

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

**Петър Руменов Живков**

**МОДЕЛИРАНЕ ВЛИЯНИЕТО НА  
КАЧЕСТВОТО НА АТМОСФЕРНИЯ ВЪЗДУХ  
ВЪРХУ ЗДРАВНИ И ИКОНИМИЧЕСКИ АСПЕКТИ**

**АВТОРЕФЕРАТ**

за придобиване на  
образователната и научна степен „доктор“  
в професионално направление  
4.6. Информатика и компютърни науки

научен ръководител  
проф. д.н. Стефка Фиданова

2024



# Глава 1

## Въведение

Замърсяването на въздуха и по-конкретно фините прахови частици (ФПЧ), е водещата екологична причина за заболяемост и преждевременна смъртност, след като през 1990 г.е било на пета позиция [48]. Смъртните случаи, причинени от ФПЧ в околната среда, са нараснали през последните 25 години [15], а излагането на по-високо замърсяване на атмосферния въздух е свързано с по-висока смъртност [9]. А последният доклад за оцен

ФПЧ се класифицират въз основа на техния размер, като частиците с размери по-малки от 10 микрометра (ФПЧ<sub>10</sub>) и 2,5 микрометра (ФПЧ<sub>2,5</sub>) 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<sub>2,5</sub> и ФПЧ<sub>10</sub> е свързано с ранна смъртност от няколко заболявания, като сърдечносъдови, респираторни, рак на белия дроб и инфекции на горните и долните дихателни пътища [2], [28],[32]. 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) [43], а епидемиологични проучвания показват, че излагането на ФПЧ увеличава риска от диабет [74].

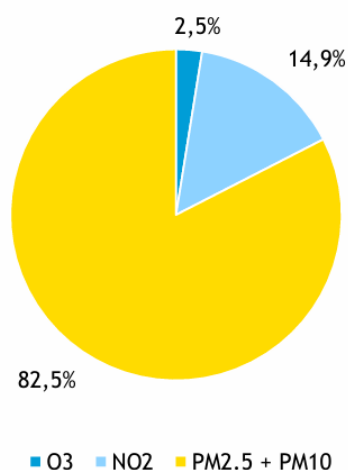
Тежестта на последиците за здравето може да варира в зависимост от концентрацията и продължителността на излагане, както и от здравословното състояние на индивида и неговата чувствителност към замърсителите. Някои групи от населението, като децата, възрастните хора и тези с вече съществуващи респираторни или сърдечносъдови заболявания, са особено уязвими към здравните ефекти от замърсяването на въздуха [58]

Изследвания показват силна връзка между замърсяването на въздуха и хоспитализациите за респираторни заболявания [66, 21, 44]. А кохортен преглед на 175 статии обобщава, че излагането на мръсен въздух 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 заболяемост въз основа на данните за качеството на въздуха [45]. 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 ФПЧ<sub>2,5</sub> 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: ФПЧ, азотен диоксид и озон [18]. Резултатите показват, че ФПЧ 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.



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

В светлината на тези констатации, нараства необходимостта от модели, които да могат да изследват заболяемостта 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. Съществуващите 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 България е лидер по загубени години на живот в добро здраве, заради мръсен въздух. Фигура 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](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 разработва 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 острият 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 заболяемост 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 характеризирание 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.

В процеса на изследване на проблема е използван широк сфера от литература. Създадените нови методи и алгоритми са публикувани в [[?], [76], [77], [78]]. Представената методология позволява цялостен анализ на въздействието на ФПЧ върху острите заболявания, калибриране на гражданските лазерни станции чрез машинно обучение и невронни мрежи и оптимизиране на велосипедни маршрути с оглед на замърсяването от ФПЧ. В съответствие с поставените задачи са разработени алгоритми и методологии решаващи конкретните проблеми. Направена е софтуерно приложение на всеки от разработените алгоритми и визуализиране на данните чрез IoT платформа. Платформата и софтуърът са написани на езика Python и използват рамката Django.

## Глава 2

# Връзка между замърсяването с ФПЧ и здравни индикатори за остра заболяемост

Връзките между градското въздушно замърсяване (ВЗ) и човешкото здраве са последователно и ясно установени от редица изследователи [67], [60], [41], като открояващи групи заболявания са сърдечно-съдови [69], невровакуларни [65], [11] и белодробни [68]. Според скорошно усъвършенствано моделиране се предполага, че има близо 9 милиона смъртни случая годишно от ВЗ [55]. 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 [20]. Повечето от тези данни идват от анализи на времеви серии [24], сравняващи вариациите в хоспитализацията със средните вариации на прахови частици [29]. ВЗ е отговорно за възпаленията на дихателните пътища [14] и като се има предвид, че дихателната система е обща врата за влизане, минимизирането на експозицията на дихателните пътища също минимизира сърдечно-съдовите проблеми [63]. Освен това съществуват проучвания, които моделират замърсяването на множество градове, като Европейското замърсяване на въздуха и здравето: европейски подход (ARNEA) [36] и Американското национално изследване на заболяемостта, смъртността и замърсяването на въздуха (NMMAPS) [61], 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 [23]. 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.

