

УНИВЕРЗИТЕТ „СВ. КИРИЛ И МЕТОДИЈ“ ВО СКОПЈЕ

МАШИНСКИ ФАКУЛТЕТ СКОПЈЕ



ПРИМЕНА НА ВЕШТАЧКА ИНТЕЛИГЕНЦИЈА ЗА
ПРЕПОЗНАВАЊЕ И КЛАСИФИКАЦИЈА НА ЗВУЧНИ НАСТАНИ

Докторска дисертација

м-р Симона Домазетовска, дипл. маш. инж.

Скопје, 2022

НАСЛОВ: Примена на вештачка интелигенција за препознавање
и класификација на звучни настани

ИЗРАБОТИЛ: асис. м-р Симона Домазетовска, дипл. маш. инж.

МЕНТОР: проф. д-р Виктор Гаврилоски

ОБЛАСТ: Технички науки - Машинство

ДАТА НА ЈАВНА ОДБРАНА: 29.06.2022

КОМИСИЈА ЗА ОДБРАНА НА
ДОКТОРСКАТА ДИСЕРТАЦИЈА

1. Проф. д-р Виктор Гаврилоски
Машински Факултет – Скопје
Универзитет „Св. Кирил и Методиј“ во Скопје
2. Проф. д-р Златко Петрески
Машински Факултет – Скопје
Универзитет „Св. Кирил и Методиј“ во Скопје
3. Проф. д-р Даме Коруноски
Машински Факултет – Скопје
Универзитет „Св. Кирил и Методиј“ во Скопје
4. Проф. д-р Христијан Мицкоски
Машински Факултет – Скопје
Универзитет „Св. Кирил и Методиј“ во Скопје
5. Проф. д-р Дејан Мираковски
Факултет за природни и технички науки
Универзитет „Гоце Делчев“, Штип

АПСТРАКТ

Оваа докторска дисертација претставува прилог во интердисциплинарна научна област која има за цел развој на систем за препознавање и класификација на звучни настани и негова примена во реални инженерски апликации. Истражувањата во оваа докторска дисертација се насочени кон развој на системи за препознавање и класификација на звучни настани преку користење разни параметри на звучни настани и техники на вештачка интелигенција.

Мотивацијата, целите и придобивките од ова истражување се дадени во глава 1. Целите на докторската дисертација се дефинирани на почетокот на истражувањето, а методологијата за истражување е утврдена за сеопфатно и ефикасно следење на напредокот и добивање на резултати од истражувањето. Поставени се главната и помошната хипотеза, кои се потврдија и докажаа во текот на изработката на докторската дисертација.

Во глава 2 се разработени стандардните методи за проценка на бучавата во животната средина преку објективна и субјективна анализа. Овие методологии овозможуваат проценка на нивото на бучава во урбано подрачје, но истите не можат да дадат информации за доминантните извори на бучава, што јавува потреба за дополнителни истражувања за идентификација на доминантните извори на бучава преку примена на напредни технологии кои се базираат на вештачката интелигенција.

Преку систематизацијата која е обработена во глава 3 се развива таксономија врз чија основа се одбрани 10 класи на звучни настани кои ќе овозможат формирање на база на звучни настани која ќе се користи за тренирање и тестирање на алгоритмите за класификација. Дополнително, одбрани се 5 параметри на звучни настани кои се математички моделирани и се користат за екстракција на важните карактеристики на звучните настани.

Во глава 4 развиени се математички модели на три алгоритми на машинско учење: машини со потпирачки вектори, случајни шуми и наивен Баесов алгоритам, кои се применети во експерименталниот дел при моделирање на системот за препознавање и класификација на звучните настани. Дополнително, како резултат од примена на процесот на хиперпараметарска оптимизација на најуспешните модели се постигна повисока точност при класификација на звучните настани.

Конволуциските невронски мрежи како алгоритми на длабокото учење се обработени и применети во глава 5. Покрај тренирање и тестирање на алгоритмите, беше применета дополнителна валидација во две фази на моделот кој покажа најдобри перформанси со цел да се добијат сознанија и јасна претстава за применливоста и начинот на кој може да се подобрат перформансите со понатамошни истражувања.

Глава 6 е посветена на заклучоците од ова истражување и препораките за идни истражувања во оваа област.

SUMMARY

This thesis focuses on development of acoustic event detection and classification (AED/C) systems based on the artificial intelligence (AI) techniques. The main research idea is devoted to several audio parameters and AI-based algorithms so classifying of the sound events with high accuracy will be established.

The goals of this dissertation and the research methodology are defined in chapter 1 in order to establish comprehensive and effective monitoring of the progress and the obtained results in this research. Also, literature review was proposed in this chapter, as well as the main motivation for choosing the research area.

Chapter 2 focuses on the standard methods for environmental noise level estimation through objective and subjective analysis that were developed and experimentally used. It can be concluded that the proposed methodologies give accurate results for quantifying the noise pollution, but could not detect the dominant noise sources, requiring additional research for identifying the dominant urban noise sources through the application of advanced technologies based on artificial intelligence.

Through the systematization proposed in chapter 3, a taxonomy for urban sounds is created leading in choosing 10 classes of sound events. Based on the classes of sound events, a database was chosen for training and testing the classification algorithms. Furthermore, in this chapter, 5 audio parameters were chosen and modeled for the feature extraction process.

Chapter 4 develops mathematical models for three machine learning algorithms: support vector machines, random forest and naïve Bayes algorithm, that are further used in the experimental part while modeling the AED/C system. Furthermore, the process of hyperparameter optimization for the most successful models resulted in higher accuracy when classifying the sound events.

The convolutional neural networks as deep learning algorithms are elaborated in chapter 5. After training and testing the algorithms, further two-step validation was applied in order to measure the performance of the system and gain more knowledge when implementing the AED/C system in a real application.

Chapter 6 is devoted to conclusions and further research recommendations.

ПРЕДГОВОР

Ја користам оваа прилика да изразам искрена благодарност до сите кои придонесоа за успешно завршување на докторската дисертација, како и за мојот досегашен академски развој.

Најголем придонес во мојот професионален развој и во реализацијата на докторската дисертација има мојот професор и ментор, проф. д-р Виктор Гаврилоски. Би сакала да изразам огромна благодарност за неговата поддршка и доверба, времето посветено во размена на идеи и насоки за развој на тезата, како и сите долги и квалитетни дискусии во кои несебично го споделуваше своето знаење и искуство.

Голема благодарност до проф. д-р Златко Петрески за значајните совети и безрезервната поддршка изминативе години. Им се заблагодарувам на проф. д-р Христијан Мицкоски и проф. д-р Даме Коруноски за довербата и корисните совети, како и слободата во развојот на часовите како асистент по нивните предмети. Благодарност и до надворешниот член на комисијата проф. д-р Дејан Мираковски за прегледување и оценка на дисертацијата.

Би сакала да изразам огромна благодарност и до останатите членови на Институтот за Механика, особено до мојата колешка и пријателка, Маја Аначкова, која што секојдневно беше тука за да ме мотивира и охрабри кон остварување на моите цели. Искрено се надевам дека нашата професионална соработка ќе продолжи во иднина.

Моето семејство и Мартин заслужуваат посебна благодарност за поддршката и трпението во периодот на изработката на докторската дисертација. На крај, благодарност до сите мои најблиски кои се директни секојдневни сведоци на сите мои успеси и остварувања.

СОДРЖИНА

Апстракт	iii
Summary	iv
Предговор	v
Содржина	vi
Листа на слики	x
Листа на табели	xiv
1. ВОВЕД	1
1.1. МЕТОДИ ЗА ПРОЦЕНКА НА БУЧАВАТА ВО УРБАНИТЕ СРЕДИНИ	2
1.2. МОТИВАЦИЈА НА ИСТРАЖУВАЊЕТО	3
1.3. ПРЕГЛЕД НА ПРЕТХОДНИ ИСТРАЖУВАЊА	5
1.3.1. Стандардни методи за проценка на бучавата во животната средина	6
1.3.2. Напредни методи за проценка на бучавата преку примена на системи за препознавање и класификација на звучни настани	8
1.4. ПРЕДМЕТ НА ИСТРАЖУВАЊЕТО	12
1.5. ЦЕЛИ НА ИСТРАЖУВАЊЕТО	13
1.6. ХИПОТЕЗИ	15
1.5. МЕТОДОЛОГИЈА НА ИСТРАЖУВАЊЕ	15
1.6. ПРИДОБИВКИ ОД ИСТРАЖУВАЊЕТО	16
2. СТАНДАРДНИ МЕТОДИ ЗА ПРОЦЕНКА НА БУЧАВАТА ВО УРБАНА ЖИВОТНА СРЕДИНА	18
2.1. ИЗБОР НА ПОДРАЧЈЕ ЗА АНАЛИЗА	19
2.2. ПРОЦЕНКА НА НИВОТО НА БУЧАВА ПРЕКУ СТАНДАРДНИ МЕРЕЊА	20
2.2.1. Избор на мерни места	21
2.2.2. Мерна опрема	22
2.2.3. Мерни индикатори за определување на бучавата во урбана животна средина	24
2.2.4. Определување на мерна неодреденост	25
2.2.3. Резултати од мерењата	27

2.3. АКУСТИЧЕН ПРЕДИКТИВЕН МОДЕЛ	30
2.3.1. Метод за изработка на акустичен предиктивен модел	30
2.3.2. Резултати од изработениот акустичен предиктивен модел	32
2.3.3. Корелација помеѓу спроведените методологии	34
2.4. СУБЈЕКТИВНА АНАЛИЗА ЗА ВЛИЈАНИЕТО НА БУЧАВАТА	36
2.2.1. Избор на број на испитаници	36
2.2.2. Креирање анкетна форма	37
2.2.3. Евалуација на анкетата и анализа на резултатите	38
2.5. СОГЛЕДУВАЊА ОД ПРИМЕНЕТИТЕ МЕТОДИ ЗА ПРОЦЕНКА НА НИВОТО НА БУЧАВА	41
3. СИСТЕМАТИЗАЦИЈА И ПАРАМЕТРИЗАЦИЈА НА ЗВУЧНИ НАСТАНИ	43
3.1. СИСТЕМАТИЗАЦИЈА НА ЗВУЧНИ НАСТАНИ	44
3.2. АНАЛИЗА НА БАЗАТА НА ЗВУЧНИ НАСТАНИ	49
3.3. ПАРАМЕТРИЗАЦИЈА НА ЗВУЧНИ НАСТАНИ	53
3.3.1. Математички операции за екстракција на параметрите на звучни настани	55
3.3.2. Параметри во временски домен	58
3.3.3. Параметри во фреквентен домен	59
3.3.4. Параметри во цепстрален домен	62
3.3.5. Параметри во модално фреквентен домен	64
3.3.6. Ритам	65
3.3.7. Параметри во сопствен домен	66
3.3.8. Параметри во фазно просторен домен	66
3.4. СПЕКТРАЛЕН КОНТРАСТ	67
3.5. ТОНАЛЕН ЦЕНТРОИД	74
3.6. МЕЛ СПЕКТРОГРАМ	79
3.7. ХРОМАТОГРАМ	84
3.8. МЕЛ ФРЕКВЕНТНИ ЦЕПСТРАЛНИ КОЕФИЦИЕНТИ	88
4. ПРИМЕНА НА АЛГОРИТМИ НА МАШИНСКО УЧЕЊЕ	92
4.1. ОСНОВИ НА МАШИНСКОТО УЧЕЊЕ	93
4.1.1. Големи податоци	96
4.2. АЛГОРИТАМ НА МАШИНИ СО ПОТПИРАЧКИ ВЕКТОРИ	98

4.3. НАИВЕН БАЕСОВ АЛГОРИТАМ	103
4.3.1. Гаусов наивен Баесов класификатор	106
4.3.2. Мултиномен наивен Баесов класификатор	106
4.3.3. Бернулиев наивен Баесов класификатор	107
4.4. АЛГОРИТАМ НА СЛУЧАЈНИ ШУМИ	107
4.5. ХИПЕРПАРАМЕТАРСКА ОПТИМИЗАЦИЈА ПРИ НАДГЛЕДУВАН ПРИСТАП НА УЧЕЊЕ	111
4.6. РЕЗУЛТАТИ ОД ПРИМЕНА НА АЛГОРИТАМИТЕ НА МАШИНСКО УЧЕЊЕ	114
4.7. РЕЗУЛТАТИ ОД ОПТИМИЗАЦИЈА НА АЛГОРИТАМОТ НА МАШИНИ СО ПОТПИРАЧКИ ВЕКТОРИ	118
4.8. РЕЗУЛТАТИ ОД ОПТИМИЗАЦИЈА НА АЛГОРИТАМОТ НА СЛУЧАЈНИ ШУМИ	120
4.9. РЕЗУЛТАТИ ОД ПРИМЕНА НА НАИВЕН БАЕСОВ АЛГОРИТАМ	122
4.10. КОМПАРАТИВНА АНАЛИЗА НА РЕЗУЛТАТИТЕ	123
5. ПРИМЕНА НА АЛГОРИТМИ НА ДЛАБОКО УЧЕЊЕ	126
5.1. ДЛАБОКО УЧЕЊЕ	127
5.2. КОНВОЛУЦИСКИ НЕВРОНСКИ МРЕЖИ	128
5.3. МАТЕМАТИЧКИ МОДЕЛ НА КОНВОЛУЦИСКИ НЕВРОНСКИ МРЕЖИ	130
5.4. ADAM ОПТИМИЗАЦИЈА	136
5.5. ПРИМЕНА НА КОНВОЛУЦИСКИ НЕВРОНСКИ МРЕЖИ ВО СИСТЕМИТЕ ЗА ПРЕПОЗНАВАЊЕ И КЛАСИФИКАЦИЈА НА ЗВУЧНИ НАСТАНИ	138
5.6. КОНВОЛУЦИСКИ НЕВРОНСКИ МРЕЖИ ПРИ ПРИМЕНА НА МЕЛ ФРЕКВЕНТНИ ЦЕПСТРАЛНИ КОЕФИЦИЕНТИ	141
5.7. КОНВОЛУЦИСКИ НЕВРОНСКИ МРЕЖИ ПРИ ПРИМЕНА НА МЕЛ СПЕКТОГРАМ	145
5.8. КОНВОЛУЦИСКИ НЕВРОНСКИ МРЕЖИ ПРИ ПРИМЕНА НА MFCC И МЕЛ СПЕКТОГРАМ	149
5.9. КОНВОЛУЦИСКИ НЕВРОНСКИ МРЕЖИ ПРИ ПРИМЕНА НА MFCC, ХРОМАТОГРАМ И СПЕКТРАЛЕН КОНТРАСТ	152
5.10. АНАЛИЗА НА ПРИМЕНЕТИТЕ АЛГОРИТМИ НА ВЕШТАЧКА ИНТЕЛИГЕНЦИЈА	156
5.11. ВАЛИДАЦИЈА НА СИСТЕМОТ ЗА ПРЕПОЗНАВАЊЕ И КЛАСИФИКАЦИЈА НА ЗВУЧНИ НАСТАНИ	157

6. ЗАКЛУЧОЦИ И ПРЕПОРАКИ ЗА ИДНИ ИСТРАЖУВАЊА	161
6.1. ЗАКЛУЧОЦИ	162
6.1.1. Заклучоци од примената на стандардните методи за проценка на бучавата	162
6.1.2. Заклучоци од примената на напредните методи за препознавање и класификација на звучни настани	162
6.2. ПРЕПОРАКИ ЗА ИДНИ ИСТРАЖУВАЊА	165
6.2.1. Подобрвање на методите за препознавање и класификација на звучни настани	165
6.2.2. Концепт за примена во системи за континуиран мониторинг	166
ЛИТЕРАТУРА	169
ПРИЛОЗИ	177

ЛИСТА НА СЛИКИ

ГЛАВА 1

Слика 1.1. Шематски приказ (архитектура) на класичен систем за препознавање и класификација на звучни настани во животната средина

Слика 1.2. Тренирање и тестирање на системот за препознавање и класификација на звучни настани

ГЛАВА 2

Слика 2.1. Географска локација на избраното подрачје за анализа

Слика 2.2. Распоред на мерни точки

Слика 2.3. Поставеност на мерната инструментација

Слика 2.4. Оперативни фази при креирање акустичен предиктивен модел

Слика 2.5. Акустични предиктивни карти за анализираното подрачје

Слика 2.6. Акустичен предиктивен модел за временскиот период кога се направени мерења на ниво на бучава

Слика 2.7. Резултати од нивото на бучава од спроведените мерења и акустичниот предиктивен модел за мерните места ММ1, ММ2, ММ3 и ММ4

Слика 2.8. Разлика помеѓу измереното и предвиденото ниво на бучава

Слика 2.9. Анкета за влијанието на бучавата во скопската населба Дебар Маало

Слика 2.10. Местоположба на испитаниците

Слика 2.11. Степен на влијание на бучавата во животната средина кај испитаниците

Слика 2.12. Најсилен интензитет на бучавата во период според годишните сезони и деноноќие

Слика 2.13. Доминантни извори на бучава

Слика 2.14. Влијание на бучавата врз здравјето на населението

Слика 2.15. Причини за појава на бучава

Слика 2.16. Фреквентен спектар во 1/3 октава

ГЛАВА 3

Слика 3.1. Споредба помеѓу говор, музика и звучен настан во животната средина во временски и фреквентен домен

- Слика 3.2. Семантичка класа на извори на звучни настани во животна урбана средина на највисоко ниво на таксономијата
- Слика 3.3. Втора и трета поделба на изворите на бучава кои се јавуваат во природата како звучен настан
- Слика 3.4. Втора и трета поделба на изворите на бучава кои се јавуваат во транспортот како звучен настан
- Слика 3.5. Втора и трета поделба на изворите на бучава кои се јавуваат во објектот како звучен настан
- Слика 3.6. Таксономија на звучни настани во урбана звучна средина кои предизвикуваат вознемиреност кај населението
- Слика 3.7. Застапеност на одделните класи на звучни настани во базата на звучни настани
- Слика 3.8. Поделба на параметрите на звучни настани согласно доменот на обработка
- Слика 3.9. Чекори при екстракција на спектрален контраст
- Слика 3.10. Процес на конверзија на сигналот во дискретно време
- Слика 3.11. Приказ на Ханов прозорец
- Слика 3.12. Визуелен приказ на спектрален контраст за репрезентативен аудио сигнал за секоја класа на звучен настан
- Слика 3.13. Чекори за екстракција на Тонален Центроид
- Слика 3.14. Проекција која покажува како елементите на Тоналниот Центроид се обвиткуваат околу торусот. Трите кружници се визуелизација на 6-D просторот.
- Слика 3.15. Визуелен приказ на тонален центроид за репрезентативен аудио сигнал за секоја класа на звучен настан
- Слика 3.16. Чекори на обработка на аудио сигналот за добивање МЕЛ спектрограм
- Слика 3.17. Приказ на 128 МЕЛ филтер банки со триаголен облик
- Слика 3.18. Визуелен приказ на МЕЛ спектрограм за репрезентативен аудио сигнал за секоја класа на звучен настан
- Слика 3.19. Обработка на снимен звучен настан за екстракција на хроматограм
- Слика 3.20. Визуелен приказ на хроматограм за репрезентативен аудио сигнал за секоја класа на звучен настан
- Слика 3.21. Чекори за добирање на МЕЛ фреквентни цепстрални коефициенти

Слика 3.22. Визуелен приказ на МЕЛ фреквентни цепстрални коефициенти за репрезентативен аудио сигнал за секоја класа на звучен настан

ГЛАВА 4

Слика 4.1. Визуелен приказ на вештачката интелигенција, машинското учење и длабокото учење како сет од множество

Слика 4.2. Класификација и регресија

Слика 4.3. Систем за анализа на големи податоци

Слика 4.4. Хипер рамнина на линеарен класификатор за машини со потпирачки вектори (хипер рамнината H_1 не ги одделува класите; H_2 ги одделува, но со мала маргина; H_3 ги одделува со максимална маргина)

Слика 4.5. Хипер рамнина со максимална маргина помеѓу две класи. Примероците кои лежат на крајот на маргината (двете помошни линии) се наречени потпирачки вектори

Слика 4.6. Одделување на податоци со користење кернел функција

Слика 4.7. Приказ на алгоритам на случајни шуми

Слика 4.8. Архитектура на систем за препознавање и класификација на звучни настани

Слика 4.9. Систем за препознавање и класификација на звучни настани

Слика 4.10. Конфузни матрици за најуспешните модели на алгоритмот на случајни шуми, машини со потпирачки вектори и наивен Баесов алгоритам

Слика 4.11. Конфузна матрица на тестираните податоци при примена на алгоритам на машини со потпирачки вектори

Слика 4.12. Конфузна матрица на тестираните податоци при примена на алгоритам на случајни шуми

Слика 4.13. Конфузна матрица на тестираните податоци при примена на Наивен Баесов алгоритам

Слика 4.14. Аудио параметри на репрезентативни аудио сигнали на звучните класи од работа на возило во мирување, гласна музика и бушење

ГЛАВА 5

Слика 5.1. Архитектура на конволуциски невронски мрежи

Слика 5.2. Визуелен приказ на процесот на конволуција

Слика 5.3. Примена на функција на максимално поврзување (max pooling) и подредување

Слика 5.4. Илустрација на методот на Стохастички градиент на намалување

Слика 5.5. Графички приказ на ReLU функцијата

Слика 5.6. Визуелен приказ на CNN мрежата при примена на MFCC

Слика 5.7. Постигната точност при класификација на податоците низ слоевите на мрежата на CNN

Слика 5.8. Загуба при класификација на податоците низ слоевите на мрежата на CNN

Слика 5.9. Конфузна матрица при примена на модел на CNN со MFCC

Слика 5.10. ReLU активациска функција со препуст

Слика 5.11. Визуелен приказ на CNN мрежата при примена на МЕЛ спектрограм

Слика 5.12. Постигната точност при класификација на податоците низ слоевите на мрежата на CNN

Слика 5.13. Загуба при класификација на податоците низ слоевите на мрежата на CNN

Слика 5.14. Конфузна матрица при примена на модел на CNN со МЕЛ спектрограм

Слика 5.15. Визуелен приказ на CNN мрежата при примена на MFCC и МЕЛ спектрограм

Слика 5.16. Постигната точност при класификација на податоците низ слоевите на мрежата на CNN

Слика 5.17. Загуба при класификација на податоците низ слоевите на мрежата на CNN

Слика 5.18. Конфузна матрица при примена на модел на CNN со MFCC и МЕЛ спектрограм

Слика 5.19. Визуелен приказ на CNN мрежата при примена на MFCC, хроматограм и спектрален контраст

Слика 5.20. Постигната точност при класификација на податоците низ слоевите на мрежата на CNN

Слика 5.21. Загуба при класификација на податоците низ слоевите на мрежата на CNN

Слика 5.22. Конфузна матрица при примена на модел на CNN со MFCC, хроматограм и спектрален контраст

Слика 5.23. Резултати од точноста при класификација од валидацијата

Слика 5.24. Приказ на алгоритмот за валидација на податоците

Слика 5.25. Резултати од валидација на звучните настани снимени во континуиран период

ГЛАВА 6

Слика 6.1. Визуелен приказ на предложениот концепт за имплементација на системот за препознавање и класификација на звучни настани во IoT инфраструктура за паметен град

Слика 6.2. Логички тек на системот за една сензорска единица

ЛИСТА НА ТАБЕЛИ

ГЛАВА 1

Табела 1.1. Преглед на истражувања во полето на препознавање и класификација на урбани звучни настани

ГЛАВА 2

Табела 2.1. Информации за инструмент за мерење на бучава

Табела 2.2. Информации за акустичен калибратор

Табела 2.3. Резултати од извршената интерна калибрација на инструментот

Табела 2.4. Буџет на мерна неодреденост

Табела 2.5. Пресметана мерна неодреденост од спроведените мерења

Табела 2.6. Дополнителни индикатори од спроведените мерења

ГЛАВА 3

Табела 3.1. Листа на карактеристики за говор, музика и звуци во животната средина

Табела 3.2. Поделеност на класите на податоци во датотеките

Табела 3.3. Процентуална застапеност на податоци согласно стапката на семплирање

Табела 3.4. Резолуцијата со која се снимени аудио податоците

Табела 3.5. Основни својства на параметрите на звучни настани

Табела 3.6. Трансформации

Табела 3.7. Филтри

Табела 3.8. Агрегации и детектори

Табела 3.9. Применети математички операции на издвоените параметри во временски домен

Табела 3.10. Применети математички операции на издвоените параметри во фреквентен домен

Табела 3.11. Применети математички операции на издвоените параметри во цепстрален домен

Табела 3.12. Применети математички операции на параметрите во модално фреквентен домен

Табела 3.13. Применети математички операции на издвоените параметри на ритам

Табела 3.14. Применети математички операции на издвоените параметри во сопствен домен

Табела 3.15. Применети математички операции на издвоените фазно просторни аудио карактеристики

ГЛАВА 4

Табела 4.1. Вид на кернел функции

Табела 4.2. Процентуален приказ на точноста на системот при тестирање во зависност од бројот на параметри на MFCC Табела

Табела 4.3. Анализа на точноста на тестираните податоци во зависност од примена на различни параметри

Табела 4.4. Ранг на вредности на искористените хипер параметри

ГЛАВА 5

Табела 5.1. Архитектура на длабоко учење и нивна примена

Табела 5.2. Приказ на ADAM алгоритмот на оптимизација

Табела 5.3. Архитектура на мрежата на алгоритмот на конволуциски невронски мрежи

Табела 5.4. Постигната точност при класификација на одделните класи

Табела 5.5. Архитектура на мрежата на алгоритмот на конволуциски невронски мрежи

Табела 5.6. Постигната точност при класификација на одделните класи

Табела 5.7. Архитектура на мрежата на алгоритмот на конволуциски невронски мрежи

Табела 5.8. Постигната точност при класификација на одделните класи

Табела 5.9. Архитектура на мрежата на алгоритмот на конволуциски невронски мрежи

Табела 5.10. Постигната точност при класификација на одделните класи

Табела 5.11. Резултати од точноста при тестирање на применетите модели

1. ВОВЕД

Воведната глава има за цел да ја прикаже главната идеја на оваа докторска дисертација и процесот на нејзина реализација. Методите за проценка на бучавата во урбаните средини се дадени на почетокот на главата, во насока кон примена на напредни методи базирани на техниките на вештачката интелигенција.

Можностите за развој на напредни технички системи со високи перформанси како предизвик за современото инженерство и главен двигател на техничко-технолошкото иноваторство беа главна мотивација за изборот на истражувачката област. Потребата од употреба на современи технологии претставува мотив за примена на вештачка интелигенција за развој на системи за препознавање и класификација на звучни настани кои ќе дадат прецизни и навремени резултати во динамичните урбани средини. Имајќи ја предвид интердисциплинарноста на темата, неопходно беше да се направи систематски преглед на литературата со цел да се навлезе во проблематиката од интерес.

Прегледот на литературата од поедините области ги анализира стандардните методи за проценка на бучавата и нивно подобрување и надградување преку примена на напредни методи кои се базираат на вештачката интелигенција. Прегледот на литература е поделен по области кои е потребно да се совладаат за да се овозможи фундаментално истражување, како и примена на концептите на конкретен експериментален систем.

Целите на докторската дисертација се дефинирани на почетокот на истражувањето, а методологијата за истражување е утврдена за сеопфатно и ефикасно следење на напредокот и добивање на резултатите од истражувањето.

Поставени се главната и помошната хипотеза, кои се потврдија и докажаа во текот на изработката на докторската дисертација.

На крајот од оваа глава, накратко се дадени придобивките од оваа докторска дисертација во развојот на истражувачкото подрачје со цел да се добие јасна слика за квалитетот на ова истражување.

1.1. МЕТОДИ ЗА ПРОЦЕНКА НА БУЧАВАТА ВО УРБАНИТЕ СРЕДИНИ

Стандардните методи за проценка на бучавата вклучуваат определување на нивото на бучава преку објективна анализа која ја проучува бучавата од физичка гледна точка, како и примена на субјективна студија за анализа на влијанието на бучавата кај населението. Овие методи претставуваат основна алатка за обезбедување релевантни информации за загадувањето од бучава во урбаните средини со цел изработка на локални и глобални акциони планови кои ќе придонесат за намалување на нивото на бучава. Бројните истражувања покажуваат дека преку употреба на методи за нормирање на бучавата вклучувајќи мерења на нивото на бучава и изработка на акустичен предиктивен модел се потврдува зголеменото ниво на бучава во урбаните средини [1, 2, 3]. Традиционалните пристапи вклучуваат изведба на краткотрајни мерења согласно Европската Директива и националните регулативи преку користење букомер од класа 1 [4, 5, 6, 7, 8]. Со цел да се визуелизира дисперзијата на бучавата во дефинирано подрачје за анализа, научниците користат техники за мапирање на бучавата преку картографска репрезентација на нивото на бучава. Од друга страна, преку примена на пристап за субјективна анализа може да се процени влијанието на бучавата и да се детектира изворот на бучава кој многу пречи, односно секојдневно ја нарушува општата здравствена состојба на населението. Преку употребата на гореспоменатите пристапи за анализа на бучавата кои вклучуваат изведба на мерења, изработка на карти на бучава, како и употреба на статистички научен метод за субјективна анализа, се овозможува проценка на нивото на бучава и нејзиното штетно влијание во урбаните средини, што може да доведе до постигнување значајно и корисно карактеризирање на бучавата.

Во последно време, како резултат на динамичниот развој на урбаните средини, стандардните методи за проценка на бучавата не можат да дадат доволно прецизни и навремени резултати при детектирање на изворите кои предизвикуваат бучава. Барањата за поточна оцена на состојбата и прецизна детекција на звучните настани кои предизвикуваат зголемено ниво на бучава наметнуваат потреба од примена на нови пристапи во техничките системи за следење на состојбата со бучавата. Во современото општество, системите за препознавање и класификација на звучните настани базирани на вештачка интелигенција стануваат сè повеќе актуелни за истражување. Идејата и можноста за следење на состојбата на звучните извори кои предизвикуваат загадување од бучава е мотив за интензивно академско истражување во насока кон развивање на нови техники и методи засновани на вештачката интелигенција. Важноста која ја има системот за препознавање и класификација на звучните настани води кон постојан развој преку примена на ефективни инженерски технологии и решенија со цел точна и сигурна детекција на звучните настани.

Системите за препознавање и класификација на звучните настани се дел од областа на машинското учење, дефинирани како способност на компјутерите да слушаат како луѓето, разликувајќи го говорот од музика и звуци во позадина, вклучувајќи ја и можноста за откривање дополнителни звучни настани. При креирање на систем кој служи за препознавање

и класификација на звучните настани, потребно е да се изградат подсистемите за детектирање и параметризација преку дигитална обработка на звучните сигнали со цел понатамошна класификација преку употреба на алгоритми засновани на вештачка интелигенција [9]. Главната цел на системите за препознавање и класификација на звучните настани е да се изработат алгоритми кои сами можат да ги детектираат и класифицираат звучни настани што се случуваат во нивната околина, давајќи соодветен одговор на корисниците. Системите се развиваат преку примена на вештачка интелигенција, која се потпира на идејата дека системите можат да учат од податоците, да идентификуваат карактеристики и да донесуваат одлуки со минимална интервенција од страна на човекот. Брзиот развој во полето на вештачката интелигенција ги прави компјутерите способни да решаваат проблеми кои се од голема важност на животната средина, како што е развојот на алгоритми за откривање на звучни настани во полето на бучавата.

Поради комплексната природа на звукот, системите за препознавање и класификација на звучните настани треба да бидат детално истражувани со цел да се постигне нивна интегрираност во комплексна урбана средина за подобрување на нејзините параметри. Со денешниот интензивен развој на системите базирани на вештачката интелигенција и „Интернет на нештата“, како и развојот на апликации подржани од машинско учење, овие системи сè повеќе го наоѓаат своето место во многу истражувачки области посветени на имплементација и анализа во реално време.

1.2. МОТИВАЦИЈА НА ИСТРАЖУВАЊЕТО

Развојот на напредни технички системи со високи перформанси претставува предизвик за современото инженерство и главен двигател на техничко-технолошкото иноваторство. Потребата од оценка на состојбата на нивото на бучава во урбаните средини доведува до развој на стандардни методи за проценка на нивото на бучава преку употреба на методологии кои одземаат многу време за спроведување и нивна верификација е потребен тим на експерти за звучна проценка на бучавата. Дополнително, стандардните методи за проценка на нивото на бучава не можат постојано да ја следат состојбата со бучавата и не можат да дадат резултати во реално време за изворите кои предизвикуваат бучава. Потребата за поточна и побрза оценка на состојбата со бучавата претставува мотив за употреба на напредни техники за развој на системи кои со висока прецизност ќе ги препознаваат и класифицираат звучните настани.

Денес, во светски рамки, многу научни институции и научници од различни области вложуваат огромни напори и ресурси за истражување на концептот на системи засновани на вештачка интелигенција и нивна примена во различни апликации. Системите за препознавање и класификација на звучните настани наоѓаат широка примена во апликации кои служат за поедноставување на стандардните методи за проценка на бучавата при развој на системи за мониторинг на бучавата предизвикана од патниот сообраќај [10, 11]. Оваа примена

овозможува развој на динамични карти на бучава кои даваат податоци за нивото на бучава во реално време и го олеснуваат процесот при креирање стратешки карти на бучава. Од друга страна, научниците во [12] креираат мрежа од ниско-буџетни сензори засновани на концептот на вештачка интелигенција за постојано мониторирање и определување на доминантните извори на бучава како дел од концептот на паметни градови и „Интернет на нештата“.

Иако концептот за системи засновани на вештачката интелигенција е актуелен во последните три децении, неговата применливост станува сè повеќе актуелна со појавата на моќни дигитални процесори за обработка на сигнали, со што се постигнува брза обработка и анализа на податоците. Технологијата заснована на вештачка интелигенција има големо значење во современото инженерство и е мотивација за истражување со можност за добивање подлабоки сознанија и придонес во полето на системите за препознавање и класификација на звучни настани. Високата точност која може да се постигне преку анализа и избор на параметрите кои се извлекуваат со дигитална обработка на сигналите и алгоритмите за машинско учење претставува мотивација за истражување во оваа докторска дисертација.

Техниката на машинско слушање вклучува детекција на широк спектар на звуци, а идеалниот систем за препознавање и класификација на звучните настани треба успешно да ги идентификува класите во кои звуците припаѓаат. При употреба на овие системи, најпрвин потребно е самите звучни настани да бидат систематизирани согласно предложена таксономија на звукот. Врз основа на предложената таксономија, потребно е да се креира база на податоци за избраните звучни настани во животната средина која се анализира. Звучните настани претставени како аудио сигнали од избраната база на податоци треба да бидат претходно обработени преку примена на технологии за обработка на дигитални сигнали. Вака обработените податоци формираат комплетна база на податоци која претставува влез во системот за препознавање и класификација на звучните настани. Преку поделба на звучните настани во соодветни класи се подобрува точноста при класификација и се овозможува побрзо процесирање на податоците. Согласно спроведените анализи, во оваа докторска дисертација ќе се примени база на податоци од звучни настани кои предизвикуваат бучава во урбаната животната средина.

При изградба на системите за препознавање и класификација на звучни настани, голем предизвик е да се направи правилна параметризација, односно да се идентификуваат и екстрахираат соодветни карактеристики на аудио сигналите кои ќе овозможат голема точност при препознавање на класата на звучниот настан. Процесот се состои од обработка на акустични сигнали и нивно претворање во симболични описи кои одговараат на различните звучни настани присутни во самиот сигнал. Сигналите се обработени користејќи ја субјективната перцепцијата на човекот за звучните настани преку анализа на сензорно-фигуративни, концептуално-логички (семантички) и параметарски информациски модели за различни звучни настани од објективна реалност во човечкиот мозок. Изборот на карактеристиките на сигналот за звучните настани има клучна улога при постигнување на успешен систем за препознавање и класификација на звучните настани. Сите овие карактеристики кои се добиени од обработка на сигналите се поврзуваат во единечен вектор на карактеристики кој го претставува звучниот сигнал како предмет за препознавање со одредени карактеристики. Повикувајќи се на истражувањата во [13, 14], ќе се анализираат

параметри кои постигнуваат голема ефикасност при изградба на овој систем, и согласно овие анализи, ќе се изберат неколку параметри кои ќе се користат во експерименталниот дел при изградба на системот за детекција и класификација на звучните настани.

Користејќи ја вештачката интелигенција и машинското учење, создадени се алгоритми за идентификација и препознавање настани што се случуваат во нивната околина, така што тие ќе можат да дадат соодветен одговор на потребите на нивните корисници. Преку користење на обработените звучни сигнали, врз основа на информациите обезбедени од компактниот вектор на карактеристики, ќе се примени процес за препознавање на звучниот сигнал преку тренирање и тестирање на обработените податоци. Тренирањето и тестирањето на системот е овозможено преку користење на постоечки сет на податоци кои се соодветно поделени во класи и имаат одредени карактеристики кои даваат дополнително објаснување на класите. При изградба на системот со надгледуван пристап на учење, за целите на ова истражување, ќе се употребат повеќе алгоритми за машинско учење кои даваат прецизни резултати при анализа на звучните настани согласно истражувањата во [15, 16, 17, 18]. Напредните технологии базирани на вештачка интелигенција се предводници во полето на имплементација на системи за класификација и препознавање на звучните настани бидејќи постигнуваат висока точност и имаат минимална потреба од човечки ресурси за нивно работење. Развојот на системи засновани на вештачка интелигенција претставува инспирација за истражување со мотив да се поедностави процесот на класифицирање на звучните настани и истиот да може да се употреби во реални апликации во современите урбани средини.

1.3. ПРЕГЛЕД НА ПРЕТХОДНИ ИСТРАЖУВАЊА

На почетокот на ова истражување се направи преглед на достапна литература, вклучувајќи книги и публикации од светски познати бази на податоци со цел запознавање со областа и следење на трендовите во развој широм светот. По детален преглед на достапната литература беше согледано дека постојат стандардни и напредни методи за проценка на бучавата во животната средина. Согласно спроведените анализи, прегледот на литературата е поделен во неколку потесни подрачја од интерес:

- Стандардни методи за проценка на бучавата во животната средина
- Напредни методи за проценка на бучавата користејќи ги техниките на вештачка интелигенција

1.3.1. СТАНДАРДНИ МЕТОДИ ЗА ПРОЦЕНКА НА БУЧАВАТА ВО ЖИВОТНАТА СРЕДИНА

Бучавата во современото општество е важен јавен здравствен проблем кој има сè поголемо влијание врз квалитетот на живеење како резултат на динамичниот развој на урбаните средини. Согласно Европската комисија [19]: „Бучавата во животната средина предизвикана од сообраќајот, индустријата и рекреативните активности е една од главните еколошки проблеми во Европа, која сè повеќе резултира со поплаки и вознемиреност од страна на јавноста“. Според Светската здравствена организација [20], загадувањето предизвикано од бучавата се наоѓа помеѓу трите највлијателни параметри на животната средина кои штетно влијаат и предизвикуваат загадување во околината.

Бучавата во животната средина е бучава предизвикана од несакан или штетен надворешен звук, создаден од човековите активности кој е наметнат од блиската средина и предизвикува непријатност и вознемирување. Бучавата е значаен фактор кој доколку ги надмине пропишаните вредности дефинирани со закон, може да предизвика деградација на квалитетот на живеење. Согласно анализата спроведена од страна на Европската Унија во 32 земји во Европа за влијанието на бучавата врз здравствената состојба на популацијата, резултатите покажуваат нарушување на здравјето и појава на хипертензија, кардиваскуларни болести, како и зголемена вознемиреност [21]. Спроведените истражувања потврдуваат дека најмалку 25% од популацијата која живее во Европа има намален квалитет на животот поради аномалијата предизвикана од бучавата во животната средина и дека помеѓу 5% и 15% од населението страда од сериозно нарушување на сонот предизвикано од бучавата [22]. Понатаму, се проценува дека поради влијанието на бучавата во животната средина, Европската Унија троши помеѓу 13 и 28 милјарди евра на годишно ниво за медицински трошоци, изгубени работни денови, намалување на цените на куќите и намален потенцијал за искористување на земјиштето.

Истражувањата во [23, 24, 25, 26] потврдуваат дека бучавата предизвикана од автомобилскиот сообраќај е доминантен извор на бучава во урбаните средини што може да доведе до намалување на слухот, несоница, намалена продуктивност и појава на сообраќајни несреќи. Покрај бучавата предизвикана од сообраќајот, постојат и други извори на бучава кои имаат штетно влијание врз здравствената состојба на населението. Со анализа на бучавата предизвикана во урбана животна средина, потврдено е постојат повеќе извори на бучава кои негативно влијаат на населението и ја загадуваат животната средина [27].

Прв чекор за решавање на загадувањето на бучава е нејзино идентификување преку примена на стандардни методи за проценка на бучавата кои вклучуваат мерење, изработка на карти на бучава и субјективна анализа, како и напредни методи преку примена на концептот на вештачка интелигенција и паметни градови.

Стандардните методи за проценка на бучавата во животната средина вклучуваат спроведување на објективна и субјективна анализа. Објективната анализа ја проучува бучавата од физичка гледна точка анализирајќи го значењето на звучниот бран, додека субјективната студија се однесува на влијанието на бучавата врз индивидуалното чувство на населението.

Преку употреба на стандардни методи за нормирање на бучавата вклучувајќи мерења на нивото на бучава и изработка на акустичен предиктивен модел, студиите во повеќе земји ја потврдуваат зголемената изложеност на бучава во урбаните области [28,29]. Овие студии преставуваат основна алатка за обезбедување релевантни информации за загадувањето од бучава во урбаните средини со цел изработка на глобални и локални акциони планови кои ќе придонесат до намалување на бучавата.

Со цел правилна проценка на загадувањето од бучава во животната средина потребно е да се направат мерења на бучавата користејќи анализатор на бучава од 1ва класа. Според истражувањето во [30], мерењето на бучавата е важна дијагностичка алатка во технологијата за контрола на бучавата и проценка на загадувањето од бучавата. За да се опише квантитативно бучавата во околината се користи параметарот за еквивалентно ниво на бучава L_{eq} , користејќи A пондерирана крива. Многу веројатно е дека измерената вредност на еквивалентното ниво на бучава ќе се разликува од вистинската поради појава на мерна неодреденост која зависи од условите и изворите на бучава. Мерната неодреденост ја прикажува разликата помеѓу измерената и „вистинската вредност“, а точноста при мерењето се одредува преку разликата на овие две вредности. Резултатот од мерењето на бучавата во животната средина е само проценка на вистинската вредност и единствено е комплетен само кога е даден заедно со мерната неодреденост. Согласно ISO стандардот за одредување на бучавата предизвикана во животната средина [6], мерната неодреденост при мерење на бучава во надворешна животна средина може да биде предизвикана од: видот на букомерот, звукот во позадина, видот на извор, метеоролошки услови и локација на изворот. Постојат многу истражувања како [31,32,33], кои имаат дадено значаен придонес во областа на одредување на мерната неодреденост при мерење на бучава во животна средина.

Со цел да се процени влијанието на бучавата во урбана животна средина, научниците во [34] користат техники за мапирање на бучавата преку картографска репрезентација на нивото на бучава и негова дисперзија во претходно дефинирано подрачје. Картите на бучава претставуваат средство за прикажување на пресметаното или предвиденото ниво на бучава на репрезентативен начин во одредена географска област [35]. Под пресметано ниво се подразбира спроведување на традиционални мерења и нивно инкорпорирање во софтвер за изработка на карти на бучава, додека под предвидено ниво на бучава се подразбира внесување на проток на сообраќај за анализираното подрачје и добивање акустични предиктивни карти. Согласно Европската Директива 2002/49/ЕС за проценка и управување со бучавата во животната средина [4], наведено е дека сите членки на ЕУ мора да направат карта на бучава за густо населените агломерации. Согласно ова барање, Англија во 2004 година ја објавува првата карта на бучава од Лондон, која претставува референца за граѓаните на Лондон да избегнат бучни средини, а понатаму и останатите големи градови на земјите членки во ЕУ пристапуваат кон изработка на карти на бучава. Истражувачите во [36] потенцираат дека покрај нивото на бучава, во предвид треба да се земат и изворите кои предизвикуваат бучава. Согласно зголемениот интерес за изготвување на карти на бучава во урбаните средини, постојат разни софтвери (IMMI, Sound PLAN, Predictor, Premium) кои овозможуваат моделирање на местоположбата и внес на разните извори на бучава. Може да се заклучи дека картите за бучава прикажуваат временска и просторна распределба на бучавата и можат да помогнат при одредување на главното влијание на изворите на бучава, а се исклучително

корисни за детална проценка на влијанието на разни локации во една урбана средина [37]. Карта на бучава може да се користи за следните цели: да ги квантифицира главните извори на бучава, јасно да го илустрира загадувањето од бучава за да можат да се превземат активности за нејзино намалување, да иницира контрола на бучавата и нејзино намалување, како и да ги подобри националните и регионалните планови за да заштита на околината. Сепак, овие студии објективно го анализираат влијанието на бучавата, но не и ефектот што таа го предизвикува врз населението.

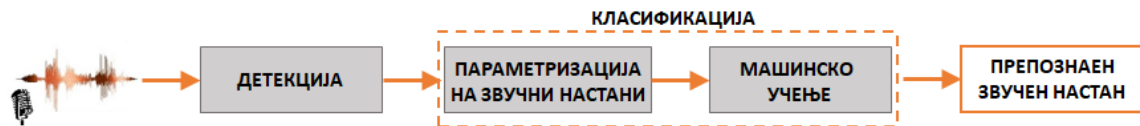
За проценка на ефектот на бучавата врз населението, во истражувањето [38] е спроведена анализа за субјективното влијание на бучавата врз населението преку метод на интервјуирање, нагласувајќи ги проблемите кои се појавуваат како замор, главоболка, депресија и несоница. Според извршената анализа за субјективното влијание на бучавата, авторите во [39] презентираат модел за нејзиното влијание, земајќи го во предвид влијанието на бучавата во различните временски периоди (ден и ноќ). Преку спроведување на субјективна анализа во [40], научниците покажале дека се добиваат дополнителни сознанија за ставот околу бучавата и највлијателните извори кои предизвикуваат штетни ефекти. Субјективното согледување за бучавата од страна на населението кое живее во урбани средини е главен фактор при анализа на критичните и највлијателните извори на бучава. Во [41], бучавата во животната средина е опишана и анализирана на универзитетски кампус преку спроведени мерења, изработка на карти на бучава и изработка на субјективна анализа.

Со анализа на објективните и субјективните аспекти на бучавата може да се постигне значајно и корисно дефинирање на загадувањето предизвикано од бучавата. Дополнително, преку пресметување на мерна неодреденост може да се најде корелација помеѓу резултатите добиени од картите на бучава со резултатите од мерење на нивото на бучава, а со тоа да се потврди и точноста на спроведената методологија.

1.3.2. НАПРЕДНИ МЕТОДИ ЗА ПРОЦЕНКА НА БУЧАВАТА ПРЕКУ ПРИМЕНА НА СИСТЕМИ ЗА ПРЕПОЗНАВАЊЕ И КЛАСИФИКАЦИЈА НА ЗВУЧНИ НАСТАНИ

Напредните методи за проценка на бучавата вклучуваат употреба на вештачка интелигенција преку примена на системи за препознавање и класификација на звучните настани кои служат за поедноставување на класичните традиционални методи за проценка на загадувањето од бучавата [42]. Главната цел на овие системи кои се потпираат на концептот на вештачка интелигенција е да се развијат алгоритми способни да ги препознаат и класифицираат звучните настани во класи соодветни за избраната акустична средина.

Карактеристичниот систем за препознавање и класификација на звучните настани во животната средина кој ќе се користи во оваа докторска дисертација е прикажан на слика 1.1., каде што може да се забележат три главни процеси: детекција, параметризација на звучни настани и класификација преку примена на вештачка интелигенција.



Слика 1. 1. Шематски приказ (архитектура) на класичен систем за препознавање и класификација на звучни настани во животната средина

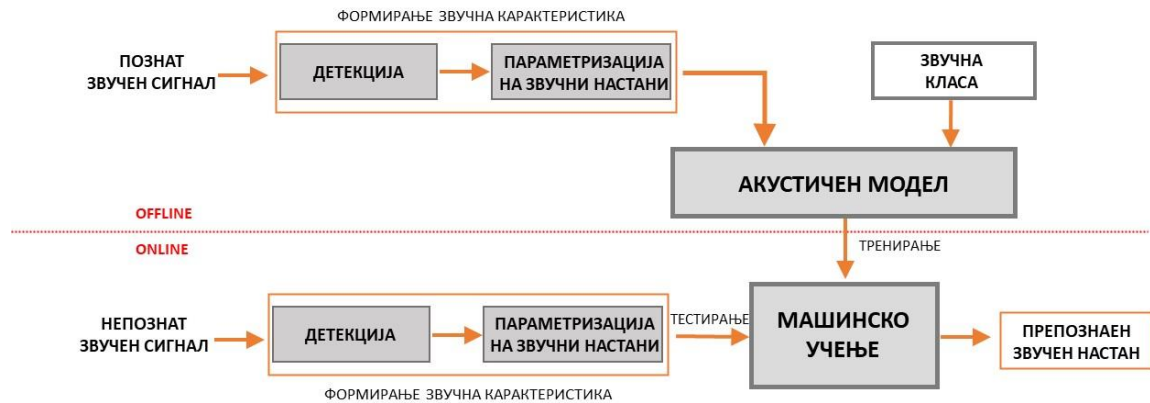
При употреба на овие системи за препознавање и класифицирање на звучните настани, потребно е самите звучни настани да бидат најпрвин систематизирани согласно предложена таксономија на звукот. Врз основа на предложената таксономија, потребно е да се креира база на податоци за избраните звучни настани во животната средина која се анализира. Звучните настани претставени како аудио сигнали од оваа база на податоци треба да бидат претходно обработени преку примена на технологии за обработка на дигитални сигнали. Вака обработените податоци формираат комплетна база на податоци која се применува како влез во системот за препознавање и класификација на звучните настани.

Во првата фаза се детектираат звучните настани од континуиран звук кој е снимен со микрофон. Детекцијата на звучните настани овозможува сегментирање и одвојување на звучните настани од интерес. Звучната детекција може да биде направена користејќи два пристапи: откривање на звучниот настан во континуиран временски период и откривање на звучниот настан во ограничен временски период. Првиот пристап се состои од детекција на промените на карактеристиките од аудио звукот преку користење на систем за откривање на нови настани дефинирани како секоја брза промена во однос на долгорочниот шум во позадина [43, 44]. Вториот пристап овозможува детекција на звучниот настан во ограничена временска рамка преку детекција на почетните и завршните точки на звучниот настан од интерес [45]. Целта на звучната детекција е да се открие звучниот настан преку пронаоѓање на почетните и крајните точки на настанот од целосниот спектар на бучава.

Следниот чекор е процесот на класификација. Најпрвин, карактеристиките на звучниот сегмент се параметризирани и компактно претставени преку употреба на параметризација на звучни настани. Изборот на карактеристиките на сигналот за звучните настани има клучна улога при креирање успешен систем за препознавање и класификација на звучните настани. Во истражувањето [46] се потврдува дека временскиот распоред на настаните е од големо значење при препознавање на звучните настани. Сите овие карактеристики кои се добиени од параметризацијата на звучните настани се поврзуваат во единствен вектор на карактеристики што го претставува звучниот настан како предмет за препознавање со одредени карактеристики.

Следно, врз основа на информациите обезбедени од параметризацијата на звучните настани, се применуваат техники на машинско учење. Зависно техниката која се применува, постои надгледуван и ненадгледуван процес при машинско учење. Надгледуван процес на машинско учење започнува со постоечки сет од податоци кои се соодветно поделени во класи и имаат одредени карактеристики кои даваат дополнително објаснување на класите. Ненадгледуваното учење е техника на машинско учење во која податоците со кои се работи немаат означени карактеристики.

Од интерес во ова истражување се системите за препознавање и класификација на звучни настани при надгледуван пристап на учење. На слика 1.2 се прикажани два карактеристични процеси за работа при примена на машинско учење при надгледуван пристап на учење, процес на тренирање и процес на тестирање (валидација).



Слика 1.2. Тренирање и тестирање на системот за препознавање и класификација на звучни настани

Независно од употребениот алгоритам за машинско учење со надгледуван пристап на учење, системот најпрвин треба да помине низ процес на тренирање. За таа цел, мора да се креира сет од познати звучни настани поделен во класи кои што сакаме системот да ги препознае. Употребувајќи го овој пристап се гради сет од примероци за секој звучен настан од секоја звучна класа. Со параметризација на звучните настани се формира акустичен модел кој се користи за тренирање на алгоритмот за машинско учење (овој процес е познат како офлајн процес). Следниот чекор е тестирање, односно препознавање на непознати звучни настани и нивна класификација во соодветна класа врз основа на информациите стекнати за време на тренирањето. Крајната цел е да се добие препознаен звучен настан врз основа на обработените тренирани податоци преку применетите алгоритми за машинско учење, кој ќе биде соодветно класифициран во класата на која припаѓа.

Параметрите на звучни настани што може да се извлечат во процесот на екстракција на карактеристики се анализирани од група истражувачи во [47]. Согласно анализата, тие можат да бидат поделени во 6 домени: временски, фреквентен, цепстрален, сопствен домен, модулно-фреквентен домен и домен на фаза-простор. Иако содржината на звучните настани од животната средина е поразновидна од говорните и музичните сигнали, параметрите кои се користат за препознавање на говорот и музиката, во големи рамки се користат и за класификација на звучните настани во животната средина. Истражувачите во [48] ја анализираат прецизноста на системот користејќи збирни параметри на звучните настани и алгоритми засновани на длабокото учење. Преку тестирање на точноста за класификација на вид на патни возила, применети се 13 параметри и 4 алгоритми на машинско учење, формирајќи 52 комбинации кои се тестираат за да се процени нивната ефикасност [49]. Преку примена на 5 класификатори и 6 параметри за класификација на урбани звучни настани, тестирана е точноста на изградените системи во [50]. Слично на ова истражување, анализирани се урбаните звучни настани преку примена на комбинација на параметри на звучни настани за

обработка на сигналите кои се користат како влез во модели базирани на длабокото учење [51]. Преку користење на неколку техники за екстракција на параметри како влез во алгоритмот за машинско учење, истражувачите во [52] проучуваат кој сет на параметри ќе даде најголема ефикасност на системот при класификација на урбаните звуци. Овој пристап користи МЕЛ фреквентни цепстрални коефициенти за екстракција на аудио карактеристиките во надгледувани алгоритми за класификација (машини со потпирачки вектори, случајни шуми и KNN) за идентификација на бучавата. Со користење на неколку бази на податоци со различни звучни класи во животната средина, научниците во [53, 54] го анализираат и визуелизираат параметарот на МЕЛ спектрограм како влез во мрежата на конволуциски невронски мрежи. Може да се забележи дека повеќето традиционални и често применувани техники за екстракција на аудио карактеристики што се користат за класификација на звучните настани во животната средина се параметрите на МЕЛ фреквентни цепстрални коефициенти и МЕЛ спектрограм. При изградба на системот за класификација на урбани звучни настани, најчесто применувани алгоритми се машини со потпирачки вектори [55], скриен Марков модел [56], како и длабокото учење преку примена на длабоки невронски мрежи [57] и конволуциски невронски мрежи [58]. На табела 1.1 е прикажан преглед на истражувања кои моделираат системи за препознавање и класификација на звучни настани во урбана средина, на која се прикажани параметрите на звучни настани и користениот алгоритам, а како резултат е прикажана точност добиена при тестирање на овие системи. Од табелата може да се забележи дека постигнатата точност во истражувањата е помеѓу 53% и 97%.

Табела 1.1. Преглед на истражувања во полето на препознавање и класификација на урбани звучни настани

Труд	Дата	Карактеристики	Алгоритам	Точност
[59]	Август, 2015	Мел Спектрограм	Случајни шуми	73.7%
[60]	Септември, 2016	MFCC	Машини со потпирачки вектори	59%
[61]	Август, 2017	МЕЛ спектрограм	Конволуциски невронски мрежи	81.9%
[62]	Септември, 2017	МЕЛ спектрограм	Конволуциски невронски мрежи	72.7%
[63]	Септември 2017	MFCC - спектрограм	Конволуциски невронски мрежи	92%
[64]	Јуни 2018	Raw data - LM	Конволуциски невронски мрежи	92.2%
[65]	Август 2018	LM-GS	Конволуциски невронски мрежи	83.7%
[66]	Септември, 2018	MFCC	Случајни шуми	90.43%
[67]	Февруари, 2019	МФЦЦ, хроматограм, спектрален контраст, тонал центроид, МЕЛ Спектрограм	Наивен Баесов алгоритам	53.77%
[67]	Февруари, 2019	МФЦЦ, хроматограм, спектрален Контраст	Машини со потпирачки вектори	85.91%
[68]	Април, 2019	MFCC и LMC	Конволуциски невронски мрежи	97.2%
[69]	Април 2019	Raw data – LM	Конволуциски невронски мрежи	89%
[70]	Декември, 2020	Кратка Фуријеова трансформација	Машини со потпирачки вектори	78.14%

Инспирирани од претходните истражувања, мотивацијата во оваа докторска дисертација е да се истражат техниките за екстракција на карактеристиките преку примена на процесот на параметризација на звучни настани кои би се тествале со цел да се одреди кој сет од параметри води кон постигнување највисока точност. Ефикасноста на овие карактеристики ќе се анализира преку користење на повеќе алгоритми на машинско учење, со цел да се избере алгоритам кој постигнува најдобри перформанси.

Користејќи го концептот на паметни градови преку примена на техниките на „Интернет на нештата“, системите за препознавање и класификација на урбаните звучни настани може да се реализираат во реално време. Примената на ниско-буџетни сензорски единици кои служат за препознавање и класификација на звучните настани се применети во бројни истражувања [71,72,73,74]. Инспирирани од современите истражувања, научниците во [75] имплементирале ниско-буџетни сензорски единици кои служат за мониторирање и препознавање на урбаната бучава. Согласно истражувањето во [76], во урбаната средина во Њујорк се распоредени 56 ниско-буџетни сензорски единици за детектирање на урбаната бучава во реално време. Преку развој на уред кој е поставен на повеќе локации во градот, авторите во [77] презентираат пристап за откривање на возила за итни случаи (амбулантни возила, полиција) преку детекција на звукот генериран од нивните сирени, каде целта е да се овозможи полесен транзит на овие возила. Истражувачите во [78] тествале три конфигурации за изведба на системите за препознавање и класификација на звучните настани базирана на „Интернет на нештата“ со цел да се искористат во апликација во реално време. Примената на овој систем овозможува следење на загадувањето предизвикано од бучавата во реално време и препознавање на изворите на бучава во големите и густо населените градови.

Според истражувањето во [79], истражувачите направиле развој на динамични карти за бучава врз основа на детектор за кој ја разликува бучавата предизвикана од сообраќајот од останатите настани на бучава. На овој начин, системите за препознавање на звучните настани можат да се користат како алатка за ажурирање на картите на бучава преку идентификација на изворите на бучава и нивото кое тие го предизвикуваат [80].

1.4. ПРЕДМЕТ НА ИСТРАЖУВАЊЕТО

Предмет на истражувањето во оваа докторска дисертација се техниките засновани на вештачката интелигенција при препознавање и класификација на звучните настани. Истражувањето е поттикнато од потребата за развој на ефикасен систем кој точно и прецизно ќе ги детектира и класифицира звучните настани во урбаните средини. Од посебен интерес на ова истражување ќе биде дигиталната обработка на сигналите и примената на разните алгоритми за машинско учење за добивање на систем кој со голема точност ќе ги предвидува резултатите.

За да се определи бучавата во урбаните средини, предмет на истражување ќе биде и анализа и примена на стандардните методи за проценка на бучавата преку мерења и определување на основни и дополнителни индикатори на бучава, креирање акустичен предиктивен модел и субјективна анализа на бучавата.

Примената на напредните технологии ќе овозможи поедноставување и надградување на стандардните методи преку развој на системи за препознавање и класификација на звучните настани кои се засновани на техниките на машинското учење. За изработка на овој систем, потребно е да се систематизираат звуците и да се избере соодветна база на податоци. Со цел правилна параметризација на звучните настани, предмет на истражувањето е дигитална обработка на звучните сигнали за идентификација и екстракција на соодветните карактеристики. Согласно спроведените анализи, ќе се изберат неколку параметри на звучни настани кои ќе се користат во експерименталниот дел при изградба на системот за препознавање и класификација на звучните настани.

Предмет на истражување се системите за препознавање и класификација на звучни настани при надгледуван пристап на учење за добивање систем кој со голема точност и прецизност ќе ги класифицира звучните настани во класите во кои тие припаѓаат. Со цел примена на процес за препознавање на звучниот сигнал преку тренирање и тестирање на податоците, предмет на истражувањето ќе биде и употреба на алгоритми на машинско учење. Преку експериментален развој на процесите кои го сочинуваат системот за препознавање и класификација на звучните настани, ќе се анализираат параметрите кои влијаат на точноста на системот и ќе се избере систем со најповолни перформанси кој понатаму може да се користи во комплексни системи кои ги обработуваат сигналите во реално време.

1.5. ЦЕЛИ НА ИСТРАЖУВАЊЕТО

За остварување на целите на оваа докторска дисертација ќе се реализираат детални теоретски и експериментални истражувања. Основната цел на дисертацијата е определување висок степен на точност при предвидување на звучните настани преку примена на техниките на вештачка интелигенција при развој на систем за препознавање и класификација на звучни настани. Преку дигитална обработка на звучните сигнали и употреба на разни алгоритми за машинско учење се очекува да се развие систем кој со голема точност ќе ги препознава и класифицира звучните настани.

Примената на стандардните методи за проценка на нивото на бучава во урбана средина ќе овозможат да се добијат основни сознанија за загадувањето предизвикано од бучава. Преку користење на стандардниот метод на мерење на нивото на бучава во животната средина се овозможува обработка на резултатите и пресметка на мерна неодреденост, каде целта е да се оцени нивото на бучава во анализирано подрачје. Преку изработката на акустичен предиктивен модел, целта е да се добијат карти на бучава кои го карактеризираат

подрачјето и овозможуваат детална картографска репрезентација на резултатите. Ќе се употреби компаративна анализа на добиените резултати од применетите пристапи со цел валидација на точноста на резултатите. Од друга страна, употребата на субјективна анализа за влијанието на бучавата ќе доведе до детектирање карактеристични извори на бучава кои предизвикуваат вознемиреност кај населението. Употребата на стандардните методи за определување на нивото на бучава во урбана средина ќе послужат како инспирација во понатамошните истражувања кон главната цел, односно примена на современи методи и техники за изградба на систем за препознавање и класификација на звучните настани заснован на вештачката интелигенција. Споменатите стандардни методи за проценка на нивото на бучава се употребуваат со цел да се покаже дека овие методи имаат ограничувања при определување на бучавата и дека може да се надградат и поедностават користејќи ги напредните технологии.

Напредните технологии за креирање на систем за препознавање и класификација на звучните настани се применуваат со цел развој на систем кој успешно ќе ги препознае доминантните извори на бучава. Истражувањето ќе даде осврт на теоретското и експерименталното моделирање на системите за препознавање и класификација на звучните настани, анализирајќи ги базите на податоци, дигиталната обработка на звучните сигнали и примената на вештачката интелигенција со цел развој и имплементација на процесите кои се составен дел од овие системи.

Преку анализа на звучната таксономија, целта е да се изберат класи на звучни настани кои ја формираат базата на податоци која се користи за тренирање и тестирање на развиените системи. Теоретската анализа на аудио параметрите ќе овозможи анализа на различните математички интерпретации кои се користат при екстракција на одредени аудио карактеристики со цел да се изберат соодветни параметри кои ќе се користат во развојот на системот за препознавање и класификација на звучните настани. Аудио параметрите претставуваат влезен параметар во системот заснован на вештачка интелигенција, согласно на тоа, правилниот избор на карактеристики е многу важен при постигнување прецизност на системот. Теоретската анализа на техниките засновани на вештачката интелигенција ќе доведе до избор на алгоритми за машинско учење кои можат да постигнат висока точност при експериментално тестирање на системот.

Експерименталната анализа која вклучува развој на системи на препознавање и класификација на звучните настани би овозможила поширок поглед и анализа на различните аудио параметри и алгоритми за машинско учење. Преку менување на аудио параметрите и карактеристиките на алгоритмите за машинско учење се очекува да се добие систем со висока точност на предвидените резултати.

Теоретските анализи ќе бидат потврдени со експериментални истражувања користејќи реални податоци со цел да се изврши верификација на применетите методологии и да се оцени степенот на подобрување на изградениот модел. Разработениот систем за препознавање и класификација на звучните настани во урбаните средини може да се користи за научноистражувачки и практични цели.

1.6. ХИПОТЕЗИ

ГЛАВНА ХИПОТЕЗА: Примената на вештачката интелигенција при развој на систем за препознавање и класификација на звучните настани може да послужи за навремено и прецизно детектирање на изворите кои предизвикуваат загадување од бучава.

