

Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті

ӘОЖ 621.39;004.93;004.032.26 (043)

Қолжазба құқығында

ДОСБАЕВ ЖАНДОС МАХСУТУЛЫ

**Аудиосигналдар негізінде төтенше жағдайларды анықтау және
классификациялауды зерттеу**

6D071900 – Радиотехника, электроника және телекоммуникациялар

Философия докторы (PhD) дәрежесін
алу үшін дайындалған диссертация

Ғылыми кеңесшілері

доктор PhD

Омаров Б.С.,

доктор PhD,

қауымдастырылған профессор

Сулиман А.

(INTI Халықаралық

Университеті, Малайзия)

Қазақстан Республикасы
Алматы, 2023

МАЗМҰНЫ

НОРМАТИВТІК СІЛТЕМЕЛЕР.....	4
БЕЛГІЛЕУЛЕР МЕН ҚЫСҚАРТУЛАР.....	5
КІРІСПЕ.....	7
1 ТӨТЕНШЕ ЖАҒДАЙЛАРДЫ АУДИО АРҚЫЛЫ АНЫҚТАУ БОЙЫНША ӘДЕБИЕТТЕРГЕ ШОЛУ.....	11
1.1 Қауіпсіздікті бақылау жүйелері.....	11
1.2 Аудиоклассификация тапсырмасын орындау сатылары.....	12
1.3 Аудио мәліметтерді алдын-ала өңдеу.....	13
1.4 Аудио мәліметтерден белгілерді алу.....	14
1.5 Аудио мәліметтер жинағы (датасет).....	22
1.6 Модельді бағалау метрикалары.....	26
1.7 Аудиосигналдар негізінде оқиғаларды анықтауда қолданылатын әдістер.....	31
2 ТӨТЕНШЕ ЖАҒДАЙЛАРДЫ АНЫҚТАУ ҮШІН АУДИО МӘЛІМЕТТЕРДІ МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІ НЕГІЗІНДЕ ӨНДЕУ.....	37
2.1 Жасанды интеллект және машиналық оқыту әдістерін төтенше жағдайларды анықтауда қолдану.....	37
2.2 Төтенше жағдайларды анықтау үшін аудио мәліметтерді өңдеуге терең оқыту әдістерін қолдану.....	39
2.2.1 Терең оқытудың жұмыс істеу принципі.....	39
2.2.2 Терең оқытудың ерекшеліктері.....	44
2.3 Нейрондық желілердің математикалық негіздері.....	46
2.3.1 Тензорларды анықтауға қолданылатын атрибуттар.....	47
2.3.2 Тензорларға қолданылатын амалдар.....	49
2.3.3 Нейрондық желілер механизмі: градиент негізінде оптимизациялау	50
2.4 Терең оқытудың құрылымдық блоктары.....	55
2.5 Машиналық оқыту негізі және машиналық оқыту алгоритмдерінің категориялары.....	56
2.6 Машиналық оқыту модельдерін бағалау.....	56
2.7 Мәліметтерді өңдеу, белгілерді құру және белгілерді оқыту.....	58
2.8 Артық оқыту, жеткіліксіз оқыту және толықтырып оқыту.....	59
2.9 CNN – конволюциялық нейрондық желілер.....	61
2.10 RNN – рекуррентті нейрондық желілер.....	63
2.10.1 LSTM және GRU қабаттары.....	64
2.11 Төтенше жағдайларды анықтау үшін CNN көмегімен аудио мәліметтер тізбектерін өңдеу.....	67
2.12 Төтенше жағдайларды анықтау үшін өңдеу үшін конволюциялық және рекурренттік желілерді біріктіру.....	67
3 АУДИОСИГНАЛДАРДЫ ӨНДЕУ АРҚЫЛЫ ТӨТЕНШЕ ОҚИҒАЛАРДЫ НАҚТЫ УАҚЫТ РЕЖИМІНДЕ АНЫҚТАУҒА ЖӘНЕ КЛАССИФИКАЦИЯЛАУҒА	

	АРНАЛҒАН МОДЕЛЬДІ ҚҰРУ	69
3.1	Төтенше жағдайларды анықтауға арналған модельді оқыту үшін дайындалған аудио мәліметтер жинағы (датасет).....	69
3.2	Төтенше жағдайларды анықтауға арналған модельді оқытуда қолданылатын аудио мәліметтерді алдын-ала өңдеу.....	72
3.3	Төтенше жағдайдың импульстік дыбыстарын анықтау.....	75
3.4	Төтенше жағдайларды анықтау үшін жинақталған аудио мәліметтерден белгілерді алу.....	76
4	АУДИОСИГНАЛДАРДЫ ӨНДЕУ АРҚЫЛЫ ТӨТЕНШЕ ЖАҒДАЙЛАРДЫ АНЫҚТАУҒА ЖӘНЕ КЛАССИФИКАЦИЯЛАУҒА АРНАЛҒАН МОДЕЛЬДІ ОҚЫТУ НӘТИЖЕЛЕРІ	83
4.1	Төтенше жағдайларды анықтау үшін аудио мәліметтерді өңдеуге ұсынылатын CNN-BiLSTM моделінің архитектурасы.....	83
4.2	CNN-BiLSTM моделін оқыту үшін гиперпараметрлерді таңдау және бағалау метрикалары.....	86
4.3	CNN-BiLSTM комбинациясына негізделген төтенше оқиғаларды анықтауға арналған модельді оқыту бойынша алынған нәтижелер	86
4.4	Ұсынылатын CNN-BiLSTM моделін ашық қолданыстағы urbansound8K және ESC-10 аудио мәліметтер жинағымен тексеру..	89
	ҚОРЫТЫНДЫ	93
	ПАЙДАЛАНЫЛҒАН ӘДЕБИЕТТЕР ТІЗІМІ	95
	ҚОСЫМША А – Төтенше жағдайларды анықтауға арналған модельді оқытуда қолданылатын аудиомәліметтерден белгілерді алу.....	102
	ҚОСЫМША Ә – Ұсынылатын CNN-BiLSTM комбинацияланған архитектурасы.....	103
	ҚОСЫМША Б – Нәтижелерді оқу үрдісіне ендіру актісі.....	104

НОРМАТИВТІК СІЛТЕМЕЛЕР

Диссертациялық жұмыста келесідей мемлекеттік үлгіқалыптарға сілтемелер жасалды

ҚР МЖМБС 5.04.034-2011. Қазақстан Республикасының Мемлекеттік жалпыға міндетті білім беру стандарты. Жоғары оқу орынан кейінгі білім. Докторантура». Негізгі ережелер /ҚР білім және ғылым министрімен бекітілген. «17» маусым 2011ж. №261.

Диссертацияларды және авторефераттарды рәсімдеу бойынша нұсқаулық. ҚР БҒМ, Жоғары аттестаттау комитеті. – Алматы, 2014.

ҚР СТ 1.1-94 ҚР. МӨЖ - терминдер мен анықтамалар.

ҚР СТ МЕМСТ Р 51264-2010. Электрондық байланыс, информатика және сигнализация құрылғылары. Жалпы техникалық шарттар.

МЕМСТ 6.38.90. Сәйкестендірілген құжаттама жүйелері. Ұйымдық өкімдік Құжаттама жүйесі. Құжаттарды ресімдеуге қойылатын талаптар.

МЕМСТ 7.32.2017. Ақпарат, кітапхана және баспа стандарттары жүйесі.

БЕЛГІЛЕУЛЕР МЕН ҚЫСҚАРТУЛАР

ЖЖ	– жоғарғы жиілік
ТЖ	– төменгі жиілік
ЖИ	– жасанды интеллект
Backpropagation algorithm	– қатенің кері таралу алгоритмі
Batch	– пакет
BiRNN	– екі бағыты бар рекурренттік нейрондық желілер
Broadcasting	– кеңею
CCTV	– (closed circuits of television)жабық телекөрсету жүйелері
Chain rule	– тізбектік ереже
CNN	– (convolutional neural network) қабаттық нейрондық желілер
Decision trees	– шешім ағаштары
Depth	– мәліметтер моделі бөлінетін қабаттардың тереңдігі
Depth-wise	– бөлінетін қабаттар
Dropout	– кесу
Feature	– белгі
Feedforward network	– тура таралуы бар желілер
Fine –tuning	– толықтырып оқыту
Freezing	– тоқтатылған қабаттар
Gradient boosting machines	– градиентті бустинг
Gradient descent	– градиентті түсу
Gradient propagation	– градиенттік таралу
Hold-out validation	– іріктемені қарапайым бөлуі бар тексеру
Hypothesis space	– гипотез кеңістігі
Information leaks	– ақпараттың шығуы
Kernel methods	– ядролық әдіс
K-fold validation	– K блоктар бойынша қиылысу арқылы тексеру
Label	– таңба
Layer compatibility	– қабаттардың сәйкестігі
Loss function	– жоғалту функциясы
Loss score	– жоғалту бағасы
MFCC	– (Mel-frequency cepstral coefficient)Mel сүзгілеріне негізделген жиіліктік спектралдық коэффициент
Minibatch SGD	– стохастикалық градиенттік түсу
Naïve Bayes algorithm	– Найв Байес алгоритмі
Overfitting	– артық оқыту
PDF	– (probability distribution functions) болжамның таралу функциясы
Precision	– дәлдік
Probabilistic	– ықтималдық модельдеу

modeling	
Random forests	– кездейсоқ орман
Shallow learning	– таяз оқыту
Space of possibilities	– мүмкіндіктер кеңістігі
STE	– (Short time energy) қысқа уақыттық энергия
SVM	– тірек векторлары әдісі
TFD	– (time frequency domain) уақыттық жиіліктік домен
Underfitting	– жеткіліксіз оқыту
Weight decay	– салмақтардың қысқаруы
Weights	– салмақтар
ZCR	– (Zero crossing rate) нөлдік осьті қию саны

КІРІСПЕ

Жұмыстың өзектілігі. Азаматтардың қауіпсіздігін қамтамасыз ету мемлекет үшін маңызды тапсырма болып табылады. Алайда әлемде және елімізде отбасылық зорлық-зомбылық, ұрлық, қару-жарақпен шабуыл жасау, терроризм, әлеуметтік сипаттағы төтенше жағдайлардың алдын алу өткір проблема ретінде тиімді шешімдерді талап етеді. Әсіресе балаларға және әйелдерге қатысты зорлық-зомбылық бойынша статистикалар соңғы жылдары артуда. Әлемдегі пандемия COVID-19 кезінде бұндай оқиғалардың саны артқандығы белгілі болып отыр. БҰҰ хабарлауынша 300 млн ға жуық 2-4 жас аралығындағы бүлдіршіндер зорлық-зомбылыққа ұшырайды [1]. Тек 2017 жылы 18 жасқа дейінгі 40 150 балалар осы себептен қаза тапқан. Қазақстан ЮНИСЕФ деректері бойынша оқушылардың 44% мектептегі зорлық-зомбылыққа ұшыраған және 24% оқушы өзге балаларға зорлық-зомбылық әрекеттерін жасаған [2]. Сонымен қатар 13-15 жас аралығындағы 150 млн жасөспірім мектеп аумағында зорлық-зомбылыққа ұшырағанын хабарлайды [3].

Әйелдерге қатысты зорлық-зомбылық бойынша қоғам белсенділері дабыл қағуда. 13 мемлекетте өткізілген сауалнама барысында әр екінші әйел өзі немесе жақынының тұрмыстық зорлық зомбылықтың құрбаны болғандығы анықталды. 2020 жылы ҚР реформалары және стратегиялық басқару агенттігінің ұлттық статистикалық бюросы ұсынған ақпарат бойынша 2020 жылы 63 447 оқиға және 2021 жылы 61 464 әйелдерге қатысты тұрмыстық зорлық зомбылық оқиғасы тіркелген [4].

ҚР ІІМ есебі бойынша 2020 жылы 163 226, 2021 жыл 157 884, ал 2022 жылдың 10 айында 140 592 қылмыстық заң бұзушылық: 53 386 ұрлық, 417 зорлық, 2044 бұзақылық және т.б. тіркелген. 2021 жылдың бірінші жартысында елімізде экстремизм және терроризммен байланысты 139 құқық бұзушылық анықталды [5]. Ал әлемде жылына 250 000 адам қару-жарақ себебінен қаза табады [6, 7].

Әлем бойынша және еліміздегі статистика бюролары ұсынған нәтижелерге сүйенсек, қару-жарақ, жасөспірімдер арасындағы және әйелдерге қатысты зорлық-зомбылық, әлімжеттік пен ұрлық, азаматтардың мүліктеріне зиян келтіру және т.б. әлеуметтік сипаттағы төтенше оқиғалардың саны кезек күттірмес проблеманың бар екендігін көрсетіп, қауіпсіздікті қамтамасыз етуге бағытталған тиімді шешімдердің қажет екендігіне нұсқайды.

Бүгінгі күнде автоматтандырылған технологиялар күнделікті адам өмірінің ажырамас бөлігіне айналып отыр. АТ-ды тағы бір қолдану бағыты ретінде қоғамдық орындарда қауіпті оқиғалардың белгілерін анықтау және қырағылықты сақтай отырып, азаматтардың қауіпсіздігін қамтамасыз етуді айтуға болады. Автономды бақылау жүйелерін зерттеу бірінші кезекте оқиғаларды анықтау үшін компьютерлік көрумен тығыз байланысты. Акустикалық мониторинг жүйелерін жабық телекөрсету жүйелерімен (CCTV)

біріктіру, әрекеттерді тану эффективтілігін арттырып, қосымша мәліметтер көзімен қамтамасыз етер еді.

Мамандар 90%-ға жуық физикалық агрессияның алдында айқай, дөрекі сөздер және т.б. түріндегі вербалды агрессияның болатындығын айтады. Яғни аудиомәліметтер арқылы төтенше оқиға туралы құнды ақпараттар мен деректерді алуға болады. Тиісінше, автоматтандырылған аудиосараптау жүйелері, нақты уақыт режимінде жұмыс істей отырып, төтенше оқиғаларды анықтау, алдын-алу, тергеу мақсатында қолданылады.

Қазіргі кезде қауіпсіздікті қамтамасыз ету үшін бейнебақылау жүйелері кеңінен қолданылады. Алайда, бейнебақылау жүйелері кейбір кемшіліктеріне байланысты қауіпсіздікті толықтай қамтамасыз ете алмайды. Мысалы, түнгі уақытта жарықтандыру жүйелерімен жабдықталмаған аумақтарда бейнебақылау жүйелерін қолдану мүмкін емес. Ал экономикалық тұрғыдан аудиоға негізделген жүйелердің қызмет көрсету шығыны, құрылғылар мен құраушы бөлшектерінің құны бейнебақылау жүйелерімен салыстырғанда аз болады. Сонымен қатар, қауіпсіздікті қамтамасыз ету жүйелері нақты уақыт режимінде жұмыс істейтіндіктен, аудио мәліметтер көлемі бейне жазбалармен салыстырғанда аз болады. Тиісінше, өткізу жолағына қойылатын техникалық талаптарды жеңілдейді. Ал жазба жүргізетін құрылғы ретінде акустикалық сенсорлар немесе микрофондар, тіпті смартфондарды да пайдалануға болады. Аудио мәліметтерге негізделген жүйелердің аталған артықшылықтары, мұндай жүйелердің қауіпсіздікті қамтамасыз ету саласында да сұранысқа ие екендігін көрсетеді. Бұл өз кезегінде нақты уақыт режимінде төтенше оқиғаларды анықтауға арналған автоматтандырылған жүйелерді зерттеудің өзекті екендігін айқындайды.

Диссертациялық жұмыстың тақырыбы бойынша жүргізілетін зерттеулердің саны жыл сайын артуда. Аудиосигналдар негізінде төтенше оқиғаларды терең оқыту әдістерін қолдану арқылы анықтау, классификациялау зерттеушілердің ерекше ықыласын аудартып отыр. Себебі дәл қазіргі кезең машиналық оқытудың үш техникалық қозғаушы күштері толық болуы есебінен, көптеген зерттеу жұмыстарының арқауы болуда: өнімділігі жоғары қондырғы; интернеттің дамуы себебінен мәліметтер жиынтығы және тесттердің жеткілікті болуы; алгоритмдік жетістіктер. Мысалы, түрлі семинарлар ұйымдастырылып қойылған тапсырмалар бойынша қатысушыларға өз шешімдерін ұсынуға және тестілеуге мүмкіндік беріледі. Солардың ішінде ең белгілілерінің бірі DCASE challenge-ді атауға болады. Іс-шарада классификациялау және анықтау бойынша дыбысқа байланысты тапсырмалардың алуан түрлерін қарастыру мақсатында, тапсырмалар мен мәліметтер жинағын жаңарта отырып, әр жыл сайын өткізіледі. Аталған тапсырмалардың арасында дыбыстық оқиғалар мен олардың орнын анықтау, дыбысты белгілеу, қоршаған орта дыбыстарының сараптамасы, акустикалық оқиғалардың классификациясы, сонымен қатар сирек кездесетін дыбыстық оқиғаларды анықтау сияқты тапсырмаларды кездестіруге болады. Алайда мұндай семинарларда ұсынылатын тапсырмалар нақты төтенше оқиғаларды анықтауға бағытталмағандықтан, төтенше

оқиғаларды неғұрлым кең сипаттайтын аудио мәліметтер жинағы жасалмаған. Ал өзге зерттеушілердің жұмыс нәтижелері тек аздаған класстар санымен шектеледі. Сондықтан аудиосигналдарды өңдеу негізінде төтенше оқиғаларды бірнеше класстар анықтауға арналған жүйелер зерттеуді талап етеді.

Зерттеу пәні – төтенше оқиғаларды анықтау үшін аудиосигналдарға сараптама жүргізу және аудио мәліметтерді алдын-ала өңдеу әдістері, нейрондық желі архитектуралары мен алгоритмдері.

Жұмыстың мақсаты – нақты уақыт режимінде жұмыс істей отырып, аудиосигнал импульстары негізінде төтенше оқиғаны анықтауға және классификациялауға арналған, нақтылығы жоғары модельді құру.

Зерттеудің міндеттері. Зерттеу жұмысының мақсатына жету үшін төмендегідей тапсырмалар қойылды:

1. Аудиосараптама негізінде төтенше оқиғаны анықтауға арналған аудио мәліметтерді өңдеу әдістері мен модельдеріне салыстырмалы талдау жүргізу.

2. Аудиосигналдар негізінде төтенше оқиғаны анықтауға арналған модельді оқытуға қолданылатын аудио мәліметтер жинағын (датасет) жасау.

3. Аудиосигнал импульстары негізінде төтенше оқиғаларды анықтауға және мультиклассификация тапсырмасын орындауға арналған терең оқыту моделін құру.

4. Құрылған модельді тестілеу және бағалау.

Зерттеу әдістері. Зерттеу жұмысында қойылған тапсырмаларды орындау үшін машиналық оқыту, аудиосигналдарды цифрлық өңдеу, аудио мәліметтерді жазудың акустикалық әдісі қолданылды.

Зерттеу объектісі. Аудиосигналдар.

Зерттеу жұмысының ғылыми жаңалығы мен маңыздылығы. Төтенше оқиғалар класстары бойынша жинақталған мәліметтер жиынтығын қолдана отырып оқытылған, нақты уақыт режимінде жұмыс істеуге бағытталған, төтенше оқиғаларды аудиосигнал импульстары негізінде анықтап, мультиклассификация тапсырмасын орындайтын, нақтылығы жоғары аудиомәліметтерді өңдеуге арналған CNN-BiLSTM комбинацияланған нейрондық желі архитектурасы жасалды. CNN-BiLSTM комбинациясын пайдалану арқылы мәліметтерді өңдеу жылдамдығы арттырылды.

Зерттеудің практикалық маңыздылығы. Ұсынылған модель темір жол, әуе, көлік бекеттерінде, сауда-ойын сауық орталықтарында, білім беру мекемелерінде, тұрғын үй кешендерінде және т.б. қоғамдық орындар мен инфрақұрылымдарда аудиосигналдарды өңдеу негізінде төтенше оқиғалардың сегіз класын анықтап, тиісті мекемелерге алдын-алу туралы хабар береді.

Алынған нәтижелердің ғылым мен технологияның дамуына әсері.

Зерттеу жұмысы барысында алынған мәліметтер жиынтығы мен нәтижелерді аудиосигналдар негізінде төтенше оқиғаларды анықтау бойынша зерттеу жұмысын жүргізуде, мәліметтерді жинақтау және алдын ала өңдеу, нейрондық желі модельдерін таңдау және оқыту барысында, әдістемелік нұсқаулық ретінде өзге зерттеушілер пайдалана алады. Диссертациялық жұмыс барысында алынған нәтижелер бойынша дайындалған автореферат «Қазақстан

Республикасы ПМ М.Есболатов атындағы Алматы академиясының» оқу үрдісіне енгізілді.

Автордың жеке үлесі. Диссертациялық ғылыми зерттеу жұмысындағы барлық нәтижелерді автордың өзі алды. Зерттеудің мақсатына жетудегі қойылған тапсырмаларды бекіту, зерттеу әдістерін талдау және ұсынылған жүйені енгізу, ғылыми зерттеудің нәтижелерін талдау автор, оның ғылыми жетекшісі және шетелдік ғылыми кеңесшісінің бағыт-бағдарымен бірге орындалды. Зерттеу жұмыстары барысында пайдаланылған өзге зерттеушілердің нәтижелеріне тиісті әдебиеттерге сілтемелер қойылды.

Зерттеу жұмысының басқа ғылыми-зерттеу жұмыстарымен байланысы. Зерттеу жұмысының тақырыбы бойынша 2022-2024 жылдарға арналған «Жас ғалым» жобасы бойынша жас ғалымдардың зерттеулерін гранттық қаржыландыру бойынша «AP149715555 Машиналық оқыту әдістерін қолдану арқылы нақты уақыт режимінде ғимарат ішіндегі қауіпсіздікті қамтамасыз ету жүйесін жобалау және енгізу» жобасы орындалуда.

Қорғауға шығарылатын нәтижелер.

1. Төтенше оқиғаларды сипаттайтын сегіз түрлі класс бойынша мәліметтер жиынтығы.

2. Аудиосигналдарды өңдеу негізінде, төтенше оқиғаларды сегіз түрлі класс бойынша 83-95% аралығындағы нақтылықпен анықтайтын нейрондық желі моделі.

3. аудио мәліметтерді өңдеу жылдамдығын арттыруға мүмкіндік беретін CNN-BiLSTM комбинацияланған нейрондық желі архитектурасы.

Жарияланымдар. Зерттеу жұмысының нәтижелері бойынша 8 ғылыми мақала, Scopus деректер базасына енетін нөлдік емес импакт-факторы бар халықаралық рецензияланатын ғылыми журналдарда 2 мақала (процентиль 79, 31%), Scopus деректер базасына енетін конференцияларда 2 мақала, ҚР Ғылым және жоғары білім министрлігінің Ғылым және жоғары білім саласындағы сапаны қамтамасыз ету комитеті ұсынған басылымдарда 3 мақала, басқа халықаралық конференцияларда 1 мақала жарияланды.

Зерттеу жұмысының нәтижелерін апробациялау. Диссертациялық зерттеу жұмысының негізгі нәтижелері Қ.И. Сәтбаев атындағы ҚазҰТЗУ (Қазақстан, Алматы) "Радиотехника, электроника және ғарыштық технологиялар" кафедрасының семинарларында баяндалып, талқыланды; "Сәтбаев оқулары-2020" халықаралық ғылыми-практикалық конференциясы, Satbayev University, (ҚР, Алматы, 2020); 2020 8th International Conference on Information Technology and Multimedia (ICIMU) – IEEE, (Selangor, Malaysia 2020); 13th International Conference, ICCCI 2021, (Kallithea, Rhodes, Greece, 2021) ұсынылып, нәтижелер Springer International Publishing, 2021 баспасында жарияланды.

Диссертациялық жұмыстың құрылымы мен көлемі. Диссертациялық жұмыс кіріспеден, 4 бөлімнен, қорытындыдан, пайдаланылған әдебиеттер тізімінен және 2 қосымшадан тұрады. Жұмыстың жалпы көлемі 104 беттен, 56 суреттен, 6 кестеден, пайдаланылған әдебиеттер тізімі 99 атаудан тұрады.

1 ТӨТЕНШЕ ЖАҒДАЙЛАРДЫ АУДИО АРҚЫЛЫ АНЫҚТАУ БОЙЫНША ӘДЕБИЕТТЕРГЕ ШОЛУ

1.1 Қауіпсіздікті бақылау жүйелері

Елімізде және әлемде орын алып жатқан зорлық-зомбылық, нысандарға қару жарақпен шабуыл жасау сияқты тұрғындардың мүлкі мен денсаулығына зиян келтіретін оқиғалар қалаларда, тұрғын үй кешендері, әуежай, метро, көлік және теміржол станциялары, сауықтыру және оқу орындары, сауда және ойын-сауық орталықтары сияқты халық көп шоғырланатын қоғамдық орындардағы қауіпсіздікті қамтамасыз ету жүйелерінің маңыздылығын айқындайды. Аталған фактілердің алдын алу немесе себептерін анықтауда заманауи шешімдерді қолдану маңызды.

1.1-суретте соңғы бес жылдағы диссертациялық жұмыс тақырыбы бойынша ScienceDirect Elsevier базасында жарияланған мақалалардың саны ұсынылады. 1.1-суреттен жыл сайын жарияланымдардың артатындығын байқауға болады. Бұл зерттеу жұмысы тақырыбының зерттеушілердің қызығушылығына ие және өзекті екендігін көрсетеді.



Сурет 1.1 – ScienceDirect Elsevier-де соңғы 5 жылда жарияланған мақалалар саны

Ескерту – Кілттік сөздер: emergency sound, event detection, sound event classification

Syed A.S. және өзге авторлар [8] шолу жұмысында «Ақылды қалалардағы IoT» технологияларын қолдануды ұсынады. Аталған жұмыста түрлі салаларда қолдануға арналған технологиялардың негізгі архитектурасын сипаттай отырып, сенсорлардың «Ақылды қала» технологиясында ерекше орын алатындығын айтады.

Автоматтандырылған технологиялар күнделікті адам өмірінде кеңінен қолданыс табады. АТ қолданудағы тағы бір мақсат ретінде қоғамдық орындарда

қауіптің кез-келген белгілерін анықтау және қырағылықты сақтау отырып азаматтардың қауіпсіздігін қамтамасыз етуді айтуға болады. Автономды бақылау жүйелерін зерттеу бірінші кезекте оқиғаларды анықтау үшін компьютерлік көрумен тығыз байланысты. Акустикалық мониторинг жүйелерін жабық телекөрсету жүйелерімен (CCTV) біріктіру әрекеттерді тану эффектілігін арттырып, мәліметтердің қосымша көзімен қамтамасыз етер еді.

Қазіргі кезде қауіпсіздікті қамтамасыз ету үшін бейнебақылау жүйелері кеңінен қолданылады [9, 10]. Алайда, бейнебақылау жүйелері кейбір кемшіліктеріне байланысты қауіпсіздікті толықтай қамтамасыз ете алмайды. Мысалы, түнгі уақытта жарықтандыру жүйелерімен жабдықталмаған аумақтарда бейнебақылау жүйелерін қолдану мүмкін емес. Экономикалық тұрғыдан аудиоға негізделген жүйелердің қызмет көрсету шығыны, құрылғылар мен құраушы бөлшектерінің құны бейнебақылау жүйелерімен салыстырғанда аз болады. Сонымен қатар, қауіпсіздікті қамтамасыз ету жүйелері нақты уақыт режимінде жұмыс істейтіндіктен, аудио мәліметтер көлемі бейне жазбалармен салыстырғанда аз болады. Тиісінше, өткізу жолағына қойылатын техникалық талаптарды жеңілдейді. Аудио мәліметтерге негізделген жүйелердің аталған артықшылықтары, мұндай жүйелердің қауіпсіздікті қамтамасыз ету саласында да сұранысқа ие екендігін көрсетеді [11, 12].

Күдікті оқиғаларды тану тапсырмасын орындау үшін машиналық оқыту алгоритмдерін қолдануға болады. Машиналық оқытуда күдікті оқиғаны анықтау бойынша кең таралған бірнеше тапсырмалар бар: әлеуметтік желі немесе БАҚ-ғы хабарламалардан күдікті мәтіндерді анықтау [13, 14], бейне бойынша эмоцияларды тану арқылы күдікті оқиғаны анықтау, мазмұнын тану [15-17] және аудио сенсорлар мәліметтері негізінде анықтау және т.б.

Құқық қорғау органдары қару-жарақпен байланысты қылмыстарды тез және тиімді тергеу үшін қаладағы қауіпті дыбыстарды анықтау жүйелерін пайдалы әрі замануи құрал ретінде қолдана алады. Қауіпсіздікті қорғау саласында ақпараттық технологияларды енгізу, белгісіз оқиғаларды, энергия деңгейі детекторын қолдана отырып, аудиожазбаларды тану блогы арқылы оқиғаларды анықтауға мүмкіндік береді.

1.2 Аудиоклассификация тапсырмасын орындау сатылары

Аудиоклассификация және тану негізгі үш бағытта кең қолданылуда: дыбыстық оқиғаны тану, дыбыстық мәтінді автоматты тану, әуеннің категориясын анықтау.

Жоғарыда келтірілген тапсырмалармен салыстырғанда аудио негізінде дыбыстық оқиғаны анықтау күрделі тапсырма болып табылады. Біріншіден, тапсырмалардың құрылымдарында айырмашылық бар. Екіншіден, жабық ғимаратта және ашық ортада дыбысты жазатын құрылғы аудио сенсоры мен дыбыс көзінің арақашықтығы алыс болуы себебінен сигнал/шуыл қатынасы аз болады [18]. Негізгі дыбыстан бөлек қосымша қабатталған дыбыстардың және дыбыс көздерінің болатындығы белгілі. Осындай құрылымдар алдыңғы

тапсырмалар мен дыбыс негізінде оқиғаны анықтау тапсырмасының ерекшеліктерін айқындайды.

[19] жұмыста Nogueira A.F.R. және өзге авторлар үлкен қалалар мен ғимараттардағы, қоғамдық орындардағы күдікті деп танылған дыбыстарды анықтау, азаматтардың қауіпсіздігін қамтамасыз етуде және әлеуметтік сипаттағы төтенше оқиғалардың алдын-алуда аудио негізіндегі тану тапсырмасын маңызды деп танып, нәтижелерге шолу жасаған.

Бір немесе бірнеше орыннан түрлі уақыт мезетінде алынған дыбыстарға автоматты классификация орындауға арналған есептеу алгоритмін жасау, нақты дыбыс жиынтықтарын классификациялау және дыбыс сипаттамаларын алу бойынша жұмыстар жасалған [20-22]. Алайда бірнеше дыбыс класстарымен бірге жұмыс істеуде, қалыпты емес шуыл шарттары кезінде және дыбыстық оқиғалардың қабаттасуының кең спектрімен жұмыс бойынша шектеулердің бар екендігін атап өткен жөн.



Сурет 1.2 – Дыбысты классификациялау процесі

Аудио дыбыстарға классификация жүргізу негізгі үш қадамды орындау арқылы жүзеге асырылады (1.2-сурет). Бірінші саты алдын ала өңдеу деп аталады. Екінші қадамда аудио мәлімет белгілерін (feature extraction) алуды орындау нәтижесінде, шағын параметрлері алынады. Классификация модельдерін тиімді оқыту үшін қолданылатын белгілерді таңдау және алу, күрделілігі жоғары әрі маңыздылығы зор тапсырма болып табылады. Бастапқы кезеңдерде аудиомәліметтердің белгілері қолмен анықталатын. Ал қазіргі уақытта тиімді әдіс ретінде жиі қолданылатын терең оқыту архитектураларында бұл кемшілік жойылған. Үшінші қадамда, жинақталған параметрлер негізінде ең қарапайым болып табылатын Эвклидтік аралық әдісінен, күрделі статистикалық әдістерді қолдану арқылы, классификация жүргізу және категорияларды анықтау тапсырмалары орындалады. Аудио классификация және категорияны анықтауға арналған модельдің тиімділігі, тиісінше алынған аудио белгілерге (feature) және классификациялау нақтылығына тәуелді. Яғни, дыбыстарды классификациялау екі негізгі компонентті жүйелендіруден тұрады: неғұрлым тиімді акустикалық белгілерді қолдану және жоғары нәтиже көрсететін классификаторды енгізу.

1.3 Аудио мәліметтерді алдын-ала өңдеу

Алдын-ала өңдеу сатысы аудио сигнал сипаттамаларын айқындау арқылы әдістің сенімділігін арттыруға және сараптаманы жеңілдетуге жауап береді. Бұл кезеңде шуыл деңгейін азайту, тегістеу, ТЖ сүзгілеу, аудиосигналды дыбысы бар және дыбыссыз оқиғаларға сегменттеу түріндегі тапсырмалар орындалады

[21, p. 1883-1-1183-12]. Аудиосигналды дұрыс және сенімді ұсыну үшін кіріс аудиосигналдарға алдын-ала өңдеу жасау аса маңызды. Микрофонның көмегімен жазылған аудиосигналда фондық шуылдар мен кедергі болатын акустикалық объектілер кездестіндігі айқын. Мұндай дыбыс машиналық оқыту негізіндегі классификациялауға кіріс мәлімет ретінде бірден қолданыла алмайды. Себебі сигналдарда жоюды талап ететін асқындылық (redundancy) болады [22, p. 3206-1-3206-18].

Subramanian A.S. және өзге авторлар жүргізген зерттеу жұмысында [23] тану нақтылығын арттыру үшін акустикалық сәулені немесе дереверберация тудыру және көпарналы аудиосигналды акустикалық әдіс интеграцияланған препроцессорды қолданған.

Kim N.K. және өзге авторлар [24], сонымен қатар Almaadeed N. және өзге авторлар [25] ұсынған екі жұмыс туннельдегі оқиғаларды анықтау бойынша зерттеу жұмыстарын жүргізіп, алдын-ала өңдеу сатысында акустикалық оқиға сигналын фондық сигнал шуылынан бөлу үшін негативті емес тензорлық факторизация (NTF) технологиясын қолданған. Осылайша нейрондық желілер оқыту барысында кездесетін шуылдарды азайтуға мүмкіндік береді. Аталған жұмыста шуылды онлайн зерттеу мүмкіндігінің болуы есебінен өнімділік жоғары болады.

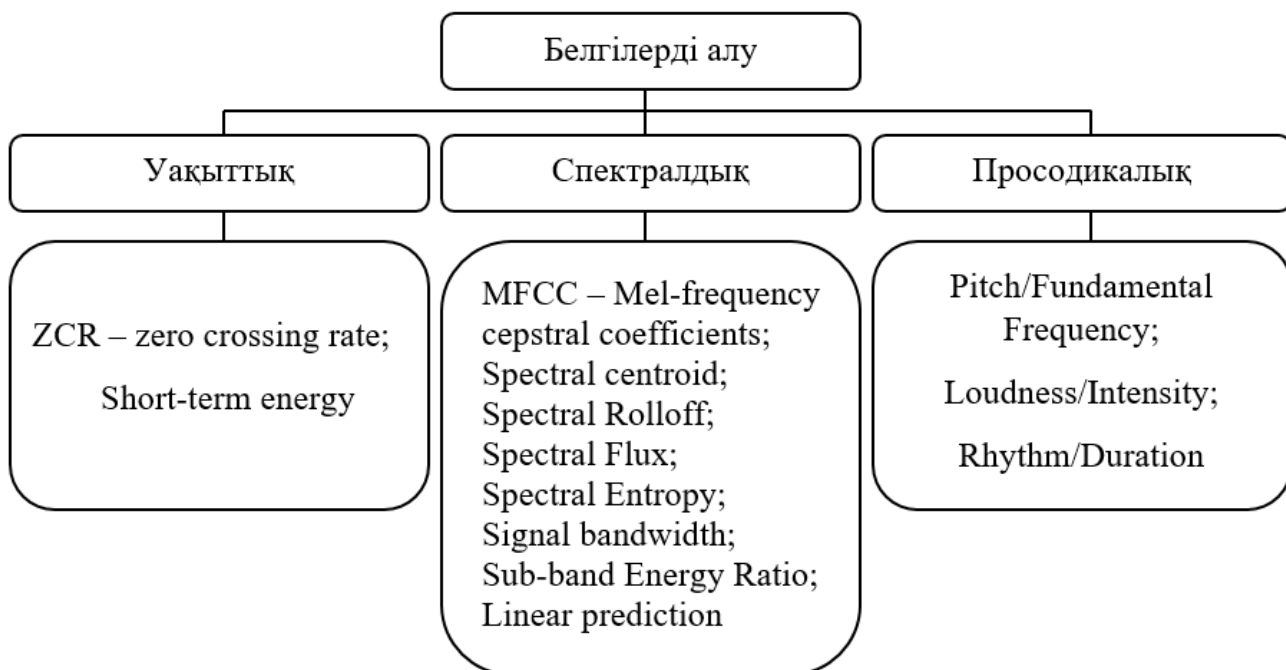
Phetcharat Parathai және өзге авторлар ұсынған [26] жұмыста оқиға шуылдары қабатталуынан тұратын қоспаны классификациялау үшін дыбыстық оқиғаларды бөлуді орындаған. Дәстүрлі теріс емес комплекстік матрицалар факторизациясы (CMF) L1 оптималды адаптивті кесумен біріктірілген. Ұсынылған CMF адаптивті алгоритм спектрлер құрылымын кодалайды және кіріс сигналдардың фазасын жиіліктік-уақыттық ұсынуға бағалайды. Басты ерекшелігі уақыттық декомпозиция процесін эффективті жылдамдатуында болып табылады. Эксперименталдық нәтижелер қолданылған әдіс ең замануи әдістерден тиімді екендігін растады.

1.4 Аудио мәліметтерден белгілерді алу

Белгілерді алу аудио контент сипаттамасында және оны бағалауда маңызды рөл атқарады. Идеал белгі сипаттамалары келесідей:

- жеңіл икемділік;
- шуылда тұрақтылық;
- жүзеге асыру жеңілдігі;
- қажетті тегістеу сипаттамаларына ие болуы.

Белгілер кеңістігі өлшемдерінің саны, алынған белгілер санына тең. Егер алынған белгілер саны көп болса, өлшемділік проблемасы туындайды. 1.3-суретте белгілердің классификациясы ұсынылған.



Сурет 1.3 – Белгілер классификациясы

Белгілерді алу сатысында өңделген аудиосигнал атрибуттық белгілерге түрлендіріледі. Әр белгі белгілер кеңістігіндегі вектор элементі болып табылады. Әдетте, дыбыс белгілері, қарастырылатын аудиосигналдарды хэмминг терезесі бар кадрларға бөлу арқылы алынады. Белгілер жиынтығы әрбір кадр үшін алынып, оқыту және тестілеу жүргізіледі. Соңғы зерттеу жұмыстарында жиі қолданылып жүрген белгілер:

1. Жиіліктік белгілер: қанықтылық белгілері (chroma features), қанықтылыққа негізделген қысқа уақыттық Фурье түрлендіруі, тоналды центроид [27, 28], спектралдық контраст белгілері [29].

2. Mel сүзгілеріне негізделген белгілер: MFCC (Mel-frequency cepstral coefficient), Mel, Log-Mel спектрограммасы [26, p. 4368-15; 30].

3. Gammatone сүзгілеріне қатысты белгілер: олар тәуелді белгілер ретінде де белгілі. GFCC (Gammatone frequency cepstral coefficient) ең жиі қолданылатындарының бірі [31].

4. Толқындық формаға негізделген белгілер: wavelet [32].

Уақыттық белгілер немесе уақыт амплитудалары, уақыт бойынша амплитуданың өзгеруі түрінде ұсынылады (толқын формасындағы сигнал). Уақыттық дыбыстық белгілер, алдыңғы мәліметтерді қолданбай, тікелей аудио сигналдардан алынады. Уақыттық белгілердің мысалдары ретінде нөлдің қиылысу жылдамдығы, амплитудаға және қуатқа негізделген белгілер. Мұндай белгілер, әдетте, аудио сигналдарды зерттеудің қарапайым әдісі болып табылғанымен оларды спектрлік сипаттамалармен бірге қолданған тиімді. Демек, уақыттық белгілердің есептеу күрделілігі спектрлік белгілерге қарағанда төмен:

1. STE (Short time energy) – N сигнал үлгілері квадраттарының қосындысы ретінде анықталатын $x[n]$ дискреттік сигнал [26, p. 4368-15; 33]:

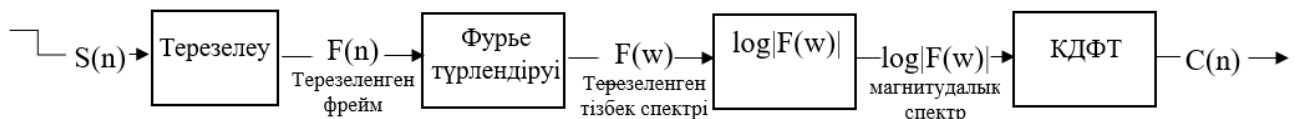
$$STE = \frac{1}{N} \cdot \sum_{n=0}^{N-1} x^2[n] \quad (1.1)$$

2. CR (Zero crossing rate) – бірлік уақыт мезетінде функция таңбасының ауысу саны (функцияның ОХ осін қиып өтуі). Басқаша айтқанда, сигналдың оңнан теріске және керісінше өгеру санын кадрдың жалпы ұзындығына бөледі. Белгілі бір деңгейде бұл белгі сигнал спектрінің спецификасын білдіреді, сондықтан ол сигналдың спектралдық табиғатына жақын. ОХ осін қию саны үлкен болған жағдайда сигнал ақпараттары жоғары жиілікті, төмен болса төменгі жиілікті ақпараттардан тұрады. ZCR сигналдың жиіліктік құрамы туралы мағлұматты қамтиды және шуылдың шамасын көрсетеді [26, p. 4368-15; 32, p. 107050-5; 33, б. 71]:

$$Z^{(t)} = \sum_n |sign[x[n]] - sign[x[n+1]]| \omega(m - n) \quad (1.2)$$

мұнда $\omega(m) = \begin{cases} \frac{1}{2N} & 0 \leq m \leq N - 1 \\ 0 & otherwise \end{cases}$, ал $sign$ таңба функциясы болып табылады.