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

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

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

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.

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  $50 \mu\text{g}/\text{m}^3$ , ЕС и СЗО се различават значително в годишните си средни препоръки, като стандартите на СЗО са двойно по-строги (20  $\mu\text{g}/\text{m}^3$ , срещу 40  $\mu\text{g}/\text{m}^3$  от ЕС). За ФПЧ2,5, ЕС нямат концентрационен праг за среднодневни 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.

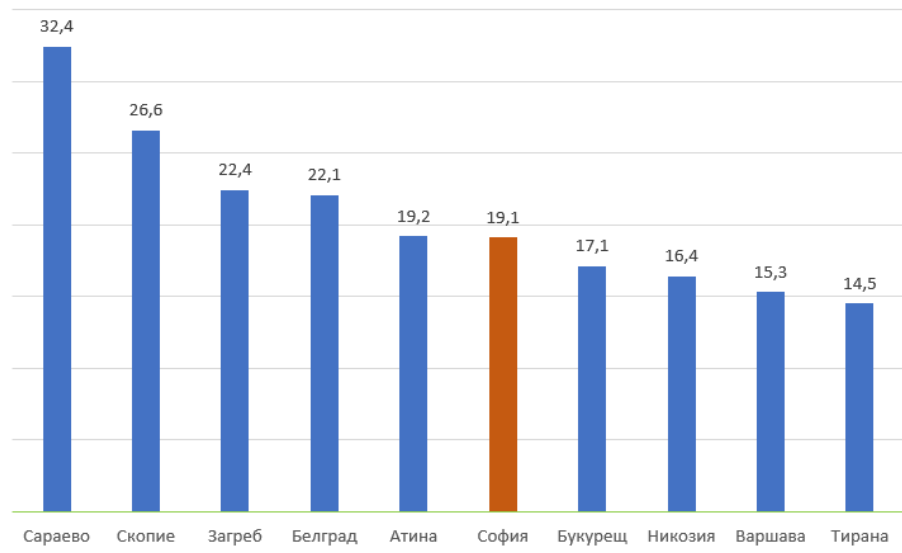
В последния си доклад СЗО полага специално внимание на замърсяването от ФПЧ и смята излагането на ФПЧ 2,5 за най-големия убиец свързан с въздуха в световен план. Около 80 % от смъртните случаи в световен мащаб, дължащи се на излагане на ФПЧ2,5, могат да бъдат предотвратени, ако страните спазват годишна 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 застоини метеорологични условия. Градът е

с население от 1,2 милиона души [49] 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 ФПЧ<sub>2,5</sub> 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.



Фигура 2.1: Средна концентрация на ФПЧ<sub>2,5</sub> в най-замърсените столици в Европа през 2022 г. (в микрограма на кубичен метър въздух)

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 [54]. 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 [25].

## 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 бърза помощ; 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<sub>2,5</sub> и ФПЧ<sub>10</sub> 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 ФПЧ<sub>10</sub> и ФПЧ<sub>2,5</sub> 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 [10]. Освен това евтините станции могат да надградят официалните измервателни уреди и по-специално ограничено количество данни и разпространение на станции [79]. 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 [47]. 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 диагностики 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 заболяемост. 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 [50],[51].

## 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 Манн – 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 [35]. 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 3.1 е илюстриран методът експозиция-ефект, където дозата е експозицията, умножена по времето.

## 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 [33]. 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/m<sup>3</sup> с диапазон от 2 до 136 g/m<sup>3</sup>. 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/m<sup>3</sup> за ФПЧ10 и 25 g/m<sup>3</sup> за ФПЧ2,5). В данните 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

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

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.

От анализирани 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 [75], [64], [70].

Но най-висока корелация се забелязва при острите инфекции на горните дихателни пътища се увеличават съответно с 47% и в частност пневмониите 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 (таблица 2.5).

## 2.5 Заключение

В резултат на проведените анализи и тестове се доказва, че острата заболеваемост 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,

Таблица 2.5: Сравнение на непараметрични данни от амбулаторни прегледи в болници с нормите на СЗО за среднодневна концентрация на ФПЧ10

Забол.	ICD-10	без отг.		отг. 1 ден		отг. 2 дни		отг. 3 дни	
		Z	p	Z	p	Z	p	Z	p
Дихателна система	C30-C39	-1.515	0.130	-2.950	0.003	-2.941	0.003	-2.423	0.015
Ушни заболявания	H65,H81	-0.754	0.451	-0.976	0.329	-1.580	0.114	-2.322	0.020
Миокарден инфаркт	I20-I25	-0.244	0.807	-1.393	0.164	-1.876	0.061	-2.104	0.035
Сърдечна недост.	I50	-0.771	0.440	-2.840	0.005	-2.559	0.010	-2.972	0.003
Горни дих. пътища	J00-J06	-3.218	0.001	-3.507	0.000	-3.910	0.000	-4.605	0.000
Бактериална пневм.	J13-J18	-3.171	0.002	-3.395	0.001	-2.556	0.011	-3.683	0.000
Остър бронхит	J20-J21	-3.096	0.002	-2.848	0.004	-1.606	0.108	-1.588	0.112
Долни дих. пътища	J40-J47	-3.988	0.000	-4.049	0.000	-2.842	0.004	-3.481	0.001
Инфекции на кожата	L00-L08	-2.552	0.011	-1.379	0.168	-0.267	0.789	-0.489	0.625
Микоплазмена пневм.	J20.0	-2.330	0.020	-2.327	0.020	-0.430	0.667	-0.431	0.667
Остър бронхит	J20.9	-2.496	0.013	-1.449	0.147	-0.039	0.969	-1.292	0.197
Астма, алерг.	J45.0	-2.426	0.015	-2.808	0.005	-3.577	0.000	-3.643	0.000
Ангина	I20.8	-1.971	0.049	-0.763	0.445	-0.773	0.440	-0.674	0.500

