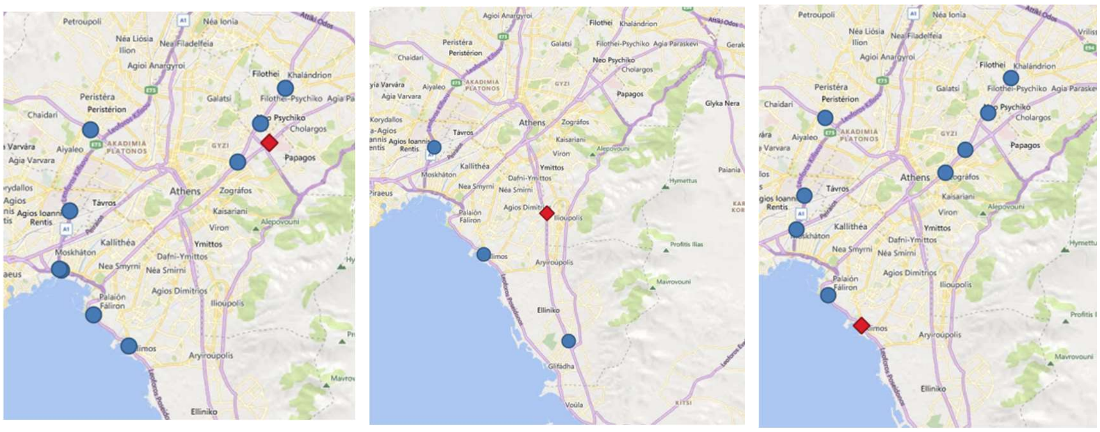
It seems that more causal relations can be found between distant locations than between neighboring locations. Considering that the look-back window is one hour, this is reasonable and indicates the validity of the findings. Moreover, it also suggests that feature selection methods based on the proximity of the locations may not be the most suitable for network-level forecasting. Moreover, one may notice that the locations close to the city center are Granger-caused mostly by locations at the perimeter, which serve as entrances to the city center and significantly less by other points that are also close to the city center. On the other hand, perimeter locations are Granger-caused by locations that are also at the perimeter and, secondarily, by locations close to the city center, which is ought to strong patterns of cross-city trips (e.g., commuting) that may as well last for up to an hour.

Finally, there are some locations that were found to Granger-cause most of the rest, which indicates that significant flows of vehicles pass from them and traffic patterns, e.g., congestion, may initiate there and propagate to the rest of the road network. These locations are very important for traffic management purposes and should always be monitored by relevant authorities. The specific locations in Athens lay in the western and southwestern regions (Kifissos Avenue).

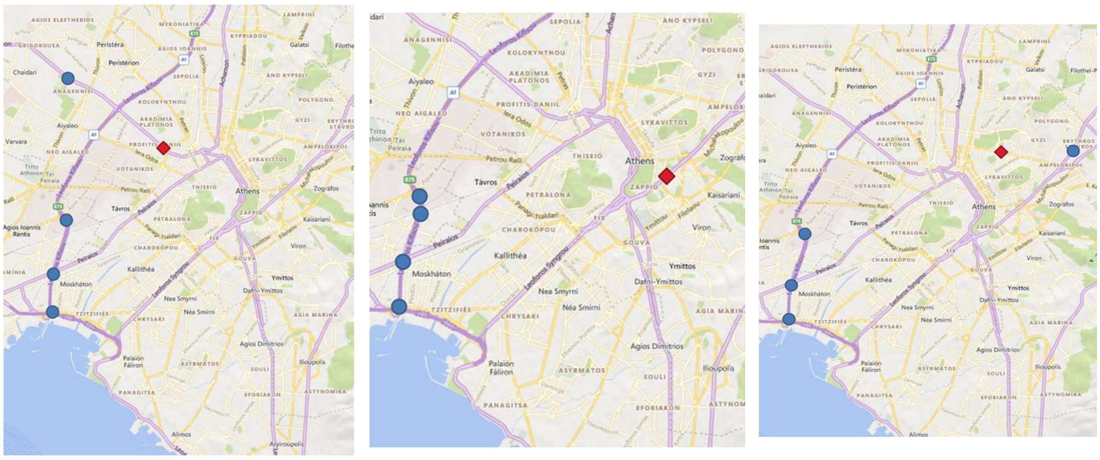


MS106 MS127 MS295

**Figure 44. Emerging locations (blue circle) that Granger-cause the target locations (red diamond) for high free flow speed locations**

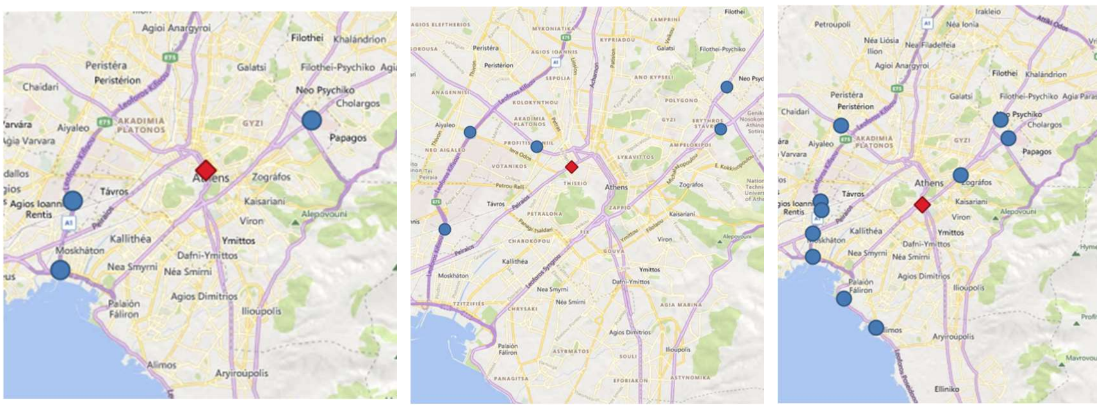


MS443 MS872 MS791



MS634 MS239 MS414

**Figure 45. Emerging locations (blue circle) that Granger-cause the target locations (red diamond) for medium free flow speed locations**



MS367 MS701 MS230

**Figure 46. Emerging locations (blue circle) that Granger-cause the target locations (red diamond) for low free flow speed locations**

### *8.3.2 Forecasting Results*

In Table 16, Table 17 and Table 18 the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE) of the predictions are presented for the two loss functions overall and for each variable separately for the target locations.

As one may observe, the TFTI loss aids the model in achieving a lower overall error for all locations examined. The mean MAPE for the TFTI loss is 10.9%, versus 13.0% for the MSE loss for the 12 target locations. The latter is mainly ought to the fact that the model with the MSE loss fails in almost all cases to accurately forecast the speed values, although it performs satisfyingly in predicting the traffic volume. A possible explanation for this is that the dataset is highly imbalanced, namely most of the points belong to the uncongested branch, which adds a bias to the model towards predicting higher speed values. Moreover, the value range of traffic volume is the same for both branches and, thus, traffic volume predictions are not affected in the same way as speed. On the other hand, the TFTI loss incorporates the information about the branch the point should belong to (i.e., distance from the fundamental diagram); consequently, although the models with the TFTI loss also predict more accurately points of the uncongested branch, they retain a decent performance for the congested branch, too, as it is discussed in more detail in the following subsection.

**Table 16. Forecasting results comparison (high free flow speed locations)**

| | *MSE loss* | ***TFTI loss*** | *MSE loss* | ***TFTI loss*** | *MSE loss* | ***TFTI loss*** |
|---|---|---|---|---|---|---|
| *Location* | **MS106** | | **MS127** | | **MS295** | |
| *Volume MAE* | 40.6 | 62.3 | 43.7 | 44.8 | 38.2 | 45.8 |
| *Volume MAPE* | 6.8% | 9.1% | 7.7% | 8.1% | 9.1% | 13.2% |
| *Speed MAE* | 8.0 | 3.3 | 4.3 | 2.1 | 5.3 | 3.3 |
| *Speed MAPE* | 17.3% | 6.4% | 7.7% | 3.5% | 13.5% | 7.7% |
| ***Overall MAPE*** | 12.0% | **7.8%** | 7.7% | **5.8%** | 11.3% | **10.4%** |

**Table 17. Forecasting results comparison (medium free flow speed locations)**

| | *MSE loss* | ***TFTI loss*** | *MSE loss* | ***TFTI loss*** | *MSE loss* | ***TFTI loss*** |
|---|---|---|---|---|---|---|
| *Location* | **MS443** | | **MS872** | | **MS791** | |
| *Volume MAE* | 40.4 | 49.2 | 42.6 | 58.9 | 41.9 | 55.6 |
| *Volume MAPE* | 9.6% | 13.0% | 11.1% | 18.6% | 8.3 | 14.7% |
| *Speed MAE* | 4.3 | 2.8 | 7.1 | 3.9 | 4.0 | 2.7 |
| *Speed MAPE* | 17.0% | 9.3% | 27.5% | 14.0% | 15.1 | 9.7% |
| ***Overall MAPE*** | 13.3% | **11.1%** | 19.3% | **16.3%** | **11.7%** | 12.2% |
| *Location* | **MS634** | | **MS239** | | **MS414** | |
| *Volume MAE* | 29.5 | 31.0 | 39.9 | 44.0 | 31.1 | 31.5 |
| *Volume MAPE* | 8.6% | 9.7% | 9.6 | 10.4 | 8.8% | 9.2% |
| *Speed MAE* | 3.2 | 1.8 | 3.6 | 2.1 | 2.7 | 1.9 |
| *Speed MAPE* | 11.1% | 6.7% | 14.0 | 7.3 | 11.5% | 6.7% |
| ***Overall MAPE*** | 9.8% | **8.2%** | 11.8% | **8.9%** | 10.1% | **7.9%** |

**Table 18. Forecasting results comparison (low free flow speed locations)**

| | *MSE loss* | ***TFTI loss*** | *MSE loss* | ***TFTI loss*** | *MSE loss* | ***TFTI loss*** |
|---|---|---|---|---|---|---|
| *Location* | **MS367** | | **MS701** | | **MS230** | |
| *Volume MAE* | 39.9 | 45.7 | 34.7 | 35.4 | 35.4 | 39.5 |
| *Volume MAPE* | 13.6% | 19.2% | 18.8% | 21.9% | 9.2% | 10.5% |
| *Speed MAE* | 4.1 | 2.6 | 2.8 | 2.0 | 3.9 | 2.1 |
| *Speed MAPE* | 21.8% | 13.9% | 13.4% | 9.5% | 23.0% | 9.4% |
| ***Overall MAPE*** | 17.7% | **16.6%** | 16.1% | **15.7%** | 15.6% | **9.9%** |

Furthermore, as mentioned earlier, the models of each location are identical, except for the loss function, of course, but may be trained for different numbers of epochs, as we used an early stopping scheme. It was observed that the models with the TFTI loss were trained faster, requiring about 92 epochs on average to converge, versus 129 for the MSE loss, which indicates that the proposed loss function can increase the model's training efficiency as well.

### *8.3.3 Trustworthiness Assessment*

Trustworthiness is a property of a model (or a system) that goes beyond its correctness, which can be assessed using popular error metrics; it expresses whether the model's performance remains reliable and stable under any circumstances (Ghobrial et al., 2023). A simple, yet accurate, definition of a trustworthy model is that it is a model that retains a correct behavior that is aligned with the training data and, in general, with domain experts' understanding of the factors that affect the outcome of the predictions (de Bie et al., 2021). Therefore, it becomes clear that an accurate model is not always trustworthy. In recent literature, however, the respective models are almost always assessed based exclusively on their accuracy, which leads to misleading conclusions. Besides, most models are not multivariate.