Аудиосигналдар, оның ішінде дыбыстық мәтіндер, сөйлеуші, тілді анықтау және т.б. қысқа уақыттық спектралдық сипаттамалар көмегімен алынған спектралдық/кепстралдық белгілерге негізделеді. Кепстралды есептеу үш процесстен тұрады: Фурье түрлендіруі, логарифмдеу және кері Фурье түрлендіруі. Бұл базалық жиілікті анықтап, аудиосигналды дискретті түрде тазалауға мүмкіндік береді. 1.4-суретте аудиосигналды кепстралдық облыста ұсыну түрлендіруіне қатысты сатылар көрсетілген. Аудиосигнал әдетте алдынала анықталып, плавный терезе функциясы көмегімен көбейтіледі. Терезе функциясы дискреттік Фурье түрлендіруі нәтижелерінің шектеулі ұзындығы себебінен қажет. Дискреттік Фурье түрлендіруі қарапайымдылығы мен өнімділігі есебінен жиі қолданылады. Процесс маңыздылығы аз болғандықтан тек амплитуда спектрі сақталады. Жылдам Фурье түрлендіруі аудиосигналды жиіліктік элементтерге бөледі.



Сурет 1.4 – Кепстралдық есептеудің блок диаграммасы

3. MFCC (Mel-frequency cepstral coefficient), дискреттік косинустық түрлендіру негізінде анықталады. Ол дыбыстарды классификациялауды жүзеге асыруда, көбінесе, дыбыстық мәтіндерді тануда кеңінен қолданылады.

MFCC-ді алу үшін келесі өңдеулерді орындау арқылы жүзеге асырылады: сигнал кадрлерге бөлініп, Хемминг терезесі функциясы қолданыла отырып әр

кадр үшін амплитудалық спектр алынады. Осы спектрлердің логарифмі анықтағаннан соң, Mel шкаласына түрлендіріледі. Ақырғы қадамда дискреттік косинустық түрлендіру жүзеге асырылады [26, p. 4368-15; 33, б. 71]:

$$c_n = \sqrt{\frac{2}{K}} \cdot \sum_{k=1}^K (\log S_k) \cdot \cos\left(\frac{1}{K} \cdot (n(k-0.5)\pi)\right) \quad (1.3)$$

мұнда K – жолақтық сүзгінің саны.

4. Сигналдың таралу уақыты ортасындағы нүкте немесе спектралдық энергияның ортасы спектралдық центроид (spectral centroid) деп аталады. Ол уақыттың өтуімен басым сигналды жиілігінің өзгерісін ұсынатын шуылды сенімді бағалауды қамтамасыз етеді. Дыбыстық мәтіндердегі сигналдарды өңдеуде кең таралған құрал [25, p. 1858-5]:

$$SC_{(f)} = \frac{\sum_{n=1}^K n|Y_x[n]|}{|Y_x[n]|} \quad (1.4)$$

Жиілік диапазонының максималды қуаты:

$$MP_{(f)} = \sum_n^K |Y_x[n]|^2 \quad (1.5)$$

мұнда Y_x – сигналдың уақыт доменіндегі x -ші кадрының ФТ.

5. Spectral Rolloff спектр магнитудасы орналасқан шамдағы жиілікті есептейді. Сонымен қатар, спектралдық форманың ассиметриясын есептейді [25, p. 1858-5]:

$$SR_{(f)} = \lambda \sum_{n=1}^K |Y_x[n]| \quad (1.6)$$

мұнда λ – шектік деңгейді білдіреді;

Y_x – сигналдың уақыт доменіндегі x -ші кадрының ФТ.

6. Spectral Flux аудиосигнал қуаты спектрінің өзгеру жылдамдығын және қорытқы спектрлер бойынша барлық спектр амплитудасының тұрақтылығына түрленуін есептейді [25, p. 1858-5]:

$$SF_{(f)} = \sum_{n=1}^{K/2} |Y_x[n] - Y_{x-1}[n]| \quad (1.7)$$

7. Spectral entropy ақпарат көзінің белгісіздік орташасы ретінде түсіндірілетін ақпараттық мазмұнды өлшейді [25, p. 1858-5]:

$$SE_{(f)} = - \sum_{n=1}^K S_x[n] \log_2([S_x[n]]), \quad (1.8)$$

$$S_x[n] = \frac{|Y_x[n]|^2}{\sum_{n=1}^K |Y_x[n]|^2} \quad (1.9)$$

8. Спектрдің орталық нүктесі айналасындағы буынның сигнал ені жиілігі сигналдың өткізу жолағы (Signal bandwidth) деп аталады. Буын өлшемі ДФТ кадрларының өткізу қабілетінің орташасы ретінде есептеледі.

9. Sub-band Energy жалпы энергия жолағының энергия диапазонын есептеу үшін сипаттама ретінде қолданылады. Дауыстық сигналдың энергетикалық спектрі бірінші диапазонда орналасқан.

10. Linear prediction ДФТ үшін спектрді бағалаудың тиімді әдісі. Ол уақыттық және жиіліктік облыста аудиосигналдардағы асқындылықты қолдануды түсіндіреді.

11 Predictor coefficient жеке белгі ретінде қолданылады. Алайда, ол неғұрлым сенімді, LPCC (linear predictive cepstral coefficient) және PLP (perceptual linear prediction) аз түзетілген белгі.

Просидикалық белгілер немесе қабылдаудың жиіліктік белгілер адам контекстіндегі семантикалық мағынаға ие ақпараттарды нұсқайды. Ал физикалық белгілер аудиосигналдарды математикалық, статистикалық және аудиосигналдың физикалық қасиеттері терминінде сипаттайды. Олар дыбыстың семантикалық мәнді аспектілеріне сәйкес ұйымдастырылған, яғни тон биіктігі/негізгі жиілік, жоғарылығы/интенсивтілік және ритм/ұзақтығы:

– тон биіктігі/ негізгі жиілік супрасегменталдық сипаттама және дыбыстық және дыбыстық мәтіндердің маңызды просидикалық қасиеті болып табылады. Мәліметтер неғұрлым ұзақ уақыт масштабтарында өзге сегментарлық аудиокорреляторлармен қабатталып таратылады, мысалы спектралдық иілу белгісі. Тиісінше тон биіктігі шамасын қолдану үшін, тұтас аудиосигнал бойынша тон биіктігінің глобалды статикасын аппроксимациялау мүмкін болады.

– жоғарылығы/интенсивтілік әр аудиосигналдың дыбыс жоғарылығын (энергия) модельдеп, түрлі кідірістердегі дыбыс амплитуасын есептеу арқылы адам құлағымен тану жолын имитациялайды. Осылайша екі негізгі сипаттаманы есепке алуға болады. Біріншіден, стимул интенсивтілігінің өсу уақытында, есту реакциясы логарфимдық түрде артады. Екіншіден, дыбысты түсіну спектралдық таралу мен ұзақтығына байланысты болады. сонымен қатар, дыбыс жоғарылығы функциясы белгілілікке негізделген және жоғарылық контуры векторына біріктірілген;

– ритм/ұзақтығы дыбысталған және дыбысталмаған бөліктердегі уақыттық қасиеттерді өңдеп, уақыттық перспективаларды модельдейді.

Түсінікті болу үшін аудио құрылымдардың уақыт детализациясындағы бірліктерін анықтаған жөн:

– audio frame (аудиокадр): аудионың тұрақсыз кездейсоқ процесс екендігі, ал оның сипаттамалары уақыт бойынша өзгертіндігі белгілі. Бірақ аталған сипаттамалар баяу өзгереді. Сол себептен де дыбыстық сигналды қысқа периодтарға бөлу арқылы өңдеу өте тиімді. Мұндай сегменттердің ұзындығы 20-30 мс және бұл бөлшектер дыбыс өңдеудің ең кіші бірлігі ретінде белгілі [33, б. 71; 34];

– аудиоклип: жоғарыда атап өткен аудиокадрлардың өлшемі өте кіші болғандықтан олардың семантикалық сипаттамаларын алу қиындық туғызады. Сол себептен уақыт бірлігі үлкен жаңа аудио бірлігі (audio clip) қолданылады. Олар, өз кезегінде бірнеше аудиокадрлардан тұрады. Аудиоклиптердің сипаттамалары аудиокадрлар арқылы анықталады [33, б. 71; 34, р. 3293-9];

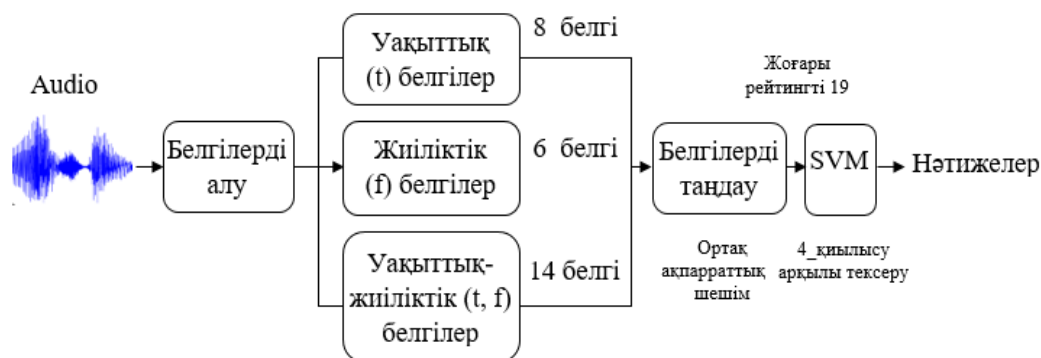
– audio shot (аудиобөлшек): Бұл видеомәліметтерден келіп шығатын ұғым. Аудиоклиптің ұзақтығы өте қысқа болуы себебінен, құрамдық семантикалық анализ жасауда қолдану ыңғайсыздық тудырады. Аудиобөлшек, бірдей санатқа тиесілі, бірнеше аудиоклиптерден тұрады [33, б. 71; 34, р. 3293-9];

– жоғары деңгейдегі семантикалық аудиобірлік: аудиобөлшектердің түрлі комбинацияларынан құралады. Тиісінше олар толыққанды семантикалық құрылымға ие аудиоқұрылым бірлігі [33, б. 71; 34, р. 3293-9].

Ал [11, р. 1822-2-1822-6] жұмыста зерттеушілер түрлі белгілерді біріктіру жеке белгілерді қолданғанға қарағанда жақсы нәтиже алуға мүмкіндік беретіндігін атап өткен. Олар LM, Mel, Mel сүзгілеріне негізделген белгілерді, сонымен қатар осы белгілерден алынған L2M, L3M белгілерін қолданған. Зерттеу барысында бұл белгілер біріктірілген және жинақталған түрінде қолданылған.

Son J.-Y. және Chang J.-H., ұсынған жұмыста [18, р. 6718-7] Log-Mel спектрограммасын қолданады. Ұзындығы 40 мс болатын терезе, секіру ұзындығы 20 мс, Хэмминг терезесінің өлшемі 40 мс және Фурье түрлендіруінің өлшемі 2048 нүктені құрады. Log-Mel спектрограммасы 128 өлшемді белгілерді алды. 10 секундты құрайтын мәлімет жиынтығынан 500 кадр алынған. Оқытылатын модель кірісіне Log-Mel спектрограммасы объектілер картасы $128 \times T$, яғни T кадрлар саны және ол 64-ке тең деп алынған.

[25, р. 1858-4] жұмыста әр аудиосегмент үшін 28 белгі сарапталған. Белгілерді алу үшін қолданылған әдіс 1.5-суретте көрсетілген. Қолданылатын белгілерді үш категорияға бөлуге болады: t-домені белгілері; f-домені белгілері; біріккен (t, f) домендері белгілері. t-домені белгілері уақыттық облыстан, f-домені белгілері спектралдық облыстан, (t, f) домендері белгілері аудиосигнал TFD-нан алынады. Спектралдық және уақыттық сигнал сипаттамалары белгілерді алуға, негізгі сигналды өңдеуге және сараптауға қолданылды. Жұмыста орташа мәні, дисперсия, ассимметрия, энтропия, спектралдық ағын, орташа жиілік және т.б. уақыттық және жиіліктік белгілер таңдалған.



Сурет 1.5 – Белгілерді алуға арналған әдістің орындалуы

Тәжірибеде 32 кГц жиіліктегі аудиосигналдар іріктеліп, ұзындығы 0,75 с сегменттерге бөлінген. Сигналдар 200 мс және 50% қабатталуы бар Хэмминг терезесін қолдану арқылы кадрларға бөлінген. Зерттеуде уақыттық және спектралдық сипаттамаларға негізделген төменгі деңгей белгілерінің жиынтығы қолданылған.

Аудиосигналдарды өңдеу сигналдың нақты уақыт режиміндегі және стационарлы емес сипаттарын ескере отырып, кадрлар бойынша тізбектік ретпен өңделеді. Іс жүзінде, дыбысты өңдеу үшін сигнал кідірісін қабылдау маңызды және ақырғы қолданушы қажетті төменгі деңгей қамтамасыз етілуі керек [35, 36]. Сонымен қатар, бейне мәліметтермен салыстырғанда, аудиосигналдар бірнеше мс уақыттық масштабтағы кенет өзгерістерді көрсететіндіктен стационарлы емес қасиетке ие. Сол себептен, аудиоағын шағын және бөліктері бір-бірімен қабаттасатын белгілі уақыт периодтары болатын сегменттерге бөлінеді. Тиісінше, төменгі- және жоғарғы жиіліктік сигнал құраушыларын сараптауда ортақ шешім табу керектігін ұмытпаған жөн [25, р. 1858-3; 37]. Егер TF тым ұзын болса өңдеуде сигналдың спектралдық құрамының жылдам өзгерістері ескерілмей қалады. Жүйе ЖЖ коипоненттерді олардың ұзақ уақыт интарвалы бойында нормализациянатындықтан ескерілмей қалады. Екінші жағынан, кадрдың тым қысқа өлшемі сигналдың ТЖ құрауыштарын қарастыруға қажетті қолжетімділікті қамтамасыз ете алмайды. TF қысқа слоты қажетті белгілерді алу үшін жеткілікті ақпаратты ұсына алмайды. Бұл сигнал сараптамасының біркелкі емес әрі тануда қателіктерге ұшырауына жол береді [38]. Сигналдың спектралдық құрамы неғұрлым жылдам өзгерген сайын кадр ұзындығы соғұрлым қысқа болуы тиіс. Сигнал кадрларға сегменттелгеннен кейін кадр деңгейінің сипаттамалары уақыттық және жиіліктік облыстарда есептеледі.

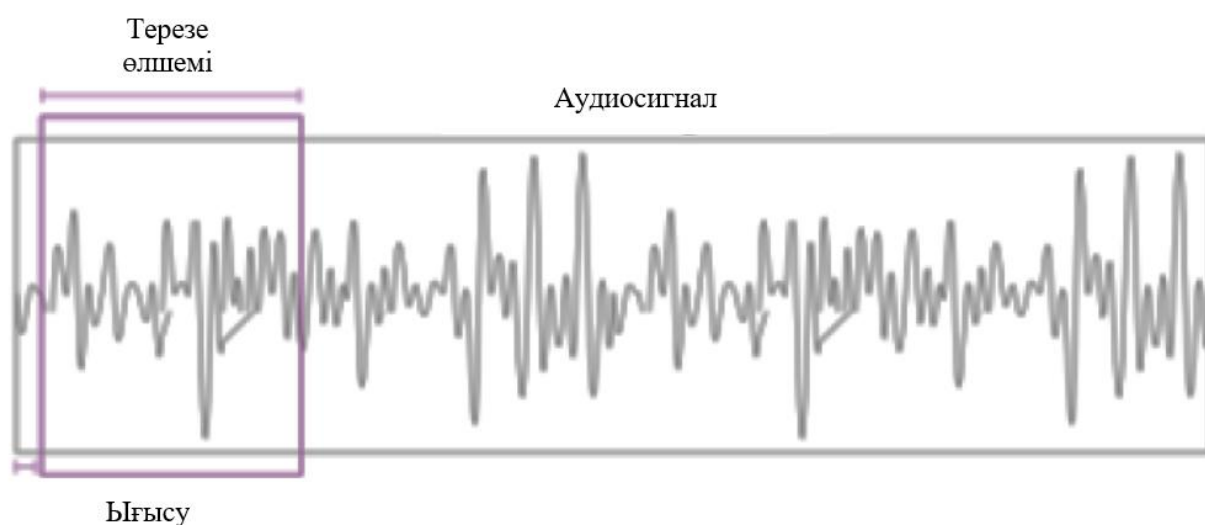
Ұзақтығы қысқа жарылыс дыбысы, бірнеше секундқа созылатын ұзақ дыбыс сигналдарының түрлі типтерін қарастырғанда, ықтималдылықтардың таралу функциясы (PDF) сияқты статистикалық детальдар да түрлі болады деген болжам бар [25, р. 1858-2; 39-41]. PDF-тер орташа мән, дисперсия, нөлді кию жылдамдығы, асимметрия, эксцесс, вариация коэффициенті, энергия сияқты белгілермен сипатталады. Жиіліктік облыстағы белгілер аудиосигналдың спектрінен спектралдық вариация негізінде сигналдарды

ажыратуу үшін алынады. Спектралдық ағын (spectrum flux) уақыт бойынша сигналдың жиіліктік спектрінің өзгеруін сипаттайды. Spectrum roll-off сигналдың жиіліктік спектрінің ассимметриясын анықтайды. Спектр центроиды сигналдың қуат спектрінің таралу ортасы ретінде анықталады. Ол аномалды және нормалды оқиғалар үшін түрлі мәндер береді. Спектралдық энтропия және жазықтық сигнал энергиясы таралуының кездейсоқтық деңгейін өлшейді.

t - және f -домендерінің өздеріне тән ерекшеліктері, түрлі энергияның таралуына ие болғандықтан, сигнал класстарын ажырату үшін қолданылады. Мысалы, таржолақты және кеңжолақты сигналдар спектралдық жазықтық сияқты f -облыстың қарапайым белгілерімен эффективті дифференциалдана алады [39, р. 74]. Алайда мұндай белгілер f -облыста ұқсас энергия тарлуына ие сигналдарды нақты ажырата алмайды. Спектралдық жазықтық екі кеңжолақты сигналды эффективті ажыратуға мүмкіндік бермейді, мысалы: түрлі сигнатураға ие сызықтық ЖМ-сы (LFM) бар сигнал мен ақ шуылды.

TFD ақпараттың едәуір үлкен көлемін қамтиды. Өлшемділік проблемаларының алдын-алу үшін белгілерді алуда (t, f) барлық нүктелері қолданылмайды. Яғни, сигнал классификациясына қатысты ақпаратты неғұрлым тиімді сипаттайтын, белгілердің шағын жиынтығы алынуы керек. [35, р. 459] жұмыста авторлар t - және f -домендерін (t, f) доменіне кеңейту әдісін ұсынған. Егер екі айнымалының арасындағы ақпарат үлкен болса, онда тиісінше олардың тығыз байланысқандығын білдіреді. Мұндай тапсырманың негізгі мақсаты үлкен K жиынтығынан, максималды релеванттылыққа және өзара минималды асқындылыққа ие, (m) функциялар жиынтығын таңдау.

Сонымен көптеген белгілерді алу үшін қысқа-мерзімді өңдеу технологиясы қолданылады [42]. Жоғарды айтылғандай, аудиосигналдар фреймдерге бөлініп, алынған фреймдерді қолдану арқылы талдау жасалады (1.6-сурет).



Сурет 1.6 – Фреймдер негізінде қысқа-мерзімді өңдеу

Қысқа-мерзімді өңдеу, фреймдерге бөлінген сигналдың әр уақыт бірлігін ескереді, яғни аудиосигнал $w(n)$ ұзақтықтағы ығысқан терезе функциясына көбейтіледі [39, р. 67-82]. i -ші өңдеудегі алынатын қорытқы сигнал $x_i(n)$ келесі теңдеуде ұсынылған:

$$x_i(n) = x(n) \cdot w(n - m_i), \quad i = 0, \dots, K - 1 \quad (1.10)$$

мұнда K – фреймдердің саны;

m_i – ығысу (i -ші фреймді алу мақсатында орындалатын фреймнің ығысу санын көрсетеді).

$x(n)$ индекстері $m_i, \dots, m_i + W_L - 1$ іріктемеден өзге облыстарда 0-ге тең, ал W_L – ығысатын терезе ұзындығы. m_i -дың ұзындығы W_S терезе қадамына тәуелді. Көбінесе, W_L ығысатын терезенің ұзақтығы 10-50 мс аралығында болады. Басқаша айтқанда, терезенің S қадамы кезекті кадрлардың аралығындағы беттесу деңгейін көрсетеді. Мысалы, 75% беттесу қажет бол, ан болса, ал терезенің ұзақтығы 40 мс, ендеше терезенің қадамы 10 мс құрауы керек. Демек, қысқа-мерзімді терезелердің K саны келесі 1.11-ші теңдеуге сәйкес есептеледі:

$$K = \begin{cases} \frac{N - W_L}{W_S} + 1 \\ 0, \text{ немесе} \end{cases} \quad (1.11)$$

Терезенің типіне тоқталатын болсақ, тік бұрышты терезені пайдаланған тиімді, тиісінше, терезенің ішіндегі сигнал еш өзгермейді, ал сырттағы сигнал кесіледі. Мұны 1.12-ші теңдеуден көре аламыз:

$$w(n) = \begin{cases} 1, & 0 \leq n \leq W_L - 1 \\ 0, & \text{немесе} \end{cases} \quad (1.12)$$

Тікбұрышты тереземен қатар Хемминг, Бартлетт сияқты т.б. терезелердің түрлері зерттеу жұмыстарында қолданылады [43, 44].

1.5 Аудио мәліметтер жинағы (датасет)

Қазіргі кезде зерттеу жұмыстарын жүргізуде авторлар дыбыс категорияларының кең спектрін қолданады. Бұл дыбыстар қала дыбыстары, эмоциялар, дыбыстық мәтіндер, айтылған сөздер, адамдар мен жануарлардың дыбыстары, медицинадағы дыбыстар және т.б. Тіпті осы бағытта семинарлар мен тапсырмалар ұйымдастырылып, зерттеушілерге өз шешімдерін жасауға және тестілеуге оң ықпалын тигізеді. Ең белгілі семинарлардың қатарында DCASE challenge-ді ерекше атап өтуге болады. Аталған семинарда түрлі тапсырмаларға классификациялау және дыбысқа қатысты түрлі тапсырмаларды қарастыру үшін жыл сайын, тапсырмалар мен датасеттерді жаңарта отырып өткізіледі. Бұл тапсырмалардың ішінде қоршаған орта дыбысын есептеу сараптамасы, акустикалық оқиғаларды классификациялау, дыбыстық

оқиғаларды және орнын олардың анықтау, дыбысты белгілеу және сирек кездесетін дыбыстық оқиғаларды анықтау бар.

[19, p. 8608-25] жұмыста зерттеушілер жасаған шолуда қалада кездесетін дыбыстар бойынша датасеттер қарастырылып, олар туралы ақпараттар ұсынылған. Авторлардың ұсынған ақпараттары 1.1-кестеде көрсетілген.

[18, p. 6718-6] және [21, p. 1883-2] жұмыста авторлар реалды ортада жазылған шуыл мәліметтерін қосымша пайдалан отырып, «TAU Spatial Sound Events 2019» мәліметтер жиынтығын оқиғаларды анықтау үшін қолданған [45]. Зерттеушілер аталған датасеттен microphone array жиынтығымен жұмыс істеген. Жиынтықта 500 аудиомәлімет, 400 аудио өңдеуге және 100 аудиомәлімет бағалау үшін қолданылған. Әрбір аудиоклип 1 минутқа созылса, дискретизациялау жиілігі 48 кГц, сигнал/шуыл қатынасы қоршаған орта шуылы және дыбыстық оқиғалар үшін 30 дБ-ді құрайды. [18, p. 6718-6] жұмыста жазбалар ғимараттың ішіндегі 5 орыннан, 504 уникалды комбинациямен (азимут-биіктік-арақашықтық) жинақталған, кеңістіктік импульстік сипаттамаларын қолдану арқылы синтезделді. Өңдеу үшін әр 60 секундтық аудиоклип 10 секундтан тұратын бөліктерге бөлінді. Ал [21, p. 1883-7] зерттеушілер кросс валидация үшін әр қайсысы 100 жазбадан тұратын 4 бөлікке бөлген. Сонымен қатар, қоршаған орта шуылын Сеулдағы Ханьян университетінің кампусінде жазылған, сигнал/шуыл қатынасы 10 дБ төмен шуылдармен араластырған. Артық оқытуды азайту үшін модельді жалпылау үшін оқыту процесінде мәліметтерді арттырудың екі қарапайым әдісі, монофониялық аудиоклиптерді блоктық микширлеу және тон биіктігін ығыстыру қолданылды.

Кесте 1.1 – Зерттеушілер қолданған датасеттері туралы ақпараттар

Датасет	Өлшемі	Класстар	Ұзақтығы, с	Сипаттама
UrbanSound8k	8732	10	≤ 4	Метадата, баланссыз класстар
ESC-50	2000	30	≤ 5	Метадата, балансталмаған класстар
ESC-10	400	10	≤ 5	Метадата, балансталмаған класстар
ESC-US	250 000	-	≤ 5	Таңбаланбаған датасет
DCASE Task 2	8	-	≤ 10	Механикалық аномалды дыбыстар
CREMA-D	7442	-	≤ 10	Эмоциямен қатар 12 сөйлем
AudioSet	2 млн астам	632	≤ 10	Категориялардың үлкен жиынтығы
TUT Sound Event	24	15	≤ 15	Шағын бөлшектерге кесілген
MIREX	903	15	-	Эмоция классификациясы үшін көңіл күй датасеті
FSD50K	51 197	200	≤ 30	Таңбаланған дыбыстық оқиғалар
Ескерту – Әдебиет негізінде құралған [46-61]				

«DCASE challenge» 2020 үшін Тампере Университеті «TAU Urban Acoustic Scenes 2020 Mobile and Development dataset» [62], «TAU Urban Acoustic Scenes 2020 3Class, Development dataset» [63] мәліметтер жинағын дыбыстық оқиғаларды анықтау тапсырмасы үшін ұсынған. Аудио мәліметтер жинағы жыл сайын толықтырып, жетілдіріліп жаңа тапсырмаларға қолданады.

Мысалы, «TAU Urban Acoustic Scenes 2020 Mobile, Development dataset»-нің аудиомәліметтері Еуропада 12 түрлі қаладан, 10 түрлі акустикалық орында, құрылғының 4 түрін пайдалана отырып жинақталған. Түпнұсқадағы жазбаларға сүйеніп 11 мобильді құрылғы үшін синтетикалық мәліметтер дайындалды. Жазу үшін төрт құрылғы қолданылған: электретті бинауральды микрофон және Zoom F8 аудиомагнитофоны (24 бит мүмкіндігі бар, дискретизация жиілігі 48 кГц құрайды); Soundman II Classic/studio A3; Samsung Galaxy S7, iPhone SE, GoPro Hero5 Session құрылғылары қолданылды. Мәліметтер жиынтығының ұзақтығы 64 сағат, сегменттер саны 23040 құрады, оның ішінде оқытуға қолданылатын сегменттер 13965, тестілеу үшін 2970 сегмент пайдаланылған.

[24, p. 26958-12] зерттеушілер әрқайсысы 200 аудиоклиптан құралған екі шинаның сырғыған дыбысы мен автокөлік соқтығысы оқиғаларынан тұратын «MIVIA road» датасетін туннельдегі жол көлік оқиғаларын дыбыс арқылы тану үшін қолданған. Екі оқиғаның ұзындығы тиісінше 326,38 с және 522,6 с құрайды. Сонымен қатар 2732 с болатын фондық шуыл қарастырылған. Барлық клиптер дискретизация жиілігі 32 кГц, барлық жақа бағытталған микрофонды қолдану арқылы жазылған. Жиынтықты бағалау үшін төрт рет кросс валидация орындалды, соңғы нәтиже барлық кросс валидация орта мәнін табу арқылы алынды. Басқаша айтқанда, шинаның мырғуы және көліктің соқтығысуы төрт топқа топталып, әр топ 50 аудиоклиптен әр оқиғаға 100 клип алынды. Сонымен қатар фондық шуыл төрт топқа бөлініп, әр топтың ұзындығы 700 с құрады. Осыдан соң төрт топтың үшеуі GMM-HMM-ті (Gaussian Mixture Models-Hidden Markov Models) немесе нейрондық желіні оқытуға пайдаланылды. Ал қалған топтар тестілеуге қолданылды. Оқытуға қолданылған ешбір оқиға клипі немесе шуылы тестілеудегі клиптермен қабатталмады. Осы кросс валидация төрт рет орындалды. Жазба уақытының ұзақтығына қарамастан, оқиға фрагменттерінің сан классификаторды оқытуға жеткіліксіз болған. Себебі туннельдегі соқтығыстар дыбысының жиілігі төмен болған. Осы себептен туннельдік ортаны имитациялау арқылы қосымша оқиға клиптері генерацияланған. Модельдік мәліметтерді генерирлеу нәтижесінде 311 шинаның сырғуы және 93 автомобильдің соқтығысуының дыбыстық фрагменттері жинақталған [64]. Содан соң оларға ортаның импульстік сипаттамаларын өзгерте отырып бұрмалану жасалған. Жасалынған бөгеуілдерді имитациялау үшін жазылған фондық дыбыс бұрмаланған оқиға дыбыстарымен араластырылды. Бағалау жиынтығы оқиғалар орташа алғанда фондық шуылда бір сағатта бір рет пайда болып отырған. Фондық шуыл ұзақтығы 48 сағатты құрады, яғни 48 дыбыстық оқиғадан тұрды.

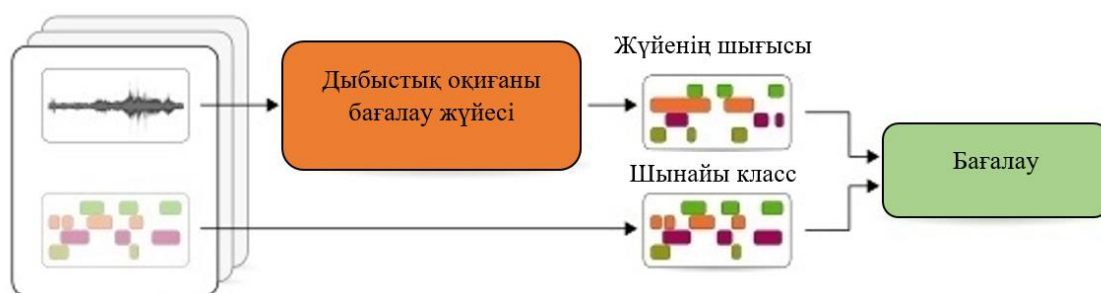
[34, p. 3293-8] жұмыста зерттеушілер «УОНО» алгоритмінің сенімділігін бірнеше жиынтықтарды қолдану арқылы бағалағап, мәліметтердің түрлі

жиынтығына алгоритмді адаптациялау жолын сипаттаған. Алғашқыда музыка мен дыбысталған мәтінді анықтау үшін «MuSpeak» мәліметтер жиынтығынан ұзақтығы 5 сағатты құрайтын аудио алынған [65]. Сонымен қатар авторлар таңбалаған «BBC Radio Devon» ұзақтығы 18 сағат аудиожазба жинақталды. Жиынтықтардың екеуі де 50% оқытуға, 30% тексеруге және 20% тестілеуге бөлінді. Сонымен қатар 46 сағат ұзақтықты құрайтын синтетикалық үлгілер қосылған [66]. Барлық аудиофайлдар 16 кГц қайта дискретизацияланады. Олар алдын-ала өңдеуге дейін арналарды орташаландыру арқылы моно-ға түрлендіріледі. Нәтижесінде, өту (hop) өлшемі 10 мс және терезе өлшемі 25 мс ие 64 log-mel ұяшығы алынаған. Оқыту нәтижесінде алынған модель екі түрлі жинақ негізінде бағаланған. Біріншісі, 4,5 сағаттық «BBC Radio Devon»-нан алынған аудиодан тұратын «MuSpeak» тестілік жинағы. Екіншісі, 27 сағат түрлі телевизиялық бағдарламалардан алынған аудиолардан тұратын «MIREX» мәтіндік дыбыстар мен музыканы тануға арналған жинақ. Екінші жинақ ретінде авторлар [34, 3293-8] «TUT Sound Event Detection» датасетін таңдаған. Аталған жинақ қоршаған орта дыбыстарын анықтауға арналған [67]. Ол CASE challenge 2017 үшінші тапсырмасы таңдалған еді. Әр дыбыс 2,56 с созылған. Тежегіштің дыбысы, автомобиль, балалар, үлкен транспорттық көлік, сөйлеуші адамдар және жүріп келе жатқан адам сияқты класстардан тұрады. Алты классты және олардың басталу-аяқталу уақыттарын болжау үшін 18 шығыс нейрон қажет болған. Мәліметтер жиынтығының жалпы ұзындығы 1,5 сағат. Мәліметтер жинағы кросс валидация 4 рет бапталған. Жиынтық көлемі аз болғандықтан бірінші L2 нормализациясы қолданылған. Мәліметтерді арттыру үшін кіріс мәліметтерден кездейсоқ түрде жиіліктік блок тізбектері немесе уақыттық қадамдарын жоятын SpecAugment қосылған [68, 26]. Зерттеушілер таңбалауды қайта анықтау қиындық тудыратындықтан ешқандай уақыттық бұрмалануларды қолданбаған. Сонымен қатар, уақытты үнемдеу мақсатында SpecAugment жеке үлгілерге емес пакеттерге қолданылған. Мәліметтер базасы дискретизация жиілігі 44,1 кГц-ке тең стереофондық аудиофайлдардан тұрады. Алдын-ала өңдеудің алдында ол моно-ға түрлендірілді. Нәтижесінде 0-ден 22 050 Гц аралығында 40 log-mel жолағы алынды. Өту өлшемі 10 мс, терезенің өлшемі 40 мс. Желінің кіріс сигналы 2,56 с аудио болғандықтан кіріс сигнал формасы 257×40 болды. Шығыс сигналдың формасы 9×18. Яғни 9 реттік қадам және 6 акустка классы. Бұл жағдайда әрбір уақыттық қадам 0,284 с тең [68]. Зерттеушілер қолданған соңғы датасет «Urban sound event detection» деп аталады. Ол генерирленген синтетикалық мәліметтер жиынтығы. Әр дыбыс 10 с созылған. Кондиционер, автомобильдің сигналы, ойнап жүрген балалар, иттің үруі, бұрғылау, қозғалтқыш дыбысы, сирена және көше дыбысы сияқты он класстан тұрады. Мәліметтер жиынтығының жалпы көлемі 30 сағатты құрайды және алдын-ала оқытуға, тексеруге, тестілеуге бөлінген. Класстардың саны он болғандықтан шығыс нейрондардың саны 30-ға тең. Мәліметтер көлемі жеткілікті болғандықтан ешқандай қосымша толықтыру жасалмады. Кіріс сигнал 10 с болғандықтан, кіріс сигналдың формасы 1001×40, яғни 1001 реттік

қадам және 40 log-mel жолғаны сәйкес. Шығыс сигналдың формасы 32×30 құрады. Демек 32 реттік қадам және он акустика классы [68, р. 2617].

1.6 Модельді бағалау метрикалары

Mesaros, A және өзге авторлар [69] зерттеу жұмысында дыбыстың бірнеше көзі активті болған реалды оқиғаларға қолданылатын, дыбыстық полифониялық оқиғаларды анықтауға арналған жүйелерді бағалау үшін қолданылуға ұсынылатын түрлі көрсеткіштерді зерттеген. Бұл жағдайда жүйенің шығысы, бір мезетте активті болатын бірнеше дыбыс ретінде таңбаланып, оқиғалардың қабатталуынан тұрады. Полифониялық жүйенің шығысы эталонмен салыстыру үшін сәйкес процедураны талап етеді. Бұл жағдайда дыбыстық мәтіндерді тану, сөйлеушіні анықтау сияқты тапсырмаларда қолданылатын көрсеткіштерді пайдалануға болады, бірақ қабатталған оқиғаларды анықтау үшін, оларды тану қажеттілігін түсіну керек. Аталған жұмыста авторлар неғұрлым кең қолданыс тапқан метрикаларға шолу жасап, олардың адаптациялануын және түсіндірілуін ұсынады. Сегменттер мен оқиғаларға негізделген әр метрикаға анықтамалар ұсыныла отырып полифониялық оқиғаларға қолданылуы, түсіндірілген. Бағалау процесі 1.7-суретте сипатталған.



Сурет 1.7 – Оқиғаларды анықтауға арналған жүйесін бағалау

Дыбыстық оқиғаларды автоматты сараптау жүйелерінің анықтау нақтылығын сандық бағалау, жүйенің шығыс мәліметін тестілік мәліметтермен салыстыру арқылы орындалады. Мұндай аудиомәліметтер эталонды қолдан аннотациялануы мүмкін немесе синтетикалық мәліметтер мен оқшауландырылған дыбыстардың сәйкес аннотацияларын қолдану арқылы алынады. аннотацияның кең таралған форматы сәйкес баталуы мен ығысуы бар дыбыстық оқиға үлгілерінің тізімі. Оқиғалардың уникалды таңбаларының жиынтығы дыбыстық оқиғаны анықтау тапсырмасына сәйкес оқиғалар класын құрайды. Нәтижелерді бағалау класс атауымен қатар уақыттық ақпаратты да ескереді [70].

Жүйе шығысы мен шынайы оқиға эталонды салыстыру анықталған ұзындық интервалы немесе оқиға үлгісі деңгейінде орындалады. Бұл өнімділікті өлшеудің екі түрлі әдісіне алып келеді: сегменттерге негізделген және оқиғаларға негізделген көрсеткіштер. Олардың әрқайсысы үшін дұрыс анытаудың не екендігін және жүйе қандай қателік түрін беріп тұрғандығын

анықтау маңызды [69, p. 162-5] жұмыста зерттеушілер оларды жүйенің дұрыс және қате шығыс мәліметтерін жеке ескеретін, аралық статистикалық мәліметтер деп атаған.

Сегменттерге негізделген метрикалар. Сегменттер негізіндегі метрикалар жүйенің шығыс мәліметтер мен шынайы классы қысқа уақыт сегменттерінде салыстырады. Әр оқиға классының активті/активті емес жағдайы сегмент болып табылатынбекітілген ұзындық интервалында анықталады. Активтілікті ұсыну негізінде келесі аралық статистикалық мәліметтер анықталады [69, p. 162-6; 70, p. 632]:

- шынайы-оң (TP-true positive): жүйенің шығысы мен шынайы класс оқиғаның осы сегментте активті екендігіне нұсқайды;

- жалған-оң (FP-false positive): шынайы класс оқиға активті еместігіне, ал жүйе шығысы активті деп көрсетеді;

- жалған-теріс (FN): шынайы класс оқиғаның сегментте активті болу керектігіне, ал жүйенің шығысы активті емес екендігіне нұсқайды;

- кейбір метрикалар шынайы-теріс (TN) мәндерді ескереді: шынайы класс пен жүйе шешімі оқиғаның активті еместігіне нұсқайды.

Бағалау үшін неғұрлым ұзын сегментті қолданғанда активтілік индикаторы неғұрлым ұзын аудио фрагментті қамтиды. Бұл оқиғаның басталуы мен шығысуын белгілеу кезінде, аннотатор субъективтілігіне қатысты проблемаларды жоюға ықпал етеді.

Оқиғаға негізделген метрикалар. Оқиғаға негізделген метрикалар жүйенің шығыс мәліметі мен сәйкес шынайы класс оқиғаларын кезектеп салыстырады. Аралық статистикалық мәліметтер келесідей анықталады [69, p. 162-6; 70, p. 633]:

- шынайы-оң: жүйе шығысындағы оқиғаның уақыттық позициясындағыдай тағбаланған шынайы класс оқиғасымен қабатталады;

- жалған-оң: жүйе шығысындағы оқиға рұқсат етілген аралықта шынайы класс оқиғасына сәйкес келмейді;

- жалған-теріс: шынайы класс оқиғасы рұқсат етілген аралықта жүйе шығысындағы оқиғаға сәйкес келмейді.

Оқиғаға негізделген метрикалар шынайы-теріс мәнге тек ұзындық терминінде фактілік уақыттық қателіктерде өлшеу кезінде ғана қолданылады. Бұл жағдайда шынайы класс пен жүйе шешімі активті оқиғаға ие емес уақыттық сегменттерінің жалпы ұзындығы өлшенеді.

Метрикаларды өлшеудегі орташалау параметрлері. Аралық статистика түрлі әдістермен агрегацияланады: глобалды – класстарға тәуелсіз немесе екі кадамды орындау арқылы: әуелі класстар бойынша, кейін жеке класстар нәтижелері бойынша орта мәні ретінде [69, p. 162-6; 70, p. 633].

Үлгі негізінде орташаландыру немес микроорташаландыру барлық шешімге тең салмақ береді. Бұл жағдайда дыбыстық оқиғаның әр үлгісі немесе әрбір сегмен үшін активті үлгі жүйе өніміділігіне тең әсер етеді. Шынайы-оң нәтижелер, жалған-оң нәтижелер және жалған-теріс нәтижелердің саны барлық тестілік мәліметтер бойынша қосындыланып, жалпы мән негізінде метрикалар

есептеледі. Бұл қарастырылатын тапсырмадағы ірі класстардың өнімділігіне әсер етеді [69, p. 162-6-162-7; 70, p. 633].