ПРИДРУЖНА ХИПОТЕЗА: Изборот на параметри на звучни настани и алгоритми за машинско учење влијаат врз постигнатата точност и прецизност при изградба на ефикасен систем за препознавање и класификација на звучните настани.

1.7. МЕТОДОЛОГИЈА НА ИСТРАЖУВАЊЕ

Имајќи ја во предвид комплексноста на поставените цели на докторската дисертација, ќе се спроведат теоретски, аналитички и експериментални истражувања, кои се опфатени со прикажаната методологија:

- Анализа на досегашните достигнувања во областа на бучавата преку преглед на стандардни пристапи засновани на Европските стандарди и националните регулативи за проценка на нивото на бучава во урбана средина.
- Експериментални мерења на бучавата преку примена на букомер од класа 1 и нивна софтверска анализа за пресметка на основните индикатори на бучава и стандардната мерна неодреденост во избрано подрачје.
- Визуелизација на дисперзијата на бучавата во животната средина преку изработка на модел на акустична предиктивна карта на бучава, кој како влезен параметар го користи протокот на возила.
- Субјективна анализа преку примена на статистички научен метод за прибирање на податоци за изложеноста на бучава и детекција на доминантни извори кои предизвикуваат загадување од бучава.
- Компаративна анализа преку споредба на резултатите од применетите стандардни методи за валидација на резултатите.
- Анализа на досегашните достигнувања во областа на вештачката интелигенција во системите за препознавање и класификација на звучните настани со цел да се добие објективна слика за заклучоците, тврдењата и сознанијата.
- Дефинирање на звучна таксономија за избор на соодветна база на податоци од звучни настани поделени во класи која ќе се користи при тренирање, тестирање и валидирање на развиениот систем.

- Избор на параметри на звучни настани. Екстракција на аудио сигналите од базата на податоци преку користење повеќе параметри на звучни настани: МЕЛ Фреквентни Цепстрални коефициенти, МЕЛ спектрограм, Тонален Центроид, Хромаграм и Спектрален Контраст.
- Примена на математичка метода заснована на машинското учење за развој на алгоритам за тренирање и тестирање на податоците. Употреба на повеќе алгоритми за машинско учење: алгоритам на Случајни Шуми, алгоритам на Машини со Потпирачки Вектори, Наивен Баесов алгоритам и Конволуциски Невронски мрежи.
- Експериментално моделирање и развој на систем за препознавање и класификација на звучните настани преку имплементација во софтвер.
- Тренирање и тестирање на податоците од избраната база на податоци за валидација и верификација на резултатите.
- Оптимизација на системот преку менување на параметрите на звучни настани и подесување на алгоритмите на машинско учење преку менување на нивните параметри за добивање систем со висока точност при предвидување на податоците.
- Компаративна анализа за споредба на изградените модели и оцена на најефикасните параметри на звучни настани и алгоритми за машинско учење.
- Валидација на системот за препознавање и класификација на звучни настани со цел негова примена во концептот на паметен град базиран на „Интернет на Нештата“.

1.8. ПРИДОБИВКИ ОД ИСТРАЖУВАЊЕТО

Барањата за поточна оцена на состојбата и попрецизна детекција на звучните настани кои предизвикуваат бучава ја наметнува потребата од употреба на современи пристапи во техничките системи за препознавање и класификација на звучните настани. Стандардните методи на проценка на бучавата ќе овозможат приказ на нивната комплексност, од каде што произлегува употребата на напредните технологии за проценка на бучавата.

Ова истражување со својот теоретски и експериментален дел ќе даде посебен придонес во примена на вештачката интелигенција при изградба на системи за препознавање и класификација на звучните настани. Развојот на системите за препознавање и класификација на звучните настани бара мултидисциплинарен пристап кој треба да опфати теоретски анализи на системи засновани на вештачка интелигенција, обработка на дигитални сигнали и алгоритми за машинско учење, експериментални анализи преку развој на модели засновани на машинско учење и софтверско решение преку користење на бази на податоци при тренирање и тестирање на системите за споредба на точноста на предвидените резултати.

Спроведените теоретски анализи на системот за препознавање и класификација на звучните настани ќе дадат јасна претстава за влијанието на одредени параметри при избор на параметри на звучни настани и алгоритми на машинско учење кои ќе се користат при експериментално моделирање на системот.

Експерименталното моделирање преку изградба на системот за препознавање и класификација на звучните настани е релативно нова техника со голем потенцијал за примена во современи апликации. Од анализата на параметрите на звучни настани кои се добиени со дигитална обработка на сигналите и нивна екстракција ќе се предложат параметри кои ќе се користат при изградба на векторот на карактеристики. Преку тренирање и тестирање на системот користејќи повеќе алгоритми за машинско учење, ќе се изврши оценка на перформансите при изградба на системите засновани на вештачка интелигенција. Придобивката од употребата на разните параметри на звучни настани и алгоритми на машинско учење е да се добие систем кој ќе постигне висока точност при класификација на звучните настани и може да се користи за мониторинг во реално време.

Креирањето на интегриран систем за препознавање и класификација на звучните настани вклучува иновативен инженерски пристап за решавање на актуелни проблеми во паметните градови изградени на концептот на „Интернет на нештата“ кој е поекономичен за изградба и одржување и бара минимална интервенција од човекот. Имплементацијата на овие системи во реални апликации може да резултира во намалување или елиминирање на конвенционалните системи за проверка и контрола на бучавата, кои може да бидат доста скапи и да одземат многу време при нивно извршување. Поради тоа, системите за препознавање и класификација на звучните настани се исклучително важни во инженерски апликации за обработка на податоците во реално време.

2. СТАНДАРДНИ МЕТОДИ ЗА ПРОЦЕНКА НА БУЧАВАТА ВО УРБАНА ЖИВОТНА СРЕДИНА

Во оваа глава се разработени стандардните методи за проценка на бучавата во животната средина преку објективна и субјективна анализа. Објективната анализа го детектира интензитетот на бучавата преку мерење на нивото на бучавата и изработка на карти на бучава, додека субјективната анализа го анализира влијанието на бучавата во урбаната животна средина врз населението.

Развојот и експерименталната примена на споменатите методологии се примени на избрано подрачје за анализа. Преку спроведување на методите за објективна анализа преку мерење на нивото на бучава во избраното подрачје, пресметани се основните и дополнителните индикатори на бучава со нивната мерна неодреденост. Понатаму, преку анализа на бројот на возила и нивно внесување во софтвер за изработка на карти на бучава, како и внесување на топографијата на избраното подрачје, добиени се акустични предиктивни модели за нивото на бучава. Споредбата на резултатите овозможи потврдување на точноста на развиените методологии и проценка на нивото на бучавата во подрачјето кое се анализира.

Како резултат од развојот на методологија за субјективна анализа и нејзина примена во подрачјето за анализа, добиени се сознанија за влијанието на бучавата врз населението. Од субјективната анализа може да се заклучи дека постојат повеќе извори на бучава кои штетно влијаат на населението, од каде што се појавува потреба за дополнителни истражувања за идентификација на доминантните извори на бучава преку примена на напредни технологии кои се базираат на вештачката интелигенција.

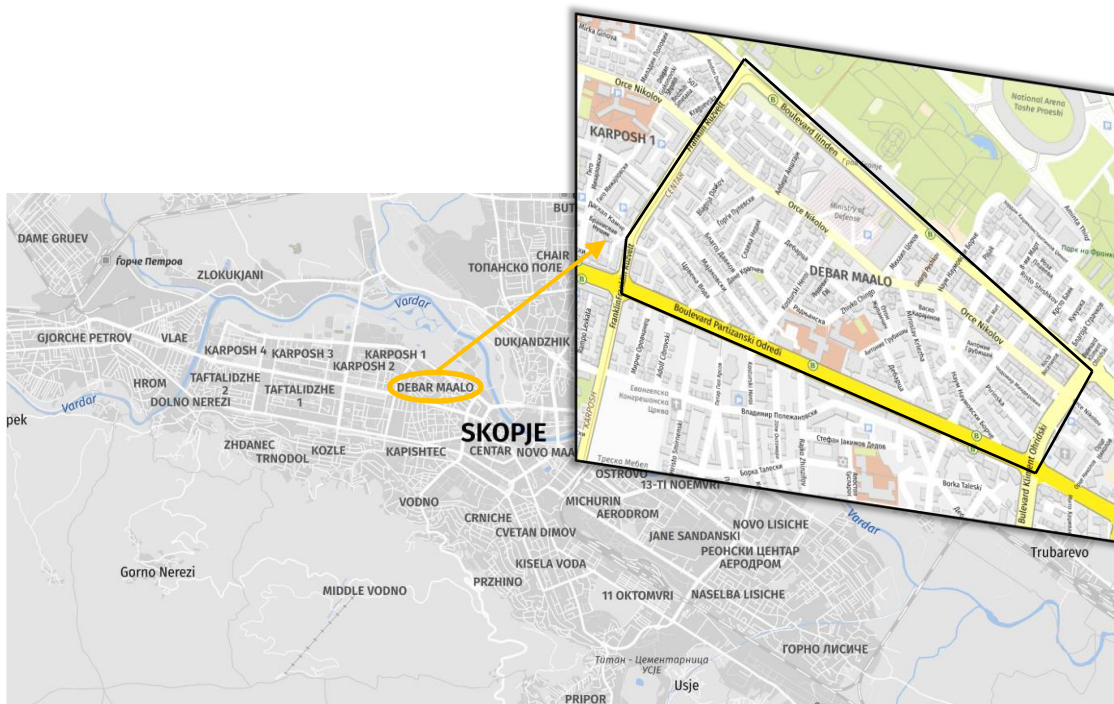
2.1. ИЗБОР НА ПОДРАЧЈЕ ЗА АНАЛИЗА

Урбаниот начин на живеење генерира бучава, чиј интензитет е во директна зависност од густината на населението, непосредната близина на објектите за домување, индустриските капацитети и сообраќајот. Поради зголемената концентрација на различните видови на енергија, интензивирање на сообраќајот, употреба на моќни апарати и сообраќајни средства има значителен пораст на нивото на бучава во урбаните средини.

Градот Скопје, како главен град на Република Северна Македонија претставува густо населена урбана средина која се соочува со проблеми кои се јавуваат поради зголеменото ниво на бучава. Во 2018 година, согласно Правилникот за локации на мерните станици и мерните места [81], нивоата на бучава се измерени на 14 мерни места, од кои 11 покажуваат зголемена гранична вредност на бучава. Субјективна анализа за влијанието на бучавата е спроведена во [82], потврдувајќи дека бучавата во урбаната средина предизвикува вознемиреност, особено за повозрасната популација во Скопје. Истражувањето покажува дека бучавата во урбаната животна средина во градот Скопје има висока опасност по здравјето на населението, а проценката преку мерење на нивото на бучава и анализа на вознемиреноста може да се покаже како корисна алатка при креирање урбанистички план.

Населението кое живее во урбаниот дел од градот Скопје е постојано изложено на високо ниво на бучава. Како репрезентативна општина која е постојано изложена на зголемено ниво на бучава поради големата фреквенција на автомобили и население е избрана општина Центар. Подрачје од посебен интерес во општина Центар претставува населбата „Дебар Маало“ која е избрана поради централната локација во градот, големиот број на угостителски објекти кои доведуват до голема посетеност и густо сообраќај. Дополнително, преку анализа на резултатите од ниско-буџетните сензори во градот Скопје во времетраење од три години [83] беше заклучено дека мерната станица во подрачјето „Дебар Маало“ има највисоко ниво на бучава во последните три години. Овие резултати се дополнителен индикатор кој потврдува дека ова подрачје е постојано изложено на зголемено ниво на бучава.

Населбата е ограничена со Булеварите „Илинден“ на север, „Партизански Одреди“ на југ и „Климент Охридски“ на исток, а со улицата „Френклин Рузвелт“ на запад. Булевар „Илинден“ го дели Градскиот Парк кој претставува тивка зона од големиот број на угостителски објекти и кафулиња кои во текот на целиот ден имаат голема посетеност. Покрај резиденцијалните и угостителските објекти, во населбата „Дебар Маало“ се наоѓаат и објекти наменети за воспитна и образовна дејност: основното училиште „Коле Неделковски“, Детска градинка „13-ти Ноември“, два факултети (Градежен факултет и Архитектонски факултет) и Библиотеката „Браќа Миладиновци“. На слика 2.1. е прикажана географската локација на урбаното подрачје во кое ќе се анализира бучавата.



Слика 2.1. Географска локација на избраното подрачје за анализа

Потребно е да се спроведе детално истражување во урбаното подрачје, каде што, преку споредување на објективната и субјективната анализа, ќе се направи проценка на нивото на бучава. Преку примената на овие стандардни методи за проценка на нивото на бучава во животната средина ќе се воспостави корелација помеѓу истражувањата со цел подобро разбирање и карактеризирање на загадувањето во избрано подрачје за анализа.

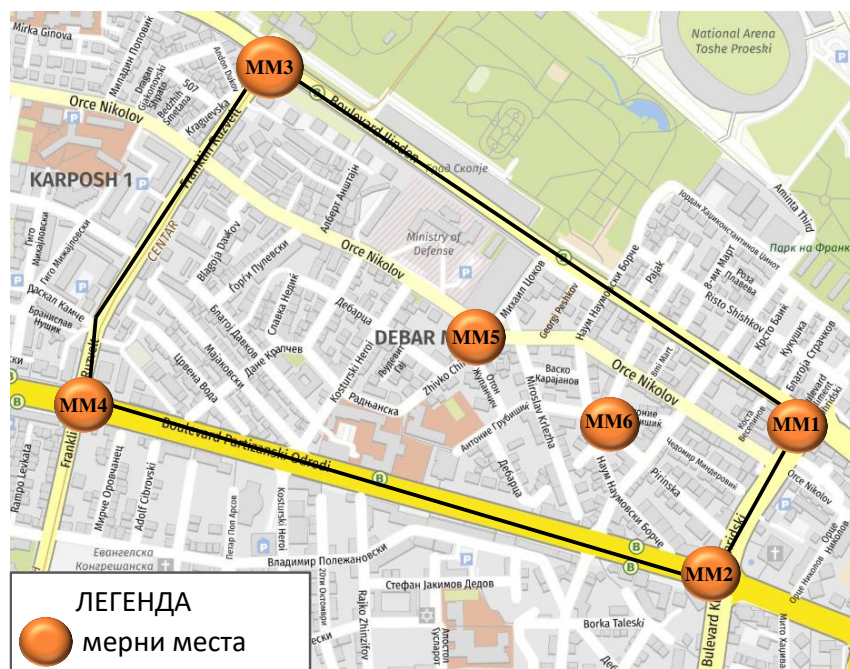
2.2. ПРОЦЕНКА НА НИВОТО НА БУЧАВА ПРЕКУ СТАНДАРДНИ МЕРЕЊА

Стандардни мерења на нивото на бучава се изведуваат преку директни мерења со букомер од 1ва класа на претходно одбрани локации. Согласно стандардот [6], при мерење на бучава во животната средина, потребно е да се измери бучавата од патниот сообраќај како доминантен извор на бучава. По спроведување на краткотрајни мерења на нивото на бучава, се пресметуваат основните и дополнителните индикатори на бучава кои ја нормираат и квантифицираат бучавата. Преку определување на мерната неодреденост, резултатот од спроведените мерења и анализи се потврдува со 95% ниво на доверба.

2.2.1. ИЗБОР НА МЕРНИ МЕСТА

Мерењето и определувањето на нивото на бучава од патен сообраќај во населбата Дебар Маало е извршено во согласност со стандардот МКС ISO 1996-2:2018 – Акустика – Опис, мерење и проценување на бучавата во животната средина, Дел 2: Одредување на нивоата на бучава во животната средина [6]. Изборот на мерните места е направен согласно барањата на стандардот МКС ISO 1996-2:2010 и Правилникот за локациите на мерните станици и мерните места [81]. Одбраната локација припаѓа во подрачје со втор степен на заштита на бучава, кое согласно Правилникот за локации на мерни станици и мерни места е примарно наменето за престој, односно станбен реон, подрачје во околина на објекти наменети за воспитна и образовна дејност, објекти во кои се вршат угостителски дејности, постојат јавни зеленила и рекреациjsки површини и подрачја на локални паркови. Гранични вредности за подрачје од втор степен за заштита од бучава се: 60 dB во дневниот период, 60 dB во вечерниот период и 55 dB во ноќниот период.

За да се процени бучавата потребно е да се постават најмалку пет мерни места кои ќе бидат рамномерно поставени на локации каде што има континуирана бучава. За целите во ова истражување, одбрани се 6 мерни места, распределени како што е прикажано на слика 2.2.



Слика 2.2. Распоред на мерни точки


Согласно сликата 2.2., мерните точки MM1, MM2, MM3 и MM4 се поставени на четири фреквентни раскрсници, односно на места каде што има постојан проток на возила. Останатите две мерни точки, MM5 и MM6 се поставени во централниот дел на избраното подрачје и истите се поставени на места каде што покрај протокот на возила, има многу ресторани кои придонесуваат за зголемен број на посетители на овие локации.

За проценка на бучавата во одбраното подрачје, направени се краткотрајни мерења во три временски периоди (ден, вечер и ноќ). Минималното време за мерење е 10 минути, или време помало од 10 минути, доколку тоа го наложува технолошкиот процес или настанот кој се анализира. Согласно локацијата на која се направени мерењата и изворот на бучава, индивидуалните мерења се во времетраење од 10 минути. Следејќи ги препораките од стандардот [6], мерењата беа направени на висина од 1.5 метри на оддалеченост од 3.5 метри од ѕидот на најизложената фасада и другите рефлектирачки површини.

2.2.2. МЕРНА ОПРЕМА

Мерна инструментација која се користи за мерење на нивото на бучава се состои од калибриран рачен анализатор на бучава кој ги исполнува барањата за инструмент од класа 1 според стандардот IEC 61672-1 со неговата дополнителна опрема. На табела 2.1. е прикажан инструментот за мерење на бучава и неговите основни информации.

Табела 2.1. Информации за инструмент за мерење на бучава

Рачен анализатор за бучава со микрофон			
Назив	Анализатор	Микрофон	
Модел	2250	4190	
Производител	Brüel & Kjær	Brüel & Kjær	
Стандард	IEC 61672-1	IEC 61672-1	

Потребно е да се направи интерна калибрација на инструментот за мерење на бучава за негова проверка, која се прави со акустичен калибратор кој обезбедува функции и карактеристики согласно стандардот IEC 60942, а основните информации се прикажани на табела 2.2.

Табела 2.2. Информации за акустичен калибратор

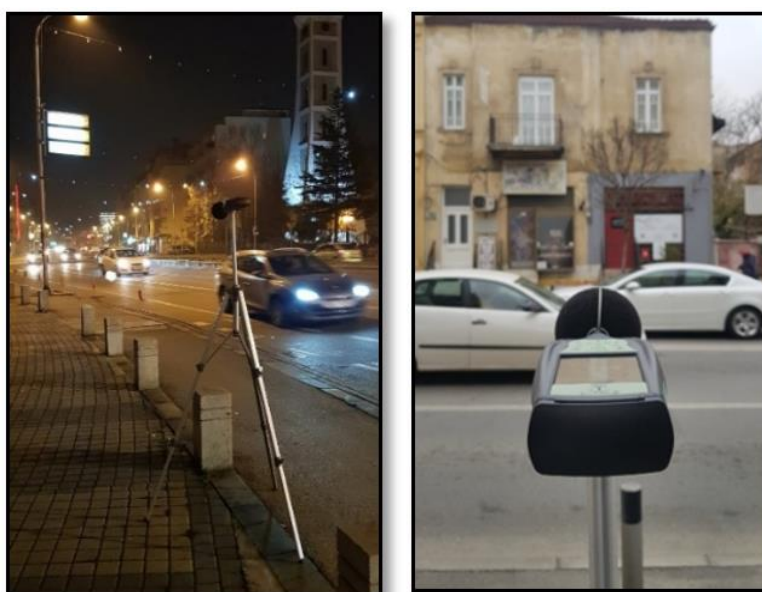
Назив	Интерен калибратор	
Модел	SV36	
Производител	Svantek Sound and Vibration	
Стандард	IEC 60942	

Корекциониот фактор на бучава е определен преку интерна калибрација на мерниот систем пред и по мерењето, за да се потврди дали инструментот ги исполнува условите, односно дали направените мерења ќе се прифатат. Системот за мерење на нивото на бучава се проверува на почетокот и крајот на секое мерење, при што се прави калибрација на фреквенција на 94 dB и 114 dB со помош на калибратор на звук. Разликата помеѓу отчитувањата на две последователни проверки треба да биде помала или еднаква на 0.5 dB без дополнително прилагодување. Доколку се надмине оваа вредност, резултатите од мерењата ќе се сметаат за невалидни и ќе бидат отфрлени. Согласно резултатите прикажани на табела 2.3, може да се забележи дека нивото на корекционен фактор е прифатливо и резултатите добиени од мерењата се валидни.

Табела 2.3. Резултати од извршената интерна калибрација на инструментот

Инструмент		Hand-held analyzer Brüel & Kjær 2250			
Калибратор		SV36			
Ред. бр.	Фреквенција (dB)	Корекционен фактор (dB)		Разлика помеѓу проверките ($\leq 0,5$ dB)	Прифатливо/ Неприфатливо
		Пред мерење	После мерење		
1	94	0,10	0,06	0,04	<input type="checkbox"/> ДА ($\leq 0,5$ dB) <input type="checkbox"/> НЕ
2	114	0,02	-0,04	0,06	<input type="checkbox"/> ДА ($\leq 0,5$ dB) <input type="checkbox"/> НЕ

Мерната опрема се постави на мерните локации на кои се спроведуваат мерењата. На слика 2.3. е прикажана поставеноста на мерниот инструмент при мерењата во избраното урбано подрачје за анализа.



Слика 2.3. Поставеност на мерната инструментација

При мерење на бучава во животната средина, многу често се појавува резидуален, односно дополнителен звук од околината кој предизвикува појава на грешка во измерените резултати. Појавата на овој проблем се појавува најчесто при мерење во надворешна средина поради потребата за одделување на бучавата според изворот кој ја предизвикува истата. За определување на мерната неодреденост и коригирање на резултатите од мерењата како резултат од резидуалниот звук создаден од околината, потребно е истиот да се детектира и измери. При мерење на нивото на бучава на одреден настан во животната средина, треба да постои доволна разлика помеѓу бучавата од самиот настан и резидуалниот звук. Оваа разлика треба да биде најмалку 3 dB, а по можност повеќе од 5 dB. Доколку разликата е помалку од 3 dB, мерната неодреденост на резултатот ќе биде голема и мерењето треба да се отфрли.

2.2.3. МЕРНИ ИНДИКАТОРИ ЗА ОПРЕДЕЛУВАЊЕ НА БУЧАВА ВО ЖИВОТНА СРЕДИНА

Определување на нивото на бучава во животната средина се прави преку определување мерните големини и параметри, односно детектирање на основните и дополнителните индикатори на бучава. Со цел да се определат мерните индикатори, потребно е да се земат во предвид резултатите од мерењата. Согласно Правилникот за примена на индикаторите за бучава, дополнителни индикатори за бучава, начинот за мерење на бучава и методите за оценување со индикаторите за бучава во животната средина [8], постојат 4 основни индикатори на бучава:

1. **Индикатор за бучава за ден-вечер-ноќ ($L_{\text{двн}}$)** во dB(A) претставува еквивалентно ниво на бучава за периодот ден, вечер, ноќ или за едно деноноќие пресметан според изразот:

$$L_{\text{двн}} = 10 \lg \frac{1}{24} \left(12 \cdot 10^{\frac{L_{\text{ден}}}{10}} + 4 \cdot 10^{\frac{L_{\text{в}}+5}{10}} + 8 \cdot 10^{\frac{L_{\text{ноќ}}+10}{10}} \right) \quad (2-1)$$

2. **Индикатор за бучава преку ден ($L_{\text{ден}}$)** во dB(A) претставува A-долготрајно просечно еквивалентно ниво на бучава определено преку целокупните дневни периоди (од 07:00 до 19:00 часот). За N кратки мерења, дневното ниво на бучава се определува според усреднето еквивалентно ниво на бучава (L_i) и параметрите (Δ_i и p_i) кои се одредуваат согласно метеоролошките услови со изразот:

$$L_{\text{ден}} = 10 \lg \left(\sum_{i=1}^N p_i 10^{0,1(L_i + \Delta_i)} \right) \text{ dB} \quad (2-2)$$

3. **Индикатор за бучава во текот на вечерта ($L_{\text{в}}$)** во dB(A) претставува A-долготрајно просечно еквивалентно ниво на бучава определено преку целокупните вечерни периоди (од 19:00 до 23:00 часот). За одредување на вечерното ниво на бучава се користат истиот израз како при определување на дневното ниво на бучава.

4. **Индикатор за бучава преку ноќ или ноќно ниво ($L_{\text{ноќ}}$)** во dB(A) претставува A-долготрајно просечно еквивалентно ниво на бучава определено преку целокупните ноќни периоди (од 23:00 до 07:00 часот). За одредување на ноќното ниво на бучава се користат истиот израз како при определување на дневното ниво на бучава.

Покрај основните индикатори за бучава, во одредени случаи може да се користат и дополнителни индикатори за бучава: еквивалентно ниво на бучава во одреден временски период (L_{eq}) и максимално ниво на бучава (L_{max}).

1. **Еквивалентното ниво на бучава за одреден временски период (L_{eq})** претставува пресметана големина од измерените нивоа на бучава ($(L_{A,eq,T})_i$) за одреден временски интервал (ΔT_i) за N мерења, со енергија еднаква на енергијата на измерената бучава пресметана според следниот израз:

$$L_{eq} = 10 \lg \frac{\sum_{i=1}^N \Delta T_i 10^{0,1L_{A,eq,i}}}{\sum_{i=1}^N \Delta T_i} \quad (2-3)$$

2. **Максималното ниво на бучава (L_{max})** претставува максимално одредено ниво на бучава според А тежинската крива за одреден временски период.

Индикаторите за бучава служат за определување на нивото на бучава во избраната животна средина во одделни периоди, како и проценка на изложеноста на населението на бучава.

2.2.4. ОПРЕДЕЛУВАЊЕ МЕРНА НЕОДРЕДЕНОСТ

Мерната неодреденост е параметар кој ја прикажува разликата помеѓу измерената и „вистинската вредност“, а точноста при мерењето се одредува преку разликата на овие две вредности. Резултатот од мерењето на бучавата во животната средина е само проценка на вистинската вредност и единствено е комплетен само кога е даден заедно со мерната неодреденост. Математичкиот генерален модел за определување на мерната неодреденост е согласно стандардот [6], за кој е потребно да се идентификува и да се определи мерна неодреденост на секој извор на несигурност. Формулата за определување на вредноста L која треба да се определи е функција f од измерените параметри x_j е:

$$L = f(x_1, x_2, x_3, \dots, x_j) \quad (2-4)$$

Доколку секој од измерените параметри x_j има стандардна неодреденост u_j , комбираната стандардна неодреденост изнесува:

$$u(L) = \sqrt{\sum_{j=1}^n (c_j u_j)^2} \quad (2-5)$$

Коефициентот на чувствителност c_j се пресметува како функција од измерениот параметар x_j :

$$c_j = \frac{df}{dx_j} \quad (2-6)$$

Вкупната мерна неодреденост е прикажана како проширена мерна неодреденост U , која се добива како производ од комбинираната стандардна неодреденост u и нумеричкиот фактор на покриеност k со изразот:

$$U = ku \quad (2-7)$$

Вообичаено се користи фактор на покриеност 2 ($k = 2$), што одговара на веродостојност на резултатот од 95%. На коригираниот резултат се додава проширената мерна неодреденост во следниот облик: $L \pm 2u$.

При определување на мерната неодреденост на бучава во животната средина, многу е комплексно да се определи формула за функцијата f на која влијаат измерените параметри. За едно независно мерење, пресметаната вредност на нивото на бучава L изразена во dB се пресметува преку измерената вредност L' , вредноста на резидуалниот звук L_{res} , мерната неодреденост предизвикана од изворот на бучава δ_{sou} , метеоролошките карактеристики δ_{met} и локацијата на ресиверот δ_{loc} , дадена со изразот:

$$L = L' + 10 \lg(1 - 10^{-0,1(L' - L_{res})}) dB + \delta_{sou} + \delta_{met} + \delta_{loc} \quad (2-8)$$

Пресметката на изразот (2-8) е едноставна, во која секој извор на мерна неодреденост е функција од неколку извори на неодреденост. Оваа формула може да се примени за било кое мерење во времетраење од неколку секунди до неколку години.

Мерната неодреденост која зависи од изворот и од метеоролошките карактеристики ($\delta_{sou} + \delta_{met}$) се определува директно од мерењата. Измереното ниво на звучен притисок L' и нивото на резидуален звук L_{res} зависат од мерната неодреденост која што произлегува од мерниот инструмент δ_{slm} .

Зависно од изворите кои предизвикуваат мерна неодреденост при мерењето на бучавата во животната средина и изразот (2-8), комбинираната мерна неодреденост се пресметува како квадратен корен од збирот на квадратите на вредностите на стандардната неодреденост u и коефициентот на чувствителност c за секој извор кој придонесува за појава на бучава според изразот:

$$u(L) = \sqrt{(c_{slm}u_{slm})^2 + (c_{sou}u_{sou})^2 + (c_{met}u_{met})^2 + (c_{loc}u_{loc})^2 + (c_{res}u_{res})^2} \quad (2-9)$$

На табелата 2.4 е прикажан буџетот на мерна неодреденост, кој се применува за определување на мерната неодреденост при мерење на нивото на бучава.

Табела 2.4. Буџет на мерна неодреденост

Вредност	Проценета вредност dB	Стандардна мерна неодреденост u_j dB	Големина на коефициентот на чувствителност, c_j	Пресметана мерна неодреденост $c_j u_j$, dB
$L' + \delta_{slm}$	L'	$u(L')$ 0,5	$\frac{1}{1 - 10^{-0,1(L' - L_{res})}}$	
δ_{sou}	0	u_{sou}	1	
δ_{met}	0	u_{met}	1	
δ_{loc}	0,0 – 6,0	u_{loc}	1	
$L_{res} + \delta_{res}$	L_{res}	u_{res}	$\frac{10^{-0,1(L' - L_{res})}}{1 - 10^{-0,1(L' - L_{res})}}$	
Комбинирана мерна неодреденост			$u(L) = \sqrt{\sum_1^n (c_j u_j)^2}$	
Проширена мерна неодреденост (95% сигурност $k=2$), $2u$				
Краен резултат				$L \pm 2u$

2.2.5. РЕЗУЛТАТИ ОД МЕРЕЊАТА

Согласно резултатите добиени од мерењата на нивото на бучава, пресметани се основните и дополнителните индикатори кои даваат информација за нивото на бучава во анализираното подрачје. Дополнително, пресметана е мерната неодреденост за основните индикатори на бучава во текот на дневниот, вечерниот и ноќниот период за сите мерни точки, а резултатите се прикажани на табела 2.5. На овој начин, може да се потврди дали нивото на бучава ја надминува граничната вредност на мерните места со 95% ниво на доверба.

Од табела 2.5. може да се забележи дека мерната неодреденост се движи во граници помеѓу $4,16 \div 4,41$ dB. Нивото на бучава ја надминува граничната вредност со 95% ниво на доверба преку пресметка на мерната неодреденост за речиси сите мерења на сите мерни места (оваа констатација важи за 17 од 18 мерења), додека за мерното место ММ6 во текот на дневниот период нивото на бучава ја надминува граничната вредност, но не може да се потврди со ниво на доверба од 95% за проширената мерна неодреденост.

Табела 2.5. Пресметана мерна неодреденост од спроведените мерења

МЕРНО МЕСТО ММ1	ден	вечер	ноќ
Коригирана вредност $L_{A,eq}$ [dB(A)]	77,40	71,97	70,28
Комбинирана мерна неодреденост (u)	2,16	2,08	2,09
Проширена мерна неодреденост 2u (95% сигурност)	4,32	4,16	4,18
Краен резултат	77,40 ± 4,32	71,97 ± 4,16	70,28 ± 4,18
МЕРНО МЕСТО ММ2	ден	вечер	ноќ
Коригирана вредност $L_{A,eq}$ [dB(A)]	69,65	69,70	68,61
Комбинирана мерна неодреденост (u)	2,19	2,10	2,10
Проширена мерна неодреденост 2u (95% сигурност)	4,38	4,21	4,21
Краен резултат	69,65 ± 4,38	69,70 ± 4,21	68,61 ± 4,21
МЕРНО МЕСТО ММ3	ден	вечер	ноќ
Коригирана вредност $L_{A,eq}$ [dB(A)]	66,73	66,57	64,87
Комбинирана мерна неодреденост (u)	2,20	2,08	2,09
Проширена мерна неодреденост 2u (95% сигурност)	4,41	4,16	4,19
Краен резултат	66,73 ± 4,41	66,57 ± 4,16	64,87 ± 4,19
МЕРНО МЕСТО ММ4	ден	вечер	ноќ
Коригирана вредност $L_{A,eq}$ [dB(A)]	73,55	70,53	66,65
Комбинирана мерна неодреденост (u)	2,22	2,10	2,09
Проширена мерна неодреденост 2u (95% сигурност)	4,44	4,20	4,19
Краен резултат	73,55 ± 4,44	70,53 ± 4,20	66,65 ± 4,19
МЕРНО МЕСТО ММ5	ден	вечер	ноќ
Коригирана вредност $L_{A,eq}$ [dB(A)]	67,15	71,99	66,66
Комбинирана мерна неодреденост (u)	2,10	2,11	2,14
Проширена мерна неодреденост 2u (95% сигурност)	4,20	4,22	4,29
Краен резултат	67,15 ± 4,20	71,99 ± 4,22	66,66 ± 4,29
МЕРНО МЕСТО ММ6	ден	вечер	ноќ
Коригирана вредност $L_{A,eq}$ [dB(A)]	58,76	67,19	57,55
Комбинирана мерна неодреденост (u)	2,12	2,14	2,15
Проширена мерна неодреденост 2u (95% сигурност)	4,24	4,28	4,30
Краен резултат	58,76 ± 4,24	67,19 ± 4,28	57,55 ± 4,30

Врз основа на резултатите добиени од извршените мерења и пресметката на индикаторите на бучава на мерните места во населбата Дебар Маало, може да се констатира дека има надминување бучавата на сите мерни места помеѓу 3,5% ÷ 17,8% во однос на граничните вредности.

Дополнително, на табела 2.6., прикажани се дополнителните индикатори на бучава, односно индикаторот на бучава за ден-вечер-ноќ и максималното ниво на бучава во трите временски периоди на мерење.

Табела 2.6. Дополнителни индикатори од спроведените мерења

МЕРНО МЕСТО ММ1	Вредност dB(A)
Индикатор на бучава ден-вечер-ноќ	78,53
Максимално ниво на бучава (ден)	84,7
Максимално ниво на бучава (вечер)	90
Максимално ниво на бучава (ноќ)	86,6
МЕРНО МЕСТО ММ2	
Индикатор на бучава ден-вечер-ноќ	75,28
Максимално ниво на бучава (ден)	82,4
Максимално ниво на бучава (вечер)	88,4
Максимално ниво на бучава (ноќ)	86,1
МЕРНО МЕСТО ММ3	
Индикатор на бучава ден-вечер-ноќ	71,75
Максимално ниво на бучава (ден)	82,6
Максимално ниво на бучава (вечер)	83,8
Максимално ниво на бучава (ноќ)	78,9
МЕРНО МЕСТО ММ4	
Индикатор на бучава ден-вечер-ноќ	75,14
Максимално ниво на бучава (ден)	95,4
Максимално ниво на бучава (вечер)	84,8
Максимално ниво на бучава (ноќ)	83,4
МЕРНО МЕСТО ММ5	
Индикатор на бучава ден-вечер-ноќ	74,21
Максимално ниво на бучава (ден)	81,7
Максимално ниво на бучава (вечер)	97,6
Максимално ниво на бучава (ноќ)	81,6
МЕРНО МЕСТО ММ6	
Индикатор на бучава ден-вечер-ноќ	67,02
Максимално ниво на бучава (ден)	74,7
Максимално ниво на бучава (вечер)	78
Максимално ниво на бучава (ноќ)	76,5

Од резултатите, може да се забележи дека максималното ниво на бучава е во дозволените граници, односно нивото е помало од 110 dB.

Добиените резултати од применетата методологија ја квантифицираат бучавата во анализираното подрачје и покажуваат драстично надминување на граничните вредности на нивото на бучава.

За да се потврди точноста на резултатите, во следното поглавје ќе се изработи акустичен предиктивен модел за анализираното подрачје во истиот временски период во кој се направени мерењата. Резултатите за нивото на бучава добиени од акустичниот предиктивен модел за бучава кој го користи протокот на возила како главен извор на бучава ќе послужи за компарација на вредностите за нивото на бучава и потврдување на нивната точност.

2.3. АКУСТИЧЕН ПРЕДИКТИВЕН МОДЕЛ

2.3.1. МЕТОД ЗА ИЗРАБОТКА НА АКУСТИЧЕН ПРЕДИКТИВЕН МОДЕЛ

Картите на бучава претставуваат појдовна основа за информирање на јавноста за состојбата со нивото на бучава и за изработка на акциони планови за намалување на нивото на бучава. Согласно Европската Директива, како главен и основен концепт кој треба да се разработи за проценка на нивото на бучава е изработка на стратешка карта на бучава за глобално оценување на утврдената изложеност на бучавата во определено подрачје како резултат од различни извори на бучава и/или за општи предвидувања во тоа подрачје. Во Р. Северна Македонија е донесен Закон за заштита од бучава во животната средина, кој како појдовна основа за заштита од бучава предлага изработка и користење на стратешки карти за бучава. Стратешките карти на бучава се изработуваат преку процес на мапирање и моделирање на дисперзијата на бучава на секои 5 години. Со анализа на изготвените стратешки мапи се овозможува локализирање на подрачја кои се постојано изложени на бучава. На овој начин се определуваат приоритетните подрачја во кои е забележано надминување на вредностите на бучавата преку определен индикатор, а може да предизвика штетно дејство врз здравјето на населението.

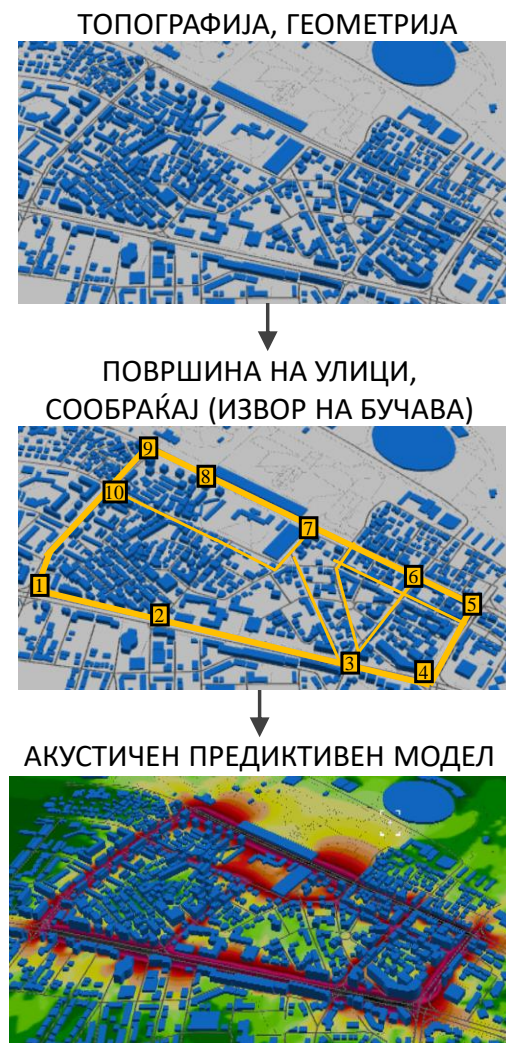
За изработка на стратешка карта на бучава, најпрвин потребно е да се изработи акустичен предиктивен модел кој е модел на карта на која е прикажана бучавата која е предизвикана од патниот сообраќај. Со изработка на акустичен предиктивен модел може да се добијат иницијални резултати за нивото на бучава предизвикана од патниот сообраќај. Анализата на патниот сообраќај е согласно Францускиот национален метод за пресметување на индикатори на бучава предизвикана од уличен (патен) сообраќај NMPB Routes [84,85]. Изложеноста на средината со бучава предизвикана од патниот сообраќај се оценува со пресметување на дневното, вечерното и ноќното ниво на бучава за карактеристичен временски период од едно деноноќие.

Оценката на нивото на бучава при креирање на акустичен предиктивен модел се состои од неколку фази за определување на еквивалентното ниво на бучава од автомобилскиот сообраќај на одделни делници во зависност од интензитетот и структурата на протокот на возила, корекција на еквивалентно ниво на бучава според брзината на движење, видот на патната облога, нагиб и влијанието на регулираните раскрсници.

Користејќи го Францускиот метод за предвидување на бучавата во сообраќајот на патиштата транспониран во националните законски акти, протокот на сообраќај се смета за главен параметар при креирање на акустичен предиктивен модел. За таа цел, обезбеден е бројот на возила на 10 раскрсници во времетраење од четири седмици од понеделник до недела за четирите временски сезони (пролет, лето, есен и зима) за 2018 година од Центарот за управување и контрола на сообраќајот (ЦУКС). Согласно локацијата на точките за каде што се превземени податоците за протокот на возила (приказ на слика 2.4), најпрвин неопходна е обработка на податоците за протокот на возилата. Овие резултати беа статистички обработени и анализирани со цел пресметување на проток на возила кои поминуваат во еден час. Протокот

на возила како линиски извор на бучава е основен влезен податок кој служи за креирање акустичен предиктивен модел.

За реализација на оцена на бучавата во урбаната средина се искористи софтверскиот пакет IMMI Premium, кој се користи за професионални и истражувачки цели и служи за статичко мапирање и предвидување на распространетоста на бучавата во животната средина. На сликата 2.4. се прикажани применетите оперативни фази за развој на акустичниот предиктивен модел. Добиени се геореференцирани картографски податоци, а дополнително внесена е топографијата и геометријата на подрачјето. Висините на објектите се многу важен параметар за дисперзија на бучавата и истите беа измерени и мануелно внесени во софтверот. За определување на просечната брзина на возилата на улиците кои се разгледуваат, се внесува максималната дозволена брзина (50 km/h на булеварите и 30 до 40 km/h на споредните патишта и улици), која понатаму се користи за определување на просечната брзина на возилата.



Слика 2.4. Оперативни фази при креирање акустичен предиктивен модел

По спроведените симулации, добиена е основна анализа за бучавата во анализираното подрачје, каде што како основен индикатор може да се одреди еквивалентното ниво на бучава во секоја точка во едно деноноќие и индивидуално во тек на ден, вечер и ноќ, како и приказ на максималното ниво на бучава.

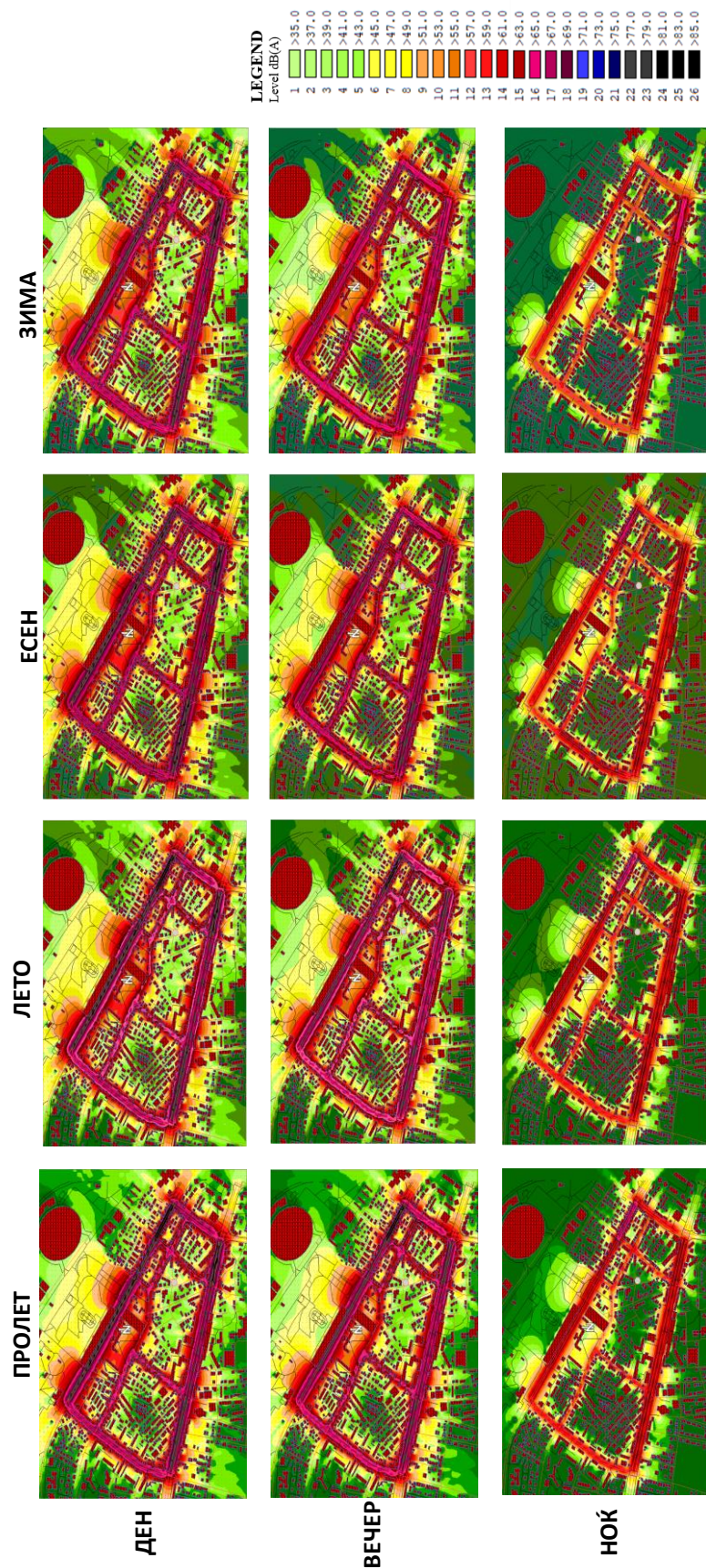
2.3.2. РЕЗУЛТАТИ ОД ИЗРАБОТЕНИОТ АКУСТИЧЕН ПРЕДИКТИВЕН МОДЕЛ

Согласно спроведената методологија, изработени се 12 акустични предиктивни модели на бучава, каде што единствениот инкрпориран извор е протокот на возила кој е причина за појава на сообраќајна бучава. Акустичните предиктивни модели кои се прикажани на слика 2.5., визуелно даваат информација нивото на бучава во 4 временски сезони (пролет, лето, есен и зима) во текот на дневниот, вечерниот и ноќниот временски период. На слика 2.5. се прикажани креираните карти на бучава, каде што може да се забележи дека нивоата на бучава на булеварите и улиците ги надминуваат граничните вредности.

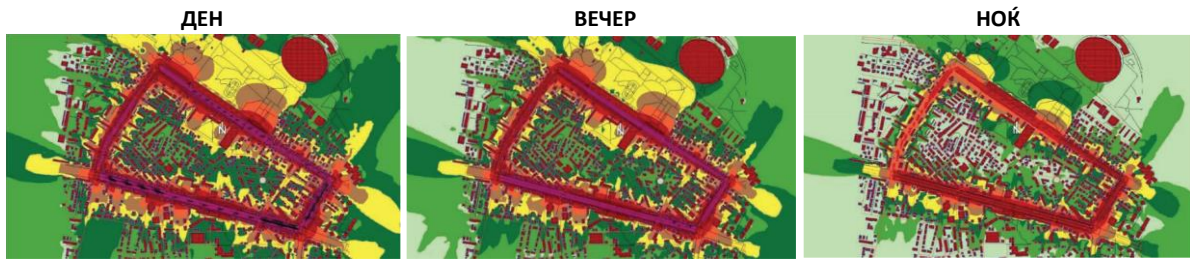
Преку споредба на резултатите од акустичниот предиктивен модел со граничните вредности, може да се заклучи дека нивоата на бучава се надминати на локациите блиску до главните улици и булевари како резултат на густоот сообраќаен метеж и постојаниот проток на возила.

При анализа на нивото на бучава во текот на дневниот, вечерниот и ноќниот период може да се забележи дека највисоко ниво на бучава се појавува во дневниот период, помеѓу 07:00 и 19:00 часот. Доколку резултатите се споредат согласно сезоната, може да се заклучи дека максималното ниво на бучава се јавува во текот на пролетниот период, потоа летниот, есенскиот и на крај, зимниот период. Анализата на овие резултати укажуваат на тоа дека во избраното подрачје постои сериозно загадување од бучавата.

Дополнително, изработен е акустичен предиктивен модел на бучава во анализираниот период за кој што се направени мерењата на нивото на бучава кои се прикажани во глава 2.2.5. Согласно внесениот проток на возила како линиски извор на бучава, добиени се три карти на бучава за дневниот, вечерниот и ноќниот период на бучава кои се прикажани на слика 2.6.



Слика 2.5. Акустични предиктивни карти за анализираното подрачје



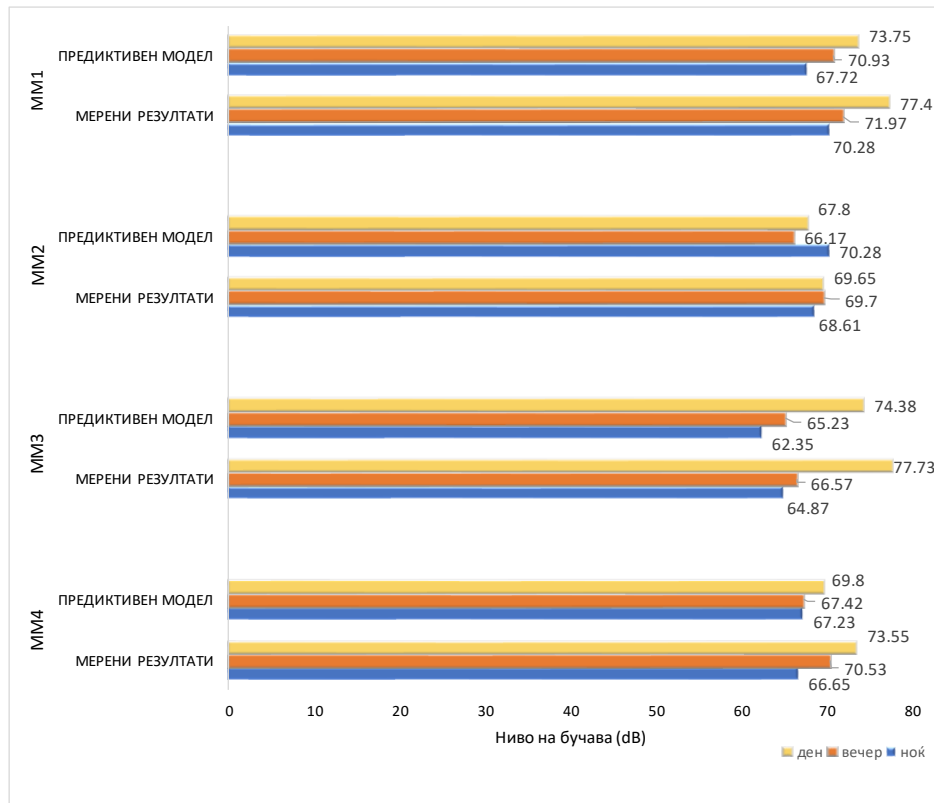
Слика 2.6. Акустичен предиктивен модел за временскиот период кога се направени мерења на ниво на бучава

Од акустичниот предиктивен модел прикажан на слика 2.6, може да се добијат информации за основните индикатори на бучава на мерните места на кои беа направени мерењата. Со следното поглавје ќе се направи споредба на резултатите од измереното ниво на бучава и нивото на бучава добиено од акустичниот предиктивен модел.

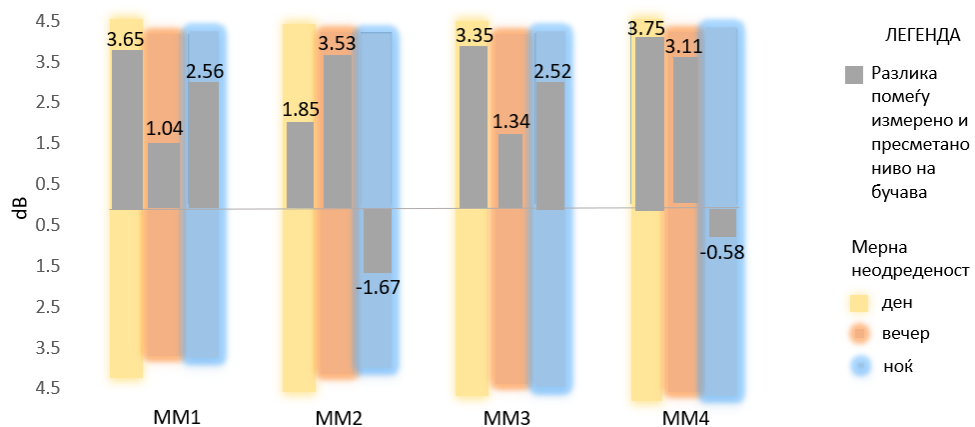
2.3.3. КОРЕЛАЦИЈА ПОМЕЃУ СПРОВЕДЕНИТЕ МЕТОДОЛОГИИ

Согласно двете методологии за објективна проценка на нивото на бучава во избраното урбано подрачје, може да се заклучи дека основните индикатори за нивото на бучава во текот на дневниот, вечерниот и ноќниот период покажуваат зголемено ниво на бучава кое ги надминува дозволените гранични вредности. Резултатите добиени од мерењата и акустичниот предиктивен модел за мерните места ММ1, ММ2, ММ3 и ММ4 се прикажани на слика 2.7.

Дополнително, преку споредба на резултатите од спроведените методологии, вклучувајќи ја мерната неодреденост, може да се потврди точноста на методологиите. Од сликата може да се забележат различни вредности на индикаторите на бучава во текот на дневниот, вечерниот и ноќниот период. Доколку се земе во предвид пресметаната мерна неодреденост од мерењата и се направи разлика помеѓу измерените и предвидените вредности, може да се види дали вредностите од акустичниот предиктивен модел се во границите на мерната неодреденост (слика 2.8).



Слика 2.7. Резултати од нивото на бучава од спроведените мерења и акустичниот предиктивен модел за мерните места MM1, MM2, MM3 и MM4



Слика 2.8. Разлика помеѓу измереното и предвиденото ниво на бучава

Од слика 2.8. може да се заклучи дека разликата од овие вредности е во границите на мерна неодреденост, од каде што може да се забележи дека нивото на бучава е повисоко на измерените вредности, кое може да биде повисоко поради тоа што при мерењата се појавуваат дополнителни извори на бучава (свирки од кола, зборување, музика). Бидејќи разликата помеѓу добиените вредности на индикаторите за нивото на бучава од применетите методологии не ја надминува границата на мерната неодреденост, може да се потврди точноста на добиените резултати.

2.4. СУБЈЕКТИВНА АНАЛИЗА ЗА ВЛИЈАНИЕТО НА БУЧАВАТА

Субјективното влијание што го има бучавата врз населението е важен показател за штетното влијание на бучавата врз населението. Основната цел од оваа анализа е да се процени влијанието на бучавата и да се детектира изворот на бучава кој намногу пречи, односно предизвикува вознемиреност кај населението. Еден од начините за субјективна анализа на бучавата е преку спроведување анкета со која би се анализирале доминантните изворите на бучава. Овие иницијални истражувања се основа за понатамошни детални анализи и истите ќе послужат за дефинирање на доминантните извори на бучавата во анализираното подрачје.

За тие цели, спроведена е јавна анкета во избраното подрачје за анализа, населбата „Дебар Маало“. За спроведување на анкетата за субјективното влијание на бучавата, потребно е да се дефинираат следните параметри: број на испитаници, форма и број на прашања што ќе ги содржи анкетата, како и начин на нејзино спроведување.

2.4.1. ИЗБОР НА БРОЈ НА ИСПИТАНИЦИ

Согласно истражувањето во [86, 87], бројот на испитаници зависи од повеќе фактори: вкупен број на население, степен на сигурност, маргина на грешка и пропорција на населението. За да се добие бројот на испитаници (n'), потребно е да се земе во предвид вкупниот број на население (N), факторот заснован на ниво на доверба (z), маргина на грешка (ε) и пропорција на населението (\hat{p}), прикажан со следниот израз:

$$n' = \frac{\frac{z^2 \cdot \hat{p}(1-\hat{p})}{\varepsilon^2}}{1 + \frac{z^2 \cdot \hat{p}(1-\hat{p})}{\varepsilon^2 \cdot N}} \quad (2-10)$$

Вкупен број на население е бројот на население во подрачјето кое е предмет на истражување. Според направените анализи, во населбата Дебар Маало живеат околу 10.000 жители. Степенот на сигурност е процент кој открива колкава е сигурноста дека спроведената анкета ќе соодветствува за целото подрачје. При спроведување на анкети за анализа во научни истражувања, нивото на доверба треба да биде најмалку 95%. Доколку нивото на доверба е 95%, факторот заснован на нивото на доверба има вредност $z = 1,96$.

Дозволена маргина на грешка е процент кој покажува колку резултатите можат да отстапат од ставовите на вкупното население. Колку е помала маргината на грешка, толку е поголема точноста на резултатот. Дозволен процент на маргината на грешка е 10%, односно $\varepsilon = 0,01$.

Пропорција на населението е процентуален износ кој покажува колку населението е засегнато од самиот проблем. Бидејќи не постои официјален податок за оваа вредност, се усвојува 50%, односно $\hat{p} = 0.5$. Согласно овие вредности, пресметано е дека за спроведување на анкета за влијанието на бучавата во скопската населба „Дебар Маало“ како дел од истражувањето во оваа докторска дисертација, се потребни 96 испитаници.

2.4.2. КРЕИРАЊЕ АНКЕТНА ФОРМА

Анкетата е спроведена во стандардизирана форма, каде што прашањата и предложените одговори се дадени однапред. Прашалникот е дизајниран со цел добивање информации за влијанието на бучавата, нејзин најсилен интензитет во период од денот и годината, извори на бучава кои најмногу влијаат врз субјективното чувство на населението и како влијанието на бучавата врз здравјето согласно истражувањата во [38, 88].

Анкетата е спроведена во усна форма, преку слободен разговор и интервју во периодот од 03.07.2019 до 07.07.2019 година. Дизајност на анкетата (обликот и прашањата) се прикажани на слика 2.9.

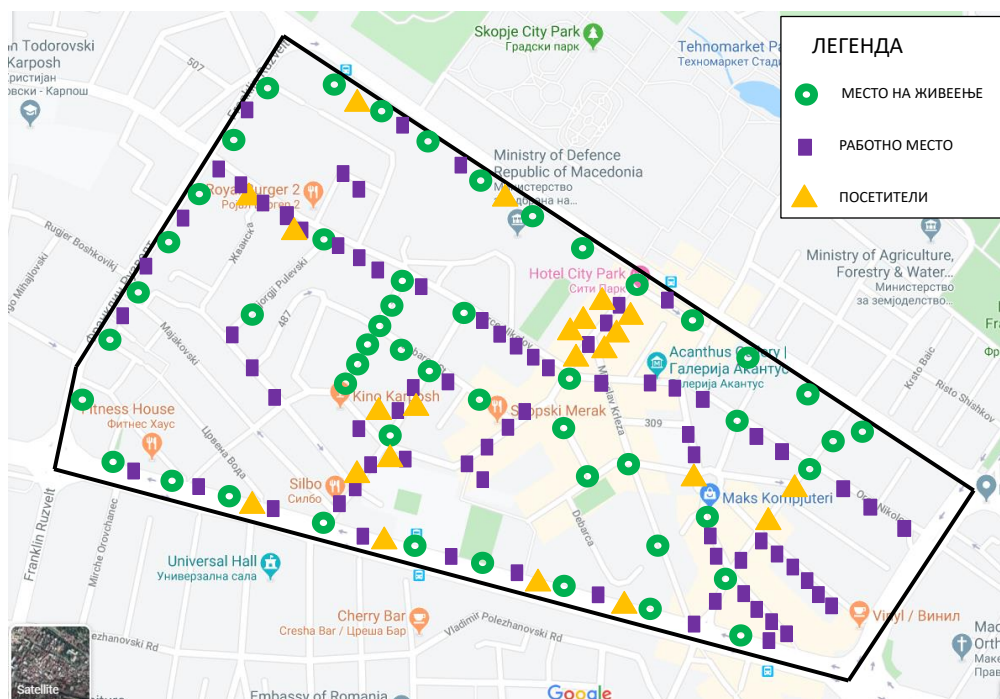
АНКЕТА ЗА ВЛИЈАНИЕТО НА БУЧАВАТА ВО НАСЕЛБАТА „ДЕБАР МААЛО“

- | | |
|--|--|
| <p>1. Во која возрастна група припаѓате?</p> <p><input type="checkbox"/> 15-29 години</p> <p><input type="checkbox"/> 30-44 години</p> <p><input type="checkbox"/> 44-59 години</p> <p><input type="checkbox"/> Над 60 години</p> <p>2. Пол:</p> <p><input type="checkbox"/> машки</p> <p><input type="checkbox"/> женски</p> <p>3. „Дебар Маало“ е Вашето:</p> <p><input type="checkbox"/> живеалиште
* Улица (незадолжително) _____)</p> <p><input type="checkbox"/> работно место
*(Улица (незадолжително): _____)</p> <p><input type="checkbox"/> посетител</p> <p>4. Вид на градба во која престојувате/ работите?</p> <p><input type="checkbox"/> Станбена градба (спрат: _____)</p> <p><input type="checkbox"/> Индивидуален објект (спрат: _____)</p> <p><input type="checkbox"/> Деловен објект (спрат: _____)</p> <p>5. Колку ве засега нивото на бучава во животната средина?</p> <p><input type="checkbox"/> Малку</p> <p><input type="checkbox"/> Средно</p> <p><input type="checkbox"/> Многу</p> <p><input type="checkbox"/> Воопшто не ме засега</p> <p>6. Во кој период од годината бучавата има најсилен интензитет?</p> <p><input type="checkbox"/> Пролет</p> <p><input type="checkbox"/> Лето</p> <p><input type="checkbox"/> Есен</p> <p><input type="checkbox"/> Зима</p> | <p>7. Во кој период од денот бучавата има најсилен интензитет?</p> <p><input type="checkbox"/> Наутро</p> <p><input type="checkbox"/> Претпладне</p> <p><input type="checkbox"/> Попладне</p> <p><input type="checkbox"/> Навечер</p> <p><input type="checkbox"/> Постојано</p> <p>8. Кои извори на бучава најмногу ви пречат?</p> <p><input type="checkbox"/> Сообраќај</p> <p><input type="checkbox"/> Застој во сообраќај, сообраќаен шпиц</p> <p><input type="checkbox"/> Бучава од градежни работи</p> <p><input type="checkbox"/> Бучава од сирени</p> <p><input type="checkbox"/> Бучава од гласна музика и зборување од кафулиња и барови</p> <p><input type="checkbox"/> Бучава од соседството</p> <p><input type="checkbox"/> Играње на улица</p> <p>9. Како влијае бучавата врз вашето здравје?</p> <p><input type="checkbox"/> Вознемирува</p> <p><input type="checkbox"/> Тахикардија</p> <p><input type="checkbox"/> Нарушен сон</p> <p><input type="checkbox"/> Други видови на болести</p> <p><input type="checkbox"/> Не влијае на моето здравје</p> <p>10 Според Вас, кои се причините за зголемено ниво на бучава?</p> <p><input type="checkbox"/> Лоша изолација на угостителските објекти</p> <p><input type="checkbox"/> Непочитување на законски обврски</p> <p><input type="checkbox"/> Ниска свест кај угостителите</p> <p><input type="checkbox"/> Ниска свест кај гостите и посетителите</p> <p><input type="checkbox"/> Непланско искористување на просторот и намена на дејностите</p> <p><input type="checkbox"/> Недоволна инспекциска контрола</p> <p><input type="checkbox"/> Неорганизирана градска инфраструктура</p> |
|--|--|

Слика 2.9. Анкета за влијанието на бучавата во скопската населба Дебар Маало

2.4.3. ЕВАЛУАЦИЈА НА АНКЕТАТА И АНАЛИЗА НА РЕЗУЛТАТИТЕ

Анкетата е спроведена на население на возраст од 15 години па нагоре. Од вкупно 96 испитаници, 57% се испитаници од женски пол, а 43% се машки пол. За 55% од населението, Дебар Маало претставува нивно работно место, за 31% живеалиште, а 14% се посетители. На слика 2.12 е прикажана местоположбата на испитаниците (слика 2.10).