Especially in the case of imbalanced datasets, the classic error metrics can be very misleading in assessing both the accuracy and the trustworthiness of a model. In terms of congested and uncongested conditions, most traffic datasets are imbalanced, i.e., highly congested conditions may occur at certain periods during the day and, usually, last for a relatively short time. Specifically, in the dataset used in this work, on average about 11% of the measurements correspond to congested conditions, which is about 2.5 hours per day. As mentioned previously, indeed, the difference between the values of the error metrics of the models trained with different loss functions is probably ought to how they handle imbalanced datasets. To examine the above hypothesis, we provide indicative fundamental diagrams that emerge from the predictions compared to the actual of the test set for both losses in Figure 47, Figure 48 and Figure 49.

**MSE loss** **TFTI loss**

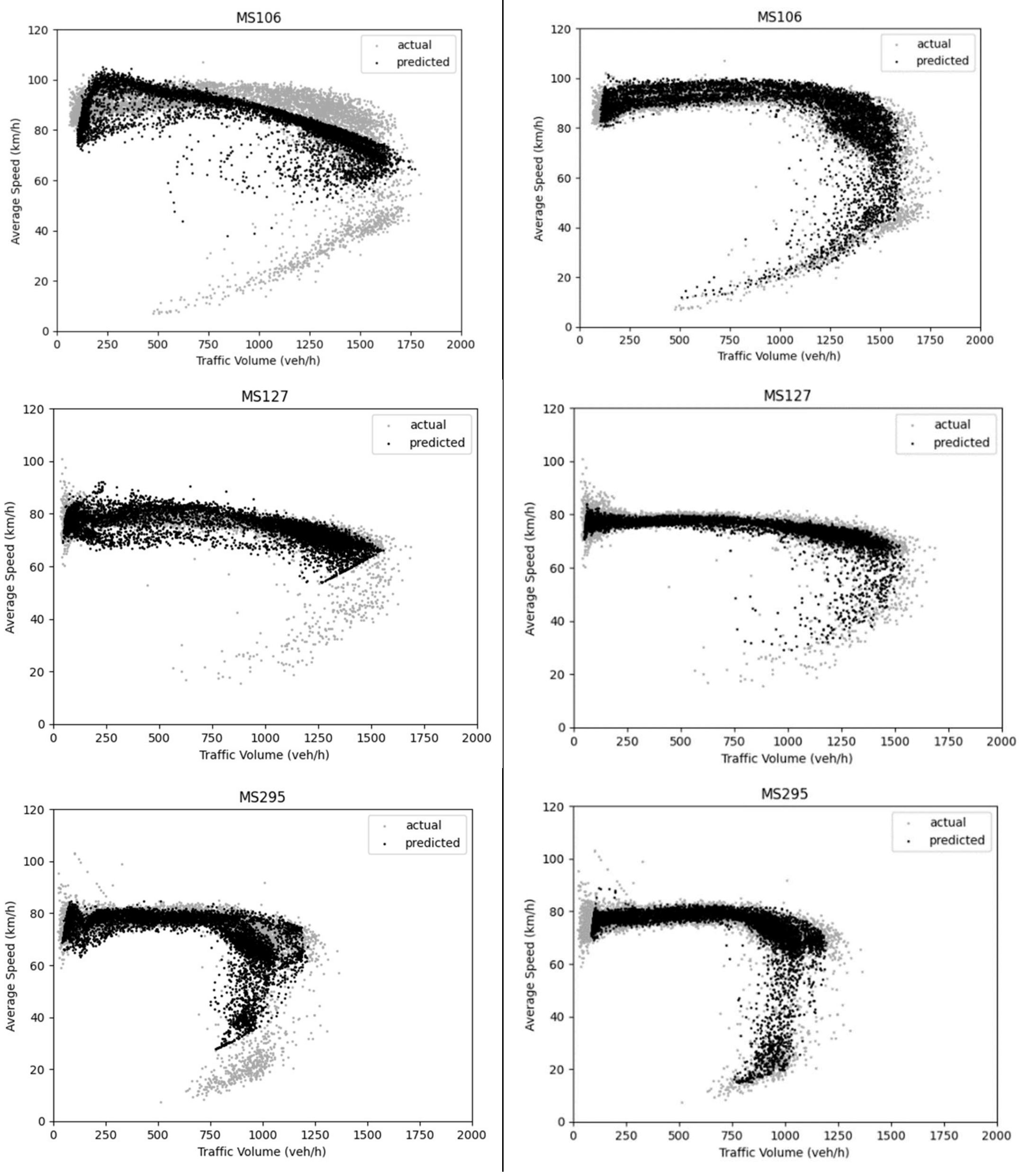



**Figure 47. Emerging predicted fundamental diagrams versus actual for MSE loss (left) and TFTI loss (right) (high free flow speed locations)**

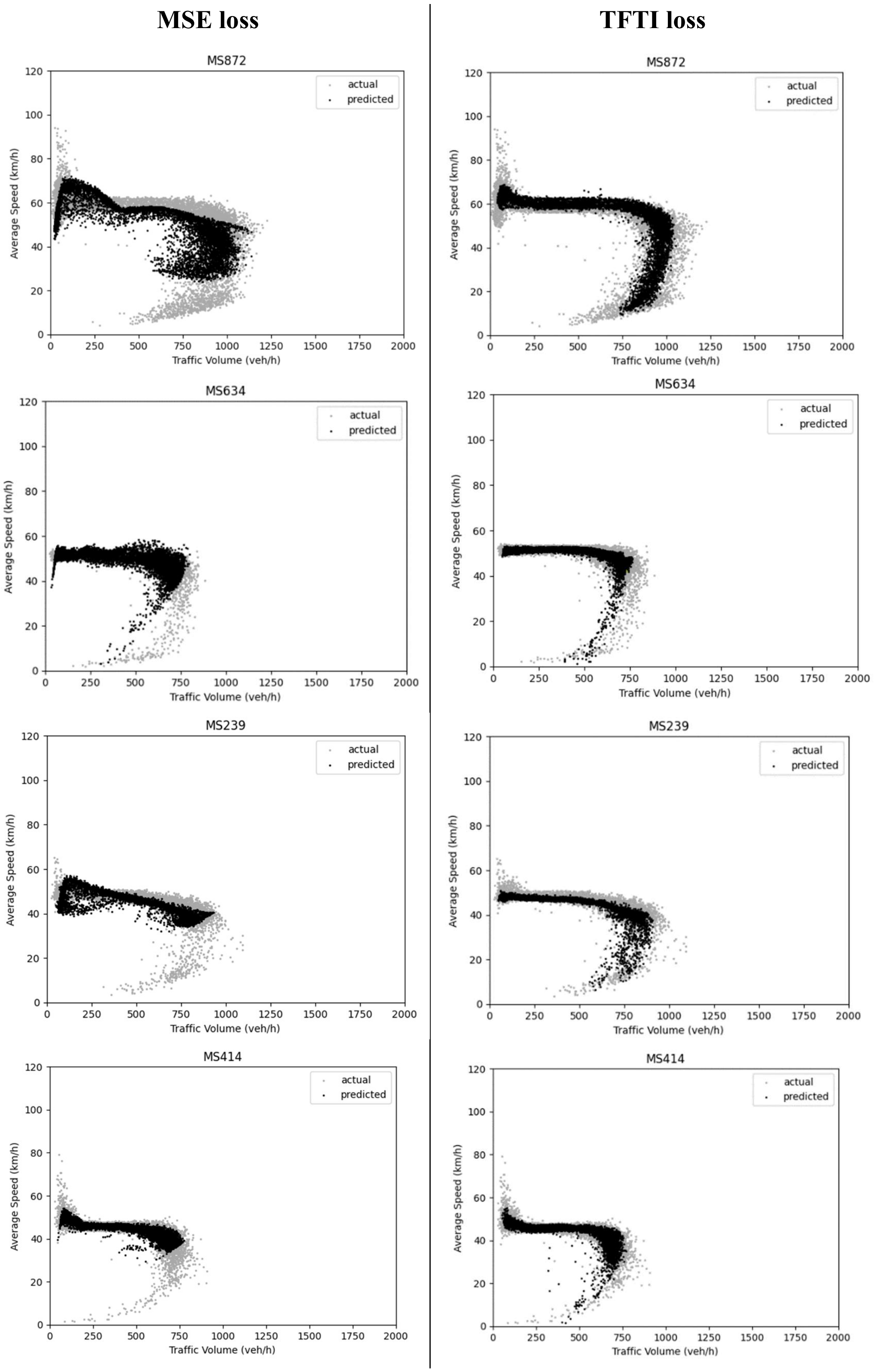



**Figure 48. Emerging predicted fundamental diagrams versus actual for MSE loss (left) and TFTI loss (right) (medium free flow speed locations)**

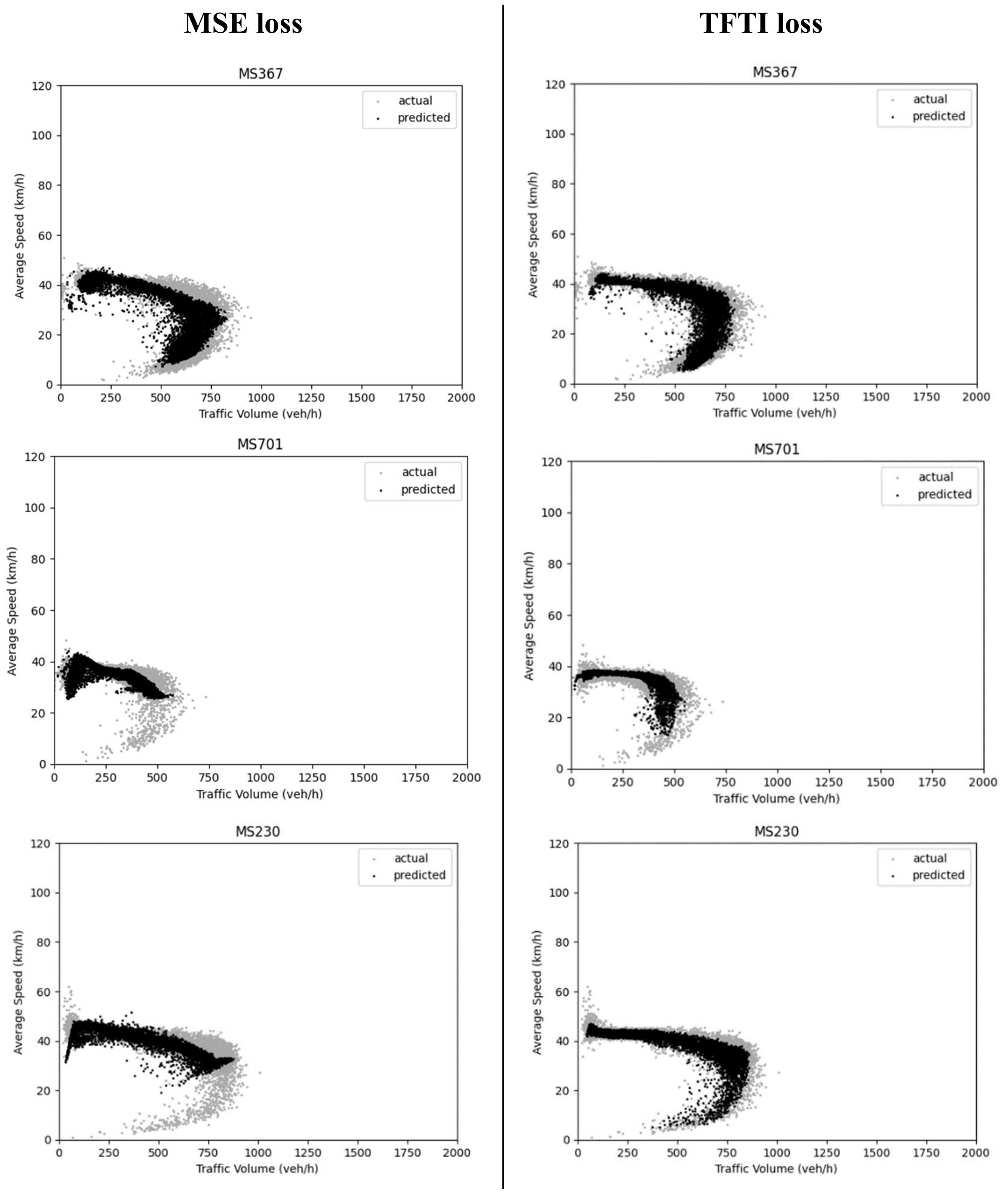



**Figure 49. Emerging predicted fundamental diagrams versus actual for MSE loss (left) and TFTI loss (right) (low free flow speed locations)**

The graphs indicate that the predictions made by the models trained with the TFTI loss are more trustworthy, as the distribution of the prediction points is closer to the actual and the corresponding fundamental diagram, while the MSE loss models fail systematically to predict the congested branch. Moreover, the TFTI loss is less vulnerable to noisy data and outliers, e.g., points that are between the two branches of the anticipated fundamental diagram. We consider assessing these plots as a very important aspect of the present work, as for most MSE loss-trained models the error metrics are quite decent, which can be very misleading regarding the trustworthiness of the models.