Класстар негізінде орташаландыру немесе макроорташаландыру барлық классқа тең салмақ береді. Жеке класстардағы жүйенің өнімділігі – бұл жағдайда аралық статистика тестілік мәліметтерден бөлек әр оқиға класына жеке агрегацияланады. Класстар бойынша агрегацияланған метрикалар (TP, FP, FN) класстар бойынша метрикаларды есептеу үшін қолданылады. Содан соң жалпы өнімділік класс бойынша өнімділіктің орта мәні ретінде есептеледі. Бұл тапсырмада қарастырылатын шағын класстардағы жүйенің әрекетін айқындайтын мәндерге алып келеді [69, p. 162-6-162-7; 70, p. 633].

Кросс валидация. Нәтижелерді салыстыру және жаңғыртуда тәжірибелерді жүргізу және кросс валидация үшін баптау маңызды аспект болып табылады. Класстар дисбалансы жоғары болған кезде өлшеу ығысуынан бөлек көпклассты тапсырмаларда әр жинақтарында барлық класс орналасады деп кепілдік беру қиын. Қалыс қалған класстар кейбір метрикаларды есептеуде нөлге бөлінуге алып келеді. мұндай жағдайлардың алдын алу үшін шынайы және жалған-оң нәтижелердің санын есептеп, содан соң метрикаларды есептеу ұсынылады. Сонымен қатар, бұл тестіленетін категориялардың ішкі жиындары бойынша орташаландыру нәтижесінде пайда болатын, жалпы орта мәндердің айрықшалануын болдырмауға әсер етеді. Мысалы, мәліметтерді түрлі оқыту/тестілеу жинақтарына бөлуді айтуға болады. Бұл тұрғыда кросс валидация жиынтықтарын, барлық жиынтықтарды тестілеуден соң ақырғы метрикаларды есептемен біртұтас тәжірибе ретінде қарастырған жөн. Бұл метрикалар тізбектей бірдей мәліметтер бойынша есептелетіндігіне кепілдік береді.

Precision, Recall, F-Score. Precision (P-дәлдік) мен Recall (R) ақпаратты іздеуде тиімді метрикалары болып табылады. Бірақ классификация тапсырмасында «оң болжам мәні» және «сезімталдық» деп те аталады. Бұл метрикаларды аралық статистика анықтамаларына сүйене отырып адаптациялау еш қиындық тудырмайды. F-Score P және R метрикалары арқылы есептеледі [71, 72]:

$$P = \frac{TP}{TP+FP} \quad (1.13)$$

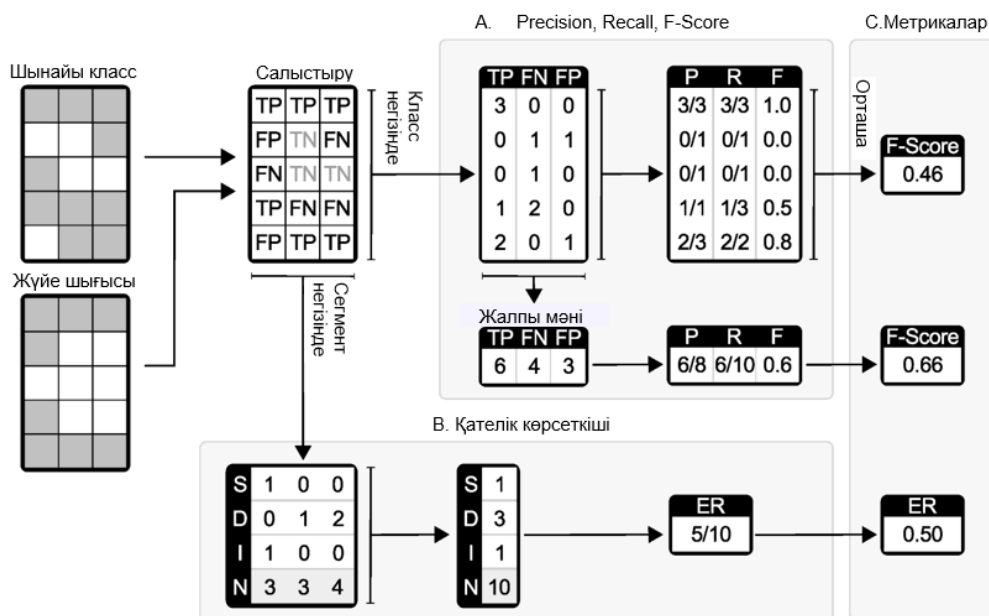
$$R = \frac{TP}{TP+FN} \quad (1.14)$$

$$F = \frac{2 \cdot P \cdot R}{P+R} \quad (1.15)$$

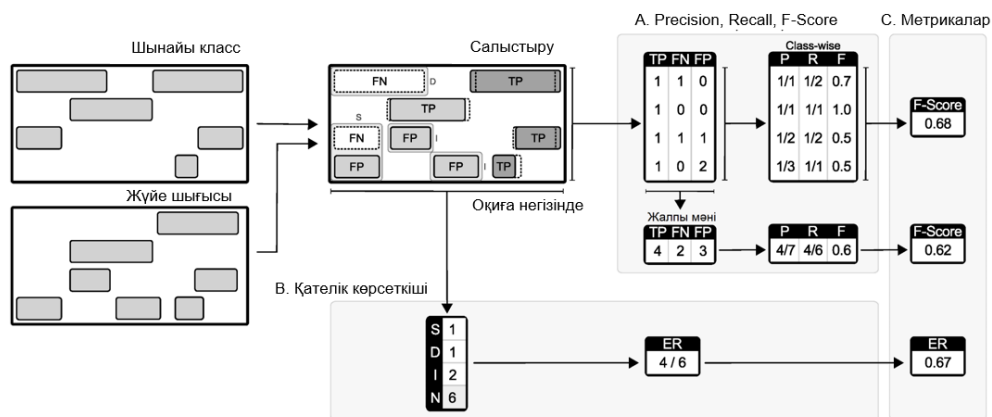
немесе аралық өлшемдер негізінде [73]:

$$F = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (1.16)$$

Сегмент негізіндегі P, R және F, класс негізінде немесе үлгі негізіндегі орташандыруды қолданып, сегмент негізіндегі аралық статистикаға сүйеніп есептеледі. Есептеу 1.8-суреттегі А пунктінде келтірілген. Оқиға негізіндегі P, R және F, дәл солай оқиға негізіндегі аралық статистикаға сүйеніп есептеледі. Есептеу 1.9-суреттегі А пунктте келтірілген[69, р. 162-8].



Сурет 1.8 – Сегментке негізделген метрикаларды есептеу



Сурет 1.9 – Оқиға негізінде метрикаларды есептеу

Дыбыстық оқиғаны анықтау эффективтілігін бағалау үшін F-Score-ды қолдану түсінуге оңайлығы есебінен кеңінен қолданыс тапқан. Кемшілігі ретінде орташандырудың ерекше маңызды болуында. F-Score мәні шынайы-оң нәтижелер бойынша анықталатындықтан, үлгі негізіндегі орташандыру кезінде ірі класстар кіші класстарға қарағанда басымдылыққа ие болады. класстар негізінде орташандыру кезінде барлық тестілік мәліметтер класстарының болуын қамтамасыз ету керек. Яғни Recall-дың анықталмайын болдырмау керек ($TP+FN=0$). Реалды өмірдегі жазбалардың кез-келген

жиынтығы балансталмаған оқиғалар кластарынан тұрады, сондықтан орташаландыруды таңдау үнемі маңызды фактор болып есептеледі [69, p. 162-7].

Қателік көрсеткіші. Қателік көрсеткіші қою (I), жою (D) және ауыстыру (S) терминдерінде қателіктердің санын өлшейді. Сегмент негізіндегі қателік көрсеткішін есептеу үшін, сегменттерді тізбектей отырып қателікті есептейді. k сегментіндегі ауыстыру қателігінің саны $S(k)$ – бұл дұрыс оқиға шығыста алынбаған, шынайы класс оқиғаларының саны. Бұл жалған-оң және жалған-теріс нәтижелерді қайсы қате оқиға қайсысын ауыстыратындығын көрсетпей салыстыру арқылы мүмкін болады. Қалған оқиғалар – бұл қою және жою:

1. $D(k)$ – дұрыс анықталмаған шынайы класс оқиғаларының саны (ауыстырудан кейінгі жалған-теріс нәтижелер ескеріледі).

2. $I(k)$ – жүйе шығысындағы мәліметтерде дұрыс болмаған оқиғалар саны (ауыстырудан кейінгі жалған-оң нәтижелер ескеріледі). Бұл келесі теңдеуге алып келеді:

$$S(k) = \min(FN(k), FP(k)) \quad (1.17)$$

$$D(k) = \max(0, FN(k) - FP(k)) \quad (1.18)$$

$$I(k) = \max(0, FP(k) - FN(k)) \quad (1.19)$$

Жалпы қателік көрсеткіші K сегменттерінің жалпы саны бойынша сегментке негізделген есептеулерді интеграциялау арқылы анықталады. $N(k)$ k сегментіндегі шынайы класстарда активті деп белгіленген дыбыстық оқиғалар санымен анықталады [69, p. 162-9]:

$$ER = \frac{\sum_{k=1}^K S(k) + \sum_{k=1}^K D(k) + \sum_{k=1}^K I(k)}{\sum_{k=1}^K N(k)} \quad (1.20)$$

Есептеулер 1.7-суреттің В пунктінде көрсетілген.

Оқиғаларға негізделген қателік дыбыстық оқиғаға қатысты көрсеткіші уақыттық орналасу мен шынайы класста көрсетілген нәтиже мен жүйе шығысындағы әр дыбыстық оқиға таңбасын салыстыру арқылы анықталады. Уақыт бойынша дұрыс орналасқан, бірақ таңбалануы дұрыс емес дыбыстық оқиғалар ауысу болып саналады [74]. Ал қою және жою дұрыс емес немесе жүйе шығысында немесе шынайы класс таңбасында ауыстырылған дыбыстық оқиғалар. Жалпы қателік көрсеткіші (1.20) теңдеу бойынша анықталады және 1.8-суреттің В пунктінде келтірілген. қателіктердің класстық көрсеткіші әр дыбыстық оқиға класы үшін жеке есептеледі, ауыстыру қарастырылмай тек қою және жою ескеріледі.

Sensitivity (сезімталдық) және specificity (ерекшелік) шынайы-оң және шынайы-теріс көрсеткіштер болып табылады. Олар өлшенетін екі компоненттің арасындағы компромиссті көрсету үшін бірге қолданылады:

$$Sensitivity = \frac{TP}{TP+FN} \quad (1.21)$$

$$Specificity = \frac{TN}{TN+FP} \quad (1.22)$$

Қарастырылған көрсеткіш қабылдағыштың жұмыс сипаттамасы (ROC- Receiver Operating Characteristic) қисығы және қисық ауданымен (AUC- Area Under the Curve) беріледі. ROC қисығы шынайы-оң нәтижелерді жалған-оң нәтижелермен салыстыра отырып көрсетеді, яғни түрлі шекті мәндегі сезімталдылық.

Ассигасу (нақтылық) жүйенің дұрыс шығыс мәліметтер санының, жалпы шығыс мәліметтер санына қатынасымен анықталады:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (1.23)$$

Ассигасу жүйенің шынайы-оң нәтижелерді анықтау қабілетін жақсы сипаттай алмайды: оқиға активтілігі кесілгенде, шынайы-теріс нәтижелердің саны Ассигасу мәнінен басым болады. Балансталған ассигасу-ды ала отырып, TP және TN салмақтарын еншілеуге болады:

$$BACC = w \cdot \frac{TP}{TP+FN} + (1 - w) \cdot \frac{TN}{TN+FP} \quad (1.24)$$

Ассигасу сегмент негізіндегі және оқиға негізіндегі дыбыстық оқиғаларды анықтауға қолданылады:

$$ACC_{MIR} = \frac{TP}{TP+FP+FN} \quad (1.25)$$

(1.23)-(1.25) теңдеулер бойынша анықталатын ассигасу нұсқаларының артықшылығы 0 және 1 шектелген. Жүйе қателеспегенде 1 алынады. Жүйе шығысы тек қателерден тұрса, (1.23), (1.24) теңдеулерден 0 алынады. Ал (1.25) теңдеуден егер жүйе еш жасап шығармаса 0 алынады. Себебі оның мәндері шынайы-теріс көрсеткіштермен байланыспаған. Ассигасу кемшілігі ретінде қателер типі арасындағы компромисс туралы еш ақпарат бере алмайтындығын атауға болады. Алайда ол кейбір қолданбаларда тек ауыстыру ұсынылатын жағдайларда, қою мен жоюды есептеуге қабілетті [75].

1.7 Аудиосигналдар негізінде оқиғаларды анықтауда қолданылатын әдістер

Зерттеу жұмысында «Қауіпсіз қала» платформасын енгізу дайындық жұмысы жүргізілді. Бұл қызметтің негізгі мақсаты – заңға қайшы әрі азаматтардың өмірі мен мүлкіне қауіп төдіретін төтенше оқиғаларды реалды уақыт мезетінде анықтау. Қолданыстағы қауіпсіздікті қамтамасыз етуге арналған бейнебақылау жүйелері микрофондармен жіне аудиосараптама

жүйелерімен жабдықталуда. Бейнебақылау жүйелерімен салыстырғанда аудиоға негізделген сараптама уақыт бойынша жылдам әрі жоғары деңгейдегі құрылғылар мен әдістерді талап етпейді.

Америкалық ShotSpotter жүйесі аудиосараптамадағы жетімтіктердің бірі болып табылады. Аталған жүйе Вашингтонда орнатылған бір жылдан бері 39 000 қару жарак қолданылу жағдайларын анықтап, оқиғаның басталуы туралы тиесілі органдарға оперативті түрде хабарлаған [76].

Ұлыбританияда жасалған Audio Analytical Ltd жүйесі дыбыс арқылы анықтау жүйесін енгізудің тағыда бір мысалы [77]. Компанияда жасалған «Ақылды үй» датчиктері қару жарак, адамдардың қатты айқайы, жылаған балалар, автомобиль сигнализациясы, тезенің сынуы сияқты т.б. дыбыстарды анықтай алады. Егер аталған оқиғалар тіркелген жағдайда қауіпсіздік органдарына, қолданушыға хабар бере алады.

Соңғы жылдардағы ақпараттың күн санап артуы, соның ішінде аудио мәліметтердің санының артуы аталған бағытта зерттеу жүргізуге жол ашылып, зерттеушілердің назарын аудартуда.

Терең оқытуға дейін дыбыстық оқиғаларды тану тапсырмасы үшін GMM, НММ, кездейсоқ орман, тірек векторлары машинасы мен негативті емес матрицалық факторизация сияқты дәстүрлі әдістер қолданылып келді. Қазіргі уақытта терең нейнордық желілерге негізделген CNN, RNN және CRNN сияқты шешімдер қолданылып, зерттеушілер жоғары өнімділікке қол жеткізуде.

[24, p. 2695-7] жұмыста зерттеушілер туннельдегі бақыланбайтын келесі жол оқиғаларын болдырмау үшін дыбыстық оқиғаларды тану әдісін ұсынады. Қалыпты емес дыбыс ретінде шинаның сырғуы және автокөліктің соқтығысуы қарастырылған. Авторлар туннельде шуылдың жоғары деңгейі байқалатындығын және қолданыстағы шешімдердің тиімділін арттыру мәселесін келтіре отырып, тапсырманы шешу үшін туннельдегі дыбысты алдын ала өңдеуден өткізіп, акустикалық оқиғаны анықтау үшін классификатор қолданған. Алдын ала өңдеу сатысында негативті емес тензорлық факторизацияны (NTF – non negative tensor factorization) қолданып, акустикалық оқиға сигналын шуылдан ажыратып алған. Содан соң оқиға және шуыл параллель қабаттарынан тұратын кеңетілген конволюциялық рекурренттік нейрондық желіні (CRNN) қолданған. Кіріс белгілер ретінде mel-filterbank параметрлер жиынтығы алынды. Ұсынылатын әдісті бағалау екі мәліметтер жиынтығы арқылы жүргізілді: баршаға қолжетімді жолдағы дыбыстық оқиғалардың мәліметтер жиынтығы және алты ай бойы реалды транспорттық туннельде жазылған мәліметтер жиынтығы. Бірінші, фондық шуыл төмен болған бағалауда, CRNN негізінде шуыл бойынша онлайн оқытуы бар дыбыстық оқиғаны анықтау әдісі қарапайым шуылмен бірге CRNN негізіндегі әдіспен салыстырғанда қателік көрсеткішін 56,25% төмендетуге мүмкіндік берген. Туннельдегі шуыл деңгейі жоғары болған CRNN негізіндегі әдіс, екінші бағалауда бірінші бағалауға қарағанда өте жоғары өнімділік көрсеткен. Нақты айтқанда, шуыл бойынша ұсынылатын онлайн оқыту әдісі танудың жоғары жылдамдығын 91,07%, қаматамасыз етеді және қателік

көрсеткішін GMM-HMM дәстүрлі әдістермен салыстырғанда 47,40% және қарапайым CRNN салыстырғанда 28,56% төмендеткен. Есептеу күрделілігін өлшеу CRNN негізінде ұсынылға әдіс өңдеу үшін NTF негізінде шуыл бойынша онлайн оқытуы бар дыбыс көдерін бөлуге және туннельдік шуыл сигналы 1 с соқылатын CRNN классификациясы үшін де 599 мс қажет болады. Яғни ұсынылатын әдіс нақты уақыт режимінде анықтай алады. Тиісінше, ұсынылған әдіс бойынша дұрыс анықталған класстардың анықтау көрсеткіші RR 98,5% (recognition rate), анықталмаған оқиғалар көрсеткіші MDR (missed detection rate) 1,0% құрады. Ең жоғарғы AUC (area under the curve) 98.39% көрсетті [24, p. 2695-17].

[34, p. 3293-10-3293-12] зерттеу жұмысында авторлар YOHO (you only hear once) деп аталатын жаңа шешімді ұсынған. Олар акустикалық шекараларды анықтау тапсырмасын фреймдер негізінде классификациялаудың орнына регрессияға түрлендіруді ұсынады. Бұл аудиокласстың бар екендігін анықтау және оның бастапқы нүктесі мен соңғы нүктесін болжау үшін жеке шығыс нейрондарының болуы есебінен жүзеге асырылады. Замануи конволюциялық нейрондық желілермен салыстырғанда F-өлшемдердің жақсаруы, дыбысты сегменттеу және дыбыстық оқиғаларды анықтауда бірнеше мәліметтер жиынтығын үшін 1%-дан 6%-ға дейін құрады. Зерттеушілердің айтуынша, YOHO шығысындағы болжам нейрондары санының аз болуы, жүйе жұмысының жылдамдығы классификация бойынша сегменттеумен салыстырғанда 6 есе жоғары, ал акустикалық шекараларды тікелей болжайтындықтан ақырғы өңдеу мен тегістеу 7 есе жылдам болған. Тиісінше, музыка-дыбыстық мәтінді анықтау тапсырмасы бойынша $F_{жалпы}$ - 97,22%, $F_{муз}$ - 98,2%, $F_{дм}$ - 94,89% құрады. MIREX датасеті негізінде музыка-дыбыстық мәтінді анықтау тапсырмасы бойынша $F_{жалпы}$ - 90,20%, $F_{муз}$ - 85,66%, $F_{дм}$ - 93,18%. TUT sound event detection датасеті бойынша қателік көрсеткіші 0,58 құрайды. Ал Urban-SED датасеті бойынша $F_{жалпы}$ - 59,5% құрай отырып, дыбыстық оқиғаның иілу бағалауын қолданылған CRNN-нен [78, 79] кейінгі орынға орналасқан.

YOHO CNN негізделген алгоритм [80] жұмыста ұсынылған желінің архитектурасы тапсырмаға адаптацияланып, соңғы қабаттар YOHO алгоритмін жүзеге асыру үшін модификацияланған.

[18, p. 6718-10] жұмыста зерттеушілер, сигнал/шуыл деңгейі төмен болғанда дыбыстық оқиғаны анықтау өнімділігін арттыру үшін бастапқы қабатта терең нейрондық қабаттар негізінде уақыттық облыстағы шуылды біріктіре отырып басуды (NS моделі) ұсынады. Ал соңғы қабаттарда CRNN негізіндегі классификация моделі ұсынылады. Нәтижесінде NS және дыбыстық оқиғаны анықтау модельдері каскад ретінде біріктіріліп, түрлі мақсаттарда оқытылады. Сонымен қатар ақырғы жоғалту арқылы оқытуды орындап, біріктірілген баптау әдісін ұсынған. Нәтижесінде тығыз қабаттардың салмақтарын қатыру арқылы назар аудару механизмін қолданып, F-score мәнін 8,6 және 10% арттыру мүмкін болды.

[81] жұмыста терең нейрондық желілер негізіндегі дыбыстық оқиғаларды анықтау үшін туынды функцияларын қолдануды ұсынады. Желілер үшін кіріс мәліметтер ретінде әр аудиосигнал кадры үшін log-mel-filterbank, оның бірінші және екінші туынды функциясы алынады. Туынды функцияларының эффективтілігін бағалау үшін екі терең нейрондық желі қолданылған. Нақты айтқанда, конволюциялық және рекурренттік нейрондық желіні біріктіру арқылы CRNN алынып, соңынан классификация деңгейі ретінде әрекет ететін тура байланысты нейрондық желі (FNN) қосылған. Сонымен қатар CRNN назарына негізделген орта-оқытушы моделі қолданылды. Модельдердің екеуі де шығыста орта бірігу деңгейіне ие, әлсіз таңбаланған және таңбаланбаған аудиомәліметтерді модельді оқыту барысында қолдану мүмкін болды. Туынды функцияларын қолдану есебінен түрлі оқыту шарттарында, нейрондық желі архитектурасы және мәліметтер жиынтығына байланысты өнімділіктің үнемі артуына алып келді.

Алдын-ала өңдеу сатысында аудио сигнал 16 кГц жиілікте дискреттелді. Қабаттасуы 360 (41,5 мс), ұзындығы 1024 (64 мс) Хэмминг терезесі қолданылып қысқауақыттық Фурье түрлендіруі (STFT) орындалды. STFT қолдану арқылы 0-ден 16 кГц-ке дейінгі кішімасштапта фильтр банкасының шығыс мәліметтері 60 жолақта алынды, содан соң әр кадрға 41,5 мс бірдей өлшемдегі LMFV алу үшін логарифмдық түрлендіру орындалды. Белгілерді алу барысында 64 өлшемді 10 с кадрлар үшін 240 фрейм генерирленді. Кейіннен, LMFV есептеліп, одан орта мәнін азайтып, барлық оқыту мәліметтері бойынша орташа ауытқуына бөлу арқылы нормализацияланды. Нәтижесінде ол CNN үшін кіріс мәлімет ретінде қолданылды. Датасет ретінде DCASE 2018 and 2019 challenge ұсынған мәліметтер жиынтығы қолданылды [82]. Жиынтық әлсіз және қатаң таңбаланған, таңбаланбаған оқытушы мәліметтер комбинациясынан тұрады, 10 түрлі типтегі тұрмыстағы дыбыстар қарастырылған.

CRNN өнімділігі F-score және ER арқылы өлшенген. Жоғалту функциясы ретінде бинарлы кросс-энтропия қолданылған, базалық CRNN-ды оқыту үшін Adam оптимизаторы 0,001 оқыту жылдамдығымен пайдаланылды. Нәтижесінде DCASE 2018 негізінде жоғарғы өнімділікті [әлсіз+таңбаланбаған+қатаң] комбинациясы қамтамасыз еткен. Бұл оқытушы мәліметтердің неғұрлым көп болуымен түсіндіріледі. Алайды таңбаланбаған мәліметтер барлық мәліметтердің 80% құрайтынын ескерсек, өнімділік мәні қанағаттандыратын емес. Демек базалық CRNN таңбаланбаған мәліметтерді параметрлерді жаңарту үшін эффективті қолдана алмайды. Сонымен қатар туынды функцияларын қолдану F-score тұрғысынан өнімділіктің үнемі артуына алып келеді. [әлсіз+таңбаланбаған+қатаң] комбинациясында 7,2% жақсару байқалады. Барлық комбинациялардағы орташа жақсару 11,6% құрайды. Алайда [әлсіз+таңбаланбаған] ER мәні жақсы емес. Оны туынды функциясын қолдану арқылы жақсартуға болады. DCASE 2019 тестілік жиынтығында да жағдай дәл осындай. F-score-дың орташа жақсаруы 5,3%, ал [әлсіз+таңбаланбаған+қатаң] комбинациясы 6%-ға жақсарған.

Орта-оқыту моделі үшін [әлсіз+таңбаланбаған+қатаң] комбинациясы таңдалды. Ерте тоқтату орнына модель 100 және 200 эпоха үшін оқытылды, тестілік мәліметтерді бағалау үшін валидация мәліметтері бойынша үздік модель таңдалды.

Adam оптимизаторы 0,001 оқыту жылдамдығымен пайдаланылды, ал классификация деңгейінің шығыс мәліметтеріне ұзындығы 5 медианалық фильтрация қолданылды. Нәтижесінде DCASE 2018 тестілік жиынтығында F-score 3% жақсарса, DCASE 2019 тестілік жиынтығы үшін 4,4% жақсару байқалды.

Кесте 1.2 – Төтенше оқиғаларды анықтауға бойынша өзге зерттеушілердің нәтижелерін талдау

Модель	Датасет	Оқиға	Белгілер	Нәтижелер	Зерттеушілер
Bagged tree	-	Қару жарақ дыбыстары	LPC, MFCC, GTCC	0.97 true positive rate	Kabir және т.б.
Binary classification simple NN	-	Қару жарақ дыбыстары	Antilog energy features	0.931 true positive	Rahman және т.б.
Модель	Датасет	Оқиға	Белгілер	Нәтижелер	Әдебиет
Single and ensemble NN	-	Қару жарақ дыбыстары	MFCC	0.95	Sigmund, Hrabina
Machine learning	-	Қару жарақ дыбыстары	MFCC	0.96 accuracy	Gupta және т.б.
YOHO (CRNN)	MuSpeak, TUT Sound Event Detection, Urban-SED	Музыка, дыбысталған мәтін, көлік, сөйлеу, қадамдар т.б.	log-mel spectrogram	0.97 F-score	Satvik Venkatesh және т.б.
SED (CNN)	MIVIA	Туннельдегі көліктің соқтығысуы	mel-filterbanks	0.98 recognition rate	Nam Kyun Kim және т.б.
CRNN	DCASE 2018 and 2019 challenges.	Сөйлеу, ит, мысық, ыдыстар, блендер, шаңсорғыш	log-mel filterbank, STFT	F-score 5% жақсартылған	Kwak J.-Y және т.б.
Зерттеу жұмысында ұсынылған модель CNN-BiLSTM	Төтенше оқиғаларды анықтау үшін жинақталған датасет (10000 аудиофайл)	Қару жарақ, жылау, терезенің сынуы, иттің үруі, от жалыны, сирена, жарылыс, айқай дыбыстары	Хромограмма; Спектралдық контраст; Mel-frequency cepstrum; MFCC; Tonnetz		
Ескерту – Әдебиет негізінде құралған [24, p. 2695-1-2695-18; 34, p. 3293-2-3293-15; 81, p. 4911-1-49-10; 83-86]					

1.2-кестеде өзге авторлардың жұмыстарын талдау бойынша нәтижелер мен ұсынылатын жұмыстың ерекшеліктері көрсетілген. Диссертациялық жұмыста төтенше оқиғаларды анықтау үшін қару жарақ, жылау, терезенің сынуы, иттің үруі, от жалыны, сирена, жарылыс, айқай дыбыстары сияқты сегіз класстан тұратын мәліметтер жинағы қолданылды. Аудио мәліметтерден хромограмма, спектралдық контраст, mel-frequency cepstrum, mel-frequency cepstral coefficients және tonal centroid features (Tonnetz) белгілері алынды. Ұсынылатын модель CNN-BiLSTM комбинацияланған архитектурасына негізделді.

2 ТӨТЕНШЕ ЖАҒДАЙЛАРДЫ АНЫҚТАУ ҮШІН АУДИО МӘЛІМЕТТЕРДІ МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІ НЕГІЗІНДЕ ӨНДЕУ

2.1 Жасанды интеллект және машиналық оқыту әдістерін төтенше жағдайларды анықтауда қолдану

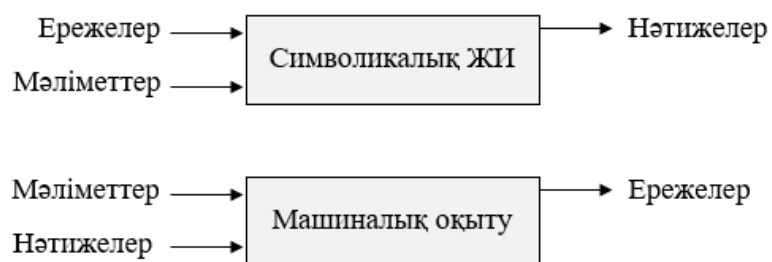
Соңғы бірнеше жылдарда, жасанды интеллект (ЖИ) саласы зерттеушілер арасында үлкен қызығушылыққа ие болғандықтан, машиналық оқыту, терең оқыту және ЖИ тақырыбындағы мақалалардың саны күн сайын артууда. Машиналық оқыту саласындағы маман пайдалы сигналды шуылдан бөліп алу тапсырмасын орындай алуы тиіс (2.1-сурет).



Сурет 2.1 – Жасанды интеллект, машиналық және терең оқыту

Жалпыға түсінікті болу үшін, жасанды интеллект саласына, әдетте адамдар орындайтын интеллектуалдық тапсырмаларды автоматизациялау деген қысқаша анықтама беруге болады. Алғашқы жасалған бағдарламалар, программистер қойған нақты ережелер бойынша логикалық тапсырмаларды орындады. Символикалық ЖИ деп аталатын бұл тәсіл, 1950 жылдардан 1980 жылдарға дейін басым парадигма ретінде қолданылып келді. Алайда суреттерді классификациялау, сөздерді тану және өзге тілдерге аудару сияқты күрделі тапсырмаларға, нақты ережелер қолдану мүмкін еместігі анықталды. Осылайша символикалық ЖИ интеллекттің орнын машиналық оқыту ауыстырды.

Символикалық ЖИ парадигмасында, енгізілген мәліметтер алдын –ала бекітілген ережелер бойынша өңделіп, нәтижелер алынады. Машиналық оқытуда, тиісінше, мәліметтер мен алынатын нәтижелер енгізіліп, шығыстан ережелер алынады. Алынған ережелер, түпнұсқалық жауаптар алу үшін жаңа мәліметтерге қолданылады (2.2-сурет).



Сурет 2.2 – Машиналық оқыту және символикалық ЖИ салыстыру

Машиналық оқыту кезінде жүйе оқытылады. Жүйеге шешілуге тиісті тапсырмаға қатысты үлгілер беріледі, ал нәтижесінде тапсырманы автоматты түрде шешуге арналған ережелерді алуға мүмкіндік беретін статистикалық құрылымдарды табады.

Машиналық оқытудың дамуы тек 1990 жылдары қарқын ала бастады. Бұл тенденция неғұрлым жылдам әрекет ететін аппаратуралардың пайда болуы және мәліметтердің үлкен жиынтығымен тұспа-тұс келіп, ЖИ-тің кең таралған және сәтті бөліміне айнала бастады. Машиналық оқыту математикалық статистикамен тығыз байланысты болғанымен, кейбір ерекшеліктері де бар. Машиналық оқыту үлкен әрі күрделі мәліметтер жиынтығымен жұмыс істейді. Ал мәліметтердің мұндай көлеміне статистикалық анализдің классикалық әдістерін қолдану мүмкін емес. Нәтиже ретінде, машиналық оқыту, әсіресе терең оқыту, терең математикалық платформаларға ие емес болғандықтан, көбінесе инженерлік шешімдерге негізделеді. Бұл, иедялар теориялықтан гөрі эмпирикалық әдістермен жиі дәлелденетін пән болып табылады.

Терең оқытуға анықтама беру және терең оқыту мен басқа да машиналық оқыту әдістерімен айырмашылығын түсіну үшін жалпы машиналық оқыту алгоритмдерінің қандай тапсырмаларды орындайтығын қарастырған жөн. Жоғарыда айтып өткеніміздей, машиналық оқыту күтілетін нәтижелер үлгілері бойынша мәліметтерді өңдеу тапсырмасының ережелерін анықтайды. Яғни машиналық оқыту жүргізу үшін үш құрауыш қажет:

- кіріс мәліметтер – мысалы, егер дыбысталған сөздерді тану тапсырмасы орындалу керек болса, кіріс мәліметтер ретін түрлі адамдардың дыбысталған сөздерінің жазбалары бар файл бола алады; егер суреттерді классификациялау тапсырмасы жағдайында, мұндай кірістер суреттер болады;

- күтілетін нәтижелердің үлгілері – дыбысталған сөздерді тану тапсырмасында дыбыстық файлдардың транскрипциясы болады; суреттерді классификациялау тапсырмасында күтілетін нәтиже «ит», «мысық» және т.б. тегтер болады;

- алгоритмнің жұмысы сапасын бағалау әдістері – бұл алгоритм қайтаратын нәтижелердің, күтілетін нәтижелерден ауытқуын анықтауға қажет.

Машиналық оқыту моделі бастапқы мәліметтерді кіріс белгілердің белгілі үлгілер мен нәтижелердің негізінде оқытылып, маңызы бар нәтижелерге түрлендіреді. Яғни, машиналық және терең оқытудың негізгі тапсырмасы мәліметтерді маңызды бар түрлендіруді немесе күтілетін нәтижеге

жақындататын кіріс мәліметтерді ұсынуды оқыту. Мысалы, түрлі түсті суретті RGB (red-green-blue) немесе HSV (hue-saturation-value) форматында коддалауға болады. Яғни, мәліметтер бір болғанымен, ұсыну екі түрлі. Мәліметтерді өңдеудегі кейбір тапсырмаларды орындау бір ұсыну түрінде қиын болса, өзге ұсынуда жеңіл болады. Машиналық оқыту моделінің негізгі тапсырмасы кіріс мәліметтерді тиімді ұсынуды іздеу – мәліметтерді классификация сияқты тапсырманы шешуге икемді ұсынуға түрлендіру.

Машиналық оқыту контекстіндегі оқыту неғұрлым тиімді ұсынуды автоматты іздеу процесі. Бұл операциялар координатаны өзгерту, сызықтық проекцияларды орындау, трансляция, сызықтық емес операциялар және т.б. болуы мүмкін. Машиналық оқыту алгоритмдері әдетте гипотездер кеңістігі деп аталатын алдын ала анықталған операциялар жиынтығында іздеуді орындайды.

Техникалық тұрғыда, машиналық оқыту – кері байланыс сигналын қолдана отырып, алдын-ала анықталған мүмкіндіктер кеңістігінде әлдебір мәліметтер кірісінің маңызы бар ұсынуын іздейді. Аталған қарапайым идея, сөздерді танудан бастап, көлікті автоматты жүргізуге дейінгі, кең ауқымдағы интеллектуалдық тапсырмаларды орындайды.

2.2 Төтенше жағдайларды анықтау үшін аудио мәліметтерді өңдеуге терең оқыту әдістерін қолдану

Терең оқыту – неғұрлым маңызы бар қабаттар (деңгейлер) тізбегін зерттеуге негізделген, мәліметтерді ұсынуды іздеудің жаңа шешімі, машиналық оқытудың ерекше бөлімі. Терең оқытудың негізгі идеясы көпқабатты ұсынуға бағытталады. Мәліметтер моделі бөлінетін қабаттардың саны - тереңдігі (depth) деп аталады [87]. Машиналық оқытудың бұл саласына көпқабатты оқыту және иерархиялық оқыту атаулары сәйкес келеді. Замануи терең оқыту ондаған тіпті жүздеген ұсыну тізбек қабаттарынан тұрады және барлығы да оқытушы мәліметтер есебінен автоматты түрде анықталады. Алайда машиналық оқытудың кейбір шешімдері бір-екі ұсыну қабаттарын зерттеуге бағытталған. Осы себептен оларды таяз оқыту (shallow learning) деп аталады [88].

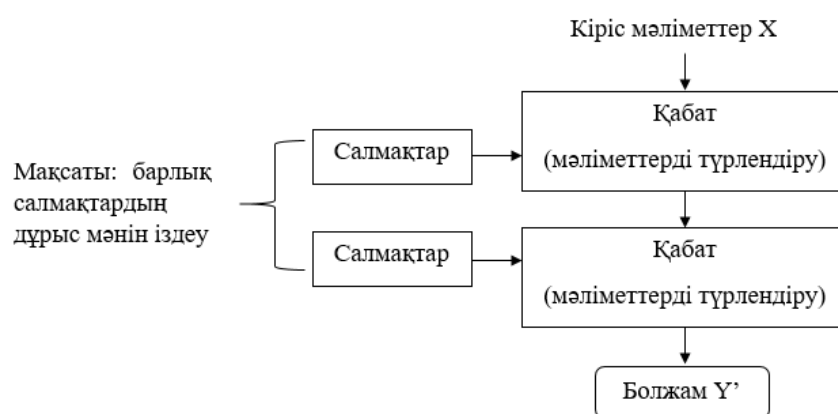
Терең оқытуда көпқабатты ұсыну, құрылымы бір біріне қабаттастырылған нейрондық желілер деп аталатын модельді қолдану арқылы зерттеледі. Терең желіні, орындау барысында ақпарат фильтрлер тізбегінен тазарып шығатын, ақпаратты тазалаудың көпсатылы операциясы ретінде қарастыруға болады. Әрбір сатыдан өткен сайын бастапқы үлгісінен ерекшеленіп, нәтиже туралы неғұрлым пайдалы ақпаратты таситын ұсынуға түрлендіреді. Техникалық тұрғыда мәліметтер ұсынуды алудың көпсатылы әдісі [89].

2.2.1 Терең оқытудың жұмыс істеу принципі

Машиналық оқыту үлкен көлемдегі кіріс мәліметтер үлгісін және нәтижелерді зерттей отырып, кірісті (мысалы сурет) нәтижеге («мысық» деген жазба) түрлендіруге негізделеді. Ал терең нейрондық желілер қарапайым

түрлендірулердің ұзақ тізбегін орындай отырып үлгілер негізінде осы түрлендірулерге оқытылады, кіріс мәліметтерді нәтижеге түрлендіреді.

Кіріс мәліметтерге тиісті деңгейдің қандай амалдар қолданатындығынын, іс жүзінде сандар жинағы болып табылатын, салмақтары арқылы анықталады. Техникалық тұрғыда қабаттарда жүзеге асырылатын түрлендірулерді, 2.3-суретте көрсетілгендей, салмақтар параметрлейді (салмақтарды қабат параметрлері деп те атайды). Бұл контекстте оқыту дегеніміз желідегі барлық қабаттар салмағының мәндер жиынтығын іздеу. Желі, әрбір нәтижеге тиесілі кіріс мәліметтер жиынтығы үлгілерін дұрыс бейнелуі қажет. Басты мәселе, терең нейрондық желі ондаған миллион параметрден тұрады. Олардың әрқайсысына дұрыс мәнді іздеу өте қиын тапсырма, әсіресе, бір параметрдің өзгерісі басқаларының барлығына әсер ететіндей болса.



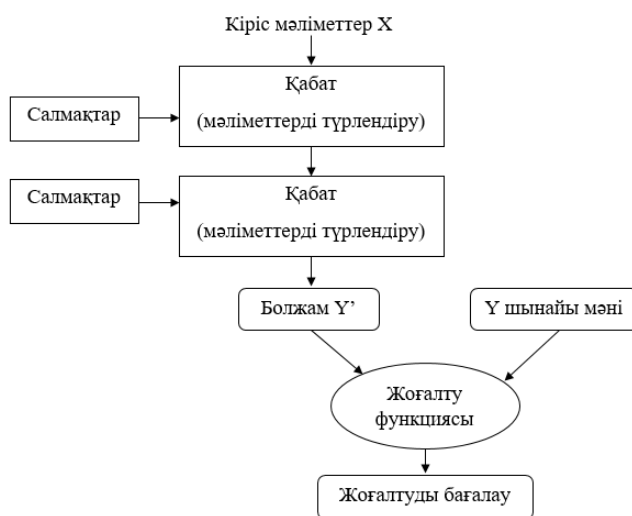
Сурет 2.3 – Нейрондық желінің салмақтармен параметрленуі

Нейрондық желі жұмысының нәтижелерін басқару үшін алынған нәтиже күтілетін нәтижеден қаншалықты алшақ кендігін анықтау мүмкіндігі қажет. Бұл тапсырманы желінің жоғалту функциясы (мақсаттық функциясы деп те аталады. Жоғалту функциясы желіден алынған болжамды және шынайы мәнді (желі қайтаруға тиісті болған) қабылдап, олардың арасындағы арақашықтықты бағалауды (жоғалтуды бағалау) есептейді. Бұл нақты үлгіде, 2.4-суретте көрсетілгендей, мәліметтермен қаншалықты жақсы жұмыс орындағандығын көрсетеді.