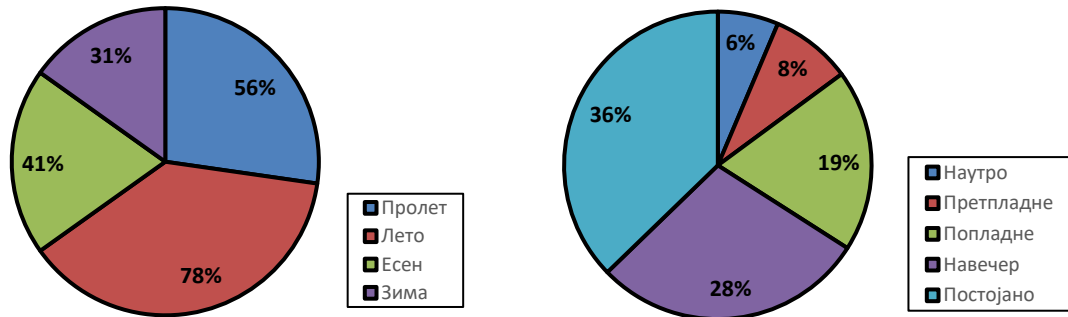
Слика 2.10. Местоположба на испитаниците

Во однос на прашањето колку се засегнати од нивото на бучава во животната средина, најголем број од населението е многу засегнато (59%), односно бучавата за нив претставува сериозен проблем (слика 2.11).



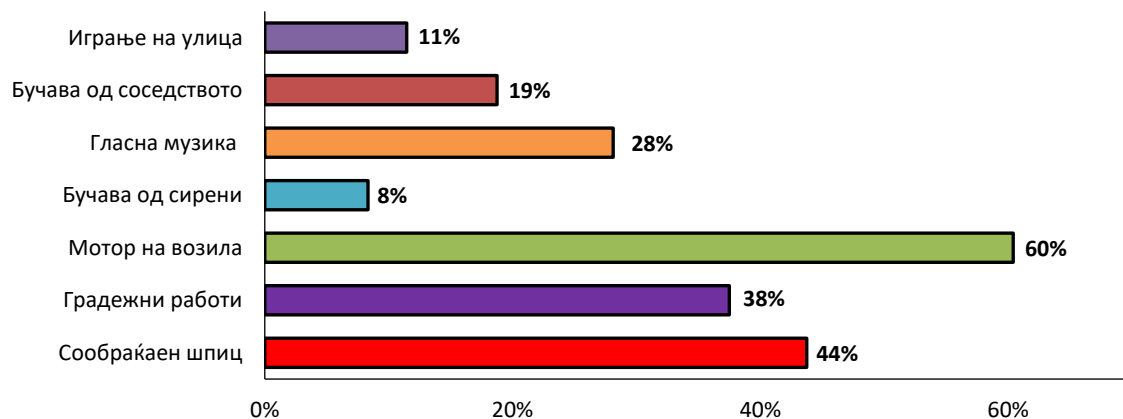
Слика 2.11. Степен на влијание на бучавата во животната средина кај испитаниците
 Согласно анализите од резултатите од анкетата, бучавата има најсилен интензитет во летниот (76%) и пролетниот период (56%). На ова прашање, дел од граѓаните дополнително

објаснија дека во овој период од годината бучавата има најголем интензитет поради поволните метеоролошки услови резултирајќи во зголемена фреквентност на движење на населението. Дополнително, на прашањето „во кој период од денот бучавата има најсилен интензитет“, 36% од населението одговорија постојано, а 28% навечер (слика 2.12).



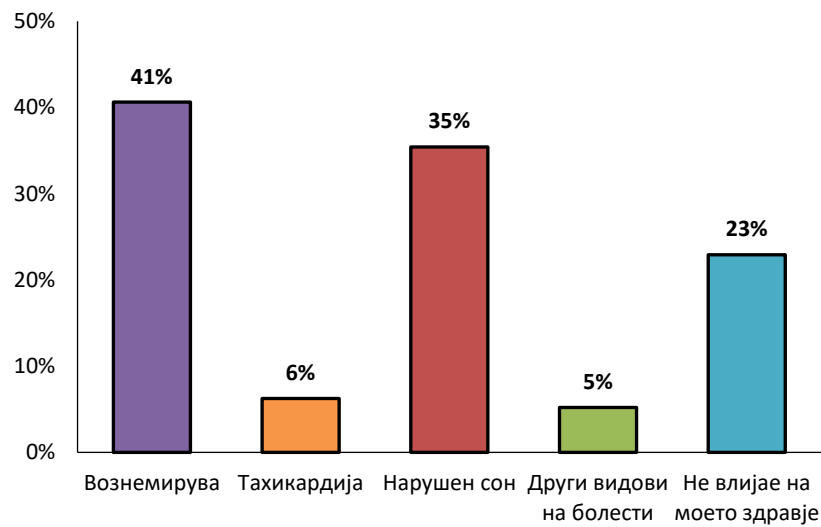
Слика 2.12. Најсилен интензитет на бучавата во период според годишните сезони и деноноќие

Анализата покажува дека бучавата предизвикана од возила и мотори кои возат со голема брзина е доминантен извор на бучава и најчеста причина која предизвикува разни форми на болести како вознемиреност и нарушен сон кај населението. Највознемирувачки звуци се звуците предизвикани од сообраќајот, бучавата што ја предизвикува моторот на возилата во мирување поради сообраќаен шпиц и брзото возење на возилата и моторите. На слика 2.13. се прикажани резултатите од изворите на бучава кои најмногу го вознемируваат населението. Може да се забележи дека сообраќајната бучавата е главна причина за појава на вознемиреност и штетно влијание врз населението.



Слика 2.13. Доминантни извори на бучава

Евидентирано е дека кај 41% од испитаниците бучавата предизвикува вознемиреност, додека 35% од испитаниците страдаат од нарушен сон од штетното влијаниена бучавата (слика 2.14).



Слика 2.14. Влијание на бучавата врз здравјето на населението

На слика 2.15. е прикажано мислењето на испитаниците за можните причини за појава на бучава. Основни причини за појава на бучава според испитаниците е непочитување на законските обврски (54%), неорганизирана градска инфраструктура (42%), недоволна инспекциска контрола (39%) и ниска свест кај населението (39%).



Слика 2.15. Причини за појава на бучава

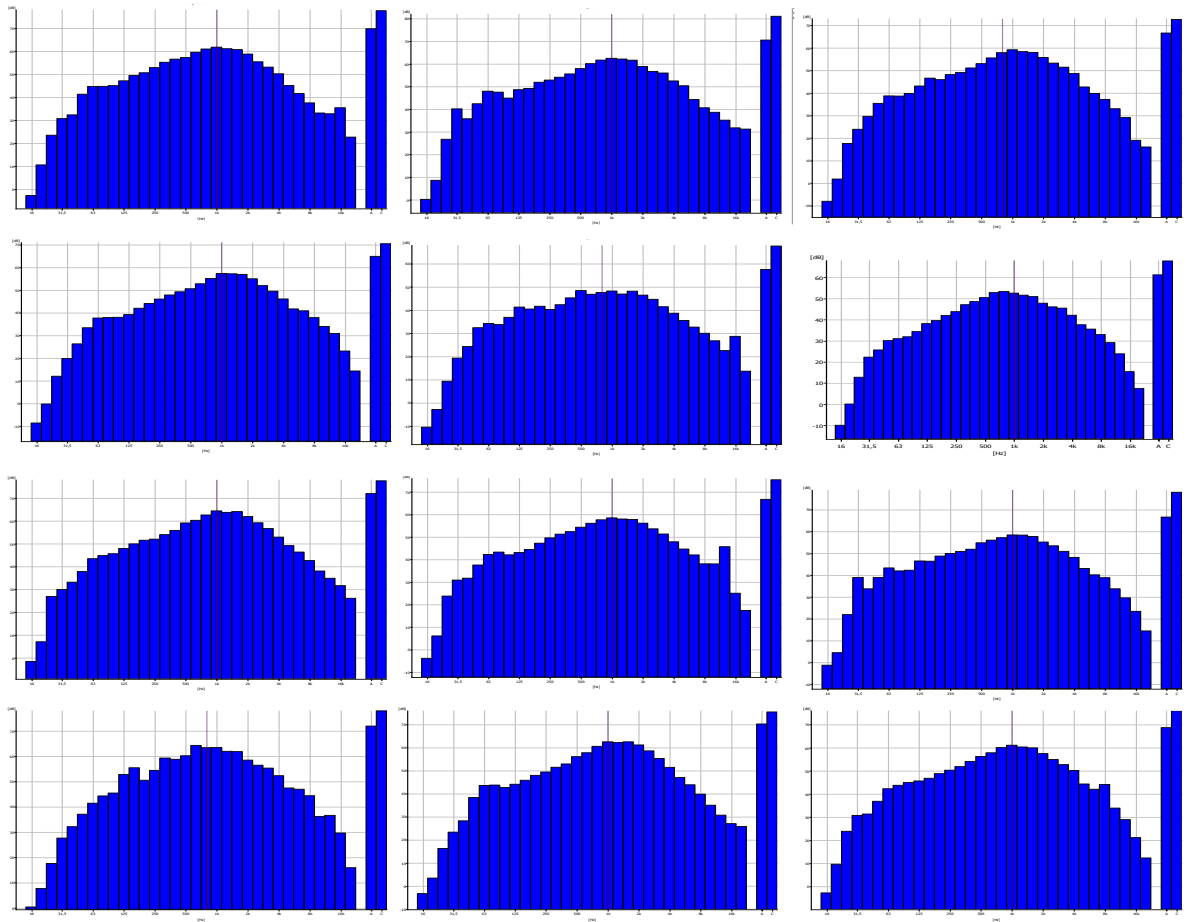
Согласно извршената анкета може да се заклучи дека сообраќајната бучава е примарен извор кој има најмногу пречи на населението, но може да се забележи дека постојат и други доминантни извори на бучава во анализираната урбана средина.

2.5. СОГЛЕДУВАЊА ОД ПРИМЕНЕТИТЕ МЕТОДИ ЗА ПРОЦЕНКА НА НИВОТО НА БУЧАВА

Резултатите од применетите методологии за проценка на нивото на бучава во животната средина, спроведување на стандардни мерења на нивото на бучавата и креирањето на акустични предиктивни модели потврдуваат дека изложеноста на бучава ги надминува граничните вредности. При анализа на резултатите од акустичните предиктивни карти засновани од протокот на сообраќај како влезна големина, може да се заклучи дека сообраќајот сам по себе предизвикува високо ниво на бучава. Од резултатите од мерењето на нивото на бучава, дополнително беа анализирани фреквентните спектри за секое мерење индивидуално. Преку анализа на спектарот во 1/3 октавен спектар може да се заклучи дека 14 од 18 мерења покажуваат максимална вредност на енергијата на фреквенција од 1000 Hz. Вообичаено, пикот на фреквенција околу 1000 Hz се јавува кога доминантен извор на бучава е сообраќајот. Бучавата што се јавува на оваа фреквенција е поврзана со звукот од гумите, пониските фреквенции се поврзани со бучавата од издувните гасови, а повисоките фреквенции се создадени од поминувањето на ветерот [89, 90]. Понатаму, научниците во [91] потврдуваат дека максималните вредности при 1/3 октавен спектар кај бучавата предизвикана од урбаниот сообраќај се појавуваат на фреквенција од 1000 Hz. На слика 2.16. се прикажани фреквентните спектри во 1/3 октавен спектар од мерењата на четири мерни места во текот на дневниот, вечерниот и ноќниот период, потврдувајќи дека максималната вредност е на 1000 Hz, од каде што може да се заклучи дека доминантниот извор на бучава е предизвикан од сообраќајот.

Традиционалните методи за проценка на нивото на урбана бучава беа прикажани како успешни методологии за квантификација и валидација на бучавата. Може да се заклучи дека методот за проценка на нивото на бучава преку мерење и преку креирање акустичен предиктивен модел можат да помогнат при проценка на загадувањето предизвикано од бучавата во урбаните области. Може да се извлечат две констатации, дека бучавата во анализираната средина значително е повисока од граничните вредности и дека главен извор на бучава претставува бучавата предизвикана од сообраќајот. При користење на овие методи, се одредува нивото на бучава, но не и идентификација на различните извори на бучава.

При споредба на овие резултати со субјективната анализа може да се потврди дека бучавата претставува сериозен проблем во анализираната урбана средина и дека таа е константно присутна. При анализа на главните извори на бучава кои се вознемирувачки за населението, може да се забележи дека доминантен е изворот од сообраќајната бучава. Постојат повеќе извори кои предизвикуваат сообраќајна бучава сообраќајниот застој и постојаниот проток на возила. Од друга страна, може да се забележи дека постојат и други извори на бучава кои предизвикуваат вознемирување на населението, како градежни работи, гласна музика, гласно зборување на луѓето и бучавата предизвикана од сирени. Сите овие извори учествуваат во креирање на комплексен спектар од звуци кој резултира во креирање на бучава.



Слика 2.16. Фреквентен спектар во 1/3 октава

Резултатите од субјективната анализа покажаа дека бучавата не е предизвикана само од протокот на сообраќај, а има и други вознемирувачки извори на бучава што предизвикуваат вознемиреност. Поради овие причини, следните чекори во истражувањето ќе бидат насочени кон идентификување и класификација на акустични настани. Развој на систем за препознавање и класификација на звучни настани е неодамнешна дисциплина во современите истражувања на бучавата. Се состои од обработка на акустични сигнали и нивно претворање во симболични описи што одговараат на перцепцијата на слушателот за различните звучни настани присутни во сигналите и нивните извори. Користејќи ја вештачката интелигенција и машинското учење, создадени се алгоритми за идентификација и препознавање настани што се случуваат во нивната околина, така што тие ќе можат да дадат соодветен одговор на потребите на нивните корисници. Со проценување на учеството на одредени извори на бучава, лесно може да се упрости спектарот од звуци за да се најде моменталниот доминантен извор на бучава со цел спречување на неговото влијание. Субјективната анализа ќе послужи за креирање на таксономија која ги дефинира класите на звучни настани кои ќе послужат како предмет за препознавање во системот за препознавање и класификација на звучните настани.

3. СИСТЕМАТИЗАЦИЈА И ПАРАМЕТРИЗАЦИЈА НА ЗВУЧНИ НАСТАНИ

Препознавањето и класификацијата на звучните настани е дисциплина од областа на компјутерска анализа на звукот која вклучува креирање под системи за детектирање и параметризација преку дигитална обработка на звучните сигнали со цел понатамошна класификација преку употреба на алгоритми на машинско учење.

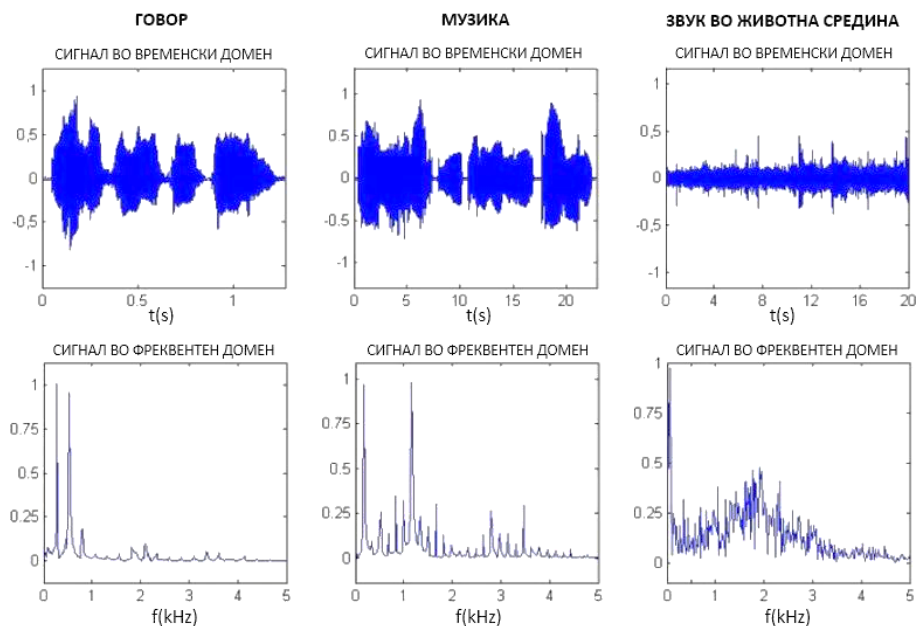
Систематизацијата на звучните настани е прв чекор при изградба на системите за препознавање и класификација на звучни настани што овозможува поделба на звуците во класи кои помагаат при креирање на базата на податоци која се користи за тренирање и тестирање на алгоритмите на машинското учење. Од интерес е да се направи систематизација на звучните настани во урбаните средини кои се причина за појава на загадување предизвикано од бучава. Согласно спроведено истражување и систематизација на класите на звучни настани, креирана е таксономија на звучни настани во анализирана урбана средина. Врз основа на усвоената таксономија, избрани се 10 репрезентативни класи на звучни настани кои помогнаа при избор на база на звучни настани составена од податоци кои ги претставуваат 10те избрани класи на звучни настани.

Следната фаза вклучува примена на параметризација на звучните настани. Параметризација овозможува екстракција на важни карактеристики на звучните настани кои доведуваат до намалување на времето за обработка и постигнување висока точност при препознавање на класата на звучниот настан. Направена е анализа на 50 параметри на звучни настани кои се применуваат за екстракција на важни карактеристики од аудио сигналите и формирање на вектор на карактеристики. Согласно постапката која се користи за обработка, одбрани се пет параметри кои се применуваат за екстракција на важните карактеристики на класите на звучни настани од базата на звучни настани. За секој параметар е развиена математичка интерпретација, а како резултат се прикажани визуелни репрезентации за секоја класа на звучен настан.

3.1. СИСТЕМАТИЗАЦИЈА НА ЗВУЧНИ НАСТАНИ

Поделбата и систематизирањето на звучниот настан независно од природата на звукот е комплексен проблем, чија цел е да се подели звукот на поедноставни и помали класи на звучни настани за полесно препознавање при употреба на алгоритмот за машинско учење. Од тие причини, потребно е да се направи звучна таксономија за систематизирање на звукот која помага во подобро разбирање на податоците кои се обработуваат. Преку поделба на звучните настани во соодветни класи се подобрува точноста при класификација и се овозможува побрзо процесирање на податоците. Концептот на звучната систематизација е субјективен и зависи од избраниот домен за поделба [92].

За идентификација на звукот во урбана средина, најпрвин треба да се воочат разликите помеѓу звучните настани во животната средина, говорот и музиката. Важно е да се напомене дека говорот и музиката имаат ограничено количество на звучни настани, додека звукот во животната средина, теоретски има бесконечен број на звучни настани, чие меѓусебно делување создава бучава во животната средина. Втората главна разлика е периодичноста во временскиот домен. Речиси кај секој звук на животната средина не постои периодичност, додека кај музиката и говорот може да се забележи периодичност. За да се потврдат овие разлики, како дел од ова истражување е обработен снимен звучен настан од сите три категории во временски и фреквентен домен (слика 3.1). Со анализа во временскиот домен, може да се забележи дека говорот и музиката имаат хармоничен и периодичен сигнал. Од фреквентниот спектар може да се заклучи дека звучниот настан во животната средина е комплексен и има повеќе звучни настани и густе и чести пикови за разлика од останатите два типа на сигнали.



Слика 3.1. Споредба помеѓу говор, музика и звучен настан во животната средина во временски и фреквентен домен

Согласно анализата на овие три категории на звуци, на табела 3.1. се прикажани нивните главни карактеристики. За разлика од говорот и музиката, кај звукот од животната средина не се познати важни информации како времетраењето на звучните настани или пропорцијата помеѓу хармоничната и нехармоничната спектрална структура, зголемувајќи ја комплексноста на препознавањето на звуците во животната средина.

Табела 3.1. Листа на карактеристики за говор, музика и звуци во животната средина

КАРАКТЕРИСТИКИ	ГОВОР	МУЗИКА	ЗВУЦИ ВО ЖИВОТНА СРЕДИНА
АНАЛИЗА НА ЗВУЧНАТА ЕДИНИЦА	Феномен	Нота	Настан
ИЗВОР	Човечки вокален тракт	Инструменти	Секој звук што се појавува и создава настан
БРОЈ НА МОЖНИ ИЗВОРИ	Конечен	Конечен	Бесконечен
ВРЕМЕНСКА СТРУКТУРА	Кратка (40-200ms). Статична. Временски константна, можна е варијација.	Кратка (600-1200ms). Стабилна состојба. Периодичност.	Долга (500-3000 ms). Непозната пропорција на статични и динамични карактеристики.
СПЕКТРАЛНА СТРУКТУРА	Најголем број хармониски. Дел нехармониски	Најголем број хармониски. Дел нехармониски.	Непозната пропорција на хармонична и нехармонична структура.
СЕМАНТИЧКА СТРУКТУРА	Симболична, продуктивна	Симболична, продуктивна	Непродуктивна. Можност за постоење на низи со значење.

Областа од интерес во ова истражување е звукот во животната средина кој е предизвикан во урбаните средини и е причина за појава на загадување предизвикано од бучава. Урбаните средини се справуваат со проблем предизвикан од загадувањето на бучавата, кој има се поголемо влијание во последните години. Систематизацијата на звучните настани преку креирање на таксономија е прв чекор за изградба на системи за препознавање и класификација на звучни настани.

Правилната таксономија има клучна улога за правилна систематизација на звучните настани која доведува до ефикасна класификација при примена на алгоритмите засновани на вештачка интелигенција. За потребите на ова истражување, ќе се креира таксономија на звучни настани кои предизвикуваат вознемиреност во урбана средина преку употреба на следниот пристап:

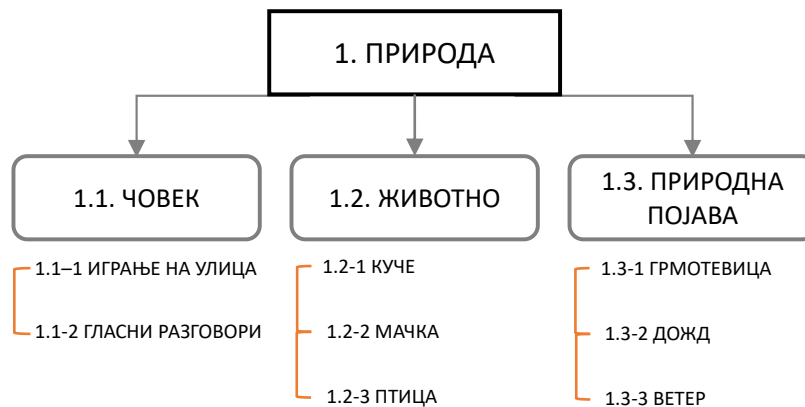
- Систематизацијата ќе се направи согласно спроведената субјективната анализа за влијанието на звучни настани кои предизвикуваат вознемиреност кај населението;
- При креирање на таксономијата ќе се дефинираат конкретни извори на бучава кои предизвикуваат бучава (пример: бушење наместо градежни работи);
- Систематизацијата ќе се направи согласно претходни истражувања и таксономии;

Согласно истражувањето и анализите, на највисокото ниво се дефинирани три групи на звучни настани: природа, транспорт и објект согласно изворите на звучни настани во анализираната урбана средина. Овие три групи претставуваат семантички класи на највисокото ниво на поделба (слика 3.2).



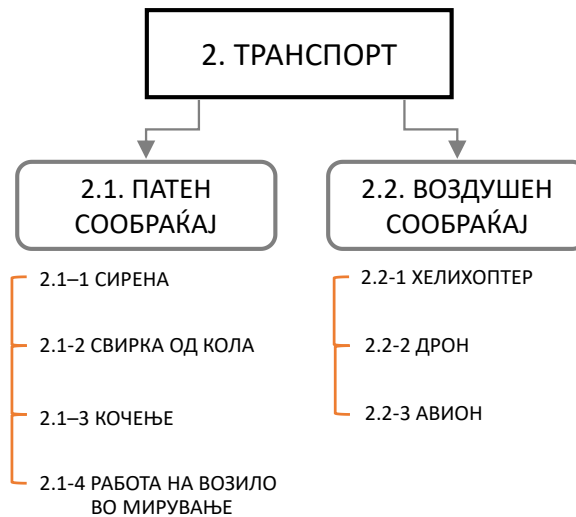
Слика 3.2. Семантичка класа на извори на звучни настани во животна урбана средина на највисоко ниво на таксономијата

Следен чекор е поделба на одделните класи на извори на второ ниво, каде што доаѓа до подетално дефинирање на изворот. Природата како звучен настан се дели на три класи: настани предизвикани од човекот, животните и природните појави (слика 3.3.). Под секој од овие извори, детално се дадени изворите на звучни настани кои предизвикуваат вознемиреност во анализираната урбана средина. Важно е да се напомене дека третата, односно последната поделба е согласно направената субјективна анализа за штетното влијание на бучавата кај населението прикажана во глава 2. Доколку вознемирувачките звучни настани се дефинирани согласно изворите кои се споменати во субјективната анализа, поради поопштата поделба тешко би дошло до нивна класификација. Од тие причини, изворите беа конкретизирани, односно беа поделени на најниско семантичко ниво, за полесно да може се препознае класата на звучниот настан.



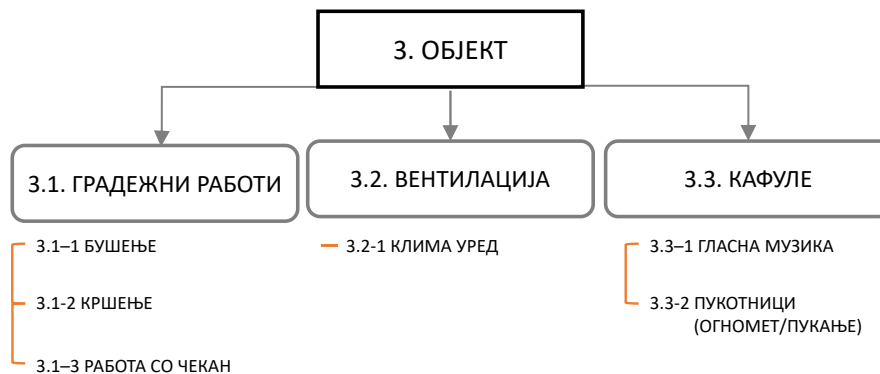
Слика 3.3. Втора и трета поделба на изворите на бучава кои се јавуваат во природата како звучен настан

Во групата на звучни настани предизвикани од транспорт се внесени изворите на патен сообраќај и воздушен сообраќај (слика 3.4). Железничкиот сообраќај не е вклучен во оваа систематизација, поради тоа што во анализираното подрачје не постојат звучни настани од овој извор. Како вознемирувачки извори во патниот сообраќај се издвоени: сирена, свирка од кола, кочење и работа на возило во мирување. Воздушниот сообраќај е редок извор на бучава во анализираното подрачје, а како вознемирувачки извори може да се вклучат струењата во воздухот што се јавуваат поради поминување на хеликоптер, дрон и авион.



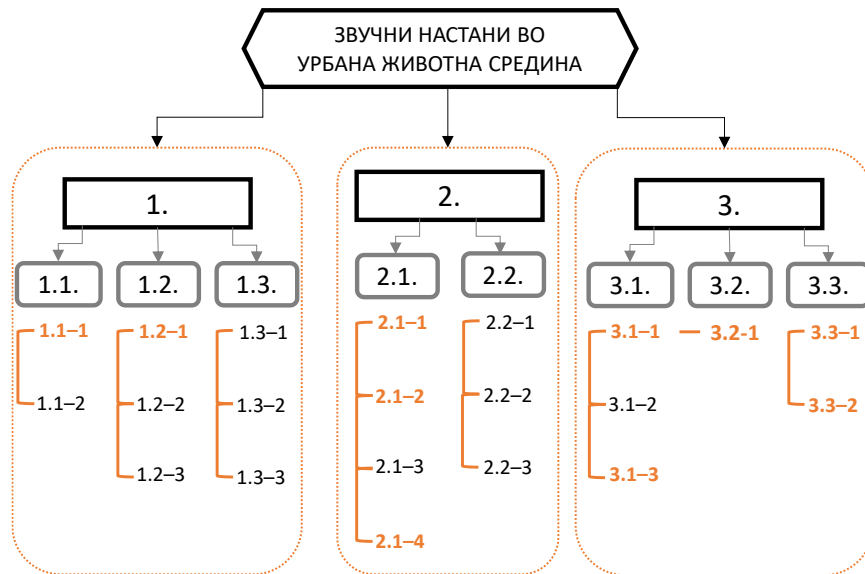
Слика 3.4. Втора и трета поделба на изворите на бучава кои се јавуваат во транспортот како звучен настан

На сликата 3.5 е прикажана поделбата на објектот како извор на звучен настан, каде што како извори се јавуваат градежни работи, вентилација и кафулиња. Звучни настани кои се издвоени во класата на градежни работи се: бушење, кршење и работа со чекан, додека работата на клима уредите е извор на бучава како дел од вентилацијата. Кафулињата се објекти кои се често застапени во анализираната урбана средина. Гласната музика предизвикана од забави и прослави, како и пукотниците како импулсни звуци се чести извори на бучава кои предизвикуваат вознемиреност кај населението.



Слика 3.5. Втора и трета поделба на изворите на бучава кои се јавуваат во објект како звучен настан

На сликата 3.6. прегледно е прикажана целосната таксономија, каде наместо имињата на звучните извори, се користат ознаките со кои се означени овие извори од сликата 3.2 до сликата 3.5. Изворите на бучава кои се означени се одбраните 10 карактеристични класи на звучни настани кои предизвикуваат вознемиреност кај населението. Одбраните класи на звучни настани кои предизвикуваат вознемиреност кај населението се: играње на улица, лаење на куче, сирена, свирка од кола, работа на возило во мирување, бушење, работа со чекан, клима уред, гласна музика и пукотници.



Слика 3.6. Таксономија на звучни настани во урбана звучна средина кои предизвикуваат вознемиреност кај населението

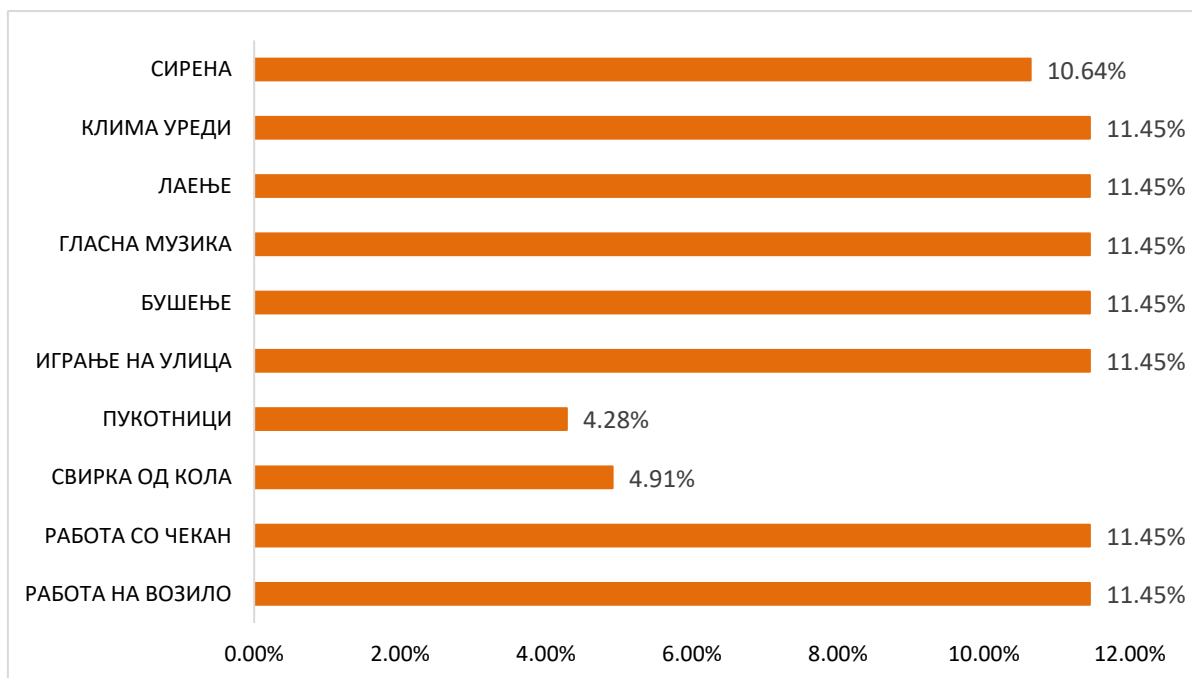
Согласно субјективната анализа прикажана во глава 2, играњето на децата и сирената се звучни настани кои предизвикуваат вознемиреност само кај 10% од населението, но се земени во оваа таксономија поради тоа што се појавуваат често и доколку не се одделат како посебни класи, може да доведат до мешање на класите и постигнување помала точност при препознавање на класите. Лаењето на куче и клима уредите се класи кои се издвоени од настанот бучава од соседството која предизвикува вознемиреност кај 19% од населението. Гласната музика предизвикува вознемиреност од 28%. Дополнително, согласно современите истражувања, додадени се и пукотниците како импулсни звуци, бидејќи иако имаат кратко времетраење, нивната појава влијае негативно врз населението. Следен извор на бучава кој има негативно влијание кај 38% од населението се градежните работи. Бидејќи постојат повеќе извори кои имаат различна природа, одвоени се бушењето и работата со чекан како извори кои често се појавуваат во класата на градежни работи. За крај, бучавата од патниот сообраќај е дефинирана како највознемирувачки извор на бучава. Покрај тоа што звучните настани предизвикани од патниот сообраќај се постојани, тие предизвикуваат и најголемо вознемирување, конкретно кај 44% од населението влијае сообраќајниот застој што доведува до дефинирање на класата свирка од кола, а дури кај 60% од населението самата работа на возилата во мирување предизвикува голема вознемиреност.

Правилната таксономија има клучна улога при правилно одделување на звучните настани и ефикасна класификација при примена на алгоритмите засновани на вештачка интелигенција. Преку креирање на оваа таксономија, целта е да се дефинира база на податоци со издвоените 10 класи на звучни настани која ќе се користи за препознавање и класификација на избраните звучни настани.

3.2. АНАЛИЗА НА БАЗАТА НА ЗВУЧНИ НАСТАНИ

Врз основа на предложената таксономија се формира база на податоци од звучни настани која ќе се користи за целите на истражувањето во оваа докторска дисертација, конкретно во експерименталниот дел при параметризација на звучните настани и тренирање и тестирање на системите базирани на вештачка интелигенција. Базата на звучни настани е составена од 10 класи на звучни настани кои предизвикуваат вознемиреност кај населението што живее во урбана средина. Базата е составена од 8732 звучни настани со означена звучна класа и е во времетраење од 8 часа и 45 минути. Секој податок е во времетраење од 4 секунди или помалку и означува една класа на звучен настан. Податоците се реални и снимени со различни уреди во урбана средина и се преработени, односно направено е нивно филтрирање, излекување само на звучниот настан и означување на неговата класа. Во прилог А-1 е прикажан кодот кој се искористи за анализа на базата на звучни настани.

На слика 3.7. може да се забележат класите на звучни настани и нивната процентуална застапеност во базата на податоци, од каде што се гледа дека осум класи на податоци имаат приближно подеднаква застапеност, додека останатите две класи на податоци (пукотници и свирка од кола) се застапени помалку поради тоа што овие звуци се инцидентни и можат полесно да се детектираат. Овие две класи се застапени помалку од 50% од средната вредност на податоци од останатите осум класи, а нееднаквата застапеност може да биде причина за појава на помала точност при препознавање на овие две класи на звучни настани. При изборот на звучни настани кои предизвикуваат вознемиреност кај населението, сите класи на звуци се додадени согласно субјективната анализа, изоставувајќи ги звучните класи на играње на улица и сирена, кои се додадени поради тоа што често се појавуваат како извор.



Слика 3.7. Застапеност на одделните класи на звучни настани во базата на звучни настани

Секој податок е однапред означен и содржи информации за: класата на звучниот настан во кој податокот припаѓа и бројот на датотеката во која тој припаѓа. Податоците се поделени во 10 датотеки за полесно управување при нивно тренирање и тестирање. На табела 3.2. се прикажани класите на звучни настани и нивната застапеност по датотеките.

Табела 3.2. Поделеност на класите на податоци во датотеките

	Сирена	Клима уред	Лаење	Гласна музика	Бушење	Играње на улица	Пукотници	Свирка од кола	Работа со чекан	Возило во мирување
Датотека1	86	100	100	100	100	100	35	36	120	96
Датотека2	91	100	100	100	100	100	35	42	120	100
Датотека3	119	100	100	100	100	100	36	43	120	107
Датотека4	166	100	100	100	100	100	38	59	120	107
Датотека5	71	100	100	100	100	100	40	98	120	107
Датотека6	74	100	100	100	100	100	46	28	68	107
Датотека7	77	100	100	100	100	100	51	28	78	106
Датотека8	80	100	100	100	100	100	30	30	78	88
Датотека9	82	100	100	100	100	100	31	32	82	82
Датотека0	83	100	100	100	100	100	32	33	96	83

Пред да се започне со параметризација на звучните настани, најпрвин треба да се анализираат четири важни својства кои помагаат за подобро разбирање на снимените сигнали на звучни настани како: број на аудио канали, стапка на семплирање, длабочина и времетраењето на звучниот настан. Овие својства се важен показател при анализа на базата на звучни настани која ќе се користи при параметризација и класификација на звучните настани. Доколку аудио сигналите од избраната база на звучни настани имаат разни вредности за споменатите својства, постои голема веројатност за неуспешна параметризација на звучните настани и добивање мала прецизност при тренирање и тестирање на податоците. Со други зборови, обработката на сигналот нема да биде точна, и, следствено на тоа, информациите кои ќе ги добива системот за препознавање и класификација на звучните настани ќе бидат погрешни.

Бројот на аудио канали (Number of channels) означува дали сигналот е снимен со еден микрофон (моно сигнал) или два микрофони (стерео сигнал). Од обработката може да се заклучи дека 91.54% од податоците се снимени со два одделни микрофони, додека 8.46% се снимени со еден микрофон. При дигиталната обработка на сигналите, потребно е сигналот да биде снимен со еден звучник. Од тие причини доаѓа до преработување на сигналите на звучните настани од стерео каналите во еден аудио канал. Процесот на конверзија на сигналот на звучен настан не доведува до намалување на квалитетот, ниту пак до губење на важни карактеристики.

Стапката на семплирање (Sample rate) означува колку пати сигналот е прочитан во една секунда. Бидејќи сигналот е семплиран n пати во секунда, периодот на семплирање е на секои $1/n$ секунди. Овој податок е даден во мерна единица Херци (Hz) и го опишува квалитетот на

аудио сигналот. На табела 3.3. се прикажани стапките на семплирање кои се пронајдени во податоците на звучните настани, од каде што може да се заклучи дека стапката на семплирање е помеѓу 8 000 и 192 000 Hz.

Табела 3.3. Процентуална застапеност на податоци согласно стапката на семплирање

Стапка на семплирање (Hz)	Процентуална застапеност
192 000	0.19%
96 000	6.99%
48 000	28.65%
44 100	61.49%
32 000	0.05%
24 000	0.94%
22 050	0.50%
16 000	0.52%
11 025	0.45%
11 024	0.08%
8 000	0.14%

Звучните настани снимени со стапка на семплирање од 8000 Hz до 11 000 Hz означуваат податоци со екстремно низок квалитет, 22 050 Hz означува низок квалитет, стапка на семплирање во вредности од 44 100 Hz до 48 000 Hz се користи за снимање на аудио плејбек (CD/DVD/Streaming), додека стапка на семплирање од 96 000 Hz до 192 000 Hz се користи при професионално аудио снимање. Доколку се анализира стапката на семплирање како референтна вредност која го означува квалитетот на податокот на звучниот настан, може да се забележи разлика помеѓу вредностите. Од табелата може да се забележи дека најголем процент на звучни настани од околу 90% се снимени со вообичаена стапка на семплирање од 44 100 Hz и 48 000 Hz, што значи дека поголемиот дел од звучните настани се снимени со стапка на семплирање која означува висок квалитет снимените податоци. Вообичаено, при конвертирање на податоци со екстремно различна стапка на семплирање може да направи деградација и губење на слушните фреквенции. Од друга страна, примероците со лош квалитет можат да помогнат при тренирање и тестирање на системот со тоа што со нивниот шум ќе предизвикаат потешко распознавање на класата на звучниот настан и ќе спречат појава на презаситување (overfitting).

Длабочината (Bit depth) ја опишува резулцијата, односно степенот на детали со кој е опишан звучниот настан. Оваа анализа покажа дека податоците на звучни настани се снимени со различна длабочина, помеѓу 4 bit и 32 bit, чија што разлика покажува екстремна разлика во квалитетот на снимените податоци. На табела 3.4. е прикажана застапеноста на податоците согласно длабочината со која е снимен звучниот настан.

Табела 3.4. Резолуцијата со која се снимени аудио податоците

Длабочина (Bit depths)	Процентуална застапеност
32	1,94%
24	65,94%
16	31,53%
8	0,49%
4	0,1%

Податоците кои се снимаат со длабочина помеѓу 32bit/24bit се користат за професионални аудио снимки, длабочина од 16bit укажува на квалитет на плејбек аудио сигнал, додека податоците со 8bit/4bit имаат многу ниска резолуција и се смета дека го уништуваат снимениот податок. Најголем број на податоци, околу 65,94% имаат длабочина од 16bit, додека застапеноста на податоците со ниска длабочина е многу мала, околу 0.6% од вкупниот број на податоци. Различните вредности резултираат во различен облик на податоците, што укажува дека е потребна нормализација на оваа карактеристика (сите податоци да имаат иста вредност на длабочина).

Времетраењето на звучниот настан (Time duration) го опишува времето на снимениот податок дадено во мерна единица секунди. Може да се забележи дека најголемиот број на податоци, односно 85,52% (7468 податоци) имаат времетраење помеѓу 3 и 4 секунди, 6% (479 податоци) имаат времетраење помеѓу 1,5 и 3 секунди, додека 8.5% (785 податоци) имаат времетраење помалку од 1 секунда. Од оваа анализа може да се заклучи податоците имаат различно времетраење. При параметризација на звучните настани и формирање на вектор на карактеристики многу е важно сите податоци да имаат еднакво времетраење. Поради тоа, при обработката ќе се примени порамнување на податоците кои имаат времетраење помало од 4 секунди преку примена на процес на додавање на нули во векторот (zero padding). Оваа техника не влијае на алгоритмите при препознавање на карактеристики, напротив, со нејзина примена се подобрува перформансот на алгоритмот на машинско учење.

Вообичаено, при обработка на база на звучни настани со различни својства на податоците, постои голема веројатност за појава на деградација на аудио сигналот и интерпретирање на грешни сигнали. Различни вредности на својствата на снимените звучни настани резултираат во различен облик и структура на самиот податок. Од тие причини, потребно е да се направи нормализација на податоците на звучните настани, односно истите да се обработат на начин што ќе имаат иста форма. Доколку не се направи нормализација, одредени аудио сигнали ќе бидат доминантни во однос на други со што може да се предизвика погрешна информација која ќе доведе до низок степен на предвидување и неефикасен систем за препознавање и класификација на звучните настани.

Согласно анализата на базата на звучни настани која покажа дека снимените податоци имаат различни својства, пред примена на процесот на параметризација на звучните настани, најпрвин ќе се примени техника на нормализација, односно стандардизирање на сигналите кои ги претставуваат звучните настани.

Примена на повисока стапка на семплирање при снимање на звук означува снимање на поширок ранг на фреквенции. Врз основа на теоремата на Никвист, стапката на семплирање овозможува да се дефинира максималниот фреквентен ранг (F_{max}) што може да се содржи во податокот, преку примена на стапката на семплирање (SR) користејќи го односот $F_{max} = SR/2$. Со други зборови, највисоката фреквенција која може да се сними е двојно помала од стапката на семплирање. При ресемплирање на овој податок со стапка од 22 050 Hz, преработениот податок ќе содржи 11 025 Hz. Бидејќи се анализираат звучни настани во урбана средина, значителни промени за овие 10 класи на звуци се случуваат на фреквенции помали од 11 025 Hz, што е доволно да се формира параметар кој ќе биде препознатлив при примена на алгоритмите засновани на вештачка интелигенција. Длабочината со која е снимен сигналот лесно се конвертира и се прави нормализација на сите податоци. Вообичаено секој податок има различна максимална вредност на амплитудата. Преку нормализирање на максималната вредност на амплитудата, се добиваат исти скали на амплитуда помеѓу сите примероци, со што се овозможува полесна екстракција и препознавање на аудио параметрите. Преку примена на постапката на израмнување со додавање на нули во векторот на карактеристики (zero padding), сите податоци се преработени на тој начин што имаат исто времетраење. За оваа преработка се применува библиотеката „LIBROSA“ која е дел од програмскиот јазик „Python“ и го конвертира аудио сигналот со 22 050 Hz стапка на семплирање и длабочина од 16 бита.

3.3. ПАРАМЕТРИЗАЦИЈА НА ЗВУЧНИ НАСТАНИ

Под параметризација на звучни настани се подразбира екстракцијата на карактеристики од аудио сигналот за разбирање важни информации за класата на звучниот настан. Параметрите кои се екстрахираат означуваат компактен збир од карактеристики кој помага при креирање на единствен вектор на карактеристики кој во процесот на машинско учење претставува звучен настан за препознавање. Параметризацијата на звучните настани е прв чекор при изградба на системи за препознавање и класификација на звучните настани, каде што обработените податоци се користат како влезен параметар во алгоритмот на машинско учење. Клучен услов при препознавање и класификација на звучниот настан е екстракција на соодветни параметри на звучни настани кои јасно ја претставуваат класата на звучниот настан и го одвојуваат од останатите класи. Количината на необработени податоци која се наоѓа во снимениот звучен настан е многу голема за директна обработка, а самата параметризација овозможува намалување на количината на податоци и екстракција на карактеристични информации на влезниот податок. Бидејќи карактеристиките на звучниот настан не може да се забележат во временски домен, потребно е тој да се обработи и визуелно да се претстави. По извршената параметризација на звучните настани, формираните вектори на карактеристики се користат како влезен параметар за тренирање и тестирање на алгоритмот на машинско учење.

Согласно анализите, на табела 3.5. се предложени важни својства на параметрите на звучни настани кои доведуваат до појава на ограничувања од аспект на компјутерска комплексност, димензија и статистички својства зависно генерираните информации.

Табела 3.5. Основни својства на параметрите на звучни настани

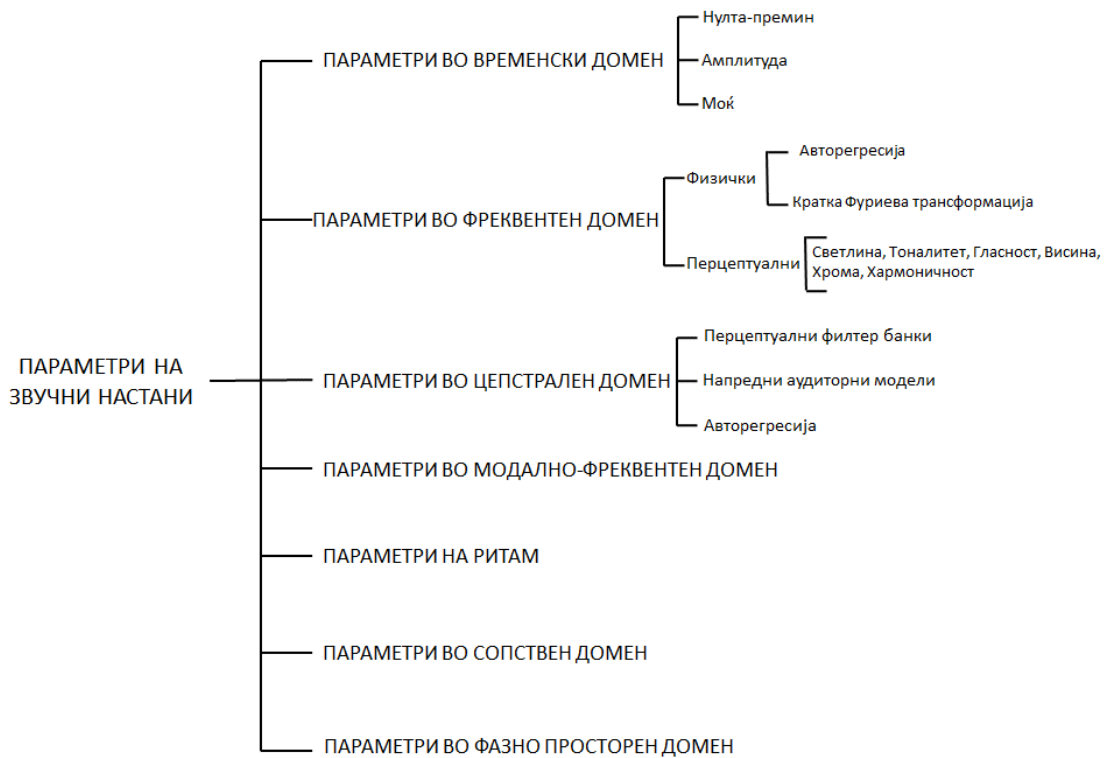
СВОЈСТВА	ОПИС
ВРЕМЕНСКА СКАЛА	Внатрешна рамка, надворешна рамка и глобална рамка
СЕМАНТИЧКО ЗНАЧЕЊЕ	Човечката перцепција, физичкото значење на сигналот
ОСНОВЕН МОДЕЛ	Психоакустичен, Непсихоакустичен
ДОМЕН	Временски домен, фреквентен домен, цепстрален домен, домен на фреквенција на модулација, домен на ритам, сопствен домен и фазен простор

Временската скала е поделена на: внатрешна, надворешна и глобална рамка. Внатрешната рамка го обработува снимениот звучен настан во мал број на независни рамки, додека надворешната рамка овозможува обработка во поголема временска скала со цел да се долови динамиката. Вообичаено, внатрешните рамки се добиваат од надворешните рамки. Глобалната карактеристика се користи кога се параметризира целиот снимен податок.

Семантичката интерпретација покажува дали параметарот е поврзан со човечката перцепција или физичкото значење на снимениот звучен настан, додека основните модели дефинираат психоакустичноста, односно даваат информација дали постои релација помеѓу звучниот настан и аудиторниот систем.

Едно од најважните својства е доменот во кој е преработен звучниот настан бидејќи тој дава информации за процесот на екстракција и компјутерска комплексност. Параметрите на звучните настани се класифицираат во следните домени: временски домен, фреквентен домен, цепстрален домен, модално-фреквентен домен, домен на ритам, сопствен домен и фазно просторен простор. Параметарот може да има трансформација во повеќе домени, но неговиот домен е означен согласно неговата последна трансформација. Предложена е систематизација на класите на параметрите на звучните настани во урбана средина која е прикажана на слика 3.8. и е направена согласно најновите истражувања во оваа област [14,15].

Оваа поделба има неколку нивоа, каде на највисокото ниво е поделбата на параметрите согласно нивниот домен. На следното ниво поделбата е според пресметката на карактеристиките (семантички и пресметковни концепти), кои се поделени согласно структурата која најдобро го опишува својството на параметарот.



Слика 3.8. Поделба на параметрите на звучни настани согласно доменот во кој се обработени

Во рамките на оваа поделба, анализирани се 50 параметри на звучни настани кои се користат во системите за препознавање и класификација на звучните настани во животната средина. Во глава 3.3.1 се прикажани основните математички операции кои се применуваат за екстракција на параметрите на звучните настани, а понатаму е направен преглед во кој се опишани параметрите кои се одвоени за ова истражување. Понатаму, во следните поглавија дадена е математичка репрезентација на избраните параметри кои ќе се користат за целите на ова истражување. За визуелизација на искористените параметри на звучни настани се искористи кодот прикажан во прилог А-2.

3.3.1. МАТЕМАТИЧКИ ОПЕРАЦИИ ЗА ЕКСТРАКЦИЈА НА ПАРАМЕТРИТЕ НА ЗВУЧНИТЕ НАСТАНИ

При екстракција на параметрите на звучните настани се користат повеќе математички операции кои доведуваат до комплексен приказ на математичката интерпретација на параметрите на звучни настани. За да се поедностави приказот, секој параметар може да се опише преку користење на потпис кој претставува компактен приказ на применетите математички операции. Постојат три групи на математички операции кои се користат при екстракција на параметрите на звучните настани: трансформации, филтри и агрегации. На табелите во прилог се прикажани најприменетите математички операции на трансформации (табела 3.6), филтри (табела 3.7) и агрегации (табела 3.8).

Табела 3.6. Трансформации

Трансформации	
A	Автокорелација
R	Крос-корелација
B	Филтер банка
F	Фуриева трансформација
C	(Инверзна) Косинусна трансформација
Q	Константна Q трансформација
M	Модулирана комплексна трансформација
V	Адаптивна временско фреквентна трансформација
W	Дискретна бранова трансформација
E	Временско-фазна трансформација
I	Независна анализа на компонентните
P	Ориентирана анализа на компонентите
S	Декомпозиција на единствена вредност

Табела 3.7. Филтри

Филтри	
b	Филтер на високи звуци
c	Комбиниран филтер
o	Филтер на ниски звуци
f	Врамување
w	Линеарна тежинска функција
d	Деривација, Разлика
e	Спектрална енергетска густина
g	Функција на доцнење
l	Логаритам
x	Експоненцијална функција
n	Нормализација
a	Авторегресија (Линеарна предиктивна анализа)
r	Цепстрална анализа

Табела 3.8. Агрегации и детектори

Агрегации и детектори	
χ	Максимум
ι	Минимум
μ	Средна вредност (нормирана, аритметичка, геометриска)
ϕ	Медијана
Σ	Сума
σ	Девијација, сума на разлики
ϖ	Корен од средна вредност
ω	Можност
H	Ентропија
π	Перцентил
ρ	Регресија
L	Хистограм
β	Спектрална анализа
k	Детекција на пик
ψ	Хармонично детектирање на пик
θ	Полиномен корен
ζ	Детектор на нулта-премин

Секоја математичка операција е претставена со различен карактер (буква), односно трансформациите се опишани со големи латински букви, филтрите со мали латински букви, а агрегациите со мали и големи грчки букви. Трансформациите се операции кои прават трансформација на податоците во различни домени, додека операциите на филтрите ги претставуваат операциите кои се прават во самиот домен во кој се наоѓа параметарот. Целта на филтрите е да го намалат опсегот на вредности на оригиналната серија. Филтрите во доменот на екстракција на параметрите на звучни настани често се базираат на Барк, МЕЛ и ЕРБ скалата, кои служат за доближување до субјективната перцепција на звукот, каде што се истакнуваат фреквенциите кои доведуваат до појава на вознемиреност кај населението. Филтер (или филтер банка) се воспоставува доколку при излезот од филтерот, повторно постои голема серија вредности и служи за делење на звучниот настан во повеќе опсези. Агрегациите се математички операции кои овозможуваат претставување на низа на вредности во еден скалар што доведуваат до намалување на димензионалноста на податоците. Детекторите како дел од агрегациите ја намалуваат димензионалноста на податокот на звучен настан преку откривање на специфични карактеристики од интерес во серија на вредности.

Дополнително на трансформациите, филтрите и агрегациите, потписите содржат два структурни елементи: мала и средна заграда. Малата заграда укажува на алтернативни операции, а средната заграда се применува во случаи кога постои можност за повторување на пресметката во неколку рамки.

За секој од параметрите дополнително ги одредуваме својствата (од табелата 3.5.): временска скала, семантичко значење и психоакустичност. Временската скала се прикажува преку изразите: „I“, „X“ и „G“ кои соодветно ја означуваат внатрешната рамка, надворешната рамка и глобалната рамка. Ознаките „Y“ (да) и „N“ (не) во колоната „перцептуална“ означуваат дали некоја одлика е перцептуална или не. Истото се прави во колоната за „психоакустичен модел“. Понатаму, ја оценуваме пресметковната комплексност на секоја карактеристика, каде што „L“ означува ниска комплексност, „M“ означува средна комплексност и „H“ означува висока комплексност при пресметка.

Дополнително, се прикажува димензијата на параметарот, односно бројот на коефициенти кој се екстрахира за секој параметар, каде ознаката „V“ означува дека бројот на коефициенти е променлив. Во самиот потпис на параметарот е наведен и „доменот на апликација“, каде ознаката „ASR“ значи автоматско препознавање на говор, „ESR“ е препознавање на звук во животната средина, „MIR“ е пребарување на музички информации, „AS“ е аудио сегментација, а „VAR“ означува дека параметарот е применлив во неколку домени на апликации.

Преку примена на овие математички операции, во прилог ќе се прикажат потписите на параметрите на звучните настани кои се анализирани во ова истражување. Параметрите се одвоени согласно доменот во кој се наоѓаат. Овој начин на приказ ќе овозможи компактност и полесен избор на параметри кои ќе се користат при препознавање и класификација на звучните настани.

3.3.2. ПАРАМЕТРИ ВО ВРЕМЕНСКИ ДОМЕН

Параметрите во временски домен (Temporal domain) ги претставуваат промените на сигналот во текот на времето. На апцисата е прикажано времето, додека на ординатата е прикажана амплитудата на звучниот сигнал. Бидејќи самиот звучен настан веќе се наоѓа во временски домен, компјутерската комплексност за пресметка на овие параметри е мала.

Во временскиот домен постојат три групи на параметри поврзани со одредено физичко својство на сигналот: амплитуда, моќ и нулта-премин. На табела 3.9. се прикажани најкористените карактеристики и применетите математички операции за нивна екстракција.

Табела 3.9. Применети математички операции на издвоените параметри во временски домен

Параметар	Временски домен	Перцептуалност	Психоакустичност	Комплексност	Димензија	Апликативен домен	Потпис
Нулта стапка на премин	I	N	N	L	1	VAR	$f \zeta$
Линеарен однос на нулта премин	I	N	N	L	1	ASR	$f \zeta a \zeta$
MPEG-7 аудио бранови	I	N	N	L	2	MIR	$f \chi i$
Дескриптор на амплитудата	I	N	N	L	9	ESR	$f \mu \sigma \zeta \mu \sigma$
Енергија во краток временски период	I	N	N	L	1	VAR	$f \omega$
Јачина на звук	I	N	N	L	1	VAR	$f \omega$

- ПАРАМЕТРИ НА НУЛТА ПРЕМИН

Параметрите на нулта премин (Zero crossings parameters) се основни параметри кои се обработуваат со едноставна временска анализа која ги следи премините на сигналот со нулата на апцисата, обезбедувајќи груба проценка на доминантната фреквенција и спектралниот центар [93]. Карактеристики на нулта-премин се:

- Нулта стапка на премин - параметар го екстрахира вкупниот број на нулта премини во временски интервал од една секунда [94].
- Линеарен однос на нулта премин - се анализира односот на бројот на нулта премини на апцисата и нулата на пресек при линеарна анализа на предвидување [95].

- ПАРАМЕТРИ БАЗИРАНИ НА АМПЛИТУДА

Параметрите базирани на амплитуда (Amplitude-based parameters) се едноставни за пресметување, но се ограничени во нивната брзина. Познати параметри кои се користат во полето на препознавање на звучните настани се:

- MPEG-7 аудио бранови – го прикажуваат обликот на брановата должина преку пресметка на минималните и максималните вредности на амплитудите кои не се повторуваат [96].
- Дескриптор на амплитудата - се пресметува преку собирање на вкупната енергија и времетраењето на сегментите на сигналот кои се генерирани врз основа на нивната максимална и минимална амплитуда.

- ПАРАМЕТРИ БАЗИРАНИ НА МОЌНОСТА

Параметри базирани на моќноста (Power-based parameters) ја претставуваат варијацијата на големината на енергијата и јачината на звукот во текот на времето. Параметри базирани на моќноста се:

- Енергија во краток временски период - средна енергија од моќноста на аудио сигналот [97].
- Јачина на звукот - се пресметува преку апроксимација на средниот квадратичен сигнал RMS во одреден временски интервал [98].

3.3.3. ПАРАМЕТРИ ВО ФРЕКВЕНТЕН ДОМЕН

Во фреквентниот домен (Frequency domain) има најголем сет на параметри кои ја прикажуваат спектралната дистрибуција на снимениот податок, анализирајќи ги хармоничните структури, опсегот на фреквенцијата и тоналноста. Популарни трансформации кои се користат во фреквентниот домен се: Фуриевата трансформација, Косинусната трансформација, Брановата трансформација и Константната Q трансформација. Согласно семантичкото значење, овие параметри се делат на перцептуални и физички параметри во фреквентен домен. Перцептуалните параметри преставуваат информации кои имаат семантичко значење за слушателот, додека физичките карактеристики ги опишуваат сигналите во однос на математичките, статистичките и физичките својства на аудио сигналот. На табелата 3.10. се прикажани издвоените параметри на звучни настани во фреквентен домен.

Табела 3.10. Применети математички операции на издвоените параметри во фреквентен домен

Домен	Параметри	Временски домен	Перцептуалност	Психоакустичност	Комплексност	Димензија	Апликативен домен	Потпис
ФРЕКВЕНТНО - ФИЗИЧКИ	Коефициенти на линеарно предвидување	I	N	N	L	V	ASR	$f(b) a F$
	Побудени карактеристики на линеарно предвидување	I	N	N	M	V	VAR	$f a \theta Q$
	Опсег на однос на енергии	I	N	N	L	V	VAR	$f F \beta e n$
	Спектрални пикови	X	N	N	L	V	MIR	$[F f \chi] d$
ФРЕКВЕНТНО-ПЕРЦЕПТУАЛЕН	Спектрален центроид	G	Y	N	L	1	MIR	$F \mu$
	Спектрален центар	I	Y	N	L	1	MIR	$f F e \phi$
	Спектрален контраст	I	Y	Y	M	7	VAR	$F f l \Sigma \sigma$
	Опсег	I	Y	N	L	1	VAR	$f F \beta (l) \sigma$
	Тонален центроид	I	Y	N	L	6	VAR	$Q f l \beta$
	Спектрална рамнина	I	Y	N	M	V	FP	$f F \beta \mu (l)$
	Специфична сензација на гласност (СОН)	I	Y	Y	H	V	MIR	$f F \beta o l w$
	Интегрална гласност	I	Y	Y	H	1	MIR	$f F l \Sigma w x \Sigma$
	Фундаментална фреквенција	I	Y	N	L	2	VAR	$F A \chi$
	МЕЛ-спектограм	I	N	Y	M	V	VAR	$f F \beta \omega$
	Хроматограм	I	Y	N	L	12	VAR	$f F \beta \omega \Sigma$
	Висински профил	I	Y	N	H	12	MIR	$f Q k \Sigma \chi \Lambda \chi \chi \Sigma$
	Хармоничен коефициент	I	Y	N	L	1	AS	$f A \chi$
	Нехармоничност	I	Y	N	M	1	MIR	$f A \psi \sigma$

- **ФРЕКВЕНТНО ФИЗИЧКИ ПАРАМЕТРИ**

Фреквентно - физичките параметри (Frequency physical parameters) ги опишуваат физичките својства на фреквентниот состав на сигналот. Параметрите се групираат согласно нивниот процес на екстракција, кој се заснова на авторегресија, адаптирана временско-фреквентна карактеристика на декомпозиција и кратка Фуриеова трансформација.

Познати параметри на авторегресија се:

- Коефициенти на линеарно предвидување – се прави линеарна предиктивна анализа на сигналите во фреквентен домен (прикажани се форми и резонанси).
- Побудени карактеристики на линеарно предвидување (CELP) - се воведени од [99] и истите даваат концизни и точни информации звучниот настан преку примена на линеарна анализа на сигналот преработен со константна Q трансформација.

Параметрите на кратка Фуриеова трансформација се пресметуваат преку претставување на сигналот во временски и фреквентен домен, а како најкористени се:

- Опсег на однос на енергии - се дефинира како мерка за нормализирана енергија на сигналот по должина на различните однапред дефинирани групи на фреквенции.

- Спектрални пикови - ги прикажуваат компонентите на енергетскиот спектар на сигналот и можните нарушувања на аудио сигналот. Погодни се за автоматско препознавање на музика (во апликација Шазам - Shazam).

- ФРЕКВЕНТНО ПЕРЦЕПТУАЛНИ ПАРАМЕТРИ

Согласно анализите кои се прават, фреквентно перцептуалните параметри се базираат на светлината, тоналноста, гласноста, висината, хромата и хармоничноста.

Светлината е карактеристика поврзана со рамнотежа на енергијата на сигналот во однос на високите и ниските фреквенции. Се користи за звуци кај кои високата фреквенција е доминантна во однос на ниската фреквенција. Параметри базирани на светлина се:

- Спектрален центроид - ја претставува првата позиција на фреквенцијата на средната вредност на спектарот на големината на рамката на сигналот [100]. Овој центар ја прикажува доминантната фреквенција на сигналот. Се добива преку пресметка на спектарот на моќноста кој е добиен од анализа на октавно-фреквентната скала и дава груб опис за остријата на звукот.
- Спектрален центар - се дефинира како средна фреквенција на сигналниот спектар, каде пониската и повисоката енергија се избалансирани [101]. Овој параметар ја опишува дистрибуцијата на енергија и е во корелација со спектралниот центроид и доминантната фреквенција на сигналот.
- Спектрален контраст – ја прикажува разлика помеѓу врвовите (кои генерално одговараат на хармонични елементи) и долините (каде што нехармоничните или компонентите на бучавата се поддоминантни) преку користење октавен спектар.