Although being the minority class, accurately forecasting the occurrence of congested conditions is vital for most traffic management purposes. Therefore, high accuracy in forecasting observations of the congested branch is a prerequisite for the model's actionability.

Consequently, to further investigate the issue indicated by the above graphs, we visually compare the boxplots and the density plots of the distribution of the actual values (test set) and the predicted ones for both output variables. An indicative example is presented in Figure 50. As one may observe, the three volume distributions are very similar, while, for the speed distributions, the MSE-trained models fail to capture the lower values. However, as these values are seen as “outliers”, Figure 50 does not reveal the full extent of the specific issue.

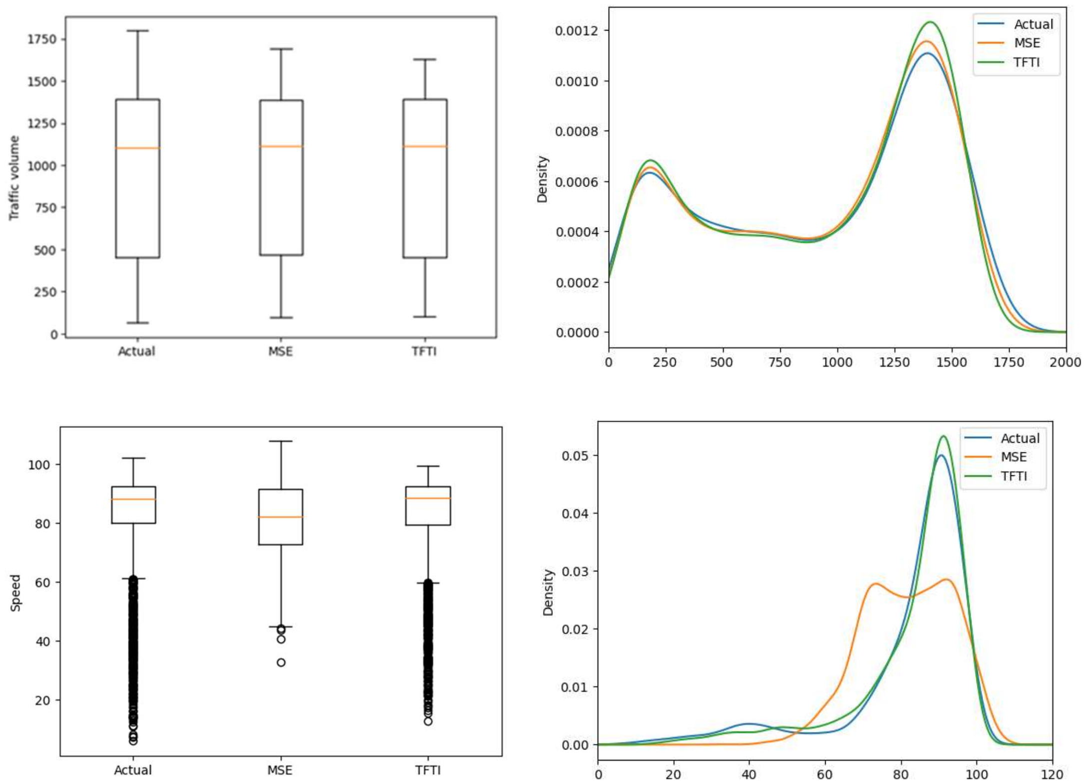



**Figure 50. Comparison of predictions (MSE and TFTI loss models) versus actual distributions for traffic volume (up) and speed (down) of indicative location (MS106)**

To quantify the above-observed divergence between the predictions made by the models trained with the two loss functions, we proceed to estimate the accuracy of the models in predicting the correct class of conditions, regardless of the exact values of the indicator variables. It is reminded that the measurements of each location were classified as congested or uncongested based on the K-Means clustering approach followed before fitting the curves for the fundamental diagrams.

Since congested conditions stand for about 11% of the dataset, the impact of misclassified congested conditions is not so significant on the overall accuracy. So, we additionally estimate the overall F1-score and the F1-score of the congested class. The results for four indicative locations are presented in Table 19. The corresponding mean values for the 12 target locations are for the MSE loss: Accuracy = 0.92, F1-score (congested class) = 0.43, Overall F1-score = 0.69, while for the TFTI loss: Accuracy = 0.96, F1-score (congested class) = 0.74, Overall F1-score = 0.86.

**Table 19. Classification metrics comparison for different indicative locations**

| | MSE loss | **TFTI loss** |
|---|---|---|
| *Location* | **MS106** | |
| *Accuracy* | 0.93 | 0.95 |
| *F1-score (class: congested)* | 0.59 | 0.79 |
| *Overall F1-score* | 0.77 | **0.88** |
| *Location* | **MS230** | |
| *Accuracy* | 0.91 | 0.95 |
| *F1-score (class: congested)* | 0.19 | 0.73 |
| *Overall F1-score* | 0.57 | **0.85** |
| *Location* | **MS443** | |
| *Accuracy* | 0.93 | 0.96 |
| *F1-score (class: congested)* | 0.55 | 0.75 |
| *Overall F1-score* | 0.76 | **0.86** |
| *Location* | **MS634** | |
| *Accuracy* | 0.97 | 0.97 |
| *F1-score (class: congested)* | 0.41 | 0.70 |
| *Overall F1-score* | 0.70 | **0.84** |

The TFTI loss-trained models provide improved performance compared to those trained with the MSE loss. Specifically in terms of the F1-scores, the latter are outperformed, especially for the congested class, as this metric was estimated without considering the percentage of the instances belonging to each class. Indicatively, in most cases, the MSE loss models managed to detect less than 1/3 of the actual congestion incidents. Taking all the above into account, the models trained with the TFTI loss seem to be more trustworthy, in addition to being more accurate.

Finally, it should also be noted that, according to the definitions given above, the exclusive exploitation of causal features in the input space further increases the trustworthiness of the proposed framework, as the significance of their impact on the output can be considered independent of the specific dataset used in this work.

## 8.4 Summary of Findings

In this work, we propose a holistic approach for actionable and trustworthy short-term traffic forecasting consisting of the following modules: a causal Granger LSTM network that is used to detect causal relations among the traffic data and reduce the dimensionality of the input space, a multitask LSTM network, enhanced with the novel TFTI loss function, and an inclusive assessment framework. By using a multitask approach, the input of the forecasting model is a traffic state, instead of a single variable, and the model can distinguish between congested and uncongested conditions, while also learning the corresponding dynamics of each state. According to recent literature and the outcomes of our experiments, multitask learning is very suitable for traffic forecasting applications.

Moreover, the use of the TFTI loss, which is the most important contribution of this work, was found to increase both the overall accuracy of the two variables and the training efficiency. In addition, although it is compatible with any deep learning model, we showcase that it can achieve very satisfying performance even when using a relatively simple structure.

Finally, a deeper evaluation of the trustworthiness of the models showed that decent values of error metrics do not always guarantee the trustworthiness of the predictions, especially in the case of imbalanced datasets, such as traffic datasets, where congested conditions are underrepresented. Thus, a dedicated evaluation strategy is proposed, which considers both congested and uncongested classes equivalently. The predictions of the models that were trained with the TFTI loss were found to be more trustworthy, as their distribution is closer to the actual and the models succeeded in forecasting the emergence of congested conditions with a significantly higher precision.

# 9 CONCLUSIONS

## 9.1 Overview

In this dissertation, we presented a toolkit of actionable methods and techniques to deal with different forecasting tasks, namely multimodal predictions, network-wide multitask predictions, traffic theory-informed forecasting, statistical and causal relations detection, as well as a framework for assessing the predictions' trustworthiness. Except for the different scales, the proposed modules can be deployed for different predicting horizons (short and long-term predictions), using higher or lower resolution data. Moreover, it was established that, using the above methods, one may achieve state-of-the-art forecasting performance, even with relatively simple modeling architectures, although the entire framework is compatible with any Deep Learning model. Thus, a traffic management authority equipped with the above toolkit would be able to apply predictive traffic management, planning and decision-making to handle almost any situation at a city-level and under any conditions.

To achieve its objectives, the dissertation employed a multi-stage analysis; we first provided a thorough review of recent literature regarding the challenges that are related to the development of actionable traffic forecasting schemes. The first challenge is linked to the limitations of state-of-the-art Deep Learning models, including the extensive computational resources and data requirements, which are not always available, and the low interpretability, especially for the more complex structures. Another important aspect is the representation of the road network and the spatiotemporal relations of the input space. The three most popular representations are the stacked vector, the image (or grid) and the graph. Although the stacked vector representation is the simplest one and is suitable for problems with few locations, the grid and, especially, the graph representations allow the modeling of more complex spatial relations, based on the connectivity, proximity and similarity of the traffic patterns of the locations of the road network. Further analysis of the published works on short-term traffic forecasting indicated that, although simpler representations and modeling techniques can be effective in road networks with few locations, they are outperformed by the graph representation and the corresponding modeling techniques in the case of bigger and more complex ones. In addition, an accurate and meaningful representation of the road network can lead to reduced dimensionality of the input space and, thus, to a prediction model of lower complexity. Such a representation also provides insights into the spatial and temporal dependencies between the traffic conditions at different locations of the network, which are vital for interpreting the results of the model and traffic management.

Moreover, causation detection methods should be preferred compared to statistical for modeling the spatiotemporal relations of road networks, as they provide a more accurate and generalizable (not dataset-specific) outcome, while also increasing the trustworthiness of the developed model. Granger causality test is a very suitable method, as it is designed for time series data and it is multivariate, i.e., examines the existence of causal relation between two variables considering the rest variables as well, in contrast to other methods that are pairwise.

In addition, the concept of Multitask Learning is proposed for traffic state forecasting (two variables constitute a traffic state) instead of a single variable, which is also capable of increasing both the accuracy and the trustworthiness of the forecasts. Multitask Learning can also be utilized for multimodal forecasting, which has significant implications for integrated traffic management of all means of transport at a city level. Novel Deep Learning concepts, such as Physics-Informed Neural Networks, are also capable of increasing the forecasting accuracy, while enhancing the trustworthiness, explainability and robustness of the model, by incorporating aspects of traffic flow theory in the modeling process.

Finally, based on the findings of recent literature, some future directions of research in traffic forecasting are proposed. To increase the actionability of the models, researchers should focus on less complex and more interpretable modeling techniques. Moreover, the development of multi-task prediction models for different traffic variables, e.g. multimodal demand, as well as the analysis of the relations between them will be a significant breakthrough with many implications in traffic management. The same applies also to efficient, network-wide modeling approaches, enabled by innovative concepts of Computational Science, such as Federated Learning, which remain a quite under-researched topic, concerning their application in the area of traffic forecasting. Finally, methods for transparently and accurately extracting causal spatial and temporal relations between the locations of the road network and describing their nature and effects are deemed necessary for decision-making under extreme and non-recurrent conditions but would also enhance the trustworthiness and actionability of the model.