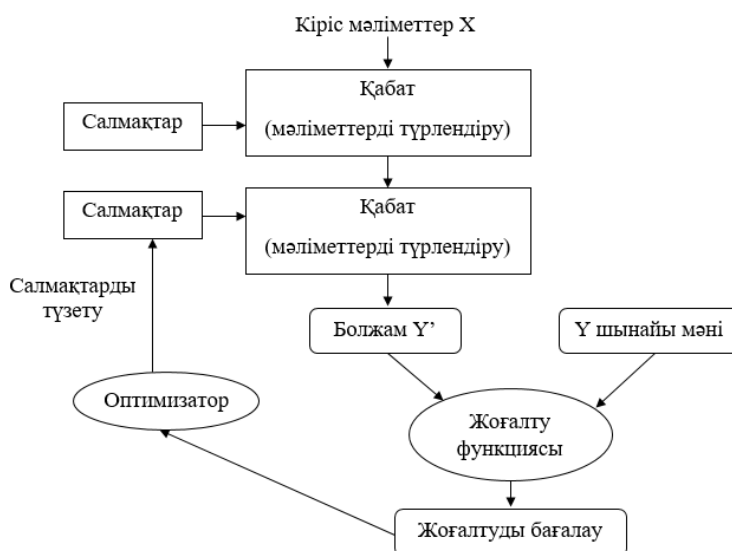
Терең оқытудың ерекшелігі - есептелген жоғалтуды бағалау, 2.5-суреттегі үлгіде көрсетілгендей, жоғалтуды азайту мақсатында салмақтар мәнін түзету үшін қолданылуы болып табылады [90]. Бұл түзету, қатенің кері таралу алгоритмін жүзеге асыратын - оптимизатордың тапсырмасы болып табылады: терең оқытудың орталық алгоритмі.

Алғашында желі салмақтарына кездейсоқ мәндер еншіленеді, яғни, іс жүзінде желі кездейсоқ түрлендірулер тізбегін жүзеге асырады. Әлбетте, алынған нәтиже идеалдан алшақ, жоғалтуды бағасы өте жоғары болады. Желіде әрбір кезекті үлгі өңделген сайын, салмақтар тиісті бағытта түзетіледі де, жоғалту бағасы азаяды. Бұл оқыту циклы бірнеше рет қайталанып (әдетте мыңдаған үлгімен ондаған итерация), жоғалту функциясын минимизациялайтын

салмақ мәндері алынады. Шынайыға жақын нәтиже алынатын, минималды жоғалтуға ие желі – оқытылған желі деп аталады.



Сурет 2.4 – Жоғалтуды бағалау



Сурет 2.5 – Жоғалту бағасының салмақтарды түзетуге кері байланыс ретінде қолданылуы

Терең оқыту машиналық оқытудың бөлімі болғанымен, оның дамуы 2010 жылдардың басында бастау алды. Ондаған жылда көру, есту және қабылдауды модельдеуде айтарлықтай жетістіктерге жетті. Терең оқыту машиналық оқытудың келесідей дәстүрлі күрделі салаларында серпілістер жасалды:

- адам деңгейінде суреттерді классификациялау;
- адам деңгейінде сөздердің дыбысталуын тану;
- адам деңгейінде қолжазбаларды тану;
- бір тілден екінші тілге машиналық аударудың сапасын арттыру;
- мәтінді машиналық дыбыстап оқудың сапасы арттыру;

- Google Now және Amazon Alexa сияқты цифрлық көмекшілердің пайда болуы;
- адам деңгейінде автокөлікті жүргізу;
- Google, Baidu, Bing компаниялары қолданатын мақсатты жарнамалардың нақтылығы артты;
- интернеттегі іздеудің релеванттығын арттыру;
- дыбыстық сұрақтарға жауап беру мүмкіндігінің пайда болуы.

Терең оқытудың әлі де ашылмаған мүмкіндіктері әлі де зерттелуде. Қзіргі кезде табиғи тіл мен машиналық қабылдаудың шекарасынан әрі жатқан формалды пайымдау сияқты кең ауқымды бағыттарда қолдану бойынша зерттеулер жүргізілуде. Бұл саладағы сәтті нәтижелер терең оқыту адамдарға ғылымда, программалаық қамсыздандыруларды жасауда және т.б. көмектесу арқылы жаңа дәуірдің басталуына жол ашады.

Нейрондық желілер қоғамдық назар мен инвестицияға қол жеткізгенімен, машиналық оқытудың алғашқы формасы емес. Өндірісте қолданылатын машиналық алгоритмдердің көбісі терең оқыту алгоритмдері емес. Терең оқыту барлық тапсырмаға қолайлы құрал бола бермейді. Терең оқытудың машиналық оқытудағы алатын орнын көру және қандай себептерден пайда болғандығын және не себептен ерекше маңызға ие екендігін түсіну үшін, оған дейін қолданылған алгоритмдерді білу маңызды.

Ықтималдық модельдеу – мәліметтер анализіне статистика принциптерін қолдану. Бұл машиналық оқытудың ең алғашқы формаларының бірі болғанымен, әлі күнге дейін кеңінен қолданылады. Бұл категориядағы кең қолданыс тапқан «Найв Байес алгоритмі».

Найв Байес алгоритмі – кіріс мәліметтердің тәуелсіздігін анық болжауға арналған Байес теоремасын қолдануға негізделген машиналық оқыту классификаторының түрі [71, р. 83-86; 72, р. 329; 91]. Мәліметтер анализінің бұл формасы компьютерлер әлі де пайда болмаған кезде қолмен есептелетін. Аталған алгоритммен тығыз байланысты логистикалық регрессия (қысқартылған түрде logreg) моделі, классификация алгоритмі болып табылады. Модель әлі күнге дейін қарапайымдылығы және әмбебап табиғаты есебінен сұранысқа ие.

Қарапайым түрдегі нейрондық желілердің негізгі идеялары 1950 жылдарда зерттеле бастады. Үлкен нейрондық желілерді оқытудың эффективті әдісі болмағандықтан ұзақ уақыт бойы дамуы тежеліп тұрды. 1980 жылдары бірнеше зерттеушілердің тәуелсіз зерттеулері нәтижесінде қателердің кері таралу алгоритмі – градиентті түсу әдісін қолдану арқылы параметрлік операциялар тізбегін оқыту әдісі ашылып, нейрондық желілерге қолданылды.

Нейрондық желінің алғашқы сәтті практикалық қолданылуы - 1989 жылы Bell Labs-тағы Ян Лекун қабаттық нейрондық желілер (CNN) мен қателердің кері таралу алгоритмін біріктіріп, қолжазбалық сандарды тану тапсырмасын шешуге қолдануы болды. Нәтижесінде алынған нейрондық желі LeNet деп аталып 1990 жылдары АҚШ-тың пошта қызметінде конверттердегі пошталық индекстерді автоматты тануға қолданылды [89, р. 15].

Ядролық әдістер – классификация алгоритмдерінің тобы. Неғұрлым кең қолданыс тапқан алгоритм – тірек векторлары әдісі (Support vector Machine, SVM) [71, p. 70-75; 89, p. 15]. Тірек векторлары әдісі, түрлі категорияларға тиесілі нүктелердің екі жиынын бөлетін тиімді шешуші шекараларды іздеу арқылы, классификация тапсырмаларын шешуге арналған. Шешуші шекара екі категорияға тиесілі кеңістіктегі оқыту мәліметтері жиынтығын бөлетін сызық немесе жазықтық бола алады. Жаңа нүктелерді классификациялау үшін екараның қай бетінде жатқандығын тексеру жеткілікті.

Мәліметтерді неғұрлым жоғары өлшемдегі кеңістікте бейнелеу үшін kernel trick процедурасы қолданылады. Жаңа кеңістікте тиімді шешуші гипержазықтықты табу үшін нүктелердің сол кеңістіктегі координаталарын есептеуді қажет етпейді, жұп нүктелері арасындағы арақашықтықты есептеу жеткілікті. Бұл тапсырманы ядро функциясы арқылы эффективті орындауға болады. Ядро функциясы – жаңа ұсынуды есептеуді толығымен айналып өтіп, бастапқы кеңістіктегі екі нүктені бейнелеп, олардың арасындағы мақсатты ұсыну кеңістігіндегі арақашықтықты есептеуге мүмкіндік беретін, шығынсыз есептеу операциясы. Ядро функциясы әдетте қолмен анықталады. Тірек векторлары әдісі жағдайында мәліметтер бойынша тек бөлуші гипержазықтық анықталады.

Тірек векторлары әдісі машиналық оқыту әдістері ішіндегі теориялық база және математикалық анализ арқылы түсіндіруге мүмкін болатын әдістердің бірі болғандықтан кең қолданыс тапты. Алайда, таяз әдістерге жататындықтан, суреттер классификациясы сияқты үлкен мәліметтермен жұмыс істеуге қолдану қиындық туғызады.

Шешім ағаштары - алдын ала берілген мәндер бойынша шығыс мәндерді болжау немесе кіріс мәліметтерді классификациялауға мүмкіндік беретін иерархиялық құрылым [71, p. 83-86; 89, p. 16]. Олар оңай визуализацияланады және түсіндіріледі. Мәліметтер негізінде жасалатын шешім ағаштары 2000 жылдары зерттеушілердің қызығушылығын тудырып, 2010 жылдары ядролық әдістерге қарағанда тиімді деп бағаланды.

Жекелей алғанда, кездейсоқ орман алгоритмі шешім ағаштары негізінде оқытудың тиімді шешімін ұсынды. Ол арнайы шешім ағаштарының үлкен санын жасап, олардан алынған нәтижелерді біріктіруді орындайды. Кездейсоқ орман тапсырмалардың кең шеңберіне қолданылатындықтан, таяз машиналық оқыту тапсырмаларының кез-келгеніне тиімді болып табылады. 2010 жылы баршаға белгілі машиналық оқыту бойынша Kaggle конкурстық веб-сайты бастау алғанда, 2014 жылы градиентті бустинг пайда болғанға дейін неғұрлым белгілі платформа ретінде қолданылды. Градиентті бустинг кездейсоқ орманға ұқсас болып келеді. Болжаудың әлсіз модельдерін, әдетте шешім ағаштары, біріктіруге негізделген. Ол кез-келген алдыңғы модельдердегі әлсіз тұстарды жоюға арналған жаңа модельдерді итеративті оқыту арқылы, машиналық оқыту модельдерін жақсарту әдісін - градиентті бустингті қолданады. Градиентті бустинг, шешім ағаштарын қолдану барысында аналогтық қасиеттерін сақтай отырып, кездейсоқ орманнан басым модельдерді алуға мүмкіндік береді.

2.2.2 Терең оқытудың ерекшеліктері

2010 жылы ғылыми қоғам тарапынан нейрондық желілерге қызығушылықтың толықтай жойылғандығына қарамастан, біршама зерттеушілер жұмыстарын жалғастырып, маңызды жетістіктерге қол жеткізді: Торонто университетінен Джеффри Хинтонның тобы, Монреал университетінен Йошуа Бенгио, Нью-Йорк университетінен Яна Лекун және Швейцариядағы IDSIA жасанды интеллект ғылыми-зерттеу институтының зерттеушілері.

2011 жылы IDSIA-дан Ден Киресан GPU-да оқытылатын терең нейрондық желілерді қолдану арқылы суреттерді классификациялау бойынша академиялық конкурсты жеңіп алды. Бұл замануи терең оқытудағы алғашқы практикалық жетістік еді. Алайда 2012 жылы Хинтон тобы жыл сайын өтетін ірі масштабтағы үлгілерді тану бойынша ImageNet жарысына қатысты. Тапсырма ретінде 1,4 млн. жоғары сападағы түрлі түсті суреттен және 1000 категориядан тұратын, классификациясы берілді. 2011 жылғы жеңімпаз модель 74,3% ғана болса, Джеффри Хинтон кеңесшілігімен Алекса Крижевский командасы 83,6% нақтылыққа жетті. Содан бері аталған жарыста алдыңғы позицияларды терең қабаттық нейрондық желілер алуда. 2015 жылы нақтылық 96,4%-ды құрап, ImageNet ұсынған классификация тапсырмасы шешілді деп танылды.

2012 жылдан бері терең қабаттық нейрондық желілер үлгілерді тану тапсырмаларында алдыңғы разрядты алгоритм ретінде танылуда. Бұл әдіс көптеген тапсырамдарда тірек векторлары әдісі және шешім ағаштары әдістерін толықтай ауыстырды. Мысалы: Ядролық зерттеулер бойынша Еуропалық ұйым (European Organization for Nuclear Research, CERN) бірнеше жылдар бойы үлкен адрондық коллайдердегі ATLAS бөлшектер детекторынан алынатын мәліметтер үшін шешім ағаштарын қолданатын. Алайда үлкен көлемдегі мәліметтер жиынтығын оқытудағы қарапайымдылығы мен жоғары өнімділігі себебінен Keras негізіндегі нейрондық желілерді қолдануға өту туралы шешім қабылдады.

Терең оқыту машиналық оқытудағы бұрын қолмен орындалатын маңызды қадам белгілерді конструкциялауды толықтай автоматтандырып, тапсырманы шешуді жеңілдетеді. Яғни, бастапқы мәліметтерді өңдеуге тиімді түрге қолмен түрлендіруге, мәліметтерді ұсыну қабаттарын жақсартуға күш жұмсалды. Бұл белгілерді конструкциялау деп аталады. Ал терең оқыту әдістерін қолдану арқылы бір айналымда барлық белгілерді алуға болады.

Терең оқытудың әдістемесі екі маңызды сипаттамаға ие: саты бойынша неғұрлым күрделі ұсынуларды қабат бойынша конструкциялайды және аралық ұсынуларды бірге зерттейді. Нәтижесінде әр қабат төменгі және жоғарғы қабаттардың ұсыну қажеттіліктеріне сәйкес жаңарып отырады. Осы екі қасиет терең оқытуды басқа шешімдерге қарағанда сәтті болуына негізгі себеп болып табылады.

Терең оқытудың үлгілерді тану тапсырмасын шешудегі негізгі екі идея – қабаттық нейрондық желілер мен қатенің кері таралу алгоритмі 1989

жылдардан-ақ белгілі болған. Уақыт қатарларын болжауға арналған, терең оқытудың негізін құраушы ұзақ қысқа мерзімді жады (Long Short-Term Memory, LSTM) алгоритмі 1997 жылдан бері ешқандай өзгеріске ұшыраған жоқ. Осы ретте «терең оқыту не себептен 2012 жылдан бастап қана қолданыла бастады» деген сұрақтың туындауы орынды.

Жалпы, машиналық оқытудың үш техникалық қозғаушы күші белгілі:

- қондырғы;
- мәліметтер жиынтығы және тесттер;
- алгоритмдік жетістіктер.

Машиналық оқыту инженерлік ғылым болғандықтан, теориялықтан гөрі эксперименталдық шешімдерге сүйенеді. Алгоритмдік жетістіктер, идеяны тексеруге жарамды қондырғы мен мәліметтер жиынтығының болуына тәуелді. 1990 және 2000 жылдар мәліметтер және қондырғы тұрғысынан өте тар түйін болды. Алайда соңғы жылдарда жоғары өнімділікке ие графикалық процессорлардың жасалуы және интернеттің қарқынды дамуы бұл салаға үлкен жол ашуда.

1990 жылдар мен 2010 жылдар аралығында стандартты процессорлардың әрекет жылдамдығы 5000 есе өскен. 2015 жылдың аяғында, құны 1000 АҚШ долларын құрайтын NVIDIA TITAN X замануи графикалық процессоры 6,6 терафлопс өнімділікті, секундына float32 сандар типіндегі 6,6 триллион операцияны орындау мүмкіндігіне ие болды. Сонымен қатар терең оқыту индустриясы GPU шеңберінен тыс арнайы процессорларға инвестиция жасауда. 2016 жылы Google I/O конференциясында Google компаниясы тензорлық процессор (Tensor Processing Unit, TPU) жобасын ұсынды. Жаңа архитектураға ие процессордың өнімділігі GPU-дің алдыңғы қатарлы модельдерінен 10 есе жоғары.

Мур заңына сәйкес соңғы 20 жылда байқалған есте сақтау құрылғылары сыйымдылығының экспоненциалдық артуы, машиналық оқыту үшін үлкен көлемдегі мәліметтерді таратуға, сақтауға үлкен мүмкіндік туғызды.

2000 жылдардың аяғында мәліметтер мен қондырғылардан бөлек, өте терең нейрондық желілерді оқытуға арналған сенімді алгоритмдер де тапшы болды. Негізгі мәселе қаттардың терең пакеті арқылы градиенттің таралуы болып табылды. Нейрондық желіні оқытуға қолданылатын кері байланыс қабаттар саны көбейген сайын азайып қалып отырды. Жағдай тек 2009-2010 жылдары градиенттің таралуын жақсартуға мүмкіндік берген қарапайым әрі маңызды алгоритмдік жетілдірудің пайда болуымен өзгерді:

- жақсартылған активация функциялары;
- салмақ инициализациялары сұлбасының жақсаруы;
- RMSProp және Adam сияқты жақсартылған оптимизация сұлбалары.

Нәтижесінде, 2014, 2015, 2016 жылдары градиентті таратудың неғұрлым дамытылған әдістері, пакеттік нормализация, айналым байланыстары, бөлінетін қабаттар пайда болды. Қазіргі таңда жаңа модельді мындаған қабат тереңдікпен оқыту мүмкіндігі бар.

Терең оқытудың қасиеттерін үш категорияға бөлуге болады:

1. Қарапайымдылығы – күрделі, қарама-қайшы, ауыр конвейерлерді, бес-алты тензорлық операция ғана қолдану арқылы оқытылатын қарапайым модельдермен ауыстыра отырып, белгілерді конструкциялау қажеттілігінен босатады.

2. Масштабталуы – терең оқыту GPU немесе TPU параллельденуіне икемді, сондықтан ол толықтай Мур заңын қолдана алады. Сонымен қатар, модельдерді оқытуды шағын мәліметтер пакетінде итеративті түрде орындауға болады. Бұл кез-келген көлемдегі мәліметтерде оқытуды жүргізуге болады дегенді білдіреді.

3. Бірнеше рет қолдануға дайын болуы және икемділігі – алдыңғы шешімдерге қарағанда, терең оқыту модельдері, қосымша мәліметтерді оқыту үшін оқытуды басынан қайта орындамай-ақ жүргізуге мүмкіндік береді, яғни үздіксіз және кезеңдік оқытуға жарамды. Бұл үлкен өндірістік модельдердің маңызды қасиеті. Сонымен қатар оқытылған модельдердің мақсатын ауыстырып, бірнеше рет қолдануға да болады: мысалы: сурет классификациясын орындауға оқытылған модельді, бейнені өңдеу конвейеріне қосуға болады. Бұл алдыңғы алынған қабілетті неғұрлым күрделі әрі қуатты модельдерге қолдануға мүмкіндік береді. Бұл терең оқытуды өте кіші көлемдегі мәліметтерге қолдануға мүмкіндік береді.

2.3 Нейрондық желілердің математикалық негіздері

Терең оқытуды түсіну үшін көптеген қарапайым математикалық ұғымдарды білу қажет: тензорлар, тензорлармен орындалатын операциялар, градиентті түсу, дифференциалдау, векторизация және т.б.

Замануи машиналық оқыту жүйелерінің барлығы дерлік мәліметтердің негізгі құрылымы ретінде тензорларды қолданады. Іс жүзінде тензор бұл – сандық мәліметтер контейнері. Басқаша айтсақ, сандарға арналған контейнер. Тензор – бірнеше өлшемдер санын жалпыландыратын матрица (тензорлар терминологиясында өлшемді осьтер деп атайды).

Жалғыз саннан тұратын тензорды скаляр деп атайды (скаляр немесе нөлдік ранг тензоры). Numpy массивіндегі float32 немесе float64 сандар типі – скаляр тензор болып табылады. Numpy тензорындағы осьтер санын ndim атрибуты арқылы анықтауға болады. Скаляр тензордың осьтер саны 0-ге тең. Тензордың осьтер саны ранг деп аталады.

Сандардың бірөлшемді массивін вектор немесе бірінші рангты тензор деп атайды. Оның бір ғана осі болады.

Бұл вектор төрт элементтен тұратындықтан, төрт өлшемді вектор деп аталады. Төрт өлшемді вектор мен төрт өлшемді тензор екі түрлі ұғым екендігін ескерген дұрыс. Төрт өлшемді вектор бір ғана осьтен және осы осьтегі төрт мәннен, ал төртөлшемді тензор төрт осьтен тұрады (әр ось бірнеше мәннен тұруы мүмкін).

Векторлар массиві – матрица немесе екінші ранг тензоры деп аталады. Матрица екі осьтен тұрады (қатарлар мен бағандар). Осы матрицаны жаңа

массивке орналастырса, сандық куб түрінде ұсынуға болатын үш өлшемді тензор алынады.

Үш өлшемді тензорды келесі жаңа массивке орналастырсақ, төртінші рангты тензор аламыз және т.с.с. Терең оқытуда бір және төрт ранг арасындағы тензорлар жиі қолданылады. Ал кейде, мысалы бейне мәліметтерді өңдеуде бесінші ранг тензоры қолданылуы мүмкін.

2.3.1 Тензорларды анықтауға қолданылатын атрибуттар

Тензор негізгі үш атрибут арқылы анықталады:

1. Осьтер саны (ранг) – мысалы, үшінші ранг тензоры үш, ал матрица екі оське ие. Numpy сияқты Python библиотекасы үшін бұл тензор атрибуты `ndim` деп аталады [88, p. 100045-1-100045-12; 90, p. 240-247].

2. Форма – әр тензор осіндегі өлшемдер санын сипаттайтын, бүтін сандар кортежі. Мысалы 2.3-суреттегі матрица формасы (3, 5), ал 2.4-суреттегі үшінші ранг тензорының формасы (3, 3, 5). 2.2-суреттегі вектордың формасы (4), ал скалярдың формасы бос формада болады.

3. Мәліметтер типі (Python үшін библиотекада `dtype` деген ат берілген) – тензорды құрайтын мәліметтер типі, мысалы, тензор `float32`, `uint8`, `float64` және т.б. Сирек жағдайларда `char` типтегі тензорларды кездестіруге болады. Numpy-да қатар тензорлары болмайтынын ескеру керек. Себебі тензорлар алдын-ала бөлінген, үзіліссіз қатар және жады сегменттерінде, ұзындығы өзгертін мән ретінде сақталады [88, p. 100045-1-100045-12; 90, p. 240-247; 91, p. 88-90].

Жалпы жағдайда терең оқытудың барлық тензорларындағы бірінші ось (индексі 0 болатын ось, себебі нөмірлеу 0-ден басталады) үлгілер осі (кейде үлгілер өлшемі деп те аталады) болады. Сонымен қатар терең оқыту модельдері мәліметтер жиынтығын бүтіндей емес, бірнеше шағын пакеттерге бөле отырып өңдейді.

Басқаруға тиісті мәліметтерәрдайым келесі категориялардың біріне тиесілі болады:

– векторлық мәліметтер – екі өлшемді тензорлар, формасы (үлгілер, белгілер);

– уақыт қатарлары мен тізбектері – үш өлшемді тензорлар, формасы (үлгілер, уақыт таңбалары, белгілері);

– суреттер – төрт өлшемді тензорлар, формасы (үлгілер, биіктігі, ені, түсі) немесе (үлгілер, түсі, биіктігі, ені);

– бейне – бес өлшемді тензорлар, формасы (үлгілер, кадрлар, биіктігі, ені, түсі) немесе (үлгілер, кадрлар, түсі, биіктігі, ені).

Векторлық мәліметтер ең жиі кездесетін мәліметтер формасы. Мұндай жиынтықтарда үлгілер вектор түрінде, ал пакет екі өлшемді тензор түрінде ұсынылады, бірінші ось – үлгілер осі, екіншісі – белгілер осі. Мысалы, адамдар туралы мәліметтер жиынтығында жасы, посталық индексі және кірісі туралы ақпараттар бар делік. Әр адам үш мәні бар вектормен сипатталады, яғни 100000 адамды сипаттайтын бүтін мәліметтер жиынтығын формасы (100000, 3) болатын екі өлшемді тензорда сақтауға болады.

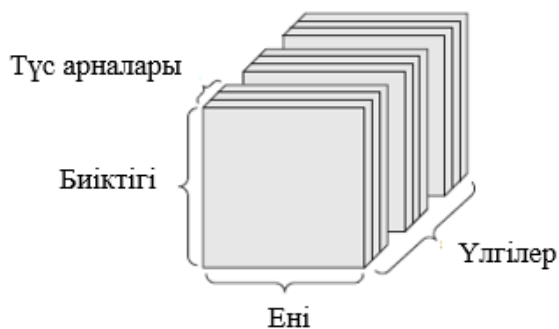
Уақыт маңызды рөл атқаратын мәліметтерді сақтау үшін үш өлшемді нақты уақыт осі бар тензорды пайдаланған дұрыс. Әр үлгі векторлар тізбегі түрінде, ал пакеттің өзі үш өлшемді тензор түрінде ұсынылады (2.6-сурет).



Сурет 2.6 – Уақыт қатары бар үш өлшемді тензор

Келісімге сәйкес уақыт осі әрдайым екінші ось болады (ось индексі 1). Мысалы, акция бағаларынан тұратын мәліметтер жиынтығын алынсын. Әр минуттағы акция бағасын, сол минуттағы ең жоғарғы және ең төменгі бағаны сақтап отырайық. Яғни әр минут үш өлшемді вектормен ұсынылады, ал сауда күні формасы (390, 3) екі өлшемді тензормен (390 – сауда күнінің минуттағы ұзақтығы), ал 30 күндегі мәліметтер формасы (30, 390, 3) үш өлшемді тензор ұсынылады. Бұл жағдайда әр үлгі бір сауда күніндегі мәліметтерді ұсынады.

Әдетте суреттер үш өлшемнен тұрады: биіктігі, ені, түсі. Тіпті қара-ақ сурет тек 1 ғана түс арнасынан тұрса да, келісім бойынша сурет бар тензорлар үш өлшемнен тұрады. Тиісінше, 256x256 өлшемді 128 ақ-қара суреттен тұратын пакет, формасы (128, 256, 256, 1), ал 128 түрлі түсті суретті пакет, формасы (128, 256, 256, 3) тензор болып сақталады (2.7-сурет).



Сурет 2.7 – Суреттен тұратын төрт өлшемді тензор

2.7-суреттен тұратын төрт өлшемді тензорлар формасы үшін екі келісім бар: арна соңғы тұрады (TensorFlow үшін), арна бірінші тұрады (Theano үшін). Keras фрейворкы екі форматты да қамтиды.

Бейнемәліметтер – сақтау үшін бес өлшемді тензорлар қолданылатын мәліметтер типі болып табылады. Бейнені, әрбір кадр түрлі түсті сурет болатын, кадрлар тізбегі ретінде қарастыруға болады. Әр кадрды үш өлшемді тензор ретінде сақтауға болады (биіктігі, ені, түсі), ал олардың тізбегін төрт өлшемді тензорда (кадрлар, биіктігі, ені, түсі) сақтауға болады. Түрлі бейне

роликтер пакеті, формасы (үлгілер, кадрлар, биіктігі, ені, түсі) болатын бес өлшемді тензорда сақталады.

2.3.2 Тензорларға қолданылатын амалдар

Оқыту барысында терең оқыту желілері орындайтын түрлендірулердің барлығы сандық мәліметтерден тұратын тензорларға қолданылатын тензорлық операцияларға негізделеді. Мысалы, тензорларды көбейту, қосу және т.б.

Бір-біріне қабаттастырылған екі dense қабатынан тұратын желіні мысал ретінде қарастырайық. Бұл жағдайда функция келесі түрде жазылады [89, p. 38]:

$$output = relu(dot(W, input) + b) \quad (2.1)$$

мұнда W – екі өлшемді тензор;

b – вектор.

Өлшемдердің екеуі де қабат атрибуттары болып табылады. (2.1) функцияда тензорлармен үш операция орындалады: $input$ кіріс тензорынан және W тензорынан скалярлы туынды алу (dot); алынған екі өлшемді тензор мен b векторын (+) қосу; $relu.relu(x)$ операциясы.

$Relu$ операциясы мен қосу жеке элементтерге қолданылатын операциялар, яғни тензордың әр элементі үшін жеке қолданылады. Бұл операциялар жаппай параллельдендіруге икемді болып табылады (векторизация). Python-да жеке элементтер операциясын орындау үшін for циклі қолданылады [92, 93].

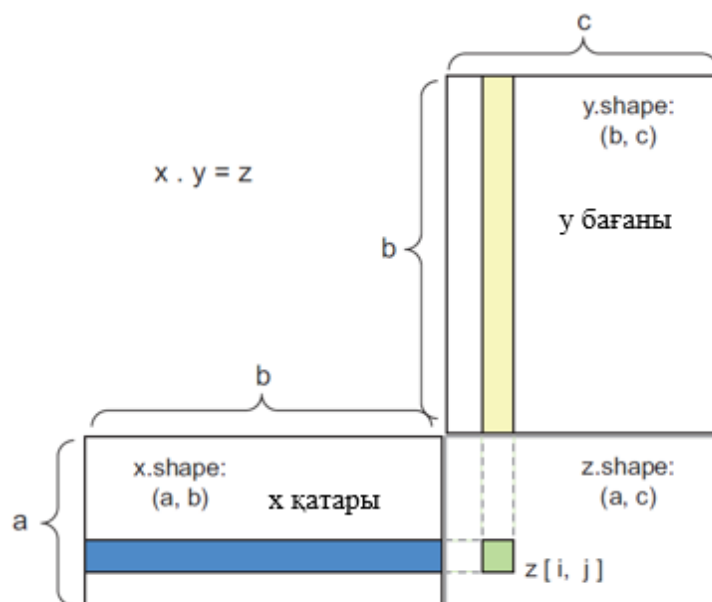
Қосылатын тензорлардың формалары әр түрлі болған жағдайда кіші тензордың формасы, үлкен тензордың формасына сәйкестендіріліп, жаңа формаға кеңейеді (broadcasting). Кеңею екі саты арқылы орындалады:

1. Кіші тензорға $ndim$ атрибутының мәні үлкен тензордың осы атрибутының санынан сәйкес келетіндей етіп, осьтер қосылады (кеңею осьтері).

2. Кіші тензор, үлкен тензордың формасына толық сәйкестендірілгенге дейін жаңа оське көшіріледі.

Іс жүзінде жаңа екі өлшемді тензор жасалмайды. Көшіру операциясы тек виртуалды түрде, яғни жады емес, тек алгоритмдік деңгейде орындалады.

Скаляр көбейту, тензорлық көбейту – неғұрлым жалпы және тиімді тензорлар операциясы. Жеке элемент операцияларынан ерекшелігі, ол кіріс тензорлар элементтерін біріктіреді. Екі векторды скаляр көбейтудің нәтижесінде скаляр алынады және операцияларға бірдей элементтер саны бар векторлар қатыса алады. Сонымен қатар x матрицасын y векторына скаляр көбейтуге болады. Скаляр көбейту үшін матрица формаларының сәйкестігін анықтау түсінікті болу үшін 2.8-суретте мысал келтірілген.



Сурет 2.8 – Матрицаларды скаляр көбейту диаграммасы

Тіктөртбұрыш түрінде x , y және z матрицалары келтірілген. x матрицасындағы қатар саны мен y матрицасының баған саны, яғни x -тің ені мен y -тің биіктігі сәйкес келуі тиіс. Жалпы жағдайда үлкен өлшемдер санына тұратын тензорларды скаляр көбейту үшін осы мысалда көрсетілгендей формалардың сәйкестігі ережесі орындалуы керек.

Тензорлармен орындалатын операцияның үшінші түрі – тензор формасын өзгерту. Бұл операция желіге беру үшін кіріс мәліметтерді дайындауда қолданылады. Тензор формасын өгерту бекітілген формаға келтіру үшін қатарлар мен бағандарды реттеу. Әрине, формасы өзгертілген тензор бастапқыдағы элементтер санына ие болады.

Практикада жиі кездесетін форма өзгертудің ерекше жағдайы – транспонирлеу. Транспонирлеу – қатарларды бағанға, ал бағандарды қатарларға ауыстыру арқылы матрицаны түрлендіру.

Тензорлар мазмұнын белгілі бір геометриялық кеңістіктегі нүктелер координатасы ретінде түсіндіруге болатындықтан, тензорлармен операциялардың барлығы геометриялық түсінікті де қамтиды. Аффиналық түрлендіру, бұру, масштабтау сияқты т.б. элементарлы геометриялық операцияларды, тензорлармен операция түрінде ұсынуға болады.

2.3.3 Нейрондық желілер механизмі: градиент негізінде оптимизациялау

Нейрондық желінің әр қабаты (2.1) теңдеуде көрсетілгендей түрлендіреді. W және b – қабат атрибуттары болып табылатын тензорлар. Олар қабаттың салмақтары немесе оқытылатын параметрлері деп аталады. Бұл салмақтар оқытатын мәліметтер желісінен алынған ақпараттарды қамтиды. Алғашқыда салмақтық матрицалар кездейсоқ мәндермен толтырылады (бұл қадам кездейсоқ инициализация деп аталады). Алғашқы ұсыну ешқандай мағынаға ие болмағанымен, бастапқы нүкте болып белгіленеді. Әрі қарай, кері байланыс

сигналы негізінде салмақтарды сатылап түзету орындалады. Бұл сатылы түзетулер машиналық оқытудың негізі болып табылатын оқыту деп аталады. Оқыту циклында қолданылатын қадамдар төменде көрсетілген [94]:

- x оқытылатын үлгілер және сәйкес нәтижелер пакеті алынады;
- желі x пакетін өңдейді (тура таралу) және болжамдар (y_{pred}) пакетін алады;
- u пен $u_{\text{болж}}$ арасындағы айырмашылықты бағалайтын, пакеттегі желі жоғалтуы есептеледі;
- пакеттегі жоғалтуды біршама азаятындай, желі салмақтары түзетіледі.

Нәтижесінде оқыту мәліметтері жинағында өте төмен жоғалтуға ие желі алынады. Желі кірістегі мәліметтерді дұрыс ақырғы мәндер түрінде ұсынуға үйренеді.

Бірінші қадам – қарапайым енгізу/шығару операциясы. Екінші және үшінші қадамдар бірнеше тензорлар операциясын ғана қолдану арқылы орындалады.

Төртінші желі салмақтарын түзету, салыстырмалы түрде неғұрлым күрделі қадам. Қарапайым шешімдердің бірі – бір салмаққа ғана түрлі мәндерді қолдану, ал басқа салмақтарды өзгертпей арқылы. Алайда бұл шешім әр салмақ (әдетте мыңдаған, ал кейде бірнеше миллион) үшін тура таралуды орындау керек болатындықтан, эффективті емес. Ал желідегі операциялардың барлығы дифференциаланатын ескеріп, әр салмақ коэффициентіне жоғалту градиентін есептеу тиімді болып табылады. Содан соң коэффициенттерді градиентке қарама-қарсы бағытқа бұру жылжыту арқылы, жоғалтуды азайтуға болады. Мысалы, x нақты санын, жаңа u нақты санында көрсететін, $f(x)=u$ тегіс үзіліссіз функциясы алынсын. Функция үзіліссіз болғандықтан, x -тің аздаған өзгерісі u -тің тек аздаған өзгерісіне әкеледі. Яғни x -ті ε аз шамаға артса $x + \varepsilon$, нәтижесінде u та ε аз шамаға $u + \varepsilon$ артады. Сонымен қатар функция тегіс болғандықтан, ε аз ғана ауытқу шамасы, p нүктесі аймағында a еңкеюі бар f сызықтық функциясымен жуықтауға болады ($a \approx \varepsilon$). Мұндай сызықтық аппроксимация x , p нүктесіне айтарлықтай жақын орналасқанда ғана мүмкін болатындығы анық. a еңкеюі – p нүктесіндегі f туындысы деп аталады. Егер a -ның мәні теріс болса, p нүктесі аймағындағы x -тың аз өзгерісі $f(x)$ -тың азаюына әкеледі. Ал егер a -ның мәні оң болса, p нүктесі аймағындағы x -тың аз өзгерісі $f(x)$ тың артуына әкеледі. Сонымен қатар a абсолютті мәні (туынды шамасы) бұл арту мен азаюдың қаншалықты екендігінен хабар береді.

Егер $f(x)$ мәнін азайту үшін x -ті ε шамаға өзгертуге, f бойынша туындыны білу жеткілікті. Туынды, $f(x)$ -тың x -ті жылжытқандағы өзгерісін толық сипаттайды. $f(x)$ -тың мәнін азайту үшін x -ті туынды бағытына қарама-қарсы жылжыту жеткілікті.

Градиент тензорлар операциясының туындысы. Кірісінде тензорларды қабылдайтын функцияға, яғни көпөлшемді кіріс мәліметтер функциясының туындысы. Мысалы, x кіріс вектор, u нәтижесі, W матрицасы және loss жоғалту функциясы бар. Матрица арқылы $u_{\text{болж}}$ нәтижеге жуықтауды есептеуге және u пен $u_{\text{болж}}$ арасындағы айырмашылықты немесе жоғалтуды анықтауға болады.

Егер кірістегі x және y белгіленсе, онда мұны W мәнін жоғалту мәнінде көрсететін функция деп түсінуге болады. W -ның бастапқы мәні W_0 деп алынсын. W_0 нүктесіндегі f функциясының туындысының тензоры формасы, дәл W формасына сәйкес болады. Ол W_0 $[i, j]$ өзгеруінде байқалатын, әр элементі жоғалту элементінің шамасын және бағытын анықтайды.

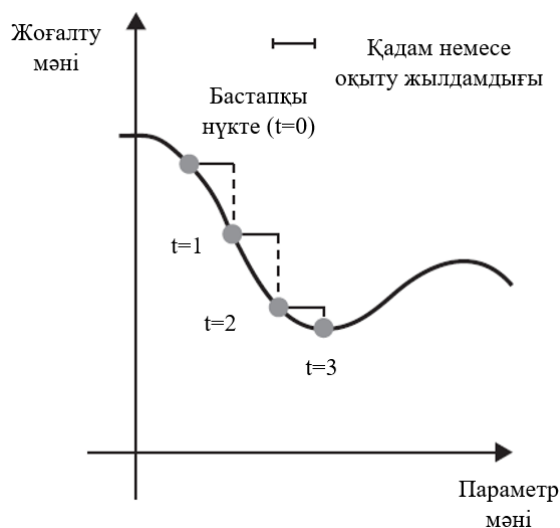
Дифференциаланатын функцияның теориялық минимумын аналитикалық түрде табуға болады. Функцияның минимум нүктесі – туынды 0-ге тең болатын нүкте. Тиісінше, туынды 0-ге ұмтылатын барлық нүктелерді табу және сол нүктелердің қайсысында функция неғұрлым аз мәнге ие екендігін анықтау жеткілікті. Нейрондық желілерге қатысты, бұл салмақтар мәнінің комбинациясын аналитикалық іздеу дегенді білдіреді, яғни W үшін $\text{gradient}(f)(W)=0$ теңдеуін шешу арқылы қол жеткізуге болады [89, р. 48]. Бұл N (желідегі салмақтар сан) айнымалыдан тұратын полиномдық теңдеу. Екі немесе үш айнымалыдан тұратын теңдеуді шешу қиындық тудырмағанымен, кем дегенде бірнеше мыңдаған және ондаған миллионнан параметрден тұратын нейрондық желілер үшін іс жүзінде шешілмейтін тапсырма болып табылады. Сондықтан, практикада төрт қадамнан тұратын алгоритм қолданылады: кездейсоқ мәліметтер пакетіндегі жоғалту мәніне сүйене отырып, параметрлерді аз-аздан өзгерту. Функция дифференциалданатындықтан, осы төрт қадамды эффективті орындауға мүмкіндік беретін, градиентті есептеуге болады. Егер салмақтардың бағытын, жоғалту градиентіне қарама-қарсы бағытқа өзгертсе, әрбір цикл сайын азайып отырады:

- x оқытылатын үлгілер және сәйкес нәтижелер пакеті алынады.
- желі x пакетін өңдейді (тура таралу) және болжамдар (y_{pred}) пакетін алады;
- y пен $y_{\text{болж}}$ арасындағы айырмашылықты бағалайтын, пакеттегі желі жоғалтуы есептеледі;
- желі параметрлері үшін жоғалту градиенті есептеледі (кері таралу);
- параметрлер градиентке қарама-қарсы бағытта аздаған шамаға түзетіле отырып, жоғалтулар азаяды.

Бұл стохастикалық градиенттік түсу деп аталады. Стохастикалық деп аталу себебі мәліметтер пакеті кездейсоқ алынуымен түсіндіріледі. 2.9-суретте бір өлшемді мәліметтер мысалында, бір параметрден және бір ғана үлгіден тұратын желі көрсетілген.

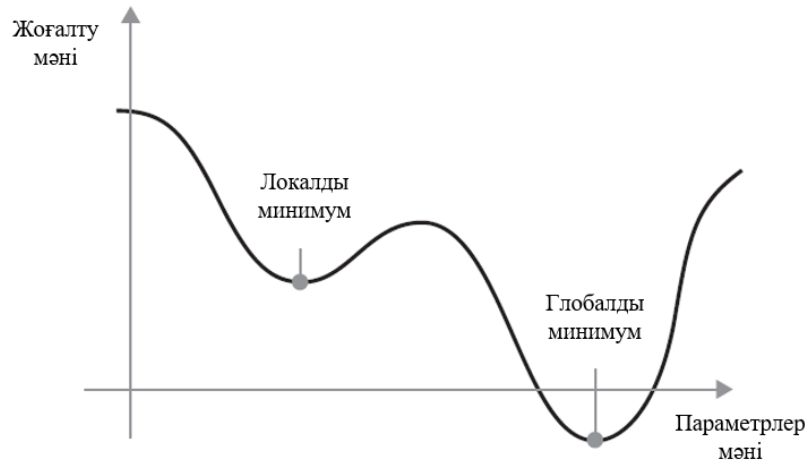
Қадамның тиімді шамасын таңдау өте маңызды. Егер өте аз шама таңдалса итерация саны артып, локалды минимумда тұрып қалу мүмкіндігі бар. Егер қадам тым үлкен болса, тзетулер мақсатсыз сипатқа ие болып, қисықтағы кездейсоқ нүктелерге әкелуі мүмкін. Стохастикалық градиентті түсу алгоритмі әр итерацияда жалғыз үлгі мен нәтижені қолданған. Іс жүзінде бұл шынайы (true) стохастикалық градиентті түсу. Ал әр қадамда барлық қолжетімді мәліметтерді қолданатыни алгоритм пакеттік (batch) стохастикалық градиентті түсу деп аталады. Бұл жағдайда өзгерістер нақты болғанымен, шығынды болып келеді. екеуінің арасындағы тиімді шешім үлкен емес пакеттерді қолдану болып табылады. Мысалда бір өлшемді параметрлер кеңістігіндегі градиентті түсу

келтірілді. Іс жүзінде өлшемдер санының әлдеқайда үлкен кеңістіктегі градиентті түсу қолданылады: нейрондық желідегі әрбір салмақ коэффициенті – кеңістіктегі тәуелсіз өлшемдер.