свързани със замърсяването на въздуха, като сърдечно-съдови и мозъчно-съдови, също се увеличават.

Това е сред първите подобни проучвания в България и може да бъде полезно за специалистите в сферата на здравеопазването, учените по околната среда и политиците.

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 по-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 [40].

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

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 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. 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 съответни-

те 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.

### 3.1 Наблюдение на качеството на въздуха

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.

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 [38].

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.

### 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 [16].

### 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 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.

### 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 ИАОС. 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 РМ10 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 подобрене 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 [31], [80].

### 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 [59]. GBDT е итеративен алгоритъм за дървета на решенията, съставен от множество дървета на решенията, при който решенията на всички дървета се сумират, за да се получи крайният отговор [39]. 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 [8]. Изкуствените невронни мрежи са математически модел, който симулира поведението на невроните и се променя автоматично чрез обратно разпространение на грешките [22].

### 3.2.5 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 фините 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.

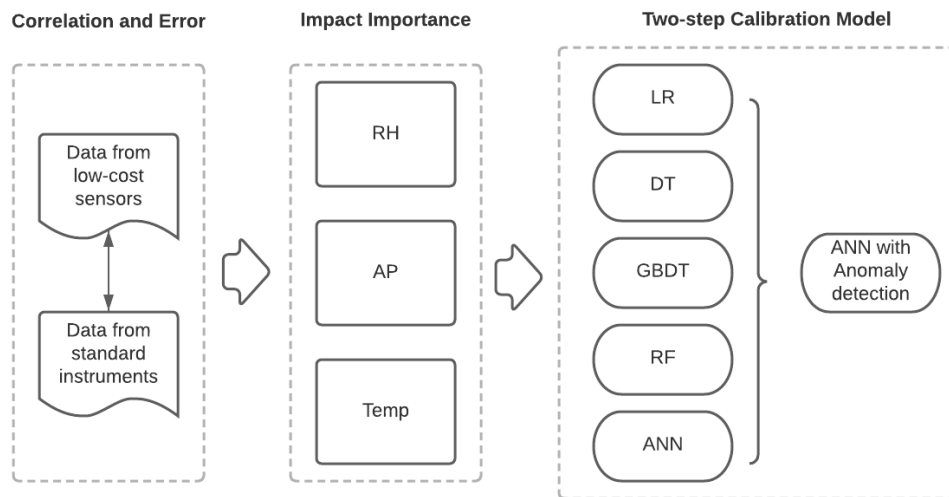
Моделът е обучен върху набор от данни, обхващащ период от три месеца, като входните данни представляват почасови средни стойности на измерванията. Преди обучението всички входни характеристики се нормализират, за да се ускори процесът на конвергенция и да се осигури стабилност на обучението. Като функция 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.

### 3.2.6 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 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 [62]. 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.



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

### 3.3 Приложение на модела върху безжична сенсорна 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.

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 захранването); 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.

Безжичната сенсорна 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

Таблица 3.1: 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

Таблица 3.2: Важността на параметрите въз основа на RF.

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

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.

### 3.4 Резултати и оценка

В Таблица 3.1 са показани резултатите от петте контролирани ML техники. За по-добра оценка на евтините сензори заедно с коефициента на детерминация ( $R$ -квадратично) са изчислени MAE и 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 [57], [56].

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 [26], [31], [34].

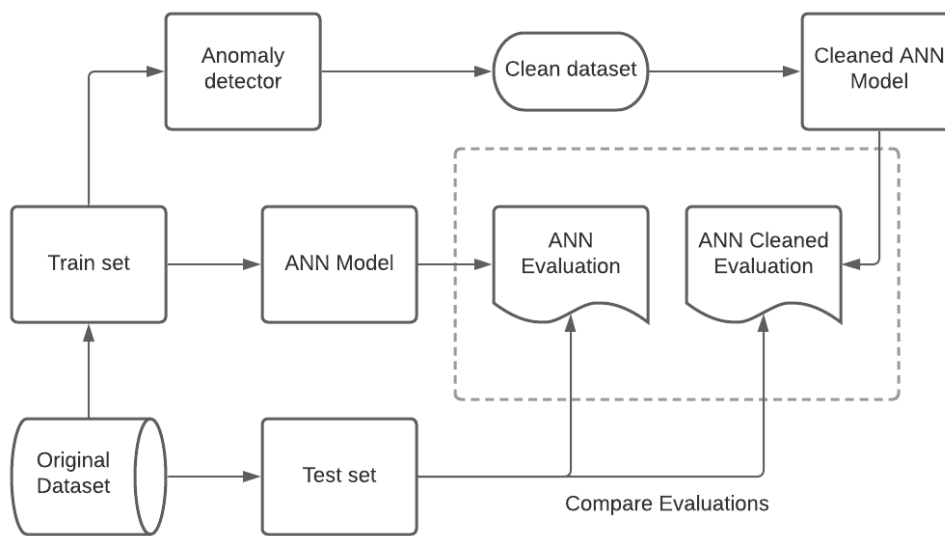
#### 3.4.1 Оценка на резултатите за относителната влажност

