

**БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ
ИНСТИТУТ ПО ИНФОРМАЦИОННИ И
КОМУНИКАЦИОННИ ТЕХНОЛОГИИ**

Петър Руменов Живков

**МОДЕЛИРАНЕ НА КАЧЕСТВОТО
НА АТМОСФЕРНИЯ ВЪЗДУХ ВЪРХУ
ЗДРАВНИ И ИКОНОМИЧЕСКИ АСПЕКТИ**

дисертационен труд за придобиване на
образователна и научна степен "доктор"
в професионално направление 4.6
"Информатика и компютърни науки"

научен ръководител:
проф. д.н. Стефка Фиданова

2024

Съдържание

1	Въведение	9
1.1	Актуалност и мотивировка на темата	10
1.2	Цел и задачи на дисертационния труд	11
1.3	Методология на изследването	11
1.4	Представяне на резултатите	12
1.5	Структура на дисертационния труд	13
2	Въздействие на ФПЧ	15
2.1	Насоки на СЗО	16
2.2	Географска зона на изследване	18
2.3	Методология на изследването	19
2.4	Резултати	22
2.5	Заклучение	24
3	Калибриране	27
3.1	Наблюдение на качеството на въздуха	29
3.2	Модел за калибриране на данни	34
3.3	Приложение на модела върху безжична сензорна мрежа	38
3.4	Резултати и оценка	40
3.5	Заклучение и бъдещи изследвания	42
4	Велосипедно трасе	45
4.1	Формулиране на задачата	47
4.2	Работа с данни в модела	47
4.3	Проект на изследването	48
4.4	Резултати и обсъждане	56
4.5	Заклучение и бъдещи изследвания	61

5	Разработка на софтуер	63
5.1	Софтуерна архитектура	63
5.2	Модул Събиране на Данни	67
5.3	Модул Съхранение на Данни в Облака	67
5.4	Модули за обработка, анализ и прогнозиране на данни	68
5.5	Използвани технологии	69
5.6	Пример за използване на софтуерната инфраструктура	73
5.7	Заклучение	76
6	Заклучение	77
6.1	Научно-приложни приноси	77
6.2	Приложни приноси	78
6.3	Публикации	79
6.4	Декларация за оригиналност	80

Списък на таблиците

2.1	Замърсяване на въздуха: препоръки на СЗО и законодателство на ЕС (в $\mu g/m^3$)	17
2.2	Дневната концентрация на ФПЧ10 и ФПЧ2,5, разделена на работни и неработни дни	22
2.3	Сравнение на непараметрични данни от хоспитализирани пациенти в дни с превишения на нормите на СЗО за ФПЧ10	22
2.4	Сравнение на непараметрични данни от хоспитализирани пациенти в тежки състояния в дни с превишения на нормите на СЗО за ФПЧ2,5	23
2.5	Сравнение на непараметрични данни от амбулаторни прегледи в болници с нормите на СЗО за среднодневна концентрация на ФПЧ10 без отместване и с 1 ден отместване	23
2.6	Сравнение на непараметрични данни от амбулаторни прегледи в болници с нормите на СЗО за среднодневна концентрация на ФПЧ10 с 2 и 3 дни отместване	24
3.1	Характеристики и техническо описание на SDS011-TRF.	33
3.2	Резултати от супервизирани ML модели	40
3.3	Важността на параметрите въз основа на RF.	41
3.4	Сравнение на едностъпков и двустъпков модел.	41
4.1	Концентрации на ФПЧ1 и ФПЧ2,5 от мобилни измервания в работни дни по време на интензивен трафик (HT)	56
4.2	Концентрации на ФПЧ1 и ФПЧ2,5 от мобилни измервания в работни дни със слаб трафик (LT)	57
4.3	Концентрации на ФПЧ1 и ФПЧ2,5 от мобилни измервания през неработни дни	57

4.4	Осреднени показатели от полеви измервания	58
4.5	Доза за вдишване на ФПЧ1 и ФПЧ2,5 по време на обиколките	58
4.6	Концентрации на ФПЧ2,5 от стационарни сензори, разположени в близост до маршрута, и мобилни измервания	61

Списък на фигурите

1.1	Принос на трите основни градски замърсители към общите щети за заболяемост и смъртност в 432 европейски града .	10
2.1	Средна концентрация на ФПЧ _{2,5} в най-замърсените столици в Европа през 2022 г. (в микрограми на кубичен метър въздух)	18
3.1	Двустъков модел на калибриране чрез изкуствена невронна мрежа и намиране на аномалии.	38
3.2	Модел на изкуствени невронни мрежи с откриване на аномалии	42
4.1	Отсечка А - предложена от нашия софтуер, който търси най-малката инхалационна доза	55
4.2	Отсечка Б - предлагана от повечето навигационни софтуери, който следва обособена велосипедна алея	56
4.3	Маршрут за колоездене с експозиция от фиксирани сензори .	60
5.1	Софтуерна архитектура	64
5.2	Източници за измерване на качеството на въздуха	66
5.3	Архитектурна диаграма показваща потока на данни между модулите за обработка на данни	70
5.4	AirLief App - едно от приложенията на разработената софтуерна система, което показва актуално въздушно замърсяване от стационарни сензори	75
5.5	AirLief App - мобилното приложение, което превръща суровите данни в разбираем Индекс за Качеството на Въздуха (ИКВ)	76

Глава 1

Въведение

Замърсяването на въздуха и по-конкретно фините прахови частици (ФПЧ), е водеща екологична причина за заболяемост и преждевременна смъртност, след като през 1990 г.е било на пета позиция [63]. Смъртните случаи, причинени от ФПЧ в околната среда, са нараснали през последните 25 години [18], а излагането на по-високо замърсяване на атмосферния въздух е свързано с по-висока смъртност [11].

ФПЧ се класифицират чрез техния размер, като частиците с размери по-малки от 10 микрометра (ФПЧ₁₀) и 2,5 микрометра (ФПЧ_{2,5}) предизвикват най-голямо безпокойство поради способността им да проникват дълбоко в белите дробове и да причиняват неблагоприятни последици за здравето. ФПЧ представляват сложна смес от твърди частици и течни капки, суспендирани във въздуха, с различни размери и химичен състав. Тези частици могат да произхождат от природни източници като прах и полени, както и от човешки дейности като промишлени емисии, отработени газове от превозни средства и изгаряне на биомаса и твърди горива.

Според проучвания излагането на ФПЧ_{2,5} и ФПЧ₁₀ е свързано с ранна смъртност от няколко заболявания, като сърдечносъдови, респираторни, рак на белия дроб и инфекции на горните и долните дихателни пътища [2], [37],[42]. В това число замърсяването на атмосферния въздух в големите градове увеличава риска от белодробен и системен оксидативен стрес, хипоксемия,имунологични модификации, атеросклероза и ускорена прогресия на сърдечносъдовите заболявания и хроничната обструктивна белодробна болест (ХОББ) [54], а епидемиологични проучвания показват, че излагането на ФПЧ увеличава риска от диабет [98].

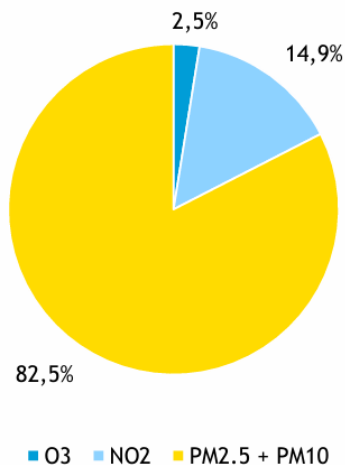
Тежестта на последиците за здравето може да варира в зависимост от концентрацията и продължителността на излагане, както и от здравословното състояние на индивида и неговата чувствителност към замърсителите. Някои групи от населението, като децата, възрастните хора и тези с вече съществуващи респираторни или сърдечносъдови заболявания, са особено уязвими към здравните ефекти от замърсяването на въздуха [75]

Изследванията показват силна връзка между замърсяването на въздуха и хоспитализациите за респираторни заболявания [88, 27, 55, 102]. А кохортен преглед на 175 статии обобщава, че излагането на замърсяване на въздуха увеличава вероятността от развитие на различни заболявания като респираторни заболявания (астма, хронична обструктивна белодробна болест (ХОББ) и рак на белия дроб), съдови заболявания (инфаркти, инсулти и хипертония), бъбречни, чернодробни и други [1].

1.1 Актуалност и мотивировка на темата

Нарастващата загриженост по отношение на въздействието на замърсяването на въздуха върху общественото здраве налага разработването на прогнозни модели за предвиждане нивото на остра заболеваемост въз основа на данните за качеството на въздуха [56]. В последния доклад на Европейския Съюз (ЕС) за качеството на въздуха се оценява, че усилията на ЕС за подобряване на качеството му са само "частично ефективни" и специално изтъква, че действията за намаляване на $\text{FP}_2,5$ са недостатъчни. На фигура 1.1 е представен резултат от проучване с 432 града в ЕС, което анализира икономическите последици от трите основни замърсители в европейските градове: $\text{FP}_2,5$, азотен диоксид и озон [24]. Резултатите показват, че $\text{FP}_2,5$ са основната причина (82,5 %) за икономическите щети в следствие на болести и смъртност от мръсен въздух. Изследването използва програмата за моделиране AIRQ_+ , създадена от СЗО, като включва стойности за замърсяване и икономически данни за всеки от градовете.

В светлината на тези констатации, нараства необходимостта от модели, които да могат да изследват заболеваемостта при остри състояния, базирани на данни за качеството на въздуха. Въпреки нарастващата осведоменост за рисковете за здравето, свързани със замърсяването на въздуха, съществува необходимост от задълбочен анализ на конкретните замърсители и тяхното въздействие върху различни здравни резултати. Същест-



Фигура 1.1: Принос на трите основни градски замърсители към общите щети за заболяемост и смъртност в 432 европейски града

вуващите проучвания показват връзки между замърсяването на въздуха и респираторните заболявания, сърдечносъдовите заболявания, неблагоприятните последици при раждане и дори проблемите с психичното здраве. Въпреки това все още е необходимо цялостно проучване на мащаба и механизмите на тези връзки, за да се информират основаните на доказателства интервенции и политики на държавно и локално ниво.

У нас нивата на ФПЧ често надхвърлят нормите на ЕС за качество на въздуха. ЕК през април 2017 г. спечели дело срещу България заради наднормено замърсяване на въздуха в редица градове. Също така българските градове са лидери по загубени години на живот и икономически ползи, заради мръсен въздух [24]. А според международен доклад, гражданите у нас скъсяват с 2,5 години живота си в следствие на въздушно замърсяване, докато средно за Европейския съюз (ЕС) този показател е 0,7 години (https://www.eca.europa.eu/Lists/ECADocuments/SR18_23/SR_AIR_QUALITY_BG.pdf). Качеството на въздуха превишава нормите основно поради битовото отопление, автомобилния трафик и географските фактори, които задържат замърсителите в атмосферата. Липсата на солидни изследвания и статистически данни за здравните ефекти от замърсяването с ФПЧ в София подкопава способността на града да разра-

ботва основани на доказателства стратегии за намаляване на въздействието му. Без точна и подробна информация създателите на политики и служителите в областта на общественото здраве са изправени пред значителни предизвикателства при прилагането на целенасочени интервенции и повишаването на обществената осведоменост относно сериозността на проблема. Ето защо е налице спешна необходимост от всеобхватни изследователски усилия и подобрени методи за събиране на данни, които да преодолеят съществуващите пропуски в знанията и да проправят пътя за вземане на информирани решения в борбата със замърсяването с ФПЧ и за подобряване на качеството на здравето в София.

1.2 Цел и задачи на дисертационния труд

Целта на дисертационния труд е изследване влиянието на фините прахови частици (ФПЧ) върху острите заболявания в София и намиране на методи за превенция.

За постигането на тази цел са формулирани следните 4 задачи:

- Връзка между фините прахови частици и здравните показатели за остра заболеваемост в София;
- Подобряване на данни от граждански станции за измерване качеството на въздуха чрез калибриране с машинно обучение: двустъпков метод;
- Разработване на софтуерен инструмент за оптимизиране и оценка на велосипедни маршрути чрез характеризирани на излагането на велосипедистите на замърсяване на въздуха;
- Разработване на IoT платформа за агрегация и моделиране на сензорни данни за качество на въздуха.

1.3 Методология на изследването

В настоящия дисертационен труд е използван мултидисциплинарен подход, съчетаващ литературен обзор, количествен анализ и усъвършенствано статистическо моделиране. Извършен е систематичен преглед на съществуващи изследвания, за да се обобщят и синтезират настоящите доказателства

за връзката между качеството на въздуха и здравните резултати. Освен това са анализирани епидемиологични проучвания и данни от агенции за здравно наблюдение, за да се определи количествено връзката между конкретни замърсители и последиците за здравето. Използвани са съвременни статистически техники, като регресионно моделиране и пространствен анализ, за да се изследват сложните взаимодействия и пространствените модели на замърсяването на въздуха и здравните последици.

Методологията, използвана в този дисертационен труд, включва систематичен подход за изследване на въздействието на фините прахови частици върху острите заболявания в София и разработване на стратегии за смекчаване на последиците. Методологията включва събиране на данни от болници и измервателни станции, анализ на данните с цел установяване на корелации, калибриране на граждански лазерни станции, разработване на IoT платформа за визуализация на данните и оптимизиране на велосипедни алеи.

Събирането на данни за постъпленията и хоспитализациите от две големи болници в София, данни от регистъра на бърза помощ и измервания на ФПЧ както от официални измервателни станции на Изпълнителната агенция по околната среда (ИАОС), така и от граждански лазерни станции. Събрани са и метеорологични данни, включително влажност, атмосферно налягане и температура, данни за трафик и други.

Приложени са техники за анализ на данните, за да се изследва връзката между нивата на ФПЧ и острите заболявания. Извършен е статистически анализ и сравнения на нивата на ФПЧ под и над здравните стандарти на Световната здравна организация (СЗО).

Процесът на калибриране се съсредоточи върху калибрирането на данните от гражданските лазерни станции, като са използвани модели за машинно обучение с контролиран и неконтролиран метод. За точното калибриране са използвани референтни данни от измервателни станции на МОСВ, като са взети в предвид влиянието на фактори като влажност, атмосферно налягане и температура.

Разработен е модел за оптимизиране на велосипедните трасета за минимално вдишване на ФПЧ. Това включва използване на модифициран алгоритъм за намиране на най-краткия път и провеждане на реални полеви тестове за валидиране на методологията.

Разработена е софтуерна система, която визуализира местоположението и показателите на измервателните станции върху географска карта.

Приложен е нов подход за агрегиране, организиране, обработка, моделиране и обмен на данни в IoT системата.

Представената методология позволява цялостен анализ на въздействието на ФПЧ върху острите заболявания, точно калибриране на гражданските лазерни станции, визуализиране на данните чрез IoT платформата и оптимизиране на велосипедните пътеки за намаляване на вдишването на ФПЧ.

1.4 Представяне на резултатите

В съответствие с поставените задачи са разработени алгоритми и методологии решаващи конкретни проблеми. Направена е програмна реализация на всеки от разработените алгоритми. Програмите са написани на езика Python и използват рамката Django.

Публикации

Авторът на дисертационния труд има над 7 публикации. Резултатите от дисертационния труд са публикувани 4 публикации, от които 3 с импакт ранг, реферирани в световната система за реферирание и индексирание и също така публикувани в международни издания и токове от конференции.

1.5 Структура на дисертационния труд

Дисертационният труд се състои от шест глави: уводна глава, последвана от четири глави, които разглеждат всяка една от задачите и заключение.

Дисертацията съдържа:

- 115 страници;
- 17 таблици;
- 12 графики;
- 108 литературни източника.

Във втора глава е направен аналитичен обзор на влиянието на ФПЧ върху остри заболявания в София. Разгледани са данни от 2 големи софийски болници (УМБАЛСМ Пирогов и УМБАЛ Токуда) и случаите в

столична бърза помощ. Сравнени са измененията на болничните постъпления и хоспитализации в дни на въздушно замърсяване под и над здравословните норми на СЗО. Мотивирана е необходимостта от използване на граждански лазерни станции за допълване на представителните данни от официалните измервателни станции на МОСВ. Резултатите от тази глава са публикувани в [102].

В трета глава е представен дву-стъпков метод на автоматична калибровка на данни от граждански лазерни станции за измерване на ФПЧ посредством използването на контролирани и неконтролирани machine learning модели. За реперни данни са използвани официални измервателни станции на МОСВ. В моделите са изследвани и включени влиянието на влажност, атмосферно влияние и температура. За калибриране на данните са използвани 5 machine learning модела, сравнена е точността на техните резултати. Резултатите от тази глава са публикувани в [101].

В четвърта глава е описан модел за намиране на най-подходящ велосипеден път с оглед минимална инхалация на ФПЧ от страна на колоездача. За целта е използван модифициран алгоритъм за намиране на най-кратки пътища. Направена е експериментална реализация и валидиране на разработената методология чрез реални тестове в полеви условия. Разгледани са и са сравнени резултатите от замърсяване и инхалация на ФПЧ на два маршрута - този, който е предложен от софтуера и маршрут, който следва велоалеи. Резултатите от тази глава са публикувани в [103].

В пета глава е описан процесът на разработване на архитектура на IoT платформа, при която е използван нов подход за агрегация, организация, интелигентна обработка и обмен на данни в IoT системата. Описано е изграждането на системата, която показва местоположението и показателите на измервателните станции върху географска карта в [104].

Шеста глава е заключението, където са обобщени получените резултати в дисертационния труд и са посочени научните и научно-приложните приноси. Определени са бъдещи насоки за изследвания и развитие. Предложени са стратегии за справяне с предизвикателствата, свързани със замърсяването на въздуха и неговото въздействие върху човешкото здраве.

Резултатите от този дисертационен труд ще допринесат за научното познание в областта на ФПЧ, разработването на политики и информираността на обществеността, което се очаква да доведе до подобряване на управлението на качеството на въздуха и до по-добра защита на човешкото здраве.

Глава 2

Връзка между замърсяването с ФПЧ и здравни индикатори за остра заболяемост

Резултатите от тази глава са публикувани в [102].

Връзките между градското въздушно замърсяване (ВЗ) и човешкото здраве са последователно и ясно установени от редица изследователи [89], [80], [52], като открояващи групи заболявания са сърдечно-съдови [91], невроаскуларни [87], [13] и белодробни [90]. Според скорошно усъвършенствано моделиране се предполага, че има близо 9 милиона смъртни случая годишно от ВЗ [71]. Около 25% от преждевременните смъртни случаи, свързани с ВЗ, са респираторни по презумция [6].

Значителна литература за епидемиологични проучвания предполага връзка между острата заболяемост и излагането на замърсяване на въздуха от ФПЧ [26]. Повечето от тези данни идват от анализи на времеви серии [33], сравняващи вариациите в хоспитализацията със средните вариации на прахови частици [39]. ВЗ е отговорно за възпаленията на дихателните пътища [17] и като се има предвид, че дихателната система е обща врата за влизане, минимизирането на експозицията на дихателните пътища също минимизира сърдечно-съдовите проблеми [83]. Освен това съществуват проучвания, които моделират замърсяването на множество градове, като Европейското замърсяване на въздуха и здравето: европейски подход (ARNEA) [47] и Американското национално изследване на заболяемостта, смъртността и замърсяването на въздуха (NMMAPS) [81], като и двете

предоставят последователни доказателства за връзка здраве и замърсители на въздуха за множество градове чрез покриване на голяма географска област.

Последиците от замърсяването на въздуха могат да се разглеждат като повишаване на риска от заболяване или нараняване на дадено лице или като допълнителен общ риск за благосъстоянието на населението [31]. Целта на управлението на качеството на въздуха е да контролира или избягва неблагоприятните въздействия на замърсяването на въздуха върху общественото здраве. Следователно е важно, чрез стандартите за ФПЧ, да се дефинират такива въздействия, които се считат за „неблагоприятни“, и да се разграничат от тези въздействия, които не се считат за неблагоприятни, като по този начин усилията за защита се концентрират върху замърсителите, които причиняват най-екстремни въздействия върху здравето.

2.1 Стандарти за ФПЧ: Анализ на законодателството на ЕС и насоки на СЗО

Този раздел изследва процесите за установяване и изпълнение на стандартите за качество на въздуха с ФПЧ в ЕС и насоките, изготвени от СЗО. Основната цел е да предостави комплексно разбиране за начина, по който тези стандарти се разработват, изпълняват и основните различия между тях. В края на раздела ще разгледаме, кои са стойностите на замърсяване от фини прахови частици, над които се счита, че въздухът е нездравословен.

2.1.1 Законодателство и изпълнение на директивите за качество на въздуха с ФПЧ

Законодателната рамка на ЕС за нивата на ФПЧ във въздуха се основава на Директива за Качество на Въздуха (2008/50/ЕО) и се допълва от Директива за Национални Емисионни Прагове (2016/2284/ЕС). Тези регулации задължават държавите-членки да изпълняват стандартите за ФПЧ във въздуха в съответствие със специфични прагови стойности, извлечени от научни оценки. Съответствието се наблюдава от Европейската комисия, а неспазването може да доведе до правни действия и глоби. В сравнение,

СЗО предоставя общи указания за качество на въздуха без законодателни задължения, оставяйки изпълнението на индивидуалните държави.

Насоките на СЗО за замърсяването на въздуха могат да се разглеждат по-скоро като препоръки, те нямат задължителен характер. Докато Директивата на ЕС за качеството на въздуха може да се разглежда като законодателство, което всеки член на ЕС трябва да следва.

Изпълнението на стандартите на ЕС за ФПЧ във въздуха е главно отговорност на държавите-членки, които адаптират регулациите към националния си контекст. Те разработват планове за управление на качеството на въздуха и редовно представят данни за въздушното качество и емисиите, докато Европейската комисия следи спазването. Несъответствия могат да доведат до правни действия и глоби. В сравнение, указанията на СЗО зависят от инициативите на индивидуалните държави за изпълнение и не разполагат с формален механизъм за наблюдение. Основното различие е в това, че стандартите на ЕС са правно обвързващи.

2.1.2 Методология за стандартите за качество на въздуха с ФПЧ

И ЕС, и СЗО използват конкретни методологии за установяване на стандартите за качество на въздуха с ФПЧ. ЕС, ръководен от Европейската комисия, започва регулации на базата на оценки за въздействие, като се вземат предвид икономически, социални и околни въздействия.

Методологията, която стои зад разработването на насоките за качеството на въздуха (НКВ) на СЗО, включва строг и систематичен процес. Събират се експерти от цял свят, за да оценят най-новите научни изследвания и данни за въздействието на замърсяването на въздуха върху здравето. Те преглеждат наличните доказателства, оценяват тяхното качество и значимост и разработват препоръки за нивата на замърсителите, които са безопасни за човешкото здраве. Основен фокус при СЗО и препоръките са здравните фактори, за разлика от директивата на ЕС, която освен здравни, включва икономически и социални фактори. Директивата на ЕС се разграничава с това, че стандартите се насочват към регионални и национални нужди на страните членки, докато стандартите на СЗО са универсални и приложими глобално из целия свят.

В двете методологии се вземат предвид различните видове замърсители, като например ФПЧ, озон, азотен диоксид и серен диоксид, както и

Таблица 2.1: Замърсяване на въздуха: препоръки на СЗО и законодателство на ЕС (в $\mu g/m^3$)

Норми за замърсяване на въздуха	СЗО		ЕС	
	ФПЧ10	ФПЧ2,5	ФПЧ10	ФПЧ2,5
среднодневна	50	25	50	–
средногодишна	20	10	40	25

техните източници, като например промишлени емисии, транспорт и природни източници. Насоките дават препоръки за нивата на замърсителите, които са безопасни за дишане от хората през различни периоди от време, като например средни стойности за деня или годината. Насоките на СЗО за качеството на въздуха осигуряват рамка за политиците и лицата, вземащи решения, за разработване и прилагане на мерки за подобряване на качеството на въздуха и опазване на човешкото здраве.

2.1.3 Контрол на качеството на въздуха - разлики в нормите на СЗО и европейското законодателството

В сравнителния анализ между законодателството на ЕС и Указанията на СЗО относно нивата на ФПЧ във въздуха се извяват няколко ключови разлики (Таблица 2.1). Докато ЕС и СЗО и двете препоръчват дневен среден концентрационен праг за ФПЧ10 от $50 \mu g/m^3$, ЕС и СЗО се различават значително в годишните си средни препоръки, като стандартите на СЗО са двойно по-строги (20 $\mu g/m^3$, срещу 40 $\mu g/m^3$ от ЕС). За ФПЧ2,5, ЕС нямат концентрационен праг за средендневни норми и има значителни разлики в годишните препоръки, като СЗО настоява за по-строги лимити. Освен това стандартите на ЕС са специфични за региона, докато указанията на СЗО са необвързващи и универсални, предоставяйки насоки на страни по целия свят.

За целите на тази дисертация ще използваме насоките за качество на въздуха на СЗО, тъй като те са препоръчителни в световен мащаб, имат ограничение за дневни норми на ФПЧ2,5 и се базират на чисто здравни аспекти, в сравнение с директивата на ЕС, която е с фокус към регионални и национални нужди и включва политико-икономически и социални аспекти

в директивите.