Сурет 2.9 – Бір өлшемді жоғалту қисығы бойымен градиентті төмен түсу

Стохастикалық градиентті түсудің келесі нұсқалары салмақтардың келесі өсімін есептеуде ағымдағы градиент мәндерін ғана емес, алдыңғы мәндерін де есепке алады. Мысалы импульстік стохастикалық градиентті түсу, RMSProp, Adagrad және т.б. бұл нұсқалар оптимизациялау әдістері немесе оптимизаторлар деп аталады. Импульс екі проблеманы шешу үшін қолданылады: сырғу жылдамдығының төмендігі және локалды минимумға түсу. 2.10-суретте көрсетілгендей нақты параметр мәні үшін локалды минимум бар: бұл нүктеден оңға немесе солға жылжу жоғалтудың артуына әкеледі. Егер қарастырылатын параметрді түзету оқыту қадамы кіші болатын градиентті түсу әдісімен орындалса, оптимизация процесі глобалды минимумды таба алмай, локалды минимумда тұрып қалады. Бұл жағдайда импульс идеясын пайдалану аталған проблемаларды шешуге мүмкіндік береді. Оптимизация процесі жоғалту қисығынан төмен сырғыған кішкене шар түрінде қарастырылады. Егер шардың импульсі жеткілікті болса ғана, кішкене еңісте қалып қоймай, глобалды минимумға жетеді. Импульс ағымдағы еңкею шамасы мен (үдеу) жылдамдықты негізге ала отырып, әр қадамда шарды жылжыта (алдыңғы қадамдағы үдеу әрекеті нәтижесі) отырып жүзеге асырылады. Бұл іс жүзінде W параметрін арттыру, ағымдағы градиент мәні ғана емес, алдыңғы параметр шамасын ескеру арқылы анықталады дегенді білдіреді.



Сурет 2.10 – Локалды және глобалды минимумдер

Жоғарыда айтып кеткендей, егер функция дифференциалданатын болса, туындыны есептеу мүмкін болады. Практикада нейрондық желілер, әрқайсысы қарапайым, белгілі туындылары бар, бірнеше тензорлардың тізбетік операциясы біріктірілген бір тізбектен тұрады. Мысалы W_1, W_2, W_3 салмақ матрицаларынан, a, b, c тензорлар операцияларынан тұратын f желісі бар. Функциялардың тізбегін тізбектік ереже деп аталатын келесі теңдеуді пайдалану арқылы мүмкін болады [89, p. 51]:

$$f(g(x)) = f'(g(x)) * g'(x) \quad (2.2)$$

Нейрондық жедінің градиент мәнін есептеу үшін тізбектік ережені қолдану қатені кері таралуы (backpropagation) немесе кері дифференциалдау деп аталатын алгоритмге әкеледі [95]. кері таралу ақырғы жоғалту мәнінен басталады және кері бағытта, жоғарғы қабаттардан төменгі қабаттарға жүреді. Тізбектік ереже қолданылып, әр параметрдің жоғалту мәніне үлесі есептеледі.

Сонымен оқыту – берілген оқыту мәліметтер жиынтығы және сәйкес нәтижелер үшін жоғалту функциясын минимизациялайтын, модель параметрлерінің комбинациясын іздеу.

Оқыту кездейсоқ мәліметтер үлгісі және нәтижелер пакетін алып, пакеттегі жоғалтуды ескере келе, желі параметрлерінің градиентін есептеу. Содан соң желі параметрлері градиент бағытына қарама-қарсы бағытта аз шамаға (ығысу оқыту жылдамдығымен анықталады) ығысады.

Нейрондық желінің дифференциалданатын тензорлар операциялар тізбегі болуы, оқу процесінің орындалуына негізгі себеп. Тиісінше, ағымдағы параметрлер мен мәліметтер пакетін градиент мәнінде көрсететін градиент функциясын алу үшін тізбектік ережені қолдануға мүмкіндік береді.

Жоғалту функциясы мен оптимизаторлар мәліметтерді желіге бергенге дейін анықталуы тиіс.

Жоғалту функциясы – оқыту барысында минимумға келтірілуге тиісті шама. Сондықтан шешілетін тапсырма үшін сәттілік өлшемі болып табылады.

Оптимизатор параметрлер өзгеруі үшін жоғалту градиентін қолдану әдісін нақты анықтайды. Ол импульстік градиентті түсуді орындайтын RMSProp немесе т.б. болуы мүмкін [96].

2.4 Терең оқытудың құрылымдық блоктары

Түрлі қабаттарға тензорлардың әр түрлі форматтары мен мәліметтерді өңдеу түрлері сәйкес келеді. мысалы, формасы (үлгілер, белгілер) болатын екі өлшемді тензорларда сақталатын қарапайым векторлық мәліметтер тығыз байланысқан, толықбайланысты немесе тығыз қабаттармен өңделеді. Мәліметтер қатары формасы (үлгілер, уақыт таңбалары, белгілер) болатын, үш өлшемді тензорлармен сақталып, LSTM сияқты рекурренттік қабаттармен өңделеді. Суреттер төртөлшемді тензорларда сақталы, екі өлшемді коволюциялық қабаттармен өңделеді. Сонымен қатар әрбір қабат тензорлардың тек белгілі бір түрлерін ғана қабылдап және қайтара алады. Мұны қабаттардың сәйкестігі деп атайды.

Терең оқыту моделі бағытталған, циклдық емес қабаттар гарфі болып табылады. Әдетте, практикада бір ғана кіріс пен бір ғана шығыстан тұратын сызықтық стектер қабаты қолданылады. Алайда өзге де желі топологиялары бар кездесуі мүмкін:

- екі тармақты желі;
- бірнеше басы (multihead) бар желі;
- кіріс блоктар.

Желі топологиясы гипотеза кеңістігін анықтайды. Желі топологиясы таңдала отырып, кіріс және шығыс мәліметтерді көрсететін тензор операцияларының тізбегінің мүмкіндік кеңістігі шектеледі. Бұл тензорлар операциясын қолданылатын тензорлық салмақтардың жақсы мәндер жиынтығын табу кезекті тапсырма болып табылады. Желінің архитектурасын дұрыс таңдау өте маңызды.

Бірнеше шығыстары бар нейрондық желі бірнеше жоғалту функциялары болуы мүмкін. Алайда градиентті түсу процесі бір ғана скалярлы жоғалту мәніне негізделуі тиіс, яғни бірнеше жоғалту функциясы бар желідегі барлық жоғалтулар бір скалярлы мәнге біріктіріледі (орташа мән).

Нақты тапсырамны шешуде мақсаттық функцияны дұрыс таңдау өте маңызды рөл атқарады. Желі жоғалтуды минимизациялаудың түрлі мүмкіндіктерін қолданады. Егер мақсаттық функция тапсырманы сәтті шешуде толықтай түзете алмайтын болса, қажетті нәтижені ала алмау қаупі бар. Дегенмен классификация, регрессия, тізбектерді болжау сияқты жалпы тапсырмалар үшін жоғалту функциясын таңдау бойынша қарапайым нұсқаулар ұсынылады. Мысалы, екі категориялы классификация үшін бинарлы қиылыс энтропиясы, бірнеше категориялы классификация үшін көпмәнді қиылыс энтропиясы, тізбектерді оқыту үшін ассоциативті уақыттық классификация және т.б.

2.5 Машиналық оқыту негізі және машиналық оқыту алгоритмдерінің категориялары

Машиналық оқыту бөлімдерге күрделі бөлінуі бар кең облыс. Машиналық оқыту алгоритмдері негізгі төрт категорияға бөлінеді.

1. Бақыланатын оқыту ең кең таралған машиналық оқыту бөлімі. Оның мағынасы мәліметтерді үлгілер жиынтығында белгілі нәтижелер (аңдатпа) түрінде көрсетуге модельді оқыту [89, р. 94]. Бұл категорияға үлгілерді тану, дыбыстық мәтіндерді тану, суреттерді классификациялау және бір тілден екінші тілге аударма жасау сияқты терең оқытудың барлық замануи әдістері жатады. Бақыланатын оқыту тапсырмалараның басым көпшілігі классификация мен регрессияға тиесілі, сонымен қатар төмендегі тапсырмалар да жатқызылады:

– тізбектерді генерациялау – берілген сурет сипаттайтын тақырыпты болжау;

– синтаксис ағашын болжау – берілген сөйлемнің синтаксистер ағашында бөлінуін болжау;

– объектілерді тану – берілген суреттегі белгілі объектілер айналасында шекараларды салу;

– суретті сегменттеу – берілген суреттегі нақты объекті үшін пиксельдік маска тұрғызу.

2. Бақыланбайтын оқыту – мәліметтерді визуализациялау, сығу, шуылдардан тазалау және мәліметтер арасындағы өзара байланысты жақсы түсіну қажеттіліктерін ешбір нәтижелік мәндердің көмегінсіз кіріс мәліметтерді түрлендіруге негізделген машиналық оқыту бөлімі [89, р. 94]. Ол бақыланатын оқыту әдістерін қолданар алдында мәліметтерді зерттеудегі қажетті қадам және мәліметтер сараптамасының негізі. Кең таралған мысалдар ретінде өлшемді төмендету және кластеризацияны атауға болады.

3. Өздігінен бақыланатын оқыту адам белгілеген таңбаларды қолданбайды, таңбалар эвристикалық алгоритмдерді қолдану арқылы кіріс мәліметтерден генерацияланады. Алдыңғы кадрлер бойынша келесі кадрдағы бейнені болжау немесе мәтіндегі келесі сөзді алдыңғы сөздер бойынша болжау мысалдарын келтіруге болады.

4. Күшейтілген оқытуда агент қоршаған ортадан ақпаратты алып, пайданы арттыру үшін әрекеттерді таңдауға үйренеді [89, р. 95]. Қазіргі таңда бұл сала әлі зерттелуде, нақты қолдану мысалдары жоқ. Алайда автопилоттық автомобильдерде, робототехника, ресурстарды басқару, білім беру сияқты т.б. салаларда қолданылады деп күтілуде.

2.6 Машиналық оқыту модельдерін бағалау

Жоғарыда атап өткеніміздей, оқыту жүргізілген мәліметтер негізінде модельдің сапасын анықтауға болмайды. Бірнеше эпоха өткеннен соң артық оқыту эффектісі орын алады. Басқалай айтқанда, оқытуға қатыспаған мәліметтерге болжам жасау сапасы артуын тоқтатады, тіпті төмендеуі де мүмкін. Машиналық оқытудың мақсаты оқытуға қатыспаған мәліметтерге сапалы болжам жасайтын жалпылама модельді жасауға негізделеді. Артық оқыту осы мақсатқа басты кедергі болып табылады. Бақылау арқылы ғана

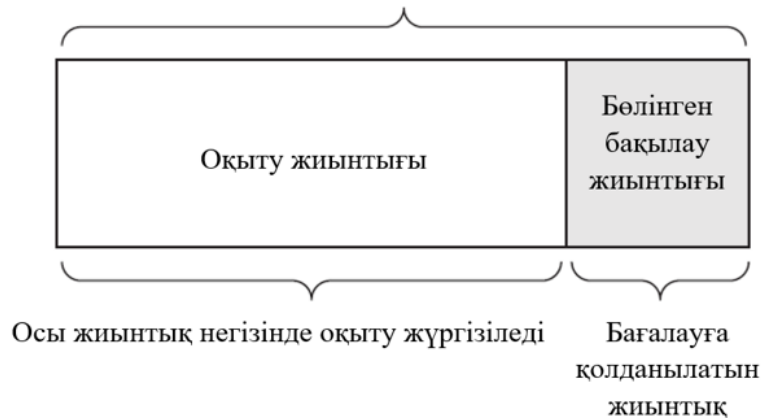
оқыту процесін басқара алатындықтан, модельдің сапасын сенімді бағалау өте маңызды.

Модельді бағалау әрдайым қолжетімді мәліметтерді үш жиынтыққа бөлу арқылы жүргізіледі: оқыту, тексеру және бақылау. Модель оқыту мәліметтері арқылы оқытылады, тексеру арқылы бағаланады. Модельдің соңғы нұсқасы жасалғаннан соң бақылау мәліметтерін пайдалану арқылы тестіленеді. Мәліметтерді үш жиынтыққа бөлу себебі келесідей: модельді құрастыру оның параметрлерін баптаумен үнемі байланысты. Мысалы, қабаттардың саны таңдау немесе өлшемін өзгерту (салмақтық коэффициенттер – параметрлерінен ерекшелендіру үшін, мұндай баптауларды гиперпараметрлер деп атайды) [97]. Баптау үшін кері байланыс сигналы ретінде тексеру мәліметтері қолданылады. Іс жүзінде баптаудың өзі оқу процесі бір түрі болып табылады, белгілі бір параметрлер кеңістігінде неғұрлым тиімді конфигурацияны іздеу. Тексеру мәліметтері негізінде болжау сапасы бойынша модель конфигурациясын баптау, бұл мәліметтерде тікелей оқытылмаса да, лезде артық оқытуға әкелуі мүмкін. Басты себеп ақпараттың шығуы. Тексеру мәліметтері бойынша болжам сапасына сүйене отырып модель гиперпараметрлерін баптау барысында сол мәліметтердің кейбіреуінің шығуына жол беріледі. Егер бұл бір параметр үшін бір рет қана орын алса, шығатын ақпарат көлемі аз болып, тексеру мәліметтері жиынтығы модель сапасының сенімді өлшемі бола алады. Алайда, егер тексеру мәліметтері бойынша модельді бағалап, соның негізінде түзету жүргізу эксперименті барысында баптау бірнеше рет қайталанса, модель үлкен көлемдегі ақпаратты шығарып алады. Нәтижесінде тексеру мәліметтері бойынша болжам сапасының жоғарғы нәтижесіне жасанды түрде бапталған модель алынады. Себебі модель осы мәліметтер негізінде бапталған. Алайда шынайы мақсат жаңа мәліметтерді болжаудағы сапасының жоғары болуы. Сондықтан модель сапасын бағалау үшін оқытуға қатыспаған жеке мәліметтер жиыны – бақылау жиынтығын қолданған дұрыс. Модельге бақылау жиынтығындағы ақпарат жанама түрде де қолжетімсіз болуы тиіс.

Кіріс мәліметтер көлемі шектелген жағдайда қажет болуы мүмкін бірнеше шешімдер бар. Бағалаудың үш классикалық мәзірі келесідей: іріктемені қарапайым бөлуі бар тексеру, K блоктар бойынша қиылысу арқылы тексеру және K блоктар бойынша итерациялық тексеру.

Іріктемені қарапайым бөлуі бар тексеру кезінде мәліметтердің бір бөлігі бақылау жиынтығына бөлінеді. Оқыту қалған мәліметтер арқылы жүргізіледі, сапа бақылау мәліметтері арқылы бағаланады. Ақпарат шығуының алдын алу үшін модель бақылау мәліметтерін болжау нәтижесі бойынша бапталмауы керек. Сол себептен жеке тексеру жиынтығы дайындалады (2.11-сурет).

Классификацияланған мәліметтердің толық қолжетімді көлемі



Сурет 2.11 – Іріктемені қарапайым бөлуі бар тексеру кезіндегі мәліметтерді бөлу

Бұл ең қарапайым бағалау хаттамасы болғанымен мәліметтер көлемі жеткіліксіз болғанда статистикалық репрезентативті болу үшін тексеру және бақылау мәліметтері өте аз үлгілерден тұруы мүмкін. Егер бөлу алдында мәліметтерді кездейсоқ орналастыру модель сапасының бағасы бір бірінене алшақ болса онда осыпроблема туындады дегенді білдіреді. Оны жою үшін келесі екі шешім ұсынылады.

K блоктар бойынша қиылысу арқылы тексеру шешімінде мәліметтер бірдей өлшемдегі K блоктарға бөлінеді. Әрбір i блок үшін қалған $K-1$ блоктармен оқыту, ал бағалау i блогында орындалады. Ақырғы бағалау K аралық бағалаулардың орташасы ретінде есептеледі. Бұл әдіс модель сапасы мәліметтерді оқыту/бақылау жинақтарына бөлуге өте тәуелді жағдай қолдану тиімді.

K блоктар бойынша итерациялық тексеру шағын мәліметтер жиынтығы бар және модельді максималды нақты бағалау жағдайында тиімді. Әр K блоктарға бөлу алдында мәліметтерді араластыра отырып K блоктары бойынша қиылысу арқылы тексеруді бірнеше рет пайдалануға негізделеді. Ақырғы бағалау алынған бағалаулардың орташасын табу арқылы анықталады.

2.7 Мәліметтерді өңдеу, белгілерді құру және белгілерді оқыту

Мәліметтерді алдын-ала өңдеудегі мақсат – кіріс мәліметтерді нейрондық желіге беруге жарамды ету. Мәліметтерді алдын-ала өңдеу векторизация, нормализация, жетіспейтін мәндерді өңдеу және белгілерді алудан тұрады.

Нейрондық желінің барлық кірістері мен шығыстары жылжымалы нүктелі сандар тензоры болуы тиіс (ерекше жағдайларда нақты сандар тензоры). Қандай бір мәліметтер болмасын, дыбыс, сурет, мәтін барлығын тензорға түрлендіру керек. Бұл қадам векторизация деп аталады.

Мәліметтерді желіге жіберер алдында, орташа мәндері 0-ге тең, ал стандартты ақытқуы 1-ге тең болатындай, әрбір белгіні жеке нормализациялау қажет болады. Жалпы жағдайда нейрондық желіге өте үлкен мәндерді

қабылдайтын немесе түрлі мәліметтерді беру қауіпті. Бұл желі үйлесімділігіне кедергі жасайтындай градиенттің айтарлықтай өзгерісіне алып келеді. желіні оқытуды жеңілдету үшін мәліметтер келесідей сипаттамаларға ие болуы керек:

– аз мәндерді қабылдау – ережеге сәйкес мәндер 0-1 диапазонында болуы керек;

– біртекті болу – барлық белгілер бір диапазондағы мәндерді қабылдауы тиіс.

Кейде кіріс мәліметтерде кейбір мәндер болмауы мүмкін, яғни белгі барлық үлгідек кездеспеуі мүмкін. Нейрондық желі жағдайында 0-дің мағыналық мәні болмаса, жетіспейтін мәндер орнын 0-мен ауыстыруға шектеу жоқ. Желі мәліметтерді өңдеу барысында 0- мәліметтің жоқ екендігін түсініп, ол мәнге назар аудармайды.

Белгілерді құру – модельге жібермес бұрын, алгоритм эффективтілігін арттыру үшін белгіленген түрлендірулерді қолданып, машиналық оқыту алгоритмдері мен мәліметтер туралы білімдерді пайдалануға негізделеді. Мәліметтер модельге оның жұмысын жеңілдететіндей түрде берілуі тиіс:

– жақсы белгілер тапсырманы ұтымды әрі аз ресурс шығынымен шешуге мүмкіндік береді;

– жақсы белгілер тапсырманы әлдеқайда аз көлемдегі кіріс мәліметтер арқылы шешуге мүмкіндік береді. Терең оқыту моделінің белгілерді анықтау қабілеті мәліметтердің көлеміне тәуелді болады. Егер үлгілер саны аз болса, белгілердің ақпараттық құндылығы арта түседі.

2.8 Артық оқыту, жеткіліксіз оқыту және толықтырып оқыту

Артық оқыту машиналық оқытудың барлық тапсырмаларында кездеседі [89, p. 94-95; 98]. Эффектпен күресе алу машиналық оқытуда маңызды рөл атқарады. Машиналық оқытудағы негізгі проблема оптимизация мен сәйкестік арасындағы қарама-қайшылық. Оптимизация оқыту мәліметтерінде максималды сапаны алу үшін модельді баптау процесі, ал сәйкестік – бұрын кездеспеген мәліметтегі оқытылған модельдің сапасы. Негізгі мақсат сәйкестіктің жоғары деңгейі болғанымен, оған тек оқыту мәліметтеріне сүйене отырып, модельді баптау арқылы ғана қол жеткізуге болады. Оқытудың бастап кезеңінде оптимизация мен сәйкестік түзетіледі: оқыту мәліметтеріндегі жоғалту аз болған сайын, бақылау мәліметтеріндегі жоғалту соғұрлым аз болады. Бұл жағдай орындалып тұрса, модель әлі жеткілікті оқытылмаған, прогресс әлі мүмкін емес, желі оқыту мәліметтеріндегі релевантты шаблондардың барлығын модельдеп болмады. Алайда оқыту мәліметтеріндегі бірнеше итерациядан соң сәйкестік артуын тоқтатып, тексеру метрикалары өсуін тоқтатады, содан соң төмендейді, яғни артық оқыту эффектісі орын алады. Басқаша айтқанда, модель оқыту мәліметтеріне тән шаблондарға оқытыла бастайды, алайда ол жаңа мәліметтерге тән болмайды. Оқыту мәліметтерінде орын алатын спецификалық немесе релевантты емес шаблондарға модельді оқытудың алдын алудың тиімді әдісі – оқыту мәліметтерінің көлемін арттыру. Үлкен көлемде оқытылған үлкен сәйкестікке

ие болады. Егер мүмкін болса, модельге сақтауға мүмкін болатын, ақпаратқа шектеу қосу және ақпараттың сапасын реттеу тағы бір тиімді әдіс. Егер желі шаблондардың аз санын ғана сақтай алса, оптимизация процесі оны неғұрлым мағыналы шаблондарға бағыттап, сәйкестіктің жоғары деңгейіне қол жеткізуге мүмкіндік береді. Осындай жолмен артық оқытумен күресі регуляризация деп аталады. Желі өлшемін төмендету, салмақтар регуляризациясын қосу және кесу.

Артық оқытуға жол бермеудің ең қарапайым әдісі – модель өлшемін, оқытылатын параметрлердің санын азайту (қабаттар мен әр қабаттағы нейрондар санымен анықталады) [71, р. 98-100; 85, р. 97; 93, р. 102]. Терең оқытуда модельдегі зерттелетін параметрлер санын модель сыйымдылығы деп атайды. Модель неғұрлым үлкен параметрлер көлемінен тұрса, жады көлемі де соншалық үлкен болатындығы анық және оқыту үлгілерін тура көрсету де оңай болады. Келесі тұрғыда, егер желінің сақтауға үшін ресурсы шектелген көлемде болса, ол тура көрсетуді ала алмайды. Өте үлкен мен жеткіліксіз көлемнің арасындағы теңдікті табу маңызды. Қабаттардың саны мен әр қабат өлшемінің дұрыс өлшемін есептеуге ешқандай формула ұсынылмайды. Мәліметтер үшін модельдің дұрыс өлшемін анықтау үшін түрлі архитектура массивтерін бағалауға (тексеру жиынтығына сүйеніп) тура келеді. Жалпы жағдайда модельдің тиісті өлшемін іздеу процесі қабаттар мен параметрлердің аздаған көлемінен бастап, тексеру мәліметтерінде жоғалту артқанға дейін бірте-бірте үлкейту керек.

Бір оқыту мәліметтері жиынтығы және желі архитектурасы үшін мәліметтерді түсіндіретін көптеген салмақ мәндері (модельдің) жиынтығы болады. Қарапайым модельдерде күрделі модельдерге қараған артық оқытуға бейімділігі аз болып келеді. Қарапайым модель – параметрлер мәнінің таралуы аз энтропияға ие модель (немесе параметрлер саны аз). Артық оқыту проблемасын жұмсартудың әдісі – салмақ коэффициенттері мәнін шектеу арқылы желі күрделілігін төмендету. Бұл жағдайда олардың таралуы біркелкі болады. Бұл шешім салмақтарды регуляризациялау деп аталады. Ол салмақ артқаны үшін желінің жоғалту функциясына баға (cost) қосады және регуляризацияның екі түрі бар [89, р. 94]:

1. L1 регуляризация – қосылатын баға салмақ коэффициенттерінің абсолюттік мәніне тура пропорционал.

2. L2 регуляризация – қосылатын баға салмақ коэффициенттері мәнінің квадратына тең. Оны басқаша салмақтардың қысқаруы деп те атайды.

Ескеру керек жағдай, баға тек оқыту сатысында қосылатындықтан, желінің жоғалту шамасы оқыту сатысында, бақылау сатысына қарағанда әлдеқайда жоғары.

Кесу – нейрондық желілер үшін регуляризация әдістерінің кең таралған әрі тиімді әдістерінің бірі. Қабаттарға қолданылатын кесу әдісі, оқыту сатысында кездейсоқ таңдалған белгілерді жоюға (нөл еншіленеді) негізделген. Кесу коэффициенті – бұл нөлге айналатын белгілер үлесі, әдетте 0,2-ден 0,5-ке дейінгі диапазонда таңдалады. Тестілеу сатысында кесу әдісі орындалмайды.

Оның орнына деңгейдің шығыс мәндері кесу коэффициентіне тең коэффициентке азайтылады. Бұл тестілеу және оқыту сатыларында белгілер активтілігінің айырмашылығын толықтырады.

Толықтырып оқыту модельді қайта қолданудың кең таралған әдісі. Белгілерді айқындау үшін қолданылған бірнеше жоғарғы тоқтатылған қабаттарды қайта қолданып, қосылған модель бөлігімен оқытуды жалғастыруға негізделеді. Тоқтату - бір немесе бірнеше қабаттарды оқыту процессінде олардың салмақтық коэффициенттерінің өзгеруін болдырмауды білдіреді. Желіні толықтырып оқыту келесі қадамдардан тұрады:

1. Оқытылған базалық желінің үстіне өз желімізді қосу.
2. Базалық желіні тоқтату.
3. Қалған бөлікті оқыту.
4. Базалық желінің тоқтатылған бірнеше қабаттарын қайта іске қосу.
5. Осы қабаттарды оқытып, қалған бөлікке қосу.

Толықтырып оқыту кезінде төменде көрсетілген жағдайларды ескеру керек:

1. Қабаттық негіздегі бастапқы қабаттар қайта қолдануға жарамды неғұрлым жалпылама белгілерді, ал жоғарғы қабаттар неғұрлым нақты белгілерді кодалайды. Жаңа тапсырмаларды шешуде жиі қайта профильденетіндіктен нақты белгілерді баптай неғұрлым тиімді болып табылады.

2. Оқылатын параметрлер көп болған сайын, артық оқыту орын алуы да ықтималдығы да соғұрлым жоғары.

2.9 CNN – конволюциялық нейрондық желілер

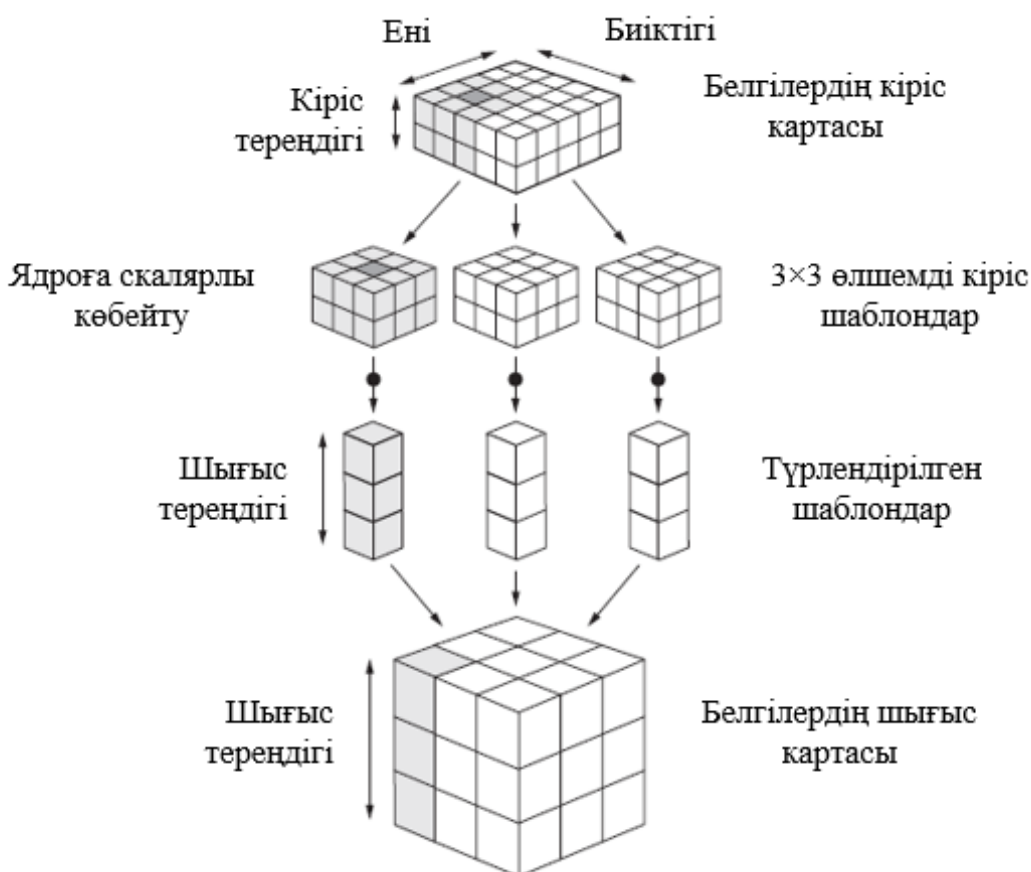
Толық байланысқан қабаттардың конволюциялық қабаттапдан ерекшелігі, Dense қабаттары кіріс белгілер кеңістігінде глобалды шаблондарды, ал конволюциялық қабаттар локалды шаблондарды зерттейді. Бұл негізгі сипаттама конволюциялық нейрондық желілердің (CNN) екі негізгі қасиетіне себеп болады [89, p. 119]:

1. Зерттелетін шаблондар тасымалдау тұрғысынан инвариантты болып табылады. CNN белгілі бір шаблонды зерттегеннен соң, оны кез-келген мәлімет бөлігі қатарынан таба алады. Ал толық байланысты желі жағдайында шаблонды қайта зерттеуге тура келер еді. Бұл конволюциялық желілердің суреттерді өңдеу тапсырмаларындағы эффективтілігін арттырады. Мұндай желілерге сәйкестендіру деңгейі жоғары ұсыну жолын алу үшін аз көлемдегі оқыту үлгілері жеткілікті болады.

2. Бұл желілер шаблондардың кеңістіктік иерархияларын зерттей алады. Бірінші конволюциялық қабат үлкен емес локалды шаблондарды (мысалы, шекара), екінші қабат бірінші қабат қайтарған белгілерден тұратын неғұрлым үлкенірек шаблондарды зерттейді. Бұл конволюциялық нейрондық желілерге неғұрлым күрделі және абстрактілі визуалды ұсынуды эффективті зерттеуге мүмкіндік береді.

Қабаттау, белгілер картасы деп аталатын, екі кеңістіктік (ені және биіктігі) және тереңдік осі бар, үш өлшемді тензорларға қолданылады. Қабаттау операциясы өз кірісіндегі белгілер картасынан шаблондарды шығарады. Барлық шаблондарға бірдей түрлендіруді қолдана отырып, шығыс белгілер картасын жасайды. Белгілердің шығыс картасы да үш өлшемді тензор болып табылады. Оның тереңдігі кез келген өлшемде болуы мүмкін, себебі шығыс тереңдік, қабат параметрі болып табылады және фильтрлерге сәйкес болады.

Қабаттау жылжушы терезе әдісі арқылы орындалады (2.12-сурет). Мысалы, ол өлшемі 3×3 болатын белгілердің үш өлшемді картасы бойынша, әр позицияға тоқтала отырып, формасы (биіктігі, ені, тереңдігі) болатын үш өлшемді белгілер шаблонын алады. Әр осындай үш өлшемді шаблон формасы (шығыс тереңдігі) болатын бір өлшемді векторға (тензорды, салмақтар матрицасына (оқыту барысында алынатын қабаттау ядросы) көбейту арқылы) түрлендіріледі. Бұл векторлардың барлығы формасы (биіктігі, ені, шығыс тереңдігі) болатын үш өлшемді шығыс картасына жинақталады. Белгілердің шығыс картасындағы кеңістіктік орналасуы кіріс белгілер картасындағы орналасуына сәйкес келеді.



Сурет 2.12 – Қабаттаудың жұмыс принципі

2.12-суретте көрсетілгендей шығыстағы карта ені мен биіктігі кірістегіден ерекшеленеді. Оған екі түрлі себеп бар:

– шекара эффе́ктісі (белгілердің кіріс картасын толықтыру арқылы алдын алуға болады);

– қабаттау қадамын қолдану.

CNN зерттеген ұсынуларды визуализациялау мүмкіндігі бар. Бұл ұсынуларды түсіндіру және визуализациялау әдістерінің кең таралғандары:

– конволюциялық нейрондық желілердің аралық нәтижелерін визуализациялау (аралық активация) – желі қабаттарының тізбегі кіріс мәліметтерді қалай түрлендіретіндігін түсінуге көмектеседі және жеке фильтрлардың мағынасын көрсетеді;

– конволюциялық нейрондық желі фильтрларын визуализациялау – әр фильтрдің нақты қай ұғымға және визуалды шаблонға жауап беретіндігін білуге көмектеседі;

– 2.12-суреттегі класс активацияларының жылулық карталарын визуализациялау – 2.12-суреттегі объектілерді анықтауға мүмкіндік беретін, белгіленген классқа 2.12-суреттің қай бөліктері тиісті екендігін анықтауға көмектеседі.

Егер мәліметтердің глобалдық реттілігіне маңызды емес болса конволюциялық желілерді қолдану, ресурс шығын аз әрі сәтті шешімге жеткізеді. Бұл әсіресе, сөйлем басында табылған шаблон мен сөйлем соңынан табылған шаблонның құны тең мәтіндік мәліметтерге қатысты.

2.10 RNN – рекуррентті нейрондық желілер

Уақыт тізбектері мен мәліметтер тізбектерін өңдеуге арналған терең оқытудың негізгі іргелі алгоритм – рекуррентті нейрондық желілер (RNN) болып табылады [89, р. 196]. Осыған дейін қарастырылған толық байланысқан және конволюциялық нейрондар жадының болмауымен ерекшеленген болатын. Олардың кірістері арасындағы күйлер еш сақталмай, бір бірінен тәуелсіз өңделеді. Мұндай желілердің көмегімен тізбектерді немесе уақыт қатарларын өңдеу үшін, тізбекті толығымен бәр пакетке орналастырып желіге беру қажет болатын. Мұндай желілерді тура таралуы бар желілер деп атаймыз.

Рекуррентті нейрондық желілер – алдыңғы элементтерді өңдеу кезіндегі күйлерді сақтап, тізбек элементтерін талдай отырып өңдейді. Іс жүзінде рекуррентті нейрондық желілер ішкі циклі бар, нейрондық желінің түрі. RNN екі түрлі, тәуелсіз тізбектер арасындағы өңдеу күйлерін қабылдамайды. Сондықтан бір тізбек мәліметтердің біртұтас болғы ретінде түсіндіріледі. Алайда желі мәліметтер блогын элементтер тізбегін талдай отырып, ішкі циклды орындайды.

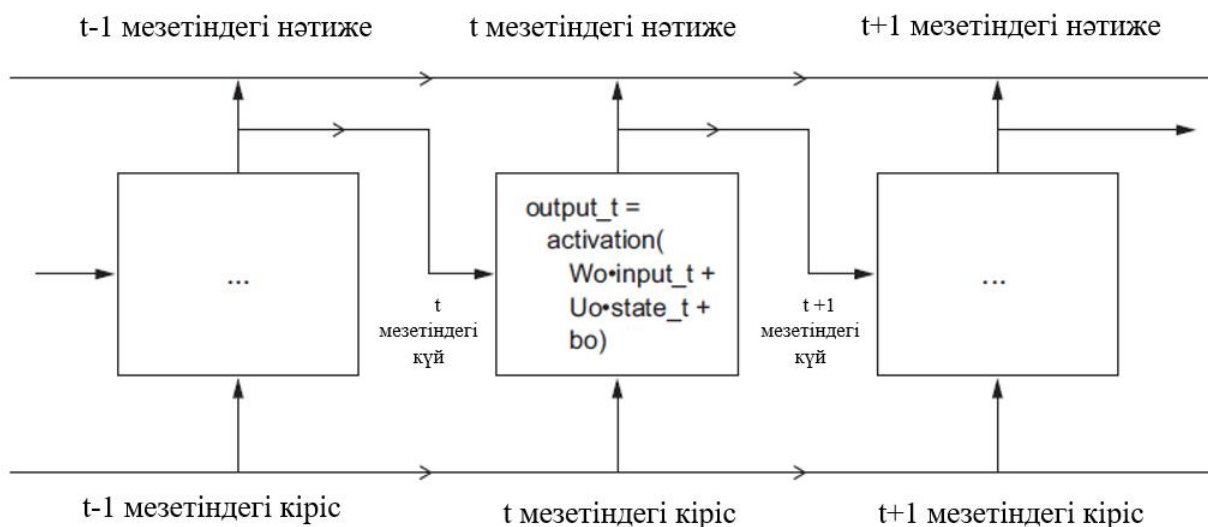
Алдымен кірісте бірнеше өлшемді тензорлар түріндегі векторлар тізбегі қабылданады, уақыт интервалдарын талдау арқылы t уақытындағы күй мен кіріс белгілерді ескере отырып, t уақытына сәйкес шығыс нәтижені құрастырады. Бұл нәтиже келесі итерацияға дайындық ретінде ішкі күйде сақталады. Бірінші уақыттық интервал үшін алдыңғы шығыс нәтиже анықталматындығы түсінікті. Сондықтан алғашқы күй, бастап күй деп аталатын, элементтердің нөлдік мәнімен вектор түрінде инициализацияланады.

Егер тізбектегі мәліметтердің глобалдық реті маңызды болса, мұндай мәліметтерді рекурренттік желілерді қолдану арқылы өңдеу тиімді. Мысалы, алыс өткенге қарағанда жақын өткен маңызды ақпаратқа ие болатын уақыт тізбектері.

2.10.1 LSTM және GRU қабаттары

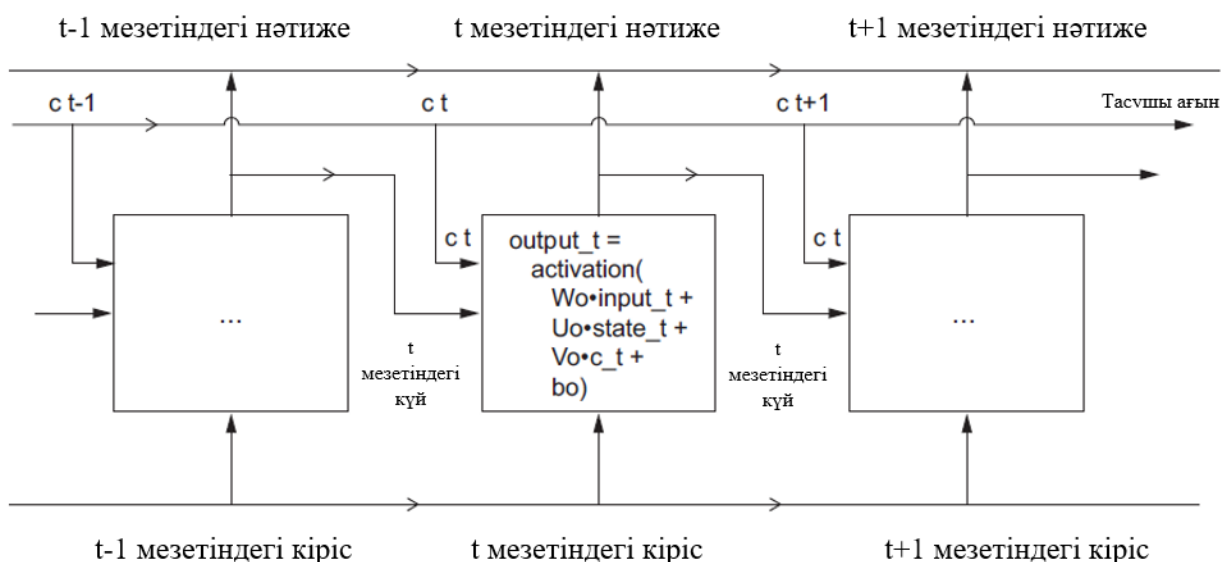
Қарапайым рекурренттік желілердің басты кемшілігі адыңғы интервалдардың әр t уақыт үшін кіріс мәліметтер туралы ақпаратты сақтауы тиіс. Ал мұндай ұзақ тәуелділіктерді оқыту мүмкін емес. Бұл үлкен көлемдегі қабаттардан тұратын рекуррентті емес желілерде кездесетін градиент өшулігіне ұқсас эффекттің есебінен болады. Желі қабаттарының тым көп болуы, желіні оқытуға икемсіз етеді. LSTM және GRU қабаттары осы проблеманы шешу үшін жасалған.

LSTM қабатының жұмысын келесідей түсіндіруге болады: келесі қолдану үшін ақпаратты сақтау арқылы, өңдеу уақытында алдыңғы сигналдар өшуінің алдын алады. 2.13-суретте SimpleRNN қабаты сипатталған [89, p. 202].