В таблица 3.2 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 [42]. 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 съотношението ФПЧ2,5/ФПЧ10, тъй като, както беше споменато по-горе, от петте стандартни уреда

Таблица 3.3: Сравнение на едностъпков и двустъпков модел.

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



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

само един измерва ФПЧ2,5.

### 3.4.2 Резултати от модела за калибриране

В Таблица 3.2 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.

### 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.



## Глава 4

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

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 [81], [27] и високите нива на физическа активност [7], [53]. Изследователите са се съсредоточили върху оценката на действителните нива на експозиция на велосипедистите по предварително избрани маршрути с помощта на персонални пробовземачи [17] 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 [52]. 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 [71]. Едно изследване дори установява, че велосипедистите абсорбират по-голяма част от фините частици ФПЧ<sub>2,5</sub> и черния въглерод, отколкото водачите на моторизирани видове транспорт [30].

### 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 качество 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.

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.

## 4.2 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.

За да изчислим VR на велосипедиста (в l/min), използваме уравнението на модела [19], 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 [19], 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  $\text{ФПЧ}_{inh}$  ( $\mu\text{g}$ ) е масата на замърсителите, които попадат в дихателните пътища на велосипедистите по време на пътуването (в двете посоки);  $\text{ФПЧ}_{conc}$  ( $\mu\text{g}/\text{m}^3$ ) е средната експозиция на замърсителите.

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  $\text{ФПЧ}_{inh}$ ; изисква по-малко усилия. HR

се увеличава по време на изкачване и висока скорост. Търсетe маршрути с ниска денивелация. Преминува се през по-малки концентрации на ФПЧ; пътищата с интензивно движение трябва да се пропускат. Предпочитат се малки улици и алеи в паркове.

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.

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], [72]. 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 [46].

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 [12]. 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 [13]. 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  $\theta$ , 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  $\theta$ ;

всички  $k$  маршрута са възможно най-кратки.

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.

#### 4.2.1 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, Openaq и много други. Той получава, записва, почиства и калибрира данните за качеството на въздуха от фиксирани евтини сензори.

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 [76], който използва изкуствени невронни мрежи и откриване на аномалии.

#### 4.2.2 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 [73].

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..

#### 4.2.3 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; маршрута са двойно несходни по отношение на прага на сходство  $\theta$ ; 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 намират 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.

#### 4.2.4 Полеви тестове и проучвани маршрути

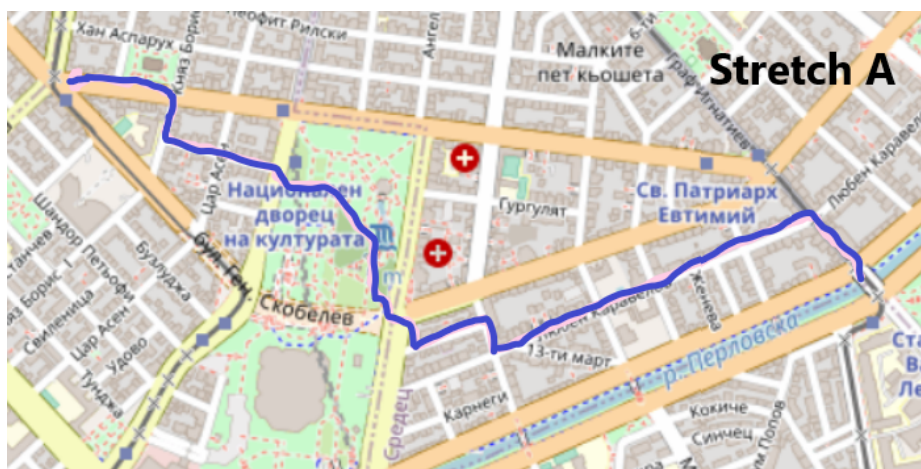
В тази част се разглежда оценката на софтуерният модел с реално теренно проучване в София и разглеждане на два велосипедни маршрута. Двата проучвани маршрута започват от националния стадион "Васил Левски" и завършват в Петте къшета. Тези две места се намират в центъра на София и имат активен велосипеден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

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 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 погълниха



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

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 тестовите.

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

## 4.3 Резултати и обсъждане

### 4.3.1 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



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

$\mu\text{g} / \text{m}^3$ ) 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 (ЛТ)

	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
Участък А	3	7	4	4	11	7
Участък 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.