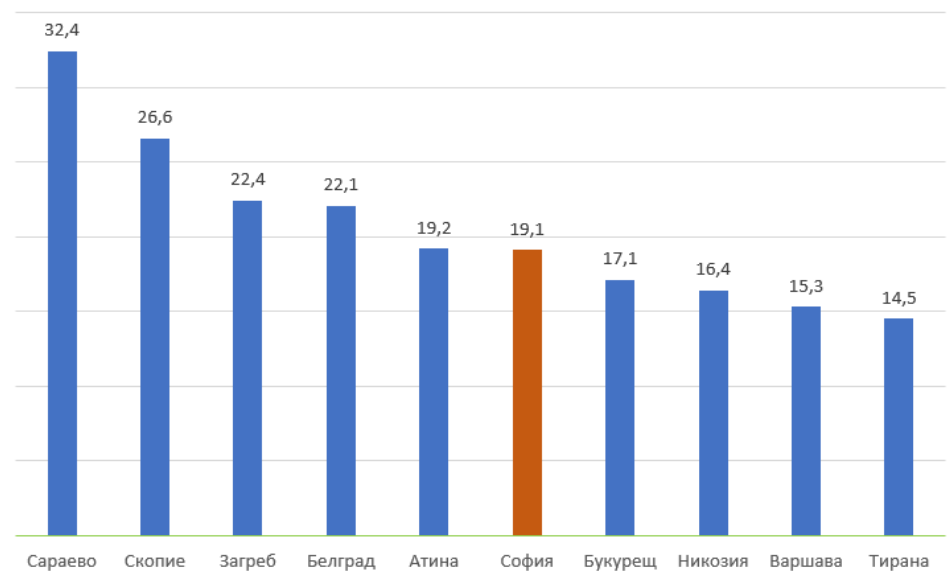
В последния си доклад СЗО полага специално внимание на замърсяването от ФПЧ и смята излагането на ФПЧ 2,5 за най-големия убиец свързан с въздуха в световен план. Около 80 % от смъртните случаи в световен мащаб, дължащи се на излагане на ФПЧ2,5, могат да бъдат предотвратени, ако страните спазват годишна прагова стойност за ФПЧ2,5 от $5 \mu\text{g}/\text{m}^3$. Постигането на междинните цели също ще има значителни ползи за здравето. В случая с ФПЧ2,5 постигането на междинна цел 4 ($10 \mu\text{g}/\text{m}^3$) би довело до приблизително 48% намаление на общия брой смъртни случаи, дължащи се на експозиция на ФПЧ2,5.

2.2 Географска зона на изследване

София е разположена в котловина и се характеризира с високо количество антропогенни емисии и честа поява на застойни метеорологични условия. Градът е с население от 1,2 милиона души [65] и е разположен в Софийското котловинно поле. Районът е признат за проблемно място, където особено през зимата има многобройни превишения на нормите в европейското законодателство за контрол на замърсяването на въздуха.

От години София се бори със значителни нива на замърсяване на въздуха, като ФПЧ постоянно надвишават препоръчителните граници, определени от СЗО и ЕС. Тези повишени нива на ФПЧ във въздуха на града представляват сериозна заплаха за здравето и благосъстоянието на неговите жители. На Фигура 2.2 е показана средна концентрация на ФПЧ2,5 в най-замърсените столици в Европа през 2022 г. На нея се вижда, че София заема една от призовите позиции в тази класация. Също важен е фактът, че данните за измерване на качеството на въздуха са недостатъчни за София. За нивата на ФПЧ 2.5 в столицата се подават данни от една единствена станция - АИС Хиподрума, което е недостатъчно за качествени статистически анализи за град с размерите и населението на София.

Комбинацията от студени зими с минусови температури и разположението на града, предразполагат към температурни инверсии, които продължават от няколко дни до седмица. Много градски райони, разположени в долина без подходящ въздухообмен, срещат значителни проблеми със замърсяването на въздуха, които са свързани с местните атмосферни особености [78]. Силна инверсия и слабо наличие на валежи и/или вятър са



Фигура 2.1: Средна концентрация на ФПЧ_{2,5} в най-замърсените столици в Европа през 2022 г. (в микрограма на кубичен метър въздух)

основните причини за задържане на замърсители във въздуха предимно през зимата. Исторически показателите на качеството на въздуха в София през зимата са значително по-лоши от тези през лятото. Освен температурните инверсии, които са типични при по-студени условия, по време на отоплителния сезон, изгарянето на изкопаеми горива за битово отопление като дърва и въглища, но понякога и боклук, повишават значително нивата на ФПЧ.

Следователно тази разлика между качеството на въздуха в летни и зимни условия ще бъде добър модел за тестване на нашата хипотеза за значителна връзка между замърсяването и последиците за здравето в София. Литературата изобилства с модели, които могат да се използват за оценка на хипотезата за замърсяване/здраве. Конкретно за София не са правени подобни проучвания. Корелация между температурата и смъртността около София се прави с помощта на линейни и нелинейни модели [70]. Други изследвания, които изследват генотоксичността на атмосферния въздух в 3 европейски града, включително София, показват, че замърсяването на

въздуха през зимата е от 6 до 10 пъти по-високо в сравнение с летния въздух [34].

2.3 Методология на изследването

До колкото е известно това е първото подобно изследване с реални данни от официални и граждански източници на информация за замърсяването на въздуха с прахови частици, изчерпателни данни от дейността на центъра за спешна медицинска помощ (бърза помощ), данни от две от най-големите болници, свързани с достъпа до спешна помощ в София - УМБАЛСМ Пирогов и УМБАЛ Токуда. Резултатите могат да бъдат сравнени и контрастирани с други международни проучвания с местни данни.

Целта на това изследване е да се установи коя остра заболеваемост и с колко се увеличава в дните, когато нивата на ФПЧ не отговарят на насоките на СЗО, като се използва анализ на времеви серии.

Решени са следните четири подзадачи:

- Сравняване на данните за качеството на въздуха със здравните данни, получени от болници и линейки;
- Сравнителен анализ и подчертаване на ключовите проблемни области;
- Тестване на хипотезата, че данните от евтини сензори могат да бъдат полезни при такива епидемиологични изследвания;
- Идентифициране на бъдещи перспективи и обобщение на ключовите области, в които са необходими допълнителни изследвания за подобряване на ефективността на модела.

Данни за замърсяването на въздуха

Първо ще открием кои дни, в рамките на наблюдавания период, концентрациите на ФПЧ надвишават средно дневните лимити на СЗО. След това ще сравним здравните данни за дните, в които качеството на въздуха е в рамките на указанията на СЗО, с дните, които не са.

Данните за качеството на въздуха под формата на средночасови концентрации на ФПЧ_{2,5} и ФПЧ₁₀ са за 2018 г. и 2019 г. и представляват

независими серии за всеки елемент от официални сензори, предоставени от (а) Изпълнителната агенция по околна среда (ИАОС) и (б) лазерни станции от мрежата luftdaten (глобална сензорна мрежа, управлявана от доброволци, която създава отворени данни за околната среда).

Данните се събират от 5 официални наблюдателни площадки (Дружба, Надежда, Хиподрума, Павлово и Младост). За измерване на ФПЧ10 и ФПЧ2,5 се използва метод за бета атенюация съгласно Европейската директива 2008/50/ЕС (Директива, 2008). Според предишни проучвания използването на данни за качеството на въздуха от евтини лазерни станции има за цел да увеличи потенциалните ползи от традиционните мрежи за наблюдение с допълнителна географска и времева резолюция на измерване [12]. Освен това евтините станции могат да надградят официалните измервателни уреди и по-специално ограничено количество данни и разпространение на станции [106]. Смята се, че евтините сензори са подходящи за много специфични цели, включително разширяване на разговорите с общностите и общата гражданска осведоменост по проблема на ВЗ [60]. В София към момента на изследването има няколко граждански мрежи от лазерни станции, които надграждат, с допълнителна географска и времева разделителна способност на измерване, официалните станции на ИАОС.

Здравни данни

Поради доброволния характер на предоставянето на данни има различни времеви характеристики и формати от различни източници:

- Обобщени данни за дейността на Центъра за спешна медицинска помощ София по диагнози на центъра - от 01.01.2017 г. до 14.03.2019 г.;
- Информация за дейността по диагностика от ДКЦ Токуда - 02.01.2018 г. до 31.12.2018 г.;
- Информация за дейността по диагностика от УМБАЛСМ Пирогов - 01.01.2018 г. до 31.05.2019 г.;
- Информация за дейността на хоспитализираните диагнози от Пирогов - 01.01.2018 г. до 31.12.2018 г.

Изследването използва Международната класификация на болестите (ICD) и по-конкретно ICD-10 за сегментиране на заболяванията и идентифицираната заболеваемост. ICD-10 е 10-та ревизия на ICD, медицински класификационен списък от Световната здравна организация (СЗО), който се използва до момента на провеждане на това проучване. Той съдържа кодове за заболявания, признаци и симптоми, необичайни находки, оплаквания, социални обстоятелства и външни причини за нараняване или заболявания [66],[67].

2.3.1 Методи

В това изследване се използва анализ на времеви редове с корелационни методи за анализ на качеството на въздуха и данните за здравето. Статистическите методи в изследването попадат в две категоризации: параметрични и непараметрични. Параметричните методи, като корелационен коефициент на Пийърсън, се предпочитат, когато данните са нормално разпределени или има достатъчно голям обем на измерванията, което позволява да се приеме, че данните следват нормално разпределение. Тези методи са подходящи за оценка на линейната връзка между две непрекъснати променливи. Параметричният корелационен тест на Пийърсън (2.1) се използва за сравняване на двата източника на данни за качеството на въздуха. Той осигурява мярка за линейната връзка между двете непрекъснати променливи (обикновено наричана само коефициент на корелация). Коефициентите на корелация за всяка двойка (x, y) се определят за извършване на оценката и стойностите на x и y съответно се заменят с техните рангове. Резултатите от теста са приложени към коефициент на корелация в диапазона от -1 до 1 .

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2.1)$$

Непараметричните сравнителни методи, като Spearman's rho и U тест на Mann – Whitney, се използват, когато данните не са нормално разпределени или не е уверен, че може да се приеме, че следват нормално разпределение. Тези методи са подходящи за оценка на ранговата връзка между променливи или за сравнение на разликите между две групи, без да се изисква специфично оценяване на линейната връзка. Изборът в тази дисертация

е следният: дескриптивен анализ за индивидуалните вариационни редове; корелационен анализ (непараметричен - Spearman's rho) между двойки променливи; междугрупово сравнение (непараметрично Mann-Whitney U test) между средни стойности.

Моделът, който е разработен в тази задача, е за взаимовръзка на показатели за качество на въздуха и здраве въз основа на данни за фини прахови частици и регистрирани постъпления в болници и Бърза помощ. По съвет на лекари, професионалисти за анализ на здравните данни, в това изследване използваме отместване със забавяне от 1, 2 и 3 дни. Като причината за това е, че след експозицията на ВЗ се изисква време, за да развият ефекти върху здравето [46]. По този начин изследването на приемането в болница и записите от Бърза помощ ще покажат не само кои случаи заболявания търпят промяна поради изменения на качеството на въздуха, но можем да видим и как те се развиват във времето. На Фигура ?? е илюстриран методът експозиция-ефект, където дозата е експозицията, умножена по времето. Някои ефекти върху здравето се нуждаят от повече време, за да проявят симптоми, в сравнение с други.

2.4 Резултати

В едно скорошно проучване се предполага, че евтините сензори могат да предложат груби детайли за установеното качество на въздуха, но не са готови за приложения, включващи висока точност [44]. В тази глава се проверява хипотезата, че евтините сензори за качество на въздуха на Luftdaten могат да бъдат добро допълнение към официалните станции за качество на въздуха на ИАОС. Ако изчистим и обобщим данните, можем да имаме корелация, която ще бъде полезна за разглеждане на някои груби детайли. При сравнението сме изчистили данните, когато влажността на въздуха е била над 70%. Тези условия не означават непременно, че данните са неверни, но при такива условия производителят не гарантира прогнозираната точност от 10%.

Резултатите от дневни концентрации за ФПЧ10 и ФПЧ2,5 са обобщени в таблица 2.2. Средната дневна концентрация на ФПЧ 2,5 е 11,7 g/m³ с диапазон от 2 до 136 g/m³. През периода на изследването 13,2% от дневните концентрации на ФПЧ10 или ФПЧ2,5 не са отговаряли на нормите на СЗО за качество на въздуха (50 g/m³ за ФПЧ10 и 25 g/m³ за ФПЧ2,5). В дан-

ните от ИАОС имаше липси за няколко часа до няколко дни (проблеми със хранване, повреди, профилактики), а някои от източниците за болнични прегледи бяха с липсващи данни през почивните дни (събота, неделя, официални празници), поради което беше предприет анализ на пропуските, за да се изключи възможността за отклонение.

Таблица 2.2: Дневната концентрация на ФПЧ10 и ФПЧ2,5, разделена на работни и неработни дни

Ден от седмицата	N	Средно	Медиан	Мин	Макс	Персентили
ФПЧ10 Работни дни	522	34.17	27.5	5	336	20
ФПЧ10 Неработни	230	37.1	26	6	290	18
ФПЧ2,5 Работни дни	522	11.15	9	2	104	4
ФПЧ2,5 Неработни	230	12.93	8	2	113	3

В данните за хоспитализирани пациенти с по-тежко състояние наблюдаваме следните корелации (Таблицы 2.3 и 2.4). Констатира се увеличение (спрямо фоновите нива) на респираторни и белодробни заболявания от 1-ви до 3-ти ден след превишение на ФПЧ10 със 120%, както и сърдечна недостатъчност в същия период от време с 19%. Превишението на ФПЧ2,5 се свързва с 59% увеличение на честотата на белодробна емболия на дни 2 и 3.

Резултатите показват увеличение на контактите с Бърза помощ с 11% в дни с превишаване на ФПЧ10 и 13,5% в дни с превишение на ФПЧ2,5, за период от най-малко 3 дни. Средно увеличение (спрямо фоновите нива) на неврозите на ден 2 след превишаване на ФПЧ10 е 1%, при инфаркти на ден 3 след превишаване е 8%, при инсулти веднага след превишаване с 9%, продължаващи до ден 3 след превишение. Същото важи и за хипертоничните състояния, които нарастват с 5% непосредствено около превишението и до третия ден след събитието.

Таблица 2.3: Сравнение на непараметрични данни от хоспитализирани пациенти в дни с превишения на нормите на СЗО за ФПЧ10

Забол.	ICD-10	без отм.		отм. 1 д.		отм. 2 д.		отм. 3 д.	
		Z	p	Z	p	Z	p	Z	p
Респ. сист.	C30-C39	-1,504	0,133	-2,211	0,027	-2,870	0,004	-2,862	0,004
Сърд. удар	I50	-1,729	0,084	-3,656	0,001	-3,475	0,001	-2,821	0,005

Таблица 2.4: Сравнение на непараметрични данни от хоспитализирани пациенти в тежки състояния в дни с превишения на нормите на СЗО за ФПЧ2,5

Забол.	ICD-10	без отм.		отм. 1 ден		отм. 2 дни		отм. 3 дни	
		Z	p	Z	p	Z	p	Z	p
Респ. сист.	C30-C39	-0,361	0,718	-1,270	0,204	-1,265	0,206	-2,119	0,034
Бел. емб.	I26	-1,601	0,109	-2,302	0,021	-2,894	0,004	-1,852	0,064
Сърд. удар	I50	-1,586	0,113	-2,831	0,005	-3,195	0,001	-2,799	0,005

От анализираните данни се забелязва значителна корелацията на сърдечносъдовите и мозъчносъдовите заболявания, като напълно съвпада с бързото нарастване на риска при ниски нива на фракции на замърсяване с ФПЧ2,5, съвпадаща с подобни открития в други подобни проучвания [99], [84], [92].

Но най-висока корелация се забелязва при острите инфекции на горните дихателни пътища се увеличават съответно с 47% и в частност пневмонията с 60%. Хроничната обструктивна белодробна болест (ХОББ) след ден е с 36% повече. Що се отнася до астмата, записите за алергична астма нарастват повече в дни с повишено замърсяване в сравнение с неалергичната астма. Амбулаторните прегледи в болниците за леки състояния, които не изискват хоспитализация, логично откриваме повишено присъствие на респираторни диагнози (таблици 2.5 и 2.6).

2.5 Заключение

В резултат на проведените анализи и тестове се доказва, че острата заболяемост се увеличава в дните, когато концентрациите на фините прахови частици са над нормата на СЗО за качество на въздуха. Това се установи сравнявайки данни за постъпления в болници и случаите на Бърза помощ в дни със и без превишенията на нормите за ФПЧ в София.

Респираторните заболявания показват по-високо и по-бързо развитие, когато замърсяването на въздуха надхвърли здравословните граници. Други заболявания, свързани със замърсяването на въздуха, като сърдечносъдови и мозъчно-съдови, също се увеличават.

Това е сред първите подобни проучвания в България и може да бъде

Таблица 2.5: Сравнение на непараметрични данни от амбулаторни прегледи в болници с нормите на СЗО за среднодневна концентрация на ФПЧ10 без отместване и с 1 ден отместване

Забол.	ICD-10	без отм.		отм. 1 ден	
		Z	p	Z	p
Дихателна система	C30-C39	-1.515	0.130	-2.950	0.003
Ушни забол.	H65.0, H66.0, H81	-0.754	0.451	-0.976	0.329
Миокарден инфаркт	I20-I25	-0.244	0.807	-1.393	0.164
Сърдечна недостатъчност	I50	-0.771	0.440	-2.840	0.005
Горни дихателни пътища	J00-J06	-3.218	0.001	-3.507	0.000
Бактериална пневмония	J13-J18	-3.171	0.002	-3.395	0.001
Остър бронхит	J20-J21	-3.096	0.002	-2.848	0.004
Долни дихателни пътища	J40-J47	-3.988	0.000	-4.049	0.000
Инфекции на кожата	L00-L08	-2.552	0.011	-1.379	0.168
Микоплазмена пневмония	J20.0	-2.330	0.020	-2.327	0.020
Остър бронхит, неуточнен	J20.9	-2.496	0.013	-1.449	0.147
Астма, алерг.	J45.0	-2.426	0.015	-2.808	0.005
Ангина	I20.8	-1.971	0.049	-0.763	0.445

Таблица 2.6: Сравнение на непараметрични данни от амбулаторни прегледи в болници с нормите на СЗО за среднодневна концентрация на ФПЧ10 с 2 и 3 дни отместване

Забол.	ICD-10	отм. 2 дни		отм. 3 дни	
		Z	p	Z	p
Дихателна система	C30-C39	-2.941	0.003	-2.423	0.015
Ушни забол.	H65.0, H66.0, H81	-1.580	0.114	-2.322	0.020
Миокарден инфаркт	I20-I25	-1.876	0.061	-2.104	0.035
Сърдечна недостатъчност	I50	-2.559	0.010	-2.972	0.003
Горни дихателни пътища	J00-J06	-3.910	0.000	-4.605	0.000
Бактериална пневмония	J13-J18	-2.556	0.011	-3.683	0.000
Остър бронхит	J20-J21	-1.606	0.108	-1.588	0.112
Долни дихателни пътища	J40-J47	-2.842	0.004	-3.481	0.001
Инфекции на кожата	L00-L08	-0.267	0.789	-0.489	0.625
Микоплазмена пневмония	J20.0	-0.430	0.667	-0.431	0.667
Остър бронхит, неуточнен	J20.9	-0.039	0.969	-1.292	0.197
Астма, алерг.	J45.0	-3.577	0.000	-3.643	0.000
Ангина	I20.8	-0.773	0.440	-0.674	0.500

полезно за специалистите в сферата на здравеопазването, учените по околната среда и политиците. Освен подобряване на качеството на въздуха, политиците трябва да се съсредоточат върху механизмите за информиране на гражданите за предстоящи дни с високо замърсяване на въздуха и поддържане на медицинско наблюдение на хората от рисковите групи.

Методът на ефекта на експозиция е сложен и прецизните измервания са много трудни за постигане дори в експозиционна камера. С помощта на методите на епидемиологичните изследвания и статистиката на населението обаче става ясно, че замърсяването на въздуха допринася за очакваните промени в някои показатели на заболяванията. В заключение, подобно на данните от публикувани проучвания в градове в Европа, Америка и Азия, повишените нива на замърсяване на въздуха се свързват и водят до високи нива на диагностицирани заболявания.

С малки изключения, краткосрочните ефекти върху здравето се измерват чрез използване на осреднени концентрации на замърсяване на въздуха в целия град за показатели за експозиция. Това може да доведе до погрешна класификация на експозицията и по този начин до отклонение [51].

Анализът на данни за качеството на въздуха показва, че използването на по-обширни данни от мрежи с евтини лазерни сензори, обогатява информацията, предоставяна от високо-точните официални станции на ИАОС. Освен, че станциите на ИАОС са малко на брой, те често имат и липсващи данни за дадени периоди. Допълването на данните от официални станции с данни от гражданските мрежи, води до подобрени измервания на връзките между здравето и замърсяването на въздуха, като осигурява по-обстойна експозиция конкретно на ФПЧ2,5.

Глава 3

Калибриране на граждански сензори чрез машинно обучение

Резултатите от тази глава са публикувани в [101].

В предишната глава разгледахме важността на включването на допълнителни данни в рамките на епидемиологични проучвания, свързани с въздушното замърсяване. Подчертахме ключовата роля, която играят данните от мрежите с лазерни сензори за ФПЧ в този контекст, особено когато стандартните инструменти за измерване се оказват недостатъчни. Въпреки това, следва да се обърне внимание на общия недостатък на станциите с лазерни сензори - техните данни не достигат до същата степен на точност като тези от стандартните измервателни станции, особено при високи нива на влажност във въздуха.

За да разберем как можем да подобрим данните, които получаваме от станциите с лазерни сензори за ФПЧ, ще трябва да измерим тяхната грешка в реални условия. За целта поставяме станции с лазерен сензор в непосредствена близост до станциите на МОСВ за измерване на ФПЧ. Сравняваме техните данни и правим калибровъчен модел на база машинно обучение, който подобрява данните от евтините лазерни станции.

Традиционните станции за наблюдение на качеството на въздуха, оборудвани със сложни и скъпи инструменти, са основният източник на данни за оценка на нивата на замърсяване на въздуха. Високата им цена, ограниченото им пространствено покритие и логистичните предизвикателства обаче доведоха до появата на евтини сензори за качество на въздуха като алтернативно решение за наблюдение. Тези сензори предлагат потенциал

за разширяване на мрежите за наблюдение на качеството на въздуха, като предоставят данни с по-висока разделителна способност и позволяват по-локални оценки.

Евтините лазерни сензори в последните години получават популярност, поради тяхната достъпност, преносимост и лекота на използване. Тези сензори работят на принципа на разсейване на светлината, като откриват и определят количествено концентрацията на ФПЧ във въздуха. Те са много по-рентабилни от традиционните станции за измерване, но им липсва надеждност на измерваните данни, поради чувствителността им към температурата и относителната влажност. Тези слабости са особено очевидни, когато става въпрос за високи нива на ФПЧ, в комбинация с висока относителна влажност - тогава стойностите измерени за ФПЧ са неестествено завишени. В заключение, въпреки че лазерните сензорни станции предлагат обещаващи предимства по отношение на цената и достъпността, тяхната точност и надеждност са обект на проверка.

Калибрирането е от съществено значение за осигуряване на точността и последователността на евтините лазерни сензори. В тази глава представяме нов двустепенен модел за калибриране на евтини лазерни сензори за наблюдение на въздуха, като използваме техники за ML с наблюдение и без наблюдение. Предложеният от нас метод за калибриране има за цел да подобри точността на отделните лазерни сензори, като използва данни от официални станции за наблюдение на качеството на въздуха като референтни.

Първата стъпка от нашия модел за калибриране включва контролирана ML, при която обучаваме прогнозен модел, използвайки данни, събрани както от евтини лазерни сензори, така и от референтни станции за наблюдение на качеството на въздуха. Алгоритъмът за контролиран ML научава връзката между показанията на сензорите и съответните референтни измервания, което позволява калибриране на данните от сензорите. Като използваме изчерпателните и точни данни от референтните станции, можем да преодолеем разликата между показанията на евтините сензори и истинските измервания.

Втората стъпка на нашия модел използва неконтролирани ML техники за по-нататъшно усъвършенстване на процеса на калибриране. Алгоритмите за неконтролирано обучение, като например групиране и откриване на аномалии, помагат да се идентифицират и коригират отклоненията и несъответствията в данните от сензорите. Като откриваме и отстраняваме

нередностите, можем да повишим общата точност и надеждност на евтините лазерни сензори.

Моделът за калибриране, който предлагаме, има няколко предимства пред традиционните подходи за калибриране. Първо, той използва възможностите на алгоритмите за ML, за изучаване и моделиране на сложни връзки между показанията на евтини сензори и референтни измервания. Този подход, основан на данни, позволява по-точно калибриране, като намалява присъщите отклонения и грешки, свързани с методите за ръчно калибриране.

Второ, нашият модел за калибриране осигурява икономически ефективно решение, като използва съществуващите станции за наблюдение на качеството на въздуха като референтни точки. Вместо да разчитаме единствено на скъпо оборудване за калибриране, можем да се възползваме от богатството от данни, събрани от установените мрежи за наблюдение. Това не само намалява финансовата тежест на калибрирането, но и осигурява последователна и стандартизирана референтна точка за оценка на точността.

Освен това предложеният двустъпков модел подобрява времевата и пространствената разделителна способност на данните за замърсяването на въздуха. Чрез интегрирането на евтини лазерни сензори в мрежата за наблюдение можем да получим данни от по-широк кръг от места и в реално време. Този увеличен обхват подобрява разбирането ни за моделите на замърсяване на въздуха, като улеснява целенасочените интервенции и политическите решения.