To this end, the proposed toolkit of the developed modules attempted to fill the identified literature gaps. First, using statistical and information-theoretic methods, we detected significant spatiotemporal relations that, indeed, when used for feature selection, increase the corresponding models' performance. However, it is not guaranteed that these relations are generalizable; thus, we proceeded to detect causal relations that are independent of the specific dataset used and increase the model's interpretability and trustworthiness, using the concept of Neural Granger. Moreover, using these relations, we achieved enhanced prediction accuracy in more complex and challenging tasks, e.g., network-wide short-term forecasting, while dramatically decreasing the dimensionality of the input space. In addition, significant traffic patterns were revealed at a city level, namely that the perimeter and distant locations are more strongly correlated with the locations close to the city center than neighboring locations and that certain significant locations affect most of the rest.

Then, a Multiplex Network, adapted from the research area of Social Network Analysis, is proposed for representing the spatiotemporal relations between a multimodal input space, consisting of measurements of traffic volume of loop detectors and ticket validations from metro stations. A community detection approach is followed to identify significant spatiotemporal patterns between both types of locations, resulting in a very satisfying performance for both modes, even when using simpler, shallow Machine Learning models.

For predicting the network-level traffic conditions at the entire network or at least a subnetwork, we proposed an efficient multitask approach (multivariate output model) based on a relatively shallow neural network that achieved adequate accuracy in jointly forecasting the hourly travel times of 30 routes close to the city center of Athens, using as input loop detector measurements of hourly volume from neighboring road sections. The results of the explainability (SHAP) analysis indicated that the model detects significant relations between not only the travel times of the routes and loop detectors that lie within them but also with more distant detectors, which justifies the deployment of the multitask model. This framework can be very useful for traffic management purposes, as it is significantly more efficient than using a separate model for each location of interest; in practice, it provides the decision-makers or the users with all the necessary information for the entire road network, without requiring much time or effort from their side or extreme amounts of data that are difficult to obtain in real time.

Finally, to deal with the challenging task of short-term forecasting at specific locations, we incorporated traffic flow theory aspects in the following ways: First, we used a multivariate approach, i.e., the model's output and input includes both traffic volume and speed measurements, the pairs of which constitute a traffic state, instead of a simple measurement and, thus, the model learns to distinguish between congested and uncongested traffic states and their different dynamics. Secondly, we used the novel TFTI loss for training the forecasting model, which considers the distance of the predictions from the corresponding fundamental diagram. As indicated by the dedicated evaluation framework that was also developed for this dissertation, the two above-mentioned approaches not only significantly increased the overall accuracy of the forecasting model, but also led to more trustworthy predictions that are less vulnerable to noisy data and outliers.

## 9.2 Main Contributions and Innovation

The main contribution of this work is a novel traffic flow theory-informed multitask framework, which is used for the joint short-term forecasting of two traffic variables (Q4, Q5). For training the model, we propose a custom-made loss function, which incorporates the distance of the emerging multivariate forecast (pairs of traffic variables) from the real traffic curve of the corresponding location. To enhance the model's performance and interpretability, network-level traffic information is selected from the most relevant locations, specific to each target location, using the Neural Granger adaptation of classic Granger causality, which is also capable of revealing significant traffic patterns at a city level. For the experiments presented in this work, we deploy a Long Short-Term Memory (LSTM) Network; nevertheless, the entire methodology (including the loss function) is compatible with any Deep Learning structure.

The anticipated advantages of the proposed framework compared to existing literature are the following:

- It is more generalizable (not data-specific), i.e., it is expected to perform better with new, previously unseen data, as it is causal and theory-driven.
- It is less vulnerable to noisy data and outliers and, thus, more resilient, as the loss function used to train the model relies on the fundamental diagram of the corresponding location and not only on the available measurements.
- It is expected to have increased overall accuracy for both output variables, which is the main advantage of multitask models, as the inputs will be traffic states and not simple variables.
- Due to the above aspects, the model has increased trustworthiness, as indicated by the corresponding results.
- It enables accurate predictions with less complex modeling structures.

Moreover, we proposed a novel approach for combining multimodal data (road traffic and public transport demand), to increase the predictability of the future values of both modes (Q1, Q2). Indeed, it was shown that there exist significant relations between the two datasets at a city level and the proposed framework can be used for effective multimodal traffic management. To enable efficient and accurate forecasting, an innovative road network and input data representation is proposed for multimodal settings, based on the concept of Multiplex Networks from the research area of Social Network analysis, which is utilized for the first time in traffic forecasting. Each layer corresponds to an hour of the day and, using the multi-layer graph adaptation of the Louvain algorithm, we were able to detect inter-layer communities that correspond to complex traffic patterns. The described feature selection strategy leads to very accurate predictions using a shallow Machine Learning model.

Moreover, we also presented one of the first attempts for network-wide forecasting, in the sense that, not only does the model's input include information from several or all locations of the road network, but also, and most importantly, its output refers to all locations as well, which are predicted simultaneously using a single model, based on the concept of Multitask Learning (Q3). The model that is deployed for this task is relatively simple and efficient in terms of training time; thus, the proposed framework is also suitable for real-world operation.

Another important contribution of the present work is that we propose a novel evaluation framework that assesses the model's trustworthiness and robustness to noisy and imbalanced datasets, which are related to the model's actionability as well (Q5). The evaluation is based on the distribution of the predictions and the corresponding traffic states and requires a multivariate output. In contrast, in recent literature, models are only evaluated based on the classic error metrics of a single variable, which can be very misleading in several cases, as discussed earlier. Especially for traffic forecasting tasks, where the distinction between congested and uncongested conditions is essential, the above framework can provide valuable insights and lead to significant conclusions regarding a model's properties.

Finally, the framework that we proposed is designed not to rely too much on complex Deep Learning structures, while succeeding in achieving state-of-the-art performance. The latter is ought to the spatiotemporal and causation analysis previously conducted, the theory aspects incorporated and the problem representation, respectively, that enable accurate predictions.

## 9.3 Limitations

In this work, we attempted to develop several forecasting tools that may be used for different traffic management and decision-making tasks, spanning from network-wide to multi-modal settings. As with most data-driven frameworks, the main limitations of this work are related to the data availability and quality. First, in case this approach is to be used in real-world conditions, a long sequence of data and extended network coverage is necessary, for the effective training of the model, the detection of meaningful causal (or statistical) correlations and for network-level forecasting. Moreover, for the multimodal forecasting framework, equivalently high amounts of data for other modes, e.g., public transport or other mobility services demand, are required. Such data are often more difficult to obtain and, especially, in high temporal resolution. More specifically, in this work we had sufficient availability of low-resolution data (1-hour timestep), but relatively low availability of high-resolution data: The dataset that is used in Section 8 for the theory-informed forecasting module covers a period of 40 days, which is considered relatively short. Moreover, they refer exclusively to traffic volume measurements, not including any other modes of transportation. High-resolution data are necessary for shorter-term forecasting which is a more challenging and interesting task from a research perspective, but, due to the above limitations, it was not possible to use shorter predicting horizons for the modules of multimodal and network-wide forecasting.

Regarding the quality of the data, the proposed solutions heavily rely on the validity of the input data and, thus, they are vulnerable to technical malfunctions of the sensing system that may provide noisy data (or fail to collect data at all). Thus, the performance of the modules developed may has been underestimated, due to the existence of erroneous data. Besides, for the multitask traffic state prediction, we used traffic volume and speed pairs, instead of volume-occupancy pairs, as there was a systematic missing rate in the occupancy measurements, at least for the available dataset. Although the proposed framework is generalizable to any pair of traffic variables and is expected to have similar performance, regardless of the variables used, occupancy measurements would have been preferred over speed, as they are considered to be more representative, especially in high-resolution datasets, where a relatively small number of vehicles pass from the corresponding location during each timestep.

Another limitation of using the developed modules for traffic management purposes is the high computational requirements for training the prediction models, but also for the data preparation and feature engineering tasks, e.g., Causal Granger and development of a multiplex network with a lot of nodes and edges. Throughout the present dissertation, we attempted to reduce complexity as much as possible, especially for the forecasting models; however, it is still not guaranteed that an average computer would be able to handle the framework's complexity efficiently. During the development of the modules, the issue of the time-consuming fitting process prevented us from testing more complex modeling structures and, using the corresponding datasets, training models for other cities to compare the results, which are now left for future research.

Finally, although the methodological framework is transferable to other road networks (as soon as the corresponding data become available), the trained models are only compatible with the specific road network. Therefore, new models should be trained to be exploited, even at other locations of the same road network. Given the amount of data, time and effort required to train the models, the transferability of the already trained ones would be essential and would significantly increase the chances of them being used. It should be noted though that the transferability of the developed models was not assessed in the present dissertation and is possible that with a little effort in the parameters' calibration and fine-tuning, they may provide decent results in other locations as well.

## 9.4 Future Research

For the purposes of this dissertation, several significant steps were made towards more actionable traffic forecasting models, in terms of trustworthiness, efficiency, performance and the way the forecasting problem is posed and handled (e.g., multimodal and network-wide forecasting). However, several aspects of the proposed framework can be further improved and new paths can be investigated.

Firstly, we will attempt to incorporate causality and theory-informed features into the multimodal forecasting module. The mutual information relational graph that is used at each layer of the multiplex network can be replaced by a causal graph, i.e., connections will be established between nodes that share a causal relationship. The theoretical aspects can be incorporated via the loss function, in a similar way to the proposed in this dissertation, for each individual mode that is considered, but also as a submodule that would provide predictions for each mode based on an analytical or time series analysis model, which would consequently be combined with the predictions of a Deep Learning model, in an ensemble scheme.

Furthermore, traffic flow theory aspects can be incorporated into the network-level travel time forecasting module. For example, the Bureau of Public Roads (BPR) function is an analytic function that expresses the travel time to cross a road section as a function of traffic volume. The BPR function could be incorporated into the loss function of a Deep Learning model to enhance the predictability of travel times. Moreover, according to the most often reported advantages of Physics-Informed Neural Networks, it has the potential to improve the model's performance and trustworthiness.

Consequently, the most significant modules developed for this dissertation can be combined to create a causal, multimodal, theory-informed and network-wide forecasting framework. Although such a model would be inclusive (i.e., provides predictions for all modes and for all locations simultaneously), it should be noted that it be very complex and its training and operation would lack the efficiency of the separated modules; therefore, it should only be used if the corresponding application justifies its necessity.

Our future work will also include the exploitation of more complex modeling structures, to investigate whether even lower forecasting errors can be achieved. In the present work, we preferred to focus on techniques to improve a forecasting framework's actionability, which would reach state-of-the-art performance even when utilizing simple modeling structures, in contrast to the more common practice of constructing a complex deep learning structure and comparing it with previous approaches from the literature, neglecting its efficiency, trustworthiness and actionability. For example, as the road network can be effectively represented as a single- or multi-layer graph, an obvious direction would be to use a Graph Convolutional Neural Network instead of the simpler structures deployed in this work. Moreover, other related techniques, such as an Attention Mechanism which is considered very effective in cases of large input space, could also improve the forecasting accuracy, without excessively increasing the complexity.

Another aspect that could also improve the performance of the proposed framework is the inclusion of geometrical/physical properties of the road network, instead of just the relational (causal or statistical) information. A combination of physical connectivity and causal relationships could be worth examining. Moreover, we will assess the effectiveness of the proposed approach in multi-step and longer-term forecasting. In general, long-term forecasting requires a quite different approach both in terms of the modeling and the data preparation and engineering methodology, which also has significant implications for planning and design purposes.

If any module of the present framework is going to be used for traffic management or decision-making purposes, human supervision of the outcomes, the data collection and the training processes are considered vital; otherwise, erroneous and harmful decisions for the road network's conditions may be made. The same applies to the case of non-recurrent conditions, e.g., in cases of heavy congestion due to a road accident, road closure, or a special event: although causal models are expected to perform well even under such extreme conditions, human supervision is also necessary as any Deep Learning model's performance may become unstable and unreliable when previously unseen conditions occur. Therefore, a Human-in-the-loop framework will be considered to supervise the training process and allow the user to intervene when there is a high risk of a highly erroneous prediction during the operation of the model.