Сурет 2.13 – LSTM қабатының бастапқы нүкте: SimpleRNN қабаты

Осы сұлбаға уақыт аралықтары бойынша ақпаратты таситын мәліметтер ағыны қосылады. Оны түрлі уақыт аралықтарында белгілеу үшін S_t белгісін қолданамыз (жылжытылған). Бұл ақпарат кіріс және рекуррентті байланыстармен бірігіп (ығысумен қоса отырып салмақ матрицаларына скаляр көбейту және активация функциясын қолдану арқылы), келес уақыт аралығына берілетін күйге әсер ету (активация функциясы және көбейту операциясы) арқылы ұяшыққа әсер етеді. 2.14-суретте көрсетілгендей, ақпаратты тасу ағыны келесі нәтиже мен күй модуляциясын орындайды.

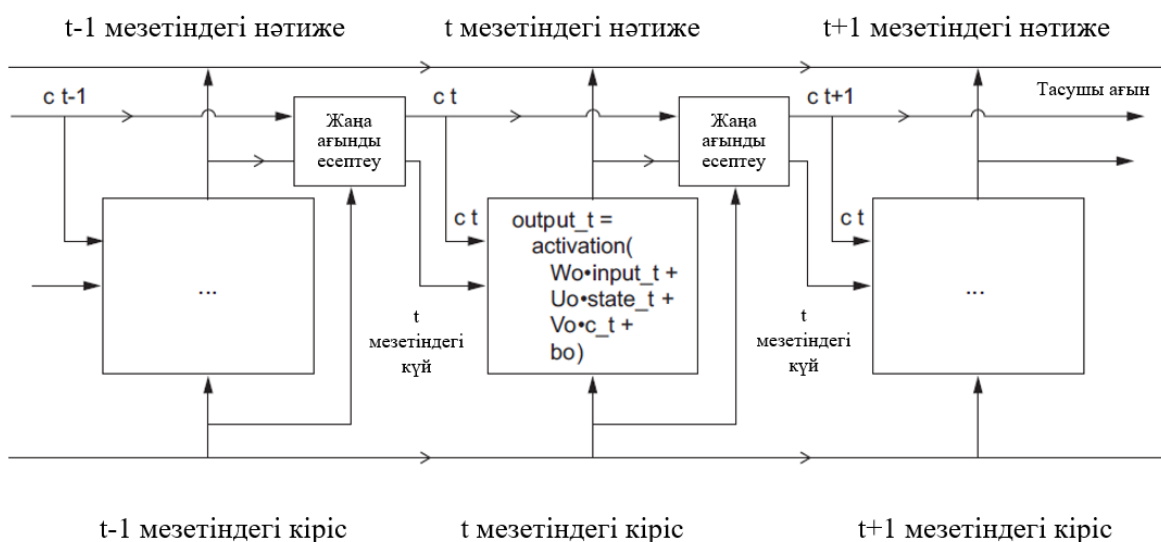


Сурет 2.14 – SimpleRNN-нан LSTM-ге өту: тасушы ағынды қосу

Мәліметтерді тасушы ағындағы келесі мәнді есептеу SimpleRNN ұяшықтары формасындағы үш түрлі түрлендіруге негізделеді. Бұл түрлендірулердің i , f , k индекстерімен белгіленетін өз салмақ матрицалары болады. i_t , f_t , k_t біріктіру арқылы, жаңа тасушы күйді аламыз. LSTM архитектурасының жүзеге асырылуы келесі теңдеуде көрсетілген [89, p. 202; 99]:

$$c_{t+1} = i_t * k_t + c_t * f_t \tag{2.3}$$

c_t мен f_t көбейтіндісі мәліметтер тасушы ағынындағы қажетсіз ақпаратты жою әдісі болса, i_t мен k_t көбейтіндісі тасушы ағынға жаңа ақпаратты қоса отырып, нақты ақпаратты ұсынады (2.15-сурет). Алайда, операция әрекеті олардың салмақтарын параметрлеуші мәні бойынша анықталады, ал салмақтар оқытудың әр циклында үздіксіз жіне қайта есептеледі. Осы себептен бұл және өзге операциялардың нақты мақсатын бекіту мүмкін емес. RNN ұяшықтары, оқыту барысында модельдің тиімді баптауларын іздейтін гипотеза кеңістігін анықтайды. Алайда ұяшықтың жұмысын оның салмақтары анықтайды. Түрлі салмаққа ие бірдей ұяшықтардың жұмысы бір-бірінен өзгеше болады. сондықтан RNN ұяшықтарын жасайтын операциялар жиынтығын, іздеуді орындауда шектеулер жиынтығы деп қарастырған дұрыс болады.



Сурет 2.15 – LSTM анатомиясы

LSTM ұяшығы, алдыңғы ақпараттың оқыту процесіне қайта еніп, градиенттің өшуі проблемасына кедергі жасауға арналған.

Рекурренттік нейрондық желілердің жалпылау деңгейін және сапасын арттырудың жетілдірілген әдістері бар:

- рекурренттік кесу – рекурренттік қабаттардағы артық оқытумен күресуге үшін кесуді қолданудың ерекше енгізілген әдісі. Барлық уақыт аралықтарына бір ғана кесу маскасы қолданылуы керек, аралықтарда өзгермеуі тиіс. LSTM және GRU сияқты рекурренттік қабаттардан алынған ұсынуларды регуляризациялау үшін кесудің уақыттық-тұрақты маскасы қабаттың ішкі рекуррентті активацияларна қолданылуы керек. Уақыттың әр аралығына бір масканы қолдану, желінің уақыт бойынша оқыту қателігін дұрыс таратуына мүмкіндік береді, яғни уақыттық-кездейсоқ маска қате сигналын бұзып, оқыту процесіне залалын тигізеді. Әрбір рекурренттік қабат екі аргументтен, dropout – кесілетін қабаттың кіріс мәнінің үлесін және recurrent_dropout – кесілетін рекурренттік мәндер үлесін анықтайтын нақты сан;

- рекурренттік қабаттарды қабаттастыру – желі ұсынуын арттыру әдісі (есептеу көлемін арттыру арқылы). Артық оқыту эффектісінен арылған соң сапаның төмендігі байқалады. Ол үшін бірінші орынға артық оқыту шыққанға дейін желі сыйымдылығын арттыру қажет (кесу сияқты алдын алудың әдістері қолданылған жағдайда). Желі сыйымдылығын арттыру қабаттар параметрінің санын арттыру немесе қосымша қабаттар қосу арқылы орындалады. Рекурренттік қабаттарды қабаттастыру неғұрлым қуатты рекурренттік желіні алудың классикалық әдісі болып табылады;

- екі бағыты бар рекурренттік қабаттар – рекурренттік желідегі бір ақпаратты, нақтылығын арттырып, ұмытуға байланысты проблемаларды әлсіздендіріп, бірнеше әдістер арқылы ұсынады. Бұл желілер табиғи тілді өңдеу сияқты тапсырмаларда өте жоғары сапаны қамтамасыз етуге қабілетті рекурренттік желілердің кең таралған түрі.

Рекурренттік желілер ретке немесе уақытқа тәуелді болады: олар кіріс тізбектерді рет бойынша өңдейді және мәлімет ретінің кез-келген өзгерісі, рекурренттік желі тізбектен алатын, ұсынуды толықтай өзгертеді. Сол себептен де реттілік маңызды болып табылатын тапсырмаларды тиімді шешеді. Екі бағыты бар рекурренттік желілер RNN-ның реттілікке сезімталдығын пайдаланады. Ол LSTM және GRU қабаттары сияқты екі қарапайым желіден тұрады. Олардың әрқайсысы кіріс тізбектерді бір бағытта өңдеп (тура немесе кері), алынған нәтижелерді біріктіреді. Тізбекті екі бағытта өңдеу отырып, екі бағыты бар рекурренттік желі бірбағытты желіге байқалмайтын шаблондарды анықтауға қабілетті.

2.11 Төтенше жағдайларды анықтау үшін CNN көмегімен аудио мәліметтер тізбектерін өңдеу

Конволюциялық нейрондық желілер параметрлерді қабаттау, локалды кіріс шаблондардан белгілерді алу, мәліметтерді модульдік ұсыну қабілеттері себебінен үлгілерді тану тапсырмаларын шешуде қолайлы. Осы қасиеттер тізбектерді өңдеуде де тиімді етеді. Уақыт кеңістіктік өлшем ретінде қарастырылады. Мұндай бірөлшемді конволюциялық нейрондық желілер аз есептеу ресурстарын талап етсе де, дыбыс генерациясы, машиналық тіл аудару сияқты кейбір тізбекті өңдеу тапсырмаларының оң шешімін алуға мүмкіндік береді. Мәтіндерді классификациялау және уақыт тізбектерін болжау сияқты қарапайым тапсырмаларды шешуде рекурренттік желілердің орнын баса алады. Мұндай бір өлшемді конволюциялық желілер тізбектегі локалды шаблондарды тануға қабілетті. Әр шаблонға бірдей өңдеу жүргізілетіндіктен, белгілі бір позициядан табылған шаблон кейінгі өзге позициялардан да табылуы мүмкін. Осылайша орындалатын өңдеу инвариантты бола алады (уақыт бойынша). Мысалы, мәтіннің әр символын өңдейтін бір өлшемді конволюциялық желі, сөздер морфологиясын зерттеуге қабілетті.

2.12 Төтенше жағдайларды анықтау үшін өңдеу үшін конволюциялық және рекурренттік желілерді біріктіру

Бір өлшемді конволюциялық желі кіріс шаблондарды тәуелсіз өңдейтіндіктен, рекурренттік желілерге қарағанда, уақыт аралықтарының реттілігіне сезімталдығы жоқ. Әрине, ұзақ шаблондарды тану үшін бірнеше конволюциялық желілерді және көрші мәндерді таңдау қабаттарын қабаттастыру әдісін қолдануға болады. Нәтижесінде жоғарғы қабаттар кіріс мәліметтердің ұзын фрагменттерін көре алады, бірақ бұл реттілікке сезімталдылықты қосудың әлсіз әдісі болар еді.

Конволюциялық желілердің жылдамдығы мен жеңілдігін, рекурренттік желінің реттілікке сезімталдығын біріктіру стратегияларының бірі – рекурренттік желіге бермес бұрын бір өлшемді желіні мәліметтерді алдын-ала өңдеу үшін пайдалану (2.16-сурет). Бұл шешім тізбектер рекурренттік желі көмегімен өңдеу мүмкін емес ұзындыққа (бірнеше мың аралық және одан да көп) ие болғанда қолдану ерекше тиімді болып табылады. Конволюциялық

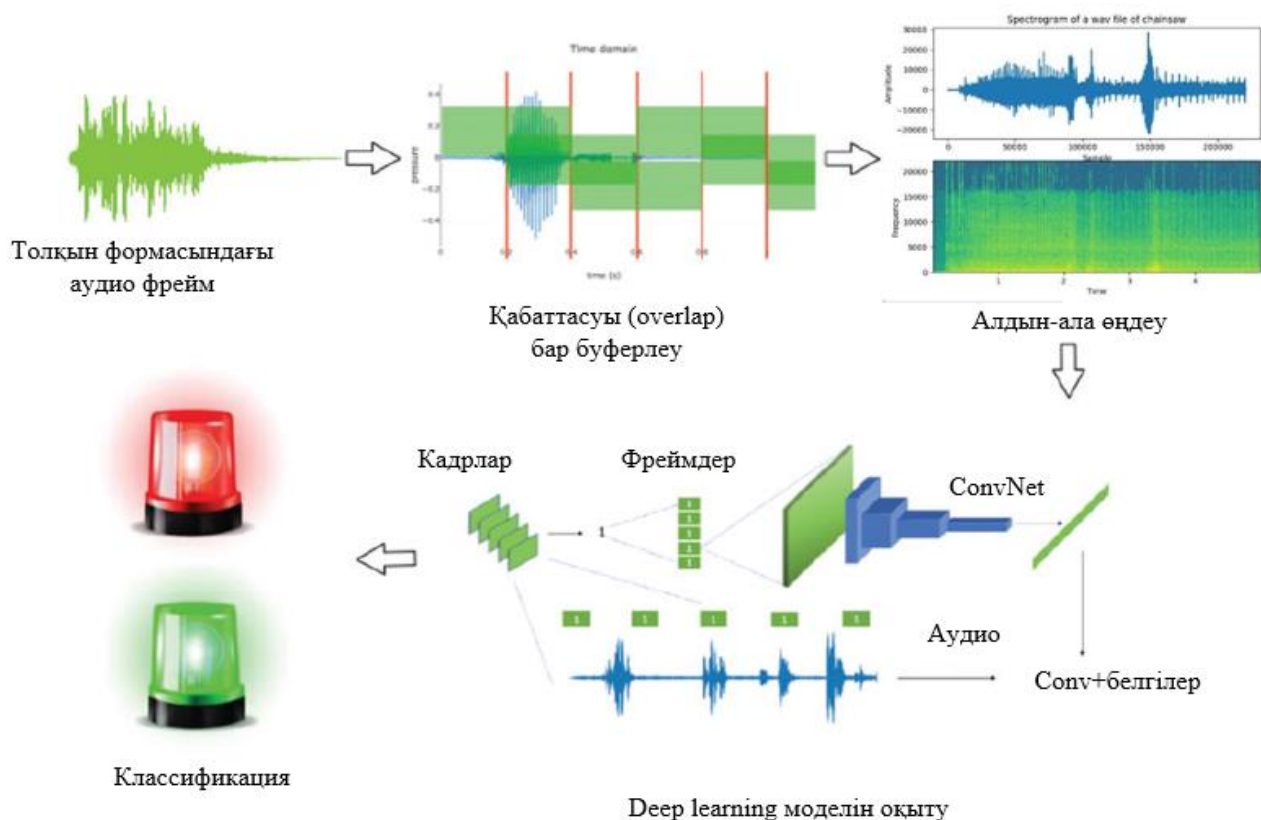
бөлік кірістегі қзын тізбекті неғұрлым қысқа жоғарыдеңгейлі белгілер тізбегіне (рұқсатын азайту арқылы) түрлендіреді. Содан соң, бөлініп алынған тізбектер желінің рекурренттік бөлігіне беріледі. Аталған әдіс ғылыми мақалалар мен қолданбаларда сирек кездесетіндігіне қарамастан эффективтілігі жоғары. Бұл стратегия неғұрлым ұзын тізбектерді басқаруға мүмкіндік береді. Осылайша неғұрлым алыс мәліметтерді қолдана аламыз (мәліметтер генераторының lookback параметрін арттыру арқылы) немесе уақыт тізбектерінің рұқсатын арттырамыз (генератордың step параметрін азайту арқылы) [89, p. 228].



Сурет 2.16 – Ұзын тізбектерді өңдеу үшін бір өлшемді конволюциялық және рекурренттік желілерді біріктіру

3 АУДИОСИГНАЛДАРДЫ ӨҢДЕУ АРҚЫЛЫ ТӨТЕНШЕ ОҚИҒАЛАРДЫ НАҚТЫ УАҚЫТ РЕЖИМІНДЕ АНЫҚТАУҒА ЖӘНЕ КЛАССИФИКАЦИЯЛАУҒА АРНАЛҒАН МОДЕЛЬДІ ҚҰРУ

Бұл бөлімде зерттеу жұмыстарында қолданылған материалдар мен әдістер, қауіпті деп саналатын қала дыбыстарын идентификациялау падаланылған мәліметтер жиынтығы, мәліметтерді жинақтау сатылары, қаладағы төтенше оқиғалар дыбысын анықтауға арналған терең оқыту моделін жасау және алдын-ала өңдеу сатылары сипатталады. 3.1-суретте нақты уақыт режимінде жұмыс істейтін қаладағы төтенше оқиғалардың дыбысын анықтауға арналған ұсынылатын жүйенің архитектурасы көрсетілген. Келесі бөлімдерде зерттеудің материалдары мен әдістері саты бойынша түсіндіріледі, мәліметтер жиынтығы ұсынылады, мәліметтерді жинақтау процесі, модельге, импульстік дыбыстарды анықтау проблемасына шолу жасалады және терең CNN-RNN архитектурасы ұсынылады.



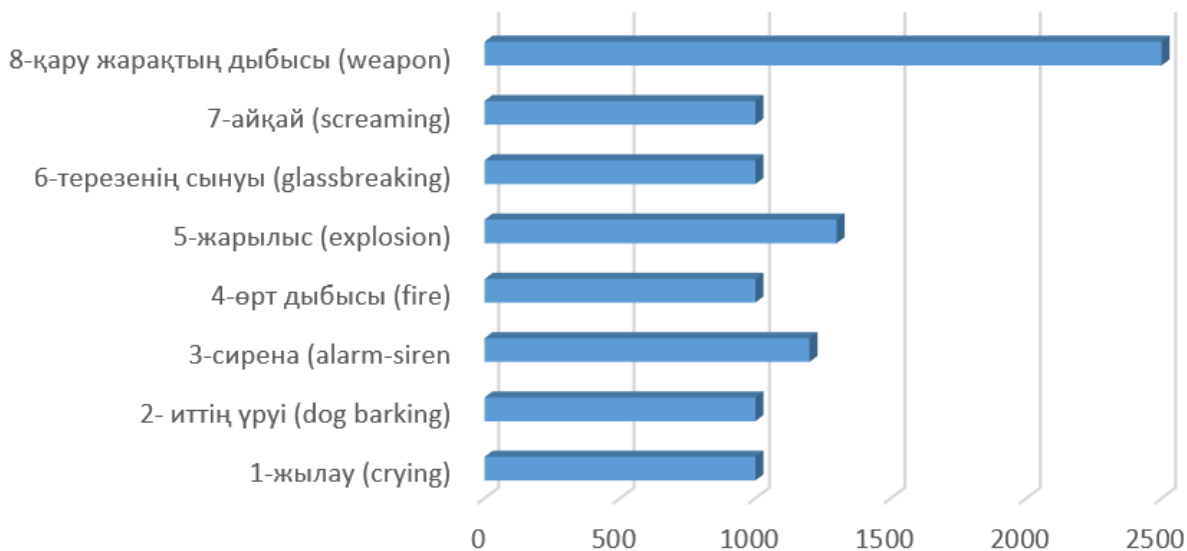
Сурет 3.1 – Ұсынылатын жүйенің архитектурасы

3.1 Төтенше жағдайларды анықтауға арналған модельді оқыту үшін дайындалған аудио мәліметтер жинағы (датасет)

Кез-келген зерттеуді жүргізу үшін үлкен көлемдегі ақпарат керек болатындықтан, зертеудің бірінші сатысы мәліметтерден жинақтаудан тұрады.

Мәліметтер жиынтығы 8 класс бойынша аудиоформаттағы (.mp3, .aiff, .flac, .wav, .m4a) 10 000 түрлі қаладағы қауіпті дыбыстар жинақталды (3.2-сурет). Ұсынылған жиынтық қаладағы қауіпті дыбыстарды анықтау және

классификациялау үшін машиналық оқыту және терең оқыту модельдерін оқытуға және тестілеуге қолданылды. Мәліметтер жиынтығы қару жарақтың атылуы және терезенің сынуы сияқты фондық дыбыстардан тұрды. Қолданбаның ерекшелігіне байланысты дыбыстар ғимарат ішінен және сыртынан алынды. Зерттеу жұмысында сигналдар 1 с интервалдарға, әр интервал 200 мс болатын, қабаттасуы 50% кадрлерге бөлінді. Жекелей алғанда, әр интервал 9 кадрдан тұрды.



Сурет 3.2 – Жинақталған датасеттің көлемі

3.1-кестеде мәліметтер жинағын құраған аудиосигнал үлгілерінің интервалдары мен кадрлары, сигналдар туралы қысқаша ақпарат пен визуализациясы ұсынылған.

3.1-кесте қауіпті дыбыстардың сипаттамасын және автокөлік трезесінің сынуы, иттің үруі, полицей сиренасы, жедел жәрдем сиренасы, қарудың бір рет атылуы, жарылыс, баланың жылауы, күзет сигнализациясы, өрт дабылы және түгін сигнализациясы сияқты үлгілердің спектрограммасы келтірілген. Кестеден ұсынылатын мәліметтер жиынтығының тапсырманы шешуде маңызды екендігін байқаймыз.



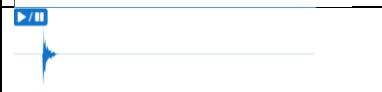
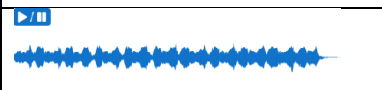




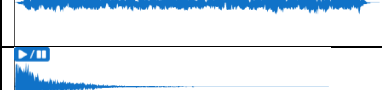

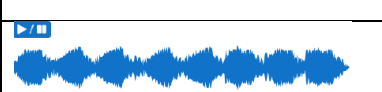
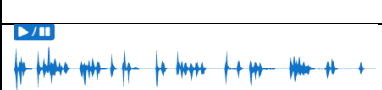
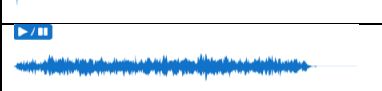

Алайда жинақталған датасеттегі мәліметтер түрлі форматтарда (.mp3, .aiff, .flac, .wav, .m4a) әрі 30 ГБ көлемді қамтитындықтан нормализацияны қажет етеді.

Модельді бұған дейін тексерілген мәліметтер жиынтығымен «қауіпті» деп аталатын дыбыстарды сараптау үшін, ашық қолданыстағы, ірі масштабтағы мәліметтер жиынтығы қолданылды. Модельді тестілеу үшін Urbansounds-8K датасеті жетілдіріліп қолданылды және ESC-10 деп аталатын мәліметтер жиынтығы таңдалды. Датасет әрқайсысы шамамен .wav форматындағы 900 файлдан тұратын 10 классқа бөлінген. Бұл мәліметтер жиынтығындағы дыбыстар тізімі жануарлардың дыбысы, табиғаттың дыбысы, адамдардың

дыбысы, тұрмыстық және күнделікті дыбыстар және қауіпті дыбыстар сияқты категориялардан тұрады:

- жануарлар дыбысы (ит, мысық, сиыр және т.б.);
- табиғат дыбыстары (жаңбыр, теңіз, құстар және найзағай);
- адамның дыбыстары (баланың жылағаны, адымдар, жөтел, дем);
- тұрмыстық және күнделікті дыбыстар (есікті қағу, клавиатурада мәтінді теру, оятқыш, сынған терезе);
- қауіпті дыбыстар (полицей сиренасы, поезд, иттің үруі, қозғалтқыш, электрлік ара, ұшақ, от шашу, жарылыс, иттің үруі, қару жарақтың атылуы және т.ю. импульсивті дыбыстар).

Кесте 3.1 – Жинақталған датасеттегі импульстік дыбыстардың үлгілері

Импульстық дыбыс типі	Уақыт, с	Спектрограммасы
Автокөлік терезесі сынуы	3.84	
Артиллериялық снарядтың жарылуы	4	
Бір рет атылған қару дыбысы	3.84	
Жедел-жәрдемнің сиренасы	15.41	
От дабылы айқайы	2.3	
Түтін дабылы	0.99	
Баланың жылаған дауысы	6.66	
Полицияның сиренасы	24.19	
Жарылыс	7.78	
Полиция сиренасы тұрақты дыбысы	56.87	
От туралы дабыл	1.41	
Ит үргендегі дыбыс	22.15	
Күзеттің дабылы	11.13	
От туралы дабыл	1.59	

Ақпараттың көптігіне қарамастан, бастапқы сатыда тек қауіпті дыбыстар ғана қолданылды. Жасалған мәліметтер жиынтығының техникалық параметрлері 3.2-кестеде көрсетілген.

Кесте 3.2 – Модельді тексеру үшін қолданылған ашық қолданыстағы UrbanSounds-8K датасетінің жетілдірілгеннен кейінгі параметрлері

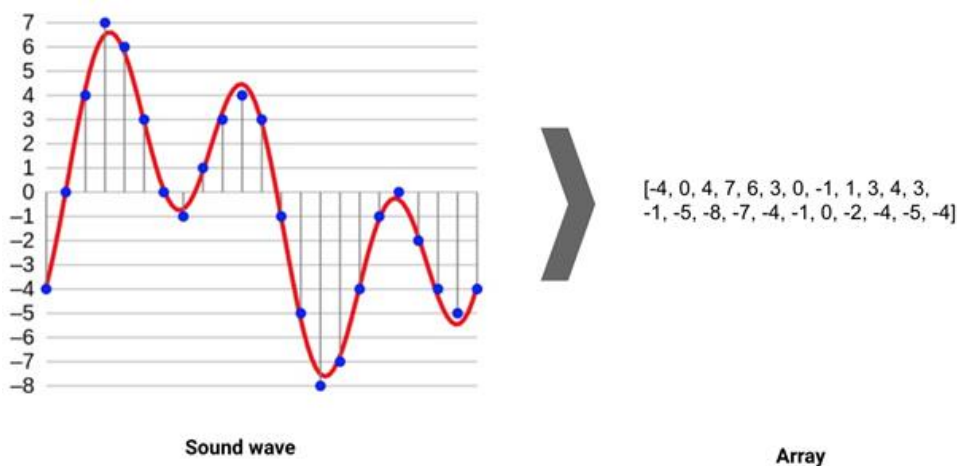
Сипаттамалары	Параметрлері
Жалпы көлемі	661 MB
Алдын-ала өңдеуден кейінгі өлшемі	45 MB
Файлдардың саны	2000
Алдын-ала өңдеуден кейінгі файлдар саны	301
Файлдардың форматы	.wav

Қару-жарақтың атылуы, айқай және терезенің сынуы зерттелетін аймақта «күдікті» оқиғалар қатарына жатқызылды. Осылайша ұсынылған жүйенің өнімділігі автоматты бақылау қолданбасы үшін бағаланды. Осы мақсатта бірнеше аудио үлгілерді қолдана отырып, түрлі шарттарда жазылған мәліметтер жиынтығы жасалды.

Сондықтан бұл мәліметтер жиынтығын негізгі модельді тексеріп болған соң қолданатын боламыз. Бұл жиынтықтағы класстар төтенше оқиғаларды ескерту үшін біздің тапсырмаға сәйкес келеді.

3.2 Төтенше жағдайларды анықтауға арналған модельді оқытуда қолданылатын аудио мәліметтерді алдын-ала өңдеу

Жоғарыда аталған аудиоформаттардың сақтау мен жаңғырту қасиеттері түрлі болғанымен, барлығының тапсырмасы бір - аудиосигналды цифрлық жүйенің ішінде тасымалдайды (3.3-сурет).



Сурет 3.3 – Дыбыс толқынын массивті түрге түрлендіру

Әрі қарай дыбыстық файл кітапханасы көмегімен `sf.read()` функциясын пайдаланып дыбысты `.ogg` форматындағы файл түрінде қарастырамыз (3.4-сурет). Файл, функция арқылы `float32` форматында оқылып, толқын сандық

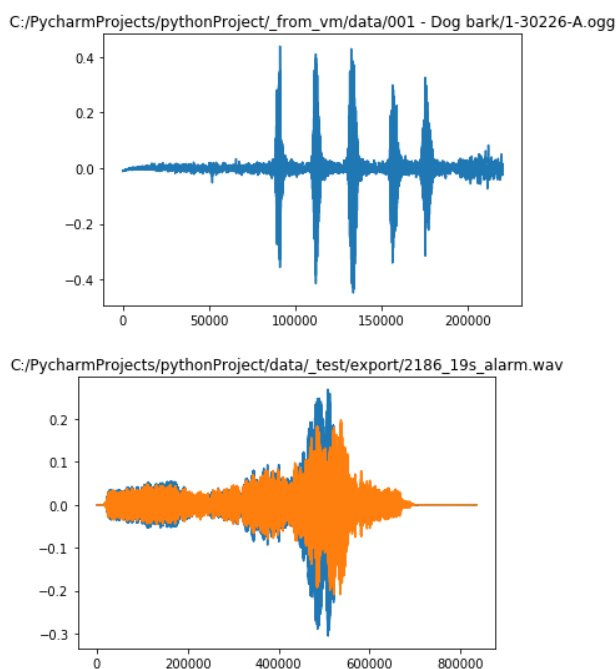
форматқа түрлендірілді. Суретте көргеніміздей екі өлшемді массив алынды, яғни бұл шығыстың екі каналы бар дегенді білдіреді және ол «.shape» қатары арқылы расталып, екі өлшемді массивте 835 200 мән бар екендігі анықталды. Содан соң дискретизация жиілігі анықталады. Басқаша айтқанда, бір секундта жаңғыртылатын мәндердің саны. Әдетте ол 44 100 құрайды.

```
alarm_path = 'C:/PycharmProjects/pythonProject/data/_test/export/2186_19s_alarm.wav'
alarm_path_x, alarm_path_sample_rate = sf.read(alarm_path, dtype='float32')
print(alarm_path_x)
print(alarm_path_x.shape)
print(alarm_path_sample_rate)

[[ 0.0000000e+00  0.0000000e+00]
 [ 0.0000000e+00  0.0000000e+00]
 [ 0.0000000e+00  0.0000000e+00]
 ...
 [ 6.1035156e-05 -9.1552734e-05]
 [ 9.1552734e-05 -9.1552734e-05]
 [ 9.1552734e-05 -6.1035156e-05]]
(835200, 2)
44100
```

Сурет 3.4 – .wav форматтағы сигнализация дыбысы үлгісін .ogg форматына түрлендіру

Аудионың дыбысты тарату каналдары болады: оң және сол шығыс, яғни дыбыс екі жақа бірден тарайды. Бұл моноканалды деп аталады. Мұндай дыбыстарды жазу тек бір ғана құрылғы арқылы жүзеге асырылады. Алайда түрлі типтегі және эффекттегі дыбыстар стерео дыбыс арқылы орындалады. Оның негізгі ерекшелігі қабылданған дыбыс сол арнаға жеке, оң арнаға жеке таратылады. осылайша каналдағы дыбыстар белгілі бір деңгей жоғарылығына ие болады. Иттің үру дыбысы мысалындағы моно және стерео дыбыстарды салыстыру 3.5-суретте көрсетілген.



Сурет 3.5 – Моно және стерео дыбыстарды визуализациялау

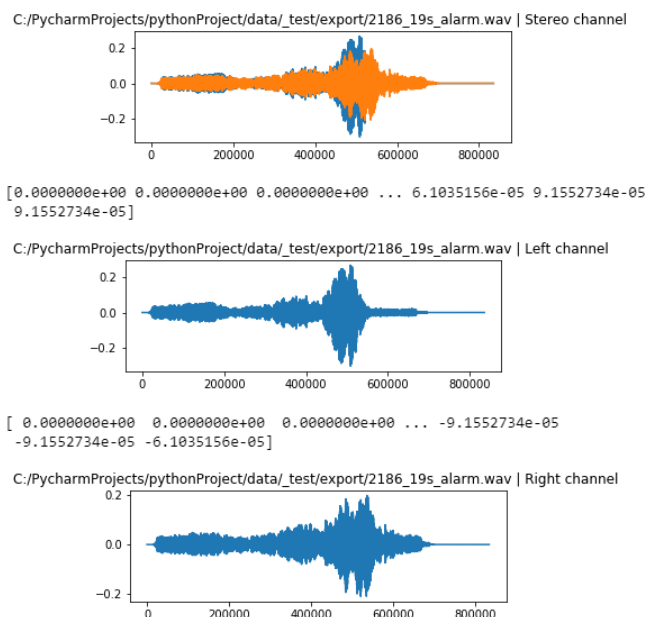
Мәліметтер жинағындағы аудио файлдардың барлығы стерео дыбыстар болмағандықтан зерттеу жұмысы үшін моно дыбыстар таңдалды, сонымен қатар стерео дыбысты жазу үшін қосымша құрылғылар қажет болатындығын да айта кеткен жөн. Келесі тапсырма оң және сол арналарды бір арнаға біріктіру, яғни екі өлшемді массивті бірөлшемді массивке түрлендіру.

Әрі қарай екінші ұяшықта екі өлшемді массивтің соңғы элементінің индексіне сілтеме жасап, екі мән шығарылды (3.6-сурет). Екі арнаның жеке визуалды түрі 3.7-суретте көрсетілген. Нәтижесінде оң (1 аргументі) және теріс (0 аргументі) алынып, бір каналда жеке массив жасаймыз. Өлшемді белгілеу үшін plt.subplot() қолданылады.

```
print(alarm_path_x)
[[ 0.0000000e+00  0.0000000e+00]
 [ 0.0000000e+00  0.0000000e+00]
 [ 0.0000000e+00  0.0000000e+00]
 ...
 [ 6.1035156e-05 -9.1552734e-05]
 [ 9.1552734e-05 -9.1552734e-05]
 [ 9.1552734e-05 -6.1035156e-05]]
(835200, 2)
44100

print(alarm_path_x[835199][0],alarm_path_x[835199][1] ) #accessing to the last element of array
9.1552734e-05 -6.1035156e-05
```

Сурет 3.6 – Екі өлшемді массивтің соңғы екі элементін шығару

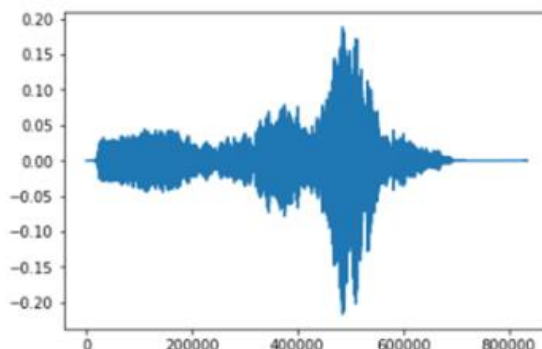


Сурет 3.7 – 2 каналдан тұратын стерео аудио (үстіңгі), сол жақ арна (ортаңғы), оң жақ арна (астыңғы)

Бөлінген екі стерео каналды бір моно арнаға түрлендіру екеуін қосып, орташа мәнін табу арқылы орындалды (3.1). Бұл ретте шуылдардың жоғалатындығы түсінікті, бірақ алғашқы тапсырмада бізге шуылдың емес дыбыстың өзі маңызды.

$$mono = \frac{stereo_{left_i} + stereo_{right_i}}{2} \quad (3.1)$$

мұнда i -каналдың әр толқынындағы итерация саны. Стерео форматтағы аудио үлгінің моно форматқа түрлендірілгеннен кейінгі визуализациясы 3.8-суретте, ал моно файлдың шығысы 3.9-суретте көрсетілген.



Сурет 3.8 – Конвертацияланған мәлімет

```
new_mono_file = sf.read('converted_stereo_to_mono.wav')
print(new_mono_file)
print(mono_or_stereo('converted_stereo_to_mono.wav'))

(array([ 0.00000000e+00,  0.00000000e+00,  0.00000000e+00, ...,
        -3.05175781e-05,  0.00000000e+00,  0.00000000e+00]), 44100)
(array([ 0.00000000e+00,  0.00000000e+00,  0.00000000e+00, ...,
        -3.05175781e-05,  0.00000000e+00,  0.00000000e+00]), 44100)
mono
```

Сурет 3.9 – конвертацияланған моно файл шығысы

3.3 Төтенше жағдайдың импульстік дыбыстарын анықтау

Бұл сатыдағы негізгі тапсырма дыбысты анықтау әдістеріне негізделді. Күдікті оқиға дыбыстарын анықтау екі тапсырмадан тұрады:

- аудио мәліметтер ағынында фондық шуылдардан ерекше жеке импульстік сигналдарды анықтау және бөлу;
- қабылданған сигналды бір немесе бірнеше акустикалық оқиғаларға классификациялау.

Қабатталмайтын аудиосигнал блоктары тізбегінің жиынтығы үшін қуатты анықтау түрлі әдістер үшін негіз болып табылады. N есептеуден тұратын сигналдың k -шы блогындағы қуатын есептеуге арналған теңдеу келесідей жазылады:

$$e(k) = \frac{1}{N} \sum_{n=0}^{N-1} x^2(n + kN), k = 0,1 \quad (3.2)$$

Мысалы: қарудың атылу дыбысы 4,6 с құрайды, $N=4000$ жазба блоктары үшін әр блоктың ұзақтығына қатысты қуат мәндерінің диапазоны 90 мс құрады. Жылдам импульстік шуылға жауап ретінде автономды идентификацияланатын блок, әдіске байланысты түрлі жолдармен орындалады:

- қуат бірліктері үшін стандартталған мәліметтердің стандартты ауытқуына сәйкес;
- қуат бірліктері үшін медиандық фильтрдің орта мәнін қолдану;
- қуат бірліктері үшін динамикалық шектеулер қою. Аталған әдіске терең үнілу, әдістеме негізінде қуаты бірліктерінің нормализацияланған стандартты ауытқуы жатқандығын байқауға болады. $[0, 1]$ диапазонында жататын қуат бірліктерінің нормализацияланған мәні, бұл стратегияда неғұрлым маңызды компонент екендігі анықталды:

$$e_{norm}(j) = \frac{e_{win}(j) - \min(e_{win}(j))}{\max_j(e_{win}(j) - \min(e_{win}(j)))} \quad (3.3)$$

Келесі қадам дисперсия деп аталатын стандартты ауытқуды есептеу:

$$\text{var}(k) = \frac{1}{L-1} \sum_{j=0}^{L-2} [e_{norm}(j, k) - \bar{e}_{norm}(k)]^2 \quad (3.4)$$

Шуыл кездескен сайын қуат блогы $[0, 1]$ аралығында біркелкі таралу тенденциясына ие. Аудио бірлік үшін қуаттың жаңа мәні, аталған диапазон аралығындағы қайта нормализацияланған болғандықтан, егер алдыңғы орнатылған фон бірліктері қуатының мәнімен салыстырғанда қуаттың неғұрлым жоғары деңгейі туындаса, импульстік сигнал арқылы бірлік анықталады. Баяу өзгертін сигнал, нормализацияланған қуат көзінің орташа мәні зерттеу арқылы анықталуы мүмкін, тиісінше бұл шешім шуыл деңгейінің өзгерісіне тұрақты.

3.4 Төтенше жағдайларды анықтау үшін жинақталған аудио мәліметтерден белгілерді алу

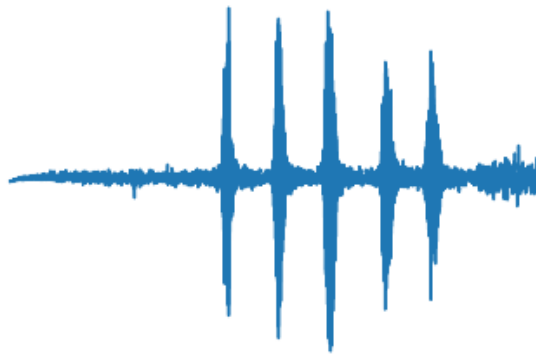
Дыбыстарды қарапайым түрлендіру арқылы пайдалануға болады, Алайда модель бұл түрлендірулерге уақыт жұмсайтындықтан, сонымен қатар модель қабылдай алмайтын диапазондағы деңгейі өте төмен немесе тым жоғары дыбыстар болатындықтан, дыбыстарды белгілер түрлендіру ыңғайлы.

1. Фурье түрлендіруі периодтық функцияларды синусоидалық құраушыларға бөлу. Кез-келген функцияны синусоидалық функциялар суммасына бөледі:

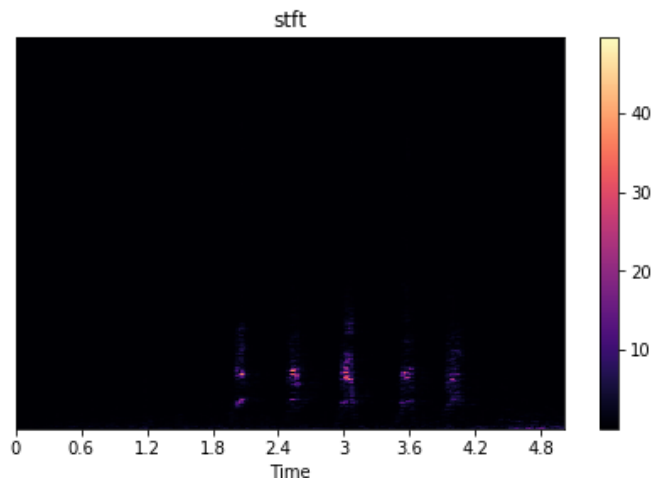
$$\hat{f}(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-ixw} dx \quad (3.5)$$

2. STFT (қысқа уақыттық Фурье түрлендіруі). Қарапайым ортадағы Фурье түрлендіруі барлық жиілік үшін қолданылса, STFT жағдайында синусоидалық фазалық жиілікті анықтау үшін қысқа уақыт аралығы қолданылады (3.10-сурет):

$$F(t, \omega) = \int_{-\infty}^{\infty} f(\tau) W(\tau - t) e^{-i\omega\tau} d\tau \quad (3.6)$$



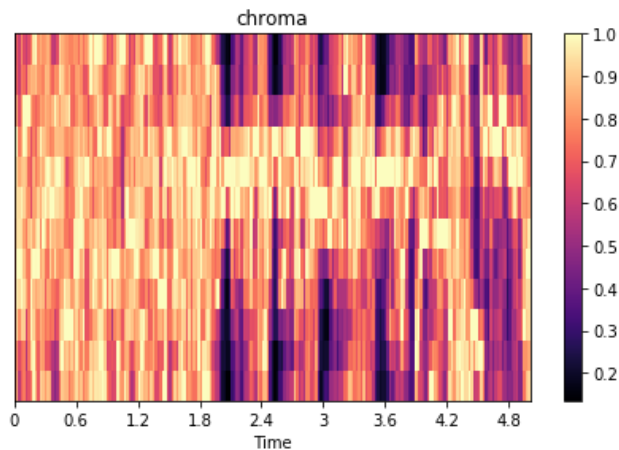
Сурет 3.10 – Иттің үру дыбысының визуализациясы



Сурет 3.11 – STFT түрлендірілген иттің үру дыбысының үлгісі

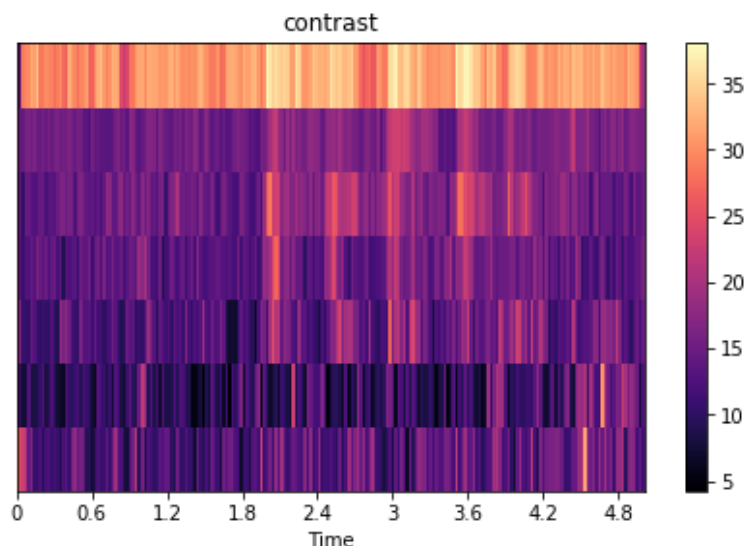
3.11-суретте STFT 1.8 бен 4.2 уақыт аралығындағы белгілерді қабылдағандығын көреміз.