В обобщение, целта на тази глава е да се разработи нов двустъпков модел за калибриране на евтини лазерни сензори за наблюдение на въздуха. Чрез комбиниране на контролирани и неконтролирани техники за машинно обучение, ние целим да повишим точността и надеждността на тези сензори, превръщайки ги в ценен инструмент за оценка на качеството на въздуха. Използването на официални станции за наблюдение на качеството на въздуха като референтна база гарантира надеждността на модела за калибриране и съвместимостта му със съществуващата инфраструктура за наблюдение. В крайна сметка този метод за калибриране допринася за напредъка на евтиното наблюдение на качеството на въздуха и дава възможност за по-всеобхватни и локални оценки на нивата на замърсяване на въздуха.

След съкратено представяне на основите на наблюдението на качество-

то на въздуха в раздел 3.1, включително кратко описание на стандартните инструменти и евтини сензори, които ще бъдат използвани в анализа (в подраздел 3.1.6) , се обсъжда как метеорологичните условия като температура, влажност и атмосферно налягане влияят на измерванията и как трябва да се вземат предвид преди изграждането на модела. Двустъпковият модел е описан и представен в раздел 3.2, където първо, ние тестваме и прецизираме 5 контролирани ML метода, използвайки 5 сензора, разположени точно до стандартните инструменти. След това прилагаме същите методи и други 13 евтини сензора в безжичната сензорна мрежа (WSN) (в раздел 3.3). В раздел 4.4 са представени и обсъдени резултатите от модела за калибриране. И накрая, заключението е дадено в раздел 3.5.

3.1 Наблюдение на качеството на въздуха

Най-честият метод за наблюдение на качеството на въздуха е използването на стандартизирани станции за наблюдение на околната среда, които са точни, но и скъпи за закупуване и експлоатация. В резултат на това станциите за наблюдение на околната среда често са разпръснати на рядко, което води до наблюдения с ниска географска резолюция. Последното десетилетие са разработени различни видове евтини лазерни сензори за наблюдение на качество на въздуха като опция, която може да увеличи детайлността и надежността на наблюдението. Използването на евтини сензори за качество на въздуха обаче има редица недостатъци: те се влияят от кръстосана чувствителност между различни замърсители на околната среда, както и от външни променливи като трафик, колебания на времето и човешка дейност, и тяхната точност намалява с времето.

В тази секция ще разгледаме някои особености на наблюдението на качеството на въздуха, които по-късно ще бъдат взети предвид при проектирането на модела за калибриране.

3.1.1 Системи за наблюдение качеството на въздуха

ФПЧ са сложна смес от твърди и течни частици, суспендирани във въздуха, които могат да причинят редица здравословни проблеми, особено при уязвими групи като деца, възрастни хора и хора с респираторни или сърдечносъдови заболявания. Поради това наблюдението на концентрацията

на ФПЧ е важна част от наблюдението на качеството на въздуха. В България и в ЕС ФПЧ се измерват с различни методи и инструменти, включително с метода на бета-абсорбация. Тук ще обсъдим как се измерват ФПЧ с помощта на тези стандартни станции за наблюдение.

Станциите за наблюдение използват няколко метода за измерване на концентрациите на ФПЧ, включително:

- Гравиметрично вземане на проби: Гравиметричното вземане на проби е най-широко използваният метод за измерване на концентрацията на ФПЧ. Той включва събиране на ФПЧ върху филтър и претеглянето им, за да се определи масата на ФПЧ на единица обем въздух. Филтърът се поставя в пробовземач, който вкарва въздух през филтъра с постоянен дебит за определен период от време, обикновено 24 часа. След вземането на проби филтърът се изважда внимателно от пробовземача и се претегля с помощта на микробаланс. Разликата в теглото преди и след вземането на пробата дава масата на ФПЧ, събрани върху филтъра;
- Оптично разсейване: Оптичното разсейване е техника, която измерва свойствата на разсейване на светлината на ФПЧ във въздуха. Този метод включва светлинен източник, който излъчва светлинен лъч през проба от въздух. Светлината се разсейва от ФПЧ във въздуха и количеството разсеяна светлина се измерва от детектор. Количеството разсеяна светлина е пропорционално на концентрацията на ФПЧ във въздуха;
- Бета абсорбация: За измерване на концентрацията на ФПЧ във въздуха може да се използва и методът на бета-анализа. Този метод включва събиране на проби от въздуха върху филтър, както при гравиметричното вземане на проби, но вместо да се претегля филтърът, той се анализира с помощта на бета-детектор. Бета детекторът измерва броя на бета частиците, излъчени от радиоактивните изотопи във филтъра. След това концентрацията на ФПЧ във въздуха се изчислява въз основа на броя на откритите бета частици;
- Микробаланс: Микробалансът е високочувствително тегловно устройство, което може да измерва масата на ФПЧ с висока точност. Микровезни се използват за претегляне на филтри, които са събрали

проби от ФПЧ по време на гравиметричното вземане на проби. Тези везни могат да претеглят с точност до микрограм, което позволява точно измерване на концентрацията на ФПЧ.

3.1.2 Сертифициране на станции и анализ на предимства и недостатъци

В ЕС има законодателство и сертифициране на станции и инструменти за наблюдение на качеството на въздуха. Европейският съюз е определил стандарти за качество на въздуха за ФПЧ и други замърсители, които трябва да бъдат спазвани от държавите членки. Директивата за качеството на въздуха на Европейския съюз определя пределно допустимите и целевите стойности за ФПЧ10 и ФПЧ2,5, а държавите членки трябва да гарантират спазването на стандартите за качество на въздуха чрез наблюдение и мерки за контрол. Директивата определя и изискванията към станциите за наблюдение на качеството на въздуха, включително необходимостта от стандартизирани и валидирани методи за измерване.

Сертифицирането на станциите и инструментите за наблюдение на качеството на въздуха е от съществено значение за гарантиране на точността и надеждността на измерванията. Европейският съюз е създал Европейската референтна лаборатория за замърсяване на въздуха (ERLAP), която отговаря за валидирането на методите за измерване на замърсителите на въздуха, включително на ФПЧ. ERLAP осигурява сертифициране на инструментите и системите за наблюдение на качеството на въздуха, като гарантира, че те отговарят на необходимите стандарти и са подходящи за използване в програмите за наблюдение на качеството на въздуха.

Използването на бета-анализ и други утвърдени методи за измерване на концентрацията на ФПЧ има няколко предимства пред евтините лазерни сензори. Едно от основните предимства на утвърдените методи е тяхната точност и прецизност. Гравиметричните методи за вземане на проби и методите за бета-анализиране са широко валидирани и стандартизирани, а тяхната точност и прецизност са установени в резултат на дългогодишни изследвания и опит. От друга страна, евтините лазерни сензори може да не осигурят същото ниво на точност и прецизност, особено при трудни условия на околната среда, като висока влажност или екстремни температури.

Друго предимство на утвърдените методи е способността им да измерват различни размери на частиците. Гравиметричните методи за вземане

на проби и методите за бета-анализ могат да измерват ФПЧ10, ФПЧ2,5 и други размери, като по този начин осигуряват по-цялостна картина на замърсяването с ФПЧ. От друга страна, евтините лазерни сензори обикновено са проектирани да измерват само един размер на ФПЧ, обикновено ФПЧ2,5 или ФПЧ10, което ограничава полезността им при оценката на качеството на въздуха.

Установените методи обаче имат някои недостатъци, включващи цената и сложността на апаратурата и необходимостта от обучен персонал за работа и поддръжка на оборудването. Цената на уредите за бета-анализ и други утвърдени методи обикновено е по-висока от тази на евтините лазерни сензори, което ги прави по-малко достъпни за по-малките организации и обществени групи. Освен това утвърдените методи може да изискват по-обширна инфраструктура, включително захранване и предаване на данни, както и необходимост от редовно калибриране и поддръжка, за да се гарантират точни измервания.

Високите разходи за оборудване и обслужване обаче правят невъзможни [61], изпълнението на изискванията за наблюдение с висока разделителна способност и оценката на степента на лично облъчване [14]. Тъй като необходимостта от по-кондензирано наблюдение постепенно нараства, евтините сензори за качество на въздуха се използват широко в наблюдението на въздуха през последните години поради предимствата на ниската цена, ниската консумация на енергия, бързата работа и бързата реакция [93] .

3.1.3 Възможности и недостатъци на безжичните евтини станции

Безжичните сензорни мрежи свързват голям брой фиксирани сензори на множество места в една мрежа, което позволява дългосрочно наблюдение с висока разделителна способност на замърсители на въздуха [76]. Техните приложения често се наблюдават в проучвания, свързани със здравето, и при проследяване на индивидуалната експозиция. Една от първите научни разработки по темата експозиция на външно въздушно замърсяване - здраве е мрежа от сензори, изградена по маршрута на маратона в Хонг Конг през 2015 г., която се използва за измерване на здравния индекс на качеството на въздуха (AQHI) и определяне на индивидуалните нива на експозиция на спортистите [86]. В Рочестър, Ню Йорк, САЩ, проучване

използва мрежа от евтини лазерни сензори за измерване на ФПЧ във въздуха на различни места едновременно и непрекъснато, за да се определи времевата и пространствена вариация на ФПЧ, както и въздействието на трафика и изгарянето на дърва върху концентрациите на ФПЧ на открито [105]. Чрез инсталиране на безжични сензори на движещи се носители като превозни средства, мотоциклети и дронове, мобилната сензорна мрежа може да предостави по-компактни пространствени данни от фиксираните сензори и може да постигне стереоскопично проследяване на качеството на въздуха и източниците на емисии [58].

3.1.4 Влияние на влажност и височина при лазерни сензори

Въпреки широкото използване на евтиния лазерен сензор, прецизността на данните е оспорима [19, 74]. Някои държави са провели тестове за оценка и калибриране на сензори и са написани препоръки за употреба и оценка на сензори [85, 9].

Само няколко учени обаче досега са тествали и калибрирали евтини лазерни сензори и тяхното представяне при различни условия на околната среда и времеви мащаби все още не са известни [100]. Доказано е, че относителната влажност (RH) и разпределението на размера на частиците имат значителен ефект върху резултатите от наблюдението на сензорите [62]. Други изследвания са направени в мъгливи условия при висока относителна влажност (RH > 80%), проследената производителност на сензора е по-висока от тази на нормалния инструмент [44].

Повърхностните концентрации на замърсители се разреждат чрез конвективно смесване на въздушната маса, причинено от слънчевото нагряване на полето. Повишени струи замърсители (напр. от промишлен комин, логаритмичен диапазон на пренасяне на прах на по-високи нива) може да повишат концентрациите близо до повърхността, ако височината на смесващия слой (MLH) се увеличи над нивото на струите прах/дим.

Няколко проучвания установяват, че барометричното налягане е важно за моделиране на прахови частици, тъй като могат да възникнат сложни ветрови потоци, водещи до стагнация/стационарни условия с малка циркулация. Замърсителите се натрупват близо до земята в резултат на тези условия [64]. Когато барометричното налягане беше включено в модела, връзката между праховите частици и сърдечно-съдовата смъртност беше

незначително засилена [43]. Намирането на връзки между MLH и концентрациите на замърсители близо до повърхността, представителни за град като Берлин (равнинен терен), изглежда непрактично, особено когато емисиите от трафика са доминиращи [29].

3.1.5 Референтен инструмент

Това изследване използва пет станции за контрол на качеството на въздуха с традиционни методи за измерване като ръководство. За изчисляване на ФПЧ10 са използвани автоматичните станции за измерване на МОСВ, които използват гравиметрични методи за събиране на проби от ФПЧ върху филтър и претегляне на филтъра, за да се определи масата на събраните ФПЧ. Тези методи са подобни на метода на бета-анализа, но не разчитат на откриване на радиация. Гравиметричните методи са много точни, но изискват специализирано оборудване и обучен персонал за събиране и анализиране на пробите. Измервателните станции използвани в проучването се намират в София, съответно в районите Младост, Дружба, Надежда, Хиподрума и Красно село. Само една от тези станции измерва ФПЧ 2,5 и поради това ФПЧ 2,5 не се използва за справка в това изследване.

3.1.6 ФПЧ лазерни сензори

Сензорът за ФПЧ, използван в това изследване, е лазерният сензор за частици NovaFitness SDS011. Този тип сензори използват принципа на лазерна дифракция. Лазерът осветява уловените йони, когато улавя разпръснати светлинни вълни под определен ъгъл, докато въздухът преминава през фоточувствителната област на сензора. Създава се континуум на размера на частиците чрез класифициране на тези импулсни сигнали в различни интервали на размера на частиците, за да се измери масовата концентрация на частиците [49]. Таблица 3.1 изброява параметрите на SDS011. Тук е важно да се види, че максималната влажност за точни измервания е 70%. Това означава, че производителят гарантира точност до тази граница, но не означава непременно, че над 70% влажност измерванията не са точни.

Таблица 3.1: Характеристики и техническо описание на SDS011-TRF.

Съдържание	Параметър
Изходни данни:	ФПЧ10, ФПЧ2.5
Обхват на мониторинг:	0 - 999.9 $\mu\text{g}/\text{m}^3$
Захранващо напрежение:	15V (4.7 - 5.3V)
Консумация на енергия (работа):	70mA \pm 10mA
Температура на работа:	-10 до +50°C
Влажност (работа):	Макс. 70%
Точност:	70% за 0.3 μm 98% за 0.5 μm
Налягане:	300-1100 hPa

3.1.7 Използвана безжична мрежа

В това изследване е използвана безжичната сензорна мрежа (WSN) на Luftdaten. Тя е гражданска мрежа и се състои от 300 стационарни сензора, покриващи София. Всеки сензор е инсталиран в пластмасова тръба, която може да се монтира на стени, балкони, стълбове за улично осветление и други конструкции.

Разработени са определени насоки, за получаване на най-доброто представяне на емисиите на прахови частици в града с възможно най-малко сензори. WSN използва фиксирани сензори, които са монтирани в мрежа от 1 km, за да се гарантира, че по-голямата част от центъра на града е покрита с подходяща плътност.

Следващият раздел описва двустепенния модел на калибриране, който се предлага за подобряване на качеството на данните на евтините сензори, използвани в тази мрежа.

3.2 Модел за калибриране на данни

Този раздел описва двустепенния модел на калибриране за евтини лазерни сензори, който използва комбинация от контролирани и неконтролирани техники за машинно обучение (supervised and unsupervised ML). Този модел използва данни от измервания от евтини сензори и стандартни станции за наблюдение на въздуха като референтен инструмент.

Други изследователи доказват, че моделите с корекция на сензорни

данни чрез машинно обучение дават добри резултати, когато са обучени спрямо референтен инструмент [20].

3.2.1 Въведение в контролираното машинно обучение

Контролираното машинно обучение е вид изкуствен интелект (ИИ), при който се използват алгоритми за обучение от маркиран набор от данни и за изготвяне на прогнози или решения въз основа на това обучение. Целта на контролираното обучение е да се намери връзката между входните характеристики на набор от данни и съответните им резултати и да се използва тази връзка, за да се правят прогнози за нови, непознати данни.

Алгоритмите за контролирано обучение могат да бъдат широко категоризирани на алгоритми за регресия и класификация. Алгоритмите за регресия се използват за прогнозиране на непрекъснати числови стойности, като например цената на акция или температурата в даден град. Обичайните алгоритми за регресия включват линейна регресия, полиномна регресия и регресия с дърво на решенията. Алгоритмите за контролирано обучение се обучават върху маркиран набор от данни, който съдържа както входните характеристики, така и съответните изходи. Алгоритмите използват тези данни за обучение, за да научат връзката между входните признаци и изходите и след това използват тази връзка, за да правят прогнози за нови данни. Точността на прогнозите зависи от качеството на данните за обучение и от избора на алгоритъм.

Едно от основните предимства на обучението с наблюдение е способността му да се учи от големи количества данни и да прави прогнози с висока точност. Това е особено полезно в приложения, в които е налично голямо количество данни и ръчният анализ е невъзможен. Например алгоритмите за контролирано обучение се използват в медицинските изображения за автоматично откриване на ракови и други заболявания, както и във финансите за прогнозиране на цените на акциите и откриване на измамни сделки.

Въпреки многобройните си предимства, контролираното обучение има и някои ограничения, които трябва да се вземат предвид. Едно от най-големите ограничения е, че алгоритмите могат да правят прогнози само въз основа на връзките, които са научили от данните за обучение. Ако данните за обучение са предубедени или ограничени, алгоритмите може да не са в състояние да направят точни прогнози за нови данни. Това е

известно като свръхприспособление и може да доведе до лошо представяне при реални данни.

Друго ограничение е, че алгоритмите за обучение под наблюдение изискват маркирани данни, чието получаване може да отнеме време и да е скъпо. В много случаи етикетирването на данни изисква човешки експерти, а може да няма достатъчно експерти, които да етикетират данните своевременно и по рентабилен начин. За да се преодолеят тези ограничения, е важно внимателно да се подберат данните за обучение и алгоритъмът и непрекъснато да се наблюдава и оценява работата на алгоритъма. Това може да включва използване на техники като кръстосано валидиране и регулиране за намаляване на риска от прекомерно приспособяване и използване на алгоритми за обучение без надзор за идентифициране и отстраняване на отклоненията в данните за обучение. В заключение, контролираното машинно обучение е мощен инструмент за изготвяне на прогнози и вземане на решения въз основа на данни. То се използва широко в различни приложения - от медицински изображения и калибриране на данни до разпознаване на реч и обработка на естествен език.

3.2.2 Методология на изследването

Методологията, използвана за калибриране на концентрацията на ФПЧ10 в София, включва няколко стъпки:

1. Почистване на данните: Данните от гражданските станции се пречистват допълнително, чрез ограничаване на нивата на концентрация на ФПЧ10 въз основа на официалните почасови измервания. Станциите с по-малко от предварително зададена прагова стойност (90 дни) също се отстраняват от набора от данни;
2. Корекция на отклоненията: Изследването започва с коригиране на данните от измерванията на граждани, получени от luftdaten. Тези измервания се сравняват с официалните измервателни станции на Изпълнителната агенция по околна среда (ИАОС), за да се отстранят всички инструментални отклонения;
3. Подготовка на данните: Изчисляват се разстоянията между двойките станции и се създават групи от станции в рамките на определено разстояние. За всяка двойка се изчисляват измервания на различията

и станциите с високи измервания на различията се отстраняват от набора от данни;

4. Анализ на факторите и характеристиките: Извършва се проучвателен анализ на данните, като се използват официалните данни на ИА-ОС. Изследва се значението на метеорологичните променливи и инженерните характеристики, които влияят на нивата на замърсяване на въздуха;
5. Модел за калибриране: Изграден е двустъпков модел с машинно обучение, включващ допълнителни фактори от метеорологични параметри и топографски данни. Характеристиките, използвани за калибриране, включват температура, влажност, налягане, концентрация на PM10 от предишния ден, скорост на вятъра и различни фиктивни променливи. Повече информация за двустъпковия метод с машинно обучение ще бъде изложена в тази глава;
6. Оценка и валидиране: Точността на прогнозните модели се оценява с помощта на квадратична грешка. Моделите са сравнени помежду си, както и с първоначалните данни преди калибриране и показват значително подобрение на точността.

3.2.3 Данни и модел

Моделът използва 5 стандартни станции за наблюдение на въздуха, разположени в София и 5 евтини лазерни сензора със спецификациите, показани в предишния раздел. Малкият размер на евтините лазерни станции позволява възможност да бъдат поставени точно до официалните станции на ИАОС.

Входните променливи са относителна влажност (RH), атмосферно налягане (AP), температура (temp). Стойността, за която калибрираме данните, е ФПЧ10. Тези входни данни получаваме, както от стандартни инструменти, така и от евтини сензори. Тъй като времевата разделителна способност на сензора се различава от тази на обикновения инструмент, за изчисляване и оценка е използвана средната стойност за час.

След като температурата, въздушното налягане и влажността влияят върху стойностите на много сензори за качество на въздуха, измерванията

на тези променливи често се записват на място и използват в коригиращия модел, както е направено и от други изследователи [41], [107].

3.2.4 Кратко описание на петте използвани ML техники

Тук описваме на кратко петте контролирани техники за машинно обучение, използвани в първата стъпка на модела за калибриране. Тези техники са: линейна регресия (LR), дърво на решенията (DT), дърво на решенията с градиентен тласък (GBDT), случайна гора (RF) и изкуствена невронна мрежа (ANN).

Линейната регресия е прост и широко използван алгоритъм за контролирано обучение, който моделира връзката между зависима променлива и една или повече независими променливи. Дървото на решенията е модел, който използва дървовидна структура на решенията и техните потенциални последици, включително последиците от случайни събития [79]. GBDT е итеративен алгоритъм за дървета на решенията, съставен от множество дървета на решенията, при които решенията на всички дървета се сумират, за да се получи крайният отговор [50]. RF е хибридно дърво за прогнозиране, при което всяко дърво се основава на стойностите на случайна променлива, избрана независимо и с едно и също разпределение за всички дървета в гората [10]. Изкуствените невронни мрежи са математически модел, който симулира поведението на невроните и се променя автоматично чрез обратно разпространение на грешките [30]. Тези пет техники се използват широко в алгоритмите за машинно обучение с наблюдение и могат да се прилагат за различни задачи, като класификация и регресия, включително и машинно калибриране на данни. Важно е да се отбележи, че случайна гора и изкуствените невронни мрежи са стохастични техники, следователно са извършени няколко серии от итерации в модела ни, за да се получат оптимални резултати.

Тип на невронната мрежа и трениране на модела

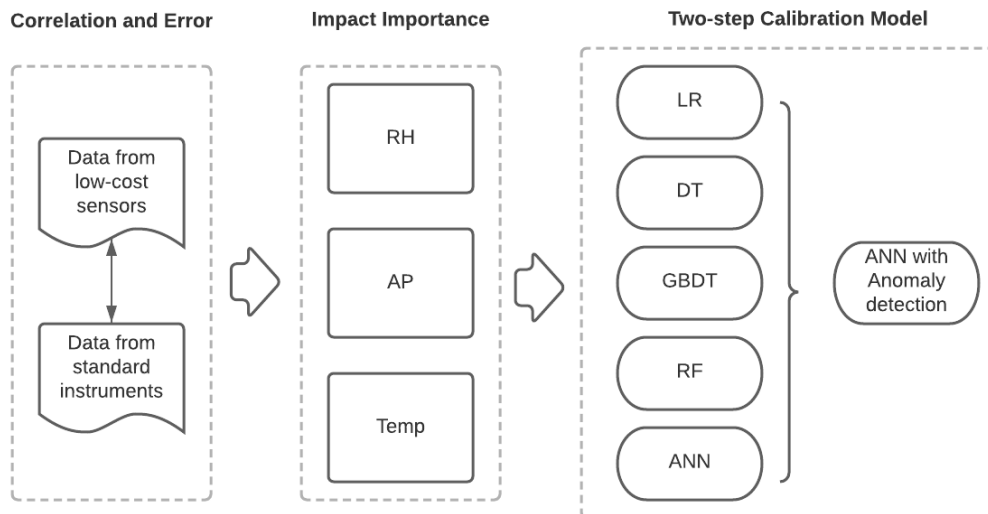
За целите на калибрирането на данните от лазерни сензори за качество на въздуха към данните от официалните станции, е използвана изкуствена невронна мрежа тип многослоен перцептрон (MLP). Този модел е избран, тъй като е подходящ за регресионни задачи, при които се предсказват

непрекъснати стойности. Входните данни за модела включват влажност, атмосферно налягане и температура, както и 'сурови' стойности на финните прахови частици (ФПЧ), измерени от нискобюджетни сензори. Целта на мрежата е да предсказва коригирани стойности на ФПЧ, които да съответстват на измерванията от официалните станции за мониторинг на въздуха, използвани като еталон. За тази цел, изходният слой на модела съдържа един неврон с линейна активационна функция, която е подходяща за регресионни задачи, тъй като резултатът е непрекъснатата стойност. Архитектурата на модела включва три слоя – два скрити слоя с по 64 и 32 неврона, които използват активационната функция ReLU, и изходен слой с един неврон за предсказване на коригираните стойности на ФПЧ.

Моделът е обучен върху набор от данни, обхващащ период от три месеца, като входните данни представляват почасови средни стойности на измерванията. Преди обучението всички входни характеристики се нормализират, за да се ускори процесът на конвергенция и да се осигури стабилност на обучението. Като функция на загуба е използван средна квадратична грешка (MSE), която е стандартна при регресионни задачи, тъй като тежко наказва големи отклонения между предсказаните и реалните стойности. За оценка на производителността на модела са използвани метриците R^2 , която измерва колко добре моделът обяснява вариациите в данните, и средна абсолютна грешка (MAE), която дава информация за средната разлика между предсказаните и реалните стойности на ФПЧ. Обучението е извършено с помощта на Adam оптимизатор, който е широко използван поради своите адаптивни способности. За да се избегне пренатоварване на модела, е приложена техника за ранно спиране, като процесът е контролиран чрез наблюдение на валидационната загуба, за да се предотврати прекомерното обучение и влошаване на резултатите върху нови данни.

3.2.5 Дизайн на модела

На Фигура 3.1 се демонстрира моделирането на двустъпковия метод за калибриране на данни от станции за наблюдение на въздуха. В първия етап на модела се използват 5 контролирани техники за машинно обучение (supervised ML) и се оценяват техните резултати. Най-ефективният от тези пет метода се използва отново във втората стъпка, която включва откриване на аномалии (anomaly detection) за изчистване на отклоненията. На всяка стъпка оценяваме дали това подобрява резултатите на модела



Фигура 3.1: Двустъпов модел на калибриране чрез изкуствена невронна мрежа и намиране на аномалии.

и с колко.