Тоналноста е својство кое ја разликува бучавата од тоналните звуци. Звучите кои произлегуваат од бучава имаат континуиран спектар, додека тонските звуци вообичаено имаат линиски спектар. Тоналноста е поврзана со јачината на тонот, а звучите со изразени синусоидални компоненти имаат тенденција да создадат поголема јачина на тонот [102]. Параметри засновани на тоналноста се:

- Опсег - пондиран просек согласно големината на разликите помеѓу спектралните компоненти и спектралниот центроид [103].
- Тонален центроид – проекција на хроматските карактеристики на 6-димензионална основа што ја претставува петтата, помалата трета и поголемата трета проекција, како и дводимензионалните координати.
- Спектрална рамнина - мерка за униформност во фреквентната дистрибуција на спектарот на моќност и може да се пресмета како однос помеѓу геометриската и аритметичката средина на опсегот [104].

Гласноста е поврзана со субјективната перцепција на интензитетот на звук. Параметри на гласноста се:

- Специфичната сензација на гласноста (СОН) – ја опишува перцептивната гласност во дефиниран опсег на фреквенција [105].
- Интегрална гласност - параметар што ја поврзува човечката субјективност со спектралната интеграција на гласноста во повеќе фреквентни групи [106]. Спектралната интеграција на гласноста во повеќе фреквенции овозможува проценка на гласноста на посложените тонови.

Параметар која ја опишува висината на звукот е фундаментална фреквенција која е дефинирана како најниската фреквенција на хармоничниот сигнал.

Најупотребувани хрома параметри кои се користат во системите за детектирање на звуци во животната средина се:

- МЕЛ спектрограм - визуелизација на спектрумот од фреквенциите на звукот во текот на времето базирани на МЕЛ скалата [107].
- Хроматограм – ги опишува 12те различни класи на висина на звук кои ги карактеризираат хармоничните и мелодиските карактеристики на звучниот настан.
- Висински профил - го претставува музичкиот тон, земајќи ги во предвид ударните звуци.

Хармоничноста е својство кое ги разликува периодичните сигнали (хармонични звуци) од непериодичните сигнали (нехармонични звуци и бучава). Параметри на хармоничност се:

- Хармоничен коефициент - карактеристика корисна за разликување на тонски од хармоничност и бучава. При пресметка на хармоничноста се пресметува коефициент на хармоничен однос. Доколку овој коефициент има вредност 1 тогаш се работи за чист хармоничен сигнал, а доколку вредноста е помалку од 0,5 сигналите се слични на бучава.
- Нехармоничност - ги покажува карактеристиките помеѓу пресметаните и теоретските вредности на карактеристиките.

3.3.4. ПАРАМЕТРИ ВО ЦЕПСТРАЛЕН ДОМЕН

Параметрите во цепстралниот домен (Cepstral domain) се добиваат со употреба на Фуриевата трансформација на логаритамот на големината на спектарот [108]. Втората Фуриева трансформација може да се замени со Инверзна Фуриева, Косинусна и Инверзна Косинусна трансформација. Најприменувана е Косинусната трансформација поради подобро сортирање на податоците во однос на Фуриевата трансформација. Цепстралните параметри се компактни претстави на спектарот користејќи ја логаритамската скала. Цепструмот се дефинира како Фуриева трансформација на логаритмот на големината на спектарот на оригиналниот сигнал [109].

На табела 3.11. се прикажани математичките операции на издвоените параметри на звучни настани во цепстрален домен. Постојат три групи на цепстрални параметри согласно

начинот на екстракција: перцептуални филтри, напредни психоакустични модели и модели на авторегресија.

Табела 3.11. Применети математички операции на издвоените параметри во цепстрален домен

Параметар	Временски домен	Перцептуалност	Психоакустичност	Комплексност	Димензија	Апликативен домен	Потпис
МЕЛ фреквентни цепстрални коефициенти	I	N	Y	N	V	VAR	$f F \beta l C$
Барк фреквентни цепстрални коефициенти	I	N	Y	N	V	VAR	$f F \beta l C$
Коефициенти на дискретни брановидни Мел фреквенции	I	N	Y	N	V	ASR	$f A o F l C$
Звучни робусни аудио параметри	I	N	Y	N	256	ESR	$f B w d o l C$
Гаматонски цепстрални коефициенти	I	N	Y	N	V	VAR	$f F \beta l C$
Перцептивно линеарно предвидување	I	N	Y	N	V	ASR	$f F \beta w w C a r$
Коефициенти на линеарно предвидување на цепструмот	I	N	N	M	V	ASR	$f b(a) r$

- ПЕРЦЕПТУАЛНИ ФИЛТЕР БАНКИ

Пресметката на параметрите на перцептуални филтер банки (Perceptual Filter bank-based parameters) е преку пресметување на логаритам на Фуриевата трансформација и користење на филтер банка за делење на сигналот во повеќе опсези.

Најкористени карактеристики од овој домен се:

- Мел фреквентни цепстрални коефициенти (MFCC) - најкористени параметри за класификација на аудио сигналите бидејќи нивната пресметка се базира на перцептивно-субјективна скала на фреквенција (МЕЛ скала) [110].
- Барк фреквентно цепстрални коефициенти - се пресметуваат идентично како MFCC коефициентите, но наместо МЕЛ скалата, се користи Барк скалата [111].
- Коефициенти на дискретни брановидни МЕЛ фреквенции - се добиваат со примена на Дискретна трансформација на енергијата на банката за филтрирање во сигналната рамка [112].

- НАПРЕДНИ АУДИТОРНИ МОДЕЛИ

Најкористени параметри на напредни аудиторни модели (Advanced Auditory Model-based parameters) за анализа на звуци од животната средина кои се наоѓаат во цепстрален домен, а се засновани на напредните аудиторни модели се:

- Звучни робусни аудио параметри - вклучуваат специфичен човечки аудитивен модел заснован на процес во три фази преку доближување до субјективната перцепција, намалување на бучавата и фокусирање на доминантните фреквенции [113].
- Гаматонски цепстрални коефициенти - се пресметуваат преку формирање шема за пресметување преку MFCC коефициентите, но наместо МЕЛ скалата се користи Гаматон скалата. Гаматонските филтри првично се дизајнирани за да го моделираат човечкиот аудитивен спектрален одговор, со оглед на неговото добро приближување во однос на одговор на импулс, одговор на големина и ширина на опсег на филтерот [114].

- АВТОРЕГРЕСИЈА

Цепстрални параметри кои се засновани на моделите на авторегресија (Autoregression based parameters) се:

- Перцептивно линеарно предвидување - претставува спектралната контура преку користење на линеарен пристап заснован на предвидување, кој вклучува и некои специфични својства инспирирани од човечкиот слух [115].
- Коефициенти на линеарно предвидување на цепструмот - се пресметуваат преку примена на инверзна Фуриева трансформација на логаритамската големина на сигналот со спектрален комплекс на линеарно предвидување [116].

3.3.5. ПАРАМЕТРИ ВО МОДАЛНО ФРЕКВЕНТЕН ДОМЕН

Доменот на модална фреквенција (Modal Frequency domain) дава информации за временските модуляции кои се содржат во еден сигнал. На апсцисата се претставени модалните фреквенции, а на ордината акустичните фреквенции. Информациите за модулацијата можат да ја прикажат структурната еволуција на сигналот во текот на времето за секој фреквентен опсег. Математичките операции на издвоените параметри се прикажани на табела 3.12.

Најприменувани параметри во овој домен се:

- Енергија на модулација од 4 Hz - се пресметува преку филтрирање на секоја сигнална спектрална анализа со филтер со опсег од 4 Hz во текот на времето [117].
- Модален спектограм – целта е да се прикаже сигналот во однос на дистрибуцијата на бавните модуляции низ времето и фреквенцијата [118].

Табела 3.12. Применети математички операции на параметрите во модално фреквентен домен

Параметар	Временски домен	Перцептуалност	Психоакустичност	Комплексност	Димензија	Апликативен домен	Потпис
Енергија на модулација од 4 Hz	X	N	Y	M	1	AS	$[f F \beta] b e n \Sigma$
Модален спектограм	X	N	Y	H	V	VAR	$[f F \beta o] W \Sigma$

3.3.6. РИТАМ

Ритмот (Rhythm) претставува важен елемент во музиката, говорот, но исто така и во животната средина и ја дефинира структурна организација на звучните настани (промени во енергијата, висината, итн.). Ритмот вклучува анализа на нискофреквентни модулации, бидејќи може да се опише како повторлив сигнал. Ритмот претставува важен елемент во музиката, говорот и во животната средина и карактеризира со структурна организација на елементите во звучните настани. На табела 3.13. се прикажани математичките операции на издвоените параметри засновани на ритам.

Табела 3.13. Применети математички операции на издвоените параметри на ритам

Параметар	Временски домен	Перцептуалност	Психоакустичност	Комплексност	Димензија	Апликативен домен	Потпис
Цикличен спектар на ритам	X	Y	N	H	V	MIR	$[f F l o] R A$
Шеми на ритам	X	N	Y	H	80	MIR	$[[f F \beta o l w] F w o] \phi$

Најкористени параметри на ритам се:

- Цикличен спектар на ритам - робусен приказ на темпото на музички сигнал во една класа на темпо.
- Шеми на ритам - дводимензионална претстава на фреквенцијата на акустична наспроти модулација, која е изградена врз основа на специфична сензација на гласност, а се добива со Фуриеова трансформација на на критичните опсези во текот на времето.

3.3.7. ПАРАМЕТРИ ВО СОПСТВЕН ДОМЕН

Параметри во сопствениот домен (Eigen domain features) овозможуваат лесно и ефикасно намалување на податоците (на пример, со отстранување на главните компоненти со ниски сопствени вредности). Тие се добиваат од звучни сегменти во времетраење од неколку секунди и се постпроцесирани од алгоритми за намалување на димензионалноста со цел да се добие компактна претстава на информациите за главните сигнали.

Табела 3.14. Применети математички операции на издвоените параметри во сопствен домен

Параметар	Временски домен	Перцептуалност	Психоакустичност	Комплексност	Димензија	Апликативен домен	Потпис
Основа на аудио спектар MPEG-7	X	N	N	H	V	ESR	$[f F \beta l n] S (I)$
Дискриминатор на дисторзии	X	N	N	H	64	FP	$[f M l P] P$

Најкористени параметри прикажани на табела 3.14. се:

- Основа / проекција на аудио спектар MPEG-7 - комбинација од аудио-спектарска основа и проекцијата на аудио-спектарот што го претставува сигналниот спектрограм добиен преку декомпозиција на единечна вредност на аудио сигналот.
- Анализа на дискриминатор на дисторзии - е компактна временска непроменлива претстава на аудио сигналот која се заснова на примена на модулирана комплексна трансформација [119].

3.3.8. ПАРАМЕТРИ ВО ФАЗНО ПРОСТОРОН ДОМЕН

Параметрите во фазно просторен домен се алтернативна перспектива на класичната линеарна застапеност на фреквенцијата. Овие карактеристики ја опишуваат субјективната состојба и нелинеарниот динамичен модел и даваат важни информации за одделување на звучните класи [120]. На табела 3.15. се прикажани математичките операции искористени за екстракција на фазно просторните параметри.

Табела 3.15. Применети математички операции на издвоените фазно просторни аудио карактеристики

Параметар	Временски домен	Перцептуалност	Психоакустичност	Комплексност	Димензија	Апликативен домен	Потпис
Фазно просторни параметри	I	N	N	H	V	ASR	$f E$

3.4. СПЕКТРАЛЕН КОНТРАСТ

Спектралниот контраст претставува фреквентно-перцептуален параметар кој се фокусира на светлината како својство на звучниот настан што е поврзано со рамнотежата на енергијата на сигналот во однос на високите и ниските фреквенции. Се користи за звучни настани кај кои високата фреквенција е доминантна во однос на ниската фреквенција (до 500 Hz). Поради тоа што класите на звучни настани кои се анализираат генерално имаат доминантна фреквенција над 500 Hz, овој параметар ќе се искористи за да се направи екстракција на карактеристиките на звучните настани.

Спектралниот контраст е дефиниран како разлика помеѓу врвовите (кои генерално одговараат на хармонична содржина на звучниот настан) и долините (каде што нехармоничните или компонентите на бучавата се поддоминантни) поделени во октавни филтри. Параметарот ја прикажува релативната дистрибуција на хармоничните и нехармоничните компоненти во фреквентниот спектар. На слика 3.9. се прикажани чекорите за екстракција на спектралниот контраст.



Слика 3.9. Чекори при екстракција на спектрален контраст

Секој звучен настан може да се прикаже во WAV форма (Waveform Audio File), кој претставува континуиран временски сигнал со временски интервал t_N [s]:

$$x_c(t) \quad 0 \leq t_N \leq 4 \tag{3-1}$$

Примената на периодично семплирање со период на семплирање (T) се овозможува трансформација на континуираниот временски сигнал во дискретно време, прикажана преку изразот:

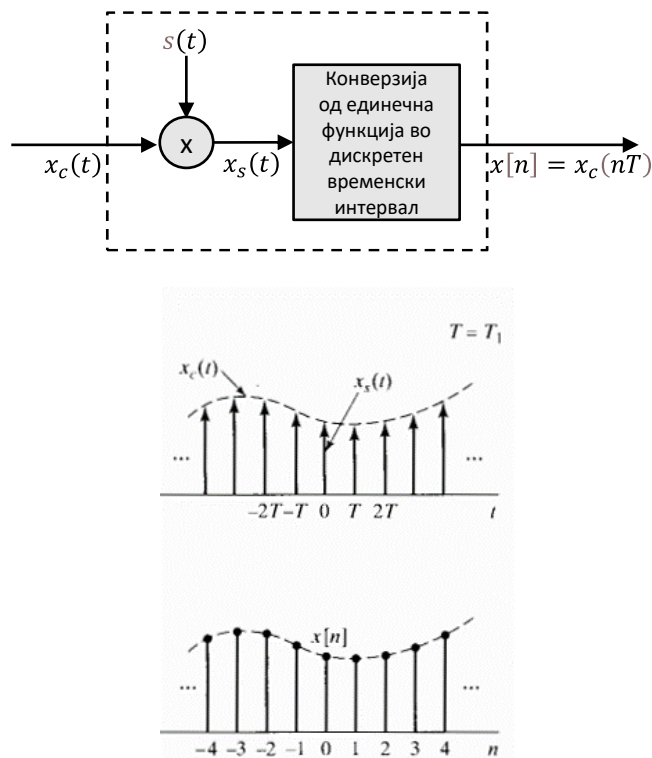
$$x[n] = x_c(nT) \quad -\infty < n < \infty \quad (3-2)$$

Стапката на семплирање (f_s) е реципрочна вредност на периодот на семплирање:

$$f_s = \frac{1}{T} \quad (3-3)$$

За стапка на семплирање со фреквенција $f_s = 22050$ Hz, периодот на семплирање е $T = 0.045$ ms.

Процесот на семплирање на континуиран временски сигнал започнува со преработка на сигналот во сума од единечни отскочни функции, а потоа во трансформирање во дискретен сигнал со дефиниран временски сегмент $x[n]$ (слика 3.10). Процесот на семплирање на континуираниот временски сигнал се прикажува преку конверзија на единечна отскочна функција (Дирак) $x_s(t)$ во временски интервал (секвенца) $x[n]$. Вредностите добиени од семплирање на $x_c(t)$ се прикажани како бројчани вредности во $x[n]$, додека во $x_s(t)$ се прикажани како единечни отскочни функции. Дискретниот временски сигнал $x[n]$ е индексирани согласно променливата n што претставува цел број за конкретниот временски интервал (T).



Слика 3.10. Процес на конверзија на сигналот во дискретно време

Согласно применетата стапка на семплирање f_s и времетраење на сигналот $t_N = 4 s$, дискретниот временски сигнал ќе има $N = f_s \cdot t = 88200$ податоци, односно:

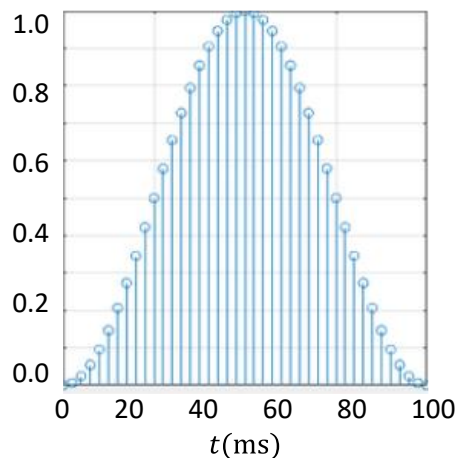
$$x[n] \in R^{88200} \tag{3-4}$$

На овој начин, сигналот се претставува како дводимензионална матрица, каде што редовите го прикажуваат вкупниот број на податоци во текот на времето, а колоните вредноста на амплитудата.

Сигналот кој претставува звучен настан е променлив во текот на времето. За да може да се направи екстракција на акустични карактеристики, потребно е сигналот да се подели на кратки временски периоди, односно да се подели на временски прозорци (w). Од овие причини, следен чекор е делење на дискретниот временски сигнал во кратки временски прозорци кои овозможуваат ефективна обработка на сигналот и спречуваат губење на спектралните елементи при трансформација. Временските прозорци се дефинираат преку избор на тип на прозорец и негово времетраење.

За целите за ова истражување, се применуваат временски рамки во времетраење од $w_f = 100 ms$, која е доволна за да се направи добра спектрална резолуција, додека преклопувањето на $w_p = 50 ms$, овозможува прецизно следење на овие карактеристики во текот на времето и прави прецизна поделба во секоја рамка. Се применува Ханов прозорец, кој е косинусна функција што започнува на амплитуда со вредност 0, достигнува вредност максимална вредност и повторно се спушта со 0, прикажана со изразот:

$$w(t) = 0.5 - 0.5 \cos\left(\frac{2\pi t}{T}\right) \tag{3-5}$$



Слика 3.11. Приказ на Ханов прозорец

Овој прозорец овозможува подобра екстракција на хармониците и изедначување на крајните вредности при понатамошна Фуриеова трансформација.

Доколку $x[n]$ е сигналот кој се дели во временски прозорци, тогаш поделениот сигнал $X[n]$ е претставен со:

$$X[n] = x[n] \cdot w(t) \quad (3-6)$$

Како резултат на делењето во временски прозорци се добива тридимензионална матрица во облик:

$$X[n] \in R^{d \times s \times w} \quad (3-7)$$

Односно, за секој временски прозорец w се добива 2 димензионална матрица во која редовите ја прикажуваат димензијата d , а колоните применетиот чекор s :

$$d = w_f \cdot f_s = 2205 \quad (3-8)$$

$$s = w_p \cdot f_s = 1103 \quad (3-9)$$

$$w = \frac{N - d}{s} = 78 \quad (3-10)$$

Новиот облик на сигналот е $X[n] \in R^{2205 \times 1103 \times 78}$, што значи дека за еден прозорец се добива временски сигнал во облик:

$$\{X[n]\}_w \in R^{2205 \times 1103} \quad 1 \leq w \leq 78 \quad (3-11)$$

За да се добие спектрален прозорец, потребно е да се примени Дискретната Фуриеова трансформација за секој временски прозорец, каде за N коефициенти на фреквенција се добива:

$$\tilde{X}(k) = \sum_{n=0}^{N-1} \{X[n]\} \cdot e^{-i \frac{2\pi k}{N} n} \quad 0 \leq k \leq N - 1 \quad (3-12)$$

Изразот (3-12) во матрична форма е прикажан со:

$$\tilde{X}(k) = \bar{F} \cdot \{X[n]\} \quad (3-13)$$

Во изразот (3-13), \bar{F} е матрица на комплексната Фуријеова трансформација во следниот облик:

$$\bar{F} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & e^{-i\frac{2\pi k}{N}} & e^{-i\frac{4\pi k}{N}} & \dots & e^{-i2\pi\frac{N-1}{N}k} \\ 1 & e^{-i\frac{4\pi k}{N}} & e^{-i\frac{8\pi k}{N}} & \dots & e^{-i2\pi\frac{2(N-1)}{N}k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-i2\pi\frac{N-1}{N}k} & e^{-i2\pi\frac{2(N-1)}{N}k} & \dots & e^{-i2\pi\frac{(N-1)^2}{N}k} \end{bmatrix} \quad (3-14)$$

Примената на Дискретната Фуријеова трансформација на целиот сигнал дава фреквентен спектар, додека пресметката за секоја временска рамка дава спектар во краток временски интервал кој овозможува подобра визуелизација. За да се добие Спектралниот Контраст, потребно е Дискретната Фуријеова трансформација да се пресмета на секој временски прозорец.

Следно, се применуваат октавни филтри кои го делат фреквентниот домен на седум подопсези: 0 – 180 Hz, 180 – 360 Hz, 360 – 720 Hz, 720 – 1440 Hz, 1440 – 2880 Hz, 2880 – 5760 Hz и 5760 – 11025 Hz (за избраната фреквенција на семплирање од 22050 Hz).

За добивање на Спектралниот Контраст, се извлекува големината на спектралниот врв, долина и нивната разлика во секој под опсег, што овозможуваат презентирање на релативните спектрални карактеристики. Релативноста на спектралните карактеристики дава многу повеќе информации за спектарот за разлика од другите карактеристики кои ја прикажуваат средната спектрална моќност. Примената на логаритмирање на добиениот спектар доведува до појава на релативност помеѓу димензиите на пресметаните вредности.

Се претпоставува дека векторот на Фуриевата трансформација од k -тиот опсег е $\{x_{k,1}, x_{k,2}, \dots, x_{k,N}\}$. Големината на спектралните врвови (P_k) и спектралните долини (V_k) се добиваат со користење на изразите:

$$P_k = \log \left\{ \frac{1}{\alpha N} \sum_{i=1}^{\alpha N} x_{k,i} \right\} \quad (3-15)$$

$$V_k = \log \left\{ \frac{1}{\alpha N} \sum_{i=1}^{\alpha N} x_{k,N-i+1} \right\} \quad (3-16)$$

Со цел да се обезбеди точност на самата карактеристика на спектрален контраст, големината на спектралните врвови и спектралните долини се проценува според просечната вредност околу максималната и минималната вредност преку воведување фактор на просечната вредност $\alpha = 0.02$.

Разликата помеѓу спектралните врвови и спектралните долини го дефинира спектралниот контраст (SC_k) за секој октавен подопсег:

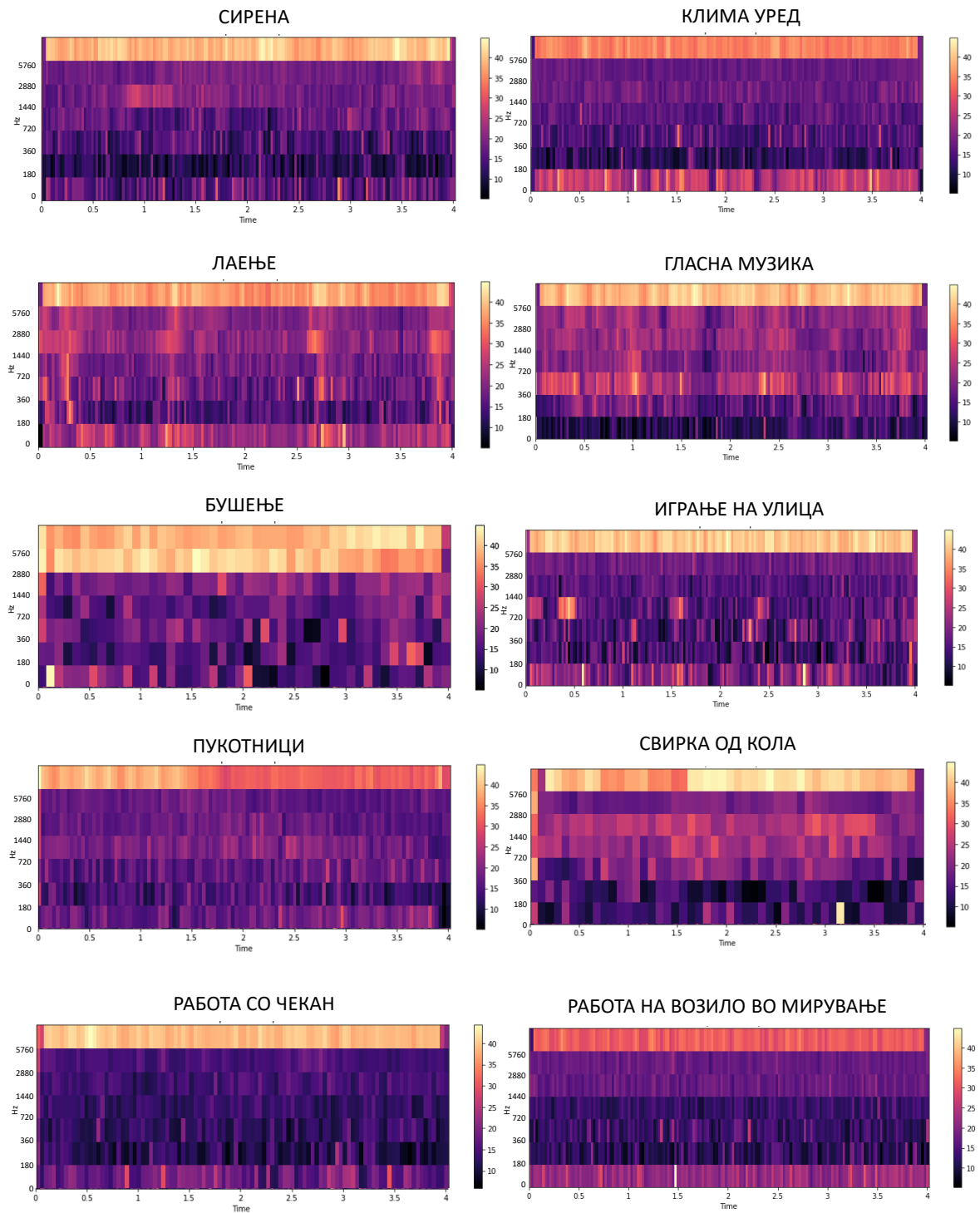
$$SC_k = P_k - V_k \quad (3-17)$$

Пресметаниот спектрален контраст претставува еден вектор на k -тиот подопсег SC_k , а за целосна пресметка на овој параметар, потребно е пресметката да се повтори за секој подопсег $k \in [1,7]$.

За елиминација на релативноста која се јавува во димензионалноста на податоците од сигналот, се применува Кархунен-Лов (Karhunen-Loeve K-L) трансформација, која го претставува стохастичкиот сигнал како бесконечен број на линеарни комбинации на ортогонални функции кои овозможуваат полесно одделување на карактеристиките на звучниот настан. Оваа трансформација го прави процесот на класификација поедноставен и доведува до повисока точност на предвидените резултати.

На слика 3.12. визуелно се прикажани 10 класи на звучни настани кои се избрани од дефинираната база на податоци користејќи го Спектралниот Контраст. Параметрите се добиени преку користење на изведената постапка за екстракција на овој параметар. Одбрани се 10 репрезативни звучни настани, секој претставувајќи една класа на звучен настан. Доколку се погледне визуелниот приказ на параметарот, се забележува дека на апцисата е прикажано времето, додека на ординатата се прикажани октавните подопсези. Боите ги прикажуваат вредностите на Спектралниот Контраст, каде што високите вредности генерално одговараат на јасни сигнали со тесен опсег, додека ниските вредности одговараат на шум со широк опсег.

СПЕКТРАЛЕН КОНТРАСТ

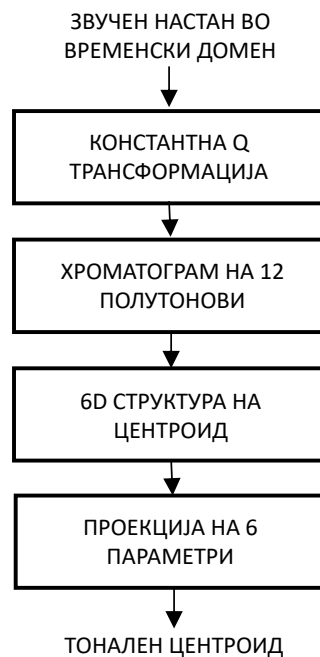


Слика 3.12. Визуелен приказ на спектрален контраст за репрезентативен аудио сигнал за секоја класа на звучен настан

3.5. ТОНАЛЕН ЦЕНТРОИД

Тоналниот центроид е фреквентно-перцептуален параметар кој се базира на тоналноста како својство кое ја разликува бучавата од тоналните звуци. Тоналноста е поврзана со јачината на тонот, а звуците со изразени синусоидални компоненти имаат тенденција да создадат поголема јачина на тонот. За разлика од останатите одбрани 4 параметри кои ја користат Фуриеовата трансформација, овој параметар ја користи Константната Q трансформација за трансформација во фреквентен домен. Тоналниот центроид ќе се користи за параметризација на звучниот настан за да се види дали примената на Константна Q трансформација ќе помогне во подобро одделување на класите на звучни настани.

Процесот на екстракција на тоналниот центроид од снимениот звучен настан има мала компјутерска пресметка и потребни се 6 коефициенти за екстракција на овој параметар. На слика 3.13. се прикажани чекорите за екстракција на тоналниот центроид од снимениот звучен податок.



Слика 3.13. Чекори за екстракција на Тонален Центроид

Звучните настани може да се прикажат во WAV форма како континуиран временски сигнал $x_c(t)$ со временски интервал t_N :

$$x_c(t) \quad 0 \leq t_N \leq 4 \quad (3-18)$$

Континуираниот временски сигнал е преработен во сигнал со дискретно време $x[n]$, а нивната релација е дадена преку изразот:

$$x[n] = x_c(nT) \quad -\infty < n < \infty \quad (3-19)$$

Со стапка на семплирање $f_s = 22050$ Hz, периодот на семплирање е $T = 0.045$ ms. Бидејќи звучниот настан е во времетраење од $t_N = 4$ s, дискретниот временски сигнал има $N = f_s \cdot t = 88200$ податоци, односно:

$$x[n] \in R^{88200} \quad (3-20)$$

Дискретниот временски сигнал е поделен на временски прозорци (w) со примена на Ханов прозорец и временски рамки во времетраење од $w_f = 100$ ms со преклопувањето на $w_p = 50$ ms. Поделениот сигнал $X[n]$ е претставен преку производот на дискретниот временски сигнал и временскиот прозорец:

$$X[n] = x[n] \cdot w(t) \quad (3-21)$$

Како резултат од делењето на сигналот на временски прозорци се добива тридимензионална матрица во облик:

$$X[n] \in R^{2205 \times 1103 \times 78} \quad (3-22)$$

Односно, за еден временски прозорец се добива временски сигнал во облик:

$$\{X[n]\}_w \in R^{2205 \times 1103} \quad 1 \leq w \leq 78 \quad (3-23)$$

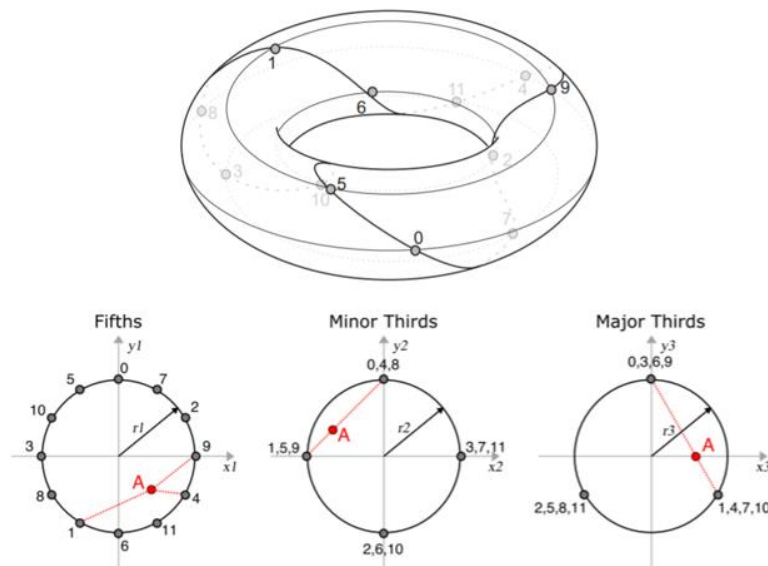
За трансформација во фреквентен домен се применува константната Q трансформација $\tilde{Q}(k)$ прикажана со изразот:

$$\tilde{Q}(k) = \sum_{n=0}^{N-1} \omega(b, k) \{X[n]\} e^{-i\omega b} \quad 0 \leq k \leq N - 1 \quad (3-24)$$

Константната Q трансформација го трансформира сигналот од временски во фреквентен домен која користи централни фреквенции за b филтри во една октава. При пресметка на оваа трансформација за N коефициенти на фреквенција се користи централната фреквенција (ω) во секој подопсег од сигналот поделен во 6 октави со различен фреквентен ранг. Оваа постапка се повторува за секој временски прозорец и за секоја октава.

За една октава се добива дводимензионална матрица (c) со l класи на хроматизам кои ги претставуваат 12те полутонови кои формираат една октава ($0 \leq l \leq 11$). Бидејќи за добивање на Тоналниот Центроид се анализира интонацијата на звучниот настан, математички

ова својство би било прикажано како бесконечна рамнина. Но, доколку се претпостави дека секоја нота од звукот ќе се прикаже во различен ред и дека тие ноти се на еднакво растојание, тогаш со превиткување на оваа рамнина се формира хеликсоидна структура (слика 3.14.), а нотите на звукот се преклопуваат и формираат торус околу структурата. На овој начин се овозможува проектирање на акордите (група од три ноти кои се присутни во ист период) и нивната висината во торусната структура.



Слика 3.14. Проекција која покажува како елементите на тоналниот центроид се обвиткуваат околу торусот. Трите кружници се визуелизација на 6-D просторот.

Параметрите кои го претставуваат тоналниот центроид се наоѓаат во 6 димензионален просторот составен од површината на торусна структура. Поради тоа што 6 димензионалниот простор не може директно да се визуелизира, се претпоставува дека тој е проекција на трите кружни структури во торусот: кругот на петтата проекција, кругот на малата трета проекција и кругот на големата трета проекција. Преку проектирање на двете оски (x и y) на секоја од трите кружници на точката A која се добива со поврзување на 1,4 и 9та класа на висина на звучниот настан, се добиваат 6 димензии (коефициенти) што го претставуваат секој подопсег на овој параметар: $5x$, $5y$, $m3x$, $m3y$, $M3x$ и $M3y$. Секоја димензија прикажува хармоничност на различен фреквентен ранг, земајќи ја во предвид и висината на тоновите. Петтата проекција претставува интервал кој одговара на висина на тонот со фреквентен однос 3:2, малата трета проекција е составена од два полутона кои имат фреквентен однос 6:5, додека големата трета проекција има фреквентен однос 5:4, што е еднаква на четири последователни полутоноти.

Оваа проекција доведува до формирање на векторот на тоналниот центроид (ϑ) за секој опсег (k). За еден временски прозорец (w), векторот на тоналниот центроид се добива преку производот од матрицата (c) и трансформационата матрица (Φ) со изразот:

$$\vartheta_n(k) = \frac{1}{\|c_w\|_1} \sum_{l=0}^{11} \Phi(k, l) c_w(l), \quad 0 \leq k \leq 5, \quad 0 \leq l \leq 11 \quad (3-25)$$

Трансформационата матрица Φ ја претставува проекцијата на 6-димензионалниот простор и е прикажана со изразот:

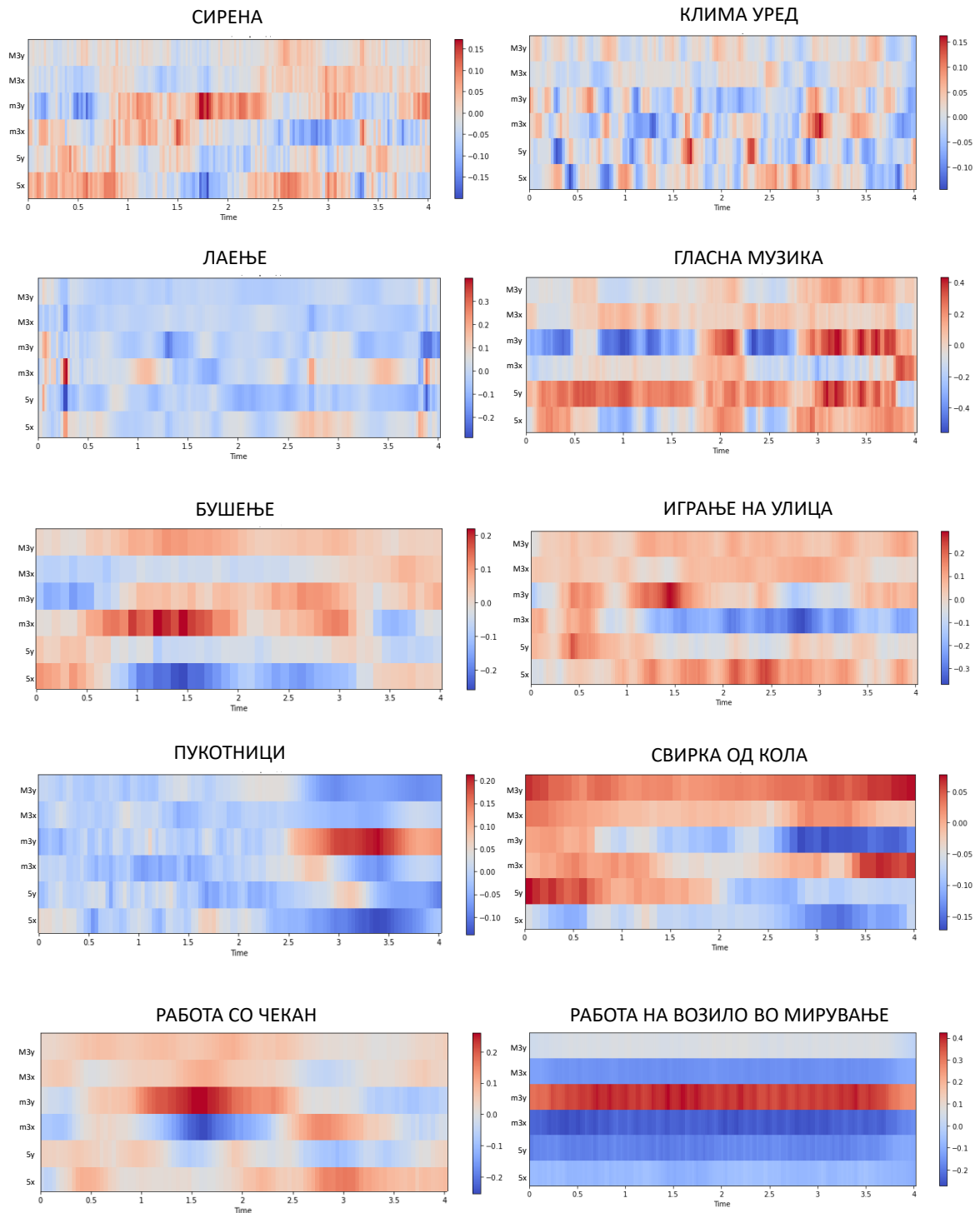
$$\Phi = [\phi_0, \phi_1 \dots \phi_{11}] \quad (3-26)$$

$$\phi_i = \begin{bmatrix} \Phi(0, l) \\ \Phi(1, l) \\ \Phi(2, l) \\ \Phi(3, l) \\ \Phi(4, l) \\ \Phi(5, l) \end{bmatrix} = \begin{bmatrix} r_1 \sin l(7\pi/6) \\ r_1 \cos l(7\pi/6) \\ r_2 \sin l(2\pi/3) \\ r_2 \cos l(2\pi/3) \\ r_3 \sin l(2\pi/3) \\ r_3 \sin l(2\pi/3) \end{bmatrix} \quad 0 \leq l \leq 11 \quad (3-27)$$

Вредностите r_1 , r_2 и r_3 се радиуси на трите проекции прикажани на слика 3.14, а нивните вообичаени вредности се 1,1 и 0.5.

Тоналниот центроид ги проектира карактеристиките на хроматограмот на 6-димензионална основа што ги претставува проекциите на петтата (5y и 5x), помалата трета (m3y и m3x) и поголемата трета проекција (M3y и M3x). На слика 3.15. е прикажан параметарот на Тоналниот Центроид за 10те репрезентативни класи на звучни настани. На апцисата е прикажано времето, а на ординатите се прикажани подопсезите на бте димензии (M3y, M3x, m3y, m3x, 5y и 5x). Бојата прикажува позитивни и негативни вредности во ранг од -0.4 до 0.4, каде што позитивните вредности прикажуваат хармоничност, а негативните вредности нехармоничност во самиот звучен настан што овозможува да се детектираат промени на хармоничната структура на звучните настани.

ТОНАЛЕН ЦЕНТРОИД



Слика 3.15. Визуелен приказ на тонален центроид за репрезентативен аудио сигнал за секоја класа на звучен настан

3.6. МЕЛ СПЕКТРОГРАМ

МЕЛ спектрограм претставува фреквентно-перцептуален параметар кој овозможува визуелизација на спектарот од фреквенции согласно МЕЛ скалата во текот на времето. Анализата на најновите истражувања покажува дека овој параметар често се применува во системите за препознавање и класификација на звучните настани поради неговата мала компјутерска комплексност при пресметка и високата точност која ја постигнува при примена на алгоритмите базирани на длабоко учење. Поради тоа што во ова истражување се користат и стандардни алгоритми на машинско учење, овој параметар ќе се искористи за да се спореди неговата ефективност при употреба на стандардните алгоритми и алгоритмите на длабоко учење. На слика 3.16. се прикажани чекорите за екстракција на МЕЛ спектрограм.



Слика 3.16. Чекори на обработка на аудио сигналот за добивање МЕЛ спектрограм

Секој звучен настан може да се прикаже во WAV форма како континуиран временски сигнал со временски интервал t_N [s]:

$$x_c(t) \quad 0 \leq t_N \leq 4 \quad (3-28)$$

Процесот на семплирање овозможува трансформацијата на континуираниот временски сигнал во дискретно време, а нивната корелација е прикажана со изразот:

$$x[n] = x_c(nT) \quad -\infty < n < \infty \quad (3-29)$$

Имајќи во предвид дека стапката на семплирање е реципрочна вредност на периодот на семплирање, за стапка на семплирање $f_s = 22050$ Hz, периодот на семплирање е $T =$

0.045 ms. Бидејќи звучниот настан е во времетраење од $t_N = 4 s$, дискретниот временски сигнал има $N = f_s \cdot t = 88200$ податоци, односно:

$$x[n] \in R^{88200} \quad (3-30)$$

За целите за ова истражување, дискретниот временски сигнал се дели на временски прозорци (w) со примена на Ханов прозорец и временски рамки во времетраење од $w_f = 100 ms$ со преклопувањето на $w_p = 50 ms$. Сигналот поделен на временски рамки $X[n]$ е претставен со изразот:

$$X[n] = x[n] \cdot w(t) \quad (3-31)$$

Како резултат на делење во временски прозорци се добива тридимензионална матрица со во облик:

$$X[n] \in R^{2205 \times 1103 \times 78} \quad (3-32)$$

Односно, за секој временски прозорец w се добива 2 димензионална матрица во која редовите ја прикажуваат димензијата, а колоните применетиот чекор. За еден прозорец се добива временски сигнал во облик:

$$\{X[n]\}_w \in R^{2205 \times 1103} \quad 1 \leq w \leq 78 \quad (3-33)$$

Се применува Дискретна Фуријеова трансформација на секој временски прозорец, каде за N коефициенти на фреквенција се добива:

$$\tilde{X}(k) = \sum_{n=0}^{N-1} \{X[n]\} \cdot e^{-i\frac{2\pi k}{N}n} \quad 0 \leq k \leq N-1 \quad (3-34)$$

Примената на Дискретната Фуријеова трансформација на целиот сигнал дава фреквентен спектар, додека пресметката за секоја временска рамка дава спектар во краток временски интервал кој овозможува подобра визуелизација. Со пресметка на Дискретната Фуријеова трансформација за секоја рамка се добива спектрограм.

Како резултат на применетата Дискретна Фуријеова трансформација, сигналот во еден временски прозорец се преработува во фреквентен домен и се добива нова матрица, каде што редовите се индексирани според фреквенцијата, а колоните според времето. Секоја вредност во $\{X[n]\}$ одговара на точка согласно одредена фреквенција и време. Димензијата на еден прозорец преработен во фреквентен домен е:

$$\tilde{X}(k) \in R^k \times 1103 \quad (3-35)$$

Спектрограмот го прикажува преработениот звучен настан во линеарна скала. Поради тоа што човекот го перцепира звукот логаритамски, потребно е спектрограмот да се преработи со цел да може да се забележат клучни елементи кои ќе доведат до одвојување на класите на звучниот настан. За таа цел, се користи МЕЛ скала, која го преработува линеарниот спектрограм во МЕЛ спектрограм. Студиите покажале дека луѓето не ги перципираат фреквенциите на линеарна скала, односно полесно е да се откријат разлики во пониски фреквенции отколку повисоки фреквенции. На пример, лесно може да се детектира разлика од 500 Hz и 1000 Hz, но тешко може да се направи разлика помеѓу 10 000 Hz и 10 500 Hz. Референтната точка помеѓу оваа скала и нормалното пресметување на фреквенцијата се дефинира со доделување на перцептивна висина на звук со големина од 1000 МЕЛ на звук со фреквенција од 1000 Hz и висина од 40 dB.

Спектарот на МЕЛ скала се пресметува преку конверзија на сигналот во фреквентен домен преку сет од филтри кои се наречени МЕЛ филтер банки. Конверзијата на фреквенцијата (f) од Херци во МЕЛ скала (f_{Mel}) е претставена со изразот:

$$f_{Mel} = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \quad (3-36)$$

Формата на МЕЛ филтерот е триаголна. При конверзија на фреквенциите во МЕЛ скала, најпрвин се избира бројот на МЕЛ филтер банки, а потоа истите се конструираат и го преработуваат спектрограмот од линеарна во МЕЛ скала. Креирањето на овие филтер банки започнува со конверзија на највисоката (f_{max}) и најниската фреквенција (f_{min}) во МЕЛ скала преку употреба на изразот (3-34).

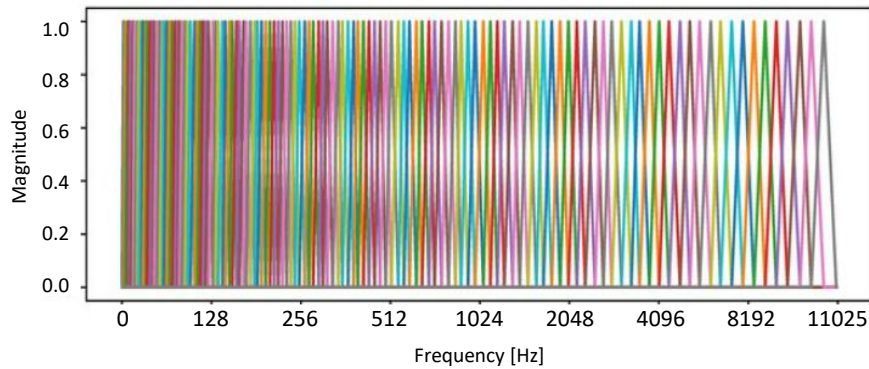
Преку дефинирање на рангот на највисоката и најниската фреквенција, и нивна подеднаква поделба согласно бројот на МЕЛ филтри m , се добиваат МЕЛ филтер банките, односно:

$$M^{-1}(\text{linspace}(M(f_{min}), M(f_{max}), M + 2)) \quad (3-37)$$

За екстракција на карактеристиката МЕЛ спектрограм се излекуваат 128 МЕЛ филтер банки ($M = 128$) кои се прикажани како дводимензионална матрица во облик:

$$M \in R^{128 \times K} \quad (3-38)$$

На слика 3.17. се прикажани 128 МЕЛ филтер банки кои се користат за преработка на спектрограмот.



Слика 3.17. Приказ на 128 МЕЛ филтер банки со триаголен облик

Доколку се погледне сликата 3.17, ќе се забележи дека филтрите во фреквентен домен не се распоредени линеарно, но во МЕЛ скалата тие се линеарно поделени. За да се направи пресметката, треба да се внимава бројот на колоните на матрицата на МЕЛ филтер банка да ист со бројот на редовите на матрицата на спектралниот прозорец.

МЕЛ спектрограмот се пресметува од линеарниот спектрограм, $\tilde{X}(k)$, преку множење на неговата големина со секој од триаголните МЕЛ филтер банки (M):

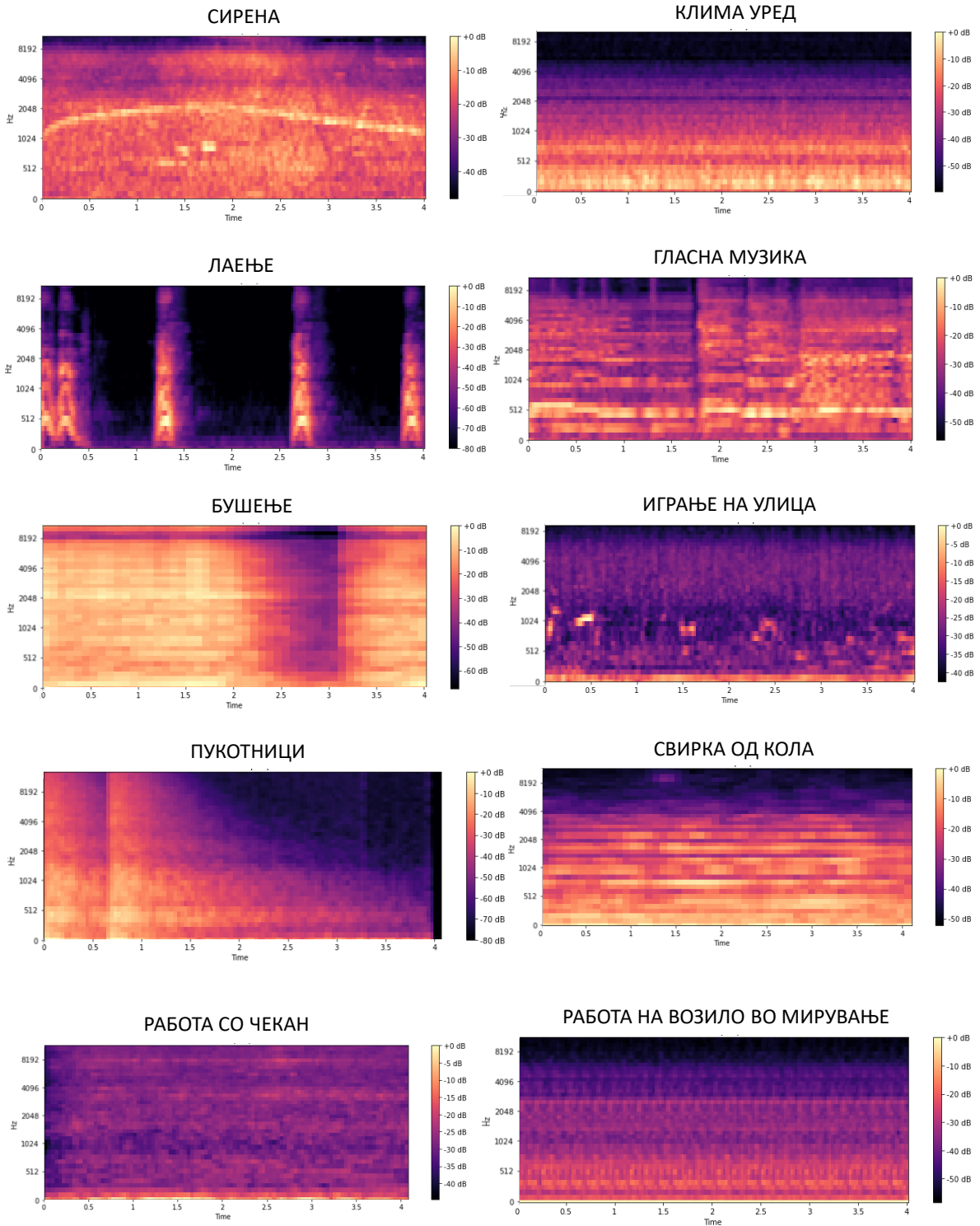
$$s(m) = \sum_{k=0}^{N-1} [|\tilde{X}(k)|^2 H_m(k)] \quad 0 \leq m \leq M - 1 \quad (3-39)$$

Тежинскиот фактор прикажан $H_m(k)$ го претставува $k^{\text{ТИОТ}}$ енергетски спектар кој влијае на $m^{\text{ТИОТ}}$ излезен филтер, прикажан со изразот:

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1) - k}{f(m+1) - f(m)} & f(m) \leq k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases} \quad (3-40)$$

Примената на горенаведената постапка овозможува приказ на МЕЛ спектрограмот за 10те класи на репрезентативни звучни настани од анализираната база на податоци (слика 3.18). Преку приказ на времето на апцисата, фреквентниот спектар согласно МЕЛ скалата на ординатата и моќноста на амплитудата во dB прикажана преку различни бои на оската се формира МЕЛ спектрограм, кој овозможува визуелизација на концентрирани фреквенции и амплитуди. Секоја временска рамка одговара на вертикалната линија на визуелизираниот параметар што ја прикажува амплитудата наспроти фреквенцијата за секој момент во времето. Рангот на моќноста е од 0 до -80 dB, каде што 0 dB означува однос 1 помеѓу моќноста и амплитудата, а пониските вредности означуваат помала моќност, односно помала застапеност на одредени фреквенции.

МЕЛ СПЕКТРОГРАМ



Слика 3.18. Визуелен приказ на МЕЛ спектрограм за репрезентативен аудио сигнал за секоја класа на звучен настан

3.7. ХРОМАТОГРАМ

Хроматограмот претставува фреквентно-перцептуален параметар кој ги анализира 12те висински класи на тонот во рамките на една октава. Една висинска класа ги означува сите висини на звукот кои имаат иста хрома. Едно од главните својства хроматограмот е тоа што овој параметар ги прикажува хармоничните и мелодиските карактеристики на аудио сигналот. Овој звучен параметар е одбран поради малата компјутерска комплексност за пресметка и различните хармонични структури кои се јавуваат во секоја класа на звучен настан. Главната идеја на хрома параметрите е да ја агрегираат цела спектрална информација која одговара на дадена класа на висина на тонот во еден коефициент. На слика 3.19. се прикажани чекорите за екстракција на хроматограмот.



Слика 3.19. Обработка на снимен звучен настан за екстракција на хроматограм

Звучните настани може да се прикажат во WAV форма како континуиран временски сигнал со временски интервал t_N [s]:

$$x_c(t) \quad 0 \leq t_N \leq 4 \tag{3-41}$$

Континуираниот временски сигнал се преработува во сигнал со дискретно време преку релацијата:

$$x[n] = x_c(nT) \quad -\infty < n < \infty \tag{3-42}$$

За стапка на семплирање од $f_s = 22050$ Hz, периодот на семплирање е $T = 0.045$ ms. Бидејќи звучниот настан е во времетраење од $t_N = 4$ s, дискретниот временски сигнал има $N = f_s \cdot t = 88200$ податоци, односно:

$$x[n] \in R^{88200} \quad (3-43)$$

Дискретниот временски сигнал е поделен на временски прозорци (w) со примена на Ханов прозорец и временски рамки во времетраење од $w_f = 100$ ms со преклопувањето на $w_p = 50$ ms. Поделениот сигнал $X[n]$ е производ од сигналот во дискретен временски период и применетиот временски прозорец:

$$X[n] = x[n] \cdot w(t) \quad (3-44)$$

Како резултат на делење во временски прозорци се добива тридимензионална матрица во облик:

$$X[n] \in R^{2205 \times 1103 \times 78} \quad (3-45)$$

За еден временски прозорец се добива временски сигнал во облик:

$$\{X[n]\}_w \in R^{2205 \times 1103} \quad 1 \leq w \leq 78 \quad (3-46)$$

Се применува Дискретна Фуриеова трансформација на секој временски прозорец, каде за $N = 128$ коефициенти на фреквенција се добива:

$$\tilde{X}(k) = \sum_{n=0}^{N-1} \{X[n]\} \cdot e^{-i\frac{2\pi k}{N}n} \quad 0 \leq k \leq N-1 \quad (3-47)$$

Како резултат на применетата Дискретна Фуриеова трансформација, сигналот во еден временски прозорец се преработува во фреквентен домен и се добива нова матрица, каде што редовите се индексирани според фреквенцијата, а колоните според времето. Секоја вредност во $\{X[n]\}$ одговара на точка согласно одредена фреквенција и време. Се добива матрица за преработениот сигнал во фреквентен домен во следниот облик:

$$\tilde{X}(k) \in R^{128 \times 1103} \quad (3-48)$$

Хроматограмот се добива преку одделување на енергијата содржана во спектрограмот на 12 висински класи (секоја класа претставува еден полутон). Овие 12 висински класи кои формираат една октава се претставени:

$$c \in [C, C\#, D, D\#, E, F, F\#, G, G\#, A, A\#, B] \quad (3-49)$$

Една висинска класа е дефинирана со сет на висини кои делат иста хрома. На пример, во висинската класа на хромата C се вклучени сите висини на тонот кои се одделени со цел број на октави:

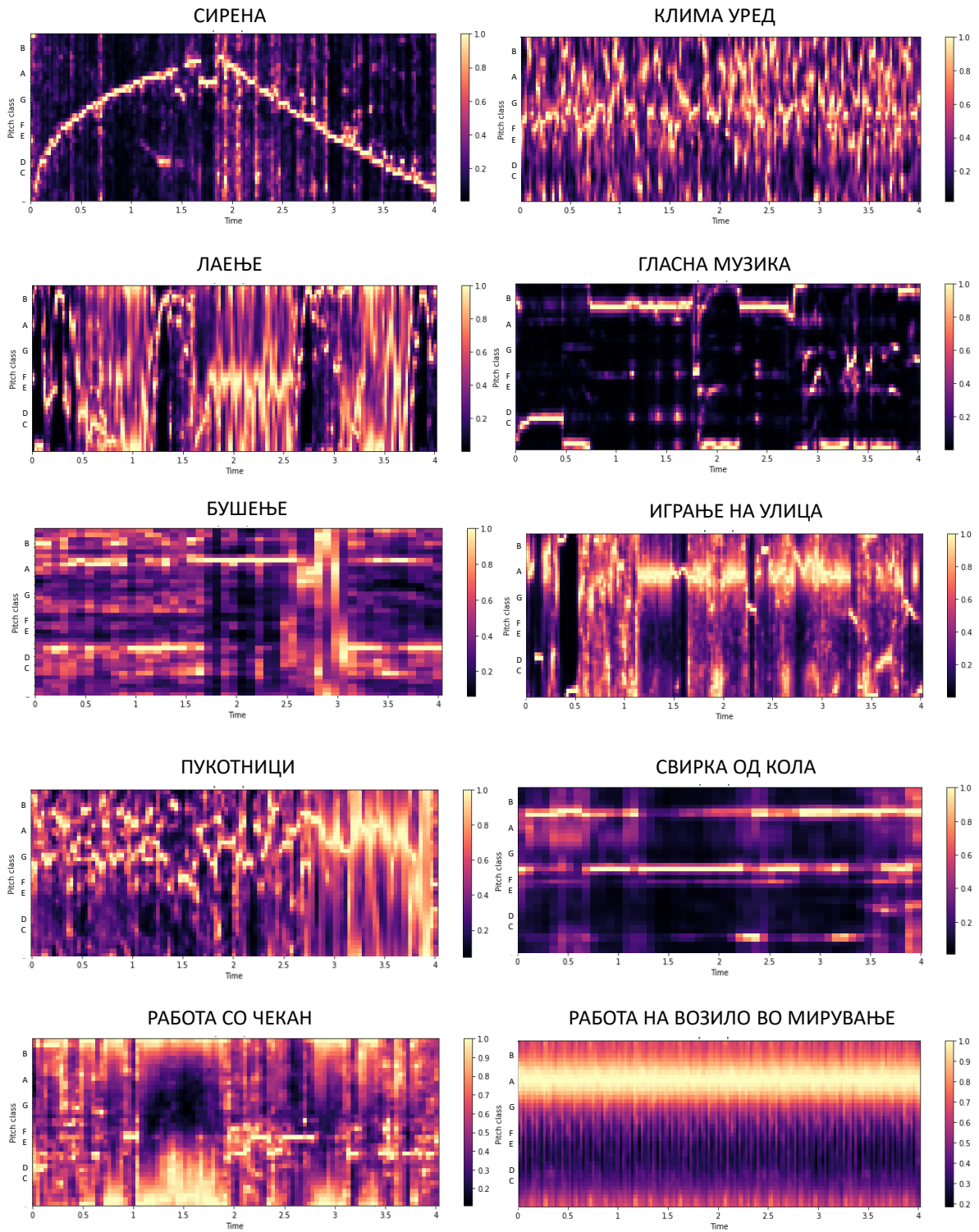
$$C \in \{\dots, C0, C1, C2, C3, \dots\} \quad (3-50)$$

Коефициентите на хроматограмот $\tilde{C}(c)$ се екстрахираат од спектрограмот $\tilde{X}(k)$ преку собирање на сите висински коефициенти кои припаѓаат на иста хрома:

$$\tilde{C}(c) = \sum_{\{k \in [0:127]: k \bmod 12 = c\}} \tilde{X}(k) \quad 0 \leq c \leq 11 \quad (3-51)$$

Пресметка на хроматограмот се повторува 12 пати со цел да се пресмета содржината на енергијата во секоја висинска класа. На слика 3.20. визуелно е прикажан Хроматограм за 10те репрезентативни звучни настани за секоја класа на звучен настан. Хроматограмот ја прикажува распределбата на енергијата низ времето и хромата. На апцисата е прикажано времето, додека на ординатата се прикажани 12те коефициенти, односно висинските класи. Енергијата е претставена преку боите, а на скалата таа е во вредност од 0 до 1, каде што 1 означува постоење на одредена висинска класа.

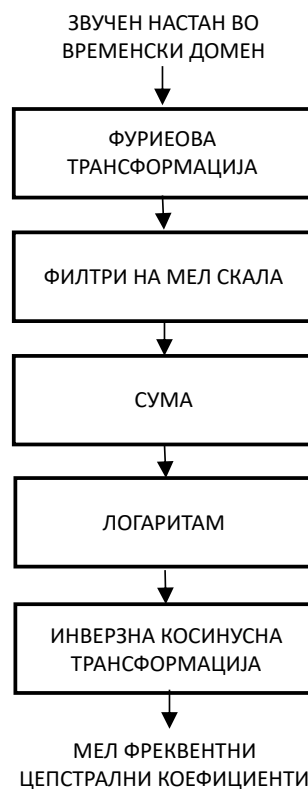
ХРОМАТОГРАМ



Слика 3.20. Визуелен приказ на хроматограм за репрезентативен аудио сигнал за секоја класа на звучен настан

3.8. МЕЛ ФРЕКВЕНТНИ ЦЕПСТРАЛНИ КОЕФИЦИЕНТИ

МЕЛ Фреквентни Цепстрални Коефициенти (MFCC) се најкористени параметри за класификација на звучните настани бидејќи постигнуваат висока ефикасност поради двојната трансформација која го прикажува сигналот во цепстрален домен, базиран на МЕЛ скалата. Цепстралниот домен се добива со примена на две трансформации, првата, Дискретна Фуриеова трансформација која служи за добивање на спектар на фреквенции, а втората, Инверзна Косинусна Трансформација која служи за преработка на овој спектар во цепструм [99]. Параметрите кои се обработуваат во цепстрален домен имаат висока компјутерска комплексност при пресметка и постигнуваат висока точност во системите за препознавање и класификација на звучни настани. На слика 3.21. се прикажани чекорите за екстракција на овој параметар.