3. Хромограмма – STFT және Фурье түрлендіруі көмегімен дыбыс қуатын алу (3.12-сурет). Дыбыс қуаты арқылы дыбыс деңгейінің төмен не жоғары болғандығына қарамастан дыбысты нақты классификациялай аламыз. Хромограмма біз қолданатын бірінші белгі.



Сурет 3.12 – Хромограмма нәтижесі

4. Спектралдық контраст музыка типтерін алуда қолданылатын белгі (3.13-сурет). Әр спектрограмма бөлігі блоктарға бөлініп, дыбыс энергиясы жоғарғы мәні бойынша төменгі мәнін салыстыру арқылы орындалатын түрлендіру. Жалпы жағдайда жоғары сигналдарда таза немесе тар жолақты, ал кең жолақты шуылдарда төмен мәндер аламыз. Спектралдық контраст біз қолданатын екінші белгі.



Сурет 3.13 – Спектралды контраст

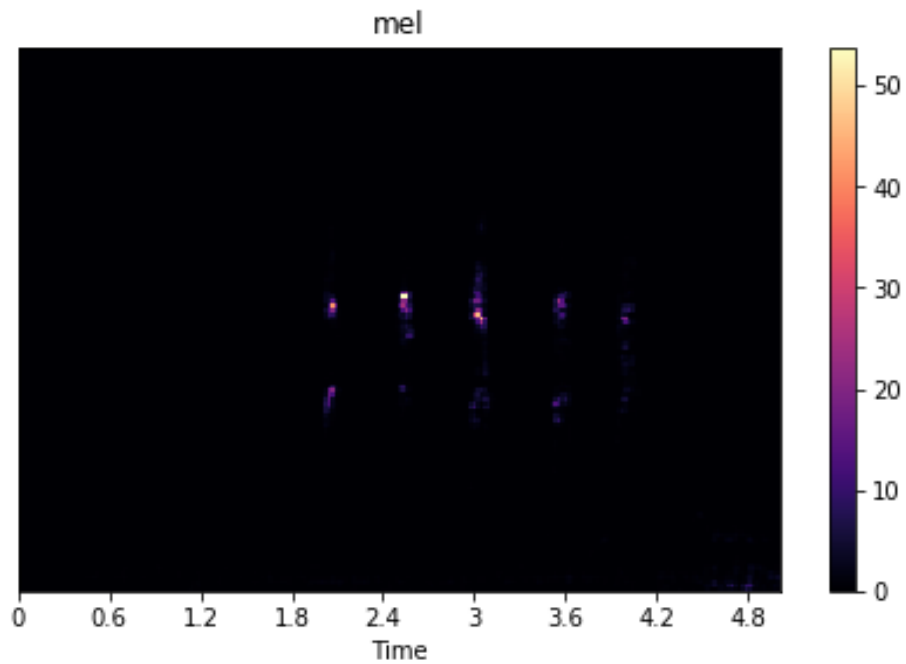
5. Mel-frequency cepstrum (MFC) және Mel-frequency cepstral coefficients (MFCC). Mel шкаласы дыбыстық сигналдар көмегімен логарифмдік түрлендіруді білдіреді (жиілік). Адамдар 10 және 100 Гц айырмашылығын түсіне алады, ал 60 пен 65 Гц ажырату қиынға соғады. Сәйкесінше, алгоритмге де дыбысты Гц мәндерінде сараптау ыңғайлы. Mel-frequency келесі теңдеу бойынша түрлендіруді орындайды (Гц-тен Mel түрлендіру):

$$mel = 1127 * \log\left(1 + \frac{f}{700}\right) \quad (3.7)$$

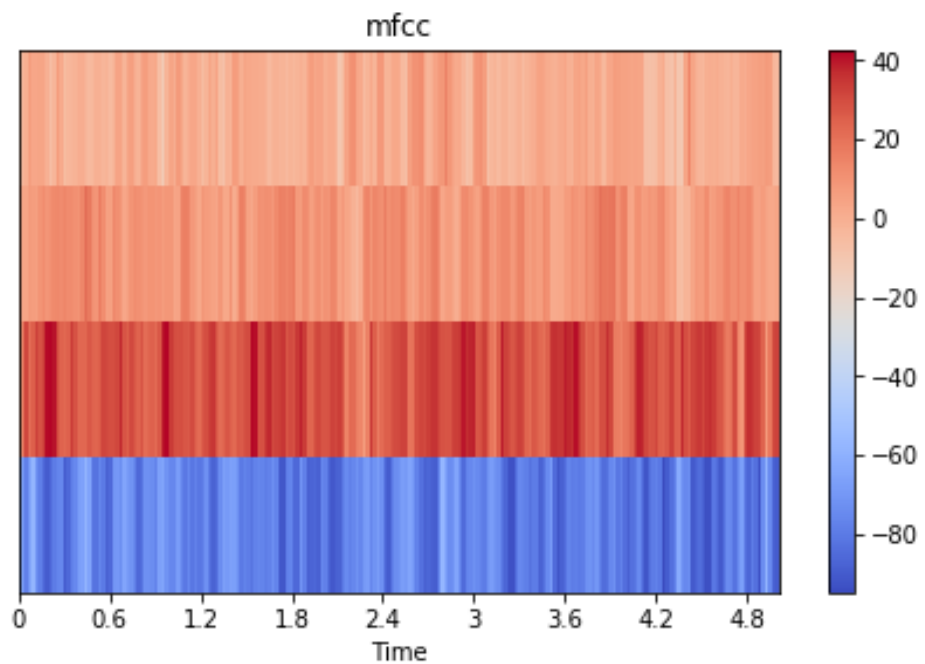
3.14-суреттен оның STFT ұқсас екендігін бірақ мәндері жоғары екендігін байқаймыз (20 және одан жоғары).

Mel-frequency cepstral coefficients mel-ді өзге коэффициенттермен қолданатындықтан солай аталады (3.15-сурет). MFCC алгоритмі келесідей:

1. Герц-ті mel-шкаласына түрлендіру.
2. Аудионы логарифмдеу.
3. Логарифмдік магнитуданы және дискреттік косинустық түрлендіруді (DCT) есептеу.



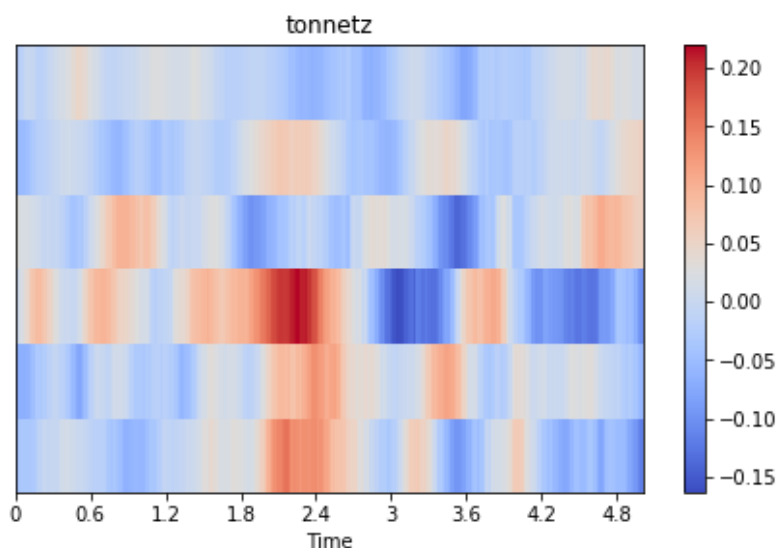
Сурет 3.14 – Mel-frequency



Сурет 3.15 – Mel-frequency cepstral coefficients

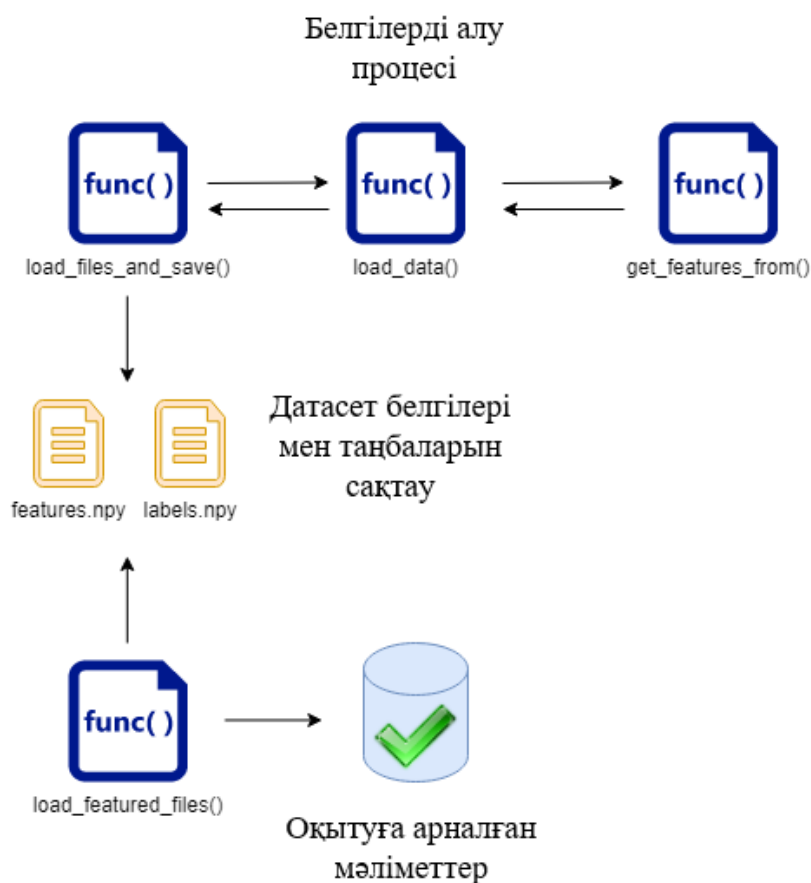
MFCCs және MFC аудио өңдеу үшін машиналық оқытуда маңызды болып табылады. Сондықтан зертеу жұмысында қолданылды.

4. Соңғы қолданылатын белгі tonal centroid features (Tonnetz, неміс тілінен «tone network»). 6 өлшемді координаттарды ұсынуға негізделген. Дыбыстың әр тонының аралығында байланысқан дыбыстарды алуға қолданылады (гармоникалық қатынастағы дыбыстар) (3.16-сурет).



Сурет 3.16 – Tonal centroid features (Tonnetz)

Алынған белгілерді модельді оқытуда қолдану үшін librosa дыбыстық кітапханасын пайдаланып өңдейміз.



Сурет 3.17 – Белгілерді алу диаграммасы

Кодтың аудиосигналдардан белгілерді алу бөлігі (Қосымша А)-да ұсынылған. 3.17-суретте Белгілерді алу диаграммасы келтірілген.

Төрт функция қолданылып, оның үшеуі белгілерді алу және сақтауға жауап берді. Нақтырақ айтқанда, «load_files_and_save()» - «load_data()»-ны шақырады, ал ол мәліметтерді теріп жеке дыбысты алады, содан соң көрсетілген дыбыстың белгілерін алу үшін «get_features_from()»-ды шақырады. Әрі қарай жұмысты жалғастыру үшін барлығы «.npy» форматында екі файлда сақталады (белгілер, таңбалар). Мәліметтерден сақталғаннан соң «load_featured_files()» функциясының көмегімен мәліметтер алынады. Оқыту сатысы 3.18, 3.19-суреттерде көрсетілгендей RNN-CNN моделін оқытудан басталды. Ол екі конволюциялық қабаттан: global maximum pool және global averaging pool.

```
def load_files_and_save():
    features, labels = load_data('./urban_sound/Audio Dataset/');
    print("Shape:", labels.shape)
    np.save('./urban_sound/_audio_dataset_features.npy', features)
    np.save('./urban_sound/_audio_dataset_labels.npy', labels)
```

```
def get_features_from(filename=None):
    if filename:
        X, sample_rate = sf.read(filename, dtype='float32') #more accurate
        #X, sample_rate = librosa.load(filename, dtype='float32')
    else:
        print('Filename is not defined.')
    if X.ndim > 1: X = X[:,0]
    X = X.T #in case of stereo
    short_term_ft= np.abs(librosa.stft(X))
    mfccs_t = np.mean(librosa.feature.mfcc(y=X, sr=sample_rate, n_mfcc=40).T,axis=0)
    chromagram = np.mean(librosa.feature.chroma_stft(S=short_term_ft, sr=sample_rate).T,axis=0)
    mel_spect = np.mean(librosa.feature.melspectrogram(X, sr=sample_rate).T,axis=0)
    spectral_contrast = np.mean(librosa.feature.spectral_contrast(S=short_term_ft, sr=sample_rate).T,axis=0)
    tonnetz_harmonic = np.mean(librosa.feature.tonnetz(y=librosa.effects.harmonic(X), sr=sample_rate).T,axis=0)
    return mfccs_t,chromagram,mel_spect,spectral_contrast,tonnetz_harmonic
```

```
def load_data(parent_dir):
    #parent_dir = "./urban_sound/Audio Dataset/"
    print(len(os.listdir(parent_dir)))
    features = np.empty((0,193))
    labels = []
    labels_to_cat = []
    count_labels = 1
    sub_dirs = os.listdir(parent_dir)
    data_size = len(sub_dirs)
    for sub_dir in sub_dirs:
        full_path = parent_dir + sub_dir + '/'
        if os.path.isdir(full_path):
            date_for_start_parse = datetime.now()
            print('Started parsing dir:' + sub_dir + ' . Time started:' + date_for_start_parse.strftime("%d/%m/%Y %H:%M:%S"))
            for x in tqdm(os.listdir(full_path), position=1, leave=True ):
                file = full_path + os.path.basename(x)
                try:
                    mfccs_t, chromagram, mel_spect, spectral_contrast, tonnetz_harmonic = get_features_from(file)
                except Exception as exception:
                    print("File name:", file)
                    print("Exception:", exception)
                    continue
                ext_features = np.hstack([mfccs_t, chromagram, mel_spect, spectral_contrast, tonnetz_harmonic])
                features = np.vstack([features, ext_features])
                labels_to_cat.append(count_labels) #for Label.npy
            count_labels += 1
            print("PARSE SUCCESS:", sub_dir)
    return np.array(features), np.array(labels_to_cat)
```

Сурет 3.18 – Белгілерді алу

```
PARSE SUCCESS: street_music
Shape: (8674,)
Time elapsed: 5439
```

Сурет 3.19 – Белгілерді алу уақыты және сәтті сарапталған дыбыстар саны

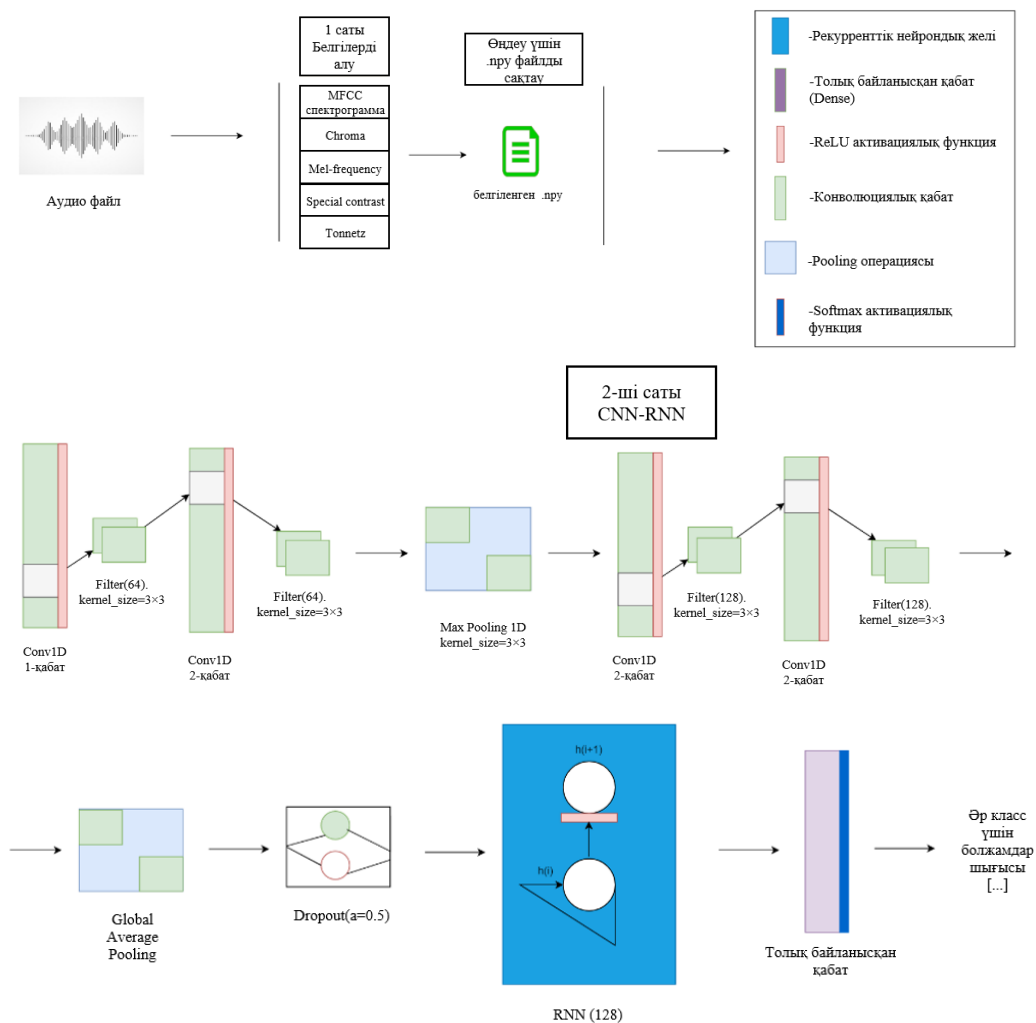
Ашық қолданыстағы Urbansound8K жетілдіріліп, көлемі 6,6 ГБ құрайтын, мәліметтер жиынтығынан белгілерді алу процесі 1,5 сағат уақыт алды. Алғашқыда тек мәліметтер жинағының тек 6,6 ГБ ғана қолданылды. Ал екінші бөлікте әдістерді қолданып, мәліметтерді толықтай жүргізіп өтетін боламыз. Дыбыстарды іріктегеннен соң 8674 дыбыс алынып, 5439 с немесе 90 минутты құрады.

4 АУДИОСИГНАЛДАРДЫ ӨНДЕУ АРҚЫЛЫ ТӨТЕНШЕ ЖАҒДАЙЛАРДЫ АНЫҚТАУҒА ЖӘНЕ КЛАССИФИКАЦИЯЛАУҒА АРНАЛҒАН МОДЕЛЬДІ ОҚЫТУ НӘТИЖЕЛЕРІ

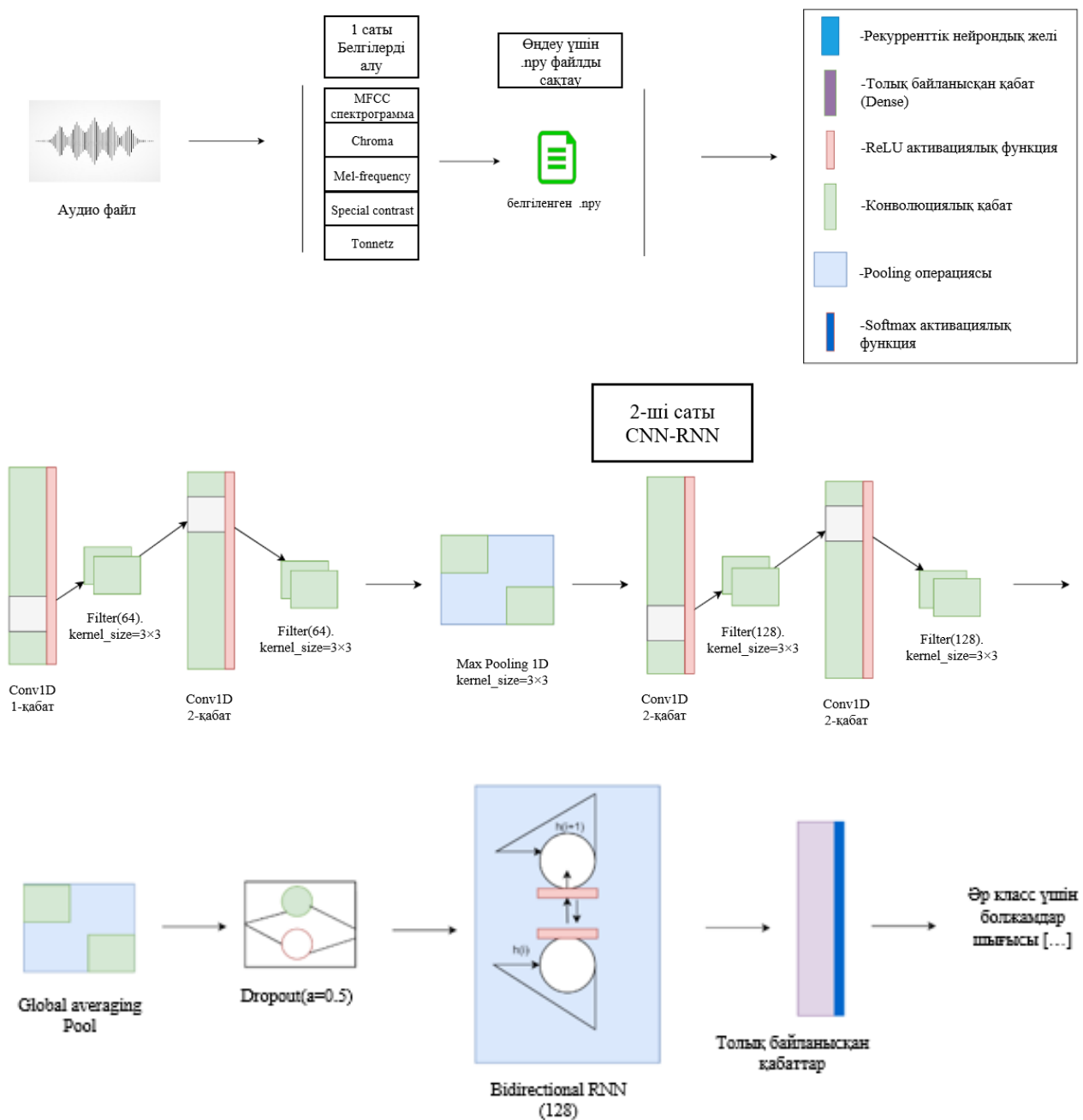
4.1 Төтенше жағдайларды анықтау үшін аудио мәліметтерді өңдеуге ұсынылатын CNN-BiLSTM моделінің архитектурасы

Алғашқыда CNN қолданып, екі нейрондық қабат алынып, олардың өлшемін «фильтр жадысы» ретпен арттырылды. 1-ші саты белгілерді алу, содан соң CNN қабаты орналастырылды. CNN архитектурасы және соңғы averaging pool мен dropout кейбір нейрондарды кездейсоқ түрде алып тастайды, содан соң толық байланысқан қабат арқылы шығыс мәндер алынады (softmax активациясының функциясы).

Зерттеу жұмысында CNN архитектурасының RNN-мен комбинациялана отырып ұсынылады. RNN ақпаратты алу үшін ReLU активациясы бар жеке қабат ретінде қарастырылды. RNN өлшемі 128-ге тең және архитектурасы 4.1, 4.2-суреттерде көрсетілген.

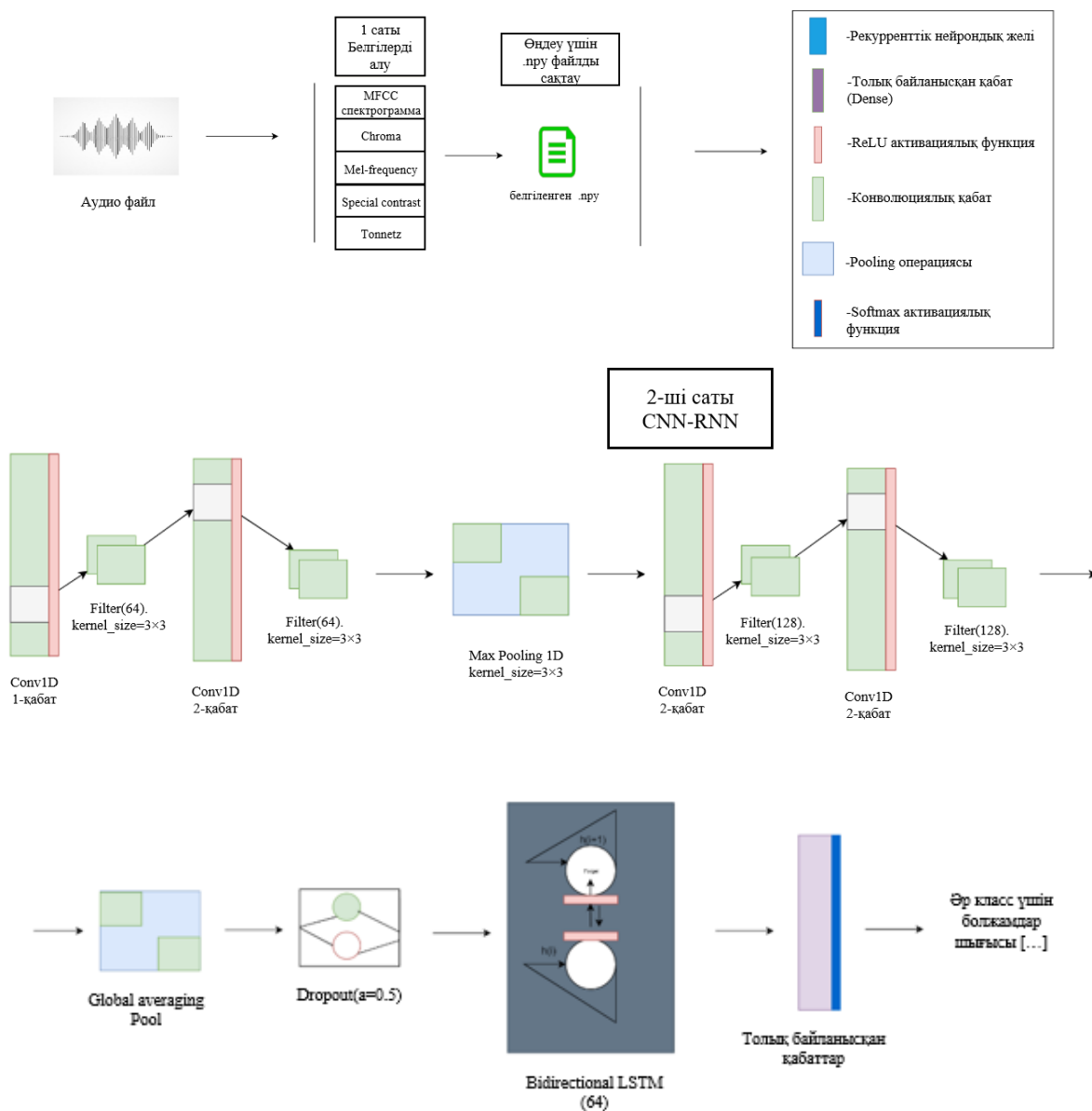


Сурет 4.1 – CNN-RNN моделінің архитектурасы



Сурет 4.2 – CNN-BiRNN моделінің архитектурасы

LSTM қысқа-ұзақ уақытты жады желісі. Оның негізгі ерекшелігі векторлардың ескі ақпаратты жаңартып, жаңа ақпаратты сақтау үшін қажет ақпаратты жадыдан өшіре алуы. Зерттеу жұмысында оның екі жақа бір мезгілде өтуі бар модификациясы қолданылды (4.3-сурет).



Сурет 4.3 – Ұсынылатын CNN-BiLSTM желісі қолданылған модельдің архитектурасы

Call backs (кері шақыру) міндетті емес шақырулар, алайда олар бізге ең оптималды модельді алуға мүмкіндік береді: Ерте тоқтату, сақтау нүктесі және платоға төмендету.

Ерте тоқтату – оқыту кезінде эпохалардың оптималды санын білмеген кезде, артық оқытуға жол бермес үшін қолданылады.

Келесі шақыру түрі сақтау нүктесі – оқыту кезінде максимум нақтылық және минимум жоғалтуды бақылау нүктесі. Осылайша модельді қажетті сегментте сақтау мүмкін болады.

Платоға төмендету – оқыту максимумына жетіп, оқыту жылдамдығын азайту мүмкіндігі бар. Нәтижесінде модельді белгілі бір периодта жақсарта аламыз.

4.2 CNN-BiLSTM моделін оқыту үшін гиперпараметрлерді таңдау және бағалау метрикалары

Alpha параметрлерді таңдау модельді таңдаудағы негізгі саты. Бірінші альфа параметр learning rate (оқыту жылдамдығы) деп аталады. Ол оқытудың интенсивтілігіне жауап береді және 1-ге дейінгі мәндер таңдалады. Alpha параметрдің мәні аз болған сайын оқыту интенсивтілігі жоғары болады. Зерттеу жұмысында Alpha 0,0025-ге тең деп алынды. Сонымен қатар Alpha параметр деп эпоха санын да атауға болады. ол толық айналымдар санын көрсетеді және ол мәліметтердің күрделілігіне байланысты. Соңғы параметр batch size (партиялар өлшемі). Бір айналымда нейрондық желіні қаншалықты толтыруға болатындығын көрсетеді. Әдетте ол жоғары мәндерге ие бола алады, алайда зерттеу жұмысында 64-ке тең деп алынды.

Қателіктер оптимизациясы үшін RMSProp (орташа квадраттық таралу) таңдалды. Ол минимумды табудың ең жылдамы әрі зерттеу жұмысында еш кедергісіз жұмыс істей алады.

Импульстік дыбыстарды анықтау үшін accuracy, precision, recall, бағалау қателігінің санын хабарлайтын F1-score және AUC-ROC сияқты параметрлер қолданылды. Әр бағалау параметрлерінің (4.1)-(4.4) теңдеулері төмнде келтірілген. Сонымен қатар, ұсынылатын CNN-RNN терең моделінің эффективтілігін түсіндіру үшін, зерттеуде оқыту және тестілеу жоғалтуы, эпох саны келтірілген. Мұндағы TP (true positive) шынайы-оң, TN (true-negative) шынайы теріс, FP (false positive) жалған-оң, FN (false negative), P оң мәндер, ал N теріс мәндерді білдіреді:

$$accuracy = \frac{TP+TN}{P+N} \quad (4.1)$$

$$precision = \frac{TP}{TP+FP} \quad (4.2)$$

$$recall = \frac{TP}{TP+FN} \quad (4.3)$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (4.4)$$

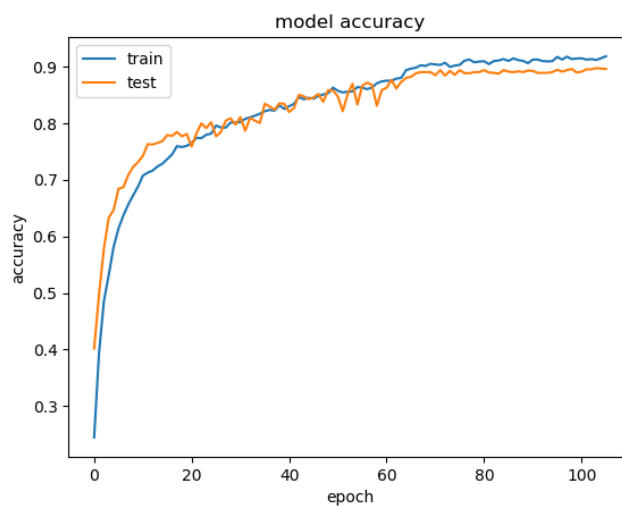
4.3 CNN-BiLSTM комбинациясына негізделген төтенше оқиғаларды анықтауға арналған модельді оқыту бойынша алынған нәтижелер

Екінші датасет үшін орындалған тапсырмаларды көлемі 30 ГБ артық жинақталған датасет үшін қайталаймыз. Алдымен белгілерді (4.4-сурет) алып Bidirectional LSTM моделін оқытамыз. CNN-BiLSTM комбинациясына негізделген модельдің архитектурасы (Қосымша Ә)-да ұсынылған.

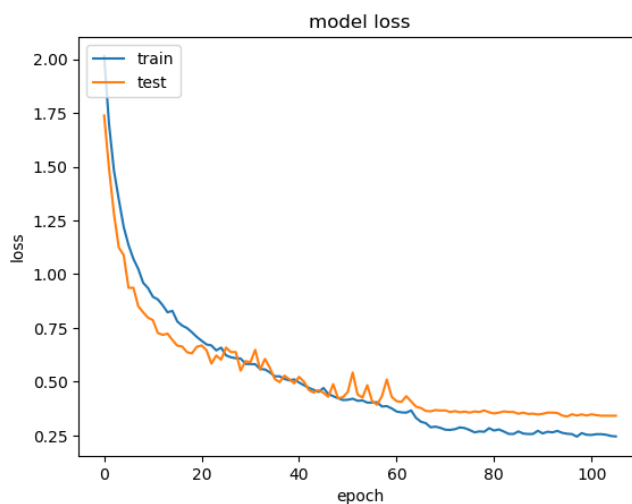
```
sample: 10000
-----
PARSE SUCCESS: weapon
Shape: (7230,)
Time elapsed: 30313
```

Сурет 4.4 – Жинақталған датасеттен белгілерді алу, мәліметтер жиынтығының жалпы ұзақтығы мен сәтті сарапталған дыбыстар саны

Белгілерді алу 30 313 с немесе 505 мин-қа, яғни шамамен 8 сағат уақытқа созылды. Жинақталған датасет негізінде оқытылған модельдің accuracy мен loss графиктері 4.5-суретте көрсетілген.



а

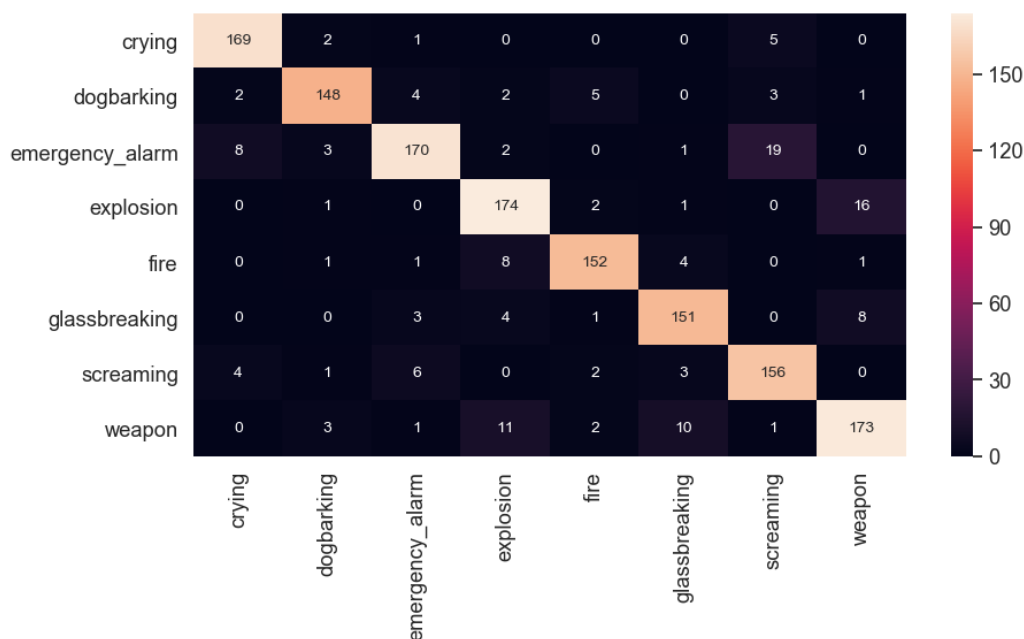


ә

а – оқыту/тестілеу кезіндегі нақтылығы; ә – оқыту/тестілеу кезіндегі жоғалтуы

Сурет 4.5 – Жинақталған датасет негізінде CNN-BiLSTM моделін оқыту және тестілеу кезіндегі нақтылығы мен жоғалтуы

4.6-суретте оқытылған модель бойынша алынған confusion matrix, ал 4.7-суретте болжау пайызы көрсетілген.



Сурет 4.6 – CNN-BiLSTM моделі бойынша алынған confusion matrix

Оқытылған модель жалған-оң, жалған-теріс, шынайы-оң және шынайы-теріс іріктемелер бойынша түрлі дыбыстар типіне және болжау процентіне қатысты confusion matrix алуға мүмкіндік берді. Қолданылған модель қауіпті қала дыбыстарының сегіз типін классификациялады.

```
{'crying': 169, 'dogbarking': 148, 'emergency_alarm': 170, 'explosion': 174, 'fire': 152, 'glassbreaking': 151, 'screaming': 156, 'weapon': 173}
{'crying': 177, 'dogbarking': 165, 'emergency_alarm': 203, 'explosion': 194, 'fire': 167, 'glassbreaking': 167, 'screaming': 172, 'weapon': 201}
{'crying': '0.9548022598870056%', 'dogbarking': '0.896969696969697%', 'emergency_alarm': '0.8374384236453202%', 'explosion': '0.8969072164948454%', 'fire': '0.9101796407185628%', 'glassbreaking': '0.9041916167664671%', 'screaming': '0.9069767441860465%', 'weapon': '0.8606965174129353%'}
```

Сурет 4.7 – Барлық класстар бойынша болжау пайызы

Графиктен көргеніміздей нәтижелердей айтарлықтай жақсы: минималды 83% emergency alarm және максималды 95% cry категориясы бойынша болжам алынды. 4.1-кестеде қолданылған модельдің нақтылығы көрсетіліп, әр классты ассығасу precision, recall, F-Score сияқты түрлі параметрлер бойынша бағалау ұсынылған.

Кесте 4.1 – Импульстік аудио оқиғаларды анықтау нәтижелері

Төтенше оқиғалар	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Қару-жарактың атылуы	0.8606	0.9012	0.9206	0.9107	0.9557
Терезенің сыну дыбысы	0.9041	0.9534	0.9004	0.9260	0.9334
Өрттің жалыны	0.9101	0.9112	0.9097	0.9104	0.9345
Сиренаның дыбысы	0.8374	0.9223	0.9669	0.9440	0.9402
Жарылыс кезіндегі дыбыс	0.8969	0.8025	0.8141	0.8082	0.9117
Адам жылағандағы дыбыс	0.9548	0.8909	0.8933	0.8920	0.9228
Ит үргенде шығатын дыбыс	0.8969	0.8108	0.8322	0.8213	0.9264
Адамның айқайы	0.9069	0.8221	0.8347	0.8283	0.9253

Нәтижесінде ұсынылған терең нейрондық желі қауіпті дыбыстарды тізбектей анықтауда барлық бағалау критерийлері бойынша жоғарғы өнімділік көрсетті.

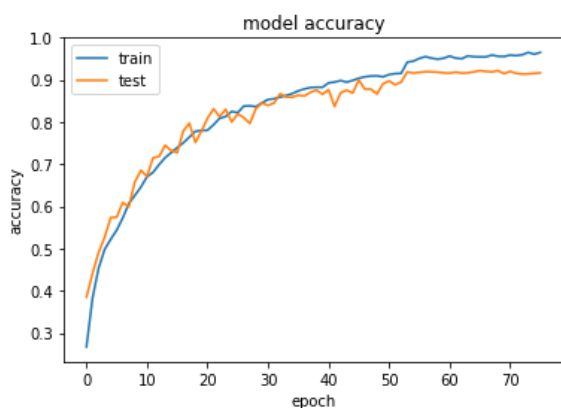
4.4 Ұсынылатын CNN-BiLSTM моделін ашық қолданыстағы urbansound8K және ESC-10 аудио мәліметтер жинағымен тексеру

Бұл саты CNN моделін оқытудан басталды. Ол global maximum pool және global averaging pool екі кноволуциялық қабаттан тұрады (4.8-сурет).

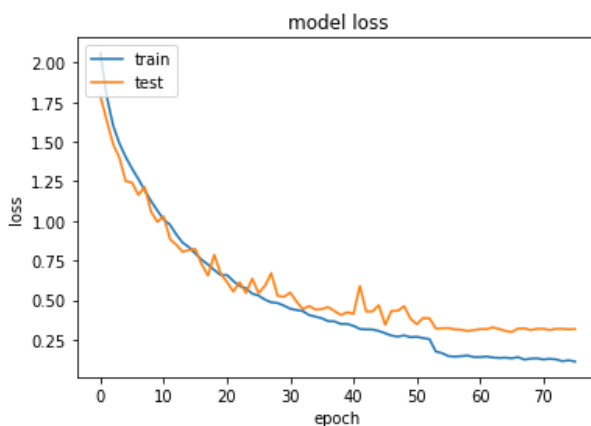
```
=====
Total params: 87,882
Trainable params: 87,882
Non-trainable params: 0
```

Сурет 4.8 – CNN моделінің жалпы параметрлері

Алынған параметрлер саны 87 822. Оқыту 267 с, яғни 4 мин созылды (4.9-сурет).