#### 4.3.2 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.

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

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

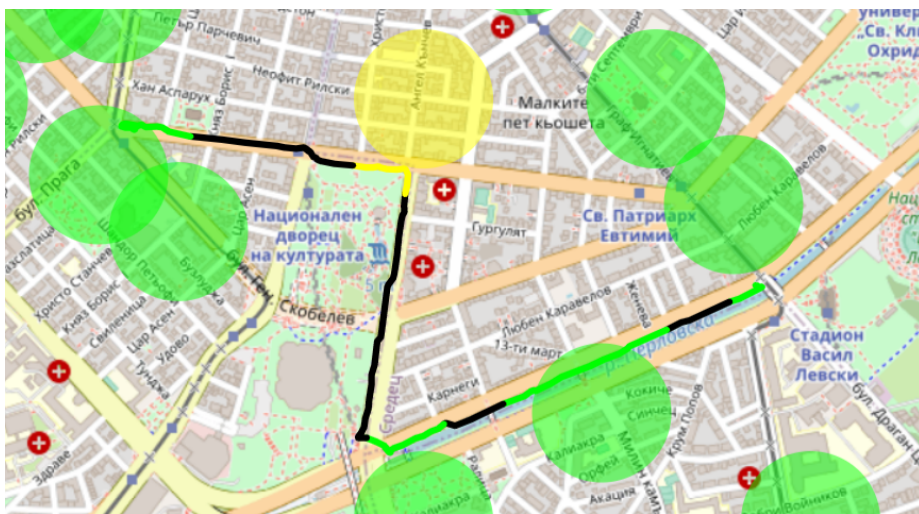
Период	Замърсител	Участък А (оптимизиран)	Участък Б (велосипедни алеи)
Интензивен трафик	ФПЧ 1 <sub>inh</sub>	29.74	46.45 (↑56%)
	ФПЧ2.5 <sub>inh</sub>	40.56	63.04 (↑55%)
Слаб трафик	ФПЧ 1 <sub>inh</sub>	27,04	39,82 (↑47%)
	ФПЧ2.5 <sub>inh</sub>	35.15	49.77 (↑42%)
Неработни дни	ФПЧ 1 <sub>inh</sub>	10.82	13.27 (↑23%)
	ФПЧ2.5 <sub>inh</sub>	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 SpO<sub>2</sub> 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 [56], [15].

### 4.3.3 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. На фиг.4.3 могат да се видят концентрациите за участък Б по време на ЛТ, където на участъка се прилагат измервания от стационарни сензори в близост до 200 м или по-близо до маршрута. Черно оцветената линия означава, че в тази част от маршрута няма фиксиран сензор, който да е по-близо от 200 м, докато зелените и жълтите цветни линии представят концентрациите, измерени от фиксираните сензори в близост. За стойностите на концентрацията на ФПЧ2,5 между 0 и 12 - се използва зелен цвят, докато жълтият цвят означава стойности на ФПЧ2,5 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. Благодарение на гъстата мрежа от евтини сензори в София, за участък А се използват 5 фиксирани сензора, а за участък Б - 6 сензора, тъй като те отговарят на критериите за подбор.



Фигура 4.3: Маршрут за колоездене с експозиция от фиксирани сензори

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.

#### 4.3.4 Прецизиране на модела и оценка на данните от мобилни и стационарни сензори

Средните стойности за концентрациите на ФПЧ<sub>2,5</sub> от стационарните сензори, които са разположени в близост до изследваните маршрути (разстояние 200 м), са сравнени с измерванията на ФПЧ<sub>2,5</sub> от мобилните сензори, както е показано в Таблица 4.6 по-долу. 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 фиксираните 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 фиксираните 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.

## 4.4 Заключение

В тази глава е представена разработката на софтуерен инструмент, който оптимизира велосипедните маршрути въз основа на алгоритми, които предвиждат най-малко

Таблица 4.6: Концентрации на ФПЧ<sub>2,5</sub> от стационарни сензори, разположени в близост до маршрута, и мобилни измервания

Период	Маршрут	$2,5_{fixed}$		$2,5_{mobile}$	
		mean	min	max	mean
WD-HT	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

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 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 ФПЧ<sub>1</sub> и ФПЧ<sub>2,5</sub> по двата маршрута.

Концентрациите на експозиция по специализираните велосипедни алеи се оказваха 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 ФПЧ<sub>2,5</sub> се намалява с 23%, в неработни дни до 55% по време на интензивен трафик 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. 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.

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. Съществува значителна разлика между изследванията със слаб трафик и с интензивен трафик. Велоалелите без физически бариери между велосипедния маршрут и пътя имат по-висока експозиция - заключение, което се установява в друго проучване [37]. Нашите констатации допринасят за по-доброто разбиране на проблемите на София, свързани със замърсяването от мръсен въздух и в частност от автомобилен трафик.