Слика 3.21. Чекори за добирање на МЕЛ фреквентни цепстрални коефициенти

Секој звучен настан може да се прикаже во WAV форма како континуиран временски сигнал со временски интервал t_N [s]:

$$x_c(t) \quad 0 \leq t_N \leq 4 \quad (3-52)$$

Периодичното семплирање го трансформира континуираниот временски сигнал во дискретно време преку релацијата:

$$x[n] = x_c(nT) \quad -\infty < n < \infty \quad (3-53)$$

Имајќи во предвид дека стапката на семплирање е реципрочна вредност на периодот на семплирање, за користена стапка на семплирање од $f_s = 22050$ Hz, периодот на семплирање е $T = 0.045$ ms. Бидејќи звучниот настан е во времетраење од $t_N = 4$ s, дискретниот временски сигнал има $N = f_s \cdot t = 88200$ податоци, односно:

$$x[n] \in R^{88200} \quad (3-54)$$

За целите за ова истражување, дискретниот временски сигнал се дели на временски прозорци (w) со примена на Ханов прозорец и временски рамки во времетраење од $w_f = 100$ ms со преклопувањето на $w_p = 50$ ms. Поделениот сигнал $X[n]$ е претставен со:

$$X[n] = x[n] \cdot w(t) \quad (3-55)$$

Тридимензионалната матрица како резултат на делење во временски прозорци е во облик:

$$X[n] \in R^{2205 \times 1103 \times 78} \quad (3-56)$$

Односно, за секој временски прозорец w се добива 2 димензионална матрица во која редовите ја прикажуваат димензијата, а колоните применетиот чекор. За еден прозорец се добива временски сигнал во облик:

$$\{X[n]\}_w \in R^{2205 \times 1103} \quad 1 \leq w \leq 78 \quad (3-57)$$

Се применува Дискретна Фуриеова трансформација на секој временски прозорец, каде за N коефициенти на фреквенција се добива:

$$\tilde{X}(k) = \sum_{n=0}^{N-1} \{X[n]\} \cdot e^{-i\frac{2\pi k}{N}n} \quad 0 \leq k \leq N-1 \quad (3-58)$$

Како резултат на применетата Дискретна Фуриеова трансформација, сигналот во еден временски прозорец се преработува во фреквентен домен и се добива нова матрица, каде што

редовите се индексирани според фреквенцијата, а колоните според времето. Секоја вредност во $\{X[n]\}$ одговара на точка согласно одредена фреквенција и време. Димензијата на еден прозорец преработен во фреквентен домен е:

$$\tilde{X}(k) \in R^{k \times 1103} \quad (3-59)$$

Спектрограмот го прикажува преработениот звучен настан во линеарна скала. Поради тоа што човекот го перцепира звукот логаритамски, спектрограмот се преработува користејќи ја МЕЛ скалата. За екстракција на МЕЛ фреквентни цепстрални коефициенти се излекуваат 40 МЕЛ филтер банки ($M = 40$) кои се прикажани како дводимензионална матрица во облик:

$$M \in R^{40 \times K} \quad (3-60)$$

Преку множење на големината на спектрограмот со секој од триаголните МЕЛ филтер банки (M) претставени преку тежинскиот фактор $H_m(k)$ се добива МЕЛ спектрограм:

$$s(m) = \sum_{k=0}^{N-1} [|\tilde{X}(k)|^2 H_m(k)] \quad 0 \leq m \leq M - 1 \quad (3-61)$$

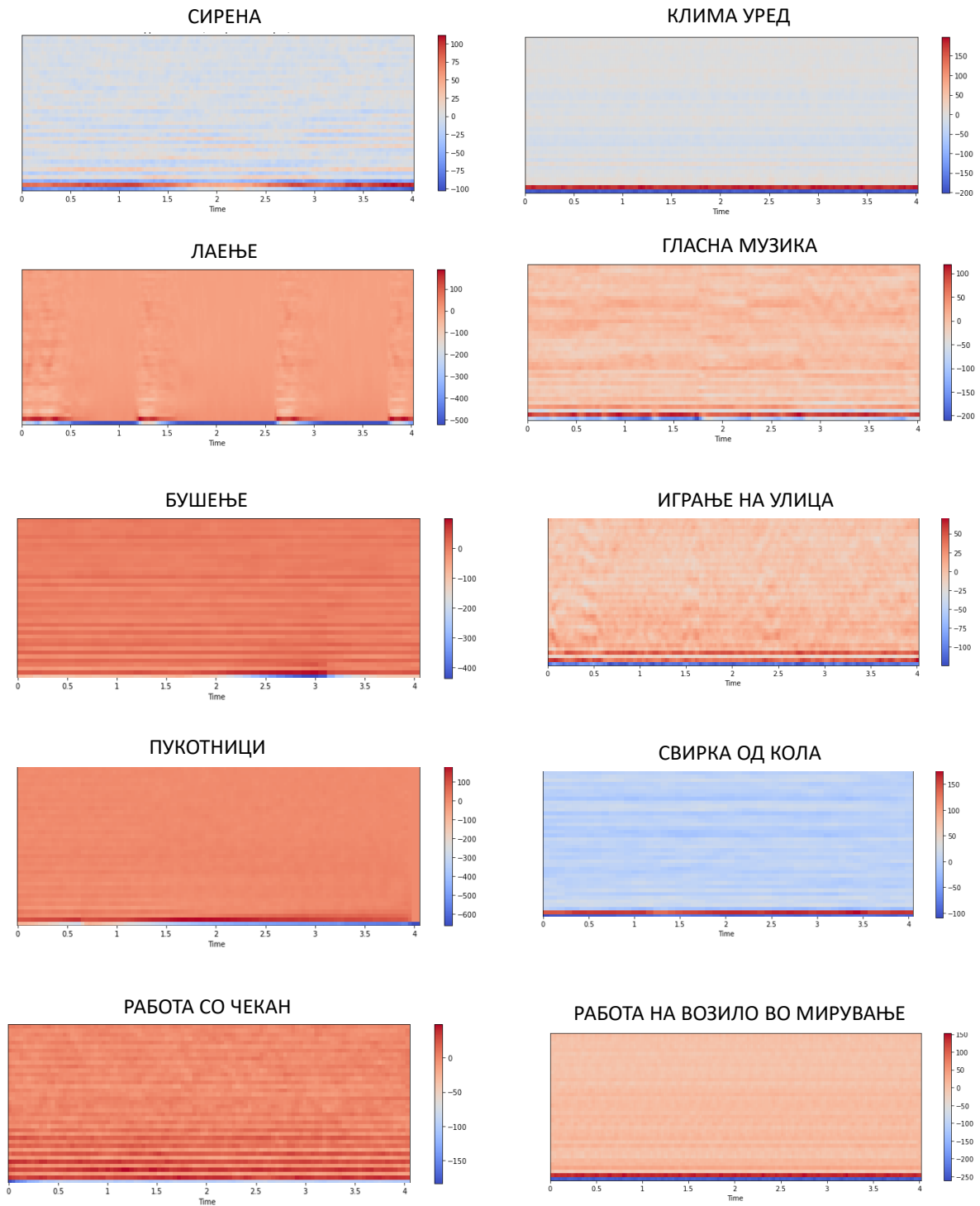
Со примената на повторна трансформација се добива спектар на спектарот, кој е наречен цепструм, кој се добива со примена на логаритам и Инверзна Косинусна трансформација. Косинусна трансформација е поедноставена верзија на Фуриевата трансформација, каде што коефициентите се претставени со реални броеви. МЕЛ Фреквентните Цепстрални коефициенти $\hat{C}(n)$ се добиваат со користење на изразот:

$$\hat{C}(c) = \sum_{m=0}^{M-1} (\log(s(m))) \cos\left(\frac{\pi n(m - 0,5)}{M}\right) \quad 0 \leq c \leq C - 1 \quad (3-62)$$

Постапката за пресметка на MFCC дадена со изразот (3-60) се повторува за сите $C = 40$ коефициенти.

Согласно горенаведената постапка се прави екстракција на овој параметар за 10те репрезентативни класи на звучни настани (слика 3.22). На апцисата е прикажано времето, а на ординатата се прикажани 40те коефициенти согласно поделбата на МЕЛ скалата. Боите ја означуваат моќноста, која се движи во граници од 150 до - 600 dB. Високите вредности на моќноста покажуваат дека најголемиот дел од спектралната енергија е концентрирана на ниски фреквенции, а негативните вредности покажуваат дека спектралната енергија е концентрирана на високи фреквенции во звучниот настан.

МЕЛ ФРЕКВЕНТНИ ЦЕПСТРАЛНИ КОЕФИЦИЕНТИ



Слика 3.22. Визуелен приказ на МЕЛ фреквентни цепстрални коефициенти за репрезентативен аудио сигнал за секоја класа на звучен настан

4. ПРИМЕНА НА АЛГОРИТМИ НА МАШИНСКО УЧЕЊЕ

При креирање на системот за препознавање и класификација на звучни настани, потребно е да се конструираат алгоритми засновани на машинско учење. Како алгоритми за препознавање и класификација на звучни настани се користат класични алгоритми на машинско учење преку примена на техники на ненадгледувано и надгледувано учење, како и алгоритми засновани на длабокото учење.

Согласно избраната база на податоци која ќе се користи за експериментални резултати, ќе се разгледува надгледуван пристап за класификација на звучните настани. Во оваа глава се разработени три класични алгоритми на машинско учење: машини со потпирачки вектори, случајни шуми и наивен Баесов алгоритам. Примената на повеќе алгоритми на машинско учење овозможува споредба на резултатите и избор на систем со висока точност при предвидување на податоците.

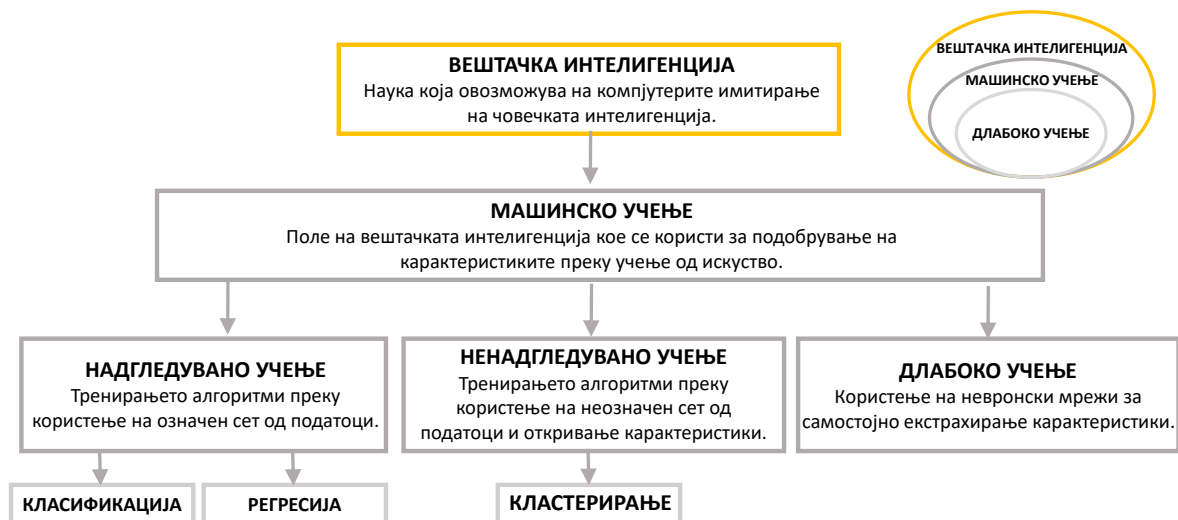
Како влез во алгоритмот се податоците на звучни настани кои се параметризирани со избраните параметри на звучни настани во глава 3. Преку примена на петте параметри на звучни настани и трите алгоритми на машинско учење се формираат 48 модели кои се користат при тестирање на системот за препознавање и класификација на звучните настани и овозможуваат споредба на резултатите со цел избор на параметри и алгоритам на машинско учење кој постигнува највисока точност. Дополнително, преку примена на процесот на хиперпараметарска оптимизација, се овозможува зголемување на перформансот на алгоритмот на машинско учење.

Од применетата методологија може да се забележи дека клучна улога при градење на успешен систем за препознавање и класификација на звучни настани има изборот на параметри на звучни настани, како и избор на алгоритми на машинско учење и нивна хиперпараметарска оптимизација.

4.1. ОСНОВИ НА МАШИНСКОТО УЧЕЊЕ

Вештачка интелигенција како составен дел од системите за препознавање и класификација на звучните настани е наука која овозможува компјутерите да ја имитираат човечката интелигенција и самите да донесуваат одлуки, да обработуваат текст и да даваат визуелна перцепција за нештата без човечка интервенција. Вештачката интелигенција, машинското учење и длабокото учење се термини што често се споменуваат кога се дискутира за големи податоци, аналитика и напредна технологија. Вештачката интелигенција е широко множество кое содржи неколку подмножества, како машинско учење, роботика и компјутерска визуелизација.

Како што е прикажано на слика 4.1, машинското учење е подмножество на вештачката интелигенција кое овозможува подобрување на машините преку учење од искуство со нивна континуирана работа и користење на расположливи податоци. Сите техники за машинското учење се класифицирани како техники на вештачка интелигенција, но не сите техники на вештачката интелигенција се засновани на правилата на машинско учење, бидејќи во основните правила на вештачката интелигенција, не сите системи учат од тренирани податоци.



Слика 4.1. Визуелен приказ на вештачката интелигенција, машинското учење и длабокото учење како сет од множество

Излезот од модел на машинско учење е резултат од тренирањето на алгоритмот со веќе постоечки податоци. Кога моделот е обезбеден со податоци за тренирање, како излез се добива предвидување засновано на овие податоци. Валидација/тестирање на системот се прави со нови податоци кои служат за проверка на излезниот трениран модел на машинско учење. Одредени модели за машинско учење се „онлајн“ и континуирано се прилагодуваат при учење од нови податоци. „Онлајн“ моделите постојано ги подобруваат нивните карактеристики и можат да се користат за апликации во реално време каде што системот

постојано учи и ги подобрува карактеристиките од новите генерирани податоци. Од друга страна, другите модели наречени „офлајн“ модели на машинско учење се добиени од алгоритми за машинско учење со одредени податоци и нивниот алгоритам не се менува.

При тренирање на алгоритмот на машинско учење, потребно е да се користат податоци чија содржина е соодветна за апликацијата, со цел да се овозможи подобрување на точноста на моделите за машинско учење. За постигнување поголема прецизност и точност на моделите од машинско учење, тренирањето на моделите треба да биде реализирано користејќи голема база на податоци. Големата база на податоци претставува извор на податоци со различна содржина, која треба да ги има следните карактеристики: исклучително голем обем на податоци, способност на обработка со голема брзина и веродостојност.

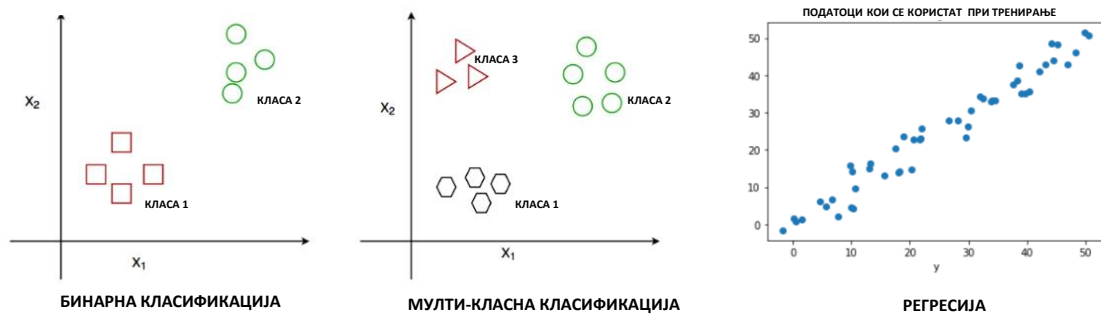
Посебна област на машинското учење која користи невронски мрежи за учење на системот е длабокото учење. Невронските мрежи се модели за обработка на информации инспирирани од човечкиот мозок, кој се состои од милјарди неврони кои комуницираат едни со други користејќи електрични и хемиски сигнали, овозможувајќи им на луѓето да учат и да носат одлуки. Принципот на невронски мрежи функционира на тој начин што користи математички функции кои го имитираат човековиот мозок и поврзуваат вештачки неврони во повеќе слоеви.

За разлика од машинското учење, при примена на длабокото учење како техника која се заснова на вештачката интелигенција, процесот на параметризација на звучни настани и класификација е повратен и повторлив. Машинското учење користи алгоритми за да ги обработи податоците, да научи од нив и да прави одлуки базирани на тоа што го има научено. Длабокото учење ги структурира алгоритмите во нивоа за да создаде “вештачка невронска мрежа” која е способна да учи и сама да прави интелигентни одлуки. Разликата помеѓу длабокото учење и техниките за машинско учење е во нивната способност автоматски да екстрахираат карактеристики.

Зависно од техниката која се применува за машинско учење, постои надгледуван и ненадгледуван процес при машинско учење. При надгледувано учење, процесот започнува со постоечки сет од податоци кои се соодветно поделени во класи и имаат одредени карактеристики кои даваат дополнително објаснување на класите. Надгледуваното учење има за цел да креира модели од веќе постоечките познати податоци кои може да се применат во аналитичкиот процес. Овие информации се зададени при набљудување од страна на човекот и се влезен параметар при употреба на надгледувана техника за машинско учење. Се користи алгоритам во форма на функција $y = F(x)$, кој задава корелација помеѓу влезната променлива (x) и функцијата (F) на излезната променлива (y). Преку примената на надгледуваното учење, целта е прецизно да се приближиме на функцијата (F) на начин што секогаш при влез на нови податоци (x), може да се дефинира излезната променлива (y) која е предвидена од веќе внесената базата на податоци.

При примена на процесот на надгледувано машинско учење, постојат два типа на учење, класификација и регресија, кои се применуваат зависно од податоците и нивната примена. Класификацијата и регресијата имаат ист концепт на користење на познат збир на податоци за предвидување. Излезот при регресија е нумеричка континуирана променлива,

додека при класификацијата излезот е определување на некоја класа (слика 4.2.). Во зависност од бројот на класи, разликуваме бинарна класификација (разлика помеѓу две класи) и мулти-класна класификација (разлика помеѓу n број на класи).



Слика 4.2. Класификација и регресија

Во машинското учење, алгоритмите за класификација се обидуваат да ја проценат функцијата (F) од влезните променливи (x) во дискретни излезни променливи (y). Добиената функцијата може да се демонстрира во форма на правила „АКО-ТОГАШ“.

Регресија е процес на наоѓање на модел или функција за разликување на податоците во континуирани реални бројни вредности. Од друга страна, алгоритмите на регресија се обидуваат да ја проценат функцијата (F) од влезните променливи (x) во нумеричките или континуираните излезни променливи (y). Во случајот на регресија, (y) е вистинска бројчана вредност.

Ненадгледуваното учење е техника на машинско учење во која корисниците немаат потреба да го надгледуваат моделот, односно базата на податоци нема означени карактеристики. Техниката на ненадгледувано учење му овозможува на моделот да работи самостојно за да открие информации кои претходно не биле откриени. Ненадгледуваното учење се користи при употреба на големо количество необележани податоци. Алгоритмите за ненадгледуваното учење ги сегментираат податоците во групи кои се наречени кластери или групи на кластери. Неозначените податоци ги создаваат вредностите на параметрите и класификацијата на податоците. Во суштина, овој процес овозможува дефинирање на карактеристики на податоците, а на тој начин тие стануваат надгледувани. Ненадгледуваното учење може да се користи само кога постои голема база на податоци. Постои хибриден процес на учење (полунадгледувано учење) кој користи мала количина на означени податоци и голема количина на неозначени податоци за време на тренирањето.

За проценка на перформансите на креираниот модел, потребно е да се направи тестирање над независно множество податоци. При примена на големи податоци за конструирање на системот за препознавање и класификација на звучните настани, вообичаено 90% од податоците се користат за градење на акустичниот модел за учење на системот, додека

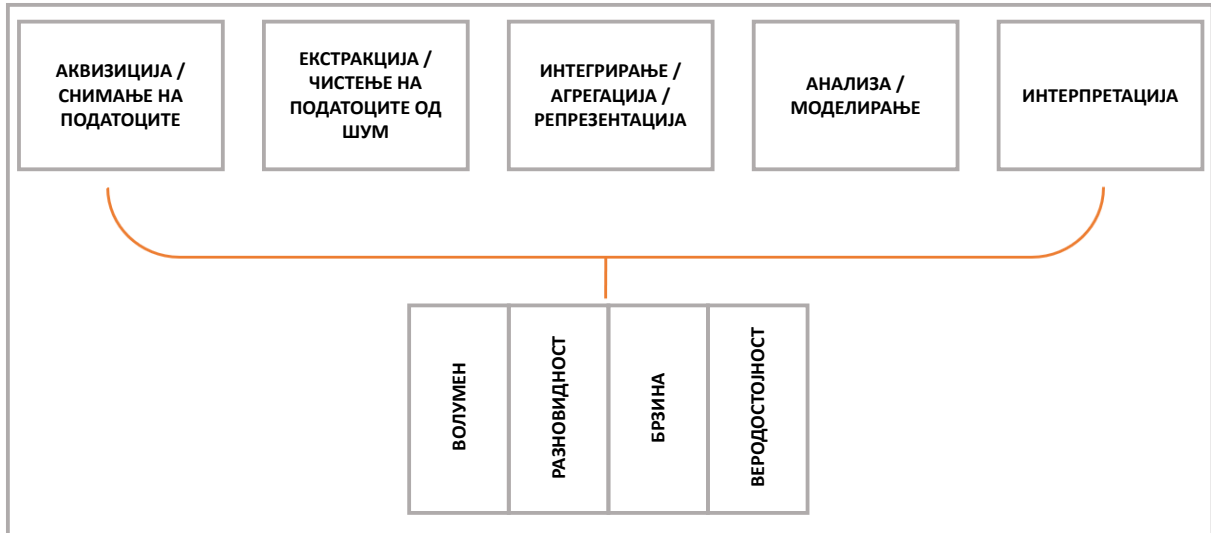
10% се користат за тестирање и валидација на системот, од каде што се извлекува информација за точноста и успешноста на системот.

Согласно базата на податоци која ќе се користи за ова истражување, бидејќи класите на звучните настани се веќе означени, ќе се користи надгледуван пристап на учење при класификација.

4.1.1. ГОЛЕМИ ПОДАТОЦИ

Големите податоци претставуваат множество од податоци кои се комплексни за складирање, управување и процесирање, чија цел е разбирање на нивното значење за зголемување на точноста при донесување на одлуки во процесот на машинско учење. Во современото општество, забрзаниот раст на информации е причина за создавање на нови технологии за анализа на големи количества на податоци. Развојот на информатичките технологии се соочува со големи предизвици во процесот на обработка на големи податоци поврзани со нивната ефективна и ефикасна обработка. Секојдневно се креираат околу 2.5 трилиони бајти на податоци, од кои 90% од податоците во светот се креирани само во последната деценија кои треба да се обработат за да се подобри функционирањето на општеството [121]. За таа цел, се користат разни алгоритми со цел анализа на множествата на податоци и извлекување корисни релации и информации. Анализата на податоците се користи за да се извлечат претходно непознати и скриени релации и информации од големи множества на податоци. При анализа на податоците, зависно од нивната природа, може да се користат различни модели на машинско учење.

За анализа на важните карактеристики на големите податоци, потребно е да се помине низ неколку фази: снимање на податоците, екстракција, агрегација, моделирање и анализирање и нивна интерпретација [122]. Овие фази водат кон дефинирањето на важни карактеристики на големите податоци: волумен, разновидност, брзина и веродостојност. Најважната карактеристика е волуменот, односно големината на податоците кои постојано се генерираат и дополнуваат во базата на податоци. Големите податоци имаат покомплексна природа при нивна обработка, управување и анализа. Со зголемување на податоците се појавува проблемот на т.н. слепа зона, односно неможност за процесирање на сите податоци што придонесува до намалување на точноста при анализа и предвидување. Појавата на разновидни податоци доведува до можност за погрешна анализа на податоците при процесот на донесување на одлуки. Брзината на податоците се однесува на брзината за генерирање, складирање и анализирање на податоците од различни извори. Со зголемување на големината на податоци, потребно е истите да се обработуваат во реално време со цел нивна целосна обработка. Веродостојноста го опишува квалитетот, односно точноста на податоците. Алатките кои се користат за управување на веродостојноста на големите податоци во најголем дел го отфрлаат шумот (непотребните податоци) и ги трансформираат истите во веродостојни.



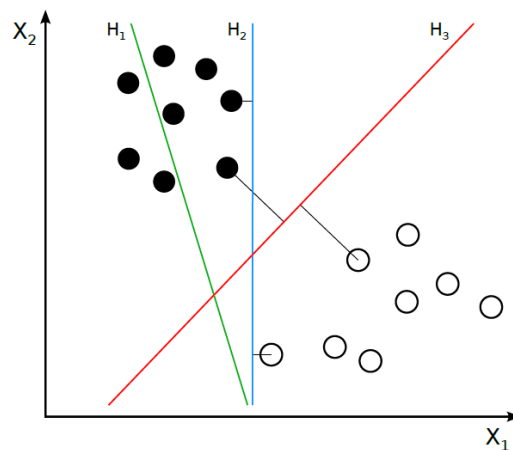
Слика 4.3. Систем за анализа на големи податоци

При обработка на податоците, тие се филтрираат и се компресираат преку креирање на вектор на карактеристики, а останатиот дел од податоците се отфрла, водејќи кон креирање база од податоци со помала големина која е полесна за обработка. Големите податоци се анализираат на платформа која треба да има способност за нивно управување, зачувување и анализа. Друга основна цел на платформата за анализа на големи податоци е можноста за намалување на времето на аналитичкиот циклус, односно времето што е потребно за откривање и трансформирање на податоци, креирање и развој на модели и анализа на добиените резултати. Секоја платформа за обработка и анализа на големи податоци е потребно да обезбеди визуелизација на податоците и да користи алгоритми за машинско учење при процесот на нивно препознавање и класификација [123].

4.2. АЛГОРИТАМ НА МАШИНИ СО ПОТПИРАЧКИ ВЕКТОРИ

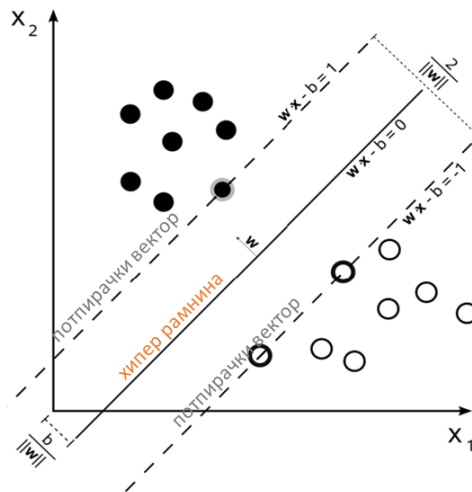
Машините со потпирачки вектори (Support vector Machines) се алгоритми на машинско учење кои се користат при надгледувано учење. Алгоритмот што ги одделува класите на податоци е развиен од истражувачите Vladimir N. Vapnik и Cortes C. во [124], преку нивно одделување во просторот користејќи математички функции од променливи кои служат за поедноставно градење и намалување на комплексноста на алгоритмот.

При бинарна класификација со две класи на податоци, на даден сет од податоци за тренирање, алгоритмот на машини со потпирачки вектори ги претставува податоците како точки во просторот, кои се означени во одделните класи во кои тие припаѓаат. Класите се одделени со хипер рамнина, што во овој случај е права која подеднакво ги одделува двата најблиски податоци на одделните класи. На слика 4.4. се прикажани неколку хипер рамнини (H_1 , H_2 , H_3). Логичен избор за најдобра хипер рамнина е линијата која прави најголемо одделување (маргина) помеѓу двете класи. Согласно сликата 4.4, најголемо растојание и најдобро одделување помеѓу двете класи на податоци прави хипер рамнината H_3 , каде што растојанието од рамнината до најблиската точка на податоци од секоја страна е максимално.



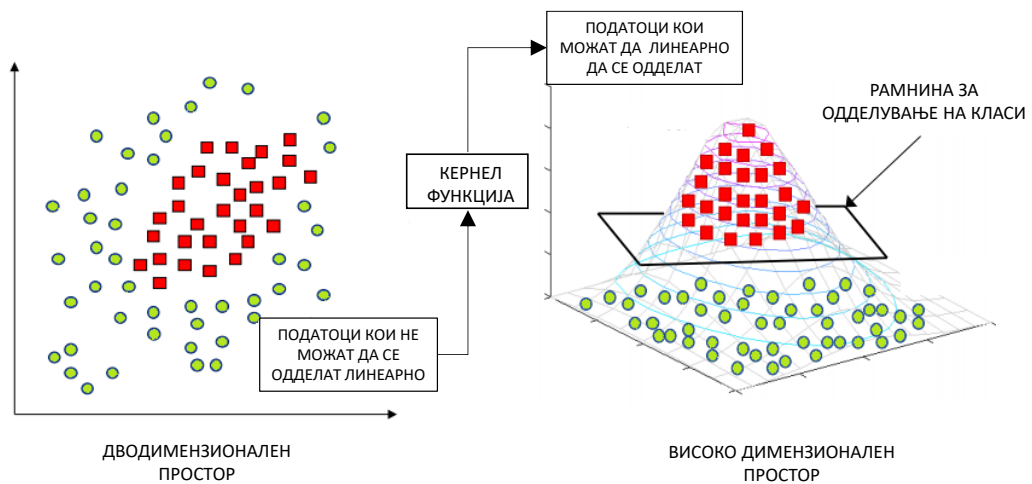
Слика 4.4. Хипер рамнина на линеарен класификатор за машини со потпирачки вектори (хипер рамнината H_1 не ги одделува класите; H_2 ги одделува, но со мала маргина; H_3 ги одделува со максимална маргина)

На слика 4.5. е прикажана хипер рамнина со максимална маргина на бинарен класификатор, а правите кои се паралелни на хипер рамнината се потпирачките вектори кои се подеднакво поделени од хипер рамнината. Меѓусебно растојание помеѓу главната хипер рамнина и потпирачкиот вектор е многу важно за определување на точноста при одделување на податоците по класи.



Слика 4.5. Хипер рамнина со максимална маргина помеѓу две класи. Примероците кои лежат на крајот на маргината (двете помошни линии) се наречени потпирачки вектори

Главниот проблем кој се јавува при употреба на алгоритмот на машини со потпирачки вектори е појава на множества од податоци кои се нелинеарни, односно не можат да се одделат линеарно. Од тие причини, имплицитно се мапираат влезовите во високо-димензионалните простори на карактеристиките на податоците и се формира хипер рамнина или збир на хипер рамнини во висок или бесконечен димензионален простор што може да се користи за класификација или регресија на податоците. Доброто одделување се постигнува при најголемо растојание на хипер рамнината од најблиската точка (податок) од било која класа, формирајќи функционална маргина. Колку е поголемо ова растојание, толку е помала грешката при класификација. Оригиналниот димензионален простор се преработува во повисок димензионален простор, зголемувајќи го растојанието помеѓу различните класи на кои припаѓаат податоците со користење на соодветна математичка кернел функција (слика 4.6) [125].



Слика 4.6. Одделување на податоци со користење кернел функција

Ефективноста на алгоритмот зависи од изборот на кернел функцијата која зависи од видот на податоци и нивната распределеност во просторот [126]. Базата на податоци се претставува со p димензионален вектор, а класите се одделуваат со $(p - 1)$ димензионална хипер рамнина. При постоечки сет од податоци D , се формира сет од одделните податоци n прикажан со изразот:

$$D = \{(x_i, y_i) | x_i \in R^p, y_i \in \{-1, 1\}\}_{i=1}^n \quad (4-1)$$

За да се најде хипер рамнината помеѓу две класи на податоци со максимална маргина мора се одделат податоците на класите на кои тие припаѓаат со променлива y_i , која зависно од видот на класата има вредност $y_i = 1$ или $y_i = -1$. На овој начин се прикажува класата во која податокот x_i припаѓа, а секој x_i податок е p димензионален реален вектор. Бидејќи целта е да се најде хипер рамнина која ги одделува класите на податоците со максимална маргина, било која хипер рамнина се запишува како сет од податоци x кои го задоволуваат условот прикажан со изразот:

$$w \cdot x - b = 0 \quad (4-2)$$

Од производот единичниот податок x со нормалниот вектор на хипер рамнината w се одзема растојанието помеѓу двата потпирачки вектори b и се изедначува на нула. Како што може да се забележи на слика 4.6, поместувањето на хипер рамнината од координатиот почеток долж нормалниот вектор w е прикажано со параметарот $\frac{b}{\|w\|}$, каде $\|w\|$ е скаларна големината на нормалниот вектор на хипер рамнината.

Доколку тренираните податоци се линеарно одделени, може да се изберат две помошни хипер рамнини (потпирачки вектори) така што ќе ги одделат податоците и нема да има мешање на класите, а притоа ќе се постигне најголемо растојание. Растојанието помеѓу овие помошни потпирачки вектори се нарекува маргина. Помошните потпирачки вектори може да се опишат со изразите:

$$w \cdot x - b = 1 \quad (4-3)$$

$$w \cdot x - b = -1 \quad (4-4)$$

Со користење на геометриска зависност, може да се најде растојанието помеѓу двата потпирачки вектори, односно се претпоставува дека $b = 2$, па од тука се добива растојание $\frac{2}{\|w\|}$. Потребно е да се сведе до минимум големината на скаларот $\|w\|$. Се отфрлат податоци кои се наоѓаат во маргината и се додава ограничување за секој i -ти податок во согласност од y_i дали е позитивна (израз 4-5) или негативна вредност (израз 4-6):

$$w \cdot x_i - b \geq 1 \text{ за } x_i \quad (4-5)$$

$$w \cdot x_i - b \leq -1 \text{ за } x_i \quad (4-6)$$

Изразите (4-5) и (4-6) може да се запишат како:

$$y_i(w \cdot x_i - b) \geq 1 \text{ за сите податоци } 1 \leq i \leq n \quad (4-7)$$

Следно, потребно е да се направи оптимизација, односно да се најде решение кое ќе постигне најефективен перформанс преку зголемување на точноста на предвидување, а сведување на минимум несакани ефекти.

Проблемот со оптимизација е комплексен бидејќи зависи од големината на нормалниот вектор $\|w\|$, чија цел е да се сведе на минимум. Постои можност да се замени големината на нормалниот вектор $\|w\|$ со изразот $\frac{1}{2}\|w\|^2$ (Факторот $\frac{1}{2}$ се користи за математичка сигурност) без да се промени решението.

Користејќи ја зависноста дадена во изразот (4-7), за секој податок $i = 1, \dots, n$ се бара минимум на големината на нормалниот вектор:

$$\arg \min_{(w,b)} \frac{1}{2} \|w\|^2 \quad (4-8)$$

За да се најде екстремот при зададените услови, потребно е да се воведат Лагранжови множители α , а проблемот со ограничување може да се изрази како:

$$\arg \min_{(w,b)} \max_{\alpha \geq 0} \left\{ \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i [y_i(w \cdot x_i - b) - 1] \right\} \quad (4-9)$$

Бидејќи Лагранжовиот множител α_i се изедначува со нула, преку примена на „стационарната“ состојба на Каруш – Кун – Такер решението може да се изрази како линеарна комбинација на тренирани вектори. Со диференцирање на изразот (4-9) по w и b , се добиваат изразите:

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (4-10)$$

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad (4-11)$$

Само неколку Лагранжови множители α_i ќе бидат поголеми од нула. Во изразот (4-10), коефициентот x_i претставува потпирачки вектор кој лежи на маргината и го исполнува условот:

$$y_i(w \cdot x_i - b) = 1 \quad (4-12)$$

Од тука, може да се потврди дека потпирачките вектори го задоволуваат и изразот:

$$w \cdot x_i - b = \frac{1}{y_i} \quad \leftrightarrow \quad b = w \cdot x_i - \frac{1}{y_i} \quad (4-13)$$

Изразот (4-13) овозможува дефинирање на вредноста b која зависи од x_i и y_i и варира за секој поединечен податок. Со воведување средна вредност на потпирачките вектори N_{SV} , вредноста b може да се пресмета преку изразот:

$$b = \frac{1}{N_{SV}} \sum_{i=1}^{N_{SV}} (w \cdot x_i - y_i) \quad (4-14)$$

Со замена на изразите (4-10) и (4-11) во изразот (4-9), се добива:

$$\arg \min_{(w,b)} \max_{\alpha \geq 0} \left\{ \frac{1}{2} \sum_{i=1}^n \alpha_i y_i x_i \sum_{j=1}^n \alpha_j y_j x_j - \sum_{i=1}^n \alpha_i y_i x_i \left(\sum_{j=1}^n \alpha_j y_j x_j \right) - \sum_{i=1}^n \alpha_i y_i b + \sum_{i=1}^n \alpha_i \right\} \quad (4-15)$$

$$\arg \min_{(w,b)} \max_{\alpha \geq 0} \left\{ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j \right\} \quad (4-16)$$

Од тука, може да се заклучи дека главната оптимизација може да се случи со менување на влезните податоци $x_i x_j$, што се постигнува со различни видови на кернел функција $f(x) = k(x_i, x_j)$. Алгоритмите на машините со потпирачки вектори користат сет од математички кернел функции чија функција е трансформација на податоците на влез во алгоритмот во посакуван облик кој ќе постигне најдобро одделување. Постојат повеќе видови на кернел функции кои се користат зависно од видот на податоци и класификација прикажани на табела 4.1.

Табела 4.1. Вид на кернел функции

Функција	Израз
Хомогена полиномна кернел функција	$k(x_i, x_j) = (x_i \cdot x_j)^d$
Нехомогена полиномна кернел функција	$k(x_i, x_j) = (x_i \cdot x_j + 1)^d$
Гаусова радијална функција	$k(x_i, x_j) = \exp(-\gamma \ x_i - x_j\ ^2)$
Хиперболична тангента	$k(x_i, x_j) = \tanh(kx_i \cdot x_j + c)$
Радијална базна функција	$k(x_i, x_j) = \exp(-\ x_i - x_j\ ^2 / 2\sigma^2)$

Покрај изведување на бинарна класификација помеѓу две класи на податоци, алгоритмот на машини со потпирачки вектори има способност за ефикасно спроведување на мултикласна класификација. При постоење на повеќе класи на податоци се користи истиот принцип на бинарна класификација каде што доаѓа до повторување на постапката на самиот алгоритам се додека не се одделат сите класи на податоци.

Алгоритмот на машини со потпирачки вектори постигнува висока точност при моделирање на системите за препознавање и класификација на звучните настани поради јасната маргина која ја постигнува при делење на класите на звучни настани. Базата на податоци која се користи за класификација има многу шум во неа и голем број на карактеристики во звучните настани поради самата природа на класите. Согласно истражувањата во [127], алгоритмите на машини со потпирачки вектори покажуваат висока ефикасност и имаат голема толеранција при работа со бази на податоци со голем шум, како што е базата на податоци која ќе се користи во експерименталниот дел.

4.3. НАИВЕН БАЕСОВ АЛГОРИТАМ

Во машинското учење, Наивните Баесови класификатори (Naïve Bayes classifiers) се семејство на едноставни класификатори на веројатност кои ја користат Баесовата теорема со силни наивни претпоставки помеѓу карактеристиките на различните класи на податоци [127, 128]. Баесовата теорема на веројатност претставува начин за пресметување на веројатноста на настан врз основа на бројот на тоа колку пати во минатото тој настан се случил. Со едноставни математички функции се претпоставува дека присуството на одредена карактеристика во една класа не е поврзана со карактеристика од друга класа. Овој модел е лесен за градење и е многу користен за големи множества на податоци, а за негова изградба се користат променливи карактеристики. Наместо да се користи повторувачка апроксимација која трае долго време и е многу комплексна за анализа, максималната веројатност може да се постигне со проценка во затворена линеарна форма во дефинирано време [129]. Се прави хистограм од карактеристиките кои се повторуваат во одреден податок, а од хистограмот се добива веројатноста со која една карактеристика се повторува во самиот податок. За секоја карактеристика се повторува процесот на пресметка на повторливост од податокот за секоја класа и се определува класата со највисоката веројатност за повторливост на карактеристиката. Зборот наивен е додаден поради тоа што за овој алгоритам не е важен редоследот на случување на настаните (карактеристиките), туку важно е само карактеристиката да постои во самата класа на податок.

Наивниот Баесов алгоритам се формира со примена на едноставна техника за конструирање на класификатор чии податоци на секоја класа се означени како вектори на карактеристики, а карактеристиките се извлекуваат од ограничено множество на податоци [130]. Овој класификатор се заснова на претпоставката дека вредноста на одредена карактеристика е независна од вредноста на било која карактеристика, без оглед на променливата. На пример, овошје може да се смета за јаболко ако е црвено, кружно и има дијаметар од околу 3 см. Наивниот Баесов класификатор смета дека секоја од овие карактеристики придонесува независност на веројатноста дека ова овошје е јаболко, без оглед корелациите кои може да се јават помеѓу карактеристиките на бојата, заобленоста и дијаметарот.

Еден податок x кој припаѓа во одредена класа е претставен како вектор со n број на независни карактеристики преку изразот:

$$x = (x_1, \dots, x_n) \quad (4-17)$$

За секоја C_k класа, веројатноста на карактеристиката p е прикажана со изразот:

$$p(C_k|x) = (C_k|x_1, \dots, x_n) \quad (4-18)$$

Доколку бројот на карактеристики е голем, се појавува комплексност при изведба на самиот модел, па затоа, изразот се преформулира користејќи ја Баесовата теорема за условна веројатност дадена со изразот:

$$p(C_k|x) = \frac{p(C_k)p(x|C_k)}{p(x)} \quad (4-19)$$

Важноста на Баесовата формула е во тоа што таа овозможува пресметка на веројатност $p(C_k|x)$ на класа C_k застапена во податок со карактеристики x , $p(C_k)$ е претходна претпоставка, $p(x|C_k)$ е веројатност на карактеристиките на податок x да е застапен во класа C_k , а $p(x)$ е доказ.

При одлуката влијае броителот, бидејќи именителот не зависи од карактеристиката на независност и тој е постојано константен. Броителот е еквивалентен на моделот на заедничката веројатност прикажан со изразот:

$$p = (C_k, x_1, \dots, x_n) \quad (4-20)$$

Изразот (4-20) може да биде запишан и во следната форма користејќи го правилото за веројатност:

$$\begin{aligned} p(C_k, x_1, \dots, x_n) &= p(C_k)p(x_1, \dots, x_n|C_k) \\ &= p(C_k)p(x_1|C_k)p(x_2, \dots, x_n|C_k, x_1) \\ &= p(C_k)p(x_1|C_k)p(x_2|C_k, x_1)p(x_n|C_k, x_1, x_2 \dots x_{n-1}) \end{aligned} \quad (4-21)$$

Се започнува со користење на “наивни” претпоставки за условна независност, односно се претпоставува дека секоја карактеристика x_i е условно независна од секоја друга карактеристика x_j за $j \neq i$, за дадена класа C_k , што доведува до формирање на изразите:

$$p(x_i|C_k, x_j) = p(x_i|C_k) \quad \text{за } i \neq j \quad (4-22)$$

$$p(x_i|C_k, x_j, x_k) = p(x_i|C_k) \quad \text{за } i \neq j, k \quad (4-23)$$

$$p(x_i|C_k, x_j, x_k, x_l) = p(x_i|C_k) \quad \text{за } i \neq j, k, l \quad (4-24)$$

Од изразите (4-22), (4-23) и (4-24) се добива заеднички модел прикажан со изразот:

$$\begin{aligned} p(C_k|x_1, \dots, x_n) &\propto p(C_k, x_1, \dots, x_n) \\ &\propto p(C_k)p(x_1|C_k)p(x_2|C_k)p(x_3|C_k) \dots p(x_n|C_k) \\ &\propto p(C_k) \prod_{i=1}^n p(x_i|C_k) \end{aligned} \quad (4-25)$$

Според горенаведените претпоставки за независност, условната распределба над класата C_k е:

$$p(C_k|x_1, \dots, x_n) = \frac{1}{Z} p(C_k) \prod_{i=1}^n p(x_i|C_k) \quad (4-26)$$

Со $Z = p(x)$ е прикажан доказот кој зависи само од променливите карактеристики x_1, \dots, x_n , кој претставува константа доколку вредностите на векторот на карактеристики се познати.

Наивниот Баесов класификатор го комбинира овој прикажан модел на веројатност во изразот (4-26) со правило за одлука. За C_k класи со k број на класи, Наивниот Баесов класификатор е функција \hat{y} прикажана со изразот:

$$\hat{y} = \operatorname{argmax}_{k \in \{1, \dots, K\}} p(C_k) \prod_{i=1}^n p(x_i|C_k) \quad (4-27)$$

Предноста при одлучување ја има класата која има поголема веројатност и истата може да се пресмета преку проценување на веројатноста за постоење на карактеристиката во одредената класа, т.е. веројатноста на одредена класа е односот помеѓу бројот на карактеристики во класата и вкупниот број на карактеристики во сите класи. За да се проценат параметрите на распределба на одредена карактеристика мора да се претпостави распределбата или да се генерираат модели на карактеристиките од тренираните податоци. Од овие причини, Наивниот Баесов класификатор може да се кориси само при надгледуван пристап на учење, односно мора да бидат познати класите и карактеристиките на податоците. Претпоставките за распределба на карактеристиките се нарекуваат модели на настани на

Наивниот Баесов класификатор и постојат повеќе модели меѓу кои Гаусов, мултиномен и Бернулиев наивен Баесов класификатор.

4.3.1. ГАУСОВ НАИВЕН БАЕСОВ КЛАСИФИКАТОР

Моделот на Гаусов наивен Баесов класификатор се користи кога се работи со континуирани податоци, каде што се претпоставува дека континуираните вредности за секоја класа се распределуваат според Гаусовата распределба.

Доколку μ_c е средната вредност на карактеристиките x во класата C_k , а σ_c^2 е веројатноста на карактеристиките x кои припаѓаат на класата C_k , тогаш веројатноста на дадената класа $p(x = v|C_k)$ може да биде пресметана преку додавање на нормална распределба v добиена од средната вредност и веројатноста на карактеристиките во класата (μ_c и σ_c^2), од каде што се добива изразот:

$$p(x = v|c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{-\frac{(v-\mu_c)^2}{2\sigma_c^2}} \quad (4-28)$$

4.3.2. МУЛТИНОМЕН НАИВЕН БАЕСОВ КЛАСИФИКАТОР

Со мултиномниот модел на Наивниот Баесов класификатор, податоците во форма на вектори на карактеристики ја претставуваат фреквенцијата со која единечни настани се повторуваат а се генерирани од веројатност (p_1, \dots, p_n) , каде p_i е веројатноста со која настанот i се повторува. Векторот на карактеристики за одреден податок $x = (x_1, \dots, x_n)$ е претставен како хистограм, а x_i ги брои настаните кои се случиле во одредено време. Мултиномниот наивен Баесов класификатор се користи при повеќекласна класификација, а веројатноста е дадена со изразот:

$$p(x|C_k) = \frac{(\sum_i x_i)!}{\prod_i x_i!} \prod_i p_{ki}^{x_i} \quad (4-29)$$

Мултиномниот наивен Баесов класификатор станува линеарен при примена на логаритам:

$$\log p(C_k|p) = \log \left(p(C_k) \prod_i p_{ki}^{x_i} \right) = \log p(C_k) + \sum_{i=1}^n x_i \cdot p_{ki}^{x_i} \quad (4-30)$$

Доколку при тренирањето на податоците за дадена класа не се појави некоја карактеристика, тогаш проценката на веројатноста која е заснована на повторливоста на настаните ќе биде нула. Оваа појава може да пројави проблем поради тоа што ги брише сите информации и затоа потребно е да се вклучи корекција на мал примерок преку додавање на

коефициент α кој вообичаено има вредноста 1, и на тој начин никогаш нема да се појави вредноста нула. Овој начин на регулација се нарекува Лапласово филтрирање.

4.3.3. БЕРНУЛИЕВ НАИВЕН БАЕСОВ КЛАСИФИКАТОР

При примена на моделот на Бернулиев наивен Баесов класификатор, карактеристиките на влез се независни бинарни променливи. Ако x_i е независна бинарна променлива кој го опишува постоењето или недостатокот на карактеристиката на податокот во класа C_k , се добива изразот:

$$p(x|C_k) = \prod_{i=1}^n p_{ki}^{x_i} (1 - p_{ki})^{(1-x_i)} \quad (4-31)$$

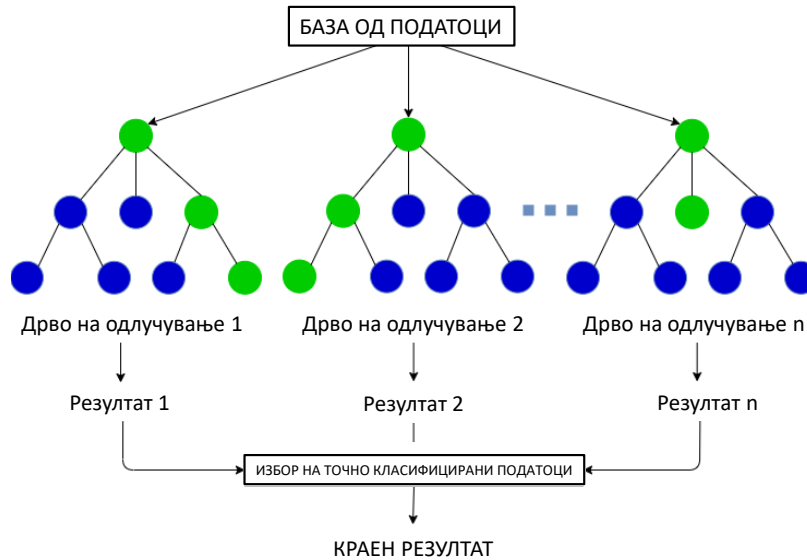
Наивниот Баесов класификатор се користи при класификација на настани во повеќе класи поради правилото за независност на секоја карактеристика. Неговата примена се зголемува поради малата комплексност која води до брза пресметка и лесното моделирање на овој алгоритам. Поради тоа што се смета дека секоја класа на податок од избраната база на податоци има независни карактеристики, се смета дека примената на овој алгоритам во системот за класификација и препознавање на звучни настани ќе постигне висока точност при класификација.

4.4. АЛГОРИТАМ НА СЛУЧАЈНИ ШУМИ

Алгоритмот на случајни шуми (Random Forest Algorithm) е модел на машинско учење претставен со група на класификациони или регресиони дрва на одлуки креирани од случајна селекција на примероци од тренирачкото множество. Алгоритмот е претставен од Breiman L. во [131,132], а е составен од голем број на дрва на одлуки што создаваат случаен примерок од повеќе дрва на одлука и заедно ги анализираат резултатите со цел да се добие постабилна и точна предикција. Алгоритмот на случајни шуми прави комбинација од едноставната интерпретација на алгоритмот на дрва на одлучување и голема флексибилност која резултира во подобрување на точноста на моделот.

При формирање на дрвата, моделот на случајни шуми не ги користи сите независни променливи, туку само оние кои што се случајно избрани. Случајните параметри се селектирани во фаза на индукција, додека пак претпоставките се креирани по пат на агрегација на самите претпоставки [133]. За разлика од алгоритмот на дрва на одлуки, алгоритмот на случајни шуми паралелно учи и комбинира меѓусебно независни дрва, а притоа, секое наредно обучено дрво ги означува слабостите на претходното. Користејќи широк спектар на

различни дрва со различни одлуки, се добива разновидност на резултатите што доведува до постигнување на многу поголема ефективност за разлика од користење на индивидуални дрва на одлуки. На слика 4.7. визуелно е прикажан моделот кој се гради при користење на алгоритмот на случајни шуми за класификација на податоци.



Слика 4.7. Приказ на алгоритам на случајни шуми

Класификациониот алгоритам на случајни шуми со едноставна структура се базира на следните принципи: сетот за тренирање е случајно генериран од оригиналниот сет на податоци; и на секој јазол, податоците и карактеристиката според која се прави поделбата се случајно избрани, а длабочината на податоците ја избираат максималната длабочина на разделување на податоците.

Даден сет од n број на податоци е прикажан со:

$$D = \{(x_1, y_1), \dots, (x_i, y_i)\} \quad (4-32)$$

Секој податок од базата е означен со x_i чиј излез е класата во која тој припаѓа y_i е прикажан со изразот:

$$\{x_i\}_{i=1}^n \quad (4-33)$$

Секој податок е претставен преку вектор на карактеристики со изразот:

$$x_{ki} = (x_{k1}, \dots, x_{kd}) \quad (4-34)$$

Дрвото на одлуки за класификација како дел од алгоритмот на случајни шуми во секој негов јазол има бинарна одлука во зависност од параметарот според кој се прави одлуката ($x_i > a$). Во горниот јазол се содржат сите карактеристики (x_k, y_k) , додека во понатамошните јазли при поделбата се поделени сетови од карактеристики. Во секој јазол од дрвото на

одлуки, карактеристиката x_i и параметарот a се одбрани преку критериумот на Гини (Gini criterion), кој го дели секој јазол се додека не дојде до избор на една класа. Критериумот на Гини е даден со изразот:

$$g(x_i) = \sum_{i=1}^n (\hat{P}(C_k|x_i)(1 - \hat{P}(C_k|x_i))) \quad (4-35)$$

Во изразот (4-35), веројатноста на примерокот x_i да припаѓа на класата C_k е дадена со изразот:

$$\hat{P}(C_k|x_i) = \frac{|x_i \cup C_k|}{|x_i|} \quad (4-36)$$

Потребно е да се генерира вектор од случајни карактеристики кој е добиен од податокот во (4-34):

$$X = (X_1, \dots, X_d) \quad (4-37)$$

На овој начин, се формира нова база на податоци D' со случајно одбрани карактеристики, кои се добиени случајно преку веројатност на дистрибуција на карактеристиките $(x_i, y_i) \sim (X, Y)$.

Целта на алгоритмот на случајни шуми е да се класифицира влезниот податок во излезна класа преку користење на базата на податоци D' . За таа цел, се одбираат повеќе дрва на одлуки h_k прикажани со изразот:

$$h = \{h_1(X), \dots, h_k(X)\} \quad (4-38)$$

Доколку се собере секое дрво на одлука $h_k(X)$, нивниот збир формира случајни шуми. За секое дрво на одлука се формираат параметри кои ја вклучуваат структурата на дрвото, поделбата на јазлите, длабочината и останатите важни параметри, дадени со изразот:

$$\lambda_k = (\lambda_{k1}, \dots, \lambda_{kp}) \quad (4-39)$$

Со дадени параметри λ_k , дрвото на одлуки $h_k(X)$ се запишува како:

$$h_k(X) = h(X|\lambda_k) \quad (4-40)$$

Карактеристиките на податокот кои ќе се појават во едно дрво на одлуки се одбрани случајно.

Ако C_k е класата од едно дрво на одлуки $h_k(X)$ од случајните шуми, веројатноста $\hat{P}(C_k)$ е дефинирана како пропорција од резултатите од дрвата на одлуки h_k ($1 \leq k \leq K$) за настанот во кој се појавила класата C_k . Се дефинира функцијата на маргината со изразот:

$$\hat{m}(X, y) = \hat{P}_k(h_k(X) = y) - \max_{j \neq y} \hat{P}_k(h_k(X) = j) \quad (4-41)$$

Општата грешка што се јавува во едно дрво на одлуки е дадена со изразот:

$$e = P_{X,y}(\hat{m}(X,y) < 0) \quad (4-42)$$

Големината на грешката се стреми кон нула со зголемување на бројот на податоците во користената база.

Може да се заклучи дека случајните шуми се алгоритми на класификација кои се потпираат на дрвата на одлуки, но користат параметри со вектор со случајни променливи. Пресметката на точноста при крајната класификација, $f(X)$, зависи од одлуката за класата во која припаѓа одреден податок X согласно секое дрво на одлуки, пракажана преку изразот:

$$f(X) = \sum_{k=1}^K \frac{1}{K} h_k(x) \quad (4-43)$$

Алгоритмот кој се користи за тренирање на податоците ја применува општата техника за агрегација на карактеристиките. По тренирањето, предвидувањата на примероците \hat{f} се добиваат преку просекот од предвидувањата од сите индивидуални дрва на одлуки:

$$\hat{f} = \frac{1}{M} \sum_{m=1}^M \hat{f}_B(h_k) \quad (4-44)$$

По тренирање на алгоритмот со базата на податоци за тренирање, истиот се користи за тестирање со непозната база на податоци, каде што одлуките согласно карактеристиките на класите се веќе познати и моделот на случајни шуми може да се тестира.

При употреба на алгоритмот на случајни шуми во класификациони проблеми се постигнува висока точност поради тоа што се смета дека изборот на случајни променливи доведува до елиминирање на податоците кои имаат шум и не можат да се класифицираат во соодветната класа. Компјутерската комплексност при пресметка на овој алгоритам е мала во однос на алгоритмот на машини со потпирачки вектори. Бидејќи се работи со податоци кои се снимени во надворешна животна средина и не сите јасно ја означуваат својата класа, се смета дека овој алгоритам успешно ќе може да се примени при класификација на звучните настани во оваа докторска дисертација.

4.5. ХИПЕРПАРАМЕТАРСКА ОПТИМИЗАЦИЈА ПРИ НАДГЛЕДУВАН ПРИСТАП НА УЧЕЊЕ

Хиперпараметарска оптимизација (ХПО) е процес на подобрување на ефективноста на алгоритмот на машинско учење преку подесување на параметрите врз чија основа се гради алгоритмот. Изборот на најдобрата конфигурација на вредностите на параметрите кои се користат при моделирање на алгоритмот на машинско учење директно влијае врз перформансите на моделот изграден согласно користената база на податоци. При избор на вредности на параметри потребно е да се има познавање на применетиот алгоритам на машинско учење и на процесот на хиперпараметарска оптимизација и вредностите на параметрите (континуирани или дискретни). Наједноставен начин да се добие модел на машинско учење со висока точност е преку тестирање и споредување на повеќе алгоритми за истиот проблем и база на податоци, од каде што се избира алгоритмот кој што постигнува највисока ефикасност. Дополнително, преку примена на ХПО на секој применет алгоритам на машинско учење се добиваат најоптимални резултати со висока ефикасност.

Постојат неколку техники за автоматска оптимизација кои ја зголемуваат динамичката комплексност со цел да се добие висока точност на предвидување. Овој процес трае долг временски период и потребна е голема моќ за негово процесирање поради големиот број на можни комбинации за тестирање и потребните ресурси за извршување на пресметките. При примена на ХПО за континуирани вредности се применува методот на намалување на градиенти (gradient descent-based method) [134,135] додека за дискретни вредности се применуваат два методи: Баесовата метода и методата на теоретско одлучување [136]. Баесовите алгоритми на ХПО го балансираат процесот на истражување при избор на хиперпараметарски конфигурации и искористување на конфигурацијата на поставките со цел да се добијат подобри резултати. Овие методи можат да резултираат во поголема ефективност на алгоритмот со примена на само неколку итерации. Еден од најважните Баесови методи е оптимизацијата заснована на моделот (Model-based Optimization), во која начинот на конструирање на функцијата за моделирање е преку анализа на распределбата на грешки и обид за нивно намалување. Гаусовиот процес на ХПО е модел кој се заснова на дрвото на одлуки преку пребарување на хиперпараметри во хиперпараметарскиот простор, пресметување на точноста при проценка и избор на вредност на најдобар параметар.

При анализа на ефикасноста на автоматската хиперпараметарска оптимизација при надгледуван пристап на машинско учење, потребно е да се земе во предвид и големината на користената база на податоци за да може да се примени ефикасна оптимизација од поглед на квалитет и динамичка комплексност. Во однос на квалитетот, се зема точноста при предвидување на податоците за проценка на успешноста на изградениот модел на машинско учење. Динамичката комплексност е во директна корелација со големината на базата на податоци која влијае на сложеноста на времето и просторот (количината на меморија и времето потребно за извршување на процесот). Основната идеја е да се лоцираат најважните

карактеристики од базата на податоци кои го одржуваат и/или подобруваат квалитетот на точноста на алгоритмите на машинското учење.

Се претпоставува дека имаме модел на машинско учење A , алгоритам на хиперпараметарска оптимизација H и база на податоци D . Вообичаено, постои само една база на податоци D кој се дели на два дела, еден за тренирање, односно оптимизација на хиперпараметрите на моделот на машинско учење, а другиот за тестирање на моделот со зададените хиперпараметри.

За математичко дефинирање на процесот на хиперпараметарската оптимизација, согласно истражувањето во [136], се дефинира функција на распределување \mathcal{P} . Алгоритмот на машинско учење, означен со A го означува секој податок преку функција на распределување \mathcal{P} од дадена база на податоци за тренирање D^{tr} како функција прикажана со изразот:

$$f_{A,D^{tr}} = A(D^{tr}) \quad (4-45)$$

Функцијата припаѓа на одреден простор од функции и има очекувана загуба L :

$$L(D^{tr}, f_{A,D^{tr}}) \quad (4-46)$$

Простор од функции што го дефинира алгоритмот на машинско учење зависи од одредени параметри кои може да имаат континуирани или дискретни вредности и мора да бидат определени пред да се примени алгоритмот, дадени со изразот:

$$\lambda = (\lambda_1, \dots, \lambda_n) \quad (4-47)$$

Со цел да се направи експлицитна зависност од параметарот λ , алгоритмот се означува како A_λ , а функцијата f_λ ја означува функцијата на податокот која зависи од параметарот λ . Од друга страна, преку примена на друга база на податоци која служи за тестирање на податоците определена од \mathcal{P} , D^t е дадена за да ја валидира загубата $L(D^t, f_{A,D^{tr}})$ при определување на точноста на предвидување преку функцијата која е генерирана од базата на податоци која служи за тренирање на алгоритмот D^{tr} .

Хиперпараметарскиот простор, Λ , е просторот каде што параметрите λ имаат одредена вредност и ја подобруваат точноста при примена на базите на податоци за тренирање и тестирање на алгоритмите на машинско учење. Доколку се направи унија помеѓу базите на тренираните и тестираните податоци, $D = D^{tr} \cup D^t$, во тој случај единствени податоци кои влијаат на загубата $L(D^{tr}, f_{A,D^{tr}})$ се хиперпараметрите $\lambda = (\lambda_1, \dots, \lambda_n)$. Во случај на примена на ХПО при класификација на податоци користејќи алгоритми на машинско учење, може да се заклучи дека загубата L е стапката на грешка при предвидување на податоците, дефинирана со функцијата:

$$\begin{aligned} \Phi_{A,D} : \Lambda &\rightarrow [0,1] \\ \lambda &\rightarrow \text{mean}_{D^t} L(D^t, f_{A,\lambda,D^{tr}}) \end{aligned} \quad (4-48)$$

Хиперпараметарската оптимизација јавува проблем при барање минимум на функцијата $\Phi_{A,D}$, а алгоритмот на ХПО е процес кој пробува да достигне најмала загуба:

$$\lambda^* = \min_{\lambda} (\text{mean}_{D^t} L(D^t, f_{A,\lambda,D^{tr}})) \quad (4-49)$$

Преку примена на хиперпараметарската оптимизација на проблемот дефиниран со функцијата $\Phi_{A,P(D)}$, се бараат оптималните хиперпараметри λ^j што можат да се користат на базата на податоци за тренирање D^{tr} за да се креира алгоритам за класификација кој ќе се тестира на базата на податоци D^t . Од тука, се добиваат вредностите на параметрите за одделните алгоритми со кои се постигнува највисока точност при тестирање на податоците. Важно е да се напомене дека при примена на разни алгоритми на машинско учење постојат различни параметри кои се оптимираат, согласно важноста која ја имаат при конструирање на моделот.

4.6. РЕЗУЛТАТИ ОД ПРИМЕНАТА НА АЛГОРИТМИТЕ НА МАШИНСКО УЧЕЊЕ

За целите на оваа докторска дисертација, моделиран е систем за препознавање и класификација на звучни настани при надгледуван пристап на учење користејќи го софтверот Python. Базата на податоци која се користи за тренирање и тестирање на системот е обележана, односно секој податок е структуриран согласно класата во која тој припаѓа. Податоците кои претставуваат влез во алгоритмот на машинско учење се обработени согласно параметризацијата на звучните настани развиена во глава 3.

Во процесот на примена на машинското учење, 90% од податоците се користат за тренирање на системот, односно се внесуваат како познати класи на звуци, додека 10% се користат за тестирање на системот. При тестирање на ефикасноста на системот, класата на влезните податоци не е позната, а истата е препознаена користејќи го алгоритмот на машинско учење. На слика 4.8. е прикажана архитектурата на системот за препознавање и класификација на звучни настани.



Слика 4.8. Архитектура на систем за препознавање и класификација на звучни настани

Крајната цел е од широк спектар на звучни настани, преку примена системот за препознавање и класификација на звучни настани да се добие препознаен звучен настан, односно да се најде класата во која припаѓа доминантниот звучен настан (слика 4.9).



Слика 4.9. Систем за препознавање и класификација на звучни настани

Основен параметар кој ќе се користи за проценка на успешноста на моделот е точноста при класификација на множеството од тестирани податоци, која претставува однос помеѓу точно предвидени податоци и вкупниот број на предвидени податоци во проценти. Како резултат на точноста се добиваат конфузните матрици, кои ги прикажуваат резултатите од класификацијата за секоја класа одделно. На конфузните матрици се прикажани 10те класи на звучни настани како вистинска и предвидена класа, а при висока точност дијагоналата во конфузните матрици треба да има најголем број на податоци.