а

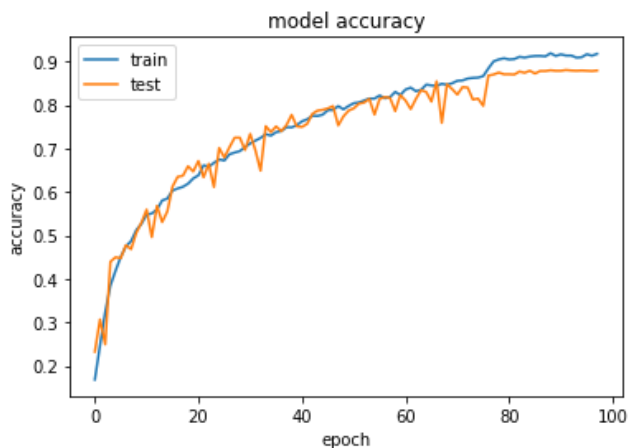


ә

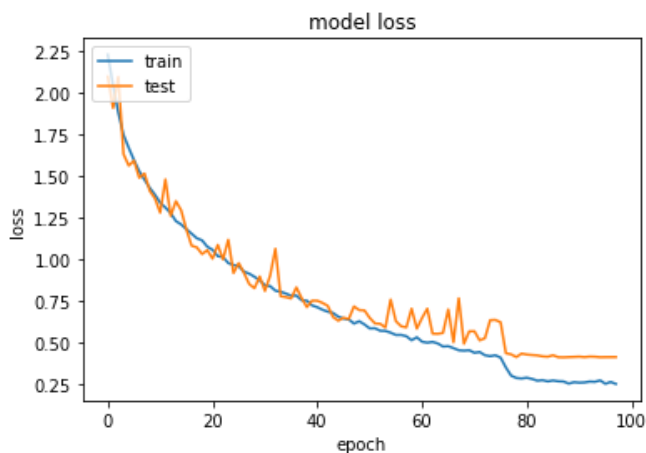
а – оқыту/тестілеу кезіндегі нақтылығы; б – оқыту/тестілеу кезіндегі жоғалтуы

Сурет 4.9 – CNN моделін оқыту және тестілеу кезіндегі нақтылығы мен жоғалтуы

4.9-суреттен артық оқыту болу мүмкіндігінің бар екендігін көруге болады. сондықтан early stopping қолданылды. Содан соң бірнеше сәтсіз CNN-SimpleRNN модельдерінен соң, 128 өлшемді рекуррентті желі алынды. (4.10-сурет).



а

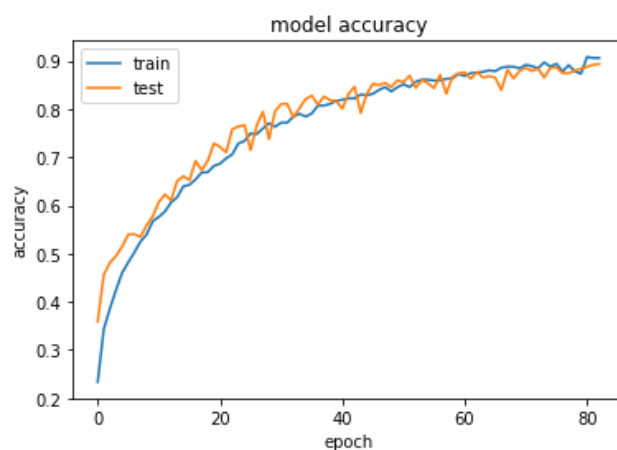


б

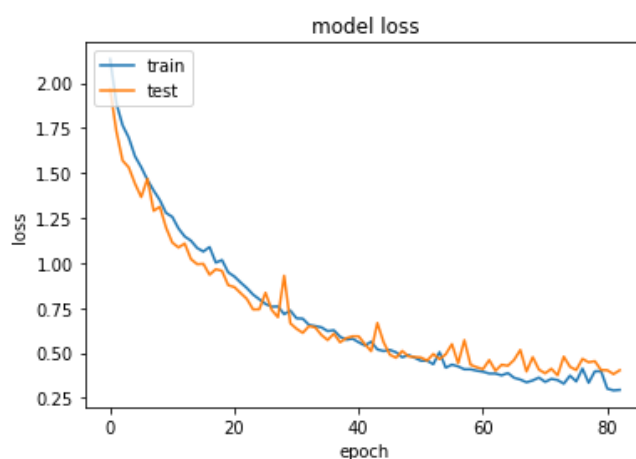
а – оқыту/тестілеу кезіндегі нақтылығы; б – оқыту/тестілеу кезіндегі жоғалтуы

Сурет 4.10 – CNN-RNN моделін оқыту және тестілеу кезіндегі нақтылығы мен жоғалтуы

CNN бойынша соңғы жоғалтуды бағалау 0.31 (30%), нақтылық 91%, ал CNN-RNN бойынша жоғалту 0.41 (41%), нақтылық 87% көрсетті. Сонымен қатар, CNN-Bi LSTM желісі параметрлер санын 325 330-ға дейін күрт арттырды. Бұл теориялық тұрғыда жақсы нәтиже болып есептеледі. Жалпы оқыту уақыты 871 с немесе 14 мин құрады (4.11-сурет).



а



ә

а – оқыту/тестілеу кезіндегі нақтылығы; б – оқыту/тестілеу кезіндегі жоғалтуы

Сурет 4.11 – CNN-BiLSTM моделін оқыту және тестілеу кезіндегі нақтылығы мен жоғалтуы

Шығыстағы соңғы сегментте accuracy 89% ал loss 40% жетті (4.2-кесте).

Кесте 4.2 – Модельді оқыту нәтижелері

Модельдің аты	Эпоха(max/stopped)	Accuracy (train/test)	Error (train/test)	Алынған параметрлер саны
CNN	100/83	96%/91%	11%/31%	87 882
CNN-SimpleRNN	100/98	91%/87%	24%/41%	120 778
CNN-Bidirectional LSTM	100/83	90%/89%	29%/40%	352 330

Бұл тапсырмада CNN-BiRNN әдеттегідей оқытылмағанымен, жақсы нәтижелер көрсетті. Алайда нәтижелер (нақтылық/қателік жиілігі) бойынша оптималды модель ретінде CNN таңдалды.

Модельді тексеру үшін келесі аудио мәліметтер жиынтығы ретінде шағын ESC-10 датасеті таңдалды. Датасет иттің үруі және балғамен ұрғандағы дыбыстардан тұрды. Иттің үруі 40 дыбыстан тұратын датасет (4.12-сурет).

```
: evaluate_wild_count('CNNRNN.h5')
   evaluate_wild_count('CNN.h5')
   evaluate_wild_count('BidirectRNN.h5')
   evaluate_wild_count('BidirectLSTM.h5')

Modelname CNNRNN.h5
Finished. Correct: 37
Modelname CNN.h5
Finished. Correct: 37
Modelname BidirectRNN.h5
Finished. Correct: 38
Modelname BidirectLSTM.h5
Finished. Correct: 38
```

Сурет 4.12 – «Иттің үруі» датасеті бойынша модельдің нәтижелері

Бұл жолы ұсынылатын CNN-BiLSTM моделі 38 жағдайда оқиғаны дұрыс анықтай алды (барлығы 40 оқиға).

ҚОРЫТЫНДЫ

Диссертациялық зерттеу жұмысының нәтижелері бойынша қысқа **тұжырымдар**. Төтенше оқиғаларды анықтау үшін акустикалық құрылғылардан алынған аудиосигналдарды қолдану ұтымды шешім болып табылады. Сондай шешімді ұсына отырып, зерттеу жұмысында төмендегі тапсырмалар орындалды:

1. Төтенше оқиға дыбыстарын анықтауда қолданылған машиналық оқыту модельдері мен архитектураларына, қолданылған мәліметтер жиынтықтарына, аудио мәліметтерді алдын-ала өңдеу әдістеріне, модельді бағалау метрикаларына әдеби шолу жүргізіліп, салыстырмалы талдау жасалды.

2. Төтенше оқиға дыбыстарын анықтауға арналған машиналық оқыту моделін оқыту және тестілеуге арналған 10 000 аудио мәліметтен, көлемі 30 ГБ құрайтын және жылау, айқайлау, иттің үруі, терезенің сынуы, жарылыс, қару жарақ дыбысы, өрт жалыны, сирена сияқты әлеуметтік сипаттағы төтенше жағдайлардың сегіз классы бойынша аудио мәліметтер жиынтығы жасалды. Аудио мәліметтер акустикалық әдіс бойынша жазылып және интернет желісінен, youtube желісінен алынған мәліметтермен толықтырылды.

3. Жинақталған датасет негізінде төтенше оқиғаларды анықтайтын және мультиклассификация тапсырмасын орындауға арналған терең оқыту моделі құрылды.

4. Құрылған CNN-BiLSTM моделі тестіленді және accuracy, precision, recall, F1-score метрикаларымен бағаланды. Модельдің ең жоғарғы нақтылықты - 95% жылау, ең төменгі нақтылықты - 83% сиренаның дыбысы класстары бойынша көрсетті.

CNN-BiLSTM комбинациясы архитектурасы қарапайым RNN желілерімен салыстырғанда аудиосигналдың ерекшелік белгілерін терең үйрене отырып, оқытылатын параметрлер санын арттыруға мүмкіндік берді. Бұл CNN желісінің уақыт аралықтарының реттілігіне сезімталдығының болмауы және оның неғұрлым маңызды белгілерге ғана назар аударуымен және BiLSTM желісінің аудио мәліметті екі бағытта өңдейтіндігімен түсіндіруге болады. Параметрлер санының артуы модель нақтылығының жоғары болуына ықпал етеді.

Сонымен қатар, ашық қолданыстағы ESC-10 және Urbansound8K датасеті жетілідіріле отырып, ұсынылған CNN-BiLSTM моделін тексеруге қолданылды.

Ұсынылған модель темір жол, әуе, көлік бекеттерінде, сауда-ойын сауық орталықтарында, білім беру мекемелерінде, тұрғын үй кешендері сияқты қоғамдық орындарда және т.б инфрақұрылымдарда аудиосигналдар негізінде төтенше оқиғаларды сегіз класс бойынша анықтап, тиісті мекемелерге алдын-алу туралы хабар береді.

Диссертациялық жұмыстың тақырыбы бойынша 2022-2024 жылдарға арналған «Жас ғалым» жобасы бойынша жас ғалымдардың зерттеулерін гранттық қаржыландыру бойынша «AP149715555 Машиналық оқыту әдістерін

қолдану арқылы нақты уақыт режимінде ғимарат ішіндегі қауіпсіздікті қамтамасыз ету жүйесін жобалау және енгізу» жобасы орындалуда.

Диссертациялық жұмыс барысында алынған нәтижелер бойынша дайындалған автореферат «Қазақстан Республикасы ПМ М. Есболатов атындағы Алматы академиясының» оқу үрдісіне енгізілді. Оқу үрдісіне ендіру актісі (Қосымша Б)-да ұсынылған.

ПАЙДАЛАНЫЛҒАН ӘДЕБИЕТТЕР ТІЗІМІ

- 1 Почти 300 млн. детей от 2 до 4 лет подвергаются физическим наказаниям // <https://news.un.org/ru/story/2020/06/1380372>. 10.01.2023.
- 2 66% детей в Казахстане сталкиваются со школьным насилием и дискриминацией // <https://www.inform.kz/ru/66-detey-v-kazahstane>. 10.01.2023.
- 3 Половина подростков в мире испытывает насилие на территории школы и за ее пределами // <https://www.unicef.org/kazakhstan>. 10.01.2023.
- 4 Количество зарегистрированных случаев бытового насилия в отношении женщин // <https://gender.stat.gov.kz/page/frontend/detail?>. 10.01.2023.
- 5 Общее число зарегистрированных преступлений, всего // https://gender.stat.gov.kz/page/frontend/detail?id=89&slug=-74&cat_id. 11.01.2023.
- 6 Worldwide gun deaths reach 250,000 yearly; US ranks high // <https://www.cnbc.com/2018/08/29/worldwide-gun-deaths-reach-250000>. 11.01.2023.
- 7 Ежегодно 250 тысяч человек погибают в результате применения огнестрельного оружия // <https://news.un.org/ru/story/2018/05/1329842>. 11.01.2023.
- 8 Syed A.S., Sierra-Sosa D., Kumar A. et al. IoT in Smart Cities: A Survey of Technologies, Practices and Challenges // *Smart Cities*. – 2021. – Vol.4, Issue 2. – P. 429-475.
- 9 Vijeikis R., Raudonis V., Dervinis G. Efficient Violence Detection in Surveillance // *Sensors*. – 2022. – Vol. 22, Issue 6. – P. 2216-1-2216-15.
- 10 Sreenu G., Saleem Durai M.A. Intelligent video surveillance: a review through deep learning techniques for crowd analysis // *J Big Data*. – 2019. – Vol. 6, Issue 1. – P. 48-1-48-27.
- 11 Mushtaq Z., Su S.F. Efficient Classification of Environmental Sounds through Multiple Features Aggregation and Data Enhancement Techniques for Spectrogram Images // *Symmetry*. – 2020. – Vol. 12, Issue 11. – P. 1822-1-1822-34.
- 12 Wu N., Haruyama S. The 20k Samples-Per-Second Real Time Detection of Acoustic Vibration Based on Displacement Estimation of One-Dimensional Laser Speckle Images // *Sensors*. – 2021. – Vol. 21, Issue 9. – P. 2938-1-2938-15.
- 13 Ferdaous Jenhani, Mohamed Salah Gouider, Lamjed Ben Said. Streaming Social Media Data Analysis for Events Extraction and Warehousing using Hadoop and Storm: Drug Abuse Case Study // *Procedia Computer Science*. – 2019. – Vol.159. – P. 1459-1467.
- 14 Pujol F.A., Mora H., Pertegal M.L. A soft computing approach to violence detection in social media for smart cities // *Soft Comput*. – 2020. – Vol. 24. – P. 11007-11017.
- 15 Ye L., Liu T., Han T. et al. Campus Violence Detection Based on Artificial Intelligent Interpretation of Surveillance Video Sequences // *Remote Sens*. – 2021. – Vol. 13, Issue 4. – P. 628-1-628-17.
- 16 Baba M., Gui V., Cernazanu C. et al. A Sensor Network Approach for Violence Detection in Smart Cities Using Deep Learning // *Sensors*. – 2019. – Vol. 19, Issue 7. – P. 1676-17-1676-17.

- 17 Choqueluque-Roman D., Camara-Chavez G. Weakly Supervised Violence Detection in Surveillance Video // *Sensors*. – 2022. – Vol. 22. – P. 4502-1-4502-29.
- 18 Son J.-Y., Chang J.-H. Attention-Based Joint Training of Noise Suppression and Sound Event Detection for Noise-Robust Classification // *Sensors*. – 2021. – Vol. 21, Issue 20. – P. 6718-1-6718-13.
- 19 Nogueira A.F.R., Oliveira H.S., Machado J.J.M. et al. Sound Classification and Processing of Urban Environments: A Systematic Literature Review // *Sensors*. – 2022. – Vol. 22, Issue 22. – P. 8608-1-8608-30.
- 20 Giannakopoulos T., Spyrou E., Perantonis S.J. Recognition of Urban Sound Events Using Deep Context-Aware Feature Extractors and Handcrafted Features // *Proced. 15th IFIP internat. conf. on Artificial Intelligence Applications and Innovations (AIAI)*. – Greece: Hersonissos, 2019. – P. 184-195.
- 21 Noh K., Chang J.H. Joint optimization of deep neural network-based dereverberation and beamforming for sound event detection in multi-channel environments // *Sensors*. – 2020. – Vol. 20, Issue 7. – P. 1883-1-1883-13.
- 22 Zhou Q. et al. Adaptive Noise Reduction for Sound Event Detection Using Subband-Weighted NMF // *Sensors*. – 2019. – Vol. 19. – P. 3206-1-3206-19.
- 23 Subramanian A.S., Wang X., Watanabe S. et al. An Investigation of End-to-End Multichannel Speech Recognition for Reverberant and Mismatch Conditions // <https://arxiv.org/pdf/1904.09049.pdf>. 28.03.2020.
- 24 Kim N.K., Jeon K.M., Kim H.K. Convolutional Recurrent Neural Network-Based Event Detection in Tunnels Using Multiple Microphones // *Sensors*. – 2019. – Vol. 19. – P. 2695-1-2695-21.
- 25 Almaadeed N., Asim M., Al-Maadeed S. et al. Automatic Detection and Classification of Audio Events for Road Surveillance Applications // *Sensors*. – 2018. – Vol. 18. – P. 1858-1-1858-19.
- 26 Parathai P., Tengtrairat N., Woo W.L. et al. Efficient Noisy Sound-Event Mixture Classification Using Adaptive-Sparse Complex-Valued Matrix Factorization and OvsO SVM // *Sensors*. – 2020. – Vol. 20. – P. 4368-1-4368-24.
- 27 Su Y., Zhang K., Wang J. et al. Environment Sound Classification Using a Two-Stream CNN Based on Decision-Level Fusion // *Sensors*. – 2019. – Vol. 19. – P. 1733-1-1733-15.
- 28 Veena S., Aravindhar D.J. Sound Classification System Using Deep Neural Networks for Hearing Impaired People // *Wireless Pers Commun.* – 2022. – Vol. 126. – P. 385-399.
- 29 Shukla V., Singour M. Multimodal Learning for Early Detection of Explosive Sounds using Relative Spectral Distribution // *Proced. Sensor Signal Processing for Defence conf. (SSPD 2020)*. – Edinburgh, UK, 2020. – P. 1-5.
- 30 Arpitha Y., Madhumathi G.L., Balaji N. Spectrogram analysis of ECG signal and classification efficiency using MFCC feature extraction technique // *J Ambient Intell Human Comput.* – 2022. – Vol. 13. – P. 757-767.
- 31 Dua M., Aggarwal R.K., Biswas M. GFCC based discriminatively trained noise robust continuous ASR system for Hindi language // *Journal of Ambient Intelligence and Humanized Computing*. – 2019. – Vol. 10. – P. 2301-2314.

- 32 Su Y. et al. Performance analysis of multiple aggregated acoustic features for environment sound classification // *Applied Acoustics*. – 2020. – Vol. 158. – P. 107050-1-107050-33.
- 33 Досбаев Ж.М., Омаров Б.С., Садықова Б.М. және т.б. Машиналық оқыту әдістерін қолдана отырып дыбысты жіктеу // *АЭБУ хабаршысы*. – 2022. №2(57). – Б. 70-81.
- 34 Venkatesh S., Moffat D., Miranda E.R. You Only Hear Once: A YOLO-like Algorithm for Audio Segmentation and Sound Event Detection // *Appl. Sci.* – 2022. – Vol. 12. – P. 3293-1-3293-17.
- 35 Boubchir L., Daachi B., V. Pangracious. A review of feature extraction for EEG epileptic seizure detection and classification // *Proccrd. 40th internat. conf. on Telecommunications and Signal Processing (TSP)*. – Barcelona, 2017. – P. 456-460.
- 36 Parathai P., Tengtrairat N., Woo W.L. et al. Single-Channel Signal Separation Using Spectral Basis Correlation with Sparse Nonnegative Tensor Factorization // *Circuits Syst Signal Process.* – 2019. – Vol. 38. – P. 5786-5816.
- 37 Magron P., Virtanen T. Towards Complex Nonnegative Matrix Factorization with the Beta-Divergence // *Procced. 16th internat. Workshop on Acoustic Signal Enhancement (IWAENC)*. – Tokyo, 2018. – P. 156-160.
- 38 Mesaros A., Heittola T., Ellis D. Datasets and Evaluation // *In book: Computational Analysis of Sound Scenes and Events*. – Cham: Springer, 2017. – P. 147-179.
- 39 Mesaros A., Heittola T., Virtanen T. et al. Sound Event Detection: A tutorial // *Signal Processing Magazine (IEEE)*. – 2021. – Vol. 38, Issue 5. – P. 67-83.
- 40 Hayashi T.S., Watanabe T., Toda T. et al. Duration-Controlled LSTM for Polyphonic Sound Event Detection // *ACM Transactions on Audio, Speech and Language Processing (IEEE)*. – 2017. – Vol. 25, Issue 11. – P. 2059-2070.
- 41 Mnasri Z., Rovetta S., Masulli F. Anomalous sound event detection: A survey of machine learning based methods and applications // *Multimed Tools Appl.* – 2022. – Vol. 81. – P. 5537-5586.
- 42 Purwins H. et al. Deep learning for audio signal processing // *IEEE Journal of Selected Topics in Signal Processing*. – 2019. – Vol. 13, Issue 2. – P. 206-219.
- 43 Chavdar M. et al. Towards a system for automatic traffic sound event detection // *Procced. 28th Telecommunications forum (TELFOR - 2020)*. – Belgrade, 2020. – P. 1-4.
- 44 Mendoza J.M. et al. Audio event detection using wireless sensor networks based on deep learning // *Wireless Internet: proceed. 11th EAI internat. conf., WiCON*. – Taipei, 2019. – P. 105-115.
- 45 Adavanne S., Politis A. et al. Multi-Room Reverberant Dataset for Sound Event Localization and Detection // <https://arxiv.org/pdf/1905.08546.pdf>. 10.02.2023.
- 46 Demir F. et al. A new pyramidal concatenated CNN approach for environmental sound classification // *Applied Acoustics*. – 2020. – Vol. 170. – P. 107520.

- 47 Guzhov A. et al. Audioclip: Extending clip to image, text and audio // *Proced. IEEE internat. conf. on Acoustics, Speech and Signal Processing (ICASSP - 2022)*. – Singapore, 2022. – P. 976-980.
- 48 Salamon J., Jacoby C., Bello J.P. UrbanSound8K dataset // <https://urbansounddataset.weebly.com/urbansound8k.html>. 10.02.2023.
- 49 Nanni L. et al. An ensemble of convolutional neural networks for audio classification // *Applied Sciences*. – 2021. – Vol. 11, Issue 13. – P. 5796-1-5796-18.
- 50 Liu Y. et al. Cross-modal spectrum transformation network for acoustic scene classification // *Proced. 2021 IEEE internat. conf. on Acoustics, Speech and Signal Processing (ICASSP - 2021)*. – Toronto, 2021. – P. 830-834.
- 51 Zhang Z. et al. Deep convolutional neural network with mixup for environmental sound classification // *Pattern Recognition and Computer Vision: proced. 1st Chinese conf. (PRCV 2018)*. – Guangzhou, 2018. – P. 356-367.
- 52 Tsalera E., Papadakis A., Samarakou M. Comparison of pre-trained cnns for audio classification using transfer learning // *Journal of Sensor and Actuator Networks*. – 2021. – Vol. 10, Issue 4. – P. 72-1-72-22.
- 53 Zhang R. et al. Transformer based unsupervised pre-training for acoustic representation learning // *Proced. 2021 IEEE internat. conf. on Acoustics, Speech and Signal Processing (ICASSP - 2021)*. – Toronto, 2021. – P. 6933-6937.
- 54 Piczak K. J. ESC: Dataset for environmental sound classification // *Proceed. of the 23rd ACM internat. conf. on Multimedia*. – Brisbane, 2015. – P. 1015-1018.
- 55 Koizumi Y. et al. Description and discussion on DCASE2020 challenge task2: Unsupervised anomalous sound detection for machine condition monitoring // *arXiv preprint arXiv:2006.05822*. 10.02.2023.
- 56 Cao H., Cooper D.G., Keutmann M.K., Gur R.C. et al. CREMA-D: Crowd-Sourced Emotional Multimodal Actors Dataset // *Transactions on Affective Computing (IEEE)*. – 2014. – Vol. 5, Issue 4. – P. 377-390.
- 57 Gemmeke J.F. et al. Audio Set: An ontology and human-labeled dataset for audio events // *Proced. internat. conf. on Acoustics, Speech and Signal Processing (ICASSP)*. – New Orleans, 2017. – P. 776-780.
- 58 Mesaros A., Heittola T., Virtanen T. TUT database for acoustic scene classification and sound event detection // *Proced. 24th European Signal Processing conf. (EUSIPCO)*. – Budapest, 2016. – P. 1128-1132.
- 59 Rachman F.H., Sarno R., Fatichah C. Music Emotion Classification based on Lyrics-Audio using Corpus based Emotion // *International Journal of Electrical & Computer Engineering*. – 2018. – Vol. 8, Issue 3. – P. 1720-1730.
- 60 Fonseca E., Favory X., Pons J. et al. FSD50K: An Open Dataset of Human-Labeled Sound Events // *Transactions on Audio, Speech, and Language Processing (IEEE/ACM)*. – 2022. – Vol. 30. – P. 829-852.
- 61 Zhang Y., Zeng J., Li Y. et al. Convolutional Neural Network-Gated Recurrent Unit Neural Network with Feature Fusion for Environmental Sound Classification // *Aut. Control Comp. Sci.* – 2021. – Vol. 55. – P. 311-318.

- 62 Heittola T., Mesaros A., Virtanen T. TAU Urban Acoustic Scenes 2020 Mobile, Evaluation dataset // <https://zenodo.org/record/3685828#>. 13.05.2019.
- 63 Heittola T., Mesaros A., Virtanen T. TAU Urban Acoustic Scenes 2020 3Class, Development dataset // <https://zenodo.org/record/3670185#>. 13.05.2019.
- 64 BBC Historical Sound Effects Library // <https://www.sound-ideas.com/Product/159/BBC-Historical-Sound-Effects-Library>. 13.05.2019.
- 65 MIREX MuSpeak Sample Dataset 2005 // <http://mirg.city.ac.uk/datasets/muspeak/>. 02.03.2022.
- 66 Venkatesh S., Moffat D., Miranda E.R. Investigating the Effects of Training Set Synthesis for Audio Segmentation of Radio Broadcast // *Electronics*. – 2021. – Vol. 10. – P. 827-1-827-21.
- 67 Mesaros A., Heittola T., Diment A. et al. DCASE 2017 challenge setup: Tasks, datasets and baseline system // *Proceed. of the Workshop on Detection and Classification of Acoustic Scenes and Events*. – Munich, 2017. – P. 1-9.
- 68 Park D.S., Chan W., Zhang Y. et al. Specaugment: A simple data augmentation method for automatic speech recognition // *Proceedings of the Interspeech*. – Graz, 2019. – P. 2613-2617.
- 69 Mesaros A., Heittola T., Virtanen T. Metrics for Polyphonic Sound Event Detection // *Appl. Sci*. – 2016. – Vol. 6. – P. 162-1-162-17.
- 70 Ferroni G. et al. Improving Sound Event Detection Metrics: Insights from DCASE 2020 // *Proceed. IEEE internat. conf. on Acoustics, Speech and Signal Processing (ICASSP - 2021)*. – Toronto, 2021. – P. 631-635.
- 71 Brownlee J. *Machine learning mastery with Python: understand your data, create accurate models, and work projects end-to-end*. – Vermont: Machine Learning Mastery, 2016. – 170 p.
- 72 Albon C. *Machine learning with python cookbook: Practical solutions from preprocessing to deep learning*. – Sebastopol: O'Reilly Media, Inc., 2018. – 366 p.
- 73 Kong Q., Xu Y., Wang W. et al. Sound Event Detection of Weakly Labelled Data With CNN-Transformer and Automatic Threshold Optimization // *Transactions on Audio, Speech, and Language Processing (IEEE/ACM)*. – 2020. – Vol. 28 – P. 2450-2460.
- 74 Bilen Ç., Ferroni G., Tuveri F. et al. A Framework for the Robust Evaluation of Sound Event Detection // *Proceed. IEEE internat. conf. on Acoustics, Speech and Signal Processing (ICASSP - 2020)*. – Barcelona, 2020. – P. 61-65.
- 75 Liu H. et al. Segment-level metric learning for few-shot bioacoustic event detection // *arXiv preprint arXiv:2207.07773*. 02.02.2023.
- 76 Audible Crime Scenes: ShotSpotter as Diagnostic, Policing, and Space-making Infrastructure // <https://journals.sagepub.com/doi/abs/10>. 02.02.2023.
- 77 Audioanalytic // <https://www.audioanalytic.com/about/index>. 02.02.2023.
- 78 Martín-Morató I., Mesaros A., Heittola T. et al. Sound Event Envelope Estimation in Polyphonic Mixtures // *Proceed. IEEE internat. conf. on Acoustics, Speech and Signal Processing (ICASSP - 2019)*. – Brighton, 2019. – P. 935-939.

- 79 Dinkel H., Wu M., Yu K. Towards Duration Robust Weakly Supervised Sound Event Detection // Transactions on Audio, Speech, and Language Processing (IEEE/ACM). – 2021. – Vol. 29. – P. 887-900.
- 80 Howard A.G. et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications // arXiv preprint arXiv:1704.04861. 20.02.2023.
- 81 Kwak J.-Y., Chung Y.-J. Sound Event Detection Using Derivative Features in Deep Neural Networks // Appl. Sci. – 2020. – Vol. 10. – P. 4911-1-4911-11.
- 82 Datasets // https://dcase-repo.github.io/dcase_datalist/datasets/list. 20.02.2023.
- 83 Kabir M., Mir J., Rascon C., Shahid M., Shaukat F. Machine learning inspired efficient acoustic gunshot detection and localization system // University of Wah Journal of Computer Science. – 2022. – Vol. 4, Issue1. – P. 1-17.
- 84 Rahman S., Khan A., Abbas S. et al. Hybrid system for automatic detection of gunshots in indoor environment // Multimedia Tools and Applications. – 2021. – Vol. 80. – P. 41413-4153.
- 85 Sigmund M., Hrabina M. Efficient feature set developed for acoustic gunshot detection in open space // Elektronika Ir Elektrotehnika. – 2021. – Vol. 27, Issue 4. – P. 62-68.
- 86 Gupta N., Deshpande P., Diaz J. et al. F-alert: Early fire detection using machine learning techniques // International Journal of Electronics Engineering and Applications. – 2021. – Vol. 9, Issue 3. – P. 34-43.
- 87 Benkendorf D.J., Hawkins C.P. Effects of sample size and network depth on a deep learning approach to species distribution modeling // Ecological Informatics. – 2020. – Vol. 60. – P. 101137.
- 88 Xu Y., Zhou Y. et al. Machine learning in construction: From shallow to deep learning // Developments in the built environment. – 2021. – Vol. 6. – P. 100045-1-100045-13.
- 89 Chollet F. Deep learning with Python. – NY.: Simon and Schuster, 2021. – 478 p.
- 90 Sudre C. H. et al. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations // Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: proced. 3rd internat. Workshop (DLMIA 2017), and 7th internat. Workshop (ML-CDS 2017). – Québec City, 2017. – P. 240-248.
- 91 Saritas M.M., Yasar A. Performance analysis of ANN and Naive Bayes classification algorithm for data classification // International journal of intelligent systems and applications in engineering. – 2019. – Vol. 7, Issue 2. – P. 88-91.
- 92 Harris C.R. et al. Array programming with NumPy // Nature. – 2020. – Vol. 585, Issue 7825. – P. 357-362.
- 93 Brownlee J. Deep learning with Python: develop deep learning models on Theano and TensorFlow using Keras. – Vermont: Machine Learning Mastery, 2016. – 245 p.
- 94 Raschka S., Mirjalili V. Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2. – Birmingham; Mumbai: Packt Publishing Ltd, 2019. – 772 p.

95 Baldi P., Sadowski P., Lu Z. Learning in the machine: Random backpropagation and the deep learning channel // Artificial intelligence. – 2018. – Vol. 260. – P. 1-35.

96 Shlezinger N., Eldar Y.C., Boyd S.P. Model-based deep learning: On the intersection of deep learning and optimization // IEEE Access. – 2022. – Vol. 10. – P. 115384-115398.

97 Yang L., Shami A. On hyperparameter optimization of machine learning algorithms: Theory and practice // Neurocomputing. – 2020. – Vol. 415, Issue 1. – P. 295-316.

98 Li H. et al. Research on overfitting of deep learning // Proceed. 15th internat. conf. on computational intelligence and security (CIS). – Macao, 2019. – P. 78-81.

99 Brownlee J. Better deep learning: train faster, reduce overfitting, and make better predictions. – Vermont: Machine Learning Mastery, 2018. – 575 p.

ҚОСЫМША А

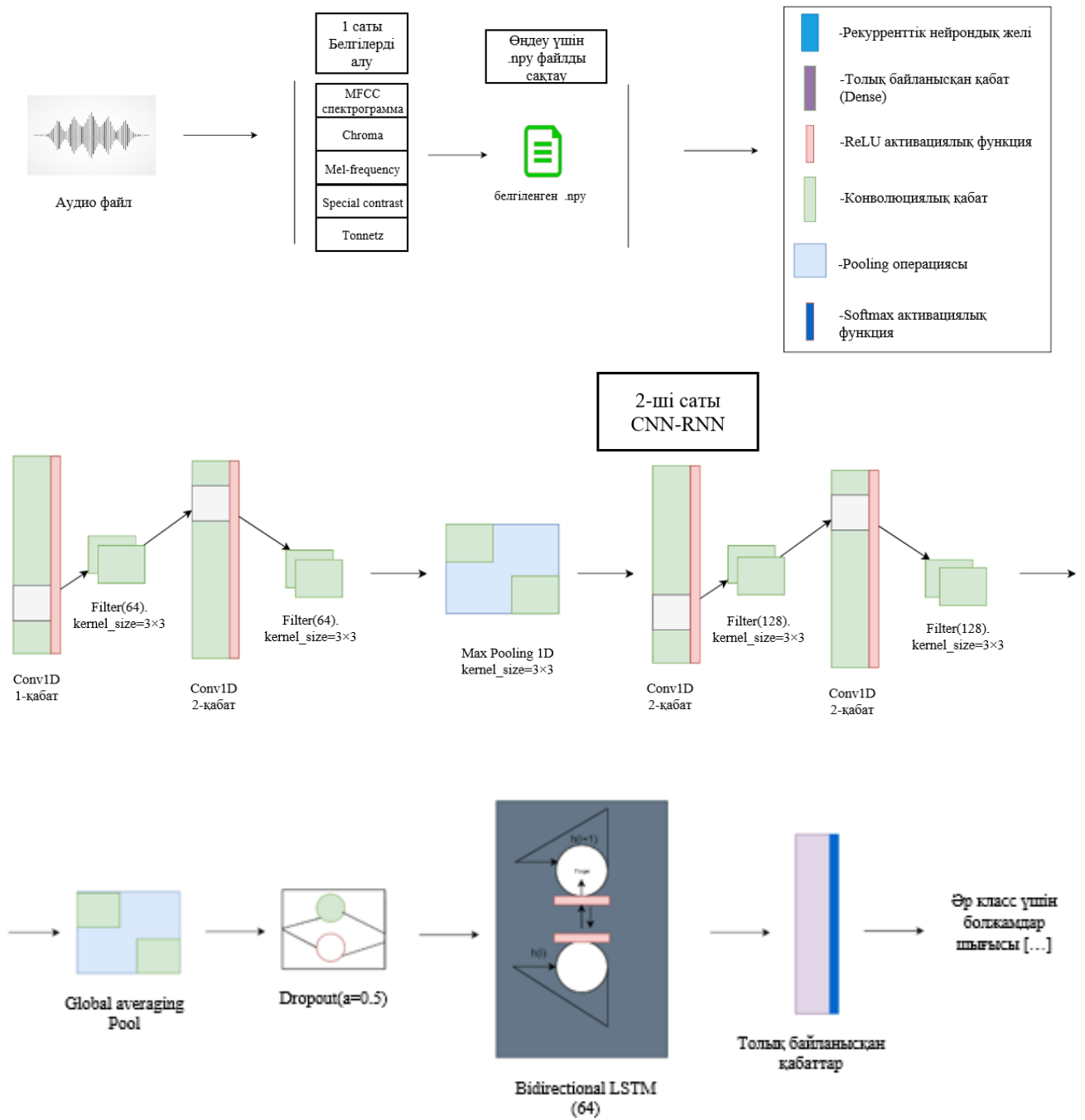
```
def load_files_and_save():
    features, labels = load_data('./urban_sound/Audio Dataset/');
    print("Shape:", labels.shape)
    np.save('./urban_sound/_audio_dataset_features.npy', features)
    np.save('./urban_sound/_audio_dataset_labels.npy', labels)
```

```
def get_features_from(filename=None):
    if filename:
        X, sample_rate = sf.read(filename, dtype='float32') #more accurate
        #X, sample_rate = librosa.load(filename, dtype='float32')
    else:
        print('Filename is not defined.')
    if X.ndim > 1: X = X[:,0]
    X = X.T #in case of stereo
    short_term_ft= np.abs(librosa.stft(X))
    mfccs_t = np.mean(librosa.feature.mfcc(y=X, sr=sample_rate, n_mfcc=40).T,axis=0)
    chromagram = np.mean(librosa.feature.chroma_stft(S=short_term_ft, sr=sample_rate).T,axis=0)
    mel_spect = np.mean(librosa.feature.melspectrogram(X, sr=sample_rate).T,axis=0)
    spectral_contrast = np.mean(librosa.feature.spectral_contrast(S=short_term_ft, sr=sample_rate).T,axis=0)
    tonnetz_harmonic = np.mean(librosa.feature.tonnetz(y=librosa.effects.harmonic(X), sr=sample_rate).T,axis=0)
    return mfccs_t,chromagram,mel_spect,spectral_contrast,tonnetz_harmonic
```

```
def load_data(parent_dir):
    #parent_dir = "./urban_sound/Audio Dataset/"
    print(len(os.listdir(parent_dir)))
    features = np.empty((0,193))
    labels = []
    labels_to_cat = []
    count_labels = 1
    sub_dirs = os.listdir(parent_dir)
    data_size = len(sub_dirs)
    for sub_dir in sub_dirs:
        full_path = parent_dir + sub_dir + '/'
        if os.path.isdir(full_path):
            date_for_start_parse = datetime.now()
            print('Started parsing dir:' + sub_dir + ' . Time started:' + date_for_start_parse.strftime("%d/%m/%Y %H:%M:%S"))
            for x in tqdm(os.listdir(full_path), position=1, leave=True ):
                file = full_path + os.path.basename(x)
                try:
                    mfccs_t, chromagram, mel_spect, spectral_contrast, tonnetz_harmonic = get_features_from(file)
                except Exception as exception:
                    print("File name:", file)
                    print("Exception:" , exception)
                    continue
                ext_features = np.hstack([mfccs_t, chromagram, mel_spect, spectral_contrast, tonnetz_harmonic])
                features = np.vstack([features, ext_features])
                labels_to_cat.append(count_labels) #for Label.npy
                count_labels += 1
            print("PARSE SUCCESS:", sub_dir)
    return np.array(features), np.array(labels_to_cat)
```

Сурет А.1 – Төтенше жағдайларды анықтауға арналған модельді оқытуда қолданылатын аудиомәліметтерден белгілерді алу

ҚОСЫМША Ә



Сурет Ә.1 – Ұсынылатын CNN-BiLSTM комбинацияланған архитектурасы

ҚОСЫМША Б

Нәтижелерді оқу үрдісіне ендіру актісі

Қазақстан Республикасы ПМ М.Есболатов атындағы Алматы академиясы оқу үрдісіне «Аудиосигналдар негізінде төтенше жағдайларды анықтау және классификациялауды зерттеу» тақырыбы бойынша Философия докторы (PhD) дәрежесін алуға арналған диссертациялық жұмысы бойынша авторефератын ендіру актісі

1. Жұмыстың тақырыбы: «Аудиосигналдар негізінде төтенше жағдайларды анықтау және классификациялауды зерттеу».

2. Шығу нәтижесінің түрі: Автореферат.

3. Жұмысты орындаушы(лар): Қ. И. Сәтбаев атындағы ҚазҰТЗУ, «Автоматика және ақпараттық технологиялар» институтының докторанты Досбаев Жандос Махсұтулы.

4. Жұмысты орындауға негіздеме: Философия докторы (PhD) дәрежесін алуға арналған диссертациялық жұмысты қорғауға.

5. Ендіру туралы мәлімет: Кәсіби-дайындық факультетінің бастығы, з.ғ.к., полиция полковнигі Д.В. Еремеев, Жоғарғы оқу орнынан кейінгі факультетінің бастығы, з.ғ.д., доцент, полиция полковнигі А.Т. Байсеитова, Оқу-әдістемелік орталығының бастығы, полиция полковнигі Д.А. Юлдашев «Аудиосигналдар негізінде төтенше жағдайларды анықтау және классификациялауды зерттеу» тақырыбы бойынша Философия докторы (PhD) дәрежесін алуға арналған диссертациялық жұмысы бойынша авторефераты оқу үрдісінде «Радиотехника, электроника және телекоммуникациялар» оқу бағдарламасы бойынша пәндерді оқыту кезінде қолданылатынын растайды.

6. Ендірудің тиімділігі туралы мәлімет: Қазақстан Республикасы ПМ М.Есболатов атындағы Алматы академиясы кәсіби дайындық факультеті киберқауіпсіздік және ақпараттық технологиялар кафедрасының оқу үрдісінде «Аудиосигналдар негізінде төтенше жағдайларды анықтау және классификациялауды зерттеу» тақырыбы бойынша Философия докторы (PhD) дәрежесін алуға арналған диссертациялық жұмысы бойынша авторефераты «Радиотехника, электроника және телекоммуникациялар» оқу бағдарламасы бойынша пәндерді оқыту кезінде қолданылады.

Қазақстан Республикасы ПМ
М.Есболатов атындағы
Алматы академиясы
бастығының орынбасары,
з.ғ.к., қауымдастырылған профессор
полиция полковнигі
« ____ » _____ 2023 ж.



Е. Бимолданов