Finally, another interesting direction is the application of the methodology on road networks of other cities, to examine whether similar results will emerge in terms of the forecasting performance and the detection of causal relations and associated traffic patterns. Furthermore, we will assess the transferability of the already trained models, by using them to forecast the traffic conditions at other locations with similar characteristics, both in the same road network (Athens) and in other cities where data are available.

# REFERENCES

Aas, K., Jullum, M., & Løland, A. (2021). Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *Artificial Intelligence*, *298*, 103502. https://doi.org/10.1016/j.artint.2021.103502

Abarbanel, H. (2012). *Analysis of Observed Chaotic Data*. Springer Science & Business Media.

Afrin, T., & Yodo, N. (2022). A Long Short-Term Memory-based correlated traffic data prediction framework. *Knowledge-Based Systems*, *237*, 107755. https://doi.org/10.1016/j.knosys.2021.107755

Amato, R., Kouvaris, N. E., Miguel, M. S., & Díaz-Guilera, A. (2017). Opinion competition dynamics on multiplex networks. *New Journal of Physics*, *19*(12), 123019. https://doi.org/10.1088/1367-2630/AA936A

Bai, J., Zhu, J., Song, Y., Zhao, L., Hou, Z., Du, R., & Li, H. (2021). A3T-GCN: Attention Temporal Graph Convolutional Network for Traffic Forecasting. *ISPRS International Journal of Geo-Information*, *10*(7), 485. https://doi.org/10.3390/ijgi10070485

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Berndt, D., & Clifford, J. (1994). Using Dynamic Time Warping to Find Patterns in Time Series. *Proceedings of AAAI-94 Workshop on Knowledge Discovery in Databases*, 229–248.

Bogaerts, T., Masegosa, A. D., Angarita-Zapata, J. S., Onieva, E., & Hellinckx, P. (2020). A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data. *Transportation Research Part C: Emerging Technologies*, *112*(March), 62–77. https://doi.org/10.1016/j.trc.2020.01.010

Bonaccorso, G. (2017). *Machine Learning Algorithms*. Packt Publishing.

Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., Kiddon, C., Konečný, J., Mazzocchi, S., McMahan, B., Van Overveldt, T., Petrou, D., Ramage, D., & Roselander, J. (2019). Towards Federated Learning at Scale: System Design. *Proceedings of Machine Learning and Systems*, *1*, 374–388.

Boukerche, A., & Wang, J. (2020). Machine Learning-based traffic prediction models for Intelligent Transportation Systems. *Computer Networks*, *181*(August), 107530. https://doi.org/10.1016/j.comnet.2020.107530

Bramich, D. M., Menéndez, M., & Ambühl, L. (2022). Fitting Empirical Fundamental Diagrams of Road Traffic: A Comprehensive Review and Comparison of Models Using an Extensive Data Set. *IEEE Transactions on Intelligent Transportation Systems*, *23*(9), 14104–14127. https://doi.org/10.1109/TITS.2022.3142255

Bródka, P., Chmiel, A., Magnani, M., & Ragozini, G. (2018). Quantifying layer similarity in multiplex networks: A systematic study. *Royal Society Open Science*, *5*(8). https://doi.org/10.1098/rsos.171747

Cai, P., Wang, Y., Lu, G., Chen, P., Ding, C., & Sun, J. (2015). A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. *Transportation Research Part C: Emerging Technologies*, *62*, 21–34. https://doi.org/10.1016/j.trc.2015.11.002

Cao, K., Liu, Y., Meng, G., & Sun, Q. (2020). An Overview on Edge Computing Research. *IEEE Access*, *8*, 85714–85728. https://doi.org/10.1109/ACCESS.2020.2991734

Chang, S., Zhang, Y., Han, W., Yu, M., Guo, X., Tan, W., Cui, X., Witbrock, M., Hasegawa-Johnson, M., & Huang, T. S. (2017). Dilated recurrent neural networks. *Advances in Neural Information Processing Systems*, *2017-Decem*(Nips), 77–87.

Chen, W., Chen, L., Xie, Y., Cao, W., Gao, Y., & Feng, X. (2020). Multi-range attentive bicomponent graph convolutional network for traffic forecasting. *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*, *34*(04), 3529–3536. https://doi.org/10.1609/aaai.v34i04.5758

Cheng, S., Lu, F., & Peng, P. (2021). Short-term traffic forecasting by mining the non-stationarity of spatiotemporal patterns. *IEEE Transactions on Intelligent Transportation Systems*, *22*(10), 6365–6383. https://doi.org/10.1109/TITS.2020.2991781

Chiabaut, N., Buisson, C., & Leclercq, L. (2009). Fundamental Diagram Estimation Through Passing Rate Measurements in Congestion. *IEEE Transactions on Intelligent Transportation Systems*, *10*(2), 355–359. https://doi.org/10.1109/TITS.2009.2018963

Cui, Z., Ke, R., Pu, Z., Ma, X., & Wang, Y. (2020). Learning traffic as a graph: A gated graph wavelet recurrent neural network for network-scale traffic prediction. *Transportation Research Part C: Emerging Technologies*, *115*, 102620. https://doi.org/10.1016/j.trc.2020.102620

Cui, Z., Ke, R., Pu, Z., & Wang, Y. (2020). Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values. *Transportation Research Part C: Emerging Technologies*, *118*, 102674. https://doi.org/10.1016/j.trc.2020.102674

Daganzo, C. F. (1997). *Fundamentals of Transportation and Traffic Operations*. NY: Elsevier.

Dai, X., Fu, R., Zhao, E., Zhang, Z., Lin, Y., Wang, F. Y., & Li, L. (2019). DeepTrend 2.0: A light-weighted multi-scale traffic prediction model using detrending. *Transportation Research Part C: Emerging Technologies*, *103*, 142–157. https://doi.org/10.1016/J.TRC.2019.03.022

de Bie, K., Lucic, A., & Haned, H. (2021). *To Trust or Not to Trust a Regressor: Estimating and Explaining Trustworthiness of Regression Predictions* (arXiv:2104.06982). arXiv. http://arxiv.org/abs/2104.06982

Di, X., Shi, R., Mo, Z., & Fu, Y. (2023). Physics-Informed Deep Learning for Traffic State Estimation: A Survey and the Outlook. *Algorithms*, *16*(6), Article 6. https://doi.org/10.3390/a16060305

Do, L. N. N., Taherifar, N., & Vu, H. L. (2019). Survey of neural network-based models for short-term traffic state prediction. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(1), 1–24. https://doi.org/10.1002/widm.1285

Do, L. N. N., Vu, H. L., Vo, B. Q., Liu, Z., & Phung, D. (2019). An effective spatial-temporal attention based neural network for traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, *108*, 12–28. https://doi.org/10.1016/j.trc.2019.09.008

Drake, J. S., Schofer, J. L., & May, A. D. (1967). A Statistical Analysis of Speed Density Hypotheses. *Highway Research Record*, *154*, 53–87.

Du, S., Li, T., Gong, X., & Horng, S.-J. (2019). *A Hybrid Method for Traffic Flow Forecasting Using Multimodal Deep Learning* (arXiv:1803.02099). arXiv. https://doi.org/10.48550/arXiv.1803.02099

Dunne, S., & Ghosh, B. (2012). Regime-based short-term multivariate traffic condition forecasting algorithm. *Journal of Transportation Engineering*, *138*(4), 455–466. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000337

Ermagun, A., & Levinson, D. (2018). Spatiotemporal traffic forecasting: Review and proposed directions. *Transport Reviews*, *38*(6), 786–814. https://doi.org/10.1080/01441647.2018.1442887

Ermagun, A., & Levinson, D. (2019). Spatiotemporal short-term traffic forecasting using the network weight matrix and systematic detrending. *Transportation Research Part C: Emerging Technologies*, *104*(May), 38–52. https://doi.org/10.1016/j.trc.2019.04.014

Fafoutellis, P., Mantouka, E. G., & Vlahogianni, E. I. (2022). Acceptance of a Pay-How-You-Drive pricing scheme for city traffic: The case of Athens. *Transportation Research Part A: Policy and Practice*, *156*, 270–284. https://doi.org/10.1016/J.TRA.2022.01.009

Fafoutellis, P., & Vlahogianni, E. I. (2023a). Traffic Demand Prediction Using a Social Multiplex Networks Representation on a Multimodal and Multisource Dataset. *International Journal of Transportation Science and Technology*. https://doi.org/10.1016/j.ijtst.2023.04.006

Fafoutellis, P., & Vlahogianni, E. I. (2023b). Unlocking the Full Potential of Deep Learning in Traffic Forecasting Through Road Network Representations: A Critical Review. *Data Science for Transportation*, *5*(3), 23. https://doi.org/10.1007/s42421-023-00083-w

Fafoutellis, P., Vlahogianni, E. I., & Del Ser, J. (2020). Dilated LSTM Networks for Short-Term Traffic Forecasting using Network-Wide Vehicle Trajectory Data. *2020 IEEE 23rd International Conference on Intelligent Transportation Systems, ITSC 2020*. https://doi.org/10.1109/ITSC45102.2020.9294752

Fang, W., Zhuo, W., Yan, J., Song, Y., Jiang, D., & Zhou, T. (2022). Attention meets long short-term memory: A deep learning network for traffic flow forecasting. *Physica A: Statistical Mechanics and Its Applications*, *587*, 126485. https://doi.org/10.1016/j.physa.2021.126485

Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian Network Classifiers. *Machine Learning*, *29*(2), 131–163. https://doi.org/10.1023/A:1007465528199

Frye, C., Rowat, C., & Feige, I. (2020). Asymmetric Shapley values: Incorporating causal knowledge into model-agnostic explainability. *Advances in Neural Information Processing Systems*, *33*, 1229–1239. https://proceedings.neurips.cc/paper/2020/hash/0d770c496aa3da6d2c3f2bd19e7b9d6b-Abstract.html

Geron, A. (2017a). *Hands-on Machine Learning with Scikit-Learn, Keras and Tensorflow.* O'Reilly Media, Inc.

Geron, A. (2017b). *Hands-on Machine Learning with Scikit-Learn, Keras and Tensorflow.*

Ghobrial, A., Hond, D., Asgari, H., & Eder, K. (2023). *A Trustworthiness Score to Evaluate DNN Predictions*. https://doi.org/10.1109/AITest58265.2023.00011

Greenshields, B. D., Bibbins, J. R., Channing, W. S., & Miller, H. H. (1935). A Study of Traffic Capacity. *Highway Research Board Proceedings*, *14*(1), 448–477.

Gu, J., Jia, Z., Cai, T., Song, X., & Mahmood, A. (2023). Dynamic Correlation Adjacency-Matrix-Based Graph Neural Networks for Traffic Flow Prediction. *Sensors*, *23*(6), Article 6. https://doi.org/10.3390/s23062897

Gu, Y., Lu, W., Qin, L., Li, M., & Shao, Z. (2019). Short-term prediction of lane-level traffic speeds: A fusion deep learning model. *Transportation Research Part C: Emerging Technologies*, *106*, 1–16. https://doi.org/10.1016/J.TRC.2019.07.003

Gulli, A., & Pal, S. (2017). *Deep Learning with Keras*.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI—Explainable artificial intelligence. *Science Robotics*, *4*(37), eaay7120. https://doi.org/10.1126/scirobotics.aay7120

Guo, J., Liu, Y., Yang, Q. (Ken), Wang, Y., & Fang, S. (2021). GPS-based citywide traffic congestion forecasting using CNN-RNN and C3D hybrid model. *Transportmetrica A: Transport Science*, *17*(2), 190–211. https://doi.org/10.1080/23249935.2020.1745927

Guo, S., Lin, Y., Li, S., Chen, Z., & Wan, H. (2019). Deep Spatial-Temporal 3D Convolutional Neural Networks for Traffic Data Forecasting. *IEEE Transactions on Intelligent Transportation Systems*, *20*(10), 3913–3926. https://doi.org/10.1109/TITS.2019.2906365

Harary, F. (1962). The Determinant of the Adjacency Matrix of a Graph. *SIAM Review*, *4*(3), 202–210.

Hartmann, A. K., & Weigt, Martin. (2005). *Phase transitions in combinatorial optimization problems: Basics, algorithms and statistical mechanics*. 348.