Откриването на аномалии и тяхното премахване е втората стъпка в модела и тя може да подобри данните, но това не е сигурно. Това е метод за откриване на необичайни обекти или събития в набори от данни, които са извън обичайното [82]. Детекторът на аномалии е използван за премахване на отклонения само от обучителния набор от данни. Тъй като това е неконтролирано обучение, е направена оценка със същата настройка на ANN, преди и след почистване на набора от данни, за да се определи дали неконтролираното обучение е подходящо за този набор от данни.

Процесът на обучение на модела е разделен на два етапа: обучение и тестване. Необработените данни са разделени на два набора от данни на случаен принцип, с 80% за обучение и 20% за тестване. Моделът първо е обучен с помощта на данните за обучение, а след това изходът му е оценен от тестовия набор от данни.

3.3 Приложение на модела върху безжична сензорна мрежа

След калибрирането на 5 сензора, разположени точно до стандартните инструменти, ние прилагаме същия модел към други сензори от мрежата Luftdaten. В анализа са добавени данни от измервания от 13 други сензора, които са в близост до 500 метра до регулаторните станции.

Целта на тази корекция е да се прогнозира „реалната“ концентрация на ФПЧ10 на място, като се пресъздаде концентрацията на ФПЧ10 възможно най-точно.

3.3.1 Данни и модел

За целите на това изследване ние вземаме данни от евтини сензори, които са инсталирани и се експлоатират от гражданите. Поради това има липсващи данни. Процентът на липсващите данни сред евтините сензори е 9,4% в резултат на повредени сензори не са получени данни, като причините за това включват: повреда в работата на кутията (напр. спиране на хранването); проблеми, свързани с времето (няма валидна времева променлива); и невъзможност за предаване на данни. Средният процент пълнота на данните за сензорите е 89,2%, с медиана от 96,7 %.

Обсъдените по-горе пет контролирани ML техники са използвани в първата стъпка от анализа: LR, DT, GBDT, RF модели и ANN. След което е изследвани тяхната ефективност. За оценка на моделите се използва корелационният коефициент за квадратична грешка R^2 като индикатор за ефективност при избора и оценката на нашите модели. По отношение на подбора на променливи, ние избрахме само променливи в етапа на валидиране, които се използват за подобряване на констатациите.

Коефициентът на детерминация, известен също като R-квадратично (R-squared), е статистическа мярка, която се използва за оценка на обяснителната способност на статистически модел. Този коефициент показва колко добре моделът обяснява вариацията на зависимата променлива. R-квадратично се изчислява като съотношението между обяснената вариация и общата вариация на данните. Обяснената вариация представлява вариацията, която моделът обяснява и предсказва успешно, докато общата вариация представлява пълната вариация в данните. Стойността на R-квадратично варира между 0 и 1. Ако R-квадратично е 0, това означава,

че моделът не обяснява никаква вариация в данните, докато стойността 1 указва, че моделът обяснява пълната вариация. По-високите стойности на R-квадратично показват по-добра прогнозна способност на модела.

3.3.2 Моделиране на измерването на налягане

Безжичната сензорна мрежа се състои от сензори, монтирани на различни височини. За оценка на измерването на налягането от евтините сензори и идентифициране на всякакви отклонения използваме барометричната формула и коефициента за корелация на Пиърсън. Като входни данни са използвани както стойностите от евтини ФПЧ сензори, така и референтни инструменти. Височината на монтаж за всеки евтин сензор е известна и добавена към набора от данни. Следователно можем също така да идентифицираме точно разликата във височината между всеки сензор и да анализираме измерванията за атмосферно налягане.

За оценка на резултатите от атмосферно налягане от евтини сензори е използван параметричният корелационен коефициент на Пиърсън (2.1). Той осигурява мярка за линейната връзка между двете непрекъснати променливи. А барометричната формула (3.1) се използва за моделиране, как налягането на въздуха се променя с надморската височина и е както следва:

$$P = P_b \cdot \left[\frac{T_b + L_b \cdot (h - h_b)}{T_b} \right]^{\frac{-g_0 \cdot M}{R^* \cdot L_b}} \quad (3.1)$$

където:

- P_b = референтно налягане (Pa);
- T_b = референтна температура (K);
- L_b = скорост на спадане на температурата (K/m) в ISA;
- h = височина, на която се изчислява налягането (m);
- h_b = височина на референтно ниво b (метри; напр. $h_b = 11\,000$ m);
- g_0 = гравитационно ускорение: $9,80665 \text{ m/s}^2$;
- R^* = универсална газова константа: $8,3144598 \text{ J/(mol}\cdot\text{K)}$;
- M = моларна маса на земния въздух: $0,0289644 \text{ kg/mol}$.

Таблица 3.2: Резултати от супервизирани ML модели

ТИП МОДЕЛ	MAE	MSR	R^2
LR:	11,19	288,12	0,77
DT:	8,89	170,03	0,86
GBDT:	8,68	145,22	0,89
RF:	7,96	125,57	0,90
ANN:	6,27	83,90	0,94

3.4 Резултати и оценка

В Таблица 3.2 са показани резултатите от петте контролирани ML техники. За по-добра оценка на евтините сензори заедно с коефициента на детерминация (R -квадратично) са изчислени средната абсолютна грешка (MAE), средната квадратична грешка (MSE).

Средната стойност на квадратична грешка между сензорите за ФПЧ и стандартните инструменти без калибриране е 0,62. Моделът с линейна регресия показва най-лошия резултат след калибриране със средна стойност R^2 от 0,77.

Най-добрата корелация за ФПЧ10 е получена от модела с изкуствени невронни мрежи. Средната стойност на R^2 е 0,94 (ФПЧ10), което съвпада с констатациите от предишни проучвания [74], [73].

За дългосрочно сравнение евтиният сензор и обикновеният инструмент са поставени на едно и също място, което е общ подход за оценка на сензорите в предишни проучвания [35], [41], [45].

3.4.1 Оценка на резултатите за относителната влажност

В таблица 3.3 се вижда, че RH, последвана от температурата се считат за най-важните фактори, оказващи влияние върху ефективността на сензора за частици. В предишни проучвания е доказано, че високата RH е катализатор за предизвикване на хигроскопичен растеж на частиците и промяна на оптичните свойства, което води до съществени смущения за лазерните сензори за ФПЧ [53]. Това се потвърждава и в моделите ни за RF и ANN за стойностите на ФПЧ10, където RH се оказва с най-голямо значение.

За съжаление, нормалните уреди не могат да бъдат сравнени за съот-

Таблица 3.3: Важността на параметрите въз основа на RF.

Съдържание	Параметър
RH:	46,62%
Температура:	30,71%
Атмосферно налягане:	22,67%

Таблица 3.4: Сравнение на едностъпков и двустъпков модел.

ТИП МОДЕЛ	MAE	MSE	R ²
едностъпков	6.27	83.90	0.94
двустъпков с откриване на аномалии:	5.62 (↑5.16%)	65.37 (↑14.69%)	0.95

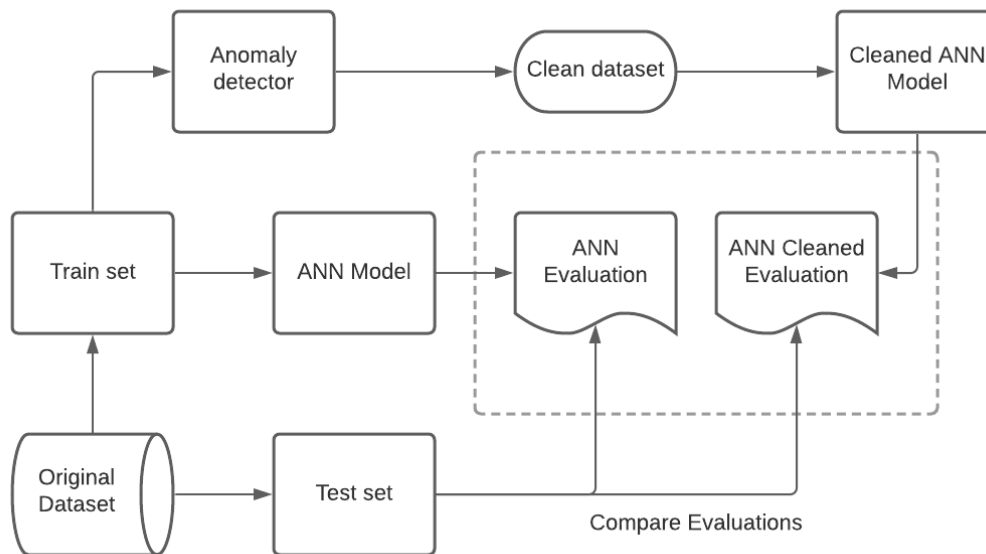
ношението ФПЧ2,5/ФПЧ10, тъй като, както беше споменато по-горе, от петте стандартни уреда само един измерва ФПЧ2,5.

3.4.2 Резултати от модела за калибриране

В Таблица 3.3 са представени статистическите резултати от тестването на всеки модел, където са определени средната абсолютна грешка (MAE), средната квадратична грешка (MSE) и R^2 .

Резултатите от останалите пет отделни модела показват, че ANN моделът се е представил най-добре. Моделът RF показва малко по-лоши резултати. R^2 за ФПЧ10 се увеличи от 0,62 до 0,9 и 0,94 съответно за RF и ANN. Моделът ANN се представи най-добре от 5-те модела, малко по-добре от модела RF, поради което беше избран да бъде използван за втората стъпка - с откриване на аномалии. Следователно използването на неконтролирано обучение в това проучване се счита за полезно и подобрява резултата от калибрирането.

В заключение, окончателният модел на ANN (ANN с откриване на аномалии) има най-добър резултат от калибрирането, с най-малка грешка и най-добра корелация, което показва, че предложеният двустепенен модел е по-точен от единичния модел при калибрирането на модела на евтини сензори.



Фигура 3.2: Модел на изкуствени невронни мрежи с откриване на аномалии

3.4.3 Резултати от приложението в сензорната мрежа

Резултатите за качеството на въздуха от 13 сензора, инсталирани на различни височини в радиус от 500 m, са оценени, като е използвана барометричната формула в изчисленията и данните от референтния инструмент. Височината на инсталиране на сензорите беше 3, 6, 8, 12 и 18 m, докато височината на референтните инструменти е 2 m. Изчисленията показва корелация между сензорите и еталонните инструменти със средна стойност на коефициента на корелация на Пиърсън $r = 0,91$. Този резултат може да се определи като много висока корелация. Въпреки това един сензор, разположен на височина 6 m в близост до станцията в Павлово, показва по-ниска корелация със стандартния инструмент. Този сензор показва по-висока AP след корекцията и е оставен за по-нататъшна оценка като отклонение.

Оценихме резултатите след калибриране с 5-те ML техники и отново моделът с ANN беше най-добре представящият се. Така този модел беше използван за втората стъпка. На първата стъпка след калибрирането на 13-те станции с помощта на ANN, R^2 беше 0,92. След прилагането на

откриването на аномалии резултатите се подобриха и R^2 се увеличи до $R^2 = 0,93$. Същият сензор, който имаше по-ниска корелация със стандартните станции за измерване на атмосферното налягане, имаше най-ниска стойност за $R^2 = 0,87$. Може да се заключи, че моделът не само може да калибрира данните, но и да открива отклонения. При отклонение, този сензор трябва да се изследват допълнително причините за това несъответствие в корелацията с останалите сензори, участвали в модела. RH отново е най-същественният фактор в резултатите от калибрирането.

3.5 Заключение и бъдещи изследвания

Ефективността на сензорите за ФПЧ беше измерена чрез сравняване на стандартни инструменти с помощта на безжичната сензорна мрежа. За калибриране на стационарните сензори беше разработен процес в две стъпки и бяха оценени резултатите от модела.

Резултатите от двустъпковия модел са окуражаващи. R^2 на фиксираните сензори за ФПЧ10 се увеличи от 0,62 до 0,95. Моделът ANN имаше най-силно въздействие от петте независими модела, следван от модела RF, докато моделът LR е неефективен.

Детекторите на аномалии могат да бъдат неконтролирана алтернатива на класификаторите при небалансирана съвкупност от данни. В резултат на тях се открива нежелано поведение на сензора и се отстранява от набора от данни. Откриването на аномалии подобри крайните резултати в това изследване.

Относителната влажност се оказа най-важният фактор за резултатите от калибрирането. Това беше нещо очаквано, тъй като високата влажност е условието, при което евтините сензори показват най-много слабости в качеството на данните. Това се оказа така и в това изследване, като имаше по-висок фактор на въздействие от температурата и атмосферното налягане.

Стойностите на атмосферното налягане на стандартните станции и сензори бяха оценени чрез използване на изчисления с барометричната формула. Корелацията беше силна, което означава, че евтините сензори могат да се разглеждат като добър източник за моделиране на замърсяването на въздуха при вертикално планиране в по-нататъшни изследвания.

Този тип калибриране, използвайки контролирана и неконтролирана

ML, демонстрира възможността за подобряване на резултатите от евтините сензори. Освен това то може да се използва и за оценка на отклоненията. Могат да се проведат допълнителни изследвания, за да се разбере дали тези отклонения са повредени сензори, неправилно инсталирани такива или те работят добре, но съществуват хиперлокални промени в условията на атмосферата и качеството на въздуха.

За да имаме по-добро наблюдение над околната среда с висока разделителна способност, основано на надеждни резултати от наблюдението, при по-ниски разходи за оборудване и поддръжка, комбинацията от официални и граждански сензори е добро решение. Наличието само на един стандартен уред за измерване на ФПЧ_{2,5} е крайно недостатъчно за град като София с 1,2 млн. жители. Въпреки това има WSN с близо 300 сензора, които измерват ФПЧ_{2,5}. Чрез подобряване на надеждността на сензорите за ФПЧ проучвания като това могат да бъдат полезни за успешния контрол и управление на качеството на въздуха в големи градове като София.

По-нататъшните проучвания ще бъдат от полза при включването на газови сензори в мрежата WSN. Освен това е полезно е да се анализират автомобилните емисии с интегрирани мобилни сензори в превозните средства и да се използва подобрението на модела от това изследване.

Глава 4

Разработка на софтуерен модел за избор на велосипедно трасе с оглед замърсяването от ФПЧ

Резултатите от тази глава са публикувани в [103].

Лошото качество на въздуха в големите градове е допринесло за замяната на моторните превозни средства с активен начин на придвижване, като например колоездене [72]. Този метод е широко възприет от множество общности поради намалените задръствания [21] и многобройните ползи за здравето от физическите упражнения. Все по-голям брой изследвания разглеждат промените, свързани с човека, които влияят върху модалния дял на колоезденето, но все още е необходима повече информация за възприеманите опасности, тактиките за смекчаване на последиците и тяхното въздействие върху субективното чувство за безопасност на колоездачите[8]. От друга страна, велосипедната инфраструктура в близост до пътни артерии е идентифицирана като вреден сценарий за велосипедистите поради излагането на замърсяване на въздуха [7]. Въпреки че това е предизвикало интереса на научната общност [22], малко са изследванията, проведени в европейските градове, където много хора са постоянно изложени на ФПЧ от антропогенни източници, като например автомобилния трафик.

Оценките на излагането на замърсяване на въздуха за целите на изследователски проекти често се основават на измервания, направени от регулаторни стационарни станции за наблюдение, като например тези, управлявани от Европейската агенция по околна среда (ЕАОС). Въпреки че тези

станции за наблюдение са изключително точни и подходящи за осигуряване на съответствие със стандартите на ЕС за качество на въздуха, тяхната стойност за документиране на експозицията на замърсяване на индивидуално ниво е ограничена по различни причини:

- На първо място, тъй като местата на станциите рядко съвпадат с местата на експозиция (напр. дом, работа или училище), експозицията на индивида на замърсяване на въздуха може да бъде измерена само косвено чрез техники за пространствена интерполация, като например интерполация с обратно претеглено разстояние и кригиране, или чрез статистически методи, като например регресионно моделиране. [97]
- На второ място, регулаторните станции предлагат ограничена времева разделителна способност (напр. средночасови стойности при станции за прахови частици), което може да доведе до пропускане на преходни скокове в нивата на замърсяване;
- На трето място, косвените методи за оценка на експозицията обикновено оценяват експозицията за едно-единствено място на човек, като например местоживеенето му, работното място [77] или училището [38], което не обхваща експозициите, които се случват, докато хората са на различни места или по време на редовни дейности като пътуване до работа.

Ето защо личната експозиция трябва да се оценява и измерва с помощта на различни устройства или още по-добре е да се намери подходящ начин за включване на стандартни инструменти с мобилни сензори.

По-голямата част от специализираните и обособени велосипедни алеи в градовете са в близост до пътни артерии с интензивен трафик, което може да доведе до значителен риск за здравето на велосипедистите поради високия прием на замърсители, свързан с по-високите нива на вентилация [108], [36] и високите нива на физическа активност [7], [69]. Изследователите са се съсредоточили върху оценката на действителните нива на експозиция на велосипедистите по предварително избрани маршрути с помощта на персонални пробовземачи [23] или върху извеждането на личната експозиция от измерванията на замърсяването на улично ниво с помощта на мобилни лаборатории или стационарни станции за наблюдение на атмосферния

въздух [68]. Много проучвания също така се опитват да свържат определени физиологични реакции с излагането на велосипедистите на замърсяване на въздуха и откриват доказателства, че краткосрочното излагане може да доведе до вредни последици за здравето [94]. Едно изследване дори установява, че велосипедистите абсорбират по-голяма част от фините частици $\text{FP}_{2,5}$ и черния въглерод, отколкото водачите на моторизирани видове транспорт [40].

Въпреки обсъжданите констатации, общинските власти в София и други български градове нямат практиката да отчитат проблемите с качеството на въздуха при разработването и изграждането на велосипедна инфраструктура. Повечето от специализираните велосипедни алеи в градовете са изградени върху пътни артерии с интензивно движение. В резултат на това е необходима повече научна информация, за да се опише излагането на велосипедистите на замърсяване на въздуха и да се оптимизират велосипедните маршрути. В това проучване се разглежда разработването и оценката на инструмент, който предлага оптимизирани велосипедни маршрути чрез изчисляване на приемната доза на околните ФПЧ. Ще подчертаем също така въздействието на транспорта върху качеството на въздуха и това, че велосипедните трасета трябва да бъдат на малки улици, вместо да се изгражда велосипедна инфраструктура в близост до пътни артерии с интензивен трафик.

Съдържанието на тази глава е както следва: Раздел 2 представя формулировката на проблема. В раздел 3 са представени входните данни, използвани в софтуерния модел. В раздел 4 е направен кратък преглед на методологията, използвана в това изследване. Също така в него се обсъжда как е изграден софтуерът, как са избрани маршрутите за изследване, кои са участниците в това изследване и как са използвани мобилни и стационарни сензори. Освен това е показано как се изчисляват скоростта на вдишване и приетата доза. В раздел 5 са представени резултатите от изчисленията, анализът и обсъждането. И накрая, в раздел 6 са представени заключенията и препоръките за бъдещи изследвания.

4.1 Формулиране на задачата

Тук е разгледан въпросът за недостигът на количествено определяне на експозицията на градските велосипедисти на ФПЧ в райони с лошо качес-

тво на въздуха. В град София, известен с високите си нива на замърсяване на въздуха, липсва цялостен метод за оценка и количествено определяне на нивото на експозиция на велосипедистите. Този пропуск в знанията възпрепятства разработването на ефективни стратегии за минимизиране на експозицията и осигуряване на по-здравословни велосипедни маршрути.

За да се справим с този проблем, предлагаме разработване и оценка на софтуерен модел, който използва модифициран метод за определяне на маршрута, който отразява очакваното вдишване на ФПЧ от велосипедиста, осигурявайки по-благоприятен за здравето маршрут. Чрез включването на данни за качеството на въздуха, концентрациите на замърсителите и моделите на движение в софтуерния модел можем да определим количествено излагането на велосипедистите на ФПЧ и да предложим препоръки за маршрути с минимално излагане. Този новаторски подход има за цел да запълни съществуващата празнина в количественото определяне на експозицията и да предостави практически инструмент за градските велосипедисти в замърсени градове като София.

Изследванията, проведени в тази глава, се основават на предишни проучвания, които подчертават влиянието на замърсяването на въздуха върху честотата на заболяванията [102]. В други проучвания се отбелязва, че въпреки силното усещане за замърсен въздух, пътуващите до работа продължават да използват активни форми на транспорт [28]. Наличието на замърсяване на въздуха обаче оказва влияние върху избора им на маршрут [32]. Ето защо е от решаващо значение да се предостави на велосипедистите точна информация за качеството на въздуха по различните маршрути, за да могат те да вземат информирани решения по отношение на своето здраве.

Като предлага практическо решение на проблема с количественото определяне и справяне с излагането на градските велосипедисти на ФПЧ, този софтуерен модел има потенциала значително да подобри условията за колоездене в замърсени градове като София. Той дава възможност на велосипедистите да вземат информирани решения за своите маршрути, като избират такива, които свеждат до минимум излагането им на замърсители на въздуха. В крайна сметка този инструмент има за цел да допринесе за опазване на здравето на градските велосипедисти и насърчаване на устойчив избор на транспорт.

4.2 Работа с данни в модела

Софтуерът реализира системата за вземане на решения, която е изложена в тази глава. Той е разгърнат в облачна инфраструктура, а за разработването му са използвани Python, Django и GraphQL (за базата данни). Той комбинира *gis* данни с данни за трафика и качеството на въздуха, за да приложи модифицирания алгоритъм $k - SPwLO$ за намиране на най-подходящия велосипеден маршрут и да го изобрази на картата *Openstreet*.

Входните параметри са:

- Измервания на ФПЧ от фиксирани сензори;
- измервания на ФПЧ от мобилни сензори;
- данни за трафика;
- температура и влажност от стационарни сензори;
- начална и крайна точка на пътуването;
- денивелация на терена;
- сърдечен ритъм;

Системата за вземане на решения се калибрира, като се използват данни за HR, IR, скорост на колоездене, измервания на ФПЧ от мобилни и стационарни сензори и данни за трафика. Софтуерът обединява данните за трафика и качеството на въздуха, взети от няколко източника, със специфични данни за велосипедния маршрут, като точното местоположение, времетраене на каране и денивелацията.

Следващият раздел описва разработването на софтуерния инструмент, който оптимизира велосипедните маршрути, и метода за изчисление на оценка на излагането на велосипедистите на замърсяване на въздуха.

4.3 Проект на изследването

В това проучване се разглежда създаването на софтуерен инструмент, който има за цел да избере оптимален маршрут за колоездене, който осигурява

най-малката доза вдишване на ФПЧ за велосипедист, който се опитва да стигне от точка А до точка Б.

Инхалационната доза (ID) на замърсителите на въздуха е променлива величина, която зависи от концентрациите на замърсителите, времето и степента на вентилация (VR) [min]. Изчисляваме инхалационната доза, като включваме в модела експозицията на ФПЧ за всеки велосипедист с биомаркери, като сърдечна честота, и времето, необходимо за изминаване на всеки маршрут. Следващият подраздел ще предостави по-задълбочена информация начините на изчисляване, използвани в модела.

4.3.1 Метод за изчисляване на скоростта на вентилация и инхалационната доза

За да изчислим VR на велосипедиста (в l/min), използваме уравнението на модела [25], което се основава на честотата на сърдечния ритъм (HR) [min] (уравнение 4.1).

$$VR = 0,00070724 \times HR^{2,17008537} \quad (4.1)$$

За да определим количествата прахови частици, които въздействат на велосипедистите, използваме уравнение 4.2 [25], за да изчислим дозата на вдишване на ФПЧ за всеки участък:

$$\text{ФПЧ}_{inh} = \text{ФПЧ}_{conc} \times VR \times time \quad (4.2)$$

където ФПЧ_{inh} (μg) е масата на замърсителите, които попадат в дихателните пътища на велосипедистите по време на пътуването (в двете посоки); ФПЧ_{conc} ($\mu\text{g}/\text{m}^3$) е средната експозиция на замърсителите.

4.3.2 Концепция при създаване на софтуера

Формулите за изчисляване на инхалационната доза (уравнения 4.1 и 4.2) водят до следната хипотеза: ако искаме да създадем инструмент, който намалява инхалационната доза на ФПЧ, той трябва да избере трасе, което е:

- Бързо и кратко. Колкото по-малко е времето, толкова по-малко е ФПЧ_{inh}

- Изисква по-малко усилия. NR се увеличава по време на изкачване и висока скорост. Търсете маршрути с ниска денивелация.
- Преполова се през по-малки концентрации на ФПЧ. Пътищата с интензивно движение трябва да се пропускат. Предпочитат се малки улици и алеи в паркове.

Тези формули (уравнения 4.1 и 4.2) се използват и при валидирането на инструмента, което ще бъде разгледано по-късно. В следващия раздел ще разгледаме алгоритмите за намиране на път.

4.3.3 Алгоритъм за намиране на k-кратки пътища с ограничено припокриване

Алгоритмите за намиране на път са изградени върху алгоритмите за търсене в графове и изследват връзките между възлите, като започват от един възел и се движат чрез връзки, докато достигнат целта си. Тези алгоритми определят най-евтиния маршрут от гледна точка на теглото или скоковете. Теглата могат да бъдат всичко, което може да бъде измерено, включително капацитет, цена, време и разстояние.

Както беше обсъдено в предишния раздел, за да се изчисли най-малкото вдишване на мръсен въздух, от този софтуер се изисква преди всичко да изчисли най-краткия път в линейно време [4], [95]. Освен това той трябва да представя варианти, които са по-дълги от най-краткия път, но имат различни желани характеристики, като например по-малък трафик на превозни средства и малка денивелация. Проблемът за k-кратките пътища е прост метод за изчисляване на алтернативни маршрути [57].