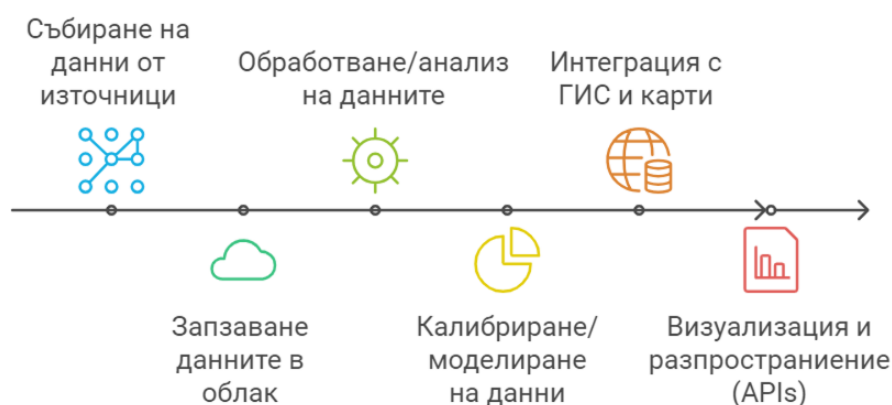


## Глава 5

# Софтуерна система за ВИЗУАЛИЗАЦИЯ НА ВЪЗДУШНОТО ЗАМЪРСЯВАНЕ

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 позволява събирането, обработката и моделирането на данни.

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. Архитектурата на софтуера се състои от следните компоненти: модул за свързване с API; Централен репозиториум в облака; модул за обработка на данни; модул за прогнозиране; модул за измерване на замърсяване от ФПЧ по зададено трасе; визуализация и разпространение на данни.



Фигура 5.1: Софтуерна архитектура

Осигуряването на точността и надеждността на нашата система за вземане на ре-

шения 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 температурата и влажността от стационарни сензори; данни за трафика.

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 по между си в една обща, позволява на нашата система да прави информирани прогнози и оценки на заърсяването и съответно да взима по-надеждни решения.

Исходните данни включват: 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.

## 5.1 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 (NO<sub>2</sub>), серен диоксид (SO<sub>2</sub>) 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 времето 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 запрашаване. 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.

## 5.2 Модул Съхранение на данни в облака

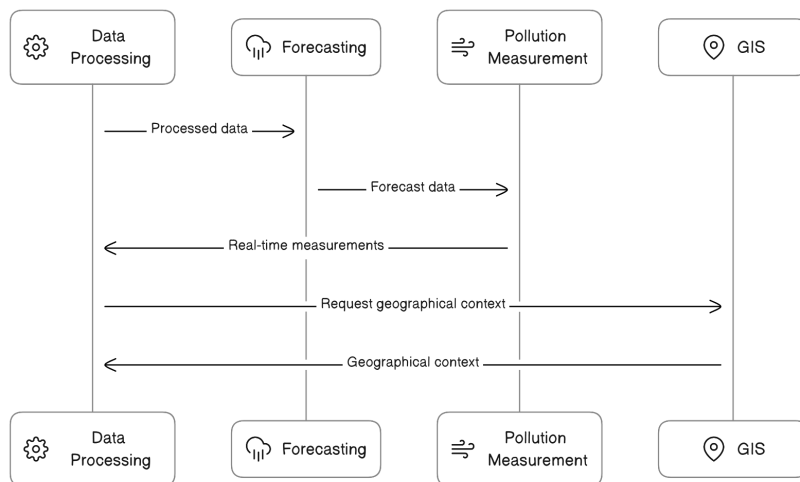
Системата за наблюдение на качеството на въздуха и влиянието му върху здравето на хората се основава на интегриран подход, който включва събиране и обработка на разнообразни данни. Ние събираме данни за замърсители във въздуха, движението по пътищата, температурата, влажността и налягането от различни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

Чрез събирането на данни за замърсители във въздуха, движението по пътищата, температурата, влажността и налягането, ние можем да предоставим обстойна представа за качеството на въздуха. Този подход не само засилва нашето разбиране за качеството на въздуха, но и ни позволява да вземем информирани решения и да предприемем своевременни мерки за защита на общественото здраве и околната среда. Данните за трафика са интегрирани, тъй като те предлагат информация за емисиите от превозните средства и тяхното въздействие върху качеството на въздуха. Чрез комбиниране на данните за трафика с измерванията на качеството на въздуха можем да определим областите с високо замърсяване, дължащо се на задръствания, и да разработим стратегии за намаляване на замърсяването.

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

## 5.3 Модул Анализ, прогнозиране и измерване на въздушно замърсяване

Модулът за Обработка на Данни е сърцето на системата и изпълнява няколко ключови функции. Фигура 5.2 предоставя изчерпателен архитектурен преглед на системата за измерване на качеството на въздуха. Тя визуално обрисова 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. На първо място, този модул предоставя основна предподготовка на данните, включително почистване и структуриране. Това означава, че данните, получени от различни източници и API-та, се подлагат на анализ и обработка, за да се гарантира тяхната точност и цялост. При почистването се извършва отстраняване на невалидни или повредени данни, а структурирането им в подходящ формат осигурява по-лесна навигация и анализ на информацията. Важно е да се подчертае, че модулът включва двустъпков метод за калибриране с машино обучение, който увеличава точността на данните и ги подготвя за анализ.