He, S., Luo, Q., Du, R., Zhao, L., & Li, H. (2022). *STGC-GNNs: A GNN-based traffic prediction framework with a spatial-temporal Granger causality graph* (arXiv:2210.16789). arXiv. http://arxiv.org/abs/2210.16789

Heskes, T., Sijben, E., Bucur, I. G., & Claassen, T. (2020). *Causal Shapley Values: Exploiting Causal Knowledge to Explain Individual Predictions of Complex Models* (arXiv:2011.01625). arXiv. https://doi.org/10.48550/arXiv.2011.01625

Hu, H., Lin, Z., Hu, Q., & Zhang, Y. (2021). Attention Mechanism With Spatial-Temporal Joint Model for Traffic Flow Speed Prediction. *IEEE Transactions on Intelligent Transportation Systems*. https://doi.org/10.1109/TITS.2021.3113935

Huang, A. J., & Agarwal, S. (2023). *Physics Informed Deep Learning: Applications in Transportation* (arXiv:2302.12336). arXiv. https://doi.org/10.48550/arXiv.2302.12336

Huang, R., Chen, Z., Zhai, G., He, J., & Chu, X. (2023). Spatial-temporal correlation graph convolutional networks for traffic forecasting. *IET Intelligent Transport Systems*, *n/a*(n/a). https://doi.org/10.1049/itr2.12330

Huo, G., Zhang, Y., Wang, B., Gao, J., Hu, Y., & Yin, B. (2023). Hierarchical Spatio–Temporal Graph Convolutional Networks and Transformer Network for Traffic Flow Forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 1–13. https://doi.org/10.1109/TITS.2023.3234512

Jiang, W., & Luo, J. (2021). *Graph Neural Network for Traffic Forecasting: A Survey*. https://arxiv.org/abs/2101.11174v3

Kaddour, J., Lynch, A., Liu, Q., Kusner, M. J., & Silva, R. (2022). *Causal Machine Learning: A Survey and Open Problems* (arXiv:2206.15475). arXiv. https://doi.org/10.48550/arXiv.2206.15475

Kantz, H., & Schreiber, T. (2004). *Nonlinear Time Series Analysis*. Cambridge University Press.

Karlaftis, M. G., & Vlahogianni, E. I. (2009). Memory properties and fractional integration in transportation time-series. *Transportation Research Part C: Emerging Technologies*, *17*(4), 444–453. https://doi.org/10.1016/j.trc.2009.03.001

Karlaftis, M. G., & Vlahogianni, E. I. (2010). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies*, *19*(3), 387–399. https://doi.org/10.1016/j.trc.2010.10.004

Karlaftis, M. G., & Vlahogianni, E. I. (2011). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies*, *19*(3), 387–399. https://doi.org/10.1016/j.trc.2010.10.004

Kendall, A., Gal, Y., & Cipolla, R. (2018). *Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics*. 7482–7491.

https://openaccess.thecvf.com/content_cvpr_2018/html/Kendall_Multi-Task_Learning_Using_CVPR_2018_paper.html

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, *2*(3), 203–271. https://doi.org/10.1093/comnet/cnu016

Knoop, V. L., & Daamen, W. (2017). Automatic fitting procedure for the fundamental diagram. *Transportmetrica B: Transport Dynamics*, *5*(2), 129–144. https://doi.org/10.1080/21680566.2016.1256239

Koch, J., Maxner, T., Amatya, V., Ranjbari, A., & Dowling, C. (2022). *Physics-informed Machine Learning of Parameterized Fundamental Diagrams* (arXiv:2208.00880). arXiv. https://doi.org/10.48550/arXiv.2208.00880

Kumar, N., & Raubal, M. (2021). Applications of deep learning in congestion detection, prediction and alleviation: A survey. *Transportation Research Part C: Emerging Technologies*, *133*, 103432. https://doi.org/10.1016/j.trc.2021.103432

Lana, I., Del Ser, J., Velez, M., & Vlahogianni, E. I. (2018). Road Traffic Forecasting: Recent Advances and New Challenges. *IEEE Intelligent Transportation Systems Magazine*, *10*(2), 93–109. https://doi.org/10.1109/MITS.2018.2806634

Laña, I., Lobo, J. L., Capecci, E., Del Ser, J., & Kasabov, N. (2019). Adaptive long-term traffic state estimation with evolving spiking neural networks. *Transportation Research Part C: Emerging Technologies*, *101*(April 2018), 126–144. https://doi.org/10.1016/j.trc.2019.02.011

Laña, I., Sanchez-Medina, J. J., Vlahogianni, E. I., & Del Ser, J. (2021). From Data to Actions in Intelligent Transportation Systems: A Prescription of Functional Requirements for Model Actionability. *Sensors*, *21*(4), Article 4. https://doi.org/10.3390/s21041121

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), Article 7553. https://doi.org/10.1038/nature14539

Lee, K., Eo, M., Jung, E., Yoon, Y., & Rhee, W. (2021). Short-Term Traffic Prediction with Deep Neural Networks: A Survey. *IEEE Access*, *9*, 54739–54756. https://doi.org/10.1109/ACCESS.2021.3071174

Lee, M., Lee, S., Choi, M. J., Moon, Y. S., & Lim, H. S. (2017). HybridFTW: Hybrid computation of dynamic time warping distances. *IEEE Access*, *6*, 2085–2096. https://doi.org/10.1109/ACCESS.2017.2781464

Leiser, N., & Yildirimoglu, M. (2021). Incorporating congestion patterns into spatio-temporal deep learning algorithms. *Transportmetrica B*, *9*(1), 622–640. https://doi.org/10.1080/21680566.2021.1922320

Li, J., Xie, N., Zhang, K., Guo, F., Hu, S., & Chen, X. (Michael). (2022). Network-scale traffic prediction via knowledge transfer and regional MFD analysis. *Transportation Research Part C: Emerging Technologies*, *141*(May), 103719. https://doi.org/10.1016/j.trc.2022.103719

Li, L., Qin, L., Qu, X., Zhang, J., Wang, Y., & Ran, B. (2019). Day-ahead traffic flow forecasting based on a deep belief network optimized by the multi-objective particle swarm algorithm. *Knowledge-Based Systems*, *172*, 1–14. https://doi.org/10.1016/J.KNOSYS.2019.01.015

Li, L., Su, X., Wang, Y., Lin, Y., Li, Z., & Li, Y. (2015). Robust causal dependence mining in big data network and its application to traffic flow predictions. *Transportation Research Part C: Emerging Technologies*, *58*, 292–307. https://doi.org/10.1016/j.trc.2015.03.003

Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Processing Magazine*, *37*(3), 50–60. https://doi.org/10.1109/MSP.2020.2975749

Li, Z., Zheng, Z., & Washington, S. (2019). Short-Term Traffic Flow Forecasting: A Component-Wise Gradient Boosting Approach With Hierarchical Reconciliation. *IEEE Transactions on Intelligent Transportation Systems*, *PP*, 1–13. https://doi.org/10.1109/tits.2019.2948381

Liang, Y., Huang, G., & Zhao, Z. (2022). Joint demand prediction for multimodal systems: A multi-task multi-relational spatiotemporal graph neural network approach. *Transportation Research Part C: Emerging Technologies*, *140*, 103731. https://doi.org/10.1016/j.trc.2022.103731

Liao, W., Zeng, B., Liu, J., Wei, P., & Cheng, X. (2022). Taxi demand forecasting based on the temporal multimodal information fusion graph neural network. *Applied Intelligence*, *52*(10), 12077–12090. https://doi.org/10.1007/s10489-021-03128-1

Lin, G., Lin, A., & Gu, D. (2022). Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient. *Information Sciences*, *608*, 517–531. https://doi.org/10.1016/j.ins.2022.06.090

Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, *16*(3), 31–57.

Liu, H., Zhu, C., Zhang, D., & Li, Q. (2023). *Attention-based Spatial-Temporal Graph Convolutional Recurrent Networks for Traffic Forecasting* (arXiv:2302.12973). arXiv. http://arxiv.org/abs/2302.12973

Liu, X., He, P., Chen, W., & Gao, J. (2019). *Multi-Task Deep Neural Networks for Natural Language Understanding* (arXiv:1901.11504). arXiv. https://doi.org/10.48550/arXiv.1901.11504

Liu, Y., Yu, J. J. Q., Kang, J., Niyato, D., & Zhang, S. (2020). Privacy-Preserving Traffic Flow Prediction: A Federated Learning Approach. *IEEE Internet of Things Journal*, *7*(8), 7751–7763. https://doi.org/10.1109/JIOT.2020.2991401

Liu, Z., & Chen, H. (2022). Short-Term Online Taxi-Hailing Demand Prediction Based on the Multimode Traffic Data in Metro Station Areas. *Journal of Transportation Engineering, Part A: Systems*, *148*(6), 05022003. https://doi.org/10.1061/JTEPBS.0000681

Liu, Z., Li, Z., Wu, K., & Li, M. (2018). Urban Traffic Prediction from Mobility Data Using Deep Learning. *IEEE Network*, *32*(4), 40–46. https://doi.org/10.1109/MNET.2018.1700411

Lu, J., Li, C., Wu, X. B., & Zhou, X. S. (2023). Physics-informed neural networks for integrated traffic state and queue profile estimation: A differentiable programming approach on layered computational graphs. *Transportation Research Part C: Emerging Technologies*, *153*, 104224. https://doi.org/10.1016/j.trc.2023.104224

Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., & Wang, Y. (2017). Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. *Sensors (Switzerland)*, *17*(4). https://doi.org/10.3390/s17040818

Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, *54*, 187–197. https://doi.org/10.1016/j.trc.2015.03.014

Magnani, M., Hanteer, O., Interdonato, R., Rossi, L., & Tagarelli, A. (2021). Community Detection in Multiplex Networks. *ACM Computing Surveys*, *54*(3). https://doi.org/10.1145/3444688

Magnani, M., Rossi, L., & Vega, D. (2021). Analysis of multiplex social networks with r. *Journal of Statistical Software*, *98*(8). https://doi.org/10.18637/jss.v098.i08

Manibardo, E. L., Lana, I., & Ser, J. D. (2021). Deep Learning for Road Traffic Forecasting: Does It Make a Difference? *IEEE Transactions on Intelligent Transportation Systems*, 1–25. https://doi.org/10.1109/TITS.2021.3083957

Mantouka, E., Barmpounakis, E., Vlahogianni, E., & Golias, J. (2021). Smartphone sensing for understanding driving behavior: Current practice and challenges. *International Journal of Transportation Science and Technology*, *10*(3), 266–282. https://doi.org/10.1016/j.ijtst.2020.07.001

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics 1943 5:4*, *5*(4), 115–133. https://doi.org/10.1007/BF02478259

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Mo, Z., Shi, R., & Di, X. (2021). A physics-informed deep learning paradigm for car-following models. *Transportation Research Part C: Emerging Technologies*, *130*, 103240. https://doi.org/10.1016/j.trc.2021.103240

Modi, S., Bhattacharya, J., & Basak, P. (2022). Multistep traffic speed prediction: A deep learning based approach using latent space mapping considering spatio-temporal dependencies. *Expert Systems with Applications*, *189*, 116140. https://doi.org/10.1016/J.ESWA.2021.116140

Molnar, C. (2019). Interpretable Machine Learning. In *Book*. https://christophm.github.io/interpretable-ml-book

Nair, R., & Dekusar, A. (2020). Keep it simple stupid! A non-parametric kernel regression approach to forecast travel speeds. *Transportation Research Part C: Emerging Technologies*, *110*, 269–274. https://doi.org/10.1016/j.trc.2019.11.018

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(23), 8577–8582. https://doi.org/10.1073/PNAS.0601602103

Nguyen, L. V., Laval, J.-P., Chainais, P., Iop, A., Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Olah, C. (2015). *Understanding LSTM Networks*. https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Oord, A. van den, Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). *WaveNet: A Generative Model for Raw Audio*. 1–15.