Ето защо в това проучване решихме да използваме алтернативен маршрут, и по-специално k-Shortest Paths with Limited Overlap (k-SPwLO), който е представен по-рано в [15]. Приложението k-SPwLO търси пътища, които са:

- възможно най-кратки;
- достатъчно различни един от друг.

Въпреки че методът се представя по-добре от базовото решение, което изброява пътищата по реда на нарастващата им дължина, OnePass не е

полезен дори за средни по размер пътни мрежи [16]. За тази цел използваме MultiPass - по-точен метод, който чрез добавяне на втори критерии за подрязване, разширява и подобрява OnePass. MultiPass обхожда мрежата $k-1$ пъти, но оценява и разширява само най-обещаващите пътища, за разлика от OnePass, който обхожда пътната мрежа веднъж и разширява само онези пътища, които отговарят на критерия за сходство. Подрязването се извършва за всеки път, който не може да доведе до решение.

Нека P е съвкупност от маршрути в пътната мрежа G , които свързват възлите s и t . Път $p'(s \rightarrow t)$ се нарича "алтернатива" в P , когато p' е достатъчно различен от всеки път $p \in P$. Формално, коефициентът на припокриване между p' и p определя доколко те са подобни:

$$Sim(p', p) = \frac{\sum_{(n_x, n_y) \in p' \cap p} w_{xy}}{l(p)}, \quad (4.3)$$

където $p' \cap p$ показва групата ребра, които са общи за p' и p . Като се има предвид прагът на сходство θ , маршрутът p' е алтернативен на множеството P , ако $Sim(p', p) \leq \theta$.

При даден изходен възел s и целеви възел t заявката k -SPwLO предоставя колекция от k маршрута от s до t , подредени в нарастващ ред по дължина, така че:

- най-краткият маршрут $p'(s \rightarrow t)$ винаги е включен;
- всички k маршрута са двойно несходни по отношение на прага на сходство θ ;
- всички k маршрута са възможно най-кратки.

В тази глава се разглежда използването на нов подход към k -SPwLO. Повечето приложения, които използват множество алтернативни маршрути, оставят крайния избор на маршрута на потребителя. Докато в този софтуер ние добавяме тегла (коефициенти), за да изберем най-краткия алтернативен маршрут с най-малък трафик и най-малка денивелация. Последователността е следната: първо, определяме два или повече маршрута с помощта на k -SPwLO, след което преценяваме кой от тези маршрути минава през по-малко натоварени пътища и къде денивелацията е най-малка, така че вентилацията да е възможно най-ниска. Крайният резултат е един велосипеден маршрут.

Крайният резултат от функцията k-SPwLO-modified е единичен път. Този път представлява най-краткия път от началния връх до крайния връх, като се вземат предвид определени ограничения, като избягване на големи булеварди и локации с високи нива на ФПЧ2,5 или голям трафик. Алгоритъмът преминава през съседите на всеки връх, като разглежда техните подходящи характеристики, базирани на зададените критерии, и избира пътя с най-ниска цена. Ако алгоритъмът успее да достигне до крайния връх, връща този единичен път.

```
[language=Python]
# Функция за извличане на данни за трафика за дадена локация
def get_traffic_data(location):
# Извличане на данни за трафика
pass

# Функция за извличане на данни за ФПЧ2,5 за дадена локация
def get_pm25_data(location):
pass

# Функция за проверка дали дадена локация е зелена зона
def is_green_area(location):
pass

# Функция за проверка дали дадена локация е голям булевард
def is_large_boulevard(location):
pass

# Реализация на алгоритъма за k-SPwLO
def k_SPwLO_modified(graph, start, end):
# Опашка за съхранение на върхове и техните съответстващи пътища
fringe = PriorityQueue()
# Поставяне на началния връх заедно с празен път в опашката
fringe.push((start, []), 0)
#Множество за проследяване на изследваните върхове
explored = set()

# Продължаване на цикъла докато опашката не е празна
while not fringe.is_empty():
# Извличане на върха и съответстващия му път от опашката
```

```

node, path = fringe.pop()

# Проверка дали текущият връх е целевият връх
if node == end:
# Ако да, връщане на намерения път
return path

# Проверка дали текущият връх е вече изследван
if node not in explored:
# Добавяне на текущия връх към множеството с изследвани върхове
explored.add(node)

# Итериране през съседите на текущия връх
for neighbor in graph.neighbors(node):
# Пропускане на съседите, които са големи булеварди или имат високо ниво
на ФПЧ2,5 или трафик
if is_large_boulevard(neighbor) or get_pm25_data(neighbor) >
    SOME_THRESHOLD or get_traffic_data(neighbor) > TRAFFIC_THRESHOLD:
continue

# Изчисляване на разхода за преминаване от текущия връх до съседа
cost = graph.cost(node, neighbor)
# Ако съседът е зелена зона, умножаване на разхода по съответния фактор
if is_green_area(neighbor):
cost *= GREEN_AREA_COST_FACTOR

# Добавяне на съседа към опашката с новия път и обновен разход
fringe.push((neighbor, path + [neighbor]), cost)

# Ако целевият връх не може да се достигне от началния, връщане на None
return None

```

4.3.4 Измервания от фиксирани сензори

Софтуерът използва инструмент за агрегиране, който извлича данни за качеството на въздуха както от стандартни инструменти за мониторинг на околната среда на ИАОС, така и от евтини сензорни мрежи като Luftdaten, Smog, Орепаqи и много други. Той получава, записва, почиства и калибрира данните за качеството на въздуха от фиксирани евтини сензори.

Традиционно концентрациите на емисиите във въздуха се наблюдават от станции за мониторинг на въздуха, оборудвани със стандартно оборудване, което позволява получаването на високонадеждни резултати от мониторинга. София разполага с 5 такива и се оперират от ИАОС. Високите разходи за оборудване и обслужване обаче правят невъзможно изпълнението на изискванията за наблюдение с висока резолюция и оценка на степента на лична експозиция [61], [14]. Тъй като нуждата от по-кондензационно наблюдение постепенно нараства, през последните години в наблюдението на въздуха широко се използват евтини сензори за качество на въздуха поради предимствата на ниската цена, ниската консумация на енергия, бързата работа и бързата реакция [93]. В резултат на това станциите за наблюдение на околната среда често са слабо разпръснати, което води до наблюдения с недостатъчна географска разделителна способност. Неотдавна бяха разработени евтини устройства за измерване качеството на въздуха като възможност, която може да увеличи детайлността на наблюдението. Използването на евтини устройства за качеството на въздуха обаче е свързано с редица недостатъци: Те се влияят от кръстосаната чувствителност на различните замърсители на околната среда, както и от външни променливи, като например трафика, колебанията на времето и човешката дейност, а точността им намалява с времето.

За да се намалят гореспоменатите недостатъци на фиксираните евтини сензори, ние калибрираме получените от тях данни. Това се прави, като се изследват както данните от станциите за наблюдение на околната среда, така и от мрежата от евтини сензори Luftdaten. За да калибрираме данните от евтините сензори и да повишим тяхната надеждност, използваме двустъпковия метод за калибриране [101], който използва изкуствени невронни мрежи и откриване на аномалии.

4.3.5 Измерване на експозицията на ФПЧ посредством мобилни сензори

Евтините преносими сензори за измерване на замърсяване са устройства за наблюдение на околната среда, които хората могат да носят или носят, докато извършват ежедневните си дейности. Тъй като те откриват нивата на замърсяване директно и в реално време, те могат да позволят на доставчиците на здравни услуги и изследователите да наблюдават експозицията на индивидуално ниво и да дадат възможност на гражданите да управляват

личната си експозиция на замърсители отвъд това, което регулаторните станции за наблюдение могат да направят [96].

Сензорът за частици в Airbeam2 - PMS 7003 е седмо поколение от серията PMSx003, разработена от Plantower. PMS7003 е цифров и универсален сензор за концентрация на частици, който се използва за измерване на количеството суспендирани частици във въздуха и извеждане на резултатите чрез цифров интерфейс. Той може да се използва с различни инструменти, за да предоставя точни данни за концентрацията в реално време. Осигурява солидна ефективност на броене на частици от 98% и може да работи при температурен диапазон между -10 и 60 C..

4.3.6 ФПЧ измервания и калибриране на модела за вземане на решения

Както вече беше обсъдено, софтуерът е програмиран да търси маршрут, за който:

- най-краткият маршрут $p'(s \rightarrow t)$ винаги е включен;
- всички k маршрута са двойно несходни по отношение на прага на сходство θ ;
- всички k маршрута са възможно най-кратки.

Изчислителният модел се основава на споменатите по-рано формули за намиране на скоростта на вентилация и инхалация, които са използвани за разработване на теглата за алгоритъма за търсене.

Алгоритъмът е допълнително усъвършенстван чрез оценка на маршрутите, комбинираща данни за трафика заедно с данни за качеството на въздуха от мобилни и стационарни сензори. Обобщаването помага за по-прецизно прогнозиране за изчисляване на най-добрия маршрут. Стационарните сензори, които са разположени в близост до велосипедните маршрути, изготвят базовата линия на концентрациите на ФПЧ в района - $PM_{baseline}$. В същото време данните за трафика дават информация, която помага да се оцени концентрацията на ФПЧ от превозните средства и транспорта - $PM_{traffic}$. Очакваната експозиция на ФПЧ PM_{exp} по велосипедния маршрут се намира чрез комбиниране на двете променливи.

$$PM_{exp} = PM_{baseline} + PM_{traffic} \quad (4.4)$$

Колкото по-близо минава велосипедният маршрут до главен път, толкова по-голямо е влиянието на замърсяването от трафика върху общите концентрации на ФПЧ. Следователно променливата $PM_{traffic}$ се увеличава. Времето, в което се предприема пътуването, също влияе върху тази променлива - преминаването по време на задръствания и в близост до силно натоварени улици и булеварди увеличава концентрациите на ФПЧ, както е отбелязано по-нататък в главата. Действителните концентрации на ФПЧ се оценяват с помощта на измервания на ФПЧ от мобилни сензори. Мобилните сензори осигуряват много точни измервания, тъй като са прикрепени към самия велосипед, предават измерванията на ФПЧ на всяка минута чрез мобилен телефон и отразяват ултралокално замърсяване, като например да се намиращ непосредствено зад автобус или огромен камион. По-късно в полевите изследвания е показано как се извършва тази оценка.

Особеностите за оценка на велосипедните маршрути са силно зависими от местоположението и особеностите на конкретното място [5]. Ето защо ние извършихме много подробно полево проучване, като изследвахме два велосипедни маршрута, описани в следващия раздел.

4.3.7 Проучвани маршрути

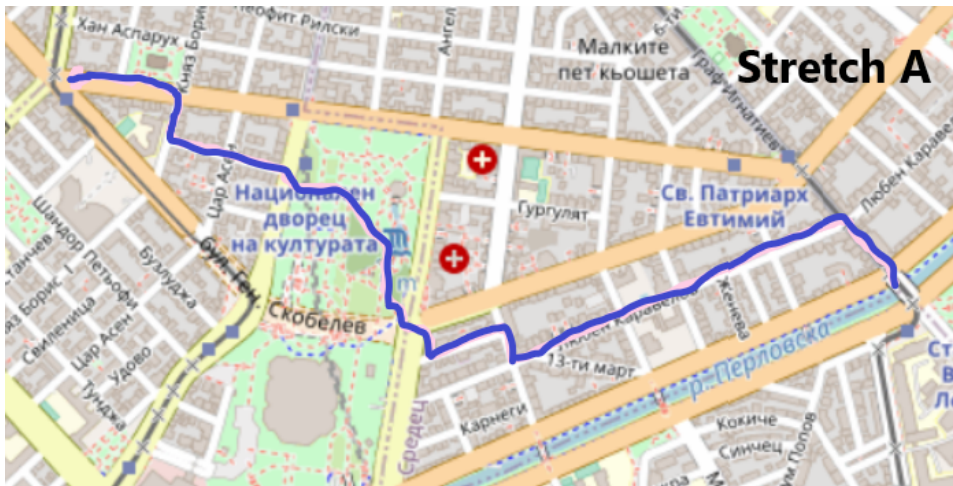
В тази част се разглежда оценката на софтуерният модел с реално теренно проучване в София и разглеждане на два велосипедни маршрута. Двата проучвани маршрута започват от националния стадион "Васил Левски" и завършват в Петте къшета. Тези две места се намират в центъра на София и имат активен велосипеден поток.

Двата проучвани маршрута са избрани, тъй като отговарят на следните критерии:

- имат еднакви начална и крайна точка;
- единият е предложен от софтуера, а другият следва обособени велосипедни алеи;
- маршрутите трябва да се използват активно от велосипедисти;

- и двата маршрута минават в близост до стационарни сензори от мрежата Luftdaten;
- и двата маршрута трябва да са безопасни за редовните велосипедисти.

На следващите изображения се виждат двата пътя, използвани за оценката: Участък А (фиг.4.1), който е оптимизираният велосипеден маршрут, предложен от софтуерния инструмент, и преминава по най-краткия път през малки централни улици и паркове; докато участък Б (фиг.4.2) използва развитата велосипедна инфраструктура (която включва предимно колездене в близост до пътни артерии с интензивно движение), по-дълъг е като разстояние и е предложен от навигационния софтуер, тъй като следва обособени велосипедни алеи.



Фигура 4.1: Отсечка А - предложена от нашия софтуер, който търси най-малката инхалационна доза

4.3.8 Полеви тестове и участници

Полевите тестове са проведени с 10 участници в изследването. Мобилното оборудване бе свързано към предната част на всеки велосипед, което позволяваше на линиите за вземане на проби да улавят замърсителите, без



Фигура 4.2: Отсечка Б - предлагана от повечето навигационни софтуери, който следва обособена велосипедна алея

да бъдат възпрепятствани; то също така беше укрепено отдолу, за да се намалят вибрациите.

Избрахме 10 души (на възраст 27-41 г.) чрез контакт с местната велосипедна мрежа (8 мъже и 2 жени). Преди да се включат в проучването, участниците попълниха предварително скринингово проучване. Критериите за изключване от групата бяха респираторни (включително астма), сърдечносъдови или други хронични заболявания, както и тютюнопушене (настоящо или скорошно). Включени бяха само лица, които пътуват ежедневно с велосипед в София.

Тези фактори бяха използвани, за да се намали рискът от увреждане поради непознаване на улиците в София, липса на опит в карането и неблагоприятни остри последици за здравето. Освен това участниците бяха помолени да се въздържат от алкохол и кофеин в продължение на 48 часа преди тестовете.

Кръговите пътувания бяха извършени в работни дни сутрин в часовете на интензивно движение (HT) (8:00-9:30 ч.) и в часовете на слабо движение (LT) (10:30-12:00 ч.) и в неработни дни (NWD): почивни и празнични дни. Резултатите от трите сценария са оценени поотделно в следващия раздел. Тестовете на терен са проведени през април и май, когато времето е уме-

рено и замърсяването на въздуха от битовото отопление не оказва влияние върху данните. Следователно увеличението на концентрациите на ФПЧ се дължат главно на транспорта.

Десетимата участници извършиха кръгови пътувания по двата участъка по време на НТ, ЛТ и NWD. Сертифициран пулмолог обучи участниците в проекта как да определят сърдечната си честота (НТ) [НВ]/min с помощта на пулсоксиметър.

4.4 Резултати и обсъждане

4.4.1 Концентрации на ФПЧ1 и ФПЧ2,5 от мобилни измервания

В таблица 4.1 са показани минималните, максималните и средните (медианни) стойности на концентрациите на ФПЧ1 и ФПЧ2,5 за двата изследвани участъка в работни дни по време на интензивен трафик (НТ). Концентрациите на ултрафините частици с диаметър под 1 микрон (ФПЧ1) са измерени в диапазона между 8 и 24 $\mu\text{g}/\text{m}^3$ (средно 11 $\mu\text{g}/\text{m}^3$) за участък А и между 8 и 41 $\mu\text{g}/\text{m}^3$ (средно 14 $\mu\text{g}/\text{m}^3$) за участък Б. Концентрациите на фини частици под 2.5 микрона (ФПЧ2,5) са между 12 и 29 $\mu\text{g}/\text{m}^3$ (средно 15 $\mu\text{g}/\text{m}^3$) за участък А и между 12 и 45 $\mu\text{g}/\text{m}^3$ (средно 19 $\mu\text{g}/\text{m}^3$) за участък Б.

Таблица 4.1: Концентрации на ФПЧ1 и ФПЧ2,5 от мобилни измервания в работни дни по време на интензивен трафик (НТ)

	ФПЧ1			ФПЧ2,5		
	min	max	mean	min	max	mean
Участък А	8	24	11	12	29	15
Участък Б	8	41	14	12	45	19

В таблица 4.2 са представени концентрациите на ФПЧ1 и ФПЧ2,5 за участък А и Б в работни дни по време на слаб трафик (ЛТ). Измерванията от участък А показват по-ниски нива на концентрация както за ФПЧ1, така и за ФПЧ2,5. Концентрациите на ултрафини частици с диаметър под 1 микрон (ФПЧ1) са измерени между 6 и 19 $\mu\text{g}/\text{m}^3$ (средно 10 $\mu\text{g}/\text{m}^3$)

за участък А и между 5 и 34 $\mu\text{g}/\text{m}^3$ (средно 12 $\mu\text{g}/\text{m}^3$) за участък Б. Концентрациите на фини частици под 2,5 микрона (ФПЧ2,5) са между 10 и 24 $\mu\text{g}/\text{m}^3$ (средно 13 $\mu\text{g}/\text{m}^3$) за участък А и между 9 и 38 $\mu\text{g}/\text{m}^3$ (средно 15 $\mu\text{g}/\text{m}^3$) за участък Б.

Таблица 4.2: Концентрации на ФПЧ1 и ФПЧ2,5 от мобилни измервания в работни дни със слаб трафик (LT)

	ФПЧ1			ФПЧ2,5		
	min	max	mean	min	max	mean
Учасък А	6	19	10	10	24	13
Учасък Б	5	34	12	9	38	15

Таблица 4.3 илюстрира минималните, максималните и медианните стойности на концентрациите на ФПЧ1 и ФПЧ2,5 за двата изследвани участъка през почивните и празничните дни. Измерванията за участък А показват почти същите ниски нива на концентрация за ФПЧ1 и ФПЧ2,5. Концентрациите на ФПЧ1 са между 3 и 7 $\mu\text{g}/\text{m}^3$ (средно 4 $\mu\text{g}/\text{m}^3$) за участък А и между 3 и 8 $\mu\text{g}/\text{m}^3$ (средно 4 $\mu\text{g}/\text{m}^3$) за участък Б. Концентрациите на ФПЧ2,5 са между 4 и 11 $\mu\text{g}/\text{m}^3$ (средно 7 $\mu\text{g}/\text{m}^3$) за участък А и между 5 и 11 $\mu\text{g}/\text{m}^3$ (средно 7 $\mu\text{g}/\text{m}^3$) за участък Б.

Таблица 4.3: Концентрации на ФПЧ1 и ФПЧ2,5 от мобилни измервания през неработни дни

	ФПЧ1			ФПЧ2,5		
	min	max	mean	min	max	mean
Учасък А	3	7	4	4	11	7
Учасък Б	3	8	4	5	11	7

Въпреки факта, че специализираният велосипеден маршрут е през по-голяма част от времето открит, големият брой автомобили, автобуси и камиони в този коридор е основната причина за повишените концентрации на замърсяване, като по-големият трафик увеличава концентрациите на ФПЧ. В резултат на това измерванията на ФПЧ2,5 са почти еднакви през почивните дни, но с почти 20% по-високи концентрации между двата маршрута през работните дни.

4.4.2 Количество на вентилацията

За да измерим скоростта на вентилация (VR) при дишане, направихме измервания на сърдечния ритъм (HR), кислородното насищане (SpO2) и дихателната честота (RR) на всеки участник в проекта. Средната стойност на всеки параметър има ниска променливост и крайните резултати за VR са представени в Таблица 4.4. Не се наблюдават разлики в тези три фактора в зависимост от това дали маршрутът е изминат в работен или неработен ден, тъй като те не се влияят пряко от трафика или различните нива на излагане на краткосрочно замърсяване на въздуха. Данните за скоростта и денивелацията, показани в таблицата, са взети от разработения софтуерен инструмент благодарение на интеграцията със Strava API.

Таблица 4.4: Осреднени показатели от полеви измервания

	Участък А	Участък Б
VR (L/min)	10,14	11,06
HR (удари/min)	82,30	85,66
Денивелация (m)	3	15
Скорост (km/h)	14,3	14,7
Разстояние (km)	3,8	4,4
Време (min)	15:52	18:04

Въпреки еднаквите начални и крайни точки, разстоянията са различни и участък А е по-къс - 3,8 км в сравнение с 4,4 км от участък Б и е по-бърз - 15:52 мин в сравнение с 18:04 мин от участък Б. Времето, необходимо за изминаване на двата участъка, е по-скоро постоянно през работните и неработните дни. Специализираният велосипеден маршрут е имал по-дълги и прави коридори с по-малко кръстовища и пресечки. Това е довело до по-високи максимални и средни скорости. Освен това участък Б има по-голяма денивелация - 15 м в сравнение с 3 м в участък А. Всички тези данни водят до повишена скорост на велосипедистите в участък Б - 11,06 /мин в сравнение с 10,14 л/мин за участък А.

В таблица 4.5 са показани резултатите от инхалационната доза на велосипедистите за ФПЧ1 и ФПЧ2,5. Те са изчислени въз основа на средната стойност на експозицията на ФПЧ по време на обиколката, измерена от

мобилни сензори, заедно с времето, необходимо за нейното извършване, и скоростта на вентилация за всеки велосипедист.

Таблица 4.5: Доза за вдишване на ФПЧ1 и ФПЧ2,5 по време на обиколките

Период	Замърсител	Участък А (оптимизиран)	Участък Б (велосипедни алеи)
Интензивен трафик	ФПЧ 1_{inh}	29.74	46.45 (↑56%)
	ФПЧ2.5 $_{inh}$	40.56	63.04 (↑55%)
Слаб трафик	ФПЧ 1_{inh}	27,04	39,82 (↑47%)
	ФПЧ2.5 $_{inh}$	35.15	49.77 (↑42%)
Неработни дни	ФПЧ 1_{inh}	10.82	13.27 (↑23%)
	ФПЧ2.5 $_{inh}$	18.93	23.23 (↑23%)

Участък Б показва повишена инхалационна доза на ФПЧ, тъй като завършването на маршрута отнема повече време, изисква повече усилия и велосипедистът е изложен на по-високи концентрации на ФПЧ поради изгорелите газове на превозните средства. Дори и през почивните дни, когато концентрациите на ФПЧ са сходни по двата маршрута - участък Б показва по-високи инхалационни дози поради по-продължителното време и по-високата скорост на вентилация (заради по-висока денивелация, по-висока средна скорост).

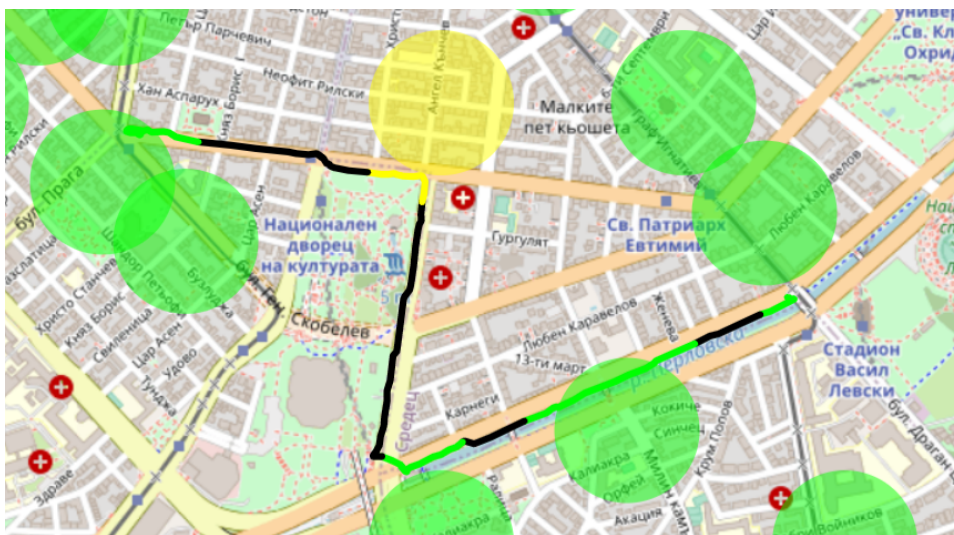
Измерванията за SpO₂ и RR не показват особени краткосрочни ефекти върху здравето. Това се очакваше, тъй като участниците в проучването, от съображения за безопасност, са подбрани непушачи без хронични заболявания и редовни велосипедисти. Като това не означава задължително, че хората с хронични заболявания и чувствителност към замърсяването на въздуха може да не получат някои симптоми или дразнения, както се наблюдава в някои проучвания [73], [18].

4.4.3 Визуализация на велосипедни маршрути чрез включване на данни от стационарни сензори

Ще използваме мрежата от фиксирани евтини сензори на Luftdaten, за да демонстрираме технологията за нашите фиксирани сензори в нашия софтуер. Мрежата, базирана в ЕС, е граждански научен проект, в който въздушните станции се приемат и поддържат от гражданите и се разполагат на техните балкони. София разполага с гъста мрежа от над 300

нискобюджетни станции за качество на въздуха от мрежата на Luftdaten, които осигуряват пространствена и времева разделителна способност за концентрациите на ФПЧ_{2,5} и ФПЧ₁₀.