Фигура 5.2: Архитектурна диаграма, показваща потока на данни между модулите за обработка на данни

## 5.4 Използвани технологии

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.

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.

Python предоставя обширна и гъвкава схема от библиотеки, които са важни инструменти за обработка на данни. 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.

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 мощна и устойчива рамка за разработка на уеб приложения, подходяща за целите на софтуера.

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.

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 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.

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.

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.

## 5.5 Пример 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 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. 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.

Интегрирането на тези разнообразни източници на данни в съгласувана рамка ни дава възможност да направим значими заключения за качеството на въздуха и взаимодействието му с околната среда, като допринася за дълбочината и всеобхватността на нашите изследвания.

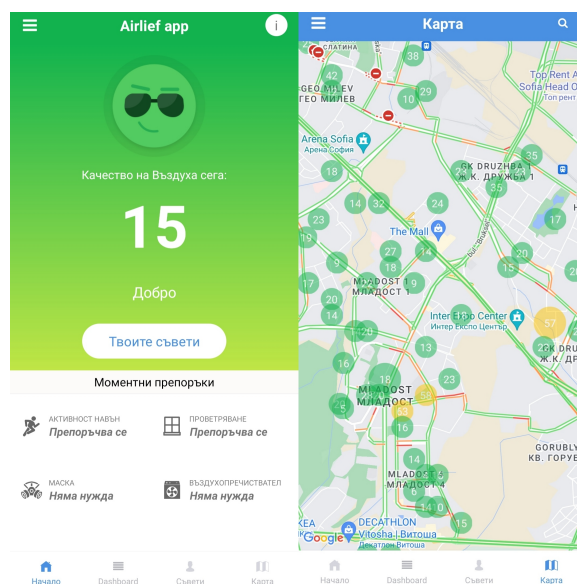
AirLief е мобилно приложение за качество на въздуха, разработено за платформите на iOS и Андроид. То е безплатно приложение с отворен достъп, като показва на потребителя колко е замърсен въздуха около него на база неговата геолокация. AirLief взима данните от софтуерната система, използвайки API-то,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

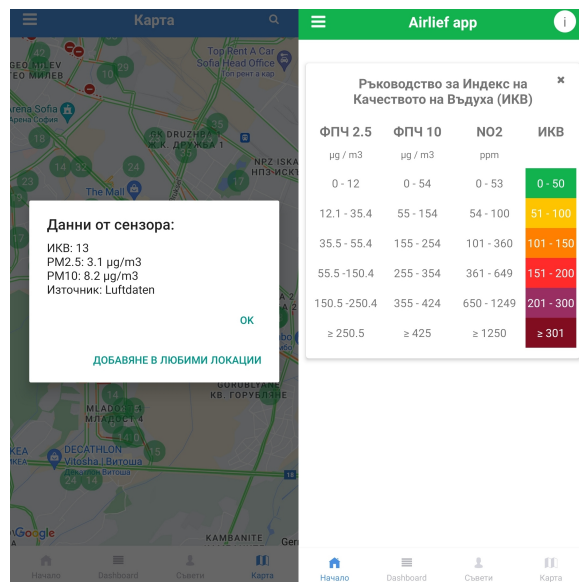
Тези инфраструктурни крайни точки предоставят гъвкавост и удобство при извличане на данни за замърсяването на въздуха и допринасят към обширната функционалност на софтуерната система за следене и управление на качеството на въздуха. В мобилното приложение са показани на потребителите по лесен и достъпен 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 5.3. 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 5.4 е изобразено как самото мобилно приложение превръща суровите данни получени от софтуерната система в Индекс за Качеството на Въздуха (ИКВ). Индекс, който директно корелира с въздействието на въздушното замърсяване върху човешкия организъм, като в него влизат освен измарванията на ФПЧ и други, като озоново замърсяване, азотен диоксид и други.



Фигура 5.3: AirLief App - едно от приложенията на разработената софтуерна система, което показва актуално въздушно замърсяване от стационарни сензори

На фигура 5.4 се вижда как приложението.



Фигура 5.4: AirLief App - мобилното приложение, което превръща суровите данни в разбираем Индекс за Качеството на Въздуха (ИКВ)

## Глава 6

# Заклучение

В дисертационния труд е дефинирана цел и 4 задачи за нейното постигане.

Дисертационният труд изследва връзката между въздушното замърсяване и острите заболявания, базирайки се на данни от две болници и бърза помощ в София. Разработено е решение за калибриране на данни от евтини IoT станции за измерване на ФПЧ чрез машинно обучение, което подобрява точността на измерванията. В тестовите с изкуствени невронни мрежи  $R^2$  се подобрява значително (от 0.62 на 0.95). Изработен е и метод за оптимизация на велосипедни маршрути според нивата на замърсяване, като тестовите показват 67% по-добри резултати спрямо алтернативни решения. Всички задачи са интегрирани в мащабируема софтуерна система за обработка и визуализация на данни, използвана и от мобилно приложение с хиляди потребители.