Parr, T., Wilson, J. D., & Hamrick, J. (2020). Nonparametric Feature Impact and Importance. *ArXiv*. http://arxiv.org/abs/2006.04750

Pavlyuk, D. (2019). Feature selection and extraction in spatiotemporal traffic forecasting: A systematic literature review. *European Transport Research Review*, *11*(1). https://doi.org/10.1186/s12544-019-0345-9

Pearl, J. (2000). *Causality: Models, reasoning, and inference* (pp. xvi, 384). Cambridge University Press.

Polson, N. G., & Sokolov, V. (2017). Deep Learning for Short-Term Traffic Flow Prediction. *Transportation Research Part C: Emerging Technologies*, 1–29.

Rahman, R., Zhang, J., Tirtha, S. D., Bhowmik, T., Jahan, I., Eluru, N., & Hasan, S. (2022). A Data-Driven Network Model for Traffic Volume Prediction at Signalized Intersections. *Journal of Big Data Analytics in Transportation*, *4*(2), 135–152. https://doi.org/10.1007/s42421-022-00059-2

Ramezani, M., & Geroliminis, N. (2015). Queue Profile Estimation in Congested Urban Networks with Probe Data. *Computer-Aided Civil and Infrastructure Engineering*, *30*(6), 414–432. https://doi.org/10.1111/mice.12095

Ranjan, N., Bhandari, S., Zhao, H. P., Kim, H., & Khan, P. (2020). City-Wide Traffic Congestion Prediction Based on CNN, LSTM and Transpose CNN. *IEEE Access*, *8*, 81606–81620. https://doi.org/10.1109/ACCESS.2020.2991462

Ribera, M., & Lapedriza, A. (2019). Can we do better explanations? A proposal of user-centered explainable AI. *IUI Workshops*, *2327*, 38.

Ryu, U., Wang, J., Kim, T., Kwak, S., & Juhyok, U. (2018). Construction of traffic state vector using mutual information for short-term traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, *96*(December 2017), 55–71. https://doi.org/10.1016/j.trc.2018.09.015

Ryu, U., Wang, J., Pak, U., Kwak, S., Ri, K., Jang, J., & Sok, K. (2021). A clustering based traffic flow prediction method with dynamic spatiotemporal correlation analysis. *Transportation*, 1–38. https://doi.org/10.1007/s11116-021-10200-9

Salvador, S., & Chan, P. (2007). *FastDTW: Toward accurate dynamic time warping in linear time and space*.

Schwab, P., & Karlen, W. (2019). CXPlain: Causal Explanations for Model Interpretation under Uncertainty. *Advances in Neural Information Processing Systems*, *32*. https://proceedings.neurips.cc/paper/2019/hash/3ab6be46e1d6b21d59a3c3a0b9d0f6ef-Abstract.html

Shi, R., Mo, Z., Huang, K., Di, X., & Du, Q. (2021). *Physics-Informed Deep Learning for Traffic State Estimation* (arXiv:2101.06580). arXiv. https://doi.org/10.48550/arXiv.2101.06580

Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge Computing: Vision and Challenges. *IEEE Internet of Things Journal*, *3*(5), 637–646. https://doi.org/10.1109/JIOT.2016.2579198

Shi, X., Qi, H., Shen, Y., Wu, G., & Yin, B. (2021). A Spatial-Temporal Attention Approach for Traffic Prediction. *IEEE Transactions on Intelligent Transportation Systems*, *22*(8), 4909–4918. https://doi.org/10.1109/TITS.2020.2983651

Shin, Y., & Yoon, Y. (2022). *PGCN: Progressive Graph Convolutional Networks for Spatial-Temporal Traffic Forecasting* (arXiv:2202.08982). arXiv. https://doi.org/10.48550/arXiv.2202.08982

Stathopoulos, A., & Karlaftis, M. G. (2003). A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies*, *11*(2), 121–135. https://doi.org/10.1016/S0968-090X(03)00004-4

Sun, Z., Hu, Y., Li, W., Feng, S., & Pei, L. (2022). Prediction model for short-term traffic flow based on a K-means-gated recurrent unit combination. *IET Intelligent Transport Systems*, *16*(5), 675–690. https://doi.org/10.1049/itr2.12165

Tank, A., Covert, I., Foti, N., Shojaie, A., & Fox, E. (2021). Neural Granger Causality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–1. https://doi.org/10.1109/TPAMI.2021.3065601

Tebaldi, C., West, M., & Karr, A. F. (2002). Statistical analyses of freeway traffic flows. *Journal of Forecasting*, *21*(1), 39–68. https://doi.org/10.1002/for.804

Tedjopurnomo, D. A., Bao, Z., Zheng, B., Choudhury, F., & Qin, A. K. (2020). A Survey on Modern Deep Neural Network for Traffic Prediction: Trends, Methods and Challenges. *IEEE Transactions on Knowledge and Data Engineering*, *14*(8), 1–1. https://doi.org/10.1109/tkde.2020.3001195

Tveite, H. (2014). *The QGIS NNJoin Plugin*. http://arken.nmbu.no/~havatv/gis/qgisplugins/NNJoin/#

Usama, M., Ma, R., Hart, J., & Wojcik, M. (2022). Physics-Informed Neural Networks (PINNs)-Based Traffic State Estimation: An Application to Traffic Network. *Algorithms*, *15*(12), Article 12. https://doi.org/10.3390/a15120447

Vlahogianni, E. I. (2009a). Enhancing predictions in signalized arterials with information on short-term traffic flow dynamics. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, *13*(2), 73–84. https://doi.org/10.1080/15472450902858384

Vlahogianni, E. I. (2009b). Enhancing predictions in signalized arterials with information on short-term traffic flow dynamics. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, *13*(2), 73–84. https://doi.org/10.1080/15472450902858384

Vlahogianni, E. I., Geroliminis, N., & Skabardonis, A. (2008). Empirical and Analytical Investigation of Traffic Flow Regimes and Transitions in Signalized Arterials. *Journal of Transportation Engineering*, *134*(12), 512–522. https://doi.org/10.1061/(ASCE)0733-947X(2008)134:12(512)

Vlahogianni, E. I., Golias, J. C., & Karlaftis, M. G. (2004). Short-term traffic forecasting: Overview of objectives and methods. *Transport Reviews*, *24*(5), 533–557. https://doi.org/10.1080/0144164042000195072

Vlahogianni, E. I., & Karlaftis, M. G. (2013). Temporal aggregation in traffic data: Implications for statistical characteristics and model choice. *Http://Dx.Doi.Org/10.3328/TL.2011.03.01.37-49*, *3*(1), 37–49. https://doi.org/10.3328/TL.2011.03.01.37-49

Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2005). Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach. *Transportation Research Part C: Emerging Technologies*, *13*(3), 211–234. https://doi.org/10.1016/j.trc.2005.04.007

Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, *43*, 3–19. https://doi.org/10.1016/j.trc.2014.01.005

Wang, M., Lai, B., Jin, Z., Lin, Y., Gong, X., Huang, J., & Hua, X. (2018). *Dynamic Spatio-temporal Graph-based CNNs for Traffic Prediction*. http://arxiv.org/abs/1812.02019

Wang, S., Zhang, M., Miao, H., Yu, P. S., Zhang, M., Miao, H., Yu, P. S., & Peng, Z. (2022). Multivariate Correlation-aware Spatio-temporal Graph Convolutional Networks for Multi-scale Traffic Prediction. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *13*(3), 1–22. https://doi.org/10.1145/3469087

Wang, Y., Zhang, D., Liu, Y., Dai, B., & Lee, L. H. (2019). Enhancing transportation systems via deep learning: A survey. *Transportation Research Part C: Emerging Technologies*, *99*(December), 144–163. https://doi.org/10.1016/j.trc.2018.12.004

Wu, N. (2002). A new approach for modeling of Fundamental Diagrams. *Transportation Research Part A: Policy and Practice*, *36*(10), 867–884. https://doi.org/10.1016/S0965-8564(01)00043-X

Wu, X., Guo, J., Xian, K., & Zhou, X. (2018). Hierarchical travel demand estimation using multiple data sources: A forward and backward propagation algorithmic framework on a layered computational graph. *Transportation Research Part C: Emerging Technologies*, *96*, 321–346. https://doi.org/10.1016/j.trc.2018.09.021

Wu, Y., & Tan, H. (2016). *Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework*. http://arxiv.org/abs/1612.01022

Xia, M., Jin, D., & Chen, J. (2022). Short-Term Traffic Flow Prediction Based on Graph Convolutional Networks and Federated Learning. *IEEE Transactions on Intelligent Transportation Systems*, 1–13.

Yao, H., Tang, X., Wei, H., Zheng, G., & Li, Z. (2019). Revisiting Spatial-Temporal Similarity: A Deep Learning Framework for Traffic Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*, 5668–5675. https://doi.org/10.1609/aaai.v33i01.33015668

Ye, J., Xue, S., & Jiang, A. (2021). Attention-based spatio-temporal graph convolutional network considering external factors for multi-step traffic flow prediction. *Digital Communications and Networks*. https://doi.org/10.1016/j.dcan.2021.09.007

Ye, J., Zhao, J., Ye, K., & Xu, C. (2022). How to Build a Graph-Based Deep Learning Architecture in Traffic Domain: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, *23*(5), 3904–3924. https://doi.org/10.1109/TITS.2020.3043250

Yin, X., Wu, G., Wei, J., Shen, Y., Qi, H., & Yin, B. (2021a). Deep Learning on Traffic Prediction: Methods, Analysis and Future Directions. *IEEE Transactions on Intelligent Transportation Systems*, 1–15. https://doi.org/10.1109/TITS.2021.3054840

Yin, X., Wu, G., Wei, J., Shen, Y., Qi, H., & Yin, B. (2021b). Multi-stage attention spatial-temporal graph networks for traffic prediction. *Neurocomputing*, *428*, 42–53. https://doi.org/10.1016/j.neucom.2020.11.038

Yin, Y., & Shang, P. (2016). Forecasting traffic time series with multivariate predicting method. *Applied Mathematics and Computation*, *291*, 266–278. https://doi.org/10.1016/j.amc.2016.07.017

Yu, H., Wu, Z., Wang, S., Wang, Y., & Ma, X. (2017). Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. *Sensors (Switzerland)*, *17*(7), 1501. https://doi.org/10.3390/s17071501

Yu, J. J. Q. (2022). Graph Construction for Traffic Prediction: A Data-Driven Approach. *IEEE Transactions on Intelligent Transportation Systems*, *23*(9), 15015–15027. https://doi.org/10.1109/TITS.2021.3136161

Yu, J. J. Q., Markos, C., & Zhang, S. (2021). Long-Term Urban Traffic Speed Prediction With Deep Learning on Graphs. *IEEE Transactions on Intelligent Transportation Systems*. https://doi.org/10.1109/TITS.2021.3069234

Yuan, P. C., & Lin, X. X. (2017). How long will the traffic flow time series keep efficacious to forecast the future? *Physica A: Statistical Mechanics and Its Applications*, *467*, 419–431. https://doi.org/10.1016/j.physa.2016.10.020

Zeng, T., Guo, J., Kim, K. J., Parsons, K., Orlik, P., Di Cairano, S., & Saad, W. (2021). Multi-Task Federated Learning for Traffic Prediction and Its Application to Route Planning. *2021 IEEE Intelligent Vehicles Symposium (IV)*, 451–457. https://doi.org/10.1109/IV48863.2021.9575211

Zhang, H., Guo, X., & Chang, X. (2022). Randomized Spectral Clustering in Large-Scale Stochastic Block Models. *Journal of Computational and Graphical Statistics*, 1–52. https://doi.org/10.1080/10618600.2022.2034636/SUPPL_FILE/UCGS_A_2034636_SM5596.PDF

Zhang, K., He, F., Zhang, Z., Lin, X., & Li, M. (2021). Graph attention temporal convolutional network for traffic speed forecasting on road networks. *Transportmetrica B*, *9*(1), 153–171. https://doi.org/10.1080/21680566.2020.1822765