На фиг.4.3 могат да се видят концентрациите за участък Б по време на LT, където на участъка се прилагат измервания от стационарни сензори в близост до 200 м или по-близко до маршрута. Изображеният маршрут показва точно къде са минали участниците в проучването, благодарение на gis интеграцията на Strava и Openstreetmap API. Черно оцветената линия означава, че в тази част от маршрута няма фиксиран сензор, който да е по-близко от 200 м, докато зелените и жълтите цветни линии представят концентрациите, измерени от фиксирани сензори в близост. За стойностите на концентрацията на ФПЧ_{2,5} между 0 и 12 - се използва зелен цвят, докато жълтият цвят означава стойности на ФПЧ_{2,5} между 12 и 35. Тези цветови категории са вдъхновени от индекса за качество на въздуха на ЕРА и са същите за визуализиране на маршрутите на Aircasting, измерени с преносими сензори.



Фигура 4.3: Маршрут за колоездене с експозиция от фиксирани сензори

Благодарение на гъстата мрежа от евтини сензори в София, за участък А се използват 5 фиксирани сензори, а за участък Б - 6 сензори, тъй като те отговарят на критериите за подбор. Сравнявайки данните от мобилните сензори, ние изследвахме, че (1) стационарните сензори на Luftdaten,

особено ако не са разположени точно на маршрута, не могат да открият ултралокални пикове в ФПЧ и не успяват да определят зоната с най-интензивен трафик и концентрация на ФПЧ; и (2) сензорите също така не могат да идентифицират времево излагане, като например преминаване до автобус, камион или мотопед, докато мобилните сензори го откриват много успешно. Основната причина за тези две констатации е, че сензорите на Luftdaten са разположени на тихи улици и не предават данни за качеството на въздуха всяка секунда.

За да се смекчат горепосочените проблеми, авторът предлага по-гъста мрежа от сензори за качеството на въздуха по пътищата с интензивен трафик в градовете. Те могат да бъдат прикрепени или интегрирани до светофарите или лампите за улично осветление. Това ще доведе до точност при количественото определяне на излагането на ФПЧ, особено на потенциално опасните за здравето места и натоварените пътни участъци.

Софтуерът има постоянен достъп до концентрациите на ФПЧ от фиксираните сензори, а данните за трафика могат динамично да променят входните стойности за механизма за вземане на решения. Инструментът разполага и с функцията да изобразява един и същ маршрут в различни времеви рамки и съответно да показва различни концентрации на замърсяване на въздуха. Това може да позволи разработването на инструмент за прогнозиране на замърсяването с ФПЧ, който да позволи на потребителя да избере най-подходящото време за пътуване. Софтуерът може да намери приложение и за избор на пешеходни маршрути. И все пак, той ще има по-голямо въздействие върху намирането на велосипедни маршрути поради по-високата вентилация на велосипедистите и честото им съседство с отработените газове от превозните средства.

4.4.4 Прецизиране на модела и оценка на данните от мобилни и стационарни сензори

Средните стойности за концентрациите на ФПЧ_{2,5} от стационарните сензори, които са разположени в близост до изследваните маршрути (разстояние 200 м), са сравнени с измерванията на ФПЧ_{2,5} от мобилните сензори, както е показано в Таблица 4.6 по-долу. Както беше разгледано в предишния раздел, данните от стационарните сензори са калибрирани с помощта на двустъпковия модел за машинно обучение, за да се повиши тяхната точност, разгледан в миналата задача. Установяваме, че съществува силна

корелация между средните стойности на $PM_{2.5_{fixed}}$ и минималните стойности за $PM_{2.5_{mobile}}$, което отчасти доказва предположението, че концентрацията от близките фиксирани сензори може да се използва като изходна за изчисляване на общата оценка на ФПЧ.

Данните от фиксирани сензори са близки до минималните концентрации по маршрута. Обикновено тази част от маршрута е с незначителен трафик и нейната стойност често е близка до $PM_{baseline}$, която се измерва от фиксирани сензори. В същото време максималните концентрации на ФПЧ по велосипедния маршрут се наблюдават, когато се преминава покрай по-значителни замърсители като камиони, превозни средства със стари дизелови двигатели, мотопеди или колони от автомобили, чакащи на светофарите. Както се очакваше, разликата между минималните и максималните концентрации на ФПЧ е по-значителна по време на НТ.

Таблица 4.6: Концентрации на ФПЧ_{2,5} от стационарни сензори, разположени в близост до маршрута, и мобилни измервания

Период	Маршрут	$2,5_{fixed}$		$2,5_{mobile}$	
		mean	min	max	mean
WD-НТ	Участък А	10	12	29	15
	Участък Б	10	12	45	19
WD-LT	Участък А	8	10	24	13
	Участък Б	8	9	38	15
NWD	Участък А	4	4	11	7
	Участък Б	4	5	11	7

4.5 Заключение и бъдещи изследвания

В тази глава е представена разработката на софтуерен инструмент, който оптимизира велосипедните маршрути въз основа на алгоритми, които предвиждат най-малко вредни замърсители на въздуха. Алгоритъмът използва модифицирана реализация на метода за намиране на алтернативни маршрути - k-кратките пътища с ограничено припокриване. Той се основава на експериментални данни и уравнения, които изчисляват общата доза вдишани замърсители от велосипедиста. Заедно с това изследването е оценено чрез два велосипедни маршрута: Участък А - предложен от

новоразработения софтуер, който преминава през малки улици в София, и участък Б - предложен от навигационни приложения, който преминава по обособена велосипедна алея. В полевия тест от изследването десет велосипедисти правят кръгови пътувания по двата маршрута през 3 периода: (1) с интензивен трафик, (2) със слаб трафик в работни дни и (3) в неработни дни. Въз основа на събраните в проучването данни от мобилен сензор са изчислени експозицията на велосипедистите и потенциалната инхалационна доза на замърсяване с ФПЧ1 и ФПЧ2,5 по двата маршрута.

Концентрациите на експозиция по специализираните велосипедни алеи се оказаха по-високи от нивата на експозиция по оптималното трасе, особено в работни дни. Дори в случаите, когато средните концентрации са били почти еднакви, инхалационната доза за велосипедиста винаги е била по-висока по маршрута с велосипедна алея, тъй като той е по-дълъг като време и разстояние, с по-висока денивелация и изисква по-интензивно колоездене. При избора на оптимизирания велосипеден маршрут дозата на вдишване на ФПЧ1 се намалява с 23%, в неработни дни до 56% по време на интензивен трафик в работни дни. За ФПЧ2,5 процентните разлики са приблизително същите 23% в почивни дни и достигайки до 55% по време на пиковите часове в работни дни. По време на работните дни резултатите показват допълнително повишаване на риска за здравето на велосипедистите, използващи изследваните велоалеи, поради появяването на замърсители директно свързани с трафика.

Резултатите от това проучване и разработеният софтуер са полезни за медицинските експерти, велосипедистите и пешеходците. Те доказват, че оптимизирането на мобилността и вземането на решения, основани на данни, могат да намалят излагането на замърсяване на въздуха. Освен това последиците от изследването могат да бъдат полезни за политиците и специалистите по околна среда. Резултатите от това проучване надграждат предишни констатации [3], които показват, че препроектирането на улици за мултимодален трафик с ниска скорост без бариери е по-устойчив и прагматичен подход от изграждането на велосипедна инфраструктура върху пътни артерии с интензивен трафик. В изследваните велосипедни алеи излагането на велосипедистите на ФПЧ е било тясно свързано с автомобилния трафик, тъй като проучването не е проведено през отоплителния сезон. Съществува значителна разлика между изследванията със слаб трафик и с интензивен трафик. Велоалеите без физически бариери между велосипедния маршрут и пътя имат по-висока експозиция - заклю-

чение, което се установява в друго проучване [48]. Нашите констатации допринасят за по-доброто разбиране на проблемите на София, свързани със замърсяването от мръсен въздух и в частност от автомобилен трафик. Също така се подчертават важноста на отчитането на качеството на въздуха при разработването и изграждането на велосипедни мрежи в българските градове.

Това проучване се фокусира върху най-проблемните градски замърсители - фините и свръх-фините прахови частици. По-нататъшните проучвания ще бъдат полезни с включването на данни за други замърсители, свързани с превозните средства, като например емисии на черен въглерод и NO₂. Освен това е полезно да се анализират праховите частици от автомобилните емисии с химичен анализ и допълнително да се открият различните метали в ПЧ. Интегрирането на повече мобилни и стационарни сензори по пътищата с интензивно движение и използването на софтуерния инструмент от това проучване ще подобри допълнително разбирането на свързаното с транспорта замърсяване на въздуха и ще намали експозицията.

Глава 5

Софтуерна система за визуализация на въздушното замърсяване

Резултатите от тази глава са публикувани в [104].

За да можем да използваме всички описани до този момент изследвания и да предадем по достъпен начин информация как замърсяването на въздуха влияе върху здравето, е необходимо да създадем комплексна софтуерна система, която да помогне за събирането, обработката и моделирането на данни. В процеса на работа забелязахме, че различните източници на качество на въздуха (Официални, граждански станции) имат различен начин и структура, по който подават данните си. Съответно е трудна интеграцията в една система и има нужда от значителна доработка. Агрегацията им отваря много възможности за анализ на тези данни.

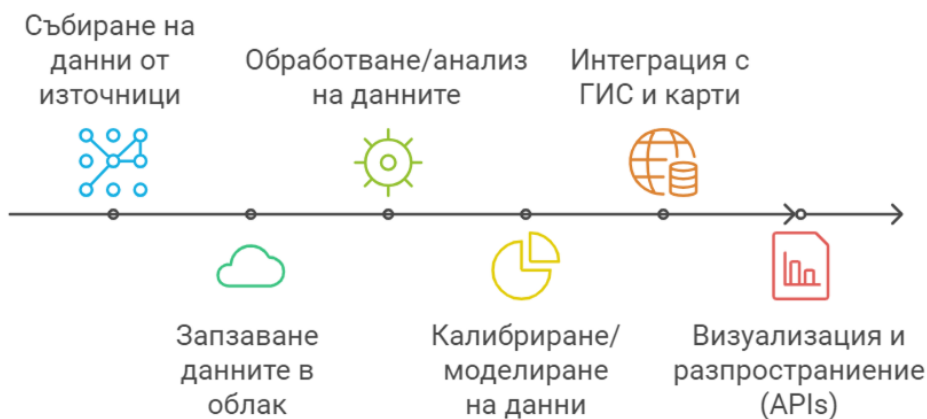
Софтуерната система описана в тази глава позволява събирането, обработката и анализирането на данни за качество на въздуха, метеорологични данни, трафик, географските информационни системи (ГИС). Също така тя поднася информацията по лесно достъпен за използване начин от различни проекти, продукти и услуги. При разработването на тази система е използвана съвременна комбинация от технологии с отворен достъп като Django, Python, GraphQL и Rest API. Тази глава предоставя обширно описание на процеса на разработка и прилагане на специализирана софтуерна система, както и са разгледани на кратко проектите, които я използват.

Цел на задчата е събирането на данни за детайлно измерване на за-

мърсяването на въздуха в реално време, запазване на данните в облачен репозиториум, анализ и намеринето на подходящ начин за визуализация на въздушно замърсяване и трафик върху популярни платформи за картографиране като Google Maps и OpenStreetMap. Накрая се търси достъпен начин софтуерната система да бъде интегрирана в приложения за интелигентни градове.

5.1 Софтуерна архитектура

Архитектурата на софтуера е модуларна и тя определя начина, по който различните компоненти взаимодействат, как се обработват данните и как се визуализират резултатите. На фигура 5.1 е представена как изглежда тя.



Фигура 5.1: Софтуерна архитектура

Архитектурата на софтуера се състои от следните компоненти:

- **Модул за събиране на данни и свързване с API:** Централна част от архитектурата на софтуера е модулът за свързване с различни източници на входни данни. Този модул е отговорен за изграждането и поддържането на връзката с разнообразни външни източници на атмосферни данни. Тук се включват API-та от официални и граждански станции, предоставящи информация за концентрации на различни замърсители във въздуха, температура, влажност, атмосферно

налягане и други фактори, влияещи на качеството на въздуха, като автомобилен трафик.

- Централен репозиториум в облака: След свързването с API, данните се прехвърлят и съхраняват в централен репозиториум в облака. Този облачен репозиториум е избран поради високата му скалируемост и надеждност. В него, съхраняваните данни от различните източници са окрупнени и подходящи за обработка, като облачната система осигурява актуализация на данните в реално време.
- Модул за обработка/анализ на данни: След като данните се съхраняват в облачния репозиториум, те преминават през модула за обработка. Този модул има задачата да почисти и структурира данните, като изключи невалидни или повредени данни и ги преобразува в подходящ формат за анализ. Анализ на данните спрямо специфични нужди и параметри. В този модул например се смята индекса за качеството на въздуха (ИКВ).
- Модул за прогнозиране/калибриране [незадължителен]: След обработката на данните, архитектурата включва модул за прогнозиране. Този е по изборен модул и използва данните от миналото и текущите условия за да генерира прогнози за качеството на въздуха в бъдеще. Тук се прилагат различни алгоритми за анализ и моделиране, които предоставят предупреждения и прогнози за замърсяването на въздуха. Чрез този модул интегрираме и методът, който разработихме за калибриране на данните от лазерни сензори - двустъпковия метод за калибриране с машинно обучение, описан в глава 3.
- Модул за измерване на замърсяване от ФПЧ по зададено трасе [специализиран]: Този модул включва измерване на замърсяване с интегрирани ГИС данни. Това позволява визуализацията на данните, за замърсяването, на картата и ги свързва с географските данни за трасето. Модулът използва данни измервания на ФПЧ, трафик данни, топографски данни, а получените маршрути ги налага върху карти на Openstreetmap и Google maps. Този модул е използван за изчислението на най-подходящо трасе за колоездачи, описан в глава 4.
- Визуализация и разпространение на данни: Накрая, архитектурата включва и модули за визуализация и разпространение на данни. Тези

модули предоставят възможност за визуално представяне на данните и предоставят интерфейси за достъп до информацията за различни потребители и регулаторни органи.

Архитектурата на софтуера се основава на модулна и интегрирана концепция. Тя осигурява ефективно събиране, обработка и прогнозиране на данни за качество на въздуха, което е ключово за управлението на околната среда и за защитата на общественото здраве.

5.1.1 Входни параметри в системата и тяхното съхранение

Осигуряването на точността и надеждността на нашата система за вземане на решения е от първостепенно значение за успеха на нашето изследване. За да постигнем това, нашият софтуер разчита на редица входни параметри. Тези параметри включват:

- измервания на ФПЧ от стационарни лазерни сензори
- измервания на ФПЧ от официални станции за измерване.
- измервания на прахови частици ФПЧ от преносими сензори.
- данни за температура, влажност, атмосферно налягане, посока и сила на вятър от стационарни сензори.
- данни за графика.

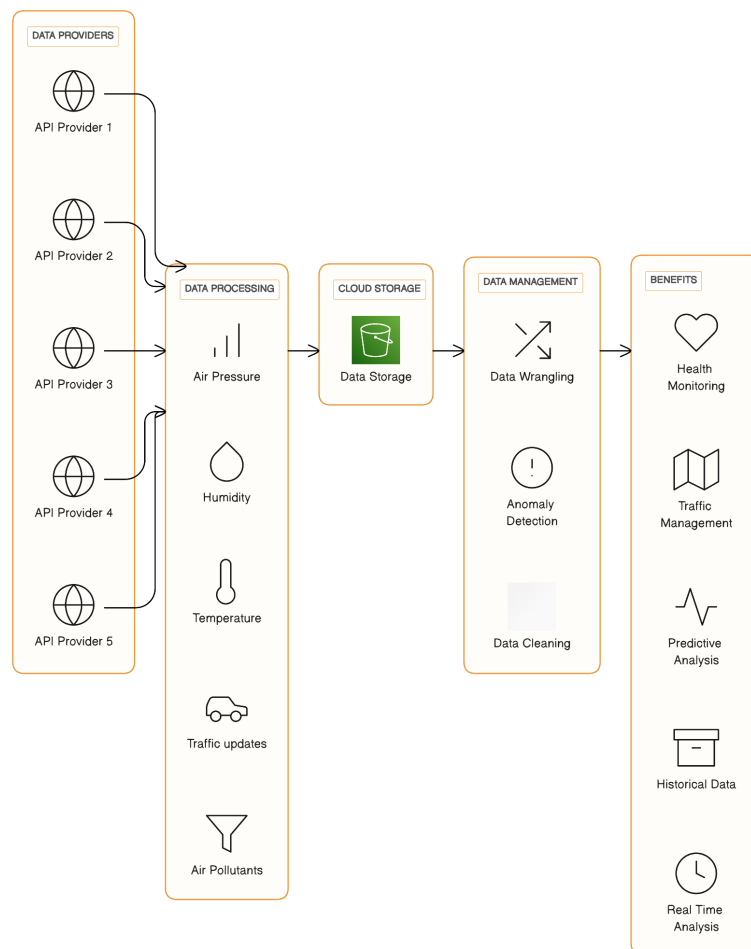
Всеки от тези входни параметри допринася за по-цялостно разбиране на качеството на въздуха. Комбинацията на данни от системи, които са разделени и не си комуникират помежду си в една обща, позволява на нашата система да прави информирани прогнози и оценки на заърсяването и съответно да взема по-надеждни решения.

5.1.2 Изходни параметри от системата

Параметрите, които излизат от системата са следните:

- данни за качеството на въздуха спрямо локация, географски радиус или град в реално време.

- разположение на станциите с техните стойности върху географски карти.
- прогноза за качество на въздуха.
- калибрирани данни от определени станции.
- избор на най-подходящ път спрямо замърсяване от ФПЧ и автомобилен трафик.



Фигура 5.2: Източници за измерване на качеството на въздуха

Графика 5.2 илюстрира разнообразния подход към събиране на данни за качеството на въздуха от множество източници. Ние получаваме информация, свързана със замърсители във въздуха, актуална информация за трафика по основните пътни артерии, температура, влажност и въздушно налягане от различни източници.

5.2 Модул Събиране на Данни

Данните от официални и граждански мрежи със станции за измерване на качеството на въздуха се достъпват посредством заявки към API-та (Luftdaten, aqicn, airthings, openaq и др.). Като системата прави тези заявки през различни периоди. Тези данни се получават в различен формат и времева рамка, за което е нужна допълнителна обработка от системата.

- **Замърсители във въздуха:** Получаваме информация за различни замърсители във въздуха, включително частици (ФПЧ2.5 и ФПЧ10), азотен диоксид (NO₂), серен диоксид (SO₂) и въглероден оксид (CO). Всеки замърсител допринася за общия индекс на качеството на въздуха.
- **Данни за трафик по пътищата:** Информация за трафика по пътищата в реално време ни помага да разберем как емисиите от автомобилите се влияят на качеството на въздуха. Този източник на данни ни помага да идентифицираме точките със задръствания и времевото изменение на емисиите.
- **Температура:** Знанието за температурата е от съществено значение, тъй като влияе на разпространението и концентрацията на замърсители във въздуха. Често по време на зимния отоплителен сезон се наблюдават по-високи нива на замърсяване, както и температурни инверсии.
- **Влажност:** Нивата на влажност влияят на качеството на въздуха, тъй като оказват влияние върху концентрацията и разпространението на частиците и замърсителите във въздуха. Високата влажност може да подобри качеството на въздуха, докато ниската влажност може да засилва проблемите обвързани със запращаване. Влажността е също

важен фактор при сравняване и обработка на данни за замърсяване идващи от различни измервателни уреди. Лазерните измерващи станции показват по-високи нива на грешка, когато има висока влажност.

- **Налягане на въздуха:** Данните за налягането на въздуха помагат при прогнозирането на метеорологични условия, които могат да влияят на качеството на въздуха. Това е важен фактор при моделирането на прогнози за замърсяване на въздуха.

5.3 Модул Съхранение на Данни в Облака

Система за наблюдение на качеството на въздуха и влиянието му върху здравето на хората се основава на интегриран подход, за събиране и съхранение данни. Ние събираме данни за замърсители във въздуха, движението по пътищата, температурата, влажността и налягането от различни източници, които не комуникират по между си и ги централизираме и съхраняваме в облаковата инфраструктура. Този подход осигурява скалируемост и гъвкавост, позволявайки ни да следим нарастващите изисквания към системата.

Чрез събирането и съхранението на данни за замърсители във въздуха, движението по пътищата и атмосферни данни на едно място, ние можем да предоставим обстойна информация за качеството на въздуха. Например интегрирайки данни за трафика с данни от стационарни измервателни станции се получава информация за емисиите от превозните средства и техния принос за замърсяването на въздуха. Като обединим данните за движението и измерванията на качеството на въздуха, ние можем да идентифицираме областите с висока замърсеност, произтичащи от трафика, и да изготвим стратегии за намаляване на замърсяването или неговото избягване.

Запазвайки данните от тези източници на едно място прави анализа и извличането на информация по-лесни, по-бързи и ефективни, предоставяйки обединена платформа за обработка на данни. Потребители с различни права имат лесен достъп до информацията от всяко място с интернет връзка. Данните се актуализират в реално време, което ни позволява да предоставяме актуална информация за качеството на въздуха на потребителите.

Сигурността на съхранението в облака е под строги стандарти, гарантирайки, че данните са поверителни и цялостни.

5.4 Модули за обработка, анализ и прогнозиране на данни

Модулът за Обработка на Данни е сърцето на системата и изпълнява няколко ключови функции. На първо място, този модул предоставя основна предподготовка на данните, включително почистване и структуриране. Това означава, че за да се гарантира тяхната точност и цялост, данните се хармонизират, уеднаквяват времевите диапазони, като добър пример за това една от системите граждански станции предава на 5 минути данните, а системата от официални станции - на кръгъл час. При почистването се извършва отстраняване на невалидни или повредени данни, а структурирането им в подходящ формат осигурява по-лесна навигация и анализ на информацията.

Пример за анализ на данните е оценката за на риска за здравето чрез ИКВ от 'суровите' данни за замърсяване. ИКВ е стандартизиран инструмент, предназначен да измерва и комуникира общото качество на въздуха в определен район. Основната му цел е да предложи проста и ясна представа за текущите нива на замърсяване на въздуха, като направи данните леснодостъпни за широката общественост. ИКВ преобразува сложни екологични измервания в едно число или цветово-кодирана категория, варираща от „добро“ до „опасно“. Това дава възможност на хората да вземат информирани решения за своите дейности, особено на открито, въз основа на потенциалните здравни рискове от замърсяването на въздуха. ИКВ е особено полезен за уязвимите групи от населението, включително деца, възрастни хора и хора с предварително съществуващи респираторни или сърдечно-съдови заболявания, тъй като осигурява своевременни предупреждения, когато качеството на въздуха се влоши.

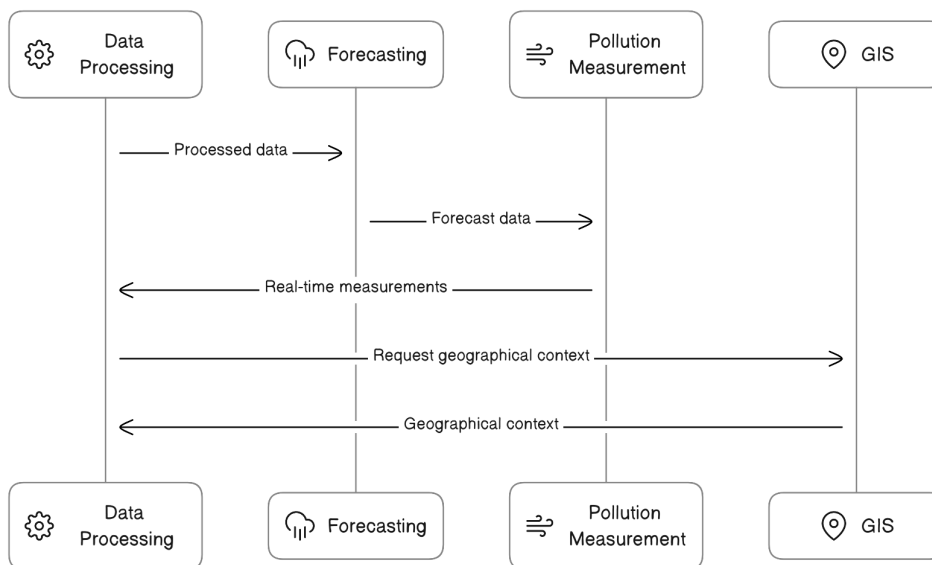
Изчисляването на ИКВ се основава на концентрациите на няколко основни замърсители на въздуха, за които е известно, че оказват значително въздействие върху човешкото здраве. Тези замърсители обикновено включват прахови частици (PM2.5 и PM10), приземния озон (O3), азотния диоксид (NO2), серния диоксид (SO2) и въглеродния оксид (CO). Станциите за мониторинг на качеството на въздуха измерват концентрацията на всеки

замърсител в атмосферата в реално време. ИКВ за всеки замърсител се изчислява отделно и най-високата от тези индивидуални стойности става общата стойност на ИКВ за съответното място. Този подход гарантира, че ИКВ отразява най-лошия възможен сценарий за общественото здраве в даден момент. Например, ако нивата на озона са особено високи в сравнение с другите замърсители, ИКВ ще отразява опасността от озона.