### 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.

## Приложни приноси

1. Разработен е софтуер за калибриране на данните за ФПЧ от лазерни сензори, вземайки 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.

## 6.1 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~abulkZUncpTN5Hj4TAjM9gME\\_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA](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~abulkZUncpTN5Hj4TAjM9gME_&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>

## 6.2 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<sub>2.5</sub> 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] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [9] 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.
- [10] N. Castell et al. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environment international*, 99:293–302, 2017.
- [11] 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.

- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] CE Delft. Health costs of air pollution in european cities and the linkage with transport. *CE Delft: Delft, The Netherlands*, 2020.
- [19] 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.
- [20] F. Dominici, L. Sheppard, and M. Clyde. Health effects of air pollution: a statistical review. *International Statistical Review*, 71(2):243–276, 2003.
- [21] 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.
- [22] 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.
- [23] 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.

- [24] 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.
- [25] 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.
- [26] 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.
- [27] Walter Ham, Abhilash Vijayan, Nico Schulte, and Jorn D Herner. Commuter exposure to pm<sub>2.5</sub>, bc, and ufp in six common transport microenvironments in sacramento, california. *Atmospheric Environment*, 167:335–345, 2017.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] David M Holstius, A Pillarissetti, 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.
- [32] Jalil Jaafari, Kazem Naddafi, Masud Yunesian, Ramin Nabizadeh, Mohammad Sadegh Hassanvand, Mansour Shamsipour, Mohammad Ghanbari Ghoskhal, 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.
- [33] 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.
- [34] 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.

- [35] Marilena Kampa and Elias Castanas. Human health effects of air pollution. *Environmental pollution*, 151(2):362–367, 2008.
- [36] K. Katsouyanni et al. Short term effects of air pollution on health: a european approach using epidemiologic time series data: the aphea protocol. *Journal of Epidemiology & Community Health*, 50(Suppl 1):S12–S18, 1996.
- [37] 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.
- [38] 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.
- [39] Suk Wah Kwok and Chris Carter. Multiple decision trees. In *Machine Intelligence and Pattern Recognition*, volume 9, pages 327–335. Elsevier, 1990.
- [40] 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.
- [41] 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.
- [42] 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.
- [43] 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 pm<sub>2.5</sub> and pm<sub>10</sub> pollution in the chinese population. *Environmental research*, 136:196–204, 2015.
- [44] 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.
- [45] Kamal Jyoti Maji, Wei-Feng Ye, Mohit Arora, and SM Shiva Nagendra. Pm<sub>2.5</sub>-related health and economic loss assessment for 338 chinese cities. *Environment international*, 121:392–403, 2018.
- [46] 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.
- [47] 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.

- [48] Christopher JL Murray. The global burden of disease study at 30 years. *Nature medicine*, 28(10):2019–2026, 2022.
- [49] National Statistical Institute of Bulgaria (NSI). Population by cities <https://www.nsi.bg/en/content/2981/population-towns-and-sex>, 2019.
- [50] World Health Organization. *International classification of diseases:[9th] ninth revision, basic tabulation list with alphabetic index*. 1978.
- [51] World Health Organization. International classification of diseases (icd) information sheet. 2018.
- [52] 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.
- [53] 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.
- [54] S. Pattenden, B. Nikiforov, and B. Armstrong. Mortality and temperature in sofia and london. *Journal of Epidemiology & Community Health*, 57(8):628–633, 2003.
- [55] 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.
- [56] 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.
- [57] 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.
- [58] 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.
- [59] 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.
- [60] 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.
- [61] 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.
- [62] David Savage, Xiuzhen Zhang, Xinghuo Yu, Pauline Chou, and Qingmai Wang. Anomaly detection in online social networks. *Social Networks*, 39:62–70, 2014.

- [63] J. Shi et al. Cardiovascular benefits of wearing particulate-filtering respirators: a randomized crossover trial. *Environmental health perspectives*, 125(2):175–180, 2017.
- [64] 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.
- [65] 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.
- [66] 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.
- [67] 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.
- [68] 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.
- [69] J. Vermylen et al. Ambient air pollution and acute myocardial infarction. *Journal of Thrombosis and Haemostasis*, 3(9):1955–1961, 2005.
- [70] 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.
- [71] 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.
- [72] 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.
- [73] Sherrie Xie and Blanca E Himes. Personal environmental monitoring. In *Precision in Pulmonary, Critical Care, and Sleep Medicine*, pages 305–320. Springer, 2020.
- [74] Antonella Zanobetti, Francesca Dominici, Yun Wang, and Joel D Schwartz. A national case-crossover analysis of the short-term effect of pm<sub>2.5</sub> on hospitalizations and mortality in subjects with diabetes and neurological disorders. *Environmental Health*, 13(1):1–11, 2014.
- [75] R. Zhang et al. Acute effects of particulate air pollution on ischemic stroke and hemorrhagic stroke mortality. *Front Neurol*, 9:827, 2018.
- [76] 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.

- [77] 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.
- [78] 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.
- [79] 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.
- [80] Marina Zusman, Cooper S Schumacher, Amanda J Gassett, 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.
- [81] 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.