Zhang, K., Wu, L., Zhu, Z., & Deng, J. (2020). A Multitask Learning Model for Traffic Flow and Speed Forecasting. *IEEE Access*, *8*, 80707–80715. https://doi.org/10.1109/ACCESS.2020.2990958

Zhang, L., Fu, K., Ji, T., & Lu, C.-T. (2022). Granger Causal Inference for Interpretable Traffic Prediction. *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, 1645–1651. https://doi.org/10.1109/ITSC55140.2022.9922211

Zhang, Q., Chang, J., Meng, G., Xiang, S., & Pan, C. (2020). Spatio-Temporal Graph Structure Learning for Traffic Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(01), 1177–1185. https://doi.org/10.1609/AAAI.V34I01.5470

Zhang, W., Wu, Z., Zhang, X., Song, G., Wang, Y., & Chen, J. (2022). Robust and Hierarchical Spatial Relation Analysis for Traffic Forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 1–17. https://doi.org/10.1109/TITS.2022.3217054

Zhang, W., Yu, Y., Qi, Y., Shu, F., & Wang, Y. (2019). Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning. *Transportmetrica A: Transport Science*, *15*(2), 1688–1711. https://doi.org/10.1080/23249935.2019.1637966

Zhang, Y., Cheng, T., Ren, Y., & Xie, K. (2020). A novel residual graph convolution deep learning model for short-term network-based traffic forecasting. *International Journal of Geographical Information Science*, *34*(5), 969–995. https://doi.org/10.1080/13658816.2019.1697879

Zhang, Z., Li, Y., Song, H., & Dong, H. (2021). Multiple dynamic graph based traffic speed prediction method. *Neurocomputing*, *461*, 109–117. https://doi.org/10.1016/j.neucom.2021.07.052

Zhao, Z., Chen, W., Wu, X., Chen, P. C. Y., & Liu, J. (2017). LSTM network: A deep learning approach for Short-term traffic forecast. *IET Intelligent Transport Systems*, *11*(2), 68–75. https://doi.org/10.1049/iet-its.2016.0208

Zheng, C., Fan, X., Pan, S., Wu, Z., Wang, C., & Yu, P. S. (2021). *Spatio-Temporal Joint Graph Convolutional Networks for Traffic Forecasting*. https://arxiv.org/abs/2111.13684v2

Zheng, C., Fan, X., Wang, C., & Qi, J. (2020). GMAN: A graph multi-attention network for traffic prediction. *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*, *34*(01), 1234–1241. https://doi.org/10.1609/aaai.v34i01.5477

Zhou, X., Ke, R., Cui, Z., Liu, Q., & Qian, W. (2022). STFL:Spatio-temporal Federated Learning for Vehicle Trajectory Prediction. *2022 IEEE 2nd International Conference on Digital Twins and Parallel Intelligence (DTPI)*, 1–6. https://doi.org/10.1109/DTPI55838.2022.9998967

Zhu, Z., Xu, M., Ke, J., Yang, H., & Chen, X. (Michael). (2023). A Bayesian clustering ensemble Gaussian process model for network-wide traffic flow clustering and prediction. *Transportation Research Part C: Emerging Technologies*, *148*, 104032. https://doi.org/10.1016/j.trc.2023.104032

# APPENDIX

## Relevant Publications

*Peer-Reviewed International Journals*

1. Fafoutellis, P., & Vlahogianni, E. I. (2023). Unlocking the Full Potential of Deep Learning in Traffic Forecasting Through Road Network Representations: A Critical Review. *Data Science for Transportation*, 5(3), 1-21.
2. Fafoutellis, P., & Vlahogianni, E. I. (2023). Traffic demand prediction using a social multiplex networks representation on a multimodal and multisource dataset. *International Journal of Transportation Science and Technology.*
3. Fafoutellis, P., & Vlahogianni, E. I. (2024). A Theory-Informed Multivariate Causal Framework for Actionable Short-Term Urban Traffic Forecasting. *Transportation Research Part C: Emerging Technology* (submitted for review).
4. Fafoutellis, P., & Vlahogianni, E. I. (2024). A Multitask Learning Framework for Network-Wide Travel Times Forecasting. (working paper)

*International full paper reviewed Conferences*

1. Fafoutellis, P., Laña, I., Del Ser, J., Vlahogianni, E. I. (2023). A Causal Deep Learning Framework for Traffic Forecasting. In the *IEEE 26th International Conference on Intelligent Transportation Systems (ITSC 2023)*, September 24-28, Bilbao, Bizkaia, Spain.
2. Fafoutellis, P., Karakitsou, E., Vlahogianni, E. I. (2023). Multimodal Traffic Demand Prediction Using a Community Detection Framework on Social Multiplex Networks. In the *102nd Annual Meeting of Transportation Research Board (TRB)*, Washington D.C., US.
3. Fafoutellis P. Vlahogianni E.I. and Del Ser J. (2020). Dilated LSTM Networks for Short-Term Traffic Forecasting using Network-Wide Vehicle Trajectory Data. In the *IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC 2020)*, September 20-23, Rhodes, Greece.
4. Fafoutellis P., Kampitakis E., Vlahogianni E., Koziris N., Yannis G. and Golias J. (2019). Mining spatiotemporal features of city traffic. In the *9th International Congress on Transportation Research (ICTR2019)*, October 24-25, Athens, Greece.

*International abstract reviewed Conferences*

1. Karakitsou E., Fafoutellis P., Vlahogianni E. I. (2022). Efficient Traffic Demand Forecasting Using A Meaningful Representation With Social Multiplex Networks and Community Detection. In the *10th symposium of the European Association for Research in Transportation (hEART)*, Leuven, Belgium.

## Other Publications

*Peer-Reviewed International Journals*

1. Mourtakos, V., Mantouka, E. G., Fafoutellis, P., Vlahogianni, E. I., & Kepaptsoglou, K. (2024). Reconstructing mobility from smartphone data: Empirical evidence of the effects of COVID-19 pandemic crisis on working and leisure. *Transport Policy*, 146, 241-254.
2. Fafoutellis, P., Mantouka, E. G., Vlahogianni, E. I., & Fortsakis, P. (2023). Investigating the impacts of the COVID-19 pandemic on Eco-driving behavior. *Safety science*, 166, 106251.
3. Kampitakis, E. P., Fafoutellis, P., Oprea, G. M., & Vlahogianni, E. I. (2023). Shared space multi-modal traffic modeling using LSTM networks with repulsion map and an intention-based multi-loss function. *Transportation Research Part C: Emerging Technologies*, 150, 104104.
4. Mantouka, E. G., Fafoutellis, P., Vlahogianni, E. I., & Oprea, G. M. (2022). Understanding user perception and feelings for autonomous mobility on demand in the COVID-19 pandemic era. *Transportation Research Interdisciplinary Perspectives*, 16, 100692.
5. Fafoutellis, P., Mantouka, E. G., Vlahogianni, E. I., & Oprea, G. M. (2022). Acceptability modeling of autonomous mobility on-demand services with on-board ride sharing using interpretable Machine Learning. *International Journal of Transportation Science and Technology*, 11(4), 752-766.
6. Fafoutellis, P., Plymenos-Papageorgas, J., & Vlahogianni, E. I. (2022). Enhancing lane change prediction at intersections with spatio-temporal adequacy information. *Journal of Big Data Analytics in Transportation*, 4(1), 73-84.
7. Fafoutellis P., Mantouka E. & Vlahogianni E. (2022). Acceptance of a Pay-How-You-Drive Pricing Scheme for City Traffic: The Case of Athens. *Transportation Research Part A: Policy and Practice*, 156, 270-284.
8. Mantouka E.G., Fafoutellis P., and Vlahogianni E.I. (2021). Deep survival analysis of searching for on-street parking in urban areas. *Transportation Research Part C: Emerging Technologies*, Volume 128, 2021, 103173, ISSN 0968-090X, https://doi.org/10.1016/j.trc.2021.103173.
9. Fafoutellis, P., Mantouka, E.G., Vlahogianni, E.I. (2020). Eco-Driving and Its Impacts on Fuel Efficiency: An Overview of Technologies and Data-Driven Methods. *Sustainability*, 13, 226, https://doi.org/10.3390/su13010226.

*International full paper reviewed Conferences*

1. Konstantinou, C., Fafoutellis, P., Mantouka, E. G., Chalkiadakis, C., Fortsakis, P., & Vlahogianni, E. I. (2023). Effects of Driving Behavior on Fuel Consumption with Explainable Gradient Boosting Decision Trees. In 2023 *8th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)* (pp. 1-6). IEEE.
2. Fafoutellis, P., Kampitakis, E. P., & Vlahogianni, E. I. (2023). A Deep Learning Framework For Modeling Pedestrian–Vehicle Interactions In Shared Space. In 2023 *8th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)* (pp. 1-6). IEEE.
3. Mantouka, E. G., Fafoutellis, P., Tselentis, D., Papadimitriou, E., Vlahogianni, E. I., Yannis, G. (2023). A Multi-level Approach to Link Smooth Driving with Safe Driver

Behavior. In the *102nd Annual Meeting of Transportation Research Board (TRB)*, Washington D.C., US.

4. Mantouka, E., Fafoutellis, P., Katzilieris, K., Vlahogianni, E., & Hoogendoorn-Lanser, S. (2023). Impact Assessment in the Era of Multimodality and Cooperation in Transport. *Transportation Research Procedia*, 72, 3964-3971.
5. Mantouka E., Fafoutellis P., Vlahogianni E. & Fortsakis P. (2022). Investigating the impacts of COVID-19 pandemic on Eco-driving behavior, In the *8th Road Safety and Simulation (RSS 2021).*
6. Mourtakos V., Fafoutellis P., Mantouka E., Vlahogianni E. (2022). Extracting Activity Patterns and Trip Chains from Smartphone Data: A Case Study on the effects of COVID-19 on Greek Mobility. In the *101st Annual Meeting Transportation Research Board (TRB)*, Washington D.C., US.
7. Fafoutellis P., Mantouka E., Vlahogianni E. I., Haxhi A., Perakis H., Gikas V., Pnevmatikou A., Kostoulas G., Frantzola E. K. (2021). What is the Impact of Driving Behavior on Fuel Efficiency? Theoretical Aspects and Empirical Evidence. In the *International Congress on Transportation Research ICTR 2021*, Rhodes, Greece.
8. Haxhi A., Perakis H., Gikas V., Fafoutellis P., Mantouka E., Vlahogianni E. I., Pnevmatikou A., Fortsakis P. (2021). Towards Smartphone-based enhanced GNSS positioning for Eco Driving ITS services. In the *International Congress on Transportation Research ICTR 2021*, Rhodes, Greece.
9. Konstantinou C., Mourtakos V., Fafoutellis P., Mantouka E., Vlahogianni E. I. (2021). Mining Human Mobility Patterns, Activity Chains and Duration using Smartphone Sensing. In the *International Congress on Transportation Research ICTR 2021*, Rhodes, Greece.
10. Mantouka E. G., Fafoutellis, P. and Vlahogianni, E. I. (2019). Data Science Models for Analyzing Cruising for On-Street Parking. In the *99th Annual Meeting Transportation Research Board (TRB)*, Washington, D.C.

*International abstract reviewed Conferences*

1. Katzilieris K., Fafoutellis P., Hryhoryeva M., Leclercq L., Vlahogianni E. I. (2022). Multi-actor fair cooperation; A reality or fallacy?. In the *Transport Research Arena Conference Lisbon 2022.*
2. Mantouka E., Fafoutellis P., Vlahogianni E. & Hoogendoorn-Lanser S. (2022). Impact Assessment in the Era of Multimodality and Cooperation in Transport. In the *Transport Research Arena Conference Lisbon 2022.*
3. Fafoutellis P. and Mantouka E. G. (2019). Major Limitations and Concerns Regarding the Integration of Autonomous Vehicles in Urban Transportation Systems, Data Analytics: Paving the Way to Sustainable Urban Mobility. CSUM 2018 Advances in Intelligent Systems and Computing, vol 879, Nathanail E., Karakikes I. (eds). Springer, Cham