Модулът за прогнозиране е в процес на изграждане и трениране на моделите. Той използва обработените данни, за да генерира прогнози за качеството на въздуха. Целта на този модул е да използва различни алгоритми за анализ и моделиране, които включват методи за времеви редове, регресия, машинно обучение и други. Те се прилагат върху историческите данни и текущите условия, което позволява на системата да предскаже какви промени в качеството на въздуха може да се очакват в бъдеще. Резултатите от модула предоставят ценни предупреждения и прогнози, които са от съществено значение за регулаторните органи и обществото, които трябва да предприемат действия за опазване на общественото здраве.

Модулът за измерване на замърсяване от ФПЧ по зададено трасе предоставя уникален аспект на системата, като извършва измерване на замърсяването на въздуха в реално време в определени географски области. Този модул се възползва от ГИС данни, което позволява визуализацията на данните върху географска карта и ги свързва с географските данни за трасето. Това предоставя по-детайлна представа за качеството на въздуха в конкретни местоположения и подкрепя бърза реакция в случай на открити аномалии или проблеми.

Фигура 5.3 предоставя изчерпателен архитектурен преглед на системата за измерване на качеството на въздуха. Тя визуално показва сложния поток на данни между модулите за обработка на данни, прогнозиране и измерване на замърсяването. Диаграмата ясно демонстрира как данните се обработват, използват за прогнозиране и се измерват в реално време. Освен това тя подчертава интеграцията на географски контекст чрез данни от ГИС, което улеснява холистичното разбиране на функционалностите на системата.



Фигура 5.3: Архитектурна диаграма показваща потока на данни между модулите за обработка на данни

5.5 Използвани технологии

5.5.1 Език за програмиране Python

В този раздел ще се фокусираме върху причината за избор на Python и свързаните с него библиотеки за моделиране на данните в рамките на нашата софтуерна система за измерване на качеството на въздуха. Ще предоставим информация за това как Python и тези библиотеки се интегрират със системата и какви ключови функционалности предоставят.

Python се използва като основен програмен език за разработка на системата, осигурявайки гъвкавост и мощност в обработката на данни. Неговата лесна четимост и богатият набор от библиотеки го правят подходящ избор за обработка и анализ на данни.

Библиотеките за моделиране на данните, като калибриране и прогнозиране, предоставят възможност за изграждане и тестване на различни модели, които се използват в системата. Те позволяват използването на машинно обучение и статистически методи за анализ и прогнозиране на данните за качеството на въздуха.

Използването на Python и библиотеките за моделиране носи няколко значителни предимства за нашата софтуерна система. Python предоставя удобна среда за разработка и тестване на модели, като позволява бързо прототипиране и ефективно моделиране на данните.

Библиотеките за моделиране разширяват възможностите на Python, предоставяйки готови инструменти и алгоритми за калибриране, прогнозиране и анализ на данните. Това ни позволява да генерираме точни прогнози и предупреждения относно качеството на въздуха.

Python предоставя обширна и гъвкава стема от библиотеки, които са важни инструменти за обработка на данни, включително данни, получени от финансови и физически измервания. Тези библиотеки, като NumPy, pandas и SciPy, предоставят мощни инструменти за анализ, манипулиране и обработка на числови данни. NumPy, например, предоставя възможност за ефективни операции с многомерни масиви и матрици, което е от решаващо значение при работата с данни, получени от физически измервания, където структурата на данните е сложна и изисква бързо и ефективно обхождане.

Съчетаването на библиотеките pandas и SciPy предоставя възможност за анализ на данни, статистически обработки и моделиране, които са от съществено значение при интерпретацията на резултатите от финансови изследвания и физически измервания. С използването на тези библиотеки можем да извличаме ценни заключения и да подкрепим научните ни хипотези с данни.

TensorFlow е водеща библиотека за машинно обучение и дълбоко обучение с широко разпространение. Тя предоставя среда за създаване и обучение на невронни мрежи, което е от съществено значение при извличането на модели и прогнози от физически данни. Тези модели могат да помогнат за разкриване на тенденции, прогнозиране на бъдещи събития и подпомагане на научни изследвания и иновации. TensorFlow се използва в нашата система при двустъпковото калибриране на данни от лазерни станции, както и за прогноза на въздушното замърсяване.

Използването на Python и съответните библиотеки в този контекст предоставя мощен инструментариум за научни изследвания и анализ на данни, което допринася за по-доброто разбиране на сложни процеси и явления в областите на финансите и физическите науки. Това сътрудничество между програмния език и библиотеките за обработка на данни и машинно обучение спомага за постигане на по-добра научна продуктивност и развитие в

съответните области.

5.5.2 Рамката Django

Django предоставя мощна и устойчива рамка за разработка на уеб приложения. Построен върху Python, Django предоставя комплексни инструменти за бърза и удобна разработка на уеб базирани приложения. Важен аспект на този инструментариум е вградената обработка на данни и възможността за съхранение и обработка на информация в бази данни.

Django предоставя мощен и интуитивен интерфейс за работа с бази данни, позволяващ удобно създаване на модели за данни и лесно управление на информацията в тях. Този инструмент осигурява безпроблемна работа с различни бази данни и автоматично генерира SQL заявки, което улеснява разработката и оптимизира заявките към базите данни.

Съчетаването на Django с инструменти за обработка на данни, като NumPy и pandas, може да бъде изключително полезно при разработка на уеб приложения, които изискват манипулиране и анализ на данни. Интеграцията на тези инструменти помага за създаването на мощни уеб базирани приложения, които работят с данни от различни източници и осигуряват ценни аналитични възможности.

Освен това, Django предоставя възможности за изграждане на публични API, което е изключително полезно за създаването на уеб приложения, които предоставят достъп до данни чрез стандартизирани интерфейси. Това улеснява споделянето и интеграцията на данни между различни системи и приложения.

Използването на Django като основна рамка за разработка на уеб приложения, съчетан с инструменти за обработка на данни, предоставя силно оръдие за създаване на уеб приложения, които не само предоставят богата потребителска услуга, но и могат да обработват и анализират данни в реално време, като по този начин обогатяват функционалността и стойността на приложението.

5.5.3 GraphQL за взаимодействие с бази данни

GraphQL представлява мощен и гъвкав заявителен език за работа с данни, който се използва за взаимодействие с различните видове източници на данни, включително бази данни, уеб услуги и други. Чрез използването на

този подход към работа с данни осигуряваме голяма гъвкавост и контрол върху заявките за данни, позволявайки точното извличане на само тези данни, които са необходими за конкретната задача или приложение.

GraphQL позволява структура на данните, която да отговаря на специфичните изисквания на различен вид приложения. Това е особено полезно при работа с данни от различни източници и схеми, тъй като GraphQL позволява да се свържат и комбинират данни от различни източници чрез една обща точка на вход.

Съчетаването на GraphQL с инструменти за обработка на данни като Apollo Client предоставя надежден начин за взаимодействие с данни и предоставя богати възможности за обработка, анализ и предоставяне на информацията към приложенията. Този подход позволява създаването на динамични и високо-производителни приложения, които могат да отговорят на разнообразни потребителски заявки.

GraphQL не само опростява работата с данни, но и подпомага ефективното взаимодействие между клиентската и сървърната част на приложението, като предоставя оптимизирани заявки и отговори. Този абстрактен и гъвкав подход към работа с данни съчетан с инструменти за обработка на данни, като NumPy и pandas, може да бъде от съществено значение при създаването на приложения, които изискват сложни и динамични заявки към данни.

Използването на GraphQL ни помогна да създаваме приложения, които не само предоставят потребителско удовлетворение, но и улесняват извличането и анализа на данни, което е от съществено значение в днешния свят на информация и данни.

5.5.4 Използване на REST API

REST (Representational State Transfer) е архитектурен стил за създаване на уеб услуги и API, който предоставя сигурен и ефективен начин за комуникация между различни системи и приложения. REST API предоставят лек и прост начин за достъп до ресурси чрез обичайни HTTP заявки, като GET (за извличане на данни), POST (за създаване на нови данни), PUT (за актуализация на данни) и DELETE (за изтриване на данни).

REST API осигурява лесен начин за достъп и управление на данни чрез стандартизирани и интуитивни заявки, което го прави подходящ за интеграция в различни приложения и платформи. Този структуриран и

прозрачен начин на работа с данни осигурява гъвкавост и скалируемост на приложенията.

Съчетаването на REST API с инструменти за обработка на данни, като Python и библиотеките NumPy и pandas, позволява обработката и анализа на данни посредством заявки към API, което може да бъде от изключителна полза при създаването на аналитични и интелигентни приложения. REST API предоставят стандартизиран начин за извличане и обработка на информация, която след това може да бъде анализирана и предоставена в подходящия формат.

Този архитектурен стил не само опростява работата с данни, но и подпомага ефективната комуникация между различни части на системите и приложенията. REST API се използват широко и предоставят основата за взаимодействието между различни системи, което улеснява интеграцията и обмена на данни в света на софтуерната разработка.

5.5.5 Облачна инфраструктура

Облачният репозиториум служи като централно хранилище за данните, идващи от различни източници. Това включва информация за концентрации на замърсители, метеорологични данни, трафик и много други фактори, влияещи върху качеството на въздуха. Облачната инфраструктура осигурява сигурно съхранение и управление на огромни обеми данни.

Използването на облачна инфраструктура е ключово за обработката на големи обеми от данни и предоставянето на услуги, които изискват гъвкавост и скалируемост. Използвани са облачните платформи на Amazon Web Services (AWS) и Microsoft Azure, които предоставят ресурси и услуги за съхранение, обработка и анализ на данни в глобален мащаб. А в комбинация с Python, Django, GraphQL и REST API съвместно с облачната инфраструктура предоставяме инструменти за анализ на качеството на въздуха и влиянието му върху здравето.

Това става по следния начин: облачната инфраструктура предоставя място за съхранение на огромни обеми от данни за качеството на въздуха и метеорологични данни, които могат да бъдат лесно достъпни и обработвани, а с използването на Python и библиотеките му за обработка на данни като NumPy и pandas, данните могат да бъдат анализирани, обработени и подготвени за визуализация и моделиране. Облачната инфраструктура укрепва скалируемостта и надеждността на системата. Системата може

да приеме и обработи информация от различни източници, като запазва непрекъснат достъп и актуализации в реално време. Джанго предоставя удобен начин за създаване на уеб приложения, които могат да предоставят данни за качеството на въздуха на потребителите. Чрез интегриране на GraphQL и REST API в Django приложението, данните могат да бъдат предоставяни в реално време на потребителите чрез гъвкави и стандартизирани заявки. GraphQL позволява точно извличане на необходимата информация, като в същото време предоставя възможност за комбиниране на данни от различни източници. REST API осигуряват структуриран достъп до информацията.

Този синергичен подход позволява наблюдение на качеството на въздуха в реално време и анализиране на данните, което е от съществено значение за опазването на общественото здраве. Python и свързаните библиотеки съчетани с облачната инфраструктура и средата за уеб разработка предоставят средствата за създаване на интелигентни приложения, които предупреждават и информират гражданите за влиянието на качеството на въздуха в реално време. Този подход не само помага за подобряване на общественото здраве, но и подкрепя научните изследвания и решения, свързани с околната среда.

5.6 Примери за използване на софтуерната инфраструктура

В основата на функционалността на нашия софтуер стои способността му да интегрира данни от ГИС, измервания на качеството на въздуха от стационарни измервателни станции и данни за трафика. Тази интеграция се постига чрез усъвършенствани техники за сливане на данни, които ни позволяват да получим цялостна представа за качеството на въздуха в различни пространствени и времеви мащаби. Данните от ГИС осигуряват съществен географски контекст, който ни позволява да определим точно измерванията на качеството на въздуха. Чрез наслагване на данни за качеството на въздуха и трафика върху ГИС слоеве можем да идентифицираме пространствени модели, горещи точки и корелации, които иначе биха останали незабелязани. Тази интеграция е от решаващо значение за нашите изследвания, тъй като ни позволява да оценим въздействието на моделите на движение и географските характеристики върху нивата на

качеството на въздуха.

Софтуерът има способността му да визуализира информация за качеството на въздуха върху популярни платформи за картографиране, като Google Maps и OpenStreetMap. Тази функция трансформира необработените данни във формат, който може да се тълкува визуално, като улеснява разпространението на важна информация сред широка аудитория. Чрез наслагване на данни за качеството на въздуха върху карти, ние предоставяме на потребителите интуитивно и географски контекстуално представяне на нивата на качеството на въздуха. Можем да анализираме вариациите на качеството на въздуха във времето, като идентифицираме тенденции и аномалии, които да послужат като основа за нашите изследователски резултати. Този времеви анализ дава представа за това как се променя качеството на въздуха през различните часове на деня, сезони или при определени метеорологични условия.

Изобразяването на качеството на въздуха върху картите също така подпомага обществената осведоменост и информираност за проблема. То прави сложните данни достъпни за широка аудитория, като повишава осведомеността за качеството на въздуха и неговото въздействие върху здравето и околната среда. Освен това то подпомага усилията за реагиране при извънредни ситуации, като предоставя информация в реално време на съответните органи по време на кризи, свързани с качеството на въздуха, или екологични инциденти.

AirLief е мобилно приложение за качество на въздуха, разработено за платформите на iOS и Андроид. То е безплатно приложение с отворен достъп, като показва на потребителя колко е замърсен въздуха около него на база неговата геолокация. AirLief взима данните от софтуерната система, използвайки API-то, създадено за комуникация на данни извън системата.

Данни за състоянието на въздуха и атмосферни данни се разпространяват от софтуерната система към мобилното приложение чрез създадени за целта инфраструктурни крайни точки (ендпойнти). За лесно подаване на данни към нашата система влиза информация за локация (GPS координати), географски радиус или град и излизат данни за качеството на въздуха в реално време спрямо направената заявка. За приложението AirLief са направени 3 инфраструктурни крайни точки (ендпойнти):

- Измервания с най-близката точка (входни данни: GPS координати):

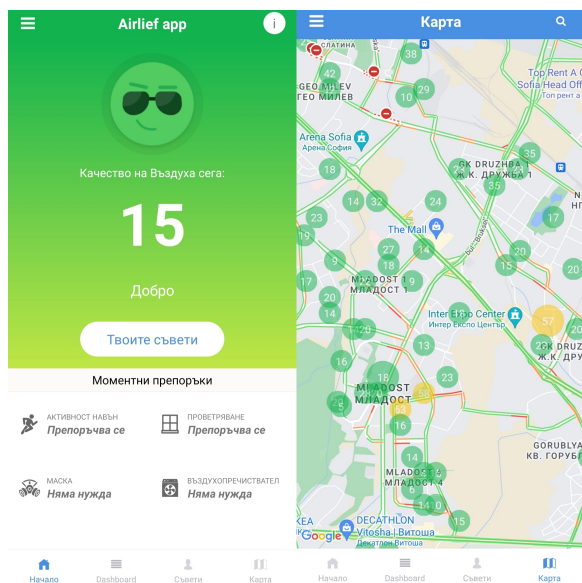
Този ендпойнт позволява потребителите да получат резултати от най-близката измервателна станция, разположена до мястото със зададените GPS координати. Това осигурява информация за текущото качество на въздуха в конкретната локация.

- Измервания по местоположение (входни данни: GPS координати, радиус в км): Този ендпойнт позволява на потребителите да получат средноаритметични стойности на измервания от всички измервателни станции в радиус от определени километри около зададените GPS координати. Това дава обобщена представа за качеството на въздуха в определен район.
- Измервания по град (входни данни: име на града): Този ендпойнт позволява на потребителите да получат средноаритметични стойности на измерванията от всички измервателни станции в определен град или населено място. Това предоставя информация за качеството на въздуха в рамките на конкретния град или населено място.

Тези инфраструктурни крайни точки предоставят гъвкавост и удобство при извличане на данни за замърсяването на въздуха и допринасят към обширната функционалност на софтуерната система за следене и управление на качеството на въздуха. В мобилното приложение са показани на потребителите по лесен и достъпен начин.

AirLief app получава данни и препредава по достъпен начин информация за въздушното замърсяване за над 100 държави от около 15000 станции. А самото приложение се използва в 70 държави от над 70000 потребителя, Фигура 5.4. То се състои от дашборд, който показва на потребителя по достъпен и лесен за разбиране начин моментното състояние на въздуха; карта, която показва стойностите от стационарните станции; възможност за доавяне на любими станции и съвети за персонално предпазване от въздушно замърсяване. На Фигура 5.5 е изобразено как самото мобилно приложение превръща суровите данни получени от софтуерната система в ИКВ. Индекс, който директно корелира с въздействието на въздушното замърсяване върху човешкия организъм, като в него влизат освен измерванията на ФПЧ и други, като озоново замърсяване, азотен диоксид и други.

На фигура 5.5 се вижда как приложението.



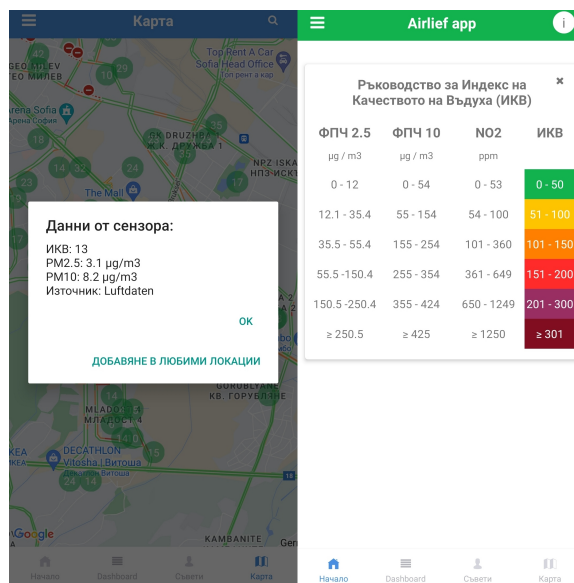
Фигура 5.4: AirLief App - едно от приложенията на разработената софтуерна система, което показва актуално въздушно замърсяване от стационарни сензори

Освен гореописаното мобилно приложение, други по-малки проекти използват софтуерната система с научни, некомерсиални цели. Екип от Португалия използва API от нас за корелция на назамните данни за качество на въздуха със сателитни данни по програмата на Европейската космическа агенция. Друг проект от Холандия използва API за пилотен проект за "умни пейки" които информират гражданите на Амстердам за качеството на въздуха в реално време.

5.7 Заключение

Нашият иновативен подход към измерването на качеството на въздуха и влиянието му върху здравето на хората използва съчетание от най-новите технологии и инструменти, които работят в синхрон с облачната инфраструктура и различни програмни елементи, за да предостави надеждни и реалнозначими данни.

Изборът на модулерна архитектура и използването на облачни техно-



Фигура 5.5: AirLief App - мобилното приложение, което превръща суровите данни в разбираем Индекс за Качеството на Въздуха (ИКВ)

логии ни осигурява интегриран и синхронизиран подход, позволяващ на системата ни да предоставя в реално време измерване на качеството на въздуха и да анализира данните с цел предупреждаване и информиране на гражданите за въздействието на въздушната среда върху тяхното здраве. Този комплексен подход не само подобрява общественото здраве, но и подкрепя научните изследвания и решения, свързани с околната среда.

Освен това внимателното калибриране и използване на входните параметри гарантират надеждността на нашата система за вземане на решения, което ѝ позволява да прави информирани прогнози и оценки по отношение на качеството на въздуха и неговото въздействие върху човешкото здраве.

Глава 6

Заклучение

В дисертационния труд е дефинирана цел и 4 задачи за нейното постигане. Първоначално е извършен детайлен анализ на данните за въздушното замърсяване и острите заболявания от 2 софийски болници и бърза помощ. Набелязани са ползи от комбиниране на данните за замърсяване от официални измервателни станции и лазерни сензорни станции (евтини IoT устройства, разпространени чрез граждани).

Дисертационният труд разработва и внедрява решение за калибриране на данни за ФПЧ от лазерни станции чрез машинно обучение, който изчисля грешките и аномалиите на база сравнение с данни от стандартизирана станция. В таблица 3.4 показва подобрението на резултатите от лазерни станции, където с използването на комбинация от изкуствени невронни мрежи и метод за детекция на аномалиите, квадратичната грешка от 0,62 преди калибриране се покачва на на 0,95 след калибриране.

Разработен е метод за избор на маршрут спрямо замърсяване от ФПЧ, който използва модифициран за целта . В таблица 4.5 се вижда, че в направения полеви тест с 10 колоездача, маршрутът от изработения софтуер показва с 67% по-добри резултат от други решения за намиране на велосипеден маршрут.

Всички тези задачи и решения са разработени в обща мащабируема софтуерна система с модулна структура за събиране и обработка на данни от различни станции и IoT устройства. Тази система допринася за ефективната работа с разнообразни хардуерни устройства за измерване на ФПЧ, които използват различни комуникационни протоколи. Практическите резултати от функционирането на системата се визуализират в потребителс-

ки интерфейс, който представлява точка на взаимодействие между потребителите и системата. Тази система има възможност както да обработва и калибрира данни, така и да генерира карти за въздушното замърсяване, като се използва от различни проекти, като например мобилно приложение, достъпвано от хиляди потребители седмично.

Научните и приложни изследвания, представени в дисертационния труд, дават следните приноси:

6.1 Научно-приложни приноси

1. Направен е статистически анализ за корелация между въздушното замърсяване и постъпленията в болници и бърза помощ, с които са определени важни причинно-следствени връзки.

2. Разработен е алгоритъм за калибриране на лазерни сензори чрез двустъпков метод с помощта на изкуствени невронни мрежи и детекция на аномалии. Този иновативен метод поддържа по-надеждни и точни измервания от лазерните сензори и има потенциал да бъде използван в широк спектър от сензорни приложения.

3. Разработен е алгоритъм, който оценява влиянието на влажността, височината и атмосферното налягане върху качеството на данните за въздушно замърсяване от лазерни сензори. Това позволява по-пълно и точно разбиране на факторите, влияещи върху замърсяването на въздуха и може да бъде приложим и във вертикално планиране.

4. Разработен е алгоритъм за изчисление на най-подходящ път за преминаване на велосипедист, основавайки се на концентрацията на ФПЧ. Този метод на маршрутизация подпомага здравословния начин на живот и насърчава устойчивия начин на придвижване в градовете.

6.2 Приложни приноси

1. Разработен е софтуер за калибриране на данните за ФПЧ от лазерни сензори, взимайки реперни данни от официалните измервателни станции на МОСВ. Този софтуер позволява на организациите за контрол на качеството на въздуха да използват по-надеждни данни за вземане на решения и управление на замърсяването.

2. Разработен е софтуер, който отчита влиянието на влажността, височината и атмосферното налягане и подобрява данните за ФПЧ от лазерни сензори. Този софтуер може да бъде използван от градските власти и здравни организации, за да наблюдават и управляват качеството на въздуха в реално време.

3. Разработен е софтуер за намиране на най-подходящ път за преминаване на велосипедист базирайки се на инхалацията на ФПЧ. Този софтуер може да бъде полезен за индивидуални велосипедисти и градски инфраструктури, които насърчават велосипедното движение.

4. Разработена е IoT платформа за агрегация и анализиране на сензорни данни за качество на въздуха. Този софтуер подпомага автоматизацията и улеснява интеграцията на данни от различни източници, подобрявайки ефективността и точността на наблюдението за качеството на въздуха. Освен метеорологични данни интегрирани са и данни за трафик и GIS. Платформата визуализира карти, като ги допълва с измервания от сензорите за качество на въздуха и данни за настоящо състояние на трафика.

С оглед на изследванията и резултатите, представени в този дисертационен труд, можем да заключим, че този проект носи важни научни и приложни приноси, които се отразяват на развитието и подобряването на системите за измерване на въздушното замърсяване и съответните технологии.

6.3 Публикации по темата на дисертационния труд

- Zhivkov, P., & Simidchiev, A. (2024). Software Tool for Optimizing Cycling Route by Defining Cyclist Air Pollution Exposure Studies in Computational Intelligence, 1158 SCI, pp. 152–170, SJR 0.208 (2023)

<https://www.scopus.com/record/display.uri?eid=2-s2.0-85190651936&origin=resultslist&sort=plf-f&src=s&sid=3eca39362dd1584d77f2fb4c5901e3b2&sot=aut&sdt=a&sl=18&s=AU-ID%2857221769514%29&relpos=0&citeCnt=0&searchTerm=>

- Zhivkov, P., & Simidchiev, A. (2022). Development of Software Tool for Optimization and Evaluation of Cycling Routes by Characterizing Cyclist Exposure to Air Pollution Annals of Computer Science and Information Systems, 32, 105-112.

https://d1wqtxts1xzle7.cloudfront.net/99417790/230-libre.pdf?1677962458=&response-content-disposition=inline%3B+filename%3DDevelopment_of_Software_Tool_for_Optimiz.pdf&Expires=1725292377&Signature=QB4TSFkNK4~abulkZUnpTN5Hj4TAjM9gMEF~6w-SnXjxPm4iR~R0vvbDioIeFVGoHobKRbEI2ntm90Rc_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

- Zhivkov, P. Optimization and Evaluation of Calibration for Low-cost Air Quality Sensors: Supervised and Unsupervised Machine Learning Models Annals of Computer Science and Information Systems, Volume 25, FedCSIS 2021, 2021, pp. 255–258

<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85117793346&doi=10.15439%2f2021F95&partnerID=40&md5=2f0e02da6b0d78aaa6523df4f48308c9>

- Zhivkov, P., & Simidchiev, A. (2020). Relationship between Particulate Matter and Health Indicators for Acute Morbidity in Sofia. In Proceeding of 1st international Conference on Environmental Protection and Disaster Risks (pp. 180-193).

<https://www.ceeol.com/search/chapter-detail?id=926651>

6.4 Декларация за оригиналност

Декларирам, че дисертационният труд съдържа оригинални резултати, получени, при проведени от мен, научни изследвания с подкрепата и съдействието на научния ми ръководител.