МЕЛ Фреквентните Цепстрални коефициенти (MFCC) се користат како основни параметри при изградба на системот за препознавање и класификација на звучните настани поради високата точност што ја постигнуваат при тренирање и тестирање на системот. На табела 4.2. се прикажани резултати од точноста добиена при тестирање на податоците согласно бројот на коефициенти на овој параметар. Согласно бројот на МЕЛ филтер банки, системот се тестираше помеѓу 10 и 40 коефициенти. Од резултатите прикажани на табелата 4.2, може да се забележи дека највисока точност при тестирање се постигнува при употреба на 40 коефициенти за сите три алгоритми на машинско учење: 55,07% за алгоритмот на случајни шуми (RF), 51,05% за алгоритмот на машини со потпирачки вектори (SVM) и 47,19% при примена на наивен Баесов (NB) алгоритам.

Табела 4.2. Процентуален приказ на точноста на системот при тестирање во зависност од бројот на параметри на MFCC

		Број на МЕЛ фреквентни цепстрални карактеристики				
		10	20	30	40	50
алгоритам	Случајни шуми	30,12 %	48,02 %	50,13 %	55,07 %	55,02 %
	Машини со потпирачки вектори	27,72 %	47,12 %	49,02 %	51,05 %	50,85 %
	Наивен Баесов класификатор	25,61 %	44,52 %	46,3 %	47,19 %	47,19 %

Понатаму, користејќи ги МЕЛ Фреквентните Цепстрални коефициенти, се додава уште еден параметар на звучни настани со цел да се анализира како секој параметар влијае на точноста на системот. Од еден податок на звучен настан се екстрахираат по два параметри кои се поврзуваат во еден вектор на карактеристики. Вака формираните вектори на карактеристики претставуваат влез во алгоритмот на машинско учење. Од резултатите прикажани на табела 4.3, може да се забележи дека МЕЛ спектрограмот ја намалува точноста на тестираните податоци, особено кај наивниот Баесов алгоритам, каде што точноста се намалува за 21,39%, што е речиси половина помала точност од примената само на параметарот MFCC. При примена на алгоритмот на случајни шуми точноста се намалува за 8,12%, додека кај машините со потпирачки вектори за 0,87%. При тестирање на системот со употреба на параметрите MFCC и Спектрален контраст доаѓа до подобрување на точноста за алгоритмот на случајни шуми за 2,15%, додека кај останатите модели постигнатата точност се намалува.

Примената на Хроматограм и Спектрален контраст доведуваат до зголемување на точноста при тестирање на системот. Највисока точност кај алгоритмот на СШ од 64,04% се постигнува со примена на аудио параметрите MFCC и Хроматограм, додека кај моделите кои го користат алгоритмот на МПВ и НБ алгоритам се постигнува највисока точност при комбинирање на параметрите MFCC и Спектрален Контраст (58,78% за МПВ и 50,53% за НБ).

Табела 4.3. Анализа на точноста на тестираните податоци во зависност од примена на различни параметри

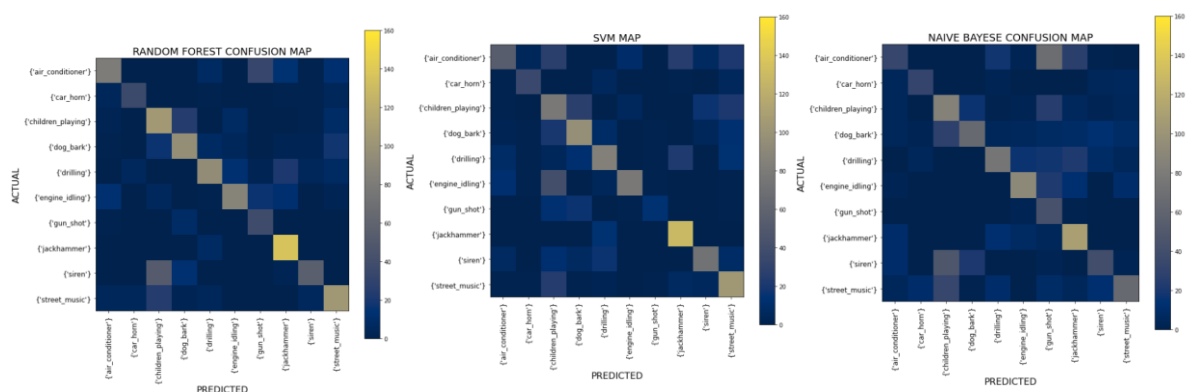
	Случајни шуми	Машини со потпирачки вектори	Наивен Баесов класификатор
MFCC	55,07%	51,05%	47,19%
MFCC + Спектрален Контраст	63,56%	58,78%	50,53%
MFCC + Тонален Центроид	57,22%	50,29%	45,04%
MFCC + Хромаграм	60,08%	53,17%	47,72%
MFCC + МЕЛ спектрограм	42,16%	49,91%	22,33%
MFCC + Спектрален Контраст+ТЦ	65,47%	58,66%	47,91%
MFCC + Спектрален Контраст + Хроматограм	66,31%	59,38%	50,78%
MFCC + Спектрален Контраст + МЕЛС	58,12%	55,45%	26,33%
MFCC + ТЦ+ Хроматограм	61,31%	52,78%	48,92%
MFCC + ТЦ + МЕЛС	62,22%	52,92%	48,02%
MFCC + Хромаграм +МЕЛС	57,58%	52,21%	26,16%
MFCC + СК + ТЦ + Хроматограм	65,88%	58,42%	48,98%
MFCC + СК + ТЦ + МЕЛС	60,45%	55,13%	29,92%
MFCC + СК + Хроматограм + МЕЛС	61,41%	57,59%	26,28%
MFCC + ТЦ + Хроматограм + МЕЛС	58,13%	56,72%	26,22%
Сите 5 параметри	62,60%	55,67%	26,40%

Во следната фаза се прави комбинација на повеќе од два параметри за да се провери дали може да се подобри точноста при препознавање и класификација на звучните настани. Може да се забележи дека најдобри резултати за трите алгоритми на машинско учење се постигнуваат при екстракција на МЕЛ Фреквентни Цепстрални Коефициенти, Хроматограм и Спектрален Контраст. При користење на овие три параметри се користат вкупно 59

коефициенти: 40 на MFCC, 12 на Хроматограм и 7 на Спектрален Контраст. Овие 59 коефициенти се излекуваат за секој аудио сигнал, и притоа тие го формираат векторот на карактеристики, кој понатаму претставува влезен параметар во алгоритмот на машинско учење. Може да се забележи дека параметрите МЕЛ спектрограм и тонален центроид ја намалуваат точноста на моделот.

Следно, комбинирани се четири параметри на звучни настани од каде што се добиени четири различни комбинации на вектор на карактеристики. Од резултатите може да се заклучи дека точноста се намалува, особено кога се користи МЕЛ спектрограмот во комбинација со другите параметри. Иако е познато дека МЕЛ спектрограмот покажува ветувачки резултати при користење на техниките на длабоко учење, од резултатите прикажани во табелата може да се потврди дека овој параметар значително ја намалува точноста при користење на класичните алгоритми на машинско учење. Дури и при примена на сите пет параметри на звучни настани со кои се излекуваат 193 коефициенти за секој аудио сигнал, може да се забележи дека точноста на моделот е повторно пониска за разлика од точноста добиена при примена на трите споменати карактеристики.

Од оваа анализа, може да се заклучи дека при изградба на системот за препознавање и класификација на звучни настани, најповолно е да се користат параметрите MFCC, хроматограм и спектрален контраст бидејќи постигнуваат највисока точност при тестирање на моделот. Највисока точност од 68,34% се постигнува при примена на алгоритмот на случајни шуми, понатаму со 59,38% при примена на алгоритмот на машини со потпирачки вектори, а најниска точност од 50,78% постигнува Наивниот Баесов алгоритам. На слика 4.10 се прикажани конфузните матрици за најуспешните модели за сите три алгоритми, каде што на вертикалата е прикажана вистинската класа на податокот, а на хоризонталата предвидената класа на податокот. Од сликата може да се забележи дека дијагоналата што ги означува точно предвидените класи има најголем број на податоци. Конфузните матрици од останатите модели чии резултати се прикажани на табела 4.3 се прикажани во прилог Б-1.



Слика 4.10. Конфузни матрици за најуспешните модели на алгоритмот на случајни шуми, машини со потпирачки вектори и наивен Баесов алгоритам

Кај алгоритмот на случајни шуми може да се забележи дека класата на сирена е помешата со деца што играат, како и клима уред со работа со чекан. Кај алгоритмот на машини со потпирачки вектори најголеми грешки се јавуваат кај класата на работа на мотор во

мирување со деца што играат, додека кај наивниот Баесов класификатор може да се забележат поголем број на грешно предвидени класи на звучни настани.

Бидејќи постигнатата точност на најуспешните модели е мала, во следната фаза, целта е да се подобри точноста преку примена на хиперпараметарска оптимизација на алгоритмите на машинско учење, каде преку дефинирање гранични вредности на алгоритмот на машинско учење се добива оптимален модел кој постигнува висока точност. Преку примена на автоматска хиперпараметарска оптимизација се овозможува ефективно моделирање на алгоритмот на машинско учење.

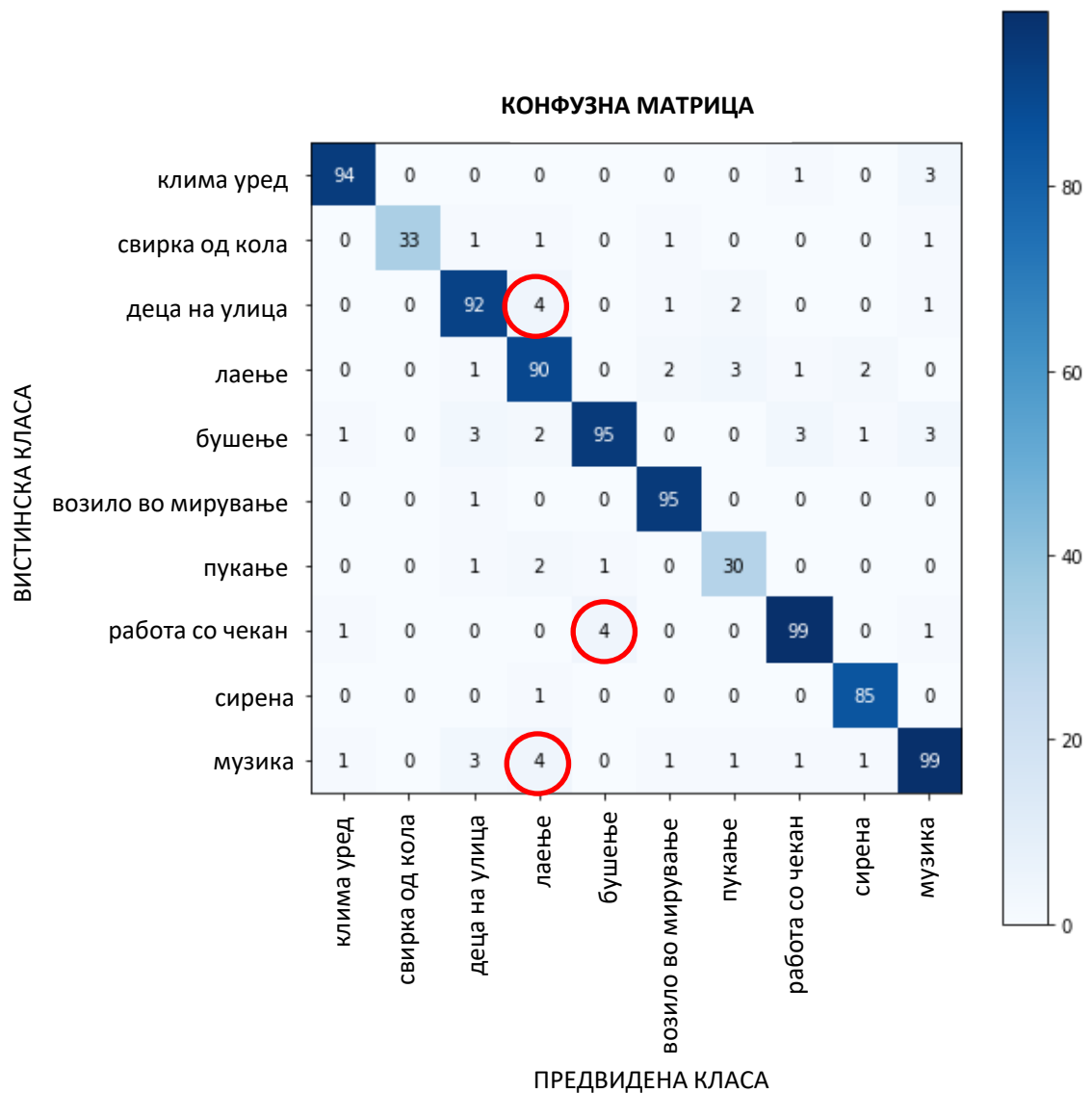
4.7. РЕЗУЛТАТИ ОД ОПТИМИЗАЦИЈА НА АЛГОРИТМОТ НА МАШИНИ СО ПОТПИРАЧКИ ВЕКТОРИ

Примената на хиперпараметарска оптимизацијата на алгоритмот на машини со потпирачки вектори ја зголемува точноста при класификација. За оптимизација се користи најуспешниот модел преку користење на три параметри за екстракција на звучните настани. Искористени се три хипер параметри за оптимизација на алгоритмот на машини со потпирачки вектори: регулационен параметар C , 'kernel' коефициент и 'gamma' коефициент. Регулациониот параметар ја означува јачината на регулацијата. Вредноста на овој параметар мора да биде позитивна, а стандардната вредност која се задава е 1. Согласно оптимизацијата, избран е коефициент со вредност 30. 'Kernel' коефициентот го означува начинот на дефинирање на хипер-рамнината, која може да биде зададена како функција на радијална основа, полиномна или сигмоидна функција. Поради тоа што се работи со 10 класи на податоци, најповолна функција е функцијата на радијална основа. Радијалната базна функција се користи како активациска функција, чија мрежа претставува линеарна комбинација од радијални базни функции кои овозможуваат лесна поделба при класификација на повеќе класи на податоци. Гама коефициентот ја дефинира вредноста на 'Kernel' коефициентот и избрана е вредност 0,0001. Во прилог Б-1 е прикажан применетиот модел за оптимизација на алгоритмот на машини со потпирачки вектори применувајќи ги параметрите MFCC, хроматограм и спектрален контраст.

По направената оптимизација и подесувањето на трите хипер-параметри добиена е точност од 98,8% на тренираните податоци, додека на тестираните податоци точноста е 92,9%. Со примена на хипер параметарска оптимизација зголемена е точноста на предвидување на податоците за 33,52%.

На слика 4.11 е прикажана конфузна матрица, каде што може да се направи споредба помеѓу вистинската и предвидената класа на тестираните податоци. На сликата може да се

забележи дека доминантен е бројот на точно предвидени податоци (дијагоналата со најголем број на податоци).



Слика 4.11. Конфузна матрица на тестираните податоци при примена на алгоритам на машини со потпирачки вектори

Најдобро предвидени класи се звучните настани од работа на возило во мирување и сирена, каде што има само по 1 грешно предвиден податок. Најголема грешка при предвидување на податоците се јавува кај звучниот настан предизвикан од гласна музика. Од табелата може да се забележи дека најголеми грешки при предвидување се случуваат помеѓу гласна музика (вистинска класа) и лаење на куче (предвидена класа), деца што играат на улица (вистинска класа) и лаење на куче (предвидена класа), како и работа со чекан (вистинска класа) и бушење (предвидена класа).

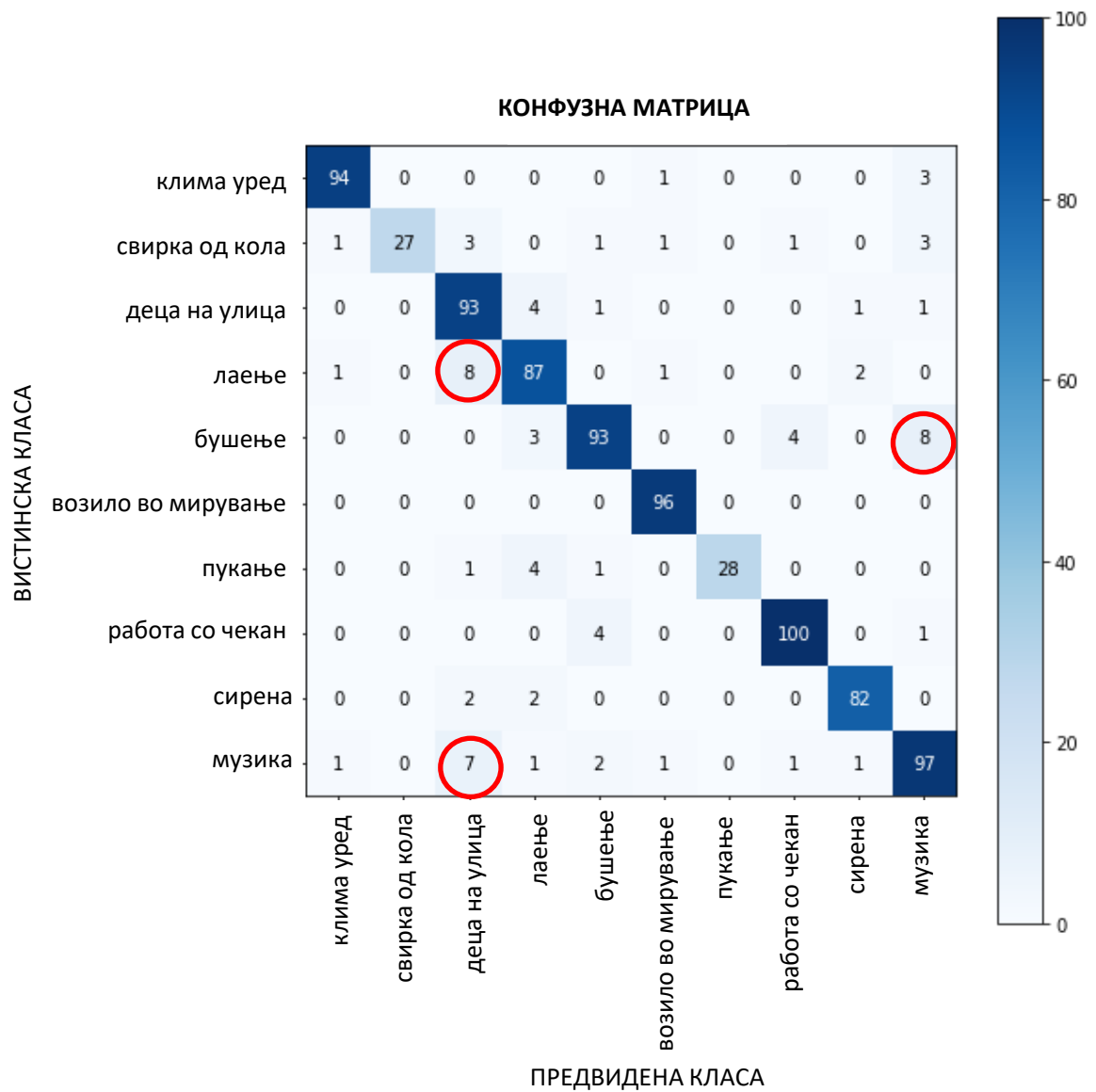
4.8. РЕЗУЛТАТИ ОД ОПТИМИЗАЦИЈА НА АЛГОРИТМОТ НА СЛУЧАЈНИ ШУМИ

При користење на алгоритмот на случајни шуми се постигнува највисока точност од 68.34%, од каде се јавува потреба за примена на оптимизација за зголемување на точноста. При хипер параметарската оптимизација се користат три параметри за оптимизација: максимална длабочина, минимален број на листови и број на естиматори. На табела 4.4. е прикажан изборот на бројот на параметри, од кој е избран означениот број за секој од параметрите. Бројот на естиматори го означува бројот на дрвја во алгоритмот, стандардната вредност е 100, а согласно оптимизацијата избрани се 1500 естиматори. Максималната длабочина претставува колку се расчленува едно дрво, а за овој параметар е избрана вредност 60. Минималниот број на листови го означува минималниот број на јазли (поделби) на едно дрво, а согласно оптимизацијата избран е 1. Постојат повеќе параметри (критериум на одлучување, максимален број на карактеристики, број на максимални примероци...), но за нив е земена стандардна вредност дадена од самиот алгоритам поради тоа што се смета дека овие карактеристики нема да направат значителна промена во класификацијата на податоците. Во прилог Б-2 е прикажан применетиот модел за оптимизација на алгоритмот на случајни шуми применувајќи ги параметрите MFCC, хроматограм и спектрален контраст.

Табела 4.4. Ранг на вредности на искористените хипер параметри

	Ранг на вредности	Избрана вредност
Број на естиматори	100:3000	1500
Максимална длабочина	10:80	60
Минимален број на листови	1:8	1

По направената оптимизација и подесувањето на трите хиперпараметри, добиена е точност од 99,92% на тренираните податоци, додека на тестираните податоци точноста е 91,53%. Со примена на хипер параметарска оптимизација зголемена е точноста на предвидување на податоците за 23,53%. На слика 4.12 е прикажана конфузна матрица, каде што може да се направи споредба помеѓу вистинската и предвидената класа на тестираните податоци.

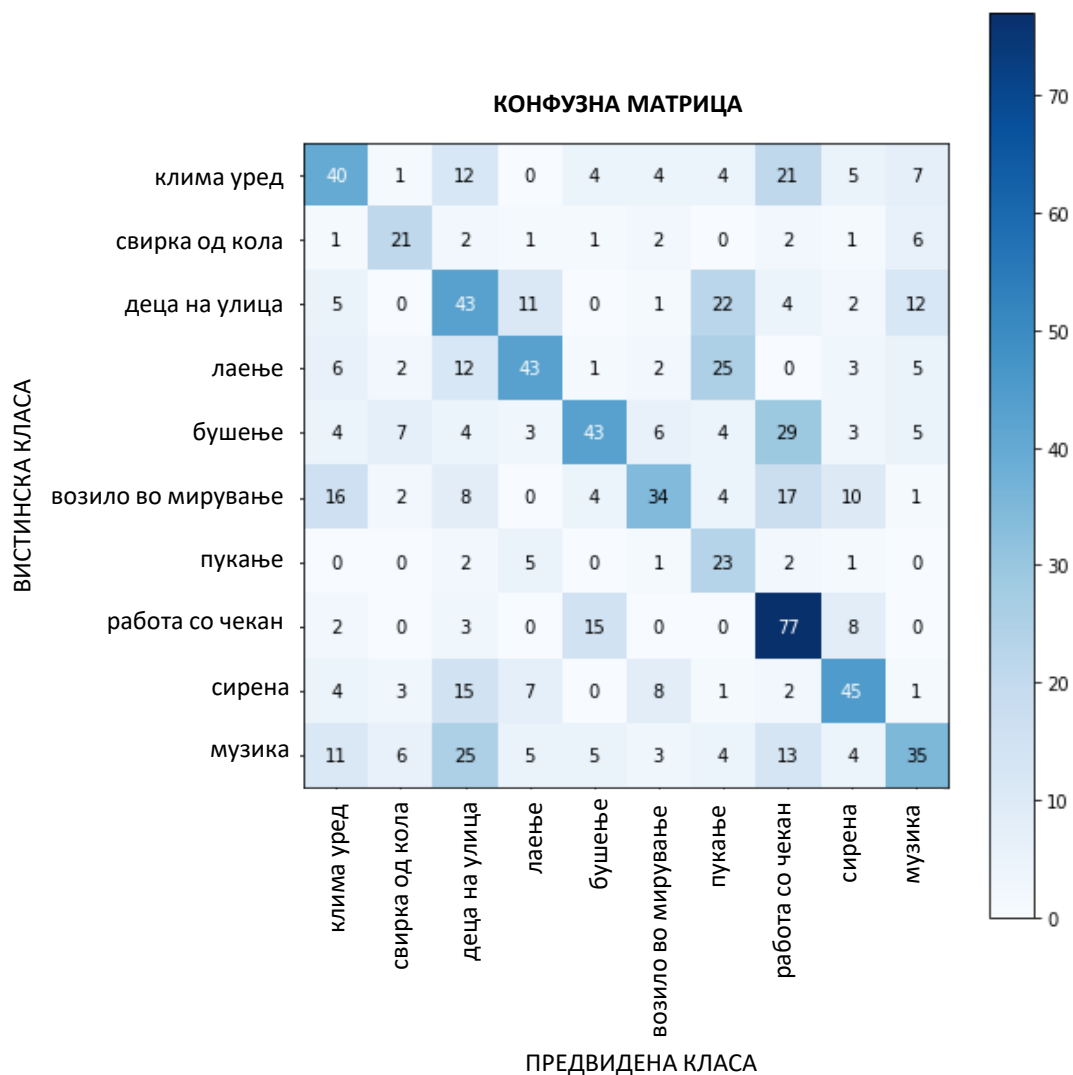


Слика 4.12. Конфузна матрица на тестираните податоци при примена на алгоритам на случајни шуми

Најдобро предвидена класа е звучниот настан од работа на возило во мирување, каде што има 1 погрешно предвиден податок. Од табелата може да се забележи дека најголеми грешки се случуваат помеѓу гласна музика (вистинска класа) и деца што играат на улица (предвидена класа), лаење на куче (вистинска класа) и деца што играат на улица (предвидена класа) и бушење (вистинска класа) и гласна музика (предвидена класа).

4.9. РЕЗУЛТАТИ ОД ПРИМЕНА НА НАИВЕН БАЕСОВ АЛГОРИТАМ

Наивниот Баесов алгоритам постигнува најниска точност од 50,78% при предвидување на податоците при употреба на полиномната распределба како основна распределба која се применува при употреба на овој алгоритам. При примена на овој алгоритам може да се примени оптимизација при распределбата, односно да се изберат три различни распределби: Гаусова, полиномна и мултиномна распределба. Најефикасна е Гаусовата распределба при примена на Наивен Баесов алгоритам, каде што е постигната точност од 53,68% на тренираните податоци, а 51,22% на тестираните податоци. Во прилог Б-3 е прикажан применетиот наивен Баесов алгоритам применувајќи ги параметрите MFCC, хроматограм и спектрален контраст. На слика 4.13 е прикажана конфузна матрица, каде што може да се направи споредба помеѓу вистинската и предвидената класа на тестираните податоци.



Слика 4.13. Конфузна матрица на тестираните податоци при примена на Наивен Баесов алгоритам

Од табелата може да се забележи погрешна класификација на податоците за сите 10 класи. Најголемите грешки се јавуваат при препознавање на звучната класа: деца што играат, работа на возило во мирување и гласна музика. Постигнатата точност при примена на Наивниот Баесов алгоритам го прави овој алгоритам неприменлив во системите за препознавање и класификација на звучните настани.

4.10. КОМПАРАТИВНА АНАЛИЗА НА РЕЗУЛТАТИТЕ

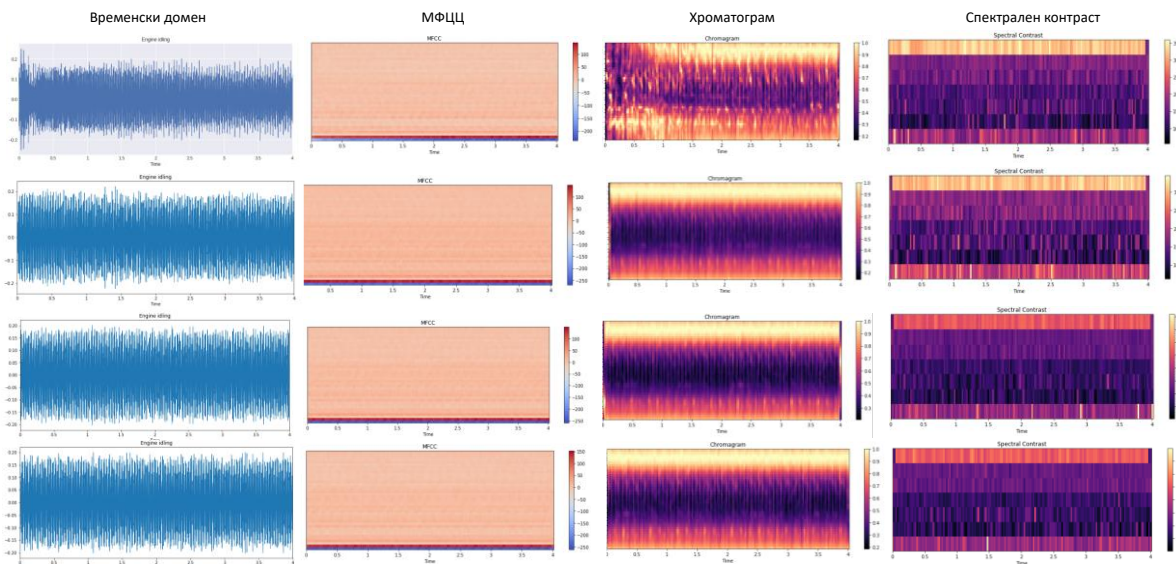
При анализа на применетите стандардни алгоритми на машинско учење, може да се забележи дека алгоритмот на машини со потпирачки вектори постигнува најголема ефикасност со 92,9% точност при предвидување на класа во која припаѓа тестираниот податок. Од останатите алгоритми, алгоритмот на случајни шуми постигнува висока точност од 91,53%, додека Наивниот Баесов класификатор има многу мала прецизност од 51,22%. Бидејќи Наивниот Баесов класификатор покажува многу мала прецизност при тренирање и тестирање на податоците, може да се заклучи дека овој алгоритам не наоѓа примена во системите за препознавање и класификација на звучни настани во урбана средина. Поради тоа, овој систем нема да се споредува во понатамошната анализа.

Од применетата методологија може да се забележи дека клучна улога при градење на успешен систем за препознавање и класификација на звучни настани има изборот на аудио параметри и правилната дигитална обработка на сигналите, како и примената на процесот на хипер параметарска оптимизација на применетите алгоритми.

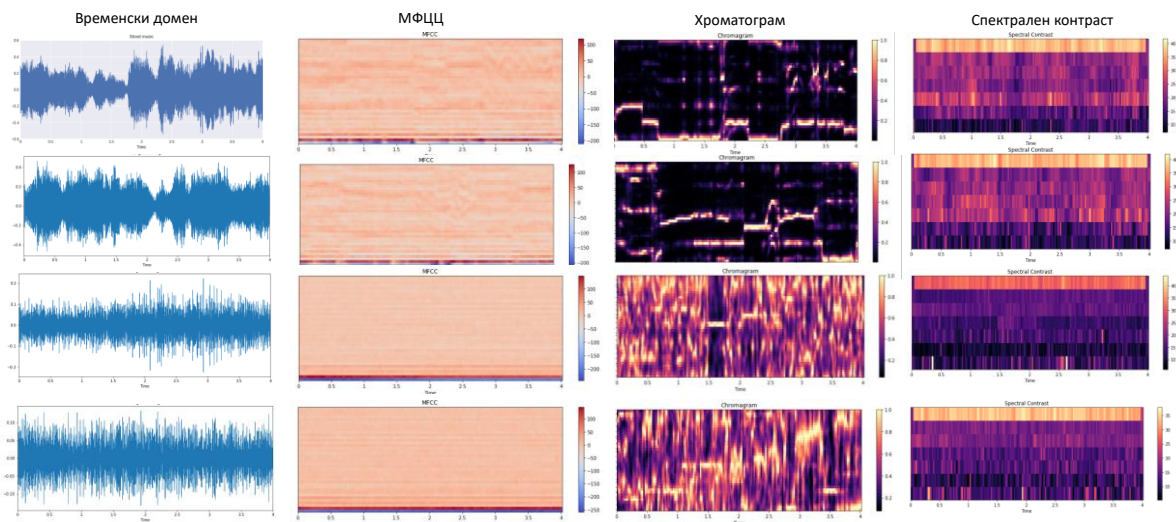
Од резултатите може да се забележи дека најточно предвидена класа е класата на звучниот настан што ја претставува работата на возилото во мирување, каде што се појавува само еден грешно предвиден податок. Звукот од сирена од кола и работа на клима уред се настани кои појавуваат малку грешки при нивно препознавање. Најголеми грешки при предвидувањето кај сите три алгоритми се јавуваат за исти класи на податоци: бушење и гласна музика. При анализа на параметрите на звучни настани, односно при нивна визуелизација може да се забележи дека звучната класа која ја претставува музиката содржи многу елементи во неа, и секој податок кој ја претставува оваа класа има различна визуелна репрезентација. Кај звучната класа која го претставува бушењето, се појавуваат звуци на ниски фреквенции и самиот звук од бушење има многу шум во него, што најверојатно е причината за појава на грешка при предвидување. Исто така, при примена на алгоритмот на случајни шуми, може да се забележи дека најголемиот број на грешно предвидени податоци се случува кај звучната класа на бушење (како вистинска класа) со гласна музика (како предвидена класа). Овие две класи на звук предизвикуваат намалување на точноста на предвидување на податоците. При аудио визуелизација на повеќе репрезентативни звуци работа на возило во мирување како класа со најпрецизно предвидување и споредба со класата на бушење и гласна музика,

екстрахирани се аудио параметрите кои се применети за трите анализирани модели за препознавање и класификација на звучните настани.

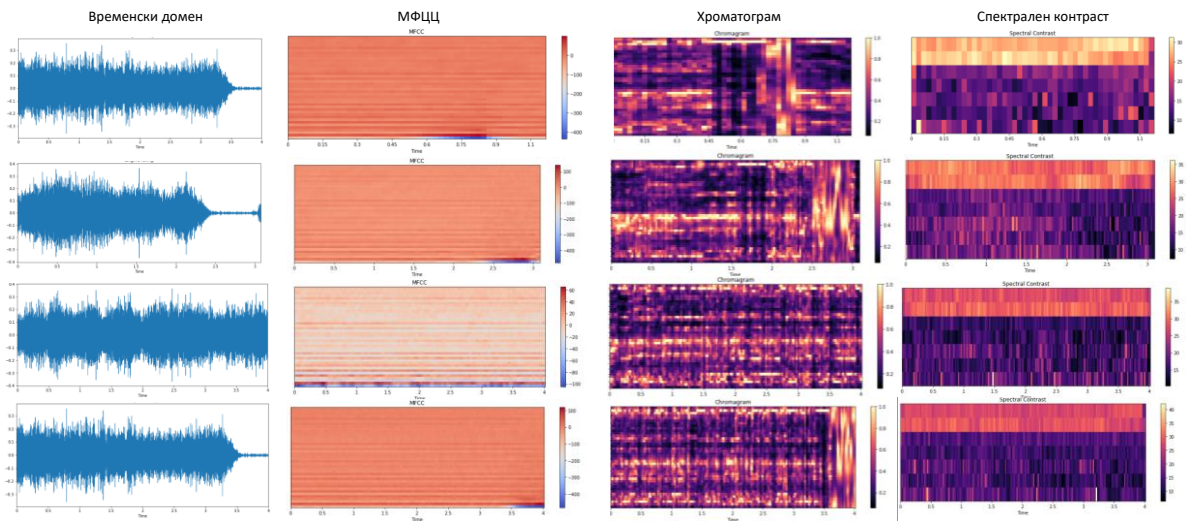
РАБОТА НА ВОЗИЛО ВО МИРУВАЊЕ



ГЛАСНА МУЗИКА



БУШЕЊЕ



Слика 4.14. Аудио параметри на репрезентативни аудио сигнали на звучните класи од работа на возило во мирување, гласна музика и бушење

Од сликата 4.14 може визуелно да се забележи дека класата на работа на возило во мирување има слични карактеристики, додека кај класите на гласна музика и бушење може да се забележат различни визуелни репрезентации. Една од причините зашто се појавуваат различни визуелни репрезентации е поради тоа што музиката има различни елементи и доста репрезентативни звуци, додека пак при бушење се појавуваат ниски фреквенции кои предизвикуваат шум. Грешката при класификација на овие два звучни настани најверојатно настанува уште при екстракција на аудио параметрите, а не при примена на алгоритмите на машинско учење.

5. ПРИМЕНА НА АЛГОРИТМИ НА ДЛАБОКО УЧЕЊЕ

Како најприменувани алгоритми кои ги користат на техниките на вештачката интелигенција за препознавање и класификација на звучни настани во литературата можат да се најдат алгоритмите базирани на длабокото учење. Во ова поглавје разработени се конволуциските невронски мрежи како алгоритми кои бележат голем развој во полето на длабокото учење и вештачката интелигенција во последните децении поради големата ефикасност при нивна примена. Даден е теоретски осврт на моделирање на алгоритми на конволуциски невронски мрежи со цел да се разработи математички модел на алгоритмот кој ќе се користи за препознавање и класификација на звучните настани кои беа параметризирани во глава 3.

Развиениот алгоритам е имплементиран на реален систем користејќи 4 различни вектори на карактеристики, резултирајќи во развој на 4 алгоритми на конволуциски невронски мрежи. Дополнително, примената оптимизација која се базира на нестационарни проблеми со градиенти кои имаат шум во нив, овозможи подобар перформанс на системот на препознавање и класификација на звучни настани. При тестирање на алгоритмите, два модели покажаа одлични перформанси при нивна работа постигнувајќи точност повеќе од 90%.

Преку анализа и имплементацијата на овие алгоритми ќе се овозможи компарација помеѓу системите базирани на длабоко учење и системите базирани на машинско учење, кои беа разработени и имплементирани во глава 4. Моделирањето со помош на техниките на длабокото учење има предност пред класичните алгоритми на машинско учење доколку системот што треба да се моделира е комплексен и треба да се примени во апликации во реално време. Споредбата на експерименталните резултати доведува до избор на систем за препознавање и класификација на звучните настани кој ќе се користи за понатамошна валидација.

Валидација на систем базиран на техниките на вештачка интелигенција е процес на дополнително тестирање на системот со непознати податоци од независна база на податоци. Валидацијата беше спроведена во 2 фази, преку валидација со снимени податоци, и валидација со снимање на податоците во реално време. Во процесот на валидација, точноста на системот беше тестирана во 10 итерации.

5.1. ДЛАБОКО УЧЕЊЕ

Длабоко учење (Deep learning) е техника на вештачка интелигенција која користи алгоритми кои се структурно и концептуално инспирирани од човековиот биолошки нервен систем. Алгоритмите на длабоко учење се составени од повеќе слоеви за обработка на податоците кои овозможуваат приказ преку повеќе нивоа на апстракција со цел постигнување на висока прецизност при класификација.

Основите на длабокото учење се поставени во 1965 година, а истите се развиени врз основа на методот на индуктивно статистичко учење. Во 1980 година, длабокото учење почнува да се применува во системи за препознавање и класификација на разни карактеристики, додека забележителен раст во неговото користење се забележува во 2012 година поради самата моќност и точност што ја постигнуваат алгоритмите на длабокото учење при препознавање на слики и звучни настани. Со развојот на напредните технологии се зголемува капацитетот на алгоритмите на длабоко учење, па една длабока невронска мрежа може да содржи повеќе од илјада нивоа на поделба што овозможува успешно меморирање на многу карактеристики научени при тренирање на алгоритмот.

При примена на длабокото учење, податоците се обработуваат преку серија на процесирања, каде што еден чекор на процесирање се означува како еден слој. Основниот алгоритам на длабоко учење е составен од мрежа на податоци составени од влезен слој (влезни податоци), скриени слоеви (применета архитектура за поделба на класите) и излезен слој (резултати). Постојат пет вида на архитектури кои се применуваат во скриените слоеви на алгоритмите на длабоко учење: рекурентни невронски мрежи (RNN), долги краткотрајни мемории (LSTM), конволуциски невронски мрежи (CNN), длабоки невронски мрежи на доверба (DBN) и длабоки натрупувачки мрежи (DSN), чија примена е дадена на табела 5.1. Согласно видот на невронската мрежа, мрежно се поврзуваат невроните кои ги преземаат влезните податоци и вршат процеси над нив, а понатаму ги препуштаат кон наредното ниво.

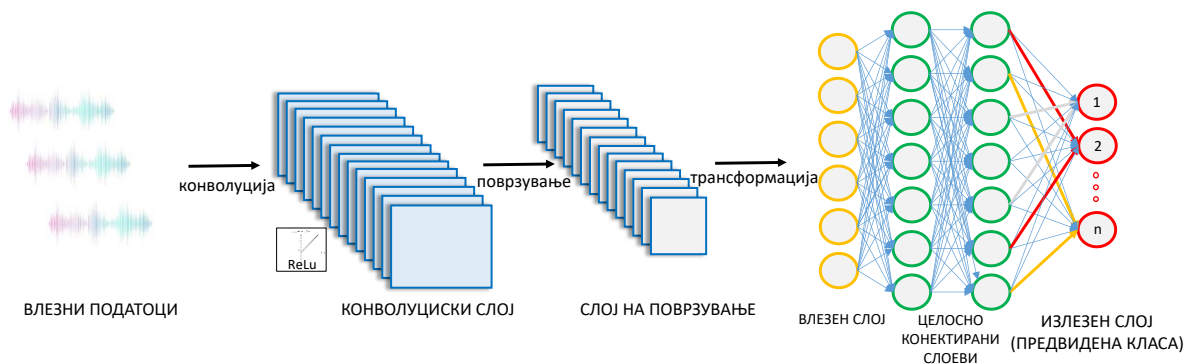
Табела 5.1. Архитектура на длабоко учење и нивна примена

АРХИТЕКТУРА (МЕТОД)	ПРИМЕНА
Рекурентна невронска мрежа RNN (Recurrent neural network)	Препознавање на говор, препознавање на рачно напишан текст
Долги краткотрајни мемории LSTM (Long short-term memory)	Препознавање на природен говор, рачно пишување, препознавање на мимики
Конволуциски невронски мрежи CNN (Convolutional Neural Network)	Препознавање на слики, анализа на видео, препознавање на звучни настани
Длабоки невронски мрежи на доверба DBN (Deep belief network)	Препознавање на слики, разбирање на природен говор, предвидување на грешки
Длабоки натрупувачки мрежи DSN (Deep Sparse-coded Network)	Препознавање на говор во континуирано говорење

За целите на оваа докторска дисертација ќе се примени алгоритам на конволуциски невронски мрежи кој е детално анализиран во следното поглавје.

5.2. КОНВОЛУЦИСКИ НЕВРОНСКИ МРЕЖИ

Конволуциските невронски мрежи (Convolutional Neural Networks - CNN) се алгоритми чија архитектура е составена од влезен, излезен и скриен слој кој вклучува повеќе конволуциски слоеви, слоеви на поврзување, слоеви за нормализација и слој за целосна конекција (слика 5.1) [137]. Овие алгоритми имаат широка примена во полето на длабокото учење кај апликации за препознавање на слики и звучни настани. Влезните податоци можат да бидат во форма на повеќекратни низи: 1D за сигнали и секвенци; 2D за слики или аудио спектограми; и 3D за видеа [138].



Слика 5.1. Архитектура на конволуциски невронски мрежи

Најважниот слој во архитектурата на конволуциски невронски мрежи е конволуцискиот слој бидејќи во овој слој се користат филтри кои формираат карактеристики од влезните податоци. Примарната функција на конволуцијата е да најде карактеристики во самиот податок (слика) преку користење на детектор на карактеристики и да креира мапа на карактеристики која покрај формата, го задржува и текот на карактеристиките. Во процесот на конволуција се формира слој од мапи на карактеристики преку примена на n -број на детектори на карактеристики (со димензии [3x3], [5x5] или [7x7]), кој уште се нарекува кернел функција или филтер. Се преработува влезниот податок и се намалува неговата димензионалност преку добивање на поедноставна форма на податок која е претставена во форма на мапа на карактеристики (слика 5.2). Со оваа трансформација се отфрлуваат карактеристики кои не се важни во процесот на идентификација на карактеристиките и се овозможува побрзо процесирање на влезните податоци во мрежата на CNN. На сликата 5.2 визуелно е прикажан конволуцискиот слој од каде што може да се забележат бројни слоеви на конволуција кои произлегуваат од примена на различни кернел функции.



Слика 5.2. Визуелен приказ на процесот на конволуција

По креирање на конволуцискиот слој, следен чекор е примена на активациска функција Rectifier Linear Unit (ReLU) за зголемување на нелинеарноста во конволуцискиот слој. Податоците во форма на слики кои се користат како влез во CNN имаат висока нелинеарност (граници, бои, елементи) и поради тоа мора да се зголеми нелинеарноста на системот за да не се добие линеарен модел кој погрешно ќе ги предвидува податоците.

Следен чекор е примена на слој на поврзување, каде што сличните карактеристики за класата на податокот се поврзуваат во една карактеристика. Најупотребувана функција е максимално поврзување (max pooling), каде што податокот се дели на блокови, а блокот со највисоката вредност (најдоминантната карактеристика) се зема како референтна вредност на излезниот податок. Оваа функција може да ја намалува димензионалноста на податокот до 75%. Следно, податоците се исправуваат, односно се трансформираат за да може да се искористат како влезен слој на едноставна вештачка невронска мрежа. На слика 5.3 се прикажани чекорите на целосно поврзување и подредување.



Слика 5.3. Примена на функција на максимално поврзување (max pooling) и подредување (flattening)

Последниот чекор се состои од целосна конекција каде што влезните податоци кои се користат за крајната вештачка невронска мрежа се целосно конектирани со секој неврон од скриените слоеви, кои при примена на конволуциски невронски мрежи се нарекуваат целосно поврзани слоеви. Разликата помеѓу вештачката невронска мрежа и конволуциската невронска мрежа е во тоа што секој неврон и карактеристика која е извлечена од податоците целосно се конектира со крајните слоеви, додека кај вештачката невронска мрежа се конектираат оние неврони кои во себе содржат важни карактеристики за класата на податокот. Постои целосно поврзување помеѓу предпоследниот слој и последниот слој кој ги дефинира класите на податоци, а преку примена на тежинските криви кои означуваат веројатност на класата на податок, самата мрежа учи кои карактеристики ќе бидат доминантни за одредена класа на податоци. Преку ова поврзување се овозможува правилна и прецизна класификација и идентификација на податоците. Дополнително, со примена на повратна контрола се овозможува постојано подобрување на архитектурата со постојана оптимизација на тежинските криви и детекторот на карактеристиките.

5.3. МАТЕМАТИЧКИ МОДЕЛ НА КОНВОЛУЦИСКИ НЕВРОНСКИ МРЕЖИ

Податоците кои се користат за тренирање и тестирање на конволуциски невронски мрежи се тридимензионални вектори – тензори од трет ред (слики со големина $H \times W \times D$, каде што H е број на редови, W е број на колони, а D го означува бројот на серија), прикажани со:

$$\{x^i\}_{l=1}^L \in R^{H_i \times W_i \times D_i} \quad 1 \leq i \leq n \quad (5-1)$$

Во изразот 5-1, со i е прикажан бројот векторите од првиот до l -тиот слој на процесирање. Постојат l слоеви на процесирање w^l кои овозможуваат обработка на податоците со цел нивна класификација:

$$\{w^l\}_{l=1}^L \quad (5-2)$$

Апстрактно, структурата на конволуциски невронски мрежи може да се прикаже со изразот:

$$x^1 \rightarrow \boxed{w^1} \rightarrow x^2 \rightarrow \dots \rightarrow x^{L-1} \rightarrow \boxed{w^{L-1}} \rightarrow x^L \rightarrow \boxed{w^L} \rightarrow z \quad (5-3)$$

Изразот 5-3 илустрира како конволуциските невронски мрежи се пресметуваат слоевито. Влезот x^1 вообичаено го прикажува податокот како тензор од 3ти ред. Најпрвин се процесира првиот слој прикажан преку w^1 . Излезот од првиот слој е x^2 , што претставува влезен податок при процесирање на вториот слој. Овој процес трае се додека не заврши процесирањето на сите слоеви, чиј излез е x^L .

Се претпоставува дека се прави класификација на податоците во C број на класи. Во овој случај, излезот x^L се претставува како вектор со C можност за излези, чиј L -ти влез ја предвидува класата (веројатност дека x^1 доаѓа од l -тата класа). За да се добие крајниот резултат од последниот излезен слој x^L , потребно е да се направи трансформација употребувајќи функција на веројатност (SoftMax) во $(L - 1)$ -тиот слој на процесирање.

Последниот слој е слој на загуби и со него се откриваат вредностите помеѓу влезната и излезната веројатност за припаѓање во одредена класа C . Доколку се претпостави дека постои t вредност која ја прикажува веројатноста на припаѓање на одредена класа на влезниот слој x^1 , тогаш функцијата на загуба ќе ја измери разликата помеѓу предвидениот податок x^L и таргетираната (крајна) вредност t . Функцијата на загуба може да се прикаже како:

$$z = \frac{1}{2} \|t - x^L\|^2 \quad (5-4)$$

Доколку се претпостави дека сите слоеви w^1, \dots, w^{L-1} од моделот на конволуциски невронски мрежи се истренирани и имаат научено од податоците, тогаш алгоритмот е подготвен за тестирање и класификација. Со процесирање на податоците низ слоевите на конволуциската невронска мрежа, крајниот излез е податок кој припаѓа на одредена класа C :

$$x^L \in R^C \quad (5-5)$$

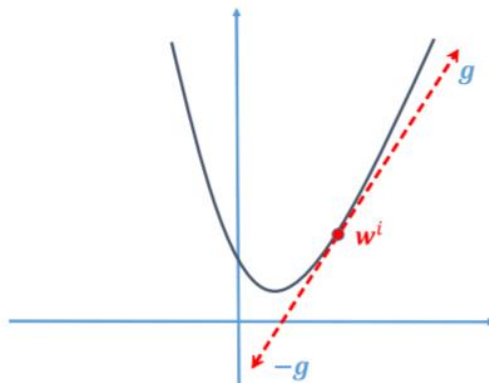
Класификацијата на податокот во одредена класа се прави преку избор на највисоката веројатност за класата со зголемување на таа веројатност до максимум:

$$\arg \max_i x_i^L \quad (5-6)$$

Параметрите на слоевите се оптимираат со цел да се намали загубата z , односно да се добие повисока прецизност при предвидување на податоците. Еден примерок при тренирање x^1 , наместо да даде предвидување во некоја класа, најпрвин треба да ја спореди предвидената класа со таргетираната (вистинската) класа t која одговара на податокот x^1 за да може да се добие загубата z .

Загубата z претставува надгледуван сигнал кој покажува како параметрите од моделот треба да се модифицираат за да се добие попрецизен модел на предвидување. Ова се овозможува со користење на Стохастички градиент на намалување (SGD) (илустрација на слика 5.4), кој ги трансформира параметрите на слоевите на следниот начин:

$$w^l \leftarrow w^l - \eta \frac{\partial z}{\partial w^l} \quad (5-7)$$



Слика 5.4. Илустрација на методот на Стохастички градиент на намалување

Во изразот 5-7, знакот \leftarrow имплицитно прикажува дека параметрите w^l од l -тиот слој се обновени од време t во време $t + 1$. Ако временскиот индекс t се користи експлицитно, равенката ќе се прикаже во следниот облик:

$$(w^l)^{t+1} = (w^l)^t - \eta \frac{\partial z}{\partial (w^l)^t} \quad (5-8)$$

Парцијалниот извод $\partial z / \partial w^l$ ја мери стапката на пораст на загубата z во однос на промените на разните димензии на слоевите w^l и се користи за обновување на параметрите во l -тиот слој. Векторот на парцијалниот извод се нарекува градиент на оптимизација во кој има мал локален регион околу моменталната вредност w^l . За да се помести w^l во насока која е определена од градиентот, треба да се зголеми објективната вредност на загубата z . Со цел да се намали функцијата на загуба, потребно е да се обнови w^l кон спротивната насока од градиентот. Ова правило се нарекува градиент на намалување. Градиентот на намалување (g) е илустриран на слика 5.4.

Доколку се помести вредноста многу подалеку кон негативната вредност на градиентот може да дојде до зголемување на функцијата на загуба. Во секоја обнова се менуваат параметрите во мали граници кон негативниот градиент, а истите се контролирани од стапката на учење η :

$$\eta \ll 1 \quad (5-9)$$

Едно обновување кое што се заснова на влезните податоци x^l ќе направи помала загуба при тренирање доколку стапката на учење има мала вредност. Може да се повтори стапката на градиентно намалување за обновување на епохите (процес на обновување на параметрите на сите тренирани податоци) за да се определат параметрите на конволуциски невронски мрежи. Една епоха генерално ќе ја намали средната загуба на сетот на тренирање се додека системот се тренира, па со зголемување на бројот на епохи ќе се намали загубата и ќе се зголеми прецизноста при предвидување.

Доколку се претпостави дека во l -тиот слој влезните податоци се тензори од трет ред $x^l \in R^{H^l \times W^l \times D^l}$, потребно е сетот од карактеристики (i^l, j^l, d^l) на секој слој за да се поврзе за елемент во податокот x^l за $0 \leq i \leq H^l = H^{l+1}$, $0 \leq j \leq W^l = W^{l+1}$, и $0 \leq d \leq D^l = D^{l+1}$. Во последниот l -ти слој, влезните податоци x^l се трансформираат во излезни податоци $x^{l+1} = y$.

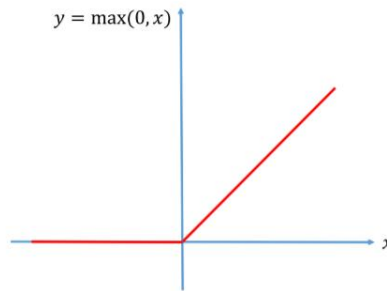
Може да се заклучи дека излезот од последниот слој е предвидената класа ($y = x^{l+1}$) преку користење на изразите:

$$\frac{dy_{i,j,d}}{dx_{i,j,d}^l} = [x_{i,j,d}^l > 0] \quad (5-10)$$

$$\left[\frac{\partial z}{\partial x^l} \right]_{i,j,d} = \begin{cases} \left[\frac{\partial z}{\partial y} \right]_{i,j,d} & \text{ако } x_{i,j,d}^l > 0 \\ 0 & \text{иначе} \end{cases} \quad (5-11)$$

По креирање на конволуцискиот слој, се користи линеарна исправувачка функција (ReLU) која е прикажана на слика 5.5, и може да се претстави како:

$$y_{i,j,d} = \max\{0, x_{i,j,d}^l\} \quad (5-12)$$



Слика 5.5. Графички приказ на ReLU функцијата

При користење на ReLU функцијата не се користат параметри, бидејќи нема учење во овој слој, само зголемување на нелинеарноста на податоците. Како излез од оваа функција, податоците може да имаат позитивна вредност или нула (позитивна кога има слични карактеристики, а нула кога има регион кој не ги препознава карактеристиките). Излезот $y_{i,j,d}^l$ се активира кога има карактеристики кои се препознатливи во одреден регион. Постојат и други нелинеарни трансформации кои се користат за да се постигне нелинеарност, како што е сигмоидна функција или експоненцијална линеарна функција.

Конволуцискиот слој (Convolutional layer) овозможува намалување на димензионалноста на векторот. За тензори од трет ред, процесот на конволуција се дефинира слично. Доколку се претпостави дека влезот во l -тиот слој е тензор од трет ред со големина $H^l \times W^l \times D^l$, а конволуцискиот кернел е исто така тензор со големина $H \times W \times D^l$, конволуцијата се прави преку производот на елементите кои одговараат на влезниот податок во каналите D^l и сумата на HWD^l за да се добие резултатот од конволуцијата на просторната локација. Потоа, се преместува кернелот за да се комплетира процесот на конволуција.

Во конволуцискиот слој, постојат повеќе конволуциски кернели. Доколку се претпостави дека постојат D кернели и секој кернел е искористен во просторното јадро $H \times W$, сите кернели ќе се претстават како тензор од 4-ти ред f во $R^{H \times W \times D^l \times D}$. Слично, се користат индексите на променливите $0 \leq i < H, 0 \leq j < W, 0 \leq d < D$ и $0 \leq d^l < D^l$ за прецизно да се посочи специфичен елемент во кернелите. Преку користење на конволуцискиот слој се намалува бројот на параметри, димензионалноста на векторот, а со тоа и времето за пресметка.

Се употребува стандардна векторска операција за да се направи конверзија на сетот од конволуциски кернели f (тензор од 4ти ред) во матрица F прикажана со изразот (5-13):

$$vec(y) = vec(x^{l+1}) = vec(\phi(x^l)F) \quad (5-13)$$

$$vec(y) \in R^{H^{l+1} \times W^{l+1} \times D} \quad (5-14)$$

$$\phi(x^l) \in R^{(H^{l+1} \times W^{l+1}) \times (HWD^l)} \quad (5-15)$$

$$F \in R^{(HWD^l) \times D} \quad (5-16)$$

Производот на матриците $\phi(x^l)$ и F во изразот (5-13) одговара на димензионалноста на векторот $vec(y)$.

Следно, потребно е да се пресметаат двата изводи на загубата за матрица: $\frac{\partial z}{\partial vec(x^l)}$ и $\frac{\partial z}{\partial vec(F)}$, каде првиот израз ќе биде искористен за повратна пропација на преходниот слој ($l - 1$), а вториот израз ќе определи како параметрите на моменталниот l -ти слој ќе се обноват.

Изразот $\frac{\partial z}{\partial vec(F)}$ се пресметува на следниот начин:

$$\frac{\partial z}{\partial (vec(F))^T} = \frac{\partial z}{\partial (vec(Y))^T} \frac{\partial vec(y)}{\partial (vec(F))^T} \quad (5-17)$$

Изразот $\frac{\partial z}{\partial vec(x^l)}$ во $(l + 1)$ -тиот слој е еквивалентен на изразот $\frac{\partial z}{\partial (vec(x^{l+1}))^T}$. Може да се заклучи дека:

$$\frac{\partial z}{\partial F} = \phi(x^l)^T \frac{\partial z}{\partial Y} \quad (5-18)$$

Ова е едноставно правило кое служи за обновување на параметрите во l -тиот слој: градиентот со помош на конволуциските параметри е производ од $\phi(x^l)^T$ и надгледуваниот сигнал кој е добиен од $(l - 1)$ -тиот сигнал $\frac{\partial z}{\partial Y}$.

По примена на конволуцискиот слој се применува слојот на поврзување. Се претпоставува дека $x^l \in R^{H^l \times W^l \times D^l}$ е влез во l -тиот слој, што сега го претставува слојот на поврзување. Операцијата на поврзување нема потреба од параметри. Доколку се заклучи дека изразот $H \times W$ е во дизајнот на CNN структурата, ќе се претпостави дека H е поделено со H^l , а W е поделено со W^l , а излезот на x^l ќе биде x^{l+1} или у што претставува тензор од трет ред со големина $H^{l+1} \times W^{l+1} \times D^{l+1}$.

$$H^{l+1} = \frac{H^l}{H} \quad W^{l+1} = \frac{W^l}{W} \quad D^{l+1} = D^l \quad (5-19)$$

Слојот на поврзување овозможува процесирање на податокот x^l во секој канал поединечно, односно во секој канал матрицата со $H^l \times W^l$ елементи е поделена на $H^{l+1} \times W^{l+1}$ подрегиони, каде секој има големина $H \times W$. Во слојот на поврзување се користат две операции: максимално поврзување (max pooling) и нормално поврзување (average pooling). Математички, овие функции се прикажуваат на следниот начин:

$$y_{i^{l+1},j^{l+1},d} = \max_{0 \leq i < H, 0 \leq j < W} x_{i^{l+1},j^{l+1},d}^l \quad (5-20)$$

$$y_{i^{l+1},j^{l+1},d} = \frac{1}{HW} \sum_{0 \leq i < H, 0 \leq j < W} x_{i^{l+1},j^{l+1},d}^l \quad (5-21)$$

За изразите (5-20) и (5-21), важат следните релации: $0 \leq i^{l+1} \leq H^{l+1}, 0 \leq j^{l+1} \leq W^{l+1}, 0 \leq d \leq D^{l+1} = D^l$.

По примена на функцијата на максимално поврзување, карактеристиките се исправуваат и се креира вештачка невронска мрежа во која постои целосно поврзување на карактеристиките. Целосната конекција целосно ги поврзува последните слоеви, со што се овозможува обновување на коефициентите на тежинските криви кои и помагаат на мрежата да учи кои карактеристики ќе бидат доминантни за одредена класа на податок. Преку ова поврзување се овозможува правилна и прецизна класификација и идентификација на податоците.

5.4. ADAM ОПТИМИЗАЦИЈА

ADAM алгоритмот се користи за оптимизација на алгоритмите засновани на длабоко учење преку адаптивна проценка на моментот (Adaptive moment estimation), кој пресметува индивидуални стапки на адаптивно учење за различни параметри од проценките на првиот и вториот момент на учење. Алгоритмот е предложен од страна на истражувачи во 2015 година [139], а има суштинска практична примена во многу области во науката и инженерството кои ги користат напредните техники на вештачка интелигенција [140, 141]. Методот за имплементација на овој алгоритам е едноставен, пресметковно ефикасен и зафаќа мала меморија, а се користи при решавање на проблеми поврзани со голема база на податоци и параметри (прикажан на табела 5.2). Овој метод се користи и за нестационарни проблеми со градиенти кои не се добро дефинирани и имаат многу шум во нив. Овој метод е дизајниран за да ги комбинира предностите на методот на адаптивен градиент [142] и методот на пресметка на средна квадратична вредност RMS (root mean square) [143].

Табела 5.2. Приказ на ADAM алгоритмот на оптимизација

ADAM алгоритам на оптимизација	
Потребна големина	α – големина на чекорот
Потребна големина	$\beta_1, \beta_2 \in [0,1]$ – експоненцијални стапка за естимација на моментот
Потребна големина	$f(\theta)$ – стохастичка објективна функција со параметри θ
Потребна големина	θ_0 вектор на иницијалниот параметар
	$m_0 \leftarrow 0$ (Иницијализација на векторот на првиот момент)
	$v_0 \leftarrow 0$ (Иницијализација на векторот на вториот момент)
	$t \leftarrow 0$ (Иницијализација на чекорот на времето)
	while θ_t not converged do
	$t \leftarrow t + 1$
	$g_t \leftarrow \nabla_{\theta} f_T(\theta_{t-1})$ – се земаат градиенти со временски чекор t
	$m_t \leftarrow \beta_1 m_{t-1} + (1 - \beta_1) g_t$ – обнова на првиот момент
	$v_t \leftarrow \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$ – обнова на вториот момент
	$\hat{m}_t \leftarrow m_t / (1 - \beta_1^t)$ – пресметка на првиот коригиран момент
	$\hat{v}_t \leftarrow v_t / (1 - \beta_2^t)$ – пресметка на вториот коригиран момент
	$\theta_t \leftarrow \theta_{t-1} - \alpha \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon)$ – обнова на параметрите
	end while
	return θ_t – резултирачки параметри

При математичко дефинирање на ADAM алгоритмот, најпрвин се дефинираат променливите, каде што $f(\theta)$ е објективна функција со шум: стохастична скаларна функција која има θ параметри. Целта е да се доведе до минимум посакуваната вредност на оваа функција $E[f(\theta)]$ со параметри θ . Со приказ на функциите $f_1(\theta), \dots, f_T(\theta)$ се дефинираат остварувањата на стохастичната функција како збир од временски чекори $1, \dots, T$. Стохастичноста може да се пресмета од евалуација на податоците добиени од функциите на временските чекори. Со $g_t = \nabla_{\theta} f_T(\theta)$ се дефинира градиентот, односно векторот од парцијални диференцијали од f_T , со карактеристики θ и временски чекор t .

Алгоритмот овозможува обновување на експоненцијалите кои ги променуваат средните вредности на градиентот m_T и квадратот на градиентот v_T каде хиперпараметрите

$\beta_1, \beta_2 \in [0,1]$ ги контролираат стапките на експоненцијален раст на овие променливи. Променливите средни вредности го претставуваат првиот момент (средната вредност), а вториот момент го претставува моментот со нецентрализирана променлива на градиентот. Овие променливи средни вредности се преставени како нулти вектори кои водат до момент кој се стреми кон нула, резултирајќи во добивање на коефициенти кои се коригирани согласно биасот (грешката). Ефикасноста на алгоритмот може да биде подобрена преку промена на редот при пресметка, односно преку додавање на функцијата:

$$\alpha_T = \alpha \sqrt{1 - \beta_1^T} / (1 - \beta_1^T) \text{ и } \theta_T \leftarrow \theta_{T-1} - \alpha_T m_T / (\sqrt{v_T} + \hat{\epsilon}) \quad (5-22)$$

При работа со Adam алгоритмот за оптимизација многу е важно да се избере правилна стапка на чекор. Ефективниот чекор во зададена временска рамка е:

$$\Delta_t = \alpha \hat{m}_t / \sqrt{v_T} \quad (5-23)$$

Постојат две гранични вредности:

$$|\Delta_t| \leq \alpha(1 - \beta_1) \sqrt{1 - \beta_2} \text{ за случај } (1 - \beta_1) > \sqrt{1 - \beta_2} \quad (5-24)$$

$$|\Delta_t| \leq \alpha \text{ за случај } (1 - \beta_1) < \sqrt{1 - \beta_2} \quad (5-25)$$

Кога $(1 - \beta_1) = \sqrt{1 - \beta_2}$, значи дека $\hat{m}_t / \sqrt{v_T} \approx \pm 1$. Ефективната големина на чекорите е ограничено со α каде $|\Delta_t| \leq \alpha$.