Резултатите, които са получени, описани и/или публикувани от други учени, са коректно и подробно цитирани в библиографията.

Настоящият дисертационен труд не е прилаган за придобиване на научна степен в друго висше училище, университет или научен институт.

Подпис:

Библиография

- [1] Norfazillah Ab Manan, Azimatun Noor Aizuddin, and Rozita Hod. Effect of air pollution and hospital admission: a systematic review. *Annals of global health*, 84(4):670, 2018.
- [2] Stacey E Alexeeff, Noelle S Liao, Xi Liu, Stephen K Van Den Eeden, and Stephen Sidney. Long-term pm_{2.5} exposure and risks of ischemic heart disease and stroke events: review and meta-analysis. *Journal of the American Heart Association*, 10(1):e016890, 2021.
- [3] Emma Barnes and Marc Schlossberg. Improving cyclist and pedestrian environment while maintaining vehicle throughput: Before-and after-construction analysis. *Transportation research record*, 2393(1):85–94, 2013.
- [4] Hannah Bast, Daniel Delling, Andrew Goldberg, Matthias Müller-Hannemann, Thomas Pajor, Peter Sanders, Dorothea Wagner, and Renato F Werneck. Route planning in transportation networks. In *Algorithm engineering*, pages 19–80. Springer, 2016.
- [5] Jörg Becker, Daniel Beverungen, Martin Winter, and Sebastian Menne. Umwidmung und weiterverwendung von traktionsbatterien. *Szenarien, Dienstleistungen und Entscheidungsunterstützung*, 2019.
- [6] M.L. Bell, J.M. Samet, and F. Dominici. Time-series studies of particulate matter. *Annu. Rev. Public Health*, 25:247–280, 2004.
- [7] Alexander Y Bigazzi and Miguel A Figliozzi. Review of urban bicyclists’ intake and uptake of traffic-related air pollution. *Transport Reviews*, 34(2):221–245, 2014.

- [8] Jennifer Bosen, Hanna E Fuchte, and Carmen Leicht-Scholten. Cycling to work and making cycling work: What makes committed utility cyclists despite perceived risks of air pollution and traffic? *Journal of Transport & Health*, 28:101519, 2023.
- [9] HL Brantley, GSW Hagler, ES Kimbrough, RW Williams, S Mukerjee, and LM Neas. Mobile air monitoring data-processing strategies and effects on spatial air pollution trends. *Atmospheric measurement techniques*, 7(7):2169–2183, 2014.
- [10] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [11] Iain M Carey, Richard W Atkinson, Andrew J Kent, Tjeerd Van Staa, Derek G Cook, and H Ross Anderson. Mortality associations with long-term exposure to outdoor air pollution in a national english cohort. *American journal of respiratory and critical care medicine*, 187(11):1226–1233, 2013.
- [12] N. Castell et al. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environment international*, 99:293–302, 2017.
- [13] J.W. Cherrie et al. Effectiveness of face masks used to protect beijing residents against particulate air pollution. *Occupational and environmental medicine*, 75(6):446–452, 2018.
- [14] Jessie Yao Chin, Tobias Steinle, Thomas Wehler, Daniel Dregely, Thomas Weiss, Vladimir I Belotelov, Bernd Stritzker, and Harald Giessen. Nonreciprocal plasmonics enables giant enhancement of thin-film faraday rotation. *Nature communications*, 4(1):1–6, 2013.
- [15] Theodoros Chondrogiannis, Panagiotis Bouros, Johann Gamper, and Ulf Leser. Alternative routing: k-shortest paths with limited overlap. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 1–4, 2015.
- [16] Theodoros Chondrogiannis, Panagiotis Bouros, Johann Gamper, and Ulf Leser. Exact and approximate algorithms for finding k-shortest paths with limited overlap. In *20th International Conference on Extending Database Technology: EDBT 2017*, pages 414–425, 2017.

- [17] L. Clancy et al. Effect of air-pollution control on death rates in dublin, ireland: an intervention study. *The lancet*, 360(9341):1210–1214, 2002.
- [18] Aaron J Cohen, Michael Brauer, Richard Burnett, H Ross Anderson, Joseph Frostad, Kara Estep, Kalpana Balakrishnan, Bert Brunekreef, Lalit Dandona, Rakhi Dandona, et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the global burden of diseases study 2015. *The Lancet*, 389(10082):1907–1918, 2017.
- [19] Ashley Collier-Oxandale, Joanna Gordon Casey, Ricardo Piedrahita, John Ortega, Hannah Halliday, Jill Johnston, and Michael P Hannigan. Assessing a low-cost methane sensor quantification system for use in complex rural and urban environments. *Atmospheric measurement techniques*, 11(6):3569–3594, 2018.
- [20] Ellen M Considine, Colleen E Reid, Michael R Ogletree, and Timothy Dye. Improving accuracy of air pollution exposure measurements: Statistical correction of a municipal low-cost airborne particulate matter sensor network. *Environmental Pollution*, 268:115833, 2021.
- [21] Cyrille Médard de Chardon. The contradictions of bike-share benefits, purposes and outcomes. *Transportation research part A: policy and practice*, 121:401–419, 2019.
- [22] Audrey De Nazelle, Olivier Bode, and Juan Pablo Orjuela. Comparison of air pollution exposures in active vs. passive travel modes in european cities: A quantitative review. *Environment international*, 99:151–160, 2017.
- [23] Audrey De Nazelle, Mark J Nieuwenhuijsen, Josep M Antó, Michael Brauer, David Briggs, Charlotte Braun-Fahrlander, Nick Cavill, Ashley R Cooper, Hélène Desqueyroux, Scott Fruin, et al. Improving health through policies that promote active travel: a review of evidence to support integrated health impact assessment. *Environment international*, 37(4):766–777, 2011.
- [24] CE Delft. Health costs of air pollution in european cities and the linkage with transport. *CE Delft: Delft, The Netherlands*, 2020.

- [25] Inês D Do Vale, Ana S Vasconcelos, and Gonçalo O Duarte. Inhalation of particulate matter in three different routes for the same od pair: A case study with pedestrians in the city of lisbon. *Journal of Transport & Health*, 2(4):474–482, 2015.
- [26] F. Dominici, L. Sheppard, and M. Clyde. Health effects of air pollution: a statistical review. *International Statistical Review*, 71(2):243–276, 2003.
- [27] Jiyuan Dong, Yurong Liu, and Hairong Bao. Revalue associations of short-term exposure to air pollution with respiratory hospital admissions in lanzhou, china after the control and treatment of current pollution. *International Journal of Hygiene and Environmental Health*, 231:113658, 2021.
- [28] Mitch J Duncan, Hannah M Badland, and W Kerry Mummery. Applying gps to enhance understanding of transport-related physical activity. *Journal of Science and Medicine in Sport*, 12(5):549–556, 2009.
- [29] Alexander Geiß, Matthias Wiegner, Boris Bonn, Klaus Schäfer, Renate Forkel, Erika von Schneidemesser, Christoph Münkel, Ka Lok Chan, and Rainer Nothard. Mixing layer height as an indicator for urban air quality? *Atmospheric Measurement Techniques*, 10(8):2969–2988, 2017.
- [30] Muriel Gevrey, Ioannis Dimopoulos, and Sovan Lek. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological modelling*, 160(3):249–264, 2003.
- [31] M. Gochfeld and J. Burger. Disproportionate exposures in environmental justice and other populations: the importance of outliers. *American Journal of Public Health*, 101(S1):S53–S63, 2011.
- [32] Stefan Gössling, Andy Choi, Kaely Dekker, and Daniel Metzler. The social cost of automobility, cycling and walking in the european union. *Ecological Economics*, 158:65–74, 2019.
- [33] T. Guan et al. The effects of facemasks on airway inflammation and endothelial dysfunction in healthy young adults: a double-blind, randomized, controlled crossover study. *Particle and fibre toxicology*, 15(1):30, 2018.

- [34] A. Gábelová et al. Genotoxicity of environmental air pollution in three european cities: Prague, košice and sofia. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 563(1):49–59, 2004.
- [35] David H Hagan, Gabriel Isaacman-VanWertz, Jonathan P Franklin, Lisa MM Wallace, Benjamin D Kocar, Colette L Heald, and Jesse H Kroll. Calibration and assessment of electrochemical air quality sensors by co-location with regulatory-grade instruments. *Atmospheric Measurement Techniques*, 11(1):315–328, 2018.
- [36] Walter Ham, Abhilash Vijayan, Nico Schulte, and Jorn D Herner. Commuter exposure to pm_{2.5}, bc, and ufp in six common transport microenvironments in sacramento, california. *Atmospheric Environment*, 167:335–345, 2017.
- [37] Ghassan B Hamra, Neela Guha, Aaron Cohen, Francine Laden, Ole Raaschou-Nielsen, Jonathan M Samet, Paolo Vineis, Francesco Forastiere, Paulo Saldiva, Takashi Yorifuji, et al. Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. *Environmental health perspectives*, 2014.
- [38] Marissa Hauptman, Jonathan M Gaffin, Carter R Petty, William J Sheehan, Peggy S Lai, Brent Coull, Diane R Gold, and Wanda Phipatanakul. Proximity to major roadways and asthma symptoms in the school inner-city asthma study. *Journal of Allergy and Clinical Immunology*, 145(1):119–126, 2020.
- [39] A.J. Hedley et al. Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in hong kong: an intervention study. *The lancet*, 360(9346):1646–1652, 2002.
- [40] Mario A Hernández, Omar Ramírez, John A Benavides, and Juan F Franco. Urban cycling and air quality: Characterizing cyclist exposure to particulate-related pollution. *Urban Climate*, 36:100767, 2021.
- [41] David M Holstius, A Pillarisetti, KR Smith, and EJAMT Seto. Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in california. *Atmospheric Measurement Techniques*, 7(4):1121–1131, 2014.

- [42] Jalil Jaafari, Kazem Naddafi, Masud Yunesian, Ramin Nabizadeh, Mohammad Sadegh Hassanvand, Mansour Shamsipour, Mohammad Ghanbari Ghosikali, Hamid Reza Shamsollahi, Shahrokh Nazmara, and Kamyar Yaghmaeian. The acute effects of short term exposure to particulate matter from natural and anthropogenic sources on inflammation and coagulation markers in healthy young adults. *Science of The Total Environment*, 735:139417, 2020.
- [43] NAH Janssen, P Fischer, M Marra, C Ameling, and FR Cassee. Short-term effects of pm_{2.5}, pm₁₀ and pm_{2.5-10} on daily mortality in the netherlands. *Science of the Total Environment*, 463:20–26, 2013.
- [44] Rohan Jayaratne, Xiaoting Liu, Phong Thai, Matthew Dunbabin, and Lidia Morawska. The influence of humidity on the performance of a low-cost air particle mass sensor and the effect of atmospheric fog. *Atmospheric Measurement Techniques*, 11(8):4883–4890, 2018.
- [45] Wan Jiao, Gayle Hagler, Ronald Williams, Robert Sharpe, Ryan Brown, Daniel Garver, Robert Judge, Motria Caudill, Joshua Rickard, Michael Davis, et al. Community air sensor network (cairsense) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern united states. *Atmospheric Measurement Techniques*, 9(11):5281–5292, 2016.
- [46] Marilena Kampa and Elias Castanas. Human health effects of air pollution. *Environmental pollution*, 151(2):362–367, 2008.
- [47] K. Katsouyanni et al. Short term effects of air pollution on health: a european approach using epidemiologic time series data: the apha protocol. *Journal of Epidemiology & Community Health*, 50(Suppl 1):S12–S18, 1996.
- [48] Christine M Kendrick, Adam Moore, Ashley Haire, Alexander Bigazzi, Miguel Figliozzi, Christopher M Monsere, and Linda George. Impact of bicycle lane characteristics on exposure of bicyclists to traffic-related particulate matter. *Transportation research record*, 2247(1):24–32, 2011.
- [49] Kirsten A Koehler and Thomas M Peters. New methods for personal exposure monitoring for airborne particles. *Current environmental health reports*, 2(4):399–411, 2015.

- [50] Suk Wah Kwok and Chris Carter. Multiple decision trees. In *Machine Intelligence and Pattern Recognition*, volume 9, pages 327–335. Elsevier, 1990.
- [51] O. Laurent et al. Air pollution, asthma attacks, and socioeconomic deprivation: a small-area case-crossover study. *American journal of epidemiology*, 168(1):58–65, 2008.
- [52] P. Li et al. The acute effects of fine particles on respiratory mortality and morbidity in beijing, 2004–2009. *Environmental Science and Pollution Research*, 20(9):6433–6444, 2013.
- [53] Xiaoting Liu, Rohan Jayaratne, Phong Thai, Tara Kuhn, Isak Zing, Bryce Christensen, Riki Lamont, Matthew Dunbabin, Sicong Zhu, Jian Gao, et al. Low-cost sensors as an alternative for long-term air quality monitoring. *Environmental research*, 185:109438, 2020.
- [54] Feng Lu, Dongqun Xu, Yibin Cheng, Shaoxia Dong, Chao Guo, Xue Jiang, and Xiaoying Zheng. Systematic review and meta-analysis of the adverse health effects of ambient pm2. 5 and pm10 pollution in the chinese population. *Environmental research*, 136:196–204, 2015.
- [55] Lisha Luo, Yunquan Zhang, Junfeng Jiang, Hanghang Luan, Chuanhua Yu, Peihong Nan, Bin Luo, and Mao You. Short-term effects of ambient air pollution on hospitalization for respiratory disease in taiyuan, china: a time-series analysis. *International journal of environmental research and public health*, 15(10):2160, 2018.
- [56] Kamal Jyoti Maji, Wei-Feng Ye, Mohit Arora, and SM Shiva Nagendra. Pm2. 5-related health and economic loss assessment for 338 chinese cities. *Environment international*, 121:392–403, 2018.
- [57] Ernesto QV Martins and Marta Pascoal. A new implementation of yen’s ranking loopless paths algorithm. *Quarterly Journal of the Belgian, French and Italian Operations Research Societies*, 1(2):121–133, 2003.
- [58] MI Mead, OAM Popoola, GB Stewart, Peter Landshoff, M Calleja, M Hayes, JJ Baldovi, MW McLeod, TF Hodgson, J Dicks, et al. The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmospheric Environment*, 70:186–203, 2013.

- [59] Charles N Mock, Rachel Nugent, Olive Kobusingye, and Kirk R Smith. Disease control priorities, 2017.
- [60] L. Morawska et al. Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone? *Environment international*, 116:286–299, 2018.
- [61] Petros Mouzourides, Prashant Kumar, and Marina K-A Neophytou. Assessment of long-term measurements of particulate matter and gaseous pollutants in south-east mediterranean. *Atmospheric Environment*, 107:148–165, 2015.
- [62] Arideep Mukherjee and Madhoolika Agrawal. World air particulate matter: sources, distribution and health effects. *Environmental Chemistry Letters*, 15(2):283–309, 2017.
- [63] Christopher JL Murray. The global burden of disease study at 30 years. *Nature medicine*, 28(10):2019–2026, 2022.
- [64] BS Murthy, R Latha, Arpit Tiwari, Aditi Rathod, Siddhartha Singh, and G Beig. Impact of mixing layer height on air quality in winter. *Journal of Atmospheric and Solar-Terrestrial Physics*, 197:105157, 2020.
- [65] National Statistical Institute of Bulgaria (NSI). Population by cities <https://www.nsi.bg/en/content/2981/population-towns-and-sex>, 2019.
- [66] World Health Organization. *International classification of diseases:[9th] ninth revision, basic tabulation list with alphabetic index*. 1978.
- [67] World Health Organization. International classification of diseases (icd) information sheet. 2018.
- [68] Luc Int Panis, Bas de Geus, Grégory Vandenbulcke, Hanny Willems, Bart Degraeuwe, Nico Bleux, Vinit Mishra, Isabelle Thomas, and Romain Meeusen. Exposure to particulate matter in traffic: a comparison of cyclists and car passengers. *Atmospheric Environment*, 44(19):2263–2270, 2010.
- [69] Hye-Youn Park, Susan Gilbreath, and Edward Barakatt. Respiratory outcomes of ultrafine particulate matter (ufpm) as a surrogate measure

- of near-roadway exposures among bicyclists. *Environmental Health*, 16(1):1–7, 2017.
- [70] S. Pattenden, B. Nikiforov, and B. Armstrong. Mortality and temperature in sofia and london. *Journal of Epidemiology & Community Health*, 57(8):628–633, 2003.
- [71] C.A. Pope, D.W. Dockery, and J. Schwartz. Review of epidemiological evidence of health effects of particulate air pollution. *Inhalation toxicology*, 7(1):1–18, 1995.
- [72] Michael Pratt, Olga L Sarmiento, Felipe Montes, David Ogilvie, Bess H Marcus, Lilian G Perez, Ross C Brownson, Lancet Physical Activity Series Working Group, et al. The implications of megatrends in information and communication technology and transportation for changes in global physical activity. *The Lancet*, 380(9838):282–293, 2012.
- [73] Xiaoliang Qin, Lujian Hou, Jian Gao, and Shuchun Si. The evaluation and optimization of calibration methods for low-cost particulate matter sensors: Inter-comparison between fixed and mobile methods. *Science of The Total Environment*, 715:136791, 2020.
- [74] Aakash C Rai, Prashant Kumar, Francesco Pilla, Andreas N Skouloudis, Silvana Di Sabatino, Carlo Ratti, Ansar Yasar, and David Rickerby. End-user perspective of low-cost sensors for outdoor air pollution monitoring. *Science of The Total Environment*, 607:691–705, 2017.
- [75] Sanjay Rajagopalan, Sadeer G Al-Kindi, and Robert D Brook. Air pollution and cardiovascular disease: Jacc state-of-the-art review. *Journal of the American College of Cardiology*, 72(17):2054–2070, 2018.
- [76] Abdul Rachman Rasyid, Netra P Bhandary, and Ryuichi Yatabe. Performance of frequency ratio and logistic regression model in creating gis based landslides susceptibility map at lompobattang mountain, indonesia. *Geoenvironmental Disasters*, 3(1):1–16, 2016.
- [77] Stefan Reis, Tomáš Liška, Massimo Vieno, Edward J Carnell, Rachel Beck, Tom Clemens, Ulrike Dragosits, Samuel J Tomlinson, David Leaver, and Mathew R Heal. The influence of residential and workday

- population mobility on exposure to air pollution in the uk. *Environment international*, 121:803–813, 2018.
- [78] A.M. Rendón et al. Effects of urbanization on the temperature inversion breakup in a mountain valley with implications for air quality. *Journal of Applied Meteorology and Climatology*, 53(4):840–858, 2014.
- [79] S Rasoul Safavian and David Landgrebe. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, 21(3):660–674, 1991.
- [80] J. Samet and D. Krewski. Health effects associated with exposure to ambient air pollution. *Journal of toxicology and environmental health, Part A*, 70:227–242, 2007.
- [81] J.M. Samet et al. The national morbidity, mortality, and air pollution study. part ii: morbidity and mortality from air pollution in the united states. *Res Rep Health Eff Inst*, 94(pt 2):5–79, 2000.
- [82] David Savage, Xiuzhen Zhang, Xinghuo Yu, Pauline Chou, and Qingmai Wang. Anomaly detection in online social networks. *Social Networks*, 39:62–70, 2014.
- [83] J. Shi et al. Cardiovascular benefits of wearing particulate-filtering respirators: a randomized crossover trial. *Environmental health perspectives*, 125(2):175–180, 2017.
- [84] S. Shin et al. Ambient air pollution and the risk of atrial fibrillation and stroke: A population-based cohort study. *Environ Health Perspect*, 127(8):87009, 2019.
- [85] L Spinelle, M Aleixandre, and M Gerboles. Protocol of evaluation and calibration of low-cost gas sensors for the monitoring of air pollution. *Publication Office of the European Union, Luxembourg*, 2013.
- [86] L Sun, J Wei, DH Duan, YM Guo, DX Yang, C Jia, and XT Mi. Impact of land-use and land-cover change on urban air quality in representative cities of china. *Journal of Atmospheric and Solar-Terrestrial Physics*, 142:43–54, 2016.

- [87] L.A. Tallon et al. Cognitive impacts of ambient air pollution in the national social health and aging project (nshap) cohort. *Environ Int*, 104:102–109, 2017.
- [88] Yan Tao, Shengquan Mi, Shuhong Zhou, Shigong Wang, and Xiaoyun Xie. Air pollution and hospital admissions for respiratory diseases in lanzhou, china. *Environmental pollution*, 185:196–201, 2014.
- [89] G.D. Thurston et al. A joint ers/ats policy statement: what constitutes an adverse health effect of air pollution? an analytical framework. *European Respiratory Journal*, 49(1), 2017.
- [90] H. Traboulsi et al. Inhaled pollutants: the molecular scene behind respiratory and systemic diseases associated with ultrafine particulate matter. *International Journal of Molecular Sciences*, 18(2):243, 2017.
- [91] J. Vermeylen et al. Ambient air pollution and acute myocardial infarction. *Journal of Thrombosis and Haemostasis*, 3(9):1955–1961, 2005.
- [92] W. Wang et al. Particulate air pollution and ischemic stroke hospitalization: How the associations vary by constituents in shanghai, china. *Sci Total Environ*, 695:133780, 2019.
- [93] Yang Wang, Jiayu Li, He Jing, Qiang Zhang, Jingkun Jiang, and Pratim Biswas. Laboratory evaluation and calibration of three low-cost particle sensors for particulate matter measurement. *Aerosol Science and Technology*, 49(11):1063–1077, 2015.
- [94] Scott Weichenthal, Ryan Kulka, Aimee Dubeau, Christina Martin, Daniel Wang, and Robert Dales. Traffic-related air pollution and acute changes in heart rate variability and respiratory function in urban cyclists. *Environmental health perspectives*, 119(10):1373–1378, 2011.
- [95] Lingkun Wu, Xiaokui Xiao, Dingxiong Deng, Gao Cong, Andy Diwen Zhu, and Shuigeng Zhou. Shortest path and distance queries on road networks: An experimental evaluation. *arXiv preprint arXiv:1201.6564*, 2012.
- [96] Sherrie Xie and Blanca E Himes. Personal environmental monitoring. In *Precision in Pulmonary, Critical Care, and Sleep Medicine*, pages 305–320. Springer, 2020.

- [97] Sherrie Xie, Jessica R Meeker, Luzmercy Perez, Whitney Eriksen, Anna Localio, Hami Park, Alicia Jen, Madison Goldstein, Akua F Temeng, Sarai M Morales, et al. Feasibility and acceptability of monitoring personal air pollution exposure with sensors for asthma self-management. *Asthma research and practice*, 7(1):1–11, 2021.
- [98] Antonella Zanobetti, Francesca Dominici, Yun Wang, and Joel D Schwartz. A national case-crossover analysis of the short-term effect of pm_{2.5} on hospitalizations and mortality in subjects with diabetes and neurological disorders. *Environmental Health*, 13(1):1–11, 2014.
- [99] R. Zhang et al. Acute effects of particulate air pollution on ischemic stroke and hemorrhagic stroke mortality. *Front Neurol*, 9:827, 2018.
- [100] Bo Zheng, Dan Tong, Meng Li, Fei Liu, Chaopeng Hong, Guannan Geng, Haiyan Li, Xin Li, Liqun Peng, Ji Qi, et al. Trends in china’s anthropogenic emissions since 2010 as the consequence of clean air actions. *Atmospheric Chemistry and Physics*, 18(19):14095–14111, 2018.
- [101] Petar Zhivkov. Optimization and evaluation of calibration for low-cost air quality sensors: Supervised and unsupervised machine learning models. In *2021 16th Conference on Computer Science and Intelligence Systems (FedCSIS)*, pages 255–258. IEEE, 2021.
- [102] Petar Zhivkov and Aleksandar Simidchiev. Relationship between particulate matter and health indicators for acute morbidity in sofia. In *Proceeding of 1st international conference on environmental protection and disaster risks*, pages 180–193, 2020.
- [103] Petar Zhivkov and Alexander Simidchiev. Development of software tool for optimization and evaluation of cycling routes by characterizing cyclist exposure to air pollution. *Annals of Computer Science and Information Systems*, 32:105–112, 2022.
- [104] Petar Zhivkov and Alexander Simidchiev. Software tool for optimizing cycling route by defining cyclist air pollution exposure. In *The Workshop on Computational Optimization*, pages 152–170. Springer, 2024.
- [105] Nadezda Zikova, Mauro Masiol, David C Chalupa, David Q Rich, Andrea R Ferro, and Philip K Hopke. Estimating hourly concentrations

- of pm_{2.5} across a metropolitan area using low-cost particle monitors. *Sensors*, 17(8):1922, 2017.
- [106] N. Zimmerman et al. Closing the gap on lower cost air quality monitoring: Machine learning calibration models to improve low-cost sensor performance. *Atmos. Meas. Tech. Discuss*, pages 1–36, 2017.
- [107] Marina Zusman, Cooper S Schumacher, Amanda J Gasset, Elizabeth W Spalt, Elena Austin, Timothy V Larson, Graeme Carvlin, Edmund Seto, Joel D Kaufman, and Lianne Sheppard. Calibration of low-cost particulate matter sensors: Model development for a multi-city epidemiological study. *Environment international*, 134:105329, 2020.
- [108] Moniek Zuurbier, Gerard Hoek, Marieke Oldenwening, Virissa Lenters, Kees Meliefste, Peter Van Den Hazel, and Bert Brunekreef. Commuters' exposure to particulate matter air pollution is affected by mode of transport, fuel type, and route. *Environmental health perspectives*, 118(6):783–789, 2010.