За многу алгоритми на машинско учење однапред се знае дека добрата оптимизација е со висока веројатност во одреден регион во параметарскиот простор. Бидејќи α воспоставува горна граница на големината на чекорите во параметарскиот простор, може да се заклучи дека правилниот избор на големината на α може да се постигне така што ќе се достигнат θ вредности во одреден број на повторувања. Ефективната големина на чекорите Δ_t е непроменлива на скалата на градиентите.

5.5. ПРИМЕНА НА КОНВОЛУЦИСКИ НЕВРОНСКИ МРЕЖИ ВО СИСТЕМИТЕ ЗА ПРЕПОЗНАВАЊЕ И КЛАСИФИКАЦИЈА НА ЗВУЧНИ НАСТАНИ

Врз основа на истражувањата и анализата на ефикасноста на параметрите на звучни настани претставени во глава 3 и 4, при примена на мрежа на конволуциски невронски мрежи ќе се искористат следните параметри: МЕЛ Фреквентни Цепстрални Коефициенти, Хроматограм и Спектрален Контраст. Дополнително, ќе се примени и МЕЛ Спектрограм поради тоа што согласно релевантните истражувања при примена на алгоритми засновани на длабокото учење, МЕЛ спектрограмот е параметар кој ја зголемува прецизноста и точноста при препознавање на звучните настани. Ќе се изградат четири модели на конволуциски невронски мрежи согласно употребените параметри на звучни настани:

1. МЕЛ Фреквентни Цепстрални Коефициенти
2. МЕЛ Спектрограм
3. МЕЛ Фреквентни Цепстрални Коефициенти + МЕЛ Спектрограм
4. МЕЛ Фреквентни Цепстрални Коефициенти + Хроматограм + Спектрален Контраст

Согласно архитектурата за секој модел на конволуциски невронски мрежи изграден во програмот Python која е прикажана во следните поглавја (5-6 до 5-11), ќе се објаснат поединечно сите блокови кои се користат:

1. **Секвенцијален модел (Sequential).** Преку употреба на оваа функција се овозможува креирање на невронска мрежа на која можат да се додаваат разни слоеви.
2. **Конволуциски слоеви.** Секој конволуциски слој учи од различни карактеристики од податоците кои се користат за тренирање на системот. Првите слоеви секогаш учат од карактеристики со понизок степен (како линии и точки на слика), додека следните слоеви учат од повисок степен на карактеристики (основни форми). Комплексноста на податоците за обработка е пропорционална со бројот на конволуциски слоеви. За целите на ова истражување, искористени се дводимензионални конволуциски слоеви (Conv2D), кои овозможуваат директно користење на сликите добиени од применетата параметризација за тренирање на мрежата. Важни параметри кои треба да се дефинираат при конструирање на Конволуциски невронски мрежи се:
 - **Големина на филтри.** Филтрите го одредуваат бројот на кернели (филтри) за преработка на влезниот волумен. Секој филтер овозможува креирање на 2D мапа за активација. При примена на повеќе конволуциски слоеви, потребно е да се зголеми бројот на филтри во мрежата.
 - **Кернел големини.** Овој параметар овозможува да се определат димензиите на кернелот. Се користат разни големини на кернелите, од 1x1 до 7x7. Системот е

тестиран и со повисока големина на кернелот како 7×7 , но со оваа база на податоци најефикасна големина е 2×2 и 3×3 . Кернел големината е број кој ја дефинира висината и ширината на 2D конволуциски прозорец.

- **ReLU активациска функција.** Активациска функција која се употребува во секој конволуциски слој.
 - **Нормализација.** Се нормализира излезот од претходниот активациски слој преку одземање на средната вредност на серијата од податокот и делење со стандардното отстапување.
 - **Чекор (strides).** Овој параметар е цел број кој го дефинира чекорот на конволуција заедно со висината и ширината на влезниот волумен. Стандардната големина која се применува е (1,1), што значи дека дадениот филтер на конволуциската невронска мрежа се применува на моменталната локација на влезниот волумен и дадениот филтер прави чекор од 1 пиксел.
 - **Порамнување (padding).** Овој параметар има две вредности 'valid' која се однесува на податоци кај кои не е применето нула порамнување или 'same' за податоци кај кои е применето нула порамнување.
3. **Максимално поврзување (Max Pooling).** Ова е често употребувана техника за намалување на димензионалноста на податокот.
 4. **Отфрлување (Dropout).** Метод на регуларизација кој се користи во невронските мрежи, а наоѓа примена во конволуциските слоеви: исфрла цели мапи на карактеристики за одредена стапка, спречувајќи силна корелација помеѓу слоевите. Оваа техника спречува преоптоварување на моделот, каде што по случаен избор се отфрла одреден број на активациски мапи. Вредностите се задаваат процентуално, па доколку се избере вредност 0.1, тоа значи дека во тој слој ќе се отфрлат 10% од активациските мапи на излез.
 5. **Глобален просечен слој на поврзување (Global Average Pooling).** Густите слоеви се заменети со додавање на овој слој кој значително ја намалува димензионалноста, а служи за пресметка на просечниот излез од секој претходен слој. Глобалниот просечен слој на поврзување ги изедначува вредностите според последната оска, односно ја претвора мапата во карактеристики со број на серии и број на канали. На пример, доколку последниот конволуциски слој има димензии (16,7,7,64) тој би се преработил во слој со димензии (16,64).
 6. **Слој на подредување (Flatten).** Овој слој конвертира активациска мапа од повеќедимензионален објект во едnodимензионален објект преку преуредување на елементите. Доколку овој слој се применува после глобалниот просечен слој на поврзување, тогаш бројот на параметри нема да се промени.
 7. **Слој на класификација (Dense).** Овој слој е последен слој што овозможува класификација на податоците. Крајниот слој содржи SoftMax функција на излез која овозможува да се добие највисока веројатност за одредената класа и класификација на податокот.

Со цел да се подобрат перформансите на алгоритмот на CNN, во програмот Python се применува оптимизација преку примена на ADAM алгоритмот и се користи за да ги обнови

тежинските криви чија вредност се заснова на тренираните податоци. Постојат 4 параметри при конфигурирање на ADAM алгоритмот на оптимизација:

1. Алфа (Alpha) – Ја означува стапката на чекор (стапката на учење). Најчесто применувана вредност е 0,001. Повисоки вредности (0,3) резултираат во побрзо иницијално учење, додека помали вредности (0,00001) означуваат мала стапка на учење за време на тренирањето.
2. Бета1 (Beta1) – Експоненцијална стапка за естимација во почетните слоеви (вообичаена вредност е 0,9).
3. Бета2 (Beta2) – експоненцијална стапка за естимација во последните слоеви (вообичаена вредност е 0,999). Ова вредност треба да биде блиску до 1 при препознавање на слики и звучни настани.
4. Епсилон (Epsilon) – Коефициент со мала вредност кој превенира појава на грешка при имплементирање на овој алгоритам (0,00000001).

При составување на моделот, потребно е да се дефинира загубата. Постојат два вида на загуби: бинарна и категорична вкрстена ентропија. Поради тоа што се работи за класификација на 10 класи, ќе се избере модел на загуба со категорична вкрстена ентропија, која се користи како функција на загуба при повеќекласна класификација.

Пред да се започне со тренирање и тестирање на системот, се прави негова визуелизација од каде што може да се забележи колона со број на параметри (params) која го покажува бројот на параметри кои се применети за тренирање на секој слој. Вкупниот број на параметри е прикажан на крајот и е еднаков на бројот на параметри искористени за тренирање.

Бројот на параметри за тренирање за секој слој (p_n) се пресметува преку бројот на влезните и излезните канали (c_i и c_o), и димензиите на кернел функцијата (k_h и k_w):

$$p_n = c_o \cdot (c_i \cdot k_h \cdot k_w + 1) \quad (5-26)$$

Слоевите кои користат параметри за тренирање се слоевите на конволуција и слојот на класификација. Дополнително, во процесот на дефинирање на конволуциската невронска мрежа, се прикажува и излезниот облик на податоците, кој што е прикажан во 4D низа во облик на (број на серија, висина, ширина, длабочина). Првата димензија ја прикажува големината на серијата на сликата, висината и ширината се согласно видот на аудио карактеристиката, а длабочината е дефинирана преку бројот на филтри во самата мрежа. Бидејќи бројот на серии не е дефиниран во влезниот аргумент, тој не е дефиниран и е означен како (None). По дефинирање на слоевите на конволуциските невронски мрежи, се дефинираат излезните слоеви на класификација кои користат дводимензионална низа дадена во обликот (број на серија, број на класи). Оваа трансформација од 4D во 2D низа се прави преку користење на слојот на исправување или преку глобалниот просечен слој на поврзување, што овозможуваат збивање на 3те димензии на сликата во единечна димензија, која понатаму може да се користи во слоевите на класификација.

По дефинирање на архитектурата на конволуциски невронски мрежи, оптимизацијата и стапката на загуба, следува процесот на тренирање и тестирање на системот, каде што е потребно е да се дефинираат два параметри:

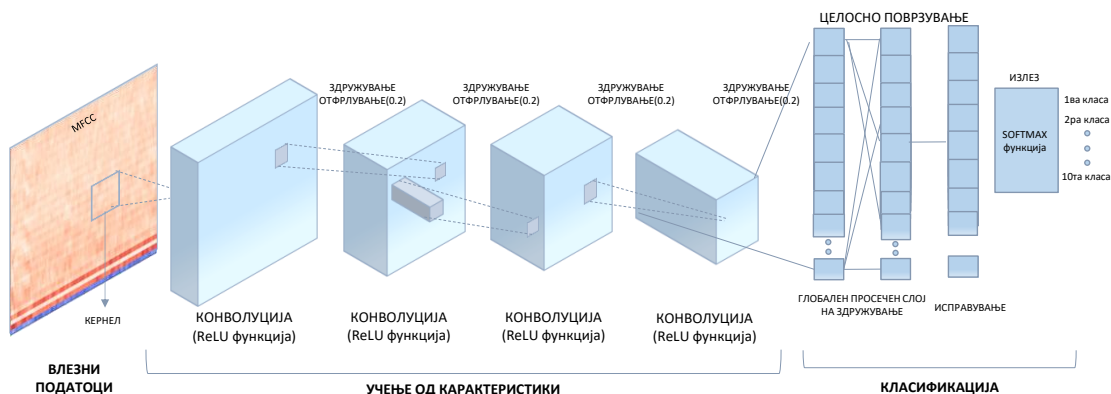
- **Број на епохи (epochs).** Целокупен број на поминување на целата база на податоци низ изградената мрежа на конволуциски невронски мрежи.
- **Број на серии (batches).** Број на примероци кои се процесираат пред да се обноват перформансите на моделот. Бројот на серии мора да биде поголем или еднаков на 1, а помал од целокупниот број на податоци кои се користат за тренирање.

Доколку базата на податоци има x број на примероци, се избира n број на епохи и m број на серии. Ова значи дека целата база на податоци ќе биде поделена на x/m број на серии, а секоја серија ќе има по m број на податоци. Тежинските криви на моделот ќе се обновуваат после секоја серија од m број на податоци, од каде што може да се заклучи дека една епоха во себе ќе има x/m број на серии, односно толку пати ќе се обнови самиот модел. За n број на епохи, целата база на податоци ќе помине низ моделот на конволуциски невронски мрежи n пати во целиот процес на тренирање.

5.6. КОНВОЛУЦИСКИ НЕВРОНСКИ МРЕЖИ ПРИ ПРИМЕНА НА МЕЛ ФРЕКВЕНТНИ ЦЕПСТРАЛНИ КОЕФИЦИЕНТИ

За целите на ова истражување се искористи алгоритам на конволуциски невронски мрежи кој се користи при употреба на МЕЛ Фреквентни Цепстрални Коефициенти чија архитектура визуелно е прикажана на слика 5.6, со детално објаснување на табела 5.3. Се користат 4 конволуциски слоеви, а бројот на филтри варира од 16 до 128 филтри, почнувајќи од помал број на филтри за првиот слој, па се до 128 филтри за последниот конволуциски слој. Согласно истражувањата, бројот на филтри треба да се зголемува при примена на повеќе конволуциски слоеви. Големината на кернелот е 2×2 и се користи ReLU активациска функција во сите филтри. Поради големиот број на податоци, се применуваат функциите на максимално поврзување и отфрлување со цел намалување на податоците. После секоја конволуциска мрежа, коефициентот на отфрлување на податоците е 0,2. По конволуцијата, се применува глобален просечен слој на поврзување и исправување, по што се применува слојот за финална класификација кој користи SOFTMAX функција.

Во конволуциските мрежи и последниот слој на класификација искористени се 44 602 параметри за тренирање. Димензиите на влезните податоци се (40,174), каде што 40 е висината, односно бројот на коефициенти, а 174 е висината на рамките.



Слика 5.6. Визуелен приказ на CNN мрежата при примена на MFCC

Табела 5.3. Архитектура на мрежата на алгоритмот на конволуциски невронски мрежи

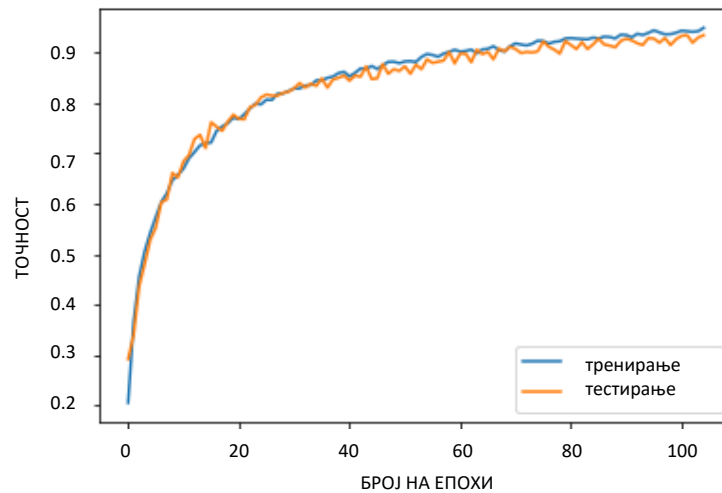
ВИД НА СЛОЈ	Облик на излез	Параметри
Convolution2D (16 filters, 2x2 kernel size, ReLU activation unit)	(None,39,173,16)	80
MaxPooling	(None,19,86,16)	0
Spatial Dropout (0,2)	(None,19,86,16)	0
Convolution2D (32 filters, 2x2 kernel size, ReLU activation unit)	(None,18,85,32)	2080
MaxPooling	(None,9,42,32)	0
Spatial Dropout (0,2)	(None,9,42,32)	0
Convolution2D (64 filters, 2x2 kernel size, ReLU activation unit)	(None,8,41,64)	8256
MaxPooling	(None,4,20,64)	0
Spatial Dropout (0,2)	(None,4,20,64)	0
Convolution2D (128 filters, 2x2 kernel size, ReLU activation unit)	(None,3,19,128)	32896
MaxPooling	(None,1,9,128)	0
Spatial Dropout (0,2)	(None,1,9,128)	0
GlobalAveragePooling2D	(None,128)	0
Flatten	(None,128)	0
Dense 10 units (SOFTMAX output)	(None,10)	1290
ВКУПЕН БРОЈ НА ПАРАМЕТРИ ПРИ ТРЕНИРАЊЕ: 44 602		

Искористена е ADAM оптимизација со зададените вредности на коефициентите: alpha – 0,00001, beta1 – 0,99, beta2 – 0,999 и epsilon – 0,000000001.

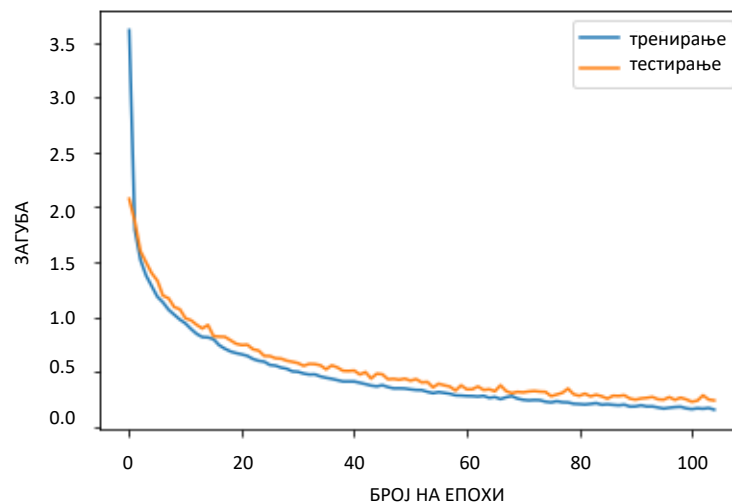
По тренирање и тестирање на системот, може да се заклучи дека при тренирање е постигната точност од 97,95%, додека при тестирање 92,68%. На дијаграмите прикажани на слика 5.7 и 5.8 може да се забележи точноста и загубата на моделот согласно епохите. Искористени се 105 епохи, бидејќи со зголемување на бројот на епохи, доаѓа до појава на презаситување на моделот. Бројот на серии е 256. Од кривите прикажани на слика 5.7, може да се забележи дека точноста која се постигнува при тестирање е повисока од точноста при тренирање, а тоа е нормално поради тоа што при тренирање, не се користат сите карактеристики (согласно процентот што ќе се зададе со функцијата на отфрлање), додека при тестирање се користат сите карактеристики. Истиот случај се јавува и при анализа на загубата

прикажана на слика 5.8, од каде што може да се забележи дека загубата при тестирање е помала од загубата при тренирање на податоците. Со користење на функцијата каде што се отфрлаат дел од карактеристиките на моделот, се подобрува неговата точност и се спречува појавата на презаситување.

Кривите одат по блага патека и нема нивно препоклопување, што е добар знак при дизајнирање на архитектурата на конволуциски невронски мрежи. Повисоките слоеви додаваат фактор на регулација што овозможува посигурна мрежа. Грешката при тестирање е помала од грешката при тренирање на моделот, а оваа појава е нормална бидејќи за време на тренирањето грешката се пресметува додека моделот користи отстапување што доведува до појава на потешкотии и поголема загуба.



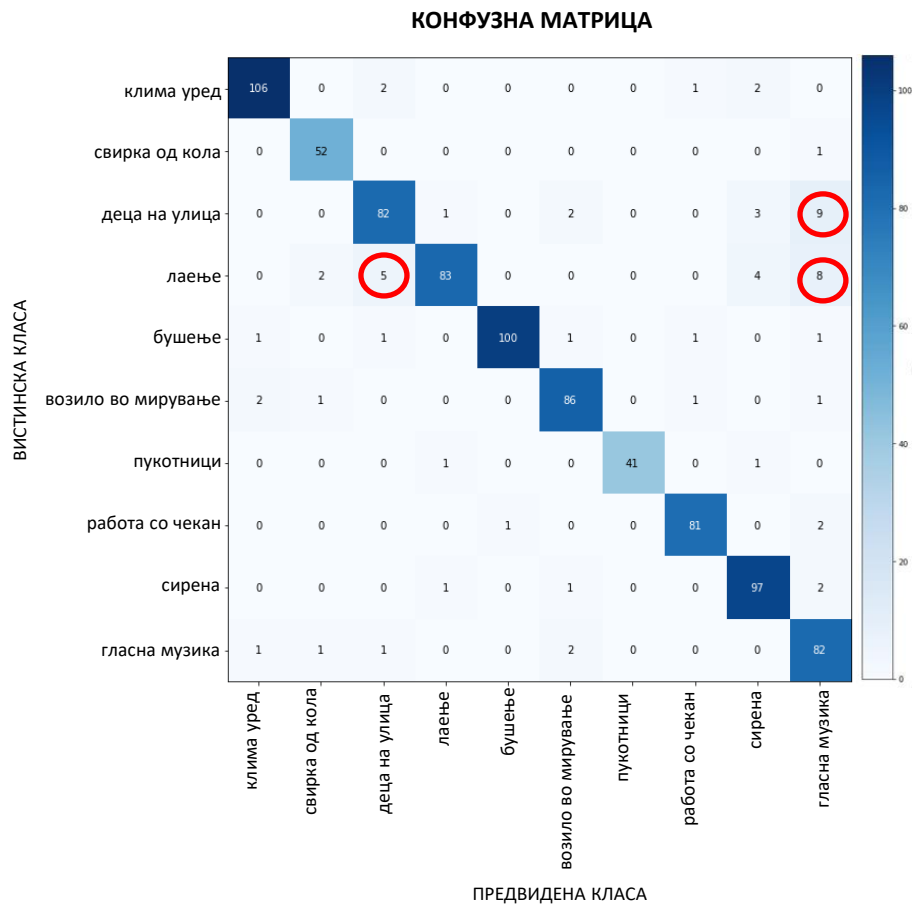
Слика 5.7. Постигната точност при класификација на податоците низ слоевите на мрежата на CNN



Слика 5.8. Загуба при класификација на податоците низ слоевите на мрежата на CNN

На слика 5.19 е прикажана конфузната матрица од тестираните резултати, а табела 5.4. процентуално е прикажана точноста за секоја класа одделно. Согласно анализата, може да се

забележи дека најголемата грешка се појавува кај звучниот настан деца што играат на улица (84,53%) и лаење на куче (81,37%), каде што најголемата конфузија е направена со звучниот настан гласна музика. Најпрецизно предвиден звучен настан е свирката од кола со точност од 98,11%. Искористениот модел изработен во програмот Python е прикажан во прилог B-1.



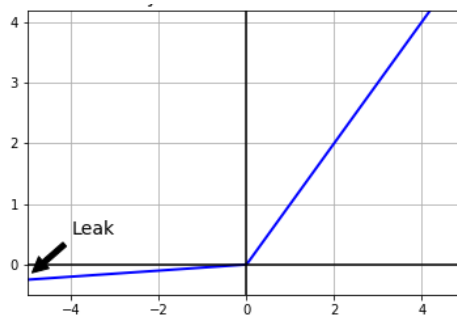
Слика 5.9. Конфузна матрица при примена на модел на CNN со МЕЛ фреквентни цепстрални коефициенти

Табела 5.4. Постигната точност при класификација на одделните класи

КЛАСА	ТОЧНОСТ
Свирка од кола	98,11%
Работа со чекан	96,42%
Сирена	96,04%
Клима уред	95,49%
Пукотници	95,34%
Бушење	95,24%
Работа на возило во мирување	94,51%
Гласна музика	94,25%
Деца на улица	84,54%
Лаење	81,37%

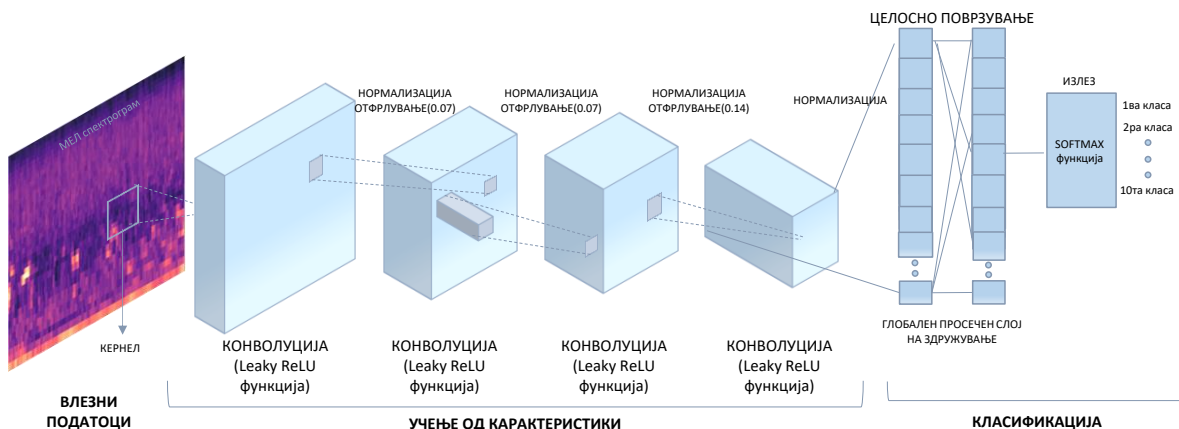
5.7. КОНВОЛУЦИСКИ НЕВРОНСКИ МРЕЖИ ПРИ ПРИМЕНА НА МЕЛ СПЕКТРОГРАМ

При изградба на мрежата на конволуциски невронски мрежи при примена на МЕЛ спектрограм за параметризација на звучните настани, започнато е со наједноставниот модел, еден скриен конволуциски слој без регулација и без целосно поврзани слоеви. Поради појава на презаситување, применета е покомплексна архитектура составена од четири конволуциски блокови, намалување на примерокот и регуларизација. Активациската функција која се користи во конволуциските слоеви е ReLU активациска функција со препуст (Leaky ReLU), функција која се базира на ReLU, но има мал наклон кој води до појава на негативни вредности. Функцијата за стандардната ReLU активациска функција е $R(x) = \max(0, x)$, додека за ReLU активациската функција со препуст е $R(x) = \max(0, 0.1x, x)$. Од тука може да се констатира дека ако се користи стандардна функција, доколку се појави негативна вредност, нема да се примени учење, додека функцијата со препуст дозволува и негативните вредности да влијаат на учењето на конволуцискиот слој. На слика 5.10 е прикажана функцијата на ReLU активациската функција со препуст.



Слика 5.10. ReLU активациска функција со препуст

Првиот конволуциски слој учи од карактеристиките кои се на пониско ниво. Вториот конволуциски слој ја намалува големината на податоците, додека третиот и четвртиот конволуциски слој учат од карактеристиките од целиот аудио сигнал (големината на филтерот е 3x3). Комбинирањето на конволуциските слоеви со примена на максимално поврзување е најкористена техника во CNN архитектурата и истата се применува за овој систем. Функција на регулација во облик на отфрлување на податоците е додадена поради појава на презаситување на моделот, а служи за да ги отфрли целосните филтрирања и да овозможи ажурирање на сите слоеви. Со зголемување на капацитетот на слојот се зголемува и стапката за просторно отфрлање. Дополнително е додадена регулација со мала вредност која служи за воведување на ограничување на коефициентите на тежинските криви на тој начин што ниту една тежинска крива нема да може да се ажурира со вредност надвор од ова ограничување. На слика 5.11 е визуелно прикажан моделот на конволуциски невронски мрежи, додека на табела 5.5 е подетално прикажан изградениот алгоритам.



Слика 5.11. Визуелен приказ на CNN мрежата при примена на МЕЛ спектрограм

Вкупниот број на параметри при тренирање со оваа мрежа е 66 410, од кои вкупен број на параметри кои се користат при тренирање се 65 642, а служат како влез во конволуциските мрежи и последниот слој на класификација.

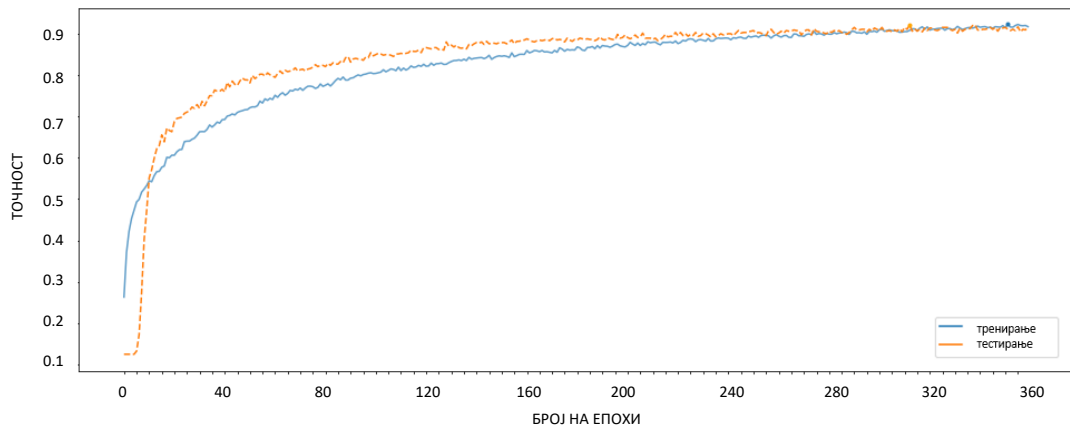
Табела 5.5. Архитектура на мрежата на алгоритмот на конволуциски невронски мрежи

ВИД НА СЛОЈ	Облик на излез	Параметри
Convolution2D (32 filters, 3x3 kernel size, Leaky ReLU activation unit)	(None,38,172,32)	320
Batch Normalization	(None,38,172,32)	128
Spatial Dropout (0,07)	(None,38,172,32)	0
Convolution2D (32 filters, 3x3 kernel size, Leaky ReLU activation unit)	(None,36,170,32)	9248
Batch Normalization	(None,36,170,32)	128
MaxPooling	(None,18,85,32)	0
Spatial Dropout (0,07)	(None,18,85,32)	0
Convolution2D (64 filters, 3x3 kernel size, Leaky ReLU activation unit)	(None,16,83,64)	18496
Batch Normalization	(None,16,83,64)	256
Spatial Dropout (0,14)	(None,16,83,64)	0
Convolution2D (64 filters, 3x3 kernel size, Leaky ReLU activation unit)	(None,14,81,64)	36928
Batch Normalization	(None,14,81,64)	256
GlobalAveragePooling2D	(None,64)	0
Dense 10 units (SOFTMAX output)	(None,10)	650
ВКУПЕН БРОЈ НА ПАРАМЕТРИ ПРИ ТРЕНИРАЊЕ: 66 410		

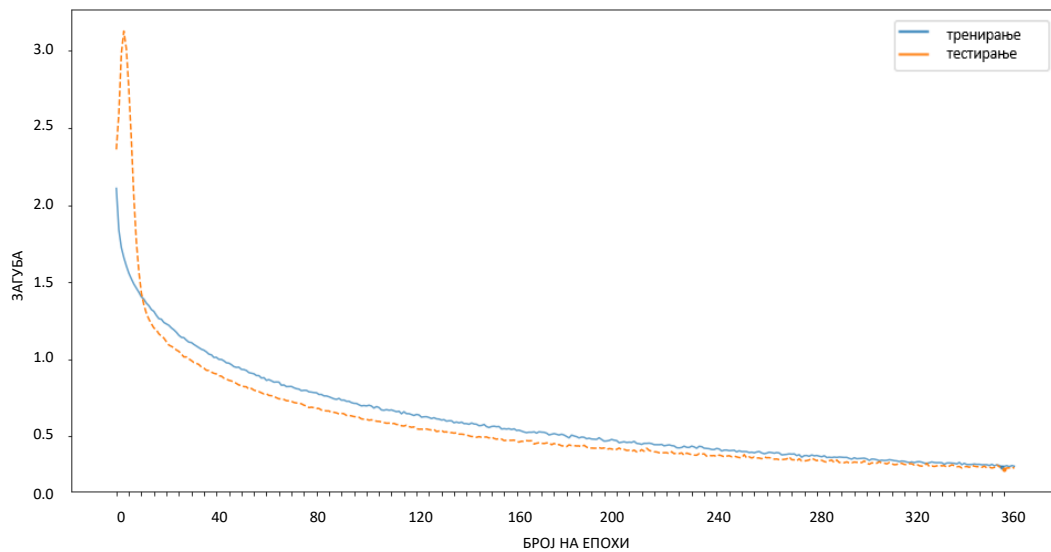
При изградба на конволуциски невронски мрежи за целите на ова истражување, за оптимизацијата се искористени следните коефициенти: alpha – 0,00001, beta1 – 0,99, beta2 – 0,999 и epsilon – 0,000000001. Изградениот алгоритам во програмот Python за овој модел е прикажан во прилог B-2.

Примената на МЕЛ спектрограм во традиционалните алгоритми на машинско учење не покажа добри резултати при препознавање на звучните настани, но согласно истражувањата,

се смета дека овој параметар постигнува висока точност при примена во алгоритми засновани на длабоко учење. При примена на МЕЛ спектрограмот во конволуциски невронски мрежи, може да се забележи висока точност, односно при тренирање се јавува точност од 95,1%, додека при тестирање на податоците се постигнува точност од 91,98%. При тренирање на податоците искористени се 360 епохи, а бројот на серии е 128. На слика 5.12 и 5.13 се прикажани дијаграмите на точност и загуба во текот на поминување низ разните слоеви на CNN.



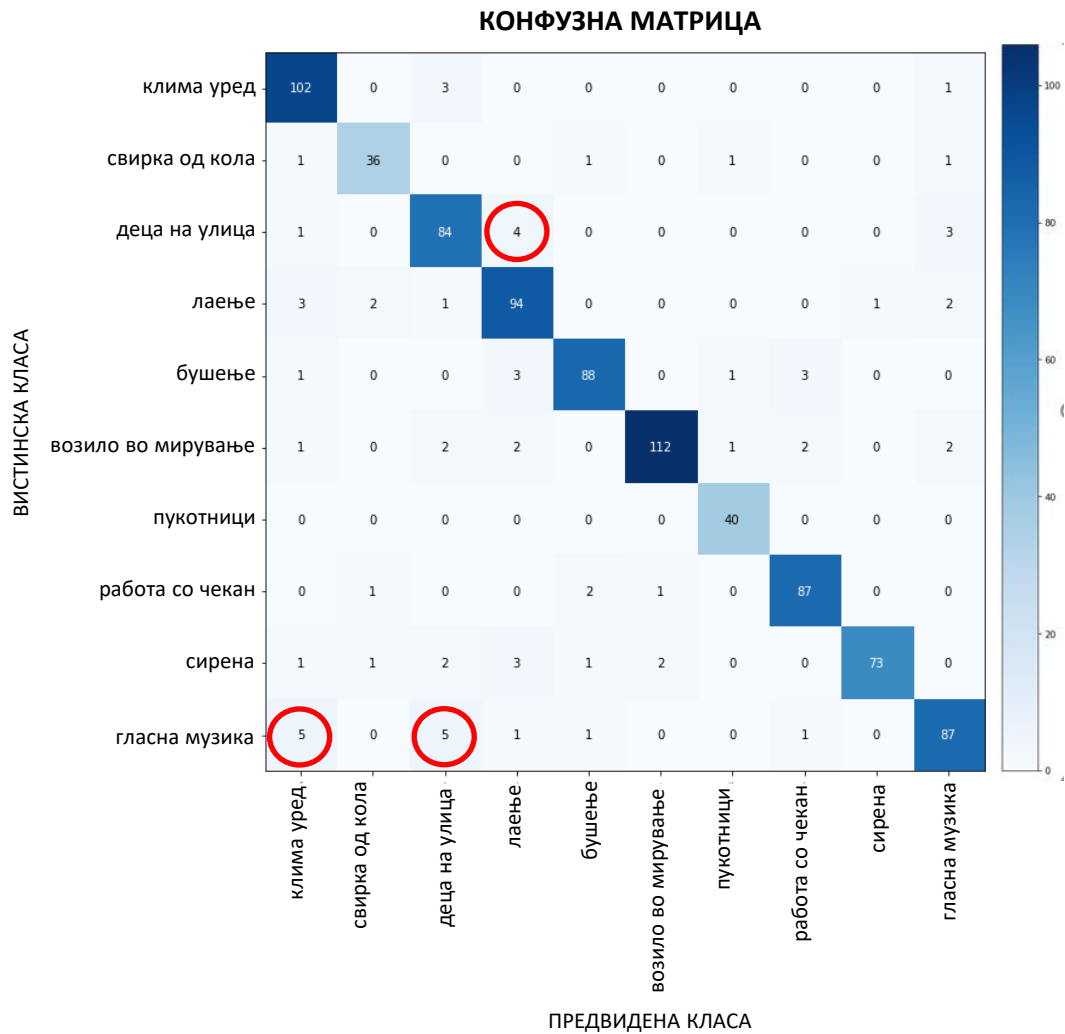
Слика 5.12. Постигната точност при класификација на податоците низ слоевите на мрежата на CNN



Слика 5.13. Загуба при класификација на податоците низ слоевите на мрежата на CNN

На сликата 5.14 е прикажана конфузната матрица помеѓу класите на звучните податоци на тренираните и тестираните податоци, додека на табела 5.6 процентуално е прикажана на постигнатата точност одделно за сите класи на звучни податоци. Од анализата на конфузната матрица и на постигнатата точност на одделните класи, може да се забележи дека звучниот настан што означува пукотници нема ниту еден погрешно предвиден податок (100% точност), додека најмала точност се јавува кај звучниот настан на гласна музика (87%) и сирена (87,95%). Од конфузната матрица може да се забележи дека најголема грешка се јавува кај звучниот

настан гласна музика, кој е помешан со клима уред и деца што играат, како и деца што играат со лаење.



Слика 5.14. Конфузна матрица при примена на модел на CNN со МЕЛ спектрограм

Табела 5.6. Постигната точност при класификација на одделните класи

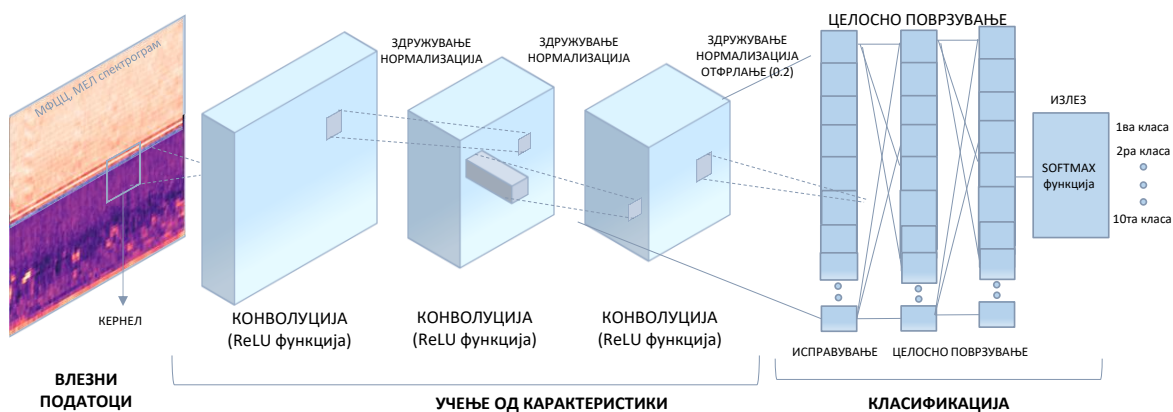
КЛАСА	ТОЧНОСТ
Пукотници	100,00%
Клима уред	96,25%
Работа со чекан	95,60%
Работа на возило во мирување	91,80%
Бушење	91,67%
Деца на улица	91,30%
Лаење	91,26%
Свирка од кола	90,00%
Сирена	87,95%
Музика	87,00%

5.8. КОНВОЛУЦИСКИ НЕВРОНСКИ МРЕЖИ ПРИ ПРИМЕНА НА MFCC И МЕЛ СПЕКТРОГРАМ

При примена на конволуциските невронски мрежи како алгоритам на вештачка интелигенција кој како влезни параметри ги користи податоците преработени со MFCC и МЕЛ спектрограм се користат три конволуциски слоеви и два слоеви на целосно поврзување. На слика 5.15 и табела 5.7 е прикажана архитектурата која е искористена при примена на параметрите на MFCC и МЕЛ спектрограм. При примена на конволуциските слоеви искористена е ReLU активациска функција, а бројот на применети филтри е 32,64 и 128 соодветно за секој слој. Во сите три слоја применета е функција на максимално поврзување и нормализација, додека во третиот слој дополнително е применето отфрлување од 20% на податоците за да се спречи појава на презаситување на моделот. По трите конволуциски слоеви, се применува слој за исправување на карактеристиките, по што следуваат последните два слоја на целосно поврзување. Првиот слој користи 64 филтри и ReLU активациска функција, каде што е применето дополнително отфрлување од 30% од податоците за тренирање, додека вториот слој е воедно и последниот слој кој овозможува класификација на податоците и ја користи SOFTMAX активациската функција.

Искористени се 83 882 параметри, од кои за тренирање се искористени 83 434 параметри. Влезната големина на векторот на карактеристики е [40,2], каде што 40 е бројот на коефициенти што е извлечен за секоја аудио карактеристика, а 2 го означува бројот на параметри на звучни настани применети за креирање на овој вектор. Со зголемување на бројот на епохи се појавува презаситување на моделот, па поради тоа се користат само 60 епохи. Бројот на серии е 128.

При изградба на конволуциски невронски мрежи за целите на ова истражување, за оптимизацијата се искористени следните коефициенти: $\alpha = 0,00001$, $\beta_1 = 0,99$, $\beta_2 = 0,999$ и $\epsilon = 0,00001$.

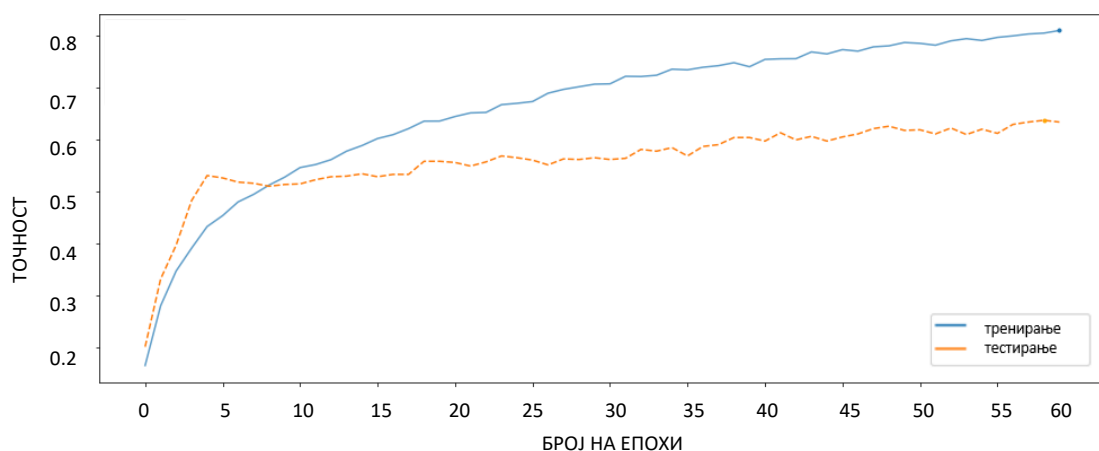


Слика 5.15. Визуелен приказ на CNN мрежата при примена на MFCC и МЕЛ спектрограм

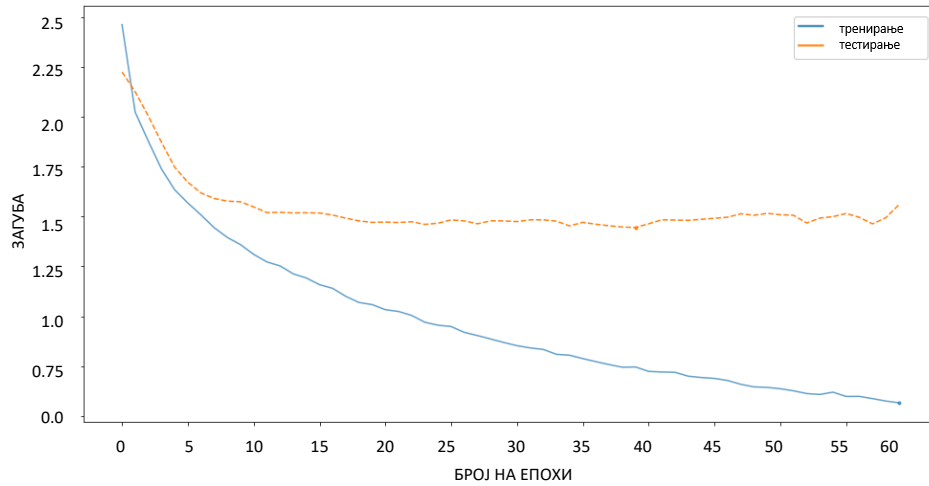
Табела 5.7 Архитектура на мрежата на алгоритмот на конволуциски невронски мрежи

ВИД НА СЛОЈ	Облик на излез	Параметри
Convolution2D (32 filters, 2x2 kernel size, ReLU activation unit)	(None,40,2,32)	160
Max pooling	(None,20,1,32)	0
Batch Normalization	(None,20,1,32)	128
Convolution2D (64 filters, 2x2 kernel size, ReLU activation unit)	(None,20,1,64)	8256
MaxPooling	(None,10,1,64)	0
Batch Normalization	(None,10,1,64)	256
Convolution2D (128 filters, 2x2 kernel size, ReLU activation unit)	(None,10,1,128)	32896
MaxPooling	(None,5,1,128)	0
Spatial Dropout (0,2)	(None,5,1,128)	0
Batch Normalization	(None,5,1,128)	512
Flatten	(None,640)	0
Dense (64 filters, ReLU activation unit)	(None,64)	41024
Spatial Dropout (0,3)	(None,64)	0
Dense 10 units (SOFTMAX output)	(None,10)	650
ВКУПЕН БРОЈ НА ПАРАМЕТРИ ПРИ ТРЕНИРАЊЕ: 83 882		

Изградениот алгоритам во програмот Python за овој модел е прикажан во прилог В-3. Со комбинирање на параметрите МЕЛ Фреквентни Цепстрални коефициенти и МЕЛ спектрограм се добива точност од 89,21% при тренирање, додека 63.34% при тестирање на системот. На слика 5.16 и 5.17 се прикажани кривите на точност и загуба при тренирање и тестирање на моделот. Може да се забележи дека со зголемување на бројот на епохи доаѓа до зголемување на растојанието помеѓу кривата на тренирање и тестирање, односно при тренирањето точноста се зголемува, но при тестирањето кривата не бележи раст, што е показател за појава на презаситување. Поради тоа, за овој модел се искористени само 60 епохи. Истата појава се повторува и кај кривите на загуба (слика 5.17).

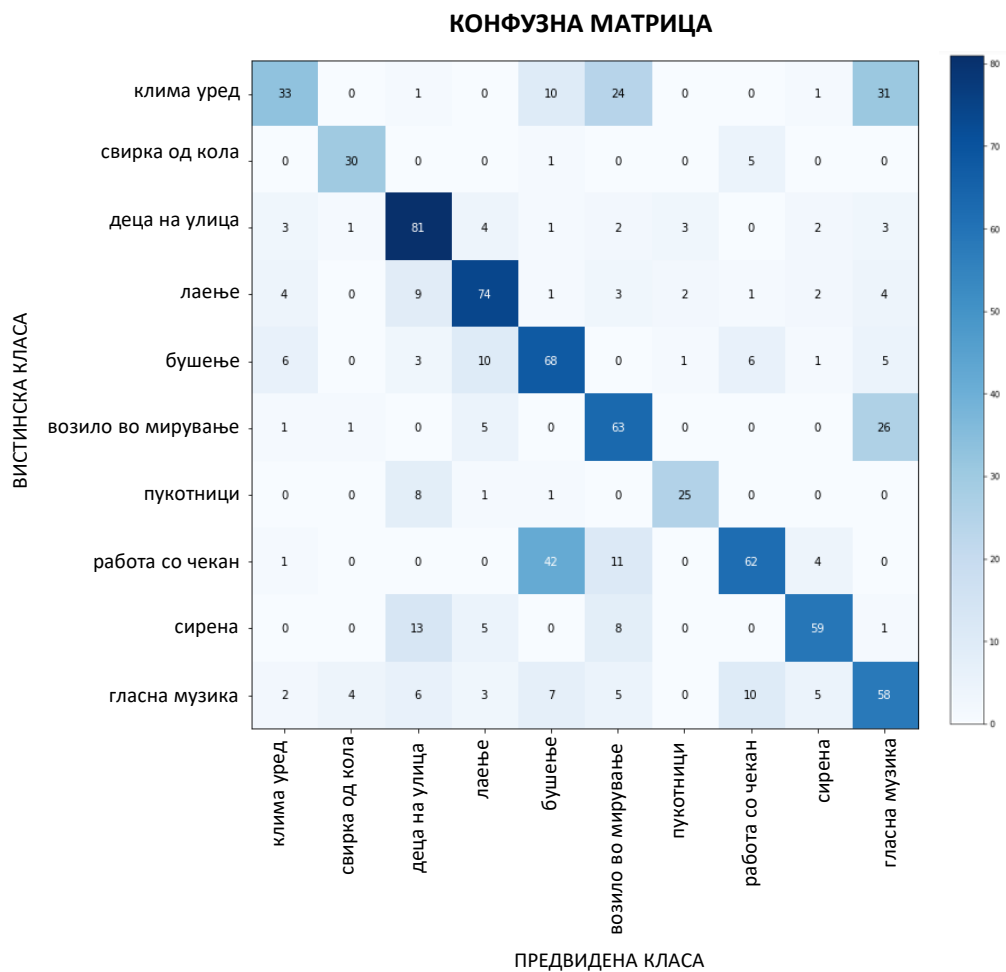


Слика 5.16. Постигната точност при класификација на податоците низ слоевите на мрежата на CNN



Слика 5.17. Загуба при класификација на податоците низ слевите на мрежата на CNN

На слика 5.18 е прикажана конфузната матрица помеѓу вистинските и предвидените класи при тестирање на моделот, а табелата 5.8 ја прикажува точноста за секоја одделна класа.



Слика 5.18. Конфузна матрица при примена на модел на CNN со MFCC и МЕЛ спектрограм

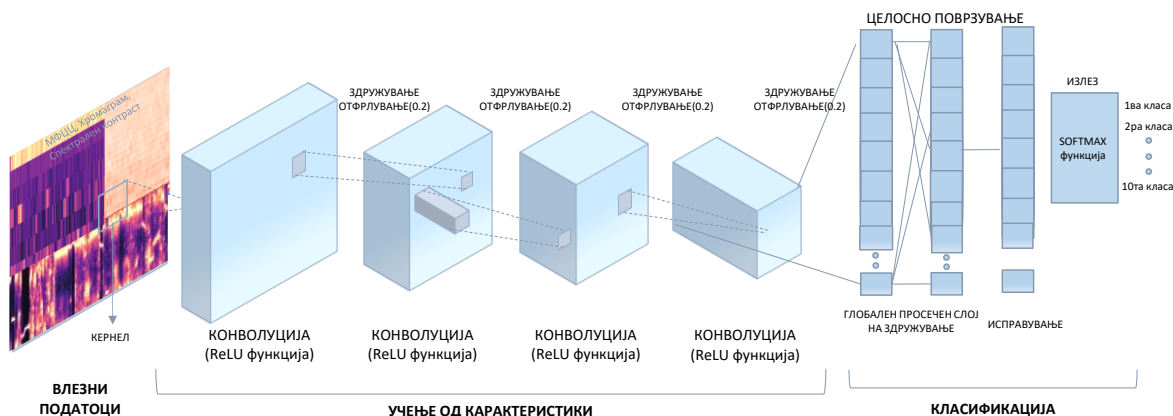
Табела 5.8. Постигната точност при класификација на одделните класи

КЛАСА	ТОЧНОСТ
Свирка од кола	83,33%
Деца на улица	81,00%
Лаење на куче	74,00%
Пукотници	71,43%
Сирена	68,60%
Бушење	68,00%
Работа на возило во мирување	65,63%
Гласна музика	58,00%
Работа со чекан	51,67%
Клима уред	33,00%

5.9. КОНВОЛУЦИСКИ НЕВРОНСКИ МРЕЖИ ПРИ ПРИМЕНА НА MFCC, ХРОМАТОГРАМ И СПЕКТРАЛЕН КОНТРАСТ

Архитектурата на применетиот алгоритам е прикажана на слика 5.19, додека на табелата 5.9 подетално се објаснети применетите функции и слоеви. Искористени се четири слоеви на конволуциски невронски мрежи кои ја користат ReLU активациска функција. Во секој конволуциски слој е применета функцијата на максимално поврзување и слој на отфрлување од каде што се отфрлени по 20% од податоците за да се спречи појава на презаситување. Првиот конволуциски слој користи 16 филтри, вториот 32, третиот 64 и последниот слој 128 филтри, што овозможуваат постепено тренирање и учење од податоците. По четвртиот конволуциски слој, применета е функција за глобално порамнување и исправување на податоците, по што е применета мрежата за класификација која ја користи SOFTMAX активациската функција.

Искористени е 44 602 параметри при тренирање на мрежата, а обликот на влез на векторот на карактеристики е (7,3), каде што 7 го претставува бројот на коефициенти извлечени од секој параметар, додека 3 го означува вкупниот број на параметри на звучни настани кои се применети за креирање на векторот на карактеристики. Алгоритмот кој што се моделираше во програмот Python е прикажан во прилог B-4.



Слика 5.19. Визуелен приказ на CNN мрежата при примена на MFCC, хроматограм и спектрален контраст

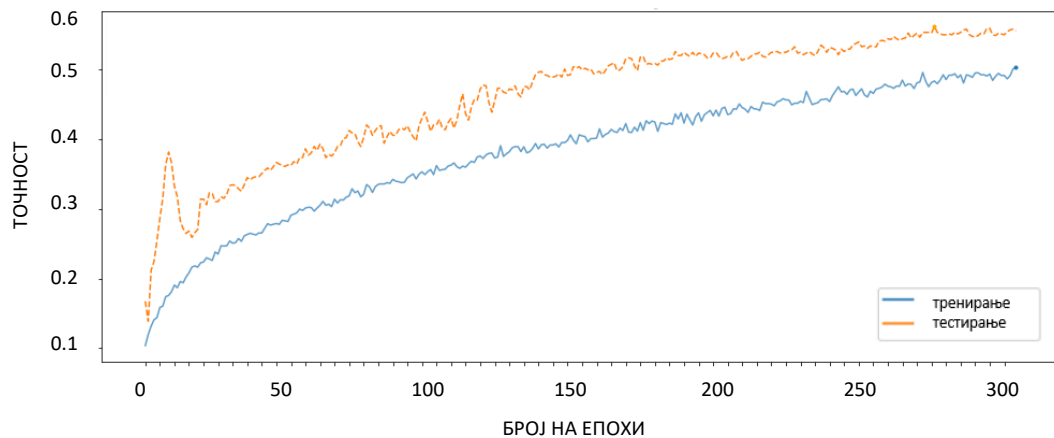
Табела 5.9. Архитектура на мрежата на алгоритмот на конволуциски невронски мрежи

ВИД НА СЛОЈ	Облик на излез	Параметри
Convolution2D (16 filters, 2x2 kernel size, ReLU activation unit)	(None,7,3,16)	80
MaxPooling	(None,4,2,16)	0
Spatial Dropout (0,2)	(None,4,2,16)	0
Convolution2D (32 filters, 2x2 kernel size, ReLU activation unit)	(None,4,2,32)	2080
MaxPooling	(None,2,1,32)	0
Spatial Dropout (0,2)	(None,2,1,32)	0
Convolution2D (64 filters, 2x2 kernel size, ReLU activation unit)	(None,2,1,64)	8256
MaxPooling	(None,1,1,64)	0
Spatial Dropout (0,2)	(None,1,1,64)	0
Convolution2D (128 filters, 2x2 kernel size, ReLU activation unit)	(None,1,1,128)	32896
MaxPooling	(None,1,1,128)	0
Spatial Dropout (0,2)	(None,1,1,128)	0
GlobalAveragePooling2D	(None,128)	0
Flatten	(None,128)	0
Dense 10 units (SOFTMAX output)	(None,10)	1290
ВКУПЕН БРОЈ НА ПАРАМЕТРИ ПРИ ТРЕНИРАЊЕ: 44 602		

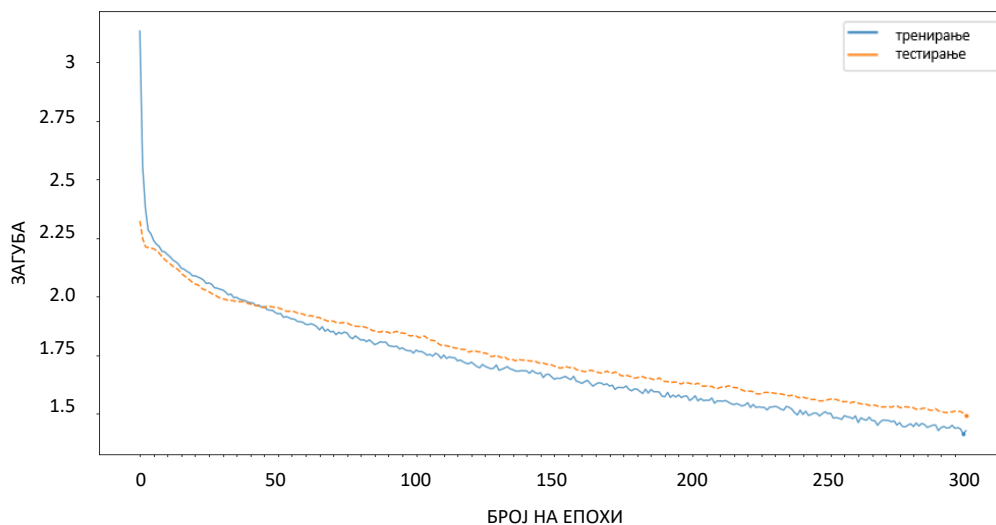
При изградба на конволуциски невронски мрежи за целите на ова истражување, за ADAM оптимизацијата се искористени следните коефициенти: alpha – 0,00001, beta1 – 0,99, beta2 – 0,999 и epsilon – 0,000000001.

При примена на МЕЛ Фреквентни Цепстрални Коефициенти, Хроматограм и Спектрален Контраст, точноста која што се постигнува согласно евалуацијата на генералниот модел е 60,57% при тренирање, додека 55,56% при тестирање на моделот. На дијаграмите прикажани на слика 5.20 и 5.21 може да се забележи точноста и загубата на моделот во текот на поминување низ разните слоеви од кои е составена мрежата на конволуциски невронски мрежи. При тренирање на податоците искористени се 300 епохи, а бројот на серии е 128.

Согласно кривите кои се прикажани на сликите, можеме да потврдиме дека не се појавува презаситување во моделот. Од кривите прикажани на слика 5.20, може да се забележи дека точноста која се постигнува при тестирање е повисока од точноста при тренирање, а тоа е нормално поради тоа што при тренирање, не се користат сите карактеристики (согласно процентот што ќе се зададе со функцијата на отфрлување), додека при тестирање се користат сите карактеристики.

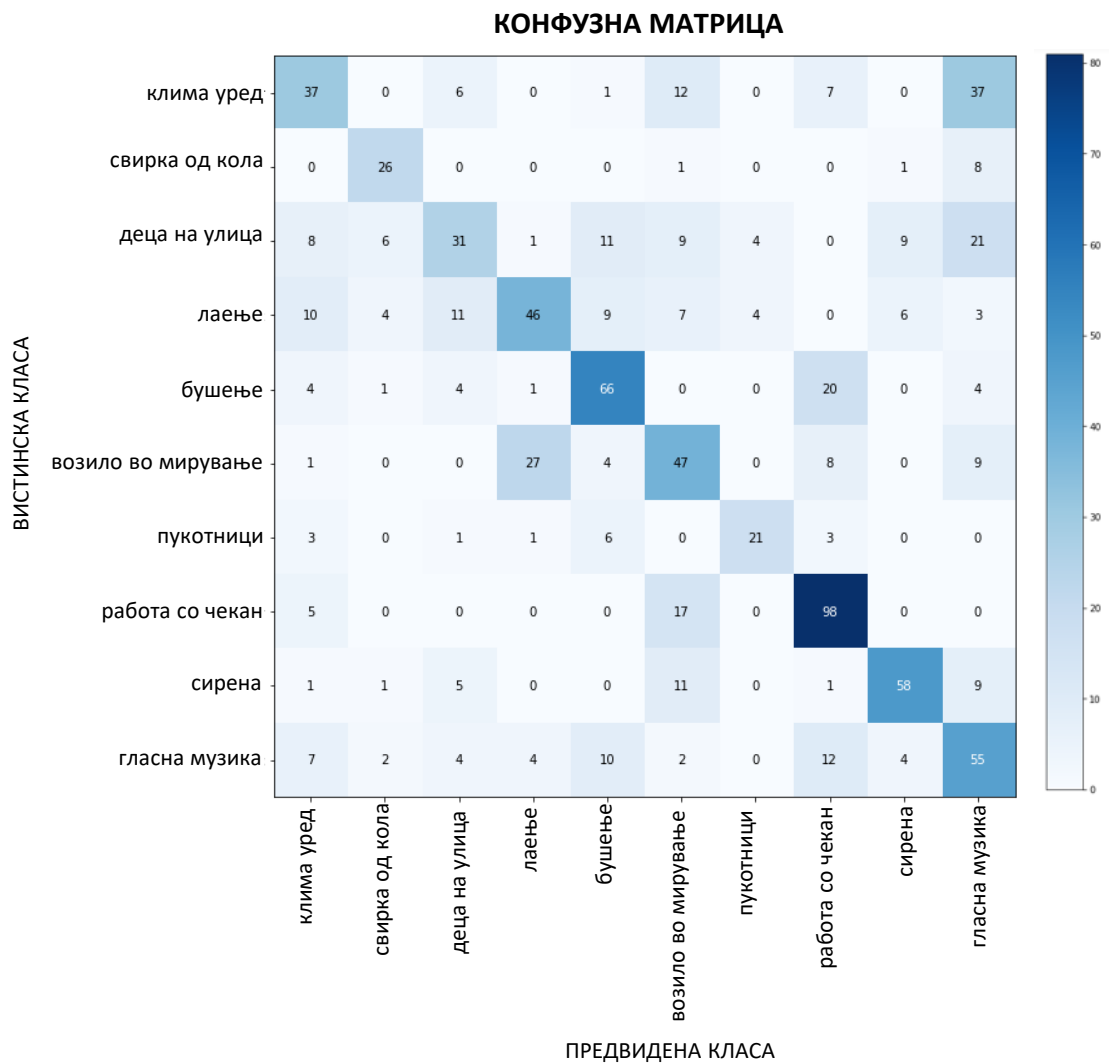


Слика 5.20. Постигната точноста при класификација на податоците низ слоевите на мрежата на CNN



Слика 5.21. Загуба при класификација на податоците низ слоевите на мрежата на CNN

Од резултатите прикажани на конфузната матрица на слика 5.22 и табелата 5.10 може да се забележи дека се појавуваат доста грешки при препознавање на одделните класи поради малата точноста која е добиена при тренирање и тестирање на моделот. Иако кривите на точноста и загубата имаат добар тек, сепак резултатите кои што се постигнуваат се ниски во споредба со останатите резултати при примена на останатите модели на конволуциски невронски мрежи.



Слика 5.22. Конфузна матрица при примена на модел на CNN со MFCC, хроматограм и спектрален контраст

Табела 5.10. Постигната точност при класификација на одделните класи

КЛАСА	ТОЧНОСТ
Работа со чекан	81,67%
Свирка од кола	72,22%
Сирена	67,44%
Бушење	66,00%
Пукотници	60,00%
Гласна музика	55,00%
Работа на возило во мирување	48,96%
Лаење	46,00%
Клима уред	37,00%
Деца на улица	31,00%

5.10. АНАЛИЗА НА ПРИМЕНЕТИТЕ АЛГОРИТМИ НА ВЕШТАЧКА ИНТЕЛИГЕНЦИЈА

При анализа на резултатите, ќе се споредат резултатите при тестирање на најуспешните алгоритми во глава 4 и глава 5, кои се прикажани на табела 5.11. Може да се заклучи дека 4 од вкупно 7 модели постигнуваат точност над 90%. Примената вектор на карактеристики составен од МЕЛ Фреквентните Цепстрални Коефициенти, Хроматограм и Спектрален Контраст во алгоритмот на Машини со Потпирачки Вектори постигнува највисока точност од 92,9%, додека примената само на МЕЛ фреквентни цепстрални коефициенти во мрежата на Конволуциски Невронски Мрежи постигнува точност од 92,68% при тестирање на системот.

Табела 5.11. Резултати од точноста при тестирање на применетите модели

Алгоритам на машинско учење	Параметар на звучен настан	коэф	Постигната точност
Случајни шуми	MFCC + хроматограм + спектрален контраст	59	91,53%
Машини со потпирачки вектори	MFCC + хроматограм + спектрален контраст	59	92,90%
Наивен Баесов алгоритам	MFCC + хроматограм + спектрален контраст	59	53,68%
Конволуциски невронски мрежи	MFCC	40	92,68%
Конволуциски невронски мрежи	МЕЛ спектрограм	40	91,98%
Конволуциски невронски мрежи	MFCC + МЕЛ спектрограм	40	63,34%
Конволуциски невронски мрежи	MFCC + хромограм + спектрален контраст	21	55,56%

Дополнително, направена е анализа на податоците каде што се јавуваат најмногу грешки, каде целта е преку слушање на звучните настани помеѓу помешаните класи да се види дали има некакво совпаѓање на звучниот настан.

При анализа на звучниот настан гласна музика и негова анализа дали се совпаѓа со деца што играат, може да се забележи дека повеќето податоци имаат многу мал шум во позадина, но има многу податоци каде што се слуша зборување (што се појавува и кај звучниот настан деца што играат). Има неколку случаи каде што не може да се оддели музиката во позадина со амбиентниот резидуален звук.

При споредба на звучните настани бушење и работа со чекан може да се забележи дека иако овие случаи се многу слични, има некои клучни разлики како висината на звукот и динамиката на амплитудата. Кај бушењето може да се забележи константен звук и амплитуда, додека кај работата со чекан може да се забележи средно-низок звук со препознатлива карактеристика.

Кај звучниот настан куче што лае и деца што играат, при одделна анализа на звуците може да се забележи дека кај овој звучен настан се слушаат деца што играат во позадина. Во многу случаи звукот на куче што лае е измешан со висок резидуален звук или зборување. Сепак, има случаи каде што се слуша само лаење на куче, но не е правилно препознаен.

При слушање на звучните настани гласна музика и клима уред, може да се забележи дека појавата на шум во самиот звучен настан најверојатно појавува грешка при

класификација. Иако пукотниците е настан кој е помалку застапен во базата на податоци, сепак неговата класификација е прецизна, најверојатно поради импулсивната природа на настанот кој го прави впечатлив и лесен за препознавање. За разлика од него, прецизноста при предвидување на свирката од кола варира зависно од применетиот модел. Работата на возилото во мирување како звучен насан има рамен звук, без забележителни настани.

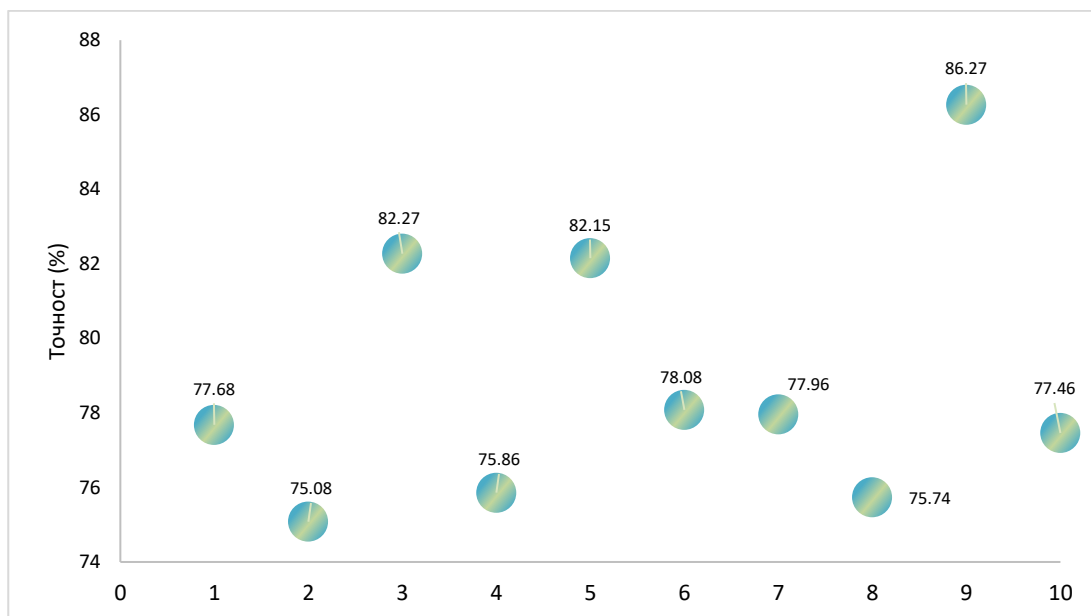
Постојат јасни знаци за лоши перформанси при класификација кога објективниот звук не е доволно гласен и има појава на други извори на звук во аудио податокот. Фактот што ова се случува често во оваа база на податоци го отежнува разликувањето помеѓу некои класи, што можеби е и причина за пониска точност во истражувањата кои ја користат оваа база на податоци.

Моделирањето со помош на техниките на длабокото учење има предност пред класичните алгоритми на машинско учење доколку системот што треба да се моделира е комплексен и треба да се примени во апликации во реално време. Бидејќи разликата помеѓу постигнатата точност на најуспешните модели е многу мала (0,22%), ќе се одбере моделот составен од MFCC како параметар на звучни настани и конволуциските невронски мрежи како алгоритам кој ги користи техниките на вештачката интелигенција за понатамошна примена за валидација преку тестирање со непознати звучни настани и негова апликација во реално базирана на „Интернет на Нештата“ за развој на паметни градови.

5.11. ВАЛИДАЦИЈА НА СИСТЕМОТ ЗА ПРЕПОЗНАВАЊЕ И КЛАСИФИКАЦИЈА НА ЗВУЧНИ НАСТАНИ

Валидација на систем базиран на техниките на вештачка интелигенција е процес на дополнително тестирање на системот, при што се задржуваат само резултатите од тренирање на алгоритмот, а валидацијата, односно дополнителното тестирање се прави со непознати податоци од независна база на податоци. Целта на валидацијата е да се измери перформансот на системот преку добивање на точноста, со тоа што ќе се добијат подобри сознанија за имплементирање на системот во реална апликација. Како најуспешен алгоритам кој постигнува висока точност од 92,68% и има применливост во апликации во реално време е алгоритмот на конволуциски невронски мрежи со користење на МЕЛ фреквентните цепстрални коефициенти. За примена на процесот на валидација, тестирањето беше направено во 2 фази, а како влез се употреби истренираната невронската мрежа. Применетиот алгоритам за валидација на системот за препознавање и класификација на звучните настани е прикажан во прилог В-5.

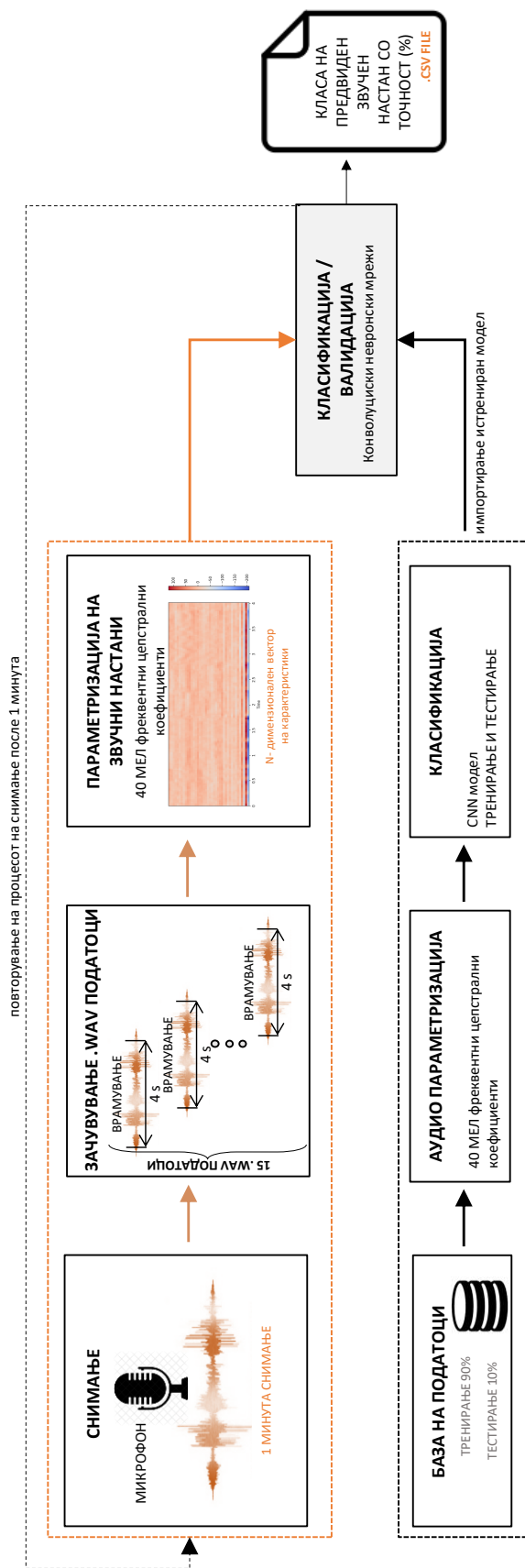
Во првата фаза на валидација, системот се тестираше со снимени податоци од 10те класи на звучни настани, при што овие податоци немаат означено класа на звучен настан. Направени се 10 итерации со цел да се спореди добиената точност на системот при валидација, односно при класификација на звучните настани. За секоја итерација, се употребија 100 снимени податоци на 10те класи на звучни настани, кои не беа подеднакво распоредени. Како резултат се доби точноста при класификација и предвидената класа на звучниот настан. На сликата 5.23 е прикажана точноста за секоја итерација, каде што беа применувани различни податоци и класи на звучни настани. Точноста варира помеѓу 75,86% – 86,27%, од каде што преку примена на методот на валидација преку 10 итерации (10K validation), просечната точност на системот е 78,86%.



Слика 5.23. Резултати од точноста при класификација од валидацијата

Добиените резултати од предвидената класа на податоци понатаму се процесираа со цел споредба со вистинската класа на податоци. Од анализата, може да се забележи дека 80% од податоците кои се грешно предвидени се класифицирани во класата на клима уред. Како класи кои имаат најниска точност и се грешно класифицирани се класите работа со чекан, работа на возило во мирување и бушење. Може да се заклучи дека континуираните извори кои генерираат повторлива бучава и податоците кои не се добро снимени често пати грешно се предвидуваат со класата на работа на клима уред. Со преслушување на податоците, може да се забележи дека грешно предвидени се оние податоци кои што имаат голем шум во нив и имаат дополнителен извор кој е во позадина.

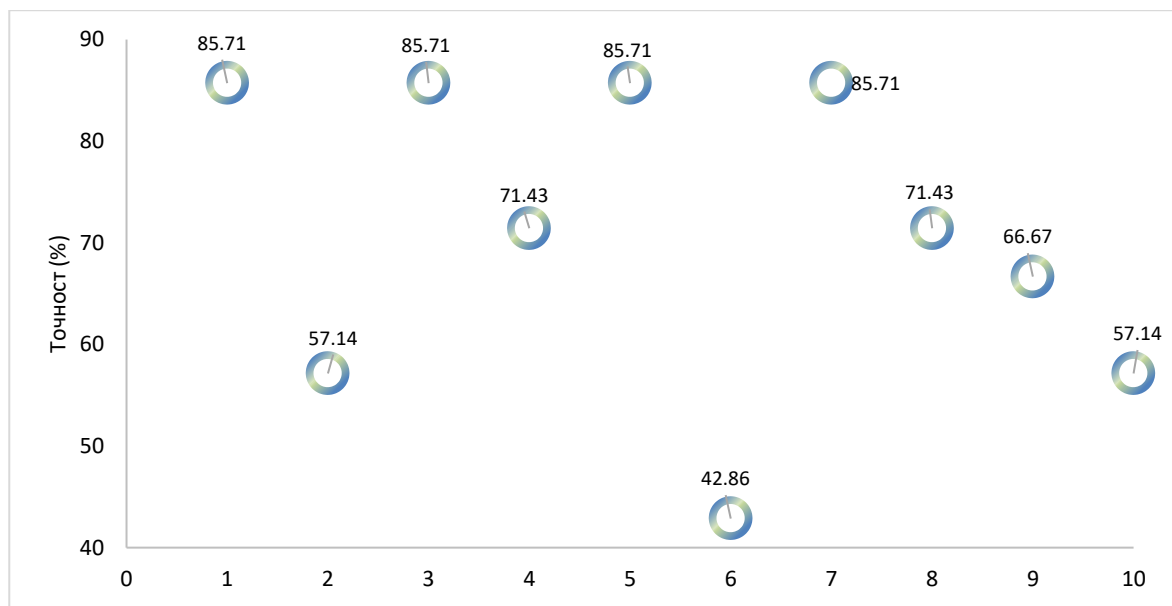
Во втората фаза на валидација, се изгради алгоритам кој автоматски ќе ги снима податоците на звучни настани во текот на 1 минута, во чие времетраење ќе се генерираат 15 податоци на звучни настани, секој податок во времетраење од 4 секунди, а потоа истите ќе се класифицираат. На слика 5.24 визуелно е прикажан алгоритмот кој служи за валидација на снимените податоци.



Слика 5.24. Приказ на алгоритмот за валидација на податоците

Повикувајќи се на базата на податоци која се искористи за тренирање и тестирање на звучните настани, се смета дека 4 секунди е доволно времетраење за да може да се карактеризира една класа на звучен настан. Во наредната минута, алгоритмот ги обработува снимените податоци и ги класифицира, а во тоа време снимањето е на мирување (sleep mode).

Во оваа фаза, повторно беа направени 10 итерации од по едно минутно снимање, каде за секоја итерација се класифицираа по 15 податоци на звучни настани. Од резултатите прикажани на слика 5.25 може да се забележи дека точноста се движи во граници помеѓу 42,86% – 85,71%.



Слика 5.25. Резултати од валидација на звучните настани снимени во континуиран период

Од анализата на предвидените резултати, повторно може да се забележи дека најголем број на предвидени податоци припаѓаат во класата која ја претставува работата на клима уред. При снимањето на звучните настани постојат периоди кога нема звучни настани (само шум од резидуален звук) или класи кои што не постојат во базата на податоци, а се класифицираат како звучен настан клима уред со ниска точност.

6. ЗАКЛУЧОЦИ И ПРЕПОРАКИ ЗА ИДНИ ИСТРАЖУВАЊА

Оваа докторска дисертација претставува прилог во интердисциплинарна научна област која има за цел развој на систем за препознавање и класификација на звучни настани и негова примена во реални инженерски апликации. Истражувањата во оваа докторска дисертација се насочени кон развој на системи за препознавање и класификација на звучни настани преку користење на разни параметри на звучни настани и техники на вештачката интелигенција.

Развојот на стандардните методи за објективна анализа на бучавата преку мерење на нивото на бучава и изработка на предиктивни карти на бучава, како субјективната анализа за проценка на вознемиреноста предизвикана од разните извори на бучава, овозможува проценка на бучавата во урбана животна средина. Стандардните методи дадоа прецизни резултати при квантификација на нивото на бучава, но истите не можат да дадат резултати за доминантните извори кои ја предизвикуваат бучавата, па затоа, преку надградба на овие системи со примена на напредни методи базирани на техниките на вештачката интелигенција се моделираа системи за препознавање и класификација на звучните настани.

Анализата на теоретскиот пристап при моделирање на параметрите на звучни настани овозможи нивна анализа и избор на параметри кои ќе се применуваат за екстракција на аудио сигналите. Од друга страна, анализата на алгоритмите на машинското и длабокото учење овозможи запознавање со математичките концепти кои може да се применат при моделирање на алгоритмите базирани на вештачка интелигенција. Примена на повеќе аудио параметри и алгоритми на машинско и длабоко учење резултираа во експериментална анализа на 52 комбинации со цел избор на сет од параметри на звучни настани и алгоритми на машинско учење кои ќе дадат најпрецизни резултати во процесот на класификација на урбаните звучни настани.

За крај, моделот кој покажа најдобри перформанси беше дополнително валидиран за да може да се добијат сознанија и јасна престава за применливоста и начинот на кој може да се подобрат перформансите со понатамошни истражувања.

6.1. ЗАКЛУЧОЦИ

6.1.1. ЗАКЛУЧОЦИ ОД ПРИМЕНАТА НА СТАНДАРДНИТЕ МЕТОДИ ЗА ПРОЦЕНКА НА БУЧАВАТА

Теоретската анализа на бројните истражувања, стандардите и националните регулативи овозможи креирање на методологија за проценка на бучавата која послужи за експериментална анализа, обработка на резултатите и нивна компарација.

Преку анализа на резултатите од мерењето на нивото на бучава и пресметка на основните и дополнителните индикатори се доби сознание за нивото на бучава во анализираното урбано подрачје. Дополнително, преку развој на алатка за статистичка пресметка на мерната неодреденост, резултатите беа пресметани и прикажани со 95% доза на сигурност. Развојот на методологија за изработка на предиктивни карти на бучава кои го користат протокот на сообраќајот како основен извор на бучава доведе до изработка на акустични предиктивни модели за бучавата, овозможувајќи визуелна репрезентација на нивото на бучава во определеното подрачје.

Компаративната анализа помеѓу резултатите од двете методологии се спроведе користејќи го параметарот за мерна неодреденост, од каде што се потврди точноста на применетите методологии. Анализите покажуваат дека методологиите може да се применат за проценка и квантифицирање на нивото на бучава во урбана животна средина.

Дополнително, статистичкиот научен метод за субјективна анализа овозможи проценка на штетното влијание на бучавата, од каде што се заклучи дека постојат голем број на влијателни извори на бучава кои не можат да се определат со примена на стандардните методи за проценка на бучавата. Од тие причини, се пристапи кон употреба на современи напредни техники во системите за следење на состојбата со бучава преку развој на системи за препознавање и класификација на звучните настани. Истражувањата од субјективната анализа дадоа придонес при креирањето на таксономијата на звучни настани кои предизвикуваат вознемиреност на населението во урбана животна средина, која е основа за креирање база на податоци која ќе се користи во системот за препознавање и класификација на звучните настани.

6.1.2. ЗАКЛУЧОЦИ ОД ПРИМЕНАТА НА НАПРЕДНИТЕ МЕТОДИ ЗА ПРЕПОЗНАВАЊЕ И КЛАСИФИКАЦИЈА НА ЗВУЧНИ НАСТАНИ

За развој на системите за препознавање и класификација на звучните настани се употреби мултидисциплинарен пристап кој опфати теоретски анализи за обработката на дигиталните сигнали и развој на алгоритми на машинско и длабоко учење, експериментални анализи преку развој на модели засновани на вештачката интелигенција и софтверско решение преку користење на бази на податоци за тренирање и тестирање на системите.

Анализата за досегашните достигнувања во областа на вештачката интелигенција во системите за препознавање и класификација на звучните настани овозможува добивање на објективна слика за сознанијата и начинот на моделирање на овие системи.

Експерименталните резултати од субјективната анализа овозможува систематизација на звучните настани преку развој на таксономија која резултираше во избор на 10 класи на звучни настани кои предизвикуваат вознемиреност во урбана животна средина, врз чија основа се одбра база на звучни настани која се искористи за тренирање и тестирање на развиените системи.

Параметризацијата на звучните настани преку дигиталната обработка на аудио сигналите и екстракција на нивните карактеристики е важен чекор во процесот на препознавање и класификација на звучните настани. Теоретската анализа овозможи избор на пет параметри на звучни настани кои се искористија за екстракција на важните карактеристики на аудио сигналите. Развиените математички модели за употребените параметри помогнаа при екстракција на карактеристиките на аудио сигналите во експерименталното моделирање. Процесот на параметризација се реализираше преку трансформација на континуираниот временски сигнал во дискретно време со одредена стапка на семплирање, делење на временски рамки и трансформација на сигналот во друг домен, согласно избраниот параметар. Како резултат се добија визуелни репрезентации на аудио сигналите од базата на звучни настани, кои помогнаа при толкување на точноста на предвидување на класите.

Изборот на алгоритми кои се базираат на техниките на вештачката интелигенција и нивна имплементација претставува инженерски предизвик со многу фактори кои треба да се земат во предвид. За примена на вештачката интелигенција се одбра надгледуван процес на учење при класификација која како влезни параметри користи означени информации за податоците. Математичката метода на која се базира вештачката интелигенција даде придонес за избор на три алгоритми на машинско учење и еден алгоритам на длабоко учење. Преку експериментално моделирање на системот за препознавање и класификација на звучните настани користејќи го софтверот Python се тестираа 52 тест модели. Употребата на хиперпараметарската оптимизацијата на применетите алгоритми резултираше во висока точност при предвидување на резултатите.

Машинското учење користи алгоритми за да ги обработи податоците, да научи од нив и да прави одлуки базирани на тоа што го има научено. Примената на алгоритмот на машини со потпирачки вектори, користејќи 3 параметри на звучни настани резултираше во висока точност од 92.9%, што го прави овој алгоритам применлив за целите на препознавање на звучните настани во урбана средина. Висока точност од 91.53% постигна и алгоритмот на случајни шуми, додека Наивниот Баесов алгоритам постигна точност од 55.68%, што го прави овој алгоритам неприменлив за посакуваните цели.

За разлика од машинското учење, при примена на длабокото учење како техника која се заснова на вештачката интелигенција, процесот на параметризација на звучни настани и класификација е повратен и повторлив. Примената на алгоритмот на конволуциски невронски мрежи како техника на длабокото учење и МЕЛ фреквентни цепстрални параметри постигна точност од 92.68%.

Анализата на конфузните матрици за секој модел даде насоки на класи на звучни настани кои не се добро предвидени. Од анализата на снимените податоци, се забележа дека постојат јасни знаци за лоши перформанси при класификација кога објективниот звук не е доволно гласен и има појава на други извори на звук во аудио податокот. Фактот што оваа појава се случува често во користената база на податоци го отежнува разликувањето помеѓу некои класи, што можеби е и причина за пониска точност при тестирање на системите кои ја користат оваа база на податоци.

Компаративната анализа преку споредба на развиените системи за препознавање и класификација на звучните настани доведе до избор на модел со највисока точност, кој понатаму се искористи за дополнителна валидација и имплементација на сензорска единица. Моделирањето со помош на техниките на длабокото учење има предност пред класичните алгоритми на машинско учење доколку системот што треба да се моделира е комплексен и треба да се примени во апликации во реално време. Бидејќи разликата помеѓу постигнатата точност на најуспешните модели е многу мала (0.22%), се одбра длабокото учење како алгоритам кој ги користи техниките на вештачката интелигенција за понатамошна валидација преку тестирање со непознати звучни настани и понатамошна апликација за развој на паметни градови.

Валидација е процес на дополнително тестирање на системот со база на податоци за кои не е позната класата во која тие припаѓаат. Преку примена на процесот на валидација кој се одвиваше во 2 фази, се добија сознанија за рангот на точноста која би се добила при имплементирање на моделот во реални технички системи. Развиениот алгоритам за валидација на системот овозможува снимање на звукот во реално време, делење во временски секвенци и преработка со цел класификација на звучните настани. Добиената точност од валидираните резултати се движи во граници помеѓу 42.86% – 85.71%. По применетата валидација, системот беше имплементиран на ниско-буџетна сензорска единица која потенцијално може да се користи во паметните градови кои се базираат на техниките на „Интернет на нештата“.

Истражувањето во оваа докторска дисертација со својот теоретски и експериментален дел има посебен придонес во примена на вештачката интелигенција при изградба на системи за препознавање и класификација на звучните настани. Важноста која ја има развиениот систем за препознавање и класификација на звучните настани води кон постојан развој преку примена на ефективни инженерски технологии и решенија со цел точна и сигурна детекција на звучните настани. Овој систем може да се имплементира во реални инженерски апликации и да се искористи за детекција на звучните настани во една урбана средина во реално време.

6.2. ПРЕПОРАКИ ЗА ИДНИ ИСТРАЖУВАЊА

6.2.1. ПОДОБРУВАЊЕ НА МЕТОДИТЕ ЗА ПРЕПОЗНАВАЊЕ И КЛАСИФИКАЦИЈА НА ЗВУЧНИ НАСТАНИ

Во процесот на истражувањата за исполнување на целите на докторската дисертација беа согледани правци за понатамошен развој во насока на зголемување на перформансите и применливоста на системите за препознавање и класификација на звучните настани.

Стандардните методи за објективна анализа на бучавата може да се употребат за квантификација на нивото на бучава и проценување на загаденоста во било кое урбано подрачје. Субјективната анализа како стандардна техника помага во развој на таксономија на звучни настани со цел избор на доминантни класи на вознемирувачки звучни настани. Со подетална анализа на класите на звучни настани може да се дефинира фактор на вознемиреност за секоја класа на звучен настан, кој може да послужи за појасно и прецизно дефинирање на класите на звучните настани.

Примената на напредните техники за проценка на бучавата го потврдија потенцијалот за нивна примена за дефинирање на бучавата преку употреба на системи за препознавање и класификација на звучните настани. Преку изработка на овој систем се појавија проблеми поради пониска точност кои може да се решат со дополнителни анализи и надградување на базата на податоци, со што би се подобрил перформансот на овие системи. Може да се заклучи дека во иднина е потребно да се елиминираат податоците кои имаат шум и присуство на повеќе звучни настани во нив, а базата на податоци да се надгради со звучни настани кои јасно ќе ги одделат класите и нема да предизвикаат конфузија при препознавање на звучните настани.

Резултатите од валидацијата дадоа сознанија за точноста која може да се постигне при употреба на системите во реално време, од каде што може да се забележи дека звучните настани кои имаат шум и/или претставуваат резидуален звук се класифицираат во класата која се однесува на работа на клима уред. Поради тоа што алгоритмот ќе се имплементира на стационарни ниско буџетни сензорски единици кои ќе бидат поставени на одредени локации во урбана животна средина, во иднина би можел да се исклучи изворот на работа на клима уред, и таа класа да се замени со класа на резидуален звук, односно шум, кој доколку алгоритмот го класифицира звучниот настан во оваа класа, тогаш тој резултат не би се прикажал. Дополнително, алгоритмот треба да се конструира на тој начин што доколку при класификација за одреден звучен настан се добие точност пониска од 60%, тогаш би се изолирал овој звук и не би се прикажал во крајните резултати.

Имплементацијата на овие системи во реални апликации може да резултира во намалување или елиминирање на конвенционалните системи за проверка и контрола на бучавата, кои може да бидат доста скапи и да одземат многу време при нивно извршување. Поради тоа, системите за препознавање и класификација на звучните настани се исклучително

важни во инженерски апликации од најразлични области, а нивната примена се зголемува со развојот на моќни системи за обработка на податоците во реално време.

Идните истражувања треба да се фокусираат кон примена на овие системи во концептот на паметен град преку нивна имплементација во повеќе ниско-буџетни сензорски единици кои би биле поставени на различни локации во урбаните средини. Врз основа на концептот на „Интернет на нештата“, резултатите можат да се обработуваат во реално време што овозможува постојано да се следи и контролира бучавата во урбаните средини. Примената на овие системи би резултирала во намалување на загадувањето со бучавата во динамичните урбани средини.

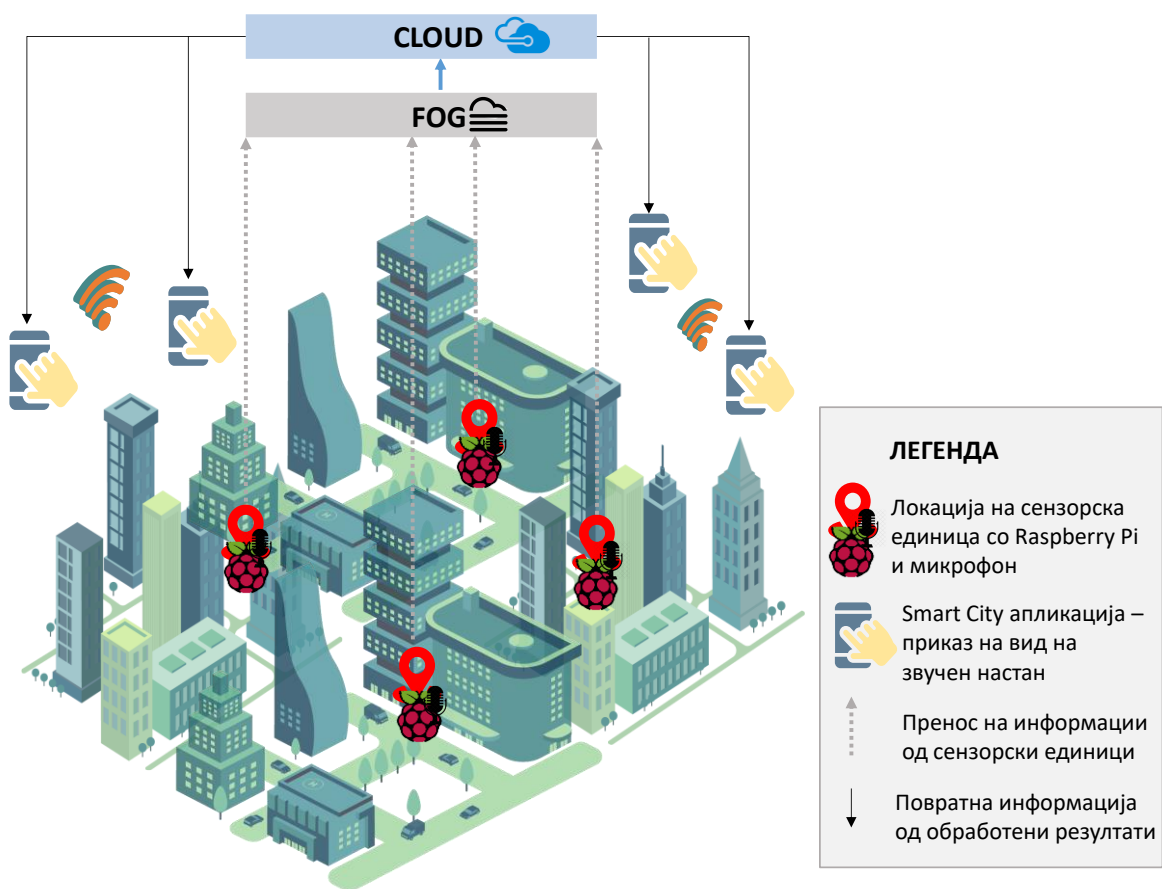
6.2.2. КОНЦЕПТ ЗА ПРИМЕНА ВО СИСТЕМИ ЗА КОНТИНУИРАН МОНИТОРИНГ

Паметен град е концепт кој користи информациска и комуникациска технологија (ИКТ) за да ја подобри оперативната ефикасност, да споделува информации со јавноста и да обезбеди подобар квалитет на услуги и информираност на населението. Клучната технологија за креирање на паметен град е „Интернет на нештата“ (Internet of Things - IoT), со чија помош се создава платформа на која се поставуваат сите функционалности и апликации за паметен град.

Интернет на нештата (IoT) е мрежа која е составена од уреди кои се безжично поврзани преку интернет. Во центарот на IoT-базирани апликации се наоѓа cloud сервер, кој овозможува складирање на податоци. Преку интегрирање на „cloud“ и „fog“ пресметка во IoT технологијата, овозможено е инкорпорирање на алгоритми кои подржуваат преработка на слики, превод, класификација на звучни настани. „Fog“-базираната аналитика поддржува донесување на локални одлуки во однапред дефинирани услови, додека „cloud“ аналитиката е способна да покрива поголеми географски региони со разни управувања и карактеристики. Мотивацијата зад оваа архитектура е дека може да се добијат подлабоки нивоа на апстракција на податоци и знаење додека податоците патуваат низ инфраструктурата на паметниот град. На највисоко ниво, потребен е развој на управување со ресурсите и услугите на градот на долгорочна основа. Од друга страна, на најниско ниво, податоците генерирани од сензорските единици се користат за управување на краткорочна основа. Од кога платформата ќе се воспостави, таа има многу можности за приказ и анализа.

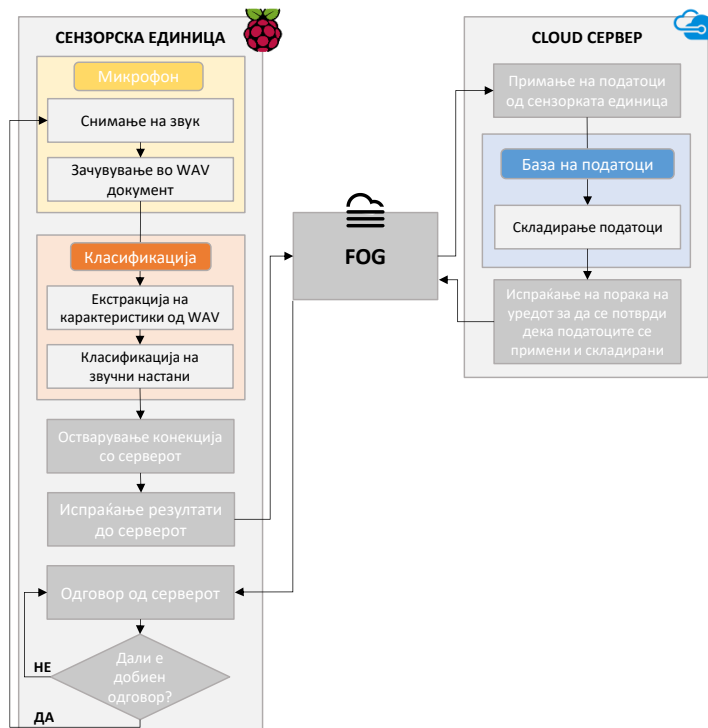
Големата динамичност на паметните градови има потреба од примена на високо ниво на вештачка интелигенција и алгоритми на машинско учење кои се флексибилни и приспособливи за да се справат со динамичноста на податоците, да анализираат и да учат од податоци во реално време. Инкорпорирањето на интелигентни мехатронички системи кои имаат вградено техники на вештачка интелигенција би овозможила голема достапност на обработени податоци кои имаат корисни информации за населението. Преку можноста да се креира архитектура со повеќе сензорски единици за да собираат податоци во поголемо подрачје, паметните градови имаат се поголема и поверодостојна имплементација.

Согласно анализата за паметни градови и „IoT“ инфраструктура, во прилог ќе се објасни предложениот концепт за имплементација на развиениот систем за препознавање и класификација на звучни настани во еден паметен град. Целта е да се постават сензорски единици во најбучните делови од урбаната средина, кои ќе бидат составени од нискобуџетен микроконтролер Raspberry Pi на кој е имплементиран алгоритмот за снимање и валидација на резултатите, и микрофон со кој ќе се овозможи снимањето на податоците. По извршената класификација на снимените податоци, преку „fog“ мрежа, резултатите ќе се префрлат на „cloud“ сервер кој овозможува достапност на информациите до крајните корисници. Предложениот концепт е прикажан на слика 6.1, и е составен од „cloud“ слој кој е поврзан со „fog“ слој што е поврзан со повеќе сензорски единици. Секој слој е програмиран да комуницира еден со друг безжично.



Слика 6.1. Визуелен приказ на предложениот концепт за имплементација на системот за препознавање и класификација на звучни настани во IoT инфраструктура за паметен град

На слика 6.2 е прикажана конфигурација на логички тек кој ќе се примени на секоја сензорска единица. Во оваа поставка, уредот го врши процесот на снимање и класификација на звучните настани. Најпрвин, сензорската единица ги снима и зачувува звучните податоци од околината. Класификација на звучните настани се прави во рамките на сензорската единица, а резултатите безжично се испраќаат преку „fog“ мрежа и се складираат во cloud серверот. Следно, „cloud“ серверот го известува уредот преку порака дека резултатите се примени и зачувани, овозможувајќи повторно префрлување на нови податоци.



6.2. Логички тек на системот за една сензорска единица

Базата на податоци која се користи за тренирање и тестирање на системот за препознавање и класификација има големина од 8 GB, а самиот алгоритам на конволуциски невронски мрежи и процесот на параметризирање на податоците има потреба од голема моќност за процесирање и контролер со голема меморија. Од тие причини е избран Raspberry Pi микроконтролерот кој има висока моќност на процесирање на податоците, вграден WiFi модул и голема меморија: вграден 4 GB SDRAM и мемориска картичка од 64 GB, кој има голема моќност за процесирање и мала потрошувачка на енергија.

За снимање на податоците може да се користи дигитален микрофон кој ги снима податоците со стапка на семплирање од 48 000 Hz и длабочина од 32 bit. Понатаму, вака снимениот податок преку параметарот MFCC се преработува од .wav податок во вектор од карактеристики што го претставува параметарот на звучен настан, со цел понатамошна класификација во алгоритмот на конволуциски невронски мрежи.

Од кога ќе се направи класификација на звучните настани, резултатите се испраќаат на cloud сервер преку WI-FI со помош на рутер кој е претставен како „fog” слој. Резултатите во форма на .csv документ на која е прикажана класификацијата на снимените звучни настани и постигнатата точност се складираат на „Cloud” серверот преку користење на Google Drive.

Со цел докажување на концептот за имплементација на системот за препознавање и класификација на звучните настани во апликација за паметен град, развиена е една сензорска единица која успешно ги снима податоците во реално време и ги класифицира во една од 10те класи на звучни настани. Сознанијата од ова истражување може целосно да се применат во апликација на паметен град со што би се добил континуиран мониторинг кој би овозможил следење на звучните настани во една урбана средина.

ЛИТЕРАТУРА

- [1] Zannin, P.H.T., Engel, M.S., Fiedler, P.E.K. and Bunn, F., 2013. Characterization of environmental noise based on noise measurements, noise mapping and interviews: A case study at a university campus in Brazil. *Cities*, 31, pp.317-327.
- [2] Bąkowski, A., Radziszewski, L., Dekýš, V. and Švietlik, P., 2019, May. Frequency analysis of urban traffic noise. In *2019 20th International Carpathian Control Conference (ICCC)* (pp. 1-6). IEEE.
- [3] Stansfeld, S.A. and Matheson, M.P., 2003. Noise pollution: non-auditory effects on health. *British medical bulletin*, 68(1), pp.243-257.
- [4] Directive END 2002/49/EC of the European parliament and of the council of 25 June 2002 relating to the assessment and management of environmental noise
- [5] ISO 1996:1 Acoustics-Description, measurement and assessment of environmental --noise-Part 1: Basic quantities and assessment procedures
- [6] ISO 1996:2 Acoustics-Description, measurement and assessment of environmental --noise-Part 2: Determination of environmental noise levels
- [7] Закон за заштита од бучава во животна средина (Службен весник на РМ бр. 79/2007; 124/2010; 47/2011)
- [8] Правилник за примена на индикаторите за бучава, дополнителни индикатори за бучава, начин за мерење на бучава и методи за оценување на за бучава во животната средина (Службен весник на РМ 107/08)
- [9] Temko, A., Nadeu, C., Macho, D., Malkin, R., Zieger, C. and Omologo, M., 2009. Acoustic event detection and classification. In *Computers in the human interaction loop* (pp. 61-73). Springer, London.
- [10] Socoró, J.C., Alías, F., Alsina, R.M., Sevillano, X. and Camins, Q., B3-Report describing the ANED algorithms for low and high computation capacity sensors.
- [11] Sevillano, X., Socoró, J.C., Alías, F., Bellucci, P., Peruzzi, L., Radaelli, S., Coppi, P., Nencini, L., Cerniglia, A., Bisceglie, A. and Benocci, R., 2016. DYNAMAP–Development of low cost sensors networks for real time noise mapping. *Noise mapping*, 3(1).
- [12] Bello, J.P., Silva, C., Nov, O., Dubois, R.L., Arora, A., Salamon, J., Mydlarz, C. and Doraiswamy, H., 2019. Sonyc: A system for monitoring, analyzing, and mitigating urban noise pollution. *Communications of the ACM*, 62(2), pp.68-77.
- [13] Salamon, J., Jacoby, C. and Bello, J.P., 2014, November. A dataset and taxonomy for urban sound research. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 1041-1044).
- [14] Sharma, G., Umopathy, K. and Krishnan, S., 2020. Trends in audio signal feature extraction methods. *Applied Acoustics*, 158, p.107020.
- [15] DYNAMAP Report: State of the art on sound source recognition and anomalous event elimination. Project: Dynamic Acoustic Mapping – Development of low cost sensors networks for real time noise mapping LIFE Dynamap report A1, LIFE13 ENV/IT/001254, 2015
- [16] Gygi, B., 2001. *Factors in the identification of environmental sounds* (Doctoral dissertation, Indiana University).
- [17] Bountourakis, V., Vrysis, L. and Papanikolaou, G., 2015. Machine learning algorithms for environmental sound recognition: Towards soundscape semantics. In *Proceedings of the Audio Mostly 2015 on Interaction With Sound* (pp. 1-7).

- [18] Caruana, R. and Niculescu-Mizil, A., 2006, June. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning* (pp. 161-168).
- [19] European Commission, 1996. Green Paper on Future Noise Policy.
- [20] World Health Organization. (2003). The World health report: 2003: shaping the future
- [21] Blanes, N., Fons, J., Houthuijs, D., Swart, W., de la Maza, M., Ramos, M., Castell, N. and van Kempen, E., 2017. Noise in Europe 2017: updated assessment. *European Topic Centre on Air Pollution and Climate Change Mitigation (ETC/ACM): Bilthoven, The Netherlands*.
- [22] Handbook on the Implementation of EC Environmental Legislation, Section 9 - Noise Legislation
- [23] Murat, O. and Ebru, S.E., 2016. Determination of traffic noise pollution of the city of Tekirdag. *Journal of Environmental Protection and Ecology*, 17(4), pp.1276-1284.
- [24] Saadu, A.A., Onyeonwu, R.O., Ayorinde, E.O. and Ogisi, F.O., 1996. Community attitudinal noise survey and analysis of eight Nigerian cities. *Applied Acoustics*, 49(1), pp.49-69.
- [25] Lam, W.H. and Tam, M.L., 1998. Reliability analysis of traffic noise estimates in Hong Kong. *Transportation Research Part D: Transport and Environment*, 3(4), pp.239-248.
- [26] Hagler, L. and Goiner, L., 2007. Noise pollution: a modern plague. *Southern Medical Journal*, 100(3), pp.287-294.
- [27] Li, B., Tao, S. and Dawson, R.W., 2002. Evaluation and analysis of traffic noise from the main urban roads in Beijing. *Applied acoustics*, 63(10), pp.1137-1142.
- [28] King, E.A. and Rice, H.J., 2009. The development of a practical framework for strategic noise mapping. *Applied Acoustics*, 70(8), pp.1116-1127.
- [29] Farçaş, F. and Sivertunb, Å., 2010. Road traffic noise: GIS tools for noise mapping and a case study for Skåne region. *Sweden: Citeseer*.
- [30] Prašćević, M., Cvetković, D. and Mihajlov, D., 2011. The uncertainty sources in environmental noise measurements and the uncertainty estimation. *Facta universitatis-series: Mechanical Engineering*, 9(2), pp.183-192.
- [31] Alves, S. and Waddington, D., 2014. Estimation of Uncerta ISO 1996-2. *Acoustics Bulletin*, 39(2), pp.34-41.
- [32] Hannah L, Wyatt P, McLaren S. An introductory guide to uncertainty in acoustic measurements. New Zeland acoustics
- [33] Manvell, D., 2020. Uncertainty in International Acoustics Standards. In *Uncertainty in Acoustics: Measurement, Prediction and Assessment* (pp. 431-443). CRC Press.
- [34] Pinto, F.A.C., 2010. Urban Noise Pollution Assessment Techniques. *Methods and Techniques in Urban Engineering*, p.237.
- [35] Mohammadi, G., 2009. An investigation of community response to urban traffic noise. In *Global Perspective for Competitive Enterprise, Economy and Ecology* (pp. 673-680). Springer, London.
- [36] Steele, C., 2001. A critical review of some traffic noise prediction models. *Applied acoustics*, 62(3), pp.271-287.
- [37] Tsai, K.T., Lin, M.D. and Chen, Y.H., 2009. Noise mapping in urban environments: A Taiwan study. *Applied Acoustics*, 70(7), pp.964-972.
- [38] Belojević, G., Jakovljević, B. and Aleksić, O., 1997. Subjective reactions to traffic noise with regard to some personality traits. *Environment international*, 23(2), pp.221-226.
- [39] Miedema, H.M. and Oudshoorn, C.G., 2001. Annoyance from transportation noise: relationships with exposure metrics DNL and DENL and their confidence intervals. *Environmental health perspectives*, 109(4), pp.409-416.
- [40] Klæboe, R., Amundsen, A.H., Fyhri, A. and Solberg, S., 2004. Road traffic noise—the relationship between noise exposure and noise annoyance in Norway. *Applied Acoustics*, 65(9), pp.893-912.
- [41] Zannin, P.H.T., Engel, M.S., Fiedler, P.E.K. and Bunn, F., 2013. Characterization of environmental noise based on noise measurements, noise mapping and interviews: A case study at a university campus in Brazil. *Cities*, 31, pp.317-327.
- [42] Socoró, J.C., Sevillano, X. and Alías, F., 2016, August. Analysis and automatic detection of anomalous noise events in real recordings of road traffic noise for the LIFE DYNAMAP project.

- In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings* (Vol. 253, No. 6, pp. 1943-1952). Institute of Noise Control Engineering.
- [43] Dennis, J.W., 2014. Sound event recognition in unstructured environments using spectrogram image processing. *Nanyang Technological University, Singapore*.
- [44] Temko A. Acoustic Event Detection and Classification [Report] : PhD thesis / Universitat Politècnica de Catalunya. - 2007.
- [45] Zhuang, X., Zhou, X., Hasegawa-Johnson, M.A. and Huang, T.S., 2010. Real-world acoustic event detection. *Pattern recognition letters*, 31(12), pp.1543-1551.
- [46] Gygi, B., 2001. *Factors in the identification of environmental sounds* (Doctoral dissertation, Indiana University).
- [47] Mitrović, D., Zeppelzauer, M. and Breiteneder, C., 2010. Features for content-based audio retrieval. In *Advances in computers* (Vol. 78, pp. 71-150). Elsevier.
- [48] Su, Y., Zhang, K., Wang, J., Zhou, D. and Madani, K., 2020. Performance analysis of multiple aggregated acoustic features for environment sound classification. *Applied Acoustics*, 158, p.107050.
- [49] Valero X, Al'ias F, 2011. Automatic monitoring of environmental noise sources. In: *Proceedings Tecnicacustica*, Caceres.
- [50] Da Silva, B., W Happi, A., Braeken, A. and Touhafi, A., 2019. Evaluation of classical machine learning techniques towards urban sound recognition on embedded systems. *Applied Sciences*, 9(18), p.3885.
- [51] Agarwal, I., Yadav, P., Gupta, N. and Yadav, S., 2021. Urban Sound Classification Using Machine Learning and Neural Networks. In *Proceedings of 6th International Conference on Recent Trends in Computing* (pp. 323-330). Springer, Singapore.
- [52] Alsouda, Y., Pllana, S. and Kurti, A., 2019, May. lot-based urban noise identification using machine learning: performance of SVM, KNN, bagging, and random forest. In *Proceedings of the international conference on omni-layer intelligent systems* (pp. 62-67).
- [53] Zhang, Z., Xu, S., Zhang, S., Qiao, T. and Cao, S., 2019. Learning attentive representations for environmental sound classification. *IEEE Access*, 7, pp.130327-130339.
- [54] Mushtaq, Z., Su, S.F. and Tran, Q.V., 2021. Spectral images based environmental sound classification using CNN with meaningful data augmentation. *Applied Acoustics*, 172, p.107581.
- [55] Mulimani, M. and Koolagudi, S.G., 2019. Segmentation and characterization of acoustic event spectrograms using singular value decomposition. *Expert Systems with Applications*, 120, pp.413-425.
- [56] Schroder, B. Cauchi, M. R. Schadler, 2013. *Acoustic event detection using signal enhancement and spectro-temporal feature extraction*. in IEEE Workshop on Applcat. Signal Process. Audio Acoust.(WASPAA), pp. 1–3.
- [57] Liu, C., Hong, F., Feng, H., Zhai, Y. and Chen, Y., 2021. Environmental Sound Classification Based on Stacked Concatenated DNN using Aggregated Features. *Journal of Signal Processing Systems*, 93(11), pp.1287-1299.
- [58] Pang, C., Liu, H. and Li, X., 2019. Multitask learning of time-frequency CNN for sound source localization. *IEEE Access*, 7, pp.40725-40737.
- [59] Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- [60] Chang, C. and Doran, B., 2016. Urban Sound Classification: With Random Forest SVM DNN RNN and CNN Classifiers. *CSCI E-81 Machine Learning and Data Mining Final Project Fall*.
- [61] Zhang, X., Zou, Y. and Shi, W., 2017, August. Dilated convolution neural network with LeakyReLU for environmental sound classification. In *2017 22nd international conference on digital signal processing (DSP)* (pp. 1-5). IEEE.
- [62] Piczak, K.J., 2015, September. Environmental sound classification with convolutional neural networks. In *2015 IEEE 25th international workshop on machine learning for signal processing (MLSP)* (pp. 1-6). IEEE.

- [63] Boddapati, V., Petef, A., Rasmusson, J. and Lundberg, L., 2017. Classifying environmental sounds using image recognition networks. *Procedia computer science*, 112, pp.2048-2056.
- [64] Li, S., Yao, Y., Hu, J., Liu, G., Yao, X. and Hu, J., 2018. An ensemble stacked convolutional neural network model for environmental event sound recognition. *Applied Sciences*, 8(7), p.1152.
- [65] Zhang, Z., Xu, S., Cao, S. and Zhang, S., 2018, November. Deep convolutional neural network with mixup for environmental sound classification. In *Chinese conference on pattern recognition and computer vision (prcv)* (pp. 356-367). Springer, Cham.
- [66] Alsouda, Y., Pllana, S. and Kurti, A., 2018. A machine learning driven IoT solution for noise classification in smart cities. *arXiv preprint arXiv:1809.00238*.
- [67] Afshankaleem, Santi Prabba. Enhancement of Urban Sound Classification Using Various Feature Extraction Techniques
- [68] Su, Y., Zhang, K., Wang, J. and Madani, K., 2019. Environment sound classification using a two-stream CNN based on decision-level fusion. *Sensors*, 19(7), p.1733.
- [69] Abdoli, S., Cardinal, P. and Koerich, A.L., 2019. End-to-end environmental sound classification using a 1D convolutional neural network. *Expert Systems with Applications*, 136, pp.252-263.
- [70] Demir, F., Turkoglu, M., Aslan, M. and Sengur, A., 2020. A new pyramidal concatenated CNN approach for environmental sound classification. *Applied Acoustics*, 170, p.107520.
- [71] Gorla, S., Comai, S., Masciadri, A. and Salice, F., 2017. BigEar: Ubiquitous Wireless Low-Budget Speech Capturing Interface. *Journal of Computer and Communications*, 5(04), p.60.
- [72] Baucas, M.J. and Spachos, P., 2020. A scalable IoT-fog framework for urban sound sensing. *Computer Communications*, 153, pp.302-310.
- [73] Manvell, D., 2015. Utilising the strengths of different sound sensor networks in smart city noise management. In *Proceedings of the EuroNoise*.
- [74] Alías, F. and Alsina-Pagès, R.M., 2019. Review of wireless acoustic sensor networks for environmental noise monitoring in smart cities. *Journal of sensors*, 2019.
- [75] Mydlarz, C., Salamon, J. and Bello, J.P., 2017. The implementation of low-cost urban acoustic monitoring devices. *Applied Acoustics*, 117, pp.207-218.
- [76] Bello, J.P., Silva, C., Nov, O., Dubois, R.L., Arora, A., Salamon, J., Mydlarz, C. and Doraiswamy, H., 2019. Sonyc: A system for monitoring, analyzing, and mitigating urban noise pollution. *Communications of the ACM*, 62(2), pp.68-77.
- [77] Hollosi, D., Nagy, G., Rodigast, R., Goetze, S. and Cousin, P., 2013, August. Enhancing wireless sensor networks with acoustic sensing technology: use cases, applications & experiments. In *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing* (pp. 335-342). IEEE.
- [78] Baucas, M.J. and Spachos, P., 2020. Using cloud and fog computing for large scale IoT-based urban sound classification. *Simulation modelling practice and theory*, 101, p.102013.
- [79] Benocci, R., Molteni, A., Cambiaghi, M., Angelini, F., Roman, H.E. and Zambon, G., 2019. Reliability of Dynamap traffic noise prediction. *Applied Acoustics*, 156, pp.142-150.
- [80] Bellucci, P., Peruzzi, L. and Zambon, G., 2017. LIFE DYNAMAP project: The case study of Rome. *Applied Acoustics*, 117, pp.193-206.
- [81] „Правилник за локации на мерни станици и мерни места“ (Службен весник на РМ бр. 120/08)
- [82] Ristovska, G., Gjorgjev, D., Polozhani, A., Kočubovski, M. and Kendrovski, V., 2009. Environmental noise and annoyance in adult population of Skopje: a cross-sectional study. *Arhiv za higijenu rada i toksikologiju*, 60(3), pp.349-355.
- [83] <https://skopjepulse.mk/>
- [84] ETRA-CERTU-LCPC-CSTB) од 'Arrete du 5 mai 1995 relatif au bruit des infrastructures routieters, Journal Officiel du 10 mai 1995, Article 6' и во францускиот стандард 'XPS 31-133'
- [85] Zannin, P.H.T., Diniz, F.B. and Barbosa, W.A., 2002. Environmental noise pollution in the city of Curitiba, Brazil. *Applied Acoustics*, 63(4), pp.351-358.
- [86] Krejcie, R.V. and Morgan, D.W., 1970. Determining sample size for research activities. *Educational and psychological measurement*, 30(3), pp.607-610.

- [87] Sample size calculator (<https://www.calculator.net/sample-size-calculator.html>) (last accessed on 6 July 2019)
- [88] Jackson, S.L., 2009. *Research methods and statistics*. Wadsworth, Cengage Learning.
- [89] Bąkowski, A., Radziszewski, L., Dekýš, V., & Šwietlik, P. (2019, May). Frequency analysis of urban traffic noise. In 2019 20th International Carpathian Control Conference (ICCC) (pp. 1-6).
- [90] Road traffic and Aircraft Noise Spectrums, Jeffrey Mahn, Acoustic Research Group, Department of Mechanical Engineering, Universitu od Canterbury, Report
- [91] Watts, G. R., & Stait, R. E. (2009). Characteristics of vehicles producing excessive noise and groundborne vibration-phase 1. TRL Published Project Report. E. Author: "The Title of the Conference Paper," Proceedings of the International Symposium on Music Information Retrieval, pp. 000–111, 2000.
- [92] 76
- [93] Downie, J.S., 2003. Music information retrieval. *Annual review of information science and technology*, 37(1), pp.295-340.
- [94] Kedem, B., 1986. Spectral analysis and discrimination by zero-crossings. *Proceedings of the IEEE*, 74(11), pp.1477-1493.
- [95] El-Maleh, K., Klein, M., Petrucci, G. and Kabal, P., 2000, June. Speech/music discrimination for multimedia applications. In *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100)* (Vol. 4, pp. 2445-2448). IEEE.
- [96] Carpentier, G., 2005, July. Information technology—Multimedia content description interface—Part 4: Audio, AMENDMENT 2: High-level descriptors. In *Motion Picture Expert Group (ISO/IEC JTC 1 SC29)*.
- [97] Jiang, H., Bai, J., Zhang, S. and Xu, B., 2005, October. SVM-based audio scene classification. In *2005 International Conference on Natural Language Processing and Knowledge Engineering* (pp. 131-136). IEEE.
- [98] Liu, Z., Wang, Y. and Chen, T., 1998. Audio feature extraction and analysis for scene segmentation and classification. *Journal of VLSI signal processing systems for signal, image and video technology*, 20(1), pp.61-79.
- [99] Schroeder, M. and Atal, B.S., 1985, April. Code-excited linear prediction (CELP): High-quality speech at very low bit rates. In *ICASSP'85. IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 10, pp. 937-940). IEEE.
- [100] Tzanetakis, G., Jones, R. and McNally, K., 2007, September. Stereo Panning Features for Classifying Recording Production Style. In *ISMIR* (pp. 441-444).
- [101] Sethares, W.A., Morris, R.D. and Sethares, J.C., 2005. Beat tracking of musical performances using low-level audio features. *IEEE Transactions on speech and audio processing*, 13(2), pp.275-285.
- [102] Zwicker, E. and Fastl, H., 1999. *Psychoacoustics Facts and Models* Springer Heiderberg. *Technische Universität München. München: Germany*.
- [103] Wold, E., Blum, T., Keislar, D. and Wheaton, J., 1996. Content-based classification, search, and retrieval of audio. *IEEE multimedia*, 3(3), pp.27-36.
- [104] Ramalingam, A. and Krishnan, S., 2006. Gaussian mixture modeling of short-time Fourier transform features for audio fingerprinting. *IEEE Transactions on Information Forensics and Security*, 1(4), pp.457-463.
- [105] Pampalk, E., Flexer, A. and Widmer, G., 2005, September. Improvements of Audio-Based Music Similarity and Genre Classification. In *ISMIR* (Vol. 5, pp. 634-637).
- [106] Pfeiffer, S., 1998, December. Importance of perceptive adaptation of sound features in audio content processing. In *Storage and Retrieval for Image and Video Databases VII* (Vol. 3656, pp. 328-337). International Society for Optics and Photonics.
- [107] Tran, H.D. and Li, H., 2011, May. Probabilistic distance SVM with Hellinger-Exponential Kernel for sound event classification. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2272-2275). IEEE.

- [108] Bartsch, M.A. and Wakefield, G.H., 2005. Audio thumbnailing of popular music using chroma-based representations. *IEEE Transactions on multimedia*, 7(1), pp.96-104.
- [109] Bogert, B.P., 1963. The quefrency alanyis of time series for echoes; Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. *Time series analysis*, pp.209-243.
- [110] Liang, S. and Fan, X., 2014. Audio content classification method research based on two-step strategy. *Int. J. Adv. Comput. Sci. Appl.(IJACSA)*, 5, pp.57-62.
- [111] Li, Q., Soong, F.K. and Siohan, O., 2001. An auditory system-based feature for robust speech recognition. In *Seventh European Conference on Speech Communication and Technology*.
- [112] Rabaoui, A., Lachiri, Z. and Ellouze, N., 2004, December. Automatic environmental noise recognition. In *2004 IEEE International Conference on Industrial Technology, 2004. IEEE ICIT'04*. (Vol. 3, pp. 1670-1675). IEEE.
- [113] Ravindran, S., Schlemmer, K. and Anderson, D.V., 2005. A physiologically inspired method for audio classification. *EURASIP Journal on Advances in Signal Processing*, 2005(9), pp.1-8.
- [114] Valero, X. and Alías, F., 2010. Applicability of MPEG-7 low level descriptors to environmental sound source recognition. In *Proceedings 1st Euroregio Conference, Ljubjana*.
- [115] Hermansky, H., 1990. Perceptual linear predictive (PLP) analysis of speech. *the Journal of the Acoustical Society of America*, 87(4), pp.1738-1752.
- [116] Atal, B.S., 1974. Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification. *the Journal of the Acoustical Society of America*, 55(6), pp.1304-1312.
- [117] Scheirer, E. and Slaney, M., 1997, April. Construction and evaluation of a robust multifeature speech/music discriminator. In *1997 IEEE international conference on acoustics, speech, and signal processing* (Vol. 2, pp. 1331-1334). IEEE.
- [118] Srinivasan, S., Petkovic, D. and Ponceleon, D., 1999, October. Towards robust features for classifying audio in the CueVideo system. In *Proceedings of the seventh ACM international conference on Multimedia (Part 1)* (pp. 393-400).
- [119] Malvar, H., 1999, March. A modulated complex lapped transform and its applications to audio processing. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258)* (Vol. 3, pp. 1421-1424). IEEE.
- [120] Lindgren, A.C., 2003. *Speech recognition using features extracted from phase space reconstructions* (Doctoral dissertation, Marquette University).
- [121] Cardenas, A.A., Manadhata, P.K. and Rajan, S.P., 2013. Big data analytics for security. *IEEE Security & Privacy*, 11(6), pp.74-76.
- [122] Labrinidis, A. and Jagadish, H.V., 2012. Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, 5(12), pp.2032-2033.
- [123] Shalev-Shwartz, S. and Ben-David, S., 2014. *Understanding machine learning: From theory to algorithms*. Cambridge university press.
- [124] Cortes, C. and Vapnik, V., 1995. Support-vector networks. *Machine learning*, 20(3), pp.273-297.
- [125] Chih-Wei, H., Chih-Chung, C. and Chih-Jen, L., 2003. A practical guide to support vector classification (Technical Report). *Department of Computer Science and Information Engineering, National Taiwan University*.
- [126] Noble, W.S., 2006. What is a support vector machine?. *Nature biotechnology*, 24(12), pp.1565-1567.
- [127] Zhang, H., 2004. The optimality of naive Bayes. *Aa*, 1(2), p.3.
- [128] Rish, I., 2001, August. An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence* (Vol. 3, No. 22, pp. 41-46).
- [129] Rennie, J.D., Shih, L., Teevan, J. and Karger, D.R., 2003. Tackling the poor assumptions of naive bayes text classifiers. In *Proceedings of the 20th international conference on machine learning (ICML-03)* (pp. 616-623).
- [130] John, G.H. and Langley, P., 1995. Estimating continuous distributions in bayesian classifiers. in proceedings of the eleventh conference on uncertainty in artificial intelligence.

- [131] Breiman, L., 1996. Bagging predictors. *Machine learning*, 24(2), pp.123-140.
- [132] Breiman, L., 2001. Random forests. *Machine learning*, 45(1), pp.5-32.
- [133] Murphy, K.P., 2012. *Machine learning: a probabilistic perspective*. MIT press.
- [134] Bengio, Y., 2000. Gradient-based optimization of hyperparameters. *Neural computation*, 12(8), pp.1889-1900.
- [135] Luketina, J., Berglund, M., Greff, K. and Raiko, T., 2016, June. Scalable gradient-based tuning of continuous regularization hyperparameters. In *International conference on machine learning* (pp. 2952-2960). PMLR.
- [136] DeCastro-García, N., Muñoz Castañeda, Á.L., Escudero García, D. and Carriegos, M.V., 2019. Effect of the sampling of a dataset in the hyperparameter optimization phase over the efficiency of a machine learning algorithm. *Complexity*, 2019.
- [137] LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *nature*, 521(7553), pp.436-444.
- [138] Wu, J., 2017. Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology. Nanjing University. China*, 5(23), p.495.
- [139] Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- [140] Deng, H., 2013. Guided random forest in the RRF package. *arXiv preprint arXiv:1306.0237*.
- [141] Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N. and Kingsbury, B., 2012. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine*, 29(6), pp.82-97.
- [142] Duchi, J., Hazan, E. and Singer, Y., 2011. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7).
- [143] Hinton, G.E., 2012. A practical guide to training restricted Boltzmann machines. In *Neural networks: Tricks of the trade* (pp. 599-619). Springer, Berlin, Heidelberg.
- [144] D. Zhe, W. Qinghong, S. Naizheng, Z. Yuhan, Study on data security policy based on cloud storage, in: 2017 IEEE 3rd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (Hpsc), and IEEE International Conference on Intelligent Data and Security (IDS), 2017, pp. 145–149. doi: 10.1109/BigDataSecurity.2017.12
- [145] R. Muoz, R. Vilalta, N. Yoshikane, R. Casellas, R. Martnez, T. Tsuritani, I. Morita, Integration of iot, transport sdn, and edge/- cloud computing for dynamic distribution of iot analytics and efficient use of network resources, *Journal of Lightwave Technology* 36 (7) (2018) 1420–1428. doi:10.1109/JLT.2018. 2800660.
- [146] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, M. Ayyash, Internet of things: A survey on enabling technologies, protocols, and applications, *IEEE Communications Surveys Tutorials* 17 (4) (2015) 2347–2376. doi:10.1109/COMST. 2015.2444095.
- [147] M. Shiraz, A. Gani, R. H. Khokhar, R. Buyya, A review on distributed application processing frameworks in smart mobile devices for mobile cloud computing, *IEEE Communications Surveys Tutorials* 15 (3) (2013) 1294–1313. doi:10.1109/ SURV.2012.111412.00045.
- [148] Alsouda, Y., Pllana, S. and Kurti, A., 2018. A machine learning driven IoT solution for noise classification in smart cities. *arXiv preprint arXiv:1809.00238*.
- [149] Hernandez-Jayo, U. and Goñi, A., 2021. ZARATAMAP: Noise Characterization in the Scope of a Smart City through a Low Cost and Mobile Electronic Embedded System. *Sensors*, 21(5), p.1707.
- [150] Alias, F. and Alsina-Pagès, R.M., 2019. Review of wireless acoustic sensor networks for environmental noise monitoring in smart cities. *Journal of sensors*, 2019.

ПРИЛОЗИ

ПРИЛОГ А

- **ПРИЛОГ А-1** – Програмски код во Python за анализа на базата на звучни настани
- **ПРИЛОГ А-2** – Програмски код во Python за визуелизација на параметрите на звучни настани за еден звучен настан

ПРИЛОГ Б

- **ПРИЛОГ Б-1** - Конфузни матрици од тестирањата на алгоритмите чии резултати се прикажани на табела 4.3. во глава 4
- **ПРИЛОГ Б-2** – Програмски код во Python за хиперпараметарска оптимизација на алгоритмот на машини со потпирачки вектори преку користење на параметрите MFCC, хроматограм и спектрален контраст
- **ПРИЛОГ Б-3** – Програмски код во Python за хиперпараметарска оптимизација на алгоритмот на случајни шуми преку користење на параметрите MFCC, хроматограм и спектрален контраст
- **ПРИЛОГ Б-4** – Програмски код во Python за наивен Баесов алгоритам преку користење на параметрите MFCC, хроматограм и спектрален контраст

ПРИЛОГ В

- **ПРИЛОГ В-1** – Програмски код во Python за конволуциски невронски мрежи при примена на MFCC
- **ПРИЛОГ В-2** – Програмски код во Python за конволуциски невронски мрежи при примена на МЕЛ спектрограм
- **ПРИЛОГ В-3** – Програмски код во Python за конволуциски невронски мрежи при примена на MFCC и МЕЛ спектрограм
- **ПРИЛОГ В-4** – Програмски код во Python за конволуциски невронски мрежи при примена на MFCC, хроматограм и спектрален контраст
- **ПРИЛОГ В-5** – Програмски код во Python за снимање на звучни настани и нивна валидација преку примена на алгоритмот на конволуциски невронски мрежи со примена на MFCC